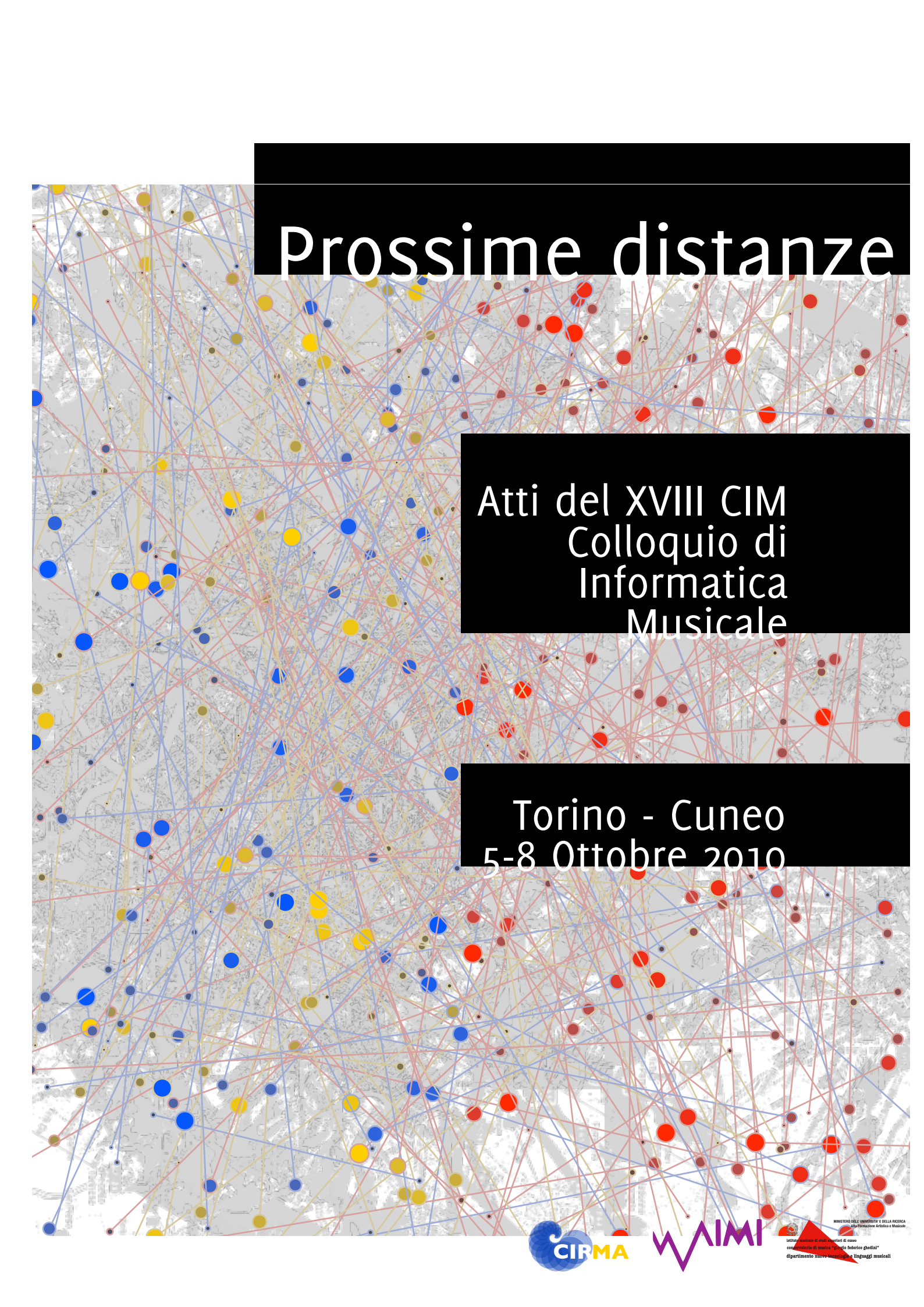




Prossime distanze

Atti del XVIII CIM
Colloquio di
Informatica
Musicale

Torino - Cuneo
5-8 Ottobre 2010



« Prossime distanze »

Atti del XVIII CIM - Colloquio di Informatica Musicale
Torino – Cuneo, 5–8 Ottobre 2010

A cura di Andrea Valle e Stefano Bassanese

XVIII CIM - Colloquio di Informatica Musicale
Torino – Cuneo, 5–8 Ottobre 2010

AIMI - Associazione Informatica Musicale Italiana

CIRMA - Centro Interdipartimentale di Ricerca su Multimedia e Audiovisivo, Università degli Studi di Torino

Dipartimento Nuove Tecnologie e Linguaggi Musicali, Conservatorio di Musica “Giorgio Federico Ghedini” di Cuneo

Con il supporto di



Atti del XVIII CIM - Colloquio di Informatica Musicale

A cura di Andrea Valle e Stefano Bassanese

ISBN: 9788890341311

sito: <http://www.aimi-musica.org/>

Riferimento BibTeX:

```
@proceedings{XVIII CIM,
  Editor = {Andrea Valle and Stefano Bassanese},
  Organization = {AIMI - Associazione Informatica Musicale Italiana},
  Publisher = {DADI - Dip. Arti e Design Industriale. Università IUAV di Venezia},
  Title = {Prossime distanze. Atti del XVIII CIM - Colloquio di Informatica Musicale},
  Year = {2011}}
```

Copyright

These proceedings, and all the papers included in it, are an open-access publication distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.



Comitato scientifico

Federico Avanzini, Università di Padova
Stefano Bassanese, Conservatorio di Cuneo
Nicola Bernardini, Conservatorio di Padova
Roberto Bresin, KTH, Stoccolma
Antonio Camurri, Università di Genova
Sergio Canazza, Università di Padova
Giovanni Cospito, Conservatorio di Como
Giovanni De Poli, Università di Padova
Riccardo Dapelo, Conservatorio di Genova
Agostino Di Scipio, Conservatorio di Napoli
Roberto Doati, Conservatorio di Genova
Federico Fontana, Università di Verona
Goffredo Haus, Università di Milano
Vincenzo Lombardo, Università di Torino
Michelangelo Lupone, Conservatorio de L'Aquila
Veniero Rizzardi, Conservatorio di Castelfranco Veneto
Davide Rocchesso, Università IUAV di Venezia
Sylviane Sapir, Conservatorio di Como
Augusto Sarti, Politecnico di Milano
Benjamin Thigpen, Conservatorio di Cuneo
Andrea Valle, Università di Torino
Alvise Vidolin, Università di Padova
Gualtiero Volpe, Università di Genova
Laura Zattra, Università di Padova

Keynote presentations

- | | | |
|------------------|----|--|
| James Harley | 5 | Iannis Xenakis: Electroacoustic music and polytopes |
| Mark Trayle | 11 | Remote Utopias - Illusions of Space and Community in Networked and Telematic Art |
| Joshua Parmenter | 15 | Composer's Toolkit: A Synthesis Of Musical Language |

Papers

1. Tecnologia, memoria e interpretazione

- | | | |
|---|----|---|
| Federico Avanzini, Sergio Canazza, Maddalena Novati, Antonio Rodà | 20 | A music bar for active listeners: an example of virtual electronic lutherie for a history 50 years long |
| Cat Hope | 27 | Interpreting old electronics: the reinterpretation of technology and concert techniques for electronic instruments in chamber music performance |
| Marco Marinoni | 32 | Anthèmes 2: un approccio monodirezionale al live electronics |
| Giuseppe Cabras, Sergio Canazza, Pier Luca Montessoro, Roberto Rinaldo | 57 | Audio documents restoration of ethnic music based on Non-negative Matrix Factorization and perceptual suppression rule |
| Gianmario Borio, Giacomo Albert, Nicola Bizzaro, Alessandro Bratus, Alessandro Cecchi, Maurizio Corbella, Matteo Giuglioli, Stefano Lombardi Vallauri, Marida Rizzuti, Federica Rovelli | 66 | Presentazione delle attività del gruppo di ricerca Worlds of AudioVision |

2. Gesto, interfaccia e controllo

Stefano Fumagalli, Saverio Monti	70	EJP – Electro Jamming Project
Carlo Massarelli, Andrea Valle	75	e-Zampognè – A Southern-Italian Bagpipe Controller
Pietro Polotti, Maurizio Goina	81	Per una rappresentazione audiovisiva del gesto
Stefano Delle Monache, Davide Rochesso	87	Experiencing Sonic Interaction Design: Product Design Activities at the SID Summer School 2010

3. Modellizzazione psicoacustica e spazio sonoro

Simone Spagnol, Michele Geronazzo, Federico Avanzini	92	Structural modeling of pinna-related transfer functions for 3-D sound rendering
Enrico Marchetto, Federico Avanzini	102	Modellazione fisica della glottide e inversione acustico-articolatoria
Sergio Canazza, Antonio Rodà, Daniele Salvati	107	A microphone array approach for browsable soundscapes
Mattia Schirosa, Jordi Janer, Stephan Kersten, Gerard Roma	115	A system for soundscape generation, composition and streaming

4. Distribuzione tra prossimità e distanza

Marco Gasperini	122	Theoretical foundations of a laptop orchestra
Federico Costanza, Marco Gasperini, Alessio Rossato	129	Meccanica/Azione/Sonora: composing networked music

5. Apprendimento e tecnologia

Serena Zanolla, Antonio Camurri, Sergio Canazza, Corrado Canepa, Antonio Rodà, Gualtiero Volpe, Gian Luca Foresti	134	Un ambiente didattico aumentato tecnologicamente
Tiziano Bole	143	Un videogioco per il training ritmico

- Nicola Montecchio, Nicola Orio **148** Visual and Aural Tools for Music Education Based on Audio to Score Alignment

6. Composizione assistita

- Massimo Avantaggiato **152** Composizione assistita e processi di trasferimento di dati musicali da PWGL a Csound
- Giorgio Klauer **160** Un sistema integrato di progettazione, composizione ed esecuzione dedicato a un brano per pianoforte e sintesi per modelli fisici real time
- Mauro Lanza, Gianluca Verlingieri, Nicola Biagioni **166** La libreria OpenMusic om4Csound - introduzione e progetto di documentazione
- Giorgio Bianchi, Pier Daniel Cornacchia **173** Composizione elettroacustica e tecnica waveset, un approccio creativo di ricostruzione algoritmica
- Sergio Canazza, Antonio Rodà, Lauro Snidaro, Ingrid Visentini **176** Ensemble methods for description of musical expressive intention

Posters

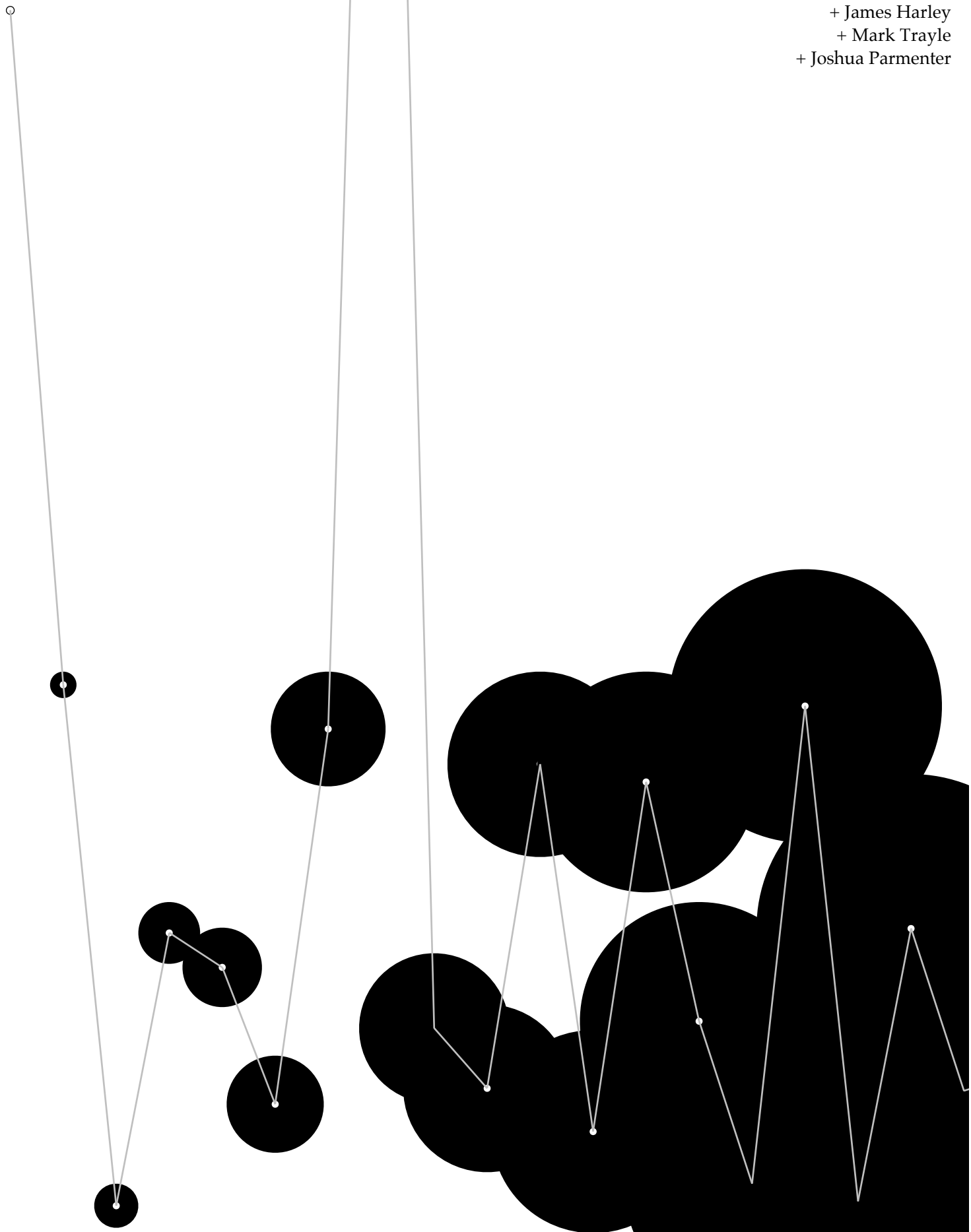
- Andrea Arcella, Agostino Di Scipio **180** Waveshaping and distortion synthesis, from Licklider (1946) to current perspectives
- Alberto Boem, Sergio Canazza, Antonio Rodà **186** Progettazione e realizzazione di un auditory display per la riabilitazione motoria
- Federico Costanza, Marco Gasperini, Alessio Rossato **189** S. Giorgio laptop ensemble
- Claudio Dionisi, Eliana Carboni **195** Sintesi sonora e grafica frattale

Amedeo Casella, Andrea Valle	199	Simulation of a soundscape using an analysis/resynthesis model
Ivano Morrone	203	The Ringer – software audio e video in tempo reale
Antonio Scarcia, Francesco Abbrescia, Francesco Scagliola	206	Un tutorial per la produzione di opere audio-visuali tramite ambienti integrati
Stefano Silvestri	209	Introduzione alla sintesi Wavetable Switching per Multiplexing di segnali

Keynote presentations

+ James Harley
+ Mark Trayle
+ Joshua Parmenter

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
[
\
]
>
_
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z



IANNIS XENAKIS: ELECTROACOUSTIC MUSIC AND THE POLYTOPES

James Harley
University of Guelph
jharley@uoguelph.ca

ABSTRACT

Iannis Xenakis composed an important series of electroacoustic works over the course of his career. These compositions are widely known through recordings and concert presentations. What is less well known is that many of these works were conceived for specific situations or concerns involving multimedia, architecture, spatialization, or algorithmic programming. This article presents these compositions within the contexts for which they were conceived with the aim of shedding light on the wider concerns that also engaged Xenakis and took his work beyond the purely musical domain.

1. INTRODUCTION

Iannis Xenakis (1922-2001) composed approximately 150 works over the course of 45 years of activity. Of those, just 17 are electroacoustic compositions. This relative handful of works has proven to be highly influential, and the music was created throughout his career, from *Diamorphoses* in 1957 to *S.709* in 1994 (a later work, *Erod*, from 1996, was withdrawn from the catalogue soon after its premiere). Xenakis composed his electroacoustic works using the means available to him at the time, from the *musique concrète* pieces produced at the Groupe de Recherches Musicales (GRM) to those produced to invitation at independent studios, to the digital works produced at the Centre d'Études Mathématique et Automatique Musicales (CEMAMu). [1]

A perhaps surprising aspect of Xenakis's electroacoustic output is that so much of it was created for multimedia presentation in one form or another, presentations he went on to define as "polytopes" (this term more precisely refers to those works created out of a unified formal conception relying on multi-modal perceptual capacity). Even those works not conceived for a multimedia context (which would include film, dance, and theatre) were very often concerned with issues going beyond the sounds, including spatialization and formal organizational principles (compositional algorithms).

2. INFLUENCES

Xenakis did not undertake traditional compositional training until he was an adult. While he did participate in musical activities, he in fact studied civil engineering, graduating from the Athens Polytechnic in 1947 (after many interruptions due to war and revolution). [2] His

scientific training, which emphasized mathematical skills above all, was achieved at the same time as Xenakis was heavily involved in the resistance in Greece. The sights and sounds of war, expressions of intense emotional distress, proved to be the touchstones of this artist's creative aesthetic, tempered by the engineer's capacity for abstraction and concern for materials.

When Xenakis arrived in Paris in late 1947, an exile from his homeland, he quickly found work in the architecture studio of Le Corbusier. One of the projects he worked on was the Philips Pavilion for the 1958 World Fair in Brussels, in which Le Corbusier collaborated with Edgard Varèse and others on a multimedia presentation, *Le Poème électronique*. Xenakis also began to pursue his path toward music, eventually receiving encouragement from Olivier Messiaen, whose course on analysis he attended as an auditor in the early 1950s. This connection led to his acceptance by Pierre Schaeffer into the studios of GRM. While he did not follow Schaeffer's directives regarding the taxonomy of sound and the "objet sonore," he nonetheless benefited from the opportunity to experiment with all manner of studio technology and to have contact with the other musicians and technicians associated with GRM, including Pierre Henry, Luc Ferrari, François Bayle, and especially François-Bernard Mâche, who would become a lifelong friend.

3. EARLY WORKS

3.1 *Diamorphoses*

Xenakis completed his first electroacoustic composition at GRM in 1957. *Diamorphoses* is in some ways a classic piece of *musique concrète*, being created from a variety of recorded sounds manipulated in the studio. Where it diverges from the norm is in the concern for continuities and densities. The sound sources are layered to create ongoing sonorities that evolve gradually, individual sounds submerged in thicker composite textures.

3.2 Philips Pavilion and *Concret PH*

At the same time as he was carrying out work at GRM Xenakis was involved in the design and construction of the Philips Pavilion, serving as project manager for Le Corbusier, who was during that time heavily implicated

in a much larger project for the city of Chandigarh, India. [3] As mentioned above, the pavilion hosted a multimedia presentation produced by Le Corbusier and Varèse, *Le poème électronique*, a ‘spectacle’ that included film sequences and lights projected onto the hyperparaboloid curved surfaces of the edifice, suspended sculptural symbols, and electroacoustic music routed through several hundred loudspeakers installed throughout the pavilion. In 1958, this show would have been quite unique, and many tens of thousands of visitors experienced it. As an interlude between showings, a short electroacoustic work by Xenakis was performed: *Concret PH*. This piece contains layered recordings of the crackling of burning charcoal. The density of the overall texture evolves imperceptibly, giving the effect of a natural process even though the piece was carefully crafted in the studio. The high-frequency energy of the attack transients of the crackling sounds ensures that the movement of the sounds through the loudspeakers of the pavilion would have been clearly perceived by listeners.

While Xenakis was not directly involved in the creation of *Le poème électronique*, aside from the contribution of *Concret PH*, the production showed him both what was possible, in terms of combining architecture with sound and light, and what did not fit with his own aesthetic sensibility (particularly the representational elements present in the film sequence).

3.3 Notes sur un geste électronique

In 1958, the same year as the Philips Pavilion, Xenakis wrote an article that coalesced his thoughts on the new possibilities of multimedia creation: “Notes sur un geste électronique.” Here, he makes a number of points that would guide his work throughout the course of his life:

1. Abstraction: conscious manipulation of laws and pure ideas, and not of concrete objects.
2. The play of forms and colors, once separated from their physical context, implies conceptual networks of a higher level.
3. Continuous variation may apply to all sound components, ... and is a complementary aspect of its temporal existence.
4. Cinematic Stereophonics: sound that can move over a line of loudspeakers.
5. Architectural form: exerts influence on the acoustic quality of the space. [3]

It would be several years before Xenakis would have the opportunity to put his concept of the “polytope” into practice.

3.4 Intervening years

Xenakis worked on a number of projects through the auspices of GRM. In the years 1960-61, he produced music for three documentary films: *Orient-Occident*; *Neg-Ale*; and *Formes rouges*. The latter two were soundtracks of instrumental music and were subsequently withdrawn from his catalogue (*Neg-Ale* has been released on DVD

by Mode Records in 2008). *Orient-Occident*, though, was electroacoustic, and Xenakis went back into the studio to produce a concert work from the soundtrack. [4] Although originally intended as music for film, this work fits closely to the GRM style, utilizing a variety of discrete sources, shaped in the studio.

Prior to embarking on these more commercial projects, Xenakis produced an experimental work, *Analogique B* (1959). This piece is built from electronic “grains,” organized by means of Markov chains (linked probabilities), an early attempt to implement granular synthesis. [5] In effect, *Analogique B* is a more calculated, electronic version of *Concret PH*, with its single sound type, short crackles of burning charcoal. It is also intended to be interleaved with a companion work, *Analogique A* (1958), a score for strings that is also built from short discrete sounds, bowed or plucked sounds.

It is worth noting that during this period between the Philips Pavilion and the *Polytope de Montréal*, Xenakis entered the world of stage as well as film. The first music for theatre was *Hiketides (The Suppliants)* from 1964, a work by Aeschylus presented at the ancient amphi theatre in Epidauros (Xenakis, being still a condemned exile, was not able to attend the premiere). The music is instrumental, with sung choruses (dropped for the instrumental concert suite he produced), the singers also being called to play hand percussion to create massed “granular” textures undoubtedly derived from the composer’s electroacoustic work. Xenakis’s collaboration with Greek stage director Alexis Solomos was reprised a few years later for *Oresteia* (1966), produced for a festival in Ypsilanti, Michigan. Again, the music is instrumental and choral, and Xenakis derived a concert suite from the stage music, as before. This score has seen numerous performances over the years, and Xenakis contributed two more parts to it later.

3.5 Spatialization experiments

The final electroacoustic work Xenakis produced at GRM was *Bohor* (1962). Like *Diamorphoses*, and even more so, this work explores continuities and densities. In addition, *Bohor* engages spatialization independently of the architectural space within which the music is intended to be heard. This work is for eight tracks, each diffused through its own loudspeaker (or speakers) surrounding the audience. In 1962, GRM did not possess an eight-track tape recorder, so *Bohor* was presented on multiple recorders. Precise synchronization would have been difficult if not impossible, and in this case, the music does not require it, the dense, noisy sonorities evolving gradually with no sudden shifts or breaks.

Xenakis’s interest in the spatialization of sound extended to the realm of instrumental music as well. In *Eonta* (1963), for five brass and piano, the brass players are required to move to different positions onstage as well as to point their instruments in different directions. In *Terretektorh* (1966) and *Nomos gamma* (1969), both for orchestra, the individual players of the orchestra are seated amongst the audience so that the music sounds

from all directions, the placement of the sounds being as important an element of the composition as anything else.

4. THE POLYTOPES AND RELATED WORKS

4.1 *Polytope de Montréal*

In 1967, Xenakis was able to return to multimedia work as envisaged back in 1958, producing his first polytope. In this case, he was invited to create an installation for the French Pavilion (designed by architect Jean Fautouren) at the 1967 World Expo in Montreal. The installation consisted of five sets of steel cables hung on angled vertical paths through the large, five-story atrium space of the pavilion, creating interlocking curved hyperparaboloid outlines. On these cables were attached several hundred flashbulbs that could be switched on and off independently. Xenakis created a lightshow lasting approximately six minutes, stored as programmed instructions controlling photocells. He also created music to run in parallel to the lights. While the music was intended to be played from loudspeakers located in various places throughout the atrium, he ended up creating an instrumental score for four identical ensembles, recorded for presentation at the pavilion but which could also be performed independently. The music is thus not strictly speaking electroacoustic, but the focus on unusual, at times noise-like, sonorities and the panning of amplitude fluctuations (dynamic swells) around the four ensembles, makes this music resemble a studio work. Altogether, the polytope, with four tracks of music filling the huge atrium of the pavilion, and bright flashbulbs assaulting the eyes while unfolding both linear and stochastic patterns, set onto a backdrop of sculptural form delineated by dozens of steel cables, was striking and intense. As with the Philips Pavilion in 1958, tens, probably hundreds, of thousands of people experienced the six minutes of *Polytope de Montréal* that year.

Xenakis would have to wait another 11 years to be able to add his own architectural design to the lights and sounds of his polytope conception. In the meantime, the Montreal installation remained in place for 25 years or so, as the French Pavilion remained standing, one of just two structures from Expo 67 that was not dismantled.

4.2 Music for Dance

The following year, Xenakis received a major commission from the National Ballet of Canada to create a full-length ballet. He worked with French choreographer Roland Petit, Hungarian designer Victor Vasarely, American conductor Lukas Foss, and Romanian composer-conductor Marius Constant. *Kraanerg* was intended to celebrate the gala opening of the National Arts Centre in Ottawa, designed by Montreal architect Fred Lebensold. In addition to being Xenakis's first dance work, it was also his first mixed composition, scored for large ensemble and four-channel tape. Carrying on from *Polytope d Montréal*, the tape part for *Kraanerg* is made up of instru-

mental sounds entirely, built from recordings of the same ensemble as the score, although Xenakis treated the ensemble recordings extensively, using filtering, reverberation, and distortion techniques, primarily, to extend the sounds into the realm of electroacoustics. The choreographer gave the composer free reign to compose the music how he saw fit, creating the dance elements after the music was complete and recorded. One of the most important characteristics of the piece is the alternation of live ensemble and recorded material, with occasional superpositions. The spatialization of the recorded sounds through four channels surrounding the audience is also a striking element of the piece. [6]

The premiere of *Kraanerg* created something of a stir in the dance world, and Xenakis was soon after approached by George Balanchine of New York City Ballet to compose another ballet (Balanchine had already created a choreography to Xenakis's two early orchestral scores, *Metastaseis* and *Pithoprakta*). *Antikhthon* was completed in 1971, a 25-minute score for full orchestra. Balanchine never set the music to dance, but it has seen life as an orchestral concert work. Xenakis never completed another dance work, but did preliminary work on a ballet for robots in the late 1980s that was abandoned for lack of funding.

4.3 *Hibiki-Hana-Ma* (1970)

Xenakis continued his exploration of spatialization in both electroacoustic and instrumental compositions. *Persephassa* (1969), for six percussionists surrounding the audience, was premiered at the Shiraz Festival in Iran, and the success of this performance led to a commission for a major multimedia work there two years later. First, though, Xenakis traveled to Japan to produce an electroacoustic work for the Osaka World Fair. *Hibiki-Hana-Ma* (1970), like the tape part for *Kraanerg*, was produced from instrumental recordings, this time involving both symphony orchestra and traditional Japanese instruments, produced at the NHK studios in Tokyo. Like the earlier piece, the recorded sounds are treated in the studio, in similar ways, but this time mixed onto 12 channels rather than four. The pavilion (Japanese Steelworkers Federation) contained something like 250 loudspeakers, with the individual channels able to be routed through the pavilion independently. Xenakis was not involved in designing the pavilion or the sound system, but, while in Osaka for the World Fair that year, he learned about laser projections, an element he would integrate into his polytopes in the future.

4.4 *Persepolis* (1971)

The presentation of *Persepolis* at the 1971 Shira Festival was intended to be site-specific, and to pay homage to the historic location. The eight-channel electroacoustic music was the starting point, diffused over several loudspeakers placed throughout the site. To the music, lasting close to an hour, was added spotlights, torch-lit processions, and bonfires in the distance. The patterns created in the sky

by the (92) spotlights were enhanced by two lasers, the beams also pointed to the sky and set into motion manually. The music represents a shift away from manipulated instrumental sounds toward a richer sound world. There are instrumental sounds, but they are extended sounds (such as clarinet multiphonics) that are extensively treated and layered. The density and noisy complexity of the composite textures draw attention away from the sources toward the slow unfolding of the temporal architecture of the work over its extended duration.

The performance in Iran required the marshalling of large forces, something possible there with the Empress, as principal sponsor of the festival, able to procure any and all resources required. In his next multimedia project, automation would be a necessity.

4.5 *Polytope de Cluny* (1972)

The Festival d'Automne in Paris, launched in 1972 under the direction of Michel Guy, initially approached Xenakis about commissioning an opera. Xenakis instead proposed a polytope. The *Polytope de Cluny* opened in the fall of 1972 in the historic vaulted space (a former Roman baths) of Cluny in the heart of Paris. The installation could do no damage to the ancient site, so instead of sculptural cables to affix flashbulbs to, as in *Polytope de Montréal*, scaffolding was erected instead. In addition to the hundreds of flashbulbs, Xenakis added a number of small controllable mirrors that provided paths for three laser beams (coloured red, green, and blue). A great deal of effort was expended developing digital control instructions for the lights and lasers. The seven channels of electroacoustic music was routed through 12 loudspeakers, and the control signals for the visuals as well as the sound spatialization were stored on the eighth track of the tape. The installation was thus able to be presented with minimal supervision. As it turned out, *Polytope de Cluny* ran for two seasons, with many tens of thousands of people turning out to experience the 22-minute multi-sensorial event.

The music for this work draws on material used in *Persepolis*, and adds new studio-produced sonorities, including Xenakis's first digitally synthesized sounds. He had long foreseen the possibility of applying stochastic generative procedures (that he had been using to create instrumental music since the 1950s) to synthesis. The recent acquisition of a then-rare digital-to-analog converter by his newly-created research centre, EMAMu (which would be renamed CEMAMu), enabled this early integration of computer-generated sounds with the studio-produced analog sounds.

4.6 *Le Diatope—La légende d'Er* (1978)

In 1974, the military junta that had controlled Greece since 1947 finally fell, the death sentence that had been outstanding against Xenakis was lifted and he was able to return to his homeland. Plans began to be set in place for a large multimedia event at the historic site of Mycenae, which would be unveiled in the summer of 1978. Prior to

that (or, concurrently with the planning for the event in Greece), Xenakis had been commissioned to create another 'spectacle' like *Polytope de Cluny* for the inauguration of the Centre Pompidou in Paris. At long last, 20 years after the Philips Pavilion, Xenakis was able to not only create the multimedia event itself but the architectural space within which it would be presented.

Le Diatope consisted of a small pavilion, rather like a tent with its red nylon-type covering stretched over a frame, the curvilinear shape bearing a familial resemblance to the 1958 structure, but designed to optimize the acoustical properties of the interior. A network of cables attached to the structural frame of the pavilion anchored the flashbulbs and movable mirrors in the same fashion as the Cluny installation. A new feature for *Le Diatope* was the installation of floor lights that would shine up through the audience members, enhancing the immersive quality of experience. The sound was again mixed onto seven tracks and diffused in a similar way to the earlier work. There are sonic materials adapted from both *Polytope de Cluny* and *Persepolis*, but new ones are added, both analog and computer-generated. The electroacoustic music portion of *Le Diatope* was actually commissioned by WDR in Cologne, and Xenakis worked in the studios there to create the music early in 1978. *La légende d'Er* was first performed at a concert in Cologne in an eight-channel version, but a seven-channel mix was produced in order to save the eighth channel of the tape for control instructions for the automated installation, as for *Polytope de Cluny*. It is this version of the piece that has been subsequently used for concert presentations and recordings.

The final element that Xenakis added to *Le Diatope* was an extensive program booklet, which included, uniquely in his output, a set of five thematically related texts (reflections on the cosmos) by different writers taken from across the centuries: Plato (who describes in the closing section of *The Republic* a vision of the afterlife from which the title of the composition is taken), Hermes Trismegistus, Blaise Pascal, Jean-Paul Richter (text also used later in *Nekuia*, a large work for choir and orchestra from 1981), and Robert P. Kirshner (an article on supernovas published in *Scientific American*). These programmatic elements add another dimension to the work, and may help lead the audience's perception of the work as an abstracted representation of war (loud noisy sounds, lights signaling explosions, lasers tracing missile paths across the sky) to a representation of meteors, stars, and galaxies.

Le Diatope was created with the aim of being portable. After the installation outside of the Centre Pompidou through the summer of 1978, it was moved to Bonn, but has languished since.

4.7 *Polytope de Mycènes* (1978) and the UPIC

With *Le Diatope* up and running in Paris, Xenakis turned to Mycenae. The *Polytope de Mycènes* turned out to resemble *Perspolis* rather than his more recent polytopes, no doubt due to the ancient, outdoor location. The searchlights, lasers, torch-lit processions, and bonfires were en-

hanced with goats equipped with lights and bells, and amplified musicians performing alongside computer-generated electronic music. This event, which ran for four days, was witnessed by several thousand people per day, and became a major cultural event in Greece and a triumphant return to his homeland for Xenakis. By contrast, *Persepolis*, while an impressive event, was isolated from mainstream culture in Iran and was aimed at a select audience of international dignitaries and invitees.

The program for *Polytope de Mycènes* included a number of Xenakis's vocal works set to ancient Greek texts: *À Hélène* (1977), *À Colonne* (1977), and *Oresteia* (1966), this latter work also including an instrumental ensemble. Along with recitations of passages from Homer's *Iliad*, tying the event in a direct way to the location (Mycenae being home to Agamemnon), two of the composer's percussion works were presented: the spatialized *Persephassa* for six players, and *Psappha* (1976), a recent score for solo percussion. The new work introduced as part of this event was *Mycenae Alpha*, a relatively short electronic work that was played seven times as an interlude between the other elements involving live performers. This piece marks the launch of a new stage in Xenakis's electroacoustic output. By 1978, he and his engineers at CEMAMu had developed a working version of what Xenakis called the UPIC (Unité Polygogique Informatique de CEMAMu), a computer system for creating sound with a powerful graphic interface. Xenakis created a striking series of graphic designs that were then converted to sound and woven together to produce the work. This new technology was virtually unique at the time, and represented the culmination of a vision Xenakis had held since he first began creating designs for music in the 1950s, such as the complex string glissandi in *Metastaseis*. In a sense, Xenakis's music has always been "multimedia," as so many of his scores are based on graphic sketches.

Mycenae Alpha is also Xenakis's first electroacoustic work that is entirely built from electronic sounds. There are no common materials between this piece and any of the earlier ones. As such, it represents a striking shift of aesthetic aim and compositional strategy.

5. THE OTHER UPIC WORKS

Xenakis never returned to either the classic analog studio or to the sonic materials he had developed in his earlier electroacoustic compositions. Instead, he focused on the UPIC and on a new algorithmic synthesis method.

The next studio composition Xenakis produced was *Pour la Paix* (1981), a radiophonic work utilizing speaking voices, choir, and electronic episodes and supportive material created with the UPIC.

After that, there was a gap of six years. The next UPIC work came in 1987, created as part of a multimedia event (to be discussed below). After that came a final work, *Voyage absolu des Unari vers Andromède* (1989), a commission from Japan for the inauguration of the International Exposition of Paper Kites. This 15-minute work has a programmatic element (interstellar voyage), but the sonorities and temporal shape are the primary concerns, with no lights or other visuals, no text (beyond the evoc-

ative title), and no specific spatialization (all the UPIC works are monophonic or stereo). Furthermore, there were no published graphic designs for *Voyage absolu*, making it as pure an electroacoustic work as any in the composer's output (looking back to *Diamorphoses* from 1957).

5.1 *Tauriphanie* (1987)—the final 'polytope'

In 1987, Xenakis was invited to create a 'spectacle' for a summer festival based in Montpellier in the south of France. For the site, he was given the ancient Roman stadium in Arles (still used for bullfights and other such events), where spotlights were installed to create patterns of light on the floor of the ring, along with performance platforms around the seating area for 12 percussionists to perform excerpts from Xenakis's three works for percussion ensemble: *Persephassa*, *Pleiades* (1978), and *Idmen B* (1985). A control tower was constructed in the middle of the ring and the latest version of the UPIC, capable of real-time playback and manipulation of sound, was installed there. In addition, the ring was graced with the presence of some of the bulls ordinarily engaged for bullfights, alternating with a group of the famed white horses of the nearby Camargue region. The bulls were originally to be equipped with contact radio mics, their snorting and other sounds to be amplified and diffused over the sound system with the UPIC interacting with them by playing back sampled or created sounds loaded into the system's memory. As it turned out, the bulls could not be amplified, so Xenakis's pre-produced work, *Taurhiphanie*, was played back with interjections or interventions by the composer and his technicians manipulating the UPIC in real time.

Taurhiphanie turned out to be Xenakis's last polytope-type project. While there were discussions and plans put forward for other events (in Athens, Mexico City, and perhaps elsewhere), the funding never came through and the projects never came to fruition.

6. DYNAMIC STOCHASTIC SYNTHESIS

The final phase of Xenakis's electroacoustic work came about in the early 1990s. With access to a powerful mini-computer capable of running a modern programming language such as C, along with a custom-built sound card, he and his engineers at CEMAMu were able to develop software that would implement not only the algorithmic composition programs he first implemented in 1962 (his ST algorithm, originally written in FORTRAN), but an extension to that work to carry the algorithm down to the level of digital waveforms so that sound could be generated directly (rather than be transcribed into a musical score and performed by instruments as Xenakis had to do back in 1962). He called this approach 'dynamic stochastic synthesis,' [7] and he produced two works using this method: *GENDY3* in 1991, and *S.709* in 1994. As with the UPIC pieces, these electroacoustic compositions contain only computer-generated sounds; they invoke no spatialization or other multimedia components. The primary aim was to fully implement algorithms that would gener-

ate not only the musical structure but the sounds themselves.

7. CONCLUSION

Over the course of close to 40 years, Xenakis produced a series of electroacoustic works that for the most part continue to be performed and have all been issued on recording. Most of these works were originally conceived as part of a larger experience than the music itself, involving sophisticated spatialization, multimedia, graphic design, or algorithmic organization. What is perhaps especially significant is that the music has continued to have a life beyond the specific contexts or concerns of the works' creation. This fact speaks to the power and originality of Xenakis's creative force.

8. REFERENCES

- [1] J. Harley: "The Electroacoustic Music of Iannis Xenakis," *Computer Music Journal*, Vol. 26, No. 1, pp. 33-57, 2002.
- [2] N. Matossian: *Xenakis*, Nicosia, Cyprus: Moufflon Publications, 2005.
- [3] I. Xenakis: *Music and Architecture*, Pendragon Press, Hillsdale, New York, 2008.
- [4] M. Solomos: "Orient-Occident: From the film version to the concert version," *Iannis Xenakis: Das elektroakustische Werk*, pp. 118-131.
- [5] A. Di Scipio: "Formalization and Intuition in *Analogique A et B*," *Definitive Proceedings of the "International Symposium Iannis Xenakis" (Athens, May 2005)*, <<http://www.iannis-xenakis.org/Articles/Di%20Scipio.pdf>>, 2006.
- [6] J. Harley: *Kraanerg*, Farnham, Surrey, Ashgate Publishing, forthcoming.
- [7] I. Xenakis: *Formalized Music*, revised edition, Pendragon Press, Hillsdale, New York, 1992.

REMOTE UTOPIAS / PHANTOM ROOMS

Mark Trayle

California Institute of The Arts
mtrayle@calarts.edu

ABSTRACT

Electronic networks have been a useful medium for artistic production and collaboration. Besides their function as a conduit for the bits, bytes, and waves of artistic output, networks provide engaging metaphors that shape the art forms that use them. The relatively brief

history of telematic art is knit together with the metaphor of 'cyberspace', a utopian vision of global communications that connects distant people by shrinking space and time. Telematic music projects have adopted this metaphor, creating musical performances by connecting musicians in different locales. But does collapsing spatially distant and acoustically distinct spaces into a supposedly unified whole add anything to the *sonic* experience of the music? Does the inclusion of the now standard video link privilege the visual over the sonic? The composer/author explores these questions in a composition for networked laptops, *Phantom Rooms*.

1. INTRODUCTION

I'd like to talk about the nature of digital networks, the way they're used and applied and extended in my work and in the work of some other musicians, and along the way touch on some of the history of networked art and music and the metaphors we use when talking about these forms. I'll briefly explore two ideas that have helped shape our perceptions of the internet and so are woven into networked music: space and community.

Searching for a mental picture of the Internet, we borrow and modify geographical terms to describe networks... "cyberspace", "topology", "sites"... and travel across the http-scape on the "information superhighway". Maps of the Internet vary according to the needs of their makers, but most ascribe a vague physicality to an infrastructure that has a tenuous grip on the material world. Images of major network nodes branching to smaller nodes spanning (most of) the globe with straight lines evoke an industrial hierarchy of pipelines, roadways, and distribution centers. The seemingly limitless capacity of the network to store text, images, and sound evokes the feeling of a vast space.

Sited on the network are virtual *communities* such as The Well, social networking sites like Facebook and MySpace, multi-user games and blogs. Prior to these colorful, graphically high-fidelity sites, text-based virtual meeting points in the form of newsgroups and chatrooms tapped into a pioneering idealism at the dawn of the Internet Revolution. These social networks were previewed in the early days of radio. In its origins as a point-to-point

communications channel between amateurs, where every station was a transmitter as well as a receiver, radio was driven by utopian impulses and do-it-yourself technology. Hugo Gernsback's seminal "Radio For All" from 1922 is not only a very detailed manual for do-it-yourself radio enthusiasts, it predicts a futuristic paradise where everything including roller skates is enhanced by radio.

Some years later the formation of a global telephonic network not only allowed people around the world to talk to each other directly, it enabled artists to share images, text, video, and audio. Beginning in the late 1970s artists such as Bill Bartlett, Liza Bear, Robert Adrian, and Roy Ascott began creating collaborative works sited on telephonic networks, using primarily images and text. Some of the pioneers who participated in these early days of telematic art explored the nature of community as the body of the work itself. In this article from 1984, *Art and Telematics: towards a network consciousness*, Roy Ascott writes:

"...the transformations of "creative data", are in perpetual motion, an unending process. In this sense art itself becomes not a discrete set of entities, but rather a web of relationships between ideas and images in constant flux, to which no single authorship is attributable and whose meanings depend on the active participation of whoever enters the network. In a sense there is one wholeness, the flow of the network in which every idea is a part of every other idea, in which every participant reflects every other participant in the whole. This grand reciprocity, this symmetry of sender and receiver is such that a mirror image is exchanged in which sender is receiver and receiver sender." [1]

Earlier, in 1966, Max Neuhaus employed the telephone network and radio station WBAI in New York in the project Public Supply 1. Neuhaus mixed incoming telephone calls, radio broadcasts, and feedback to create a sonic portrait of a community of listeners.

"...it seems that what these works are really about is proposing to reinstate a kind of music which we have forgotten about and which is perhaps the original impulse for music in man: not making a musical product to be listened to, but forming a dialogue, a dialogue without language, a sound dialogue." [7]

Later, in 1977, Neuhaus expanded on Public Supply 1 with Radio Net. Still sited on radio and telephone networks, Neuhaus used the National Public Radio Corpora-

tion loop to create a spatially enormous feedback and signal processing system.

"It created a sound-transformation <box> that was literally fifteen hundred miles wide by three thousand miles long." [7] (pp. 25-26)

Connectivity and collaboration were key elements of early telematic art and music. As the number of participants in telematic art increased worldwide, artists began to create network infrastructures of their own. Robert Adrian, with the help of the I.P. Sharp Company, developed the Artex computer network in 1980. Later, Carl Loeffler and Fred Truck developed the Art Com Electronic Network, enabling artists (and the public) to communicate in a text-based environment. These digital networks allowed for a decentralization of transmission and reception, promoting "... an awareness of the medium as a new kind of space in which to work..." [6].

Geographer Michael Curry terms these two artistic/cultural imperatives cyberspace and cyberplace. **Cyberspace** as the network, an extension of the Cartesian/Newtonian model of space; **cyberplace** as the network nodes created by social and cognitive interaction [3]. Curry's analysis is in harmony with Marshall McLuhan's ideas of visual vs. acoustic space. Cyberspace as visual space, the highly linear space of cognition and causality molded by the phonetic alphabet and perspective, and cyberplace as acoustic space, a context for simultaneity, nonlinearity, and resonance.

While the metaphors of cyberspace and cyberplace, visual and acoustic space, appear to work smoothly in the realm of telematic *art*, they collide in the nomadic wanderings of networked and telematic *music*.

2. TOWARD A CRITIQUE OF TELEMATIC MUSIC

Most projects that we would now identify as "network music" are, or have historically been, tele-collaborative projects sited on telephonic or digital networks: telematic music. These projects share a number of features. At each location the sonic space is constructed from local and remote sources. These sound sources are mixed together to form a unified sound field. These projects often include some kind of visual component that shows what's happening in the remote performance space. The visual component provides a kind of visual validation or confirmation of the telematic nature of the piece and helps create the illusion of a coherent, continuous, performance space. These elements have been features of telematic music performances since at least the early 1990s.

During the 1990s the Electronic Café International in Santa Monica, California, hosted a number of transcontinental music projects over high-speed phone lines. Michael Redolfi, in France, jammed with Terry Riley, in Los Angeles. For the Brainspace project, ensembles in Freiburg, Germany and at the Electronic Café donned brainwave sensors as part of their distant improvisation, literally connecting brains across the Atlantic. Morton Subotnick, David Rosenboom, Leo Smith and others tele-

collaborated in the Three Cities / Multimedia Tele-concert, with audio and video links between Los Angeles, Santa Fe, and New York.

In each of these cases, connectivity and community were the foreground elements, bringing the physically distant to the physically tele-present, creating the illusion of erasing time zones and kilometers, as amateur radio had in the 1920s and 30s, and as telematic art did in the 1970s and 80s.

My first experience with telematic music came in 1987 as a member of The Hub, a network music ensemble. We started as a diverse bunch of electronic musicians, inspired by a previous group, The League of Automatic Music Composers. Chris Brown organized a pair of performances in San Francisco in 1986 called The Network Muse, which included members of The League of Automatic Music Composers as well as Larry Polansky, Phil Stone, Scot Gresham-Lancaster, Chris Brown, and myself. Shortly thereafter The Hub was formed from members of The League of Automatic Music Composers, John Bischoff and Tim Perkis, plus Chris Brown, Scot Gresham-Lancaster, Phil Stone, and myself.

Shortly after we formed we were invited by Nicolas Collins and Phill Niblock to create a linked musical performance between The Clocktower and Experimental Intermedia. The ensemble was split into two groups of three, each trio performing at each space. We used two identical "hubs" (microcomputers with modems) to send performance data back and forth between the two venues. There was no audio link. At one point in the concert some members of the audience left one venue and walked the few blocks to the other venue to resume listening. Kyle Gann's review in the Village Voice, "Musica Telephonica", noted both the idea of bridging space ("someday anyone with a modem can have the same experience of a concert as the people who are actually there") and creating community [5]. The wave of interest in telematic projects led to several opportunities, including the Points of Presence project (1997, CalArts, Mills College, and Arizona State University, it broke up the band), and a performance at the Dutch Electronic Arts Festival (2004, Rotterdam, our reunion concert). Still, we approached each of them with a healthy skepticism about telematic music. Our focus would always be "...a way to extend compositional ideas from the solo electronic performer to an ensemble, creating a new form of chamber music. (The fact that the chamber could be expanded in distance was not entirely irrelevant, but never really the point)" [2].

Telematic music projects have tended to highlight the network-collaborative aspects of the project rather than the music itself. Foregrounding the connection requires representing it as such, a pipeline of sound and image from a distant location. While sounds may be piped in on multiple channels/streams from somewhere else, *someplace* else, they don't bring their enclosing space with them. While audio electronics can create simulations of remote or imaginary acoustic spaces, these artificial rooms just form new, alien, spaces for sounds that are essentially excised from the space of their own origination and streamed in to another acoustic space. Acoustic spaces don't intermingle and resonate together in the prevailing

modes of telematic music. The visual component of the connection usually serves as a kind of proof of collaboration. In an age saturated with musical reproduction how do you know if the sounds you hear are “real”, being played by the people onstage in front of you? Displayed on video monitors, often onstage next to the “real” performers, disembodied ensembles prove that yes, it really is working. Network delays and technical glitches aside, what breaks the illusion of community is the cognitive dissonance of *seeing* and *hearing* the stream, but not the space in which the stream originated. Connectivity, represented as a two-dimensional image on-screen, requires clear sight-lines and forward focus while sound propagates through the hall in three dimensions surrounding the listener. Would it sound any different with my eyes closed? And if seeing is believing, how am I to integrate the visually dominant paradigm of cyberspace with the resonance of cyberplace?

3. PHANTOM ROOMS

When Stefano Bassanese invited me to create a piece for this conference, one involving distance collaboration between the cities of Torino and Cuneo, I took it as an opportunity to revisit telematic music. I chose to integrate my recent interest in using acoustic and digital feedback in my solo work with the idea of “close distance”. I came across the idea of using room resonance, the spectral profiles of different rooms used by the participants in the project, and network connectivity, as the central points of the piece. I chose the title Phantom Rooms to foreground the use of room acoustics in the piece and playfully suggest that rooms could make ghostly appearances inside other rooms.

My interest in feedback comes from a renewed interest in the music of David Tudor. Tudor’s work as an electronic musician centered around the use of feedback networks. He interconnected guitar pedals, consumer audio equipment, and homebrewed circuits to create complex circuits whose non-linear behavior stood in contrast to the linear construction and more predictable action of most analog synthesizers. In a mode more oriented towards discovery than execution, Tudor performed by playing with the emergent behaviors of these feedback networks. As Tudor himself said, “Well, I let it play itself as much as I can, but if it doesn’t, then I interfere” [4].

For Phantom Rooms I developed a Max/MSP patch that could be used as a prototype for the players in the piece.¹ This patch was based on a digital feedback circuit with inputs for injection of network and microphone audio signals. In the middle of the circuit was a finite impulse response filter (buffir~) allowing the feedback signal to be filtered by the impulse response of a room (see Figure 1). Each player could take this prototype and add other signal processing elements as he saw fit. Players were asked to take impulse responses of rooms where they make music, both public and private. These impulse responses

were shared among all the players and used in everyone’s version of the patch.

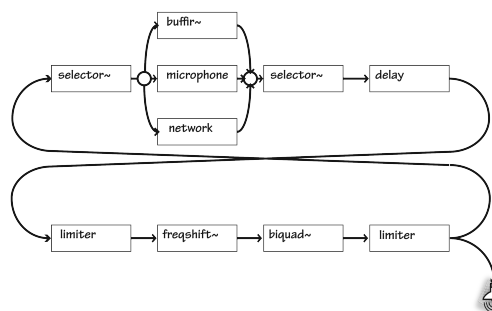


Figure 1. Prototype for Phantom Rooms circuit.

Additional audio could be injected into the system through the network or microphone inputs. Audio networking was accomplished using Olaf Matthes’ netsend~ and netreceive~ objects, which I had recompiled and tested for the latest versions of the OS X operating system and Max/MSP. Players who also played guitar were asked to occasionally play long tones (using an E-bow) based on the resonant frequencies found in the concert halls. Optionally, other players could connect microphones to their audio interfaces to work with acoustic feedback. A very simple score in the form of a timeline was produced, but only as a general guide for improvisation. As with Tudor’s performing strategy, I wanted the players to discover the emergent behavior of this complex digital/analog circuit, so a light compositional touch was all that was needed.

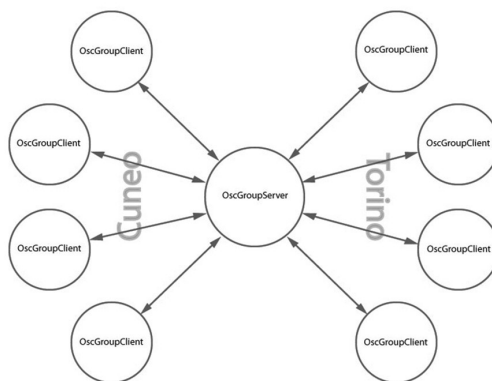


Figure 2. Phantom Rooms network topology.

The Phantom Rooms network used the same peer-to-peer topology we use in The Hub. The network infrastructure is built on Ross Bencina’s OSCGroups, a simple-to-use system for routing UDP messages between users². It solves the problems of constantly changing IP addresses and addressing computers behind firewalls, among oth-

¹ Players included faculty and students from the Conservatorio “G. F. Ghedini” in Cuneo, Italy: Benjamin Thigpen, Alessio Dutto, Bruno Fabrizio Sorba, Giuseppe Mercuri, Marco Chiavarino, Gianluca Delfino, Francesco Torelli, Giulio Beccaria, Nicola Biagioni, and Mark Trayle.

² <http://www.audiomulch.com/~rossb/code/oscgroups/>

ers. The OSCGroupServer served as the central routing mechanism between the players, each connected to the server by the OSCGroupClient (see Figure 2). Performers interacted with their feedback circuits through a number of control points. In the prototype circuit I provided, most of the shaping of sound in the piece is done by changing the characteristics of the filter, pitch shifter, limiters and delays. Each performer added additional controls as they customized the prototype. Performers interacted with each other by sending commands to change another player's impulse response, and receiving audio signals from other players to add into their feedback loops. I thought of this as a rather direct way of sharing one's acoustic environment with another: a bit like being invited into a friend's home and bringing your own room, your own acoustic space, into the house with you. Phantom Rooms engages the metaphor of space through the interaction and connectivity of the players, and of place through the sharing of the spectral characteristics of the rooms they work and perform in.

4. JOINING SPACE AND PLACE

Network music projects can take many forms: remote collaborations using streaming audio, interconnected improvisations in a shared physical space, soundings of the network's physical dimensions, elisions of all of these plus other permutations. Here I've looked at the possibility of joining two metaphors (cyberspace and cyberplace) and two methodologies (feedback and emergence) as the starting point, and extend the more idiomatic features of network behaviors into the physical realm. Network music, it's roots in the surround-sound simultaneity of acoustic space and the social dynamics of music-making, is in a unique position to discover collaboration and spontaneity on the digital landscape.

5. REFERENCES

- [1] Ascott, R., "Art and Telematics: towards a network consciousness/ Telematik/L'Art et le Télématique." In: *Telecommunication*, H. Grundmann (ed.), The Western Front, Vancouver, pp. 25-67.
- [2] Brown, Chris, & Bischoff, John, "Indigenous to the Net", http://crossfade.walkerart.org/brownbischoff/hub_texts/hub_aesthetics_f.html
- [3] Curry, Michael R., "Cyberspace and cyberplaces: Rethinking the identity of individual and place", paper presented at the *20th General Assembly and Scientific Conference of the International Association for Mass Communication Research*, Sydney Australia, 1996
- [4] Fulleman, John, "Interview with David Tudor", <http://davidtudor.org/Articles/fullemaan.html>.
- [5] Gann, Kyle, "Musica Telephonica (The Hub)", *The Village Voice*, XXXII, 25, p. 83.

- [6] Gidney, Eric, "Art and Telecommunication – 10 Years On", *Leonardo*, 24: 2, p. 148.
- [7] Max Neuhaus, "Rundfunkarbeiten und Audium", *Transit*, Zeitgleich, Vienna 1994, pp. 21–23

COMPOSER'S TOOLKIT: A SYNTHESIS OF MUSICAL LANGUAGE

Joshua Parmenter

Center for Digital Arts and
Experimental Media, University of Wash-
ington, Seattle
josh@realizedsound.net

ABSTRACT

The Composer's ToolKit (Ctk) extension library for the SuperCollider Synthesis language brings together aspects of many older synthesis and algorithmic composition languages and environments.

1. THE COMPOSER'S TOOLKIT

One of the amazing things about the research and creation of computer music languages has been the fact that so much of the technology has been created by musicians and other people who are familiar with musical ideas. As a result, many programming languages and environments for music that can be found have much in common with the history of musical development itself (at least in regards to the western classical tradition). The SuperCollider language in particular, which I have used as my primary tool for composition, research and development for almost a decade now, has some especially strong affinities with musical practice and development. The SuperCollider language is well over a decade old now with the most recent versions released as SuperCollider Server (often also referred to as SuperCollider 3). As of the time of this writing, the latest stable version is SuperCollider 3.4.2, released in January of 2011. The language is released as an open-source project under the GNU Public License. Previous versions of SuperCollider were created by James McCartney as a closed source product. In 2002, the sources for the still in development SuperCollider Server were released by James, and a number of users and class developers from earlier versions of SuperCollider immediately began developing the new project. As before, the language was an object-oriented, extensible interpreted language featuring realtime garbage collection. It is the extensible class library that, in my mind, brings a strong connection to the history of western music, specifically written traditions and how those traditions have grown to preserve history while also expanding to make new discoveries.

As a system, western classical music's system of music notation has proven extremely flexible and expand-

able. If we look at a score (Figure 1) from the 18th century, we can see a system that had developed to include information about pitch (though limited to the 12 note chromatic scale, and note that nothing is really stated about tuning!) and rhythm. While we tend to divide rhythms usually into halves (a quarter note is half of a half note for instance) the system of meter allowed for more complex rhythmic relations (the thirds of a compound meter for instance) and if a different subdivision was desired, a special notation was added (tuplets). Information about pitch hierarchy (key) is included that also helps musicians shape phrasing and pitch accent, but there is little other detail in the score. Even tempo is left to the performer. But as composers started to standardize more parameters of musical thought, the notation expanded while at the same time retaining the ability to hold onto its previous abilities and knowledge. In Figure 2, dynamics are apparent, as well as markings that indicate bowing and phrasing that may be counter to what was expected out of normal performance practice. Figure 3 shows further expansion of the system to allow written directions to describe timbre and directions to the performers about how to play their instruments, and there are even moments here where the resulting pitch isn't even accurately described anymore. It takes a little more specialized knowledge for a musician to know what the resulting pitch will be from the artificial harmonic that is given to the cello in measure 5, but we now have symbols telling the player how to play more than simply what to play. Figure 4 shows an even more recent and extreme example. The point is that, compared to Figure 1, we are still dealing with the same notation system. In fact, this is just a small snapshot of the whole picture since music written well before the 1700s can still be represented very well in the system, retaining the knowledge that it has discovered in its past, while at the same time allowing for some very detailed expansion. I wouldn't go so far as to say that western notation is able to express *anything* musical, but composers and theorists over the past 600 years have found ways to express some very complex thinking within and around this system.

Copyright: © 2010 Parmenter, Joshua et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



Figure 1: Bach, J.S. Flute Sonata in b, BWV 1030



Figure 2: Beethoven, String Quartet Op. 59 No. 1 I



Figure 3: Webern, Op. 9



Figure 4: Ferneyhough, *La chute d'Icare*

While computer music (especially programming languages for music) has a history of over 50 years, I would like to offer my own experiences as a small microcosm that will hopefully link the ideas about notation to my development work with SuperCollider and the Composers ToolKit (Ctk). I came to the University of Washington for a Masters of Music in Composition in 2000, and after a couple of meetings with Richard Karpen, I asked if I could study with him. However, at the time I had very little interest in learning computer music beyond some fundamentals that would help me get a job later. At our first lesson he asked if I was taking the computer music series that quarter, and I told him I was going to wait until the next year. Richard told me if I didn't take it right now, I had to find someone else to study with, so I signed up for the class (taught at the time by Juan Pampin). The series at the University of Washington (then taught in the School of Music and now taught at the Center of Digital Arts and Experimental Media) is a wonderful combination of music and art history, strong composition critique and an intensive course in programming. In 2000, CSound and LISP (with Rick Taube's wonderful library 'Common Music') was used for teaching the course. Over the course of that year, I went from someone who was just starting to use email to someone who was starting to program, from a composer that was concerned with just notes on the page to one that began to open up to the world of sound, timbre and extended techniques. While my masters project had no computer part in it, the thinking that went into the project was certainly algorithmic in many respects, and the sonic result was more than I had ever guessed could be done with the instruments I was using. This was, of course, a direct result of my experience with algorithmic composition and signal processing in the computer music course.

CSound comes in part of the long line of 'Music-N' languages where the code work is divided into two main categories—an orchestra that describes synthesis and DSP algorithms and a score that describes note events. At the University of Washington, we used CSound mostly for its synthesis capabilities, and Common Music was our main tool for creating scores. Even for programmers coming into the series, one of the major tasks for students was the acquisition of CSound's syntax and the use of LISP (already a rare language for just about anyone at that time to have any serious contact with). Also mixed into the lab environment at UW was a group of students and staff that, along with Richard Karpen, were starting to use SuperCollider 2. I had some brief contact with SuperCollider that first year, and took a course with Chad Kirby the following year. SuperCollider 2 had a couple of major differences from CSound and Common Music, the primary thing being that the language itself was also the synthesizer (in a very similar fashion to Bill Shodtstadt's 'Common Lisp Music'). While a UGen in SuperCollider had slots that were similar to CSound's p-fields, those slots could also have other signals patched into them on the fly. An Instr in CSound could set up a p-

field (let's say p4) that allowed you to pass a parameter from the score into the Instr, but what that p-field could contain was limited to a single numerical value. If you wanted to define an envelope from the score, you had to have a separate instrument that took more p-fields and filled the values into the Instr in the proper way. In SuperCollider 2, you just had to change the single value into an envelope and everything still worked.

In SuperCollider 3, there was a separation of the program into language and synthesis programs. Along with this came the introduction of the SynthDef structure as a way to describe synthesis graphs (which, in SuperCollider 2 was done through ALL the code that was sent to the interpreter). As a result, the elegant functionality described above that SuperCollider 2 gave us was in some ways lost again. A SynthDef is a fixed structure in both the language and the server, more like an Instr in CSound, where the number of function arguments and the number of memory slots allotted to them (like p-fields in CSound) are static. For example:

```
SynthDef(\test, {arg env;
    Out.ar(0, SinOsc.ar(440, 0, env))
}).add;
```

the code above says that the argument 'env' will be a single value, it can't suddenly be given an array of values representing a Env object or a dynamic LFO. That argument does update at the control rate in the system however, and as a result you can have another synthesis process writing values from an Env or an LFO out to a control bus. The output of that process could then be mapped to the 'env' parameter to create a similar effect to what was available in SuperCollider 2, though it is a much more involved process that requires much more code and even some knowledge that this can happen at all (a common question among new SuperCollider users!).

Finally, the system for non-realtime synthesis in SuperCollider 2 was amazingly simple. The enclosing Synth object that was created in SuperCollider 2 simply needed to call a different method to run in realtime or non-realtime (Synth.play versus Synth.write). The division of the language and synthesis work in SuperCollider 3 into two different programs made this impossible. Non-realtime rendering in SuperCollider 3 requires that you create a score-like listing of time-stamped OSC commands that can be sent to the rendering server. While all communication with sserver in SuperCollider 3 is already OSC based, the OSC messages themselves are mostly hidden within classes that take care of the formatting and sending for you. The class abstractions that do this work do have methods that let you capture OSC messages into a Score, but to run a piece in realtime versus non-realtime you need two different sets of code.

The main motivation for Ctk was to try and recapture some of the elegance of the SuperCollider 2 language, while at the same time attempting to bring some of the algorithmic power of CommonMusic and other languages into SuperCollider 3. The above problem of non-realtime performance versus realtime performance

was the first motivator. As the development of classes that were more self aware of their own parameters progressed, other desired features also became apparent fairly early on. Thinking back these were the four biggest areas of attention:

- 1) Must be able to work in real-time and non-real-time almost seamlessly with little change to syntax. In addition, a syntax structure that resembled the rest of the language was desirable.
- 2) Should take advantage of an object-oriented environment. Objects must be able to set values in an object instance and query values that may be used further down the line for algorithmic decisions.
- 3) Algorithmic output must be able to be saved in a text readable score-like format that someone could still edit by hand.
- 4) It should be modular and expandable.

The result is the current Ctk library. The primary objects in Ctk consist of CtkScore, CtkNoteObject (and a few classes that allow for different ways of storing SynthDefs and creating the base note object including CtkSynthDef), CtkNote for note representations, CtkAudio and CtkControl for audio and control signals and CtkEvent for modular functions that control large scale gestures of many notes. In current development are classes that take on larger scale functionality (and are meant as replacements for the ProcMod and ProcEvents classes that have been used to control a large number of pieces of mine, as well as Richard Karpen, Juan Pampin and a number of our students). One of the biggest conveniences that these classes provide is their ability to recognize each other. The classes in the library can expect certain kinds of information to be present. For instance, when using the regular Score class that is in the SuperCollider main distribution, you fill it with OSC messages by hand. Then it is necessary to add a 'dummy event' that simply consists of a time-stamp. The time-stamp of this final event tells the NRT engine when to finish rendering. It is then up to you to keep track of note onsets and durations and then add this line to the end of your score. With Ctk, every Ctk object is able to take a starttime and duration parameter, and when objects are added to a CtkScore, the instance of CtkScore will take these parameters into account and will calculate the correct endtime for you. While this seems like a small thing to have a library do for you, I can't express how much headache this saves for the students of my computer music class (which means we have more time to talk about synthesis techniques and composition rather than how to set up variable and function to keep track of these times for you!).

Some of the features that have grown out of the modular and expandable requirements start to make the library also let SuperCollider 3 act more and more like SuperCollider 2. Arguments to synthesis functions held in a CtkNoteObject can take other Ctk objects as parameters and the different objects will take care of the underlying requirements and OSC messages to make them work. Putting a CtkControl into a synthesis parameter will control the creation of a unique control bus, a second syn-

thesis process that is then written out to that control bus, and finally the necessary OSC message to map the output of that control synth into the main one. While the logic to do this inside the classes has to account for many cases (which also means that the language suffers a slight performance hit), I feel the time saved and the flexibility that is readily available to you more than makes up for this. And since it all happens in the language the performance hit is a one time cost. In SuperCollider 2 a library such as this would have been very expensive in terms of CPU usage since the synthesis graph was the same as the language. These kinds of conditional statements would have been evaluated over and over again every control period. Finally, since CtkScore can add other CtkScores, the expandability of the system in the organizational domain really is limited more by system resources than conceptual ones. In modern operating systems where the amount of RAM that can be addressed is moving into the 64-bit realm, this will become less and less of a concern as well.

While CtkScore allows for a smooth connection between the real-time and non-real-time modes of the SuperCollider synthesis server, the library itself also brings interaction with the server into a similar realm as the rest

```
SynthDef(\test, {arg gate = 1, freq;
  var env, envGen;
  env = \env.kr(Env([0, 1, 0], [1, 1], \sin, 1));
  envGen = EnvGen.kr(env, gate, doneAction: 2);
  Out.ar(0, SinOsc.ar(freq, 0, 0.1 * envGen));
}).add;

// no sound! 'freqs' should be 'freq', and 'env' has to have
// special treatment

a = Synth(\test, [\freqs, 440])
  .setn(\env, Env([2, 1, 0], [0.5, 2], \sin, 1));
a.free;

a = Synth(\test, [\freq, 440])
  .setn(\env, Env([2, 1, 0], [0.5, 2], \sin, 1));
// release the note
a.set(\gate, 0);

b = CtkSynthDef(\test, {arg gate = 1, freq;
  var env, envGen;
  env = \env.kr(Env([0, 1, 0], [1, 1], \sin, 1));
  envGen = EnvGen.kr(env, gate, doneAction: 2);
  Out.ar(0, SinOsc.ar(freq, 0, 0.1 * envGen));
});

// this doesn't play, and throws an error that 'freqs' is not
// understood and the Env is passed in as just another argument
c = b.new.freqs_(440)
  .env_(Env([2, 1, 0], [0.5, 2], \sin, 1)).play;

c = b.new.freq_(440)
  .env_(Env([2, 1, 0], [0.5, 2], \sin, 1)).play;

c.release;
```

Figure 5: Comparison of standard and Ctk library for note creation

of the language. Where the stock server abstraction classes (e.g. Synth, Bus, Buffer) mostly wrap their arguments into the appropriate OSC messages, the Ctk library resembles the syntax of the rest of the SuperCollider language. This simplifies learning the language for new users, and also supplies important debugging tools to experienced users. Figure 5 shows a comparison between the two different approaches for creating the same note.

In the first (using SynthDef and Synth), values to the SynthDef's parameters are set using an array of symbol and value pairs, the array representation of an Env has to be passed into the note as a separate message to the Synth object and a misspelling in a symbol would fail silently. In the second example using CtkSynthDef, the 'new' note that is created from the CtkSynthDef knows about its parameters and will throw an error on the misspelling, the Env is converted for you and it can all be done using the setter syntax available to the rest of the language.

The note prototypes in Ctk (and the arguments that are described in their synthesis graph functions) behave more like the classes that make up the SuperCollider language. Values that are set within Ctk can also be retrieved inside the language. In the above example, the value set with the 'freq_' setter method can later be polled by other objects with the 'freq' getter method. However, this functionality is still limited. For instance, if I was to place one of CtkControl's 'lfo' objects into the 'freq' argument to add randomness to that parameter, polling that argument would only return to me the CtkControl object itself rather than the current output value of that synthesis process. Programming this kind of interaction is part of what is on the horizon for Ctk, as well as a couple of other features.

As mentioned earlier, Ctk versions of my classes ProcMod and ProcEvents are under development for a project that will have performances in early 2011. CtkPMod and CtkPEvents, like the other objects in the system, will work in both real-time and non-real-time, and should make the rendering of electronic parts of performer and live electronics pieces much easier. In addition, each layer will have the ability to render out its output as a single layer soundfile so the final mixing of a recording can still take place inside a more sophisticated digital audio workstation. Also in development for a couple of years (but still not satisfactorily complete in my opinion) is the companion Notation Tool Kit (Ntk). Ntk will eventually connect into Ctk rather seamlessly and will allow you to output notation representation of computer generated events as well as the ability to use the algorithmic strengths of the SuperCollider language for instrumental writing. Currently Ntk can output MIDI files, Guido and Lilypond markup and MusicXML. Bugs abound however (especially in MusicXML output) and I hope to someday have the time, or the grad students, to really complete this work.

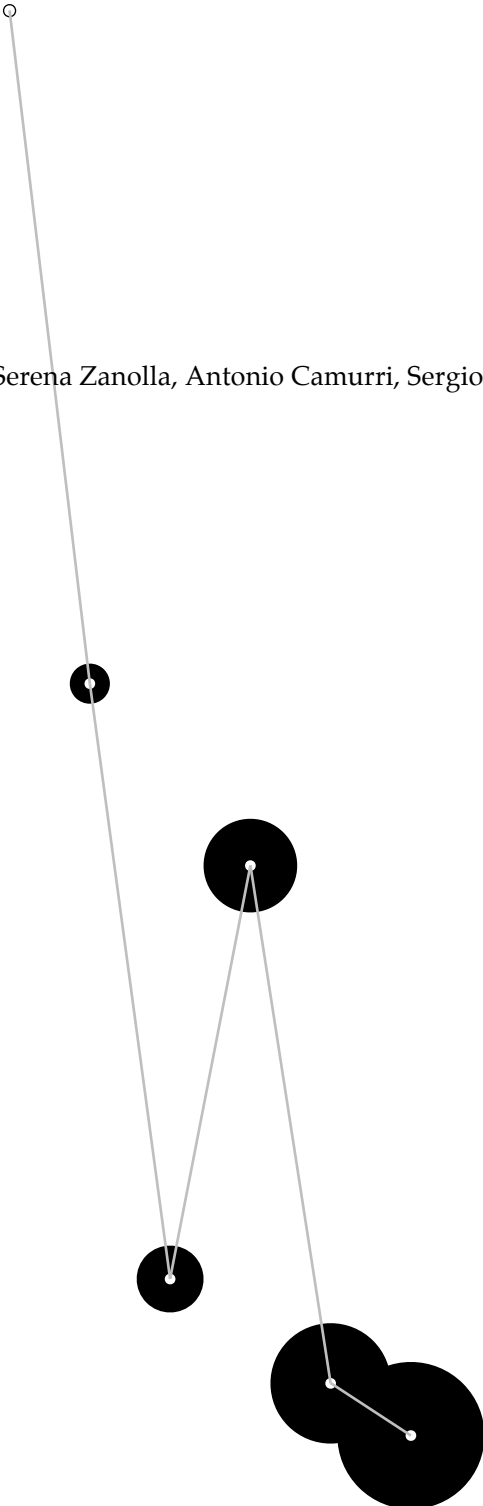
Like music notation, I feel that Ctk has given the SuperCollider language an interface that both captures the strengths of past languages and environments while presenting an expandable possibility for the future. Maybe someday the ability to simply pass in a String to Ctk like Ctk.string("Play with a soft touch and gradually get noisier") will become possible as computer parsing and learning becomes more complex. If it can't (and I'm pretty sure it won't be able to given the speed and development of new languages, machines and systems) then hopefully some yet undiscovered technology will be able to someday. Hopefully that same system will also give us

the ability to capture and express the inventions and knowledge of all that has come before.

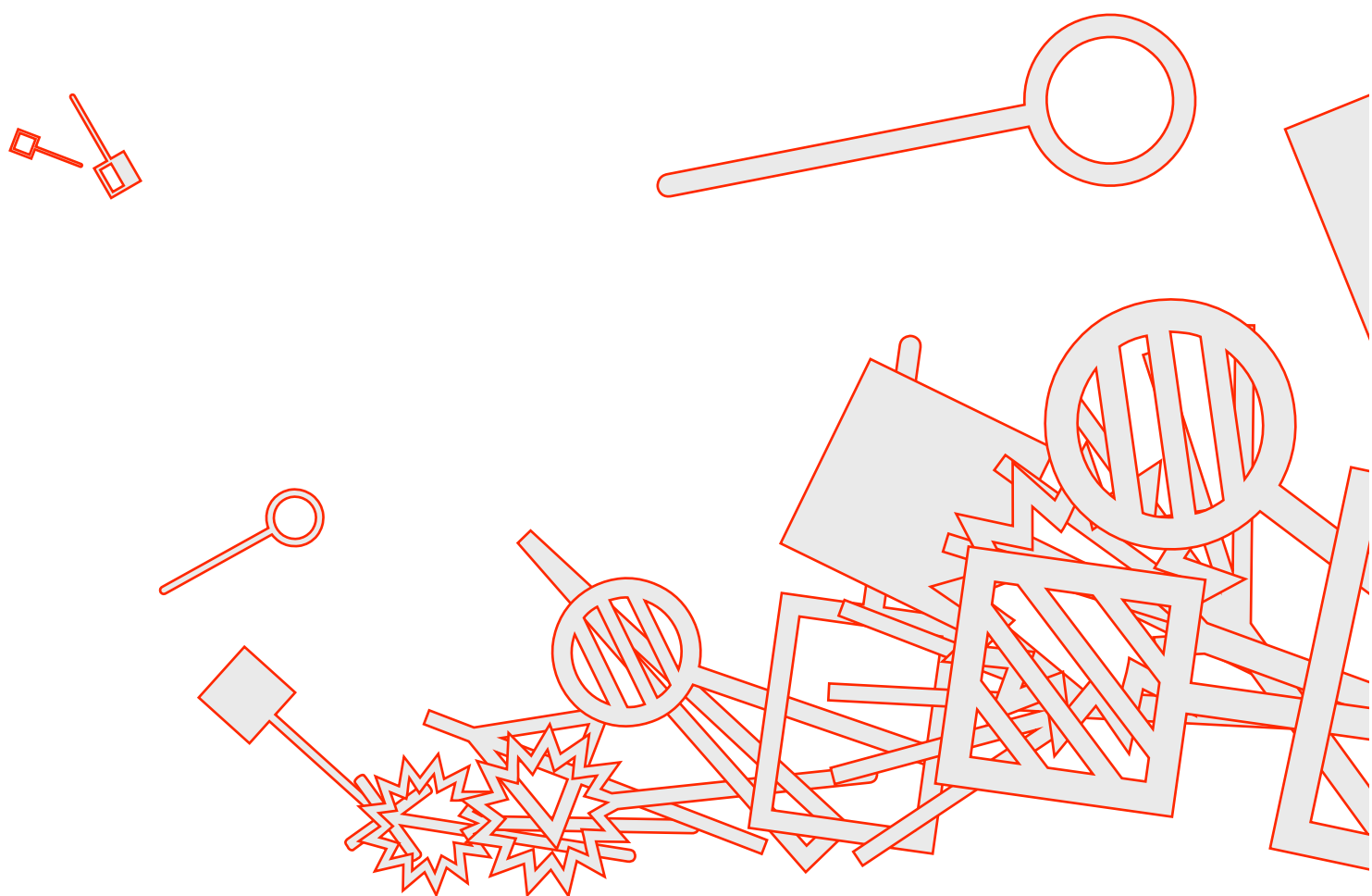
Papers

+ Sergio Canazza, Federico Avanzini, Maddalena Novati, Antonio Rodà
+ Cat Hope
+ Marco Marinoni
+ Giuseppe Cabras, Sergio Canazza, Pier Luca Montessoro, Roberto Rinaldo
+ Gianmario Borio, Giacomo Albert, Nicola Bizzaro, Alessandro Bratus, Alessandro Cecchi, Maurizio Corbella, Matteo
Giuggioli, Stefano Lombardi Vallauri, Marida Rizzuti, Federica Rovelli
+ Stefano Fumagalli, Saverio Monti
+ Carlo Massarelli, Andrea Valle
+ Pietro Polotti, Maurizio Goina
+ Stefano Delle Monache, Davide Rocchesso
+ Simone Spagnol, Michele Geronazzo, Federico Avanzini
+ Enrico Marchetto, Federico Avanzini
+ Sergio Canazza, Antonio Rodà, Daniele Salvati
+ Mattia Schirosa, Jordi Janer, Stephan Kersten, Gerard Roma
+ Marco Gasperini
+ Federico Costanza, Marco Gasperini, Alessio Rossato
+ Serena Zanolla, Antonio Camurri, Sergio Canazza, Corrado Canepa, Antonio Rodà, Gualtiero Volpe, Gian Luca Foresti
+ Tiziano Bole
+ Nicola Montecchio, Nicola Orio
+ Massimo Avantaggiato
+ Giorgio Klauer
+ Mauro Lanza, Gianluca Verlingieri, Nicola Biagioni
+ Giorgio Bianchi, Pier Daniel Cornacchia
+ Sergio Canazza, Antonio Rodà, Lauro Snidaro, Ingrid Visentini

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
[
\
]
^
_
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z



1. Tecnologia, memoria e interpretazione



+ Sergio Canazza, Federico Avanzini, Maddalena Novati, Antonio Rodà

+ Cat Hope

+ Marco Marinoni

+ Giuseppe Cabras, Sergio Canazza, Pier Luca Montessoro, Roberto Rinaldo

+ Gianmario Borio, Giacomo Albert, Nicola Bizzaro, Alessandro Bratus, Alessandro Cecchi, Maurizio Corbella, Matteo
Giuggioli, Stefano Lombardi Vallauri, Marida Rizzuti, Federica Rovelli

A Music Bar for active listeners: An example of Virtual Electronic Lutherie for a history 50 years long

Sergio Canazza, Federico Avanzini

SMC Group, Dept. of Information Engineering
University of Padova
Via Gradenigo 6/B, 35131 Padova
canazza@dei.unipd.it
avanzini@dei.unipd.it

Maria Maddalena Novati

RAI, Milano
Archivio di Fonologia
novati@rai.it

Antonio Rodà

AVIRES Lab., Dept. of Informatics
Viale delle Scienze, Udine
University of Udine
antonio.roda@uniud.it

ABSTRACT

This paper presents initial results of an ongoing project devoted to the analysis and virtualization of the analog electronic devices of the “Studio di Fonologia Musicale”, one of the European centres of reference for the production of electroacoustic music in the 1950’s and 1960’s. After a brief summary of the history of the Studio, the paper discusses a particularly representative musical work produced at the Studio, namely the analogue tape work *Scambi* composed in 1957 by Henri Pousseur. Finally, the paper presents initial results on the analysis and simulation of the electronic lutherie used by Pousseur in this composition, and the ongoing work finalized at developing an installation that re-creates such electronic lutherie.

1. INTRODUCTION

“I can easily imagine that in the not too distant future the possibility of doing such work will become generally available. [...] The whole material could be made available to amateurs in some kind of ‘music bar’. The joy of turning the volume control – maybe in company – or even of a little spatial distribution, [...] would help to give the now active listener the experience of how the course of an event can be influenced and raised to the form of a vital, creative freedom.” [1, p. 54]

A classical musical composition (a Beethoven symphony, a Mozart sonata, or Stravinsky Rite of Spring) posits an assemblage of sound units that the composer arranged in a closed, well-defined manner before presenting it to the listener. He converted his idea into conventional symbols, obliging (more or less) the (eventual) performer to reproduce the format devised by the composer himself. On the contrary, a number of music (or, more generally, multimedia) works are linked by a common feature: the considerable autonomy left to the individual performer in the way he chooses to play the work [2]. Thus he is not merely free

to interpret the composer’s instructions following his own discretion (as happens in traditional music), but he must impose his judgment on the form of the piece, as when he decides in what order to group the sounds: an act of “improvised creation” [3]. An example is *Klavierstück XI* by Karlheinz Stockhausen, where the composer gives to the performer a single sheet of music with note groupings, and the performer has to choose among groupings and mount the sequence of musical units in the order he chooses, thus altering the combinative structure of the piece. One second example is the Third Sonata for piano by Pierre Boulez, where the first section (*Antiphonie, Formant 1*) is composed of ten different pieces on ten corresponding sheets of music which can be arranged in different sequences (although with some constraints in possible permutations).

A particularly representative example of musical open work is *Scambi*, an analogue tape work created in 1957 by the Belgian composer Henri Pousseur (1929 – 2009) at the *Studio di Fonologia Musicale* in Milan. An analysis of this work is provided in [4]: by means of a specific process, termed “dynamic filtering” and realized through a custom device (the *Selezionatore di ampiezza*, or “amplitude selector”, designed by Alfredo Lietti [5], the engineer of the studio), the composer extracted complex time structures from noise, and further processed them eventually producing 32 sequences. These sequences can be arranged by anyone who had access to them and who wished to realize the work, according to certain rules regarding their order and possible overlapping. Together with the composer and the performer, the third main actor, in this sense is the listener.

Today’s technology increasingly enables novel forms of interaction between users and multimedia content. To a certain extent this is also true for music content: high quality music can be enjoyed on low-cost and even mobile devices, can be selected by users depending on their taste and mood, and so on. However, to date music listening remains a passive and non-interactive experience: there is the need for novel approaches that can encourage an active music listening experience, in which the listener can interact with the music content, modify it, and ultimately recreate it. In this perspective the listener will become a prosumer. (as predicted in the 1980’s by futurologist Alvin Toffler, who coined this portmanteau by contracting the word “producer” with “consumer”).

The aim of this paper is to report on initial results of an

ongoing project devoted to the analysis and virtualization of the analog electronic devices of the Studio di Fonologia Musicale. The final goal is to develop an installation consisting of a SW-HW system that re-creates the electronic lutherie of the Studio, allowing users to interact with such lutherie. In particular, the production setup originally employed to compose *Scambi* is considered as a relevant case study. Achieving the goal of the project implies (i) analyzing the original devices through both project schemes and direct inspection; (ii) validating the analysis through simulations with *ad-hoc* tools (particularly Spice – Simulation Program with Integrated Circuit Emphasis, a software especially designed to simulate analog electronic circuits [6]); (iii) developing physical models of the analog devices, which allow efficient simulation of their functioning (according to the *virtual analog* paradigm [7]); (iv) designing appropriate interfaces to interact with the virtual devices.

The remainder of the paper is organized as follows. Section 2 briefly summarizes the history of the Studio di Fonologia. Section 3 discusses in some detail the form and the realization of *Scambi*. Finally, Sec. 4 presents initial results on the analysis and simulation of the electronic lutherie used by Pousseur for the composition of *Scambi*.

2. THE STUDIO DI FONOLOGIA MUSICALE

The Studio di Fonologia Musicale [8] was founded in 1955 at the Milan offices of the Italian Radio-Television (RAI), under the initiative of the Italian composers Luciano Berio and Bruno Maderna. In a few years, the Studio became one of the European centres of reference for the production of electroacoustic music, by deploying cutting-edge devices for the generation and processing of sound. Often these devices were especially designed and crafted by Alfredo Lietti: oscillators, noise generators, filters, dynamic and frequency modulators. These were unique pieces, created with great care to meet the needs of the composers who attended the Studio.

In 1967 the Studio underwent a partial renovation. As a consequence, much of the older equipment was dismantled and has been lost. However, thanks to records kept in archives (photographs, schemas, drawings and articles) it is possible, in many cases, to know the characteristics and the functionality of most equipments that no longer exist. The Studio was closed in 1983 and the devices were disassembled and transported to Turin, where they remained packed in storage until 2003, when they were returned in the RAI headquarters in Milan.

The electronic lutherie of the Studio di Fonologia Musicale has recently been transferred to the Milan Museum of musical instruments: this inestimable technological and cultural heritage is now accessible to the general public in a permanent museum exhibition. However the electronic devices and sound generators are not currently functioning, they are exhibited as mute instruments and visitors can only listen to recordings of musical compositions that made use of these devices. As a consequence, it is not possible to fully appreciate the variety of timbral and expressive possibilities of these instruments, which could be

experienced only by directly interacting with them. The museum is thus seeking for solutions that allow visitor to experience such an interaction, possibly by a virtual recreation of the devices.

The cultural value of the electronic lutherie of the Milan Studio is testified by the efforts devoted jointly by RAI and the Milan Museum of musical instruments, in order to set-up the current exhibit. In fact the Milan Studio, together with its counterparts in Paris and Cologne, was one of the topical places of early electronic music, and during the 1950's and 1960's was able to attract composers of international stature. Henry Pousseur was among them.

3. SCAMBI

We propose now the analysis of an interesting case study: Pousseur's electronic work *Scambi*, following [4]. Writing about his work in 1959, Pousseur ended by envisaging the day when technology would allow listeners to make their own realizations of the work (either following his connecting rules or not) and to give the, now active, listener the experience of a temporal event open to his intervention and which could therefore be elevated in type, as vital, creative freedom [1]. The active listener becomes, in effect, a composer; reception and interpretation are expressed as (musical) production.

In our work, Pousseur's invitation to creatively interpret and re-compose his work is extended to other types of appropriation that were not only permitted but welcomed by the composer [1] (a position that associates him with popular-music culture in which such freedom is assumed). In our case study, we have collected the original 32 audio sequences realized by Pousseur, thanks to the *Scambi Project*, Lansdown Centre for Electronic Arts, School of Arts, Middlesex University, UK (<http://www.scambi.mdx.ac.uk>).

3.1 Form and realization of *Scambi*

Pousseur focused on material which could be easily produced through real-time processes (he only had six weeks for the realization of the work). The work is based on a set of sequences, connected to each other based on a set of rules. Sequences differed from each other in terms of four musical parameters: the statistical tempo (from slow to fast), the relative pitch (from low to high), the "homogeneity" of the sound pattern (from dry to long reverb), and the "continuity" (from long silence gaps to continuous sound) (see Fig. 1).

Sequences were constructed as follows. The composer extracted irregular impulse patterns from input signals made of white noise band-passed at different center frequencies. These signals were processed through "dynamic filtering" with the *Selezionatore di ampiezza*. The device acted approximately as a noise gate (see Sec. 4), therefore depending on the settings the output signals extended from a isolated impulses to complex swarms of pseudo-random impulses. Through mixing and montage of the tapes obtained from this process, he constructed four basic sequences, in which only one parameter among tempo and pitch is

Family	Sequence	Pitch	Speed	Hom.	Cont.	Dur.
1	1-2	0/1	1/1	1/0	0/0	42"
2	3-4	0/1	1/1	0/1	1/1	42"
3	5-6	1/0	1/1	0/0	0/1	42"
4	7-8	1/0	1/1	1/1	1/0	42"
5	9-10	1/1	1/0	1/0	1/0	30"
6	11-12	1/1	1/0	0/1	0/1	30"
7	13-14	1/1	0/1	0/0	0/0	30"
8	15-16	1/1	0/1	1/1	1/1	30"
9	17-18	0/0	0/1	1/0	0/1	30"
10	19-20	0/0	0/1	0/1	1/0	30"
11	21-22	0/0	1/0	0/0	1/1	30"
12	23-24	0/0	1/0	1/1	0/0	30"
13	25-26	1/0	0/0	1/0	1/1	42"
14	27-28	1/0	0/0	0/1	0/0	42"
15	29-30	0/1	0/0	0/0	1/0	42"
16	31-32	0/1	0/0	1/1	0/1	42"

Figure 1. Pousseur, *Scambi*, table with transformation processes; characteristic per sequence [4].

pitch	1	1 1 1 0	1 1 1 0 0 0	1 0 0 0	0
speed	1	1 1 0 1	1 0 0 1 1 0	0 1 0 0	0
homogeneity	1	1 0 1 1	0 1 0 1 0 1	0 0 1 0	0
continuity	1	0 1 1 1	0 0 1 0 1 1	0 0 0 1	0
selected sequences	*	*	*	**	* **

Figure 2. Pousseur, *Scambi*, table with transformation processes; all possibilities of characteristics and Pousseur's choice [4].

changing. These sequences could be played backwards, could be reverberated, and finally processed again with dynamic filtering. Pousseur executed only a part of the possible combinations in each work process, obtaining 32 sequences (see Fig. 2).

3.2 A Music Bar

The so-obtained sequences could then be arranged by the composer, according to rules defining their order and overlapping. The identity of the work would be maintained even between markedly different arrangements, due to the homogeneity of the sound material which would always create the same general impression [1].

Pousseur writes in the booklet accompanying the compact disc: "Several sequences (which begin with a same character) can even be superposed and thus lead to a polyphony of divergent situations – until they come together again on a common point." [9]. An example of branching structure leading to polyphony is shown in Fig. 3: the performance starts with sequence 1 and ends with 23. The arrows indicate which sequences connect with each other. At one point there are four layers proceeding simultaneously.

The analysis outlined above allows to identify some characteristics of the open-form [2]: the interchange of sections within certain constraints, and the possibility of combin-

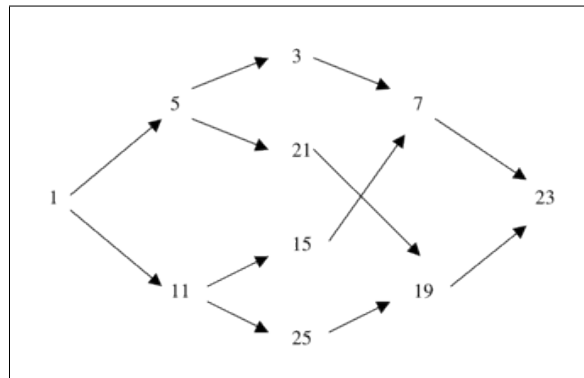


Figure 3. Example of branching structure.

ing sections. In particular Pousseur specified "connecting rules", according to which only certain sequences beginning with the same character could be rearranged. In fact several versions of *Scambi* exist. Two were realized by Pousseur himself, one by Marc Wilkinson [10], and two by Luciano Berio (who seems to have disregarded the connecting rules suggested by Pousseur).

Pousseur's conception of open form can be summarized in the definition of 'music bar' given by the composer himself:

"(...) I can easily imagine that in the not too distant future the possibility of doing such work will become generally available. All one needs at home is some splicing tape and leader tape and a simple tape recorder, or preferably two or three (and with several amateurs this should not be too difficult), which need not be perfectly synchronized, the small errors here leading to constantly new figures; else, the whole material could be made available to amateurs in some kind of 'music bar'. The joy of turning the volume control – maybe in company – or even of a little spatial distribution, requiring no expensive apparatus, would help to give the now active listener the experience of how the course of an event can be influenced and raised to the form of a vital, creative freedom." ([1, p. 54])

Therefore, as a consequence from his work on *Scambi*, Pousseur envisaged a new way of musical praxis in which the composer provides source materials which have to be further developed by each prosumer [4]. In this sense, Pousseur anticipates certain aspects of contemporary commercial developments where consumers today can, as an example, construct their own albums (often by downloading individual tracks) rather than passively accepting the products marketed by large corporations.

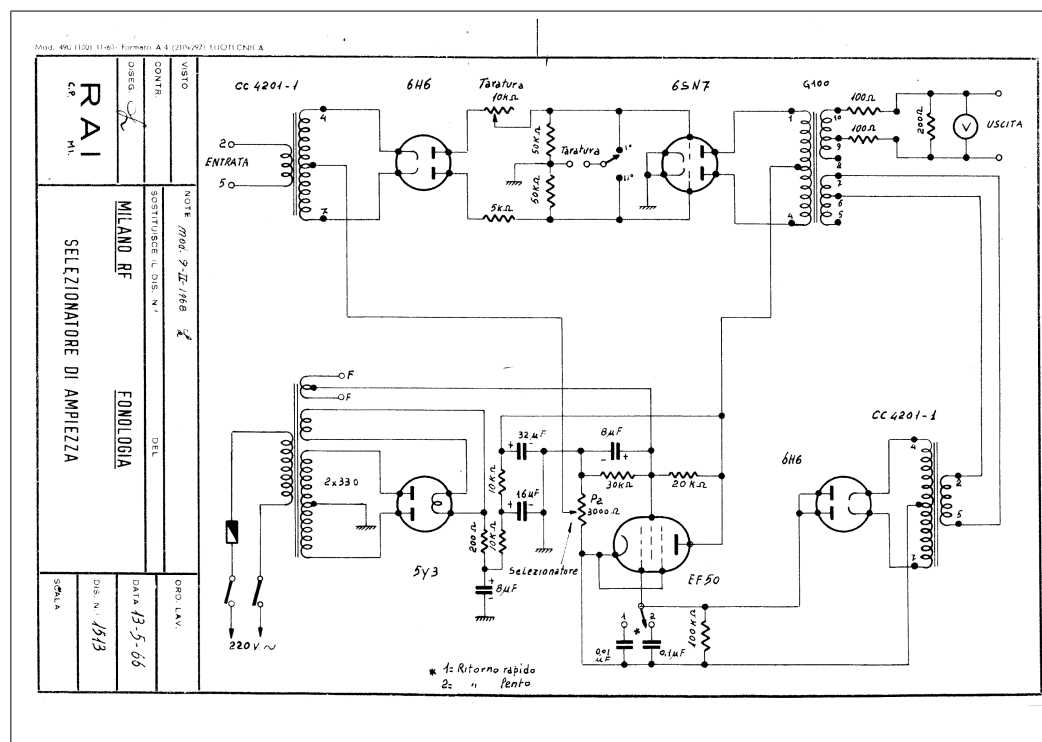


Figure 4. Original project scheme of the Selezionatore di Ampiezza.

4. VIRTUALIZATION OF THE ELECTRONIC LUTHERIE

This section presents the ongoing work finalized at realizing an installation at CSC (Centro di Sonologia Computazionale, University of Padova) consisting of a *music bar* [1, pag. 54], i.e. a SW-HW system that re-creates the electronic lutherie and the production setup that were originally used to compose *Scambi*: an equivalent of the concept of the *access copy* in the field of audio document preservation [11, 12]. The installation will be at disposal of visitors of CSC, who will have the possibility to directly interact with the virtual versions of the electronic instruments (sound generators, filtering devices, etc.) in the same way as electronic musicians did in the 1950's and 1960's, and will, therefore, be able to better understand the functioning and the possibilities of this kind of equipment. In particular, visitors will be able to compose their own versions of *Scambi* using the virtual equivalents of the original equipment and material employed by the composer, eventually realizing Pousseur's vision after more than 50 years.

4.1 Analysis and reverse engineering

Analysis of the analogue electronic devices is being conducted using the original projects of the devices, which are available at the Milan museum of musical instruments. Preliminary inspections indicate that the original project schemes are generally well documented in terms of characteristics of the basic components and of circuitry. In some



Figure 5. The front panel of the *Selezionatore di ampiezza* (photo courtesy of M. Novati [8]).

cases where the documentation is not detailed enough, it is necessary to inspect directly the original instruments through *ad-hoc* measurements on some components. Finally, in order to assess the accuracy of the analysis, the behavior of the devices has to be simulated using electronic engineering tools (particularly Spice [6]).

The circuit of the *Selezionatore di ampiezza* utilized by Pousseur in the composition of *Scambi* is depicted in Figure 4. The figure reproduces the RAI project schemes, which are slightly different from the ones originally presented by Lietti in [5].

The circuit has two operating modes, which depend on the activation status of the EF50 pentode.

1. When the pentode is off, no current flows through the potentiometer P2, so that the secondary of the input transformer CC4201 is connected to ground. In



Figure 6. Rear view of the *Selezionatore di ampiezza* (photo courtesy of M. Novati [8]).

this case, the input signal, scaled by the input transformer, passes unchanged through the twin diode 6H6. The following bridge, composed by three resistances and the potentiometer P1, renders the signal symmetric: by means of the connectors and the switch positioned in the rear of the device (see Figure 6) it is possible to tune the potentiometer P1 until the amplitudes in the upper and in the lower side of the bridge are equal. Finally, the dual triode 6SN7 amplifies the signal to drive the output stage.

2. When the pentode is on, the current flowing through the potentiometer P2 polarizes the secondary of the input transformer to the voltage V_p (depending on the position of the potentiometer). As a result, the current will flow through one of the diodes of the 6H6 tube only when the voltage of the input signal is, in absolute value, greater than the bias voltage V_p . If on the other hand the amplitude of the input voltage is less than V_p , the twin diode 6H6 is off and the output voltage will be zero. The knob at the bottom left of the front panel of the device (see Figure 5) lets the operator control the resistance value of P2 and the V_p threshold.

The activation status of the pentode EF50 depends by the feedback circuit: the output signal is drawn from the connectors 6 and 7 of output transformer G100, it is rectified by the twin diode 6H6, it is filtered by the RC circuit and, finally, is applied to the suppression grid of the pentode EF50.

If a signal is present in the output stage, the twin diode 6H6 is on and the current flows through the RC circuit, biasing the suppression grid to a negative potential, in respect to the cathode. In this condition, the flow of current is inhibited and the pentode is off. Conversely, when there is no signal in the output stage, no current flows through the RC circuit and then the grid will be at the same potential of the cathode. Under these conditions, the pentode is on. The biasing of the pentode is provided by the power supply circuit, that rectified the alternate power supply through the tube 5Y3. The speed at which changes the pentode is

switched on and off depends on the speed at which the RC circuit responds to changes in the feedback signal, i.e. on the time constant of the circuit $\tau = RC$. The switch at the bottom right of the front panel (see Figure 5) lets the operator select between two time constants: $\tau_1 = 0.001s$ and $\tau_2 = 0.01s$.

4.2 Simulations

The circuit of the *Selezionatore di ampiezza* has been replicated in Spice. To this end, datasheets and libraries for all the circuit components have been found. Figure 7 shows a snapshot of the resulting Spice replica of the original circuit.

The final version of the paper will provide results from simulations, in particular the output of the circuit in response to

- sinusoidal signals with slowly varying amplitude and frequency;
- white noise with slowly varying amplitude.

5. CONCLUSION

The advent of digital technologies allowed to overcome many of the technical limitations of analog electroacoustic devices. However the question is whether the electroacoustic community is exploiting these digital resources for new experiments in form. The authors strongly believe that now the composers are able to explore in exhaustive way the potential of open forms using new media and new Human Computer Interfaces. But, in order not to constantly “re-invent the wheel”, works such as *Scambi* must be regarded as being more important now than fifty years ago.

In this sense, the authors are developing the *Music Bar* for active listeners. starting from the original project and schemas of *Selezionatore di ampiezza*, the authors developed a system that allows the user-performer-composer to surf among the existing performances of *Scambi* and to create his own. Specifically, the installation will allow users to creatively interact with (i) virtual counterparts of the electronic devices of the Studio di Fonologia, and (ii) the production system of *Scambi* realized by Pousseur. The user-performer-composer will be able to surf among the existing performances of *Scambi* (e.g. by Luciano Berio and others), and to create his own, by selecting the original audio sequences used by Pousseur, and following (or not) the connecting rules proposed by the composer.

Future work will be devoted to the development of accurate and efficient virtual analog models of the original devices. Recently proposed techniques for the efficient simulation of nonlinear electric systems will be employed [13], and results from spice simulations of the circuits will be used to evaluate the accuracy of the virtual analog models.

A second key point for the effectiveness of the final installation is the design of the user interface. As future work, the authors intend to develop a tangible interface, able to recreate the corporeity, the materiality of the original interfaces: the inherent latencies between the user gestures and the corresponding effects on sound generation; the resistance and viscosity of the tape, which was slowed

by hand by the composer-performer; and so on. All these physical characteristics influenced the composer and his way of interacting with the devices, and need to be preserved in their virtual counterparts.

6. ACKNOWLEDGEMENT

This work was partially supported by Culture2007 project EA- CEA 2010-1174/001-001: *DREAM – Digital Reworking/reappropriation of ElectroAcoustic Music*.

7. REFERENCES

- [1] H. Pousseur, "Scambi," *Gravesaner Blätter*, no. IV, pp. 36–54., 1959.
- [2] J. Dack, "The 'open' form – literature and music." Paper presented at the 'Scambi Symposium', 2005.
- [3] U. Eco, *The role of the reader: explorations in the semiotics of texts*. USA: Indiana University Press, 1979.
- [4] P. Decroupet, "Studio di fonologia musicale della rai milano," *Elektronische Musik*, vol. 2, no. VI, pp. 99–104, 1997.
- [5] A. Lietti, "Soppressore di disturbi a selezione di ampiezza," *Elettronica*, vol. 5, pp. 1–3, Sep. 1955. In italian.
- [6] L. W. Nagel and R. A. Rohrer, "Computer analysis of nonlinear circuits, excluding radiation," *IEEE Journal of Solid State Circuits*, vol. SC, no. 6, pp. 166–182, 1971.
- [7] V. Valimaki, F. Fontana, J. O. Smith, and U. Zölzer, "Introduction to the special issue on virtual analog audio effects and musical instruments," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, pp. 713–714, Apr. 2010.
- [8] M. Novati, ed., *Lo Studio di Fonologia – Un diario musicale 1954-1983*. Milano, Italy: BMG Ricordi Publications, 2009. In italian.
- [9] H. Pousseur, "Henri pousseur." Booklet accompanying CD, BVHAAST 9010, 2005.
- [10] M. Wilkinson, "Two months in the 'studio di fonologia'," *The Score*, pp. 41–48, February 1958.
- [11] S. Canazza and A. Vidolin, "Preserving electroacoustic music," *Journal of New Music Research*, vol. 30, no. 4, pp. 351–363, 2001.
- [12] IASA-TC 03, *The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*. IASA Technical Committee, 2005.
- [13] F. Fontana and F. Avanzini, "Computation of delay-free nonlinear digital filter networks. Application to chaotic circuits and intracellular signal transduction," *IEEE Trans. Sig. Process.*, vol. 56, pp. 4703–4715, Oct. 2008.

NEW POSSIBILITIES FOR ELECTROACOUSTIC MUSIC PERFORMANCE

Cat Hope
Western Australian Academy of Performing Arts
Edith Cowan University, Western Australia
c.hope@ecu.edu.au

ABSTRACT

Western Australian new music ensemble Decibel has an ongoing research project dedicated to performing music that combines acoustic and electronic instruments. In the process of revitalising pieces that have been considered un-performable due to limitations in technology at the time of composition, or certain technologies becoming obsolete, Decibel has developed a unique approach to new music performance involving electronic and acoustic instruments. This has also involved the re-working of electronic pieces not intended to be performed live, works that have previously proved difficult to perform, and the 'electroacoustification' of acoustic works. The ensemble combines old technologies such as reel-to-reel tape machines with newer approaches to music making using interactive programming and networked environments.

This paper investigates possibilities for the configuration of electronic devices in chamber music with acoustic instrument performers, arguing that through the development and implementation of a series of methodologies for performance, Decibel is able to create a new kind of pure 'electroacoustic music' where electronics and acoustics are truly blended on a live concert platform.

1. INTRODUCTION

Everywhere we remain unfree and chained to technology, whether we passionately affirm or deny it. But we are delivered over to it in the worst possible way when we regard it as something neutral; for this conception of it, to which today we particularly like to do homage, makes us utterly blind to the essence of technology [1].

In the article 'Electroacoustic Performance Practice', created from a lecture given in 1961, German composer Karlheinz Stockhausen attempted to define different areas of electroacoustic performance practice. He described six different techniques; recording, transformation (understood as amplification), pre-formed (or pre-fabricated) music, electronic instruments and combinations of all electroacoustic possibilities known [2]. This list describes the different ways in which electronics may be employed in performance environments and it is re-

markable how, despite the many developments in the nature of electronic technology, these categories still hold fast. Many developments in electronic technologies have meant that music only be experienced using playback can now be adapted for live performance. This applies particularly to works using magnetic tape. The physical splicing of analogue tape is now replaced with digital audio editing, and playback no longer needs the linear medium of tape to be unspooled. This opens up possibilities for the recreation of tape works through digitisation. But there is an important question that arises here, what if the timbral qualities of the original mechanisms, such as tape, are an important part of the composition? These are not unlike the practice of 'authentic performance' for early music instrumentalists. It seems we are entering a new era of period performance; that of 'authentic electronic instruments'. As with other early music performance, research must be done to uncover each composers intentions with the technologies employed. And, unlike other types of period performance (such as Baroque), electronic technologies are not always employed for their timbral qualities, but also for their ability to facilitate certain details of a composition. Once this distinction has been made, the process is somewhat demystified. Electronics can be employed as an instrument; with its own colour, controls and characteristics - or as a facilitator, a mechanism to enable a compositional process. If this distinction is made, the approach to any revisioning of technologies becomes somewhat demystified.

2. PREPARING WORKS FOR PERFORMANCE

Most extant works using electronic components require some degree of revision before performance. This can be because:

- 1) The electronic apparatus or media is no longer available or very unreliable due to its age;
- 2) Software programs are unreadable by contemporary versions, media players or hardware;
- 3) There are improvements to the electronic technology that can provide a better service to the piece;
- 4) The piece needs to be arranged to suit the group lineup.

Apart from elements of reliability, convenience and ease, how do we measure if new technology improves on old?

When it is appropriate to adopt new mechanisms, and when is it appropriate to use original ones? If the categorization proposed above is applied, where the electronic component is identified as an instrument or a mechanism, it becomes somewhat easier to fathom.

Many composers have written about the use of electronics in their works, describing the reasons why they chose electronics at all, especially in earlier works from the sixties and seventies. So preliminary background research often provides fruitful insights to understanding the reason electronic mediums were employed at the composition stage. It may be for timbral reasons: the different sounds electronics generate or effect – in which case, the electronics are being used as an instrument in their own right, in the case of synthesis or electronic instruments. But it is also possible that electronics are used as a mechanism, the facilitator of a compositional idea.

Works such as Alvin Lucier's seminal *"I am Sitting in a Room"* (1969) exemplify this division. The score suggests live performance, which would have been very difficult at the time of its composition [3]. Using tape to record and rerecord, the splicing or even rewinding of the recording would create a dead space in the performance of the work that does not exist on any of the recordings. However, the availability of digital – non linear - tools facilitate this compositional process as proposed by Lucier. When Decibel prepared this piece for live performance, MaxMSP offered a simple solution for the 'mechanisms' of the piece, that is, the playing back and re-recording of the original spoken text used in the work. But the question that arose regarding the employment of tape in this work was it's timbral quality integral to the sound world of the work, or is it employed simply as a 'facilitator' of a compositional idea? The sound of a recording on tape has a more significant noise floor, and the sound quality is very different from that reproduced from a computer sound card. Lucier has commented about how he used tape in this work:

I didn't choose to use tape, I had to, because in order to recycle sounds into a space, I had to have them accessible in some form. Tape then wasn't a medium in which to compose sounds, it was a conveyor; a means to record them and play them back one after another in chronological order [4].

This comment above implies Lucier is not particularly interested in that timbral quality difference, and the focus of the work is the way the acoustic of the room overcomes the original piece of text spoken in it, rather than the quality of the original recording or its playback. The work was originally created in mono, and Decibel kept this element of the work in tact; using one speaker and one microphone, each re-recorded iteration appearing immediately after the last.

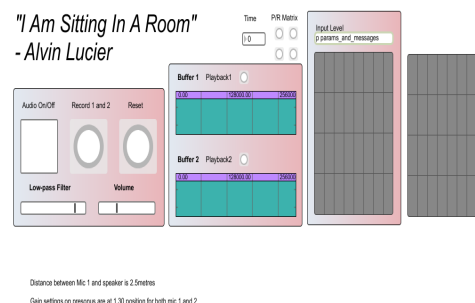


Figure 1: MaxMSP patch for Alvin Lucier's *"I am Sitting in a Room"*, developed by Stuart James, 2010.

Yet Lucier often specifies his electronic instruments. In his work *"Ever Present"* (2002) for three instruments and 2 sine waves, he includes a recording of the sine waves with the score, expressing concern for the quality and accuracy of the sine tone in performance [5]. But even since Material Press published this work, things have changed. Corresponding with the publisher about this work's performance; Daniel Wolf remarked, "MaxMSP can create an equally good quality tone nowadays [6]. In this case, this 'upgrading' of technology has opening other possibilities for the work, namely the performativity of the tone. By creating a simple MaxMSP patch for the work, a performer may control the tone as part of the chamber ensemble, rather than as an audio engineer on the periphery of the stage. This has also meant that the tones could be 'performed' on laptops, in the performance space. This way, the 'instrument' keeps its sonic integrity whilst being introduced into the ensemble as a bone fide instrument in the performance space. To facilitate this process, Decibel created a MaxMSP score reader integrated into the sine tone player, as seen in Figure 6. On Lucier's original score, the tones are notated using a series of note heads with glissando indicators between them, indicating a more inclusive, musical inclination for the electronic instruments. The new 'screen score' contains the sounds for the sine wave parts, and enables the musicians to play these tones whilst reading the score. The acoustic instrumentalists may locate and interact with the tones as they do with each other, communicating with the performers, blending amongst the colours of more traditional instruments. In this way, the electronic components become true instruments as part of the ensemble.



Figure 2: A page from the “*Ever Present*” score on a Max/MSP patch showing the notated sine tone parts (top two stave), and a simple sine tone player operated manually by a performer, relating pitch to noteheads and the tempered scale. Max/MSP patch by Lindsay Vickery.

3. SITUATING ELECTRONICS IN THE ACOUSTIC SCENE

At the end of Stockhausen lecture, he is asked, “Would you also find it an acceptable solution to perform traditional music with amplification and sound projection through loudspeakers?” [7]. A good question, considering the amount of words dedication to sound projection in the article. Stockhausen asks where is the line drawn, and goes on to discuss the qualities projected sound can offer a chamber music performance, noting “I would in every case bring out subtleties, project them with vivid transparency, bring them into a physically perceptible proximity, and strive for the audibility of the musicians [8]. Whilst the above two examples demonstrate the value of real time processing to works with electronic parts,” and the value of electronics as performable instruments with their own unique sound output in the concert space, it does not necessarily explore the possibilities of electronics for acoustic works. The potential of electronics is not always manifest in this ‘single voice’ approach – one characteristic of electronic instruments such as the laptop is the ability to diffuse sound far from the source and to multiple outputs. To this end, Decibel adapt acoustic works with spatial qualities into the electroacoustic realm.

One of these was by Australian composer, Rainer Linz, entitled “*Walk on Parts*” [9]. This work was written for a number of clarinetists, who would walk about the stage and play parts of the score as soloists, in duos, trios and other configurations depending on the size of the group, and decided before the commencement of the performance by noting combinations in a box above the

stave (Figure. 3). After consulting with Linz, Decibel decided to perform this piece with one clarinetist, performing a number of different kinds of clarinet – the Ab, Bb, Eb, bass and contrabass clarinets. The score is for ‘solo or reed ensemble’ and terms such as ‘soloist and ensemble’ “group parts” or ‘tutti’ appear throughout. Decibel replaced many of the live instruments with pre-recorder ones, sending them to different speakers placed in different parts of the auditorium. The parts were pre-recorded in a dry recording studio and then diffused throughout 5 speakers in the space by a laptop operator during the live clarinet performance, according to predetermined combinations decided by the performers. In this way the performer himself reappears around the room in multiple locations - but only sonically. He must cue his performances with the computer operator, who triggers the pre-recorded material. Whilst the clarinetist walked from one clarinet/music stand to the other, the recordings of himself playing other parts appear somewhere else, and often join him in the duo’s and trios as suggested on the score. In this way the acoustic work was re structured to become electro acoustic, with a single computer operator on stage ‘performing’ the different clarinet parts around the space, making decisions in real time where to send sounds in relation to the clarinetists movements. The laptop operator controls the selection and diffusion of the other parts, without being the creator of them, yet he is an integral part of the duet on the stage. Here the electronics create a new dimension to the piece, without interfering with compositional processes in any way.

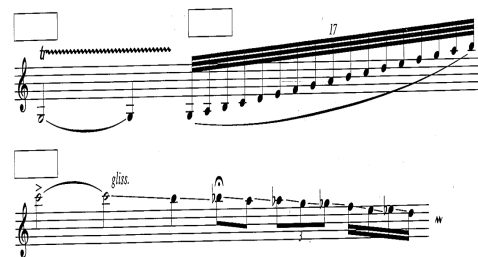


Figure 3: An excerpt from Rainer Linz’s “*Walk On Parts*”, showing the boxes that indicate what combination of instruments should be used for each phrase.

4. RE-CREATING ELECTRONIC COMPOSITIONS AS ELECTROACOUSTIC WORKS

One kind of electro-acoustic music that Stockhausen does not include could be the adoption of ‘pre formed’ electronic works to live performance. He claims that the attitude to the creation of preformed electronic music is radically different from that of live performance, claiming that such works can “never again be made-not corrected, repaired, or made over again”[10]. Many preformed compositions are created as such because this was the only way they could exist, not only because they required the high level of refinement Stockhausen proposes. Complex layering procedures, room sized computer processors, unwieldy instruments or mechanisms are just some of the

reasons pre formed works exist that way. But in addition to a simple re-creation of electronic works for the stage, as in “*I am Sitting in a Room*”, or the ‘electroacoustification’ of works such as “*Walk On Parts*”, pre formed works can also be adapted with a combinations of electronic and acoustic instruments and mechanisms.

This was the process undertaken in Decibel’s adaptation of Brian Eno’s “*Music For Airports*”, track 1/1, for live performance [11]. Decibel are not the first to arrange this work, it has been performed and recorded by US new music ensemble Bang on A Can, who created a literal transcription of the electronic score for four instruments and choir, done in consultation with Eno and released as a CD of the same name, and others [12]. Whilst a ‘score’ was reproduced on the original LP release of this work, it seems little more than an idea of the work; it in no way ‘instructs’ on how to perform the piece. Unlike for “*I am Sitting In A Room*”, the qualities of tape seemed integral to any rendition of this work, since it was the very fragility of the medium that was a key compositional component to the work. Long tape loops speeding up and slowing down ever so slightly gave this work the qualities it is famous for, and have been a corner stone of Eno’s compositional career. The work, initially collaboration with Eno, Robert Wyatt and Rhett Davies, was designed to be continuously looped as a sound installation in an airport. As Eno was defining ambient music as something “like an atmosphere, or a surrounding influence or a tint...suited to a wide variety of moods and atmospheres” [13]. It was obvious his interest in music rotated around sound, texture and noise, rather than melodic structures [14]. These would need to be key components in this work.

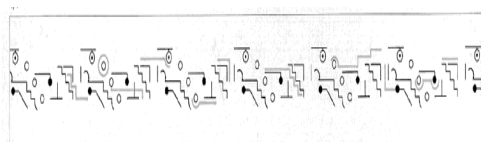


Figure 4: The ‘score’ to Brian Eno’s “*Music For Airports 1/1*”, as it appeared on the LP cover.

