

Atti del XXI Colloquio di Informatica musicale
Proceedings of the XXI Colloquium of Musical Informatics

XXI Colloquio di
Informatica Musicale

<Extending Interactivity>

Cagliari 28 settembre - 1 ottobre 2016



« Extending interactivity »

Atti del XXI CIM - Colloquio di Informatica Musicale
Cagliari, 28 settembre–1 ottobre 2016

A cura di Anna Terzaroli e Andrea Valle

XXI CIM - Colloquio di Informatica Musicale
Cagliari, 28 settembre–1 ottobre 2016

AIMI – Associazione di Informatica Musicale Italiana - www.aimi-musica.org
Spaziomusica – www.spaziomusicaproject.com

Con il contributo di:

Regione Autonoma della Sardegna, Assessorato della Pubblica Istruzione, Beni Culturali, Informazione, Spettacolo e Sport

Comune di Cagliari, Assessorato alla Cultura, Pubblica Istruzione, Sport, Spettacolo e Politiche Giovanili

Con la collaborazione di:

Università di Cagliari, Dipartimento di Storia, Beni Culturali e Territorio

Conservatorio Statale di Musica Giovanni “Pierluigi da Palestrina” di Cagliari

Labimus

10 Nodi - I festival d'autunno a Cagliari

Atti del XXI CIM - Colloquio di Informatica Musicale

A cura di Anna Terzaroli e Andrea Valle

ISBN: 9788890341342

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

Riferimento BibTeX:

```
@proceedings{XVIIICIM,  
  Editor = {Anna Terzaroli and Andrea Valle},  
  Organization = {AIMI - Associazione Informatica Musicale Italiana},  
  Publisher = {DADI - Dip. Arti e Design Industriale. Università IUAV di Venezia},  
  Title = {Extending interactivity. Atti del XXI CIM - Colloquio di Informatica Musicale},  
  Year = {2016}}
```

Copyright

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



Coordinamento generale

Roberto Zanata

Coordinamento scientifico

Federico Avanzini
Stefano Delle Monache
Federico Fontana

Coordinamento programma musicale

Fabrizio Casti
Stefano Bassanese
Roberto Zanata

Comitato di programma

Gianpaolo Antongirolami, Conservatorio di Perugia
Federico Avanzini, Università di Padova / direttivo AIMI
Stefano Bassanese, Conservatorio di Torino / direttivo AIMI
Paolo Bravi, Conservatorio di Cagliari
Fabrizio Casti, Conservatorio di Cagliari / Associazione Spaziomusica
Paolo Dal Molin, Università di Cagliari
Stefano Delle Monache, IUAV di Venezia / direttivo AIMI
Federico Fontana, Università di Udine / direttivo AIMI
Marco Lutzu, Università di Firenze
Giorgio Klauer, Conservatorio di Padova / direttivo AIMI
Ignazio Macchiarella, Università di Cagliari
Anna Terzaroli, Conservatorio di Roma / direttivo AIMI
Andrea Valle, Università di Torino / direttivo AIMI
Roberto Zanata, Associazione Spaziomusica / Conservatorio di Bolzano

Comitato di revisione (programma musicale)

Andrea Agostini, Stefano Bassanese, Nicola Bernardini, Laura Bianchini, Paolo Bravi, Lelio Camilleri, Massimo Carlentini, Fabrizio Casti, Luigi Ceccarelli, Carmine Emanuele Cellà, Antonino Chiaramonte, Walter Cianciusi, Fabio Cifariello Ciardi, Alessandro Cipriani, Giovanni Cospito, Riccardo Dapelo, Maria Cristina De Amicis, Franco Degrassi, Fabio De Sanctis De Benedictis, Agostino Di Scipio, Marco Donnarumma, Francesco Galante, Daniele Ghisi, Francesco Giomi, Giorgio Klauer, Silvia Lanzalone, Carlo Laurenzi, Marco Liuni, Marcella Mandanici, Massimo Marchi, Marco Marinoni, Elio Martusciello, Alberto Novello, Alessandro Olla, Paolo Pachini, Lorenzo Pagliei, Simone Pappalardo, Pietro Polotti, Giorgio San-cristoforo, Riccardo Santoboni, Sylviane Sapir, Riccardo Sarti, Anna Terzaroli, Gianni Trovalusci, Andrea Valle, Gianluca Verlingieri, Massimiliano Viel, Andrea Vigani, Roberto Zanata, Paolo Zavagna.

Comitato di revisione (programma scientifico)

Giacomo Albert, Federico Altieri, Alessandro Altoè, Federico Avanzini, Stefano Baldan, Marco Braggion, Roberto Bresin, Antonio Camurri, Sergio Canazza, Luca Cossettini, Giovanni Costantini, Stefano D'Angelo, Amalia de Götzen, Giovanni De Poli, Alan Del Piccolo, Michele Del Prete, Stefano Delle Monache, Carlo Drioli, Federico Fontana, Alessio Gabriele, Leonardo Gabrielli, Michele Geronazzo, Bruno Giordano, Nicola Giosmin, Roberto Girolin, Vincenzo Lombardo, Marco Lutzu, Ignazio Macchiarella, Enrico Marchetto, Fabio Morreale, Maddalena Murari, Nicola Orio, Stefano Papetti, Niccolò Pretto, Davide Rocchesso, Antonio Rodà, Simonetta Sargentì, Stefania Serafin, Simone Spagnol, Alberto Torin, Andrea Valle, Domenico Vicinanza, Gualtiero Volpe, Stefano Zambon.

Programma musicale/Musical programme

Mercoledì 28 settembre

Concerto inaugurale

Giuseppe Orrù, Launeddas; Maureen Chowning, voce; Tenores di Bitti "Mialinu Pira"

Luca Richelli – *InterplaY* launeddas, per launeddas e live electronics

Giuseppe Orrù – Launeddas, repertorio della tradizione musicale della Sardegna

John Chowning – *Phoné*, per nastro magnetico

John Chowning – *Voces*, per soprano e computer interattivo

Roberto Zanata – *Tenores Live*, per tenores e live electronics

Tenores di Bitti "Mialinu Pira" – Repertorio della tradizione musicale della Sardegna

Giovedì 29 settembre

Sessione d'ascolto 1

John Chowning – *Turenas*, acusmatico

Sessione d'ascolto 2

Antonio Scarcia – *Harvest Fields*, acusmatico

Concerto 1

Enrico Di Felice, flauto; Simone Floris, clarinetto; Gianpaolo Antongirolami, sassofoni; Mario Carraro, pianoforte

Marco Marinoni – *Finita è la terra*, per pianoforte ed elettronica

Paolo Pastorino – *Dimensione aggiuntiva*, acusmatico

Roberto Doati – *Antidinamica*, da 1 a 4 sassofoni e live electronics

Joao Pedro Oliveira – *Et ignis involvens*, Audio/Video

Silvia Lanzalone – *eRose*, acusmatico

Venerdì 30 settembre

Sessione d'ascolto 3

Panayotis Kokoras – *Mutation* for clarinet and electronics
Simone Floris, clarinetto

Sessione d'ascolto 4

Roberto Cassano – *When the Iron Bird comes to light*, acusmatico

Concerto 2

Gianpaolo Antongirolami, sassofoni

Kosmas Giannoutakis – *Contraction point* per pianoforte e live electronics

Antonio D'Amato – *Körper*, acusmatico

Alfredo Ardia / Sandro L'Abbate – *Studio 1*, Audio/Video

Riccardo Dapelo – *Adaptive_study#02, Symbolic structures* per sassofono e live electronics

Dante Tanzi – *Double miroir*, acusmatico multicanale

Demian Rudel Rey – *Cenizas del tiempo*, Audio/Video

Giulio Colangelo – *Organismo aperto N° 2* per sassofono e live electronics

Sabato 1 ottobre

Sessione d'ascolto 5

Matteo Polato – *Et vous, que cherchez-vous?*, acusmatico

Concerto 3

Enrico Di Felice, flauto; Simone Floris, clarinetto; Gianpaolo Antongirolami, sassofoni; Mario Carraro, pianoforte

Marcela Pavia – *Risonanze*, acusmatico multicanale

Daniele Pozzi – *Breakpoint*, acusmatico

Sergio Missaglia – *Obscure matter*, Audio/Video

John Nichols – *Gates*, acusmatico multicanale

Filippo Guida – *Blank*, Audio/Video

Nicoletta Andreuccetti – *Before silence and after*, per flauto, clarinetto, pianoforte e live electronics

Giovedì 29 settembre - Sabato 1 ottobre

Installazioni

Roberto Musanti – *Festo*, Installazione interattiva multimediale

Roberto Dani – *EEG*, Installazione performativa

Programma scientifico/Scientific programme

Indice/Table of Contents

Papers

Riccardo Dapelo	9	Verso un'opera musicale adattiva (e interattiva)
Alberto Barberis, Sandro Mungianu, Sylviane Sapir	17	Ormé, a tool for automated spatialization of fixed-media music based on spectrum contents annotation
Stefano Delle Monache, Davide Rocchesso	25	Understanding cooperative sound design through linkographic analysis
Giorgio Klauer, Annalisa Metus, Pietro Polotti	33	Sonic interaction design e paper wearables
Fabio Morreale, Raul Masu	41	Renegotiating responsibilities in human-computer ensembles
Simone Spagnol, Charalampos Saitis, Kyriaki Kalimeri, Omar I. Jóhannesson, Rúnar Unnþórsson	47	Sonificazione di ostacoli come ausilio alla deambulazione di non vedenti
Fabio De Sanctis De Benedictis	55	PWGL come strumento per l'analisi musicale assistita: alcuni esempi
Francesco Roberto Dani, Giorgio Klauer, Antonio Rodà, Alvise Vidolin	63	L'espressività nell'esecuzione automatica di partiture di musica contemporanea: uno studio di caso
Andrea Valle, Amedeo Casella	69	Imaginary Landscape No. 4: study and annotation of the score
Michelangelo Lupone, Laura Bianchini, Alessio Gabriele, Silvia Lanzalone, Martina De Luca	77	Struttura, creazione, interazione, evoluzione: la ricerca al CRM per la realizzazione di Forme Immateriali di Michelangelo Lupone
Marco Lutzu	85	Un dispositivo per l'analisi del rapporto tra musica e gesto nelle launeddas
Paolo Bravi	92	The instrumental analysis of the singing voice: the case of harsh/rough/raspy voices
Lorenzo Vanelli	99	Software di analisi etnomusicologica: modelli di pensiero
Alessandro Russo, Valentina Burini, Sergio Canazza, Antonio Rodà	107	il caso dell'archivio storico del Teatro Regio di Parma: applicazione di una metododologia scientifica per la conservazione attiva di documenti sonori
Valentina Burini, Federico Altieri, Sergio Canazza	114	Rilevamenti sperimentali per la conservazione attiva dei documenti sonori su nastro magnetico: individuazione delle curve di equalizzazione
Michele Del Prete	122	La riduzione dello spazio. Di una gerarchia hegeliana nei primi scritti di Stockhausen

- Simone Spagnol, Silvio Galessio, Federico Avanzini **129** Stima di feature spettrali di HRTF mediante modelli antropometrici non lineari per la resa di audio 3D
- Federico Altieri, Marcella Mandanici, Sergio Canazza, Antonio Rodà, Emanuele Menegatti, Matteo Munaro **136** Good or bad? Potenziare l'ascolto attivo mediante spazi sonori interattivi
- Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, Federico Avanzini **143** Selfie acustiche con il progetto SelfEar: un'applicazione mobile per l'acquisizione a basso costo di pinna-related transfer function
- Federico Fontana, Federico Avanzini, Hanna Järveläinen, Stefano Papetti, Giorgio Klauer, Lorenzo Malavolta **151** Un prototipo di pianoforte digitale con feedback vibrotattile

Posters/Demo

- Renato Messina **159** Interpolazioni easing: implementazione e applicazioni audio
- Paolo Pastorino **164** Implementazione su Max MSP di Solo[Nr.19] - K. Stockhausen
- Marco Amistadi, Andrea Agostini **170** Sistemi interattivi in ambito artistico performativo. Mechanical Track Project
- Roberto Zanata, Federica Pau **176** Aesthetic and phenomenological approach to soundscape
- Marco Matteo Markidis, José Miguel Fernández **181** Analisi e sintesi in tempo reale mediante riconoscimento timbrico
- Anna Terzaroli **186** La notazione della Dissonanza
- Daniele Pozzi **191** Composing exploration: a multi-agent approach to corpus-based concatenative synthesis
- Andrea Cera, Davide Andrea Mauro, Davide Rocchesso **198** Sonic in(tro)spection by vocal sketching
- Edoardo Micheloni, Niccolò Pretto, Federico Avanzini, Sergio Canazza, Antonio Rodà **203** Installazioni interattive per la valorizzazione di strumenti musicali antichi: il flauto di Pan del Museo di Scienze Archeologiche e d'Arte dell'Università degli Studi di Padova
- Riccardo Franceschini, Giovanni Sparano **209** Dynaudiorouter: software open source per il routing audio dinamico in Max/MSP
- Sonia Cenceschi, Maria Teresa Trecozzi **212** BaO: the Baby Orchestra
- Alba Francesca Battista, Carlos Maximiliano Mollo **215** A sea of stars for KM3NeT: a multimodal virtual reality experience
- Marco Ferrazza **221** Modello riepilogativo delle principali attività organizzative in una composizione elettroacustica

Papers

VERSO UN'OPERA MUSICALE ADATTIVA (e INTERATTIVA)

Riccardo Dapelo

Conservatorio "G. Nicolini" Piacenza

rdapelo@libero.it

SOMMARIO

In questo articolo si delineano alcune ipotesi di implementazione di un'opera "adattiva" musicale per uno o più esecutori e sistema di live electronics. Tali ipotesi si fondano su una serie concreta di studi, inaugurati nel 2015, che hanno incluso tutti i parametri dell'opera, dal momento del concepimento alla realizzazione della partitura, dalle istruzioni per l'esecuzione alla concezione algoritmica del live electronics. L'obiettivo, innanzitutto compositivo e al tempo stesso algoritmico, era ed è quello di progettare e realizzare un'opera non definita/chiusa a livello temporale, capace di adattarsi al mutare di alcune condizioni durante la sua esecuzione.

1. INTRODUZIONE

Per progettare questo ciclo di lavori tale è stato necessario definirne i vari aspetti:

- la capacità di reagire agli stimoli forniti dall'esecutore;
- la simulazione di una "coscienza temporale" (memoria del tempo trascorso durante la performance);
- la manipolazione di strutture simboliche (frammenti di partiture);
- la abilità di operare scelte sulla base di quanto eseguito (o prodotto dal sistema) fino al momento attuale;
- il mantenimento del controllo della densità massima, senza degenerazione entropica.

Restava da definire in quale modo e se fosse possibile che la scrittura stessa potesse venir coinvolta nel processo dalla sua fondazione. Pertanto con una tecnica di retro-dizione, analizzando le conclusioni desiderate sono state definite le condizioni tecniche necessarie per la realizzazione.

Copyright: © 2016 Riccardo Dapelo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1.1 I precedenti

A partire dagli anni 60 l'idea di un'opera musicale aperta, mutevole e che potesse cambiare ad ogni esecuzione ha avuto molti sostenitori [1] e sperimentatori [2]. Ovviamente la "apertura" si poteva verificare in sede compositiva e/o performativa, delegando agli esecutori l'assemblaggio finale di una serie di indicazioni compositive lasciate più o meno velatamente (o volutamente) indefinite. Ma l'indefinizione in sé non contiene il nucleo dell'opera aperta, né appare sufficiente a risolvere il problema (il compositore aveva comunque predeterminato il campo delle azioni possibili).

Sicuramente l'approccio stocastico, probabilistico (comunemente definito aleatorio) può liberare l'opera musicale dal carico simbolico della narrazione,¹ ma non riesce spesso a scongiurare un valore molto alto di imprevedibilità percettiva e una certa impossibilità di controllo del campo mini e meso-strutturale.²

Scrive K. Stockhausen a proposito della *moment-form*:

"In queste forme un minimo o un massimo può essere previsto in ogni momento, e nessuna direzione di sviluppo può essere prevista con certezza dato il momento attuale; hanno sempre già cominciato, e potrebbero continuare per sempre"[2].

In queste affermazioni emerge uno dei problemi che hanno afflitto alcuni tentativi di opera aperta dell'epoca: l'eccesso di imprevedibilità. Questo eccesso di imprevedibilità può peraltro riflettersi in maniera critica sia nella percezione sia nella scrittura.³ Visto lo stadio attuale della composizione algoritmica sarebbe relativamente semplice creare un sistema algoritmico-stocastico con le caratteristiche percettive (ma anche formali) descritte da Stockhausen, cioè un sistema dato di proporzioni

¹...è un approccio compositivo in cui è volutamente evitata una linea narrativa complessiva. I momenti componenti in tale forma sono legati da un principio non lineare di proporzioni. Se questo sistema di proporzioni esaurisce una serie di possibilità, la forma è definita 'chiusa'; se no, o se la serie di proporzioni non è finita, allora la forma è 'aperta'. [2]

²I. Xenakis distingue in *Formalized Music* tra microstruttura (= timbro), ministruttura (= nota), mesostruttura (= poliritmo, scale melodiche di intensità), macrostruttura (= evoluzione globale dell'ordine di decine di minuti).[3]

³Se, come illustrato successivamente, si può definire "stile" una invarianza organizzativa del materiale sonoro, con questo tipo di comportamento si può ottenere una variazione totale e continuata, che potrebbe essere definita "stile" a sua volta e che però sfugge al controllo percettivo del compositore e dell'ascoltatore.

temporali che regola eventi non correlati tra loro. D'altro canto i sistemi di live electronics si sono sempre più evoluti verso la massima e ampia reattività, abituandoci a livelli sempre più sofisticati di interazione diretta. L'unica mancanza di questi sistemi consiste nell'assenza di capacità semantiche, di interpretazione di ciò che avviene, sia in un singolo evento sia soprattutto a livello di flusso temporale.

Al giorno d'oggi, con lo sviluppo delle tecnologie di sensori, realtà espansa, aumentata ecc, su cui ci invita a riflettere anche il titolo di questo CIM - Estensione dell'interattività -, la prospettiva di introdurre ed esplorare capacità semantiche di un sistema sembra allettante e praticabile.

2. DEFINIRE L'ADATTIVITÀ

A seguito di queste considerazioni la ricerca è consistita nel progettare un'opera aperta con le seguenti caratteristiche:

- adattiva (in grado cioè di adattarsi a certe condizioni esterne e/o ai loro mutamenti), in cui però sia possibile il controllo di determinati aspetti temporali, (in primis, ma non solo, la modulazione della densità degli eventi nel micro e macro livello);
- che consenta e realizzi ciò che M. Lupone definisce "stile" (invarianza organizzativa del materiale sonoro [4]);
- interattiva, ossia in forma dialogante con uno o più esecutori;
- in cui le azioni possibili, sia da parte dell'esecutore sia da parte del sistema, non siano predeterminate; altrimenti si tratterebbe semplicemente di assegnare un coefficiente di probabilità a una delle diverse possibili soluzioni;
- che sia in grado di reagire a uno stimolo diretto e simuli un certo grado di memoria del passato prossimo (attraverso la scrittura?);
- che sia in grado manipolare strutture simboliche (frammenti di partitura sia in input che in output);
- che sia in grado di rilevare alcuni comportamenti dell'esecutore (ad es. la dinamica media, la densità di eventi prodotti ecc);

Nell'affrontare il percorso, trattandosi di un tale livello di complessità, si è preferito optare per lo studio, ossia una forma in cui poter sperimentare un sistema di ipotesi, test, retro-dizione, visti i limiti metodologici che l'opera d'arte presenta.⁴

⁴ Nel caso dell'arte si pone un problema metodologico bidirezionale: una ipotesi elegante e/o efficiente non garantisce un risultato (estetico) equivalente e/o controllabile e d'altronde un risultato eccellente non sempre può essere formalizzato o replicato. La predicitività del risultato non può essere stabilita, se non a posteriori. Scrive a proposito

Il percorso di studio e progettazione è iniziato con Adaptive_Study#01-Lexicon⁵ e proseguito con Adaptive_Study#02-Symbolic structures per Sax Contralto e Live electronics, qui di seguito descritto.⁶

2.1 Suggerimenti della biologia evolutiva

La riflessione su queste problematiche e la ricerca di una possibile metodologia in grado di unificare la realizzazione compositiva e la sua declinazione temporale attraverso il live electronics, ha portato ad esplorare e ricercare aspetti della biologia evolutiva che potessero rappresentare una fenomenologia applicabile al caso in oggetto o che, a proposito, si potessero "adattare" al progetto. Metodologicamente si può partire dalla definizione di uno o più risultati per selezionare e ricavare i processi più efficaci/efficienti. In questo caso volendo progettare un sistema che dovesse operare in input un riconoscimento e successivo concatenamento di strutture musicali (di tipo monodico, almeno per i primi passi della ricerca) sono stati mutuati alcuni concetti dalla biologia evolutiva, specificamente per trovare regole di connessione e selezione di microstrutture. Spesso nella trasmigrazione di concetti i sistemi viventi presentano caratteristiche utili e stimolanti sul piano artistico.⁷ La sperimentazione di modelli si è orientata verso una pluralità di soluzioni alternative e concomitanti, senza concentrarsi su una unica soluzione o via ottimale, cercando di cogliere i medesimi aspetti da entrambi i lati del problema, l'aspetto della elaborazione elettronica e quello della scrittura.⁸

2.2 Adattamento

Il termine **adattamento** [7] in biologia si riferisce alla facoltà degli organismi viventi di mutare i propri processi metabolici, fisiologici e comportamentali, consentendo loro di adattarsi alle condizioni dell'ambiente nel quale vivono. In particolare, l'adattamento è riferito alle modificazioni di una struttura anatomica, un processo fisiologico o un tratto comportamentale di un organismo che si è evoluto in un certo periodo di tempo come

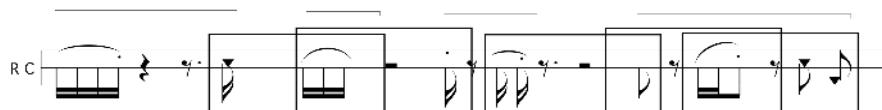
P. Veyne: "opera d'arte che per quanto oggettiva, non ha metodo e tanto meno è scientifica. Lungi dallo sboccare in una scienza o in una tipologia, essa non cessa di confermare che l'uomo è materia variabile sulla quale è impossibile pronunciare un giudizio permanente". [5]

⁵per Sax baritono e live electronics (Eseguito con G. Antongirolami a Piacenza nel 2015).

⁶Il prossimo step sarà Adaptive_Study#03_Lsystems per cello, pfte, live e immagini di prossima realizzazione (novembre 2016).

⁷A tale proposito si cita un estratto dal programma di artescienza 2016 del CRM: " l'Arte e la Scienza connessi attraverso l'intuizione, la migrazione dei concetti, le tecnologie e i metodi sperimentali, rispondono alle esigenze di una percezione più sofisticata e complessa della realtà. La bellezza, l'emozione, la comprensione e il sapere sono gli elementi immediatamente riconoscibili della loro sinergia, sono gli elementi che percepiamo come dono per la nostra intelligenza o sensibilità, [...]".

⁸In particolare in questa fase è risultato utile il modello sperimentazione-decimazione-diversificazione-riduzione secondo cui si crea una radiazione iniziale e si procede poi per adattamenti successivi alla selezione di un ristretto numero finale di candidati. [6]



ESEMPI di possibili Exaptation sul ritmo (raggruppamenti di prossimità diversi dalla lettura iniziale)

Figura 1. exaptations sulla prossimità temporale

mente cooptate per produrre suoni e si sono trasformate



ESEMPI di possibili Exaptation sulle altezze (raggruppamenti di prossimità frequenziale)

Figura 2. exaptations sulle altezze. Chunk generato dalla combinazione di lettura 1 e lettura 2 (cfr. Fig 4)

effetto della selezione naturale, in maniera tale da aumentare il successo riproduttivo di tale organismo (nelle determinate condizioni ambientali in cui si trova).

Nel cercare una riverberazione musicale di questo concetto sono stati presi in esame alcuni parametri che potevano essere strutturalmente significativi: in un adattamento diretto (inizialmente a livello piuttosto rudimentale) la valutazione della velocità di esecuzione (valore medio delle durate delle note trascritte in input) in un dato intervallo temporale influenza la velocità delle sequenze proposte dal sistema in output;⁹ in seguito sono state esplorate possibilità di adattamento più raffinate, valutando la densità media di un certo intervallo temporale. Il sistema mentre trascrive non è attualmente in grado di valutare la velocità/densità di esecuzione, se non con un certo grado di ritardo, il tempo fisiologico/algoritmico della formazione (ossia trascrizione) dell'immagine percettiva, che sarà poi accessibile come ricordo. [8] Attualmente la velocità di esecuzione delle sequenze in output ha una possibilità di aggiustamento manuale.¹⁰ In una accezione di livello superiore l'adattamento dovrà anche stabilire la percentuale di successo delle diversificazioni ottenute mediante **exaptations**.

2.3 Exaptation¹¹

Nell'exaptation un carattere evoluto per una particolare funzione ne assume una nuova, indipendente dalla primitiva: un classico esempio è costituito dalle piume degli uccelli, evolute dai dinosauri presumibilmente per scopi di isolamento termico e poi rivelatesi utilissime per il volo, oppure il primitivo polmone che si è evoluto dalla vescica natatoria dei pesci. Nella specie umana, le pieghe laringee, comparse per impedire che il rigurgito del cibo entrasse nei polmoni, sono state successivamente

⁹ Questa operazione di valutazione si aggiunge al ritardo iniziale dovuto al primo segmento di trascrizione.

¹⁰ Nelle fasi di studio l'obiettivo è comunque di ottenere un risultato artisticamente soddisfacente, pertanto sono stati inseriti aggiustamenti arbitrari ove necessario.

¹¹ Il **preadattamento** (o exattamento) è un concetto utilizzato per descrivere un particolare tipo di evoluzione delle caratteristiche degli esseri viventi; dall'espressione inglese *exaptation* introdotta da Stephen Jay Gould ed E.S. Vrba. [9]

nelle corde vocali, pur mantenendo la loro funzione originaria. Nel caso musicale le exaptations sono state implementate in scrittura variando la distanza e/o il raggruppamento tra alcune figure ritmiche, applicate alla lettura dei frammenti (Fig.1), oppure ad altezze continue di differenti figure (Fig.2), creando dei nuovi raggruppamenti di prossimità temporale o frequenziale.

È quindi possibile definire exaptation musicale, almeno a questo stadio del lavoro, la creazione arbitraria di raggruppamenti di prossimità temporale e/o frequenziale.¹² In questo modo si può esplorare il possibile adiacente, ossia l'insieme degli stati potenziali che distano un solo step dallo stato attuale. Allo stesso modo gli algoritmi di live electronics operano dei raggruppamenti simili sulle strutture da interpretare, trascritte in frammenti di partitura.

3. TECNICHE DI SCRITTURA

Dati i presupposti indicati nel paragrafo precedente per la scrittura musicale il punto di partenza è stato la conclusione, ossia il tentativo di ottenere, a partire dal materiale iniziale che ancora non esisteva¹³, una potenzialità di radiazione percettiva del materiale musicale che potesse assicurare il controllo dello stile. I requisiti per la generazione di mini-strutture (unità di articolazione) [10, 11] devono presentare:

- una alta capacità combinatoria;
- una dimensione relativamente breve;
- un certo numero di ripetizioni di alcune altezze;
- una capacità (ovviamente arbitraria) di costituirsi in "figure" percettivamente definibili e/o riconoscibili;
- la capacità delle "figure" di unirsi in "chunks", sequenze complesse di ordine superiore; [12]

Successivamente è stato elaborato un piano delle possibili conseguenze di ogni singolo tratto.¹⁴

¹² Lo studio preliminare della costruzione compositiva dei frammenti ovviamente favorisce la formazione di queste exaptations (cfr. Fig.3).

¹³ Scandagliare le dinamiche di ciò che ancora non è.

¹⁴ In senso gestaltico, di buona e/o possibile continuazione.

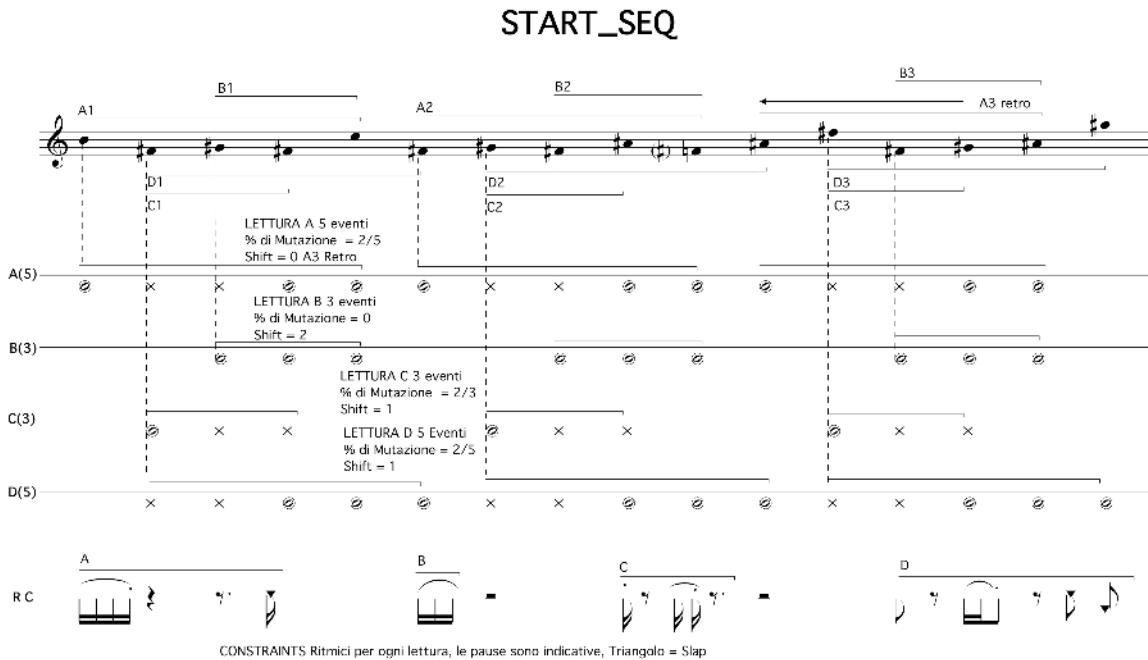


Figura 3. La matrice di generazione dei materiali

Il punto di partenza è stata una sequenza di altezze generata da un algoritmo di uno dei primi software di Computer Aided Composition¹⁵ [13] che ha delle caratteristiche morfologiche¹⁶ che soddisfano potenzialmente quanto sopra espresso, sulla quale sono state effettuate:

- quattro diverse letture della sequenza con tecniche di segmentazione indicate nella Fig.3,¹⁷ con segmenti che hanno indici di mutazione (possibilità di sostituire l'altezza originale) diversi, ottenendo tre differenti "sequenze"(Fig.4), tramite l'applicazione di uno strato di constraints (figure ritmiche obbligate);¹⁸

- una serie di mutazioni ottenute sovrapponendo le varie "sequenze/chunks", alternando un frammento di una lettura con uno dell'altra (1A, 2A, 1B, 2B ecc), sfruttando i silenzi preventivamente inseriti nelle sequenze tramite i constraints ritmici (la sequenza generata dalla sovrapposizione delle letture 1 e 2 è visibile in Fig.2);
- una ricorsività generale ottenuta riapplicando la mutazione sopra descritta;¹⁹

A questo punto della elaborazione il materiale generato presenta un grado soddisfacente di modularità adattativa e si presta ad una elaborazione confacente alle premesse, realizzando così la condizione posta all'inizio. Le operazioni di manipolazione si sono limitate a tratti



¹⁵ Open Music in una delle sue prime versioni. Non è stato possibile ricostruirne le modalità generative.

¹⁶ Si tratta evidentemente di una salita melodica difettiva dal SI3 al FA# 4, alternata con una polarizzazione attorno al FA# 3 e altezze limate.

¹⁷I numeri tra parentesi indicano il numero di step di ogni lettura, lo shift la distanza dall'inizio, le letture con X la possibilità di mutazione dell'altezza originale.

¹⁸Tecnica assimilabile alla *talea* del XV secolo, ripetizione sistematica dello stesso schema ritmico, su serie di altezze spesso non congruenti nel numero di componenti. [14]

¹⁹Dopo la sovrapposizione della lettura (1+2) con la lettura 3 si ottengono dei resti (altezze non entrate nel sistema per mancanza di spazi ritmici). A tali resti sono state riapplicate le sovrapposizioni per un totale di 24 sequenze.

molto semplici, confidando nella consapevolezza che spesso si creano sistemi complessi a partire da regole semplici, come nel caso della iterazione di funzioni non

lineari.²⁰ Inoltre nel caso specifico la semplicità strutturale dei singoli frammenti (unità di articolazione) era condizione espressamente richiesta e ricercata.

3.1 Assemblaggio

Il materiale generato dalle tecniche appena descritte è stato diviso in una serie di frammenti combinabili a piacere dall'esecutore,²¹ in cui si realizza comunque un certo grado di invarianza stilistica, mantenendo il sistema doppiamente aperto, in quanto all'esecutore è richiesto, dopo aver eseguito il primo frammento (unico determinato insieme all'ultimo) di realizzare la radiazione iniziale del materiale in completa autonomia, e successivamente, ascoltando le risposte del sistema, di cambiare atteggiamento adattandosi a sua volta nel corso dell'esecuzione.

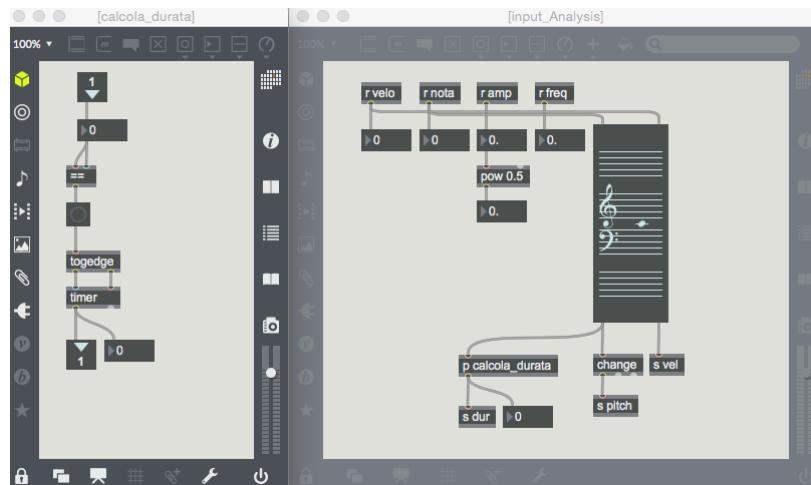


Figura 5. input analysis in MAXMsp, le informazioni su frequenza e ampiezza provengono dall'oggetto *return*²² opportunamente tarato e configurato.

4. LIVE ELECTRONICS

La libreria *bach* e il set di astrazioni *cage* [15,16] di A. Agostini e D. Ghisi sono state un contributo fondamentale per poter realizzare questo progetto in ambiente MAXMsp. L'ipotesi di affrontare un progetto di Computer Aided Composition in tempo reale compie un ulteriore passo avanti e quello che nella non lontana epoca del lavoro di G. Grisey all'Ircam doveva essere fatto in

²⁰Si ritiene che un sistema di generazione di frammenti musicali si possa definire efficace se risulta convincente sul piano percettivo (seppur arbitrario) e non necessariamente per la complessità delle sue regole.

²¹ Il totale è impaginato in modo simile al Klavierstück XI di K. Stockhausen. L'esecutore può scegliere e combinare liberamente frammenti brevi, medi e lunghi ed una serie di gesti strumentali (multifonici e rumori) che servono per interrompere la trascrizione del sistema e creare nuove condizioni di adattamento.

²²In questo e nei successivi paragrafi i termini indicati con ~ si riferiscono all'implementazione MAXMsp.

tre differenti passaggi (analisi, quantizzazione, riscrittura – quest'ultimo passaggio anche in modo manuale) può al giorno d'oggi essere affrontato nella dimensione *real time*

4.1 Struttura dell'algoritmo

L'input (monofonico in questi primi studi) viene analizzato per la stima dell'altezza e contestualmente ne viene calcolata la durata (Fig.5); l'array viene inviato ad una struttura di oggetti della libreria *bach* (tra cui è fondamentale *bach.transcribe*) (Fig.6). Il primo problema tecnico che si pone in questo stadio riguarda la segmentazione dell'ingresso. Per necessità di buffer e per poter applicare gli algoritmi di elaborazione lo stream di ingresso deve essere necessariamente segmentato. Nella prima versione si è optato per la segmentazione manuale.²³ In questo secondo studio viene sperimentata la segmentazione automatica sulla base di un valore delta di

silenzio. Resta ancora da risolvere l'alternativa tra una lunghezza costante dei frammenti e la segmentazione dell'input sulla base di pause.

A questo punto interviene il processo di quantizzazione che viene gestito dalla libreria *bach* in modo molto sofisticato e complesso, potendo controllare anche le reazioni alla densità nei vari slot in cui la trascrizione viene segmentata.

Lo stream segmentato in sequenze viene memorizzato in strati successivi. La registrazione in immagini successive può rappresentare un minimo di coscienza temporale, a livello di una Short Term Memory; ogni immagine, segmentata dall'esecutore al live o automaticamente, è una sorta di snapshot di un breve momento esecutivo. Al termine di ogni intervallo di registrazione la sequenza viene trascritta in forma di partitura e archiviata nella struttura dati *coll* (Fig.7). In questa implementazione viene calcolata una media globale e locale (per ogni snapshot) delle durate e delle dinamiche e tale

²³Non essendo possibile risolvere tutti i problemi al primo step si è scelto di risolverli con soluzioni temporanee da approfondire negli stadi/studi successivi.

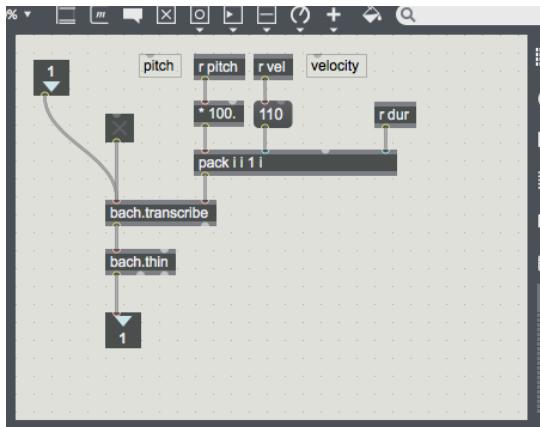


Figura 6. la struttura del trascrittore di input

valore viene aggiunto come flag nella struttura dati.²⁴ In questo punto è inserita una variabile (shift amount visibile in Fig.8) che permette la trasposizione in altezza della sequenza prima della sua memorizzazione. Il parametro è a controllo manuale ma potrebbe esser mappato

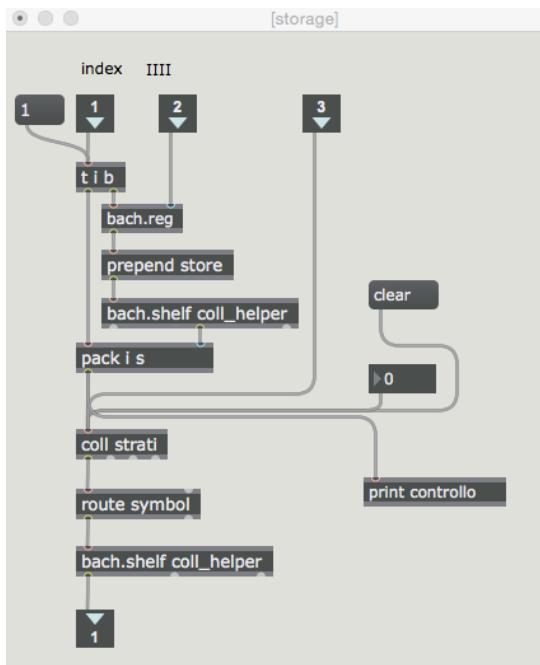


Figura 7. la struttura di memorizzazione delle sequenze

ad una successiva analisi di comportamenti rilevati dall'osservazione dell'esecutore. Un contatore tiene conto del numero progressivo di sequenze memorizzate, che possono essere richiamate individualmente in qualsiasi ordine. Ovviamente anche l'ordine delle sequenze può essere scelto dall'interprete al live electronics o può essere affidato ad un algoritmo pseudocasuale.

²⁴In modo che ad esempio al verificarsi di una certa condizione si possa richiamare una struttura simbolica di grado differente di densità o dinamica

4.2 Filtraggio e exaptations

L'ultimo oggetto *bach.score* (ultimo pentagramma di Fig.8) mostra l'applicazione di una exaptation alla sequenza memorizzata: tramite l'astrazione *cage.cascade~* alla sequenza richiamata dallo storage viene applicato un filtro passa banda simbolico, riducendo la sequenza ad altezze all'interno di una determinato range. Tramite l'applicazione di questi filtri simbolici in parallelo vengono realizzate le exaptations di tipo frequenziale.²⁵ Parallelamente vengono effettuate interpolazioni ritmiche (*cage.rhythminterp*) tra frammenti di sequenze per ottenere le exaptations ritmiche. Questi trattamenti per adesso vengono memorizzati in una diversa struttura dati e richiamati a parte.

È in corso di valutazione una opzione algoritmica per sostituire o integrare le mutazioni all'interno della struttura dati principale (locking-in) [17],



Figura 8. la struttura del manipolatore di strutture simboliche (frammenti di partitura)

²⁵E secondo il principio riduzione-decimazione-diversificazione il risultato in output, rispetto a n voci in cui si può dividere l'input, può essere una sola linea e/o una qualsiasi combinazione possibile. Applicando un principio di retroazione, si amplificano le caratteristiche di un tratto a capito di altri.

ossia riscrivere la snapshot nella struttura dati qualora sia stato applicato un certo grado di mutazione.

4.3 Strutture simboliche in output

Tutti i frammenti generati, sia quelli di trascrizione di ingresso sia quelli di mutazione delle trascrizioni vengono riproposti all'esecutore. Al termine della catena di *bach.score* si trova un sampler software che gestisce una libreria di campioni di sax di alta qualità. La velocità di lettura dei frammenti è un aspetto critico, ed è una delle variabili di adattamento del sistema (deve essere valutata attentamente la prevedibilità della risposta, ad es rispondere con lentezza ad una determinata modalità esecutiva).

Il termine del playback di una sequenza o di una sua mutazione attiva il playback di un'altra sequenza, ma l'intervallo di tempo che intercorre dipende dal comportamento dell'esecutore: ad es una grande densità di eventi suonati mette il sistema in attesa per il playback, mentre la trascrizione dell'input resta sempre attiva.

4.4 Diffusione spaziale

La dislocazione spaziale delle risposte del sistema è stata implementata (su impianto quadriphonico) sulla base della valutazione della percentuale di suoni lunghi prodotti dall'esecutore: percentuale maggiore risposta lontana, verso il fondo della sala, percentuale minore (predominanza di suoni corti) risposta vicina all'esecutore. Bisogna rilevare che la caratteristica dei frammenti da assemblare a cura dell'esecutore è principalmente a dominanza di suoni corti. Sono stati inseriti solo pochi "gesti" lunghi e dirompenti, con lo scopo di interrompere il flusso delle figure e "allontanare" le risposte del sistema. Questi "segnali", come precedentemente accennato, hanno anche la funzione di definire una sezione formale degli algoritmi di live electronics.

5. CONCLUSIONI E SVILUPPI FUTURI

5.1 Approfondimenti necessari

Al livello attuale di sviluppo restano aperte una serie di questioni che necessitano di ulteriori approfondimenti e procedure di test:

- l'applicazione di tecniche di feedback sulla trascrizione in ingresso (*locking-in*), in modo da amplificare i fenomeni emergenti all'interno dello stream musicale, dovrebbe enfatizzare una struttura che si auto stabilizza e le cui origini malgrado la semplicità dell'anello di feedback sono di fatto impenetrabili [16].
- l'applicazione di tecniche di feedback sulle exaptations di tipo frequenziale/o ritmico, in modo da selezionare e amplificare solo alcuni tratti distintivi delle figure trascritte.²⁶

²⁶Come già accennato in precedenza inserendo le modifiche in sostituzione nelle varie snapshot.

- un progressivo raffinamento delle regole interpretative per separare/definire i vari tratti e conseguentemente il tipo di segmentazione da applicare. Attualmente come descritto in precedenza si applicano elementari raggruppamenti di prossimità frequenziale e temporale.
- Un approfondimento dell'analisi qualitativa delle sequenze trascritte.
- La definizione di un repertorio di simboli attivabili (e riconoscibili dal sistema). Per questo occorre definire un simbolo e tutte le sue condizioni attinenti e determinanti [17,18,19].
- La definizione e catalogazione delle operazioni di constraints programming e predici condizionali che possono far prendere decisioni al sistema.

5.2 Questioni aperte

Lo scopo di questo articolo è di riportare le esperienze affrontate nella progettazione e nell'implementazione di un'opera musicale. Il risultato finale che si può sottoporre a giudizio è l'opera in sé. Come riferito dal titolo questo lavoro è un primo passo verso l'idea di opera adattiva. Si potrebbe definire uno studio di Computer Aided Composition in tempo reale, con caratteristiche di adattività. Resta sul tavolo la questione aperta più spinosa: controllare lo sviluppo temporale.

Può un sistema riuscire a passare dall'universo materiale delle unità di articolazione a quello delle forme? [3] Si può riuscire a creare un sistema completamente autonomo nel prendere decisioni di tipo compositivo/artistico o serve la decisione umana?

Quale può essere il livello di percezione temporale (memoria) del sistema?

Si può aggiungere al sistema un grado di consapevolezza temporale e coniugare le strutture nel tempo (la generazione dei micro eventi) e quelle fuori tempo (le articolazioni formali che si adattano alle mutate condizioni)?

5.3 Sviluppi futuri

Le prossime aggiunte previste e immaginate per il sistema sono:

- un approfondimento dell'analisi spettrale in input, con un sistema di trascrizione di immagini spettrali simile a quanto illustrato per i sistemi di altezze e durate. La accoppiata di oggetti come *sigmund*²⁷ e *bach.transcribe* ha mostrato risultati lusinghieri. Sarebbe piuttosto interessante realizzare una memoria di immagini spettrali da richiamare e trasformare, per realizzare in tempo reale una sorta di migrazione dal dominio dello spettro a quello della scrittura.

²⁷Analizzatore di spettro di M. Puckette

- ra, simile a quella applicata da G. Grisey ad esempio in *Partiels*.
- La aggiunta di un sistema di videocamera e/o kinect per rilevare anche le informazioni di movimento dell'esecutore e poter introdurre nell'analisi anche i dati di movimento, posizione ecc.
 - La generazione di una partitura al termine dell'esecuzione. Ipotesi che richiede un lavoro aggiuntivo di raffinamento per perfezionare la trascrizione in input e le procedure di quantizzazione.

6. BIBLIOGRAFIA

- [1] U. Eco, *Opera aperta*, Bompiani, Milano, 1962/2006.
- [2] K. Stockhausen. "Momentform: Neue Beziehungen zwischen Aufführungsdauer, Werkdauer und Moment". In his *Texte zur Musik*, vol. 1, pp. 189-210, DuMont Schauberg, Cologne, 1963.
- [3] Iannis Xenakis, *Formalized Music*, Indiana University Press, Bloomington, 1971.
- [4] A. Gabriele, *Ladop1: Definizione e Computazione degli adattamenti in opere d'arte intermediali e interattive* in Atti del XX CIM, pp.75-80, Roma, 2014.
- [5] P. Veyne, *Come si scrive la storia*, Laterza, Milano, 1973.
- [6] T. Pievani, *Introduzione alla filosofia della biologia*, Laterza, Bari, 2005.
- [7] T. Pievani, *La vita inaspettata*, Raffaello Cortina, Milano, 2011.
- [8] H. Bergson, *Materia e Memoria*, Laterza, Bari, 1996.
- [9] S.J.Gould, E.S.Vrba, *Exaptation. Il bricolage dell'evoluzione*, a cura di T. Pievani, tr. It Bollati Boringhieri, Torino, 2008.
- [10] F.Oppo, *Per una teoria generale del linguaggio musicale*, in Quaderni della Rivista Italiana di Musicologia, a cura di M. Baroni e L. Callegari, Modena, 1982
- [11] N. Chomsky, *Syntactic Structures*, Mouton & co, The Hague, 1957
- [12] D. Coon, J. O. Mitterer, *Psicologia Generale*, Utet, Torino, 2011
- [13] F. De Sanctis De Benedictis, *Dall'analisi musicale alla composizione e formalizzazione algoritmica: esempi applicativi con PWGL*, in Atti del XX CIM, pp.11-14, Roma, 2014.
- [14] A. F. Gallo, *La polifonia del Medioevo vol.3*, EDT, Torino, 1996
- [15] A. Agostini and D. Ghisi *bach: una libreria per la composizione assistita in Max* in Atti del XIX CIM, pp.150-157, Trieste, 2012.
- [16] A. Agostini, E. Daubresse, D. Ghisi *cage: una libreria di alto livello per la composizione assistita da computer in tempo reale* in Atti del XX CIM, pp.17-22, Roma, 2014.
- [17] D. Hofstadter *Anelli nell'io*, Mondadori, Milano, 2008.
- [18] R. Jakobson *Saggi di linguistica*, Feltrinelli, Milano, 2002.
- [19] U. Eco, *Apocalittici e integrati*, Bompiani, Milano, 1962/2006.
- [20] N. Ruwet, *introduction a la grammaire generative*, Plon, Paris, 1967

ORMÉ, A TOOL FOR AUTOMATED SPATIALIZATION OF FIXED-MEDIA MUSIC BASED ON SPECTRUM CONTENTS ANNOTATION

Alberto Barberis

Conservatorio della Svizzera italiana
Dep. Composition and Theory
Lugano (CH)
albertobarberis11@gmail.com

Sandro Mungianu

Conservatorio della Svizzera italiana
Dep. Composition and Theory
Lugano (CH)
sandromungianu@gmail.com

Sylviane Sapir

Conservatorio "G. Verdi" di Milano
Dep. of Music and New Technologies
Milano (IT)
sylviane.sapir@consmilano.it

ABSTRACT

This paper describes the ongoing development of Orm  , a spatialization tool developed with Max/MSP, which enables to interpret monophonic or stereophonic fixed-media music by means of sound projection. Our intent is, following a detailed spectral analysis, to dynamically and automatically split the sound in various spectral bands that will be sent to a multi-channel system in order to highlight the intimate sound characteristics as well as the structural aspects of the interpreted piece of music. The spatialization technique used in Orm   has mainly performative and analytical objectives, but the system also provides a sonogram annotation syntax that allows for a real-time representation of the spectral and spatial elements involved in the performance which is particularly suitable for didactic purposes, especially for music analysis or/and ear-training.

1. INTRODUCTION

According to the musical community, since electroacoustic, and more generally fixed-media music, is usually not supported by scores, its analysis has to involve either esthetic or poietic approaches [20]. The former, which is mainly based on listening activities, could be more suited for interpretative purposes. Among the various analytical methods, which deal with listening and perceptual cues, we selected the analysis founded on the principles of spectromorphology [22]. Denis Smalley's spectromorphology is a wide and very articulated theory, which investigates on sound by observing its spectrum through time and by defining shapes and motions within the spectrum. The different movements and growth processes described by Smalley evolve both in the spectral dimension and in the spatial one. Space, in literal and musical terms, is a founding concept in Smalley's approach and thus the starting point of our work. The whole idea is to isolate pertinent

Copyright: © 2016 Barberis 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.

morphologies derived from the spectral analysis through dynamic filtering techniques and then to separately spatialize them coherently with the 'intrinsic space' of music.

Our use of spatialization is very specific and is not concerned with the common practice of space processing. Indeed, a study on the use of spatialization by composers and sonic artists, has shown that the majority of the interviewed musicians takes advantage of spatialization in concert, in order to enhance the listening experience by immersing the audience in sound [16]. This highlights the dominant usage of spatialization: an effect or a post-production process in a 'multi-track metaphor'. Nowadays many spatialization systems are still mixer-oriented; this is because many masterpieces were composed for audio systems controlled by a mixer, and the history of electronic music's interpretation is strongly related with the sound-space aesthetic developed in the acousmoniums. Fewer composers yet consider space as an integral part of the main compositional parameters, even if very interesting projects, such as OMPrisma for OpenMusic [22] or the bach library for Max/MSP [1], demonstrate that a good integration of spatialization in computer-aided composition environments may improve spatial sound synthesis and symbolic control of sound source spatialization.

The above noted study [16] also indicates that a minority of musicians identified spatialization as a mean to investigate segregation and sound blending as key concepts of auditory scene analysis [5]. This last point demonstrates how few composers are aware that spatial aspects are perceptual attributes of sound, and that current spatialization techniques allow to intimately investigate into both spectral and spatial dimensions of timbre: a great potential for music composition and interpretation. Some noteworthy works in that direction may be mentioned: the Torchia and Lippe's approach which combine spectral analysis and resynthesis for spatial distribution [26]; the use of spectral delay for spectral spatialization [12]; and the Operational Spectral Synthesis (SOS), which allows to explore the perceptual spatial phenomenon of objecthood as an expressive musical tool [26]. However, as pointed out by Kendall and Ardila, misconceptions about the technical

capacities of spatialization and about the nature of spatial perception are the main reasons why the spatial potential of electroacoustic music is not always realized [11]. These authors also discuss different techniques to play artistically at the level of perceptual grouping and auditory spatial schemata while having clear concepts and terminology, which unfortunately is not always the case.

