

# SINESTESIE SONORE SONIC SYNESTHESIA



© Ivan Penov

XIX COLLOQUIO DI INFORMATICA MUSICALE  
19TH COLLOQUIUM ON MUSIC INFORMATICS

Trieste 21–24 novembre 2012  
*Trieste, November 21-24, 2012*



XIX Colloquio di Informatica Musicale  
*19th Colloquium on Music Informatics*

SINESTESIE SONORE  
*SONIC SYNESTHESIA*

---

Trieste 21–24 novembre 2012  
*Trieste, November 21–24, 2012*

AA.VV.

SINESTESIE SONORE / SONIC SYNESTHESIA  
XIX Colloquio di Informatica Musicale  
*19th Colloquium on Music Informatics*

**A cura di:**

**Editors:**

Pietro Polotti, Giorgio Klauer, Federico Fontana, Carlo Drioli

© 2012 AIMI – Associazione Informatica Musicale  
I diritti dei singoli articoli rimangono agli autori  
*The copyrights of the single articles remain with the authors*

**ISBN 9788890341304**

Per gentile collaborazione di Davide Rocchesso e Stefano delle Monache  
*Courtesy of Davide Rocchesso and Stefano delle Monache*

**Pubblicato da:**

**Publisher:**

DADI - Dip. Arti e Design Industriale  
Università IUAV di Venezia

**In co-edizione con:**

**Co-published with:**

Conservatorio di Musica “G. Tartini” di Trieste  
Università degli Studi di Udine



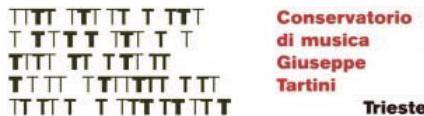
Quest’opera è sotto licenza Creative Commons Attribuzione - Non opere derivate 3.0 Unported License.  
Copia della licenza è disponibile presso <http://creativecommons.org/licenses/by-nd/3.0/>.  
*This work is licensed under the Creative Commons Attribution-NoDerivs 3.0 Unported License.*  
*To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/3.0/>.*

A Giovanni Battista Debiasi,  
pioniere dell'informatica musicale  
(1928-2012)





AIMI – Associazione di Informatica Musicale Italiana  
[www.aimi-musica.org](http://www.aimi-musica.org)



Conservatorio di Musica "Giuseppe Tartini" di Trieste  
Dipartimento di Nuove Tecnologie e Linguaggi Musicali  
[www.conservatorio.trieste.it](http://www.conservatorio.trieste.it)



In collaborazione con  
*In collaboration with*

Casa della Musica, Trieste  
[www.scuoladimusica55.it](http://www.scuoladimusica55.it)



Teatro Miela  
[www.miela.it](http://www.miela.it)



Università degli Studi di Udine  
Dipartimento di Matematica e Informatica  
[www.dimi.uniud.it](http://www.dimi.uniud.it)



Cantierzero  
Translocal collective for music and contemporary arts (I/A)  
[www.cantierzero.org](http://www.cantierzero.org)



Consortium GARR  
La Rete Italiana dell'Università e della Ricerca  
[www.garr.it](http://www.garr.it)



Con il patrocinio di  
*Endorser*

Comune di Trieste

**Con il supporto e l'amichevole collaborazione di**  
*With support by*



Fondazione Benefica Kathleen Foreman Casali  
SKE Fonds - Austro Mechana  
[www.ske-fonds.at](http://www.ske-fonds.at)



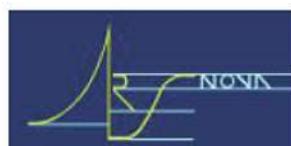
Forum Austriaco di Cultura Milano  
[www.bmeia.gov.at](http://www.bmeia.gov.at)



Scuola Superiore Internazionale di Musica da Camera "Trio di Trieste" Collegio del Mondo Unito dell'Adriatico-ONLUS  
[www.uwcad.it](http://www.uwcad.it)

**Servizio Concerti**  
Cerneaz

Servizio Concerti Cerneaz



Accademia Ars Nova  
[www.arsnovatrieste.it](http://www.arsnovatrieste.it)



Musicali Rossoni  
[www.musicalirossoni.it](http://www.musicalirossoni.it)

Civico Museo Teatrale "Carlo Schmidl" Trieste  
[www.triestecultura.it](http://www.triestecultura.it)



AUSTROMECHANIA  
[www.austromechana.at](http://www.austromechana.at)



**Media partner**  
*Media partner*

RTV Koper - Capodistria  
[www.rtvslo.si./capodistria](http://www.rtvslo.si./capodistria)

XIX COLLOQUIO DI INFORMATICA MUSICALE  
19th COLLOQUIUM ON MUSIC INFORMATICS

L'Associazione di Informatica Musicale Italiana (AIMI), fondata nel 1981, è punto di riferimento nel panorama italiano per musicisti e ricercatori interessati al rapporto tra musica, scienza e tecnologia. L'AIMI, che vanta tra i suoi associati musicisti e ricercatori di rilevanza internazionale, patrocina con cadenza biennale i Colloqui di Informatica Musicale (CIM), che rappresentano il più atteso momento di discussione scientifica e di confronto musicale della comunità italiana di informatica musicale.

Il Colloquio di Informatica Musicale costituisce il punto d'incontro tra gli scienziati, gli artisti e i musicologi che si dedicano secondo le rispettive metodologie all'interazione musica-tecnologia e ha come scopo la divulgazione delle loro ricerche sulle applicazioni artistico-musicali delle tecnologie informatiche, nei risvolti accademici, artistici, culturali e pedagogici. Come di consueto, il programma del Colloquio include comunicazioni scientifiche, attività didattiche e produzioni artistiche. Le comunicazioni scientifiche/artistiche sono state selezionate mediante una Call for papers e i lavori musicali e artistici mediante una Call for scores/works; tutte le proposte sono state vagilate dal Comitato scientifico e selezione musicale.

*The Italian Association of Music Informatics (AIMI), founded in 1981, is a reference point for Italian musicians and researchers who are interested in the relationships between music, science and technology. The AIMI, which counts internationally renowned musicians and researchers among its affiliates, organizes every two years the Colloquium on Music Informatics (CIM), representing the most important moment of scientific discussion and musical debate in the Italian community of music informatics.*

*The Colloquium on Music Informatics is a touch point for scientists, artists and musicologists who dedicate their methodologies to the interaction between music and technology, and serves the dissemination of their research on the artistic-music applications of information technology in all its academic, artistic, cultural and pedagogical facets. As it is customary, the program of the Colloquium includes scientific communications, seminal activities and artistic productions. The scientific/artistic communications have been selected through a Call for papers, and the musical and artistic works through a parallel Call for scores/works; all submissions have been reviewed by the Scientific and Musical Committee.*

SINESTESIE SONORE  
SONIC SYNESTHESIA

I temi dell'edizione 2012 sono multimodalità e interattività. Attraverso le tecnologie digitali, le diverse forme d'arte tendono oggi a stabilire forti interconnessioni estetiche e luoghi di coabitazione. In particolare il suono è sempre più presente e quasi necessario in qualsiasi tipo di espressione artistica. In una recente call for papers del Leonardo Music Journal si legge: "Art is getting noisier every day. Whether made by sculptors, video artists, composers, printmakers or installation artists, there's no question but that "Sound Art" is a genre ascendant." Il suono si integra quindi in senso multimodale nel discorso artistico e, spesso, il mezzo tecnologico è lo strumento più efficace per raggiungere questo scopo. In modo simile, la tecnologia odierna permette lo sviluppo di lavori artistici in cui l'interazione con la macchina da parte di un performer, o da parte del pubblico, diventa sempre più dinamicamente legata al corpo e a ciò che avviene nell'ambiente in cui l'opera si colloca. L'idea di embodiment e di recupero della fisicità nel rapporto con la macchina sono stimolo di nuove forme di espressione e relazione con la tecnologia, non solo in ambito artistico ma anche nella vita quotidiana.

*Multimodality and interactivity are the topics of the 2012 edition. Today, different forms of art tend to establish strong aesthetic connections and shared domains through the digital technology. In particular, sound is becoming more and more present, and somehow necessary in any type of artistic expression. A recent call for papers at the Leonardo Music Journal reads: "Art is getting noisier every day. Whether made by sculptors, video artists, composers, printmakers or installation artists, there's no question but that 'Sound Art' is a genre ascendant." Sound, then, integrates the artistic discourse in multimodal sense, and often the technological medium becomes the most effective tool for achieving this goal. Similarly, nowadays technology allows for the development of artworks in which the interaction between a machine and the performer, or the audience, becomes more and more dynamically linked to the human body and to what happens in the environment where the artistic work takes place. The idea of embodying and recovering physicality during the relationship with the machine is stimulating new forms of expression and contact with the technology, not only in the artistic realm but also in the everyday life.*

COMITATO SCIENTIFICO E DI SELEZIONE MUSICALE  
*SCIENTIFIC AND MUSICAL PROGRAM COMMITTEE*

Federico Avanzini, Università di Padova  
Stefano Bassanese, Conservatorio di Torino  
Davide Bonsi, Università di Venezia  
Nicola Buso, Conservatorio di Trieste  
Antonio Camurri, Università di Genova  
Sergio Canazza, Università di Padova  
Carmine Emanuele Cella, IRCAM, Parigi (F)  
Andrea Cera, Accademia di Brera, Milano  
Riccardo Dapelo, Conservatorio di Genova  
Amalia De Götzen, Aalborg University, Copenhagen (DK)  
Stefano Delle Monache, Università IUAV di Venezia  
Agostino Di Scipio, Conservatorio di Napoli  
Roberto Doati, Conservatorio di Genova  
Carlo Drioli, Università di Udine  
Federico Fontana, Università di Udine  
Michele Geronazzo, Università di Padova  
Antonio Grande, Conservatorio di Como  
Giorgio Klauer, Conservatorio di Como  
Michelangelo Lupone, Conservatorio dell'Aquila  
Enrico Marchetto, Università di Padova  
Paolo Pachini, Conservatorio di Trieste  
Stefano Papetti, ICST, Zurich University of the Arts, Zurigo (CH)  
Fabián Pérez Tedesco, Conservatorio di Trieste  
Pietro Polotti, Conservatorio di Trieste  
Davide Rocchesso, Università IUAV di Venezia  
Antonio Rodà, Università di Padova  
Corrado Rojac, Conservatorio di Trieste  
Sylviane Sapir, Conservatorio di Como  
Daniele Salvati, Università di Udine  
Stefania Serafin, Aalborg University, Copenhagen (DK)  
Domenico Sciajno, Conservatorio di Trapani  
Simone Spagnol, Università IUAV di Venezia  
Andrea Valle, Università di Torino  
Alvise Vidolin, Conservatorio di Padova  
Gualtiero Volpe, Università di Genova  
Stefano Zambon, Viscount SpA, Italy  
Laura Zattra, Università di Padova

Coordinamento generale

*General chair*

Pietro Polotti

Coordinamento programma musicale

*Musical program chairs*

Nicola Buso, Giorgio Klauer, Paolo Pachini

Coordinamento programma scientifico

*Scientific program chairs*

Carlo Drioli, Federico Fontana

Curatori degli atti

*Editors*

Pietro Polotti, Giorgio Klauer, Federico Fontana, Carlo Drioli



**Programma del colloquio**  
*Colloquium program (in Italian)*

**8**

**Indice degli articoli**  
*Index of papers*

FEDERICO FONTANA <i>Interactive sound synthesis by the Lotka-Volterra population model</i>	20
STEFANO BALDAN, ADRIANO BARATÈ, GOFFREDO HAUS, LUCA ANDREA LUDOVICO <i>Fruizione evoluta via web di contenuti musicali e multimediali: il portale EMIPIU</i>	26
FEDERICA BRESSAN <i>Un ambiente informatico per la gestione dei processi relativi alla conservazione attiva dei documenti sonori</i>	34
TIZIANO BOLE, CRISTINA FEDRIGO <i>ALA – Un'iper-lavagna per il training ritmico</i>	40
LUCA ROVERELLI, NICOLÓ ZEREIK <i>Movements classification on the Laban's theory of effort time axis</i>	46
CRISTINA FEDRIGO, SARA RADIN <i>EGGS (Elementary Gestalts for Gesture Sonification). Risorse informatiche per l'educazione e didattica musicali</i>	52
JULIAN SCORDATO <i>Ariadne: un sistema per l'esecuzione musicale su tablet</i>	60
STEFANO PIANA, ANTONIO CAMURRI, MAURIZIO MANCINI <i>The EyesWeb human movement analysis library</i>	66
WALTER CIANCIUSI, LORENZO SENO <i>Feed-Drum e Skin-Act: l'ultima frontiera dello strumento aumentato</i>	72
NICOLA SCATTOLIN, SERENA ZANOLLA, ANTONIO RODÀ, SERGIO CANAZZA <i>SoundingARM rappresentazione acustica di una mappa spaziale</i>	80
SERENA ZANOLLA, GIORGIO TEMPESTA, SERGIO CANAZZA, LAURO SNIDARO <i>MEMORY SONORO realtà aumentata accessibile mediante audio direzionale</i>	88
DAVID MEREDITH <i>Music analysis and Kolmogorov complexity</i>	96
SILVIA LANZALONE <i>Clavecin électrique. Studio e realizzazione dell'opera</i>	104
MARIA-LETIZIA MICHIELON <i>"Il divino principio di incessante e sapiente vita". Proposta per una rilettura live electronics dell'Anima Mundi platonica</i>	112

MARCELLA MANDANICI

*La micropolifonia ligetiana come modello di composizione algoritmica*

120

GIACOMO ALBERT

*Multimodalità e Immersione: verso una rilettura del Minimalismo attraverso la storia delle installazioni sonore*

128

GIORGIO KLAUER

*Analisi paradigmatica di composizioni acusmatiche e music information retrieval: un primo approccio*

136

ANDREA ARCELLA, STEFANO SILVESTRI

*Analogique B – A computer model of the compositional process*

144

ANDREA AGOSTINI, DANIELE GHISI

*Bach: una libreria per la composizione assistita in Max*

150

DIMITRIOS MARGOUNAKIS, DIONYSIOS POLITIS

*Revealing the hidden motives of chromaticism in music through computational semantic Visualizations*

158

SIMONE SPAGNOL

*Are spectral elevation cues in head-related transfer functions distance-independent?*

166

DIMITRIOS MARGOUNAKIS, DIONYSIOS POLITIS

*Exploring the relations between chromaticism, familiarity, scales and emotional responses in music*

172

**Indice dei poster**  
*Index of posters*

MASSIMO GRASSI, GUILLAUME LEMAITRE, DAVIDE ROCCHESSO <i>Cinquant'anni di dislocazione temporale acustica</i>	182
DANIELE SALVATI, SERGIO CANAZZA <i>Cepstrum analysis and blind system identification for localization of pseudo-periodic sounds</i>	184
MICHELE GERONAZZO, SIMONE SPAGNOL, DAVIDE ROCCHESSO, FEDERICO AVANZINI <i>Model-based customized binaural reproduction through headphones</i>	186
GIANLUCA DARIO, RICCARDO MAZZETTI, LORENZO JACOPO SCIUCA, FEDERICO FONTANA <i>EGLO – Un'applicazione per la sonificazione interattiva di tavoli di gioco per bambini</i>	188
ELIO TOPPANO, ROBERTO RANON <i>Sincresi audio-visiva: una esperienza didattica</i>	190
ANDREA VALLE <i>A sound programming strategy to survive to a complex multimedia setup</i>	192
CARLO DRIOLI, CLAUDIO ALLOCCHIO <i>LOLA: a low latency high quality a/v streaming system for networked performances and Interaction</i>	194
LEONARDO GABRIELLI, STEFANO SQUARTINI, FRANCESCO PIAZZA, MATTEO MENCARELLI <i>Towards a portable wireless platform for networked performance</i>	196
DARIO SANFILIPPO <i>LIES (distance/incidence) 1.0: a human-machine interaction performance</i>	198
ELIO CORBOLANTE, ALESSANDRO FOGAR <i>h</i>	200
ALESSANDRO FIORIN DAMIANI, FILIPPO GECELE, ALESSANDRO DE NARDI, PAOLO BOSCAROL, CHIARA OLCESE, LUCA LORENZO FERRO, FEDERICO PETREI <i>Siloë</i>	202



## mercoledì 21 novembre 2012 / Wednesday, November 21, 2012

### Teatro Miela

---

21.00 **Concerto inaugurale / Opening concert**

#### BODYSNATCHERS/ULTRACORPI

ENSEMBLE CANTIEREZERO

IVAN RABAGLIA *violino*

STEFANIA AMISANO, CLAUDIO CRISTANI *pianoforte*

GIORGIO KLAUER, STEFANO TREVISO *regia del suono*

---

GIORGIO KLAUER (1976)

*Three Men in a Wormhole* (2012) \*

*per pianoforte a quattro mani, violino, self-sensing actuators ed elettronica*

AGOSTINO DI SCIPIO (1962)

*chpn3.2* (2012) \*

*variazioni di sensibilità al sé*

*per 3 esecutori*

*rete di feedback con pianoforte, attuatori elettrodinamici auto-sensibili, e altri dispositivi*

JOHN CAGE (1912-1992)

*Two<sup>6</sup>* (1992)

*per violino e pianoforte*

\*\*\*

STEFANO TREVISO (1974)

*Still again Sealed again* (2012) \*

*per pianoforte (due esecutori), violino, self-sensing actuators ed elettronica*

FABIO NIEDER (1957)

*Requiem für einen Namen* (2010)

*versione per due pianisti*

GERHARD E. WINKLER (1959)

*Black Mirrors II* (2012) \*

*HerzStück*

*per pianoforte a quattro mani, self-sensing actuators e live electronics*

\* *prima esecuzione assoluta*

*Il progetto Bodysnatchers/Ultracorpi è realizzato con il contributo di SKE Fonds - Austro Mechana, Forum Austriaco di Cultura (Milano), ZKB - Credito Cooperativo del Carso e viene presentato grazie al supporto e all'amichevole collaborazione di Scuola Superiore Internazionale di Musica da Camera "Trio di Trieste" - Collegio del Mondo Unito dell'Adriatico O.N.L.U.S., Lorenzo Cerneaz Pianoforti (Udine), Accademia "Ars Nova" (Trieste) e Civico Museo Teatrale "Carlo Schmidl" (Trieste)*

**giovedì 22 novembre 2012 / Thursday, November 22, 2012**

**Conservatorio "G. Tartini"**

---

- 10.00 **Introduzione al convegno / Conference opening**
- Presentazioni orali / Paper sessions**  
coordinatore Federico Fontana (*Università di Udine*)
- I sessione / Session 1**  
*chair Carlo Drioli (*Università di Udine*)*
- 10.15 CLAUDIO ALLOCCHIO (GARR) *keynote speaker*  
*LOLA, from an idea to real time human interaction*
- 11.00 FEDERICO FONTANA  
*Interactive sound synthesis by the Lotka-Volterra population model*
- 11.30 **Coffee break**
- II sessione / Session 2**  
*chair Federico Fontana (*Università di Udine*)*
- 11.45 STEFANO BALDAN  
ADRIANO BARATÈ  
GOFFREDO HAUS  
LUCA ANDREA LUDOVICO  
*Fruizione evoluta via web di contenuti musicali e multimediali: il portale EMIPIU*
- 12.10 FEDERICA BRESSAN  
*Un ambiente informatico per la gestione dei processi relativi alla conservazione attiva dei documenti sonori*
- 12.35 TIZIANO BOLE  
CRISTINA FEDRIGO  
*ALA – Un'iper-lavagna per il training ritmico*
- 13.00 **Poster flash presentation**  
coordinatore *Carlo Drioli (*Università di Udine*)*
- MASSIMO GRASSI  
GUILLAUME LEMAITRE  
DAVIDE ROCCHESO  
*Cinquant'anni di dislocazione temporale acustica*
- DANIELE SALVATI  
SERGIO CANAZZA  
*Cepstrum analysis and blind system identification for localization of pseudo-periodic sounds*
- MICHELE GERONAZZO  
SIMONE SPAGNOL  
DAVIDE ROCCHESO  
FEDERICO AVANZINI  
*Model-based customized binaural reproduction through headphones*

GIANLUCA DARIO  
RICCARDO MAZZETTI  
LORENZO JACOPO SCIUCA  
FEDERICO FONTANA

*EGLO – Un'applicazione per la sonificazione interattiva di tavoli di gioco per bambini*

ELIO TOPPANO  
ROBERTO RANON

*Sincresi audio-visiva: una esperienza didattica*

13.10 **Video-listening room\* & Poster**

13.10 **Lunch break**

**III sessione / Session 3**

*chair Pietro Polotti (Conservatorio di Trieste)*

15.00 WALTER GERBINO (Università di Trieste) *keynote speaker*  
*Cross-modal congruency, attention, and visual recognition*

15.40 LUCA ROVERELLI  
NICOLÒ ZEREIK  
*Movements classification on the Laban's theory of effort time axis*

16.05 CRISTINA FEDRIGO  
SARA RADIN  
*EGGS (Elementary Gestalts for Gesture Sonification). Risorse informatiche per l'educazione e didattica musicali*

16.30 **Video-listening room\* & Poster**

16.30 **Coordinamento Nazionale Docenti di Musica Elettronica**

17.30 **I sessione d'ascolto / Listening session 1**  
*coordinatore Nicola Buso (Conservatorio di Trieste)*

NATHANIEL GOOSSEN, *No more glaciers*

GIULIANO ANZANI, A-M

FRANCESCO BIANCO, *Astratto #1*

MARCO DIBÈLTULU, *Sguardo contemporaneo*

FRANCESCO GALANTE, *Liberare la terra dalla immobilità fissata (quadro primo)*

FRANCESCO GARBO, *Nocicezione*

DIEGO GARRO, *Más Claro*

## Casa della Musica

---

20.30      **Concerto elettroacustico e live electronics**  
***Electroacoustic and live electronics concert***  
coordinatore Nicola Buso (*Conservatorio di Trieste*)

FABIÁN PÉREZ TEDESCO

*Ecos*

GIANNI BALICU  
MARCELLINO GARAU  
MARCO MARINONI  
FILIPPO MEREU  
SANDRO MUNGIANU  
ROBERTO ZANATA

*Non consumiamo Cage*

JOÃO PEDRO OLIVEIRA

*Vox sum vitae*

STEFANO VINCIARELLI

*Rifrazione di ecosistemi*

MASSIMO PALERMO

*ZAUM II*

MAURA CAPUZZO

*Come arriva l'amore così ti si libera il naso, improvvisamente ed a caso*

---

MARKO JUGOVIC, FABIÁN PÉREZ TEDESCO *percussioni*

MAURA CAPUZZO *voce*

TIZIANO BOLE, DARIO KHADEMI, IVAN PENOV

*e gli autori stessi, elettronica e regia del suono*

**venerdì 23 novembre 2012 / Friday, November 23, 2012**

**Conservatorio "G. Tartini"**

---

**Presentazioni orali / Paper sessions**

**IV sessione / Session 4**

chair Andrea Valle (Università di Torino)

- 9.30 MIHA CIGLAR (IRZU, Ljubljana) *keynote speaker*  
*Syntact – touch the sound*
- 10.10 JULIAN SCORDATO  
*Ariadne: un sistema per l'esecuzione musicale su tablet*
- 10.35 STEFANO PIANA  
ANTONIO CAMURRI  
MURIZIO MANCINI  
*The EyesWeb human movement analysis library*
- 11.00 WALTER CIANCIUSI  
LORENZO SENO  
*Feed-Drum e Skin-Act: l'ultima frontiera dello strumento aumentato*
- 11.30 **Coffee break**
- V sessione / Session 5**  
chair Federico Avanzini (Università di Padova)
- 11.45 NICOLA SCATTOLIN  
SERENA ZANOLLA  
ANTONIO RODÀ  
SERGIO CANAZZA  
*SoundingARM rappresentazione acustica di una mappa spaziale*
- 12.10 SERENA ZANOLLA  
GIORGIO TEMPESTA  
SERGIO CANAZZA  
LAURO SNIDARO  
*MEMORY SONORO realtà aumentata accessibile mediante audio direzionale*
- 12.35 DAVID MEREDITH  
*Music analysis and Kolmogorov complexity*
- 13.00 **Poster flash presentation**
- ANDREA VALLE  
*A sound programming strategy to survive to a complex multimedia setup*
- CARLO DRIOLI  
CLAUDIO ALLOCCHIO  
*LOLA: a low latency high quality a/v streaming system for networked performances and interaction*

LEONARDO GABRIELLI  
STEFANO SQUARTINI  
FRANCESCO PIAZZA

*Towards a portable wireless platform for networked performance*

DARIO SANFILIPPO

*LIES (distance/incidence) 1.0: a human-machine interaction performance*

ALESSANDRO FOGAR  
ELIO CORBOLANTE

*h*

ALESSANDRO FIORIN DAMIANI  
FILIPPO GECELE  
ALESSANDRO DE NARDI  
PAOLO BOSCAROL  
CHIARA OLCESE  
LUCA LORENZO FERRO  
FILIPPO GENESINI  
FEDERICO PETREI

*Siloë*

13.10 **Video-listening room\* & Poster**

13.10 **Lunch break**

**VI sessione / Session 6**

*chair Stefano Bassanese (Conservatorio di Torino)*

14.45 SILVIA LANZALONE

*Clavecin électrique. Studio e realizzazione dell'opera*

15.10 MARIA-LETIZIA MICHELI

*"Il divino principio di incessante e sapiente vita". Proposta per una rilettura live electro-nics dell'Anima Mundi platonica*

16.00 **Video-listening room\* & Poster**

16.00 **Assemblea Associazione di Informatica Musicale Italiana**

17.30

**Il sessione d'ascolto / Listening session 2**

coordinatore Nicola Buso (Conservatorio di Trieste)

ANTONIO D'AMATO

*Prismes électriques – hommage à Sonia Delaunay*

CHRISTIAN MARCHI

*Elaborazione del canto di un uccello*

NICOLA MONOPOLI

*Juicy Joyce – Subvocalizing (at pre-speech level)*

TOMMASO MUTO

*Attraversamenti*

BENJAMIN O'BRIEN

*Densité*

ANTONIO SCARIA

*Elegia seriale*

JEROD SOMMERFELDT

*Kernel\_panic*

---

## Casa della Musica

---

20.30

**Concerto performance interattive / Interactive performances**

coordinatore Pietro Polotti (Conservatorio di Trieste)

DARIO SANFILIPPO

*LIES (distance/incidence) 1.0*

ALESSANDRO FOGAR

FRANCESCA BERGAMASCO

*Hyle*

MARCELLA MANDANICI

*Score 1. per Disembodied voices*

CARLO MARZAROLI

FRANCESCO GARBO

FEDERICO MULLNER

MARCO GODEAS

*Gestual Interpolation Patch*

MIHA CIGLAR

*M.I.S.T.*

**sabato 24 novembre 2012 / Saturday, November 24, 2012**

**Conservatorio "G. Tartini"**

---

**Presentazioni orali / Paper sessions**

**VII sessione / Session 7**

chair Domenico Sciajno (*Conservatorio di Trapani*)

- 10.00 MARCELLA MANDANICI  
*La micropolifonia ligetiana come modello di composizione algoritmica*
- 10.25 GIACOMO ALBERT  
*Multimodalità e Immersione: verso una rilettura del Minimalismo attraverso la storia delle installazioni sonore*
- 10.50 GIORGIO KLAUER  
*Analisi paradigmatica di composizioni acustiche e music information retrieval: un primo approccio*

11.15 **Coffee break**

**VIII sessione / Session 8**

chair Gualtiero Volpe (*Università di Genova*)

- 11.45 ANDREA ARCELLA  
STEFANO SILVESTRI  
*Analogique B – A computer model of the compositional process*
- 12.10 ANDREA AGOSTINI  
DANIELE GHISI  
*Bach: una libreria per la composizione assistita in Max*

13.00 **Video-listening room\* & Poster**

13.00 **Lunch break**

**IX sessione / Session 9**

chair Stefano Delle Monache (*Università IUAV di Venezia*)

- 14.45 DIMITRIOS MARGOUNAKIS  
DIONYSIOS POLITIS  
*Revealing the hidden motives of chromaticism in music through computational semantic visualizations*
- 15.10 SIMONE SPAGNOL  
*Are spectral elevation cues in head-related transfer functions distance-independent?*
- 15.35 DIMITRIOS MARGOUNAKIS  
DIONYSIOS POLITIS  
*Exploring the relations between chromaticism, familiarity, scales and emotional responses in music*

## Casa della Musica

---

20.00

### **Concerto audiovisivo / Audiovisual concert**

coordinatore Paolo Pachini (*Conservatorio di Trieste*)

IVAN PENOV

*Edgeless*

ALESSANDRO PERINI

FABIO MONNI

*Secret Lives*

FABRIZIO FIORE

*Hi.La*

LUCA DE SIENA

ANTONELLO BELGRANO

*Concrezione*

JEROD SOMMERFELDT

*Linear*

RICHARD NELMES

DIEGO GARRO

*Visitation Adagio*

TIZIANO BOLE

*Lontano giace il mondo*

MAURIZIO GOINA

SARAH TAYLOR

*Passo a sei*

---

FABRIZIO FIORE *regia del suono*

**giovedì 22 e venerdì 23 novembre / Thursday, 22 & Friday, 23 November**  
**Conservatorio “G. Tartini”**

---

**\* Video-listening room**

coordinatore Paolo Pachini (*Conservatorio di Trieste*)

LLOYD DUNN

*55505 pareidolia*

SILVIA LANZALONE

*Alba d'oro*

IVAN PENOV

*Entropy swing*

JUN MIZUMACHI

YNGVILD ROLLAND

*Location 2: flashes of time*

---

I seguenti lavori sono stati selezionati dal Comitato scientifico e musicale, ma, per ragioni diverse, gli autori non hanno potuto presentarli al Colloquio:

*The following works were accepted by the Scientific and Musical Committee, however, for different reasons, they could not be presented at the Colloquium:*

LUCA DE SIENA

*La Messa Elettroacustica: L'utilizzo dell'Ordinarium Missae in alcune composizioni elettroacustiche (paper)*

GIOVANNI COSTANTINI

MASSIMILIANO TODISCO

RENZO PERFETTI

*Automatic piano music transcription system (paper)*

GIOVANNI COSTANTINI

MASSIMILIANO TODISCO

GIORGIO NOTTOLI

*A kinematic sensor to expressively control musical processing (paper)*

ALESSANDRO PERINI

CRISTINA ROVEREDO

BRIAN BURGAN

*“Con\_tatto”: spazializzare la sensazione tattile (poster)*

SIMONE PAPPALARDO

*Room\_3327 (Concert/live-performance for two viewers at a time)*

KYLE EVANS

*de/Rastra (interactive performance)*

COSTANTINO RIZZUTI

*Advaita (flauto e live electronics)*

GRÉGOIRE LORIEUX

*Description du blanc (fisarmonica e live electronics)*

of the 19th CIM, Trieste, November 21-24,

# Contributi *Papers*



# Interactive sound synthesis by the Lotka-Volterra population model

Federico Fontana

Università di Udine

Dipartimento di Matematica e Informatica  
via delle Scienze, 206 – 33100 Udine, Italy  
federico.fontana@uniud.it

## ABSTRACT

The predator-prey population model based on the Lotka-Volterra system of nonlinear differential equations is here solved using discrete-time signal processing techniques. This approach makes the model solution appealing for interactive sound synthesis purposes, meanwhile making its numerical integration easy to understand for the sound designer and the music composer. The resulting digital filter network requires to compute two delay-free loops sharing a signal multiplier, in practice realizing an implicit solver: this family of solvers provides robustness in front of the structural instability of the Lotka-Volterra system even in presence of abrupt changes of the evolutionary parameters, hence affording interactive model control. A real-time implementation of the resulting realization is provided in form of a Java external for the software Puredata.

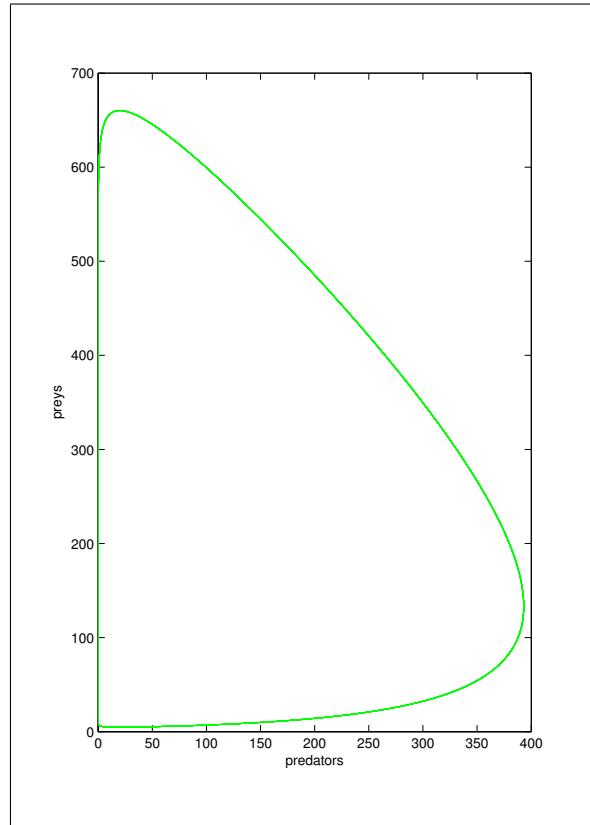
## 1. INTRODUCTION

In 1920, Lotka [1] proposed a set of coupled auto-catalytic chemical reactions whose reactants oscillated around an average concentration. The dynamics of such reactants was described by the following system of ordinary differential equations:

$$\begin{cases} x'(t) = k_x x(t) - k_m x(t)y(t) \\ y'(t) = k_m x(t)y(t) - k_y y(t) \end{cases}, \quad (1)$$

in which  $x$  and  $y$  are the concentrations of the respective reactants across time  $t$ . Six years after Lotka's research, Volterra [2] came to the same differential equation system while independently working on the predator-prey model of a biological population, trying to give sense to fish concentrations measured off the Trieste bay before and after the First World War, when the regular fishing activity had come to a stop due to the warfare.

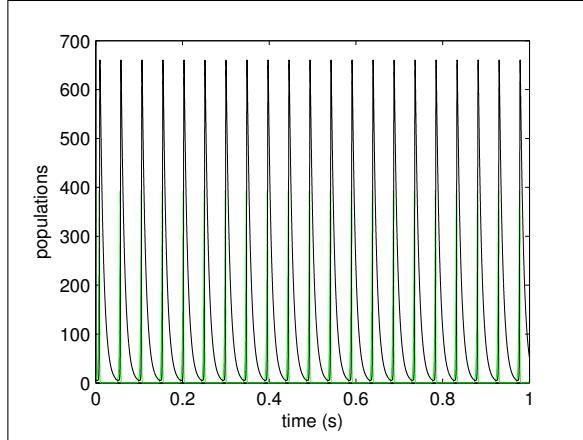
The trade-off between simplicity of the system and significance of the evolutionary solution has fertilized much research around this model, and makes it appealing still today. The Lotka-Volterra system yields strictly positive solutions that indefinitely oscillate for *any* choice of the positive evolutionary parameters  $k_x$ ,  $k_y$  and  $k_m$ , around an



**Figure 1.** Trajectories of (1) for  $k_x = 800$ ,  $k_m = 6$ ,  $k_y = 120$ ,  $x(0) = 19$ , and  $y(0) = 126$ .

equilibrium point which is determined by the parameters themselves. The system (1) does not yield an analytic solution, but for small deviations from the equilibrium point [3]. Furthermore, any numerical solver must deal with the *structural instability* affecting the Lotka-Volterra dynamics: due to this instability, even small perturbations caused by numerical errors and inaccuracies during the integration introduce irreversible deviations of the solution *trajectories* (see Figure 1) from the unperturbed dynamics. This inherent fragility of the solution implies that appropriate, usually implicit numerical solvers must be employed to integrate the system once initial conditions for  $x$  and  $y$  are set.

A filter-based solver for the Lotka-Volterra model has not yet been proposed in the literature. Although accurate numerical integrations already exist for (1), based on e.g. Runge-Kutta or multi-step schemes, here we make use



**Figure 2.** Solution of (1), using the same parameters and initial state as in Figure 1 (green: preys; black: predators).

of formal tools that are more familiar to the dsp engineer. In this way, we envisage its use for musical or everyday sound synthesis. In this specific application field the proposed model may add to a growing corpus of research and applied work, in which evolutionary systems are employed in both generation [4] and control [5] of the sound.

In the next sections it will be shown that, once removing the offset and bounding the trajectories within normalized amplitude ranges, the Lotka-Volterra system provides interesting outputs in a sonic perspective.

## 2. DISCRETE-TIME SYSTEM SOLUTION

Once the system (1) is integrated, it provides closed trajectories for  $x$  and  $y$  describing the temporal oscillations of the populations  $x(t)$  and  $y(t)$ , as shown in Figure 2. Instead of resorting to traditional implicit solvers, we proceed by defining the predator-prey product  $m(t) = x(t)y(t)$ . Then, by Laplace transforming (1) [6] we obtain:

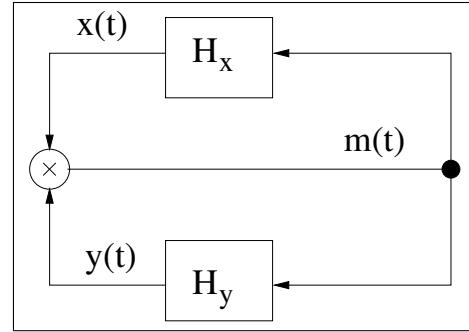
$$\begin{cases} sX(s) &= k_x X(s) - k_m M(s) \\ sY(s) &= k_m M(s) - k_y Y(s) \end{cases}, \quad (2)$$

or, equivalently,

$$\begin{cases} X(s) &= -\frac{k_m}{s-k_x}M(s) = H_x(s)M(s) \\ Y(s) &= \frac{k_m}{s+k_y}M(s) = H_y(s)M(s) \end{cases}. \quad (3)$$

This way, two linear filters  $H_x$  and  $H_y$  are figured out which are both fed by the signal  $m(t)$ , and whose respective outputs form the same signal once they are multiplied each other. Figure 3 illustrates the filter network that models (3).

This network has two notable properties: i) the oscillatory behavior arises in consequence of the existence of a *delay-free* feedback loop, conveying the signal  $m(t)$  to the filters  $H_x$  and  $H_y$ ; ii) by modeling the exponential reproduction of preys,  $H_x$  is an *unstable* linear filter. The latter fact in particular gives reason of the structural instability of the overall system, whose divergent behavior caused by the instability of  $H_x$  is compensated by the concurrent damping action of  $H_y$ , accounting for the predators' activity.



**Figure 3.** Filter network modeling (1).

Only if  $x$ ,  $y$  and  $m$  are computed one instantaneously from the others can the integration error be kept under control. This is the reason why explicit methods fail to converge.

We transform the analog transfer functions,  $H_x(s)$  and  $H_y(s)$ , to the discrete-time domain. Whatever the analog-to-digital operator used in this transformation, both such transfer functions inform discrete-time filters,  $H_x(z)$  and  $H_y(z)$ , whose state at every temporal step  $nT$  is known at the beginning of the computation. For this reason, we denote with  $q[n-1]$  the contribution of the state to the respective output at step  $nT$ , with  $F_s = 1/T$  the sampling rate of the real-time simulation. By recalling that  $H_x(z) = X(z)/M(z)$  and  $H_y(z) = Y(z)/M(z)$ , we have:

$$\begin{cases} x[n] &= a_x m[n] + q_x[n-1] \\ y[n] &= a_y m[n] + q_y[n-1] \end{cases}. \quad (4)$$

The network in Figure 3 can be computed in the discrete-time thanks to the existence of a generalized approach to the nonlinear delay-free loop problem [7, 8]. Among other results, this approach has already been successfully used to figure out efficient and accurate solutions of vacuum-tube guitar amplifier [9] as well as voltage-controlled filter [10] analog models.

Since  $m[n] = x[n]y[n]$ , by multiplying each other the equations in (4) we obtain the following 2nd-order equation in the  $m$  variable:

$$\underbrace{a_x a_y}_a m^2 + \underbrace{(a_x q_y + a_y q_x - 1)}_b m + \underbrace{q_x q_y}_c = 0, \quad (5)$$

whose roots can be calculated in closed form for every analog-to-digital transformation, once  $a$ ,  $b$ , and  $c$  have been figured out. By using in particular the *bilinear* transformation

$$s \leftrightarrow \frac{2}{T} \frac{z-1}{z+1} = \phi \frac{z-1}{z+1}, \quad (6)$$

then  $H_x(z)$  and  $H_y(z)$  preserve the order of the analog counterparts. After some algebraic manipulation we find:

$$\begin{aligned} a_x &= \frac{k_m}{k_x - \phi} \\ a_y &= \frac{k_m}{k_y + \phi} \end{aligned} \quad (7)$$

and

$$\begin{aligned} q_x[n] &= \frac{k_m}{k_x - \phi} m[n] - \frac{k_x + \phi}{k_x - \phi} x[n] \\ q_y[n] &= \frac{k_m}{k_y + \phi} m[n] - \frac{k_y - \phi}{k_y + \phi} y[n]. \end{aligned} \quad (8)$$

In conclusion, at every temporal step  $nT$  the discrete-time values  $x[n]$  and  $y[n]$  are computed through the following procedure, based on the current parameter values  $(k_x, k_y, k_m)$  and state  $(q_x[n-1], q_y[n-1])$ :

1. compute  $a$ ,  $b$ , and  $c$  in (5) from (7) and the values  $q_x[n-1]$  and  $q_y[n-1]$ ;
2. compute the positive root  $m[n]$  of (5);
3. compute  $x[n]$  and  $y[n]$  by (4);
4. update the state, by computing  $q_x[n]$  and  $q_y[n]$  through (8).

Figures 1 and 2 have been obtained by setting  $k_x = 800$ ,  $K_m = 6$ ,  $K_y = 120$ ,  $x(0) = 19$ , and  $y(0) = 126$ , and by iterating on the above procedure.

### 3. INTERACTIVE SOUND SYNTHESIS

Numerical stability of the solutions has been checked in practice, by testing the proposed solver against simple Euler and Runge-Kutta, across ranges of parameters and initial states covering spectra that are interesting for audio synthesis. In such ranges the filter-based integration provided the same solution as that computed by Runge-Kutta, whereas the output from simple Euler followed trajectories that often went progressively far from the two previous solutions.

The number of preys and predators across time can be turned into a pair of audible signals once the dc component is removed; furthermore, their intensity must be normalized within unity values.

- Concerning the dc value, the literature provides the average number of predators and preys living in a Lotka-Volterra system [3]: they respectively amount to  $k_x/k_m$  and  $k_y/k_m$ .<sup>1</sup> Hence, instead of post-processing the output signals through a dc killer, it is more efficient to offset the predator and prey numbers by the respective average value at runtime.
- As opposed to the dc component, there is no explicit formula linking the predator and prey peak values to the evolutionary parameters. Storing range information in the form of table look-ups mapping these parameters onto corresponding normalization factors is costly, in terms of memory and time to lookup. This issue often appears when multidimensional (in this case 3D) tables must be stored in the machine memory: even if their size can be halved due to the symmetry of the Lotka-Volterra system, yet an accurate

<sup>1</sup> The latter average can be interpreted in this way: the more predators proportionally dying and/or the less collisions between predators and preys, the more preys populating the ecosystem. The former average value can be interpreted symmetrically.

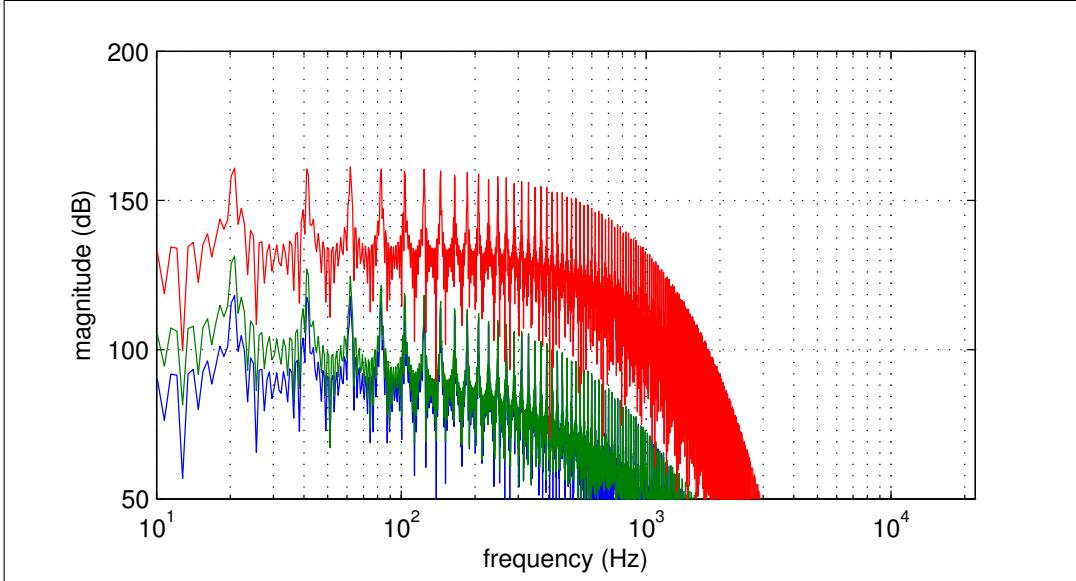
solution of the range normalization based on table look-ups requires a specific effort for the real-time hardware.

Rather, we chose to estimate the range at runtime: by triggering the end of the ascending phase in both the predator and prey signals — this corresponds to numerically evaluate the change of sign in the first derivative, for instance by computing the difference  $x[n] - x[n-1]$  — it is immediate to figure out the peak values in the corresponding signals and normalize them accordingly. This technique is potentially problematic, for it exposes to audible transients due to abrupt scaling of the signal, occurring in between adjacent oscillations that are widely different in amplitude. In practice, it can be adopted in all cases when the evolutionary parameters are changed smoothly enough.

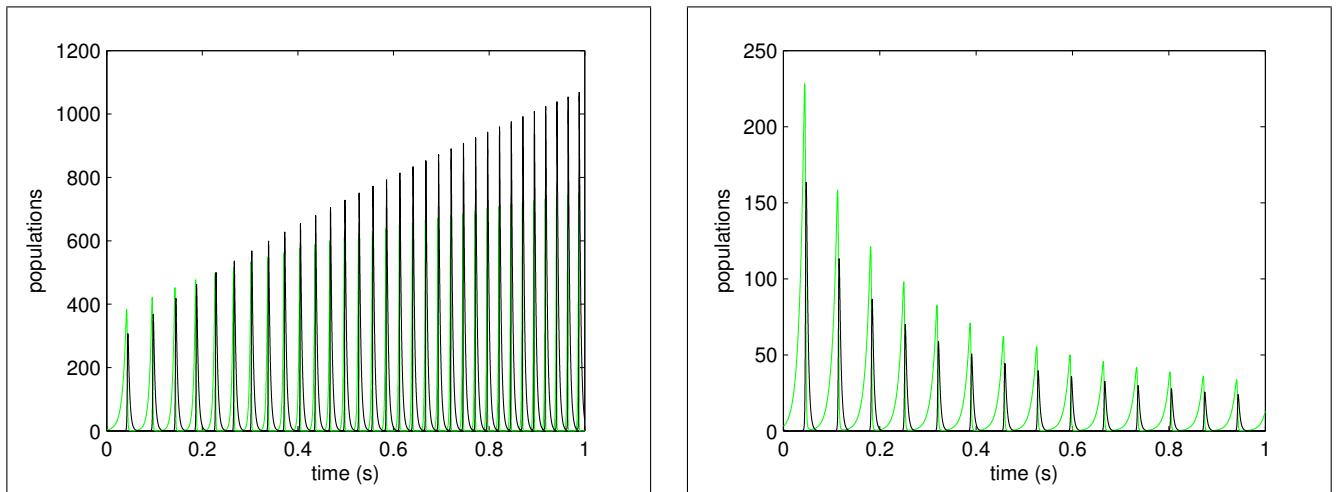
Another more restrictive yet robust solution for keeping the range under control consists of using the prey signal as a predictor of the amplitude of the predator signal. Since both signals result from a lowpass filtering of their product  $m$ , their spectra contain the same components—refer to Figure 4. Hence, it is not crucial to reproduce both such signals. Now, forcing  $k_x = k_y$  corresponds to generating two signals  $x$  and  $y$  having the same range: since the number of preys grows to the peak before the predators reach the largest population, then the range for the predators can be predicted from the preys' evolution. As we will see in the following paragraph, the above constraint on the evolutionary parameters does not pose heavy limits for the purpose of sound synthesis.

Figure 5 illustrates the evolution of the Lotka-Volterra system when the reproduction parameter  $k_x$  of the preys is linearly increased. It can be seen that not only the number of the preys (and, consequently, the predators) becomes larger: in parallel, the population peaks occur with higher frequency. Similar effects on the population are encountered if the parameter  $k_y$  is changed: the more predators die, the larger both peaks of the population and, proportionally, the frequency when such peaks occur. For this reason, imposing the constraint  $k_x = k_y$  does not turn in considerable loss of sonic expressive power of the Lotka-Volterra model. In conclusion, once the peak amplitudes are kept constant by the synthesis algorithm (for instance using one of the techniques mentioned above), then  $k_x = k_y$  can be used as a simple, nonlinear pitch control.

One may wonder what happens when the evolutionary parameter  $k_m$  is manipulated. It is convenient to note that this parameter does not change the pole position in either lowpass filter forming the system. As Figure 6 illustrates, reinforcing the “coupling” between preys and predators through the linear increase of  $k_m$  results, perhaps counter-intuitively, in an exponential decrease of their respective peaks furthermore with no appreciable changes in the corresponding frequencies. In presence of amplitude peak normalization, hence, the effect of  $k_m$  is negligible apart from subtle changes in the timbre, resulting from minor



**Figure 4.** Magnitude spectra of preys, predators, and their product using the same parameters and initial state as in Figure 1 (green: predators; blue: preys; red: product).



**Figure 5.** Solution of (1), with  $k_x$  linearly increasing from 100 to 1000,  $k_m = 5$ ,  $k_y = 300$ , and initial state as in Figure 1 (green: preys; black: predators).

**Figure 6.** Solution of (1), with  $k_x = 100$ ,  $k_m$  linearly increasing from 5 to 50,  $k_y = 300$ , and initial state as in Figure 1 (green: preys; black: predators).

shape modifications of both the predator's and prey's waveforms.

Perhaps interestingly,  $k_m$ , for its effect on the amplitude peaks, can be used as a signal magnitude control by the sound designer who prefers to disable the automatic peak normalization, and instead operate directly on the whole set of parameters: a good compromise, in this case, is that of using  $k_x = k_y$  to control the pitch meanwhile compensating its effect on the magnitude through  $k_m$ . The action on both such controls, if operated jointly with a traditional output gain, enables the sound designer to make full use of the expressive possibilities of the Lotka-Volterra model.

#### 4. CONCLUSIONS

The proposed model has been implemented as an *external* Java code for the real-time synthesis software Pure-data, using the public pdj libraries [11]. The code runs at  $F_s = 44.1$  kHz without encumbering an 2.5 GHz Intel Core 2 Duo laptop machine running Ubuntu Linux. The spectra of the Lotka-Volterra dynamics exhibit a good density of components, making the computational effort worth being done. Further post-processing stages may of course be designed at the sound designer's taste, for making best use of the signal coming out from the predator-prey model.

Besides its value as a sound synthesizer, the algebraic technique behind the solution of the proposed filter network can be generalized to a class of topologies, characterized by the existence of one geometrical nonlinearity

```

public Method dsp(MSPSignal[] ins, MSPSignal[] outs) {
    kx = 800;
    km = 6;
    ky = 120;
    x = 19;
    y = 126;
    m = x0*y0;
    Fs = ins[0].sr;

    qx = km/ (kx-2*Fs)*m - (kx+2*Fs)/ (kx-2*Fs)*x;
    qy = km/ (ky+2*Fs)*m - (ky-2*Fs)/ (ky+2*Fs)*y;

    return getPerformMethod("iterate");
}

public void iterate(MSPSignal[] ins, MSPSignal[] outs) {
    float[] out = outs[0].vec;
    int vec_size = ins[0].n;
    double newPrey, newPredators;
    for (int i=0; i<vec_size; i++) {

        a = Km*Km/ ((Kx-2*Fs)*(Ky+2*Fs));
        b = Km/ (Kx-2*Fs)*qy + Km/ (Ky+2*Fs)*qx - 1;
        c = qx*qy;

        m = (-b-Math.sqrt(b*b-4*a*c))/2/a;

        x = Km/ (Kx-2*Fs)*m + qx;
        y = Km/ (Ky+2*Fs)*m + qy;

        qx = km/ (kx-2*Fs)*m - (kx+2*Fs)/ (kx-2*Fs)*x;
        qy = km/ (ky+2*Fs)*m - (ky-2*Fs)/ (ky+2*Fs)*y;

        out[i] = (y - kx/km) / kx;
    }
}

```

(say, a signal multiplier) shared by two delay-free loops each, in principle, encoding whatever linear transfer function. This generalization is evident from the inspection of Figure 3 and the 2nd-order polynomial (5). In this sense, the proposed technique may find profitable use and rewarding research time in the filter-based modeling of other differential equation systems: more elaborate predator-prey models [12] and other circuits built around signal multiplication, such as ring modulators [13], may fall within this class.

The direct manipulation of the system parameters translates in some useful controls, especially on pitch and magnitude. Ultimately, it must be remembered that the Lotka-Volterra system *is* an oscillator and, although providing a relatively rich waveform, exhibits a traditional parametric dependence which is similar to that of other nonlinear oscillators, such as those synthesizing triangular or sawtooth waves.

The core of the external Java code for the synthesis of sound is reported in the box below, limitedly to the signals

in Figure 2. The complete real-time Java external enabling interactive control of the Lotka-Volterra model is available upon request to the author.

## 5. REFERENCES

- [1] A. J. Lotka, “Undamped oscillations derived from the law of mass action,” *J. Am. Chem. Soc.*, vol. 42, pp. 1595–1599, 1920.
- [2] V. Volterra, “Fluctuations in the abundance of a species considered mathematically,” *Nature*, vol. 118, pp. 558–560, 1926.
- [3] R. Illner, C. S. Bohun, S. McCollum, and T. van Roode, *Mathematical Modelling*. Providence, RI: American Mathematical Society, 2005.
- [4] P. Dahlstedt, *Sounds Unheard of: Evolutionary Algorithms as Creative Tools for the Contemporary Com-*

poser. PhD thesis, Chalmers University of Technology, 2004.

- [5] J. M. McDermott, *Evolutionary Computation Applied to the Control of Sound Synthesis*. PhD thesis, University of Limerick, Limerick, Ireland, 2008.
- [6] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1989.
- [7] D. T. Yeh, J. S. Abel, A. Vladimirescu, and J. Smith, “Automated physical modeling of nonlinear audio circuits for real-time audio effects – part i: Theoretical development,” *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 728–737, 2010.
- [8] F. Fontana and F. Avanzini, “Computation of delay-free nonlinear digital filter networks. Application to chaotic circuits and intracellular signal transduction,” *IEEE Trans. on Signal Processing*, vol. 56, pp. 4703–4715, Oct. 2008.
- [9] J. Pakarinen and D. T. Yeh, “A review of digital techniques for modeling vacuum-tube guitar amplifiers,” *Computer Music Journal*, vol. 33, pp. 1–16, summer 2009. In press.
- [10] F. Fontana and M. Civolani, “Modeling of the EMS VCS3 voltage-controlled filter as a nonlinear filter network,” *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, no. 4, pp. 760–772, 2010. Special Issue on Virtual Analog Audio Effects and Musical Instruments.
- [11] “pdj.” Web page, July 2012. Web published at [www.leson666.com](http://www.leson666.com).
- [12] S. Q. Liu, E. Beretta, and D. Breda, “Predator-prey model of Beddington-DeAngelis type with maturation and gestation delays,” *Nonlinear Analysis. Real World Applications*, vol. 11, no. 5, pp. 4072–4091, 2011.
- [13] R. Hoffmann-Burchardi, “Digital simulation of the diode ring modulator for musical applications,” in *Proc. Conf. on Digital Audio Effects (DAFX-08)*, (Espoo, Finland), pp. 165–168, Sep. 1-4 2008.

# FRUIZIONE EVOLUTA VIA WEB DI CONTENUTI MUSICALI E MULTIMEDIALI: IL PORTALE EMIPIU

Stefano Baldan, Adriano Baratè, Goffredo Haus, Luca A. Ludovico

Laboratorio di Informatica Musicale

Dipartimento di Informatica

Università degli Studi di Milano

{baldan, barate, haus, ludovico}@di.unimi.it

## ABSTRACT

Il presente lavoro descrive il portale EMIPIU, un'interfaccia Web per la fruizione evoluta di contenuti musicali. Esso si basa sulle caratteristiche del formato IEEE 1599, che consente di codificare l'informazione musicale eterogenea relativa ad un singolo brano all'interno di un unico documento XML. Il portale fornisce accesso a una libreria di pezzi in formato IEEE 1599, rappresentativi di diversi periodi storici, stili e organici strumentali. Per la loro fruizione è stato realizzato un lettore multimediale IEEE 1599 in HTML5, le cui peculiarità permettono di scaricare in streaming e visualizzare flussi multimediali multipli. L'articolo descrive nel dettaglio il progetto, soffermandosi sulle caratteristiche principali del formato IEEE 1599, del portale EMIPIU, della teca digitale e del player.

## 1. INTRODUZIONE

I modelli di fruizione dei contenuti musicali sono notevolmente cambiati grazie all'adozione delle tecnologie informatiche. Solo per citare un esempio, è sufficiente prendere in considerazione la cosiddetta "musica liquida" (si veda [1]) e considerare l'impatto che questa ha avuto sui formati, sulla diffusione in rete dei contenuti e sulla progettazione di hardware apposito. Un aspetto cui dedicare ancora molta ricerca riguarda una codifica dei contenuti musicali adeguata ad innovativi metodi di fruizione. Ad esempio, esistono archivi di vaste dimensioni che potrebbero essere ri-vivificati, innanzi tutto avviando opportune campagne di digitalizzazione e in secondo luogo organizzando i contenuti in modo efficiente ed efficace, nell'ottica di una loro conservazione, valorizzazione e fruizione su larga scala. Questa materia è stata già trattata in numerosi lavori, tra cui [2], [3] e [4].

Proprio in questa cornice si colloca il progetto EMIPIU (*Enhanced Music Interactive Platform for Internet User*), che mira ad applicare tecnologie informatiche al fine di valorizzare i beni culturali musicali, con particolare riferimento a teatri, musei, manifestazioni musicali, grandi archivi pubblici e privati, istituzioni per l'educazione musi-

Copyright: ©2012 Stefano Baldan, Adriano Baratè, Goffredo Haus, Luca A. Ludovico 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.

cale. Tali soggetti possono trarre nuovo impulso dall'adozione delle tecnologie proposte nel seguito dell'articolo e adottarle al fine di valorizzare le risorse disponibili. Possibili applicazioni spaziano dalla fruizione via Web dei contenuti musicali e multimediali all'edizione di nuovi media.

Il progetto EMIPIU nasce dalla collaborazione scientifica tra il *Laboratorio di Informatica Musicale (LIM)* dell'Università degli Studi di Milano<sup>1</sup> e il *Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM)* dell'Université Montpellier<sup>2</sup>. Tra gli obiettivi, esso annovera lo sviluppo di tecnologie musicali e multimediali per la valorizzazione delle teche digitali e per la ri-vivificazione di documenti musicali in formato digitale. Il progetto è stato co-finanziato dalla Regione Lombardia, e vede coinvolto come *main partner* industriale Didael KTS S.r.l.<sup>3</sup>.

Il primo passo per raggiungere gli obiettivi del progetto consiste nell'ideare e implementare strumenti per la creazione, gestione e fruizione di documenti musicali in formato IEEE 1599. Quest'ultimo rappresenta uno standard internazionale per la codifica dell'informazione musicale multi-livello in XML, le cui principali caratteristiche verranno trattate nel Paragrafo 2. In seguito alla standardizzazione del formato, lo staff del LIM sta lavorando alla piattaforma tecnologica per la produzione di documenti IEEE 1599. Trattandosi di informazioni complesse, che comprendono simboli musicali, testi, immagini statiche, audio e video, al momento la codifica di brani in formato IEEE 1599 risulta un'operazione piuttosto lenta e laboriosa. La piattaforma tecnologica in fase di implementazione mira ad abbattere i tempi di produzione dei materiali IEEE 1599 mediante processi di ottimizzazione e di automatizzazione.

Una volta codificati i materiali musicali in maniera opportuna, come mostrato nel Paragrafo 3, è necessario organizzarli all'interno di una teca digitale e renderli fruibili. Per massimizzare la loro reperibilità e diffusione, includendo calcolatori elettronici così come dispositivi mobili, si è optato per la realizzazione di un portale Web. Sarà questo l'argomento principale del Paragrafo 4, in cui verranno illustrati i contenuti e le caratteristiche del portale EMIPIU.

<sup>1</sup> <http://www.lim.di.unimi.it>

<sup>2</sup> <http://www.lirmm.fr>

<sup>3</sup> <http://www.didaelkts.it>

In questo ambito, ha una particolare importanza il modulo messo a punto dal LIM per la fruizione via Web di documenti IEEE 1599. Tale strumento introduce modelli innovativi per la visualizzazione e per l'interazione con i contenuti musicali, come illustrato nel dettaglio nel Paragrafo 5. Il lettore multimediale di documenti IEEE 1599 deve inoltre affrontare e risolvere rilevanti problematiche tecniche quali la compatibilità con i più diffusi browser e la gestione simultanea di flussi multipli. Questo argomento sarà affrontato nel corso del Paragrafo 6.

## 2. CENNI SUL FORMATO IEEE 1599

L'IEEE 1599 è un formato di rappresentazione musicale, standardizzato presso l'IEEE nel settembre del 2008. Il suo sviluppo ha seguito le direttive espresse nell'IEEE P1599 (*Recommended Practice Dealing With Applications and Representations of Symbolic Music Information Using the XML Language*), proponendosi di rappresentare simbolicamente e in modo integrato la musica in tutti i suoi diversi aspetti: simboli musicali, partiture e altre rappresentazioni grafiche, audio, video, ecc.

La codifica IEEE 1599 è basata sullo standard XML; di quest'ultimo condivide perciò alcune caratteristiche peculiari: leggibilità, strutturazione, semplicità di editing.

Per permettere di codificare i vari aspetti dell'informazione musicale all'interno di un unico documento, il formato IEEE 1599 organizza il materiale in sei strati o livelli (in inglese *layer*):

- *General* - metadati e informazioni catalografiche;
- *Logic* - descrizione logica dei simboli musicali codificati;
- *Structural* - modellazione di oggetti musicali e individuazione delle relazioni reciproche;
- *Notational* - rappresentazioni grafiche della partitura;
- *Performance* - esecuzioni automatiche e sintetiche da parte del computer;
- *Audio* - performance audio/video.

E' opportuno notare che non tutti gli strati debbano necessariamente essere presenti in ogni file IEEE 1599. Naturalmente, maggiore è il numero di strati presenti e di oggetti all'interno di ciascun livello, maggiore è la quantità d'informazione codificata e dunque la ricchezza informativa. A questo proposito, si sottolinea la possibilità di includere sia materiali eterogenei facenti capo a diversi strati, sia oggetti diversi inclusi in un singolo livello, ad esempio diverse edizioni di una partitura o diverse esecuzioni dello stesso brano.

Utilizzando una struttura dati interna denominata *spine*, il formato non solo racchiude all'interno di un unico documento l'informazione eterogenea che può essere collegata a un singolo brano, ma mette a disposizione anche un meccanismo raffinato per sincronizzare i diversi materiali.

Pur trattandosi di codifica in XML, che quindi può essere effettuata anche con applicazioni quali gli editor di testo, la produzione di materiali nello standard IEEE 1599 è agevolata dall'utilizzo di strumenti appositi. Ad esempio, nell'ambito del progetto EMPIIU sono stati messi a punto algoritmi e prototipi software per facilitare e in una certa misura automatizzare i processi di sincronizzazione dei materiali multimediali rispetto all'informazione simbolica ivi contenuta.

Trattandosi di uno standard internazionale, l'argomento IEEE 1599 è stato trattato in numerose pubblicazioni scientifiche. A titolo di esempio, si evidenzia [5] per quanto riguarda le caratteristiche generali dello standard, [6] per le sue ricadute in ambito didattico, [7] per la ri-vivificazione dei beni culturali musicali e infine [8] per quanto concerne l'applicabilità alle performance artistiche *live*.

## 3. I MATERIALI CODIFICATI

Tutti i materiali caricati nella sezione ad accesso non protetto del portale risultano non vincolati da diritti d'autore o da diritti di riproduzione fonomeccanica. Spartiti e partiture sono trascrizioni originali effettuate presso il LIM, digitalizzazioni di materiale non coperto da diritti o infine oggetti digitali tratti da archivi online specializzati. A quest'ultimo proposito merita una particolare menzione l'iniziativa *International Music Score Library Project (IM-SLP)*<sup>4</sup>, nota anche come Petrucci Music Library [9]. Analogamente, gli audio (e talvolta i video) risultano autoprodotti, oppure digitalizzati da materiale su cui sono scaduti i diritti di riproduzione, oppure ricadono nella casistica delle *Creative Commons Public Licenses (CCPL)*.

La libreria di brani è in costante evoluzione, per poter fornire all'utente una sempre più ricca esperienza di ascolto. Come mostrato in Figura 1, all'interno della libreria trovano spazio i più disparati generi musicali, dal canto gregoriano al pop/rock, dalla musica "colta" al canto folcloristico, dalla lirica al jazz. Nella fase iniziale del progetto EMPIIU si è scelto di realizzare una libreria di brani appositamente studiata per evidenziare le caratteristiche e le potenzialità del formato. Si è pertanto considerato aspetto premiante ottenere la massima eterogeneità per quanto concerne il genere musicale, il periodo storico e critico, l'ensemble, la numerosità e la tipologia di oggetti multimediali collegati ai brani. In altri termini, la libreria prototipale già a disposizione degli utenti ha uno scopo principalmente dimostrativo. La codifica di opere complete (ad esempio, tutti i preludi per pianoforte Op.10 di Chopin) caratterizza invece la seconda fase di arricchimento della teca, attualmente in corso, che porterà entro Ottobre 2012 il numero di documenti IEEE 1599 consultabili a circa un centinaio.

Si analizzano ora alcuni dei materiali codificati, significativi per cogliere l'eterogeneità dell'informazione e il numero di differenti campi applicativi.

*L'Ave Maria* di J.S. Bach e C. Gounod presenta quattro esecuzioni, di cui due "tradizionali" (del soprano Montserrat Caballé e del tenore Andrea Bocelli), e altre due meno

<sup>4</sup> <http://www.imslp.org>

convenzionali (di Antonella Ruggiero per voce non impostata e del duo Bobby McFerrin e Yo-Yo Ma per voce e violoncello solo). Questo esempio è interessante in quanto pone a confronto e sincronizza materiali audio estremamente eterogenei. Un caso simile è dato dal lied *Gottes Macht und Vorsehung* di L. van Beethoven, di cui vengono proposte una versione per canto e piano, una sintetizzata in formato MIDI e una con accompagnamento d'organo. Quest'ultima presenta un video anziché una semplice traccia audio.

Il brano *Il mio ben quando verrà*, aria tratta da *Nina o sia La pazza per amore* di G. Paisiello, è rilevante in quanto fornisce invece una molteplicità di versioni della partitura: accanto all'autografo dell'autore e a una copia manoscritta d'epoca, si trova una versione a stampa del libretto e una riduzione per canto e pianoforte. Sempre dal punto di vista notazionale, è interessante anche l'esempio dell'*Introitus* dalla prima messa *In Nativitate Domini*, che presenta sia la notazione neumatica di un graduale antico, sia la versione a stampa sempre in notazione quadra, sia infine una trascrizione moderna. Anche le intavolature per liuto sono supportate dal formato IEEE 1599, come dimostrato dal *Prélude* dalla *Suite n. 3* di S.L. Weiss un esempio di intavolatura per liuto.

*Lullaby of Birdland* di G. Shearing mostra l'applicabilità della codifica IEEE 1599 alla musica jazz, così come il *Maple Leaf Rag* di S. Joplin rappresenta un'incursione nel ragtime, *Por una Cabeza* di C. Gardel nel tango, *Ul parisien* di F. Paggia nella musica folkloristica.

Grazie alle tecnologie messe a punto presso il LIM e convogliate nel framework, è possibile arricchire a piacimento la teca digitale esistente realizzando materiali *ad hoc* per esigenze specifiche, finalizzati ad esempio alla didattica o all'alta formazione musicale.

#### 4. IL PORTALE EMIPIU

Nell'ambito del progetto EMIPIU, è stato realizzato un portale per rendere pubblicamente accessibili i contenuti della teca digitale descritta nel Paragrafo 3. La homepage, mostrata in Figura 2, è raggiungibile all'indirizzo <http://emipiudico.unimi.it>. Il portale è stato realizzato su indicazioni del LIM da parte di Didael KTS, il cui specifico background ICT ha consentito di prestare attenzione ad aspetti quali l'usabilità e la coerenza grafica.

Il portale risulta suddiviso in quattro sezioni:

- *EMIPIU* - La prima area è dedicata alla descrizione del progetto e ne illustra le ricadute nell'ambito della multimedialità, della didattica e della valorizzazione dei beni culturali.
- *IEEE 1599* - La seconda sezione, di natura più tecnica, illustra le caratteristiche del formato IEEE 1599, adottato per codificare i brani musicali nel portale. Si veda a questo riguardo quanto scritto nei Paragrafi 2 e 3.
- *Music Box* - La terza area è probabilmente quella di maggior impatto per l'utente finale. Essa ospita

Figura 1. Alcuni brani codificati in IEEE 1599 e presenti nel portale.

Figura 2. Screenshot della homepage del portale.

un media player che realizza via Web le funzionalità prima descritte, e in particolare la rappresentazione sincronizzata di contenuti multimediali eterogenei (partitura, audio, video, ecc.). Grazie ai contenuti della sezione *Music Box*, il fruitore diventa protagonista attivo della propria esperienza musicale. Questi argomenti verranno trattati più nel dettaglio nel corso dei prossimi paragrafi.

- *Community* - La quarta ed ultima sezione rappresenta l'area di social network del progetto. Essa permette di scambiare opinioni con altri utenti e con gli autori della piattaforma, di chiedere delucidazioni sulle tecnologie adottate, di condividere contenuti multimediali.

Il portale si rivolge a tutti gli amanti della musica, a vario livello di competenza e di formazione. I destinatari di tale piattaforma spaziano dall'ascoltatore occasionale al professionista, dal semplice appassionato al musicologo, e più in generale includono chiunque desideri avvicinarsi alla musica in un modo innovativo, interattivo e multimediale.

Come accennato, lo scopo primario è fornire all'utente Web un'interfaccia per la visualizzazione evoluta dei contenuti musicali. Tale piattaforma si pone l'obiettivo di rivoluzionare il modello di diffusione e fruizione della musica, non limitandosi al canale audio, ma ampliando l'orizzonte percettivo ad altri aspetti descrittivi di un brano, quali ad esempio la partitura e il video. L'approccio di EMPIIU non solo prevede un'eterogeneità di materiali collegati ad ogni brano, ma si spinge oltre fornendo - ove possibile - una gamma di differenti versioni tra cui l'utente può scegliere. E' quindi possibile confrontare in tempo reale esecuzioni e trascrizioni differenti dello stesso brano, seguendo la partitura in modo sincronizzato con l'audio.

Il modello di erogazione è basato sul Web, al fine di rendere il prodotto accessibile al maggior numero possibile di utenti, indipendentemente dalle caratteristiche hardware e software del sistema in uso. Di conseguenza, qualsiasi dispositivo dotato di un browser compatibile con lo standard HTML5 e di adeguata connessione alla rete è abilitato alla visualizzazione del portale. Tra gli esempi di dispositivi supportati vi sono i personal computer, i notebook, i palmari e i tablet.

Tra gli obiettivi del progetto EMPIIU possiamo annoverare la diffusione dello standard IEEE 1599 attraverso la pubblicazione di una vasta libreria di brani codificati in tale formato, della documentazione relativa al formato stesso e degli strumenti software per la fruizione e per la creazione di nuovi materiali. In questa ottica, la realizzazione del portale si è rivelata un passo di fondamentale importanza. Didael KTS S.r.l. ha realizzato gli *user acceptance test* tramite appositi *focus group*, che hanno evidenziato l'interesse da parte dell'utenza tanto non specializzata quanto professionale in merito ad aspetti differenti della fruizione. Questo argomento verrà trattato nel prossimo paragrafo.

## 5. IL MODELLO DI FRUIZIONE

All'interno del portale EMPIIU, l'area di particolare interesse dal punto di vista della fruizione dei contenuti musicali è la cosiddetta sezione *Music Box* (si veda il Paragrafo 4).

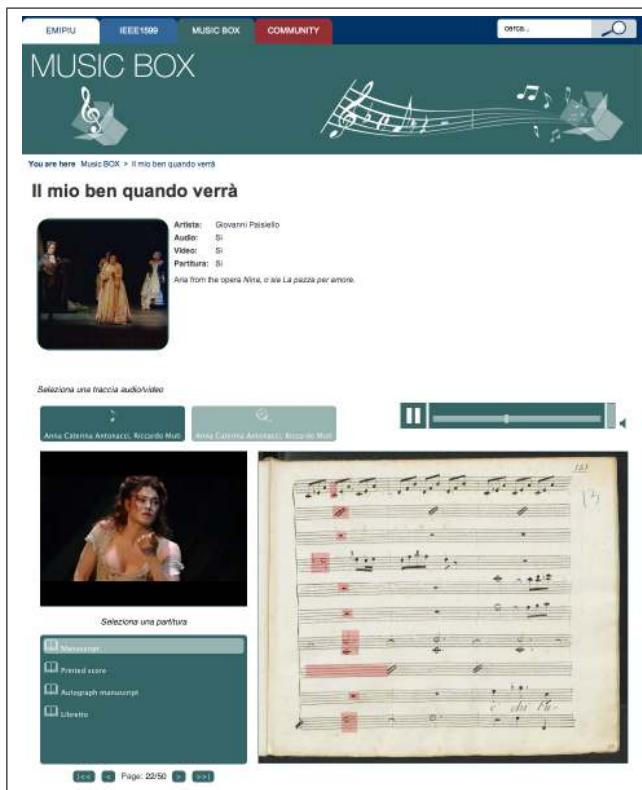
Sulla base delle precedenti esperienze scientifiche maturate presso il LIM, si desidera trasferire nell'interfaccia Web l'approccio già adottato per numerosi applicativi offline, come mostrato in [10].

L'approccio prevede tre distinte modalità di fruizione.

1. La cosiddetta *fruizione tradizionale* rappresenta una modalità di consultazione basilare dei materiali, rivolta anche a utenti senza particolari competenze musicali. Banalmente, tale modello assomiglia ad uno strumento evoluto di *score following*, simile a quelli messi a disposizione dagli editor di partitura. Ad esempio, all'interno dei software di notazione musicale è tipicamente possibile verificare gli inserimenti dei simboli musicali procedendo a un ascolto sincronizzato con la visualizzazione grafica, sincronizzazione mostrata da un'opportuna evidenziazione. L'avanzamento rispetto a tali strumenti già reperibili è dovuto alle possibilità di visualizzare partiture e audio non sintetici, derivanti ad esempio da processi di digitalizzazione di originali analogici. La principale funzionalità prevista è quella di sincronizzazione tra molteplici oggetti multimediali, ciascuno rappresentativo di uno specifico livello.
2. La *fruizione multi-oggetto* si spinge oltre, consentendo la consultazione dei materiali in modo incrociato. In altre parole, si sfruttano le relazioni sia tra oggetti omogenei (all'interno dello stesso livello) sia tra oggetti eterogenei (appartenenti a strati diversi) in modo da arricchire l'esperienza da parte dell'utente. Lo scopo di questa sezione consiste pertanto nel collegare documenti di varia natura e formato, consentendo ad esempio di sostituire in tempo reale la partitura visualizzata o l'audio in esecuzione, a patto che il documento contenga più di un oggetto per livello.
3. La *fruizione interattiva*, infine, intende mettere a disposizione dell'utente una serie di controlli finalizzati a poter alterare il flusso naturale dei materiali multimediali. Ad esempio, ogni evento musicale individuato sulla partitura deve diventare cliccabile, così come la barra di avanzamento sul player audio/video deve permettere l'immediato riposizionamento di tutti i materiali correnti.

L'implementazione delle tre modalità sopra descritte consente una fruizione innovativa dei contenuti musicali. A titolo di esempio, si considerino le Figure 3, 4 e 5. Esse mostrano il prototipo di lettore multimediale implementato nel portale EMPIIU e descritto da un punto di vista tecnico nel prossimo paragrafo. Le tre immagini si rifanno a diverse visualizzazioni dello stesso brano. Figura 3 potrebbe rispondere alla prima modalità di fruizione: l'utente lancia

l'esecuzione del pezzo, e segue i materiali inizialmente caricati (nell'esempio una partitura autografa e un video). La seconda modalità è legata alla presenza di più audio e più partiture per lo stesso brano, il che consente ad esempio di passare alla riduzione per canto e pianoforte della partitura, come mostrato in Figura 4, o ad un'esecuzione audio differente, come mostrato in Figura 5. Si osservino infine gli indicatori che evidenziano l'istante corrente di esecuzione, in termini di tempo per l'audio/video e di posizione fisica per la partitura: agendo sugli opportuni controlli, l'utente può provocare l'istantaneo riposizionamento dei media.

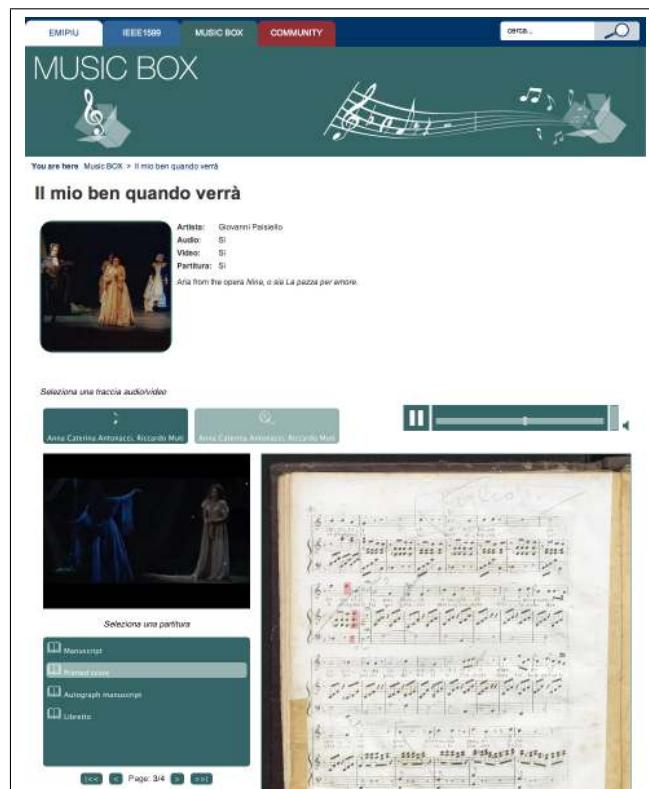


**Figura 3.** Una delle possibili combinazioni di materiali: partitura autografa e video.

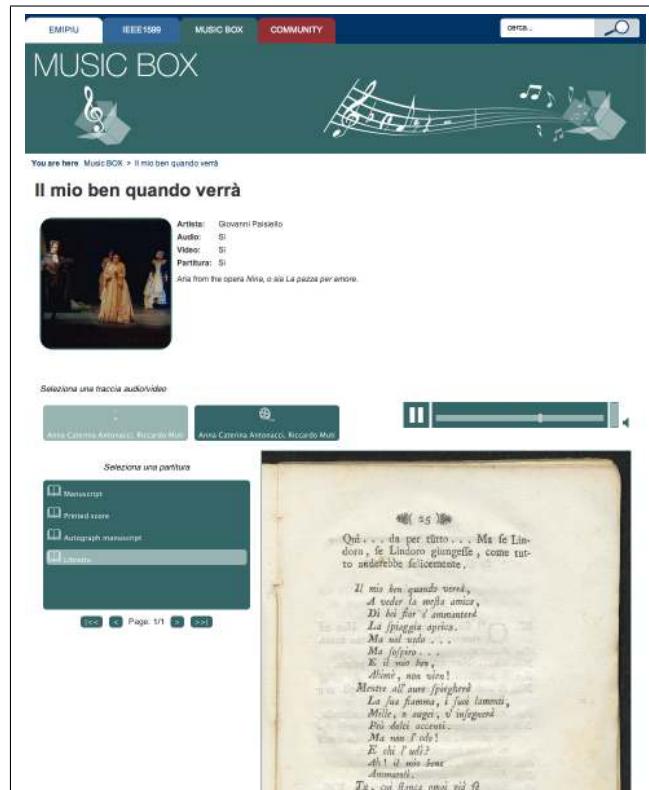
## 6. CARATTERISTICHE TECNICHE DEL LETTORE MULTIMEDIALE

Il lettore multimediale (o *media player*) IEEE 1599 sviluppato per il progetto EMIPU è un'applicazione AJAX, realizzata in linguaggio Javascript utilizzando le nuove API audio/video di HTML5. Tali tecnologie permettono di fornire tutte le funzionalità di sincronizzazione e di fruizione dei materiali audiovisivi sopra descritte all'interno di un comune browser Web, senza bisogno di installare alcun software aggiuntivo o plugin esterno.

I documenti IEEE 1599 possono essere caricati all'interno del portale sia dagli amministratori che dagli utenti, e sono quindi resi disponibili in rete tramite un comune server Web HTTP. All'atto del caricamento viene effettuata una prima fase di rielaborazione del documento, allo scopo di uniformare le codifiche dei vari file multimedia-



**Figura 4.** Un'altra possibile combinazione: riduzione per canto e pianoforte e video.



**Figura 5.** Una terza combinazione: libretto a stampa e audio.

li ad esso associati [11] e di renderle compatibili con le implementazioni di HTML5 fornite dai vari browser Web:

- JPEG per le immagini statiche;
- OGG Vorbis e MP3 per i contenuti audio;
- OGG Theora + Vorbis, WebM + Vorbis e H.264 + AAC per i contenuti video.

Tutte le risorse appartenenti al documento IEEE 1599 da eseguire vengono richieste e spedite contemporaneamente al lettore multimediale. In questo modo il lettore immagazzina nel proprio buffer una certa quantità di dati per ciascun flusso multimediale, rendendo il sistema molto reattivo rispetto alla selezione e alla navigazione dei contenuti da parte dell'utente. Lo svantaggio di questo approccio è il gran traffico generato sulla rete a causa dell'invio di numerosi flussi multimediali in contemporanea. Per risolvere il problema sono state escogitate due ottimizzazioni: in primo luogo l'utilizzo delle funzionalità di *HTTP compression* dei moderni browser Web per spedire il documento IEEE 1599 in formato compresso, tecnica particolarmente efficace con i formati testuali come XML; in secondo luogo l'implementazione di un sistema di *quality switching*, che prevede l'invio di anteprime a bassa qualità ed alto tasso di compressione per i materiali audiovisivi correntemente non utilizzati dall'utente.

La sincronizzazione dei vari contenuti multimediali avviene utilizzando il flusso audio/video selezionato come riferimento temporale: ad ogni istante di esecuzione, il lettore multimediale cerca nel documento IEEE 1599 l'evento corrispondente nella traccia selezionata. Quando un altro flusso viene selezionato il lettore lo carica, cerca per la nuova traccia nel file XML l'evento corrispondente all'ultimo memorizzato, infine usa quest'ultimo per riposizionare l'istante di esecuzione del nuovo flusso selezionato. Il risultato è che l'esecuzione del nuovo contenuto multimediale inizia dallo stesso istante logico in cui il contenuto precedente è stato fermato, ottenendo di fatto la sincronizzazione desiderata.

Per la corretta visualizzazione dei contenuti musicali del portale sono supportati i principali browser presenti sul mercato:

- Internet Explorer 9.0 o superiore;
- Mozilla Firefox 12.0 o superiore;
- Google Chrome 19.0 o superiore;
- Safari 5.1 o superiore;
- Opera 11.0 o superiore.

Il vincolo principale consiste chiaramente nella compatibilità dei browser con lo standard HTML5, relativamente recente e quindi al momento supportato in modo difforme dai diversi software.

Per maggiori approfondimenti dal punto di vista tecnico si rimanda a [12].

## 7. CONCLUSIONI E SVILUPPI FUTURI

Il presente lavoro ha affrontato nel dettaglio diversi aspetti relativi al progetto EMIPIU. Trattandosi di una piattaforma per la fruizione evoluta di contenuti musicali, il primo punto di rilievo è stato individuare un formato le cui caratteristiche si prestassero a tale finalità. In questo senso, lo standard internazionale IEEE 1599 si è rivelato una valida scelta. In secondo luogo, l'articolo ha presentato e opportunamente commentato esempi di documenti contenuti nella teca digitale prototipale del progetto, evidenziando come il formato si presti a codificare brani dalle caratteristiche eterogenee. Infine, l'attenzione si è spostata sulle peculiarità del portale EMIPIU, con particolare riferimento alle tecnologie impiegate per il lettore multimediale.

In conclusione, il progetto EMIPIU ha mostrato i vantaggi di una codifica opportuna dell'informazione musicale, unita alle potenzialità della diffusione via rete dei contenuti e all'adozione di opportune tecnologie per la fruizione. I campi di applicabilità spaziano dalla didattica alla disseminazione, dalla valorizzazione dei beni culturali all'intrattenimento.

EMIPIU è un progetto in continua evoluzione da vari punti di vista. Innanzitutto, il *corpus* di materiali musicali in formato IEEE 1599 viene costantemente incrementato, aggiungendo nuovi brani o aggiungendo ulteriori materiali ai brani già presenti nella libreria. In secondo luogo, le funzionalità stesse di EMIPIU devono necessariamente tener conto delle innovazioni tecnologiche in campo hardware e software. Ad esempio, i nuovi dispositivi mobili potrebbero richiedere adeguamenti alla parte grafica e ai controlli per l'interazione.

## 8. RINGRAZIAMENTI

Il portale EMIPIU è co-finanziato dalla Regione Lombardia ed è stato realizzato in cooperazione con lo staff di Didael KTS. Si desidera ringraziare Gianna Martinengo, Paolo Turolla, Simone Uberti ed Elena Vanini per la loro attività di progettazione grafica e di implementazione del portale. Si vuole inoltre ringraziare Massimiliano Margaglio, Francesco Palmisano e Francesco Vitale, collaboratori dello staff del LIM, per le operazioni di reperimento e di codifica dei brani in formato IEEE 1599.

## 9. BIBLIOGRAFIA

- [1] G. Sibilla, *Musica e media digitali: Tecnologie, linguaggi e forme sociali dei suoni, dal walkman all'iPod*. Bompiani, 2010.
- [2] E. Pampalk, A. Rauber, and D. Merkl, “Content-based organization and visualization of music archives,” in *Proceedings of the tenth ACM international conference on Multimedia*, pp. 570–579, ACM, 2002.
- [3] E. Pampalk, S. Dixon, and G. Widmer, “Exploring music collections by browsing different views,” in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR’03)*, pp. 201–208, 2003.

- [4] G. Haus and L. Ludovico, “The digital opera house: an architecture for multimedia databases,” *Journal of Cultural Heritage*, vol. 7, no. 2, pp. 92–97, 2006.
- [5] L. Ludovico, “Key concepts of the IEEE 1599 standard,” in *Proceedings of the IEEE CS Conference The Use of Symbols To Represent Music And Multimedia Objects, IEEE CS, Lugano, Switzerland*, pp. 15–26, 2008.
- [6] A. Baratè and L. Ludovico, “New frontiers in music education through the IEEE 1599 standard,” in *Proceedings of the 4th International Conference on Computer Supported Education (CSEDU 2012)* (J. Cordeiro, M. Helfert, and M. J. Martins, eds.), vol. 1, pp. 145–151, Porto, Portugal: SciTePress - Science and Technology Publications, 2012.
- [7] A. Baratè, G. Haus, L. Ludovico, and D. Triglionne, “Multimodal navigation within multilayer-modeled gregorian chant information,” in *Proceedings of the 18th International Conference on Virtual Systems and Multimedia (VSMM2012)*, 2012.
- [8] A. Baratè, G. Haus, L. Ludovico, and D. Mauro, “IEEE 1599 for live musical and theatrical performances,” *Journal of Multimedia*, vol. 7, no. 2, pp. 170–178, 2012.
- [9] C. Mullin, “International Music Score Library Project/Petrucci Music Library (review),” *Notes*, vol. 67, no. 2, pp. 376–381, 2010.
- [10] D. Baggi, A. Barate, G. Haus, and L. Ludovico, “Nina-navigating and interacting with notation and audio,” in *Semantic Media Adaptation and Personalization, Second International Workshop on*, pp. 134–139, IEEE, 2007.
- [11] S. Pfeiffer, “Patents and their effect on standards: Open video codecs for HTML5,” *International Free and Open Source Software Law Review*, vol. 1, no. 2, pp. 131–138, 2010.
- [12] S. Baldan, L. Ludovico, and D. Mauro, “Managing multiple media streams in HTML5: the IEEE 1599-2008 case study,” in *SIGMAP 2011 - Proceedings of the International Conference on Signal Processing and Multimedia Applications*, SciTePress, 2011.



# UN AMBIENTE INFORMATICO PER LA GESTIONE DEI PROCESSI RELATIVI ALLA CONSERVAZIONE ATTIVA DEI DOCUMENTI SONORI

Federica Bressan

Scuola Normale Superiore di Pisa

federica.bressan@sns.it

## ABSTRACT

Questo articolo presenta un ambiente informatico costituito da strumenti software originali sviluppati sulla base dell'esperienza maturata durante progetti di ricerca con partner quali la Fondazione Arena di Verona (2009-2011) e la Scuola Normale Superiore di Pisa (2011-2012), orientati alla conservazione dei beni culturali musicali, con una particolare attenzione alla definizione di algoritmi per il controllo della qualità. Parallelamente ai risultati teorici che costituiscono il contesto scientifico della ricerca, l'articolo ne porta in evidenza il carattere innovativo, che consiste nell'utilizzo di un approccio sistematico alla progettazione degli strumenti informatici per controllare in maniera automatica processi concorrenti, con riferimento al protocollo per la conservazione attiva dei documenti sonori. Tali strumenti sono attualmente utilizzati dall'archivio storico della Fondazione Arena di Verona e dall'archivio sonoro del Laboratorio di Linguistica della Scuola Normale Superiore di Pisa, ma è auspicabile che in futuro essi vengano adottati da altri archivi per incoraggiare la diffusione di corrette pratiche conservative per mezzo dell'utilizzo di strumenti adeguati e liberamente condivisi.

## 1. INTRODUZIONE

In un rapporto di lavoro ufficiale, l'UNESCO [1] afferma che oltre la metà del patrimonio culturale è a forte rischio di scomparsa nonostante l'attenzione per il tema della conservazione dei beni culturali, soprattutto da parte della Comunità Europea che ha dimostrato grande sensibilità nel finanziare numerosi progetti di ricerca in questo campo.

I fattori che ostacolano la salvaguardia dei beni sonori e visivi sono molteplici: principalmente l'ingente investimento di risorse materiali e umane richieste dalle campagne di digitalizzazione, oltre a gruppi di lavoro con competenze multidisciplinari, difficili e costosi da formare. Come conseguenza, oggi molti archivi sono di fatto sprovvisti degli strumenti metodologici e tecnologici per tutelare adeguatamente il proprio patrimonio.

Da tempo l'informatica si è imposta pressoché in ogni settore della società industrializzata, e gli operatori dei be-

Copyright: ©2012 Federica Bressan 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.

ni culturali si avvalgono di gestioni informatizzate dei cataloghi e sempre più spesso anche delle modalità di accesso dell'utenza.

Van Huis [2] critica la generale inerzia delle istituzioni archivistiche di fronte alle nuove tecnologie, sintomo che ne venga ignorato il potenziale nel raggiungere gli utenti ormai avvezzi all'immediatezza, a fronte dell'incerta autorevolezza, di fenomeni come Google. Dall'informatica possono arrivare soluzioni sia ai problemi di carattere gestionale sia di carattere filologico-documentale, tuttavia l'informatica non è sufficiente: è cruciale che gli attori dei settori disciplinari coinvolti si confrontino secondo un approccio fondato sulla permeabilità delle competenze, come teorizzato da Agosti [3].

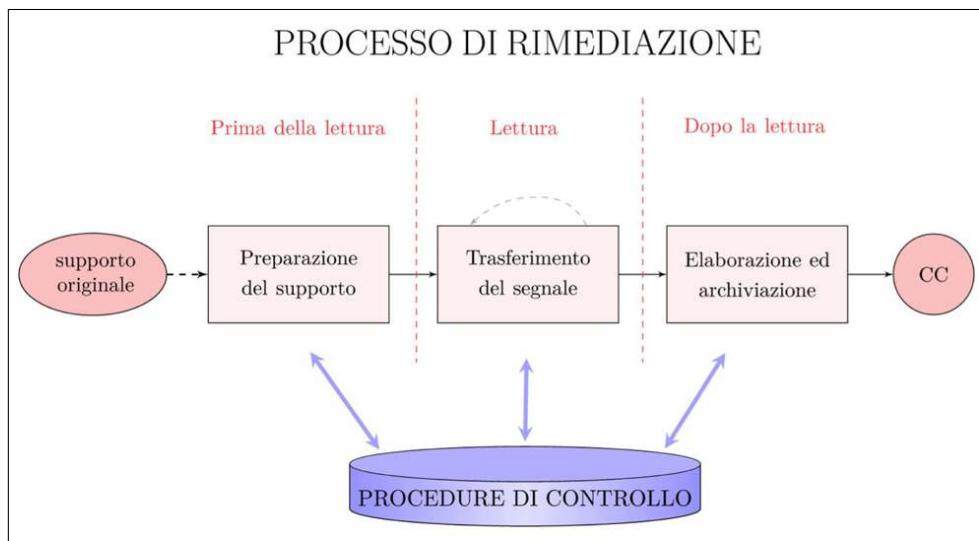
L'ambiente informatico descritto in questo articolo è il frutto della conoscenza maturata durante numerosi progetti di ricerca: quelli che hanno avuto maggior rilievo nel lavoro di sviluppo degli strumenti software sono il progetto di tutela dell'archivio storico della Fondazione Arena di Verona<sup>1</sup> e il progetto di tutela dell'archivio sonoro del Laboratorio di Linguistica "G. Nencioni" della Scuola Normale Superiore di Pisa<sup>2</sup>. Questi archivi, uno monotematico e prestigioso (opera lirica) e l'altro composto da documenti eterogenei prodotti in condizioni non ottimali e spesso conservati in condizioni inadeguate (inchieste di carattere linguistico ed etno-musicologico), costituiscono due estremi rappresentativi tra i quali si collocano le varie tipologie possibili di archivio sonoro.

Approcciando queste realtà dal punto di vista della ricerca metodologica e della riflessione scientifica, ho individuato alcune criticità nel sistema di gestione dei processi coinvolti nell'attività di conservazione del patrimonio sonoro, che ho deciso di risolvere per mezzo di strumenti informatici originali, progettati e sviluppati a valle di uno studio approfondito delle esigenze e delle risorse degli archivi. Tutti gli strumenti descritti sono condivisi liberamente<sup>3</sup>. Dal punto di vista della ricerca scientifica, tali strumenti si collocano nell'area delle *interfacce cultura-*

<sup>1</sup> Progetto congiunto REVIVAL 2009-2011 (REStauro dell'archivio Vlcentini di Verona e sua Accessibilità come audio e-Library), con l'Università di Verona e la Fondazione Arena di Verona, e il supporto scientifico di Eye-tech di Udine.

<sup>2</sup> Progetto PAR Fas Gra.fo 2011-2013 (Grammo-foni. Le soffitte della voce) finanziato dalla Regione Toscana nell'ambito della linea d'azione 1.1.a.3 relativa al periodo 2007-2013, con la Scuola Normale Superiore di Pisa e l'Università di Siena.

<sup>3</sup> Software licenziati con GNU Lesser General Public License, versione 3. I pacchetti di installazione e il codice sorgente sono disponibili nella sezione 'software' della pagina <http://www.dei.unipd.it/~bressan/>.



**Figure 1.** Schema del processo di conservazione attiva dei documenti sonori, in cui si distinguono le tre fasi principali della lavorazione e l’insieme delle procedure di controllo applicate durante il flusso di lavoro. Ciascuna fase di suddivide in procedure e sotto-procedure.

li. L’elemento innovativo consiste nell’utilizzo di un approccio sistematico nella progettazione degli strumenti per il controllo della qualità da applicare a procedure semi-automatizzate per la creazione e per la descrizione degli archivi digitali a scopo conservativo, per la distribuzione delle informazioni e per il mantenimento della loro integrità nel tempo. In altre parole, per il controllo automatico di processi concorrenti.

La sezione 2 offre alcuni riferimenti essenziali nell’ambito della conservazione dei documenti sonori, con lo scopo di rendere comprensibile il contesto, con i suoi principi, in cui è stato sviluppato l’ambiente informatico; mentre la sezione 3 descrive gli strumenti originali sviluppati dall’autore, dal punto di vista della funzionalità e della realizzazione.

## 2. CAPIRE E CONSERVARE GLI ARCHIVI SONORI

Negli ultimi trent’anni si è andata diffondendo la sensibilità verso il tema della conservazione dei documenti sonori, riconosciuti nel loro valore di bene culturale. Parallelamente, si sono arricchiti il dibattito etico e gli strumenti per attuare le pratiche conservative. Più di recente, alle registrazioni è stato anche riconosciuto un importante ruolo svolto nella ricerca scientifica di area linguistica, storica, sociologica, musicologica e in molte altre, come fonte documentale di dignità pari alle fonti bibliografiche, svelando la centralità dei problemi di carattere filologico quali l’autenticità e l’autorevolezza delle fonti [4].

Il processo di degrado fisico che caratterizza tutte le tipologie di supporti audio può essere rallentato, per mezzo di opportune azioni di conservazione, ma non arrestato. Di conseguenza, la sopravvivenza dell’informazione contenuta nel documento può avvenire soltanto rinunciando alla

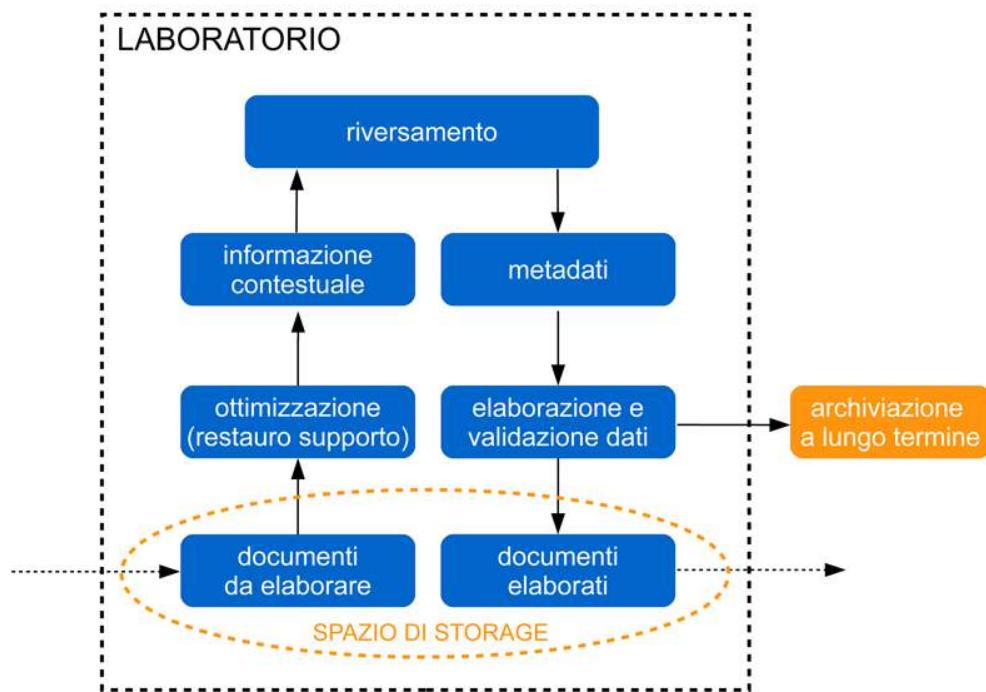
sua materialità, attraverso un continuo trasferimento dell’informazione verso nuovi supporti. Purtroppo il sistema del processo di ri-medializzazione (trasferimento dell’informazione acustica da un *medium* ad un altro *medium*) è soggetto a errori (elettronici, progettuali, operativi), oltre a essere dotato di un comportamento che facilmente asseconda i mutamenti del gusto estetico del proprio tempo. Di conseguenza non è ipotizzabile una neutralità nel trasferimento dell’informazione, ed è un punto questo che rende particolarmente delicato il problema dell’autenticità per le copie dei documenti sonori.

L’insieme delle azioni volte a conservare le memorie sonore (Figura 1) ha come scopo la creazione di *copie conservative* [5] dei documenti sonori, ovvero un insieme organizzato di dati che descriva in modo completo: l’informazione primaria costituita dal contenuto sonoro, le informazioni che si possono evincere dal segnale audio, e le informazioni contestuali (note scritte a mano, foto, video, documenti collegati, ecc.), minimizzando la perdita di informazione.

Metodologie di conservazione poco rigorose compromettono l’affidabilità delle registrazioni come fonte documentale. Diventa quindi di fondamentale importanza che la critica presti attenzione alla consistenza dei testimoni generati nel passaggio dal dominio analogico a quello numerico e, dal punto di vista informatico, che vengano progettati strumenti per il controllo della qualità del processo di ri-medializzazione dei documenti sonori.

## 3. STRUMENTI SOFTWARE

Quasi la totalità delle operazioni che caratterizzano le routine archivistiche sono fortemente ripetitive, e non fanno eccezione le operazioni relative al processo di conserva-



**Figure 2.** Schema logico delle attività che si svolgono all'interno del laboratorio di restauro.

zione attiva dei documenti sonori<sup>4</sup>. Pertanto il monte ore impiegato nell'elaborazione e nella gestione di documenti elettronici è molto elevata, e le note patologie dell'attenzione legate alle attività ripetitive e/o di basso livello possono indurre l'operatore a introdurre errori dall'effetto a cascata sul flusso di lavoro, causando il malfunzionamento degli algoritmi di controllo sulla coerenza interna dell'archivio e/o degli algoritmi di reperimento dell'informazione. Anche per il mantenimento dell'archivio nel lungo termine sono necessari strumenti adeguati, poiché da un lato la verifica periodica della totalità dei documenti è insostenibile e dall'altro la verifica a campione non è soddisfacente.

Gli strumenti software originali che ho sviluppato contribuiscono fortemente a risolvere questi problemi. Inoltre: (a) riducono la durata delle sessioni di riversamento, (b) riducono significativamente la durata della fase di produzione di copie d'accesso, e (c) introducono una serie di controlli automatizzati che garantiscono in modo pienamente affidabile l'integrità dei dati dell'archivio. In particolare, il software *Audiografo PreservationPanel* (Fig. 4) gestisce in maniera parallela le operazioni sui dati (archivio digitale) e sui metadati (popolamento della base di dati), garantendo che essi siano sempre allineati.

*Audiografo PreservationPanel* è stato utilizzato per la lavorazione dell'archivio sonoro del Laboratorio di Linguistica “G. Nencioni” della Scuola Normale Superiore di Pisa. L'analisi, l'organizzazione, il trasferimento e l'archiviazione di oltre 700 GB di dati, pari a 217 copie conservative associate a 10 archivi di provenienza diversi (oltre

<sup>4</sup> Nel campo delle memorie sonore, la conservazione si articola in passiva (difesa del supporto dagli agenti ambientali, senza alterarne la struttura) e attiva (riposizionamento dei dati su nuovi media). In questo articolo viene trattata la conservazione attiva.

430 ore di audio) hanno richiesto solo tre giorni di lavoro — a fronte di sei mesi necessari per la digitalizzazione dei documenti — grazie alla strategia di gestione per lotti descritta nella sezione 3.3.

Gli strumenti da me sviluppati si dividono, in base alla finalità d'uso, in:

- Strumenti di lavoro

1. Allineamento degli archivi (lab, server, backup)
2. Creazione e condivisione di copie d'accesso
3. Base di dati
4. Programmi per il popolamento della base di dati

- Strumenti di controllo

1. Monitoraggio dei processi
2. Validazione dei dati (medio/lungo termine)
3. Procedure di backup (base di dati, sito web, ...)
4. Monitoraggio dell'incremento dei dati

In base alla funzione svolta all'interno del *workflow* di progetto, questi strumenti si possono anche raggruppare in:

- strumenti per il processo di conservazione attiva dei documenti sonori
- strumenti per la descrizione dei contenuti
- strumenti per il monitoraggio e per il mantenimento dei dati
- strumenti per la condivisione dei dati

### 3.1 Descrizione del flusso di lavoro

In questo paragrafo viene descritto il flusso di lavoro ad alto livello, mentre nei successivi vengono descritti gli strumenti informatici, suddivisi per tipologia, che realizzano le

operazioni del flusso stesso. Le operazioni che si svolgono all'interno del laboratorio di conservazione e restauro sono schematizzate in Figura 2.

Dopo le fasi preliminari di presa in carico dell'archivio e di definizione dei criteri di priorità di intervento, ciascun documento viene sottoposto ad un'analisi finalizzata a stabilirne lo stato di conservazione, in base al quale si può pianificare un opportuno restauro del supporto fisico. Dopo aver documentato fotograficamente il supporto, esso è pronto per la fase di estrazione del segnale. A questo punto viene predisposta la struttura della copia conservativa, che verrà completata con i dati prodotti durante la fase di ri-medializzazione e con il risultato della loro analisi.

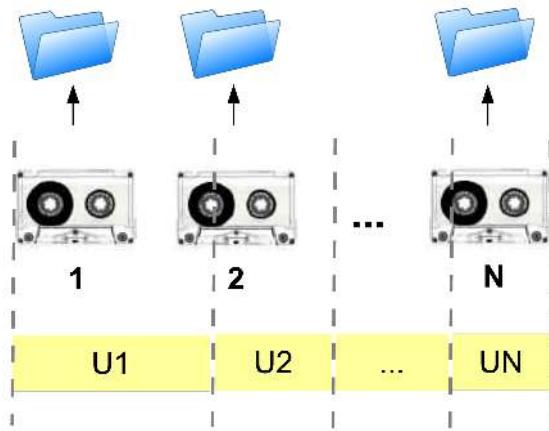
Le copie conservative complete vengono trasferite quotidianamente dal laboratorio ad un calcolatore server, che ospita il sistema di archiviazione a lungo termine e dove ha luogo il resto delle routine di elaborazione: per ciascuna copia conservativa vengono create una copia di sicurezza e una copia d'accesso (conversione dei documenti audio e selezione del materiale di corredo), quest'ultima condivisa via web con accesso regolamentato. L'équipe incaricata della fase di catalogazione (esperti di linguistica, di musicologia, ...) riceve una mail automatica con la notifica dei nuovi documenti disponibili. La condivisione via web è stata scelta perché i collaboratori cui spetta di ottenere le copie d'accesso si trovano dislocati sul territorio nazionale, infatti per il popolamento della base di dati utilizzano uno strumento software in grado di stabilire connessioni remote (*Audiografo CataloguingPanel*)

Questo protocollo si ripete quotidianamente, minimizzando il tempo necessario a coprire il percorso che le informazioni compiono idealmente dal supporto fisico alla consultazione degli utenti finali. Infine, lo stato di avanzamento del lavoro di digitalizzazione viene monitorato calcolando la durata complessiva dei riversamenti presenti in archivio al termine di ogni sessione di lavoro.

Una volta che le copie conservative sono state complete e trasferite all'archivio per il lungo termine, il processo di conservazione attiva è terminato. Tuttavia per la sopravvivenza dell'archivio è necessario programmare processi di *checking*, *copying* e *migration*, poiché la tecnologia digitale non è immune dall'azione degradante del tempo e dall'obsolescenza dei formati e dei supporti. Il calcolatore server è programmato per avviare periodicamente e in maniera automatizzata processi di monitoraggio e di backup dell'archivio e della base di dati, generando reportistica in formato testuale che viene inviata a mezzo posta elettronica agli operatori incaricati della supervisione del sistema.

### 3.2 Base di dati

Una base di dati progettata in seno a progetti di ricerca orientati alla conservazione di un archivio sonoro, come quelli descritti nella sezione 1, deve necessariamente prevedere il mantenimento dei dati di carattere tecnico-con-



**Figure 3.** La relazione tra i documenti sonori (al centro) e le copie conservative (in alto) è sempre 1:1. La relazione tra questi e le unità sonore mediate per la fruizione (U1, U2, ...) può essere complessa, e questo accade particolarmente negli archivi composti da materiale prodotto sul campo per indagini di carattere linguistico ed etno-musicologico.

servativo per la documentazione a lungo termine, assieme ai dati di descrizione dei contenuti che permettono il reperimento delle risorse da parte degli utenti finali nel breve termine. A seconda delle caratteristiche e delle esigenze specifiche dell'archivio, individuate in maniera chiara durante una fase preliminare di studio, la struttura della base di dati potrà variare sensibilmente, fermi restando i dati di tipo tecnico-conservativo enumerati dalle più importanti istituzioni archivistiche in ambito internazionale (IASA [6] e IFLA [7]).

La base di dati del progetto Gra.fo (vedi Sez. 1) è stata progettata secondo il modello relazionale in MySQL. Il popolamento avviene per mezzo delle applicazioni *Audiografo PreservationPanel* e *Audiografo CataloguingPanel* (vedi Sez. 3.3). L'obiettivo più delicato della fase di progettazione è stata senz'altro quella in cui si è definita la relazione tra le entità proprie dell'approccio al documento (approccio conservativo) e quelle dell'approccio al contenuto, poiché uno guarda al documento come testimone di una storia che va tramandata nella sua integrità senza mediazioni, l'altro si preoccupa precisamente di operare una mediazione tra il testimone e il documento nella forma in cui verrà fruito dagli utenti finali in quanto oggetto dotato di senso in un contesto specifico. In particolare, la relazione tra documento sonoro e l'unità mediata per la fruizione è rappresentata in Figura 3. Nella base di dati, il legame tra la copia conservativa (e quindi il documento sonoro) e l'unità mediata è mantenuto per mezzo della relazione tra la traccia audio associata all'unità e tutte le tracce audio provenienti dalle copie conservative che sono state utilizzate per comporla. Con questa struttura, la presentazione dei dati può avvenire sia dal punto di vista dell'utente meramente interessato al contenuto di un'unità (ad esempio, la narrazione delle antiche tradizioni legate alla mietitura), sia dal punto di vista dell'utente interessato al processo che

ha portato alla generazione della fonte, ovvero alle caratteristiche fisiche del supporto originale e al trattamento cui è stato sottoposto. Ovviamente, ed è il nocciolo della base di dati, è possibile risalire alle informazioni di carattere conservativo a partire dal risultato delle interrogazioni sul contenuto e viceversa.

### 3.3 Audiografo

*Audiografo PreservationPanel* e *Audiografo Cataloguing-Panel* sono applicazioni software originali sviluppate nell'ambito del progetto Gra.fo utilizzando il linguaggio di programmazione ad oggetti Java (circa 29K righe di codice). Il loro utilizzo è previsto in due fasi distinte del flusso di lavoro: *Audiografo PreservationPanel* offre funzionalità legate al processo di conservazione attiva dei documenti sonori, *Audiografo CataloguingPanel* alla fase di catalogazione dei contenuti.

Si tratta quindi di applicazioni con **finalità** distinte e indipendenti, viceversa il loro **funzionamento** è fondamentalmente analogo e solo in parte dipendente (*Cataloguing-Panel* contiene riferimenti a dati prodotti da *Preservation-Panel*, i quali costituiscono il legame tra la parte del progetto orientata alla conservazione e quella orientata ai contenuti – vedi Sez. 3.2).