As such, Decibel wanted the sonic and mechanical qualities of tape to be part of the ensemble. Both the facilitation (mechanism) and sound quality and fragility of tape seemed imperative to any presentation of the work. So, in addition to four acoustic instruments, three reel-to-reel tape players were situated on stage, each playing a tape loop of around sixteen meters, containing around 2 minutes of audio each. As with the Linz, these were pre-recorded, but then transferred to the tapes as loops. Each reel-to-reel tape machine had its own loud speaker placed below it, and was tuned carefully using variable speed control to maintain tuning. The directional nature of speakers in this and the aforementioned works was taken into consideration as a characteristic of the mechanism, and situated in the space accordingly, for an optimum mix with the live, unamplified players. The phasing of tape loops of a slightly different length is a key to this work – they come in and out of synchronisa-

tion due to the slightly different lengths of the tape and the natural movement of the medium. By using tape players on the stage, this characteristic was maintained as a kind of live, aleatoric compositional element. The tapes were pre-recorded by the same performers featured in the live performance - alto flute, piano, and cello- using three transcribed musical excerpts from the work. These performers started the tapes as they walked on stage. A violinist was also featured, as a soloist around half way through the work, a characteristic of the original recording. The performers improvised around the themes and keys of the original, weaving in and out of the tape parts over around fifteen minutes.



Figure 5: Putting Tape machines in place for a performance of Brian Eno’s “*Music For Airports*” for a performance in September, 2009. Photograph by KFord.

The result was a particularly moving and sensitive rendition that remained faithful to the spirit of the work: the fragility of the original tape technology combined with compositional elements inherent in the work, not only from the harmonic and melodic materials, but also from the tape loop movement and audio quality.

5. CONCLUSION

Through a process of practice as research, Decibel has created methodologies that permit a clearer and richer involvement of electronics in chamber music performance. Through the development of their own assessments for the role of electronics in different works, Decibel has revitalized and refined the electroacoustic chamber music concept. Defining electronics in electroacoustic works as facilitators or instruments assists in prioritizing the behavior of different members in the ensemble and ultimately results in a more integrated electroacoustic music presentation. Through a performance practice informed by musicology and new electronic media developments, Decibel have revitalized key works in the electronic music oeuvre, but also welcomed new ones into it.

The personification of electronic instruments such as laptops, speakers and tape players with individual performers enables them to be more musically involved in the chamber music performance experience. Connecting each performer directly to their own sound output also assists in this relationship, handing the responsibility of

sound quality directly and volume to the performer, something acoustic instrument performance have controlled for years. By avoiding the default to stereo public amplification and external operators for sound Decibel create performers that are characteristically personal and musical.

Using Stockhausen's categories declared in 1961 has illustrated the potential for electro acoustic music to fold and integrated these areas into each other. The adaption of pre-formed electronic as well as acoustic music works galvanise electroacoustic performances, attracting new audiences for electro-acoustic music performance.

Note: Decibel is a new music ensemble directed by the author. Performers and programmers include Lindsay Vickery (reeds), Cat Hope (flutes), Malcolm Riddoch (electronics), Stuart James (piano, percussion), Tristan Parr (cello). More information on them can be found at <http://decibel.waapamusic.com>

6. REFERENCES

- [1] Heidegger, M. (1977). 'The Question Concerning Technology' in *Basic Writings*. San Francisco: Harper.
- [2] Stockhausen, K. & Kohl, J. (1996). 'Electroacoustic Performance Practice', in *Perspectives of New Music*, Vol. 34, No. 1, p. 75.
- [3] Lucier, A (2005). *Reflections. Interviews, Scores, Writings 1965–1994*. (2nd Ed). Koln: Music Texte, p.83.
- [4] *Ibid*, p. 8.
- [5] Lucier, A. (2002). *Ever Present*. Music score. Berlin: Material Press.
- [6] Wolf, D. Private Communication (email). February 2010.
- [7] Stockhausen, K. & Kohl, J., p.86.
- [8] *Ibid*, p. 87
- [9] Linz, R. in Various. (1996). *Reeds: flexible scores for Woodwinds*. Music scores. Sydney: Red House Edition.
- [10] Stockhausen, K. & Kohl, J., p.91.
- [11] Eno, B. (1978). *Ambient Music 1: Music For Airports*. LP. London: EG records.
- [12] Bang On A Can. (1998). *Music For Airports*, New York: Point Music.
- [13] Eno, B. liner notes.
- [14] Tamm, Eric (1989). *Brian Eno: His Music and the Vertical colour of sound*. London: Faber & Faber, p. 14.

ANTHÈMES 2: UN APPROCCIO MONODIREZIONALE AL LIVE-ELECTRONICS

Marco Marinoni
Conservatorio di Musica “B.
Marcello” di Venezia
marco.marinoni@virgilio.it

ABSTRACT

In questo articolo si opera un’analisi di *Anthèmes 2*, un lavoro di Pierre Boulez datato 1997 per violino e dispositivo elettronico [1], della durata approssimativa di 20 minuti.

Il brano del compositore francese, di estrema rilevanza per ciò che concerne la storia della composizione legata all’uso delle nuove tecnologie, è stato fatto oggetto di analisi musicale e tecnologica condotte sulla partitura al fine di individuare le strategie e i processi compositivi messi in atto. I dati emersi sono stati in seguito posti a confronto con le dichiarazioni del compositore: la sostanziale concordanza tra queste ultime e le evidenze analitiche ha permesso di costruire un quadro integrato di inferenze circa l’approccio bouleziano al live electronics, evidenziandone linee guida (asimmetria nel rapporto tra strumento ed elettronica; monodirezionalità dell’interazione; funzione demiurgica del compositore; utilizzo della ridondanza all’interno di una drammaturgia; finalità costruttiva del dialogo tra funzione mimetica/elettronica e funzione diegetica/materiali strumentali) e finalità (manipolazione mirata dei meccanismi psicologici della fruizione).

Ampia parte delle dichiarazioni rilasciate da Boulez in occasione della prima esecuzione del brano (21 Ottobre 1997, IRCAM, Parigi) al filosofo e musicologo francese Peter Szendy è stata appositamente tradotta in italiano e inclusa in appendice.

METODOLOGIA

L’approccio analitico utilizzato è conseguente a un’impostazione parametrico-estesica basata sui cambiamenti contrastivi, secondo la direzione indicata da Michel Imberty [2] per cui il processo di segmentazione di un brano musicale viene strutturato a partire dalla percezione

di cambiamenti qualitativi più o meno pregnanti nel flusso del tempo musicale. Il cambiamento contrastivo, per essere percepito, necessita che l’Io percepisca non solo gli stati A e B ma la *transizione* da A a B. Il passaggio costituisce la realtà percettiva della relazione tra le parti. B deve presentare una *qualità diversa* rispetto ad A. Il cambiamento introduce una *discontinuità* nel tessuto temporale attraverso due possibili modalità: *gerarchia* e *giustapposizione*.

Attraverso la segmentazione in questo modo attuata, il brano viene descritto prima a livello della macro-forma, articolando una osservazione preliminare di tipo paradigmatico in cui la metodologia varia flessibilmente a seconda dell’oggetto; l’osservazione scende quindi al livello della micro-forma, con l’obiettivo di individuare le cellule tematiche strutturali e la ripartizione dei ruoli a livello morfo-sintattico.

Parallelamente all’analisi dei materiali strumentali viene indagato il ruolo dell’elettronica ponendo a confronto il suo decorso formale con quello dei materiali strumentali, alla ricerca di congruenze o divergenze cui dare significato.

Questa *metodologia mista* è coerente con quanto affermato da Boulez [3] riguardo all’importanza di osservare un brano anche in base a come esso viene *percepito* piuttosto che solamente per come esso è *costruito*. In questa sede non viene pertanto attuata una esplorazione di tipo seriale focalizzata sul parametro altezza, già esaurientemente portata a termine da Goldman [4] nella sua tesi.

Attraverso l’indagine dei rapporti gerarchici (asse sintagmatico) e delle relazioni orizzontali che interessano il decorso formale del brano (asse paradigmatico) vengono compiute inferenze qualitative circa le scelte operate dal compositore, integrando l’osservazione analitica con un intervento di tipo *ermeneutico*.

Questo approccio riprende quello utilizzato da me nell’articolo *Atomi distratti* di Mario Garuti [5] in cui, attraverso un’analisi percettivo-paradigmatica si perviene a un intervento ermeneutico operato direttamente sulle scelte compositive, le quali divengono in tal modo intelligibili e dotate di significato.

Le dichiarazioni del compositore, nel caso di *Atomi distratti*, sono state utilizzate come linee guida per

Copyright: © 2010 M. Marinoni. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

l'organizzazione dell'osservazione analitica. In questo lavoro, diversamente, le parole di Boulez vengono utilizzate come motore di verifica e confronto *post hoc*.

Nelle due APPENDICI sono inclusi i seguenti materiali:

- APPENDICE I: tabella riassuntiva della classificazione degli oggetti
- APPENDICE II: traduzione italiana dell'intervista rilasciata da Boulez in occasione della prima esecuzione di *Anthèmes 2*.

ANALISI DEI MATERIALI

Il brano è suddiviso in quattordici regioni.

REG.	DENOM.	TIPOL.	NUM. BATT. / DESCRIZIONE
1	-	Introduzione	3
2	/I	interludio 1	zona libera + 2 campi cronometrici (el.)
3	I	sezione 1	14
4	I/II	interludio 2	zona libera + 2 campi cronometrici (el.)
5	II	sezione 2	118
6	II/III	interludio 3	zona libera tripartita + 2 campi cronometrici (el.)
7	III	sezione 3	58
8	III/IV	interludio 4	zona libera + 2 campi cronometrici (el.)
9	IV	sezione 4	39
10	IV/V	interludio 5	zona libera bipartita + 2 campi cronometrici (el.)
11	V	sezione 5	29
12	V/VI	interludio 6	zona libera tripartita + 2 campi cronometrici (el.)
13	VI	sezione 6	208
14	-	coda	zona libera suddivisa in cinque eventi + 1 oggetto con funzione di chiusura

Tabella 1. Segmentazione.

Vengono ora descritte le singole regioni. Per ogni segmento del brano, il contenuto musicale viene fatto oggetto di osservazione paradigmatica prima e sintagmatica in seguito, integrando i dati relativi alla parte elettronica.

2.1 Introduzione e Coda

2.1.1 Regione 1

Il tactus parte con una croma = 92 (battuta 1), subisce un rallentamento fino a croma = 66 (battuta 2) e rimane stabile (battuta 3). L'indicazione agogica iniziale è "Libre", con la dicitura aggiuntiva "brusque". In batt. 1 tre fram-

menti melodici parzialmente scalari con direzionalità discendente (**a**) precedono un trillo (**b**) di semitono su una breve; in batt. 2 vi sono bicordi (**c**) *iterati* con variazione del parametro altezza, intervallati da pause, eseguiti con arco battuto (crine); in batt. 3 troviamo un oggetto complesso costituito da due linee discontinue (**d**) parzialmente irregolari che confluiscono in un unisono, con dinamica discendente; in chiusura di battuta tre un oggetto puntiforme (**e**) (lo stesso unisono) eseguito con arco piatto battuto in *ppp*.

2.1.2 Regione 14

Nessuna indicazione metronomica; l'indicazione agogica è "Libre", come per la Regione 1. Questa sezione si suddivide in cinque parti più un oggetto puntiforme (**e'**) con funzione di chiusura. Nella prima parte il violino esegue un trillo (**b**) di semitono con arco sul tasto; la stessa nota tenuta (**f**) senza trillo costituisce la seconda parte; nella terza parte, mentre l'altezza precedente si spegne in un diminuendo al niente sulla seconda corda, troviamo un armonico di quinta tenuto (**f'**) in pianissimo eseguito sulla quarta corda: si tratta dello stesso oggetto osservato in Regione 2; questo stesso bicordo si spegne in un diminuendo al niente nella quarta parte; la quinta parte è costituita da una corona lunga su un respiro (**z**). Il brano si chiude su un oggetto puntiforme costituito da un bicordo di semitono eseguito con legno battuto in *pppppp*. Questo oggetto è una variante dell'oggetto **e** che chiude la Regione 1, sebbene non si tratti di un unisono D5 ma di un bicordo C#5-D5, non più in *ppp* ma nella dinamica più bassa possibile. Nel passaggio dal primo al secondo, la *defocalizzazione* del parametro altezza pare aver causato una perdita di energia dinamica. Per questa sezione ritengo sia possibile parlare di tripartizione, con gli eventi raggruppati a due a due.

OGGETTI	
Classif.	Descrizione
a	frammento melodico con direzionalità discendente
b	trillo di semitono o di tono
c	bicordi crine arco battuto
d	linee discontinue confluenti
e	oggetto puntiforme unisono arco battuto c./I
e'	oggetto puntiforme bicordo arco battuto c./I
f	nota tenuta
f'	armonico 5a tenuto
z	Respiro

Tabella 2. Regioni 1, 14: classificazione degli oggetti.

Gli oggetti rinvenuti in queste due sezioni sono descritti in Tabella 2. La morfologia delle due regioni osservate è indicata in Tabella 3.

REG.	MORFOLOGIA
1	aaab/ccccccc/de

14	bff'ze'
----	---------

Tabella 3. Regioni 1, 14: morfologia.

L'elettronica, nelle regioni 2 e 14, implementa i seguenti processi:

REG.	PROCESSI
1	1. riverberazione infinita 2. sampler con riverberazione infinita 3. sampler (audio tracks) 4. frequency shifting
14	1. sampler con riverberazione infinita 2. sampler con riverberazione infinita

Tabella 4. Regioni 1, 14: live electronics, processi.

Nella Regione 2, il Processo 1 si applica ai materiali strumentali delle batt. 1-2 e permane sino alla comparsa dell'oggetto **e**, come anche gli altri processi; il Processo 2 si innesta sul primo degli oggetti **a** a batt. 1 con un oggetto scalare discendente; il Processo 3 genera elementi scalari direzionali in sincro con gli attacchi degli oggetti **b** e **c** (batt. 1,2); il Processo 4, infine, genera un evento sincronico all'evento **d** (batt. 3), di cui riproduce anche la morfologia (*mimesi* - *ridondanza*). Questa considerazione è applicabile anche all'oggetto prodotto dal Processo 2 e con l'uso del riverbero infinito sul materiale strumentale.

Nella Regione 14 i due processi generano suoni tenuti di tipo **f**, il primo dei cluster e il secondo suoni singoli, entrambi riverberati in modo da costruire elementi lineari orizzontali, più o meno densi, si cui si articola la coda strumentale.

Gli eventi assegnati all'elettronica seguono in parte (sull'oggetto **f** i materiali elettronici si *cristallizzano* e permangono immutati al variare degli oggetti strumentali) la suddivisione in cinque parti della regione, creando un parziale sfasamento tra i differenti livelli (*processo figura-sfondo*).

2.2 Gli Interludi

2.2.1 Regione 2

Questo primo breve interludio è costituito da due parti, la seconda a propria volta bipartita.

La prima parte presenta un armonico tenuto (**f'**) di quinta eseguito sulla quarta corda che quindi subisce un glissando ascendente indeterminato (**g**) con dinamiche in diminuendo al niente.

La seconda parte è formata da due campi cronometrici, di cinque e tre secondi, in cui permane la processazione elettronica dei materiali sonori.

L'elettronica implementa quattro processi:

- harmonizer 1;
- harmonizer 2;

- due modulatori ad anello filtrati;
- riverberazione infinita.

Tali processi sono finalizzati a creare un *meccanismo di defocalizzazione* applicato al materiale strumentale: i due harmonizer mediante urti di semitono distribuiti su ottave differenti e unisoni, i ring modulator processando la risultante sonora del violino su frequenze differenti e la riverberazione generando una coda che permane nei due campi cronometrici.

2.2.2 Regione 4

La struttura del secondo interludio è identica a quella del primo, ovvero un evento strumentale nella prima parte e due campi cronometrici nella seconda. In questo caso l'armonico (**f''**) è di quarta, sempre sulla quarta corda.

Anche l'intervento dell'elettronica è congruente con quello in Regione 2.

2.2.3 Regione 6

Il terzo interludio, ancora una volta bipartito, differisce dai primi due nella prima parte per durata, livello di articolazione e grado di densità; la seconda parte è sempre formata da due campi cronometrici, di durata identica a quelli trovati nelle Regioni 2 e 4.

La prima parte è a sua volta tripartita. La suddivisione interna è operata attraverso l'apposizione di due respiri (**z**). Gli oggetti eseguiti dal violino sono tre, descritti in Tabella 5.

Gli eventi semplici (primari) sono l'armonico tenuto e il glissando ascendente; gli eventi composti (secondari) sono formati dall'armonico tenuto seguito dal glissando ascendente.

L'oggetto 2 (aggregato) può essere interpretato come formato da due eventi semplici (l'armonico tenuto iniziale e il glissando finale) e due eventi composti parzialmente sovrapposti, il secondo presentato per moto retrogrado. Il meccanismo di *crossfade* avviene sul glissando ascendente, con transizione nel punto in cui le dinamiche vanno a zero.

L'oggetto 3 può a sua volta essere interpretato come formato da due eventi composti, il secondo presentato per moto retrogrado, uniti a specchio mediante parziale *crossfade*: anche in questo caso la dinamica va al niente nella parte centrale del glissando, lungo l'asse di simmetria dell'aggregato, come se i due eventi fossero stati parzialmente sovrapposti.

L'elettronica presenta gli stessi processi osservati nelle Regioni 2 e 4, con le stesse finalità e punti di attacco disposti sugli attacchi degli oggetti **f''** ed **f'**.

N° AGGR.	TIPOL.	DESCRIZIONE
1	f'g	Arm. 4° (f'), gliss. ascend. (g)
2	f''f'gf'g	Arm. 4a (f''), arm. 5a (f'), gliss. ascend. (g) → arm. 5a (f'), gliss.

		ascend. (g)
3	f'gf''	Arm. 4° (f''), gliss. ascend. (g) → arm. 4a (f'')

Tabella 5. Regione 6: classificazione degli oggetti.

2.2.4 Regione 8

Il quarto interludio riprende la struttura bipartita dei primi tre, variando il materiale incluso nella prima parte. L'oggetto in questione è un aggregato che si presenta analogo all'oggetto 3 del terzo interludio e anche in questo caso può essere letto come la parziale sovrapposizione di due eventi composti (f'g e gf'') sistemati a specchio, con simmetria centrale per quanto riguarda la morfologia ma rapporto di *anamorfosi* relativamente al parametro altezza.

L'intervento dell'elettronica è analogo a quello osservato nelle precedenti regioni.

2.2.5 Regione 10

Una nuova tipologia di evento compare nel quinto interludio: il glissando discendente (g'). Si tratta di una variante dell'oggetto g.

Questo interludio conserva la struttura bipartita con suddivisione della seconda parte in due campi cronometrici, mentre nella prima parte osserviamo due aggregati, separati da un respiro (z). Il primo aggregato è formato da un armonico di quarta tenuto (f'') seguito da un glissando ascendente (g); il secondo aggregato è formato da due oggetti, il primo composto (armonico di quarta tenuto [f'']) seguito da glissando discendente [g']) e il secondo semplice (armonico di quarta tenuto [f'']). L'azzeramento delle dinamiche a seguito di diminuendo al niente è anche in questo caso il criterio adottato per individuare le suddivisioni interne.

L'intervento dell'elettronica è analogo a quello osservato nelle precedenti regioni, con attacchi sugli oggetti f'.

2.2.6 Regione 12

Anche l'ultimo interludio è bipartito. La prima parte è internamente quadripartita, con le prime tre zone occupate da materiali strumentali e l'ultima da silenzio, con corona. Il primo oggetto è un evento semplice: armonico di quinta tenuto (f''); il secondo evento è un aggregato formato da due oggetti: il primo composto da armonico di quarta tenuto (f'') e glissando discendente (g'), il secondo semplice (armonico di quarta tenuto [f'']); il terzo evento è costituito da un armonico di quarta tenuto (f''), con dinamiche impercettibilmente decrescenti da *ppp* a *pppp*.

L'elettronica in Regione 12 è ancora una volta congruente con i precedenti interludi.

Possiamo operare una classificazione degli oggetti rinvenuti nelle Regioni 2, 4, 6, 8, 10 e 12, escludendo quelli già citati in relazione alle Regioni 1 e 14.

OGGETTI (Regioni 2, 4, 6, 8, 10 e 12)	
Classificazione	Descrizione
f''	armonico tenuto di quarta
g	glissando ascendente indeterminato di armonico
g'	glissando discendente indeterminato di armonico

Tabella 6. Interludi: classificazione degli oggetti.

Le possibilità di combinazione degli oggetti in aggregati vengono descritte in Tabella 7. Si tratta di una classificazione puramente gestuale, quindi non verrà mantenuta la distinzione tra f' e f'', basata unicamente sulle componenti frequenziali. Entrambi gli oggetti verranno, in questo caso, per maggiore chiarezza, chiamati F.

CLASSIFICAZ. AGGREGATO	MORF.	COMPARAZ.
A	Fg	originale
A'	gF	retrogrado di A
B	Fg'	inverso di A
B'	gF	retrogrado di B, retrogrado inverso di A

Tabella 7. Interludi: classificazione degli aggregati.

REG.	MORFOLOGIA	
	oggetti	aggregati
2	f'g	A
4	f''	f''
6	f'gzf''f'gf'gzf''gf''	A; f'', A---B', g; A---B'
8	f'gf''	A---B'
10	f'gzf''g'f''	A; B, f''
12	f'zf''g'f''zf''z	f'; B,f'; f'

Tabella 8. Interludi: morfologia.

Nella sintassi morfologica relativa agli interludi ritroviamo gli stilemi caratteristici della composizione seriale.

Possiamo ora descrivere i sei interludi attraverso la classificazione operata, individuando gli aggregati descritti sopra, limitatamente alla prima parte di ogni interludio, in cui sono presenti materiali sonori prodotti direttamente dallo strumento.

Nella colonna centrale viene descritta la stringa degli oggetti mentre nella colonna di destra vengono evidenziati gli aggregati, quando presenti; la linea tratteggiata orizzontale posta tra due aggregati indica la presenza di un *processo di crossfade* messo in atto tra di essi.

Per quanto riguarda il rapporto tra elettronica e materiali strumentali, è possibile rilevare la presenza di tratti distintivi che ritorneranno, come atteggiamenti e strategie compositive, nelle altre regioni del brano:

- *meccanismo di defocalizzazione*
- *processo figura-sfondo*
- *cristallizzazione*
- *mimesi – ridondanza*

2.3 Le Sezioni

2.3.1 Regione 3 (Sezione 1)

La Sezione I è formata da 16 battute, con croma compresa tra 92 e 98. L'indicazione domanda all'esecutore "beaucoup de flexibilité". Nessuna variazione del tactus occorre all'interno della sezione.

Gli eventi sonori individuabili sono suddivisibili in tre tipologie, ognuna delle quali con suddivisioni interne:

b	trillo semitono o tono	isolato
b'		seguito da una nota più grave, stessa arca-ta
h	suono breve ordinario	con acciaccature
h'		senza acciaccature
h''		gruppi irregolari con acciaccature
i	elemento scalare	direzionale
		discontinuo ascen-
		denti
i'		direzionale
		discontinuo discen-
		denti

Di seguito viene mostrata la classificazione morfologica della sezione, indicando gli aggregati mediante le parentesi quadre e la suddivisione in battute con lo slash.

[hh]i/b'b'/b'/b'/bb/bbbb/b'/b'b'/b/[hhhhh]i/b[h''h''h'']/
[h''h''h'']i'/b[h''h'']i/bh''/h''i/b[h''h'']

Dalla classificazione emerge come la Regione 3 sia suddivisibile, secondo un principio di omogeneità morfo-sintattica, in due macro-sottoregioni:

- **3.1** (batt. 1-9);
- **3.2** (batt. 11-16).

La batt. 10 può essere considerata un elemento ponte tra le due macro-sottoregioni.

La macro-sottoregione **3.1** è a sua volta suddivisibile in due micro-sottoregioni:

- **3.1.1** (batt. 1);
- **3.1.2** (batt. 2-9).

La Macro-sottoregione **3.2** è suddivisibile in quattro micro-sottoregioni, come indicato in Tabella 9.

MICRO-SOTTOREG.	BATT.	MORFOLOGIA
3.2.1	11-12	ib[h''h''h'']
3.2.2	13	i'b[h''h''h'']
3.2.3	14-15	ibh''h''
3.2.4	16	ib[h''h''h'']

Tabella 9. Macro-sottoregione **3.2**: suddivisione in micro-sottoregioni.

Le quattro micro-sottoregioni interne alla **3.2** sono bipartite, costituite da un *antecedente* a sua volta bipartito, formato da un elemento scalare direzionato e da un trillo, e da un *conseguente* che include gruppi irregolari con acciaccature organizzati in aggregati tramite la suddivisione in arcate e uniti a livello di frase, con eccezione per i due aggregati a cavallo di batt. 15.

Riprendendo il tipo di osservazione operato per gli interludi, la **3.2.2** ha funzione di inverso rispetto le altre tre micro-sottoregioni, che a loro volta sono sottoposte a *processi di micro-variazione* (es. elemento conseguente in **3.2.2** e **3.2.3**) e *diminuzione / aumentazione* (es. elemento conseguente in **3.2.1** rispetto alle altre micro-sottoregioni).

L'elettronica, in questa prima sezione, opera quattro processi:

1. quattro harmonizer
2. sampler
3. sampler con riverbero infinito
4. frequency shifting

A questi si aggiunge il processo di spazializzazione, applicato ai materiali strumentali e ai singoli processi di trasformazione e produzione elettronica del suono.

Il Processo 4 è attivo sugli oggetti **i** e **i'**.

Il Processo 3 genera una nota in pizzicato, riverberata lungo tutta la macro-sottoregione **3.1** e parte della batt. 10. Nella **3.2** la stessa nota viene prodotta quattro volte, coerentemente con la suddivisione in micro-sottoregioni e interrompendo la riverberazione in concomitanza con l'attivarsi del Processo 4.

Il Processo 2 si attiva su **3.1.2** e genera quindici eventi sincroni agli attacchi dei quindici oggetti **b** e **b'**. Nella macro-sottoregione **3.2** gli eventi sono prodotti in sincronamente con l'attacco dei quattro oggetti **b**, secondo la suddivisione formale interna.

Il Processo 1 produce un evento sull'aggregato **[hh]**, elemento antecedente in **3.1.1**, quindi in **3.1.2** genera quindici eventi, in sincro con il Processo 2 e con gli oggetti **b** e **b'**, mettendo in atto un procedimento a specchio relativamente al parametro pitch, che rimane immutato nei trigger 8-14 (batt. 5-6), suddividendo in questo modo la micro-sottoregione in due ulteriori aree, con asse di simmetria intorno alle batt. 5 e 6.

La *multi-stratificazione dei processi* nel rapporto tra il materiale strumentale, la processazione elettronica dello stesso e i processi di sintesi mettono in luce anche in questo caso un *meccanismo figura-sfondo* mirato a rendere

maggiormente percettibili da una parte le suddivisioni formali e sintattiche, dall'altra i singoli oggetti che costituiscono la tessitura del brano e i processi di mutazione / transizione che li interessano.

In questo senso, i Processi 1, 2 e 3 si interrompono all'unisono nell'ultima parte di batt. 10, sottolineando con una cesura la bipartizione formale.

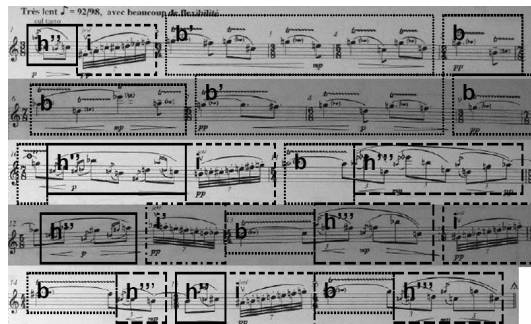


Figura 1. Regione 3: morfologia.

2.3.2 Regione 5 (Sezione 2)

La seconda sezione è costituita da 118 battute con indicazione di tempo variabile. L'indicazione agogica iniziale è "Rapide, dynamique" con croma a 172, "très rythmique, rigide". Nessuna variazione del tactus occorre all'interno della sezione, che è interamente in pizzicato.

Il materiale include note singole (l) (es. batt. 16), bi-cordi (l') (es. batt. 57), tricordi (l'') (es. batt. 30) isolati (es. batt. 57-61) o organizzati in gruppi (es. batt. 53) e figurazioni direzionali (es. batt. 67, 70-71, 88, 92-93) o a-direzionali (es. batt. 64-65). Le modalità di esecuzione del pizzicato includono "alla corda", accentato e sforzato nelle varie combinazioni possibili. La dinamica varia in un range compreso tra *mf* e *ff*.

Il motore di organizzazione è il parametro altezza e la presentazione del materiale procede senza soluzione di continuità lungo tutta la sezione, la cui "rigidità" e "ritmicità" viene messa in risalto attraverso la scelta di un'organizzazione per microframmentazione dell'ambito frequenziale.

Si tratta di un unico oggetto freddo, apparentemente privo di evoluzione, discontinuo ma in maniera continua, articolato, scandito dalla componente rumorosa dei transienti d'attacco, inorganico.

Un *cristallo*, struttura inorganica declinata in un tempo immobile.

L'elettronica implementa tre processi (oltre alla spazializzazione, di cui non si parlerà più, in quanto sempre attiva, sia sui materiali strumentali che su quelli derivanti dai processi elettronici e quindi strumento poco efficace ai fini di un'analisi parametrico-percettiva centrata sui cambiamenti contrastivi).

1. quattro harmonizer + delay

2. sampler
3. sei frequency shifting + delay

Il Processo 1 si attiva a batt. 1 quindi rimane inattivo fino a batt. 62; tra batt. 62 e batt. 73 produce sette eventi, suddivisibili in tre cluster:

- cluster 1: cue 21-24 (batt. 62-66)
- cluster 2: cue 26-27 (batt. 68-70)
- cluster 3: cue 29 (batt. 72)

Tra un cluster e l'altro sono situati eventi parzialmente direzionali, riconducibili all'oggetto **d'**, prodotti dal Processo 3.

Il Processo 3 si attiva a battuta 2, sincronicamente alla fine dell'evento prodotto dal Processo 1 e rimane attivo fino a batt. 62, processando i materiali strumentali secondo parametri fissati a batt. 2 e modificati alle batt. 7, 13, 18, 20, 25, 27, 31, 35, 37, 39, 46, 49, 51, 53, 55, 57, 59, 61. A batt. 73 questo tipo di processazione riprende con nodi nelle batt. 73, 78, 81, 84, 88, 90, 92, 95, 97, 100, 104, 106, 108, 110, 112, 114, 116 e 118.

Il Processo 2 produce eventi in sincro con il processo 3 fino a batt. 62, quindi tace per dieci battute e riprende a batt. 73, di nuovo in sincro con il Processo 3.

Dall'organizzazione dei materiali elettronici emerge una chiara suddivisione della Regione 5 a partire dalla tripartizione descritta in Tabella 10.

La macro-sottoregione 5.2 agisce come asse di simmetria tra le altre due macro-sottoregioni, in cui operano processi congruenti. Il rapporto tra 5.3 e 5.1 è approssimativamente di 4/3.

MACRO-SOTTOREG.	BATTUTE	N° BATT.
5.1	1-62	62
5.2	63-72	10
5.3	73-118	46

Tabella 10. Regione 5: suddivisione in macro-sottoregioni.

E' possibile operare una ripartizione ulteriore, a partire dai punti di attacco delle cue, e suddividere le macro-sottoregioni 5.1 e 5.3 rispettivamente in venti e diciotto micro-sottoregioni, che vengono elencate nelle Tabelle 11 e 12.

Nel grafico in Figura 2 vengono confrontati i dati, espressi in numero di battute, relativi alla durata delle micro-sottoregioni.

E' possibile osservare un movimento congruente di *contrazione-dilatazione* con quattro punti di massima espansione per ciascuna sottoregione e una zona condivisa di contrazione massima localizzata nella zona caudale delle due macro-sottoregioni.

Gli elementi direzionali tipo **d'** prodotti dal Processo 3 sono innestati su elementi analoghi prodotti dallo strumento.

Attraverso l'osservazione e la classificazione dei materiali assegnati all'elettronica è stato possibile individuare una suddivisione formale tripartita a specchio per la Regione 5.

MICRO-SOTTOREG.	BATTUTE	N° BATT.
5.1.1	1	1
5.1.2	2-6	5
5.1.3	7-12	6
5.1.4	13-17	5
5.1.5	18-19	2
5.1.6	20-24	5
5.1.7	25-26	2
5.1.8	27-30	4
5.1.9	31-34	4
5.1.10	35-36	2
5.1.11	37-38	2
5.1.12	39-45	7
5.1.13	46-48	3
5.1.14	49-50	2
5.1.15	51-52	2
5.1.16	53-54	2
5.1.17	55-56	2
5.1.18	57-58	2
5.1.19	59-60	2
5.1.20	61-62	2

Tabella 11. Sottoregione 5.1: suddivisione in micro-sottoregioni.

Questo tipo di organizzazione, pur utilizzando i transienti d'attacco di molti oggetti strumentali come punto di innesto per le cue, è in netto contrasto con l'omogeneità e l'immobilità del tessuto strumentale, fatta eccezione per gli oggetti alle battute 67 e 71.

La sovrapposizione di due livelli morfologicamente in contrasto è riconducibile alla messa in atto del *processo figura-sfondo*, di cui si è parlato in relazione alla sezione precedente.

E' inoltre interessante rilevare l'*interscambiabilità di ruoli* che sussiste a livello compositivo tra la texture strumentale e l'assemblaggio dei processi elettronici, relativamente alla costruzione dell'organizzazione formale.

MICRO-SOTTOREG.	BATTUTE	N° BATT.
5.3.1	73-77	5
5.3.2	78-80	3
5.3.3	81-83	3
5.3.4	84-87	4
5.3.5	88-89	2
5.3.6	90-91	2
5.3.7	92-94	3
5.3.8	95-96	2
5.3.9	97-99	3
5.3.10	100-103	4
5.3.11	104-105	2
5.3.12	106-107	2

5.3.13	108-109	2
5.3.14	110-111	2
5.3.15	112-113	2
5.3.16	114-115	2
5.3.17	116-117	2
5.3.18	118	1

Tabella 12. Sottoregione 5.3: suddivisione in micro-sottoregioni.

Macro-sottoregioni 5.1 e 5.3 - n° batt.

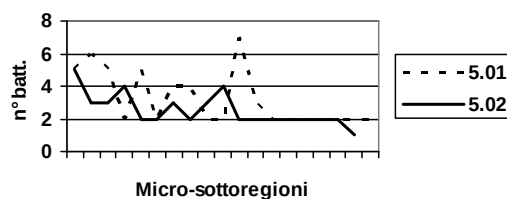


Figura 2. Macro-sottoregioni 5.1 e 5.3: durata delle micro-sottoregioni.

2.3.3 Regione 7 (Sezione 3)

La Regione 7, costituita da 58 battute con indicazione di tempo variabile da battuta a battuta, è internamente suddivisa in nove macro-sottoregioni organizzate a partire da differenti indicazioni agogiche, come mostrato in Tabella 13.

Una nuova tipologia di oggetti compare in questa regione:

n tremolo d'arco con acciaccature
n' senza acciaccature

MACRO-SOTTOR.	BATT.	TIPOL. TRANS.	INDICAZ. AGOGICA
7.1	1-4	CONT.	Lent [croma = 86] régulier
7.2	5-14	DISCONT.	Nerveux, irrégulier [croma = 116]
7.3	15-23	DISCONT.	Plus irrégulier [croma = 112]
7.4	24-33	DISCONT.	Extrêmement irrégulier [croma = 108]
7.5	34-35	CONT.	Sub. lent [croma = 86] régulier
7.6	36-42	PARZ.. DISCONT.	Très calme [croma = 92 ÷ 98], avec beaucoup de flexibilité
7.7	43-49	DISCONT.	Nerveux et

quindi si riattivano all'inizio di 7.7 e producono un nuovo evento all'inizio di ciascuna macro-sottoregione, fino alla fine della regione.

Il sampler che esegue file audio pre-registrati (Processo 7) è attivo solo su 7.6, secondo una funzione complementare rispetto ai processi 3, 4 e 5. L'alternanza di questi processi individua una bipartizione della regione, con nodo su 7.6 e la macro-sottoregione 7.5 come elemento di transizione.

In Tabella 16 vengono rappresentate graficamente le aree di attivazione dei singoli processi, relativamente alla Regione 7.

PROC.	MACRO-SOTTOREGIONE								
	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9
1									
2									
3									
4									
5									
6									
7									

Tabella 16. Regione 7: live-electronics, assegnazione dei processi.

E' possibile a questo punto ipotizzare per l'elettronica una *precisa e interscambiabile ripartizione dei ruoli a livello funzionale*, analoga a quella rilevata nella ripartizione morfologica operata sugli eventi strumentali, individuando le seguenti funzioni:

- *nodo* [ripartizione formale: macro-sottoregione 7.6]
- *transizione* / *interpunzione* [violino: elementi (i), variazioni agogiche (batt. 4, 35), transizioni nelle modalità esecutive (batt. 4, 35), oggetti usati con funzione di transizione (batt. 4, oggetto b); elettronica: frequency shifting (7.1 e 7.5), quattro harmonizer (7.1); ripartizione formale: macro-sottoregione 7.5]
- *anticipazione* / *reminiscenza* [l'oggetto b, utilizzato come elemento di transizione da 7.1 a 7.2, anticipa il contenuto della 7.6]
- *diegesi* [nell'accezione aristotelica, il contenuto narrativo proprio delle singole parti, in cui viene descritto il rapporto tra gli oggetti e le funzioni (es. 7.1, batt. 1, 2 e la prima metà di batt. 3)]
- *mimesi* [in opposizione alla *diegesi*, la "rappresentazione" dei rapporti tra oggetti e funzioni (es. l'andamento sincronico e morfologicamente congruente tra l'oggetto i', seconda metà batt. 3, e l'evento prodotto dal Processo 6, secondo un principio di ridondanza)].

2.3.4 Regione 9 (Sezione 4)

La quarta sezione si costituisce di 39 battute con indicazione di tempo variabile. La forma è di tipo ABA: la prima parte coincide con le batt. 1-12 (macro-sottoregione 9.1), la seconda con le batt. 12-24 (macro-

sottoregione 9.2) e la terza con le batt. 25-39 (macro-sottoregione 9.3).

La macro-sottoregione 9.1 include 24 oggetti che, dal punto di vista morfologico, sono delle varianti dell'oggetto b. In questo caso i trilli (b'') sono indipendenti, senza nota grave di caduta al termine ma la porzione prossimale è costituita quasi sempre da un bicordo o tricordo, strappato sul transiente d'attacco.

In Tabella 17 viene indicata la classificazione degli oggetti.

N° OGG	MORFOLOGIA
1-5	b''
6-7	B
8	b''
9-10	b
11-24	b''

Tabella 17. Macro-sottoregione 9.1: classificazione degli oggetti.

La batt. 12 è un elemento ponte tra la prima e la seconda micro-sottoregione.

Nella prima parte della battuta troviamo l'elemento i, questa volta come gruppo irregolare di 20 note in acciacatura di sessantaquattresimo. Si tratta di un *elemento di interpunzione*, allo stesso modo con cui gli interludi si frappongono alle sezioni vere e proprie del brano. Nella seconda parte della batt. 12 inizia la macro-sottoregione 9.2, con l'indicazione "Rythmiquement stable", e il tactus di 104 alla croma.

La seconda macro-sottoregione è a sua volta suddivisibile, secondo uno schema ABA, in tre micro-sottoregioni, come descritto in Tabella 18.

L'oggetto k è identificabile con un suono breve staccato.

MICRO-SOTTOR.	BATT.	MORFOL.	INDICAZ.
9.2.1	12-16	Reminescenza 9.1 (b'') + 8 oggetti tipo a	Agité, instable [croma = 112 ÷ 132]
9.2.2	17-19	3 oggetti: (a)k, (a)k, kk	Rythmiquement stable [croma = 104]
9.2.3	20-24	8 oggetti + i' + kk + Anticipazione 9.2.3	De nouveau instable

Tabella 18. Macro-sottoregione 9.2: morfologia delle micro-sottoregioni.

Gli otto oggetti inclusi nella micro-sottoregione 9.2.1 sono analoghi agli 8 oggetti inclusi in 9.2.3: frasi di semibiscrome raggruppate frammenti melodici, distinti in arcate, con gruppi irregolari e dinamiche comprese tra *f* e *fff*.

Gli oggetti 1 e 2 (fine batt. 12 e batt. 13) sono elementi borderline in quanto il primo formato da 2 sole note ed il secondo da un suono unico breve con modalità esecutiva ordinaria privo di acciaccatura (**h'**). Tutti gli oggetti di tipo (**a**) presenti in **9.2.1** sono a-direzionali (**a''**) tranne l'oggetto 3, dotato di direzionalità discendente (**a**).

Le batt. 23 e 24 hanno funzione di ponte tra **9.2.3** e **9.3**. A batt. 23 troviamo una variante dell'oggetto **i** (**i''**, elemento scalare ascendente legato, arcata unica) come interpunzione, quindi uno dei tre eventi inclusi in **9.2.2** (**kk** senza accenti) e infine, a batt. 24, un'anticipazione della macro-sottoregione **9.3** (**b''**), speculare alla reminiscenza individuata nella seconda parte di batt. 12.

La macro-sottoregione **9.3** riprende le tipologie di oggetti presenti nella **9.1**. In questo caso possiamo osservare 35 oggetti, tutti di tipo **b''** tranne gli oggetti 17, 21, 26, 29, 33, 34 che rientrano nella tipologia **b**. Le dinamiche variano tra *p* e *ff*.

Per quanto riguarda l'elettronica, i processi attivati includono due sampler con riverberazione infinita, il primo attivo su **9.1** e **9.3**, il secondo su **9.2**.

In **9.1** il sampler produce ventiquattro eventi tipo nota tenuta **f** (pizzicato con riverberazione infinita) sincroni agli attacchi dei ventiquattro oggetti della tipologia (**b**) e in **9.3** trentasei eventi analoghi che si innestano sui trentasei prodotti dal violino.

In **9.1** i suoni prodotti dall'elettronica coincidono sempre con la nota di frequenza più acuta prodotta di volta in volta dal violino, coerentemente con la *funzione mimetica* tra elettronica e materiali strumentali postulata in relazione alla precedente regione. In **9.3** i primi diciotto preservano il rapporto di unisono, l'evento 19 dell'elettronica è dissonante, quindi gli ultimi diciassette riprendono il rapporto di unisono. L'oggetto 19 è sfasato di semitono rispetto a quello prodotto dal violino (D5, Db5). Trattandosi di un evento anomalo rispetto al tessuto circostante e non giustificato a livello di strategie compositive, ritengo opportuno ipotizzare che si tratti di un errore in fase di trascrizione. Per quanto riguarda la notazione dell'evento analogo del violino, il manoscritto riporta la nota Db5, in accordo con la versione a stampa.

Se non si tratta di errore tipografico, la ripartizione morfologica degli eventi assegnati all'elettronica fanno sì che la macro-sottoregione **9.3** sia bipartita, con *elemento nodale* il suono numero 19.

Nella macro-sottoregione **9.2** il secondo sampler genera cinque eventi, in coincidenza con gli attacchi delle batt. 12, 14, 17, 21, 24. In questo caso l'elettronica sottolinea la cesura tra **9.2.1** e **9.2.2**, suddividendo la prima micro-sottoregione in due ulteriori frammenti e la seconda in tre, non coincidenti con la ripartizione formale attuata tra **9.2.2** e **9.2.3**, come evidenziato dalla Tabella 19.

BATT.	VLN.	ELETTR.
12	9.2.1	Cue 5
13		
14		Cue 6

15	9.2.2	Cue 7
16		
17		
18		
19	9.2.3	Cue 8
20		
21		
22		
23		Cue 9
24		

Tabella 19. Macro-sottoregione **9.2**: raffronto tra le suddivisioni formali.

Questo sfasamento, conseguente alla multi-stratificazione dei processi, è volto a rendere ancora più percettibile il *processo di anamorfosi* che si innesta sulla forma ABA, precedentemente individuata sulla base della morfologia degli oggetti, che era già stata negata dal differente grado di densità osservabile nella prima e nella terza sottoregione.

2.3.5 Regione 11 (Sezione 5)

La Regione 11 include la Sezione V ed è internamente suddivisa in sette macro-sottoregioni, con indicazione di tempo variabile da battuta a battuta. Le sottoregioni sono indicate in Tabella 20 e descritte di seguito.

La prima macro-sottoregione è tripartita, con dinamiche comprese tra *pp* e *mp*. **11.1.1** è un'introduzione (oggetti **i** e **b**); la forma delle restanti micro-sottoregioni è 3+1 (ABA+B') con l'ultimo aggregato ottenuto come dimezzamento del secondo.

Gli oggetti inclusi in **11.1.2** e **11.1.3** sono **h** e **h'**, suoni brevi in modalità esecutiva ordinaria con o privi di acciaccature. Un "rallentando" al termine di batt. 3 sembra introdurre un meccanismo di transizione graduale alla seconda sottoregione ma si tratta di una *flessione* nella texture, una falsa transizione, dato che il tempo della **11.2** è più veloce.

MACRO-SOTTOR.	BATT.	INDICAZ. AGOGICA
11.1	1-3	Très lent [croma = 92 ÷ 98], avec beaucoup de flexibilité
11.2	4-7	Sub. nerveux et extrêmement irrégulier [croma = 108]
11.3	8-11	A peine moins irrégulier [croma = 112]
11.4	12-15	Assez irrégulier mais rythmique [croma = 116]
11.5	16-19	Plus irrégulier [croma = 112]
11.6	20-22	Extrêmement irrégulier [croma = 108]
11.7	23-25	Lent [croma = 92 ÷ 98], très flexible

Tabella 20. Regione 11: suddivisione in macro-sottoregioni.

La seconda macro-sottoregione, quadripartita, include quattro aggregati formati da oggetti **j** e **j'** “picchettato con accenti variabili. Gli oggetti sono distribuiti nelle quattro battute con dinamica *ff* per l’acciaccatura iniziale e *p* il resto.

La macro-sottoregione è costruita, al pari della **11.1**, secondo una forma ABA+B', con il primo e il terzo aggregati congruenti e il quarto che aggiunge un prolungamento, ottenuto mediante l’iterazione dell’ultimo oggetto, in coda alla struttura del secondo.

La macro-sottoregione **11.3** è bipartita e i sei aggregati, seguiti dal respiro, sono a loro volta raggruppabili a tre a tre: nella micro-sottoregione **11.3.1** lo stesso aggregato costituito da due oggetti viene iterato tre volte mentre la seconda micro-sottoregione è costruita secondo una forma ABA, a specchio con asse di simmetria centrale. Gli aggregati sono costituiti dall’alternanza di suoni brevi e acciaccature, tutto “picchettato”, con accenti sugli oggetti **j'** privi di acciaccatura, inversione del rapporto acciaccatura nota (batt. 8) e permutazione degli eventi (batt. 9-10).

La macro-sottoregione **11.4** è quadripartita al pari della **11.2**, con massima estensione nel terzo aggregato. I primi tre aggregati sono costruiti per aumentazione reciproca a partire dal primo; il quarto è speculare al primo. La forma può essere interpretata come un 3+1.

Gli aggregati sono contraddistinti dal “picchettato” e dalla presenza di una (batt. 12-13) o più (batt. 14-15) suoni brevi senza acciaccatura (**j'**) con arco in su nella regione caudale.

La quinta macro-sottoregione è bipartita: i quattro aggregati individuano due micro-sottoregioni. Esattamente come per la **11.3.1**, la **11.5.1** è costituita dall’iterazione dello stesso aggregato mentre la seconda micro-sottoregione vede il secondo aggregato costruito come immagine trunca del primo.

La forma della macro-sottoregione **11.6** è bipartita. Nella prima parte i primi due aggregati sono il secondo il raddoppiamento del primo mentre nella seconda parte viene messa in atto una forma a specchio, con una nota tenuta al centro quale asse di simmetria tra due oggetti **j**, il secondo dei quali ibridato con l’oggetto **f** e un oggetto **h**. Un “rallentando” a batt. 22 introduce l’ultima sottoregione con un reale processo di transizione.

La macro-sottoregione **11.7**, tripartita, include le seguenti micro-sottoregioni:

- **11.7.1:** quattro oggetti **b** (batt. 23-24), i primi tre riuniti in un aggregato attraverso un’unica arcata; la suddivisione in 2+2 operata attraverso la suddivisione delle battute è negata dalla forma 3+1 ottenuta con la legatura di frase in una arcata;
- **11.7.2:** sei oggetti **b'** (batt. 25-28); la forma è ABBA, con le due battute centrali che includono un solo oggetto;

- **11.7.3:** a batt. 29 un oggetto **b** seguito da un oggetto **i'**, retrogrado speculare all’elemento **i** presente in batt. 1.

In Tabella 21 viene descritta la morfologia delle singole macro-sottoregioni.

L’osservazione di questa sezione mette in luce come Boulez utilizzi processi di *aumentazione*, *diminuzione*, *raddoppiamento* e *dimezzamento*, *iterazione*, *inversione* (oltre alle altre strategie proprie della composizione seriale) nella generazione dei materiali, così come l’organizzazione secondo simmetrie (*anamorfosi*, se si considerano tutti i parametri musicali), l’utilizzo di *forme a specchio* e di *dissimmetrie ponderate* nell’organizzazione della micro-forma.

L’elettronica implementa i seguenti processi:

- quattro harmonizer
- sampler
- frequency shifting
- sampler 1 e 2
- sampler con riverberazione infinita

I processi sono ripartiti come indicato in Tabella 22.

MACRO-STTR.	MORFOLOGIA
11.1	11.1.1 ib/ 11.1.2 [h''] [hhhh] / [h''] 11.1.3 [hh] (→)
11.2	11.2.1 [jj'j'] / [jj'j'j'] [jj'j'] / 11.2.2 [jj'j'j'j'] z
11.3	11.3.1 [j'j] [j'j] [j'j] / 11.3.2 [j'j'j'] / [j'j'j'] [j'j'j'] z
11.4	11.4.1 [jjj'] / [jjjj'] / [jjjjj'j'] / 11.4.2 [jj'j'] z
11.5	11.5.1 [j'jj] / [j'jj] / 11.5.2 [j'j'j'j'] / [j'j'j'] z
11.6	[jj'] / [jjj'j'] [j] - {f} - h →
11.7	11.7.1 [bb/b] b / 11.7.2 b'b' / b'b' / b'b' / 11.7.3 bi'

Tabella 21. Regione 11: morfologia.

				PROCESSO				
				1	2	3	4	5
SOTTOREGIONI	11.1	BATTUTE	1					
			2					
			3					
	11.2		4					
			5					
			6					
			7					
	11.3		8					
			9					
			10					
			11					

13.1.13	h''b/h''b/h''b/[k]
13.1.14	h'''b'''/[k']
13.1.15	h'''b'''/[k]
13.1.16	h'''b'''/h'''b'''z

Tabella 24. Micro-sottoregioni **13.1.3-16**: morfologia.

A partire dalla **13.1.11** l'elemento antecedente dell'aggregato subisce una mutazione per ibridazione con l'elemento **h''** che aumenta il suo grado di articolazione, perdendo i gruppi irregolari a partire da **13.1.14**, fino a **13.1.16** (batt. 52-53), in cui esso è presente in entrambe le battute, sotto forma di oggetto **h'''** (suoni brevi in gruppi regolari con acciaccature, legato); il trillo, a partire da **13.1.14**, ha un'acciaccatura sul transiente d'attacco (**b'''**). In questo caso l'elemento antecedente è di tipo a-direzionale e la flessione con falsa transizione è assente. In Tabella 24 viene descritta la morfologia delle restanti micro-sottoregioni.

Gli oggetti rinvenuti per la prima volta in questa sottoregione sono i seguenti:

b'''	trillo semitono o tono	acciaccatura sul transiente d'attacco
h'''	suono breve ordinario	gruppi regolari con acciaccature
k'	staccato	gruppi irregolari

[k] e [k'] sono aggregati formati da permutazioni degli oggetti **k** e **k'**.

La micro-sottoregione **13.1.13** (batt. 44-47) si declina quale momento di massima dilatazione per quanto riguarda la zona caudale di **13.1**, mentre sia **13.1.14** (batt. 48-49) e **13.1.15** (batt. 50-51) sono formate da due sole battute, la prima occupata dall'aggregato **h'''b'''** e la seconda dall'elemento di flessione che si innesta su [k] e [k'] e innesca la falsa transizione, risolta sempre con la propria negazione e la ripresa del tempo originale. L'elemento antecedente degli aggregati tipo (**hb**), (**ab**) e (**ib**) mantiene sempre un rapporto di contiguità con il trillo che segue e, nel caso sia presente direzionalità, la transizione tra i due elementi può essere considerata continua. In questo senso, l'aggregato può essere considerato come una lettura orizzontale (dispiegamento lungo l'asse diacronico) del bicordo o tricordo posto sul transiente d'attacco del trillo nell'aggregato **b''**.

MICRO-SOTTOR.	BATT.	CONTRAZ. / DILATAM.	TIPOL. E-LEM. DI FLESSIONE
13.1.1	6	Prima contrazione	direzionale discendente
13.1.2			a-direzionale
13.1.3			
13.1.4			
13.1.5	5		
13.1.6	4		
	3		

13.1.7	2	Seconda contrazione	direzionale discendente
13.1.8	4		
13.1.9	3		
13.1.10	2	Dilatamento	a-direzionale
13.1.11			
13.1.12	3		
13.1.13	4	Max. contrazione	[assente]
13.1.14	2		
13.1.15			
13.1.16			

Tabella 25. Macro-sottoregione **13.1**: analisi parametrica.

Dall'osservazione emerge un andamento direzionale per quanto riguarda la durata, espressa in numero di misure, delle micro-sottoregioni. Due processi di contrazione hanno luogo: il primo interessa le prime 7, in cui si passa da 6 a 2 battute, il secondo le quattro successive, con passaggio da 4 a 2.

Dalla **13.1.12** ha luogo un processo di dilatazione da 2 a 4 battute, per poi ritornare al massimo livello di contrazione (2 battute) nelle ultime tre micro-sottoregioni.

In Tabella 25 vengono descritti l'andamento del parametro durata in relazione alle singole micro-sottoregioni, espresso in numero di battute, unitamente ai movimenti contrattili/dilatatori individuati a partire dalla variazione di tale parametro e alla tipologia (direzionale o a-direzionale) dell'elemento di flessione posto alla fine di ogni micro-sottoregione, con l'eccezione della **13.1.16**, in cui è assente.

Per quanto riguarda gli elementi di flessione, dopo il primo direzionale discendente, essi sono privi di direzionalità sino alla **13.1.11**, in cui avviene la transizione verso processi direzionali e l'inversione di tendenza da contrazione a dilatamento.

La micro-sottoregione **13.1.11** si configura quindi come nodo intorno a cui hanno luogo due processi di trasformazione dei materiali:

1. da *inorganici* tendono a divenire *organici*, con processo di acquisizione di direzionalità e conseguente affermazione della freccia del tempo;
2. i movimenti contrattili subiscono un'inversione di tendenza, dando luogo a un movimento dilatatorio.

L'acquisizione di direzionalità, instabilità e la trasformazione di un moto contrattile iterato in una pulsazione formata da contrazione-dilatazione nonché l'evidente processo di ibridazione tra oggetti di tipo (**a**), (**i**), (**h**) e (**k**) che, in questa macro-sottoregione, condividono e si scambiano continuamente tratti morfologici e sintattici, rendendo ardua una classificazione rigorosa, ci dicono del *passaggio da una condizione di stabilità a una di instabilità* del sistema: una transizione che evoca quella dai *cristalli alle forme di vita organiche*, attraverso l'acquisizione dell'elemento liquido.

Tale meccanismo di transizione è reso esperibile attraverso un processo *figura-sfondo*, in cui l'esplorazione diegetica di materiali musicali secondo tipologie ricorrenti è lo sfondo immobile su cui si dispiega la transizione da uno stato ad un altro.

L'elettronica implementa due sampler.

Il primo si attiva sugli elementi conseguenti degli aggregati **i''b**, **h''b**, **i'''b**, **ab**, **a''a''a''a''b**, **h''b** e **h'''b** secondo quattro tipologie base di oggetti:

1. [ORIG.] arpeggio ascendente di suoni tenuti sommati in un cluster (es. batt. 1)
2. [RETR.] cluster che si svuota progressivamente arpeggiando dall'acuto verso il grave (es. batt. 3)
3. [INV.] arpeggio discendente di suoni tenuti sommati in un cluster (es. batt. 4)
4. [INV. RETR.] cluster che si svuota progressivamente arpeggiando dal grave verso l'acuto (es. batt. 2)

I processi elettronici in **13.1** vengono presentati, nelle prime quattro battute, seguendo la stessa logica, di tipo seriale, che viene applicata alla morfologia degli interludi.

Nelle battute seguenti, si aggiungono altre quattro tipologie, sempre di matrice seriale:

5. [I'] permutazione direzionale discendente di suoni tenuti sommati in un cluster (batt. 8)
6. [O'] permutazione direzionale ascendente di suoni tenuti sommati in un cluster (batt. 9)
7. [R'] cluster che si svuota secondo direzionalità discendente (batt. 10)
8. [IR'] cluster che si svuota secondo direzionalità ascendente (batt. 14)

Una ulteriore tipologia, ibrida, compare solo nelle due forme O e R, in quanto le altre due non sarebbero state possibili, a fronte della assenza di direzionalità o direzionalità multiple / caotiche degli elementi antecedenti e conseguenti il cluster.

9. [O''] cluster che si svuota in modo a-direzionale, defocalizzandosi dal centro o con più vettori direzionali (batt. 7, 15, 34, 42, 52)
10. [R''] permutazione a-direzionale o con più vettori direzionali di suoni tenuti sommati in un cluster (batt. 35, 53).

Il secondo sampler si attiva sugli aggregati **[k]** e produce sequenze di bicordi che anticipano la **Tipologia 6** che sarà descritta in relazione alla sottoregione **13.3**, così come gli aggregati su cui si innestano anticipano la **Tipologia 7**. La **Tipologia 5** viene anticipata dagli elementi antecedenti degli aggregati **h'''**, la cui variante **h''** era già stata presentata nella **3.2**.

Osserviamo come ancora una volta una delle strategie compositive messe in campo da Boulez riguardi la *ge-*

stione e manipolazione mirata delle aspettative e come il gioco di rimandi e riferimenti tra gli oggetti e gli aggregati, tra le tipologie morfologiche e la sintassi del brano sia finalizzata a una *manipolazione della memoria* dell'ascoltatore anche attraverso l'inclusione dell'*elemento sorpresa*, attraverso ibridazioni, transizioni reali e fasulle, contaminazioni morfo-sintattiche e continui scambi nell'assegnazione di ruoli e funzioni tra il materiale strumentale e quello elettronico, in una continua interazione che non avviene solo a livello tecnico ma che risiede profondamente radicata nella prassi compositiva dell'autore.

Macro-sottoregione 13.2

La seconda porzione della Sezione VI è suddivisibile in 37 micro-sottoregioni. Alcuni nuovi oggetti e varianti di oggetti già classificati emergono dall'osservazione di questa macro-sottoregione:

d'	linee discontinue	divergenti a partire da un unisono
	(13.2.1, 5, 13, 17, 22, 24, 27, 30, 33, 37)	
i'''	pizzicato	gruppi irregolari note singole + bicordi
	(13.2.2, 4, 6, 10, 12, 16, 20, 26, 29, 36)	
i''''	quadricordo	
	(13.2.31, 34)	
m	tricordi/quadricordi	arco, seguiti da suoni brevi, stessa arcata
	(13.2.3, 8, 14, 18, 21, 23, 28, 34)	

In Tabella 26 le micro-sottoregioni vengono descritte a livello morfologico.

Dall'osservazione emerge con chiarezza come la suddivisione in micro-sottoregioni sia stata operata dal compositore sulla base delle differenze morfologiche: ad ogni frammento, circoscritto da una precisa indicazione agogica e metronomica, corrisponde un contenuto oggettuale preciso, che si ripresenta ogni volta che si ripete quel tipo di frammento, con le stesse dinamiche, secondo quattro tipologie:

1. Calme, régulier [croma = 98÷100]
2. Agité [croma = 126]
3. Brusque [croma = 138÷140]
4. Calme, retenu [croma = 92]

In Tabella 27 viene osservato l'andamento del parametro durata, espresso in numero di battute, per quanto riguarda le quattro tipologie individuate. Nel grafico vengono rappresentati visivamente gli stessi dati.

La **Tipologia 1**, che rimane invariata per tutta la sottoregione, include unicamente oggetti **d'** e micro-sottoregioni della durata di una battuta, con dinamiche comprese tra *pp* e *p*. Si tratta di oggetti formati da due

linee di semibiscrome organizzate in gruppi irregolari che divergono in staccato a partire da un unisono, la linea superiore invariata, la linea inferiore con un glissando discendente, con arco in giù “ricochet” e dinamiche in diminuendo. L’intervallo massimo raggiunto è una settima minore (batt. 54); l’intervallo minimo è una seconda minore (batt. 163): rispettivamente la prima e l’ultima micro-sottoregione, la durata maggiore (un quarto e mezzo, insieme a batt. 124) e la durata minore (un sedicesimo in gruppo irregolare 5:4).

La **Tipologia 2** include unicamente aggregati composti da oggetti **P'''** e varia per numero di oggetti e durata, con una zona di maggiore estensione intorno a **13.2.20**. I materiali musicali sono costituiti da aggregati di suono singoli e bicordi in pizzicato, organizzati in gruppi formati da terzine e separati da pause.

La **Tipologia 3** include aggregati formati da oggetti **m** fino a **13.2.31**, quando gli aggregati divengono eterogenei, formati da oggetti **m**, **P''** e **P'''**. Gli oggetti inizialmente sono a carattere puntiforme, formati da tricordi e note singole, suonati con arco ordinario e legati nella stessa arcata; dagli oggetti binari di batt. 59 si passa a costruzioni più articolate con tricordi e quadricordi in area prossimale (*antecedente*) e porzioni caudali (*conseguente*) formate da gruppi sempre più estesi, fino alla prima ibridazione di batt. 150, in cui compaiono due antecedenti isolati che rientrano nelle tipologie **P''** e **P'''**. Le dinamiche sono sempre elevate, in *ff* con “crescendo” su tutti gli oggetti non puntiformi.

La **Tipologia 4** include aggregati formati da oggetti **P''**, tricordi in pizzicato lasciati risuonare, con dinamiche comprese tra *mp* e *pp*. I pizzicati sono arpeggiati dalla nota più grave alla nota più acuta. Gli aggregati variano, su un piano morfologico, solamente per il numero di oggetti inclusi, che variano nei parametri altezza e durata.

Come si evince dalla Figura 3, per quanto riguarda il parametro durata, le **Tipologie 2, 3 e 4** hanno un andamento ad arco, con valori minimi agli estremi e valori elevati nelle zone centrali, in generale con una flessione localizzata intorno a **13.2.31** che pare ricoprire in **13.2** lo stesso ruolo nodale della micro-sottoregione **13.1.11** in **13.1**, in quanto questo è anche il punto in cui avviene la transizione da una condizione di *omogeneità* a una di *eterogeneità*, relativamente alla composizione interna degli aggregati.

Questa transizione è interpretabile come acquisizione di *complessità*, ovvero il passaggio da forme di organizzazione elementari (organismi formati da cellule della stessa tipologia) a forme più complesse ed evolute, esattamente come in **13.1** avviene un passaggio da forme inorganiche a forme organiche.

Anche in questo caso la transizione è resa evidente da un *processo figura-sfondo*, in cui elementi permangono pressoché immutati (**Tipologie 1, 2 e 4**) al fine di rendere più percettibile il meccanismo di transizione assegnato ad altri elementi (**Tipologia 3**). Questo è valido anche all’interno dei singoli parametri del suono: nella stessa

Tipologia 3, ad esempio, la morfologia degli aggregati subisce una mutazione mentre il piano dinamico resta invariato. Fatti salvi i processi di mutazione, infatti, l’assegnazione dei parametri musicali alle tipologie di oggetti e aggregati è quasi sempre statica, rigida.

I processi messi in atto dall’elettronica sono altamente specializzati: a ciascuna tipologia morfologica corrisponde un particolare processo.

I processi attivati sono i seguenti:

1. frequency shifting
2. quattro harmonizer + delay
3. sampler con riverberazione infinita
4. sampler 1
5. sampler 2
6. ring modulator + comb filter
7. riverberazione infinita

In Tabella 28 sono descritte l’assegnazione dei processi alle tipologie strumentali, le micro-sottoregioni in cui essi sono attivi e la morfologia degli eventi generati.

Nell’accoppiamento dei Processi 2 e 3 sulla **Tipologia 2** ritroviamo la *multi-stratificazione dei processi* in funzione del *meccanismo figura-sfondo*, messo in atto tra le note tenute e i suoni brevi (sampler con riverberazione infinita) pizzicati trasposti (harmonizer), questi ultimi con *funzione mimetica* rispetto ai materiali strumentali.

PROC.	TIPOL.	MICRO-SOTTOREG.	MORFOL.
1	1	13.2.1, 5, 13, 17, 22, 24, 27, 30, 33, 37	d'
2, 3	2	13.2.2, 4, 6, 10, 12, 16, 20, 26, 29, 36	f, suoni brevi pizz.
4	3	13.2.3, 8, 14, 18, 21, 23, 28, 31, 34,	suoni brevi + pizz.
5, 6, 7	4	13.2.7, 9, 11, 15, 19, 25, 32, 35,	f

Tabella 26. Macro-sottoregione **13.2**: assegnazione delle tipologie alle micro-sottoregioni.

Un meccanismo di *anticipazione* viene messo in atto sui materiali relativi al Processo 4 / **Tipologia 3**, che anticipano i materiali inclusi nella **Tipologia 6** in **13.3**. Il bicordo generato a batt. 110 dal Processo 4 pare anticipare l’oggetto puntiforme alla fine della Regione 14, di cui è un possibile dispiegamento sull’asse del tempo (*trasposizione dello stesso oggetto su assi differenti*). Anche in questa sottoregione le unità frequency shifting (Processo 1) utilizzano il *meccanismo della ridondanza* per creare oggetti morfo-sintatticamente unitari e definiti ma timbricamente ibridi (*ibridazione timbrica*), quindi ambigui: la riconoscibilità degli oggetti viene affermata e negata, al fine di provocare un senso di straniamento nell’ascoltatore, che riceve un’informazione ambivalente. La *ripartizione funzionale dei ruoli*, sia per quanto ri-

guarda l'elettronica che per i materiali strumentali è rigorosa e coerente con i dati emersi sino ad ora. La ripresentazione di materiali congruenti segue una strategia di *iterazione / micro-variazione* volta a manipolare l'aspettativa e la fissazione mnemonica dei materiali e delle relazioni che li percorrono.

Macro-sottoregione 13.3

La terza porzione della Sezione VI è suddivisibile in 28 micro-sottoregioni. Gli oggetti inclusi si suddividono in tre tipologie, elencate e descritte in Tabella 29. In Tabella 30 viene operata la classificazione morfologica dei materiali strumentali relativamente a ciascuna micro-sottoregione.

TIPOLOGIA	MORFOLOGIA
5	h'''
6	c' (bicordi crine sul pont. non legato)
7	k

Tabella 27. Macro-sottoregione 13.3: tipologie e morfologia.

Le tre tipologie morfologiche sono costituite da oggetti già utilizzati, organizzati in figurazioni complesse.

La **Tipologia 5** utilizza il legato, con fraseggi articolati in gruppi e sottogruppi di suoni brevi (**h**).

La **Tipologia 6** include un'alternanza di bicordi non legati (13.3.2, 5, 7, 11, 27) o legati a gruppi (13.3.13, 15, 18, 21, 23), con arco al ponticello, organizzati in aggregati mediante la compresenza di due linee melodiche. La 13.3.28 è una variante a note singole. Le micro-sottoregioni 13.3.15, 18, 21 e 23 sono attraversate da un'unica legatura di frase e possono essere considerate come oggetti singoli, multi-stratificati.

La **Tipologia 7** utilizza lo staccato, organizzando le altezze a gruppi di due o tre note con direzionalità ascendente o discendente; variazioni agogiche introducono *flessioni* nella texture e *transizioni*.

Le dinamiche variano sia all'interno dei singoli aggregati che tra le micro-sottoregioni, con range compreso tra *ppp* (13.3.28) e *ff* (13.3.26).

MICRO-SOTTOR.	BATT.	N° OGG.	N° AGGR.	TIP.	DIN.
13.3.1	164 – inizio 165	6	1	5	<i>mp</i>
13.3.2	165-166	27	[6]	6	<i>pp</i>
13.3.3	167-172	79 [31]	1	7	<i>pp</i> ÷ <i>mf</i>
13.3.4	173-175	7	5	5	<i>p</i> ÷ <i>mp</i>
13.3.5	176	20	[5]	6	<i>pp</i>
13.3.6	177-179	14	6	5	<i>p</i> ÷ <i>mf</i>
13.3.7	fine 179 - 180	9	[3]	6	<i>pp</i>
13.3.8	181	2	1	5	<i>mp</i>
13.3.9	181-183	35	[7]	6	<i>pp</i>

13.3.10	184	3	1	5	<i>mp</i>
13.3.11	185-187	44	[8]	6	<i>pp</i>
13.3.12	188	5	2	5	<i>mf</i>
13.3.13	fine 188 - 189	14	[4]	6	<i>pp</i>
13.3.14	190	5	2	5	<i>mf</i>
13.3.15	190	9	1 [3]	6	<i>p</i>
13.3.16	191	9	3	7	<i>pp</i>
13.3.17	192	4	2	5	<i>mp</i>
13.3.18	fine 192 - 193	14	1 [4]	6	<i>pp</i>
13.3.19	194	9	3	7	[<i>pp</i>]
13.3.20	195	4	2	5	<i>mf</i> ÷ <i>f</i>
13.3.21	196	20	1 [5]	6	<i>mf</i> ÷ <i>p</i>
13.3.22	197 – inizio 198	8	3	5	<i>f</i>
13.3.23	198-199	27	1 [6]	6	<i>p</i>
13.3.24	200	18 [9]	1	7	<i>pp</i> ÷ <i>f</i>
13.3.25	201	1	1	i''	<i>p</i>
13.3.26	202-203	9	3	5	<i>ff</i>
13.3.27	204-207	44	8	6	<i>mf</i>
13.3.28	208	1	1	6*	<i>ppp</i>

Tabella 28. Macro-sottoregione 13.3: analisi parametrica.

L'oggetto **i''**, a batt. 201, è un elemento di interpunzione, rilevabile anche in **191** e **194**, in forma ibridata con la **Tipologia 7**: si tratta di un *meccanismo di transizione* tra le due tipologie oggettuali.

L'elettronica implementa due sampler con riverberazione infinita.

Il primo sampler genera un cluster costruito intorno a un'altezza generatrice, il secondo produce una singola altezza. Gli eventi sono generati, uno per micro-sottoregione tranne nella **13.3.27** in corrispondenza delle micro-sottoregioni **13.3.1, 3, 4, 6, 8, 10, 11** (a metà), **12, 14, 17, 20, 22, 23, 24, 26, 27** (tre eventi). Le micro-sottoregioni centrali che contengono una generazione di evento includono la **Tipologia 5**, quelle situate in zona prossimale o caudale le altre due tipologie.

Sia il rapporto tra la morfologia degli eventi elettronici (una sola tipologia) e la morfologia di quelli strumentali (tre tipologie in costante permutazione) che la suddivisione formale operata attraverso l'analisi dei materiali eseguiti dall'elettronica (tripartizione) rispetto alla segmentazione individuata sulla base dei materiali strumentali indicano chiaramente la presenza, in questo caso a livello macroscopico, dei meccanismi *figura-sfondo* e *multi-stratificazione dei processi*.

La strategia messa in atto da Boulez si configura quindi come un unico processo, coeso e unitario, in grado di applicare gli stessi strumenti compositivi tanto alla microforma che alla macroforma, alla scrittura elettroacustica che a quella strumentale.

MICRO-SOTTOREG.	BATT.	N° BATT.	INDICAZ. AGOG.	N° OGG	MORFOLOGIA	DINAM.
13.2.1	54	1	Calme, régulier [croma = 98÷100]	1	d'	pp ÷ p
13.2.2	55-58	4	Agité [croma = 126]	4	l''' l''' l''' l'''	mf ÷ f
13.2.3	59-60	2	Brusque [croma = 138÷140]	4	mmmm	ff
13.2.4	61-65	5	Agité	4	l''' l''' l''' l'''	mf ÷ f
13.2.5	66	1	Calme, régulier	1	d'	pp
13.2.6	67-71	5	Agité	3	l''' l''' l'''	mf ÷ f
13.2.7	72-73	2	Calme, retenu [croma = 92]	3	l'' l'' l''	pp ÷ p
13.2.8	74-75	2	Brusque	5	mmmmm	ff
13.2.9	76-80	5	Calme, retenu	10	l' l' l' l' l' l' l' l' l' l'	pp ÷ mp
13.2.10	81-85	5	Agité	4	l''' l''' l''' l'''	mf ÷ f
13.2.11	86-87	2	Calme, retenu	2	l'' l''	pp ÷ p
13.2.12	88-92	5	Agité	4	l''' l''' l''' l'''	mf ÷ f
13.2.13	93	1	Calme, régulier	2	d' d'	pp ÷ p
13.2.14	94-97	4	Brusque	8	mmmmmmmm	ff
13.2.15	98-99	2	Calme, retenu	4	l' l' l' l' l' l'	pp ÷ p
13.2.16	100-104	5	Agité	5	l''' l''' l''' l''' l'''	mf ÷ f
13.2.17	105	1	Calme, régulier	1	d'	pp
13.2.18	106-109	4	Brusque	6	mmmmmm	ff
13.2.19	110-112	3	Calme, retenu	6	l' l' l' l' l' l'	pp ÷ p
13.2.20	113-118	6	Agité	5	l''' l''' l''' l''' l'''	mf ÷ f
13.2.21	119-123	5	Brusque	7	mmmmmmmm	ff
13.2.22	124	1	Calme, régulier	1	d'	pp ÷ p
13.2.23	125-128	4	Brusque	5	mmmmm	ff
13.2.24	129	1	Calme, régulier	1	d'	pp
13.2.25	130-134	5	Calme, retenu	8	l' l' l' l' l' l' l' l'	pp ÷ mp
13.2.26	135-139	5	Agité	5	l''' l''' l''' l''' l'''	mf ÷ f
13.2.27	140	1	Calme, régulier	1	d'	pp
13.2.28	141-143	3	Brusque	6	mmmmmm	ff
13.2.29	144-148	5	Agité	4	l''' l''' l''' l'''	mf ÷ f
13.2.30	149	1	Calme, régulier	1	d'	pp
13.2.31	150-152	3	Brusque	4	l' mml''''	ff
13.2.32	153	1	Calme, retenu	1	l''	p
13.2.33	154	1	Calme, régulier	1	d'	pp
13.2.34	155-156	2	Brusque	3	ml''''l''	ff
13.2.35	157-158	2	Calme, retenu	7	l' l' l' l' l' l' l'	pp ÷ mp
13.2.36	159-161	3	Agité	5	l''' l''' l''' l''' l'''	mf ÷ f
13.2.37	163	1	Calme, régulier	1	d'	pp

Tabella 29. Macro-sottoregione 13.2: morfologia.

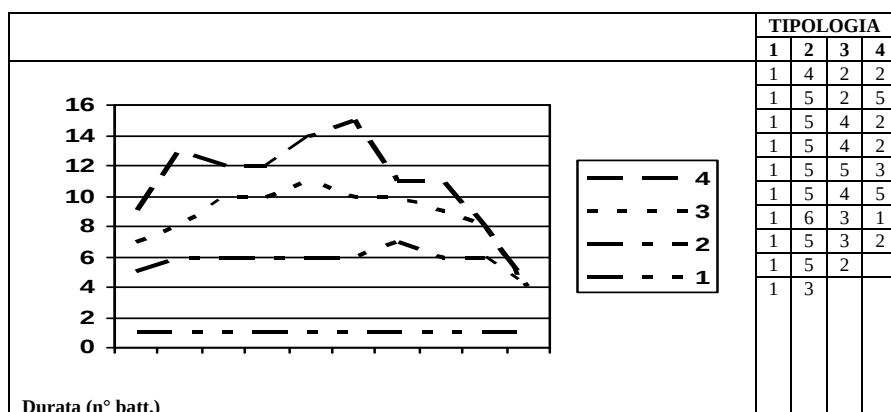


Tabella 30. Macro-sottoregione 13.2: durata delle micro-sottoregioni per tipologia.

CONCLUSIONI

Alla luce dei dati emersi, è possibile tracciare una mappa dei processi compositivi implicati in *Anthèmes 2*.

- processi di micro-variazione
- multi-stratificazione dei processi
- defocalizzazione
- iterazione
- crossfade
- anamorfosi
- uso di elementi di interpunzione
- processo figura-sfondo
- transizione semplice → complesso
- transizione cristallo → organico
- meccanismi a specchio
- flessione (falsa transizione)
- interscambiabilità dei ruoli tra strumento ed elettronica, relativamente alla costruzione dell'organizzazione formale e alla ripartizione delle funzioni strutturali
- flessibilità funzionale
- precisa ripartizione dei ruoli a livello funzionale
- meccanismo della ridondanza (elettronica: ispessimento di eventi lineari, obliqui, puntiformi)
- funzione mimetica / funzione diegetica
- trasposizione dello stesso oggetto su assi differenti (lettura orizzontale di oggetti puntiformi)
- modulazione del rapporto ambiguità / riconoscibilità

Queste linee guida emerse dall'analisi sono interamente congruenti con quanto affermato da Pierre Boulez nell'intervista rilasciata a Parigi, in occasione della prima esecuzione del brano. In quell'occasione, Boulez paragonò *Anthèmes 2* a

“...un inno in cui c'è una successione di strofe e di paragrafi che sono costruiti come inni, cioè una sorta di ritornello”.¹

Per quanto riguarda gli item *processi di microvariazione*, *precisa ripartizione dei ruoli a livello funzionale* e *modulazione del rapporto ambiguità / riconoscibilità*, in relazione anche a quanto scritto circa *gestione e manipolazione mirata delle aspettative* e dei *processi di ritenzione mnestica*, Boulez chiarisce come si tratti di strategie messe in atto coscientemente e con finalità processo-specifiche.

“...sono ora convinto che la musica debba essere basata su oggetti musicali riconoscibili. Questi non sono “temi” in senso classico ma, piuttosto, entità che, sebbene cambino costantemente nella loro forma, hanno determinate caratteristiche a tal punto identificabili da non poter essere confuse con alcuna altra

entità. Questo pezzo è costituito da simili entità, che possono essere identificate molto facilmente. Ciò che è meno facilmente identificabile è l'*ordine* secondo il quale esse hanno luogo o, piuttosto, il *disordine* secondo il quale hanno luogo. Noi riconosciamo un determinato evento ma non possiamo sapere quando esso accadrà; ce ne rendiamo conto *dopo il fatto*. Questo è ciò che mi interessa – creare un effetto di sorpresa e riconoscimento simultanei. Una variazione ha luogo per quanto riguarda l'oggetto complessivo tuttavia, quando un evento è percepito, quell'oggetto è altamente caratterizzato”.

Le strategie compositive includono la gestione dell'elemento sorpresa, l'utilizzo di processi quali la cristallizzazione degli oggetti strumentali processati e la connotazione degli stessi in termini psicologici (anticipazione / reminiscenza).

“Tutto ciò che resta [riferendosi ai materiali elettronici generati a partire dagli eventi strumentali, nella Regione 14] è una specie di reminiscenza, ed è quasi impercettibile, dato che l'ultima nota del pezzo è in armonici. Questa nota rimane ghiacciata sul posto e segnala la fine. Ciò che io vi sto *inviando*, ciò che voi *percepите*, sono *segnali*. Si tratta di segnali che veicolano le vostre emozioni, sebbene in modo inconscio. Questo è essenzialmente come io vi *manipolo*. Ogni esecuzione, ogni composizione è una sorta di manipolazione, nel senso migliore del termine, spero. E questa manipolazione deve essere preparata prima.”

I meccanismi di compressione / dilatazione osservati nella macro-forma si applicano anche nella micro-forma.

“C'è un passaggio suonato completamente in *pizzicato*, con dinamiche molto elevate, che è stato concepito come *compressione*. Ho in mente César, in particolare; César, lo scultore, chiaramente. C'è una compressione – vi sono tre linee che vengono alternate ritmicamente. Sebbene ciò sia stato concepito come tre linee, tutte e tre sono contenute in una singola linea eseguita dal violino, che sono fuse insieme in un'unica matassa.”

I processi e i meccanismi di *transizione*, *mutazione*, *variazione*, *aumentazione*, *diminuzione ecc.* (cioè tutte le strategie compositive che vengono applicate ai materiali primari, in questa analisi descritti sotto forma di oggetti) messi in atto sono centrali. I materiali di partenza sono elementi indifferenziati ai quali vengono conferiti ordine, misura e organizzazione, identità e personalità.

“Trovo che i punti di partenza non abbiano molta importanza. Ciò che è importante è la *traiettoria* che poi essi prendono”.

Boulez inoltre spiega qual è il suo approccio al live electronics, relativamente a questo brano.

“...il violinista fornisce tutto il materiale che gli richiediamo, con tutta la libertà necessaria. Non c'è nessuna forzatura su di lui, nessuna limitazione temporale. In particolare, egli non ha bisogno di preoccuparsi della sincronizzazione, che avrebbe potuto altrimenti inficiare il suo apporto immaginativo. Al con-

¹ Cfr. APPENDICE II. Tutte le altre citazioni ibid.

trario, noi prendiamo ciò che suona il violinista per trarne fuori qualcos'altro."

Se consideriamo il live electronics come interazione continua e reciproca tra due esecutori, il regista del suono e l'esecutore allo strumento, questa concezione appare meccanicistica e superata nel suo determinismo unilaterale e un po' infantile. Da una parte, la complessità del brano e la sua difficoltà di esecuzione con i mezzi elettronici forniti dall'IRCAM in quell'occasione, potrebbero fornire una sorta di giustificazione alla scelta di ridurre la scena del live electronics a una interazione unilaterale.

"Gli arpeggi sono triggerati dallo 'score follower'; seguendo le note corte, a un certo punto c'è una nota tenuta lunga trillata che è ciò che in definitiva triggera l'arpeggio. A volte – non siamo bene sicuri del perché, che rimane da analizzare – sebbene in genere funzioni, capita che il dinosauro perda per strada qualche dente... Noi ci assicuriamo che il dinosauro non li perda, i denti, e in generale, questo è fattibile, in tempo reale, seguendo la partitura. La partitura viene immessa nel computer e questo segue, con l'aiuto della partitura, tutto quello che accade, sebbene in pratica sia estremamente difficile."

Tuttavia, l'origine di questa scelta risiede a mio avviso nella funzione demiurgica che Boulez sembra riservare al compositore e a una precisa scelta poetica che prevede una relazione asimmetrica tra esecutore allo strumento tradizionale e compositore / esecutore al live-electronics. La tensione dialogica tra mimesi (rappresentazione) e intreccio (diegesi) deriva direttamente dalla tragedia classica aristotelica, in cui in ruolo centrale era ricoperto dal Fato, che disponeva liberamente e in accordo ai propri capricci dei destini dei personaggi.

"C'è sicuramente un elemento teatrale in questo pezzo, non nel senso di avere un violinista che agita le braccia in aria o che suona con lo strumento sopra alla testa ma piuttosto nel senso di avere una *drammaturgia* del pezzo. Soprattutto, l'elemento teatrale risiede nella relazione tra l'esecutore e qualcosa su cui egli non ha controllo."

L'alternanza di interludi e sezioni, elementi di interruzione ed elementi diegetici, sia nella macro-forma che nella micro-forma, è una cifra stilistica che pervade e descrive il brano. L'origine di questo tipo di strategia compositiva deve essere ricercato nelle forme antiche e, in particolare, nelle *Lamentazioni di Geremia*, brano eseguito più volte da Boulez durante l'infanzia.

"Paragonerei questi – dato che ho già parlato di strofe e paragrafi – a lettere. Ricordo quando da bambino cantavamo le lamentazioni di Geremia durante la settimana santa di Pasqua. Ciò che mi colpiva allora era che, malgrado il testo fosse ovviamente in Latino, le strofe erano separate da lettere, che erano a loro volta cantate, ma in Ebreo. Cioè *aleph*, *beit*, ecc. Questo si avvicina al sistema che ho usato qui. Dopo una breve introduzione, la prima lettera annuncia il primo paragrafo. Questa è seguita dal paragrafo, in cui c'è una certa quantità di attività. Quindi questo finisce ed è seguito da una seconda lettera, quin-

di il secondo paragrafo, ecc. Da questo, la forma del pezzo è interamente dedotta. Avete un senso della forma del pezzo che è abbastanza immediato, penso, grazie a questi *respiri* [couples] che sono completamente a-direzionali; essi sono *neutrali*, e poi continua il resto".

Come rilevato, la scelta di assegnare una funzione prevalentemente mimetica al suono elettronico è finalizzata a creare disorientamento nell'ascoltatore e a dissimulare l'alta riconoscibilità dei materiali.

"...questo [descrivendo i processi elettronici applicati ai materiali strumentali] rende estremamente difficile distinguere chi sta suonando in un determinato momento. Vi perdetevi in un labirinto di suono in cui, naturalmente, certi suoni sono più forti di altri."

Infatti, laddove i materiali strumentali rivestono preponderantemente una funzione diegetica, la finalità principale dell'elettronica, in questo brano, è di ordine mimetico e utilizza il meccanismo della ridondanza.

"Lo scopo dell'elettronica in questo pezzo è quello di ispessire il suono."

Le ragioni per la scelta del mezzo elettronico sono puramente pragmatiche e lo strumento tradizionale rimane in primo piano; il ruolo dell'elettronica è quello di amplificarne le possibilità meccaniche e, in questo modo, aumentare il grado di complessità degli oggetti musicali da questo prodotti: in pratica trasformarlo in un iper-strumento.

"...senza modificare le sonorità del violino, noi trasgrediamo le possibilità dello strumento. Abbiamo prodotto *pizzicati* che non potrebbero mai essere eseguiti a quella velocità, poiché sarebbe molto semplicemente impossibile sul piano *fisico*. Ciò che è interessante è l'utilizzo del suono strumentale nella sua nudità, nella sua esattezza ma *oltre* certe possibilità delle dita. [esempio di pizzicati prodotti elettronicamente] Questi sono semplicemente degli arpeggi ma, di fatto, non sono semplici come sembrano; essi sono arpeggi *intersecati*, di una certa complessità di costruzione, che vengono costantemente modificati in modi diversi. E ancora tutto è scritto in relazione al violino, ed è triggerato dal violino stesso."

La centralità del lavoro compositivo compiuto sulle aspettative, sui tempi di fruizione, sui meccanismi psicologici di attesa, sorpresa, straniamento sono in ultima analisi strategie relative alla psicologia della percezione: si tratta di un lavoro sulle emozioni che manifesta la lucida volontà di Boulez di restituire alla musica la sua identità di *voce*, intesa come traccia musicale delle passioni *quae sunt in anima*, per usare le parole scelte da Boezio nel suo commento aristotelico *De Interpretatione*.

RIFERIMENTI

- [1] P. Boulez: *Anthèmes 2 pour violon et dispositif électronique*, Universal Edition UE31160, Vienna, 1997.
- [2] M. Imberty: *Le scritture del tempo*, Ricordi, Milano, 1990.
- [3] P. Boulez: “Le système et l’idée”, In *Harmoniques n° 1, décembre 1986: le temps des mutations*, Ircam – Centre Georges-Pompidou, 1986.
- [4] J. Goldman: *Understanding Pierre Boulez’s Anthèmes [1991]: ‘Creating a Labyrinth out of Another Labyrinth’*, Faculty of Music, Université de Montréal, 2001.
- [5] M. Marinoni: “Atomi Distratti di Mario Garuti [2005] – Analisi musicale e tecnologica ed elaborazione di un algoritmo di sintesi in ambiente Max/MSP per l’esecuzione della parte di tastiera elettronica”, *Atti del XVII Colloquio di Informatica Musicale*, pp. 177–190, 2008.

APPENDICE 1

CLASSIFICAZIONE	MORFOLOGIA	VARIANTE
a	frammento melodico	direzionalità discendente
a'		direzionalità ascendente
a''		a-direzionale
b	trillo semitono o tono	isolato
b'		seguito da una nota più grave, stessa arcata
b''		bicordo / tricordo sul transiente d'attacco
b'''		acciaccatura sul transiente d'attacco
c	bicordi crine	arco battuto
c'		sul pont. non legato
d	linee discontinue	convergenti in un unisono
d'		divergenti a partire da un unisono
e	oggetto puntiforme	unisono
e'		bicordo
f	suono tenuto	suono ordinario
f'		armonico di quinta
f''		armonico di quarta
g	glissando	ascendente indeterminato in armonico
g'		discendente indeterminato in armonico
h	suono breve ordinario	con acciaccature
h'		senza acciaccature
h''		gruppi irregolari con acciaccature
h'''		gruppi regolari di 2,3 o 4 legato stessa arcata
h''''		gruppi regolari con acciaccature
i	elemento scalare	direzionale discontinuo ascendente
i'		direzionale discontinuo discendente
i''		direzionale continuo (arcata unica) ascendente
i'''		direzionale continuo (arcata unica) discendente
j	picchettato	con acciaccature
j'		senza acciaccature
k	staccato	gruppi regolari
k'		gruppi irregolari
l	pizzicato	suono singolo
l'		bicordo
l''		tricordo
l'''		gruppi irregolari note singole + bicordi
l''''		quadricordo
m	tricordi / quadricordi	arco, seguiti da suoni brevi, stessa arcata
n	tremolo d'arco	con acciaccature
n'		senza acciaccature
z	respiro	

APPENDICE 2

I materiali di cui ho operato una traduzione in italiano sono quelli inclusi da Jonathan Goldman nella sua tesi. Boulez, discutendo con il musicologo Peter Szendy in occasione della prima esecuzione mondiale di *Anthèmes 2*, avvenuta a Parigi all'IRCAM il 21 Ottobre 1997, fornisce delucidazioni circa i trattamenti elettronici applicati al materiale musicale e le peculiarità estetiche perseguite in questo lavoro, quali l'opposizione vicino-lontano, contrappunto compresso in una sola linea, passaggi di "caos ordinato", ecc.

Peter Szendy: Vorrei ringraziarla, Pierre Boulez, per aver acconsentito a parlarci di *Anthèmes*. Iniziamo con alcune questioni di carattere generale: potrebbe dirci qualcosa riguardo al titolo, *Anthèmes*?

Pierre Boulez: Sì, beh il titolo è una specie di gioco di parole. Si riferisce sia a "inno" che a "tema". Il termine inglese "anthem" significa "inno" come in "inno nazionale". Sebbene la parola *Anthème* non abbia significato in francese, io la uso in questo pezzo per riferirmi sia agli inni che ai temi. Si tratta di un inno in cui c'è una successione di strofe e di paragrafi che sono costruiti come inni, cioè una sorta di ritornello.

PS: Quindi il titolo *Anthèmes* ci dice qualcosa circa la sua relazione al tema, alla notazione di un tema musicale?

PB: Sì, poiché in gioventù pensai che la musica potesse essere *atematica*, completamente priva di temi. Alla fine, in ogni caso, sono ora convinto che la musica debba essere basata su oggetti musicali riconoscibili. Questi non sono "temi" in senso classico ma, piuttosto, entità che, sebbene cambino costantemente nella loro forma, hanno determinate caratteristiche a tal punto identificabili da non poter essere confuse con alcuna altra entità. Questo pezzo è costituito da simili entità, che possono essere identificate molto facilmente. Ciò che è meno facilmente identificabile è l'*ordine* secondo il quale esse hanno luogo o, piuttosto, il *disordine* secondo il quale hanno luogo. Noi riconosciamo un determinato evento ma non possiamo sapere quando esso accadrà; ce ne rendiamo conto *dopo il fatto*. Questo è ciò che mi interessa – creare un effetto di sorpresa e riconoscimento simultanei. Una variazione ha luogo per quanto riguarda l'oggetto complessivo tuttavia, quando un evento è percepito, quell'oggetto è altamente caratterizzato. Questa può essere una possibile spiegazione del titolo.

PS: Ho cercato la parola *Anthème* e ho trovato un significato al quale forse lei non stava pensando quando ha scelto il titolo. Mi riferisco a "chrysanthème" (crisantemo), il fiore. Lo cito come pretesto per ricollegarmi alla prossima domanda: il pezzo è molto ornamentale, quindi come lavora lei con quella che uno potrebbe chiamare la *fioritura*, l'ornamento musicale, nel senso migliore del termine?

PB: Questo pezzo è, come mi succede spesso, una riflessione su qualcosa di composto precedentemente, ma lasciato in uno stato di incompiutezza. In questo caso si tratta di un piccolo frammento di *...explosante-fixe...* Me lo ha mostrato lei prima, quando non avevo sottomano l'esempio. Per essere precisi, inizia semplicemente con sette note. Trovo che i punti di partenza non abbiano molta importanza. Ciò che è importante è la *traiettorie* che poi essi prendono. E questa traiettoria, come lei mi dice, è sette note che durano all'incirca cinque secondi. Considerato che il pezzo dura venti minuti, c'è molto spazio per l'invenzione. Ma questa invenzione è focalizzata su determinate note. Ad esempio, lei ha rilevato come il finale sia giocato su una singola nota, il RE. Infatti, avvicinando il finale si osserva come la texture si assottigli sino a che tutto ciò che rimane è una nota singola, la quale viene poi interrotta attraverso un gesto sul violino che è sia enfatico che ironico, come se si dicesse "E' abbastanza per adesso! Ci si vede in futuro!". Così è come intendo io il finale e la violinista lo ha eseguito molto bene, come le era stato spiegato. E' come se lei aspettasse questa nota per dissolversi e poi - "Finito!". Sta di fatto che si allude a questo finale già all'inizio del brano, in cui c'è già una polarizzazione intorno alla nota RE. C'è una costruzione, quindi questa si arresta, come a dire "No, non ancora. C'è ancora tutto un pezzo da suonare." C'è sicuramente un elemento teatrale in questo pezzo, non nel senso di avere un violinista che agita le braccia in aria o che suona con lo strumento sopra alla testa ma piuttosto nel senso di avere una *drammaturgia* del pezzo. Soprattutto, l'elemento teatrale risiede nella relazione tra l'esecutore e qualcosa su cui egli non ha controllo. Per me, la cosa interessante di un pezzo elettronico è che il violinista fornisce tutto il materiale che gli richiediamo, con tutta la libertà necessaria. Non c'è nessuna forzatura su di lui, nessuna limitazione temporale. In particolare, egli non ha bisogno di preoccuparsi della sincronizzazione, che avrebbe potuto altrimenti inficiare il suo apporto immaginativo. Al contrario, noi prendiamo ciò che suona il violinista per trarne fuori qualcos'altro. Mi pare ci sia una relazione interessante qui tra ciò che è prodotto da un essere umano e ciò che possiamo produrre con le macchine – che è, naturalmente, anch'esso un prodotto dell'attività umana, solamente ottenuto mediante altri circuiti. Ci sono, quindi, due circuiti: un circuito estremamente diretto, intuitivo, e

uno molto più analitico. Ed è in un particolare pezzo che i due circuiti si incontrano.

PS: Vorrebbe dirci qualcosa circa la struttura del pezzo?

PB: Sì, sebbene non abbia alcuna intenzione di tenere una lezione di composizione: non è questo il luogo deputato a una cosa di quel tipo; inoltre, non è nel mio stile. Ho insegnato composizione per tre anni della mia vita ed è stato sufficiente. Discuterò con voi, tuttavia, di ciò che deve colpirvi quando ascoltate il pezzo. Certamente la cosa più ovvia che deve colpirvi sono le interruzioni in cui, diciamo, non accade molto. Questi passaggi, in cui [il violino] suona note tenute in armonici, contrastano con altri passaggi, in cui vi è molta attività. Paragonerei questi – dato che ho già parlato di strofe e paragrafi – a lettere. Ricordo quando da bambino cantavamo le lamentazioni di Geremia durante la settimana santa di Pasqua. Ciò che mi colpiva allora era che, malgrado il testo fosse ovviamente in Latino, le strofe erano separate da lettere, che erano a loro volta cantate, ma in Ebreo. Cioè *aleph*, *beit*, ecc. Questo si avvicina al sistema che ho usato qui. Dopo una breve introduzione, la prima lettera annuncia il primo paragrafo. Questa è seguita dal paragrafo, in cui c'è una certa quantità di attività. Quindi questo finisce ed è seguito da una seconda lettera, quindi il secondo paragrafo, ecc. Da questo, la forma del pezzo è interamente dedotta. Avete un senso della forma del pezzo che è abbastanza immediato, penso, grazie a questi *respiri* [coupsures] che sono completamente a-direzionali; essi sono *neutrali*, e poi continua il resto. E quando non ci sono più note in armonici, potete essere certi che il pezzo è finito. Tutto ciò che resta è una specie di reminiscenza, ed è quasi impercettibile, dato che l'ultima nota del pezzo è in armonici. Questa nota rimane ghiacciata sul posto e segnala la fine. Ciò che io vi sto *inviando*, ciò che voi *percepите*, sono *segnali*. Si tratta di segnale che veicolano le vostre emozioni, sebbene in modo inconscio. Questo è essenzialmente come io vi *manipolo*. Ogni esecuzione, ogni composizione è una sorta di manipolazione, nel senso migliore del termine, spero. E questa manipolazione deve essere preparata prima.

Lo scopo dell'elettronica in questo pezzo è quello di ispessire il suono. Abbiamo parlato di crisantemi, prima e ciò non è lontano dal senso di questo. C'è un ricettacolo, con dei petali intorno. Questo è ciò che si chiama in genere 'harmonizer'. Si tratta di una procedura che in sé non ha niente di particolarmente eccezionale. Ciò che è interessante è che in questo modo non si abbandona mai il sistema temperato, si rimane molto vicini al temperamento del violino. E se aggiungiamo un accordo ad esso, questo accordo si muoverà in maniera parallela. Il che equivale a dire, ad esempio, che se questo accordo è costituito da una terza minore, una settima, ecc. qualsiasi intervallo, il violino suonerà e, come in certi

quadri, la linea diviene più spessa, senza deformazione. Questo crea una sorta di progressione aritmetica; i rapporti non mutano. Dall'altra parte, c'è un'ulteriore tipo di trasformazione conosciuta come 'Frequency shifting'. Il Frequency shifting non è basato sull'intervallo ma sul valore frequenziale in sé, calcolato in Hertz. Non c'è niente di particolarmente scientifico in questo, si tratta solo di sostituire un valore con un altro. Ma in questo caso abbiamo una progressione geometrica, cioè invece di avere una linea retta abbiamo una curva. Le frequenze gravi sono abbassate di un valore rilevante, laddove quelle acute sono solo lievemente abbassate. L'ammontare della trasformazione è in funzione del registro del suono. In questo modo, abbiamo due tipi di trasformazioni, due significati della parola 'trasformazione'. Da una parte abbandoniamo il sistema temperato; non temperati, cioè arricchiti, gli intervalli vengono sommati. Nell'altro tipo di trasformazione, si ha un suono ordinario di violino, solo trasformato. Si hanno gli armonici naturali del violino ma questi armonici sono deformati. Vengono introdotti suoni inarmonici che non hanno nulla a che fare con il suono naturale prodotto dallo strumento (fatta eccezione per la prima stanza, volta a dimostrare i due tipi di trasformazione).

Il secondo importante aspetto dell'elettronica utilizzata perviene alla struttura ritmica. Posso citarvi un passaggio eseguito che avrete certamente notato. C'è un passaggio suonato completamente in *pizzicato*, con dinamiche molto elevate, che è stato concepito come *compressione*. Ho in mente César, in particolare; César, lo scultore, chiaramente. C'è una compressione – vi sono tre linee che vengono alternate ritmicamente. Sebbene ciò sia stato concepito come tre linee, tutte e tre sono contenute in una singola linea eseguita dal violino, che sono fuse insieme in un'unica matassa. Ora, per prolungare questo, Andrei Gertzo (che è responsabile della realizzazione elettroacustica) opera attraverso mezzi elettronici. Il che è come dire che abbiamo allargato questa possibilità svolgendo la matassa e moltiplicandola. Al punto che uno sente la stessa nota suonata, non suonata, che riparte da capo, ecc. Ora ascolteremo due battute del violino da solo, che avete già ascoltato, in modo che possiate individuare il testo solamente. Dopo, ovviamente, esse vengono fatte transitare nei diffusori, e il volume dei diffusori è settato in modo da essere equivalente a quello del violino; questo rende estremamente difficile distinguere chi sta suonando in un determinato momento. Vi perdetevi in un labirinto di suono in cui, naturalmente, certi suoni sono più forti di altri. [frammento della seconda stanza, prima senza elettronica, quindi con il trattamento]

Potete sentire che vi sono note ripetute, ci sono note che cambiano e ci sono note che partono in ogni direzione. Questo significa che è molto difficile, persino per noi, distinguere quello che il violino sta suonando da quello che non sta suonando.

Per me, ciò che è importante è che, quando avete davanti un violino, il suono *prenda corpo* davanti a voi, malgrado con gli amplificatori il suono venga completamente *svuotato*. Questo dà origine a un suono che si sposta, che non è più legato direttamente a qualche forma di percezione locale. Ciò che è interessante qui è avere questo punto fissato, questo punto di coordinazione e, al suo opposto, uno spazio mobile intorno ad esso.

Per fare un altro evento, c'è un ambiente aleatorio. Il violino suona un certo numero di figurazioni, che sono altamente direzionali, e la macchina prende praticamente le stesse note, ma riarrangiandole molto velocemente. Il riposizionamento non è operato in modo caotico – il campo delle altezze è fisso, solo l'ordine è randomico. [In altri passaggi] il campo è fisso e l'ordine è *a sua volta* fisso. Ecco un esempio di opposizione, piuttosto che coordinazione, tra un gesto che è altamente intenzionale e gesti che sono totalmente non intenzionali [frammento dalla quarta sezione].

C'è un contrasto qui tra note suonate molto forti e altre suonate *piano*. C'è un tempo contrastante – ho segnato questo passaggio '*extrêmement irrégulier*' poiché ci deve essere una sorta di movimento convulso. Deve dare l'impressione di non essere più in grado di controllare i movimenti. Ma queste 'convulsioni' sono realizzate attraverso l'uso di valori numerici. Ora queste convulsioni verranno rinforzate da suoni aleatori sullo sfondo – un ulteriore tipo di disordine. E' esattamente la stessa cosa, solo interamente randomica.

Ho usato il termine aleatorio, ma dovrei aggiungere che i valori non sono scelti in maniera casuale. Per le altezze, c'è una consegna ben precisa, e la scelta viene operata all'interno di questa pianificazione. Anche per i valori ritmici esiste una pianificazione poiché, se scegliessimo valori a caso, ovviamente non otterremmo lo stesso risultato. I valori sono calcolati in modo da avere ripetizioni di valori bassi in opposizione a valori alti. Il sistema è pesato in favore dei valori bassi. E' semplicemente una questione di rapporti ma ne sto parlando per evitare di dare l'impressione che i valori siano scelti mediante una lotteria. E' sensibilmente più deterministico che non lasciando le cose al puro caso.

Il prossimo esempio riguarda una forma di contrappunto tra ritmo acustico e spazio acustico [ritmo sonoro e spazio sonoro]. Questo è impossibile senza i mezzi elettronici, dato che non ci si può aspettare che musicisti orchestrali corrano intorno ai quattro angoli della sala per far viaggiare nello spazio i suoni che producono!

Che cosa è il tempo? Possiamo velocemente definire il tempo come due categorie che sono sovrappresse e che possono essere utilizzate precisamente in tal modo. Il tempo è la prima di tutte le relazioni numeriche – una battuta da 4/4 con una nota da ottavo puntato, o una nota da ottavo, ecc. la loro è una indicazione di tempo, tempo con una pulsazione costituita da valori più gradi o più piccoli che vengono posizionati in relazione a tale

pulsazione. Il continuum temporale, d'altra parte, è velocità. Le relazioni numeriche possono essere modificate modificando tale velocità.

Lo spazio, possiamo dire, ha la stessa struttura. Abbiamo la collocazione dei diffusori che è discontinua. Ci sono sei diffusori. Ma c'è anche la funzione 'vicino → lontano', che è una funzione interamente continua. Vero, questa funzione può essere definita attraverso numeri, ma è percepita come una funzione continua piuttosto che discontinua. Per cui, se sovrapponiamo i due livelli, tempo e spazio, possiamo creare una sorta di contrappunto tra i due [frammento dalla quinta sezione].

In questo caso, i suoni non sono modificati, è piuttosto la loro spazializzazione che cambia costantemente. Se siete seduti vicino a un diffusore, sentirete qualcosa che è molto vicino a voi, molto forte e, in seguito, non sarete in grado di sentire niente; vi concentrate sul violino. E' come se lo spazio intorno al violino cambiasse, avvicinandosi o allontanandosi. La percezione di questo è assolutamente individuale, dato che dipende dal luogo in cui è posizionato l'ascoltatore.

Ho già accennato al fatto che le macchine sono in grado di compiere scelte aleatorie ma ciò che è interessante è fornire a tali scelte campi d'azione prefissati. La prossima cosa che faremo è scegliere un campo d'azione che corrisponda al campo d'azione del violino, settando quello che chiamiamo un *cluster*. La macchina sceglierà valori interni al cluster, che produrranno accordi o, se preferite, cluster *parziali*, i quali sono tenuti. Suoniamo l'accordo molto rapidamente e quindi lo lasciamo riverberare. Prendiamo una decisione, settiamo la macchina di conseguenza e la lasciamo lavorare. [esempi di tre accordi generati dal computer] Così, vedete, c'è sempre l'infame RE, di cui ho già parlato. E' un elemento base dell'accordo e deve essere percepito con chiarezza. Quindi si crea una sorta di alone, intorno a questo RE, che cambia con ogni accordo e che è definito dall'iniezione di un certo numero di note durante un periodo di tempo molto breve, provocando il continuo cambiamento dell'accordo. Se ripetiamo il pezzo, questo non si ripete mai nello stesso modo. Questo è un modo per introdurre il sistema aleatorio in un contesto deterministicamente fissato con precisione.

L'ultima cosa di cui vorrei parlare è il fatto che, senza modificare le sonorità del violino, noi trasgrediamo le possibilità dello strumento. Abbiamo prodotto *pizzicati* che non potrebbero mai essere eseguiti a quella velocità, poiché sarebbe molto semplicemente impossibile sul piano *fisico*. Ciò che è interessante è l'utilizzo del suono strumentale nella sua nudità, nella sua esattezza ma *oltre* certe possibilità delle dita. [esempio di pizzicati prodotti elettronicamente] Questi sono semplicemente degli arpeggi ma, di fatto, non sono semplici come sembrano; essi sono arpeggi *intersecati*, di una certa complessità di costruzione, che vengono costantemente modificati in modi diversi. E ancora tutto è scritto in relazione al

violino, ed è triggerato dal violino stesso. [esempio dall'inizio della settima sezione]

Gli arpeggi sono triggerati dallo 'score follower'; seguendo le note corte, a un certo punto c'è una nota tenuta lunga trillata che è ciò che in definitiva triggera l'arpeggio. A volte – non siamo bene sicuri del perché, che rimane da analizzare – sebbene in genere funzioni, capita che il dinosauro perda per strada qualche dente... Noi ci assicuriamo che il dinosauro non li perda, i denti, e in generale, questo è fattibile, in tempo reale, seguendo la partitura. La partitura viene immessa nel computer e questo segue, con l'aiuto della partitura, tutto quello che accade, sebbene in pratica sia estremamente difficile. Abbiamo iniziato facendo seguire al computer partiture di strumenti che lo permettevano. Con questo intendo strumenti quali il flauto, poiché c'è una posizione specifica delle dita per ciascuna nota. Avendo intersezioni tra due variabili, il computer non fa errori, nel novantanove per cento dei casi. Col violino, comunque, non c'è una posizione specifica delle dita triggerata da tasti, o fori, o tubi, come nel caso del clarinetto o del flauto, il che rende la cosa molto più difficile. Ecco perché, a volte, dobbiamo riferirci a categorie che sono molto più superficiali, come ad esempio semplicemente distinguere tra passaggi in cui c'è molta attività e altri in cui non ce n'è. C'è una serie di note rapide che io non sono in grado di seguire, devo aspettare di trovare il segno dalla nota lunga che viene dopo. E questo è, di fatto, come opera il computer qui: non segue le fioriture rapide, salta direttamente alla nota lunga che segue in partitura. Così vedete come a volte siamo costretti a operare con funzioni che sono molto più superficiali di quanto uno vorrebbe, e comporre con le nozioni che il computer ci fornisce.

Traduzione dall'inglese di Marco Marinoni

Un metodo di restauro per documenti sonori di musica etnica basato sulla Non-negative Matrix Factorization e su una regola di soppressione percettiva

Giuseppe Cabras

Dip. di Ingegneria Elettrica,
Gestionale e Meccanica
Università di Udine, Italia
giuseppe.cabras@uniud.it

Sergio Canazza

Sound and Music Computing Group,
Dip. di Ingegneria dell'Informazione
Università di Padova, Italia
canazza@dei.unipd.it

Pier Luca Montessoro, Roberto Rinaldo

Dip. di Ingegneria Elettrica,
Gestionale e Meccanica
Università di Udine, Italia
montessoro@uniud.it

ABSTRACT

Nell'ambito del restauro audio, il metodo più diffuso è l'attenuazione spettrale a breve termine (*Short Time Spectral Attenuation*, STSA). Tuttavia, pur riducendo il rumore e migliorando il rapporto segnale rumore (SNR), spesso tende a introdurre distorsione nel segnale e un disturbo residuo noto come rumore musicale (*musical noise*): un artefatto artificiale a tempo variante, isolato, casuale e tonale.

Questo lavoro presenta un nuovo algoritmo per il restauro audio basato sulla *Non-negative Matrix Factorization* (NMF) e su di una regola di soppressione del rumore percettiva basata su modelli psicoacustici avanzati per calcolare la soglia di mascheramento del rumore a partire dalla stima della sorgente.

Prove sistematiche a basso SNR ($< 10\text{dB}$) con misure PESQ mostrano che il metodo non introduce rumore musicale e permette di controllare il trade-off tra la soppressione di componenti indesiderate e l'attenuazione della sorgente.

In particolare, verrà provato che la tecnica NMF è appropriata a isolare il segnale audio pulito dal rumore non stazionario indesiderato in registrazioni monoaurale di musica etnica.

Inoltre, è stato effettuato un esperimento percettivo utilizzando il metodo di valutazione EBU MUSHRA per comparare NMF con lo stato dell'arte dei sistemi di restauro audio.

Gli ottimi risultati ottenuti incoraggiano ad applicare questa metodologia in modo intensivo al restauro dei documenti sonori di importanti archivi audio.

1. INTRODUZIONE

Il patrimonio della musica etnica – spesso l'unica testimonianza di culture orali estinte – rischia di scomparire: i documenti audio sono spesso registrati su supporti non professionali con sistemi di registrazione amatoriale. Quindi, per una loro fruizione appropriata o un accesso idoneo con tecniche di *Music Information Retrieval* (MIR), è necessario elaborare i segnali con algoritmi di restauro audio.

Negli ultimi dieci anni la ricerca nel campo del restauro audio si è focalizzata sulla progettazione di algoritmi che sottendono una pluralità di modelli e di ipotesi sulla realtà sonora, sviluppati in relazione al particolare approccio al documento per cui il sistema viene utilizzato:

- Approccio documentale: in questo caso gli algoritmi di restauro non devono superare il livello tecnologico dell'epoca di registrazione.
- Approccio estetico: modifica la qualità sonora in funzione dei gusti dell'utente attuale (sia per nuove edizioni commerciali sia per preparare il segnale all'applicazione di tecniche MIR).
- Approccio sociologico: è finalizzato a ottenere una ricostruzione storica della registrazione, come veniva ascoltata all'epoca (vedi Storm, Type I [1]).
- Approccio ricostruttivo: ha l'obiettivo di conservare l'intenzione dell'autore (vedi Storm, Type II [1]).

Per soddisfare uno o più dei punti sopraelencati, è necessario avere a disposizione diversi strumenti di restauro audio (di frequente, nello stesso documento audio ci sono corrotte con caratteristiche fisiche differenti, che possono essere attenuate con filtri di soppressione del rumore differenti). Gli algoritmi di restauro audio possono essere suddivisi in tre categorie [2]:

1. metodi nel dominio delle frequenze, come diverse forme di filtri di Wiener non-causale o schemi di sottrazione spettrale e algoritmi recenti che cercano di incorporare la conoscenza del sistema dell'udito umano; tutti questi metodi fanno uso di poca informazione a priori;
2. il restauro nel dominio del tempo attraverso modelli del segnale come nell'Extended Kalman Filter (EKF): in questi metodi è necessaria molta informazione a priori per poter stimare i parametri statistici degli eventi audio;
3. restauro attraverso il modello della sorgente: si usa solo informazione a priori.

I metodi nel dominio delle frequenze sono relativamente semplici da implementare. Tuttavia, sono affetti dalle seguenti limitazioni: il rumore musicale (sinusoidi

Copyright: ©2010 Giuseppe Cabras et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Trasformata di Fourier a breve termine (*Short Time Fourier Transform*, STFT) che rivela le variazioni di energia del segnale in elementi temporali (detti *frame*) ed elementi di frequenza (detti *bins*), fornendo quindi una rappresentazione matriciale non-negativa. Nel seguito, il segnale in un dominio tempo vs. log-frequenza sarà rappresentato come una STFT elevata a potenza per elemento:

$$X = |STFT\{x(t)\}|^\gamma \quad (3)$$

La linearità espressa dall'Eq. 2 si applica anche all'Eq. 3 quando $\gamma = 2$, ma la pratica sperimentale evidenzia che γ è un parametro importante per regolare l'efficacia di separazione con NMF. In particolare, risulta che $\gamma = 2$ non è una buona scelta per la separazione delle componenti, mentre una scelta ottimale è $\gamma = 2/3$, che corrisponde alla compressione radice cubica del STFT di potenza. Sorprendentemente, questo risultato è consistente con l'esponente della Legge di Potenza di Stevens per l'intensità percepita della pressione sonora di uno stimolo tonale di 3 kHz. Inoltre, la Legge di Potenza di Stevens è stata usata per modellare le non-linearità della coclea [6] e la conversione da intensità acustica a intensità percepita in tecniche di analisi della voce parlata, note come Perceptual Linear Predictive (PLP) [7]. Più di recente, Plourde e Champagne inserirono la non-linearità compressiva della coclea nella stima Bayesiana per il miglioramento della voce parlata con tecniche *Short Time Spectral Attenuation* (STSA) [8]. Questa curiosa coincidenza del valore dell'esponente γ , suggerisce l'approccio motivato dalla percezione nel denoising audio, come spiegato nella Sez. 2.5.

2.2 Voice Activity Detection

Il *Voice Activity Detector* (VAD) è un componente largamente usato nei metodi di miglioramento della voce parlata, per aggiornare la stima dello spettro del rumore frame per frame. Nella implementazione qui proposta è stato usato un VAD basato su un modello statistico per costruire due matrici quadrate binarie diagonali:

$$A(t, t) = \begin{cases} 1, & \text{se il bersaglio è presente nel frame } t \\ 0, & \text{altrimenti.} \end{cases} \quad (4)$$

ed il suo complemento $\bar{A}(t, t)$.