The last point we want to introduce deals with graphical scores and spatial parameters notation. Providing performers with an easy and readable graphic representation, comparable to a listening guide or an analytical annotation system, may actively support performers activities while playing fixed-media music. Since there is no format or standard notation systems for analytical transcription of fixed-media music, many analysts use sonogram representations as a support for their transcriptions and annotations, as a time-frequency signal visualization can often provide useful insights [3] [6][7]. Software tools, such as the Acousmograph, allow to create graphical scores by adding symbols layers above the sonogram and/or waveform layers - the whole being displayed with great temporal synchronization while playing the audio file [8]. Other software tools, such as Spear [13] or Audiosculpt [4], allow to graphically edit sound in the time-frequency domain, but not to superimpose graphical layers to the sonogram view. Our work is also an attempt to mix both of these aspects: an interactive notation system which allows to describe the morphologies to be spatialized according to a finite set of spatial patterns and trajectories, and the real-time visualization of the annotated spectrum while morphologies will be automatically extracted from the audio signal and individually spatialized during the performance.

2. PROJECT BACKGROUND

Spatialization of stereo music in concert is currently a common practice. It allows to enrich the music internal space, which is technically limited by the use of only two channels, through the diffusion/projection of the music in an external space by means of a multi-channel diffusion system, thus transforming a simple playback experience into a real live music performance. However, controlling many channels is not an easy task, especially for novice performers. Hence the idea of developing a specific tool to be used within specific hardware configurations in order to support such task.

2.1 Designing a spatialization system

Any spatialization system is usually designed to be used in a specific listening environment, and the greater its specificity and technical complexity, the lower is its applicability in the most common performative contexts. Furthermore, quantitative studies on spatialized music have shown that individual differences in approaching various compositional and interpretive contexts make impossible, if not useless, to create a system that can meet all artistic needs [16]. For these reasons we adopted a stratified approach to sound spatialization, as suggested in [17]. We separated higher symbolic layers, such as the authoring and spatialization scene descriptor layers, from lowest layers, which

directly address spatialization algorithms and physical hardware, in order to ensure the efficiency of a spatialization system. Furthermore Orm  follows the idea that authoring and rendering of spatial audio might occur at completely separate times and places as suggested in [18].

Within Orm , a macro-event consists in the management of a collection of several sound entities. Each entity represents a specific sound morphology, which will be defined through its shape and time/frequency coordinates and with its specific associated spatial action. Spatial actions may be selected through predefined static or dynamic space patterns or through user-defined trajectories. Thus, a macro-event marks the starting point of the entities evolution at a given moment of the musical piece. A more detailed data structure and list of spatial configurations will be given in the section of software implementation.

Allowing to edit (authoring level) single entities by associating a sound morphology to a spatial action helps focus on the musical and interpretative functions of the spatialization, which is the main purpose of our work. Interpreters may find a lot of hints on that topic by reading the theoretical and methodological approaches about the diffusion of acousmatic music through multi-channel equipment in [10][27][28]. However, while interpreters of acousmatic music usually play live by simply acting on the potentiometers of a console and without any sort of automation, with Orm  they can introduce automations at a very low-level (spectrum), and then adjust spatial parameters to the current setup of the performance context. These features may simplify their tasks and reduce practice and rehearsal times.

2.2 Spatialization techniques overview

There are many spatialization techniques commonly used nowadays. Since our project deals with multi-channel systems, we will not take into consideration binaural reproduction, as it addresses single listeners and imply the use of headphones. In the following paragraphs we will briefly describe the most common spatialization techniques, without going into technical details, which is not the goal of this paper, and then we will present the spatialization method used in Orm .

Panning techniques are related to the spatial distribution of a sound signal in a multi-channel diffusion system through inter-channel amplitude differences and/or time-delay differences. The former is the most frequently used. It is based on our auditory perception, especially the way we perceive the combination of two sources as ‘phantom images’ according to the difference of their amplitude levels. Common 2D setups include stereo, quadriphonic, 5.1 and octophonic configurations, but more complex and non-conventional configurations may occur, as in acousmonium-like setups, which are usually organized in concentric circles of speakers arrays.

Vector-Based Amplitude Panning (VBAP) belongs to the amplitude panning techniques [19]. It is a method to calculate gain factors for pair-wise or triplet-wise (for 3D setups) amplitude panning. It is a very simple and efficient technique to position a virtual sound source for multi-speaker outputs. However, as

with normal stereo panning, it guarantees a good quality of imaging mainly for a specific listening point, the so-called sweet spot.

Distance-Based Amplitude Panning (DBAP) extends the stereo amplitude panning technique to an array of speakers which can assume different dimensions according to the size of the listening environment [14]. DBAP was designed for situations that disregard the idea of the sweet spot and where it is not possible or desirable to position loudspeakers in a surrounding way.

Ambisonics is originally a multi-microphonic recording technique which encodes, at a specific location, a directional pressure-field of a plane wave through its spherical harmonics [9]. Sound is encoded and stored in multi-channel sound files and is decoded for playback. The encoding can be carried out in an arbitrary degree of accuracy. The accuracy is given by the so-called order of Ambisonics. Higher Order Ambisonics (HOA) can use an arbitrary number of reproduction channels and potentially provides a high spatial resolution [24].

Diversely from the previously mentioned techniques, Wave Field Synthesis (WFS) aims at reproducing the exact wavefront inside the listening space through spatial sampling [2]. In the synthesis procedure, based on the Huygens' principle, any wavefront can be synthesized through the use of a large number of speakers, each one acting as a punctual sound source. This technique is no longer dependent on the listener position as it creates a real sound-field.

For our project we selected a particular implementation of the Ambisonics technique: the Ambisonics Equivalent Panning method (AEP). The latter combines the encoding and decoding phases of Ambisonics into one process. Indeed, the original Ambisonics functions are complicated, non-intuitive and computationally expensive for higher orders. AEP uses panning functions equivalent to basic and in-phase decoding which can be used for a real-time panning in an arbitrary ambisonic order [15].

Our use of spatialization is focused on analytical and interpretative purposes without the need of moving punctual sound sources in space or reproducing a sound-field. Indeed, as pointed out before, our goal is to give the audience the means to enhance the listening experience by enlightening structural aspects of the interpreted piece through the movement of sound entities extracted from its spectral contents. For these reasons, even if VBAP and AEP could be equally exploited, we selected the AEP which allows to easily adapt the spatial audio information to different reproduction scenarios and which permits a real-time spatial rendering with low computational costs. Furthermore, simple and accessible tools dedicated to the Ambisonics are yet available for the Max/MSP environment. They allow to easily define a multi-channel reproduction setup and to distribute and move sound sources in that specific listening space.

These tools (ICST Ambisonics Tools) [15], developed by the Institute for Computer Music and Sound Technology in Zurich, are the following Max/MSP externals: `ambiencode~`, `ambidecode~`, `ambipanning~`, `ambimonitor~`, `ambicontrol~`.

`Ambiencode~` and `ambidecode~` are used only when dealing with the Ambisonics B-format, while `ambipanning~` is used for the AEP without any intermediate B-format. The other two objects are used both in the Ambisonics B-format and in the AEP method to control the spatialization parameters. The `ambimonitor~` is the GUI object for the Ambisonics spatialization. It allows to arrange and visualize both speakers and sound-sources positions. The `ambicontrol~` is the object dedicated to the sound trajectories, to whom we will refer as spatial patterns (SP). As detailed in the following paragraphs, Orm  takes advantage of these last three objects, which have proved to be reliable, simple to use, well documented and user-friendly. Furthermore they allow motion automations and an overall data structure which is compliant with the SpatDIF format [18]. This last point allows us to foresee the exportation of spatial information for a specific performance and to use it with other spatialization tools. For technical details about the ICST Ambisonic tool visit: www.zhdk.ch.

3. ORM 

3.1 Development Environment

The Tool Orm  has been developed in Max/MSP, using the AEP (Ambisonics Equivalent Panning) technique for the spatialization. We decided to develop Orm  using the visual programming language Max/MSP because its extensible design is ideal for an interactive music and visual performance project. Besides, it allowed us to create a standalone tool with an intuitive graphical user interface (GUI). Furthermore, Max/MSP is modular: an application programming interface (API) allows third-party development of new routines (external objects) and a lot of spatialization tools exist as shared libraries.

As already mentioned, we decided to use the Ambisonics techniques and, in particular, the AEP.

3.2 Spatialization Methodology

The sound projection method developed in Orm  is a multi-channel and multi-band spatialization, in which the track is dynamically and automatically split in various spectral bands to be spatialized by means of the ICST Ambisonics tools [21], after a previous and detailed spectral analysis. The current version of Orm  is designed to split the sound into a maximum of five spectral bands and to project them in a multi-channel system limited to eight speakers. The algorithm of the system is scalable, and it will be easy to add more bands and to increase the total outputs in Orm 's future implementations. We would like to point out that a limited number of spectral bands fits the objective of our work since Orm  is designed for performance and music interpretation and not as a spatialization effect. It turns out by our experience that too many bands may be confusing and misleading. This method requires the simultaneous projection of the spatialize spectral-bands and the original track, the latter always evenly diffused from the whole system. This fact allows to preserve the spectral cohesion of the track, which the single spectral-bands would not be able to ensure, and to prevent

undesirable gaps in the sonic space. This method requires three phases for the realization of a complete spatial interpretation: analytical; interpretative and editing; and performative.

In the first phase, the user has to carry out a detailed spectral analysis of the piece with a dedicated software (such as Audiosculpt, Acousmograph or Spear), determining which are the structural, timbral, gestural and formal aspects that can be highlighted through the spatialization.

In the second phase, the user identifies the sound entities which, according to personal choices, may be better extrapolated, by means of filtering, and emphasized through the spatialization process. This is done by determining and editing the data for the spectral bands and their space configurations.

During the last phase, the interpreter assumes primarily a function of control of the automation and spatialization process. Nevertheless, the user can operate in real-time on certain parameters in order to adapt the system to the acoustics of the environment. He also can follow the real-time annotation and representation of the elements involved, taking advantage of the visual aid to better understand the ongoing relationships between space and spectral characteristics.

3.3 Software implementation

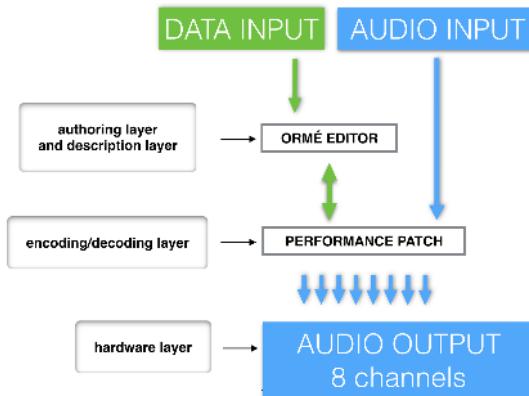


Figure 1. Orm  overall architecture design.

Orm  consists of two separate patches to be used simultaneously, as illustrated in figure 1: the first one (Orm  Editor) provides an interface to be compiled by the user during the interpretative and editing phase; the second one (Performance Patch) is dedicated to the performance and to the real-time visualization of the annotated spectrum representation. The latter reads the data produced by the former to create the automated multi-channel spatialization of the frequency bands. The audio input format is monophonic or stereophonic. The input data are organized in macro-events. With the term macro-event, we refer to different sound entities which consist of spectral bands and their spatial patterns. There is no limit to the number of macro-events that can be defined during the interpretative and editing phase. In each macro-event, the audio input is split in up to five frequency bands, and each band is in turn

independently spatialized by means of the AEP tool and superimposed to the original track. Each band is defined by an upper and a lower frequency limit, and for the moment it is frequency-wise fixed. The performance patch used during the last phase returns the audio signals for the multi-channel system, which for the moment is limited to eight speakers.

3.3.1 The Interpretation and Editor Patch

The first patch to be used, during the interpretation phase, is the Orm  editor. Its user interface is designed with a spreadsheet approach as illustrated in figure 2, in which data related to each macro-event (listed in the following table), should be inserted and saved.

ME_i	Macro-event configuration number; $i = 1 \dots +\infty;$
ST_i	Start time in milliseconds of the ME _i ;
SB_{ij}	Spectral-band number relative to the ME _i ; $j = 1 \dots 5;$
ET_{ij}	End time of the SB _{ij} spatialization process;
AT_{ij}	Attack time relative to the SB _{ij} ;
RT_{ij}	Release time relative to the SB _{ij} ;
dB_{ij}	Amplitude relative to the SB _{ij} (dB full-scale);
LLSB_{ij}	Low-limit (in Hertz) relative to the SB _{ij} ;
HLSB_{ij}	High-limit (in Hertz) relative to the SB _{ij} ;
SP_{ij}	Spatial pattern type relative to the SB _{ij} ;
XposS_{ij}	Start X-position;
YposS_{ij}	Start Y-position;
Size_{ij}	Amount of movement;
TR_{ij}	Trajectory number relative to the SB _{ij} ;
nTR_{ij}	Number of dots relative to the TR _{ij} ;
xTR_{ijw}	X-position; $w: 2 \dots nTR_{ij};$
yTR_{ijw}	Y-position; $w: 2 \dots nTR_{ij};$
t TR_{ijw}	Time between TR _{ij(w-1)} and TR _{ijw} ; $w: 2 \dots nTR_{ij};$
TRseg_{ij}	Type of segment: linear or polar;
TRLP_{ij}	Activation of loop or palindrome movement;

Table 1. List of the parameters managed by Orm . The index i is referred to a single macro-event, the index j is referred to a single spectral band, the index w is referred to a single spatialization segment.

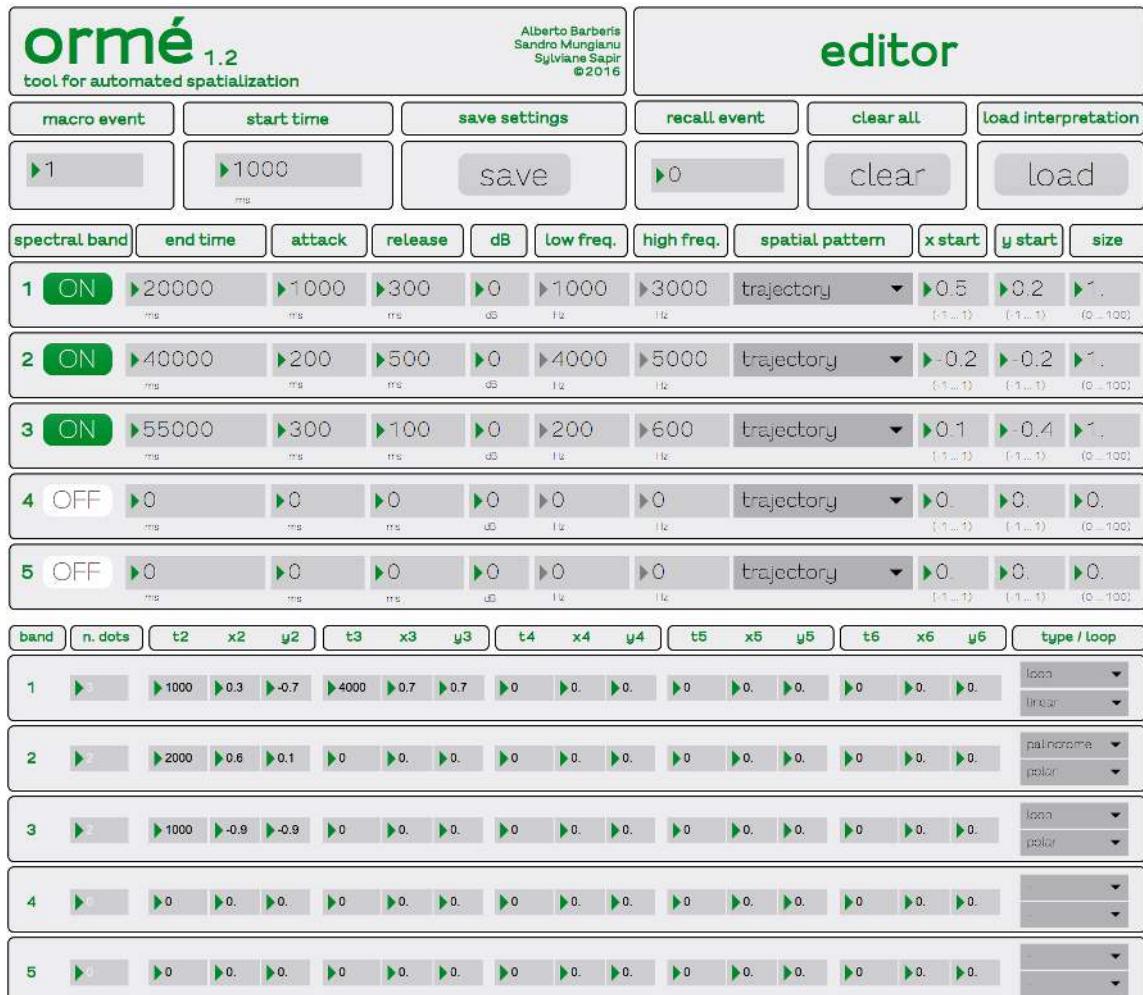


Figure 2. Orm  editor patch user interface.

As shown in figure 2, the first two data that should be inserted, related to each macro-event, are the macro-event number (ME) and the start time (ST). Subsequently, for each ME it is possible to activate up to five Spectral Bands (SB), using the On/Off button. For each SB, the user has to set up the parameters listed in Table 1: ET, AT, RT, dB, LLSB, HLSB, SP, XposS, YposS, Size. If the SB is a trajectory, the user has to insert other pertaining data: TR, nTR, (x, y, t)TR, TR, TRloop. The possible spatial patterns are those defined by the AEP tool; they can be static (fixed point) or dynamic (random, circular random, rotation, trajectory). These are the different motions that, for the moment, can be used in Orm  to spatially interpret the isolated spectral morphologies coherently with their ‘intrinsic space’.

Thus, the edited MEs set the different configurations that will be recalled during the performance phase. However, it is possible to directly recall a specific configuration from the editor during rehearsals, and eventually modify it according to the current setup or interpretation intentions. With the button Save, the user freezes the data contents of a ME

which will be saved into different Max/MSP coll objects.

3.3.2 The performance patch

After having started the spatialization process, the interpreter can act on the overall controls of the interface, which are dedicated to the audio signal levels. The role of the performance patch is first of all to playback the audio file, and secondly to recall the data referred to the current ME at the corresponding ST, in order to process the audio according to the related spatialization configuration. The graphical interface of the performance patch (figure 3) is designed to provide a representation of the spectral and spatial elements involved in this process. This occurs through a real-time spectrogram annotation, which includes: the audio waveform; its spectrogram; the SBs activated for each ME; the SPs. The waveform is directly created when the file is loaded. Instead, the spectrogram has to be generated as an image during the interpretation phase and loaded before the performance. The SBs are represented as colored rectangles directly on the spectrogram, and they appear in real-time. Each of the



Figure 3. Orm  performance patch user interface. A screenshot of the display at about 5'37" during the performance of *Jeux de Configurations* by Bernard Parmegiani.

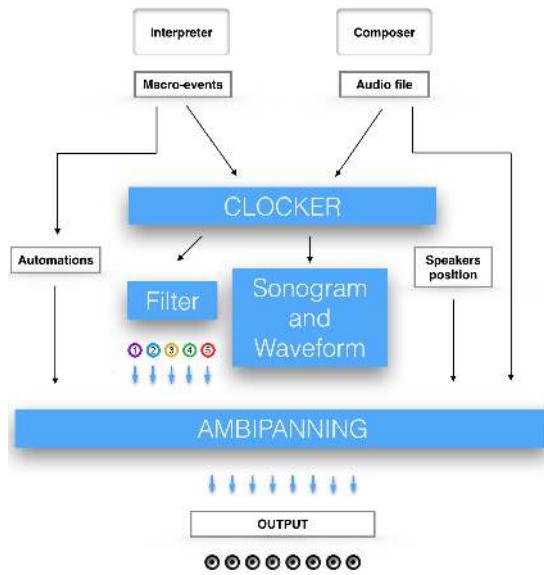


Figure 4. Block-diagram of the overall system workflow.

five bands has a specific color that is maintained also in its motion representation. The motion representation is obtained, using the AEP tool, by means of numbered and colored dots (as shown in the figure). In the lower part of the GUI there are several panels containing different controls and live indicators.

In order to use the patch, the performer should primarily set up the speaker configuration on the room

representation and arrange the position of the main L - R dots, which correspond to the original track played back without processing and which presumably will not be moved during the performance. Empirical experiments have shown us that a good position of the main L - R seems to be a frontal one, with the left and right channels rather close, and not so far from the sweet spot. Afterwards, the user should load the audio file and its spectrogram, then open the interpretation patch, which must always remain open during the performance, and finally load (Load Button) the interpretation file (that contains the macro-events data). In the player panel the playback is activated and some information are provided (time in milliseconds, ME_i). The remaining panels provide the remote control of signal levels.

Figure 4 illustrates the general organization of the performance patch, which has been developed in different blocks. The starting data consist of audio files and interpreter MEs; these data are scanned by the clocker, which synchronizes three processes: the filtering of the file audio; the visualization of the annotated sonogram; the waveform and the spatialization ICST modules, which have been previously initialized with the speaker position. In the current version of Orm , we use a 5-way crossover to extrapolate the audio bands. This filter has been realized by using ten Max/MSP third order filters (cross~). The first phase of testing in studio has brought good results, but the filter has to be tested in different concert conditions with a bigger audience.

4. CONCLUSION AND FURTHER WORK

The tool Orm  represents only our first attempt to incorporate analytical procedures in an automated multi-channel spatialization. With this tool, we provide a flexible solution for the spatialization of monophonic or stereophonic fixed-media compositions, in which the intimate sound characteristics and the structural aspects of the music could be highlighted by means of sound projection. Furthermore, we provide a sonogram annotation syntax that allows for a real-time representation of the spectral and spatial elements involved in the process. Orm  has already been used in concert situations and it has shown its potential both for performative and didactic purposes. In fact, according to our experience, we observed that Orm  could be a useful tool to support music analysis, in addition to music performance, since it is possible to directly verify analytical hypotheses by listening the spatialized tracks and comparing different kinds of interpretations. Additionally, Orm  could also be a suitable tool for analytical listening and ear-training practice.

For the time being, we considered the multi-band spatialization technique the most effective way to achieve our goals, although it requires a compromise with the absolute timbral integrity of the music. Nevertheless, an appropriate use of the multi-band filtering, awareness of psychoacoustic principles and a ‘gentle’ use of spatialization should limit the possible musical misconceptions. The planar configuration and the limitation to eight speakers are not excessive limits; in fact, they guarantee the adaptability to the most common performative contexts, without requiring the use of excessively complex hardware.

We are currently implementing a real-time representation of the spectrogram, by using the internal Max/MSP FFT and the Jitter functions. Furthermore, by using the same FFT process, at the authoring level, we would also like to graphically define by hand the spectromorphologies. We are also planning to expand the functionalities of Orm  into a 3D model and to fully adapt our trajectories annotation to the ICST Tools format.

Moreover, further implementations will concern the typecasting of the sound entities. The idea is to create spectral masks, that could either dynamically follow more complex morphologies or trace specific spectral patterns. This topic will also affect the amount of simultaneous SBs, now limited to five, and the filtering technique method. In addition to these implementations we intend to evaluate the way to adapt the system to the SpatDIF format.

5. REFERENCES

- [1] A. Agostini, D. Ghisi, “A Max Library for Musical Notation and Computer-Aided Composition”, *Computer Music Journal*, Volume 39, No. 2, pp. 11–27, 2015.
- [2] A.J. Berkhout, J. Augustinus, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis”, *The Journal of the Acoustical Society of America*, 93.5, pp. 2764-2778, 1993.
- [3] A. Blackburn, J. Penny, “Timbral notation from spectrograms: notating the un-notable?”, *Proceedings of the First International Conference on Technologies for Music Notation and Representation*, pp. 218–225, 2015.
- [4] N. Bogaards, A. R bel, X. Rodet, “Sound analysis and processing with Audiosculpt 2”, *International Computer Music Conference (ICMC)*, 2004.
- [5] A. Bregman, *Auditory Scene Analysis*, MIT Press, Cambridge, 1990.
- [6] R. Cogan, E. Pozzi, *New images of musical sound*, Harvard University Press, Cambridge, Massachusetts and London, 1984.
- [7] P. Croupie, “Graphical representation: an analytical and publication tool for electroacoustic music”, *Organised Sound 9 (1)*, Cambridge University Press, pp. 109–113, 2004.
- [8] E. Favreau, Y. Geslin, A. Lef vre, “L’acousmographe 3”, *Journ es d’Informatique Musicale (JIM 2010)*, 2010.
- [9] R. K. Furness, “Ambisonics - An Overview”, *AES 8th International Conference*, Washington, D.C., 1990.
- [10] E. Justel, “Vers une syntaxe de l’espace, in Vande Gorne”, *L’espace du son III*, Lien, 2011.
- [11] G. S. Kendall, M. Ardila, “The artistic play of spatial organization: Spatial attributes, scene analysis and auditory spatial schemata”, *Computer Music Modeling and Retrieval. Sense of Sounds*. Springer Berlin Heidelberg, pp. 125-138, 2007.
- [12] D. Kim-Boyle, “Spectral spatialization-an overview”, *Proceedings of the International Computer Music Conference, Belfast, Ireland*, 2008.
- [13] M. Klingbeil, “Software for spectral analysis, editing, and synthesis”, *Proceedings of the International Computer Music Conference*, 2005.
- [14] T. Lossius, P. Baltazar, T. d. I. Hogue, “DBAP – Distance based amplitude panning”, *International Computer Music Conference (ICMC)*, Montreal, pp. 39-46, 2009.
- [15] M. Neukom, J. C. Schacher, “Ambisonics equivalent panning”, *Proceedings of the 2008 International Computer Music Conference*, Belfast, UK, 2008.
- [16] N. Peters, G. Marentakis, S. McAdams, “Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis”, *Computer Music Journal*, 35 (1), MIT pp. 10-27, 2011.
- [17] N. Peters, T. Lossius, J. Schacher, P. Baltazar, C. Bascou, T. Place, “A stratified approach for

- sound spatialization”, *Proceedings of the 6th Sound and Music Computing Conference*, Porto, PT, 2009.
- [18] N. Peters, L. Trond, and J. C. Schacher. “The spatial sound description interchange format: Principles, specification, and examples”, *Computer Music Journal* 37.1, pp. 11-22, 2013.
- [19] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning”, *Journal of the Audio Engineering Society* 45.6, pp.456-466, 1997.
- [20] S. Roy, *L'analyse des musiques électroacoustiques, modèles et propositions*, Editions L'Harmattan, 2004.
- [21] J. C. Schacher, “Seven years of ICST Ambisonics tools for maxmsp—a brief report”, *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, pp. 2-5, 2010.
- [22] M. Schumacher, J. Bresson, “Spatial Sound Synthesis in Computer-Aided Composition”, *Organised Sound* 15 (3), Cambridge University Press, pp. 271-289, 2010.
- [23] D. Smalley, “Spectromorphology and structuring processes”, in *The Language of Electroacoustic Music*, edited by S. Emmerson, MacMillan Press, New York, pp. 61-93, 1986.
- [24] S. Spors, J. Ahrens, “A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling”, *Audio Engineering Society Convention 125*. Audio Engineering Society, San Francisco, 2008.
- [25] D. Topper, M. Burtner, S. Serafin, “Spatio-operational spectral (sos) synthesis”, *Proceedings of the International Conference on Digital Audio Effects*, Hamburg, Germany, 2002.
- [26] R. H. Torchia, C. Lippe, “Techniques for multi-channel real-time spatial distribution using frequency-domain processing”, *Proceedings of the 2004 conference on New interfaces for musical expression*, National University of Singapore, 2004.
- [27] A. Vande Gorne, “L’interprétation spatiale. Essai de formalisation méthodologique”, *Revue DEMéter*, Université de Lille-3, 2002.
- [28] H. Vaggione, “L’espace composable. Sur quelques catégories opératoires dans la musique électroacoustique”, *Chouvel J.-M., Solomos M., L’Espace: Musique/Philosophie*, L’Harmattan, Coll. Musique et Philosophie, Paris, 1998.

Understanding cooperative sound design through linkographic analysis

Stefano Delle Monache

Iuav University of Venice

sdellemone@iuav.it

Davide Rocchesso

Iuav University of Venice

roc@iuav.it

ABSTRACT

Protocol and linkographic analysis are applied to a task of cooperative vocal sketching, proposed in the scope of educational research activites. The understanding of the cognitive behaviors involved in sound creation is aimed at setting the ground for the development of rigorous, designerly evaluation practices tailored to sound design. We show how relevant qualitative and quantitative information about the creative process can be used to inform the assessment and possibly improvement of vocal sketching methods.

1. INTRODUCTION

In the last decade, Sonic Interaction Design (SID) became a widespread label in the academic and artistic domains, to give voice to the plethora of emerging sound design practices and approaches to the culture of the audible, rooted in the synthesis of art, science, and technology. Major efforts have been undertaken by the research community to devise practices, frameworks, methods and tools to channel creative processes within the streams of a proper design discipline, founded at the intersection of interaction design and sound and music computing [1].

However, it must be admitted that despite these many efforts, SID propositions¹ are hardly finding their place in the real-world practice, being mostly confined in the labs and the fictive realms of educational workshops. For example, cooperative processes based on incremental iterations of crafting and evaluation are far from entering the *de facto* workflow of sound practitioners. Instead, projects are mainly carried out individually, the sound designer is left alone, immersed in a multitude of software tools to transform early, highly conceptual ideas in refined realizations.

Design methods, proposed by the academy, fail to breach the professional practice because they are developed in playgrounds which are essentially research-led rather than practice-led. It is not only a concern of magnitude of resources to assess the effectiveness of a design method in a real-world context, it is rather a matter of user research and test-

ing. In fact, despite the general advocacy of a user-centered design, many scholars often fail to take in account designers as end-users of the methods under development [2].

In the last years, we have been collecting several workshop experiences around the themes of basic design and vocal sketching [3, 4], and several processes and activities have been conceived and structured to foster SID pedagogy and training [5, 6, 7]. Despite the optimization of careful pacing and clear learning objectives, we must say that workshop outcomes are often below expectations, in terms of both quality and quantity. The fact is that the proposed methods are mostly assessed qualitatively through observations collected in the lab environment or educational contexts, with the aim of setting points for analysis and future work. We feel that a step forward in the consolidation of SID would benefit from the application of rigorous, designerly evaluation practices. For this purpose, we attempt to apply established approaches to study design cognition in order to assess the effectiveness and appropriateness of SID methods and practices.

In this paper, we address the assessment of the creative processes involved in a task of cooperative vocal sketching. A group of six master students in sound and music computing was engaged in sketching the engine sound for a concept car, and producing a sonic overlay out of vocalizations: A video-prototype of the vehicle was produced, by adding the sonic elements as layers of a silenced video of the car motion² [4]. The design session was constrained to a maximum duration of sixty minutes.

Protocol and linkographic analyses [8, 9] are the preferred tools to generate a qualitative and quantitative understanding of the design processes, and implicitly of the effectiveness of the use of vocalizations as primary means to support sound design-thinking, when drafting the acoustic appearance and behavior of an artifact.

The present study is part of a broader research aimed at developing a fine-grained understanding of the unfolding of design-thinking in the early conceptual stage of the sound design process. A systematized understanding represents an essential step in the development of methods relevant for the practice. Indeed, the engine sound design task was proposed to six teams, under three different sound production conditions, one condition per pair of groups: 1) vocalizations only; 2) Foley³ only; 3) sound synthesis and processing only. The objective is to produce comparative analyses of the creative processes under the same condi-

¹ A comprehensive chart on SID can be found at http://mat.ucsb.edu/240/F/static/notes/Sonic_Interaction_Design.html.

Copyright: ©2016 Stefano Delle Monache et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

² Similarly to the video example available at <https://vimeo.com/128886746>.

³ [https://en.wikipedia.org/wiki/Foley_\(filmmaking\)](https://en.wikipedia.org/wiki/Foley_(filmmaking)).

tion, and across conditions, in order to assess the effectiveness and appropriateness of approaches and tools.

In Section 2, we briefly introduce protocol analysis as means to study design cognition, then we report about the preliminary study on the early five minutes of the vocal sketching session, aimed at calibrating the coding procedure and the linkographic analysis. In Section 3, we analyze the whole session qualitatively and quantitatively. Then we discuss the findings in the light of possible improvements to the design task.

2. CODING THE SOUND DESIGN PROCESS

In our previous study [10], we tested the application of linkographic analysis methods to cooperative vocal sketching. In particular, we discussed the calibration of the analysis to the early five minutes of the audiovisual recording, and showed how sentences, vocalizations, and gestures by participants can be extracted, coded, and related to each other to extract qualitative and quantitative summaries of design processes. A summary is reported in Section 2.3.

Protocol analysis is an established method to study design cognition. The design process is segmented in *design moves*, that is minimal, sequential semantic units representing increments in thought. Once the protocol is parsed in design moves, a coding scheme is applied in order to categorize the segments. In our study, we developed a coding of elementary design moves by exploiting the Function-Behavior-Structure (FBS) ontology of design, as proposed by Gero [11, 12]. According to this categorization, each design object can be conceptualized in terms of three main ontological categories, that is the Function, the Behavior, and the Structure. This FBS categorization has been developed to model the process of designing as a set of distinctive activities⁴.

2.1 The Function-Behavior-Structure scheme

The FBS generic categorization framework is showed in Figure 1, and represents the design process as valid semantic transitions in a finite-state automaton of design moves.

Design moves are coded according to:

- the **Function (F)**, that is the purpose of the artifact, set by the designer or redefined from **Reformulation III**, as a consequence of the emerging structure (S);
- the **Expected Behavior (Be)**, as **Formulation** of the function or as **Reformulation II** of the emerging structure (S);
- the **Structural Behavior (Bs)**, as derived from the **Analysis** of artifact configuration, that is the structure (S). **Evaluation** occurs when the expected and structural behaviors are compared against each other;
- the **Structure (S)**, that is the configuration of the elements and their formal relationships, as emerging from the **Synthesis** of the expected behaviors, or

⁴ See https://en.wikipedia.org/wiki/Function-Behaviour-Structure_ontology, for a comprehensive overview.

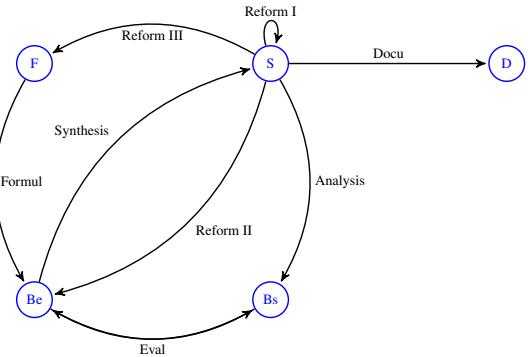


Figure 1. The FBS categorization automaton. The codes represent the states. Arcs are labeled according to the design processes.

from the **Reformulation I** of the revised Structure (S);

- the **Documentation (D)**, represents the set of design descriptions as outcomes of the FBS process.

Hence, we segmented the audiovisual recording of the vocal sketching session, and extracted the relevant verbalizations, vocalizations and gestures. These segments were put in relation and linked with each other using the finite-state automaton in Figure 1. Being this an interpretive activity based on common sense, we analyzed the transcriptions and the coding, individually and together, in order to converge to shared coding procedure and ensure intersubjective verification and consensus around the occurrence of links [10]. The emerging network of links between design moves represents a rich source of information that can be systematized through linkographic analysis [9].

2.2 A linkograph of design moves

The Excel file with the transcriptions and the code labels was imported in LINKOgrapher⁵ [13] software to generate the linkograph, and extract general statistics as well as parameters of dynamic models proposed in the literature for design processes [14, 15]. At this stage, absolute temporal information is neglected, as linkographic analysis is only based on a discrete temporal logic.

Figure 2 shows a snapshot example extracted by the linkograph of the whole session. Links emerge as nodes of the reasoning, that can be established as forward or backward, based on their contents. Forelinks denote acts of synthesis. Backlinks reveal evaluation steps in design thinking.

In particular, Figure 2 depicts the reasoning about the aerodynamic component of the braking sound: Participant 5 (P5) formulates the behavior of the braking, accompanied by a sucked woosh vocalization and an iconic gesture illustrating the sound dynamics. Both vocalization and gesture at moves 10 and 11 represent a synthesis of the expected behavior set at move 9. Participant 2 (P2) proposes

⁵ <https://sites.google.com/a/linkographer.com/linkographer/>

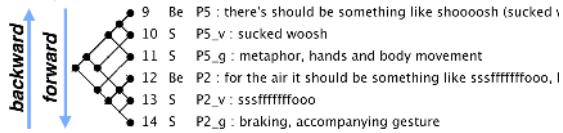


Figure 2. Snapshot of the linkograph for moves 9–14. The number of the move, the participant_type label, that is verbalization, vocalization (−v), gesture (−g), are reported for each transcription.

an alternative formulation of the braking function (set at move 5 by P4 and not displayed in Figure 2), yet the proposed structural vocalization and gesture are both a synthesis of the behavior at move 12, and a reformulation of the vocal and gestural structures at moves 10 and 11.

2.3 Summary of the early five minutes

The early five minutes, corresponding to the segments 1 – 81, are highlighted in the blue-framed box in Figure 3, which represents in turn the linkograph of the whole design session. Lapse 1 – 81 was isolated by the authors to calibrate the application of linkographic analysis methods, because it encompasses a complete cycle of design process, from the formulation of functions to the production of early documentations.

The application of linkographic analysis to the early five minutes of the design session has been extensively discussed elsewhere [10]. Here, we report only a few points to stress the richness of the information that can be extracted.

In general, this early cycle describes a process progressing smoothly, yet not well-integrated, due to the presence of several nodes with distant links that quickly resolve to documentation steps. The first three purple ellipses highlight three main clusters of links reflecting an intense and in-depth examination of the design problems at hand. Cognitive activity in conception and exploration of ideas is high at the beginning, to fade almost linearly during clarification and resolution, respectively represented by the almost linear progression in the fourth ellipse and the documentation transition.

Table 1 reports the occurrence of different FBS categories for segments 1 – 81, as extracted by LINKOgrapher. Structure codes are the most influential elements in

Code	Segments	
F	1	1.2%
Be	14	17.3%
Bs	13	16.0%
S	37	45.7%
D	16	19.8%

Table 1. FBS codes for segments 1 – 81.

this session, thus suggesting a tendency to support and confirm the behavioral elements set, as they are. This can be read in terms of functional fixation on the initial ideas, thus impairing the generation of alternative concepts from the

iterative inspection of the representations being produced (i.e., the structures).

Figure 4 represents the temporal distribution of codes, computed with a sliding window of 10 segments: Three roughly structured cycles of evaluation can be observed, right after a short and intense step of formulation and synthesis in the early 20 moves.

The analysis of the linkograph unveils an early collaborative process which is mainly focused on searching a shared and steady arrangement and progression, rather than exploring alternative ideas, to the detriment of creative and out-of-the box solutions.

3. THE WHOLE SESSION

Summary: After a brief discussion on the sonic features of the concept car, the group formulates a basic set of car sound behaviors, and namely the ignition, engine idling, engine revs-up, braking. One idea per sound is quickly sketched, based on individual proposals, and a first audio-visual documentation is produced by recording a live collaborative performance on top of the car video. Tasks are divided between participants, and the individual sounds are developed, through iterative sketching and discussion. Finally, the documentations of the single sounds are edited and added to the car video.

3.1 Qualitative analysis of the linkograph

From the combined inspection of the protocol and the linkograph (Figure 3) it is possible to roughly partition the overall session in three macro regions: 1) the early design cycle from segment 1 to 81; 2) a second region spanning the exploration and refinement of the sounds set; 3) a third region encompassing editing and montage of sounds on the video.

The first region shows a relatively higher density of links and inter-links between groups of moves, compared to the other parts, and indicates a rather solid activity leading to the early production of several design descriptions.

Groups of moves that are almost exclusively linked among themselves represent lines of thought focused on a specific design element, while interlinks between these groups can reveal the exploration of divergent thinking, and different viewpoints on the issue at hand [9]. This is not the case, as the connections between the four ellipses rather describe the generation and inspection of concepts which alternate or go in parallel. Figure 5 provides a closup on segments 1–31, wherein the presence of several moves with no fore-links can be noticed, denoting propositions which are discarded by the group and no further scrutinized. The only move with no backlinks is at segment 6 and represents a novel proposition.

The substantial absence of moves with no backlinks suggests that the activity is focused on one principal idea only, derived from the function set at move 5 (red dash). In other words, the set of sounds documented at the end of this first cycle are originated from the one main idea.

Nonetheless, the first region is the most active part of the whole session, the working plan is defined and sound

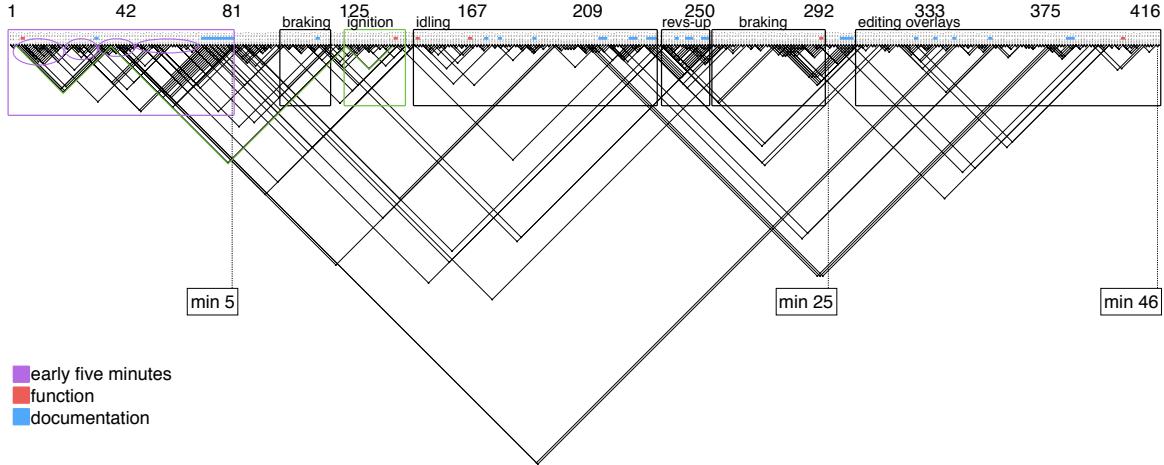


Figure 3. Linkograph for the whole session. 416 segments were retained for coding.

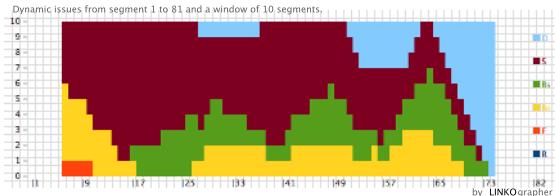


Figure 4. Temporal distribution of codes for segments 1 – 81.



Figure 5. Closeup of the early design steps. Moves with no forelinks denote lines of thought which are no further carried out. Moves with no backlinks represent novel ideas.

behaviors are structured, evaluated, and documented through sketching.

Indeed, the second part (spanning segments 100 – 295) essentially represents a refinement step, in which the individual sound behaviors are deepened sequentially, as highlighted by the most distant connections between the first region and the sub-regions of the second part. We can speculate that the group is mostly focused on improving the quality of their sketches, that is practicing and exploring the vocal articulation of the fixated sound concepts.

The rather long extension of the number of moves for the idling sound refinement may be read in terms of a high complexity of the element at hand. However, from the

closeup in Figure 6 it is possible to isolate several *sawtooths*, sequences of at least four linked moves, ending with no forelinks. This means that the idling sound process is mainly proceeding linearly with a poor engagement in the synthesis process. Propositions are built one upon another, without a widen investigation. It can be noticed how only the last two documentations out of the six produced are actually forwarded in the process.

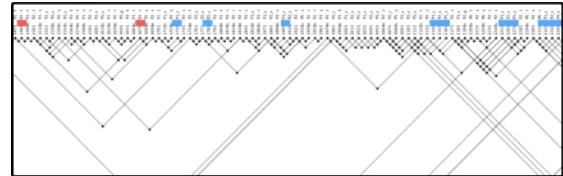


Figure 6. Closeup of the idling-sound design section. Sawtooths are linear sequences of moves and are an indicator of low engagement in synthesis processes.

Of interest is the case of the ignition sound design-thinking process, whose unfolding is highlighted in green. The trajectory clearly shows how this proposed sound function received basically no examination at all, and was left aside until move 122. As shown in Figure 7, the design process progresses through dead-end linear sequences of moves to a point in which the function (red dash) is reformulated. However this reformulation does not lead to a new FBS cycle, and the ignition sound process is abandoned at move 145. The specific sound for the car ignition is discarded, yet the function is retained and assimilated in the idling sound process, as it can be inferred in Figure 3, by the move right after the red dash linked with the next to last documentation (blue dash) of the idling region.

The third macro region in Figure 3 encompasses the editing and montage operations towards the prototyping step. At this stage, the sound elements produced so far undergo a process of evaluation in context. In some cases, the group preferred to discard the most recent documenta-

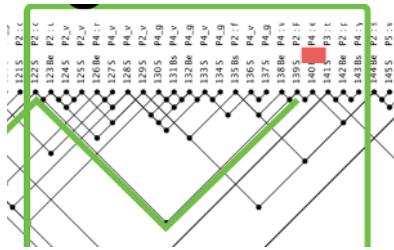


Figure 7. Closeup of the ignition-sound design section.

tions and recovered the early sound structures. These are well-exemplified by the connections towards this region, with the highest link span. In other words, the editing step reflects the evaluation and reconsideration of the previous design cycles.

Although absolute time intervals are not taken into account in linkographic analysis, the relative durations of the processes in the three main regions are worth noticing. Five minutes are devoted to conception and exploration stages (i.e., first region), twenty minutes are allocated to clarification and refinement activities, and other twenty minutes to the resolution stage. Only a small time was actually used by the group to understand and address the design task.

This can be interpreted as 1) difficulty or scarce inclination to exploit divergent thinking; 2) limited mastery in sound sketching, and therefore pressure to accomplish the task; 3) inadequacy of the software tool (i.e., the basic video editor) in affording sound sketching without sparing attention on the production process (indeed, sound editing and audiovisual syncing were mostly postponed to the end of session).

Finally, Table 2 shows the occurrence of verbalizations, vocalizations, and gestures by each participant during the whole session.

	Verbalizations	Vocalizations	Gestures
P1	40	44	7
P2	101	41	8
P3	40	2	0
P4	25	33	12
P5	9	22	5
P6	15	9	3
Total	230 (55,3%)	151 (36,3%)	35 (8,4%)

Table 2. Occurrence of verbalizations, vocalizations and gestures per participant during the sketching session.

In general verbal-thinking is prevalent, however some participants are more prone than others in using vocalizations as means to externalize sound ideas. Gestures have a limited use, mostly in a dual function, 1) as iconic means to signify aspects of sounds that cannot be produced or controlled with the voice, and 2) to physically affect the vocal sound, e.g. a damper effect obtained with the hand on the lips.

Sketching through vocalizations proved to be an effec-

tive means in structuring a complex concept in a few minutes. It can be argued that the bottleneck to prolific concept production and creativity is not in the aesthetic limitation of the human voice, but instead in sketching dexterity as opposed to the proficiency of "vocal rendering".

3.2 Quantitative description of the session

The qualitative analysis can be supported by statistics generated in LINKOgrapher. Some general statistics are reported, to give a clue about segments and links:

- Number of segments: 416 were retained for coding;
 - Number of links: 674. The link index is of 1.62 per segment. The link index is an indicator of productivity in terms of linking activity;
 - Average segment length: 7.35s, computed on non-empty segments;
 - The occurrence of different FBS categories is reported in Table 3.

Code	Segment
F	7 1.7%
Be	96 23.1%
Bs	60 14.4%
S	202 48.6%
D	51 12.3%

Table 3. FBS codes in the whole session.

The occurrence of FBS codes for the whole session is in continuity with the statistics extracted for the analysis of the early five minutes, reported in Table 1. However the slightly different percentages tell us that the overall session is essentially less effective and integrated if compared with the early five minutes.

The increase of the expected behaviors (Be) and structures (S) to the detriment of the structural behaviors (Bs) essentially signifies an increase of synthesis transitions and reformulations involving the structure or the expected behaviors. However these cycles are not the result of an iterative process of evaluation, they rather unfold through addition in a linear progression, as visually inspected in the linkograph. In fact, being evaluation associated to transitions between Be and Bs, the decrease of Bs codes results in less opportunities of analysis and evaluation.

3.2.1 Semantic processes

Relevant statistics are also found considering the semantics of the transitions between the FBS states. In particular, $A \rightarrow B$ is a valid semantic transition if A is a segment linked back by B in the linkograph [16]. The design processes (arcs of the finite automaton) are derived from such transitions, and their occurrences are reported in table 4.

Analysis and evaluation are effectively marginal compared to the combination of synthesis and reformulation I (i.e., reconsideration of the structural elements). Taken together, the design processes affecting the structure take

Process	Occurrence	Percentage
Formulation	11	1.9%
Synthesis	106	17.9%
Analysis	68	11.5%
Evaluation	67	11.3%
Documentation	65	11.0%
Reformulation I	192	32.4%
Reformulation II	75	12.6%
Reformulation III	9	1.5%

Table 4. Occurrence of design processes.

the 50% of all valid semantic transitions. The direct effect of this imbalance is a scarce productivity in terms of documentations. This analysis suggests that, despite the effort and the engagement, the team suffered from a lack of knowledge and attitude to evaluation and critical thought. Mastery of vocal production is certainly another important issue, as stressed by the large percentage of reformulations I.

3.2.2 First-order Markov model

LINKOgrapher also returns the transition probabilities of codes across links. It is a prediction of the probability of design issues coming after each other. Empirical probabilities of code-to-code transitions are computed, based on the association of the FBS states and processes with the segments representing the design moves, and constitute a first-order Markov model [16]. The transition probabilities between codes is reported in Table 5.

	F	Be	Bs	S	D
F	0.00	0.58	0.00	0.42	0.00
Be	0.01	0.02	0.17	0.80	0.00
Bs	0.00	0.90	0.06	0.04	0.00
S	0.02	0.18	0.17	0.47	0.16

Table 5. Transition probabilities between FBS codes.

In general, the transition probabilities towards the structure code (S) are the highest, except for $Bs \rightarrow S$. When Bs issues are occurring, the likelihood to be followed by Be and therefore lead to an evaluation process is very high. However, the probability to have a Bs right after a S code is very low, being only the 17%. Remember that $S \rightarrow Bs$ is the valid semantic transition, as illustrated in the scheme in Figure 1, and that evaluation processes can be bidirectional $Be \leftrightarrow Bs$.

The combined reading of $Bs \rightarrow Be$ and $Be \rightarrow Bs$ probabilities stresses that a design progression, as the one found by the group, would lead to a low rate of iterations. In other words, the design issues at hand are not really deepened, and rather resolved as they are. In practice, it is likely to have a design process stuck on one issue, or hopping from one issue to an other. In both cases, the process inefficiency results in a scarce probability of producing documentations.

3.2.3 Entropies

The strength of an idea can be measured quantitatively by using entropy [15], and this is a measure that can contribute to give a clear picture of the productivity of the sketching session.

To use the information-theoretic concept of entropy, every link between two segments is considered as an event, which does or does not occur. In the measurement of dynamic entropy only the structure of the linkograph is considered, ignoring the codes. For the reported LINKOgrapher computations over the whole session, we used a sliding window of forty-two segment. This means that, given a segment, the software checks for the presence or absence of a link with other segments not farther than forty-two positions away.

Figure 8 shows the dynamic forelink entropy of the whole session. Forelink entropy gives a measure of the opportunity for new creations or initiations. A low entropy means weak or fixated ideas, respectively due to the occurrence of a too few or too many forelinks. The drop around move 42 corresponds to the third ellipse in Figure 3, and the clarification step in Figure 4. It means that the process is not focused on creation, but rather on evaluation. We can infer that this drop actually represents the most effective evaluation step in the whole session.

The highest peaks around moves 230 and 290 correspond respectively to the last parts of the idling sound region and the second braking iteration in Figure 3.

The segment spanning approximately moves 125 – 150 represents the ignition sound process. Here, the only peak over the mean value is the link of the segment in which the group decides to include the ignition function in the idling sound.

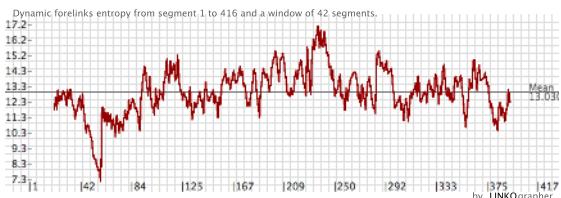


Figure 8. Forelink entropy in time.