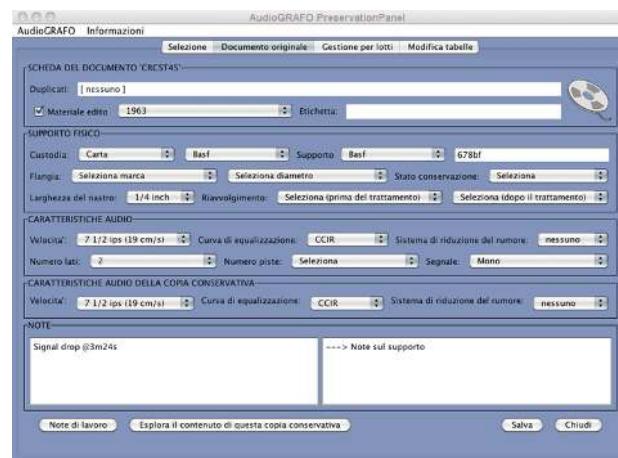
A livello utente, le funzionalità principali di *PreservationPanel* sono:

- la creazione e l'archiviazione delle copie conservative (generazione ed elaborazione di file)
- il popolamento della base di dati studiata e progettata per Gra.fo con le informazioni relative al supporto originale, alla copia conservativa, alle copie d'accesso e di sicurezza.

L'utilità di *PreservationPanel* si esprime principalmente nel controllo della qualità del processo di conservazione, in quanto è in grado di gestire l'archivio delle copie conservative e i relativi metadati memorizzati nella base di dati, mantenendo l'allineamento ed evitando incongruenze costose per l'integrità dell'archivio. L'efficienza del programma si esprime principalmente nella strategia di gestione dei documenti per lotti, permettendo l'elaborazione di grandi quantità di dati in poche operazioni ridondante dal punto di vista della sicurezza. Per l'estrazione dei metadati audio mi sono avvalsa di JHove, una collezione di algoritmi sviluppati dall'Università di Harvard che ho integrato in *PreservationPanel*<sup>5</sup>.

C'è un aspetto originale che rende *PreservationPanel* più di un'interfaccia per *data entry* o uno strumento per l'elaborazione dei documenti in lotti, e riguarda il pannello di descrizione dei documenti sonori (Figura 4). Il pannello è stato studiato per semplificare il lavoro dell'operatore umano e per minimizzare l'introduzione di errori. Il primo

<sup>5</sup> *JSTOR/Harvard Object Validation Environment* è uno strumento software modulare sviluppato per l'analisi e la validazione di oggetti digitali nell'ambito della conservazione dei beni culturali (<http://hul.harvard.edu/jhove/>).



**Figure 4.** Pannello di descrizione del documento sonoro in *Audiografo PreservationPanel* per la tipologia 'audiobobina'.

obiettivo è raggiunto (a) filtrando i componenti dell'interfaccia in base al criterio dell'applicabilità degli attributi associati al documento sonoro (modellati nella base di dati); e (b) imponendo vincoli di dipendenza tra i valori dei componenti dell'interfaccia. Il secondo obiettivo è raggiunto per mezzo del punto (a), e (c) filtrando i valori validi per ciascun attributo in base alla tipologia di supporto selezionata. Al di là della complessità che una tale gestione implica a livello di programmazione software, il momento che ha richiesto maggior sforzo è stato quello della riflessione circa le caratteristiche rilevanti dei supporti sonori e dei rispettivi valori ammessi per ciascuna di esse. A tale fine ho studiato la documentazione prodotta da altri progetti di ricerca in ambito di conservazione di documenti sonori<sup>6</sup>, ho studiato le schede descrittive di archivi sonori accreditati a livello internazionale<sup>7</sup>, e mi sono affidata all'esperienza maturata durante l'intensa attività laboratoriale che è stata la cifra della mia partecipazione ai progetti di ricerca nel corso del dottorato.

Analogamente, la funzionalità principale di *Cataloguing-Panel* è la facilitazione nell'operazione di popolamento della base di dati per mezzo di una presentazione dei dati appositamente studiata e per mezzo di vincoli di dipendenza trasparenti per l'utente. *CataloguingPanel* non interagisce con il *file system* locale. Tutti i dati prodotti ed elaborati durante una sessione di lavoro di *CataloguingPanel* vengono reperiti o inviati via internet da e all'elaboratore server ospitato presso la Scuola Normale Superiore di Pisa. Il gruppo di lavoro che ha utilizzato e sta utilizzando regolarmente *CataloguingPanel* è distribuito tra Arezzo, Firenze e Torino.

<sup>6</sup> PLANET - <http://www.planet-ict.eu/>; PRESTOSPACE project - <http://www.prestospace.org/>; PREMIS project - <http://www.loc.gov/standards/premis/>. Tutte le pagine sopra citate sono state visitate l'ultima volta il 26 luglio 2012.

<sup>7</sup> Speech & Language Data Repository (SLDR) - <http://crdo.up.univ-aix.fr/> (La pagina è stata visitata l'ultima volta il 26 luglio 2012).

### 3.4 Script e altri programmi

Nella sezione 3 abbiamo detto che la maggior parte delle routine archivistiche sono fortemente ripetitive. Gran parte di queste consiste nell’elaborazione di grandi moli di dati. Spesso la natura dell’elaborazione è definita a priori e quindi non necessita dell’intervento dell’operatore umano per valutazioni e per decisioni. Pertanto gli strumenti informatici, in questo caso in particolare gli *shell script*, sono ideali per (a) automatizzare le operazioni ripetitive sui dati, e contemporaneamente per (b) controllare in maniera sistematica l’esecuzione e l’output delle operazioni. A fianco ad *Audiografo PreservationPanel* e *Audiografo CataloguingPanel*, che sono dotati di GUI, è stata sviluppata una serie di *shell script* per eseguire specifiche operazioni sia sulle postazioni di lavoro in laboratorio sia sull’elaboratore server. Queste operazioni includono:

- backup della base di dati
- backup dell’archivio di copie conservative
- backup del sito web
- monitoraggio dei valori di alcune tabelle della base di dati (controllo sull’incremento e sulla coerenza interna dei valori da parte dell’équipe di catalogazione)
- invio di messaggi e-mail per il controllo del corretto funzionamento del server
- creazione delle copie d’accesso e loro pubblicazione sul web
- calcolo della durata complessiva dei riversamenti per ciascun archivio, fondo e serie

Dal momento che il flusso di lavoro viene generalmente definito nel protocollo redatto a monte del progetto, è possibile programmare l’esecuzione delle procedure automatizzate a cadenza temporale predefinita. Ciò è possibile grazie al carattere ciclico del flusso di lavoro, che si ripete grosso modo a rotazione giornaliera, e che inoltre ammette l’esecuzione concorrente dei processi (approccio sistematico). I processi avviati automaticamente sono specificati nel *crontab* (lista di *cron-job*) dell’elaboratore server (Unix) e si suddividono in gruppi a rotazione giornaliera, settimanale, mensile e annuale (archiviazione dei file di backup in formato compresso).

I processi che necessitano di ricevere o di passare dati prodotti durante l’elaborazione generalmente lo fanno per mezzo di file di testo, alcuni dei quali sono temporanei e alcuni dei quali costituiscono il testo, opportunamente modificato, dei *log* inviati a mezzo mail. Nella programmazione degli script è stata data particolare attenzione all’efficienza del codice, alle esigenze peculiari dell’archivio in lavorazione (la sua struttura, ecc.), alla documentazione del codice e alla ridondanza dei controlli.

## 4. CONCLUSIONI

Dal punto di vista della ricerca scientifica, gli strumenti software originali che costituiscono l’ambiente informatico presentato in questo articolo si collocano nell’area delle interfacce culturali. Essi si distinguono per lo sviluppo condotto con un approccio trans-disciplinare, raggiunto grazie al rapporto di collaborazione quotidiana di ricercatori informatici con esperti di settori scientifici diversi

(musicologi, linguisti, archivisti, ecc.), che si è tradotto in occasioni di confronto e di approfondimento nella raccolta dei requisiti, nella comprensione e nella convergenza dei vocabolari e delle metodologie, ecc. Questo tipo di approccio si è riflesso:

- sulla progettazione della base di dati (conciliazione di diversi approcci alla conservazione, modellazione di informazioni)
- sulla formalizzazione del flusso di lavoro (accordo tra teoria della conservazione e pratica di laboratorio): sostenibilità e tempi di lavorazione, supporto di sessioni di riversamento e di catalogazione su postazioni parallele.

Oltre ad aver migliorato la qualità del lavoro di laboratorio, i risultati raggiunti dimostrano che l’introduzione nel processo di conservazione attiva dei documenti sonori di strumenti informatici sviluppati *ad hoc* permette di dare una risposta precisa alle questioni metodologiche sull’affidabilità delle registrazioni come fonti documentali, chiarendo il concetto di “fedeltà” all’originale inquadrandolo nei precisi limiti della tecnologia per la riproduzione sonora.

L’architettura dell’ambiente informatico riflette i fondamenti della metodologia applicata al lavoro di conservazione attiva dei documenti sonori. Assunta la condivisione di tali fondamenti, l’ambiente può essere adattato a basso costo per l’utilizzo presso altri archivi, un obiettivo desiderabile per la diffusione di corrette pratiche di conservazione incentivate dall’uso di strumenti adeguati e condivisi liberamente.

## 5. REFERENCES

- [1] R. Edmonson, *Memory of the World: General Guidelines to Safeguard Documentary Heritage*. UNESCO, February 2002.
- [2] E. Van Huis, “What makes a good archive?”, *IASA Journal*, pp. 25–28, December 2009.
- [3] M. Agosti, *Information Access through Search Engines and Digital Libraries*, ch. Information Access using the Guide of User Requirements, pp. 1–12. Springer-Verlag Berlin Heidelberg, 2008.
- [4] D. Schüller, “Audiovisual research collections and their preservation,” tech. rep., European Commission on Preservation and Access, March 2008.
- [5] S. Canazza, G. De Poli, and A. Vidolin, “La conservazione dei documenti audio: un’innovazione in prospettiva storica,” *Archivi*, vol. VI, pp. 7–56, luglio-dicembre 2011.
- [6] IASA, *The IASA Cataloguing Rules*. IASA Editorial Group, 1999.
- [7] IFLA - Audiovisual and Multimedia Section, “Guidelines for audiovisual and multimedia materials in libraries and other institutions,” tech. rep., March 2004.

# ALA - Un'Iper-Lavagna per il Training Ritmico

**Tiziano Bole**

Conservatorio "G. Tartini" di Trieste  
Scuola di Musica e Nuove Tecnologie  
tizi.bole@gmail.com

**Cristina Fedrigo**

Conservatorio "G. Tartini" di Trieste  
Pedagogia Musicale  
cristina.fedrigo@tin.it

## ABSTRACT

Per un'efficace didattica della musica appare di centrale importanza l'impiego di risorse flessibili e adattabili alla realtà che l'insegnante affronta in classe. I sistemi interattivi multimodali presentano per loro stessa natura caratteristiche funzionali in tal senso. A questo proposito si presenta in questa sede un'interfaccia didattica denominata *Audi/Lege/Age* (ALA), finalizzata al *training* ritmico di gruppo. ALA, sviluppata da un precedente gioco di *training* ritmico, ne riprende le forme di rappresentazione non tradizionale delle figure ritmiche e lo stile di approccio, estendendo le possibilità originali di quel gioco e piegandole a un utilizzo di gruppo. ALA diventa per l'insegnante un'iper-lavagna, per eseguire essa stessa gli esercizi e offrire una serie di varianti di contesto utili a un *training* ritmico il più efficace possibile. Viene inoltre discussa una prima sperimentazione di ALA effettuata in una classe di conservatorio, durante la quale sono emersi utili spunti per delineare strategie di *training* in contesti che cambiano.

## 1. INTRODUZIONE

In questo articolo viene presentato un sistema informatico di ausilio alla didattica della musica, che nella forma attuale è stato denominato *Audi/Lege/Age* (ALA). ALA ha conosciuto una certa evoluzione, essendo stato corretto e modificato, anche sostanzialmente, per rispondere alle problematiche emerse in sede operativa e per ottimizzarne la funzionalità, in ogni caso con una costante sperimentazione in campo didattico.

### 1.1 Da un videogioco per il training ritmico...

Nella prima versione (presentata in [3]), il sistema era nato come strumento di *training* ritmico a carattere ludico, con il quale l'utente potesse misurare la correttezza e la precisione della propria esecuzione in una serie di test, attraverso un percorso a livelli. Esso si collocava, dunque, nel filone dell'*Intelligent Tutoring System* ([5]), assieme a sistemi quali, a titolo esemplificativo, *Piano Tutor* ([12] e [13]), *IMUTUS* ([14]) o *i-Maestro*<sup>1</sup> ([10], [16], [18], [19]).

<sup>1</sup> Sito web del progetto i-Maestro: [www.i-maestro.org](http://www.i-maestro.org).

Copyright: ©2012 Tiziano Bole and Cristina Fedrigo. 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 [20]), e per l'attenzione in particolare agli aspetti ritmici, iPalmas ([15]). In questa prima versione, in cui l'azione dell'insegnante era necessaria ma marginale, era prevista la presenza di un microfono con il quale il sistema catturava l'esecuzione di un *pattern* ritmico dato, da far eseguire all'utente battendo le mani o con l'ausilio di strumenti percussivi. La consegna del *pattern* richiesto avveniva tramite una doppia modalità, sia visiva (sullo schermo del computer), sia sonora (con l'esecuzione del *pattern* medesimo da parte del sistema). A ogni insuccesso dell'utente, lo stesso *pattern* era riproposto affinché si potesse migliorare la prestazione fino al livello richiesto, mentre ad ogni successo, a seconda dei casi, il *pattern* cambiava o ne veniva modificata la velocità, con un margine tollerato d'errore via via minore.

Il principale limite del taglio che era stato dato al sistema in questa prima versione (nato per una pratica di esercizio individuale) era l'impossibilità di fruizione contemporanea da parte di più utenti su un'unica interfaccia, per l'ovvia ragione che un'esecuzione collettiva avrebbe impedito la corretta valutazione della *performance* ritmica, per la molteplicità di ogni singolo evento sonoro. Essendo, peraltro, quella di classe la situazione più tipica con cui l'educatore o l'insegnante si confronta, il limite di questo primo approccio è parso significativo, al punto da far evolvere il sistema in funzione di una didattica di gruppo.

### 1.2 ...ad ALA

Nella nuova versione, ad ALA è stato tolto il ruolo valutativo ed è stata ampliata la gamma delle possibili configurazioni ritmiche, anche attraverso la sovrapposizione di *pattern* e di metri diversi, mantenendo le medesime modalità di consegna sul doppio binario sonoro/visivo. In questo modo, diventa centrale la figura dell'insegnante nell'utilizzo dell'interfaccia, acquisendo egli sia il ruolo valutativo sottratto al sistema, sia quello creativo, componendo egli stesso i vari *pattern* da sottoporre agli studenti, in modo da adeguare il livello di complessità ai diversi contesti, gruppi e situazioni in cui opera. Anche il fine dell'interfaccia dunque cambia, spostandosi da quello originale di puro *training* ritmico, con il quale l'utente poteva misurarsi a un certo punto anche autonomamente (riecheggiando modalità di istruzione programmata), a quello di ausilio all'esercizio in classe, in un certo senso un'iper-lavagna, la cui caratteristica principale diventa ora essere risorsa didatticamente flessibile.

## 2. ALCUNE CONSIDERAZIONI DI ORDINE METODOLOGICO

Le possibilità offerte dalle nuove tecnologie per l'educazione e l'insegnamento musicale, pur ritrovando un'anima antica nella restituzione dell'esperienza "immediata" che corrella suono, gestualità, rappresentazioni grafiche, costituiscono un vastissimo territorio, da esplorare. Un modo efficace per indagare le funzionalità di queste risorse, peraltro in linea con la ricerca didattica contestualizzata in sede di attività, consiste nel calarle in un reale e determinato contesto operativo e "interrogarne" le potenzialità a concreto confronto con un problema didattico o educativo [17]. "Situare" questo genere di esperienze non consente successive generalizzazioni circa gli esiti dell'attività, ma fornisce esempi necessari ad arricchire i punti di vista sulle possibilità d'azione dell'insegnante, ad ampliare comunque le risorse di cui disporre affrontando l'attività didattica e le sue criticità<sup>2</sup>. Inoltre nascono così riscontri sulle possibili vie di sviluppo in merito all'utilizzo di questo specifico tipo di strumenti per l'esperienza musicale. Durante lo scorso anno accademico 2011/2012, si è presentata una situazione di questo genere, con una classe di I anno del Triennio di *Musica e Nuove Tecnologie* al Conservatorio "G. Tartini" di Trieste, costituita da una ventina di studenti che, per livello di alfabetizzazione musicale e formazione culturale generale, si presentava assai disomogenea. Si è trattato di intraprendere in via sperimentale un percorso - di cui da tempo in seno al Dipartimento stesso si avvertiva l'esigenza - per offrire a questo tipo di studenti una formazione coerente con profilo professionale e corso di studi specifici, centrata sui fondamenti del linguaggio musicale ma diversamente "tagliata" rispetto ai più usuali corsi di teoria e percezione musicali, e simili. Il *focus* dell'attenzione avrebbe dovuto essere l'oggetto/evento sonoro, nelle sue rappresentazioni come nei suoi caratteri fenomenici, un oggetto da saper quindi ascoltare (*Audi*), ponendolo in immediata correlazione strutturale con altri oggetti consimili nel tempo, leggere (*Lege*), ossia riconoscere e collocare in partitura, sempre in relazione con gli altri eventi rappresentati, e da eseguire (*Age*).

Questa esperienza si è rivelata un contesto di continua sperimentazione e ricerca di strategie, materiali e risorse [2], dove anche ALA ha dimostrato di poter offrire un contributo efficace. Poiché non tutti questi studenti erano musicalmente alfabetizzati in modo strutturato e facilmente verificabile, è stato necessario ricollocare contenuti di una formazione musicale basilare, ripensarli in funzione delle competenze richieste nella specifica formazione in questione. Tra gli obiettivi del percorso infatti sono previsti: la competenza mirata di approccio al testo musicale e di penetrazione dell'evento sonoro nell'ascolto, e la capacità di correlare elementi in senso sia strutturale che percettivo, potendo avvalersi del sapere musicale proveniente dalla tradizione, essendo in grado, tuttavia, di fruirne con la necessaria apertura per repertori non così tradizionali nella forma e nei contenuti.

Peraltro, gli studenti di *computer music* possono mancare di una pratica musicale strumentale comunemente inte-

<sup>2</sup> Si veda in proposito la funzionalità pedagogica del problema in [1].

sa, ma hanno comunque la necessità di sviluppare competenze utili a gestire questo tipo di *performance* nel tempo. ALA è stato sperimentato proprio qui: all'incrocio tra lettura ed esecuzione di gruppo, sempre in forma estemporanea, a livelli incrementali di complessità ritmica, gestiti in tempo reale dall'insegnante. Tale attività esecutiva è stata svolta ponendo in correlazione diversi eventi contemporanei, controllando forme di rappresentazione come si vedrà "semplici", ma sviluppando necessariamente una supervisione di ascolto assai complessa.

E' ragionevole supporre che lo studente possa trarre vantaggio da detta pratica in classe, in senso di maggiore consapevolezza e autonomia d'uso dei propri specifici strumenti di espressione, a diversi livelli di mediazione, per poi usarli anche nel monitorare le proprie prestazioni ([11]). D'altro canto il "metautilizzo" dei propri mezzi creativi ci pare ben inscritto nella filosofia di esercizio in contesti interattivi e multimodali. La flessibilità che ci aspettiamo questi contesti possano offrire costituisce una precondizione necessaria per operare anche ai livelli appena indicati. Ogni ambiente culturalmente connotato offre strumenti, sintassi, disciplina, piani di senso tipici ([6], [7]). Le competenze che realizziamo in e attraverso detti contesti, utilizzandone gli specifici codici, formano anche le nostre prospettive di attività, la nostra mentalità. Come si può restare "aperti" rispetto alle proprie familiari risorse e autonomi rispetto al loro utilizzo? Quali caratteristiche, quindi, queste risorse devono possedere per sostenere il perfezionarsi delle nostre abilità, rendendoci anche partecipi e consapevoli di ciò che stiamo imparando? Ad esempio, ci devono far ricordare un'attività complessa alla nostra piena gestibilità, consentire un'efficace auto-correzione delle prestazioni, e devono permetterci di selezionare, volta per volta, i contenuti di pratica e le variabili di modo, tempo e misura che consentono un lavoro sostenibile. ALA è un sistema che cerca di offrire alcune possibilità in tal senso, nel nostro caso con un intento puramente formativo poiché l'aspetto creativo non è stato finora preso in considerazione. Non va, tuttavia, dimenticato che non è opportuno pensare gli strumenti informatici in sostituzione di quelli della tradizione musicale, ma devono auspicabilmente affiancarvisi e integrarli, poiché, a livello didattico, la realtà pone sempre problemi inediti e multiformi: abbandonare la ricchezza e la varietà delle risorse praticabili risulta in generale una scelta poco avveduta, come pure la sostituzione delle risorse invece del loro adeguamento.

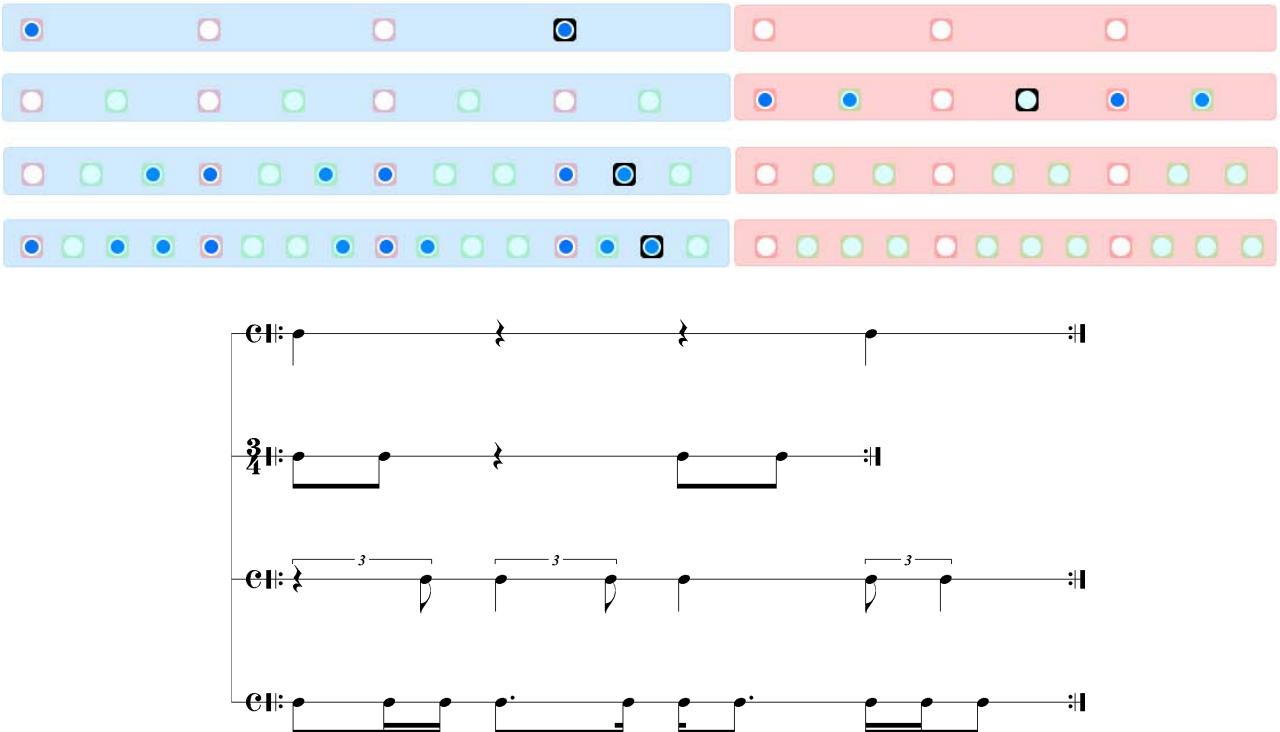
Una risorsa come ALA consente, dunque, tipi e livelli di azione musicale, nel nostro caso in forma estemporanea, efficacemente integrabili anche al *training* tradizionale del musicista "acustico" e non solo di quello "elettronico"<sup>3</sup>.

## 3. DESCRIZIONE E USO DELL'INTERFACCIA

L'interfaccia, sviluppata in MAX-MSP<sup>4</sup> propone una serie di righe di led organizzati con colori diversi e che rappresentano ognuno un'unità ritmica, seguendo il semplice principio della scrittura intermittente, come già nella pri-

<sup>3</sup> Approfondimenti a carattere bibliografico e metodologico, si veda il sito web <http://www.cristinafedrigo.it>.

<sup>4</sup> Sito web: [cycling74.com](http://cycling74.com).



**Figura 1.** Esempio di rappresentazione di un esercizio polimetrico all'interno di ALA (in alto) e relativa versione in notazione tradizionale (in basso). Si possono notare le quattro righe dalle suddivisioni via via più piccole, i led, alcuni dei quali accesi (centro blu), la scomposizione in 4+3 pulsazioni (blocco azzurro/blocco rosa) e la presenza dei cursori (quadrati neri attorno ai led). In questo caso viene considerato il solo blocco azzurro per tutte le righe meno la seconda, in cui si considera solo il blocco rosa da tre pulsazioni. Ciascun gruppo di studenti esegue in *loop* la riga attribuitagli.

ma versione illustrata in [3]. Tale scrittura, che utilizza unicamente segni del tipo acceso/spento, è stata adottata poiché capace di maggiore immediatezza rispetto alla scrittura musicale tradizionale: questa caratteristica consente di operare a livelli di complessità molto alti, senza sovraccaricare le funzioni cognitive dell'utente. La consegna dell'esercizio consiste nel far corrispondere al led acceso un elemento sonoro e al led spento un silenzio. Siccome i suoni richiesti sono di carattere percussivo (anche se non per forza ad altezza indeterminata) tale indicazione diventa sufficiente e si può escludere una notazione delle durate di tali suoni. In sede di sperimentazione, è emerso che la potenza di questa rappresentazione permette in pochi passaggi intermedi la sovrapposizione di diversi *pattern* ritmici, realizzando livelli di complessità maggiori di quanto comunemente si pratichi ad un pari livello formativo musicale. Ogni riga di led rappresenta, infatti, una scansione ritmica via via più fitta, partendo per la prima riga da un rapporto di 1:1 con la pulsazione (l'equivalente delle semiminime in un tempo di 4/4), passando per un rapporto di 1:2 nella seconda riga (similmente l'equivalente delle crome), di 1:3 nella terza riga e di 1:4 nella quarta (l'equivalente delle semicrome). E' evidente come sovrapporre *pattern* ritmici composti sulla base di queste diverse scansioni, porti a livelli di complessità ritmica elevati. Inoltre, l'insegnante può decidere per ogni riga di led se sfruttarla interamente (nel caso sperimentato ogni riga intera dura 7 pulsazioni) o solo in parte (nel caso sperimental-

tato era possibile scegliere di limitarsi alle prime 4 o alle ultime 3 pulsazioni). Nell'esecuzione in *loop* simultanea di più parti diverse per lunghezza, si sviluppano situazioni polimetriche, con conseguente ulteriore incremento della complessità ritmica. Tuttavia, detta complessità aumenta solo a livello esecutivo e coordinatorio rispetto al contesto sonoro, ma non per quanto concerne i meccanismi di lettura. Un esempio di notazione all'interno di ALA è illustrato in Fig. 1.

L'insegnante indica con le modalità di cui sopra i *pattern* ritmici da eseguire, accendendo o spegnendo i led delle righe desiderate, suddividendo eventualmente la classe in più gruppi, ognuno dei quali esegue una delle righe, indicando quanti led della propria riga ogni gruppo debba considerare.

### 3.1 Varianti utili

Il sistema consente varianti utili, che possono essere introdotte anche durante la *performance* degli studenti, in modo da modificare con il contesto di esercizio anche i punti di vista sull'esperienza musicale. Le varianti sono le seguenti:

- La possibilità di includere o escludere il metronomo, e di variarne la velocità. Non è necessaria la presenza del metronomo quando la *performance* ritmica, dopo un certo numero di ripetizioni, si è ben consolidata e la pulsazione è stata interiorizzata da-

gli studenti, oppure nel caso siano attivi i cursori (si veda il punto seguente).

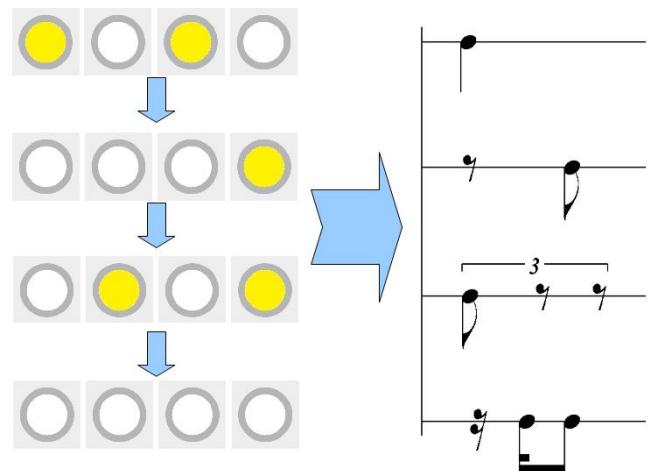
- La possibilità di introdurre i cursori, uno per riga, che indicano il led “corrente” (o i led se sono attive più righe) con l’effetto di non far “perdere il segno” e di fornire anche una versione visiva della pulsazione metronomica<sup>5</sup>.
- La possibilità di far eseguire all’interfaccia stessa i *pattern* richiesti attraverso campioni sonori. Il tipo di esperienza musicale può cambiare notevolmente al variare di tali campioni, soprattutto usandone alcuni ad altezza determinata (comunque di tipo percussivo) rispetto all’uso dei soli campioni ad altezza indeterminata. Introdurre o escludere questo ausilio sonoro, o cambiare i suoni impiegati, può modificare profondamente il senso musicale dell’esercizio ritmico, pur mantenendone inalterata la consegna.
- La possibilità di celare agli studenti la rappresentazione spaziale dei *pattern* richiesti, ossia non mostrare la consegna dei *pattern* con il sistema descritto a led accesi/spenti. Questa esclusione ha come effetto la massima concentrazione da parte degli studenti sulla sola organizzazione sonora temporale degli eventi da riprodurre. In tal caso, per la consegna dei *pattern* serve attivare l’esecuzione sonora dei *pattern* stessi da parte dell’interfaccia, o la loro visualizzazione a led singolo descritta al prossimo punto.
- La possibilità di visualizzare il *pattern* ritmico attraverso un unico led per gruppo di studenti, che si accende in occorrenza di un evento sonoro, in alternativa alla rappresentazione spaziale a più led. Un esempio di questa modalità è illustrato in Fig. 2.

Queste varianti sono state proposte in sede di sperimentazione in ogni combinazione, in particolare sono state individuate sequenze progressive di varianti, atte a fornire una direzione agli esercizi ritmici degli studenti, per consolidare gradualmente la *performance* e renderla sempre più padroneggiata. I più significativi di questi percorsi di *training* verranno illustrati nella prossima sezione.

#### 4. ASPETTI METODOLOGICI DELLA Sperimentazione

La sperimentazione si è svolta, come già accennato, con la classe di primo anno di Triennio Superiore di I livello in Musica e Nuove Tecnologie del Conservatorio “G. Tartini” di Trieste, durante alcune delle lezioni di “Ascolto, lettura e pratica di repertori musicali” tenute dalla prof.ssa Cristina Fedrigo. ALA è stata sperimentata in cinque lezioni, distribuite durante tutto l’arco dell’anno accademico 2011/2012. La classe, come già detto, era composta da una

<sup>5</sup> In questo caso, come anche per i campioni sonori citati successivamente, è importante che la latenza della scheda audio utilizzata non sia percepibile, in modo da non dar luogo a un’asincronia audio-visiva, che influirebbe negativamente sulla prestazione degli studenti.



**Figura 2.** Esempio di modalità di visualizzazione dei *pattern* a led singoli. A sinistra sono rappresentati i quattro led corrispondenti alle quattro diverse righe in istanti successivi di tempo. A destra la stessa figura ritmica è rappresentata in notazione tradizionale. Questa modalità di visualizzazione è quindi di tipo temporale e non spaziale, al contrario della rappresentazione mostrata in Fig. 1.

ventina di studenti il cui livello di alfabetizzazione musicale era disomogeneo. Nell’arco degli incontri, ALA è stata potenziata per offrire un ventaglio sempre più ampio di varianti, fino a giungere a quelle presentate nel capitolo precedente, cercando di rispondere alle esigenze di fruizione che si sono presentate durante il lavoro. Così, se inizialmente l’interfaccia offriva soltanto la modalità di rappresentazione visiva a led sulle quattro righe e il metronomo, presto si è arricchita in modo da ampliare e variare le possibilità strategiche di *training*. Utilizzare una nuova risorsa in classe, infatti, richiede un almeno duplice stile di lavoro. In primo luogo lo strumento di cui intendiamo avvalerci a fronte di un dato problema didattico va allestito in modo semplice, efficiente e aperto. Questa fase di allestimento progettuale, razionale e analitica, richiede una mappatura accurata delle caratteristiche della risorsa, e delle possibili attività che è ragionevole ideare a partire da dette caratteristiche. Quando, in secondo luogo, inizia la fase operativa, quando, cioè, la risorsa entra in confronto con la realtà didattica, con gli esiti che concorre a produrre, si valorizzano la flessibilità dello strumento che usiamo e del contesto che abbiamo progettato, nonché le capacità creative dell’insegnante. Si tratta di aspetti profondamente interdipendenti: una risorsa può avere vocazione didattica “aperta” ma rischia il silenzio se non opportunamente contestualizzata e se l’atteggiamento dell’insegnante, oltre che vigile, non è sperimentalmente aperto. Per questa ragione i tipi di percorso che seguono sono nati dal costante monitoraggio dell’attività stessa e dall’aggiustamento di consegne e strategie di lavoro, prodotte in situazione. Nascono dall’invenzione e dall’osservazione, dall’ascolto e dall’attenzione all’evidenza di ciò che accade. Pertanto, le indicazioni fornite hanno natura contestuale, non costituiscono regole ma esempi, da cui poter trarre ispirazione e non istruzioni rigidamente preordinate.

#### 4.1 Tipi di percorso

Durante la sperimentazione del *training* con ALA, tre tipi di percorso sono parsi di significativa importanza.

##### 4.1.1 Primo percorso

Una prima tipologia di percorso ha visto gli esercizi ritmici svolgersi cominciando da una iniziale ridondanza multimediale di rappresentazione dei *pattern* (rappresentazione visiva + esecuzione sonora da parte del sistema) e della pulsazione (metronomo + cursori visivi), per poi proseguire con una progressiva sottrazione in corso di esercizio di una delle modalità (sonora o visiva), fino alla sottrazione completa di entrambe le modalità (nessuna rappresentazione né dei pattern né della pulsazione ad alcun livello). Alla fine di questa progressione, non solo i gruppi di studenti hanno saputo proseguire autonomamente l'esercizio a memoria e in assenza di una qualsiasi sincronizzazione dall'esterno, ma anzi, la progressiva sottrazione di riferimenti ha dato il più delle volte luogo a una miglior prestazione ritmica. Questo fatto sembra suggerire come il carico attentivo inizialmente rivolto alla lettura e decodifica della consegna ritmica, liberatosi per effetto delle sottrazioni di cui sopra, abbia reso disponibile una maggiore concentrazione performativa e di auto-controllo, ad esempio come il musicista che appare più efficace qualora suoni a memoria anziché leggendo, o similmente l'attore più convincente qualora reciti a memoria.

##### 4.1.2 Secondo percorso

Una seconda tipologia di percorso sperimentato comincia similmente al precedente con la stessa ridondanza multimediale di rappresentazione dei *pattern* e delle pulsazioni, e parimenti conosce una graduale sottrazione di riferimenti. Tuttavia, anziché finire come il percorso precedente con una totale assenza di indicazioni, esso finisce con la sola rappresentazione dei *pattern* nello spazio attraverso le quattro righe di led. Questo punto d'arrivo risulta così molto vicino, come attività di decodifica da parte dell'utente, alla tradizionale lettura di una partitura a più voci, se non fosse per l'uso della scrittura intermittente anziché di quella simbolica. E' parso, dunque, molto importante questo percorso in prospettiva di un successivo passaggio proprio alla scrittura musicale tradizionale, al conseguente esercizio di lettura, decodifica ed esecuzione poliritmica e polimetrica di gruppo.

##### 4.1.3 Terzo percorso

Una diversa (e inversa) tipologia di esercizio è stata affrontata richiedendo agli studenti di individuare la pulsazione ricavandola dal solo ascolto dei *pattern* eseguiti dal sistema, o dalla sola loro visualizzazione tramite il led singolo. Si è potuto osservare così, in termini di tempo impiegato, una maggiore facilità nel ricostruire la pulsazione, cioè nell'organizzare internamente il materiale ritmico, nei casi del solo canale uditorio o della compresenza di entrambi i canali, rispetto al caso del solo canale visivo. Questo sembra suggerire che l'organizzazione ritmica, per lo meno di tale ordine, sia facilitata se avviene per mezzo della modalità sonora rispetto a quella visiva.

#### 4.2 Consegne verbali

Per quanto riguarda le consegne verbali date agli studenti, come già accennato queste si limitano, nel caso di ALA, ad esempio, all'indicazione di eseguire un suono di tipo percussivo (con la voce, con le mani o con l'ausilio di strumenti musicali) per ogni led acceso, e poco altro, come l'identità tra singolo led e singola pulsazione nella prima riga. Da questa il passaggio alle righe successive, con le loro suddivisioni sempre più fitte, appare spontaneo e non richiede particolari chiarimenti verbali. C'è infatti da osservare che una didattica musicale efficace richiede che strumenti, risorse, strategie adottate durante il lavoro, portino a interferire verbalmente il meno possibile durante l'esecuzione. Ciò a tutto vantaggio di una condizione di attenzione, ascolto e coordinazione non "disturbata" da fattori distraenti, ma anche frutta con continuità, quindi con minor dispendio di tempo ed energie. Un sistema interattivo che riduca la necessità delle istruzioni verbali per condurre l'azione musicale, e che consenta i margini di fisiologico "aggiustamento" della propria prestazione, può opportunamente garantire una miglior esperienza temporale della propria azione, quindi non comprometterne continuamente la dinamica tipicamente musicale. Quando la prestazione stessa è immediato indicatore dell'efficacia dei propri processi di lettura, o di quelli del gruppo e del gruppo rispetto agli altri, il sistema presenta motivi di interesse didattico. ALA ha mostrato di consentire un lavoro continuo (sotto ogni profilo) nel tempo, costituendo una buona simulazione per situazioni anche complesse di esecuzione di gruppo, in forma strettamente correlata all'azione altrui. La *performance* risulta un indicatore dei livelli di apprendimento quando è osservata rispettando i tempi "fisiologici" di apprendimento e consolidamento della prestazione stessa: il sistema deve essere utilizzato con flessibilità e adattamenti capaci di offrire i tempi necessari al sintonizzarsi di tutti. Il tempo collettivo è un'esperienza squisitamente corale, ossia un tempo di ascolto integrato all'azione altrui in modo che anche la propria si coordini a quella degli altri<sup>6</sup>. ALA par offre buone condizioni d'esercizio perché ciò si realizzi spontaneamente, dopo il breve tempo necessario all'assestamento di tutti.

### 5. OSSERVAZIONI CONCLUSIVE

ALA è nato come gioco-verifica di abilità ritmiche. Si è evoluto come strumento di *training* per la lettura di gruppo, con cambiamenti estemporanei che sollecitano un alto livello di attenzione, come pure di ascolto della prestazione propria e altrui. I contenuti ritmici sono diventati via via sempre più complessi, consentendo, come si era già osservato da un punto di vista metodologico, la pratica su combinazioni molto impegnative in una condizione di rappresentazione tuttavia semplice.

La sottrazione a volte dell'informazione visiva, a volte del suono di riferimento metronomico, a volte del *pattern* eseguito anche da computer, ha consentito interessanti osser-

<sup>6</sup> L'operatività di gruppo implica il tenere in considerazione differenze anche sostanziali negli stili, nelle preferenze, e nei tempi di apprendimento di ciascun membro del gruppo. In proposito vedasi [8].

vazioni sulla prestazione esecutiva da diversi punti di vista, al variare dei livelli di combinazione e complessità delle attività proposte. Per il momento ALA resta un sistema per l'esercizio e la verifica, capace di evidenziare quanto, nel vivo della prestazione musicale, entrino componenti varie e differenziate, capaci di incidere ora sulle prestazioni percettive, ora attente, ora mnestiche [4], ora di più generale controllo coordinatorio, costringendo così l'insegnante a ben più approfondita riflessione su quanto "chiede" di realizzare. "Andare a tempo" è un'espressione che usiamo abitualmente. Ma cosa significa dal punto di vista operativo? In altri termini: come si concorre a costruire prestazioni musicali *altri* a tempo? Sembra ragionevole considerare [9] la "stratificazione" di azioni interdipendenti che realizzano la prestazione musicale, come espressione di diversi e coordinati livelli di gestione del pensiero-azione musicale. Il sistema ALA è stato sviluppato proprio per rendere percepibili, manipolabili e variabili nei diversi livelli le azioni che costituiscono l'andare a tempo. Per l'insegnante, pertanto, pensare l'andare a tempo come azione complessa costituisce un orientamento metodologico utile a decentrare l'azione dalle molte e consolidate pratiche didattiche usuali, dove i livelli d'intervento variano per quantità, qualità e rapporto delle diverse componenti in gioco. Il contributo delle nuove tecnologie consiste nel far da lenze per cercare strategie di lavoro in grado di integrare approssimi alla complessità, di amplificare la rappresentazione che di tale complessità abbiamo e di allontanarci dall'illusione di "risolverla". Una visione più ampia dell'educazione musicale dovrebbe concorrere ad abrogare la paura, forse ancora residuale in tanti docenti, di tentare vie diverse, come stile di lavoro propriamente musicale e non come *extrema ratio* di fronte a un problema.

## 6. BIBLIOGRAFIA

- [1] F. Blezza. *Studiamo l'educazione oggi. La pedagogia generale del nuovo evo*. Osanna Venosa, Potenza, 2005.
- [2] F. Blezza. *Scritti di Pedagogia professionale*. Università degli Studi di Trieste, a.a. 1998/99.
- [3] T. Bole. Un videogioco per il training ritmico. In *Atti del XVIII CIM - Colloquio di Informatica Musicale*, pp. 143–147, 2010.
- [4] S. Bonino. *Dizionario di psicologia di sviluppo*. Einaudi, Torino, 1994.
- [5] M. Brandao, G. Wiggins, Á. G. Wiggins, e H. Pain. Computers in music education. In *Proceedings of the AISB Symposium on Musical Creativity*, 1999.
- [6] J. Bruner. *La mente a più dimensioni*. Laterza, Roma-Bari, 2000.
- [7] J. Bruner. *La cultura dell'educazione*. Feltrinelli, Milano, 2001.
- [8] A. Cadamuro. *Stili cognitivi e stili di apprendimento. Da quello che pensi a come lo pensi*. Carocci, Roma, 2004.
- [9] M. Critchley e R.A. Henson. *Music and Brain. Studies in the Neurology of Music*. Heinemann Medical B., London, 1987.
- [10] S. Cunningham. Potential for unification of musical standards with i-Maestro. In *Proceedings of the Second International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS)*, volume for Workshops, Tutorials, Applications and Industrial, pp. 140–141, 2006. [www.axmedis.org/axmedis2006](http://www.axmedis.org/axmedis2006).
- [11] E. Damiano. *L'azione didattica. Per una teoria dell'insegnamento*. Armando, Roma, 1993.
- [12] R. B. Dannenberg, M. Sanchez, A. Joseph, R. Joseph, R. Saul, e P. Capell. A computer-based multimedia tutor for beginning piano students. *Journal of New Music Research*, 19(2-3):155–173, 1990.
- [13] R. B. Dannenberg, M. Sanchez, A. Joseph, R. Joseph, R. Saul, e P. Capell. Results from the piano tutor project. In *Proceedings of the Fourth Biennial Arts and Technology Symposium*, pp. 143–150, 1993.
- [14] D. Fober, S. Letz, e Y. Orlarey. IMUTUS an interactive music tuition system. In *Proceedings of SMC Conference*, 2004.
- [15] 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.
- [16] T. Koerselman, O. Larkin, B. Ong, N. Leroy, J. Lambert, D. Schwarz, F. Guedy, N. Schnell, F. Bevilacqua, e K. Ng. SDIF integration in i-Maestro gesture tools. In *Proceedings of the 4th i-Maestro workshop on technology-enhanced music education*, pp. 15–20, 2008. All'interno dell'8th International Conference on New Interfaces for Musical Expression (NIME).
- [17] K. Lewin. *La teoria, la ricerca, l'intervento*. Il Mulino, Bologna, 2005.
- [18] N. McKenzie e D. Crombie. Creating accessible interfaces for i-Maestro learning objects. In *Proceedings of the Second International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS)*, volume for Workshops, Tutorials, Applications and Industrial, pp. 132–139, 2006. [www.axmedis.org/axmedis2006](http://www.axmedis.org/axmedis2006).
- [19] K. Ng, B. Ong, T. Weyde, e K. Neubarth. Interactive multimedia technology-enhanced learning for music with i-Maestro. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp. 5673–5678, 2008.
- [20] K. Ng, T. Weyde, e P. Nesi. I-Maestro: Technology-enhanced learning for music. In *International Computer Music Conference Proceedings*, 2008.

# Movements Classification on the Laban's Theory of Effort Time Axis

**Luca Roverelli**

Casa Paganini - InfoMus, Universitá di Genova  
luca.roverelli@gmail.com

**Nicoló Zereik**

Casa Paganini - InfoMus, Universitá di Genova  
nicolo.zereik@gmail.com

## ABSTRACT

The work is part of the European EU ICT project M.I.R.O.R. (Musical Interaction Relying On Reflexion) [1], that aims at developing innovative interactive systems to learn and teach music, based on the reflexive interaction paradigm, to children under the age of 12. The specific objective of this work is to build a classifier able to recognize whether a specified gesture is quick or sustained, according to Laban's "Time" axis, as defined in "Laban's Theory of Effort". Computational models to measure automatically such a Laban's Time feature contribute to the development of systems for embodied music learning based on full-body non-verbal communication.

## 1. INTRODUCTION

In a world in which the technology is introduced in different fields, completing them, a technological turn in educational tasks has been tried, to help children (in particular those who have disabilities) to learn in a new interactive way. The MIROR Project deals with the development of an innovative adaptive system for music learning and teaching based on the reflexive interaction paradigm. The platform will be developed in the context of early childhood music education. It will act as an advanced cognitive tutor, designed to promote specific cognitive abilities in the field of music improvisation, both in formal learning contexts (kindergartens, primary schools, music schools) and informal ones (at home, kinder centres, etc.). The reflexive interaction paradigm is based on the idea of letting users manipulate virtual copies of themselves, through specifically designed machine-learning software referred to as interactive reflexive musical systems (IRMS). By definition IRMS are able to learn and configure themselves according to their understanding of learner's behaviour. The M.I.R.O.R. project proposes to extend the IRMS paradigm with the analysis and synthesis of multisensory expressive gestures to increase its impact on the musical pedagogy of young children, by developing new multimodal interfaces. The project will contribute to promoting the reflexive interaction paradigm not only in the field of music learning but more generally as a new paradigm for establishing a synergy between learning and cognition in the context of child/machine interaction.

*Copyright: ©2012 Luca Roverelli 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.*

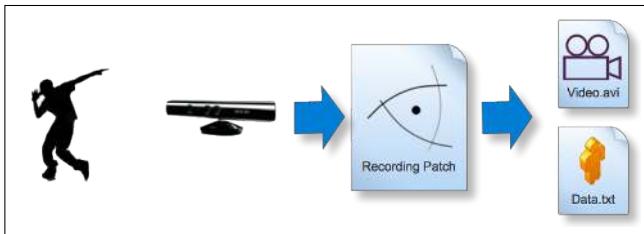
Our work is inserted in the M.I.R.O.R. project and based on "Laban's Theory of Effort", introduced by choreographer Rudolf Laban in 1947 [2, 3], to characterize and explain the different expressiveness of different gestures [4], according to 4 dimensions: Space, Weight, Flow and Time. The Space dimension focuses on whether the trajectory of a gesture is direct or flexible. Weight measures the "strength" of a gesture. Flow shows whether a gesture is bounded or free. The Time axis measures whether a gesture is controlled or not: it is related to the concept of impulsiveness of a gesture. In particular, this axis has as extremes gestures being whereas quick or sustained. Quick movements are rapid and impulsive, while sustained ones are controlled (i.e. they can be suspended and restarted without destroying the gesture itself). For example a "jump" is classified as "quick".

This research aims at developing a software system for the classification of two predefined gestures in the Time dimension. The work has been organized in different steps. The first step was to choose the gestures to analyze; it should be significant for the educational tasks of the related MIROR application. In order to permit the analysis it was necessary to capture some videos of different subjects performing the chosen gestures. Then an evaluation of the recorded videos has been made in order to choose the relevant characteristics of the gestures. The successive step was to implement a system able to automatically extract the chosen features in order to classify the gestures. The last step was try the classification with the chosen characteristics and validate the whole system, elaborating the results.

The research work adopts the Microsoft Kinect sensor [5] to extract 3D movement data in real-time, the Eyesweb Environment [6] for the real-time analysis of the extracted data, and the R environment for classification [7].

## 2. EXPRESSIVE GESTURES IN LABAN'S TIME DIMENSION

The first step is the choice of the movement the final user will have to make in the specific application on music education. Considering that final users of the outcome application of this project are children under the age of 10-12 years old, it is necessary to choose gestures that are natural and easy for them to perform and to conceive. It is also important to take into account the context in which the application will be used. In our scenario, children are told a story to make them pay attention to their full-body activity and to the sonic environment they explore, so the gesture should be easy and intuitively emerging in the story. Fur-



**Figure 1.** The "Recording" EyesWeb patch: it creates a video and a text file with the 3D position of the skeleton joints of human body.

thermore, it is important that gestures can be clearly classified as *quick* or *sustained* according to the way in which they are performed. We do not enter here in further details of the music education application. In compliance with the previous considerations and to the storyboards selected in the specific educational environments, we have chosen a few types of gesture. One of the considered gestures is the gesture of a woodcutter, who at first cuts down a tree with the axe and then cuts the tree branches with the saw. The first is considered as a *quick* gesture, while the later is the *sustained* movement.

### 3. GESTURES RECORDINGS

In this step we captured the performances of these gestures by five subjects, who freely played the predefined quick and sustained gestures. The first thing to consider is that these five subjects are all adults and males. In fact children who are asked to play the same movements may perform them in a different way. The subjects were told only to play the movement of the woodcutter and nothing more, not even which movement were *quick* and which *sustained*.

Video capture and movement analysis have been performed with the EyesWeb [6] patch. This patch uses the information extracted by Kinect [5] and saves it on a text file. This information consists of the 3D position of the skeleton joints of the human body frame by frame. Hence this patch creates a text file and a related video file. Figure 1 shows how it works.

Once all the recordings were available, all the text files describing the gestures have been cut by creating a segmentation file to indicate the interesting part of the videos and then such interesting parts were extracted from the video. These two operations are made with two Java programs, that we have expressly created: "TexterMiro" and "GenerateDataset". The segmentation file is a text file consisting of a number (0 or 1) for each frame of the video, 1 meaning the corresponding frame is interesting and 0 that it is not. Then, using the information in the segmentation file, the text files describing the videos have been cut to extract the interesting parts.

### 4. FEATURES SELECTION

Initially, we have searched for some well known gesture features in literature. For example in [8] are described some features that can be used to classify gesture, like

*Quantity of Motion* or *Impulsiveness*, and which may be useful also for our work. Otherwise, in [8], the sensors used to track the gesture of the user are normal video cameras, while for our work is available a more precise tracking of the user due to the data extracted by the Kinect sensor.

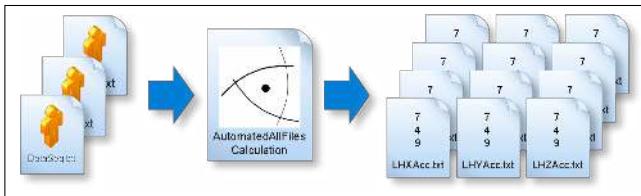
To select what gesture features are most significant in our context, we carried out a qualitative analysis on the selected video segments. By viewing the videos, we noted that only few parts of the body are involved in performing the movement: the main parts are hands and arms. By observing differences between the *quick* gesture and the *sustained* one we have noted that the first one is a rapid and clean gesture, while the second one is somehow repeated. The main characteristic is the change of direction of the hand in the *sustained* movement, that is repeated many times. In this case the velocity and acceleration of the hand changing direction should be 0.

$$\underline{V}_2 = \underline{X}_2 - \underline{X}_1 \quad (1)$$

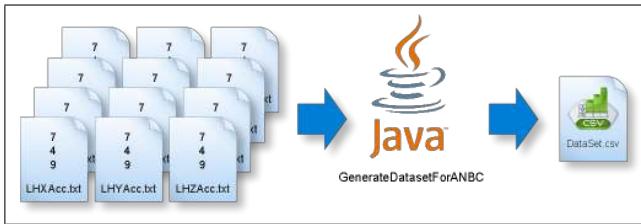
$$\underline{A}_2 = \underline{V}_2 - \underline{V}_1 \quad (2)$$

Since the 3D position of the hand is known at any time it is easy to compute the 3D velocity and the 3D acceleration, as showed in Formula 1 and 2, considering that the hand is in  $\underline{X}_1 (x_1, y_1, z_1)$  in  $t_1$  and in  $\underline{X}_2 (x_2, y_2, z_2)$  in the successive instant of time  $t_2$ .

Since the change of direction implies that acceleration and velocity should be 0 in that instant, there would be a change of sign of these two features, thus implicating a *zero-crossing* in the velocity and in the acceleration. Since in the *sustained* movement the gesture is repeated, *zero-crossings* should be somehow periodic. Thus we chose to extract the time distance between a *zero-crossing* and the previous one. This distance in the *sustained* gesture should be almost equal each time, while for *quick* gesture should be almost random. This way a total of 12 features were extracted: the distances between *zero-crossings* of the 3 component of velocity and acceleration for both hands. The use of such features generates an issue: a different number of values (*zero-crossing*) may be extracted in the same time window. So for a feature we can have a higher number of data with respect to another one, because distances are registered when a *zero-crossing* arises. Hence each of these 12 features can be considered as an array of variable size. For applying machine learning techniques it is important to have data with the same size. The solution is to compute the variance of these distances: the variance of distances coming from the *sustained* gesture should be low, while the one coming from the *quick* movement should be higher. This method implies the variances to be calculated at the end of the performance. However, such a choice would not allow the real time execution of the application. This issue is avoided by using a method called *windowing*. This technique consists in considering parts of the video of  $N$  frames each one overlapped for  $\mu$  seconds to the previous and to the following one.



**Figure 2.** The patch "AutomatedAllFilesCalculation" takes as input the text files describing a gesture and produces as output 12 files consisting of the distances between the zero-crossings of velocity and acceleration of both hands.



**Figure 3.** The program "GenerateDatasetForANBC" takes as input the 12 text files consisting of the distances between zero-crossings and generates the dataset.

## 5. IMPLEMENTATION

To extract the features we created an EyesWeb patch called "AutomatedAllFilesCalculation". This patch is able to compute the requested measures on all the text files created by Kinect describing a gesture, but only if the files are named in a standard way ("Quick" or "Sustained" plus an incremental number ".txt"). Hence the patch takes as input the path of the text files on which it calculates the distances between *zero-crossings* of velocity and acceleration of both hands and save them as a list in a file, thus creating 12 files. Figure 2 shows how it works.

The next step is to calculate variances, thus creating the final dataset on which classification is carried out. This operation is performed with a Java program we created, named "GenerateDatasetForANBC". This program is also responsible for windowing the videos; it takes as input the files with the distances between the *zero-crossings* and the parameters for the windowing operation and generates automatically the final dataset to be used for classification from all the files in a folder (again named in a standard way). Its functioning is shown in Figure 3.

## 6. CLASSIFICATION AND RESULT

The final step is to proceed with classification and validation. There are different possibilities deriving from the opportunity to specify different parameters for the windowing operation. In addition there is one more chance: to use the dataset generated without the windowing operation as training set. For these reasons we chose to generate different datasets named as:

- dataset #1: with window size of 1.5 seconds and overlap of 0.5 seconds.

Dataset	Naive Bayes	SVM	KNN opt
1	19.56	18.47	11.41
2	17.11	13.77	13.55
3	13.51	10.81	9.72
4	7.02	7.56	7.02
5	10.42	10.42	11.30

**Table 1.** Percentage of errors using as training set, the dataset generated by not windowed videos.

Dataset	Naive Bayes	SVM	KNN k=1
1	8.9–23.2	10.7–17.8	8.9–17.8
2	10.4–17.0	10.4–12.6	3.7–8.9
3	7.1–17.9	3.5–10.7	5.3–14.3
4	3.6–10.7	7.1–10.7	3.6–10.7
5	8.8–13.9	8.1–11.0	3.7–8.8

**Table 2.** Percentage of error generated by the classification made splitting dataset in training and validation set with proportion of 70/30.

- dataset #2: with window size of 2.0 seconds and overlap of 0.25 seconds.
- dataset #3: with window size of 2.0 seconds and overlap of 0.5 seconds.
- dataset #4: with window size of 3.0 seconds and overlap of 0.5 seconds.
- dataset #5: with window size of 3.0 seconds and overlap of 0.25 seconds.

All the applied classification methods were tested on these 5 datasets to check which one performed better.

Firstly a method which consists of using as training set the dataset built on the non-windowed videos was tested. The validation sets are the windowed datasets. The used classification methods are Naive Bayes [9], SVM [10] and KNN (Nearest Neighbour with k optimized) [9]. Table 1 reports the percentage of errors for each method on each dataset.

It must be said that the parameter K that generated the best result with KNN was set to 1 for all the datasets. Hence we decided to fix this parameter for the further classifications.

The second method consists of the use of 70% of the dataset as training and the other 30% as validation. The same classification algorithm produced the results shown in Table 2.

The values reported are the minimum and the maximum error observed in different execution of the methods; this is because of the randomness splitting of the dataset. At first sight can be observed that the use of the dataset generated from not windowed videos seems to perform worse than the other method. A further piece of information that can be extracted from this first results is that datasets #2 and #5 seem to perform better than the others. Also datasets #3 and #4 perform well but they have a greater range of

Dataset	Naive Bayes	SVM	KNN k=1
1	18.5 – 19.7	13.6 – 15.4	13.0 – 14.5
2	15.0 – 15.6	8.7 – 9.8	6.8 – 7.8
3	10.8 – 12.5	9.1 – 11.8	8.4 – 10.3
4	8.6 – 9.3	7.3 – 8.9	7.5 – 8.7
5	11.9 – 12.2	7.2 – 8.5	5.9 – 7.4

**Table 3.** Percentage of error generated by the classification made with the bootstrap method.

Dataset	Naive Bayes	SVM	KNN k=1
1	16.7 – 19.4	11.1 – 13.9	5.5 – 16.7
2	14.5 – 16.7	6.7 – 8.9	3.3 – 6.7
3	5.6 – 13.9	8.3 – 13.9	2.8 – 13.3
4	5.6 – 11.1	5.5 – 8.3	2.8 – 8.3
5	10.0 – 12.2	6.7 – 7.8	2.2 – 5.5

**Table 4.** Percentage of error generated by the classification made with the Cross validation method.

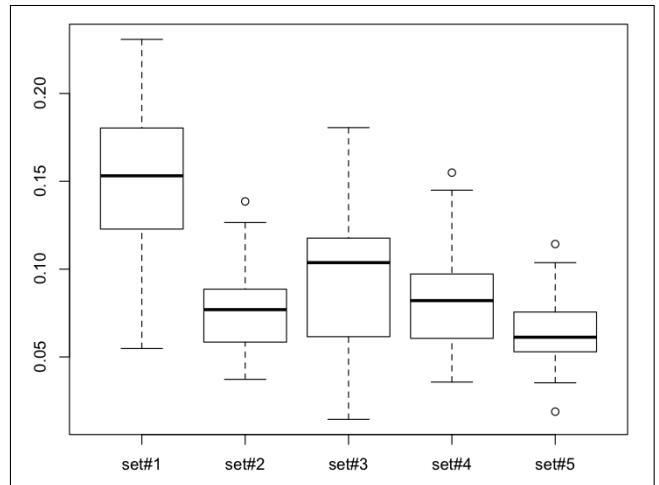
error, thus meaning they are more sensitive to randomness. Dataset #1, instead, seems to have very poor performances with respect to the other.

The next method used is *Bootstrap* [9], a technique used to approximate the generalization error, i.e. the error generated by the classification made on samples not belonging to the training set. Another similar method is *Cross Validation*. These two techniques have been used to estimate more precisely the error without having too wide ranges as experienced for previous method; they consist in different iterations in each of which a smaller replica of the dataset is used as training set and the other one is used as validation set. The outcomes of these algorithms are shown in Tables 3 and 4. Also in this case minimum and maximum error values are reported because these methods include a degree of randomness too. In this case it can be noticed that for some datasets ranges are lower. Datasets that have generated high ranges also with these methods are probably less robust.

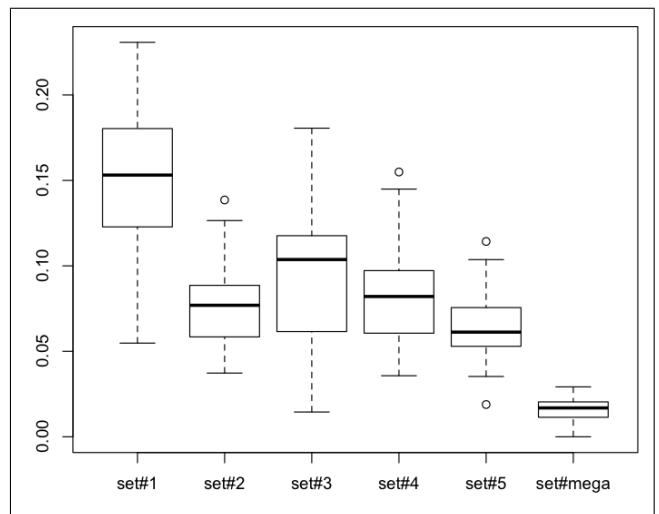
With these more precise methods the best datasets seem to be #2, #4 and #5. Figure 4 shows the boxplot of the KNN classification on the five datasets using *Bootstrap*.

Boxplot is a particular plot that shows six values: the two horizontal lines at the extremes are minimum and maximum values; the two horizontal lines that delimit the box are the first and second quartile (value that divides the population in 2 groups of proportions respectively 25/75 and 75/25), the black marked line is the median (value that divides the population in 2 groups of proportion 50/50), and the circles are the outliers, values considered anomalous. Thus the "boxes" represent the IQR (Inter Quartile Range), the values between which stands the 50% of the population nearer the median value.

From this plot it seems that the best dataset is #5, with a median of the errors of about 6%. It is important also to notice that the range of errors shown by this dataset is lower than the others, thus meaning that this dataset is more robust, even if results obtained by dataset #2 are really simi-



**Figure 4.** Boxplots of the error distributions of classification with Bootstrap on the five datasets with KNN.



**Figure 5.** Boxplots of the error distributions of classification on the five datasets and dataset #mega.

lar.

It must be remembered that these outcomes are based on data obtained by gesture performed by adults. It is possible that children perform these gesture in a completely different way. In addition it must be considered that these results are obtained by videos on only 5 subjects; other people may perform the same gestures with different timing, thus implicating a different best solution in the research of optimal windowing parameters. In such a case the outcomes may be very different. Considering this, a new technique has been thought up: to unify the 5 datasets to create a more heterogeneous one; we have called it #mega set. With the KNN algorithm, the best one, a greatly better result was observed. Figure 5 shows the result of this method compared with the same method on the previous five datasets. The median of the percentage of error is 1.68% with a maximum value of 2.91%, obtaining then 98% of correctness in classification.

## 7. CONCLUSIONS

Most important conclusion to highlight is that the primary goal has been reached with a great percentage of correctness. It must be noted, however, that there has been few data to build the classifier as well as to validate it. Methods to overcome this problems have been used, as described before. But it must be taken into account that different persons can make the same gestures in very different ways. For this very reason the #mega set has been built. But this set can't be necessary the best solution when running the application in the context for which it has been created. For this reason, and also because the analyzed gestures have been performed only by adults, a preliminary test to choose the best algorithm and the best dataset is necessary.

## 8. ACKNOWLEDGEMENTS

We would like to thank the staff of InfoMus lab, in particular Gualtiero Volpe, Barbara Mazzarino and professor Antonio Camurri, who guided us during the progress of the project.

This work is partially supported by the EU 7FP ICT MIROR Project n 258338 (2010 - 2013, [www.mirorproject.eu](http://www.mirorproject.eu)).

## 9. REFERENCES

- [1] "M.i.r.o.r. project, <http://www.mirorproject.eu>."
- [2] R. von Laban, *Effort*. Macdonald & Evans, 1967.
- [3] R. Laban, L. Delfini, and F. Zagatti, *La danza moderna educativa*. I libri dell'icosaedro, Ephemeria, 2009.
- [4] A. Camurri, G. D. Poli, M. Leman, and G. Volpe, "A multi-layered conceptual framework for expressive gesture applications," in *In Proceedings of MOSART: Workshop on Current Directions in Computer Music*, pp. 29–34, 2001.
- [5] "Microsoft kinect, <http://www.xbox.com/it-it/kinect>."
- [6] "Eyesweb platform, [http://www.infomus.org/eyesweb\\_eng.php](http://www.infomus.org/eyesweb_eng.php)."
- [7] "R project for statistical computing, <http://www.r-project.org/>"
- [8] G. Volpe and A. Camurri, "A system for embodied social active listening to sound and music content," *J. Comput. Cult. Herit.*, vol. 4, pp. 2:1–2:23, Aug. 2011.
- [9] R. Duda, P. Hart, and D. Stork, *Pattern classification*. Pattern Classification and Scene Analysis: Pattern Classification, Wiley, 2001.
- [10] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sept. 1995.



# EGGS (Elementary Gestalts for Gesture Sonification)

## Risorse informatiche per l'educazione e didattica musicali

Cristina Fedrigo

Conservatorio di Musica

“Giuseppe Tartini” di Trieste - Italia

cristina.fedrigo@tin.it

Sara Radin

Conservatorio di Musica

“Giuseppe Tartini” di Trieste

Scuola Media “Giancarlo Roli” di Trieste

Italia

sara.radin@virgilio.it

### ABSTRACT

EGGS è un sistema sviluppato da Maurizio Goina e Pietro Polotti (Scuola di Musica e Nuove Tecnologie, MNT, Conservatorio di Trieste) [14] e testato nell'ambito del Programma di ricerca e sviluppo di risorse per la multimedialità in ambito pedagogico dalle autrici (Scuola di Didattica della Musica, DdM, Conservatorio di Trieste). Il presente contributo si basa sulla sperimentazione svolta nelle scuole primarie della Provincia di Trieste per testare EGGS in reale contesto educativo e didattico. Focalizzando l'attenzione su alcuni aspetti inerenti ai processi di ascolto [1], apprendimento [13], controllo [2] e coordinazione di prestazioni complesse come quella musicale [3], il progetto ha inteso porre in evidenza le potenzialità di EGGS (nato per usi professionali e creativi) e svilupparne caratteristiche e accessibilità (uso facile e intuitivo, non ambiguità nelle istruzioni, carattere auto-correttivo<sup>1</sup>, ecc.) per renderlo pienamente funzionale, nei reali contesti di insegnamento ed educazione musicali, a contribuire allo sviluppo di competenze musicali, offrendo una strumentazione flessibile per operare in forma interattiva tra i processi di ascolto, rappresentazione e prestazione musicale. Nel presente contributo l'uso di EGGS viene presentato con riferimento all'esperienza di ascolto e controllo gestuale e motorio dell'altezza sonora, fondamentale a vari livelli di esercizio della competenza musicale.

Copyright: © 2012 Fedrigo Cristina and Sara Radin. 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.

<sup>1</sup>Nel definire auto-correttivo un carattere saliente di EGGS, ci riferiamo, ispirandoci alla lezione montessoriana sul "Materiale di sviluppo", al meccanismo di controllo dell'errore, non operato dall'insegnante ma ottenuto realizzando materiali didattici «in modo tale da denunciare lo sbaglio commesso, sbaglio reso nei materiali più semplici sensibilmente evidente, materializzato, ma che gradualmente richiede invece, per essere rilevato, l'esercizio delle attività superiori.». In [12]: D. Parigi e A. Pettini: *I Metodi di Ovidio Decroly e Maria Montessori*, O.D.C.U., Bologna, 1954, p. 95. Sul materiale di sviluppo, vedasi anche [11]: M. Montessori: *Educare alla libertà*, Mondadori, Milano, 2008, pp. 97-101

### 1. INTRODUZIONE

Poter esprimere la propria musicalità con competenza implica la gestione di una relazione controllabile e consapevole col suono: ciò si realizza in azioni che definiamo musicali ad ogni livello di complessità, anche a partire dalle forme più elementari di espressione.

Il presente lavoro, riferendosi a un progetto di ricerca realizzato nel contesto della scuola di base, offre contenuti di riflessione e indicazioni operative utili ai livelli primari della formazione musicale, ma anche e più ampiamente ai problemi inerenti la formazione del “popolo della musica”, fruttore o interprete che sia.

Indipendentemente dall'età del soggetto che ne è protagonista, un aspetto fondamentale del comportamento musicale è che questo è esercitato con controllo e autonomia (nei limiti del livello padroneggiato di competenza musicale), e mediatori di tale gestione sono i processi d'ascolto e gesti/movimenti intenzionali.

Il suono e le sue caratteristiche sono alla base del processo di creazione ed espressione musicali. E' materia viva di ricerca per educatori e ricercatori l'indagare contenuti e modi per sviluppare un rapporto ricco e consapevole con il materiale della musica. Più che di sola conoscenza di aspetti del fenomeno sonoro, infatti, serve attrezzarsi per offrire esperienze utili del suono, anche in forma ludica e creativa, e gestite consapevolmente.

Dove inizia un uso cosciente del suono e per quali caratteristiche? Che significa cantare e suonare con adeguato controllo dell'intonazione? Quando lo studente dovrebbe scoprire la bellezza del suono che può creare o interpretare? Ci è parso che la sola risposta ragionevole sia fin dall'inizio, sfruttando opportuni strumenti e strategie per procedere gradualmente (dalle pratiche più semplici ed elementari, ma parimenti musicali a quelle via via più complesse). Auspiciamo che questa ricerca possa offrire un esempio di come le risorse multimediali e, in questo specifico caso l'uso di suoni puri generati informaticamente, possano offrire territorio di sviluppo di controllo e consapevolezza dei fenomeni sonori di base (e non solo). Abbiamo, in quest'occasione, focalizzato la nostra attenzione sulla competenza correlata al controllo, attraverso i processi di ascolto, dell'altezza del suono prodotto e gestito attraverso gesto e movimento. Ci siamo rivolti a gesti elementari, semplici, che rappresentano il correlato di come il suono cambia nel tempo in ordine alla sua frequenza, caratteristica strutturale del suono come dell'organizzazione musicale. Ascendendo e discendendo, restando costante o muovendosi in paral-

lelo, componendo un lessico elementare di profili attraverso la combinazione di gesti semplici, abbiamo cercato di rendere "manipolabili" le relazioni assolute e relative che i suoni intrattengono tra loro e per cui costruiamo attraverso l'esperienza rappresentazioni dei fenomeni di cui siamo partecipi.

## 2. IL PROGETTO DI RICERCA

Intitolato *Gestualità, suono, rappresentazione*, e co-finanziato dal Servizio università, ricerca e innovazione Regione Friuli Venezia-Giulia e dalla Fondazione Cassa di Risparmio di Trieste, il Programma di ricerca e sviluppo di risorse per la multimedialità in ambito artistico e pedagogico riservato alle istituzioni dell'alta formazione, ha avuto come responsabili dell'attività di ricerca per aree di competenza, Pietro Polotti (MNT) e Cristina Fedriga (DdM) insieme a Maurizio Goina (MNT) e Sara Radin (DdM); coordinatore del progetto: Paolo Rodda. Il progetto, focalizzandosi sulla relazione tra gesto e suono e le rappresentazioni che, operando in forma estemporanea con EGGS, si sviluppano e vengono gestite, è stato condotto per ottenere indicazioni utili a trarre diversi scenari e funzionalità di applicazione, in contesti di educazione generale e musicale, da potersi rivolgere ad età diverse (in questo progetto circoscritte a quelle della scolarizzazione primaria) e a livelli pure diversi di competenza musicale<sup>2</sup>.

### 2.1 La sperimentazione a scuola

La fase di sperimentazione ha coinvolto, durante l'anno scolastico 2010/11, 16 Scuole Primarie nella Provincia di Trieste, 96 classi (dalla prima alla quinta), 1586 alunni, di età compresa tra i 6 e gli 11 anni). In dettaglio, alla tabella che segue.

<b>CLASSI e NUMERO ALUNNI PARTECIPANTI</b>			
<b>Totale numero prestazioni: 2049</b>			
<b>Classe (di Scuola Primaria)</b>	<b>Numero delle classi</b>	<b>Numero dei bambini</b>	<b>Numero prestazioni per classe</b>
I	20	356	467
II	19	290	324
III	22	403	428
IV	17	243	485
V	18	294	345
<b>Totale</b>	<b>96</b>	<b>1586</b>	<b>2049</b>

<sup>2</sup>Il progetto si è ispirato alle lewiniane linee guida della ricerca, ma l'indagine sul campo circa le caratteristiche di EGGS testate, non ha inteso verificare soluzioni d'uso dello stesso, né condizioni ed esiti di apprendimento dei bambini o loro competenze, quanto unicamente mettere il sistema a confronto con reali contesti e condizioni d'uso, per raccogliere informazioni necessarie a ipotizzare scenari didattici efficaci e flessibili. L'idea di flessibilità - fondamentale per l'azione in contesto con una risorsa culturale - costituisce, tra gli altri, uno dei più preziosi contributi di Kurt Lewin [7] [8] [9].

**Tabella 1.** Sono riportati i numeri relativi (totali e parziali) alle classi, agli studenti e alle prestazioni della fase di sperimentazione.

### 2.2 Fasi di progettazione in sintesi

1. Sviluppare un'interfaccia-utente configurata in modo intuitivo e flessibile, per meglio aderire a contesti, caratteristiche ed esigenze d'uso diversi.

Adattare la risorsa per l'uso nei contesti individuati per la sperimentazione.

2. Progettare la sperimentazione definendone il protocollo.

- Ipotizzare gli utilizzi di EGGS.

- Individuare sedi per la sperimentazione.

- Adeguare e controllare il materiale necessario: computer, amplificatore, torce elettriche, telecamere wii (propriamente il controller della console di gioco Wii Nintendo, ossia il *wii remote*) e treppiedi, batterie ricaricabili e caricabatterie.

- Scegliere i contenuti sonori: tipi e caratteristiche di suono, scala, intervalli, tessiture, in rapporto alle caratteristiche del gesto/movimento.

- Definire i tipi di attività, tempi e modalità di intervento in classe.

- Definire le caratteristiche del *setting* di lavoro, perfezionare comunicazioni adeguate e modalità coerenti di lavoro (anche attraverso la fase di pre – test in classe).

- Definire criteri, strategie e strumenti per verificare e documentare l'esperienza.

2. Coordinare la fase di sperimentazione con la partecipazione degli insegnanti delle scuole coinvolte:

a) offrendo loro criteri e informazioni utili per preparare gli studenti a partecipare e per gestire le condizioni di lavoro e osservazione in classe.

b) quindi, dopo l'esperienza in classe, incontrare gli insegnanti stessi per raccogliere le loro osservazioni e considerazioni in forma libera

3. Analizzare i risultati per ottimizzare il sistema, come pure riflettere sul suo utilizzo con gli insegnanti coinvolti nell'esperienza.

4. Realizzare una sintesi per pubblicare i risultati, per rendere la risorsa disponibile agli interessati accompagnandola con istruzioni e indicazioni d'uso verificate per gli utilizzi in contesti di insegnamento/apprendimento, offrendo suggerimenti per differenti sviluppi nell'uso del sistema.

**Nota:** nel contesto della sperimentazione, EGGS è stato utilizzato seguendo un protocollo rigorosamente controllato ma, accanto a ciò, sono state anche raccolte in maniera informale osservazioni riguardo l'esperienza in classe, quelle degli utilizzatori, problemi inaspettati o risultati interessanti non previsti ...

### 2.3 EGGS in azione

EGGS è un sistema che agisce trasformando i gesti e i movimenti del corpo in suono e segno grafico: la ricezione di un input luminoso (le sorgenti luminose sono applicate al corpo del *performer* o impugnate dagli uti-

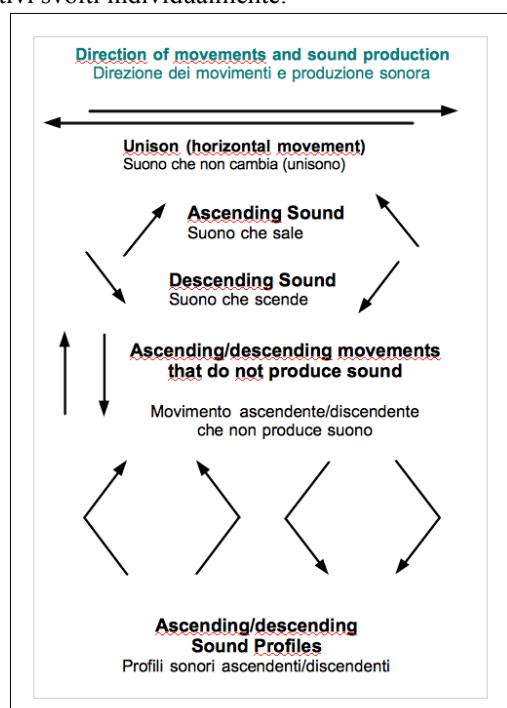
lizzatori) attraverso la/le telecamera/e wii. Per la fase sperimentale a scuola e in altre sedi extrascolastiche, sono sempre state utilizzate due torce luminose, adattate per poter essere facilmente impugnate e maneggiate anche da bambini. In questo stesso contesto non è stata utilizzata la parte grafica del sistema e si è scelto di utilizzare suoni puri per valorizzare la massima chiarezza possibile nell'informazione percettiva rispetto all'altezza sonora, nell'ambito dei registri "centrali" per la nostra funzionalità uditiva.

Il sistema è stato usato chiedendo agli studenti di agire sia ad occhi aperti che chiusi, o, comunque, sia in condizioni di esclusione dell'informazione visiva (ad esempio rispetto al gesto della sperimentatrice o del compagno di prestazione), privilegiando così esclusivamente comportamenti controllati attraverso i processi di ascolto. Le istruzioni verbali sono state strettamente controllate per non influenzare impropriamente le azioni degli studenti.

### 2.3.1 La struttura di ogni intervento in classe

1. Attività preliminare con l'intero gruppo classe (per una durata di max 10') per consentire agli studenti di praticare collettivamente le istruzioni in modo da essere pronti individualmente ad agire col sistema in modo immediato, escludendo quindi fasi di apprendimento, producendo i suoni richiesti secondo istruzioni precise e controllate.

Questa fase è stata resa necessaria dal tempo molto breve che ciascun alunno aveva di lavoro col sistema (max 5'), per chiarire a tutti la relazione tra parole della consegna e corrispondente azione sonora-gestuale; per testare l'immediatezza e comprensibilità d'uso di EGGS a partire da poche istruzioni uguali per tutti, per far partire ogni studente dallo stesso livello quanto a utilizzo della risorsa stessa, senza prevedere apprendimenti preventivi svolti individualmente.



**Figura 1.** Nella figura che precede sono rappresentati i movimenti base per produrre suono / profilo sonoro.

2. Lavorando individualmente o in coppia, gli studenti hanno imitato le azioni della ricercatrice o ne hanno realizzato le istruzioni/consegne.

Gli studenti hanno partecipato all'attività usando sia una che due luci, muovendo una o due mani, sia in sequenza che contemporaneamente, utilizzando tre tipi di gesti elementari: verticale, orizzontale e obliquo, sia con direzione ascendente che descendente, sia da destra a sinistra che da sinistra a destra, muovendo, quando operavano con entrambe le mani, sia le stesse per moto parallelo che contrario (speculare), all'altezza delle tre regioni della dimensione corporea (dal bacino in giù, lungo il tronco, dalla testa in su).

NOTA: la gestione spaziale e collocazione di registro del suono devono essere proporzionati all'altezza dello studente. Pertanto, nel passare da un bambino all'altro, nel caso di stature diverse, la posizione delle telecamere wii veniva riaggiustata, o in termini di distanza maggiore/minore dalla fonte luminosa, o puntando più in alto/basso la telecamera stessa, o entrambe le cose insieme.

### 2.3.2 Setting, criteri di rilevazione e modalità di lavoro

Ogni intervento, della durata di un'ora per ogni classe/gruppo, ha richiesto 10 minuti prima dell'inizio per allestire il *setting* di lavoro, costituito da uno spazio vuoto (di un minimo di 4 x 3 metri ca.), dove collocare le telecamere e i bambini, a seconda delle diverse attività.

Per usare EGGS, è necessario che lo spazio di lavoro sia abbastanza silenzioso e non troppo illuminato. Queste condizioni non sono sempre usuali a scuola, tuttavia, il sistema ha dimostrato di essere flessibile quanto necessario per far fronte a qualche problema incontrato in merito.

Per testare l'efficacia, la flessibilità e facilità d'uso, l'accessibilità e l'auto-correzione del sistema abbiamo assunto come indicatori i comportamenti degli scolari di seguito riportati.

Riprodurre il suono/profilo sonoro all'altezza richiesta.  
Produrre il profilo sonoro nello stesso range, mantenendone pure lo sviluppo nel tempo e nella direzione.  
Coordinare la propria azione nel tempo e nello spazio con quella del partner (producendo suoni/profilo sonori).  
Coordinare e riprodurre suoni/profilo sonori simultaneamente con partner.

Agire con entrambe le mani e due luci individualmente. Esiti favorevole o sfavorevole di ogni singola prestazione sono stati assunti come indicatori di facilità e accessibilità d'uso, auto-correzione del sistema: ogni bambino ha usato EGGS per non più di 5 minuti, così che i risultati non concernono i processi di apprendimento.

TIPO DI ATTIVITA' E NUMERO DI PRESTAZIONI						
ATTIVITA'	NUMERO DI PRESTAZIONI			Distribuzione delle prestazioni in percentuale	Distribuzione della modalità visione/non visione rispetto a ciascuna attività	
	TOTALE	VISIONE	SENZA VISIONE		TOTALE	VISIONE
1 IMITAZIONE SUONO	559	185	374	27,28%	33%	67%
2 IMITAZIONE UN PROFILO	434	202	232	21,18%	47%	53%
3 IMITAZIONE DUE PROFILI	130	72	58	6,34%	55%	45%
4 UN PROFILO IN DUE	516	282	234	25,18%	55%	45%
5 UNISONO IN DUE	120	74	46	5,86%	62%	38%
6 DUE BAMBINI (UN PROFILO)	79	53	26	3,86%	67%	33%
7 DUE BAMBINI (DUE PROFILI)	10	8	2	0,49%	80%	20%
8 DUE BAMBINI (UNISONO)	16	10	6	0,78%	63%	38%
9 IMITAZIONE SUONO (2 MANI)	26	21	5	1,27%	81%	19%
10 IMITAZIONE PROFILO (2 MANI)	28	22	6	1,37%	79%	21%
11 IMITAZIONE 2 PROFILI (2 MANI)	21	17	4	1,02%	81%	19%
12 PARALLELISMO UNISONO (2 MANI)	25	20	5	1,22%	80%	20%
13 PARALLELISMO PROFILO (2 MANI)	22	18	4	1,07%	82%	18%
14 PARALLELISMO 2 SUONI (2 MANI)	8	8	0	0,39%	100%	0%
15 PROFILI PER MOTO CONTRARIO (2 MANI)	14	12	2	0,68%	86%	14%
16 PROFILO E SUONO (2 MANI)	11	9	2	0,54%	82%	18%
17 TELECAMERE DISPETTOSE	30	30	0	1,46%	100%	0%
<b>TOTALE</b>	<b>2049</b>	<b>1043</b>	<b>1006</b>	<b>100,00%</b>	<b>51%</b>	<b>49%</b>

**Tabella 2.** Sono riportati: in nero, il numero totale delle prestazioni e quello per ciascun tipo di attività. Le attività svolte in modalità non visiva sono indicate in blu, quelle con visione in rosso.

#### 2.4 Per leggere le tabelle

E' riportata la descrizione delle attività svolte dagli studenti. Naturalmente, tali attività sono state direttamente proposte per essere imitate o richieste verbalmente dalla ricercatrice.

Gli studenti hanno operato in coppia, nello stesso spazio di lavoro, sia a turno che agendo simultaneamente. Hanno lavorato sia con visione che senza, ad esempio, imitando quanto proposto dalla ricercatrice o dal compagno di coppia, potendo vedere quanto proposto dall'altro o basandosi esclusivamente sul proprio ascolto.

Le prime sei attività elencate nella colonna di sinistra rappresentano la quantità più rilevante di dati riguardo quanto ciascuno studente ha svolto imitando o producendo a richiesta un suono/profilo sonoro. Le attività successivamente indicate si riferiscono a ciò che ogni studente, operando con due mani, o due studenti, lavorando insieme, hanno sperimentato. Esse ci offrono prime indicazioni per indagare diversi e vari utilizzi didattici di EGGS e se lo stesso può effettivamente rispondere con efficacia in situazioni in cui entrano in gioco coordinazioni più complesse nel tempo (ad esempio, giocando a due mani, lavorando in coppia). Queste situazioni d'uso del sistema, in contesto educativo, sono preziose poiché consentono agli studenti maggior libertà nell'esplorare e provare le loro possibilità di interazione con l'esperienza sonora, anche in forma creativa.

Tuttavia, durante la fase di sperimentazione del progetto, abbiamo proporzionalmente ridotto lo spazio dedicato ad attività come queste ultime per focalizzare l'atten-

zione sulle caratteristiche più strettamente basilari dell'utilizzo del sistema, preferendo un tempo breve di prestazione per ciascun bambino a favore del maggior numero di prestazioni possibili. In altre, libere, situazioni, fuori dal programma di ricerca, abbiamo, al contrario, lasciato gli utenti liberi di esplorare e sperimentare, ad ogni età e livello di competenza musicale, in molti modi e creativamente.

Per miglior chiarezza, ecco una descrizione sintetica delle attività richieste.

**1 Imitazione di un suono:** operando con una luce lo studente riproduce

il suono proposto dalla ricercatrice o dal partner.

**2 Imitazione di un profilo sonoro:** operando con una luce lo studente riproduce il profilo sonoro proposto dalla ricercatrice o dal partner.

**3 Imitazione di due profili sonori:** operando con una luce lo studente riproduce il doppio profilo sonoro proposto dalla ricercatrice o dal partner.

**4 Un profilo in due:** operando con una luce lo studente completa il profilo sonoro iniziato dal partner.

**5 Un unisono in due:** operando con una luce lo studente prolunga lo stesso suono iniziato dal partner.

**6 Due bambini e un profilo sonoro:** operando con una luce e simultaneamente, ogni studente esegue lo stesso profilo sonoro.

**7 Due bambini e due profili sonori:** operando con una luce e simultaneamente, ogni studente esegue gli stessi due profili sonori.

**8 Due bambini in unisono:** operando con una luce simultaneamente, ciascuno studente produce lo stesso suono per altezza.

**9 Imitazione di un suono (a due mani):** operando con due luci (entrambe le mani in successione) uno studente riproduce il suono richiesto dalla ricercatrice.

**10 Imitazione di un profilo sonoro (a due mani):** operando con due luci (entrambe le mani in successione) uno studente riproduce il profilo sonoro richiesto dalla ricercatrice.

**11 Imitazione di due profili sonori (a due mani):** operando con due luci (entrambe le mani in successione) uno studente riproduce due profili sonori successivi richiesti dalla ricercatrice.

**12 Parallelismo di unisono (a due mani):** uno studente riproduce l'unisono, operando con entrambe le mani e due luci nello stesso momento.

RISULTATI DELLE ATTIVITA' IN MODALITA' VISIVA					
ATTIVITA'	POSITIVO	MEDIO	NEGATIVO	TOTALE	PARAMETRI OSSERVATI
1 IMITAZIONE SUONO	116	64	5	185	Colloca il suono nella regione delle altezze
2 IMITAZIONE UN PROFILO	184	13	5	202	Riproduce la direzione
	168	28	6	202	Colloca nella regione delle altezze
	172	19	11	202	Riproduce l'estensione temporale
3 IMITAZIONE DUE PROFILI	68	3	1	72	Riproduce la direzione
	51	17	4	72	Colloca nella regione delle altezze
	51	19	2	72	Riproduce l'estensione temporale
4 UN PROFILO IN DUE	183	87	12	282	Mantiene la continuità
	144	106	32	282	Mantiene la velocità
5 UNISONO IN DUE	62	0	12	74	Mantiene l'altezza
	48	23	3	74	Coordina inizio/fine
6 DUE BAMBINI (1 PROFILO)	34	16	3	53	
7 DUE BAMBINI (2 PROFILI)	2	4	2	8	
8 DUE BAMBINI (UNISONO)	9	1	0	10	
9 IMITAZIONE SUONO(2 MANI)	18	3	0	21	
10 IMITAZIONE PROFILO (2 MANI)	19	3	0	22	
11 IMITAZIONE DUE PROFILI(2 MANI)	12	4	1	17	
12 PARALLELISMO UNISONO(2 MANI)	18	2	0	20	
13 PARALLELISMO PROFILO(2 MANI)	17	1	0	18	
14 PARALLELISMO 2 SUONI(2 MANI)	5	3	0	8	
15 PROFILI PER MOTO CONTRARIO (2 MANI)	7	5	0	12	
16 PROFILO E SUONO(2 MANI)	6	3	0	9	
17 TELECAMERE DISPETTOSE	20	8	2	30	
TOTALE	1414	432	101	1947	

RISULTATI DELLE ATTIVITA' IN MODALITA' NON VISIVA					
ATTIVITA'	POSITIVO	MEDIO	NEGATIVO	TOTALE	PARAMETRI OSSERVATI
1 IMITAZIONE SUONO	173	157	44	374	Colloca il suono nella regione delle altezze
2 IMITAZIONE UN PROFILO	205	9	18	232	Riproduce la direzione
	188	34	10	232	Colloca nella regione delle altezze
	194	26	12	232	Riproduce l'estensione temporale
3 IMITAZIONE DUE PROFILI	36	3	19	58	Riproduce la direzione
	25	14	19	58	Colloca nella regione delle altezze
	28	12	18	58	Riproduce l'estensione temporale
4 UN PROFILO IN DUE	151	71	12	234	Mantiene la continuità
	121	92	21	234	Mantiene la velocità
5 UNISONO IN DUE	38	0	8	46	Mantiene l'altezza
	23	18	5	46	Coordina inizio/fine
6 DUE BAMBINI (1 PROFILO)	18	7	1	26	
7 DUE BAMBINI (2 PROFILI)	0	2	0	2	
8 DUE BAMBINI (UNISONO)	5	1	0	6	
9 IMITAZIONE SUONO (2 MANI)	5	0	0	5	
10 IMITAZIONE PROFILO (2 MANI)	6	0	0	6	
11 IMITAZIONE DUE PROFILI (2 MANI)	3	1	0	4	
12 PARALLELISMO UNISONO (2 MANI)	4	1	0	5	
13 PARALLELISMO PROFILO (2 MANI)	4	0	0	4	
14 PARALLELISMO 2 SUONI (2 MANI)	1	1	1	1	
15 PROFILI PER MOTO CONTRARIO (2 MANI)	1	1	0	2	
16 PROFILO E SUONO (2 MANI)	1	1	0	2	
17 TELECAMERE DISPETTOSE	1	1	1	1	
TOTALE	1229	450	187	1866	

**13 Parallelismo di due profili sonori (a due mani):** uno studente riproduce due profili sonori richiesti simultaneamente, operando con entrambe le mani e due luci con moto parallelo.

**14 Parallelismo di due suoni diversi per altezza (a due mani):** uno studente riproduce due suoni di altezza diversa richiesti, operando simultaneamente con entrambe le mani e due luci con moto parallelo.

**15 Due profili sonori a specchio (a due mani):** uno studente riproduce due profili sonori richiesti simultaneamente, operando con entrambe le mani e due luci per moto contrario.

**16 Un profilo sonoro e un suono (a due mani):** uno studente riproduce la combinazione simultanea di un suono e un profilo sonoro, usando entrambe le mani e luci.

**17 Telecamere dispettose:** uno studente ricerca l'unisono girandosi da una telecamera a un'altra, ma queste sono poste a diversa altezza.

Le tabelle (3 e 4) a sinistra presentano i risultati, in termini di successi o fallimenti, per ciascuna attività proposta, sia in **modalità visiva** (scritta in rosso, Tabella sopra) che **non visiva** (scritta in blu, Tabella sotto).

## 2.5 Criteri di rilevazione e valutazione

La valutazione è stata espressa con positivo, medio, negativo. Valutazione positiva è stata attribuita a prestazioni pienamente rispondenti alla consegna per ciascun aspetto osservato. Valutazione media è stata attribuita a prestazioni parzialmente rispondenti alla consegna, per l'aspetto osservato, e negativa per prestazioni non rispondenti.

Seguono alcune precisazioni per le diverse attività.

**1 Imitazione di un suono.** Positivo: lo studente riproduce esattamente la stessa altezza sonora; medio: lo studente colloca il suono in zona prossimale per altezza (entro un intervallo di terza maggiore sopra/sotto il suono); negativo: lo studente colloca il proprio suono oltre tale limite.

**2 Imitazione di un profilo sonoro.**

**3 Imitazione di due profili sonori.** I parametri osservati sono: a) riproduce la direzione del profilo (correttamente / parzialmente / inverte direzione); b) colloca il profilo nella regione delle altezze (correttamente / parzialmente / non colloca) c) riproduce l'estensione temporale (correttamente / accelera o rallenta di poco il modello / altera il tempo di esecuzione del profilo).

**4 Un profilo in due.** a) mantiene la continuità del profilo (scambio coordinato / scambio impreciso / interruzione profilo); b) mantiene la velocità del profilo (correttamente / accelera o rallenta di poco il profilo / altera il tempo di esecuzione del profilo).

**5 Un unisono in due:** a) mantiene l'altezza (sì / no); coordina inizio-fine (scambio coordinato / scambio impreciso / interruzione suono).

**Per tutte le ulteriori attività (da 6 a 17):** sì / parzialmente / no.

Dopo ogni attività la ricercatrice ha sempre commentato positivamente, ma senza valutazione propriamente intesa, la partecipazione dello studente, ringraziandolo per questa. Spesso i compagni di classe hanno commentato spontaneamente la prestazione dei compagni, sottolineandone i risultati, specie il successo.

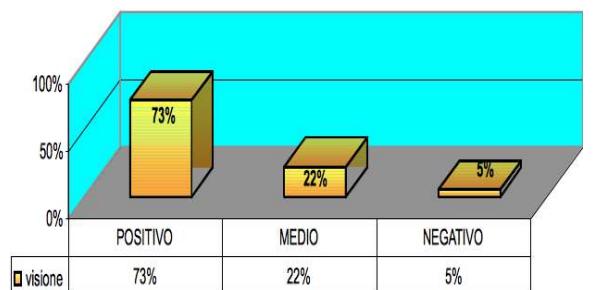
Riteniamo sia importante, dopo questa esperienza di sperimentazione osservare i risultati e riflettere su questi, senza trarre conclusioni definitive ma preziosi spunti di lavoro e approfondimento.

Sia in modalità visiva che non visiva, EGGS è risultato facile da usare per gli studenti in ogni classe. La quantità di successi nelle differenti prestazioni proposte è piuttosto alto e ci incoraggia a considerare il sistema semplice ed efficace. Nessun bambino ha dimostrato di incontrare difficoltà nel suo immediato utilizzo.

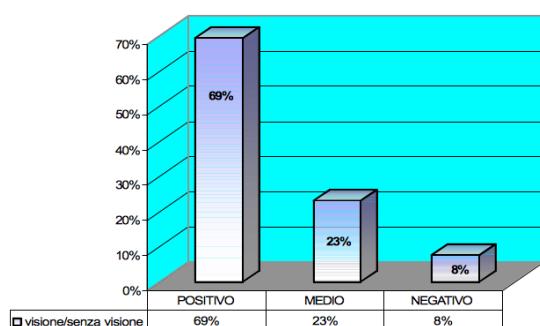
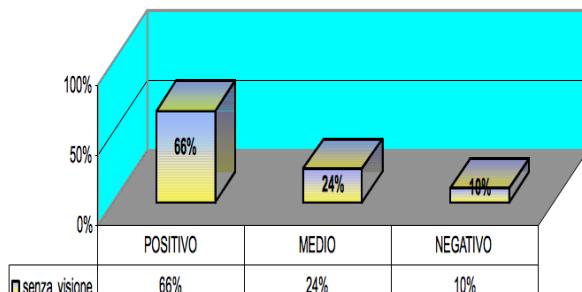
ATTIVITA'	CON VISIONE			SENZA VISIONE			PARAMETRI OSSERVATI
	POSITIVO	MEDIO	NEGATIVO	POSITIVO	MEDIO	NEGATIVO	
	63%	35%	3%	46%	42%	12%	Colloca il suono nella regione delle altezze
1 IMITAZIONE SUONO	91%	6%	2%	88%	4%	8%	Riproduce la direzione
2 IMITAZIONE UN PROFILO	83%	14%	3%	81%	15%	4%	Colloca nella regione delle altezze
	85%	9%	5%	84%	11%	5%	Riproduce l'estensione temporale
3 IMITAZIONE DUE PROFILI	94%	4%	1%	62%	5%	33%	Riproduce la direzione
	71%	24%	6%	43%	24%	33%	Colloca nella regione delle altezze
	71%	26%	3%	48%	21%	31%	Riproduce l'estensione temporale
4 UN PROFILO IN DUE	65%	31%	4%	65%	30%	5%	Mantiene la continuità
	51%	38%	11%	52%	39%	9%	Mantiene la velocità
5 UNISONO IN DUE	84%	0%	18%	83%	0%	17%	Mantiene l'altezza
	65%	31%	4%	50%	39%	11%	Coordina inizio/fine
6 DUE BAMBINI ((UN PROFILO))	64%	30%	6%	69%	27%	4%	
7 DUE BAMBINI (DUE PROFILI)	25%	50%	25%	0%	100%	0%	
8 DUE BAMBINI (UNISONO)	90%	10%	0%	83%	17%	0%	
9 IMITAZIONE SUONO (DUE MANI)	88%	14%	0%	100%	0%	0%	
10 IMITAZIONE PROFILO (DUE MANI)	86%	14%	0%	100%	0%	0%	
11 IMITAZIONE DUE PROFILI (DUE MANI)	71%	24%	6%	75%	25%	0%	
12 PARALLELISMO UNISONO (DUE MANI)	90%	10%	0%	80%	20%	0%	
13 PARALLELISMO PROFILO (DUE MANI)	94%	6%	0%	100%	0%	0%	
14 PARALLELISMO 2 SUONI (DUE MANI)	63%	38%	0%				
15 PROFILI PER MOTO CONTRARIO (2 M.)	58%	42%	0%	50%	50%	0%	
16 PROFILO E SUONO (DUE MANI)	67%	33%	0%	50%	50%	0%	
17 TELECAMERE DISPETTOSE	67%	27%	7%				
<b>TOTALE</b>	<b>73%</b>	<b>22%</b>	<b>5%</b>	<b>66%</b>	<b>24%</b>	<b>10%</b>	

**Tabella 6.** Risultati totali in percentuale per ogni attività.

Abbiamo potuto rilevare che concentrandosi solo sul fenomeno sonoro, senza il contributo delle informazioni visive, un rapido e buon livello di risposta da parte dei bambini non è stato compromesso. Il numero dei successi resta considerevole, come si può vedere dai dati riportati nelle relative tabelle.



**Tabella 5. A e B (sopra e sotto):** Sintesi dei risultati per il totale delle attività, con e senza visione.



**Tabella 5. C** Sintesi dei risultati generale.

### 3. OSSERVAZIONI CONCLUSIVE

#### 3.1 EGGS e utilizzatori

Il sistema è stato testato per entrambe le categorie di utizzatori: per quanto concerne a) esigenze, modalità d'uso, strategie e funzionalità didattiche/educative, da parte di insegnanti/educatori e per quanto riguarda b) gli utizzatori "diretti", ossia gli studenti, nella presente ricerca con mirato riferimento alla produzione controllata e ai meccanismi di percezione del suono che varia per altezza, sia con modalità discrete che continue.

Pensando ai docenti/educatori, abbiamo testato il sistema dalle esigenze di *setting* (caratteristiche dello spazio di lavoro e attrezzatura necessaria), a quelle dell'interfaccia, configurata in modo chiaro ed essenziale anche per un operatore non esperto in materia, comprensibile e flessibile nel consentire di modificare/adeguare le varie caratteristiche del suono prodotto in rapporto al gesto/movimento e all'utilizzo da parte di uno o due utizzatori diretti che operano in forma coordinata e/o simultanea.

#### 3.2 EGGS, risultati e osservazioni fuori protocollo

Nella presente ricerca EGGS è stato utilizzato per operare sui processi di ascolto e produzione sonori con suo-

ni puri, scelti proprio per la chiarezza e non ambiguità dell'informazione percettiva, ai registri utilizzati. EGGS ha dimostrato di consentire un utile lavoro di esercizio/produzione sonora strettamente controllato attraverso i processi d'ascolto. E' utile per operare consapevolmente con la rappresentazione dell'altezza sonora e le strutture musicali che da essa possono essere derivate poiché fornisce feedback immediato per gli aspetti necessari al controllo consapevole delle prestazioni musicali.

Fuori dal più stretto protocollo di rilevazione, in particolare grazie al confronto coi docenti coinvolti nel progetto, abbiamo potuto raccogliere osservazioni libere, spontanee che si sono rivelate molto utili e interessanti. Alcune di queste riguardavano l'uso di suoni puri, un tipo di suono per nulla naturale al nostro orecchio che sono risultati per alcuni piuttosto strani, perfino non gradevoli, ma ciò ha riguardato l'inizio dell'esperienza. Tuttavia, l'informazione in merito all'altezza si è confermata chiara e facile da individuare per la maggior parte dei partecipanti. Aspetto interessante, che non abbiamo espressamente indagato in questo contesto progettuale, è rappresentato dalla risposta emotionale degli studenti a questo strano genere di suoni. Possiamo qui limitarci a considerare che i suoni puri usati, dopo un breve processo di *abituazione*, non risultavano più così poco familiari.

Un altro aspetto interessante riguarda le prestazioni di studenti che, come riferito dai loro insegnanti, presentano abitualmente difficoltà di apprendimento o comportamento: spesso, a detta sempre dei loro stessi docenti, le loro *performance* si sono dimostrate inaspettatamente di alto livello, o addirittura migliori di quelle degli altri compagni di classe.

Le azioni di imitare e riprodurre un suono/profilo sonoro, coordinare la propria azione con quella del compagno o completare la sua, agendo sia con il supporto delle informazioni visive, sia facendo esclusivo affidamento sui processi d'ascolto, sono in generale risultate azioni facilmente realizzabili per la maggioranza dei bambini.

E' stato cruciale sviluppare un protocollo per fornire istruzioni e comunicare le consegne in modo essenziale, semplice, chiaro e "pulito" per gestire le attività in classe. EGGS opera efficacemente se i suoi utilizzatori usano e necessitano di poche parole.

Consideriamo quest'ultimo un risultato importante che pone in evidenza come l'insegnamento musicale, specie in forma operativa e a partire dai livelli di base, abbia bisogno di poche parole, molte meno di quelle che spesso si usano o si reputano indispensabili, per non interferire, disturbare, offuscare i processi di ascolto [16], [6], concentrazione, memoria [18], che presiedono e rendono possibile la prestazione musicale (di pensiero e d'azione). Quanto questo aspetto illumini la necessità di una didattica musicale capace di incentivare processi che avvengono quasi spontaneamente se non inquinati dalla fretta di spiegare, dal bisogno di istruire, dalla scarsa abitudine a rispettare i processi "fisiologici" nell'esperienza di musica.

Per verificare l'utilità di EGGS come strumento didattico e risorsa educativa, abbiamo necessariamente dovuto pensare alla didattica, alle strategie, ai modi, oltre che ai contenuti, del suo possibile utilizzo. Ogni strumento didattico costringe a questa riflessione, in ciò consiste la sua prima utilità: non tanto nel rendersi sostituto di qualche altro strumento, quanto a forzare felicemente e funzionalmente i nostri punti di vista verso approcci che ci erano meno usuali o agevoli. EGGS apre orizzonti interessanti e funzionali per la ricerca e le sue applicazioni nel campo dello sviluppo dei processi di ascolto e attraverso l'ascolto in classe. Ha dimostrato di poterlo fare in modo efficace e semplice per gli utilizzatori, senza escludere, ma integrandosi o addirittura abbinandosi ad altri ordini di esperienza sonora e musicale, cui magari il nostro orecchio è maggiormente abituato.

Aspetto fondamentale nelle esperienze condotte con EGGS, anche fuori dal contesto della presente sperimentazione, è l'evidente e fortissima relazione che immediatamente si instaura con la corporeità, la rappresentazione che ne abbiamo, in rapporto allo spazio e al tempo, attraverso il gesto, il movimento. Il sistema consente, richiedendo un bagaglio di azioni motorie elementari e accessibili a chiunque, di "manipolare" eventi sonori, dal più semplice a quelli più complessi, e porli in relazione tra loro in maniera sempre più ricca e musicale. I livelli di discriminazione dell'esperienza sonora che l'utente realizza in tempo reale possono salire per qualità e finezza fino a soglie molto sottili.

Possiamo ipotizzare, per il futuro, di pensare, ad esempio, a EGGS come a una risorsa per sviluppare specifici apprendimenti su aspetti dell'esperienza sonora utili a chi studia musica, o per offrire a tutti occasioni per visualizzare, sperimentandolo fisicamente, il proprio rapporto col mondo dei suoni, e sue caratteristiche e rappresentazioni. EGGS è felicemente biunivoco: agisce dal movimento al suono ma anche dal suono al movimento, chiudendo un circuito coordinativo e di controlli che può efficacemente ricadere tanto a livello psicomotorio che cognitivo, che strettamente musicale. La parte grafica di cui EGGS dispone, i cui potenziali a livello educativo musicale promettono sviluppi di grande interesse, non è stata oggetto della presente indagine, ma la possibilità di correlare simultaneamente gesto, percezione sonora e rappresentazione grafica è di sicure e ampie risonanze formative.

### 3.3 EGGS e la "nuova" immediatezza

Abbiamo osservato 1586 scolari muovere le mani nell'aria impugnando torce elettriche, per imitare o realizzare una consegna sonora [19]. EGGS è risultato di immediato utilizzo, le consegne, verbali e sonore, non ambigue per la maggioranza degli studenti. Usare EGGS è quindi facile? No, è qualcosa di diverso, è immediato, dove usare il concetto di immediatezza, quando l'utente è in realtà "distante" dal cuore della produzione sonora (non tocca la corda, non percuote la membrana, non mette in pressione l'aria ...), esige una rilettura.

Le nuove tecnologie, infatti, ci stanno abituando a sempre più sofisticati livelli di simulazione e di mediazione

del reale, al punto da consegnarci "analogati" [4] di realtà così manipolabili direttamente da risultare concreti. Suonare suoni puri, a tutto vantaggi dell'esercizio di precisa e sottile individuazione delle altezze usate è solo un esempio, tra i più semplici, che apre all'utilità di queste risorse multimodali per la percezione, l'intellegibilità del proprio prodotto sonoro, quindi al controllo consapevole e immediato della propria azione [15]. Senza la necessità del tipico lungo training strumentale per ottenere un simile controllo consapevole, questi sistemi possono offrire strumenti di sperimentazione precoce di queste competenze, da integrare, affiancare all'esperienza della musica, con risorse tradizionali. Quel che è notevole è che con queste risorse è possibile ritrovare comunque, anziché perdere, la potenza espresiva del gesto (cui EGGS si è legato fin da principio), correlabile di volta in volta con diverse caratteristiche dell'esito sonoro. In questo modo la possibilità di esplorare, dandovi corpo, l'esperienza del suono si amplia a dismisura. Ma ciò che va approfondito è lo spettro di possibilità che si sviluppa da una complessità di prestazione che sale mantenendosi tuttavia gestibile in forma immediata e quasi intuitiva. Una risorsa come EGGS consente di pensare da ulteriori e nuove prospettive alla complessità dell'esperienza musicale: l'interdipendenza dei molti e diversi aspetti che esigono una costante ricerca metodologica risuona, per altri versi, anche sull'approccio educativo, sull'orizzonte di senso, personale e sociale, che la musica esprime [5]. Le nuove tecnologie stanno dilatando il rapporto con la musica e attraverso la musica, contribuendo a rimodularne contesti di produzione e fruizione, pertanto di educazione. A conoscenza delle autrici, nello scenario ampio e articolato di ricerche e studi in materia<sup>3</sup> [10][17], il progetto EGGS presenta una propria originalità e contribuisce ad arricchire la riflessione.

#### 4. REFERENCES

- [1] Eric F. Clarke: *Ways of Listening. An ecological Approach to the Perception of musical Meaning*, Oxford University Press, New York, 2005.
- [2] C. Cornoldi: *Metacognizione e apprendimento*, Il Mulino, Bologna, 1995.
- [3] M. Critchley and R.A. Henson: *Music and Brain. Studies in the Neurology of Music*, Heinemann Medical B., London, 1987.
- [4] E. Damiano: *L'azione didattica. Per una teoria dell'insegnamento*, Armando, Roma, 1993.
- [5] C. Fedrigo and R. Calabretto: *Introduzione alle problematiche dell'ascolto musicale*, Pordenone, Libreria al Segno, 1998.
- [6] E. Bruce Goldstein: *Sensation and Perception*, Wadsworth, Belmont, California, 2010.

<sup>3</sup>I contributi cui si fa riferimento aiutano a farsi un'idea di quanto ampio e diversificato possa essere il campo di studi sulle applicazioni educative musicali delle nuove risorse tecnologiche.

- [7] K. Lewin (ed. by D. Cartwright): *Field theory in social science: selected theoretical papers*, Harpers, Oxford, England, 1951.
- [8] K. Lewin: *Il bambino nell'ambiente sociale*, La Nuova Italia, Firenze, 1976.
- [9] K. Lewin: *La teoria, la ricerca, l'intervento*, Il Mulino, Bologna, 2005.
- [10] Lee CHENG (a cura di): "Software Development and Applications for Music". *Postgraduate Research Students Seminar*, HKIEd, 13/05/2011, in: [www.ied.edu.hk/](http://www.ied.edu.hk/) Settembre 2012.
- [11] M. Montessori (a cura di Lamparelli C.): *Educare alla libertà*, Mondadori, Milano, 2008.
- [12] D. Parigi e A. Pettini: *I Metodi di Ovidio Decroly e Maria Montessori*, O.D.C.U., Bologna, 1954.
- [13] M. Piattelli Palmarini, *Le scienze cognitive classiche: un panorama*, Einaudi, Torino, 2008.
- [14] P. Polotti and M. Goina: "EGGS in Action", *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME 2011), Oslo, Norway, 2011.
- [15] L.D. Rosenblum: *See what I'm saying. The extraordinary powers of our five senses* (it. *Lo straordinario potere dei nostri sensi*, Bollati Borighieri, Torino, 2011), 2010.
- [16] R. Murray Schafer: *The Soundscape. Our Sonic Environment and the Tuning of the World*, Destiny Books, Rochester, Vermont, 1994.
- [17] M. Sedláček (ed. by): *The multimedia Technologies Applications In Music Education*, Masaryk University Faculty of Education, Brno, 2010.
- [18] John A. Sloboda: *The Musical Mind. The Cognitive Psychology of Music*, Oxford University Press, Oxford, 1985.
- [19] Esempi video del lavoro con gli scolari durante la fase di sperimentazione in classe e approfondimenti di ordine metodologico e bibliografico nel sito: <http://www.cristinafedrigo.it/progetti.html>

# ARIADNE: UN SISTEMA PER L'ESECUZIONE MUSICALE SU TABLET

Julian Scordato

Conservatorio di Venezia

info@julianscordato.com

## ABSTRACT

Nel seguente testo vengono presentati i percorsi e i risultati di una ricerca tecnologica-musicale rivolta all'utilizzo del tablet come interfaccia gestuale per la generazione, la riproduzione e il controllo in tempo reale della musica elettronica, con l'intento specifico di valorizzare il contributo del gesto e della sua scrittura, quale mezzo per l'intendimento dell'organizzazione sonora e musicale, attraverso il diretto coinvolgimento dell'interprete informatico [14]. In questo caso in mancanza di una conoscenza a priori della corrispondenza fra gesto e risultato percepito dall'ascoltatore [10], come invece avviene con gli strumenti acustici, l'obiettivo principale posto da questo sistema è quello di fornire sia all'interprete di musica informatica sia all'ascoltatore un approccio alla materia sonora simil-strumentale ed espressivo [4] mediante l'impiego di relazioni semplici e intelligibili di causa-effetto, utili alla memorizzazione delle differenti tipologie di gesti ed eventi sonori correlati [15].