Questo ci permette di apprendere il dizionario delle componenti indesiderate, calcolando NMF sul segnale:

$$Z(f, t) = X(f, t)\bar{A}(t, t) \quad (5)$$

nei periodi in cui il bersaglio è assente, e successivamente separare il dizionario delle componenti di interesse, calcolando una NMF modificata (NMF^*) sul segnale:

$$Y(f, t) = X(f, t)A(t, t) \quad (6)$$

nei periodi in cui il bersaglio è presente. Assumendo che il bersaglio e le componenti indesiderate siano additive (come assunto nell'Eq. 1), il modulo VAD deve decidere, per ogni

frame t , a favore di una tra due ipotesi:

$$H_0 : X_f = N_f : \quad \text{bersaglio assente,} \quad (7)$$

$$H_1 : X_f = S_f + N_f : \quad \text{bersaglio presente.} \quad (8)$$

È stata usata la seguente regola di decisione basata sul Test del Rapporto di Verosimiglianza (LRT, Likelihood Ratio Test):

$$\frac{1}{L} \sum_{f=1}^{L-1} \log \Lambda_f \underset{H_0}{\overset{H_1}{\geq}} \eta \quad (9)$$

dove il rapporto di verosimiglianza per l' f mo bin è:

$$\Lambda_f = \frac{p(X_f|H_1)}{p(X_f|H_0)} = \frac{1}{1 + \xi_f} \exp\left(\frac{\gamma_f \xi_f}{1 + \xi_f}\right) \quad (10)$$

dove ξ_f e γ_f definiscono rispettivamente il SNR a priori e a posteriori. In particolare, ξ_f viene stimato usando un approccio *decision directed* (con $\alpha = 0.99$) come in [9], L è la lunghezza della FFT e η è una soglia definita dall'utente. UN algoritmo di *hang-over* basato su Hidden Markov Model (HMM) estende e rende più *smooth* la decisione del VAD per recuperare periodi bersaglio che sono mascherati dalla componente indesiderata. La Fig. 2 mostra un esempio di VAD applicato a una registrazione reale (musica etnica su disco Shellac a 78 rpm) con voce cantata e accompagnamento musicale¹.

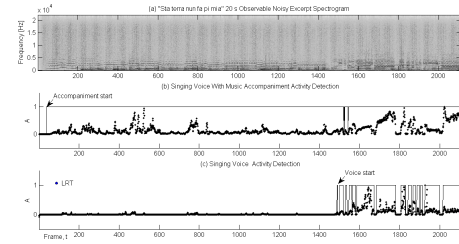


Figure 2. VAD statistico al lavoro. (sopra) Nello spettrogramma dell'estratto pre-elaborato della registrazione originale, si riconosce la struttura armonica immersa in rumore a banda larga (ciclo)stazionario. (centro) Classificazione del contenuto musicale fornendo al VAD l'informazione a priori dell'"estensione del tempo iniziale di assenza del bersaglio" = 0.46 s e soglia $\eta = 0$. (sotto) Classificazione della voce cantata fornendo al VAD l'informazione a priori dell'"estensione del tempo iniziale di assenza del bersaglio" = 17 s e soglia $\eta = 0.1$.

2.3 Apprendimento delle componenti indesiderate

Assumendo la disponibilità di frame con bersaglio-assente, rilevati applicando il VAD al segnale osservabile $X(f, t)$ durante la fase di apprendimento, il segnale risultante $Z(f, t)$ dell'Eq. 5 sarà equivalente a $X(f, t)$ con i frame bersaglio-presente soppressi. Applicando a $Z(f, t)$ una NMF regolarizzata e con distanza Euclidea, otteniamo le matrici positive del dizionario $D_n(f, k)$ e del codice sparso $H_n(k, f)$,

¹ *Sta terra nun fa pi mia*, di R. Gioiosa, arr. R. Romani - 78 rpm 10" Brunswick 58073B (E 26621/2), registrato a New York, 23/02/1928, durata 3'22".

dove k è il numero degli elementi del dizionario di interferenza definito dall'utente. Seguendo la semplificazione proposta da [5], viene qui definita la seguente elaborazione moltiplicativa e iterativa di H_n e D_n :

$$\hat{X}_n = \bar{D}_n H_n; H_n \leftarrow H_n \bullet \frac{\bar{D}_n^T Z}{\bar{D}_n^T \hat{X}_n + \lambda_n}; \quad (11)$$

$$D_n \leftarrow \bar{D}_n \bullet \frac{Z H_n^T + \bar{D}_n \bullet (1(\hat{X}_n H_n^T \bullet \bar{D}_n))}{\hat{X}_n H_n^T + \bar{D}_n \bullet (1(Z H_n^T \bullet \bar{D}_n))}. \quad (12)$$

Dove $\bar{D}_n$ è la normalizzazione Euclidea per colonne di D_n nell'iterazione corrente (vedi Sez. 2.4), l'operatore $\bullet$ indica la moltiplicazione per elemento, la linea di frazione indica la divisione per elemento e 1 è una matrice quadrata di 1. Il parametro di regolarizzazione λ_n pondera l'importanza del termine di parsimonia (sparsity) nella ricostruzione.

La matrice risultante D_n rappresenta il dizionario dell'interferenza appresa dai dati e verrà usato nel modulo successivo per stimare le due sorgenti additive che compongono il segnale miscelato.

2.4 Stima delle sorgenti bersaglio e interferenza

Per stimare le sorgenti, utilizziamo nuovamente una NMF vincolata (che indichiamo con NMF*) per calcolare il restante dizionario della sorgente bersaglio e il codice sparso di entrambe le sorgenti. Assumendo, come di consueto, l'additività delle sorgenti, il dizionario del segnale miscelato può essere visto come la concatenazione per colonne dei dizionari delle singole sorgenti. Inoltre, il codice sparso del segnale miscelato può essere visto come la concatenazione per righe dei codici sparsi delle singole sorgenti:

$$X = X_s + X_n = [D_s D_n] \begin{bmatrix} H_s \\ H_n \end{bmatrix} + E = DH + E \quad (13)$$

Nell'equazione precedente, E è una matrice ignota che rappresenta l'errore di approssimazione. Non possiamo risolvere direttamente con NMF l'Eq. 13, a causa di una ambiguità di permutazione. Infatti, possiamo scrivere:

$$DH = (DP)(P^{-1}H) \quad (14)$$

dove P è una matrice di permutazione generalizzata, cioè con un solo elemento positivo non nullo in ogni riga e in ogni colonna della matrice.

Schmidt, Larsen e Hsiao [10] suggeriscono di pre-calcolare D_n come abbiamo fatto nella precedente sezione per l'interferenza nel segnale $Z(f, t)$; quindi apprendere $D_s(f, m)$, $H_s(m, t)$ e $H_n(k, t)$, con m numero degli elementi del dizionario della sorgente bersaglio definito dall'utente, attraverso una NMF vincolata modificata che viene applicata a $Y(t, f)$ nell'Eq. 6, cioè il segnale osservato nei frame bersaglio-presente. Descriviamo di seguito l'algoritmo vincolato da 1-dizionario (D_n^*):

1. Inizializza $D_s(f, m)$, $H_s(m, t)$ e $H_n(k, t)$ con valori casuali compresi tra $[0 \div 1]$; moltiplica $H_s(m, t)$ e $H_n(k, t)$ con A per sopprimere i frame bersaglio-assente.

2. Definisci una normalizzazione Euclidea per colonne del dizionario bersaglio per prevenire una deriva numerica congiunta in H_s e D_s :

$$\bar{D}_s(f, m) = \frac{D_s(f, m)}{\sqrt{\sum_f D_s(f, m)^2}} = \frac{D_s(f, m)}{\|D_s(m)\|_2}. \quad (15)$$

3. Calcola la ricostruzione complessiva con:

$$\hat{X} = \bar{D}_s H_s + \bar{D}_n H_n. \quad (16)$$

4. Aggiorna il codice sparso del bersaglio con la regola:

$$H_s \leftarrow H_s \bullet \frac{\bar{D}_s^T Y}{\bar{D}_s^T \hat{X} + \ell_s}. \quad (17)$$

5. Calcola la ricostruzione complessiva come nell'Eq. 16.

6. Aggiorna il codice sparso dell'interferenza con la regola:

$$H_n \leftarrow H_n \bullet \frac{\bar{D}_n^T Y}{\bar{D}_n^T \hat{X} + \ell_n}. \quad (18)$$

7. Calcola la ricostruzione complessiva come nell'Eq. 16.

8. Aggiorna il dizionario non normalizzato del bersaglio con la regola:

$$D_s \leftarrow \bar{D}_s \bullet \frac{Y H_s^T + \bar{D}_s \bullet (1(\hat{X} H_s^T \bullet \bar{D}_s))}{\hat{X} H_s^T + \bar{D}_s \bullet (1(Y H_s^T \bullet \bar{D}_s))}. \quad (19)$$

9. Ripeti dal passo 2 fino a convergenza della funzione costo Euclidea da minimizzare:

$$C^{(i)} = \frac{1}{2} \sum_{f,t} (Y(f, t) - \hat{X}(f, t))^2 + \ell_n \sum_{k,t} H_n(k, t) + \ell_s \sum_{m,t} H_s(m, t). \quad (20)$$

L'algoritmo termina all'iterazione i quando $|C^i - C^{i-1}| < \varepsilon C^i$. I parametri di regolarizzazione ℓ_s e ℓ_n determinano il grado di *sparsity* nelle matrici di attività. Il dizionario delle componenti indesiderate, D_n , non viene aggiornato dall'algoritmo perchè predefinito e fissato dalla precedente fase di apprendimento; inoltre, non viene cercato un codice sparso per il dizionario fissato, ma il codice che minimizza l'errore di ricostruzione, ponendo $\ell_n = 0$. In generale, λ_n , ℓ_s , k ad m dipendono dalle sorgenti incognite; con i nostri dati di prova, abbiamo ottenuto buoni risultati con $\lambda_n = 0.2$ e $\ell_s = 0.05$, $k = 256$ ed $m = 256$, confermando, in un ambito applicativo più vasto, i risultati di Schmidt et al. [10].

2.5 Regola di soppressione Bayesiana percettiva

Le stime di D_s , H_s , D_n e H_n , ottenute con i due precedenti passi, ci permette di stimare gli spettrogrammi bersaglio e interferenza nei frame in cui il bersaglio è presente:

$$\hat{X}_s = D_s H_s \quad (21)$$

$$\hat{X}_n = D_n H_n \quad (22)$$

La figura 3 mostra gli spettrogrammi stimati $\hat{X}_s$ e $\hat{X}_n$ dove la componente indesiderata è il rumore a banda larga ciclo-stazionario presente nell'estratto originale numero 4 della Sez. 3.

Equindi possibile ricostruire la sorgente bersaglio tempo-variante tramite una regola di soppressione del rumore, tecnica ben nota nell'ambito del miglioramento della voce parlata e nella riduzione del rumore dell'audio in generale. Una regola di soppressione può essere vista come un guadagno $G(f, t)$, variabile in tempo-frequenza a valori reali non-negativi, applicato allo spettro complesso dell'osservabile $X(f, t)$ per ottenere lo spettro complesso della sorgente bersaglio:

$$\hat{S}(f, t) = G(f, t) \bullet Y(f, t) \text{ con } 0 \leq G(f, t) \leq 1 \quad (23)$$

le lettere in grassetto indicano numeri complessi. Sebbene in molti casi, con SNR elevati, possiamo ottenere buone ricostruzioni della sorgente bersaglio con il semplice filtro di Wiener, per SNR modesti, aumentano sia la distorsione del bersaglio che l'ancor più fastidioso rumore musicale (rumore tonale variabile nel tempo in modo isolato e casuale). Incorporando il fenomeno del mascheramento del modello dell'udito umano nella stima della sorgente bersaglio, possiamo ridurre la soppressione del rumore e favorire una migliore fedeltà audio o intelligibilità della voce. Un uditore, infatti, tollera l'interferenza additiva sino a che l'energia dell'interferenza rimane al di sotto della soglia di mascheramento definita dall'energia della sorgente bersaglio: non è necessario sopprimere l'interferenza mascherata, in quanto non udibile; in questo modo, viene soppresso solo la parte in eccesso dell'interferenza non mascherata.

Un modello di mascheramento efficace, semplice largamente impiegato, venne proposto da Johnston [11] per mascherare la distorsione introdotta nei processi vocali e audio, in seguito adottata con successo in [12] nel miglioramento della voce parlata. In questo modello psicoacustico, una interferenza debole ad una certa frequenza è resa non udibile da un bersaglio di maggiore intensità e simultaneo, cioè nello stesso frame temporale, all'interno dello stesso intervallo percettivo di frequenze, detta Banda Critica (Critical Band) o Bark, e al di fuori della stessa Banda Critica, applicando una convoluzione con una funzione di *spread*. Il calcolo della soglia di mascheramento di Johnston non tiene conto del mascheramento temporale in avanti o in dietro.

Secondo Wolfe e Godsill [3], è possibile formulare una misura di distorsione basata sulla percezione come una generalizzazione della funzione costo dell'Errore Quadratico

Medio Minimizzato (MMSE, Minimum Mean Square Error) che incorpora esplicitamente la soglia di mascheramento:

$$C_{WG}(S, \hat{S}, T) = \begin{cases} \left(\hat{S} - S - \frac{T}{2} \right)^2 - \left(\frac{T}{2} \right)^2, & \text{se } \left| \hat{S} - S - \frac{T}{2} \right| > \frac{T}{2}; \\ 0, & \text{altrimenti.} \end{cases} \quad (24)$$

dove S è l'ampiezza della STFT della sorgente ideale incognita, $\hat{S}$ è l'ampiezza della STFT della sorgente stimata e T è la soglia di mascheramento; per semplicità sono stati omessi gli indici di frequenza f e di frame t . Nella Eq. 24, si vede che se l'errore di stima è inferiore alla soglia di mascheramento, allora il costo assegnato è zero, mentre il costo è penalizzato (ossia incrementato) solo se l'errore di stima supera la soglia di mascheramento. Ciò previene l'attenuazione non voluta della sorgente quando la componente indesiderata è mascherata e sopprime solo la componente indesiderata percepibile dall'udito umano. Sfortunatamente, la minimizzazione di $E[C_{WG}(S, \hat{S}, T)]$ non è trattabile analiticamente e infatti gli autori hanno adottato una implementazione numerica [3].

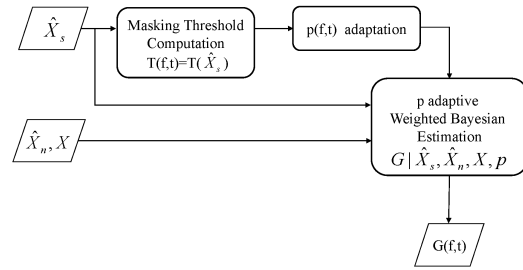


Figure 4. Schema proposto per la regola di soppressione percettiva, basato su uno stimatore Bayesiano pesato con peso adattivo secondo la soglia della maschera.

Nel nostro modello motivato dalla percezione, schematizzato in fig. 4, abbiamo seguito un approccio diverso, proposto da Loizou [13], dove il criterio percettivo è implementato implicitamente, gravando l'ampiezza dell'errore STFT con un filtro che ha la forma dell'inverso dell'ampiezza STFT della sorgente, in modo tale che si pone minore enfasi in presenza di picchi delle formanti (implicitamente mascherate) e si pone maggior enfasi in presenza delle valli spettrali (implicitamente non mascherate):

$$C_{WE}(S, \hat{S}, p) = (\hat{S} - S)^2 \bullet S^p \text{ con } -2 < p \leq 0 \quad (25)$$

dove p è un parametro in tempo-frequenza a valori reali che enfatizza le valli spettrali quando è negativo. Questa funzione costo è nota come misura di distorsione Euclidea pesata; assumendo $S(f, t)$ modellata con variabili casuali Gaussiane a media nulla e statisticamente indipendenti, la minimizzazione analitica di $E[C_{WE}(S, \hat{S}, p)]$ determina il seguente guadagno in tempo-frequenza $G_{WE}(f, t)$:

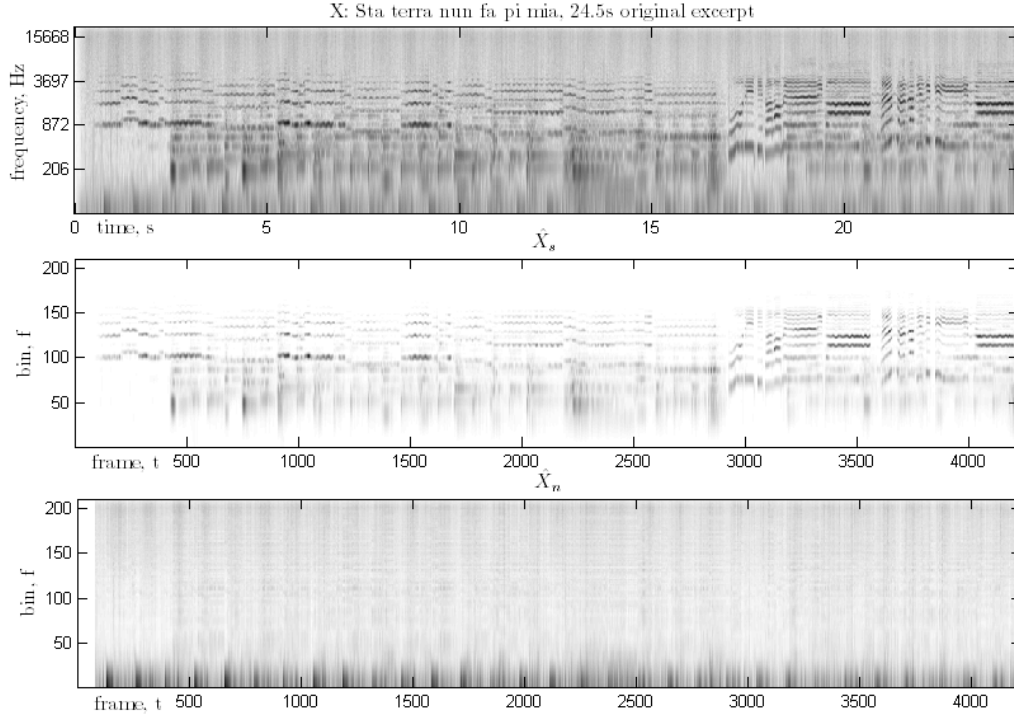


Figure 3. Spettrogrammi del segnale rumoroso originale X (top), sorgente target stimata $\hat{X}_s$ (center) ed interferenza stimata $\hat{X}_n$ (bottom) dall'estratto autentico di 24.5 secondi intitolato 'Sta terra nun fa pi mia' (vedi Sez. 3 voce 4), gli spettrogrammi sono rappresentati in log frequenza, da $f_{\min} = 50\text{Hz}$ e risoluzione 24bin/octave. La struttura armonica e il rumore a banda larga ciclo-stazionario sono chiaramente separati.

$$G_{WE} = \frac{\sqrt{\nu}}{\gamma} \bullet \frac{\Gamma\left(\frac{p+1}{2} + 1\right) \bullet \Phi\left(-\frac{p+1}{2}, 1; -\nu\right)}{\Gamma\left(\frac{p}{2} + 1\right) \bullet \Phi\left(-\frac{p}{2}, 1; -\nu\right)}, \quad p > -2 \quad (26)$$

dove:

$$\gamma = \frac{X}{\hat{X}_n}; \quad \xi = \frac{\hat{X}_s}{\hat{X}_n}; \quad \nu = \frac{\xi}{1 + \xi} \bullet \gamma; \quad (27)$$

$\Gamma(\cdot)$ è la funzione gamma e $\Phi(a, b; z)$ è la funzione ipergeometrica confluyente. Se $p = 0$, l'Eq. 26 si riduce al classico stimatore MMSE-STSA (Minimum Mean Square Error Short Time Spectral Amplitude) di Ephraim e Malha [14].

Consideriamo ora una funzione costo simile, anch'essa proposta da Loizou [13], chiamata misura di distorsione COSH pesata (Weighted Cosh):

$$C_{WCOSH}(S, \hat{S}, p) = \left(\frac{S}{\hat{S}} + \frac{\hat{S}}{S} - 1 \right) \bullet S^p \quad \text{con } -1 < p \leq 0 \quad (28)$$

anche questa funzione costo enfatizza le valli spettrali quando p è negativo e determina il guadagno $G_{WCOSH}(f, t)$:

$$G_{WCOSH} = \frac{\sqrt{\nu}}{\gamma} \bullet \sqrt{\frac{\Gamma\left(\frac{p+3}{2}\right) \bullet \Phi\left(-\frac{p+1}{2}, 1; -\nu\right)}{\Gamma\left(\frac{p+1}{2}\right) \bullet \Phi\left(-\frac{p-1}{2}, 1; -\nu\right)}}, \quad p > -1 \quad (29)$$

Il terzo stimatore Bayesiano implementato è noto come stimatore MMSE STSA di ordine β (β SA), proposto in [15] e sviluppato successivamente in [16]. Si tratta di una generalizzazione dello stimatore MMSE-STSA classico [14] [9] ottenuto introducendo l'esponente β (che per uniformità con gli stimatori precedenti continueremo a chiamare p):

$$C_{SA}(S, \hat{S}, p) = (\hat{S}^p - S^p)^2 \quad \text{con } -2 < p < 0 \quad (30)$$

preso $p < 0$, il meccanismo di penalizzazione della funzione costo è simile alla misura di distorsione Euclidea pesata dell'Eq. 25 ed entrambe producono una stima accurata della sorgente nelle valli spettrali. La funzione guadagno $G_{SA}(f, t)$ determinata dallo stimatore β SA è la seguente:

$$G_{SA} = \frac{\sqrt{\nu}}{\gamma} \bullet \left[\Gamma\left(\frac{p}{2} + 1\right) \bullet \Phi\left(-\frac{p}{2}, 1; -\nu\right) \right]^{1/p}, \quad p > -2 \quad (31)$$

Se $p \rightarrow 0$, allora la β SA corrisponde allo stimatore MMSE log-STSA di Ephraim e Malah [9].

Verifiche dei tre stimatori Bayesiani, condotte su frasi di voce parlata a basso SNR ($< 10\text{dB}$) e valutate con misure PESQ e prove di ascolto informali, mostrano prestazioni simili e di buona qualità: non introducono rumore musicale e permettono il controllo del trade-off tra soppressione della componente indesiderata e attenuazione della sorgente variando il parametro p . Alcune valutazioni delle prestazioni ed i corrispettivi esempi audio di comparazione

sono disponibili in <http://dialogo.fisica.uniud.it/BASS/ComparisonWithGustafsson02>.

Indubbiamente, la scelta ottimale del parametro reale $p(f, t)$ è un elemento centrale per ottimizzare la prestazione complessiva ottenibile; a questo scopo è stata esplicitata, in modo euristico, la relazione tra la soglia di mascheramento $T(f, t)$ e il parametro $p(f, t)$.

Si è visto che le componenti indesiderate possono essere ridotte diminuendo p . Tuttavia, ciò comporta una maggiore distorsione della sorgente. Si può quindi definire un meccanismo adattivo sulla base della seguente considerazione: se la soglia di mascheramento è alta, l'interferenza sarà mascherata e quindi non udibile; di conseguenza, non c'è bisogno di ridurla, contenendo la distorsione ai valori minimi. In questo caso, il parametro p viene mantenuto al suo valore massimo: $p = p_{max}$. Però, se la soglia di mascheramento è bassa, la componente indesiderata può essere particolarmente fastidiosa all'udito ed è quindi necessario ridurla. Ciò si ottiene diminuendo progressivamente p verso il suo valore minimo: $p = p_{min}$. Per ogni frame t , il valore minimo della soglia di mascheramento $T(f, t)$ corrisponde al minimo del parametro $p(f, t)$. Allo scopo di evitare discontinuità nella funzione guadagno G dovute al meccanismo adattivo, viene utilizzata una funzione di *smooth*, controllata dal parametro x definito dall'utente. Il meccanismo adattivo del parametro $p(f, t)$ è quindi realizzato dalla seguente equazione:

$$p(f, t) = \left(\frac{T(f, t) - T_{min}(t)}{T_{max}(t) - T_{min}(t)} \right)^x (p_{max} - p_{min}) + p_{min} \quad (32)$$

dove $T_{max}(t)$ e $T_{min}(t)$ sono rispettivamente il valore massimo e minimo della soglia di mascheramento $T(f, t)$ nel frame corrente t . In questo modo, $p(f, t)$ si adatta alla minima riduzione di interferenza in corrispondenza del valore massimo della soglia di mascheramento (cioè in corrispondenza dei picchi delle formanti della sorgente) e alla massima riduzione per il valore minimo della soglia di mascheramento (cioè in corrispondenza delle valli spettrali). La figura 5 mostra le semplici curve di *smoothing* ottenute con l'Eq. 32 variando il parametro di *smoothing* x .

Il valore minimo e massimo di p ed x determinano il trade-off tra rumore residuo e distorsione della sorgente. Sono stati condotti diversi esperimenti con differenti tipi di rumore a diversi livelli per scegliere il valore appropriato per questi parametri.

Nel caso di interferenza additiva ed un SNR prossimo a 0 dB, per ottenere un buon trade-off per l'uditore, sono stati scelti i seguenti valori per ogni stimatore Bayesiano considerato:

1. Per lo stimatore Euclideo pesato: $p_{WE_{max}} = 0$, $p_{WE_{min}} = -1.98$;
2. Per lo stimatore COSH pesato: $p_{WCOSH_{max}} = 0$, $p_{WCOSH_{min}} = -0.99$;
3. Per lo stimatore β SA: $p_{max} = -0.001$, $p_{min} = -1.98$;

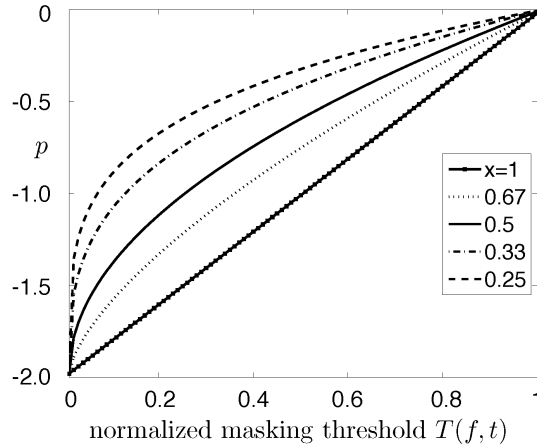


Figure 5. Il parametro p al variare della soglia di mascheramento normalizzata T e parametro di *smoothing* $x = (0.25, 0.33, 0.5, 0.67, 1)$, $p_{min} = -1.98$, $p_{max} = 0$.

L'esponente di *smoothing* è lo stesso per tutti gli stimatori: $x = 0.5$.

Questo trade-off può essere facilmente modificato in base all'applicazione; in genere, per ottimizzare il trade-off è sufficiente regolare solo p_{min} .

L'uso della stima della soglia di mascheramento simultaneo di Johnston permette la costruzione di regole di soppressione del rumore sofisticate ed efficaci. Tuttavia, se la soglia non è stimata correttamente, le prestazioni che si ottengono soffrono pesantemente in termini di fastidioso rumore musicale introdotto nella sorgente audio stimata, compromettendo la bontà di qualsiasi regola di soppressione del rumore. Per stimare correttamente la soglia, è necessario effettuare una stima estremamente accurata dello spettro della sorgente bersaglio $\hat{X}_s(f, t)$, che abbiamo ottenuto attraverso l'uso di NMF.

3. VALUTAZIONE SOGGETTIVA

Per validare il sistema, è stato condotto un esperimento percettivo. Come materiale audio di prova, sono stati utilizzati alcuni documenti sonori di musica etnica.

Materiale. Sono stati usati quattro brani musicali registrati su dischi Shellac. Per minimizzare la fatica dell'ascolto e massimizzare l'attenzione dei soggetti partecipanti (valutatori), sono stati estratti solo i 20 secondi iniziali di ogni brano. Poiché la valutazione è più orientata alla comparazione tra gli stimoli che alla loro analisi individuale, questi brevi estratti sembrano essere adeguati allo scopo:

1. *Chi campa deritto campo afflito* di Perrocato e Canoro, Eduardo Migliaccio (voce) - 78 rpm 10" Victor 14-81712-B (BVE 46692-2), registrato a New York il 14/08/1928, durata 3'36". Nell'estratto considerato: voce cantata e musica.
2. *Il funerale di Rodolfo Valentino*, Compagnia Columbia (2 cantanti uomo, 2 cantanti donna, campane ed orchestra) - 78 rpm 10" Columbia 14230-F (w 107117 2), registrato a New York nel settembre 1926, durata

2'55". Nell'estratto considerato: voci parlate, campanone ed orchestra.

3. *La signorina sfinciusa*, Leonardo Dia (voce), Alfredo Cibelli (mandolino), sconosciuti (2 chitarre) - 78 rpm 10" Victor V-12067-A (BVE 53944-2), registrato a New York il 24/07/1929, durata 3'20". Nell'estratto considerato: voce cantata e musica.
4. *Sta terra nun fa pi mia* di R. Gioiosa, arr. R. Romani, Rosina Gioiosa Trubia (voce), Alfredo Cibelli (mandolino), sconosciuti (2 chitarre) - 78 rpm 10" Brunswick 58073B (E 26621/2), registrato a New York, il 23/02/1928, durata 3'22". Nell'estratto considerato: voce cantata e musica.

Tutti gli stimoli rumorosi sono stati pre-elaborati per mezzo del Filtro di Kalman Estes (EKF), presentato in [17], in modalità *de-click*; il restauro a banda larga è stato eseguito utilizzando il nostro metodo con la regola di soppressione esposta in 2.5, e con i seguenti tre prodotti commerciali, scelti tra i prodotti più apprezzati negli archivi audio e studi di post-produzione:

1. X-Noise del Waves Restoration bundle (Waves V6 Update 2);
2. Denoiser (con filtro *Musical noise suppression* abilitato) di iZotope RX v1.06;
3. Auto Dehiss di CEDAR Tools;

I plug-ins CEDAR sono usati in un sistema Pro Tools HD. I parametri utente usati per controllare i diversi sistemi sono stati regolati in modo soggettivo al fine di ottenere il miglior trade-off tra rimozione del rumore e conservazione del segnale musicale. Sono stati prodotti quindi 16 stimoli restaurati.

Metodologia. Le prove sono state condotte utilizzando il metodo di test EBU MUSHRA [18], metodo di valutazione raccomandato dall'ITU [19]. Il protocollo si basa sul metodo del "triplo-stimolo a doppio-cieco con riferimento nascosto", che è un metodo stabile e permette di distinguere, in modo accurato, differenze sottili tra gli stimoli da valutare. Una caratteristica importante di questo metodo risiede nella inclusione, tra gli stimoli da confrontare, del riferimento nascosto e di uno o più segnali *ancora*: nel nostro caso due segnali a banda limitata (a 7 kHz e a 3.5 kHz).

Tutti gli stimoli rumorosi di prova sono tratti da registrazioni reali. Ciò implica l'impossibilità di confrontare gli stimoli migliorati con uno stimolo di riferimento di alta qualità (valutato 5.0 in cima alla scala di valutazione), ma con lo stimolo di riferimento rumoroso (valutato 0.0). Inoltre, deve essere possibile assegnare valutazioni negative per valutare stimoli peggiori del riferimento rumoroso. Il riferimento nascosto deve essere almeno valutato 0.0 dall'ascoltatore; tutti gli altri stimoli in prova e le *ancore* possono essere valutati soggettivamente con un punteggio (compreso tra -5.0 e 5.0) che rispecchi la qualità complessiva assegnata dal valutatore ad ogni singolo stimolo.

Fase di addestramento. Lo scopo della fase di addestramento dei valutatori, secondo la specifica MUSHRA, è di permettere ad ogni uditor di: a) prendere confidenza con tutti gli stimoli della prova e farsi un'idea complessiva dell'intervallo di qualità attribuibile agli stimoli in esame; b) apprendere l'uso dell'apparecchiatura per condurre le prove e della scala di valutazione.

Partecipanti. Sono stati selezionati due diversi gruppi di soggetti:

1. Addestrati musicalmente (MT, Musically trained): 12 ricercatori (musicisti e/o musicologi) dell'Università di Padova e 12 tecnici di diversi archivi internazionali dell'audio.
2. Non-addestrati musicalmente (MU, Musically untrained): 16 studenti in Ingegneria dell'Informazione dell'Università di Padova.

Equipaggiamento. Gli stimoli sono stati registrati a 44.1 kHz/24 bit (files sonori non compressi) e riprodotti su un computer Apple PowerBook Pro 2.4 GHz Intel Core 2 Duo con 2 GB di RAM DDR3 a 1067 MHz, fornito di convertitore D/A RME Fireface 400 e cuffia audio AKG K 501. Gli uditori possono riprodurre in qualsiasi ordine gli stimoli di prova, compresi il riferimento nascosto e le due *ancore* a banda limitata.

Durata della prova. La sessione di addestramento per ogni uditor dura approssimativamente 40 minuti ed include una spiegazione sull'apparecchiatura e sulle prove, oltre a una sessione di esercitazione con la scala di valutazione. La fase di valutazione vera e propria è costituita da 4 sessioni di prova (una per ogni brano musicale), ogni sessione contiene 9 stimoli (1 segnale rumoroso, 6 segnali restaurati e 2 *ancore*). Ogni sessione dura, in media, 8 minuti circa. I valutatori possono effettuare una pausa tra diverse sessioni, ma non all'interno della stessa.

Risultati. Per elaborare i risultati della prova, è stata utilizzata l'analisi statistica descritta nelle specifiche di MUSHRA. I risultati sono riassunti nella Tab. 1 con il punteggio di valutazione medio. I risultati forniti da sei uditori (cinque del gruppo MU e uno del gruppo MT) non sono stati utilizzati perché la media dei propri punteggi dei soli riferimenti nascosti risultava maggiore di $+/- 0.5$.

L'intervallo di qualità individuato dal migliore e peggiore sistema di restauro è solo di 0.80 (gruppo MT) e 0.40 (gruppo MU). In generale ci sono solo due sistemi che raggiungono un punteggio > 3.5 : CEDAR e quello sviluppato dagli autori. L'algoritmo qui presentato raggiunge punteggi simili a CEDAR in tutte le sessioni di prova (meglio in quelle del gruppo MU) e risulta essere migliore degli altri software considerati.

4. CONCLUSIONI

Questo studio è centrato sul restauro sonoro di registrazioni monoaurali di musica etnica. In particolare, è stato studiato l'uso della NMF e mostrato che NMF è una tecnica adeguata per estrarre il segnale audio da una registrazione disturbata da rumore additivo a banda larga non stazionaria in uno scenario di basso SNR. Inoltre, è stata introdotta

Table 1. Media dei punteggi degli stimoli e ancora, 34 soggetti. MT = Musically trained; MU = Musically untrained.

Sistema di restauro	Gruppo MT	Gruppo MU	Media
Nostro	+3.00	+4.20	+3.60
CEDAR	+3.40	+4.00	+3.70
Waves	+2.80	+3.80	+3.30
iZotope RX	+2.20	+3.80	+3.00
Ancora 7 kHz	-2.69	+0.20	-1.02
Ancora 3.5 kHz	-5.00	-4.20	-4.60

una regola di soppressione del rumore percettiva basata su modelli psicoacustici avanzati (Sez. 2.5). Al fine di valutare le prestazioni del metodo proposto, è stata prodotta una prova di valutazione soggettiva di alcuni segnali audio restaurati (Sez. 3). L'esito di queste prove mostrano che il metodo proposto migliora sensibilmente la qualità sonora ed è una valida alternativa ai classici metodi STSA.

Come sviluppi futuri, si prevede l'applicazione sistematica di questo ambiente di restauro audio su due archivi di musica etnica registrata su dischi fonografici.

5. BIBLIOGRAFIA

- [1] W. Storm, "The establishment of international re-recording standards," *Phonographic Bulletin*, vol. 27, pp. 5–12, 1980.
- [2] S. Canazza and A. Vidolin, "Preserving electroacoustic music," *Journal of New Music Research*, vol. 30, no. 4, pp. 351–363, 2001.
- [3] P. J. Wolfe and S. J. Godsill, "Towards a perceptually optimal spectral amplitude estimator for audio signal enhancement," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. II, (Istanbul, Turkey), pp. 821–824, 2000. ISBN 0-7803-6296-9.
- [4] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, no. 401, pp. 788–791, 1999.
- [5] J. Eggert and E. Körner, "Sparse coding and nmf," in *IEEE International Conference on Neural Networks*, pp. 2529–2533, IEEE, 2004.
- [6] R. Meddis, L. P. O'Mard, and E. A. Lopez Poveda, "A computational algorithm for computing nonlinear auditory frequency selectivity," *Journal of the Acoustical Society of America*, vol. 109, no. 6, pp. 2852–2861, 2001.
- [7] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *Journal of the Acoustical Society of America*, vol. 87, pp. 1738–1752, Apr 1990.
- [8] E. Plourde and B. Champagne, "Integrating the cochleas compressive nonlinearity in the bayesian approach for speech enhancement," in *15th EUSIPCO, Poznan, Poland*, pp. 70–74, 2007.
- [9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, 1985.
- [10] M. N. Schmidt, J. Larsen, and F. T. Hsiao, "Wind noise reduction using non-negative sparse coding," in *IEEE Workshop on Machine Learning for Signal Processing*, pp. 431–436, Aug 2007.
- [11] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, 1988.
- [12] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech, Audio Process.*, vol. 7, no. 2, pp. 126–137, Mar. 1999.
- [13] P. C. Loizou, "Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5-2, pp. 857–869, 2005.
- [14] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, 1984.
- [15] S. H. K. C. H. You and S. Rahardja, " β -order mmse spectral amplitude estimation for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 4, pp. 475–486, 2005.
- [16] E. Plourde and B. Champagne, "Further analysis of the β -order mmse stsa estimator for speech enhancement," in *Canadian Conference on Electrical and Computer Engineering. CCECE 2007*, pp. 1594–1597, 2007.
- [17] S. Canazza, G. De Poli, and G. Mian, "Restoration of audio documents by means of extended kalman filter," *IEEE Transactions on Audio, Speech and Language Processing*, vol. in press, 2010.
- [18] ITU-R, "Methods for the subjective assessment of small impairments in audio systems including multi-channel sound systems," *Recommendation BS.1116-1*, 2000.
- [19] EBU Project Group B/AIM, "EBU report on the subjective listening tests of some commercial internet audio codecs," October 2000.

PRESENTAZIONE DELLE ATTIVITÀ DEL GRUPPO DI RICERCA *WORLDS OF AUDIOVISION*

Gianmario Borio
Università degli Studi di Pavia

Giacomo Albert
Università degli Studi di Pavia

Nicola Bizzaro
Università degli Studi di Pavia

Alessandro Bratus
Università degli Studi di Pavia

Alessandro Cecchi
Università degli Studi di Siena

Maurizio Corbella
Università degli Studi di Milano

Matteo Giuggioli
Università degli Studi di Pavia

Stefano Lombardi Vallauri
IULM Milano

Marida Rizzuti
IULM Milano

Federica Rovelli
Università degli Studi di Cagliari

Contact
www.worldsofaudiovision.org
contact@worldsofaudiovision.org

ABSTRACT

Il paper presenta le attività del gruppo di ricerca *Worlds of AudioVision*, costituitosi nel 2007 presso il Dipartimento di Musicologia dell'Università di Pavia (sede di Cremona). Componenti del gruppo sono giovani studiosi afferenti a diverse Università coordinati da Gianmario Borio ed Elena Mosconi, docenti di Teoria e Storia della musica del XX secolo presso l'Università di Pavia e di Storia del cinema presso le Università di Pavia e Cattolica di Milano. Oggetto d'indagine del gruppo di ricerca sono le tematiche audiovisive, affrontate con un taglio interdisciplinare, che coinvolge musicologia, filmologia e mediologia. Particolarità di *Worlds of AudioVision* è la spiccata dialettica interna che consente di affrontare un argomento così multiforme e in continua evoluzione, dando origine a un network di idee in formazione, confronto e aggiornamento costanti, il cui punto di arrivo sono i lavori pubblicati sul sito www.worldsofaudiovision.org.

Il paper è strutturato in due sezioni: nella prima si discutono le metodologie e i campi d'indagine affrontati dal gruppo, nella seconda sono discusse alcune fra le ricerche svolte in seno al gruppo.

1. PRESENTAZIONE DEL GRUPPO

Il gruppo di ricerca *Worlds of AudioVision* si è costituito nel 2007 presso il Dipartimento di Musicologia dell'Università di Pavia (sede di Cremona), proseguendo un percorso iniziato con il seminario dottorale *La comunicazione audiovisiva: immagine e suono nel cinema e nei testi multimediali*, coordinato da Gianmario Borio ed Elena

Copyright: © 2010 Borio et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](http://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Mosconi. Durante questa iniziativa, i cui risultati sono stati pubblicati sulla rivista *Philomusica on-line*¹, è emersa da più parti la necessità di affrontare tali fenomeni in maniera interdisciplinare, rispettandone la testualità specificamente legata ai *media* elettronici, in cui suono e immagine, pur se separati sul piano tecnico nelle arti pre-digitali, si organizzano reciprocamente in un insieme dotato di una struttura e di un significato peculiari. Da queste premesse si è originata la volontà di aggregarsi in un gruppo di studio in cui siano rappresentati differenti indirizzi disciplinari, in modo da assicurare un confronto costante sui metodi e sui risultati raggiunti, indirizzare le ricerche future e individuare i punti più problematici di ciascuno dei temi trattati dal maggior numero di punti di vista.

Attualmente *Worlds of AudioVision* comprende 13 studiosi, per la maggior parte dottorandi o dottori di ricerca afferenti alle Università di Pavia, Bologna, Siena, Cagliari e alle Università Statale, Cattolica e IULM di Milano. Ogni componente è impegnato su un diverso argomento, che diviene oggetto di discussione collettiva a seguito della presentazione al gruppo. La discussione e il confronto svolgono un ruolo fondamentale nella definizione dei risultati, che sono via via presentati nel sito Internet www.worldsofaudiovision.org.

Scopo del gruppo è affrontare prodotti multimediali di diverse arti, generi e forme della rappresentazione artistica, dal cinema alla videoarte, passando attraverso le installazioni, il *musical*, l'opera-video, il film-opera, la ripresa cinematografica di melodrammi e concerti e gli spettacoli intermediali e multimediali in generale. Si è scelto un approccio interdisciplinare, che consente di impiegare gli strumenti sviluppati in varie discipline, quali la musicologia, la filmologia e la mediologia. I metodi di ricerca, dunque, sono costantemente confrontati attraverso la loro applicazione a casi specifici. Scelta qualificante

¹*Philomusica on-line*, Vol. 6 N. 3 (2007), reperibile all'indirizzo Internet: <http://riviste.paviauniversitypress.it/index.php/phi/issue/view/06-03>

è proprio quella di affrontare una panoramica di esempi poco omogenea, procedendo da una serie di presupposti condivisi, che permette di far emergere l'esistenza di processi comuni nell'organizzazione degli elementi sonoro e visivo in contesti molto diversi, in senso tanto tecnico-formale, quanto semantico e comunicativo.

Il gruppo di studio si caratterizza per il proposito di accompagnare lo studio delle diverse forme di interazione tra le componenti sonora e visiva con la consapevolezza del ruolo essenziale che nei testi audiovisivi svolge la tecnologia, focalizzando l'attenzione sulla funzione specifica del suono e della musica. Punto di partenza delle indagini sono, quindi, l'evoluzione tecnica e tecnologica, in particolar modo quelle legate all'audio, e il loro rapporto con le tecniche compositive e le strategie audiovisive adoperate. Si prendono in considerazione inoltre le ricadute della tecnologia nei diversi ambiti: dalla nascita di figure professionali apposite, allo sviluppo di tecniche specifiche, coinvolgendo anche la genesi di nuove forme di fruizione.

2. PRESENTAZIONE DELLE ATTIVITÀ

Nelle attività del gruppo si possono distinguere due principali aree di interesse, corrispondenti agli indirizzi di ricerca finora emersi:

1. l'elaborazione di metodi per l'analisi di oggetti d'arte e di progetti performativi nei quali le dimensioni di immagine, parola e suono sono strettamente intrecciati;

2. l'indagine sulla terminologia in uso per descrivere e interpretare le opere audiovisive.

Il primo di questi punti rappresenta un nodo comune a tutte le ricerche portate avanti in ogni singolo ambito. A questo tema si è consapevolmente riservato un approccio pragmatico, centrato sull'analisi di opere provenienti sia dal panorama del cinema *mainstream* e indipendente, sia dalla sperimentazione multimediale per cui il grande schermo non è il referente obbligato (ad esempio le installazioni di Bill Viola, oppure collaborazioni tra compositori di musica d'arte e cineasti, come la realizzazione televisiva di *Die Schachtel* di Franco Evangelisti nella regia di Gregory Markopoulos, analizzata da Gianmario Borio).

In tale ambito si inserisce la ricerca condotta da Stefano Lombardi Vallauri, che si concentra su uno studio comparativo della *temporalità* in musica e nel "video astratto". La ricerca è condotta su due fronti: uno analitico-strutturale e l'altro estetico-filosofico, e tratta in particolare della "video-opera" di Fausto Romitelli e Paolo Pachini *An Index of Metals* (2003). L'opera è analizzata sezione per sezione e nel suo insieme, al fine di evidenziarne le caratteristiche salienti, che ne fanno un campione rappresentativo di alcune tendenze della musica e della videoarte contemporanee, nonché un ambito di studio privilegiato per la comparazione tra le espressioni nei *media* diversi del suono e dell'immagine. Lo studio della musica nell'intreccio multimediale è utile per una definizione del campo musicale all'interno del generale campo estetico, poiché il confronto con altri *media* può servire a indivi-

duare ciò che nella musica è più specifico e irriducibile. Per questo le "arti temporali" astratte, e all'interno di esse il video astratto, sono trattate come il genere artistico non-musicale più simile alla musica, rispetto al quale la differenza specifica si definisce precisamente la musica stessa.

Sul piano musicale *An Index of Metals* è qualificata da due tratti particolari, tra quelli che possono appartenere in generale al repertorio post-tonale: "centricità" e "vettorialità". Occorre domandarsi se nel video astratto si diano modi di scorrimento temporale analoghi a quelli musicali. Pure sul piano video in *An Index of Metals* si riscontrano in effetti costantemente direzionalità formali, quindi parimenti temporali. Esse sono principalmente di due tipi: aggregazione e disgregazione. Anche nel video astratto è possibile istituire tensioni differenziali tra forme e colori nello spazio affini e isomorfe alle tensioni sussistenti in musica lungo le varie dimensioni dello spazio sonoro (dalle più elementari: dinamica, timbrica, diastematica, alle più complesse: armonica, testurale, ecc). In *An Index of Metals* tuttavia le durate temporali coinvolte in processi formali vettoriali sono più brevi rispetto a quelle musicali, a favore di una maggiore frammentarietà e di quella che Christine Ross ha definito "extended temporality".

Oltre all'analisi di singoli film e opere multimediali una parte delle ricerche si è orientata su problematiche più generali, che hanno a che fare con ampie formazioni culturali come i generi cinematografici, oppure con la rappresentazione di determinati soggetti, di identità in senso nazionale o locale. Alessandro Bratus, ad esempio, lavora sull'immagine degli artisti della *popular music* anglo-americana trasmessa nel cinema, prendendo in esame le specifiche strategie audiovisive che coinvolgono il momento della *performance* dal vivo e la sua rappresentazione. Tale momento è un oggetto di particolare interesse perché mostra in che modo possa essere rovesciata la gerarchia che, nel linguaggio cinematografico più diffuso, vuole il suono subordinato alle esigenze dell'immagine, assumendo un ruolo organizzativo di primo piano.

A un simile filone di ricerca si è avvicinato anche Matteo Giuggioli, che si è occupato del rapporto tra musica e generi cinematografici (in particolare il *noir* e la commedia). A partire dalla considerazione che la dimensione sonora penetra ed è in grado di influenzare la costruzione del testo filmico sul piano semantico, sintattico e pragmatico, Giuggioli si propone quindi di indagare l'apporto della musica alla determinazione del genere cinematografico inteso come "negoziato" e istituzione culturale. Questo tipo di approccio consente di non trascurare l'aspetto dinamico del genere e al contempo di far trasparire il suo carattere evanescente di dimensione presente nelle aspettative, ma aperta alla ridefinizione ogni volta che si attui in un testo. Come chiarisce Rick Altman, teoria e storia dovranno necessariamente incontrarsi in tale prospettiva, pena il rinvenimento di categorie arbitrarie, seppur affascinanti e apparentemente inattaccabili.

Il secondo indirizzo di ricerca include a sua volta due direzioni d'indagine: da un lato l'interrogazione della terminologia che non ha ancora raggiunto una stabilità se-

mantica e dall'altro lato la revisione dei concetti fondamentali.

La prima direzione d'indagine ha come obiettivo la sistemazione di una terminologia entrata nell'uso, tanto comune quanto accademico, sulla base delle pratiche di costruzione del suono video-cinematografico che negli ultimi trent'anni hanno individuato con autocoscienza critica sempre maggiore nuove aree d'azione e di competenza. Il concetto di *sound design* rientra in questa categoria. Maurizio Corbella ne ha tentato un primo inquadramento terminologico, tecnico e storico-culturale. Se da una parte il termine nasce nel corso degli anni Settanta nell'ambiente della *New Hollywood*, legato individualmente all'attività di Walter Murch e di Ben Burt, con significati peraltro differenziati nei due casi, dall'altra parte è assurto a emblema della svolta stereofonica del cinema a partire dal decennio successivo. Il termine è riferito a un ampio spettro semantico, che oscilla dall'implicita assunzione di un'idea di progettualità autoriale sull'intera componente sonora del testo audiovisivo, alla confezione "artigianale" degli effetti speciali sonori. A partire dagli anni Trenta nel cinema di Hollywood si afferma un principio di organizzazione del suono che stabilisce confini abbastanza netti tra parola, rumore e musica, sulla base delle funzioni narrative che essi ricoprono. Nel corso degli anni Sessanta, in seguito al riconoscimento delle potenzialità creative insite nelle pratiche di manipolazione e sintesi elettroacustica del suono, tale principio organizzativo va incontro a una fase di assestamento e ridefinizione. Il suono viene così ripensato secondo un paradigma psico-percettivo di stampo comportamentista, che sfrutta consapevolmente la dimensione evocativa del timbro, della dinamica e della spazialità, oltre alle risorse di montaggio orizzontale e verticale delle tecnologie elettroacustiche che, con l'avvento del digitale, diventano pressoché illimitate.

Anche il concetto di immersione affrontato da Giacomo Albert rientra nel medesimo ambito: si tratta di una delle più importanti categorie della fruizione, studiata soprattutto in ambito mediologico, ma adoperata anche nel linguaggio quotidiano. L'interesse connesso con questa categoria risiede nella sua capacità di mettere in risalto la ricaduta dell'evoluzione delle tecnologie sulla creazione dell'immaginario. L'immersione del fruitore nell'opera comporta il passaggio dal paradigma della conoscenza di tipo cognitivista, basato sulla distanza critica tra fruitore e opera, a quello basato sulla costituzione della presenza e sull'appartenenza del fruitore e dell'opera al medesimo mondo, al medesimo spazio, in cui rappresentazione e realtà si confondono, per cui l'orizzonte semantico all'interno del quale si interpreta l'opera non è distinto dall'esperienza sensoriale e non mediata del fruitore. Tale cambiamento di paradigma rimette in discussione le categorie basilari dell'audiovisione, come la multimedialità, a favore della multimodalità. L'audiovisione stessa viene così intesa non più quale somma di due *media*/sensi, ognuno dotato di specifiche forme simboliche, bensì come convergenza di tutti i sensi in un'esperienza che coinvolge totalmente il fruitore. Attraverso l'analisi di esempi selezionati fra diversi generi artistici si enucleano più forme dell'immersione, in cui la relazione non mediata tra opera e fruitore si concretizza con finalità differenti e crea in tal modo esperienze eterogenee. Lo studio si concentra sul

rapporto tra tali forme dell'immersione e le strategie messe in atto per la loro reificazione, attraverso l'osservazione delle tecniche compositive e più in generale audiovisive adoperate nei singoli casi.

La seconda direzione d'indagine si avvale con profitto della metodologia ormai consolidata della storia dei concetti (*Begriffsgeschichte*). Questa viene coniugata a partire dalle specificità del gruppo di studio, che conducono a una doppia esigenza: contribuire alla formulazione di una teoria dell'audiovisione coerente e consapevole dei suoi presupposti filosofici e vagliare gli ambiti di applicabilità analitica dei concetti esaminati. Rientra in questo filone di ricerca il confronto, intrapreso da Alessandro Cecchi, con l'opposizione concettuale diegetico/extradiegetico. La distinzione risulta cruciale per la definizione del rapporto tra narrazione audiovisiva e orizzonte finzionale del cinema; tale centralità colloca il concetto di diegesi tra i concetti fondamentali (*Grundbegriffe*) della disciplina. La revisione critica del concetto, sollecitata dagli interrogativi sorti dalle applicazioni analitiche di tale opposizione – soprattutto quando si tratta di precisare la posizione del suono o della musica rispetto alla diegesi – suggerisce i *desiderata* di una teoria dell'audiovisione la cui formulazione è ancora allo stato embrionale: un'impostazione coerentemente costruttivista consentirebbe di affinare gli strumenti di analisi in modo da cogliere le complesse articolazioni delle strutture audiovisive. Ciò induce a riconoscere che la narrazione è funzione della costruzione audiovisiva e non viceversa, che quindi la diegesi è una costruzione analitica secondaria e soggettiva, il cui ambito di applicabilità deve essere drasticamente ridimensionato.

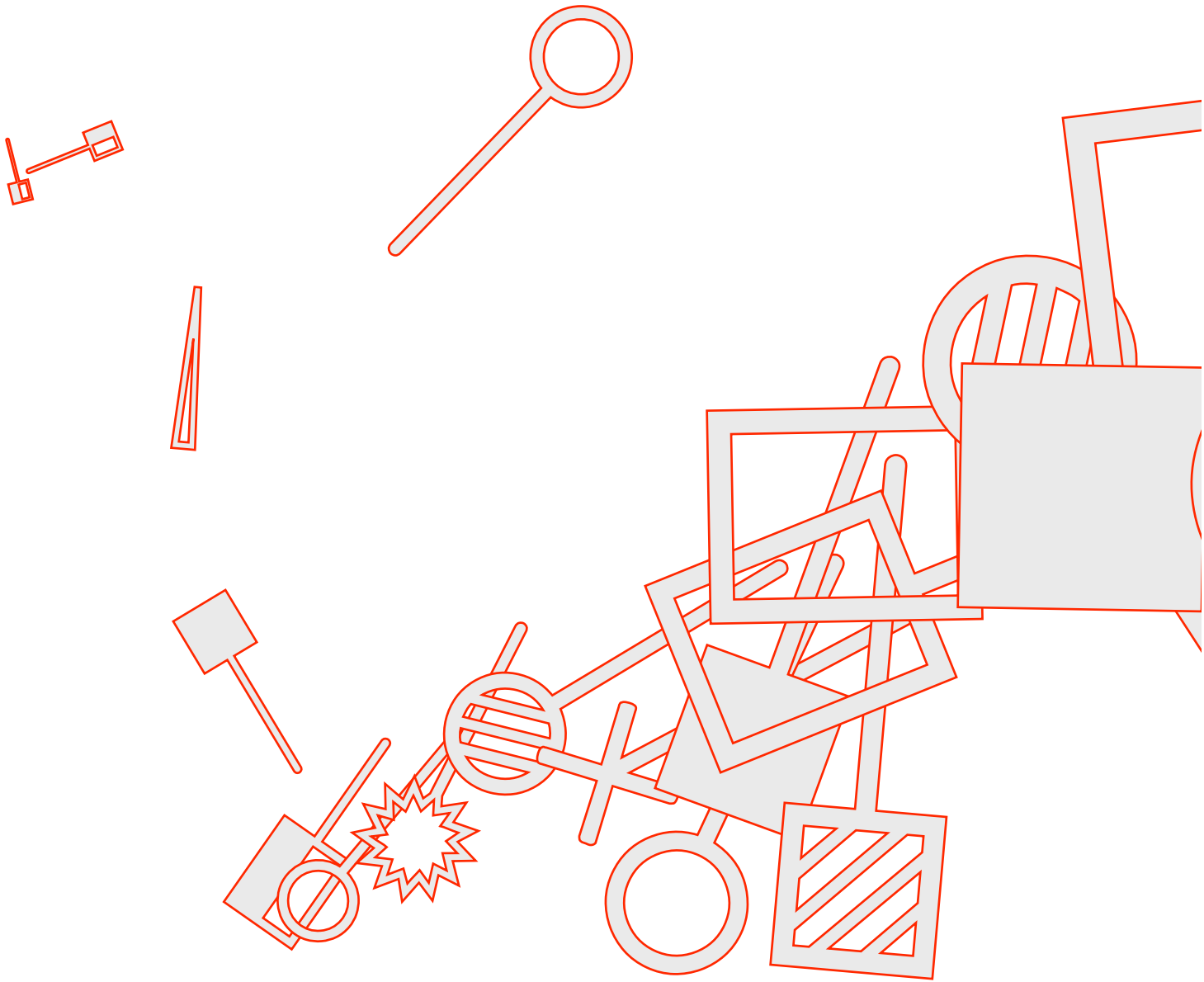
Ad esempio, nella scena più celebre di *Psycho* (1960) di Alfred Hitchcock, "The Murder", lo sguardo analitico dovrebbe essere rivolto non al contenuto raffigurato, bensì in primo luogo alla costruzione della sequenza audiovisiva intesa come narrazione filmica, in cui rientra a pieno titolo la musica di Bernard Herrmann. Se priviamo la sequenza del sonoro, le immagini rivelano tutta la loro insufficienza diegetica rispetto allo *choc* audiovisivo della sequenza non schermata: il *mi bemolle* stridulo dei violini – che nell'arco di otto battute, con il concorso di tutti gli archi, si organizza in un *cluster* sui successivi tre gradi della scala cromatica dislocato su quattro ottave – risulta sincronizzato perfettamente con l'apertura della tenda per mano del killer; e proprio dal suo stridulo rumore metallico, udito appena prima, quando la vittima aveva chiuso la tenda, trae lo spunto timbrico. I *clusters* delle successive otto battute, resi ancora più dissonanti da rapidi *glissando* ascendenti in tutte le parti, risultano non solo narrativamente, ma anche diegeticamente più convincenti dell'inquadratura delle coltellate, la cui successione è astratta, depurata dal potenziale lesivo dei colpi. La musica, per quanto sia chiara la sua posizione rispetto alla diegesi, non esaurisce in questo la sua funzione; al contrario determina l'audiovisione proprio nella sua capacità raffigurativa, cui naturalmente collaborano in egual misura le urla della donna, il rumore dell'acqua che scorre, lo strappo metallico della tenda e la caduta del corpo agonizzante. Il senso stesso dei suoni diegetici viene riconfigurato nell'interazione con la musica extradiegetica. Risulta dunque evidente che la distinzione diegetico/extradiegeti-

co, dal punto di vista narrativo, ha una funzione del tutto marginale nella sequenza in questione. Noi facciamo esperienza di tale soglia nella stretta misura in cui ne constatiamo la secondarietà dal punto di vista dell'esperienza estetica del film, che è esperienza dell'audiovisione.

Nel medesimo ambito, ma con lo scopo di effettuare una ricostruzione storica piuttosto che di aprire nuovi fronti di dibattito teorico, si situa lo studio di Nicola Bizzaro, che ha approfondito la teoria dell'audiovisione di Pierre Schaeffer; questi è principalmente noto quale padre della musica concreta e teorico dell'oggetto sonoro, ma in realtà è stato anche a lungo attivo nel campo della teoria della comunicazione e della produzione audiovisiva. Grazie alle posizioni strategiche da lui ricoperte nei contesti dell'emittente radiofonica e radiotelevisiva francese, fu in grado di promuovere e dirigere una serie di iniziative variamente mirate alla produzione di documenti audiovisivi di marcata impronta sperimentale, sia per garantire materiali pubblicabili per l'ente patrocinatore, sia per incentivare percorsi di ricerca che, partendo da alcune idee fondamentali sviluppate dallo stesso Schaeffer in ambito radiofonico e musicale, potessero inscrivere il fenomeno audiovisivo in una solida e articolata impalcatura teorica. L'attività più che trentennale di Schaeffer e del suo gruppo in questo ambito è a oggi quasi completamente ignorata e merita di essere presa in considerazione, se non altro in virtù della mole di contributi originali prodotti (anche se non sempre facilmente reperibili) e per la centralità del teorico all'interno del panorama artistico e culturale francese ed europeo. I filoni d'indagine che qui si aprono richiedono ovviamente un lavoro ad ampio raggio, di cui l'attuale ricerca non può che rappresentare un sintetico preludio. Ciononostante, sono già a disposizione sufficienti elementi che hanno permesso di individuare alcune problematiche di fondo e avanzare ipotesi utili per la ricostruzione della teoria schaefferiana dell'audiovisione.

Quelli citati sono solo alcuni degli studi portati avanti in seno al gruppo di ricerca; altri lavori, precedentemente non menzionati, si rifanno solo in parte ai due indirizzi di ricerca. Ad esempio Marco Monzio Compagnoni analizza la sequenza finale di *Prenom Carmen* di Godard e mette in evidenza il modo in cui il Quartetto op. 135 di Beethoven si integra nella costruzione audiovisiva; Federica Rovelli mette a confronto le tre edizioni di *Touch of Evil* (1958, 1976, 1998) indagando sul complesso status testuale dell'opera cinematografica e sugli effetti che su di esso esercitano le fasi di negoziazione intercorse tra sistema produttivo e volontà registica. Elena Mosconi si occupa della costruzione dell'identità italo-americana nei film incentrati sul personaggio del "tenore italiano", prendendo come spunto il caso di Mario Lanza. Marida Rizzuti si occupa attraverso l'esempio di *Carmen Story* dei concetti di adattamento, trasposizione e traducibilità, e infine Gaia Varon indaga il ruolo delle scelte registiche nella resa audiovisiva delle performance musicali.

2. Gesto, interfaccia e controllo



+ Stefano Fumagalli, Saverio Monti
+ Carlo Massarelli, Andrea Valle
+ Pietro Polotti, Maurizio Goia
+ Stefano Delle Monache, Davide Rocchesso

EJP – Electro Jamming Project

Stefano Fumagalli

Conservatorio di musica “G.Verdi” – Como
info@stefanofumagalli.it

Saverio Monti

Conservatorio di musica “G.Verdi” – Como
monti.saverio@gmail.com

ABSTRACT

Il progetto EJP - *Electro Jammin' Project* - nasce dalla ricerca e dalla sperimentazione di nuove soluzioni hardware e software come ausilio strategico alla composizione, improvvisazione ed esecuzione della musica elettronica attraverso la tecnologia multitouch. Il sistema realizzato - hardware e software - è in grado di assistere compositori, performer e esecutori attraverso un'intuitiva interfaccia multi-tattile. In ambito hardware l'interfaccia è stata sviluppata con tecnologia DSI - *Diffuse Surface Illumination* - seguendo le indicazioni redatte dal NUI Group all'interno del libro *Multi-Touch Technologies*[2]. In ambito software l'interfaccia grafica è stata sviluppata in ambiente VVVV¹ mentre per la manipolazione generazione e sintesi degli eventi sonori ci si è avvalso del software di programmazione Max/Msp². Il risultato che ne consegue è un sistema intuitivo e di facile utilizzo che può “aprire nuove vie” performative, compositive ma anche di carattere pedagogico. Per ora il sistema EJP è ancora un prototipo, ma è già stato testato in pubblico in performance dal vivo presso il Conservatorio G. Verdi di Como e al Festival Cinque Giornate presso il Palazzo Reale di Milano dimostrando molta affidabilità e concretezza nel suo utilizzo.

1.INTRODUZIONE

1.

L'interaction design degli ultimi anni mira all'eliminazione di qualsiasi artificialità e barriera non necessaria tra utente e contenuto dell'interazione, realizzando sistemi d'interazione sempre più intuitivi e naturali. Il valore dei contenuti digitali virtuali ha ormai un valore equivalente, se non superiore (considerata la facilità di distribuzione), ad eventi e contenuti reali. Si ricercano quindi sistemi che permettano di accedere ai contenuti virtuali in maniera sempre più veloce e diretta. Questa ricerca ha portato alla creazione di nuove interfacce che superano la tipica combinazione di mouse e tastiera. L'esempio più lampante di questa tendenza è costituito dalla forte espansione dell'utilizzo di dispositivi multi-tattili, i quali consentono di operare direttamente a contatto con il feedback grafico dell'applicazione. Iphone, Ipad

¹ vvvv: a multi purpose kit, <http://vvvv.org>

² Cycling 74 Tools for media, <http://cycling74.com/>

touch, smartphones in genere, Ipad e i Tablet PC sono prodotti che testimoniano il grande successo delle tecnologie Multi-Touch. Anche nel campo musicale e artistico sono state create numerose opere con interfaccia multi-tattili. Queste applicazioni, però, sono limitate ad un'unica funzione o ad uno scopo specifico: un'installazione, una performance, un mixer virtuale...

L'obiettivo del nostro progetto è la realizzazione di un sistema multi-tattile (*Electro Jammin' Project*, da ora EJP) in grado di gestire sia l'organizzazione temporale e collaborativa di eventi multimediali che l'elaborazione in tempo reale di eventi interattivi. Questa ricerca ha portato gli autori ad occuparsi, nello specifico, dello sviluppo di due aspetti: creazione di un software per la gestione dei contenuti e sviluppo di un hardware stabile e funzionale. Nell'articolo seguente saranno presentate le soluzioni software e hardware adottate per la realizzazione del primo prototipo di EJP.

2.EJP: SOFTWARE

EJP è un'interfaccia multi-tattile per la gestione di un insieme di eventi multimediali. Con il semplice utilizzo delle dita, consente di posizionare lungo la superficie di una *timeline* scorrevole dei segni grafici abbinabili a eventi multimediali quali: suoni campionati, suoni di sintesi, elaborazioni audio/video in real-time, eventi video e quant'altro. Come è possibile tutto questo? EJP è costituito da diversi moduli, i principali sono: il modulo per la gestione grafica della *timeline* e il modulo per l'analisi e la comunicazione degli eventi presenti su questa linea. Il primo permette l'interazione con l'interfaccia grafica multi-tattile, il secondo, invece, è un server OSC³ in grado di analizzare i dati della linea del tempo (come la posizione e la durata degli eventi) e comunicare con client OSC residenti su software esterni. In questo modo è possibile interfacciare EJP con svariati software per la gestione di media in real-time: Max/Msp, Jitter, Pure Data, Ableton Live, Resolume, etc... Il feedback grafico utilizza la metafora dello scorrimento di un nastro, come succede per gli ambienti multitraccia audio. L'aspetto caratterizzante del progetto è quello di unire in un'unica interfaccia il concetto di organizzazione temporale asincrono (tipico dell'approccio della composizione musicale) con

³ Open Sound Control, <http://opensoundcontrol.org/>

quello in *real-time* (tipico dell'approccio improvvisativo). L'utilizzo di EJP può essere di supporto alla composizione di musica elettronica, facilitare attività di gruppo nella pianificazione e esecuzione di performance (musicali, teatrali, coreutiche), servire da partitura per l'esecuzione di performance, essere utilizzato per il sound design di video, oppure trasformato in un'installazione audiovisuale interattiva.

2.1 Come funziona EJP?

EJP si avvale di una superficie multi-tattile costruita con tecnologia DSI (vedere paragrafo 3). Funziona con il contatto delle dita e non è previsto il *tracking* di oggetti e *fiducials*, come nel caso della Reactable [3]. L'architettura di EJP è costituita da molteplici moduli residenti su diverse piattaforme software. I dati in input sono costituiti dall'immagine di una videocamera a infrarossi. Queste immagini vengono analizzate dal software *Community Core Vision*⁴ (CVV), piattaforma open-source in ambito della Computer Vision, il quale estrapola informazioni riguardanti la posizione delle dita sulla superficie tattile. I dati sono passati a vvvv tramite il protocollo TUIO⁵, un protocollo aperto usato in ambito delle applicazioni multi-tattili. vvvv lavora su due livelli: una parte del programma si occupa del feedback grafico, l'altra della trasmissione dei dati dell'interfaccia grafica via OSC. Il feedback grafico viene visualizzato sulla superficie multi-tattile mentre i dati OSC raggiungono dei client residenti su altri software applicativi che utilizzeranno i dati per la produzione di media. Ora descriviamo meglio questi passaggi.

Gestione dell'interfaccia grafica

I dati in uscita da CCV vengono letti da vvvv grazie al modulo TUIO Decoder in grado di ricevere i pacchetti che arrivano via UDP e di suddividerli in liste di dati utilizzabili. In questo modo otteniamo uno *spread* (in vvvv corrisponde a una lista di dati) per ogni parametro in ingresso. L'interfaccia grafica è composta da elementi statici che chiameremo interfaccia di controllo (figura 1) e da elementi dinamici, in grado di scorrere, che costituiscono la timeline vera e propria. Vediamo ora quali elementi costituiscono l'interfaccia di controllo. Questi elementi consentono azioni interattive di base e universalmente utilizzabili in ogni applicazione client.

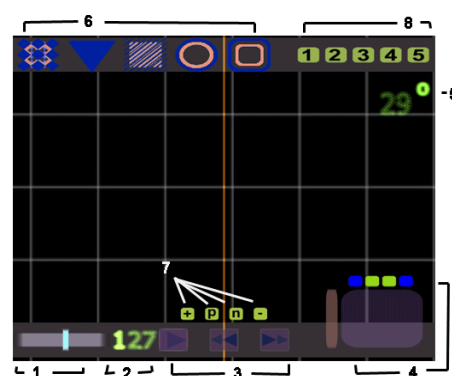


Figura 1: Interfaccia grafica di controllo.

Successivamente, vengono descritte le varie zone dell'interfaccia in base alla numerazione indicata in figura:

1. Barra del tempo: questo *slider* permette di determinare la durata totale della timeline, nel prototipo va da un minimo di 10 secondi e un massimo di 3 minuti. Basta far scorrere un dito su questo oggetto per modificarlo;
2. Indicatore tempo totale: qui viene indicata la durata totale della *timeline*. Nel caso in figura la durata è di 1 minuto e 12 secondi;
3. Pulsanti di navigazione: questi pulsanti permettono di accedere alle funzioni di play/pausa e scorrimento veloce in avanti e in dietro. Premendo il tasto play la *timeline* inizia a scorrere al centro della proiezione, il tasto "pausa" la ferma e i tasti di scorrimento veloce permettono di raggiungere velocemente parti lontane della *timeline*;
4. Controlli "Live mode": questa sezione permette di modificare in tempo reale dei parametri di elaborazione, rimandiamo un approfondimento alla sezione dedicata;
5. Tempo attuale: questi numeri indicano il tempo corrente della *timeline*;
6. Palette di oggetti: in questa sezione sono visualizzate delle icone con forme diverse. Muovendo un dito sopra di esse si può trascinare all'interno della *timeline* un elemento con forma corrispondente all'icona selezionata. Sarà sempre possibile riposizionare questi elementi in un'altra posizione. Il posizionamento sull'asse orizzontale sottende un posizionamento sull'asse temporale: elementi a sinistra della *timeline* saranno eseguiti prima di elementi posizionati a destra.
7. Markers: questi quattro pulsanti permettono di inserire dei marker temporali e di memorizzare, quindi, delle posizioni temporali precise. Questa funzione è utile per organizzare e scandire temporalmente la performance e rende possibile il passaggio veloce da un marker all'altro. Con il tasto "+" si aggiunge un marker nella posizione attuale dalla barra di scorrimento. Il marker viene visualizzato con una linea verticale verde. Con il pulsante "-" si eliminano i marker, l'ordine di eliminazione è inverso a quello di creazione. Il

⁴CVV, <http://ccv.nuigroup.com/>

⁵ TUIO, <http://www.tuio.org/>

tasto “n” consente di saltare al marker successivo. Il tasto “p” consente di tornare al marker precedente.

8. Preset: questi numeri servono per la gestione di preset sonori. Si può selezionare un numero per volta. Via OSC viene mandato un numero corrispondente al preset scelto, questo numero può essere utilizzato per modificare stati del programma client.

La parte centrale della proiezione è costituita dalla *timeline* vera e propria. Essa è in grado di scorrere e di visualizzare man mano gli elementi posizionati precedentemente sulla linea del tempo. Non è possibile posizionare elementi durante la fase di scorrimento. Quando la *timeline* scorre, una barra verticale (barra di scorrimento) arancione indica la posizione corrente, mentre una griglia di sfondo ci aiuta ad orientarci nella successione temporale degli eventi. Nel momento in cui la barra di scorrimento incontra una figura posizionata sulla *timeline* viene trasmesso, all'applicazione client, un pacchetto OSC contenente i dati relativi all'evento (categoria dell'evento, posizione xy). Quando la timeline non scorre è possibile eseguire delle operazioni interattive avanzate:

1. *Stretching* (allargamento) delle figure: utilizzando due dita è possibile modificare le dimensioni degli elementi sulla *timeline*. Ovviamente, il risultato di questa operazione dipenderà dall'applicazione client utilizzata. Ad esempio, se modifico un oggetto collegato ad un generatore di bande di rumore bianco, la lunghezza della mia figura corrisponderà alla durata del suono, l'altezza, invece, alla larghezza di banda. Attualmente è funzionante lo stretching orizzontale, mentre quello verticale è in fase di implementazione.
2. Attivare “Live Mode”: è possibile attivare una modalità di elaborazione in tempo reale. I controlli “Live Mode” prevedono un pad quadrato (il quadrato in basso a sinistra in figura 1) per il rilevamento xy della posizione e la possibilità di cambiare 4 preset, schiacciando i 4 quadratini superiori. Per determinare quali eventi sono elaborati nella modalità live è necessario premere il rettangolo a sinistra del pad di rilevamento xy. Una volta premuto, il rettangolo diventa di colore rosso indicando che è possibile scegliere quali eventi dovranno essere elaborati. Ora se si mette il dito su un evento della timeline quest'ultimo diventa bianco. Questo indica che, al momento della sua esecuzione, quell'evento sarà processato dall'algoritmo d'elaborazione live selezionato. Per deselectare un evento dalla modalità live è sufficiente riappoggiare un dito sull'evento ed esso riassumerà il suo colore originale. Al termine dell'operazione è consigliato premere di nuovo il rettangolo a destra del pad xy in modo tale da bloccare la modalità di scelta degli eventi “live mode”

I dati di esecuzione provenienti dalla *timeline* sono trasmessi via OSC a applicazioni audio/video esterne. Si verifica la trasmissione di tre categorie di messaggi: i messaggi inerenti all'interfaccia di controllo (play, pause, stop, position), i messaggi relativi agli eventi sulla *timeline*, i messaggi inerenti al “live mode”.

La ricerca sul lato software si è svolta in parallelo con l'implementazione e il testing di hardware sempre più stabili. In una prima fase è stata utilizzata un'ampia interfaccia a tavolo con tecnologia FTIR [1], simile a ReacTable [3]. Questa soluzione permetteva una notevole dimensione della superficie interattiva ma presentava notevoli complicazioni in fase di allestimento: calibrazione del sistema di proiezione, montaggio e smontaggio del tavolo. La sperimentazione ci ha quindi portato a prediligere un'interfaccia di dimensioni più ridotte e costruita con tecnologia DSI. Nel prossimo paragrafo descriveremo in dettaglio le differenze tra questi due approcci.

3. SOLUZIONI HARDWARE

Come precedentemente esposto, EJP si avvale di una superficie *multi-touch* costruita con tecnologia DSI -*Diffused Surface Illumination*. La tecnologia DSI è una variante della tecnologia FTIR- *Frustrated Total Internal Reflection*-, tecnica che si avvale di una metodologia di *Opti-cal multi-touch* sviluppata da Jeff Han[1]. La tecnica FTIR[2] consiste nell'inondare un pezzo di acrilico con luce infrarossa che per sua intrinseca costituzione intrappola al suo interno i raggi IR -*infrared* - e genera così l'effetto di *Total Internal Reflection*. Quando si crea un contatto con la superficie, i raggi infrarossi vengono “*Frustrated*” (“perturbati”) generando una riflessione non più totale in quel punto.

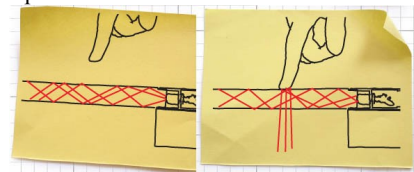


Figura 2 : Particolare dei raggi infrarossi Frustrated. complexidadeorganizada.files.wordpress.com

Questo avviene perché, una volta toccata la superficie Multi-touch, il fascio di luce *Frustrated* viene riflesso verso il basso in direzione della webcam. La camera è in grado di riconoscere solamente il fascio di luce infrarossa grazie all'utilizzo di un filtro IR-pass.

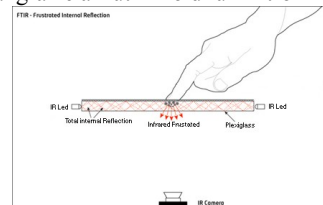


Figura 3: Elementi del sistema FTIR. wiki.nuigroup.com

Nella tecnologia FTIR le immagini riprodotte sulla superficie Multi-Touch vengono generate attraverso un videoproiettore posto al di sotto della superficie tattile, solitamente vicino alla IR webcam.

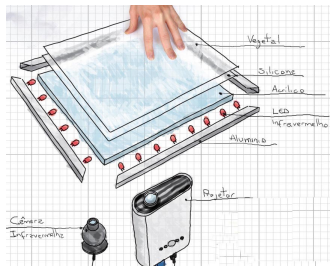


Figura 4 : Elementi del sistema FTIR .
complexidadeorganizada.files.wordpress.com

La tecnologia DSI[2] si differenzia dalla FTIR per due aspetti :

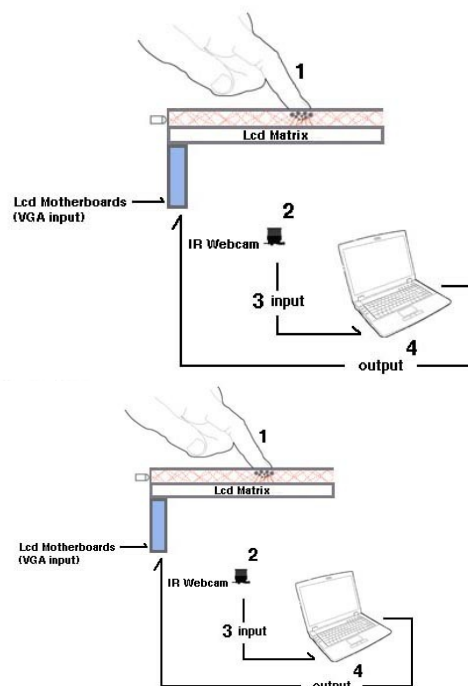
1. Tipologia di acrilico utilizzato ;
2. Assenza di videoproiettore .

La tecnica *Diffused Surface Illumination* utilizza un acrilico speciale denominato *EndLight* che è in grado di distribuire i raggi IR in modo uniforme su tutta la superficie. Questo acrilico utilizza piccole particelle che si trovano al suo interno come migliaia di piccoli specchi che reindirizzano i raggi a infrarossi diffondendoli in maniera omogenea per tutta la superficie della resina; per questo motivo tale tecnica viene detta *Diffused Surface Illumination*. L'assenza del videoproiettore è dovuta al fatto che la tecnologia DSI utilizza uno schermo LCD modificato, posizionato al di sotto dell'acrilico *EndLight* e genera lui stesso l'immagine. Lo schermo LCD viene smembrato dal suo contenitore per sfruttarne unicamente la sua circuiteria elettronica, la superficie composta di cristalli liquidi e la superficie riflettente.



Figura 5 : Elementi interni di uno schermo LCD . www.diniro.net

In questo modo la webcam è in grado di ricevere i raggi *Frustrated* generati dal tocco sull'acrilico posto al di sopra dello schermo LCD ed allo stesso tempo si ha la possibilità di inviare le immagini tramite cavo VGA allo schermo a cristalli liquidi e quindi si elimina la necessità di utilizzare un videoproiettore per il feedback grafico. In questo modo il sistema rimane molto compatto, facilmente trasportabile e configurabile in pochi minuti. Inoltre va sottolineato che la realizzazione di un sistema DSI comporta una spesa contenuta rispetto ad un sistema costruito con tecnologia FTIR.



1. Controllo fisico
2. Ripresa della zona di pressione
3. Invio dei dati al software CCV
4. Invio dei dati allo schermo LCD per un feedback visivo

Figura 6: Elementi del sistema DSI .

4.CONCLUSIONE

Nonostante il sistema sia ancora un prototipo è già stato testato in performance live e si è rilevato affidabile e di facile configurazione.

Si sono riscontrati alcuni problemi di eccessiva irradiazione dei raggi IR lungo il perimetro dell'acrilico *EndLight* generando una non precisa ripresa della zona di pressione. Ulteriori sviluppi potrebbero appunto concentrarsi nell'ottimizzazione dell'illuminazione della superficie tattile lungo il perimetro suo perimetro. Dal punto di vista software, invece, si devono ancora sviluppare nuove funzioni ed applicazioni. Alcuni esempi possono essere: l'implementazione di spazi tridimensionali rapportabili alla spazializzazione sonora, testare l'interfacciamento con sequencer commerciali, ed infine lavorare sull'ottimizzazione delle prestazioni.

Riferimenti Bibliografici

- [1] Han, J. Y. 2005. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th Annual ACM Symposium on User interface Software and Technology* (Seattle, WA, USA, October 23 - 26, 2005). UIST '05. ACM Press, New York, NY, 115-118.
- [2] NUI Group Authors. Multi-Touch Technologies .1 st edition [Community Release] : May 2009.
- [3] www.reactable.com
- [4] Fumagalli S. , "Electro jammin' project 2 -

timeline multitouch per la composizione e l'improvvisazione musicale", Tesi di laurea nel biennio di Musica Elettronica e Tecnologie del Suono

Copyright: ©2010 First author et al. This is an open-

access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

E-ZAMPOGNĚ

A SOUTHERN-ITALIAN BAGPIPE CONTROLLER

Carlo Massarelli
massarelli.carlo@gmail.com

Andrea Valle
CIRMA, Università di Torino
andrea.valle@unito.it

ABSTRACT

This paper describes *e-ZampognĚ*, a project of “virtualization” of a traditional musical instrument: the Southern-Italian bagpipe (the *zampogna*). The project starts from the desire to re-assess the musical potential of the *zampogna*. Apart from ethnomusicological interest, the digital implementation of the *zampogna* allows to extend its features, e.g. introducing chromatic fingering, pitch and tonality shifting, etc., thus providing a new tool for music application.

e-ZampognĚ makes use of low cost hardware and software technologies.

1. INTRODUCTION

The *e-ZampognĚ* projects aims at creating a simple and inexpensive prototype of a digital musical interface that simulates the behavior of a particular family of ancient instruments: the Southern-Italian bagpipes (*zampogna*, as in Italian language). The choice of this instrument was initially motivated by the long-time interest in the bagpipes of the first author.

The bagpipe is an instrument of popular origins, namely pastoral, and is denoted by different vernacular terms depending on its geographic diffusion. In Basilicata and Cilento (in the Southern Italy), for example, someone calls it using a very impressive term, “i suoni” (the sounds). This designation indicates not only the polyphonic capability, but also the instrument's harmonic richness, as well as the ability to produce strong, high-pitched sound [1]. By contrast, the bagpipe is suffering from severe limitations in terms of intonation. The *zampogna* underwent a long process of ghettoization that relegates it to few appearances during the traditional religious festivities, particularly Christmas. Such a strong characterization has negatively affected its musical uses and diffusion. One of the aims of the digital implementation of the *zampogna* is thus to extend its capabilities in order to increase its usage.

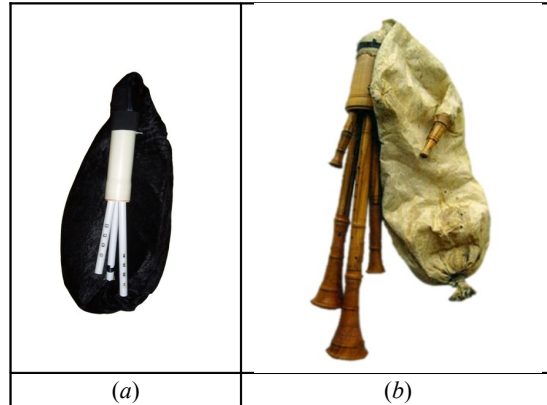


Figure 1. E-ZampognĚ digital MIDI controller (a) and a traditional *zampogna* “*zoppa*” (b).

As all the bagpipes, the Italian ones have specific limitations in terms of reduced pitch range, related to the size of the pipes. More, they require the user to provide a constant amount of energy in order to generate sound and to maintain tuning. Such amount is significantly higher than that required for other wind instruments, a peculiarity affecting the learning curve, generally involving many years. Moreover, as weather conditions relevantly affect the natural components (wood, cane) of the instruments, the *zampogna* requires a frequent processes of re-tuning: far from being an easy and well-defined process, retuning involves special techniques (more generally, bagpipe construction and tuning are kept hidden by master builders), so that the detuning of the instrument is often a cause of abandonment for beginners who have not the necessary knowledge and skills to maintain the instrument correctly tuned.

The electronic implementation of a *zampogna* (see the *e-ZampognĚ* example in figure 1a) introduces a “flexibility” factor (not present in real ones) that allows the selection by the user of one of n basic types of bagpipe, the adjustment of intonation and tuning, and the control of sound simulation. In this way, the project allows the automatic correction of the inaccuracies often present in real instruments and increases the user basis by providing easier access to the instrument and its maintenance. More, the digital instruments shows other features that can be exploited in musical applications.

In the next sections we briefly analyzes *zampogna*’s models, features and operating principles, both from an acoustic and performing perspective. We describes a simplified model of the instrument, designed on the basis

of the approximations of its operating principles. The last section describes suitable technologies (hardware) to realize the controller, and some examples of synthesis techniques suitable to obtain a realistic zampogna's sound.

2. WHAT ZAMPOGNA MEANS?

The term “zampogna” refers generally to all musical instruments whose sound is produced by the vibration of air, contained in a flexible tank placed between the player's mouth and one or more aerophones' reed. The air supply is contained in a bag (traditionally made of animal skin) that allows constant blowing even while the player is breathing. Also, the bag allows to feed more than a reed at the same time. From this basic type, many different models can be derived, as the result of the combination of already existing reed instruments with the bag. In this way, a variety of combinations have appeared (and disappeared) during the centuries [2].

The most relevant difference between common European bagpipes and the Southern-Italian ones' lies in the number of *chanter*s (the pipes that usually play melodies). While in European model the player usually makes use of only one chanter, in the Italian model s/he plays on two chanter, one for each hand. Moreover the zampogna has one or more drones (generally not chromatic) with fixed notes. Unlike the European one, all the pipes in the zampogna are planted in the same wood block connected to a breath-powered bag.

A very ancient instrument, the zampogna is popular, in its many varieties, in Central and Southern Italy (from northern Latium to Calabria) and Sicily. While it is possible identify at least 13 models of zampogna [2], in realation to construction and timbre characteristics, for our project we chose the “*A chiave* (with key)”, the “*Surdullina*” and the “*Zoppa/A paro*” models as they are considered the most representative.

2.1 “*A chiave*” from Lucania

This model of bagpipe, common in Lucania (Basilicata, Cilento and a part of Calabria) is identified by four unequal pipes: two chanter and two drones, all characterized by the presence of double reeds. This type is closely linked to bagpipes from Lazio and Molise. The four pipes are separated from each other and sound different, because of a variable, but always truncated, conical, internal profile (the conical factor depends on the length of the bore). The particular drilling technique, performed with a lathe and various steel reamers, introduces a common irregularity of the inner profile, not perfectly conical but cylindrical “in steps” (i.e. not always accurately rounded).

The right chanter is the little one, it has several tone-holes (usually nine) but only five are really used by the hand. The left chanter is the biggest one and, at the same time, it is the biggest pipe of the entire instrument. It has only four tone-holes and the last one is covered with a metal key. The chanter have tone-holes and conical bells to emphasize the sound with middle and high

frequencies and high volume. The two drones (differing in size) have no tone-holes and no true bells.

In this model of zampogna the left chanter plays five notes (from the 1st to the 5th of the only major scale available on the instrument), the right one plays the same plus one (at the upper octave), and the drones play only one: the dominant (5th) of the major scale. The pitch range and the fingerings are shown in Table 1.

“A chiave” C Major Zampogna model					
Left chanter					
G3	F3 (or F#3)	E3	D3	C3	
“A chiave” C Major Zampogna model					
Right chanter					
G4	F4 (or F#4)	E4	D4	C4	B3

Table 1. Traditional “*a chiave*” zampogna fingerings with relative notes (optional notes in brackets).

2.2 “*Surdullina*” from Pollino area.

Very different in timbre and size from the first model, this bagpipe has the same number of pipes: two chanter of the same size but different pitch range, and two drones, all characterized by the presence of single reeds, cylindrical inner profiles and no bells (actually, it is possible to find the bell in some models). While the right chanter plays five notes (from the 1st to the 5th of the major scale), the left one is tuned on the lower 4th and not in octave with the other chanter. As in the first model, the drones play the dominant note (5th). Pitch range with fingerings are shown in Table 2.

2.3 “*Zoppa*” and “*a paro*” zampogne.

It is possible to define these two typologies of zampogna -different from each other just in the profile factor and in the use of double (for the “*zoppa*”) and single (for the “*a paro*”) reeds- as a perfect mix between the “*a chiave*” model and “*surdullina*”. Featuring two chanter with bells, with all the pipes bearing a conical shape, and tuned by an interval of 4th in the scale, maybe these are the most typical and ancient zampogne. Pitch ranges and fingerings are shown in Table 3.

All the models of zampogna have a limited extension, because of the use of a double chanter. Regardless of the size of the instrument, scale, and in particular the interval relation among the notes, are unchanged. What changes, however, is the fundamental tone which is closely dependent on the size of the instrument.

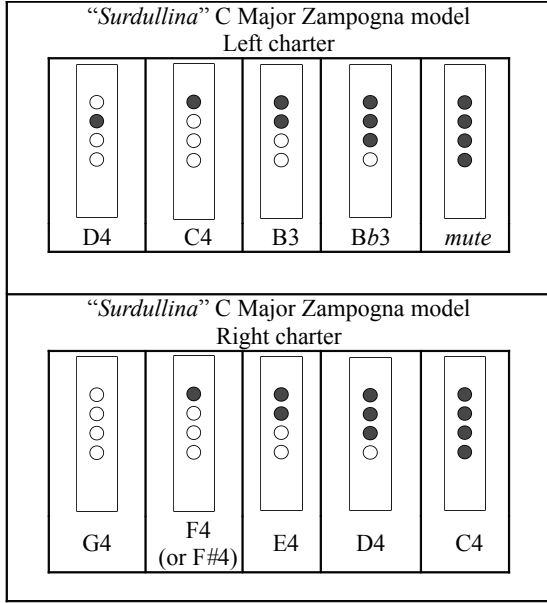


Table 2. Traditional “surdullina” zampogna fingering with relative notes (optional notes in brackets).

3. A SIMPLIFIED GENERIC CONTROLLER MODEL

In this section we discuss the controller we developed to provide zampogna players an instrument-like interface to sound synthesis.

3.1 The bag

We first studied the physical instruments in order to identify the relevant features that needs to be implemented in the controller as interfaces for the user. Concerning the use of the bag, it resulted obvious that changing the air pressure in the bag has no real relevance in performing with real instruments: it is not possible to obtain controlled modulation of sound like effects of loudness or vibrato and no real dynamics are available on this kind of instruments. This happens because of the different response that any reed inside of the instrument has. Moreover, according to the bagpipe tuning and playing technique traditions, the best players are the ones able to obtain a constant perfect tuning (usually with 0 hertz unison between notes of same pitches). Thus, eventual variations in pitch that could be induced

residually by air pressure are not favored by traditional practice.

This feature let us to introduce a first approximation in the model: if no multiple states bag control is needed, then we can implement a simple two-state control, that allows to turn on/off the performance. In our model, no air needs to be provided in input, and the bag is simply pressed, not breathed into. A piezoelectric vibration sensor is used to sense the bag filled with air, because it also reacts to the stress of incident air pressure and it has extremely lower cost than a gas pressure sensor.

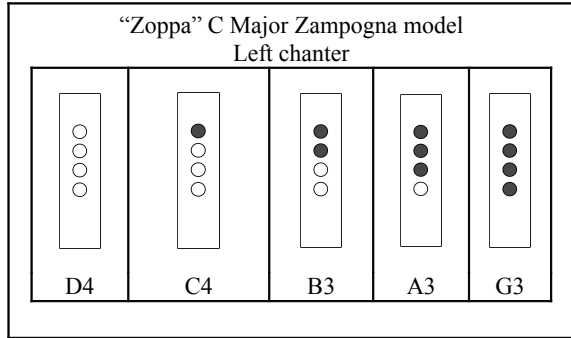
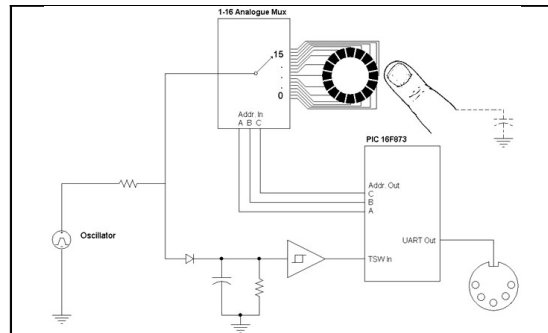


Table 3. Traditional “zoppa” zampogna left chanter fingering with relative notes. Right chanter follow the same example of the “a chiave” model (see table 1.).

3.2 The pipe

Regarding tone-holes effects, bagpipes, as well as oboe and clarinet, allow partial occlusions. Some suitable technologies that simulate realistic tone-holes behavior include sensing resistors (as in the Scavone example [3]) or multiple state capacitive sensors (as in the Cannon-Hughes-O Modhràin example [4], see Figure 2) installed in the controller in substitution of the tone-holes. In the zampogna case, because of the extremely simple traditional Italian performing techniques, an approximation of two-state tone-holes (open/closed) is sufficient. Always by using fully open/close particular fingerings, traditional players are able to create pitch bending in order to obtain a vibrato effect (as we discussed, air pressure is not involved). In short, cheap two-state capacitive sensor or push button technologies were considered for the prototype development.



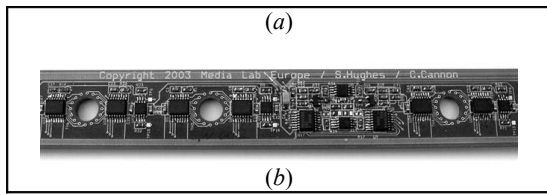


Figure 2. EpipE multiple-state tone-hole sensor: working scheme (a) and PCB example photo (b).

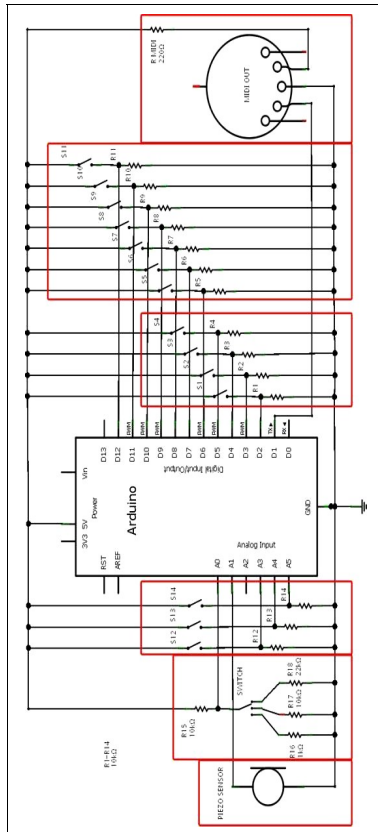


Figure 3. E-Zampognë controller electric schematics. From left to right: piezo sensor for the air pressure; working mode switch; tone-holes sensors for the drone and the left and the right chanter; MIDI-out port.

In our e-zampognë project, an experimental all-in-one electronic zampogna controller, with chromatic fingering on the chanter and a modern melodic drone (in according to most state of art experimental luthier models [2]) was developed. Various fingering schemes, related to the typologies explained in chapter 2. and shown in the table 4, are available.

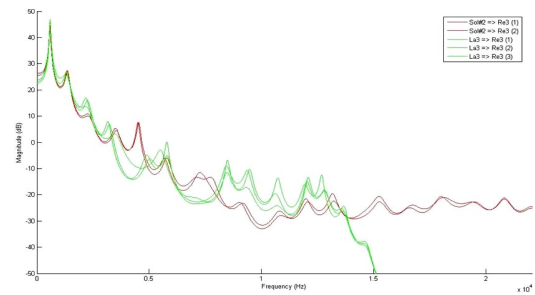


Figure 4. Comparison between two bagpipe frequency spectrums (formants extracted with 30 order LPC filter after shifting sound samples in pitch/time, to obtain the same pitch frequency). The correlation is not absolute, but just indicative: when trasposing single pipe sounds, sampled from pipes different by length, but with same fingering, most important formants often match. So frequency shifting for bagpipe synthesis purpose becomes acceptable.

Among the communication method, standard MIDI protocol and MIDI port were chosen because of their particular simplicity of implementation and their universal diffusion. The control parameters of the bagpipes are not particularly sophisticated; so they do not require high resolutions or speed. In terms of band rate, as we do not require any kind of continuous breath control, the limit of 31,250 bps imposed by the MIDI standard is acceptable.

The processing platform necessary to sense the gestural activity on the interface, and translate it into MIDI messages to the output port, is an Arduino Duemilanove. Its analog and digital inputs allow to detect in real time the state of 14 hole sensors located on the three pipes, the bag pressure sensor, one bagpipe-type selector. At each processing cycle, the microcontroller detects the states of the electronic components and, depending on the combination detected, generates the corresponding MIDI messages on three separate channels (two chanter and one drone). In the current implementation the controller is just a controller, that is, it is not able to generate sound as stand-alone instrument): it communicates with the synthesizer software through its MIDI port, placed at the end of the drone. The schematics is shown in Figure 3.

	A4 v.		G4 v.		E4 v.		C4
	A4 v.		F#4		Eb4		B3

	Ab4		F4		Eb4 v.		Bb3
	Ab4 v.		F4 v.		D4		
	G4		E4		C#4		

Table 4a. Example of chromatic e-Zampognè fingering (right chanter, “a chiave” mode). Note that vibrato effect (masked as “v.”) is available and two tone-holes are added in this scheme. They refers to the newest models of some “a chiave” and “zoppa” bagpipes [2]; these features became available adding two extra sensors in the controller.

	B3		A3		Ab3		G3
--	----	--	----	--	-----	--	----

Table 4b. E-Zampognè drone allows chromatic fingering and let the player to vary the accompaniment tone, usually fixed without tone-holes.

4. ZAMPOGNA SOUND FEATURES

In order to provide synthesis capabilities to the digital instrument, many traditional instruments were recorded and sampled; analyzing the notes from different pipes, and in particular comparing pipes of the same type but with different length, some particular spectral features emerged (see Figure 4).

There are some correlations between the frequency spectrum formant of the notes played with same fingering on different pipes (to find out this, pitch/time shifting was performed). This correlation allows to understand why simple pitch/time shifting processing can be very effective in transforming a zampogna into a bigger/smaller model. This kind of behavior allows to synthesize many different pitches just from few sampled examples.

The traditional tuning techniques for zampogna is very interesting: by changing reeds’ size and applying wax inside of the tone-holes it is possible to change both single note tuning and the entire tonality of the instrument (even if in a limited range, ± 3 semitones).

Moreover, because of its particular “sound”, the zampogna needs to be tuned in a non-equal tuning scheme. Usually, the typical tuning scheme is natural tuning, but many little regional differences exist. A modern semi-chromatic zampogna (following the Ricci model [2], with melodic drone) needs to be tuned in a different scheme (equal tuning or natural with some exceptions), that in all cases does not guarantee the typical perfect unison effect between pipes.

5. ZAMPOGNA SYNTHESIS

As a test tool, we developed a simple Max/MSP synthesizer that allows the user to select the type of bagpipe between the available typologies (see section 2), to customize the native tonality, the musical scale and single notes’ tuning. The synthesis algorithm implement an analysis/resynthesis subtractive technique: it allows to extract the timbre characteristics from original sound samples and assign them, with the possibility of frequency-shifting transformations, to the synthesized zampogna.

The most common synthesis algorithms (all with signal-based approaches) that are usually implemented in other digital bagpipe¹ are sampling and wave-table synthesis, in particular as the simple and computationally inexpensive implementation can be ported to a micro-controller core where the interface software loop can run too, thus creating a complete stand-alone instrument. Moreover the wave-table technique allows to implement another essential feature: timbre shifting (in frequency domain). A database of several wave-table samples representing every distinctive sounds from each pipe of each recorded zampogne (actually, minimum three examples, different by size, e.g: a little one, a medium one and a big-sized one for each typology) were created. As discussed before, because of intrinsic musical limits of the zampogna instruments (e.g. the absence of dynamics) the synthesizer needs few wavetables to simulate a complete instrument, obtaining a realistic sound.

One of the most influent limits of wave-table technique is that the typical tuning imperfections of a real zampogna inevitably get lost, so that a deeply realistic sound is not really possible to obtain. In order to overcome these limitations, we are now approaching the acoustic behavior of zampogne through physical modeling approaches.

6. EVALUATION

During a preliminary testing phase, carried out involving young traditional bagpipe players, some problems in controller stability emerged: in particular, tone-holes sensors (that, in the prototype were simply push-buttons, even if extremely sensible) revealed a delay effect that introduced a non-realistic tactile feeling, deceiving the player. This feeling of unease was mitigated by some

¹ like *Midi-Pipe* SGM200, *Fagerstrom’s Technopipe*, *Master gaita*, *Degerpipes*, *V-pipes*, The Fred Morrison Sound VST plugin by ePipes, etc.

practice on the instruments. Bag sensor introduced problems too: some unintentional signal noises and wrong response at the air pressure caused wrong sound starts and stops (about one error per minute). We solved these issues by using higher level technologies. The first prototype of the controller was appreciated by the Italian bagpipe community, that shared our documents and demo videos, showing also an interest into commercial production of e-Zampognë.

7. CONCLUSIONS

While electronic models using innovative pipelike controller that implements single-chanter control are diffused and commercially available (e.g.: *Technopipe*, *vPipes*, *Epipes*), e-Zampognë represents the unique electronic model of double or triple-chanter bagpipe. Eventually, there are several reason for which e-Zampognë project could be useful:

- *training*: digital controller has no tuning problems, allows a simplified technique and, also, volume adjustment;
- *cost*: the digital instrument costs less than traditional instruments, while implementing many models and many tonalities all in one.
- *ethnological research*: after building an complete database of original zampogna sounds, the instrument allows to compare different models using traditional songs or live performances transcribed in MIDI format;

Finally, you can create new experimental models of zampogna, combining different pipes or playing different timbres from other kind of instruments with bagpipes' musical techniques.

Some video examples of e-Zampognë are available at <http://ezampogne.altervista.org>.

8. REFERENCES

- [1] F. Guizzi, R. Leydi: *Le zampogne in Italia*. Ricordi, 1985.
- [2] M. Gioielli: *La zampogne. Gli aerofoni a sacco in Italia*. Iannone, 2005.
- [3] G. P. Scavone: The Holey Controller (see <http://www.music.mcgill.ca/~gary/>)
- [4] S. Hughes, C. Cannon, S. O'Modhrain: "Epipes: A Novel Electronic Woodwind Controller" in *proceedings of NIME'04*, Hamamatsu, Japan, 2004 (see <http://www.sarc.qub.ac.uk/~somodhrain/>).
- [5] C. Massarelli: *E-Zampognë – Analisi delle zampogne italiane e sviluppo di uno strumento musicale che ne recupera le caratteristiche*, Final dissertation, Politecnico di Torino, 2010 (see <http://ezampogne.altervista.org>)

PER UNA RAPPRESENTAZIONE AUDIOVISIVA DEL GESTO

Pietro Polotti

Conservatorio “G. Tartini” di Trieste
pietro.polotti@conts.it

Maurizio Goina

Conservatorio “G. Tartini” di Trieste
maurizio@goina.it

ABSTRACT

In questo articolo vengono presentati alcuni risultati dell'indagine che stiamo conducendo sulla rappresentazione audiovisiva del gesto in chiave artistica. Ci concentreremo in particolare sul lavoro Visual Sonic Enaction (VSE), un'installazione interattiva in forma di *public art* che permette ai visitatori di scoprire ed esplorare l'espressività del proprio gesto. A differenza di un normale sistema di pittura interattiva, VSE vuole essere uno stimolo a creare una rappresentazione audio-visiva dei propri movimenti fisici. Come di fronte a una sorta di “specchio” interattivo, ognuno può “guardare” ed “ascoltare” il proprio gesto. Il suono e l'immagine agiscono da feedback, stimolando e guidando il movimento in modo enattivo. I risultati audiovisivi generati dal pubblico durante una sessione dell'installazione VSE vengono registrati. Dopo l'evento, questa sorta di autoritratti multi-modali sono resi disponibili su internet in modo che chiunque possa dare una “occhiata audiovisiva” all'espressività corporea delle persone che hanno sperimentato il sistema.

Questo lavoro è uno tra le diverse realizzazioni inerenti lo studio della rappresentazione audiovisiva del gesto sviluppate nel corso degli ultimi tre anni ed è parte di un progetto più ampio di carattere metodologico, mirante a definire nuove strategie di indagine artistica tramite l'adozione di tecniche e modalità operative tipiche del design.

1. INTRODUZIONE

VSE è stata concepita come diversivo sonico all'interno di una rassegna di video-arte [1]. In contrasto alla passività intrinseca dell'atto di guardare un video nel contesto di una mostra, VSE ha proposto un'esperienza (en)attiva offrendo ai visitatori la possibilità di scoprire la propria espressività gestuale per mezzo di un feedback multi-modale ovvero audiovisivo.

Copyright: © 2010 Polotti and Goina. This is an open-access article distributed under the terms of the **Creative Commons Attribution License 3.0 Unported**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

In VSE il suono gioca un ruolo fondamentale, a dispetto della maggior evidenza dell'atto visivo del dipingere. Da un punto di vista enattivo, infatti, il suono fornisce un feedback immediato e più embodied di quanto offra il risultato visivo. Mentre sta “dipendendo”, l'utente “percepisce il suono nelle articolazioni” e modifica di conseguenza la propria azione. In questo senso il suono costituisce il perno tramite il quale vengono interconnessi i tre elementi di VSE, gesto, suono e immagine.

VSE si sviluppa partendo da un consolidato filone di ricerca sui sistemi interattivi per la produzione ed il controllo del suono (vedi [2] o, in generale, il dibattito all'interno della comunità *New Interfaces for Musical Expression*¹) e sull'espressività del gesto (vedi per esempio [3], [4]). VSE si basa inoltre sui concetti e la ricerca nel campo dell'Auditory Display (AD)² e del Sonic Interaction Design (SID)³, da un lato, e i principi del Basic Design, dall'altro [5], [6]. Per quanto riguarda il primo aspetto, l'idea fondamentale è di usare il suono non verbale come rappresentazione del gesto in maniera interattiva, o, meglio, enattiva. Nell'area di ricerca del SID, ciò è definito come sonificazione interattiva. Il secondo aspetto riguarda invece la semplicità dei principi adottati nell'analisi del gesto e nel mapping sonico e visivo. Tali aspetti rimandano inoltre alla volontà di rifarsi ai principi del design per dotare l'indagine artistica in ambito interattivo di linee guida forti mediante le quali poter perseguire delle finalità di ricerca precise in un panorama tecnologico in continua e frenetica evoluzione. Ciò costituisce il tema di discussione delle Sezioni 2 e 3.

In Sezione 4 vengono illustrate le modalità di sonificazione e visualizzazione interattive adottate in VSE. Nella stessa sezione, inoltre, si discute brevemente come la combinazione di principi ed elementi attinti da differenti discipline ed esperienze fanno di VSE uno strumento innovativo, le cui applicazioni possono andare al di là dell'ambito

¹ www.nime.org

² www.icad.org

³ www.cost-sid.org

artistico. Nelle Sezioni 5 e 6 si descrivono rispettivamente l'analisi del gesto e il tipo di *mapping* impiegati. La Sezione 7 fornisce una descrizione sintetica del sistema. Infine, in Sezione 8 si discutono i risultati e traggono alcune conclusioni.

2. RICERCA ARTISTICA E RICERCA SCIENTIFICA

Il significato più generale di questo lavoro consiste nel sondare la valenza di arte come ricerca e, in particolare, di arte interattiva come luogo di avvicinamento tra umanesimo e tecnologia. Quest'ultimo aspetto si colloca nell'ambito della rivoluzione epistemologica introdotta dalla *Human-Computer Interaction* (HCI): quando la tecnologia non è più impiegata solamente in un contesto di scienze esatte e naturali, ma entra in contatto con la complessità dell'essere umano, la natura non univoca dei fatti e dei comportamenti umani devono essere presi in considerazione per poter ideare e dare forma ad una tecnologia sensata e sostenibile per l'uomo.

In questo senso è interessante come il nome di una disciplina nata in ambito puramente informatico, e quindi scientifico, come l'interazione uomo-macchina (HCI) oggi abbia assunto la denominazione Interaction Design (ID), eliminando la "macchina" e introducendo un termine che potrebbe risultare ambiguo come quello del Design, ma che, se preso alla lettera, indica in realtà proprio una disciplina che sta a cavallo tra scienze esatte, umanistiche, tecnologia e arte.

Nel dibattito all'interno della comunità HCI/ID alcune delle domande che ci si pone di frequente sono: "Siamo ingegneri o designers? Siamo scienziati o filosofi?". O anche: "una validazione quantitativa di quanto produciamo e studiamo è sempre necessaria e congruente con il tipo di oggetti di studio e argomenti che trattiamo?". Più in generale, il pensiero euristico, un approccio olistico alle scienze e un ragionamento di tipo qualitativo/argomentativo trovano sempre più spazio nell'ambito delle scienze accanto al pensiero strettamente logico e deduttivo e al metodo sperimentale rigorosamente inteso, dove il principio di riproducibilità dei dati è garanzia di oggettività (vedi per esempio il testo fondamentale di Perelman e Olbrechts-Tyteca [7] per quanto riguarda la teoria dell'argomentazione e [8] per ciò che concerne specificatamente il nostro campo e in particolare il design del suono per il prodotto). Quando infatti nell'osservazione di un oggetto di studio il contesto gioca un ruolo fondamentale e imprescindibile, il principio riduzionistico della fisica e la definizione di modelli generali e semplici diviene impresa tanto più ardua quando l'oggetto di studio e il suo contesto implicano fattori umani, che siano psicologici, sociali o culturali. Esempi di questo dilemma si possono trovare nel rapporto tra le neuroscienze e l'approccio fenomenologico della psicologia sperimentale o, ancora, tra la stessa

psicologia sperimentale e la psicanalisi o tra la sociologia fondata su metodi quantitativi e l'etnografia, le prime basate sulla somministrazione di test circostanziati e miranti ad un'oggettività di tipo scientifico e le seconde fondate sull'osservazione dei fenomeni nella loro complessità (vedi per esempio [9] per quanto concerne il carattere qualitativo della ricerca etnografica). Una ricerca quindi più prossima ai paradigmi dell'indagine umanistica e altresì una ricerca in cui l'indagine artistica e lo "strumento artistico" possono dare un contributo in termini di costruzione di evidenze: arte come scoperta e definizione di nuove prospettive di comprensione della realtà.

Il lavoro qui presentato di indagine sul gesto mediante la sua rappresentazione sonora e visiva si colloca in un'ottica di questo tipo. In particolare, trattandosi di un'installazione interattiva in forma di *public art* emerge anche una valenza esplicitamente didattico-esplorativa del lavoro artistico, un punto di vista che del resto appartiene a tanta produzione di *public art* e *body art* fin dagli anni 70 (vedi per esempio l'opera di Dennis Oppenheim, ed altri artisti di quel periodo [10]), in cui ci si concentra sull'esperienza del pubblico e sulla sua partecipazione attiva.

A sottolineare ulteriormente questa situazione di confine tra arte e scienza, in cui il design si colloca a pieno titolo, è utile ribadire che questo lavoro è fortemente influenzato dai fondamenti delle discipline dell'AD e del SID. In particolare, la comunità che studia il display uditivo, cioè le modalità di impiego del suono non verbale per la rappresentazione di informazione è una comunità scientifica in senso stretto. La seconda comunità, quella del SID è invece per sua natura maggiormente ibrida e multidisciplinare nella stessa accezione di quanto detto circa l'ambito più generale dell'ID. La tesi che l'indagine attraverso il mezzo artistico possa essere interlocutrice della ricerca scientifica trova corrispondenza nel nostro lavoro in cui si vuol fornire uno spunto, un'evidenza delle potenzialità del suono non verbale come elemento di rappresentazione e comprensione della realtà: in questo caso particolare della gestualità umana. D'altro canto l'interconnessione tra i diversi campi procede anche lungo la direzione inversa, in quanto nel presente lavoro facciamo nostro un approccio ispirato al design e in particolare al SID.

3. PER UNA PRATICA ARTISTICA DESIGN-ORIENTED

A chi si cimenta con la creazione di una nuova interfaccia o sistema sonico/musicale non si richiede di essere necessariamente un ingegnere elettronico o un informatico. D'altro canto, se il lavoro avviene in ambito musicale o artistico in genere, l'etichetta di designer risulterebbe inappropriata.

Quello che ci domandiamo, tuttavia, è se le metodologie del design non possano fungere da punto di riferimento per un artista che lavora con le tecnologie digitali: fare arte ispirati da un approccio tipico del design, dove alle idee iniziali devono fare seguito molteplici ed alternative realizzazioni affinché queste possano essere confrontate e, attraverso il confronto, possano rivelare il maggior numero di punti critici riguardo all'idea di partenza.

Allo stesso modo il principio di iterazione ciclica di realizzazione, valutazione e ridefinizione della realizzazione (o dell'idea stessa) in base ai risultati della valutazione può costituire un paradigma forte per stabilire una pratica artistica strutturata e volta ad un'indagine sistematica di un'idea.

Inoltre, il principio base del design di passare attraverso delle fasi di sketching rapido e/o di messa in opera di cosiddetti *mock-up* fornisce una linea guida potente di fronte alla inevitabile rapidità di evoluzione della tecnologia e all'impossibilità di definire ed ancorarsi a degli standard. Un fatto questo che determina una mancanza di strumenti stabili e duraturi sulla base dei quali poter sviluppare "nuove tradizioni" e inedite peculiarità culturali e stilistiche, che, nel caso della nostra disciplina, si riflette per esempio nella difficoltà di creare nuovi strumenti musicali.

Il design, quindi, come linea guida per un rapporto "sostenibile" con il mezzo tecnologico in ambito artistico, da una parte, e strumento metodologico per una ricerca artistica strutturata, dall'altra.

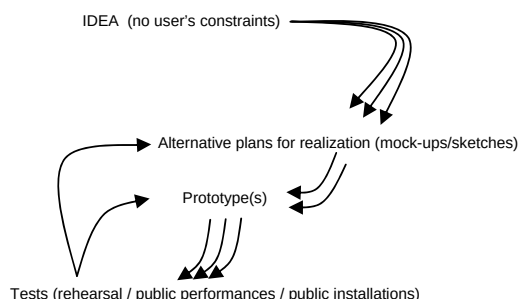


Figura 1. Un paradigma di pratica artistica ispirata al design.