The novelty of ideas can be measured also through backlink entropy. If an idea is very novel, it will not have backlinks, and therefore it will have low entropy. The backlink entropy of the session is shown in Figure 9. It is possible to isolate three main iterations starting at moves 1, 84, and 300, and roughly corresponding to the macro-regions, highlighted in the qualitative analysis of the linkograph. Segments 125 – 150 reflect the ignition sound region, where the apparent novelty is related to the fact that the embryonic idea, set at the very beginning of the design process, was basically left in the background. The low backlink and average forelink entropies for the set of documentations occurring at segments 300 – 310 suggest that the novelty mainly arises from an alternative arrangement of the structural elements at hand, as a consequence

of one of the few $Bs \leftrightarrow Be$ transitions. That is, the group actually managed to produce design a description when enabled by an iterative evaluation.

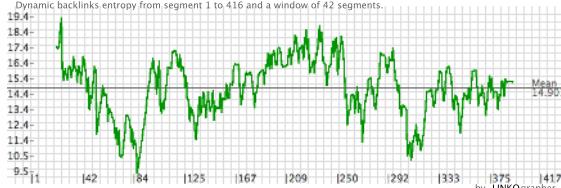


Figure 9. Backlinks entropy in time.

Finally, horizonlink entropy provides a measure of the opportunities for ideas generation based on the horizontal link span between inter-connected moves. This measure takes in account the cohesiveness and the incubation of segments. Low entropy indicates a strong cohesiveness. The graph (Figure 10) provides an overview of the process unfolding based on the distance between inter-connected moves.

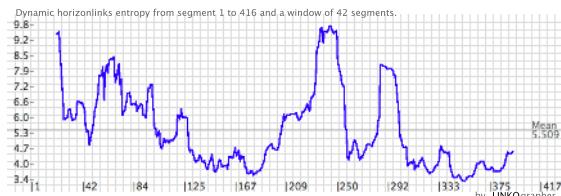


Figure 10. Horizonlinks entropy in time.

From the combined reading of the graphs of entropies, some interesting observations can be drawn regarding the revs-up sound region. As shown in Figure 11, this region is characterized by the highest incubation of segments, and high values of backlink and forelink entropies. Given the high backlinks entropy, the novelty of the ideas does not depend from the formulation of new and original propositions. Rather, the revs-up sound design process is effectively carried out and documented, by transforming and rearranging a few structural sound elements produced farther away.

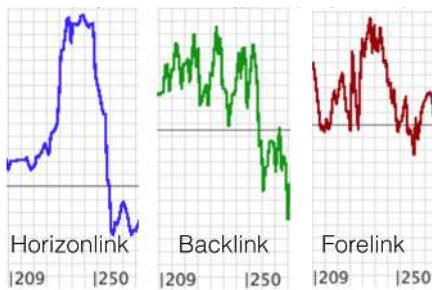


Figure 11. Entropies for the Revs-up sound region.

4. DISCUSSION AND CONCLUSIONS

The information collected from the linkographic analysis stresses several issues deserving a further in-depth analysis. A consideration concerns the critical review of the proposed task, from the viewpoint of its rationale. The proposed assignment is framed in a set of incremental activities aimed at acquiring and developing cognitive abilities and compositional skills in sketching sonic interactions and designs.

The first issue is understanding whether the conception and pacing of the assignment meet their underlying rationale, and whether alternative measures can be undertaken to effectively achieve the objective. For this purpose, protocol analysis and linkographic representations are exploited to develop a fine-grained understanding of the cognitive behaviors and the possible bottlenecks emerging in sound creation processes. Time-constraints, and economy of means are devices that are largely used in design exercises to foster creativity.

The task at hand was conceived around a few requirements: the production of poliphonic sketches in order to promote cooperation, no speech sounds nor onomatopoeia in order to foster the creation and design of non-verbal auditory displays, no Foley nor sound processing in order to encourage the practice of vocal articulation and production. Loose time-constraints are counter-balanced by the requirement of producing at least two alternative designs. In the early formulation of this task, it was required to produce at least five propositions.

The linkographic analysis of this specific team work showed that the group effectively created a self-contained sketch in the early five minutes of the design session. Therefore, a measure to foster effective productivity is to reduce the overall duration of the task, and allocate constrained slots of five minutes to actually produce one sketch. This pacing is similar to the use of speed-dating for design ideas generation and exploration of divergent thinking opportunities [17].

The exploitation of divergent thinking certainly represents a fundamental issue, and preparatory exercises around creative sound associations should be conceived in order to develop design-thinking and imagery. In order to do that, the cognitive activity emerged from the analysis of the revs-up sound process stresses the relevance of learning and mastering mental transformations. That is the ability to shift configurations and relations in the time-frequency domain, as opposed to shifting spatial relations in visual sketching.

As a consequence of focusing on the improvement of dexterity in vocal sketching, it is necessary to develop a practice of vocal articulation tailored to design, as emphasized by the frequent occurrence of structure codes and reformulation I processes across the session.

On the other side, we must not forget the assumption stated in the introduction, that is the effective grounding of methods development in user-centered design. From this viewpoint, we recently run a 48 Hours Sound Design workshop at Chateau La Coste art park and vineyard, in south France. Each of five professional sound designers

was invited to work on one of the site-specific art pieces located in the park, and design an accompanying sound signature for the chosen art installation, in 48 hours. Vocal sketching methods and tools, developed in the scope of the EU project SkAT-VG⁶ (Sketching Audio Technologies using Vocalizations and Gestures) were the exclusive means available for sound ideas generation and sketching. The documentary of the workshop is available at: <https://vimeo.com/169521601>. We plan to apply protocol and linkographic analysis to sound design processes involving both expert and novice designers.

5. ACKNOWLEDGMENTS

The work described in this paper is part of the project SkAT-VG, which received the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission under FET-Open grant number: 618067. We are thankful to Cumhur Erkut and Stefania Serafin for organizing and hosting the workshop.

6. REFERENCES

- [1] K. Franinović and S. Serafin, *Sonic interaction design*. Mit Press, 2013.
- [2] G. Dickson and E. Stolterman, “Why design method development is not always carried out as user-centered design,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 4056–4060, ACM, 2016.
- [3] S. Delle Monache and D. Rocchesso, “Bauhaus legacy in research through design: The case of basic sonic interaction design,” *International Journal of Design*, vol. 8, no. 3, pp. 139–154, 2014.
- [4] S. Delle Monache, D. Rocchesso, S. Baldan, and D. Mauro, “Growing the practice of vocal sketching,” in *ICAD 2015 Proc. of the 21th International Conference on Auditory Display* (K. Vogt, A. Andreopoulou, and V. Goudarzi, eds.), pp. 59–65, 2015.
- [5] D. Hug and N. Misdariis, “Towards a conceptual framework to integrate designerly and scientific sound design methods,” in *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, AM ’11, (New York, NY, USA), pp. 23–30, ACM, 2011.
- [6] D. Hug and M. Kemper, “From Foley to function: A pedagogical approach to sound design for novel interactions,” *Journal of Sonic Studies*, vol. 6, no. 1, 2014.
- [7] C. Erkut, S. Serafin, M. Hoby, and J. Särde, “Product sound design: Form, function, and experience,” in *Proceedings of the Audio Mostly 2015 on Interaction With Sound*, AM ’15, (New York, NY, USA), pp. 10:1–10:6, ACM, 2015.
- [8] J. S. Gero and T. Mc Neill, “An approach to the analysis of design protocols,” *Design studies*, vol. 19, no. 1, pp. 21–61, 1998.
- [9] G. Goldschmidt, *Linkography: unfolding the design process*. MIT Press, 2014.
- [10] S. Delle Monache and D. Rocchesso, “Cooperative sound design: A protocol analysis,” in *Proc. of Audio Mostly 2016, a conference on interaction with sound*, (Norrköping, Sweden), 2016.
- [11] J. W. Kan and J. S. Gero, “Using the FBS ontology to capture semantic design information in design protocol studies,” in *About: Designing. Analysing Design Meetings*, pp. 213–229, CRC Press, 2009.
- [12] J. S. Gero and U. Kannengiesser, “An anthology of theories and models of design: Philosophy, approaches and empirical explorations,” in *An Anthology of Theories and Models of Design* (A. Chakrabarti and L. T. M. Blessing, eds.), ch. The Function-Behaviour-Structure Ontology of Design, pp. 263–283, London: Springer London, 2014.
- [13] M. Pourmohamadi and J. S. Gero, “LINKOgrapher: An analysis tool to study design protocols based on FBS coding scheme,” in *ICED 2011 Proc. of the 18th International Conference on Engineering Design, Impacting Society through Engineering Design* (S. Culley, B. Hicks, T. McAloone, T. Howard, and J. Clarkson, eds.), vol. 2: Design Theory and Research Methodology, pp. 294–303, 2011.
- [14] J. W. Kan and J. S. Gero, “Acquiring information from linkography in protocol studies of designing,” *Design Studies*, vol. 29, no. 4, pp. 315–337, 2008.
- [15] J. W. Kan, Z. Bilda, and J. S. Gero, “Comparing entropy measures of idea links in design protocols: Linkography entropy measurement and analysis of differently conditioned design sessions,” *AI EDAM: Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*, vol. 21, pp. 367–377, 11 2007.
- [16] J. S. Gero, J. W. Kan, and M. Pourmohamadi, “Analysing design protocols: Development of methods and tools,” in *Proc. of the 3rd International Conference on Research into Design Engineering* (A. Chakrabarti, ed.), pp. 3–10, 2011.
- [17] S. Davidoff, M. K. Lee, A. K. Dey, and J. Zimmerman, “Rapidly exploring application design through speed,” in *Proc. 9th International Conference on Ubicomp* (J. Krumm, G. D. Abowd, A. Seneviratne, and T. Strang, eds.), pp. 429–446, Springer, 2007.

⁶ <http://www.skatvg.eu/>.

SONIC INTERACTION DESIGN E PAPER WEARABLES

Giorgio Klauer
Conservatorio C. Pollini di Padova
giorgio.klauer@conservatoriopollini.it

Annalisa Metus
Artista freelance
annalisa.metus@gmail.com

Pietro Polotti
Conservatorio G. Tartini di Trieste
pietro.polotti@conts.it

ABSTRACT

In questo articolo vengono presentati i risultati di un workshop svoltosi presso il Conservatorio di Musica "C. Pollini" di Padova, ideato e condotto dagli autori. Il tema principale è stato l'utilizzo del suono nel processo del design del prodotto nell'ambito dell'abbigliamento e degli accessori da indossare. Ne sono risultati tre progetti che hanno indagato ruoli interattivi del suono in tre scenari diversi: a) astratto e relazionale, b) strettamente funzionale, c) estetico/performativo nell'ambito della moda. La presentazione e discussione dei tre progetti è preceduta da una serie di considerazioni sull'approccio metodologico adottato.

1. IMPLICAZIONI ESTETICHE NELL'AMBITO DEL SONIC INTERACTION DESIGN

Quale può essere la relazione tra ciò che viene insegnato nei corsi di musica elettroacustica e lo sviluppo di una pratica della composizione del suono, che deve soddisfare degli specifici requisiti funzionali ed estetici in un contesto di design del prodotto? Quale può essere il nesso tra *Musique Concrète*, ascolto ridotto, ascolto musicale, ascolto ecologico, suoni funzionali, design del suono, ovvero tra musica elettroacustica e sonificazione? Tali domande portano a una riflessione sul valore che l'esperienza e le conoscenze maturate nella tradizione della musica elettroacustica possono avere per il crescente utilizzo di suoni artificiali in chiave di realtà aumentata nell'ambito del design industriale [1]. Il workshop svolto al Conservatorio di Padova, a cui hanno partecipato studenti del corso di Musica elettronica dei Conservatori di Padova e di Venezia oltre ad alcuni *freelance*, aveva lo scopo di far mettere in pratica ai partecipanti le proprie conoscenze in ambito musicale elettroacustico in un contesto diverso dal solito, vale a dire quello del design del prodotto. È stato infatti richiesto a persone abituate a pensare in termini di "suono organizzato" mirato alla realizzazione di lavori musicali e istruite a pensare al suono in termini di oggetti sonori [2, 3] di applicare le proprie competenze nell'ambito del design dell'interazione finalizzato alla creazione di "oggetti sonanti" [4]. Dal punto di vista metodologico, si è trattato di sfruttare le teorie e la pratica sviluppate a partire dal pensiero Schaefferiano per la progettazione e realizzazione di

suoni interattivi inseriti in un ambito d'uso di oggetti quotidiani. Suoni, quindi, ideati ed elaborati seguendo i metodi analitici dell'Ascolto ridotto per essere poi impiegati in un contesto di ascolto ecologico e causale [5, 6, 7].

In una prospettiva in cui il *soundscape* "antropofono" è sempre più caratterizzato da suoni artificiali digitali riprodotti attraverso altoparlanti, lo studio del cosiddetto ascolto ecologico o quotidiano [8] rappresenta una via fondamentale per poter disporre di strumenti adeguati per la valutazione dal punto di vista psicologico di tali suoni artificiali in un contesto di design del prodotto [9]. Accanto all'aspetto della valutazione è necessario considerare anche quello della progettazione e produzione del suono, altrettanto fondamentale in un processo di design che includa le qualità soniche degli oggetti d'uso quotidiano. Il workshop intendeva concentrarsi su quest'ultimo aspetto, introducendo le problematiche del *Sonic Interaction Design* (SID) in un Conservatorio di musica ed esplorando, tra l'altro, un ambito di applicazione relativamente nuovo, quello degli indumenti [10]. È infatti convinzione degli autori che le conoscenze e le pratiche della musica elettroacustica possano fornire un contributo determinante alla formulazione di una metodologia per l'aspetto poetico del design dei suoni non meccanici riprodotti da altoparlanti, secondo le finalità del SID.

2. CONTESTO DI SVOLGIMENTO

Il workshop è stato promosso dal Conservatorio di Padova nell'ambito di SaMPL¹, una realtà cross-dipartimentale creata nel 2010 dalla collaborazione con il Dipartimento di Ingegneria dell'informazione dell'Università. Realizzata seguendo il modello del *living lab*, essa promuove attività di ricerca, produzione musicale e formazione legate alle tecnologie e ai *topic* del *Sound and Music Computing* (SMC). In quanto *living lab*, SaMPL è da considerarsi più come una metodologia e una comunità ristretta che coinvolge le istituzioni ospitanti, che non un'infrastruttura.

Rinviano al sito internet per quanto riguarda i settori della produzione e della ricerca, in questo contesto è sufficiente soffermarsi sulle attività di formazione, che consistono nel trasferimento della conoscenza derivata dalla ricerca sui *topic* del SMC attraverso una serie di laboratori, articolati in base a bisogni manifestati nell'ambito di una pratica artistica riflessiva nel merito dei propri metodi, risultati e rapporti con il contesto accademico - in breve, nell'ambito di ciò che è stato definito negli ultimi anni a livello

Copyright: ©2016 Giorgio Klauer et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ <http://www.sampl-lab.org>

della comunità delle accademie musicali europee², ricerca artistica in musica. I workshop sono denominati SaMPL School, si basano su chiamata pubblica - adottando eventualmente processi di revisione - e privilegiano la forma di lavoro collaborativa e l'interdisciplinarità. Essi hanno la caratteristica di orientare le attività alla presentazione pubblica dei risultati nell'ambito di manifestazioni di contenuto tecnologico/scientifico e artistico combinanti la forma del concerto, del *talk* e dell'installazione interattiva³.

I workshop hanno in comune l'attenzione al processo progettuale. Normalmente essi constano di tre fasi. La fase iniziale serve a trasmettere gli apparati concettuali e a formulare ipotesi di lavoro - o nell'ambito della revisione dei progetti proposti o all'interno di un esteso momento di carattere seminariale. La seconda consiste nell'attività laboratoriale in senso stretto. La terza fase è costituita dall'implementazione dei progetti.

In genere la fase centrale del workshop è circoscritta in due-quattro giornate; se la fase iniziale è ricompresa nel workshop, tra questa e la fase centrale intercorre almeno una settimana di tempo. All'implementazione segue la pubblicazione dei risultati, in tempi e modalità non predefiniti.

Ulteriore caratteristica comune dei workshop è la scelta di separare in modo netto quelle che sono le istanze di carattere artistico dai fondamenti metodologici delle discipline cui afferiscono le tematiche affrontate, considerato che i partecipanti posseggono livelli di competenza spesso differenti, essendo ammessi alle attività studenti di primo e secondo livello, diplomati/laureati, ricercatori e professionisti.

In buona sostanza, con le School si intende creare nell'ambito accademico di formazione artistica un formato di workshop ispirato ai cicli di vita della progettazione, in particolare secondo il modello del design dell'interazione [11, p. 165–197], con l'obiettivo di progettare artefatti artistici quali installazioni, strumenti musicali, oggetti sonanti e lavori sonori connessi a uno specifico apparato concettuale e/o supporto materiale con rilevanti implicazioni estetiche, e la finalità generale di proporre un *testbed* di innovazione nel contesto dell'alta formazione musicale.

3. STRUTTURA DEL WORKSHOP

Il workshop in oggetto è stato preceduto da due altri laboratori di *Sonic Interaction Design*, “Sounding Popables” (2011)⁴ e “Bozzetti sonori, voce e gesto nel design dell'interazione sonora” (2014)⁵, centrati rispettivamente sulla narratività mediante artefatti interattivi basati su supporti cartacei (*pop-up*) e sul *vocal sketching*. Un'ulteriore

premessa è stata “Data Sonification as Artistic Practice” (2015)⁶.

Per similarità negli aspetti materiali e ciclo di vita è utile richiamare in particolare il workshop “Sounding Popables” (2011), in cui l'interazione sonora doveva riprendere le possibilità narrative non-lineari implicite nella fruizione del *pop-up* dimostrandone le potenzialità estetiche e riflettendo sulla possibilità di rendere i *pop-up* aumentati mediante interazione sonora delle vere interfacce tangibili strumentali al design dell'interazione uomo-macchina. Obiettivo del workshop era infatti valutare il grado di permanenza nel tempo di un libro *pop-up* aumentato e riflettere se questo possa considerarsi un'interfaccia affettiva e risultare una tecnologia valida per la vita di ogni giorno [12].

3.1 Scelta di utilizzare la carta

Il workshop del 2016 intendeva approfondire l'aspetto metodologico del processo di design ponendo attenzione al momento ideativo, al materiale cartaceo quale strumento di prototipazione rapida e alle possibilità di *embedding* della tecnologia.

Si è pensato di utilizzare la carta e il cartone come materiali per la realizzazione dei prototipi con l'intento principale di facilitare il passaggio dallo schizzo all'oggetto, essendo la carta un materiale con il quale tutti hanno un certo grado di dimestichezza.

Un prototipo di carta permette inoltre una rapida sostituzione o adattamento di singole parti (grazie a tecniche di disegno basilari come il ricalco di una sagoma) o la loro duplicazione (ad esempio mediante scansione o fotocopia). L'assemblaggio e smontaggio durante la fase di prototipazione può avvenire attraverso nastro adesivo (ri-posizionabile, biadesivo o a forte tenuta) e punti e graffette metalliche di diverse forme e dimensioni.

L'utilizzo della carta a discapito della stoffa, inibendo la possibilità di aumentare abiti già esistenti, aveva inoltre l'obiettivo di forzare l'immaginazione di capi da indossare diversi dagli indumenti cui siamo abituati e di far ragionare in termini di funzioni da assolvere (anche tramite il suono) e movimenti/azioni da assecondare e tracciare.

Coerentemente, ciascun gruppo aveva al suo interno un componente con spiccata familiarità con la modellazione per mezzo della carta (una sarta modellista, una *paper designer* e un *designer*).

3.2 Programma

Il workshop è stato pensato in due segmenti rispettivamente di una e due giornate con un intervallo di dieci giorni nel mezzo. Una *call for participation* si è aperta un mese prima dell'inizio, completa di descrizione degli obiettivi minimi, modalità di svolgimento e riferimenti bibliografici⁷.

Il primo segmento era costituito da una parte seminariale sul SID svolto dal terzo autore e da ogni riferimento

² <http://www.aec-music.eu>

³ <http://www.living-lab-music.org>

⁴ http://www.sampl-lab.org/it/index.php/Sounding_Popables e <https://soundingspopables.wordpress.com>, relatori D. Cestaro, J. Qi, D. Rocchesso, coordinamento A. de Götzen e S. Delle Monache.

⁵ http://www.sampl-lab.org/it/index.php/Bozzetti_sonori,_voce_e_gesto_nel_design_dell'_interazione_sonora, relatori D. Rocchesso, S. Delle Monache, D. Mauro e S. Baldan, coordinamento G. Klauer.

⁶ http://www.sampl-lab.org/it/index.php/Data_Sonification_as_Artistic_Practice, relatori F. Cifariello Ciardi e M. Rumori, coordinamento G. Klauer e A. Vigani.

⁷ http://ocs.sme-ccppd.info/index.php/SaMPL_Schools_2016/SID_PAA_2016/schedConf/overview

utile allo svolgimento dei lavori, e da una seconda parte articolata in approfondimenti metodologici che hanno dato il via al lavoro concettuale e pratico, finalizzato a programmare gli obiettivi da raggiungere nelle giornate del secondo segmento e le eventuali attività preparatorie da svolgere individualmente.

L'intervallo tra il primo e il secondo segmento è servito a maturare le bozze progettuali⁸.

Il secondo segmento era dedicato alla verifica dei progetti iniziali e alla prototipazione. Chiudeva il secondo segmento un momento dimostrativo e di videodocumentazione.

Nella *call* si trascurava volutamente l'ambito di applicazione, concentrandosi sul vincolo di carta e cartone quali materiali privilegiati ma non esclusivi dove alloggiare altoparlanti miniaturizzati, circuitazione elettronica realizzata espressamente, sensori, microcontrollori e possibilmente *single board computer*: i materiali sarebbero stati usati come forma, supporto e *mock-up* mentre l'interazione sonora sarebbe servita a creare oggetti d'arte oppure prototipi di oggetti puramente funzionali. Come obiettivo specifico e circoscritto veniva indicato semplicemente un artefatto con il sistema di interazione software e hardware *embedded* comprensivo di sensori, trasduttori sonori e microcomputer oppure microcontrollore. Lo scopo di approfondire la nozione di *embedding* era dichiarato quale condizione tecnica necessaria per poter iscrivere l'oggetto nella logica del design; ciò avrebbe dovuto limitare l'eccentricità dell'approccio artistico rispetto a quello funzionale.

Ulteriore particolarità della *call* era l'aspetto collaborativo. Proponendo che ogni partecipante portasse oltre al proprio *know how* qualsiasi elemento tra oggetti e materiali cartacei preconfezionati, abbozzi di opere, circuiti, software per hardware *embedded*, suoni preregistrati, algoritmi, dispositivi di interazione, si richiedeva di includere nel documento di iscrizione, oltre a ogni informazione utile relativa alle competenze e ai supporti/strumenti/dispositivi, l'indicazione se questi fossero condivisibili e integrabili con quanto messo a disposizione dall'organizzazione.

4. PASSI SEGUITI NELLO SVILUPPO DEI PROGETTI

4.1 Definizione del tema

Il tema inizialmente generico sull'*interaction design* e la sonificazione di prototipi in carta di oggetti d'arte e d'uso quotidiano è stato circoscritto al tema più specifico degli indumenti e accessori solo pochi giorni prima dell'inizio del workshop. Si è quindi deciso di delimitare il territorio d'azione della fase laboratoriale per concentrare le risorse e orientare le idee verso risultati condivisi, partecipati e confrontabili. D'altra parte, quello degli indumenti e degli accessori è un argomento oltremodo vasto, che poteva essere interpretato in modo estremamente diversificato, lasciando spazio alle differenti competenze e inclinazioni delle varie personalità coinvolte nel workshop. In effetti,

⁸ Attraverso una *mailing list* i partecipanti avevano la possibilità di aggiornarsi reciprocamente e anticipare le bozze prima dell'avvio del workshop vero e proprio.

come vedremo, i tre progetti risultanti sono stati estremamente diversi tra loro, anche in virtù del fatto che in ognuno dei tre gruppi era presente una persona esterna ai corsi di Musica elettronica, con competenze che andavano dal design, al *paper engineering*, alla sartoria di moda.

4.2 Prima giornata (7 marzo 2016)

Dopo la mattinata introduttiva a carattere seminariale, la prima parte del laboratorio è consistita nella proposta di idee e scenari attraverso una fase di *brainstorming* collettivo seguito da una messa ai voti di tutte le idee prodotte. Le tre proposte più votate e quindi selezionate sono state:

1. la sonificazione della relazione tra due o più individui, in senso astratto, secondo un'accezione orientata ai criteri riduzionisti del *basic design*, che mira ad affrontare classi di problemi il più generali possibile [13]. In questo caso l'indumento è stato inteso in senso generico come strumento per l'*embedding* tecnologico.
2. La sonificazione delle azioni funzionali di una figura professione che contempla l'uso della divisa e la cui gestualità è univocamente riconducibile ai compiti svolti (ad es. il barista).
3. La sonificazione di un abito inteso come oggetto performativo, grazie al quale condurre una *performance* vera e propria.

A questo punto si è proceduto alla suddivisione dei partecipanti in tre gruppi di cinque persone ciascuno. Ognuno dei gruppi comprendeva una persona con esperienza di design e/o di sartoria e/o di utilizzo della carta come materiale di modellazione. In vista della fase laboratoriale vera e propria, che si sarebbe svolta dopo dieci giorni, a ciascuno dei tre gruppi è stato dato il compito di lavorare a distanza, via mail o teleconferenza, per la definizione di uno *story-telling* e l'individuazione delle azioni da sonificare.

4.3 Seconda giornata (18 marzo 2016)

La seconda giornata del workshop si è svolta, come previsto, dopo circa dieci giorni. È stata interamente dedicata alle attività laboratoriali come la giornata successiva, al fine di giungere tramite un lavoro intensivo a dei prototipi funzionanti. Ogni gruppo ha ripreso il proprio soggetto discusso a distanza.

Sono stati riconsiderati ed elaborati gli *story-telling* riguardanti ciascuna delle idee prescelte, individuando e analizzando le azioni da sonificare e discutendo e descrivendo verbalmente i possibili suoni da utilizzare. Successivamente gli scenario/*story-telling* elaborati sono stati presentati da ciascun gruppo agli altri due gruppi. Questi ultimi avevano il compito di commentare e fornire eventuali suggerimenti.

La fase seguente prevedeva un'analisi e segmentazione più dettagliata degli *story-telling* in base alle finalità (*Task Analysis*), il tipo di azione necessaria per raggiungere il task (*Action Analysis*), il tipo di gesto implicato dall'azione

(*Gesture Analysis*) e, infine, il tipo di suono da utilizzare per la sonificazione interattiva (*Sound Analysis*).

Prendendo ad esempio lo scenario del barista, l'analisi seguiva questo schema:

1. task: aprire/chiudere una bottiglia;
2. azione: ruotare il tappo;
3. gesto da sonificare: ruotare il polso;
4. suono: continuo con leggero glissando ascendente o discendente a seconda del verso di rotazione.

Mediante un'ulteriore fase di *brainstorming* a gruppi separati, il tema 2 si è focalizzato sulla figura della maschera teatrale, cioè l'addetto all'accoglienza in un teatro lirico o di prosa, mentre il tema 3 è diventato il *défilé* di moda.

Come passo successivo, ogni gruppo ha lavorato sulla proposta di più suoni (almeno tre) per la sonificazione dei gesti individuati e analizzati, e ha realizzato una versione in chiave *mock-up* dell'interazione usando eventualmente la tecnica *Wizard of Oz* (una persona del gruppo eseguiva i gesti mentre un'altra riproduceva al computer in playback le varie possibili sonificazioni interattive). Conclusa la definizione dei suoni e del *mapping* interattivo, ogni gruppo ha presentato agli altri due gruppi le versioni *mock-up* della sonificazione e i secondi hanno espresso un voto per le diverse alternative di suoni proposti per ogni gesto. Questa procedura è stata considerata come una sorta di test mediante esposizione a un pubblico di potenziali *user*.

Infine, al termine della giornata, ogni gruppo ha discusso quali sensori avrebbe dovuto utilizzare nella giornata seguente, per la realizzazione effettiva del prototipo di sonificazione interattiva.

4.4 Terza giornata (19 marzo 2016)

La terza e ultima giornata è stata dedicata alla realizzazione dei prototipi. La sensoristica dedicata ai *wearable* e una certa quantità di dispositivi come altoparlanti e amplificatori in miniatura, *single board computer*, schede Arduino e simili sono stati forniti dal Conservatorio. Alcuni dei partecipanti hanno utilizzato altra tecnologia propria⁹.

Il Conservatorio ha fornito inoltre l'occorrente per la lavorazione della carta come elemento plastico per l'effettiva costruzione dei prototipi.

Nella fase di elaborazione dei prototipi, ovviamente la parte più critica è risultata essere il *mapping* tra i suoni prescelti e i dati dei sensori. In un così breve tempo, i risultati sonori non hanno potuto eguagliare le dimostrazioni dove il suono era controllato in tempo reale da un "esecutore" al computer. Ad ogni modo, ogni gruppo ha potuto presentare i propri risultati agli altri due gruppi. Come nelle altre fasi, i due gruppi che costituivano il pubblico hanno liberamente commentato i risultati prototipali.

⁹ Un gruppo del workshop ha usufruito in anteprima della piattaforma di prototipazione audio *embedded Bela* (<http://bela.io>) sviluppata presso la Queen Mary University di Londra.

Infine ogni gruppo ha realizzato una breve presentazione in forma audiovisiva del prototipo, operazione con la quale il workshop si è concluso¹⁰.

5. I PROGETTI

5.1 Défilé

Il progetto del primo gruppo¹¹ prevedeva come scenario il *défilé* di moda per un abito realizzato in carta e cartone. L'abito di carta andava inteso sia come prototipo del vestito di stoffa sia come manufatto artistico.

La struttura dell'oggetto è stata concepita modularmente, con un corpetto in cartone usato come alloggio per i supporti tecnologici meno leggeri la cui presenza era dissimulata da elementi decorativi applicati di carta. Al corpetto sono state attaccate delle code composte da dischi di carta sovrapposti e inanellati attorno a un filo di tessuto conduttore. Le code, disegnate come strascico mobile in sospensione, attribuivano all'indumento un certo peso; potevano essere staccate durante la sfilata, in modo da evidenziare la molteplicità del capo (gonna corta/gonna lunga) e possibili utilizzi alternativi delle estensioni stesse (ad es. come accessorio - sciarpa).

Anche se lo scenario del *défilé* non era così funzionale come l'abito considerato nel suo utilizzo quotidiano, si presentavano elementi d'uso fortemente codificati e prestabiliti su cui fondare il processo di design, quali l'ingresso sulla passerella, il passaggio fino al fondo della stessa, movimenti quali piroette, fermate improvvise (messa in mostra dell'abito al pubblico) e passaggio di ritorno. Di conseguenza la sfilata è stata trattata più come una sorta di *performance/coreografia* che uno scenario quotidiano ed è stato necessario stabilire uno *storyboard* per un'acattivante sonorizzazione dei gesti nei tempi e modalità dell'esibizione del capo d'abbigliamento, attraverso l'abito stesso.

Il *sound design* si è basato principalmente sul campionamento sonoro dei processi di lavorazione del vestito, in questo caso rumori di taglio e strappo, piegatura e accartocciamento, frizioni e sfregamenti vari atti a evidenziare le tipologie di superficie e consistenza del materiale. I campioni sonori sono stati utilizzati sia come elementi ritmici da associare, in modo automatico e interattivo, al movimento, sia come elementi da editare per creare *texture* sonore continue o granulari a densità variabile, ai fini della creazione di un ulteriore piano semantico. Erano stati ipotizzati anche passaggi graduati tra un modello e l'altro. È stato inoltre contemplato un elemento di contrasto quale un suono sintetico, da associare a uno dei momenti topici che contraddistinguono l'uso dell'indumento nel *défilé* - in questo caso, la sosta davanti al pubblico/ai fotografi - e ulteriori elementi sonori di appoggio di funzione contestuale.

Lo scenario composto da gesti/movimenti e suoni è stato il seguente:

¹⁰ <http://repo.sme-ccppd.info/prog/sid2016>

¹¹ Francesca Tavarado, Freelance; Giulio Moro, Queen Mary University London; Lucia Violetta Gasti, Matteo Polato, Nicola Sanguin, Conservatorio di Padova.



Figure 1. Test del prototipo da *défilé*. Sullo sfondo, Beagle Bone Black+Bela prima di essere alloggiati nel corpetto.

- quale situazione di base, prima dell'entrata sulla passerella, si è prevista una sonorizzazione astratta, di struttura caotica e non ritmica; i suoni utilizzati sono stati impatti di materiali vetrosi/metallici e gli esiti di processi di sintesi granulare;
- all'ingresso dell'indossatrice si è prevista una seconda sonorizzazione collegata agli aspetti materiali e al processo di fabbricazione dell'indumento;
- durante la camminata, tipicamente rapida da parte degli indossatori, il movimento dei passi determinava un processo di bilanciamento alternativo delle due soluzioni sonore ma, progressivamente, a ogni passo si sottraeva intensità alla sonorità che identificava la situazione prima dell'ingresso; all'avvicinamento dell'indossatrice al pubblico (cioè al punto di svolta prima del ritorno) il procedimento è stato reso in modo più esplicito: la cadenza ritmica degli eventi sonori diventava complessa ma, allo stesso tempo, emergeva sempre più evidente la connessione con il gesto;
- alla fermata l'indossatrice poggiava il palmo della mano sul fianco attivando un suono differente che, nello specifico scenario, enfatizzava il momento degli scatti fotografici complementandone, peraltro, il suono;
- la rimozione delle code e in generale lo spostamento degli elementi del vestito quali accessori era associata a una *texture* sonora ad hoc caratterizzata da una morfologia con un preciso orientatamento, ad esempio una "cascata" di suoni o glissandi;
- alla piroetta seguivano la camminata di ritorno e l'inversione, con modificazioni, dei processi eseguiti nella prima parte: attenuazione/disaccumulo al posto di intensificazione/accumulo, presenza di suoni differenti quali elementi intonati o musicali, maggiore continuità e persistenza del suono.

Nell'implementazione, il cuore del sistema interattivo è stato realizzato mediante un *single board computer* Beagle

Bone Black combinato alla scheda Bela. Sensori di luminosità servivano a determinare la collocazione dell'indossatrice in scena/fuori scena e in maggiore prossimità dei riflettori. La rilevazione dei passi e l'analisi della camminata era svolta sulla base delle informazioni estraiibili da sensori piezoelettrici collocati nelle calzature e accelerometri posti a un fianco dell'indumento. Gestì specifici quali la posa con la mano sul fianco erano rilevati da sensori di pressione e interruttori realizzati mediante tessuto conduttivo.

Il tessuto conduttivo utilizzato per le sospensioni a strascico presentava elevata resistività e rendeva la struttura di carta un sensore elastico. La rimozione delle code implicava l'interruzione del circuito, così come la loro collocazione o differente collocazione/utilizzo determinava una nuova inserzione circuitale, con trasformazioni elettriche facilmente utilizzabili come segnali d'avvio di processi sonori. Le cospicue variazioni elettriche legate all'interazione manuale erano altresì usate come elemento di controllo.



Figure 2. Elementi sospesi dell'abito da *défilé*.

Il suono generato dal dispositivo *embedded* era trasmesso via bluetooth ad altro dispositivo di amplificazione e diffusione, nel rispetto dello scenario della sfilata di moda.

5.2 Maschera di teatro lirico o di prosa

Il gruppo della maschera teatrale¹² ha individuato il proprio scenario dopo aver valutato le alternative dell'*hostess* e il barista, della guida museale, dell'uomo *sandwitch*, abbozzando per ciascuno l'analisi dei *task* e dei gesti. Lo scenario individuato riguardava dunque una maschera che agisce in sala o nelle immediate vicinanze e si relaziona con gli spettatori singolarmente o per comunicazioni di gruppo assieme ai colleghi.

Dopo aver valutato sistematicamente i *task* della maschera di teatro, una sorta di *Storyboard* ha permesso di focalizzare i quattro compiti relativamente ai quali perfezionare l'analisi di azioni e gesti ed effettuare il design del suono:

Mi viene incontro uno spettatore, lo saluto...
Mi chiede indicazioni sul posto: glielo indi-

¹² Annalisa Metus, Freelance; Sebastiano Aleo, Francesco Roberto Dani, Eric Guizzo, Conservatorio di Padova; Gianmarco Bellato, Conservatorio di Venezia.



Figure 3. Struttura dell'abito della maschera teatrale.

co/lo accompagnano... Segnalo assieme ai miei colleghi che lo spettacolo sta per iniziare... Ammonisco chi non fa silenzio...

Come nel caso del *défilé*, ogni compito doveva svolgersi in assenza di comunicazione verbale, quindi unicamente mediante l'interazione coi suoni, in questo caso diffusi da altoparlanti *embedded*:

- l'accoglimento delle persone è stato combinato con il richiamo dell'attenzione per mezzo di gesti estremamente stilizzati di saluto e benvenuto (distensione o innalzamento delle braccia, inchino) sia individuati che eseguiti in gruppo;
- l'indicazione e l'accompagnamento al posto sono stati realizzati con gesti deittici del braccio, da fermi o in movimento;
- l'indicazione dell'inizio dello spettacolo è stata realizzata per mezzo di un gesto iconico, il battito del dito sul polso, eseguito in gruppo;
- l'ammonizione è stata rappresentata dal gesto di portare il dito indice davanti alla bocca o dal gesto “no” eseguito con lo stesso dito.

Nel design, l'espressività soggettiva associata normalmente ai gesti doveva essere mutuata dal suono. Nell'implementazione, i diversi compiti sono stati realizzati sia con interazioni sonore continue che discrete, scelte in base al livello di soggettività da esprimere. Interazioni discrete sono stati ad esempio il ticchettio del metronomo attivato dal battito (contatto elettrico) del dito sul polso e la stilizzazione sintetica del suono “sssh” (alveo-palatale) attivata da un sensore di prossimità collocato al posto del *papillon*.

Per l'indicazione e l'accompagnamento al posto è stata utilizzata una sonificazione di tipo continuo: in questo caso, il dispositivo di sintesi sonora era controllato mediante l'analisi dell'informazione estratta da un sensore accelerometrico e giroscopico posto al polso; l'algoritmo sonoro prevedeva due livelli di complessità, rispettivamente un rumore a banda limitata e una variante di una storica modellizzazione *waveguide*¹³. In quest'ultimo caso le possibi-

¹³ Lo *slide flute* di Perry Cook.



Figure 4. Maschera teatrale: UDOO sulla schiena.

lità del *wearable* interattivo si avvicinavano a quelle di una performance.

La struttura dell'indumento riprendeva quella del progetto *défilé*, con una giubba al posto del corpetto quale alloggio per la piattaforma di prototipazione UDOO.

5.3 Indumento come strumento relazionale

Il lavoro del terzo gruppo¹⁴ è cominciato con un intenso *brainstorming* sull'opportunità di applicare alcuni modelli della psicologia ambientale, delle teorie dell'attaccamento e della teoria dei giochi, focalizzandosi infine su uno scenario minimale concentrato su tre fattori elementari quali a) uno spazio delimitato, b) un set di condizioni psico-percettive per indurre processi decisionali e c) l'interazione tra due soggetti.

Lo scenario consisteva in un gioco/*performance* nel quale due soggetti entravano in uno spazio delimitato e i loro movimenti e interazione (avvicinamento/allontanamento e contatto) consentivano di riorganizzare la sonorità dell'ambiente conferendogli una forma specifica: un *soundscape* inizialmente confuso e frastornante cominciava a variare in base alle loro azioni ed essi intuivano come raggiungere una posizione e rapporto tali da generare un suono armonioso e gradevole, o il silenzio, o altra singolarità sonora decisa in funzione della qualità emozionale da attribuire all'esperienza.

Anche in questo progetto, l'interazione sonora doveva sia assolvere a funzioni di orientamento e indirizzo precise, sia alimentare la soggettività, in questo caso l'arbitrio nei processi decisionali.

Più si avvicinano e più sentono che il volume cala, allora continuano ad avvicinarsi, fino a quando sono vicinissimi, ma ancora si sente qualcosa. Si allontanano, l'intensità aumenta, allora si avvicinano di nuovo. Si avvicinano sempre di più. Sono vicinissimi, ma ancora si sente qualcosa. Iniziano a muovere le braccia fino a che poi si toccano e il suono sparisce del tutto.

¹⁴ Arcangelo Di Donato, Freelance; Simone Sacchi, Conservatorio di Venezia; Razvan Hanganu, Giovanni Salice, Conservatorio di Padova; Filippo Zampinetti, Freelance.



Figure 5. Brainstorming del terzo gruppo.



Figure 6. Test dello strumento relazionale.

L'applicazione dello scenario poteva essere pensata per livelli dal più facile al più difficile o in maniera maggiormente estetica/esperienziale, posto comunque l'obiettivo di realizzare un'esperienza intima tra due persone che esplorano se stesse e il rapporto con l'altro tramite il suono e lo spazio.

Nell'ideazione, gli aspetti estetici erano stati posti in relazione ad alcuni modelli di *performance* situazionista ed *endurance art*, i cui fenomeni superficiali non sembravano contraddirsi i criteri riduzionisti del *basic design* cui il gruppo metodologicamente si ispirava. Tali criteri hanno permesso in ultima analisi di considerare l'ambiente delimitato e il processo di avvicinamento come fattori della costruzione dell'identità affettiva (psicologia ambientale e teorie dell'attaccamento) e i meccanismi decisionali come prodotto di *swarm intelligence*.

L'indumento è stato inteso pertanto in senso astratto-funzionale e le qualità estetiche dell'oggetto messe in secondo piano. L'implementazione ha seguito un approccio minimale a partire da un *mock-up* video realizzato filmando in *stop motion* un teatrino di carta. Nel prototipo, la rilevazione spaziale dei soggetti è stata realizzata mediante sensori di distanza cablati a un microcontrollore a sua volta collegato a un computer: i sensori erano alloggiati in una giubba di cartone. Il *sound design* è stato progettato in modo da poter identificare tramite *pattern* ritmici in particolar

modo il processo di avvicinamento reciproco.

6. CONCLUSIONI

L'introduzione delle problematiche del SID all'interno di un Conservatorio può risultare uno stimolo all'innovazione della metodologia dell'insegnamento artistico [1]: in tal senso questo workshop costituisce un'esperienza a carattere pionieristico ed esplorativo con esito complessivamente positivo.

Oltre alla portata strutturale del trasferimento metodologico, è da notare come la visione delle versioni alternative e della progettazione come strumento per soddisfare bisogni di portatori di interesse anche solo potenziali, non è estranea al processo creativo musicale in sé.

La forma collaborativa imposta dal workshop ha permesso di impiegare le competenze artistiche e musicali individuali in progetti condivisi, in particolare modo da parte dei più esperti.

L'uso della carta quale materiale facilmente manipolabile e neutro ha effettivamente permesso di stimolare l'immaginazione dei partecipanti nel progettare l'interazione sonora evitando possibili inibizioni o ingenuità dovute alle limitate conoscenze negli ambiti specifici della produzione dei tessuti, artigianato sartoriale e tecnologie dell'abbigliamento.

Allo stesso modo, lo sviluppo dei progetti a partire dal *mock-up* e la possibilità di modulare il livello di *embedding* ha impedito alle problematiche tecniche di prendere il sopravvento rispetto al momento ideativo.

7. REFERENCES

- [1] P. Polotti, "Closing the loop of sonification and electroacoustic music practices through gesture and embodied interaction," *The New Soundtrack*, vol. 4, no. 2, 2014.
- [2] M. Chion, *Guide des objets sonores*. INA/Buchet Chastel, 1983.
- [3] P. Schaeffer, *Traité des objets musicaux*. Seuil, 1966.
- [4] D. Rocchesso and F. Fontana, eds., *The Sounding Object*. Mondo Estremo, 2003.
- [5] W. Gaver, "What in the world do we hear? an ecological approach to auditory source perception," *Ecological Psychology*, vol. 5, no. 1, pp. 1–29, 1993.
- [6] W. Gaver, "How do we hear in the world? explorations in ecological acoustics," *Ecological Psychology*, vol. 5, no. 1, pp. 285–313, 1993.
- [7] F. Avanzini, "Interactive sound," in *Sound to Sense, Sense to Sound. A State of the Art in Sound and Music Computing* (D. Rocchesso and P. Polotti, eds.), pp. 345–396, Logos, 2008.
- [8] B. Giordano, "Everyday listening: an annotated bibliography," in *The Sounding Object* (D. Rocchesso and F. Fontana, eds.), pp. 1–13, Mondo Estremo, 2003.

- [9] B. Giordano, P. Susini, and R. Bresin, “Experimental methods for the perceptual evaluation of sound-producing objects and interfaces,” in *Sonic Interaction Design* (K. Franinović and S. Serafin, eds.), pp. 151–197, MIT, 2013.
- [10] L. Elblaus, V. Tsaknaki, V. Lewandowski, and R. Bresin, “Nebula: An interactive garment designed for functional aesthetics,” in *CHI'15 extended abstracts*, (Seoul), 2015.
- [11] J. Preece, Y. Rogers, and H. Sharp, eds., *Interaction Design*. Wiley, 2002.
- [12] S. Delle Monache, D. Rocchesso, J. Qi, L. Buechley, A. De Götzen, and D. Cestaro, “Paper mechanisms for sonic interaction,” in *Proceedings of the 6h International Conference on Tangible, Embedded and Embodied Interactions*, (Kingston), 2012.
- [13] G. Anceschi, “Basic design, fondamenta del design,” in *L'ambiente dell'apprendimento - Web design e processi cognitivi* (G. Anceschi, M. Botta, and M. Garito, eds.), pp. 56–67, Mc Graw-Hill, 2006.

RENEGOTIATING RESPONSIBILITIES IN HUMAN-COMPUTER ENSEMBLES

Fabio Morreale

Centre for Digital Music, EECS
Queen Mary University of London
f.morreale@qmul.ac.uk

Raul Masu

interAction Lab
University of Trento
raul.masu@unitn.it

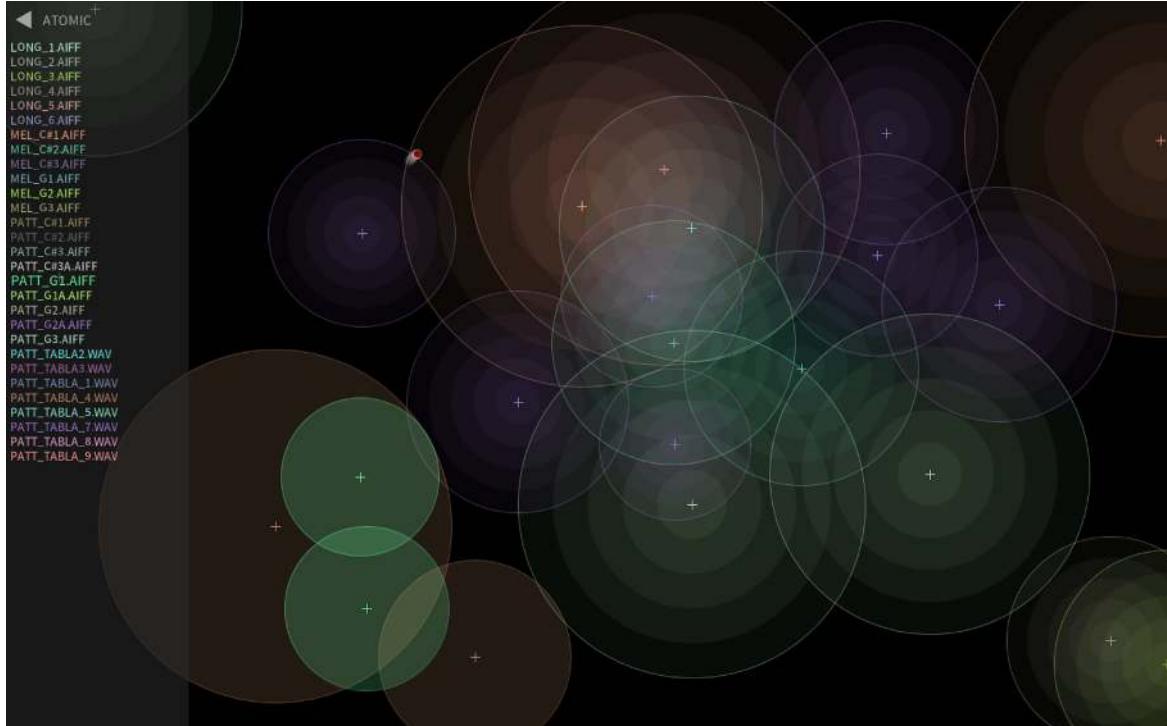


Figure 1. Chimney, the software used as a probe to reflect on new scenarios of human-machine collaborations.

ABSTRACT

This paper proposes discussions for the (re)negotiation of responsibilities in performances that involve a collaboration between human and computer agents. Most of current research is human-hierarchical, leaving the machine the role of a mere tool at performer's service. In this paper a more balanced distribution of responsibilities between the two agent is proposed. Chimney, a software developed by one of the authors, is proposed as a design probe to reflect on this topic. Chimney allows the composer to control only the musical material of a piece, leaving its evolution to an algorithmic agent whose deci-

sions are unpredictable. This redistribution of responsibilities results in a compositional shift that causes the roles of the composer / performer to overlap. The implications for the experience of the performer operating Chimney are also discussed.

1. INTRODUCTION

At the dawn of the last century, the digital revolution offered musicians and researchers the possibility to explore new forms of musical creativity using computers. Besides offering unprecedented interaction possibilities thanks to novel instruments and interfaces, musicians started experimenting with new compositional strategies by detaching human responsibilities to the compositional process, which were partially or completely delegated to computers. Such strategies had been extensively explored by George Lewis, one of the pioneer musical pieces that "emerges from a nonhuman intelligence" [1].

Copyright: © 2016 Morreale and Masu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

In his work, he explored performance interactions between improvising musicians and computer programs that create new music material at performance time.

The work described in this paper keeps up with the encounter between human and algorithmic agents to collaboratively operate to create new music. In particular, we propose discussions about the role of the machine in such hybrid ensemble. Traditionally, computer agents have been relegated as tools to serve the performer. We here advocate a more equal redistribution of the roles of the two agents. Chimney, a software that shares the control over the performance between the musician and the computer, is presented as a probe for discussion. In Chimney the musician-part of the control over the composition is reduced to (i) selecting the musical material to be played during the execution; and (ii) deciding their likelihood of being played temporally closer to each other.

To emphasise the non-hierarchical nature of the ensemble, a visual identity was assigned to the computer agent. An algorithmic *random walker* unsystematically roams throughout the screen: as it encounters an object, the sound associated to that specific object starts playing with a sound level which is proportional to its distance to the walker. The human agent can only decide the objects to be placed on the canvas, their position, and their size. Under these conditions, the musician cannot organise a temporal structure, which is entirely controlled by the algorithmic agent.

Such renegotiation of responsibilities results in two principal compositional shifts: (i) the music is no longer organised according to the phraseological-temporal structure; (ii) the sound objects become the focus of the composition. As a consequence, the system fosters the musician to elaborate new compositional strategies.

It is worth remarking that the scope of interest of this paper lies beyond music and art domains. Chimney belongs to a recent corpus of work of the authors which aims at as well as other works from the authors [4] [5] that encompass reflections on philosophical and cultural concerns. Our pieces are probes that pose questions and speculate about possible future scenarios of computational art. For instance, what would be the consequences of (partially) delegating artistic creation to an autonomous agent? How would the role of art and the artist change?

The remainder of the paper is structured as follows. The next section presents the technical details of Chimney; Section 3 describes the computational shift that a composer needs to face when utilising this system; Section 4 presents Alinearity, a piece written by one of the authors for Chimney and a trumpet; Section 5 discusses more insights about the experience of the trumpet player that interacted with the system.

2. CHIMNEY

Chimney is a multi-platform open source application developed with Processing¹. Quite an efforts was made to

build a strong visual identity to the algorithm for two reasons: (i) to provide the algorithmic agent a *character*, thus emphasising the absence of hierarchy between the two agents; (ii) to help the audience understand the performance. The visual interface is composed of a *canvas* that displays the status of the algorithmic walker and the sonic material.

2.1 The random walker: the fly

The canvas is initially blank with the exception of a particle - the random walker - that moves throughout the screen. The idea was to give the walker a behaviour that recalled that of a *fly*. For this reason, we based the *fly* on an adapted version of the Perlin Noise [6]. The Perlin Noise is a random generator function originally developed to produce natural looking textures on computer generated surfaces. This function produces a more natural, harmonic succession of numbers than that of pure random functions.

To the original Perlin Noise function we added a few features to match our requirements. In particular, the *fly* had:

1. to cover the entire space of the canvas;
2. to be independent of the presence or absence of other objects in the canvas;
3. to have non-deterministic movements: the path taken by the *fly* cannot be predicted nor controlled in any possible ways by the musician.

Figure 2 shows a possible path taken by the *fly* over a period of two minutes.