Per la parte performativa si è scelto di esplorare lo strumento del tablet in quanto soluzione economica ed ampiamente diffusa dotata di schermo multitouch e sensore di accelerazione, ritenuti interessanti dal punto di vista delle possibilità e delle varietà gestuali [13].

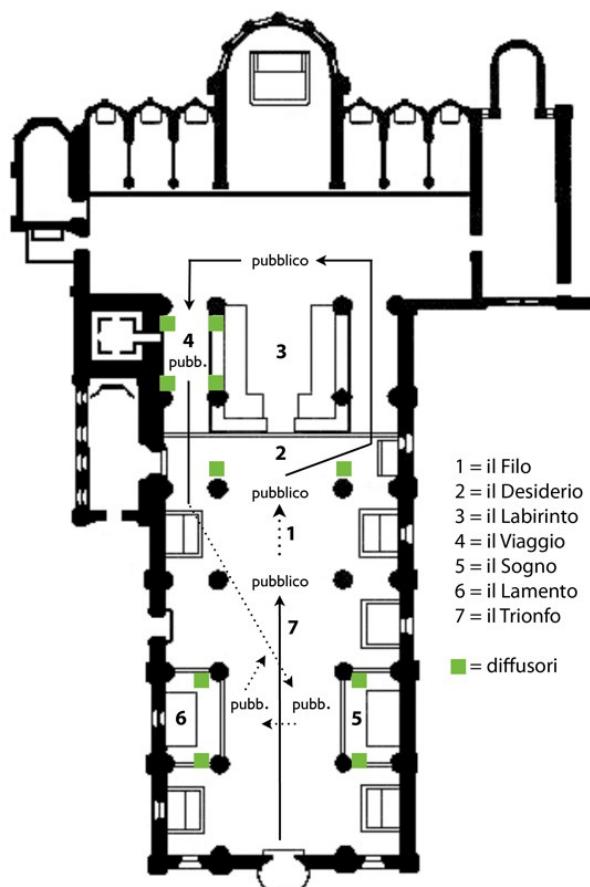
## 1. CONTESTO ARTISTICO

Il lavoro in questione è stato in larga parte progettato in occasione della prima rappresentazione assoluta de *La signora del labirinto* (2012), opera in 7 quadri di Victor Nebbiolo di Castri, Davide Gagliardi, Alvise Zambon, Julian Scordato, Marco Marinoni, Stefano Alessandretti e Giovanni Sparano, avvenuta il 5 giugno 2012 presso la Basilica dei Frari a Venezia.

La drammaturgia, curata dal regista Marco Bellussi, prevedeva la suddivisione del mito classico di Arianna in sette diverse sezioni – il Filo, il Desiderio, il Labirinto, il Viaggio, il Sogno, il Lamento, il Trionfo – ciascuna delle quali veniva messa in scena in un luogo differente della basilica ed elaborata dal punto di vista musi-

*Copyright: © 2012 Scordato 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.*

cale da uno dei sette compositori. Mentre sei tra questi momenti consideravano nell'organico la presenza di gruppi vocali e strumentali, il tema del viaggio invece, elaborato per la parte musicale dall'autore di questo articolo, è stato pensato per elettronica sola. Si è così posta la questione di come tale parte potesse essere integrata organicamente con il resto dell'opera, in cui il contributo scenico dei musicisti, variamente disposti nello spazio della basilica, avrebbe avuto un ruolo rilevante. La decisione presa comportava che la parte elettronica fosse 'suonata' con il tablet da un esecutore al centro di uno spazio di ascolto quadrifonico (cfr. Figura 1), all'interno del quale il pubblico sarebbe stato condotto da un attore.



**Figura 1.** Schema generale della prima dell'opera *La signora del labirinto*.

## 2. HARDWARE E SOFTWARE

Nella Tabella 1 vengono elencate le principali attrezzature (HW) e i software (SW) usati per l'implementazione del sistema Ariadne.

La lista ha funzione di documentazione ed è da ritenersi puramente esemplificativa anche se, ai fini della descrizione, nei paragrafi successivi viene fatto riferimento esplicito ad essa. Allo stesso tempo vengono indicate soluzioni alternative e suggerimenti che si rivolgono a possibili sviluppi futuri del sistema, indipendenti dalle apparecchiature specificate.

TIPO	MODELLO	FUNZIONE
HW / SW	Prestigio PMP5080B 8", 1 GHz, 512 MB RAM / Android 4.0.3, TouchOSC 1.3	Interfaccia gestuale per l'esecuzione e superficie di controllo (OSC)
HW / SW	Apple iMac 21,5", dual core 3 GHz, 4 GB RAM / OS X 10.7.3, Max/MSP 6.0	DSP e gestione dei dati di controllo
HW	TC Electronics Kon- nekt 24D	Interfaccia audio
HW	D-Link DSL-2740R	Comunicazione dei dispositivi over IP
HW	4x Genelec 8030A	Monitor da studio

**Tabella 1.** Attrezzature impiegate dall'autore.

## 3. DISPOSITIVI DI INPUT E DESCRIZIONE DEL GESTO

Durante l'esecuzione il gesto del performer viene captato da due dispositivi del tablet: lo schermo multitouch e l'accelerometro.

I dati gestuali, codificati in messaggi Open Sound Control (OSC) e trasportati in rete WLAN via protocollo UDP, vengono trasmessi ad un computer per il controllo in tempo reale della generazione/riproduzione e dell'elaborazione degli eventi sonori.

EVENTO	MESSAGGIO TRASMESSO
Pressione pad <i>n</i>	/2/pushn 1.
Rilascio pad <i>n</i>	/2/pushn 0.
Variazione di inclinazione	/accxyz <i>x y z</i>

**Tabella 2.** Implementazione OSC.

### 3.1 Schermo multitouch

Lo schermo del tablet serve all'utente da superficie di controllo per la generazione di suoni percussivi mediante 7 *drum pads* posti lateralmente in modo da favorire l'usabilità. Il software TouchOSC [5] [6] offre un'inter-

faccia utile a tale scopo (cfr. Figura 2) e dispone inoltre di altre tipologie di controlli gestuali. L'esecutore può percuotere i *pads* con il pollice e l'indice (ed eventualmente il medio) di entrambe le mani usando una diteggiatura mobile e sostenendo il tablet con le altre dita e il palmo della mano. In alternativa è possibile usare una diteggiatura invariabile premendo ciascun *pad* con un dito diverso e adoperando i pollici per il sostegno. Il tablet a disposizione dell'autore riconosce contemporaneamente fino a 5 punti di contatto, reputati più che sufficienti per il tipo di esecuzione proposta. Per garantire una stretta corrispondenza fra il tocco delle dita sullo schermo e i suoni prodotti dal sistema, tenendo conto anche dell'apporto visivo-scenico, si è ipotizzato che gli eventi sonori dovessero essere brevi (< 1000 ms).



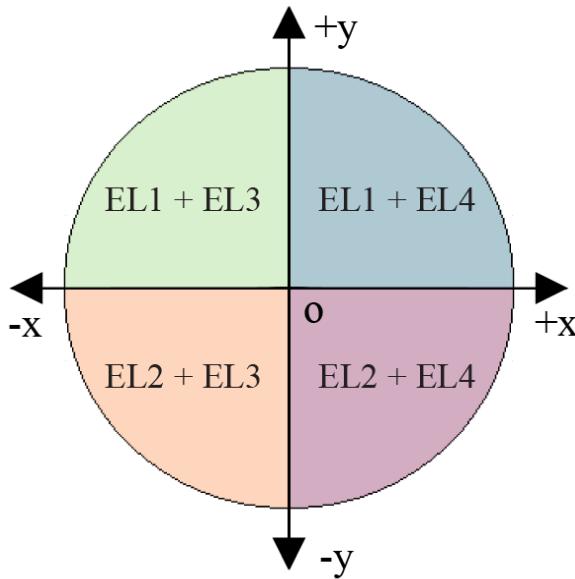
**Figura 2.** Schermata del software TouchOSC con indicazione dei *drum pads* usati per l'esecuzione.

Un aspetto critico del sistema è la latenza nella trasmissione del segnale [7]. Nel caso considerato è stato misurato un valore di latenza complessiva (input gestuale/output sonoro) compreso fra 105 e 130 ms, ritenuto accettabile solo per alcune tipologie di esecuzione e notazione musicale [8]; i valori sono stati calcolati in base al ritardo fra il suono prodotto concretamente dal tocco della superficie di controllo ed un evento sonoro puntuale generato dal software Max/MSP alla ricezione del messaggio OSC corrispondente (cfr. Tabella 2), con gli strumenti disponibili (cfr. §2) e un microfono per la registrazione. Per ottimizzare il tempo di trasmissione era previsto che il computer venisse collegato al router con una connessione LAN cablata e l'impostazione della dimensione del buffer audio di Max/MSP (*I/O Vector Size*) fosse pari a 64 campioni, con una frequenza di campionamento di 44,1 KHz. Un ulteriore miglioramento delle prestazioni complessive può essere ottenuto con l'aggiornamento del sistema Android [2] o impiegando un tablet con sistema iOS, come viene rilevato anche nei test di [9] e in [3].

### 3.2 Accelerometro

L'inclinazione della posizione del tablet rispetto agli assi cartesiani  $x$  e  $y$  viene utilizzata per il *processing* dei suoni generati con i *drum pads* e per la spazializzazione mentre i movimenti nell'asse  $z$  servono al riconoscimento di un gesto specifico – la rotazione del tablet in posizione orizzontale capovolta – utile al controllo di un sistema di eventi e di parametri preimpostati (*presets*).

Il sistema di elaborazione comprende quattro diverse unità (EL1, EL2, EL3, EL4) controllate rispettivamente dall'inclinazione nei semiassi  $+y$ ,  $-y$ ,  $-x$ ,  $+x$  (cfr. Figura 3), quindi direttamente dal gesto dell'esecutore, oltre che in maniera indiretta dall'attivazione da parte dell'esecutore dell'automazione definita in fase preliminare dal sistema di eventi.



**Figura 3.** Interazione fra unità di elaborazione in relazione ai quadranti.

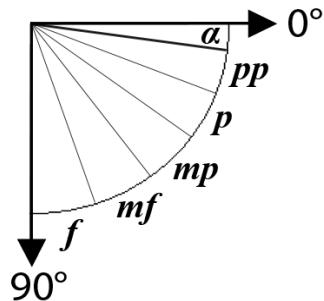
Le singole unità, inattive in posizione del tablet orizzontale ( $0^\circ$ ) ed entro una soglia di errore pari ad un angolo  $\alpha$ , vengono attivate e restano in funzionamento con un angolo di inclinazione  $\gamma$  tale che

$$\alpha < \gamma < 180^\circ \quad (1)$$

per consentire al sistema di disattivarsi in posizione capovolta ( $180^\circ$ ) durante la variazione dei *preset*.

Oltre a definire la tipologia di trattamento dell'audio e la relativa collocazione spaziale, l'angolo di inclinazione determina allo stesso tempo la dinamica sonora in uscita da ciascuna unità di elaborazione (cfr. Figura 4). Per una questione pratica di esecuzione è opportuno che in condizioni normali l'angolo  $\gamma$  sia inferiore o uguale a  $90^\circ$  e in tale ambito ci sia una proporzionalità diretta fra ampiezza dell'angolo e dinamica. Nella porzione fra  $90^\circ$  e  $180^\circ$  occorre invece che ci sia una proporzionalità inversa, per favorire la disattivazione del sistema. Relati-

vamente all'esecuzione di quest'ultima azione, a seconda della diteggiatura usata per la generazione dei suoni percussivi, il musicista in un caso ruota il tablet verso di sé, nell'altro necessariamente in senso contrario rivolgendo il dorso delle mani verso il basso.



**Figura 4.** Rapporto fra dinamica e angolo di inclinazione del tablet.

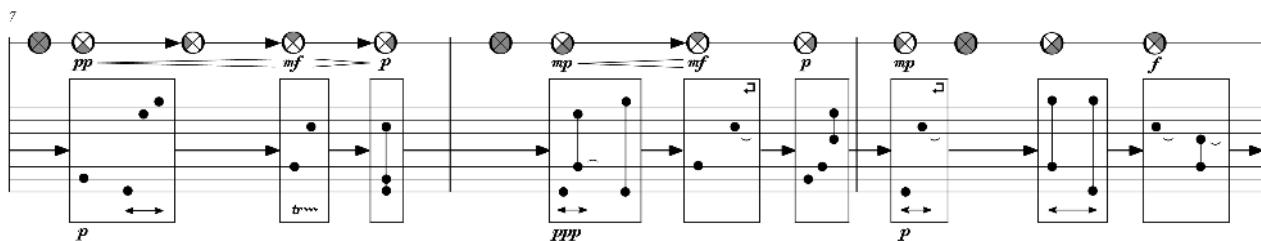
Nelle medesime condizioni di misurazione della risposta del multitouch, la latenza totale determinata dal sistema con l'input dell'accelerometro varia fra 40 e 75 ms. La misura è stata effettuata calcolando la differenza di tempo tra il suono concreto causato dalla collisione del tablet con una superficie orizzontale in seguito alla rotazione lungo un asse e quello generato da Max/MSP alla ricezione del messaggio OSC relativo alla posizione orizzontale. Va tenuto in considerazione che le prestazioni ottenute potrebbero decrementare nell'eventualità in cui sia presente un rumore elevato; in tal caso è richiesta l'implementazione di un algoritmo di *smoothing* che comporta l'aggiunta di un ulteriore ritardo nella trasmissione del segnale.

## 4. NOTAZIONE

Fatta eccezione per la pratica improvvisativa, in questo sistema la scrittura del gesto si pone come necessaria per la strutturazione e la memorizzazione di un contenuto linguistico-musicale organizzato, soprattutto in considerazione della difficile relazione che intercorre fra gesto e suono. Nel tentativo di aderire alla citata concezione simil-strumentale dell'elettronica, la notazione viene impiegata con funzionalità realizzativa, attraverso l'esecuzione dal vivo, oltre che progettuale [16].

Di seguito viene proposta una possibile soluzione al problema semiografico, legato alla specificità del carattere 'strumentale'.

Gli eventi del *touchscreen* possono essere scritti in modo analogo a note poste su due diversi trigrammi distinguendo i suoni prodotti dalla mano sinistra (trigramma inferiore) e quelli prodotti dalla mano destra (trigramma superiore). In condizioni di elevata latenza è inopportuno ricorrere ad una precisa misurazione delle durate dei suoni, lasciando piuttosto un maggior grado di libertà nel flusso temporale.



**Figura 5.** Frammento della partitura esecutiva per l'opera *La signora del labirinto* [12].

La simbologia relativa alla posizione del tablet e al verso di inclinazione (cfr. Figura 6), corrispondente alla localizzazione spaziale del suono elaborato, può essere notata in un ulteriore rigo musicale costituito da una sola linea. L'angolazione invece viene indicata con i segni dinamici tradizionali (cfr. Figure 4 e 5).

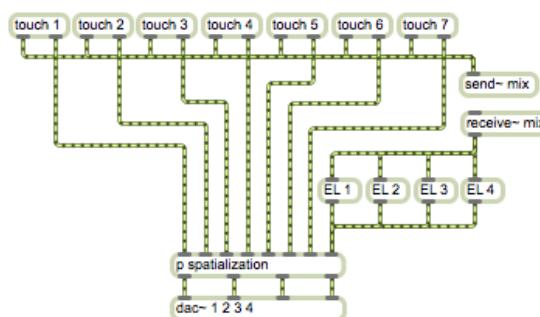
- ⊗ = posizione del tablet orizzontale
- ⊗ = posizione del tablet sottosopra
- ⊗ = inclinazione in avanti
- ⊗ = inclinazione indietro
- ⊗ = inclinazione a sinistra
- ⊗ = inclinazione a destra

**Figura 6.** Indicazioni per l'esecuzione.

## 5. IMPLEMENTAZIONE IN MAX/MSP

Complessivamente il DSP comprende 7 unità di generazione/riproduzione sonora tramite multitouch, 4 unità di elaborazione e 1 unità di spazializzazione.

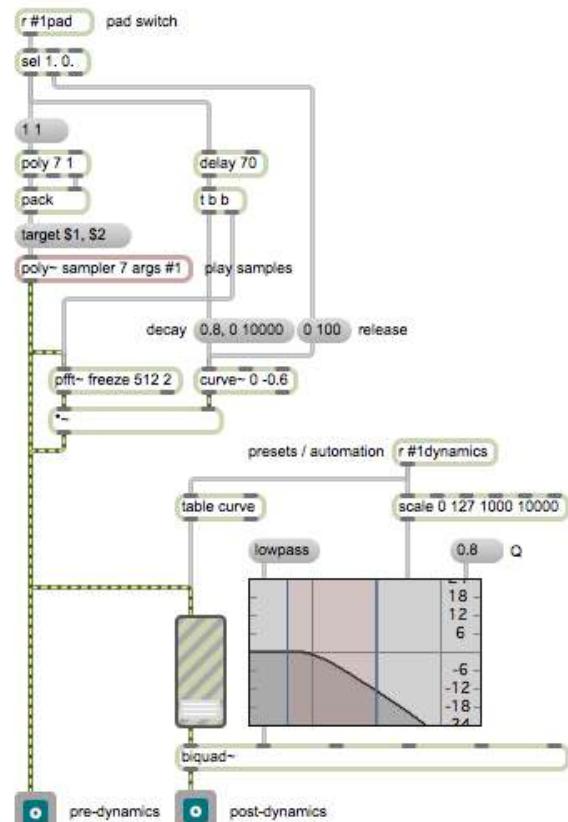
Le unità *touch* possiedono un'uscita post-elaborazione dinamica (cfr. Figure 7 e 8) che viene inviata direttamente al sistema di spazializzazione e un'uscita pre-dinamica diretta al sistema di elaborazione (*EL*).



**Figura 7.** Instradamento dei segnali audio.

### 5.1 Unità di generazione

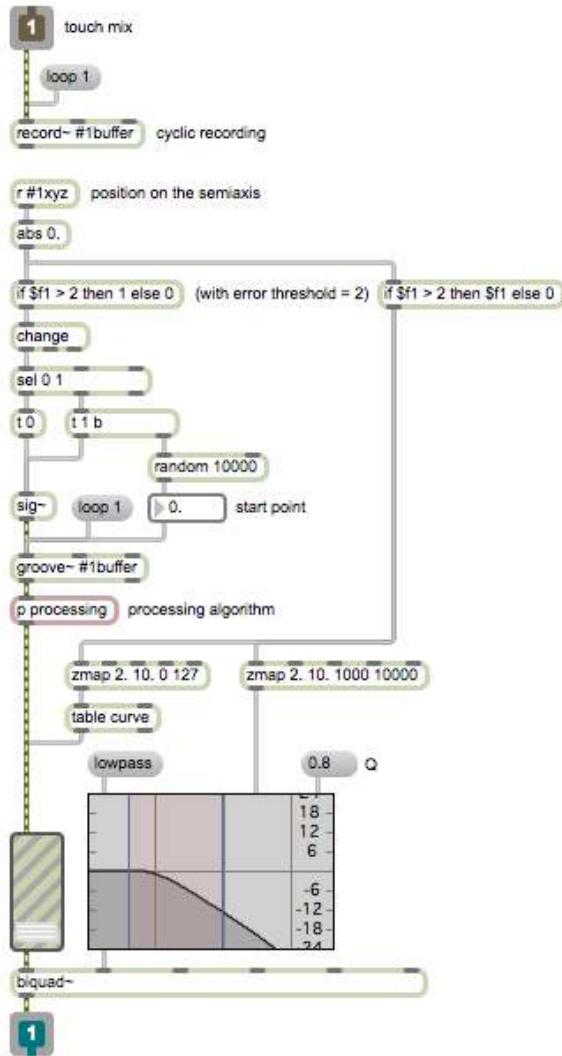
La scelta dell'algoritmo di generazione/riproduzione va aldilà della descrizione del sistema, essendo estremamente caratterizzante dal punto di vista compositivo. A titolo di esempio viene proposto un campionatore polifonico con effetto *freeze* per simulazione della risonanza e controllo della dinamica (cfr. Figura 8).



**Figura 8.** Esempio di unità *touch*.

### 5.2 Unità di elaborazione

Nel caso descritto in questo paragrafo, il sistema di elaborazione è composto da 4 istanze di un unico processo (cfr. Figura 9). In alternativa è possibile utilizzare algoritmi differenti tenendo presente che l'uscita audio di ciascuno di essi si colloca in una zona precisa dello spazio di ascolto, a seconda della posizione del tablet.

**Figura 9.** Unità *EL*.

### 5.3 Unità di spazializzazione

La localizzazione del segnale audio ricevuto dalle unità *EL* è determinata dai valori di inclinazione ricevuti dal tablet tramite l'oggetto *udpreceive*; ciò comporta che il movimento della sorgente sia strettamente vincolato al gesto dell'interprete che, trovandosi al centro del sistema di diffusione, favorisce l'identificazione di tale legame.

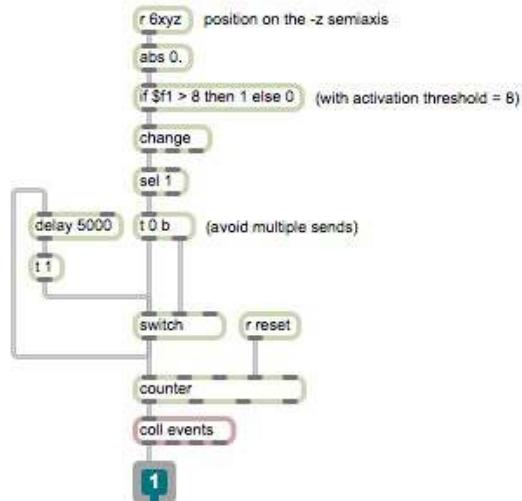
Per l'implementazione dell'unità di spazializzazione è possibile usare gli oggetti *ambipanning~* (configurato con 8 input e 4 output) e *ambimonitor* [1]; la posizione di ciascun input (*n*) è definita dalla seguente sintassi:

*xyz n x y 0. n*

Nell'ingresso 8 *x* e *y* corrispondono alle coordinate cartesiane relative all'inclinazione del tablet riscalate fra i valori di -1 e +1. Gli altri ingressi, provenienti dalle unità *touch*, possono essere spazializzati come sorgenti sonore statiche, simulando la disposizione di un ensemble strumentale.

### 5.4 Preset e automazione di eventi

Con il capovolgimento del tablet l'esecutore può controllare il procedere di una 'partitura' di eventi (ad esempio variazione di parametri, riproduzione di campioni e ulteriori elaborazioni) implementabili come messaggi comprensibili all'utente inseriti nell'oggetto *coll* e successivamente codificati nella sintassi appropriata.

**Figura 10.** Algoritmo per il riconoscimento del gesto e il controllo degli eventi automatici.

## 6. CONCLUSIONI

La presenza fisica del musicista informatico sulla scena, ammessa dalla strutturazione di un linguaggio gestuale ancora prima che musicale, si pone come tramite e immagine stessa della percezione acustica [11] e non come soggetto (vittima) della rappresentazione in una musica che, priva di una fonte strumentale concreta di emissione sonora, non consente di essere rappresentata. In questo senso il sistema *Ariadne* nasce dalla ricerca del corpo del suono, dove la gestualità descritta attraverso la notazione musicale contribuisce direttamente alla definizione sonora. In altri termini il musicista, disponendo il proprio corpo in una struttura, è parte integrante di quello stesso processo di elaborazione che porta il suono a scriversi: in un ascolto tendenzialmente acusmatico la sua presenza sottolinea l'importanza e l'urgenza della domanda di come le cose si odono, rendendo l'ascoltatore partecipe al processo.

Il tablet, adoperato tenendo in considerazione le proprie limitazioni in termini tecnologici tra cui il problema della latenza e l'impossibilità di definire la dinamica mediante il gesto di pressione, ha il vantaggio di consentire una gestualità differente da quella di altri controller e degli strumenti tradizionali e, senza tentare di porsi come loro imitazione ma piuttosto rivolgendosi al gesto performativo da inventare in relazione al gesto sonoro, può cercare di dare un nuovo e diverso contributo al linguaggio musicale.

## 7. RIFERIMENTI

- [1] Ambisonics Externals for MaxMSP [luglio 2012], <http://www.icst.net/research/downloads/ambisonics-externals-for-maxmsp/>
- [2] Android 4.1 for Developers [ottobre 2012], <http://developer.android.com/about/versions/jelly-bean.html>
- [3] Android Lag, Attributed To Core System Framework, Here To Stay [ottobre 2012], <http://appadvice.com/appnn/2011/12/android-lag-attributed-to-core-system-framework-here-to-stay>
- [4] S. Benford: "Performing Musical Interaction: Lessons from the Study of Extended Theatrical Performances", *Computer Music Journal*, 34:4, pp. 49-61, 2010.
- [5] G. Dubus, K. F. Hansen e R. Bresin: "An overview of sound and music applications for Android available on the market", *Proceedings of the 9th Sound and Music Computing Conference*, pp. 541-546, 2012.
- [6] Hexler.net | TouchOSC for Android [luglio 2012], <http://hexler.net/software/touchosc-android>
- [7] N. Lago e F. Kon: "The quest for low latency", *Proceedings of the International Computer Music Conference*, pp. 33-36, 2004.
- [8] T. Maki-Patola e P. Hamalainen: "Latency tolerance for gesture controlled continuous sound instrument without tactile feedback", *Proceedings of the International Conference on Digital Audio Effects*, pp. 11-16, 2004.
- [9] M. Montag, S. Sullivan, S. Dickey e C. Leider: "A Low-Cost, Low-Latency Multi-Touch Table with Haptic Feedback for Musical Applications", *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 8-13, 2011.
- [10] G. Paine: "Gesture and Morphology in Laptop Music Performance", *The Oxford Handbook of Computer Music*, pp. 214-232, Oxford University Press, 2009.
- [11] J. C. Schacher: "The Body in Electronic Music Performance", *Proceedings of the 9th Sound and Music Computing Conference*, pp. 194-200, 2012.
- [12] Julian Scordato - Note [ottobre 2012] <http://www.julianscordato.com/note.php?n=ariadne>
- [13] A. Tanaka: "Mapping Out Instruments, Affordances, and Mobiles", *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 88-93, 2010.
- [14] A. Tanaka: "Musical Performance Practice on Sensor-based Instruments", *Trends in Gestural Control of Music*, pp. 389-405, IRCAM, Parigi, 2000
- [15] A. Tanaka, A. Altavilla e N. Spowage: "Gestural Music Affordances", *Proceedings of the 9th Sound and Music Computing Conference*, pp. 318-325, 2012.
- [16] A. Valle: *La notazione musicale contemporanea*, pp. 122-127, EDT, Torino, 2002.

# The EyesWeb human movement analysis library

**Stefano Piana**

University of Genoa, Italy  
steto84@infomus.org

**Antonio Camurri**

University of Genoa, Italy  
antonio.camurri@unige.it

**Maurizio Mancini**

University of Genoa, Italy  
maurizio@infomus.org

## ABSTRACT

A framework and software system for real-time tracking and analysis of human motion. The objective is to design and create a multimodal interactive system for the automated analysis of human movements. The system is based on the EyesWeb XMI open software platform and on Kinect depth sensors.

## 1. INTRODUCTION

This paper will introduce a software platform that offers various tools to perform human tracking and movement analysis, in particular we will illustrate software modules for user tracking and movements analysis using the Microsoft Kinect sensor.

In Section 2 we briefly introduce the EyesWeb XMI platform a visual-based development platform.

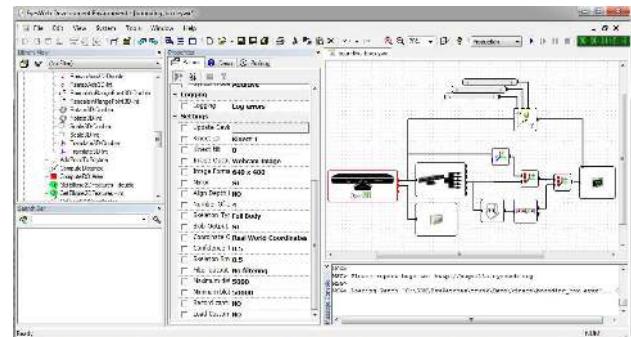
In Section 3 we describe newly developed software modules that extend the EyesWeb library. The new modules perform human tracking and human motion analysis. In Section 3.1 we present software modules that interfacing to Microsoft Kinect can extract information about the user position and her movements. Finally in section 3.2 we present some modules that starting from the data captured by the Kinect sensor compute low and mid-level motion features.

## 2. EYESWEB XMI

The EyesWeb XMI platform is a modular system that allows both expert and non-expert users to create multimodal installations in a visual way [1]. The platform provides modules, called *blocks*, that can be assembled intuitively (i.e., by operating only with mouse) to create programs, called *patches*, that exploit system's resources such as multimodal files, webcams, sound cards, multiple displays and so on.

The EyesWeb platform consists of a number of integrated hardware and software modules which can be easily interconnected and extended. The EyesWeb software consists of a development environments and a set of libraries of reusable software components which can be assembled by the user in a visual language to build patches as in common computer music languages inspired to analog synthe-

*Copyright: ©2012 Stefano Piana 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.** The eyesWeb XMI GUI, it includes, starting from the left to the right : blocks catalogs and search tools, property dialog, patch overview and console dialog

sizers. The aspect of the GUI and an example of a simple EyesWeb patch is shown in Figure 1. The patch shows a typical application on movement analysis based on the Kinect sensor.

A patch can be created choosing blocks from the blocks catalog (shown on the left in Figure 1) and dragging it on the designing area (on the right in Figure 1) the patch is then completed linking the blocks together.

The behavior of the patch can be modified acting on the various parameters that characterize each block of the patch; the tuning of the parameters can be performed through the use of a menu of the GUI denominated property window

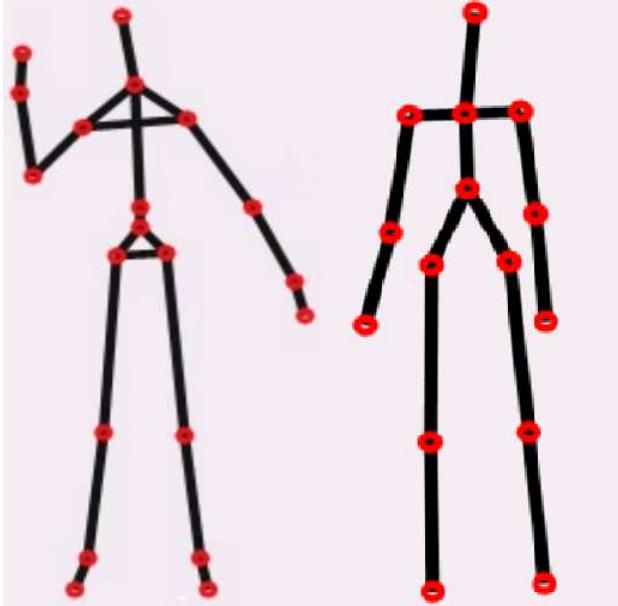
We now describe the implementation of a subset of the blocks belonging to the EyesWeb environment catalog, consisting in 2 classes of new EyesWeb XMI blocks: the first one (see Section 3.1) consists in blocks for multiple user detection and tracking; the second one (see Section 3.2) consists in blocks for the computation of users' movement features.

## 3. USER TRACKING AND MOVEMENT ANALYSIS

We now introduce some of the modules included in the EyesWeb development environment developed specifically for user tracking and movement analysis

### 3.1 User Tracking

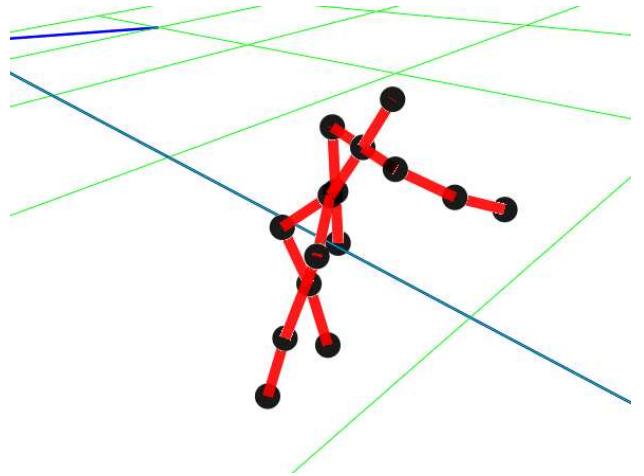
Multiple users detection and tracking is performed thanks to different EyesWeb XMI software modules (blocks). To communicate with the Kinect sensor, EyesWeb XMI supports both the OpenNi [2] framework and Microsoft Kinect



**Figure 2.** 2D coordinates of the tracked joints produced by the Microsoft SDK (on the left) and by the OpenNI API (on the right).

SDK [3]. The two APIs support the streaming of both color images and *depthmaps* captured by the Kinects optical sensors. A *depthmap* is a gray-scale image where the color intensity represents the distance from the sensor measured in millimeters. Both OpenNI and Microsoft Kinect SDK support user segmentation and tracking by providing 2D and 3D measurements of multiple users joints: the two API are similar in term of speed and real-time performances (the Microsoft API is slightly less influenced by occlusions), the sets of joints tracked by the two APIs are similar but the Microsoft one can track a bigger number of joints (see Figure 2), furthermore the Microsoft SDK enables to capture also an audio stream using the Kinects microphone array and offers face tracking functionalities: it can detect up to six users for each sensor but can only track the body parts of up to two users at the same time on each sensor. The OpenNI API doesn't support audio streaming nor face tracking, but can theoretically track an unlimited number of users (limited only by the field of view of the sensor).

Using the OpenNI Framework, the EyesWeb environment supports the automatic calibration of the user tracking system, and provide functionalities to save configuration files. This feature allows to avoid the tuning phase of the Kinect, which consists of the automatic calibration phase requiring from ten to fifteen seconds (during the tuning phase the tracking measurements are less precise). To avoid this tuning process, the system can save calibration files to reduce the time needed for calibration, consisting in this case only of the time to load the calibration file. OpenNI offers functionalities to save and load video recordings of the participants using .oni files. Oni is a file format which allow storing both video information from the Kinect color camera and depth information from the Kinect depth sensor. The recorded streams can then be played and analyzed in order to extract user tracking data. These func-



**Figure 3.** 3D visualization of the tracked user's skeleton tracked using the OpenNI API.

tions are automatically done by specific software modules in the Kinect EyesWeb Library. The support for Oni files simplifies data recording and off-line analysis.

The Microsoft SDK doesn't need any time for calibration, the tracking procedure is started as soon as a user is detected by the sensor, it doesn't provide tools for record and playback video streams but these functionalities can be performed using the EyesWeb software platform that offers a wide selection of modules to record and playback multimedia streams.

Different EyesWeb XMI modules have been developed to manage the communication between the EyesWeb environment and one or multiple Kinect sensors, and to manipulate data coming from those devices. Two different blocks, called Kinect Extractor OpenNI and Kinect Extractor SDK, were developed to interface with Kinect using the OpenNI or the Microsoft APIs; both of the two blocks provide input data from the Kinects sensors. Multiple instances of these blocks may be instantiated in a single application in order to use several Kinect devices at the same time; for each device, the outputs provided by the blocks are: a set of tracked users the image from the color camera or, alternatively, the image from the infrared camera (available only using OpenNI), an image representing the reconstructed depthmap, where the distance from the sensor is mapped to a gray-level in the image. The block developed to support the Microsoft SDK can also output information about face tracking and an audio stream. Both EyesWeb blocks can be configured through the setting of some parameters, including the following: parameters to identify the sensor to which the application has to interface in case of multiple connected devices; settings to control the device tilt angle; settings regarding webcam image and depthmap (e.g., resolution, alignment); settings regarding the performances of the tracking algorithms: e.g., the quantity of tracked users, the number of tracked joints for every user, the quantity of smoothing applied to the measurements, the type of coordinates computed.

The Extractor blocks output a list of detected users. In order to analyze every user's data (e.g., the coordinates of

the tracked joints), a block that manages the users list has been developed. This block selectively outputs every information regarding a subset of the users contained in a list. The block, called KinectDataManager manages the data received from both OpenNi and Microsoft SDK Extractor blocks. The number of users managed can be specified through a parameter. The block has a variable number of outputs that depends on the number of users that will be tracked simultaneously. At the beginning, when all outputs are free, the block orders the incoming recognized users according to a specified criteria (e.g., label, index, distance from a given point), and assigns each of them to an available output. When one output is assigned to a user, the output remains reserved for that user (the id is the way to uniquely identify the user) independently from the sorting criteria. Only when the user with the given id disappears, a new user may be assigned to the corresponding output. The assignment to an output is based on the user-specified sorting criteria; however, users already assigned to some outputs are excluded from the list of candidates. A reset command is available to remove the binding between user identifiers and outputs, and restore the block to its initial condition. For each tracked user, the block provides the following information:

- Label (i.e., its unique identifier);
- Status with respect to the system (it can be detected, tracking, calibrating, Out of sight);
- 2D coordinates of the points of the skeleton (2D coordinates are normalized in the range 0.0-1.0, (0.0, 0.0) being the top-left corner of the image and (1.0, 0.0) being the bottom-right corner);
- 3D coordinate of the points of the skeleton in normalized image coordinates or real word coordinates, according to the setting of the Extractor block;
- Blob2D, an image containing white pixels where the user is visible in the scene, and black pixels if the corresponding pixel is not occupied by the

In addition to the previous outputs, if the Microsoft SDK Extractor block is used the DataManager block can output the following further data:

- Face 3D coordinates containing a set of three dimensional points that identify position, orientation and the shape of the user's face
- Face 2D coordinates containing the same set of points as the three dimensional one but projected in a way that they can fit on the color image
- Face Animation units a set of indexes each of which represents a single parameter of the face's expression (e.g., the brows position, mouth openness and shape)

The EyesWeb environment provides a large number of modules to manipulate the captured data, to visualize them using two dimensional or three dimensional representations,

and to analyze data using a wide selection of different methods. The following session will describe some blocks that, starting from the data collected by the motion tracking modules, perform feature extraction and movement analysis.

### 3.2 Movement features extraction

Movement analysis is based on the *multilayered conceptual framework* developed by Camurri and colleagues [4] [5]. Such framework supports multimodal systems, and includes a *Physical Layer*, in which data captured by sensors (e.g., images) is analyzed and early processing is performed. The use of Kinect enables full-body (2D as well as 3D) user tracking so the *Physical Layer* will include both two and three dimensional measurements of coordinates of users joints.

Starting from the measurements performed at the *Physical Layer*, *low-level* movements features are extracted. The low-level features here considered, besides velocities and accelerations of singles or groups of points, are the following:

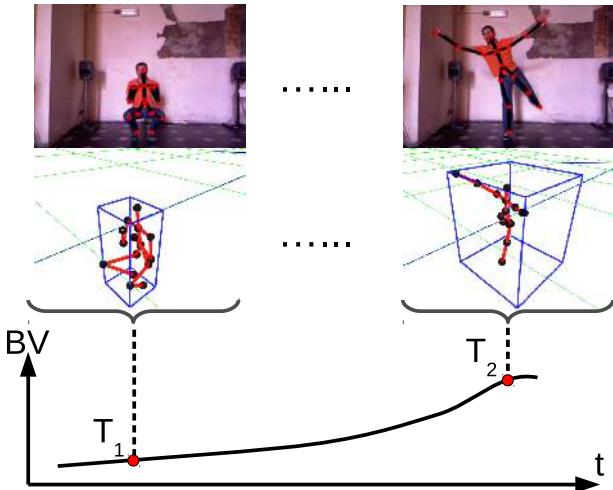
- Bounding Volume: the minimum parallelepiped that encloses each tracked user
- Kinetic Energy: It can be measured at different levels, from a single joint to an overall feature considering the whole body (see Section 3.2.1 ).
- Curvature index: it can be measured both on 2D and 3D trajectories of a single joint or a set of joints and gives clues on the type of movement performed by the user

The physical and low-level features are used to extract the following mid-level features:

- Repetitiveness Index: how much repetitive (therefore predictable) the movement performed by users;
- Impulsivity Index: how much impulsive is the movement performed by a user (see Section 3.2.2);
- Smoothness Index: an approximation of the fluidity of the movement;
- Contraction Index: the contraction/expansion of movement;
- Symmetry Index: how much movement is symmetric with respect to vertical or horizontal axes;

Many researchers on human movement, like Wallbott & Scherer [6], Gallaher [7], pointed out the importance of body motion's low-level features (i.e., speed, amplitude, energy and so on) in the process of recognizing emotions, social behavior, and expressiveness. Wallbott demonstrated in [8] that body activity, expansiveness and power are discriminating factors in communicating with other people. The choreographer Rudolf Laban elaborated in his Theory of Effort [9], a model of expressive movement.

We consider a classification of such features on two levels of abstraction: the lower level includes those features



**Figure 4.** Bounding Volume real-time computation: at time  $T_1$  the user has a contracted posture, thus the BV value is very low; at time  $T_2$  the user reaches an expanded posture, exhibiting a higher value for the BV.

directly describing the physical characteristics of movements, such as its speed, amplitude and so on; the higher level features can be described by models and algorithms based on the low-level features, for example the movement smoothness can be computed as the correlation between the movement velocity and curvature.

### 3.2.1 Low-level features

We now describe three low-level user's full-body movement features: the *Bounding Volume*, *Kinetic Energy* and *Curvature Index*. *Bounding Volume* and *Kinetic Energy* features can be considered a 3D extension of the two previously developed 2D low-level features *Contraction Index* and *Motion Index* (or *Quantity of Motion*), see [10] for details. *Curvature Index* characterizes a trajectory followed by a single joint or a group of joints in time. the curvature can be a clue on how the user performs certain movements and help to identify their characteristics.

- *Bounding Volume (BV)* - It is the normalized volume of the smallest parallelepiped enclosing the user's body. Figure 4 show an example of BV computation. The BV can be considered as an approximation of the user's degree of body "openness": for example, if the user stretches her arms outside or upside the bounding volume, and so the BV Index, increases.
- *Kinetic Energy (KE)* - It is computed from the speed of user's body segments, tracked by Kinect, and their percentage mass as referred by [11]. In particular the full-body kinetic energy  $KE$  is equal to:

$$E_{FB} = \frac{1}{2} \sum_{i=0}^n m_i v_i^2 \quad (1)$$

where  $m_i$  is the mass of the  $i$ -th user's body segment (e.g., head, right/left shoulder, right/left elbow

and so on) and  $v_i$  is the velocity of the  $i$ -th segment, computed as the difference of the position of the segment at the current Kinect frame and the position at the previous frame.

- *Curvature Index (CI)* - It is computed from a trajectory drawn by a joint or a set of joints. as stated by [12] the curvature is defined as:

$$CI_b = (\frac{\dot{r} \times \ddot{r}}{v^3}) \quad (2)$$

where  $\dot{r}$  is the velocity of the movement,  $\ddot{r}$  the acceleration,  $v$  is the module of the velocity  $\dot{r}$  and  $b$  is the binormal unit vector, normal to the plane which locally contains the trajectory.

all the indexes presented here including velocities and accelerations can be easily computed using EyesWeb blocks developed specifically to compute them each of the developed block has different parameters to fine-tune the outcome (e.g., to smooth out noise).

### 3.2.2 Mid-level feature: Impulsivity Index

*Impulsive* body movements are those performed quickly, with a high energy and by suddenly moving spine/l Limbs in a straight direction. We adopt this definition after reviewing some literature about human movement analysis and synthesis.

Wilson, Bobick e Cassell [13] studied *beat* gestures, those produced while accompanying speech to underline the goal-conveying words during conversations, like pitch accents do for speech. Wilson and colleagues found that stroke gestures have a lower number of execution phases compared to other conversational gestures, suggesting that their execution is shorter in time.

Laban theory, as described by Shapiro in [14], is a descriptive format originally created for body dance gestures but widely adopted also in Computer Science. Laban identifies four main qualities of movement: Flow (bound Vs. free), Weight (heavy Vs. light), Time (sustained Vs. quick), and Space (direct Vs. indirect). Following this taxonomy, impulsive gestures have quick Time and free Flow, that is, they are executed quickly with energy flowing through body in a consciously uncontrolled way.

Finally, Bishko [15] defines impulsive gestures as "an accent leading to decreasing intensity".

The measure we propose for the *Impulsivity Index* mid-level feature is a combination of the two low-level features Kinetic Energy ( $KE$ ) and Bounding Volume.  $KE$  is firstly used to identify the gesture duration  $dt$ : for example, using an adaptive threshold, when the  $KE$  becomes greater than the threshold we identify the gesture beginning time; when the  $KE$  goes below the threshold we identify the ending time. Then, if the  $KE$  is higher than a fixed energy threshold and the gesture length  $dt$  is lower than a fixed time threshold then *Impulsivity Index* is equal to the ratio between the variation of  $BV$  and the gesture length  $dt$ :

```
let time_threshold = 0.45 s;
let energy_threshold = 0.02;
```

```

if (KE ≥ energy_threshold) then
    evaluate the GestureTimeDuration dt;
    if (dt ≤ time_threshold) then
        ImpulsivityIndex = Δ BV/dt;

```

#### 4. CONCLUSION

The EyesWeb modules for human tracking and movement analysis have been experimented and used in different applications, environments and works that include music learning [16], expressive gesture recognition and emotional analysis; it will be used in the EU FP7 ICT 3-year Project ASC-INCLUSION where user tracking and movement analysis will be integrated in a software platform that will assist children affected by Autism Spectrum Conditions to understand and express emotions. Future work will include the expansion of the library to offer more functionalities, support new devices and allow a higher quality movement analysis. The library will be included in the next release of the EyesWeb development platform.

#### 5. ACKNOWLEDGMENTS

This research has been partially funded by the European Community Seventh Framework Programme (FP7/2007-2013) ICT, under grant agreement No. 289021 (ASCInclusion).

#### 6. REFERENCES

- [1] A. Camurri, P. Coletta, G. Varni, and S. Ghisio, “Developing multimodal interactive systems with eyesweb xmi,” in *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07)*, pp. 302–305, ACM, 2007.
- [2] “Openni.” <http://www.openni.org/>.
- [3] “Microsoft kinect for windows sdk.” <http://www.microsoft.com/en-us/kinectforwindows>.
- [4] A. Camurri, G. Volpe, G. De Poli, and M. Leman, “Communicating expressiveness and affect in multimodal interactive systems,” *Multimedia, IEEE*, vol. 12, pp. 43 – 53, jan.-march 2005.
- [5] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, “Multimodal analysis of expressive gesture in music and dance performances,” *Lecture Notes in Artificial Intelligence*, vol. 2915, pp. 20–39, 2004.
- [6] H. G. Wallbott and K. R. Scherer, “Cues and channels in emotion recognition,” *Journal of Personality and Social Psychology*, vol. 51, no. 4, pp. 690–699, 1986.
- [7] P. E. Gallaher, “Individual differences in nonverbal behavior: Dimensions of style,” *Journal of Personality and Social Psychology*, vol. 63, no. 1, pp. 133–145, 1992.
- [8] H. G. Wallbott, “Bodily expression of emotion,” *European Journal of Social Psychology*, vol. 28, pp. 879–896, 1998.
- [9] R. Laban and F. C. Lawrence, *Effort*. USA: Macdonald & Evans, 1947.
- [10] A. Camurri, C. Canepa, P. Coletta, B. Mazzarino, and G. Volpe, “Mappe per affetti erranti: a multimodal system for social active listening and expressive performance,” 2008.
- [11] D. Winter, *Biomechanics and motor control of human movement*. Toronto: John Wiley & Sons, Inc., 1990.
- [12] P. Morasso, “Three dimensional arm trajectories,” *Biological Cybernetics*, vol. 48, pp. 187–194, 1983. 10.1007/BF00318086.
- [13] A. Wilson, A. Bobick, and J. Cassell, “Recovering the temporal structure of natural gesture,” in *Proc. of the Second Intern. Conf. on Automatic Face and Gesture Recognition*, pp. 66–71, 1996.
- [14] A. I. Shapiro, *The Movement Phrase and its clinical value in Dance/Movement Therapy*. PhD thesis, Master of Arts in Dance/Movement Therapy, MCP-Hahnemann University, 1999.
- [15] L. Bishko, “The use of labian analysis for the discussion of computer animation,” in *The 3rd Annual Conference of the Society for Animation Studies*, 1991.
- [16] G. Volpe, G. Varni, A. R. Addessi, and B. Mazzarino, “Besound: embodied reflexion for music education in childhood,” in *Proceedings of the 11th International Conference on Interaction Design and Children, IDC ’12*, (New York, NY, USA), pp. 172–175, ACM, 2012.



# FEED-DRUM E SKIN-ACT: L'ULTIMA FRONTIERA DELLO STRUMENTO AUMENTATO

Walter Cianciusi

Conservatorio di Musica  
Luigi Canepa - Sassari  
cianciusi@me.com

Lorenzo Seno

Conservatorio di Musica  
Alfredo Casella - L'Aquila  
lorenzo.seno@mnt-aq.it

## ABSTRACT

Nelle opere più recenti del compositore Michelangelo Lupone vi è una costante interazione e un reciproco adattamento tra la costruzione dello strumento, la composizione musicale e l'esecuzione. Il confine fra i tre diversi ambiti diventa sempre più sottile nelle opere concepite sugli strumenti "aumentati" Feed-Drum e SkinAct. La stesura compositiva appare talmente correlata allo strumento da esaurirne le possibilità comportamentali, sia da un punto di vista fisico che musicale.

La pratica viene reiterata con crescente convincimento e puntualità: uno strumento della tradizione diventa piattaforma base per un'operazione di "aumentazione" delle opportunità di modulazione timbrica e comportamentale grazie all'utilizzo dell'elettronica; un brano musicale condiviso con uno speciale interprete (parte dell'aumentazione dello strumento a tutti gli effetti) conduce all'enunciazione e immediatamente dopo all'esaurimento delle possibilità vibrazionali ed emozionali dello strumento.

Questo breve saggio si propone di esplorare dalla prospettiva scientifica ed estetica i postulati comuni alla ultima produzione di Lupone, che rappresenta senza dubbio lo stato dell'arte della ricerca connessa all'idea di strumento aumentato in Italia e oltre.

## 1. INTRODUZIONE

La musica occidentale ha legato la sua evoluzione alla trasformazione degli strumenti e delle tecniche esecutive attraverso la costante interazione tra le esigenze espressive del linguaggio musicale, le esigenze acustiche, le tecniche di emissione del suono. Consideriamo come dalla Viella del secolo X si passa alla famiglia delle Viole rinascimentali e si giunge alla famiglia dei Violini. In termini compositivi questo coincide con il transito dalle forme monodiche che doppiano la voce e il ritmo sillabico all'autonomia formale della musica strumentale (con conseguente allargamento del *range* frequenziale) alle grandi forme e agli organici della musica barocca. La tecnica esecutiva è integrata in questo processo perché l'interprete non ha solo il ruolo di veicolo per la traduzione acustica, ma anche di esperto che verifica i criteri di agilità e di ergonomia dello

strumento, d'inventore delle soluzioni di adattamento e virtuosistiche [5].

Viste in termini temporali, oggi le trasformazioni del linguaggio musicale, che si misurano in pochi decenni, ci appaiono decorrelate dalla trasformazione fisica degli strumenti d'orchestra. La tecnica esecutiva tenta di allineare le caratteristiche vibrazionali dello strumento con quelle del linguaggio ma, in molti casi, sollecita le aberrazioni del sistema fisico (e.g. suoni multipli dei fiati, eccitazioni non convenzionali delle corde e dei corpi risonanti), che rendono fortemente complicata o aleatoria la riproducibilità dei fenomeni acustici, e di conseguenza la notazione e la predizione da parte del compositore. Basti far riferimento al concetto di "hyperinstrument" introdotto da Tod Machover [4]: l'elettronica risulta ancora una giustapposizione forzata al sistema violino, violoncello o pianoforte. Non una trasformazione dello strumento fisico ma un macchinoso cambiamento di destinazione d'uso.

## 2. ELETTRONICA E AUMENTAZIONE

L'utilizzo dell'elettronica come elaborazione del suono strumentale ha prodotto trasformazioni profonde soprattutto nel lessico compositivo e uditivo, ha aderito con immediatezza alle esigenze espressive del linguaggio musicale ma solo su pochi strumenti ha favorito una reale trasformazione, chiarendo che con trasformazione s'intende l'ampliamento o la caratterizzazione esecutiva e acustica (e.g. chitarra elettrica). Consideriamo che nella musica colta, dove è evidente la forza di trasformazione posta sul sistema linguistico e tecnico, l'elettronica e lo strumento tradizionale cercano da molto tempo di interagire, di integrarsi, di condividere lo sviluppo dei suoni non riuscendo però a perdere la reciproca identità. In molti brani l'elettronica è posta in parallelo, dialoga, si fonde, prolunga lo strumento, ma non ne cambia le caratteristiche acustiche e d'esecuzione. In questi brani si avverte la presenza di un nuovo organismo sonoro, ma l'azione e il controllo dell'esecutore strumentale restano parziali e non necessariamente sono causa riconoscibile e modulante del suono. Soprattutto quando il brano musicale utilizza l'elaborazione in tempo reale dello strumento, e in generale quando si esegue con il "live electronics", la difficoltà più evidente per l'esecutore è di rendere il proprio comportamento, tecnico ed espressivo, coerente ai fenomeni acustici risultanti. I destini percettivi dello strumento acustico e dell'elettronica restano separati o decorrelati; anche quando l'ascolto effettua una sensata

ricostruzione delle informazioni musicali e il linguaggio musicale sostiene la coerenza delle informazioni, riconosciamo la distanza tra il sistema vibrazionale dello strumento, il processo d'elaborazione elettronica, la tecnica esecutiva.

Ed allora lo studio dei comportamenti fisici degli strumenti, la traduzione in modelli matematici e poi la simulazione di essi con sistemi di calcolo numerico, appare come la strada più perseguitabile per approdare a forme specifiche di integrazione, alla trasformazione o invenzione di nuovi strumenti coerenti alle attuali esigenze musicali. La ricerca che opera in questo senso affronta il complesso scenario della realtà fisica, con l'intento di costruire metodi analitici e sintetici adeguati alla rappresentazione dei fenomeni. Ovviamente lo scopo non è l'emulazione degli strumenti d'orchestra per un loro utilizzo virtuale e facile (tipica e limitante deriva commerciale), bensì la conoscenza estesa del fenomeno vibrazionale, la verifica dei modelli, la validazione acustica del processo logico e numerico [3].

Questi sono stati i presupposti con cui è iniziato il lavoro musicale e scientifico che ha prodotto l'opera "Gran Cassa" (Lupone, 1999) e che successivamente ha permesso la realizzazione dello strumento *Feed-drum*.

### 3. LO STRUMENTO DI BASE

La *grancassa sinfonica*, lo strumento a percussione più grave, è stata introdotta nell'orchestra solo nel XVIII secolo e nel secolo successivo ha raggiunto la forma che oggi conosciamo. In particolare la forma stretta e lunga ancora in uso nelle bande militari è stata portata a 80-90 cm di diametro e 35-50 cm di fusto; questo è chiuso da uno o da ambedue i lati con membrane di pelle naturale tenute in diversi modi da sistemi che provvedono a regolarne anche la tensione. La versione più grande, adeguata ai massimi organici orchestrali, è chiamata *grancassa imperiale*, ha due membrane e raggiunge 102 cm di diametro. Su questo tipo di strumento, messo a disposizione dal Conservatorio di L'Aquila e utilizzato per la prima esecuzione dell'opera *Gran Cassa*, sono state effettuate le sperimentazioni preliminari al lavoro di composizione musicale e di progetto del *Feed-drum*.

L'utilizzo della grancassa, per quanto sostanziale in orchestra e costante da Mozart in poi, è considerato secondario e ristretto a pochi modi di emissione del suono: il rullo (nota lunga), spesso finalizzato al crescendo e il colpo di riempimento nelle sequenze ritmiche. Non sono studiate tecniche specifiche, come nel caso dei timpani, e i battenti tipici sono la mazza e le bacchette da timpano.

L'idea di un'opera musicale, interamente basata su questo strumento, è nata dall'osservazione dei modi vibrazionali della membrana e dalle esperienze precedentemente fatte sui Planofoni® e sul modello fisico della corda e dell'archetto (Palumbi, Seno 1997) presso il CRM - Centro Ricerche Musicali di Roma.

Nonostante la membrana permetta l'eccitazione di un numero elevato di modi ad alta frequenza, la loro durata nel tempo non è apprezzabile all'ascolto, fatto salvo il contributo timbrico della fase d'attacco del suono. Le variazioni possibili del modo di emissione, adeguate ad

una sufficiente risposta acustica del risonatore (fusto), sono limitate e poco modulabili. La frequenza di base<sup>1</sup>, ottenuta con la tensione delle membrane, superiore e inferiore, vincolate al bordo con 16 tiranti meccanici ognuna, risente della distribuzione non omogenea delle forze, che contribuisce a rendere complesso lo spettro dei modi reali [2].



**Figura 1.** Prima esecuzione dell'opera *Gran Cassa*, Alessandro Tomassetti esegue su grancassa imperiale, Corpi del suono 1999 - L'Aquila

#### 3.1 Il Percorso Sperimentale

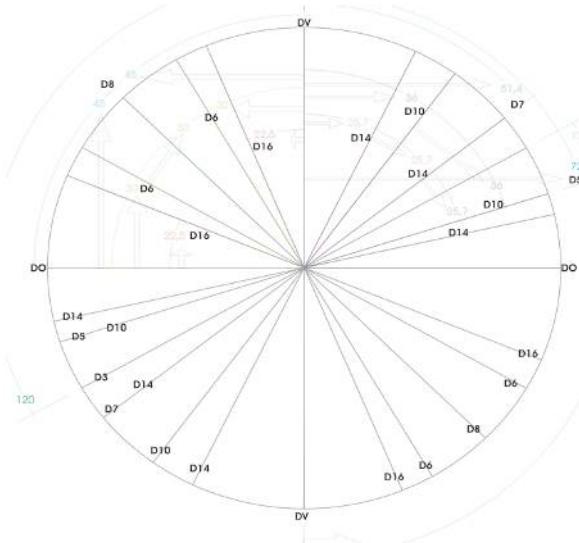
Michelangelo Lupone non sembra ricevere ispirazione dalle alte sfere o da chissà quale mistica derivazione ma da una concreta, attenta, scrupolosa e forse anche maniacale sperimentazione. Nel caso di specie dopo una fase di ascolto e analisi delle caratteristiche sonore dello strumento, utilizzando anche modi di eccitazione non convenzionali come strusciati e jetée di spazzole, è stato necessario, per la scrittura dell'opera musicale, individuare e catalogare un'ampia gamma di possibilità sonore con diversi gradi di contiguità. Le sperimentazioni sono state fatte con l'intento di raggiungere i seguenti obiettivi:

1. variazione della frequenza di base attraverso l'introduzione di vincoli nodali posti sulla membrana,
2. distinzione di più timbri in base al tipo, al modo e alla posizione dell'eccitazione,
3. modulazione del suono attraverso glissandi, vibrati, portamenti e micro-articolazione ritmica,
4. variazioni continue e a gradini della dinamica, in base al tipo di smorzamento imposto alla membrana.

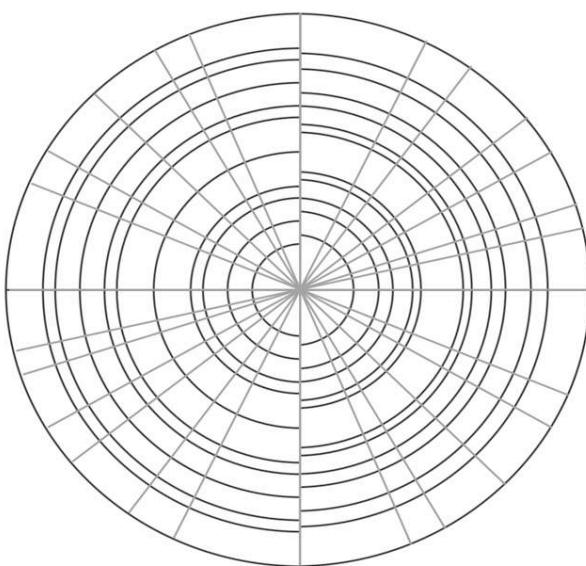
Le caratteristiche della *grancassa* tradizionale ovviamente non permettono di ottenere una gamma di risultati acustici rilevanti sugli obiettivi preposti. Per esplorare la ricchezza timbrica della fase d'attacco e isolare i mo-

<sup>1</sup> Abbiamo preferito, qui e nel seguito, questo termine a quello talvolta utilizzato di "fondamentale", perché quest'ultimo dovrebbe essere più propriamente riservato alla frequenza di pitch degli strumenti intonati.

di vibrazionali, è stato di conseguenza creato un sistema di condizionamento elettronico della membrana. Attraverso il principio del feedback, il segnale prodotto dall'eccitazione della membrana è riportato sulla membrana stessa sotto forma di pressione acustica. Il risultato è il prolungamento infinito del suono. Il sistema controlla lo smorzamento del movimento della membrana, quindi il decadimento del suono, e permette di isolare i modi di alta frequenza con l'azione combinata dei nodi posti sulla membrana e della quantità di energia immessa nel feedback. La stabilità del segnale, ottenuta con il sistema di condizionamento, ha permesso di sperimentare e disegnare sulla superficie della membrana una prima mappa semplificata dei modi oscillatori basata sulle funzioni di Bessel [1]. La mappa è stata limitata a 13 diametri e 8 cerchi nodali (Fig. 2, Fig. 3), questi ultimi divisi in semicerchi pari (a sinistra), semicerchi dispari (a destra).



**Figura 2.** Mappa dei primi 13 diametri nodali



**Figura 3.** Feed-drum: prima mappa con 13 diametri e 8 cerchi nodali

L'intervento elettronico sullo strumento ha lasciato inalterati la topologia e i tratti acustici primari ma ha esteso i criteri e i controlli vibrazionali fino a permettere evidenti differenze di timbro tra i modi, l'emissione di note lunghe e modulabili come su una corda, l'adattamento dell'energia acustica indipendentemente dalle frequenze emesse.

Per mantenere l'agilità e un'adeguata riproducibilità dei fenomeni, la prima catalogazione dei suoni e delle tecniche d'esecuzione è stata limitata all'uso delle dita, delle mani e delle braccia (Fig. 4).



**Figura 4.** Feed-drum: una delle tecniche di esecuzione per l'eccitazione di modi ad alta frequenza

Durante la scrittura dell'opera *Gran Cassa* sono stati sperimentati anche oggetti di diversa forma e dimensione in grado di occupare porzioni nodali più ampie o multiple; questo ha permesso di ampliare ulteriormente le possibilità sonore, ma la complessità dei fenomeni vibrazionali ha reso necessaria un'analisi anche delle parti meccaniche dello strumento per comprendere e ridurre le dispersioni e il contributo non lineare introdotto dalle vibrazioni dei materiali costruttivi e dei loro accoppiamenti.

In base a tali complicazioni si è ritenuto opportuno procedere alla progettazione e alla realizzazione di un nuovo strumento, il *Feed-drum* (Fig. 5), sia per estendere le possibilità acustiche, sia per permettere un uso ergonomico delle nuove tecniche esecutive.

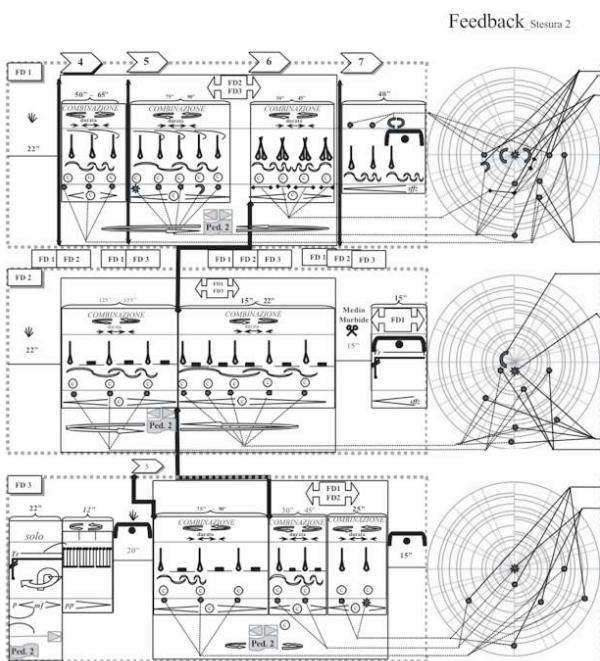


**Figura 5.** Feed-drum

In particolare è stato trasformato l'assetto vibrazionale eliminando la membrana inferiore, cosa che ha semplificato l'accordatura della frequenza di base dello strumento (30 Hz) e ha ridotto il tempo di salita dell'eccitazione dei modi alti. È stata applicata una membrana sintetica con caratteristiche isotrope ed elevata flessibilità e su di essa è stata realizzata la mappa precedentemente descritta, con colori che rendono più visibili le aree di esecuzione. Il fusto e la centina di tiro della membrana sono stati realizzati in acciaio e alluminio; in particolare la centina è stata resa più rigida, è stata ridotta l'altezza e aumentata la superficie di aderenza. Il sistema di sospensione è stato realizzato in modo da separare completamente il *Feed-drum* dalla struttura di appoggio a terra; tutte le parti meccaniche, che si trovano in contatto tra loro, sono state divise con uno strato intermedio di materiale antivibrazione.

#### 4. FEED-DRUM

Sul *Feed-drum*, nonostante ci siano molti aspetti ancora da studiare, è stato possibile verificare la riproducibilità dei suoni catalogati e delle modulazioni, l'agilità dei modi di eccitazione e di controllo, l'estensione in frequenza e le caratteristiche timbriche. Questo ha facilitato la stesura della partitura di esecuzione dell'opera *Gran Cassa* e successivamente dell'opera *Feedback* (per 3 *Feed-drum*) (Fig. 6), dove sono presenti, oltre alle usuali indicazioni di comportamento ritmico, le forme e i punti di eccitazione della membrana, la quantità di energia da immettere nel feedback, le frequenze e le durate dei suoni, l'intensità puntuale, i tipi di modulazione (vibrati, glissandi, portamenti), il range e la velocità dell'azione modulante.



**Figura 6.** Feedback per 3 feed-drum (2002) estratto della partitura

#### 4.1 Sistema di Condizionamento e Attuazione

L'eccitazione della membrana avviene per mezzo di un altoparlante (diametro = 45 cm) attraverso un tratto di guida d'onda di lunghezza 11 cm (realizzata per cogliere la massima pressione acustica tra il centro e 1/3 del raggio), dunque piuttosto corta, come fattore di forma. Mentre è agevole ottenere, oltre alla frequenza di base a 30 Hz, la frequenza di 68.9 Hz corrispondente al modo (0,2), è stato per converso impossibile ottenere quella a 47.8 Hz corrispondente al modo (1,1). A queste frequenze il moto dell'aria eccitata dall'altoparlante è presumibilmente ben schematizzabile con un moto a pistone, che esercita quindi una pressione pressoché uniforme sulla membrana. Un'eccitazione uniforme è poco compatibile con la forma modale (1,1).

L'altoparlante è pilotato da un segnale elettrico di potenza, ottenuto attraverso un feedback che preleva un segnale di deflessione della membrana mediante un sensore piezoceramico posto sul bordo. In questo modo si ottiene un oscillatore "multimodale" generando un feedback su di un elemento risonante, la membrana. Il guadagno di anello è controllabile mediante un pedale.

#### 4.2 L'Intonazione dei Modi Altì

L'intonazione avviene mediante l'azione combinata del guadagno di controllazione e della pressione su di uno o due punti di qualche linea nodale. L'effetto della pressione può essere schematizzato in prima approssimazione come duplice: da un lato come l'introduzione di un vincolo nei punti di pressione, dall'altro come uno spostamento del "punto di lavoro" della membrana attorno ad una tensione lievemente più alta, con un innalzamento della velocità di propagazione delle onde trasversali. Come conseguenza, tutte le frequenze si spostano verso l'alto. Si tratta di un meccanismo di spostamento a "pitch-shift", nel senso che le frequenze dei modi vengono tutte moltiplicate per un comune fattore, lasciando quindi invariati i loro rapporti. Questo effetto è stato riscontrato nella pratica, ed è utilizzato per ottenere il vibrato. Il termine "pitch-shift" è però in questo caso improprio perché lo spettro delle parziali della membrana non è armonico, e pertanto non è definibile un pitch.

L'apposizione di punti di vincolo ( $z = 0$ ) ha l'effetto, in linea di principio, di inibire tutti i modi che non abbiano alcun set di linee nodali passanti per tutti i suddetti punti. Ad esempio, premendo sul centro della membrana si rendono impossibili tutti i modi con ( $m = 0$ ) [1], dato che questo punto è per questi sempre un ventre. La pressione su di un qualunque altro punto della membrana, sempre in linea di principio, rende possibili tutti i modi con  $m \geq 1$ , perché sarà sempre possibile far passare un diametro nodale per quel punto. In linea pratica, poiché il vincolo non è perfetto, sarà favorito quell'unico modo che possegga un diametro e un cerchio nodali passanti per quel punto. La circostanza che

le funzioni di Bessel non hanno radici coincidenti ha come conseguenza che modi di ordine  $m$  diverso non possono avere cerchi nodali coincidenti. Anche modi con medesimo  $m$  e diverso  $n$  non possono, ovviamente, avere cerchi nodali coincidenti. Due modi distinti possono per converso avere diametri coincidenti, se i loro indici sono in rapporto intero. Un singolo punto di pressione diverso dal centro individua quindi un solo modo avente un diametro e un cerchio passante per quel punto. I punti che "discriminano" meglio i modi di frequenza vicina sono però quelli vicini al centro, perché verso la periferia i cerchi nodali si infittiscono e quindi un singolo punto tende ad averne molti vicini. E' quindi il primo cerchio nodale, il più interno, quello che discrimina meglio i modi, come mostra anche un'analisi della varianza.

In linea di principio, la pressione su due punti qualsiasi della membrana potrebbe porre vincoli incompatibili con qualunque modo.

Tuttavia è bene limitare tutte queste considerazioni solo a modi di ordine relativamente basso. E' infatti presumibile che, con l'aumentare dell'ordine dei modi, risulti meno valida l'approssimazione della membrana non rigida, dato che il passo internodale diventa comparabile con lo spessore della membrana stessa.

Vi sono inoltre ulteriori considerazioni. La classica equazione della membrana generalmente utilizzata per ricavare i modi (si veda 3.7, p. 69, di [2]) è, come già detto, del tutto conservativa e non tiene conto né delle dissipazioni per attrito interno, né dell'irraggiamento, entrambi meccanismi che sottraggono energia al moto delle parziali, provocandone il decadimento nel tempo in assenza di forzante.

Affrontare in modo simbolico la soluzione dell'equazione corrispondente al moto vibro-acustico descritto è una strada irta di difficoltà. E' certamente possibile risolverla con metodi numerici (quali FEM, BEM, ecc.) ma anche in questo caso, volendo tenere conto dell'accoppiamento acusto-elastico e delle dissipazioni interne della membrana, il problema resterebbe comunque molto delicato, e i risultati dovrebbero essere oggetto di attenta validazione sperimentale.

Anche in assenza di una soluzione, è però possibile considerare che il decadimento delle parziali è in ogni caso legato al fattore di merito ( $Q$ ) della relativa risonanza, e provoca un allargamento della riga spettrale, tanto più accentuato quanto più il relativo modo è smorzato. Gli attriti interni sono proporzionali alla velocità di variazione della curvatura locale, che aumenta con la frequenza. E' quindi presumibile che, in analogia con le corde, lo smorzamento dei modi aumenti con la loro frequenza. Ne consegue che nelle alte zone spettrali, dove i modi sono ravvicinati e densi, la funzione di trasferimento della membrana si presenta più continua che rigata, con lievi picchi sulle frequenze modali. In queste zone, i modi eccitabili risultano meno nettamente

definibili, e dipendenti dal guadagno di anello e dalle caratteristiche di frequenza del circuito elettronico di controreazione. Per converso, il passaggio da un modo ad uno frequentemente adiacente ha poca influenza sulla frequenza risultante.

La costruzione di una prossima mappa perfezionata per l'eccitazione di modi deve quindi prevedere un'accorta scelta di coppie di punti in grado di compiere un'alta discriminazione fra modi.

Inoltre le frequenze modali vanno verificate sperimentalmente, perché è presumibile che le frequenze di alcuni modi si discostino da quelle teoriche a causa della presenza dell'attuatore con relativa guida d'onda, che intercettano una luce pari a 1/3 del diametro della membrana. La misura di questi scostamenti non è determinabile in modo affidabile da considerazioni teoriche, perché il modello complessivo è troppo complesso e risolvibile, come già detto, solo con metodi numerici.

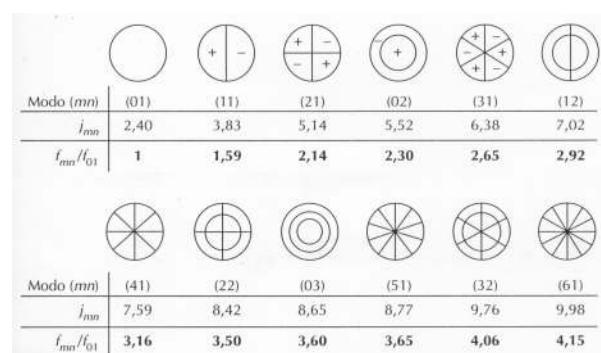


Figura 7. Modi 01, 11, 21, 02, 31, 12, 41, 22, 03, 51, 32, 61

## 5. SKIN-ACT E SPAZIO CURVO

Gli SkinAct (Skin Active) sono strumenti aumentati a percussione. Come per il feed-drum siamo di fronte ad una membrana di grancassa (questa volta non imperiale) posta in feedback attraverso sensori e attuatori. L'attuatore nel caso specifico è un piccolo shaker applicato sulla pelle stessa, laddove il feed-drum prevedeva un woofer per porre in vibrazione la membrana.

La recente opera Spazio Curvo di Michelangelo Lupone, la cui prima esecuzione risale al 26 giugno 2012, prevede l'utilizzo di 3 SkinAct da parte di un solo interprete. Philippe Spiesser con gesti decisi esplora le membrane, porta le vibrazioni della materia al loro limite e trasforma i suoni da impulsivi a risonanti, fino ad un canto incrociato di voci che con un flusso continuo, migra tra i diversi momenti espressivi dell'opera.

L'opera parte dall'idea di riprodurre lo spazio acustico, che è intorno all'interprete, anche intorno all'ascoltatore, con caratteristiche coerenti di mobilità e localizzazione. Ciò può rendere percepibile le differen-

ze e le trasformazioni più intime dei suoni affinché diventino espressive. Per realizzare una scrittura musicale che potesse evidenziare questa idea, Lupone si è ispirato al concetto di spazio percepito e misurato attraverso il tempo. Una semplice metafora può chiarire il concetto: siamo inclini a misurare in termini di durata lo spazio attraversato da un ascensore, in tal modo possiamo percepire uno spazio acustico attraverso le articolazioni ritmiche e le durate dei suoni che in esso sono presenti. I ritmi e le durate rappresentano il materiale compositivo di base di "Spazio curvo" sul quale vengono articolati i timbri eterogenei dello SkinAct. L'opera è formalmente divisa in cinque sezioni, ognuna delle quali propone una diversa concezione del ritmo:

1) "nel suono" (battimenti ottenuti con le frequenze parziali della membrana), 2) "nello spazio" (ritmi derivati dalla percezione dei movimenti e delle localizzazioni dei suoni nello spazio acustico), 3) "del suono" (ritmi ottenuti con tecniche di scomposizione temporale del suono e ricomposizione a densità diverse), 4) "con il suono" (ritmo organizzato con accenti e pause), 5) "poliritmia" (ritmi diversi sovrapposti temporalmente).

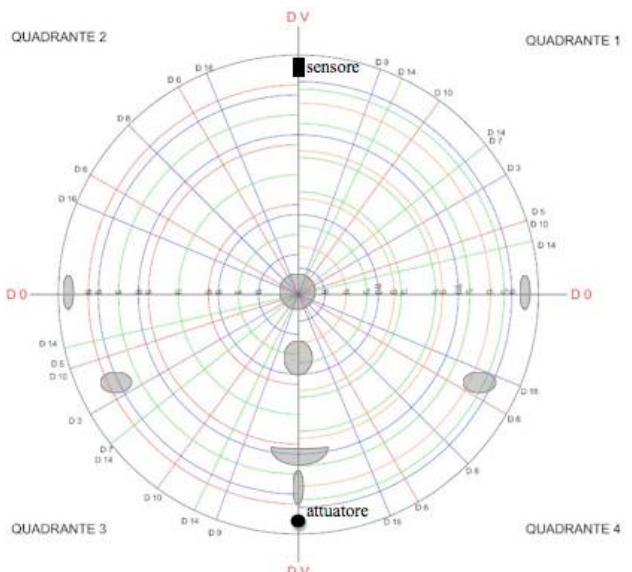
I tre SkinAct posti a distanza ravvicinata, si influenzano reciprocamente dando luogo a una complessità elevata del fenomeno sonoro, e delle tecniche di controllo. Tale influenza viene utilizzata in termini compositivi per ottenere devianze delle altezze dei singoli strumenti con conseguenti variazioni del timbro.

Le tre membrane vengono accordate su una frequenza base di, rispettivamente, 35, 38 e 42 Hz.

Questa condizione iniziale finisce per enfatizzare i battimenti come elemento fondante del percorso musicale in Spazio Curvo. In figura 9 si possono osservare le aree di controllo delle frequenze parziali poste in battimento.



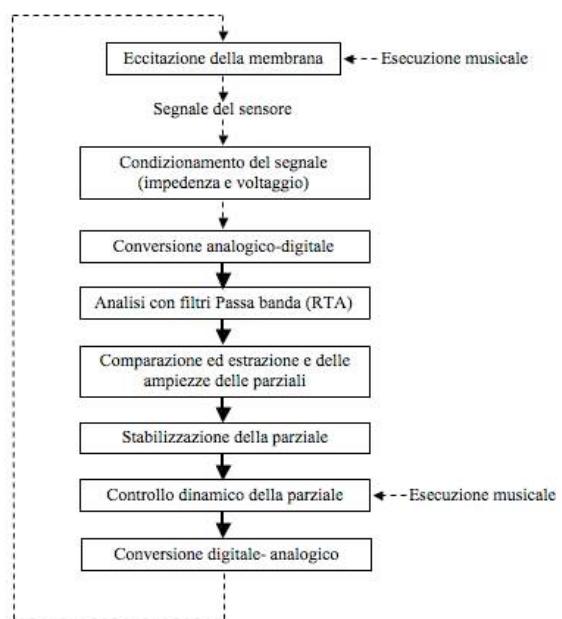
**Figura 8.** I tre SkinAct durante una sessione di test presso il CRM - Centro Ricerche Musicali di Roma



**Figura 9.** La mappatura dello SkinAct

Le frequenze di risonanza utilizzate per l'articolazione del timbro nell'opera Spazio Curvo sono Hz:  
 66.2 – 97 – 103.4 – 124.6  
 138 – 140.5 – 180 – 218 – 275  
 294.3 – 375 – 420.7 – 438  
 473.5 – 537.9 – 720.7

Lo schema di principio del processo di controllo delle frequenze parziali è molto simile a quello utilizzato nelle opere per feed-drum



**Figura 10.** Anche per lo SkinAct vale il principio di contoreazione già enunciato a proposito del Feed-Drum. Per esecuzione musicale si intende ogni gesto dell'interprete sulla membrana.

## 6. SOLIPSISMO E IDEALISMO TRASCENDENTALE

Osservando col senso del poi i percorsi sperimental-compositivi di Michelangelo Lupone si ha l'impressione che egli assuma la coscienza empirica, individuale, come fondamento di ogni forma di conoscenza. Il mondo esterno, quello dell'ascoltatore e del pubblico in genere, appare solo come una momentanea percezione della realtà. L'opera, sempre più vicina al dominio della coscienza (si pensi ad esempio agli ultimi lavori che il Maestro ha realizzato insieme all'artista Licia Galizia presso il Centro Ricerche Musicali di Roma - [www.crm-music.it](http://www.crm-music.it)), prodotto di un iter interattivo, adattivo ed infine evolutivo, è la sola realtà. Le gesta sonore di essa fan parte di una morale prestabilita dall'io algoritmico, al di là delle leggi sintattiche condivee dal mondo esterno, al di là di altre soggettività. Come potrebbe mai un compositore terzo avvicinarsi agli strumenti aumentati creati da Lupone? Essi vivono una speciale condizione di estensione della coscienza dell'autore. La fisicità del loro apparire, le forme, lo spazio occupato rispondono ad esigenze composite talmente stringenti che ogni tentativo di modulazione del loro destino risulta vano ed inessenziale. Nonostante possa apparire accattivante ad un primo sguardo l'estetica degli strumenti è dettata da esigenze fisiche/compositive. E' un "bello" - quello dello strumento musicale - che deriva da leggi naturali, non da una ricerca a sé stante.

Lupone è nato nello stesso giorno di Immanuel Kant. Per entrambi la metafisica, che indaga sull'essenza delle cose, non ha valore come scienza perché la sua sostanza non è accessibile ai nostri sensi e non possiamo quindi comprenderla. Ogni conoscenza parte dall'esperienza: non c'è conoscenza senza esperienza sensibile.

### 6.1 Volontà di Potenza

Si è artisti al prezzo di percepire come contenuto, come "la cosa stessa" quello che tutti i non-artisti chiamano "forma". Con ciò ci si trova naturalmente in un mondo alla rovescia: perché ormai il contenuto diventa qualcosa di assolutamente formale.

E' così che Lupone, interrogato più volte dagli autori di questo saggio in merito ai contenuti delle sue opere, ha finito per descriverne profili di basso livello, ciò che appariva come mera forma: grandezze e specifiche tecniche delle singole parti che compongono lo strumento, mai un riferimento alla costruzione musicale.

D'altra parte lo diceva già Nietzsche: non si può ritrovare ciò che è la causa del fatto che generalmente si dà sviluppo, sulla via della ricerca sullo sviluppo; non si deve voler comprendere ciò in quanto "diveniente", ancor meno in quanto divenuto...la "volontà di potenza" non può essere divenuta.

Bisogna allora raggiungere un'altezza e una prospettiva a volo di uccello per la contemplazione, dalla quale si

possa cogliere che tutto va proprio come *dovrebbe andare*; e che ogni genere di imperfezione e di dolore nell'arte appartiene alla suprema desiderabilità.

Lo strumento musicale, pure se aumentato, resiste in qualche modo alla volontà del compositore. In rapporto alle resistenze che una forza trova per dominarle, deve crescere la misura del fallimento e della fatalità in tal modo sfidati; e nella misura in cui ogni forza può scaricarsi soltanto contro qualcosa che oppone resistenza, in ogni azione vi è necessariamente un ingrediente di dispiacere. Solo questo dispiacere funziona da stimolo e fortifica la volontà di potenza. Solo questo dispiacere dà vita all'opera d'arte.

## 7. L'INTERPRETE AUMENTATO

Chi scrive ha avuto l'opportunità di seguire innumerevoli sessioni di prova in presenza di Lupone e degli interpreti delle sue opere. E' certo che la scelta dell'interprete per Lupone è parte integrante del processo di aumentazione dello strumento. Così come in passato Sylvano Bussotti preferiva scrivere per David Tudor piuttosto che per pianoforte, oggi Lupone cala l'interprete al centro del meccanismo di emancipazione dello strumento tradizionale. La partitura di Skin-Act ad esempio era solo allo stato embrionale quando Philippe Spiesser ha iniziato a prender confidenza con gli strumenti in gioco. La sperimentazione del compositore aveva a che fare con la naturale fisicità di quel particolare interprete di fronte a quegli specifici strumenti. Come può un interprete terzo avvicinarsi ad una partitura di Lupone scritta per un altro strumento/interprete? La sua finirebbe per essere una goffa imitazione della fisicità altrui. L'interpretazione non potrebbe mai liberarsi dall'ombra dell'imitazione. Si ha insomma l'impressione che nel bel contenitore rinforzato utilizzato di solito dal compositore per far viaggiare la propria opera (persino esso, il flight-case, costruito da artigiani amici su istruzioni - dettagliatissime - del compositore), andrebbe inserito a forza l'interprete che della specifica partitura si rende protagonista assoluto, conoscitore attento di particolari invisibili e segreti per l'ascoltatore e per la gran parte dei musicologi.

## 8. BIBLIOGRAFIA

- [1] – S.Finch, Bessel Functions Zeroes, INRIA, Oct. 2003, <http://pauillac.inria.fr/algo/csolve/bs.pdf>
- [2] – M. Lupone, Corpi del Suono, Istituto Gramma - L'Aquila 1999
- [3] – M. Palumbi, L. Seno, Physical Modeling by Directly Solving Wave PDE, In Proc. of the 1999 International Computer Music Conference, ICMA 1999
- [4] - T. Machover, Shaping Minds Musically, BT Technology Journal 22 (4), 2004
- [5] – N.H. Fletcher, T.D. Rossing, The Physics of Musical Instruments, Springer 1991



# SoundingARM

## RAPPRESENTAZIONE ACUSTICA DI UNA MAPPA SPAZIALE

Nicola Scattolin, Antonio Rodà, Sergio Canazza  
Università di Padova  
nicola.scattolin@studenti.unipd.it  
{roda, canazza}@dei.unipd.it

Serena Zanolla  
Università di Udine  
serena.zanolla@uniud.it

### SOMMARIO

La capacità di costruire la mappa cognitiva, la rappresentazione mentale, degli ambienti è fondamentale per l'acquisizione delle competenze di orientamento e mobilità. I vedenti dispongono di mappe cognitive con immagini visive, mentre le persone con minorazioni visive si avvalgono di mappe mentali costruite mediante immagini tattili. Dato che le informazioni spaziali derivano principalmente dal canale visivo, le persone con disabilità visive incontrano difficoltà nel: a) riconoscere immediatamente il tipo di stanza, b) muoversi con sicurezza negli spazi di varia grandezza che richiedono l'esplorazione locomotoria e la capacità di "ascoltare" l'ambiente e c) individuare rapidamente la posizione di uno o più oggetti (esplorazione tattile).

In questo articolo presentiamo SoundingARM (Rappresentazione Acustica di una Mappa spaziale), un sistema in grado di fornire la mappa acustica degli ambienti interni. Questo sistema è stato progettato sulla base dell'assunto che la mancanza della vista può essere supportata da un'adeguata informazione spaziale derivante dai canali sensoriali compensativi. Mediante SoundingARM l'utente può esplorare l'ambiente restando sulla soglia della stanza ed eseguendo semplici gesti con le braccia (indicazione).

Il sistema SoundingARM utilizza il sensore Microsoft Kinect per determinare la posizione dell'utente e un software, sviluppato dagli autori presso i laboratori del Centro di Sonologia Computazionale del Dipartimento di Ingegneria dell'Informazione dell'Università di Padova, che usa questi dati per produrre informazione acustica sull'ambiente e gli oggetti contenuti.

### 1. INTRODUZIONE

Muoversi autonomamente nello spazio, riconoscere immediatamente un ambiente, pianificare un tragitto o scegliere il percorso più breve sono competenze spaziali di estrema importanza perché consentono lo svolgimento delle attività quotidiane in autonomia [1]. Queste competenze sono essenziali per tutti ma lo sono ancor di più per le persone con gravi disabilità visive (cecidità o ipovisione). In questi casi la conoscenza spaziale, e quindi la rappresentazione mentale dello spazio topografico, normalmente viene acquisita

*Copyright: ©2012 Scattolin 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.*

mediante strumenti compensativi, come le mappe tattili, o attraverso l'esplorazione diretta dell'ambiente [2]. L'insieme dei soggetti con disabilità visive può essere suddiviso in due grandi sottoinsiemi: le persone con ipovisione e le persone con cecità (anche se la disabilità visiva è in realtà un continuum che va dall'ipovisione alla cecità assoluta). Le persone con ipovisione hanno delle limitazioni funzionali che comprendono: elevata sensibilità alla luce, visione distorta, assenza di visione centrale o periferica, perdita dell'acuità visiva e del focus, ridotta capacità di discriminazione cromatica e offuscamento della vista. Le persone con cecità invece possono essere in grado di percepire la forma degli oggetti, il contrasto cromatico e la differenza buio/luce o essere affetti da cecità assoluta. Per ottenere sufficienti informazioni sull'ambiente, le persone con disabilità visive molto gravi devono compensare la mancanza della vista utilizzando altri canali sensoriali. Le persone con cecità assoluta, per esempio, dipendono totalmente dall'udito, dal tatto, dall'olfatto e dalle informazioni derivanti dall'interazione motoria con l'ambiente [3].

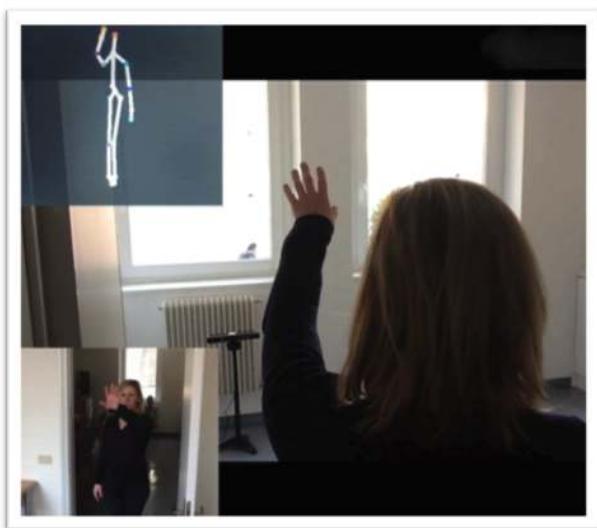
La capacità di muoversi autonomamente nello spazio richiede lo sviluppo di svariate strategie di elaborazione delle informazioni che vengono impiegate, ad esempio, per la localizzazione degli oggetti in ambienti non familiari; per cercare un oggetto l'utente non vedente deve eseguire l'esplorazione locomotoria e tattile di tutto l'ambiente. Di solito, nell'esplorazione di uno spazio si realizzano due tipi di ricerca:

1. l'esplorazione del perimetro, che fornisce informazioni sulle dimensioni e sulla forma dello spazio;
2. l'esplorazione attraverso una serie di movimenti in linea retta da un lato a quello opposto dell'ambiente.

Avvenuta l'individuazione degli oggetti, vengono messe in campo altre strategie volte a memorizzarne la posizione nello spazio. Queste strategie consistono nel: camminare in avanti e in dietro lungo la linea retta che virtualmente unisce due oggetti, o, camminare in avanti e in dietro da un punto del perimetro e un oggetto. Un'altra strategia consiste nell'utilizzare un oggetto come "punto di riferimento" da cui dirigersi in linea retta verso altri oggetti. Attraverso queste strategie l'utente costruisce una mappa mentale, rettangolare o a forma di cella Braille, che approssima la forma dello spazio e la posizione degli oggetti in esso contenuti [4]. E' facile immaginare che l'attuazione di questo tipo di strategie richiede una lunga, in ordine di tempo,

esplorazione locomotoria e tattile.

Il sistema SoundingARM (Rappresentazione Acustica di una Mappa Spaziale) intende offrire un “colpo d’occhio” dell’ambiente, una mappa acustica dello spazio, per permettere agli utenti di riconoscerne rapidamente la tipologia (cucina, camera da letto, soggiorno) e di localizzare velocemente gli oggetti in esso contenuti. Il sistema SoundingARM consente di esplorare una stanza restando fermi sulla soglia ed eseguendo il semplice gesto dell’indicazione. Se l’utente indica con il braccio una zona dello spazio, il sistema risponde producendo il feedback audio relativo all’oggetto che si trova in quella direzione. L’utente ottiene le informazioni di localizzazione dell’oggetto dal gesto stesso: l’oggetto è infatti collocato nella direzione di puntamento del braccio (Fig.1). Al momento, il feedback audio è costituito dalla sintesi vocale che descrive il tipo di oggetto o, semplicemente, ne riproduce il nome. Si prevede comunque di utilizzare presto anche altre tipologie di informazioni acustiche: *auditory icons, earcons, spearcons*, ecc. SoundingARM utilizza il sensore



**Figura 1.** Utilizzo di SoundingARM

Microsoft Kinect per determinare la posizione dell’utente e un software, sviluppato dagli autori presso i laboratori del Centro di Sonologia Computazionale del Dipartimento di Ingegneria dell’Informazione dell’Università di Padova, che impiega questi dati per produrre informazione acustica sullo spazio topografico.

I paragrafi seguenti documentano nel dettaglio il sistema SoundingARM. In particolare, nella Sez. 2 presentiamo lo stato dell’arte dei dispositivi parlanti e dei sistemi di *way-finding*; nella Sez. 3 illustriamo l’utilizzo, i principali obiettivi (Sec. 3.1) e i requisiti di sistema (Sec. 3.2) di SoundingARM. La Sez. 4 descrive l’architettura del sistema (i dispositivi di ingresso in Sez. 4.1, le componenti di elaborazione dei dati in tempo reale in Sec. 4.2, i dispositivi di uscita in Sec. 4.3, l’applicazione nella Sez. 4.4 e la patch in Pure Data nella Sez. 4.5). Nella Sez. 5 vengono riportate le osservazioni ricavate in seguito alla somministrazione di un test di usabilità ad un campione ristretto di

utenti con disabilità visive. Da ultimo, la Sez. 6 riporta le conclusioni e gli sviluppi futuri.

## 2. STATO DELL’ARTE

Negli ultimi anni, l’utilizzo di tecniche sempre più avanzate nel campo degli *Auditory Display* ha incrementato notevolmente il numero delle tecniche di compensazione della disabilità visiva che affligge milioni di persone con cecità e ipovisione. Gli Auditory Display possono infatti fornire un prezioso supporto nelle attività di orientamento e mobilità nello spazio. Esistono già diversi strumenti che fornendo informazioni acustiche, consentono alle persone con cecità o con ipovisione di compiere attività che prima non erano alla loro portata; i primi ad essere realizzati sono stati i *Talking Device*: dispositivi parlanti.

### 2.1 Talking Device

Di norma, gli ausili per l’autonomia quotidiana delle persone non vedenti o ipovedenti utilizzano le tecniche della sintesi vocale [5]. I principali sono software per la lettura; quelli più comunemente utilizzati sono: *JAWS (Job Access With Speech)* prodotto da “*Blind and Low Vision Group of Freedom Scientific*”; *Window-Eyes*, sviluppato da “*GW Micro*”; *VoiceOver*, la tecnologica della Apple che consente l’accesso alla lettura dello schermo; *SAToGO (System Access To Go)*, un programma online per la lettura dello schermo sviluppato da *Serotek* che, in collaborazione con la fondazione *AIR (Accessibility is a Right)*, ne ha realizzato una versione completamente gratuita; *NVDA (NonVisual Desktop Access)*, sviluppato da *NV Access* e *Zoom Text*, sviluppato da *Ai Squared* che consente di accedere a tutte (testo e immagini) le informazioni visive dello schermo. Vengono definiti Talking Device anche gli ausili per la lettura costituiti da sistemi portatili o desktop, computer-based o soluzioni standalone, uno scanner, un software OCR e un sintetizzatore vocale; fra i molti, si citano *OPENBook*<sup>1</sup> e *EYE-PAL*<sup>2</sup>. Si possono facilmente trovare anche applicazioni per iPhone, iPod Touch, e iPad, come *Voice Brief*<sup>3</sup>, che legge i testi in formato elettronico tramite la sintesi vocale.

Per estendere il campo dei Talking Device e avvicinarsi alla tipologia di sistemi più simili a SoundingARM, presentato in questo articolo, ora si considerano le tecnologie di *way-finding* che possono essere suddivise in due grandi categorie: i sistemi per ambienti interni e quelli per ambienti esterni. Generalmente le tecnologie per ambienti esterni utilizzano il GPS (*Global Positioning System*) per localizzare l’utente: di questo tipo si ricorda, per esempio, l’*Atlas system*<sup>4</sup>, una mappa parlante digitale che fornisce informazioni vocali su posizioni, direzioni e punti di interesse. I sistemi per ambienti esterni si basano spesso anche sulla tecnologia *infrared* (IR), come *Talking Signs®*<sup>5</sup> costituito

<sup>1</sup> <http://www.freedomscientific.com/>

<sup>2</sup> <http://www.abisee.com/>

<sup>3</sup> [http://www.voicebriefweb.com/Voice\\_Brief/](http://www.voicebriefweb.com/Voice_Brief/)

<sup>4</sup> <http://www.csun.edu/cod/conf/2003/proceedings/>

<sup>5</sup> <http://www.talkingsigns.com/>

da una emittente permanentemente installata nell'ambiente e un ricevitore da tenere in mano.

I sistemi per ambienti interni di solito utilizzano gli infrarossi [6], gli ultrasuoni [7], l'identificazione a radio frequenza (RFID) [8] o la *Computer Vision* [9].

Ultimamente anche la tecnologia mobile viene utilizzata come strumento per la navigazione: integrando i dati GPS con quelli derivanti da un magnetometro, queste applicazioni forniscono importanti dati direzionali; l'utente deve solo puntare il dispositivo in una specifica direzione per ottenere informazioni sullo spazio topografico selezionato [10]. *Ariadne GPS*<sup>6</sup> è un'applicazione iOS che ha ricevuto un encomio dalla Comunità dei Ciechi e degli Ipovedenti. Ariadne GPS utilizza "mappe parlanti" che consentono alla persona non vedente di esplorare lo spazio circostante toccando con le dita la mappa sullo schermo del dispositivo. Se VoiceOver è attivo, questa applicazione legge i nomi delle strade e i numeri civici delle case della zona in cui si trova l'utente; le strisce pedonali invece vengono segnalate da una vibrazione. Ariadne GPS è utilizzabile solo dove GoogleMaps è disponibile.

La maggior parte di queste soluzioni prevedono comunque che l'utente, per ricevere le informazioni di cui ha bisogno, debba a) spostarsi all'interno dell'ambiente e b) indossare/tenere in mano un sensore [11]. A differenza dei sistemi che abbiamo brevemente illustrato in questo paragrafo, SoundingARM permette all'utente di esplorare un interno rimanendo sulla soglia ed eseguendo il gesto dell'indicazione senza dover indossare o tenere in mano dei sensori. Eseguendo queste due semplici attività, la persona non vedente può velocemente costruire una mappa spaziale per: a) capire immediatamente di che tipo di ambiente si tratta, b) localizzare un oggetto o c) intraprendere con maggiore sicurezza l'esplorazione dello spazio interno.

### 3. SOUNDINGARM (RAPPRESENTAZIONE ACUSTICA DI UNA MAPPA SPAZIALE)

Il sistema SoundingARM, basato sul riconoscimento della posizione e del movimento dell'utente, è finalizzato ad assistere le persone con gravi disabilità visive nell'esecuzione di attività di orientamento e mobilità in spazi interni noti o sconosciuti. La principale finalità del sistema consiste nel risparmiare all'utente, se non necessario, l'esplorazione locomotoria e tattile dell'ambiente fornendo acusticamente informazioni topografiche essenziali: a volte infatti la persona non vedente necessita solamente di sapere di che tipo di ambiente si tratta e/o se in quella stanza c'è l'oggetto che sta cercando.

Rimanendo sulla soglia e indicando con il braccio in una direzione dello spazio, l'utente non vedente riceve in tempo reale l'informazione acustica dell'oggetto posto in quella direzione. In questo modo, l'utente ha la possibilità di costruire la mappa spaziale di tutte le caratteristiche essenziali dell'ambiente.

<sup>6</sup> <http://www.ariadnegps.eu/>

#### 3.1 Obiettivi

SoundingARM può essere utilizzato con i seguenti obiettivi:

- Facilitare la re-integrazione negli ambienti privati dell'utente che ha acquisito la cecità durante l'adolescenza/età adulta.
- Promuovere la conversione della rappresentazione visiva dello spazio nella rappresentazione di uno spazio di azione.
- Incoraggiare l'evoluzione dalla condizione di totale dipendenza attraverso l'acquisizione delle competenze di orientamento e mobilità nello spazio.
- Facilitare la decodifica delle informazioni ambientali e la loro organizzazione in un sistema di riferimento spaziale.
- Promuovere l'acquisizione di informazioni multi-sensoriali sull'ambiente.

#### 3.2 Requisiti di Sistema

I requisiti di sistema di SoundingARM sono elencati di seguito.

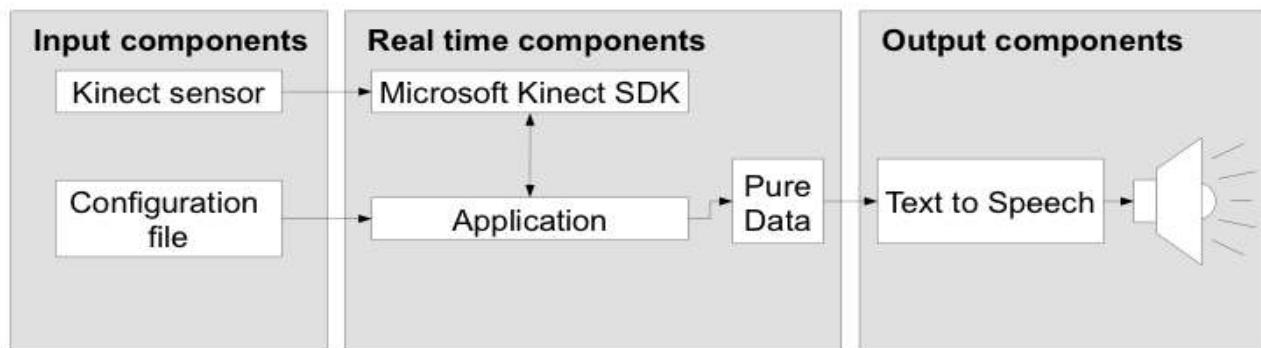
##### 3.2.1 Hardware

1. Il sensore Kinect della Microsoft per il tracciamento della posizione delle braccia e della testa dell'utente.
2. Un personal computer con i seguenti requisiti:
  - Sistema operativo Windows 7, Windows 8 Consumer Preview, Windows Embedded Standard 7 o Windows Embedded POSReady 7.
  - Processore a 32 bit (x86) o a 64 bit (x64).
  - Processore dual core a 2,66-GHz o più veloce.
  - Bus USB 2.0 dedicato.
  - 2 GB di RAM.
3. Un cavo USB.
4. Due diffusori acustici.

##### 3.2.2 Software

Di seguito sono riportati i principali elementi software utilizzati da SoundingARM.

1. Il pacchetto SDK (Software Development Kit) 1.0 per Windows utile all'acquisizione dei dati mediante il sensore Microsoft Kinect; questo software permette di sviluppare applicazioni basate sul riconoscimento del gesto e della voce.
2. Un software (Applicazione), sviluppato in C++, per la gestione del protocollo di comunicazione con il sensore Kinect e i dati relativi a: a) la posizione dell'utente, b) le misure della stanza e c) la posizione degli oggetti.



**Figura 2.** L'architettura di sistema di SoundingARM.

3. Un software sviluppato in Pure Data (Patch Pure Data) per l'implementazione del feedback acustico. La versione base di SoundingARM impiega il sintetizzatore vocale di Windows ma la flessibilità del sistema consente comunque di utilizzare anche altre tipologie di feedback acustico.

#### 4. L'ARCHITETTURA DEL SISTEMA

La Fig 2 descrive l'architettura di sistema di SoundingARM.

- Le componenti di ingresso: a) il sensore Kinect della Microsoft connesso al computer (su cui è installato Windows 7) mediante una porta USB e b) il File di Configurazione che contiene le informazioni relative alle misure e all'arredamento della stanza; questo file viene caricato all'avvio dell'applicazione.
- Le componenti di elaborazione in tempo reale dei dati: a) il Software Development Kit (SDK) per il sensore Kinect della Microsoft utile al riconoscimento in tempo reale dello scheletro, b) l'Applicazione che analizza lo scheletro dell'utente ricavando dati sulla direzione dell'indicazione e c) la Patch Pure Data che fa da intermediario tra l'Applicazione e il server della sintesi vocale.
- Le componenti multimediali di uscita: il server della sintesi vocale che, utilizzando il Microsoft Speech API Engine (un pacchetto completo per la sintesi e il riconoscimento vocali), attraverso due diffusori acustici produce il feedback audio di ciò che l'utente sta indicando.

##### 4.1 Le Componenti di Ingresso

Il sensore Microsoft Kinect è uno scanner 3D che può essere installato su un computer Windows 7 grazie al Microsoft Kinect SDK. I driver sono progettati in modo da essere in grado di riconoscere la forma del corpo umano all'interno di un ambiente e fornirne una ricostruzione digitale, sotto forma di scheletro, basandosi sui dati di profondità rilevati. Lo scheletro è costituito da un insieme di punti a coordinate tridimensionali; ogni punto descrive le articolazioni del corpo umano e il sensore Kinect costituisce l'origine degli assi.

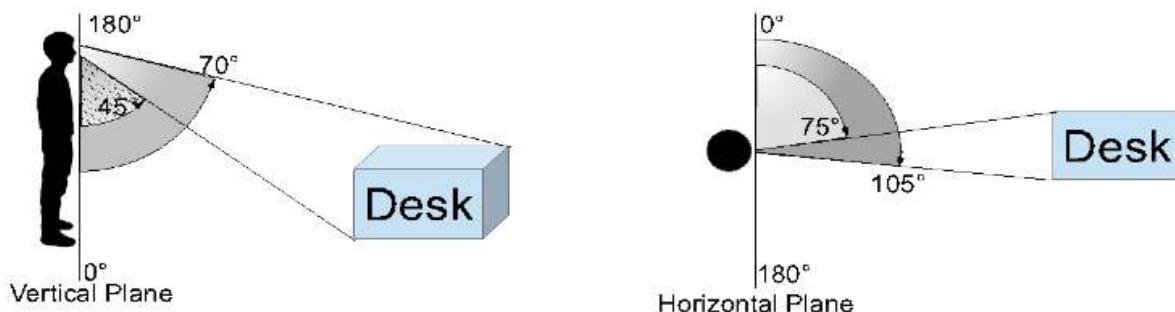
Il File di Configurazione contiene le informazioni sull'arredamento e sugli oggetti della stanza che il sistema "vede" come figure solide. Per esempio, una scrivania è definita dal parallelepipedo che la contiene; questo parallelepipedo viene elaborato utilizzando i valori di quattro punti della base, le misure dell'altezza della scrivania e l'altezza della scrivania dal pavimento. I punti della base vengono calcolati in riferimento alla posizione del sensore Kinect. Il File di Configurazione descrive tutti gli oggetti statici della stanza (scrivanie, tavoli, armadi, divani) che possono essere di ostacolo alla deambulazione. Gli oggetti mobili (come le sedie) non vengono mappati: per ora Soundin-gARM non è progettato per la rilevazione delle variazioni dell'ambiente.

##### 4.2 Le Componenti di Elaborazione in Tempo Reale dei Dati in Ingresso

Il software di sviluppo SDK contiene driver e API che consentono all'Applicazione di controllare e ottenere dati dal sensore Kinect. Il cuore del sistema è costituito dall'Applicazione che, all'avvio, analizza i dati relativi all'ambiente forniti dal File di Configurazione. L'Applicazione a) rileva la presenza di una persona, b) ricostruisce il piano prospettico della stanza - calcolato in funzione della posizione della testa - e c) verifica se l'utente sta indicando con il braccio in qualche direzione. Se effettivamente l'utente sta puntando il braccio in una direzione, l'Applicazione invia alla Patch Pure Data il nome (in formato *raw text*) dell'oggetto ivi posizionato. La Patch Pure Data ha due principali funzioni: a) creare l'interfaccia grafica per facilitare l'operatore informatico che installerà il sistema e b) interfacciare l'Applicazione con il *Microsoft Text-to-Speech server* utilizzando il protocollo OSC.

##### 4.3 Le Componenti di Uscita

La componente software di uscita è composta dal sintetizzatore vocale che utilizza le librerie *Microsoft Speech API* fornite da Windows 7. La componente hardware è invece costituita da due diffusori acustici che riproducono il nome dell'oggetto indicato.



**Figura 3.** Una persona è posizionata davanti ad una scrivania: la rappresentazione del punto di vista laterale è a destra mentre la rappresentazione del punto di vista dall’alto è a sinistra. Per l’utente la scrivania è posizionata in un angolo la cui ampiezza è in funzione della sua posizione rispetto la posizione della scrivania. Questo angolo può essere individuato mediante componenti verticali e orizzontali. Il piano verticale e il piano orizzontale sono piani prospettici immaginari che definiscono l’orientamento della testa dell’utente.

#### 4.4 L’Applicazione

Il kernel del sistema è costituito da un’*Application Server* che in background comunica con la Patch Pure Data utilizzando il protocollo OSC. In fase di caricamento, l’Applicazione riceve dal File di Configurazione i dati relativi all’ambiente da inserire in una struttura personalizzata di sola lettura. Poi l’Applicazione riceve ed elabora i dati della posizione dell’utente al fine di calcolarne la vista prospettica. La posizione dell’utente viene calcolata per mezzo dei dati relativi alla posizione della testa forniti dal sensore Kinect. Calcolata la posizione della testa, l’applicazione computa gli angoli che l’utente deve eseguire con le sue braccia per indicare un oggetto. Per ogni oggetto si ottiene così un insieme di dati: l’angolo minimo/massimo sul piano orizzontale e l’angolo minimo/massimo sul piano verticale. Questi dati vengono inseriti in una matrice dove gli indici rappresentano gli angoli mentre le celle corrispondenti contengono l’informazione relativa all’oggetto. In figura 3 una persona è posizionata davanti a una scrivania. Rispetto a lui la scrivania è situata a min 75° - max 105° sul piano orizzontale e a min 45° - max 70° sul piano verticale; in questo caso, l’Applicazione riempie le celle della matrice bidimensionale con l’ID della scrivania nell’intervallo ([75:105], [45:70]). Questa operazione viene ripetuta per ogni oggetto presente nel File di Configurazione: l’ID dell’oggetto è unico e consiste in un numero che cresce in relazione alla posizione dell’oggetto nell’elenco del File di Configurazione. Il risultato è una matrice bidimensionale di 180 x 180 gradi che descrive la vista prospettica della stanza. La matrice viene aggiornata continuamente quando l’utente si muove per più di 10.0 centimetri in ogni direzione, un compromesso tra la precisione del sensore Kinect e i costi di ricalcolo.

Il principale scopo dell’Applicazione consiste nell’identificare ciò che l’utente sta indicando; per ottenere quest’informazione l’Applicazione: a) considera un punto della testa e un punto della mano, poi, b) calcola il vettore

tra questi due punti. Questo vettore ha due componenti angolari:

1. quella orizzontale, la cui grandezza varia da sinistra a destra a seconda del punto di vista dell’utente rispetto il sensore Kinect;
2. quella verticale la cui grandezza varia dal basso verso l’alto.

Gli angoli orizzontale e verticale sono utilizzati come indici della matrice bidimensionale che restituisce il valore dell’oggetto indicato.

Supponendo che l’utente stia indicando la scrivania, come nell’esempio di Fig. 3, il vettore immaginario che unisce la testa e la mano sinistra (o destra) deve creare un angolo con una grandezza compresa tra il 75 e i 105 gradi in relazione al piano orizzontale e fra i 45 e i 70 gradi in relazione al piano verticale; in questo caso la richiesta “(80, 50)” restituirà l’ID della scrivania. Infine, l’ID viene utilizzato per recuperare il nome dell’oggetto selezionato che viene inviato al sintetizzatore vocale dalla Patch Pure Data.