Il quadro di una pratica artistica ispirata ai procedimenti del design è rappresentato dal diagramma di Figura 1, dove l'idea di partenza non è il frutto di una riflessione sulle esigenze del consumatore/cliente/committente, ma di una libera indagine artistica. Allo stesso modo, nel caso delle arti interattive, la fase di validazione/valutazione mediante *user-test* viene sostituita dalle prove con i performer oppure dall'esposizione del sistema al pubblico (in fondo una sorta di *user test*), non implicando delle valutazioni quantitative come nel caso del design, bensì puramente qualitative.

Aggiungiamo che, per quanto riguarda il nostro lavoro sulla sonificazione del gesto, VSE rappresenta una realizzazione particolare dei principi di sonificazione mediante elementi semplici e che altri luoghi di sperimentazione alternativi sono stati considerati, sia nell'ambito di performance interattive, di cui si fa cenno nella sezione successiva, che in una recente installazione pubblica, dove ci si è invece concentrati sulla gestualità ed espressività del camminare [11].

4. GENERAZIONE ENATTIVA DELL'AUDIO-VISUALIZZAZIONE DEL GESTO

VSE è un sistema multimediale che permette all'utente di esperire una rappresentazione sonico-visiva della propria espressività gestuale. Il sistema funziona in tempo reale, cosicché l'utente può vedere ed ascoltare l'effetto dei propri movimenti nel mentre li sta eseguendo.

Il sistema è stato presentato in un'accezione ludica, proponendo come scenario la metafora della pittura di graffiti: agli utenti viene proposto di dipingere su di un muro usando una "torcia elettrica/bomboletta spray" che controlla un algoritmo di generazione grafica e degli algoritmi di sintesi ed elaborazione del suono. Il suono stimola e guida i movimenti dell'utilizzatrice/utilizzatore e la/lo immerge in un'esperienza corporea-visiva-uditiva, producendo complessivamente un feedback multi-modale e continuo del gesto. Come già affermato, infatti, il suono è l'elemento legante e di fusione delle tre componenti di VSE.

Come prodotto finale, i visitatori hanno a disposizione un file audiovisivo con la registrazione della rappresentazione sonico-visiva astratta dell'espressività delle loro performance. Inoltre chiunque può prendere visione dell'espressività corporea dei visitatori consultando un sito internet pubblico, in cui sono reperibili le registrazioni dei ritratti sonico-visivi.

Si possono immaginare molte altre potenziali applicazioni per tale sistema, che spaziano dalla performance artistica all'educazione o al recupero senso-motorio, alla didattica musicale e altro ancora. In effetti il sistema è stato concepito inizialmente come strumento per la sonificazione del gesto di un performer e denominato EGGS (Elementary Gestalts for Gesture Sonification) [12], [13]. Nello spirito di EGGS, il performer/danzatore non segue un pezzo musicale, non controlla l'esecuzione di un pezzo musicale, e neppure genera alcuna musica col proprio movimento. Piuttosto ella/egli ascolta il proprio gesto, enattivamente, modificando e controllando la propria azione performativa in base al suono prodotto. Il suono, quindi, è una rappresentazione del movimento, una conseguenza sonora e un feedback continuo, in nessun modo esterno al gesto stesso, che, in qualche modo,

aumenta la propriocezione del performer. È facile immaginare ulteriori applicazioni di questi principi in altri campi, dove siano cruciali il movimento corporeo ed il suo controllo nel tempo.

In VSE lo stesso principio è applicato anche al dominio visivo. Lo scopo non è dipingere. Piuttosto, ciò che appare sul muro o sullo schermo del computer è una visualizzazione della espressività gestuale del fruitore. Allo stesso tempo, in modo enattivo, il feedback visivo induce l'utente a modificare e controllare il proprio gesto anche a seconda del tipo di grafica utilizzata.

5. SUONI E SEGNI GRAFICI ELEMENTARI

L'obiettivo più generale di VSE è quello di indagare la relazione tra il gesto e la sua rappresentazione sonora e visiva per mezzo di un approccio elementare. Questo lavoro prende ispirazione dagli ideali del Bauhaus e la ricerca di Paul Klee sulle forme e la rappresentazione pittorica [14] e dalla già menzionata scuola di Basic Design. L'idea principale è quella di ridurre la complessità della realtà a gruppi di componenti semplici ed elementari che possano essere maneggiati e ricombinati agevolmente per formare strutture più complesse. In linea con queste idee, abbiamo definito diverse strategie per ridurre il gesto ad una sequenza di unità piccole ed elementari (gestalts) usate per controllare degli insiemi basilari di suoni e di segni grafici.

Nella versione corrente, sono stati definiti tre insiemi di suoni elementari e tre gruppi di segni grafici semplici, usati per la sonificazione e la visualizzazione di due categorie basilari di gesti: movimenti dritti e movimenti circolari. Al visitatore era consentito sperimentare a scelta tutte e tre le tipologie di segni grafici, ma solamente un gruppo sonico alla volta. All'interno di un singolo ritratto, corrispondente ad uno solo dei tre suoni, il visitatore poteva passare a sua scelta da un set grafico all'altro, fra i tre disponibili, scuotendo delle bottiglie colorate dotate di sensori di movimento e appoggiate a lato della postazione. Le tre tipologie grafiche sono illustrate schematicamente in Figura 2. I tre tipi di suono venivano invece sperimentati separatamente uno dopo l'altro in tre diversi ritratti sonico-visivi. I tre insiemi sonori, rappresentati iconicamente in Figura 3, includevano suoni sibilanti e metallici, suoni gravi di sintesi per modulazione di frequenza e tintinnii di vetri e cristalli sintetizzati mediante i modelli fisici del Sound Design Toolkit (SDT) [15], [16]. La tipologia di suono cambiava solo quando l'utilizzatore decideva di salvare il ritratto audiovisivo corrente e di passare al successivo. Sono stati implementati diversi tipi di mapping, più o meno differenziati, discreti o continui, in modo da rendere varie ciascuna delle nove possibili combinazioni sonico-grafiche. I diversi mapping sono

descritti in dettaglio nella sezione seguente. Un esempio di risultato grafico è riprodotto in Figura 4.

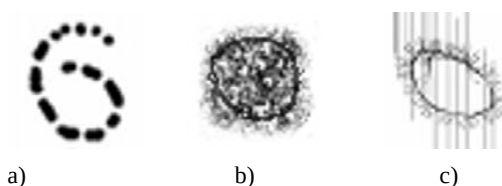


Figura 2. I set di segni visivi: a) a guisa di punti, b) nubi colorate, c) geometrico/grafico.

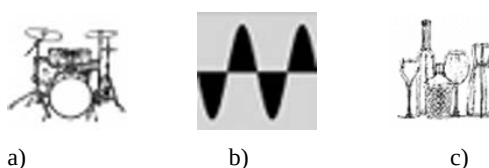


Figura 3. Rappresentazione mediante icone dei gruppi sonori, a) suoni sibilanti e metallici, b) suoni gravi di sintesi per modulazione di frequenza, c) suoni di oggetti di cristallo.

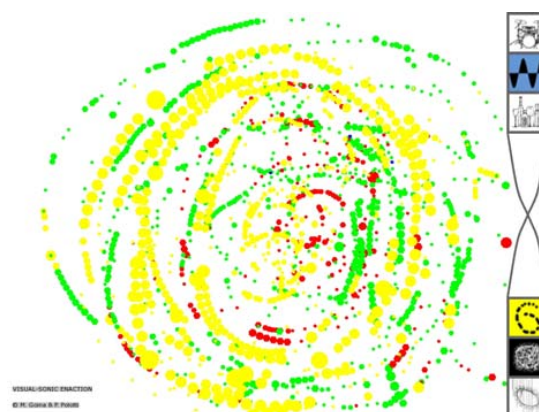


Figura 4. La “tela” VSE con un autoritratto sonico-visivo di uno dei partecipanti.

6. MAPPING

In VSE, il comportamento gestuale viene analizzato da un programma di motion tracking, che rileva una sorgente luminosa, valuta la traiettoria e discrimina tra le seguenti categorie basilari di movimento: rettilineo, circolare in senso orario, circolare in senso anti-orario, di inversione della direzione, e di quiete (assenza di movimento).

Come già anticipato, sono stati definiti diversi mapping sonici e visivi per differenziare in maniera più o meno accentuata le categorie di cui sopra. Il set visivo a guisa di punti è quello più elementare e distingue tutte le categorie di movimento utilizzando quattro diversi colori fondamentali per rappresentarle. Le nubi colorate

cambiano gradualmente tonalità di colore dal movimento circolare orario a quello antiorario, mentre il movimento rettilineo è trattato come caso intermedio. Il terzo gruppo di segni grafici, quello geometrico, distingue solamente tra movimento rettilineo e circolare, senza differenziare la rotazione oraria da quella antioraria.

Riguardo la sonificazione, sia il gruppo dei suoni sibilanti/metallici che quello dei suoni di cristallo distinguono in maniera discreta (non continua) tra traiettoria rettilinea e circolare mediante due sottoinsiemi di suoni ben distinti. Inoltre, i suoni sibilanti/metallici cambiano timbro variando i parametri del suono in accordo con un mapping di tipo spaziale, relativamente cioè alla posizione sul muro che viene dipinto. I suoni di cristallo vengono invece modulati timbricamente in base all'angolo di curvatura del gesto circolare ed alla posizione spaziale quando il movimento è rettilineo. Infine i suoni FM cambiano timbro allo stesso modo del caso grafico delle nuvole colorate. Tale varietà ha come fine quello di evitare un mapping costante che potrebbe pregiudicare la spontaneità dell'esperienza a causa di uno schema ripetitivo, e porta quindi ad una maggiore ricchezza e varietà di risultati.

Oltre agli aspetti geometrici, un elemento chiave nella rappresentazione sonico-visiva è l'energia, cioè il contenuto dinamico del gesto. Una combinazione di valori istantanei della velocità e dell'accelerazione controlla la dimensione del segno grafico e l'intensità del suono prodotti in una modalità differente per ciascuna delle coppie sonico-visive possibili (per una migliore comprensione si veda il video disponibile sul sito del progetto [17]).

7. ASPETTI TECNICI DELL'ALLESTIMENTO

L'allestimento di VSE prevede i seguenti elementi: un video-proiettore collegato al computer, per proiettare su di una parete bianca il ritratto di dimensioni approssimative 3 per 2 metri; un diffusore acustico amplificato, connesso al computer, per diffondere il suono nello spazio pubblico; una cuffia senza fili indossata dall'utilizzatore necessaria a realizzare un'esperienza audio più coinvolgente e immersiva; una torcia modificata, funzionante come una sorta di bomboletta spray luminosa, tenuta in mano dal visitatore per dipingere sulla parete i graffiti digitali; una video-camera senza fili per rilevare la luce della torcia; tre bottiglie colorate per scegliere, scuotendole, uno dei tre tipi di segni grafici disponibili; un grande bottone rosso del tipo di quelli di emergenza che funge da interruttore: premuto dall'utente, salva il ritratto sonico-visivo corrente e permette di passare al set sonoro successivo; l'algoritmo per la produzione dei segni grafici e quello per l'analisi del gesto sono implementati in Processing [18]; la sintesi del suono è invece realizzata in ambiente Max-MSP [19]; le due

applicazioni comunicano tra di loro tramite il protocollo Open Sound Control (OSC) [20]. L'intero set-up viene illustrato in Figura 5 e 6.

Al pubblico vengono presentate le seguenti istruzioni:

Visual Sonic Enaction

Il suono guida la pittura

- *Dipingi sul muro con la torcia-spray*
- *Scuoti le bottiglie per cambiare lo stile grafico*
- *Premi il bottone per salvare il tuo ritratto sonico-visivo e iniziare un nuovo dipinto guidato da un nuovo suono.*

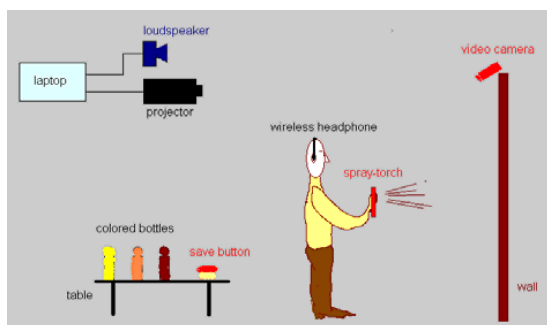


Figura 5. Vista laterale del setup dell'installazione.

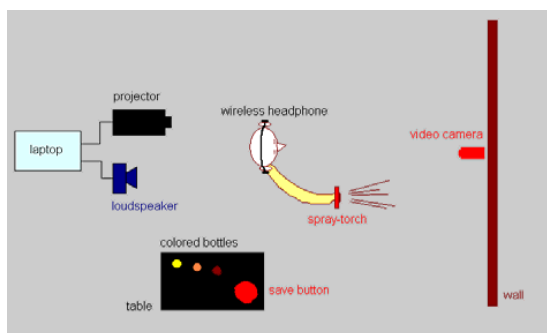


Figura 6. Vista dall'altro del setup dell'installazione.

8. CONCLUSIONI

Il tipo di partecipazione e di coinvolgimento del pubblico è ben evidente nella documentazione video. Un estratto delle riprese video e tutti gli autoritratti sonico-visivi prodotti dal pubblico durante la presentazione al festival video Catodica sono disponibili sul sito internet del progetto [17]. La varietà dei ritratti prodotti illustra bene le diverse personalità di coloro che hanno utilizzato VSE. In generale, l'installazione ha avuto successo in termini di apprezzamento e partecipazione del pubblico.

Il risultato principale è stato quello di constatare come, nella maggior parte degli utilizzatori di VSE, il suono realmente influenzi il gesto. In molti casi è stato evidente un progressivo aumento del coinvolgimento del partecipante, stimolato dalla presenza inusitata di un

feedback sonoro, a cui si abituava gradualmente e da cui si lasciava poi influenzare. Consideriamo ciò un risultato positivo nello studio del suono come mezzo di rappresentazione e stimolo dell'espressività corporea, applicabile sia in un contesto artistico che in altri e più generali ambiti, in cui si possa fare uso di sistemi interattivi. Come già affermato, non rientra nelle nostre finalità fornire una stima quantitativa dei risultati di questo lavoro, ma piuttosto produrre un'evidenza e, eventualmente, uno spunto per uno studio più rigoroso sul tema.

9. RICONOSCIMENTI

I nostri ringraziamenti vanno a Maria Campitelli del Gruppo78, al festival Catodica [1] e a Fucine.it.

10. REFERENCES

- [1] www.catodica.it (Dec. 12, 2010).
- [2] J. Drummond: "Understanding interactive systems," *Organised Sound* 14, 2, pp. 124-133, Aug. 2009.
- [3] A. Camurri, C. Drioli, B. Mazzarino, and G. Volpe: "Controlling Sound with Senses: multimodal and crossmodal approaches to control of interactive systems," in *Sound to Sense, Sense to Sound - A State of the Art in Sound and Music Computing*, P. Polotti and D. Rocchesso Eds. Logos Verlag, Berlin, pp. 213-243, 2008.
- [4] S. Kopp, and I. Wachsmuth: *Gesture in Embodied Communication and Human-Computer Interaction*. Book series: Lecture Notes in Computer Science. Vol. 5934/2010 Springer, Berlin / Heidelberg, 2010.
- [5] G. Anceschi: "Basic design, fondamenta del design" [Basic design, foundations of design], in *L'ambiente dell'apprendimento - Web design e processi cognitivi [Learning environment - Web design and cognitive processes]*, G. Anceschi, M. Botta, M. A. Garito Eds. McGraw Hill, Milano, Italy. 57-67, 2006.
- [6] D. Rocchesso, P. Polotti, and S. Delle Monache, "Designing Continuous Sonic Interaction," *International Journal of Design (IJD)* 3, 3, pp. 13-25, Dec. 2009.
- [7] C. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric, A Treatise on Argumentation*. University of Notre Dame, Indiana, 1969.
- [8] U. Jekosch, "Assigning Meaning to Sounds - Semiotics in the Context of Product-Sound Design," in J. Blauert, ed. *Communication Acoustics*, Berlin, Springer, pp.193-221, 2004.
- [9] Editoriale a cura della Direzione, "Una nuova rivista", *Etnografia e ricerca qualitativa*, Il Mulino, vol. I, pp. 3-8, 2008.
- [10] R. L. Goldberg, *Performance Art. From Futurism to the Present*, Thames & Hudson, pp. 156-163, 2001.
- [11] <http://visualsonic.eu/sw.html> (Dec. 12, 2010).
- [12] M. Goina, and P. Polotti: "Elementary Gestalts for Gesture Sonification," *Proceedings of the 2008 International Conference on New Interfaces for Musical Expression (NIME-08)*. Genova, Italy, pp. 150-153, 2008.
- [13] M. Goina, P. Polotti, and S. Taylor: "Swish & Break (Geschlagene-Natur), interactive performance in the *Concert around Freesound, SMC 2010, 7th Sound and Music Computing Conference*, (video disponibile su http://visualsonic.eu/eggs_in_action.html) Universitat Pompeu Fabra, Sala Polivalent, Barcelona, Spain, 22 July 2010.
- [14] P. Klee *Pedagogical Sketchbook*, trans. Sibyl Moholy-Nagy. Frederick A. Praeger, New York, 1965.
- [15] Closing the Loop of Sound Evaluation and Design (CLOSED), FP6-NEST-PATH project no: 29085, *Deliverable 2.3*, disponibile online: http://closed.ircam.fr/uploads/media/CLOSED_D2.3.pdf, 2009.
- [16] S. Delle Monache, P. Polotti, and D. Rocchesso, "A Toolkit for Explorations in Sonic Interaction Design", *Proceedings of Audiomostly '10*, Pitea, Sweden. Settembre, 15 - 17, 2010.
- [17] <http://visualsonic.eu/vse.html> (Dec. 12, 2010).
- [18] <http://processing.org/> (Dec. 12, 2010).
- [19] <http://cycling74.com/> (Dec. 12, 2010).
- [20] <http://opensoundcontrol.org/> (Dec. 12, 2010).

Experiencing Sonic Interaction Design: Product Design Activities at the SID Summer School 2010

Stefano Delle Monache

IUAV - University of Venice
Venezia, Italy
stefano.dellemonache@gmail.com

Davide Rocchesso

IUAV - University of Venice
Venezia, Italy
roc@iuav.it

ABSTRACT

Educational activities represent important means of dissemination and experimentation of research outcomes. The Product Sound Design Summer School 2010, organized in the scope of the EU-funded COST Action on *Sonic Interaction Design* (SID), aimed at bringing together the state of the art of methods and tools investigated and developed so far in the SID community, and starting an integration process of SID with Product Development and Design. The paper describes the learning environment and its aims, the activities carried on and the impact of the outcomes in terms of inspirational sources for future investigations in SID.

1. INTRODUCTION

Sonic interaction design is the activity of shaping the relationship between human and objects, and between humans through objects, by means of sound [1]. Artifacts that integrate computational capabilities are increasingly inhabiting our everyday environment, from household appliances, entertainment systems, to mobile communication devices, clothes and furniture. Objects turn multifaceted, and communication meets information. As technology becomes more and more miniaturized and embeddable, devices get smaller and visual displays tend to disappear. Sound can provide a good alternative and opportunity to convey information, communicate and share the intimate, inner qualities of use and interaction with and through objects. Notwithstanding, to provide interactive artifacts with appropriate, expressive, functional acoustic behaviour is not an easy task. Indeed, how many annoying auditory displays can we account, so that after a while one prefers to turn the device silent?

The design question is about how to make a product sensible to manipulative actions, not as an intelligent or knowledgeable subject, but as an object capable of feeding the stimuli back in a dialogic form. For this reason, prominent experts of interaction design claim that the development of a literacy on understanding, interpreting, and building up of a repertoire of sketches and prototypes is a

core issue in interaction design [2, 3, 4]. Designing for manipulative interaction requires a shift of thinking and approach to practices, research and educational activities from the product itself towards interaction. Designers need tools and techniques to explore, visualize, interpret, refine interaction and accompanying products. On the other side research and education through shared practices of creation, manipulation, interpretation are aimed at being sensitized to and exploit the richness and expressiveness of movement, of both users and objects, and sound.

So far, SID related research investigated a plethora of approaches, and tools with the aim of developing a corpus of knowledge on sound to be exploited in interaction design activities [5, 6, 7, 8, 9, 10]. The development of such a literacy represents a crucial step towards the education of the future product design team members with a specific competence on interactive sound. Since 2008, the COST-Action IC0601 on SID¹ organized several training activities devoted to the creation and consolidation of this innovative and interdisciplinary domain. The SID Summer School took place in August 2010 and it was aimed at summarizing some results of three years of research in *sonic interaction design*, widening the knowledge of the Action activities, and providing intensive training on product sound design in the scope of product development.

The paper has the following structure: in Section 2, we introduce the Product Sound Design Summer School 2010, the learning environment and its aims; Section 3 describes the organization and program of the activities carried on; in Section 4, we reflect on the impact of the training school; in Section 5, we draw our conclusions.

2. THE PRODUCT SOUND DESIGN SUMMER SCHOOL 2010

The Product Sound Design Summer School (see fig. 1) took place at the Aalto Design Factory, Espoo, Finland, from the 23th to 26th of August 2010². A project of the Aalto University³, the Design Factory⁴ is a multidisciplinary experimental platform focusing on *product development and design*⁵ [11], and provides a framework where educational, research, and business opportunities are

Copyright: ©2010 Stefano Delle Monache et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹<http://www.cost-sid.org/>

²<http://trac.sme-ccppd.org/SID/wiki/SIDTrainingSchoolProductSoundDesign201008>

³<http://www.aalto.fi/en/>

⁴<http://aaltodesignfactory.fi/>

⁵<http://www.ulrich-eppinger.net/>

integrated in mutual collaborative projects. In this framework, the Summer School program was constructed around two main guiding principles:

1. to start an integration process of SID product sound design with *product development and design*;
2. to bring together the state of the art of the several tools, techniques and interdisciplinary guidelines that are constantly under development and investigation in SID community.



Figure 1. The poster announcing the Product Sound Design Summer School

Participants were recruited with an international call for applications, with the objective of selecting 16-20 students in the last year of masters studies, or first two years of PhD studies in design, engineering, economics, or sound-related studies. Applicants were required to send a CV, a letter of motivation, and an official transcript of their studies. Eventually, 20 participants were selected out of 56 applications.

The summer school was positioned at the intersection between *product development and design* and *sonic interaction design*: problem-based learning activities – that involve ideas to realize and problems to solve by planning, searching for information, decision making, manufacturing, assembling and testing – were merged with SID activities, such as making designers sensitive to sonic interactions by means of soundwalks, by analyzing and imitating sonic gestures, and by sonic explorations devoted to concept development, sound sketching and scenario enactment.

For this purpose, two partnering companies, *Hipui*⁶ and *Powerkiss*⁷, were involved in order to provide participants with a real scenario to be confronted with. For the summer school, Hipui was interested in exploring the use of non-verbal sound to support the execution of gestural input in handheld and eyes-free devices. Powerkiss' open problem was in auditory signalling their wireless charging technology, so that the latter could be invisible and merge completely with the host artifact, a table for instance. Sonic branding aspects and context of use, private or public, had to be considered.

In figure 2, the proposed design themes frame the contents of the whole training activities.

⁶ <http://www.hipui.com/>

⁷ <http://www.powerkiss.com/>

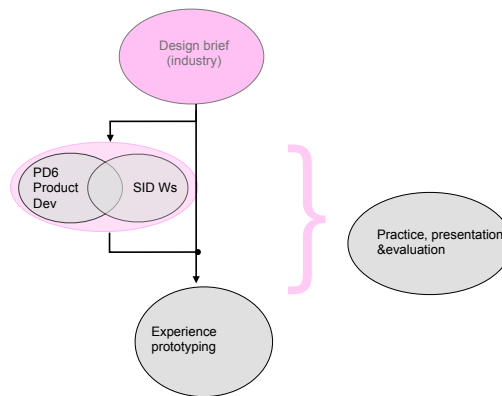


Figure 2. The structure of the Summer School

3. ORGANIZATION AND PROGRAM

3.1 Day One

The first day was devoted to provide some basic skills and tools aimed at being sensitized to sound in interaction, at collecting and enacting fast and rough ideas, and at generating quick and dirty sonic prototypes:

- Soundwalk and listening session (F. Behrendt): it encompasses a series of listening and sound walking exercises, aimed at sensitizing to sonic interactions, improving listening skills, expanding the vocabulary to talk about sounds, representing visually and orally the temporal, spatial, embodied aspects of sound. In particular, guidance was given in order to reflect on the role of sound in both urban and indoor space, how sound can facilitate or disturb navigation and orientation, and how it is possible to design sonic interactions for complex multisensory environments [12].
- Vocal sketching session (I. Ekman): it introduces the use of vocal sounds as means for rapid sonic prototyping of design ideas, especially in the early stage of the design process. Can sound designers make use of their voice, in the same way visual designers sketch on paper with pencil? The session motivates participants to exploit their vocal capabilities in a series of warming up exercises and design-oriented task. For instance, groups were asked to use their voice to sketch the sonic behaviour of a given artifact [7].
- PD6 - product development in 6 hours (W. Raduma): it is a workshop format developed at Design Factory. A product development process is squeezed in 6 hours. PD6 is a tool for accelerating team building, enacting fast and rough ideas generation, improving communication and effective problem solving. Prototyping and hands-on are at the center of a non-linear, iterative process of planning, concept development, design, testing and refinement [13, 14].

Two teams (1 and 2) were asked to create a system that makes wireless charging intuitive, easy and informative, while the remaining teams (3 and 4) had to create an interactive alternative for a display and keyboard/touch interface. Requirements for both briefs were that feedback and interaction should be handled by means of non-verbal sound and gestures.

3.2 Day Two

The second day was dedicated to provide analytical tools for concept development and to experiment with various sound making techniques for sonic prototyping:

- Narration and performativity in sonic interactive commodities (D. Hug): this session provided analytical tools for interpreting and developing narratives around sound, starting from a systematic analysis of film or game sound cases. In particular, the concept of narrative metatopics was introduced, namely abstracted themes and attributes associated with narratively significant artifacts and interactions in fictional media [9]. Participants learned to identify design ideas, and apply narrative strategies to sound design. In practical sessions they were asked to experiment with Foley sound making techniques, and to refine and prototype with a Wizard-of-Oz approach the outcomes of the PD6.
- Sound - from perception to basic design (S. Delle Monache): this session introduced an analytical approach to sound perception and listening via experimenting various manipulations of sound feedback in continuous and multisensory interaction. According to a basic design approach, the groups were asked to solve a design problem, with well-defined objectives and constraints, and to prototype the sonic interaction by manipulating the control parameters of some sound models, available in the Sound Design Toolkit, a physics-based sound generation application [8, 10].

3.3 Days Three and Four

Day three and four were finally dedicated to independent work of the groups. The briefs given for the PD6 workshop were refined and better specified as follows:

- Powerkiss: Create a system that makes wireless charging intuitive, easy and informative. Extend you device for social information sharing;
- Hipui: make an interactive alternative for a display and keyboard/touch interface. Extend your device for information navigation;
- Requirements for both: feedback and interaction are handled by the use of non-verbal sound and gestures. Demonstrate the discovery of its functions and capabilities in an extreme situation with experience prototyping.

4. OUTCOME

To run the school, the twenty selected participants were grouped in four multidisciplinary teams, each composed of two designers, one engineer, one economist, one musician. Teams 1 and 2 dealt with the design theme suggested by Powerkiss, while the remaining two teams (3 and 4) concentrated on the Hipui theme.

During the first two days, specific workshops exposed the participants to the basics of product design and development, and gave them the opportunity to experience various perspectives on sonic sketching and prototyping, in order to share a common starting platform. In the second half of the school, the groups were asked to explore the design themes indicated by the companies, and to design mock-ups and prototypes to demonstrate their exploration. Finally, their realizations were evaluated by companies, instructors, tutors, and evaluators (see fig. 3).

As seen from the participant, the School took a tele-scopic form, with an initial introduction that tried to encompass the whole workshop development. The School chief organizer, Cumhur Erkut, initially explained the aims and strategy by devoting just a couple of sentences to each of the fundamental components (soundwalk, toolkit, etc.). This priming allowed the participants to start thinking about how to gain the most from the planned activities. Just as important, there was the description of the environment and facilities, as given by the Design Factory director Kalevi Ekman. The partner companies were also called to present their expectations to the participants, although these were kept quite open to the unexpected. Then the activities of sound walking and vocal sketching were conducted, as described in section 3, before the crash exposition to the product development cycle.

As activities such as sound walking and vocal sketching require some practice and reflexion, they were also continued in the early second day, with a more analytic attitude. This turned out to be beneficial to introduce theory and practices in sound narration and performativity. By using physical objects of a toolkit, the participants were asked to produce sounds with some qualities (e.g., searching / finding with hesitation), or to mimic salient film sounds. It is interesting to notice that software sound models were introduced only after extensive direct experience of physical sounding objects. Actually, sound models were introduced with a rethoric trick. The participants were asked to reproduce, by using physical objects, a given recorded sound. Only after their direct, physical attempt, the fact that the reference sound was synthetic was actually unveiled. On the one hand, this made the participants trust the models. On the other hand, it encouraged them to adopt a similar approach when designing the sound using software models.

Two basic design assignments were accomplished in the second day. One was to create a soundscape with two impact sounds and two friction sounds, where none of them stands out (the Antiprimadonna). It is remarkable that all groups got different yet interesting and balanced soundscapes. The second assignment was to sonify the gesture of stirring flour and sugar with water, and it made the par-

ticipants aware of the performative potential of everyday objects, when they are properly augmented with sound.

In days three and four the groups were left largely alone in developing their projects. They were allowed to expand or deviate from the preliminary sketches, but checkpoints were established in both days to keep them on a secure track. It was agreed that prototypes could still be based on wizards, but the sounds, as compared to the initial sketches should be refined.

A short video documentation of the activities carried on in the summer school can be found at <http://www.vimeo.com/16655747>.



Figure 3. Evaluation by Maija Itkonen, CEO of Powerkiss

5. EVALUATION

The Product Sound Design Summer School was announced on various social, professional and research networks, blogs, institutional websites, and mailing lists. 56 applications were received from all over the world: 39% Engineering, 36% Design, 9% Economics, and 16% Music. Musical skills, amateurish or professional, are the binding factor between the diverse backgrounds. Compared to previous educational activities organized by the SID-COST Action, the applicants profile was much more design- and less scientific- or artistic- oriented. Conversely, almost nobody had specific skills on *sonic interaction design*.

The letters of motivation served to frame the expectations of candidate participants and highlighted a strong demand of understanding by doing what a sonic interactive, aesthetic experience is, and how it is actually possible to shape it designarily, with respect to:

- the product itself (product designers);
- the technological development and implementation (engineers);
- the impact of product sound on users' experience, brand positioning and customers' loyalty (managers);
- the inner dynamic relationships between sound, shape and materials beyond a musical application, and to environmental sustainability (musicians).

For this purpose, the workshop modules were constructed in order to ensure a ratio of 80-90% of practice and 20-10% of related contact teaching. This explains why an intensive training on sound-related methods and techniques was given in the early stage of the design process. The specific interests raised in the letters of motivations combined naturally in the group exploration in the second half of the training school, integrating various aspects of the product development process.

After completing the summer school, students were asked to evaluate, in a feedback form, the workshops, the instructors, their experience of the training activities, and participation to groups work. In general, the training school gained an overall high score. The activities and the various perspectives on SID, though condensed in two very intensive, and sometimes hectic, days, were almost unanimously evaluated useful learning experiences, with a high degree of applicability to own work. Considered the multidisciplinary background of the participants, and the different levels of access, such a result confirms a value and trust in the tools and techniques that are being investigated in the SID community.

As seen from the instructors, tutors, and evaluators, the school was a great source of inspiration. It was understood that giving much details about software tools (for example, detailing the sophisticated mapping strategies that can be implemented) diverts the attention from the design aspects of sound in interaction. It was recognized that the diverse backgrounds that were initially mixed in all groups are actually needed to develop good product sound design, and that a common platform for effective team work can be found. The problems and hints giving by the partner companies were at the same time concrete and broad and they elicited a number of research questions that are likely to animate the community for several months to come.

6. CONCLUSION

The Product Sound Design Summer School 2010 represented an important means of dissemination and experimentation in context of a series of methods and tools developed so far in the SID community. In four days, participants had the opportunity to experience and practice *sonic interaction design*. A strong, hands-on, pedagogical approach to the different matters enabled students to start quickly a discourse around sonic interaction and product design. Moreover, it gave a measure of the effectiveness of an integration between SID and product development, towards the definition of a format of activities and disciplines in SID education.

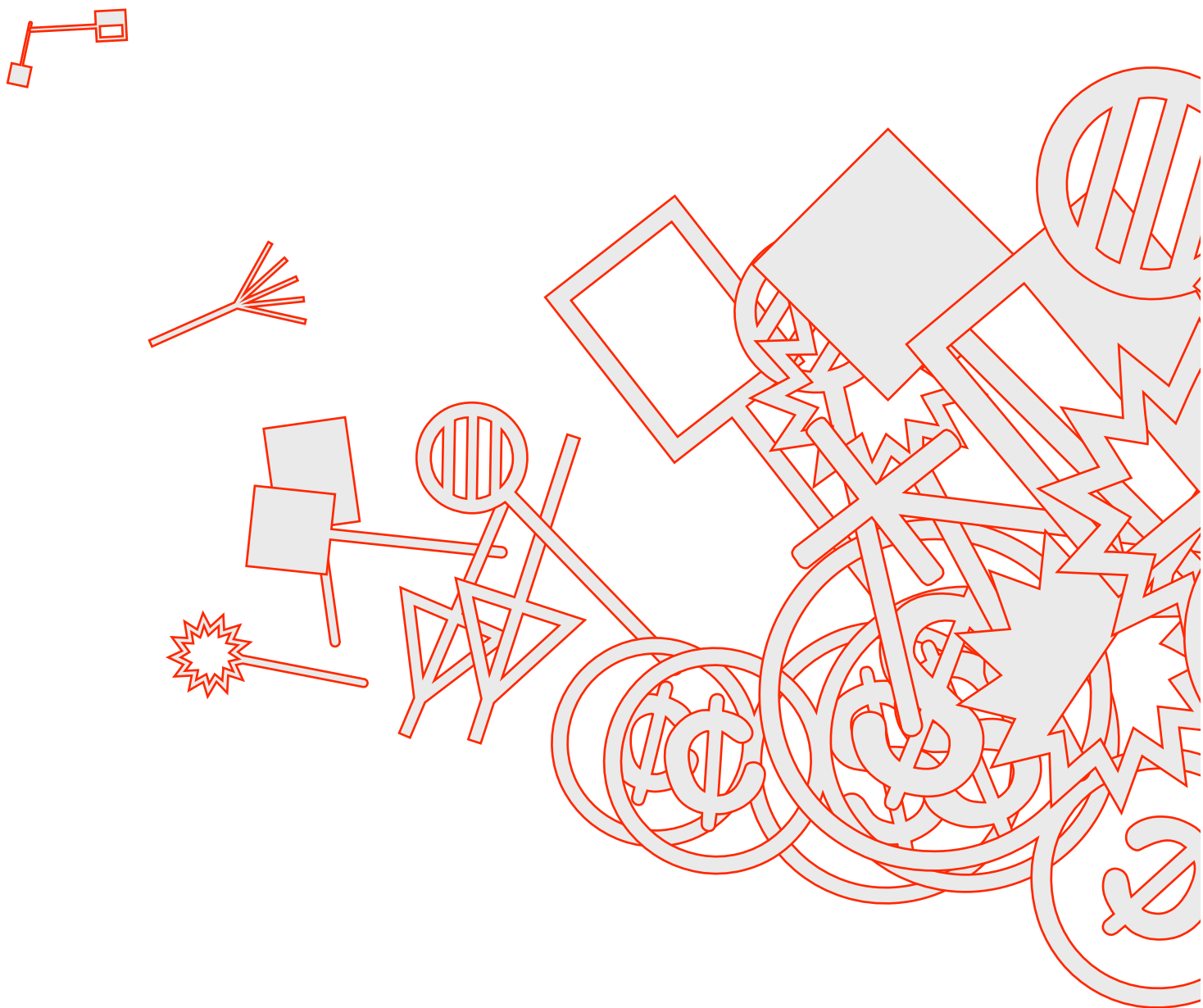
7. ACKNOWLEDGEMENTS

Thanks to: Stephen Barrass (evaluator), Frauke Behrendt (tutor), Inger Ekman (tutor), Kalevi Ekman, Cumhur Erkut (chief organizer), Daniel Hug (instructor), Antti Jylhä, Sandra Pauletto (evaluator), Wycliffe Raduma, and Sylviane Sapir (evaluator). The authors were, respectively, instructor and evaluator.

8. REFERENCES

- [1] D. Rocchesso and S. Serafin, "Sonic interaction design," *International Journal of Human-Computer Studies*, vol. 67, 2009. Special issue editorial, to appear.
- [2] J. Bardzell, J. Bolter, and J. Löwgren, "Interaction criticism: three readings of an interaction design, and what they get us," *Interactions*, vol. 17, no. 2, pp. 32–37, 2010.
- [3] Y.-K. Lim, E. Stolterman, and J. Tenenbergh, "The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas," *ACM Trans. Comput.-Hum. Interact.*, vol. 15, no. 2, pp. 1–27, 2008.
- [4] J. Löwgren, "Toward an articulation of interaction esthetics," *New Review of Hypermedia and Multimedia*, vol. 15, no. 2, pp. 1361–4568, 2009.
- [5] E. Brazil, "A review of methods and frameworks for sonic interaction design: Exploring existing approaches," in *Auditory Display - 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009, Revised Papers* (S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, eds.), vol. 5954 of *Lecture Notes in Computer Science*, pp. 41–67, Springer, 2010.
- [6] K. Franinovic, L. Gaye, and F. Behrendt, "Exploring sonic interactions with artifacts in everyday contexts," in *Proceedings of the 14th International Conference on Auditory Display*, (Paris, France), 2008. inproceedings.
- [7] I. Ekman and M. Rinott, "Using vocal sketching for designing sonic interactions," in *DIS '10: Proceedings of the 8th ACM Conference on Designing Interactive Systems*, (New York, NY, USA), pp. 123–131, ACM, 2010.
- [8] D. Rocchesso, P. Polotti, and S. Delle Monache, "Designing continuous sonic interaction," *International Journal of Design*, vol. 3, December 2009.
- [9] D. Hug, "Investigating narrative and performative sound design strategies for interactive commodities," in *Auditory Display - 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009, Revised Papers* (S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, eds.), vol. 5954 of *Lecture Notes in Computer Science*, Springer, 2010.
- [10] S. D. Monache, P. Polotti, and D. Rocchesso, "A toolkit for explorations in sonic interaction design," in *AM '10: Proceedings of the 5th Audio Mostly Conference*, (New York, NY, USA), pp. 1–7, ACM, 2010.
- [11] K. T. Ulrich and S. D. Eppinger, *Product Design and Development*. McGraw-Hill, 2008 (1st. ed. 1995).
- [12] H. Westerkamp, "Soundwalking," *Sound Heritage*, vol. 3, no. 4, 1974.
- [13] M. Reinikainen and T. Björklund, "PD6, a method for interdisciplinary product development training and education," in *SEFI 2008: Proceedings of the SEFI 36th conference on Quality Assessment, Employability and Innovation*, 2008.
- [14] M. Reinikainen and T. Björklund, "PD6, an idea generation and evaluation method," in *SEFI 2008: Proceedings of the SEFI 36th conference on Quality Assessment, Employability and Innovation*, 2008.

3. Modellizzazione psicoacustica e spazio sonoro



+ Simone Spagnol, Michele Geronazzo, Federico Avanzini
+ Enrico Marchetto, Federico Avanzini
+ Sergio Canazza, Antonio Rodà, Daniele Salvati
+ Mattia Schirosa, Jordi Janer, Stephan Kersten, Gerard Roma

Structural Modeling of Pinna-Related Transfer Functions for 3-D Sound Rendering

Simone Spagnol

spagnols@dei.unipd.it

Michele Geronazzo

Università di Padova

geronazz@dei.unipd.it

Federico Avanzini

avanzini@dei.unipd.it

ABSTRACT

This paper considers the general problem of modeling pinna-related transfer functions (PRTFs) for 3-D sound rendering. Following a structural approach, we present an algorithm for the decomposition of PRTFs into ear resonances and frequency notches due to reflections over pinna cavities and exploit it in order to deliver a method to extract the frequencies of the most important spectral notches. Ray-tracing analysis reveals a convincing correspondence between extracted frequencies and pinna cavities of a bunch of subjects. We then propose a model for PRTF synthesis which allows to control separately the evolution of resonances and spectral notches through the design of two distinct filter blocks. The resulting model is suitable for future integration into a structural head-related transfer function model, and for parametrization over anthropometrical measurements of a wide range of subjects.

1. INTRODUCTION

At the beginning of the last century, Lord Rayleigh's studies on the scattering of sound waves by obstacles gave birth to the extensive and still partially misunderstood field of 3-D sound. Within the context of his notable Duplex Theory of Localization [1], a commonly known formula that approximates the behaviour of sound waves diffracting around the listener's head provided indeed a first glance of the today-called head-related transfer function (HRTF). Alas, despite the importance and applicative potential of such a centenary theory, most of the efforts towards efficient modeling of HRTFs were spent in the last few decades only.

Throughout these years, low-order rational functions [2] and series expansions of HRTFs [3] were proposed as tools for HRTF modeling. Albeit the straightforward nature and intrinsic simplicity of both techniques, real-time HRTF modeling requires fast computations which cannot undergo the complexity of filter coefficients and weights, respectively. Oppositely, structural modeling [4] represents nowadays the ultimate alternative approach for real-time HRTF rendering: if we isolate the contributions of the user's head, pinnae and torso to the HRTF in different subcomponents,

each accounting for some well-defined physical phenomenon, then thanks to linearity we can reconstruct the global HRTF on-the-fly from a proper combination of all the considered effects. What we have is then a model which is both economical (if we assume that each physical phenomenon depends from few parameters) and well-suited to real-time implementations; as a further advantage, the intuitive nature of physical parameters enforces the chance to relate the model to simple anthropometrical measurements.

The present work exclusively deals with the contribution of the pinna to the HRTF. Even though head motion is perceptually a better discriminant, pinna cues are still of great importance in sound localization. A number of experiments have shown that, conversely to azimuth effects that can be reduced to simple binaural quantities, elevation effects - which are the result of a superposition of scattering waves influenced by a number of resonant modes - are basically monaural and heavily depend on the listener's anthropometry. Finding a suitable model for representing the pinna contribution to the HRTF (whose transfer function we commonly refer to as Pinna-Related Transfer Function - PRTF) is thus a crucial task, with the ultimate challenge in this direction being relating the model's parameters to easily obtainable anthropometric measurements on the user's pinnae. The resulting model, cascaded to a simple Head-and-Torso (HAT) model [5], will allow us to achieve a complete structural HRTF representation.

This paper lies its foundations on an iterative algorithm that separates resonance effects from pinna reflections in experimentally measured PRTFs. Moving from this start point, a method for extracting the frequencies of the most important notches is here developed, followed by a discussion on the possible relation between notch frequencies and anthropometry. Finally, a structural model of the pinna is proposed.

2. PREVIOUS WORKS

According to Batteau [6], high-frequency tones are typically reflected by the outer ear, as long as their wavelength is small enough compared to the pinna dimensions. Consequently, interference between the direct and reflected waves causes sharp notches to appear in the high-frequency side of the received signal's spectrum with a periodicity that is inversely proportional to the time delay of each reflection. Such observation led to a first rough double-path model of the pinna [7]. Unhappily, this model lacks the description of pinna resonant modes: as Shaw argued [8],

since pinna cavities act as resonators the frequency content of both the direct and the reflected sound waves is significantly altered. Batteau's model has accordingly been improved by Barreto *et al.*, with a new reflection structure [9] represented by four parallel paths cascaded to a low-order resonator block. Furthermore, the model parameters were associated to eight measured anthropometric features by means of multiple regression analysis [10]. The trouble is as well as providing no cloudless evidence of the physics behind the scattering phenomenon, the considered measures can only be acquired through the use of a 3-D laser scanner. In any case, these works surely endorse our final PRTF model's "resonance-plus-delay" architecture.

A different approach for reflection modeling, acting both in the time and frequency domains, was pursued by Raykar *et al.* [11]. Robust digital signal processing techniques are used here to extract the frequencies of the spectral notches due to the pinna alone: first the autocorrelation function of the HRIR's windowed LP residual is computed; then, frequencies of the spectral notches are found as the local minima of the group-delay function of the windowed autocorrelation. What's more, the authors advance a ray-tracing argument to attest that the so found spectral notches are related to the shape and anthropometry of the pinna. Specifically, knowing that the elevation-dependent temporal delay $t_d(\phi)$ between the direct and the reflected wave at the ear canal puts the point of reflection at a distance

$$d(\phi) = \frac{ct_d(\phi)}{2}, \quad (1)$$

where c is the speed of sound (approximately 343 m/s), and assuming the reflection coefficient to be positive, then each extracted frequency f_0 is considered as the first of a periodic series

$$f_n(\phi) = \frac{(2n+1)}{2t_d(\phi)} = \frac{c(2n+1)}{4d(\phi)}, \quad n = 0, 1, \dots, \quad (2)$$

in particular

$$f_0(\phi) = \frac{c}{4d(\phi)}. \quad (3)$$

The corresponding distance $d(\phi)$ was then projected onto the 2-D image of the pinna, resulting in a mapping consistent with reflections on the crus helias and concha wall.

Another important contribution on PRTF modeling was provided by Satarzadeh *et al.* [12]. In this work, PRTFs for elevation $\phi = 0^\circ$ are synthesized through a model composed of two second-order bandpass filters and one comb filter, which respectively approximate the two major resonances (Shaw's resonant modes 1 and 4) and one main reflection. The frequency of the comb filter's first tooth, f_0 , is estimated from the spacing of consecutive notches in the PRTF spectrum: consequently, if the filter takes the form $[1 + \rho \exp(-st_d)]$ (where ρ is the reflection coefficient), then the time delay between direct and reflected wave is calculated as

$$t_d = \frac{1}{2f_0} \quad (4)$$

if $\rho > 0$ (according to Raykar *et al.*), or as

$$t_d = \frac{1}{f_0} \quad (5)$$

if $\rho < 0$. Once the sign of the reflection coefficient is determined from the impulse response, the distances inferred from Eq. (1) put the point of reflection either at the back of the concha or at the edge of the rim. In addition, a cylindrical approximation of the concha is used with the purpose of directly parameterizing the resonances' coefficients. In conclusion, such a low-order anthropometry-based filter provides a good fit to the experimental PRTF in all cases where the pinna has an approximately cylindrical shaped concha and a structure with a dominant reflection area (concha or rim). However, besides considering solely the frontal direction of the sound wave, taking into account a single reflection appears as a limiting factor.

3. PRTF ANALYSIS

Taking the last two works described in the previous section as an inspiration and a "resonance-plus-delay" PRTF model as starting point, the main and final goal of our work is the construction of an essential multi-notch filter suitable for anthropometric parametrization. This obviously requires a PRTF analysis step. In order to analyze PRTFs, we consider measured HRIRs from the CIPIC database [13], a public domain database of high spatial resolution HRIR measurements at 1250 directions for 45 different subjects along with their anthropometry. We choose to investigate the behaviour of pinna features in subjects 010, 027, 134, and 165 in order to facilitate comparison with previous works on notch frequencies extraction (the same subjects' PRTFs were analyzed in [11]).

3.1 The Separation Algorithm

For purpose of analysis we focus on HRIRs sampled on the median plane, with elevation varying from -45° to 90° . As a matter of fact, since sensitivity of PRTFs to azimuth is weak [12], we roughly expect PRTFs to be elevation dependent only. Such an assumption makes the PRTF model suitable for all azimuths.

Knowing that the magnitude response of an earless head with respect to a sound source in the median plane is ideally flat if the head is modeled as a rigid sphere, the only preprocessing step we apply to obtain a raw estimate of the PRTF is windowing the corresponding HRIR using a 1.0 ms Hann window [11]. In this way, spectral effects due to reflections caused by shoulders and torso are removed from the PRTF estimate.

In order to isolate the spectral notches in the so built PRTFs we exploit an ad-hoc designed algorithm that returns an estimate of the separated resonant and reflective components. Figure 1 reports the complete flow chart of this analysis algorithm. The idea beyond it is to iteratively compensate the PRTF magnitude spectrum with an approximate multi-notch filter until no significant notches are left. Once convergence is reached (say at iteration i), the PRTF spectrum $H_{res}^{(i)}$ will contain the resonant component, while the combination $H_{refl}^{(i)}$ of the multi-notch filters will provide the reflective component. The algorithm's initial conditions heavily influence the final result; three parameters have to be chosen:

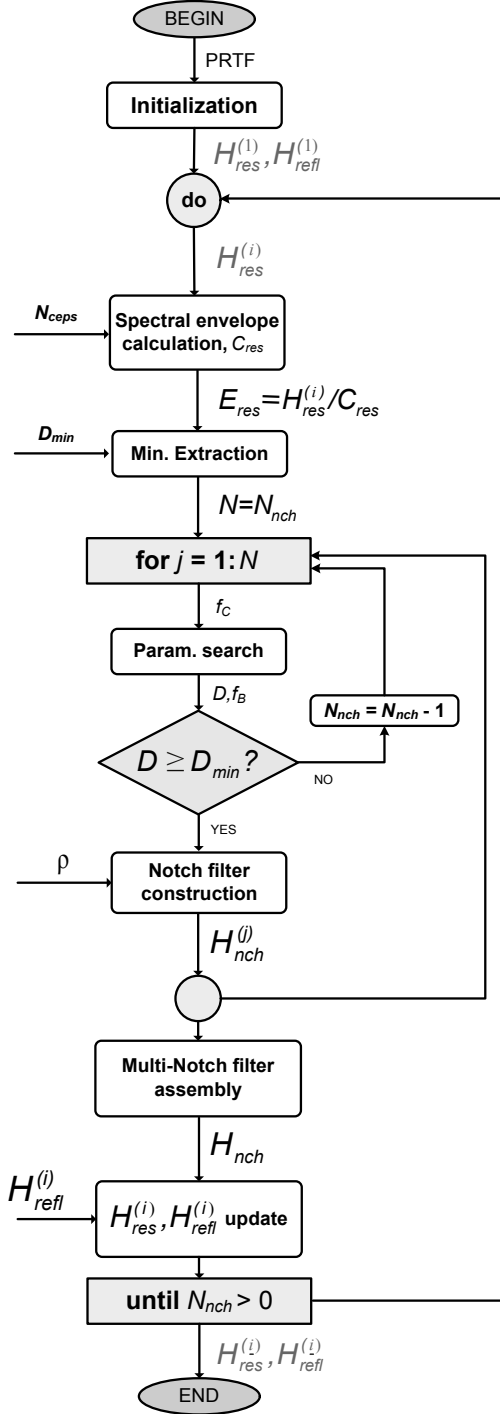


Figure 1. Flow chart of the analysis algorithm.

- N_{ceps} , the number of cepstral coefficients used for estimating the PRTF spectral envelope at each iteration;
- D_{min} , the minimum dB depth threshold for notches to be considered;
- ρ , the reduction factor for every notch filter bandwidth (its purpose will be discussed below).

Before entering the core of the algorithm, let $H_{res}^{(1)}$ match the PRTF and set $H_{refl}^{(1)}$ to 1. These two frequency responses will be updated at each iteration, resulting in $H_{res}^{(i)}$ and $H_{refl}^{(i)}$ at the beginning of the i -th iteration. If $N_{nch}^{(i)}$ is the number of “valid” notches algebraically identified at the end of it, the algorithm will terminate at iteration i if $N_{nch}^{(i)} = 0$, while $H_{res}^{(i)}$ and $H_{refl}^{(i)}$ will respectively contain the resonant and reflective components of the PRTF. As one may expect, both the number of iterations and the quality of our decomposition strongly rely on a good choice of the above parameters. For instance, choosing D_{min} too close to zero may lead to an unacceptable number of iterations; conversely, a high value of D_{min} could result in a number of uncompensated notches in the resonant part of the PRTF. In the following, we present the step-by-step analysis procedure on $H_{res}^{(i)}$, assuming that $N_{nch}^{(i-1)} > 0$. For the sake of simplicity, in the following the apex (i) indicating iteration number is dropped from all notation.

3.1.1 Residue computation

First, in order to extract properly the local minima due to pinna notches in the PRTF, the resonant component of the spectrum must be compensated for. To this end, the real cepstrum of H_{res} is calculated; then, by liftering the cepstrum with the first N_{ceps} cepstral coefficients and performing the FFT, an estimate of the spectral envelope of H_{res} is obtained, which we call C_{res} .

The parameter N_{ceps} must be chosen adequately, since it is crucial in determining the degree of detail of the spectral envelope. As N_{ceps} increases, the notches’ contribution is reduced both in magnitude and in passband while the resonance plot becomes more and more detailed. We experimentally found that the optimal number of coefficients that capture the resonant structure of the PRTF while leaving all the notches out of the spectral envelope is $N_{ceps} = 4$. This number also matches the maximum number of modes identified by Shaw which appear at one specific spatial location: for elevations close to zero, modes 1, 4, 5, and 6 are excited. Once C_{res} is computed, we subtract it from the dB magnitude of H_{res} and obtain the residue E_{res} .

3.1.2 Multi-notch filter parameter search

At this point E_{res} should present an almost flat spectrum with a certain number of notches. Parameter N_{nch} is first set to the number of local minima in E_{res} deeper than D_{min} , extracted by a simple notch picking algorithm. Our aim is to compensate each notch with a second-order notch filter, defined by three parameters: central frequency f_c , bandwidth f_B , and notch depth D .

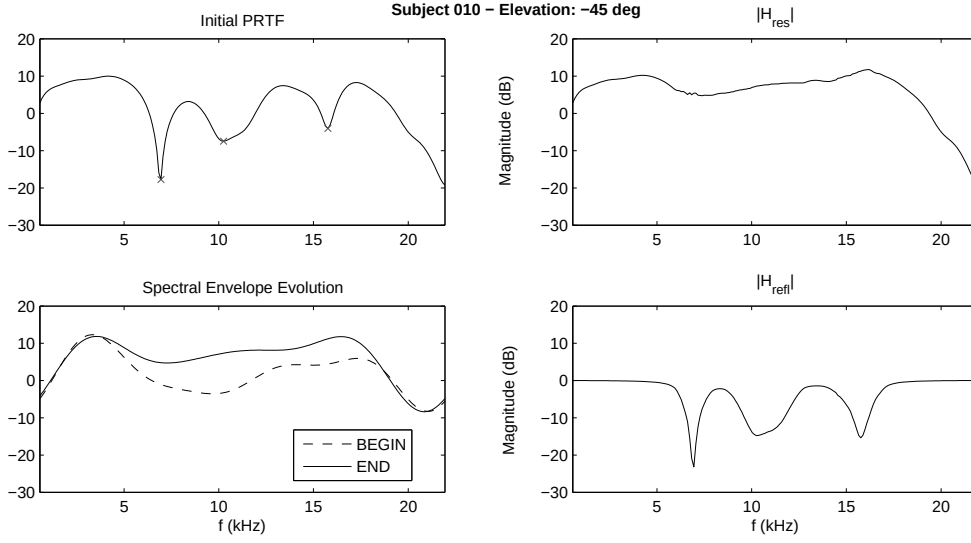


Figure 2. An example of the algorithm's evolution. The PRTF magnitude in the top left panel is decomposed into resonances (top right panel) and frequency notches (bottom right panel). The bottom left panel shows the evolution of the PRTF spectral envelope from the first iteration to convergence.

Consider the j -th local minimum. The central frequency of the corresponding notch filter f_C is immediately determined, while notch depth is found as $D = |E_{res}(f_C)|$. Computation of f_B is less straightforward. Indeed, f_B is calculated as the standard 3-dB bandwidth, i.e. $f_B = f_r - f_l$, where f_l and f_r are respectively the left and right +3 dB level points relative to f_C in E_{res} , except for the following situations:

1. if $D < 3$ dB, the 3-dB bandwidth is not defined. Then f_r and f_l are placed at an intermediate dB level, halfway between 0 and $-D$ in a linear scale;
2. if the local maximum of E_{res} immediately preceding (following) f_C does not lie above the 0-dB line while the local maximum immediately following (preceding) does, f_B is calculated as twice the half-bandwidth between f_C and f_r (f_l);
3. if both local maxima do not lie above the 0-dB line, we vertically shift E_{res} until the 0-dB level meets the closest of the two. Then, f_B is calculated as before except if the new notch depth is smaller than D_{min} in the shifted residue plot, in which case the parameter search procedure for the current notch is aborted and N_{nch} is decreased by one.

Note that case 1 may occur simultaneously with respect to case 2 or 3: in this situation, both corresponding effects are considered when calculating f_B .

3.1.3 Multi-notch filter construction

The so found parameters f_C , D , and f_B need to uniquely define a filter structure. To this end, we use a second-order

notch filter implementation of the form [14]

$$H_{nch}^{(j)}(z) = \frac{1 + (1+k)\frac{H_0}{2} + l(1-k)z^{-1} + (-k - (1+k)\frac{H_0}{2})z^{-2}}{1 + l(1-k)z^{-1} - kz^{-2}}, \quad (6)$$

where

$$k = \frac{\tan(\pi \frac{f_B}{f_s}) - V_0}{\tan(\pi \frac{f_B}{f_s}) + V_0}, \quad (7)$$

$$l = -\cos(2\pi \frac{f_C}{f_s}), \quad (8)$$

$$V_0 = 10^{\frac{D}{20}}, \quad (9)$$

$$H_0 = V_0 - 1, \quad (10)$$

and f_s is the sampling frequency. Using such an implementation allows us to fit our parameters directly to the filter model. Clearly, not every combination of the three parameters is accurately approximated by the second-order filter: if the notch to be compensated is particularly deep and sharp, the filter will produce a shallower and broader notch, having a center frequency which is slightly less than f_C .

Although moderate frequency shift and attenuation is not detrimental to the estimation algorithm (an underestimated notch will be fully compensated through the following iterations), an excessive notch bandwidth could lead to undesired artifacts in the final resonance spectrum. Here is where parameter ρ comes into play: if we divide f_B by $\rho > 1$, the new bandwidth specification will produce a filter whose notch amplitude will be further reduced, allowing us to reach a smaller bandwidth. Typically, in order to achieve a satisfactory trade-off between the size of ρ and the number of iterations, we set it to 2.

Consequently, the parameters to be fed to the filter are $(f_C, D, f_B/\rho)$, yielding coefficients vectors $\mathbf{b}^{(j)}$ and $\mathbf{a}^{(j)}$ for $H_{nch}^{(j)}$. We iterate the parameter search and notch filter construction procedures for all N_{nch} notches. In order to build the complete multi-notch filter H_{nch} ,

$$H_{nch}(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{a_0 + a_1 z^{-1} + a_2 z^{-2}} = \prod_{j=1}^{N_{nch}} H_{nch}^{(j)}(z), \quad (11)$$

it is now sufficient to convolve all the coefficient vectors computed during iteration i :

$$\mathbf{b} = [b_0, b_1, b_2] = \mathbf{b}^{(1)} * \mathbf{b}^{(2)} * \dots * \mathbf{b}^{(N_{nch})} \quad (12)$$

$$\mathbf{a} = [a_0, a_1, a_2] = \mathbf{a}^{(1)} * \mathbf{a}^{(2)} * \dots * \mathbf{a}^{(N_{nch})}. \quad (13)$$

Finally, before considering the next iteration, we must update the global multi-notch filter $H_{refl}^{(i+1)} = H_{refl}^{(i)} \cdot H_{nch}$ and compensate the PRTF by applying $H_{res}^{(i+1)} = H_{res}^{(i)} / H_{nch}$.

3.1.4 Algorithm evolution example

Figure 2 illustrates the algorithm's evolution for a particular PRTF. The specific choice of the initial parameters was $N_{ceps} = 4$, $D_{min} = 0.1$ dB, and $\rho = 2$. The top left panel illustrates Subject 010 PRTF for an elevation of -45 degrees. The bottom left panel reports the spectral envelope evolution, where we can see how interfering spectral notches negatively influence the initial estimate. The panels on the right represent the resonant (H_{res}) and reflective (H_{refl}) parts of the PRTF at the end of the algorithm.

Consider the range where acoustic effects of the pinna are relevant, i.e. the range from 3 to 18 kHz approximately. Figure 2 shows that inside such range the algorithm has produced a realistic decomposition: the gain of the reflective component is unitary outside the notch regions, while the peaks appearing in the resonant component have a good correspondence to Shaw's modes (this point is further discussed in the next section). Outside the relevant range for the pinna, there is a sharp gain decrease in the resonant part and further imperfections that appear for different subjects and elevations. Nevertheless, this is not a problem as long as we consider the pinna contribution to the HRTF alone.

The behavior exemplified in figure 2 is observed for different elevations and subjects too.

3.2 Resonances

We now discuss the PRTF features identified by the decomposition carried out through the separation algorithm. From the 3-D plots in Figure 3 we can study how the resonances' contribution for Subjects 010 and 165 varies throughout all available elevations. The center frequency of each resonance was extracted with the help of an identification system based on a sixth-order ARMA model [15] and spatially tracked along elevation, resulting in the dotted tracks superposed on the plots.

We can easily identify two major hot-colored areas in these plots. The first one, centered around 4 kHz, appears to be very similar amongst subjects since it spans all elevations. One may immediately notice that this area includes

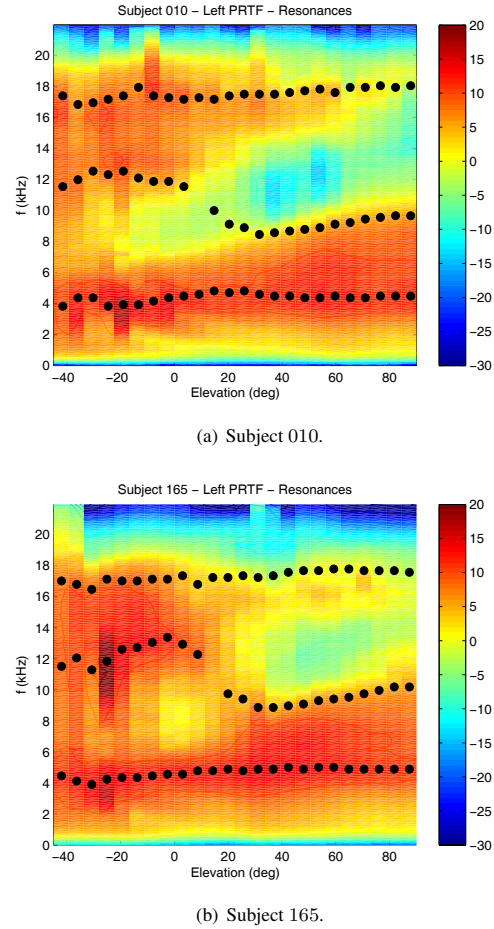


Figure 3. Resonance plots for different elevations.

Shaw's omnidirectional mode 1. The resonance's bandwidth appears to increase with elevation; however, knowledge of pinna modes implies that a second resonance is likely to interfere within this frequency range, specifically Shaw's mode 2 (centered around 7 kHz with a magnitude of 10 dB). On the other hand, the second hot-colored area differs both in shape and shade amongst subjects. Still it is most prominent at low elevations between 12 and 18 kHz, a frequency range which is in general agreement with Shaw's horizontal modes 4, 5, and 6.

Note that the higher resonance may be perceptually irrelevant since it lies near the upper limit of the audible range. In addition, since the resonances at 12 and 7 kHz are excited in mutually exclusive elevation ranges, we may look forward to a double-resonance filter design.

3.3 Notches

Similarly to the resonance plots, those in Figure 4 represent the frequency notches' contribution for Subjects 027 and 134. As expected, reflection patterns strongly depend on elevation and pinna shape. While PRTFs generally exhibit poor notch structures when the source is above the head, as soon as the elevation angle decreases the number and

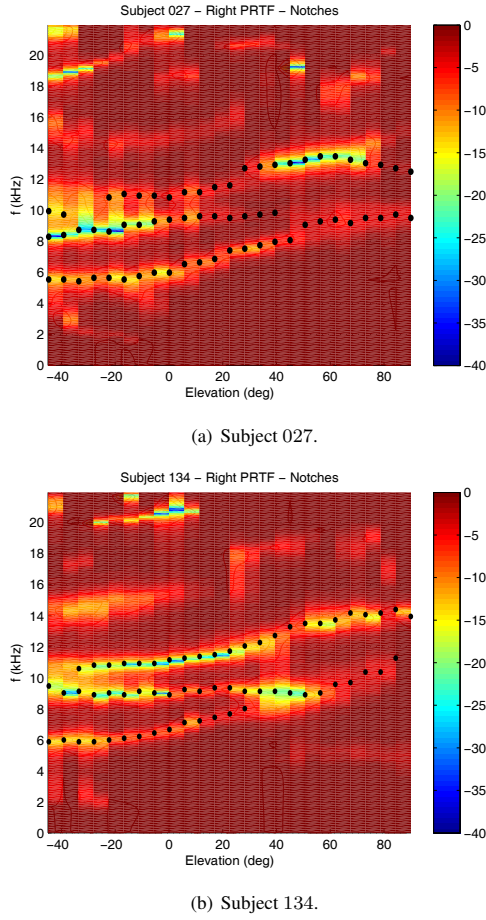


Figure 4. Spectral notch plots for different elevations.

depth of frequency notches grows to an extent that varies among subjects.

However, several analogies can be noticed here too. In order to investigate such common trends, we inherit an analysis tool that is widely used in the field of sinusoidal modeling, specifically the McAulay-Quatieri partial tracking algorithm (see [16] for details), to track the most prominent notches' patterns along all elevations. Originally, this algorithm was used to group sinusoidal partials (extracted through a peak detection algorithm) along consecutive temporal windows according to their spectral location. We implemented the original version [16] of the algorithm; obviously, since in our case elevation dependency replaces temporal evolution and spectral notches take the role of partials, we call it "notch tracking" algorithm. The notch detection step simply locates all of the local minima in the reflective component's spectrum, while the matching interval for the notch tracking procedure is set to $\Delta = 3$ kHz.

Since it is preferable to restrict our attention to the frequency range where reflections due to the pinna alone are most likely seen, and ignore notches which are overall feeble, two post-processing steps are performed on the obtained tracks:

- delete the tracks which are born and die outside the range 4 – 14 kHz;
- delete the tracks that do not present a notch deeper than 5 dB.

The outputs of the notch tracking algorithm are the dotted tracks superposed on the plots in Figure 4. Results are definitely akin to those found in [11] with the use of an elaborated DSP-based algorithm. Three major tracks are seen for both subjects, whereas the shorter track in Subject 027's plot very probably represents the continuation of the missing track at those elevations. Reasonably, the gap between tracks is caused by the algorithm's impossibility of locating proper minima in that region (due e.g. to superposition of two different notches or the presence in the magnitude plot of valleys which are not notch-like). However, the three longer tracks suggest that similar reflection patterns occur in different PRTFs.

4. REFLECTIONS AND ANTHROPOMETRY

We now move to a possible explanation of the physical mechanism lying behind the production of frequency notches in the PRTF spectrum. As already pointed out, we relate each major notch to a distinct reflection, assuming it to be the first and most marked notch of a periodic series.

4.1 Reflection coefficient sign

Reflection models usually assume all reflection coefficients to be positive. If this were the case, the extra distance travelled by the reflected wave with respect to the direct wave must be equal to half a wavelength in order for destructive interference to occur, which translates into spectral notches in the frequency domain (see Eq. (4)). This was the assumption taken by [11] when tracing reflection points over pinna images based on the extracted notch frequencies.

Nevertheless, Satarzadeh [17] drew attention to the fact that the majority of CIPIC subjects exhibit a clear negative reflection in the HRIR. He motivated this result by hypothesizing a boundary created by an impedance discontinuity between the pinna and air which could produce its own reflection, reversing the phase of the wave. In this latter case, destructive interference would not appear for half-wavelength delays anymore, yet only for full-wavelength delays (see Eq. (5)).

4.2 Ray tracing

Following Satarzadeh's hypothesis, we choose to use the negative reflection assumption in establishing a relation between notches and pinna geometry through a simple ray-tracing procedure, very similar to the one described in [11].

Right pinna images are taken from the CIPIC database and uniformly rescaled in order to match parameters d_5 (pinna height) and d_6 (pinna width) [13]. The distance of each reflection point with respect to the entrance of the ear canal is calculated through Eqs. (1) and (5), leading to the relation

$$d(\phi) = \frac{c}{2f_0(\phi)}, \quad (14)$$

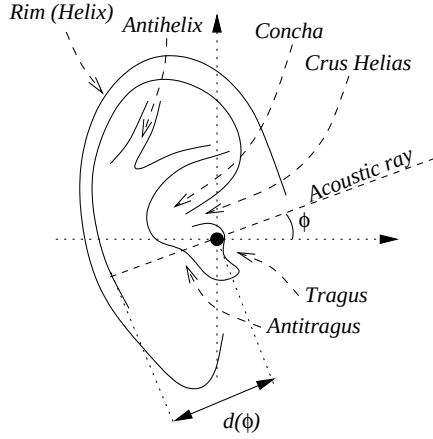


Figure 5. Anatomy of the pinna.

where $f_0(\phi)$ represents the frequency of the current notch at elevation ϕ . The negative reflection coefficient assumption causes distances to be approximately doubled with respect to those calculated in [11]. Then, if we consider the 2-D polar coordinate system illustrated in Figure 5 having the right ear canal entrance as origin, each notch is mapped to the point $(d(\phi), \pi + \phi)$.

Results for subjects 010, 027, 134, and 165 are reported in Figure 6. For all these subjects, the so-obtained mapping shows a high degree of correspondence between computed reflection points and pinna geometry. One can immediately notice that the track nearest to the ear canal very closely follows the concha wall of each subject for all elevations, except for a couple of cases:

- at low elevations, displacement of points may be caused by the little extra distance needed by the wave to pass over the crus helias;
- Subject 010's track disappears at around $\phi = 60^\circ$ probably because of the insufficient space between tragus and antitragus that causes the incoming wave

to reflect outside the concha.

The intermediate track falls upon the area between concha and rim, with variable length among subjects:

- in the case of subjects 010 and 165 the track is faint and probably due to the antihelix;
- conversely, subjects 027 and 134 present a longer and deeper track, that we visually associate to a reflection on the rim's edge.

Finally, the furthest track follows the shape of the rim and is likely to be associated to a reflection in the inner wall of it, except for Subject 010 whose reflection occurs at the rim's edge. A strong evidence that validates the track's connection to the rim structure lies in the fact that the rim terminates in the vicinity of the point where the track disappears.

4.3 Model fitting to anthropometry

Further refinements should be applied to the above preliminary analysis for a more detailed account of the reflection structures of a vast test bed of subjects to be performed, in particular the use of a 3-D model of the pinna that allows to investigate its horizontal section. As a matter of fact, in most cases the pinna structure does not lie on a parallel plane with respect to the head's median plane, especially in subjects with protruding ears. Hence plotting distances on the side-view images should take into account the displacement caused by the flare angle of the pinna.

Nevertheless, our preliminary analysis has revealed a satisfactory correspondence between computed reflection points and reflective structures over the pinna. This opens the door for a very attractive approach to the parametrization of the structural PRTF model based on individual anthropometry. Indeed, given a 2-D image or a 3-D reconstruction of the user's pinna, one can easily trace the contours of the concha wall, antihelix and rim, compute each contour's distance with respect to the ear canal for all elevations, and extrapolate the notch frequencies by reversing

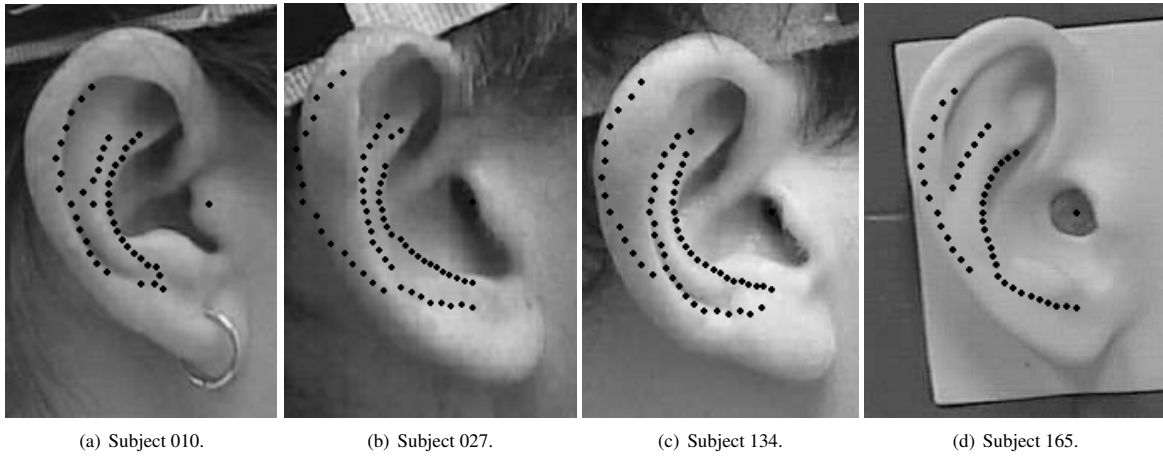


Figure 6. Reflection points on four CIPIC subjects' right pinnae.

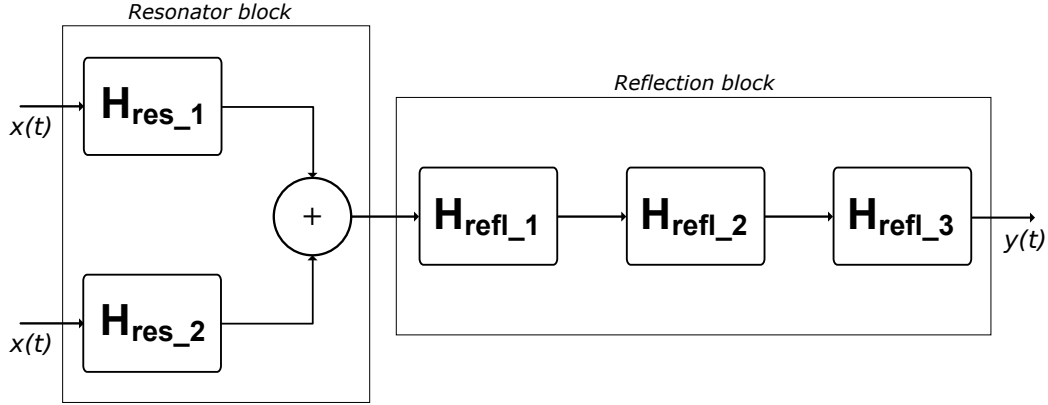


Figure 7. General model for the reconstruction of PRTFs.

Eq. 5. Obviously, since notch depth strongly varies within subjects and elevations, the reflection coefficient must also be estimated for each point. This problem theoretically requires strong physical arguments; alternatively, psychoacoustical criteria could be used in order to evaluate the perceptual relevance of notch depth, and potentially simplify the fitting procedure.

5. A STRUCTURAL MODEL OF THE PINNA

The information gathered from the outputs of the decomposition and notch tracking algorithms allows to model the PRTF with two resonances and three spectral notches. As Figure 7 depicts, our final aim is to design two distinct filter blocks, one accounting for resonances and one for reflections. Clearly, in order to reach complete control of the filter parameters, full parametrization of the model on anthropometrical measurements is needed. Hence for the moment we shall present the PRTF re-synthesis procedure driven by the outputs of the two above algorithms.

5.1 Filter design

In Section 3.2 we have shown that a PRTF at one specific elevation includes two main resonances in the frequency range of interest for the pinna. It is then possible to approximate the effective resonances by deducing center frequency f_C and magnitude G of each resonance from the dotted tracks and directly using the so found parameters to design two second-order peak filters with fixed bandwidth $f_B = 5$ kHz of the form [18]

$$H_{res}(z) = \frac{V_0(1-h)(1-z^{-2})}{1+2dhz^{-1}+(2h-1)z^{-2}}, \quad (15)$$

where

$$h = \frac{1}{1 + \tan(\pi \frac{f_B}{f_s})}, \quad (16)$$

$$d = -\cos(2\pi \frac{f_C}{f_s}), \quad (17)$$

$$V_0 = 10^{\frac{G}{20}}, \quad (18)$$

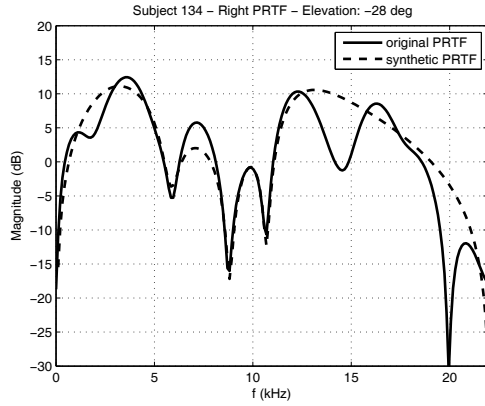
and f_s is the sampling frequency. A posteriori analysis of the synthesized resonances has revealed that PRTFs for high elevations only need the first resonance to be synthesized, being the second very close to it. We thus choose to bypass the second resonant filter when $\phi \geq 20^\circ$.

Similarly, for what concerns the reflection block, we feed the center frequency f_C , notch depth D , and bandwidth f_B parameters coming from the notch tracking algorithm to three second-order notch filters of the form in Eq. 6, each accounting for a different spectral notch. The three notch filters must be placed in series and cascaded to the parallel of the two peak filters, resulting in an eighth-order global filter.

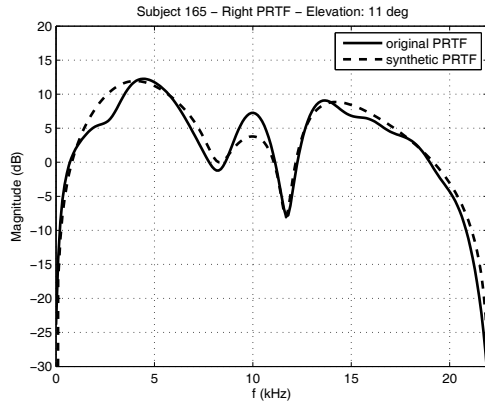
5.2 Results

Figure 8 reports the comparison between original and re-synthesized PRTF magnitudes for three distinct subjects, each at a different elevation. Adherence rate to the original PRTFs is overall satisfactory in the frequency range up to 14 kHz. Still, several types of imperfections need to be adjusted: as a first example, deep frequency notches that appear at low elevations complicate the notch filter design procedure. In point of fact, if the notch to be approximated is particularly deep and sharp, the second-order filter will produce a shallower and broader notch whose bandwidth may interfere with adjacent notches, resulting in underestimating the PRTF magnitude response in the frequency interval between them. Figures 8(a) and 8(b) show this behaviour around 7.5 and 10 kHz, respectively. Using a filter design procedure which forces to respect the notch bandwidth specification during re-synthesis would grant a better rendering of resonances, at the expense of worsening notch depth accuracy.

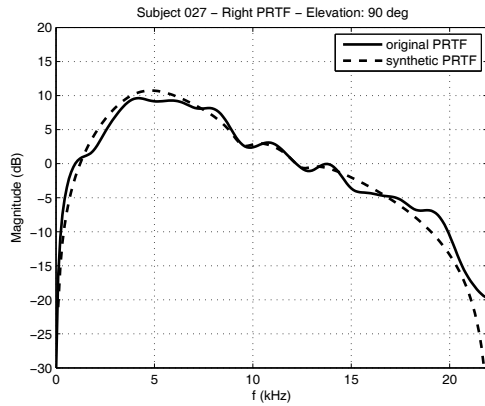
The absence of modeled notches over the upper frequency threshold is another cause of imprecision. For instance, Figure 8(a) presents an evident mismatch between original and modeled PRTF just after the 12.5-kHz peak, due to the cut of the frequency notch at 14.5 kHz. This problem may be corrected by increasing the 14-kHz threshold in order to take into account a higher number of notches. However, being the psychoacoustic relevance of this fre-



(a) Subject 134, elevation -28° .



(b) Subject 165, elevation 11° .



(c) Subject 027, elevation 90° .

Figure 8. Original vs Synthetic PRTF plots.

quency range relatively low, the effective weight of the mismatch is reduced.

Last but not least, resonance modeling may bring approximation errors too. In particular, the possible presence of non-modeled interfering resonances and the fixed-bandwidth specification both represent a limitation to the re-synthesis procedure. Furthermore, center frequencies extracted by the ARMA identification method mentioned in Section 3.2 do not always coincide with peaks in the PRTF. Thus a stronger criterion for extracting the main parameters of each resonance is needed. Nevertheless, the approximation error seems to be negligible in all those cases where resonances are distinctly identifiable in the PRTF.

In conclusion, the above presented re-synthesis model appears to be overall effective, especially for PRTFs which clearly show one or two main resonant modes and moderately deep notches. Figure 8(c) supports this assertion.

6. CONCLUSIONS AND FUTURE WORK

In this paper we presented an approach for structural PRTF modeling, which exploits an algorithm that separates the resonant and reflective parts of the PRTF spectrum. We used such decomposition to re-synthesize the original PRTF through a low-order filter model, whose results show an overall suitable approximation. In a parallel manner, our attempt towards the explanation of the scattering process resulting in the most important spectral notches in the PRTF provided visually convincing results. Besides improving the synthesis step, ongoing and future work includes understanding of the reflection coefficient and relating the resonant component of the PRTF to anthropometry.

7. REFERENCES

- [1] J. W. Strutt, "On our perception of sound direction," *Philosophical Magazine*, vol. 13, pp. 214–232, 1907.
- [2] E. C. Durant and G. H. Wakefield, "Efficient model fitting using a genetic algorithm: pole-zero approximations of HRTFs," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 1, pp. 18–27, 2002.
- [3] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1637–1647, 1992.
- [4] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998.
- [5] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The use of head-and-torso models for improved spatial sound synthesis," in *Proc. 113th Convention of the Audio Engineering Society*, (Los Angeles, CA, USA), 2002.
- [6] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. London. Series B, Biological Sciences*, vol. 168, pp. 158–180, August 1967.

- [7] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, vol. 63, pp. 1152–1165, April 1978.
- [8] E. A. G. Shaw, *Binaural and Spatial Hearing in Real and Virtual Environments*, ch. Acoustical features of human ear, pp. 25–47. Mahwah, NJ, USA: R. H. Gilkey and T. R. Anderson, Lawrence Erlbaum Associates, 1997.
- [9] K. J. Faller II, A. Barreto, N. Gupta, and N. Rishe, "Time and frequency decomposition of head-related impulse responses for the development of customizable spatial audio models," *WSEAS Transactions on Signal Processing*, vol. 2, no. 11, pp. 1465–1472, 2006.
- [10] N. Gupta, A. Barreto, and M. Choudhury, "Modeling head-related transfer functions based on pinna anthropometry," in *Proc. of the Second International Latin American and Caribbean Conference for Engineering and Technology (LACCEI)*, (Miami, FL, USA), 2004.
- [11] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.*, vol. 118, pp. 364–374, July 2005.
- [12] P. Satarzadeh, R. V. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proc. 122nd Convention of the Audio Engineering Society*, (Vienna, Austria), May 5-8 2007.
- [13] R. V. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (New Paltz, New York, USA), pp. 1–4, 2001.
- [14] U. Zölzer, ed., *Digital Audio Effects*. New York, NY, USA: J. Wiley & Sons, 2002.
- [15] P. A. A. Esquef, M. Karjalainen, and V. Välimäki, "Frequency-zooming ARMA modeling for analysis of noisy string instrument tones," *EURASIP Journal on Applied Signal Processing: Special Issue on Digital Audio for Multimedia Communications*, no. 10, pp. 953–967, 2003.
- [16] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.
- [17] P. Satarzadeh, "A study of physical and circuit models of the human pinnae," Master's thesis, University of California Davis, 2006.
- [18] S. J. Orfanidis, ed., *Introduction To Signal Processing*. Prentice Hall, 1996.

Modellazione fisica della glottide e inversione acustico-articolatoria

E. Marchetto, F. Avanzini

Dip. di Ingegneria dell'Informazione

Università di Padova

{marchet1|avanzini}@dei.unipd.it

SOMMARIO

Questo lavoro presenta una tecnica per la stima del modello a due masse della corda vocale a partire da un dato flusso glottale tempo-variante. Il modello a due masse è specificato da un certo numero di parametri meccanici di basso livello, calcolati in funzione di quattro parametri articolatori (livelli di attivazione di tre muscoli laringali e pressione subglottale). Le forme d'onda del flusso glottale, sintetizzate dal modello, sono caratterizzate da un insieme di parametri acustici per la quantificazione della sorgente vocale. Misurando un flusso glottale di riferimento viene data una sequenza di parametri acustici e, impiegando la programmazione dinamica e l'interpolazione con reti RBF (Radial Basis Function Networks), si derivano i parametri di attivazione muscolare che portano alla risintesi del flusso glottale di partenza.

1. INTRODUZIONE

Un problema di ricerca aperto nella modellazione fisica delle corde vocali a bassa dimensionalità è la relazione tra i parametri dei modelli ed i parametri acustici relativi alla *voice quality*.

Un recente lavoro [1] ha studiato la sensibilità dei parametri acustici del flusso alla variazione dei parametri fisici di un modello a due masse, dando indicazioni sul comportamento del modello per la simulazione delle diverse *voice qualities*. I parametri del modello (di basso livello: masse, costanti elastiche, ecc.) non sono però controllati in modo volontario dal parlatore: è necessario uno spazio di controllo fisiologicamente motivato per il modello. Una questione affine è il cosiddetto "problema inverso", ovvero il problema di stimare parametri di controllo tempo-varianti da usare come ingresso al modello fisico, così da risintetizzare un segnale acustico *target*. Questo implica l'inversione di un sistema dinamico non lineare con un elevato numero di parametri; la soluzione, inoltre, può non essere univoca. A questo proposito, per evitare la non-univocità, è possibile lavorare su sequenze temporali di *frames* acustici e stimare i parametri articolatori mediante la minimizzazione di una funzione di costo che includa una componente di "sforzo articolatorio". Questo approccio è stato applica-

to in [2] alla soluzione del problema inverso relativo ad un modello articolatorio del tratto vocale.

Questo articolo presenta una procedura per la stima di un modello a due masse delle corde vocali [3] a partire dai parametri acustici, tempo-varianti, di un flusso glottale target; il modello è specificato da un vasto numero di parametri fisici di basso livello. Questi parametri fisici, calcolati da un livello aggiuntivo di modellazione, vengono ottenuti come funzione di quattro parametri articolatori (tre livelli di attivazioni di muscoli laringali e la pressione subglottale) [4]. Le forme d'onda dei flussi glottali sintetizzati dal modello sono caratterizzate da un insieme di parametri acustici: frequenza fondamentale F_0 , *open quotient* OQ, *speed quotient* SQ, *return quotient* RQ, *normalized amplitude quotient* NAQ [5], ecc. Questi vengono riconosciuti in letteratura come una tipica quantificazione della sorgente vocale [6].

Esistono quindi tre distinti spazi di parametri, legati tra loro: articolatorio, fisico ed acustico. Questo lavoro affronta il problema della mappatura dei parametri acustici nelle loro controparti articolatorie. *Frames* temporali del segnale di flusso glottale vengono caratterizzati mediante sequenze di parametri acustici; viene poi sviluppata una metodologia per derivare le corrispondenti sequenze di parametri articolatori usando tecniche di programmazione dinamica. La procedura è ulteriormente migliorata usando le *Radial Basis Function Networks* (RBFN, reti neurali a funzioni base radiali) per interpolare i punti dello spazio articolatorio. I risultati ottenuti mostrano che il modello fisico, controllato con i parametri stimati, è in grado di risintetizzare un segnale di flusso target con buona accuratezza.

La Sezione 2 descrive il modello fisico usato nel presente lavoro mentre la Sez. 3 dettaglia le tecniche usate per stimare il modello a partire da una segnale glottale tempo-variante target. I risultati, le limitazioni e le lacune dell'approccio proposto vengono discusse in Sez. 4

2. IL MODELLO FISICO

L'analisi sviluppata nelle prossime sezioni è basata sul modello a due masse presentato in [3] e raffigurato in Fig. 1. Il modello assume il flusso, dalla regione subglottale al *punto di separazione* z_s nella glottide, come unidimensionale, quasi-stazionario, privo di attriti e incompressibile; nella glottide avviene la separazione del flusso e la formazione del getto libero. Non è prevista alcuna risalita di pressione all'uscita della glottide. Il punto di separazione z_s si ha quando l'area glottale $a(z)$ supera il suo valore mini-

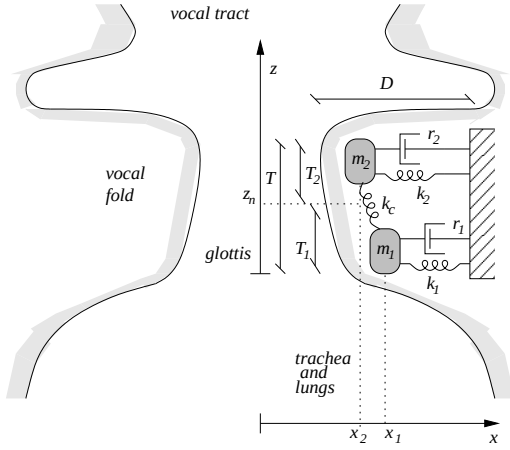


Figura 1. A destra: diagramma della corda vocale, trachea e tratto vocale supraglottale; a sinistra: modello a due masse della corda vocale.

mo di una certa quantità (10 – 20%) [3]. Introducendo una *costante di separazione* s (nell'intorno 1.1 – 1.2) la separazione si ha quando l'area glottale assume il valore $a_s = \min(sa_1, a_2)$.

Il tratto vocale è modellato come un carico ineritivo. La colonna d'aria in esso presente, assumendo frequenze fondamentali molto più basse del primo formante, agisce approssimativamente come una massa accelerata in modo unitario; la pressione all'ingresso del tratto vocale può essere scritta come $p_v(t) = Ru(t) + I\dot{u}(t)$, dove R, I sono rispettivamente la resistenza e l'inerzia di ingresso. I valori di R, I sono dati da [7]. Essendo un sistema del primo ordine, questo modello non tiene conto delle risonanze del tratto vocale; i suoi effetti più rilevanti sull'oscillazione delle corde vocali vengono comunque descritti con sufficiente accuratezza, in particolare l'abbassamento della soglia di pressione per l'oscillazione [7].

I parametri fisici di basso livello (masse, costanti elastiche, ecc.) non sono controllati dal parlatore: è necessario quindi uno spazio di controllo fisiologicamente motivato, che richiede di stabilire una mappatura tra la fisiologia (attivazioni muscolari) e la fisica (parametri del modello a due masse). Un insieme di regole empiriche, derivate da [8], sono state usate in [4] per il controllo del modello fisico. Le regole legano la geometria delle corde vocali ai livelli di attivazione di tre muscoli: cricotiroideale (a_{CT}), tiroaritenoidale (a_{TA}) e cricoaritenoidale laterale (a_{LC}). Si assume che questi livelli siano normalizzati nell'intorno $[0, 1]$. In questo articolo, inoltre, consideriamo la pressione subglottale p_s . In conclusione il modello fisico è completamente controllato da un insieme di quattro *parametri articolatori*: $a_{CT}, a_{TA}, a_{LC}, p_s$.

3. STIMA DEL MODELLO

3.1 Il codebook articolatorio

Il primo passo della procedura di stima consiste nel definire e popolare un *codebook diretto*, in cui ogni vettore di

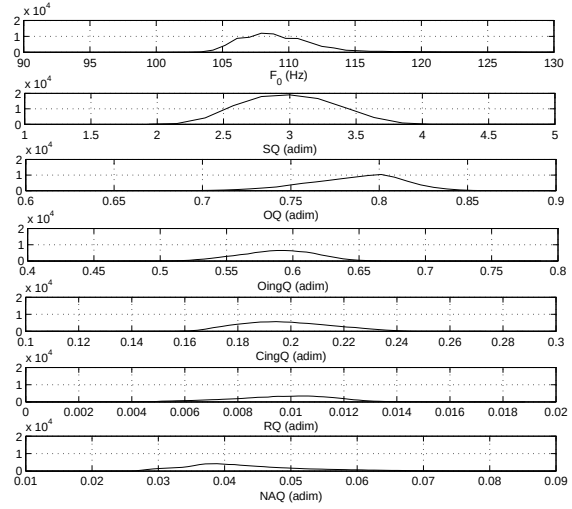


Figura 2. Distribuzione dei parametri acustici, visti singolarmente, nel codebook diretto.

parametri articolatori $[a_{CT}, a_{TA}, a_{LC}, p_s]$ è una “chiave” associata con uno ed uno solo vettore di parametri acustici. A questo scopo è stato condotto un grande numero di simulazioni numeriche del modello a due masse, variandone i parametri articolatori in una griglia densa di valori. Per ogni simulazione i parametri acustici di rilievo del flusso sintetizzato sono stati misurati mediante il toolkit APARAT [9].

Il codebook diretto usato in questo lavoro è stato ottenuto con una griglia in cui a_{CT} e a_{TA} variano nell'intorno $[0, 1]$ con passo fisso 0.05, mentre l'intorno di a_{LC} è $[0.25, 0.5]$ con un passo di 0.025 (la fonazione è ottenibile solo in questa regione). p_s varia nell'intorno $[500, 1500]$ Pa con un passo fissato di 50 Pa. Il codebook risultante contiene 86125 coppie di vettori articolatori/acustici; la Fig. 2 mostra la distribuzione dei 7 parametri acustici nel codebook.

3.2 Inversione del codebook ed accesso dinamico

Per risolvere il problema inverso il codebook diretto è stato invertito, ottenendo il *codebook inverso*. Quest'ultimo però manifesta il problema della non-univocità, ovvero un medesimo vettore acustico può essere associato a uno o più vettori articolatori. Affrontiamo il problema lavorando, piuttosto che su singoli vettori, su sequenze temporali di vettori acustici. Queste possono essere ottenute, ad esempio, analizzando un flusso glottale tempo-variante per sezioni (*frame-by-frame*). Data una sequenza di vettori acustici x_k vogliamo ottenere una sequenza “ottima” di vettori articolatori v_k^j nel codebook inverso: come già spiegato, ogni x_k è in principio associato a molti vettori candidati v_k^j a causa del problema di non-univocità. In particolare, eseguiamo una ricerca nello spazio acustico del codebook inverso per trovare i vettori più vicini (distanza euclidea) ai dati x_k ; i vettori v_k^j sono quindi i vettori articolatori associati ai vettori acustici del codebook più vicini agli x_k .

La sequenza ottima di parametri articolatori è ottenuta minimizzando una *funzione di costo* formata da tre termini. Un termine *acustico* tiene conto della distanza euclidea tra x_k e la sua versione discretizzata nello spazio acustico del codebook (i vettori trovati mediante ricerca). Un termine *articolatorio* minimizza la distanza euclidea tra v_k^j e v_{k-1}^j , ovvero tra tutte le coppie di vettori articolatori *consecutivi nel tempo*. Questo è il termine chiave della procedura, che permette di ottenere variazioni fluide dei parametri: esso minimizza lo “sforzo articolatorio”, in accordo con il comportamento fisiologico dei muscoli. Un termine di *accumulazione* estende il dominio della funzione di costo all’intera sequenza di input, così da garantire l’ottimalità della sequenza articolatoria in senso globale.

La funzione di costo (semplificata) è:

$$f(v_k^j) = \min_{\gamma, \delta} [\tau_1 \|x_k - c_k^\delta\|^2 + \tau_2 \|v_k^j - v_{k-1}^\gamma\|^2 + f(v_{k-1}^\gamma)]$$

dove $\tau_{1,2}$ sono i pesi per i termini acustico ed articolatorio, rispettivamente; c_k^δ sono i vettori acustici discretizzati più vicini agli x_k . Le tecniche di programmazione dinamica sono gli strumenti essenziali per la minimizzazione della funzione di costo: in particolare il termine di accumulazione porterebbe ad una complessità esponenziale, che però è evitata usando questo approccio; la complessità computazionale rimane quindi polinomiale.

3.3 Clustering del codebook e interpolazione con RBFN

Un problema della procedura proposta è che i vettori target x_k tipicamente non sono presenti nel codebook inverso, che è discretizzato; ogni v_k^j trovato non è quindi associato con x_k , ma solo con un vettore simile (vicino) ad x_k . Le limitazioni del codebook discreto possono essere superate interpolando lo spazio articolatorio; questo permette di calcolare i vettori articolatori associabili esattamente ad ogni dato x_k .

L’interpolazione usa RBFN (Radial Basis Function Networks, reti a funzione base radiale) [10]. Dal momento che le RBFN interpolano solo funzioni e non possono gestire multimappe, il codebook inverso deve essere manipolato in modo da evitare il problema di non-univocità. Abbiamo sviluppato un innovativo algoritmo che suddivide il codebook in insiemi (*clusters*) acustici ed in sottoinsiemi (*subclusters*) articolatori; ogni cluster è associato ad uno o più subclusters. L’algoritmo garantisce che per tutti i vettori acustici in un dato cluster ci sarà solo un (o nessun) vettore articolatorio in ogni subcluster associato. In questo modo in ogni subcluster il codebook, così suddiviso, dà una mappatura univoca, necessaria per il corretto funzionamento delle RBFN.

L’algoritmo di *clustering* prima suddivide lo spazio acustico in clusters C_i usando una tecnica standard. Vengono generati dei vettori acustici casuali, tanti quanti il numero di clusters desiderato, che seguito vengono variati con una procedura iterativa [11], diventando così i centroidi dei clusters. I centroidi vengono ripetutamente spostati in modo tale che la somma delle distanze tra ogni centroide ed i vettori a lui associati (ovvero associati al cluster del centroide) sia minimizzata. I clusters C_i sono infine costituiti

associando ciascun vettore acustico al centroide più vicino. Per ottenere una distribuzione uniforme dei vettori in ogni cluster, la procedura iterativa è applicata in due passi; inoltre, per assicurare un certo grado di sovrapposizione tra i clusters (necessario per il corretto funzionamento dell’ottimizzazione), i vettori più vicini al confine tra due clusters adiacenti vengono replicati in entrambi.

Quando i clusters acustici C_i sono costituiti, l’algoritmo di clustering determina gli s subclusters articolatori S_j^i ($j = 1 \dots s$) associati a ciascun C_i . s è uguale al massimo numero di vettori articolatori associati al medesimo vettore acustico x^* in C_i . Ogni vettore articolatorio associato ad x^* è posto in un subcluster distinto ed usato come “seme”. I rimanenti vettori articolatori vengono allocati come segue. Quando diversi vettori articolatori v_k^j sono associati al medesimo vettore acustico x_k , ogni v_k^j è assegnato ad un differente subcluster, scegliendo quello che ha il centroide *articolatorio* più vicino. La posizione di tale centroide è aggiornata in seguito ad ogni aggiunta di nuovi vettori.

Avendo determinato i clusters C_i ed associato ciascuno con uno o più subclusters S_j^i , all’interno di ogni S_j^i costruiamo quattro diverse reti RBFN per interpolare ogni dimensione dello spazio articolatorio. Ogni vettore acustico associato al subcluster è usato come centro per una funzione base della RBFN (nel nostro caso funzioni gaussiane). I valori per i parametri di ciascuna funzione base (deviazione standard, ecc.) sono stati determinati con una estesa serie di sperimentazioni sul codebook. Dopo aver determinato tutte le RBFN è possibile interpolare lo spazio articolatorio. La seguente procedura è utilizzata per passare alla programmazione dinamica i vettori interpolati. Dato un vettore acustico troviamo i k cluster acustici a lui più vicini, e tutti i subclusters ad essi associati. Il vettore acustico è quindi usato come dato di ingresso per l’insieme delle RBFN presenti in ciascun subcluster. Infine, tutti i vettori articolatori ottenuti dall’interpolazione (tanti quanti il numero di subclusters individuati) vengono passati alla procedura di programmazione dinamica, che procede all’ottimizzazione.

4. RISULTATI E DISCUSSIONE

Gli algoritmi proposti sono stati inizialmente testati e tarati con sequenze artificiali di vettori acustici target. Queste sono state usate come ingresso al sistema per ottenere i corrispondenti parametri articolatori. I risultati ottenuti da questi test preliminari hanno dato due indicazioni; per prima cosa, si è verificato che i segnali sintetici ottenuti controllando il modello fisico con i parametri articolatori ottenuti inseguono bene i vettori acustici target. La seconda indicazione è stata che le attivazioni muscolari e la pressione subglottale, ottenute dal sistema, hanno evoluzioni fisiologicamente attendibili, ovvero presentano variazioni fluide nel tempo. Questi risultati iniziali confermano la validità della funzione di costo usata e della interpolazione con RBFN.

Per verificare gli algoritmi proposti su segnali reali abbiamo realizzato una procedura completa di “sintesi mediante analisi” (*synthesis-by-analysis*). Partendo da una voce registrata (una vocale sostenuta con altezza e voce

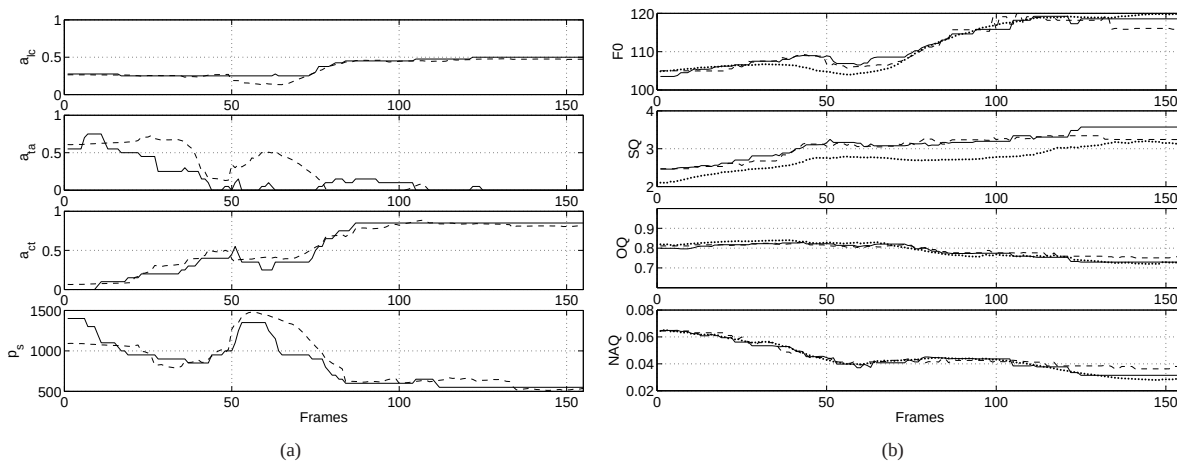


Figura 3. Esempio di procedura di sintesi mediante analisi. (a) Sequenze temporali di parametri articolatori ottenute dalla procedura di ottimizzazione (linea solida: senza RBFN; linea tratteggiata: con RBFN). (b) Sequenze temporali di parametri acustici del flusso glottale (linea a punti: sequenze target estratte da una frase registrata; linea solida: risintesi senza RBFN; linea tratteggiata: risintesi con RBFN).

quality variabili) il segnale è stato sottoposto a filtraggio inverso mediante APARAT. Il flusso glottale stimato è stato analizzato *frame-by-frame*, ottenendo una sequenza di vettori acustici misurati. I corrispondenti vettori articolatori, derivati mediante i procedimenti esposti in Sez. 3, vengono usati per guidare il modello fisico; il flusso glottale risintetizzato viene quindi convoluto con il filtro tempo-variante dei formanti del tratto vocale. Il risultato finale è la risintesi del segnale vocale di partenza, in cui l'evoluzione dell'altezza e della voice quality è simile a quella originale.

La Fig. 3 mostra le prestazioni della procedura di sintesi mediante analisi su una fonazione reale (una /e/ sostenuta). I vettori acustici tempo-varianti ottenuti nella risintesi inseguono con buona accuratezza quelli target; test di ascolto informali confermano che la risintesi è qualitativamente simile al segnale originale. In particolare, *NAQ* è solitamente ben inseguito, come visibile in Fig. 3(b). Questo è un risultato positivo poiché *NAQ* è noto essere fortemente correlato alla voice quality [5]. L'effetto dell'impiego delle RBFN può essere notato in Fig. 3(a): le sequenze di vettori articolatori interpolate dalle RBFN hanno variazioni più fluide rispetto alle sequenze ottenute usando la sola programmazione dinamica. Un secondo vantaggio dell'impiego delle RBFN è che il numero di vettori che vengono forniti alla procedura di programmazione dinamica viene significativamente ridotto, portando così ad una riduzione del tempo di calcolo necessario all'ottimizzazione.

I risultati riportati in questo lavoro indicano che l'approccio proposto è efficace nella stima dei parametri di controllo del modello fisico, sia con dati target sintetici, sia con segnali vocali reali; tuttavia, alcune limitazioni affliggono le prestazioni della procedura di stima descritta nell'articolo. Esse sono principalmente legate alle limitazioni intrinseche del modello a due-masse; gli intorni di variazione per i parametri acustici, infatti, sono in genere ridotti (si veda Fig. 2) ed, in alcuni casi, non realistici. *RQ* e *NAQ* in particolare assumono valori eccessivamente

bassi; la causa è la limitata capacità del modello di descrivere il flusso con piccole aperture glottali. Questo causa la simulazione di chiusure glottali improvvise e un picco della derivata del flusso eccessivamente alto. La relazione tra i parametri fisici del modello ed i parametri acustici necessita inoltre di essere verificata meglio: ad esempio, la relazione tra p_s ed F_0 osservata nel modello non è in accordo con i risultati riportati in letteratura. Infine, per sfruttare appieno i benefici dell'interpolazione del codebook, è necessario un approccio più sistematico alla determinazione dei parametri delle RBFN.

5. BIBLIOGRAFIA

- [1] D. Sciamarella and C. D'Alessandro, "On the acoustic sensitivity of a symmetrical two-mass model of the vocal folds to the variation of control parameters," vol. 90, pp. 746–761, July 2004.
- [2] J. Schroeter and M. Sondhi, "Speech coding based on physiological models of speech production," in *Advances in Speech Signal Processing* (S. Furui and M. Sondhi, eds.), pp. 231–263, New York: Dekker, 1992.
- [3] N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis, and A. Hirschberg, "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," vol. 84, pp. 1135–1150, 1998.
- [4] F. Avanzini, S. Maratea, and C. Drioli, "Physiological control of low-dimensional glottal models with applications to voice source parameter matching," vol. 92, pp. 731–740, Sept. 2006.
- [5] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," vol. 112, pp. 701–710, Aug. 2002.

- [6] P. Alku and E. Vilkman, "A comparison of glottal voice quantification parameters in breathy, normal and pressed phonation of female and male speakers," vol. 48, pp. 240–254, Sept. 1996.
- [7] I. R. Titze and B. H. Story, "Acoustic interactions of the voice source with the lower vocal tract," vol. 101, pp. 2234–2243, Apr. 1997.
- [8] I. R. Titze and B. H. Story, "Rules for controlling low-dimensional vocal fold models with muscle activation," vol. 112, pp. 1064–1027, Sept. 2002.
- [9] M. Airas, H. Pulakka, T. Bäckström, and P. Alku, "A toolkit for voice inverse filtering and parametrisation," in *Proc. 9th European Conf. on Speech Communication and Technology (Interspeech'2005 - Eurospeech)*, (Lisbon), pp. 2145–2148, Sept. 2005.
- [10] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE*, vol. 78, pp. 1481–1497, Sept. 1990.
- [11] A. Gercho and R. M. Gray, *Vector quantization and signal compression*. The Kluwer international series in engineering and computer science, Kluwer, 1992. Boston.

A microphone array approach for browsable soundscapes

Sergio Canazza

Sound and Music Computing Group
Dep. of Information Engineering
University of Padova, Italy
canazza@dei.unipd.it

Antonio Rodà

AVIRES Lab.
Dep. of Math. and Computer Science
University of Udine, Italy
antonio.roda@uniud.it

Daniele Salvati

AVIRES Lab.
Dep. of Math. and Computer Science
University of Udine, Italy
daniele.salvati@uniud.it

ABSTRACT

This article presents an innovative architecture for the recording and the interactive browsing of soundscapes. The system uses a limited set of microphone arrays to capture sound signals from an open space (eg a square or a street). Then, the user can select a point or draw a trajectory in the plane of interest and beamforming techniques are used to attenuate all the signals that do not come from the desired point. The system was tested by simulating a soundscape captured by two linear arrays. The results show that even with only two arrays, you can select different sources in the soundscapes, exploring the space from one source to another.

1. INTRODUCTION

Although the word *soundscape* can be used in several scientific fields with different meanings (see [1] for a review), the concept of soundscape concerns, in any case, sounds pertinent to a place, i.e. sounds that are spatially and/or geographically organized. In the late sixties, R. Murray Schafer gave birth to the World Soundscape Project, an educational and research group aimed at studying the sonic environments. With the collaboration of colleagues and students, Schafer picked hundreds of recordings of American and European soundscapes, using a portable magnetic tape recorder. In recent years, the spread of digital audio technologies and telecommunications networks has given new impetus to the collection and dissemination of soundscapes. Participants in many collaborative projects have started to capture and share through Internet a large amount of field sound recordings from around the world¹ or collected with the aim to create a sound map of a particular city². The recordings are made in mono or stereo format and are usually geographically tagged. Each recording represents a single subjective point of view, or better a point

of listen, of the soundscape. This implies that a very high number of recordings (points of listen) are needed to have a global representation of the soundscape. Moreover, users can access a recording by selecting the geographical coordinates on a map, but this can be done only for those points of listen where the recordings was made. Therefore, it is not usually possible to browse with continuity through the sound map, like in a real context. Some works attempt to overcome this limitation. For example, Valle et al. [1] proposed a graph-based system for the dynamic generation of soundscapes that can allow an interactive and real-time exploration of a soundscape. The soundscape is generated by defining a graph structure, named GeoGraphy, whose nodes represent the sound sources and are geographically positioned. The user can navigate freely around the map where the graph is defined, moving towards or away the spatially organized nodes. This system, while allowing you to navigate with continuity within the sound environment, requires a prior analysis of the soundscape, the definition of a number of points of listen, and the recording or simulation of any sound source corresponding to those points. The LISTEN project [2] aims to define a hardware and software architecture for creating an immersive audio-augmented environment. It consists in a series of sound objects (sound files, audio effects, etc...) together with the description of their spatial organization, updated in real-time with respect to the listener's position and orientation. The system allows you to navigate interactively within a soundscape, always seen, however, as a collection of spatially distributed audio files. For example, to simulate the soundscape of a marketplace, you must separately capture the sounds produced by different vendors, the sound of people walking, the noise of cars on the road, the sound of a fountain, and so on, saving the information about where each recording took place.

This paper presents a different approach to the recording and fruition of soundscapes. The idea is to record a soundscape using a small number of microphone arrays, instead of a relatively high number of mono or stereo recordings. In consequence of the principle that sound waves coming from different directions will arrive to the array sensors with different delay times, the signals captured by a microphone array also contains information about the spatial location of the sources. Then, a soundscape composed from multiple sources located in different places can be captured by a limited number of arrays because is then possible to separate the sources coming from different di-

¹ E.g., RADIO APOREE MAPS (<http://aporee.org/maps/>);
SOUNDCITIES (<http://www.soundcities.com/>);
LOCUSTREAM SOUNDMAP (<http://locusonus.org/>)

² (E.g., SONS DE BARCELONA
(<http://barcelona.freesound.org/>);
SOUND-SEEKER <http://soundseeker.org/>;
LONDON SOUND SURVEY <http://www.soundsurvey.org.uk/>)

rections using beamforming techniques. Indeed, the array can be steered according to a desired beam pattern, which is modeled by processing the signals captured by the microphones. Changing the direction of the beam pattern, you can explore the sound field, highlighting a source or the other.

Many techniques for processing the signals from microphone arrays have been developed in recent years, with application to various contexts as, for example, the tracking of the speaker during a conference [3], the reduction of noise coming from concurrent sources [4] or the acoustical analysis of a mechanical device [5]. The application of these techniques to the capturing and browsing of soundscapes requires to adapt them to the constraints of the new applicative scenario: i) the far-field condition (it is often necessary to locate sources at a distance of tens of meters), in which the acoustic pressure wave can be approximated to a plane wave; ii) the need to monitor sources that are moving on a two-dimensional space (the plane of a square, a street or a monitored park); iii) the need to place sensors on a plane different from that monitored, in order to avoid damage by pedestrians or vehicles; iv) the need to have a reduced number of arrays, not to invade the public spaces in an excessive way. Whereas in the near-field case would be sufficient a linear array of at least three microphones to locate the sources position in a two-dimensional space, in the far-field case the estimation of the source position is extremely difficult, if not almost impossible, using a single array: from the Time Difference Of Arrival (TDOA) among the microphones we can estimate the Direction Of Arrival (DOA) of the sound, but not its distance. Therefore, the two-dimensional position of the source can be estimated using two linear arrays, by means of the triangulation of the DOA estimations (see Figure 1).

The rest of this paper is organized as follows: after presenting the system architecture (Section 2), we briefly summarize the adopted algorithm for the beamforming of the microphone array (Section 2.1). Finally, Section 3 illustrates some preliminary experimental results, obtained in a simulated scenario.

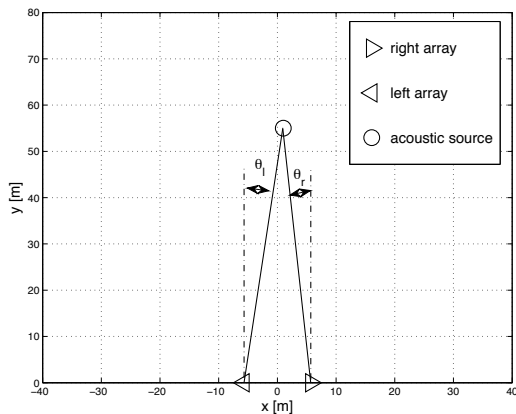


Figure 1. Single source localization; x, y axes reference.

2. SYSTEM ARCHITECTURE

A key feature of the microphone arrays is the ability to direct (to steer) the array towards a specific direction. I.e., the signals captured by the microphones can be processed in order to attenuate the sound waves from all directions except the desired one. After recording the signals captured by microphones, the proposed system takes as input the spatial coordinates of a point in the plane of interest and proceed to the attenuation of all the sound signals except those from that point. While using a single array you can select audio signals from a specific direction, to select those from a point you must use at least two arrays: if each of the two arrays is steered toward a specific direction, the selected point is positioned at the intersection of those directions (it is necessary to put some constraints on directions, e.g. they should not be parallel). Though two arrays are sufficient to direct the playback to a point, the discriminatory capacity increases with the number of the arrays.

The user specifies the coordinates of the point (x, y) towards which to steer the array (see Figure 2). Through the function `pos2tdoa()`, the system maps the coordinates of the point in TDOA values, which correspond to the Time Difference Of Arrival of an audio signal that reaches the array from the specified point. Since the arrays are located in different places, you must calculate a TDOA value for each array. These values are used to steer each array to the point (x, y) , by means of beamforming techniques. The signals processed by the appropriate beam pattern are finally synchronized and summed.