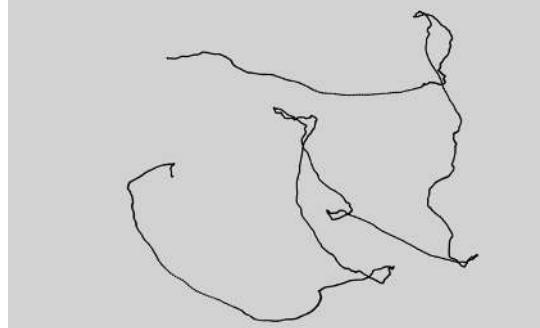


Figure 2. A possible trajectory of the fly.

2.2 The compositional material: sound circles

The performer can interact with the canvas by adding sound sources to it. These sound sources are displayed as circles (Figure 1), whose placement and radius can be controlled by the user interacting with the mouse. Once positioned in the canvas, the performer can interact with it by changing its size, by repositioning it and by deleting it. Every time a new circle is added to the canvas, the system sets it in idle state and mutes it. As the *fly* enters a circle, the sound connected to that circle increases. The maximum level is reached when the *fly* is at the centre of the circle.

The sound sources are made of pre-recorded excerpts stored as audio files. The user can access these files through a sidebar that lists all the sound files contained in a specific folder (Figure 1, on the left). Once selected

¹ <https://processing.org>

a new sound source, the next circle to be positioned in the canvas will be associated to that sound source. A colour code helps identifying what sound each circle is associated to.

3. COMPOSITIONAL SHIFT

The theoretical ideas that lie behind the compositional approach to Chimney was based on the early approach to composition and improvisation of John Cage. In particular, during his first improvisation period (following the Feisst's analysis of Cages' work [2]), Cage categorised compositional process as a fourfold activity composed of: structure, material, method, and form. The *structure* is the temporal division of the sections, the *material* consists of the sound objects (including noise and silence), the *method* is the “note to note procedure”, and the *form* is the “morphological line of the sound-continuity”. Following Cage's opinion, material and methods can be both improvised and composed. Form cannot be composed but only improvised, as it results from the sum of the other activities. Finally, structure is the only element that cannot be improvised.

Having control on the structure allows a composer to be in charge of the evolution of the piece. In fact, he can move back and forth on the music timeline. Sarath formulated this concept proposing the term *expanding temporality* of a compositional process [7]. Given that the composer does not work at performance time, the musical decisions are not directly influenced by the previous events. The composer can *freeze* time, re-think passages, and anticipate future musical phrases. As a result, the development of the piece is based not only on previous events, but also on the anticipation of the material to be presented at successive moments. A different kind of temporality is the *inner directed temporality* [7], a concept proposed to describe the evolution typical of improvisation. When improvising, a musician cannot precisely anticipate how the music will develop. Thus, improvisers can mostly focus their attention on the present, with little attention to the near past and future.

Chimney offers the composer the possibility to redefine the edge of composing and improvising. Given that the temporal evolution of the piece is not under his or her control, the structure can no longer be composed. Chimney restrains human control over the development of the piece as the composer cannot step back and forth and recompose a particular event. As the music is shaped at *run time*, the performance itself represents the final compositional process. As a consequence, performing with Chimney requires an *inner directed temporality* approach. However, at the same, it is not pure improvisation: despite the freedom to take real time decisions the performer cannot improvise the actual musical material.

The musical *material* itself is the element that bounds performance to composition. The musical gesture, limited to organising pre-composed music, is composed out-

side before the performance, thus it is closer to the concept of *expanding temporality*. The musical material becomes the very core of the piece and can be seen as a meta composition of the music, which ultimately generates at performance time.

The *method* touches both the compositional and the improvisational aspects of music making. It is involved both in the creation of the sound sources and the strategies adopted to place them on the canvas. The *form* and the *structure* are both determined during the performance, and consequently are improvised elements.

4. ALINEARITY

This section exemplifies the innovative aspects of Chimney by presenting *Alinearity*, a piece of music composed for Chimney by one of the authors. *Alinearity* was composed to provide the performer with the necessary degree of freedom to properly interact with Chimney while maintaining a degree of harmonic coherence. The piece can be performed as a solo piece for Chimney or as a duo for Chimney and a melodic instrument. *Alinearity* was performed in 2015 at the International Society for Improvised Music Festival (Château-d'Oex, Switzerland) in a version for Chimney and 10 string electric guitar and in 2016 at the International Conference on the Design of Cooperative Systems 2016 (Trento, Italy) for Chimney and Trumpet.

4.1 Description of the piece

The most important aspect of the compositional process adopted for *Alinearity* is the creation of sound objects as the main elements of the piece. The musical form is not determined by the development of musical material moving toward a specific section. Thus, there are no cadences and the material itself does not evolve. The development of the piece is based on the overlapping of different sound excerpts that create different situations. The musical form is shaped by the very transit of the *fly* through different cluster of sounds (the circles in the canvas). This transit can be gradual or sudden, depending on the behaviour or the *fly*.

The musical material was composed to be congruous and at the same time to open a wide range of possible combinations that guarantee coherence and expressiveness to the piece. It consists of 20 short monodic musical excerpts, each with a timbral, melodic, harmonic, and rhythmic value pre-recorded using a synthesiser made with Max-Msp². The synthesiser is controlled with standard MIDI messages.

4.2 Harmony

The harmony of the excerpts is organised as a politonality over two tonality centres: Gmin and C#. The note of the excerpts can either belong to one specific scale or to both scales (C, which enharmonically corresponds to B#, F#, and Bb, which enharmonically corresponds to A#). More specifically, the musical excerpts are harmonically organised as follows:

² <https://cycling74.com/products/max/>

- 7 excerpts belonging to Gmin;
- 7 excerpts belonging to C#;
- 6 excerpts belonging to both tones.

Their combination can create situations with distinct key notes, polytonal harmonies, or situations with a modal appearance that lacks a clear tonic centre.

4.3 Melody and rhythm

From a melodic point of view, the excerpts are clustered into three sets. They can be long pedal notes, very short patterns, and short themes:

- 6 long pedal notes;
- 6 short melodies;
- 8 short patterns.

The harmony of the six long pedal note belongs to both tonal centres. The short patterns have sharper rhythm and can create particularly complex polyrhythmic structures.

4.4 Timbre and register

The excerpts spread across five octaves. Long notes have low register and short notes have high register. As a consequence, the six long pedal notes are the lowest register, the six short melodies lays in the medium register, and the eight short patterns have the highest register.

The core of the synthesiser is a bank of eight resonant band pass filters, which processes white noise in input. The central frequencies of the filters are composed of a fundamental tone and its seven higher harmonic partials. The resulting sound can be accurately manipulated by interacting with its nosiness and brightness, as well as its envelope. Lower notes have a slow transients and a high level of nosiness. Higher notes have fast transients and low nosiness.

5. EXAMINING EXPERIENCES

Alinearity can be performed as a solo interface or in a duo with a melodic instrument. This sections analyses the different experiences of the musicians who (a) performs Chimney, and (b) plays the melodic instrument playing along Chimney.

5.1 Performing Chimney

Given that Chimney has been mainly operated by one of the authors, the following discussion is based on self-analysis of the experience. When using Chimney, composition and performance are superimposed. The emergence of this hybrid composer-performer was thoroughly described by Vallis [8][9]. The implications of this superimposition on the iteration strategies adopted during the performance are substantial. Four main gestures, or approaches to interaction, with Chimney emerged.

Placing excerpts and spectating the movements of the fly
In this first approach, the performer does not consider the movements of the *fly* at all. The gesture is guided by his knowledge of the musical material, thus by placing the excerpts on the canvas. The activities of the human performer have no effects on the musical output, who is relegated to being a spectator of the performer. Depending

on the movements of the *fly*, the actions of the musician can also fail to influence the music at all as the *fly* could potentially never reach the circles. From a musical perspective, this approach is the closest approach to traditional composition. Indeed, it operates in an *expanded temporality*, and all the performers' activities result in long term effects. This approach is typically adopted at the beginning of a performance.

Fostering a gradual passage to new situations

In this approach, the performer positions in the canvas a limited amount of excerpts and follows the movement of the *fly*. As a consequence, the control is more balanced between the human and the algorithm. The performer follows the changes introduced by the transit of the *fly* to a new circle by adding or subtracting elements from the canvas. This approach reduces the time spanning to a more localised present. The performer can focus on the previous events and on the changes that are occurring. The time window could be fairly wide but always centred on the present.

Following the fly to increase music complexity

In this approach, the performer closely follows the movement of the *fly* by adding circles to its trajectory. The human has the highest control over the composition, thus the stochastic aspect is consequently less influent. As the performer gesture are closely related to the movement of the *fly* the temporality is more focused on the present, similarly to the *inner directed* case earlier described (Section 3). This approach is normally adopted to create local musical climax and results in a complex structure. Such complexity results from the number of overlapping excerpts that is achieved by adding several circles that follow the *fly*. The outcome is a much more complex texture articulated by a bigger number of musical objects. Furthermore, the music becomes richer and the quantity of rest decreases.

Removing all the musical excerpts

Chimney consents to delete all the excerpts that are visible on the canvas at once. In this case, the musical gesture is similar to what typically occurs with traditional instruments: to one action corresponds one musical outcome. The temporality is focused on the very time of the key pressing, but is influenced by recent developments in the music. This kind of interaction normally occurs only once during a performance at the end of the last climax. The sudden rest that results from this gesture resolves the tension and gently goes towards the end of the piece.

When Alinearity was performed as a duo piece for Chimney and Trumpet at COOP 2016 in Trento (Figure 3)³, peculiar human-human interactions occurred. The shifting among the different approaches was determined both by the personal taste on the musical output as well as by the input received by the other musician. The trumpeter sometimes proposes shifting that did not follow the movement of the *fly*. The operator of Chimney - one of

³ An excerpt of the exhibition can be found at <https://youtu.be/8-zVa1tN3Ak>

the authors of the paper - who was controlling Chimney was thereby required to pay careful attention on the trumpet. In these case, gradual transits to new situations were not determined by the fly movements but rather by the other musician.



Figure 3. A trumpeter improvising along to Chimney.

5.2 Improvising along to Chimney

This section describes the trumpeter experience of playing along to Chimney. The trumpeter had no score to follow and was instructed about how the musical excerpts were composed. He was also given some time rehearsing the piece to become familiar with the software and the musical material.

After the concert, he was asked to elaborate on his experience of playing along to Chimney. The objective was to understand his musical strategies and to collect reflections about the interaction among the three agents (the trumpeter, Chimney, and the performer of Chimney).

The trumpeter reported that understanding all the details of the music played by Chimney was particularly demanding. In particular, as opposed to traditional improvisation, he did not feel free to propose new musical material. He was always answering to Chimney proposals. Each note, or phrase, was indeed primarily based on listening. As a result, his gestures were mostly about imitating, completing, or opposing what Chimney proposed.

From a musical point of view, the trumpeter was mostly influenced by harmony and rhythms proposed by Chimney. The approaches he adopted can be clustered into three main categories:

- using exactly the same notes of the excerpts;
- using the same scale of the excerpts;
- moving gradually from one scale to the other.

The rhythm could be influenced by the single pattern or by the overlapping of more than one. In the first case, the excerpts themselves determined the rhythmical result. In the second case, he was instead mostly influenced by the movement of the *fly*. With respect to the form, the trumpeter tended to follow the output of Chimney, enhancing his rhythm and loudness coherently with the overlapping of more excerpts.

When prompted to discuss his relation with the two other agents, the trumpeter reported that he experienced two distinct interactions. He perceived to be in closer contact with Chimney as he felt he was duet-ing with it. The performer controlling Chimney was only the access to ask something to the computer agent. Only in a few

cases the main connection was between the two humans. This was particularly the case when the performer was following the *fly* or when he cleared the canvas. In these cases, Chimney was considered closer to a musical instrument.

He also commented that he would have preferred a higher complexity in the music generated by Chimney. In those situations, the trumpeter compensated for this perceived deficiency by increasing the rhythm complexity or the loudness of his instrument. To conclude, the inherent limitation of Chimney of having a limited material and being non-responsive fostered the trumpeter to find novel musical strategies.

6. CONCLUSIONS

Chimney was mainly developed as a design probe to reflect on the consequences of delegating part of the compositional process to an algorithmic agent. In this context, a number of negotiations between the human and the computer agent have to take place. The implication of such negotiations are evident from a musical perspective and from the performer's experience alike. With respect to the musical perspective, the focus of compositional effort shifts to the interaction strategies more than the form, and a number of compositional decisions are left to performance time. As a consequence, the distinction between composition and improvisation, which is neat in traditional performances, becomes blurred.

The disruption of built-in hierarchies of human-leads-computer-follows also caused a series of implications in the experience of the performer who is controlling Chimney and of other musicians that play along to it. Whereas some of these implications can be attributed to the non-responsive nature of the system, others are attributable to the distinctive character of the interaction. In particular, both Chimney performer and the musician that improvises along to it, are withhold the possibility to control all the aspects of the performance. Freed by a precise control on musical structure and declined the possibilities to have *veto buttons*, they need to surrender a more balanced collaboration with the algorithmic counterpart.

7. ACKNOWLEDGEMENTS

We are grateful to Riccardo Terrin for having performed Alinearity, and to the organisers of the COOP and ISIM conference.

8. REFERENCES

- [1] D. Charles. *For the Birds*. Boston: Marion Boyars, 1981.
- [2] S. M. Feisst. John Cage and Improvisation: An Unresolved Relationship. *Musical Improvisation: Art, Education, and Society*: 38-51. 2009.
- [3] G. E. Lewis. "Interacting with Latter-Day Musical Automata". *Contemporary Music Review* 18(3): 99–112, 1999

- [4] R. Masu, A. Conci, Z. Menestrina, F. Morreale, A. De Angeli. "Beatfield: An Open-Meaning Audio-visual Exploration", 2016
- [5] F. Morreale. "Generative Everything 2083". In *Proceedings of the XVIII Generative Art Conference*, 2015
- [6] K. Perlin. "An image synthesizer." *ACM Siggraph Computer Graphics* 19, no. 3: 287-296, 1985
- [7] Sarath, Ed. "A new look at improvisation." *Journal of Music Theory* (1996): 1-38.
- [8] O. S. Vallis, Contemporary approaches to live computer music: the evolution of the performer composer. Victoria University of Wellington, 2013.
- [9] O. S. Vallis, A. Kapur. "Community-based design: The democratization of musical interface construction." *Leonardo music journal* 21, 2011: 29-34, 2011.

SONIFICAZIONE DI OSTACOLI COME AUSILIO ALLA DEAMBULAZIONE DI NON VEDENTI

Simone Spagnol
University of Iceland
Reykjavík, Islanda
spagnols@hi.is

Charalampos Saitis, Kyriaki Kalimeri
ISI Foundation
Torino, Italia
kyriaki.kalimeri@isi.it

Ómar I. Jóhannesson, Rúnar Unnþórsson
University of Iceland
Reykjavík, Islanda
runson@hi.is

SOMMARIO

In questo articolo proponiamo uno schema di sonificazione che codifica informazione visiva nel dominio uditorio tramite suoni che simulano l'impatto tra il bastone e gli ostacoli che un non vedente incontra durante la deambulazione. Lo schema è valutato da sette utenti non vedenti in uno scenario virtuale semplificato. La valutazione non si limita all'accuratezza delle risposte secondo misure psicofisiche, includendo analisi del coinvolgimento e dello stress sperimentato dai partecipanti tramite segnali fisiologici. I risultati psicofisici mostrano che lo schema di sonificazione è in grado di comunicare informazioni significative riguardo la scena virtuale, anche se soltanto nei casi nei quali la procedura di training sia risultata sufficiente. I risultati psicofisiologici suggeriscono la necessità di modifiche allo schema di sonificazione che rendano lo stesso più coinvolgente.

1. INTRODUZIONE

La tecnica della sonificazione è spesso usata in applicazioni audio per rappresentare azioni, oggetti o situazioni che descrivono una scena virtuale. Recenti applicazioni della sonificazione rientrano nell'ambito dell'assistenza sanitaria, ad esempio sistemi per la riabilitazione motoria [1], ausili elettronici per la navigazione (*Electronic Travel Aids*, ETA - ovvero dispositivi che supportano la mobilità indipendente attraverso il riconoscimento di ostacoli e/o guide all'orientamento e alla deambulazione sicura) [2] e altre tecnologie assistive per non vedenti e ipovedenti. Ciononostante, la maggior parte di tali sistemi è tuttora in uno stadio di sviluppo prematuro, e i prodotti attualmente in commercio hanno funzionalità limitate, un modesto valore scientifico e costi generalmente non accessibili [2].

Gli ETA attualmente disponibili sono in grado di comunicare al non vedente informazioni che spaziano dal semplice riconoscimento di ostacoli tramite un sensore di distanza a complesse rappresentazioni visive della scena in esame acquisite tramite camera digitale. Gli output uditi di tali sistemi spaziano a loro volta da rozzi allarmi binari che indicano la presenza (o assenza) di un ostacolo

Copyright: ©2016 Simone Spagnol 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.

nelle vicinanze a pattern audio contenenti un carico di informazione paragonabile a quello di un'immagine [3, 4]. Il problema principale è in questo caso riuscire a comunicare tramite suoni informazioni sull'ambiente che circonda l'utente che siano quanto più significative e utilizzabili possibile. Trovare il miglior *trade-off* tra accuratezza e semplicità è quindi un problema centrale nel design della tecnica di sonificazione.

Le più comuni (e, di pari passo, naturali) relazioni tra le proprietà dell'oggetto e quelle del suono prodotto sono legate alla posizione dell'oggetto nello spazio circostante, ad esempio

- azimut → panning stereo / filtraggio tramite *Head-Related Transfer Function* (HRTF) [5, 6, 7];
- elevazione → HRTF [6] / altezza [8];
- distanza → ampiezza [5, 9] / altezza [5, 9].

Il presente studio si propone di esplorare uno schema per la traduzione di rappresentazioni 3D di una scena o un ambiente visivo - sotto forma di liste di oggetti con proprietà - in un linguaggio uditorio. La scelta delle relazioni e dei parametri utilizzati segue l'analisi preliminare basata su metodi di valutazione psicofisica riportata in uno studio precedente [10]. Lo schema utilizzato è introdotto nella Sezione 2, mentre la Sezione 3 descrive un esperimento finalizzato all'analisi della performance dello schema stesso in un task virtuale, effettuata tramite misure sia psicofisiche che psicofisiologiche. I risultati dell'esperimento sono riportati nella Sezione 4, e discussi nella Sezione 5.

2. SONIFICAZIONE TRAMITE SUONI D'IMPATTO

Lo schema qui proposto tratta ogni oggetto presente nell'emisfero frontale dell'utente alla stregua di una sorgente sonora virtuale indipendente che emette ciclicamente suoni d'impatto, come se l'utente non vedente lo stesse percuotendo con il bastone. L'altezza di tale suono d'impatto viene considerata dipendente dalla larghezza dell'oggetto in esame, mentre il timbro dipende dalla tipologia di oggetto. La distanza tra oggetto e utente è codificata nell'ampiezza e nella frequenza di ripetizione del suono d'impatto, entrambe crescenti all'avvicinarsi dell'oggetto. Infine, ogni sorgente sonora virtuale viene spazializzata a seconda della direzione dell'oggetto rispetto all'utente.

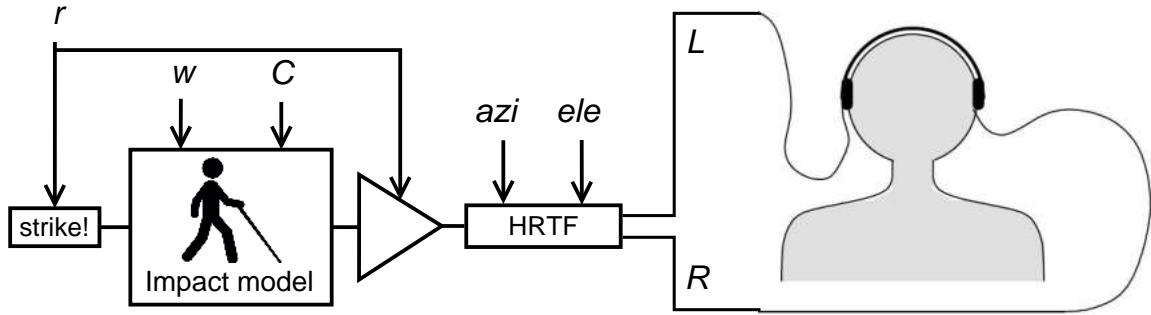


Figura 1. Lo schema di sonificazione proposto. I parametri in ingresso sono azimut (*azi*), elevazione (*ele*), distanza (*r*), larghezza (*w*) e categoria (*C*) dell’oggetto.

I singoli suoni d’impatto sono generati tramite un modello fisico di impatto nonlineare tra due oggetti modali. Tale modello è parte della collezione inclusa nel *Sound Design Toolkit* (SDT),¹ un pacchetto software *open-source* (GPLv2) realizzato come ausilio alla ricerca e all’educazione nell’area del *Sonic Interaction Design* [11]. L’SDT consiste in una libreria di algoritmi di sintesi del suono basati su o ispirati da fenomeni fisici, disponibili sotto forma di *external* e *patch* per Max e Pure Data (nel presente lavoro abbiamo usato quest’ultima versione).

Il modello fisico riceve parametri relativi all’eccitatore (oggetto modale 1), al risonatore (oggetto modale 2) e all’interazione tra i due. Alcuni dei parametri sono considerati costanti; tra i più significativi figurano la velocità d’impatto, fissata a 1.85 m/s, e la massa dell’eccitatore, fissata a 0.6 kg (valori considerati ragionevoli nel caso di un bastone per non vedenti prolungato e l’atto di colpire l’oggetto con esso). I parametri del risonatore, ovvero l’oggetto da rappresentare tramite il suono, variano invece rispetto alla larghezza e alla tipologia dell’oggetto.

In particolare, la larghezza dell’oggetto è mappata direttamente nella frequenza f del singolo modo di risonanza del risonatore. Allo scopo di massimizzare l’intervallo di frequenze considerate, è stato scelto di far variare tale parametro tra 50 Hz (per oggetti molto larghi, ad esempio un muro) e 4 kHz (per oggetti larghi 20 cm) attraverso la seguente relazione,

$$f = \frac{840}{w} \text{ [Hz]} \quad (1)$$

dove w è la larghezza reale dell’oggetto (in metri).

Diverse categorie di oggetti sono invece rappresentate da diversi tempi di decadimento del modo di risonanza. Varie scelte riguardo la categorizzazione di oggetti possono essere effettuate. Una possibilità è dividere gli oggetti per materiale (dove ad esempio gomma, legno, vetro e metallo sono ordinati per tempo di decadimento crescente [12]), oppure per tipologia (oggetti generici, muri, pali o alberi, buche o specchi d’acqua, e così via). Definite le categorie $C = 1, 2, 3, \dots$, la relazione con il corrispondente parametro acustico, in questo caso il tempo di decadimento

t_d , è

$$t_d = 0.02C \text{ [s]} \quad (2)$$

definita euristicamente allo scopo di permettere l’associazione con impatti su diversi materiali [12]. Nel test riportato in questo articolo non è stata tuttavia effettuata alcuna categorizzazione, con la categoria 5 assegnata a tutti gli oggetti (generici).

La distanza assoluta r tra soggetto e oggetto è anch’essa utilizzata come parametro. Assumendo che tutti gli oggetti da sonificare siano posizionati a distanza maggiore di 1 m dall’ascoltatore (oggetti più vicini sono direttamente raggiungibili con il bastone), e quindi nel campo acustico lontano [13], la distanza stessa è mappata direttamente nell’ampiezza del suono attraverso la classica legge di attenuazione della pressione $1/r$ [14]. Inoltre, r influenza la frequenza di ripetizione del suono d’impatto: il periodo T tra due impatti consecutivi è calcolato euristicamente come

$$T = 0.1(r + n - 1) \text{ [s]}, \quad (3)$$

dove n è il numero totale di oggetti considerati nella scena. Occorre precisare che il punto associato all’oggetto è definito come il baricentro nel caso di oggetti relativamente piccoli, oppure l’intersezione tra la superficie più vicina dell’oggetto e la propria normale che attraversa l’ascoltatore nel centro dell’asse interaurale nel caso di oggetti più voluminosi, ad esempio un muro.

Infine, la direzione dell’oggetto rispetto all’ascoltatore, definita in coordinate angolari (azimut, elevazione), viene mappata direttamente nei parametri corrispondenti di un paio di HRTF generiche fornite dall’external Pure Data *earplug~* per la sintesi binaurale. In particolare, tali filtri rendono la posizione angolare della sorgente sonora relativa all’ascoltatore attraverso una semplice convoluzione del segnale in ingresso con le HRTF sinistra e destra del database MIT KEMAR [15].² In questo modo, il suono viene teoricamente spazializzato lungo la reale direzione dell’oggetto. Precisiamo che in questo caso la spazializzazione non è individuale; tuttavia, l’integrazione di modelli per la personalizzazione di HRTF [16, 17] oppure di HRTF individuali misurate acusticamente è facilmente attuabile nel caso in cui si voglia una resa spaziale più accurata [18].

¹ <http://soundobject.org/SDT/>

² <http://sound.media.mit.edu/resources/KEMAR.html>

La sonificazione proposta è schematizzata in Figura 1. L'utilizzo di suoni d'impatto per comunicare informazioni riguardo gli ostacoli è motivato da diverse ragioni. Innanzitutto la validità ecologica dei suoni generati tramite modelli fisici, la cui natura permette di associare in maniera diretta il suono all'atto virtuale di incontrare un oggetto percuotendolo con il bastone. In secondo luogo la peculiarità dei suoni d'impatto, il cui ricco contenuto in frequenza e la breve durata della fase di attacco sono ingredienti ottimali per un'accurata localizzazione del suono sul piano orizzontale [19]. Inoltre, le scelte riguardanti la mappatura tra le proprietà dell'oggetto e quelle del suono sono state ispirate dalla precedente letteratura (distanza e direzione) o basate sulla natura del modello d'impatto. Infatti, l'associazione di altezze maggiori a oggetti più piccoli e di diversi tempi di decadimento a diverse categorie, ad esempio materiali, ha basi fisiche [12].

Lo schema di sonificazione è stato implementato in Pure Data. La versione corrente del software supporta sia scene virtuali statiche che semplici scene dinamiche con un numero fissato di oggetti. La *patch* principale riceve come input un file di testo che descrive gli oggetti presenti nella scena, uno per riga. Ogni riga contiene le seguenti informazioni relative all'oggetto, separate da spazi:

- ID dell'oggetto (stringa);
- azimut (incluso tra -90 and 90 gradi [15]);
- elevazione (inclusa tra -40 e 90 gradi [15]);
- distanza (1 m e oltre);
- larghezza (20 cm e oltre);
- categoria (1, 2, 3, ...);
- modalità (statica = 0, dinamica = 1).

Inizialmente le sorgenti virtuali vengono ordinate per azimut crescente, da sinistra a destra. Al fine di evitare impatti simultanei durante il primo ciclo, il primo impatto relativo a un determinato oggetto viene presentato 100 ms dopo l'impatto relativo all'oggetto alla sua immediata sinistra. Successivamente, la generazione ripetuta di suoni d'impatto per un determinato oggetto è indipendente dagli altri oggetti nella scena. Qualora la scena fosse dinamica, gli impatti relativi a un oggetto cessano non appena questo termina nell'emisfero posteriore dell'ascoltatore (ovvero al di fuori dell'intervallo di azimut $[-90^\circ, 90^\circ]$).

3. METODOLOGIA Sperimentale

L'esperimento riportato in questa sezione è stato originariamente concepito come procedura di test preliminare per la valutazione di un setup sperimentale di acquisizione dati multimodale con partecipanti non vedenti e ipovedenti. I metodi utilizzati sono di tipo psicofisico e psicofisiologico, precisamente tempo di risposta e accuratezza, elettroencefalografia (EEG), e monitoraggio dell'attività elettrodermica (EDA). L'esperimento consiste nella navigazione di

un percorso virtuale sonificato tramite lo schema proposto nella precedente Sezione 2, e la conseguente valutazione della funzionalità (psicofisica) e della performance cognitiva (psicofisiologia) dello schema stesso. Il presente studio è stato accettato dal Comitato Bioetico Nazionale d'Islanda, con numero di protocollo VSN-15-107.

3.1 Partecipanti

Sette partecipanti (di cui 4 donne, età media 38 anni, compresa tra 22 e 50 anni) hanno preso parte all'esperimento su base volontaria. Quattro di loro erano affetti da cecità totale o cecità funzionale, due erano ciechi legali (residuo visivo binoculare non superiore a 1/20), e uno era ipovedente grave (residuo visivo binoculare tra 1/20 e 1/10). Cinque di loro erano ciechi dalla nascita o entro i primi 2 – 3 anni di vita, mentre due lo erano diventati dopo i 3 anni di età. Tutti i partecipanti parlavano fluentemente in lingua inglese e hanno riportato di non avere alcun impedimento uditivo, né problemi di salute generici. Tutti i partecipanti hanno sottoscritto il loro libero consenso informato.

3.2 L'approccio psicofisiologico

L'attività elettrodermica è un noto indicatore di eccitazione fisiologica e attivazione da stress [20]. Contrariamente agli agenti stressanti fisici, i quali sono particolarmente rintracciabili in misure di attività cardiovascolare quali la frequenza cardiaca, l'EDA è particolarmente sensibile alle variazioni di eccitazione relative alle emozioni. Dall'altro lato, l'elettroencefalografia è in grado di fornire marker neurofisiologici relativi a processi cognitivi ed emotivi indotti dallo stress e indicati da variazioni nell'attività ritmica cerebrale [21]. Utilizzando un approccio multimodale che combina le proprietà inerenti e complementari di EEG ed EDA per studiare l'esperienza emotiva e cognitiva di soggetti non vedenti durante la navigazione di ambienti non familiari [22], abbiamo acquisito e analizzato i relativi segnali congiuntamente alle più tradizionali misure psicofisiche quali tempo di risposta e accuratezza.

Una misurazione EDA è caratterizzata da due diversi tipi di risposte: *risposte fasiche*, di breve durata (visualizzabili come rapidi picchi nel segnale) e *livelli tonici*, di lunga durata (l'andamento del segnale in assenza di attività fasica) [20]. Le risposte fasiche sono principalmente provocate da stimolazioni esterne, e vengono tipicamente osservate sovrapposte negli stati di elevata eccitazione.

EDA è stata acquisita con il braccialetto Empatica E4 [23], il quale misura la conduttanza cutanea al polso tramite due elettrodi interni ($f_s = 4$ Hz). I segnali acquisiti sono stati analizzati con Ledalab, una *toolbox* per Matlab.³ Ledalab implementa un metodo di decomposizione del segnale (Continuous Decomposition Analysis - CDA) basato sulla deconvoluzione dello stesso, risultante in una misura continua la quale integrazione temporale risulta a sua volta in un indicatore obiettivo dell'attività fasica, denominato *integrated skin conductance response* (ISCR) [24]. La nostra ipotesi è che un feedback uditivo piacevole, facile

³ <http://www.ledalab.de>

da comprendere e poco stressante provochi generalmente un'attività fasica minore secondo l'indicatore ISCR.

L'attività cerebrale è caratterizzata da pattern ritmici osservabili in diverse bande di frequenza, la cui definizione può variare da studio a studio. In questo lavoro abbiamo analizzato EEG in cinque bande, precisamente theta (4–7 Hz), alpha-1 (7.5–10 Hz), alpha-2 (10–12.5 Hz), beta (13–30 Hz) e gamma (30–60 Hz). L'EEG è stato registrato attraverso il dispositivo wireless Emotiv EPOC+, caratterizzato da 14 elettrodi passivi (canali) che rilevano l'attività nelle posizioni AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 e FC4 (frequenza di campionamento $f_s = 128$ Hz) [25].

Per l'analisi dei segnali abbiamo utilizzato la Performance Metrics Detection Suite (PMDS) di Emotiv. PMDS utilizza algoritmi proprietari che analizzano automaticamente i segnali EEG registrati ed estraggono cinque variabili associate a stati emotivi, in particolare agitazione istantanea, agitazione a lungo termine, frustrazione, coinvolgimento e meditazione.⁴ Abbiamo deciso per questo esperimento di analizzare le prime quattro metriche. Recenti studi hanno dimostrato l'affidabilità di PMDS nell'osservare stati emotivi [26, 27].

3.3 Apparato sperimentale

All'inizio dell'esperimento ogni partecipante viene invitato ad utilizzare la propria mano dominante per rispondere, premendo i tasti direzionali di una tastiera: freccia a sinistra per girare a sinistra, freccia in alto per proseguire diritto e freccia a destra per girare a destra. La freccia in basso è stata rimossa preventivamente dalla tastiera per evitare confusione. Il feedback uditorio viene presentato attraverso un paio di cuffie in-ear.⁵ File audio pre-registrati sono inoltre utilizzati per comunicare al partecipante ulteriori istruzioni e feedback riguardo il progresso dell'esperimento.

EDA viene registrata al polso della mano non dominante del partecipante al fine di minimizzare artefatti dovuti alla pressione delle dita sui tasti [20]. EEG viene acquisito in modo continuo dai 14 elettrodi del dispositivo EPOC. I segnali vengono trasmessi attraverso un ricevitore USB wireless a un computer che esegue il software proprietario. I dati PMDS sono ottenuti in tempo reale dalla API di Emotiv via OSC (Open Sound Control), in particolare tramite l'applicazione *MindYourOSCs*,⁶ e memorizzati attraverso uno *script* Processing adattato da un precedente lavoro [27].

3.4 Stimoli

Utilizzando lo schema di sonificazione proposto in questo articolo, sono stati creati stimoli sonori rappresentanti due oggetti che appaiono uno dopo l'altro (a una distanza di 3 m e 5 m dal piano frontale dell'ascoltatore, rispettivamente), il primo alla sinistra o destra dell'ascoltatore e il

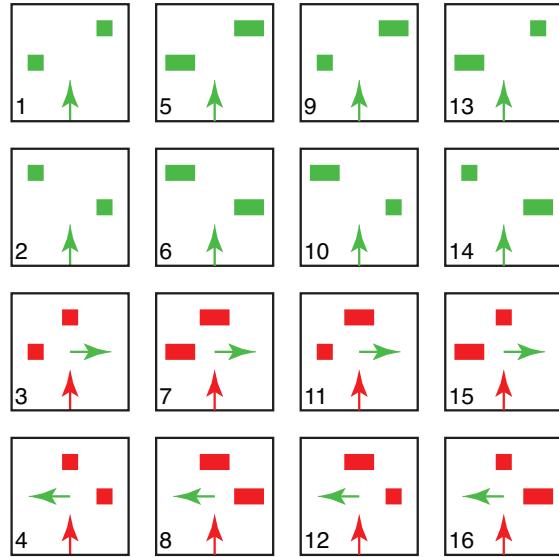


Figura 2. Disposizione degli oggetti nelle scene virtuali usate per l'esperimento. Le frecce verdi in ogni scena indicano la direzione relativa all'uscita corretta, mentre le frecce rosse indicano le situazioni in cui il partecipante deve girare a destra o sinistra.

secondo o dal lato opposto rispetto al primo oppure frontalmente. Ognuno dei due oggetti è un cubo largo 50 cm oppure un parallelepipedo largo 100 cm.

I due oggetti appaiono in quattro diverse configurazioni spaziali, rappresentate nella Figura 2: primo oggetto a sinistra e secondo a destra; viceversa; primo oggetto a sinistra e secondo di fronte; primo oggetto a destra e secondo di fronte. Conseguentemente, $2 \times 2 \times 4 = 16$ scene sono state sonificate per questo esperimento. Durante la presentazione del feedback uditorio, il partecipante deve quindi decidere se girare a sinistra (risposta corretta nelle scene 4, 8, 12, 16), a destra (risposta corretta nelle scene 3, 7, 11, 15), o proseguire diritto (risposta corretta nelle scene 1, 2, 5, 6, 9, 10, 13, 14).

3.5 Protocollo sperimentale

Inizialmente ognuna delle 16 scene virtuali viene riprodotta, e la corretta risposta associata spiegata dallo sperimentatore. Successivamente il partecipante indossa i sensori fisiologici, prestando particolare cura nel trovare una posizione confortevole evitando così movimenti non necessari. Il partecipante viene quindi invitato a rilassarsi per 300 s, allo scopo di registrare l'attività fisiologica durante lo stato di riposo spontaneo. Subito dopo le 16 scene vengono riprodotte una seconda volta nello stesso ordine utilizzato precedentemente, ma questa volta il partecipante risponde e riceve feedback riguardo la correttezza o meno delle sue risposte. Il partecipante viene contestualmente invitato a proseguire diritto nel suo cammino virtuale quando possibile, e a rispondere più velocemente e accuratamente

⁴ Per una descrizione esaustiva delle variabili stesse si rimanda alla guida di Emotiv: <https://emotiv.zendesk.com/hc/en-us/>

⁵ <https://earhero.com>

⁶ <http://sourceforge.net/projects/mindyourosc/>

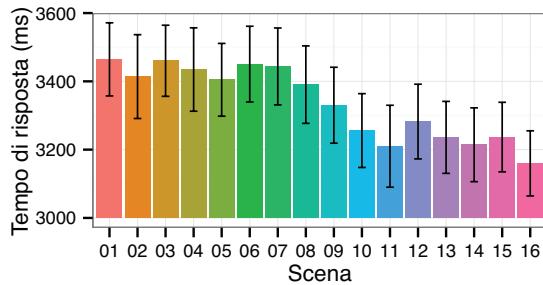


Figura 3. Tempo medio di risposta per scena. Le barre di errore rappresentano l'errore standard della media.

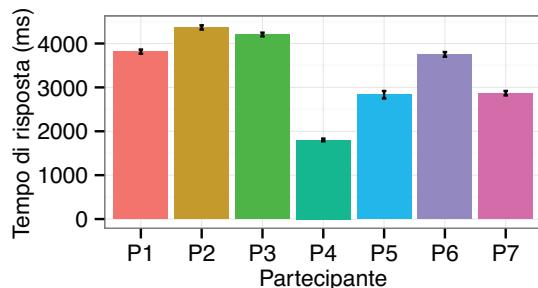


Figura 4. Tempo medio di risposta per partecipante. Le barre di errore rappresentano l'errore standard della media.

possibile. Al termine di tale fase di training introduttiva, la fase sperimentale ha inizio.

Ogni *trial* sperimentale (o percorso) include 4 delle scene rappresentate in Figura 2, presentate in rapida successione. L'ordine delle scene è scelto pseudocasualmente, fermo restando che tutte le scene vengano presentate lo stesso numero di volte, ossia 20, per un totale di 320 risposte. La velocità del percorso virtuale è stata fissata a 1.0 m/s. Per ogni scena il partecipante deve prendere una decisione entro 5 m (5 s): svoltare a destra o sinistra, oppure continuare diritto. In caso di mancata risposta, questa viene considerata errata. All'inizio di ogni *trial* viene riprodotto un breve tono alto a indicare l'inizio di un nuovo percorso. All'interno di un percorso, la scena successiva viene presentata o subito dopo la risposta del partecipante o non appena il suono relativo alla scena precedente (di durata pari a 5.3 s) termina. Durante la fase sperimentale il partecipante non riceve alcun feedback riguardo l'accuratezza delle proprie risposte. Lo sperimentatore siede all'esterno della stanza, tenendo tuttavia in costante controllo l'andamento dell'esperimento e dell'acquisizione dei dati fisiologici.

Alla fine della fase sperimentale il partecipante viene invitato a rilassarsi per 300 s, allo scopo di registrare una seconda volta l'attività fisiologica durante lo stato di riposo spontaneo. Successivamente, i sensori EPOC ed E4 vengono disinstallati. Infine, al partecipante vengono richiesti liberi commenti sull'esperimento e sul feedback uditorio. La durata totale dell'esperimento è di circa un'ora.

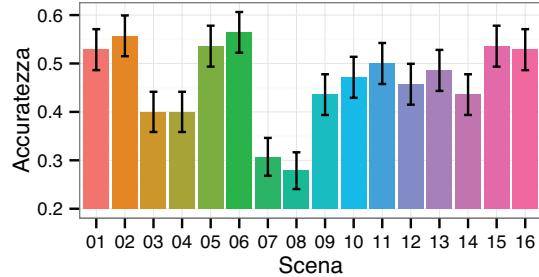


Figura 5. Accuratezza media per scena. Le barre di errore rappresentano l'errore standard della media.

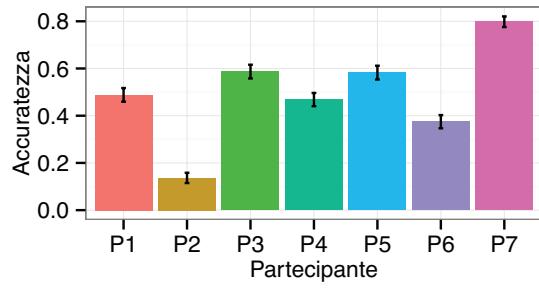


Figura 6. Accuratezza media per partecipante. Le barre di errore rappresentano l'errore standard della media.

4. RISULTATI

4.1 Tempo di risposta e accuratezza

Dati statistici riguardo il tempo di risposta e l'accuratezza sono stati calcolati per ogni scena e paragonati usando un'ANOVA per misure ripetute. Dopo aver rimosso le mancate risposte (169 occorrenze) e i tempi di risposta che deviavano più di 3 SD dalla media di ogni partecipante (20 occorrenze), il tempo medio di risposta risulta essere di 3.3 s ($SD = 1.2$ s). La Figura 3 mostra media ed errore standard del tempo di risposta diviso per scena. Non è stato rilevato alcun effetto significativo della scena sul tempo di risposta ($F(15, 90) = 1.31, p = 0.216$). Tuttavia, dalla Figura 3 emerge che le scene con due oggetti di dimensioni diverse (9 – 16) hanno registrato tempi di risposta leggermente inferiori. Al contrario, come riportato in Figura 4, sono state rilevate differenze considerevoli tra i partecipanti, i cui tempi medi di risposta variano tra 1.8 s (P4) e 4.4 s (P2).

L'accuratezza media è pari a 0.46 ($SD = 0.5$) e non è purtroppo diversa da una performance completamente casuale ($paired-t(6) = 1.76, p = 0.126$), la cui accuratezza media è pari a 0.33. Media ed errore standard dell'accuratezza divisi per scena sono riportati in Figura 5. Non è stato rilevato alcun effetto significativo della scena sull'accuratezza ($F(15, 90) = 1.21, p = 0.279$). Tuttavia, dalla Figura 5 pare emergere una relazione tra la dimensione degli oggetti in ogni scena e le risposte dei partecipanti. In particolare, si nota una tendenza a non svoltare correttamente nelle scene che includono due oggetti di dimensio-

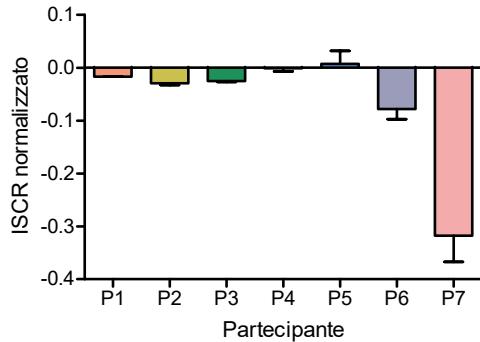


Figura 7. ISCR normalizzato medio per partecipante. Le barre di errore rappresentano l'errore standard della media.

ni uguali (scene 3, 4, 7, 8), e una più leggera tendenza a non proseguire correttamente diritto nelle scene che includono due oggetti di dimensioni diverse (scene 9, 10, 13, 14). L'influenza della dimensione degli oggetti sulla corretta direzione da intraprendere, seppur statisticamente non significativa (effetto delle diverse dimensioni degli oggetti sulla scelta di proseguire diritto $p = 0.197$ in un'ANOVA a misure ripetute con dimensioni e direzioni come fattori), indica la presenza di fattori che hanno impedito un apprendimento efficace del modello, realisticamente legati ad un'insufficiente fase di training.

Al contrario, come riportato in Figura 6, anche in questo caso sono state rilevate differenze considerevoli tra i partecipanti, la cui accuratezza media varia tra 0.14 (P2, peggiore di una performance casuale) e 0.8 (P7). Il valore medio di accuratezza per il partecipante P7 indica che qualora la fase di training sia risultata esauriente ed efficace per il partecipante, lo schema di sonificazione è in grado di fornire indicazioni utili e precise riguardo la navigazione del percorso virtuale.

4.2 Risposta elettrodermica

I dati ottenuti dal sensore EDA E4 sono stati innanzitutto filtrati con un filtro passa-basso di Butterworth del primo ordine, con frequenza di taglio di 0.6 Hz. Tale operazione si è resa necessaria al fine di rimuovere alcuni artefatti (picchi) dovuti alla pressione sugli elettrodi [20]. La sequenza di dati filtrata è stata quindi analizzata con Ledalab come specificato in Sezione 3.2. Poiché gli intervalli tra gli stimoli variavano a seconda del tempo di risposta, per le singole risposte abbiamo considerato una finestra temporale variabile che dura da 1 s dopo l'attacco dello stimolo a 4 s dopo la terminazione dello stesso. L'operazione di integrazione temporale risultante nella misura ISCR è stata quindi effettuata su tale finestra e normalizzata sulla durata della stessa. Inoltre, al fine di ridurre le variazioni interindividuali, abbiamo sottratto dai vari campioni il valore medio individuale di ISCR durante gli stati di riposo spontaneo.

L'ISCR normalizzato diviso per partecipante è riportato in Figura 7. È innanzitutto interessante notare come l'ISCR decresca dalla fase di riposo a quella sperimentale

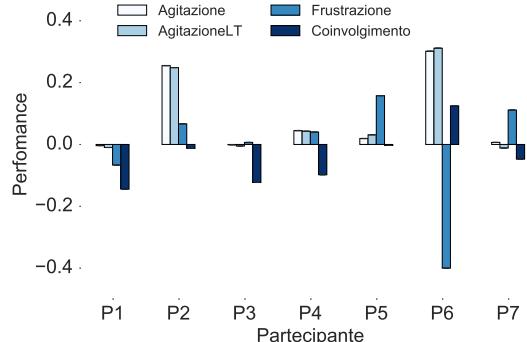


Figura 8. Metriche di performance PMDS medie per partecipante.

per la maggior parte dei partecipanti (6 su 7). Viene inoltre rilevata una moderata correlazione negativa con i valori medi di accuratezza in Figura 6 ($r = -0.57$); in particolare il partecipante P7, la cui accuratezza risulta migliore, registra un calo medio di ISCR rispetto allo stato di riposo di gran lunga superiore agli altri partecipanti. I commenti post-esperimento confermano che P7 ha trovato il task particolarmente piacevole e divertente. Al contrario, il secondo miglior partecipante in termini di accuratezza, P5, registra un incremento medio di ISCR rispetto allo stato di riposo. In questo caso, i commenti post-esperimento rivelano che P5 ha trovato il task molto stressante.

4.3 Metriche PMDS

Le metriche della PMDS di Emotiv descritte in Sezione 3.2 sono disponibili con la precisione di 1 ms. Al fine di rimuovere le differenze individuali da ognuno dei quattro segnali, abbiamo rimosso la media individuale di ogni partecipante prendendo come riferimento le registrazioni EEG durante gli stati di riposo spontaneo. Quindi, per ogni metrica e partecipante, e al fine di non perdere la continuità dell'informazione nel segnale EEG, abbiamo mediato i valori lungo l'intera sessione sperimentale.

Occorre precisare che, contrariamente all'analisi dei dati psicofisici, sia per quanto riguarda l'analisi EEG che per la precedente EDA abbiamo scelto di considerare tutti i trial sperimentali. Tale decisione è dovuta al fatto che i processi cognitivi che portano a una risposta corretta, sbagliata o mancata contribuiscono globalmente al profilo della performance cognitiva del modello. Ciò non toglie la possibilità di esaminare risposte corrette, errori e risposte mancate separatamente in analisi future.

I valori medi delle quattro metriche, divisi per partecipante, sono riportati in Figura 8. Dalla figura si nota un'evidente correlazione negativa tra le metriche di agitazione (a breve e lungo termine) e l'accuratezza media in Figura 6 (entrambe $r = -0.76$): in particolare, le performance peggiori (P2 e P6) registrano un valore di agitazione considerevolmente aumentato rispetto allo stato di riposo. Per quanto riguarda la metrica di frustrazione, non è stata trovata nessun particolare correlazione con le misure psicofisiche.

siche, né con i commenti post-esperimento. Ad esempio, i partecipanti con valore di frustrazione più alto (P5) e più basso (P6) hanno entrambi riportato a fine esperimento di aver sofferto l'eccessiva durata e la monotonia dell'esperimento. È tuttavia interessante notare come, secondo la metrika di coinvolgimento, l'esperimento sia risultato poco coinvolgente per la grande maggioranza dei partecipanti (6 valori negativi su 7), e ciò emerge anche dalla maggioranza dei commenti post-esperimento.

5. CONCLUSIONI E SVILUPPI

L'obiettivo principale dello studio riportato in questo articolo era il design di uno schema di sonificazione di ostacoli efficace e di una metodologia di valutazione sperimentale dello stesso che andasse oltre le classiche misure psicofisiche. I risultati psicofisici, seppur non conclusivi a causa di una procedura di training insufficiente, hanno dimostrato che lo schema di sonificazione è in grado di comunicare informazioni utili e discriminabili durante la navigazione uditiva di un semplice percorso virtuale. I risultati psicofisiologici hanno invece evidenziato le diverse reazioni da parte dei partecipanti, la maggioranza delle quali testimoniano come il feedback uditivo utilizzato, seppur gradevole, vada migliorato dal punto di vista del coinvolgimento.

Tali risultati si sono rivelati fondamentali per il design della successiva versione dello schema di sonificazione, descritta in [28]. In questa versione la principale modifica riguarda la frequenza di ripetizione del suono d'impatto, la quale non dipende più dalla distanza del singolo oggetto bensì dal solo numero di oggetti presenti nella scena. Evitando la sovrapposizione di impatti provenienti da diversi oggetti, il pattern ritmico dei suoni risulta più regolare e meno caotico.

L'esperimento riportato in questo articolo ha inoltre fornito preziose indicazioni riguardo il design del nuovo esperimento descritto in [28], dove lo schema di sonificazione qui presentato è stato paragonato a due ulteriori schemi precedentemente utilizzati in letteratura. Particolare attenzione è stata rivolta a un'efficace fase di training e ad un'analisi più trasparente dei dati EEG. I risultati del nuovo esperimento hanno evidenziato l'efficacia e il modesto carico cognitivo del feedback uditivo.