#### 4.5 La Patch Pure Data

L’Applicazione, utilizzando il protocollo OSC, trasmette i dati alla Patch Pure Data che offre grande flessibilità operativa grazie alla versatilità e modularità della piattaforma su cui si basa. Il protocollo OSC comunque non preclude l’utilizzo di altri tipi di linguaggi.

Attualmente, la Patch Pure Data ha il compito principale di inviare il nome dell’oggetto selezionato dall’utente al sintetizzatore vocale ma, nel prossimo futuro, è previsto che possa gestire anche la spazializzazione del suono. La Patch Pure Data e l’Applicazione sono già predisposte per questa modifica, infatti, l’Applicazione è in grado di inviare alla Patch Pure Data tutte le informazioni sull’oggetto selezionato (posizione e dimensioni) e la posizione dell’utente. Questi dati sono sufficienti per consentire la riproduzione di effetti audio 3D.

La Patch Pure Data fornisce inoltre un'interfaccia di controllo che l'operatore informatico utilizzerà per la gestione dell'*Application Server*; nel dettaglio, l'interfaccia di controllo consente: a) l'avvio dell'Applicazione e della sintesi vocale, b) l'attivazione/disattivazione del riconoscimento dello scheletro ed, eventualmente, c) la chiusura del programma. Per avviare i file di esecuzione, la Patch utilizza *Createprocess*, una specifica funzione esterna che consente di eseguire in background un'applicazione Windows 7 evitando l'apertura di finestre di terminale.

## 5. TEST DI USABILITÀ

SoundingARM è stato sottoposto ad un test di usabilità semplificato allo scopo di testare il livello di performance del sistema. Con questa finalità SoundingARM, il 6 marzo 2012, è stato installato nella cucina dell'Istituto Regionale "Rittmeyer" per i Ciechi di Trieste (Italy) ed è stato testato da dieci utenti adulti con diversi livelli di disabilità visiva e con deficit cognitivi/motori associati. Gli utenti non avevano avuto precedente esperienza dell'ambiente designato. Il test di usabilità (informale, qualitativo) ha fornito indicazioni sugli elementi che costituiscono un ostacolo al corretto utilizzo del sistema da parte di questo gruppo eterogeneo (per tipologia di disabilità) di utenti.

Prima di iniziare il test, gli utenti hanno ricevuto alcune istruzioni sulle principali modalità di utilizzo del sistema in modo da familiarizzare con i compiti da eseguire. Ad ognuno è stato chiesto di: a) entrare nella stanza, b) fermarsi sulla soglia della cucina, c) e indicare con il braccio teso in varie direzioni. Gli oggetti da identificare erano dodici: tavolo, frigorifero, cappa aspirante, scolapiatti, pensile, fornello, lavandino, interruttore della luce, due termosifoni e due finestre. I partecipanti al test non conoscevano il numero di oggetti da identificare e potevano effettuare l'esplorazione per mezzo del sistema SoundingARM senza limiti di tempo. L'osservazione diretta dei partecipanti al test ha consentito di raccogliere delle importanti considerazioni riguardo il loro approccio a SoundingARM (azioni, difficoltà e commenti) e le prestazioni del sistema con soggetti molto diversi.

Anche se tutti gli utenti hanno identificato la maggior parte degli oggetti della cucina, si è potuto osservare che:

1. Al primo utilizzo del sistema gli utenti con cecità e disabilità cognitive associate necessitano: (1) di ulteriori spiegazioni orali in aggiunta a quelle previste; (2) di una guida che fisicamente li assista nella corretta esecuzione del gesto dell'indicazione; (3) di un segnale tattile a pavimento che li aiuti a mantenere la posizione sulla soglia (tendono infatti a entrare nella stanza per effettuare l'esplorazione locomotorio/tattile, come sono abituati a fare). L'intervento dell'operatore esterno (2) nella fase di familiarizzazione ha interferito con le operazioni di riconoscimento dello scheletro causando errori di tracciamento del gesto.
2. Le persone non vedenti con disabilità motorie associate che utilizzavano la sedia a rotelle non hanno

ricevuto il feedback acustico di tutti gli oggetti presenti; ciò è accaduto molto probabilmente perché a) la sedia a rotelle ha costituito un ostacolo al riconoscimento dello scheletro e b) il punto di riferimento della testa stava quasi sullo stesso piano del sensore Kinect della Microsoft.

3. Gli utenti con cecità senza disabilità associate hanno facilmente identificato tutti gli oggetti della stanza.

Generalmente, le persone con cecità non sono abituate a localizzare un oggetto (sonoro/non sonoro) mediante l'indicazione che è un'azione maggiormente correlata alla senso della vista più che all'udito o al tatto. Le persone non vedenti, quando devono raggiungere un oggetto, anche sonoro, di solito lo cercano a tentoni invece di raggiungerlo direttamente [12]. L'abilità di raggiungere direttamente un punto specifico dello spazio è molto difficile da acquisire senza punti di riferimento visivi. Lo sviluppo di questa competenza si ottiene attraverso la costruzione della mappa cognitiva dell'ambiente [3] che, in questo contesto, viene agevolata dall'esperienza acustica fornita da sistema SoundingARM.

Nella comunicazione interpersonale il gesto dell'indicazione è un'abilità motoria molto importante che, se non naturalmente presente, viene insegnata nell'ambito dei corsi di mobilità e orientamento. Da questo punto di vista SoundingARM può essere uno strumento molto utile per l'acquisizione/miglioramento di questa competenza.

## 6. CONCLUSIONI E SVILUPPI FUTURI

SoundingARM è un sistema che permette alla persona con grave disabilità visiva di familiarizzare rapidamente con un ambiente interno senza dover fare grandi spostamenti: dalla soglia della stanza, indicando con il braccio un punto dello spazio, l'utente riceve in tempo reale l'informazione acustica relativa all'oggetto posto in quella direzione. Con questa modalità, il soggetto ha la possibilità di costruirsi la mappa spaziale delle caratteristiche fisse dell'ambiente. Questa mappa spaziale può essere utilizzata per: a) riconoscere immediatamente la tipologia di ambiente interno (cucina, salotto, sala da pranzo, ecc.), b) muoversi con maggiore sicurezza nello spazio, e/o c) localizzare velocemente un oggetto.

SoundingARM, offrendo affordance spaziali attraverso il suono (house-hold sounds), permette all'utente di "scoprire/riscoprire acusticamente" le principali caratteristiche di un ambiente attraverso l'esecuzione di semplici e limitati movimenti. Le attività di orientamento e mobilità vengono facilitate dall'informazione verbale sul posizionamento degli oggetti fornita dal sistema. Nello stesso tempo, l'utente utilizza le informazioni motorie derivanti dalla posizione del braccio nel gesto dell'indicazione (propriocezione) per ottenere le coordinate spaziali degli oggetti.

Il sistema necessita di essere migliorato soprattutto per soddisfare i bisogni delle persone non vedenti con associate disabilità motorie che utilizzano la sedia a rotelle. Il sistema SoundingARM sarà inoltre implementato con la funzione di spazializzazione degli eventi sonori al fine di offrire all'utente affordance acustiche di localizzazione mag-

giamente informative (distanza degli oggetti). Con la sperimentazione prevista nel prossimo futuro, si intende verificare: a) se con stimoli fissi e senza interazione l'utente acquisisce una rappresentazione mentale dell'ambiente e una conoscenza dello spazio efficaci e b) se e come l'interazione favorisce l'acquisizione della rappresentazione spaziale.

## 7. RICONOSCIMENTI

Questo lavoro è stato parzialmente supportato dall'Istituto Regionale "Rittmeyer" per i Ciechi di Trieste (Ts - Italia) attraverso il Progetto "Evoluzione dei sistemi di feedback acustico per la prima domiciliazione dei soggetti con inabilità visiva acquisita" finanziato dalla Regione Autonoma Friuli Venezia Giulia con decreto n. 1265/AREF dd. 25 novembre 2010.

## 8. BIBLIOGRAFIA

- [1] S. Zanolla, F. Romano, F. Scattolin, A. Rodà, S. Canazza, and G. L. Foresti, "When sound teaches," in *Proceedings of the SMC 2011 - 8th Sound and Music Computing Conference* (S. Zanolla, F. Avanzini, S. Canazza, and A. de Götzen, eds.), pp. 64–69, 2011.
- [2] C. Thinus-Blanc and F. Gaunet, "Representation of space in blind persons: Vision as a spatial sense?," *Psychological Bulletin*, vol. 121, no. 1, pp. 20–42, 1997.
- [3] Y. Hatwell, *Psicologia cognitiva della cecità precoce*. Biblioteca Italiana per i Ciechi Regina Margherita - ONLUS, 2010.
- [4] E. W. Hill and J. J. Rieser, "How persons with visual impairments explore novel spaces: strategies of good and poor performers.," *Journal of Visual Impairment and Blindness*, vol. 87, p. 295, October 1993.
- [5] R. W. Massof, "Auditory assistive devices for the blind," in *Proceedings of the International Conference on Auditory Display*, (Boston, MA, USA), pp. 271–275, July 2003.
- [6] Y. Sonnenblick, "An indoor navigation system for blind individuals," in *Proceedings of the 13th annual Conference on Technology and Persons with Disabilities*, 1998.
- [7] L. Ran, S. Helal, and S. Moore, "Drishti: An integrated indoor/outdoor blind navigation system and service," in *IEEE International Conference on Pervasive Computing and Communications*, 2004.
- [8] S. Mau, N. A. Melchior, M. Makatchev, and A. Steinfeld, *BlindAid: An Electronic Travel Aid for the Blind*. 2008.
- [9] A. Hub, J. Diepstraten, and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," *SIGACCESS Access. Comput.*, no. 77-78, pp. 147–152, 2004.
- [10] C. Magnusson, M. Molina, K. Rassmus-Gröhn, and D. Szymczak, "Pointing for non-visual orientation and navigation," in *Proceeding of NordiCHI, Reykjavik, Iceland*, October 2010.
- [11] A. Fusiello, A. Panuccio, V. Murino, F. Fontana, and D. Rocchesso, "A multimodal electronic travel aid device," in *Proceedings of ICMI*, pp. 39–46, 2002.
- [12] M. C. Wanet and C. Veraart, "Processing of auditory information by the blind in spatial localization tasks," *Perception and Psychophysics*, no. 38, pp. 91–96, 1985.
- [13] S. Zanolla, A. Rodà, F. Romano, F. Scattolin, G. L. Foresti, S. Canazza, C. Canepa, P. Coletta, and G. Volpe, "Teaching by means of a technologically augmented environment: the Stanza Logo-Motoria," in *Proceedings of INTETAIN 2011 Conference*, 2011.



# MEMORY SONORO REALTÀ AUMENTATA ACCESSIBILE MEDIANTE AUDIO DIREZIONALE

Serena Zanolla, Giorgio Tempesta, Lauro Snidaro

Università degli Studi di Udine

serena.zanolla@uniud.it

tempesta.giorgio@spes.uniud.it

lauro.snidaro@uniud.it

Sergio Canazza

Università degli Studi di Padova

canazza@dei.unipd.it

## SOMMARIO

I recenti sviluppi delle tecnologie utilizzate nei sistemi di Realtà Aumentata permettono di realizzare applicazioni interattive e multimodali basate su interfacce che ne estendono l'accessibilità agli utenti con disabilità sensoriali. Il contenuto virtuale può infatti compensare le situazioni in cui la percezione non è completa, come nel caso delle persone con disabilità visiva. Parallelamente, la grande disponibilità di software in grado di elaborare il segnale audio in tempo reale rende possibile l'utilizzo congiunto di tecniche di spazializzazione del suono. In questo articolo si intende illustrare come queste tecnologie siano state adottate nella realizzazione del Memory Sonoro, un sistema interattivo multimodale accessibile a utenti affetti da cecità o ipovisione perché permette di svolgere il gioco di memoria "Memory" a livello tattile e acustico. Nel Memory Sonoro la modalità di interazione classica, basata esclusivamente sul canale visivo, è stata sostituita con a) la fruizione acustica (la riproduzione di suoni spazializzati) e b) la fruizione tattile (utilizzo di interfacce tattili). L'articolo documenta anche le prime osservazioni derivanti dal test di usabilità somministrato ad un target eterogeneo di utenti con disabilità visive gravi.

## 1. INTRODUZIONE

Le applicazioni di Realtà Aumentata ("Augmented Reality", abbreviato AR) sovrappongono alla scena reale informazioni virtuali supplementari che tradizionalmente sono costituite da scritte, simboli o immagini generate al computer attraverso tecniche di *compositing video* o la proiezione su superfici trasparenti. Gli oggetti virtuali, che vengono sovrapposti alla realtà attraverso dispositivi mobili, pc dotati di webcam (o altri sensori) e dispositivi di visione (occhiali), di ascolto (auricolari) e di manipolazione (guanti), veicolano informazioni aggiuntive supportando l'utente nella fruizione di contenuti altrimenti non accessibili. L'AR può dunque essere considerata un esempio di ciò che Fred Brooks definisce *Intelligence Amplification* [1].

Copyright: ©2012 Serena Zanolla 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.

La ricerca sull'AR si è focalizzata principalmente su componenti "grafiche", sovrapponendo immagini reali a immagini virtuali ma, de facto, la Realtà Aumentata può avvalersi anche di altri canali sensoriali (in questo caso tattile ed uditorio).

Il Memory Sonoro, descritto in questo articolo, è un sistema interattivo multimodale che estende le tecniche dell'AR al dominio dell'audio direzionale. Il Memory Sonoro utilizza infatti le tecnologie di tracking ottico e le tecniche della spazializzazione audio in tempo reale attraverso diffusori acustici per realizzare un'interfaccia acustico-tattile, con valenza sia ludica sia educativa, accessibile a utenti con ipovisione o cecità. L'adozione di queste tecnologie permette agli utenti di: a) interfacciare il sistema tramite oggetti reali (tessere munite di marker) e b) agire liberamente (senza indossare sensori) all'interno dello spazio sonoro tridimensionale.

Le persone non vedenti, per sopperire alla mancanza della vista, sviluppano maggiore sensibilità negli altri sensi: l'udito e il tatto, in particolare, diventano i sensi principali per l'orientamento e la mobilità all'interno dell'ambiente [2]. In questo modo, i non vedenti mettono in atto un processo di riequilibratura plurisensoriale valorizzando il potenziale esistente e compensando ciò che è definitivamente perduto [3].

L'udito è un sistema telerecettore che apporta informazioni utili alla localizzazione delle sorgenti sonore ed ha una notevole importanza per l'orientamento spaziale perché permette alla persona con cecità di individuare la posizione degli oggetti nello spazio. La percezione acustica della direzione e della distanza di una fonte sonora, grazie alla resonanza delle pareti, consente di avere informazioni sulla grandezza e la forma della stanza e sulla presenza di altre persone. Il tatto invece, recettore di contatto con un campo percettivo molto ristretto, permette di conoscere la forma, la grandezza, la localizzazione spaziale, la distanza, il peso, la temperatura degli oggetti esplorati attraverso la manipolazione [3].

La capacità di orientarsi e muoversi nello spazio però è il risultato di una specifica educazione sensoriale che permette di sfruttare a pieno le stimolazioni sensoriali e di servirsi per esplorare l'ambiente, muoversi con sicurezza, cogliere le posizioni e i rapporti spaziali delle cose. Le sensazioni tattile-uditiva devono essere esercitate per favorire la costruzione di immagini mentali, la percezione degli ostacoli

coli, l'orientamento nello spazio, la capacità di individuare punti di riferimento. La perdita della vista in età precoce o adulta è una variabile importante, con conseguenze funzionali specifiche. Le persone che hanno perso la vista in età precoce sono allenate a rappresentarsi mentalmente gli spazi con punti di riferimento, prestano molta attenzione ai suoni, agli odori, alle vibrazioni, all'eco della voce, più che alle informazioni riferite da un'accompagnatore. Coloro che, invece, hanno perso la vista in età adulta, trovano molta difficoltà nell'attuare il processo di riequilibrizzazione del potenziale esistente avvalendosi dei sensi vicari [4].

Il Memory Sonoro trova applicazione proprio in questo contesto ponendosi come sussidio tiflodidattico per lo sviluppo e l'esercizio dei sensi vicari della vista e nella riduzione degli effetti secondari della cecità. In particolare, il Memory Sonoro può essere utilizzato per:

1. sviluppare l'integrazione percettiva, tattile e uditiva;
2. identificare e analizzare un oggetto e i suoi dettagli;
3. apprendere delle strategie esplorative e di orientamento nello spazio;
4. denominare e descrivere i suoni;
5. verbalizzare i processi logici e le strategie operative e spaziali;
6. abbinare per collegamento semantico (suono/oggetto);
7. potenziare le capacità attentive e mnemoniche.

## 2. LA REALTÀ MEDIATA DALL'ELABORATORE (LA REALTÀ AUMENTATA)

Le tecniche di AR, potendo essere facilmente applicate a diverse piattaforme tecnologiche di supporto, hanno ottenuto ampia diffusione in molti ambiti di vita. PC dotati di webcam, cellulari di ultima generazione (iPhone o Android) e Tablet sono tutti dispositivi in grado di supportare tecnologie di AR. La nascita dell'AR coincide con la ricerca di Sutherland [5] che realizza il primo sistema binoculare di tipo "see-through" (l'HMD, *Head-Mounted Display*). E' però Thomas Caudell a coniare nel 1989 il termine AR per un sistema che permette di progettare diagrammi di cablaggio dei velivoli senza il bisogno di *wiring board* o di schemi cartacei [6]. In campo medico, Bajura [7] realizza un sistema di AR che permette di visualizzare, grazie ai dati acquisiti tramite le analisi con ultrasuoni, le immagini del feto sopra al ventre di una futura madre. Nel campo della progettazione architettonica i sistemi di AR possono agevolare i processi di costruzione, rinnovo e ristrutturazione di opere edili [8]. Liarokapis utilizza l'AR per l'insegnamento della chitarra [9] e per offrire un'esperienza museale aumentata alle persone con disabilità [10]: l'interfaccia costituita, come nel Memory Sonoro, da tessere con marker, permette di scegliere il museo di cui si vogliono vedere le opere in 3D e il tipo di illuminazione delle stanze.

L'utilizzo multi-utente dell'AR è stato realizzato dalla Vienna University of Technology [11] con "Studierstube", un sistema di AR che, mediante l'utilizzo di un HMD, permette di sovrapporre dei *dataset* virtuali agli oggetti reali

senza limitare, anzi favorendo, l'interazione naturale fra le persone e con l'ambiente.

Il sensore Kinect della Microsoft è già stato ampiamente utilizzato in molti sistemi di AR come *KinectShop*<sup>1</sup>, un vero e proprio negozio interattivo dove l'utente, assistito dalla periferica Kinect, può provare virtualmente qualsiasi capo di abbigliamento oppure *FaceOSC*<sup>2</sup> che, attraverso il tracciamento del volto (*face tracking*), consente di controllare vari parametri di un software di editing audio. Considerando le applicazioni di AR non visiva, ci soffermiamo in particolare sulla *Audio-Based AR*, particolarmente adatta a fornire informazioni sull'ambiente alle persone con disabilità sensoriali come la cecità o l'ipovisione. I sistemi di AR rivolti alle persone con disabilità visive non devono ovviamente fornire informazioni virtuali grafiche ma possono utilizzare informazioni acustiche e/o tattili. Di norma, lo scopo principale di questo tipo di applicazioni consiste nel supportare i non vedenti nella deambulazione autonoma in spazi interni ed esterni. NAVI (*Navigational Aids for the Visual Impaired*) [12] è un sistema di AR che utilizza il sensore Kinect della Microsoft installato su un casco e una cintura vibro-tattile per fornire informazioni spaziali e indicare la presenza di ostacoli. NAVI, grazie ai marker posizionati nell'ambiente, fornisce anche istruzioni vocali sui diversi elementi/oggetti presenti. Il *Multimodal Travel Aid System* [13] invece, progettato per agevolare la mobilità dei non vedenti, utilizza un puntatore laser, occhiali con microcamere, un palmtop computer e un paio di cuffie. Il puntatore laser viene usato come se fosse un bastone per non vedenti: il sistema rileva i punti laser sugli oggetti e ne fornisce la descrizione spaziale (distanza e posizione) attraverso segnali acustici. Il concetto di "vedere attraverso il suono" è applicato anche al progetto "*The vOICE*"<sup>3</sup>, un sistema che realizza un'esperienza di audio-based AR per persone con cecità assoluta attraverso una tecnologia che trasforma le immagini in informazioni acustiche.

Interessante, artistico e assolutamente innovativo esempio di Audio-Based AR è "*The National Mall*", il primo album *location-aware* pubblicato dal duo *Bluebrain* sotto forma di app gratuita per iPhone; utilizzando le funzionalità del GPS, la musica, ascoltata all'interno dei confini del parco National Mall, evolve secondo la posizione topografica dell'utente che avrà così un'esperienza musicale unica. Questa esperienza, con altri album, è stata ripetuta al Central Park (New York City), lungo la Sesta Strada ad Austin (Texas) e sarà realizzata anche per la Pacific Coast Highway in California<sup>4</sup>. Questo dimostra che anche i dati di localizzazione possono essere usati per creare esperienze multimediali.

Il primo gioco di memoria, Memory, progettato per persone con disabilità visive è stato realizzato da [14]: si tratta di una versione per pc in cui il piano di gioco e le carte sono visualizzate sullo schermo del computer. L'utente

<sup>1</sup> <http://emergingexperiences.com/2011/06/kinectshop>

<sup>2</sup> <https://github.com/kylemcdonald/ofxFaceTracker>

<sup>3</sup> <http://www.seeingwithsound.com/>

<sup>4</sup> <http://www.bluebra.in/>

sceglie la carta: a) attraverso comandi verbali (che il sistema riconosce attraverso un *speech recognition software* o b) digitando le coordinate sulla tastiera. Il sistema risponde fornendo il nome dell'oggetto illustrato sulla carta selezionata. Interessante la soluzione adottata per rappresentare acusticamente la posizione delle carte sul piano di gioco: la posizione orizzontale è ottenuta attraverso la presentazione stereo degli eventi sonori mentre la posizione verticale è indicata mediante la variazione del *pitch* della sintesi vocale.

Il Memory Sonoro, il sistema presentato in questo articolo, fa parte della categoria dei sistemi Audio-Based AR, infatti, ha l'obiettivo di estendere il dominio dell'AR al suono. Nel Memory Sonoro la posizione degli oggetti nello spazio viene utilizzata per "aumentare" la scena sonora e per modificare in tempo reale la percezione del suono. Il Memory Sonoro si basa su un sistema di tracking ottico che utilizza unicamente processi di analisi dell'immagine per ricavare informazioni sugli oggetti presenti sulla scena. Il tracking ottico è abbinato all'utilizzo di marker (ARtags), oggetti bidimensionali su cui è stampata un'immagine che può essere riconosciuta da appositi algoritmi di *image processing*. Questa soluzione è facilmente implementabile perché non necessita di grandi risorse a livello hardware. I marker inoltre permettono di utilizzare anche il canale tattile e, al contrario di altri sistemi basati sul tracking ottico con *feature extraction*, dimostrano buone prestazioni anche in diverse condizioni di illuminazione.

### 3. MEMORY SONORO

Il Memory Sonoro è un sistema di AR che utilizza il suono per fornire informazioni altrimenti inaccessibili a persone con cecità o ipovisione [4]. Si basa sulle regole di un gioco di memoria molto semplice e conosciuto (il Memory) che consiste nel trovare coppie di figure uguali. Le coppie di immagini del Memory classico sono sostituite da coppie di marker (collocati su delle tessere) che il sistema riconosce e associa a coppie di suoni. Nel Memory Sonoro i giocatori devono quindi trovare coppie di suoni uguali (fig. 1). Inoltre, i suoni possono essere spazializzati in tempo reale se l'utente muove le tessere sul piano orizzontale e sul piano frontale.

Lo scopo principale del sistema consiste nel permettere alle persone con gravi disabilità visive di dedicarsi a un tipo di attività che di solito utilizza principalmente il canale visivo sviluppando, attraverso la gestione delle tessere, le abilità tattili e acustiche fondamentali per l'acquisizione dell'autonomia personale.

#### 3.1 Requisiti di Sistema

Il Memory Sonoro è costituito dalle seguenti componenti hardware:

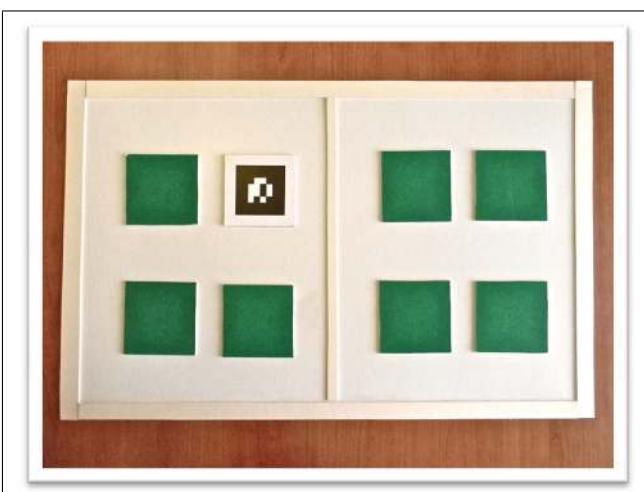
1. La webcam (fig. 1), posizionata sopra l'utente su uno stativo a circa un metro dal piano del tavolo su cui si svolge il gioco in modo che la direzione dell'inquadratura sia perpendicolare al pavimento: questa disposizione a) permette ampia libertà di movimento all'utente e b) consente prestazioni di tracking



**Figura 1.** Un bambino con cecità assoluta dalla nascita gioca con il Memory Sonoro.

di alto livello (si evitano distorsioni nel tracciamento della posizione dei marker).

2. Le tessere (fig.2) sono l'elemento principale attraverso cui gli utenti interagiscono con il sistema. Sono costituite da un supporto in cartoncino di dimensioni 8 cm x 8 cm: il verso ha una superficie liscia su cui è collocato un marker cartaceo; il dorso è invece costituito da una superficie ruvida o vellutata. Questo accorgimento è necessario affinché l'utente sappia sempre se le tessere sono coperte o girate.
3. L'area di gioco dev'essere delimitata chiaramente in modo che i giocatori abbiano l'esatta percezione delle dimensioni dell'area di azione. L'area di gioco è costituita da un supporto in cartoncino (58 cm x 24 cm) con i bordi e la metà in rilievo (fig.2). Queste dimensioni garantiscono che i marker restino sempre all'interno del campo di ripresa della webcam e del tracker.



**Figura 2.** Le tessere, i marker e il piano di gioco.



**Figura 3.** L'architettura di sistema del Memory Sonoro.

4. Tutte le operazioni a livello software sono state eseguite da un MacBook con sistema operativo Mac OS X, versione 10.8.6.
5. Scheda audio.
6. 6 diffusori acustici disposti secondo la configurazione del Dolby Surround 5.1.

#### 4. L'ARCHITETTURA DI SISTEMA

La figura 3 descrive l'architettura di sistema del Memory Sonoro.

##### 4.1 Le componenti di Ingresso.

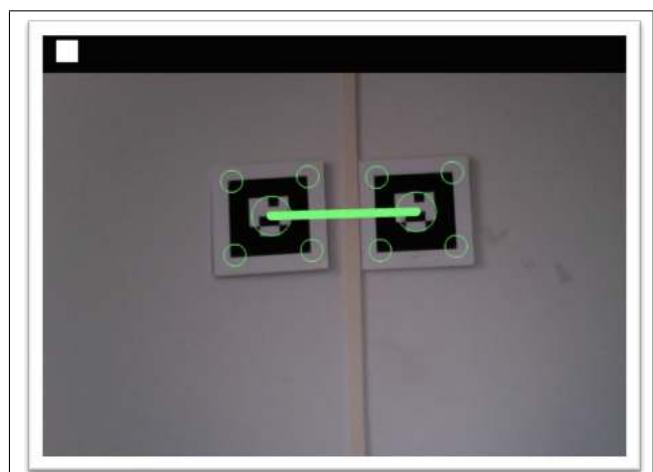
- Il Memory Sonoro utilizza il tracking di tipo ottico: tutte le informazioni sulla posizione degli oggetti presenti sulla scena derivano dall'analisi del flusso video che viene fornito dalla webcam.
- L'interfaccia utente è costituita da coppie di tessere (di numero variabile). Sul verso di ogni tessera sono posizionati dei marker, immagini in bianco e nero che il sistema "riconosce" mediante un algoritmo di *image-processing*. La posizione della tessera sul piano di gioco determina la spazializzazione del suono.

##### 4.2 Le componenti di elaborazione in tempo reale dei dati

- Un software di AR (swAR), un'applicazione creata con Flash e la libreria FLARManger, che si occupa di: a) realizzare il tracking ottico dei marker presenti sulla scena, b) sovrapporre le immagini virtuali alle riprese video della scena reale e c) trasmettere le informazioni sui marker al software audio. Le immagini riprese dalla webcam vengono binarizzate: operazione necessaria affinché il sistema possa riconoscere correttamente il marker. La soglia di binarizzazione va scelta accuratamente e modificata secondo le condizioni di illuminazione della stanza in cui ci si trova ad operare. Una volta localizzato il marker, il swAR vi sovrappone le immagini virtuali (in 2D o in 3D): cerchi colorati virtuali, uno grande al centro e quattro più piccoli sugli angoli (fig.4).

Quando due marker uguali vengono avvicinati, sullo schermo appare anche una linea colorata virtuale che ne unisce il centro.

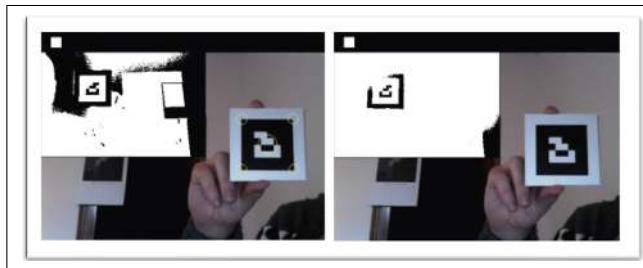
- Un software audio sviluppato in Pure Data (swPD) che si occupa di a) riprodurre i suoni corrispondenti ai marker e b) di spazializzarli in base alla loro posizione sul piano di gioco. Quando sulla scena viene introdotto un nuovo marker, il swPD manda in esecuzione il suono corrispondente e, se il marker viene spostato, il suono cambia la direzione di provenienza a seconda della nuova posizione.
- La scheda audio si occupa di elaborare il flusso di dati inviato dal swPD: i dati vengono separati in 6 segnali audio indipendenti e inviati ai diffusori acustici.



**Figura 4.** Gli elementi grafici virtuali di controllo.

##### 4.3 Le componenti multimediali di uscita

- Il display visualizza la scena reale del piano di gioco e le immagini virtuali sovrapposte generate dal swAR. Per avere un controllo maggiore del funzionamento del sistema è possibile visualizzare anche l'immagine binarizzata da cui il swAR estrae i marker (fig. 5).



**Figura 5.** Display di controllo.

- I sei altoparlanti che realizzano la spazializzazione del suono devono essere disposti secondo lo schema 5.1: di fronte, sinistra, destra, dietro a sinistra e dietro a destra; il *sub-woofer* deve essere posizionato in basso.

Il Memory Sonoro è quindi costituito da una componente di AR (swAR) che invia messaggi di tipo OSC al software per l'elaborazione audio del segnale (swPD). Questi messaggi contengono le informazioni relative al tipo di marker individuato e alla sua posizione e sono veicolati attraverso un *server Flosc* (usato appunto per l'invio di messaggi di tipo OSC da/verso il programma Flash). In particolare, la componente AR è gestita dalla libreria *FLARManager* per Adobe Flash. Il framework FLARManager permette quindi di analizzare e tracciare i marker ripresi dalla webcam; i dati ottenuti sono poi utilizzati per “aumentare la scena reale” con suoni e immagini. Gli elementi grafici virtuali vengono visualizzati sullo schermo e sono generati direttamente dalle librerie di Adobe Flash, mentre gli eventi sonori sono riprodotti e spazializzati da una patch sviluppata in Pure Data (swPD) in base alle informazioni inviate dal swAR tramite messaggi OSC. Il swPD riproduce i suoni corrispondenti ai marker presenti sul tavolo da gioco e li spazializza in base alla loro posizione. Il flusso dati viene inviato alla scheda audio che si occupa di suddividerlo in 6 canali indipendenti. I passi fondamentali che compongono questo processo sono i seguenti:

1. individuazione del marker presente nel flusso video;
2. confronto tra i marker individuati;
3. sovrapposizione di immagini virtuali a quelle ottenute dalla webcam;
4. invio dei messaggi OSC al server;
5. riproduzione del suono;
6. spazializzazione del suono.

## 5. IL GIOCO

In questa sezione si descrive come gli utenti (l'operatore vedente che gestisce il sistema e il bambino/adulto non vedente che gioca con le tessere) interagiscono con il Memory Sonoro.

Il compito principale dell'operatore consiste nel verificare

in tempo reale che il sistema sta funzionando correttamente. A questo scopo il sistema attraverso il display rende visibile a) il flusso video del piano di gioco con le grafiche aumentate e b) la *dialog box* delle immagini binarizzate (fig. 5) fornendo informazioni su:

1. La soglia di binarizzazione; l'operatore, secondo l'iluminazione della stanza, regola la soglia tramite i tasti freccia del computer (freccia in alto e freccia in basso).
2. La correttezza nella rilevazione dei marker; la finestra del flusso video con la grafica sovrapposta permette all'operatore di rendersi conto dello stato del sistema nel riconoscimento del marker: la presenza della grafica indica che il sistema sta effettuando correttamente il riconoscimento; l'assenza della grafica segnala un problema di tracciamento che può essere risolto a) modificando la soglia oppure b) regolando alcuni elementi dell'installazione, come le condizioni di illuminazione o la disposizione della telecamera.
3. La correttezza nella rilevazione delle coppie di marker; quando il bambino avvicina due carte uguali (suoni uguali) il sistema unisce, virtualmente, con una linea colorata dello spessore di 8 pixel il centro dei due marker individuati. L'insegnante deve verificare la presenza di questa grafica e, in caso contrario, modificare le impostazioni del software o alcuni elementi dell'installazione.

L'utente principale (il bambino non vedente) interagisce con il sistema attraverso la manipolazione delle tessere. Di seguito le azioni compiute dal bambino:

1. Girare una carta ed ascoltare il suono corrispondente;
2. Girare un'altra carta ed ascoltarne il suono; questa operazione si ripete fino a quando viene individuato il suono gemello.
3. Interrogazione del sistema: il bambino volge alla telecamera entrambe le carte avvicinandole. Se la scelta è corretta, il sistema manda in esecuzione un suono di conferma. Se la scelta non è corretta il sistema non manda in esecuzione nessun suono: secondo gli esperti di tiflodidattica, infatti, il suono di segnalazione dell'errore è un'informazione ridondante che potrebbe inficiare il compito di memorizzazione degli stimoli sonori principali.

## 6. TESTING

Il prototipo del Memory Sonoro è stato testato su un campione di 12 persone con ipovisione o con cecità, di età compresa tra i 4 e i 45 anni. È stato somministrato un test di usabilità con valutazione basata su feedback verbale degli utenti: questionario [15]. Data l'elevata eterogeneità del campione di utenti (per età, tipo di disabilità, capacità di usare i sensi vicarianti), ogni sessione di testing è stata di

volta in volta adattata al soggetto (numero di tessere impiegate e tempistica). Il test di usabilità ci ha consentito di valutare la risposta del sistema al fine di evidenziare eventuali carenze nella progettazione che possono essere affrontate prima di effettuare uno studio su larga scala. Oltre a esaminare la performance, lo studio di usabilità ha permesso di rilevare ciò che gli utenti pensano del Memory Sonoro; infatti, per quanto i punteggi di performance possano essere positivi, l'opinione dei partecipanti sulla tecnologia è molto importante nel progettare gli sviluppi futuri del sistema. Il test di usabilità è servito a:

1. identificare quali fossero le azioni più congeniali e quelle più difficili da espletare;
2. osservare i limiti e le potenzialità del sistema in condizioni di uso diverse da quelle immaginate durante la fase dello sviluppo.

Per la somministrazione del test sono stati utilizzati stimoli sonori complessi, come: il rumore di un tuffo, di un elicottero, di un aereo, di una macchina da corsa, il verso di alcuni animali e rumori ambientali. Seguendo le linee guida dei test di usabilità [16], sono stati somministrati due questionari: uno da compilare prima dello svolgimento del test, finalizzato alla descrizione della tipologia di utente (genere, età, l'istruzione, visus e il livello di dimestichezza con il gioco di memoria Memory) e l'altro, finalizzato a valutare il feedback fornito dal sistema a livello tattile ed uditivo e a ottenere una serie di opinioni e consigli sull'esperienza di gioco nel suo insieme, da compilare dopo la somministrazione del test.

Il test di usabilità ha permesso di individuare le seguenti problematiche:

1. L'utente manipola liberamente le tessere, girandole e toccandole e, di conseguenza, copre il marker con le dita impedendone il corretto tracciamento; per ovviare a questo problema, l'utente deve sempre appoggiare le tessere sul piano di gioco e allontanare le mani: azioni non spontanee che limitano sia l'interazione sia l'esperienza percettiva del suono spazializzato (in particolare sull'asse frontale). La collocazione della webcam sotto il piano di gioco (in questo caso trasparente), utilizzando, per esempio, la tecnologia TUO<sup>5</sup>, potrebbe risolvere questa problematica.
2. L'utente, dopo aver girato una carta e ascoltato il suono corrispondente, di solito la lascia scoperta per andare alla ricerca del suono gemello; quando però si muove sul piano di gioco, spesso, passando con la mano sopra la prima carta ne riattiva il suono. Questa situazione genera ambiguità e confusione all'utente. Nell'immediato, questo problema è stato risolto chiedendo all'utente di ri-girare la carta subito dopo averne ascoltato il suono. Per assicurare una fruizione spontanea delle tessere è comunque necessario implementare il sistema con una logica di funzionamento diversa.

<sup>5</sup> <http://www.tuio.org/>

3. A causa delle difficili condizioni di illuminazione (luci al neon sopra l'area di gioco) le prestazioni del tracker ottico non sono state ottimali. La soluzione a questa problematica, nell'immediato, si è ottenuta attraverso la regolazione della soglia di binarizzazione o chiedendo agli utenti di spostare le carte in altre zone dell'area di gioco.
4. Tutti gli utenti hanno definito il gioco piacevole e di facile comprensione e gestione.
5. Solo 4 utenti su 12 hanno dichiarato che la spazializzazione del suono li ha aiutati nella ricerca dei suoni sul piano di gioco. Tutti gli altri, per memorizzare le posizioni delle tessere e dei suoni corrispondenti, hanno utilizzato strategie di memoria spaziale. Molto probabilmente la problematica descritta nel punto 1 ha influito anche sulla scelta della strategia di memorizzazione: se gli utenti avessero potuto muovere le tessere liberamente mentre ascoltavano il suono avrebbero percepito più chiaramente gli effetti della spazializzazione in tempo reale.

## 7. CONCLUSIONI E SVILUPPI FUTURI

Il Memory Sonoro è un sistema interattivo multimodale che estende le tecniche della Realtà Aumentata al dominio dell'audio direzionale ed è accessibile a utenti con ipovisione o cecità. L'utilizzo congiunto delle tecnologie di tracking ottico e della spazializzazione audio in tempo reale ha consentito di realizzare un sistema basato su un'interfaccia acustico-tattile con valenza sia ludica sia educativa: la soluzione adottata permette agli utenti di interfacciare il sistema tramite oggetti reali (tessere) e di gestire, senza indossare sensori, uno spazio sonoro spazializzato. La scelta di utilizzare diffusori per la riproduzione del suono, e non cuffie acustiche, ha dato modo agli utenti di muoversi liberamente all'interno dell'installazione. L'architettura del Memory Sonoro prevede l'utilizzo di marker, creati per i sistemi di tracking ottico, che una webcam rileva sul piano di gioco. Il flusso video fornito dalla webcam viene inviato a un software di AR che si occupa di tracciare i diversi tipi di marker. Le informazioni così ottenute sono:

1. utilizzate per arricchire il flusso video con delle grafiche virtuali (cerchi e linee colorate ad uso dell'operatore);
2. inviate ad un software audio che gestisce il mapping e la spazializzazione dei suoni.

Il sistema, mediante un test di usabilità, è stato testato su un campione eterogeneo di 12 utenti. Sono emerse modalità di fruizione diverse da quelle ipotizzate in fase di sviluppo e anche alcuni limiti sia a livello di architettura sia di implementazione del sistema.

Nel prossimo futuro si intende in primis attuare le modifiche elencate nel paragrafo 6. Inoltre, dato che le tecnologie utilizzate possiedono funzionalità che in questo contesto non sono state sfruttate appieno, una di queste è la possibilità di inviare messaggi OSC nelle due direzioni, si

intende implementare anche la comunicazione dei dati dal software PD al software AR fornendo in questo modo un ulteriore elemento di controllo sullo stato del sistema.

Al fine di facilitare l'installazione del sistema, si potrebbero apportare alcune modifiche all'interfaccia grafica di controllo come, per esempio, l'aggiunta di un feedback grafico relativo allo stato della connessione delle componenti software con il server Flosc.

Le funzionalità del Memory Sonoro possono essere ampliate anche: a) utilizzando categorie di suoni complessi che l'utente deve associare, non per somiglianza, ma per area semantica o b) implementando una nuova dinamica di gioco in cui il compito dell'utente consiste nella localizzazione della sorgente sonora in uno spazio acustico tridimensionale a 360 gradi.

## 8. RICONOSCIMENTI

Si ringraziano gli utenti, gli operatori e il direttivo dell'UICI (Unione Italiana dei Ciechi e degli Ipovedenti) di Pordenone (Pn) per la partecipazione al test di usabilità e per la gentile concessione degli spazi necessari. Si ringraziano inoltre per la collaborazione anche alcuni utenti e operatori dell'Istituto Regionale "Rittmeyer" per Ciechi di Trieste (Ts).

## 9. BIBLIOGRAFIA

- [1] F. P. Brooks, "The computer scientist as toolsmith ii," *Magazine Communications of the ACM*, vol. 39, pp. 61–68, March 1996.
- [2] Y. Hatwell, *Psicologia cognitiva della cecità precoce*. Biblioteca Italiana per i Ciechi Regina Margherita - ONLUS, 2010.
- [3] G. Virga, "Considerazioni sperimentali sulla rappresentazione mentale dello spazio nei non vedenti," *Quaderni di Ricerca in Didattica*, no. 10, 2001.
- [4] C. Frauenberger and M. Noisternig, "3d audio interfaces for the blind," in *Proceedings of the International Conference on Auditory Display*, (Boston, MA, USA), July 2003.
- [5] I. E. Sutherland, *A head-mounted three dimensional display*, ch. Reprinted with permission from *Proceedings of the AFIPS Fall Joint Computer Conference*, pp. 757–764. Washington, D. C.: Thompson Books, 1968.
- [6] R. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, pp. 355, 385, August 1997.
- [7] M. Bajura, H. Fuchs, and R. Ohbuchi, "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient," in *Proceedings of SIGGRAPH*, pp. 203–210, July 1992.
- [8] A. Webster, S. Feiner, B. MacIntyre, W. Massie, and T. Krueger, "Augmented reality in architectural construction, inspection, and renovation," in *Proc. Of ASCE 3 Congr. on Computing in Civil Eng.*, (Anaheim, CA), pp. 913–919, 1996.
- [9] F. Liarokapis, "Augmented reality scenarios for guitar learning," *Theory and Practice of Computer Graphics*, 2005.
- [10] F. Liarokapis, S. Sylaio, A. Basu, N. Mourkoussis, M. White, and P. Lister, "An interactive visualisation interface for virtual museums," in *The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST*, 2004.
- [11] D. Schmalstieg, A. Fuhrmann, Z. Szalavari, and M. Gervautz, "Studierstube - an environment for collaboration in augmented reality," in *Proceedings of CVE Conference*, (Nottingham, UK), pp. 37–48, 1998.
- [12] S. Huber, H. Jetter, H. Reiterer, and M. Zollner, "A proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect," in *Proceedings of the 13th IFIP TC International Conference*, (Lisbon, Portugal), September 2011.
- [13] A. Fusillo, A. Panuccio, V. Murino, F. Fontana, and D. Rocchesso, "A multimodal electronic travel aid device," in *Proceedings of ICMI*, pp. 39–46, 2002.
- [14] V. Delić and N. Vujnović Sedlar, "Stereo presentation and binaural localization in a memory game for the visually impaired," *COST 2102 Int. Training School 2009, LNCS 5967*, pp. 354–363, 2010.
- [15] M. Obrist, F. Förster, D. Wurhofer, M. Tscheligi, and J. Hofstätter, "Evaluating first experiences with an educational computer game: A multi-method approach," *Interaction Design and Architecture(s) Journal - IxDa*, no. 11-12, pp. 26–36, 2011.
- [16] D. A. Norman, *The Design of Everyday Things*. The MIT Press, 1998.



# MUSIC ANALYSIS AND KOLMOGOROV COMPLEXITY

David Meredith  
Aalborg University  
dave@create.aau.dk

## ABSTRACT

The goal of music analysis is to find the most satisfying explanations for musical works. It is proposed that this can best be achieved by attempting to write computer programs that are as short as possible and that generate representations that are as detailed as possible of the music to be explained. The theory of Kolmogorov complexity suggests that the length of such a program can be used as a measure of the complexity of the analysis that it represents. The analyst therefore needs a way to measure the length of a program so that this length reflects the quality of the analysis that the program represents. If such an effective measure of analysis quality can be found, it could be used in a system that automatically finds the optimal analysis for any passage of music. Measuring program length in terms of number of source-code characters is shown to be problematic and an expression is proposed that overcomes some but not all of these problems. It is suggested that the solutions to the remaining problems may lie either in the field of concrete Kolmogorov complexity or in the design of languages specialized for expressing musical structure.

## 1. INTRODUCTION

Bent [1, p. 1] defined *music analysis* as the “resolution of a musical structure into relatively simpler constituent elements, and the investigation of the functions of those elements within that structure”. When attempting to find explanations for the structures of musical works, music analysts typically have two goals: first, they want to find explanations that are as *simple* as possible; and second, they want to account for as much *detail* as possible. These two goals often conflict: in order to account for more detail, a more complex explanation is usually required. The music analyst must therefore attempt to find an *optimal* explanation that strikes just the right balance between simplicity and level of detail. For some musical passages, there may be two or more distinct, but equally good, explanations. For example, two equally simple explanations might account for different (but equally important) aspects of the music’s structure; or there may be two or more equally good ways of explaining the same structural aspects of a musical passage. In such cases, the music may give rise to a *multitable* percept, where it can be interpreted equally

satisfactorily in more than one way [2].

The view adopted here is that a musical analysis can be thought of as being an *algorithm* or *program* that, when executed, generates as output a *representation* of the music being analysed. Such a program therefore embodies an explanation for those structural aspects of the music that are encoded in the output representation. Kolmogorov complexity theory [3–7] suggests that the *length* of such a program can then be used as a measure of the complexity of its corresponding explanation: the shorter the program, the simpler—and, in general, the *better*—the explanation. The level of structural detail that the explanation accounts for corresponds to the level of detail with which the music is encoded in the representation generated by the program. Typically, much of the detailed structure in the music will not be encoded in the output of the program and will therefore go unexplained. It is assumed that the representation generated by the program will be an explicit, *in extenso* description of certain aspects of the structure of the music. The program can therefore be seen as being a compressed or compact encoding of the representation that it generates. On this view, the music analyst’s goal is to find the shortest possible programs that generate the most detailed representations of musical passages, works and corpora.

Music analysis is about finding the best ways of interpreting musical works. In other words, it is concerned with finding the most satisfying *perceptual organizations* that are consistent with the musical “surface”. This surface could be either the notated score or a particular performance. Of the two, a score will typically permit a higher number of consistent perceptual organizations, since the micro-structure of a performance will usually reflect the particular perceptual organization (i.e., interpretation) of the performer. Most theories of perceptual organization have been based on one of two principles: the *likelihood* principle of preferring the most *probable* interpretations (originally due to Helmholtz [8]); and the *minimum* [9] or *simplicity* [10] principle of preferring the *simplest* interpretations. Typically, statistical approaches to musical structure analysis (e.g., [11]) have applied the likelihood principle, whereas theories in the tradition of Gestalt psychology (e.g., [12]) apply the minimum principle. Indeed, as van der Helm and Leeuwenberg [9, p. 153] point out, the fundamental principle of Gestalt psychology, Koffka’s [13] law of *Prägnanz*, which favours the simplest and most stable interpretation, can be seen as an “ancestor” of the minimum principle. For many years, the likelihood and minimum principles were considered by psychologists to be in competition. However, in 1996, Chater [10], drawing on results in Kolmogorov’s [3] theory of complexity, pointed

Copyright: ©2012 David Meredith. 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.

out that the two principles are mathematically identical.

The work presented here relates closely to psychological *coding theories* of perceptual organization that employ the minimum or simplicity principle. In the coding theory approach, a *coding language* [9] or *pattern language* [14] is devised to represent the possible structures of patterns in a particular domain. The preferred organization is then the one that has the shortest encoding in the language. Coding theories of this type have been proposed to explain the perception of serial patterns [14–17], visual patterns [18, 19] and musical patterns [12, 16, 20, 21].

Chater [10] points out that two shortcomings of coding theories are, first, that each domain needs its own pattern language; and, second, that the length of an encoding depends to a certain extent on the specific design of the language used. Chater claims, however, that these problems can be overcome by applying the *invariance theorem* [5, p. 104–107], a central result of Kolmogorov complexity theory. This theorem states that the shortest description of any object is invariant up to a constant between different *universal languages*, a universal language being one that is rich enough to express partial recursive functions. Fortunately, all standard computer programming languages (e.g., C, Lisp, Java) are universal languages. This suggests that one might be able to meaningfully compare analyses of a given musical passage by comparing the lengths of programs representing these analyses, written in the same programming language. Moreover, if an objective, effective method can be found for measuring program length appropriately, then it becomes possible in principle to automate the process of searching for the best analysis of any given passage, piece or corpus of music.

In the remainder of this paper, I introduce the idea of representing a musical analysis (and therefore a particular interpretation of a passage of music) as a computer program. I then address the problem of using the *length* of such a program to evaluate its quality, relative to programs representing alternative analyses of the same musical structure.

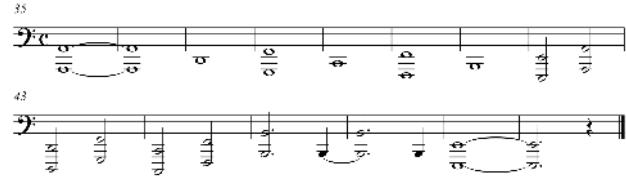
## 2. REPRESENTING A MUSICAL ANALYSIS AS A PROGRAM

Consider Figure 1, which shows the left-hand part of bars 35 to 48 of Chopin’s *Étude* in C major, Op. 10, No. 1. Suppose we segment the passage so that each note onset starts a new segment and then label each segment with the pitch class of the notes in it. If we then merge adjacent segments with the same pitch class content, then the resulting structure can be represented by the sequence of pitch classes,

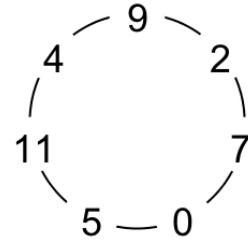
$$9 \ 2 \ 7 \ 0 \ 5 \ 11 \ 4 \ 9 \ 2 \ 7 \ 0 \ 5 \ 11 \ 4 . \quad (1)$$

One possible way of understanding this sequence is as 14 unrelated numbers. Writing the sequence out *in extenso* (as in (1)) expresses this interpretation. This interpretation requires 14 unrelated pieces of information to be remembered and encoded.

Sequence (1) can also be thought of as being con-



**Figure 1.** The bass part of bars 35 to 48 of Chopin’s *Étude* in C major, Op. 10, No. 1.



**Figure 2.** The diatonic fourths pitch class cycle in C major.

structed by repeating the much shorter sequence,

$$9 \ 2 \ 7 \ 0 \ 5 \ 11 \ 4 . \quad (2)$$

This seems to be a simpler and more satisfying way of understanding the sequence, since it recognizes some of the structure in it and requires us to remember only 8 pieces of information in order to reproduce it, namely, 7 unrelated numbers and the single operation of repeating this 7-number sequence.

A little more study reveals that sequence (2) is one turn around the diatonic fourths cycle in C major, shown in Figure 2. If  $p_i$  is the  $(i + 1)$ th element in sequence (1), then

$$p_i = (((i + 2) \bmod 7) * 5 + 11) \bmod 12 . \quad (3)$$

Sequence (1) could therefore be encoded using just this one formula, together with a specification that  $i$  should take values from 0 to 13.

However, Eq. 3 requires us to remember and encode 5 unrelated numbers, 5 unrelated operations and the two boundary values of  $i$ , making a total of 12 pieces of information. This interpretation therefore seems to be less parsimonious than expressing the structure as two copies of sequence (2).

Indeed, one could argue that the description in terms of Eq. 3 is even more complex than the literal description in (1). To reconstruct the structure from sequence (1), the only additional piece of procedural information required is that each element in the sequence is to be printed as it is scanned. Whereas, to reconstruct the pitch class structure using Eq. 3, one also needs to have an understanding of precedence rules, the definitions of the various operations used and a definition of some kind of iterating “for” loop construct that allows the formula to be applied to successive values of  $i$  between the boundary values.

The advantages of using the formula in Eq. 3 over a literal description like sequence (1) only become apparent if the sequence to be described is longer or if the formula can also be used to describe other musical passages. For

```
#include<stdio.h>
main(){
    printf("9 2 7 0 5 11 4 9 2 7 0 5 11 4");
}
```

**Figure 3.** C program that literally prints the pitch class structure of the passage in Figure 1.

```
#include<stdio.h>
main(){
    char i;
    for(i=0;i<14;i++)
        printf("%d ", (((i+2)%7)*5+11)%12);
}
```

**Figure 4.** C program that generates the pitch class structure of the passage in Figure 1 using the diatonic fourths cycle in C major, shown in Figure 2.

example, an *in extenso* description of a sequence consisting of *four* turns around the C major diatonic fourths circle would require encoding 28 pieces of information, whereas a description in terms of Eq. 3 would require no more information than that needed to encode sequence (1). In other words, if the length of such a sequence is  $n$ , then the length of an explicit description is  $O(n)$  whereas the length of a description using Eq. 3 is constant, i.e.,  $O(1)$ .

Another advantage of a description in terms of Eq. 3 is that it can be parametrized and *reused* (with only minor changes) to describe *any* sequence of pitch classes formed by circulating around a diatonic fourths cycle *of any length and in any key*.

To make the discussion more concrete, consider Figure 3, which shows a C program that literally prints out sequence (1), and Figure 4 which uses Eq. 3 to calculate each value in the sequence. If we format these programs so that they are as short as possible and ignore the initial line that loads stdio.h, then the program in Figure 4 is 66 characters long, whereas that in Figure 3 is only 48 characters long. Kolmogorov complexity theory tells us that the complexity of an object is related to the length of the shortest program that can generate it. The fact that the program in Figure 4 is apparently longer than that in Figure 3 therefore seems to support the suggestion made above that a literal description of sequence (1) is actually more parsimonious than one that employs Eq. 3.

However, if we were to cycle around the C major diatonic fourths circle four times rather than just twice, then we would generate the pitch class sequence

$$9 \ 2 \ 7 \ 0 \ 5 \ 11 \ 4 \ 9 \ 2 \ 7 \ 0 \ 5 \ 11 \ 4 \ 9 \ 2 \ 7 \ 0 \ 5 \ 11 \ 4. \quad (4)$$

This sequence is literally printed out by the C program in Figure 7 and generated using Eq. 3 by the program in Figure 5.

If the programs in Figure 5 and Figure 7 are formatted so that they are as short as possible and the initial line that loads stdio.h is ignored, then the program in Figure 5 is shorter (66 characters) than that in Figure 7 (78 characters), supporting the point made above that the advantage of using Eq. 3 only becomes apparent when the sequence to be generated is somewhat longer than sequence (1).

As suggested above, we can also refactor Figure 5 so that we define a separate function that encapsulates Eq. 3

```
#include<stdio.h>
main(){
    char i;
    for(i=0;i<28;i++)
        printf("%d ", (((i+2)%7)*5+11)%12);
}
```

**Figure 5.** C program that generates sequence (4) using the diatonic fourths cycle in C major, shown in Figure 2.

```
#include<stdio.h>
f(char k, char s, char l) {
    char i;
    for(i=0;i<l;i++)
        printf("%d ", (((i+s)%7)*5+k+11)%12);
}
main(){
    f(0,2,14);
}
```

**Figure 6.** C program that generates sequence (1) using a function,  $f(k, s, \ell)$ , such that  $k$  is the pitch class of the tonic of the major key of the diatonic set,  $s$  is the starting index in the cycle (0 being 11 more than the tonic pitch class) and  $\ell$  is the length of the sequence.

and allows us to generate any sequence formed by circulating around a diatonic fourths cycle in any key. This has been done in Figure 6. Note that, given the function  $f$ , defined in Figure 6, we can now describe sequence (1) using the very short expression  $f(0, 2, 14)$ . We can also describe any other diatonic fourths cycle sequence using a similarly short expression. For example, the expression  $f(5, 4, 12)$  generates the sequence

$$0 \ 5 \ 10 \ 4 \ 9 \ 2 \ 7 \ 0 \ 5 \ 10 \ 4 \ 9 \quad (5)$$

which is 12 consecutive steps around the F major diatonic fourths circle, starting on C.

However, the program in Figure 6 is even longer (102 characters) than that in Figure 4. Thus, the ability to encode any member of a whole class of commonly-occurring sequences by a parsimonious expression comes at the price of having to define a custom function that must be included in the “background” knowledge required for these parsimonious expressions to be decoded.

Nevertheless, a music analyst and a software engineer would probably agree that, of the descriptions offered above for sequence (1), the program in Figure 6 is the most elegant, insightful and satisfying, despite being the longest (at least in terms of characters). This seems to be in large part due to the following:

1. the function  $f$  provides a very compact description of an infinite set of structures of a type that commonly occurs in music; and
2. the function  $f$  allows each member of the set that it defines to be encoded with a very short description, given the definition of  $f$ .

In other words, a music analyst would prefer the description in Figure 6 because the function  $f$  is a short program that allows *many* musical structures to be described parsimoniously and thus provides a very simple explanation for

```
#include<stdio.h>
main(){
    printf("9 2 7 0 5 11 4 9 2 7 0 5 11 4 9 2 7 0 5 11 4 9 2 7 0 5 11 4");
}
```

**Figure 7.** C program that literally prints sequence (4).

a large set of musical passages of which the one in Figure 1 is an example.

However, if the pitch class structure of the passage in Figure 1 were the *only* passage in the musical literature that could be generated by a call to  $f$ , then we could not reasonably consider the program in Figure 6 to be a good analysis of this structure. And this seems to be reflected in the fact that Figure 6 is longer than, for example, Figure 3. In order to justify the claim that function  $f$  is a good analytical idea, we would have to show that it could be used to produce parsimonious descriptions (i.e., good explanations) of some other musical passages. That is, we would have to write another program that uses the function  $f$  to generate the pitch class structures of several passages (that actually occur in interesting musical works). This program would demonstrate the value of function  $f$  and this new program as a whole would represent a higher quality analysis of the structures that it generates as output than the program in Figure 6. Moreover, the superiority of this new program would be reflected in the fact that the length of its output would be considerably longer than that of Figure 6, while the length of the two programs would be almost the same.

The foregoing discussion suggests that the quality of the analysis represented by a program  $P$  that generates an output  $X$ , increases with the value  $\ell_{P_0}/\ell_P$  where  $\ell_{P_0}$  is the length of a program that just literally prints the same output as  $P$  (i.e.,  $X$ ), and  $\ell_P$  is the length of  $P$  itself.  $P$  is clearly an *encoding* of  $X$  and  $\ell_{P_0}/\ell_P$  is the *compression ratio* achieved on  $X$  by the encoding  $P$ . The music analyst's goal is therefore to losslessly *compress* music as much as possible, or, more simply, *to find the shortest program that generates the music to be explained*. This is a special case of Occam's razor which states that one should prefer the simplest theory that explains the facts, and Kolmogorov complexity explicates the notion of theory complexity as the length in bits of the shortest description of the theory in terms of a binary program on a universal computer [5, pp. 341–343].

### 3. MEASURING THE LENGTH OF A PROGRAM

If we accept that the goal of music analysis is to find the shortest program that generates the music to be explained, then a music analyst is frequently going to be faced with the problem of deciding which of two programs that generate the same output is the shorter. In cases where the output of one program is a proper subset of the output of another, then the analyst will need to decide which of the two programs (that now generate *different* outputs) achieves the better compression. The analyst therefore needs to be able to measure and compare the lengths of programs.

In the examples above, the length of each program has been given in terms of the number of characters in the

most compactly formatted version of the program's source code, written in C. Clearly, one would not count unnecessary whitespace characters or characters within comments; notwithstanding, there are still several reasons why source-code length in characters cannot be used directly to measure the quality of the analysis that a program represents.

One obvious reason for this is that user-defined identifiers for variables, constants and functions can be of arbitrary length. Thus, if program  $P_1$  has fewer characters than  $P_2$ , we can easily make  $P_1$  longer than  $P_2$  simply by lengthening the names of the variables in  $P_1$ , which clearly does not change the quality of the analysis represented by  $P_1$ . Suppose we try to solve this by stipulating that each user-defined identifier only counts as 1 character. This rule has been used by Zenil and Delahaye [22] in their competition to find the shortest universal Turing machine implementation. But what if a program has more identifiers than there are characters in the character set being used? Indeed, not even all the characters in the character set can be made available for use as identifiers—digit characters cannot be used and others must be reserved for operators and reserved words (e.g., '+', '-', '[' etc. in the C programming language). In fact, in ANSI C, there are only 53 permissible 1-character identifiers (the Roman letters and '\_') [23, p. 192].

It is actually only justified to count 1 character for each user-defined identifier if the number of distinct tokens used by the program (excluding literal strings and numbers) is no more than the number of characters in the character set. Such tokens include

- the names of functions, constants and variables—both predefined (e.g., ‘printf’) and user-defined (e.g., ‘f’, ‘i’ in Figure 6);
- reserved words for constructs and types (e.g., ‘for’, ‘int’);
- required punctuation marks, such as the ‘end of statement’ character (‘;’ in C) and commas;
- operators (e.g., ‘=’, ‘+’).

C uses seven-bit ASCII [23, p. 229] so the character set in this case contains only 128 characters. It is therefore easy to see how the number of distinct tokens may exceed the number of available characters.

Another way of measuring the length of a token is in terms of the number of bits required to uniquely identify it among the complete set of distinct tokens used by (or available to) the program. We would therefore need to know the minimum value of  $n$  such that we can assign a unique bit string of length  $n$  to each distinct token (excluding literal numbers and strings). If there are  $N_{\text{tok}}$  such distinct tokens, then this minimum value of  $n$  is given by

$n_{\text{tok}} = \lceil \log_2 N_{\text{tok}} \rceil$ , which is the amount of Shannon information in the token, assuming all tokens are equally probable [24]. Similarly, the length of a string can be measured in terms of the amount of information in it. Thus, if  $N_{\text{chr}}$  is the size of the set of permissible string characters, then the least number of bits capable of encoding a string character is  $n_{\text{chr}} = \lceil \log_2 N_{\text{chr}} \rceil$  and the length in bits of (i.e., the information in) a string containing  $\ell$  characters is  $\ell n_{\text{chr}}$ . The length in bits of a *literal* number can be defined to be the size of the number's type. For example, in C, typically a long uses 8 bytes or 64 bits of memory, an int uses 32 bits and a char uses 8 bits.<sup>1</sup> Putting all this together, we get the following expression for estimating the *length* of a C program,  $P$ , in bits:

$$\ell(P) = n_{\text{tok}} N_{\text{tok}}(P) + n_{\text{chr}} \sum_{i=1}^{N_{\text{str}}(P)} \ell_{\text{chr}}(s_i) + \sum_{i=1}^{N_{\text{num}}(P)} \ell(n_i), \quad (6)$$

where

- $N_{\text{tok}}(P)$  is the number of tokens in  $P$  that are neither literal strings nor literal numbers;
- $N_{\text{str}}(P)$  is the number of literal strings in  $P$ ;
- $\ell_{\text{chr}}(s_i)$  is the length *in characters* of the  $i$ th string in  $P$ ;
- $N_{\text{num}}(P)$  is the number of literal numbers in  $P$ ; and
- $\ell(n_i)$  is the size in bits of the type of the  $i$ th literal number in  $P$ .

The total number of distinct tokens,  $N_{\text{tok}}$ , is the sum of the number of tokens defined within the language itself (including any libraries loaded), which we can denote by  $N_{\text{tok}}^L$ , and the number of user-defined tokens in the program,  $N_{\text{tok}}^P$ . Therefore  $n_{\text{tok}} = \lceil \log_2(N_{\text{tok}}^L + N_{\text{tok}}^P) \rceil$ . If we assume that the program is written in C and that the stdio.h library is loaded, then  $N_{\text{tok}}^L$  is about 150. So for all the programs given in this paper,  $n_{\text{tok}} = 8$ . Table 1 shows the lengths of the programs presented above in terms of number of characters ( $\ell_{\text{chr}}(P)$ ) and in terms of bits using Eq. 6 ( $\ell(P)$ ). Note that the initial line in each program that loads the stdio.h library has been ignored when calculating the values in this table.

As can be seen in the third column of Table 1, measuring program length using Eq. 6 generally produces a lower value than measuring in terms of the number of characters in the most compactly formatted version of the source code. The values in the third column also show that the proportional reduction in the length value is greater for those programs that capture some of the structure in the pitch class sequence (Figures 4, 5 and 6) than for those that literally print the output (Figures 3 and 7). Moreover, measuring in terms of Eq. 6 solves the problem of arbitrary identifier length.

<sup>1</sup> These values are actually implementation-dependent. The sizeof function can be used to determine the sizes of these types in bytes on a particular system.

Program, $P$	$7\ell_{\text{chr}}(P)$	$\ell(P)$	$\ell(P)/(7\ell_{\text{chr}}(P))$
Figure 3	336	291	0.87
Figure 4	462	381	0.82
Figure 5	462	381	0.82
Figure 7	546	501	0.92
Figure 6	714	573	0.80

**Table 1.** Lengths of programs in this paper.  $\ell_{\text{chr}}(P)$  is the length in characters of the most compactly formatted version of the program. This value is multiplied by 7 because C uses the 7-bit ASCII character set.  $\ell(P)$  is the length in bits calculated using Eq. 6. Both measures ignore the line that imports the stdio.h library.

#### 4. DEPENDENCY OF PROGRAM LENGTH ON LANGUAGE

Although Eq. 6 overcomes the problem of identifiers being of an arbitrary length, it does not solve a deeper problem caused by the fact that a program's length depends on the syntax of the language in which it is written and the constructs that this language makes available. For example, programs for symbolic manipulation tend to be shorter in Lisp than in Fortran whereas the reverse is true for programs that carry out numerical computations. If every program in Lisp were, say, 70% of the length of the “equivalent” program in Fortran, then this would not be a problem in the present context, because we are here only concerned with the *relative* lengths of programs written in the *same* language. However, the situation is not as simple as this. In fact, if we have two Lisp programs,  $P_1^L, P_2^L$ , and two equivalent Fortran programs,  $P_1^F, P_2^F$ , then, in general, it is possible for  $P_1^L$  to be shorter than  $P_2^L$  while  $P_1^F$  is longer than  $P_2^F$ , even if we measure program length in a way that takes into account the arbitrariness of identifier length. We are also here glossing over the non-trivial problem of defining when two programs in different languages are “equivalent”.

We therefore need a method of measuring the complexity of a musical analysis that is either independent of the language in which the analysis is expressed or based on some generally-agreed ‘reference’ language. An obvious place to look for such a measure is in the theory of Kolmogorov complexity.

#### 5. KOLMOGOROV COMPLEXITY

Let  $\phi_0$  be an additively optimal universal partial recursive function that we call the *reference function* and let  $U$  be a Turing machine that computes  $\phi_0$  that we call the *reference machine*. The *conditional Kolmogorov complexity* of an object  $x$  given an object  $y$  is defined to be

$$C_{\phi_0}(x|y) = \min\{l(p) : \phi_0(\langle y, p \rangle) = x\}, \quad (7)$$

where

- $l(p)$  is the length in bits of the program  $p$  that computes  $x$  on the reference machine  $U$  when given input  $y$ ; and

- $\langle y, p \rangle$  is a function that maps the input  $y$  and the program  $p$  onto a single binary string.

$x$ ,  $y$  and  $p$  are understood to be represented as binary strings. The (*unconditional*) Kolmogorov complexity of an object  $x$  is then given by

$$C(x) = C_{\phi_0}(x|\epsilon) \quad (8)$$

where  $\epsilon$  is the empty string. Note that the Kolmogorov complexity of an object is always defined relative to some specified reference function which defines a particular description method which must be an additively optimal universal partial recursive function [5, pp. 104–106].

The *invariance theorem* [5, pp. 105] tells us that, within the set of partial recursive functions that compute  $x$  given  $y$ , there exists an additively optimal universal partial recursive function,  $\phi_0$ . It follows that  $C_{\phi_0}(x|y) \leq C_\phi(x|y) + c_\phi$  for all partial recursive functions  $\phi$  and all  $x$  and  $y$ , where  $c_\phi$  is a constant that depends only on  $\phi$ . It can then be shown that, if  $\phi$  and  $\phi'$  are two additively optimal functions, then

$$|C_\phi(x|y) - C_{\phi'}(x|y)| \leq c_{\phi\phi'} \quad (9)$$

where  $c_{\phi\phi'}$  is a constant that depends only the choice of  $\phi$  and  $\phi'$ . In other words, the complexity of  $x$  given  $y$  is independent of the description method up to a fixed constant for all inputs and outputs. Such a description language could be a standard programming language such as Lisp or C. Thus if  $C_{\text{Lisp}}(x|y)$  and  $C_C(x|y)$  are the complexities of  $x$  given  $y$  in Lisp and C, respectively, then the invariance theorem tells us that  $|C_{\text{Lisp}}(x|y) - C_C(x|y)| \leq c$  where  $c$  is a constant. In other words, the complexity of  $x$  given  $y$  will be the same, up to a constant, regardless of the language in which we express the program that computes  $x$ . However, crucially, we do not know whether the shortest C program that computes  $x$  will be longer or shorter than the shortest Lisp program that computes  $x$ . The invariance theorem therefore helps us if we are interested in knowing approximately the length of the shortest program that computes a given object (i.e., its Kolmogorov complexity). However, it does not help us directly if we want to be able to meaningfully compare the lengths of a number of different programs that compute some given object  $x$ , even if all the programs are written in the same universal programming language. This casts doubt on Chater's [10] claim that the invariance theorem solves the problem of the language-dependency of encoding length.

One way forward might be to adopt the approach taken by Chaitin [25,26] and Tromp [27] who have defined short, concrete universal machines implemented from weak elementary operations (see also [5, pp. 206–211]). However, such an approach would require the expression of musical analyses in a rather low-level language based on lambda calculus or combinatory logic, which a music analyst might find to be too cumbersome. A more practical approach might therefore be to return to the development of coding languages specially designed for expressing musical structure, as has been done recently by Meredith [21].

## 6. CONCLUSIONS AND FUTURE DIRECTIONS

I have proposed that the goal of music analysis should be to devise the shortest possible programs that generate the music to be explained. The reasoning behind this is that a short program that generates a given passage, gives more insight into the structure of that passage and a more satisfying explanation of it than a longer program. This is just a special case of Occam's razor. If we accept this as the goal of music analysis, then, given two programs that generate the same passage of music, an analyst needs to be able to decide which of the two is shorter. This, in turn, implies that the music analyst must be able to measure the length of a program in such a way that this length accurately reflects the quality of the analysis represented by the program. I have shown that measuring program length in terms of number of source-code characters is problematic and I have tentatively proposed an expression for measuring program length that overcomes the problem of arbitrary identifier length. However, there remains a deeper problem caused by the fact that program length depends on the syntax of the programming language and the constructs that the language makes available. This problem is as follows: if  $P_1^A, P_2^A$  are two programs in language  $A$  and  $P_1^A$  is shorter than  $P_2^A$  and  $P_1^B, P_2^B$  are two programs in language  $B$  equivalent to  $P_1^A, P_2^A$ , respectively, then it is possible for  $P_1^B$  to be longer than  $P_2^B$ . This means that whether or not one program is shorter than another depends on the language in which the two programs are written. I have shown that the invariance theorem in Kolmogorov complexity theory does not help to resolve this problem. However, it may be possible to find a solution to the program-length measuring problem in the field of concrete Kolmogorov complexity where the goal is to design concrete universal machines implemented from weak elementary operations. An alternative approach would be to return to the development of coding languages that are specially designed for expressing musical structure.

## 7. REFERENCES

- [1] I. Bent, *Analysis*. The New Grove Handbooks in Music, Macmillan, 1987. (Glossary by W. Drabkin).
- [2] P. Kruse and M. Stadler, eds., *Ambiguity in Mind and Nature: Multistable Cognitive Phenomena*, vol. 64 of Springer Series in Synergetics. Berlin: Springer, 1995.
- [3] A. N. Kolmogorov, “Three approaches to the quantitative definition of information,” *Problems of Information Transmission*, vol. 1, no. 1, pp. 1–7, 1965.
- [4] G. J. Chaitin, “On the length of programs for computing finite binary sequences,” *Journal of the Association for Computing Machinery*, vol. 13, no. 4, pp. 547–569, 1966.
- [5] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*. Berlin: Springer, third ed., 2008.

- [6] R. J. Solomonoff, “A formal theory of inductive inference (Part I),” *Information and Control*, vol. 7, no. 1, pp. 1–22, 1964.
- [7] R. J. Solomonoff, “A formal theory of inductive inference (Part II),” *Information and Control*, vol. 7, no. 2, pp. 224–254, 1964.
- [8] H. L. F. von Helmholtz, *Treatise on Physiological Optics*. New York: Dover, 1910/1962. Trans. and ed. by J. P. Southall. Originally published in 1910.
- [9] P. A. van der Helm and E. L. Leeuwenberg, “Accessibility: A criterion for regularity and hierarchy in visual pattern codes,” *Journal of Mathematical Psychology*, vol. 35, pp. 151–213, 1991.
- [10] N. Chater, “Reconciling simplicity and likelihood principles in perceptual organization,” *Psychological Review*, vol. 103, no. 3, pp. 566–581, 1996.
- [11] D. Temperley, *Music and Probability*. Cambridge, MA.: MIT Press, 2007.
- [12] D. Deutsch and J. Feroe, “The internal representation of pitch sequences in tonal music,” *Psychological Review*, vol. 88, no. 6, pp. 503–522, 1981.
- [13] K. Koffka, *Principles of Gestalt Psychology*. New York: Harcourt Brace, 1935.
- [14] H. A. Simon, “Complexity and the representation of patterned sequences of symbols,” *Psychological Review*, vol. 79, no. 5, pp. 369–382, 1972.
- [15] E. L. L. Leeuwenberg, “Quantitative specification of information in sequential patterns,” *Psychological Review*, vol. 76, no. 2, pp. 216–220, 1969.
- [16] F. Restle, “Theory of serial pattern learning: Structural trees,” *Psychological Review*, vol. 77, no. 6, pp. 481–495, 1970.
- [17] P. C. Vitz and T. C. Todd, “A coded element model of the perceptual processing of sequential stimuli,” *Psychological Review*, vol. 76, no. 5, pp. 433–449, 1969.
- [18] J. Hochberg and E. McAlister, “A quantitative approach to figural “goodness”,” *Journal of Experimental Psychology*, vol. 46, no. 5, pp. 361–364, 1953.
- [19] E. L. J. Leeuwenberg, “A perceptual coding language for visual and auditory patterns,” *American Journal of Psychology*, vol. 84, no. 3, pp. 307–349, 1971.
- [20] H. A. Simon and R. K. Sumner, “Pattern in music,” in *Formal representation of human judgment* (B. Kleinmuntz, ed.), New York: Wiley, 1968.
- [21] D. Meredith, “A geometric language for representing structure in polyphonic music,” in *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, (Porto, Portugal), 2012.
- [22] H. Zenil and J.-P. Delahaye, “The shortest universal machine implementation,” 2008. <http://www.mathrix.org/experimentalAIT/TuringMachine.html>. Accessed 28 July 2012.
- [23] B. W. Kernighan and D. M. Ritchie, *The C Programming Language*. Englewood Cliffs, NJ.: Prentice Hall, 1988.
- [24] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [25] G. J. Chaitin, “A new version of algorithmic information theory,” *Complexity*, vol. 1, no. 4, pp. 55–59, 1995/1996.
- [26] G. J. Chaitin, “How to run algorithmic information theory on a computer: Studying the limits of mathematical reasoning,” *Complexity*, vol. 2, no. 1, pp. 15–21, 1996.
- [27] J. Tromp, “Binary lambda calculus and combinatory logic,” 2011. Available at <<http://homepages.cwi.nl/~tromp/cl/LC.ps>>. Accessed 30 July 2012.



# *Clavecin électrique. Studio dello strumento e realizzazione dell'opera.*

Silvia Lanzalone

Conservatorio di Musica "G. Martucci" di Salerno  
Dipartimento di Nuove Tecnologie e Linguaggi Musicali  
mail@silvialanzalone.it

## ABSTRACT

Il presente articolo descrive il lavoro svolto per la composizione di *Clavecin électrique*, per clavicembalo ed elettronica (2011).

Il brano è stato prodotto attraverso un percorso di sperimentazione sulle possibilità acustiche e tecniche dello strumento svolto con l'aiuto del clavicembalista Giorgio Spolverini e comprende un'installazione di tubi risonanti ad esso integrata.

*Clavecin électrique* è il primo prototipo di un progetto di più ampio respiro rivolto alla realizzazione di un clavicembalo con caratteristiche 'aumentate'.

## 1. INTRODUZIONE

Recenti sperimentazioni sulle sonorità producibili con risonatori di diverse forme e materiali mi hanno portato alla realizzazione di installazioni sonore caratterizzate dalla modifica o dall'integrazione delle condizioni meccaniche ed acustiche dello strumento tradizionale, come per il clarinetto, o per la simulazione dei suoni vocali. Altre 'opere di arte elettroacustica', prodotte con la collaborazione di un'artista visiva, sviluppano un'integrazione tra il suono elettronico ed il 'mezzo' di diffusione sonora [17, 18, 19]. In queste opere il 'mezzo' di produzione ed il 'mezzo' di diffusione, sia acustico che tecnologico, sono parti integrate nell'opera. Il nuovo 'strumento', o lo strumento 'aumentato', viene concepito e progettato insieme alla composizione musicale, al punto da non poter essere scisso da questa. Il carattere parzialmente improvvisativo, comune alle modalità di esecuzione di questo tipo di opere, si rende premessa indispensabile per una sperimentazione tecnica ed espressiva. L'interazione tra strumento musicale e interprete si rinnova rispetto allo strumento acustico tradizionale di riferimento e le tecniche esecutive si reinventano sulla base delle nuove caratteristiche strutturali. Il 'nuovo strumento' richiede spesso un'esplorazione tecnico-espressiva condotta secondo un principio di estemporaneità e questo tipo di approccio conduce ad una naturale teatralizzazione del gesto che scaturisce direttamente dalla sua natura fisico-acustica.

Copyright: © 2012 Lanzalone. 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. GLI STRUMENTI AUMENTATI

Le tecniche di live electronics sviluppatesi dagli anni '70 ad oggi prima con tecnologie analogiche, poi digitali, costituisce senz'altro il punto di partenza per una cultura musicale rinnovata che ricerca modi originali per ampliare le possibilità degli strumenti musicali, sia dal punto di vista dell'invenzione timbrica, sia per quanto riguarda le proprietà di articolazione sonora legate al gesto. L'utilizzo di tecniche estese sugli strumenti tradizionali integrato con le elaborazioni elettroacustiche e la manipolazione dei parametri di controllo del suono, possono rivestire un ruolo fondamentale per le variazioni microstrutturali che danno luogo a trasformazione timbriche complesse [19]. Lo sviluppo parallelo di meta-strumenti, iper-strumenti, o strumenti aumentati, reso sempre più accessibile dall'evoluzione dell'industria tecnologica, ha raggiunto negli ultimi anni e per le stesse esigenze una consistente crescita e una grande attenzione da parte di musicisti e ricercatori scientifici. Tale linea di ricerca costituisce una delle più importanti aree di interesse nella produzione musicale contemporanea che si occupa di interattività e di sistemi multimodali. Le caratteristiche dei 'nuovi strumenti' sono diversificate, in quanto possono includere almeno due tipologie di approccio. Il 'nuovo strumento' può essere interamente digitale, arricchito con interfacce più o meno sofisticate che l'interprete utilizza realizzando gestualità con un grado di complessità paragonabile a quello degli strumenti acustici tradizionali; oppure può consistere in un sistema hardware e software che, attraverso sensori applicati allo strumento tradizionale, o ad altro tipo di 'strumenti', ne aumenta le possibilità tecniche ed espressive rilevando parametri legati al movimento dell'interprete e dello strumento stesso. L'interfaccia può non essere visibile e/o essere affidata interamente o parzialmente al pubblico attraverso azioni che possono richiedere un intervento di tipo consapevole o inconsapevole, prevedendo articolazioni semplificate con sistemi complessi di regole [6]. Molte opere con live electronics risolvono l'interazione con lo strumento tradizionale attraverso criteri di gestione del software che vanno dall'organizzazione strutturata di interventi gestuali minimi fino ad algoritmi che richiedono una condizione esecutiva articolata. Questa può essere

condotta tramite le normali devices di input [13, 14, 15, 16], tramite l'integrazione dell'ambiente software con controllers digitali o analogico-digitali più o meno sofisticati [24, 26, 7], oppure tramite la sovrapposizione di sensori allo strumento tradizionale per l'interpretazione del gesto complesso ad esso correlato [4, 10].

Le due modalità di approccio, apparentemente simili, presentano una certa differenza di risultati: il 'nuovo strumento', quando completamente digitale, può raggiungere una complessità ed un livello di 'embodiment' che lo strumento tradizionale può invece quasi del tutto perdere se il suo suono, elaborato attraverso le tecnologie digitali, viene emesso in maniera decorrelata dal processo di produzione originario. Lo strumento, in tal caso, può essere percepito come 'disincarnato'<sup>1</sup>, secondo un processo percettivo paragonabile a quello che si instaura con l'ascolto della musica acusmatica [28].

Una terza strada pone l'aspetto dell'integrazione tra tecnologia e natura acustica del 'nuovo strumento' come uno degli obiettivi primari da raggiungere nella realizzazione dell'opera musicale. In questo caso il 'meta-strumento' è il risultato di un orientamento estetico che utilizza la tecnologia in modo integrato allo strumento tradizionale tecnicamente modificato, affinché ne risultino ampliati non solo gli aspetti gestuali, ma anche le possibilità acustico - espressive ad essi insindibilmente collegate [17, 18, 19, 23, 8].

### 3. CLAVECIN ÉLECTRIQUE

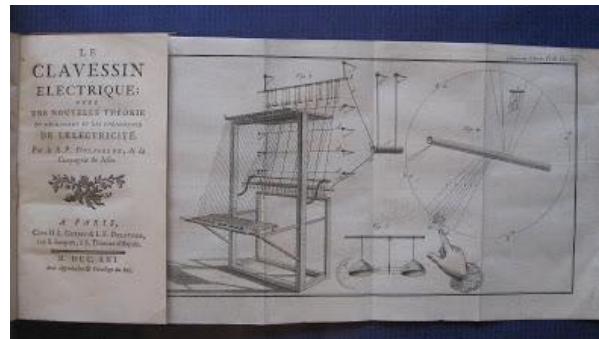
#### 3.1 Il titolo

*Clavecin électrique* prende il titolo dall'omonimo strumento sperimentale inventato a Parigi nel 1759, uno dei primi strumenti a tastiera controllato da principi elettrostatici.

Il clavicembalo elettrico, inventato nel 1759 dal matematico e fisico gesuita francese Jean-Baptiste de La Borde, era caratterizzato da un meccanismo che metteva in oscillazione un batacchio tra due piccole campane intonate all'unisono, appese l'una ad un filo di ferro, l'altra ad un filo di seta, e caricate con carica opposta tramite un generatore di elettricità elettrostatica. Il musicista, alla pressione di un tasto, effettuava lo spostamento di una leva provocando una "messa a terra" di una delle due campane della coppia corrispondente al tasto premuto. Tale azione causava che il batacchio oscillasse rapidamente tra la campana carica e la campana non carica, producendo il suono desiderato [2]. Lo strumento, paragonato ad una sorta di carillon elettrico forse per le sue caratteristiche acustiche, se ne differenziava notevolmente dal punto di vista esecutivo,

in quanto il meccanismo era azionato tramite una tastiera a due ottave<sup>2</sup>.

Il clavecin électrique, anche se fu estremamente innovativo ed ebbe grande successo di pubblico e di stampa, come diversi strumenti aumentati anche più recenti, venne presto dimenticato senza dar luogo ad una specifica letteratura musicale<sup>3</sup>.



**Figure 1.** Schema del clavicembalo elettrico pubblicato da Jean-Baptiste de La Borde (Paris, 1761).

Il brano è un omaggio a tale strumento, ma anche al clavicembalo storico, di cui si vogliono esplorare nuove risorse dinamiche e timbriche.

#### 3.2 Il clavicembalo in *Clavecin électrique*

In *Clavecin électrique* ho voluto creare una composizione musicale per un clavicembalo il cui suono fosse 'aumentato' attraverso l'applicazione integrata di elementi elettroacustici. Per questa prima operazione compositiva non ne ho modificato la struttura meccanica, ma ne ho amplificato le caratteristiche di diffusione sonora e di risonanza. L'interesse ad esplorare il timbro del clavicembalo mi ha portato a privilegiare l'utilizzo di strumenti storici o loro riproduzioni. Il suono dei clavicembali più antichi è infatti, sia per semplicità della meccanica che per caratteristiche timbriche, maggiormente vicino al suono di altri strumenti con corde pizzicate come il salterio<sup>4</sup>, suo antenato, o anche il virginale e la spinetta [3], rispetto al suono dei clavicembali più moderni costruiti a partire dai ultimi anni del XIX secolo<sup>5</sup> [22].

<sup>2</sup> Lo strumento fu descritto in Jean-Baptiste de La Borde, *Le clavessin électrique; avec une nouvelle théorie du mécanisme et des phénomènes de l'électricité*. Paris, 1761. Una riedizione del facsimile del testo è stata pubblicata da Éditions Minkoff, Ginevra, 1997.

<sup>3</sup> Guy-Philippe Ayer, del Conservatorio di Friburgo, ha realizzato, nel 2009, una ricostruzione funzionante dello strumento. Cfr.: [www.synthe.ch](http://www.synthe.ch).

<sup>4</sup> Al salterio, strumento antico di origine orientale a corde pizzicate o percritte, viene aggiunta una tastiera che aziona leve munite di un uncino (penna) per pizzicare la corda. Il termine clavicembalo deriva da clavis (chiave) e compare verso la fine del 1300.

<sup>5</sup> La nota clavicembalista Wanda Landowska, collabora con Pleyel a partire dal 1912 per un progetto di clavicembalo moderno, che si distinse notevolmente dai clavicembali precedenti attraverso l'introduzione di alcuni criteri costruttivi presi in prestito dal pianoforte. Gli strumenti di Pleyel, molto più grandi e pesanti, producevano un suono più potente e meno risonante, richiedevano un tocco più pesante, ma avevano molta più varietà nei registri, che potevano essere cambiati con una pedaliera. Con altri accorgimenti di

<sup>1</sup> Il termine 'disincarnato' (disembodied) è adoperato da Risset per esprimere lo straniamento acustico ottenuto con la musica acusmatica, quando la sorgente sonora di un evento acustico non è riconoscibile né visibile.

Il brano è stato tuttavia scritto, sia per estensione che per combinazioni di registri, in modo da poter essere eseguito su strumenti con caratteristiche diverse, anche prevedendo che alcune tecniche esecutive, come alcuni tipi di articolazioni e alcune velocità, vadano di volta in volta adattate allo specifico strumento usato.

Le prime sperimentazioni sul suono, sulle tecniche esecutive e la prima esecuzione del brano sono state effettuate su un modello italiano messo a disposizione dal Conservatorio Giuseppe Martucci di Salerno<sup>6</sup>. Lo strumento è una riproduzione di un modello italiano del '600 a singolo manuale. Non tutti i brani moderni per clavicembalo possono essere eseguiti su questo tipo di strumento in quanto, a differenza di altri modelli barocchi di epoca posteriore, possiede un solo manuale e una minore estensione, da Fa 0 a MI 5.



**Figura 2.** Clavicembalo Mascheroni, Conservatorio Giuseppe Martucci di Salerno, Italia.

Il clavicembalo adoperato per la seconda esecuzione del brano<sup>7</sup> è stato messo a disposizione dal Conservatorio di Santa Cecilia di Roma. Lo strumento, a doppio manuale, ha un'estensione maggiore e una maggiore quantità di contrasti timbrici e dinamici<sup>8</sup>. L'utilizzo di più manuali amplia le possibilità timbriche ed espressive dello strumento: la possibilità di abbinare diversamente più file di salterelli ai manuali che li azionano infatti, aumenta la quantità di combinazioni timbriche possibili.

Dolmetsch e Challis, fu introdotta la tavola armonica di alluminio per una maggiore stabilità di intonazione e plettri di cuoio per consentire maggiori sfumature nel tocco. La letteratura per clavicembalo del Novecento si riferisce molto spesso a questo tipo di strumento, radicalmente diverso dallo strumento originario.

<sup>6</sup> Prima esecuzione 15 maggio 2011, Giardino della Minerva di Salerno, clavicembalo Giorgio Spolverini, installazione e live electronics Silvia Lanzalone. Manifestazione *Terra delle Risonanze, sviluppi, connessioni, interazioni*, nell'ambito della Settimana Italiana delle Arti e Festival Arti di Maggio 2011 di Salerno.

<sup>7</sup> 12 ottobre 2011, Sala Accademica del Conservatorio S. Cecilia di Roma, clavicembalo Giorgio Spolverini, installazione e live electronics Silvia Lanzalone. Nell'ambito di EmuFest 2011, 4° Festival Internazionale di Musica Elettroacustica del Conservatorio S. Cecilia di Roma.

<sup>8</sup> Nel corso di circa tre secoli, fino al XVIII secolo, l'evoluzione porta i cembalari ad ampliare le possibilità dello strumento. Nel 1700 Vengono costruiti molto più frequentemente cembali a due o tre manuali, di cui viene anche ampliata l'estensione fino a raggiungere cinque ottime nei cembali francesi.

### 3.3 L'approccio metodologico

La realizzazione del brano è stata affrontata attraverso aspetti metodologici che hanno consentito di strutturare un percorso di ricerca applicato a tutti i settori della produzione musicale, ossia tecnico, linguistico, stilistico ed espressivo. Struttura e materiali sono stati esplorati, sperimentati e scelti sulla base di un progetto compositivo emerso dopo una serie di approfondimenti sulla letteratura contemporanea per clavicembalo, sulla tecnica strumentale, sulle caratteristiche acustiche, timbriche e sui metodi di sintesi e di elaborazione del suono degli strumenti a corda pizzicata presenti in letteratura.

La tecnologia applicata, concepita come un mezzo di espansione delle possibilità timbriche del clavicembalo, è stata progettata in modo da enfatizzare ed integrare le caratteristiche di risonanza e di irradiazione spaziale del suono dello strumento. L'installazione che ne è derivata è composta da un algoritmo per la produzione di suoni di sintesi e l'elaborazione in tempo reale dei suoni di clavicembalo, integrato con un sistema di tubi risonanti a cui è stata affidata la diffusione di tali suoni.



**Figura 3.** Clavicembalo Kurt Wittmayer, Conservatorio Santa Cecilia di Roma, Italia.

### 3.4 Riferimenti stilistici e idea formale

La composizione del brano è stata preceduta da una fase di approfondimento della letteratura moderna per clavicembalo solo [22], soffermando l'attenzione su alcune opere del Novecento maggiormente di riferimento per gli aspetti stilistici del brano.

*Rounds* (UE, 1965)<sup>9</sup> di Luciano Berio e *Continuum* (Schott, 1968)<sup>10</sup> di György Ligeti, sono i brani da cui più significativamente ho attinto per la composizione di *Clavecin électrique*. I due brani, entrambi di circa quattro minuti, utilizzano tecniche di articolazione che esaltano aspetti timbrici legati alle caratteristiche di decadimento e di attacco del suono, come i clusters di *Rounds*<sup>11</sup> e i velocissimi pattern melodici di *Continuum*,

<sup>9</sup> Prima esecuzione: Basilea 1965, clavicembalo Antoinette Vischer.

<sup>10</sup> Prima esecuzione: Basilea 1968, clavicembalo Antoinette Vischer.

<sup>11</sup> Si riportano le note di sala scritte da Luciano Berio: <<Rounds si scinde in quattro segmenti, eseguiti dapprima con registrazioni contrastanti. La partitura viene poi capovolta, provocando in questo modo dei cambiamenti di altezze e di timbro; successivamente l'interprete la riporta nella posizione iniziale, eseguendola però con un

costruendo un tessuto fittissimo e rapido. Un particolare interesse ho rivolto alla concezione del clavicembalo che emerge dalla composizione di Ligeti, per il quale il clavicembalo è considerato <<really like some strange machine>>, <<an instrument with a non-continuous sound, the twang of the string is of short duration, followed by silence.>>. L'idea, apparentemente paradossale, consiste nel cercare di creare un suono continuo attraverso <<innumerable thin slices of salami>> [21]. Essa si avvicina ad una concezione "granulare" del suono, in cui la percezione della pulsazione del brano non coincide con il ritmo dell'articolazione, ma dipende dalla distribuzione delle altezze e dalla loro frequenza di ripetizione [5]. Ligeti indica di eseguire il brano al tempo "Prestissimo", e di farlo durare meno di quattro minuti<sup>12</sup>, eseguendo le note senza accenti ed <<extremely fast, so that the individual tones can hardly be perceived, but rather merge into a continuum>> [20]. L'attenzione alla struttura meccanica in relazione alle caratteristiche acustiche consentono a Ligeti di realizzare una partitura imprescindibilmente legata allo strumento per cui è stata concepita. Ligeti afferma: <<A harpsichord has an easy touch; it can be palyed very fast, almost fast enough to reach the level of continuum, but not quite (it takes about eighteen separate sounds per second to reach the threshold where you can no longer make out individual notes and the limit set by the mechanism of the harpsichord is about fifteen to sixteen notes a second). As the string is plucked by the plectrum, apart from the tone you also hear quite a loud noise. The entire process is a series of sound impulses in rapid succession which create the impression of continuous sound>> [21].

Le parole di Ligeti confermano che il transitorio d'attacco prodotto dall'urto del plettro sulla corda e l'inerzia del meccanismo di leve che muove il plettro sono elementi strutturali per la composizione di *Continuum*.

Un analogo principio di 'esasperazione' della tecnica strumentale è stato applicato in *Clavecin électrique*, con l'intento di ottenere condizioni di 'non equilibrio' tali da influenzare la percezione del suono mettendo in evidenza aspetti normalmente considerati secondari o complementari al timbro dello strumento. L'evoluzione formale del brano si sviluppa a partire da questo principio. Il paragrafo successivo illustra la struttura del brano mettendo in evidenza le tecniche esecutive adoperate e le relative finalità espressive.

tempo più rapido. In questa composizione tutte le possibilità offerte dal clavicembalo sono sfruttate: effetti di pedale, clusters, registrazioni di liuto. Rounds è una specie di fuoco d'artificio di linee e timbri, confusi e ramificati all'infinito.>> Testo pubblicato in: <http://www.lucianoberio.org/>.

<sup>12</sup> Cambouropoulos e Tsougras [5] osservano come la velocità di esecuzione indicata da Ligeti implica che ciascuna nota duri almeno 75 millisecondi. Essi paragonano le note di *Continuum* ai grani nella sintesi granulare, che hanno durate inferiori a 100 msec.

### 3.5 Tecniche esecutive e struttura

Una seconda fase preliminare ha riguardato lo studio sistematico delle possibilità tecniche dello strumento in relazione al principio descritto.

Il clavicembalo antico possiede un timbro poco flessibile, che decade molto rapidamente e con poca escursione dinamica, data per lo più dalla quantità di tasti abbassati contemporaneamente e dal tipo di registro utilizzato. Le caratteristiche di attacco e la breve durata del suono mettono in evidenza i dettagli del fraseggio e delle modalità di articolazione scelte dall'esecutore. La meccanica, risultando più leggera di quella del pianoforte, consente di raggiungere una maggiore velocità.

L'interesse che ho rivolto allo strumento riguarda la possibilità di mettere in risalto, di volta in volta, le seguenti caratteristiche del suono dello strumento: il transitorio d'attacco, il rumore prodotto dal sistema meccanico, la risonanza delle corde e della tavola armonica dopo l'eccitazione. Per ottenere questo, in alcune parti del brano ho dovuto cercare il massimo smorzamento delle corde, ed in altre la massima libertà di vibrazione.

La collaborazione con l'esecutore<sup>13</sup> è stata fondamentale per individuare modalità esecutive finalizzate ad ottenere i risultati timbrici desiderati. Oltre al recupero di stilemi tipici della letteratura per clavicembalo dei secoli precedenti, quali trilli, acciaccature, scale, seppure riproposti all'interno di un diverso sistema di organizzazione delle altezze, vengono utilizzati criteri di 'preparazione' dello strumento, o particolari tecniche esecutive già utilizzate nel Novecento sul pianoforte, ma strettamente aderenti alla necessità di intervenire sulle caratteristiche strutturali del clavicembalo.

Sono di seguito elencate le soluzioni più significative presenti in partitura:

a. Timbro smorzato, senza alcuna risonanza delle corde: l'esecutore pone sulle corde un panno di feltro spesso e pesante che ha la funzione di bloccare la vibrazione di tutte le corde all'interno di un'estensione prestabilita<sup>14</sup>. Il risultato sonoro consiste in un timbro composto quasi esclusivamente dal suono del plettro che pizzica la corda e del meccanismo di leve di legno che ne aziona il movimento.

b. Cluster, cromatico o diatonico, da eseguire con una o due mani, oppure con uno o due avambracci, sulla base dell'estensione richiesta.

c. Cluster cromatico da eseguire ponendo una barra di materiale di peso sufficiente ad abbassare un numero di tasti pari a un'ottava, oppure a tutta l'estensione dello strumento. La particolarità della barra consiste nel poter essere lasciata sui tasti per evitare che lo smorzatore di feltro del salterello ricada sulla corda

<sup>13</sup> Giorgio Spolverini.

<sup>14</sup> Nel caso di un clavicembalo in cui il registro da utilizzare abbia due file di corde, sarà importante disaccoppiare la seconda fila, in modo da evitare che i plettri pizzichino le corde, non potendo queste essere smorzate perché collocate in una posizione non raggiungibile dal panno di feltro.

bloccandone la vibrazione. In questo modo il suono della corda pizzicata, che nel clavicembalo termina molto rapidamente anche a causa del rapido abbassarsi dello smorzatore quando l'esecutore rilascia il tasto, risulta molto più prolungato del normale.

- d. Scale, ribattuti e trilli da eseguirsi alla massima velocità, con o senza alterare il sistema di smorzamento dello strumento.
- e. Suoni ottenuti facendo scorrere delicatamente il polpastrello sulla cordiera, sfiorando le corde in modo progressivo così da ottenere un suono continuo, in glissando ascendente o descendente, come è consuetudine produrre sulla cordiera del pianoforte. L'effetto va realizzato con la barra abbassata su tutta la tastiera del clavicembalo, in modo da eliminare completamente qualsiasi smorzamento delle corde.

Il brano risulta composto da sei sezioni chiaramente delineate dal succedersi di queste cinque modalità esecutive, da sole o in combinazione.

Il percorso formale conduce da una completa assenza di risonanza ad un massimo grado di risonanza dell'intero strumento. Le tecniche esecutive e le scelte stilistiche sono state sempre stabilite sulla base di tale percorso.

I riferimenti alla letteratura classica per clavicembalo sono rivisitati con un linguaggio contemporaneo: le scale non appartengono al sistema tonale, i trilli si sovrappongono a distanza di semitono o di tono procurando dissonanze e battimenti, i mordenti sono applicati insieme ai suoni smorzati, gli accordi sono in realtà dei clusters. Il brano è stato concepito per poter essere eseguito anche senza elettronica e la ragione di tale scelta, apparentemente contraddittoria, risiede nella natura delle tecniche esecutive adoperate. La versione per clavicembalo solo mette in evidenza le caratteristiche articolatorie e le diverse tipologie di emissione del suono sullo strumento. L'integrazione del suono elettronicamente prodotto o modificato, invece, oltre a potenziare le sonorità naturali, introduce nuove idee timbriche correlate a quelle originarie e nuove caratteristiche espressive.

#### *Clavecin électrique*

**Figura 4.** *Clavecin électrique*, partitura, estratto pag. 1  
Esempio dalla sezione 1: suoni smorzati, cat. a.

**Figura 5.** *Clavecin électrique*, partitura, penultima ed ultima pagina. Esempio dalla sezione 5: cluster isolati, cat. b + c (pag. 9) e dalla sezione 6: suoni sfiorati, cat. e (estr. pag. 10).

#### 3.6 Analisi dello strumento

Il clavicembalo può essere descritto in modo semplificato come un sistema composto da elementi multirisonanti, le corde, che vibrano alla loro frequenza fondamentale e a tutte le corrispondenti armoniche, in accoppiamento, tramite i due punti fissi alle estremità, detti ponte e capotasto, con la tavola armonica [25]. Quest'ultima introduce delle formanti nello spettro del suono prodotto ed è in accoppiamento con l'aria con una maggiore efficienza di irradiazione del suono rispetto alla corda, nonostante una certa percentuale di energia dissipata. Il 'range' di un clavicembalo va da una frequenza fondamentale grave di circa 50 Hz, per un'estensione di 4 o 5 ottave, fino a parziali nello spettro

prodotto da una corda superiori a 10 KHz. La risposta in frequenza di una tavola armonica eccitata da una sollecitazione impulsiva al ponte si estende per circa un terzo di ottava con un picco intorno ai 200 Hz<sup>15</sup>, ma l'estensione reale dipende dalle dimensioni della tavola e dal suo spessore, mediamente intorno a 3 mm [9, 1]. La lunghezza delle corde, mantenendo costanti diametro, materiale e tensione, dovrebbe raddoppiare ad ogni ottava inferiore. Per evitare lunghezze eccessive, i cembalari adottano materiali di diversa densità o diametro per le corde più gravi e più acute<sup>16</sup>, dando luogo a particolari lunghezze e forme della tavola armonica in base alla scuola di provenienza<sup>17</sup>. Così per le catene, la cui consistenza e disposizione regolava in modo diverso, a seconda delle idee costruttive, la velocità di trasmissione del suono nella tavola modificandone le caratteristiche di risonanza [2].

Per una prima progettazione dei processi di sintesi ed elaborazione del suono, compreso il sistema di diffusione acustica, alcuni suoni del primo clavicembalo sono stati sottoposti all'analisi FFT, da cui sono state estratte le principali formanti e le differenze spettrali più significative tra le categorie precedentemente elencate, in particolare: suono ordinario, suono smorzato, cluster, suono sfiorato. Il suono è stato ripreso con un microfono DPA 4061 posto in prossimità delle corde e della tavola armonica.

Le differenze spettrali individuate hanno dato luogo ad una prima impostazione dei principali parametri di controllo dell'algoritmo, che sono stati poi sottoposti a verifica e leggermente modificati per via empirica durante la successiva fase sperimentale.

### 3.7 Sintesi ed elaborazione del suono

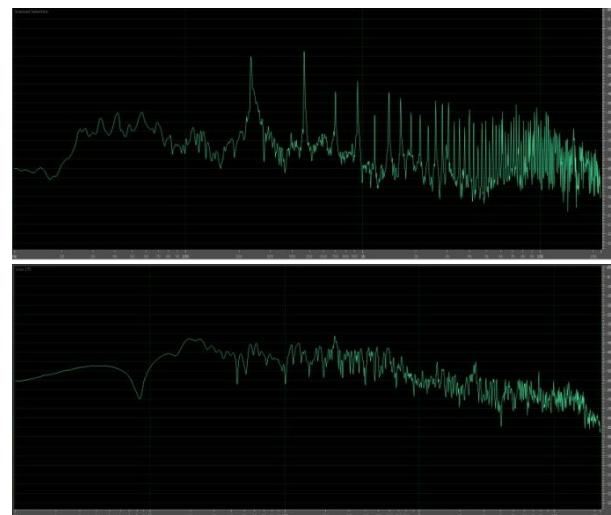
La parte elettronica del brano è costituita in parte da suoni di sintesi controllati in tempo reale dall'esecutore. L'algoritmo è stato prodotto a partire dal modello di K. Karplus e A. Strong [12, 11, 1, 29], al cui ingresso è stata applicata, per alcuni suoni, una sorgente di rumore, oppure, per altri, un diverso segnale di eccitazione<sup>18</sup>, costituito da campioni tratti da suoni di clavicembalo smorzato.

<sup>15</sup> Per misurazioni più accurate cfr. [27].

<sup>16</sup> La corda vibra ad una determinata frequenza di risonanza secondo la formula  $f = v/2L$ , dove  $f$  è la frequenza fondamentale,  $L$  è la lunghezza della corda,  $v$  è la velocità di propagazione del suono sulla corda, calcolabile secondo la formula  $v = \sqrt{T/\mu}$ , dove  $T$  è la tensione della corda in Newton,  $\mu$  la densità lineare della corda, ossia la massa in grammi al metro. La velocità del suono in una corda è dunque maggiore per corde di massa minore, o sottoposte a maggiore tensione. Cfr.: <http://hyperphysics.phy-astr.gsu.edu/hbase/hframe.html>

<sup>17</sup> La legge, detta anche di 'proporzione giusta', è applicata dai cembalari solo per una certa estensione. Il cembalo italiano, ad esempio, la utilizza per le prime 3 ottave alte e, per evitare un'eccessiva lunghezza, le corde acute venivano costruite molto corte, condizionandone il timbro; al grave, invece, si ricorreva al 'ponte spezzato', poiché si abbandonava la proporzione in modo brusco.

<sup>18</sup> Cfr. Vesa Välimäki, Henri Penttinen, Jonte Knif, Mikael Laurson, Cumhur Erkut [30].

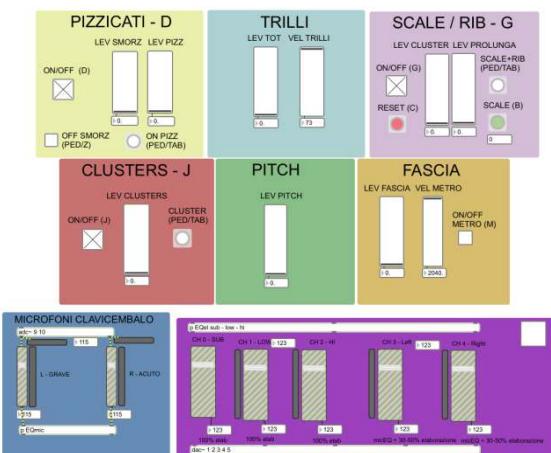


**Figura 6.** Analisi FFT del LA# 231 Hz, con emissione normale (sopra) e smorzata (sotto).

Tra le altre varianti introdotte nell'algoritmo: diverse tipologie di filtri LP e HP del primo e secondo ordine all'interno della controreazione e diverse forme di inviluppo del segnale di eccitazione, talvolta iterato a velocità anche elevate<sup>19</sup>.

L'elaborazione del suono consiste principalmente di filtri risonanti e filtri di equalizzazione, in integrazione al sistema di diffusione progettato.

La variazione dei parametri dell'algoritmo di sintesi è affidata al clavicembalista tramite pedaliera MIDI, mentre la variazione dei parametri dei filtri è gestita dall'esecutore al live electronics.



**Figura 7.** Patch di controllo dell'algoritmo di *Clavecin électrique*, realizzata con Max/MSP.

<sup>19</sup> Si precisa che, in questo caso, l'algoritmo di sintesi non è finalizzato alla riproduzione del suono di clavicembalo, ma alla ricerca di varianti con appropriate valenze espressive. Le varianti, dopo un esame delle diverse implementazioni citate, sono state effettuate con criteri empirici durante le fasi di sperimentazione delle tecniche strumentali.

### 3.8 Diffusione integrata

Il sistema di diffusione integrato allo strumento consiste in un set di otto tubi di diverse dimensioni, come si vede nella figura 8. I tubi, al cui interno è posto un altoparlante di diametro leggermente inferiore, sono 'accordati' proporzionalmente alla loro lunghezza<sup>20</sup>. Cinque di essi sono progettati per essere impiegati come risuonatori<sup>21</sup>; tre di essi sono stati invece progettati per avere una risposta in frequenza lineare, come dei normali diffusori<sup>22</sup>.



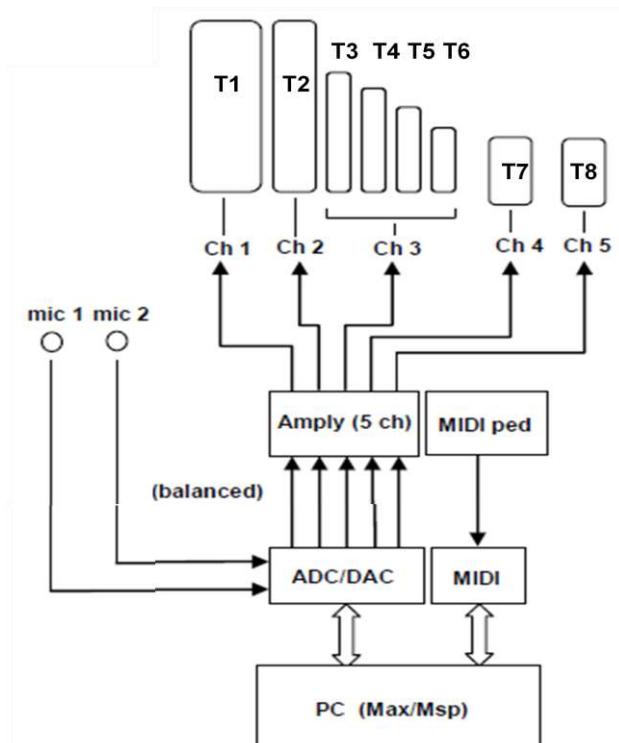
**Figura 8.** *Clavecin électrique*, una foto dell'installazione. Conservatorio di Santa Cecilia, Roma, Sala Accademica, 2012.

I tubi risonanti sono posti sotto al clavicembalo allo scopo di far convergere la zona di emissione del suono elettronico, elaborato o di sintesi, con la zona di principale irradiazione dello strumento costituita dalla tavola armonica, nonché di trasmettere le vibrazioni alla tavola stessa per enfatizzare le risonanze che essa produce con maggiore efficienza. Tale diffusione, pur essendo limitata dal mancato accoppiamento tra l'estremità radiante dei tubi e l'estremità inferiore del corpo del clavicembalo<sup>23</sup>, produce, se sono diffuse alcune frequenze con energia adeguata<sup>24</sup>, la vibrazione simpatica di alcune corde. L'ultima parte del brano, precedentemente indicata come sezione 6, fa largo uso di questo fenomeno. Lo schema in figura 10 riporta il set-up dell'installazione. La duplice funzione dei tubi è realizzata tramite i due diversi tipi di diffusori: il suono elettronico diffuso tramite i tubi risonanti (T2, T3, T4, T5, T6) è 'accordato' con le principali formanti dello strumento; gli altri tubi (T1, T7, T8) hanno la funzione di diffondere nello spazio d'ascolto il suono elaborato e

il suono amplificato dello strumento in modo omogeneo<sup>25</sup>.

## 4. CONCLUSIONI

*Clavecin électrique* costituisce una prima realizzazione di clavicembalo aumentato. Un futuro progetto di collaborazione con altri ricercatori in ambito scientifico, unitamente all'apporto di alcuni interventi di 'luteria' sul clavicembalo tradizionale, potrebbe dar luogo ad un ulteriore approfondimento dello strumento per perfezionare i risultati ottenuti e per raggiungere nuovi obiettivi, sulla base delle misure e delle intuizioni provenienti da questa esperienza. Un certo numero di interventi si rende necessario per la modifica di alcune caratteristiche di eccitazione delle corde; maggiore attenzione dovrebbe essere rivolta ai criteri di accoppiamento del sistema di risuonatori per migliorare la trasmissione delle vibrazioni alla tavola armonica. Un sistema di sensori potrebbe poi contribuire ad una più sofisticata gestione dei controlli, in funzione di una maggiore aderenza al gesto esecutivo.