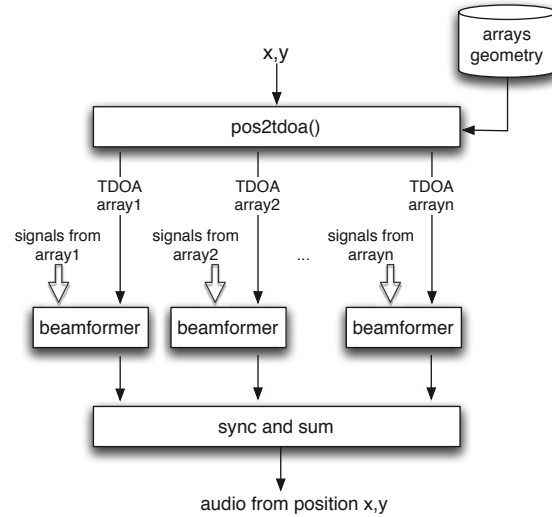


Figure 2. The system architecture.

In the case the two arrays do not lie on the plane of interest, as is recommended when the recording takes place in public spaces, it is necessary to derive the equations that relate the points on the plane with the arrival angles of the sound waves. The possible points identified by desire angle are located on a cone surface, whose vertex is placed in the array and whose axis is the straight line joining the

two arrays. Every array presents a cone: the intersection of the two cones is represented by a circumference. The intersection point between the circumference and the plane of interest is the estimation of the source distance from arrays. Hence, considering d_a the distance of the arrays, h the height of arrays above the plane of interest, ϕ_l and ϕ_r the desire angle of left and right beamformer, we obtain:

$$x = \frac{d_a}{2} \left(\frac{\tan \phi_l + \tan \phi_r}{\tan \phi_l - \tan \phi_r} \right) \quad (1)$$

$$y = \sqrt{\left(\frac{d_a}{\tan \phi_l - \tan \phi_r} \right)^2 - h^2} \quad (2)$$

2.1 Beamforming techniques

The beamforming [6] can be seen as a combination of the delayed signals from each microphone in such a way that an expected pattern of radiation is preferentially observed. The process can be subdivided in two sub-tasks: synchronization and weight-and-sum. The synchronization task consists in delaying (or advancing) each sensor output of an adequate interval of time, so that the signal components coming from a desired direction are synchronized. The information required in this step is the angle corresponding to the desired direction. The weight-and-sum task consists in weighting the aligned signals and then adding the results together to form a single output. The output signal of beamformer allows to enhance a desired signal from its detection corrupted by noise or competing sources. The Delay & Sum Beamforming (DSB) is the classical technique for realizing directional array systems. In general, the DSB output y at time k is:

$$y[k] = \frac{1}{N} \sum_{n=1}^N x_n[k + F_n(\tau(\phi))] \quad (3)$$

where N is the number of microphones, x_n is the received signal at microphones n and $F_n(\tau(\phi))$ is the TDOA between the n_{th} microphone and the reference and depends on the microphone array geometry and on the angle ϕ corresponding to the desired direction. For a linear and equispaced arrays, i.e. Uniform Linear Array (ULA), we have:

$$F_n(\tau(\phi)) = (n-1)\tau(\phi), \quad n = 2, \dots, N \quad (4)$$

In far-field condition, in which the acoustic pressure wave can be approximated to a plane wave, the TDOA between two microphones can be express as:

$$\tau(\phi) = \frac{d \sin(\phi)}{c} \quad (5)$$

where c is the speed of sound and d the distance between microphones.

In the frequency domain, the DSB output from (3) becomes:

$$Y[k, f] = \frac{1}{N} \sum_{n=1}^N X_n[k, f] e^{-j2\pi f F_n(\tau(\phi))} \quad (6)$$

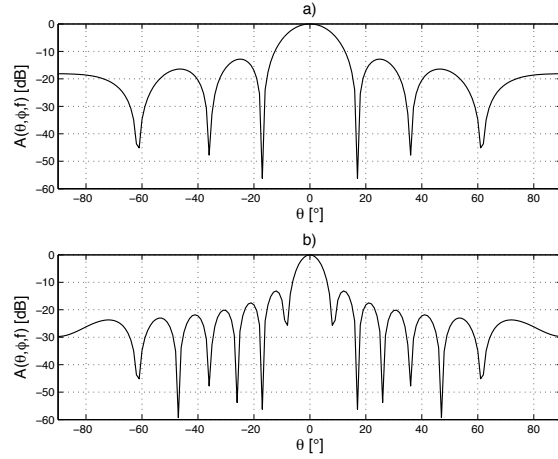


Figure 3. The beam pattern of ULA when $d = 10$ cm $\phi = 0^\circ$ and $f = 1.5$ kHz. a) eight sensors b) sixteen sensors.

where $Y[k, f]$ and $X_n[k, f]$ are the Discrete Fourier Transform (DFT) of the signals. The frequency response of the DSB is defined as:

$$R(\phi, f) = \frac{1}{N} \sum_{n=1}^N e^{-j2\pi f F_n(\tau(\phi))} \quad (7)$$

In this case, the response depends only from the geometry of the array: the number of microphones, the distance between the microphones, the placement of the microphones.

In general, introducing a weights filter $\mathbf{w} = [w_1 w_2 \dots w_N]^T$, and defining $\mathbf{r}(\phi, f) = [e^{-j2\pi f F_1(\tau(\phi))} \dots e^{-j2\pi f F_N(\tau(\phi))}]^T$ the frequency response can be expressed as:

$$R(\phi, f) = \mathbf{w}^T \mathbf{r}(\phi, f) \quad (8)$$

Then, the beam pattern on desire direction ϕ , representing the gain of beamformer, is written as:

$$A(\phi, f) = |R(\phi, f)| \quad (9)$$

In case of DSB (where the vector $\mathbf{w}$ is equal to one), in case of ULA and far-field environment, and assuming an angle range as: $-90^\circ + 90^\circ$ ($-\pi/2 < \theta < \pi/2$) (where zero is in front of the array, and the microphone reference is the first from left), the beam pattern becomes:

$$A(\theta, \phi, f) = \left| \frac{1}{N} \sum_{n=1}^N e^{\frac{-j2\pi f (n-1)d(\sin(\theta) - \sin(\phi))}{c}} \right| \quad (10)$$

Figure 3 shows the beam pattern for an equispaced linear array of eight and sixteen microphones, microphone distance $d = 10$ cm, frequency $f = 1.5$ kHz, and desired direction $\phi = 0^\circ$. The beam on desired direction with the highest amplitude is named mainlobe and all the others are called sidelobes. The sidelobes represent the gain pattern for noise and competing sources along the directions other than the desired one. The beamforming techniques

aim to make the sidelobes as low as possible so that signals coming from other directions would be attenuated as much as possible. For this reason, to improve the beamforming performance, some filter methods have been developed in order to define the weights vector w , e.g. least-squares technique [7] for data independent beamforming, and minimum variance distortionless response technique [8] for adaptive beamforming.

3. RESULTS

To verify if the proposed approach is applicable to the recording and browsing of soundscapes, we rendered a virtual soundscape, simulating a recording by means of arrays. We carried out two simulations, both made using two arrays: the first simulation is based on two arrays composed by eight microphones each one; the second, two arrays with sixteen microphones each. We consider the sources located in a virtual plane of about 50x50 meters, so the far-field condition is generally satisfied. The distance between the arrays is assumed to 11.4 m. The sample rate of sounds is 44.1 kHz and the observation time for the Short Time Fourier Transformer (STFT) is 4096 samples, with an overlap-add of 512 samples. The simulated soundscape is composed by three sound sources, whose waveforms and spectrograms are visible in Figure 4. The three sources were placed in a virtual acoustic scenario, following the map plotted in Figure 5. The two-dimensional coordinates coordinates are: source 1 (-5.7, 9.2), source 2 (4, 10.1), and source 3 (-19.3, 32). We assumed the user draws a trajectory in the virtual space that, starting from the position of source 2, reaches source 1 and source 3, passing through the points P1 and P2. According to Section 2, for each point in the trajectory, the signals coming from the arrays are processed by means of a DSB. Then, the beamformed signals are synchronized and summed (see Figure 2).

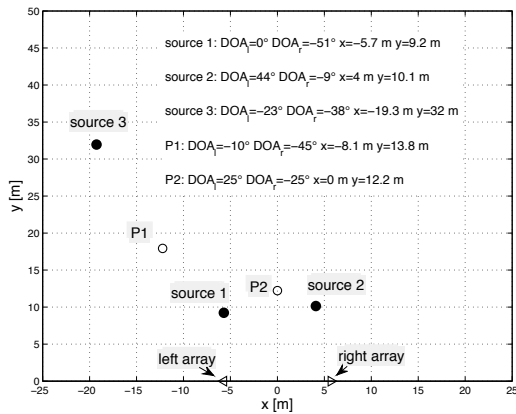


Figure 5. The acoustic map scenario.

In this scenario, the signal received by the first microphone of left array is shown in Figure 6.

We analyze now in detail the output signal corresponding to the 5 points: source 1, source 2, source 3, P1, and

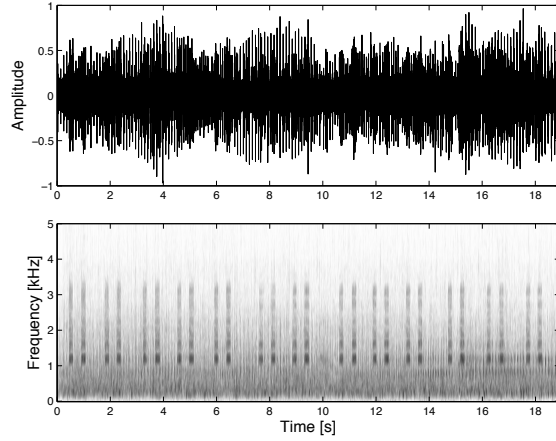


Figure 6. The signal received by the first microphone of the left array.

P2. The position of source 1 corresponds to the steering angles $\phi_l = 0^\circ$ (for the left array) and $\phi_r = -51^\circ$ (for the right array). Figure 7 shows the waveform and the spectrogram of the output signal, obtained with 2x8 microphones (on the left) and 2x16 microphones (on the right). Comparing it with Figure 4, it is possible to see the capability of the system to enhance the source 1 and to separate it from the other sounds. The same is done by pointing the array towards the source 2 ($\phi_l = 44^\circ$ and $\phi_r = -9^\circ$) and source 3 ($\phi_l = -23^\circ$ and $\phi_r = -38^\circ$). Figure 8 and 9 show the output signals in these cases. Regarding the positions P1 ($\phi_l = -10^\circ$ and $\phi_r = -45^\circ$) and P2 ($\phi_l = 25^\circ$ and $\phi_r = -25^\circ$), which are intermediate points, the output signal is characterized, as one might expect, a combination of all three sound sources (see Figure 10 and 11), even if the signal amplitude is quite low.

As concern the number of microphones, the results show that the sidelobes are attenuated by increasing the number of microphones, giving a better separation of the sources. Instead, looking at the results shown in Figure 8 and 9, we can see the best performance of beamforming with more sensors.

4. CONCLUSIONS

This paper presented an architecture based on microphone arrays to record and browse soundscapes. The purpose of this system is to obtain a highly directional microphone antenna, based on the use of two linear arrays and a Delay & Sum Beamforming technique. Combining the output of the two arrays, the system can emphasize the sound coming from any point of a two-dimensional plane on which the acoustic sources are located. This approach can be apply to the soundscape of open spaces of large dimensions, as is the case of a square or a park.

We verified the functionality of the system with a simulated soundscape composed by three sources. The results showed the system's capacity to enhance the source of interest and to separate it from other sounds, underlining the limitations due to the presence of sidelobes in

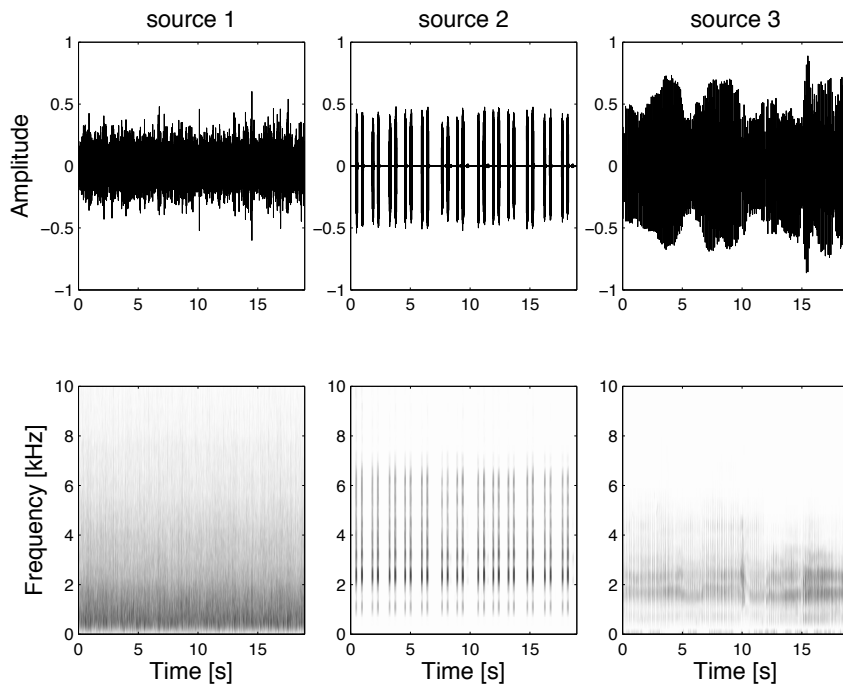


Figure 4. The waveforms and spectrograms of the three sources used in the simulation.

the spatial response filter of the beamforming. The system performance can be improved by increasing the number of microphones of array and the number of arrays. Other improvements concern the use of filter beamforming techniques and adaptive beamforming methods: these algorithms allow to reduce the interferences of competitive sounds and to enhance the observation of the pointed soundscape. This will be the subject of future investigations.

5. ACKNOWLEDGMENTS

This work is partially supported by the Smart resource-aware multi- sensor network project (SRSnet), an Interreg IV research project funded by the European Community.

6. REFERENCES

- [1] A. Valle, V. Lombardo, and M. Schirosa, "A graph-based system for the dynamic generation of soundscapes," in *Proceedings of the 15th International Conference on Auditory Display (ICAD2009)* (M. Aramaki, R. Kronland-Martinet, S. Ystad, and K. Jensen, eds.), (Copenhagen, Denmark), 18–21 May 2009.
- [2] O. Warusfel and G. Eckel, "Listen-augmenting everyday environments through interactive soundscapes," in *Virtual Reality for Public Consumption, IEEE Virtual Reality 2004 Workshop*, vol. 27, (Chicago IL), 2004.
- [3] N. Strobel and R. Rabenstein, "Robust speaker localization using a microphone array," in *In Proceedings of the X European Signal Processing Conference, volume III*, pp. 1409–1412, 2000.
- [4] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *The Journal of the Acoustical Society of America*, vol. 76, no. 1, pp. 84–84, 1984.
- [5] S. R. Venkatesh, D. R. Polak, and S. Narayanan, "Beamforming algorithm for distributed source localization and its application to jet noise," *AIAA journal*, vol. 41, no. 7, pp. 1238–1246, 2003.
- [6] H. Johnson and D. E. Dudgeon, eds., *Array Signal Processing: Concepts and Techniques*. Simon & Schuster, 1993.
- [7] S. Doclo and M. Moonen, "Design of far-field and near-field broadband beamformers using eigenfilters," *Signal Processing*, vol. 83, pp. 2641–2673, 2003.
- [8] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, 1969.

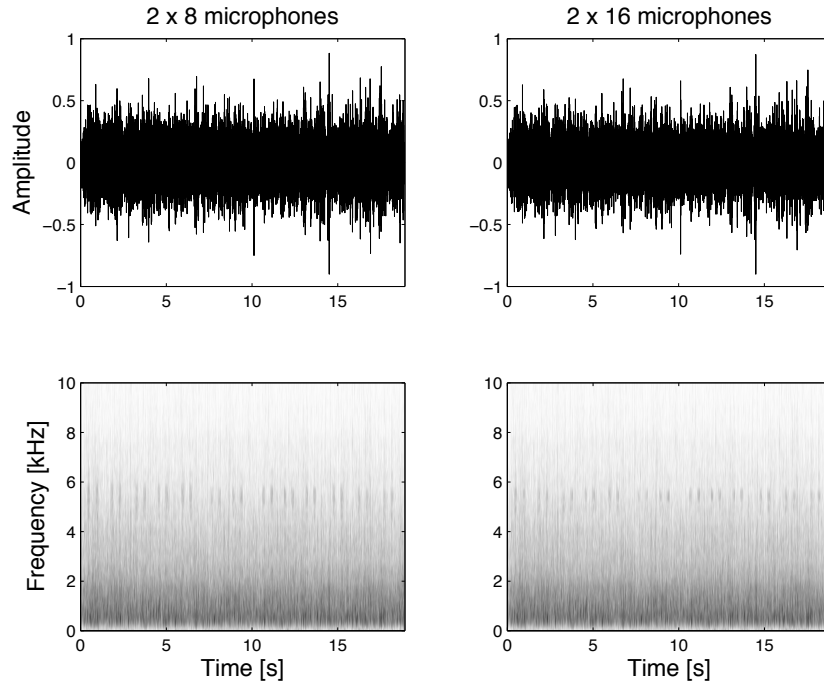


Figure 7. The beamforming output on desired angles $\phi_l = 0^\circ$ and $\phi_r = -51^\circ$ (source 1).

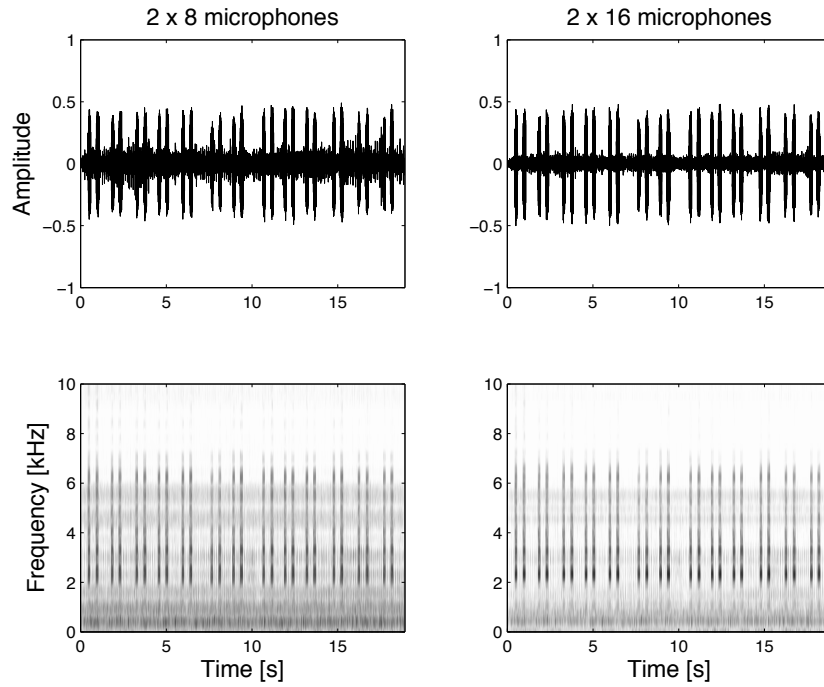


Figure 8. The beamforming output on desired angles $\phi_l = 44^\circ$ and $\phi_r = -9^\circ$ (source 2).

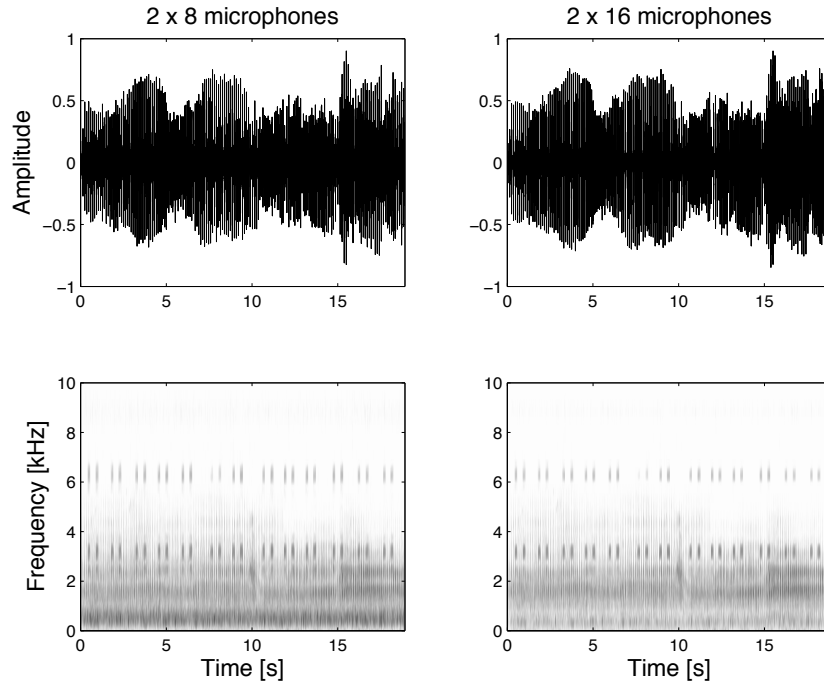


Figure 9. The beamformings output on desired angles $\phi_l = -23^\circ$ and $\phi_r = -38^\circ$ (source 3).

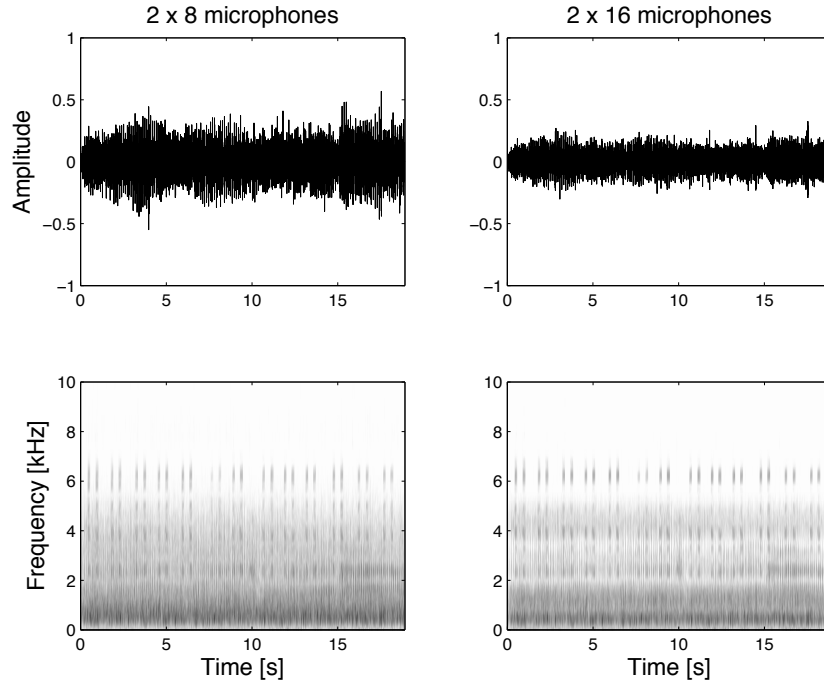


Figure 10. The beamformings output on desired angles $\phi_l = -10^\circ$ and $\phi_r = -45^\circ$ (P1).

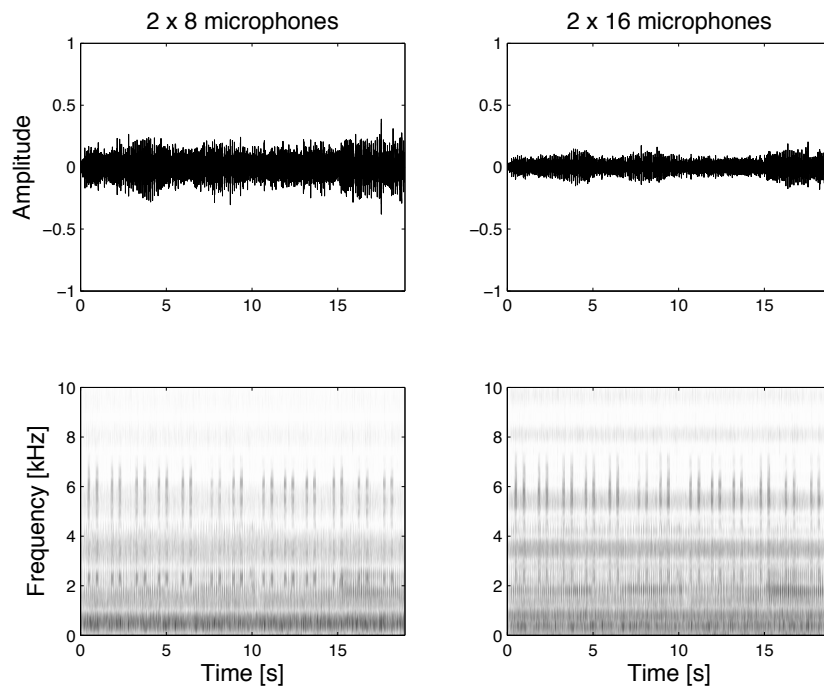


Figure 11. The beamformings output on desired angles $\phi_l = 25^\circ$ and $\phi_r = -25^\circ$ (P2).

A system for soundscape generation, composition and streaming

Mattia Schirosa, Jordi Janer, Stefan Kersten, Gerard Roma

Universitat Pompeu Fabra

Music Technology Group

Roc Boronat 138, Barcelona

firstname.lastname@upf.edu

ABSTRACT

Soundscape design is beginning to receive considerable attention in virtual environments and interactive media developments. Current trends (e.g. online communities and games, web and mobile technologies and augmented-reality tourism platforms, 2D and 3D virtual cartography and urban design) might require new paradigms of soundscape design and interaction.

We propose a generative system that aims at simplifying the authoring process, but offering at the same time a realistic and interactive soundscape. A sample-based synthesis algorithm is driven by graph models. Sound samples can be retrieved from a user-contributed audio repository. The synthesis engine runs on a server that gets position update messages and the soundscape is delivered to the client application as a web stream. The system provides standard format for soundscape composition. Documentation and resources are available at [1].

1. INTRODUCTION

Audio is a crucial element for building immersive virtual environments. In this context, a principal role of audio is the creation of a sound ambience or soundscape, in fact, during the last decade, several technologies emerged to provide a more realistic and interactive audio content. We can include, among others, real-time synthesis techniques (e.g. physical models [2], sound ambience textural synthesis [3], [4]) or spatial audio reproduction systems (e.g. 5.1, Ambisonics).

Audio generation systems generally feature different characteristics depending on the type of media production (e.g. animation films, virtual worlds, arcade games, etc.). They can differ in terms of interaction possibilities, latency requirements, or the achieved sonic realism.

At the same time, we observe that current developments such as online communities, web and mobile technologies, might bring applications that make use of a new paradigm of soundscape generation. Precisely in this new context, we believe that the authoring process should be simplified, also encompassing user-generated content.

Our system is positioned in this new scenario. It takes advantage of user-contributed sound assets and it provides a technology that allows the system to be part of web-based tools for the design and interaction of the soundscape.

The system is formed by three components: composition format, generation module and streaming web service. Figure 1 shows the whole system overview.

The composition format is a series of rules and parameters that describes the sound space, the sound concepts that live within that space, their generation probabilities and several types of sequencing behaviour between events. This information is encoded in a KML score file: KML is an XML schema aimed at describing and displaying geographic data [5]. The composition is also aimed at producing the database: an XML document that stores the sound concept events annotation mapped to segmented sound samples. Composition format is addressed to designer, composers and users. Composition or design are human processes, it is a challenge to think to a complete automatic design.

The generation module parses the score and the database annotations, it generates the graph structures, performs the sequencing and the multi-listener spatialisation synthesis, provide an OSC (Open Sound Control) interface for real-time soundscape performance interaction and listener position control.

The actual generation runs on the streaming server, the listeners position and orientation are communicated through HTTP request messages, and the soundscape is accessed as a MP3 stream. The developed web API includes session management, allowing personalized streams for simultaneous listeners in the same soundscape.

2. BACKGROUND

When we speak about “Soundscape” we mean an interactive and explorable audio environment that completely differs from linear sound design works (e.g. cinema, animation, tv). We elaborate the soundscape definition presented in [6] and [7]: “*Soundscape*” is a complex temporal-spatial structure of sound objects that composes the perception of an environment in a listener throughout its hearing, moving and discovering process.

The soundscape is composed by a set of sound zones, which are populated by a set of sound concept classes which are described by a sequencing structure and realised in a set of sound events.

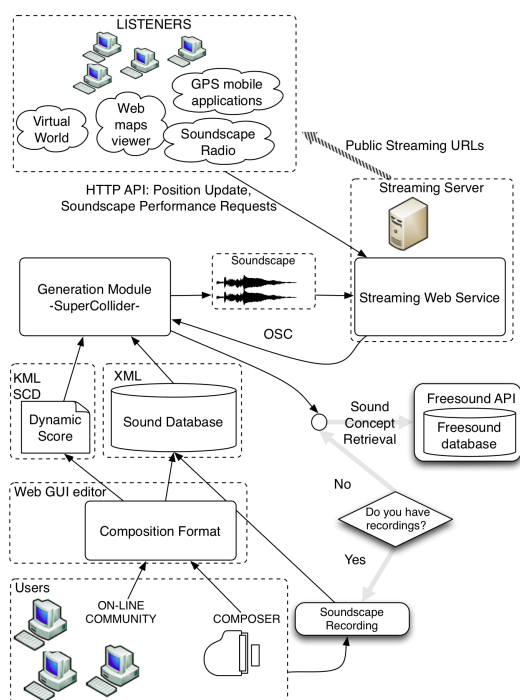


Figure 1. System overview

Sound zone is specific sub-part of the soundscape that presents characteristic sound ambiances and sound sources that allow to distinguish it from the other sub-parts. For instance, in one of our experiment, the sonification of the Second Life re-creation of the Canary Islands Las Palmas city, we identify 4 zones which are managed independently: the main city square, the city beach, the most important museum, and the biggest park just outside the museum. Zones have their own description scale that can differ from zone to zone. They have parameters that describe the area of space they refer to (geometry, closed ambience, scale etc.). Each Zone is composed by a set of sound concepts.

A **sound concept** describes the sonic behaviour of a “relevant” set of signals that are assumed to be perceived with the same meaning by the listener. Each concept is described by a set of sound events (simulated with sound samples), a sequencing structure stored in graph object (see figure 2) and parameters (see section 3) that affect the sequencing structure and the synthesis engine in real-time. The set of events collected in a concept class could represent several nuances of the concept. The user must decide where to put the threshold in concepts definition. For instance, describing the human sound activity of a soundscape, considering a coarse scale, we could choose just one concept: “people”. Instead, using a higher detail, four concepts: “man”, “woman”, “children” and “elderly people” voices; finally, with a deeper detail, sixteen concepts: “laughs”, “cheers”, “screams”, “quite voices”, each one divided in male, female, young and elderly. The choice of sound concepts is fundamental for the system because

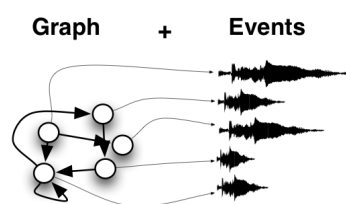


Figure 2. The sound concept

it is the actual mapping performed by the user in order to control the temporal evolution of the soundscape through the interaction with the concept parameters. Certainly, this definition is related to the target soundscape and depends on the scene the user is interested to describe. This means that just one “people” concept is enough in a soundscape where humans are very rare, such as a forest. In this case, the user has no convenience in mapping all the specific voice sounds, instead the presence of human activity is a more interesting concept. On the contrary, in the soundscape of a pedestrian area of a european city, probably the user would be able to control the generation of each type of sound independently (smiles, cheers, screams, woman, children, etc.) as they could be related to specific soundscape status (e.g. during the morning children probability can increase, while during the night the concept could be almost deactivated).

Atmosphere is a particular type of concept, it is an overall layer of sound, which cannot be analytically decomposed into single sound objects, as no particular sound signal emerges. Atmosphere characterizes quiet states without relevant sound events. Each zone must have at least an active atmosphere because it is the background layer from which concept signals emerge.

Certainly the concept declaration is a subjective process, [6] it presents an interesting elaboration of the soundscape studies in order to have analysis procedures and listening exploration methodologies that assist the user in this task and in the definition of “relevant” signals.

Sound events are instances of a concept class. Each event links to a sound sample, thus it contains the information about sound sample URL, rate, duration, format, distance of recording, etc. Event recordings are completely focused on a specific source and try to avoid the presence of any background sound, thus it is preferred to select recordings performed with a highly directional microphone. Instead, atmospheric recordings represent a quite state of the soundscape with no relevant events, better performed with omnidirectional microphones. Events copy all the parameters from the parent class, but the user could specify their own specific parameters, for instance position.

The concept sequencing structure is represented by a **Graph** [8]: each vertex represent a sound sample. The edges represent a possible sequencing relation on a pair of vertices, the edge duration represents the pause between the triggering time of a vertex and the next one in the generative sequencing. Edge duration and vertex duration can be equal but do not necessarily have to coincide. If they are

not equal the graph models triggering sequence were the next sound will be activated first or after the previous one stops. This behaviour allows to model pauses and cross-fades between vertices. Each edge has a specific probability to be chosen in the sequencing path. Actants are dynamic elements that navigate the graph and perform the triggering process according to the generative path based on probabilities. Each graph could have several actants.

3. COMPOSITION FORMAT

The composition format provides rules and parameters that assist the user in the soundscape composition through the sonic space description, the soundscape sequencing behaviour definition and the events database search. The user should sketch a metric map of the landscape, in order to have a reference system where to locate sounds. It is important to determine if the soundscape is composed by zones, this implies to name and define zone areas. Then the user chooses the sound concepts of each zone and, for each concept, he provides the annotation of its instances: audio samples that realise the events of the general concept class.

3.1 Sonic space

First, the user defines the sonic space, as shown in figure 3. The spatial parameters for zones are “geometry”, “scale” and “closed ambient”. Geometry describes the area where its features have effects, the synthesis engine supports only rectangular geometries. Scale is the ratio between the unit of the virtual space representation and meters. Each zone of the soundscape could have a different scale, this parameter affects the spatialisation engine. The user can assign a zone geometry to a closed ambient, in this case the spatialisation engine will mute all the other zones when the listener enters a closed ambient, while no closed ambient is audible when the listener is outside its geometry. Note that, if a zone has no spatial related feature, the geometry could also not be specified, in this case the zone is described by the set of its concepts positions, this allows to not rely on the limit of the rectangle geometry (for instance, the user can create a whole zone for all the streets of a city, in this case the zone is not a closed ambient but a container layer that allow to control all its concept parameters at the same time).

The second step focuses on concepts. Concepts have three position typologies: “point source”, “not point source” and “point source random generated in an area”. “Point” source is the most general type that represents a source position through a pair of coordinates. “Not point source” is defined by a rectangle, when the listener is inside it the concept amplitude does not change, while when the listener is outside it, the spatialisation engine computes the distance between him and the closest point of the rectangle; this position type is useful for specific sources or atmospheres that cannot be assumed to be located in a point (e.g. a non point source is a waterfall, whose sound is the compound of a multitude of parts). “Point source random generated in an area” is a useful position type to model

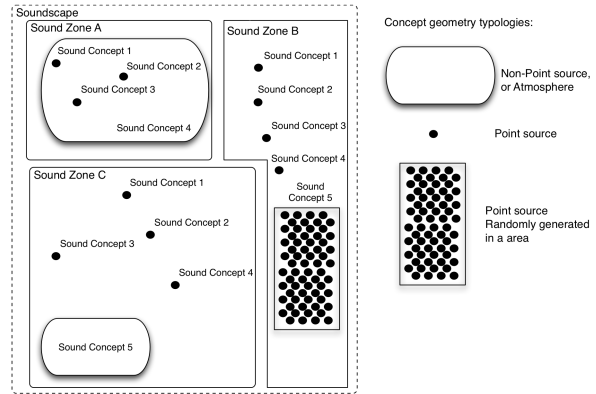


Figure 3. Sonic space description

common soundscape point sources that frequently appear in several positions moving along a specific area. For instance it is a powerful tool to model voices or waves. Concepts also has two position attributes: “clone” and “listened area”. If a concept is located in several fixed positions, clone is a more sensible choice than “point source random generated in an area” to model multiple concept occurrences in space. Clone allows to copy the parameters of a concept and to reuse its samples to place the same concept in another location. For instance it is useful to model several flags flapping under the action of wind. “Listened area” attribute de-emphasizes the perception of distance, it controls the distance from which the spatialisation engine applies a Low Pass Filter to the sound event in order to set a target concept more clearly audible also from far distances.

3.2 Sequencing and mix

Once the sonic space is defined, the user focuses on sequencing and mixing. The mix, basically, is the initialisation of the zones and each concept “gain” parameters. Those values must be expressed just if the user needs to amplify or de-amplify the elements, thus when the gain differs from one ¹. Concepts can be also “active” or “inactive”. The sequencing is described by a set of concept parameters. A “continuous” parameter produces a never ending stream of events. The sequencing of a continuous concept does not present pauses, but it can be deactivated, while when it is active it is always audible in the soundscape, if the listener is close to it. Examples of continuous concepts are wind, fountains or general water streams. Non-continuous concepts are described by “probability”, “multiple path” and “arrhythmic generation”. “Probability” describes the number of concept occurrences per hour, but the same probability could represent very different sequencing behaviour depending on the pauses between the generation of two consecutive events and the number of possible contemporaries triggering process. “Multiple path parameter” informs about the number of contemporaries triggering process (i.e. the number of graph actants, as described in section 2). “Arrhythmic generation” is a param-

¹ The software uses an linear amplitude representation where 0.125 = -18dB, 0.25 = -12dB, 0.5 = -6dB, 1 = 0dB, 2 = +6dB.

eter that controls the irregularity of the triggering pattern. The graph generation algorithm is detailed in section 4.

The spatial and sequencing information can be both written in the SuperCollider language (see section 4) or in KML files. The KML, namely Keyhole Markup Language, is a description open format based on XML used to display geographic data developed by Google and used on Google Earth and Google Maps. KML files can be created with the Google Earth user interface, or using an XML or a simple text editor to enter raw KML from scratch. As KML allows to declare extension data, we created a **soundscape KML** scheme that augments the basic KML informations adding all the compositional parameters previously explained.

Refer to the system documentation page [1] for a practical guide for editing the scheme instances, here we provide a general introduction.

We use the tag “Placemark”, a tag with associated geometry, to declare zones and concepts. Placemark has name, description and two types of geometry elements, “Point” models point sound sources and “Polygon” models area sound sources and zone geometries. KML allows to create “Folder”, a container element, we use a parent Folder to declare the soundscape, which in turn contains the zone Folders and each zone Folder contains a collection of Placemarks. We declare three extended elements with scheme identifiers “soundscape”, “zone” and “concept” that allows to explicitly generate system specific data. Soundscape has four attributes: name, width, height and database type (the two supported database types are explained above). Zone has four attributes: gain, close ambient, scale and geometry, which allows to use a meter relative reference system instead of the KML standard latitude and longitude. Concept has nine attributes: geometry (again, meter relative reference system), gain, random generation, continuous, multiple path, probability, arrhythmic generation, listened area and clone. The possibility to use relative and absolute coordinate system allows the system to serve two types of application, respectively based on real locations or virtual spaces.

We chose the KML because it is becoming a popular format on online maps and earth browsers, thereby enabling interoperability of earth browser implementations. Programs such as AutoCad or 3D model editors easily export or convert their data in this format. We are developing a web GUI editor that allows users to place zones and concepts in a virtual space, to define parameters and to export a KML file.

3.3 Events database

Once the two previous step are accomplished, in a final step the user searches sounds that represent the concept events, creating the database. The system supports two types of databases: “pre-segmented events” or “annotated on-site recordings”. “Pre-segmented event” format allows the user to specify a list of segmented audio samples that populate the concept instances. The XML minimal schema starts with the element “soundscapeDatabase”, which contains the nested “soundConcept” elements that in turn have the child elements “name” and a set of ele-

ments “event”. The “event” element has six attributes that store the information about the database path (URL²), the segmentation (start and end frame), the sample rate of the audio file (sampleRate), and two types of normalisation attribute, the event distance of recording in case it is different from the standard of 5 meters (recDistance) or the normalisation volume. The system uses the latter attributes to normalise all the concept event with the same loudness before the spatialisation. Users that performed on-site recordings could prefer the distance of recording while users that search audio materials on online repository are more comfortable with a perceptive loudness normalisation.