6. RINGRAZIAMENTI

Gli autori desiderano ringraziare la Dott.ssa Vigdís Vala Valgeirsdóttir per il suo aiuto nell'esecuzione dell'esperimento e il Prof. Árni Kristjánsson per la gentile concessione degli spazi sperimentali. Sentiti ringraziamenti ai partecipanti per la preziosa collaborazione, e al National Institute for the Blind, Visually Impaired, and Deafblind in Islanda per la generosa assistenza. Questo progetto ha ricevuto il finanziamento del programma di ricerca e innovazione dell'Unione Europea Horizon 2020 in virtù del contratto di sovvenzione n.643636.⁷

⁷ This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636. <http://www.soundofvision.net/>

7. BIBLIOGRAFIA

- [1] F. Avanzini, S. Spagnol, A. Rodá, and A. De Götzen, "Designing interactive sound for motor rehabilitation tasks," in *Sonic Interaction Design* (K. Franinovic and S. Serafin, eds.), ch. 12, pp. 273–283, Cambridge, MA, USA: MIT Press, March 2013.
- [2] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," *IEEE Trans. Syst. Man Cybern.*, vol. 40, pp. 25–35, January 2010.
- [3] A. Csapó, G. Wersényi, H. Nagy, and T. Stockman, "A survey of assistive technologies and applications for blind users on mobile platforms: A review and foundation for research," *J. Multimod. User Interf.*, vol. 9, pp. 275–286, December 2015.
- [4] A. Kristjánsson, A. Moldoveanu, O. I. Jóhannesson, O. Balan, S. Spagnol, V. V. Valgeirsdóttir, and R. Unnthórsson, "Designing sensory-substitution devices: Principles, pitfalls and potential," *Restor. Neurol. Neurosci.*, 2016.
- [5] S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the Navbelt - A computerized travel aid for the blind," *IEEE Trans. Syst. Man Cybern.*, vol. 28, pp. 459–467, August 1998.
- [6] J. L. González-Mora, A. Rodríguez-Hernández, L. F. Rodríguez-Ramos, L. Díaz-Saco, and N. Sosa, "Development of a new space perception system for blind people, based on the creation of a virtual acoustic space," in *Engineering Applications of Bio-Inspired Artificial Neural Networks*, vol. 1607 of *Lecture Notes in Computer Science*, pp. 321–330, Springer Berlin Heidelberg, 1999.
- [7] F. Fontana, A. Fusiello, M. Gobbi, V. Murino, D. Rocchesso, L. Sartor, and A. Panuccio, "A cross-modal electronic travel aid device," in *Human Computer Interaction with Mobile Devices*, vol. 2411 of *Lecture Notes in Computer Science*, pp. 393–397, Springer Berlin Heidelberg, 2002.
- [8] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 112–121, February 1992.
- [9] E. Milios, B. Kapralos, A. Kopinska, and S. Stergiopoulos, "Sonification of range information for 3-D space perception," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, pp. 416–421, December 2003.
- [10] M. Bujacz, K. Kropidłowski, G. Ivanica, A. Moldoveanu, C. Saitis, A. Csapó, G. Wersényi, S. Spagnol, O. I. Jóhannesson, R. Unnthórsson, M. Rotnicki, and P. Wittek, "Sound of Vision - Spatial audio output and sonification approaches," in *Computers Helping People with Special Needs - 15th International Conference (IC-CHP 2016)* (K. Miesenberger, C. Bühlert, and P. Penaz, eds.), vol. 9759 of *Lecture Notes in Computer Science*,

- pp. 202–209, Linz, Austria: Springer Int. Publishing, July 2016.
- [11] S. Delle Monache, P. Polotti, and D. Rocchesso, “A toolkit for explorations in sonic interaction design,” in *Proc. 5th Audio Mostly Conference (AM '10)*, no. 1, (New York, NY, USA), ACM, September 2010.
 - [12] F. Avanzini and D. Rocchesso, “Controlling material properties in physical models of sounding objects,” in *Proc. Int. Computer Music Conf. (ICMC'01)*, (La Habana, Cuba), September 2001.
 - [13] S. Spagnol, “On distance dependence of pinna spectral patterns in head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 137, pp. EL58–EL64, January 2015.
 - [14] D. H. Ashmead, D. LeRoy, and R. D. Odom, “Perception of the relative distances of nearby sound sources,” *Percept. Psychophys.*, vol. 47, pp. 326–331, April 1990.
 - [15] W. G. Gardner and K. D. Martin, “HRTF measurements of a KEMAR,” *J. Acoust. Soc. Am.*, vol. 97, pp. 3907–3908, June 1995.
 - [16] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini, “Synthetic individual binaural audio delivery by pinna image processing,” *Int. J. Pervasive Comput. Comm.*, vol. 10, pp. 239–254, July 2014.
 - [17] S. Spagnol and F. Avanzini, “Anthropometric tuning of a spherical head model for binaural virtual acoustics based on interaural level differences,” in *Proc. 21st Int. Conf. Auditory Display (ICAD 2015)*, (Graz, Austria), pp. 204–209, July 2015.
 - [18] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hamershøi, “Binaural technique: Do we need individual recordings?”, *J. Audio Eng. Soc.*, vol. 44, pp. 451–469, June 1996.
 - [19] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 2nd ed., October 1996.
 - [20] W. Boucsein, *Electrodermal Activity*. New York, NY, USA: Springer, 2nd ed., 2012.
 - [21] W. Klimesch, “EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis,” *Brain Res. Rev.*, vol. 29, pp. 169–195, April 1999.
 - [22] C. Saitis and K. Kalimeri, “Identifying urban mobility challenges for the visually impaired with mobile monitoring of multimodal biosignals,” in *Universal Access in Human-Computer Interaction - 10th International Conference* (M. Antona and C. Stephanidis, eds.), no. 9739 in Lecture Notes in Computer Science, pp. 616–627, Berlin, Germany: Springer-Verlag, June 2016.
 - [23] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, “Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition,” in *EAI 4th Int. Conf. Wirel. Mob. Commun. Healthcare (Mobihealth)*, (Athens, Greece), pp. 39–42, November 2014.
 - [24] M. Benedek and C. Kaernbach, “A continuous measure of phasic electrodermal activity,” *J. Neurosci. Methods*, vol. 190, pp. 80–91, June 2010.
 - [25] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, “Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs,” *PeerJ*, vol. 19, 2013.
 - [26] P. Aspinall, P. Mavros, R. Coyne, and J. Roe, “The urban brain: Analysing outdoor physical activity with mobile EEG,” *Br. J. Sports Med.*, vol. 49, pp. 272–276, February 2015.
 - [27] D. Kalogianni and T. Wolfgang, “The city soundscape and the brain,” in *Proc. Invisible Places*, (Viseu, Portugal), pp. 5–33, July 2014.
 - [28] S. Spagnol, C. Saitis, M. Bujacz, O. I. Jóhannesson, K. Kalimeri, A. Moldoveanu, A. Kristjánsson, and R. Unnþórsson, “Model-based obstacle sonification for the navigation of visually impaired persons,” in *Proc. 19th Int. Conf. Digital Audio Effects (DAFx-16)*, (Brno, Czech Republic), September 2016.

PWGL come strumento per l'analisi musicale assistita: alcuni esempi

Fabio De Sanctis De Benedictis

ISSM "P. Mascagni" - Livorno

fdesanctis@teletu.it

SOMMARIO

È possibile classificare la composizione algoritmica secondo un approccio costruttivo o dichiarativo. L'analisi musicale assistita dal computer può essere distinta in modo analogo, a seconda che l'elaboratore elettronico sia principalmente utilizzato per verificare elementi musicali e fornire/costruire tabelle o altro materiale di consultazione utile all'indagine, oppure conduca esso stesso un'analisi musicale completa. Nel corso della storia troviamo esempi dell'uno e dell'altro approccio, ma solo in casi relativamente sporadici abbiamo il computer ad effettuare per intero l'indagine, o l'implementazione dei metodi tradizionali di analisi musicale, così come praticati dai musicisti. Di frequente poi i risultati analitici non sono di immediata leggibilità secondo le procedure classiche di analisi della musica. In questo contesto PWGL può svolgere un ruolo efficace, principalmente sfruttando le sue caratteristiche di programmabilità e le sue possibilità di contrassegnare in maniera opportuna la partitura. Dopo una sintetica e necessariamente incompleta rassegna su esperienze pregresse, si propongono alcune implementazioni in PWGL concernenti l'analisi statistica, insiemistica, motivica e armonico-funzionale. In conclusione qualche considerazione sull'opportunità e il vantaggio di usare PWGL in un contesto analitico didattico e professionale.

1. INTRODUZIONE

Riprendendo alcune nozioni esposte in un precedente articolo,¹ è opportuno ricordare come la composizione musicale algoritmica possa essere classificata secondo due approcci: costruttivo e dichiarativo.² Nel primo caso l'elaboratore elettronico viene utilizzato principalmente per lo sviluppo del materiale compositivo, nel secondo è il computer stesso a generare un brano intero o porzioni di esso.³ In analogia si può definire l'analisi musicale assistita: abbiamo un approccio costruttivo quando il calcolatore elet-

¹ Vedi [1].

² Vedi [2], p. 5

³ A questi approcci corrispondono le classificazioni in CAC e CGM (*Computer Aided Composition* e *Computer Generated Music*) di Puckette e in CGS, CGA, CGC, rispettivamente *Computer Generated Sounds*, *Computer Generated Assistance* e *Computer Generated Composition* di Cope. Si vedano rispettivamente [2] pp. ix-xiv e [3], pp. ix-xv.

Copyright: ©2016 Fabio De Sanctis De Benedictis 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.

tronico viene principalmente utilizzato per verificare elementi musicali, o per fornire/costruire tabelle o altro materiale di consultazione utile all'indagine;⁴ dichiarativo quando è il computer stesso a condurre l'analisi. Come si vedrà nella sezione seguente, senza pretesa di esaustività, nel corso degli anni si sono avuti esempi dell'uno e dell'altro tipo di approccio. Tuttavia sono relativamente pochi i casi in cui è la macchina ad effettuare per intero l'analisi e/o in cui si implementano algoritmi concernenti le metodologie storiche di indagine analitica. Una difficoltà ulteriore consiste nella visualizzazione dei risultati prodotti dal *software*, che non sempre sono di immediata comprensione e utilizzo per il musicista, almeno secondo i precetti dell'analisi musicale comunemente praticata. In questo contesto PWGL (acronimo per *PatchWork Graphical Language*) può svolgere un ruolo efficace, principalmente sfruttando le sue caratteristiche di programmabilità, ben documentate in letteratura e nelle funzioni di aiuto integrate nel *software*, congiuntamente alla possibilità di contrassegnare in maniera opportuna la partitura.⁵ Dopo una sintetica rassegna su alcuni esempi di quanto effettuato in passato, si propongono alcune *patch* in PWGL relative a metodologie analitiche quali l'analisi statistica, insiemistica, motivica, armonica, per passare poi alle conclusioni. Non ci occupiamo in questa sede di analisi di musiche elettroacustiche né di analisi dell'interpretazione, limitandoci quindi all'analisi musicale centrata sulla partitura, sul livello neutro.

2. CENNI SULL'ANALISI MUSICALE ASSISTITA

In primo luogo occorre specificare che un'impostazione analitica è comunque presente in quegli studi che si occupano della ricostruzione e realizzazione algoritmica dei processi compositivi di opere del repertorio.⁶ Il procedimento analitico viene rovesciato in direzione compositiva, conseguentemente a quanto asserito da Bent sulla posizione intermedia dell'analisi musicale, tra la teoria della composizione e l'estetica (vedi [12], p. 2). Schüller ci offre un'utile classificazione (si veda [13]), distinguendo

⁴ È questo ad esempio il caso di alcune pagine web, che offrono strumenti di analisi principalmente nell'ambito della dodecafonia e della *Pitch-class Set Theory*. Possiamo ricordare: <http://composertools.com/> di Paul Nelson; i programmi di Dmitri Tymoczko, scaricabili dalla seguente pagina web: <http://dmitri.tymoczko.com/software.html>; il materiale di Reginald Bain, alla pagina <http://in.music.sc.edu/fs/bain/software.html>; il *software* Jdubiel: <http://music.columbia.edu/~akira/JDubiel/>; la pagina di Nico Schüller: <http://www.musicanalysis.org/> (ultima visita per ogni pagina web citata in questo articolo in data 4/8/2016).

⁵ A questo proposito il lettore interessato potrà consultare con profitto [4], [5], [6], [7] e [8].

⁶ Si vedano [9], [10], [11], rispettivamente rivolti a composizioni di Hiller e Isaacson, Boulez, Birtwistle.

do vari tipi di metodologie analitiche assistite dal computer: statistiche e collegate alla teoria dell'informazione; insiemistiche (*Pitch-class Set Theory*); matematiche; basate su metodologie gerarchiche, ovvero implicanti procedure riduzionistiche; trasformazionali; Schenkeriane; spettrali; cognitive e legate all'intelligenza artificiale; combinate. L'analisi statistica e/o collegata alla teoria dell'informazione risulta tra le metodologie più adatte ad un'implementazione informatica, tanto è vero che già Bent ne riporta alcuni esempi ([12], pp. 250-274). Per l'analisi insiemistica, oltre a quanto indicato in nota 4, è possibile fare riferimento alle funzioni e classi specifiche presenti sia in OpenMusic sia in PWGL. Analisi musicali basate su altre metodologie matematiche sono assai diffuse: possiamo ricordare le ricerche condotte da Guerino Mazzola e quelle del *Séminaire MaMux* dell'Ircam, (<http://repmus.ircam.fr/mamux/home>), la nascita relativamente recente del *Journal of Mathematics and Music* (<http://www.tandfonline.com/toc/tmam20/current>), nonché rimandare agli atti delle conferenze *Mathematics and Computation in Music* (per informazioni ulteriori: <http://www.smcm-net.info/>), pubblicati biennalmente per le edizioni *Springer* a partire dal 2007; frequentemente questi studi originano o sono basati su medesimi presupposti di *software* che ne implementano gli strumenti, come ad esempio Rubato ed hexachord (<http://www.rubato.org/> e <http://www.lacl.fr/lbigo/hexachord>). Per i metodi analitici che implicano processi di riduzione troviamo rappresentate principalmente l'analisi Schenkeriana⁷ e la teoria generativa della musica tonale di Lerdhal e Jackendoff (si vedano [16] e [17]). Tra altri metodi di analisi musicale legati al linguaggio compare l'analisi paradigmatica (vedi [18]). L'analisi tematica è un'altra metodologia tradizionale implementata piuttosto spesso al calcolatore (si rimanda a [19], [20] e [21]), non sempre con adeguata visualizzazione dei risultati. Per quanto riguarda la segmentazione, primo approccio al testo musicale e tra le fasi più delicate dell'analisi di un brano, possiamo ricordare come esempio il saggio di Cambouropoulos (si veda [22]), che presenta un modello formale, poi implementato, per la segmentazione di melodie, e quello di Bresson e Pérez-Sancho (vedi [23]), che descrive strumenti di OpenMusic per la segmentazione e l'analisi della partitura. In taluni casi si cerca un'integrazione tra *software* differenti, come nel sistema OHR di Garbers (descritto in [24]), in cui Open-Music, Humdrum e Rubato sono utilizzati in parallelo. Un cenno a parte merita infine SOAL, una libreria per Open-Music specificamente indirizzata all'analisi musicale secondo particolari principi e metodi, di cui troviamo un esempio applicativo nell'articolo di Guigue, Fagner Onofre e Rolim (si veda [25]). I casi citati, pur rappresentando una minima parte di quanto presente in questo campo, offrono stimoli e suggerimenti validissimi e molto interessanti, ma nel contempo evidenziano alcuni problemi: non esiste uno

⁷ Vedi [14] e [15]. In particolare Marsden propone il *software* Schenker, scritto in Java, virtualmente eseguibile su qualsiasi piattaforma, che permette riduzioni schenkeriane di frasi musicali, usando un'interfaccia che si avvicina maggiormente alle consuetudini d'uso del musicista, con una visualizzazione in forma di Piano-roll. Il programma può essere scaricato al seguente indirizzo: <https://www.lancaster.ac.uk/people/marsdena/software/schenker/index.htm>.

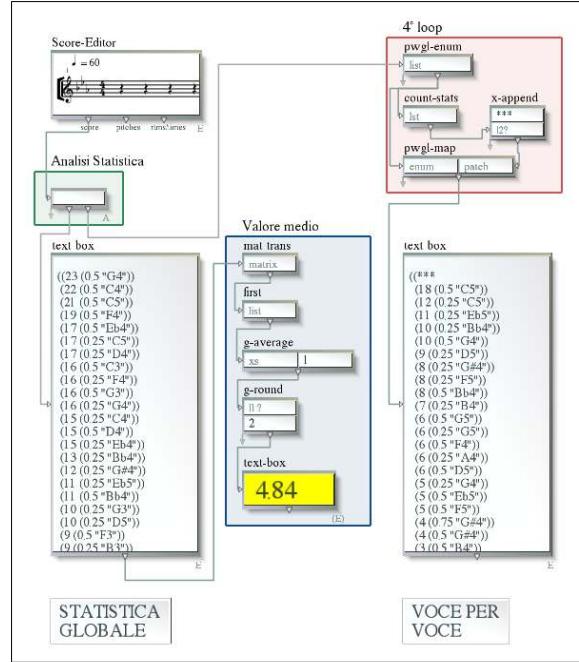


Figura 1. Patch relativa all'analisi statistica di note e durate.

standard condiviso per la codifica delle partiture da analizzare, per cui si va dal formato MIDI al MusicXML, sino a diversi codici di tipo testuale, a seconda del *software* utilizzato; di conseguenza non abbiamo ancora una banca dati di partiture musicali, adeguatamente formattate, di pubblico dominio, che comprenda la totalità del repertorio, pronte per l'analisi assistita; in alcuni casi il computer procede autonomamente ed in altri è l'analista ad effettuare la segmentazione e le principali scelte analitiche; l'output dei programmi non sempre è in un formato comodamente leggibile o nella notazione musicale tradizionale. Certamente PWGL non risolve tutti questi aspetti, né si ha la pretesa di avere la capacità e la possibilità di poterlo fare autonomamente. Tuttavia si ritiene e si auspica che gli esempi che seguono, frutto di un lavoro e una ricerca in continua evoluzione, possano essere di stimolo e utilità.

3. IMPLEMENTAZIONE DI STRUMENTI ANALITICI IN PWGL

Il primo esempio che si propone è riportato in Fig. 1 e concerne una semplice analisi statistica. In Fig. 2 la *abstraction* (termine che in PWGL designa le *subpatch*) denominata "Analisi Statistica", contenente il motore del calcolo. Nello Score-Editor di Fig. 1 è importato un file MIDI relativo alla Fuga in do minore del secondo volume del *Clavicembalo ben temperato* di Bach, introdotto poi nella *abstraction* per il computo. In Fig. 2 l'oggetto collect-enp-objects seleziona le singole voci, che sono immesse, una alla volta, in un primo *loop*, di livello superiore ed esterno rispetto ai seguenti due, che estrae i valori delle altezze e delle durate (1 = semimi-

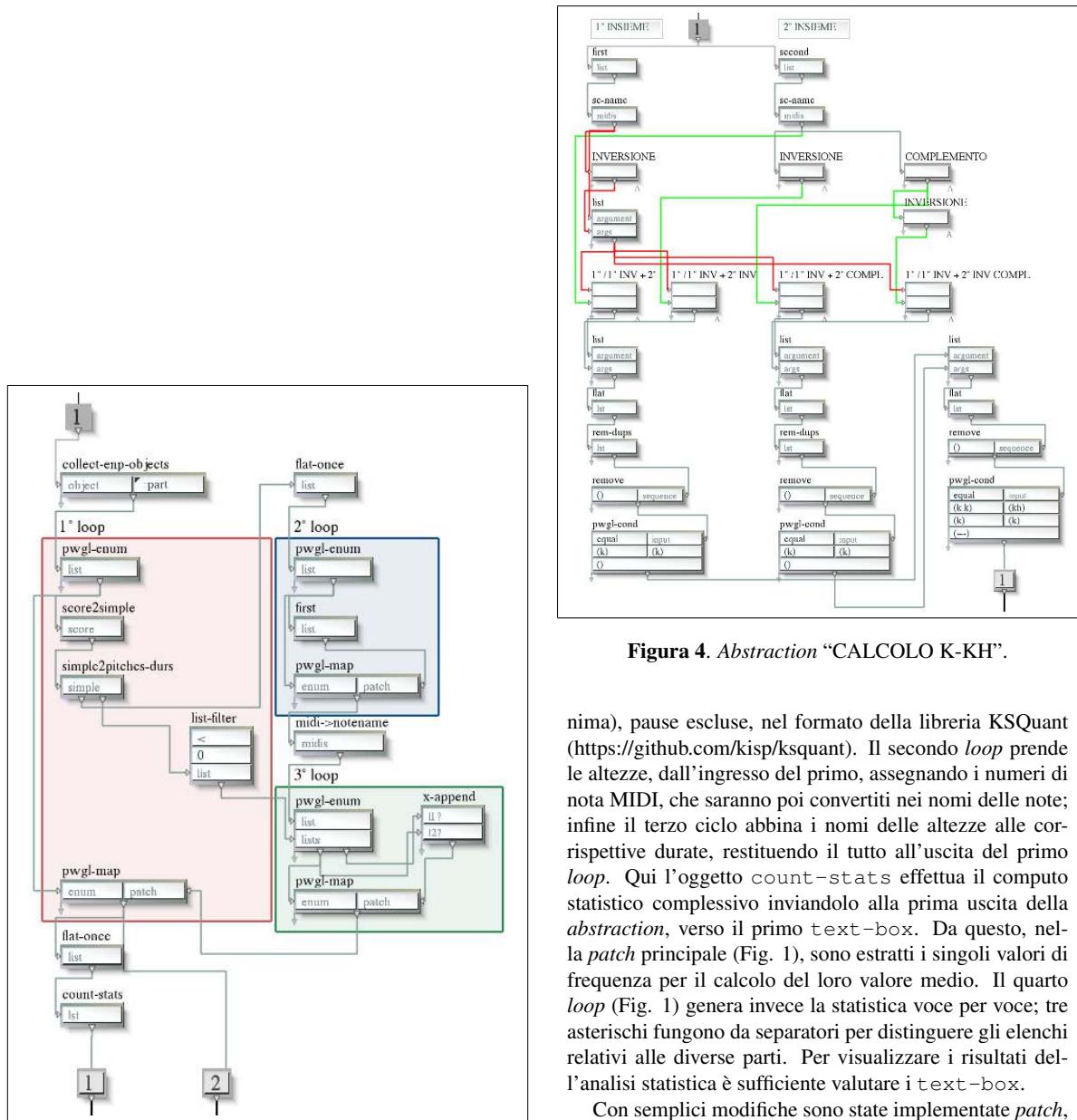


Figura 2. Abstraction “Analisi Statistica”.

Figura 4. Abstraction “CALCOLO K-KH”.

nima), pause escluse, nel formato della libreria KSQuant (<https://github.com/kisp/ksquant>). Il secondo *loop* prende le altezze, dall’ingresso del primo, assegnando i numeri di nota MIDI, che saranno poi convertiti nei nomi delle note; infine il terzo ciclo abbina i nomi delle altezze alle corrispettive durate, restituendo il tutto all’uscita del primo *loop*. Qui l’oggetto *count-stats* effettua il computo statistico complessivo inviandolo alla prima uscita della *abstraction*, verso il primo *text-box*. Da questo, nella *patch* principale (Fig. 1), sono estratti i singoli valori di frequenza per il calcolo del loro valore medio. Il quarto *loop* (Fig. 1) genera invece la statistica voce per voce; tre asterischi fungono da separatori per distinguere gli elenchi relativi alle diverse parti. Per visualizzare i risultati dell’analisi statistica è sufficiente valutare i *text-box*.

Con semplici modifiche sono state implementate *patch*, non riportate in questa sede, che effettuano l’analisi statistica delle sole altezze, delle sole durate, degli intervalli melodici, in quest’ultimo caso contando anche il numero complessivo di intervalli ascendenti e discendenti, con la possibilità ulteriore di visualizzare direttamente la frequenza di specifiche note, durate, o intervalli.

L’esempio seguente (Fig. 3) riguarda la *Pitch-class Set Theory* e lo illustreremo sulla base della segmentazione e analisi del sesto dei *Six kleine Klavierstücke* op. 19 di Schönberg, così come riportata nel testo di Nicholas Cook (Vedi [26], pp. 158–174).⁸ Per questa *patch* ci siamo

⁸ La *patch* è stata verificata anche con alcuni schemi di [27], precisamente quelli delle figure 109, 111 e 113, rispettivamente alle pagine 115, 118 e 122. Si sono trovate due sole incongruenze, nell’ultimo schema, in cui Allen Forte tra gli insiemi 5-19 e 4-16 non pone alcuna relazione, mentre PWGL restituisce la relazione K, e nella coppia di insiemi successiva, 5-19 e 4-19, Forte individua la relazione K, PWGL nessuna. Una verifica manuale e attraverso il *software* di Paul Nelson, citato alla nota

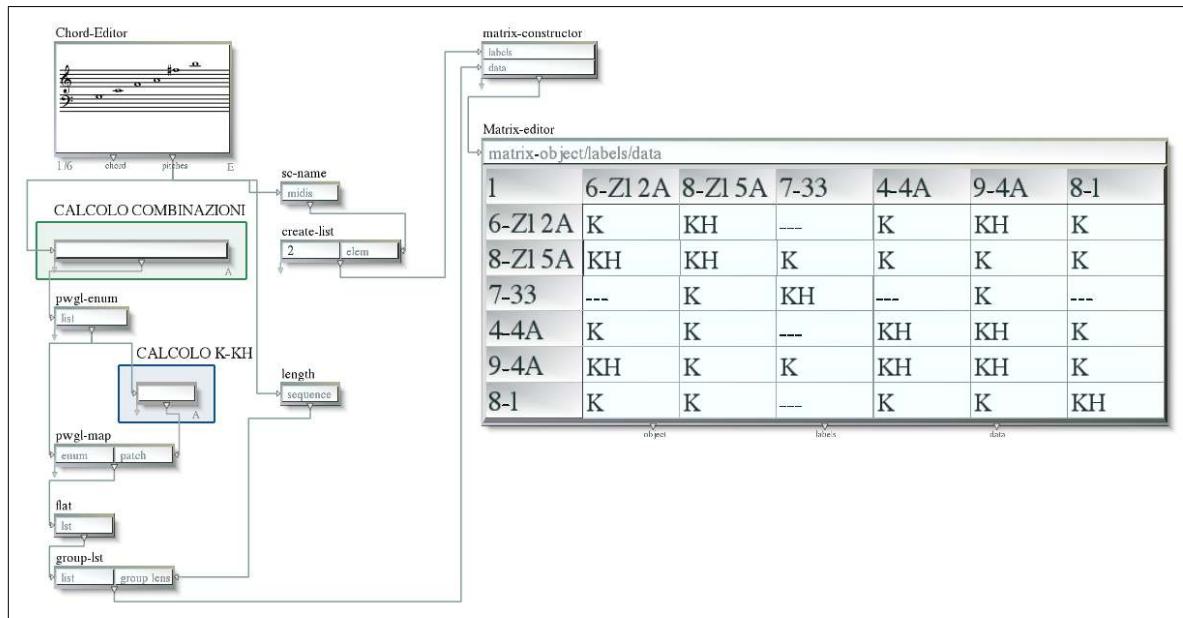


Figura 3. Patch per l'individuazione dell'insieme cardine.

liberamente ispirati, come idea, ad una pagina web con un *software* java di Marcus Uneson, purtroppo al momento in cui scriviamo non più disponibile. L'algoritmo implementato permette di compilare una matrice K-Kh per l'individuazione dell'insieme cardine.⁹

Nel Chord-Editor vanno immessi gli insiemi di classi di altezze segmentati dal brano.¹⁰ L'uso di una tastiera MIDI collegata all'elaboratore elettronico può permettere di velocizzare l'immissione di dati. La *abstraction* denominata "CALCOLO COMBINAZIONI" abbina ogni insieme a se stesso e a tutti gli altri creando una serie di sottoliste composte ciascuna di due elementi, corrispondenti alle note MIDI dei due insiemi di volta in volta presi in esame. Queste coppie sono immesse in un *loop* in cui la *abstraction* denominata "CALCOLO K-KH" (vedi Fig. 4) effettua la valutazione della relazione K o Kh afferente ai due insiemi di classi di altezze, confrontando le forme *recta* e inversa di un insieme di classi di altezze con le forme *recta* e inversa, e i loro complementari, dell'altro in-

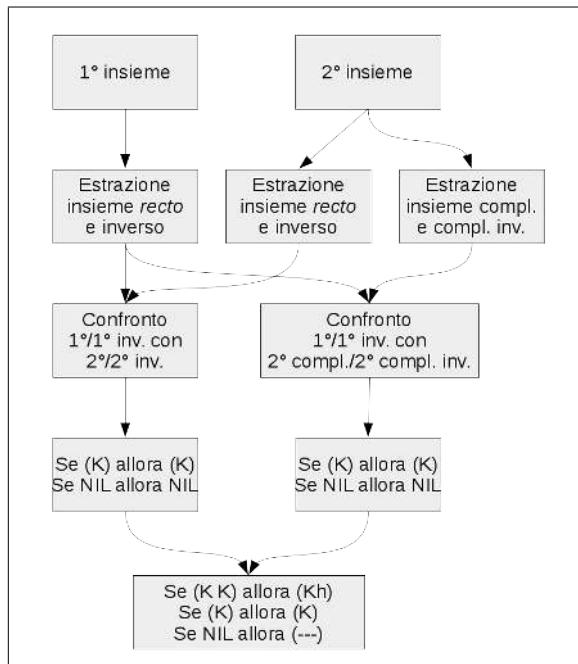


Figura 5. Diagramma relativo all'individuazione delle relazioni K-Kh (*abstraction* di Fig. 4).

⁴, darebbe ragione a PWGL. Si suppone quindi che nel testo di Forte ci sia un refuso, con una inversione di relazioni contigue sulla stessa riga. È opportuno precisare anche che nell'uso di PWGL i nomi degli insiemi di classi di altezze hanno una lettera A o B alla fine, diversamente da altre consuetudini, per indicare l'insieme *recto* o quello inverso. In una variante della *patch* illustrata, che utilizza i nomi degli insiemi piuttosto che il Chord-Editor, occorre indicare dette lettere (A o B indifferentemente, poiché le relazioni K e Kh sono simmetriche), quando presenti nel nome dell'insieme, poiché PWGL non restituisce errore, ma le relazioni in matrice appaiono false.

⁹ Le relazioni K e Kh sono relazioni di inclusione. L'insieme cardine è quello collegato a tutti gli altri per mezzo di questi rapporti. Siamo consapevoli del fatto che la matrice non rispecchia la tradizionale forma a "elle" degli schemi analoghi nelle analisi insiemistiche, configurazione particolare dovuta al fatto che ordinariamente non è necessario raffrontare insiemi di classi di altezze dello stesso numero di note. Sarà l'utente a interpretare debitamente i risultati escludendo le parti di matrice non necessarie.

¹⁰ Come detto in nota 8, in un'altra versione di questa *patch* gli insiemi di classi di altezze sono introdotti non in forma di note, bensì coi loro nomi, potendo così figurare in ordine, dai tricordi agli esordi.

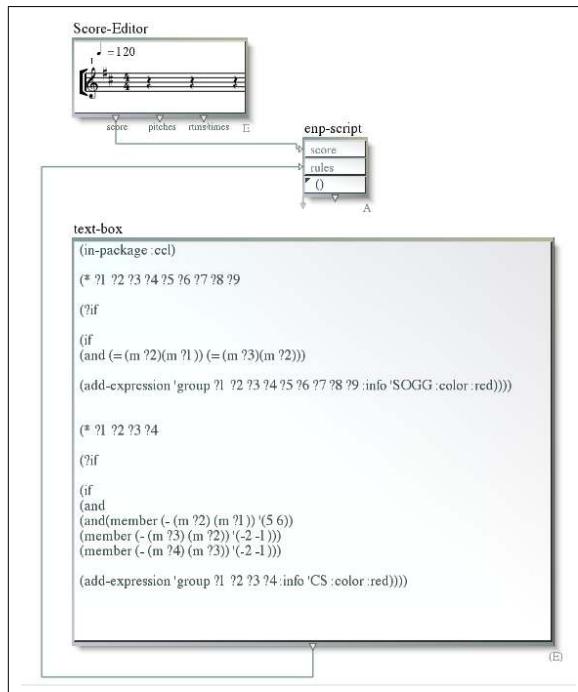


Figura 6. Analisi motivica di una Fuga bachiana.

sieme. Una serie di test condizionali, a seconda degli esiti restituiti in ogni calcolo, seleziona infine il simbolo corrispondente alla relazione risultante, ovvero K o Kh. Il tutto viene poi visualizzato nel Matrix-editor. In figura 5 un diagramma della logica sottesa al funzionamento della abstraction di Fig. 4. Come è possibile verificare in Fig. 3 l'insieme cardine risulta essere 9-4, come asserito nel testo di Cook. Sono state sviluppate altre patch nell'ambito dell'analisi insiemistica, per la compilazione di complessi Kh (Vedi [27] pp. 200–208), e il calcolo di relazioni K e Kh di un insieme con tutti gli altri di una specifica cardinalità.

Con l'esempio successivo (Fig. 6) entriamo in una delle caratteristiche più utili di PWGL: la possibilità di notare direttamente in partitura i risultati dell'analisi.¹¹ Non è questa la sede per descrivere in dettaglio il linguaggio di programmazione utilizzato in PWGL, per il quale rimandiamo all'ampia letteratura disponibile; ci limiteremo a descrivere il codice utilizzato.

Nello Score-Editor viene importato un file MIDI relativo alla Fuga numero 5 dal secondo volume del *Clavicembalo ben temperato* di Bach. Il text-box contiene il codice usato, mentre valutando l'enp-script si ottiene l'analisi e l'apposizione dei relativi contrassegni in

¹¹ Non è possibile, nei limiti di questo contributo, attuare un confronto dettagliato tra l'implementazione di strumenti analitici in PWGL rispetto a quanto disponibile in altri software. Mediamente possiamo però affermare che l'interfaccia più moderna e la notazione musicale più raffinata rendono già in questo PWGL più amichevole rispetto ad alternative come OpenMusic e Rubato, ad esempio, che conservano indubbiamente innumerevoli punti di forza. La versatilità di PWGL, con numerose librerie disponibili, funzioni e classi, la possibilità di utilizzare file MIDI (caratteristica presente anche negli altri programmi citati), congiuntamente ad un linguaggio di scripting non eccessivamente complicato, rendono questo software un buon candidato, per i musicisti in particolare, per l'assolvimento di compiti analitici generali e specifici.

partitura, ossia le etichette testuali “SOGG” per il Soggetto e la Risposta reale, e “CS” per il Controsoggetto, entrambe in colore rosso. In Figura 7 si riporta l'inizio della partitura come risulta dopo l'analisi. Occorre chiarire, come apparirà dalla discussione che segue, che il Soggetto e il Controsoggetto sono determinati e scelti dall'analista, che dovrà configurare conseguentemente i parametri del codice; l'algoritmo si occupa della loro ricerca, individuazione ed etichettatura lungo tutta la partitura.

In questo esempio si sono utilizzati due gruppi di istruzioni differenti per la ricerca e la marcatura del Soggetto/Risposta e del Controsoggetto. Si illustrano in ordine le istruzioni principali del codice contenuto nel text-box di Fig. 6. Nel caso del Soggetto, verificato che risulta composto di nove altezze e che inizia con tre note ripetute, l'istruzione (* ?1 ?2 ?3 ?4 ?5 ?6 ?7 ?8 ?9 :info 'SOGG :color :red)) imposta la ricerca su nove elementi consecutivi in qualsiasi punto di ogni voce,¹² in termini di note MIDI se non specificato altrimenti. Con l'istruzione (?if inizia il vero e proprio codice dell'algoritmo. (and (= (m ?2) (m ?1)) (= (m ?3) (m ?2))) stabilisce che debbano essere vere contemporaneamente due condizioni, ossia che le note MIDI 1 e 2 e quelle 2 e 3, in ogni stringa di 9 note consecutive prese in esame, abbiano lo stesso valore, il che equivale alle tre note ribattute tipiche della testa del Soggetto. Se la condizione è vera si applica l'istruzione successiva (add-expression 'group ?1 ?2 ?3 ?4 ?5 ?6 ?7 ?8 ?9 :info 'SOGG :color :red) che ordina al programma di apporre sotto le nove note MIDI consecutive l'etichetta “SOGG” in colore rosso. In maniera analoga si stabiliscono le condizioni per la ricerca del Controsoggetto, solo che adesso, oltre a variare il numero di note consecutive da analizzare, quattro nello specifico, devono essere valide contemporaneamente tre condizioni: (member (- (m ?2) (m ?1)) '(5 6)), (member (- (m ?3) (m ?2)) '(-2 -1)) e (member (- (m ?4) (m ?3)) '(-2 -1)), ossia che la differenza intervallare tra la seconda altezza e la prima sia di 5 oppure 6 semitonni ascendenti, per includere tutte le varianti possibili nella Fuga; poi che la differenza tra la terza e la seconda altezza sia di 2 o 1 semitono discendenti; infine stessa cosa tra la quarta e la terza nota. È opportuno specificare alcuni limiti da risolvere nel lavoro futuro: per ogni composizione musicale, che presenta motivi e temi specifici, occorre scrivere codice apposito, o modificare quello qui presentato, mentre sarebbe auspicabile una procedura più universale e un'interfaccia più comoda per immettere le caratteristiche intervallari dei temi/motivi da cercare. Ugualmente potrebbe essere utile unire istruzioni di segmentazione automatica, almeno per gli elementi iniziali, a funzioni di analisi motivica, in maniera da poter utilizzare il codice con qualsiasi tipo di configurazione tematica.¹³

¹² L'asterisco funge da carattere *jolly* esattamente come in altri programmi.

¹³ Come accennato in precedenza, la segmentazione è un processo preliminare indispensabile, tanto importante quanto delicato, implicando scelte interpretative, quantunque ridotte al minimo, alla base dell'analisi che seguirà. Un aiuto in questo senso potrebbe venire dalla formalizzazione algoritmica in PWGL di quanto espresso da Cifariello Ciardi (si vedano [28] e [29]). Si ritiene altresì che una segmentazione totalmente automatica non sia del tutto possibile, in particolare quando convenzioni

Figura 7. Risultato dell’analisi motivica automatica: prime battute della partitura.

L’ultimo esempio che presentiamo (Fig. 8) è relativo ad un’analisi armonica di un Corale bachiano, secondo i precetti della teoria funzionale dell’armonia.¹⁴ Il codice immesso permette di apporre i simboli relativi alle possibili funzioni rappresentate dagli accordi. Ove non risultante alcuna funzione il programma restituisce NIL.

La logica espressa dal codice indica che quando l’armonia è completa – ($(m ?1 :complete? t)$) – si ponga alla voce del basso – ($(m ?1 :parts 4 :object t)$) – il simbolo funzionale corrispondente all’accordo, selezionato sulla base dei test condizionali multipli ottenuti dall’istruzione cond, fondati sul risultato dell’istruzione sc+off.¹⁵ Al momento il codice non tiene conto delle modulazioni, caratteristica che dovrà essere implementata nel lavoro successivo. Nel caso di più configurazioni armoniche per movimento, l’ordine di successione degli accordi si legge dal basso verso l’alto. In Fig. 9 le prime battute analizzate automaticamente.

4. CONCLUSIONI

Le implementazioni analitiche realizzate in PWGL necessitano sicuramente di ulteriori perfezionamenti che ci met-

stilistiche, soprattutto nella musica del passato, infrangono raggruppamenti e cesure che dovrebbero essere invece realizzati sulla scorta degli articoli citati.

¹⁴ Per la notazione degli accordi secondo questa teoria ci siamo basati su [30].

¹⁵ Ovvero il nome dell’insieme e il numero di semitonni di differenza rispetto alla sua forma primaria. Ad esempio, dato che l’insieme $[3 - 11B]$ rappresenta la triade maggiore, e che in forma primaria la prima nota è un do, se lo scostamento in semitonni dalla forma primaria dà 7, questo implica che la prima nota dell’insieme trasposto sia un sol. Abbiamo quindi una triade maggiore con fondamentale sol, e possiamo apporre il simbolo T, corrispondente appunto alla funzione di Tonica nella tonalità del corale esaminato. È questa la prima delle condizioni successive all’istruzione cond.

```

Score-Editor
↓ - 76
[Musical Staff]
score patches rtmal/times E

enp-script
score rules () A

text-box
(in-package :ecl)
(* ?1 :harmony
(?if
(when (m ?1 :complete? t)
(add-expression 'group
(m ?1 :parts 4 :object t)
:info
(cond
((equal (sc+off(m ?1)) (list 'B-1 1 B'(7))) 'T)
((equal (sc+off(m ?1)) (list 'B-1 1 B'(0))) 'S)
((equal (sc+off(m ?1)) (list 'B-1 1 B'(2))) 'D)
((equal (sc+off(m ?1)) (list 'B-1 1 A'(9))) 'Sp)
((equal (sc+off(m ?1)) (list 'B-1 1 A'(11))) 'Tg/Sp)
((equal (sc+off(m ?1)) (list 'B-1 1 A'(4))) 'Tp/Sg)
((equal (sc+off(m ?1)) (list 'A-2 7 B'(6))) 'D7)
((equal (sc+off(m ?1)) (list 'B-1 0 '6)) 'D7_no_fond.)
((equal (sc+off(m ?1)) (list 'A-2 6 '4)) 'S5/6)
((equal (sc+off(m ?1)) (list 'B-8 A'(0))) 'D7_inc.)
).color:red
)))) (E)

```

Figura 8. Analisi armonico-funzionale.



Figura 9. Risultato dell'analisi armonico-funzionale.

tano in condizione di poterle facilmente e immediatamente utilizzare con qualsiasi brano musicale. Allo stesso tempo i problemi sollevati dalla letteratura presa in esame, attinente all'analisi musicale assistita dal computer, sono risolti solo in parte da quanto mostrato in questa sede. Ciò nonostante riteniamo che l'uso di PWGL permetta comunque di velocizzare e agevolare il lavoro analitico in più direzioni:

1. capacità di analizzare grandi quantità di opere in un lasso di tempo relativamente breve, una volta che queste siano codificate opportunamente in formato MIDI
2. possibilità di affiancare la didattica dell'analisi della musica, utilizzando il *software* sia come strumento di lavoro, sia come mezzo di verifica dei risultati
3. affrancandoci in parte dal lavoro meccanico di classificazione degli elementi musicali, PWGL mette l'analista nelle condizioni di concentrarsi maggiormente sulle relazioni tra le parti, quindi sulla sintesi e le conclusioni
4. il parallelismo metodologico con la codifica delle tecniche compositive del Novecento (Vedi [1]) permette una maggiore consapevolezza dei legami tra analisi e composizione.

5. RINGRAZIAMENTI

L'autore vuole ringraziare Mika Kuuskankare per gli utili suggerimenti sul codice Lisp, senza i quali non sarebbe stato possibile realizzare la *patch* di Fig. 8.

6. BIBLIOGRAFIA

- [1] F. De Sanctis De Benedictis, "Dall'analisi musicale alla composizione e formalizzazione algoritmica: esempi applicativi con PWGL," in *Proceedings of XX CIM*, 2014.
- [2] C. Agon, G. Assayag, and J. Bresson, *The OM Composer's Book. Volume One*. Editions Delatour France/Ircam, Parigi, 2006.
- [3] J. Bresson, C. Agon, and G. Assayag, *The OM Composer's Book. Volume Two*. Editions Delatour France/Ircam, Parigi, 2008.
- [4] M. Kuuskankare and M. Laurson, "Extending VIVO as a MIR System," in *Proceedings of the 15th International Conference on Auditory Display*, 2009.
- [5] M. Laurson and M. Kuuskankare, "Extensible Constraint Syntax through Score Accessors," in *Journées d'Informatique Musicale*, pp. 27–32, 2005.
- [6] M. Laurson, M. Kuuskankare, and K. Kuitunen, "Introduction to Computer-Assisted Music Analysis in PWGL," in *Proceedings of XV CIM*, 2005.
- [7] M. Kuuskankare and M. Laurson, "MIR in ENP-Rule-Based Music Information Retrieval from Symbolic Music Notation," in *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, 2009.
- [8] M. Laurson, M. Kuuskankare, and K. Kuitunen, "The visualization of computer-assisted music analysis information in pwgl," in *Journal of New Music Research*, pp. 61–76, 2008.

- [9] Örjan. Sandred, M. Laurson, and M. Kuuskankare, “Revisiting the Illiac Suite – a rule based approach to stochastic processes,” in *Sonic Ideas/Ideas Sonicas*, 2, pp. 42–46, 2009.
- [10] Y.-K. Ahn, C. Agon, and M. Andreatta, ““Structures Ia pour deux pianos” by Boulez: towards creative analysis using OpenMusic and Rubato,” in *Mathematics and Computation in Music*, 2009.
- [11] R. V. Barbosa, “Análise, modelo e composição,” in *IV Seminario Música Ciência Tecnologia: Fronteiras e Rupturas*, 2000.
- [12] I. Bent, *Analisi musicale*. EDT, Torino, 1990.
- [13] N. Schüller, “Towards a General Methodological Classification of Computer-Assisted Music Analysis,” in *International Computer Music Conference Proceedings*, 2000.
- [14] P. B. Kirlin and P. E. Utgoff, “A Framework for Automated Schenkerian Analysis,” in *ISMIR 2008, Ninth International Conference on Music Information Retrieval – Session 3b – Computational Musicology*, 2008.
- [15] A. Marsden, “Software for Schenkerian Analysis,” in *Proceedings of the International Computer Music Conference 2011*, 2011.
- [16] L. Camilleri, F. Carreras, and F. Giomi, “Sistemi esperti in musicologia: un prototipo per l’analisi Time-Span Reduction,” in *Atti del X Colloquio di Informatica Musicale*, 1993.
- [17] P. Principi, “Un’indagine sulle regole preferenziali di raggruppamento di Lerdahl e Jackendoff,” in *Atti dell’XI Colloquio di Informatica Musicale*, 1995.
- [18] L. Camilleri, D. Bencini, and M. Ignelzi, “Analisi paradigmatica di repertori monofonici,” in *Atti del IX Colloquio di Informatica Musicale*, 1991.
- [19] P.-Y. Rolland, “FlExPat: A Novel Algorithm For Musical Pattern Discovery,” in *Atti dell’XII Colloquio di Informatica Musicale*, 1996.
- [20] C. Buteau, “Melodic Clustering Within Topological Spaces of Schumann’s Träumerei,” in *International Computer Music Conference, ICMC 2006*, 2006.
- [21] C. Buteau and J. Vipperman, “Melodic Clustering Within Motivic Spaces: Visualization in OpenMusic and Application to Schumann’s Träumerei,” in *Mathematics and Computation in Music*, 2009.
- [22] E. Cambouropoulos, “Musical Parallelism and Melodic Segmentation,” in *Atti dell’XII Colloquio di Informatica Musicale*, 1996.
- [23] J. Bresson and C. Pérez-Sancho, “New Framework for Score Segmentation and Analysis in OpenMusic,” in *Proceedings of SMC Conference 2012*, 2012.
- [24] J. Garbers, “An Integrated MIR Programming and Testing Environment,” in *ISMIR 2006, 7th International Conference on Music Information Retrieval*, 2006.
- [25] D. Guigue, M. F. Onofre, and A. Rolim, “SOAL for Music Analysis: a Study Case with Berio’s Sequenza IV,” in *JIM 2005, Journées d’Informatique Musicale*, 2005.
- [26] N. Cook, *Guida all’analisi musicale*. Milano, Guerini Studio, 1991.
- [27] A. Forte, *The Structure of Atonal Music*. New Haven-London, Yale University Press, 1973.
- [28] F. C. Ciardi, “Appunti per un modello generale di segmentazione melodica,” in *Rivista di Analisi e Teoria Musicale*, pp. 75–112, 2002/1.
- [29] F. C. Ciardi and L. Curinga, “Segmentazione melodica e interpretazione: Syrinx, un esempio applicativo,” in *Rivista di Analisi e Teoria Musicale*, pp. 61–85, 2003/2.
- [30] D. de La Motte, *Manuale di armonia*. Roma, Astrolabio Ubaldini, 2007.

L'espressività nell'esecuzione automatica di partiture di musica contemporanea: uno studio di caso

Francesco Roberto Dani, Giorgio Klauer

Conservatorio "Cesare Pollini", Padova

f.r.d@hotmail.it

giorgio.klauer@conservatoriopollini.it

Antonio Rodà, Alvise Vidolin

CSC-DEI, Università di Padova

roda@dei.unipd.it

vidolin@dei.unipd.it

ABSTRACT

È possibile imprimere un'intenzione espressiva ad un'esecuzione musicale informatica tale da venire correttamente riconosciuta da una popolazione anche se applicata a brani del repertorio contemporaneo? Una cinquantina di soggetti sono stati sottoposti all'ascolto di più versioni di due brani del secondo '900 suonati al Disklavier mediante il programma CaRo 2.0 sviluppato presso l'Università di Padova. I soggetti dovevano valutare il grado di corrispondenza di un set di attribuzioni con le esecuzione proposte; le attribuzioni erano le stesse impostabili dal programma ai fini del conferimento dell'intenzione espressiva. Un secondo esperimento è stato svolto per rilevare il gradimento rispetto alle esecuzioni ai fini di un possibile utilizzo del programma come esecutore virtuale.

1. INTRODUZIONE

Nel lavoro qui presentato è stata valutata l'applicazione nel contesto della musica contemporanea - uno scenario dove l'approccio interpretativo dell'esecutore si differenzia dalle tradizionali prassi esecutive - di uno strumento informatico risultante dalla ricerca nel campo dell'espressività musicale. Partendo dal presupposto la una macchina può eseguire solo le istruzioni per cui è stata predisposta, risulta chiaro che le informazioni contenute nella partitura sono sufficienti solamente a creare un'esecuzione nominale della stessa, ossia un'esecuzione carente di caratterizzazione espressiva. Per ottenere una autentica esecuzione di una partitura sono necessarie altre informazioni, quali deviazioni di tempo, di intensità e di articolazione [1]: ammettendo che la partitura sia il mezzo che il compositore ha a disposizione per comunicare la propria musica in maniera sufficientemente efficace per un essere umano, il sistema di notazione utilizzato non contiene tutti gli elementi necessari a dare una descrizione completa dell'espressività. Per riuscire ad avere una performance automatica a partire da una partitura sono stati costruiti modelli capaci di rappresentare l'espressività da conferire al brano, che uniti

alle informazioni provenienti dalla partitura hanno reso plausibile l'esecuzione da parte dell'elaboratore. Vari centri di ricerca sulla computer music in tutto il mondo hanno sviluppato la propria soluzione al problema, si vedano per esempio il KTH di Stoccolma [2] [3] [4], l'Università di Vienna [5], il Kansei presso l'Università di Osaka [6] ed il CSC presso l'Università di Padova [7] [8] [9] [10]. Come modello di riferimento per gli esperimenti svolti si è preso in esame il programma CaRo 2.0 sviluppato presso l'Università di Padova [11] un sistema interattivo per il rendering espressivo in tempo reale di partiture nel quale il tipo di espressione che si vuole ottenere durante l'esecuzione automatica è controllato attraverso l'interazione col piano Kinetics/Energy [7] [10]. Il piano Kinetics/Energy è derivato dal modello Valence/Arousal [12] [13], da cui differisce per l'uso di attributi sensoriali quali pesante, leggero, duro, morbido, più comuni in ambito musicale per descrivere le emozioni, anzichè attributi affettivi. Sono stati svolti dunque due esperimenti al fine di indagare la valenza delle intenzioni espressive possibili mediante il suddetto programma se applicate a brani del repertorio contemporaneo. Nel primo esperimento si è voluto indagare l'effettiva efficacia che le intenzioni espressive conferite dal programma CaRo possono avere se applicate a partiture musicali del secondo Novecento valutando statisticamente la corretta percezione delle stesse: ad una popolazione campione è stato chiesto di qualificare il contenuto espressivo di tre esecuzioni renderizzate mediante CaRo comparate con un'esecuzione nominale della stessa partitura. Nel secondo esperimento si è voluto verificare se una popolazione sottoposta all'ascolto di più versioni dello stesso brano di musica contemporanea eseguite mediante elaboratore con e senza l'intervento di CaRo sia stimolata più favorevolmente dalle esecuzioni alterate mediante CaRo o meno. Sono state fatte valutare tre esecuzioni per ognuno dei brani proposti: due esecuzioni espressive ed un'esecuzione nominale.

2. CORRETTA PERCEZIONE DELLE INTENZIONI ESPRESSIVE

2.1 Popolazione

Sono stati interpellati 49 soggetti, di cui 25 musicisti e 24 non musicisti. 26 soggetti erano di età inferiore ai 21 anni e 23 superiore ai 21 anni.

Copyright: ©2016 Francesco Roberto Dani, Giorgio Klauer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2.2 Stimoli

Si è scelto di proporre all'ascolto due brani appartenenti a correnti artistiche diverse dello stesso periodo storico, in modo da valutare il funzionamento di CaRo in contesti simili dove la metodologia interpretativa deriva principalmente dalle scelte composite. Sono stati proposti due brani dei compositori Morton Feldman [14] e Karlheinz Stockhausen [15], considerati due pilastri della musica del secondo '900. La scelta dei brani è stata condizionata sia dai limiti del programma CaRo 2.0, in quanto necessita in ingresso una partitura completamente notata, sia dalla necessità che i brani fossero scritti per lo stesso strumento utilizzabile per l'esecuzione informatica, ossia il pianoforte Yamaha Disklavier. Di Morton Feldman è stato individuato il brano *Intermission 4* (1952); di Karlheinz Stockhausen il brano *Klavierstück III* (1952). Per ciascuno sono state create 4 esecuzioni: una nominale, ossia senza alcun apporto aggiuntivo che non fosse già scritto in partitura, sfruttando la funzione "Play Mechanical" di CaRo, e tre esecuzioni renderizzate tramite la funzione "Play Expressive" denominate rispettivamente Soft, Heavy e Bright, come proposte dal programma e considerate rappresentative delle connotazioni espressive che il programma è ritenuto in grado di enfatizzare. Nel programma questi tre attributi esprimono in maniera esaustiva le estremità del modello utilizzato: Soft, caratterizzato sia da una bassa componente di Energy che da una bassa componente di Kinetic; Heavy, composto da una media componente di Energy e da una bassa componente di Kinetic; Bright, composto da una media componente di Energy e da un alto valore di Kinetic. Si è scelto di testare solamente 3 attributi dei 5 offerti dal programma (ossia Hard, Heavy, Soft, Light e Bright) per limitare il numero degli stimoli e quindi il rischio di affaticamento da parte dei soggetti; peraltro, alcuni degli aggettivi possono essere considerati simili (si pensi alla stretta relazione fra Hard e Heavy e fra Soft e Light).

2.3 Procedura

La durata dell'esperimento per ciascun soggetto è stata di circa 20 minuti. Esso era composto dalle seguenti fasi:

1. introduzione verbale nella quale vengono spiegate le finalità e le modalità di svolgimento del test;
2. compilazione da parte del soggetto del modulo con le informazioni personali;
3. ascolto comparato delle versioni del brano *Intermission 4*. In questo caso, siccome la durata del brano variava notevolmente a seconda dell'intenzione espressiva, considerato che ciò poteva infierire nella percezione veniva lasciata al soggetto la possibilità di ascoltarne anche solo una parte, a sua discrezione;
4. Ascolto comparato delle versioni del brano *Klavierstück III*.