**Figura 10.** Schema del setup dell'installazione.

<sup>20</sup> La frequenza di risonanza è per i tubi chiusi è data da:  $f = v / 4L$  [25, 9, 1].

<sup>21</sup> L'altoparlante è posto a circa 1/3 della lunghezza del tubo. Per la realizzazione di tali diffusori sono stati utilizzati i criteri costruttivi sperimentati dal CRM - Centro Ricerche Musicali di Roma per la progettazione delle "guide di suono". Cfr. <http://www.crm-music.it/>.

<sup>22</sup> L'altoparlante è posto alla fine dell'estremità inferiore del tubo. <http://www.quarter-wave.com/>

<sup>23</sup> La tavola armonica costituisce la parete superiore di una 'scatola' di legno, detta cassa di risonanza, chiusa al di sotto dal fondo e ai lati dalle fasce laterali. L'unica unica apertura è spesso costituita da un foro circolare, simile ad un 'rosone', posto nella zona più larga, il quale provoca una specifica risonanza, collocata usualmente sotto i 100 Hz.

<sup>24</sup> I valori delle frequenze dipendono dal tipo di clavicembalo impiegato e dalla sua accordatura.

<sup>25</sup> I tubi risonanti hanno altezze rispettivamente di 60, 55, 50, 45 e 38 cm; i diametri sono di 20 cm per T2 e 10 cm per gli altri quattro tubi. La banda passante di T2 va da 40 a 2000 Hz, mentre per gli altri quattro tubi va da 150 a 1500 Hz.

## 5. RIFERIMENTI

- [1] A.A. V.V., a cura di S. Cingolani e Renato Spagnolo, *Acustica musicale e architettonica*, UTET, Torino, 2005.
- [2] A. A. V. V., *Dizionario Enciclopedico della Musica e dei Musicisti*, UTET, Torino, 1983.
- [3] A. Bellasich, E. Fadini, F. Granzera, S. Leschiutta, *Il Clavicembalo*, EDT, Torino 2005.
- [4] F. Bevilacqua, N. Rasamimanana, E. Flety, S. Lemouton, F. Baschet, *The augmented violin project: research, composition and performance report*. Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France, 2006, pp. 402-406.
- [5] E. Cambouropoulos and C. Tsougras, *Auditory Streams in Ligeti's Continuum: A Theoretical and Perceptual Approach*, Journal of Interdisciplinary Music Studies, spring/fall 2009, volume 3, issue 1&2, pp. 119-137.
- [6] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, G. Volpe, *EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems*, Computer Music Journal, 24:1, Massachusetts Institute of Technology, Spring 2000, pp. 57-69.
- [7] L. Dahl, N. Whetsell, J. Van Stoecker, *The WaveSaw: A Flexible Instrument for Direct Timbral Manipulation*. Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07), New York, NY, USA, 2007, pp. 270-272.
- [8] N. Farwell, *Adapting the trombone: a suite of electro-acoustic interventions for the piece Rouse*, 358-363. Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France, 2006, pp.358-363.
- [9] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, New York, NY, USA, 1991.
- [10] A. Freed, D. Wessel, M. Zbyszynski, F. M. Uitti, *Augmenting the Cello*, Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France, 2006, pp. 409-413.
- [11] D. A. Jaffe and J. O. Smith, "Extensions of the Karplus-Strong plucked-string algorithm," *Computer Music Journal*, vol. 7, no. 2, 1983, pp. 56-69.
- [12] K. Karplus and A. Strong, "Digital synthesis of plucked string and drum timbres," *Computer Music Journal*, 1983, vol. 7, no. 2, pp. 43-55.
- [13] S. Lanzalone, *The Manipulation of <<Body Sound>>, Iteration between Music and Dance. Analysis of <<Contropasso>> by Michelangelo Lupone*, in A.A. V.V., "XIII CIM – Colloquium on Musical Informatics. Atti". L'Aquila, 2000, pp.112-115.
- [14] S. Lanzalone, *SDP – Sonorous Drawing Plane AND Laboratory <<The Sonic Draw>>. Timbric Esploration in Real Time through Expressive Gesture Tracks*, in A.A. V.V., "XIII CIM – Colloquium on Musical Informatics. Atti", L'Aquila, 2000, pp.209-212.
- [15] S. Lanzalone, *HIDDENS GRIDS. Paths of expressive gesture between instruments, music and dance*, in *Organized Sound, International Journal of Music and Technology*, Cambridge University Press, 2000, vol.5, n°1, pp.17-26
- [16] S. Lanzalone, *Post-Prae-Ludium per Donau for tuba and live electronics by Luigi Nono. The algorithm for real time performance implemented with Max-MSP*, in "Creativity & Sciences, Proceedings of 5th Conferences UCM - Understanding and Creating Music", Seconda Università degli Studi di Napoli, Caserta, 27-30.11.05, pp.105-111
- [17] S. Lanzalone, *The 'suspended clarinet' with the 'uncaused sound'. Description of a renewed musical instrument*, in Proceedings of 8th International Conference on New Interfaces for Musical Expression (NIME08), Genova, 2008, pp.273-276.
- [18] S. Lanzalone, *Suoni scolpiti e sculture sonore: alcuni esempi di installazioni d'arte elettroacustica*, in "Radici Futuro, Biennale Musica - Atti del XVII Colloquio di Informatica Musicale", AIMI - Associazione di Informatica Musicale Italiana, Conservatorio di Musica 'Benedetto Marcello', La Biennale di Venezia, Venezia 2008, pp. 149-156.
- [19] S. Lanzalone, "Suoni scolpiti / sculture sonore. Un percorso (est)etico tra re-incarnazione e inte(g)razione", in *Le arti del suono*, semestrale, Numero 2 / 2009, installazioni sonore e altri ascolti, pp. 109-132, Edizione Orizzonti Meridionali, pp. 109-132.
- [20] G. Ligeti, *Continuum*, B. Schott's Söhne, Mainz, 1970.
- [21] G. Ligeti, P. Várnai, J. Häusler, C. Samuel, *György Ligeti in Conversationwith Péter Várnai, Josef Häusler, Claude Samuel and Himself*. Eulenberg: London, 1983.
- [22] J. Z. Lindorff, *Contemporary Harpsichord Music: Issues for Composers and Performers*, Graduate Council of The Juilliard School, Doctor of Musical Arts Degree, April 1982.
- [23] M. Lupone, L. Seno *Gran Cassa and the adaptive instrument Feed-Drum*, in *Le Feedback dans la création musicale Rencontres Musicales Pluridisciplinaires*, Lyon, Marzo 2006.
- [24] Y. Nishibori, T. Iwai, *TENORI-ON*, Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France, 2006, pp. 172-175.
- [25] H. F. Olson, *Music, Physics and engineering*, Dover, New York, 1967.
- [26] L. Pizzaleo. *Beyond the Mouse. An Approach to Interfaces and New Lutherie*. Intervento al II International Congress of Contemporary Music and Technologies (CIMTECO). Siviglia, Novembre 2008.
- [27] E. Ravina and P. Silvestri, *Vibration And Acoustics Of Harpsichord Soundboards Through Experimental Modal Analysis*, 3rd International Conference on Integrity, Reliability and Failure, Porto/Portugal, 20-24 July 2009, Paper Ref: S1106\_P0246.
- [28] J. C., Risset, "Musica, calcolo segreto?", in <<Musica ed Elaboratore. Orientamenti e prospettive>>, a cura di A. Vidolin,Venezia, 1980.
- [29] C. Roads. *The computer music tutorial*, MIT Press. Cambridge Massachusetts, 1996.
- [30] V. Välimäki, H. Penttinen, J. Knif, M. Laurson, C. Erkut, *Sound Synthesis of the Harpsichord Using a Computationally Efficient Physical Model*, EURASIP Journal on Applied Signal Processing 2004:7, Hindawi Publishing Corporation, 2004, pp. 934-948.

# “IL DIVINO PRINCIPIO DI INCESSANTE E SAPIENTE VITA”.

## PROPOSTA PER UNA RILETTURA LIVE ELECTRONICS DELL’ANIMA MUNDI PLATONICA.

**Letizia Michielon**  
Conservatorio di Trieste  
[info@letiziamichielon.it](mailto:info@letiziamichielon.it)

### ABSTRACT

Matematica, filosofia, musica e astronomia si intrecciano indissolubilmente nel *Timeo* platonico che raggiunge uno dei propri vertici poetici ed epistemici nella creazione demiurgica dell’*Anima Mundi*. In questo *trait d’union* dialettico, lanciato tra l’universo sensibile e quello intelligibile, pulsano le stesse leggi matematiche, di ascendenza pitagorica, che regolano l’armonia tra anima e corpo, l’ordine cosmico e il sistema modale.

Utilizzando una metodologia di ricerca interdisciplinare, ci proponiamo in questa sede un duplice obiettivo: da una parte la ricostruzione poliprospettica della rete di conoscenze che fonda la centralità speculativa dell’*Anima Mundi*, prendendo le mosse dal ruolo determinante svolto dalla vocalità all’interno del pensiero platonico; dall’altra suggerire alcune possibili applicazioni dei principali nodi concettuali racchiusi nel *Timeo* all’interno di uno *Studio* per pianoforte in prima assoluta che si avvale del contributo elettronico. Attraverso l’uso di otto altoparlanti e la valorizzazione delle tecniche di spazializzazione e di elaborazione del suono, la procedura informatica si rivela infatti decisiva per esaltare la stratificazione gnoseologica del testo, raccordare scienza e musica e restituire l’impalpabile e dinamica spiritualità dell’*Anima Mundi*.

### 1. IL CANTO, MEMORIA E RIVELAZIONE DELLA DIVINA ARMONIA.

Dalla voce delle Muse alla “canzone dialettica”, vertice del percorso epistemico, Platone immagina un itinerario di educazione all’ascolto e alla *phonè* che inconsapevolmente oltrepassa la ricerca metafisica attuata dal *logos*, lasciando trapelare le potenzialità più eversive del linguaggio musicale.

*Copyright: © 2012 Michielon. 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.*

Una prima occasione di approfondimento ci è offerta dalla rilettura del mito delle cicale narrato nel *Fedro*, testo che incarna, secondo Matassi, la *summa* della riflessione sulla musica del pensatore ateniese<sup>1</sup>. Il racconto si svolge in un luogo ameno, abitato dalle ninfe<sup>2</sup>, divinità gentili che attraverso il canto e la danza esprimono poeticamente lo stupore provato dall’uomo greco di fronte al mistero sacro emanato dalla bellezza della natura (*physis*)<sup>3</sup>. Il *canto* sorge cioè dall’*incanto*, dal pudore (*aidos*) che avvolge l’animo afasico immerso nel silenzio naturale<sup>4</sup>.

La stessa meraviglia presiede alla nascita delle Muse, le nove fanciulle figlie di Zeus e di *Mnemosyne* (la Memoria), venute alla luce per lodare la prodigiosa magnificenza dell’universo<sup>5</sup>.

Anche nella *paideia* platonica il canto (*aoide*) appare come una sorta di incantesimo (*ep-aoide*)<sup>6</sup>, in grado di riportare armonia nel cuore dell’uomo.

Il potere ammaliante della voce rivela in Platone una triplice valenza: magico-esoterica, salvifica e terapeutica. Nelle *Leggi* i canti corali vengono definiti infatti “incantesimi collettivi”<sup>7</sup>, che nulla hanno però in comune con i riti che evocano le forze demoniache, giudicati irrazionali e dunque condannati dal filosofo ateniese<sup>8</sup>. L’aspetto salvifico si collega invece direttamente alla tradizione orfica, cultura eminentemente orale, in grado di suscitare una

<sup>1</sup> E. MATASSI: *Musica*, Napoli , Guida , 2004, p. 9.

<sup>2</sup> PLATONE: *Fedro*, Bari-Laterza, 1989, 230b, p. 216.

<sup>3</sup> W.F.OTTO: *Die Musen und der göttliche Ursprung des Singens und Sagens* (1954), trad.it. *Le Muse e l’origine divina della parola e del canto*, a cura di S.Mati, Roma, Fazi Editore , 2005, p. 8. Per un approfondimento sul tema delle ninfe e delle muse, cfr. anche H.KOLLER: *Ninfe, Muse, Sirene*; A. QUEYREL: *Le Muse a scuola*, in *Musica e mito nella Grecia antica*, a cura di D.Restani, Bologna, Il Mulino , 1995, pp. 97-124.

<sup>4</sup> W.F.OTTO, cit., p. 10. Sul ruolo del canto nella *paideia* platonica, cfr. CHAILLEY J.: *La Musique grecque antique*, Pairs, Les Belles Lettres, 1979; sulla vocalità nella cultura musicale dell’Antica Grecia, cfr. BÉLIS A.: *Les Musiciens dans l’Antiquité*, Paris, Hachette Littératures, 1999, pp.179-191.

<sup>5</sup> OTTO, cit., p. 85.

<sup>6</sup> MOUTSOPoulos ricorda come in origine gli incantesimi fossero dei canti (E.MOUTSOPoulos: *Le musique dans l’œuvre de Platon* (1959), trad.it. *La musica nell’opera di Platone*, a cura di F.Filippi, Milano, Vita e Pensiero , 2002, p. 32).

<sup>7</sup> PLATONE: *Leggi*, Roma- Bari, Laterza , 2001, II 659e (p. 71).

<sup>8</sup> MOUTSOPoulos, cit., p. 33.

magnetica immedesimazione tra l'animo dell'ascoltatore e la voce del rapsodo<sup>9</sup>. Il potere sovrumanico e terapeutico della musica assume infine un ruolo di rilievo nella scuola pitagorica<sup>10</sup> e in quella ippocratica<sup>11</sup>, il cui influsso si rivela determinante per elaborazione della teoria platonica della formazione<sup>12</sup>. La facoltà curativa del canto è strettamente connessa alla sua capacità di ricreare nell'uomo il perduto equilibrio psico-fisico attraverso il movimento suscitato nell'anima e nel corpo.

Nelle *Leggi*, ad esempio, Platone sostiene che le ninne-nanne, accompagnate dal dondolio delle braccia, esercitano un'azione benefica e rasserenante sull'anima dei neonati<sup>13</sup>, contribuendo a formare già nei primi anni di vita una preziosa sensibilità verso la nobile armonia. Analogi effetti sono ottenuti dalla danza, disciplina artistica che, come la poesia e il canto, ha il proprio fulcro nel ritmo, cardine della sinergia tra parola, voce e movimento realizzata dalla *mousikè*<sup>14</sup>.

In termini strettamente musicali, il concetto di "armonia" viene utilizzato da Platone per definire "la caratteristica principale di ogni musica, in quanto fattore di ordine e unità, che conduce l'anima umana alla somiglianza, all'amore della Bellezza, alla concordanza con il Bello"<sup>15</sup>.

Simmia esalta la natura divina dell'armonia<sup>16</sup>, mentre Socrate ne sottolinea la capacità di realizzare un legame unificante tra elementi contrari<sup>17</sup>. È proprio l'armonia, infatti, che rende possibile il legame tra l'anima e il corpo e che consente la convivenza delle diverse parti dell'anima tra loro<sup>18</sup>.

## 2. CREAZIONE DELL'ANIMA MUNDI.

### 2.1. Influssi pitagorici sul *Timeo* di Platone.

La relazione armoniosa tra l'anima e il corpo consente di intuire, per analogia, il rapporto che si instaura tra dio e il cosmo.

Nel *Timeo* il demiurgo, sorta di *Nous* anassagoreo che ordina l'universo, dona vita alla materia amorfa e fluttuante introducendovi la forma delle Idee e dei Numeri<sup>19</sup> e l'Armonia<sup>20</sup>. La sua azione plasmante si

<sup>9</sup> Su questo aspetto della tradizione orfica, cfr. SEGAL, cit., p. 295.

<sup>10</sup> MOUTSOPoulos, cit., p.32.

<sup>11</sup> Cfr., a tale proposito, la *Parte II* del saggio di A.BARKER: *Psicomusicologia nella Grecia antica*, Napoli, Guida, 2002 (pp. 75-98).

<sup>12</sup> Per un ulteriore approfondimento sui rapporti tra la tradizione pitagorica e ippocratica e la paideia platonica, cfr. W.JAEGER: *Paideia. Die Formung des griechischen Menschen* (1947), trad.it. *Paideia La formazione dell'uomo greco*, a cura di A.Setti, Firenze, La Nuova Italia, 1967, vol. II e III.

<sup>13</sup> PLATONE: *Leggi*, cit., VII 790 d-e, 791 a-b, p. 212.

<sup>14</sup> Sull'importanza della danza nella *paideia* platonica, cfr. ancora MOUTSOPoulos, cit., *Parte II*, capp. 1-5. Cfr. anche il rilievo della formazione ginnica e coreutica nella tradizione spartana, testimoniata da PLUTARCO: *De Musica*, a cura di R.Ballerio, Milano, BUR, 2001, 1134b, p.35.

<sup>15</sup> Ivi, p. 352. Cfr. PLATONE: *Repubblica*, cit., II 401c, p. 113.

<sup>16</sup> ID.: *Fedone*, Bari, Laterza, 1988, 86 a-c, p. 141.

<sup>17</sup> Ivi, 92e -93a, p. 151.

<sup>18</sup> Ivi, 86b-c, p. 141.

<sup>19</sup> PLATONE: *Timeo*, Bari, Laterza, 1991, 53b, p. 399. Sulla numerologia e l'importanza delle scienze matematiche nel pensiero

sprigiona in particolare attraverso la creazione dell'*Anima mundi*, "il divino principio di incessante e sapiente vita"<sup>21</sup>, inteso come *trait d'union* dialettico lanciato tra l'universo sensibile e quello intelligibile<sup>22</sup>. Per Moutsopoulos l'intero dialogo va interpretato come una "trasposizione parallela di nozioni matematiche attestate tanto nel campo dell'armonia che in quello della scienza delle rivoluzioni celesti"<sup>23</sup>.

Barker osserva come già nel V sec. a.C. le concezioni pitagoriche dell'ordine del cosmo fossero connesse con la musica e il numero<sup>24</sup>. Filolao definisce *harmonia* il principio ordinatore che presiede alla conciliazione e composizione delle due categorie di elementi costitutivi dell'universo, quelli che "limitano" e quelli che "vengono delimitati": tra questi ultimi vi è anche l'intonazione musicale<sup>25</sup>. Il tratto distintivo dell'analisi musicale pitagorica risiede nella rappresentazione dei rapporti tra le intonazioni come rapporti tra numeri interi. Nel Fr. 1, Archita (IV sec. a. C.) concepisce i suoni come movimenti impressi nell'aria; le loro intonazioni vengono definite quali valori misurabili di una proprietà del movimento. Tale proprietà è intesa come velocità o forza o frequenza degli impulsi nel mezzo di propagazione (per es. la corda di uno strumento musicale)<sup>26</sup>. Le misure sul *kanon* (il monocordo, lo strumento più adatto per realizzare esperimenti sulle corde vibranti) dimostrano come il rapporto di 2:1 corrisponda all'intervallo di ottava, quello di 3:2 alla quinta giusta, quello di 4:3 alla quarta giusta. Il fatto che i quattro numeri che stanno alla base dei rapporti musicali di ottava, quinta e quarta giusta siano gli stessi che formano la *grande tetrattì* induce a credere che in essa sia contenuta l'essenza dell'armonia musicale<sup>27</sup>. Ottava, quinta e quarta giusta costituiscono i cardini del sistema musicale greco e, insieme al rapporto di 9:8, che produce il tono (*toniaion diastema*), i punti di riferimento dei modelli di accordatura (*harmoniai*). Di tutti gli intervalli all'interno dell'ottava questi sono gli unici a produrre *symponia* (consonanza) e questa loro proprietà armonizzante ed equilibrante conserva la medesima efficacia anche oltre l'ambito specificatamente musicale, poiché sorge da proprietà matematiche. Dunque utilizzando le strutture matematiche che presiedono alle intonazioni musicali si possono applicare analoghe riflessioni al cosmo e all'anima dell'uomo.

platonico, cfr. anche G.REALE: *Platone. Alla ricerca della sapienza segreta*, Milano, BUR, 2005, cap. VIII

<sup>20</sup> Ivi, 36e-37a, pp. 376-377.

<sup>21</sup> Ivi, 37°, p. 377.

<sup>22</sup> Per un approfondimento sul concetto di *Anima mundi*, cfr. MOUTSOPoulos, cit., p. 348 e pp. 373-376; utile anche G.GUANTI: *Estetica musicale. La storia e le fonti*, Firenze, La Nuova Italia, 1999, pp. 1-10.

<sup>23</sup> MOUTSOPoulos, cit., p. 381.

<sup>24</sup> A.BARKER: *Psicomusicologia nella Grecia antica*, cit., p. 114.

<sup>25</sup> FILOLAO, Framm. 6, in *I Presocratici. Frammenti e testimonianze*, Bari, Laterza, 1986, Tomo I, p. 467.

<sup>26</sup> ARCHITA: Framm. 1, in *I presocratici*, cit., pp. 492 - 493.

<sup>27</sup> Per un approfondimento sulla *tetrattì*, cfr. P.KUCHARSKI: *Étude sur la doctrine pythagoricienne de la Tétrade*, Paris, Les Belles-Lettres, 1952.

Platone aveva una concezione della teoria musicale vicina a quella pitagorica. Nel libro VII della *Repubblica* tra le cinque discipline matematiche propedeutiche alla dialettica compare proprio la teoria musicale<sup>28</sup>, di cui si distinguono due diverse scuole di pensiero: quella dei pitagorici, che cercano di trovare “i numeri delle consonanze che si odono”, e quella degli empiristi, liquidati da Socrate con disprezzo. Ma anche i Pitagorici si sono impegnati, secondo Socrate, in un’impresa inutile in quanto l’autentica teoria musicale dovrebbe occuparsi non di suoni ma di numeri, accessibili solo alla mente razionale. Più che ai motivi che rendono piacevoli al nostro udito certi intervalli musicali, a Platone interessa infatti indagare le forme strutturali dei sistemi matematicamente perfetti che vi sono sottesi.

## 2.2. La genesi dell’ *Anima Mundi*.

Un’esemplare applicazione di tale teoria scientifica è realizzata nel *Timeo*.

Quando il demiurgo crea l’ *Anima Mundi*, ovvero l’entità vivente che anima il corpo del cosmo e che gli preesiste<sup>29</sup>, si coglie con evidenza come i principi che la caratterizzano siano eminentemente matematici, in quanto dipendono dalle tre medie matematiche codificate da Archita<sup>30</sup>.

Secondo Moutsopoulos è necessario distinguere due momenti successivi che caratterizzano la formazione dell’ *Anima Mundi*: uno è quello della concezione (l’intenzione, in virtù della quale il demiurgo può considerarsi un *inventore*), l’altro è quello della realizzazione (l’esecuzione, resa possibile grazie all’azione di un demiurgo *artefice*)<sup>31</sup>. Una volta che il *nous* ha concepito nella sua mente il mondo come immagine intelligibile, egli dovrà realizzare concretamente il suo progetto agendo sulla materia bruta per donarle ordine, forma e coerenza.

Esiste dell’universo un solo modello ideale, perfetto, accessibile grazie alla scienza dialettica, e l’ *Anima Mundi* rappresenta l’ordine ideale impresso alla materia fluttuante. Grazie all’ *Anima Mundi*, dunque, la materia si sottomette all’Intelligenza del demiurgo che le conferisce una struttura attraverso l’azione di forme e numeri<sup>32</sup>. Armonia, proporzione e misura rappresentano i tratti caratterizzanti dell’ *Anima Mundi*, essere il cui *status ontologico*, sorto dall’azione combinata tra l’Identico, il Diverso e l’Esistenza, è superiore a quello della materia ma inferiore a quello dell’Essere Assoluto. La nozione di *Anima Mundi* rappresenta un simbolo cosmografico, una sorta di mondo funzionale ai calcoli dell’astronomia matematica. In sostanza essa risponde alla necessità dialettica della presenza di un intermedio ontologico generato da due tipi di mescolanze di opposti.

<sup>28</sup> PLATONE, *Repubblica*, Bari, Laterza, 1991, 530d-531c, p.

<sup>29</sup> J.N.FINDLEY: *Plato. The written and Unwritten Doctrines*, (1974), trad. it. *Platone. Le dottrine scritte e non scritte*, a cura di G. Reale, Milano, Vita e Pensiero, 1994, p.283.

<sup>30</sup> ARCHITA, Framm. 2, in : *I presocratuci*, cit., pp.493-94.

<sup>31</sup> MOUTSOPoulos, cit., p. 373.

<sup>32</sup> PLATONE, *Timeo*, cit. 37a, p.

Nella *prima mescolanza* agiscono tre coppie di opposti.

- a) l’Essenza indivisibile/divisibile
- b) l’Identico indivisibile/ divisibile
- c) il Diverso indivisibile/divisibile.

Nella *seconda mescolanza* si combinano Essenza intermedia, l’Identico intermedio e il Diverso intermedio<sup>33</sup>.

La mescolanza degli opposti porta alla costruzione di una scala musicale assoluta e l’ *Anima Mundi*, divenendo sostrato dei numeri ideali, comprenderà tutti gli intervalli di questa scala musicale. In quanto *mescolanza* di opposti essa rappresenta un legame ideale espresso dai numeri ideali; in quanto *proporzione* presente nella mescolanza di opposti, cioè in quanto “accordo”, è armonia musicale; in quanto *successione* di intervalli è espressione di una scala musicale assoluta e acquista autonomia dialettica divenendo *Harmonia Mundi*.

Una volta composta attraverso le mescolanze, l’ *Anima Mundi*, il demiurgo continua la sua opera creatrice effettuando tre *suddivisioni*.

Nella *prima* fase viene fissata la struttura dell’ *Anima*: delle masse ancora amorfe vengono estratte dalla mescolanza della materia di cui l’ *Anima Mundi* è costituita e la sua sostanza viene divisa in sette parti secondo la successione matematica 1, 2, 3, 4, 9, 8, 27.

Nella *seconda* fase Platone inserisce la medietà *armonica* e *aritmetica* in ciascun intervallo della serie delle doppie (2, 4, 8) e delle triple (3, 9, 27), operazione grazie alla quale si ottengono le seguenti tre serie:

- A) 1, 4/3, 3/2, 2, 8/3, 3, 4, 16/3, 6, 8
- B) 1, 3/2, 2, 3, 9/2, 6, 9, 37/2, 18, 27
- C) dalla combinazione delle due serie precedenti: 1, 4/3, 3/2, 2, 8/3, 3, 4, 9/2, 16/3, 6, 8, 9, 27/2, 18, 27.

Si vengono così a creare nuove porzioni di materia dell’ *Anima Mundi* che riempiono come una sorta di legami gli intervalli iniziali, generando dei termini che stanno con il precedente in rapporto o di 4/3, o di 3/2 o di 9/8.

Nella *terza* fase l’intervallo di quarta (4/3) viene coperto dall’intervallo più piccolo di 9/8. L’intervallo di quarta può essere coperto da due intervalli di tono più un residuo. Intercalando i nuovi valori negli intervalli ottenuti nella fase precedente, si otterranno nell’intervallo 1-2 i seguenti valori: 9/8, 81/64, 3/4, 3/2, 27/16, 243/128, 2, corrispondenti a una serie di note ascendente procedente da *ut* a *ut*, o discendente da *mi* a *mi*, cioè una eptatonica regolare, identificabile con gli intervalli dei suoni dell’ottacordo diatonico dorico.

## 2.3. I circoli dell’ Identico e del Diverso

L’aspetto musicale è solo uno dei tratti fondamentali dell’ *Anima Mundi* che è insieme “la costruzione di una

<sup>33</sup>F.M.CORNFORD: *Plato’s Cosmology. The Timaeus*, Indianapolis, Hackett Publishing Company, 1997, p.61. REALE identifica i principi supremi del medesimo e del diverso con l’Uno e il Due, a lungo indagati da Platone nelle dottrine non scritte. A questo proposito cfr. G. REALE: *Storia della filosofia antica*, Milano, Vita e Pensiero, 1985, vol. II; AA.VV.: *Verso una nuova immagine di Platone*, Milano, Vita e pensiero, 1994.

scala musicale, di un ordine planetario e di un supremo intelletto sensibile”<sup>34</sup>.

Dopo queste prime suddivisioni e mescolanze l’*Anima Mundi* viene divisa in due senso della lunghezza in due strisce, che il demiurgo dispone a forma di X; le due metà vengono poi curvate per creare circoli che si incontrano di nuovo nel punto opposto alla loro prima giuntura<sup>35</sup>.

Il circolo “esterno” (quello dell’“identico”) gira verso destra a velocità costante, è indivisibile e quello dominante, responsabile del movimento diurno dei Cieli da Est a Ovest; la striscia “interna” (quella del “diverso) viene divisa sei volte fino ad ottenere sette circoli disuguali - i circoli dei pianeti - dotati di diversa grandezza, le cui misure sono determinate da due progressioni geometriche (le stesse impiegate nella prima fase della costruzione della scala musicale). La sequenza delle lunghezze relative è, ancora una volta: 1,2, 3, 4, 9, 8, 27. Si tratta di circoli concentrici che girano a velocità differenti “diagonalmente verso sinistra”<sup>36</sup>, nonostante vengano spinti simultaneamente verso destra dal movimento rotatorio primario della circonferenza esterna. Quando ognuno dei circoli viene collegato a una parte del corpo dell’universo, i suoi movimenti sono visibili nel moto rotatorio delle sfere celesti<sup>37</sup>. I primi tre pianeti sono quelli del Sole, Venere e Mercurio; gli altri quattro quelli della Luna, Marte, Giove e Saturno. La serie dei numeri 1, 2, 3, 4, 9, 8, 27 indica i rapporti delle distanze di questi pianeti dalla Terra. Teoria musicale razionale e astronomia razionale (la scienza senza oggetti visibili di cui si accenna nella *Repubblica*<sup>38</sup>) creano dunque la struttura perfettamente integrata dell’*Anima Mundi*. Come nota Findley, essa “non è semplicemente un legame estrinseco e conscio tra l’*eidos* e l’ambito degli esempi: essa è anche un canale vivente attraverso il quale l’una incide sull’altro”.<sup>39</sup> Platone cioè “crede che la stessa serie di numeri che organizza le distanze dal centro dei vari circoli planetari organizzzi anche i processi psichici che avvengono nel cosmo”. Dunque “il cerchio dell’Identico che conduce i contenuti dell’universo in modo uniforme da Est a Ovest è anche l’organo attraverso il quale l’Anima del Mondo è in grado di conoscere tutte le verità e le nature eterne (37a), mentre il circolo del Diverso, con i suoi circoli planetari subordinati, è anche l’organo attraverso il quale l’Anima ha la certezza di entrare in contatto con tutte le questioni meramente contingenti e con i fatti mutevoli (37b-c)”<sup>40</sup>.

## 2.4 Il Tempo, l’Eterno, l’anima umana.

La generazione del cosmo organico viene perfezionata dalla generazione del Tempo<sup>41</sup>.

<sup>34</sup> FINDLEY, cit., p. 283.

<sup>35</sup> PLATONE, *Timeo*, cit., 37c e d, p. 376.

<sup>36</sup> Ivi, 36 b-d, pp. 375-6.

<sup>37</sup> Ivi, 38c-39e, pp. 378-9.

<sup>38</sup> ID., *Rep.*, cit. 528e-530d, pp. 244- 246.

<sup>39</sup> FINDLEY, cit. p. 284.

<sup>40</sup> Ivi, p. 286.

<sup>41</sup> PLATONE, *Timeo*, cit., 37 c-d, p 377.

Nel caso delle stelle e dei pianeti, che comprendono sia permanenza che divenire, l’intelletto eterno può trattenere la sua creazione dall’evanescenza, ma il tempo vero e proprio sorge, come nota Findley, “quando il flusso privo di significato viene limitato, quando esso rientra all’interno di modelli regolarmente ricorrenti e riconoscibili, e questo sfondo di ricorrenza e di permanenza è ciò che i corpi celesti devono offrire”<sup>42</sup>. Come l’eternità ha il suo riflesso nella persistenza del tempo, così tale persistenza si riflette negli *eide* del movimento e del divenire in sé, della lentezza e della rapidità in sé.

Sempre Findley osserva che il tempo non deve essere considerato un *medium* del divenire o qualcosa di sostanziale: esso è “un aspetto della permanente possibilità dell’esserci degli esempi, di cui lo spazio e il luogo sono l’altro aspetto”<sup>43</sup>.

Il tempo cosmico viene misurato dalla Luna, dal Sole dagli altri cinque pianeti. Il Sole è responsabile della successione del giorno e della notte; calcoli più complessi consentono lo studio delle dinamiche celesti (studio del periodo dei vari corpi planetari e calcolo dell’anno perfetto alla fine del quale tutti i corpi celesti ritorneranno alla loro relativa posizione<sup>44</sup>).

La struttura dell’anima umana è collegata strettamente a quella dell’universo e molteplici sono i nessi tra musica umana e armonia cosmica. Quando l’anima viene attaccata al corpo umano, attraverso “numerosi e invisibili chiodi”, essa viene immessa nel flusso del cambiamento corporeo, in particolare dal movimento del nutrimento che consente al corpo di crescere. Le sensazioni agitano i moti circolari dell’anima e il loro flusso ostacola il moto del circolo più esterno dell’anima (quello dell’“identico”, che dovrebbe invece svolgere il ruolo di dominio). Anche i circoli del “diverso” vengono turbati così che i tre intervalli in rapporto doppio (2:1) triplo (3:1) ed emiolico, epitrito ed epogdoo (3:2; 4:3; 9:8), si contorcono in tutte le direzioni, non potendosi sciogliere tra loro. Si tratta degli stessi rapporti numerici che definivano l’*harmonia* dell’*Anima Mundi*.

Nel pensiero di Platone l’aspetto metafisico dell’armonia si intreccia con la sua valenza etica. La virtù stessa è armonia, e l’arte dei suoni - in particolare il canto, che si avvale dell’apporto concettuale racchiuso nella parola - contribuisce in modo determinante a rendere l’anima virtuosa<sup>45</sup>. La musica, infatti, sorta dall’eros che si sprigiona tra armonia e ritmo<sup>46</sup>, incarnava l’immagine più viva delle sottili assonanze che relazionano l’anima al cosmo.

Occorre però distinguere tra i due diversi tipi di godimento generato dalla musica<sup>47</sup>: vi è una sensazione piacevole, che trae gioia dal dinamismo dei suoni, pur senza saperne cogliere il nesso; e vi è un piacere razionale, che si sprigiona dai suoni puri, emessi da una

<sup>42</sup> FINDLEY, cit., p. 286.

<sup>43</sup> Ivi, p. 287.

<sup>44</sup> PLATONE, *Timeo*, cit., 39 d, p. 380.

<sup>45</sup> PLATONE: *Protogora*, Roma- Bari, Laterza, 2003, 326b, p. 87; *Repubblica*, cit., III, 401d , p. 113; IV, 442a, p. 156.

<sup>46</sup> ID.: *Simposio*, Bari, Laterza, 1989, 187c, p. 162.

<sup>47</sup> ID.: *Timeo*, cit., 80b, p. 432.

“voce unica”, belli in sé e non in relazione ad altro<sup>48</sup>. La nota unica e pura a cui allude Platone evoca l’unità interna del suono, caratterizzata da un rapporto numerico. La predilezione dimostrata per i suoni puri si giustifica dunque con la capacità che essi hanno di lasciar trasparire in grado elevato l’idea di cui partecipano. Le emozioni estetiche sorte da questo tipo di ascolto conducono alla conoscenza del Vero, a cui è strettamente connessa la conoscenza del Bene<sup>49</sup> e del Bello<sup>50</sup>.

La musica, dunque, nel momento stesso in cui rivela la struttura del mondo<sup>51</sup>, instilla nell’animo l’intuizione e il desiderio di un modello migliore di vita. Di qui la sua importanza cruciale nell’ambito dell’organizzazione statale e il rilievo pedagogico affidatole nella *Repubblica* e nelle *Leggi*.

Prendendo spunto dalla tradizione pitagorica, il cui paradigma, radicato nell’orfismo, attribuisce all’educazione musicale un ruolo privilegiato all’interno della *paideia* in quanto consente la comprensibilità e la comunicazione dei segreti racchiusi nella *tetratti* (la dottrina riservata agli acusmatici, ai quali vengono rivelate le leggi che governano la musica celeste e umana<sup>52</sup>), Platone invita a non anteporre le orecchie alla mente. Cercando i numeri che danno origine alle consonanze<sup>53</sup>, si rende cioè necessaria l’elaborazione di una teoria musicale *sorda*<sup>54</sup>, in linea con la tradizione anti-empirista dei pitagorici<sup>55</sup>. Attraverso questo silenzio intenso, che rifiuta voci e strumenti per nutrirsi di pura astrazione matematica, è possibile accedere alla “canzone dialettica”, la sola capace di elevare la parte migliore dell’anima alla contemplazione dell’Essere più sublime<sup>56</sup>.

### 3. ANIMA MUNDI, ÉTUDE PER PIANOFORTE ED ELETTRONICA

Primo di una serie di sette studi per pianoforte ognuno dei quali dedicati a un compositore e a un filosofo, il brano si ispira alla concezione platonica dell’*Anima Mundi* descritta nel *Timeo*.

<sup>48</sup> ID.: *Filebo*, Bari, Laterza, 1989, 51d, p. 118.

<sup>49</sup> ID.: *Repubblica*, cit., VI, 508e, p. 223.

<sup>50</sup> ID.: *Simposio*, cit., 212a, p. 192.

<sup>51</sup> ID.: *Timeo*, cit., 47b-d, pp. 391-392.

<sup>52</sup> JAEGER, cit., p. 194.

<sup>53</sup> PLATONE: *Repubblica*, cit., VII, 531 a-c, p. 247.

<sup>54</sup> Secondo MERIANI, “la teoria musicale ‘sorda’ auspicata da Platone si configura, in sostanza, come uno strumento euristico e astratto, che dalla musica prende soltanto le mosse, per poi staccarsene progressivamente, definendo campi e metodi di indagine del tutto autonomi” (MERIANI, cit., p. 105). Su questo argomento, cfr. anche BARKER, cit., pp.119-120.

<sup>55</sup> Per un approfondimento sulla scuola pitagorica e le sue diverse fasi di sviluppo, cfr. P. CIARLANTINI: “*Phonos ed ethos. Aspetti di estetica musicale greca*”, Macerata, Olmi Ed., 1985, pp. 9 – 17; sul rapporto tra Platone e la scuola pitagorica del tempo di Socrate, cfr. MERIANI, cit., pp. 83- 114.

<sup>56</sup> PLATONE: *Repubblica*, cit., VII 532c, p. 248. Nonostante tale impostazione metafisica e la rigida normativa che regola la formazione del compositore, Platone non disconosce il ruolo decisivo svolto dall’ispirazione nella creazione artistica (cfr. PLATONE: *Ione*, Roma-Bari, Laterza, 2003, 533, pp.365-6).

L’utilizzo dell’elettronica, e in particolare della spazializzazione, attraverso l’uso di otto altoparlanti (elementi che giustificano la seconda dedica dello studio, quella a K. Stockhausen), contribuiscono a esaltare la stratificazione del testo filosofico e a porne in luce l’impalpabile spiritualità.

La disposizione degli altoparlanti è la seguente (fig.1):

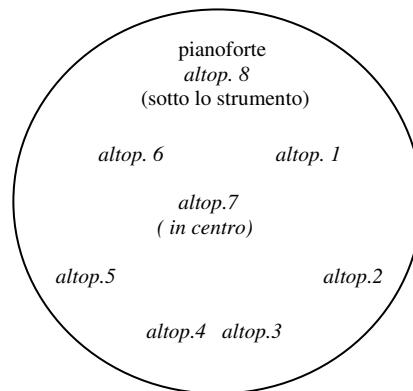


fig.1. Disposizione degli altoparlanti

Formalmente composto da sette sezioni - macrostruttura che riflette la prima delle tre suddivisioni dell’*Anima Mundi* realizzata dal demiurgo, quella delle “doppie” e delle “triple” – il brano segue da vicino i momenti salienti del *Timeo*.

Nell’introduzione, “Chora”, viene evocata la materia indistinta che il demiurgo utilizza per plasmare l’universo. Il blocco accordale affidato al pianoforte è costituito dalla serie originaria del brano, amplificata con spazializzazione circolare per esaltare la natura avvolgente dell’*Anima Mundi* (fig.2):

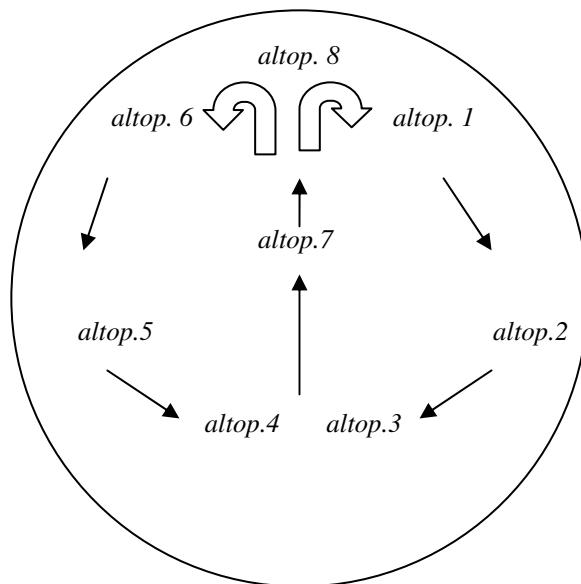


Fig.2. Spazializzazione di Chora

Seguono i quattro elementi (terra, acqua, fuoco e aria), tratti dalle quattro possibili forme della serie

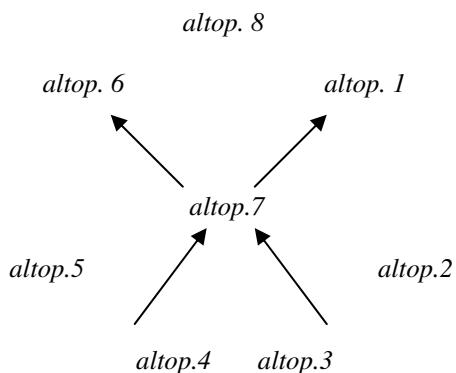
(Originaria, Inversa, Retrograda e Retrograda Inversa, cfr. fig. 3),

Prospecto delle serie *Anima Mundi*

Originale  
Inversa  
Retrograda  
Retrograda Inversa

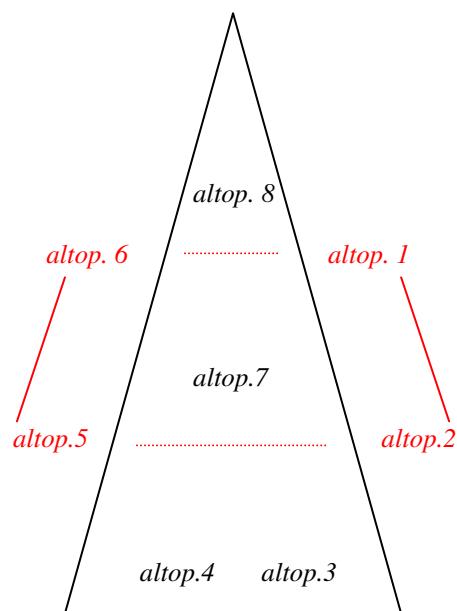
**Fig. 3.** Serie dell'*Anima Mundi*

fino alla comparsa, dalla rarefazione dell'aria, del tema dell'*Anima Mundi*, dal carattere eminentemente lirico. In questa prima creazione ‘concettuale’ immagino che l’essenza trasparente dell’*Anima Mundi* contenga, come in una sorta di sogno, tutto ciò che seguirà (compresi alcuni frammenti degli studi dell’intera raccolta, pre-registrati e lanciati dagli altoparlanti). Alla fase della mescolanza segue quella delle tre suddivisioni operate dal demiurgo, nelle quali ho utilizzato le note corrispondenti alla numerologia platonica. La spazializzazione agisce in questa sezione sottolineando i due rami rovesciati della lambda (prima suddivisione, fig.4):



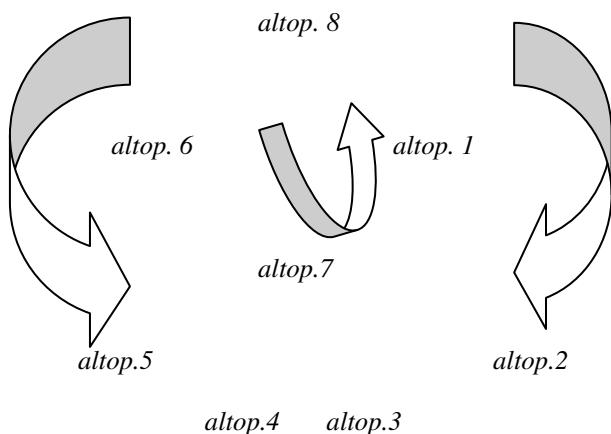
**fig. 4.** Spazializzazione della prima suddivisione

le varie triangolazioni della seconda suddivisione (in fig.5 sono evidenziate le traiettorie percorse dal suono tra gli altoparlanti 3+4 , 7+8; e in rosso la sezione di triangolo formato da 5+2 e 1+6: ricordiamo che il triangolo rappresenta una figura geometrica fondamentale all’interno di tutto il *Timeo*):



**Fig. 5.** Esempi di triangolazione nella seconda suddivisione

e frammenti di circolarità (terza suddivisione, quella che porta alla scala dorica esaltata dai glissandi; in fig.6 si vedono alcuni es. dei frammenti di rotazione formata dagli altoparlanti 8,1,2; 7,1,8; 8,6,5):



**Fig. 6.** Esempi di rotazione circolare nella terza suddivisione

Riguardo all’uso del materiale tematico all’interno delle tre suddivisioni, segnaliamo nella prima fase l’impiego delle seguenti note derivate dalla suddivisione delle doppie e delle triple ascende, individuata da Moutsopoulos: do1, do2, sol2, do3, re4, do4, si5; e descendente: la5, la4, re4, la3, sl2, la2, do1<sup>57</sup>; nella seconda la successione do1, fa1, sol1, do2, fa2, sol2,

<sup>57</sup> MOUTSOPoulos, cit., p. 395.

do3, re3, fa3, sol3, do4, re4, la4, mi5, la<sup>5</sup><sup>58</sup>; nella terza, infine, l'ottacordo dorico, in glissando.

La sezione successiva è puramente elettronica e utilizza una spazializzazione rotante in *loop* per potere rendere al meglio il dinamismo innescato dai due “circoli”, quello *esterno* e quello *interno*, di cui parla Platone. Il circolo *esterno*, simbolo dell’eterno, gira verso la destra dell’ascoltatore e permane regolare ed equilibrato fino alla fine del brano; si avvale di una cellula extrapolata dal tema che identifica *Anima Mundi* e utilizza la rapida successione degli altoparlanti 8,6,5,4,3,2,1. Il circolo *interno*, invece, gira verso la sinistra dell’ascoltatore, è irregolare e spezzato in tre frammenti tratti dai temi dei quattro elementi, in libera alternanza tra loro; è caratterizzato da forti contrasti e coinvolge in rotazione gli altoparlanti 3,4,5,6,1,2 (fig. 7):

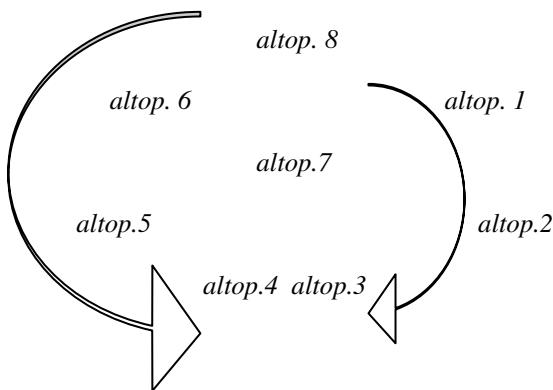


Fig. 7. Spazializzazione dei due circoli “esterno” e “interno”

Inizia quindi la sesta fase, quella che presiede alla creazione dei pianeti, incastonati in sette orbite derivate da un’ulteriore suddivisione del circolo interno, quello del “diverso”.

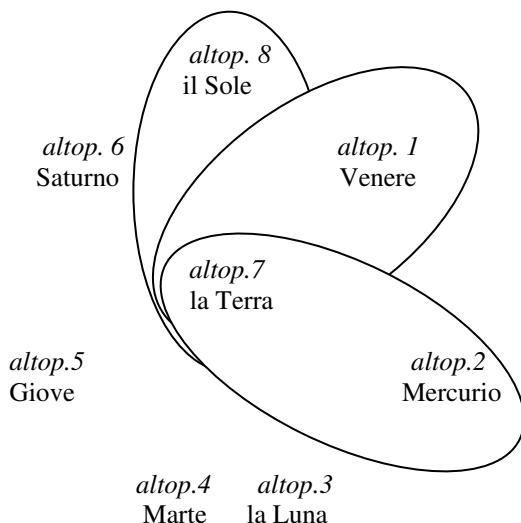


Fig. 8. Es. di alcune rivoluzioni dei pianeti

Il materiale tematico è tratto dall’interpolazione dei quattro elementi originari; la spazializzazione si

<sup>58</sup> Ivi, p. 396

complica ed evidenzia come ognuno degli otto altoparlanti rappresenti un pianeta, mentre la Terra è posta, in linea con le concezioni astronomiche del tempo, al centro (altoparlante n. 7). L’eco del materiale acustico che caratterizza ogni corpo celeste viene catturato *live*, elaborato e posto in rotazione con l’altoparlante centrale. Si vengono così a creare delle rivoluzioni planetarie fittizie, espresse dalle diverse velocità del *loop*, sempre più lento a seconda della distanza che separa ogni pianeta dalla Terra (vedi fig.8). Nella coda avviene un’accumulazione dinamica e agogica, oltre che timbrica, grazie alla granulazione, attraverso l’utilizzazione di materiale pre-registrato lanciato radialmente dall’altoparlante centrale (n.7) a tutti gli altri altoparlanti (fig. 9):

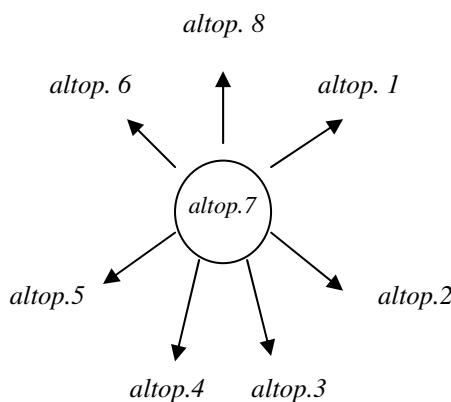


Fig. 9. La deflagrazione e l’irruzione nel tempo

La deflagrazione conclusiva rappresenta il momento in cui l’opera del demiurgo irrompe nella realtà del tempo: comincia la vita, ovvero, simbolicamente, l’intera serie dei sette studi (sette appunto, come i pianeti).

Alla radice delle proporzioni architettoniche temporali, vi è la numerologia derivata dalla prima suddivisione, quella che utilizza le due progressioni geometriche in ragione del doppio (2,4,8) e del triplo (3,9,27).

Si sono prese in considerazione sia le successioni numeriche singole (2,4,8; 3,9,27), precedute o meno dall’unità; sia la successione mista (1,2,3,4,9,8,27)<sup>59</sup>; sia, infine, le varie combinazioni che sorgono dal disporre le progressioni in ragione del 2 e del 3 sui rami rovesciati della lambda, figura della grande *tetratti*<sup>60</sup> (fig.10):

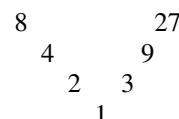


Fig.10. Le progressioni geometriche poste sui due rami rovesciati della lambda

<sup>59</sup> Platone inverte il 9 con l’ 8 per mantenere intatto l’ordine interno della progressione. Il 9, quadrato di 3, è infatti una potenza più piccola di 8, cubo di 2 (ivi, pp. 376-377).

<sup>60</sup> Sulla raffigurazione della grande *tetratti* attraverso i due rami della lambda, cfr. SPITZER L.: *Classical and Christian Ideas of World Harmony*, trad.it. *L’armonia del mondo, storia semantica di un’idea*, a cura di V.Poggi, Bologna, Il Mulino, 2006, p.14.

Ecco, in sintesi, le proporzioni temporali delle sette sezioni:

- 1) Chora: 10'' (1+2+3+4)
- 2) Quattro Elementi: 17'' (9+8)
- 3) Anima Mundi: 27'' (27)
- 4) Le tre suddivisioni: 35'' (da 27+8). La fase è articolata come segue:

- a) prima suddivisione: 6''=2+4;
  - b) seconda suddivisione: 12''=3+9;
  - c) terza suddivisione: 17'', cifra sorta dalla differenza tra il totale 35''- 18'', quantità composta dalla somma di a)=6 +b)= 12
- 5) I due circoli: 18'', così ripartiti:
- a) 5''(2+3) durata di rotazione del circolo esterno da solo
  - b) 13'' (4+9) - con innesco al 6''- corrispondente alla durata del circolo interno
  - 6) I sette pianeti: durata complessiva di 53'' (2+3+4+9+8+27)
  - a) Sole: 2''
  - b) Venere: 3''
  - c) Mercurio: 4''
  - d) Luna: 5''
  - e) Marte: 6''
  - f) Giove: 7''
  - g) Saturno: 10''
  - 7) Coda e deflagrazione: 10''

La componente elettronica viene utilizzata principalmente per esaltare la spazializzazione ma vengono realizzate anche altre operazioni decisive per la struttura compositiva del brano: inserimenti di materiali preregistrati, operazioni *live*, anche di elaborazione del suono che arricchiscono la tavolozza timbrica ed esaltano la natura anfibia dell' *Anima Mundi*, onnipresente *ousia* che riflette nel suo ordine la perfezione del mondo sovrasensibile e ne consente l'innesto nella vita dell' "animale cosmico".

## BIBLIOGRAFIA

- [1] AA.VV.: *Verso una nuova immagine di Platone*, Milano, Vita e Pensiero, 1994
- [2] BARKER A.: *Timaeus on music and the liver*, in M.R. WRIGHT (cur.): *Reason and Necessity: Essays on Plato's Timaeus*, London, 2000.
- [3] BARKER A.: *Psicomusicologia nella Grecia antica*, Napoli, Guida, 2002.
- [4] BÉLIS A.: *Les Musiciens dans l'Antiquité*, Paris, Hachette Littératures, 1999.
- [5] CHAILLEY J.: *La Musique grecque antique*, Pairs, Les Belles Lettres, 1979.
- [6] CIARLANTINI P.: "Phonos ed ethos. Aspetti di estetica musicale greca", Macerata, Olmi Ed., 1985
- [7] CORNFORD F.M.: *Plato's Cosmology. The Timaeus*, Indianapolis, Hackett Publishing Company, 1997.
- [8] FINDLEY J.N.: *Plato. The written and Unwritten Doctrines*, (1974), tr. it. *Platone. Le dottrine scritte e non scritte*, a cura di G. Reale, Milano, Vita e Pensiero, 1994
- [9] GUANTI G.: *Estetica musicale. La storia e le fonti*, Firenze, La Nuova Italia, 1999.
- [10] JAEGER W.: *Paideia. Die Formung des griechischen Menschen* (1947), tr.it. *Paideia La formazione dell'uomo greco*, a cura di A. Setti, Firenze, La Nuova Italia, 1967.
- [11] I PRESOCRATICI: *Frammenti e testimonianze*, Bari, Laterza, 1986, Tomo I.

- [12] KOLLER H.: *Ninfe, Muse, Sirene*; A. Queyrel: *Le Muse a scuola*, in *Musica e mito nella Grecia antica*, a cura di D. Restani, Bologna, Il Mulino, 1995.
- [13] KUCHARSKI P.: *Étude sur la doctrine pythagoricienne de la Tétrade*, Paris, Les Belles- Lettres, 1952.
- [14] MATASSI E.: *Musica*, Napoli , Guida , 2004.
- [15] MOUTSOPoulos E.: *Le musique dans l'ouvre de Platon* (1959), tr.it. *La musica nell'opera di Platone*, a cura di F. Filippi, Milano, Vita e Pensiero , 2002.
- [16] OTTO W.F.: *Die Musen und der göttliche Ursprung des Singens und Sagens* (1954), tr.it. *Le Muse e l'origine divina della parola e del canto*, a cura di S. Mati, Roma, Fazi Editore , 2005
- [17] PLATONE: *Fedone*, Bari, Laterza, 1988.
- [18] ID.: *Fedro*, Bari, Laterza, 1989.
- [19] ID.: *Filebo*, Bari, Laterza, 1989.
- [20] ID.: *Ione* , Roma-Bari, Laterza, 2003.
- [21] ID.: *Leggi*, Roma- Bari, Laterza , 2001.
- [22] ID.: *Meneseno*, Roma-Bari , Laterza, 2003.
- [23] ID.: *Protagora*, Roma- Bari, Laterza, 2003.
- [24] ID.: *Repubblica*, Bari, Laterza, 1991.
- [25] ID.: *Simposio*, Bari, Laterza, 1989.
- [26] ID.: *Timeo*, Bari, Laterza, 1991.
- [27] PLUTARCO: *De Musica*, tr. it. a cura di R. Ballerio, Milano, BUR, 2001.
- [28] REALE G.: *Platone. Alla ricerca della sapienza segreta*, Milano, BUR, 2005
- [29] REALE G.: *Storia della filosofia antica*, Milano, Vita e Pensiero, 1985, vol. II.
- [30] SPITZER L.: *Classical and Christian Ideas of World Harmony*, tr.it. *L'armonia del mondo, storia semantica di un'idea*, a cura di V. Poggi, Bologna, Il Mulino, 2006.

# LA MICROPOLIFONIA LIGETIANA COME MODELLO DI COMPOSIZIONE ALGORITMICA

Marcella Mandanici

Conservatorio di Musica "Luca Marenzio" - Brescia, Italia  
marcella@mandanici.com

## SOMMARIO

Questo articolo prende le mosse dall'analisi di un particolare caso di concezione polifonica nella musica del secondo Novecento, la cosiddetta "micropolifonia" Ligetiana. György Ligeti ha utilizzato l'esperienza estetica e percettiva delle particolari tessiture sonore della musica elettronica per trasferirle nella sua scrittura strumentale. Questo ha reso la sua produzione musicale particolarmente originale ed innovativa e ha fatto di lui una delle figure più interessanti del panorama musicale europeo del secondo dopoguerra.

L'esempio musicale principale scelto per l'analisi è tratto dalla composizione *Lux Aeterna*, scritta nel 1966. Il brano è analizzato, reinterpretato e messo in relazione con altri esempi tratti dalla produzione contemporanea dello stesso autore. Dall'analisi vengono derivate tutta una serie di caratteristiche algoritmiche utili alla riproduzione del modello e sono illustrate alcune possibilità di variazioni che, opportunamente combinate, consentono di scrivere e di eseguire una vera e propria partitura algoritmica. Infine sono presentate alcune considerazioni sull'utilità di questo approccio prospettandone anche futuri ulteriori sviluppi.

## 1. INTRODUZIONE

La musica del Novecento, e in particolare quella prodotta dalle avanguardie successive alla Seconda scuola di Vienna [1], ha subito importanti evoluzioni strutturali che hanno portato a concepire la polifonia in modo completamente nuovo.

Partendo dalla rottura degli schemi propri dell'armonia classica e del contrappunto che vedevano la polifonia come un insieme ordinato e distinto di voci, si assiste, per esempio nella scrittura orchestrale, ad un rigonfiamento delle sezioni, ad una moltiplicazione delle parti con conseguente stratificazione e polverizzazione della scrittura strumentale. Infatti, mentre da una parte l'anichilimento dell'orizzontalità tematica delle micro-cellule Weberniane fa sì che le dimensioni melodica e armonica siano perfettamente equidistanti fra loro, dall'altra parte l'eclettismo e il distacco dal materiale tipico

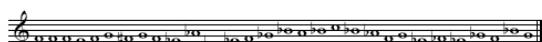
*Copyright: © 2012 Mandanici. This is an open-access article distributed under the terms of the , which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.Creative Commons Attribution License 3.0 Unported*

dello stile di Stravinsky [2], rende perfettamente sovrapponibili entità armoniche senza alcun riguardo alla loro originaria provenienza. Nascono così organismi completamente nuovi come masse pulviscolari, fasce sonore di spessore variabile, sommatorie di micro-cellule in costante movimento, fenomeni materici da misurare statisticamente in termini di densità, velocità di spostamento, coerenza degli elementi interni, etc. Di certo i vari stati verticali della musica d'avanguardia, elettronica e contemporanea che noi oggi sperimentiamo a che spesso denominiamo col termine di "polifonia", poco hanno a che fare col significato originario del termine. Eppure la parola è ancora valida ed esprime la presenza di "oggetti plurimi", risultato dell'azione più o meno coordinata di molte parti, sovrapposizione di diversi strati, universo pulsante di molti organismi.

I due compositori che, all'alba degli anni '60, hanno praticato e teorizzato questo nuovo modo di concepire l'evento musicale, sono stati principalmente Iannis Xenakis [3] e György Ligeti [4]. Partendo dalla critica del determinismo esasperato dovuto all'applicazione delle rigide regole della dodecafonia, Xenakis propone un trattamento di massa dei suoni guidato da calcoli statistici e probabilistici. La rottura radicale degli antichi meccanismi di scrittura così operata, mentre apre le porte al pensiero scientifico e a una concezione pienamente algoritmica della composizione, introduce implicitamente il caos come parte integrante del linguaggio musicale. Più o meno negli stessi anni Ligeti arriva a perseguire lo stesso risultato partendo in un certo senso dal lato opposto rispetto a Xenakis. Mentre quest'ultimo si preoccupa di controllare l'evento musicale dall'esterno, gestendo la macro-struttura della massa sonora, Ligeti arriva a produrre agglomerati di vario spessore ed entità partendo da processi di micro-germinazione interna che assomigliano alla divisione cellulare o al movimento delle amebe. Il preziosismo della scrittura richiama l'antica arte polifonica ma lo sfalsamento dei movimenti, la casualità degli incontri verticali, la gestione puramente quantitativa degli allargamenti/restringimenti intervallari, reintegrano anche qui la componente caotica dell'insieme degli accadimenti. Esaminerò ora più da vicino le caratteristiche di un tipico processo di costruzione di una fascia Ligetiana, prendendo a modello le prime ventiquattro battute di *Lux Aeterna* [5] per coro misto di sedici voci a cappella (1966).

## 2. ANATOMIA DI UNA FASCIA POLIFONICA NELLO STILE DI LIGETI

Quello che descriverò qui di seguito è un processo compositivo riscontrabile assai frequentemente nelle opere di Ligeti, a partire dalle opere dei primi anni '60 (per esempio *Atmosphères* [6] per orchestra, 1961) e che coincide con la tecnica da lui chiamata "micropoli- fonia". Nel caso specifico si tratta di un canone costituito dalla sequenza di note riportata in Figura 1, rispettata da tutte le otto voci che entrano successivamente (quattro soprani e quattro contralti).



**Figura 1.** Note del canone delle prime ventiquattrre battute di *Lux Aeterna*

Ciò che rende il procedimento interessante è lo sfasamento temporale che caratterizza sia le distanze di entrata delle varie voci, sia le durate delle varie note, sia le pause fra una nota e l'altra del canone. Le divisioni ritmiche utilizzate sono le più varie (includendo oltre alla quartina anche cinquine e sestine) e anche la scelta

tempo che può essere analizzata e descritta in base ai seguenti parametri:

- 1) Altezza (o gruppo di altezze) di partenza e di arrivo: partendo dall'unisono sul suono "fa" la fascia si espande nell'arco di tredici battute fino a raggiungere la sua massima densità (contrassegnata nel grafico di Figura 3 dalla striscia tratteggiata, battuta 13), costituita da un accordo di otto suoni nello spazio massimo di una settimana maggiore (re bemolle-do).
- 2) Limiti estremi (acuto e grave) del punto di massima espansione: il grafico di Figura 3 è stato realizzato riportando la presenza dei vari suoni nelle otto voci e nella durata di tutta la battuta. Come si può vedere la fascia non è uniforme né nella crescita né nella distribuzione interna.
- 3) Densità e distribuzione armonica del punto di massima espansione: anche la densità, una volta raggiunto il massimo di otto note a battuta 13, non è costante in quanto continua a oscillare, nelle battute seguenti, fra le sette e le otto note, senza tuttavia diminuire.
- 4) Velocità e modalità dello spostamento: il grafico di Figura 3 indica una stabilità iniziale (prime tre battute), una zona di migrazione (dieci battute dalla 3 alla 13) e il raggiungimento del nuovo stato (pur con modificazioni interne), da battuta 14 a 24.

## LUX AETERNA

$J = 56$ , SOSTENUTO, MOLTO CALMO, „WIE AUS DER FERNE“ \*  
"FROM AFAR" \*

György Ligeti, 1966

Sopr. 1-4:  
stets sehr weich einsetzen / all entries very gentle  
*pp sempre*

Soprano 1: Lux lux lux lux ae - ter -

Soprano 2: Lux lux lux lux ae -

Soprano 3: Lux lux lux lux ae -

Soprano 4: Lux lux lux lux ae -

Alt 1-4:  
stets sehr weich einsetzen / all entries very gentle  
*pp sempre*

Alt 1: Lux lux lux lux ae - ter -

Alt 2: Lux lux lux lux ae -

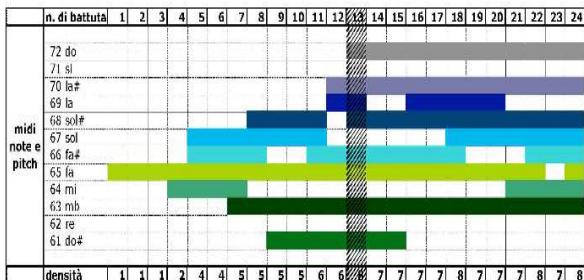
Alt 3: Lux lux lux lux ae -

Alt 4: Lux lux lux lux ae -

**Figura 2.** Partitura della pagina iniziale di *Lux Aeterna*.

delle durate sembra indirizzata all'assoluta non coincidenza delle voci e a costruire una tessitura il più possibile uniforme nella sua più totale imprevedibilità (Figura 2). In sostanza ciò che avviene è la costituzione graduale di una fascia di suoni di densità variabile nel

- 5) Ampiezza del movimento melodico delle parti: come si può vedere dalla sequenza di note di Figura 2, i movimenti melodici sono di piccola entità (prevalentemente nell'ambito della terza minore) tranne attorno al punto minimo (re bemolle).



**Figura 3.** Grafico della densità armonica delle prime ventiquattro battute di *Lux Aeterna*.

6) Tipologia di articolazione strumentale: parametro abbastanza ininfluente nel caso specifico, visto che l'articolazione è unica (nota lunga) e si differenzia soprattutto per la sillaba, ma molto indicativo nel caso di scritture strumentali specifiche (per esempio nella sezione degli archi di *Atmosphères*, lettera C della partitura, dove ogni strumento ad arco esegue un diverso tremolo di terza minore).

7) Durata dei valori ritmici (in rapporto alla densità delle parti): in mancanza di un'analisi particolareggiata dei valori impiegati per le note e le pause che si frappongono fra una nota e l'altra del canone, un semplice paragone visivo fra la prima pagina e la pagina contenente battuta 13, indica come i valori ritmici delle note tendano ad allungarsi mentre la densità dei loro attacchi tenda a diradarsi notevolmente rispetto all'inizio.

### 3. INTRODUZIONE A POÈME SYMPHONIQUE

*Poème symphonique* [7] per cento metronomi (1962) schematizza e riassume assai bene i modi di procedere finora descritti. Se le prime ventiquattro battute di *Lux Aeterna* sono rappresentate da un percorso che parte da un unisono e va verso un accordo, da situazioni armoniche meno dense a situazioni più dense, da eventi frequenti (come intensità di apparizione, dipendente dai valori ritmici impiegati) a eventi più rari (corrispondenti a valori più lunghi), *Poème symphonique*, almeno per alcuni aspetti, rappresenta esattamente il procedimento opposto. Dal cluster indifferenziato dei cento metronomi che partono contemporaneamente si arriva al suono dell'ultimo metronomo superstite (il più lento, secondo una delle modalità esecutive indicate nella partitura di Ligeti), dalla massima densità sonora si va verso la rarefazione, dalla massima frequenza di eventi si va alla minima, fino all'estinzione. Un percorso che si muove dal caos al silenzio attraverso tutta una serie di stadi intermedi in cui al posto delle altezze determinate e dei timbri strumentali ci sono le "voci", la localizzazione spaziale, i differenti timbri e le differenti dinamiche dei diversi metronomi. Le infinite cellule ritmiche risultanti

che affiorano qua e là, sono l'espressione autentica di questa "polifonia casuale" e rappresentano una specie di sguardo profondo nella complessità delle trame così generate:

"... Era una bella idea. Mi ero immaginato il suono di cento metronomi che si fermavano uno dopo l'altro. Cento era solo una cifra approssimativa: pensavo ad un numero sufficiente di metronomi perché il rumore, inizialmente uniforme, desse luogo successivamente a costellazioni ritmiche che mutavano a poco a poco".<sup>1</sup>

*Poème symphonique* dunque è il simbolo massimamente contratto di un gesto compositivo che Ligeti ha declinato molte volte e in vari contesti strumentali ed espressivi. Essendo descritto a parole e non attraverso una partitura, è di fatto una specie di codice verbale che, mancando di una veste musicale definita, chiede di essere interpretato, riscritto, modificato ed evoluto. Occorre non dimenticare che un oggetto musicale come il flusso sonoro originato dai cento metronomi di Ligeti è sicuramente semplificato rispetto al canone di *Lux Aeterna*, non foss'altro perché manca di altezze determinate e dell'elemento armonico, almeno nel senso convenzionale del termine. Tuttavia è un oggetto percettivamente multidimensionale, polifonico perché costituito da cento "parti", poliedrico perché può cambiare molto in base al punto di osservazione dell'ascoltatore, magmatico e attraente perché sfida l'uditivo a riconoscere (e ad immaginare) trame fugaci e irripetibili. Un oggetto che per essere pienamente compreso necessita di forme di visualizzazione che, anche parzialmente (come ad esempio in Figura 4), ne evidenzino almeno le principali caratteristiche spaziali, temporali e cinematiche e che aiutino l'immaginazione ad intravvederne possibili varianti.

### 4. ANALISI E RESINTESI DEI MODELLI LIGETIANI

Partendo quindi dalle caratteristiche algoritmiche dei cento metronomi, verranno in questa sezione descritti tutti i cambiamenti necessari a sintetizzare le prime ventiquattro battute di *Lux Aeterna*, per arrivare successivamente allo sviluppo di ulteriori modelli derivati da quello originale. La visualizzazione dei dati come pure l'implementazione di *Poème symphonique* e di *Lux Aeterna* sono stati realizzati tramite un'applicazione sviluppata dall'autore in Max/MSP atta a riprodurre i modelli studiati.

#### 4.1 Caratteristiche Algoritmiche del Modello dei cento Metronomi

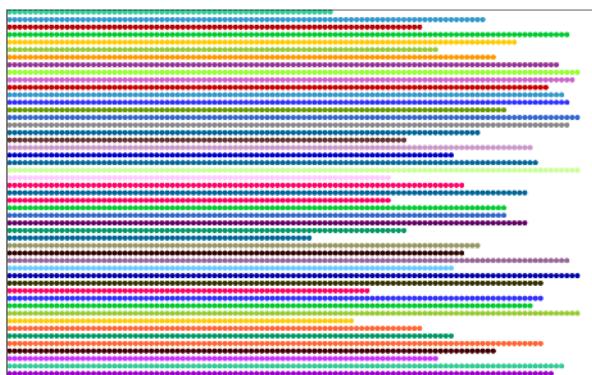
Per sintetizzare il modello dei cento metronomi occorre stabilire quali elementi sono funzionali a una sua ripro-

<sup>1</sup> Tratto da un'intervista di G. Ligeti riportata in "Ilenia Buzzù, Meditazioni sui 100 metronomi" [8]

duzione algoritmica e come essi possono essere modificati al fine di produrne tutta una serie di derivati.

Questi elementi sono:

- 1) Innanzitutto un metodo per generare il suono del metronomo (suono campionato o di sintesi).
- 2) Un meccanismo di temporizzazione che determini la ripetizione del suono in maniera continuativa nel tempo.
- 3) Un sistema che provveda ad avviare e a fermare l'esecuzione. In Figura 4 è rappresentato il risultato finale di un'esecuzione di *Poème symphonique*. Ogni riga indica il percorso di un metronomo il cui colore è scelto casualmente al fine di differenziare le diverse tracce. Sull'asse delle "y" sono ordinati i percorsi dei cento metronomi in dipendenza del loro indice, sull'asse delle "x" il numero di battiti di ogni metronomo, rappresentati graficamente da una piccola sfera disegnata utilizzando l'oggetto "lcd" di Max/Msp.



**Figura 4.** Visualizzazione dell'esecuzione di *Poème symphonique* di Ligeti.

In un metronomo reale le durate dell'esecuzione è inversamente proporzionale al valore di "BPM"<sup>2</sup>: maggiore è il "BPM" più breve è la durata. Nella visualizzazione i metronomi corrispondenti alle linee più lunghe sono i più veloci (quelli che hanno un maggior numero di impulsi) e quindi quelli che si fermano prima, mentre i più lenti hanno un minor numero di impulsi e perciò si fermano per ultimi. Nei vecchi metronomi con l'esaurimento della carica si verifica anche un effetto di rallentamento del "BPM", particolarmente utile a realizzare il desiderio dell'autore:

"...Mi ero immaginato il suono di cento metronomi che si fermavano uno dopo l'altro." [8]

Perciò, una volta esauritosi il tempo di esecuzione di ogni metronomo, si attiva un meccanismo che provvede ad effettuare un rallentamento progressivo fino ad arrivare a 30 "BPM" nel tempo di 20 secondi. Alla fine di questi il metronomo si ferma.

Questo meccanismo è molto interessante in quanto crea dinamicità all'interno dell'algoritmo. Analizzando infatti il decorso temporale del "BPM" senza il rallentamento, abbiamo un dato statico: una volta che il metro-

<sup>2</sup> "BPM", acronimo di "beats per minute" (numero di pulsazioni al minuto).

nomo è partito esegue i battiti sempre alla stessa velocità. Col meccanismo di rallentamento la funzione che descrive il "BPM" di ciascun metronomo comprende invece due segmenti rappresentabili attraverso quella che viene chiamata una "breakpoint function"<sup>3</sup>. La "Bpf" è una funzione definita mediante una serie di segmenti di retta. Ogni segmento è rappresentato da un punto iniziale e da un punto finale determinati a loro volta da due coordinate cartesiane "x" e "y". La funzione sarà quindi descritta attraverso tante coppie di dati quanti saranno i suoi segmenti (Figura 5).



**Figura 5.** "Breakpoint function" della variabile "BPM".

Nel caso di Figura 5 il "BPM" parte a una data velocità (130-0), rimane stabile per una certa durata, qui denominata "begdur" (durata iniziale), scaduta la quale il "BPM" comincia a muoversi verso un nuovo valore (30), in un dato tempo, qui denominato "linedur" (durata di spostamento). I parametri della funzione così descritta sono quindi :

parametri	coordinate
"BPM" (valore iniziale)	130 - 0
"begdur" (durata in cui il valore iniziale si mantiene stabile)	130 - begdur
"targetBPM" (valore finale) e "linedur" (durata di migrazione del valore)	30 - linedur

**Tabella 1.** parametri di definizione della "Breakpoint function" del "BPM"

La funzione utilizzata effettua un'interpolazione lineare qui impiegata per soli motivi di praticità e a scopo esemplificativo. Un'accurata analisi del modello potrebbe portare a scegliere altri tipi di interpolazione (interpolazione logaritmica o esponenziale)<sup>4</sup>.

Oltre alla "Bpf" relativa al "BPM", l'algoritmo dei cento metronomi contiene un'altra importante funzionalità: il ritardo iniziale. Questo dato viene stabilito casualmente all'interno di un determinato intervallo di valori, sufficientemente largo per consentire una certa dif-

<sup>3</sup> Letteralmente "funzione a linea spezzata", d'ora in poi "Bpf".

<sup>4</sup> L'utilizzo di varie funzioni di interpolazione e di distribuzione probabilistica nella generazione di eventi stocastici rientra nella nota teoria di composizione algoritmica per maschere di tendenza formalizzata da G.M. Koenig (Project I e II, 1964-1966) e successivamente utilizzata anche da B. Truax (POD Programs, 1977).

ferenziazione del momento di attacco dell'esecuzione. Concepito per evitare che l'attacco dei cento metronomi sia perfettamente coincidente, producendo un picco di intensità iniziale molto alto, il ritardo della partenza (con conseguente spostamento del punto di arrivo) sarà poi interpretato - negli algoritmi seguenti - come un mezzo efficace per distribuire in modo più fluido l'esecuzione delle varie istanze e in definitiva per dosare il grado di intensità dell'evento complessivo.

- 4) Un sistema che, una volta stabilite le caratteristiche del singolo metronomo (o strumento), le moltiplicherà per il numero di metronomi (o voci, o istanze) provvedendo così a generare la polifonia. I parametri utili a definire ogni singolo metronomo sono: a) il valore di "BPM" scelto all'interno di una fascia che va da 40 a 200, il ritardo iniziale (dai 50 ai 300 ms) e le caratteristiche timbriche del suono del metronomo, in base a piccole differenze che i vari strumenti possono avere come frequenza, inviluppo, dinamica e posizione nello spazio.
- 5) Un meccanismo automatico che provveda ad assegnare alle varie istanze (realizzate in Max/Msp utilizzando l'oggetto "poly~"), la serie di valori stabilita in un determinato intervallo per ogni parametro richiesto.

#### 4.2 Da Poème symphonique a Lux Aeterna

Naturalmente la "Bpf" può essere rappresentata da una linea spezzata con più di due segmenti e può essere applicata a qualsiasi parametro esecutivo. Pensando per esempio a un'esecuzione "Midi"<sup>5</sup>, si può prevedere una "Bpf" per il valore di "midi pitch", "velocity" e durata.

Per riprodurre una composizione vocale come *Lux Aeterna*, l'algoritmo dei cento metronomi può essere così modificato:

- 1) La sintesi del suono del metronomo è sostituita dall'esecuzione di campioni vocali dipendenti da messaggi MIDI, con "midi pitch" e "velocity" controllati ognuno da una sua "Bpf".
- 2) Il meccanismo di temporizzazione - sostanzialmente un metronomo - è controllato da un valore di "BPM" dipendente anch'esso dalla sua "Bpf".
- 3) L'avvio dell'esecuzione è dipendente da un valore di ritardo, utile a sfasare la partenza delle varie istanze. La fermata dell'esecuzione avviene arrestando il metronomo a conclusione del tempo complessivo assegnato alla durata dell'evento.
- 4) Le istanze, simulando la disposizione delle voci di un coro, sono gestite a gruppi e in numero dipendente da quante sono le voci che partecipano all'evento. *Lux Aeterna* però non è costituita dal suono di un metronomo, cioè dallo stesso suono ripetuto a intervalli regolari, ma da un procedimento a canone, il che implica che ogni voce segua un preciso percorso melodico. Tuttavia, c'è una certa differenza fra voce fisica e le varie istanze dell'algoritmo. Innanzitutto ogni voce fisica ha una sua caratteristica timbrica che la differenzia dalle altre; oltre a ciò l'articolazione sillabica

<sup>5</sup> L'acronimo "MIDI" sta per "Musical Instrument Digital Interface"

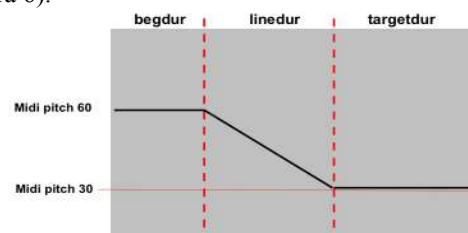
rende evidente l'ingresso di una parte anziché di un'altra, senza contare il fatto che, come è bene evidenziato nell'analisi di J. Jarvlepp [9], vi è una precisa corrispondenza fra nota del canone e sillaba pronunciata. Tutto ciò tuttavia va un po' al di là dell'obiettivo che si vuole raggiungere. Ciò che importa infatti non è la riproduzione fedele del modello, ma la comprensione e l'implementazione di ciò che del modello è essenziale. Quindi l'approccio prescelto non è tanto quello di considerare la linea melodica delle singole parti ma di considerare il percorso armonico complessivo, definendolo in termini di punti di partenza e di arrivo della "Bpf".

5) Il meccanismo di assegnazione dei valori è particolarmente importante per il parametro di "midi pitch", poiché la gestione delle altezze deve prevedere diverse possibilità oltre la semplice migrazione da nota a nota, come si può vedere nello schema riportato in Tabella 2.

Punto di partenza	Punto di arrivo
Singola nota	Accordo/cluster
Singola nota	Singola nota
Accordo	Singola nota
Accordo	Accordo/cluster
Cluster	Accordo/cluster
Cluster	Singola nota

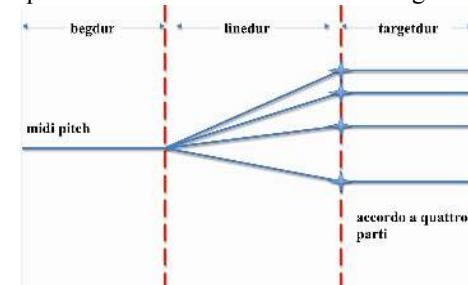
**Tabella 2.** Schema delle possibilità di percorso dialettico della "Bpf" relativa al "midi pitch".

6) La "Bpf" inoltre può essere modificata aggiungendo alla funzione un altro punto relativo al tempo di esecuzione della condizione di arrivo (targetdur, vedi Figura 6).



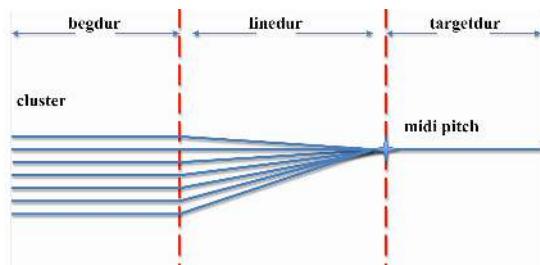
**Figura 6.** "Breakpoint function" con tre linee spezzate.

Una "Bpf" di un percorso "midi pitch-accordo" potrebbe quindi essere visualizzata come in Figura 7:



**Figura 7.** "Breakpoint function" di un evento "midi pitch – accordo".

Una "Bpf" di un percorso "cluster - midi pitch" invece appare come in Figura 8:



**Figura 8.** "Breakpoint function" di un evento "cluster - midi pitch".

#### 4.3 La Resintesi della Micropolifonia di *Lux Aeterna*

Come evidenziato dall'analisi di *Lux Aeterna* descritta nella sezione n. 2, le prime otto voci partono da una nota singola ("midi pitch" 65 - fa sopra il do centrale) e arrivano a un cluster re bemolle - do ("midi pitch" da 61 a 52). Per quanto riguarda la temporizzazione, ci sono quasi 13 secondi di "begdur" (corrispondenti a tre battute e mezza circa della partitura), 36 secondi di "linedur" (batt. 4-13) e 52 secondi di "targetdur" (batt. 13-24) per un totale di 1' e 43" di esecuzione (cfr. Figura 3). Questi dati sono i parametri di definizione della "Bpf" rispetto alle altezze; rispetto alla "velocity" non c'è "Bpf" poiché il dato è statico (pianissimo).

Il "BPM" utilizza gli stessi parametri temporali della "Bpf" relativa al pitch. La fascia di valori da assegnare ad ogni istanza è stata scelta considerando i limiti di frequenza della comparsa di un nuovo evento nelle prime battute delle varie voci. Poiché la densità degli eventi va diradandosi, anche il valore di "BPM" va progressivamente diminuendo.

Il ritardo iniziale, fondamentale in questo caso poiché le voci partono molto distanziate l'una dall'altra, va da un minimo di 0 ms a un massimo di 5890 ms (corrispondente alla durata della pausa che precede l'ultima entrata del quarto contralto a 56 di metronomo), precisando che tutti i valori sono assegnati a caso all'interno dell'intervallo stabilito. Ecco la visualizzazione di uno dei possibili risultati dell'algoritmo delle prime ventiquattro battute di *Lux Aeterna* appena descritto:

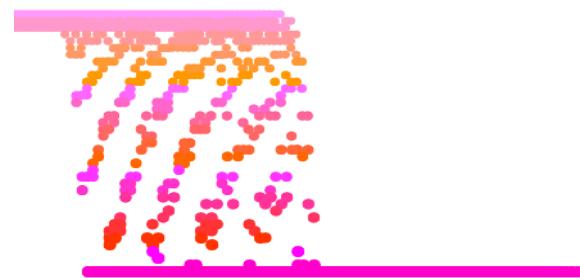


**Figura 9.** Visualizzazione dell'algoritmo delle prime ventiquattro battute di *Lux Aeterna*.

Sull'asse delle "y" è rappresentato il "midi pitch" delle varie voci, sull'asse delle "x" è rappresentato il loro movimento nel tempo. Sono bene evidenti le tre fasi temporali della "Bpf" e il risultato della migrazione delle altezze. Inizialmente non si vede il ritardo iniziale delle singole istanze, dal momento che esse partono tutte dalla stessa nota. Più avanti però se ne notano le conseguenze osservando la sfasatura del movimento di migrazione verso il suono obiettivo, nonostante che i parametri temporali della "Bpf" siano gli stessi per tutte le istanze.

#### 4.4 Variazioni su Modello

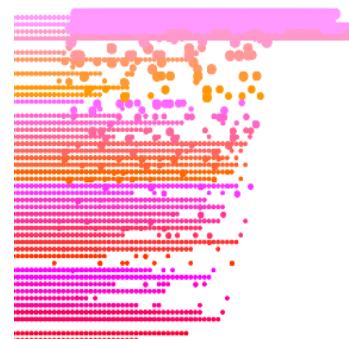
Una volta definito il modello algoritmico per l'esecuzione della micropolifonia Ligetiana, si aprono diverse possibilità compositive. Come prima cosa si può agire sui parametri di definizione della "Bpf" relativi al "midi pitch", realizzando alcune delle diverse tipologie indicate in Tabella 2.



**Figura 10.** Visualizzazione di un movimento cluster-singola nota"

Nell'esempio di Figura 10 si ha la visualizzazione del risultato di un movimento da un cluster acuto a una nota singola grave. In questo caso il valore di "begdur" è piuttosto basso mentre il "linedur" è piuttosto lungo.

C'è anche un piccolo aumento della "velocity" visualizzabile nell'aumento delle dimensioni dei punti colorati che rappresentano le note.

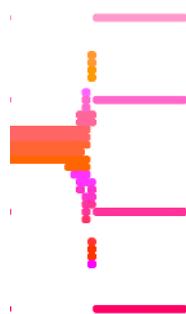


**Figura 11.** Visualizzazione di un movimento cluster-cluster"

L'esempio di Figura 11 mostra la visualizzazione di un movimento che parte da un cluster a largo spettro,

comprendente cioè un gran numero di suoni su un registro molto ampio, per andare verso un cluster ristretto nell'ambito di pochi suoni raggruppati nel registro acuto. La "velocity" ha anch'essa un'escursione piuttosto ampia, passando la dinamica musicale dal piano al forte.

Infine, l'esempio di Figura 12 mostra un percorso che parte da un cluster nel registro medio e che arriva a un accordo. L'aspetto interessante di questo esempio riguarda però il concetto di pendenza. La pendenza in un grafico rettilineo è il rapporto fra la variazione dell'ordinata "y" (pitch) rispetto alla variazione dell'ascissa "x" (tempo). Nel nostro caso è evidente che il parametro da cui dipende la pendenza del movimento del "midi pitch" è la durata dello spostamento ("linedur").



**Figura 12.** Visualizzazione di un movimento cluster-acordo”

La pendenza incide molto sulla percezione dell'evento. Infatti, in una situazione con "linedur" molto breve, oltre allo schiacciamento temporale dello spostamento, si possono verificare dei "buchi" intervallari nel percorso delle singole voci, poiché la velocità di lettura non consente il rilevamento di tutti i gradini (semitoni), come si può bene osservare nell'esempio di Figura 12. Al contrario, in una "Bpf" con "linedur" molto lungo (esempio di Figura 11) c'è la possibilità di percepire molto di più tutti gli stati armonici di transizione. Per sviluppare altre idee partendo sempre dallo stesso modello algoritmico si può immaginare un movimento in un certo senso opposto a quello originario. Se il modello di Ligeti è una fascia più o meno densa che si muove o verso un'altra fascia o verso un punto in modo più o meno veloce, il suo contrario potrebbe essere un evento “puntuale” molto breve e senza evoluzione nel tempo (un accordo o una nota singola o amplificata su diverse ottave, per esempio), che comunque può sempre essere considerato come un caso limite dell'evento originale in versione massimamente compressa. Ma la parte più creativa è sicuramente la combinazione e la sovrapposizione di tutti questi elementi compositivi che possono dare luogo a una vera e propria partitura esecutiva costituita sostanzialmente da eventi continui (nota statica, fascia statica, fascia in movimento, etc.) e da eventi istantanei (accordi, note brevi, veloci portamenti, etc.) prefigurando una situazione assimilabile al principio “Gestaltico” della figura-sfondo [10].

## 5. CONSIDERAZIONI FINALI

In questo paragrafo esporrò alcune valutazioni sulla composizione algoritmica, sul metodo sopra illustrato e sui suoi possibili utilizzi, prospettando anche qualche ipotesi di ricerca futura.

### 5.1 La Composizione Algoritmica

Il passaggio dall'individuazione, all'analisi e successiva implementazione di un modello di composizione algoritmica non è operazione né banale né scontata. Esso implica infatti una profonda consapevolezza del linguaggio musicale e delle sue caratteristiche ed ha quindi un grande valore conoscitivo, analitico e interpretativo.

Implementare un algoritmo compositivo significa inoltre individuare le caratteristiche ricorrenti del modello, costruendo una macchina atta non a riprodurre il modello stesso, ma a fornirne un numero infinito di diverse versioni. Ciò implica arrivare al cuore del processo creativo, sottraendo l'atto compositivo al controllo esclusivo del compositore. Significa anche conquistare una piena padronanza del risultato musicale che, dipendendo da variabili note, può essere manipolato con piena consapevolezza anche da persone senza alcuna preparazione musicale.

### 5.2 Usabilità dei Modelli Compositivi

La costituzione attraverso l'analisi di un repertorio di modelli di composizione algoritmica può consentire la possibilità di esplorazione da parte di musicisti non compositori di mondi sonori ancora poco conosciuti ai più. Inoltre, l'individuazione di strutture di controllo come la “Bpf” incoraggia l'accesso alla manipolazione di processi compositivi complessi altrimenti irraggiungibili dall'utente non acculturato musicalmente. Infine, la possibilità di integrare la composizione algoritmica in nuovi strumenti elettronici da utilizzare con sistemi multi-touch o di tracciamento del movimento<sup>6</sup> può costituire un notevole arricchimento espressivo della produzione musicale in questi ambienti.

## 6. UN'IPOTESI DI RICERCA: COMPORRE TRAMITE IL GESTO

Il compito della “Bpf” è di disegnare una direttrice lungo la quale eventi caotici e casuali si muovono andando da un punto ad un altro in un dato tempo secondo una determinata maschera di tendenza. La parola “disegnare” conduce immediatamente al concetto di gesto. Poiché i sistemi di tracciamento danno la possibilità di tradurre il movimento in dato numerico, ciò significa che un movimento della mano può tracciare una “Bpf” e che

<sup>6</sup> Si veda ad esempio Disembodied voices: a kinect virtual choir conductor (2012) (<http://www.youtube.com/watch?v=oYf7GrMMrL8>) (ultima consultazione 14/10/2012)

questa funzione può essere utilizzata per pilotare interattivamente un algoritmo compositivo.

## 6.1 La Dimensione Temporale

Un problema è costituito dalla dimensione temporale in cui ciò dovrebbe avvenire. Il gesto umano è fisiologicamente eseguito all'interno di determinate durate mentre gli algoritmi compositivi possono richiedere anche tempi che vanno molto al di là di questi limiti (per esempio una "Bpf" può durare anche diversi minuti). Ciò potrebbe richiedere un sistema di archiviazione della funzione e un sistema di assegnazione del suo tempo di esecuzione. Questo potrebbe a tutti gli effetti essere considerato come un sistema interattivo di composizione tramite il gesto che permette di provare varie soluzioni, descrivendo nuove funzioni e ascoltando il risultato in un'esecuzione successiva.<sup>7</sup>

## 6.2 Relazione tra Gesto e Risultato Musicale

Un altro problema riguarda il grado di coerenza fra il gesto e il risultato musicale. Storicamente i gesti a mano libera nella musica sono stati utilizzati dai direttori di coro e d'orchestra per dare gli attacchi, scandire il tempo dell'esecuzione e per comunicare agli esecutori la dinamica e le intenzioni espressive del direttore. Se si escludono gli esempi tratti dalla chironomia gregoriana [12] e dalla simbologia impiegata da Zoltan Kodály per l'intonazione delle altezze nel canto corale<sup>8</sup>, non esiste nella tradizione musicale alcun riferimento ad indicazioni gestuali che riguardino strutture organizzative della musica, essendo questa una funzione da sempre delegata al compositore ed espressa nella partitura<sup>9</sup>.

## 6.3 Contiguità con altri tipi di Interazione Gestuale

Si è già detto come l'interazione gestuale sia stata da sempre un metodo efficace per imprimere carattere ed espressività ad un'esecuzione musicale, come testimoniato anche recentemente da varie esperienze compiute nel campo della direzione virtuale<sup>10</sup>.

<sup>7</sup> Protocolli di interazione esecutiva (non compositiva) fra gesto e partitura sono descritti ad esempio nell'ambito del progetto Inedit ([http://www.ircam.fr/306.html?&tx\\_ircamprojects\\_pi1\[pType\]=p&tx\\_ircamprojects\\_pi1\[showUid\]=81&cHash=b2d11a32f74336b463049ae5301ebe5e](http://www.ircam.fr/306.html?&tx_ircamprojects_pi1[pType]=p&tx_ircamprojects_pi1[showUid]=81&cHash=b2d11a32f74336b463049ae5301ebe5e)) (ultima consultazione 14/10/2012)

<sup>8</sup> <http://www.britannica.com/EBchecked/topic/320927/Zoltan-Kodaly> (ultima consultazione 6/10/2012)

<sup>9</sup> Controlli gestuali ispirati alla chironomia sono stati utilizzati in Chorus digitalis ([http://vida.limsi.fr/doku.php?id=wiki:projet\\_chorus\\_digitalis](http://vida.limsi.fr/doku.php?id=wiki:projet_chorus_digitalis)) (ultima consultazione 14/10/2012). Si tratta però anche qui di controlli esecutivi e di sintesi della voce cantata, non di composizione interattiva.

<sup>10</sup> Si vedano in proposito, fra i vari esempi, Personal orchestra di J. Borchers (2002) (<http://hci.rwth-aachen.de/actibits/po/virtualconductor.html>) o

Assegnare al gesto una funzione compositiva oltre che espressiva significa far coesistere in uno stesso spazio funzioni diverse, rischiando sovrapposizioni e confusioni. Sono dunque necessarie approfondite analisi e test di usabilità che, partendo da schemi già noti, estendano il repertorio e la funzionalità della gestualità tradizionale alla conquista non solo dell'espressività ma anche degli aspetti strutturali più profondi della musica.

## 7. BIBLIOGRAFIA

- [1] O. Neighbour, G. Perle, P. Griffiths: *La seconda scuola di Vienna. Schoenberg, Berg, Webern da The new grove*, Universal Music MGB, 1986.
- [2] Aa. Vv.: *Stravinskij*, a cura di G. Vinay, Il Mulino, Bologna, 1992.
- [3] Aa. Vv.: *Xenakis*, a cura di E. Restagno, EDT, Torino, 1988.
- [4] Aa. Vv.: *Ligeti*, a cura di E. Restagno, EDT, Torino, 1986.
- [5] György Ligeti: *Lux Aeterna*, for mixed choir a cappella, Edition Peters, 1966.
- [6] György Ligeti: *Atmosphères*, for orchestra, UE Universal Edition, 1961.
- [7] György Ligeti: Poème Symphonique, 1962, Fluxus debris, <http://www.artnotart.com/fluxus/gligeti-poemesymphonique.html> (ultima consultazione 6/10/2012)
- [8] Ilenia Buzzi: "Meditazioni sui 100 metronomi", La Critica, 2008, [http://www.lacritica.net/ilenia\\_buzzi.htm](http://www.lacritica.net/ilenia_buzzi.htm) (ultima consultazione 5/10/2012)
- [9] Jan Jarvlepp: "Pitch and Texture Analysis of Ligeti's Lux Aeterna", Ex tempore, <http://www.extempore.org/jarvlepp/jarvlepp.htm> (ultima consultazione 5/10/2012)
- [10] E. Rubin: *Figure and Ground* in S. Yantis (Ed.), Visual Perception, Philadelphia, Psychology Press, 2001, pp. 225-229.
- [11] M. Mandanici, S. Sapir: "Disembodied voices: a kinect virtual choir conductor," *SMC 2012 Proceedings*, pp. 271 – 276.
- [12] J.R. Carroll: *The technique of Gregorian Chironomy*, Church Music Association of America, (2009)
- [13] S. Le Beux, L. Feugère, C. d'Alessandro: "Chorus Digitalis": experiment in chironomic choir singing", *Proc. Interspeech 2011*.

il sistema commerciale Wii Music (2008) (<http://www.youtube.com/watch?v=itMOykSfutQ&feature=related>) (ultima consultazione 6/10/2012)

# MULTIMODALITÀ E IMMERSIONE: VERSO UNA RILETTURA DEL MINIMALISMO ATTRAVERSO LA STORIA DELLE INSTALLAZIONI SONORE

Giacomo Albert

Università degli Studi di Torino

gialb@hotmail.com

## ABSTRACT

Nel saggio è analizzata la struttura modale di un gruppo di installazioni sonore create a partire dalla fine degli anni Sessanta, che si rifanno a un modello simile. Tali opere, cui capostipite è la *Dream House* di La Monte Young, sono per lo più attribuibili al panorama culturale del cosiddetto “minimalismo” americano. Sono composte da una ridotta quantità di materiale, adoperano tecniche di sintesi e diffusione semplici, lavorano sovente sull’elaborazione di principi acustici elementari, nonché sull’amplificazione di effetti psicoacustici e sono basate su tecniche di privazione sensoriale. Si osserverà come nel panorama coeve delle installazioni sonore il gruppo di opere qui analizzate costituisca una novità e un insieme conchiuso, dotato di caratteristiche comuni, che lo differenziano rispetto alla produzione precedente e a quella coeva legate ad altri ambienti culturali, ad altri stili. Si metterà in luce come le caratteristiche peculiari della struttura modale di tali opere siano la multimodalità e l’immersione: la manipolazione non-mediata della percezione del fruttore, che viene qui intesa non come la somma dei diversi modi, ma come l’interazione di essi. Tali proprietà, a ben vedere, caratterizzano la produzione minimalista e post-minimalista e sono sfruttate anche in generi diversi da quello installativo, come la *virtual art* e la musica da film.

## 1. INTRODUZIONE

*In the late sixties and seventies, Eye and Spectator negotiate some transactions. Minimal objects often provoked perceptions other than the visual. Though what was there instantly declared itself to the eye, it had to be checked: otherwise, what was the point of three-dimensionality? There are two kinds of time here: the eye apprehended the object at once, like painting, then the body bore the eye around it. This prompted a feedback between expectation confirmed (checking) and hitherto subliminal bodily sensation. Eye and Spectator were not fused, but cooperated for the occasion. The finely tuned Eye was impressed with some residual data from its abandoned body (the kinesthetics of gravity, tracking, etc.). The Spectator’s other senses, always there in the raw, were infused with some of the Eye’s fine discriminations. The Eye*

*Copyright: © 2012 Albert. 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.*

*urges the body around to provide it with information – the body becomes a data-gatherer. There is a heavy traffic in both directions on this sensory highway – between sensation conceptualized and concept actualized. In this unstable rapprochement lie the origins of perceptual scenarios, performance and Body Art. ([24] pp. 50-52)*

Brian O’Doherty nel 1976, nel secondo dei rivoluzionari articoli incentrati sugli spazi dell’arte contemporanea poi raccolti in *Insight the White Cube*, definiva così le specificità della fruizione dell’arte minimalista. Nel saggio sarà analizzato il rapporto tra opera e fruizione in alcune installazioni sonore, facendo riferimento ai concetti qui delineati. L’elemento più interessante da rilevare della citazione proposta è l’attenzione verso il lavoro di negoziazione tra Occhio e Spettatore, dunque tra singolo modo di percezione e intero corpo, che viene messo in atto anche in un’arte monomediale come la pittura minimalista: caratteristica importante a tal punto da essere condizione necessaria per la comprensione dell’opera stessa. Si tratta di un lavoro di negoziazione capace di ingenerare quella che Michael Fried aveva definito quale *teatralità* [7]<sup>1</sup> dell’opera plastico-pittorica di stampo minimalista, che ingenera una dimensione temporale immanente anche a opere di tipo plastico: una temporalità legata alla fruizione, ma nel contempo inscritta nello statuto dell’opera. Secondo l’interpretazione di O’Doherty, dunque, gli oggetti minimalisti provocano percezioni ulteriori rispetto a quella visiva e la loro fruizione si differenzia rispetto alla contemplazione/ interpretazione tipica di un’opera classica: cambiano le modalità per la fruizione dell’opera nel suo complesso, non solo i simboli da interpretare, ma le stesse forme della conoscenza stimolate.

Fulcro dell’interesse del presente saggio è dunque la struttura modale di alcune installazioni sonore che si richiamano al medesimo ambito culturale. Si intende osservare le modalità di fruizione implicate da tali opere, che condividono un modello simile. A tal fine si farà riferimento in particolar modo a due concetti: la multimodalità (cfr. [20] pp. 63-66), ossia l’integrazione dei modi di percezione in un’esperienza unitaria che interessa l’intero corpo del fruttore e può aver luogo in un contesto mono- o multimediale; e l’immersione (cfr. [10, 22, 5, 1]), una tipologia di fruizione basata sull’appartenenza del fruttore e dell’opera al medesimo

<sup>1</sup> Per una panoramica più ampia sul concetto friediano di *theatricality*, vedi [6].

mondo, che viene da questi percepito come naturale, sebbene sia artificiale, tecnologico e composto, un'esperienza dell'opera non mediata da forme di rappresentazione.

## 2. LE INSTALLAZIONI MINIMALISTE

### 2.1 Il panorama coevo

Il genere installazione sonora ha origine nella prima metà del Novecento; almeno dal 1955, a partire da *Work (Bell)* di Atsuko Tanaka, si può parlare in maniera compiuta di una nuova forma artistica incentrata sulla dimensione sonora e nel contempo basata su una concezione spaziale che include la presenza del fruitore nello spazio e nello statuto dell'opera.

I primi anni di storia del genere, compresi tra la fine degli anni Cinquanta e la prima metà degli anni Sessanta, vedono susseguirsi una variegata serie di opere, che afferiscono a modelli differenti: installazioni interattive (da *Work (Bell)* di Atsuko Tanaka alla *Caverna dell'Antimateria* di Gallizio), installazioni concorrenti (il *Philips Pavilion* di Le Corbusier, Xenakis e Varèse), installazioni-assemblage (da *Oracle* di Robert Rauschenberg a *Destroyed Piano with 28 Documents* di Nam June Paik), *Environment (Words)* di Allan Kaprow), "esposizioni musicali" (*Exposition of Music. Electronic Television* di Paik), le "stanze musicali" e le stanze risonanti (da *Das Urteil* di Dieter Schnabel alle numerose *Raummusiken* di Stockhausen), sonorizzazioni di mostre (*La casa abitata* di Pietro Grossi), le installazioni partecipative (da *Word event fluxusversion 1* di George Brecht a *Public opinion descends upon demonstrators* di Robert Ashley) e relativi sottogeneri come gli ibridi tra installazione e happening (come il *Technological-Happening-Room* di Wolf Vostell), le installazioni di *land art* sonora (*Semaforo* di Akira Kanayama), etc. Tutti sottogeneri non ben delimitati, né tanto meno definiti nei confini formali; generi e sottogeneri che si incrociano in un panorama vasto e variegato di produzioni interessate a mettere in relazione la dimensione sonora con lo spazio e la partecipazione del fruitore, in molteplici direzioni. All'interno di questa poliedricità di modelli irrompe nel corso degli anni Sessanta un nuovo modello, l'installazione minimalista.

Si può definire mediante tale locuzione un gruppo di opere, che condividono alcune peculiarità, che analizzeremo di seguito. L'individuazione di tale insieme non dipende da affermazioni degli artisti o da una loro vicinanza geografica e personale, bensì da proprietà formali e strutturali condivise dalle opere, non soltanto dal punto di vista mediale, ma soprattutto per il modo in cui costruiscono l'esperienza del fruitore.

### 2.2 La *Dream House* di La Monte Young

La prima opera nella storia delle installazioni sonore che corrisponde in tutte le sue caratteristiche con la definizione del sottogenere minimalista è la *Dream House* di La Monte Young, compositore a sua volta ritenuto padre del minimalismo musicale *tout court* (ad es. in [32] p. 13 e [23] pp. 138-144). Con *Dream House* non si intende unicamente l'opera compiuta, ma anche le diverse versioni di installazioni performative che la pre-

cedono. Il compositore stesso, infatti, quando parla della *Dream House* non si riferisce esclusivamente alle versioni dell'opera successive al 1966-69, a partire da *Map of 49's dream the two systems of eleven sets of galactic intervals ornamental lightyears tracery*, o alla stessa *Dream House* compiuta, ma anche alla concezione generale dell'ambiente sonoro periodico, basilare nella sua poetica, su cui sono fondate opere precedenti, come *The Four Dreams of China*, oppure *The Tortoise, His Dreams and Journeys* e *The Lower Map of the Eleven's Division from the Romantic Symmetry in Prime Time from 144 to 112 with 129* (vedi [9] e [31] pp. 28-29). D'altronde *Map of 49's dream* viene definita una sottosezione di *The Tortoise, His Dreams and Journeys* ([37] p. 5). Lo stesso artista, inoltre, ha a più riprese affermato che «By 1962 La Monte had formulated the concept of Dream House in which a work would be played continuously and ultimately exists in time as "living organism with a life and tradition of its own..."» ([40] p. 10).<sup>2</sup> Si può in questa sede operare con leggerezza una simile semplificazione, perché fulcro del saggio non sono gli studi sull'intonazione e sui rapporti armonici, che cambiano di volta in volta e costituiscono punti nodali tanto nella poetica e nella prassi artistica di Young quanto nelle opere in questione (cfr. [9, 38, 12]), bensì la concezione generale delle opere stesse, in particolar modo per quanto concerne la struttura modale.

Per indicare la sua opera Young adopera la locuzione «Acoustic Environment» (si vedano [40, 39]). In tal modo ricorre a una parola che, nella teoria artistica degli anni Sessanta, era connotata in maniera chiara: l'*Environment*, infatti, è un genere teorizzato proprio tra la seconda metà degli anni Cinquanta e gli anni Sessanta da Allan Kaprow<sup>3</sup>. Questi lo ha adoperato per classificare opere basate sulla costruzione di uno spazio e di una situazione, create da artisti dell'avanguardia newyorkese come James Dine, Claes Oldenburg e dallo stesso Kaprow. In realtà questi, nei suoi scritti, insiste, anche in riferimento al genere *environment*, sulla dimensione partecipativa della fruizione<sup>4</sup>; Dine e Oldenburg, inoltre, pur meno interessati agli aspetti riguardanti il coinvolgimento dei *participants*, fanno ampiamente ricorso a tale termine ([27] pp. 16-21). In realtà Young non mostra interesse, né nei suoi scritti, né nella sua opera, verso il coinvolgimento del fruitore attraverso strategie partecipative. L'opera di Young, dunque, si discosta dall'estetica kaprowiana, e, pur definita mediante una categoria fortemente connotata, fa riferimento a un significato neutro del termine, a un generico "ambiente acustico" e non a un genere specifico.

Young stesso esplicita la sua concezione dell'"ambiente", in tal modo:

<sup>2</sup> Young, peraltro, esprime un concetto analogo anche in altri scritti, tra cui [39] p. 218 e [41] p. 483.

<sup>3</sup> Particolarmente celebre è [14], volume del 1966, ma la teorizzazione del genere accompagna anche le prime opere dell'artista, fin dal 1958 e dall'*Untitled Environment* esposto il medesimo anno presso la Hansa Gallery di New York.

<sup>4</sup> Già in [16, 15], entrambi saggi del 1958. Per una recente disamina della categoria della partecipazione, vedi [8, 2].

*the concept of an environment directly relates to this concept of the discipline of the body in order to achieve this high-spiritual state, because the various senses must be dealt with. [...] This puts you in a special state that allows you to have visual stimulus that is supportive of the music and the music can be supportive of the visual stimulus. And the entire environment is simply an extension of this same concept.* ([42] p.29)

Dalla citazione si può desumere come secondo l'interpretazione di Young la categoria *Environment* sia legata in primo luogo alla dimensione spirituale-meditativa dell'opera, una tipologia di concezione dell'esperienza artistica che accomuna numerose esperienze intorno al volgere degli Settanta, anche afferenti a scuole apparentemente lontane: basti pensare al lavoro teorico e pratico di Pauline Oliveros (vedi [25] pp. 138-174) o alle modalità della prassi esecutivo-improvvisativa dello Stockhausen di *Aus den sieben Tagen*. Young intende realizzare l'obiettivo, il raggiungimento di uno stato spirituale attraverso la disciplina del corpo, investendo il fruttore con stimoli di diversa natura, ma congruenti, ossia non necessariamente sincronizzati (cosa che non avviene nell'opera), ma con una concezione della temporalità e della fruizione simili: un flusso ridotto, quasi costante e invariante di informazione, che lo induce ad acuire l'attenzione verso i minimi cambiamenti e verso le relazioni tra le poche componenti di cui l'opera è composta. In tal modo il fruttore, esposto per lunghi periodi di tempo al medesimo "ambiente", entra in una relazione più profonda con esso e con i suoi modi di percezione. Il suo obiettivo è quello di far entrare il fruttore nel suono, ossia da una parte dargli la sensazione di essere immerso in esso, dall'altra far sì che la sua attenzione si sposti dall'articolazione formale dei suoni nel tempo, inesistente nell'opera di Young, alla vita interna dei suoni stessi. Afferma, infatti, in un'intervista a Kostelanetz che il suo scopo con la *Dream House* sia:

*trying to get inside a sound [...] There are several ways you can approach it. One is someone concentrates so heavily upon a given sound – he gives himself over to it such a degree – that what's happening is the sound. Even though I could be sitting here, all I am is an element of the sound. Another approach is to walk into an area in which the sound is so abundant that you actually are in a physical sound environment. This happens when someone walks into one of my concerts [...] you could be enveloped.* ([17] p. 35)

Per tale motivo, dunque, diffonde in un ambiente in maniera ininterrotta, virtualmente infinita, una serie limitata di sinusoidi. Queste concorrono a formare un timbro continuo, solo molto lentamente cangiante (le sinusoidi cambiano gradualmente e in maniera impercettibile); un timbro composto da relazioni periodiche tra le componenti, anche se con una *ratio* elevata<sup>5</sup>, che

<sup>5</sup> Con "ratio elevata" si indicano parziali acute, frequenze multiple di una fondamentale comune grave. La periodicità dell'evento sonoro risultante è dunque espressa attraverso un rapporto complesso tra le frequenze; in tal modo la percezione di tale periodicità è resa più difficile e richiede un maggior tempo di esposizione al suono di quanto non richiederebbe un evento sonoro continuo e stabile in cui i rapporti tra le parziali siano determinati mediante proporzioni più semplici.

concorre a fornire una caratterizzazione specifica al suono, che Young interpreta come espressione di potenzialità emotive, e rende più complessa per il fruttore la percezione della periodicità e della relazione tra le componenti.