A soundscape XML database annotation file example is:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE MTG soundscape composition system database>
<soundscapeDatabase>
  <soundConcept>
    <name>Wave</name>
    <event URL="http://www.freesound.org/samples/id101429.aiff"
      start="0" end="12381696" sampleRate="44100"/>
  <!-- New File-->
  <event URL="http://www.freesound.org/samples/id101459.aiff"
    start="23456" end="234556" sampleRate="44100" recDistance="2"/>
  <!--Same file, further event, further sound concept instance -->
  <event URL="http://www.freesound.org/samples/id101459.aiff"
    start="234999" end="333000" sampleRate="44100"/>
  </soundConcept>
  <soundConcept>
    <name>Children</name>
    <event URL="http://www.freesound.org/samples/id18929.aiff"
      start="0" end="12381696" sampleRate="44100"/>
  <!-- Local Host path-->
  <event URL="file:///Users/mattia/samples/smiling.aiff"
    start="23456" end="234556" sampleRate="44100" recDistance="10"/>
  </soundConcept>
  <!--...many concepts -->
</soundscapeDatabase>
```

Currently the XML annotation should be created using an XML editor. His step will be extended with a web GUI interface where users can easily select events, segment interesting audio file regions and export the annotation. The web interface shall make use of the Freesound [9] repository using an extended concept based search we previously developed [10].

The second format is “annotated on-site recordings”. It is an annotation specifically for users that are interested in recreating the sonic environment of a real location. The annotation procedure is achieved using SonicVisualiser³. This software allows to create annotation of “Region” layer on audio file. Users have to manually select and label the segments in the recordings that belong to a concept with the associated name he previously defined. The annotation is exported as CSV file. The generative module creates an event for each concept segment annotation and it counts the number of occurrence per concept computing its probability.

4. GENERATION MODULE

The generative software is implemented in the audio programming language SuperCollider (SC) [11], which features a high-level, object-oriented, interactive language together with a real-time, efficient audio server. In the system documentation web page [1], we provide the gener-

² A sample path expressed as URL is in the form [subprotocol]://[node]/[sample path].

³ An application for viewing and analysing audio files content <http://www.sonicvisualiser.org/>

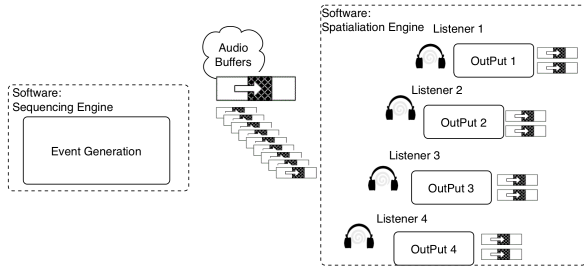


Figure 4. Sound events sequencing engine and spatialisation engine

ative software as a SC library, that uses two other external libraries in turn: “GeoGraphy” for the sound object sequencing management [8] and the “XML” SC library that implements a subset of the DOM-Level-1 specification for XML parsers.

As the application is designed for being a streaming web service module it presents high performance technical features: low requirements of CPU and RAM, high stability and robustness. Thanks to the sequencing engine that streams sound directly from disk using a small buffer window (32768 frames), the system allows for a minimal RAM usage while simultaneously playing dozens of audio files. The sequencing engine also manages the sound event generation process driven by graphs: each time an actant triggers a new sound object, the sequencing retrieves the associated sample from the database and adds it to the list of active sound object initializing its streaming buffer. The engine applies to each activated sample a control envelope that performs a fade in and a fade out proportional to the sample duration. The spatialisation engine manages the list of listener objects: each listener processes the list of active buffers, creates its own temporary copy of the buffers and performs its relative soundscape mix considering its position and orientation. The spatialisation engine produces a stereophonic stream but being split from the sequencing, the engine could easily re-use external libraries for other spatialisation set-ups or models. Both engines deal with multi-threading events synchronization.

The application implements parsing methods that initialise the soundscape software representation starting from the user space, sequencing and database annotations in the several formats previously discussed. During the initialisation phase the application performs the automatic graph generation per each concept. For continuous concept, it creates a graph whose edges have durations proportional to the duration of the vertex (i.e. the sample) they start from, so that the resulting triggering path will generate a continuous stream of concept events. The durations are not equivalent but proportional because the edge duration is slightly shorter in order to superpose events and to provide a cross-faded concatenation. The arrhythmic generation (AR) is a parameter that controls the number of input and output edges from each graph vertex. The number of edges is balanced: the system always sets the connection between the vertices that are less connected. The more

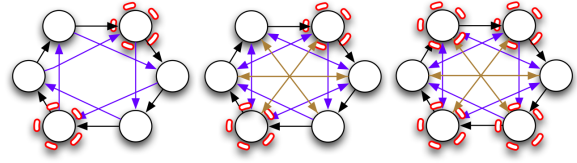


Figure 5. The type of sequencing for a 3 graph having the same global probability = 80, just changing the AR and MP parameters

edges each vertex has the less predictable the sequencing will be.

For a non continuous concept, AR also represents the irregularity in the sound object generation process. In fact, for this type of concept, the edge duration is proportional to the general concept probability and the multiple path parameter (MP), considering that each further actant doubles the effective probability. The edge durations are computed recursively, in the sense that each further cycle of edge connections between vertexes have the same values. Considering the simple case where edges probability are uniform, this means that edges probability are equal to the number of output edges from a vertex (i.e. $1/AR$). Starting from the case where $AR = 1$, each vertex has just one output (and input) edge, this means that all the pauses between sample activations are the same and the triggering pattern is completely regular, like a metronome. The number of vertices (i.e samples) a concept has, the less repetitive and poor in nuances its generation will result. First we consider the effective probability p taking into account the actual number of actants, which differs from the declared probability p_d provided by the user in the concept parameter:

$$p = \frac{p_d}{MP}$$

Thus, the duration of all graph edges per $AR = 1$, considering d = edge duration in second and $h = 3600$ seconds, is:

$$d = \frac{h}{p}$$

If $AR = n$, vertexes has n number of input/output edges and each further cycle i of edge creation has a duration that is proportionally lower than the previous one, always granting the overall probability:

$$\sum_{i=1}^{AR} d_i = \frac{h}{p}$$

where

$$d_i = d_{i-1} * 2^i$$

Summarizing, in figure 5 and in table 1 we show how the sequencing can generate completely different triggering pattern having the same global probability. In the example $p = 80$ occurrences per hour.

Once the application parsed the annotation, thus creating the soundscape and the graph associated to each concept, the audio generation starts using the sequencing and

MP	AR	d_1	d_2	d_3	d_4	d_5
2	2	120	60			
2	5	58	29	14.5	7.25	3.62
4	5	464	232	116	58	29

Table 1. The edge duration d_i per each further creation cycle i with AR the number of cycle (input/output edges per vertex)

spatialization engines. At this stage the application provides a OSC interface along with a GUI to allow both user or external application to interact in real-time with the graphs and the synthesis models controlling the soundscape performance. Externals requests can:

- Create listeners
- Update listeners position and orientation
- Control zones amplitude
- Control concepts amplitude
- Control concepts probability requiring more multiple generation processes (i.e. increasing MP)
- Control concepts probability requiring to change the edge durations (i.e. recomputing the graph)
- Update concept position
- Create pre-set: map a global soundscape status to an high-level general description

Pre-sets allow to save all the soundscape parameters (positions, mix and probability) mapping them to an high-level description that could be activated on request in real-time. For instance it is a powerful tool to model typical time period or particular state of the soundscape, like: night, day, windy, holiday, winter and so on.

As a conclusive note, Iain McGregor⁴ conducted a qualitative experiment [12] to test the perception of spatialisation in a Canary Island soundscape use-case generated with the proposed application. The experiment was conducted with 20 listeners. The results were very positive but the main problem observed was the perception of distance. Often users wrongly perceived very far sound objects considering them closer than the actual soundscape design. We believe that this issue relies on the lack of reverberation cues in our spatialisation engine that just uses the implicit audio sample reverberation without computing a procedural re-synthesised reverberation based on distance and on the acoustic space features specific to each zone.

5. STREAMING MODULE

The generative application runs on a public server hosted by the UPF at <http://mtg110.upf.es>. The server

⁴ PhD Student at School of Computing Edinburgh Napier University working on a mapping tool for comparing auditory display interfaces with listeners experiences

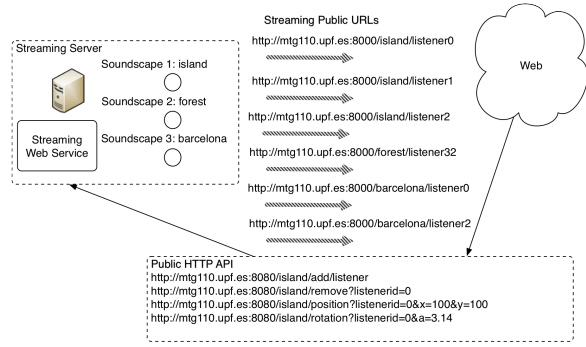


Figure 6. Streaming server and public API

service is developed using Icecast⁵ as audio streaming server and the python based Twisted framework to manage the networking engine⁶. This module delivers the produced soundscapes as MP3 streams and it implements a public interface for the SuperCollider application that allows for easily access, interact and control the soundscape generation through public HTTP API. The server has 2 types of API, one that creates new stream on demand and one that modifies the streams. Each new soundscape has its own streaming URL, then each “Add Listener” request creates a nested streaming URL as figure 6 shows. The “Remove listener” request frees the streaming resource. Instead the “Listener position update” requests, that controls the spatialisation engine, and the “Performance update” requests, that controls the sequencing engine, are translated on OSC and routed to the generative application.

6. CONCLUSIONS

There are not completely similar approaches in soundscape generation. What we propose is an innovative work, both in the idea of providing a web service for delivering listener based spatialised soundscape, and an interface to simplify the soundscape design also encompassing user-generated recordings. An interesting work that proposes an automatic sonification system of geographic sound activity controlled by acoustic, social and semantic informations is [13], but it is not aimed to high quality sound design standard and the system can not be used directly in multimedia productions, in spite of being an interesting sound events geographical browsing application. Outside from the environmental sound context, [14] presents a powerful tool to control the generation of musical events from a web map interface, mapping geo-related data to a musical interface. Furthermore, a related approach that focuses on the techniques for environmental sound retrieval is [15] but it does not take into account the generation part.

We imagine that possible clients of our system are Virtual Social Worlds (e.g we design a dedicated sonification in Second Life for a Metaverse project use-case, see section 7), Online games, web Mapping Viewer (e.g. Google maps) and augmented reality tourist application delivering

⁵ <http://www.icecast.org/>

⁶ <http://twistedmatrix.com/>

Soundscape Radio through the GPS technology (e.g. cars, mobile)⁷. Also architectural rendering for project presentation or urban planning makes use of video game technology to create 3D real time graphics simulations and do not consider the audio cues. Finally expert users, sound designer and composer, could also deploy and integrate the system in concert, sound art or interactive installation using the generative module as a local application.

The proposed system provides a standard composition format, making soundscape design more easily accessible and controllable by others applications or clients.

6.1 Future steps

An online platform that dynamically manages several generative module instantiations along with a soundscape web graphic editor are the future steps to realise the soundscape web service. From this platform clients could explore the already available composition scores searching for a desired space, and in case of need, proceed with a dedicated new design.

7. ACKNOWLEDGMENTS

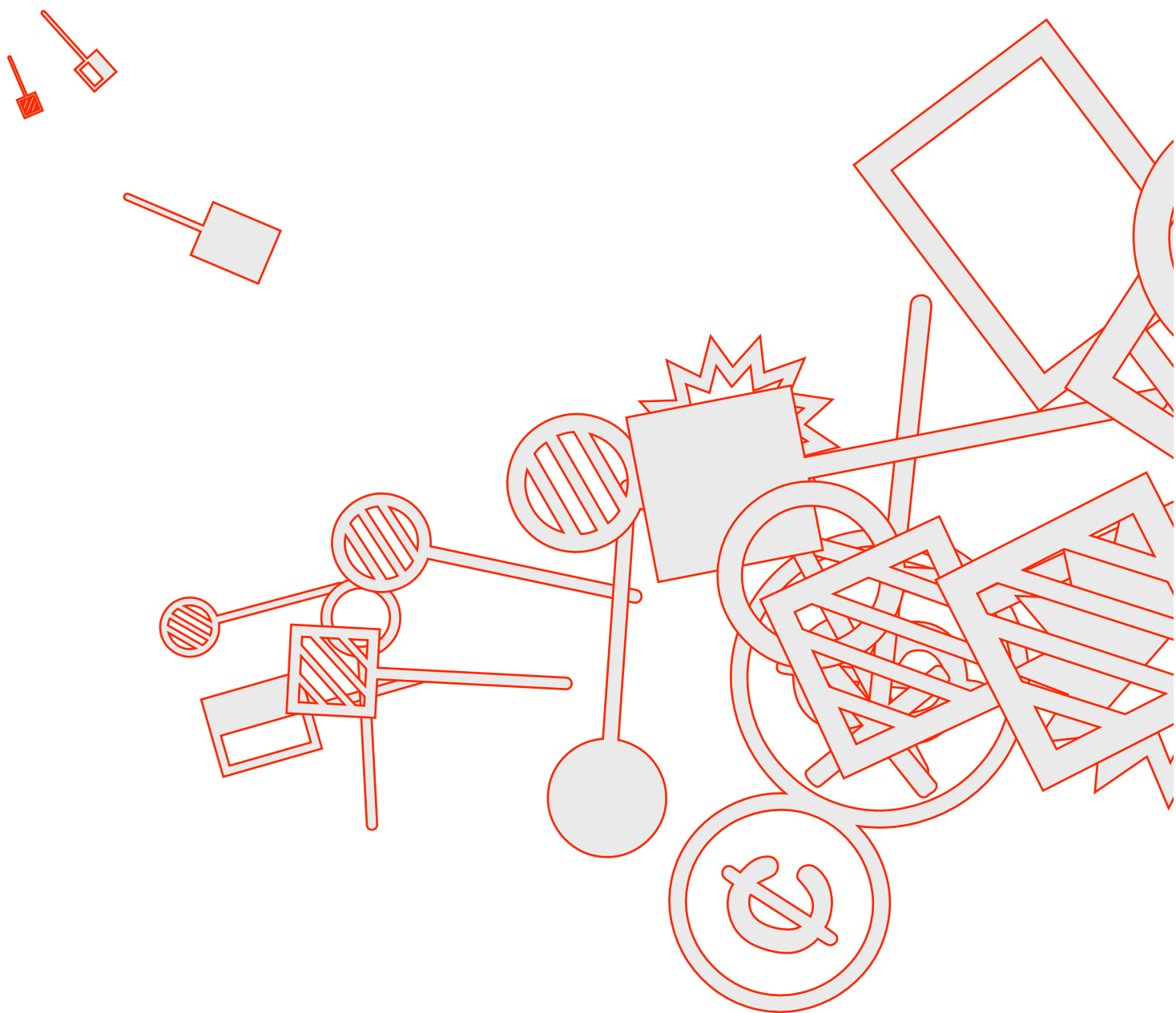
This work was partially supported by the ITEA2 Metaverse1 Project <http://www.metaverse1.org>.

8. REFERENCES

- [1] Universitat Pompeu Fabra, "<http://www.mtg.upf.edu/technologies/soundscapes>," 2010. MTG Soundscapes project web page with documentation and resources.
- [2] A. Farnell, *Designing Sound - Practical synthetic sound design for film, games and interactive media using dataflow*. London: Applied Scientific Press Ltd., 2008.
- [3] S. Kersten and H. Purwins, "Sound texture synthesis with hidden markov tree models in the wavelet domain," in *The 7th Sound and Music Computing Conference*, (Barcelona), 2010.
- [4] D. Schwarz and S. Norbert, "Descriptor-based sound texture sampling," in *The 7th Sound and Music Computing Conference*, (Barcelona), 2010.
- [5] Google, "<http://code.google.com/intl/ca/apis/kml/>," 2010. KML format.
- [6] A. Valle, V. Lombardo, and M. Schirosa, *Auditory Display 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009. Revised Papers*, vol. 5954, ch. Simulating the Soundscape through an Analysis/Resynthesis Methodology, pp. 330–357. Berlin: Springer, 2010.
- [7] R. Murray Schafer, *The Tuning of the World*. Toronto and New York: McClelland & Stewart and Knopf, 1977.
- [8] A. Valle, "Geography: a real-time, graph-based composition environment," in *NIME 2008: Proceedings*, pp. 253–256, 2008. Antonio Camurri and Stefania Serafin and Gualtiero Volpe.
- [9] Universitat Pompeu Fabra, "<http://www.freesound.org>," 2005. Repository of sounds under the Creative Commons license.
- [10] G. Roma, J. Janer, S. Kersten, M. Schirosa, and P. Herrera, "Content-based retrieval from unstructured databases using an ecological acoustics taxonomy," in *Proceedings of the International Community for the Auditory Display Conference*, 2010.
- [11] S. Wilson, D. Cottle, and N. Collins, eds., *The SuperCollider Book*. Cambridge, Mass.: The MIT Press, 2008.
- [12] I. McGregor, G. LePlâtre, P. Turner, and T. Flint, "Soundscape mapping: a tool for evaluating sounds and auditory environments," in *The 16th International Conference on Auditory Display*, (Washington D.C.), 2010.
- [13] A. Fink, B. Mechtley, G. Wichern, J. Liu, H. Thornburg, A. Spanias, and G. Coleman, "Re-sonification of geographic sound activity using acoustic, semantic, and social information," in *The 16th International Conference on Auditory Display*, (Washington D.C.), 2010.
- [14] S. Park, S. Kim, S. Lee, and W. S. Yeo, "Composition with path: musicla sonification of geo-referenced data with online map interface," in *Proceedings of the International Computer Music Conference (ICMC)*, (New York), 2010.
- [15] B. Mechtley, G. Wichern, H. Thornburg, and A. Spanias, "Combining semantic, social, and acoustic similarity for retrieval of environmental sounds," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 2402 – 2405, 2010.

⁷In this sense an interesting example is <http://www.soundwalk.com/>

4. Distribuzione tra prossimità e distanza



+ Marco Gasperini
+ Federico Costanza, Marco Gasperini, Alessio Rossato

THEORETICAL FOUNDATIONS OF A *LAPTOP ORCHESTRA*

Marco Gasperini

Meccanica Azione Sonora
marcogsp@yahoo.it

ABSTRACT

In this article some theoretical issues about the foundations of a laptop orchestra will be presented, prompted by the actual involvement in the design of the *S. Giorgio Laptop Ensemble* in Venice. The main interest is in the development of an *eco-systemic* [10] kind of communication/interaction between *performers*. First the main phenomenological features of *orchestra* and *laptop* will be reviewed, followed by a summary of the history of *laptop orchestras* up to now. A coherent system according to the premises will then be developed defining what will be an *orchestral player* and what will be a *conductor* and how they will be musically interacting. Technical issues regarding the set-up of the orchestra and the means of communication between the *elements* of the orchestra will then be approached. The most versatile solution for this last issue (audio communication) has been found exploiting possibilities of Local Area Network communications on ethernet interfaces. Finally the partial results of the research will be summarized.

1. INTRODUCTION

Based on recent research in the development of laptop orchestras [5] [13] [19] [21] [22], digital ensembles and diffused network performances [1] [3] [6] [7] [12] [14] [24], a laptop section has been developed according to the hereafter described architecture. A similar musical network has already been developed and tested in the frame of the *S. Giorgio Laptop Ensemble* in the course of the 2009 in Venice [8]. The author will try to show how to develop from scratch a musical ensemble made up by laptops, taking care of evolving from the basic consideration on the *instruments* and inserting the least possible number of heterogeneous elements.

Copyright: © 2010 Last name et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2. ORCHESTRA VS. LAPTOP

Orchestra is a Greek word whose etymon may be traced back to the verb *to dance*. According to this, it was used to name the space between the stage and the audience in the ancient Greek theatre, a space where there used to take place the performances of the *chorus*.

As for the use of the word in our days, it is possible to trace some constant features:

- it is a relatively wide instrumental group (there are no voices);
- it is divided in sections of instruments of similar timbres or of similar playing techniques (strings, woods, brass, percussions);
- it generally needs a director.

Starting from these considerations it can be said that an orchestra is characterized by a specific sound quality which depends on the *mix* of all the instrumental sounds produced by the musicians, that is to say by all the different timbres of the instruments, by their distribution in the performance place and, last but not least, by the quality of the performers and of their instruments.

An orchestral player is responsible:

- of the playing of his/her *part*, as close as possible to the indications contained in the score;
- of the musical quality of the produced sounds.

Each orchestral should have developed, through years of practice and studies, technical and musical skills in a very high degree on an instrument which, besides a more or less wide range of variations of its timbre and playing techniques, sounds always the same, it always manifests its *systemic identity*, and for this reason is recognized.

Beside all this the musicians are subject to a performing stress caused by the care in following the score and the director, and a physical stress caused by the playing position and by the gestural expressiveness typical of each instrument. It can be said that the good quality of an execution depends on the way a similar stream of strains (by no way to be intended in a negative sense: a string must be *strained* to be musically useful) is musically managed.

The director is responsible of this and of all the other global sounding factors of the whole ensemble:

- the acoustic mixture, which depends on the intonation, the use of the playing techniques of the various instruments, the dynamic balance between sections, etc.;
- of the rhythmic precision and synchrony which is a fundamental feature for the perception of the orchestra as a unified sounding body, internally articulated;
- of the actual realization *in sounds* of what is contained as *signs* in a musical score.

The figure of the director, as has been defined from the nineteenth century onward, does not play a sounding part in the performance, that is to say that he does not have a musical part in the score but he has to control and manage all the parts.

As for the first term it is a definitely primitive term: *laptop* is the abbreviation for *laptop computer* which is referred (as his predecessor *desktop computer*) to the position in which a portable computer may be held.

As for its acoustical and playing qualities a laptop (cfr. [22]):

- is a neutral instrument (or, said otherwise, *passive*), not directly linked to a specific intrinsic sonority (apart from that of the CPU fan or that produced by pressing the keyboard);
- is potentially ever-changing: the algorithm used for the local production of sounds may change during the same performance, making it impossible to univocally link the source and its acoustical appearance, even for the fact that its actual sound source (the loudspeaker/s) may be placed far from the sound-computing engine;
- does not need the development of specific playing techniques, at least not comparable to the instrumental technical skills requested to an orchestral player;
- it does not induce any physical nor performing stress. About this fact, Dan Trueman reports the critics made by a spectator of a performance of the *Princeton Laptop Orchestra*: “As far as I could tell, they were all just checking their e-mail”[22];
- in any case there is no direct nor necessary link between the gestural expressiveness and the sounding result.

3. A BRIEF HISTORY OF LAPTOP ORCHESTRAS

3.1. From the Eighties...

With the term *laptop orchestra* is intended a reasonable group of musicians (*orchestra*) each of them exerting his/her musical skills by means of a portable digital computer (*laptop*). This kind of musical group may be seen as one of the possible outcomes of the musical

research developed through technological instruments in the last seventy years.

This perspective was foreshadowed (how much unconsciously?) in the first historical article on computer music made widely available: “The Digital Computer as a Musical Instrument” by Max Mathews [16].

The first experiments in this field were carried on as soon as personal computers were made available on the market, in the second half of the Seventies; this large availability made it possible to use the computers outside of the great institutions and laboratories. At least the work of *The League of Automatic Composers* and *The HUB* [2] should be cited how the first examples of musical ensembles made up exclusively by a *network* of computers, considered as an infrastructure for communication between digital machines. The main interesting fact is the shift of the function of the computer from a *compositional instrument* to a *performance instrument*: the members of these ensembles must design the network of relationships which should exist between their devices, according to the performance strategies (and so *instrumental*) they have adopted; moreover they should design their sound production algorithms which they will be *playing* and not merely use to produce sounds to be recorded on tape.

3.2. ...to the Third Millennium

However not before the XXI sec. was it possible to conceive a *laptop orchestra* for the wide diffusion of the technological facilities (network communication, wide bandwidths, stable and quite standardized protocols, powerful processors, compatibility issues, etc.).

Now some *laptop orchestra* experiences will be reviewed.

The most active make-up, both for the documentation produced on their activity and for the public performances, is the *Princeton Laptop Orchestra* (PLORK), founded in 2005 by Dan Trueman and Perry Cook [19] [22]. The main features of this ensemble, compounded by approximately fifteen people, are:

- the orchestra is placed on the stage;
- each member has its own laptop, controllers and sound diffusion system (a six channels hemispherical speaker, permitting different patterns of sound radiation) placed nearby;
- the overall sound of the orchestra is localized;
- control signals may be transmitted through a Local Area Network by means of the Open Sound Control (OSC) protocol [25];
- on some occasions acoustical instruments have been included in the performance.

The composer who wants to write for them should at first imagine and design a possible interaction between the performers and then design the environment itself which will become the *instrument* of an orchestral. The compositions written for PLORK use the figure of the conductor in heterogeneous ways.

Others *laptop orchestras* whose activity may be connected with that of PLORK are the *Oslo Laptop Orchestra* (OLO), founded in 2007¹, and the *Stanford Laptop Orchestra* (SLORK), founded in 2008 [23].

A different attitude is that of the *Milwaukee Laptop Orchestra* (MiLO), which defines itself as “a collective of musicians and artists dedicated to live audio/visual performance”². It is a non-hierarchical ensemble, inclined toward the free improvisation, in which each member is free to bring his/her musical approach, may use the software of his/her choice, is free to play alongside acoustical instruments as well or use some video processing. They developed custom network facilities with the software pure-data [19] to improve the interaction between performers, named NRCI [5] [21] and based on the protocol OSC. This facilities are used to broadcast to the orchestra the performance data of each performer: rhythm, melodic envelope, amplitude, etc.; each member is free to choose which data to receive and how to relate the received data to his/her sound processing.

Others *laptop orchestras* are: the *Carnegie Mellon Laptop Orchestra* (CMLO), founded by Roger B. Dannenberg, addressed to didactical issues [9]; the *Worldscape Laptop Orchestra* (WLO), the largest ensemble of this type, set-up for just one concert [13]; the *Linux Laptop Orchestra* (L2ORK)³⁴.

Finally I would like to mention the *S. Giorgio Laptop Ensemble*, founded in Venice in 2009⁵, who gave its first concert in 11.07.2009 at the Foundation “G. Cini” in the island of S. Giorgio Maggiore [8]. This experience has gave birth to the research which will be here presented. The work of this group is witnessed by a printed score for string quartet and laptop ensemble⁶.

4. THE LAPTOP AS A MUSICAL INSTRUMENT

A laptop musical ensemble will be, by definition, an ensemble of musicians (*performers*) each using a laptop computer, leaded by a *conductor*. The computers are connected to form a network of musical information.

¹ **OLO**: <http://fourms.wiki.ifi.uio.no/Oslo_Laptop_Orchestra/> [10.1.2011]

² **MiLO**: <<https://pantherfile.uwm.edu/gssurges/www/milo/>> [10.1.2011]

³ **L2ORK**: <<http://l2ork.music.vt.edu/main/>> [10.1.2011]

⁴ This list is not intended as a complete catalog nor a personal choice between experiences. It is a brief review of the major significant contribution founded in the scientific literature.

⁵ Founding members of *S. Giorgio laptop ensemble*: Stefano Alesandretti, Federico Costanza, Marco Gasperini, Marco Marinoni, Monica Karina Lopez Lau, Luca Richelli, Alessio Rossato, Julian Scordato, Paolo Zavagna.

⁶ “*WAV*, *Ad Alvisé Vidolin per il suo 60° compleanno*, for string quartet and laptop ensemble”, Avenza, Italy, Ars Publica, 2009. The score includes five *fragments* for string quartet and the description of the whole performance environment. Titles and composers are: *Osso*, Federico Costanza; *Schegge*, Marco Gasperini; *13.glifi.exit*, Marco Marinoni; *Inclinato*, Alessio Rossato; *Interlinea*, Julian Scordato.

The whole complex of algorithms used locally by a computer will be called *instrument*; each *instrument* will be *played* by the *performer* through some variable control parameters. One element (built-up by a computer and an *instrument*, *played* by a *performer*) of the network will be called *client*, while the *conductor*’s instrument will be called the network’s *server*.

The aim is that to create a system capable of some autonomous dynamics in sound starting from the intrinsic features of the basic elements. To this end it is necessary to consider the devices in their true nature, that is to catch their phenomenological *essence*; doing so will permit, in a constructivist view, to develop consequently and coherently an ensemble. Although it was not explicitly present in the premises of this work, the outcome get very near to the research on sound *eco-systems* carried on by the Italian composer Agostino Di Scipio [10] [11].

Let’s consider the setting-up of an audio network with an input signal $x(t)$. In relationship to an audio input signal a computer may be considered as a delay unit with an arbitrary transfer function which, in the simplest case, would be an unitary pulse: in this case the system will be linear and its output, given a fixed delay of N samples, will be:

$$y(t) = x(t-N) \quad (1)$$

The delay is due to the size of the memory buffer (I/O vector) used by the audio interface to carry on the A/D and D/A conversions. The I/O vector is linked to the CPU clock: given a same amount of operations the I/O vector size could be decreased as the processor power will increase.

5. BUILDING THE ENSEMBLE

5.1. Step one

By placing in a row several elements as that previously described an elementary hierarchical chain of N computers could be obtained (see Figure 1). The first element **S** (*server*) generates the signal which pass through each client. Each element **Cn** has a local output **outCn** which reflects the instantaneous state of the *client*, and will introduce in the audio chain a certain **dCn** samples delay.

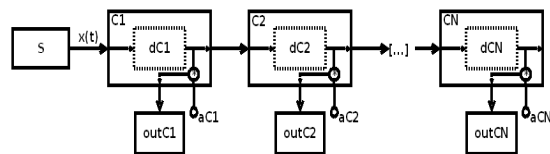


Figure 1: audio chain of N clients with control over the amplitude of the local output

The choice of a single source of audio signal is due to considerations about economy of means and unity of results: given a certain audio input each *client* will repeat it after a certain delay time and the whole ensemble will emerge as the succession of a same acoustical sign (*canone all'unisono*).

5.2. Some improvements

According to considerations about the lack of vitality at both global and local levels in the audio chain of Figure 1 the system thus obtained must be improved. Particularly the need is felt for the performers to exert some control on the qualities of the sound they produce (not only on loudness).

A possible improvement of the *client* behaviour is shown in Figure 2, where internally to each *client* a feedback has been applied, said *local* feedback for it is exerted at the *client* level.

A system as that in Figure 2 is said an IIR comb filter [26], described by the well known frequency response. It should be noted that this response displays resonant peaks and attenuation valley: given a certain amount of **dCn** samples of delay (*local* I/O vector) at the sampling rate sr , the system will have **dCn** peaks evenly spaced in the whole spectrum from 0 Hz to $sr/2$; the distance ΔP between peaks will be inversely proportional to the delay **dCn**, according to the following relationship:

$$\Delta P = sr / dCn [Hz]. \quad (2)$$

The peaks are thus in harmonic relationship to a fundamental frequency ΔP . The relationship between peaks (**pCn**) and valleys (**vCn**) of the frequency response, which affects the perception of the zones of resonance, depends on the gain factor **gCn**:

$$pCn/vCn = (1+gCn) / (1-gCn). \quad (3)$$

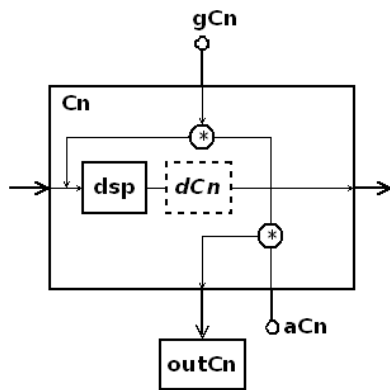


Figure 2: client with local feedback and control over the gain factor **gCn**.

According to the actual gain value the *client* will have a stronger or weaker presence of its resonances (i.e. on timbre).

By imposing *local* feedbacks we have thus obtained two instrumental possibilities controlled by a single parameter (gain):

1. *clients* relative independence to input signal (*resonance*);
2. timbric presence of the *instrument* determined by the fundamental frequency ΔP of the comb filter and by ratio (3) above.

It should be noticed how systems as that in Figure 2 are the building blocks of what is known as physical modelling synthesis or, more generally, waveguides synthesis [4] [15] [26].

In a system composed of such elements emerges a first possibility of proper subdivision of the ensemble in several resonance sections, determined by a contingent factor (I/O vector size), and unrelated to the actual signal processing locally implemented. Moreover a further improvement may be that to configure each section of the ensemble according to some physical model of resonance (for example: string, pipe, plate, etc. See Figure 3).

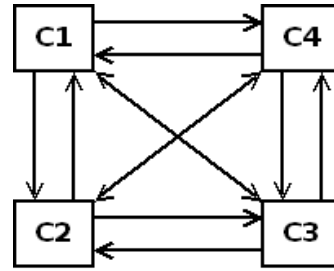


Figure 3: a section of the *laptop orchestra* arranged to form a simplified model of a resonant plate [17].

5.3. Stability

The stability of the designed networks is not obtained by resolving the proper polynomial of the particular system but with the use of non-linear energy compensation functions which may be traced back to dynamics compression issues [4] [10] [26].

Before being transmitted to the next *client* the envelope of the output signal is extracted and its ones' complement is used to compensate the actual output (see Figure 4). The particular technique of envelope extracting depends on two variable factors:

1. t : delay [sec.], memory of the envelope extractor;
2. g : feedback gain [0, 1], sensitivity to positive derivative of input signal (growth of energy) and length of the decay curve.

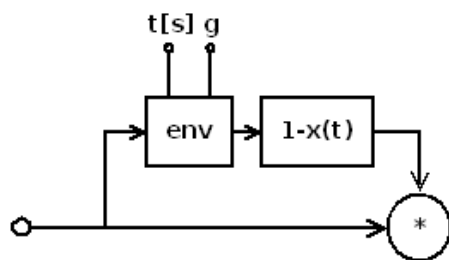


Figure 4: energy compressor.

These two variables have been unified in a single factor called *sensitivity* (**sCn**), directly proportional to both variables on a 0 to 1 scale.

The sensitivity could thus be used as a measure of the dynamicity of an instrument, as is the case with instrumental sections in a classical ensemble where some of them are, historically, characterized by a greater versatility (for example the strings' section) than others (for example the brass section).

5.4. The complete *client*.

The complete implementation of a single *client* can be seen in Figure 5. The performer will be able to control at least three parameters of his instrument (see Table 1), in addition to those proper to the digital signal processing (*dsp*) algorithm locally implemented. These three parameters may be related to the musical parameters of intensity (*local* loudness, **aCn**), rhythmic articulation (*local* sensitivity, **sCn**) and timbric presence (*local* feedback gain, **gCn**).

For the *local* output of the laptop performers non-standard transducers may be used, made by the coupling of an acoustic magnetostrictive actuator [20] with a resonant surface. These actuators may be used to put in resonance any surface on which they are placed (wood, glass, metal, etc.) and are currently widely used for environmental sonifications.

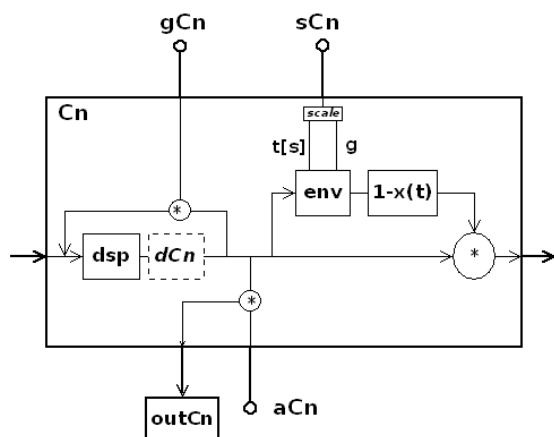


Figure 5: *client*'s architecture

By using different resonators there will be different colouring of the excitation audio signals and the

possibility to distinguish each performer by the particular timbre of its resonator, independently from the actual signal processing locally implemented.

Variable	Exerted Control	Perceptible effect
gCn	<i>local</i> feedback gain	<i>Resonance</i> to input signal and timbric presence
sCn	<i>local</i> sensitivity	Rhythmic articulation of the audio stream
Can	<i>local</i> loudness	Loudness produced on the <i>local</i> output

Table 1: *client*’s parameters.

Besides this controls each performer should manage the parameters of the signal processing carried on locally on his/her machine. This will be a second order level of the orchestra design: among the sound elaboration tools used in the *S. Giorgio Laptop Ensemble* there were *vocoders*, granular processing, synthesis tools, transpositions, etc., all acting on the incoming sound produced by the string quartet.

6. THE DIRECTOR

Having thus defined the *clients* level it is felt the need to improve accordingly the *server* level. Up to now the only function of the element **S** has been that to feed the network with an audio signal. For a conductor to exert its control over an ensemble it is necessary that (s)he knows and hears the players' parts: to obtain a similar condition in the network of audio signals implemented, the *clients* should feed-back their output signal to the *server*. This will be called *global* feedback whose gain will be controlled by the director. The stability of this system is obtained at this global level by means of non-linear compensation functions as it is at the *local* level, and will be called *global* sensitivity (**sS**) (see Figure 6).

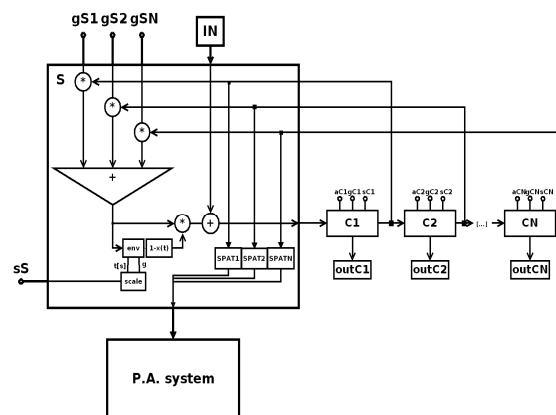


Figure 6: the complete audio network.

The conductor will also be responsible of the *global* sound of the ensemble diffused via a standard P.A. system (*global* loudness level). The global signal will not be exactly as the sum of the performers' *local* outputs because of the differences in the transduction elements which are not transmitted in the audio stream. This differences will emerge as spatial dynamics (*local* output vs. *global* output) due to timbric factors and not on any pre-arranged simulation of sound movement.

The laptop performers and the conductor will thus exert their control by means of the same kind of variables which have, however, different effects depending on the level at which this controls are exerted (see Table 2).

Laptop Performer	Conductor	Laptop Performer
Gain	Timbric presence	Formal articulation
Sensibility	Rhythmic texture	Dynamics of the stream
Loudness	Local loudness	Global loudness

Table 2: comparison between *client* and *server* parameters.

The *server* should also manage the actual configuration of the orchestra, namely the nodes of communication between *client* and *client*, in order to configure the sections, for example, as a simple row of *clients* (see Figure 1) or as a particular physical model (see Figure 3).

As for the **IN** signal fed to the network by the *server*, there are some possible observations:

- it could come from some acoustical source (one or more instruments, ambience noise, etc.);
- it could be read from an audio file;
- it could be synthesized on some *node* of the network;
- different signals may be sent to different *nodes* of the network, in order to feed different section with distinct signals.

7. TECHNICAL DETAILS

In the *S. Giorgio Laptop Ensemble* a similar network has been implemented with audio signals being broadcasted on a Local Area Network using UDP protocol on ethernet interfaces. After some test the more reliable solution was that of using an open source VST plug-in (*wormhole2*¹) to manage the audio communication between *server* and *clients*. This choice was made to avoid the complexities of a physical set-up using audio interfaces, and musically managing the shortcomings of network communications (intrinsic delays).

The network has been used even to broadcast control data streams carrying information on main features of the audio stream (input signal envelope, *global* audio envelope, etc.) using the OSC protocol; to avoid conflict with audio signals broadcasting, the control data were sent through a wireless network.

Each orchestral will thus be equipped:

- n.1 laptop (with an ethernet interface);
- n.1 acoustic actuator;
- n.1 resonance surface;
- n.1 ethernet cable;
- n.1 MIDI controller.

The director:

- n.1 laptop (with an ethernet interface);
- n.1 ethernet cable;
- audio interface;
- n.1 MIDI controller.

For the whole system to work are needed:

- n.1 Public Address System;
- one or more ethernet switch(es) for audio communication;
- n.1 wireless access-point for data communication.

8. FINAL REMARKS

Control parameters have been founded which are relevant to the nature of the network: this parameters are in a certain degree automatic, but their actual response depends on the combined actions of performers and conductor. This first control level regards audio communication between laptops, and to describe it have been used terms borrowed from network technology: the *global* behaviour will be managed by the *server* that will also have to grant the stability of the whole audio network (avoiding saturation), of its configuration (connecting *clients*) and feeding the *clients* with an audio signal. The *clients* will manage the modality of their own *local resonance* and *stability*. The *server* level is managed by the conductor while the *clients* level by the performers: this is the second control level, the human level of interaction which is not automatic, depends on performance practice and must be finely tuned through rehearsals. It should be noted that the *server/director* does not directly take part in the acoustical matter (i.e. sounds are produced/transformed by the orchestra) but just manages the relationships between the orchestra elements and the *global* features (loudness, rhythmic articulation, etc.) of the audio stream.

9. ACKNOWLEDGEMENTS

This research has been carried on independently by the author, who would like to thank Agostino Di Scipio, Riccardo Vaglini and Alvis Vidolin for their support.

¹ <<http://code.google.com/p/wormhole2/>> [10.1.2011]

10. REFERENCES

- [1] Barbosa, A. "Displaced Soundscapes: A Survey of Network Systems for Music and Sonic Art Creation", «Leonardo Music Journal», vol.13, 2003, pp. 53–59.
- [2] Brown, C., Bischoff, J., *Indigenous to the Net: Early Network Music Bands in the San Francisco Bay Area*, 2002, available at <<http://crossfade.walkerart.org/brownbischoff/IndigenoustotheNetPrint.html>> [10.1.2011].
- [3] Burk, P. L. "Jammin' on the Web—A New Client/Server Architecture for Multi-User Performance", *Proceedings of the International Computer Music Conference*, Berlin, Germany, 2000.
- [4] Burns, C. "Emergent Behavior from Idiosyncratic Feedback Networks", *Proceedings of the International Computer Music Conference*, Singapore, 2003.
- [5] Burns, C., Surges, G. "NRCI: Software Tools for Laptop Ensemble", *Proceedings of the International Computer Music Conference*, Belfast, UK, 2008.
- [6] Caceres, J. P., Renaud, A. B. "Playing the network: the use of time delays as musical devices", *Proceedings of the International Computer Music Conference*, Belfast, UK, 2008.
- [7] Costanza, F., Gasperini, M., Rossato, A. "Meccanica Azione Sonora: composing networked music", *Proceedings of the XVIII° Colloquio di Informatica Musicale (XVIII° CIM)*, Torino, Italy, 2010.
- [8] Costanza, F., Gasperini, M., Rossato, A. "S. Giorgio Laptop Ensemble:", *Proceedings of the XVIII° Colloquio di Informatica Musicale (XVIII° CIM)*, Torino, Italy, 2010.
- [9] Dannenberg, R. D. et al., "The Carnegie Mellon Laptop Orchestra", *Proceedings of the International Computer Music Conference*, Copenhagen, 2007.
- [10] Di Scipio, A. "Sound is the interface: from interactive to ecosystemic signal processing", «Organised Sound», 8(3), 2003, pp. 269-277.
- [11] Di Scipio, A. "L'émergence du son, le son de l'émergence", «Revue de l'Association pour la Recherche Cognitive», CNRS 1-2, n. 48-49, 2008.
- [12] Duckworth, W. "Making Music on the Web", «Leonardo Music Journal», vol. 9, 1999, pp. 13–18.
- [13] Harker, A., Atmadjaja, A., Bagust, J., Field, A. "Worldscape Laptop Orchestra", *Proceedings of the International Computer Music Conference*, Belfast, UK, 2008.
- [14] Kane, B. "Aesthetic Problems of Net-Music", *SPARK Festival Proceedings*, Minneapolis-St.Paul, USA, 2007.
- [15] Karplus, K., Strong, A. "Digital synthesis of plucked string and drum timbres", «Computer Music Journal», 7(2), 1983, pp. 43-55.
- [16] Mathews, M. V. "The Digital Computer as a Musical Instrument", «Science», 142(3592), 1963, pp.553-557.
- [17] Mikelson, H. "Mathematical Modeling with Csound: From Waveguides to Chaos", in Boulanger, R. (editor) *The Csound book*, MIT Press, Cambridge and London, 2000, pp. 369-384.
- [18] Puckette, M. "Pure Data", *Proceedings of the International Computer Music Conference*, Hong Kong, China, 1996.
- [19] Smallwood, S., Trueman, D., Cook, P. R., Wang, G. "Composing for Laptop Orchestra", «Computer Music Journal», 32(1), 2008, pp. 9-25.
- [20] G.A.V. SOWTER, "Soft Magnetic Materials for Audio Transformers: History, Production, and Applications", «Journal of the Audio Engineering Society», 35(10), 1987, pp. 760-777.
- [21] Surges, G., Burns, C. "Networking Infrastructure for Collaborative Laptop Improvisation", *SPARK Festival Proceedings*, Minneapolis-St.Paul, USA, 2008.
- [22] Trueman, D. "Why a Laptop Orchestra?", «Organised Sound», 12(2), 2007, pp. 171-179.
- [23] Wang, G., Bryan, G., Oh, J., Hamilton, R. "Stanford Laptop Orchestra (SLORK)", *Proceedings of the International Computer Music Conference*, Montreal, Canada, 2009.
- [24] Weinberg, G. "Interconnected Musical Networks: Toward a Theoretical Framework", «Computer Music Journal», 29(2), 2005, pp. 23-29.
- [25] Wright, M. "Open Sound Control: an enabling technology for musical networking", «Organised Sound», 10(3), 2005, pp. 193-200.
- [26] Zölzer, U. (editor), *DAFX - Digital Audio Effects*, John Wiley and Sons Ltd., Chichester, UK, 2002.

MECCANICA AZIONE SONORA: COMPOSING NETWORKED MUSIC.

Federico Costanza

Marco Gasperini
Meccanica Azione Sonora
meccanicazione sonora@lib
ero.it

Alessio Rossato

ABSTRACT

This paper presents *Meccanica Azione Sonora*, an electro-acoustic trio formed by Federico Costanza, Marco Gasperini, Alessio Rossato and its research into the development of new musical syntaxes, prompted by the use of technologies (digital signal processing, live electronics, network audio communication, non-standard sound diffusion systems, etc.). We will present a compositional strategy named here *functional music*. The use of this term must be considered in opposition to its current use indicating a music that is used with a specific (mainly practical) purpose. First a definition of *functional music* will be given and then it will be exemplified through the description of several of our compositions in which this compositional attitude has been developed.

1. INTRODUCTION

Meccanica Azione Sonora is an electro-acoustic trio formed in 2007 by the composers/instrumentists Federico Costanza (keyboards), Marco Gasperini (guitars) and Alessio Rossato (percussions) active in the composition/performance of electro-acoustic music. In these years of activity we have developed from free improvisation with live electronic processing of instrumental sounds to the definition of a compositional strategy named here *functional music*. Our musical works range from compositions for instruments and live electronics to audio visuals¹: despite the technical innovations that currently have a strong prominence in creating new sounds and forms, we wish to emphasize how the action of the musician/composer is increasingly important in contemporary music research.

2. FUNCTIONAL MUSIC

There are many evidences of the timbre being one of the main interest in the musical thought of the last 60 years and is, at least, one of the main field of research in

Copyright: © 2010 Last name et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

electronic and computer music issues. However an univocal definition of *timbre* seems somehow elusive [4] [6].

What we mean with *timbre* is the sign of a *systemic identity* (a definition borrowed from A. Di Scipio). A *system* is defined according to the following *spaces*:

1. *instrumentation* or *treatment of instruments* or, more generally, *treatment of acoustical matter*;
2. *sound processing tools* and their *intrinsic dynamics* (planned interactions, shortcomings, etc.);
3. *sound diffusion system*.

Once that a certain number of variables for each space has been chosen and that a *field of definition* for the *variables* of each of these three *spaces* has been *designed*, we have traced the *fields of definition* of a *musical function*, capable of forms generation² (it should be outlined that what we are dealing here with is not what is known as *algorithmic composition*). The *arguments* to this *function* are the more or less defined executive canvas which we set-up for the different *application* of the function itself.

A single function, being by itself complex, has its own formal possibilities, which may or may not be manifested in the actual musical work, whose formal development will be given, for this same reason, by the order of the *functionalities* made audible.

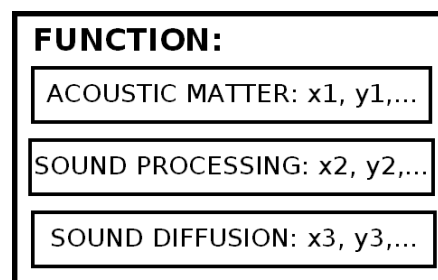


Figure 1: composition space for *functional music*.

It is not our will to apply mathematical terms inappropriately nor are we attempting to construct a musical theory out of mathematical issues, we are just borrowing a concept, as a suggestion, shifting its meaning into a musical domain to describe and better understand ourselves an emerging procedure of

¹ a selection of works may be found at the address <http://www.youtube.com/user/GruppoMAS>.

² For the definitions of *space* and *field* see *La costruzione del suono* (1963) and *Note di teoria della composizione* (1970) by G. M. Koenig in [9].

composition. We are aware that we must “ground musical systems exclusively on musical criteria”¹ [1]: our aim is thus to give a theoretical status to our musical research. A mathematical function is univocal: given a certain input value there’s just one possible result. This is for sure not the case in music composition or, at least, not our goal. The mathematical suggestions must thus be combined with the use of the term *function* in music theory to define the hierarchical and syntactical features of a chord in tonal harmony.

We will clarify our definitions with an example. We *design* the following *fields*:

1. **x1** = plucked strings;
2. **x2** = Delay; **y2** = Granulation;
3. **x3** = Loudspeaker(s).

The first field defines the use of a certain class of instruments for which a *performance* canvas should be written (detailed score or improvisation plan).

The second space has several possible *behaviours* according to the combination of the two variables **x2** and **y2**:

1. no processing;
2. Delay alone;
3. Granulation alone;
4. Delay and Granulation simultaneously;
5. Succession Delay, Granulation;
6. Succession Granulation, Delay;

We call these *functionalities*. Musical form is given by displacement of *functionalities*. More precisely a *functionality* is given by the combination of the actual *behaviours* of the several *variables* of the three *fields*.

We may consider the two terms *system* and *function* as synonymous, but with slightly different meanings: *system* gives much relevance to the objects that produce a *timbre*, i.e. *the instruments*, while *function* gives relevance to the *transformation* of something into something else, it suggests the *becoming*, movement and hence *form*.

Composing means to define *spaces for actions*, actual *actions* (both human and electronic) and succession of *behaviours* in each *space*.

Up to now we have defined three main *functions* which have been used to produce several works:

1. *sovra – posizione*²;
2. *butchery*;
3. *rumore – silenzio – metallo*³;

This list represent also a scale with three different degree of versatility of the musical *functions*: the first is the most flexible and adaptable (i.e. formally *open*), while the last is the most formally *closed*. They will be now described in reverse order.

3. RUMORE – SILENZIO – METALLO

With regard to works in *bianco/nero/metallo*⁴ by the venetian artist Emilio Vedova, we developed a multi-sensorial performance through intensive experimentation of sound, visual and gestural elements: the three musicians, using both an instrumental and electronic set, act on the entire space of the stage. Behind each player’s station there is a suspended copper plate illuminated by headlights. The musical layer interacts, like a *pas de deux*, with a video by Lino Budano (a multimedia artist and sculptor).

3.1. Instrumentation and sound diffusion

The instrumental set used in the *rsm* project includes:

- Percussions (including vibraphone, cymbals, snare drum);
- Guitars (electric and acoustic);
- 2 Toy pianos;
- 3 resonant copper plates;
- 3 magnetostrictive acoustic actuators;
- 3 laptops;
- 3 digital tracks (3 channels);
- Loudspeakers;
- Video.

In Figure 1 is displayed the on stage arrangement of performers and instrumentation (**AMP**, electric guitar amplifier; **L1-2**, loudspeakers; **P1-3**, copper plates with applied acoustic actuators; **SUB**, sub-woofer).

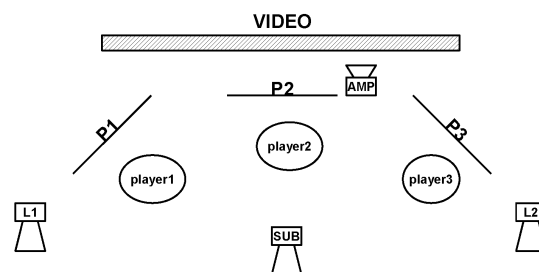


Figure 2: on stage arrangement for *Rumore – Silenzio – Metallo*.

For the sound diffusion, besides a standard public address stereophonic system with a central sub-woofer, three non-standard transducers are used, each made-up by the coupling of a magnetostrictive acoustic actuator [8] with a suspended copper plate (ca. 190x40 cm). The sound thus diffused will be *coloured* by the particular resonance characteristics of each plate.

3.2. The instrumental score

The musical score for this composition is divided in nine sections with the following possible *behaviours*:

¹ English translation by the authors.

² Super-position, overlap.

³ Noise – silence – metal.

⁴ *black/white/metal*. We refer to the cycles *Frammenti/Schegge* and *Arbitrii*, (1977/1978), *Laminati*, (1991), *Dischi Non Luogo*, (1987/1997).

1. digital tracks;
2. digital tracks and instruments;
3. instruments and live electronics;

Sections are completely written out but may comprehend some part of controlled improvisation (*bridges*), based on some defined instrumental gestures (see Figure 3). The *functionality* of *bridges* is to explore the actual *behaviour* of sound processing and diffusion, i.e. to let emerge the *timbre* of the whole *system*.

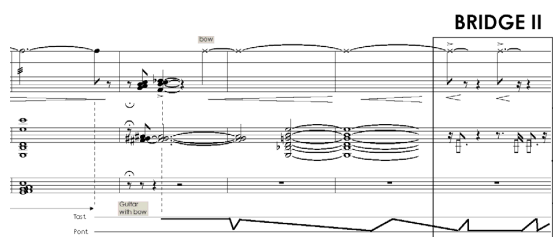


Figure 3: score for part A of *rsm* with improvisation bridge.

3.3. Digital tracks

There are three digital three tracks tapes to be played during the performance. Each track of the three tapes is played locally by *clients* on the respective copper plate (*client1*: track 1 on **P1**; etc.). The first tape is played at the beginning of the piece, as a prelude for the acoustical instruments. The second tape slowly emerges from the instrumental section **C**. The third grows out after a quite long silence in the final section. In this last section the plates became the only instruments of both the electronic and human layer of the trio as well as of the visual side of the performance. The performers leave, one at a time, the instruments used up to that point and act directly on the plates, by variously moving them; the video become progressively white, the light's reverb, caused by reflections of the stage lights on the plates, are the main attractions for the eyes. The action of the performers on the plates emerges as spatial movement of sounds because of the relative uniformity of sound materials between the three sources (filtered white noise). In the apex ten seconds of white noise conclude *Rumore-Silenzio-Metallo*(2009).

3.4. Sound processing

Each performer uses also a laptop computer which carries-on some processing on incoming signals (see Figure 4). Each computer has even one or more microphone inputs to capture the instrumental sounds and a direct local output to the copper plate (**Pn**). Moreover each *client* may send to and receive from the *server* two signals. On one of the three *clients* resides also the *server*, which is used to configure the network of audio signals by means of a *global* matrix. Incoming and outgoing signals may be routed internally by *clients* by means of a similar *local* matrix.

The sound processing *variables* are:

- Delay unit (**DEL**);
- Granular processing unit (**GRAIN**) [6] [10];
- Reverb unit (**REV**);
- Ring modulation unit (**RM**);
- Transposition unit (**TRANSP**);
- Vocoder unit (**VOCODER**) [7] [11].

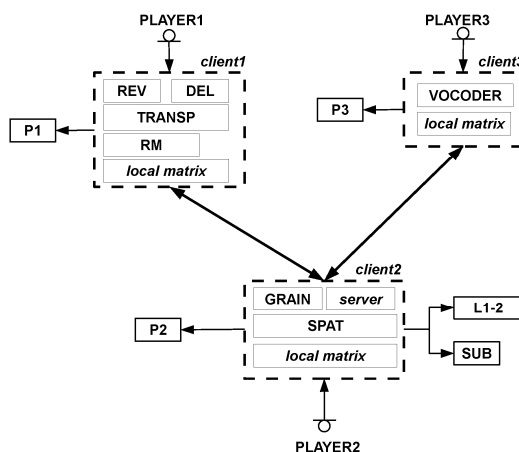


Figure 4: *rsm* sound processing field.

All these various processing modules will assume different configurations (*behaviours*) in the course of the piece. This succession together with that of the configurations of the audio network, forms the computers *score* [1]. The *part* of each computer is contained in its (*local*) memory as a list of commands; the reading into the score is controlled by the players in precise and determined attacks. The advancement of this *playlist* is synchronized between the three players by means of the OSC protocol [12].

Some word must be spent on the *spatializer* (SPAT): this unit is not used to simulate sound movements but to move sounds from one source to another, and since these differ so much in timbre, the unit is used to move sounds (*excitations*) between timbres. Due to different delays introduced in some path to the output units (remote from the *server* copper plates) the unit is non-linear.

Audio communication between *clients* and *server* is done by means of the UDP protocol, on an ethernet interface [3]; this kind of communication imply some delay in the transmission line (between 0.2 and 1 second) that have been musically employed to compose complex and non linear time/space relationships between various signals [2]. Let's see, for example, how in section B (see Figure 5) of the piece the audio chain (implying several exchanges of audio signals between *clients*) affects the delay composition (delay values are given only as likely examples):

- $DEL(1-2) = 0.2''$ (client1 to client2 delay time);
- $DEL(2-1) = 0.1''$ (client2 to client1 delay time);

- $DEL(2-3) = 0.3''$ (client2 to client3 delay time);
- $DEL(3-2) = 0.5''$ (client3 to client2 delay time).

Player1 and *Player3* are reverberated; *Player3* is slightly delayed ($0.6''$) for the **REV** unit is in *client1*, and its signal must travel to *server*, in *client2*, and then to *client1*. The reverb output is routed to the **SPAT** unit in *client2* and thus delayed by $0.2''$; the **SPAT** unit routes the **REV** output to **L1**, **L2** and **P2** in *client2* real time, but to **P1** and **P3** the signal is further (and differently) delayed. *Player1* and *Player3* are also amplified on their local copper plate in their respective real time, but on **P2** with the delay time due to the communication with *client2*.

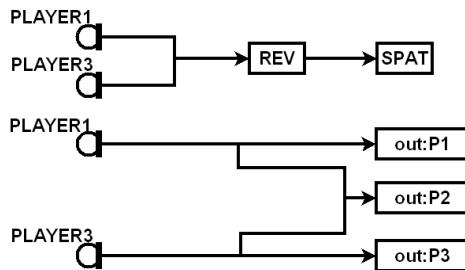


Figure 5: behaviour of sound processing space in section B of *rsm*.

According to this example it is clear that spatial dynamics must emerge from distinct *timbre colourations* of related acoustic signs (acoustical sounds and their possible transformations) and from combination of transmission delays (each communication leave its *traces*).

This conception of space composition draws on some statements by Pierre Boulez regarding the misconception between “movement and mean of *transport*” in space composition: “the space is instead a polyphonic distribution potential, an index for structures distribution”¹ [1].

The *rsm* function has been applied (up to now) to two works: *Rumore – Silenzio – Metallo*(2009), 2009, 30', for instruments, three laptops, three copper plates and video, intended for live performance, and *Rumore – Silenzio – Metallo*(2010), 2010, 20', for stereophonic audio and video, in which the whole score of the first version have been revised, adapted and recorded and the sound processing/diffusion *behaviours* wholly reconsidered for the different *medium*.

4. BUTCHERY

For this project we have *condensed* some of the main features of *rsm* to obtain a particular *function* suitable for what we call *table electronic music*, which represents an extremely adaptable performance modality.

¹ English translation by the authors.

A *corpse* is chosen to be *butchered* between any available and suitable sound producing object. The three performers will act on this *instrument* according to some previously established improvisation plan, one by one, by two or all together.

Performers will even manage a *sample and play* environment to record and repeatedly play *cuts* extracted by the acting performer(s). The audio loops thus produced are sent to the actuators placed on different resonating surfaces (see Figure 6).

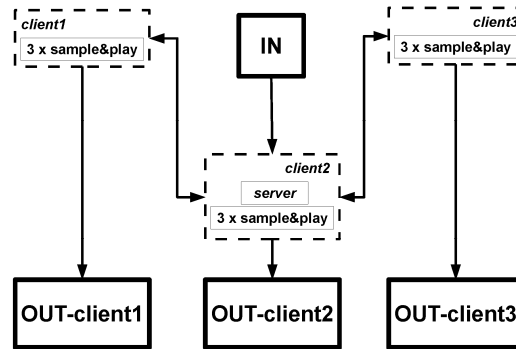


Figure 6: fields of definition for *butchery*.

The first piece realized with this *function* has been *Guitar butchery*, 2010, played at the *Biennale Musica* of Venice in 2010, in which the *corpse* was an electric guitar and the acoustic actuators were placed two on two metal surfaces and one on the wooden ground of the stage to provide a better response on low frequencies.

In *Guitar butchery* the body of the instrument is *sectioned* in three parts and each of these parts is acted on by a different performer. The performance score (see Figure 7) is in three parts and each is composed by two layers, one indicating the actions to be done on the guitar and one the actions to be done on the *laptop*.

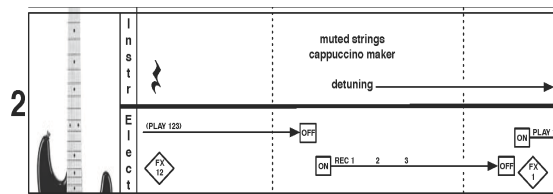


Figure 7: excerpt from the performance score of *Guitar Butchery*.

5. SOVRA-POSIZIONE

This *function* is born as an environment for sound installations: simplifying the description of the actual process, several inputs signals (instruments, ambience noises, etc.) are recorded at different times in distinct memory buffers. After sometime this buffers are

randomly read in chunks of variable length (from 20" to 0.1") and eventually further processed by some sound processing tool. Alongside with this process some pre-composed audio file may be read, according to the actual occasion. The *score* defines the actual length of audio chunks in the different *sections*, their density (simultaneity of chunks), conditions which trigger the reading of audio files and the dynamics for the sound diffusion.

Sovra-posizione is our most versatile *function* (see Figure 8); it was conceived initially to be used in concerts as an accompaniment for our improvisations or as an *intermezzo* between sections of a performance, and thus used several times.

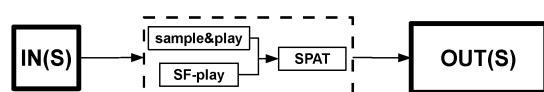


Figure 8: *sopra-posizione*.

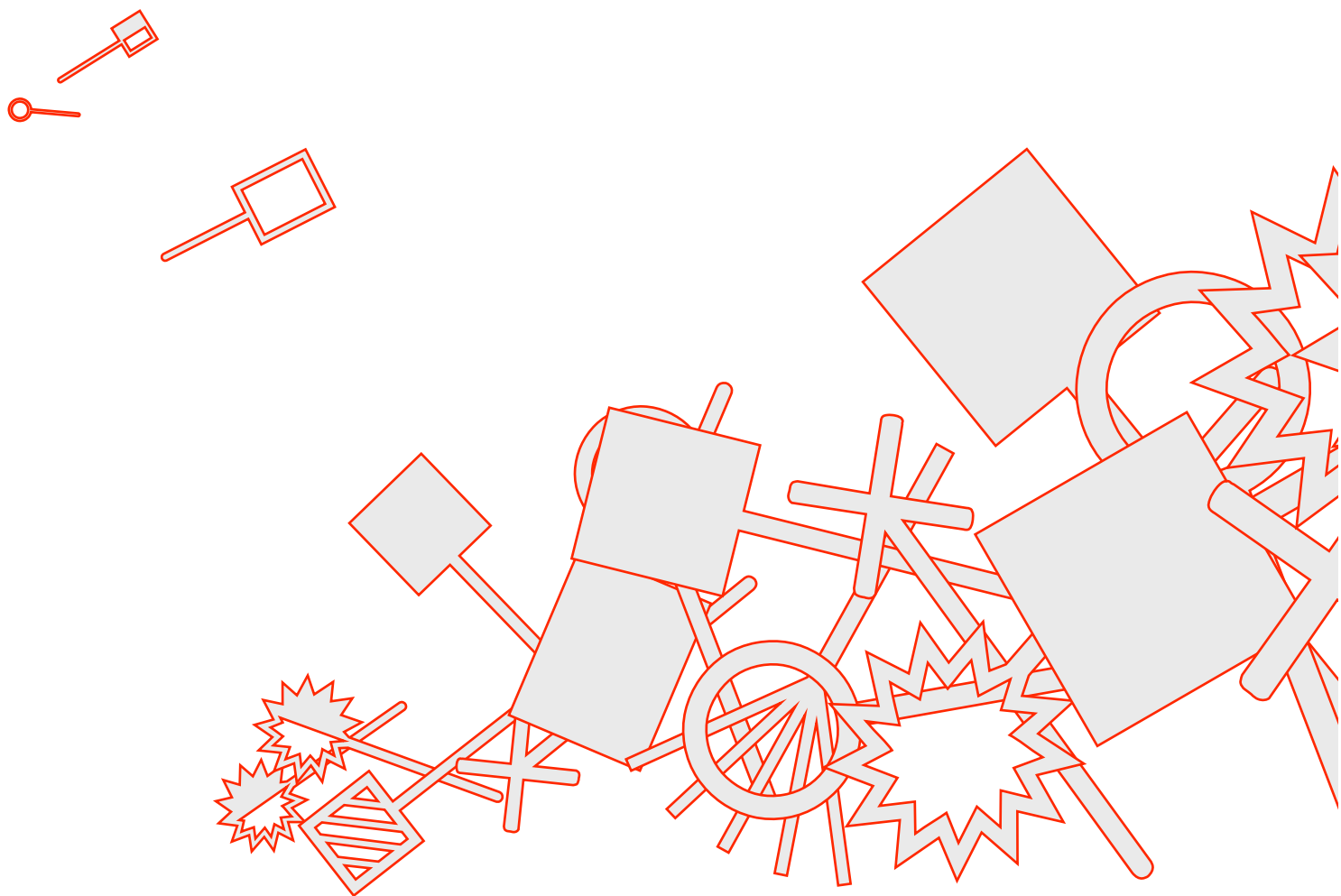
Since then it has even been used to compose electro-acoustical pieces such as *sopra-posizione(rsm)*, 2010, for multi-channels audio system and video¹, in which the *acoustical matter* and *sound processing* are taken from the production material for *Rumore – Silenzio – Metallo*(2010).

6. REFERENCES

- [1] Boulez, P. *Pensare la musica oggi*, Giulio Einaudi Editore, Torino, 1979.
- [2] Caceres, J. P., Renaud, A. B. "Playing the network: the use of time delays as musical devices", *Proceedings of the International Computer Music Conference*, Belfast, UK, 2008.
- [3] Chafe, C., Wilson, S., Leistikow, R., Chisholm, D., Scalone, G. "A simplified approach to high quality music and sound over IP", *Proceedings of the Conference on Digital Audio Effects (DAFx-00)*, Verona, Italy, 2000.
- [4] Di Scipio, A. "Formal Processes of Timbre Composition Challenging the Dualistic Paradigm of Computer Music", *Proceedings of the International Computer Music Conference*, Aarhus, Denmark, 1994.
- [5] Di Scipio, A. "Real-Time Polyphonic Time-shifting of Sound with Interactive Systems", *Proceedings of the XI° Colloquio di Informatica Musicale (XI CIM)*, Bologna, Italy, 1995.
- [6] Di Scipio, A. "Tecnologia dell'esperienza musicale nel Novecento", *Rivista Italiana di Musicologia*, 35(1-2), 2000.
- [7] Roads, C. *The Computer Music Tutorial*, MIT Press, Cambridge, 1996.
- [8] G.A.V. SOWTER, "Soft Magnetic Materials for Audio Transformers: History, Production, and Applications", *Journal of the Audio Engineering Society*, 35(10), 1987, pp. 760-777.
- [9] Koenig, G. M. *Genesi e forma*, Semar, Roma, 1995.
- [10] Truax, B. "Real-Time Granular Synthesis with a Digital Signal Processor", *Computer Music Journal*, 12(2), 1988, pp. 14-26.
- [11] Vidolin, A. "Manuale di Vocoder", *Laboratorio Musica*, a.II, n.17.X, 1980, pp.9-11.
- [12] Wright, M. "Open Sound Control: an enabling technology for musical networking", *Organized Sound*, 10(3), 2005, pp. 193-200.

¹ Selected for the *XVIII Colloquio di Informatica Musicale*, Turin, 2010, video by Lino Budano.

5. Apprendimento e tecnologia



+ Serena Zanolla, Antonio Camurri, Sergio Canazza, Corrado Canepa, Antonio Rodà, Gualtiero Volpe, Gian Luca Foresti
+ Tiziano Bole
+ Nicola Montecchio, Nicola Orio

UN AMBIENTE DIDATTICO AUMENTATO TECNOLOGICAMENTE

Antonio Camurri, Corrado Canepa, Gualtiero Volpe

Università di Genova

{antonio.camurri,gualtiero.volpe}@unige.it
corrado@infomus.dist.unige.it

Sergio Canazza

Università di Padova

canazza@dei.unipd.it

Gian Luca Foresti, Antonio Rodà, Serena Zanolli

Università di Udine

{gianluca.foresti,antonio.roda,serena.zanolli}@uniud.it

SOMMARIO

La *Stanza Logo-Motoria* è un ambiente interattivo multimodale per l'apprendimento e la comunicazione sviluppato mediante la piattaforma EyesWeb XMI (www.eyesweb.org). È installato in modo permanente in una Scuola Primaria dove viene utilizzato come strumento alternativo e/o addizionale alle tradizionali tecniche di insegnamento. La *Stanza Logo-Motoria* viene usata da tutti gli studenti della scuola compresi i bambini in situazione di handicap. Questo articolo intende descrivere il sistema *Stanza Logo-Motoria* e presentare i primi risultati delle attività di insegnamento che fino a ora vi si sono attuate.

1. INTRODUZIONE

Il Sistema Educativo Europeo è costituito da una popolazione scolastica estremamente eterogenea: coesistono vari livelli di apprendimento, una percentuale molto alta di studenti stranieri e un crescente numero di allievi in situazione di handicap [1]. Il percorso educativo degli studenti con disabilità è molto lungo e complesso e viene affrontato dagli istituti scolastici degli Stati dell'Unione Europea con modalità differenti. L'Agenzia Europea per lo Sviluppo dell'Educazione Speciale sostiene che le istituzioni scolastiche debbano in generale volgere all'implementazione di politiche educative finalizzate all'inserimento degli studenti con disabilità nelle scuole statali dotando gli insegnanti di diverse tipologie di supporto: personale scolastico aggiuntivo, nuovi materiali di insegnamento, corsi di formazione e nuovi strumenti tecnologici. Si ritiene quindi che la tecnologia possa essere di grande aiuto nel promuovere una maggiore adattabilità dell'offerta formativa e nel incrementare il livello della domanda culturale del sistema educativo [2].

Nel 1983 H. Gardner inizia a studiare le principali manifestazioni dell'intelligenza umana sviluppando la Teoria delle Intelligenze Multiple [3]. Con questa teoria Gardner sostiene che la nozione tradizionale di intelligenza, ba-

sata sulla misurazione del Q.I., sia troppo limitata. Egli ipotizza infatti l'esistenza di otto differenti tipi di intelligenza spiegando in questo modo l'ampio range di potenzialità umane presenti sia nei bambini sia negli adulti. Le tipologie di intelligenza che Gardner individua sono le seguenti: linguistica, logico-matematica, spaziale, cinestesico-corporea, musicale, interpersonale, intrapersonale, naturalistica. Gardner afferma che le nostre scuole e la nostra cultura si focalizzano quasi esclusivamente sulle intelligenze linguistiche e logico-matematica privilegiando così solo i soggetti che possiedono queste inclinazioni. Egli sostiene invece che è necessario porre la stessa attenzione nei confronti delle persone che dimostrano maggiore propensione verso le altre tipologie di intelligenza: gli artisti, i musicisti, i designer, i ballerini e le persone con disabilità. Sfortunatamente nell'ambito scolastico gli allievi che dispongono di queste attitudini spesso non ricevono un adeguato supporto. La Teoria delle Intelligenze Multiple di Gardner suggerisce di inserire nel curriculum scolastico ambiti o campi dell'attività umana (quali, ad esempio, la musica o la psicomotricità) che invece di solito vengono marginalizzati o non considerati affatto. È infatti estremamente importante conoscere e nutrire le menti umane in tutte le loro combinazioni al fine di incoraggiare l'interazione con il mondo, la crescita globale della persona e il raggiungimento del massimo livello di apprendimento possibile [4].

In linea con tale approccio, questo articolo presenta un sistema tecnologicamente aumentato, la *Stanza Logo-Motoria*, in grado di offrire uno strumento alternativo e/o addizionale alle tradizionali tecniche di insegnamento che spesso non si adattano alle individuali modalità di apprendimento. Il sistema analizza in tempo reale i movimenti globali del corpo umano e i suoi gesti all'interno di un ambiente sensorizzato, li elabora e li mette in relazione con un determinato contenuto sonoro e/o visivo. Particolare attenzione viene posta alle caratteristiche espressive dei gesti in grado di convogliare importanti significati emotivi e affettivi. La *Stanza* può essere utilizzata dagli insegnanti per trasmettere un contenuto mediante una modalità alternativa o per verificare il livello di apprendimento delle conoscenze nei bambini che esprimono meglio le loro capacità utilizzando le intelligenze visiva, spaziale e corporea.

Le basi teoriche e lo stato dell'arte degli ambienti interattivi vengono proposti nella Sez. 2 mentre l'architettura del

sistema viene descritta nella Sez. 3. Nella Sez.4 vengono presentate dettagliatamente le componenti della fase *feature extraction*. La Sez.5 spiega come vengono utilizzate le caratteristiche del movimento estratte nella fase *feature extraction* nell'ambito di una specifica applicazione della Stanza: il *Resonant Memory*. La Sez.6 illustra nel dettaglio le attività realizzate e i primi risultati derivanti dall'utilizzo del sistema interattivo. Seguono, in Sez.7, le conclusioni e le prospettive future di sviluppo.

2. FONDAMENTI TEORICI E STATO DELL'ARTE

Il sistema della *Stanza Logo-Motoria* affonda le radici neurofisiologiche nella scoperta del sistema dei neuroni specchio e si basa sulle teorie dell'*Enaction* e dell'*Embodied Cognition*. I neuroni specchio costituiscono il substrato neurale del riconoscimento e della comprensione delle azioni eseguite dagli altri individui. Rizzolatti e Voza [5] danno molta importanza all'aspetto motorio dell'attività cognitiva, sostenendo che l'apprendimento dipende dall'agire e che alla base della conoscenza sta il "compiere delle azioni". Esistono due tipologie di conoscenza: quella scientifica e oggettiva e quella esperienziale; quest'ultima è la modalità di conoscenza primaria dell'essere umano basata sul sistema motorio e sulle esperienze; la partecipazione del corpo è la condizione necessaria per lo sviluppo dei processi cognitivi. Questo approccio [6] si fonda sull'associazione multi-sensoriale di percezione e azione, di imitazione motoria e manifestazioni emotive e affettive. Ogni attività cognitiva è "contestualizzata", è associata in modo inestricabile a "ciò che stiamo facendo fisicamente" e alla struttura e alla dinamica dell'ambiente [7]. "Imparare facendo" è una dimensione importante anche per le teorie della cognizione *enattiva* che si basa su attività motorie - come la manipolazione di oggetti - dove le rappresentazioni *enattive* vengono acquisite mediante "il fare" [8].

Nella *Stanza Logo-Motoria* è possibile riscoprire l'importante aspetto motorio della conoscenza e utilizzarlo per situazioni di difficoltà di apprendimento. La *Stanza Logo-Motoria* è un ambiente in cui l'utente è motivato a "compiere delle azioni" per ricevere un contenuto: deve entrare all'interno di uno spazio, scegliere una zona, ascoltare attentamente, eseguire delle attività; l'utente ha un motivo per imparare, diversi modi di agire e di percepire all'interno di un ambiente sensibile. La conoscenza non viene imposta dall'alto ma viene offerta. Lo studente impara in movimento cercando un contenuto, muovendosi fisicamente nello spazio: idee, pensieri, concetti e categorie vengono modellati dalle posizioni e dai movimenti del corpo [6]. La *Stanza Logo-Motoria* diventa un "agente istruzionale" [2] perché produce un trasferimento di informazione e conoscenza come quando un insegnante tiene una lezione ai suoi studenti.

Alla fine degli anni Sessanta Myron Krueger inizia la sperimentazione sulle immagini elettroniche interattive; Krueger intende creare un ambiente artificiale attraverso la manipolazione dell'informazione audio-visiva. Egli presenta [9] il concetto di ambiente connesso allo spazio fisico in cui l'osservatore ha la possibilità di intervenire. Krueger utilizza spazi modificati in tempo reale dalla presen-

za dell'utente che non deve indossare particolari dispositivi. Nei primi esperimenti, usando un tappeto sensorizzato, Krueger estrae la posizione dell'utente all'interno dell'ambiente (*Glowflow*, *Metaplay*, *Psycic Space*) e poi elabora i gesti dell'utente mediante la tecnica del *video tracking*.

La *Stanza Logo-Motoria*, come lo spazio di Krueger, è un ambiente in cui la posizione e i gesti dell'utente vengono rilevati da una video-camera. Le coordinate della posizione del corpo nello spazio e le misure del gesto di apertura delle braccia sono sincronizzate con uno specifico feedback audio e/o video. Krueger [10] dà molta importanza al ruolo del corpo e pensa che gli "ambienti reattivi" siano strumenti finalizzati alla riappropriazione delle facoltà sensoriali sacrificate al potere della rappresentazione audiovisiva. Allo stesso modo la *Stanza Logo-Motoria* è uno strumento per recuperare le abilità attentive e di concentrazione; questo ambiente interattivo offre ai bambini la possibilità di re-imparare a sentirsi bene con se stessi durante l'ascolto attivo.

Un altro importante esempio di utilizzo dell'ambiente interattivo multimediale da parte dei bambini (in particolare quelli in situazione di handicap) è SOUND=SPACE. Nel 1984 Rolf Gehlhaar sviluppa SOUND=SPACE [11], un ambiente interattivo musicale multiuso, in cui gli utenti attivano e influenzano la produzione di suoni muovendosi all'interno di uno spazio vuoto rilevato da un sistema di ecolocalizzazione a ultrasuoni. Questo sistema ha suscitato un interesse a livello mondiale in particolare per la fruizione da parte di gruppi di persone con disabilità che ancora ne fanno esperienza durante workshop dedicati. In seguito, Gehlhaar e colleghi [12] hanno elaborato una nuova installazione multimediale, CaDaReMi, che risponde al problema della difficoltà da parte dell'utente di "ancorare" spazialmente le sue azioni; CaDaReMi fornisce infatti una serie di indizi visivi del suono che aiutano l'utente a comprendere più velocemente come funziona il sistema.

La *Stanza Logo-Motoria*, come CaDaReMi, è un ambiente: stimolante in quanto permette di esplorare un ampio range di suoni; collaborativo perché può essere utilizzato da più persone nello stesso momento; impegnativo e interessante in quanto offre sia suoni familiari sia suoni insoliti; visivamente attraente perché incoraggia l'utente a fare esperienze, a migliorare la sua abilità di collocarsi nello spazio e, allo stesso tempo, a localizzare spazialmente gli eventi sonori; socializzante in quanto favorisce l'interazione fra utenti e il consolidamento del senso comunitario dell'attività e del luogo; intuitivo perché il suo utilizzo non richiede complesse spiegazioni e per avere immediati risultati non è necessario possedere particolari prerequisiti; apprendibile perché sufficientemente complesso da fornire un'esperienza divertente ma anche sufficientemente facile in modo che i principianti abbiano la possibilità di sperimentarlo subito con successo.

3. L'ARCHITETTURA DEL SISTEMA

Diversamente dagli ambienti sopra descritti la *Stanza Logo-Motoria* è uno spazio interattivo installato in modo permanente in una scuola e permette il trasferimento delle conoscenze mediante il movimento. È un contenitore di espe-

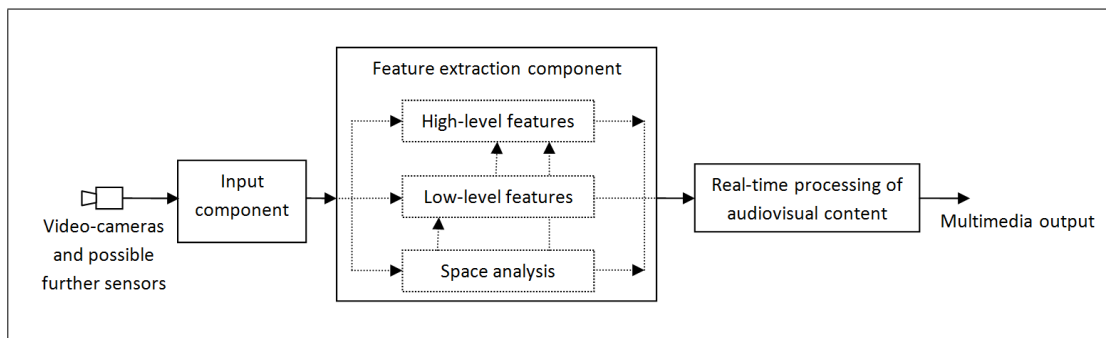


Figura 1. L'architettura del sistema della *Stanza Logo-Motoria*.

rienze da riempire con qualsiasi argomento: storia, matematica, geografia, scienze, etc. Il sistema crea un ambiente interattivo per la comunicazione, un luogo in cui l'utente, spostandosi nello spazio e con un semplice movimento delle braccia, genera informazione. Lo schema in fig. 1 illustra le tre componenti principali dell'architettura del sistema della *Stanza Logo-Motoria*:

- La componente *input* che riceve da una video-camera il segnale video; questa componente è anche responsabile del processo di pre-elaborazione dei dati (e.g. la riduzione del rumore e la sottrazione dello sfondo per estrarre la silhouette dell'utente).
- La componente *feature extraction* che analizza i dati in input al fine di ottenere informazioni su (i) la modalità di occupazione dello spazio da parte dell'utente (e.g. dove si trova, quanto tempo rimane in una determinata zona) e (ii) l'espressività del gesto. Nella Sez. 4 verranno forniti ulteriori dettagli riguardanti questa specifica componente.
- La componente per l'*elaborazione in tempo reale del contenuto audio-visivo* che dipende dalle caratteristiche estratte dalla componente *feature extraction*.

Le componenti dell'architettura del sistema verranno descritte dettagliatamente anche nella Sez. 5 che approfondisce una particolare applicazione della *Stanza Logo-Motoria*: la modalità *Resonant Memory*.

4. ESTRAZIONE DELLE CARATTERISTICHE DEL MOVIMENTO

Nella *Stanza Logo-Motoria* i movimenti del corpo e i gesti dell'utente vengono rilevati da una video-camera. Il segnale video viene elaborato per ricavare descrittori qualitativi (*features*) del movimento. Le caratteristiche del movimento estratte rappresentano sia la modalità di occupazione dello spazio da parte dell'utente sia le qualità espressive dei suoi gesti (e.g., se un gesto è fluido o impulsivo, preciso o incerto, etc.). Le analisi dell'occupazione dello spazio e dell'espressività del gesto vengono inizialmente pre-elaborate per ottenere la silhouette dell'utente. Le analisi del movimento dell'utente nello spazio si basano sul computo della traiettoria da lui seguita all'interno della zona

attiva (viene calcolata la traiettoria del 'centro della massa' dell'utente).

È possibile definire delle regioni all'interno della zona attiva e identificare quelle che vengono occupate più a lungo dall'utente mediante il calcolo del tasso di occupazione (informazione che può essere usata, per esempio, per attivare un contenuto audio-visivo); questa misurazione permette di associare a un alto tasso di occupazione di una determinata regione un alto interesse dell'utente per il contenuto associato a essa.

Nell'individuazione delle caratteristiche del movimento abbiamo seguito un approccio a più livelli: dai descrittori delle misure di basso livello (posizione, velocità, accelerazione delle parti del corpo) si passa ai descrittori delle caratteristiche complessive del movimento (fluidità, direzionalità, impulsività dei movimenti) [13]. Le analisi dell'espressività del gesto infatti permettono di estrarre diversi descrittori qualitativi che possono essere disposti su vari livelli di complessità: le caratteristiche gestuali molto semplici corrispondono al livello più basso (*low-level features*), altre invece, più articolate, vengono fatte corrispondere al livello più alto (*high-level features*). Le *low-level features* includono:

- le caratteristiche cinematiche di base (la posizione, la velocità e l'accelerazione del baricentro della silhouette);
- le caratteristiche estratte dalla silhouette - calcolate direttamente sulla silhouette dell'utente - che includono (i) il *Motion Index*, la quantità di movimento rilevata dalla telecamera, e (ii) il *Contraction Index*, un indice che misura la contrazione/espansione del corpo calcolando il rapporto tra l'area della silhouette e l'area del rettangolo che la delimita, il *Bounding Rectangle*;
- l'orientamento del corpo: calcolato considerando l'orientamento dell'asse maggiore di un'ellisse che approssima il contorno della silhouette.

Il concetto di *low-level feature* nasce dagli studi sulla percezione visiva e viene utilizzato per costruire modelli computazionali basati sulla percezione [14]. Le *high-level features* vengono calcolate a partire dalle *low-level features* sulla base di teorie, modelli ed esperimenti effettuati nel

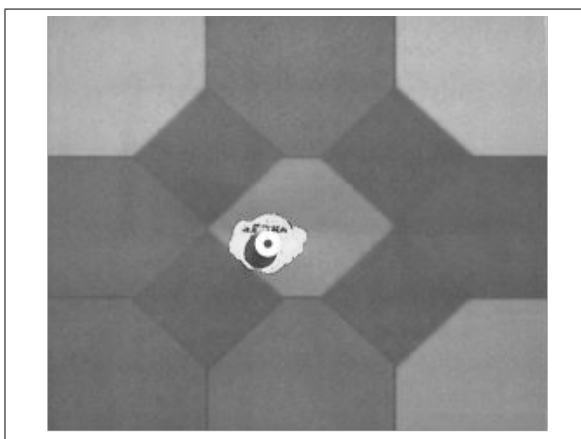


Figura 2. La riproduzione sonora della storia inizia quando il bambino entra nella zona centrale.

campo della psicologia, della biomeccanica e degli studi umanistici. I descrittori di alto livello, oltre ad avere un aggancio con una consolidata tradizione biomeccanica, derivano dallo studio del movimento umano effettuato da ricercatori con una formazione legata alle “performing arts” come il coreografo e kinesiologo Rudolf Laban che, con la Teoria dello Sforzo (inteso come portata dinamica del movimento), classifica il gesto a seconda del modo in cui un essere umano investe la propria energia cinetica; allora il gesto può essere forte o leggero, subitaneo o sostenuto, diretto o flessibile, controllato o libero [15]. Alcune fonti di rilievo provengono anche dalla ricerca nel campo della psicologia come gli studi di Wallbott [16], De Meijer [17] e Boone e Cunningham [18]. Le high-level features includono la misura della direttività del gesto, *Directness Index* e la misura della fluidità, *Fluidity*, che definisce se un gesto è sicuro o incerto.

L'applicazione Resonant Memory, descritta nella seguente sezione, rappresenta un esempio di come si può utilizzare la *Stanza Logo-Motoria*; in questo caso vengono usate solo low-level features estratte direttamente dal segnale video: (i) il Motion Index - il movimento globale del corpo umano rappresentato dalla misura della traiettoria del baricentro - e (ii) il Contraction Index - il gesto di apertura delle braccia rappresentato dalla misura della variazione delle dimensioni del Bounding Rectangle. In questo contesto sono state utilizzate solo low-level features - descrittori di movimento molto semplici - perché sono relativamente facili da calcolare e, allo stesso tempo, sufficienti per definire un set di gesti di controllo.

5. L'APPLICAZIONE RESONANT MEMORY

Il gioco Resonant Memory è una specifica applicazione della *Stanza Logo-Motoria*; lo spazio rilevato dalla videocamera viene suddiviso in nove zone: otto periferiche e una centrale. In questo caso le zone sono nove ma il numero può variare a seconda delle esigenze didattiche; a ogni zona corrisponde un'informazione sonora. La misura della traiettoria del centro della silhouette viene impiegata

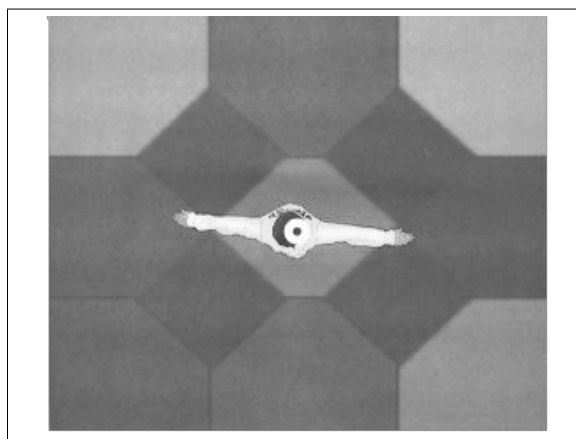


Figura 3. Il Contraction Index permette all'utente di fermare la riproduzione sonora della storia semplicemente allargando le braccia.

per sincronizzare un suono a una precisa zona dello spazio. Il bambino esplora lo spazio “risonante” (area attiva) muovendosi liberamente senza utilizzare alcun genere di sensori:

- alle zone periferiche vengono associati rumori, musiche o suoni ambientali che il sistema genera nel momento in cui il bambino li occupa fisicamente (fase di esplorazione);
- alla zona centrale viene associata la riproduzione sonora di una storia (Fig.2) che contiene gli elementi verbali/concettuali corrispondenti ai suoni collocati nelle zone periferiche (fase della storia).

Il bambino, ascoltando la storia, si diverte a ritrovare i suoni ascoltati nella fase di esplorazione e, allo stesso tempo, ne realizza la colonna sonora. Le analisi del segnale video e la generazione del feedback audio vengono eseguiti da una patch sviluppata in ambiente EyesWeb XMI. La Figura 4 presenta la patch intera che può essere suddivisa in tre stadi come spiegato di seguito. Nello stadio di *input* il segnale derivante dalla videocamera viene elaborato per estrarre alcune low-level features in relazione con i movimenti dell'utente. La sottrazione dello sfondo viene effettuata con approccio statistico mediante il *brightness/chromaticity distortion method* [19]. In questo modo vengono estratte la traiettoria del centro della silhouette, l'indice di movimento e l'indice di contrazione. Nello stadio di *mapping* (Fig.4 e 5) la patch analizza le caratteristiche del movimento estratte e, decodificando le azioni dell'utente, gestisce quattro fasi del gioco: l'esplorazione, la storia, la pausa e il reset. Infine, nel terzo stadio, quello di *output*, avviene il controllo della riproduzione del set di file audio pre-registrati. Dopo l'avvio dell'applicazione e nel momento in cui l'utente entra per la prima volta nell'area attiva, l'applicazione Resonant Memory avvia la fase di esplorazione. Ogniqualevolta l'utente raggiunge una delle otto zone periferiche il sistema memorizza tale spostamento. Se durante la fase di esplorazione l'utente raggiunge

la zona centrale prima di aver “toccato” tutte le otto zone periferiche non si attiva la riproduzione sonora; quando, completata la fase di esplorazione, l’utente entra nella zona centrale l’applicazione consente la riproduzione della storia (fase della storia). Se, durante l’ascolto, l’utente apre le braccia il sistema mette in pausa la riproduzione sonora fino a quando le abbassa (Fig.3). Se l’utente lascia l’area attiva e non vi rientra entro un tempo stabilito l’applicazione avvia la fase di reset che cancella la memoria delle zone visitate. Attualmente la rilevazione del movimento viene effettuata solo mediante video tracking e il sistema fornisce un feedback esclusivamente sonoro per le seguenti ragioni:

- il percorso di insegnamento/apprendimento mediante la *Stanza Logo-Motoria* è una assoluta novità sia per gli studenti sia per gli insegnanti quindi abbiamo pensato di proporre uno strumento molto semplice, diretto e intuitivo anche in considerazione delle limitate risorse finanziarie della scuola;
- in accordo con gli insegnanti abbiamo deciso che i bambini necessitavano prioritariamente di recuperare l’abilità di ascolto e di fare esperienza dello spazio senza utilizzare riferimenti visivi;
- il raggiungimento della multimodalità del sistema è l’obiettivo finale: la possibilità di interagire con il mezzo tecnologico in modi diversi, alternativi ed equivalenti verrà introdotta quando i bambini, nel rispetto dei loro tempi di crescita e dei loro bisogni, avranno ampiamente sperimentato la modalità sonora.

6. PROGETTI REALIZZATI

La *Stanza Logo-Motoria* è attualmente installata in un’aula della Scuola Primaria “E. Frinà” a Gorizia (Go) dove viene usata per favorire l’espressione delle intelligenze alternative e la comunicazione, promuovere la motivazione all’apprendimento e sviluppare i differenti stili cognitivi degli allievi. Il primo periodo di utilizzo del sistema è stato impiegato per far conoscere lo strumento agli insegnanti e agli allievi (quasi 170 bambini, dalla prima alla quinta classe). Vi si sono svolti numerosi progetti, descritti brevemente di seguito, di cui abbiamo raccolto documentazione video e molte osservazioni.

- Il Canguro della Continuità: per la classe prima e la sezione dei grandi di due Scuole dell’Infanzia del Circolo Didattico di Via Codelli; la classe prima ha svolto anche la sessione I Musicanti di Brema per favorire l’integrazione di un bambino con lieve autismo.
- Infolibro: in cui la *Stanza* è stata inserita come strumento per attivare la produzione di un testo di fantasia da parte dei bambini di seconda.
- I Suoni della Preistoria: sessione finalizzata allo studio della storia (classe terza).

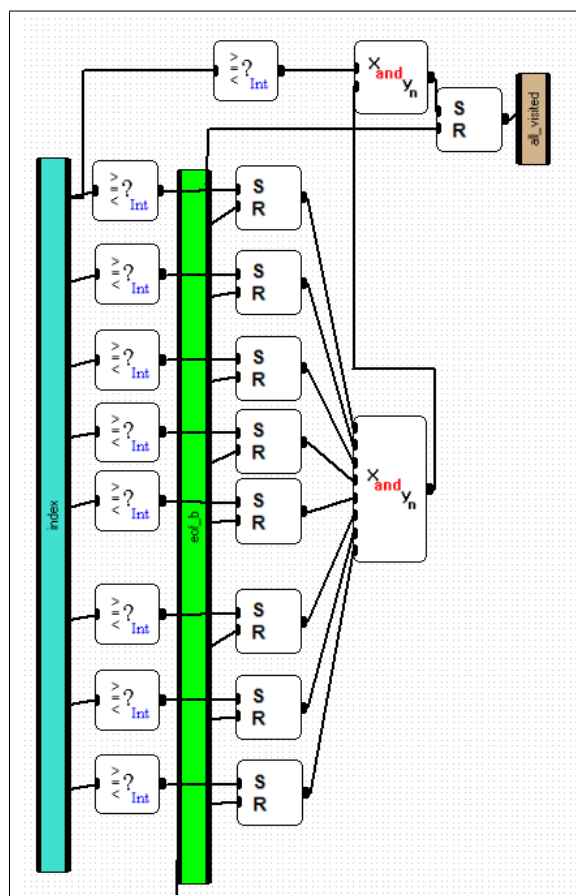


Figura 5. La subpatch che gestisce la transizione fra gli stati operativi dell’applicazione.

- Pierino e il Lupo di Sergej Prokof’ev: sessione di ascolto e analisi musicale (classe quarta).
- Nelle classi quinte la *Stanza* è stata utilizzata da due ragazzi con dislessia certificata come strumento compensativo/dispensativo nello studio delle scienze.
- Tutti gli allievi della scuola hanno potuto utilizzare la *Stanza Logo-Motoria* per l’apprendimento della lingua inglese.

Nella sezione 6.1 saranno descritti nel dettaglio due particolari percorsi didattici svolti mediante l’utilizzo della *Stanza*: l’apprendimento della lingua inglese e lo studio delle scienze da parte di studenti con dislessia.

6.1 L’apprendimento della Lingua Inglese

Ogni anno durante il secondo quadrimestre in questa scuola vengono organizzati, con l’intervento di un’insegnante madrelingua, corsi di lingua inglese che si articolano su due livelli di difficoltà. La *Stanza Logo-Motoria* viene utilizzata in questo contesto come strumento aggiuntivo e ha una principale finalità: abituare l’orecchio del bambino all’ascolto di fonemi molto diversi da quelli in uso



Figura 6. Imparare l'inglese nella *Stanza Logo-Motoria*.

nel proprio linguaggio. Utilizzare l'ambiente tecnologicamente aumentato per l'apprendimento della lingua straniera accresce la consapevolezza verso gli strumenti comunicativi non-linguistici: il corpo, le immagini, i suoni, e i simboli per il loro valore trans-culturale. La *Stanza Logo-Motoria*, usata come strumento veicolare, permette di strutturare le attività in modo da creare genuine connessioni fra le discipline. In questo modo si incoraggia sia il miglioramento delle abilità cognitive sia l'apprendimento integrato. Segmenti di linguaggio e intere frasi vengono utilizzate per 'fare' delle esperienze linguistiche che contribuiscono ad ampliare il lessico del bambino. Imparare l'inglese qui diventa spontaneo, divertente e naturale (Fig.6). Questo metodo di apprendimento coinvolge diversi linguaggi e metodologie:

- il linguaggio musicale: ascoltare la musica, discriminare i suoni e le caratteristiche timbriche degli strumenti musicali;
- il linguaggio teatrale: caratterizzare i personaggi, mimare azioni e situazioni;
- l'utilizzo delle lingue comunitarie per l'apprendimento delle discipline - la metodologia CLIL (*Content and Language Integrated Learning*);
- l'impiego delle nuove tecnologie.

I bambini esplorano le otto zone periferiche e memorizzano le coordinate spaziali dei suoni ascoltati (rumori, suoni ambientali o musica); poi, entrando nella zona centrale, attivano la riproduzione sonora di una storia in inglese. I bambini devono ascoltare con molta attenzione per comprendere velocemente quale suono introdurre all'interno del racconto e spostarsi nello spazio per attivare il suono suggerito dalla storia. Alle volte, in una o più zone periferiche, vengono inseriti i file audio di alcune sequenze della storia in cui sono stati omessi i nomi, gli aggettivi o i verbi che i bambini devono individuare e inserire con la loro voce. L'alta concentrazione nell'ascolto associata al movimento nello spazio assicurano un effettivo apprendimento. Anche dopo un certo lasso di tempo i bambini sono in grado di recuperare i contenuti appresi durante una

sessione di *Stanza Logo-Motoria*. Dalle analisi delle video riprese e dalle osservazioni sistematiche effettuate dagli insegnanti abbiamo constatato che l'ambiente tecnologicamente aumentato favorisce la motivazione all'ascolto e di conseguenza l'apprendimento di parole nuove e frasi. I test di valutazione somministrati in seguito dagli insegnanti hanno dimostrato che i bambini hanno assimilato correttamente tutti i contenuti proposti.

6.2 Supporto ai soggetti con dislessia

Ogni anno le istituzioni scolastiche riscontrano un certo numero di studenti con dislessia. In questi casi gli insegnanti devono utilizzare misure compensative e/o dispensative per facilitare l'apprendimento e adottare protocolli di valutazione specifici. La *Stanza Logo-Motoria* in modalità Resonant Memory è stata utilizzata anche da questi bambini come strumento alternativo per l'apprendimento delle materie di studio. L'attività si è svolta come segue:

- nella prima sessione di *Stanza Logo-Motoria* il testo da studiare (trasformato in file audio) è stato suddiviso in sequenze sonore; a ogni zona è stata sincronizzata una sequenza audio e associato un cartellone vuoto; gli studenti dovevano entrare in una zona, ascoltare attentamente il testo e attaccare sul cartellone le immagini collegate al testo scartando quelle non coerenti con i contenuti;
- nella seconda sessione gli studenti dovevano ascoltare il testo di ogni sequenza osservando attentamente le immagini precedentemente posizionate e ripetere con le proprie parole il contenuto riprodotto;
- nella terza fase gli insegnanti di classe hanno valutato oralmente l'apprendimento dei contenuti.

La scuola è frequentata anche da bambini ipovedenti e da bambini con problemi di comportamento che partecipano regolarmente alle attività organizzate nella *Stanza Logo-Motoria*. Questi bambini si integrano nel gruppo dei pari e raggiungono gli obiettivi previsti dato che l'ambiente tecnologicamente aumentato richiede abilità di apprendimento che possiedono: non ci sono testi da leggere, domande scritte a cui rispondere, operazioni matematiche da risolvere ma solo suoni da ascoltare e movimenti da eseguire. L'utilizzo della *Stanza Logo-Motoria* nel soggetto con disabilità può favorire:

- il potenziamento delle abilità comunicative già in atto;
- l'interazione con gli altri e l'ambiente;
- il potenziamento dell'intenzionalità e l'espressività del gesto;
- la diminuzione del movimento incontrollato (sincinesia) dovuto alla patologia;
- l'aumento dei tempi di attenzione e concentrazione;
- una maggiore motivazione allo spostamento autonomo nello spazio;
- l'acquisizione dei concetti spatio-temporali.

6.3 Risultati e prime valutazioni del sistema

Durante lo svolgimento delle attività nella *Stanza Logo-Motoria* è stato possibile immediatamente osservare nei bambini un particolare entusiasmo e una spiccata motivazione che si traducono in alta concentrazione e in un significativo aumento dei tempi di attenzione. L'ambiente interattivo, utilizzato dai bambini con dislessia per l'apprendimento delle materie di studio, ha contribuito a bypassare il codice linguistico scritto che in questi casi costituisce un ostacolo alla comprensione. Questi bambini inoltre hanno dimostrato di avere assimilato correttamente i contenuti proposti e hanno riacquisito la motivazione ad apprendere chiedendo spontaneamente di esporre alla classe ciò che avevano imparato. Allo stesso tempo hanno ottenuto punteggi migliori nei test di valutazione sugli argomenti trattati. Anche le insegnanti si sono dimostrate particolarmente motivate a utilizzare la *Stanza* proponendo spontaneamente argomenti e percorsi da realizzare. Alla fine dello scorso anno scolastico abbiamo rilevato che:

- rispetto il primo quadrimestre (da settembre a dicembre quando la *Stanza* non veniva utilizzata) i bambini dislessici hanno migliorato sia le prestazioni nei test di valutazione sia il loro approccio verso tutte le altre materie dimostrando maggiore sicurezza; le insegnanti hanno osservato che all'interno della *Stanza Logo-Motoria* questi bambini sono in grado di riassumere i contenuti con maggiore sicurezza facendo anche collegamenti con le altre discipline, un'abilità che non riescono a esplicitare lavorando solo in classe;
- per quanto riguarda l'insegnamento della lingua inglese l'insegnante ha rilevato che, grazie alla percezione del suono così totale e immersiva, i bambini hanno dimostrato un miglioramento della comprensione, della pronuncia e della produzione orale.

Questi risultati sono stati ricavati da:

- l'osservazione diretta;
- l'analisi delle videoregistrazioni;
- i test scolastici di valutazione.

Durante il corrente anno scolastico avremo la possibilità di ottenere maggiori informazioni sulle potenzialità didattiche ed educative della *Stanza*; intendiamo infatti attuare un protocollo di validazione mediante un gruppo di controllo e un gruppo sperimentale che utilizzino rispettivamente l'ambiente ordinario di apprendimento e la *Stanza Logo-Motoria*. I due gruppi saranno poi sottoposti alla stessa prova di valutazione che metterà in evidenza la qualità degli apprendimenti ottenuti nei due ambienti differenti. Sarà possibile anche verificare la presenza di cambiamenti significativi nelle medie dei voti ottenuti negli anni in cui la *Stanza* non era presente rispetto gli anni successivi in cui è stata utilizzata.

6.4 Conclusioni

Le attività realizzate con i 170 bambini della Scuola Primaria hanno dimostrato che la *Stanza Logo-Motoria* permette agli utenti di imparare a risolvere problemi di organizzazione attraverso esperienze gradualmente più complesse. I contenuti didattici appresi vengono assimilati correttamente e permangono nel tempo perché sono offerti attraverso un percorso di conoscenza corporea all'interno di uno spazio interattivo.

Per fornire un'esperienza maggiormente completa a livello sensoriale intendiamo ampliare il sistema implementando il feedback visivo e la spazializzazione del suono. Stiamo anche pensando di offrire ai bambini maggiori opportunità di interazione con i compagni sviluppando un'applicazione in cui la qualità del gesto di un utente dipenda da quella dell'altro. Questa caratteristica è essenziale quando l'utente ha difficoltà di relazione come quelle tipiche dell'autismo.

7. RIFERIMENTI BIBLIOGRAFICI

- [1] C. J. Meijer, V. Soriano, and A. Watkins, *Special Needs Education in Europe*. European Agency for Development in Special Needs Education, in collaboration with Eurydice, 2003.
- [2] G. Olimpo, "I nuovi ambienti interattivi per l'apprendimento," *ITD-CNR*, 1996.
- [3] H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*. Basic, 1983.
- [4] H. Gardner, *Educazione e sviluppo della mente. Intelligenze multiple e apprendimento*. Centro Studi Erickson, 2005.
- [5] G. Rizzolatti and L. Voza, *Nella mente degli altri. Neuroni specchio e comportamento sociale*. Zanichelli, 2008.
- [6] M. Leman, *Embodied Music Cognition and Mediation Technology*. The MIT Press, 2007.
- [7] F. Morganti and G. Riva, *Conoscenza, comunicazione e tecnologia. Aspetti cognitivi della Realtà Virtuale*. LED Edizioni Universitarie, 2006.
- [8] J. Bruner, *Toward a theory of instruction*. Belknap Press of Harvard University Press, 1966.
- [9] M. Krueger, *Responsive Environments. The New Media Reader*. The MIT Press, 1977.
- [10] M. Krueger, T. Gionfriddo, and K. Hinrichsen, "Video-place - an artificial reality. human factors in computing systems," *ACM press*, 1985.
- [11] R. Gehlhaar, "Sound=space: An interactive musical environment," *Contemporary Music Review*, vol. 6, no. 1, pp. 59-72, 1991.
- [12] R. Gehlhaar, L. M. Girao, and P. Rodriguez, "Cadaremi - an educational interactive music game," *ICDVRAT with Art Abilitation, Maia, Portugal*, 2008.

- [13] A. Camurri, G. De Poli, M. Leman, and G. Volpe, "Toward communicating expressiveness and affect in multimodal interactive systems for performing art and cultural applications," *IEEE Multimedia*, 2005.
- [14] A. Camurri and T. B. Moeslund, *Visual Gesture Recognition. From motion tracking to expressive gesture*. Appears as chp. 10 in the book *Musical Gestures. Sound, Movement, and Meaning*. Rolf Inge Godøy and Marc Leman (Eds.). Published by Routledge, ISBN: 9780415998871, 2010.
- [15] R. Laban and F. C. Lawrence, *Effort*. Macdonald & Evans Ltd., 1947.
- [16] H. Wallbott, "Bodily expression of emotion," *European Journal of Social Psychology*, 1998.
- [17] M. D. Meijer, "The contribution of general features of body movement to the attribution of emotions," *Journal of Nonverbal Behavior*, 1989.
- [18] R. T. Boone and J. G. Cunningham, "Children's decoding of emotion in expressive body movement: The development of cue attunement," *Developmental Psychology*, 1998.
- [19] T. Horprasert, D. Harwood, and L. Davis, *A Robust Background Subtraction and Shadow Detection*, vol. 1. In 4th ACCV, Taipei, Taiwan, 2000.

UN VIDEOGIOCO PER IL TRAINING RITMICO

Tiziano Bole

Conservatorio di Musica “G. Tartini” di Trieste
Scuola di Musica e Nuove Tecnologie
Biennio Superiore Sperimentale ad indirizzo videomusicale
tizi.bole@gmail.com

ABSTRACT

La figura dell’insegnante durante l’apprendimento musicale di uno studente è sicuramente importante, e in particolar modo si rende necessaria la presenza di un controllo esterno nell’ambito del *training* ritmico. Infatti, questo è uno dei campi in cui maggiormente appare centrale la ripetizione prolungata di tentativi seguiti da *feedback* circa la precisione raggiunta o meno. Ecco che lo strumento informatico, sia affiancandosi alla figura dell’insegnante in classe, sia supportando l’esercizio domestico autonomo, può facilmente prestarsi a questo genere di apprendimento. Infatti, se da un lato il computer non si sfinisce a far provare all’allievo un qualsiasi numero di volte lo stesso materiale, dall’altro l’allievo stesso, confrontandosi con la macchina, accetta più facilmente il necessario alto numero di prove, come avviene nei molti videogiochi in cui lo “sbagliare” e il ricominciare da capo sono un meccanismo consolidato e accettato. A tal proposito verrà qui presentata una *patch* in MAX-MSP mirata al training ritmico e rivolta ad un’utenza musicalmente non preparata, e saranno inoltre discussi i risultati di una prima sperimentazione effettuata in ambito scolastico.

1. INTRODUZIONE

Nel presente articolo si vuole illustrare un lavoro nato all’interno della Scuola di *Didattica della Musica* presso Conservatorio “G. Tartini” di Trieste¹. Si tratta di una *patch* in MAX-MSP rivolta al *training* di consolidamento ritmico per imitazione, con supporto visivo e procedimento per *prove ed errori* con difficoltà progressiva. Verranno anche discussi i risultati di una prima sperimentazione della *patch* stessa effettuata con ragazzi in ambito scolastico.

1.1 Importanza del mezzo informatico

Il *training* ritmico, in quanto esercizio di un’abilità, richiede un’accentuata componente ripetitiva ed un continuo

feedback circa la bontà della prestazione stessa. La ripetitività richiesta rende necessario lo studio anche autonomo, ma in assenza di un *feedback* sulla prestazione si assiste comunemente, a seguito dell’esercizio musicale errato in assenza del docente, all’assimilazione di errori che poi si rivelano gravosi da correggere. Vi è quindi la necessità di esercitarsi anche in autonomia, ma sapendo se quel che si sta eseguendo sia corretto o meno, e lo strumento informatico si può dimostrare adatto a tale scopo. E’ questo lo scenario in cui si colloca il presente lavoro: sebbene, come si vedrà avanti, in questa versione non sia ancora previsto un utilizzo del tutto autonomo del mezzo informatico, è preciso intento di sviluppare una futura versione adatta allo studio individuale in completa autonomia, per offrire attraverso il computer un ausilio didattico in una fase così critica dell’apprendimento.

Accanto a questa potenzialità, lo strumento informatico si rivela utilizzabile anche in presenza del docente, e questo è il caso elaborato attualmente. Infatti, prendendo ispirazione dai videogiochi, si può constatare come essi sappiano far ben accettare le ripetizioni, con il consolidato (e spesso usato) meccanismo del ricominciare un livello da capo a seguito di un errore. L’utente accetta questa modalità di gioco, e nei casi meglio riusciti non perde interesse, attenzione e motivazione nel ripetere a lungo, per esempio, uno stesso livello. Questo vantaggio motivazionale è proprio del misurarsi direttamente con la macchina, e l’insegnamento della musica, ed in generale ogni situazione educativa (citando Malone²), possono trarre sicuramente beneficio da un tale tipo di approccio, indipendentemente da un utilizzo domestico e autonomo ovvero scolastico in presenza del docente. Inoltre, come riferisce Calvani [1], lo stesso *feedback*, se fornito dalla macchina, essendo “impersonale” rende l’eventuale errore qualcosa che non si deve temere, il che diventa importante per il senso di sicurezza e di autostima dell’utente.

1.2 I videogiochi musicali

Il *training* ritmico, e in generale il *training* musicale, sono campi in cui lo strumento informatico è stato già ampiamente utilizzato, tuttavia con finalità spesso non educative ma di semplice divertimento, con decine di titoli di cosiddetti “videogiochi musicali”, dalla fine degli anni ’90 fino ad oggi, con esempi anche molto popolari quali i celebri *Guitar Hero*, *Rock Band* ed altri ancora, alcuni dei quali

¹ Conservatorio di Musica “G. Tartini” di Trieste, Scuola di Didattica della Musica, corso di Elementi di Pedagogia Musicale, docente prof.ssa Cristina Fedrigo.

Copyright: ©2010 Tiziano Bole et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

² Mason, L. e Varisco, B. M., *Mente umana, mente artificiale*, Angeli, Milano 1990. Cit. in Calvani [1].

disponibili addirittura per dispositivi portatili quali l'iPhone e l'iPod. Questi videogiochi costituiscono il più delle volte una sorta di evoluzione del *karaöke*, con una rappresentazione grafica scorrevole nel tempo di azioni da compiere (come premere dei tasti su di un apposito *controller*) in relazione a una musica di sottofondo. Per quanto riguarda specificamente i videogiochi ritmici, nati inizialmente come *dancing games* per le sale da gioco attrezzate e trasferitisi poi su versioni virtuali di strumenti musicali, essi ereditano da quel primo approccio un giocare basato per parte rilevante su un comportamento di risposta immediata, in cui è la variabilità e l'imprevedibilità dei *pattern* ritmici a costituire l'ambito della sfida, con l'assegnazione di un punteggio proporzionale alla correttezza dell'esecuzione. Vi sono almeno tre aspetti nella modalità di gioco di questi videogiochi che non li rende buoni strumenti di *training* ritmico. In primo luogo, è proprio la ripetizione di uno stesso *pattern* a coprire un ruolo centrale nel consolidamento della *performance* ritmica, elemento invece negato in favore della variabilità del gioco, il quale così diventa sicuramente più avvincente ma meno didattico. In secondo luogo, un gioco basato in modo così massiccio sul comportamento di risposta rapida distoglie inesorabilmente la concentrazione dell'utente dagli aspetti più musicali della *performance*. In terzo luogo infine, l'assegnazione di un punteggio riduce il gioco a un test, senza per contro apportare alcun beneficio alla prestazione. Un'alternativa più "didattica" deve invece rinunciare all'assegnazione di un punteggio, conservare possibilmente la struttura per livelli di difficoltà progressiva per la capacità motivazionale che ciò comporta, deve mantenere a lungo costante il *pattern* ritmico, variando piuttosto altri parametri "di contorno" come la velocità di esecuzione o altri, e deve rinunciare a tutti quegli aspetti che sovraccaricano sensorialmente e cognitivamente l'utente e che non siano strettamente necessari per il gioco stesso, come sicuramente il gioco basato sui comportamenti di risposta rapida, ma anche, come si vedrà nel nostro caso, la notazione ritmica tradizionale. Si può inoltre considerare rilevante la possibilità per l'utente di eseguire la sua prestazione non solo premendo tasti o percuotendo *pad*, come avviene solitamente nei videogiochi musicali, ma con una vera e propria esecuzione musicale attraverso strumenti in questo caso percussivi.

Esempi di "giochi" che corrispondano a questa descrizione sono sicuramente meno celebri di quelli citati precedentemente. Recentemente, ad esempio, è stata presentata "iPalmas" (di A. Jylhä ed altri [3]), una *patch* in PUREDATA rivolta al *training* dei *pattern* ritmici tipici del flamenco. In questo lavoro, il cui focus è rivolto allo studio delle modalità di iterazione tra insegnante e allievo piuttosto che al *training* ritmico in sé stesso, si possono individuare tutti gli aspetti di cui si è appena discusso: utilizzo di *pattern* ritmici costanti nel tempo, possibilità di scelta (sebbene manuale, dalla *user interface*) tra vari livelli di difficoltà, assenza di punteggio, esecuzione da parte dell'utente utilizzando una reale fonte sonora come il battito delle proprie mani, variabilità della velocità d'esecuzione (sebbene tale variabilità non sia prefissata ma segua in modo adattivo la *performance* dell'utente).

La *patch* che ora si presenterà porterà avanti per tutti i motivi sopra citati queste stesse caratteristiche, per un utilizzo dunque inteso in senso didattico e non puramente ludico.

2. FUNZIONAMENTO

2.1 Il ruolo dell'insegnante

La *patch* che qui viene presentata non costituisce uno strumento autoesplicativo, in grado cioè di far passare all'utente tutte le informazioni necessarie per il suo corretto utilizzo, ovvero come emettere il suono durante la *performance*, come interpretare la rappresentazione grafica del *pattern* che viene richiesto, e soprattutto cosa si richieda per la corretta esecuzione in generale di un *pattern* ritmico. E' richiesta così una spiegazione di questi aspetti, in particolare anche di quelli - semplici - musicali, per cui questa risorsa si configura come uno strumento di esercizio di consolidamento di abilità ritmiche già apprese. Per questo è prevista, accanto al mezzo informatico, la presenza di un *tutor* (tipicamente un insegnante di musica) non solo per dare all'utente le necessarie spiegazioni, ma anche per interagire poi con la *patch* nello stabilire volta per volta il numero delle ripetizioni del *pattern* da far eseguire all'utente prima di far valutare alla stessa *patch* la prestazione. La figura dell'insegnante durante il *training* non viene quindi eliminata ma diventa per così dire "nascosta", attraverso un'interazione - per la verità minima - non più con l'utente ma soltanto con il mezzo informatico. Egli sembra così osservare in disparte ma in verità stabilisce i tempi della *performance*, in un ruolo insostituibile ma non invasivo per l'utente.

2.2 Overview

In fig. 1 è illustrato il funzionamento generale della *patch*. Vi è innanzitutto una struttura per livelli e sottolivelli, di cui si parlerà a breve. Ad ogni livello la *patch* genera un *pattern* ritmico, lo rappresenta graficamente attraverso una scrittura intermittente basata su un certo numero di *led* che possono essere accesi o spenti, e a seconda del sottolivello può eseguire o meno *in loop* il *pattern* appena generato; l'utente deve quindi eseguire tale *pattern* ritmico su un metronomo stabilito, battendo le mani o con un qualche strumento percussivo; attraverso un microfono il sistema acquisisce la *performance* e valuta, nel momento deciso dall'insegnante, quanto questa sia corretta. Se essa viene valutata positivamente, vi è un progresso di livello o di sottolivello e si ricomincia da capo con l'invenzione di un nuovo *pattern* o con la variazione dei parametri di contorno, come si vedrà avanti. Se la prestazione viene invece valutata negativamente, la *patch* continua a rilevare la *performance* dello stesso *pattern* finché non vi è un successo.

Verranno ora illustrate nel dettaglio le varie fasi indicate in fig. 1.

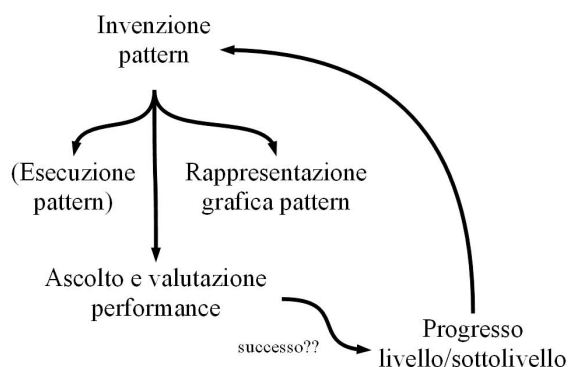


Figura 1. Schema di funzionamento generale della patch.

2.3 Livelli e sottolivelli

Come si è detto, vi è una struttura per livelli e sottolivelli, con la logica di una difficoltà progressivamente maggiore ad ogni livello, all'interno del quale il *pattern* presentato resta sempre lo stesso, e di una variazione dei parametri di contorno ad ogni sottolivello. La variazione di tali parametri è pensata per fissare al meglio nella mente dell'utente il *pattern* ritmico del livello in questione, presentandolo sotto forme diverse. Così, ad ogni livello viene presentato un unico *pattern*, ma a seconda del sottolivello variano sia la velocità (nell'ordine, media, lenta e veloce) sia la presenza o meno dell'esecuzione del *pattern* stesso da parte della *patch*; tale esecuzione, avvenendo *durante* la prestazione dell'utente, fornisce un aiuto acustico alla *performance* stessa. Dalla combinazione di queste possibilità, si hanno 6 sottolivelli per ogni livello, come illustrato in fig. 2.

velocità:	media		lenta		veloce	
aiuto:	si	no	si	no	si	no

6 sottolivelli

Figura 2. Schema illustrativo dei 6 sottolivelli di ogni livello, con la variazione della velocità e la presenza o meno dell'aiuto acustico, cioè l'esecuzione del *pattern* richiesto da parte della *patch*.

Superato il sesto sottolivello, si passa quindi al livello successivo, e conseguentemente viene "inventato" dalla *patch* un nuovo *pattern* ritmico. Con il progredire dei livelli, che in totale sono 20, l'errore accettato nella prestazione, come si vedrà più avanti, si riduce sempre più, e ciò costituisce la principale fonte di incremento di difficoltà perché si richiede via via una precisione maggiore all'utente. Inoltre, a partire dal settimo livello si "sbloccano" i valori ritmici più brevi, della durata di un *led* (equivalente, come si vedrà, ad una croma), cosa che dà fatalmente luogo a sincopi fino a prima assenti dai *pattern*.

2.4 Invenzione dei pattern ritmici

L'invenzione di un nuovo *pattern* ritmico, come si è detto, avviene all'inizio di ogni livello, durante il quale poi il *pattern* da eseguire rimane sempre lo stesso. L'invenzione da parte della *patch* avviene componendo casualmente dei valori ritmici corrispondenti a quelli di croma, semiminima, minima e minima puntata, per una durata complessiva equivalente ad una battuta in 4/4; vengono usati qui i termini "corrispondente" e "equivalente" perché il *pattern* così generato in verità non viene mai presentato in notazione tradizionale, bensì, come si vedrà più avanti, semplicemente attraverso una sequenza di *led* accesi o spenti. Inoltre, come si è già avuto modo di dire, i valori più brevi (per così dire le "crome") non vengono utilizzati prima del settimo livello, rappresentando la loro presenza fonte di un certo arricchimento ritmico, con la conseguente comparsa di sincopi e simili: in tal modo queste figure vengono evitate nei primi livelli per favorire un arricchimento progressivo delle abilità richieste.

2.5 Rappresentazione grafica dei pattern

Si è già detto che la rappresentazione grafica non avviene per notazione tradizionale bensì per notazione intermittente a *led*. Il motivo di questa scelta sta nel fatto che una tale rappresentazione, poiché strettamente essenziale, si rivela la più funzionale. Laddove la notazione tradizionale rappresenta infatti anche la durata dei suoni, nel caso percussivo che riguarda la presente applicazione una tale informazione appare inutile. Scegliendo così una scrittura intermittente, che quindi non rappresenta la durata dei suoni, si evita il sovraccarico di lavoro e elaborazione che sarebbero invece richiesti per leggere e interpretare scritture più complesse, come quella tradizionale, i cui simboli offrono come detto più informazioni ma in questo caso inutili; in ultima analisi si evitano anche le ambiguità che possono venire da una sovrainformazione, basti pensare alla pluralità di scritture per così dire "sinonimiche" del medesimo ritmo permesse in notazione tradizionale. Rifacendosi alla teoria di Salomon³, inoltre, la scrittura intermittente rappresenta, per un generico utente, un qualcosa di sicuramente più vicino alle esperienze pregresse - e quindi ad un sistema simbolico già acquisito - rispetto alla notazione musicale tradizionale, e comporta così una minore attività di elaborazione.

Nel dettaglio, il *pattern* ritmico viene rappresentato mediante 8 *led*, ciascuno stante per un *click* del metronomo, con la convenzione *led acceso* / suonare, *led spento* / non suonare. Questi *led* sono disposti in fila orizzontale che va letta da sinistra a destra, a diverse elevazioni in modo da rappresentare le corrispondenti altezze del suono del metronomo: se infatti l'intero *pattern* va pensato come una battuta in 4/4, il metronomo suggerisce una suddivisione per crome scandendo così con un suono più acuto i *click* che cadono in batture su una delle 4 pulsazioni, con uno più grave i rispettivi *click* in levare, e, poiché il primissimo *click* costituisce il punto di riferimento di inizio battuta, esso è reso unico attraverso un suono più grave di tutti

³ Salomon, G., *Interaction of media, cognition and learning*, Jossey-Bass Publisher, San Francisco 1979, cit. in Calvani [1].

gli altri. Così anche l'elevazione grafica dei *led* segue questo stesso andamento, con il primo posto più in basso e gli altri alternati più sopra, in modo da facilitare la lettura durante l'ascolto del metronomo. Inoltre, il *led* di volta in volta "attuale" viene evidenziato tramite un colore diverso del bordo, per facilitare ancora maggiormente la tenuta del segno in lettura. Per un esempio, si veda in fig. 3.

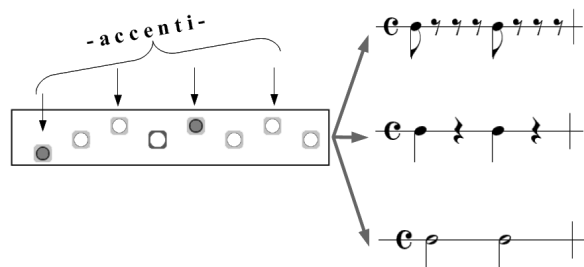


Figura 3. Sulla sinistra, esempio della rappresentazione tramite *led* di un semplice *pattern* ritmico: sono accesi soltanto il primo ed il quinto *led*, mentre il quarto è contornato in modo diverso in quanto è il *led* "attuale"; sulla destra, tre possibili rappresentazioni equivalenti attraverso scrittura tradizionale. Sono inoltre indicati con le frecce i *led* corrispondenti agli accenti del metronomo.

2.6 Esecuzione del pattern ritmico

L'esecuzione o meno da parte della *patch* del *pattern* ritmico dipende dal sottolivello, come si è già detto. Nel caso questa sia prevista, essa avviene tramite campioni di *clap* di mani, differenziati in accentuati (per gli eventi sonori in battere) e non (per gli altri), in modo da rievocare un'esecuzione umana realistica ed il più vicina possibile a quella richiesta all'utente.

2.7 Ascolto e valutazione della performance dell'utente

Come detto, la prestazione dell'utente va effettuata attraverso un qualsiasi strumento percussivo, come anche il battito delle mani. Il suono viene acquisito tramite un qualsiasi microfono e analizzato con un *attack detector* tarato per reagire ad incrementi maggiori di 6 dB in 10 ms. In questo modo si introduce un ritardo costante di esattamente 10 ms, che non incide nella successiva valutazione della correttezza della *performance*, in quanto gli intervalli di tempo tra gli eventi sonori acquisiti rimangono ovviamente invariati. Una volta che il tutor/insegnante attiva la funzione di valutazione durante l'esecuzione, alla conclusione di ogni *pattern* viene costruito il vettore degli intervalli temporali tra eventi sonori successivi dell'utente. Sia $(t_0, t_1, \dots, t_{n-1})$ questo vettore, e sia invece $(\tau_0, \tau_1, \dots, \tau_{n-1})$ il vettore degli intervalli temporali corretti (quelli generati dalla *patch* all'inizio), allora l'errore viene valutato attraverso la formula:

$$E = \frac{1}{n} \sum_{k=0}^{n-1} |t_k - \tau_k| \quad (1)$$

equivalente alla media delle *distanze L1* (la distanza "Manhattan") tra ogni intervallo temporale della versione corretta ed il rispettivo intervallo di quella eseguita. Come si dirà nelle conclusioni, la scelta della funzione distanza appropriata da utilizzare in questo luogo, in modo da approssimare il più possibile la sensibilità ritmica umana, è da considerarsi ancora oggetto di futuro studio.

Calcolato l'indice di errore E , esso viene confrontato con la soglia di accettazione, e se ne risulta minore vi è un successo, rappresentato acusticamente con il suono di un applauso, e l'utente conseguentemente avanzerà di livello o sottolivello. Altrimenti la valutazione di E prosegue, senza soluzione di continuità, per una nuova esecuzione del *pattern* ritmico, fino ad un successo. Il valore di questa soglia diminuisce con il progredire dei livelli, richiedendo così una sempre maggiore precisione all'utente per poter avanzare ulteriormente. Essa passa da un valore iniziale di 0.5 s ad uno finale (al livello 20) di 0.05 s, pari al valore della soglia cronoestesica relativa⁴. Costituendo quest'ultima la più piccola differenza di durata apprezzabile tra due suoni consecutivi, può rappresentare a buon titolo il limite di precisione realisticamente richiedibile all'utente. Si è potuto testare con la *patch* stessa che una buona esecuzione umana, utilizzando la *distanza L1*, può tranquillamente superare una tale richiesta, cioè superare l'ultimo livello. Per il valore della soglia nei livelli intermedi, è stata utilizzata un'interpolazione esponenziale tra il valore iniziale e quello finale. Questo perché, per la legge di Weber⁵, volendo mantenere costante l'incremento percepito di precisione richiesta da un livello al successivo, è necessario che il rapporto tra le soglie dei due livelli sia costante, e quindi che i valori stessi della soglia si dispongano su una curva esponenziale. In simboli, sia x il numero di livello, $S(x)$ la soglia al livello x , e $P(S)$ la precisione richiesta dalla soglia S , si avrà per la legge di Weber:

$$\begin{aligned} P(S(x+1)) - P(S(x)) &= \text{cost. sse} \\ S(x+1)/S(x) &= \text{cost.} \end{aligned} \quad (2)$$

La curva così ottenuta per la soglia S al variare del livello x segue quindi l'equazione:

$$S(x) = 564.4e^{-0.1212x} \text{ ms} \quad (3)$$

ed il grafico è quello riportato in fig. 4.

3. SPERIMENTAZIONE

Questa *patch* è stata sperimentata anche in sede didattica⁶ con progetto dedicato⁷. Una delle lezioni, della durata di

⁴ Aa.vv., *Dizionario di psicologia*, Edizioni Paoline, Torino 1986 (*Le-xikon der Psychologie*, Verlag Herder KG, Freiburg in Breisgau 1980), ad vocem "Soglia cronoestesica", pag. 1094, a cura di W. Traxel, cit. in Fedrigo e Calabretto [2].

⁵ *ibidem*, ad vocem "Weber/legge di", pag. 1267, a cura di G. D. Wilson.

⁶ Progetto "Scuole Aperte", laboratorio musicale-matematico "Contare la Musica e Cantare la Matematica", realizzato da settembre 2009 a febbraio 2010, a cura della Scuola Media Statale "G. Corsi" di Trieste (coordinatrice prof.ssa Rossana Divo) e della Scuola di Didattica della Musica del Conservatorio di Musica "G. Tartini" di Trieste (coordinatrice prof.ssa Cristina Fedrigo).

⁷ Realizzato con la prof.ssa Cristina Fedrigo presso il Conservatorio "G. Tartini" di Trieste.

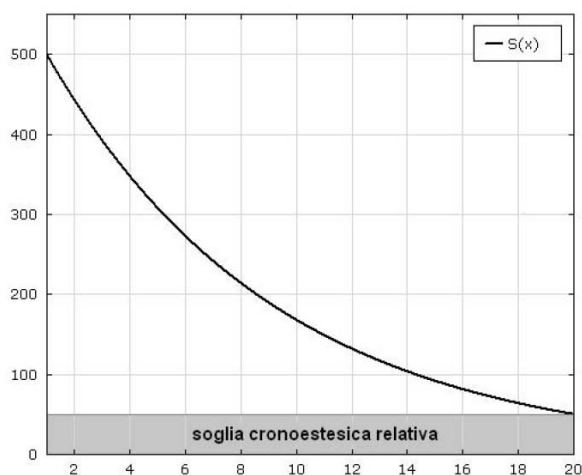


Figura 4. Grafico del valore della soglia S di accettazione dell'indice di errore sulla performance dell'utente al variare del livello x .

un paio d'ore, è stata completamente dedicata all'utilizzo della *patch*, in una classe di 20 ragazzi, tutti aventi una preparazione musicale di base ricevuta all'interno della scuola, alcuni inoltre instradati allo studio di uno strumento al di fuori della scuola stessa. In veste di insegnante, il mio intervento si è limitato a illustrare il funzionamento della *patch* e a chiarire le modalità del "gioco". E' stata quindi utilizzata dapprima la versione "sonora", che è la modalità con cui la *patch* è stata in questa sede presentata; tale modalità non permette la fruizione multipla contemporanea nello stesso ambiente per motivi di interferenza acustica, così i ragazzi in questa fase sono stati invitati a interagire a turno per mezzo del battito delle proprie mani; è stata poi utilizzata la versione "muta", in cui l'interazione sonora dell'utente con il computer è stata sostituita con la pressione di un tasto, e l'ascolto del metronomo è stato allestito in cuffia. In questo secondo modo si è potuto far "giocare" i ragazzi contemporaneamente, ognuno su un proprio computer.

Il risultato di questa sperimentazione è stato incoraggiante, innanzitutto poiché il livello di concentrazione si è mantenuto molto alto per più di 30 minuti, cioè l'intera durata della sessione "muta"; inoltre si è potuto notare che, facendo lavorare in parallelo i ragazzi, chi aveva già una certa esperienza musicale è riuscito a raggiungere livelli superiori rispetto agli altri, fatto che testimonia la consistenza del lavoro indotto dalla *patch* rispetto alle esperienze pregresse e la funzionalità del metodo operativo utilizzato; tutti i ragazzi sono comunque riusciti a superare i primi livelli, aspetto questo che evidenzia la sostenibilità del compito richiesto anche per chi aveva un livello di competenze musicali inferiore. L'incremento delle abilità nella prestazione ritmica è stato ulteriormente verificato anche con successivi esercizi ritmici e poliritmici collettivi. La valutazione dei ragazzi infine è stata generalmente da favorevole a molto favorevole.

4. CONCLUSIONI

La *patch* si è rivelata un buon strumento didattico e sufficientemente accattivante anche nell'attuale versione, per la verità un po' grezza e "fredda" nell'interfaccia utente. Tra i miglioramenti ancora da apportare c'è senz'altro tutto il lavoro sulla presentabilità grafica, con un'ipotesi di sviluppo per la risorsa in ambito scolastico. Questo genere di intervento cercherà comunque di non appesantire il carico cognitivo richiesto, mantenendo quindi un certo "minimalismo" a ogni livello operativo, e ogni elemento andrà presentato nel modo più *child-friendly* possibile, sia a livello grafico ma anche a quello acustico. Inoltre, una possibilità da aggiungere è quella di poter variare la durata dei pattern ritmici, magari introducendo "battute-equivalenti" diverse dai classici 4/4. Come già accennato nell'introduzione, poi, la presente risorsa si presta a essere sviluppata anche in un'ottica di fruizione domestica, in autonomia dall'insegnante, come supporto allo studio individuale. Infine, per quanto riguarda la ricerca, sarà da valutare attentamente la funzione di distanza da utilizzare nel calcolo dell'indice di errore, la più prossima possibile a quello che darebbe un musicista "in carne ed ossa".

5. BIBLIOGRAFIA

- [1] A. Calvani. *Dal libro stampato al libro multimediale*. La Nuova Italia, Firenze, 1990.
- [2] C. Fedrigo e R. Calabretto. *La porta incantata*. Libreria al Segno, Pordenone, 1998.
- [3] A. Jylhä, C. Erkut, M. Pesonen, e I. Ekman. Simulation of rhythmic learning: a case study. In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, September 2010.

Un aiuto preziosissimo è stato fornito inoltre dalla prof.ssa Cristina Fedrigo, la cui persona va citata in questa bibliografia come fonte rilevante di sapere e di conoscenze nel settore della pedagogia, il cui significativo apporto in taluni casi non mi è stato possibile ricondurre ad adeguate fonti scritte.

Visual and Aural Tools for Music Education Based on Audio to Score Alignment

Nicola Montecchio, Nicola Orio
Department of Information Engineering
University of Padova

ABSTRACT

We investigate a set of novel applications that are based on audio-to-score alignment technologies. While such technologies have been historically exploited for the most part in the context of automatic accompaniment, thus relegating their usage principally to the performance of electronic music by professional musicians, here the focus is on users who do not possess a rigorous music training background and would consequently benefit from an improved listening experience by exploiting a synchronization to the information provided by a score, even if they are unable to read music notation.

1. INTRODUCTION

An audio-to-score alignment system is a software tool that is capable of aligning a digital audio signal recorded from a music performance with a symbolic representation of its score: given a recording/score pair, an alignment is a match of each region of the audio source with the musical event that ideally originated it (Figure 1).

Most existing systems are based on statistical models. In particular hidden Markov models (HMMs) are the most common approach and their effectiveness has been investigated in [1]; another frequently used technique is Dynamic Time Warping [2], a mathematically equivalent approach which is not endowed with a statistical formalism. Alternative approaches make use of more sophisticated techniques, such as hybrid Markov/semi-Markov chains [3], conditional random fields [4], sequential Montecarlo inference [5]. Even though each approach has its own peculiarities, the discussion of this paper transcends the particular system model and focuses instead on the usage of the alignment data.

The typical application of such alignment systems is an “automatic accompanist”, a software allowing solo players to practice their part while the computer synchronously plays a predefined accompaniment; in the case of contemporary music, the aim is usually to trigger the playback of pre-recorded samples or to process and transform the sound produced by the soloist. These tasks require that the system is able to operate in real time, that is to infer the position along the score *while* the human player is perform-

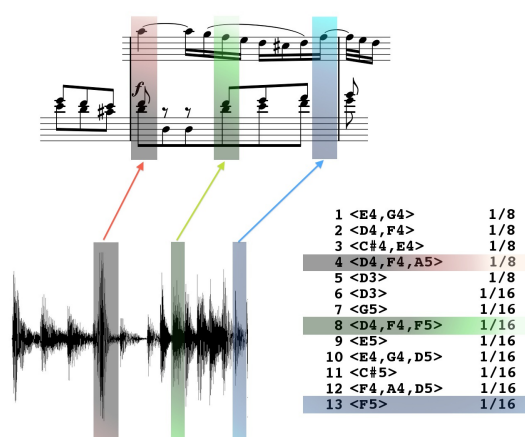


Figure 1. Alignment of an audio recording to symbolic score, represented both graphically and as a list of events.

ing, making a decision on the most probable position also in case of mismatches between the expected performance and the incoming signal and synchronize accordingly.

It is worth mentioning that the real-time requirement has two kinds of implications. On the one hand, in the case of systems that react at individual note events, the computation time should be below the perceptual threshold at which listeners perceive two events as distinct, even though there is a slightly delay between them. In this case efficiency is more important than effectiveness. On the other hand, on-line synchronization means that no information on future events can be used for computing the local position: a mistake made by the musician can be reflected by a mistake of the system in computing the actual position. In the latter case, the system can correctly realign with the performer only after a number of correct notes are observed.

In many applications alignment does not need to be carried out in real-time. In particular, we focus on applications where alignment can be carried out on a recording of a performance, which is more commonly available for non musicians users. It can be noted that the relaxation of real-time requirements has a positive impact on the quality of local alignment, because the synchronization between the audio signal and the events in the score can be carried out using, at any point, the complete information of the performance (i.e. the choice of current position can be based also on the knowledge of future events).

Off-line alignment can be used to highlight some characteristics of a music performance by exploiting the information carried in the score. For instance, the presence of a repetition, the directions about dynamics, the presence of a *fermata* or of a *legato*, are usually clearly marked in the score, but may be difficult to infer by simply listening to a recording, especially for musically untrained users.

To this end, this contribution intends to propose tools aimed at filling the gap between the access of music content by musicians and musicologists, who normally take most of the information about a music work directly from a structured music score, and musically untrained listeners, who are usually not able to read a music score but enjoy listening to recordings.

2. APPLICATIONS

Methodologies and systems for the automatic alignment between a music performance and the corresponding score have usually been developed as tools for musicians. Aligning a music score and an audio performance might seem at first to be of little use for a person without music training, because the latter is incapable to interpret the information expressed by the score and is thus unable to make any use of it. On the contrary, we argue that even such listener, being guided through a performance on the corresponding music score, could be facilitated in many task, such as:

- Accessing annotations made by the composer, such as indications for the performers, description of the tempo, and changes in time and key signatures.
- Visualizing comments by a reviser in an appropriate context.
- Identifying patterns in the compositions, repetitions, reprises of the main theme, and variations.
- Anticipating forthcoming events, actively listening to particular cues, recognizing the instruments that are playing.
- Selectively following one or more instruments, even if they are not playing the main melody, or one or more themes in a counterpoint.

It should be noted that trained listeners are able to accomplish most of these tasks also without a music score, yet it is not common for persons without a musical education to have these abilities. On the other hand, the possibility of easily accessing all this information may improve the quality of a listening experience, and consequently open novel possibilities in music education. We might even go as far as arguing that such tools might help overcoming the diffidence of many potential students towards the steep learning curve that characterizes, at least in the common conception, formal music education.

In this paper we propose two sets of tools which aim at visually and aurally enriching the listening experience.

2.1 Visual Tools

The simplest approach to a visual aid to listening experience consists in presenting to the user a series of textual annotations, made by an expert, in synchronous with the listening of a performance. Annotations could be added using a custom graphical user interface, or could be directly added to a MIDI file when using such format for representing the score, using the available textual fields provided by the standard. Part of the information, in particular changes in key and time signatures could be automatically detected and presented to the user¹, in addition to the name of the instruments that start or stop playing. Ideally, the user would be allowed to activate or deactivate these functionalities. Information on the instruments playing in a particular moment, could be represented using different colors on a piano roll representation. Hence a pictorial representation of the set of instruments that are employed in a particular composition can be exploited by highlighting the active instruments at each moment, or the instant in which they start playing, or the compositional material that they are playing.

Commercial software implementing many of these functionality already exists, such as *iAnalyse*², depicted in Figure 2. Even though the existence of commercial software is an indicator of the interest for such tools, to the best of our knowledge none of them executes automatic synchronization to an audio recording, rendering the synchronization process necessarily manual and relegating the usage of the software to sole musicians. Moreover, the task of manual alignment is extremely tedious, thus it can be seen how this absence is limiting the market for a software that is already directed at a restricted niche.

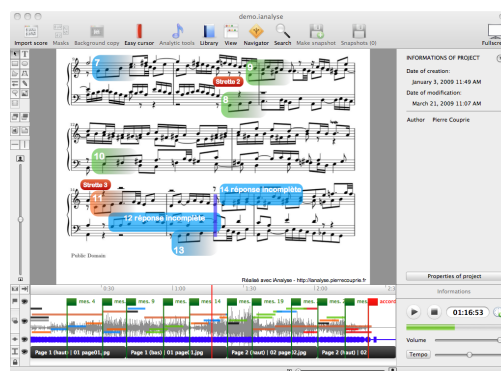


Figure 2. Screenshot of the music analysis tool *iAnalyse*.

Another set of new possibilities is opened by algorithms being researched in the field of Music Information Retrieval, such as the identification of recurrent patterns and musical phrases [6, 7]. All these tools would clearly benefit considerably from their integration in a graphical user interface that links their results directly to audio material.

¹ The automatic detection of key and tempo changes in MIDI files is dependent on the correct usage of the control messages provided by the format.

² <http://logiciels.pierrecouprie.fr/>

In addition to the analysis of a score, studies often revolve around the comparison of different interpretations of a piece. The aspect that benefits the most from an alignment to a recording is the analysis of tempo: Figure 3 depicts a simple tool that we prototyped in order to visualize the difference of interpretation between two or more performances from the point of view of tempo fluctuations.

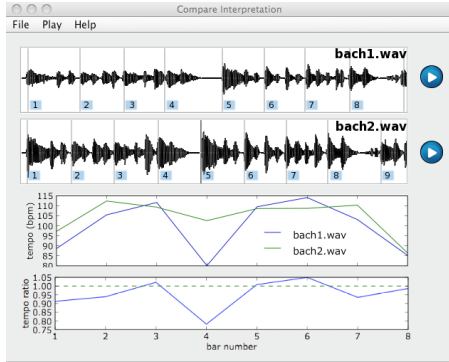


Figure 3. Comparison of tempo at the bar level between two performances.

2.2 Auditory Tools

As was argued above, an annotation of an audio recording in the form of tagging of all the entry points of each instrument, could be created by a teacher in order to show students details of instrumentation and orchestration which might not appear evident at a first hearing. Such temporization of a performance to visual cues in the score may be exploited also for the development of auditory tools.

In particular, we developed a simple tool called *AudioZoom*, a system for the auditory highlight of an instrument combination in a complex polyphony. The basic idea is that, through alignment, a polyphonic musical performance can be decomposed into its individual components. The problem is commonly known as source separation, which is usually blind because it is assumed that (almost) no information is available about the role of the sources. In our case, having the score as a reference, the identification of the different acoustic sources is straightforward, because, after alignment, the system has a complete knowledge about the notes played, at each instant, by each instrument. The user can select one or more instruments, one or more particular musical themes or patterns, or any combination, and the system can selectively amplify the selected elements. The final effect is to put on the front the interested elements.

Even though more advanced approaches are currently being researched, for instance in the field of *informed source separation*, we obtained satisfactory results with a basic model based on a bank of bandpass filters centered at specific times around the harmonics of a selected instrument track, as in Figure 4 (the group delay of each filter at its center frequency is compensated by an appropriate fractional delay-line in order to obtain a constant delay on all

bands). In particular, the user can select one channel from the MIDI file representing the score, and the system synchronizes the opening of the different filter banks with the audio recording, selectively filtering out the other signals.

This process is depicted in Figure 5.

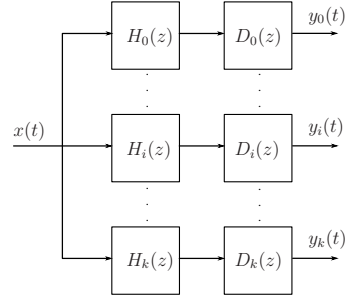


Figure 4. Filter bank structure used by AudioZoom.

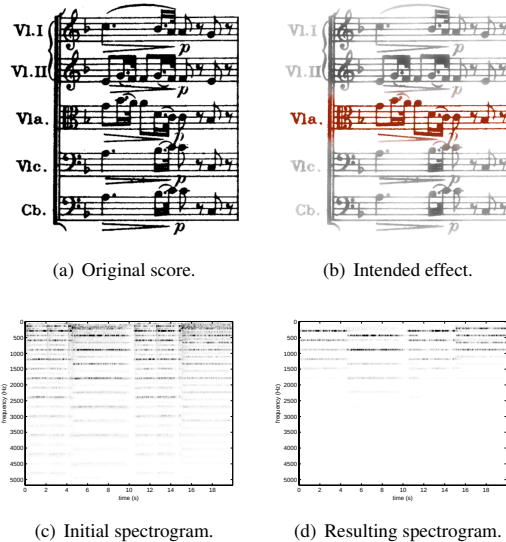


Figure 5. Effect of AudioZoom on an excerpt of Haydn's Symphony Op. 104 – Highlighting of the viola part.

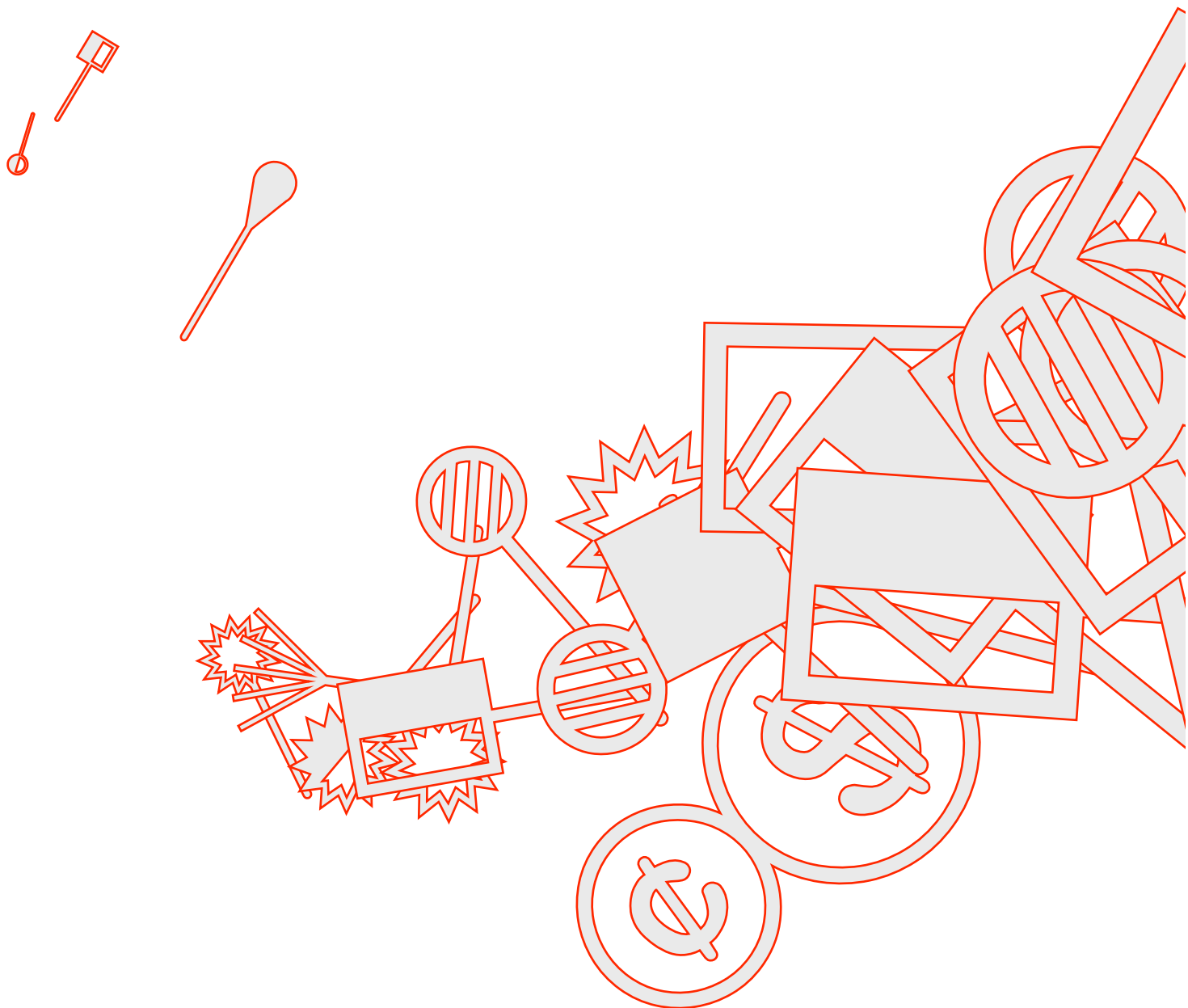
3. CONCLUSIONS

We briefly introduced a set of tools that are based on audio-to-score alignment, aimed at bridging the gap between the enjoyment of music by trained musicians and untrained final users. Our goal is to enrich the information carried by the recording of a performance with the information carried in the score. The main application of the proposed tools is music education, where on the one hand the study of music interpretation can be improved by accessing structural information reported in the score, and on the other hand the study of compositions based on score analysis can be eased by a direct access to the corresponding parts of one or more performances.

4. REFERENCES

- [1] N. Orio and F. Déchelle, “Score Following Using Spectral Analysis and Hidden Markov Models,” in *Proceedings of the ICMC*, (Havana, Cuba), 2001.
- [2] S. Dixon, “Live tracking of musical performances using on-line time warping,” in *Proceedings of the 8th International Conference on Digital Audio Effects*, 2005.
- [3] A. Cont, “A coupled duration-focused architecture for realtime music to score alignment,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.
- [4] C. Joder, S. Essid, and G. Richard, “A conditional random field viewpoint of symbolic audio-to-score matching,” in *Proceedings of the international conference on Multimedia*, MM ’10, ACM, 2010.
- [5] N. Montecchio and A. Cont, “A unified approach to real time audio-to-score and audio-to-audio alignment using sequential montecarlo inference techniques,” in *Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing*, 2011.
- [6] T. Collins, J. Thurlow, R. Laney, A. Willis, and P. Garthwaite, “A comparative evaluation of algorithms for discovering translational patterns in baroque keyboard works,” in *Proceedings of the International Symposium on Music Information Retrieval*, 2010.
- [7] D. Conklin and M. Bergeron, “Discovery of contrapuntal patterns,” in *Proceedings of the International Symposium on Music Information Retrieval*, 2010.

6. Composizione assistita



+ Massimo Avantaggiato

+ Giorgio Klauer

+ Mauro Lanza, Gianluca Verlingieri, Nicola Biagioni

+ Giorgio Bianchi, Pier Daniel Cornacchia

+ Sergio Canazza, Antonio Rodà, Lauro Snidaro, Ingrid Visentini

“COMPOSIZIONE ASSISTITA E PROCESSI DI TRASFERIMENTO DI DATI MUSICALI DA PWGL A CSOUND”

Massimo Avantaggiato
Conservatorio G. Verdi – Milano
mavantag@yahoo.it

ABSTRACT

Fin dalla sua introduzione nei corsi di Musica Elettronica ad opera del compositore J. Torres Maldonado, PWGL [1] si è imposto come uno strumento importante per lo sviluppo della creatività musicale degli allievi del Conservatorio G. Verdi.

Le applicazioni di PWGL sono molteplici: oltre ad essere un valido ausilio didattico nella ricostruzione del percorso compositivo di brani storici, esso offre nuove possibilità nella creazione di brani originali e nell'applicazione delle varie tecniche di sintesi sonora.

La mia attenzione si è subito focalizzata sull'esigenza di creare collegamenti funzionali tra PWGL e Csound, collegamenti che avrebbero consentito di integrare i processi di composizione algoritmica con quelli di sintesi sonora.

1. INTRODUZIONE

In un brano, normalmente, si scrivono migliaia di righe di istruzioni. La maggior parte dei musicisti, però, non compone brani scrivendo una nota alla volta, ma utilizza programmi per generare le score.

Tali programmi, che si chiamano score generator [2], hanno l'obiettivo di sollevare il compositore dalle attività ripetitive di digitazione di lunghe righe di codice. PWGL, programma di composizione algoritmica basato

sul Common Lisp [3], può essere impiegato come score generator e interfacciato con Csound, diffuso sound renderer basato sul linguaggio C++ , può diventare anche un potente motore di sintesi [4].

PWGL senza l'ausilio di librerie esterne consente anche a chi non è un esperto programmatore di:

- 1) poter realizzare gesti elettronici notati e con un pieno controllo dei parametri sonori coinvolti nel processo “compositivo” del suono: strumento, attacchi, note, ampiezze, frequenze, pan, riverberi, delay etc.;
- 2) poter effettuare un rapido rendering in c-sound, con possibilità di reiterazione dei processi e di correzione delle patch di PWGL;
- 3) poter realizzare, mediante semplici algoritmi, un'efficiente implementazione di diverse tecniche di sintesi.

2. SINTESI ADDITIVA E KARPLUS

Come primo esempio, partiamo dalla sintesi Karplus e Strong: la patch che illustriamo permette di optare tra alcuni spettri di partenza, che saranno sottoposti ad un processo di accelerazione o decelerazione.

Il gesto risultante verrà sintetizzato in Csound, utilizzando l'opcode seguente [5]:

```
ar pluck kamp, kcps, icps, ifn, imeth[iparm1,iparm2] [6]
```

1 “PWGL could be seen as an attempt to fill the gap between several different aspects of music tuition. It is our belief that PWGL could be established as a pedagogical tool for the academia”[7].

2 Gli Score Generator sono stati scritti in molti linguaggi: C, C++, Java, Lisp, Lua e Python. Gogins in un suo scritto elenca diversi tipi di Score Generator e Composing Environment tra cui elenca AthenaCl, Blue, Common Music, Pure Data [4].

Copyright: © 2010 Avantaggiato. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

3 Per approfondimenti: [1,21].

4 Questa è un'alternativa all'impiego del modulo PWGLSynth già in dotazione con PWGL [8-10].

5 Per approfondimenti sulla tecnica di sintesi Karplus e Strong si rimanda alla bibliografia [5, 6].

6 ove:

- kamp = ampiezza del suono;

- kcps = freq. desiderata.

- icps = valore in Hz che determina la lunghezza della tabella.

Solitamente = kcps, ma può essere posto maggiore o minore per ottenere particolari effetti timbrici

- ifn = numero della funzione usata per inizializzare la tabella.

Quando ifn = 0, la tabella viene riempita con una seq. random di valori (algoritmo originale di Karplus e Strong)

-imeth = metodo usato per la modifica dei valori della tabella durante la generazione del suono. Ve ne sono sei [3].

Scriviamo il nostro strumento:

```
instr 1
idur = p3
iamp = p4
ifrq = p5
ipanl = sqrt (p6)
ipanr = sqrt (1-p6)
kenv linseg 1, p3*.50, 1, p3*.25, 0
a1 pluck iamp, ifreq, ifreq, 0, 6
outs a1*kenv*ipanl, a1*kenv*ipanr
endin
```

e procediamo a creare la patch illustrata nella fig. 2, attraverso i seguenti passi:

- in 1) individuiamo la sequenza di note di partenza;
- in 2) costruiamo l'abstract- box, che ha la funzione di generare i dati per la sintesi e, eventualmente, di procedere con l'accelerazione o la decelerazione del materiale;
- in 3) raccogliamo i dati in una text-box per realizzare la sintesi in Csound.

L'abstract-box, vista dall'esterno, si compone di una serie di slider cui sono collegati degli ingressi(Fig.1).

Ad ogni parameter-field possono corrispondere uno o più input (Fig.3):

- 1) **INSTRUMENT (p1):** i valori assegnati allo strumento (i1) vengono ripetuti un numero di volte pari a quello delle frequenze che compongono lo spettro.

- 2) **CREATION TIME (p2):** usando una funzione parabolica vengono riscaldati o meno gli attacchi della sequenza originale, per dare origine ad accelerazioni o decelerazioni del materiale iniziale;

- 3) **DURATE DELLE NOTE (p3):** comprese tra un valore minimo e massimo, sono definite usando g-scaling in percentuale rispetto alle distanze relative tra i creation time prima definiti;

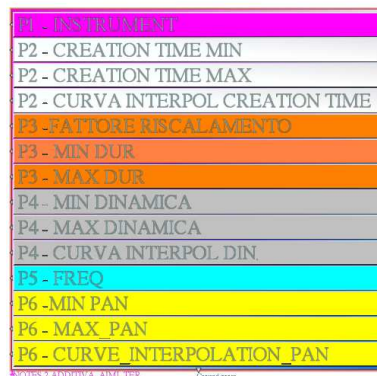


Fig.1: Gli slider componenti l'abstract-box per effettuare il trasferimento dei dati a Csound.

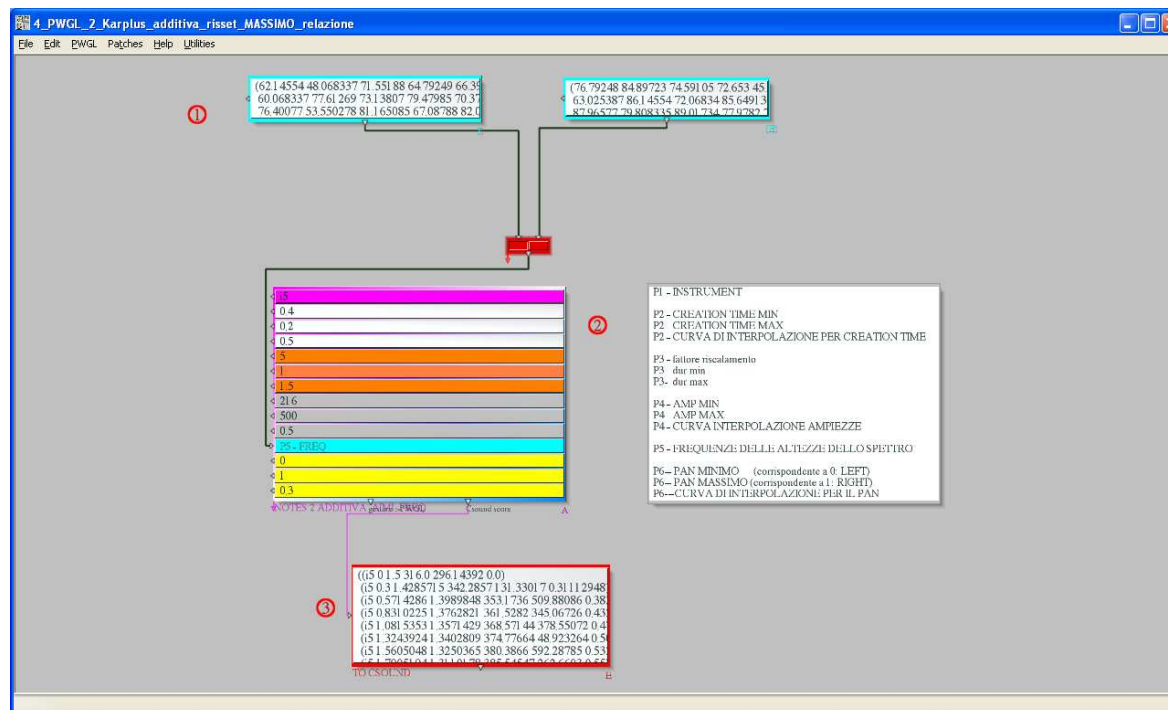


Fig.2: La patch in PWGL che consente di trasferire i dati in Csound. In 1) sono selezionabili gli spettri iniziali (armonici o inarmonici); in 2) è visibile l'abstract- box il cui dettaglio è fornito dalla figura 1). In 3) la score relativa allo spettro optato in 1).

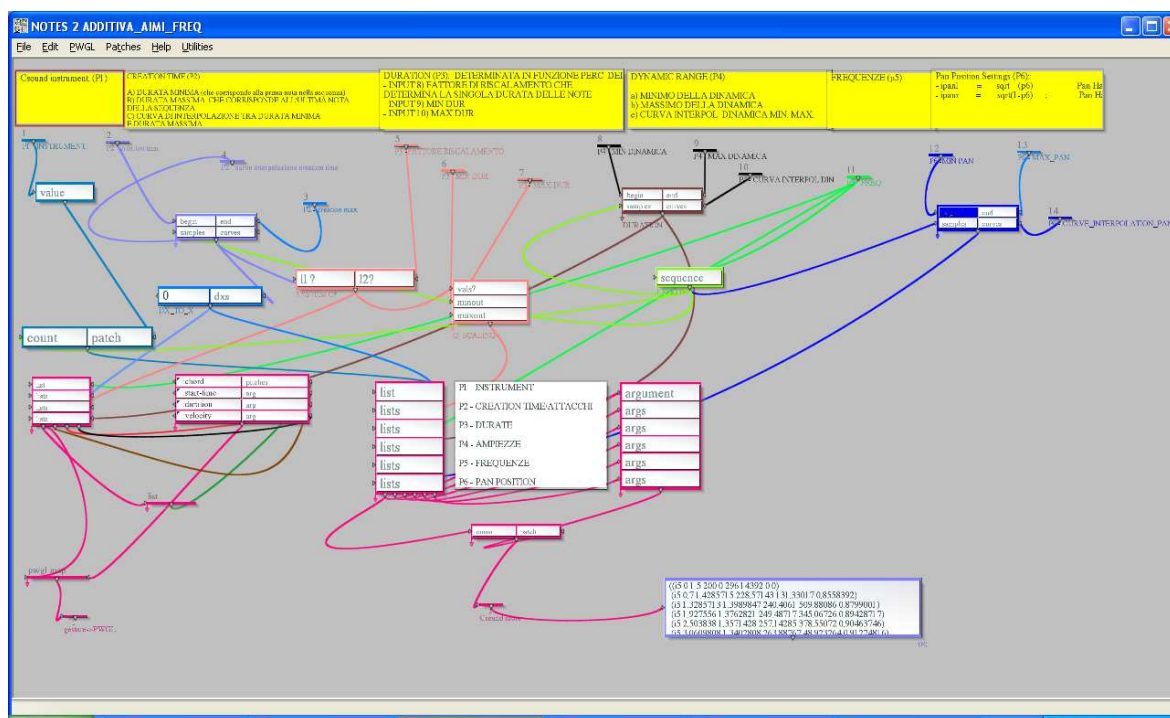


Fig.3: L'abstract- box vista dall'interno. Sono visibili gli ingressi relativi a: 1) strumenti; 2) attacchi; 3) durate; 4) ampiezze; 5) frequenze; 6) pan. In 7) PWGL-enum box è collegato al primo ingresso di PWGL-map: in accoppiata consentono di creare loop e processare più liste di valori simultaneamente, nella fattispecie quelli relativi ai parameter-field. Ad ogni interazione il risultato viene depositato in una text-box che funge da "result-list": questa è la score di Csound.

4) AMPIEZZA (p4): compresa tra un minimo e un massimo dinamico: può essere espressa in valori assoluti o essere indicata in Db;

5) FREQUENZE (p5): sono le frequenze degli spettri iniziali; in caso di impiego di grandezze midi occorrerà prima effettuare una opportuna conversione mediante la funzione "Patchwork Midi to Frequencies";

6) POSIZIONAMENTO SUL FRONTE STEREOFONICO [7] (p6): sempre tramite una funzione parabolica si interpola tra un valore massimo pari ad uno (suono tutto a sinistra) ed uno minimo pari a zero (suono tutto a destra).

Per i parametri relativi a creation time, durate, ampiezze e pan, si utilizza dunque l'interpolazione basata su una curva concava/convessa, anche se questa non è l'unica forma di interpolazione possibile.

Definita la sequenza di note di partenza, è possibile sostituire la patch per la sintesi Karplus prima descritta con una per la sintesi additiva e passare da una

tecnica di sintesi ad un'altra, [8], impiegando ad esempio il seguente strumento:

```
giamp_fact      = 16      ; Amplitude factor

Instr 2
  idur          = 1/p3
  iamp          = p4 * giamp_fact
  ifreq         = p5
  ipanl         = sqrt(p6)      ; Pan left
  ipanr         = sqrt(1-p6)    ; Pan right
  aamp          oscili iamp, idur, 2
  aout          oscili aamp, ifreq, 1
  outs          aout*ipanl, aout*ipanr
endin
```

Dagli esempi illustrati risulta evidente che gli strumenti, una volta creati, possono essere reimpiegati in altre situazioni, indipendentemente dal materiale di partenza e dai processi algoritmico-compositivi (Fig. 9-10).

⁷ "The most popular spatial illusion are horizontal panning – lateral sound movement from speaker to speaker – and reverberating – adding a dense and diffuse pattern of echoes to a sound to situate it in a larger space.

Vertical panning (up and down and overhead) can also create striking effects in electronic music [19]."

Mickelson parla di strategie per realizzare il pan e offre una classificazione di alcune tecniche di base: Simple Pan; Square Root Pan; Sine Pan; Equal Power Pan; Delay and Filtered Panning [15]

⁸ Per un approfondimento delle tecniche di sintesi additiva [11-14] e [16-18].

2.1 SINTESI ADDITIVA: TRA RICOSTRUZIONE E REINTERPRETAZIONE

PWGL offre la possibilità di ripercorrere il processo compositivo di un brano storico, di enucleare l'idea concettuale individuando regole compositive ed eventuali eccezioni.

Tra i lavori che ho personalmente ricostruito, la mia attenzione si è concentrata su brani come Studio II di K. Stockhausen. In questo secondo studio il compositore tedesco impiega 81 frequenze partendo dal limite grave di 100 Hz.

Da questa frequenza base egli ottiene una scala di successive 80 frequenze, una scala di 25 intervalli uguali, partendo dall'armonico 1 all'armonico 5; mentre nel temperamento equabile il rapporto di frequenza è l'esatta radice dodicesima di 2.

Stockhausen utilizza intervalli più ampi del semitono, con un rapporto di frequenza dato dalla radice venticinquesima di 5: ogni frequenza successiva è pertanto ottenuta moltiplicando la precedente per 1.066494.

Cinque è un numero ricorrente nel brano: 5 note formano una sequenza; 5 sequenze formano un set o gruppo ("sequenzgruppe"). Individuate le frequenze, le durate in rapporto alla lunghezza del nastro (2.5 cm sono pari a 0.039 sec), l'intensità, per ciascuna sequenza o gruppo si possono ricostruire le diverse sezioni che compongono il brano: 5 parti più una coda finale [9], come si può vedere dalla fig. 6 relativa alla ricostruzione della prima sezione del brano.

Oltre alla ricostruzione di brani storici, PWGL consente di reinterpretare situazioni sonore del passato: ciò può rappresentare una buona palestra soprattutto per chi approccia Csound per la prima volta.

In questo secondo esempio, adattato da un esercizio del compositore J.Torres Maldonado, creiamo un tessuto micropolifonico di natura spettrale partendo da uno strumento disegnato da J.C.Risset. [10] e procediamo, come di consueto, a creare una nostra patch per trasferire il risultato in Csound.

Questo bel suono, simile al canto armonicale tibetano, produce un arpeggio con la serie degli armonici: si possono sentire distintamente la fondamentale e le parziali da 5 a 9.

La score di Risset è stata modificata impiegando un pitch di partenza di C2 - 65.41 Hertz., invece della frequenza originaria di 96 Hertz.

Si utilizzeranno le note non presenti nell'arpeggio per generare nuovi spettri, per cui, data la frequenza di partenza di 65.41 Hertz, otteniamo il nuovo set di altezze (Fig.5): 69.3; 77.78; 87.31; 92.5; 103.83; 110.0; 123.47 Hertz, che insieme alla frequenza iniziale costituiranno le fondamentali dei nuovi spettri

alla base del tessuto micropolifonico.

Anche in questo caso l'abstract (Fig.8) collezionerà le informazioni musicali della score di Csound: strumento; start time, definiti da una serie di Fibonacci; durata, derivata dal retrogrado della serie; frequenze, ampiezze; offset.

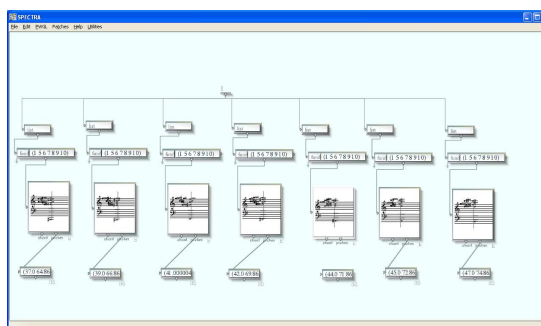


Fig.4: generazione dei nuovi spettri con fondamentale 69.3;77.78; 87.31; 92.5; 103.83; 110.0; 123.47

F1	0	1024	10	1	0	0	.7	.7	.7	.7	.7	.7
; start dur freq amp offset												
i1	1	68	65.41	1500	.03							
i1	2	42	69.3	1500	.03							
i1	3	26	77.78	1500	.03							
i1	5	16	87.31	1500	.03							
i1	8	10	92.5	1500	.03							
i1	13	6	103.83	1500	.03							
i1	21	4	110	1500	.03							
i1	34	2	123.47	1500	.03							
endin												

Fig.5: Informazioni musicali della score di Csound

3. CONCLUSIONI

In questo articolo si descrive come integrare PWGL e Csound, al fine di costruire una collezione di strumenti Csound. Ciò consente ai musicisti di sottrarsi all'uso di programmi e librerie a pagamento, la cui realizzazione è spesso influenzata dalle idee e dagli approcci, anche compositivi, dei realizzatori e permette invece di farne di proprie.

Con queste poche indicazioni, ogni musicista può infatti creare delle librerie dedicate alla sintesi sonora, modificabili ¹¹, a seconda delle esigenze che via via si presentano.

Ciò che è illustrato è il primo risultato di un work in progress che porterà al rilascio di una suite di

⁹E' fondamentale il riferimento alla lettura della partitura [22] e all'analisi di H. Silberhorn [20].

¹⁰Lo strumento, creato dal compositore francese Jean Claude Risset per il brano "Mutation", è contenuto in The Amsterdam Catalog of Csound Computer Instrument, disponibile sul sito http://www.music.buffalo.edu/hiller/accci/02/02_43_1.txt.html [16]

¹¹ E' vero quanto dice M. Stroppa: "Since the estetical needs of a musician cannot be guessed every attempt at searching for a more general solution must generate a System both as open as possible and very easy to personalize. The musician's first task will then be to adapt it to his or her own particular way of thinking about sonic potentials"[23].

sintetizzatori o virtual synthesizer per gestire orchestre e sintesi in Csound; questi strumenti si affiancheranno ai moduli già esistenti dei PWGLSynth. PWGL diventerà nelle nostre intenzioni anche un ambiente per il disegno e la rappresentazione grafica degli strumenti.

E' facile prevedere che PWGL, grazie al supporto di una comunità sempre più estesa di sviluppatori e utilizzatori finali, acquisterà un ruolo sempre più importante nell'ambito della composizione algoritmica e della sintesi sonora

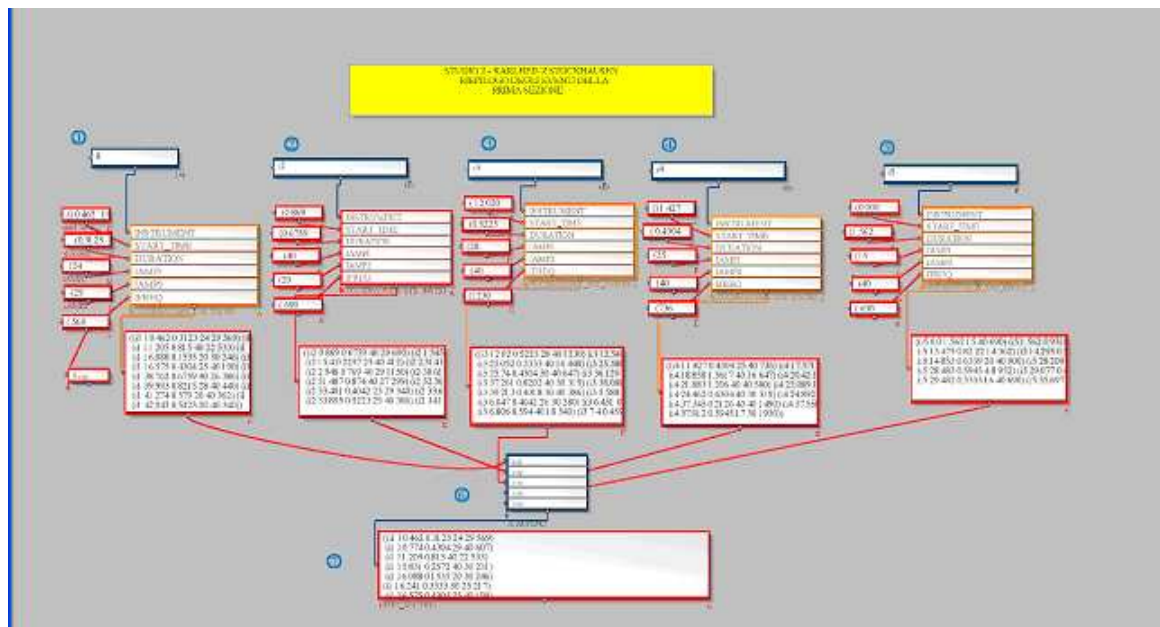


Fig.6: Prima sezione di Studio II. In (1-5) è visibile il materiale relativo ai 5 strumenti o "mixture" che compongono i "gruppen". Nella result list (7) è indicata la partitura finale ottenuta dalla aggregazione mediante x-append delle score relative a ciascuno strumento (6).

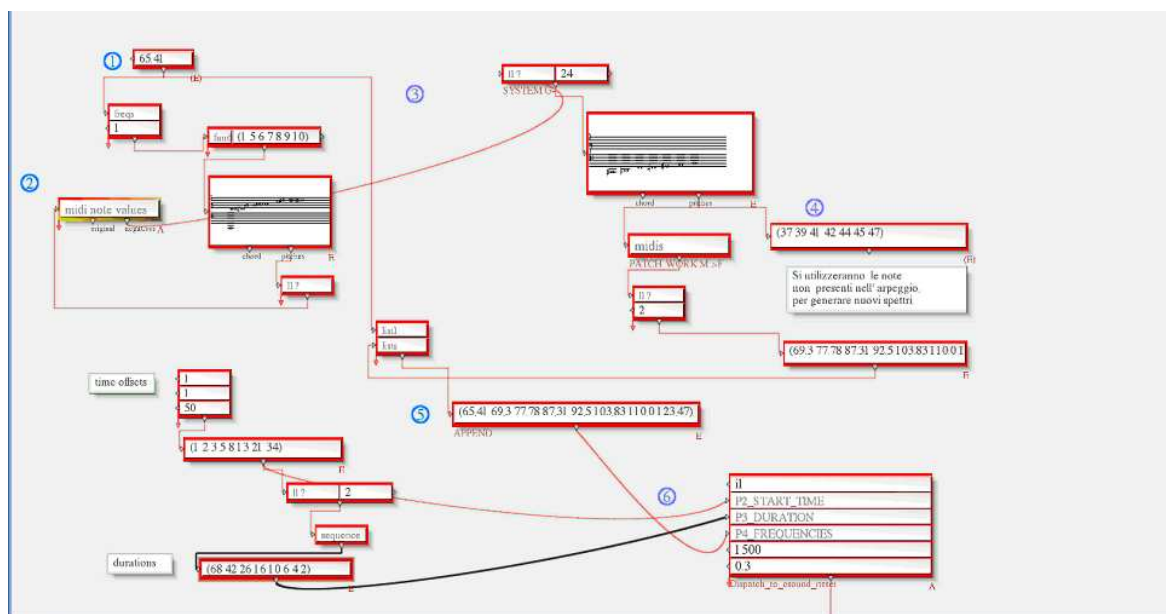


Fig.7: in (1) si individua la fondamentale dell'arpeggio o frequenza base (65,41 Hertz) sulla quale si costruiranno gli armonici (2). In (3) lo spettro viene "negativizzato" ottenendo set di frequenze che costituiranno le fondamentali per nuovi arpeggi (4). In (5) la frequenza base iniziale viene aggiunta alle frequenze da esse derivate, che vanno ad alimentare l'abstract- box (6)

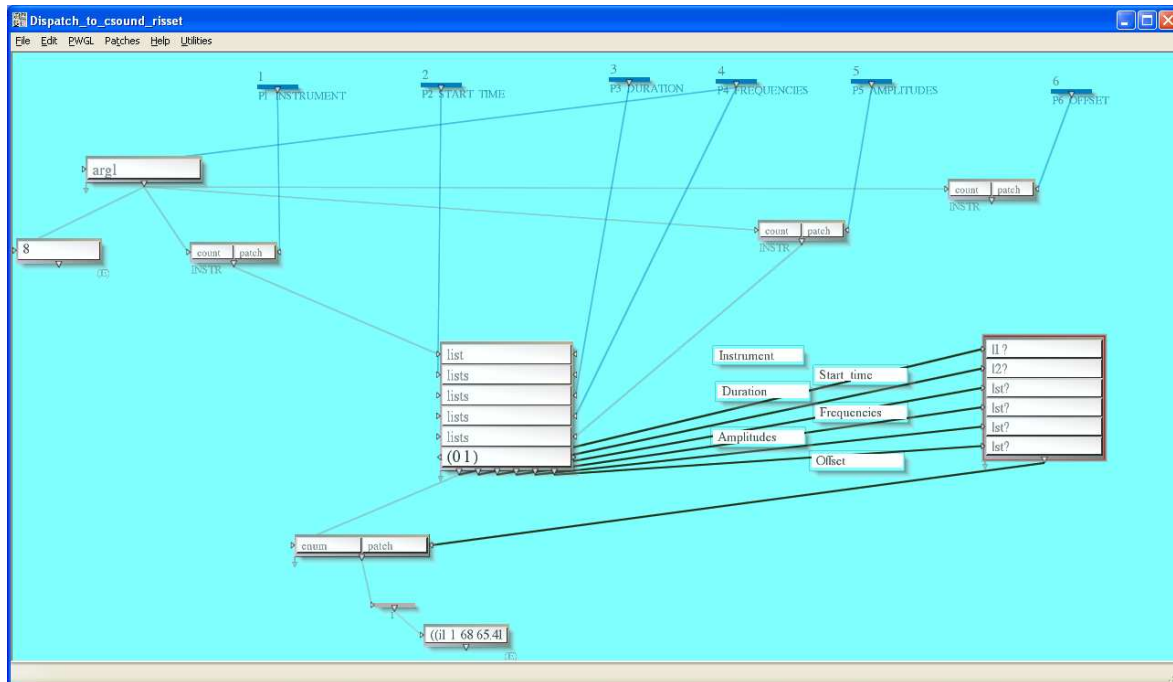


Fig.8: Dettaglio dell'abstract- box di Fig.6: sono visibili tutte le informazioni musicali da trasferire a Csound

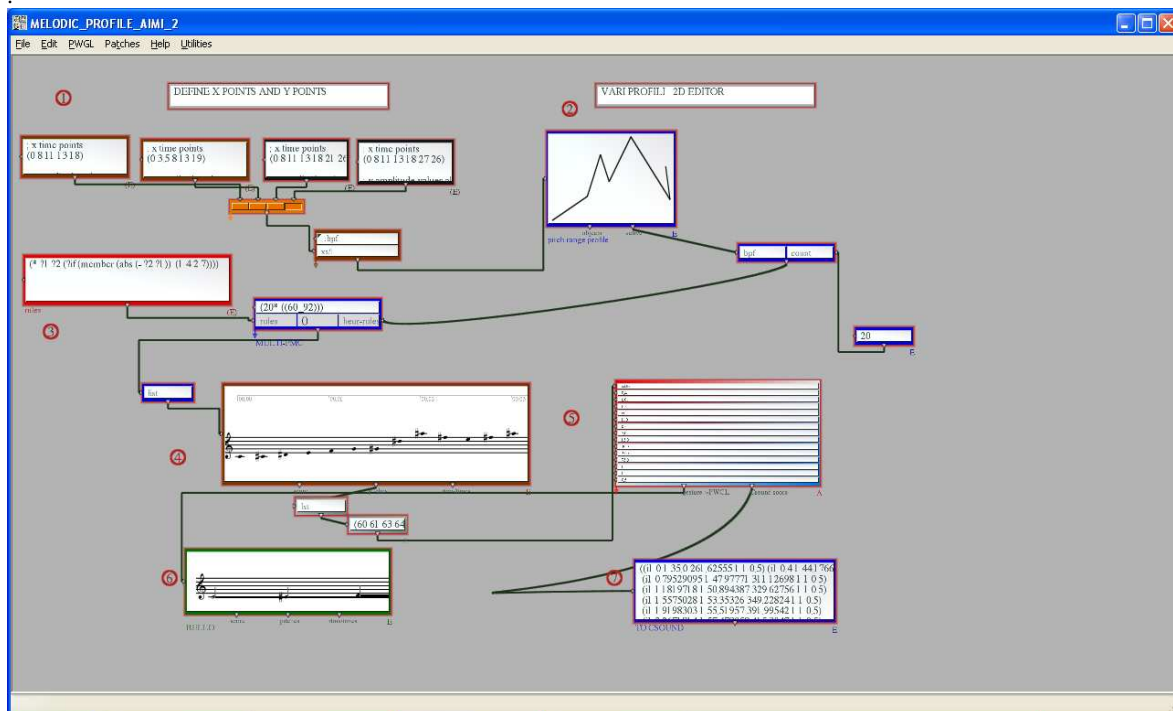


Fig.9: Una sequenza di note midi compresa tra degli estremi, ricalca un profilo melodico scelto tra quelli disponibili (2). Il profilo melodico è costruito mediante valori cartesiani (x,y) assegnati in una text-box (1). Il risultato viene influenzato dal tipo di intervallo prescelto ricorrendo alle regole euristiche definite in (3). La sequenza risultante viene accelerata o decelerata e fornisce la sequenza in (6) con la relativa score in Csound (7).

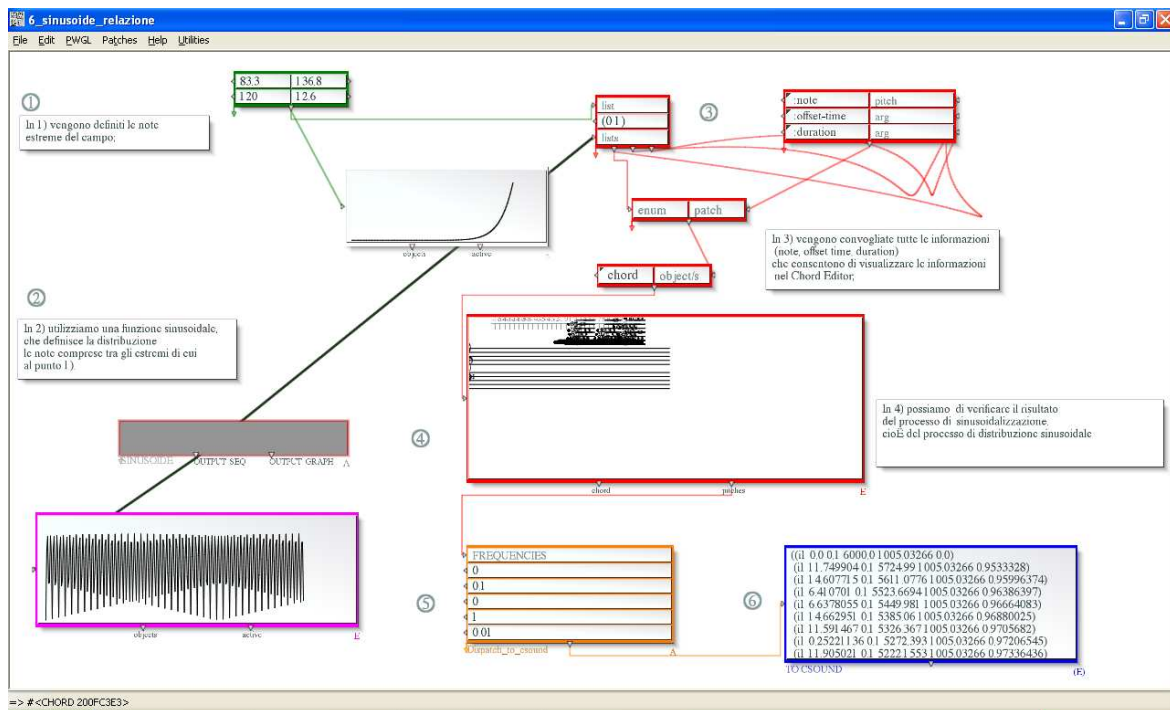


Fig.10: Una sequenza di note definite tra degli estremi segue un profilo melodico di tipo sinusoidale incrociato con una curva concava/concava. In 1) vengono definite le note estreme della sequenza; in 2) utilizziamo una funzione sinusoidale, per distribuire le note; in 3) usando PATCH-WORK INTERPOLATION interpoliamo tra due valori midi, che sono gli estremi in cui si “muove” la fascia sonora. In 3) vengono convogliate tutte le informazioni (note, offset time, duration) che consentono di visualizzare le informazioni nel “Chord Editor”; in 4) possiamo verificare il risultato del processo di distribuzione sinusoidale. In 6) otteniamo il risultato da trasferire su Csound, tramite un’abstract-box (5).

4. RIFERIMENTI BIBLIOGRAFICI

- [1] Association of Lisp Users (A.L.U), <http://www.lisp.org/alu/home>.
- [2] R. Boulanger, *The Csound Book - Perspectives in software synthesis, sound design, signal processing, and programming*, R. Boulanger Ed., p. 370, 2000.
- [3] R. Bianchini, A. Cipriani, *Il suono virtuale - Sintesi ed elaborazione del Suono - Teoria e pratica con Csound*, pp. 343 - 362, 2007.
- [4] M.Gogins, *A Csound Tutorial*, pp. 54-55, 2006.
- [5] D. Jaffe, J.O. Smith “Extensions of the Karplus-Strong Plucked-String Algorithm”, *Computer Music Journal* 7(2), 1983. Reprinted in C.Roads *The Music Machine*. MIT Press, pp. 481-494, 1989.
- [6] K. Karplus, and A. Strong “Digital Synthesis of plucked string and drum timbres”, *Computer Music Journal* 7 (2): pp. 43-55, 1983.
- [7] M.Kuuskankare, M. Laurson. “PWGL, Towards an Open and Intelligent Learning Environment for Higher Music Education”, *Proceedings of the 5th European Conference on Technology Enhanced Learning*, EC- TEL 2010 – Barcelona, Spain, p.520, 2010.
- [8] M. Kuuskankare, M. Laurson, V. Norilo. “PWGLSynth, A Visual Synthesis Language for Virtual Instrument Design and Control”, *Computer Music Journal*, vol. 29, no. 3, pp. 29-41, 2005.
- [9] M. Laurson and V. Norilo. “Copy-synth-patch: A Tool for Visual Instrument Design”, *Proceedings of ICMC04*, Miami, 2004.
- [10] M. Laurson and V. Norilo. “Recent Developments in PWGLSynth”, *Proceedings of DAFx 2003*, pp. 69-72, London, England, 2003.
- [11] D. Lorrain, “Inharmonique, Analyse de la Bande de l’Oeuvre de Jean-Claude Risset”, *Rapports IRCAM*, 26, 1980.
- [12] M. Mathews, J.-C. Risset. “Analysis of Instrument Tones”, *Physics Today* 22(2): pp. 23-30, 1969.
- [13] F.R. Moore, 1977, 1985. “Table Lookup Noise for Sinusoidal Digital Oscillators”, *Computer Music Journal* 1(2):26-29.

Reprinted in C. Roads and J. Strawn, *Foundations of Computer Music*. MIT Press, pp. 326-334, 1985.

- [14] J.A. Moorer, “*Analysis-based Additive Synthesis*” in Strawn, J., *Digital Audio Signal Processing: An Anthology*. A-R Editions, pp. 160-177, 1985.
- [15] H. Mickelson “Panorama” , *Csound Magazine*, Autumn 1999.
- [16] J.-C. Risset, *An Introductory Catalogue of Computer Synthesized Sounds*, reprinted in “The Historical CD of Digital Sound Synthesis”, *Computer Music Currents* n° 13, Wergo, Germany, 1969.
- [17] J.-C. Risset, *Additive Synthesis of Inharmonic Tones* in M.V. Mathews and J.R. Pierce, eds. 1989. *Current Directions in Computer Music Research*. MIT Press, pp. 159-163, 1989.
- [18] J.-C. Risset, *Computer Music Experiments 1964-...* in C. Roads, “The Music Machine” MIT Press, pp. 67-74, 1989.
- [19] C. Roads, *Computer Music Tutorial*, MIT Press, p. 452, 1996.
- [20] H. Silberhorn, *Die Reihentechnik in Stockhausens Studie II*, 1980.
- [21] G.L. Steele, “*Common Lisp The Language*”, 2nd Edition. Digital Press, 1990.
- [22] K. Stockhausen, *Nr.3 Elektronische Studien, Studie II, Partitur*, Universal Edition, Zürich-London, 1954. Wien
- [23] M. Stroppa, *Paradigms for the high-level musical control of digital signal Processing*, Hochschule für Musik und darstellende Kunst Stuttgart, Germany, p.2, 2000.

UN AMBIENTE INTEGRATO DI PROGETTAZIONE, COMPOSIZIONE, ESECUZIONE E NOTAZIONE DI MUSICA MISTA. L'ESPERIENZA DI *DOPPELGÄNGER* (*HEIMLICH/UNHEIMLICH*) PER PIANOFORTE (DUE ESECUITORI) E SINTESI REAL-TIME PER MODELLIZZAZIONE FISICA DI UNA CORDA ALLE DIFFERENZE FINITE.

Giorgio Klauer

Conservatorio di Como, via Cadorna 4, I-22100 Como
klauer@alice.it

ABSTRACT

Sviluppato attorno a un algoritmo di sintesi real-time per modellizzazione fisica di una corda alle differenze finite, è stato realizzato in *SuperCollider* un ambiente integrato compositivo ed esecutivo, volto a estremizzare fino ai limiti del parossismo la ricerca timbrica delle morfologie sonore e a inserirla coerentemente nei formalismi del linguaggio musicale che contraddistingue l'opera per pianoforte ed elettronica scaturita da questa esperienza.

Il duplice proposito è stato attuato grazie a una serie di GUI di controllo programmate per agevolare l'invenzione delle morfologie e una SCORE alfanumerica dove disporre i necessari complessi di istruzioni. Sulla base della SCORE, in particolare, è stato possibile ricavare automaticamente la rappresentazione in partitura sia della parte pianistica che di quella elettronica. E' stato attuato infine un processo ulteriore di interazione, ricavando dall'analisi del suono di sintesi i parametri atti a informare la scrittura strumentale.

L'ambiente integrato ha pertanto reso possibile e accompagnato ogni fase del processo creativo, dalla progettazione alla composizione alla notazione all'esecuzione musicale, mentre l'uso esclusivo di *SuperCollider* ha radicalizzato la dialettica tra creazione e realizzazione, ripulendola dal maggior numero possibile di fattori eccentrici dovuti al passaggio fra diversi linguaggi e logiche degli applicativi.

1. INTRODUZIONE

Lo sviluppo dell'ambiente integrato per la composizione e l'esecuzione di *Doppelgänger*¹ risponde a istanze tecnico-artistiche molto specifiche, maturate nel costante confronto con le metodologie e i concetti dell'elaborazione digitale del suono e della composizione assistita dall'elaboratore.

L'essere funzione di un singolo prodotto artistico pone ovviamente dei limiti a questa realizzazione come apparato autonomo; alla sua astrazione in vista di composizioni future, si è voluto anteporre infatti la soddisfazione

di esigenze personali sulle quali è necessario fornire almeno un cenno di chiarimento.

Nel mio approccio alla composizione musicale mista, l'elaboratore viene considerato un macrostrumento sia compositivo che interpretativo atto a gestire coerentemente le prerogative e le specificità della parte elettroacustica/informatica rispetto a quelle della parte strumentale acustica; esse vengono messe in relazione secondo una serie di condizioni che ho progressivamente maturato, che sono:

- l'esclusione dei processi di analisi, sintesi ed elaborazione del segnale non attuabili in tempo reale (o quasi);
- la subordinazione della scelta dei metodi di sintesi e dei processi di elaborazione alla strategia compositiva e all'organico;
- l'annotazione simbolica, assieme a quella degli strumenti, dell'intera parte elettroacustica su di una partitura nel senso storico del termine, ovvero un compromesso grafico operativo/funzionale che serva da supporto di ideazione, composizione ed esecuzione;
- l'utilizzo del solo contenuto concettuale di contributi di terzi maturati in specifici ambiti di ricerca e non degli strumenti operativi contestualmente offerti;
- l'utilizzo di un'unica applicazione (*SuperCollider*) in quasi tutte le fasi della composizione musicale, per radicalizzare la dialettica tra creazione e realizzazione ripulendola dal maggior numero possibile di fattori eccentrici dovuti al passaggio fra diversi linguaggi e logiche degli applicativi.

2. RIFERIMENTI E MODELLI

Un modello importante nello sviluppo dell'ambiente integrato è stato il sistema Cordis-Anima descritto da Cadoz, Luciani e Florens [2] [3], unito all'ambiente interattivo di progettazione Genesis, descritto in Castagne e Cadoz [4]. L'ambiente realizzato per *Doppelgänger* ne costituisce in un certo senso l'interpretazione particolare, discostandosene per le finalità anche esecutive e l'implementazione real-time. A questo proposito, è stato importante il confronto con contributi che spostano la centralità del modello fisico dal suono all'interazione aptica, fra cui gli algoritmi *PMPD* proposti da Henry [11] e il modello di *scanned synthesis* di Verplank e Mathews [17], ripreso anche da Couturier [8].

¹ *Doppelgänger* (*heimlich/unheimlich*) per pianoforte (due esecutori) ed elettronica è stato commissionato e composto nel 2010 per il Duo Novalis (Stefania Amisano e Claudio Cristani) grazie al contributo dell'ÖGZM (Österreichische Gesellschaft für zeitgenössische Musik).

Copyright: © 2010 Giorgio Klauer. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

E' stato altresì utile il confronto con le composizioni musicali sorte dai diversi metodi di modellizzazione fisica, di cui è stata svolta una prima comparazione da Chafe [7] e nuovamente da Chafe assieme a Kojs e Serafin [13]². Per quanto riguarda il metodo di modellizzazione alle differenze finite, si tratta in pratica dei brani realizzati grazie al sistema di Cadoz: *résorption/couputre* (Tutschku, 2000), *pico..TERA* [5] e *Gaea* (Cadoz, 2003 e 2007), *At and Across* e *Zvonenje* (Kojs, 2007), *Shine* (Brümmer, 2007) e pochi altri. A questi si possono aggiungere le opere *Corda di metallo* (1998) [14] e *Canto di madre* (1999) di Lupone, dove il modello di corda eccitata per sfregamento implementa l'algoritmo di Hiller e Ruiz. Delle opere citate, le uniche composizioni per organico sia acustico che elettroacustico sono quelle di Kojs e di Lupone. Molto più numerosi sono i brani realizzati con sintesi modale, mediante le piattaforme di composizione assistita sviluppate presso l'IRCAM, e mediante guida d'onda, per le molteplici implementazioni real-time³.

L'ambiente realizzato per *Doppelgänger* integra un algoritmo di redazione automatica della partitura cartacea, che riceve in input la SCORE che serve ad eseguire la parte elettroacustica: in riferimento a questo specifico problema, del coordinamento immediato tra istruzione informatica e rappresentazione/istruzione per gli interpreti, sono state molto importanti le esperienze di realtime-scoring attuate da Gerhard Winkler [18] [19]: nelle composizioni per organico misto *KOMA* (1995/96) e *Hybrid II (NetWorks)* (1996-2001) in particolare, l'elaboratore elettronico serve sia per concretizzare algebricamente la forma musicale, anche mediante processi di analisi del segnale sonoro, sia per gestire le elaborazioni real-time, sia per pilotare l'interpretazione, attraverso una notazione ricavata in tempo reale e proiettata sugli schermi di computer portatili adibiti a leggio⁴.

3. L'ALGORITMO DI SINTESI

La sintesi avviene calcolando col metodo delle differenze finite le oscillazioni di una corda. Precisamente lo strumento in questione è un monocordo privo dell'accoppiamento di qualsiasi sistema di risonanza - perché questo dovrebbe essere fornito dall'ambiente di diffusione - composto da 32 elementi massa-molla-

smorzatore. Confrontando le implementazioni di Henry [11] con il porting in *SuperCollider* di Crabbe [9] è stato realizzato su questo software un algoritmo dove l'escursione delle 32 masse viene computata ricorsivamente su un array audio multicanale: i valori di massa, rigidità, viscosità e coefficiente di attenuazione, aggiornabili a frequenza di controllo, sono gli argomenti della funzione di sintesi che contiene la ricorsione.

Per mettere in vibrazione la corda, è stata modellizzata l'azione di un pizzicato, nella forma di un inviluppo audio attack-release applicato come limitatore al segnale di singoli canali. Variante di questa funzione, definita PLUCK, è la funzione STOP, un limitatore continuo che serve in particolare a simulare l'azione di sfioramento.

Per moderare l'apporto di segnale proveniente dai singoli elementi, in vista della diffusione elettroacustica, è stata modellizzata la trasduzione di un pickup per chitarra a riluttanza variabile con posizione e direttività arbitrarie e variabili dinamicamente. In attesa di una implementazione per array di altoparlanti, la diffusione risulta attualmente stereofonica, dove il panorama stereo del segnale di ciascuno dei 32 elementi coincide semplicemente con la sua posizione sulla corda.

4. CONTROLLO E RAPPRESENTAZIONE

Grazie all'algoritmo descritto si sono potute mettere in rilievo sonorità inudibili in condizioni d'ascolto reali e, specie assegnando in maniera dinamica e disomogenea i valori di massa e rigidità, attuando opportunamente le funzioni PLUCK e STOP, ottenere chimere sonore a dir poco sorprendenti.

Per agevolare l'invenzione delle morfologie sonore, sono state implementate alcune GUI atte al controllo puntuale dei parametri di sintesi; poiché, inoltre, le sonorità più interessanti e funzionali alla composizione si producevano controllando i parametri in maniera complessa, è stata formulata una specifica partitura alfanumerica (SCORE) con le istruzioni cifrate e disposte a blocchi. La medesima SCORE viene utilizzata dal regista del suono nell'esecuzione in pubblico.

Un riferimento per la sintassi della SCORE è stata la notevole proposta di Chafe [6] volta a formalizzare le azioni compiute nell'esecuzione al violoncello e implementare come macro alcune regole di prassi esecutiva. In quest'ultimo aspetto differisce l'implementazione di *Doppelgänger*, laddove si è preferito rendere organici i parametri di sintesi in modo inedito, secondo strategie individuate nel progresso della composizione. La coerenza dell'approccio operativo, legata all'utilizzo di una SCORE siffatta, da una parte ha permesso di estremizzare la ricerca timbrica ai limiti del parossismo, dall'altra ha rafforzato la coerenza dello linguaggio musicale stesso ed esaltato la possibilità di limare i dettagli.

Nella SCORE sono state inserite anche le informazioni relative ai parametri altezze-intensità-durate della parte del pianoforte: questo è servito a verificare il risultato d'assieme, accoppiando alla lettura automatica della SCORE da parte dell'algoritmo di sintesi, quella della notazione pianistica da parte di un campionario.

La SCORE contiene pertanto tutti i dati necessari a una rappresentazione grafica: tramite un algoritmo sono

² Raccogliendo le dichiarazioni dei rispettivi autori, l'articolo di Chafe del 2004 individua, sotto il denominatore comune della fascinazione rispetto alle morfologie sonore, tre tipi di approccio: l'estensione virtuale della liuteria tradizionale; la costituzione di uno spazio parametrico polarizzato tra realistica e irrealistica; la generalizzazione del modello estendendolo alla forma musicale. L'articolo del 2007 pone maggiormente l'attenzione sul metodo di modellizzazione, fornendo un'interpretazione globale in chiave cyber; dalla pubblicazione di questo articolo è sorto l'utile strumento di catalogazione online www.cybermusik.net.

³ Non sono note opere realizzate grazie al sistema TAO di Pearson [15].

⁴ L'ottica di Winkler è quella di formare interattivamente la composizione a partire da modelli matematici di sistemi dinamici complessi, tuttavia la sua tendenza a integrare in un unico ambiente informatico l'intero processo di produzione musicale ha indotto ad alcune utili riflessioni di ordine realizzativo. Nonostante la diversità di pensiero, si segnala anche una certa affinità per l'ambizione a conferire la dignità di strumento musicale all'elaboratore elettronico, restituendo in tal modo un margine di qualità auratiche alla composizione musicale elettroacustica.

state ricavate le immagini in formato pict, dalla cui stampa si è potuto confrontare bozze e differenti versioni, correggendo il codice in corso d'opera e chiosando i risultati in vista della redazione finale della partitura, portata a termine integrando manualmente una serie di indicazioni interpretative la cui numerizzazione non era oggetto d'interesse.

5. TECNICA MUSICALE ED ESEGESI

La commissione di *Doppelgänger* (*heimlich/unheimlich*) da parte del Duo Novalis si inserisce nel progetto *visibile(in)visible* centrato sul tema del doppio, che ho volto nell'organico stesso, individuando nel pianoforte il *doppelgänger* della parte elettroacustica, nel senso specificamente freudiano di una proiezione narcisistica derivata da irrisolti complessi⁵.

L'invenzione morfologica della parte elettroacustica ha portato a una struttura musicale dominata dalla coazione a ripetere, dove i corsi e ricorsi esprimono nelle forme più svariate l'autocompiacimento per il timbro prodotto: anche se all'ascolto non è sempre evidente, basta un rapido sguardo alla partitura per rendersene conto.

L'autocompiacimento del suono sintetico è tale da rendere la presenza del pianoforte profondamente destabilizzante, proprio perché la sua parte è *sintomaticamente* il preciso rispecchiamento di quella elettronica; questo fatto rende peraltro difficile e contraddittorio il riportare *Doppelgänger* ai brani appartenenti al repertorio ormai "storico" della composizione mista assistita dall'elaboratore, nei cui confronti vi sono i maggiori riferimenti a livello tecnico⁶. Inoltre, adottando il sistema tradizionale di notazione pentagrammatica, le tecniche compositive rendenti l'idea del rispecchiamento hanno portato all'estremo (e forse superato) il limite di eseguibilità pianistica.

Il primo tipo, più elementare, di rispecchiamento consiste nell'esplicitare linearmente in un segmento melodico al pianoforte, una dopo l'altra e a ritroso, le 32 autofrequenze degli elementi massa-molla, avulsi dal contesto della corda. La variante di questa tecnica è l'utilizzo della derivata delle autofrequenze, come deviazione temporale della regolare progressione nel tempo di 32 classi d'altezza disposte cromaticamente ("volate" cromatiche). In partitura, i segmenti pianistici vengono disposti simultaneamente alla sintesi attuata con quegli specifici parametri.

Il secondo tipo di tecnica consiste nel ricavare lo spettro medio di una sezione elettroacustica e mapparla sulle 88 classi d'altezza del pianoforte, utilizzando il valore di intensità come deviazione temporale rispetto a un asse collocato a un estremo della sezione elettroacustica che

funge da materiale. La variante di questa tecnica consiste nel mappare solo i picchi dell'involuppo spettrale.

Nell'interpretazione in chiave freudiana di quest'opera vi è anche l'attività propria del *doppelgänger*. Durante l'esecuzione il suono del pianoforte viene infatti campionato e analizzato, fornendo l'involuppo spettrale che, utilizzato come profilo di alterazione del valore di massa di singoli elementi della corda, serve a predisporla in maniera da produrre le chimere sonore e i parossismi timbrici di cui si è parlato.

6. BIBLIOGRAFIA

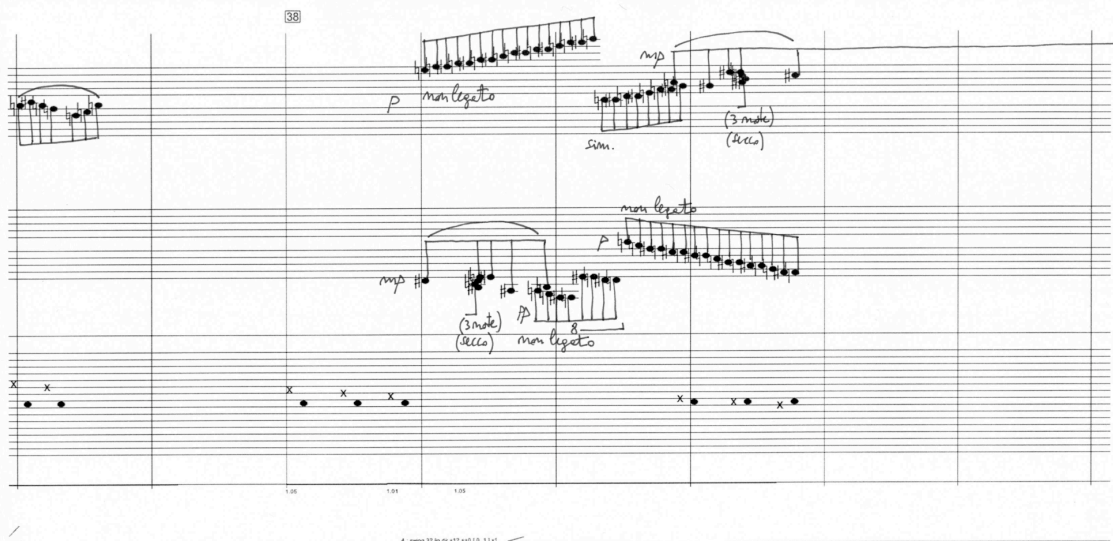
- [1] B. Bossis: "Les Chants de l'amour de Gérard Grisey. Entre rigueur formelle et jubilation humaniste", in D. Cohen-Levinas (cur.): *Le Temps de l'écoute. Gérard Grisey, ou la beauté des ombres sonores*, L'Harmattan, Paris, pp. 229-70, 2004.
- [2] C. Cadoz, A. Luciani, J.-L. Florens: "Responsive input devices and sound synthesis by simulation of instrumental mechanisms: the Cordis system", *Computer Music Journal*, Vol. 8, No. 3, pp. 60-73, 1984.
- [3] C. Cadoz, A. Luciani, J.-L. Florens: "CORDIS-ANIMA: A modeling and simulation system for sound and image synthesis - the general formalism", *Computer Music Journal*, Vol. 17, No. 1, pp. 19-29, 1993.
- [4] N. Castagne, C. Cadoz, "GENESIS: A Friendly Musician-Oriented environment for Mass-Interaction Physical Modeling", *Proceedings of the International Computer Music Conference*, Göteborg, 2002.
- [5] C. Cadoz: "The Physical Model as Metaphor for Musical Creation: pico.TERA, a Piece Entirely Generated by a Physical Model", *Proceedings of the International Computer Music Conference*, Göteborg, pp. 305-12, 2002.
- [6] C. Chafe: "Simulating performance on a bowed instrument", in M. V. Mathews, J. R. Pierce: *Current directions in computer music research*, MIT, Cambridge (MS), pp. 185-98, 1989.
- [7] C. Chafe: "Case studies of physical models in music composition", *Proceedings of the 18th International Congress on Acoustics*, Kyoto, 2004.
- [8] J.-M. Couturier: "A scanned synthesis virtual instrument", *Proceedings of the 2002 conference on New instruments for musical expression*, Dublin, 2002.
- [9] P. Crabbe: *PMSC library (physical models)*, 2007.
- [10] S. Freud: "Das Unheimliche", *Imago*, No. 5, 1919. Tr. it. C. Balducci: "Il perturbante", in *Sigmund Freud. Opere 1905/1921*, Newton Compton, Roma, pp. 1049-70, 1992.

⁵ Si è fatto riferimento in particolare all'analisi della figura di Nathanael nel *Sandmann* di Hoffmann contenuta nel saggio *Das Unheimliche* del 1919 [10].

⁶ Si tratta in particolare di *Désintégrations* di Tristan Murail (1982) e *Les Chants de l'amour* di Gérard Grisey (1985), per le cui analisi rimando, rispettivamente, a [12] e [1]. In merito alla logica delle corrispondenze tra elettroacustica e pianoforte vi sono stati comunque riferimenti anteriori, fra cui *Mantra* di Karlheinz Stockhausen (1970), cfr. [16].

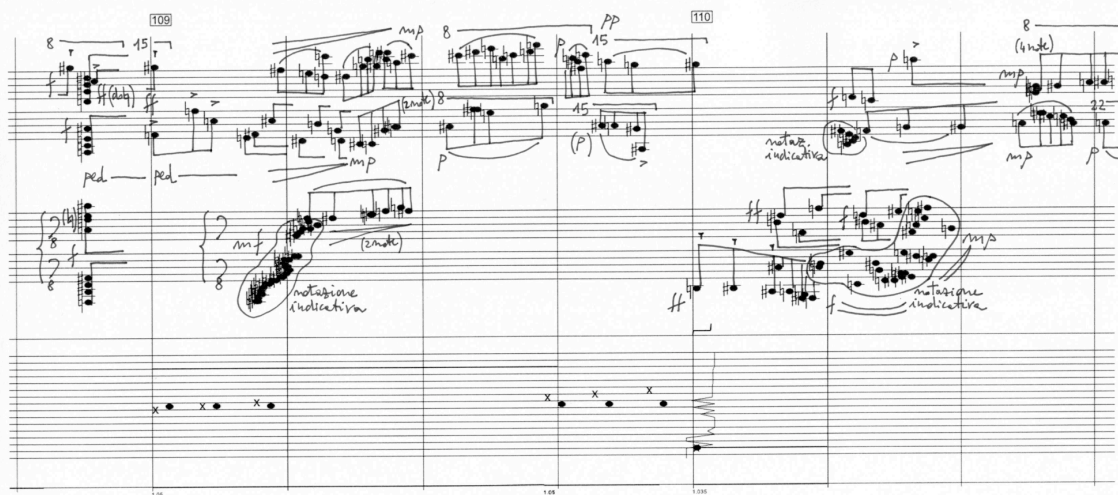
- [11] C. Henry: "pmpd. Physical Modelling for pd", *Proceedings of the first Sound and music computing conference*, Paris, 2004.
- [12] G. Klauer: *Il percorso creativo di Tristan Murail tra mimesi uditive e automatismi tecnici. 1967-1984*, tesi di laurea, Università di Pavia, 2006.
- [13] J. Kojs, S. Serafin, C. Chafe: "Cyberinstruments via physical modeling synthesis: Compositional applications", *Leonardo Music Journal*, No. 17, pp. 61-6, 2007.
- [14] M. Lupone: "Spazio, arco e metallo. Una ricerca musicale e scientifica", *Quaderni della Civica scuola di musica*, No. 26, pp 80-9, 1999.
- [15] M. Pearson: "TAO: A physical modeling system and related issues", *Organised Sound*, Vol. 1, No. 1, pp. 43-50, 1996.
- [16] K. Stockhausen: "Mantra, für 2 Pianisten", in *Texte zur Musik 1970-1977*, Vol. 4, Du Mont, Köln, pp. 153-66, 1978.
- [17] B. Verplank, M. V. Mathews, R. Shaw: "Scanned synthesis", *Proceedings of the 2000 International Computer Music Conference*, Berlin, pp. 368-71, 2000.
- [18] G. E. Winkler: "The realtime-SCORE. A missing-link in computer-music performance", *Proceedings of the first Sound and music computing conference*, Paris, pp. 9-14, 2004.
- [19] G. E. Winkler: "Hybrid II (NetWorks) or At the Edge of Musical Self-Organization", in C.-S. Mahnkopf, F. Cox, W. Schurig: *Electronics in New Music*, Wolke, Hofheim, pp. 236-49, 2006.

Handwritten musical score for two pianos, pages 36 and 37. The score is written on grand staves with various musical notations including notes, rests, and dynamic markings. Handwritten annotations include "non legato", "mp", "P", "senza pedale", and "(3 note (uic))". A tempo marking "Moderato 100-110 (200-2) 100" is visible at the bottom left. The page number 36 is circled at the bottom center.



region (H1 freq.)

37



mean frequency 25.5 (25.7 Hz)

0	10281.9	107.840
1	10281.9	107.840
2	10281.9	107.840
3	10281.9	107.840
4	10281.9	107.840
5	10281.9	107.840
6	10281.9	107.840
7	10281.9	107.840
8	10281.9	107.840
9	10281.9	107.840
10	10281.9	107.840
11	10281.9	107.840
12	10281.9	107.840
13	10281.9	107.840
14	10281.9	107.840
15	10281.9	107.840
16	10281.9	107.840
17	10281.9	107.840
18	10281.9	107.840
19	10281.9	107.840
20	10281.9	107.840
21	10281.9	107.840
22	10281.9	107.840
23	10281.9	107.840
24	10281.9	107.840
25	10281.9	107.840
26	10281.9	107.840
27	10281.9	107.840
28	10281.9	107.840
29	10281.9	107.840
30	10281.9	107.840
31	10281.9	107.840

31

Handwritten musical score for page 92. The score is written on a grand staff with a treble and bass clef. It features various musical notations including notes, rests, and dynamic markings such as *pp*, *mf*, and *mp*. There are also handwritten annotations like "motar. indicativa" and "3 note". The score is divided into measures, with some measures containing multiple notes. The bottom of the page shows a series of notes on a single staff, likely a transcription or a specific part of the score.

92

Handwritten musical score for page 93. The score is written on a grand staff with a treble and bass clef. It features various musical notations including notes, rests, and dynamic markings such as *pp*, *mf*, and *mp*. There are also handwritten annotations like "motar. indicativa" and "3 note". The score is divided into measures, with some measures containing multiple notes. The bottom of the page shows a series of notes on a single staff, likely a transcription or a specific part of the score.

93

LA LIBRERIA OPENMUSIC *om4Csound*

INTRODUZIONE E PROGETTO DI DOCUMENTAZIONE

Mauro Lanza

IRCAM
Centre Georges Pompidou
mauro.lanza@ircam.fr

Gianluca Verlingieri

Scuola di Musica e Nuove Tecnologie
Conservatorio “Ghedini” di Cuneo
verlingieri@libero.it

Nicola Biagioni

Scuola di Musica e Nuove Tecnologie
Conservatorio “Ghedini” di Cuneo
nicola.biagioni@gmail.com

ABSTRACT

Programmata e sviluppata da Mauro Lanza a partire dal 2008, la libreria *om4Csound* è concepita per integrare la versatilità e le potenzialità virtualmente illimitate del sintetizzatore *Csound* con l'interfaccia grafica “ad oggetti” del software di composizione assistita *OpenMusic (OM)*¹. Al di là dei vantaggi puramente ergonomici, la libreria *om4Csound* offre la possibilità che parametri relativi ai files *score* di *Csound*, ma anche la costruzione stessa delle orchestre, possano essere parte o risultato di processi compositivi e analitici sviluppati grazie alle potenzialità di OM e dalle sue librerie.

Dalla primavera del 2010 è in corso presso il Dipartimento Nuove Tecnologie e Linguaggi Musicali del Conservatorio “G. F. Ghedini” di Cuneo un progetto didattico a cura di Nicola Biagioni e di Gianluca Verlingieri, il cui obiettivo è la diffusione della libreria *om4Csound* col corredo di un apparato di documentazione e *tutorials* contenenti alcune esemplificazioni pratiche delle possibilità di utilizzo della stessa.

Parole Chiave

Composizione assistita all'elaboratore, composizione algoritmica, *OpenMusic*, *Csound*, analisi sintesi ed elaborazione del suono.

1. INTRODUZIONE

L'idea alla base della libreria *om4Csound* è, semplicemente, quella di integrare le potenzialità di *Csound* all'interno di OM.

Altre librerie sono state sviluppate in passato con

Copyright: © 2010 Lanza et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

intenti più o meno simili: la prima in ordine cronologico è quella programmata da Karim Haddad, Mikhail Malt e Laurent Pottier nel 1999 e denominata *Om2Csound* [1]. Essa è a sua volta il porting della libreria *Csound/Edit-sco* scritta dagli stessi Pottier e Malt, e il suo utilizzo è prevalentemente finalizzato alla compilazione di file *score* da utilizzare in *Csound* in un secondo tempo, senza cioè che il lancio del motore di sintesi avvenga dall'interno di OM, come accade invece con *om4Csound*.

Al 2000 risale la versione 1.0 di *OM-Chroma*, libreria sviluppata da Carlos Agon e dal compositore Marco Stroppa [2] sul modello dell'ambiente di lavoro *Chroma* (concepito dallo stesso Stroppa per la sintesi del suono e la composizione assistita [3]). *OM-Chroma* è intimamente legata al pensiero compositivo dell'autore e finalizzata ad un controllo di alto livello, in larga misura indipendente dal motore di sintesi impiegato.

om4Csound, più vicina come filosofia a *Om2Csound* che a *OM-Chroma*, ha l'ambizione di offrire un sistema aperto in cui l'integrazione del software di sintesi nel software di controllo risulti pressoché completa. A tal fine, alla quasi totalità degli opcodes di *Csound* corrispondono altrettanti equivalenti in OM, come appare evidente da una prima overview dei contenuti della libreria, sfogliandone i menù a tendina (v. figura 1).

Gli opcodes diventano in OM funzioni generiche [4], capaci cioè di comportamenti diversi in relazione ai loro inputs. Queste funzioni generiche sono affiancate da un numero ridotto di funzioni di “servizio” (*Lib Functions*, perlopiù relative alla combinazione di strumento e partitura e alla “chiamata” di *Csound* attraverso il terminale).

Meno immediatamente evidente è il fatto che semplici operatori di base presenti in OM vedano allargato il loro raggio d'azione e diventino capaci di interagire e modificare opcodes *Csound* all'interno di uno strumento, come evidenziato nel paragrafo successivo. Tutte le funzioni di *om4Csound*, ad eccezione di *csd->synth*², generano classi OM (*cs_opcode*, *cs_pfield*, *cs_gen*). I collegamenti tra funzioni generano un incapsulamento di classi il cui contenitore ultimo è la classe *csd*, unione di strumento e relativa partitura. Le classi che si creano

¹ *OpenMusic* è un marchio registrato dell'IRCAM.

² Cfr. paragrafo 2.1.

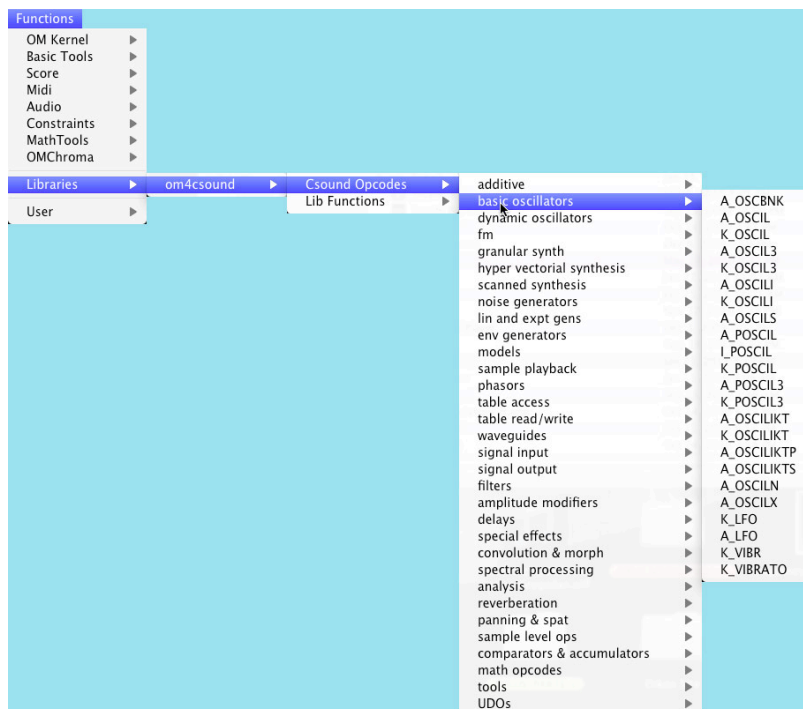


Figura 1. Panoramica su alcune delle molteplici funzioni della libreria *om4Csound*, corrispondenti ad altrettanti opcodes di *Csound*.

all'uscita degli opcodes possono essere mono-, bi- o n -dimensionali (istanzianti cioè n variabili). Se si utilizza un algoritmo per la spazializzazione su più canali (ad esempio l'opcode *vbap*), l'output sarà una classe n -dimensionale, dove n corrisponderà al numero dei canali impiegati. Il fatto che un opcode accetti in ingresso anche liste, permette inoltre di semplificare alcune routines: utilizzando ad esempio una lista di n frequenze per un opcode *a-oscil*, quest'ultimo genererà un numero n di oscillatori con frequenze corrispondenti agli elementi della lista.

2. ESEMPI

La libreria *om4Csound* non pone limiti teorici alla complessità delle patches implementabili; tuttavia, per introdurre con chiarezza i principi base del suo funzionamento, esaminiamo nel dettaglio due algoritmi volutamente semplici che la utilizzano.

2.1 Simple Oscillator

Il caso più elementare (v. il diagramma di flusso³ in figura 2) è quello di un oscillatore semplice (*a_oscil*), la cui forma d'onda può essere definita tramite la funzione *table*⁴, selezionando una *gen routine* in ingresso (di

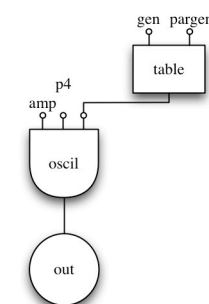


Figura 2. Esempio di diagramma a blocchi di un oscillatore semplice in Csound.

default la n . 10, generante forme d'onda di sinusoidi semplici).

La *gen* da utilizzare può essere selezionata rapidamente da un menù a tendina che elenca e descrive le opzioni disponibili (v. figura 3); vi si accede cliccando sull'inlet a sinistra della funzione *table*.

L'integrazione con l'ambiente OM (v. figura 4) rende più comodo variare anche gli altri parametri dell'oscillatore, come ampiezza e frequenza. Ad esempio, il valore di quest'ultima può essere impostato sotto forma di altezza notata su pentagramma tramite le *score-classes* di OM (*note*, *chord*, *chord-seq*, *voice* ecc.): nella parte superiore della figura 7 possiamo osservare come il valore iniziale di frequenza di 440 Hz sia stato modificato in un *Do* di ca. 261,63 Hz, inserito direttamente attraverso l'editor della classe *note*. In generale, la gestione dei parametri degli opcodes può

³ La simbologia adottata per i diagrammi di flusso di Csound è quella descritta da Bianchini, Cipriani [5].

⁴ La funzione *table* non corrisponde all'omonimo opcode Csound, bensì ad una funzione per generare una *wavetable*.

GEN01	Transfers data from a soundfile into a function table
GEN02	Transfers data from immediate pfields into a function table
GEN03	Generates a stored function table by evaluating a polynomial
GEN04	Generates a normalizing function
GEN05	Constructs functions from segments of exponential curves
GEN06	Generates a function comprised of segments of cubic polynomials
GEN07	Constructs functions from segments of straight lines
GEN08	Generate a piecewise cubic spline curve
GEN09	Generate composite waveforms made up of weighted sums of simple sinusoids
✓ GEN10	Generate composite waveforms made up of weighted sums of simple sinusoids
GEN11	Generates an additive set of cosine partials
GEN12	Generates the log of a modified Bessel function of the second kind
GEN13	Stores a polynomial whose coeffs derive from the Chebyshev polynomials of the 1st kind
GEN14	Stores a polynomial whose coeffs derive from Chebyshevs of the 2nd kind
GEN15	Creates two tables of stored polynomial functions
GEN16	Creates a table from a starting value to an ending value
GEN17	Creates a step function from given x-y pairs
GEN18	Writes composite waveforms made up of pre-existing waveforms
GEN19	Generate composite waveforms made up of weighted sums of simple sinusoids
GEN20	Generates functions of different windows
GEN21	Generates tables of different random distributions
GEN23	Reads numeric values from a text file
GEN24	Reads numeric values from another allocated function-table and rescales them
GEN25	Construct functions from segments of exponential curves in breakpoint fashion
GEN27	Construct functions from segments of straight lines in breakpoint fashion
GEN28	Reads a text file which contains a time-tagged trajectory
GEN30	Generates harmonic partials by analyzing an existing table
GEN31	Mixes any waveform specified in an existing table
GEN32	Mixes any waveform, resampled with either FFT or linear interpolation
GEN33	Generate composite waveforms by mixing simple sinusoids
GEN34	Generate composite waveforms by mixing simple sinusoids
GEN40	Generates a random distribution using a distribution histogram
GEN41	Generates a random list of numerical pairs
GEN42	Generates a random distribution of discrete ranges of values
GEN43	Loads a PVOCEX file containing a PV analysis
GEN51	This subroutine fills a table with a fully customized micro-tuning scale
GEN52	Creates an interleaved multichannel table from the specified source tables

Figura 3. Il menù a scomparsa dell'oggetto *table*, in cui si può rapidamente selezionare la gen routine desiderata.

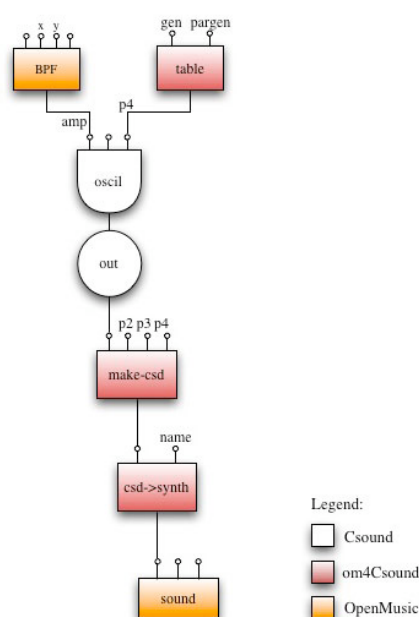


Figura 4. La sinergia fra funzioni e/o classi native di OM (in arancio), funzioni corrispondenti ad opcodes di Csound (in bianco) e funzioni “di servizio” della libreria *om4Csound* (in rosso).

avvenire tramite *p-campi*, che in *om4Csound* corrispondono agli inlets della funzione *make-csd* (v. figura 5). Quest’ultima seleziona automaticamente il

parametro in ingresso appropriato per ciascun p-campo: ad esempio, collegando l’outlet *self* di una classe *voice* a p2, *make-csd* ne estrapolerà gli onsets, mentre collegandolo a p3, ne ricaverà le durate. Gli inlets di *make-csd* si comportano quindi diversamente a seconda del tipo di dato in ingresso.

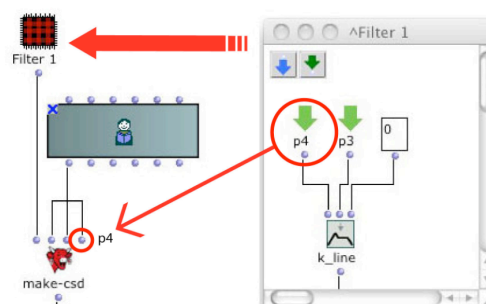


Figura 5. La gestione dei p-campi della funzione *make-csd*.

Definito un *output* per lo strumento creato (tramite le apposite funzioni *cs-out* o *easy-out*), in *om4Csound* la compilazione dei relativi files *.orc* e *.sco* avviene automaticamente, così come la generazione dei nomi delle variabili, proprio tramite la funzione *make-csd*⁵.

⁵ La dicitura *csd* nella denominazione della funzione *make-csd* non si riferisce alla corrente convenzione di Csound di creare un unico file con estensione *.csd* anziché due file separati con estensione *.orc* e *.sco*. Essa si riferisce al nome della classe OM creata (*csd*), unione di strumento e relativa partitura.

Orchestra e score così realizzati sono facilmente visualizzabili in formato di testo tramite la funzione *open-textfiles* (v. figura 6), la quale lancia l'editor di testo predefinito del sistema operativo in uso (*TextEdit* nel caso di *Mac OS X*) e apre i files *.orc* e *.sco* creati.

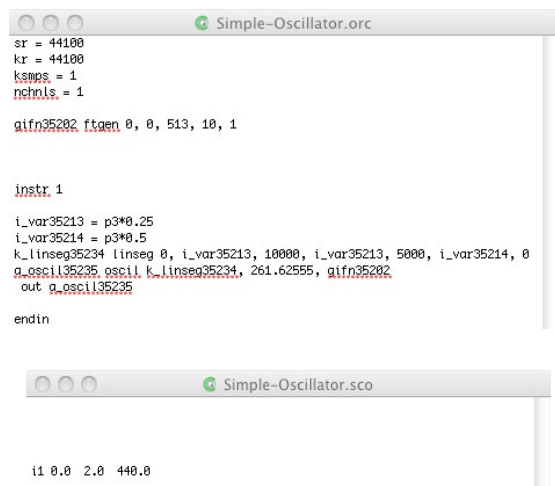


Figura 6. La visualizzazione delle linee di testo dei files *.orc* e *.sco* automaticamente creati da *om4Csound* è comodamente richiamabile dall'interno di OM, valutando la funzione *open-textfiles* della libreria.

A questo punto la funzione *csd->-synth* si occupa di eseguire il rendering, effettuando la chiamata del motore Csound tramite terminale. È quindi possibile ascoltare il risultato finale rimanendo all'interno di OM, attraverso l'editor della classe *sound*. La figura n. 7 illustra la patch di OM per l'algoritmo descritto.

2.2 Two Filters

Il secondo esempio mostra due moduli che generano rumore bianco filtrato i cui parametri di ampiezza, frequenza e durata sono determinati dalla classe *voice* di OM. Nel primo modulo (v. figure 8 e 11) abbiamo un involuppo di ampiezza del rumore bianco generato tramite l'oggetto *BPF* (break-point-function) di OM, e un involuppo sulla frequenza di taglio del filtro, generato utilizzando l'opcode *line* di Csound. Il segnale filtrato viene successivamente bilanciato tramite l'opcode *balance* (che permette di modularne il livello in base al valore di un altro segnale) e inviato alla funzione *cs_out*.

Il secondo modulo (v. figure 9 e 12) pur essendo molto simile al primo, presenta alcune differenze: il filtro è a banda passante anziché passa basso e l'involuppo è applicato alla banda passante. L'uscita dell'opcode *balance* viene modulata in ampiezza da un opcode *lfo* (low frequency oscillator) la cui frequenza è determinata da un p-campo.

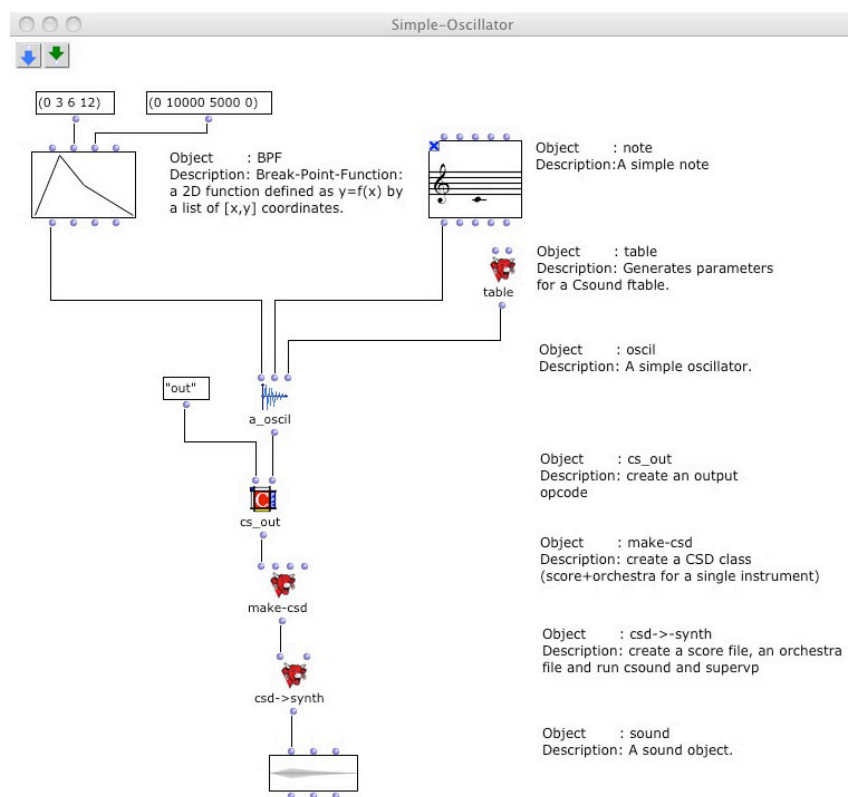


Figura 7. La patch di OM per l'oscillatore semplice descritto nell'esempio 2.1.

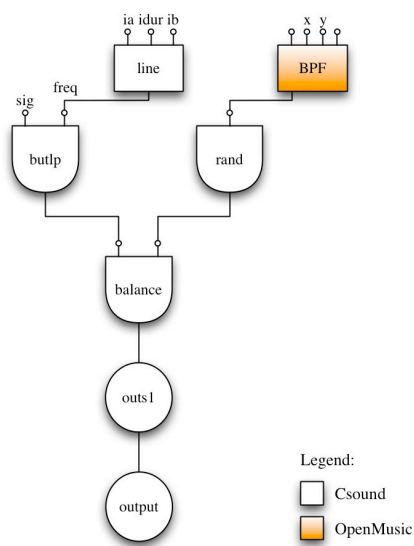


Figura 8. Diagramma a blocchi del primo modulo dell'esempio 2.2.

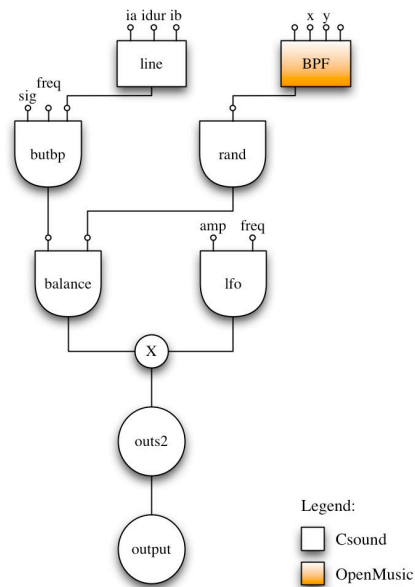


Figura 9. Diagramma a blocchi del secondo modulo dell'esempio 2.2.

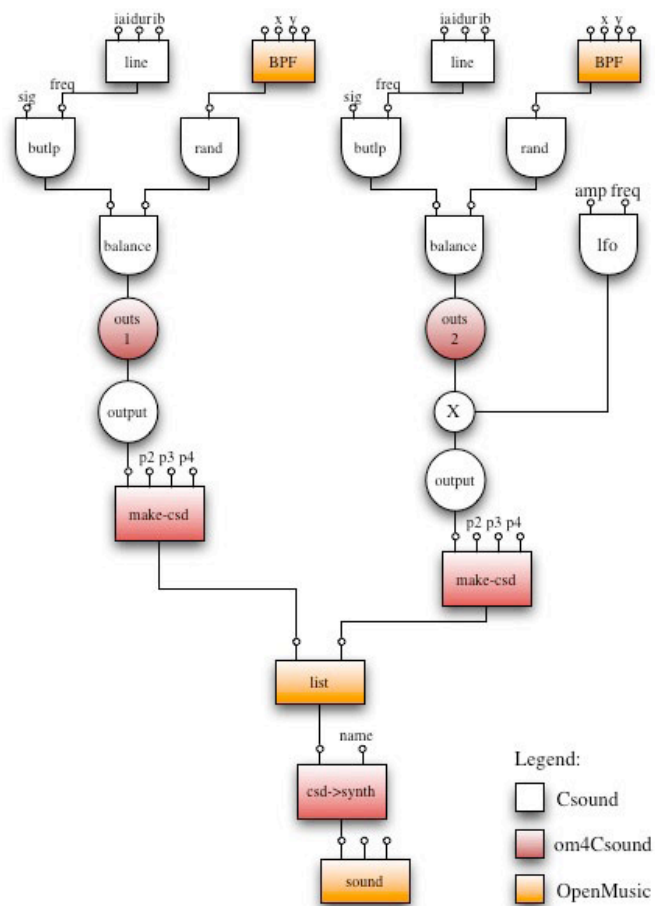


Figura 10. Diagramma a blocchi complessivo dell'esempio 2.2.

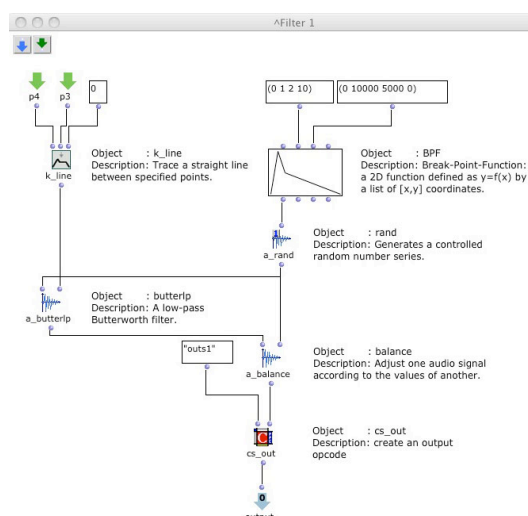


Figura 11. La sub-patch *Filter 1* dell'esempio 2.2.

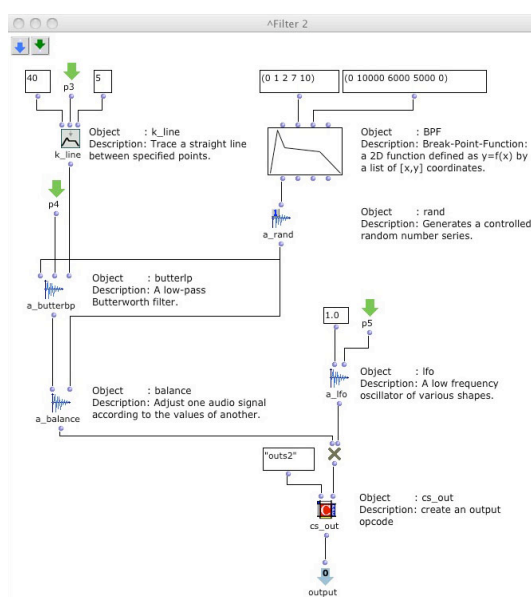


Figura 12. La sub-patch *Filter 2* dell'esempio 2.2.

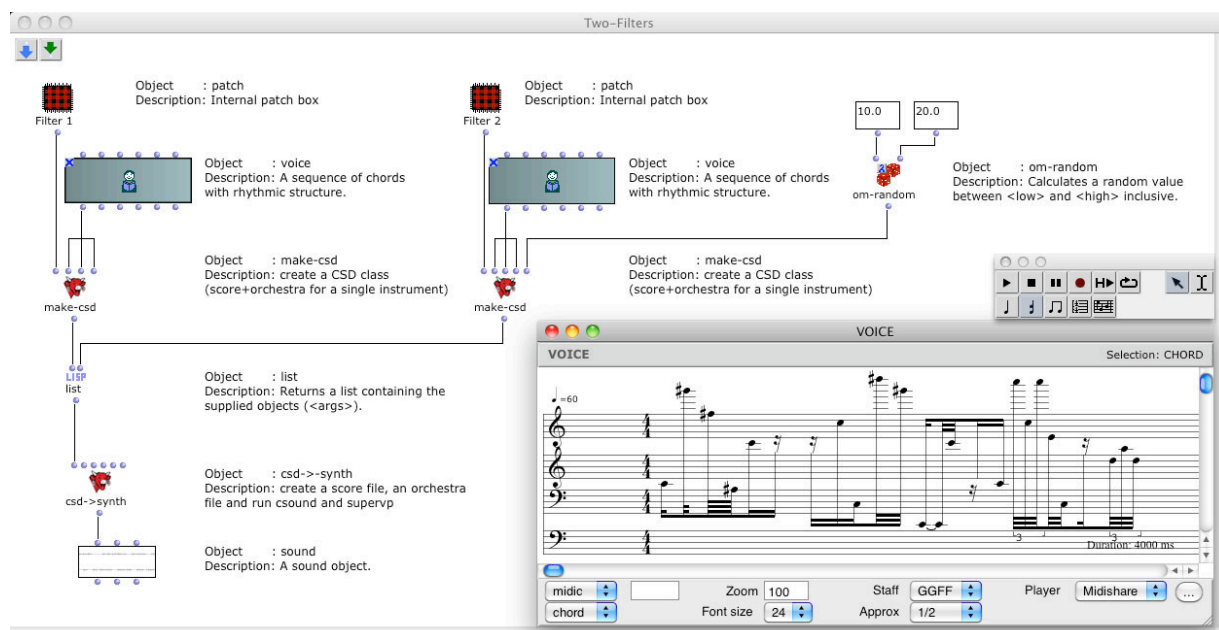


Figure 13. La patch principale dell'esempio 2.2, con in evidenza la finestra dello *score editor* di una delle due classi *voice* di OM impiegate.

Nella patch principale (v. figure 10 e 13) troviamo quindi le due sub-patches relative ai moduli appena descritti (denominate *Filter 1* e *Filter 2*), le due classi *voice* che controllano i parametri di ampiezza frequenza e durata, la funzione *om-random* che controlla la frequenza della modulazione d'ampiezza del secondo filtro (in un range variabile dai 10,0 ai 20,0 Hz), e le funzioni *make-csd*, che creano una classe *csd* per ogni strumento (in questo caso una *per Filter 1* e una *per*

Filter 2). Le classi *csd* così create vengono raccolte in una lista tramite la funzione *list*. A seguire, la funzione *csd->synth* effettua il rendering del processo, il cui risultato sarà la classe *sound*. In figura 14 è visualizzato il testo dell'orchestra e della score relative all'esempio 2.2.

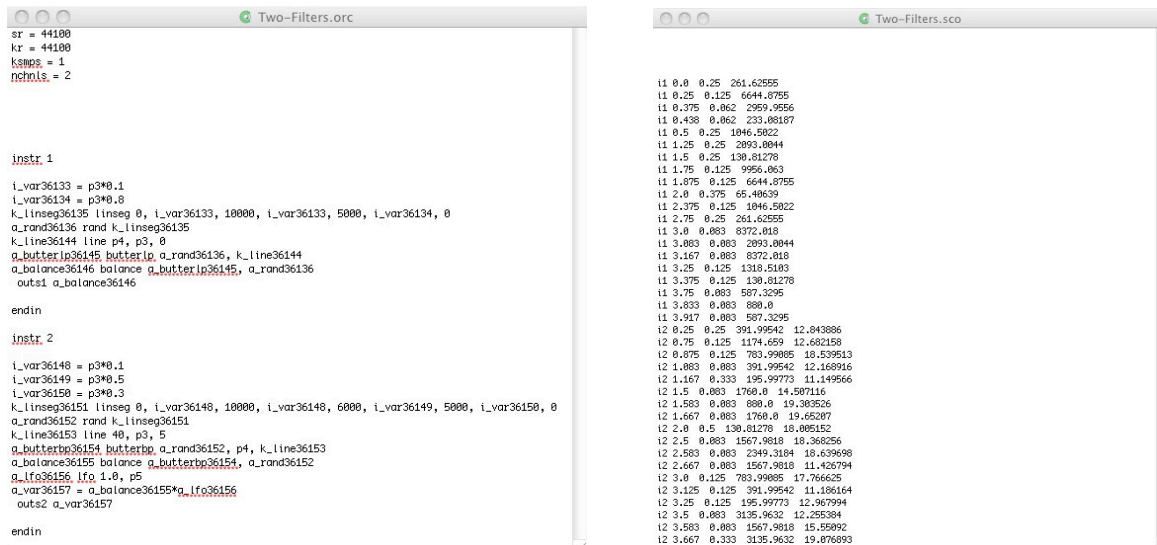


Figura 14. La visualizzazione delle linee di testo dei files .orc e .sco dell'esempio 2.2 tramite la funzione *open-textfiles*.

3. PROGETTO DI DOCUMENTAZIONE E TUTORIALS

Il progetto di documentazione della libreria *om4Csound* non si pone come obiettivo la redazione di un manuale di riferimento delle funzioni relative agli opcodes disponibili – scopo che può generalmente assolvere la relativa documentazione di Csound [6], peraltro richiamabile in *om4Csound* selezionando una funzione relativa a un opcode e premendo il tasto *d* (documentation) – ma intende piuttosto fornire all'utente della libreria, al quale è richiesta una conoscenza preventiva sia di OM che di Csound, una serie di tutorials contenenti applicazioni pratiche di alcune funzioni disponibili, abbinate o meno a strutture di controllo gestite da OM.

Le *patches* saranno accompagnate da una documentazione in formato *pdf* e *html* consistente in due parti: un'introduzione contenente le informazioni necessarie per l'installazione e la configurazione della libreria (*getting started*), e una seconda parte che illustrerà tramite testo e *screenshots* i processi e gli algoritmi implementati di volta in volta nelle singole *patches*.

La lingua di riferimento per i commenti all'interno delle *patches* sarà l'inglese, mentre la documentazione testuale verrà redatta, oltre che in inglese, anche in italiano.

4. CONCLUSIONI

Il presente articolo ha introdotto e brevemente esemplificato il funzionamento della libreria *om4Csound* di OM, attualmente disponibile per il sistema operativo *Mac OS X*. Oltre all'attività di documentazione e

tutorials in corso, gli sviluppi futuri del progetto includono possibili versioni per altri sistemi operativi (*Windows*, *Linux*) e un eventuale porting della libreria in ambiente open source *PWGL*.

5. RIFERIMENTI

- [1] K. Haddad: *OpenMusic Om2Csound, Bibliothèque de modules de génération de scores pour Csound, version 1*, IRCAM Centre Georges Pompidou, Paris 1999.
- [2] M. Stroppa: *OpenMusic omChroma, Paradigms for the high level musical control of sonic processes using omChroma + omChroma 1.0 Manual & Tutorial*, IRCAM Centre Georges Pompidou, Paris 2000.
- [3] G. Albert, *Musica e informatica: tracce di silicio nelle forme del suono*, relazione tenuta presso il XI Colloquio di Musicologia del Saggiatore Musicale (Bologna, DAMS, 23-25 Novembre 2007).
- [4] C. Agon, *OpenMusic: un langage visuel pour la composition musicale assistée par ordinateur*. these de doctorat, Université Paris 6, Paris 1998.
- [5] R. Bianchini, A. Cipriani, *Il Suono Virtuale, Sintesi ed Elaborazione del Suono, Teoria e Pratica con Csound*, ConTempo, 1998.
- [6] B. Vercoe, *The Canonical Csound Reference Manual*, Version 5.13, MIT Media Lab, <http://csounds.com/manual/html/>

COMPOSIZIONE ELETTROACUSTICA E TECNICA WAVESET

Giorgio Bianchi - Pier Daniel Cornacchia

giorgio.bianchi@email.it - cornwil@libero.it

Conservatorio di Musica di Como

ABSTRACT

I processi di elaborazione del segnale praticati da Trevor Wishart sotto la denominazione di tecniche waveset sono stati reimplementati sotto una prospettiva creativa originale. In seguito all'attenta analisi degli scritti e delle opere di Wishart e dopo il confronto con realizzazioni algoritmiche più recenti, è stata effettuata in SuperCollider una ulteriore reimplementazione di alcune fra tali tecniche. L'ambiente di programmazione utilizzato ha consentito di approntare interessanti varianti e prospettare molteplici ed efficaci soluzioni alternative di gestione e manipolazione dei dati rispetto al modello. La reimplementazione è stata quindi utilizzata in due composizioni elettroacustiche originali.

1. TREVOR WISHART E LA TECNICA WAVESET

La manipolazione di un segnale a partire dai waveset [Nota a piè pagina: Il waveset consiste nell'insieme di campioni compresi fra uno zero-crossing e il successivo zero-crossing in cui c'è un passaggio da valori dello stesso segno] permette un approccio di design sonoro eccentrico, che risolve in maniera artistica e tendenzialmente disinteressata rispetto alle implicazioni scientifiche, le problematiche legate alla dinamica e alla natura spettrale dei suoni.

Pur attuandosi nel dominio del tempo, la tecnica waveset non si basa sull'estrapolazione di frammenti di segnale né sull'imposizione di ulteriori involucri. Limitandosi ad agire sul contenuto dei waveset e sulla loro sequenzialità a livello micro-temporale, giunge comunque a sensibili alterazioni del profilo spettrale e dinamico percepito.

Wishart implementò le proprie originali tecniche nell'ambito del Composer's Desktop Project [2], community indipendente che ha avuto lo scopo di sviluppare un corpus di algoritmi di sound processing finalizzato alla composizione musicale elettroacustica.

L'impiego di tali implementazioni viene attestata da Wishart, al riguardo in particolare della propria opera dei primi anni Novanta, in *Audible Design* [7]. La peculiarità di questo testo consiste nell'affrontare le problematiche dell'elaborazione del segnale, particolarmente quello digitale, nella prospettiva di un artigianato compositivo elettroacustico, dove l'intenzione di preservare l'originalità di pensiero musicale si manifesta nell'evitare di adoperare il linguaggio formale della disciplina ingegneristica informati-

ca e nel proporre eventualmente gli algoritmi in termini prettamente figurativi (Appendice 2 al volume).

Il brano che si è rivelato maggiormente interessante per gli scopi del nostro contributo, sia per quanto attiene l'uso di queste tecniche sia per la loro formalizzazione nella tecnica di composizione musicale, è *Tongues of Fire* [5]. Realizzato a partire dal materiale di un frammento di montato di Voce Estesa, attraverso svariati stadi di manipolazione e distorsione sonora, in questo brano si crea una ben ricercata costruzione pseudo-narrativa.

A questo proposito, gli esiti delle tecniche esaminate sono stati interpretati alla luce dei contenuti di "On Sonic Art" [6], manifesto del paradigma wishartiano dove sono descritti nella maniera più minuziosa scopi e motivazioni della propria prassi compositiva degli anni Settanta e Ottanta. Fra gli elementi fondanti del suo paradigma si colloca infatti l'idea di una narratività simbolica, scaturita dalle fonti sonore stesse, siano esse reali, virtuali o immaginarie.

A dispetto del modello schaefferiano, il riconoscimento della fonte, ovvero l'individuazione attiva da parte dell'ascoltatore di una causa vera o presunta e del contesto ad essa relativo, nell'estetica della Sonic Art wishartiana diviene di fondamentale importanza a livello formale. La composizione elettroacustica di questo autore è condotta infatti secondo una logica di trasformazioni che, attraverso la percezione di identità sonore e indici di contestualizzazione, dovrebbe condurre all'immaginazione di un archetipo narrativo o addirittura a un nucleo tematico di carattere concettuale.

Wishart ha dichiarato ripetutamente [1, 8] come il proprio approccio alla composizione elettroacustica sia passata nell'ambito digitale dall'uso delle tecniche di Phase Vocoder (Vox 5) alla predilezione per le tecniche attuate nel dominio del tempo, di cui fanno parte a titolo principale le manipolazioni waveset: questi cambi di mezzo, come in precedenza il cambio di dominio dall'analogico al digitale, non hanno dunque stravolto le istanze originali di questo autore. Il rinnovamento delle tecnologie e delle tecniche ha portato anzi al loro consolidamento, seppure in un'altra versione auditiva.

2. IMPLEMENTAZIONI DEGLI ALGORITMI IN SUPERCOLLIDER

Per gli scopi di creazione originale menzionati sopra e di cui verranno successivamente discussi due prodotti, si è voluto basare la reimplementazione delle tecniche waveset attuata in ambiente SuperCollider [Nota a piè pagina: SuperCollider è un ambiente e linguaggio di programmazione per la sintesi sonora in tempo reale e per la compo-

sizione algoritmica. Si basa su un linguaggio interpretato ad oggetti che funziona come una rete client-server per la sintesi sonora real time. SuperCollider fu scritto da James McCartney per un periodo di diversi anni, ed è ora un progetto open source (GPL).] direttamente dai diagrammi e le descrizioni verbali fornite negli scritti di Wishart, specialmente nel testo *Audible Design*, verificandone il funzionamento confrontando gli esiti di prove svolte su materiale vocale specifico con similari sonorità riscontrate in *Tongues of Fire*, anche sulla scorta del commento fornitone dall'autore nel *Computer Music Journal*[5]. Ci si è confrontati comunque almeno con due implementazioni successive a quella originaria del CDP, entrambe peraltro sviluppate in SuperCollider, ovvero quella di De Campo[10], limitata alla waveset detection, e di Hochherz[9], la cui libreria di applicazioni è atta anche a riorganizzare in varie forme il segnale sonoro. Le implementazioni delle tecniche waveset sono state realizzate in SuperCollider in quanto linguaggio di programmazione versatile che, pur fortemente orientato a finalità di carattere musicale, consente un'amplissima versatilità nell'elaborazione dei dati disposti in matrici e vettori multidimensionali (nidificati). Preliminare a ogni elaborazione è la waveset analysis, avvenute come obiettivo la creazione di un vettore contenente le coordinate assolute di ogni secondo zero crossing, espressi in frazioni di campioni (subsamples/floats). La derivata dei valori contenuti in questo vettore è ovviamente la consistenza temporale dei singoli waveset. Gli algoritmi seguenti sfruttano questo vettore per ricampionare, sostituire, ridisegnare o riorganizzare temporalmente il contenuto (la forma d'onda) dei waveset, presi singolarmente o a gruppi, secondo specifici parametri.

3. DUE APPROCCI COMPOSITIVI A CONFRONTO

3.1 L'esperienza compositiva di Giorgio Bianchi

Dall'idea narrativa del "concepimento" del pensiero/creatività umano (da un dipinto di Alla Polozenko) è tratta la metafora della voce femminile, simbolo di fertilità ed evoluzione. Il lavoro si è svolto su questo materiale secondo un orientamento tecnicista finalizzato alla sperimentazione dei processi di elaborazione waveset. Su questa base sono però emersi ulteriori e sorprendenti spunti e collegamenti narrativi. La composizione, basata sul montaggio del suono processato, ha assunto un carattere fortemente aggressivo, specie per via di procedimenti come "averaging" e "substitution", che spingono facilmente ai limiti della dimensione sonora digitale; gli artefatti emersi sono spesso singolari e si è trovato utile lasciarsi ispirare dai risultati sonori di questo tipo di DSP. L'esito musicale di questa esperienza è interpretabile con la metafora del parto feroce.

3.2 L'esperienza compositiva di Pier Daniel Cornacchia

Obiettivo della composizione *Umanità* è il ricercare diversi approcci volti ad arricchire il significato espresso dalle parole: ciò è ottenuto con l'uso delle manipolazioni waveset sulla voce di un attore che recita. La composizio-

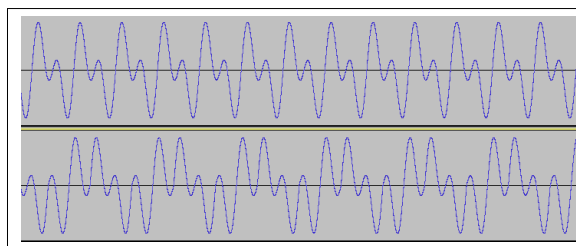


Figure 1. Esempio di Waveset Time-Stretching.

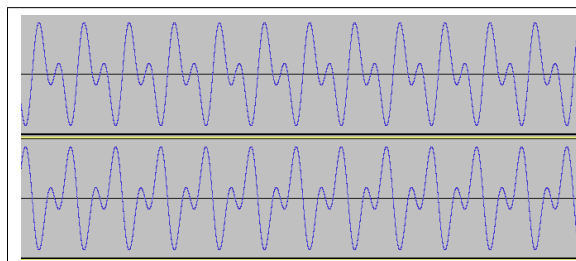


Figure 2. Esempio di Waveset Reverse.

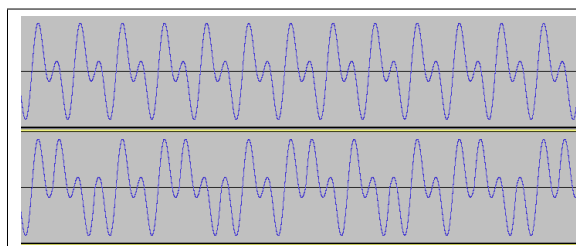


Figure 3. Esempio di Waveset Shuffling.

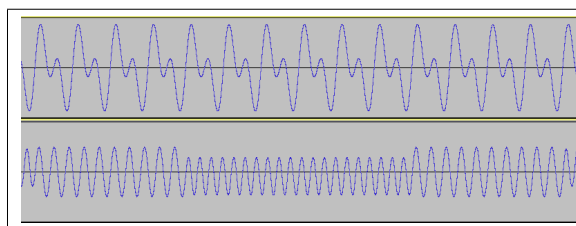


Figure 4. Esempio di Waveset Averaging.

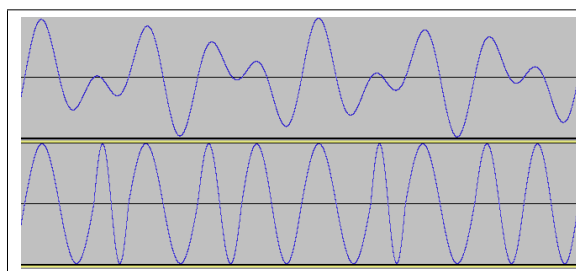


Figure 5. Esempio di Waveset Substitution.

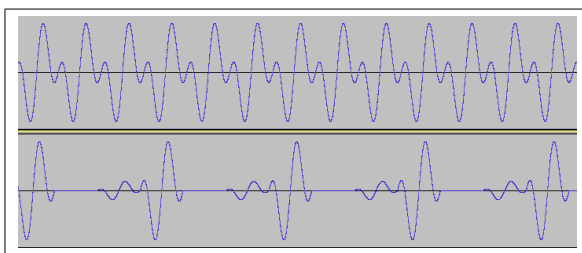


Figure 6. Esempio di Waveset Pattern Omission.

ne è costituita da cinque parti molto brevi, ognuna associata a una o poche parole, ripetute e sottoposte a diversi tipi manipolazioni, in particolare manipolazioni waveset. Alle elaborazioni di ciascuna parola è associato un mondo sonoro che può suggerire di volta in volta soundscape, stati cinetico-emotivi o immagini sonore, mentre le poche verbalizzazioni ancora intelligibili trasmutano suggestivamente la propria semantica. Il criterio compositivo è stato quindi di costruire un percorso auditivo che inviti a riflettere sulla natura umana e su come spesso gli individui si contrappongano alla stessa Umanità di cui fanno parte.

4. CONCLUSIONI

La composizione con i waveset si è rivelata un'esperienza ricca di spunti e certo non una tecnica fine a se stessa. I materiali sonori ottenibili sono di disparata natura, pur mantenendo una marcata direzionalità temporale simile al suono di origine: l'elaborazione waveset consiste nella sovversione dell'ordine microtemporale. La tecnica permette da un qualunque materiale di ottenere mondi sonori particolareggiati più o meno lontani dall'origine e pertanto efficaci per la composizione musicale. La tecnica si presta evidentemente alla poetica di Wishart, in quanto queste manipolazioni, i cui esiti sono relativamente controllabili, si prestano alla costruzione di una narrazione sonora caratterizzata da una originalità senza pari. E' da valutare una prossima implementazione che ottimizzi la velocità di calcolo, in modo da minimizzare la latenza tra processo di analisi del segnale in input e ricampionamento del segnale da passare in output, nella prospettiva di un'applicazioni in (quasi) real time che favorisca l'immediatezza dell'approccio auditivo, magari con l'associazione di una GUI dedicata al monitoraggio multidimensionale degli esiti dei processi di elaborazione.

5. REFERENCES

- [1] T Wishart - Computer Sound Transformation A personal perspective from the U.K. - <http://www.trevorwishart.co.uk/transformation.html> - 2000
- [2] <http://www.composersdesktop.com/>
- [3] S Emmerson - The language of electroacoustic music - Macmillan - 1986

- [4] J Young - Sound morphology and the articulation of structure in electroacoustic music- Organised Sound (2004), 9: 7-14 Cambridge University Press
- [5] T Wishart - Sonic Composition in Tongues of Fire - Computer Music Journal Summer 2000, Vol. 24, No. 2, Pages 22-30
- [6] T Wishart , S Emmerson - On sonic art - Harwood Academic Publishers - 1996
- [7] T Wishart - Audible Design, A plain and easy introduction to practical sound composition - Electronic Music Foundation - 1994
- [8] Matteo Milani, Federico Placidi - An interview with Trevor Wishart - U.S.O. Project, January 2009
- [9] O Hochherz - PList, a Waveset synthesis library and its usage in the composition *Draussen* - lac.linuxaudio.org/2008/download/papers/
- [10] DeCampo - A simple analysis class for Waveset Synthesis a la Trevor Wishart 2006

Ensemble methods for musical expressive intention recognition

Antonio Rodà, Ingrid Visentini, Lauro Snidaro, Gian Luca Foresti

University of Udine
antonio.roda@uniud.it

Sergio Canazza

University of Padova
canazza@dei.unipd.it

ABSTRACT

This paper analyses how automatic expressive musical intention recognition can be influenced by several parameters such as classifier type and sampling frequency. In contrast to recent literature where linear classifiers achieved slightly better than chance recognition rates, here we show promising preliminary results where high recognition percentages can be attained by ensembles of classifiers. Another novel contribution is the analysis of the influence of two parameters (window size and hop distance) upon classification performance. Experimentation has been carried on three musical excerpts each of which played with nine expressive intentions.

1. INTRODUCTION

The sharing of increasingly large digital audio libraries over the network demands sophisticated tools to enable users to easily find the requested content. The textual approach used by today's search engines has limitations in its application to audio files, because it allows only searching by metadata (i.e., title, author, genre, and so on), not by content. So if metadata, which are usually added manually, are incorrect or do not match with the content, the search can fail. Moreover, the user may not know exactly what document she is looking for, but might want to browse the audio library to search for a document that meets certain criteria: for example, a relaxing music or "something hard". In recent years, much progress has been made toward developing tools for content-based retrieval in audio documents (see [1] and [2] for a review). One of the most used approaches is to define a set of features that describe certain characteristics of sound and can be used to automatically classify the songs according to a determined list of categories.

This paper deals with features that are suitable to describe the expressive characteristics of a musical performance. The communication of expressive content by music can be studied at different levels, considering i) the expressive intentions of the performer (e.g., [3, 4]), ii) the listener's perceptual experience (see [5] for reviews), or iii) the composer's message (e.g. [6, 7]). Concerning the first aspect, Mion & De Poli [8] asked three musicians to

play several times a few short melodies, following different expressive intentions described by a set of affective and sensorial adjectives. A set of features, considered to be particularly representative of the expressive nuances of the performances, were extracted on the base of a frame size of 4 seconds. Results showed that, using the selected features, a linear classifier can recognize the expressive intentions of the songs, with an accuracy better than chance.

This paper aims to examine if and how several parameters, such as the classification function or the size of the observation window, influence the automatic recognition of expressive intentions. We have recorded some expressive performances and calculated a set of acoustic features, using observation windows of different lengths from 1.5 to 4 seconds and hop-size values from 0.125 to 0.5 seconds (Section 2). The features were used to train a set of classifiers combined into a strong classifier ensemble, whose results and performances are presented in Section 4.

2. AUDIO FEATURE EXTRACTION

A violinist was asked to play three musical excerpts several times, each one inspired by one of the expressive intentions described by the following adjectives: happy, sad, angry, calm, hard, soft, heavy, light, and normal. The adjectives were chosen among the most widely used in studies of music performance: four refer to the emotional domain and four to the sensorial one. The normal performance, i.e. a performance that lacks a specific expressive intention, was introduced as a term of comparison to better assess the changes induced by the other expressive intentions. The three musical excerpts were chosen to represent different musical genres: a piece belonging to the Western classical repertoire (the incipit of the Violin Sonata Op. 1 No. 12 by G. F. Haendel), a popular melody (*Twinkle Twinkle Little Star*), and a jazz standard (*I Got Rhythm* by G. Gershwin). The performances were captured by one microphone and the audio signal were recorded in monophonic digital form at 24 bits and 48000 Hz.

A set of acoustic features were calculated for each performance. The set was chosen among those features that in previous listening experiments [9] were found to be important for discriminating different emotions and were also used to classify the style [10] and the expressive content in musical performances [11] and [8]. See Table 1 for a formal description of the first ten features (for the others, an informal description is more convenient). The features are: a) *RMS* takes into account the global energy of the signal, computed as the root average of the square of the amplitude (root-mean-square); b) *Zerocross* consists in counting the

number of times the audio signal changes sign. It can be considered as a simple indicator of noisiness; c) *Centroid* is the first moment of the spectral amplitude. It is related with the impression of “brightness” of a sound [12], because a high centroid value means that the sound energy is concentrated at the higher frequencies; d) *Brightness* measures the amount of energy above the frequency of 1000 Hz. The result is expressed as a number between 0 and 1; e-g) *Spectral ratios* (SRs) over different frequency bands of the spectrum are other useful indications of the spectrum shape. The spectrum is divided in three regions: below 534 Hz (SRI), from 534 to 1805 Hz (SRm), and above 1805 Hz (SRh); h) *Rolloff* is the frequency such that the 85% of the total energy is contained below that frequency. It is related to the “brightness” of the sound; i) *Spectralflux* is the distance between the spectrum of each successive frame; l) *Lowenergy* is the percentage of frames showing less-than-average energy. It is an assessment of the temporal distribution of energy, in order to see if it remains constant throughout the signal, or if some frames are more contrastive than others; m) *Roughness* is calculated starting from the results of Plomp and Levelt [13], that proposed an estimation of the dissonance degree between two sinusoids, depending on the ratio of their frequency. The total roughness for a complex sound can be calculated by computing the peaks of the spectrum, and taking the average of all the dissonance between all possible pairs of peaks [14]; n) *Notes per second* Notes per second (NPS) is computed by dividing the number of onsets by the window duration. o) *Attack time* Attack time is computed as the time required to reach the RMS peak, starting from the onset instant. p) *Tempo* is the musical velocity of the performance, calculated in beat per minute. q) *Legato* is the ratio among the duration of a note and the inter-onset-interval. A value of 1 means that the note is tied to the next one; a value less than 1 means a *staccato*.

3. CLASSIFIER ENSEMBLE

Given a set of classifiers $\{h_1, h_2, \dots, h_M\}$ so that $h_m : X \rightarrow \Omega$ where X is a set of patterns and $\Omega = \{\omega_1, \dots, \omega_N\}$ is a set of labels, a classifier ensemble H is a combination of the h s obtained through a fusion rule. A pattern x is assigned to the most probable class ω if

$$P(\omega|H(x)) = \arg \max_k P(\omega_k|H(x)) \quad (1)$$

For the Bayes theorem, we have that

$$P(\omega|H(x)) = \frac{P(H(x)|\omega)P(\omega)}{P(x)} \quad (2)$$

where $P(H(x)|\omega)$ represents the joint probability distribution of the measurements extracted by the classifiers combined into H . In our case, a pattern x is assigned to the most probable class ω as per the vote by majority rule. The principle is to count the number of favourable outcomes given by the classifier ensemble, and assign the final decision to the label with the highest score.

Dietterich [15] gives some motivations why multiple classifiers systems may be better than a single classifier.

Table 1. List of the acoustic features. The signal x is blocked in M frames of N samples. Let be $x(f, n)$ the signal amplitude of the sample n at the frame f ; $X(f, k)$ the spectrum magnitude of the bin k at the frame f and $F(f, k)$ the center frequency of that bin; k_{f_i} the bin corresponding to the frequency f_i ; $I\{A\}$ the indicator function equal to 1 if A is true and 0 otherwise; $\text{sign}(x)$ a function equal to 1 if $x \geq 0$ and 0 otherwise; $\text{rms}(x(f))$ the *RMS* value over the frame f and $\text{rms}(x)$ the *RMS* value over the entire signal x .

<i>RMS</i>	$\sqrt{\frac{1}{N} \sum_{n=1}^N x(f, n)^2}, f = 1, \dots, M$
<i>Zerocross</i>	$\sum_{n=1}^{N-1} I\{\text{sign}(x(f, n)) \neq \text{sign}(x(f, n+1))\}, f = 1, \dots, M$
<i>Centroid</i>	$\frac{\sum_{k=1}^N F(f, k) X(f, k)}{\sum_{k=1}^N X(f, k)}, f = 1, \dots, M$
<i>Brightness</i>	$\frac{\sum_{k=k_{1000}+1}^N X(f, k)}{\sum_{k=1}^N X(f, k)}, f = 1, \dots, M$
<i>SRI</i>	$\frac{\sum_{k=1}^{k_{534}} X(f, k)}{\sum_{k=1}^N X(f, k)}, f = 1, \dots, M$
<i>SRm</i>	$\frac{\sum_{k=k_{534}+1}^{k_{1805}} X(f, k)}{\sum_{k=1}^N X(f, k)}, f = 1, \dots, M$
<i>SRh</i>	$\frac{\sum_{k=k_{1805}+1}^N X(f, k)}{\sum_{k=1}^N X(f, k)}, f = 1, \dots, M$
<i>Rolloff</i>	$f(k_{85}), \text{ where } k_{85} = \min(k_0) : \frac{\sum_{k=1}^{k_0} X(f, k)}{\sum_{k=1}^N X(f, k)} > 0.85, f = 1, \dots, M$
<i>Spectralflux</i>	$\sqrt{\sum_{k=1}^N [X(f+1, k) - X(f, k)]^2}, f = 1, \dots, M-1$
<i>Lowenergy</i>	$\frac{\sum_{f=1}^M I\{\text{rms}(x(f)) < \text{rms}(x)\}}{M}$

The first one is statistical: the ensemble may be not better than the single best classifier, but the risk of picking an “inadequate single classifier” while constructing the ensemble is reduced. The second reason is computational: a set of decision makers can fetch a solution working each one on a slice of the problem, in less time than a single individual. The last motivation refers to the ability of a set of classifiers to work on separate regions of the domain, or on its variations, yielding a fusion output that can be intended as the optimal classifier for that domain. Moreover, any parameter variation can be better satisfied under the constraint that many classifiers have to be tuned, while it is difficult or even impossible to adapt every time the parameters of a single classifier.

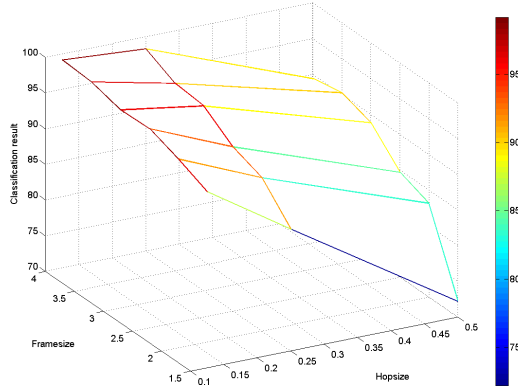
4. EXPERIMENTS

We computed the features in 18 different conditions, obtained by combining different values for the window length and the window hop-size. In particular, we used 6 values for the window length (1.5, 2, 2.5, 3, 3.5, and 4 s) and 3 values for the window hop-size (0.125, 0.25, 0.5 s). The minimum value of the window length has been chosen so that at least one musical event falls inside the window; the maximum values, instead, roughly corresponds to the size of the echoic memory. The values of the hop-size has been chosen so that no musical events are skipped moving from a window to the next one. In total, we collected a set of 13 audio features on about 20000 windows. See Table 1 for a formal description of the features.

As far as it regards the classifier ensemble, it comprised

Table 2. Performances of classifier ensemble. The maximum and the minimum are highlighted in bold.

hop-size frame-size	0.5			0.25			0.125		
	%	#Train	#Test	%	#Train	#Test	%	#Train	#Test
4	89.405	339	170	98.3631	672	336	99.1814	1466	733
3.5	90.22345	357	179	96.32765	708	354	98.9004	1545	773
3	88.82975	375	188	95.96775	744	372	97.7805	1622	811
2.5	84.6939	391	196	93.04125	776	388	97.95075	1699	849
2	83.1707	409	205	91.50245	812	406	96.51685	1779	890
1.5	72.19625	427	214	87.0283	848	424	94.66595	1856	928

**Figure 1.** Accuracy values when varying the frame-size and the hop-size parameters used to calculate the features.

a set of six weak classifiers, including a decision tree C4.5 (J48), two K-Nearest Neighbour algorithms (with $k = 2$ and $k = 4$ respectively), two Random Forests (with 3 and 15 trees respectively) and a Bayesian classifier. The combination rule, as stated in 3, is the vote-by-majority.

Varying the frame-size and the hop-size, from a minimum of 1.5 and 0.125 respectively, to a maximum of 4 and 0.5, the classifier ensemble clearly changed in performance. In Figure 1 a first overview of the performances is shown; the accuracy level rises when the framesize increases and the hop-size decreases. More in detail, as we can see in Table 2 the classification rate, calculated as the diagonal term of the confusion matrix out of the sum of the test samples, is maximum when considering a frame-size equal to 4 in correspondence to a hop-size of 0.125. Contrarily, it is minimum when considering the smallest frame-size value and the higher hop-size. #Train refers to the number of samples used to train the classifiers, while #Test are the number of samples employed for the test phase. A long frame-size can represent a smoothing factor, that filters the noise in the interval and reduces the probability of error occurrence. A short hop distance increases the number of available samples, reducing the variance of the classifier thus better generalizing the model.

The breakdown of the performances while training and testing the best and the worst combination of parameters

Table 3. Classifiers performances with frame-size = 4 and hop-size = 0.125. Fused output: **99.1814%**

Classifier	Training %	Testing %
j48	99.7271	97.8172
Bayes	91.2688	91.5416
RandomForest (3)	99.8636	98.3629
RandomForest (15)	99.9318	98.7722
KNN ($k = 2$)	91.2005	79.9454
KNN ($k = 4$)	87.2442	79.9454

Table 4. Classifiers performances with frame-size = 1.5 and hop-size = 0.5. Fused output: **72.196%**

Classifier	Training %	Testing %
j48	94.1452	67.2897
Bayes	88.9930	69.6262
RandomForest (3)	97.6581	66.8224
RandomForest (15)	100	74.7664
KNN ($k = 2$)	62.2951	29.4393
KNN ($k = 4$)	57.1429	24.2991

are presented in Table 3 and 4 respectively. In the first case, the fusion rule outperformed the best classifier in the set, while in the second case the fused output resulted slightly more inaccurate than the best classifier in the ensemble. However, in any case the fused outcome is better than the average classifier. A sample confusion matrix of classifier ensemble applied to samples obtained with frame-size = 4 and hop-size = 0.125 is presented in Table 5.

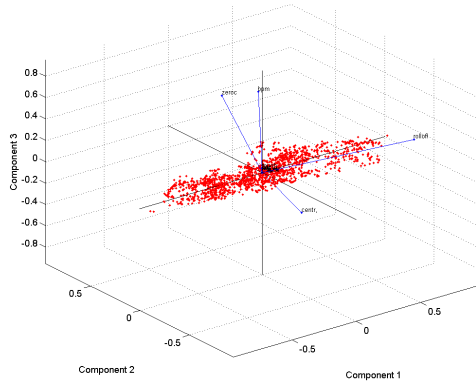
To understand which features contributed most to achieve the classification performance, we run the Principal Com-

Table 5. Confusion matrix of classifier ensemble applied to samples obtained with frame-size = 4 and hop-size = 0.125.

		Actual									
Predicted	52	0	0	0	0	0	0	0	0	0	0
	0	84	0	0	0	0	0	0	0	0	0
	0	0	141	0	0	0	0	0	0	0	0
	0	0	0	36	0	0	0	0	0	0	0
	0	0	0	0	54	0	0	0	0	0	0
	0	0	0	0	0	43	0	0	0	0	0
	0	2	0	0	0	0	63	0	0	0	0
	0	0	0	0	0	0	0	172	1	0	0
	0	1	0	0	0	0	1	1	82	0	0

Table 6. Variance explained by the corresponding principal component (first nine components shown).

Component	1	2	3	4	5	6	7	8	9
Variance	440.8967	129.7064	31.2803	24.2814	5.5024	2.2337	0.3736	0.0782	0.0655
% Variance explained	91.3413	7.9053	0.4598	0.2770	0.0142	0.0023	0.0001	0.0000	0.0000

**Figure 2.** Representation of principal component coefficients for each feature and principal component scores for each observation.

ponent Analysis (PCA) on the training data. In Figure 2 the principal component coefficients for each feature and the principal component scores for each observation are visualized in the case of the data obtained with frame-size = 4 and hop-size = 0.125. The direction and length of the vector indicates the contribution of each of the three main principal components. *Zero Crossing*, *Centroid*, *Rolloff* and *bpm* resulted as the most discriminative features. In addition, the observations are represented by red dots in the graph. This visualization gives the general picture of the displacement of the projection of the training samples with respect to the main components; the farther the sample the higher the variance. Considering that the principal component accounts for more than 91% of the overall variance (Table 6), a few other components are required to represent the data space.

5. CONCLUSIONS

In this paper, we have employed ensemble classifiers for automatic expressive musical intention recognition showing promising performance with respect to recent literature. In addition, we have analysed the influence of sampling frequency (in terms of window size and hop distance) which has never been considered so far. In this regard, our preliminary experiments show that a window size roughly equivalent to the echoic memory (4 s) and minimum hop distance (0.125 s) have attained the best results in the recognition of nine expressive intentions of three musical excerpts. To consolidate the work done here and to confirm the obtained performances we plan to run additional validation tests on the dataset used in [8].

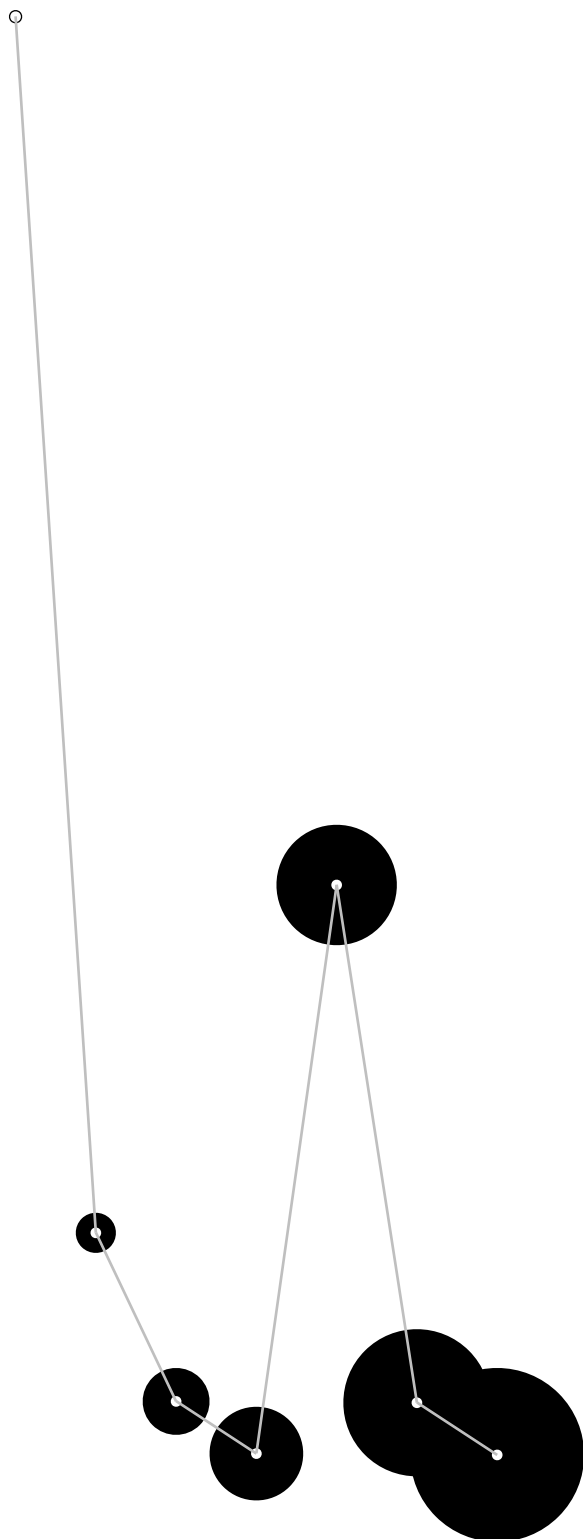
6. REFERENCES

- [1] J. Downie, "Music information retrieval," *Annual Review of Information Science and Technology*, vol. 37, pp. 295–340, 2003.
- [2] N. Orio, "Music retrieval: A tutorial and review," *Foundations and Trends in Information Retrieval*, vol. 1, no. 1, pp. 1–90, 2006.
- [3] N. P. McAngus Todd, "The kinematics of musical expression," *Journal of the Acoustical Society of America*, vol. 97, pp. 1940–1949, 1995.
- [4] S. Canazza, G. De Poli, and A. Rodà, "Analysis of expressive intentions in piano performance," *Journal of ITC Sangeet Research Academy*, vol. 16, pp. 23–62, October 2002.
- [5] P. N. Juslin and J. A. Sloboda, *Music and emotion. Theory and research*. Oxford University Press, 2001.
- [6] B. H. Repp, "A microcosm of musical expression: I. quantitative analysis of pianists' timing in the initial measures of chopin's etude in e major," *Journal of The Acoustical Society of America*, vol. 104, pp. 1085–1100, 1998.
- [7] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet, "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts," *Cognition and Emotion*, vol. 19, no. 8, pp. 1113–1139, 2005.
- [8] L. Mion and G. De Poli, "Score-independent audio features for description of music expression," *IEEE Trans. Speech, Audio, and Language Process*, vol. 16, no. 2, pp. 458–466, 2008.
- [9] P. N. Juslin, "Communicating emotion in music performance: A review and a theoretical framework," in *Music and Emotion: Theory and Research* (P. N. Juslin and J. A. Sloboda, eds.), pp. 305–333, New York: Oxford Univ. Press, 2001.
- [10] R. Dannenberg, B. Thorn, and D. Watson, "A machine learning approach to musical style recognition," in *Proc. Int. Comput. Music Conf. (ICMC97)*, (San Francisco, CA), pp. 344–347, 1997.
- [11] A. Friberg, E. Schoonderwaldt, P. Juslin, and R. Bresin, "Automatic real-time extraction of musical expression," in *Proc. Int. Comput. Music Conf. (ICMC02)*, (Goteborg, Sweden), pp. 365–367, 2002.
- [12] E. Schubert, J. Wolfe, and A. Tarnopolsky, "Spectral centroid and timbre in complex, multiple instrumental textures," in *Proceedings of the International Conference on Music Perception and Cognition* (N. W. University, ed.), (Illinois), 2004.
- [13] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560, 1965.
- [14] W. A. Sethares, *Tuning, Timbre, Spectrum, Scale*. Springer-Verlag, 1998.
- [15] T. G. Dietterich, "Ensemble methods in machine learning," in *International Workshop on Multiple Classifier Systems*, pp. 1–15, 2000.

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
[
\
]
^
_
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Posters

- + Andrea Arcella, Agostino Di Scipio
- + Alberto Boem, Sergio Canazza, Antonio Rodà
- + Federico Costanza, Marco Gasperini, Alessio Rossato
- + Claudio Dionisi, Eliana Carboni
- + Amedeo Casella, Andrea Valle
- + Ivano Morrone
- + Antonio Scarcia, Francesco Abbrescia, Francesco Scagliola
- + Stefano Silvestri



WAVESHAPING AND DISTORTION SYNTHESIS, FROM LICKLIDER (1946) TO CURRENT PERSPECTIVES

Andrea Arcella
Conservatorio S.Pietro a Majella -
Napoli
andrea.arcella@gmail.com

Agostino di Scipio
Conservatorio S.Pietro a Majella
– Napoli
discipio@tin.it

ABSTRACT(English)

Along the history of electronic (analog and digital) sound synthesis and transformation, time and again various "waveshaping" approaches have surfaced. In this paper we survey some of them, and offer insight into what could be seen, in retrospect, as the earliest ever thorough treatment of waveshaping (the application of nonlinear transfer functions to an input signal), pursued in the later 1940's by J.C.R.Licklider in the context of investigations on speech intelligibility.

Based on this survey, further work may eventually bear on the different cognitive potentials opened up for musical purposes by the different methods and perspectives. In general, older and newer methods for the electronic generation of sound reflect differences not only in the technical and theoretical frame, but also in underlying sound representation premises, that is, different ideas of what sound is and means for human beings and how it is or can be creatively dealt with. Such kind of discussion, in our opinion, is of the highest relevance in studies in the history of sound synthesis, and represents a subject matter of interest in sound studies at large.

ABSTRACT (Italiano)

Nella storia della sintesi e della trasformazione del suono, sia nel dominio analogico che digitale, vi sono stati diversi approcci alla sintesi di tipo "waveshaping" (sintesi per distorsione non lineare). In questo articolo viene fatto un excursus storico di questi approcci confrontandoli con la più antica trattazione della distorsione non lineare (da ora DNL) applicata all'audio nota in letteratura: un seminario di J.C.R.Licklider [21] sugli effetti della DNL applicata al parlato. Si ritiene di particolare interesse il documento di Licklider per due motivi: il primo di tipo storico, in quanto questo intervento è probabilmente il primo in letteratura in cui si affronta in modo esteso la DNL¹ applicata a segnali audio; il secondo è che l'approccio di Licklider alla variazione di intelligibi-

lità del parlato in funzione della distorsione può aprire nuove prospettive alle ricerche di tipo artistico con particolare riferimento alle installazioni sonore di tipo interattivo.

1. INTRODUZIONE

La distorsione non lineare (DNL) può essere intesa come una tecnica per operare modifiche su un segnale in ingresso oppure, mettendosi in una prospettiva speculare, per studiare gli effetti che segnali esterni possono avere sul segnale in oggetto nel caso in cui questi ultimi siano rappresentabili come funzioni distorcenti. I modelli che si basano sulla DNL nella teoria dei segnali sono stati proposti già prima della seconda guerra mondiale nell'ambito degli studi sulla teoria dell'informazione e sulle telecomunicazioni.

2. SEGNALI SOTTOPOSTI A FUNZIONI DISTORCENTI

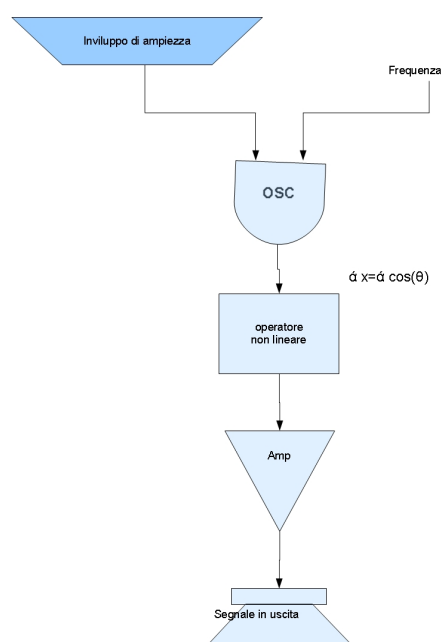


Figure 1. Schema a blocchi della DNL

¹ Più precisamente intendiamo il primo in cui venga utilizzato il framework concettuale della DNL e non approcci diversi usati talvolta con fini analoghi, come ad esempio il vocoder.

L'idea base della DNL è quella di applicare una funzione detta distorcente, ovvero un operatore, al dominio di un'altra funzione in ingresso (Fig. 1)

Un esempio tipico di DNL è quello di un amplificatore la cui funzione di trasferimento non è lineare lungo tutto il range delle ampiezze possibili per il segnale in ingresso (Fig.2).

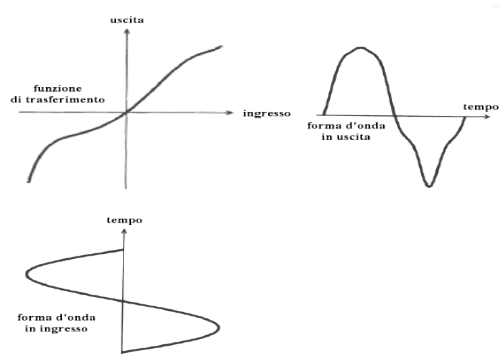


Figure 2. Funzione di trasferimento lineare non lineare [24]

È evidente che il segnale risultante da distorsioni non lineari produce segnali con un contenuto armonico diverso (generalmente più ricco) da quello presente nel segnale originario. Possiamo quindi desumere che la modifica operata sull'ampiezza del segnale induce una modifica sul contenuto in frequenza. Da un punto di vista analitico possiamo definire questo processo come un operatore distorcente $f()$ applicato ad un segnale in ingresso $x(t)$. Tale operatore può essere applicato al dominio o al codominio della funzione ovvero $f()=x(f(t))$ oppure $f()=f(x(t))$.

In linea di principio possiamo applicare qualunque operatore, ad esempio la derivata $f(x(t))=dx(t)/dt$, oppure sperimentare con operatori che non hanno una definizione analitica disegnando arbitrariamente la funzione.

3. LA DNL NELL'INFORMATICA MUSICALE

La DNL è stata applicata ai segnali audio in anni successivi rispetto alla sua formulazione nell'ambito della teoria dei segnali. Riferendoci in particolare all'utilizzo della DNL applicata alla sintesi del suono la prima notizia documentata è riferita all'uso che ne fa Risset nel 1969; nell'esempio 150 (Fig.3) del Catalog of Computer Synthesized Sound [24] c'è la descrizione di uno strumento (clarinetto) in cui l'onda seno è sottoposta ad una "funzione di trasferimento non lineare" ed il controllo di ampiezza del seno "determina l'ammontare di distorsione effettuata sul seno". Risset evidentemente usa il concetto di DNL ma non pubblica al riguardo articoli che diano

presupposti teorici per l'uso della DNL come tecnica di sintesi.

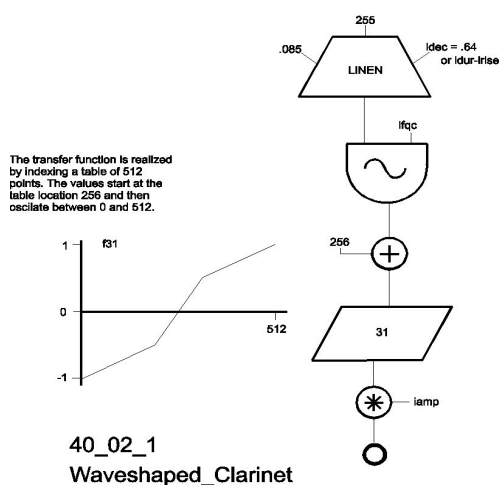


Figure 3. Strumento 150 del Catalog of Computer Synthesized Sound [25]

Bisogna ricordare che l'utilizzo empirico di tecniche riconducibili al framework concettuale della DNL erano già in uso per la sintesi del suono anche se non esisteva (o meglio non era nota) una sistemazione teorica generale. Una tecnica chiaramente riconducibile al concetto di DNL è la sintesi FM [1] introdotta da Chowning nella seconda metà degli anni '60; a posteriori possiamo interpretare la funzione modulante come un operatore applicato alla funzione portante secondo lo schema tipico della DNL. Questa interpretazione della FM come un caso particolare di DNL è però venuto solo più tardi, da quando si è pervenuti ad un assetto teorico della DNL formalizzato in modo chiaro.

Tra i requisiti tipici di qualunque tecnica di sintesi del suono vi è quello di fornire risultati in qualche misura prevedibili. La formalizzazione classica della FM ha portato alla definizione di pochi parametri (indici di modulazione, etc) disponibili direttamente come parametri di controllo del suono nella fase di sintesi. Questo ha determinato il successo della FM come tecnica di generazione sonora. Nel caso più generale della DNL il risultato è arrivato più tardi, condizionato non solo da problemi di ordine concettuale e analitico ma anche dalle trasformazioni tecnologiche in atto negli anni '70 con il passaggio dall'analogico al digitale e l'introduzione sempre più massiccia dei computer. Nella prima metà degli anni '70 Suen [2], Schaefer [3] e altri danno una prima sistemazione analitica della DNL con applicazioni nel dominio analogico. Il problema dell'uso a fini musicali della DNL è quello di trovare una classe di funzioni distorcenti il cui risultato nell'applicazione alla sintesi del suono pre-

vedibile. Dalla teoria matematica delle approssimazioni lineari abbiamo ereditato la possibilità di costruire grandi classi di funzioni partendo da combinazioni lineari di funzioni trigonometriche o polinomiali. Utilizzando questa classe di funzioni come dominio dell'operatore è noto analiticamente il codominio e si può quindi prevedere deterministicamente anche il risultato sonoro della DNL. Il problema che si pone nel dominio analogico è che la progettazione elettronica necessaria per implementare un generico operatore non è per niente semplice, salvo che in alcuni casi notevoli come il clipping, la derivazione o l'integrazione. Questo forse è stato il principale limite allo sviluppo di una teoria generale della DNL negli anni precedenti. Il dominio digitale invece ci dà la possibilità di generare operatori distortanti comunque complessi in modo relativamente semplice ma impone un problema importante assente nel dominio analogico. I segnali generati applicando un generico operatore infatti sono in generale a banda non limitata. Questo è un problema nel dominio digitale in quanto dà luogo al fenomeno dell'aliasing e quindi pone un freno all'uso generale della DNL come tecnica di sintesi.

A questo problema hanno posto rimedio Arfib[4] e Lebrun [5] che in due ricerche indipendenti pubblicate nel 1979 hanno indicato il modo di ottenere segnali a banda limitata dalla DNL. La classe di funzioni distortanti proposte sono quelle ottenibili dai polinomi di Chebishev di primo tipo della forma $T_k(x)$ dove k rappresenta l'ordine del polinomio e x è una sinusoide:

$$T_1(x) = x \quad (1)$$

$$T_2(x) = 2x^2 - 1 \quad (2)$$

.

$$T_{(n+1)}(x) = 2xT_n(x) - T_{(n-1)}(x) \quad (3)$$

Tali polinomi si ottengono quindi da una relazione ricorsiva. La proprietà interessante di questi polinomi è che quando una funzione coseno in ingresso $x(t)$ viene sottoposta ad una funzione distortante $T_k(x)$ il segnale risultante è una sinusoide con frequenza pari alla k -esima armonica del coseno in ingresso. Quindi dato che ciascun polinomio è in grado di generare un'armonica ben definita possiamo usare combinazioni lineari pesate di polinomi per creare combinazioni spettrali prevedibili e a banda limitata.

In generale quindi la funzione distortante con n termini polinomiali sarà del tipo

$$f(x) = \frac{h_0}{2} + \sum_{k=1}^n h_k T_k(x) \quad (4)$$

Se alla funzione in ingresso viene applicato un inviluppo temporale dinamico otterremo uno spettro dinamico analogamente a ciò che avviene con la variazione dell'indice

di modulazione della FM. Ricordiamo infatti che una variazione di ampiezza del segnale di ingresso determina una variazione nello spettro del segnale di uscita. In riferimento alla Fig.1 il coefficiente $\hat{a}$, che rappresenta l'ampiezza del segnale di ingresso può essere una funzione che varia nel tempo, ad esempio l'inviluppo indicato.

In anni successivi la letteratura si è arricchita di nuovi contributi tesi ad ottimizzare gli algoritmi per la sintesi DNL ed a trovare nuove soluzioni di applicazione nel vasto mondo dei Virtual Instruments [17] [18]. Un'altra direzione di ricerca ha utilizzato la DNL come motore di sintesi per strumenti "gesture controlled" [15]. Dal punto di vista della generazione sonora la DNL è stata spesso indicata come particolarmente adatta per la creazione di texture sonore [13] [14]. Infine la ricerca di Di Scipio [10] [11] [12] conduce verso un ulteriore processo di generalizzazione della DNL riconducendola all'interno del framework più generale delle funzioni non lineari iterate, di cui la DNL rappresenta il caso particolare di IFS di grado zero (ovvero senza iterazioni).

4. UN CONTRIBUTO STORICO: IL TALK DI J.C.R. LICKLIDER

4.1 Il framework concettuale della DNL in Licklider

In un talk del 1950 alla Macy Conference [21] Licklider esamina gli effetti del rumore sull'intelligibilità del parlato. Ciò che risulta interessante ai fini storici per coloro che si occupano di suono sono le rappresentazioni dell'interazione tra segnale e rumore che lo scienziato propone. Lo scienziato effettua una disamina delle tipologie di rumore che possono compromettere il parlato. Nella prima parte del testo Licklider offre una panoramica dei rumori che si possono modellare come una funzione distortrice applicata al dominio del parlato, il cui caso più semplice è una dilatazione o una compressione dell'asse temporale corrispondente ad una riproduzione rallentata o accelerata della registrazione audio. Nel seguito vengono affrontate le distorsioni applicate al codominio della funzione utilizzando il framework della DNL come lo abbiamo precedentemente illustrato. Le figure incluse nel suo articolo sono particolarmente indicative (Fig.4) (Fig.5) (Fig.6). Prima di entrare nel merito dell'intelligibilità del parlato, Licklider si sofferma su una meticolosa analisi degli effetti della distorsione armonica usando come segnali di test onde sinusoidali; su tali segnali applica varie funzioni distortrici tra cui il peak clipping simmetrico, il clipping centrale, la derivazione e l'integrazione (Fig.6). La scelta di questo tipo di funzioni non deve stupire, infatti non solo sono rappresentative di una serie di disturbi delle comunicazioni ma sono anche particolarmente semplici da riprodurre in laboratorio mediante l'elettronica analogica a componenti discreti. L'interesse storico dell'articolo risiede nel fatto che Licklider utilizza il framework concettuale della DNL applicato ad un segnale audio con venti anni di anticipo rispetto al suo utilizzo nell'informatica musicale e con quasi quarant'anni di anticipo rispetto alla sua formalizzazione definitiva in questo ambito. Inoltre osserviamo che durante il dibattito seguito all'esposizione degli argomenti principali del talk alcuni uditori fanno esplicito riferimento a possibili utilizzi musicali della DNL. Dato il contesto della conferenza e l'assenza di compositori tra gli uditori la questione viene rapidamente abbandonata ma riteniamo che anche solo questi cenni indichino l'articolo di Licklider come un antecedente storico importante.

4.2 Approcci sistemici per suoni e parlato

L'altro punto focale dell'articolo di Licklider è appunto lo studio dell'intelligibilità del parlato. Siamo perfettamente consapevoli che gli studi sul parlato (analisi, sintesi, riconoscimento, etc.) hanno una storia che, pur procedendo parallelamente a quella dell'informatica musicale, è diversa. E' diversa soprattutto la finalità di tali studi che non sono rivolti alla produzione musicale.

Ciò che invece ci sembra interessante nell'articolo di Licklider è l'approccio di tipo sistemico delle sue ricerche, che ricordiamo essere inserite in un filone culturale più generale da cui è emerso il pensiero di Bateson. Questo fatto è ben evidente sia nel modo in cui Licklider presenta la sua ricerca sia nel dibattito che segue il talk.

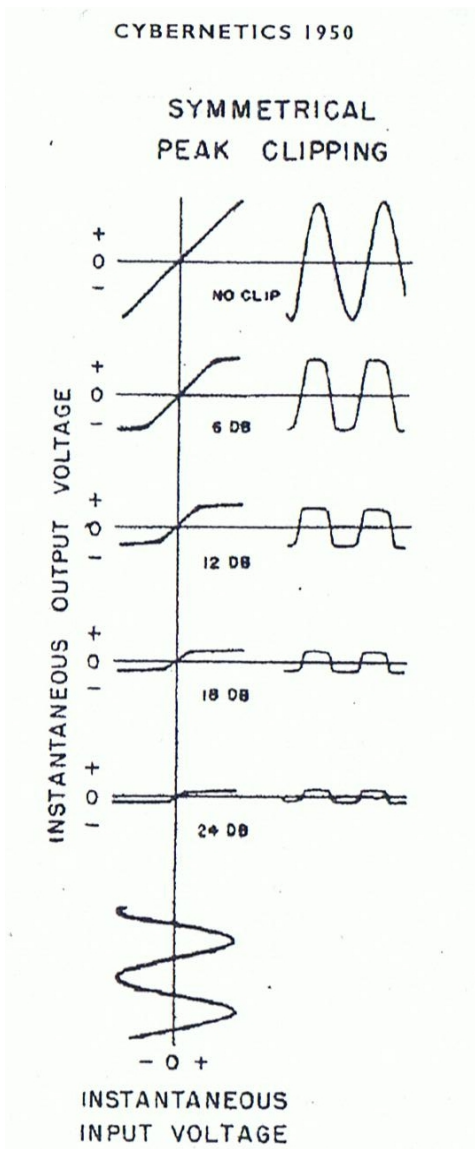


Figure 4. Effetto del peak clipping simmetrico applicato ad una sinusoide.

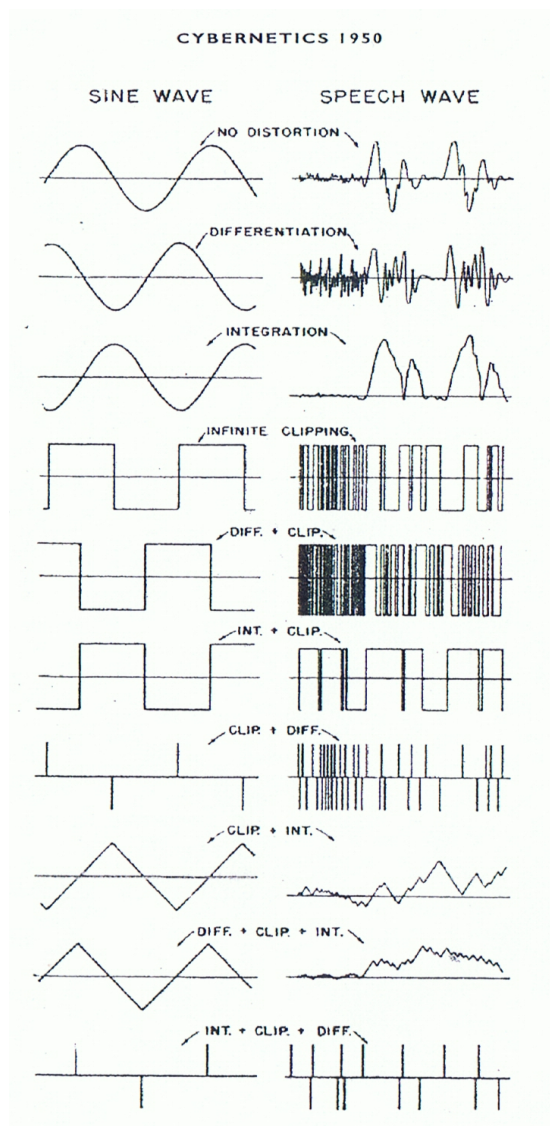


Figure 5. Effetto di varie funzioni distorcenti e di loro combinazioni rispettivamente su una sinusoide (colonna a sinistra) e sul segnale parlato (colonna di destra)

Tale approccio ricongiunge i due territori di ricerca (parlato e musica) in un nuovo campo di esplorazione delle arti del suono a partire dall'interesse sempre crescente verso il rapporto sistemico che unisce i soundscapes naturali e/o urbani con l'ascoltatore fino alla realizzazione di installazioni sonore interattive in cui il parlato gioca un ruolo di primo piano. In questi contesti, con finalità dichiaratamente artistiche, il rapporto tra segno e significato coinvolge tanto il gesto sonoro verbale quanto quello non verbale.

5. CONCLUSIONI

Questo lavoro ha indicato la necessità di espandere la ricognizione storica di alcuni aspetti della sintesi del suono con particolare riferimento alla Distorsione non Lineare. Si indica l'articolo di J.C.R. Licklider come un lavoro seminale a tal proposito. Si evidenzia inoltre un rinnovato interesse verso l'approccio sistemico alla produzione sonora già presente nel testo citato.