I soggetti disponevano di un computer con un algoritmo creato in SuperCollider [16] per ascoltare le registrazioni,

un paio di cuffie chiuse ed un plico di fogli A4 prestampati dove annotare le risposte. In via preliminare veniva chiesto di rispondere a un set di domande volte a discriminare la popolazione in categorie. In seguito, per ciascuno dei due brani i soggetti erano tenuti a comparare ognuna delle tre esecuzioni ottenute con l'ausilio di CaRo con l'esecuzione nominale dello stesso brano, e per ogni comparazione a compilare il corrispondente modulo di risposte annotando in una scala da 0 a 10 il grado di corrispondenza tra i cinque attributi Hard, Heavy, Soft, Light e Bright con lo stimolo proposto. Una volta completata la sessione d'ascolto relativa al primo brano, i soggetti procedevano con la seconda sessione. È importante notare che ai soggetti non veniva fatto sapere a quali intenzioni espressive corrispondevano le registrazioni proposte.

2.4 Metodo di analisi

Si è deciso di suddividere la popolazione in categorie di età e di competenza musicale, per capire attraverso l'utilizzo di tecniche statistiche quali l'analisi della varianza ed il test di Tukey HSD se l'appartenenza a una di queste categorie fosse determinante o ininfluente per quanto riguarda la corretta percezione delle intenzioni espressive. Il software statistico utilizzato per eseguire le analisi dei dati è R [17] [18].

2.5 Risultati

Nelle tabelle seguenti sono riportate le medie e gli errori standard dei punteggi attribuiti ai cinque aggettivi per ciascuna intenzione espressiva proposta.

Medie ± errore standard relativi al brano <i>Intermission 4</i>					
Int. espr.	Bright	Hard	Heavy	Light	Soft
Soft	3.32 ± 0.41	3.69 ± 0.37	4.26 ± 0.38	4.75 ± 0.37	6.20 ± 0.39
Heavy	3.75 ± 0.36	4.59 ± 0.34	6.32 ± 0.35	3.61 ± 0.32	4.75 ± 0.37
Bright	5.04 ± 0.37	4.73 ± 0.39	5.02 ± 0.36	4.65 ± 0.32	4.86 ± 0.33

Table 1. Medie ± errore standard relativi al brano *Intermission 4*.

Medie ± errore standard relativi al brano <i>Klavierstück III</i>					
Int. espr.	Bright	Hard	Heavy	Light	Soft
Soft	4.33 ± 0.33	3.04 ± 0.33	4.02 ± 0.38	6.04 ± 0.35	6.61 ± 0.26
Heavy	4.99 ± 0.32	4.71 ± 0.35	5.00 ± 0.32	3.94 ± 0.31	4.49 ± 0.33
Bright	5.61 ± 0.38	4.78 ± 0.33	5.10 ± 0.33	3.94 ± 0.30	4.45 ± 0.31

Table 2. Medie ± errore standard relativi al brano *Klavierstück III*.

2.5.1 Intenzione espressiva Soft

E' stato condotto un test one-way ANOVA a misure ripetute sui dati riportati in Table 1 e Table 2, da cui risulta una differenza statisticamente significativa nei giudizi dei soggetti: ($F(4, 240) = 8.475, p < .001$) nel brano *Intermission 4* e ($F(4, 240) = 19.77, p < .001$) nel brano *Klavierstück III*. Svolgendo quindi il test di Tukey HSD con livelli di confidenza del 95% risulta una significatività nella maggiore votazione attribuita all'aggettivo Soft rispetto agli altri ($p < 0.01$).

	diff	lwr	upr	p adj
HARD-BRIGHT	0.3673469	-1.13119102	1.865885	0.9618919
HEAVY-BRIGHT	0.9387755	-0.55976245	2.437313	0.4224580
LIGHT-BRIGHT	1.4285714	-0.06996653	2.927109	0.0698679
SOFT-BRIGHT	2.8775510	1.37901306	4.376089	0.0000029
HEAVY-HARD	0.5714286	-0.92710939	2.069967	0.8325737
LIGHT-HARD	1.0612245	-0.43731347	2.559762	0.2957750
SOFT-HARD	2.5102041	1.01166612	4.008742	0.0000655
LIGHT-HEAVY	0.4897959	-1.00874204	1.988334	0.8972892
SOFT-HEAVY	1.9387755	0.44023755	3.437313	0.0041071
SOFT-LIGHT	1.4489796	-0.04955837	2.947518	0.0634909

Figure 1. Tukey HSD relativo all'intenzione *Soft* del brano *Intermission 4*.

	diff	lwr	upr	p adj
HARD-BRIGHT	-1.2857143	-2.5787293	0.007300746	0.0521148
HEAVY-BRIGHT	-0.3061224	-1.5991375	0.986892583	0.9663781
LIGHT-BRIGHT	1.7142857	0.4212707	3.007300746	0.0030097
SOFT-BRIGHT	2.2857143	0.9926993	3.578729317	0.0000210
HEAVY-HARD	0.9795918	-0.3134232	2.272606868	0.2311015
LIGHT-HARD	3.0000000	1.7069850	4.293015032	0.0000000
SOFT-HARD	3.5714286	2.2784135	4.864443603	0.0000000
LIGHT-HEAVY	2.0204082	0.7273931	3.3134232195	0.0002445
SOFT-HEAVY	2.5918367	1.2988217	3.884851766	0.0000009
SOFT-LIGHT	0.5714286	-0.7215865	1.864443603	0.7428513

Figure 2. Tukey HSD relativo all'intenzione *Soft* del brano *Klavierstück III*.

Un'eccezione è riscontrabile nel brano *Klavierstück III*, dove l'intenzione espressiva *Soft* non si differenzia significativamente rispetto all'attribuzione *Light* ($p = 0.74$). Altra cosa da notare è che in entrambi i brani l'attribuzione *Hard* viene spesso confusa con l'attribuzione *Bright* e con *Heavy*; inoltre, nel brano di Feldman anche l'attribuzione *Light* viene confusa con *Heavy* e *Hard*.

2.5.2 Intenzione espressiva *Heavy*

È stato condotto anche in questo caso un test one-way ANOVA a misure ripetute sui dati riportati in Table 1 e Table 2, da cui risulta una differenza statisticamente significativa solamente riguardo al brano *Intermission 4*: ($F(4, 240) = 9.582$, $p < .001$) nel brano *Intermission 4* e ($F(4, 240) = 1.801$, $p = 0.129$) nel brano *Klavierstück III*. In base al test di Tukey HSD con livelli di confidenza del 95% si può notare come nel brano *Intermission 4* risulti una significatività nel maggior punteggio attribuito all'aggettivo *Heavy*, mentre nel brano *Klavierstück III* questo non si riscontra minimamente: ciò potrebbe essere dovuto dal fatto che il brano non si presta ad essere caratterizzato con questa intenzione espressiva.

	diff	lwr	upr	p adj
HARD-BRIGHT	0.8367347	-0.5236574	2.1971268	0.4416974
HEAVY-BRIGHT	2.5714286	1.2110365	3.9318206	0.0000043
LIGHT-BRIGHT	-0.1428571	-1.5032492	1.2175349	0.9984756
SOFT-BRIGHT	1.0000000	-0.3603921	2.3603921	0.2593406
HEAVY-HARD	1.7346939	0.3743018	3.0950860	0.0049052
LIGHT-HARD	-0.9795918	-2.3399839	0.3808002	0.2792978
SOFT-HARD	0.1632653	-1.1971268	1.5236574	0.9974305
LIGHT-HEAVY	-2.7142857	-4.0746778	-1.3538936	0.0000010
SOFT-HEAVY	-1.5714286	-2.9318206	-0.2110365	0.0144962
SOFT-LIGHT	1.1428571	-0.2175349	2.5032492	0.1455326

Figure 3. Tukey HSD relativo all'intenzione *Heavy* del brano *Intermission 4*.

	diff	lwr	upr	p adj
HARD-BRIGHT	-0.26530612	-1.5306064	0.9999941	0.9784263
HEAVY-BRIGHT	0.02040816	-1.2448921	1.2857084	0.9999991
LIGHT-BRIGHT	-1.04081633	-2.3061166	0.2244839	0.1613804
SOFT-BRIGHT	-0.48979592	-1.7550962	0.7755043	0.8247626
HEAVY-HARD	0.28571429	-0.9795860	1.5510145	0.9716836
LIGHT-HARD	-0.77551020	-2.0408105	0.4897900	0.4453835
SOFT-HARD	-0.22448980	-1.4897900	1.0408105	0.9884462
LIGHT-HEAVY	-1.06122449	-2.3265247	0.2040758	0.1467381
SOFT-HEAVY	-0.51020408	-1.7755043	0.7550962	0.8020644
SOFT-LIGHT	0.55102041	-0.7142798	1.8163207	0.7531340

Figure 4. Tukey HSD relativo all'intenzione *Heavy* del brano *Klavierstück III*.

2.5.3 Intenzione espressiva *Bright*

In questo caso il test one-way ANOVA a misure ripetute sui dati riportati in Table 1 e Table 2 non restituisce alcun risultato statisticamente significativo: ($F(4, 240) = 0.235$, $p = 0.918$) nel brano *Intermission 4* e ($F(4, 240) = 3.673$, $p < .01$) nel brano *Klavierstück III*.

Procedendo col test di Tukey HSD con livelli di confidenza del 95% si può notare come in nessuno dei due brani proposti l'aggettivo *Bright* venga attribuito coerentemente.

	diff	lwr	upr	p adj
HARD-BRIGHT	-0.30612245	-1.677419	1.0651741	0.9728429
HEAVY-BRIGHT	-0.02040816	-1.391705	1.3508884	0.9999994
LIGHT-BRIGHT	-0.38775510	-1.759052	0.9835415	0.9369840
SOFT-BRIGHT	-0.18367347	-1.554970	1.1876231	0.9960633
HEAVY-HARD	0.28571429	-1.085582	1.6570109	0.9789257
LIGHT-HARD	-0.08163265	-1.452929	1.2896639	0.9998383
SOFT-HARD	0.12244898	-1.248848	1.4937456	0.9991943
LIGHT-HEAVY	-0.36734694	-1.738644	1.0039496	0.9477954
SOFT-HEAVY	-0.16326531	-1.534562	1.2080313	0.9975092
SOFT-LIGHT	0.20408163	-1.167215	1.5753782	0.9940886

Figure 5. Tukey HSD relativo all'intenzione *Bright* del brano *Intermission 4*.

	diff	lwr	upr	p adj
HARD-BRIGHT	-0.8367347	-2.1249549	0.4514855	0.3845308
HEAVY-BRIGHT	-0.5102041	-1.7984243	0.7780161	0.8123170
LIGHT-BRIGHT	-1.6734694	-2.9616896	-0.3852492	0.0039039
SOFT-BRIGHT	-1.1632653	-2.4514855	0.1249549	0.0981469
HEAVY-HARD	0.3265306	-0.9616896	1.6147508	0.9570577
LIGHT-HARD	-0.8367347	-2.1249549	0.4514855	0.3845308
SOFT-HARD	-0.3265306	-1.6147508	0.9616896	0.9570577
LIGHT-HEAVY	-1.1632653	-2.4514855	0.1249549	0.0981469
SOFT-HEAVY	-0.6530612	-1.9412814	0.6351590	0.6324962
SOFT-LIGHT	0.5102041	-0.7780161	1.7984243	0.8123170

Figure 6. Tukey HSD relativo all'intenzione *Bright* del brano *Klavierstück III*.

Si può comunque confermare che fra le tre intenzioni espressive di entrambi i brani la media maggiore dell'attribuzione Bright si riscontra proprio sull'intenzione espressiva Bright: è quindi possibile affermare che l'intenzione espressiva Bright venga ritenuta come la più Bright dalla popolazione.

Si può notare da quanto sopra riportato che l'intenzione espressiva Soft risulta essere la più robusta in entrambi i casi in esame. Anche l'intenzione espressiva Heavy risulta essere valida, seppure con la riserva del caso: si ipotizza che l'efficacia dipenda dal brano cui l'intenzione espressiva è applicata, che si immagina possedere propria espressività intrinseca che può essere variata ma non al punto da essere contraddetta. L'intenzione espressiva Bright, contrariamente, risulta non essere stata percepita al pari delle altre. Riguardo alla varianza delle risposte fra chi ha dichiarato una competenza musicale e chi no si evince una differenza solamente in due casi su sei, pertanto si può concludere che non si riscontra una differenza significativa fra i due sottogruppi della popolazione. Lo stesso si può dire riguardo alla discriminazione della popolazione in base all'età dei soggetti. Dai risultati ottenuti si può confermare che le intenzioni espressive attuate da CaRo vengono correttamente percepite dalla popolazione, con modeste limitazioni. Si voleva appurare se la sensibilità all'ascolto musicale fosse puramente legata ai cambiamenti fisici dell'esecuzione, a prescindere dall'influenza culturale data dalla conoscenza del repertorio: anche questo sembra confermato, in quanto i brani utilizzati sono stati scelti perché estranei al sistema tonale. A riprova del fatto che l'influenza culturale non condiziona la percezione delle intenzioni espressive, si deve notare come queste siano state riscontrate correttamente a prescindere sia dall'età che dalla competenza musicale che essi hanno dichiarato di avere. Si può quindi affermare con relativa certezza che i parametri gestiti da CaRo ed il modo in cui sono alterati sono determinanti nella percezione delle intenzioni espressive nell'esecuzione di brani musicali in genere, inclusi quelli appartenenti al repertorio contemporaneo.

3. VALUTAZIONE QUALITATIVA DELL'APPORTO ESPRESSIVO

3.1 Popolazione

Per il secondo esperimento stati interpellati 49 soggetti, di cui 18 musicisti e 31 non musicisti, 34 di essi erano al di sotto i 21 anni e 15 al di sopra.

3.2 Stimoli

Sono stati utilizzati gli stessi brani dell'esperimento precedente. Per ognuno dei due brani è stata comparata un'esecuzione nominale creata tramite la funzione "Play Mechanical" di CaRo con due esecuzioni in cui il programma CaRo ha attuato le intenzioni espressive in modo dinamico sulla base di regole.

3.2.1 *Intermission 4*

Questo brano è caratterizzato da passaggi lenti e note molto lunghe. Si è quindi deciso di interpretare il brano in base alla durata delle note: note lunghe (dalla semiminima in su) consecutive sono state interpretate come pesanti, quindi con l'intenzione "Hard"; note brevi (dalla croma in giù) sono state interpretate come leggere o brillanti a seconda della dinamica loro assegnata, leggere (Soft) per i piani e brillanti (Bright) per i forti. Questa suddivisione ha permesso di avere delle variazioni di espressività in alcuni punti del brano, pur mantenendo una continuità per tutta la sua durata. La prima versione è stata creata in questo modo; la seconda è stata creata semplicemente invertendo le intenzioni, ad esempio "Hard" al posto di "Soft" e viceversa.

3.2.2 *Klavierstück III*

In questo brano si presenta il problema opposto: le note sono molto ravvicinate, il tempo è veloce e le dinamiche si succedono contrapposte anche a distanza di semicroma; si rende così impossibile applicare la regola utilizzata per il brano precedente. Si è scelto quindi di creare delle sezioni in base alla media delle dinamiche delle strutture ritmiche evidenziate nella partitura. Sequenze di piano e mezzopiano sono state considerate come leggere o brillanti a seconda della durata delle note, mentre sequenze di mezzoforte e forte sono state interpretate come dure. Brevi inserti con dinamica discordante sono stati lasciati inalterati. I passaggi tra un'intenzione espressiva e l'altra sono stati realizzati mediante interpolazione lineare sul piano Kinetics/Energy della durata di poche note. La seconda versione è stata creata invertendo gli attributi.

3.3 Procedura

La durata dell'esperimento per ciascun soggetto è stata di circa 20 minuti. La struttura di svolgimento è identica all'esperimento precedente. In questo esperimento ai soggetti interpellati è stato chiesto di ascoltare tre esecuzioni per ogni brano: una esecuzione nominale e due mediante alterazione dinamica dell'espressione; le esecuzioni sono state proposte in sequenza casuale. Dopo aver ascoltato le tre esecuzioni di ciascun brano ai soggetti è stato chiesto di valutare in una scala da 0 a 10 il

gradimento per ciascuna di esse. Poichè i soggetti non hanno avuto alcun riferimento rispetto alla sequenza di ascolto, l'esecuzione nominale è stata valutata al pari delle altre.

3.4 Metodi di analisi e risultati

I metodi di analisi sono identici all'esperimento precedente, per cui si fa riferimento alla sezione 2.4. Nelle tabelle seguenti sono riportate le medie e gli errori standard dei punteggi attribuiti alle varie esecuzioni proposte.

Medie ± errore standard relativi al brano <i>Intermission 4</i>			
Brano	Dinamica dritta	Dinamica inversa	Esecuzione meccanica
<i>Intermission 4</i>	4.63 ± 0.36	5.90 ± 0.39	4.33 ± 0.46

Table 3. Medie ± errore standard relativi al brano *Intermission 4*.

Medie ± errore standard relativi al brano <i>Klavierstück III</i>			
Brano	Dinamica dritta	Dinamica inversa	Esecuzione meccanica
<i>Klavierstück III</i>	5.35 ± 0.34	6.06 ± 0.36	4.35 ± 0.47

Table 4. Medie ± errore standard relativi al brano *Klavierstück III*.

In base ai dati la versione renderizzata mediante CaRo risulta essere migliore dell'esecuzione nominale in entrambi i casi, seppure dall'analisi della varianza non si possono derivare i fattori di tale preferenza. Si conferma anche qui che non sussiste una differenza sostanziale fra chi ha dichiarato di possedere competenze musicali o meno, e che la suddivisione per età non apporta alcuna variazione significativa. Si può confermare quindi la funzionalità di CaRo per enfatizzare l'espressività di brani musicali anche nell'ambito del repertorio del secondo '900. L'approccio rimane legato al caso specifico, non potendosi dimostrare che una determinata intenzione espressiva abbia la medesima efficacia in tutti i brani.

4. CONCLUSIONI

Da quanto è emerso dagli esperimenti, si può concludere che il programma CaRo risulta un valido strumento per l'esecuzione automatica di partiture musicali anche nel contesto della musica contemporanea, considerandone comunque la varietà e l'articolazione sul piano linguistico-musicale. L'esperimento ha dimostrato che regole sviluppate per l'esecuzione automatica di musiche del repertorio classico possono essere applicate, con modesti adattamenti, anche al repertorio contemporaneo.

5. REFERENCES

- [1] L. L. Balkwill and W. Forde Thompson, "A Cross-Cultural Investigation of the Perception of Emotions in Music: Psychophysical and Cultural Cues", *Music Perception*, University of California, pp. 43-64, 1999.
- [2] A. Friberg, R. Bresin and J. Sundberg, "Overview of the KTH rule system for musical performance", *Advances in Cognitive Psychology*, Vol. 2, No 2-3, pp.145-161, 2006.
- [3] A. Friberg, "Generative Rules for Music Performance: A Formal Description of a Rule System", *Computer Music Journal*, Vol. 15, No. 2, pp. 56-71, 1991.
- [4] A. Friberg, L. Frydén, L. G. Bodin and J. Sundberg, "Performance Rules for Computer-Controlled Contemporary Keyboard Music", *Computer Music Journal*, Vol. 15, No. 2, pp. 49-55, 1991.
- [5] G. Widmer, "Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries", *Artificial Intelligence*, Vol. 146, No. 2, pp. 129-148, 2003.
- [6] S. Inokuchi and H. Katayose, "The Kansei Music System '90", *Proc of International Computer Music Conference*, pp. 308-310, 1990.
- [7] S. Canazza, G. De Poli, A. Rodà, A. Vidolin and P. Zanon, "Kinematics-energy space for expressive interaction in music performance", *Proc. of MOSART, Workshop on current research directions in Computer Music*, pp. 35-40, 2001.
- [8] G. De Poli, A. Rodà and A. Vidolin, "Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance", *Journal of New Music Research*, Vol. 27, No. 3, pp. 293-321, 1998.
- [9] S. Canazza, G. De Poli and A. Rodà, "CaRo 2.0: An Interactive System for Expressive Music Rendering", *Advances in Human-Computer Interaction*, Hindawi Publishing Corporation, Vol. 2015.
- [10] S. Canazza, G. De Poli, A. Rodà and A. Vidolin, "An Abstract Control Space for Communication of Sensory Expressive Intentions in Music Performance", *Journal of New Music Research*, Vol. 32 No. 3, pp. 201-294, 2003.
- [11] D. Nichele, "CaRo 2.0: Progettazione e sviluppo di un sistema creativo per l'esecuzione espressiva della musica", Tesi di laurea magistrale, Universit di Padova, 2015.
- [12] L. Barrett Feldman, "Discrete Emotions or Dimension? The Role of Valence Focus and Arousal Focus", *Cognition and emotion*, Vol. 12, No. 4, pp. 579-599, 1998.
- [13] Y. Yi-Hsuan, L. Yu-Ching, C. Heng-Tze and H. Chen, "Mr. Emo: Music Retrieval in the Emotion Plane", *Proc. of the 16th ACM international conference on Multimedia*, pp. 1003-1004, 2008
- [14] M. Feldman, *Pensieri Verticali*, Adelphi, 2013.
- [15] M. Tannenbaum, *Stockhausen. Intervista sul genio musicale*, Laterza, 1985.

- [16] S. Wilson, D. Cottle and N. Collins, *The SuperCollider Book*, The MIT Press, 2011.
- [17] R. Ihaka, R. Gentleman, “R: a language for data analysis and graphics”, *Journal of computational and graphical statistics*, 1996.
- [18] M. Crawley, *Statistics: an introduction using R*, John Wiley & Sons Ltd, 2005.

IMAGINARY LANDSCAPE No. 4: STUDY AND ANNOTATION OF THE SCORE

Andrea Valle

CIRMA-Università di Torino
andrea.valle@unito.it

Amedeo Casella

Università di Torino
am3d3o.casella@gmail.com

ABSTRACT

This paper presents a study of the score of John Cage's *Imaginary Landscape No. 4* for 12 radios and 24 players. After introducing the work and its history, the paper shows the relation between the formal operations at its origin and the final score. An annotation format is then introduced, and the resulting annotated version of the score is discussed. The latter may be used as an analytical tool, as a performing aid for human players, and as source data for an automated realization of the work. Finally, a complete graphic score is presented, obtained from data processing and displaying.

1. INTRODUCTION

*Imaginary Landscape No. 4*¹ (hence on: *IL4*) is the fourth instalment of Cage's *Imaginary Landscapes* series. Written in 1951 –almost 10 years after the *No. 3* (1942)– it still shares a typical feature common to all the previous pieces, the experimental attitude towards electronic technology. In the case of *IL4*, this feature is apparent, as the piece is scored for 12 radios, each to be played by 2 performers, one controlling the frequency, the other the volume and the “tone” (see later). The piece was conducted by Cage himself at the premiere [1, p.157]. The score does not mention the need for a director, but the ratio for his/her presence is to be found in the involvement of a remarkable number of performers (24) and in the synchronisation difficulties due to the notation of time (see later). The historical relevance of *IL4* is twofold.

On one side, it explores the mediascape by radically replacing, still in a live-performed piece, acoustic sources with electronic devices. Moreover, it deals only with everyday appliances (common radios), thus proposing an *ante litteram* “lo-fi” approach to live electronic music, to be pioneered extensively by Cage in other pieces (e.g. *Cartridge music*, 1960), and then largely developed by David Tudor, Cage's close collaborator. This experimental media attitude has led to consider *IL4* as a pivotal work in relation

to music exploitation of radio and (electromagnetic) noise [2, 3, 4]. It has been considered as a forerunner of random content access available from internet streaming [5]. Following recent trends in the computer music agenda, it has been noted that radios are treated like “multi-user” instruments [6, 7]. After Weinberg [8], the work has also been considered as a pioneering experience in networked music [9, 10, 11, 12]².

On the other side, *IL4* is also important in the context of Cage's production. It follows the seminal *Music of Changes* (1951). The latter piece, for solo piano, marks the beginning of the collaboration with David Tudor [13, p.178], and had a great influence among avantgarde European composers after its Darmstadt premiere in 1956 [14, p.111]. Following the inspiration of Zen Buddhism [13], in *Music of Changes* Cage experimented for the first time with *I Ching* – the Chinese divination book– as a composition method, with the aim of detaching himself from composition (“I wrote the music for radios feeling sure that no one would be able to discern my taste in that” [1, p.63]). In relation to the issue of abandoning subjectivity, Cage has stated not to be totally satisfied with *Music of Changes*, while reporting to having reached its goal with his following work, that is, *IL4* [1]. Indeed, the two works, even if sonically so different, share the same composition process, as they belong from a group of pieces based on the same set of procedures, first devised for *Music of Changes*, that have been defined as “Chart systems” [13]. Concerning *IL4*, it must also be noted that, even if the presence of 12 radios might suggest “an extremely raucous effect” [13, p.90], the widespread use of silence, together with Cage's request to use the AM tuning (where much less signal was broadcast), resulted in a very quiet piece, coherently with the aesthetic assumptions at the basis also of *Music of Changes*³.

¹ John Cage, *Imaginary Landscape No. 4*, Edition Peters, no. 6718. The handwritten title on the score by Cage is “Imaginary Landscape No. 4 or March No. 2”. Dedicated to Morton Feldman, its composition is dated by Cage between April 4th and 27th, 1951.

Copyright: ©2016 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.

² The two last claims are highly debatable, as both multi-user instruments and networked performances focus by definition on interaction, respectively among the users and the instrument, and among the nodes in a network. There is no possible interaction among performers and/or radios in *IL4*, as the piece is strictly determined in its performance.

³ AM tuning poses a series of relevant issues for actual performances. In *IL4*' score, written events concerning tuning may indeed result in absence of signal. Moreover, AM transmission is continuously declining, while also the much more popular analog FM is already going to be dismissed in some countries (e.g. Norway, from 2017, [15]). In this paper, we will not deal with performing issues. In any case, a quick survey on WWW reveals that many versions of the piece are in FM and do not take into account the tone parameter. A thorough discussion of another radio piece by Cage, *Radio Music* (1956), dealing with aspects that are relevant also for *IL4* is [16].

2. COMPOSITION

Cage described in detail the composition process of *Imaginary Landscape* in [17], while providing a more general context in [1], the two sources both dealing with *Music of Changes* and *IL4*. Following the so-called “micro-macro-cosmic” rhythmic structure [13, p.14], he devised a time organisation based on a square number of measure (144), so that “the large lengths have the same relation within the whole that the small lengths have within a unit of it” [17, p.57]. Given this pre-organised time-canvas, the events in the score result from a double, linked mapping (Figure 1). The first mapping associates three tossed coins with lines, where [17]:

- 3 heads: broken, with circle;
- 2 tails + 1 head: straight;
- 2 heads + 1 tail: broken;
- 3 tails: straight, with a circle;

By iterating six times the tossing process, Cage was thus able to build various hexagrams, that is, figures made up of 6 horizontal lines that are labelled and used by *I Ching* as the basis for divination. Hexagrams as prescribed by *I Ching* are 64 (see Figure 1, bottom right, for the whole set). The presence of the “circle” in the above definitions indicates that a second hexagram is to be generated, in which “circled” line have to be swapped, that is, broken lines with circle in the first hexagram become straight in the second, and vice versa, as shown in Figure 1 (see “mobile”).

The second mapping associates hexagrams to music parameters. To do so, Cage prepared many “charts”, where each chart is a 64-element hash table linking a hexagram with a certain value for the selected parameter. The whole procedure is only partially described in available published sources and literature. Charts for frequencies, dynamics, tone and durations are event-related, as they are used to calculate the parameters for each event. There are eight charts for each of the parameters, because eight is the maximum number of overlapping radio parts that Cage arbitrarily decided to be possible (thus, each of the 8 possible radio parts has its set of charts). Of these 8 charts, four are “mobile” while other four are “fixed”, in accordance with single or double hexagram to be used. Values in mobile charts have to be replaced once used⁴, while values in fixed charts can be used many times:

mobile means an element passes into history once used, giving place to a new one; immobile means an element, though used, remains to be used again [17, p.58]

The type of chart (mobile or fixed) “is determined by the first toss at a large unit structural point [see later on tempi], an odd number bringing about a change, an even number maintaining the previous status” [17, p.58]. Many charts are in use simultaneously, as they refer to different voices. For each voice, charts related to sound events, half of the values (32) correspond to silence; the other 32 values are

⁴ The replacement procedure is not documented in details, as far as we know. It could be observed that such a replacement weakens the idea of mapping itself.

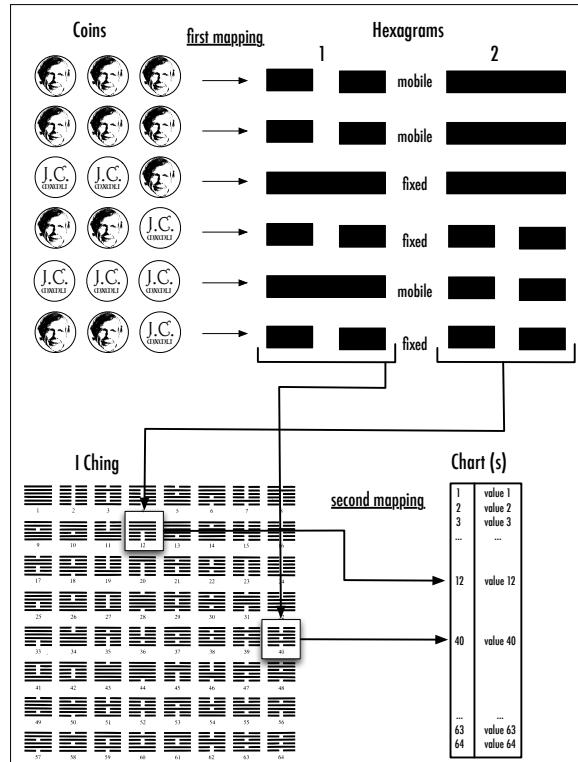


Figure 1. Schematization of the composition process.

combinations of radio frequencies. In the dynamics chart, only 16 values prescribe a new dynamics, while the other 48 values require to maintain the previous dynamics. Charts for durations specify their values by rational numbers expressed as a fraction (or as the sum of fractions) corresponding to the base time unit chosen by Cage, the whole note. Two more charts are in use, that are composition-rather than event-related, that is, they apply to all the events. They are both of fixed type. The chart for tempi “has thirty-two elements, the blanks maintaining the previous tempo” [17, p.58]. The chart includes also “accelerando” and “ritardando” indications. A new tempo is retrieved at a “structural point”, which is calculated following the pattern 2-1-3, where each integer indicates four measures of 4/4. As each page always contains four measures, the pattern applies directly to the number of pages, that is, tempo changes each 2, then 1, then 3 pages. Finally, the chart for “superimposition” indicates how many events there will be in a certain portion of the structural space. Superimposition acts like an event filter. First Cage calculates all the eight layers, then applies the superimposition parameter that may result in filtering out some of the events. This process yields an even sparser sonic texture, so that after the premiere the piece was criticised “because it was so soft”, as Cage remembered [1, p.63].

The composition process is schematized in Figure 1. To sum up, tossed coin sequences select the associated *I Ching* hexagrams. Then, hexagrams are associated to parameter values. Following his anti-subjective attitude, Cage is in-

deed quite strict in applying the procedures that he devised, so that the piece (as many others by Cage) might be considered as a pure example of algorithmic composition but, also, as a form of Parameter Mapping Sonification [18].

3. NOTATION

Music of Changes marked a radical depart for Cage. In the piece, time is no more thought of as metrical, that is, referring to a specific metric grid that, even if made flexible by various notation procedures, still acts as a discrete time quantization. Rather, Cage introduced “time notation”, where the graphic space of notation is linearly proportional to time (see [19, chap.3]). The score is thus a plot of events in time⁵. Thus, while notating *IL4*, Cage began writing the values on standard paper by means of a ruler. Later, he decided to write the piece using “traditional” notation, as he had planned the piece to be performed by 24 players, who would otherwise find difficult to read the score, both because of their reading habits related to traditional notation, and for the lack of a beat to help them progress over time (as it happens with beat counting in metric notation). At a first look, *IL4*’s score seems to be notated in an ordinary way (Figure 3). The score defines a staff for each radio, which is given a five-line staff, where note heads represent MHF frequencies, expressed using the shortened numerical values common on standard radios’ dials at time (55-160)⁶. Figure 2 shows a RCA Golden Throat: KPFA’s John Whiting, one of the performers at second and third performance, reports that in the occasion of the premiere 12 of these radios were in use (“Cage saw them in a shop window and, persuasive as ever, got the manager to lend him 12 for the premiere” [22]). In the score, the frequency range is slightly out of the dial range on both sides, extending it from 50 to 165 (reading the staff as G-clef, the boundaries are represented respectively as D under the staff and A above the staff). Pitches are clearly a rough approximations of frequencies, as furthermore Cage does not use accidentals for sake of simplicity (e.g. the same pitch represents a range of frequencies).

Frequencies are indicated numerically on top of the note heads, and standard glissando lines represent continuous transitions between adjacent values. The volume is indicated on the bottom of each staff (as traditionally with dynamics), on a scale in the range [1, 15], where 1 indicate silence and 2 is not in use⁷. The volume range is indeed arbitrarily applied to each radio, with no absolute meaning in terms of dynamics⁸. Finally, a dotted line under the notes indicates tone as “a change to high-frequency over-

⁵ Valkenburg [20] has emphasized the influence of the “smoothness” of magnetic tape manipulation on this conceptualization of time. This was already observed by Cage himself, as the composer noted that time notation “is directly analogous to the practice of cutting magnetic tapes” [17, p.29], see also [19].

⁶ That is, 540–1610 kHz in ITU region 2 (the Americas), see for USA [21]. The same range is used in *Radio Music* [16].

⁷ Cage uses 1 to indicate the lowest position of the volume knob, avoiding 0 to specifically remark that radios should not be turned off by pushing the knob to its minimum, as in analog radios this would have resulted in an audible click.

⁸ In the preface of the score, Cage proposes to use adhesive tape to mark the steps to provide a visual clue for the performers.



Figure 2. RCA’s “Golden Throat”.

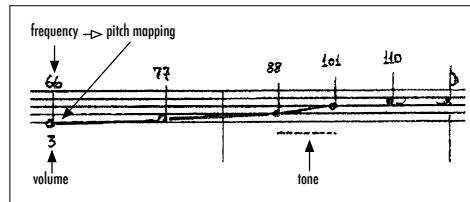


Figure 3. Notation symbols for frequency, volume, tone.

tones”, that is, it asks to switch off a discrete low pass filter commonly used on radios to avoid high frequencies in the noisy AM tuning⁹.

But a more careful inspection shows how *IL4*’s notation, at least in relation to time, results from two, partially contradictory, constraints. On one side, Cage’s purpose was to use amounts of musical durations that are not representable by the conventional notation, as they are placed in a chronometric, “smooth” extension. On the other hand, Cage tried to obtain a score that was readable by musicians accustomed to the so-called common practice notation (CPN)¹⁰. The duration charts specify event durations by means of fractions of the whole (i.e. of a 4/4 measure). By concatenating fractions, durations are added progressively without taking account quantization by measure. CPN mensural notation can be said to be “subtractive”: the measure is the time unit, and its duration is distributed among events so that at the end the former is filled. In contrast, Cage’s rhythmic organisation could be defined as “additive”, as it adds up durations without referring to the measure framework. These two ways of conceptualizing durations are mutually exclusive. In *IL4* each measure is thus deprived from its original, subtractive meaning in order to simply represent a certain amount of time, as in a time graph. Graphically, all the measures have the same length, and each page (made up of four measures) thus represents the same amount of time. In *IL4* durations are thus expressed with the CPN notation symbols, but:

⁹ This means that Cage assumes the lowpass filter standard setting as “on”. It is interesting to note that the RCA “Golden Throat” model said to be used at premiere is *not* provided with such a filter (see schematics in [23]).

¹⁰ Significantly, in the score of *Imaginary Landscape No. 5* –that does not involve live performers as it is a set of instructions for assembling tape fragments from long-playing records– Cage has decidedly opted for graph paper.

- their durations, as expressed by symbols (e.g. white vs. black note), is simply approximated, and its real value is expressed by indicating on top of the note head the associate fraction of the whole. In fact, note durations in a measure, if calculated in relation to their CPN value, rarely sums up to a whole note. But it must also be noted that fractions are *not* always notated, while sometimes they are placed on top of rests that fill entirely the measure. In this sense, this (not always consistent) notation seems a residual of the composition practice;
- their attacks depend on their actual placement in the space that stands for chronometric time (the placement of the symbols in the space depends on the sum of the durations of previous symbols). This is evident if considering the use of "X crotchets" (a term introduced by Cage in the preface of the score). Even if such symbol is provided with a traditional duration (a crotchet), Cage states that it simply "indicates the point of stopping sound and does not have any duration value": that is, the relevant element is only the placement of the symbol in space, to be used as a termination mark. Interestingly, rest symbols are placed in the middle of the time segment that they occupy (the beginning of a rest being indicated by the X crotchet). Moreover, if a note has to start exactly at the beginning of the measure, its visual placement faithfully respects the constraint and the note is thus placed *across* the bar (this is evident in the last note of Figure 3).

The situation is summed up in Figure 5. The first two measure of the score are shown in Figure 4.

In short, the 4/4 metric organisation is superimposed to time notation, and it is intended to provide a common time framework by means of the measures, and an approximate indication of the durations that it contains. In this way, the score is written in a more practically readable format without losing the original information, as contemporary events are aligned vertically by construction. In any case, the use of time notation does not imply that the ratio between time and notation space is fixed for the whole score, as it also depends on tempi. The latter are specified in beats per minute, with the quarter as a unit, in the upper left corner of the page (Figure 5), and they change following the 2-1-3 pattern, as already discussed. Such a mixed notation, while relevant to musical performance, partially hides to the analysis the overall organisation of the piece. Hence the idea of annotating the score.

4. RECONSTRUCTING THE RULED PAPER

As we saw, the actual score is notated in a partially "traditional" format, and in the score's preface Cage says the it resulted from an approximate conversion, a "rough tracing of the original" diagram (one could speculate to be written onto graph paper as in *Imaginary Landscape No. 5*), where quarter = $\frac{1}{2}$ inch. In the preface of the score, Cage also emphasizes that the same source diagram was drawn by means of an "inaccurate" ruler, so that "notation may

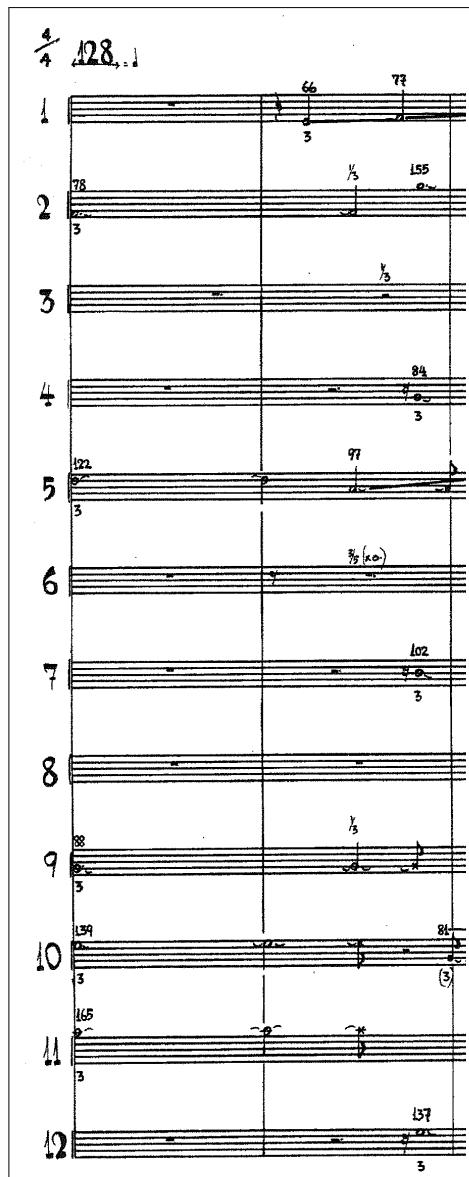


Figure 4. IL4, first two measures of the score.

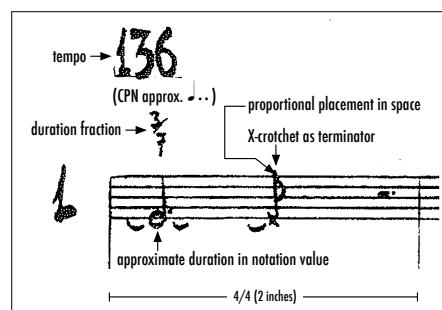


Figure 5. Duration notation.

be said to be, to say the least, approximate". These statements by Cage seem to indicate a very loose relation between the actual score and the data at its origin. In order to verify these statements and to better understand the organisation of events in *IL4*, we pursued a reverse engineering approach to the score. We inspected the printed score (a reproduction of Cage's handwritten one), by carefully measuring the notated musical signs (i.e.notes and rests as placed in the graphical space). By assuming the ratio between duration and space provided by Cage (quarter = $\frac{1}{2}$ inch) it has been possible to measure all the durations as expressed in space widths. This task was performed with the goal of verifying our reconstruction of the notation procedure and to identify eventual inconsistencies between note placements (apart from their CPN value) and declared fractions. We found the graphical score quite accurate, as we were not able to individuate any error. Considering Figure 5, it can be observed that the CPN note symbol has a graphical extension of slightly less than half the measure, as defined by the ratio $\frac{3}{7} \approx 0.43$. Curiously, its approximate duration notated by Cage in CPN –dotted half note– is by far different ($\frac{3}{4} = 0.75$)¹¹.

Thus, if considering note placement, the score is not "graphic" at all, if by graphic score we define a score that provides the user a set of indications, that are approximated within a certain range of values. Goodman [24] has proposed the couple "autographic" vs. "allographic" as a way to define the relation between a notation and its content (for music notation see [19]). An allographic notation depends on a formalized content, and the organization of the latter allows to define a symmetrical organization of the former. An autographic notation does not rely on such a well-defined organization. Its content is thus a variously extended set of performing possibilities. While many experiences in music graphic notations, in particular from the '60s and the '70s are frankly autographic (e.g. the collection published by Cage himself, [25]), *IL4*'s notation is indeed allographic, as it is defined by a clear bijection between mapped data and graphic signs, not only by construction (as explicitly declared by Cage) but also in terms of how it is actually implemented in the score, as our measurements have verified. This allographic regime allowed us to annotate the score.

5. AN ANNOTATION FORMAT

It can be said that the score is organized into events that can intuitively belong to two types: sound events and silences. We devised an annotation format that was intended to be abstract (that is, not hardware or software dependant), and easy to write and to read. The annotation process has been done by hand, and we needed a quick solution to annotate while measuring. XML encoding is rather verbose, and, while human-readable, it is typically generated by machines (or through GUI applications). To sum up, our encoding is intended as a minimal ASCII encoding that is still human-readable, partially redundant so that it allows

¹¹ A much better CPN approximation would have been a double dotted quarter note ($\frac{4+2+1}{16} = 0.4375$), as reported in Figure 5.

*	g	1/4
1	77	h
@	2/4	@
2/4 + 2/4 + 1/4 + 1/4 + 1/4	g	0
@	88	5/4
3	1/4	1
7/4	g	1/4
h	101	0
@	1/4	1/4
66	h	#
2/4	110	

Figure 6. Annotation of a score event.

fast error checking, that can be easily parsed without recurring to e.g. XML tools and converted into other formats if needed (e.g. the same XML). A sound event is described by the following parameters:

- *: beginning of the event
- ID: event unique identifier
- @: field separator
- overall duration of the event
- @: field separator
- dynamics value
- duration of the dynamics value
- dynamics type: h (hold) or g (glissando) [repeat for each value]
- @: field separator
- frequency value
- duration of the frequency value
- frequency type: h (hold) or g (glissando) [repeat for each value]
- @: field separator
- tone: [0,1], where 1 indicates presence in the score
- duration of the tone value [repeat for each value]
- #: end of the event

A silence is described by the following ones:

- *: beginning of the event both sound and silence
- ID: event unique identifier
- @: field separator
- overall duration of the event
- #: end of the event

Event IDs are progressive, and apply both to sounds and silences, regardless of the type. This allows to easily retrieve each event of the score (through the ID) and to locate it (as the ID is progressive). Figure 6 shows the resulting annotation for the event of Figure 3. In Figure 6, the annotation spans over three columns for sake of readability, and has to be read by progressive columns, from top left to right bottom.

The final annotation of *IL4*' score consists of twelve ASCII files, one for each voice.

6. DATA PROCESSING AND PLOTTING

The annotated score provides a formalized data source, compliant with Cage's composition system, that can be further processed. As an example, it may be used in case of automated performance¹². On an analytical side, our aim was to inspect visually the overall structure of the piece. A software application has been written in SuperCollider [26], as the SuperCollider language provides a rich variety of data structures, with associated visualization utilities

¹² A first, incomplete automated performance of the *IL4* has been implemented by the authors and Pietro Pasquero by controlling via Arduino digital potentiometers that replaced frequency and volume standard potentiometers in radio kits, on February 1st, 2013 in Torino. See also [16] for an automated implementation of Cage's *Radio Music*.

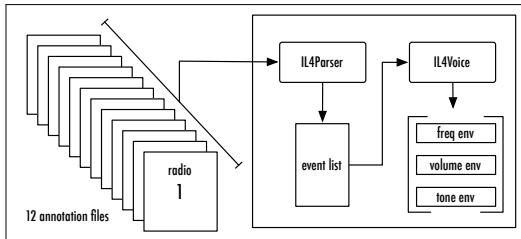


Figure 7. Software pipeline for visualization.

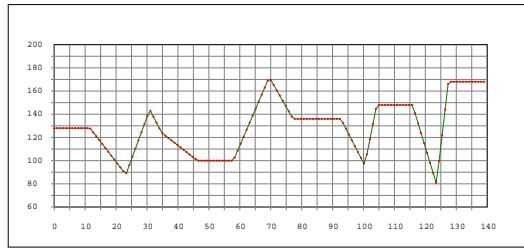


Figure 9. Envelope for tempi.

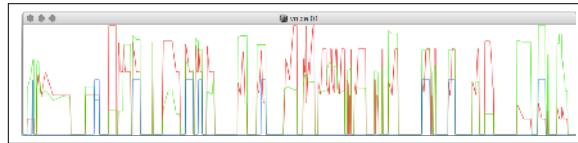


Figure 8. Synchronised envelope visualization for radio 1, whole duration.

that helped debugging. The information flow for the SuperCollider application is shown in Figure 7. Annotation files for voices are read by an instance of the IL4Parser class that outputs an event list for each radio part. The event list (that contains in a structured form all the annotated data) is passed to an instance of the IL4Voice class, that creates three envelopes (frequency, volume, tone) for the selected voice. In SuperCollider, envelopes are represented by the Env class, that codifies an envelope in form of a breakpoint table. While tone data are discrete (0 or 1), in order to take account of continuous transitions (glissando in frequency and volume), we used linear interpolation (as directly provided by the Env class). This assumption on linearity seemed reasonable because the performers do not rely on audible cues, but move the knobs only by referring to visual clues. Figure 8 shows an overlapping visualization in SuperCollider of the three envelopes for frequency (green), volume (red) and tone (blue) in the whole radio 1 part.

As discussed, even if using time notation, Cage refers to metronomic tempo as a time modulation. Because of this, absolute duration scales variably in relation to graphic width. To take into account this aspect, we have created an envelope for tempi (Figure 9, values in abscissa represent measures). Tempi have been interpolated (again, linearly) to account for accelerando/decelerando. In this case values are necessarily approximate as they depend on qualitative values to be decided during the performance (actually, this is the only “undetermined” –even if typically musical– indication provided in the score).

Envelopes for each voice represent time-stamped data regardless of time modulation prescribed by changing tempi. Thus, in order to obtain the time-modulated data, we resampled the envelopes taking into account changing ratios provided by the tempi envelope. By applying the tempi envelope, we also determined an overall absolute duration of the piece, in our case 04:37:370. The final frequency, volume and tone envelopes allow for a visual description of the score as a time graph. Figure 10 shows a representa-

tion of radio 1 part. Automated plotting has been generated by exporting data from SuperCollider into a textual format and importing them into the Python-based Nodebox package¹³. In Figure 10, on top, frequencies are reported when there is a change between adjacent frequencies greater than an (adjustable) threshold.

Page 8 shows a complete plotting of *Imaginary Landscape no. 4*, that is, including all radios with their control parameters over time. Plotting has been simplified with respect to Figure 10 so to avoid visual cluttering. The resulting graphic score allows to immediately appreciate the non uniform distribution of events among voices, even if the score is the result of chance operations. Radio 1 is the most dense part, followed by radio 2 and 8. Also, the use of the tone filter is very sparse (it is absent from 5 voices). By adjusting visual parameters, it is possible to reach the desired level of details, to fit analytical and/or performing requirements.

7. CONCLUSIONS

The accurate study of *IL4* reveals how Cage, in his first quest for anti-subjectivity, reached a -so to say- radical algorithmic approach to composition. The score clearly shows such a rigour, even if the composer himself seems to dismiss it in its preface. *IL4* is also an interesting example of how notation, as a technological means, necessarily acts as a mediator among various, sometimes contradictory, instances. The annotation of the score provides an additional analytic tool both for direct inspection and for information display, that can be directly used for performance, be it human-based or automatic.

8. ACKNOWLEDGEMENTS

We thank Francesco Richiardi for his advices on vintage radio tunings and constructions.

9. REFERENCES

- [1] R. Kostelanetz, *Conversing with Cage*. New York-London: Routledge, 1987.
- [2] A. Dunne, *Hertzian Tales. Electronic Products, Aesthetic Experience, and Critical Design*. Cambridge, Mass.: MIT Press, 2008.

¹³ <https://www.nodebox.net/code/index.php/Home>.

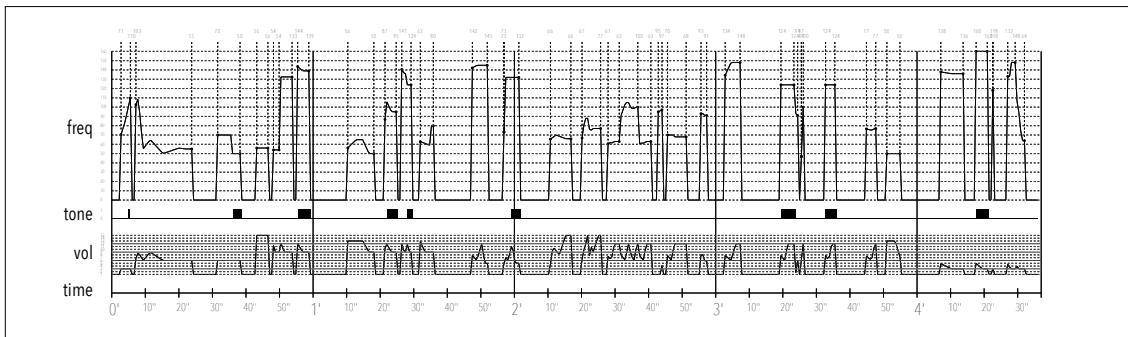
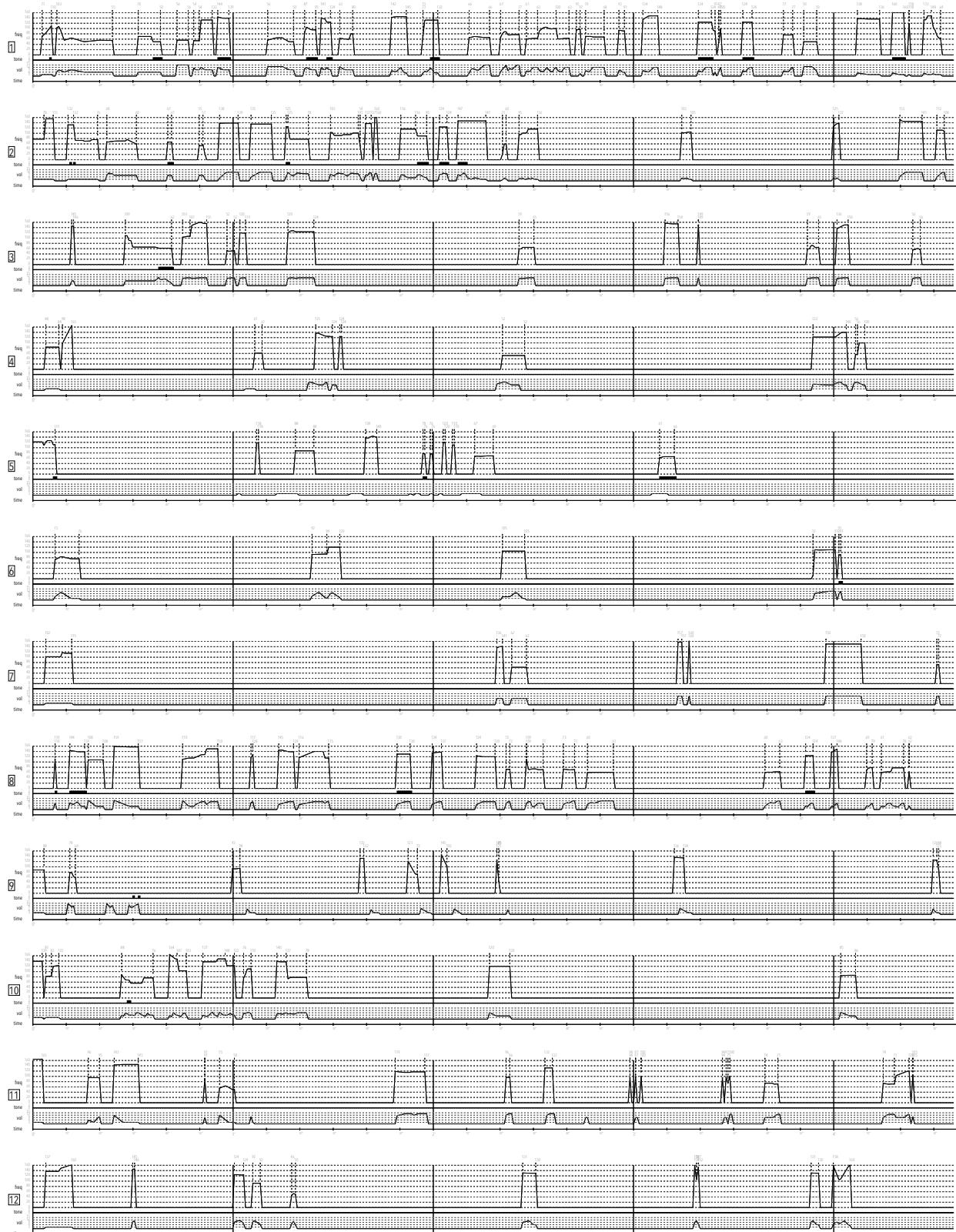


Figure 10. Visualization of *IL4*, radio 1.