L'artista è dunque interessato alle proprietà psicoacustiche di una lunga esposizione al medesimo, semplice, suono, dimostrando, peraltro, una conoscenza non banale dei fenomeni percettivi. Afferma, infatti, in un altro passo:

*The volley theory of pitch perception assumes that a sequence of electrical impulses is sent travelling along specific neurons of the auditory nerve. For frequencies up to 2000 Hz. only, these produce a more or less complete reproduction of the frequency of the vibratory motion of the basilar membrane in the case of a single sine wave and a more or less distorted reproduction of the complete waveform for more complex signals. It is presumed that this reproduction of the complete waveform will be best for sound at lower frequencies and less good for higher frequencies since an individual neuron cannot fire faster than 300 Hz. [...] The assumption of place theory and volley theory suggests that when a specific set of harmonically related frequencies is continuous, as is often the case in my music, it could more definitively produce (or stimulate) a psychological state that may be reported by the listener since the set of harmonically related frequencies will continuously trigger a specific set of auditory neurons which in turn will continuously perform the same operation of transmitting a periodic pattern of impulses to the corresponding set of fixed points in the cerebral cortex. When these states are sustained over longer periods of time they may provide greater opportunity to define the psychological characteristics of the ratios of the frequencies to each other.* ([37] pp. 8-9)

Si interessa dunque nella sua poetica dei dettagli, dei meccanismi di funzionamento dell'apparato uditivo, incentrando l'attenzione sul collegamento tra le relazioni armoniche, i registri e le proprietà psicologiche, in base al processo di trasduzione della vibrazione in stimolo nervoso. Interpreta inoltre l'opera come primo esperimento sugli effetti di una lunga esposizione del pubblico a un ambiente acustico composto da onde continue periodiche non variante nel tempo, non menzionando dunque per la sua opera una funzione artistica, ma esclusivamente scientifico-sperimentale ([37] p. 6).

In altri passi specificherà che il suo lavoro, basato sul meccanismo spiegato nella citazione appena riportata, non è mirato a instaurare uno stato psicologico preciso; al contrario, questo rimane soggettivo. Ossia, l'interpretazione dell'ambiente acustico è individuale. Ciò nonostante, afferma che i meccanismi psicologici stimolati dagli ambienti simili alla *Dream House*, sebbene provochino risultati individuali, generano degli stati d'animo astratti, che precedono l'interpretazione individuale. Egli, infatti afferma che «we differentiate the moods that we get through a particular set of frequency ratios from the more mundane moods of personal feelings» ([42] p. 62). I rapporti tra le frequenze, quando semplici e iterati per virtualmente infiniti lassi di tempo, dunque, sono in grado di trasmettere stati d'animo «specific to the mathematical abstract» ([42] p. 62), poi interpretati in diverse maniere, secondo le individualità del nostro fisico e secondo la nostra cultura.

Da quanto ora osservato si possono trarre un po' di conseguenze sulle modalità di fruizione ricercate da Young attraverso la *Dream House* (ma anche attraverso altri lavori coevi). Queste, infatti, differenziano in maniera netta la sua produzione dal panorama installativo coevo, dalle poetiche e dalle opere dei primi anni Sessanta. Young, infatti, immergendo il fruitore nel suono, intende sgretolare la distanza critica che lo separa dall'opera: l'esperienza dell'opera si basa ora sull'appartenenza di questi all'ambiente in cui è immerso e non a una sua interpretazione dello stesso. L'opera perde la sua dimensione e la sua funzione rappresentativa, ed è mirata verso la costituzione di una presenza. Il fruitore non la inserisce in un orizzonte semantico che ha studiato o appreso nel corso degli anni, ma la fruisce con l'intero corpo, senza attribuirle un senso a partire da un sistema ad essa esterno. L'opera agisce direttamente sulla percezione in maniera non mediata da forme di conoscenza razionali o simboliche, di cui il fruitore sia consci. L'opera, ambiente artificiale per definizione, ricerca una comunicazione "naturale" e non mediata dalla conoscenza e dalla cultura del fruitore. Anche per questo motivo Young si interessa di problematiche legate alla psicoacustica. Si passa dunque da un paradigma cognitivo dell'esperienza artistica a uno percettivo, basato sull'esperienza. Per tutte queste ragioni si può parlare di una fruizione non mediata e, dunque, far riferimento alla categoria dell'immersione del fruitore nell'opera. Una categoria non immanente ai *media* e ai dispositivi impiegati, ma alla poetica e alle strategie poetiche di Young, dunque a una costellazione di tecniche, al modo in cui il compositore progetta l'esperienza dell'opera, a prescindere dal materiale che ha a disposizione. La riduzione dei mezzi, ad esempio, più che essere scopo dell'opera, come la definizione stessa di "minimalismo" lascia a intendere<sup>6</sup>, è un mezzo per realizzare tale obiettivo<sup>7</sup>.

L'ascolto ricercato da Young, dunque, non è contemplativo, distaccato, e non è neppure analitico o basato sulla proiezione di conoscenze preformate sull'opera: piuttosto, è un ascolto che avviene con il corpo, che riguarda l'intero individuo. L'ascolto è interiore e, coinvolgendo il corpo, implica anche la dimensione cinestesica, pur se in presenza della sola vibrazione della pressione dell'aria. Anche solo per ciò potremmo parlare di un'opera multimodale.

Vi è, però, anche un altro motivo per affermare la natura multimodale della *Dream House*, pratico e più consistente, di cui il compositore era ben consciente (vedi [40] pp. 1-12). La diffusione statica, costante, di timbri semplici, di onde sinusoidali in un ambiente, provoca un fenomeno acustico particolare, ben noto agli stu-

<sup>6</sup> Secondo un'idea che non coinvolge solamente gli studiosi del fenomeno minimalista in musica, ma anche alcuni artisti che affriscono al movimento, tra cui Young stesso (vedi [42] pp. 23-25).

<sup>7</sup> Bill Viola, uno dei primi artisti a essersi confrontato con il modello di installazione minimalista afferma a tal riguardo: «Desert solitude is an early form of visionary technology. [...] Individuals have used it to hear voices of the past and future, to become "prophets", to receive images or, for Native Americans, to host "vision quests". It seems that when all the clutter and noise of everyday life is reduced to such brutal minimalism, the usual "control valves" are released and images well up from within» ([36] p. 164)

diosi di acustica architettonica: la configurazione della stanza determina l'amplificazione di determinate frequenze a discapito di altre, a causa di un processo di riflessione delle onde nell'ambiente e di una conseguente e statica amplificazione e cancellazione di fase. Tale processo non è costante per tutto lo spazio; al contrario, ogni frequenza presenta in alcune zone della stanza un'ampiezza maggiore e in altre risuona di meno o non si sente affatto. L'ambiente, saturo, composto di vibrazioni statiche, assume una definizione sonoro-spaziale unica, determinata dall'interrelazione tra le lunghezze d'onda delle sinusoidi diffuse e le dimensioni dello spazio stesso. Per ciò La Monte Young parla di "accordare" l'ambiente sonoro con la stanza in cui è installato e per lo stesso motivo si può richiamare, pur se solo in maniera parziale, la categoria della *site-specificity*<sup>8</sup>.

Ciò significa che muovendosi per l'ambiente il fruitore percepisce una variazione continua di ampiezza del segnale, diversa per ogni frequenza: un timbro unico, ma cangiante nello spazio (e conseguentemente anche nel tempo, in base all'interazione del fruitore con lo spazio e con la disposizione spaziale del suono). Dal punto di vista modale, dunque, c'è una relazione diretta tra la propriocezione, la cinestesia e l'udito, determinata dalla configurazione sonora dello spazio, dell'architettura in cui il fruitore è immerso. La relazione tra le modalità di percezione, inoltre, non è basata su principi esterni, come la sincronizzazione di diversi eventi che accadono nello stesso spazio e nello stesso momento, ma avviene nella costituzione stessa dello spazio sonoro. Per tal motivo non si può che parlare di una natura multimodale dell'opera e non si può non mettere in risalto come tale concezione distingua l'opera di Young in seno al panorama coevo.

Il fruitore con ogni probabilità coglie la sua relazione con lo spazio solo in maniera intuitiva, senza comprenderne i presupposti fisici (a meno che non conosca elementi basilari di acustica architettonica). Piuttosto, si farà immergere dall'ambiente, interagendo con esso. Il suo movimento nello spazio cambierà la configurazione delle riflessioni, modificando, pur se in maniera ridotta, l'architettura sonora. La sua sensazione di interagire con l'ambiente sarà inoltre modificata dal fatto che anche i semplici movimenti della testa cambiano la configurazione delle frequenze sentite, a causa della distanza variabile tra le orecchie e della conseguente cancellazione di fase. Non si può parlare di un'installazione interattiva, dal momento che il fruitore non agisce in maniera diretta sul sistema-opera: la categoria dell'interazione, a differenza di quella immersiva, non è schiacciata sulla sola dimensione modale, ma anche su quella mediale, ossia sulla struttura interna dei dispositivi, della loro disposizione e in generale delle opere e del modo in cui investono il fruitore. Ciò nonostante, il fruitore si sentirà parte dell'opera e percepirà di interagire con essa.

<sup>8</sup> Ad esempio in [38] p. 16, si legge «We will be able to have exactly the right sound in every dreamroom, playroom and workroom, further reinforcing the integral proportions resonating through structure (re: earlier Architectural Music), Dream House (shrines, etc.) [...].»

Il collegamento tra movimento del fruitore e suono comporta una concezione introspettiva, multimodale e immersiva dell'opera. Tale fruizione costituisce, nell'ambito della storia delle installazioni sonore una novità. Solo poche opere avevano provato a raggiungere una tipologia di fruizione simile: l'unica che mi viene in mente è la *Caverna dell'Antimateria* di Pinot Gallizio attraverso il Tereminofono di Gege Cocito; però, qui il proposito è realizzato solo in maniera parziale. Tale caratteristica è a mio avviso più importante della semplice quantità limitata di mezzi e materiali impiegati dall'artista; essa, inoltre, è condivisa dal repertorio minimalista di ambito plastico-pittorico e determina una svolta radicale nella storia delle installazioni sonore, schiudendo la strada a un'ampia serie di opere basate su principi affini a quelli osservati nell'analisi della *Dream House*, alcuni dei quali saranno presi ora in esame.

### **2.3 Installazioni minimaliste basate sul modello della *Dream House*.**

#### *2.3.1 Installazioni minimaliste basate su una concezione architettonica dello spazio acustico.*

A partire dalla prima metà degli anni Settanta una quantità crescente di installazioni sonore si rifà in maniera più o meno esplicita al modello elaborato da Young nella *Dream House*. Le prime installazioni di Bill Viola, *Hallway Nodes* (1973) e *The Mysterious Virtue*, (1974) ([33, 34] e [36] pp. 35-37), sono un esempio lampante di tale fenomeno: composte di soli suoni sinusoidali e basate sulla produzione di aree, zone di differente ampiezza; inoltre, lavorano sulla dimensione tattile, fisica, del suono, prodotta dall'amplificazione estrema di basse frequenze, che dà origine a una multimodalità complessa. Viola crea ambienti 'eterodini', basati sulle configurazioni spaziali delle riflessioni delle onde sinusoidali e sulle loro interferenze, che generano aree variabili di pressione dell'aria e velocità delle particelle. In seguito Viola abbandonerà tale metodo, ma continuerà a basare le sue opere sull'immersione e sulla multimodalità, nonché su una concezione meditativa della fruizione e su un'attenta calibrazione e gestione dei materiali.

Anche Michael Brewster ha lavorato a lungo sui principi riscontrati nell'analisi del modello di Young. La sua produzione installativa, suddivisa nei "Sonic Drawings" e nelle "Acoustic Sculptures" è incentrata sulla concezione architettonica dello spazio sonoro, come spiega lui stesso in [3]. Ha rielaborato tale principio in numerosi modi, rendendolo nel corso degli anni sempre più complesso, non lavorando solo sulle proprietà spaziali delle onde di pressione sonora, ma anche esplorando le potenzialità plastiche della dislocazione spaziale delle sorgenti sonore, così come quelle della morfologia dello spazio sonoro<sup>9</sup>.

Particolarmenete celebre, poi, è l'unica installazione sonora di Philip Glass, o almeno l'unica opera del suo catalogo *on-line* espressamente classificata tale<sup>10</sup>:

<sup>9</sup> Per una visione d'insieme della produzione installativa di Brewster, cfr. [35].

<sup>10</sup> In realtà Glass ha collaborato a numerose installazioni sonore e installazioni-performance non citate nel suo catalogo (<

*Pink Noise*, realizzata dal compositore nel 1987 insieme a Richard Serra presso la *Ohio State University of Fine Arts* a Columbus. Un'imponente scultura di Serra, composta da grandi piastre di metallo inclinate e piegate, riflette in molteplici direzioni una serie di alte frequenze diffuse da altoparlanti dislocati in diverse posizioni dell'ambiente. La scelta di adoperare frequenze acute deriva dalla loro capacità di propagazione direzionale: la loro diffusione, infatti, dipende dall'angolo di incidenza tra la sorgente e la scultura<sup>11</sup>. Queste, dunque, una volta riflesse, generano attorno alla statua diverse combinazioni armoniche in base al luogo di fruizione.

Sebbene dunque il principio di generazione del suono di *Pink Noise* e della *Dream House* sia difforme, la struttura modale delle due opere è simile. Il principio fisico su cui si basa l'opera non è qui prodotto dalle sole interferenze o dalla cancellazione di fase, ma anche dalla marcata direzionalità delle frequenze con lunghezza d'onda ridotta e dunque dalla relazione tra la localizzazione degli altoparlanti e la forma della scultura. Però, anche *Pink Noise*, così come *Hallway Nodes* e la *Dream House*, è basata sull'idea di usare il suono in senso plastico, per la creazione di uno spazio acusticamente composto. La relazione tra fruitore e spazio, tra fruitore ed opera e tra spazio e suono è dunque simile a quella osservata nell'opera precedentemente analizzata. Costituisce, ovviamente, un'interessante novità la stretta relazione tra il suono e la scultura e tra questa e lo spazio che la circonda.

I casi qui mostrati, solo esempi tra i molti menzionabili<sup>12</sup>, sviluppano il modello di Young verso diverse direzioni, a seconda dell'interesse degli artisti. Essi condividono un interesse verso la dimensione architet-

---

[> - \*Pink Noise\* è qui classificata quale «collaborative acoustic installation». Si deve a tal riguardo ricordare la lunga e stretta collaborazione con Richard Serra, con cui iniziò a lavorare già dalla seconda metà degli anni Sessanta: è ben conosciuta, ad esempio, l'installazione \*Long Beach Island. World Location\* del 1971. Glass, per l'occasione, ha posizionato trentadue altoparlanti intorno a un'installazione all'aperto, un'opera di \*Land Art\* di Serra; dagli altoparlanti venivano diffusi suoni prodotti da alcuni magnetofoni, che leggevano in maniera continuativa dei \*loop\* sfasati, in cui era recitata la parola «is». Si possono inoltre ricordare le performance pianistiche che Glass ha a più riprese effettuato all'interno di diverse installazioni sonore, come la \*performance\* realizzata all'interno dell'installazione sonora \*Sum of Days\* di Carlotto Carvalhos al MoMA di New York \(aperta tra il 24 agosto e il 14 novembre 2011; Philip Glass e il suo ensemble hanno suonato all'apertura e hanno realizzato li alcuni concerti\): uno spazio chiuso che riprende e diffonde in maniera continua la \*performance\* di Glass e gli altri eventi sonori che accadono con un ritardo di un giorno, in \*feedback\* continuo, in maniera tale che l'opera risulta ogni giorno più stratificata.](http://www.philipglass.com/music/compositions/index_compositions.php)

<sup>11</sup> Ovviamente una superficie concava rifletterà le onde verso il suo punto focale, mentre una superficie convessa riflette il suono in diverse direzioni.

<sup>12</sup> Ad esempio, un'altra notevole opera che si potrebbe citare, anche se un po' più lontana dalla *Dream House* rispetto a quelle sopra trattate, è *Sounding Space* di Paul Earls, opera creata per la mostra *Sound Sculpture* alla Vancouver Art Gallery nel 1974: un'installazione sonora che, in maniera simile a altre opere coeve di Earls realizzate al M.I.T. sotto la guida di Gyorgy Kepes come *Flame Orchard* e *Sound Floor*, è basata sulla rielaborazione in tempo reale del *feedback* acustico ambientale. Opera in qualche modo a sua volta modello, pur se parziale, del lavoro successivo sul *feedback* acustico nell'ambito installativo, fino agli ecosistemi udibili di Di Scipio.

tonica, spaziale, delle onde sonore e verso tipologie di fruizione multimodali e immersive. Tali caratteristiche contribuiscono a generare un insieme conchiuso di opere, che condivide caratteristiche simili, nettamente differenziato dalla produzione installativa coeva.

### 2.3.2 Altre installazioni minimaliste.

Robert Irwin e James Turrell trascorrono nel 1968 lunghi lassi di tempo nella camera anecoica della *Garrett Corporation*. Conducono lì numerosi esperimenti di psicologia cognitiva insieme a Ed Wortz, direttore del dipartimento di *Life Sciences* della *Garrett*, pianificando alcune opere/installazioni, che poi non realizzarono, se non come esperimenti (cfr. [21]). Tale lavoro ha però esercitato una notevole influenza sull'attività successiva dei due artisti [28], e con loro su un'ampia parte dell'arte installativa e della *light art* degli anni Settanta. Di primo acchito parrebbe un lavoro radicalmente diverso rispetto a quello condotto da Young con la *Dream House*, ma, a ben vedere, mostra notevoli somiglianze con esso, sia nei principi fondamentali, sia, soprattutto, nel modo di concepire la relazione tra fruitore e opera.

Irwin e Turrell non intendevano costruire un oggetto che rappresentasse o significasse qualcosa, bensì lavorare con i limiti, le soglie della percezione. Si interessano a uno spazio buio e silenzioso per studiare la percezione degli stimoli sensoriali. L'ambiente anecoico rappresenta per loro il grado zero dell'esperienza sensoriale su cui costruire l'opera, manipolando la fruizione attraverso un dosaggio controllato di stimoli percettivi minimi. Studiano la reazione dei fruitori a prolungate esposizioni alla camera anecoica, per vedere gli effetti della privazione sensoriale. Analizzano, anche su base statistica, gli esiti che tale privazione ha sulla percezione, come questa si modifichi in base al contesto. Si interessano dunque alle forme dell'ascolto conseguenti a lunghe ore di immersione nell'oscuro silenzio, al modo, differente e unico in cui vengono percepiti i suoni e i colori in questa situazione. Lo spazio assume una qualità sonora e luminosa particolare, che può essere manipolata in maniera artificiale. Immaginano poi percorsi per i fruitori in cui gli stimoli, sempre molto limitati, sono gestiti dagli artisti. Non lavorano su un materiale, bensì su «states of consciousness and awareness» ([21] p. 131), costruiscono, pianificano l'esperienza dell'opera e non la sua struttura. Come affermano, «the media are just perception» ([21] p. 132).

L'esperienza dell'opera non è per loro legata a specifici sensi, bensì all'intero corpo dei fruitori, dal momento che «Sensory experience is heightened when sense modalities act in phase» ([21] p.132). Per questo pensano di introdurre, dopo una lunga esposizione nella camera anecoica oscurata, stimoli insieme visivi e sonori controllati. Intendono inoltre costruire percezioni della temporalità distorte e iniziano a lavorare nell'ambito dell'«alpha conditioning», sui cicli di onde cerebrali che hanno luogo durante gli stati di meditazione.

Si tratta dunque di una serie di progetti diversi rispetto a quelli di Young; ciò nonostante anche loro adoperano mezzi minimi e molto controllati per manipolare in maniera diretta la percezione del fruitore. La

loro opera non rappresenta alcunché, non «sta per qualcosa», bensì agisce sulla percezione multimodale del fruitore. Questi, inoltre, è investito da stimoli sensoriali e prova un'esperienza che non riesce a controllare, a schematizzare mediante categorie da lui conosciute, a razionalizzare. Egli non interpreta un'opera, la vive.

Anche Bruce Nauman ha realizzato diverse installazioni sonore tra la fine degli anni Sessanta e gli anni Settanta ([13] pp. 86-90). Non è qui possibile operare un'analisi di tali opere, peraltro ben note; piuttosto, si vuole rimarcare una continuità nelle forme di fruizione tra loro e le «installazioni minimaliste» precedentemente analizzate. È interessante rilevare l'attenzione di Nauman verso i dettagli dell'acustica architettonica: progetta spazi invisibili, la cui risonanza, udibile solo in alcuni tratti degli stretti corridoi in cui costringe i fruitori, colpisce in maniera inaspettata lo spettatore. Questi non comprende perché l'ambiente acustico dello spazio angusto in cui è immerso cambi a tratti, ma è spinto verso un'esperienza immersiva e multimodale: vive, esperisce l'opera con tutto il corpo. È necessario sottolineare come Nauman stesso insista verso un controllo totale dell'esperienza dei fruitori<sup>13</sup>:

*The first corridor pieces were about having someone else do the performance. But the problem for me was to find a way to restrict the situation so that the performance turned out to be the one I had in mind. In a way, it was about control. I didn't want somebody else's idea of what could be done.* ([30] p. 147)

Il fruitore non prende solo parte all'installazione di Nauman, non è un semplice *participant*, ma è messo in crisi, in discussione da essa: la sua tranquillità è minata da un'eccitazione sensoriale di cui non riesce a comprendere la natura. Non è stimolato a prendere coscienza o a comprendere una situazione sociale, ma è investito sensorialmente. Anche le opere, le installazioni sonore di Nauman (a differenza, in parte, delle opere basate su testi [19]), tendono verso una manipolazione dell'esperienza e della coscienza del fruitore: un'esperienza costruita, multimodale e immersiva. Per tale motivo sono classificabili sotto la medesima categoria, quella delle installazioni minimaliste, adoperata per le altre opere affrontate nel presente saggio.

In questo paragrafo abbiamo dunque individuato alcune installazioni sonore che, pur differenti nelle forme e nell'architettura mediale rispetto a quelle precedentemente analizzate, ne condividono la struttura modale, il modo in cui creano l'esperienza del fruitore, sensoriale e cognitiva: la relazione opera-fruitore.

## 3. CONCLUSIONI: MINIMALISMO, IMMERSIONE E MULTIMODALITÀ

La definizione di minimalismo osservata attraverso la lente della storia delle installazioni sonore non si riferisce soltanto a una caratteristica esterna delle opere, il fatto di essere composte da una quantità minima di materiale, ma anche a una specifica forma di fruizione, che

<sup>13</sup> Si può vedere quanto lui stesso afferma in diverse occasioni, ad es. [29, 30 e 4]. Per uno studio complessivo e un'interpretazione sociologica della fruizione nell'opera di Nauman, vedi [18].

prevede un continuo lavoro di negoziazione tra apparato sensoriale e corpo. Le opere sono volte verso la messa in atto di un'esperienza sensoriale piuttosto che verso la messa in scena di forme di rappresentazione simbolica. In tal modo la costruzione del senso non è demandata a un orizzonte esterno alle opere, che ne coadiuvi l'interpretazione, ma è inscritto direttamente nell'esperienza stessa dell'opera. Una tipologia di fruizione che prende lentamente piede nel corso degli anni Sessanta, per affermarsi pienamente intorno al 1968, pur accanto ad altre, relative ad altri repertori.

Si deve rimarcare infine che l'immersione e soprattutto la multimodalità non caratterizzino solamente le installazioni sonore minimaliste, ma in generale l'arte minimalista: le abbiamo infatti ritrovate nella lettura di O'Doherty della pittura minimalista citata all'inizio del saggio. Inoltre una simile interpretazione potrebbe costituire una chiave attraverso cui leggere, almeno in parte, la musica minimalista *tout court*, anche le opere performative. L'incessante ripetizione di moduli ritmici che caratterizza in diverse forme i lenti processi minimalisti comporta anche un cambiamento nelle modalità e nelle forme dell'ascolto; si potrebbe ipotizzare che in tale attività sia qui coinvolta anche la dimensione cinesistica, per generare un diverso rapporto tra il modo uditorio e il corpo del fruitore, o, seguendo la terminologia di O'Doherty, tra l'orecchio e lo spettatore. Tale caratteristica, peraltro, è, pur sotto diverse forme, sfruttata anche nella tradizione post-minimalista filtrata dalla produzione cinematografica, per la sua influenza sulla psicologia percettiva: strategie compositive in cui i singoli moduli, connotati nelle proprietà di manipolazione motoria della percezione, vengono incessantemente ripetuti per esercitare un influsso non mediato sui fruitori; moduli "cineticci" collegati a determinate situazioni narrative (moduli d'azione, moduli horror, etc.).

La riduzione del materiale, dunque, secondo la nostra interpretazione, non è che uno stratagemma per instaurare una comunicazione immersiva e multimodale tra opera e fruitore. Una tipologia di costruzione dell'esperienza non del tutto nuova nella storia delle arti e in particolare della musica<sup>14</sup>, ma contrapposta a quella dominante negli anni Sessanta, di matrice kaprowiana e in parte cageana: quella dell'arte partecipativa. Una partecipazione cosciente all'opera, che demanda al fruitore la responsabilità dei meccanismi sociali che in essa si instaurano e finalizzata a forme di conoscenza e consapevolezza. Un'arte che sfrutta la collaborazione, un rapporto ludico con l'opera, forme di shock e ironia, che ebbe nel *happening* la forma artistica principale, e che fu messa da parte nell'estetica minimalista. Contemporaneamente l'ottica intermediale teorizzata da Dick Higgins [11], viene soppiantata da un interesse degli artisti verso la dimensione multimodale, dalle forme di interazione dei e tra i media e dalla ricerca in seno ai loro interstizi a quelle sui modi e sulle forme della percezione.

<sup>14</sup> Si pensi alla concezione wagneriana del *Gesamtkunstwerk* come unità delle arti nel corpo del fruitore e, nel contempo, dominio totale di questo attraverso l'arte da parte del compositore-genio (cfr., ad esempio, [26], in particolar modo pp. 3-10 e [8] pp. 21-24).

## 4. BIBLIOGRAFIA

- [1] Giacomo Albert: *L'immersione*, in corso di pubbl.: <[www.worldsofaudovision.org](http://www.worldsofaudovision.org)>.
- [2] Claire Bishop: *Participation*, Cambridge Massachussetts, MIT Press, 2006.
- [3] Michael Brewster: "Where, There and Here. An essay about sound as sculpture", Brandon LaBelle e Steve Roden [ed.], *Site of Sound: of Architecture and the Ear*, Errant Bodies Press, 1999.
- [4] Jan Butterfield: "Bruce Nauman: The Center of Yourself", «Arts Magazine», Vol. 49, no. 46, 1975.
- [5] Frances Dyson: *Sounding New Media. Immersion and Embodiment in the Arts and Cultures*, University of California Press, Berkeley-Los Angeles, 2009.
- [6] Michael Fried: *Absorption and Theatricality: Painting and Beholder in the Age of Diderot*, University of California Press, Berkeley-Los Angeles, 1980.
- [7] Michael Fried: "Art and Objecthood," *Artforum*, Vol 5, 1967, pp. 12-23.
- [8] Rudolf Frieling [ed.]: *The Art of Participation. 1950 to now*, San Francisco – New York – London, SFMOMA – Thames and Hudson, 2008.
- [9] Kyle Gann: "The Outer Edge of Consonance. Snapshots from the Evolution of La Monte Young's Tuning Installations", William Duckworth e Richard Fleming [cur.], *Sound and Light; La Monte Young Mariane Zazeela*, Fleming, Lewisburg, 1996, pp. 152-190.
- [10] Oliver Grau: *Virtual Art. From Illusion to Immersion*, MIT Press, Cambridge Massachusetts, 2003.
- [11] Dick Higgins, "Intermedia", Id., *Horizons. The Poetics and Theory of the Intermedia*, Southern Illinois University Press, Carbondale and Edwardsville, 1984; ed. or., «Something Else Newsletter», Vol. 1, No. 1, New York, Something Else Press, 1969.
- [12] Charles Johnson: *Tuning in Opposition: The Theatre of Eternal Music and Alternate Tuning Systems as Avant-Garde Practice in the 60's*, <[http://www.cirrusoxide.com/journal2/wp-content/uploads/Tuning\\_in\\_Opposition.pdf](http://www.cirrusoxide.com/journal2/wp-content/uploads/Tuning_in_Opposition.pdf)>.
- [13] Douglas Kahn: "Pléitudes vides et espaces expérimentaux", *Sons et Lumières. Une histoire du son dans l'art du XXe siècle*, Sophie Duplaix – Marcella Lista [eds.], Editions du Centre Pompidou, Paris, 2004, pp. 79-89.

- [14] Allan Kaprow: *Assemblages, Environments & Happenings*, New York, Harry N. Abrams – Inc. Publisher.
- [15] Allan Kaprow: *Notes for the Creation of a Total Art*, Id., *Essays on the Blurring of Art and Life*, Jeff Kelley [ed.], University of California Press, Berkeley and Los Angeles, 2003, pp. 10-14.
- [16] Allan Kaprow: “The Legacy of Jackson Pollock”, Id., *Essays on the Blurring of Art and Life*, Jeff Kelley [ed.], University of California Press, Berkeley and Los Angeles, 2003, pp. 1-9; ed. or. «Artnews», Vol. 57, No. 6, 1958.
- [17] Richard Kostelanetz: *Conversation with La Monte Young*, La Monte Young e Mariane Zazeela, *Selected Writings*, ubuclassics, 2004, pp. 17-63.
- [18] Janet Kraynak: “Participation: Bruce Nauman’s Environments”, «Grey Room», No. 10, 2003, pp. 22-45.
- [19] Janet Kraynak, *Please Pay Attention Please. Le parole di Bruce Nauman*, Postmedia books, Milano, 2004.
- [20] Pierre Lévy: *Cybercultura: gli usi sociali delle nuove tecnologie*, Feltrinelli, Milano, 1997, ed. or. Id., *Cyberculture*, Odile Jacobs, Paris, 1997.
- [21] Jane Livingstone: “Robert Irwin. James Turrell”, *A report on the Art and Technology Program of the Los Angeles County Museum of Art 1967-1971*, Los Angeles County Museum of Art, Los Angeles, 1971, pp. 127-143.
- [22] Mary Anne Moser – Douglas McLeod [eds]: *Immersed in Technology: Art and Virtual Environments*, MIT Press, Cambridge, 1996.
- [23] Michael Nyman: *Experimental Music. Cage and Beyond*, Cambridge University Press, Cambridge, 1999 [1<sup>a</sup> ed. 1974].
- [24] Brian O’Doherty: *Inside the White Cube. The Ideology of the Gallery Space. Expanded Edition*, University of California Press, Berkeley–Los Angeles–London, 1999.
- [25] Pauline Oliveros: *Software for People. Collected Writings 1963-80*, Baltimore, Smith Publications, 1984.
- [26] Randall Packer e Ken Jordan, *Multimedia. From Wagner to Virtual Reality*, Norton & Company, New York & London, 2001.
- [27] Julie H. Reiss: *From Margin to Center. The Spaces of Installation Art*, Cambridge Massachusetts, MIT Press, 1999.
- [28] Dawna Schuld: “White Cube and Black Box: Irwin, Turrell and the return of the subject in 1960s and 1970s art and psychology», *Beyond Mimesis and Nominalism: Representation in Art and Science*, <http://philsci-archive.pitt.edu/3134/>, London, pp. 1-23.
- [29] Willoughby Sharp: “Nauman Interview”, «Arts Magazine», Vol. 44, no. 5, 1970.
- [30] Joan Simon: “Breaking the Silence: an Interview with Bruce Nauman”, «Art in America», Vol. 76, No. 9, 1988, pp. 140-149.
- [31] Volker Straebel: “Geschichte und Typologie der Klanginstallation”, Ulrich Tadday [cur.], *Klangkunst*, «Musik-Konzepte Sonderband», Vol. XI/2008, München, edition text+kritik, pp. 24-47.
- [32] Edward Strickland: *Minimalism, Origins*, University of Indiana Press, Bloomingdale, 1993.
- [33] Peter Szendy: “Bill Viola: Hallway Nodes”, Sophie Duplaix e Marcella Lista [ed.], *Sons & Lumières. Une histoire du son dans l’art du XXe siècle*, Centre Pompidou, Paris, pp. 238-239.
- [34] Peter Szendy: ““Installations sonores?”, «Resonance», Vol. 12, Ircam, Paris, 1997.
- [35] Irene Tsatsos [ed.]: *Michael Brewster: See Hear Now. A sonic Drawing and Five Acoustic Sculptures*, Los Angeles Contemporary Exhibitions, Los Angeles, 2002.
- [36] Bill Viola e Robert Violette [ed.], *Reasons for Knocking at an Empty House. Writings 1973-1994*, Thames and Hudson, London.
- [37] La Monte Young: “Notes on the continuous periodic composite sound waveform environment realizations of “Map of the 49’s dream the two systems of eleven sets of galactic intervals ornamental lightyears tracery”, Young e Zazeela, *Selected Writings*, ubuclassics, 2004, pp. 5-9.
- [38] La Monte Young: *Notes on The Theatre of Eternal Music and The Tortoise, His Dreams and Journey*, 2010, <<http://melafoundation.org/theatre.pdf>>.
- [39] La Monte Young e Mariane Zazeela: “Continous Sound and Light Environments” William Duckworth e Richard Fleming [cur.], *Sound and Light; La Monte Young Mariane Zazeela*, Fleming, Lewisburg, 1996, pp. 218-221.
- [40] La Monte Young e Mariane Zazeela: “Dream House”, La Monte Young e Mariane Zazeela, *Selected Writings*, ubuclassics, 2004, pp. 10-16.
- [41] La Monte Young, Mariane Zazeela, Cole Gagne: “Conversation”, Cole Gagne [cur.], *Soundpiece 2. Interviews with American Composers*, Metuchen, New Jersey-London, pp. 471-505.
- [42] La Monte Young, Mariane Zazeela, Frank J. Oteri: *La Monte Young and Mariane Zazeela at the Dream House. In conversation with Frank J. Oteri*, Newmusicbox, 2003, <<http://www.newmusicbox.org/article.nmbx?id=2216>>.

# ANALISI PARADIGMATICA DI COMPOSIZIONI ACUSMATICHE E MUSIC INFORMATION RETRIEVAL: UN PRIMO APPROCCIO

Giorgio Klauer

Conservatorio di Como

[giorgio.klauer@conservatoriocomo.it](mailto:giorgio.klauer@conservatoriocomo.it)

## SOMMARIO

È stata formalizzata una procedura di analisi e associazione paradigmatica mediante elaboratore delle unità musicali segmentate da composizioni acusmatiche. Scopo della procedura è fornire un riscontro ai risultati che discendono dall'applicazione auditiva del modello di analisi paradigmatica.

La procedura utilizza il complesso di informazioni espresse da un insieme di descrittori di basso livello, raggruppate in accordo a una preesistente segmentazione dell'opera. La procedura prosegue in una variante dove la dissimilarità reciproca delle unità musicali è computata come il costo di allineamento (DTW e DDTW) delle sequenze di informazioni contenute nelle celle, e in una seconda variante dove l'allineamento sequenziale (ED) è attuato su simboli discreti che costituiscono l'indicizzazione delle celle stesse (clustering).

Sulla base delle matrici di similarità così ottenute, lo scaling multidimensionale ricolloca le unità in uno spazio le cui dimensioni sono interpretabili, seguendo il modello dell'analisi paradigmatica, come complessi di tratti componenziali. La collocazione delle unità nello spazio MDS consente di effettuare le associazioni paradigmatiche (clustering) e di stilare infine la tavola paradigmatica.

La procedura è stata attuata sulla base di composizioni di B. Parmegiani, A. Parmerdud, J.-C. Risset e H. Vagione, conducendo a risultati incoraggianti.

## 1. INTRODUZIONE E OBIETTIVI

In un notevole volume di Stéphane Roy dedicato alla metodologia dell'analisi della musica elettroacustica, si dimostra come le consuetudini dell'ascolto acusmatico e l'intensa pratica della rappresentazione iconica, possano costituire insieme a specifiche teorie di appoggio la base per segmentare in maniera neutra e pertinente il testo sonoro di composizioni appartenenti al repertorio in oggetto, e complementarne la descrizione con l'analisi mu-

tando metodi di provenienza semiologica [17]. Lo scrivente ha potuto verificare la validità della maggior parte degli accomodamenti metodologici proposti nell'applicazione al genere che è loro proprio, constatandone l'utilità nell'ambito della formazione artistico-musicale accademica, specialmente per il raggiungimento di quella comprensione formale necessaria a interpretare le opere in concerto agendo manualmente sulla console di

missaggio di un impianto di diffusione multicanale<sup>1</sup>.

L'esperienza ha evidenziato tuttavia l'intrinseco limite, nella validazione dei risultati, del dover fare affidamento a superiori e specifiche competenze estetiche, insostituibili anche adottando un'ottica comparativa, cioè disponendo di differenti analisi del medesimo brano svolte secondo il medesimo metodo, o svolgendo collettivamente le analisi stesse. Da questo dato, che riflette la caratteristica forma di trasmissione del sapere artistico, è emersa la necessità di un riscontro non soggettivo nel merito dei risultati dell'analisi, quale può essere fornito da un metodo computazionale.

L'urgenza riguarda soprattutto l'aspetto classificatorio dell'analisi ed è massimamente percepibile nell'applicazione del metodo classico dell'analisi paradigmatica, laddove la ponderazione del complesso dei tratti componenziali pertinenti conduce all'attribuzione delle unità musicali ai paradigmi di riferimento [10, 11, 16]. Infatti, se tradizionalmente l'applicazione del metodo dell'analisi paradigmatica si svolge a partire da una notazione simbolica che costituisce di per sé una formalizzazione del discorso musicale (seppure le diverse istanze descrittive/prescrittive finalizzate alla trasmissione e alla riproduzione dell'opera tendano a comprometterne la completezza e la coerenza), nel campo della musica elettroacustica i passaggi metodologici risultano quanto mai legati alle capacità percettive e cognitive individuali, rendendole talmente decisive rispetto ai risultati della procedura, da lasciare un margine di arbitrio solo parzialmente colmabile dalle teorie di appoggio che possono intervenire, ad esempio, nella valutazione di pertinenza e nella scelta dei tratti componenziali, o dalle necessarie competenze storico-tecniche che rischiano pur sempre di declinare l'analisi nel senso di una pratica esegetica.

Nonostante l'intrinseca corrispondenza del metodo paradigmatico con le opportunità offerte dal calcolatore e dal repertorio elettroacustico, si è notata una certa sparsità di soluzioni, divaricate tra quelle dove il computer non è utilizzato altrimenti che per semplificare il riscontro auditivo di associazioni paradigmatiche effettuate senza l'ausilio di strumentazione tecnologica [5], e quelle molto più numerose dove tende a sostituire completamente l'essere umano nello svolgimento di procedimenti di segmentazione e classificazione, peraltro non attinenti all'analisi paradigmatica. Singolari esperienze di analisi musicale elettroacustica dove il computer fornisce un riscontro a operazioni svolte in forza di precise competenze musicali o musicologiche sono servite da riferimen-

Copyright: ©2012 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.

<sup>1</sup> I metodi di analisi funzionale, paradigmatica e generativa sono stati applicati sistematicamente a partire dal 2009 dallo scrivente nell'ambito dei corsi accademici di Analisi della musica elettroacustica (COME/02).

to per la rimodellazione della procedura dell'analisi paradigmatica qui presentata [12].

Nel formulare la procedura si era valutato inizialmente di utilizzare descrittori morfologici di alto livello [15] sulla base di una preesistente segmentazione molto serrata e di effettuare l'associazione paradigmatica delle unità mediante clustering gerarchico. La descrizione morfologica mediante elaboratore di un flusso di informazione complesso qual è una composizione acusmatica costituiva tuttavia una problematica che superava gli scopi e si è preferito prendere in considerazione solamente descrittori di basso livello e articolare i rimanenti passaggi enfatizzando le intrinseche basi statistiche dell'analisi paradigmatica.

La procedura qui presentata consiste quindi in una successione di algoritmi che partendo da una segmentazione preesistente effettua l'associazione paradigmatica in forza di tratti componenziali virtuali. La procedura si svolge nei passaggi: acquisizione delle informazioni espresse da descrittori; misura della similarità fra le unità mediante algoritmi di sequence alignment; individuazione di fattori di caratterizzazione di alto livello mediante scaling multidimensionale (tratti componenziali); distribuzione delle unità mediante clustering in un numero arbitrario di insiemi (associazione paradigmatica).

## 2. DESCRIZIONE DELLA PROCEDURA

Oggetto dell'analisi è il testo sonoro digitale/digitalizzato di composizioni acusmatiche: documenti audio con risoluzione 16 bit e frequenza di campionamento 44.1 KHz depositati su supporti commercializzati, il cui formato originale viene ridotto a un singolo canale<sup>2</sup>.

### 2.1 Descrittori di basso livello

Si acquisisce il complesso di informazioni relative al contenuto del flusso sonoro espresso da descrittori di basso livello. Le informazioni, ricavate a partire dallo spettro, rappresentano il flusso come una sequenza di celle corrispondenti alla finestratura FFT<sup>3</sup>.

- Mel Frequency Cepstral Coefficients (1:13)
- Chromagram (14:25)
- Loudness (26)
- Pitch (27)
- Periodicity (28)
- Roughness (29)
- Brightness (30)
- Roll-Off (31,32)

<sup>2</sup> In un prossimo sviluppo le informazioni relative alla spazialità verranno integrate nella procedura come descrittore di basso livello (differenza del contenuto tra i canali). Seguendo le convenzioni dell'editoria discografica, la maggioranza delle produzioni musicali acusmatiche consiste in un master stereofonico, usato anche come base materiale nella diffusione effettuata tipicamente agendo sulla console di missaggio di un impianto multicanale. L'effettiva resa delle informazioni spaziali è delegata all'interprete, che conferirà al brano uno specifico orientamento formale anche in considerazione della catena di trasduzione elettroacustica utilizzata e dell'acustica dell'ambiente - fattori che condizionano sostanzialmente le scelte interpretative per l'impatto sulla traslucidità/opacità dei piani sonori, sul mascheramento delle sonorità stesse e quindi sull'attribuzione soggettiva delle qualità timbriche.

<sup>3</sup> FFT con finestre e hopsize differenti: i dati vengono interpolati per allineare le informazioni alla finestratura principale di 1024 punti.

- Flatness (33)
- Crest (34:36)
- Spread (37)
- Slope (38)
- Onsets (39)

Le informazioni vengono normalizzate sulla base dell'andamento complessivo del singolo descrittore e raggruppate in sequenze secondo la preesistente segmentazione delle unità musicali<sup>4</sup>.

### 2.2 Calcolo della similarità fra le unità musicali

Il calcolo della (dis)similarità reciproca tra le unità musicali viene effettuato valutando mediante algoritmo di dynamic time warping (DTW) il costo del migliore allineamento tra le sequenze di informazioni<sup>5</sup>. È stato eseguito in alternativa un algoritmo di derivative dynamic time warping<sup>6</sup> (DDTW).

### 2.3 Variante: calcolo della similarità sulla base dell'indicizzazione delle celle

Successivamente all'acquisizione del complesso di informazioni espresso dai descrittori, ogni cella viene indicizzata mediante algoritmo k-means++<sup>7</sup>. La similarità tra le unità musicali viene quindi calcolata valutando il costo di allineamento di sequenze di simboli discreti. A questo scopo è stata considerata la edit distance<sup>8</sup> (ED).

### 2.4 Fattori di caratterizzazione di alto livello (tratti componenziali)

Sulla base delle matrici di similarità così ottenute si esegue uno scaling multidimensionale classico (MDS), in modo da disporre le unità musicali in uno spazio avente per dimensioni i presunti fattori di caratterizzazione<sup>9</sup>.

<sup>4</sup> L'acquisizione dei dati è avvenuta in SuperCollider per mezzo del macro-operatore SCIMR sviluppato da Nick Collins [3] che utilizza a sua volta gli operatori MFCC, Chromagram, Loudness, Tartini, SensoryDissonance, SpecCentroid, SpecPcile, SpecFlatness, FFTCrest, FFTspread, FFTslope, Onsets. Per il funzionamento degli algoritmi si rimanda alla documentazione del software. Nella maggior parte sono stati mantenuti i settaggi di default, tranne: Roll-Off reso tramite SpecPcile in due versioni (95% e 80%), Crest reso da FFTCrest in tre versioni (banda passante completa, 0-2000 Hz, 2000-10000 Hz). L'informazione Onsets (indicazione istantanea di attacco) è stata resa più significativa mediante filtraggio passa-basso. Simili dati consentono tipicamente di desumere informazioni di più alto livello correlate a specifiche teorie della musica: la pratica dell'analisi computazionale dell'informazione musicale ha portato a modelli ampiamente cristallizzati e offerti in numerosissime distribuzioni sotto licenza GNU [7].

<sup>5</sup> L'algoritmo è stato attuato nella sua forma elementare, applicando soltanto un minimo coefficiente, nel calcolo delle corrispondenze, a favore del percorso di allineamento più breve [9].

<sup>6</sup> L'algoritmo individua il miglior allineamento valutando, al posto della distanza euclidea, la somma dei quadrati delle differenze tra le derivate delle informazioni espresse dai singoli descrittori [6].

<sup>7</sup> Esemplificato in [1], distanza euclidea. Il numero dei centroidi è stato determinato sperimentalmente, in misura doppia al numero di paradigmi cui si attribuiranno le unità musicali.

<sup>8</sup> Nella sua forma più semplice, l'algoritmo computa il minimo numero di operazioni di cancellazione, sostituzione e inserimento necessario a trasformare una sequenza di simboli in un'altra [8]. L'applicazione nell'ambito dell'analisi musicale è finalizzata tipicamente al riconoscimento di pattern melodici [14].

<sup>9</sup> Cfr. lo scaling multidimensionale (viceversa, non-metrico e INDSCAL) nella storica ricerca di J. Grey [4] volta a individuare i

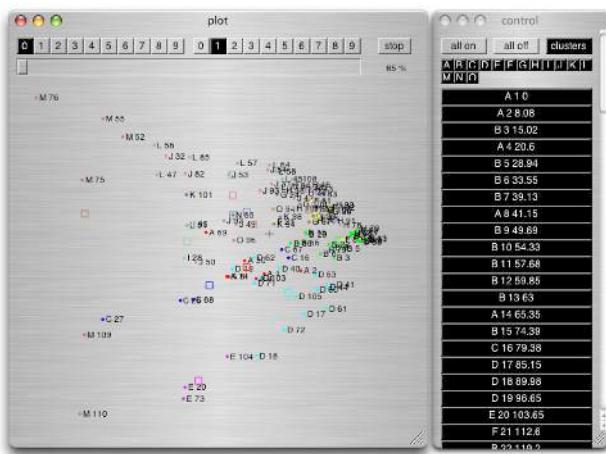
## 2.5 Clustering (associazione paradigmatica)

Le unità musicali collocate nello spazio MDS vengono distribuite in un numero arbitrario di insiemi (vale a dire, attribuite a paradigmi di riferimento) mediante algoritmo k-means++.

## 2.6 Interfaccia esplorativa

Per esplorare l'output di MDS e clustering è stata programmata un'interfaccia interattiva provvisoria che rappresenta le unità musicali e i centroidi dei paradigmi su un grafico bidimensionale.

L’interfaccia permette di impostare le dimensioni del grafico, la visualizzazione delle unità singolarmente e in gruppi, nonché l’audizione delle unità stesse cliccando sul punto che le rappresenta.



**Figura 1.** Interfaccia esplorativa provvisoria (Åke Parmerud, *Jeux Imaginaires*: spazio MDS/ED, dimensioni 0,1)

Ponendo la medesima dimensione sui due assi e scorrendo mediante l'ascolto i punti allineati sulla diagonale, è possibile valutarne la significatività e formulare ipotesi interpretative.

Visualizzando le unità per coppie di dimensioni significative e integrando nella rappresentazione i centroidi è possibile valutare la diffusione/concentrazione di quel particolare insieme e quindi, mediante l'ascolto, la consistenza del paradigma dal lato estesico.

### 3. RISULTATI

La procedura è stata attuata sperimentalmente su un insieme di brani acusmatici (Horacio Vaggione, *Agon, Ash*; Jean-Claude Risset, *Aer, Aqua, Focus, Terra*; Bernard Parmegiani, *Incidences/Résonances, Matières induites, Murmurants*; Åke Parmerud, *Jeux Imaginaires, Repulse*) segmentati auditivamente con l'ausilio di sonogrammi dallo scrivente e da musicisti in possesso di specifiche competenze sul repertorio. I risultati vengono esaminati sulla base di *Jeux Imaginaries* di Åke Parmerud<sup>10</sup>.

fattori caratterizzanti la percezione dei timbri strumentali, a partire da valutazioni soggettive di somiglianza tra suoni risintetizzati mediante elaboratore.

<sup>10</sup> Åke Parmerud, *Osynlig Musik/Invisible Music/Musique invisible*, Phono Suecia, 1994. *Jeux Imaginaires* è una commissione GMEB 1993.

un'opera particolarmente refrattaria all'analisi paradigmatica contraddistinta da permanente instabilità testurale e gestualità continue.

### 3.1 Descrittori

La sospensione del giudizio di pertinenza relativamente alle grandezze percettive messe in gioco nell'ascolto della musica acusmatica ha portato a individuare un insieme di descrittori di basso livello ampio e a optare per un calcolo della similarità senza attribuire ad essi pesi specifici.

Attribuendo pesi specifici molto eterogenei, sono state effettuate molteplici controprove per stabilire l'eventuale necessità di una maggiore caratterizzazione dei descrittori stessi rispetto al repertorio in esame: la scarsa variabilità dei risultati ha suggerito tuttavia il contrario.

Rimane comunque la possibilità di individuare "pacchetti" di descrittori di volta in volta differenti, nell'ottica di validare la procedura mediante ulteriori riscontri ottenuti dall'applicazione percepitiva del metodo paradigmatico.

### 3.2 Calcolo della similarità

Poiché l'algoritmo DTW privilegia l'individuazione di identità, le dissimilarità computate tra le unità musicali risultano generalmente elevate. L'esecuzione sulle 110 unità di *Jeux Imaginaries* (tab. 1) ha evidenziato un rapporto medio di 0.9 tra la lunghezza dei percorsi di allineamento individuati e la massima lunghezza computabile, con un minimo di 0.6 e picchi che sfiorano l'unità. Nonostante le restrizioni imposte al calcolo delle corrispondenze a vantaggio del percorso più breve, le singolarità risultano comunque eccessive.

In effetti, nell'ambito del repertorio musicale elettrao-  
custico le tecniche di sequence alignment mediante DTW  
sono state applicate proficuamente soltanto nella compa-  
razione delle *varianti* di testi sonori in studi svolti secon-  
do una prospettiva genetica [13]. Si è deciso quindi di  
eseguire un algoritmo DDTW, sensibile all'identità dei  
profili e quindi apparentemente più adatto allo scopo:  
nonostante, sono stati ottenuti punteggi analoghi.

L'algoritmo ED eseguito sulla base delle sequenze contenenti l'indicizzazione delle celle (30 insiemi di riferimento) porta a costi di allineamento inferiori rispetto agli algoritmi DTW e DDTW: in *Jeux Imaginaires* il rapporto medio tra costi di editing e massimi teorici risulta 0.87, con un minimo di 0.19 e un picco pari all'unità<sup>11</sup>. Questa soluzione produce una matrice di similarità notevolmente diversa rispetto a DTW e DDTW.

### 3.3 Scaling multidimensionale

In base alle prove effettuate, la matrice di coordinate prodotta dall'algoritmo MDS ordinata in base agli autovalori dimostra un'alta significatività mediamente per il 25% delle dimensioni (terzo quartile degli autovalori): questo avviene partendo da ciascuna matrice di similarità discendente da DTW, DDTW, ED. Anche se la percentuale varia in proporzione alla complessità della composizione

<sup>11</sup> Queste cifre variano chiaramente a seconda del numero di centroidi precedentemente impostati nell'algoritmo k-means++ di indicizzazione.

elettroacustica e al numero delle unità segmentate, si tratta tipicamente di 15-30 dimensioni (28 in *Jeux Imaginaires*): troppe per essere decifrate con sicurezza all'ascolto.<sup>12</sup>

Comparando auditivamente le unità vicini in nella singola dimensione e, soprattutto, osservando la collocazione di unità vicini in base a coppie di dimensioni, si evidenziano tuttavia importanti nessi morfologici consistenti in identità, somiglianze e analogie di comportamento sonoro.

La constatazione dei nessi morfologici è agevolato nello spazio MDS discendente da ED per via della distribuzione articolata e filamentosa<sup>13</sup> (figg. 1-2), rispetto allo spazio discendente da DTW/DDTW che porta tendenzialmente a visualizzazioni di ammassi concentrati intorno a un unico nucleo (figg. 3-4). Seguendo i singoli filamenti è possibile riscontrare nessi morfologici anche tra unità non vicine.

### 3.4 Clustering e valutazione auditiva della consistenza paradigmatica

L'algoritmo di clustering delle unità musicali eseguito sulla base delle coordinate MDS produce purtroppo risultati mutevoli anche utilizzando la variante k-means++. L'output dell'algoritmo si stabilizza moderatamente, nel caso in cui si impongano come centroidi iniziali unità particolarmente caratterizzate. Limitando le dimensioni valutate dall'algoritmo a quelle più significative i risultati si stabilizzano ulteriormente, peggiorando però la consistenza al riscontro auditivo.

L'ulteriore problema è la frequente presenza di insiemi costituiti da un singolo membro, non interpretabili in accordo al metodo di analisi paradigmatica come "resti".

Al di là delle limitazioni dell'algoritmo, in base alla valutazione dello scrivente la consistenza paradigmatica risulta da abbastanza a molto soddisfacente al riscontro auditivo in tutte le varianti procedurali, specialmente distribuendo le unità in 10-15 insiemi di riferimento.

Nella variante che discende da ED vi sono ulteriori margini di miglioramento, impostando il numero più adatto di centroidi nel passaggio procedurale iniziale di indicizzazione delle celle<sup>14</sup>.

<sup>12</sup> Come si è rilevato sopra, nel caso discendente da ED la collocazione delle unità e l'interpretabilità delle dimensioni dipende anche dal numero di centroidi utilizzati nella preliminare indicizzazione delle celle di informazione, e cioè, dalla giusta caratterizzazione del flusso sonoro, cercando di evitare sia la sovraccaratterizzazione, sia l'eccessiva semplificazione.

<sup>13</sup> In *Jeux Imaginaires* questo tipo di distribuzione si evidenzia tra le coppie di dimensioni corrispondenti ai 5-6 autovalori maggiori.

<sup>14</sup> Se per corrispondere alle sfumature del linguaggio acustico della composizione in esame si impone un numero notevole di centroidi (50), per conferire consistenza al clustering delle unità musicali si è dimostrato necessario adottare un più elevato numero di paradigmi di riferimento. Il numero di centroidi per l'indicizzazione preliminare deve determinarsi di volta in volta sperimentalmente, in funzione della complessità del linguaggio musicale, del numero di unità in cui è stata segmentata l'opera e del numero di paradigmi cui si vorranno attribuire le unità stesse. I risultati migliori su composizioni acustiche di media-elevata complessità e segmentate in 70-100 unità sono stati ottenuti impostando inizialmente 30 centroidi e distribuendo le unità in non più di 15 insiemi.

### 3.5 Valutazione delle tavole paradigmatiche

I risultati discendenti da DTW, DDTW ed ED producono tavole paradigmatiche notevolmente differenti (tab. 2). Si nota in particolare la persistente riconduzione a 3-4 paradigmi principali operata dalle varianti DTW e DDTW, rispetto ai 5 paradigmi della variante ED. Si nota inoltre la netta divaricazione tra paradigmi popolati e sparsi (DTW, DDTW) rispetto a una distribuzione più uniforme (ED).

La variante MDS/ED formalizza il flusso musicale in modo più articolato e corrisponde maggiormente alla complessità del brano *Jeux Imaginaires*. Pertanto, pur producendosi analoghi snodi macro-formali e risultando altrettanto buone consistenze paradigmatiche, delle tre varianti la soluzione MDS/ED sembra più indicata alla prosecuzione dell'analisi mediante accorpamento delle unità e formulazione di strutture paradigmatiche di più alto livello<sup>15</sup>.

## 4. CONCLUSIONI E SVILUPPI FUTURI

La caratteristica distribuzione delle unità musicali nello spazio MDS che discende dalla preliminare indicizzazione delle celle di informazione e computa la ED sembra più interessante rispetto alle possibilità esplorative e l'attribuzione delle unità ai paradigmi di riferimento più pertinente gli obiettivi di comprensione formale<sup>16</sup>.

A differenza delle soluzioni discendenti da DTW e DDTW, è stato possibile confrontare la specifica tavola paradigmatica così ottenuta con altre tavole ricavate in precedenza dallo scrivente su basi percettive, mettendole utilmente in discussione per quel che riguarda la tendenza a redistribuire a posteriori i gruppi paradigmatici in forza di superiori intuizioni formali, spieganti meglio la consistenza dell'opera, laddove la valutazione dei tratti componenziali si dimostrava particolarmente difficoltosa o portava a risultati dubbi.

Il riscontro fornito dalla procedura appare quindi come un possibile strumento di verifica nell'attribuzione percettiva delle unità musicali ai paradigmi di riferimento e un ausilio per la formulazione di ipotesi di pertinenza in merito ai tratti componenziali stessi.

I risultati ottenuti incoraggiano ad avviare una ulteriore fase di verifica nell'ambito dei corsi accademici di primo livello dell'alta formazione artistica e musicale, estendendo l'applicazione di tale procedura di analisi mediante elaboratore parallelamente alle esperienze già normalmente svolte su basi percettive, individuando un

<sup>15</sup> Cfr. ad esempio la sensibilità rispetto all'importante "ritorno" all'unità 86.

<sup>16</sup> Tale successione procedurale è fra l'altro più aderente al modello analitico, in quanto l'indicizzazione costituisce *de facto* un procedimento di associazione paradigmatica, seppure a livello microscopico, da cui viene originato un insieme finito di "sonemi" i cui raggruppamenti costituiscono a loro volta i "sintagmi" o "lessemi" caratteristici del linguaggio della composizione presa in esame. A questo proposito, cfr. la tipizzazione "atomica" proposta in [2] - un contributo ignoto allo scrivente durante la formulazione della procedura, che si muove con finalità diverse in un orizzonte metodologico non dissimile (tipizzazione mediante clustering, individuazione di archetipi mediante modelli marcoviani).

preciso programma di sperimentazione coinvolgente un maggior numero di soggetti.

Limitatamente allo sviluppo degli algoritmi, sono previste comunque ulteriori implementazioni:

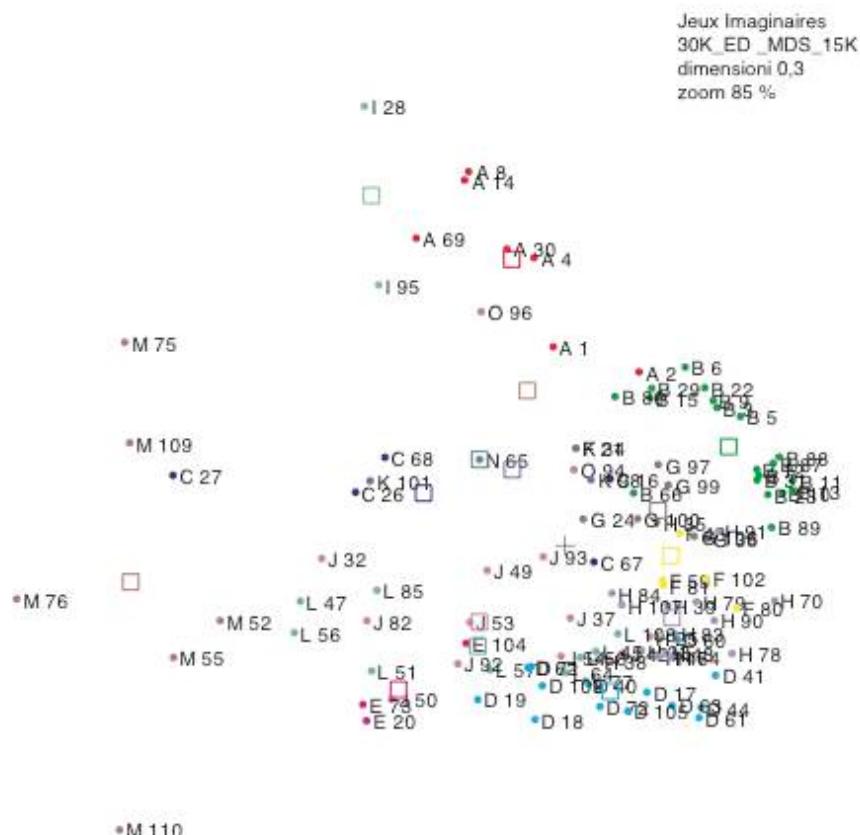
- opzione di individuare diversi “pacchetti” di descrittori di basso livello per analizzare il flusso sonoro;
- applicazione ricorsiva dell’associazione paradigmatica nella fase di indicizzazione delle celle (reimplementando quindi l’operazione di accorpamento finale del metodo classico dell’analisi paradigmatica), in modo da ottenere sequenze brevi costituite da pochi simboli e rendere più pertinente l’algoritmo ED;
- nel calcolo della edit distance, sostituzione della distanza Levenshtein con la distanza Damerau-Levenshtein; implementazione dei metodi Longest Common Subsequence (LCS) ed Elastic Matching (EM);
- miglioramento dell’output degli algoritmi di clustering mediante diverse metriche e metodi di raggruppamento;
- implementazione di Individual difference scaling (INDSCAL) e algoritmi MDS non-metrici per poter interpretare meglio i fattori di caratterizzazione delle unità come complessi di tratti componenziali;
- eventuale applicazione della procedura su unità individuate mediante segmentazione automatica;
- implementazione dell’accorpamento automatico delle unità per la formulazione di strutture paradigmatiche di più alto livello.

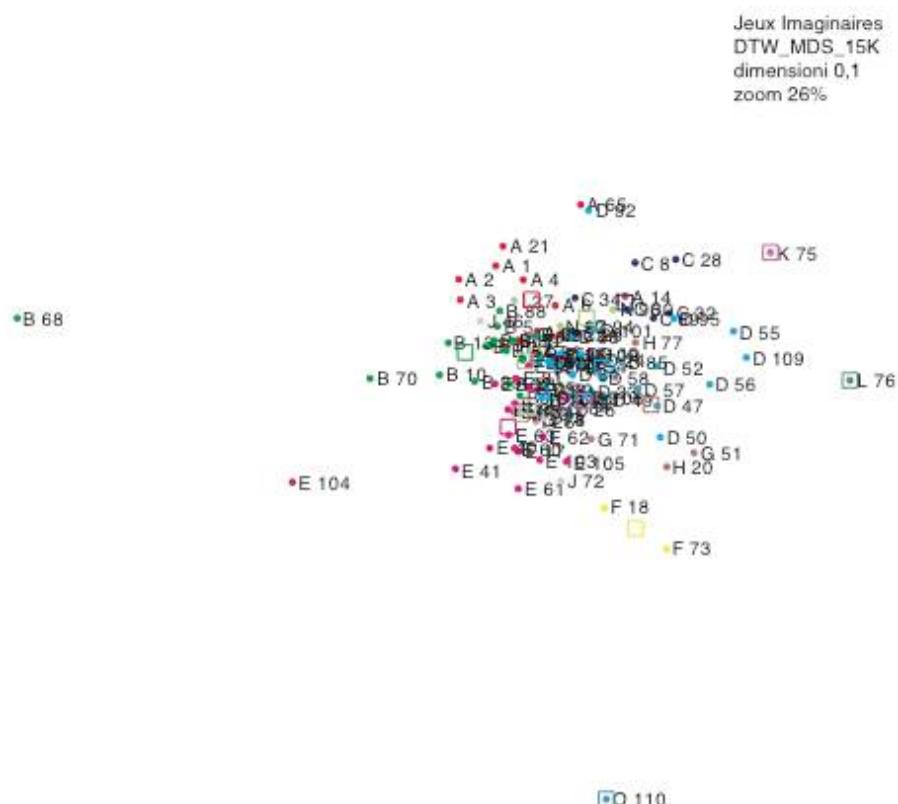
## 5. BIBLIOGRAFIA

- [1] D. Arthur and S. Vassilvitskii: “k-means++: the advantages of careful seeding”, *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027-1035, 2007.
- [2] C. E. Celli: “Sound-types: A new framework for symbolic sound analysis and synthesis”, *Proceedings of the International Computer Music Conference*, Huddersfield, pp. 179-187, 2011.
- [3] N. Collins: “SCMIR: A SuperCollider Music Information Retrieval Library”, *Proceedings of the International Computer Music Conference*, Huddersfield, pp.499-502, 2011.
- [4] J. M. Grey: *An Exploration of Musical Timbre*, Ph.D. dissertation, Center for Computer Research in Music and Acoustics, Stanford University, pp. 57-69, 1975.
- [5] N. Donin, J. Goldman: “Charting the Score in a Multimedia Context: the Case of Paradigmatic Analysis”, *Music Theory Online*, Vol. 14, No. 4, 2008.
- [6] E. J. Keogh and M. J. Pazzani: “Derivative Dynamic Time Warping”, *SIAM International Conference on Data Mining*, 2001.
- [7] O. Lartillot, P. Toivainen, and T. Eerola: “A Matlab Toolbox for Music Information Retrieval”, in C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker (eds): *Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, pp. 261-268, 2008.
- [8] V. Levenshtein: “Binary codes capable of correcting deletions, insertions, and reversals”, *Soviet Physics Doklady*, No. 10, pp. 707-710, 1966.
- [9] M. Müller: *Information retrieval for Music and Motion*, Springer, pp. 69-84, 2007.
- [10] J.-J. Nattiez: “Varèse’s «Density 21.5»: A study in semiological analysis”, *Music Analysis*, Vol. 1, No. 3, pp. 243-340, 1982.
- [11] J.-J. Nattiez: “Syrinx di Debussy: un’analisi paradigmatica”, in L. Marconi and G. Stefani (eds), *Il senso in musica*, CLUEB, Bologna, pp. 93-138, 1987.
- [12] F. Nucibella, S. Porcelluzzi, and L. Zattra: “Computer music analysis via a multidisciplinary approach”, *Proceedings of the Sound and Music Computing Conference*, Salerno, 2005.
- [13] N. Orio and L. Zattra: “Audiomatching for the phonological analysis of electroacoustic music”, *Proceedings of the International Computer Music Conference*, Copenhagen, pp. 157-164, 2007.
- [14] E. Pollastri: “Melody-Retrieval based on Pitch-Tracking and String-Matching Methods”, *Proceedings of the XII Colloquium on Musical Informatics*, Università di Udine, Gorizia, pp. 151-154, 1998.
- [15] J. Ricard and P. Herrera: “Using morphological description for generic sound retrieval”, *Proceedings of Fourth International Conference on Music Information Retrieval*, Baltimore, 2003.
- [16] N. Ruwet: *Langage, musique, poésie*, Seuil, Paris, 1972.
- [17] S. Roy: *L’analyse des musiques électroacoustiques*, L’Harmattan, Paris, 2005.

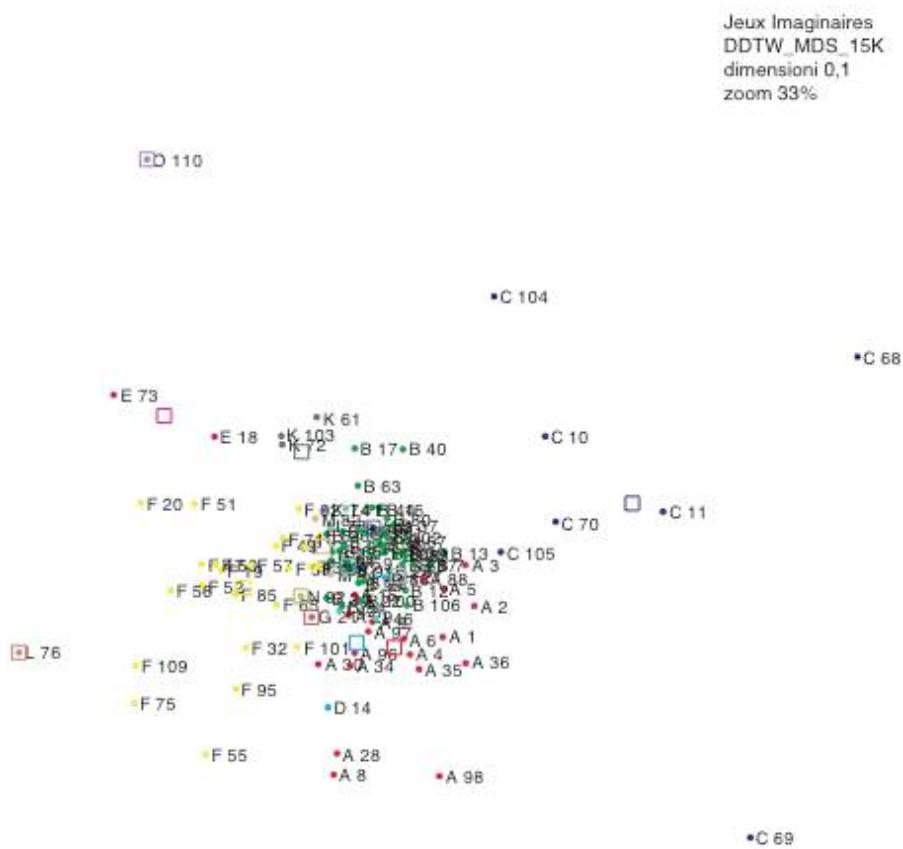
| unità : i.iniziale |
|--------------------|--------------------|--------------------|--------------------|--------------------|
| 001 : 00.00.00     | 023 : 02.04.00     | 045 : 04.03.00     | 067 : 06.19.00     | 089 : 08.31.00     |
| 002 : 00.08.10     | 024 : 02.06.10     | 046 : 04.07.60     | 068 : 06.24.75     | 090 : 08.32.75     |
| 003 : 00.15.00     | 025 : 02.13.55     | 047 : 04.11.00     | 069 : 06.34.00     | 091 : 08.34.40     |
| 004 : 00.20.60     | 026 : 02.15.40     | 048 : 04.19.45     | 070 : 06.43.40     | 092 : 08.37.00     |
| 005 : 00.28.95     | 027 : 02.24.50     | 049 : 04.24.15     | 071 : 06.46.90     | 093 : 08.44.00     |
| 006 : 00.33.55     | 028 : 02.37.30     | 050 : 04.30.70     | 072 : 06.53.00     | 094 : 08.50.25     |
| 007 : 00.39.15     | 029 : 02.47.50     | 051 : 04.38.30     | 073 : 06.58.55     | 095 : 08.56.15     |
| 008 : 00.41.15     | 030 : 02.52.35     | 052 : 04.46.40     | 074 : 07.07.55     | 096 : 09.05.50     |
| 009 : 00.49.70     | 031 : 02.59.80     | 053 : 04.56.10     | 075 : 07.11.50     | 097 : 09.13.30     |
| 010 : 00.54.35     | 032 : 03.02.10     | 054 : 05.02.80     | 076 : 07.23.65     | 098 : 09.18.30     |
| 011 : 00.57.70     | 033 : 03.10.90     | 055 : 05.07.85     | 077 : 07.36.75     | 099 : 09.24.30     |
| 012 : 00.59.85     | 034 : 03.15.00     | 056 : 05.17.90     | 078 : 07.41.35     | 100 : 09.28.90     |
| 013 : 01.03.00     | 035 : 03.21.20     | 057 : 05.26.35     | 079 : 07.44.30     | 101 : 09.33.90     |
| 014 : 01.05.35     | 036 : 03.25.05     | 058 : 05.32.40     | 080 : 07.48.00     | 102 : 09.43.25     |
| 015 : 01.14.40     | 037 : 03.29.00     | 059 : 05.37.25     | 081 : 07.50.85     | 103 : 09.46.55     |
| 016 : 01.19.40     | 038 : 03.34.35     | 060 : 05.41.95     | 082 : 07.55.25     | 104 : 09.52.80     |
| 017 : 01.25.15     | 039 : 03.38.85     | 061 : 05.46.50     | 083 : 08.03.25     | 105 : 10.00.90     |
| 018 : 01.30.00     | 040 : 03.42.40     | 062 : 05.51.10     | 084 : 08.06.45     | 106 : 10.05.80     |
| 019 : 01.36.65     | 041 : 03.47.95     | 063 : 05.57.35     | 085 : 08.11.00     | 107 : 10.09.80     |
| 020 : 01.43.65     | 042 : 03.51.40     | 064 : 06.01.35     | 086 : 08.18.55     | 108 : 10.14.00     |
| 021 : 01.52.60     | 043 : 03.56.00     | 065 : 06.06.15     | 087 : 08.24.70     | 109 : 10.18.00     |
| 022 : 01.59.20     | 044 : 03.59.00     | 066 : 06.13.60     | 088 : 08.27.70     | 110 : 10.30.30     |

**Tabella 1.** Segmentazione di *Jeux Imaginaires* di Åke Parmerud





**Figura 3.** *Jeux Imaginaries*: spazio MDS/DTW (dimensioni 0,1)



**Figura 4.** *Jeux Imaginaries*: spazio MDS/DDTW (dimensioni 0,1)

**Tabella 2.** Tavola paradigmatica di *Jeux Imaginaires* discendente da DTW, DDTW, ED

# *Analogique B*

## A computer model of the compositional process

**Andrea Arcella**

Conservatorio di Napoli

andrea.arcella@gmail.com

**Stefano Silvestri**

Conservatorio di Napoli

stefanodrive\_1@libero.it

### ABSTRACT

In this paper we propose a software reconstruction of Xenakis' composition *Analogique B* - namely, the electronically-generated tape for the mixed work *Analogique A et B* (1958-59). We briefly illustrate both Xenakis' "mechanism" (the compositional algorithm) and the sound synthesis method he implemented and explored, and some music-analytical evidence concerning the *Analogique B* tape. We then illustrate our digital implementation (using C++ and Csound as the sound-synthesis engine), and discuss questions as to the rendering of the original analog process in the digital domain.

### 1. ANALYTICAL REMARKS ON THE *ANALOGIQUE B* COMPOSITIONAL PROCESS

*Analogique A et B* resulted from the superposition of two works, initially composed as separate: *Analogique A* for nine string instruments, and *Analogique B* for electronically generated sounds (realized partly at GRM in Paris and partly in Hermann Scherchen's studio in Gravesano). As he was working, Xenakis also described his compositional process in a paper later published (1963) as Chapitre 2 of *Musique formelles* [4]. Chapitre 2 describes the theory of Markovian stochastic music, as explored in both *Analogique A* and *Analogique B*. Papers [1] and [2] clarify many aspects of these works. We will only point out, here, the main concepts as related to what is needed for a digital implementation.

Both *Analogique A* and *Analogique B* are generated from one and the same statistical process, resulting into sequences of instances of 8 items, called "screens" (in the terminology of Formalized Music[4]), labelled A to H. The time sequence of screens is conditioned by a

*Copyright: © 2012 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.*

specific selection process, namely a 8x8 state transition matrix (MPT, matrice de probabilité de transition).

	A	B	C	D	E	F	G	H
A	0.021	0.357	0.084	0.189	0.165	0.204	0.408	0.096
B	0.084	0.089	0.076	0.126	0.15	0.136	0.072	0.144
C	0.084	0.323	0.021	0.126	0.15	0.036	0.272	0.144
D	0.336	0.081	0.019	0.084	0.135	0.024	0.048	0.216
E	0.019	0.063	0.336	0.171	0.11	0.306	0.102	0.064
F	0.076	0.016	0.304	0.114	0.1	0.204	0.018	0.096
G	0.076	0.057	0.084	0.114	0.1	0.054	0.068	0.096
H	0.304	0.014	0.076	0.076	0.09	0.036	0.012	0.144

**Figure 1.** MPT used by Xenakis for *Analogique A* and *Analogique B*

For example the probability that screen A will be followed by screen B is 35.7 % whereas the probability that screen A will be followed by screen C is 8.4 %.

Each screen is a function of three variables: a density value **d**, an amplitude interval **g** and a frequency interval **f**. As a particular screen is selected, the next step is to get random amplitude and frequency values from given intervals (described below). Thus, we can view each screen as an infinite set of parameter values.

We will focus, here, on the process as designed and utilized for *Analogique B* particularly. The sound synthesis is based on the theory of acoustical quanta put forth by Dennis Gabor [5]. *Analogique B* is considered the first musical work based what is known, today, as "granular synthesis" [6].

Xenakis' screens are defined as follows:

1) the full frequency range is partitioned in 16 "regions" (Xenakis' term). The ranges for the 16 frequency regions are:

f-region 1: 42-63 Hz, f-region 2: 63-84,  
f-region 3: 84-131, f-region 4: 131-178  
f-region 5: 178-267, f-region 6: 267-355  
f-region 7: 355-532, f-region 8: 532-710  
f-region 9: 710-1065, f-region 10: 1065-1420  
f-region 11: 1420-2130, f-region 12: 2130-2850  
f-region 13: 2850-4275, f-region 14: 4275-5700

f-region 15: 5700-8550, f-region 16: 8550-11400

Two frequency sets are defined as regions linked among themselves in this way:

$$f0=(1; 2; 3; 5; 7; 10; 13; 14; 15; 16)$$

$$f1=(4; 6; 6; 8; 9; 9; 11; 11; 12; 12)$$

2) Similarly, there are 4 amplitude regions, measured in phones<sup>1</sup>. Phone ranges are:

g-region 1: 50-60 phones

g-region 2: 60-70

g-region 3: 70-80

g-region 4: 80-90

Phone regions are combined together as follows, making up two phone sets:

$$g0 = (1, 1, 1, 1; 2, 2; 3, 3; 4, 4)$$

$$g1 = (1, 1, 1; 2, 2, 2; 3, 3; 4, 4)$$

3) Finally, there are 7 density regions, with density values measured in grains per second (gps):

$$\begin{aligned} d\text{-region 1} &= 1.3 \text{ gps} \\ d\text{-region 2} &= 3.9 \\ d\text{-region 3} &= 11.7 \\ d\text{-region 4} &= 35.1 \\ d\text{-region 5} &= 105.4 \\ d\text{-region 6} &= 315.9 \\ d\text{-region 7} &= 957.7 \end{aligned}$$

Here are the two density sets:

$$\begin{aligned} d0 &= (1, 1, 1; 2, 3, 3; 4, 5; 6, 6) \\ d1 &= (1; 2, 2, 2; 4, 4; 5, 5; 6, 6) \end{aligned}$$

Observe that the highest density range (d-region 7) is not comprised in the density sets actually utilized in the composition, probably due to the restrictions of the technological means available to Xenakis. A discussion on issues related to technological restrictions, in the making of this work, is sketched later on in the paper. The 8 "screens" (A, B, ... H) are thus defined by the combinatorics of 23 variables:

$$\begin{array}{ll} A = f0, g0, d0 & B = f0, g0, d1 \\ C = f0, g1, d0 & D = f0, g1, d1 \\ E = f1, g0, d0 & F = f1, g0, d1 \\ G = f1, g1, d0 & H = f1, g1, d1 \end{array}$$

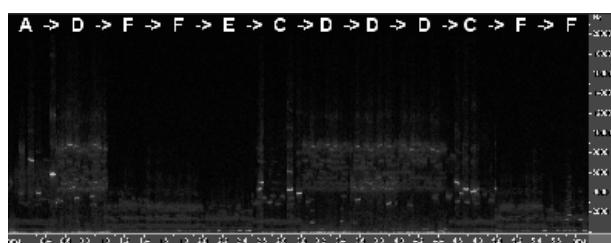
As the MPT is looped (i.e. iteratively applied), a sequence of screens is generated, that Xenakis calls a "protocol". Each subsequent screen in a protocol sets

the statistical musical content in a fixed time window of 0.5". Grain amplitudes and grain frequencies are thus selected from within the regions included in the current screen, and the number of sound grains in the time window is fixed, also based on the density range included. The entire piece is made with several protocols, i.e. several sequences of combinations of the 8 screens available.

## 2. TIME-FREQUENCY ANALYSIS OF THE TAPE SOUND

The tape sounds of *Analogique B* consist in 0.5"-long "clouds" (Xenakis' term), each made of the statistical content of one screen. There are two kinds of protocols: "equilibrium" protocols and "perturbation" protocols. "Equilibrium" means "statistical balance", i.e. protocols generated by iterating the 8x8 matrix illustrated above (Figure 1). "Perturbation" protocols are generated independent of the matrix (perturbated protocols are, in turn, of two possible types, as illustrated below). In other words, "perturbated" protocols are contradictory to the matrix process itself: they are to manifest, according to Xenakis, the logics of the algorithmic process by negating the process' own functioning [4]: the normal (statistical) behaviour of the process (protocols generated with the matrix) is alternated with behaviours which are not reducible to the matrix functioning (protocols generated with other means, not less formlisable, but not belonging to the fundamental matrix).

For our purposes, it is useful to analyse the frequency content of a sample segment<sup>2</sup>. We focus here on the first tape segment (0'00"-0'20"), consisting in two protocols (0'00"-0'15" and 0'15"-0'20"). Figure 2 shows a frequency analysis of the initial 6" of sound.

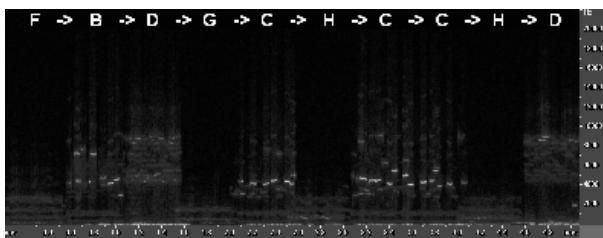


**Figure 2.** Frequency analysis of the initial 6" second of sound

This segment is clearly in "equilibrium", that is, the sequence of 12 screens correctly reflects the MTP statistical behaviour. The next sonogram (figure 3) shows the next 5", with the subsequent 10 screens:

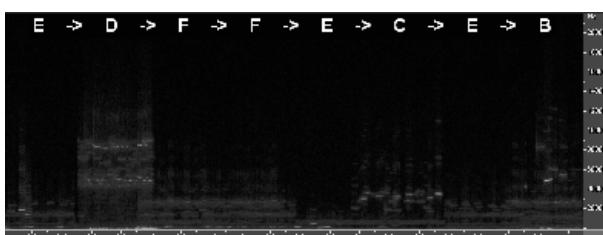
<sup>1</sup> The phon is a non-standard sound unit that is designed to reflect perceived loudness, and is based on psychoacoustic experiments in which volunteers were asked to adjust the decibel level of a reference tone of 1 kHz until it was the same loudness as the signal being measured. So for example, if a sound is 70 phons, that means it sounds as loud as a 70-dB, 1-kHz tone.

<sup>2</sup> The source sound for our sonogram analysis is a digital copy available from Salabert (Xenakis' publisher), and kindly provided by Prof. Makis Solomos and Agostino Di Scipio.



**Figure 3.** Frequency analysis of the the next 5"

Following is the sonogram of the next 4 seconds:



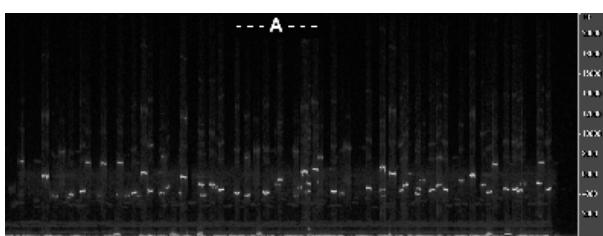
**Figure 4.** Frequency analysis of the the next 4"

From these sonograms, we can tentatively reconstruct the protocol unfolded in these initial 15", comprising 30 screens in total:

$Y = A \rightarrow D \rightarrow F \rightarrow F \rightarrow E \rightarrow C \rightarrow D \rightarrow D \rightarrow C \rightarrow C \rightarrow F \rightarrow F \rightarrow B \rightarrow D \rightarrow G \rightarrow C \rightarrow H \rightarrow C \rightarrow C \rightarrow H \rightarrow D \rightarrow E \rightarrow F \rightarrow E \rightarrow D \rightarrow F \rightarrow F \rightarrow E \rightarrow C \rightarrow E \rightarrow B$

This sequence, therefore, fully represents a possible state vector of the MPT stochastic process.

After the first 15", the tape has a completely different screen protocol, consisting of screen A repeated 10 times (figure 5). This is not to say that the sound signal repeats itself for 10 times, but that the statistical configuration is repeated, while the precise frequency content statistically reflects the prescribed (repeated) ranges. Overall, this is not anything possibly achieved with the Xenakis' MPT: this is a "perturbated" protocol, not at all reflecting the stochastic process.



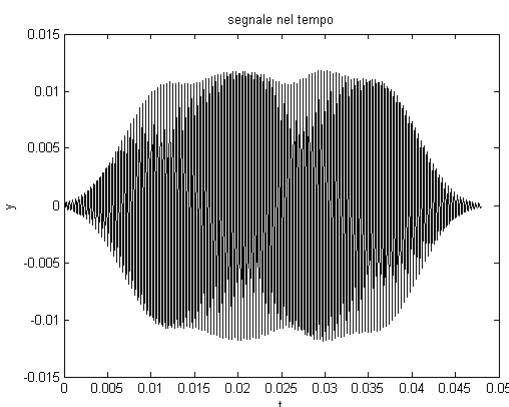
**Figure 5.** Perturbated protocol

This is a very simple perturbation (or variation) of the given MTP:  $P(A | A) = 1$  (the transition from A to A is certain, no other transitions are possible). There is a second type of "perturbation", a simple statistical distribution. We won't go into further details on the issue, here, as we will only be concerned, in our digital implementation, with the normal use of Xenakis' MTP, in "equilibrium" mode. In actuality, the two types of perturbations are still formalisable as two different MTPs.

In further work, we will accordingly implement the whole process with three different MTPs.

### 3. XENAKIS' GRAINS: THEORY AND PRACTICE

Let's take the opportunity here for a short detour on the sound grain description. Xenakis wrote that his grains were 40 ms in durations, and mentioned no grain envelope. That is not confirmed by the scrutiny of the tape signals, which evidences grain envelopes of different shapes. In Figure 6, a single grain is illustrated, isolated from the sound sequence analysed in Figure 5:



**Figure 6.** Single grain

In Xenakis' theory, the grain should have been a rectangle window<sup>3</sup>:

$$rect(t) = u\left(t + \frac{1}{2}\right) - u\left(t - \frac{1}{2}\right) \quad (1)$$

Actual envelopes are smoothed, as we clearly see in Figure 6. That can be due to several factors, especially relating to the studio technologies available to the composer at the time. Among the several hypothesis, it has been also noted [1] that Xenakis mentioned the use of "filters" in the making of this work, not going into any details. Use of band-pass filters could clearly explain envelope shapes such as the one illustrated in Figure 6<sup>4</sup>.

### 4. XENAKIS IN THE STUDIO

But how Xenakis did achieve his granular textures and how did he practically arrange the sonic values according to his Markovian process (or the negation, "perturbation" of it)? Following [1], we know that Xenakis created several separate tapes, each with granular materials corresponding to the frequency/amplitude/density regions. His basic grains were short sine-wave signals,

<sup>3</sup> That already represents a significant deviation from Gabor's theory of acoustical quanta [6].

<sup>4</sup> As we know from theory of signals, every filter that cuts high frequencies in the frequency domain also modifies the temporal shape of the signal that results more smoothed.

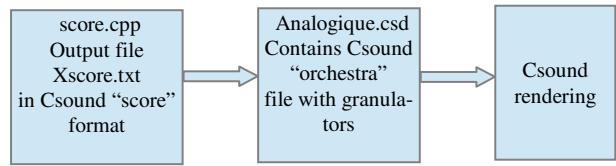
approximately lasting 0.04" (empirical evidence contradicts this datum, as grain durations seem to span over durations longer than that). Because grain densities prescribed by the composer go up to such a relevant figure of 300 gpus, Xenakis should have undergone an incredible, huge amount of mechanical cut-and-past work to achieve that. It is reasonable to agree with [1], where the hypothesis is made that he devised a much more economic procedure, simplifying his task in the studio. In **5.1** short, we can summarize the process as follows:

- 16 longer grain sequences were created, corresponding to the 16 frequency regions, all having normalized peak amplitude and minimal density (1.3 gpus);
- tape segments were extracted and mixed together to create higher grain density sequences, as **5.2** prescribed by the selected screens;
- the mix thus obtained was scaled to achieve the amplitude value as prescribed by the selected screen.

This out-of-time, additive procedure, which is anyway rather cumbersome and time-consuming, allowed Xenakis to get fractional density values (see the density regions, listed above). That represents a problematic point for our digital implementation (to be discussed later). This elaborate studio procedure is of little import for our software implementation; however, it suggested us a possible modeling strategy, maybe closer to the actual decision making process Xenakis had to deal with. Other implementations are indeed possible and may reveal more effective (in fact, one of the authors followed a different path in an independent work [7]; another attempt is described in [9]), but our goal here is to stay as close as possible to the path designed by Xenakis for himself. This is like leaning more on the analysis of the compositional process than on the analysis of the compositional results [8].

## 5. SOFTWARE IMPLEMENTATION OF ANALOGIQUE B COMPOSITIONAL ENGINE

Our software implementation factors the whole problem in two: it splits into two software modules; the first written in C++ language<sup>5</sup> generates the screen sequence (i.e. It creates the "protocols"), based on the Xenakis MPT (see Figure 1). The second module is written in Csound<sup>6</sup> and implements the granular synthesis process, driven by the screen values:



**Figure 7.** Single grain

### 5.1 The algorithm engine

The C++ program outputs a text file with the Csound "score" format. The main part of the C++ code is a loop iterating the MPT for as many times as specified by the program user. When launching the program, the user can set the first screen to start with, and the protocol length (number of MPT iterations).

### 5.2 The synthesis engine

he Csound "orchestra" file contains the sound synthesis code. This opcode takes as input the following parameters:

- a) the function that defines the waveform of the grain (the sine function) and the function that defines the windowing envelope:

```
itmp      ftgen 1, 0, 4096, 10, 1      ;generate sin table
gifnum = 1           ;grain waveform (sin)   ;called in grain
itemp2 ftgen 2, 0, 8192, 20, 9       ;generate sync table
giwf1 = 2            ;sync waveform - called in grain
```

- b) the grain duration, which in our case is constant and is 0.04"

- c) the frequency range from which the actual frequency value is randomly selected for each grain. The range is specified as a base frequency value (offset) and a range width. Here is an example for the first two regions:

; f-region definition (16 in tot)

```
gifreg1=42
gifinter1= 63-42 ; width is 11
gifreg2=63
gifinter2= 84-63 ; width is 21
```

- d) the range from which the amplitude value is randomly selected for each grain. Here is an example for the first two regions:

; g-region definition (4 in tot)

```
gigain1=2000
gigain_inter1=2000
gigain2=4000
gigain_inter2=4000
```

- e) the density values:

; d-region definition (5 in tot)

```
gidens0 = 1.3
gidens1 = 3.9
gidens2 = 11.7
gidens3 = 35.1
gidens4 = 105.4
gidens5 = 315.9
```

<sup>5</sup> Standard c++ language. We have used open source Dev-C++ IDE with mingw compiler (<http://www.bloodshed.net/dev/>)

<sup>6</sup> Csound v 5.15 <http://www.csounds.com/>

We define 8 synthesis "instruments", one for each of the 8 screens, A to H. Each instrument has 10 overlapping grain generators. Each generator has a fixed frequency range (related to the f0 set and f1 set). Here is an example for the screen B instrument:

```

instr 2                               ; B screen
aenv      linseg 0, 0.001, 1, 0.498, 1, 0.001, 0
a1       grain   gigain1, gifreg1, gidens3, gigain_inter1,
        gifinter1, 0.04, gifnum, giwfn,0.04
a2       grain   gigain2, gifreg2, gidens1, gigain_inter2,
        gifinter2, 0.04, gifnum, giwfn,0.04
a3       grain   gigain4, gifreg3, gidens4, gigain_inter4,
        gifinter3, 0.04, gifnum, giwfn,0.04
a4       grain   gigain3, gifreg5, gidens3, gigain_inter3,
        gifinter5, 0.04, gifnum, giwfn,0.04
a5       grain   gigain1, gifreg7, gidens1, gigain_inter1,
        gifinter7, 0.04, gifnum, giwfn,0.04
a6       grain   gigain1, gifreg10, gidens0, gigain_inter1,
        gifinter10, 0.04, gifnum, giwfn,0.04
a7       grain   gigain2, gifreg13, gidens1, gigain_inter2,
        gifinter13, 0.04, gifnum, giwfn,0.04
a8       grain   gigain1, gifreg14, gidens5, gigain_inter1,
        gifinter14, 0.04, gifnum, giwfn,0.04
a9       grain   gigain3, gifreg15, gidens4, gigain_inter3,
        gifinter15, 0.04, gifnum, giwfn,0.04
a10      grain  gigain4, gifreg16, gidens5, gigain_inter4,
        gifinter16, 0.04, gifnum, giwfn,0.04
out
aenv*(a1/10+a2/10+a3/10+a4/10+a5/10+a6/10+a7/10+a8/10+a
9/10+a10/10)
endin

```

## 6. CONCLUSION

The digital rendering of Analogique B could be done in two different ways:

- 1) as described in this paper
- 2) creating a digital version of Xenakis' tapes with a series of audio files corresponding to the screens' montage made with analog tapes, then writing a Csound orchestra that implements a sampler addressing those files.

Variants of the first approach would be required for realtime versions [9].

The second method would be closer to the way of work adopted by Xenakis with analog technology, but it would involve a smaller amount of automation and more handiwork. Our choice followed not merely from the available technology, but from a design strategy making full advantage of the digital domain. Tools and technologies used to produce a musical work are not neutral but incorporate knowledge that influence the choices of the composer.

## 7. REFERENCES

- [1] A. Di Scipio: "An analysis of Analogique B", *Unpublished*.
- [2] A. Di Scipio: "Le nuvole di suono e i loro meccanismi". Uno studio di Analogique A et B", *Atti del Convegno Internazionale Iannis Xenakis - Milano 2005*,

*Quaderni della Civica Scuola di Musica di Milano*, 2006.

[3] A. Di Scipio, "Da Concret PH A Gendy301 – modelli compositivi nella musica elettroacustica di Xenakis", *Sonus* 7 pp. 1-2-3, 1995.

[4] I. Xenakis, *Musique Formelles*, special issue of *La Revue Musicale*, pp. 23-24, 1963.

*Formalized Music*, translation by Christopher Butchers, G. H. Hopkins, John Challifour, Bloomington, University Press, 1971.

*Formalized Music* New revised edition, additional material compiled, edited and translated by Sharon Kanach, Stuyvesant (New York), Pendragon Press, 1992.

[5] C. Roads, *The Computer Music Tutorial*, Mit Press 1996.

[6] D. Gabor, "Acoustical Quanta and the Theory of Hearing", *Nature* 159, 591-594, 1947.

[7] S. Silvestri, "Studio e implementazione della macchina stocastica in Analogique A+B", *Unpublished*, 2011

[8] O. Laske, "Toward an Epistemology of Composition", *Interface*, Vol. 20, 1991.

[9] K. L Hagan, "Genetic Analysis of Analogique B", *Electroacoustic Music Studies Network – Montréal*, 2005.



# BACH: UNA LIBRERIA PER LA COMPOSIZIONE ASSISTITA IN MAX

**Andrea Agostini**

Conservatorio "G. Verdi" di Torino

and.agos@gmail.com

**Daniele Ghisi**

Freelance Composer

danieleghisi@gmail.com

## ABSTRACT

Viene spesso assunta una contrapposizione tra i sistemi per la composizione assistita da computer (Computer Aided Composition, o CAC), che permettono di generare e trasformare dati musicali simbolici, e gli ambienti *real-time* e i sequencer. Questa contrapposizione è essenzialmente metodologica: negli ambienti di CAC tradizionali, le interazioni con l'interfaccia non hanno effetto finché non viene eseguita un'operazione di *refresh*; al contrario, gli ambienti *real-time* reagiscono immediatamente alle interazioni dell'utente. In quest'articolo presenteremo la libreria *bach: automated composer's helper* per Max, che introduce funzionalità avanzate di notazione musicale ed elaborazione simbolica in un tipico ambiente *real-time*. La libreria si pone lo scopo di ricomporre la frattura che separa la CAC dal mondo del *real-time*, dal momento che riteniamo l'interattività un aspetto performativo essenziale nel processo di scoperta musicale.

## 1. INTRODUZIONE

Sin dagli albori dell'informatica c'è stato grande interesse su come si potesse trarre vantaggio dalla precisione, dalla velocità e dalla potenza di calcolo dei computer nelle attività musicali. Probabilmente, la direzione più nota (e di maggior successo commerciale) si è rivelata essere la generazione e trasformazione del suono. Negli ultimi anni, i personal computer (e persino alcuni *smartphone* di fascia alta) hanno raggiunto una potenza sufficiente a calcolare trasformazioni e generazioni audio in tempo reale con qualità professionale. D'altra parte, sono stati sviluppati svariati sistemi per elaborare dati simbolici piuttosto che acustici - 'note' piuttosto che 'suoni'. Questi ultimi sistemi possono essere sommariamente divisi in strumenti per la videoscrittura musicale (Finale, Sibelius, Lilypond, ...) e strumenti per la composizione assistita da computer (o *computer-aided composition*, ovvero generazione e trasformazione di dati musicali simbolici, come OpenMusic, PWGL, Common Music, ...). Inoltre, almeno due ambienti di programmazione grafica, Max e Pd (che peraltro sono strettamente imparentati tra loro) hanno tra i loro principali obiettivi il controllo MIDI e l'elaborazione e generazione audio, ma allo stesso tempo sono capaci di trattare video, collezioni di dati arbitrarie e le più disparate periferiche di

input/output. Di fatto, i confini che separano tutte queste categorie sono tutt'altro che netti: i sistemi di videoscrittura musicale spesso offrono strumenti non banali per l'elaborazione di dati; quasi tutti i sequencer permettono anche trattamenti audio, e alcuni possono gestire rappresentazioni grafiche di alta qualità di partiture; gli ambienti più moderni per la *computer-aided composition* (CAC) includono strumenti per la sintesi e la trasformazione sonora. Al tempo stesso, è importante osservare come Max e Pd offrano strumenti di sequencing molto limitati, ed essenzialmente nessun supporto per la notazione musicale.

Si può tracciare un'altra distinzione, ortogonale alla precedente, tra sistemi in tempo reale, che reagiscono 'immediatamente' alle azioni sull'interfaccia (come Finale, Max, ProTools, ...) e sistemi in tempo differito, dove queste azioni non hanno effetto finché non viene esplicitamente effettuata un'operazione di *refresh* (come Lilypond, OpenMusic [1], PWGL [2], ...). Quest'ultimo è il caso tipico degli ambienti CAC, ma noi riteniamo che non ci siano ragioni intrinseche per cui l'elaborazione simbolica non possa essere eseguita in tempo reale. Il fatto che un ambiente per la composizione musicale sia basato su un paradigma in tempo reale o in tempo differito ha un'influenza profonda sulla natura ultima del processo compositivo. Un compositore che lavori con sequencer, plug-in e strumenti elettronici vuole che questi reagiscano immediatamente a variazioni nei loro parametri; allo stesso modo, un compositore che lavori con dati simbolici può desiderare che la macchina si adatti rapidamente a nuove configurazioni dei parametri. Basandoci sulla nostra personale esperienza di compositori, abbiamo maturato l'opinione che la creazione e l'intervento creativo su una partitura musicale non siano un'attività che si svolge fuori dal tempo, ma che invece seguì il processo di scoperta del compositore e si sviluppi di conseguenza.

Questo tema è già stato toccato da Miller Puckette [3]:

"While we have good paradigms for describing processes (such as in the Max or Pd programs as they stand today), and while much work has been done on representations of musical data (ranging from searchable databases of sound to Patchwork and OpenMusic, and including Pd's unfinished "data" editor), we lack a fluid mechanism for the two worlds to interoperate."

e in seguito approfondito da Arshia Cont [4].

In quest'articolo vogliamo mostrare che la CAC in tempo reale non solo è teoricamente possibile, ma può essere realizzata nella pratica attraverso il software *bach: auto-*

Copyright: ©2012 Andrea Agostini 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.

*mated composer's helper*, una libreria per Max da noi sviluppata, contenente un ampio insieme di strumenti per l'elaborazione simbolica e la rappresentazione grafica della musica e progettata in maniera da poter interagire con gli strumenti di Max per l'elaborazione audio, l'interazione in tempo reale e la programmazione grafica.

## 2. PARADIGMI DI PROGRAMMAZIONE

*bach* segue il paradigma generale di tipo *data-flow* di Max: l'utente costruisce una sorta di 'catena di montaggio' costituita da operatori estremamente specializzati. All'interno di questa metafora, i dati in ingresso al sistema vengono sottoposti a una serie di elaborazioni successive fino ad arrivare a una o più 'porte di uscita'. Non c'è bisogno di un comando esplicito di attivazione della catena di montaggio: virtualmente nel momento stesso in cui un dato arriva questo viene sottoposto alla catena predefinita di elaborazioni, che peraltro può essere arbitrariamente ramificata e soggetta a condizioni.

Possiamo invece osservare che il paradigma di programmazione di un ambiente tradizionale di CAC (come Patch-work [5], OpenMusic o PWGL) è profondamente diverso, dal momento che l'immissione di dati non provoca una reazione immediata. L'idea è piuttosto che l'utente fornisce al sistema un'intera configurazione di dati preliminari, e solo quando questa configurazione è completamente definita invia alla macchina un comando esplicito di valutazione. La differenza è meno sottile di quanto possa apparire, sia in termini concettuali che pratici. Naturalmente, il comando di valutazione consiste nella semplice pressione di un tasto, o in un clic del mouse; ma mentre questa azione può non essere considerata critica in un contesto di tempo differito, diventa cruciale quando il sincronismo è importante, o quando i dati arrivano ad esempio da una tastiera MIDI, da un microfono o da una videocamera. L'approccio in tempo differito impedisce quindi di reagire correttamente a un flusso di eventi in entrata, siano essi una sequenza di tasti suonati da uno strumentista o una serie di istruzioni che guidano il pensiero del compositore (e che da esse sono guidate). La distinzione tra sistemi in tempo reale e sistemi in tempo differito è stata ampiamente trattata, tra gli altri, da Carl Seleborg [6].

A questo punto dovrebbe essere chiaro come la scelta di implementare un sistema di CAC in un ambiente *data-flow* in tempo reale porti a conseguenze importanti nel tipo di scenario in cui il sistema stesso può essere impiegato. Se le azioni del compositore nell'interfaccia influenzano in tempo reale il risultato simbolico, il feedback della macchina è molto più esteso e intuitivo, e permette all'autore di sperimentare e valutare un insieme molto più ampio di possibilità. In questo senso, *bach* consente un approccio 'per tentativi' all'elaborazione dei dati simbolici - non dissimile a quello adottato da alcuni compositori di musica elettroacustica. *bach* ambisce a ridurre il gap che separa l'approccio 'performativo' e quello 'speculativo' negli strumenti per la computer music.

Infine, vale la pena di sottolineare che il paradigma *real-time* è una risorsa, non un obbligo: i cambiamenti alla partitura sono gestiti dall'utente che è completamente libero

di farli avvenire immediatamente o solo dopo un qualche tipo di operazione di *refresh* (come avviene nel caso degli ambienti in tempo differito). Questo significa che, in linea di principio, nulla impedisce di utilizzare *bach* alla stessa maniera di qualsiasi altro sistema di CAC: peraltro, ci sono casi (ad esempio la quantizzazione ritmica) in cui il tempo di calcolo richiesto da un'operazione è troppo lungo per poter ottenere una risposta immediata all'azione dell'utente.

## 3. LA LIBRERIA BACH

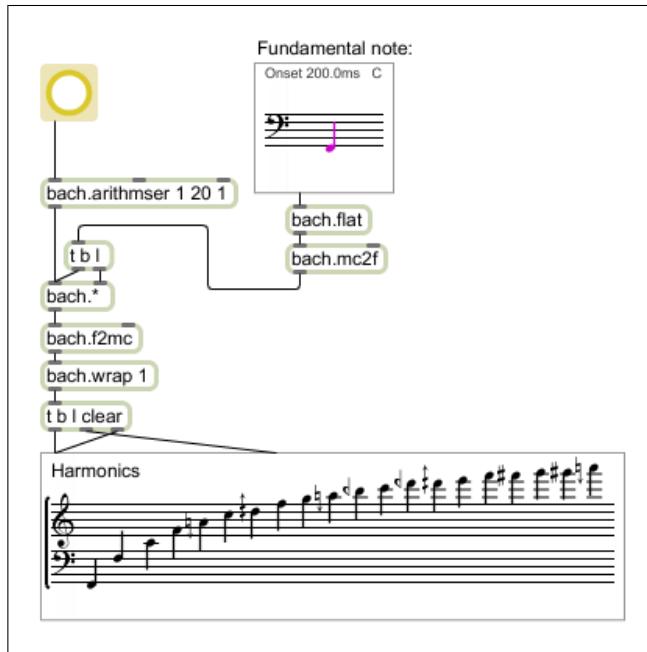
### 3.1 I componenti di *bach*

*bach* è una libreria di poco meno di 200 tra oggetti e astrazioni per il software Max. La distinzione tra oggetti e astrazioni riguarda l'implementazione di questi moduli, piuttosto che il loro effettivo utilizzo. Gli uni e gli altri appaiono come oggetti grafici che possono essere collegati l'un l'altro per costruire la 'catena di montaggio' che eseguirà le operazioni desiderate. La differenza sta nel fatto che gli oggetti sono scritti in C o C++ e, a meno che il codice sorgente non sia reso disponibile dallo sviluppatore, si comportano come 'scatole nere' non modificabili; le astrazioni al contrario sono combinazioni di oggetti e altre astrazioni (possiamo vederle come 'catene di montaggio' a loro volta), che l'utente può sempre aprire e modificare. Il costo di questa flessibilità è che un'astrazione è di solito molto meno efficiente di un oggetto che svolge la stessa funzione, e alcune funzionalità semplicemente non possono essere implementate senza ricorrere alla programmazione in C: ad esempio l'implementazione di nuovi tipi di dati e di operazioni grafiche avanzate, entrambe caratteristiche chiave di *bach*. In Fig. 1 è mostrata una patch di esempio, contenente alcuni oggetti e astrazioni della libreria *bach*.

Oltre alla vera e propria programmazione dei moduli software, è stata posta grande attenzione alla documentazione di *bach*. Ogni oggetto e astrazione ha un file di help che ne descrive completamente le funzionalità; una patch generale di help può essere richiamata dal menu *Extras* di Max; un insieme di tutorial permette di apprendere gradualmente le caratteristiche principali del sistema. Inoltre è attivo un forum internet degli utenti, e un nuovo gruppo di tutorial video è in fase di sviluppo.

### 3.2 Gli oggetti d'interfaccia

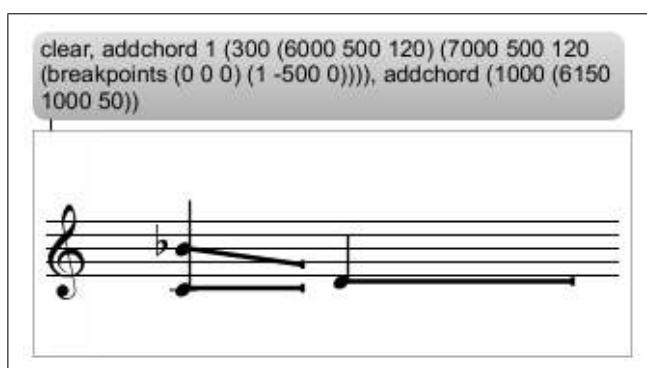
Abbiamo anticipato che una delle funzionalità principali di *bach* è la visualizzazione della notazione musicale. Questa funzionalità è implementata principalmente in due oggetti di interfaccia: *bach.score* e *bach.roll*. *bach.score* esprime il tempo in termini di unità musicali tradizionali, e include nozioni come pause, misure, metro e tempo; *bach.roll* esprime il tempo in termini di unità temporali assolute (ovvero millisecondi) e di conseguenza non ha nozione dei concetti temporali tradizionali: questo è utile per rappresentare musica non misurata, ma anche per trattare in maniera semplice materiale diastematico le cui informazioni temporali non siano note o siano irrilevanti. L'implementazione dei concetti temporali tradizionali in *bach.score* è



**Figure 1.** Una patch che calcola le prime venti armoniche di una nota. Per aggiornare il risultato, si può cliccare il pulsante giallo, o trascinare con il mouse la fondamentale in rosa.

piuttosto avanzata: permette tra l’altro metri, tempi e agogiche simultanei (vedi Fig. 5). Al di là di questa differenza fondamentale i due oggetti offrono un ampio insieme di caratteristiche comuni, tra cui:

- Editing attraverso tastiera e mouse, o messaggi Max (vedi Fig. 2): gli strumenti per l’editing combinano caratteristiche tipiche dei sistemi di videoscrittura WYSIWYG (p. es., creazione e modifica di note e accordi attraverso l’interazione grafica) e dei sequencer audio e MIDI (p. es., playback immediato di una nota insieme al suo set di meta-dati - vedi sotto), e permettono all’utente di programmare azioni algoritmiche pilotate dal computer che intervengono



**Figure 2.** L’interazione con qualsiasi oggetto d’interfaccia della libreria *bach* può avvenire sia graficamente, sia attraverso messaggi Max. In questo esempio il contenuto dell’oggetto *bach.roll* viene eliminato, e successivamente vengono inseriti due accordi.



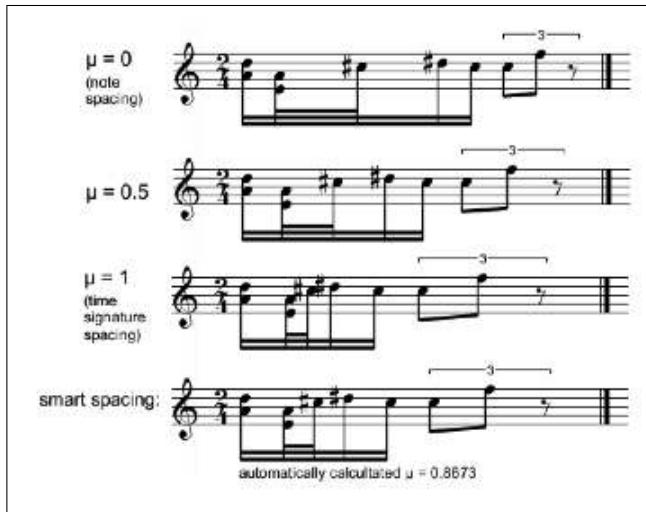
**Figure 3.** Le divisioni in semitoni, quarti e ottavi di tono sono rappresentate attraverso i simboli tradizionali di altezza (esempio in alto). Qualsiasi altra suddivisione microtonale viene invece rappresentata attraverso un riquadro contenente la frazione di tono (esempio centrale), o con la differenza in cent dalla nota diatonica (esempio in basso).

su virtualmente qualsiasi parametro dei dati musicali e della rappresentazione grafica. Inoltre, qualsiasi interazione attraverso messaggi Max provoca l’aggiornamento immediato della visualizzazione, e qualsiasi interazione con mouse e tastiera genera un messaggio di notifica: in questo modo, i due possibili modi di interazione sono completamente integrati l’uno con l’altro. Entrambe le categorie di interazione sono annullabili tramite un sistema di *undo*.

- Supporto per alterazioni microtonali di risoluzione arbitraria fino a un cent (vedi Fig. 3).
- Ampia possibilità di intervento sui parametri grafici della notazione musicale, p. es. la spaziatura (vedi Fig. 4), i font e i colori.
- Possibilità di associare a ogni nota vari tipi di meta-dati, tra cui testo, numeri, file, filtri e *breakpoint function* (vedi Fig. 9).
- Playback: *bach.score* e *bach.roll* sono sequencer a tutti gli effetti, e l’intero insieme dei dati (altezza, dinamica, durata) e dei meta-dati associati a ciascuna nota viene restituito al momento appropriato durante il playback: questo rende i due oggetti estremamente adatti al controllo di sintetizzatori e altre apparecchiature fisiche o virtuali. Entrambi gli oggetti inoltre comunicano con l’oggetto Max *setclock*, che permette di controllare in tempo reale la velocità di lettura.