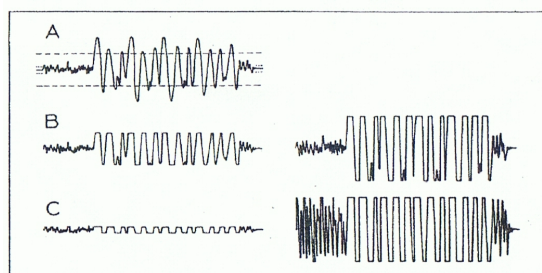


Figure 6. Effetto del peak clipping sul parlato. (A) con parola non distorta. (B) con peak clipping di 6 dB e successiva normalizzazione. (C) con clipping ancora maggiore

6. REFERENCES

- [1] J. Chowning, The Synthesis of Complex Audio Spectra by Means of Frequency Modulation, in "Journal of the Audio Engineering Society", Vol. XXI, No. 7, Sept. '73, New York, NY, 1973.
- [2] C.Y. Suen (1970), "Derivation of harmonic equations in non linear circuits", J. of the Audio Eng. Soc., Vol.18(6):675-676.
- [3] Schaefer, R. A. 1970. "Electronic Musical Tone Production by Nonlinear Waveshaping." Journal of the Audio Engineering Society 18(6):675-676.
- [4] Arfib, D. "Digital synthesis of complex spectra by means of multiplication of nonlinear distorted sine waves", Journal of the AES, 27(10), 1979.
- [5] LeBrun, . "Digital waveshaping synthesis", Journal of the AES, 27(4), 1979.
- [6] Beauchamp, J. "Brass tone synthesis by spectrum evolution matching with nonlinear functions", Computer Music Journal, 3(2), 1979.
- [7] Risset, J.C. An introductory catalog of computer synthesized sounds. Bell Telephone Laboratories, 1969.
- [8] Schaefer, R. "Electronic musical tone production by nonlinear waveshaping", Journal of the AES, 18(4), 1970.

- [9] Roads, C. "A tutorial on nonlinear distortion or waveshaping synthesis", in *Foundations of computer music* (C.Roads and J.Strawn eds.). MIT Press, 1985. [Originally appeared in *Computer Music Journal* 3(2), 1979.]
- [10] Di Scipio, A. and Prignano, I. "Synthesis by Functional Iterations. A Revitalization of Nonstandard Synthesis", *Journal of New Music Research*, 25(1), 1996.
- [11] Di Scipio, A. "Iterated nonlinear functions as a sound generating engine", *Leonardo*, 34(2), 2001
- [12] Di Scipio, A. "Synthesis of environmental sound textures by iterated non linear functions and its ecological relevance to perceptual modeling", *Journal of New Music Research*, 32(2), 2002.
- [13] Filatriau, J.J. *Réalisation d'un instrument de synthèse de textures sonores en temps réel*, DEA Université Aix Marseille II, 2004
- [14] Recht, B. and Whitman, B. "Musically expressive sound textures from generalized audio", *Proc. Int'l Conference on Digital Audio Effects (DAFx03)*, 2003.
- [15] Filatriau, J.J. And Arfib, D. "Instrumental gestures and sonic textures", *Proceedings of the Sound and Music Computing Conference*, 2005.
- [16] Filatriau, J.J., Arfib, D., Glowinski, D. and Volpe, G. "Virtual objects as mediation between sonic textures and interactive gestures", *Proceedings of the Int'l Computer Music Conference*, 2007
- [17] Lazzarini, V. and Timoney, J. "New Perspectives on Distortion Synthesis for Virtual Analog Oscillators", *Computer Music Journal*, 34(1), 2010.
- [18] Smith, J. O. "Physical Audio Signal Processing for Virtual Musical Instruments and Audio Effects". Available online at ccrma.stanford.edu/~jos/pasp, 2008.
- [19] [18] Licklider, J.C.R., "Effects of amplitude distortion on the intelligibility of speech", *Journal of the ASA*, 18, 1946.
- [20] Licklider, J.C.R. And Pollack, I., "Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech", *Journal of the ASA*, 20, 1948.
- [21] Licklider, J.C.R. "The manner in which and extent to which speech can be distorted and remain intelligible", *Transactions of the 7th Conference on Cybernetics Circular Causal and Feedback Mechanisms in Biological and Social Systems* (H.von Foerster ed.), J.Macy Foundation, 1950. [Reprinted in *Cybernetic The Macy Conferences 1946 – 1953* (C.Pias ed.), Diaphanes, 2003].
- [22] Licklider, J.C.R. "Man computer symbiosis", *IRE Transactions on Human Factors*, 1960.
- [23] Licklider, J.C.R. "The computer as a communication device", *Science and Technology*, April 1968.
- [24] *Amsterdam Catalog of Csound Computer Instrument* (Website) <http://www.music.buffalo.edu/hiller/accci/>
- [25] "Un' introduzione all'Informatica Musicale e ai suoi aspetti in ambito didattico" (Website) <http://www.fisica.unina.it/mfa/acust/materiale%20sito/Sistemi%20di%20sintesi/index6.htm>

PROGETTAZIONE E REALIZZAZIONE DI UN AUDITORY DISPLAY PER LA RIABILITAZIONE MOTORIA

Alberto Boem

Università degli Studi di Udine
boem.alberto@gmail.com

Sergio Canazza Targon

Università degli Studi di Padova
canazza@dei.unipd.it

Antonio Rodà

Università degli Studi di Udine
antonio.roda@uniud.it

ABSTRACT

I sistemi di riabilitazione motoria tecnologicamente assistita integrano sempre più spesso dispositivi robotici con ambienti multimediali, al fine di migliorare l'efficacia della terapia grazie all'uso coordinato di più canali sensoriali. In questi sistemi, l'audio viene spesso utilizzato per aumentare il realismo all'interno di scenari 3-D di *virtual reality*, mentre ancora poco sfruttate sono le sue potenzialità in termini di *auditory display*. Questo lavoro presenta la progettazione e lo sviluppo in versione prototipale di un *auditory display* finalizzato alla riabilitazione tecnologicamente assistita di pazienti affetti da deficit motorio in seguito a danni di tipo neurologico. In particolare, l'*auditory display* è stato progettato per l'integrazione in un sistema robotico di assistenza della camminata e fornisce al paziente un segnale sonoro semplice e intuitivo, basato sui timbri vocalici della lingua italiana. Una prima valutazione dell'*auditory display* è stata effettuata sonorizzando alcune traiettorie misurate per mezzo del sistema di riabilitazione per arti inferiori ARTHuR.

1. INTRODUZIONE

Negli ultimi anni sono stati sperimentati sistemi di riabilitazione tecnologicamente assistita [1] basati sull'integrazione di dispositivi robotici con ambienti multimediali, al fine di migliorare l'efficacia della terapia motoria attraverso l'utilizzo coordinato di più canali sensoriali. All'interno di questi sistemi, l'audio viene per lo più impiegato per aumentare il realismo di scenari 3-D di *virtual reality*. Ancora poco sfruttate, invece, sono le potenzialità in termini di *auditory display* [2], ossia per fornire all'utente un feedback acustico che sia rappresentativo dello stato del sistema e di come questo varia in funzione delle proprie azioni. L'utilizzo del suono nei dispositivi orientati alla riabilitazione motoria tecnologicamente assistita può avere varie finalità ([3], [4]), tra cui:

- permettere un maggior coinvolgimento del paziente nel processo riabilitativo;
- permettere al paziente una maggiore comprensione della qualità dei propri movimenti, correlandoli a eventi sonori opportunamente prodotti.

Grazie al feedback acustico, i pazienti possiedono un ausilio in più per svolgere i compiti e le azioni richieste e, al tempo stesso, i terapisti possono avere una migliore comprensione delle caratteristiche dei movimenti effettuati dal paziente. Inoltre, l'impiego di un feedback uditivo favorisce l'uso di questi dispositivi anche a soggetti ipovedenti e a pazienti che si trovano nella fase acuta successiva ad un evento traumatico oppure ad un ictus. In questo caso, infatti, i pazienti trovano spesso difficoltà a mantenere l'attenzione su uno schermo, mentre mantengono la capacità di percepire stimoli sonori.

In un recente studio, Avanzini et al. [4] hanno analizzato l'impiego del suono in numerosi dispositivi orientati alla riabilitazione motoria, da cui risulta che quasi mai il suono viene utilizzato come guida all'esecuzione di un compito di tipo motorio, preferendo impiegarlo per aumentare il realismo di scene di *virtual reality*. L'obiettivo del nostro progetto, invece, è la generazione di un segnale sonoro che evidenzii gli eventuali errori, anche minimi, compiuti dal paziente durante l'esecuzione dei compiti richiesti dalla terapia. Questo viene effettuato attraverso il continuo confronto tra la traiettoria corretta (prodotta dal dispositivo robotico che, agganciato al piede, mostra al paziente il movimento esatto) con quella prodotta in tempo reale dal paziente durante gli esercizi di riabilitazione (agganciato al robot, ma in maniera passiva) tramite la sintesi in tempo reale delle vocali. Il suono può quindi divenire utile al paziente che, in seguito al trauma, perde o diminuisce la propria capacità propriocettiva: attraverso la produzione di sequenze sonore egli può "ascoltare" il proprio movimento, che viene rappresentato attraverso un passaggio graduale tra timbri vocalici. La traduzione del movimento in suono, inoltre, permette di sviluppare una componente ludica nel processo di riabilitazione (quasi come imparare a suonare uno strumento musicale), aspetto non trascurabile data la costanza e la fatica richiesta nelle lunghe sessioni di riabilitazione. La scelta di usare timbri vocalici è dovuta alla riconosciuta capacità dell'orecchio umano di riconoscere ed identificare questi timbri [5], che sono alla base del linguaggio parlato. Diversi studi [6], inoltre, hanno sottolineato come la capacità del sistema umano di riconoscere ed identificare questa tipologia di suoni può essere d'aiuto per la rappresentazione di tipi complessi di dati.

Per valutare il funzionamento dell'*auditory display*, abbiamo utilizzato i risultati di un test condotto sul robot orientato alla riabilitazione della camminata denominato ARTHuR (Ambulation-Assisting Robotic Tool for Human Rehabilitation), una cui descrizione è esposta in Emken et

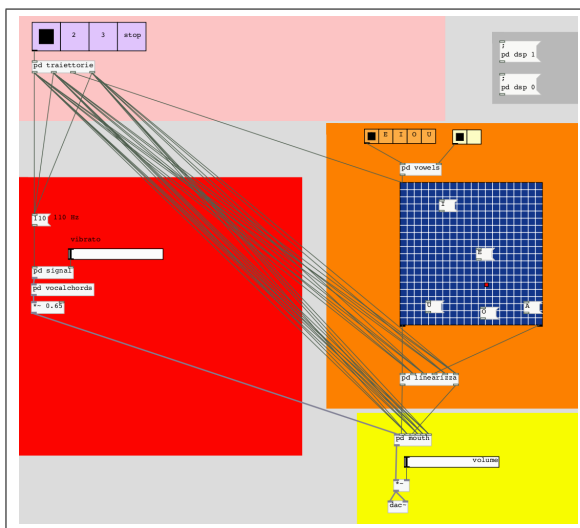


Figure 1. Screenshot della patch principale, sviluppata in Pure Data.

al. [7]. Oltre alla descrizione del robot, questo articolo riporta le variazioni nei pattern di movimento di un gruppo test di pazienti in seguito ad una fase di terapia motoria effettuata con l'ausilio di ARTHuR.

2. AUDITORY DISPLAY PER LA RIABILITAZIONE MOTORIA

L'*auditory display* consiste in un'applicazione software che permette la traduzione dei movimenti del piede – effettuati da un soggetto sottoposto ad un processo di riabilitazione della camminata con l'ausilio di un dispositivo robotico – in un segnale sonoro, sintetizzato in tempo reale, che associa timbri vocalici al movimento. Il prototipo è stato implementato in ambiente Pure Data, per la sua versatilità e portabilità. La Figure 1 mostra la schermata principale della patch che implementa l'*auditory display*.

Oltre alla patch principale, il sistema comprende alcune subpatch e sezioni relative ai differenti processi di analisi del movimento e di sintesi del suono. La patch riceve in ingresso le coordinate x e y che rappresentano la posizione del piede nel tempo. Questi valori vengono opportunamente scalati e mappati all'interno dello spazio delle vocali i cui assi corrispondono al primo (F1) e al secondo (F2) formante spettrale.

Per la generazione del suono in tempo reale è stata scelta una sintesi per formanti [8]. Questo metodo, infatti, permette di generare, con continuità, anche i timbri di transizione tra una vocale e l'altra. Inoltre, ha un costo computazionale ridotto, che consente la sonificazione in tempo reale. Un treno di impulsi è stato utilizzato come sorgente, successivamente elaborata attraverso una serie di filtri risonanti (come minimo due) in parallelo, con caratteristiche tali da simulare i formanti vocalici.

Come si può vedere in Figure 1, l'asse x è stato dimensionato in modo da variare tra 250 e 750Hz, l'asse y tra 800 a 2300Hz. Questi valori rispecchiano il range di valori

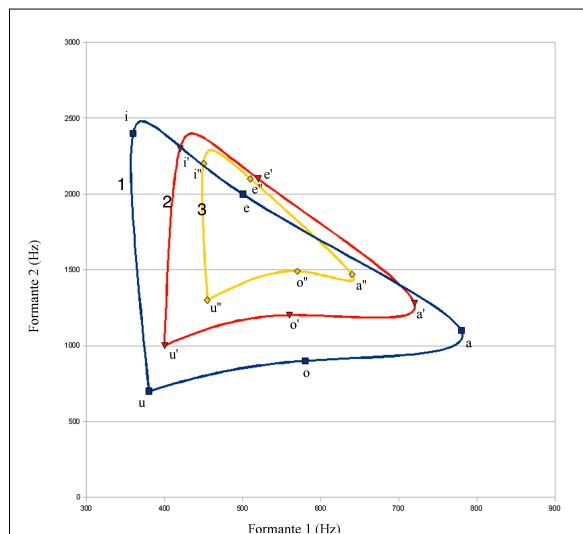


Figure 2. Le tre traiettorie utilizzate per testare il sistema mappate nello spazio delle vocali.

corrispondenti alla posizione dei primi due formanti delle vocali italiane. I due filtri formantici sono definiti tramite i parametri caratteristici: frequenza centrale, guadagno e larghezza di banda. In questa prima versione del sistema, il guadagno e la larghezza di banda dei filtri vengono fissati in fase di inizializzazione, e solo il parametro frequenza centrale viene modificato in tempo reale tramite i valori di ingresso. I valori di inizializzazione sono stati scelti seguendo le linee guida per la sintesi per formanti che possiamo trovare in [6, 8].

3. RISULTATI

Per verificare se l'*auditory display* è in grado di rendere percepibili, nel dominio sonoro, le differenti caratteristiche delle traiettorie motorie, sono state prese in esame tre traiettorie registrate nel corso di una sessione di riabilitazione motoria dal robot ARTHuR. La Figure 2 mostra le tre traiettorie mappate nello spazio delle vocali. La traiettoria di colore blu (1) è la traiettoria *target*, ottenuta mediante l'intervento dell'operatore; le traiettorie di colore arancio (2) e giallo (3) sono due traiettorie errate, prodotte dal paziente nel cercare di riprodurre la traiettoria *target*.

Le Figure 3, Figure 4 e Figure 5 mostrano i sonogrammi del segnale audio generato a partire da quelle traiettorie. Si può notare il passaggio tra le diverse formanti vocaliche. In particolare, rispetto al suono prodotto dalla traiettoria 3 (Figure 5), quello relativo alla traiettoria 2 (Figure 4) è più simile a quello relativo alla traiettoria *target* (Figure 3), essendo minore l'errore motorio commesso dal paziente.

4. CONCLUSIONI E LAVORI FUTURI

Questo studio ha presentato una proposta di feedback uditivo integrato ai sistemi robotici di riabilitazione motoria. Si tratta di uno studio preliminare alla realizzazione di un più complesso *auditory display*, che andrà sperimentato colle-

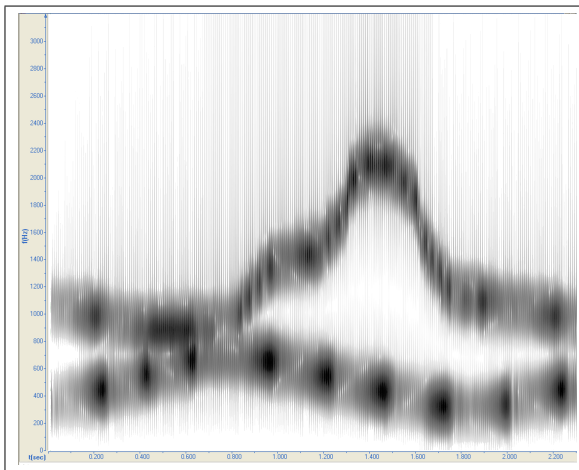


Figure 3. Sonogramma corrispondente alla sonificazione della traiettoria *target*

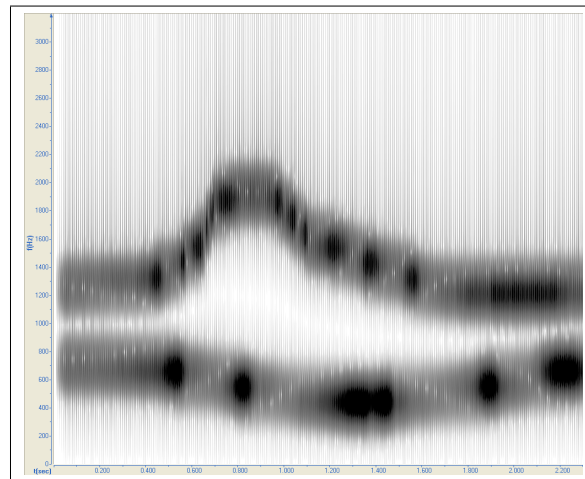


Figure 5. Sonogramma corrispondente alla sonificazione della traiettoria 3.

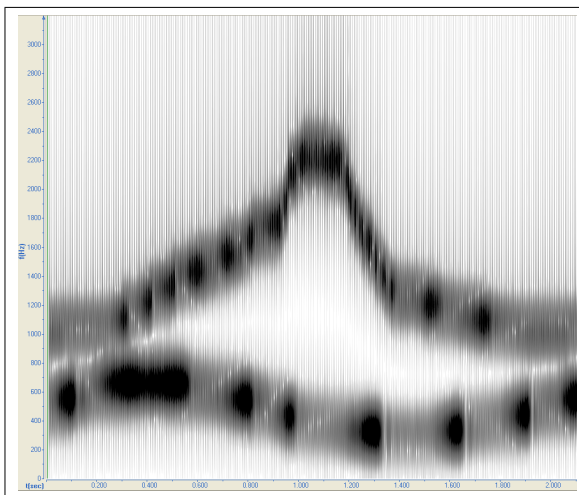


Figure 4. Sonogramma corrispondente alla sonificazione della traiettoria 2.

gandolo con un robot reale e valutandone l'effetto su alcuni casi clinici. Verranno inoltre sperimentate sia l'impiego di tecniche di mapping che prendono in considerazione parametri cinematici come velocità e accelerazione, che un modello di sintesi dei suoni vocalici più completo.

5. REFERENCES

- [1] H. I. Krebs and N. Hogan, "Therapeutic robotics: A technology push," in *Proceedings of the IEEE*, 94 (9), pp. 1727–1738, 2006.
- [2] G. Kramer, *Auditory Display : Sonification, Audification, and Auditory Interfaces*. Addison Wesley, 1994.
- [3] W. Harwin, J. Patton, and V. R. Edgerton, "Challenges and opportunities for robot-mediated neuro-

rehabilitation," in *Proceedings of the IEE*, 94 (9), pp. 1717–1726, 2008.

- [4] F. Avanzini, A. D. Götzén, A. Spagnol, and A. Rodà, "Integrating auditory feedback in motor rehabilitation system," in *Proceedings of International Conference on Multimodal Interfaces for Skills Transfer (SKILLS09), Bilbao*, pp. 1717–1726, 2009.
- [5] R. J. Cassidy, J. Berger, K. Lee, M. Maggioni, and R. R. Coifman, "Auditory display of hyperspectral colon tissue image using vocal synthesis model," in *Proceedings of 10th International Conference on Auditory Display, Sydney*, pp. 1717–1726, 2004.
- [6] M. Kleiman-Weiner and J. Berger, "The sound of one arm swinging: A model for multidimensional auditory display of physical motion," in *Proceedings of the 12th International Conference on Auditory Display, London, Uk*, pp. 278–280, 2006.
- [7] J. L. Emken, S. J. Harkema, J. A. Beres-Jones, C. K. Ferreira, and D. J. Reinkensmeyer, "Feasibility of manual teach-and-replay and continuous impedance shaping for robotic locomotor training follow spinal cord injury," in *IEEE Transactions on Biomedical Engineering*, 55 (1), pp. 322–334, 2008.
- [8] C. Roads, *The Computer Music Tutorial*. Cambridge, MA: The MIT Press, 1996.

S. GIORGIO LAPTOP ENSEMBLE

Federico Costanza

Marco Gasperini
Meccanica Azione Sonora
meccanicazione sonora@libe
ro.it

Alessio Rossato

ABSTRACT

With this paper we would like to present the first experience of a laptop orchestra in Italy, the *S. Giorgio Laptop Orchestra*, founded in Venice in 2009, with the support of the Foundation “Giorgio Cini”, by students and ex-alumni of Alvise Vidolin, whose first performance took place in July 2009 with the execution of the collective piece *.WAV, Ad Alvise Vidolin per il suo 60° compleanno*. First some aesthetic/technological aspects will be presented regarding audio communication technologies via Local Area Network, followed by the description of the score and of the final set-up of the ensemble.

1. INTRODUCTION

The idea of setting up a laptop ensemble was born in 2009 by students and ex-alumni of Alvise Vidolin as a tribute to (our) Master in the occasion of the celebrations for his 60th birthday on 11th July 2009 at Sala degli Arazzi (see Figure 1) at the Foundation “Giorgio Cini” in Venice¹ [6].

The ensemble of 8 performers² played, together with the *Paul Klee* string quartet³, the collective work *.WAV, Ad Alvise Vidolin per il suo 60° compleanno*⁴ for string quartet and laptop ensemble⁵.

The title of the score is an acclamation to Alvise Vidolin, based on the ambiguity between his acronym-signature (AV) and the well-know file extension of digital audio.

Copyright: © 2010 Last name et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ On that occasion, the book *60 dB. La scuola veneziana di musica elettronica*, Paolo Zavagna (editor), Firenze, Leo S. Olschki, 2009, was presented.

² Founding members of *S. Giorgio laptop ensemble*: Stefano Alesandretti, Federico Costanza, Marco Gasperini, Monica Karina Lopez Lau, Marco Marinoni, Luca Richelli, Alessio Rossato, Julian Scordato, Paolo Zavagna.

³ Members of *Paul Klee* string quartet were: Alessandro Fagioli and Stefano Antonello, violins, Andrea Amendola, viola, Giancarlo Trimboli, cello. <<http://www.quartettopaulklee.it/>> [30.11.2010]

⁴ *.WAV, to Alvise Vidolin for his 60th birthday*.

⁵ The publisher is **ArsPublica** <<http://www.arspublica.it/>>. Author/pieces are: Federico Costanza, *Ossò*; Marco Gasperini, *Schegge*; Marco Marinoni, *13glifi.exit*; Alessio Rossato, *Inclinato*; Julian Scordato, *Interlinea*.

2. COMPOSING THE SYSTEM

The first question to be solved in the development of this project regarded the understanding of what exactly meant to compose a laptop ensemble and *for* a laptop ensemble. It has been clear from the beginning that each issue in this field posed questions that needed decisions whose effects would be spread on the way the ensemble would have manifested itself and hence meant *composing*.

First a review of known experiences in this field was made to focus on the main questions [2] [11] [13] [15]. We then pointed our attention in considering the laptop *from the outside*, as an *instrument*, trying to understand it as a *phenomenon*. This yielded the theoretical issues treated in [6]. It suffices here to say that it emerged, according to the *nature* (i.e. intrinsic features) of the system, that laptops should constitute a network, and since it would have to be a sound producing network, the elements should communicate sound and *in* sound.



Figure 1: S. Giorgio Laptop Ensemble at S. Giorgio.

The whole development of the musical structure of the ensemble was carried on while working on a porting to *pure-data*⁶ [7] and *Max/MSP* (Cycling ‘74)⁷ of the sound processing environment of *5 interazioni cicliche alle differenze sensibili*, for string quartet and live-electronics (1997-98)⁸, by the Italian composer Agostino Di Scipio,

⁶ <<http://puredata.info/>> [30.11.2010].

⁷ <<http://cycling74.com/products/maxmspjittr/>> [30.11.2010].

⁸ Porting made by Marco Gasperini; patches available from the composer (<<http://xoomer.virgilio.it/adiscipi/>> [30.11.2010]).

originally implemented using *Kyma* (Symbolic Sound¹). Many hints emerged in making this last work that have had relevant influences on the composition of the system and its dynamics [4].

Given the considerations above we came to adopt some network technology terms (*client*, *server*) and to locate two possible levels of action:

1. *client* level: said also *local* level, corresponding to the fields of actions of a single orchestral player;
2. *server* level: said also *global* level, corresponding to the field of actions of the conductor who manages links (interactions) between performers, the whole dynamics of the audio stream and the sound diffusion through a P.A. system.

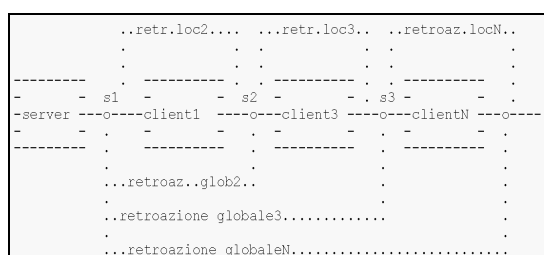


Figure 2: first draft for the *S. Giorgio Laptop Ensemble*.

Once that a first draft of the ensemble was sketched out (see Figure 2) it was clear, from our point of view, that a *conditio sine qua non* was that audio communication should be done through Local Area Network protocols and not through audio interfaces, given the complexities of setting up a large and versatile group with the last and the advantages given by the former (see below). This seemed indeed more consistent with the elements of the system, considering the fact that network communication is the *natural* form of communication between computers [1].

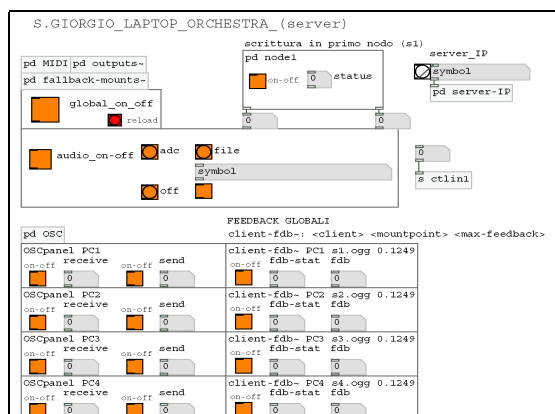


Figure 3: *pure-data* prototype of the *server*.

Several tests and prototypes were then developed to find the solution most appropriate to our frame². The first

¹ <<http://www.symbolicsound.com/>> [30.11.2010].

² Early stages of development were carried on with the precious help of Bruno Curtolo on network communication protocols and in general on network issues.

prototype was implemented in *pure-data* using this software as an interface for the *Icecast2* streaming media server³, which supports *Ogg Vorbis* and *MP3* audio streams. This tool is used mainly for Internet radio broadcasting and its use for our aims seemed to be a 'stress test' for it. The whole environment was made-up by a laptop acting as *server*, carrying on the *Icecast2* service and the *pure-data* patch which acted as an interface for the service itself (see Figure 3) and managed the audio network and its dynamics; on all connected laptops was acting the *client* patch who received and sent one channel of *OGG* audio stream from/to the *server*⁴. Since many members of the ensemble used *Max/MSP* as the main sound processing tool, the whole environment was implemented even with this software, being available the same objects used in *pd*⁵.

The main shortcomings of this system are the delay times in the transmission line, up to several seconds long: when the *server* send to a *client* a signal, it will arrive with, let's say, a two seconds delay. The *client* will process this signal and the results will be audible through its *local* output in the *client* real time, by means of a little loudspeaker placed nearby; if the result is then sent back to the *server* it will be received after some other seconds of delay and made audible through the global output (P.A. system). What may have been seen as an unacceptable defect we had considered as an intrinsic feature of the system, acting at several formal levels (both *local* and *global*) and used it as an opportunity to build-up complex and non-linear time/space relationships.

Stability issues prompted us to change the described environment (but not its phenomenological features) and we focused on uncompressed audio transmission by using the objects **net**send~ and **net**receive~, available both for *pd* and *Max/MSP*⁶; this seemed at first an easier and more reliable solution, but it turned out that there were several compatibility problems (the objects were no longer supported and on Intel based MacOS they didn't worked at all). The most reliable solution was, finally, to use a *VST* plug-in, which manages audio communication through UDP protocol (*Wormhole2*⁷), given the stability and compatibility of the *VST* standard with several operating systems and audio software packages.

The *clients* interface (see Figure 3) allows the performers to see the levels of incoming and outgoing audio streams and to vary the main performance parameters of *local* feedback (affecting timbre) and *local sensitivity* (affecting rhythmic articulation of audio stream, see [7]).

The main attractions of network technology to transmit audio communication can be thus summarized:

- Universality: *natural* communication between

³ <<http://www.icecast.org/>> [30.11.2010]

⁴ The objects used to communicate between *Icecast2* and *pure-data* were **oggcast~** and **oggamp~** by Olaf Matthes distributed with *pd-extended* (<<http://puredata.info/downloads>>).

⁵ <<http://www.nullmedium.de/dev/oggpro/index.html>> [31.11.2010]

⁶ <<http://www.nullmedium.de/dev/netsend-/>> [30.11.2010]

⁷ <<http://code.google.com/p/wormhole2/>> [30.11.2010]

computers;

- cheapness: there's no need for dedicated devices;
- simplicity: a *client* may *join in* just plugging an ethernet cable into the network switch independently from the number of channel that it will be sending/receiving;
- may be freely extended, given a same amount of cables; increasing the number of channels do not mean an increase of cables and a bandwidth of 1 Gbit/s permits an intense data stream as needed by a multichannel audio transmission.

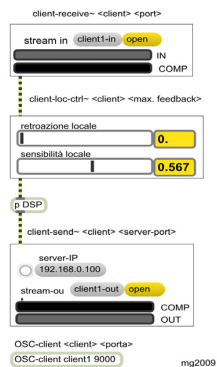


Figure 4: client final interface.

It is fundamental, in such a context, to keep the set-up as easy as possible to avoid waste of time and concentration during rehearsals due to *cables-debugging* activities. Once a stable configuration of the network communication is reached it may just be saved and recalled in rehearsals.

However it should be outlined that, given the state of the art, this solution presented problems (apart from that of transmission delays) in the fact that it doesn't exist a diffused standard, even though the matter has been variously faced and resolved (some examples in [3] [9] [17]).

3. .WAV: THE SCORE

3.1. Composing/Interfacing/Control: implications and choices

The primary goal which emerged between the members of *S. Giorgio Laptop Ensemble*, was to couple the described system with instrumental sounds, in order to create a sort of “extended chamber music” in which electronics performers do act on processes strictly related to instrumental *behaviours*.

Focusing our attention on *behaviors* and *actions* in making music together came quite naturally, because of our formation as instrumentalists as well as composers. Particularly a gestural component that linked acoustical and visual interaction between the musicians as well as between performers and audience, was needed. Traditional

players were thus considered fundamental because of the lack of gestural *necessities* in electronic music performance.

At the same time we were posing ourselves different questions: how to work with a laptop ensemble? What kinds of sounds can we create? How do we compose with this sound and the ensemble? How can we *physically* control these sounds through many laptops? And not least: how do we play a laptop ensemble with one or more acoustic instruments? How do we organize a large number of players in this context? With or without a conductor?

After some initial tests all the members have decided to develop a “collective score” (.WAV, see Figure 5) to organize the actions both of the string quartet and of the laptop ensemble.

The string quartet was chosen as the source of acoustical matter for the fact that in the concert for Vidolin should be played *5 interazioni cicliche alle differenze sensibili* by Di Scipio¹ (see Chapter 2).

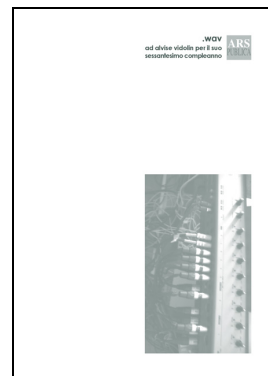


Figure 5: .Wav cover (courtesy of Editor ArsPublica, Carrara, 2009).

The members of the ensemble may thus participate at three different levels to the project (see Table 1):

- composition of short fragments for string quartet;
- design and implementation of sound processing environments;
- performance of the score in the ensemble.

3.2. Composing/Interfacing/Control: writing

The fragments should have been written according to the following common directives:

- should not have been longer than 1'30”;
- each fragment should have been made of several distinct musical *gestures*.

This kind of musical writing makes it possible the *invention* of an unitary event which is the result of the assembly of the various gestures between them and with their electronic counterparts. Moreover it should be outlined how it is possible to differently assemble the performance score for each performance and even to include newly written pieces (see in Figure 6 a possible

¹ Paul Klee String Quartet, Marco Gasperini, live-electronics and sound direction.

combination of two distinct gestures by two different authors).

Among the sound processing environments designed by the performers there were:

- a granular processing unit which processed incoming signals according to semi-automatic control signals [5];
- a *Vocoder* unit in which the incoming signal could be used both as the analysis or the synthesis sound [14];
- a stack of 24 *tuned* comb filters [12];
- an impulse generator with glottal stops characteristics whose synthesis parameters varied according to the features of the audio stream passing through the *client*.

Every single unit has been designed to establish some connection with the instrumental parts, first on a theoretical level, and then tested in several sessions, using recordings of the composed fragments. The fact that the same people were authors both of the digital processing environments and of the instrumental gestures has been a good chance to experience the real *opportunities* of a complex system as a networked laptop group may be.



Figure 6: combination of two distinct musical *gestures* (courtesy of Editor **ArsPublica**, Carrara, 2009).

The score for the laptop players was conceived to exploit different configurations of the whole ensemble. As can be seen from the score excerpt (see Figure 7) the ensemble was divided in two *choruses* (**LTR** and **LTD**, see Chapter 4); the resulting combinations exploited during performance are the following:

- string quartet alone;
- string quartet and **LTR** / **LTD** / (**LTR**+**LTD**);
- **LTD** / (**LTR**+**LTD**) without string quartet.

Possible configurations of spatial distribution were designed (and managed by the two *servers*, one for each section): the sound sources were diffused with different speeds and routes and/or distributed to the eight speakers

by means of a matrix. The *space* parameter was used to *distribute* the *polyphony* which reached a maximum of eight voices. Constant care was taken not to overwhelm the string quartet; this attention was important for us given our intention to obtain the most balanced sound in a sort of an “extended chamber music”.

Quartetto		D Schegge B → Glifi 7,11,2 → Schegge C	INTERLUDIO 3	E Exit → Inclinato	CODA
Laptop tempo reale	IN	Quartetto + LTD	LTD	Quartetto	LTD
	OUT	Alterna 4 elaborazioni	Sovrappone 4 elaborazioni	Alterna 4 elaborazioni	Sovrappone 4 elaborazioni
	SPACE	Circolare Orario lenta	1-4, 3-6 5-8, 7-2	Circolare orario veloce	1,3,5,7
Laptop tempo differito	IN	Quartetto + LTR	LTR	Quartetto	NO
	OUT	Alterna 4 elaborazioni	Alterna 4 elaborazioni	Sovrappone 4 elaborazioni	Sovrappone 4 elaborazioni
	SPACE	Circolare Antiorario veloce	Circolare Antiorario lenta	1-3, 2-4, 5-7, 6-8	2,4,6,8

Figure 7: excerpt of the laptop ensemble score.

Working procedures	
1	Write instrumental pieces/design sound processing
2	Write electronic score for laptop ensemble
3	Assemble instrumental pieces and combine them with the sound processing
4	Roles and relationships between sections of the laptop ensemble
5	Organization and management of space (for polyphonic distribution)

Table 1: general compositional procedures.

4. DEFINITIVE SET-UP

As has been touched upon in Chapter 3, the whole ensemble was finally divided in two *choruses* (see Figure 8):

- **LTR**: real time *chorus*, used audio communication through dedicated interfaces;
- **LTD**: deferred time *chorus*; described in Chapter 2.

This subdivision was made mainly to employ a parallel project developed by Paolo Zavagna and Luca Richelli. The two sections referred to distinct compositional paradigms as well as used different technologies. The **LTD** implements a hierarchical structure based on constructivist assumptions while the **LTR** was a *flat* environment inspired by improvisation groups [2] in which mappings may be made of (possibly) perceptually unrelated control signals.

The real-time chorus (**LTR**) used audio communication and a *global* matrix to diffuse *clients* audio signals through the P.A. system. Each of the two sections (each composed by four performers) depended on a conductor who managed its *global* features (levels, output routing, sound movements, etc.; moreover the two conductors even *played* their own sound processing

instruments, i.e. *clients*) and the synchronization with the string quartet; the two sections could even communicate their audio stream each other.

In the **LTD** some control signals were extracted from the audio streams and used to modulate perceptually related parameters [4]: for example, the energy envelope extracted from the quartet was related to the resonance of the synthesis filters of the *Vocoder* and to the grain duration of the granular processing (the stronger the signal the shorter the grains); the spatialization was conceived and made using *Ambisonics* modules [10] and several sound movements were planned, whose speed was in part related to the energy of the audio stream. This control signals were extracted by the *server* and broadcast to the **LDT** using the OSC protocol [16].

An excerpt of the performance of *.WAV* held at S. Giorgio Island during the first performance of the ensemble is published in a DVD¹.

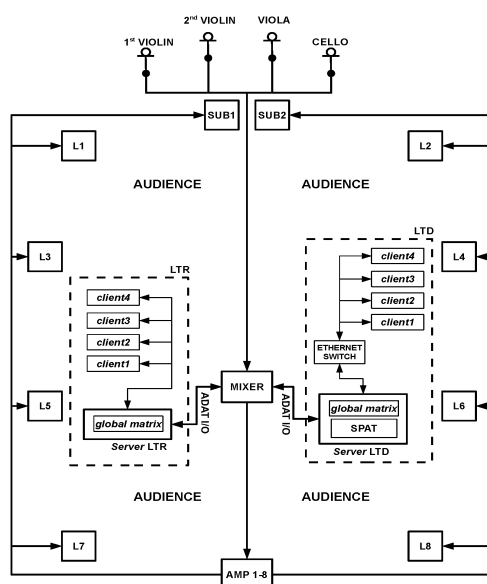


Figure 8: final set-up of the S. Giorgio laptop ensemble.

5. FINAL REMARKS

We would like to thank the Foundation "Giorgio Cini" of Venice and particularly the Director of the Institute of Music prof. Giovanni Morelli that allowed us and the other musicians, in that annual workshop held in the Foundation halls, to work for the "new musical practice" required by this new *instrumental* entity that is the Laptop Orchestra.

The work of the *S. Giorgio Laptop Ensemble* has reflected the multiplicity of attitudes toward music developed in what has been more than just an Electronic

Music Class, that held by Alvise Vidolin in Venice Conservatory from 1975 to 2009. It has been more a school of musical thought, grown exactly where the aesthetic constraints were less rigid. Fundamental and binding was instead the concept of *tèchne*, never as in these days debased and deprived of its true meaning and used in a mechanical acceptance to mean at most an attitude toward the solution of practical problems, where it should be used instead to indicate the complete awareness of instrumental, sensible and intellectual means, needed eventually to pose this same means in *chrisis*. The continuous research of this awareness seems to be the prominent lesson given by Vidolin to his students.

6. REFERENCES

- [1] Brown, C., Bischof, J. "Indigenous to the Net: Early Network Music Bands in the San Francisco Bay Area", 2002, available at <http://crossfade.walkerart.org/brownbischoff/IndigenoustotheNetPrint.html> [30.11.2010].
- [2] Burns, C., Surges, G. "NRCI: Software Tools for Laptop Ensemble", *Proceedings of the 2008 International Computer Music Conference*, Belfast, UK, 2008.
- [3] Chafe, C., Wilson, S., Leistikow, R., Chisholm, D., Scalone, G. "A simplified approach to high quality music and sound over IP", *Proceedings of the Conference on Digital Audio Effects (DAFx-00)*, Verona, Italy, 2000.
- [4] Di Scipio, A. "Sound is the interface: from interactive to ecosystemic signal processing", *Organised Sound*, 8(3), 2003, pp. 269-277.
- [5] Di Scipio, A. "Real-Time Polyphonic Time-shifting of Sound with Interactive Systems", *Proceedings of the XI° Colloquio di Informatica Musicale (XI CIM)*, Bologna, Italy, 1995.
- [6] Gasperini, M. 60 dB. "La scuola veneziana di musica elettronica. Omaggio ad Alvise Vidolin", *Finnegans, percorsi culturali*, n.17, 2009, pp. 57-60.
- [7] Gasperini, M., "Theoretical foundations of a laptop orchestra", *Proceedings of the XVIII° Colloquio di Informatica Musicale (XVIII° CIM)*, Torino, Italy, 2010.
- [8] Puckette, M. "Pure Data", *Proceedings of the International Computer Music Conference*, San Francisco, USA, 1996.
- [9] Sæbø, A., Svennson, U. P. "A low-latency Full-Duplex Audio Over IP Streamer", *Proceedings of the IVth Linux Audio Conference*, Karlsruhe, Germany, 2006.
- [10] Schacher, J. C., Kocher, P. "Ambisonics Spatialization Tools for Max/MSP", *Proceedings of the International Computer Music Conference*, New Orleans, USA, 2006.

¹ "Laboratorioarazzi 2009", Istituto per la Musica, Fondazione "Giorgio Cini", Embtool 2010_cat.0906.

- [11] Smallwood, S., Trueman, D., Cook, P. R., Wang, G. "Composing for Laptop Orchestra", *Computer Music Journal*, 32(1), 2008, pp. 9-25.
- [12] Stockhausen, K. "Electronic music for *Kathinka's Gesang as Lucifer's Requiem*", *Perspectives of New Music*, 23(2), 1985.
- [13] Surges, G., Burns, C. "Networking Infrastructure for Collaborative Laptop Improvisation", *SPARK Festival Proceedings*, Minneapolis-St.Paul, USA, 2008.
- [14] Vidolin, A., Doati, R. (editors) *Nuova Atlantide: il continente della musica elettronica 1900-1986*, La Biennale di Venezia, Venezia, 1986.
- [15] Wang, G., Bryan, N., Oh, J., Hamilton, R. "Stanford Laptop Orchestra (SLORK)", *Proceedings of the 2009 International Computer Music Conference*, Montreal, Canada, 2009.
- [16] Wright, M. "Open Sound Control: an enabling technology for musical networking", *Organized Sound*, 10(3), 2005, pp. 193-200.
- [17] Xu, A., Woszczyk, W., Settel, Z., Pennycook, B., Row, R., Galanter, P., Bary, J., Martin, G., Corey, J., Cooperstock, J. R., "Real Time Streaming of Multi-channel Audio Data over Internet", *Journal of the Audio Engineering Society*, 48(7-8), 2000.

SINTESI SONORA E GRAFICA FRATTALE

Claudio Dionisi

Claude24_01@hotmail.com

Elia Carboni

Ely24_01@hotmail.it

ABSTRACT

Il lavoro presentato è parte di un progetto di ricerca intrapreso nel corso di “Musica e Nuove Tecnologie” del conservatorio “Luigi Canepa” di Sassari a partire dall’A.A. 2008/2009, con l’obiettivo di sperimentare l’impiego di processi ed algoritmi propri della matematica frattale nell’ambito elettroacustico. Nel corso di tali sperimentazioni, è stato sviluppato un sistema di sintesi additiva in cui vengono utilizzati processi di iterazione di polinomi complessi, comunemente utilizzati nelle applicazioni grafiche dei frattali, per la costruzione di forme d’onda. In occasione de “La Terra Fertile 2010” è stata presentata una prima implementazione software di tale sistema, realizzata tramite Max/MSP 4.5. In questo testo si presenta quindi un ulteriore sviluppo del progetto iniziale.

1. INTRODUZIONE

Ciò che segue è da considerare come parte di un nostro progetto di ricerca e sperimentazione, iniziato durante l’A.A. 2008/2009 nei corsi di Elettroacustica, DSP e Laboratorio di Elaborazione Numerica del Segnale, appartenenti al corso di Musica e Nuove Tecnologie del Conservatorio “Luigi Canepa” di Sassari. Allo stato attuale, il nostro lavoro è incentrato sullo sviluppo di una fra le molte possibili metodologie di sintesi del suono correlate con i processi ed i concetti della matematica *frattale*. Questo sistema, basato sulla sintesi additiva, prevede la creazione di un materiale sonoro formato da un numero variabile di parziali, i cui parametri di frequenza e fase e le cui variazioni di ampiezza nel tempo siano poste in relazione fra loro da una determinata funzione frattale.

1.1 Brevi Cenni sui Frattali

Il termine *frattale* (dal latino fractus, frazionario) è stato introdotto nel 1975 dal matematico Benoît B. Mandelbrot nel testo *Les objects fractals*. L’autore, nel corso delle sue molteplici ricerche, ha operato un’elaborata unificazione di elementi tratti da altre discipline, sia matematiche che fisico-naturali, inerenti lo studio di fenomeni di natura principalmente caotica.

Copyright: © 2010 Last name et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Dalla riunificazione di tali teorie, originariamente non correlate fra loro, è nata quella che oggi è conosciuta come una nuova “*geometria della natura*”.

Con il termine *frattale* si intende in effetti una particolare tipologia di oggetti geometrici, caratterizzati da strutture in cui la forma principale si ripete su scale di grandezza progressivamente inferiore, secondo quella che viene definita proprietà di *autosimilarità*. A differenza delle figure e degli oggetti propri della geometria classica, i *frattali* seguono dei particolari processi di costruzione: ad una semplice figura di partenza viene applicata una determinata trasformazione che ne aumenta il grado globale di complessità. La nuova figura così ottenuta viene sottoposta nuovamente a tale trasformazione, mediante un processo di tipo *iterativo*, che viene riproposto, in termini teorici, all’infinito.

Questo nuovo modello matematico possiede dunque un’elevata capacità di rappresentazione delle forme e dei fenomeni comunemente osservabili in natura. Strutture geometriche complesse, che rassomigliano ad oggetti del mondo fisico, sono quindi realizzabili attraverso degli algoritmi ricorsivi governati da poche e semplici regole. Questa caratteristica ci ha spinto ad utilizzare un principio di natura *iterativa* e *frattale*, come criterio generativo di materiali sonori.

1.2 Iterazioni di Polinomi Complessi e Grafica Frattale

Un altro metodo per creare strutture frattali consiste nell’applicare principi iterativi a funzioni di polinomi complessi. Considerando ad esempio la legge ricorsiva $Z_{(n+1)} = Z_{(n)}^2 + C$, dove Z e C sono numeri complessi formati da parte reale e parte immaginaria, possiamo osservare il comportamento della funzione su un dato intervallo del piano complesso: sostituendo infatti a C le coordinate di un punto nel piano, e ponendo $Z_{(0)} = 0$, possiamo verificare se, dopo un determinato numero di iterazioni, la funzione diverge, superando una soglia limite prestabilita detta valore di *bailout*. Ripetendo il procedimento per tutti i punti del piano complesso definiamo, in questo caso particolare, il noto “*insieme di Mandelbrot*”, a cui appartengono tutti quei punti per i quali la funzione non diverge, restando all’interno del limite assegnato.

Come si può intuire, le funzioni complesse divergono, per i punti esterni ad un determinato insieme frattale, dopo un quantitativo di iterazioni variabile. Sulla base di questo principio si è sviluppata una metodologia di disegno grafico computerizzato, oggi largamente diffusa, che

sfrutta il quantitativo di iterazioni compiute dalla funzione per assegnare dei valori cromatici ai vari punti corrispondenti sul piano complesso, identificati in ambito software come dei pixel, che compongono quindi sullo schermo un'immagine della struttura *frattale* presa in esame.

2 SINTESI ADDITIVA ATTRAVERSO ALGORITMI FRATTALI

L'obiettivo principale del processo di sintesi qui esposto è quello di creare strette connessioni fra i parametri più importanti dei modelli di sintesi additiva e i principi, precedentemente esposti, di creazione di strutture frattali, mediante l'impiego di funzioni polinomiali complesse. E' infatti nostra intenzione tentare di utilizzare le potenzialità strutturali di questi processi al fine di realizzare uno strumento capace di creare, con un ristretto numero di istruzioni, delle strutture sonore con alto grado di complessità, in maniera analoga ai meccanismi della creazione grafica frattale.

Il nucleo centrale dell'algoritmo è formato da un blocco di calcolo nel quale vengono elaborate le funzioni iterative, elemento principale del processo, ed ospita, come default, la funzione generatrice dell'*insieme di Mandelbrot* $Z_{(n+1)} = Z_{(n)}^2 + C$.

$$\begin{aligned} Z_{(n+1)} &= Z_{(n)}^2 + C & Z_{(1)} &= Z_{(0)}^2 + C \rightarrow Z_{(1)} = C \\ & & Z_{(2)} &= Z_{(1)}^2 + C \rightarrow Z_{(2)} = C^2 + C \\ \text{con } Z_0 &= 0 & Z_{(3)} &= Z_{(2)}^2 + C \rightarrow Z_{(3)} = (C^2 + C)^2 + C \end{aligned}$$

Figura 1. Grafico esplicativo dell'insieme di Mandelbrot.

Ciascun punto del piano complesso per cui la formula viene verificata è considerato una parziale del generico materiale sonoro che si vuole sintetizzare e pertanto lo sviluppo della funzione fornisce dati relativi a *frequenza*, *fase* ed *inviluppo*, necessari a caratterizzare ciascuna componente sonora. Le informazioni così ottenute regolano la creazione di differenti onde sinusoidali, che verranno quindi scritte in buffer di memoria di grandezza definibile, e sommate fra loro mediante un processo ricorsivo per formare la risultante forma d'onda complessa finale.

2.1 Componenti Frequenziali e Punti sul Piano

Il criterio di attribuzione dei valori di frequenza delle parziali si basa sull'associazione dei concetti di spettro sonoro e di piano complesso: se nella grafica frattale i vari punti del piano sono considerati pixel sullo schermo, in questo processo di sintesi si attribuisce a ciascuno di essi un singolo e univoco valore frequenziale, che pertanto risulta stabilito a priori ed indipendente dallo sviluppo della funzione frattale. Durante lo sviluppo della formula, l'algoritmo seleziona un determinato numero di punti della regione del piano presa in esame, variando il valore della variabile complessa C con differenti *step* incre-

mentali, rendendo possibile la creazione di suoni con un numero variabile di parziali. Gli effettivi valori di frequenza vengono stabiliti sulla base della seguente relazione:

$$F_{(p)} = M + (A \times P) \quad (1)$$

con:

$$M = \left(|C_{(p \text{ Max})}| - |C_{(p)}| \right) \times R \quad (2)$$

e con :

$$\bullet \quad A = \arctan\left(\frac{\text{Im}\{C_{(p)}\}}{\text{Re}\{C_{(p)}\}}\right)$$

se $\text{Re}\{C_{(p)}\} > 0$ e $\text{Im}\{C_{(p)}\} \geq 0$

$$\bullet \quad A = \arctan\left(\frac{\text{Im}\{C_{(p)}\}}{\text{Re}\{C_{(p)}\}}\right) + 2\pi$$

se $\text{Re}\{C_{(p)}\} > 0$ e $\text{Im}\{C_{(p)}\} < 0$

$$\bullet \quad A = \arctan\left(\frac{\text{Im}\{C_{(p)}\}}{\text{Re}\{C_{(p)}\}}\right) + \pi$$

se $\text{Re}\{C_{(p)}\} < 0$

$$\bullet \quad A = \frac{\pi}{2}$$

se $\text{Re}\{C_{(p)}\} = 0$ e $\text{Im}\{C_{(p)}\} > 0$

$$\bullet \quad A = \frac{3}{2}\pi$$

se $\text{Re}\{C_{(p)}\} = 0$ e $\text{Im}\{C_{(p)}\} < 0$

(3)

dove:

- se $C_{(p)} = 0$, allora $F_{(p)} = 0$
- $C_{(p)}$ è la variabile corrispondente alle coordinate del punto preso in esame
- $C_{(p \text{ Max})}$ è il punto più lontano dall'origine nella regione selezionata
- R è un valore costante definito in partenza che altera il range frequenziale in modo da ottenere spettri più o meno compressi
- P è un moltiplicatore del valore di offset relativo all'angolo del punto esaminato.

2.4 Algoritmo Iterativo nella Formazione dei Materiali

Per ciascun punto del piano complesso il nucleo centrale dell'algoritmo compone quindi due buffer: il primo incamera una sinusoide con un proprio valore di frequenza e fase, mentre il secondo memorizza l'involuppo ad essa relativo. L'effettiva composizione dei materiali sonori avviene dunque all'interno di un terzo blocco dell'algoritmo: come primo passaggio, la sinusoide e l'involuppo relativi ad un punto $C_{(p)}$ vengono moltiplicati fra loro campione per campione e memorizzati su un terzo buffer. A questo punto si attiva un processo ricorsivo in cui la parziale appena creata viene sommata ad un quarto buffer che conserva tutte le somme svolte in precedenza secondo il medesimo processo per ogni altro punto del piano. Al termine della definizione dell'insieme frattale, in questo caso l'insieme di Mandelbrot, si avrà dunque una forma d'onda complessa di durata proporzionale alla dimensione impostata per i buffer di memoria. Dopo un opportuno riscaldamento nei valori di ampiezza, l'algoritmo avrà quindi prodotto un materiale sonoro sintetizzato sulla base della funzione frattale impiegata.

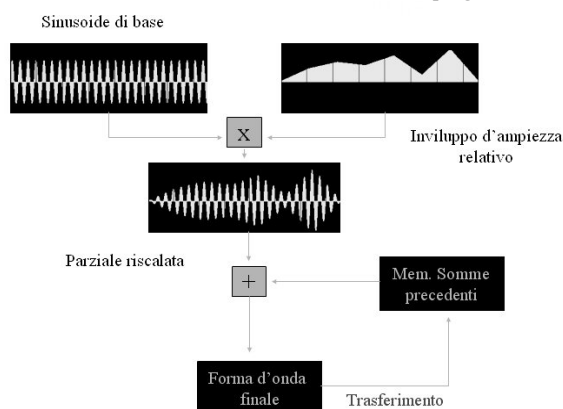


Figura 6. In questo grafico è visualizzato il processo iterativo utilizzato nella costruzione della forma d'onda finale risultante dalla somma delle sinusoidi parziali.

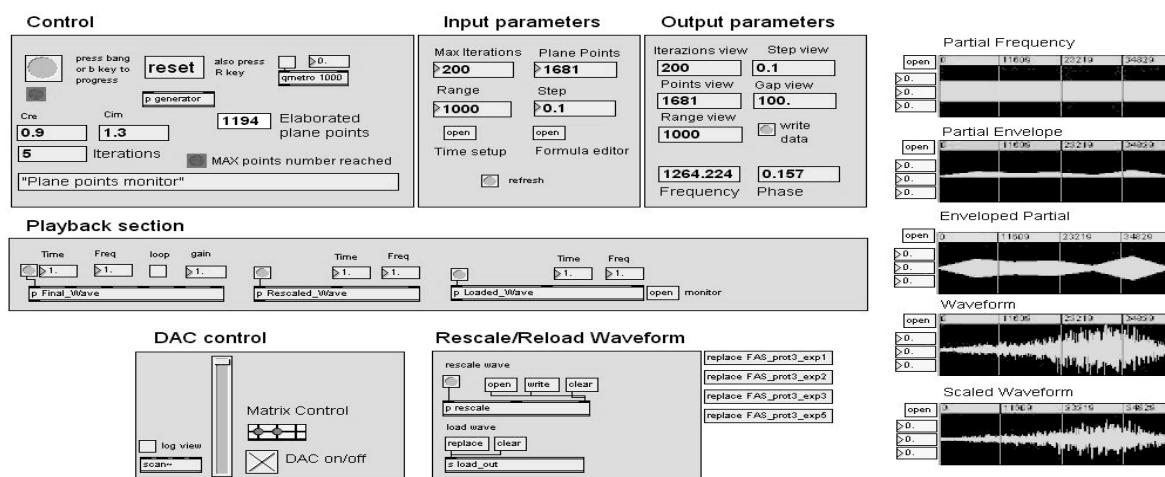
3 IMPLEMENTAZIONE SU MAX/MSP

Una prima implementazione del processo di sintesi presentato è stata realizzata con l'ausilio del software Max/MSP 4.5. Sulla base dei procedimenti illustrati abbiamo realizzato una patch che gestisce le varie fasi dell'algoritmo, monitorizzandole tramite una interfaccia grafica che permette all'utente di controllare i vari parametri in ingresso. Il nucleo centrale, principalmente utilizzato per il calcolo iterativo della funzione frattale, è stato programmato in *javascript* utilizzando l'oggetto *js* di Max.

4 REFERENCES

- [1] B. Mandelbrot: *Les Objects Fractals*, Flammarion, 1975
- [2] [Http:// ciram.unibo.it/](http://ciram.unibo.it/)
- [3] [Http://fractals.iut.u-bordeaux1.fr/](http://fractals.iut.u-bordeaux1.fr/)
- [4] [Http:// math.yale.edu/Mandelbrot/](http://math.yale.edu/Mandelbrot/)

Figura 7. L'interfaccia grafica dell'implementazione su Max/MSP 4.5. Sulla destra sono visualizzati i passaggi di formazione del materiale sonoro per ogni singola parziale generata dai punti sul piano complesso. I controlli relativi alla gestione della funzione frattale sono collocati nella parte centrale, nella sezione Input parameters.



SIMULATION OF A SOUNDSCAPE USING AN ANALYSIS/RESYNTHESIS MODEL

Amedeo Casella
Università degli studi di Torino
elblancohomicida
@gmail.com

Andrea Valle
CIRMA
Università degli studi di Torino
andrea.valle@unito.it

ABSTRACT

This paper presents the result of a simulation of an indoor soundscape, the multimedia laboratory “Guido Quazza” of the University of Torino. The simulation is carried out by using the *GeoGraphy* application that allows interactive, real-time exploration by the user.

1. INTRODUCTION

The *GeoGraphy* project aims at providing a theoretical model of soundscape simulation and at implementing it into a real-time interactive application. It is based on phenomenological and semiotic assumptions, and focuses both on perceptual and cultural features of soundscape. The model is based both on classic soundscape studies [1] and on the notion of sound object [2]. It shares with [3] an approach to sound synthesis that mixes/integrates perceptual, physical and cultural aspects. The *GeoGraphy* framework includes four aspects:

- a) representation of the space to which the soundscape is related;
- b) an audio database including each component of the soundscape;
- c) a generative algorithm for the soundscape synthesis using gathered data from audio material, space and listener path;
- d) criteria for the qualitative evaluation of the result.

To summarize, the simulation model is based on a two-step process, leading to an analysis-based resynthesis of the soundscape.

For a discussion on soundscape models and an in-depth introduction to *GeoGraphy* model and application see [4,5,6]

2. CASE-STUDY

The “Guido Quazza” multimedia laboratory is a facility dedicated to teaching and research of the “Scienze della Formazione” faculty in the University of Torino. It includes three production departments:

- Multimedia production and e-learning
- Audio and video productions

Copyright: © 2010 Casella et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

- Hardware and software systems

Since 2002 offices and labs are collocated partly below ground level, where a technologic infrastructure of broadcaster level was built, for audiovisual and multimedia production.

In order to provide services and audiovisual productions activities, G. Quazza lab offers a permanent support to teaching activities on movies, television and multimedia topics, in particular related to the Extracampus project, the television channel of the University of Torino; it provides multimedia solutions for event documentation (eg. remote and live streaming, to be uploaded on the Extracampus web site); it also produces documentaries that focuses on research activities; finally, it produces videos for the e-learning platform of the University.

A map of the lab is represented in Figure 1.

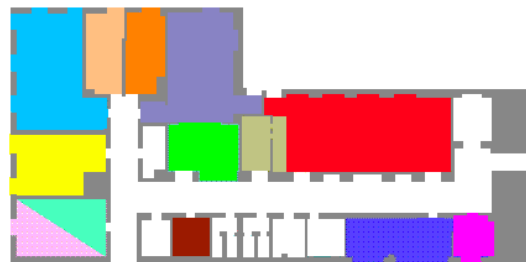


Figure 1. Planimetry of the laboratory “Guido Quazza”.

2.1 Analysis

In the analysis phase we learnt about the case study environment, gathering information about cultural and technical aspects, both by asking to the laboratory staff and observing activities taking place there.

The whole area was explored without using a particular listening strategy but simply walking and scouring every room, so that it was possible to evaluate the general acoustic composition of the laboratory.

The most relevant feature that emerged from the analysis, was the noisy presence of the ventilation pipe, caused by the lab underground collocation; the sound of the pipe is the the keynote sound of the whole environment. In every room there is at least one ventilation pipe. Being an indoor environment, the laboratory is totally decontextualized from the surrounding, as it is substantially not possible hear sounds coming from the outside; being an environment dedicated to teaching and research it is very

quiet and it lacks intense sounds, sometimes resulting almost silent. This has lead during the second in-depth listening phase to focus on low-intensity sources like computer fans. The goal of this phase was the creation of a sound map: hence, space was analyzed in a more meticulous way, taking into account all the existing sound objects.

In relation to Figure 2, the global environment was divided into six macroareas. Macroareas are acoustically homogeneous sub-spaces. Rooms belonging to the same macroarea share the same sound atmosphere. Six macroareas were identified (see Figure 2).

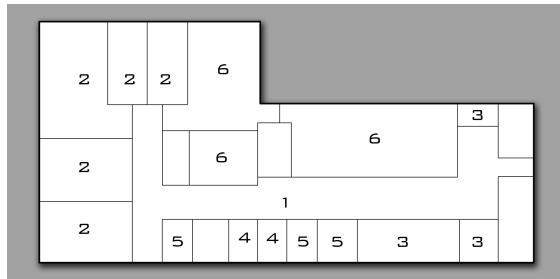


Figure 2. Macroareas distribution.

Macroareas are not strictly separated one from each other, as their boundaries overlap.

The macroareas are:

1. Corridor: connects all the rooms, the most crowded and acoustically rich area
2. Laboratories: rooms where computers sounds is predominant
3. Secretariat and offices: rooms where administrating functions are carried out
4. Restrooms: here, the soundscape is very quiet.
5. Power plant, heating plant: the most complex and noisy sources of the whole laboratory
6. Audiovisual: rooms dedicated to audiovisual projects, including the Auditorium, the soundscape includes audio materials from the production

Every room was considered as part of a room ensemble forming a macroarea; for each macroarea different kinds of recording were planned. Omnidirectional microphones were used for capturing sounding objects recognized as atmospheres, and cardioid ones to focus on a source in an clear-cut way, avoiding undesired background noise. Audio samples were recorded with a DAT recorder, with sample rate at 44.100 (CD quality)

A specific issue emerged while trying to represent such a quiet and sparsely populated place: the human voices of the lab staff were a relevant feature, as they are fully integrated into the soundscape. Thus, during the recording phase, human voices were recorded too (of course, without infringing privacy, but indeed privacy could be a relevant issue).

To summarize the soundscape of each room is the sum of hardware sounds and staff voices.

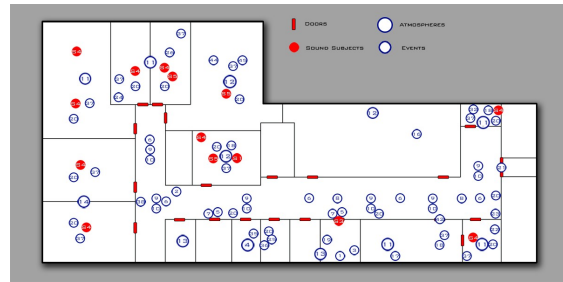


Figure 3. Sound objects in the lab.

Figure 3 shows all the sound objects (organized in atmospheres, events and sound subjects, see [6]) resulting from the analysis and implemented in the resynthesis. Sound subjects are properly subgraphs, but are represented as single graphical units. Edges are omitted for sake of clarity.

2.2 Resynthesis

In order to simulate the soundscape, the resynthesis step of the soundscape is based on data gathered during the analysis step. In GeoGraphy, sequences of sound objects (collected in the database) are generated by means of graphs defining all the possible sequencing relations, and specifying temporal details (onset times and durations). As a consequence, graph design (topology and metrics) is a crucial aspect in the whole simulation process (see [5, 6]).

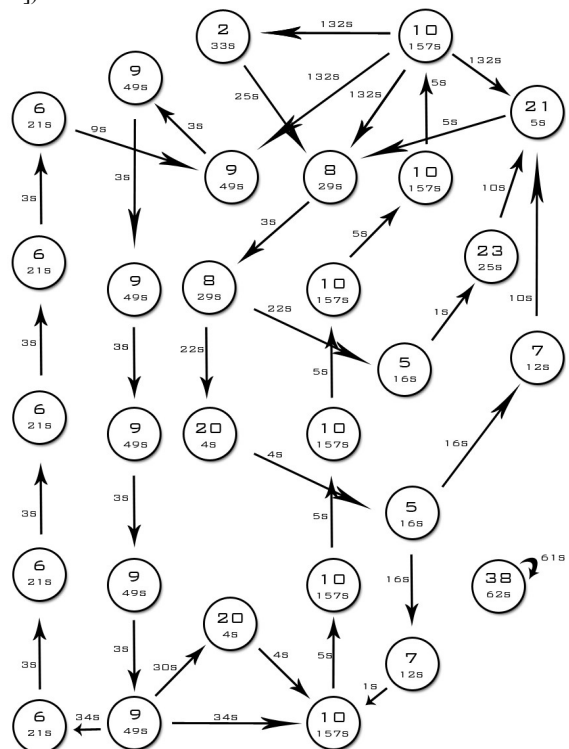


Figure 4. The corridor graph.

Thus, in the following we will not discuss sound synthesis but we will focus instead on graph design.

Each room was represented by a disconnected, circuit-shaped, sub-graph. The overall soundscape of the laboratory is thus a graph composed by disconnected sub-graphs, one for each room.

A particular case is the corridor (Figure 4). During the analysis phase, it has been possible to observe that many sound objects were repeated. As a consequence, the corridor graph is designed so that each time a vertex is activated, all its copies (identified by the same numerical ID) are turned on after him, at a specific time interval. In this way, each vertex triggers the beginning of a sequence of identical samples, which correspond to vertices with radiation area slightly greater than the width of the corridor.

A person walking at normal speed takes about 10-15 seconds to travel the entire length of the corridor. In the graph, five sound objects are connected to each other and placed in the map along the corridors, representing two people talking while walking. After the first sound object has started, every three seconds the next one in the sequence is activated. In this way it is possible to simulate the sound of footsteps along the corridor, without affecting the soundscape of the other rooms, and the listener can hear a virtual walk.

Another graph was created to represent a very peculiar sounding body, a Betacam workstation (Figure 5).

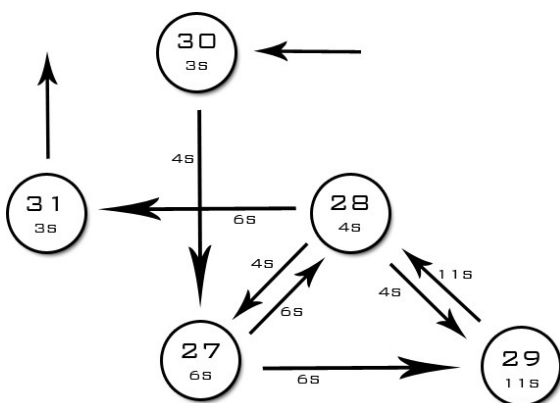


Figure 5. The Betacam workstation graph.

The analog editing activity at the Betacam features a continuous passage between magnet sound and tape sound. The typical action begins with the tape insertion in the video recorder, ending with its ejection. In the middle, different actions can happen (rewind, fast forward, play) in no specific order. Hence, the process needs to be described neither in terms of a long background sound (atmosphere) nor in terms of a single sound object (event), but, rather, as a complex subgraph. The subgraph is composed of five sound objects: tape input and output sounds from video recorder (30 and 31), magnet sounds and tape rewind/forward (27, 28 and 29). The graph is connected to a larger graph (in fact, it is a sub-graph) through vertex 31 (the disconnected arrow is linked to the rest of the graph). Each vertex is associated with an audio sample representing typical magnet sounds of the *Betacam*. The

subgraph, to be heard only in the analog editing room, represents a grammar of sounding actions that are usually performed on the Betacam station, that include variable repetitions of subsequences (one can consider the vertices 27-28-29). The Betacam sound subject is an acyclical graph, and the Betacam simulation ends with vertex 31, with the actant moving to activate other sound objects of the room graph. On the other side, the subgraph 27-28-29 is cyclical, allowing to create editing action sequences of different –theoretically infinite– length. Thus, it models a sound producing activity. Figure 6 shows a screenshot from the application implementing the graph of Figure 5.

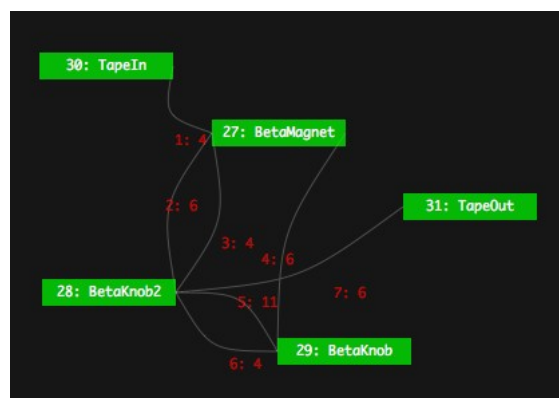


Figure 6. A screenshot from the GeoGraphy application with the Betacam graph.

To conclude, Figure 7 shows a screenshot of the application GUIs, used to simulate the whole laboratory soundscape.

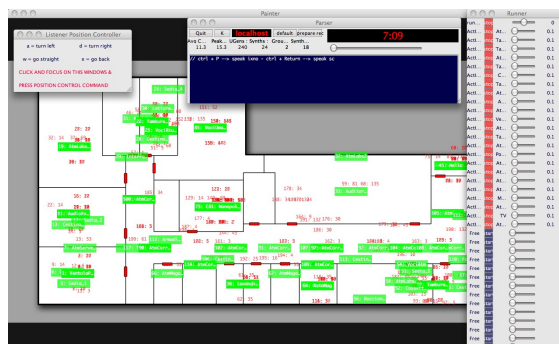


Figure 7. A screenshot from the GeoGraphy application implementing the simulation of the laboratory

3. CONCLUSIONS

GeoGraphy has been previously used to simulate outdoor environment, where areas tend to be clearly separated (an outdoor market in [3,4,5], a park in [7]).

On the contrary an indoor environment like the one analyzed in this paper, presents strong acoustic interferences among subspaces. It is still not easy to reproduce such effects through the *GeoGraphy* framework.

In any case informal evaluation tests with students and personal working at the lab have proven the substantial effectiveness of the simulation.

4. REFERENCES

- [1] Murray Schafer, R., *The Tuning of the World*, New York, Knopf, 1977.
- [2] Schaeffer, Pierre, *Traité des objets musicaux*, Paris, Seuil, 1966.
- [3] Rocchesso, D., Fontana, F., eds.: *The Sounding Object. Edizioni di Mondo Estremo*, Firenze, 2003.
- [4] Valle, A., Schirosa, M., and Lombardo, V., *A Framework for Soundscape Analysis And Re-synthesis*, Proceedings of the SMC 2009 - 6th Sound and Music Computing Conference, 2009, pp. 13-18
- [5] Valle, A., Lombardo, V., and Schirosa, M., *A Graph-based System for the Dynamic Generation of Soundscapes*, Proceedings of the 15th International Conference on Auditory Display, 2009, pp. 217- 224.
- [6] Valle, A., Lombardo, V., Schirosa, M., *Simulating the Soundscape through an Analysis/Resynthesis Methodology*, Auditory Display 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009. Revised Papers, Berlin, Springer, pp. 330-357
- [7] Minato, M., *Simulazione di un paesaggio sonoro: il Parco del Valentino*, Final Dissertation, DAMS, Università di Torino, a.a., 2009-10.

THE RINGER – SOFTWARE AUDIO E VIDEO IN TEMPO REALE

Ivano Morrone

Conservatorio di Musica di Cosenza
morronei@alice.it

SOMMARIO

E' un progetto basato sulla possibilità di immaginare una liuteria digitale audio e video in stretta connessione con specifiche prassi compositive e performative musicali e visuali. Nello specifico con: la composizione ed esecuzione di un brano per quartetto di laptop; l'implementazione di forme di creazione visuale nella performance per strumento solo; la messa in opera di una installazione sonora e visuale interattiva.

Nel presente scritto viene trattata la sola parte tecnologica. La descrizione della parte musicale e visuale del progetto è rimandata ad un prossimo documento.

1. OBIETTIVI

Realizzare un piccolo gruppo di applicazioni che in tempo reale siano in grado di:

- gestire il segnale audio modificandolo con processi di RM e FM;
- variare dinamicamente, in modo anche interattivo, i parametri grafici delle immagini in relazione alla gestione e modificazione del segnale audio;
- permettere il controllo dei parametri sonori e visuali anche per mezzo dell'interazione con i movimenti del corpo umano.

Che siano inoltre caratterizzate da:

- interfacce usabili con facilità dal vivo;
- tempi di latenza audio e video bassi e comunque compatibili con la dimensione performativa;
- predisposizione per l'utilizzo con controller MIDI
- predisposizione per l'utilizzo multi-piattaforma;

E, infine, che possano essere adoperate per attivare possibilità di interazione in rete, ad esempio nelle installazioni plurimediali a distanza.

2. STRUMENTI E METODI

La scelta dei linguaggi di programmazione da utilizzare è stata rivolta al C ed al C++, sia per le conoscenze dell'autore, sia per la possibilità di gestire in modo efficiente le diverse librerie software. Le librerie utilizzate sono tutte open-source e multi-piattaforma:

- il controllo dell'audio è stato realizzato con le funzioni di I/O disponibili in Portaudio[1] e l'uso dei driver ASIO;
- il pacchetto di Fast Lite Toolkit[2] ha permesso di implementare le interfacce grafiche;
- la Open Graphics Library[3] è servita per le modificazioni dei parametri grafici delle immagini e per con-

trollarne la forma ed il movimento nello spazio di proiezione con renderizzazione in 2D e 3D;

-le librerie Open Computer Vision[4, 5] hanno permesso di acquisire le immagini in formato jpg, implementare i filtri grafici su di esse applicati, attivare il sistema di tracciamento dei movimenti del corpo umano e la sua interazione con i controlli audio/video per mezzo di una videocamera.

Il lavoro di programmazione è consistito sia nel rimodulare le singole librerie in conseguenza delle specifiche necessità, sia nell'integrarle in un unico ambiente software. I processi in tempo reale hanno riguardato nell'ipotesi più complessa (installazione sonora e visuale) la catena di interazione così raffigurata (Fig. 1).

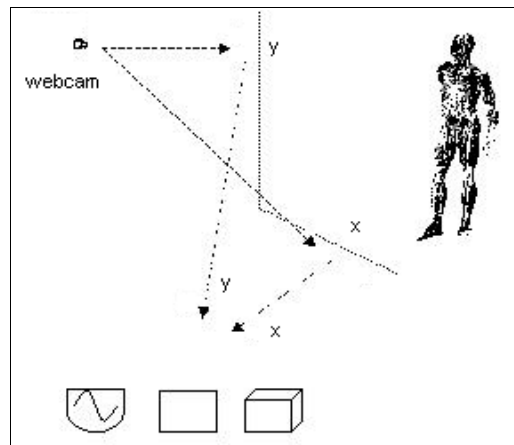


Figura 1. Catena di interazione dei processi in tempo reale.

Il tracciamento dei movimenti del corpo umano realizzato con la webcam determina, adottando adeguati valori di frame rate, una mappatura dinamica dei valori delle coordinate x, y che, normalizzati, sono trasferiti secondo specifici criteri ai parametri degli oscillatori digitali e delle renderizzazioni delle immagini in 2D e/o 3D, permettendo al fruitore dell'installazione sonora e visuale di interagire con quest'ultima in funzione anche della distanza del suo punto di osservazione (valore di z) da essa.

3. APPLICAZIONI REALIZZATE

3.1 The Ringer

E' basato sulla possibilità di modulare il segnale audio in ingresso con processi di modulazione ad anello e di caratterizzarne gli esiti attraverso forme di modulazione di frequenza. Gli oscillatori modulanti possono essere gestiti manualmente ed in parte con controlli che ne attivano l'automazione. Possono essere utilizzate quattro forme d'onda: sinusoidale, triangolare, a dente di sega e quadrata.

Al momento sono state sviluppate due versioni a due e quattro canali (Fig. 2).

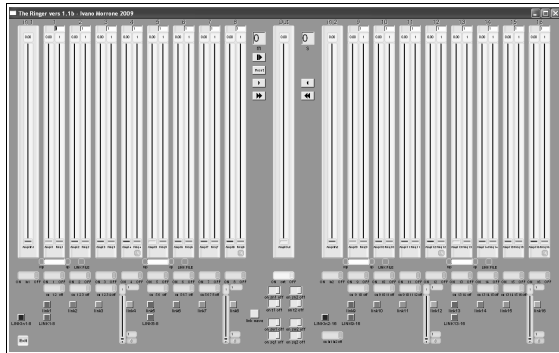


Figura 2. The Ringer versione a due canali, finestra principale

3.2 The RingerV

Il programma ha la possibilità di gestire e visualizzare immagini in formato jpg e realizzazioni grafiche in OpenGL. E' dotato di controlli manuali per la gestione dei canali di colore RGB e alfa. Nella consolle di controllo (Fig. 3) un peak meter mostra l'ampiezza del segnale audio stereo in ingresso che agisce secondo una certa gradualità su alcuni parametri di renderizzazione delle immagini: una parte di quelli relativi alla forma e movimento di queste ultime nel loro spazio di proiezione (Fig. 4). Altri controlli manuali potenziano l'interazione automatica tra segnale audio in ingresso e modificazioni delle rappresentazioni grafiche. Il programma è in grado di essere gestito utilizzando più monitor per cui la consolle dei controlli video può essere assegnata ad un display e le visualizzazioni, a schermo intero, ad un altro.

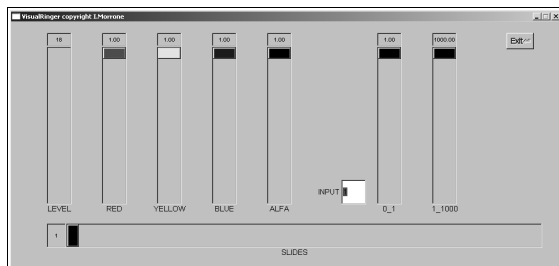


Figura 3. The RingerV, finestra principale

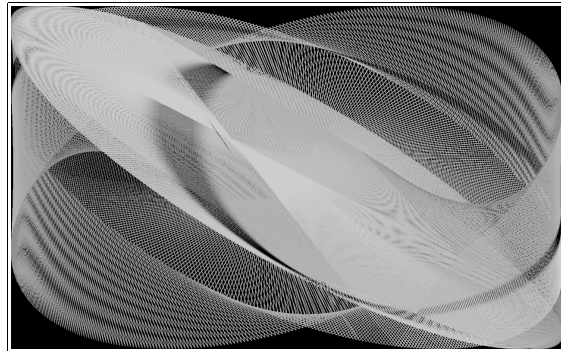


Figura 4. Still estratto dal flusso video in 3D realizzato con The RingerV

3.3 The RingerVI

Il software coniuga e coordina una serie di funzioni: l'acquisizione audio e sue modificazioni; la gestione e trasformazione degli elementi visuali; forme di interazione dei primi due aspetti con il movimento del corpo umano.

La consolle principale (Fig. 5) permette di gestire, come nel caso di *The RingerV*, i canali RGB e alfa delle immagini. Altri controlli automatici e manuali intervengono nelle modificazioni grafiche implementate. Il peak meter presente viene questa volta adoperato per monitorare le differenze di ampiezza di un segnale sinusoidale che permette di organizzare nel tempo la successione automatica delle diverse immagini.

Sono presenti i controlli per gestire l'audio in ingresso e modularlo in tempo reale con gli stessi processi già illustrati in *The Ringer*. Un altro controllo attiva la possibilità di automatizzare in modo inversamente proporzionale, anche in relazione al movimento del corpo umano, il rapporto di ampiezza tra segnale audio originale e segnale modulato.

Un'altra finestra mostra la lista delle immagini jpg acquisite ed una serie di controlli di filtri grafici che possono essere attivati sia in forma automatica, dallo spostamento del corpo e sia manuale, dall'operatore alla consolle.

Infine una terza schermata mostra lo spazio di ripresa di una videocamera (anche una semplice "webcamera" usb). Un sistema di tracciamento del movimento dei corpi nello spazio invia una serie di dati ai controlli audio e video che in tempo reale interagiscono con il suono diffuso e le immagini renderizzate e proiettate (Fig. 6, 7), modificandoli secondo le prassi accennate.

Anche in questo caso, come in *The RingerV*, le tre diverse consolle di controllo possono essere assegnate ad un monitor e le visualizzazioni, a schermo intero, ad un altro.

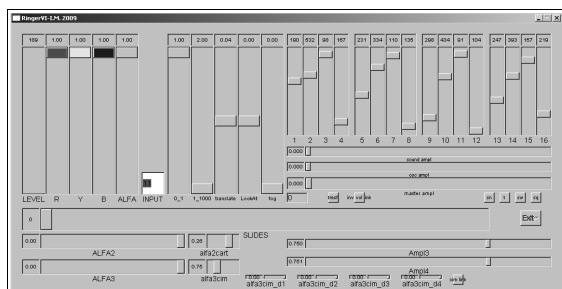


Figura 5. The RingerVI, finestra principale



Figura 6. Still estratto dal flusso video in 2D realizzato con The RingerVI



Figura 7. Still estratto dal flusso video in 2D realizzato con The RingerVI

4. PROSSIMI SVILUPPI

Sono dedicati a raggiungere gli obiettivi fino ad oggi conseguiti in modo parziale:

- migliorare alcune funzioni e renderle “visibili” all’utilizzatore finale, non solo al programmatore;
- concludere l’elaborazione del codice per il controllo delle applicazioni via protocollo MIDI;
- garantire la portabilità con la piena implementazione multi-piattaforma (attualmente le applicazioni sono attive solo su sistemi Win32 e solo sperimentate su altre);

-concretizzare le forme di interazione in rete con la realizzazione di una installazione sonora e visuale a distanza.

5. CONCLUSIONI

Il lavoro fino ad oggi svolto ha avuto inizio nel 2006. Le tre applicazioni sono state utilizzate dal vivo in più occasioni, in forma di concerto e installazione: Reggia di Caserta 2006, Biblioteca Nazionale di Cosenza 2007-08, Artech 2008, Emufest 2009, La terra fertile 2010, SMC 2010.

L’uso pubblico ha permesso di verificare le peculiarità del progetto, seppur nei suoi limiti, sia in termini di affidabilità tecnologica e sia in termini di strumento utilizzabile dal compositore e dall’esecutore.

6. RIFERIMENTI

- [1] <http://www.portaudio.com>
- [2] <http://www.fltk.org>
- [3] <http://www.opengl.org>
- [4] <http://www.opencv.willowgarage.com/wiki>
- [5] Gary Bradski, Adrian Kaehler: *Learning OpenCV – Computer Vision with the OpenCV Library*, O'Reilly Media, Sebastopol, 2008.

UN TUTORIAL PER LA PRODUZIONE DI OPERE AUDIO-VISUALI TRAMITE AMBIENTI INTEGRATI

Antonio Scarcia
antscar@gmail.com

Francesco Abbrescia
Conservatorio "N.Piccinni", Bari
f.abbrescia@tiscali.it

Francesco Scagliola
fscagliola@libero.it

ABSTRACT

In questo report si presentano le motivazioni di un tutorial in preparazione anche presso i corsi istituzionali di Musica Elettronica del Conservatorio "N.Piccinni" di Bari. Il tutorial sarà finalizzato all'utilizzo di una metodologia di composizione audio-visuale integrata con uso degli ambienti Csound e Processing, della quale si riportano alcuni esempi significativi.

1. INTRODUZIONE E MOTIVAZIONE

Sono state già sviluppate in passato presso il laboratorio della Classe di Musica Elettronica del Conservatorio di Bari - ed in particolare nell'ambito delle attività del gruppo di ricerca Sin[x]Thésis [1], diretto da Francesco Scagliola - esperienze di realizzazione di opere audiovisuali. Fin da subito si è sentita la necessità:

- di individuare e sviluppare metodologie e strumenti coerenti con le prassi compositive già adottate nella realizzazione dei brani elettroacustici in una ottica di trasversalità tra la pratica della composizione musicale e della composizione video;
- documentare e quindi condividere alcuni aspetti di questa esperienza.

2. IDEA DI RICERCA E FINALITA'

La produzione elettroacustica sviluppata nell'ambito di Sin[x]Thésis si caratterizza per l'utilizzo intensivo dell'elaboratore come strumento essenziale di assistenza alla composizione, con particolare riguardo alle possibilità offerte dall'ambiente Mathematica [2].

Lo sviluppo in tempi più recenti di lavori multimediali da parte di componenti del gruppo Sin[x]Thésis, attraverso l'uso integrato di ambienti distinti di sintesi per l'audio ed il video, ha incoraggiato l'elaborazione di una documentazione tutoriale intesa come punto di partenza per lo sviluppo di ulteriori sperimentazioni in ambito audiovisuale.

Un tutorial, di imminente presentazione, è stato concepito per l'integrazione di Processing [3][4], domain specific language basato su Java, con Csound, correntemente in uso per la sintesi audio, realizzando così una possibile integrazione tra sintesi audio e sintesi video già a livello di formalizzazione dei processi compositivi di tipo generativo.

La metodologia di lavoro presentata utilizza un approccio basato sulla sintesi per tempo differito sia audio che video, approccio che risponde all'esigenza di poter sfruttare appieno le potenzialità degli ambienti da utilizzare senza vincoli di tipo prestazionale a livello di sistema. Peraltro, è convinzione degli autori che al momento attuale tale approccio assicuri il totale controllo sui materiali nonché sui procedimenti.

3. METODI E RISULTATI

La metodologia sviluppata è articolata secondo le fasi seguenti:

- generazione algoritmica di file dati per il controllo di parametri per gli ambienti di sintesi; allo stato attuale vengono prodotti file di testo con sintassi compatibili con
- file dati per il tool Timeline, per la sintesi in ambiente Processing;
- score Csound [5];
- sintesi audio in tempo differito con CSound
- sintesi video in tempo differito in ambiente Processing
- operazioni di editing e post-produzione ove ritenute necessarie.

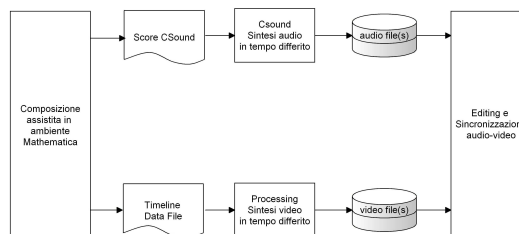


Fig. 1. Metodologia di lavoro

Il tool Timeline, sviluppato come tool grafico da Devon Rifkin [6] si è dimostrato, dopo una serie di esperienze, una soluzione ottimale anche per la gestione dei tempi sulle sequenze video operando in tempo differito.

Il tutorial in preparazione intende porre un'enfasi sulla predisposizione ed utilizzo del file dati per il Tool Timeline, fornendo indicazioni operative per un rapido start-up e nel rispetto di possibili scelte alternative per la generazione algoritmica finalizzata alla generazione dei file dati.

Per quanto concerne la sperimentazione in corso, la fase a) è caratterizzata allo stato dall'utilizzo di una libreria di composizione algoritmica per Mathematica sviluppata da Francesco Scagliola [7]; essa permette al compositore piena discrezionalità nella organizzazione strutturale dell'opera [8], nella scelta delle strutture dati e delle loro relazioni, sia a livello dei parametri della sintesi audio che della sintesi video. Il numero dei parametri generabili e le loro interdipendenze possono essere progettati secondo le necessità espressive.

Dal 2009 ad oggi sono stati realizzati alcuni lavori multimediali che condividono l'adozione di metodologie di composizione assistita con la libreria citata e sintesi audio e video in ambienti Csound e Processing:

- **A UN TEMPO** (2009), di Antonio Scarcia; la sintesi video è stata realizzata a partire da un'opera elettroacustica compiuta, attraverso un approccio ibrido, ovvero con il controllo dei parametri video ricavato sia dai dati delle score Csound che dall'estrazione di features del materiale musicale [9]. La sintesi video fa impiego della libreria MSAFluid di Memo Aitken [10]. Il lavoro, primo tra quelli citati ad essere presentato al pubblico (Mantis Festival 2010, Manchester), prevedeva per il controllo dei parametri video l'uso di codici Processing [11] per la lettura di tabelle di dati, richiedendo particolare attenzione alle problematiche di sincronizzazione;

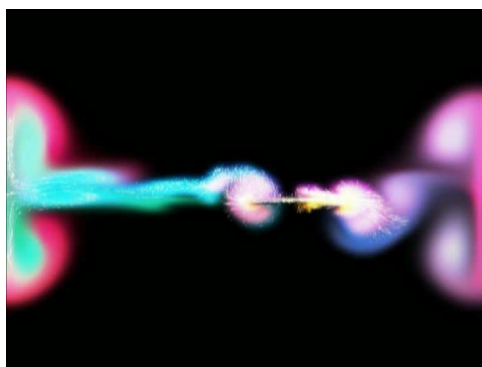


Fig. 2. Una frame da “A un tempo” di A. Scarcia

Copyright: © 2010 Scarcia, Abbrescia, Scagliola - This is an open-access article distributed under the terms of the **Creative Commons Attribution License 3.0 Unported**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source

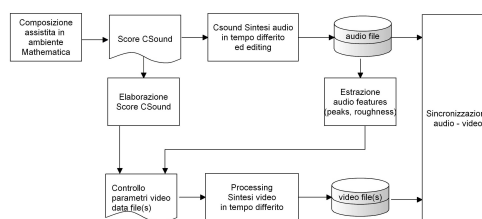


Fig. 3. “A un tempo”: fasi del lavoro

- **STUDIO SULL'INTONAZIONE DELLA CARNE** (2010, come opera multimediale, estensione di un lavoro acusmatico del 2006), di Francesco Abbrescia; per la sua realizzazione è stata operata una completa rigenerazione del brano elettroacustico unitamente alla sintesi video. Il video utilizza classi originali sviluppate dall'autore per la creazione di oggetti grafici con sintesi in tempo differito e controllo dei parametri mediante il tool Timeline;



Fig. 4. Una frame da “Studio sull'intonazione della carne” di F. Abbrescia

- **ROLLING SAND** (2010), di Antonio Mazzotti; costituisce il primo esempio di produzione contestuale di audio e video, completamente integrata, secondo la metodologia presentata in fig. 1. La sintesi video utilizza la libreria Traer-Physics 3.0 sviluppata da Jeff Traer [12].



Fig. 5. Una frame da “Rolling sand” di A. Mazzotti

4. CONCLUSIONE E SVILUPPI FUTURI

La scelta di Processing consente di accedere ad un vasto repertorio di codici (in forma di librerie o di sketch) [13] [14], messo a disposizione dalla operosa comunità di sviluppatori, consentendo così anche forme di riuso di prodotti e di idee in nuovi contesti come quello delle opere multimediali. I risultati finora conseguiti sembrerebbero forieri di interessanti sviluppi negli ambiti della composizione audiovisiva integrata. In questa prospettiva di ricerca – di particolare rilievo nell’ambito delle attività dei corsi istituzionali di Musica Elettronica del Conservatorio di Bari – si inserisce la redazione di una documentazione tutoriale.

5. REFERENCES

- [1] www.myspace.com/sinxthesis.
- [2] Wolfram Research - *Mathematica Documentation* (voll. *Core language, Visualization and Graphics, Data Manipulation*), 2008.
- [3] Casey Reas Ben Fry - *Processing: a programming handbook for visual and artists designers* - The MIT Press, 2007.
- [4] Daniel Schiffmann - *Learning Processing: A Beginner's Guide to Programming Images, Animation, and Interaction* – Kaufmann, 2008.
- [5] *The Canonical Csound Reference Manual* ver. 5.12, www.csounds.com.
- [6] Timeline Tool, <http://www.drifkin.net/timeline/>
- [7] Francesco Scagliola - *Composizione Musicale Assi-stita dal Calcolatore (Modello)*, Mathematica Italia User Group Meeting – Roma, 2008.
- [8] Francesco Scagliola - *Una struttura dati per una teoria computazionale della musica* – Mathematica Italia IV User Group Meeting - Milano, 2010.
- [9] Antonio Scarcia - “A un Tempo” *Un caso di composizione intermediale*, La Terra Fertile – Sassari, 2010.
- [10] MSAFluid library (web resources): <http://www.memo.tv/>
- [11] Ben Fry – *Visualizing Data* – O'Reilly, 2007.
- [12] Traer-Physics 3.0 (web resources): www.cs.princeton.edu/~traer/physics/
- [13] www.processing.org
- [14] www.openprocessing.org

INTRODUZIONE ALLA SINTESI WAVETABLE SWITCHING PER MULTIPLEXING DI SEGNALI

Stefano Silvestri

Conservatorio S. Pietro a Majella – Napoli
Dipartimento di Musica Elettronica e Nuove Tecnologie
stefanodrive_1@libero.it

ABSTRACT

La sintesi *wavetable switching* per *multiplexing* di segnale [1], in breve “sintesi *wavetable multiplexing*”, nasce da un lavoro di sperimentazione¹ inerente la produzione di spettri complessi dinamici nel tempo. Tale procedimento tempo-discreto, essenzialmente basato su oscillatori *wavetable look-up* [M. Mathews, 1969] e su algoritmi di emulazione di reti combinatorie semplici, rientra a far parte del gruppo dei metodi “transizionali” e più precisamente dei processi di sintesi per *wavetable switching* [W. Palm, 1975].

In linea generale alla base della tecnica vi è un meccanismo di combinazione sequenziale di più funzioni tabulate, ricomposte nel tempo [2,3] in un vettore risultante.

Tale concetto di generazione deriva in parte dalle prime implementazioni della *wavetable switching*², adottate verso la fine degli anni '70 su alcuni sintetizzatori digitali commerciali [4], e può essere visto come una forma di sintesi granulare deterministica dove operazioni quali lettura e finestratura rappresentano la generazione del grano <<quanto sonoro>>, mentre l'apparato di commutazione è ciò che guida i parametri del grano stesso.

1. INTRODUZIONE

L'articolo mostra le caratteristiche fondamentali del procedimento di sintesi introdotto. Vengono brevemente designati gli aspetti tecnici riguardanti l'implementazione del modello algoritmico e le relative possibilità sonore di

base, senza tralasciare alcuni formalismi essenziali legati al metodo generale di elaborazione.

2. PROCEDIMENTO DI SWITCHING

In linea di principio il procedimento di commutazione dei vettori segnale $v[n]$ è rappresentato graficamente da oscillatori *wavetable* accumulatori di fase $n = (n+1) \bmod N \rightarrow s_n = x[n]$ posti in parallelo.

Ad ogni $v[n]$ è applicata una funzione di inviluppo di periodo T_{env} di forma quadra, mediante il quale si effettua l'operazione ideale di commutazione. Per ottenere uno *switching* sequenziale si stabilisce un valore di sfasamento delle funzioni di commutazione tale che risulti $\varphi_n + \Delta t \forall T_{env}$.

La sequenza delle commutazioni sfasate $w[n]$ definirà il nuovo vettore somma $y[n]$ formato da brevi frazioni di campioni Q_N . Nel dominio spettrale il segnale risultante sarà composto, in istanti diversi, dal contributo frequenziale delle varie funzioni di ingresso lette in tabella e mandate con ordine prestabilito³ in uscita.

Per frequenze di commutazione $f_{cm} > 20 \div 25\text{Hz}$, corrispondenti a tempi di *switching* inferiori a $T_{cm} \approx 50 \div 40\text{ms}$, il segnale risultante sarà già caratterizzato da un certo arricchimento spettrale (presenza di bande laterali) ed il comportamento sequenziale del semplice *switching* progressivo⁴ solleciterà le capacità psico-fisiche del sistema uditivo riguardanti la *fusione percettiva* [S. Mc Adams] in dominio temporale. E' quest'ultima una particolare e ben nota conseguenza psicoacustica dell'ascolto di segnali ravvicinati di breve durata⁵.

¹ La ricerca qui presentata è parte del lavoro di tesi in cui si affrontano, tra l'altro, problematiche inerenti il controllo in tempo differito e reale dei parametri di sintesi. Le procedure di *switching* o commutazione del vettore segnale vengono coadiuvate da sistemi tempo discreti legati alla complessità; si studiano alcuni processi stocastici fondamentali quali i metodi congruenziali per la generazione di variabili pseudo-aleatorie, i moti Markoviani (catene di Markov) e Browniani discreti, basati sul progetto di generatori PRN's.

² Tale metodica di generazione era, ed è, utile per ottenere spettri complessi a partire da funzioni semplici.

Copyright: © 2010 Silvestri Stefano. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

³ Per l'ordine di commutazione si fa riferimento alla teoria del calcolo combinatorio dove risultano di grande utilità: combinazioni semplici, permutazioni ed altri raggruppamenti.

⁴ Il passaggio consecutivo, di ordine crescente, da T_n verso il successivo sottoperiodo $T_n - \text{esimo}$.

⁵ Si citano a tal proposito fenomeni psico-acustici di fusione percettiva, spettrale e temporale, le cui formalizzazioni vengono sfruttate soprattutto a partire dai modelli classici di riverberazione digitale [Schroeder, Logan, Moorer].

L'implementazione della suddetta tecnica comporta, anche ad uno stato fondamentale, una serie di artefatti da correggere. Indipendentemente dalla dimensione della commutazione del segnale tabulato di ingresso, di periodo possibile $T_n > T_{cm}; T_n = T_{cm}; T_n < T_{cm}$ ⁶, avendo adottato periodi di switching T_{env} quadri, bisognerà tenere conto dell'arricchimento spettrale apportato da ogni $\delta[n]$ ai margini del T_{env} . Infatti i periodi di inviluppo sfasati possono essere visti come una serie di gradini unitari contigui dove i punti di discontinuità $\mathfrak{S}$, ai margini del T_{env} , producono artefatti estranei all'applicazione specifica, proprio in concomitanza della delta di Dirac.

2.1 Finestratura del Segnale di Commutazione

In dominio discreto il punto di discontinuità $\mathfrak{S}$, o margine temporale (in campioni $N \geq T_c$) esistente tra periodi T_{env} consecutivi, ha corrispondenza con il passaggio di stato da parte della rete combinatoria⁷ in domino analogico, la quale commuta in risposta ad una serie di stimoli d'ingresso.

Nel caso in questione si vedrà che la rete combinatoria da adottare per la commutazione è un circuito di moltiplicazione (multiplexer) e gli stimoli di ingresso vengono definiti dai vettori segnale $v[n]$ e da vettori booleani contenenti bit di indirizzo $I = a, b, \dots$ atti ad individuare i segnali di ingresso sull'unica uscita U . Tale linea singola è idealmente il buffer del convertitore D/A, contenente il vettore $w[n]$ da porre in uscita.

Per ridurre le problematiche di arricchimento spettrale, indicate alla fine del punto 2, è conveniente definire un periodo di funzione di finestratura T_λ gaussiano⁸ tale che risulti $T_{env} \Leftrightarrow T_\lambda$. In tal modo si può osservare che:

- Per vettori tabellari semplici $v[n]$, in linea con la teoria di Fourier, si riduce⁹ il contributo spettrale apportato dalla sola funzione di commutazione T_{env} .

⁶ La possibilità $T_n > T_{cm}$ implica un segnale da commutare con durata estesa $\Delta t = N / Sr$ (switching di più periodi di tempo).

⁷ Sostanzialmente, in dominio analogico, il tempo di propagazione dello stato della porta logica interessata coincide con la saturazione/interdizione (o viceversa) dei transistor ad effetto di campo usati come "gate" per il segnale.

⁸ Una funzione Gaussiana, sfasata di $\pi/2$ e traslata sull'offset zero ha corrispondenza con l'intervallo $[0, 3\pi/2] \cup [0, 2\pi]$ della funzione sinusoidale.

⁹ Nell'algoritmo risulta fondamentale impiegare oscillatori wavetable efficienti (fasori di Array). Esperienze di compilazione in C-Sound e Pure Data dimostrano che, in caso contrario, subentrano problematiche di shifting delle fasi di commutazione-finestratura determinanti un arricchimento spettrale graduale.

È questo un caso ideale in cui la tipologia di sintesi si presta alla generazione di fasce sonore, di qualità percettiva tonica e simile al prodotto della FM semplice [J. M.- Chowning, 1967].

- Per vettori tabellari $v[n]_i$, differenti tra loro ed irregolari, il periodo di finestratura T_λ forza ogni segnale ad essere periodico e riduce il contributo spettrale apportato ai margini di T_{env} quadro.

Con tabelle discontinue si perviene facilmente ad una generazione sonora attestabile in termini di densità spettrale. È questa un'altra prerogativa della tipologia di sintesi introdotta, la quale, per determinate condizioni di generazione, può essere vista come una forma di sintesi granulare sincrona, dove le operazioni di lettura e finestratura stabiliscono la velocità e durata del quanto, mentre l'apparato di commutazione decreta l'ordinamento temporale del quanto stesso.

In tal caso le possibilità di controllo dell'apparato di commutazione-finestratura, e lo specifico contenuto tabellare, rappresentano il punto focale per la ricerca musicale di particolari variazioni e fluttuazioni spettrali di carattere microstrutturale, in linea concettualmente con la ricerca sonora del <<fin-qui-inudito>> [A. Di Scipio]. Le varianti sonore così ricavate trovano infatti applicazione in uno studio compositivo-musicale¹⁰ specifico.

3. FORMALIZZAZIONE SPETTRALE

Si analizzerà di seguito il caso basilare riguardante la commutazione di segnali semplici $T_n = T_{cm}$, di pari f_0 . La modulazione di segnali periodici in tabella, attraverso T_λ non lineari, porta ad una semplificazione del procedimento di generazione. Esso diviene formalizzabile in prima approssimazione attraverso le formule della modulazione di ampiezza semplice. Nel continuo, assumendo un singolo periodo sinusoidale T_n e modulando una portante¹¹ $\sin(\alpha)$, si ha:

$$y = \sin(\alpha)[1 + \sin(\beta)] \quad (1)$$

Supponendo una funzione di modulazione (o commutazione) sinusoidale $\sin(\beta)$.

Dalla E.1 è possibile notare il segnale modulante unipolare di finestratura (di modulo unitario); si assume

¹⁰ "Studio Sonoro III" – composizione interamente basata su algoritmi di wavetable multiplexing implementati sia per la sintesi in tempo reale che, per la parte per nastro, in tempo differito.

¹¹ La tecnica di switching del segnale comporta, ad un livello semplificato, la modulazione in tempi consecutivi di n valori tabulati in ingresso. Commutando segnali sinusoidali con stessa ampiezza, frequenza e fase è possibile semplificare il tutto analizzando un singolo intervallo di tempo, coincidente con una finestra T_λ singola.

inoltre che la fase del segnale commutabile sia coincidente:

$$\varphi_n \in T_n \Rightarrow T_{env} = T_\lambda \quad (2)$$

Per verificare la presenza delle bande laterali è possibile sviluppare la E.1 come accade per la AM semplice:

$$\begin{aligned} y &= \sin(\alpha)[\sin(\beta) + \sin(\alpha)] = \\ &= \frac{1}{2}[\cos(\alpha - \beta) - \cos(\alpha + \beta) + \sin(\alpha)] \end{aligned} \quad (3)$$

Tuttavia dovendo commutare sequenzialmente il contenuto di più di una tabella, vi insorgeranno n bande laterali attorno agli $f_0 \forall T_n$.

Supponendo di modificare deterministicamente le frequenze fondamentali $f_0 \leq \frac{Sr}{2}$ si avrà una apparente

sovrapposizione sequenziale delle bande laterali. Tale fenomeno spettrale è caratteristico di questo metodo di generazione e produce interazioni particolari tra le diverse bande laterali dei diversi spettri commutati¹². In primis il sistema uditivo può percepire, in misura della diversa posizione delle laterali e delle fondamentali f_0 dello spettro di commutazione, battimenti di primo e secondo ordine dovuti ad una vicinanza "illusoria" delle frequenze stesse.

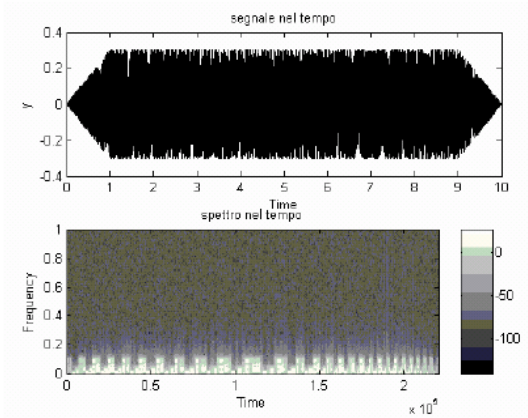


Figura 1. Test condotto con $f_0 = 40\text{Hz}$ e distanza=2

Tali formanti sono infatti distribuite temporalmente entro intervalli inferiori al periodo di discriminazione proprio del sistema uditivo.

L'elaborazione sfrutta essenzialmente il passaggio da un diverso contenuto spettrale all'altro entro brevi

¹² Fenomeni accentuabili se la modulazione di ogni f_0 , e la stessa posizione di f_0 , sono posti in condizione di "overlapping" delle bande spettrali.

tempi di azione ed ordini prestabiliti secondo leggi deterministiche e/o pseudo-caotiche.

Diviene possibile così rappresentare forme sonore non statiche, pur derivanti da punti pre-allocati in memoria (od allocati in real-time).

Da una forma di elaborazione "per modelli del segnale" [8], lineare e standard, si passa ad una forma di generazione propriamente non-standard.

In tal caso il termine sintesi è giustificato dal fatto che è lo stesso trattamento dei campioni ad influenzare gli attributi diretti del segnale musicale [9].

4. ALGORITMO DI MULTIPLEXING

La tecnica di multiplazione, derivante dall'elettronica digitale e dalla teoria delle reti combinatorie, è applicata in tal sede come metodo formale di distribuzione degli NT_c campioni commutabili.

L'algoritmo implementato permette di sintetizzare la rete di un multiplexer ad $N-bit$, tramite operazioni logiche. Porte AND, NOT, OR sono alla base di questo modello di selezione; esse permettono di smistare $U = 2^I$ ingressi dato attraverso I bit di indirizzo. I dati sono appunto rappresentati dai segnali discreti da processare.

La condizione 2^{I-bit} è tale da consentire l'individuazione univoca di un solo $v[n] \cdot v_\lambda[n]$ in uscita U .

Dalla teoria delle telecomunicazioni, in particolar modo in relazione ai trattamenti multipli dei segnali, si sottolinea l'impiego del TDM (*Time Division Multiplexing*) dove ogni parte di segnale tributario (*slots*) è collocato nel tempo con ordine prestabilito, alla frequenza f_{mux} . Da ciascuno degli N segnali semplici S_N , limitati superiormente alla frequenza $2f_M$ [Shannon-Nyquist], si prelevano campioni ad intervalli costanti:

$$\Delta t \leq \Delta t_M = \frac{1}{2f_M} \quad (4)$$

Ad istanti di campionamento non sovrapposti. Il segnale unico di uscita s_p , non finestrato, può considerarsi modulato ad onde quadre (sfasate) e per tanto di tipo PAM (*Pulse Amplitude Modulation*), con distanza costante fra campioni adiacenti pari a:

$$\Delta t_p = \frac{\Delta t}{N} \leq \frac{1}{2Nf_M} \quad (5)$$

Da cui ponendo N^{-1} si ha:

$$\Delta t_p = \Delta t(N^{-1}) \leq 2f_M^{-1}(N^{-1}) \quad (6)$$

Quindi per la E.5 risulterà:

$$\Delta t \leq \frac{1}{2} (Nf_M)^{-1} \quad (7)$$

Lo schema generale del modello algoritmico capace di svolgere il comportamento descritto è il seguente:

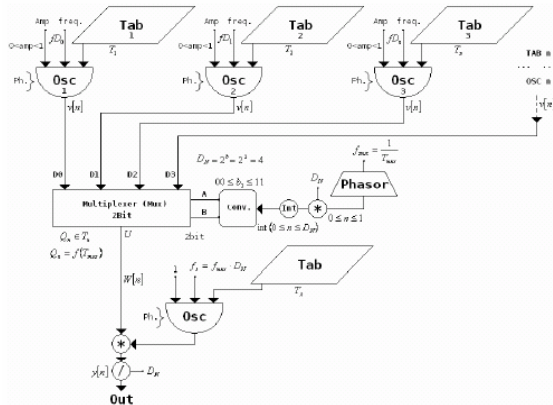


Figura 2. Schema a blocchi dell’algoritmo di sintesi

Il modulo di multiplexer è una macro contenente al suo interno la rete combinatoria predisposta per il trattamento di segnali:

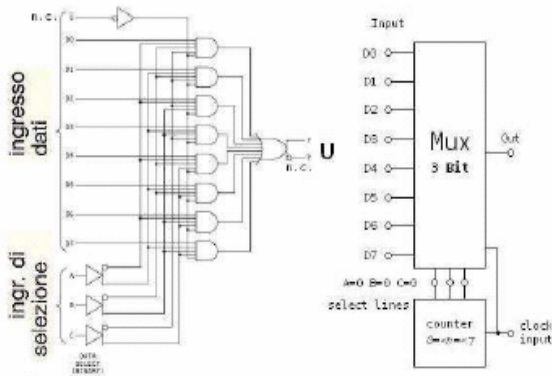


Figura 3. Contenuto schematico del sottomodulo “Mux”

Ogni segnale in uscita al sistema potrà quindi avere parametri diversi. Ad es. si potrà adottare una serie geometrica delle frequenze ed altri parametri per costruire una certa successione spettrale¹³, anche dinamica del tempo.

¹³ In Figura 1 il parametro relativo alla distanza è un semplice valore moltiplicativo che separa i centro-banda f_0 nello spettro; similmente al caso di serie matematiche adottate per la sintesi additiva a spettro fisso.

5. CONCLUSIONI E SVILUPPI

Un primo sviluppo può riguardare l’approfondimento delle tecniche di controllo dei parametri, con maggiore attenzione alle metodiche di auto-generazione aleatoria. Il modello multiplexer di base è inoltre ampliabile con numeri maggiori di ingressi dati ed è possibile effettuare il porting del codice direttamente su un dispositivo a microcontrollore (MCU ad architettura RISC per il DSP) per un’elaborazione dedicata del segnale.

Tra l’altro è possibile introdurre nuovi parametri come quello di scala per la funzione di finestratura, da rendere a sua volta funzione di una variabile indipendente.

Attualmente è in fase di studio un’espansione del modello capace di commutare sorgenti sonore acquisite in tempo reale (contenuti in tabella variabili).

Non di minore importanza è la possibilità di modificare la sequenza dei segnali di ingresso D entro le n possibilità combinatorie; ovviamente solo nei casi in cui permutazioni e scambi risultino acusticamente significativi.

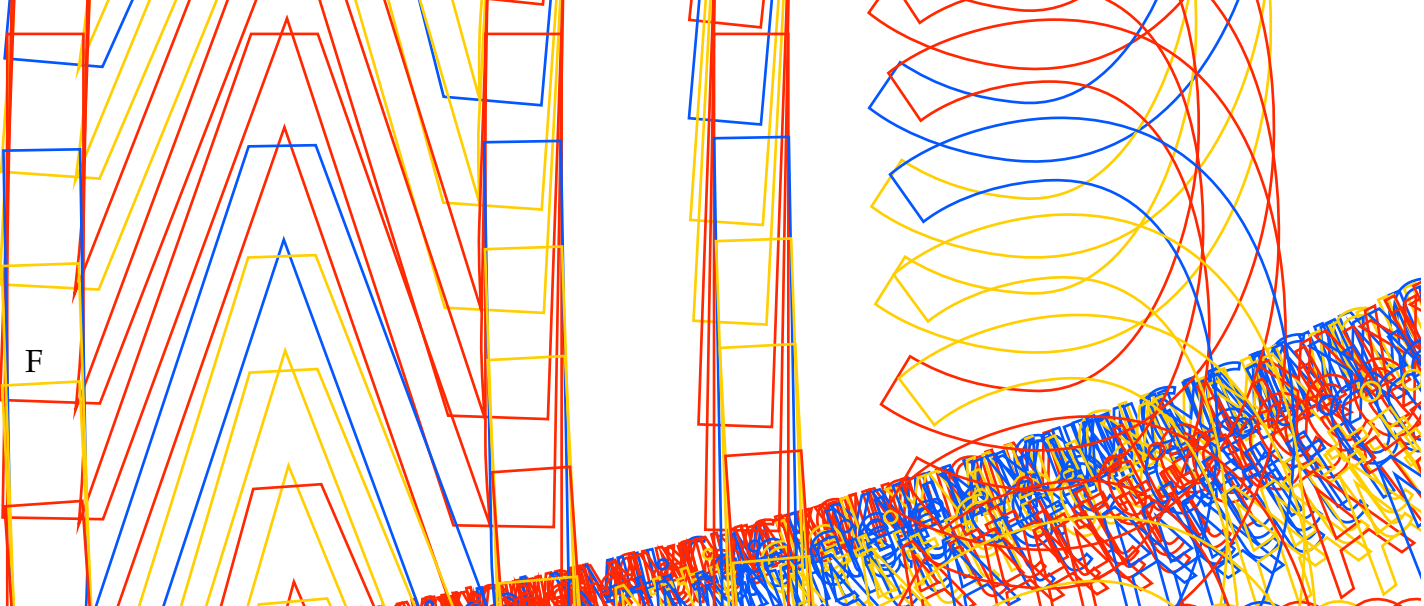
La variante di wavetable switching qui introdotta può essere quindi assimilata ad un processo *granulare-deterministico* atto a fornire una rappresentazione quantistica [Xenakis, Gabor], pur sempre ordinata degli *stream* sonori. Essa rappresenta un *front-end* nuovo-caratteristico per il controllo creativo del segnale. In tal caso parametri quali densità e spessore sarebbero attribuibili alla velocità di transizione e alle dimensioni dei segnali commutati.

In fine è opportuno evidenziare che in letteratura esistono numerosi studi ed ampliamenti della tecnica di commutazione tabellare. Spesso con nomi e metodi differenti si intendono particolari sfumature del procedimento wavetable-look up. Nelle implementazioni in hardware si fa anche riferimento alla stessa tecnica di Multiplexing. Diversamente, la procedura qui illustrata adotta alla base un commutatore emulato per lo “smistamento a velocità acustica” di diversi contenuti tabulati. In tal caso sussistono differenti e molteplici possibilità di gestione dei campioni.

6. REFERENCES

- [1] Tesi di laurea in Musica elettronica e nuove tecnologie: “*Sintesi Wavetable Switching per multiplexing di segnali e controllo dei parametri mediante processi stocastici Markoviani*”. Relatore A. Di Scipio, A.A. 2008/2009; (I Liv).
- [2] R. Bristow – Johnson, (1996). “*Wavetable Synthesis 101, A Fundamental Prospective*” – Wave Mechanics, Inc. AES convention
- [3] A. Horner, J. Beauchamp, L. Haken, (1993). “*Methods for Multiple Wavetable Synthesis of Musical Instrument Tones*” – J. Audio Eng. Doc
- [4] D. T. Horn, (1988). “*Il Sintetizzatore*”, Franco Muzzio Editore, Padova

- [5] F. Valdoni – F. Vatalaro, (1984). *“Telecomunicazioni”*, Calderini
- [6] A. Di Scipio, (2009). Dispense del corso di informatica musicale III, acustica e psicoacustica II&III
- [7] S. Cavaliere, (2009). Dispense del corso di elaborazione numerica dei segnali
- [8] G. De Poli, (1981). *“Tecniche numeriche di sintesi della musica”*, a cura di A. Vidolin, Bollettino LIMB I
- [9] M. V. Mathews, (1976). *“Tecnologia della musica di elaboratore”*, in H. Pousser, *“La musica elettronica”*– Feltrinelli



Prossime distanze. Atti del XVIII CIM