- [3] D. Kahn, *Earth Sound Earth Signal: Energies and Earth Magnitude in the Arts*. Berkeley, CA: University of California Press, 2013.
- [4] S. H. Daniel Cermak-Sassenrath, Ayaka Okutsu, “Electromagnetic Landscape. In-between Signal, Noise and Environment,” in *Proceeding of the 21st ISEA*, (Vancouver), ISEA, 2015.
- [5] J. Freeman, “N.A.G.: Network Auralization for Gnutella,” in *Proceedings of the ACM Multimedia, MULTIMEDIA ’04*, (New York), pp. 180–181, ACM, 2004.
- [6] S. Jordà, “Multi-user instruments: Models, examples and promises,” in *Proceedings of NIME 2005*, (Singapore City), pp. 23–26, National University of Singapore, 2005.
- [7] S. W. Lee and G. Essl, “Live coding the mobile music instrument,” in *Proceedings of NIME 2013*, (Daejeon, Republic of Korea), pp. 493–498, Graduate School of Culture Technology, 2013.
- [8] G. Weinberg, “Interconnected Musical Networks: Toward a Theoretical Framework,” *Computer Music Journal*, vol. 29, no. 2, pp. 23–39, 2005.
- [9] P. Rebleo and A. B. Renaud, “The frequencyliator: Distributing structures for networked laptop improvisation,” in *Proceedings of NIME 2006*, (Paris), pp. 53–56, IRCAM, 2006.
- [10] P. Mathews, N. Morris, J. W. Murphy, A. Kapur, and D. A. Carnegie, “Tangle: a flexible framework for performance with advanced robotic musical instruments,” in *Proceedings of NIME 2014*, (London), pp. 187–190, 2014.
- [11] M. Akkermann, “Computer Network Music. Approximation to a far-scattered history,” in *Proceedings of the 2014 EMS Conference*, (Berlin), EMS, 2014.
- [12] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*. Singapore: Springer, 2016.
- [13] J. Pritchett, *The Music of John Cage*. Cambridge and New York and Melbourne: Cambridge UP, 1993.
- [14] A. Trudu, *La “Scuola” di Darmstadt*. Milano: UNICOPLI-Ricordi, 1992.
- [15] M. Dumiak, “Norway Pioneers the Digital Radio Future, Abandoning FM. <http://spectrum.ieee.org/telecom/wireless/norway-pioneers-the-digital-radio-future-abandoning-fm>,” 2015.
- [16] L. Vickery, “Adapting John Cage’s Radio Music for digital performance,” in *Proceedings of ACMC 2012*, pp. 69–77, Griffith University, Brisbane: ACMC, 2012.
- [17] J. Cage, *Silence*. Middletown, CT: Wesleyan UP, 1961.
- [18] F. Grond and J. Berger, *The Sonification Handbook*, ch. Parameter Mapping Sonification, pp. 363–397. Berlin: Logos, 2011.
- [19] A. Valle, *La notazione musicale contemporanea. Aspetti semiotici ed estetici*. Torino: De Sono-EDT, 2002.
- [20] J. Valkenburg, “From Bars to Inches (to Seconds): Timekeeping in the Music of John Cage,” *Dutch Journal of Music Theory*, vol. volume 15, no. 1, pp. 68–75, 2010.
- [21] U.S. Department of Commerce. National Telecommunications and Information Administration Office of Spectrum Management, “United States Frequency Allocations. The Radio Spectrum,” 2003.
- [22] J. Whiting, “My KPFA - A Historical Footnote. Random Radio.”
- [23] J. Rider, *Perpetual trouble shooter’s manual*, vol. XVIII. New York: John Rider F. Publisher, 1949.
- [24] N. Goodman, *Languages of Art*. Indianapolis: The Bobbs-Merrill Co., 1968.
- [25] J. Cage and A. Knowles, eds., *Notations*. New York: Something Else Press, 1969.
- [26] S. Wilson, D. Cottle, and N. Collins, eds., *The SuperCollider Book*. Cambridge, Mass.: The MIT Press, 2011.



Struttura, creazione, interazione, evoluzione: la ricerca al CRM per la realizzazione di *Forme Immateriali* di Michelangelo Lupone

Michelangelo Lupone, Laura Bianchini, Alessio Gabriele, Silvia Lanzalone

CRM – Centro Ricerche Musicali, Roma, Italia

{michelangelo.lupone, laura.bianchini, alessio.gabriele, silvia.lanzalone}@crm-music.it

Martina De Luca

Galleria Nazionale d'Arte Moderna e Contemporanea, Roma

martina.deluca@beniculturali.it

ABSTRACT

Il documento presenta gli aspetti tecnici, musicali ed estetici di *Forme Immateriali* di Michelangelo Lupone, un'opera musicale adattiva site-specific totalmente integrata nello spazio architettonico ricompreso nel cortile della Galleria Nazionale d'Arte Moderna di Roma. Tale spazio è infatti completamente pervaso dalla musica, creata appositamente dal compositore per essere diffusa attraverso l'acqua all'interno di due fontane, per mezzo di due particolari lenti acustiche. La musica di *Forme Immateriali* deriva dalla manipolazione e trasformazione dei suoni provenienti dall'acqua che scorre dalle fontane nelle vasche. Una volta catturati, i suoni d'acqua vengono analizzati e utilizzati per selezionare frammenti sonori che si ricompongono in tempo reale, in maniera da produrre, insieme a suoni sintetici, una dialettica polifonica composta da sequenze musicali con altezze, durate e timbri cangianti.

La forma musicale proviene da un algoritmo, basato su un sistema di regole, che permette all'opera di evolvere nel tempo, sulla base di criteri adattivi rispetto al suo stato attuale, alla sua storia e all'ambiente.

Un elevato livello di interazione consente al pubblico di generare suoni correlati al movimento e alla forma dei gesti, tramite l'immersione in acqua di una bacchetta. L'intera struttura musicale permette di instaurare un dialogo musicale tra le due fontane e di creare, nello spazio del cortile, trame sonore di sfondo e di primo piano.

Forme Immateriali, realizzato dal CRM - Centro Ricerche Musicali di Roma, è la prima opera musicale italiana commissionata dalla Galleria Nazionale d'Arte Moderna di Roma (MiBACT - Ministero per i Beni e le Attività Culturali e Turismo) ad entrare nella collezione permanente del Museo.

1. IDEA ARTISTICA

Il fascino del suono dell'acqua ha stimolato molti compositori che hanno emulato o si sono ispirati alle articolazioni delle altezze, alle infinite trasformazioni di ritmo e di timbro. I suoni dello scorrere dell'acqua, del goc-

Copyright: © 2016 Michelangelo Lupone 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.

ciolare, dell'infrangersi delle onde, hanno stimolato non solo la metafora musicale, ma anche una più ampia concezione del suono naturalistico che dall'impressionismo giunge fino all'odierno paesaggio sonoro.

La suggestione sonora prodotta dal comportamento naturale dell'acqua è stata fonte d'ispirazione contemplativa come in *Reflets dans l'eau* di Debussy, ma anche di sviluppi speculativi: ha suggerito forme e trasformazioni insolite dei ritmi, delle velocità, delle direzioni di altezze, fino ad evidenziare il rapporto di gioco che questo elemento instaura con la nostra percezione. *Les jeux d'eau à la Villa d'Este* di Liszt, o *Jeux d'eau* di Ravel si ispirano al carattere cangiante ma anche giocoso, non predicibile, effimero e stupefacente che i nostri sensi instaurano con l'acqua. L'opera *Forme Immateriali* di Lupone, prosegue il percorso di suggestioni appena descritte e trae gli elementi primari della composizione dal rapporto con il "gioco" e dal carattere evocativo dei suoni dell'acqua. I criteri con cui si strutturano i suoni si avvalgono ora della partecipazione attiva del pubblico: i gesti estemporanei ed evanescenti e i suoni che li caratterizzano sono in costante dialettica con uno sfondo musicale che fluisce in modo continuo ed esalta, per contrasto, l'espressività del gesto fugace, la valenza informativa di un fenomeno sonoro che esiste solo perché è attuale, la forza associativa della memoria a breve termine indotta dal suono che emerge e subito si dissolve, come la scia di un gesto nell'acqua.

Forme Immateriali disegnate sull'acqua producono i suoni, i ritmi e gli andamenti di una partitura musicale che si adatta e interagisce con i gesti del pubblico.

La musica che scaturisce dalle fontane muove i suoni dal fondo alla superficie, tracciando e inseguendo le traiettorie che il gesto di una bacchetta disegna nell'acqua, lasciando scie e profili fugaci. Il gesto incorporeo e la musica intangibile si fondono al segno che appare, imprime e scompare, facendo affiorare variazioni sonore.

Il suono vibra nell'acqua e il gesto ne modella e ne imprime la forma, producendo una musica cangiante. L'acqua è una materia labile che disperde immediatamente i profili derivanti dai movimenti generati al suo interno; da essi nasce il suono che produce forma musicale.

La musica che ne deriva è un complesso immateriale di vibrazioni impalpabili e invisibili, ma nell'acqua diviene concreta, corporea e si apre a molteplici visioni.

2. DESCRIZIONE GENERALE E PRINCIPI DI COMPOSIZIONE MUSICALE

Forme Immateriali è un'opera musicale concepita per essere completamente integrata alle due fontane interne al Cortile del Partigiano, presso la Galleria Nazionale d'Arte Moderna di Roma, sia per quel che riguarda la struttura architettonica che per ciò che concerne il materiale sonoro. L'opera è entrata a far parte delle collezioni del museo aggiungendosi alla serie di opere site-specific, commissionate ad artisti differenti per generazioni e poetica chiamati a interpretare lo spazio del museo e a confrontarsi con le sue collezioni. *Forme Immateriali*, infatti, non solo interviene a suggerire una nuova lettura dello spazio, ma prevede la partecipazione e l'interazione con i visitatori, invitati a disegnare nell'acqua forme che si traducono in suoni [11].

La tessitura musicale si ottiene attraverso una dialettica stretta tra figure sonore di "primo piano" e figure sonore di "sfondo". Queste ultime sono sempre attive e lentamente cangianti, assumono un carattere diverso in funzione delle stagioni e costruiscono l'ambiente sonoro complessivo del cortile che accoglie le due fontane.

Il pubblico può immergersi in acqua una bacchetta che è disponibile ai bordi di ciascuna vasca: il suo libero movimento influenza la forma musicale e le traiettorie spaziali dei suoni.

Il fondo della fontana è illuminato da una luce diffusa che colora e mette in risalto le forme e i movimenti d'acqua.

La musica in *Forme Immateriali* è generata a partire dalla trasformazione dei suoni dei getti d'acqua che si riversano nelle vasche. I continui movimenti d'acqua dovuti al gocciolamento degli ugelli che alimentano le due fontane, forniscono il movimento costante della superficie, utile alla propagazione delle onde acustiche in ogni punto di ascolto, come pure i suoni di base con cui sono state realizzate le trasformazioni e le elaborazioni musicali dell'opera.

I suoni dell'acqua, dopo essere stati rilevati all'esterno e all'interno delle vasche in punti diversi delle fontane, sono stati analizzati, scomposti in minuscoli frammenti e poi ricomposti in brevi sequenze modulate dal gesto del pubblico in altezza, intensità, durata, ritmo, velocità e timbro.

Le sequenze sono strutturate per combinarsi tra loro come in una partitura polifonica e sono rese emergenti dinamicamente e formalmente per creare un "primo piano" percettivo e un'immediata riconoscibilità del gesto interattivo che le modula. Le sequenze coesistono con un continuum sonoro che forma lo "sfondo": una sorta di ambiente omogeneo ma sempre mutevole in altezze, durate e timbri, ottenuto con suoni, singoli o molteplici, prodotti da un algoritmo di generazione in cui il valore dei parametri attuali muta in funzione delle stagioni e del tempo che trascorre. Con tali suoni l'opera mantiene un profondo legame di coerenza acustica

con l'ambiente, e come in un'osservazione al microscopio, mette in evidenza le caratteristiche timbriche e le articolazioni ritmiche più nascoste, perché interne alla materia vibrante, e quindi non immediatamente percepibili.

Il carattere d'intimità e di osservazione introspettiva è preservato e perseguito, e la musica non ha mai carattere invasivo. Il livello sonoro della musica è, infatti, sempre coerente a quello naturale delle due fontane così come si diffondono acusticamente nel cortile, autoregolandosi in modo da non invadere gli ambienti limitrofi del museo.

La musica, generata dall'acqua e correlata alle vibrazioni che si generano nell'acqua, è ascoltabile soprattutto in prossimità delle fontane; tale vicinanza permette al fruitore di entrare in relazione dialogica con esse.

Il carattere multisensoriale dell'opera favorisce l'accesso a quelle categorie di pubblico tradizionalmente escluse dalla fruizione museale come ad esempio i non vedenti e le persone con disabilità cognitive. Si tratta di un'accessibilità ampia: non è un'opera o un percorso *per i* non vedenti o *per i* portatori di altre disabilità, ma è un'opera cui *tutti* possono avvicinarsi e interpretare secondo la propria sensibilità ed esperienza, contribuendo a fare del museo un luogo partecipato e accessibile a *tutti* che non incorre nel rischio di ripercuotere le separazioni tra i cosiddetti normodotati e chi non lo è [12].

3. ALLESTIMENTO E TECNOLOGIE

Forme Immateriali è un'opera musicale complessa che integra diverse competenze artistiche, tecniche e scientifiche. Tutti gli aspetti coinvolti nell'esecuzione sono interdipendenti e contemplano la stretta correlazione tra musica, interazione con il pubblico, adattività temporale, dialogo tra le fontane, controllo e memorizzazione delle informazioni, lavoro coordinato dei sensori, sistema di generazione dei suoni.

La realizzazione dell'opera, nelle parti meccaniche, informatiche, audio, video, luci e sensoristiche, è stata affidata al CRM - Centro Ricerche Musicali di Roma. La predisposizione elettrica, l'adeguamento del locale computer, la realizzazione dei condotti per i cavi dal locale tecnico alle fontane, i cablaggi di rete e la rifinitura muraria, sono stati curati dal Museo e realizzati da una ditta specializzata.

Alla esecuzione del progetto artistico hanno partecipato, oltre al compositore Michelangelo Lupone per l'ideazione, il progetto e la realizzazione della musica, anche un responsabile generale e di produzione (Laura Bianchini); un assistente musicale per gli algoritmi musicali e la documentazione (Silvia Lanzalone); un architetto per il progetto esecutivo e l'integrazione delle tecnologie alle parti architettoniche delle fontane (Emanuela Mentuccia); un musicista informatico per gli aspetti computazionali e il trattamento dei dati sensoristici (Alessio Gabriele); un tecnico elettronico e audio per il

montaggio, i cablaggi, le misure e le tarature delle tecnologie musicali (Maurizio Palpacelli), e un assistente audio (Leonardo Mammozzetti).

L'allestimento tecnico e le tecnologie dell'opera *Forme Immateriali* sono stati realizzati in modo da soddisfare condizioni di amovibilità, manutenibilità, modularità.

Gli apparati immersi in acqua presentano caratteristiche di protezione IP68 con contatti e raccordi a tenuta stagna e alimentazioni a basso voltaggio in corrente continua. La dimensione delle tecnologie poste nelle fontane è la più piccola oggi possibile, in modo da celare alla vista o rendere appena percepibile la presenza di elementi estranei all'architettura. Ogni parte del sistema è sotto il controllo di un computer collegato ad una rete locale e accessibile da postazione remota. Un apposito programma permette di osservare costantemente le funzionalità dell'opera e di rilevare attività in corso ed eventuali malfunzionamenti delle singole parti.

L'infrastruttura tecnologica è articolata in due sezioni: la prima, posta in un locale adiacente alle fontane non accessibile al pubblico, è composta da due computer per il calcolo e il trattamento dei segnali digitali audio e video; la seconda, collocata all'interno delle fontane, è composta da un sistema di sensori e dalle lenti sonore.

La prima sezione utilizza due computer basati su piattaforma Intel® Core i7, corredati di memoria di massa di grandi dimensioni e da due gruppi di continuità. Gli elaboratori sono connessi alle unità di condizionamento e conversione numerica dei segnali audio e video provenienti dai sensori. La trasmissione delle informazioni sonore dai computer alle fontane avviene attraverso unità di conversione digitale-analogica (DAC) e un sistema di amplificatori specifico per ogni lente sonora posta in acqua.

La seconda sezione è composta da una serie di speciali lenti sonore subacquee aderenti a lastre di acciaio e telecamere adibite al rilevamento dei movimenti della bacchetta mediante cui il pubblico interagisce. Tali tecnologie sono poste in acqua, all'interno della fontana, in posizione non accessibile e visibile al pubblico, adeguatamente trattate per lavorare in immersione con ancoretti amovibili al fondo e alle pareti della fontana. I suoni sono diffusi in acqua attraverso le lenti sonore che, ancorate al fondo della fontana, insistono su lastre di acciaio che completano e ottimizzano la trasduzione delle vibrazioni in energia acustica. Le lenti e le lastre sono quattro per ogni fontana e ciascun accoppiamento è distanziato l'uno dall'altro di circa un metro.

Le lastre di acciaio hanno geometria irregolare, con dimensioni comprese fra 40 e 80 cm e spessore 0.5 mm; sono state rese solidali alle lenti e mantengono una camera d'aria stagna tra la parte convessa della lente e la base della lastra. La realizzazione della camera stagna è risultata un efficace elemento costruttivo realizzato attraverso diverse sperimentazioni in acqua, ed ha per-

messo di ottenere una particolare caratterizzazione timbrica, assimilabile al suono della insufflazione prodotta dal fiato negli strumenti a bocchino.



Figura 1: *Forme Immateriali*, vista di una delle due fontane, da cui è possibile scorgere le lastre in acciaio che affiorano sulla superficie.

Ogni lastra è posta in posizione obliqua rispetto alla superficie e fuoriesce dall'acqua di circa 30 cm. Questa posizione rende possibile il trasferimento in aria dei suoni prodotti nell'acqua e permette un adattamento dell'impedenza acustica necessaria alla conservazione dell'energia dei segnali sonori.

L'illuminazione delle fontane è realizzata con un sistema di proiezione a led IP68 ad alta efficienza luminosa; i proiettori sono posti lateralmente e fanno convergere i fasci luminosi sul bordo posteriore della fontana con un'inclinazione di 30° verso la superficie. Tale posizionamento rende evidente, la sera, i disegni dovuti al movimento della bacchetta e alle increspature dell'acqua che si correlano alle trasformazioni musicali.



Figura 2: *Forme Immateriali*, vista di una delle due fontane con illuminazione serale e un fruitore che interagisce attraverso la bacchetta.

La forma e il materiale della bacchetta sono stati accuratamente selezionati sperimentando numerosi modelli e prototipi, al fine di individuare un'adeguata ergonomia che potesse assicurare facilità di utilizzo in termini di giusta lunghezza rispetto alla profondità dell'acqua, di comodità dell'impugnatura, di corretta posizione del

baricentro, del peso, di agilità in acqua, di gradevolezza al tatto e alla vista.



Figura 3: Forme Immateriali, disegno progettuale scelto per la realizzazione della bacchetta

Il progetto della bacchetta è stato ulteriormente condizionato da requisiti funzionali legati all'adeguata rilevazione del movimento in acqua imposto dal fruttore dell'opera. Tutto ciò ha guidato la scelta del colore, delle dimensioni e del profilo dello "strumento-bacchetta" destinato alla interazione.

L'infrastruttura informativa è interamente gestita da remoto mediante una rete di comunicazione che consente il controllo e l'interscambio di dati.

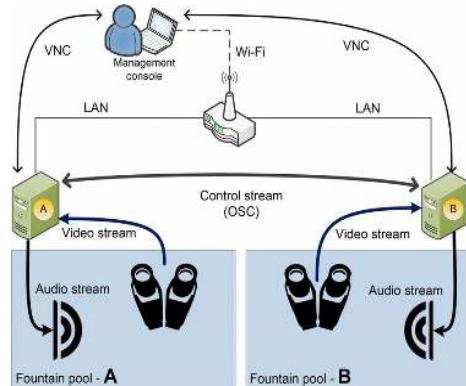


Figura 4: Forme Immateriali, flusso dati e connessioni.

Tutti i dispositivi di calcolo sono interconnessi attraverso una rete LAN (Local Area Network) dedicata: ciascun calcolatore accede alla rete con connessione cablata di tipo Fast Ethernet, mentre le unità mobili di controllo e manutenzione (computer portatili o tablet), una volta autenticate, possono accedere alla rete attraverso connessione Wi-Fi IEEE 802.11.

La comunicazione tra le fontane è implementata attraverso l'interscambio di messaggi impacchettati su protocollo OSC (Open Sound Control), mentre gli interventi manutentivi possono essere operati da remoto attraverso VNC (Virtual Network Computing).

4. ALGORITMI DI COMPUTER VISION

Le operazioni di acquisizione di immagini, elaborazione ed estrazione delle *features*, trasmissione dei dati, sono effettuate in tempo reale per mezzo di algoritmi implementati in *Max/MSP* e *Jitter*, il linguaggio di programmazione visuale di *Cycling '74* per l'elaborazione di segnali multimediali.

Un sistema di visione costituito da due videocamere subacquee poste al centro dello specchio d'acqua, a circa 25 cm di profondità, costituisce l'unico sensore mediante cui si rilevano tutte le azioni dei fruttori e alcune caratteristiche ambientali in cui l'opera vive.

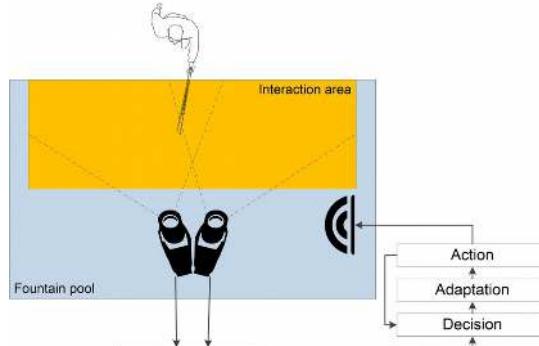


Figura 5: Forme Immateriali, sensori e attuatori, schema logico

Uno spazio sensibile (*interaction area*) costituito dallo specchio d'acqua accessibile frontalmente dai visitatori mediante la bacchetta, è stato attivato su quasi tutta la larghezza della fontana. Lo spazio è sensibile ai gesti compiuti dai visitatori sia sul pelo dell'acqua che nella parte sommersa. Le telecamere sono poste una di fianco all'altra, puntate verso il bordo frontale della vasca con un angolo di divergenza di circa 35°; tale disposizione genera una parziale sovrapposizione laterale dei quadri ripresi che consente di intervenire via software sui puntamenti con maggior precisione rispetto a quanto sarebbe possibile ottenere allineando manualmente le telecamere poste in acqua. È possibile intervenire sulla geometria del quadro catturato, mediante operazioni di zoom, traslazione, rotazione, giustapposizione, al fine di integrare i due flussi video in una scena unica in cui il gesto del fruttore può essere seguito senza soluzione di continuità, anche nelle aree di passaggio fra le inquadrature delle due telecamere. Il segnale video è sottoposto ad uno stadio di condizionamento che mira ad esaltare le caratteristiche peculiari della bacchetta utilizzata dai fruttori dell'opera e a migliorare il rapporto segnale/rumore rispetto ai dati di interesse. Si generano poi una inquadratura di superficie e una di profondità, ognuna di queste diventerà l'input di procedure specifiche di *feature extraction* nei successivi stadi di elaborazione.

Grazie ad alcuni algoritmi di percezione visuale artificiale (*Computer Vision*), l'opera prende in input delle immagini digitali e quantifica una serie di caratteristiche riferite alla scena nel suo complesso e ad oggetti specifici: ad esempio può distinguere e valutare in tempo reale i gesti con cui il fruttore perturba la superficie o compie con la bacchetta dei movimenti in profondità. Attraverso l'estrazione di tali *features*, è stato possibile gestire una rappresentazione snella e mirata del contenuto dell'immagine: inseguendo solamente le caratteristiche ritenute d'interesse per l'interazione si è ottenuta una significativa riduzione delle informazioni da amministrare. Una volta estratte in tempo reale alcune caratteristiche "di base" della scena ripresa, come ad esempio la presenza e la posizione delle bacchette dentro l'acqua, la quantità di moto e la zona di maggior perturbazione

sulla superficie dell'acqua, o ad esempio le condizioni d'illuminazione della scena e l'eventuale presenza rumore [7], di corpi estranei o fasci di luce non significativi per l'interazione, è stato possibile astrarre delle *features* di più alto livello mediante pesatura e aggregazione di informazioni semplici. Attraverso la gestione di caratteristiche di complessità e astrazione differenziate, si ricavano proprietà geometriche qualitative e quantitative della scena oltre che proprietà semantiche del gesto con cui il fruitore interagisce con l'opera.

La molteplicità di informazioni contenute nelle immagini, insieme alla frequente e mutevole presenza di "rumore" costituito principalmente da aberrazioni della prospettiva dovute alla rifrazione in acqua, dalle condizioni di illuminazione fortemente variabili nel corso della giornata e durante il passare delle stagioni, dai corpi estranei presenti nell'acqua come foglie, alghe, piccoli rami, ha reso necessaria la realizzazione di un sistema di visione capace di adattare la ripresa alle mutevoli condizioni ambientali, di isolare specifiche caratteristiche della scena e di avviare un processo decisionale per la selezione delle caratteristiche utili alla gestione dell'interazione con il pubblico. Tutti i processi costruttivi tecnici e tecnologici utilizzati per catturare l'interazione del fruitore e del contesto ambientale "intorno" all'opera, ivi compresi i dispositivi di ripresa e gli algoritmi di lettura delle immagini, l'elaborazione e il condizionamento del segnale video, l'estrazione delle features, sono stati oggetto di un impegnativo processo di evoluzione imposto in egual misura sia dal pensiero formale e musicale espresso dal compositore, sia dalle caratteristiche ambientali specifiche del sito (Cortile del Partigiano) per cui l'opera è stata concepita. L'opera, il sito e il fruitore diventano così parte integrante di un'unica esperienza sensoriale immateriale, in uno scambio reciproco di stimoli, risposte e adattamenti.

5. ALGORITMI DI SONIFICAZIONE DEL GESTO

L'implementazione dell'algoritmo di generazione musicale di *Forme Immateriali* è stato realizzato tramite il linguaggio Max/MSP secondo criteri che tengono conto della particolare caratterizzazione timbrica e spaziale dovuta alla disposizione dei diffusori in acqua e al loro accoppiamento con le strutture planari in acciaio.

La musica che emerge dalle fontane è composta da uno sfondo sonoro sempre presente e caratterizzato da micro variazioni interne al suono. In condizione di stasi dell'acqua nelle fontane, dovuta ad assenza di interazione, tali variazioni sono continue, ma statisticamente costanti. Le increspature della superficie dell'acqua dovute al vento, alla pioggia, o alla caduta di elementi naturali provenienti dal giardino centrale che sono considerate rilevanti dall'algoritmo di computer vision precedentemente descritto, comportano l'estrazione di un dato relativo alla quantità di movimento che è stato associato

alle microvariazioni dei parametri coinvolti nella realizzazione del suono di sfondo. In questo modo è possibile utilizzare tali informazioni per variare i comportamenti statistici caratterizzandoli con una maggiore o minore densità di accadimenti all'interno della tessitura sonora. La stessa tipologia di comportamenti interviene anche quando la bacchetta è utilizzata per il solo sfioramento della superficie dell'acqua che produce perturbazioni più o meno ampie.

La logica di realizzazione musicale dello sfondo è stata implementata con un algoritmo di generazione polifonica di suoni di sintesi sommati a suoni di gocce d'acqua, la cui evoluzione temporale e le cui trasformazioni sono gestite tramite la variazione semi-casuale dei parametri di step incrementale delle wave tables e dei segmenti degli inviluppi d'ampiezza. La sintesi sonora è basata su criteri non standard di modulazione d'ampiezza e di frequenza ed è differenziata per le due fontane. L'evoluzione dell'opera nel tempo durante i diversi periodi dell'anno è organizzata tramite un sistema di regole rispondente alle scelte musicali e stilistiche del compositore, ma è anche dovuta agli adattamenti conseguenti alle interazioni dei fruitori e dell'ambiente circostante.

Il sistema di rilevazione dei movimenti sotto la superficie dell'acqua consente inoltre di integrare musicalmente anche alcune interazioni dovute all'ambiente naturale, sia esterno che interno alla vasca, come brevi movimenti di foglie sul fondo, o come la caduta in acqua di piccoli rami, semi o altri elementi naturali provenienti dal giardino centrale, laddove questi presentino caratteristiche di rilevanza. Il principale contributo al destino musicale dell'opera è però affidato ad azioni volontarie effettuate tramite l'uso della bacchetta.

L'interazione realizzata nell'algoritmo di calcolo relativo alla produzione dei suoni di *Forme Immateriali* prevede l'utilizzo espressivo di eventi acustici generati dal fruitore quando immerge la bacchetta al di sotto della superficie dell'acqua. Tali eventi si inseriscono coerentemente nel tessuto sonoro cangiante di base delle due fontane, producendo figure e profili ritmico-melodici capaci di collocarsi in relazione di primo piano rispetto agli altri suoni di sfondo e di modificare quindi il destino formale dell'opera musicale. Per tali ragioni i materiali sonori affidati al fruitore sono tratti dai suoni naturali delle due fontane precedentemente registrati e sottoposti ad un primo livello di elaborazione per ricavarne grani più o meno articolati secondo strutture interne diversificate. I grani così ottenuti, di durata variabile, sono memorizzati e dati in pasto a un algoritmo di elaborazione del suono in tempo reale, i cui parametri sono correlati allo spostamento della bacchetta rilevato in acqua, attraverso particolari strategie di *mapping* prestabilite per ciascun grano. L'inserimento della bacchetta in acqua da parte del fruitore comporta la selezione di uno dei materiali sonori precedentemente memorizzati, tramite una logica che tiene conto della quantità, della durata e del profilo gestuale degli accadimenti. Le modifi-

che timbriche disponibili al fruitore tramite l'inserimento della bacchetta in acqua sono relative alla variazione di parametri di ampiezza, velocità di lettura tabellare, frequenza di taglio e pendenza di un filtro numerico del secondo ordine, che può essere di tipo band-pass, high-pass, low-pass, hi-shelving.

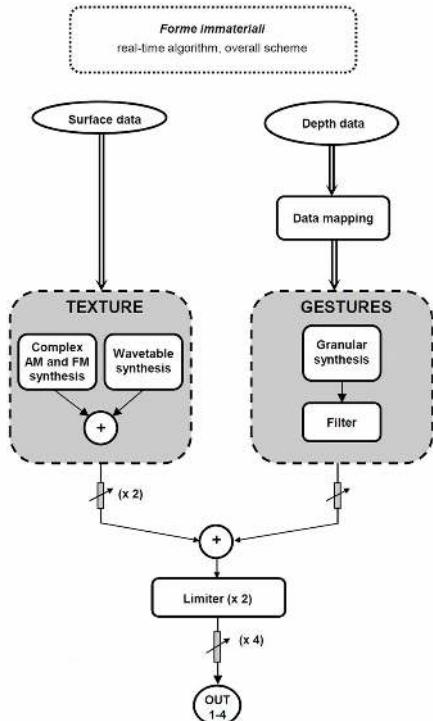


Figura 6: Schema di alto livello dell'algoritmo di elaborazione dei suoni.

Nella figura 6 è mostrato il diagramma di flusso dell'algoritmo generale di produzione della musica, simile per entrambe le fontane. Le varianti musicali tra le due fontane consistono nella tipologia di sintesi sonora, nella caratteristica dei materiali concreti memorizzati e nella diversa mappatura dei parametri. Il sistema di *Computer Vision* traduce i movimenti della bacchetta e trasmette all'algoritmo di elaborazione dei grani due coordinate fondamentali: spostamento sull'asse orizzontale parallelo al bordo della vasca e spostamento lungo l'asse verticale relativo alla profondità d'immersione. I valori di movimento della bacchetta sono normalizzati e modellati tramite una funzione la cui forma può essere rettilinea, spezzata, o curvilinea con diverse tipologie di curvatura, come ad esempio parabola, ellisse, mezzo seno o coseno, esponenziale, logaritmica, ecc. Le successive operazioni di riscalamento e di interpolazione consentono di associare l'evoluzione di ciascuno di questi valori di movimento ad un parametro specifico di elaborazione del suono. I criteri sono analoghi a quanto già utilizzato da Michelangelo Lupone per la realizzazione di precedenti opere, come *Contropasso*, composta tra il 1998 e il 1999 [4, 5, 6].

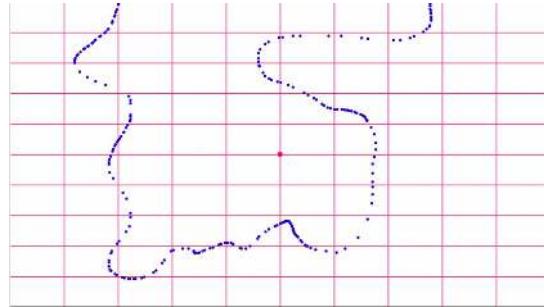


Figura 7: Forme Immateriali, esempio di simulazione grafica di un gesto compiuto in acqua tramite la bacchetta.

L'algoritmo originario rileva e converte i dati provenienti dal movimento del mouse all'interno di un'interfaccia grafica caratterizzata da un riquadro in cui vengono colorati i pixel corrispondenti ai percorsi del gesto in modo da realizzare dei tracciati visivi e valutarne le caratteristiche di evoluzione timbrica. La specificità di questo algoritmo risiede nella possibilità di associare ad un solo gesto la variazione indipendente di più di un parametro, in modo che ad un evento gestuale semplice possa corrispondere una variazione anche molto complessa dell'evento musicale. Un elemento importante di caratterizzazione espressiva dell'implementazione riguarda la risposta dell'algoritmo di elaborazione, che risulta sensibile alle differenze di velocità del movimento in relazione ai livelli e alle tipologie d'interpolazione dei parametri ad esso associati, soprattutto per quanto riguarda il filtraggio numerico. Tali aspetti consentono di riprodurre andamenti temporali e profili spaziali in modo senz'altro coerente alle caratteristiche gestuali del movimento. L'implementazione dell'algoritmo musicale di *Forme Immateriali* è stata preceduta da uno studio sulle tipologie di movimento della bacchetta in acqua in relazione alle possibili forme e velocità del gesto. Lo studio è stato suddiviso in sessioni di registrazione visiva di diversi tipi di movimento in acqua e da sessioni di simulazione grafica e acustica dei movimenti più ricorrenti e significativi. Tale studio ha consentito di individuare e produrre una sufficiente quantità di grani ottimizzandone preventivamente la "mappatura" dei parametri rispetto alle tipologie di gesto più rilevanti per il compositore. La figura 7 mostra un esempio di simulazione grafica di un gesto composto da una fase di immersione e una di emersione della bacchetta, laddove la diversa distanza tra i punti disegnati identifica una differenza di velocità.

6. CARATTERISTICHE DI FRUIZIONE E DEI MODI D'INTERAZIONE

Il carattere interattivo dell'opera consente al pubblico di inserirsi come co-artece degli accadimenti musicali, rendendo possibile non solo una comunicazione tra il fruitore e la fontana, ma anche tra le due fontane, poiché ciascuna risuona dei gesti effettuati sull'altra.

Il gesto del fruitore sulla superficie dell'acqua produce devianze, accelerazioni, condensazioni e dilatazioni dei profili sonori, con conseguenti trasformazioni musicali dell'opera. Il pubblico diventa, in questo modo, protagonista delle variazioni musicali e anche del movimento spaziale dei suoni intorno e dentro la fontana [8].

Il fruitore, compiendo un gesto nell'acqua, ascolta una variazione sonora in prossimità della fontana sulla quale è intervenuto, e trasmette all'altra fontana le caratteristiche delle vibrazioni sonore generate, stimolando ulteriori varianti musicali dell'opera. Tale condizione di dialogo permette a più fruitori di interagire musicalmente tra loro, anche se spazialmente lontani, in una sorta di contrappunto.

Per la natura stessa del luogo per cui è stata concepita, *Forme Immateriali* è un'opera disseminata: non c'è un centro definito per la fruizione e l'interazione ma una molteplicità di luoghi. La partecipazione dei visitatori rispetto all'opera è cooperativa e distribuita: seppur pienamente indipendenti dal punto di vista tecnico e computazionale, le due fontane sono funzionalmente interconnesse in quanto le interazioni e i cambiamenti di stato di una influenzano anche gli adattamenti e il comportamento musicale dell'altra, con modi e tempi che si adattano a quanto previsto dalla partitura musicale.

Forme Immateriali è anche un'opera adattiva poiché è in grado di seguire una evoluzione musicale preservando una propria identità formale ed espressiva, anche a fronte delle elaborazioni che si susseguono nel tempo e delle interazioni con i fruitori e l'ambiente circostante. Attraverso la tecnologia informatica l'opera cattura il comportamento del fruitore ed evolve conseguentemente il processo di creazione ed elaborazione musicale, rispettando una serie di regole compositive che definiscono l'invariante stilistica dell'opera [1]. L'opera dunque è un sistema che si auto-regola e che modifica il proprio stato in relazione alle azioni del fruitore e dell'ambiente circostante, restituendo al fruitore risposte che tengono conto non solo della interazione attuale, ma anche della successione delle precedenti azioni e dell'intero contesto ambientale, conservando la propria identità formale ed espressiva [2, 3].

7. L'IMMATERIALE E I NON VEDENTI

L'opera *Forme Immateriali* è concepita con il preciso intento di estendere la possibilità di fruizione e interazione ai non vedenti e agli ipovedenti.

La musica fonda le proprie caratteristiche espressive sulla percezione acustica dei suoni e delle relazioni parametriche tra questi. Un crescendo d'intensità, un glissando di altezze, per esempio, stimolano il sistema cognitivo a rappresentare questi fenomeni acustici come "forme" in un dominio immaginifico, spirituale, intangibile.

Nell'opera *Forme Immateriali* questa rappresentazione è accentuata dalla possibilità di condizionare i suoni al movimento di una bacchetta nell'acqua. La direzione

e la velocità del gesto operato per mezzo della bacchetta si correlano alle variazioni musicali ascoltate come un processo di causa-effetto; da ciò nasce la percezione profonda di forma che si incarna nel suono e ne disegna il profilo temporale e i modi di variazione, rendendo espressivo tutto ciò che è movimento astratto ed evanescente, non necessariamente visibile.

8. CONCLUSIONI

Forme Immateriali costituisce un ulteriore passo in avanti verso lo sviluppo di opere adattive che, insieme ad altre opere create da Michelangelo Lupone e realizzati dal CRM – Centro Ricerche Musicali di Roma, costituiscono un "corpus" di produzioni innovative volte a trasformare il concetto stesso di opera musicale nel rapporto con la percezione della forma nel tempo, nella relazione con le forme plastiche e con lo spazio acustico e architettonico, nella relazione con il fruitore [9, 10].

Forme Immateriali è anche la prima opera musicale italiana, committenza della Galleria Nazionale d'Arte Moderna (MiBACT - Ministero per i Beni e le Attività Culturali e Turismo), ad entrare nella collezione permanente di un Museo.

9. RIFERIMENTI

- [1] Inverardi P., Pelliccione P., Lupone M., Gabriele A.: "Ad-Opera: Music-inspired Self-adaptive Systems". In Zander J. & Mosterman P. (org.) *Computation for Humanity: Information Technology to Advance Society*, p. 359 – 379, UK CRC Press, 2012.
- [2] Gabriele A.: "LAdOp1: definizione e computazione degli adattamenti in opere d'arte intermediali adattive". *Atti del XX CIM Colloquio di Informatica Musicale*, Roma, 2014.
- [3] Lupone M.: "Interattivo, Adattivo, Evolutivo". *Lecture at University of Belgrade – Faculty of Architecture*, CRM - Centro Ricerche Musicali, 2008.
- [4] Lanzalone S.: "The Manipulation of 'Body Sound'. Interaction between Music and Dance. Analysis of 'Contropasso' by Michelangelo Lupone", in A.A. V.V., *XIII CIM – Colloquium on Musical Informatics. Atti*, pp.112-115, L'Aquila, 2000.
- [5] Lanzalone S., "SDP – Sonorous Drawing Plane and Laboratory 'The Sonic Draw'. Timbric Exploration in Real Time through Expressive Gesture Tracks", in A.A. V.V., *XIII CIM – Colloquium on Musical Informatics. Atti*, pp.209-212, L'Aquila, 2000.
- [6] Lanzalone S., "HIDDENS GRIDS. Paths of expressive gesture between instruments, music and dance". *Organized Sound, International Journal of*

- Music and Technology*, Cambridge University Press, vol.5, n°1, pp.17-26, 2000.
- [7] Xie M.: "Fundamentals of Robotics – Linking Perception to Action. Series in Machine Perception and Artificial Intelligence" – Vol. 54. Singapore: World Scientific, 2003.
 - [8] Lupone M.: "Musica e Mutazione". High-Art, rivista semestrale d'informazione dell'Alta Formazione Artistica e Musicale, apr. – ott. 2008 (nr 1). Gangemi Editore, 2008.
 - [9] Bianchini L.: "From real time performance to interactive Art Sound Installation". *Proceedings of Journée d'étude internationale MINT-OMF "musique et interactivité"*, Université Paris VIII, 2008.
 - [10] Bianchini L.: "Scenari dell'ascolto. Le nuove forme dell'arte musicale". Terra Fertile, Sassari, 2010.
 - [11] De Luca M., Gennari Santori F., Pietromarchi B., Trimarchi M. (org.): "Creazione contemporanea. Arte società e territorio", Roma, Sossella editore, 2004.
 - [12] Bellisario M. G.: "Politiche pubbliche per l'arte contemporanea: costruire una rete". Economia della Cultura, anno XXIV (nr 1). Bologna: Il Mulino, p. 19 – 30, 2014.

Un dispositivo per l'analisi del rapporto tra musica e gesto nelle launeddas

Marco Lutzu

Università di Cagliari

mlutzu@livestudio.it

ABSTRACT

Nel maggio del 2007 l'ingegnere Francesco Capuzzi presentò una versione elettronica delle *launeddas*, il triplo clarinetto di canna considerato lo strumento più importante della musica tradizionale sarda. Quando ascoltai le *elettroneddas* per la prima volta pensai subito che, con alcune modifiche, sarebbe stato possibile utilizzarle per analizzare il rapporto tra musica e gesto. Così, assieme all'ingegnere Guido Coraddu, io e Capuzzi abbiamo apportato alcune modifiche allo strumento con l'intento di realizzare un dispositivo in grado di mappare la posizione delle dita dei suonatori durante la performance.

Questo articolo presenta i primi risultati della ricerca svolta con tale dispositivo, realizzata grazie ad alcune prove effettuate con la collaborazione del suonatore di launeddas Andrea Pisu. Verrà mostrata l'efficacia della rappresentazione grafica fornita dal software *e-launedda* per lo studio delle diteggiature, inoltre si cercherà di dimostrare come la scelta della diteggiatura dipenda sia da fattori musicali che ergonomici; infine verrà evidenziata la rilevanza assunta da alcuni pattern di movimento nella musica per launeddas.

1. INTRODUZIONE: MUSICA E GESTO IN PROSPETTIVA ETNOMUSICLOGICA

Fare musica presuppone il movimento del corpo. I gesti di un musicista sul proprio strumento sono determinati da esigenze funzionali di base (per esempio evitare che lo strumento scivoli dalle mani o avere un accesso più agevole ai fori o alle corde) e da norme culturalmente determinate, riconducibili direttamente alle caratteristiche del repertorio e alle estetiche che guidano le diverse espressioni musicali, oltre che ad aspetti quali le funzioni dello strumento e della sua musica all'interno della società e, più in generale, alla cinesica culturale. L'etnomusicologia, intendendo la musica non solo come mero risultato sonoro, si è interessata fin dalle sue origini al rapporto tra musica e corpo, occupandosi non solo di danza ma concentrando anche sullo studio della gestualità dei musicisti durante la performance.

Copyright: © 2016 Marco Lutzu. 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.

Già nel 1929 Eric von Hornbostel, dopo aver ascoltato le registrazioni di alcuni suonatori di xilofoni africani, notò che questi: «realizzano melodie prima di tutto come atto motorio, considerando la loro qualità udibile come una questione secondaria, piuttosto che desiderabile» [1]. Gli studi finalizzati a identificare le relazioni tra pattern motori e musicali condotti da John Blacking sulla musica per flauto Butembo [2] e sulla kalimba Nsenga [3], così come quelli realizzati da Gerard Kubik in Africa centrale e orientale [4] portarono quest'ultimo ad affermare che «la musica africana può essere definita in una dei suoi aspetti fondamentali come un sistema di pattern motori» [5].

L'ipotesi che per uno strumentista la rappresentazione cognitiva della musica nel corso della performance possa essere di tipo motorio piuttosto che esclusivamente acustica è stata esaminata da John Baily a partire dai suoi studi sui liuti afgani [6, 7].

Più recentemente, diversi studi sul gesto musicale si sono avvalsi di un approccio interdisciplinare che ha visto i musicologi lavorare accanto a ingegneri, neuroscienziati, psicologi cognitivi, informatici e altri studiosi¹. Tali ricerche si avvalgono sempre più spesso di nuove tecnologie che affiancano all'osservazione diretta l'analisi informatica dei video o sistemi di *motion capture* meccanici, inerziali o infrarossi.

Questo articolo presenta i risultati preliminari di uno studio sul rapporto tra musica e gesto nelle launeddas realizzato con la versione modificata delle *elettroneddas* associate al software *e-launedda*.

2. LAUNEDDAS, UNA BREVE INTRODUZIONE

Il termine launeddas identifica un triplo clarinetto di canna diffuso nella Sardegna meridionale composto da un bordone e due canne melodiche di differente lunghezza. Il bordone, detto *tumbu*, e la canna melodica più grave, detta *mancosa*, sono legati assieme e tenuti con la mano sinistra. La terza canna, detta *mancoseda*, viene impugnata con la mano destra (Fig. 1).

Mancosa e *mancoseda* hanno ognuna cinque fori frontali di forma rettangolare. Quattro vengono diteggiati mentre il più basso, detto *arrefinu*, risuona quando gli altri sono chiusi. Le ance sono ricavate incidendo

¹ Per una sintesi degli studi più recenti e alcune proposte su possibili futuri sviluppi si veda [8].

una linguetta su un piccolo cannellino detto *cabitzina*, infisso sull'estremità superiore delle tre canne. Le due canne melodiche sono realizzate con una varietà di canna detta localmente *canna mascu* (canna maschio), mentre per il *tumbu* e per le tre *cabitzinas* si utilizza la *canna fēmina* (canna femmina).

Gli altri materiali impiegati per la costruzione delle launeddas sono lo spago e la cera d'api. Il primo viene utilizzato per rinforzare le parti più fragili dello strumento e per tenere assieme *tumbu* e *mancosa*. La cera si usa invece per intonare lo strumento, sia correggendo la posizione dei fori sia regolando la frequenza di vibrazione delle ance.

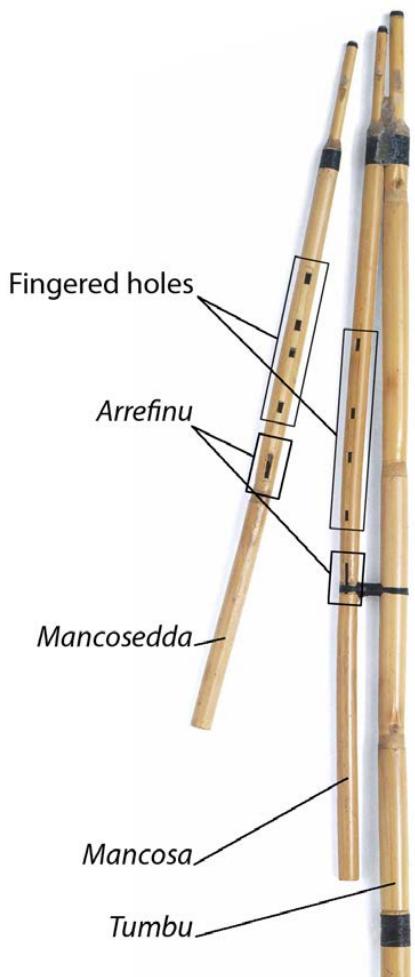


Figura 1. Launeddas.

Le launeddas non sono un unico strumento ma piuttosto una famiglia di oltre dieci tipologie di strumenti (detti *cuntzertus*) ognuno dei quali ha un proprio nome come *medianu*, *fiorassiu*, *puntu* 'e *òrganu*, *ispinellu* e altri ancora. Ciò che distingue un *cuntzertu* dagli altri è l'ambito melodico delle due canne, sempre basato su un

frammento di scala maggiore avente la nota di bordone come centro tonale.

La tecnica esecutiva delle launeddas è caratterizzata dalla respirazione circolare, che permette allo strumento di essere suonato senza soluzione di continuità per diversi minuti utilizzando la riserva d'aria accumulata nella guance mentre si inspira dal naso. Le dita devono stare distese e rilassate così che, a seconda della morfologia della mano del suonatore, i fori vengono tappati con la prima o la seconda falange.

Una statuetta di bronzo risalente al IX-VIII sec. a.C. che raffigura un suonatore di un aerofono tricalamo a canne divaricate testimonia l'antica origine delle launeddas [9,10]. Fonti più recenti attestano una ampia e radicata diffusione dello strumento a partire dai primi anni del XVIII secolo [11].

Attualmente in Sardegna sono presenti decine di suonatori di launeddas, quasi esclusivamente uomini, che suonano a livello semi-professionali perlopiù in occasione di feste patronali, quando il simulacro del santo viene portato in processione per le vie del paese al suono dello strumento, o per l'accompagnamento del ballo tradizionale.

3. LE ELETRONEDDAS

Nel 2003 l'ingegnere Francesco Capuzzi iniziò a lavorare a un prototipo di launeddas elettroniche che nel maggio del 2007 presentò sul suo blog con il nome di *elettroneddas* (Fig. 2). Tale strumento non richiede l'utilizzo del fiato. È composto da tre tubi di alluminio nei quali i fori sono sostituiti da sensori a contatto collegati a degli oscillatori (uno per ogni canna) che generano un suono di sintesi. Un *controller* esterno permette al suonatore di selezionare il *cuntzertu*, l'altezza del centro tonale, il timbro e di cambiare la scala di riferimento.

Le *elettroneddas* hanno avuto un certo successo e incuriosito diversi suonatori di launeddas, i quali hanno iniziato ad utilizzarne nei loro concerti solisti o in ensemble di musica sperimentale.



Figura 2. Elettroneddas.

4. IL SOFTWARE E-LAUNEDDA

La tecnica esecutiva delle launeddas richiede rapidi movimenti delle dita e un buon livello di coordinazione tra le due mani, in particolare per la musica da danza, considerata la parte più difficile del repertorio. Il tempo veloce, la presenza del bordone, la grande qualità di abbellimenti e ornamentazioni e il fatto che in alcuni *cuntzertus* la stessa nota è presente in entrambe le canne melodiche rende particolarmente complesso determinare la diteggiatura semplicemente ascoltando una registrazione o osservando un sonatore dal vivo ([video 1](#))².

Inoltre, le analisi attraverso l'osservazione dei video si sono rivelate insufficienti a causa del fatto che i suonatori solevano le dita molto rapidamente e solo di pochi millimetri. Quando ho sentito le *elettroneddas* per la prima volta ho subito pensato che, con opportune modifiche, sarebbero potute diventare uno strumento utili per analizzare il gesto musicale. Così chiesi all'ingegnere Capuzzi di realizzare un prototipo delle *elettroneddas* che potesse essere collegato a un computer, e all'ingegnere Guido Coraddu di creare un software capace di rappresentare graficamente il movimento delle dita sullo strumento. La frequenza di campionamento è stata fissata in 10ms. La schermata di *e-launedda* è suddivisa in due parti: nella parte superiore viene mappato il movimento delle dita sulla *mancosedda*, in quella inferiore la diteggiatura della *mancosa* (Fig. 3).