### 3.3 Tipi di dati, rappresentazioni e operatori

*bach* implementa in Max due nuovi tipi di dati: i numeri razionali e una struttura ad albero chiamata *lill*, un acronimo per *Lisp-like linked list*. I numeri razionali sono estremamente importanti nel calcolo di dati musicali, dal momento che esprimono unità temporali come  $1/2$ ,  $3/8$  o  $1/12$  (cioè un ottavo di terzina) e rapporti armonici. La *lill* è una



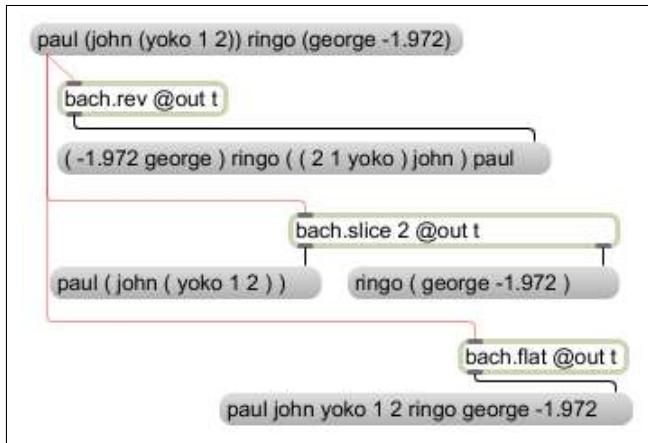
**Figure 4.** *bach.score* offre diversi algoritmi di spaziatura. Quello mostrato qui, al variare del parametro  $\mu$  tra 0 e 1, interpreta tra una spaziatura basata sulla durata delle note ( $\mu = 0$ ) e una basata sul metro ( $\mu = 1$ ). L'algoritmo di ‘spaziatura intelligente’ calcola automaticamente un valore ottimale per  $\mu$  (esempio in basso).



**Figure 5.** *bach.score* permette la coabitazione di diversi metri e tempi simultanei, e adatta l’allineamento grafico in maniera da rappresentare sempre la successione temporale degli eventi in maniera corretta.

collezione ordinata di elementi, ognuno dei quali può essere un numero (intero, in virgola mobile o razionale), un simbolo (cioè un elemento di testo, un’entità grosso modo corrispondente a una stringa di caratteri) o una *lill*: in questo modo si possono costruire strutture nidificate ricorsivamente, senza limiti esplicativi di profondità e complessità. La *lill* è stata scelta essenzialmente per tre ragioni: il fatto che qualsiasi messaggio Max può essere considerato come una *lill* non nidificata; la sua somiglianza alla lista di Lisp, che rende semplice la comunicazione con i principali sistemi esistenti di CAC; e la necessità di stabilire una struttura di dati che sia al tempo stesso abbastanza potente da rappresentare la complessità di una partitura musicale, ma abbastanza flessibile per essere un contenitore generico di dati che si presti a manipolazioni arbitrarie attraverso un insieme di primitive relativamente piccolo.

Di fatto, la grande maggioranza dei moduli della libreria



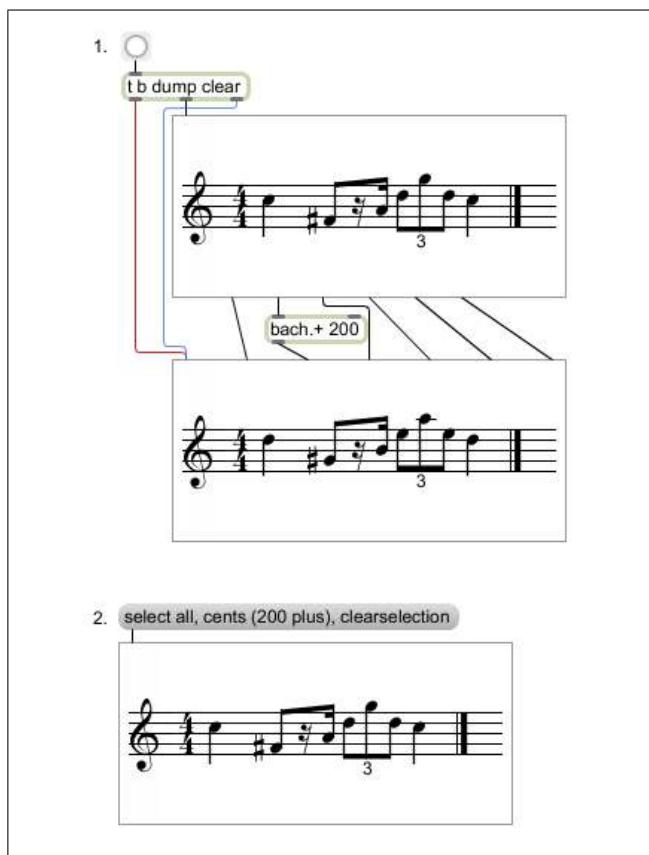
**Figure 6.** *bach* ha una grande quantità di oggetti capaci di eseguire operazioni di base sulle *lill*. In questa figura vediamo i risultati dell’inversione, la divisione in due parti e l’appiattimento di una *lill*. Inoltre, la più parte delle operazioni può essere limitata ad un particolare ambito di livelli di profondità: p. es., si possono rimuovere le parentesi di tutti i livelli intermedi, lasciando immutati il livello esterno e interno.

ria *bach* è costituita da strumenti per lavorare sulle *lill* svolgendo operazioni semplici come recupero di elementi, iterazione, inversione, ordinamento, divisione di una *lill* in più parti, unione di più *lill* in una sola, ecc. (vedi Fig. 6). Alcuni sottoinsiemi della libreria funzionano solamente su *lill* che soddisfino certe condizioni: ad esempio è possibile eseguire operazioni matematiche su *lill* composte esclusivamente da numeri; esiste un gruppo di operatori per il calcolo matriciale che funziona solo su *lill* strutturate in maniera appropriata; e così via.

All’intersezione tra i moduli per la notazione musicale e gli operatori su liste si trova una famiglia di oggetti che svolgono operazioni su *lill* che contengono dati musicali. Vale la pena notare che gli oggetti *bach* si scambiano tra loro partiture musicali in forma di *lill* strutturate in maniera specifica; i contenuti di queste *lill* sono integralmente leggibili e modificabili dall’utente. Questo è diverso da quanto avviene ad esempio in OpenMusic, dove lo scambio di dati musicali solitamente avviene attraverso oggetti opachi. Di conseguenza, in *bach* operazioni strettamente musicali come la quantizzazione ritmica sono solo operazioni estremamente specializzate su *lill*: queste operazioni ovviamente possono essere eseguite solo se la *lill* stessa è strutturata come una partitura musicale, e se il suo contenuto è consistente dal punto di vista della notazione musicale.

La struttura di una *lill* che rappresenta una partitura immagazzinata nell’oggetto *bach.score* può apparire molto complessa a prima vista, ma l’organizzazione del suo contenuto ambisce ad essere estremamente razionale: dopo una sezione di intestazione che contiene informazioni globali come le chiavi o i tipi di meta-dati che appaiono nella partitura, troviamo un sotto-albero ognuno dei cui rami corrisponde a una voce; il ramo di ogni voce contiene rami per ogni misura; il ramo di ogni misura contiene alcu-

ne informazioni specifiche alla misura (come il metro) e rami per ogni accordo; il ramo di ogni accordo contiene alcune informazioni specifiche all'accordo (come la durata) e rami per ogni nota; e il ramo di ogni nota contiene foglie per l'altezza e la dinamica, oltre ad eventuali altre informazioni come glissandi, enarmonia e meta-dati. La *lill* che rappresenta una partitura immagazzinata nell'oggetto *bach.roll* ha una struttura essenzialmente uguale, salvo il fatto che il livello delle misure non è presente. Con gli strumenti di operazioni su liste di *bach* è semplice recuperare e modificare informazioni specifiche relative a singoli elementi o sezioni della partitura. Inoltre *bach.score* e *bach.roll* offrono strumenti per accedere in maniera semplificata a specifiche categorie di valori (per esempio solo le altezze, o solo le dinamiche), in maniera da rendere più facile l'implementazione di operazioni musicali algoritmiche.



**Figure 7.** Due modi di eseguire una trasposizione in *bach*. Nel primo esempio, si ricostruisce una partitura che ha gli stessi parametri dell'originale, eccettuate le altezze (aggiungiamo un tono ad ogni altezza). Nel secondo esempio, la trasposizione è realizzata *in-place*: tutte le note vengono selezionate, quindi spostate verso l'alto e infine de-selezionate (non appena si clicca sulla *message box* grigia la partitura viene aggiornata).

### 3.4 Altre osservazioni

Si è scelto di non implementare un insieme predefinito di strumenti che eseguano trasformazioni musicali di base. Quindi, ad esempio, *bach* non contiene strumenti specifici

per la trasposizione o la retrogradazione: ma queste operazioni possono essere facilmente programmate, rispettivamente come un'addizione sull'altezza di ogni nota e l'inversione della lista che contiene gli accordi. Naturalmente questi sono casi estremamente semplici, ma la documentazione fornita con la libreria contiene informazioni approfondite ed esempi che mostrano come implementare diversi tipi di operazioni musicali: *bach* non dovrebbe essere generalmente più difficile da programmare di OpenMusic o PWGL, e lo scopo del progetto è fornire all'utente un insieme coerente e completo di strumenti con cui programmare virtualmente qualsiasi operazione desideri. Un progetto per il futuro è lo sviluppo di una libreria di strumenti ad alto livello che coprano un insieme generico di operazioni specificatamente musicali: tuttavia questa libreria non sarà una parte di *bach*, ma piuttosto un progetto separato basato su *bach*. Infine, un'API C sarà presto resa disponibile: in questo modo i programmati più avanzati avranno la possibilità di costruire nuovi oggetti esterni che accedano alle strutture di dati di *bach*, agli operatori di lista di basso livello e agli strumenti di notazione musicale.

Dovrebbe essere chiaro a questo punto che *bach* si situa all'intersezione di diverse categorie di software musicale, come accennato nell'introduzione di questo articolo. Da una parte, le sue funzionalità di rappresentazione grafica di partiture musicali appartengono tipicamente ai sistemi di videoscrittura musicale, benché si debba sottolineare il fatto che *bach* non offre alcune caratteristiche essenziali di questo tipo di programmi, come la possibilità di gestire la visualizzazione di pagine intere e, di conseguenza, di stampare partiture musicali. D'altra parte, molte delle caratteristiche di *bach* sono orientate a farne uno strumento per la CAC di potenza comparabile agli ambienti tradizionali basati su Lisp. Infine, *bach* può essere impiegato come il nucleo di un sequencer estremamente avanzato e flessibile, con la possibilità di pilotare virtualmente qualsiasi tipo di processo e di sistema di playback. E, naturalmente, *bach* si presta ad applicazioni innovative che sfruttino la convergenza di questi diversi paradigmi e del funzionamento in tempo reale.

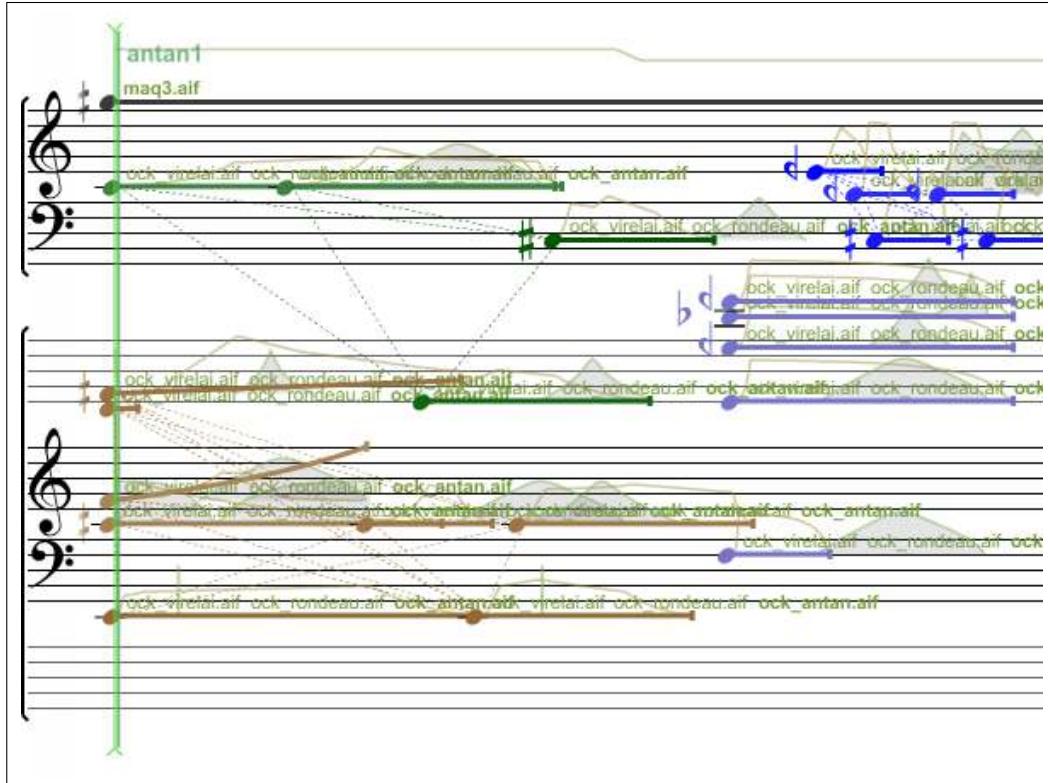
## 4. ESEMPI

Daremo ora qualche esempio di come la libreria *bach* possa essere utilizzata nella pratica.

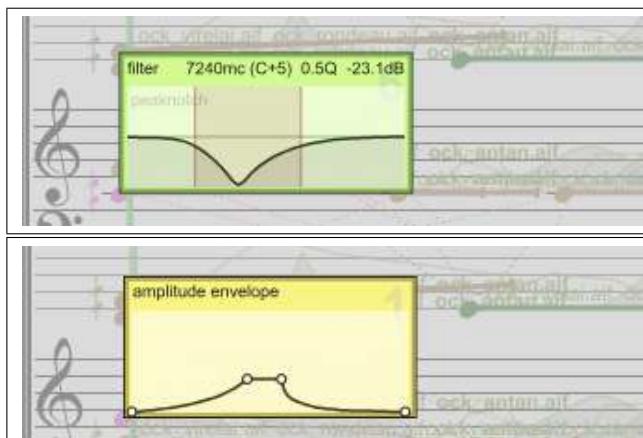
### 4.1 Un sequencer 2.0

Una partitura può essere considerata come un sequencer completamente personalizzabile, il cui contenuto è aperto a qualsiasi processo in tempo reale l'utente possa voler realizzare.

Alle note sono associate informazioni supplementari (i meta-dati descritti sopra), memorizzate in strutture chiamate *slot*, che specificano i parametri per i processi a cui sono sottoposte. Queste informazioni possono avere varie forme, tra cui funzioni d'inviluppo, dati numerici, curve di risposta di filtri audio, istruzioni testuali per un sintetizzatore, nomi di file, contenuti di matrici per il routing, *lill*, traiettorie di spazializzazione, colori. Non solo tutte que-



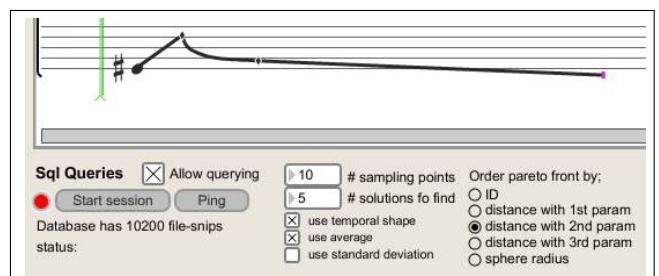
**Figure 8.** Un frammento della partitura elettronica della composizione *Chansons* di Daniele Ghisi, realizzata in un oggetto *bach.roll*. Le note sono raggruppate in strutture più ampie, rappresentate dai colori diversi.



**Figure 9.** Due esempi di finestre di *slot*.

ste informazioni sono accessibili attraverso messaggi, ma sono anche modificabili attraverso l’interfaccia: la pressione di un tasto specifico (e personalizzabile) fa apparire la finestra di uno slot e modifica le informazioni in esso contenute (vedi Fig. 9). La struttura di ogni slot (tipo di dato, nome, *hotkey*, dimensioni della finestra, dominio, codominio, curvatura...) è completamente definibile dall’utente; il contenuto di alcuni slot particolarmente importanti può essere reso sempre visibile nella partitura, non solo quando la finestra dello slot è aperta (per esempio, nella fig. fig:ex1a abbiamo sempre sott’occhio gli inviluppi di ampiezza in giallo, i nomi dei file audio associati alle note e le risposte in frequenza dei filtri).

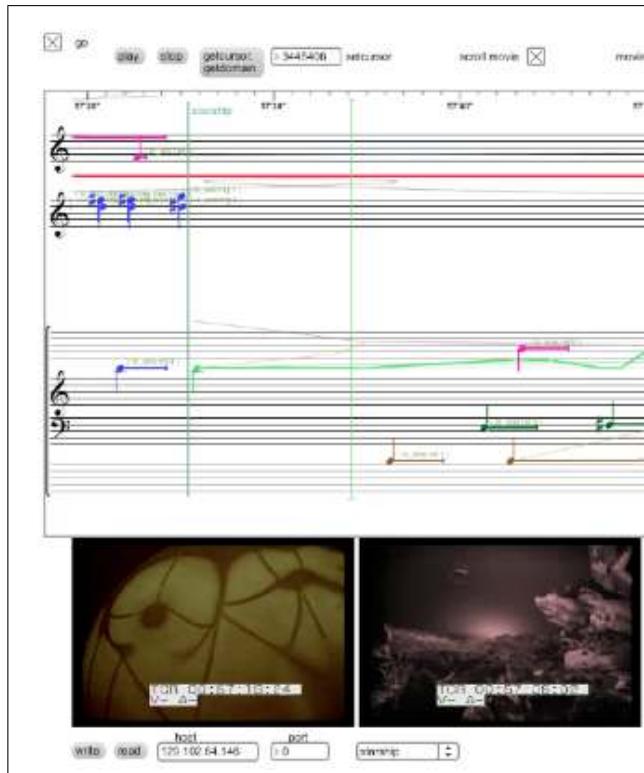
Rispetto ad altri approcci alla scrittura di musica elettronica, questa rappresentazione permette all’utente di servirsi della notazione tradizionale. Inoltre, la struttura degli slot è abbastanza flessibile da poter gestire un meccanismo di feedback simbolico. Per esempio, selezionando una nota e premendo un tasto può essere attivata una ricerca in un database di un profilo di altezza corrispondente; l’elemento trovato può poi essere re-immesso in un altro slot della stessa nota (vedi Fig. 10).



**Figure 10.** Porzione di un pannello per la composizione *abroad* di Daniele Ghisi: viene effettuata la ricerca di un profilo di altezza corrispondente in un database di file audio. Il motore di ricerca è stato sviluppato da Philippe Esling all’IRCAM.

Allo stesso tempo, grazie alla possibilità di recuperare in tempo reale tutte le informazioni collegate ai dettagli della visualizzazione grafica della partitura, è semplice mantenere una sequenza video costantemente allineata alla partitura stessa. In questo modo è sempre possibile co-

noscere l'esatto fotogramma a cui ogni evento musicale è sincronizzato (vedi Fig. 11).



**Figure 11.** Una parte della patch utilizzata da Andrea Agostini per gestire la partitura elettronica per il film *Wunder der Schöpfung*. Ogni nota contiene un insieme di istruzioni per un sintetizzatore, espresse come comandi di testo e indicazioni grafiche. Quando la partitura viene eseguita, tutte le informazioni associate ad ogni nota sono inviate alla voce corrispondente del sintetizzatore, rappresentata dal colore della nota. Sotto la partitura vengono mostrati alcuni fotogrammi del film, in maniera da offrire un riferimento visivo alla posizione degli eventi musicali. In una finestra separata, non mostrata qui, il fotogramma corrispondente alla posizione esatta del cursore di lettura (la linea verticale sottile verde) viene mostrato in tempo reale, permettendo un controllo preciso della sincronizzazione tra suono e immagine.

#### 4.2 Un motore di granulazione simbolica in tempo reale

Si può immaginare una grande categoria di processi di CAC che possono essere eseguiti in tempo reale. Ad esempio, è semplice costruire con *bach* una macchina che esegue una granulazione in tempo reale di dati simbolici (vedi Fig. 12). I parametri della granulazione possono rispecchiare quelli della granulazione elettroacustica classica: ad esempio dimensione del grano, posizione di inizio, intervallo dei grani o densità, velocità dei grani... Quando la macchina viene accesa, alcuni frammenti di una partitura vengono estratti da un originale (ognuno di questi frammenti è considerato un grano), e immediatamente integrato in un oggetto *bach.roll* che contiene il risultato complessivo del-

la granulazione. Si può implementare un sistema coerente di playback, in maniera da poter ascoltare in tempo reale un risultato simulato dell'operazione. Naturalmente tutti i parametri possono essere variati in tempo reale, eventualmente associandoli a qualsiasi tipo di flusso di dati in ingresso.

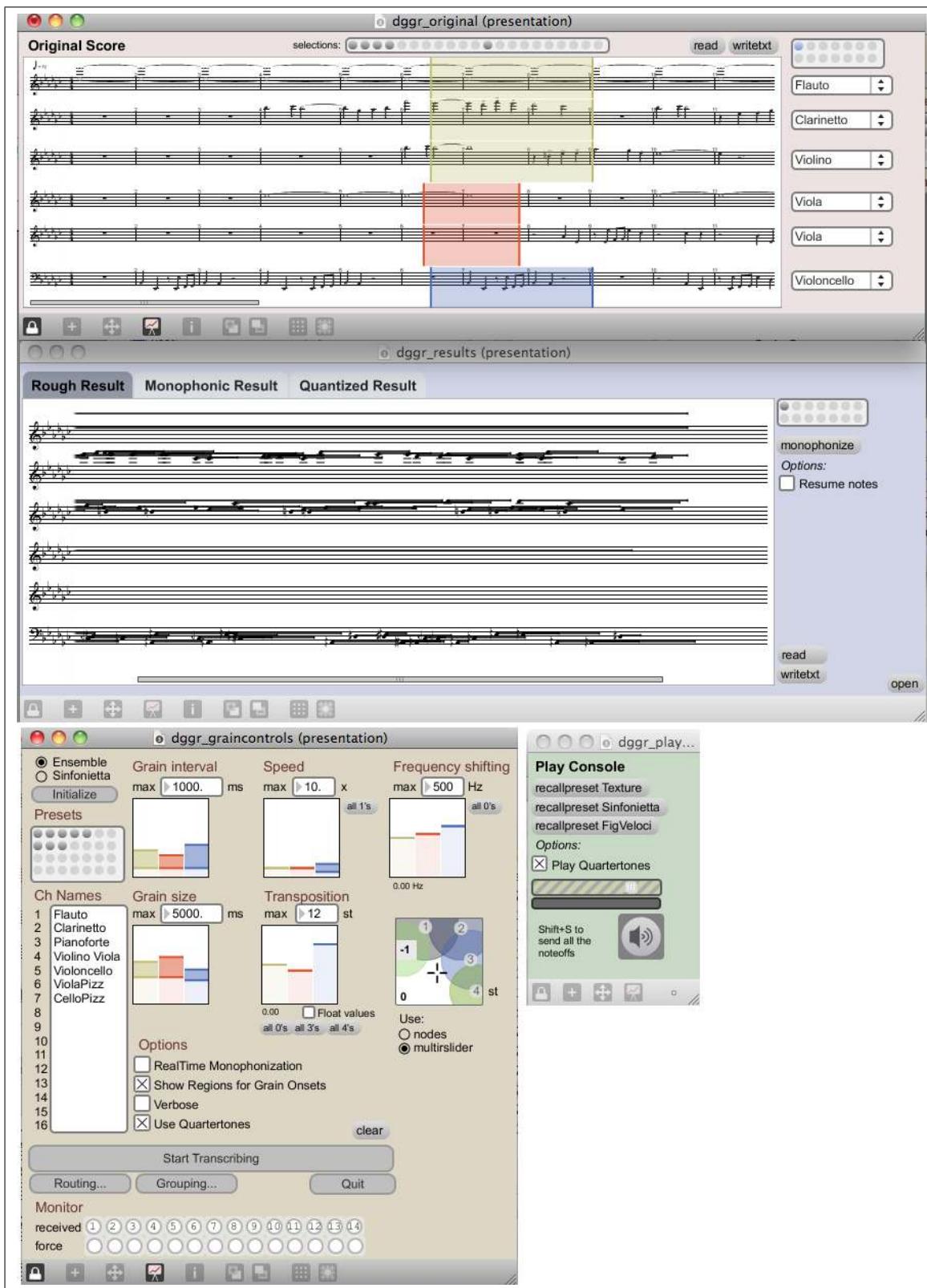
#### 5. PROSPETTIVE FUTURE

Nel momento in cui quest'articolo viene scritto, lo sviluppo di *bach* è nella fase *alpha*: benché il sistema sia utilizzabile, non tutte le caratteristiche previste sono già state implementate. Tra le funzionalità addizionali previste figurano:

- Supporto per la rappresentazione ritmica ad albero, che permetterà ad esempio la rappresentazione di gruppi irregolari (al momento, una terzina contenente una quintina viene rappresentata come una quindicesimina ‘piatta’).
- Implementazione di strutture gerarchiche nella partitura, in maniera da permettere all’utente di raggruppare i suoi elementi per nome (per elemento si intende un accordo, una nota o un altro gruppo).
- Supporto per l’importazione e l’esportazione di file MIDI, MusicXML e SDIF.
- Un risolutore di CSP (problemi di soddisfazione di constraints).

#### 6. REFERENCES

- [1] G. Assayag *et al.*, “Computer assisted composition at ircam: From patchwork to openmusic,” *Computer Music Journal*, no. 23 (3), pp. 59–72, 1999.
- [2] M. Laurson and M. Kuuskankare, “Pwgl: A novel visual language based on common lisp, clos and opengl,” in *Proceedings of International Computer Music Conference*, (Gothenburg, Sweden), pp. 142–145, 2002.
- [3] M. Puckette, “A divide between ‘compositional’ and ‘performative’ aspects of Pd,” in *Proceedings of the First Internation Pd Convention*, (Graz, Austria), 2004.
- [4] A. Cont, *Modeling Musical Anticipation*. PhD thesis, University of Paris 6 and University of California in San Diego, 2008.
- [5] M. Laurson and J. Duthen, “Patchwork, a graphical language in preform,” in *Proceedings of the International Computer Music Conference*, (Miami, USA), pp. 172–175, 1989.
- [6] C. Selegborg, “Interaction temps-réel/temps différé,” tech. rep., Université Aix-Marseille II, 2004.



**Figure 12.** Screenshot di una patch per la granulazione simbolica in real-time. La partitura originale (la finestra in alto, color violetto) contiene alcuni marker che determinano le regioni da elaborare. I parametri sono specificati nella finestra color ocra in basso. Quando l'utente clicca sul bottone 'Start transcribing' il risultato appare e viene raccolto nella finestra blu al centro. Quando l'utente lo desidera, può renderlo monofonico (se necessario), raffinarlo e infine quantizzarlo. Tutti i parametri sono modificabili dall'utente e influenzano in tempo reale il risultato, esattamente come in una macchina per la granulazione audio.

# REVEALING THE HIDDEN MOTIVES OF CHROMATICISM IN MUSIC THROUGH COMPUTATIONAL SEMANTIC VISUALIZATIONS

Dimitrios Margounakis

Dept. of Informatics,  
School of Sciences

Aristotle University of Thessaloniki  
GR-541 24, GREECE  
dmargoun@csd.auth.gr

Dionyssios Politis

Dept. of Informatics,  
School of Sciences

Aristotle University of Thessaloniki  
GR-541 24, GREECE  
dpolitis@csd.auth.gr

## ABSTRACT

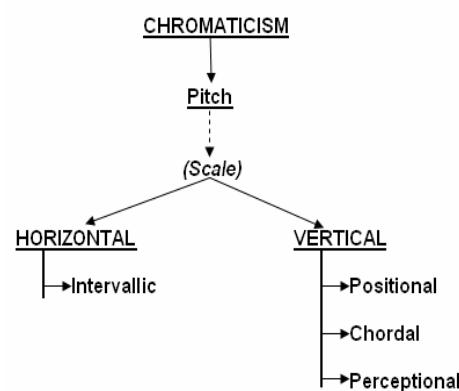
State-of-the-art ways for visualizing high-level music concepts are currently a trend in music computing, aiming at bridging the semantic gap between music information retrieval and human knowledge. This work presents the analysis of Brahms's String Quartet Op.51 No.1 concerning the concept of chromaticism by using the chromatic index (a chroma measure especially designed for calculating chromatic elements within a piece of music with useful diagnostic value) as a metric. The visualization of the chromatic analysis with actual colours exhibits chromatic variations to musicologists and can even be used as a digital signature for music. The suggested computational approach in MIDI and Audio Files reveals specific motives of the quartet that provide interesting chromatic phenomena. A comparative study between the results of analysis on score-based data and the performance data (from an indifferent live performance) brings the diversity of chromatic perception the same music can cause to light. Moreover, Allen Forte's proposed motives in his study of Brahms's Op.51 No.1 are examined for chromatic elements. Each one of the four voices is separately analyzed so that its specific role in the piece's progression will become evident. Additive techniques are used to calculate the overall chromatic indices of the quartet.

## 1. INTRODUCTION

Mathematical modeling of Chromaticism in music can be used for analysis of sound, resulting in representations that can be used in many tasks: indexing, classification and retrieval. Chromatic index is a chroma measure especially designed from the authors for calculating chromatic elements within a piece of music with useful diagnostic value.

The study of several aspects of chromaticism in mu-

sic resulted in the modeling of a theoretical framework that calculates the factors of music chromaticism by the authors and is shown in Figure 1 [12].



**Figure 1.** The factors that affect music chromaticism in a framework.

The chromatic analysis relates the chromatic index  $\chi$ , to the perceptual progress of melodic lines in music [14]. Chroma can be illustrated with colors. The final visualization is semantically important and allows useful conclusions to be observed. Different visualizations may result from different data of the same music work (score-based data or performance data) and their comparison is also interesting. This work in progress demonstrates horizontal (intervallic) chromatic analysis by applying the notion of Chromaticism in music to Brahms's quartet Opus 51, No 1, 1st movement. Vertical relations are presented here indirectly by combining the horizontal results in a vertical manner.

The object of this research consists of one remarkable piece of music work of Brahms's, which has been previously analyzed by different scholars. In 1983, Allen Forte presents a motivic design of Brahms's quartet emphasizing on his motive 'alpha' [4]. Later in 1990, Lewin discovers one Beethoven-like and one Mozart-like subject, both put together in the quartet [9]. In 2001, Huron writes a critical essay on Forte's work, which proves that motive 'alpha' is unable to distinguish quartet No. 1 from the other Brahms's quartets [6]. Moreover, other computational analysis approaches

Copyright: © 2012 Margounakis 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.

have been applied to the quartet using a variety of computational implementation techniques [3] [8].

This work presents the analysis of Brahms's String Quartet Op.51 No.1 concerning *horizontal chromaticism* by using the chromatic index as a metric. The suggested computational approach in MIDI and Audio Files reveals specific motives of the quartet that provide interesting chromatic phenomena. A comparative study between the results of analysis on score-based data and the performance data (from an indifferent live performance) brings the diversity of chromatic perception the same music can cause to light. Moreover, Allen Forte's proposed motives in his study of Brahms's Op.51 No.1 are examined for chromatic elements. Each one of the four voices is separately analyzed so that its specific role in the piece's progression will become evident. Additive techniques are used to calculate the overall chromatic indices of the quartet.

## 2. RELATED WORK

*Chromaticism* in music is the use of notes foreign to the mode or diatonic scale upon which a composition is based. Chromaticism is applied in order to intensify or colour the melodic line or harmonic texture [1] [7]. *Chromatic elements* are considered to be "elaborations of or substitutions for diatonic scale members" [2]. According to Meyer: "Chromaticism is almost by definition an alteration of, an interpolation in or deviation from this basic diatonic organization"<sup>1</sup> [2].

The definition found in The New Grove Dictionary of Music and Musicians provides a standard characterization: "In melodic and harmonic analysis the term 'chromatic' is generally applied to notes marked with accidentals foreign to the scale of the key in which the passage is written".

An extended study on chromaticism of intervallic nature has been presented by the authors in the past. The interested reader is prompted to refer to [11], [12], [13] and [14] for a comprehensive description. Previous work is not repeated here extensively due to space constraints. The key-principles of chromatic analysis follow:

Each scale (or mode) bears a *scale index (SI)*, which determines the inherent chromaticism of the scale according to its intervallic structure. Therefore, the key of a musical piece is the first benchmark for calculating its chromatic nature (Figure 1) [12].

Intervallic chromaticism is expressed by a running variable  $\chi_i$ . In general, the value of  $\chi$  is greater at times when music produces greater chromatic phenomena of intervallic nature. Chromatic pitches and microtonal deviations affect  $\chi_i$  by multiplying certain constant factors to its value.  $\chi_i$  is equal to SI at the start point of mu-

sic.  $\chi_i$  is defined by the intervallic relations between the notes in the melody, and the duration of those intervals.

The approach of computational chromatic analysis consists of five stages: (a) extraction of melodic salient pitches (for audio files) or notation for MIDI files, (b) matching a scale to the music piece, (c) segmentation, (d) calculation of the chromatic elements, and (e) visualization of Chromaticism using colorful strips. The fundamental concepts and algorithms of chromatic analysis are presented in [11] and [12]. Also several discussions can be found in [10], [14] and [15]. The analysis has been conducted with the use of our tool MEL-IRIS client V.3.0. The application's core function is the chromatic analysis of music, as described here. MEL-IRIS was mainly developed in Borland C++ Builder 6 and uses the MS-SQL database (Figure 2). Further manual mathematic processing has taken place in some cases, using the raw data of MEL-IRIS database after the analysis of the quartet. The strategy of our approach is to construct a semantic chromatic graph and identify significant chromatic segments and repetitive patterns during the piece.



**Figure 2.** Snapshots from the tool MEL-IRIS v.2 (top) and MEL-IRIS v.3 (bottom).

The goal of chromatic analysis is to reduce the gap between notational aspects and our real forms of knowledge by giving a visual representation of the evolution of chromaticism in a tonal work over time. The system is able to detect fluctuations of chromatic gradations also with reference to the same sections of the score, played a second time. This circumstance occurs because an already played music has modulated the  $\chi$ -value and this seems to optimally reflect what really happens in

<sup>1</sup> These definitions accept the major and minor scales as the diatonic scales, although this is not generally acceptable by all authors'

our perceptual activity: we are affected by learned factors. It follows that this analysis is able to bring to light interesting analytical factors which would not be detectable with a simple study of the score.

### 3. CHROMATIC PROGRESSIONS IN FORTE'S MOTIVES

In this section, we will examine whether the proposed main motives of Forte are of chromatic importance within the quartet and discuss our finding taking into account Forte's analysis [4]. Any transition of  $\chi$  greater than zero implies a chromatic phenomenon in the motive. The metrics used for calculating the transitions have been discussed in [13]. Considering that the scale of the quartet is C minor (calculation of MEL-IRIS returned Minor Harmonic as the dominant scale) the  $\chi$  transitions in Forte's motives are:

$a$	$\bar{a}$	$a'$	$\bar{a}'$	$\phi$	$\bar{\phi}$	$\phi'$	$\bar{\phi}'$	$\chi$	$\bar{\chi}$
-	-	-	-	*1.0249	*1.03	*1.0249	*1.03	%1.005	%1.005
$\beta$	$\bar{\beta}$	$\beta'$	$\bar{\beta}'$	$\gamma$	$\bar{\gamma}$	$\omega$	$\bar{\omega}$	*1.01	-
$\theta$	$\bar{\theta}$	$\varepsilon$	$\bar{\varepsilon}$	$\delta$	$\bar{\delta}$	$\delta'$	$\bar{\delta}'$	$\sigma$	$\bar{\sigma}$

**Table 1.** Importance of Forte's motives in Brahms's Quartet.

Moreover, the idiomatic pattern in meter 39, which Forte [4] includes in his table of motives (see Figure 3) adds chroma to the piece (\*1.0198), since it contains 2 tones that decrease  $\chi$  by 1.005, but also a chromatic 3/2-tone (Gb-Eb) that increases  $\chi$  by 1.03.



**Figure 3.** A chromatic pattern in Brahms's quartet (m.39).

A first observation on the  $\chi$  values of Table 1 is that motives alpha, beta, theta, omega and all their transformations cause no chroma variance. This is quite predictable for motive alpha, since it is the most prominent motive in the movement. Motive alpha is presented as the first three notes in the melodic first subject and defines the tonic lead of the movement. Its form bears no deviation from the underlying scale. In fact, it consists of the first three notes of the scale; therefore no further chromatic impression is created. The same stands for motive beta, which also appears for the first time in meter 1 just after alpha. Motives alpha and beta essentially set the tone for the rest of the piece. Motive theta is the octave motive, which deserves special attention according to Forte, since it occupies an entire measure alone.

In terms of chroma, its role is to bring back the so far created  $\chi$  value to the initial  $SI$ . Its duration allows only for chroma restoration, since motive theta is used to end music phrases in the context. Of course, such a motive could never add chroma to a music composition. Finally, motive omega is reserved as a closure and reflects the basic melodic background in the Shakerian sense [4]. As in the case of motive theta, no more chroma is added from omega either.

Motives chi and epsilon reduce the chroma value. The same stands for motive delta and its inversion (delta retrograde and its inversion present no chromatic variance – especially for these transforms, Forte states that “*delta bar and delta bar prime...are not represented in any significant way in the music*”). These motives (especially epsilon) are associated with minor to major transformations. Such transformations are translated as reduction of  $\chi$  in this work.

It becomes clear that the main chromatic motives of the movement are: phi, gamma, lambda (not lambda bar) and sigma. Sigma and sigma bar resulted in the greatest chroma deviations. According to Forte: “*gamma plays such a crucial role..*” and “*..gamma foreshadowed, in a very specific way, the large-scale tonal reorientation..*”. Indeed, the single interval of motive gamma adds chroma and “*..is reserved for special moments in the music..*”. Motive lambda implies greater  $\chi$  deviation in music than gamma, and Forte recognizes its chromatic motivic motion. He also describes an example of lambda as an event of very special significance in music. However, the most striking phenomena are produced by motives sigma. According to Forte, this musical emblem of the first movement explains certain of the otherwise enigmatic tonalities. At the end of his essay, he states that “*not only sigma as an entity but also its pitch-class components have striking roles in the music*”. Apparently, these striking phenomena as isolated motivic patterns are well reflected by the chroma measure  $\chi$ . In the next sections, the movement will be analyzed chromatically as a whole. This analysis will reveal the chromatic effect of these patterns in the movement.

### 4. CHROMATIC ANALYSIS OF THE FOUR VOICES

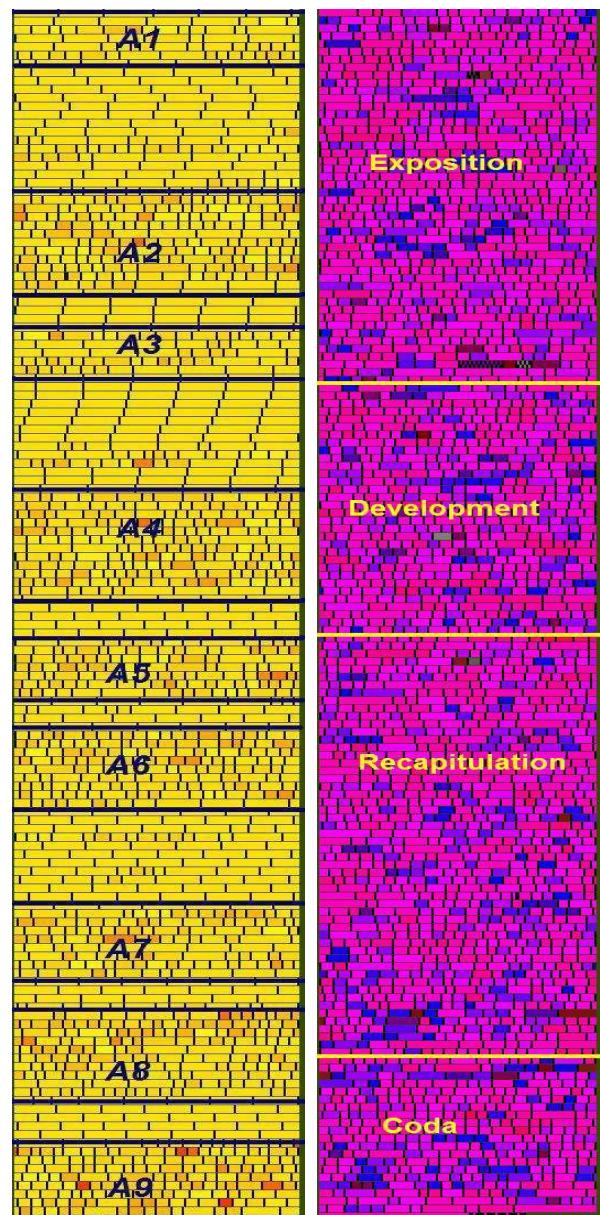
Chromatic analysis is applied on melody or the salient pitches of a musical piece. The structure of the quartet provides us 4 voices, which allow us to analyze them separately. The reason for the separation is that we can gain better results from each voice analysis than trying to extract the predominant voice. We extracted the four different parts from a FINALE sheet of the quartet. The parts were saved as four separate MIDI files. These files have been fed into MELIRIS and the chromatic diagrams have been created. Graphs of the progression of  $\chi$  during time are presented next.

Beginning with the first violin, we can see different parts of chromatic perception in the graph (Fig. 4a): chromatic-varied parts (frequent transitions of yellow) and parts that do not add any chroma to the music (purely yellow bricks). It should be also noted that yellow represents the depiction of the voice's SI. Any colour deviation from yellow depicts transitions of  $\chi$ . This leads to an extra semi-automatic segmentation of the voice in larger clusters of bricks. We excluded clusters of bricks with  $\chi$  deviations close to zero (these zero-clusters present no chromatic phenomena of importance). Using the raw data extracted from MELIRIS, we detected clusters of continuous zero  $\chi$  deviations and then added intermediate areas between them with very smooth  $\chi$  transitions. We set  $S=0.002$  as a threshold for entering or not the zero zone, which means that a  $\chi$  deviation average of a cluster greater than  $S$  bears chromatic elements of importance. After defining the zero clusters, we manually set the exact starting and ending points of them. The first yellow brick of the cluster signalized the starting point, while the last yellow brick of the cluster signalized the ending point. This way, 9 segments of importance ( $S>0.002$ ) were discovered (A1,A2,...,A9). Also, the intermediate parts represent 8 zero-clusters (Z1,Z2,...,Z8). Their exact timings and average deviations can be seen in Table 2. Moreover, these parts have also been marked on the graph of Figure 5, which represents the progression of chroma  $\chi$  during time. To all intents and purposes, the chromatic wall in Figure 4a comprises a graphic visualization of Figure 5.

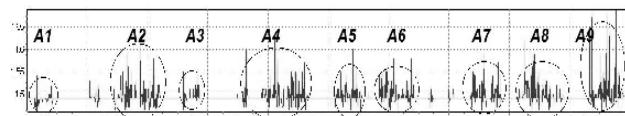
We concentrate on the 9 chromatically important segments, which have been marked in the graph. These contain the bars: A1(1-15), A2 (53-79), A3 (11-15)<sup>2</sup> [repetition], A4 (53-86) [repetition], A5 (103-114), A6 (127-146), A7 (179-197), A8 (207-230) and A9 (243-263). The aforementioned patterns are also denoted in Figure 5.

We can observe that patterns A1 and A2 are recognized again in the execution of repetition as patterns A3 and A4 with minor chromatic changes. The question that arises here is: why doesn't the same execution result in exactly the same  $\chi$  transitions? The answer is that the second repetition comes when an already played music has already modulated a perceived  $\chi$  progression (index  $\chi$  is continuously changing depending on its previous value). This continuous retrospective nature of  $\chi$  reflects the way of progressive human perception of chromaticism, which is a function of learned factors acting in conjunction with innate factors. Of course, the differences are not huge. This can be seen in Figure 5 where we can see that A3 is slightly more chromatic than A1, while A4 is slightly more chromatic than A2.

<sup>2</sup> Segments A3 and A4 actually represent the repetition of segments A1 and A2, since at the end of the exposition (in the score) there is the repetition sign :|| in bar 82.



**Figure 4.** (a)The chromatic graph of the 1st Violin melodic progression.(b) The chromatic graph of the audio recording.



**Figure 5.**  $\chi$  transition during the melodic progression of the 1st violin.

Pattern A5 is created because of the existence of Forte's sigma motive [5]. Measures 107 and 109 in the development provide the chromatic motion. As we discussed in the previous section, sigma motives are the most chromatic of all, because of the continuous progression of the voice in chromatic 3/2 tones. The rests between the chromatic intervals decrease the created

chromatic impression rapidly, and then chroma increases again. This technique prepares for the unexpected after the incident: change of scale and tonality.

<b>Segm.</b>	<b>Avg <math>\chi</math> dev</b>	<b>Start time</b>	<b>End time</b>
A1	0.002004	0	20897
A2	0.008532	71399	110292
A3	0.008203	129610	143960
A4	0.009473	174726	233355
A5	0.012864	254833	278052
A6	0.012135	292378	321008
A7	0.012435	367447	393569
A8	0.014608	405759	445232
A9	0.018981	464388	496431
<b>Zero segm.</b>	<b>Avg <math>\chi</math> dev</b>	<b>Start time</b>	<b>End time</b>
Z1	0.001519	20897	71399
Z2	0.000025	110292	129610
Z3	0.000137	143960	174726
Z4	0.000179	233355	254833
Z5	0.001719	278052	292378
Z6	0.001096	321008	367447
Z7	0.000295	393569	405759
Z8	0.000156	445232	464388

**Table 2.** Segmentation of the 1st Voice in segments of chromatic importance and zero-segments (timings are in milliseconds)

The first 10 meters in A6 achieve exactly the same thing: they act as preparative for a transition to the initial theme (and the initial scale). Although they are included in the analysis together with the initial theme, this pattern is recognized as quite similar to A1, as it will be discussed later.

A7, A8 and A9 are the three final patterns that have been extracted from the first violin score. All of them bear several chromatic elements and tonality changes. Especially, Coda in A9 contains a lot of chromatic intervals (eg. 2 ½ tones in m.250, 254, semitones in 251, chromatic chords etc.) and produces high levels of chroma.

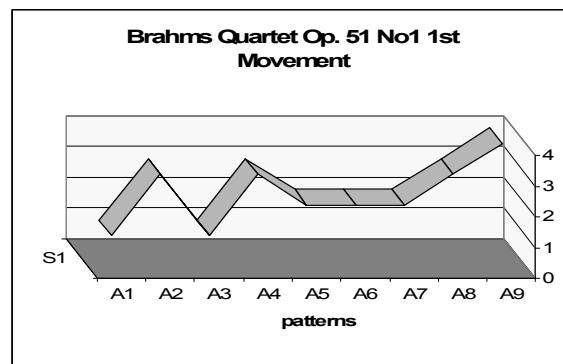
Two tests have been conducted in the analysis of the 1st violin in Brahms's quartet. The first test tried to capture repetition of the chromatic patterns in the context. We tried all the 9 patterns as input queries in MEL-IRIS MIR engine. The retrieval method for similar patterns in MEL-IRIS is based on Euclidean distance as a metric, regarding the color distribution of the analyzed musical pieces. No special similarities were found except for patterns A1, A3 and A6. Patterns A1 and A3 are similar in a percentage of 89,68% (the reasons were mentioned before), while patterns A1 and A6 are similar in a percentage of 67,63%. This similarity is merely affected by the repetition of bars 1-9 in A6. If the first ten meters of A6 were excluded from the pattern, then the similarity would be greater. But this is not the goal, since we extracted chromatic patterns and not discrete segments of

the score. A2, A4 and A8 also showed some chromatic similarity: A2-A4 (66,54%) and A2-A8 (51,12%).

The second test aimed at grouping the distinctive patterns together by means of minimum and maximum deviations from the original SI (namely maximum deviations since important minimum deviations are not observed). The groups can be seen in Table 3. A progressive chromatic movement is observed. Little chroma is added at the beginning, which is followed by an increase. Then, this schema is repeated. Afterwards, chroma is kept at the same level (neither low nor too high) to finally reach the highest level in coda. This interesting motive can be seen in Figure 6. In this graph we have defined 4 levels. Each level corresponds to one deviation group. As we can see, the Exposition lies in levels 1 and 3, the Development in level 2, while the Recapitulation begins from level 2 and passes through level 3 to reach the highest level 4. Coda moves to the most chromatic level 4.

Group	Level	Patterns
Small deviation (-0.03, +0.05)	1	A1, A3
Medium Deviation (-0.04, +0.1)	2	A5, A6, A7
Great Deviation (-0.05, +0.13)	3	A2, A4, A8
Greatest Deviaiton (-0.06, +0.19)	4	A9

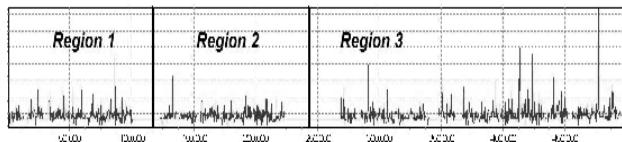
**Table 3.** Discrete chromatic patterns in the 1st voice



**Figure 6..** Graphical Representation of the Chromatic Progression in the 1st voice.

Next, the chromatic graph of the second violin is presented. As we can see, it is not possible to discriminate segments as in the case of the first violin, but 3 great regions can be seen. The boundaries have been set manually in Figure 7 within the two greatest zero-zones ([99843 ms – 122482 ms] and [222906 ms – 268764 ms]). It is obvious from the chromatic analysis that Region 1 depicts the Exposition, Region 2 the repetition of Exposition and Region 3 the rest of the movement. If we consider a splitting point around 340000 ms (which is not so obvious), then we have two sub regions in Region 3: Development and Recapitulation. Again, we can observe in this voice that greater chroma appears at the end of the movement. Especially in Coda, we can see

the greatest deviation (0,6). Also, two great peaks (deviation: 0,4) are observed between measures 200 and 220 (corresponding to pattern A8 of the 1st violin), which qualify Coda.



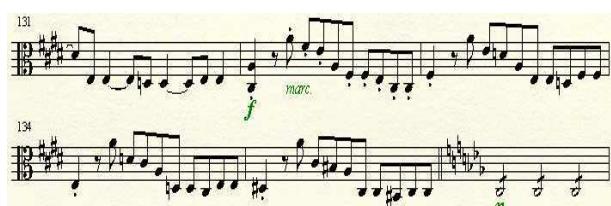
**Figure 7.**  $\chi$  transition during the melodic progression of the 2nd violin.

In contrast to the violins, the Viola and the Cello do not contain further obvious segments, so no further manual segmentation has been attempted. Their chromatic movement fluctuates in a continuous manner creating the chromatic background for violins to become salient. There are, however, some interesting peaks, where we can examine their chromatic origin.

An interesting chromatic motive V1 is observed in measures 33-35 of Viola (around 50,000 ms). The red line in Figure 8 depicts an absolutely chromatic movement D-Eb-Gb-A-Bb (semitone – 1 ½ tone – 1 ½ tone – semitone), while the blue lines show extra Chromaticism addition with the surprisingly created intervals of 2 ½ and 5 ½ tone. This great technique creates a moment of strong chromatic impression. But this is only a moment, since it is assured that all this extra Chromaticism will be disappeared in the next 2 measures with the repetitions of Bb and the interval of the octave for 24 times (a certain non-chromatic motive)! At the graph of the viola,  $\chi$  returns immediately to its fundamental value after the peak.



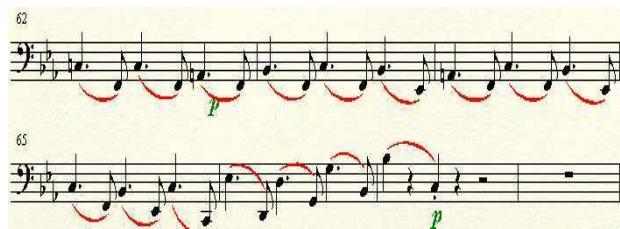
**Figure 8.** The chromatic motive V1 in the Viola (m. 33-35).



**Figure 9.** The chromatic motive V4 in the Viola (m. 132-136).

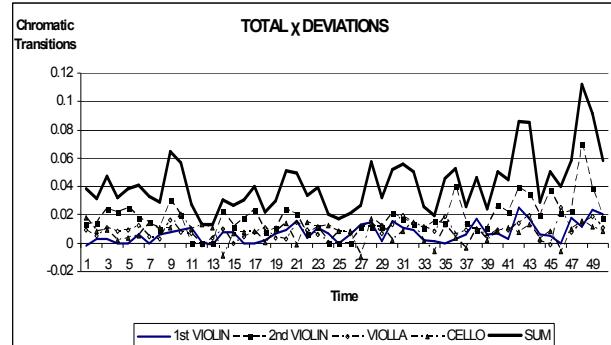
Figures 9 and 10 show two other interesting chromatic patterns: the motive V4 of Viola (132-136) at the end of the development and the beginning of the recapitulation ([312161 ms – 316968 ms]) and the motive C3 (63-66) of Cello with the interesting rhythmic mis-

quoted intervals ([105880 ms – 113113 ms]). V4 creates Chromaticism by changing rapidly tonalities, scales and by sharpening and flattening notes.



**Figure 10.** The chromatic motive C3 in the 4th Voice (m. 63-66).

After the separate analysis of the four voices, a cumulative result will be presented. By summing up the chromatic deviations during the time axis, we can get a cleaner image of total chroma through the music context. For readability purposes we split the quartet in segments of 10,000 ms (for a total of 50 segments) and we sum the average deviations from the base chroma for all voices in each segment. This way we can gain a view of the harmonic relations of the melodic lines for our MIDI file. Figure 11 shows the result of summing up all voices. The parts of the quartet have approximately the following boundaries: Exposition [segments 1-23] (repetition included), Development [segments 24-30], Recapitulation [segments 31-50] and Coda [segments 44-50].



**Figure 11.** Adding the chromatic effect of all voices in the 1st Movement.

It is obvious that the 2nd Violin creates the most chromatic situations during the piece and has the greatest effect on the overall chroma deviation. On the contrary, the Cello seems to have the most stable chroma values. Some of the cello's diatonic intervallic patterns create even negative deviations at times (segments 14, 27, 34, 38 and 46). Viola has the most natural movement within the boundaries of 0.02 with little Chromaticism adding on it. Finally the main instrument (1st Violin) leads the melodic progression to occasional peaks and occasional chromatic silence.

As it can be seen in the graph, the total harmonic effect of all the voices together can be compared with the pattern progression of the 1st Violin of Figure 6. A

closer look reveals that the first and the third part of the quartet (Exposition and Recapitulation) are close by means of Chromaticism. Differences and several chromatic variations are observed during the Development. Consequently, we can observe that in the beginning and the ending of the 1st movement of the quartet, all the instruments are designed to follow a standard chromatic perception. The participation of all voices in these parts is very important since they make the desired perception stronger to the audience. However, during the second part of the movement the instruments do not follow the same chromatic pattern, resulting in a very interesting movement with a lot of fluctuation. These successive alterations are achieved in the movement in two ways: the use of the 3 voices (mostly the Cello) against the 1st violin in some cases, or the strengthening of the chroma value of the 2nd Violin in other cases. These alterations result in a variety of feelings during the Development. The 1st movement's finale (Recapitulation and Coda) shows a gradual increase of chroma. The greatest  $\chi$  deviations appear at the end of the movement, which means that the strongest chromatic elements are then perceived.

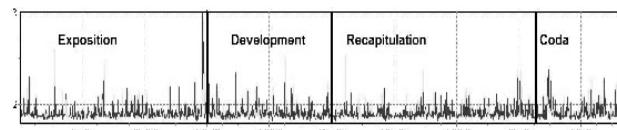
It should be noted here that the indirect presentation of 'vertical' vision cannot fully reflect the harmonic chromatic perspective of music. Moreover, the voices should not be treated in the same statistical level and more complex calculations should be modeled in order to be accurate. As it is discussed in [12], in order to calculate vertical chromaticism, four coexistent auditory relations should be examined (Figure 1). Three of them are actually written in the music notation (two of them are referred to as *positional chromaticism*, while the third one is referred to as *chordal chromaticism*). The fourth one (perceptual chromaticism) is subconscious and is perceived by the listener's music cognition. Although we are working on the modeling of these calculations, our tool is not ready yet to provide consistent results on vertical chromaticism.

## 5. CHROMATIC ANALYSIS OF THE AUDIO RECORDING

After the analysis of the four voices from the MIDI file, it would be interesting to analyze an audio recording of the quartet and compare the differences of the chromatic results. The MP3 file we supplied to MEL-IRIS, is a recording of the 1st movement of the quartet by Delirium Terens Quartet (bit rate: 128 kbps). The results of the audio analysis (the chromatic wall and the graph of  $\chi$  movement) appear in Figures 4b and 12. The segmentation of the audio recording in chromatic bricks resulted from automatic clustering.

The first observation here is that, during the whole time of the piece, there are more and much greater chromatic deviations (even deviations greater than 1!) than any of the previous graphs concerning MIDI files.

This is because of the nature of salient pitches that are detected by the system. MIDI files follow the rule of the 'well-tempered clavier' while reproducing exactly the music score. As a result, all the intervals, created in MIDI executions, are integer multipliers of the semitone. This creates limited chromatic perception compared to a live recording, where pitches from the whole frequency spectrum are indicated. This may even lead to 'odd' intervals, e.g. 35 cents or 118 cents, which encompasses added chroma. Therefore, proper playing<sup>3</sup> by the instrument players can create their intended chromatic perception to the audience. It should be pointed out here that strings (in our case the violins, the viola, and the cello) are able to reproduce any pitch, in contrast to the well-tempered piano. However, also a certain amount of 'noise' in calculating chroma arises from overtones from each instrument in the signal; these are quite prominent for bowed strings.



**Figure 12.**  $\chi$  transition of the melodic salient pitches of a live performance (MP3 format).

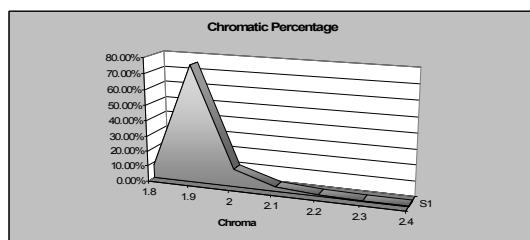
The general sonata form of the composition is manually marked in the images, according to the audio content. The four discrete parts of the quartet (written in sonata form) have been denoted: exposition (1-83), Development (84-132) and Recapitulation (133-260) with Coda (224-260) included. It should be noted that in the particular execution, there is no repetition of the Exposition. Moreover, during the execution of the movement by musicians, many rhythm fluctuations are observed (rallentando, accelerando etc.), in contrast to MIDI files.

As it can be seen in the chromatic graph, the basic color of the piece is a tint of pink. This is owned to the fact that the SBI of the quartet is 1.857, since the scale that has been matched from the 'Scale Matching' algorithm is Minor Harmonic. Moreover, the chromatic transitions that dominate the piece (as they appear in Figure 13) tend to blue and sometimes purple and brown. Tones in these transitions are more chromatic, as they move further the diatonic organizations. A certain hypothesis that should be examined here is whether this indicator shows that the performers try to sustain a rather high emotional state by using practices of increased chromaticism.

The greatest part of the quartet (76%) chromatically lies at almost 1.9 (Figure 13), which implies the prevalence of this chroma. Using Huron's terminology [6], there is presence of chroma addition by 0.1 (11.12%), as well as chroma reduction by 0.1 (8.43%). A notable

<sup>3</sup> Also, another recording by Amadeus Quartet has been analyzed. Its analysis showed a lot more Chromaticism than the presented one. It is not presented here, due to space constraints.

increase by 0.2 is observed in a percentage of 2.83%, which implies the salience of the phenomenon. An increase of chroma greater than 0.015 is statistically significant and it is not a very common phenomenon in many cases of western music. Such an increase is a result of continuous chromatic phenomena, which draw the attention for further analysis. That's why 0.015 is set as the  $\chi$  distance between different colors in our twelve-grade color scale. An increase of chroma greater than 2 usually betrays the temporary change of the underlying key or the transition to another music scale. Although the other chromatic deviations do not occur often in the context (<1% frequency), the greatest deviation that occurs at the end of the exposition should be mentioned ( $\chi \approx 3$ ). The incident represents the colorful melancholic impression of measures 75-80, as they are performed by Delirium Tremens.



**Figure 13.** Distribution of Chroma (audio recording).

By using the method of averaging the standard deviations as in the case of MIDI, we tried to compare the two graphs for similarities. Obviously, the comparison cannot be direct for two reasons. Firstly, the MIDI execution bears two repetitions of the Exposition, while the audio recording bears only one. Secondly, the times of execution in live performance are not analogical, since (as previously discussed) the musicians do not toe the mark of the rhythmic motive as in the MIDI execution. Consequently, since no total time concurrency exists, we examined each part separately by stretching the timers where needed. Similarities in chromatic patterns were detected in the Development and Coda (Figure 14).



**Figure 14.** Similarities in chromatic patterns between MIDI and audio: (a)Development, (b) Coda.

## 6. REFERENCES

- [1] V. Barski: *Chromaticism*, Harwood Academic Publishers, Netherlands, 1996.
- [2] M. Brown: "The Diatonic and the Chromatic in Schenker's 'Theory of Harmonic Relations'", *Journal of Music Theory*, Vol. 30, No. 1, pp. 1-33, 1986.
- [3] C. Buteau and J. Vipperman: "Representations of motivic spaces of a score in OpenMusic", *Journal of Mathematics and Music*, Vol. 2, No. 2, pp. 61-79, 2008.
- [4] A. Forte: "Motivic design and structural levels in the first movement of Brahms's String Quartet in C minor", *The Musical Quarterly*, Vol. 69, No. 4, pp. 471-502, 1983.
- [5] J. Godwin: *The Harmony of the Spheres: A Sourcebook of the Pythagorean Tradition*, in Music, Inner Traditions Intl. Ltd., 1993.
- [6] D. Huron: "What is a musical feature? Forte's analysis of Brahms's opus 51 no 1, revisited", *Music Theory Online*, Vol. 7, No. 4, 2001.
- [7] A. Jacobs: *The New Penguin Dictionary of Music*, Penguin, USA, 1980.
- [8] O. Lartillot: "Automated motivic analysis of Brahms' String Quartet based on multidimensional closed pattern mining", *International Workshop on Computational Music Analysis*, Paris, 2008.
- [9] D. Lewin: "Brahms, his past, and modes of music theory", in G. Bozarth (Ed), *Brahms Studies: Analytical and Historical Perspectives*, 1990.
- [10] D. Margounakis and D. Politis: "Converting Images to Music Using their Colour Properties", *Proceedings of the International Conference on Auditory Display*, London, England, 20 – 23 June 2006.
- [11] D. Margounakis, D. Politis and K. Mokos: "MEL-IRIS: An Online Tool for Audio Analysis and Music Indexing", *International Journal of Digital Multimedia Broadcasting*, vol. 2009, Article ID 806750, 2009.
- [12] D. Politis and D. Margounakis: "An Integrated Music Chromaticism Model", in *Recent Researches in Communications and Computers*, ISBN: 978-1-61804-109-8, pp. 183-187, 2012.
- [13] D. Politis and D. Margounakis: "Modeling Musical Chromaticism: The Algebra of Cross-Cultural Music Perception". *International Journal of Academic Research*, Vol. 2, No. 6, pp.20-29, 2010.
- [14] D. Politis and D. Margounakis: "Determining the Chromatic Index of Music", *Proceedings of the 3rd IEEE International Conference on Web Delivering of Music*, Leeds, pp. 95-102, 2003.
- [15] D. Politis, D. Margounakis and K. Mokos: "Visualizing the Chromatic Index of Music", *Proceedings of the 4th IEEE International Conference on Web Delivering of Music*, Barcelona, pp. 102-109, 2004.

# Are spectral elevation cues in head-related transfer functions distance-independent?

Simone Spagnol  
IUAV - University of Venice

## ABSTRACT

Since its title, this paper addresses one of the still open questions in sound localization: is our own perception of the elevation of a sound source affected by the distance of the source itself? The problem is addressed through the analysis of a recently published distance-dependent head-related transfer function (HRTF) database, which includes the responses of a single subject on a spatial grid spanning 14 elevation angles, 72 azimuth angles, and 8 distances comprised between 20 and 160 cm. Different HRTFs sharing the same angular coordinates are compared through spectral distortion and notch frequency deviation measurements. Results indicate that, even though the independence of spectral elevation cues from distance of the source can be assumed for the majority of all possible source directions, near-field HRTFs for sources close to the contralateral ear or around the horizontal plane in the ipsilateral side of the head are significantly affected by distance-dependent pinna reflections and spectral modifications.

## 1. BACKGROUND

It is undisputed that vertical localization is possible thanks to the presence of the pinnae [1]. Even though localization in any plane involves pinna cavities of both ears [2], determination of the perceived elevation angle of a sound source in the median plane is essentially a monaural process [3]. The external ear plays an important role by introducing peaks and notches in the high-frequency spectrum of the head-related transfer function (HRTF), whose center frequency, amplitude, and bandwidth greatly depend on the elevation angle of the sound source [4] and to a remarkably minor extent on azimuth [5]. Following two historical theories of localization, the pinna can be seen both as a filter in the frequency domain [6] and a delay-and-add reflection system in the time domain [7] as long as typical pinna reflection delays for elevation angles, clearly detectable by the human hearing apparatus [8], are seen to produce spectral notches in the high-frequency range.

Nevertheless, the relative importance of major peaks and notches in elevation perception has been disputed over the past years. A recent study [9] showed how a parametric HRTF recomposed using only the first, omnidirectional

peak in the HRTF spectrum (corresponding to Shaw's mode 1 [10]) coupled with the first two notches yields almost the same localization accuracy as the corresponding measured HRTF. Additional evidence in support of the lowest-frequency notches' relevance is given in [11], which states that the threshold for perceiving a shift in the central frequency of a spectral notch is consistent with the localization blur (i.e., the angular threshold for detecting changes in the direction of a sound source) on the median plane. Also, in [12] the authors judge increasing frontal elevation apparently cued by the increasing central frequency of a notch, and determine two different peak/notch patterns for representing the above and behind directions.

In general, hence, both pinna peaks and notches seem to play a primary function in vertical localization of a sound source,<sup>1</sup> even though it is difficult without extensive psychoacoustic evaluations to ascertain how importantly these features work as spatial cues. It has to be highlighted, however, that vertical localization bears little resolution compared with horizontal localization [13]. For the sake of record, the localization blur along the median plane was found to be never less than 4°, reaching a much larger threshold ( $\approx 17^\circ$ ) for unfamiliar speech sounds, as opposed to a localization blur of approximately 1° – 2° in the horizontal plane for a vast class of sounds [6]. Such a poor resolution is motivated by two basic observations:

- the need of high-frequency content (above 4–5 kHz) for accurate vertical localization [12, 14];
- the theoretically nonexistent interaural differences between the signals arriving at the left and right ear in the median plane.

Still, distance estimation of a sound source (see [15] for a comprehensive review on the topic) is even more troublesome than elevation detection. At a first level, when no other cue is available, sound intensity is the first variable that is taken into account: the weaker the intensity, the farther the source should be perceived. Under anechoic conditions, sound intensity reduction with increasing distance can be predicted through the inverse square law: intensity of an omnidirectional sound source will decay of approximately 6 dB for each doubling distance [16]. Still, a distant blast and a whisper at few centimeters from the ear could produce the same sound pressure level at the

Copyright: ©2012 Simone Spagnol. 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.

<sup>1</sup> In this context, it is important to point out that both peaks and notches in the high-frequency range are perceptually detectable as long as their amplitude and bandwidth are sufficiently marked [11], which is the case for most measured HRTFs.

eardrum. Having a certain familiarity with the involved sound is thus a second fundamental requirement [17].

However, the apparent distance of a sound source is systematically underestimated in an anechoic environment [18]. On the other hand, if the environment is reverberant, additional information can be given by the proportion of reflected to direct energy, the so-called *R/D ratio*, which functions as a stronger cue for distance than intensity: a sensation of changing distance occurs if the overall intensity is constant but the R/D ratio is altered [16]. Furthermore, distance-dependent spectral effects also have a role in everyday environments: higher frequencies are increasingly attenuated with distance due to air absorption effects.

Literature on source direction perception generally lies its foundations on a fundamental assumption, i.e. the sound source is sufficiently far from the listener. In particular, both previously discussed elevation cues as well as azimuth cues such as interaural time and level differences (ITD and ILD) are distance-independent when the source is in the so-called *far field* (approximately more than 1.5 m from the center of the head) where sound waves reaching the listener can be assumed to be plane. On the other hand, when the source is in the *near field* some of the HRTF features exhibit a clear dependence on distance. By gradually approaching the sound source to the listener's head in the near field, it was observed that low-frequency gain is emphasized; ITD slightly increases; and ILD dramatically increases across the whole spectrum for lateral sources [19]. In this paper, Brungart and Rabinowitz drew the following conclusions:

- ITD is roughly independent of distance even when the source is close;
- low-frequency ILDs are the dominant auditory distance cues in the near field;
- elevation cues are not correlated to distance-dependent features in the near field.

It should be then clear that ILD-related information needs to be considered in the near field, where dependence on distance cannot be approximated by a simple inverse square law. However, in [19] the last conclusion is supported just by graphical evidence on a limited number of HRTFs. Specifically, it is shown with the support of a single figure that the major features of the HRTFs at three distinct elevations and three distances are considerably more consistent across distance than across elevation. The authors argue that “*if this result generalizes to all elevations, it would imply that elevation cues are roughly independent of distance and that the same mechanisms that mediate elevation perception in the distal region (i.e. the far field) may also be used in the proximal region (i.e. the near field)*” but, to the best of my knowledge, this hypothesis has never been verified in the following literature. My aim in this paper is thus to investigate more deeply – through the analysis of a new distance-dependent HRTF database [20] – whether Brungart's claim on the rough independence between elevation cues and distance in the near field is well-grounded.

## 2. ANALYSIS OF DISTANCE-DEPENDENT HRTFS

Typically, HRTFs are measured by presenting a sound stimulus at several different spatial locations lying on the surface of a sphere centered in the subject's head, hence at one single distance (typically 1 m or farther). Most public HRTF databases, such as CIPIC [21] and LISTEN [22], follow this standard. Measuring HRTFs at closer distances introduces technical difficulties because a common loudspeaker cannot simulate an acoustic point source in the near field, and the sound source should be as compact as possible in order to avoid sound reflections from the loudspeaker back into the microphones [23].

Recently, Qu *et al.* [20] successfully overcame the problem by using a specialized spark gap as an appropriate acoustic point source (from the viewpoints of frequency response, signal-to-noise ratio, directivity, power attenuation and stability) to collect a spatially dense set of HRTFs of a KEMAR manikin. The database, that was updated and made available in June 2012,<sup>2</sup> includes the responses at both the left and right ears for 72 different azimuth angles, 14 different elevation angles, and 8 different distances ranging from 20 to 160 cm from the center of the manikin's head, totalling 6344 HRTFs. The so obtained HRTFs were seen to be comparable to the well known and widely used KEMAR HRTFs included in the CIPIC database. The following analysis will be based on the whole set of right-ear HRTFs.

### 2.1 Spectral distortion

In order to have a first quantification of the difference between HRTFs for various distances at fixed azimuth ( $\theta$ ) and elevation ( $\phi$ ) angles, an error measure widely used in recent literature [20, 24] is introduced: spectral distortion

$$SD(H, \tilde{H}) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( 20 \log_{10} \frac{|H(f_i)|}{|\tilde{H}(f_i)|} \right)^2} \quad [\text{dB}], \quad (1)$$

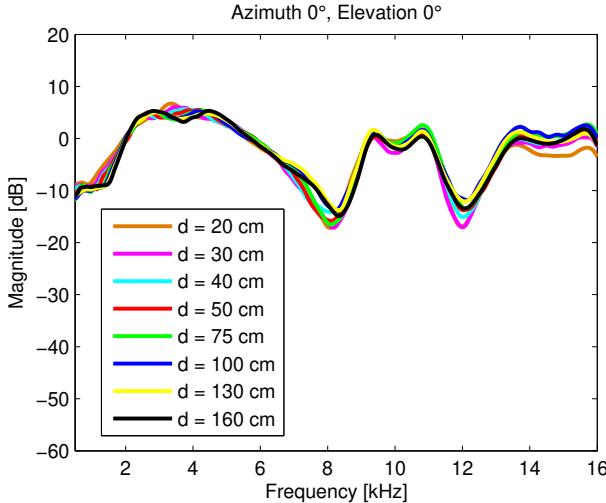
where  $H$  and  $\tilde{H}$  are the responses to be compared and  $N$  is the number of available points in the considered frequency range, that I choose to be  $R_1 = [500, 16000]$  Hz in order to include all the possible elevation cues. In the following analysis the reference HRTF,  $\tilde{H}$ , will always be the response for the farthest distance ( $d_8 = 160$  cm) approximating the far field response, while  $H$  will be the HRTF for one of the closest distances sharing the same angular coordinates  $(\theta_k, \phi_k)$ .

The analysis first requires the normalization of the responses in order to eliminate sound intensity cues. This is simply accomplished by dividing each HRTF by its mean magnitude in  $R_1$ ,

$$\hat{H}(f, \theta, \phi, d) = \frac{H(f, \theta, \phi, d)}{\sum_{i=1, N} |H(f_i, \theta, \phi, d)|} \times N. \quad (2)$$

The resulting normalized HRTFs for fixed angular coordinates are approximately aligned in magnitude, as is the case for  $(\theta, \phi) = (0^\circ, 0^\circ)$  in Figure 1.

<sup>2</sup> <http://www.cis.pku.edu.cn/auditory/Staff/Dr.Qu.files/Qu-HRTF-Database.html>



**Figure 1.** Normalized HRTFs for  $\theta = 0^\circ$  and  $\phi = 0^\circ$ .

Spectral distortion  $SD(\hat{H}(\theta_k, \phi_k, d), \hat{H}(\theta_k, \phi_k, d_8))$  can now be computed for each available angular coordinate<sup>3</sup> and each distance. The related results are represented in Figure 2, where the missing angular coordinates are conventionally assigned a 0-dB  $SD$  in the corresponding matrices' entries and coordinate  $(\theta, \phi) = (0^\circ, 90^\circ)$ , showing a  $SD$  never greater than 2.4 dB, is omitted for plotting issues.

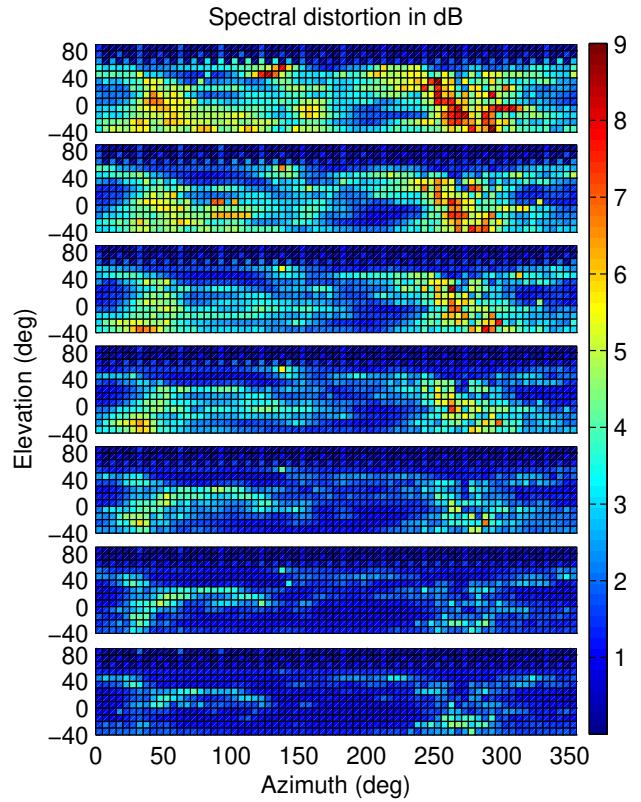
A first look at these results reveals that the initial hypothesis is verified for the vast majority of the spatial coordinates: 71% of the nonzero entries are less than 3 dB and 87% are less than 4 dB, two  $SD$  values that reflect a reasonable agreement between different HRTFs considering both the inter-measurement variability and the increasingly lowpass behaviour of the human head as the sound source approaches that can be noticed back in Figure 1 for the highest frequencies. The latter effect is thought to be responsible for the average increase in  $SD$  for decreasing distances clearly detectable in Figure 2. Also notice how the responses on the median plane ( $0^\circ$ -azimuth column) are always scarcely affected by distance:  $SD$  is never more than 4 dB.

Nevertheless, two major critical areas are shared by all of the seven plots:

1. a wide area ( $A_1$ ) extending across several azimuth angles in the ipsilateral side of the head ( $\theta = [0^\circ, 180^\circ]$ ) and concentrated around the horizontal plane, with a prominent tail around the coordinate  $(130^\circ, 40^\circ)$ ; and
2. a more compact area ( $A_2$ ) concentrated around the contralateral ear ( $\theta = 270^\circ$ ) at all elevations between  $-40^\circ$  and  $40^\circ$ .

Here  $SD$  increases up to 9 dB for the closest distances and could imply the involvement of an effect of distance on the

<sup>3</sup> Taking the vertical polar coordinate system as reference, elevation goes from  $-40^\circ$  to  $90^\circ$  in  $10^\circ$  steps, while azimuth goes from  $0^\circ$  to  $355^\circ$  in  $5^\circ$  steps except for elevation  $60^\circ$  ( $10^\circ$  steps),  $70^\circ$  ( $15^\circ$  steps),  $80^\circ$  ( $30^\circ$  steps), and  $90^\circ$  ( $\theta = 0^\circ$  only). The  $(0^\circ, 0^\circ)$  direction is right in front of the listener,  $(90^\circ, 0^\circ)$  is at the right ear, and  $(270^\circ, 0^\circ)$  is at the left ear.



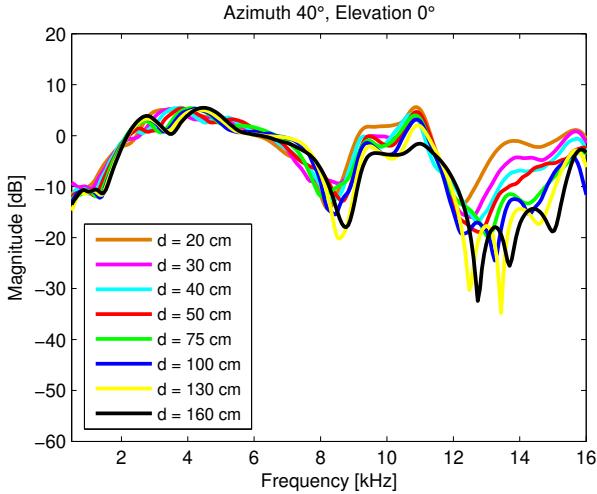
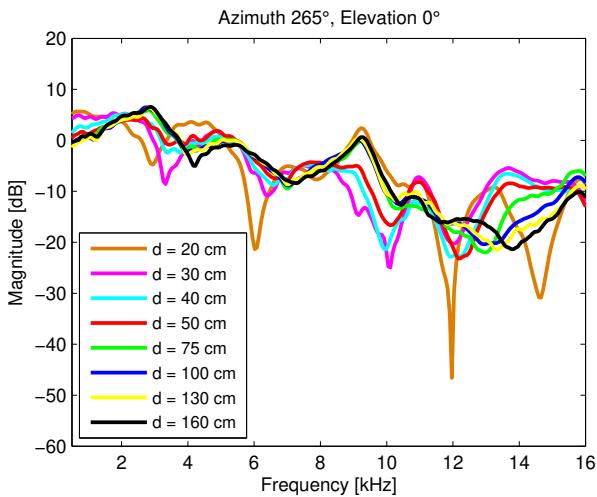
**Figure 2.** Spectral distortion between HRTFs at distance  $d = 160$  cm and distance (top-to-bottom):  $d = 20$  cm,  $d = 30$  cm,  $d = 40$  cm,  $d = 50$  cm,  $d = 75$  cm,  $d = 100$  cm,  $d = 130$  cm.

spectral features of the HRTF. Thus, further investigation is needed in order to understand the cause of such a systematic  $SD$  rise.

## 2.2 Deviation of spectral notches

We first examine what happens at two of the most critical angular coordinates,  $(\theta, \phi) = (40^\circ, 0^\circ)$  in  $A_1$  and  $(\theta, \phi) = (265^\circ, 0^\circ)$  in  $A_2$ . The corresponding normalized HRTFs are traced in Figure 3 and Figure 4, respectively. What can be immediately seen in both figures is that the greatest dissimilarities among HRTFs are caused by the difference in both magnitude and frequency among the spectral notches and, to a minor extent, among the spectral peaks (as is the case of the peak around 9 kHz in Figure 4). Furthermore, some of the notches belong just to a strict subset of the 8 HRTFs, see e.g. the spectral notch at 14 – 15 kHz appearing only for the farthest distances in Figure 3.

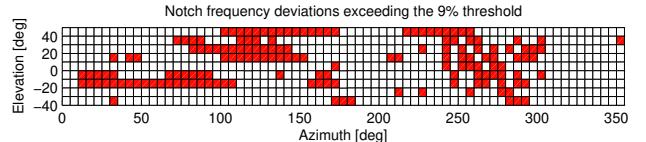
It should be then pointed out that, although the gross characteristics of the HRTF are preserved across distances, the presence/absence or displacement of some of the most important spectral cues for elevation detection could have an impact on localization. As a matter of fact, one could verify that notch shifts in the range of 1 kHz such as those appearing in Figure 4 usually correspond to an increase or decrease of the elevation angle greater than  $20^\circ$  [25]. Fur-

**Figure 3.** Normalized HRTFs for  $\theta = 40^\circ$  and  $\phi = 0^\circ$ .**Figure 4.** Normalized HRTFs for  $\theta = 265^\circ$  and  $\phi = 0^\circ$ .

thermore, since the work of Moore *et al.* [11] we know that two steady notches in the high-frequency range differing just in center frequency are distinguishable on average if the mismatch is more than approximately 9% of the center frequency  $f_c$  of the lowest, regardless of notch bandwidth.<sup>4</sup> Although these results were found for just one frequency band (around 8 kHz), I may extend the validity of the assumption to the range  $R_2 = [6, 11]$  kHz where the first two spectral notches, generally the most relevant in elevation perception [9], usually lie.

Using such assumption, another error measure is now introduced in order to attest whether HRTFs at different distances can potentially be distinguishable due to a frequency shift of one or more of their frequency notches. Having fixed the deepest notch in the range  $R_2$  appearing in  $H(\theta_k, \phi_k, d_8)$  as reference, let me define as the *notch frequency deviation* among the corresponding notches in the set of HRTFs  $H(\theta_k, \phi_k, d_1), \dots, H(\theta_k, \phi_k, d_8)$ , where  $d_1, \dots, d_8$  are the eight available distances in increasing

<sup>4</sup> By contrast, the perceptual relevance of changes in bandwidth and amplitude of a notch is little understood in previous literature.

**Figure 5.** Notch frequency deviations exceeding the 9% threshold between  $\phi = -40^\circ$  and  $\phi = 40^\circ$ .

order, and denote it as  $dev(\theta_k, \phi_k)$ , the difference in Hz between the frequency of the highest and the frequency of the lowest notch in the set, where available – if not (i.e. the notch is missing in one or more HRTFs),  $dev(\theta_k, \phi_k)$  will conventionally take infinite value.

Each notch frequency deviation shall now be related to the aforementioned 9% threshold by simply expressing it as the percentual amount of the lowest notch frequency in the set. Figure 5 reports in a simple two-value matrix all those deviations that exceed the threshold, indicating a potential significant difference in the relative HRTF set, as positive (red) entries. In the figure, elevations greater than  $\phi = 40^\circ$  are omitted because of the known lack of deep spectral notches for directions above the listener [25], where the notch frequency deviation metric loses its consistency.

Comparing this last matrix with those reported back in Figure 2 we can immediately notice a good agreement between the leftmost positive entries of the matrix and  $A_1$ , and a very good correspondence between the rightmost positive entries and  $A_2$ . The only significant differences between the two representations are

1. a greater number of positive entries around the tail of  $A_1$ , indicating that the notch frequency deviation metric begins to become inappropriate at medium-to-high elevations because notches are too shallow to be significant; and
2. a small number of positive entries for  $\phi = -40^\circ$  and  $\phi = -30^\circ$  meaning that, even though significant notch frequency deviations are rare in this range, the very deep notches appearing at low elevations greatly affect the *SD* computation even in presence of small notch frequency shifts.

Apart from these specific considerations, results suggest that the frequency deviation of spectral notches across distances is indeed the greatest source of spectral distortion among iso-directional HRTFs. Hence, the initial hypothesis on the independence between spectral elevation cues and distance is in this case not guaranteed for directions included in the previously defined areas  $A_1$  and  $A_2$ . From the viewpoint of the listener, modifications of spectral features for sources close to the contralateral ear (area  $A_2$ ) could be thought of having little effect on the perception of a sound source, as the ipsilateral ear will always receive a much louder signal from which it shall monaurally correctly decode the elevation of the source [2]. However, notch deviations in the ipsilateral side of the head (area  $A_1$ ) could as well have an effect on elevation perception.

### 3. CONCLUSIONS

By analyzing a recent database of distance-dependent head-related transfer functions, I found how the rough independence of elevation cues from distance advanced in [19] can not be attested at a purely analytical level for all directions of the sound source. The analysis was conducted on a single subject, a KEMAR manikin: KEMAR-related measurements (a smaller number of which were also used in [19] to support the incriminated claim) have the considerable advantage of being fully controllable, whereas similar measurements on a human subject would intolerably multiplicate the measurement time required to collect a standard set of HRTFs by the number of source distances, thus proportionally increasing the subject's tiredness and uneasiness. Nevertheless, a future contingent availability of distance-dependent HRTF sets measured on human subjects will allow a similar data analysis to be repeated, and the presented results to be further verified.

### 4. ACKNOWLEDGMENTS

I would like to acknowledge Michele Geronazzo and Dr. Federico Avanzini for the precious conversations throughout these last years that allowed the present work to be conceived and realized.

### 5. REFERENCES

- [1] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: Effect of pinnae cavity occlusion," *J. Acoust. Soc. Am.*, vol. 53, no. 2, pp. 400–408, 1973.
- [2] M. Morimoto, "The contribution of two ears to the perception of vertical angle in sagittal planes," *J. Acoust. Soc. Am.*, vol. 109, pp. 1596–1603, April 2001.
- [3] J. Hebrank and D. Wright, "Are two ears necessary for localization of sound sources on the median plane?," *J. Acoust. Soc. Am.*, vol. 56, pp. 935–938, September 1974.
- [4] E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Am.*, vol. 44, no. 1, pp. 240–249, 1968.
- [5] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Am.*, vol. 100, pp. 3248–3259, November 1996.
- [6] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1983.
- [7] 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.
- [8] D. Wright, J. H. Hebrank, and B. Wilson, "Pinna reflections as cues for localization," *J. Acoust. Soc. Am.*, vol. 56, pp. 957–962, September 1974.
- [9] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Appl. Acoust.*, vol. 68, pp. 835–850, 2007.
- [10] E. A. G. Shaw, "Acoustical features of human ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, pp. 25–47, Mahwah, NJ, USA: R. H. Gilkey and T. R. Anderson, Lawrence Erlbaum Associates, 1997.
- [11] B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Am.*, vol. 85, pp. 820–836, February 1989.
- [12] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1829–1834, December 1974.
- [13] A. Wilska, *Studies on Directional Hearing*. English translation, Aalto University School of Science and Technology, Department of Signal Processing and Acoustics, 2010. PhD thesis originally published in German as "Untersuchungen über das Richtungshören", University of Helsinki, 1938.
- [14] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Am.*, vol. 88, pp. 159–168, July 1990.
- [15] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: a summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, pp. 409–420, 2005.
- [16] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. San Diego, CA, USA: Academic Press Professional, Inc., 1994.
- [17] M. B. Gardner, "Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space," *J. Acoust. Soc. Am.*, vol. 45, no. 1, pp. 47–53, 1969.
- [18] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, no. 3, pp. 311–322, 1979.
- [19] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, pp. 1465–1479, September 1999.
- [20] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu, "Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, pp. 1124–1132, August 2009.
- [21] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, (New Paltz, New York, USA), pp. 1–4, 2001.

- [22] G. Eckel, "Immersive audio-augmented environments - the LISTEN project," in *Proc. 5th IEEE Int. Conf. Info. Visualization (IV'01)*, (Los Alamitos, CA, USA), pp. 571–573, 2001.
- [23] A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *J. Acoust. Soc. Am.*, vol. 125, pp. 2233–2242, April 2009.
- [24] M. Otani, T. Hirahara, and S. Ise, "Numerical study on source-distance dependency of head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 125, pp. 3253–3261, May 2009.
- [25] S. Spagnol, M. Hiipakka, and V. Pulkki, "A single-azimuth pinna-related transfer function database," in *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, (Paris, France), September 2011.

# EXPLORING THE RELATIONS BETWEEN CHROMATICISM, FAMILIARITY, SCALES AND EMOTIONAL RESPONSES IN MUSIC

Dimitrios Margounakis

Dept. of Informatics,  
School of Sciences

Aristotle University of Thessaloniki  
GR-541 24, GREECE  
dmargoun@csd.auth.gr

Dionyios Politis

Dept. of Informatics,  
School of Sciences  
Aristotle University of Thessaloniki  
GR-541 24, GREECE  
dpolitis@csd.auth.gr

## ABSTRACT

The most common form of synesthesia is the experience of colors linked to sounds and music. This paper presents the primary experimental results for creating a unique music visualization computer-based system that functions in a synesthetic way. The idea behind the development is not the mapping of sounds to colors, but actual feelings and colors to melodic elements in a semantic way. The main objective of the research is to provide results to help answer (a) if there are relations between chromaticism, familiarity, scales and degrees of emotional arousal and (b) if certain melodic implications stimulate certain emotional states. An experiment has been designed to investigate the previously stated questions. The experiment took place in May, 2012. The sample includes 27 adults regardless of their musical background. Four familiar tunes and four unfamiliar melodies in major and minor keys were randomly presented to the participants for evaluation. Eight different realized intervals for each melody tested how the change in direction, size and chromaticism of the realization affected the emotional state of the audience.

## 1. INTRODUCTION

Synesthesia is a condition in which the real information of one sense is accompanied by a perception in another sense. The most common form of synesthesia pertains to the experience of colors. Synesthesia is characterized by phenomena of cross-modal association that are involuntary, projected, durable (discrete and generic), memorable, emotional (and noetic) [4]. The term *color music* is dated back to the ages of the Greek philosopher Aristotle.

The synesthetic association of music to colors has been investigated in several studies [2] [6]. A lot of theories have been deployed about the feelings that colors provoke [1], [12]. Also, several studies have used color-emotion matching tasks [14]; matching colors

(e.g., red, yellow, blue) to a certain number of emotions (e.g., happiness, sadness, anger). There is a general assumption that different aesthetic domains, like color and music, might be related to each other through common emotional associations. The question that arises is: “To what extent can a color model provide a useful metaphor to describe melodic progression in a emotionally semantic way?”

It is discussed in literature that there exist emotional attributions to simple musical elements. Cooke suggested in 1959 that intervals have universal emotional effects [3]. Specific patterns of acoustic cues, which may be described in terms of tones and tone relations (intervals), can trigger specific emotions. Also, expectancy plays an important role in the emotional and affective responses. According to Meyer, emotion and affect are heightened when musical expectations are unfulfilled [7]. Narmour suggests that the listener’s perception of melodic structure is affected by melodic implications [8]. According to this theory, these implications are stronger when successive notes (implicative intervals) move away from tonal stability and metric stress.

Before proceeding in modeling the music-color associations, it is necessary to define the perceptual correlations that exist. Therefore, this paper presents a psychoacoustic experiment that took place in May, 2012. The experiment aims at the exploration of the emotional dimensions and the musical expectations of the audience, when only one factor is involved and affects them: pitch (and by extension the melodic intervals). The parameters, which are tested in this presentation, are melody-related (key, chromaticism) and user-related (familiarity to the melodies, emotional responses).

In a similar experiment, Palmer et. al. found that faster music and major mode are associated with lighter, more saturated, yellower colors, whereas slower music and minor mode are associated with darker, desaturated, bluer colors [9]. Their hypothesis to be tested for the associations was: *When choosing colors to “go best with” an emotional stimulus in a different modality, people choose the colors that best match the emotional content of that stimulus.* [9]

The authors have also created and presented the system MEL-IRIS for visualizing music with colors in a

synesthetic way [10]. The model of the authors is based on the fact that chromaticism in music affects the music perception of the audience by creating different emotional states.

## 2. THE EXPERIMENT

The main objective of the research is to provide results to help answer the following questions:

- Is there a relation between chromaticism, familiarity, scales and degrees of emotional arousal?
- Do certain melodic implications stimulate certain emotional states?

These answers, arising from the evaluation of the participants' emotional involvement, are descriptive of Western-trained ears since all the participants belong to that cultural context.

### 2.1 Participants

The sample included 27 adults. All of them had everyday exposure to Western music, while some of them had some special musical training (although this was not a prerequisite in order to take the test). One group of the sample ( $n = 14$ ) took the online test on their own, after our invite to participate through social networks. The other group consisted of 13 undergraduate students of computer science, who were registered in a computer music course. The second group participated in the experiment in the multimedia lab of the computer science department under the authors' supervision.

Adults in the first group were of several age groups ( $M = 30$  years, 2 months;  $SD = 4$  years, 11 months; range = 25 to 43 years). Half of them had never taken formal music lessons, 2 of them had some formal music knowledge (1-2 years in the past), 3 of them had average music knowledge (partially-completed formal music studies in the past), and 2 of them had adequate musical training (many years of musical lessons and/or a music degree or a diploma in a musical instrument). Students in the second group were generally younger ( $M = 23$  years, 8 months;  $SD = 1$  year, 11 months; range = 21 to 26 years). 6 of them had never taken formal music lessons, 5 of them had some formal music knowledge, and 2 of them had average music knowledge. None of the participants in the second group was an expert musician. Three additional participants (two in the first group and one in the second group) were excluded from the final sample, since they didn't complete the entire range of questions.

### 2.2 Apparatus

Stimulus melodies were created initially as ogg files using notation software (musescore). The exported ogg files in musescore had common settings: acoustic piano timbre, 120 bpm, A4 at 440 Hz, same reverb and chorus. Finally, the ogg files were fed into the online audio converter media.io, in order to produce their corre-

sponding mp3 files of high quality (192 kbit/s). Stimuli were presented on participants' personal computers. The students of the second group listened to the melodies at a medium volume with personal stereo headphones. It is assumed that the individual participants of the first group used either headphones or their personal computer's speakers. All of them used a mouse to initiate trials and record their responses.

### 2.3 Stimuli

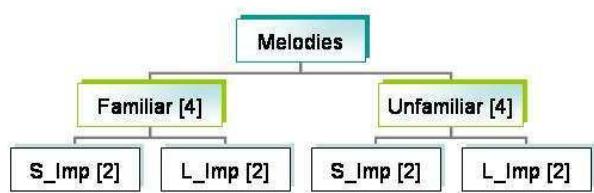
The stimulus melodies were four fragments taken from familiar Western tunes, and four fragments taken from the unfamiliar Acadian folk songs that were used in [11]. The eight melodies can be seen in Figure 1. All fragments contained at least 11 notes. This allowed for the participants to perceive well the major or the minor key of each melody. Since the fragments originate from notation software with common attributes, no other factors (such as timbre, tempo, dynamics etc) affect the auditory and emotional perception of the participants. The only factor of importance here is pitch. All fragments end in the middle of a melodic phrase, creating therefore implicative intervals, as they are defined in [8]. Two of the familiar and two of the unfamiliar melodies (namely Melodies A, B, E and F) end in small implicative intervals. All of them end in an upward implicative interval, except for melody 2 where the implicative interval is downward. Respectively, Melodies C, D, G and H end in large implicative intervals. All of the large implicative intervals are upward. The hierarchy of the melodic fragments of the experiment can be seen in Figure 2(a).



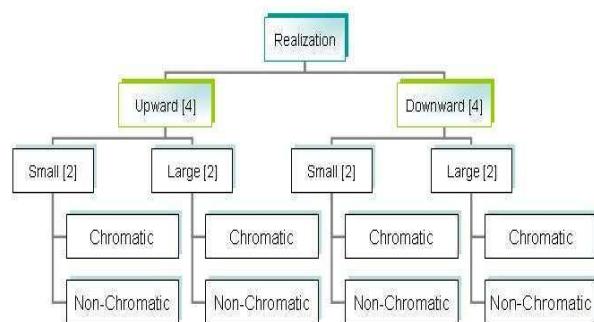
**Figure 1.** The 8 melodic fragments used in Experiment.

Eight test tones were generated as possible continuations of each unfinished melody. Figure 2(b) shows the binary tree of each possible continuation. According to this hierarchy, there are three dimensional levels of depth for each realized interval: direction, size and chromaticism (classifying variables). E.g., the first possible continuation for melody 1 would be a small upward chromatic interval (in our case F# has been used). Eight test tones for each of the eight melodies result in 64 stimulus melodies that were used in this experiment. Test tones had a longer duration than the final note of the melody to support the feeling of closure.

(a)



(b)



**Figure 2.** (a) The binary tree structure of the melodies in the experiment. The classifying variables are: familiarity and size of implicative intervals. (b) The binary tree structure of the realized intervals for each melody in the experiment. The classifying variables are: direction, size and chromaticism.

#### 2.4 Procedure

Participants of the first group were tested individually on their own PCs and the instructions were given to them on the computer screen. Students of the second group were given instructions both verbally and on the computer screen in the laboratory. After the first audition of the melody, the participants had to rate the degree of emotional arousal (caused by the particular fragment they just listened to) on a scale from 1 (*no emotional arousal at all*) to 10 (*extremely high emotional arousal*). Afterwards they had to rate specific emotional dimensions that this melody caused to them in terms of opposite concepts. The dimensions were namely: *sadness – joy*, *unhappiness – happiness*, *frustration – calmness*, *seriousness – playfulness*, *anger – tranquility*. The ratings on these concepts were on a scale from 1 (the absolutely negative emotion) to 9 (the

absolutely positive emotion). 5 on that scale means neutral (neither negative nor positive). For example, a rating of 7 at the unhappy – happy dimension means quite happy, while a rating of 4 means a bit unhappy. Before the actual experiment, participants were asked to try these questions on a trial fragment in order to practice.

The 64 fragments were put in random order, so no obvious connections between fragments could affect the participants' ratings. The actual time needed for the entire testing procedure was about one hour. Since the procedure proved to be time-consuming (64 fragments x 8 ratings = 512 ratings)<sup>1</sup>, participants in the lab were asked to have a 10-minutes break after the first 32 melodies. Individual participants of the first group were also encouraged to have a break at that point with a text message on screen. Three participants did not have the patience to complete the test, so their incomplete ratings were excluded from the sample.

### 3. RESULTS AND DISCUSSION

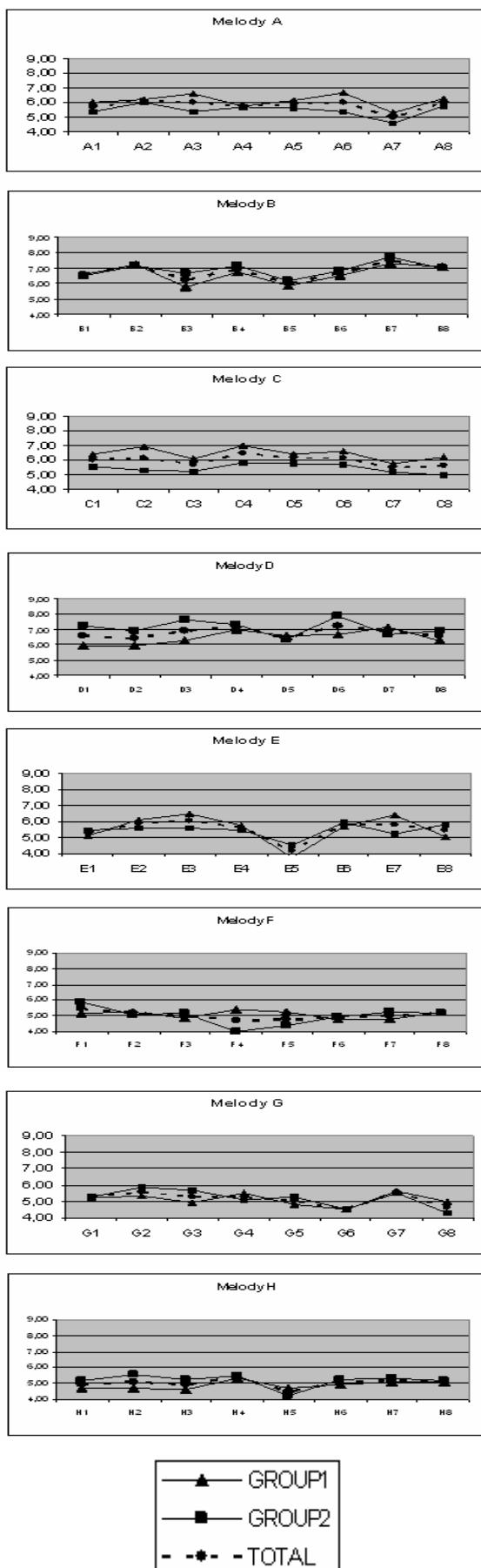
#### 3.1 Abbreviations

Certain abbreviations will be used for the rest of the experiment's presentation. The eight aforementioned melodies are symbolized with the Latin characters *A*, *B*, *C* ... *H*. Melodies *A*, *B*, *C* and *D* constitute the familiar melodies, while melodies *E*, *F*, *G* and *H* constitute the unfamiliar Acadian tunes (see Figure 1). The eight instances of each melody (according to the realized interval) are symbolized with numbers 1 to 8. Following the binary tree of Figure 2(b), instance 1 means the small, upward and chromatic realization, instance 2 means the small, upward and non-chromatic realization, instance 3 means the large, upward and chromatic realization, and so on. For example, case *C6* is the case of the small, downward, non-chromatic realized interval of the familiar melody *C*. Therefore, 64 cases are to be examined: *A1*, *A2*, ..., *H8*. The pairs of emotional dimensions under consideration (*sadness – joy*, *unhappiness – happiness*, *frustration – calmness*, *seriousness – playfulness*, *anger – tranquility*) are represented as *P1*, *P2*, ..., *P5*. Therefore, variable *P1* defines the ratings of the sadness-joy dimension, variable *P2* defines the ratings of the unhappiness-happiness dimension, and so forth. The variable *emotional arousal* is symbolized as *EM*.

#### 3.2 Average Ratings

Figure 3 shows the average ratings of emotional arousal, for all 64 melodic cases. Although the ratings of both groups are depicted, the whole sample will be used as a whole in the next sections, since no significant differences were found between the two groups' ratings in all the variables.

<sup>1</sup> There was also a second part in the experiment, which contained two more ratings (continuation and expectancy). It is not discussed in this paper.

**Figure 3.** Ratings of emotional arousal for all melodies

Emot. Dim.	A	B	C	D	E	F	G	H
<b>1-P1</b>	6.11	4.93	6.30	4.26	6.19	4.96	5.26	5.26
<b>1-P2</b>	6.11	4.85	6.22	4.30	5.81	4.78	5.04	5.11
<b>1-P3</b>	5.63	5.44	5.78	4.63	5.19	5.15	5.33	4.85
<b>1-P4</b>	6.30	4.85	6.59	3.04	6.15	5.15	5.15	5.26
<b>1-P5</b>	5.81	5.89	5.56	4.96	5.52	5.30	5.67	5.22
<b>2-P1</b>	6.63	5.37	6.78	4.07	6.59	5.48	5.48	5.81
<b>2-P2</b>	6.67	5.48	6.56	4.41	6.30	5.33	5.33	5.56
<b>2-P3</b>	5.89	6.37	5.74	4.89	5.96	5.59	5.15	5.19
<b>2-P4</b>	6.41	4.33	7.26	3.19	6.78	5.11	5.30	5.89
<b>2-P5</b>	6.22	6.56	6.33	5.48	6.11	5.74	5.56	5.52
<b>3-P1</b>	6.52	5.04	6.04	4.07	6.26	5.04	5.30	5.15
<b>3-P2</b>	6.44	5.41	5.96	3.96	6.26	4.93	4.93	5.11
<b>3-P3</b>	5.41	5.56	5.78	4.48	5.11	4.78	4.26	4.41
<b>3-P4</b>	6.70	4.56	6.48	3.30	6.44	4.96	4.96	5.04
<b>3-P5</b>	5.85	5.93	5.59	4.63	5.63	5.00	4.48	4.74
<b>4-P1</b>	6.52	5.15	6.74	3.74	6.44	4.96	4.59	5.74
<b>4-P2</b>	6.59	5.22	6.48	4.11	6.19	4.48	4.81	5.44
<b>4-P3</b>	6.00	5.44	5.67	4.74	5.56	4.37	4.70	4.85
<b>4-P4</b>	6.93	4.41	7.04	3.15	6.44	4.44	4.52	5.59
<b>4-P5</b>	6.07	5.89	6.41	5.70	5.89	4.93	4.96	5.22
<b>5-P1</b>	6.04	5.07	6.41	4.33	5.04	4.48	5.15	5.19
<b>5-P2</b>	6.26	4.81	6.33	4.22	4.81	4.63	5.04	5.07
<b>5-P3</b>	5.93	4.85	5.44	4.93	4.89	4.56	5.11	5.15
<b>5-P4</b>	6.15	4.33	6.56	3.07	5.11	4.44	4.67	5.00
<b>5-P5</b>	5.74	5.15	5.78	4.96	5.19	4.89	5.19	5.15
<b>6-P1</b>	6.67	4.96	6.96	3.74	6.81	5.11	5.04	5.41
<b>6-P2</b>	6.81	4.93	6.78	3.89	6.56	5.11	4.96	5.41
<b>6-P3</b>	6.85	5.44	6.22	6.11	6.19	5.22	5.41	5.67
<b>6-P4</b>	6.26	4.22	6.93	2.52	6.85	4.67	4.85	4.93
<b>6-P5</b>	6.89	5.59	6.59	6.11	6.26	5.30	5.41	5.59
<b>7-P1</b>	6.00	4.67	6.30	3.81	6.15	4.93	5.00	5.26
<b>7-P2</b>	6.04	4.78	6.22	3.78	5.78	4.48	4.78	4.85
<b>7-P3</b>	5.52	5.04	5.70	5.67	5.26	4.33	4.96	4.89
<b>7-P4</b>	6.15	4.15	5.96	2.96	6.22	4.26	4.37	4.93
<b>7-P5</b>	6.37	5.26	5.63	5.67	5.52	4.81	4.93	5.00
<b>8-P1</b>	6.11	5.00	6.19	3.67	6.04	5.07	5.07	5.59
<b>8-P2</b>	6.22	5.19	6.15	3.70	5.74	4.63	4.89	5.33
<b>8-P3</b>	6.04	5.22	5.56	4.81	5.41	4.41	5.44	5.00
<b>8-P4</b>	6.48	4.78	6.26	2.74	6.44	4.78	4.81	4.89
<b>8-P5</b>	6.15	5.56	6.15	5.63	5.67	4.44	5.44	5.30

**Table 1.** Average ratings of the emotional dimensions for all cases.

Table 1 shows the average ratings of the emotional dimensions for all 64 melodic cases. Columns represent the melodies (A-H), while rows represent the emotional dimensions (P1...P5) for all cases of realized intervals.

### 3.3 Results and Discussion

In this section, we present some statistical findings of this research concerning the several factors that may affect the emotions of the audience. All tests have been conducted with 0.05 level of significance.

#### 3.3.1 Chromaticism and Emotions

A certain research question is whether chromatic intervals (as closures) affect the emotional reaction of the audience. According to Meyer, chromaticism in music has an affective aesthetic power, because chromatic alterations delay or block the expected motion to the normal diatonic tones [7]. For that purpose, we examined the null hypothesis:

*H1a. There is no significant difference between the emotions that arise from chromatic closures and non-chromatic closures.*

Five more similar null hypotheses *H1b ... H1f* were tested, concerning the variables *P1... P5*<sup>2</sup>. The differences were tested using independent't-test'. The tests were conducted on average scores of the variables. The chromatic groups of the tests contained: chromatic closures (instances 1,3,5 and 7 of realized intervals) and non-chromatic closures (instances 2,4,6 and 8). The results can be seen in Table 2.

Levene's Test					
	F	Sig.	t	df	Sig. (2-tailed)
EM. Equal var.	.121	.729	-.927	62	.358
P1 Equal var.	1.997	.163	-1.027	62	.308
P2 Equal var.	.490	.487	-1.249	62	.216
P3 Equal var.	1.313	.256	-2.650	62	<b>.010</b>
P4 Equal var.	1.866	.177	-.613	62	.542
P5 Equal var.	.612	.437	-3.489	62	<b>.001</b>

**Table 2.** The comparison of chromatic and non-chromatic closures' emotional effects. Equal variances assumed for all variables (according to Levene's test).

CHR_CATEGORY	N	Mean	Std. Deviation	Std. Error Mean
P3 Chromatic	32	5.1256	.45454	.08035
	32	5.4722	.58373	.10319
P5 Chromatic	32	5.3444	.44075	.07791
	32	5.7709	.53291	.09421

**Table 3.** Group statistics for the variables P3 and P5.

As shown by the results, there were significant differences between chromatic and non-chromatic closures in relation to variables *P3* ( $t=2.65$ ,  $p<0.05$ ) and *P5* ( $t=3.489$ ,  $p<0.05$ ). That is (taking into account group statistics in Table 3), *melodies with chromatic closures cause significantly more frustration and anger to the audience in contrast to melodies with diatonic closures that cause more calmness and tranquility*. There weren't significant differences on chromaticism in the emotional variables EM, P1, P2 and P4 ( $p>.05$ ).

### 3.3.2 Scales and Emotions

Another research question is whether the scale, in which a melody is written, affects certain emotional parameters. Major keys are said to be associated with happier responses, while minor keys with sad ones [9] [13]. This statement is supported by the testing of the following null hypothesis:

*H2a. There is no significant difference between the emotions that arise from melodies in major keys and melodies in minor keys.*

Again, the null hypotheses *H2b ... H2f* (concerning the variables *P1..P5*) are tested together with *H2a* using independent't-test'. The groups for the test contained the average scores of variables for melodies written in major keys (namely *A, C, E, G* and *H*) and melodies written in minor keys (*B, D* and *F*). The results can be seen in Table 4.

Levene's Test					
	F	Sig.	t	df	Sig. (2-tailed)
EM. Not Equal	15.427	.000	-3.425	32.617	.002
P1 Equal var.	.846	.361	7.920	62	.000
P2 Not Equal	4.677	.034	7.604	57.233	.000
P3 Equal var.	.233	.631	2.527	62	.014
P4 Equal var.	.061	.805	8.557	62	.000
P5 Equal var.	.039	.844	1.994	62	.051

**Table 4.** The comparison of major and minor scales' emotional effects. According to Levene's test equal variances are not assumed for *EM* and *P2*. Equal variances assumed for all other variables.

The null hypothesis is rejected for *EM*, *P1*, *P2*, *P3* and *P4* since the results in Table 4 showed significant differences between melodies in major keys and melodies in minor keys in relation to the emotional variables. In particular, *melodies in minor keys cause significantly more intense feelings than melodies in major keys* ( $t=3.425$ ,  $p<0.05$ ). Moreover, *melodies in minor keys cause significantly more sadness* ( $t=7.92$ ,  $p<0.05$ ), *unhappiness* ( $t=7.604$ ,  $p<0.05$ ), *frustration* ( $t=2.527$ ,  $p<0.05$ ), and *seriousness* ( $t=8.557$ ,  $p<0.05$ ) *than melodies in major key, which tend to be joyful, happy, calm and playful*.

Concerning the emotional dimension *P5* (anger-tranquility), the statistical test has extracted a borderline value  $p=0.051$ . Since the confidence level is 95%, we have to accept the null hypothesis *H2f* (*the emotional pair anger-tranquility is not affected by the scale in which a melody is written*) in this case. The average score for major keys (5.66) is more inclined to tranquility than that for minor keys (5.39) (Table 5).

<sup>2</sup> Null Hypothesis *H1a* tests the *emotional arousal* variable. The variables of emotional dimensions *P1..P5* are always tested separately, whenever this variable is tested.

SCALE_		N	Mean	Std. Deviation	Std. Error Mean
EM	Major	40	5.4938	.53154	.08404
	Minor	24	6.1908	.90800	.18535
P1	Major	40	5.9035	.63345	.10016
	Minor	24	4.6629	.55827	.11396
P2	Major	40	5.7738	.65862	.10414
	Minor	24	4.6421	.52087	.10632
P3	Major	40	5.4275	.51614	.08161
	Minor	24	5.0846	.54117	.11047
P4	Major	40	5.8763	.82328	.13017
	Minor	24	4.0588	.82149	.16769
P5	Major	40	5.6578	.52299	.08269
	Minor	24	5.3908	.51080	.10427

**Table 5.** Group statistics for all the emotional variables.

### 3.3.3 Familiarity and Emotions

In this section, we test whether the familiar tunes cause different emotional responses to the audience than the unfamiliar tunes. Contemporary research has brought to light clues that the feeling of familiarity evoked by a musical excerpt could be accompanied by other processing mechanisms at the conceptual level [5]. Therefore, the following hypothesis is tested:

*H3a. There is no significant difference between the emotions that arise from familiar melodies to the audience and unfamiliar melodies to the audience.*

The null hypotheses *H3b ... H3f* are tested together with *H3a* using independent 't-test', as in the previous two tests. The average scores were grouped in familiar melodies (*A, B, C* and *D*) and unfamiliar melodies (*E, F, G* and *H*). The results can be seen in Table 6.

According to the results, there were significant differences between familiar and unfamiliar melodies in relation to variables *EM* ( $t=8.873$ ,  $p<0.05$ ), *P3* ( $t=3.592$ ,  $p<0.05$ ) and *P5* ( $t=4.433$ ,  $p<0.05$ ). Therefore, the null hypotheses *H3a, H3d* and *H3f* are rejected. Instead, we can conclude (because of Table 7) that *familiar melodies cause significantly more intense feelings than unfamiliar melodies. Moreover, unfamiliar melodies cause significantly more frustration and anger to the audience in contrast to familiar melodies that cause more calmness and tranquility.*

The latter coincides with the deduction of Section 3.3.1 about chromaticism and emotional responses. It seems that unfamiliar tunes with chromatic intervals can cause anger and frustration to the audience.

		Levene's Test				
		F	Sig.	t	df	Sig. (2-tailed)
EM.	Equal var.	3.700	.059	8.873	62	.000
P1	Not Equal	23.485	.000	.051	47.585	.960
P2	Not Equal	22.632	.000	1.130	47.459	.264
P3	Equal var.	.428	.516	3.592	62	.001
P4	Not Equal	38.031	.000	-.447	44.440	.657
P5	Equal var.	.520	.474	4.433	62	.000

**Table 6.** The comparison of familiar and unfamiliar melodies' emotional effects. According to Levene's test equal variances are not assumed for *P1, P2* and *P4*. Equal variances assumed for all other variables.

FAMILIARITY		N	Mean	Std. Deviation	Std. Error Mean
EM	Familiar	32	6.3259	.58370	.10319
	Unfamiliar	32	5.1844	.43470	.07684
P3	Familiar	32	5.5244	.53729	.09498
	Unfamiliar	32	5.0734	.46444	.08210
P5	Familiar	32	5.8159	.50400	.08909
	Unfamiliar	32	5.2994	.42477	.07509

**Table 7.** Group statistics for the variables *EM, P3* and *P5*.

### 3.3.4 Chromatic Intervals in Unfamiliar Melodies and Emotions

A certain assumption that arises from the previous tests is that chromatic closures in unfamiliar melodies are more likely to cause negative feelings. In this section, only chromatic intervals of unfamiliar melodies (*E, F, G* and *H*) are considered. Chromatic closures (cases 1,3,5 and 7) of each melody are considered as separate intervallic groups. The realized intervals for each case can be seen in Table 8.

The hypothesis to be tested is:

*H4a. There is no significant difference at the average emotional arousal because of the effect of the four different chromatic realizations.*

This hypothesis is tested separately for each unfamiliar melody together with the null hypotheses *H4b ... H4f*. Beginning with melody *E*, a single-factor ANOVA test will be conducted, since the data is normally distributed. The dependent variables are *EM, P1, P2, ..., P5*, while the factor is the intervallic group of melody *E* (first column in Table 8).

	E	F	G	H
1	semitone(Ab)	1½ tones(Gb)	semitone(Ab)	semitone(F#)
3	4 tones(Eb)	4 tones(B)	4½ tones(E)	4 tones(C#)
5	1½ tones(Gb)	1 tone(Db)	semitone(Gb)	1 tone(D#)
7	5½ tones(Ab)	4½ tones(Gb)	5½ tones(Ab)	4½ tones(G#)

**Table 8.** Chromatic realized intervals in unfamiliar melodies.

		Sum of Squares	df	Mean Square	F	Sig.
EM	Between Groups	60.028	3	20.009	4.369	.006
	Within Groups	476.296	104	4.580		
	Total	536.324	107			
P1	Between Groups	27.444	3	9.148	4.582	.005
	Within Groups	207.630	104	1.996		
	Total	235.074	107			
P2	Between Groups	30.000	3	10.000	5.474	.002
	Within Groups	190.000	104	1.827		
	Total	220.000	107			
P3	Between Groups	2.074	3	.691	.301	.824
	Within Groups	238.593	104	2.294		
	Total	240.667	107			
P4	Between Groups	28.556	3	9.519	2.934	.037
	Within Groups	337.407	104	3.244		
	Total	365.963	107			
P5	Between Groups	3.000	3	1.000	.482	.696
	Within Groups	215.852	104	2.075		
	Total	218.852	107			

**Table 9.** The ANOVA result for chromatic intervals in melody E.

As the ANOVA result (Table 9) suggests that the significance value comparing the groups is  $<.05$  for  $EM$ ,  $P1, P2$  and  $P4$ , so the null hypotheses  $H4a$ ,  $H4b$ ,  $H4c$  and  $H4e$  are rejected. However,  $H4d$  ( $p=0.824$ ) and  $H4f$  ( $p=0.696$ ) are accepted. Therefore, *there is no significant difference at the 'frustration-calmness' and 'anger-tranquility' dimensions because of the effect of the four different chromatic realizations.*

The test of homogeneity of variances showed that the variances of the groups for all six variables are not significantly different ( $p>0.05$ ). Therefore, the post-hoc

Turkey test (which relies on homogeneity of variance) can be used. Table 10 shows only the significant mean differences that resulted from the test.

*Multiple Comparisons – Turkey HSD*

	(I) GROUP Dependent Variable	(J) GROUP FOR MELODY E	Mean Difference (I-J)	Std. Error	Sig.
<i>EM</i>	1.5 Tones	4 Tones	-1.926*	.582	.007
		5.5 Tones	-1.704*	.582	.022
<i>P1</i>	1.5 Tones	Semitone	-1.148*	.385	.018
		4 Tones	-1.222*	.385	.010
		5.5 Tones	-1.111*	.385	.024
<i>P2</i>	1.5 Tones	Semitone	-1.000*	.368	.038
		4 Tones	-1.444*	.368	.001
		5.5 Tones	-.963*	.368	.049
<i>P4</i>	1.5 Tones	4 Tones	-1.333*	.490	.038

\*. The mean difference is significant at the 0.05 level.

**Table 10.** Post-hoc Turkey test for chromatic intervals in melody E.

The results of table 10 suggest that the chromatic closure of 1½ tone (namely A – Gb) produces significantly different results in emotional variables  $EM$ ,  $P1$ ,  $P2$  and  $P4$  from the other chromatic closures. In particular:

*There is a significant difference between the emotional arousal caused by the chromatic closure of 1½ tone and the chromatic closures of 4 and 5½ tones,  $F(3,104)=4.37, p<.05$ .*

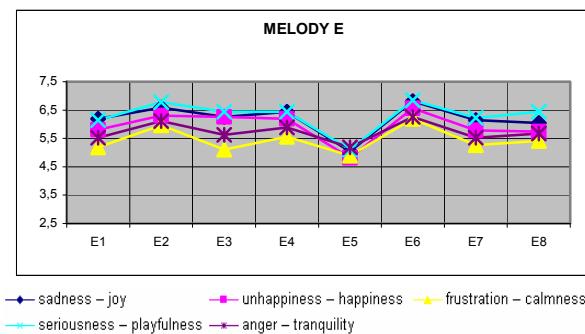
*There is a significant difference between the 'sadness-joy' and the 'unhappiness-happiness' dimensions caused by the chromatic closure of 1½ tone and the other three chromatic closures, [ $F(3,104)=4.58, p<.05$ ,  $F(3,104)=5.47, p<.05$ ].*

*There is a significant difference between the 'seriousness-playfulness' dimension caused by the chromatic closure of 1½ tone and the chromatic closure of 4 tones,  $F(3,104)=2.93, p<.05$ .*

In all of these cases, the scoring average of the chromatic closure of 1½ tone was lower than the other closures, leading to the conclusion that this kind of chromatic closure causes negative feelings to the audience. The average for emotional arousal can be seen in Figure 3 (instance of melody E), while the other dimensions are shown in Figure 4.

Testing the same hypotheses in melody G, another significant difference was discovered in variables  $P3$  and  $P5$ . In particular: *there is a significant difference between the 'frustration-calmness' and the 'anger-*

*tranquility' dimensions caused by the chromatic closure of semitone and the chromatic closure of 4½ tones, [F(3,104)=3.54, p<.05, F(3,104)=3.73, p<.05].*



**Figure 4.** Average ratings of the emotional dimensions for melody E. Case E5 tends to negative responses for all dimensions.

Table 11 shows only the significant differences that resulted from the post-hoc Turkey test (since there is homogeneity of variances). No other significant differences were detected for melodies F and H.

Dependent Variable	(I) GROUP FOR	(J) GROUP FOR	Mean Difference (I-J)	Std. Error	Sig.
	MELODY G	MELODY G			
P3	Semitone	4.5 Tones	1.074*	.349	.014
P5	Semitone	4.5 Tones	1.185*	.363	.008

**Table 11.** Post-hoc Turkey test for chromatic intervals in melody G.

#### 4. CONCLUSIONS

Some useful conclusions for the mapping of emotional parameters to melodic elements (and their proper semantic visualization) have been extracted from the presented experimental analysis. The results seem to confirm scientifically why the atonal music in the beginning of the 20th century spurred such violent reactions. Frustration and anger seem in fact to be the most relevant emotions elicited by chromaticism and unfamiliar music.

More particularly, the findings showed that: the audience is more emotionally aroused when they listen to familiar melodies, especially those written in minor key. It is easier for the unfamiliar melodies (that is the melodies that one listens to for the first time) to cause anger and frustration to the audience, especially if they contain chromatic closures. Melodies in minor key tend to cause more negative feelings, in contrast to melodies in major key that cause positive feelings.

An important finding that concerns the concept of chromaticism in music came to light after the statistical analysis of the only unfamiliar melody in major key (out of the stimuli) where the chromatic interval of 1½ tone has been tested as a possible closure. The interval of 1½ tone caused more intense (and negative) feelings than the other chromatic closures. This finding comes in agreement to the basic theory of chromaticism, on which the MEL-IRIS system for visualizing music with colors is based [10].

It should be stressed that these results are considered within the cultural context of European participants, who are exposed and trained in Western music.

#### 5. REFERENCES

- [1] F. Birren: *Color and human response*, Van Nostrand Reinhold, New York, 1978
- [2] R. Bresin: "What is the Color of that Music Performance?", *Proceedings of the International Computer Music Conference*, Barcelona, pp. 367-370, 2005.
- [3] D. Cooke: *The language of music*, Oxford University Press, London, 1959.
- [4] R. E., Cytowic: *The man who tasted shapes*, Putnam, New York, 1993.
- [5] J. Daltrozzo, B. Tillmann, H. Platel, and D. Schön: "Temporal Aspects of the Feeling of Familiarity for Music and the Emergence of Conceptual Processing", *Journal of Cognitive Neuroscience*, Vol. 22, No. 8, pp. 1754–1769, 2010.
- [6] D. Margounakis and D. Politis: "Converting Images to Music Using their Colour Properties", *Proceedings of the International Conference on Auditory Display*, London, pp. 198-205, 2006.
- [7] L. B. Meyer: *Emotion and meaning in music*, University of Chicago Press, Chicago, 1956.
- [8] E. Narmour: *The analysis and cognition of basic melodic structures*, University of Chicago Press, Chicago, 1990.
- [9] S. E. Palmer, T. Langlois, T. Tsang, K. B. Schloss and D. J. Levitin: "Color, Music and Emotion", *Journal of Vision*, Vol. 11, No.11, 2011.
- [10] D. Politis and D. Margounakis: "Modeling Musical Chromaticism: The Algebra of Cross-Cultural Music Perception", *International Journal of Academic Research*, Vol. 2, No. 6, pp.20-29, 2010.
- [11] E. G. Schellenberg, K. T. Purdy, M. Adachi, and M. C. McKinnon: "Expectancy in Melody: Tests of Children and Adults", *Journal of Experimental Psychology*, Vol. 131, No. 4, pp. 511–537, 2002.
- [12] P. Valdez and A. Mehrabian,: "Effects of color on emotions", *Journal of Experimental Psychology*, Vol. 123, No. 4, pp. 394-409, 1994.
- [13] G. D. Webster, and C. G. Weir: "Emotional Responses to Music: Interactive Effects of Mode, Texture, and Tempo", *Motivation and Emotion*, Vol. 29, No. 1, pp. 19–39, 2005.
- [14] M. R. Zentner: "Preferences for colors and color-emotion combinations in early childhood", *Developmental Science*, Vol.4, No.4, pp. 389-398, 2001.

of the 19th CIM, Trieste, November 21-24,

Poster  
*Posters*



# Cinquant'anni di dislocazione temporale acustica

**Massimo Grassi**

Università di Padova

massimo.grassi@unipd.it

**Guillaume Lemaitre**

Università Iuav di Venezia

glemaître@iuav.it

**Davide Rocchesso**

Università Iuav di Venezia

roc@iuav.it

## ABSTRACT

Nel 1962, in un breve articolo apparso sulla Rivista di Psicologia, Giovanni Bruno Vicario rivelava il suo piano sperimentale teso a chiarire le condizioni per l'insorgenza di inversioni nell'ordine temporale di eventi acustici. Un anno dopo i risultati di questa ricerca venivano pubblicati nella stessa rivista e inquadriati all'interno della teoria della psicologia strutturale di Benussi. Dopo aver guardato ad alcuni sviluppi derivati dal problema della dislocazione temporale negli ultimi cinquant'anni, proponiamo un paio di estensioni sperimentali, una facente uso di stimoli acustici ecologici, e l'altra impiegante stimoli dicotici per verificare l'eventuale insorgenza di uno *spatial agency bias*.

## 1. INTRODUZIONE

All'inizio degli anni sessanta del secolo scorso, come assistente alla cattedra di Gaetano Kanizsa a Trieste, Giovanni Bruno Vicario eseguiva alcuni esperimenti di grande importanza per la comprensione degli eventi acustici e per la costruzione di una psicologia del tempo [10, 11]. Egli scopri l'effetto tunnel acustico [6] e indagava in profondità il fenomeno della dislocazione temporale [7, 8]. E' su quest'ultimo fenomeno che vogliamo soffermarci in questo contributo, descrivendo alcuni sviluppi di ricerca che attorno ad esso si sono registrati e portando, se possibile, un contributo originale.

Per le sue indagini sulla dislocazione temporale, cioè il disallineamento del tempo fenomenico dal tempo fisico con la possibile sovversione della relazione d'ordine tra diversi eventi fisici, Vicario usò due tipi di stimoli: (i) una successione di due note interrotta da un *burst* di rumore, oppure (ii) una successione di due note interrotta da una nota di altezza significativamente diversa dalle altre due. La sperimentazione più estesa ed accurata si svolse con gli stimoli di tipo (ii), ed è ben descritta in [8].

Gli esperimenti di Vicario anticiparono di molti anni quelli di Bregman e Campbell [1] che, pur se inseriti nel paradigma di *streaming*, di fatto ridimostrarono e rispiegarono l'incapacità, in certe condizioni, di attribuire una corretta relazione d'ordine a sequenze di stimoli. Gli stessi Bregman e Campbell scrivevano di *recent finding* riferendosi ad un lavoro pubblicato su *Science* da Warren et al

*Copyright: ©2012 Massimo Grassi 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.*

nel 1969 [12], nel quale si riportava l'incapacità di ordinare correttamente sequenze di brevi stimoli diversi da voce o musica. Peraltra, il paradigma di *streaming* era già stato descritto dallo stesso Vicario negli studi condotti con Paolo Bozzi negli anni sessanta del novecento [9].

Vicario riprenderà i suoi studi sulla dislocazione temporale all'inizio degli anni novanta nell'ambito della tesi di laurea di Lisa Trasforini. In questo caso gli stimoli furono generati usando la workstation IRCAM ISPW e fu possibile misurare il tempo di reazione motoria. Si constatò che la risposta motoria al rumore avveniva con un ritardo indipendente dalla presenza della sequenza di due toni. Invece, il rumore veniva sistematicamente riportato come dislocato dalla sua effettiva posizione nella sequenza [5, 11, 4].

## 2. FENOMENI

L'ipotesi di lavoro, verificata da Vicario, è che all'interno di un breve intervallo temporale, e in dipendenza dalle caratteristiche degli eventi sonori, non si riesca a percepire una relazione d'ordine fra gli elementi di una sequenza, ma invece si acquisisca una struttura (*gestalt*) complessiva. In una sequenza costituita da due note gravi interrotte da un tratto di rumore, la vicinanza tonale delle due note gravi tende a formare una unità strutturale, percepita come separata del tratto di rumore. Quindi, la struttura percepita non è un'successione nota - rumore - nota, bensì rumore - nota - nota o nota - nota - rumore. Vicario constatava che, in tal caso, il tratto di rumore è quasi sistematicamente percepito *dopo* la struttura delle note gravi (dislocazione positiva del rumore), e che quindi avviene una ristrutturazione percettiva dello stimolo fisico.

In questo lavoro indaghiamo due effetti che potrebbero influenzare la percezione della struttura temporale di una successione di elementi brevi: le relazioni di causalità tra gli elementi, e lo *spatial agency bias*. La prima ipotesi è semplice. Per una sequenza di tre elementi brevi A1-B-A2, la struttura percepita più frequentemente è A1-B-A2 allorché ci sia una relazione di causalità del tipo A1 → B → A2. Immaginiamo ad esempio una palla che rotola su una superficie (primo suono di rotolamento A1), la quale si scontra con una seconda palla (suono di impatto B), che si mette a rotolare (secondo suono di rotolamento A2). In assenza di causalità tra gli elementi gli esperimenti di Vicario hanno mostrato che, viceversa, si percepisce più spesso la sequenza A1-A2-B. In altre parole si fa l'ipotesi che la causalità annulli la dislocazione temporale del suono centrale della sequenza, inserendolo solidamente tra i due tratti di rotolamento.

La seconda ipotesi è basata sul fenomeno di *spatial agency bias*. Questo fenomeno si presenta in circostanze diverse, ad esempio quando i soggetti immaginano o ricordano eventi collocati nello spazio. Risulta prevalente una direzione da sinistra verso destra, indipendentemente del vero verso del movimento [2, 3]. Con una sequenza A1-B-A2, per la quale Vicario riscontrò un egual numero di dislocazioni negative e positive dell'elemento B, si verifica se l'assegnazione degli elementi ai diversi canali sinistro e destro provoca una rottura della simmetria e il prevalere di un segno nelle dislocazioni temporali. Vicario ha mostrato come l'ordine temporale attribuito a una sequenza di note grave - acuto - grave sia ambiguo in certi casi. Intendiamo usare questo tipo di stimolo come *baseline* e studiare l'effetto della riproduzione dei due tipi di eventi sui canali uditivi destro e sinistro.

### 3. IMPIANTO Sperimentale

Per portare un contributo alla ricerca sulla dislocazione temporale acustica vogliamo verificare le due ipotesi:

1. Usando stimoli ecologici che possano dar luogo a relazioni di causalità la dislocazione temporale può essere diversa o sparire;
2. Facendo presentazioni dicotiche, in cui gli stimoli sono segregati alle due orecchie, la dislocazione temporale può essere diversa o sparire. In particolare, si può rivelare una asimmetria nella dislocazione temporale (*spatial agency bias*).

Per la costruzione degli stimoli per la verifica della ipotesi (1) si usano due brevi brani di suono di sintesi di rotolamento, uno con palla grande e uno con palla piccola. Lo stimolo di riferimento è una sequenza tono-rumore-tono per la quale Vicario ha riscontrato in maniera inequivocabile la dislocazione del rumore. Bisogna assicurarsi del fatto che, per durate intorno ai 100ms un suono di rotolamento mantenga la sua natura percepita, e questo risulta più facile per suoni di sintesi. Il terzo suono è un semplice impatto, questa volta estratto dalla registrazione di un impatto tra due palle di legno. La composizione dello stimolo, ottenuta per giustapposizione di suoni sintetici (rotolamento) e registrati (impatto) offre un compromesso tra naturalezza dei suoni e chiarezza percettiva del fenomeno meccanico. Verifichiamo se, con questo tipo di stimolo composto da segmenti in relazione causale tra loro, viene compromessa la dislocazione temporale del rumore/impatto.

Per la verifica della ipotesi (2) usiamo una sequenza di note grave - acuto - grave per la quale Vicario ha verificato una distribuzione simmetrica di dislocazioni positive e negative. Si verificherà se la segregazione sui canali audio introduce un bias di dislocazione.

In una ultima fase, si proverà l'effetto della distribuzione spaziale di stimoli ecologici, per vedere se essa rafforza o riduce il bias di dislocazione temporale. In effetti, in una sequenza rotolamento - impatto - rotolamento si può facilmente individuare un agente attivo e un agente passivo e, quindi, si potrebbe riscontrare la predilezione per uno sviluppo causale dell'azione da sinistra a destra.

### 4. CONCLUSIONE

Abbiamo ripreso alcuni esperimenti storici di Vicario sulla percezione dell'ordine temporale in successioni di eventi. A cinquant'anni di distanza dalle prime sperimentazioni, il fenomeno della dislocazione temporale acustica è ancora di grande interesse per psicologi e sound designer, e si sente la necessità di indagarlo meglio, usando un più ampio repertorio di stimoli e diverse modalità di fruizione. Le ipotesi qui presentate, relative all'introduzione della causalità e della segregazione spaziale, sono al momento oggetto di indagine sperimentale. I primi risultati saranno presentati al XIX CIM e i partecipanti al Colloquio avranno modo di ascoltare gli stimoli e contribuire agli esperimenti.

### 5. BIBLIOGRAFIA

- [1] Albert S. Bregman e Jeffrey Campbell. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2):244–249, 1971.
- [2] Andrea R. Halpern e Michael H. Kelly. Memory biases in left versus right implied motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2):471–484, march 1993.
- [3] Anne Maass, Caterina Suitner, Xenia Favaretto, e Marina Cignacchi. Groups in space: Stereotypes and the spatial agency bias. *Journal of Experimental Social Psychology*, 45(3):496 – 504, 2009.
- [4] Giorgio Marchetti. Commentary on Giovanni Bruno Vicario's "Il tempo. Saggio di psicologia sperimentale". 2006.
- [5] Lisa Trasforini. Dislocazione temporale: tempi di reazione e resoconti soggettivi, 1996. Tesi di Laurea, Facoltà di Psicologia, Università di Padova.
- [6] Giovanni Vicario. L'effetto tunnel acustico. *Rivista di Psicologia*, 54(2):41–52, 1960.
- [7] Giovanni Vicario. Alcune osservazioni sperimentali sulla dislocazione temporale di stimoli acustici. *Rivista di Psicologia*, 56:268–273, 1962.
- [8] Giovanni Vicario. La dislocazione temporale nella percezione di successioni di stimoli discreti. *Rivista di Psicologia*, 57(1):17–87, 1963.
- [9] Giovanni Vicario. Vicinanza spaziale e vicinanza temporale nella segregazione degli eventi. *Rivista di Psicologia*, 59(4):843–863, 1965.
- [10] Giovanni Bruno Vicario. Il tempo in psicologia. *Le Scienze*, 30(347):43–51, 1997.
- [11] Giovanni Bruno Vicario. *Il tempo. Saggio di psicologia sperimentale*. Il Mulino, Bologna, Italy, 2005.
- [12] Richard M. Warren, Charles J. Obusek, Richard M. Farmer, e Roslyn P. Warren. Auditory sequence: Confusion of patterns other than speech or music. *Science*, 164(3879):586–587, maggio 1969.

# CEPSTRUM ANALYSIS AND BLIND SYSTEM IDENTIFICATION FOR LOCALIZATION OF PSEUDO-PERIODIC SOUNDS

Daniele Salvati

AVIRES Lab.

Dep. of Mathematics and Computer Science  
University of Udine, Italy  
daniele.salvati@uniud.it

Sergio Canazza

Sound and Music Computing Group  
Dep. of Information Engineering  
University of Padova, Italy  
canazza@dei.unipd.it

## ABSTRACT

Blind System Identification (BSI) focuses on the impulse responses between the source and the microphones to estimate the Time Difference Of Arrival (TDOA) of an acoustic source in reverberant environments. Considering the Adaptive Eigenvalue Decomposition (AED) method based on Normalized MultiChannel Frequency Least Mean Square (NMCFLMS) with sparse priors imposition, this paper shows that the use of cepstrum analysis of microphone pair signals allows to locate pseudo-periodic sounds by selecting the appropriate frequency bins, reducing the effect of reverberant ambient noise that covers most of the spectral range. Experimental results of a thirteen sinusoidal components wave in different reverberant conditions are reported.

## 1. INTRODUCTION

The aim of an acoustic source localization system is to estimate the position of sound sources in space by analyzing the sound field with a microphone array, a set of microphones arranged to capture the spatial information of sound. The AED [1] is a time delay estimation method based on the BSI, which focuses on the impulse responses between the source and the microphones. The AED takes the reverberation fully into account with the advantages to obtain a better performance under highly reverberant conditions. It can be efficiently implemented with the NMCFLMS with sparse priors imposition [2]. However, when the source of interest is harmonic, or generally pseudo-periodic, the NMCFLMS converges to an incorrect value because reverberant ambient noise covers most of the spectral range.

This papers proposes the use of cepstrum spectrum to select the frequency bins of pseudo-periodic signals, and to consider only these spectrum components in NMCFLMS filter to successfully obtain the TDOA estimation of a microphone pair. The interest in locating pseudo-periodic sounds using microphone arrays may be attractive for human-computer interaction in musical applications. It has been shown in [3, 4] that there is some potential in using the

*Copyright: ©2012 Daniele Salvati 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.*

sound source localization to directly control an audio processing by moving the sound produced by its own musical instrument.

## 2. BSI CEPSTRUM ALGORITHM

AED [1] is a BSI estimation method for time delay between a microphone pair based on a reverberant model using eigenvalue decomposition. Introducing the channel impulse response  $g_m$  from the source to microphone  $m$ , we can express the reverberant model as

$$x_m(k) = g_m * s(k - k_s - \tau_m) + v_m(k) \quad (1)$$

where  $*$  denotes convolution,  $s(k)$  is the unknown uncorrelated source signal,  $k$  is the sample time index,  $k_s$  is the propagation time from the unknown source to the reference sensor,  $\tau_m$  is the TDOA of the signal between the  $m$ th microphone and the reference of array, and  $v_m(k)$  is the additive noise signal at the  $m$ th sensor, assumed to be uncorrelated with not only all of the source signals but also with the noise observed at the other sensors. AED assumes that the room is linear and time invariant; therefore, neglecting the influence of noise we can write

$$x_1 * g_2 = g_1 * s * g_2 = x_2 * g_1. \quad (2)$$

The vectors of the signal samples at the microphone outputs and the impulse response vectors of length  $L$  can be expressed as

$$\mathbf{x}_i(k) = [x_i(k), x_i(k-1), \dots, x_i(k-L+1)]^T \quad (3)$$

and

$$\mathbf{g}_i = [g_i(0), g_i(1), \dots, g_i(L-1)]^T. \quad (4)$$

Substituting the vectors, equation (2) becomes

$$\mathbf{x}_1^T \mathbf{g}_2 - \mathbf{x}_2^T \mathbf{g}_1 = 0. \quad (5)$$

Then, introducing the correlation matrix, we can define the following equation in matrix notation from (5)

$$\mathbf{R}\mathbf{u} = \mathbf{0} \quad (6)$$

where  $\mathbf{R}$  is the correlation matrix and  $\mathbf{u}$  is a  $2M \times 1$  vector formed by the juxtaposition of the two impulse responses

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} \mathbf{R}_{x_1 x_1} & \mathbf{R}_{x_1 x_2} \\ \mathbf{R}_{x_2 x_1} & \mathbf{R}_{x_2 x_2} \end{bmatrix} \\ \mathbf{u} &= [\mathbf{g}_2 - \mathbf{g}_1]^T \end{aligned}$$

and  $\mathbf{R}_{x_i x_j} = E[\mathbf{x}_i(k) \mathbf{x}_j^T(k)]$ . The equation (6) implies that  $\mathbf{u}$  is the eigenvector corresponding to the zero-valued eigenvalue of  $\mathbf{R}$ . If the two impulse responses have no common zeros and the autocorrelation matrix of  $s(n)$  is full rank, there is only a single zero-valued eigenvalue. In practice, only estimate of the sample correlation matrix is available, and, therefore, instead of the zero-valued eigenvalue, we search for the minimum eigenvalue of  $\mathbf{R}$  using the iterative method. To speed up the convergence and to achieve efficient implementation, a NMCFLMS can be used [2]. Finally, the time delay estimation is

$$\hat{\tau}_{12}^{AED} = \operatorname{argmax}_l \hat{g}_{1,l} - \operatorname{argmax}_l \hat{g}_{2,l}. \quad (7)$$

An improved AED method [2] imposes sparse priors on the responses to reduce the temporal whitening and provide a more accurate and robust time delay estimation.

The performance of the AED technique is dramatically reduced in the case of pseudo-periodic sounds because reverberant ambient noise covers most of the spectral range, and consequently the NMCFLMS converges to an incorrect value of TDOA.

Cepstrum [5] analysis provides to select the spectral components to be used in the NMCFLMS filter. The cepstrum is calculated from the FFT by taking the real logarithm and performing an IFFT, which leads to the real cepstrum

$$c(f) = \frac{1}{L} \sum_{f=0}^{L-1} \log \left| \sum_{k=0}^{L-1} x(k) e^{-\frac{2\pi j f k}{L}} \right| e^{\frac{2\pi j f k}{L}}. \quad (8)$$

The local maxima of cepstrum spectrum are the selected frequency bins of signal to be used in the NMCFLMS filter, and considering that the other values, i.e., those not selected by the cepstrum analysis, of FFT are assumed to be zero.

### 3. EXPERIMENTAL RESULTS

A thirteen sinusoidal components wave is used in experiments to test the proposed BSI cepstrum algorithm in different reverberation conditions. AED method based on NMCFLMS with sparse priors imposition and an Hann window analysis of 2048 samples at 44.1 kHz are used. The first six cepstrum components are selected to evaluate the NMCFLMS performance. The Matlab code that implements the image-source method developed by E. A. Lehmann is used to simulate reverberant audio data in room acoustics [6]. A room of  $(3.5 \times 4.5 \times 3)$  meters size is considered. The position of microphones are  $(1.4, 1, 1.3)$  m and  $(1.6, 1, 1.3)$  m, and the source is  $(2.5, 3, 1.3)$  m. In this case the acoustic source reaches the microphones with a TDOA of -11 samples. Table 1 summarizes the TDOAs estimated by NMCFLMS with sparse priors imposition, comparing the results without and with cepstrum analysis, considering different number of cepstrum components. In all reverberant conditions the NMCFLMS without cepstrum analysis converges to the an incorrect value of TDOA, while the NMCFLMS with cepstrum properly works to a minimum of five frequency components.

<b>RT60 (s)</b>	<b>NMCFLMS without cepstrum</b>	<b>NMCFLMS with cepstrum</b>			
		3	4	5	6
0.25	2	-12	-14	-11	-11
0.5	0	-10	-12	-11	-11
1.0	0	-10	-13	-11	-11

**Table 1.** The comparison of TDOA estimation (values are in samples). The cepstrum NMCFLMS is evaluated with different numbers of frequency components.

### 4. CONCLUSIONS

The use of cepstrum analysis in the AED method based on NMCFLMS with sparse priors imposition allows to estimate TDOA of harmonics, or generally pseudo-periodic, sounds in reverberant environments. Some experimental results with a thirteen sinusoidal components wave in different reverberant conditions show that the selection of frequency component bins, obtained by means of cepstrum spectrum and picking the local maxima, reduces the influence of reverberant ambient noise that covers most of spectral range. The proposed BSI cepstrum algorithm converges to the correct TDOA value with a minimum of five frequency components up to RT60 of 1 s.

### 5. REFERENCES

- [1] J. Benesty, “Adaptive eigenvalue decomposition algorithm for passive acoustic source localization,” *Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 384–391, 2000.
- [2] J. Cho and H. Park, “Imposition of sparse priors in adaptive time delay estimation for speaker localization in reverberant environments,” *IEEE Signal Processing Letters*, vol. 16, no. 3, pp. 180–183, 2009.
- [3] D. Salvati, S. Canazza, and A. Rodà, “Sound spatialization control by means of acoustic source localization system,” in *Proceedings of the 8th Sound and Music Computing Conference*, pp. 284–289, 2011.
- [4] D. Salvati, S. Canazza, and A. Rodà, “A sound localization based interface for real-time control of audio processing,” in *Proceedings of the 14th International Conference on Digital Audio Effects*, pp. 177–184, 2011.
- [5] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, “The quefrency alalysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking,” in *Proceedings of the Symposium on Time Series Analysis*, pp. 209–243, 1963.
- [6] E. Lehmann and A. Johansson, “Prediction of energy decay in room impulse responses simulated with an image-source model,” *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, 2008.

# Model-based customized binaural reproduction through headphones

**Michele Geronazzo**  
University of Padova

**Simone Spagnol, Davide Rocchesso**  
IUAV - University of Venice

**Federico Avanzini**  
University of Padova

## ABSTRACT

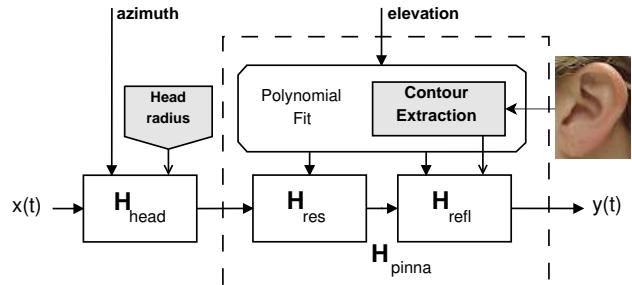
Generalized head-related transfer functions (HRTFs) represent a cheap and straightforward mean of providing 3D rendering in headphone reproduction. However, they are known to produce evident sound localization errors, including incorrect perception of elevation, front-back reversals, and lack of externalization, especially when head tracking is not utilized in the reproduction. Therefore, individual anthropometric features have a key role in characterizing HRTFs. On the other hand, HRTF measurements on a significant number of subjects are both expensive and inconvenient. This short paper briefly presents a structural HRTF model that, if properly rendered through a proposed hardware (wireless headphones augmented with motion and vision sensors), can be used for an efficient and immersive sound reproduction. Special care is reserved to the contribution of the external ear to the HRTF: data and results collected to date by the authors allow parametrization of the model according to individual anthropometric data, which in turn can be automatically estimated through straightforward image analysis. The proposed hardware and software can be used to render scenes with multiple audiovisual objects in a number of contexts such as computer games, cinema, edutainment, and many others.

## 1. A CUSTOMIZABLE HRTF MODEL

There is no doubt that, if we set the direction of the sound source with respect to the listener, the greatest dissimilarities among different people's HRTFs are due to the massive subject-to-subject pinna shape variation. The external ear plays an important role by introducing peaks and notches in the high-frequency spectrum of the HRTF, whose center frequency, amplitude, and bandwidth greatly depend on the elevation angle of the sound source. The relative importance of major peaks and notches in typical HRTFs in elevation perception has been disputed over the past years; in general, both seem to play an important function in vertical localization of a sound source.

However, in a previous work [1] we highlighted that while the resonant component of the pinna-related counterpart of the HRTF (known as PRTF) is in broad terms similar among different subjects, the reflective component of the PRTF comes along critically subject-dependent. In the same work, we exploited a simple ray-tracing law to

*Copyright: ©2012 Michele Geronazzo 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.*



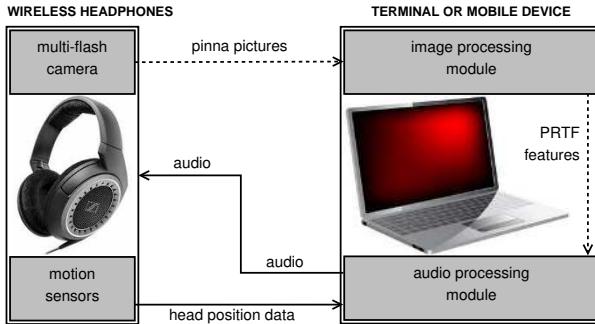
**Figure 1.** The structural HRTF model.  $H_{head}$  is the head filter;  $H_{res}$  and  $H_{refl}$  are the resonant and reflective components of the pinna model  $H_{pinna}$ , respectively.

show that in median-plane frontal HRTFs (with elevation ranging from  $\phi = -45^\circ$  to  $\phi = 45^\circ$ ) the frequency of the spectral notches, each assumed to be caused by its own reflection path, is related to the shape of the concha, helix, and antihelix on the frontal side of the median plane at least. This finding opens the door for a very attractive approach to the parametrization of the HRTF based on individual anthropometry, that is, extrapolating the most relevant parameters that characterize the PRTF just from a 2-D representation of the user's pinna.

Following this important result, a complete structural HRTF model that takes into account the user's anthropometry was proposed in [2] and is schematically depicted in Figure 1. In the model, elevation and azimuth cues are handled orthogonally: vertical control is associated with the acoustic effects of the pinna ( $H_{pinna}$ ) while the horizontal one is delegated to head diffraction ( $H_{head}$ ). The model is designed so as to avoid expensive computational and temporal steps, allowing implementation and evaluation in a real-time audio processing environment. Two instances of such model, appropriately synchronized through interaural time delay estimation methods, allow for real-time binaural rendering.

The core of the above structure is the pinna model: here two second-order peak filters (filter structure  $H_{res}$ ) and three second-order notch filters (filter structure  $H_{refl}$ ) synthesize two resonance modes and three pinna reflections respectively, with the associated parameters either derived from the subject's anthropometry or taken from average measurements on a group of subjects. Obviously, extracting the relevant features from a picture implies a mandatory image processing step. The clearest contours of the pinna and the ear canal entrance must be recognized in order to calculate distances between reflection and observation points and convert them to notch frequencies.

Intensity edge detection techniques applied to a single



**Figure 2.** The system's architecture and software.

picture of the pinna are hardly applicable. This task can be instead achieved through a technique known as multi-flash imaging [3]: by using a camera with multiple flashes strategically positioned to cast shadows along depth discontinuities in the scene, the projective-geometric relationship of the setup can be exploited to detect depth discontinuities (in our case, pinna contours) and distinguish them from intensity edges due to color discontinuities.

## 2. SYSTEM ARCHITECTURE

In order to fully exploit the potential of a customizable HRTF model in both static and dynamic listening scenarios, an appropriate audio device equipped with sensors able to detect the relevant parameters needed to fine tune the model both before and during listening has to be designed. Our idea is that a couple of common wireless headphones augmented through motion sensors and possibly a multi-flash camera could easily fit the goal. Figure 2 illustrates the architecture of the wireless system we are currently realizing, including all of the data exchanges between the headphones and a master device running both the image and audio processing software.

The headphones incorporate a number of sensors (a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis digital compass) able to continuously track the user's head pose in the 3D space thanks to the full 6-DoF motion processing they convey [4]. Moreover, a digital camera equipped with four bright LEDs acting as flash lights and positioned around the camera eye can be slotted inside one of the two cups of the headphones depending on the available space, both inside the cup and between the lens and the ear of the user wearing the headphones. Alternatively, the multi-flash camera shall be proposed as a separate wearable or interchangeable device. Synchronization between each of the flash lights and the related shot is managed by a simple microcontroller.

Storage of the resulting pictures and transmission to the master device is managed through a wireless SD card. The pictures will be received by an image processing program performing the following steps:

- *depth edge detection*: based on the available pictures and their relative differences in shadow and lighting, a depth edge map is computed through the algorithm proposed in [3];

- *pinna contour recognition*: the most prominent contours are extracted among the available depth edges based on both their shape and length and the consistency between their relative positions, and stored as a specifically designed data format;
- *ear canal detection*: the ear canal entrance is approximated by one specific point of the tragus edge;
- *computation of pinna-related features*: distances between the extracted contours and the ear canal entrance are translated into notch frequency parameters through straightforward trigonometric computation, approximated as functions of the elevation angle, and fed to the audio processing module.

This last module, implemented in a real-time environment, systematically receives at each time frame the data from the motion sensors (pitch, roll, and yaw rotations of the head) through radio transmission and translates it into a couple of polar coordinates (azimuth, elevation) of a fixed or moving sound source with respect to the user. The couple of coordinates finally represents the input to the structural HRTF model that performs the spatialization of a desired sound file through the user's customized synthetic HRTFs. This way, provided that the center of rotation of the head does not excessively translate during the rotation (distance between the user and the sound source cannot indeed be tracked in real time by the available sensors), the user will perceive the position of the virtual sound source as being independent from his or her movement.

The HRTF model currently includes a large portion of the frontal hemispace and the proposed architecture could thus be suitable for real-time control of virtual sources in a number of applications involving frontal auditory displays, such as a sonified screen. Further extensions of the HRTF model, capable of including source positions behind, above, and below the listener, may be obtained in different ways, and will be objects of future research.

## 3. REFERENCES

- [1] S. Spagnol, M. Geronazzo, and F. Avanzini, “Fitting pinna-related transfer functions to anthropometry for binaural sound rendering,” in *Proc. IEEE Int. Work. Multi. Signal Process. (MMSP’10)*, (Saint-Malo, France), pp. 194–199, October 2010.
- [2] M. Geronazzo, S. Spagnol, and F. Avanzini, “A head-related transfer function model for real-time customized 3-D sound rendering,” in *Proc. INTERPRET Work., SITIS 2011 Conf.*, (Dijon, France), pp. 174–179, November-December 2011.
- [3] R. Raskar, K.-H. Tan, R. Feris, J. Yu, and M. Turk, “Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging,” *ACM Trans. Graphics (Proc. SIGGRAPH)*, vol. 23, no. 3, pp. 679–688, 2004.
- [4] S. Nasiri, S.-H. Lin, D. Sachs, and J. Jiang, “Motion processing: the next breakthrough function in handsets,” 2010.

# EGLO – UN'APPLICAZIONE PER LA SONIFICAZIONE INTERATTIVA DI TAVOLI DI GIOCO PER BAMBINI

Gianluca Dario, Riccardo Mazzetti, Lorenzo Jacopo Sciuga

Università di Udine

Polo Didattico di Pordenone

via Prasecco, 1 – 33070 Pordenone, Italy

{dario.gianluca, mazzetti.riccardo, sciuga.lorenzo}@spes.uniud.it

Federico Fontana

Università di Udine

Dipartimento di Matematica

e Informatica

via delle Scienze, 206 – 33100 Udine, Italy

federico.fontana@uniud.it

## ABSTRACT

*Electronic Game and Learning Object*, brevemente EGLO, è un'applicazione che realizza delle semplici narrative di gioco per bambini in età prescolare con valenza educativa, in linea con l'obiettivo di fornire strumenti nuovi e intuitivi per l'apprendimento. Similmente a quanto avviene in altri sistemi d'interfaccia uomo-macchina, il più affine dei quali è Reactable, EGLO realizza un tavolo di gioco con tessere le cui firme e coordinate bidimensionali dettano a ogni istante la generazione interattiva di suoni contestuali al gioco selezionato. L'interfaccia è appositamente progettata per celare quanto più possibile i sistemi di visione, di calcolo e di generazione dei suoni. Il prototipo, ad oggi pienamente operativo, lascia intravvedere la possibilità di essere adoperato in strutture per l'infanzia con finalità anche d'indagine sperimentale.

## 1. INTRODUZIONE

Nell'ambito della parte pratica e laboratoriale dell'insegnamento di Elaborazione del suono del corso in Comunicazione Multimediale e tecnologie dell'Informazione dell'Università di Udine, è stata messa a punto una realizzazione in grado di sonificare in modo immediato, semplice e intuitivo dei giochi per bimbi di età 3-5 anni in cui i piccoli giocatori, disponendo e spostando individualmente o collettivamente delle tessere poste su di un tavolo di gioco, ricevono *feedback* uditorio come conseguenza delle interazioni eseguite sul tavolo.

Alla base del concetto di design proposto vi è il paradigma consolidato delle Tangible User Interfaces, le quali prevedono un'interazione immediata tra utilizzatore e macchina attraverso la manipolazione di oggetti di uso intuitivo, su cui è talvolta possibile effettuare l'*embodiment* della risposta della macchina attraverso un loro aumento [1]. Interfacce di questo tipo

*Copyright: © 2012 Gianluca Dario 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.*

peraltro hanno già trovato ampio utilizzo in applicazioni per specifiche classi di utenti quali i bambini [2].

Nel caso particolare in cui la risposta del sistema sia di tipo uditorio, il meccanismo d'attuazione è in buona misura dipendente dall'ambito e lo scopo che l'interfaccia si prefigge: si va così da esempi di *sound embodied* in cui prevalgono l'unità e consistenza percettiva in senso multimodale dell'oggetto manipolato [3], ad applicazioni in cui l'oggetto viceversa riveste un ruolo principalmente di controllo in un sistema di sintesi del suono ad elevata accuratezza, in cui la riproduzione è tradizionalmente affidata ad altoparlanti. Reactable [4] appartiene a questo secondo insieme di esempi, e ha svolto un ruolo d'ispirazione fondamentale per il presente progetto di cui saranno chiariti concetto, realizzazione e applicazione nelle successive sezioni.

## 2. CONCETTO

EGLO realizza semplici giochi per bimbi, costruiti attorno alla capacità di organizzare delle tessere sopra un tavolo sulla base di un sistema di regole. La disposizione delle tessere determina il livello di successo nel gioco, il quale è sottolineato da EGLO con un feedback uditorio affermativo quando una o più tessere sono disposte in modo corretto, oppure con un feedback uditorio informativo se la semantica di una tessera dipende dalla posizione, ad esempio quando il gioco preveda di suonare la nota che contraddistingue una tessera. EGLO a ogni istante distingue le tessere e ne legge costantemente la disposizione sul tavolo: le tessere, dunque, definiscono uno stato del gioco a cui corrispondono uno o più suoni, i quali informano il o i giocatori dell'esito del posizionamento.

I diversi giochi sono essi stessi selezionabili attraverso delle tessere speciali, oppure attraverso la giustapposizione sul tavolo di cornici di sfondo che formano parte integrante del gioco. Nel realizzare l'interfaccia si è cercato di minimizzare la presenza di tecnologia visibile: oltre a un pulsante di accensione, EGLO espone all'utente solamente le tessere e gli sfondi dei diversi giochi.

### 3. REALIZZAZIONE

EGLO si presenta come un cubo di legno di circa 40 cm di lato, con la faccia superiore di vetro. L'ispezione dell'interno (Figura 1) rivela la presenza di una webcam a 30 fps, una lampada, una scheda pc provvista d'interfaccia audio, un amplificatore e due *exciter* i quali, attaccati a due opposte facce laterali, ricevendo segnale audio amplificato trasmettono energia meccanica al cubo che, dunque, risuona acusticamente in presenza di opportuno feedback dal sistema.



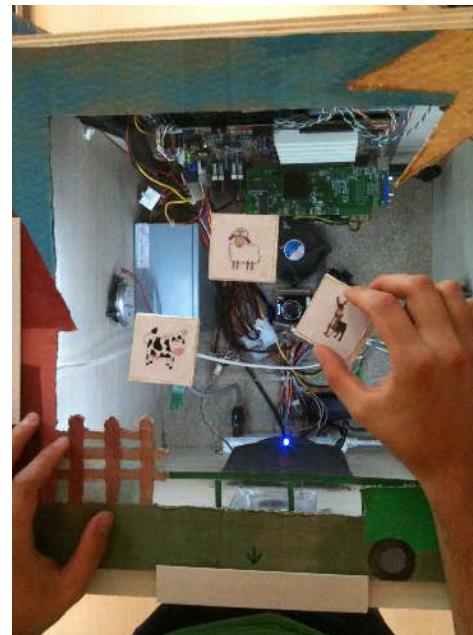
**Figura 1.** L'interno di EGLO: sono ben distinguibili la scheda pc e (sopra l'alimentatore) un *exciter*.

Facilitata dalla presenza della retroilluminazione posta all'interno del cubo, la webcam cattura l'immagine proiettata sulla superficie di vetro contenente tessere e sfondi. Se di fronte le tessere esibiscono disegni collegati al gioco, al contrario sul retro presentano delle firme (*marker*) comprensibili a Reactivision, il software libero di riconoscimento adoperato in EGLO. Reactivision invia dei messaggi TUIO contenenti le coordinate e l'identità di ogni tessera a un'istanza del software Puredata, il quale estrae i dati necessari e infine genera la risposta audio riproducendo suoni pre-caricati oppure controllando piccoli banchi di oscillatori in ampezza e frequenza.

### 4. GIOCHI

Ad oggi sono stati realizzati 5 prototipi di giochi divisibili in due categorie tematiche: "ecologici" e "musicali". Dei giochi ecologici, uno è a tema la raccolta differenziata dei rifiuti e propone un feedback affermativo quando le tessere dei rifiuti sono correttamente giustapposte a quelle dei relativi

contenitori; l'altro conferma che un animale sia stato collocato nel suo corretto habitat (Figura 2).



**Figura 2.** Il gioco *Animali e Habitat*: da notare le tessere e lo sfondo.

Dei giochi musicali, due riguardano la corretta associazione tra uno strumento musicale e la sua famiglia d'appartenenza mediante giustapposizione della tessera corrispondente sul relativo sfondo; il terzo aiuta l'insegnamento delle note, eseguendo suoni di altezza corrispondente alla nota identificata nella tessera. A queste regole si aggiungono funzionalità quali il controllo dell'intensità dei suoni mediante rotazione delle tessere, reso possibile dal riconoscimento dell'orientazione dei marker da parte di Reactivision.

### 5. RIFERIMENTI

- [1] H. Ishii: "Tangible bits: beyond pixels," *Proc. 2nd international conference on tangible and embedded interaction*, Bonn, Germany, 2008.
- [2] J. Marco, S. Baldassarri, E. C. Diana, Y. Xu, J. C. Read: "LIFELONG INTERACTIONS: Let the experts talk: an experience of tangible game design with children," *Interactions*, v. 17 n. 1, pp. 58-61, 2010.
- [3] S. Delle Monache, D. Rocchesso, J. Qi, L. Buechley, A. De Götzen, D. Cestaro: "Paper mechanisms for sonic interaction," *Proc. 6th international conference on tangible, embedded and embodied interaction*, Kingston, Ontario, Canada, 2012.
- [4] [www.reactable.com](http://www.reactable.com).

## Sincresi audio-visiva: un'esperienza didattica

**Elio Toppano**

Dipartimento di Matematica e Informatica  
Università di Udine  
elio.toppano@uniud.it

**Roberto Ranon**

Dipartimento di Matematica e Informatica  
Università di Udine  
roberto.ranon@uniud.it

### ABSTRACT

La sensorialità è un tema sempre più importante nella teoria e pratica della progettazione multimediale. Nonostante ciò gli approcci tradizionali alla realizzazione di artefatti comunicativi tendono a sottovalutare il ruolo e la ricchezza della multimodalità e della intermodalità sensoriale e impediscono lo sviluppo di una comprensione profonda e critica di questi aspetti. Il lavoro presentato in questo poster descrive alcune esperienze didattiche proposte nell'ambito del corso di Sistemi Multimediali e Laboratorio del Corso di Laurea in Tecnologie Web e Multimediali dell'Università di Udine con lo scopo di sviluppare negli studenti una maggiore consapevolezza degli aspetti morfosintattici, percettivi e semantici delle produzioni audiovisive nello spirito e nella tradizione del Basic Design.

### 1. INTRODUZIONE

Negli ultimi anni - nel settore dei sistemi multimediali - si è assistito ad un progressivo spostamento dell'attenzione dalla progettazione di prodotti e servizi alla messa in scena di esperienze [1]. Questo mutamento di prospettiva è dovuto principalmente alla diffusione di applicazioni multimediali e interattive al di fuori degli ambiti tradizionali di lavoro (per l'istruzione, il divertimento, l'arte) dove è necessario considerare, accanto ai tradizionali requisiti tecnico-funzionali basati sulla utilità e sulla usabilità, anche aspetti edonici legati al piacere estetico, alle emozioni, al divertimento, alla seduzione, ecc. [2, 3].

La necessità di sviluppare una consapevolezza verso queste tematiche ha portato a proporre - nell'ambito del corso di Sistemi Multimediali e Laboratorio del Corso di Laurea in Tecnologie Web e Multimediali dell'Università di Udine - un percorso di istruzione e sperimentazione orientato alla comprensione e al controllo dei meccanismi che sono alla base della esperienza sensoriale audio-visiva [4] in prodotti multimediali interattivi.

Ci interessa, in particolare, esplorare gli effetti di senso - sensazioni, impressioni, associazioni simboliche, ordinamenti narrativi, etc. [5, 6] - che possono essere evocati nell'utente dall'integrazione di strutture (forme) sonore e visive e come questi effetti possono essere influenzati e modificati dall'introduzione di meccanismi interattivi. Il

*Copyright: ©2012 Elio Toppano 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.*

percorso curriculare proposto si ispira alla tradizione didattica del Basic Design [7, 8]. L'obiettivo è quello di favorire lo sviluppo negli studenti di una sensibilità verso le caratteristiche plastiche/cinetiche [9] e spettromorfologiche [10] rispettivamente delle immagini e dei suoni, e la capacità di saper usare queste caratteristiche in maniera sinergica e interattiva per un preciso intento comunicativo. L'appuccio seguito si differenzia da esperienze simili [11, 12] per l'attenzione rivolta alla esperienza mediale dell'utente e il ruolo attribuito alla Semiotica come strumento critico di analisi dei processi di costruzione e articolazione del significato durante la progettazione del prodotto multimediale [5]. Di seguito, vengono illustrati alcuni esercizi assegnati agli studenti.

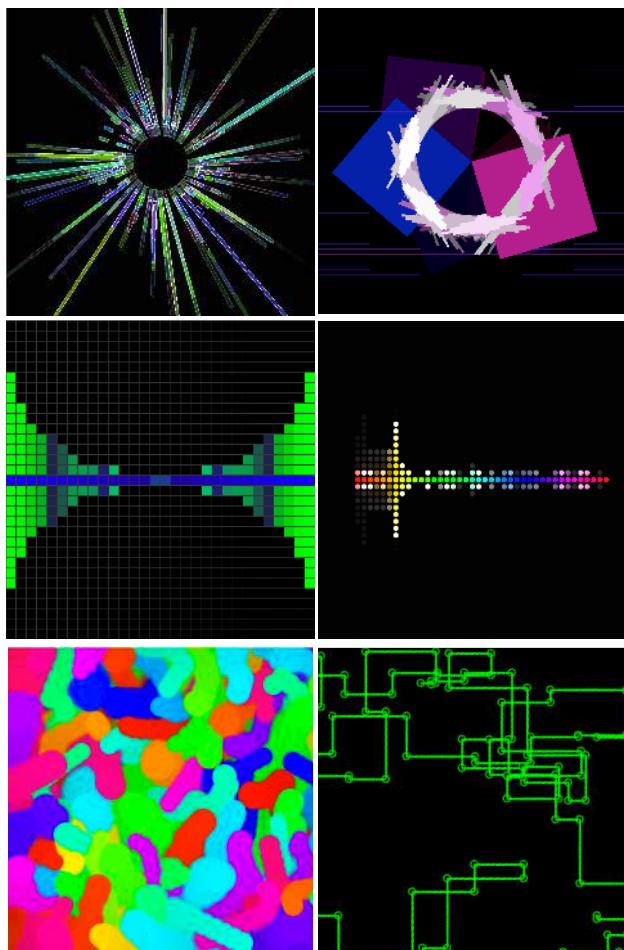
### 2. SINCRESI AUDIO-VISIVA: LE SPECIFICHE DEL PROGETTO

Il progetto consiste nell'ideazione e realizzazione, tramite l'ambiente di programmazione *Processing*<sup>1</sup>, di due animazioni audio-visive basate su uno stesso frammento sonoro scelto dallo studente (ad esempio, un suono ambientale registrato o della musica), della durata di almeno un secondo. Per la parte visiva è richiesto di usare un unico tipo di forma geometrica bidimensionale (triangolo, rettangolo, ellisse, cerchio,...) che può cambiare in dimensione, colore, posizione e orientazione. Il numero totale di elementi grafici dello stesso tipo è libero.

In entrambe le animazioni, alcune caratteristiche visive (ad esempio, l'evoluzione nel tempo della posizione, forma, colore, dimensioni, ... degli elementi grafici) devono essere congruenti con la simultanea variazione di alcune caratteristiche sonore scelte dallo studente (ad esempio, l'intensità del suono, le sue caratteristiche spettrali, il ritmo, il beat, ...). Per quanto riguarda l'interazione, nella prima animazione, si richiede che l'utente possa modificare alcune caratteristiche visive non utilizzate per la sincresi col suono attraverso azioni elementari (per esempio la posizione del cursore del mouse, un click del mouse). Nella seconda animazione, le azioni dell'utente devono poter modificare il frammento sonoro, ad esempio introducendo dei filtri.

Infine, in entrambe le animazioni, si richiede che alcune caratteristiche non utilizzate possano dipendere da elementi casuali, leggi fisiche, ecc. usando almeno una tecniche (o variazioni di essa) vista nelle esercitazioni in laboratorio. Le stesse esercitazioni hanno precedentemen-

<sup>1</sup> [www.processing.org](http://www.processing.org)



**Figure 1.** Screenshot di alcuni progetti realizzati dagli studenti.

te mostrato agli studenti come usare la libreria *Minim*<sup>2</sup> per la riproduzione, l'analisi (ad esempio, tramite FFT) e l'elaborazione di base di un frammento sonoro.

### 3. RISULTATI E CONCLUSIONI

I lavori realizzati dagli studenti sono consultabili alla pagina <http://www.openprocessing.org/classroom/1600> (alcuni di essi sono visibili in Figura 1). Anche se le specifiche del progetto non permettono un'analisi comparativa rigorosa, l'esame dei lavori eseguiti dagli studenti permette di notare l'uso ricorrente di alcune corrispondenze tra caratteristiche visive e auditive. In molti lavori, per esempio, è stato scelto di associare la dimensione dell'elemento grafico usato nell'animazione con l'energia associata a specifiche bande dello spettro del segnale oppure con il tempo (beat) del suono. Analogamente, per quanto riguarda l'interazione, l'azione dell'utente è stata spesso associata al colore o all'orientamento degli elementi grafici usati nell'animazione oppure alla frequenza di taglio del filtro usato per il suono. Va sottolineato il fatto che il progetto non è stato associato ad un preciso intento comunicativo. Pertanto gli studenti si sono focalizzati prevalentemente sugli aspetti tecnici della sincronizzazione e della interazione.

<sup>2</sup> [code.compartmental.net/tools/minim/](http://code.compartmental.net/tools/minim/)

La progettazione per la esperienza sensoriale richiede una competenza approfondita delle qualità dei materiali multimediali e degli effetti che si possono ottenere dalla loro composizione spaziale e temporale. Le esercitazioni discusse sono da ritenersi preliminari alla formazione di tale esperienza. Per il suono, si ritiene inoltre indispensabile sviluppare una concettualizzazione e un vocabolario condiviso che vada oltre alle tradizionali caratterizzazioni in termini fisici (es. intensità, inviluppo, frequenza, spettro) o percettivi (es. volume, altezza/massa, timbro) e permetta ai progettisti di parlarne come una forma unitaria (*gestalt*) dotata di coerenza interna che riassume diversi parametri.

### 4. REFERENCES

- [1] P. Desmet and P. Hekkert, "Framework of Product Experience," *International Journal of Design*, vol. 1, no. 1, pp. 57–66, 2007.
- [2] P. C. Wright, J. Wallace, and J. C. McCarthy, "Aesthetics and experience-centered design," *ACM Trans. Comput.-Hum. Interact.*, vol. 15, no. 4, 2008.
- [3] B. M. Costello and E. A. Edmonds, "A study in play, pleasure and interaction design," in *DPPI* (I. Koskinen and T. Keinonen, eds.), pp. 76–91, ACM, 2007.
- [4] M. Chion, *L'audiovisione. Suono e Immagine nel Cinema*. Lindau, 2001.
- [5] R. Eugeni, *Semiotica dei media. Le forme dell'esperienza*. Carocci Editore, 2010.
- [6] D. Ricco, *Sentire il design*. Carocci Editore, 2008.
- [7] I. Simonini, "Storia del basic design," in *L'ambiente di apprendimento. Web Design e processi cognitivi* (M. A. Garito, G. Anceschi, and M. Botta, eds.), pp. 69–88, McGrawHill, 2006.
- [8] A. Findeli, "Rethinking design education for the 21st century: Theoretical, methodological, and ethical discussion," *Design Issues*, vol. 17, no. 1, pp. 5–17, 2003.
- [9] A. J. Greimas, "Sémiotique figurative e sémiotique plastique," vol. 60, 2008.
- [10] D. Smalley, "Spectromorphology: explaining sound-shapes," *Org. Sound*, vol. 2, pp. 107–126, Aug. 1997.
- [11] D. Riccò, A. Belluscio, and S. Guerini, "Design for the synesthesia. audio, visual and haptic correspondences experimentation," in *Proceedings of the 1st International Meeting of Science and Technology of Design Sensors and Sensibility – Linking Tradition to Innovation Through Design*, pp. 25–26, 2003.
- [12] D. Rocchesso, P. Polotti, and S. Delle Monache, "Designing continuous sonic interaction," *International Journal of Design*, vol. 3, Dec. 2009.

# A sound programming strategy to survive to a complex multimedia setup

Andrea Valle

CIRMA - StudiUm - Università di Torino  
andrea.valle@unito.it

## ABSTRACT

The paper presents a software solution for audio multi-process setups, i.e. where many different audio software setups are to be fast loaded/unloaded, that has been implemented for Marcel·lí Antúnez Roca's last work *Pseudo* (2012).

## 1. INTRODUCTION

Marcel·lí Antúnez Roca (hence on, MAR) is a renown Catalan artist, founder and artistic director for the first ten years (1979-1989), of the notorious La Fura dels Baus, a collective that since its inception has worked structurally with complex "multimedia" setups [1]. Since the '90s, MAR has started an advanced research in interactive multimedia, where the body often plays a pivotal role. He has now systematized his approach developing the notion of "Sistematurgia" ("systematurity") [2]. In most of MAR's recent works, the crucial notion in managing the complexity of contents in relation to technology is the "scene". Scenes are dramaturgical units that include a set of real-time, interactive, visual features, e.g. face tracking, pressure sensors, audio capture and manipulation, and so on. The complexity of the Systematurity approach is pushed to its limit in the most recent of MAR's work, *Pseudo*, premiered in July 2012 at Grec Festival, Barcelona, that further develops the previous work *Cotrone* [3]. *Pseudo* includes interactive animated videos, real-time video capture, a mechanical head (reproducing MAR's face) controlled by an Arduino board. Moreover, in certain scenes MAR wears on his head a miniprojector (his "mind lantern") that projects a video onto a paper sheet attached to the head itself, and the "bodysound", a couple of mini loudspeakers that are placed on his breast (with an evident reference to female nature). The main control device is an exoskeleton (the "dreskeleton") equipped with various electro-mechanical controllers, that MAR himself wears in order to pilot the interactive elements of each scene, and to switch between a scene and another. Figure 1 shows MAR wearing the dreskeleton, interacting with the audience by adding animated content to live video.

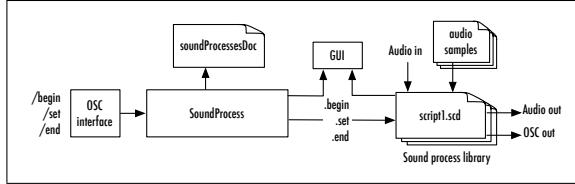
Copyright: ©2012 Andrea Valle et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



**Figure 1.** *Pseudo*: audience interacting live with the videos (courtesy MAR).

## 2. SINGLE- VS. MULTI-PROCESS AUDIO SETUPS

From the perspective of audio, the main feature that differentiates the systematrical approach from other multimedia configurations for live performance can be described in terms of single- vs. multi-process setup. A single process setup is an audio hardware-software setup targeted at performing a specific task. Substantially all the setups for live electro-acoustic music are single-process in the sense that they include a specific hardware-software configuration, of arbitrary complexity. Of course, e.g. during a live-electronics concert, many single-process setups can be used. Instead, a multi-process setup is based on the switching, fastest as possible, between processes, each one again of arbitrary complexity. Multiprocess setups are indeed present also in other contexts, e.g. in contemporary operas involving live electronics such as the recent Marco Stroppa's *Re Orso* (2012). Limiting our scope to the audio software side, MAR's requirements are particularly demanding. Sound and music processing are continuously running during the whole performance. The processes range from audio recording and manipulation to real-time algorithmic composition, from sound analysis to interactive playback triggering. The change of scene is handled live by MAR himself by means of the dreskeleton and has to be carried out in a completely automatic way, upon receiving an OSC message (OSC being the standard communication protocol among the various software involved in *Pseudo*). Robustness in the change of scene is indeed required. Finally, while the sequence of scenes is fixed during live performance, their order can dramatically change during rehearsal, so that a "free jumping" mode has to be considered as the general case for implementation. In order to cope with the complexity of these requirements, we decided to use SuperCollider (SC) as the main development

**Figure 2.** Sound Process: main components and relations.

platform [4]. SC features a very efficient and state-of-the-art audio server, and a flexible, object-oriented language as a client. The language is specialized in sound and music manipulation but implements very general features. All the required tasks can be easily and efficiently performed by SC, but they must be integrated into a global framework, in order to provide a common and unique interface, ensuring that they can be fast recalled upon request from external applications via OSC protocol. Symmetrically, also the composer/sound designer needs a framework that allows her/him to shape the composition/design into an interactive application. We thus developed the notion of “sound process”.

### 3. SOUND PROCESSES

A “sound process” is a SC script that contains code to perform real-time DSP operations (including synthesis, playback of audio files and live input processing); to schedule sound events (i.e. algorithmic composition); to send OSC messages (e.g. containing data from audio analysis to be used to trigger animations). A multi-process setup is indeed characterized by the presence of a whole set of sound processes. Sound processes are implemented in SC as *prototypes*, that is, constructs that mediate between classes and objects, as they serve both to define objects *and* to use them. Prototypes can be implemented in SC by means of *events* [5], that is, special dictionaries associating keys to arbitrary objects. The use of prototypes avoids cluttering the library itself with a huge amount of classes, each one simply representing a sound process. In a multi-process setup, the composer/sound designer thus creates a set (a library) of sound process scripts, and selects/controls them in real time. Some conventions apply to sound process creation. Each sound process contains a dictionary *nfo* storing information used to create automatic documentation and GUIs. Each sound process is supposed to implement an interface that includes three possible behaviors accessible from outside: *begin*, *end*, and *set*. The first two methods respectively specify what to do when a scene is loaded or unloaded, including allocation/deallocation of resources (buffers, busses) and process startup/stop. The *set* methods are used to publicly expose an interface for the specified parameters. In order to link sound process scripts to an application network, a SoundProcess class (in the proper sense) has been developed, that encapsulates all the functionalities required to manage a library of sound processes: sound process loading/unloading; OSC connection; automatic GUI generation for each process; automatic documentation from source code. The design of Sound Process can be seen as an application of the Strat-

egy pattern, that allows to encapsulate a family of algorithms and choose among them at runtime [6]. Figure 2 shows a general schema of the whole system. A SoundProcess instance acts as the manager on the software side of a library of sound process scripts, to be written by the composer/sound designer. Once inserted into a library by a SoundProcess instance, a sound process receives a symbolic name, that is automatically retrieved from the file where it is described. Each sound process is contained in a folder including all the data related to that process, e.g. a “sounds” subfolder where all samples to be used by the process are stored. Finally, a SoundProcess can inspect the sound processes in its library to generate automatic documentation. An instance of SoundProcess is thus the application that runs in the *Pseudo* application network, to which OSC messages can be sent or from which they can be received. Upon receiving OSC messages, SoundProcess then calls the relative methods to the selected process.

### 4. CONCLUSIONS

The SoundProcess framework has proved to be both very versatile and efficient. Each sound process is a completely autonomous block of SC code that can be controlled interactively in real-time from outside (via OSC) by means of the SoundProcess class. This allows to radically change the sound and music perspective in response to a change of scene. Scene swapping is substantially instantaneous and loading time increases only if sound samples are loaded in RAM. In any case, an extreme case where 209 buffers were allocated on the fly to store audio samples for a total of 130 MB required empirically only 0.5 sec on a MacBook Pro 2.4 Ghz with 4 GB RAM. Once revised and fully documented, the Sound Process framework is planned to be published as a *quark*, a SC extension.

### 5. REFERENCES

- [1] T. Mazali, F. Mazzocchi, and A. Pizzo, eds., *Marcel·lí Antúnez Roca e la performatività digitale*. Acireale-Roma: Bonanno, 2011.
- [2] M. Antúnez Roca, *Marcel·lí Antúnez Roca e la performatività digitale*, ch. Materia sistematurgica, pp. 13–31. Acireale-Roma: Bonanno, 2011.
- [3] A. Valle and P. Armao, “Integrating sound design and music composition in the interactive play Cotrone by Marcel·lí Antúnez Roca,” *The New Soundtrack*, vol. 2, no. 1, pp. 63–77, 2012.
- [4] S. Wilson, D. Cottle, and N. Collins, eds., *The SuperCollider Book*. Cambridge, Mass.: The MIT Press, 2011.
- [5] A. De Campo, J. Rohrhuber, and T. Bovermann, *The SuperCollider Book*, ch. Object Modeling, pp. 237–269. Cambridge, Mass.: The MIT Press, 2011.
- [6] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object Oriented Software*. Addison-Wesley, 1995.

# LOLA: A LOW LATENCY HIGH QUALITY A/V STREAMING SYSTEM FOR NETWORKED PERFORMANCE AND INTERACTION

A project by Conservatorio di Musica G. Tartini and Consortium GARR

**Carlo Drioli<sup>1,2</sup>**

<sup>1</sup>Conservatorio di Musica G. Tartini, Trieste, Italy

<sup>2</sup>University of Udine, Udine, Italy

**Claudio Allocchio**

Consortium GARR

Rome, Italy

## ABSTRACT

We present LOLA (LOw LATency audio visual streaming system), a system for distributed musical interaction over packet networks. It is intended to operate on high performance networking infrastructures, and is based on low latency audio/video acquisition hardware and on the integration and optimization of audio/video data acquisition, presentation and transmission. The extremely low round trip delay of the transmitted data makes the system suitable for remote musical education, real time distributed musical performance, and in general for any human-human interactive distributed activity in which timing and responsiveness are critical factors for the quality of the interaction. The experimentation conducted so far with professional music performers and skilled music students, on geographical distances of up to 2400 Km, demonstrated its effectiveness and suitability for distance musical interaction, even when professional players are involved and classical music repertoire is concerned.

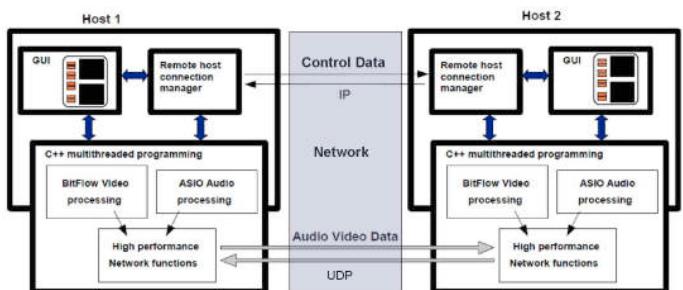
## 1. INTRODUCTION

Distributed collaborative environments have been recently the subject of considerable interest and investigations. In some specific cases, collaborative activities include tasks in which the speed of interaction plays a critical role, and the growing demands in terms of performance has highlighted the limits of the currently available multimedia data processing and transmission technology [1]. In particular, the field of interactive musical collaboration, referring to scenarios such as geographically distributed musical performance or distance music education, poses specific problems related to the management of high quality audio-video streaming, to the transmission speed and delays, and to the impact of transmission delays on musical performers accuracy [2, 3, 4].

To specifically address this class of problems, an high-quality, low transmission latency system aimed at distributed music performance over high-end packet networks, was designed and developed. LOLA (the acronymous stands for "LOw LATency") was originally conceived and designed at the Tartini Music Conservatory of Trieste, and at present

is being developed and tested with the collaboration of GARR (the Italian Education and Research Network organization). The Lola development team is composed of Massimo Parovel (conception and supervision), Paolo Pachini (general coordination), Nicola Buso (testing and musical advice), Carlo Drioli (system design and programming), Claudio Allocchio (testing and networking advice).

Low transmission latency refers to a very specific feature of an Audio/Video communication system where the transmission delay among remote sites is very small and therefore negligible for the human eye and ear. Since its origin, the main goal of the LOLA project was to create a system fulfilling this specific requirement, thus building an high quality tool for remote musical education and real time distance musical performance.



**Figure 1.** System architecture overview.

## 2. ARCHITECTURE AND IMPLEMENTATION

The LOLA A/V streaming system is conceptually a video-conferencing hardware/software system. However, the operating conditions that it was designed for, and the hardware and software design solutions adopted, makes it different from any other videoconference system available to date. It was designed to fulfill a number of fundamental requirements: 1. to be suitable for musical performances relying on both audio and visual communication, with the goal to provide a natural and transparent end-user interaction; 2. to be low cost and portable; 3. to exploit bandwidth and robustness of dedicated high performance networks (e.g., LightNet Project, GARR, GÉANT, Internet2).

In order to achieve a low transmission and presentation latency, LOLA relies on software optimization and on high performance audio and video devices. Fast video acquisition and streaming relies on a family of industrial video grabbers by BitFlow Inc., which provides high hardware

Copyright: ©2012 Carlo Drioli<sup>1,2</sup> 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 2.** Public networked live performances at three different moments of the LOLA system development. Left: piano duo playing classical music repertoire (Bach, Reger) during 2010 Network Performing Arts Production Workshop at a distance of approx. 1300 Km (Trieste, Tartini Conservatory - Paris, IRCAM) ; Center: violin-cello duo playing Haendel during 2011 Internet2 Fall Members Meeting at a distance of 1850 Km (Chicago, NIU - Raleigh, Congress Center); Right: trumpet duo playing Bozza during 2012 Performing Arts Production Workshop, at a distance of approx. 2400 Km (Chicago, NIU - Miami, NWS).

performances and a versatile programming API for low-level video processing control, and on industrial class progressive analog video cameras. Low latency audio performance is achieved by relying on robust hardware and driver equipment (RME Hammerfall and ASIO drivers). Both audio and video streaming are optimized for speed by relying on accurate audio acquisition, transmission, and rendering threads synchronization.

Current release supports audio at 44100 samples/sec, 16 bit, and 640x480 resolution video, at 60 or 30 fps, colour or black and white. Audio and video are non compressed, to avoid introducing time delays in the encoding/decoding process. This allows to reach RTT delays (not considering network delay) as low as 5 msec for audio, and as low as around 30 msec (estimated perceptually) for video. Network latency to be considered in the total RTT delay estimates is <1 ms on LANs, ~1 ms per 100 Km on WANs. Jitter might become sensibly high when operating on network branches with public traffic, thus a buffering mechanism is provided to prevent data loss due to network delay oscillations. In terms of bandwidth usage, LOLA requires at least 100 Mbps in minimal configuration (standard definition, b/w, 30 frames per second) up to 500 Mbps in full configuration (standard definition, color, 60 frames per second), and generates a very high Packet per Second (PPS) rate, as it uses 1 Kb data packets. Thus the minimal end-to-end connectivity must be at least 1 Gbps<sup>1</sup>.

### 3. DISTRIBUTED PERFORMANCES

The LOLA project was conceived in 2005, and after a set of preliminary studies, was developed between 2008 and 2010. LOLA first public showcase was in November 2010, with a performance between Trieste and Paris at the Network Performing Arts Production Workshop, and then in several other venues in Europe and USA.

If compared to inherent latencies introduced by standard videoconferencing systems (e.g. DVTS, Conference XP, Skype), usually not below 0.5 sec for both audio and video, the inherent low latency provided by LOLA was received very favourably by musicians who had former expe-

riences related to distributed musical performances. In all cases in which the RTT was kept below the 75 msec threshold, performers reported to be able to play comfortably and to feel the system as transparent after a short while (this operating situation was met in all public events, see e.g. Fig. 2). During laboratory test sessions, in which latencies were artificially raised or lowered on request, a number of observations were collected concerning the different factors participating to the perception of interaction delay and on their impact on performance quality: the musical repertoire, the timbre and dynamic characteristics of the musical instruments, the reverberation and remote instrument rendering, among others. All these factors will be systematically investigated in future research concerning the musical applications of LOLA.

### 4. CONCLUSIONS

A new audio visual streaming system, characterized by high quality audio/video rendering and by very low data acquisition, transmission and presentation delays, has been illustrated. It was designed to accomplish time-critical distributed interactive tasks such as distributed music performance. It was assessed with respect to this specific aim by involving professional music performers playing classical repertoire over wide geographical distances, and it proved suitable and effective for distance musical rehearsal, performance and production.

### 5. REFERENCES

- [1] D. Gergle, R. E. Kraut, and S. R. Fussell, "The impact of delayed visual feedback on collaborative performance," in *In Proceedings of CHI 06*, pp. 1303–1312, ACM Press, 2006.
- [2] D. Konstantas, Y. Orlarey, O. Carbonel, and S. Gibbs, "The distributed musical rehearsal environment," *IEEE MultiMedia*, vol. 6, pp. 54–64, July 1999.
- [3] R. Zimmermann, E. Chew, S. A. Ay, and M. Pawar, "Distributed musical performances: Architecture and stream management," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 4, pp. 14:1–14:23, May 2008.
- [4] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan, "Effect of time delay on ensemble accuracy," in *In Proceedings of the International Symposium on Musical Acoustics*, 2004.

<sup>1</sup> The software and documentation is freely available for non commercial use at: <http://www.conservatorio.trieste.it/artistica/lola-project>

# Towards a portable wireless platform for networked performance

**Leonardo Gabrielli, Stefano Squartini, Francesco Piazza, Matteo Mencarelli**  
Dept. Information Engineering, Università Politecnica delle Marche, Ancona, Italy

## ABSTRACT

Networked musical performances have gained increasing interest in recent years due to the high availability of computational power on common laptops, their networking capabilities and their increasing versatility. Their size and weight however can limit the freedom of the performer. This poster describes the ongoing activities concerning the setup and enhancement of a small computing platform for sound synthesis and musical composition based on the BeagleBoard xM and capable of streaming audio in a wireless network.

## 1. INTRODUCTION

In recent years networked performances and laptop orchestras have gained popularity and are well established media for contemporary music composition and improvisation. Laptops are not designed, however, to be used as musical instruments, nor they enable good physical interaction without additional interfaces. In the need to improve usability mobile phones have been used as instruments and their internal sensors as input devices [1]. Smartphones however have lower computational power with respect to personal computers and they are more limited in the input/output capabilities. Another solution emerged in the last years is the use of small microcontroller-based boards, such as Arduino<sup>1</sup>. These have a tiny footprint and can gather data from sensors and feed it to a computer for audiovisual processing.

A recent trend that emerged in the very last years is the use of solutions based on ARM processors. These are powerful enough to process sound and/or video in realtime, while they can run complex and flexible operating systems such as GNU/Linux distributions. This way all the advantages of embedded platforms are retained while the flexibility of computational systems such as CSound, Super Collider, Pure Data and Processing are available. The BeagleBoard<sup>2</sup>, an open-hardware board designed around a Texas Instruments ARM-Cortex-A8 core, is a good choice as it provides good processing capabilities in a small package. For this reason it has already been chosen for several projects [2, 3]. The only downside of this board is the

lack of analog input/output ports. For this reason in [2] an Arduino Nano board is attached to the board to provide analog and digital input/output.

In this work preliminary results are provided for a similar system which includes BeagleBoard xM boards equipped with WiFi dongles that are able to process and synthesize audio and receive or stream it from/to other devices for many purposes at a reasonably low latency. The aim of the authors is to provide a proof of concept of a low-cost network of interoperable audio/musical devices enabling the investigation of possible future scenarios, applications and use issues.

## 2. SYSTEM DESCRIPTION

The system is composed of several BeagleBoard xM embedded devices. The computational system of choice is Pure Data [4] (in short Pd) which is used for audio processing and transmission, while the Jack audio server is used to enable real-time audio scheduling.

The BeagleBoard xM is based on a TI DM3730 processor including an ARM Cortex-A8 core and a C64x fixed point DSP. Audio processing is possible thanks to SIMD instructions available on the ARM core (VFPv3 and NEON). The C64x processor is available to enhance computational power, but further effort is required to make it work with the ARM processor. A Belkin F5D7050 WiFi dongle is used to connect to the wireless network.

### 2.1 Software

Each BeagleBoard is equipped with a Ubuntu 10.10 distribution as provided by the Satellite CCRMA project [2]. Pure Data has been chosen as the processing and streaming platform because of its cross-platform availability, the support of both audio and control data formats and the ease of programming for musical applications. The audio/control streaming is done in Pure Data. Two externals are available in Pd for streaming audio through TCP or UDP called *netsend~* and *netreceive~*. These are very simple streaming functions relying on the UNIX *sys/socket.h* primitives. They allow connection to a remote machine via TCP and simple unicast or broadcast via UDP datagrams. They do not implement device discovery or clock synchronization mechanisms. For the current setup they provided sufficient functionalities while in future a more feature-rich and low level implementation is needed.

With the combination of hardware and software used in this setup acceptable latency requirements were met by using three buffers with size 128 samples at 48kHz sampling frequency. This yields approximately 8 ms of latency.

<sup>1</sup> <http://www.arduino.cc>

<sup>2</sup> <http://www.beagleboard.org>

Should latency requirements change, the framesize in Jack can be adjusted, bearing in mind that a lower framesize will require a higher CPU load due to more frequent system interrupts and buffer handling. With the current software setup the average CPU load on a BB is approx. 31%, 14.5%, 7.5% 2.8%, 1.7% for framesizes of respectively 64, 128, 256, 512, 1024 samples at 48KHz.

### 3. NETWORK TESTS

The setup has been put to test to see whether it allows acceptable streaming latencies both with audio and control data. Round Trip Times (RTT) have been measured by sending audio/control data to another board and back again to the main one. The *netsend~* and *netreceive~* externals have been used for audio streams while *netsend* and *netreceive* have been used for control streams. In both the test-cases RTTs are averaged through mean or median, leading to slightly different results due to packet losses.

Test results follows. The mean RTT of Fast Ethernet audio/control data is 12.5ms/2.3ms while the 802.11g RTT of audio/control data is 42/2.4ms. The values are listed as RTTs, hence the single link latency is approximately half. Audio tests latency is mostly due to the buffering delay (three 128-frames buffers at 48kHz sampling rate for each audio input/output). The results show that very low latency streaming is already possible with control signals and acceptable latency also shows up in wireless tests with audio signals (< 50ms).

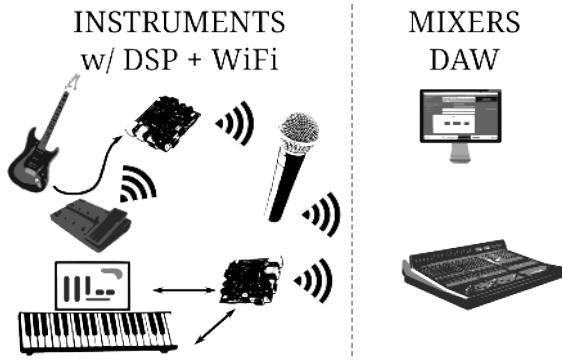
#### 3.1 Description of a real world setup

A setup similar to that shown in [5], was built where the morpher/synthesizer IBRIDA is controlled by a mobile device. An Android phone running the touchOSC app by Hexler<sup>3</sup> sends OSC data (including accelerometer data and virtual buttons, sliders and surfaces) to a master BeagleBoard through UDP packets. The master BeagleBoard runs a patch of the audio morpher/synthesizer IBRIDA. The audio generated by IBRIDA is sent to a slave BeagleBoard that receives the audio and feeds it back to the IBRIDA audio input on the master board, thus creating a feedback loop with variable delay. The loop delay changes unpredictably depending on network conditions, similarly to Internet Acoustic Echo in [6]. The feedback gain is controlled inside the master BeagleBoard IBRIDA patch in order to prevent instabilities. Approximate average latency was measured to be around 25.4 ms. This latency is acceptable as strict timings are not required for this application.

More applications are going to be built in a similar configuration by also adding touchscreens to the BeagleBoard.

### 4. DISCUSSION

As seen, such a platform can allow several new applications. BeagleBoards can act as standalone computing devices, input sources, or sinks for audio coming from other boards. Several input devices can be connected to the boards such as musical keyboards, touchscreens, microphones, or



**Figure 1.** A schematic view of possible scenarios enabled by the platform.

custom electronic interfaces. This way a BeagleBoard can also act as a MIDI/OSC controller and send commands through a wireless network. Such a device can also be used in a music production system, such as a small home studio. A resume of possible scenarios is provided in Figure 1.

The present work provided preliminary results for a wireless system for music production, performance and sound synthesis. While the feasibility of the approach as been demonstrated, more effort is required to reduce network latencies and improve the buffer latencies on the BeagleBoard DACs. At the moment the streaming is only possible within Pure Data. More advanced tools are needed that work as Jack clients, which also include automatic device discovery and connection to other similar boards.

### 5. REFERENCES

- [1] J. Oh, J. Herrera, N. Bryan, L. Dahl, and G. Wang, “Evolving the mobile phone orchestra,” in *Proc. Int. Conf. on New Instruments for Musical Expression*, Sidney, Australia, 2010.
- [2] E. Berdahl, “Satellite CCRMA: A musical interaction and sound synthesis platform,” *Proc. Int. Conf. on New Instruments for Musical Expression*, Oslo, Norway, 2011.
- [3] R. Vega and D. Gómez, “Using the Beagleboard as hardware to process sound.” <http://lac.linuxaudio.org/2012/papers/31.pdf>, online, last viewed 31/07/2012
- [4] M. Puckette, “Pure data: another integrated computer music environment,” in *Proc. Int. Computer Music Conf.*, pp. 37–41, 1996.
- [5] L. Gabrielli and S. Squartini, “Ibrida: a new DWT-domain sound hybridization tool,” in *AES 45th International Conference*, Espoo, Finland, 2012.
- [6] C. Chafe, S. Wilson, and D. Walling, “Physical model synthesis with application to internet acoustics,” in *Acoustics, Speech and Signal Processing, IEEE Int. Conf. on*, IEEE, 2002.

<sup>3</sup> <http://hexler.net/software/touchosc-android>

# LIES (distance/incidence) 1.0: a human-machine interaction performance

**Dario Sanfilippo**  
Conservatory of Naples  
[sanfilippo.dario@gmail.com](mailto:sanfilippo.dario@gmail.com)

## ABSTRACT

This paper describes *LIES (distance/incidence) 1.0*, a human-machine interaction improvised performance implementing complex dynamical systems based on analog and digital feedback networks. The author's approach to human-machine interaction will be introduced, and the concepts behind this work will be discussed.

## 1. INTRODUCTION

Human-machine interaction is often linked to the practice of designing gestural interfaces to control a system through the mapping of variables to the parameters of the processes the system performs. Nowadays we have a great variety of devices to obtain variables to be used to control a system, from MIDI to optical sensors, motion capture and so on, and these are often used in a linear communication flow design, where the human performs an action and the system reacts to it [5]. Yet, a condition of *mutual influence* is fundamental and implicit in the notion of *interaction*, and in order to achieve such a condition both the human and the machine should be capable of autonomous dynamical behaviours, so that they both can perform actions and reactions. The performance project described here explores a condition where interaction takes place in both high- and low-level [1,4,5]: between the human and the machine, and between the processes of sound creation in the machine. Feedback (aural feedback in the case of improvisation) is the key for the realisation of such a condition, a mechanism with which it is possible to establish an interdependency between human, machine, sound and environment [7,13].

## 2. FUNDAMENTAL CONCEPTS

An important concept at the basis of this project is that the human and machine are considered as inseparable: two autonomous entities which, unavoidably, will influence each other, creating a unique meta-system made up of these two elements. The human and the machine establish a dialectics, a *talking through the other*, with no attempts of subordination, creating a performance which is the result of their cooperation, where, thus, the performer creates *together* with the machine. The environment where the per-

formance takes place is a constant mediation for the sound, as the necessary condition for the sound to emerge, through acoustic feedback (Larsen phenomena), implies a sensitivity to external factors. Unlike most of the music systems where dynamical behaviours and unpredictability is achieved by means of automation and randomness generators, the approach for the design of a dynamical and autonomous entity in this project is a systemic one. Through feedback, it has been implemented a dense network of low-level sonic processes which will be recursively and nonlinearly interacting, where all the nodes in the network will directly or indirectly influence each other, and where a distributed circular causality takes place [8,10,13]. There's no randomness or high-level automated processes, yet the system is capable of chaotic, evolving behaviours and self-disorganisation [9]. Its properties are emergent phenomena coming from a complex synergy between the system's components [2,3,4,6,7,11], resulting in a whole which is different from the sum of its parts [12], a *nontrivial machine* with a causal relation between past, present and future, which will be capable of modifying itself [8]. Microphones and loudspeakers can be considered as system's terminals, the outer nodes with which, through the environment, the system establishes a connection with itself. The performer's role is that of mediating this connection, namely by varying the distance between microphones and loudspeakers, and the angle of incidence of sound, thus exploring the different behaviours of a system which has different sensitivities to the intensity and spectral profiles of sound. Moreover, if we also consider the role of the environment in this project, we realise the performer's role is actually a mediation within a mediation, so it is not just about variations of distance and angle of the microphones, rather it is an exploration of a 3D space, of a *performance ecosystem* [14] with complex characteristics of anti/resonances where the interactions between human, machine and sounds take place.

## 3. THE MACHINE

The fundamental machine's design takes into account the interaction and cooperation between an analog and a digital feedback network. Namely, the analog feedback network is responsible for the generation of the sonic source: when enough amplification is provided, the feedback is positive and self-oscillation occurs. The digital feedback network has been implemented so that the overall feedback coefficient is below the unity, so that self-oscillation does not occur. This way, the two networks are interdependent and operate as a single unit: the analog feedback network

Copyright: ©2012 Dario Sanfilippo 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.

will generate sonic source which will turn the digital feedback network into an operating state, the sonic source will be transformed and will influence the analog feedback network, recursively.

### 3.1 Analog Feedback Network

The analog feedback network consists of 2 microphones (inputs) and a number of loudspeakers and subwoofers (outputs) to be connected to. It is possible to set the number of loudspeakers to a multiple of 2, while 1 or 2 subwoofers are usually enough for most spaces, and according to the number of outputs different displacements and setups can be explored. Considering that all these elements are within the same environment, we can assume that all outputs will influence all inputs, and all inputs will in turn be sent to all outputs. Thus we have a fully-connected network topology where a high number of feedback paths can be achieved, and where their interactions can result in already complex behaviours.

### 3.2 Digital Feedback Network

The software has been implemented on Pure Data. It consists of 3 main blocks of processes in feedback configurations where all feedback coefficients have a minus sign to avoid DC, and where a delay of 1024 samples has been set for the feedback loops. The system is fed with the microphones inputs, and each of the blocks has a different sensitivity to the sound intensity and spectral profile. The first block consists of a group of four frequency shifters, where the first two are cross-coupled and perform a shift in the very high frequencies, around 10kHz, while the others, in series, shift the signal of a fraction of Hz, around 0.05-0.02, which is then summed to the original in order to obtain beats phenomena. In the first step of the second block we have a couple of comb filters bank and a couple of ring modulators. The comb filters have a decay of 10 seconds and their loop periods are prime numbers below 30 milliseconds, so that the energy spreading over the spectrum is maximised. Combs will be modulated by a signal which is the sum of the modulators' output and the microphones' inputs, and the 4 output signals will go into 2 cross-coupled nonlinear distortion units whose transfer function is represented by the first 26 Chebyshev polynomials, which will enrich the spectrum and will also act as kind of a nonlinear gate. In the first step of the third block we have 2 frequency shifters that, as in the first block, will be used to achieve beats phenomena. The sound will then be equalised by shelving filters to boost the very low frequencies, around 40Hz and up to 30 dBs, and will be sent to 2 reverberation units whose decay time is around 20 seconds, here implemented as 16th order feedback delay networks with Hadamard distribution matrixes, and that will act as a smoother on the dynamics of the system. The final step of the network is the automatic gain control algorithm, which is a lookahead limiter providing adaptive attenuation according to the input signal, here used to avoid clipping and maintain signals at a reasonable operating level. The 3 units will be sent to the outputs and will interact in the environment, as they will be fed back into the inputs.

## 4. REFERENCES

- [1] C. Anderson, Dynamic Networks of Sonic Interactions: An Interview with Agostino Di Scipio, in Computer Music Journal, vol. 29, no. 3, pp. 11-28, 2005.
- [2] E. Bonabeau and J.-L. Désalles, Detection and emergence, in Intellectica, no. 25, pp. 85-94, 1997.
- [3] P. Corning, The re-emergence of emergence: a venerable concept in search of a theory, in Complexity, vol. 7, no. 6, pp. 18-30, 2002.
- [4] A. Di Scipio, Formal processes of timbre composition challenging the dualistic paradigm of computer music, in Proceedings of the International Computer Music, 1994.
- [5] —, Sound is the interface: from interactive to ecosystemic signal processing, in Organised Sound, vol. 8, no. 3, pp. 269-277, 2003.
- [6] —, Emergence du Son, Son d'Emergence. Essai d'épistémologie expérimental par un compositeur, in Intellectica, vol. 1-2, no. 48-49, 2008.
- [7] —, Listening to yourself through the otherself: On background noise study and other works, in Organised Sound, vol. 16, no. 2, pp. 97-108, 2011.
- [8] Heinz von Foerster, Understanding Understanding: Essays on Cybernetics and Cognition. New York: Springer, 2003.
- [9] C. Gershenson, F. Heylighen, and C. L. Apostel, When can we call a system self-organizing, in Advances in Artificial Life, 7th European Conference, ECAL 2003 LNAI 2801. Springer-Verlag, 2003, pp. 606-614.
- [10] F. Heylighen and C. Joslyn, Encyclopedia of Physical Science and Technology. New York: Academic Press, 2001, ch. Cybernetics and Second-Order Cybernetics, pp. 155-170.
- [11] M. Mitchell, Complex systems: Network thinking, in Artificial Intelligence, vol. 170, pp. 1194-1212, 2006.
- [12] E. Morin, Restricted complexity, general complexity, in Intelligence de la complexité: Epistémologie et pragmatique, Cerisy-La-Salle, 2005.
- [13] D. Sanfilippo, A. Valle, "Towards a topology of music feedback systems", in Proceedings of the International Computer Music Conference, (in press) 2012.
- [14] S. Waters, "Performance Ecosystems: Ecological Approaches to Musical Interaction," Proceedings of the Electroacoustic Music Studies Network 2007. <http://www.ems-network.org/spip.php?article278>.

# h SONIFICAZIONE DI UNA CAMERA A NEBBIA

Elio Corbolante

eliocor@gmail.com

Alessandro Fogar

Conservatorio G. Tartini -Trieste  
sfogar@gmail.com

## ENGLISH ABSTRACT

We would like to present the construction and subsequent sonification of a cloud chamber. Its sonification has been realized in Supercollider, using as input data obtained via an image analysis algorithm implemented with Max/Msp/Jitter.

## ITALIAN ABSTRACT

Ci proponiamo di illustrare la nostra esperienza nella costruzione e successiva sonificazione di una camera a nebbia. La relativa sonificazione è stata realizzata in Supercollider, utilizzando come input dati ottenuti tramite un algoritmo di analisi dell'immagine implementato con Max/Msp/Jitter.

### 1. INTRODUZIONE

'h', dal simbolo che indica convenzionalmente la costante di Planck, è la sonificazione di un fenomeno fisico che è osservato (e ascoltato) dai visitatori dell'installazione tramite una camera a nebbia [1].

Abbiamo realizzato una sonificazione, basata su una serie di processi di sintesi sonora che utilizzano come dati le tracce lasciate dalle particelle elementari nella camera a nebbia, catturate tramite una telecamera: il computer interpreta le immagini, riconoscendo le tracce e genera in tempo reale il materiale sonoro, in base alle disposizioni e alle scelte del compositore e agli stimoli provenienti dal modulo di riconoscimento.

L'installazione è stata presentata per la prima volta nell'ambito della 'Notte dei ricercatori 2011' a Trieste.

### 2. LA CAMERA A NEBBIA

La camera a nebbia a diffusione (camera di Langdorf) è un dispositivo che contiene una "nebbia" di vapori supersaturi di alcool e permette di visualizzare le traiettorie delle particelle subatomiche che la attraversano.

Tali particelle ionizzano gli atomi dell'alcool in sospensione che si comportano da nuclei di aggregazione per la condensazione del vapore creando una scia che mostra la traiettoria della particella. L'utilizzo di un'illuminazione radente facilita la visione a occhio

nudo di tali tracce che possono essere osservate attraverso una parete trasparente della scatola.

Le immagini sono acquisite da una telecamera/microscopio Celestron 44302-A con messa a fuoco ravvicinata cui sono state apportate delle modifiche per ridurre i riflessi sul contenitore della camera a nebbia.

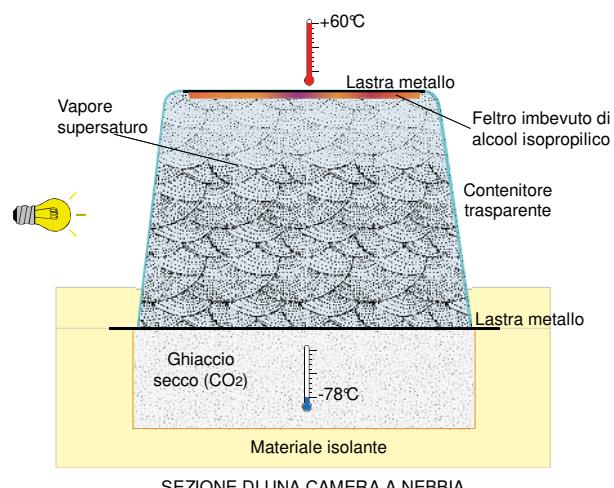


Figura 1. Schema realizzativo della camera a nebbia

La camera a nebbia presentata consiste in:

- una base termicamente isolata che contiene del ghiaccio secco (-78°C)
- un contenitore trasparente appoggiato a tale base
- un coperchio a temperatura 60°C con un filtro imbevuto di alcool isopropilico
- una luce radente

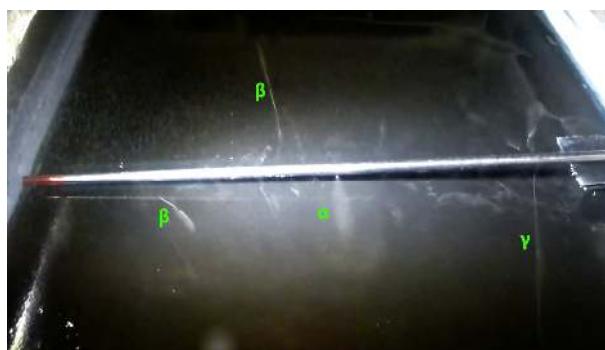
La differenza di temperatura tra la parte calda e fredda ( $>120^{\circ}\text{C}$ ) crea nel contenitore un vapore supersaturo di alcool in cui si possono rilevare le tracce del passaggio delle particelle ionizzanti. Si tratta di un progetto originale, realizzabile a partire da materiali di facile reperibilità.

### 3. RILEVAMENTO DELLE TRACCE

La parte di rilevamento delle tracce lasciate dalle particelle elementari nella camera a nebbia è implementata utilizzando le librerie cv.jit di Jean Marc Pelletier [2], un insieme di oggetti e tools per la computer vision, basati su Max/Msp/Jitter. L'external utilizzato per il rilevamento è cv.jit.lines. Le immagini sono acquisite

dalla telecamera con una risoluzione di 320 x 240 pixel e trasformate in scala di grigi. Le caratteristiche delle linee presenti nelle immagini sono estratte usando dapprima l'algoritmo di Canny [3] per estrarre i bordi, seguito poi da una trasformata di Hough [4]. Si ottiene così una matrice formata da quattro diversi piani: il numero di celle nelle matrici corrisponde al numero di linee rilevate nell'immagine. I quattro piani della matrice rappresentano la posizione delle linee. Tali dati poi sono filtrati, estratti dalle matrici e inviati all'algoritmo di sonificazione.

L'approccio scelto è quello di rilevare le linee che meglio approssimano le tracce lasciate nella nebbia alcolica. Ci sono due diversi processi di rilevamento che operano sulla stessa capture dell'immagine. Il secondo algoritmo è impostato in modo tale che rilevi solo le tracce più evidenti. L'algoritmo di riconoscimento delle tracce è impostato in modo tale da scartare le tracce dalla lunghezza minore di una determinata soglia e maggiorie di un altro limite fissato.



**Figura 2.** tracce lasciate da particelle  $\alpha$ =sfumate e a forma di piumetto, particelle beta  $\beta$ =nette e corte, raggi gamma  $\gamma$ =nitide e lunghe. La barra metallica è un elettrodo di saldatura TIG composto al 2% di Torio ( $^{232}\text{Th}$ ) ed è leggermente radioattiva.

La comunicazione tra il modulo di rilevamento delle tracce e il modulo di sonificazione avviene tramite il protocollo OSC [5]. Questo permette una facile distribuzione del carico di elaborazione tra due elaboratori, uno che si occupa delle rilevazione delle tracce, l'altro che si occupa della sonificazione. I due algoritmi di riconoscimento delle tracce inviano separatamente i loro dati al modulo di sonificazione. I dati che sono inviati dal secondo algoritmo innescano delle variazioni nei parametri che regolano il modulo di sonorizzazione.

#### 4. SONIFICAZIONE

' $h$ ' è la costante fisica (detta di Planck) il cui valore è equivalente alla quantità d'azione fondamentale. Determina la distanza tra i valori assunti dai quanti delle grandezze fisiche fondamentali, tra le quali l'energia, tema principale della "Notte dei ricercatori 2011" per la quale l'installazione è stata ideata. Le linee rilevate vengono "quantizzate" in quattro livelli in base alla loro lunghezza, cioè alla loro energia.

Il modulo di sonificazione è stato sviluppato utilizzando il linguaggio di programmazione SuperCollider, unitamente alle librerie JITLib per l'utilizzo di tecniche di Just In Time programming [6]. Si è scelto di procedere in questo modo non tanto per la necessità di modificare in realtime il modulo di sonificazione ma per la facilità di sviluppo e tuning del modulo stesso.

La sonorizzazione [7] utilizza l'unità di generazione Warp1, un granularizzatore, time stretcher e pitchshifter basato su sound buffer permettendo, in questo modo, di sonificare le tracce, le linee rilevate, ricavando da queste i parametri per la lettura granulare di una serie di soundfiles. Le particelle lasciano una traccia sonora del loro passaggio come se effettuassero uno scrubbing su un file audio virtuale. Posizione iniziale e finale di lettura nel soundfile derivano dalle coordinate rilevate sull'asse x, il pitch dalle coordinate sull'asse y, la durata e la scelta del soundfile in base all'appartenenza ai 4 "quanti" di energia. La quantità di riverbero e multitap delay derivano inversamente dalla densità di eventi. I soundfiles sono stati realizzati in precedenza dall'autore utilizzando SuperCollider con caratteristiche che li rendono adatti a quel "livello" di linea/energia.

#### 5. CONCLUSIONI

Lo scopo dell'installazione è prevalentemente di tipo didattico. Possiamo dire che, pur con le difficoltà relative all'utilizzo del ghiaccio secco, abbiamo ottenuto un ottimo risultato. Possibili evoluzioni del sistema potrebbero consistere in un diverso modo di raffreddare la camera a nebbia e miglioramenti nel modulo di rilevamento delle tracce, realizzando un metodo ad hoc.

#### 6. RIFERIMENTI

- [1] [http://en.wikipedia.org/wiki/Cloud\\_chamber](http://en.wikipedia.org/wiki/Cloud_chamber)
- [2] Pelletier, J.: "A shape-based approach to computer vision musical performance systems" Proceedings of the 2004 Conference on New interfaces For Musical Expression.
- [3] Canny, J.: "A Computational Approach To Edge Detection" IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.
- [4] Duda, R. O. and P. E. Hart: "Use of the Hough Transformation to Detect Lines and Curves in Pictures" Comm. ACM, Vol. 15, pp. 11–15 (January, 1972)
- [5] Wright, M. and Freed, A.: "Open Sound Control: A New Protocol for Communicating with Sound Synthesizers" in International Computer Music Conference Proceedings, (Thessaloniki, Hellas, 1997).
- [6] Rohrhuber, J. and de Campo, A.: "Just in time programming" The SuperCollider Book. MIT Press, Cambridge, Massachusetts, 2011.
- [7] de Campo, A., Rohrhuber, J., and Boermann, T. and Frauenberger, C.: "Sonification and Auditory Display in SuperCollider" The SuperCollider Book. MIT Press, Cambridge, Massachusetts, 2011.

# SILOË

**Paolo Boscariol**

Ph.D, DIEGM, University of  
Udine, Italy  
[pao.lo.boscariol@uniud.it](mailto:pao.lo.boscariol@uniud.it)

**Alessandro De Nardi**

**Alessandro Fiorin Damiani**  
**Filippo Gecele**  
[infodenso@gmail.com](mailto:infodenso@gmail.com)

**Luca Lorenzo Ferro**

**Federico Petrei**  
[www.deltaprocess.it](http://www.deltaprocess.it)

**Chiara Olcese**

University of Ferrara, Italy  
[infodenso@gmail.com](mailto:infodenso@gmail.com)

## ABSTRACT

Siloë is a game. An audio-visual installation artwork in which participants are actively involved in a process of creation and renewal of sounds and lights, becoming themselves part of the artwork. Players interact with their own mobile devices or their bodies to control the sounds originating from twelve speakers located on the vertices of a dodecagon. A beamer on the top spot of the installation projects visuals that are modulated by the sound.

## 1. GENESIS

This installation is the result of two years of research and metaphysic speculations. In order to allow ourselves to sense an experience of natural perception, and to let it truly happen, we have to turn ourselves back into Nature. Siloë helps us in this mission by fostering the perception of time based on the instant, rather than on the duration, and through the refusal of the common concepts of time and original culture. The most common method to measure time is based on the solar system, and the number 12 is its most recurring expression. Twelve sub-divisions of shape, feature and quality.

Inspired by the concept of "friction", intended as sacrifice [1], Siloë reproduces these aspects in their most intimate essence, in their acoustic nature. Twelve acoustic entities, which synthesis is based on the Schumann's resonances [2], constitute the model of the "supreme order of things". Twelve independent voices that can reach their absolute unity. The loop, typical representation of the eternal return, expresses the continuous regeneration of an instant and at the same time annihilates the concept of time as duration.

## 2. RATIONALE

Siloë will also be used as a medical testing facility, in order to extend the studies introduced in [3]. The mentioned work presents the results of an experimental study on the effects of sound waves on the cortisol release of human bodies. Cortisol is a steroid hormone produced by the adrenal gland. It is released in response to stress and a low level of blood glucocorticoids. For this reason it is reputed to be a good indicator of the level of stress to which a patient is subject to.

It has been shown in [3] that half an hour of exposure to low frequency sound (20-100 Hz), both analogical or digital, reduces the cortisol level of 22-23% and it is judged pleasant by the testers, while the exposure to mixed analogical radio-waves produces a cortisol decrease of 16% even if it is considered unpleasant. Analogical high frequencies (8400-16000 Hz) raise cortisol values up to 50% and result unpleasant while the same session digitally recorded diminishes cortisol concentrations and is judged neutral, therefore the emotional status does not interfere with cortisol release, as much as the physical properties of sound-waves do. This pilot study was based on a small number of testers and it needs to be confirmed with a bigger group or with similar studies, with a larger participation of individuals over 50 years old. It will also be interesting to evaluate if there is a difference in the response if the session is in the morning, when the cortisol level is not expected to diminish.

## 3. THE INSTALLATION

From the results of the said study an installation was conceived. The installation will look like the rendering in Figure 1. The purpose of this work is to supply a device that may be employed for:

- therapeutic sessions
- audio/visual compositions
- interactive improvisations
- didactic activities

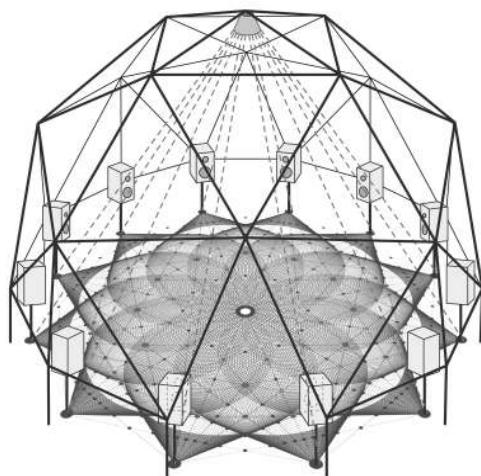
*Copyright: © 2012 C. Olcese. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

- social entertainment

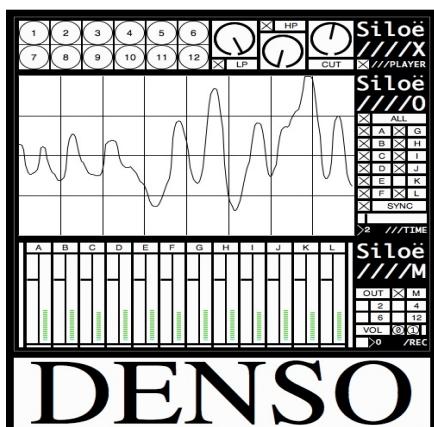
The universal language of geometry was chosen to abstract from any cultural roots. The acoustic nature was based on the Schumann's resonances [2] and the use of twelve entities [1].

#### 4. FORM OF INTERACTION

Participants will use their mobile devices and their bodies to control the synthesized audio and videos. There are twelve audio sources, corresponding to twelve tones, one for each active participant. Through the web interface that can be seen in Figure 2, she can control its timbre by selecting a variation or tweaking the filters. The overall produced audio is broadcasted via a web radio. The position of human bodies inside the circle affects the size, the color and the shape of the figures projected on the floor representing the relationship between the twelve tones and their modulations.



**Figure 1.** The structure of Siloë



**Figure 2.** The user interface for controlling the audio sources.

#### 5. PHYSICAL SPACE

The installation could be hosted inside a geodesic dome [4]. The choice of this structure was made in base of its relation with the dodecahedron and its number twelve. Also Poincaré proposed a dodecahedral space model of the universe [5].

However, instead of the dome, will be enough a space measuring not less than 6x6 meters and not less than 5 meters of height.

A virtual tour of the work is available at the URL [6].

#### 6. REFERENCES

- [1] M. Schneider: "L'idéologie de quelques instruments de musique populaire et le cycle de l'Année en Espagne", 1953.
- [2] E.R. Williams, et al.: "The Schumann resonance: A global tropical thermometer", *Science* 256 (5060), 1992.
- [3] C. Olcese, A. Fiorin Damiani, R. Dittadi, P. Borasio, L. Bartoloni, "Soundwaves' effect on hematic cortisol level: a pilot study" *Endocrine Abstracts* (2012) 29 p. 56
- [4] <http://www.geometer.org/mathcircle/s/geodesic.pdf>
- [5] <http://physicsworld.com/cws/article/e/news/18368>
- [6] <http://vimeo.com/denso/siloe-promo>