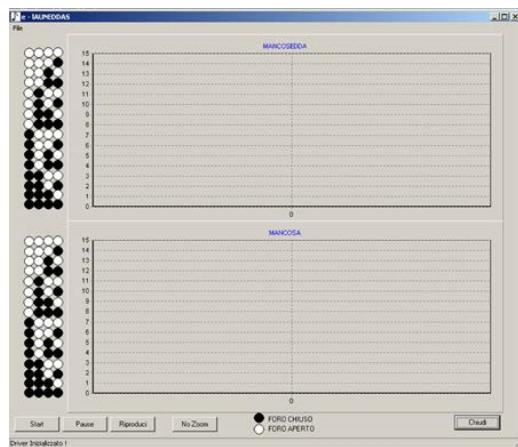


Figura 3. La schermata di *e-launedda*.

Per ognuna delle due canne melodiche l'asse X rappresenta il tempo, mentre nella Y viene riportata la posizione delle quattro dita sullo strumento. Nella parte sinistra si trova il diagramma che mostra tutte le possibili

² I link ai video sono indicati in coda all'articolo.

³ Approfitto di questa sede per ringraziare ufficialmente Francesco e Guido per la loro preziosa collaborazione.

diteggiature. I quattro cerchi rappresentano, da sinistra a destra, rispettivamente l'indice, il medio, l'anulare e il mignolo della mano. I cerchi sono neri quando il foro è tappato e bianchi quando è aperto. (Fig. 4).

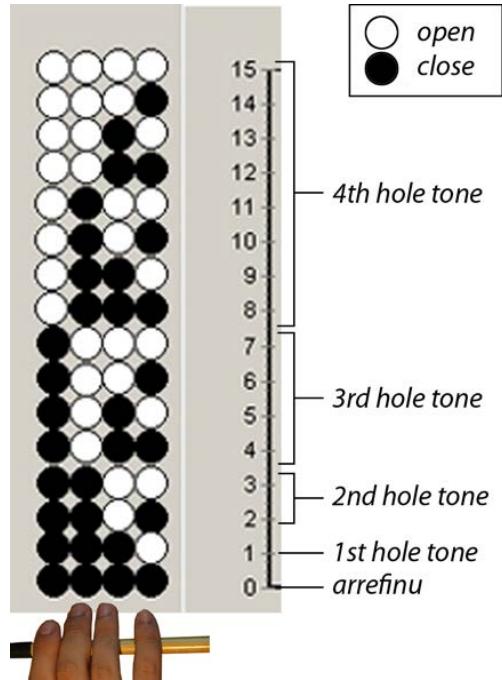


Figura 4. Diagramma della diteggiatura in *e-luanedda*.

Alcune osservazioni preliminari possono essere fatte semplicemente osservando il diagramma. In primo luogo il fatto che per produrre cinque differenti suoni sono possibili sedici diverse diteggiature. Inoltre, il diagramma mostra come la note dell'*arrefinu* e quella del foro più basso possono essere realizzate con una sola posizione (0 e 1), mentre sono possibili due posizioni per la seconda nota (2 e 3), quattro per la terza (da 4 a 7) e otto per la quarta (da 8 a 15).

5. PRIMI RISULTATI

Messo a punto il nuovo dispositivo, grazie alla disponibilità del suonatore di launeddas Andrea Pisu sono state realizzate le prime analisi i cui risultati vengono brevemente presentati nelle pagine seguenti.

5.1 L'illusione dello staccato

A causa dell'utilizzo della respirazione circolare il suono delle tre canne non può essere interrotto. Tuttavia, ascoltando un pezzo per launeddas si percepisce chiaramente un effetto di staccato sia nella *mancosa* che nella *mancosedda*. Questo perché, grazie all'impiego di una particolare tecnica, i suonatori di launeddas sono in grado di creare una illusione acustica che produce l'effetto dello staccato. Come spiega Andreas Bentzon:

La più comune tecnica di esecuzione consiste nell'aprire soltanto un foro alla volta, tenendo chiusi gli altri, così che l'*arrefinu* risuona tra l'una e l'altra nota. In questo modo il suonatore ottiene effetti di pausa e di staccato perfetti, poiché l'*arrefinu*, come già detto, è assorbito dal *tumbu* fino ad essere indistinguibile. [12, p. 26].

Il [video 2](#) mostra la tecnica dello staccato. Nella prima parte viene suonata solo la *mancosedda*, così è possibile sentire come l'*arrefinu* risuona a volume più basso tra le note prodotte dai fori diteggiati. La seconda parte mostra lo staccato eseguito con la mano destra. In questo caso il suono dell'*arrefinu* non può più essere distinto poiché viene mascherato da quello del *tumbu*.

Attualmente sono disponibili in commercio un certo numero di pubblicazioni che riportano trascrizioni di musica per launeddas realizzate sia per scopi analitici [12] che didattici [13,14,15]. Non esiste una norma condivisa per indicare l'effetto dello staccato, che viene trascritto a seconda dei casi come una punto sopra la nota, con l'alternanza tra una nota e una pausa, o ancora alternando la nota prodotta dal foro diteggiato e quella prodotta dall'*arrefinu*. (Fig. 5, in alto). In alcuni casi vengono sovrapposti due pentagrammi: il primo riporta una trascrizione semplificata e il secondo l'effetto percepito all'orecchio. Tutte queste soluzioni, oltre a non essere omogenee, non forniscono alcune informazioni sulla diteggiatura effettivamente utilizzata dal suonatore per ottenere l'effetto di staccato.

Il [video 3](#) mostra come ottenere l'effetto dello staccato con le *eletroneddas*. Osservando il tracciato (Fig. 5, in basso) si vede chiaramente come per ogni nota il suonatore scelga la posizione che richiede l'apertura di un unico foro: posizione 8 per il Mi bemolle, 4 per il Re, 2 per il Do e 1 per il Si bemolle.

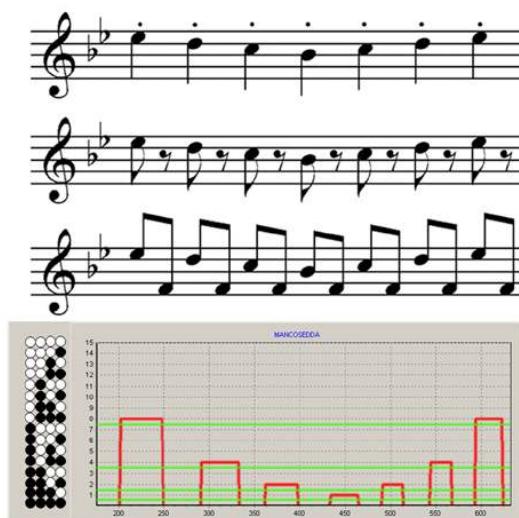


Figura 5. La parte superiore mostra diverse soluzioni grafiche per trascrivere l'effetto dello staccato; la parte inferiore il tracciato prodotto da *e-launedda*.

5.2 Il cambio di *arrefinu*

Un altro effetto che i suonatori possono ottenere grazie a una particolare diteggiatura è il cosiddetto “cambio di *arrefinu*”. Come detto, tranne in alcuni casi⁴, il suono prodotto dall'*arrefinu* non ha una funzione melodica ma viene utilizzato dai suonatori per creare l'illusione della pausa. Ciò è possibile perché l'*arrefinu* ha un volume inferiore rispetto a quello degli altri fori, inoltre è sempre intonato su una delle note dell'accordo maggiore costruito sulla nota del *tumbu*, dalla quale viene mascherato. In alcuni passaggi i suonatori eseguono una melodia sulle note più acute tenendo sollevato il dito mignolo. In questo modo il suono dell'*arrefinu* viene sostituito da quello del primo foro. L'effetto che si ottiene viene così descritto da Bentzon:

A volte si usa una tecnica un po' più complicata, perché si può aprire uno dei fori più bassi della canna mentre si esegue una figura con quelli più alti sollevando un dito alla volta. Si avrà allora l'impressione di percepire la nota del foro tenuto costantemente aperto come una voce estranea, distinta dalla frase eseguita con le dita più in alto [12, p. 27].

La prima parte del [video 4](#) mostra una frase suonata con la sola mano destra. Il suono dell'*arrefinu* si sente per un istante quando il suonatore inizia a soffiare nello strumento. Subito dopo, quando solleva il dito sul primo foro, il suono corrispondente inizia a risuonare, dando l'illusione di persistere e sovrapporsi allo staccato realizzato sulle note più acute.

La seconda parte del video mostra la stessa sequenza con l'aggiunta della mano sinistra. L'effetto finale consiste nella percezione di quattro parti: le frasi realizzate con le due canne melodiche, il bordone del *tumbu* e il “falso *arrefinu*” prodotto con la *mancosedda*. Questo effetto è stato trascritto sul pentagramma con differenti soluzioni grafiche. Le due più comuni sono rappresentate nella parte superiore della Fig. 6.

Il [video 5](#) mostra il cambio di *arrefinu* realizzato con le *eletroneddas*. Il tracciato (Fig. 6, in basso) mostra come questo effetto si ottenga sostituendo al suono dell'*arrefinu* quello del primo foro (posizione 1). Inoltre è interessante rilevare come l'utilizzo di questa tecnica richieda una diversa diteggiatura rispetto all'esempio precedente (Fig. 5). In questo caso il suonatore utilizza la posizione 5 anziché la 4 per ottenere il Re e la 3 anziché la 2 per ottenere il Do. Mancando il punto d'appoggio sul mignolo, ha bisogno di riorganizzare

⁴ Fanno eccezione alcune frasi musicali realizzate con il *cuntzertu* detto *puntu* ‘e òrganu’. Nella *mancosedda* di questo strumento l'*arrefinu* è intonato sul centro tonale, mentre i fori diteggiati producono rispettivamente il secondo, terzo, quarto e quinto grado della scala maggiore. Nelle frasi che partono dalle note più acute e terminano sul centro tonale il suono dell'*arrefinu* viene percepito nella sua funzione melodica anziché come una pausa.

l'assetto di tutta la mano - e dunque di cambiare la diteggiatura - per ottenere la medesima sequenza di note.

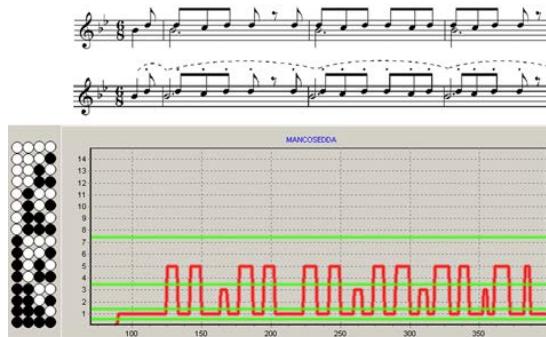


Figura 6. La parte superiore mostra diverse soluzioni grafiche per trascrivere l'effetto prodotto dal cambio di *arrefinu*; la parte inferiore mostra il tracciato in *e-launedda*.

5.3 Staccato e legato

Ascoltando un pezzo di musica per launeddas è possibile sentire la stessa frase musicale (*noda*) eseguita una volta con lo staccato e subito dopo con il legato ([video 6](#)). Ciò che potrebbe sembrare di scarso rilievo è in realtà un aspetto particolarmente importante nell'economia di una suonata, che attiene sia ai principi di organizzazione strutturale sia all'estetica della musica per launeddas. Le suonate di accompagnamento al ballo sono infatti fondate su un principio di sviluppo lineare che prevede la proposta di un determinato materiale melodico e il successivo abbandono. Bentzon sintetizza questo procedimento con il concetto di ‘continuità tematica’:

Sotto l'aspetto musicale un pezzo è giudicato quasi esclusivamente dall'eleganza della sua struttura globale. Ciascun gruppo deve essere ben sviluppato, nessuna noda deve essere ripetuta e il principio di continuità tematica non dovrebbe mai essere infranto. [12, p. 76].

L'efficacia di una suonata per launeddas dipende dunque dal giusto equilibrio tra la realizzazione di ripetizioni variate (nessuna noda deve essere ripetuta) e il principio di continuità tematica:

La struttura globale dei balli professionali è completamente retta dall'ideale estetico della continuità tematica [...] Esso esige due requisiti: a) che non ci siano ripetizioni di nodas in un gruppo, fatta eccezione per la noda principale, che può essere ripresa alcune volte; b) che la differenza tra nodas consecutive sia la minima possibile. [...] In pratica, per restare fedeli a questo ideale per tutta la durata del ballo, ci vuole un vero talento compositivo che soltanto pochi possiedono, dato che molti gruppi di frasi consistono semplicemente di varianti giostapposte di una noda principale. [12, p. 60].

Uno dei meccanismi utilizzati dai suonatori per variare una *noda* consiste nell'eseguire la melodia di cui è composta prima staccata e subito dopo legata. Il [video 7](#) mostra l'esecuzione di una *noda* da ballo con le *elettroneddas*, suonata prima con lo staccato e poi col legato. L'analisi del frammento preso in esame mostra come per ottenere i due effetti siano necessarie due diteggiature differenti. Come già evidenziato (Fig. 7, in alto), nello staccato le dita vengono sollevate una per volta (posizione 8 per il Mi bemolle, posizione 4 per il Re), lasciando risuonare l'*arrefinu* (posizione 0) tra una nota e l'altra. Invece per realizzare il legato è richiesto un movimento assai più complesso (Fig. 7, in basso). Il primo Mi bemolle viene eseguito con tre diverse posizioni: sollevando l'indice (posizione 8), poi il medio (posizione 12) e infine l'anulare (posizione 14). Si passa poi direttamente al Re chiudendo il primo e il quarto foro (posizione 6). La successiva chiusura dell'anulare produce un Do con momentanea funzione di falso *arrefinu* (posizione 2), e si termina con la chiusura di medio e mignolo per realizzare il Mi bemolle (posizione 10).

Dopo essere tornato al Do con funzione di falso *arrefinu* (posizione 2), si ripete l'intera terzina ma questa volta eseguendo il primo Mi bemolle con le posizioni 10 e 14.

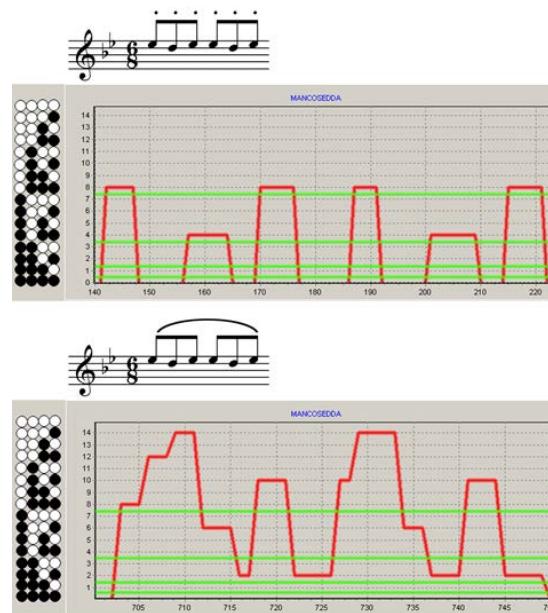


Figura 7. *E-launedda*: in alto la trascrizione su pentagramma e il tracciato dello staccato; in basso la stessa terzina eseguita con il legato.

5.4 Il *sutu 'e manu*

Negli accompagnamenti alla danza per launeddas la mano sinistra ripete spesso un pattern melodico-ritmico che prende il nome di *sutu 'e manu*, letteralmente “sotto

mano” o “mano di sotto”. Nella Fig. 8 la lettera b mostra una possibile realizzazione del *sutu ‘e manu* con due diversi *cuntzertus* entrambi intonati in Si bemolle. Confrontando le trascrizioni si nota come questi siano analoghi dal punto di vista ritmico ma presentino differenze significative dal punto di vista melodico-armonico. Nel primo esempio, realizzato col *fioràssiu*, il pattern insiste sul centro tonale (Si bemolle), mentre nel secondo, *puntu ‘e òrganu*, insiste sul secondo grado, il Do.

Il video 8 mostra il *sutu ‘e manu* suonato con il *fioràssiu*, prima con la sola mano sinistra e successivamente con lo strumento completo. Il video 9 propone la stessa sequenza con il *puntu ‘e òrganu*.

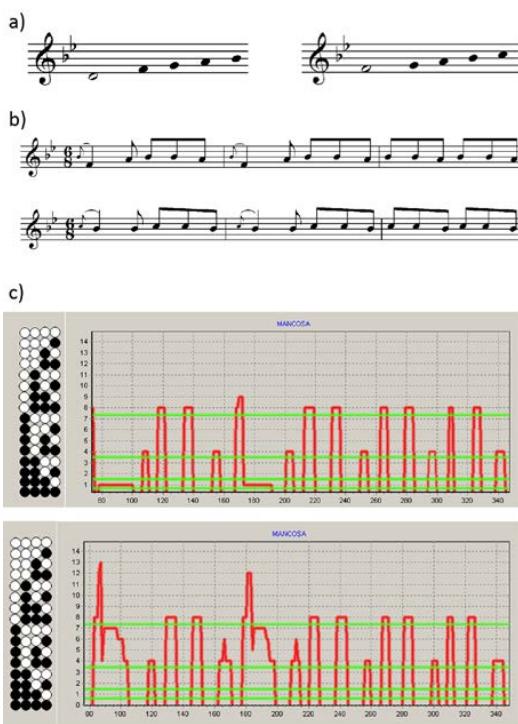


Figura 8. Fioràssiu e puntu ‘è òrganu in Si bemolle a confronto: a) note prodotte dalla *mancosa*, b) trascrizione del *sutu ‘e manu*, c) tracciato del *sutu ‘e manu* con realizzato da *e-launedda*.

Tale differenza, che contribuisce a dare un diverso sapore alle danze accompagnate con i due strumenti, può esser spiegata in termini gestuali più che musicali. I due *sutu ‘e manu* sono infatti basati sullo stesso pattern di movimento che viene riproposto in maniera simile su entrambi gli strumenti. La lettera A della Fig. 8 mostra le note prodotte dai cinque fori dalle *mancosa* di *fioràssiu* (a sinistra) e *puntu ‘e òrganu* (a destra). A parte per l’*arrefinu* (pallino bianco) che nel *fioràssiu* è intonato sul terzo grado (Re) e nel *puntu ‘e òrganu* sul quinto (Fa), nelle due canne le quattro note restanti sono trasla-

te di un grado. Ad essere conservata nel passaggio da uno strumento all’altro non è dunque la sequenza ritmico-melodica ma piuttosto il pattern motorio. Ciò viene confermato dal tracciato di *e-launedda* (Fig. 8, lettera C) nel quale si osserva chiaramente come, a parte lievi differenze, la medesima sequenza di movimento (e diteggiatura) viene riproposta su entrambe le *mancosas*.

6. CONCLUSIONI

La presente ricerca nasce dalla convinzione che per una comprensione più profonda dell’esperienza connessa al fare musica strumentale non si possa prescindere dallo studio del movimento umano inteso come atto necessario alla produzione del suono in un determinato contesto musicale e culturale.

Questo tipo di studi, già intrapresi dai pionieri della disciplina e mai messi da parte durante gli sviluppi della stessa, hanno trovato nuovo vigore in anni recenti grazie a una apertura interdisciplinare e al crescente impiego di tecnologie informatiche che consentono una più raffinata raccolta e analisi dei dati.

In questo caso la relazione tra musica e gesto nelle launeddas è stata indagata con un nuovo dispositivo che consente a una versione modificata delle *eletroneddas* di interfacciarsi con il software *e-launedda*, appositamente sviluppato.

I primi risultati emersi spaziano dalla possibilità di studiare in maniera più dettagliata la tecnica esecutiva dello strumento a una accurata indagine sulle diteggiature. Inoltre, i tracciati di *e-launedda* hanno permesso di isolare alcuni esempi in cui l’organizzazione del materiale ritmico-melodico sembra soggiacere più a logiche di ordine gestuale che meramente musicale.

Ulteriori sviluppi saranno possibili raccogliendo nuovi dati, in primo luogo con l’obiettivo di estendere lo studio del *sutu ‘e manu* ad altri *cuntzertus*, o di confrontare lo stile di Andrea Pisù con quello di altri suonatori di launeddas. Altrettanto interessante sarebbe la creazione di ulteriori dispositivi per l’analisi del rapporto tra musica e gesto in altre tipologie di aerofoni.

7. BIBLIOGRAFIA

- [1] E. Von Hornbostel: “African Negro Music”, *Africa* Vol. 1, p. 49, 1928.
- [2] J. Blacking: “Eight flute tunes from Butembo, East Belgian Congo”, *African Music*, Vol. 1, No. 2, pp. 24-52, 1955.
- [3] J. Blacking: “Patterns of *Nsenga* kalimba music”, *African Music*, Vol. 2, No. 4, pp- 3-20, 1961.
- [4] G. Kubik: “The phenomenon of inherent rhythm in East and Central African instrumental music”, *African Music*, Vol. 3, No. 1, pp. 33-42, 1962.
- [5] G. Kubik: “Pattern perception and recognition in African music”, in J. Blacking e J. Kealiinohomoku

- (a cura): *The performing arts*, Mouton Publishers, The Hague, p. 221-249, 1979.
- [6] J. Baily: "Movement patterns in playing the Herati dutâr", in J. Blacking (a cura): *The Anthropology of the Body*, Academic Press, Londra, pp. 275-330, 1977.
- [7] J. Baily: "Music structure and human movement", in P. Howell, I. Cross e R. West (a cura): *Musical Structure and Cognition*, Academic Press, Londra, pp. 237-258, 1985.
- [8] R. I. Godøy e M. Leman (a cura): *Musical Gestures. Sound. Movement, and Meaning*, Routledge, Londra, 2010.
- [9] G. Fara: "Su uno strumento musicale sardo", *Rivista Musicale Italiana*, Vol. XX, pp. 763-791, 1913 (prima parte) e Vol. XXI, pp. 12-50, 1914 (seconda parte).
- [10] P. Bernardini: "L'aulete di Ittiri", in G. Lallai (a cura) *Launeddas*, AM&D-ISRE, Cagliari-Nuoro, pp. 205-209, 1997.
- [11] M. Lutzu: *Launeddas*, in F. Casu e M. Lutzu (a cura), *Enciclopedia della Musica Sarda*, Voll. 11 and 12, L'Unione Sarda, Cagliari.
- [12] A. F. W. Bentzon: *The Launeddas. A Sardinian folk-music instrument*, Akademisk Forlag, Copenaghen, 1969.
- [13] L. Lai: *Metodo per launeddas*, B&P, Cagliari, 1996.
- [14] C. Mariani: *Suonare le launeddas*, Comune di Quartu Sant'Elena - Associazione Punta Giara, Quartu Sant'Elena - Sant'Anna Arresi, 1999.
- [15] Associazione Sonus de Canna (a cura): *Is Sonus. Costruzione e apprendimento delle launeddas*, Condaghes, Cagliari, 2008.

8. LINKS VIDEO

- Video 1:** <https://youtu.be/rNDvCKfc5yc>
- Video 2:** <https://youtu.be/sYpUSqWI-uk>
- Video 3:** <https://youtu.be/Xq3wFMR22aE>
- Video 4:** <https://youtu.be/q2cWU4t38SQ>
- Video 5:** https://youtu.be/hf90_WPZYU8w
- Video 6:** <https://youtu.be/0qLNV4e7Zfk>
- Video 7:** <https://youtu.be/N7QOaAEIVXI>
- Video 8:** <https://youtu.be/lLadidLE7BA>
- Video 9:** <https://youtu.be/MNjD36HYvkA>

The instrumental analysis of the singing voice: the case of harsh/rough/raspy voices

Paolo Bravi

ITE "Martini" – Cagliari (IT)
pa.bravi@tiscali.it

ABSTRACT

Voice quality has a major importance both as part of a musical style and as a crucial trait of a singer's performative art. Instruments like the Praat program allow to observe and analyze this aspect of vocal performance in detail and in an objective way. In this paper a brief illustration of some example cases taken from diverse musical traditions is proposed with a twofold aim. The first (and narrower) goal is to describe how a particular aspect of the voice quality, usually identified with terms like harshness, roughness or raspiness, can be visualized and analyzed on the basis of acoustic features that are correlated with the perception of harshness in vocal production. The second (and broader) goal is to show that instrumental analysis may be highly beneficial for all those who wish to have a deeper understanding of how the voice is used as an expressive means in musical contexts.

1. INTRODUCTION

For ethnomusicologists, musical transcription of singing has been a central and problematic issue for a long time. Different configurations of scales, intervals, ornaments, rhythms, as well as the impossibility to realistically describe aspects which may be referred to as timbre, make the usual Western way of writing music not completely suitable, effective and affordable for the description of the singing voice, particularly in the case of music based on different codes [1, 2, 3, 4]. For this reason, musical score transcriptions of sung lines are sometimes to be considered with caution or even suspicion. The digital revolution has dramatically changed the set of tools at hand for making music analysis. In particular, modern computer tools help to solve some of the problems related to the use of the score as a method for visualizing music and give the possibility to analyse in detail aspects of the organization of musical sounds and, more generally, to reach a deeper understanding on the acoustic surface of musical performances¹.

In this article I will examine one aspect of the problem of the musical representation of timbre in vocal performances. As is well known, the difficulty of effectively describing voice quality is one of the drawbacks both in the standard musical semiotics and in the phonetic representation of speech (cfr. [10: p. 157-164]). In particular, I will discuss how the digital technologies may help to visualize and analyse voices characterised by harshness. To illustrate the acoustic characteristics of this kind of voice, I will consider some excerpts of both well-known rock and jazz singers and of traditional singers of the Mediterranean area (in particular, traditional singers of the Marches - Italy).

2. ACOUSTIC CORRELATES OF HARSH/ROUGH/RASPY VOICES

The co-occurrence of the three adjectives harsh/rough/raspy in the title above is due to the fact that no standardized agreement is nowadays shared on the terminology to be used to describe those voices that, according to the common sense and use of the language, may be said to have a harsh (or rough, or raspy) sound quality.

In the field of phonetics, as John Laver observes, "small cycle-to-cycle variations in fundamental frequency are associated with voice judged to be *harsh*" [10: p. 127, italic mine]. In particular, in fact, *jitter*, *shimmer* and *harmonicity* (or *HNR*: Harmonics-to-Noise Ratio) are acoustic features correlated with the perception of harshness. The first two are measures of the vocal perturbation: *jitter* refers to small variations in the fundamental frequency; *shimmer* refers to perturbation in the amplitude of cycles. *Harmonicity* refers to the amount of spectral noise in the signal and is measured as the ratio between the energy in the harmonics and in the noise.

In the field of logopedics and phoniatrics, anomalies of this kind relevant to the laryngeal activities are usually considered under the term *roughness*. Both GRBAS [11, 12] and CAPE-V [13, 14], two well known scales commonly used by voice pathologists, consider "roughness" as one of the crucial parameters in the diagnosis of voice disfunctions².

¹ The instrumental approach in music analysis clearly profits from the widespread diffusion of computers and sound analysis software of today. However, the pioneering use of fundamental frequency analysers in the field of music psychology and ethnomusicology is time-honoured [5, 6, 7, 8, 9].

² An overall survey on the principal standard adopted in the analysis of voice pathology is in [15].

In the field of ethnomusicology, one of the parameters used in Alan Lomax's *cantometrics* to describe the vocal features of singing is defined by the word *raspiness*, but the same authors remark that the words "throatiness" or "harshness" could have also been used [16: p. 73-74].

In this paper, I will not enter into the details of the semantic aspects regarding these terms or other pertaining to the same area (I am referring to adjectives like hoarse, croaky, grating, raucous, throaty etc.). For investigation on the value commonly associated with the terminology used – by singers, teachers, critics, sound engineers, etc. – in order to refer to this and other type of voice quality I refer here to [17, 18, 19] and to the references thereof. Instead, I will assume the usual terminology in its basic and common meaning and I will show how the perceived quality of this type of vocal sounds can be analyzed and visualized using an instrumental approach. Analyses and graphical representations presented here are made via the Praat program [20]. Some references will also be made to the MDVP *Multi Dimensional Voice Program*, a software created by KayPENTAX and commonly used in the field of logopedics and phoniatrics for the assessment of voice pathology [21]. Both programs allow for different type of measures of the features relevant to the analysis of voice quality.

Of course, when we pass from the field of phoniatrics to that of musicology, we have to consider that in the case of sung voice, harshness / roughness / raspiness are in many cases not a form of voice disorder or a pathology, but a distinctive and sometimes highly appreciated and remarkable characteristic of particular singing styles, the 'vocal signature' of specific singers, or an expressive trait in particular vocal passages. To possess this kind of vocal quality – or to be able to produce it when required – is an essential gift of a talented singer in some particular musical styles. This is the case of many renowned rock and flamenco singers, to name just a couple of well known example.

The perturbations of the periodicity are revealed by a close observation of the waveform. Two excerpts from recordings of the jazz singers Ella Fitzgerald and Louis Armstrong are used in figure 1 as examples to show the differences in the shape of the waveform and in the spectrogram in short snippets where their voices are not accompanied by instruments and where the two singers' voice quality rapidly go from clear to rough (in the snippet from Ella Fitzgerald) and viceversa (in the one from Louis Armstrong)³. A visual inspection of the waveform zoomings shows clear anomalies in the periodicity. In particular, the local shimmer, i.e. the cycle-to-cycle variation of amplitude, is different in the rough and the clear parts of both snippet pairs.

³ The excerpts are respectively taken from recordings of the songs "One note samba" (Ella Fitzgerald, available at <https://www.youtube.com/watch?v=PbL9vr4Q2LU>) and "Black and blue" (Louis Armstrong, available at <https://www.youtube.com/watch?v=YGiUi2oir3I>) [retrieved 31.07.2016].

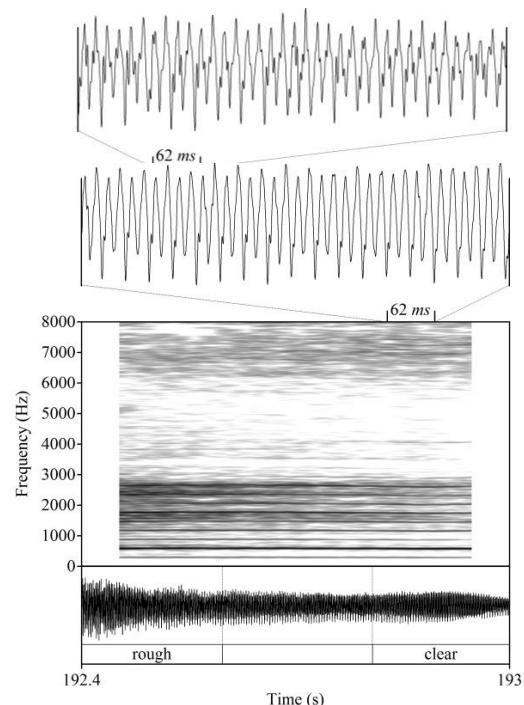
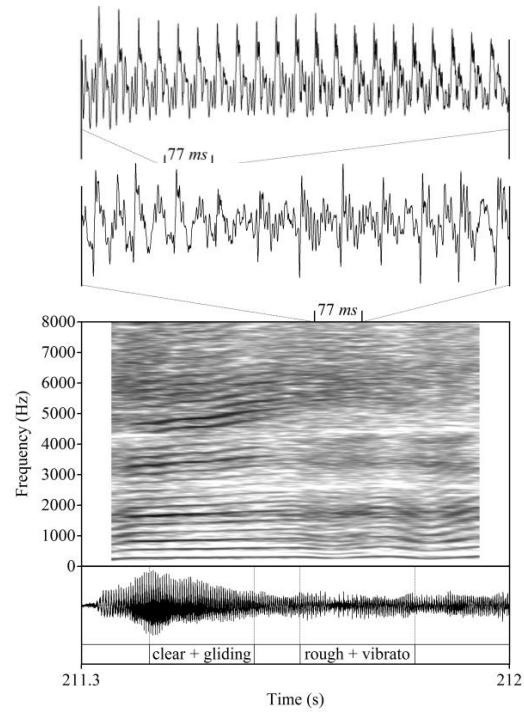


Figure 1. Two short excerpts from solo voice parts in recording of Ella Fitzgerald (top panel) and Louis Armstrong (bottom panel). In each panel, from the bottom: waveform and Praat textgrid; spectrogram (window: 0.1 s); two waveform zoomings.

Instrumental analysis allows to differentiate among various voice qualities by means of a number of measures relevant to the acoustic features jitter, shimmer and harmonicity (see Table 1)⁴.

Parameter	Voice quality	
	Rough	Clear
<i>Jitter</i> (local)	EF LA	5.883% 1.785%
	EF LA	0.287% 0.310%
<i>Shimmer</i> (local)	EF LA	14.180% 10.192%
	EF LA	11.017% 3.588%
<i>Harmonicity</i> (mean HNR)	EF LA	4.898 dB 11.277 dB
		16.800 dB 22.884 dB

Table 1. Values of jitter and shimmer (local) and mean HNR in the pair of snippets from Ella Fitzgerald and Louis Armstrong recordings with different voice quality.

Perturbations of the periodicity may also be observed in excerpts taken from recordings of rock singers characterized by harsh/rough/rasp voice. Figure 2a refers to a snippet from a performance of Ian Gillan where the singer performs a very high note through his peculiar voice⁵. The spectrogram zooming shows that cycles have uneven amplitude. The case presented in figure 2b refers to a snippet of a performance of Joe Cocker, another rock singer whose voice is well known for its peculiar harsh quality, especially on the highest register, as happens in this case⁶. Both the spectrogram and the waveform show the characteristics of a voice wherein the noisy part prevails over the harmonic one.

3. ANALYSIS AND VISUALIZATION OF HARSH/ROUGH/RASPY VOICES

Voice quality is, in many cases, the most striking feature of a singing voice. As Marco Beghelli puts it, while listening to a singing voice «l'interesse fonico è [...] riposto non tanto nella melodia in sé, quanto nel suo inveramento canoro, per il quale l'apporto

⁴ Reported values in Table 1 are been measured through the program Praat and are relevant to one measure only for each features. Further measures of jitter, shimmer and harmonicity, based on different criteria, may be computed (see “Voice” in the Praat online manual). MDVP gives the values of 1.040% for *jitter (local)* – in MDVP called *Jitt* – and 3.81% - in MDVP called *Shim* – for *shimmer (local)* as thresholds for voice pathology.

⁵ The excerpt is taken from a recording of the song “Strange kind of woman” of the rock group Deep Purple (available at https://www.youtube.com/watch?v=01-2-7_IRFA [retrieved 31.07.2016]). Values correlated to the harsh/rough/raspy voice quality – as reported in Table 1 – are the following: *jitter (local)* = 2.315%; *shimmer (local)* = 15.963%; *harmonicity (mean HNR)* = 10.706 dB.

⁶ The excerpt is taken from a recording of the song “With a little help from my friends” (available at https://www.youtube.com/watch?v=_wG6Cgmgm5U [retrieved 31.07.2016]). Values correlated to the harsh/rough/raspy voice quality – as reported in Table 1 – are the following: *jitter (local)* = 9.386%; *shimmer (local)* = 21.634%; *harmonicity (mean HNR)* = -0.675 dB.

dell’interprete, con la specifica “grana” vocale, si fa – più che prioritario – ineludibile» [22: p. 833]. Singing analysis should be able to appropriately consider this aspect. Instead, focusing primarily in melody, harmony and rhythm and essentially excluding the timbre dimension, standard music analysis is in most cases inevitably partial and unsatisfactory.

Instrumental analysis may help to overcome this problem and to avoid the well known limitations bound to the common representation of music through staff notation.

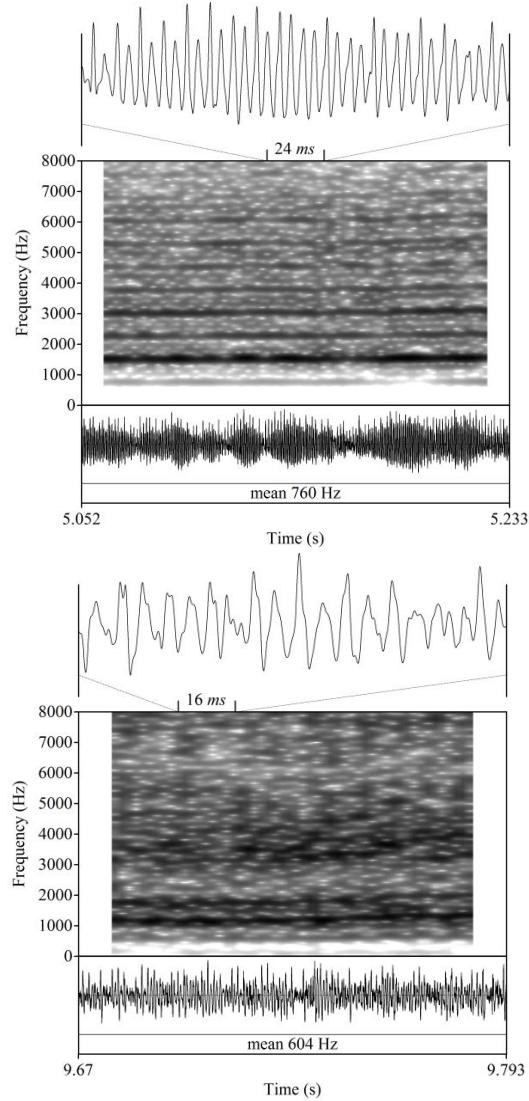


Figure 2. Two short excerpts from solo voice parts in recording of Ian Gillan (2a: top panel) and Joe Cocker (2b: bottom panel).

The parameters of jitter, shimmer and harmonicity may be captured according to the analytic needs and subjected to analysis and comparison. The two plots in figure 3 are taken from a recent research on the voices of three

traditional singers of the Marches (Italy), Concetta Piccioni Vecchioni (here CPV), Lina Marinozzi Lattanzi (LML) and Nazzareno Saldari (NS) [23]. The characteristic harsh quality of LML voice shows levels of jitter and shimmer, as well as a f_0 mean significantly higher than those of the other two singers.

An effective visual representation of this acoustic feature may be obtained by associating the variation in time of the values of shimmer – corresponding to the

perception of harshness – to the profile of the fundamental frequencies. A visual comparison of the three profiles in figure 4, which are relevant to single lines sung in the performances of the three traditional marchigian singers, shows the almost constant presence of a high level of shimmer along the sung line of LML, which corresponds to the typical harshness that differentiate her voice quality from those of the other two singers.

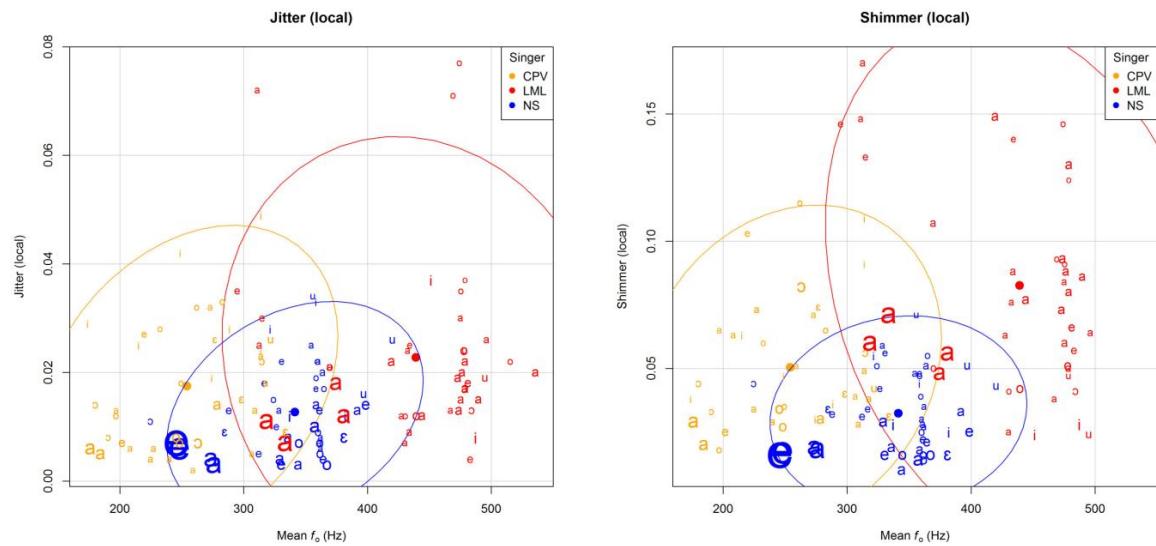
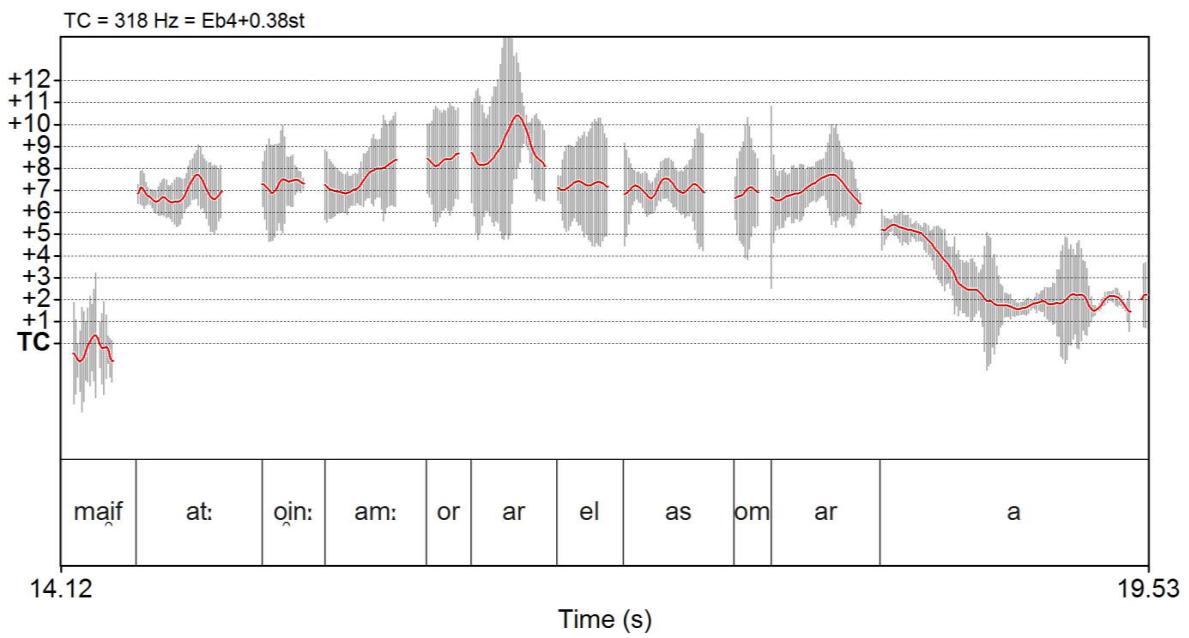


Figure 3. Plots with data relevant to mean values of frequency and *jitter* (3a: left panel) and *shimmer* (3b: right panel) in three solo recording by the traditional singers Concetta Piccioni Vecchioni [CPV, orange], Lina Marinozzi Lattanzi [LML, red] and Nazzareno Saldari [NS, blue]. The size of the IPA phonetic symbols representing each point in the plot varies according to the duration of the sung vowels.



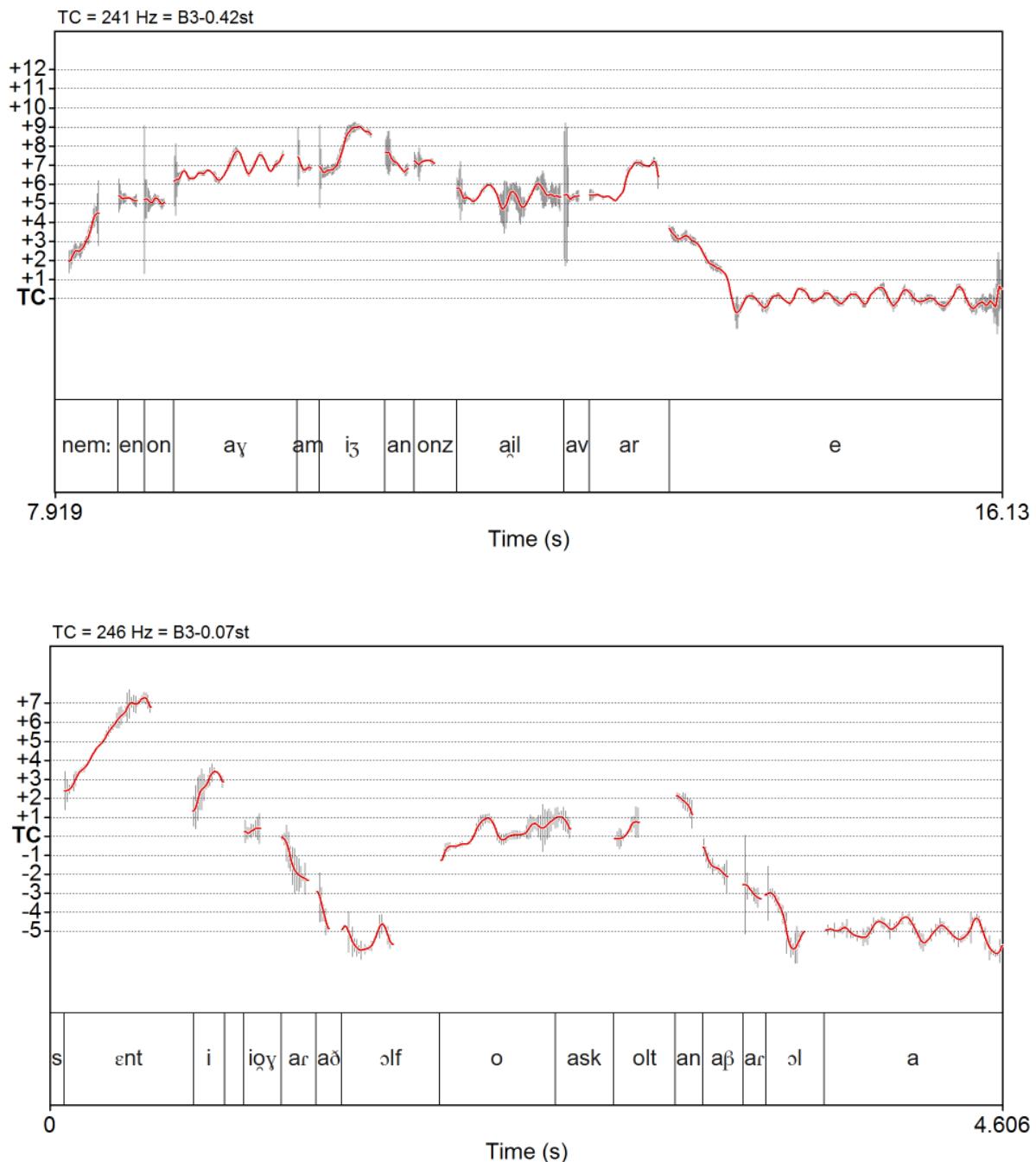


Figure 4. f_0 tracks of sung lines taken from three solo voice songs of different traditional singers of the Marches (Italy): Lina Marrazzo Lattanzi (4a: top panel [in the preceding page]; line 3); Nazzareno Saldari (4b: middle panel; line 2); Concetta Piccioni Vecchioni (4c: bottom panel; line 1). The contour line (red colour) is surrounded by a band of variable width (grey colour) which displays, in proportion, the level of shimmer over time.

4. VOICE MANIPULATION: CHANGING VOICE QUALITY BY ADDING JITTER, SHIMMER AND NOISE

Computers may be helpful not only to analyze timbre effects in existing recordings of vocal music, but also to modify – for the sake of listening experiments and/or for creative aims – existing vocal sounds by adding jitter, shimmer and noise to natural or artificial voices. As a short illustration of this area of applicability of computer to music research, figure 5 shows a visual comparison between three pairs of waveforms, observed very closely. In each pair, one can see the waveform of the original sound (in this case created via synthesis) and its manipulation realized through a Praat script which adds to it a certain (arbitrarily set) amount of jitter (5a), shimmer (5b) and noise (5c). Appropriate manipulation of vocal sound may serve both to scientists who need to construe stimuli for listening experiments with a precise control of the variables under examination, and to composers and sound engineers interested in using altered voices for artistic goals.

5. CONCLUSION

In the field of ethnomusicology, musical score transcription has been the main tool-of-the-trade for music analysis despite the acknowledged cultural bias and the limitations of staff notation, particularly as a means to represent music conceived and performed outside of the Western music culture. In the digital era, a number of analytical tools have been designed which may serve to solve some of the problems related to the use of the score as a method for describing and visualizing music. These tools allow to carry out analyses that were virtually impossible until a few decades ago. *Praat*, a software developed by Paul Boersma and David Weenink and designed for phonetic studies, may also be helpful for musical aims, in particular as far as the analysis, visualization and manipulation of singing voices are concerned. This paper has illustrated some of the possibilities provided by this program for a detailed description and analysis of one important aspect of voice quality in singing, i.e. harshness / roughness / raspiness, that is in most cases neglected or evaluated only in a pure impressionistic way by musicologists dealing with the voice as an expressive means in the musical field.

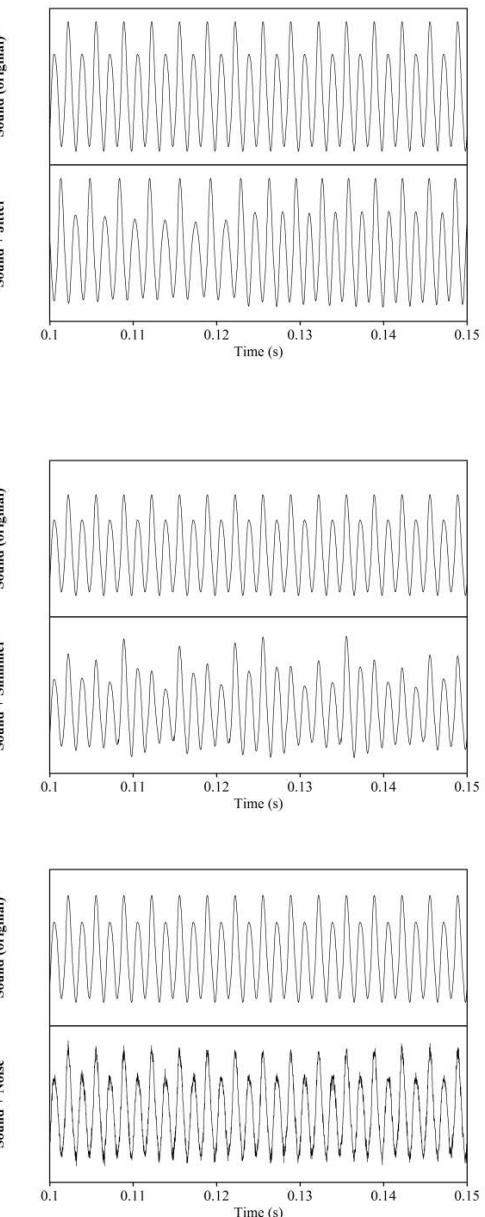


Figure 5. Waveform pairs of original sound (created by synthesis) and different kind of manipulations realized adding jitter (5a: top panel), shimmer (5b: mid panel) and noise (5c: bottom panel).

6. REFERENCES

- [1] D. Stockmann: "Il problema della trascrizione nella ricerca etnomusicologica," *Etnomusicologia. Seminari internazionali di etnomusicologia*, Siena, Accademia Musicale Chigiana, 1989, pp. 209-238.

- [2] T. Ellington: "Notation," in *Ethnomusicology: an Introduction*, H. Myers (ed), London, The MacMillan Press, pp. 153-164, 1992.
- [3] B. Nettl: *The Study of Ethnomusicology: Thirty-one Issues and Concepts*, University of Illinois Press, Champaign (Illinois), 2005.
- [4] M.-P. Baumann, J. P. J. Stock and M. Marian-Bălașa (eds), "Notation, Transcription, Visual Representation," *The World of Music*, Vol. 2, 2005.
- [5] C. E. Seashore, "The tonoscope," *The Psychological Monographs*, Vol. 16, No. 3, pp. 1-12, 1914.
- [6] M. Metfessel, *Phonophotography in Folk Music: American Negro Songs in New Notation*, University of North Carolina Press, Chapel Hill, 1929.
- [7] C. Seeger, "An Instantaneous Music Notator," *Journal of the IFMC*, No. 3, pp. 103-107, 1951.
- [8] O. Gurvin, "Photography as an aid in folkmusic research," *Norveg*, Vol. 3, pp. 181-196, 1953.
- [9] K. Dahlback, *New methods in vocal folk music research*, Oslo University Press, Oslo, 1958.
- [10] J. Laver, *The Phonetic Description of Voice Quality*, Cambridge University Press, Cambridge, 1980.
- [11] N. Isshiki, M. Olamura, M. Tanabe and M. Morimoto, "Differential Diagnosis of Hoarseness," *Folia phoniatrica*, Vol. 21, No. 1, pp. 9-19, 1969.
- [12] M. Hirano, *Clinical Examination of Voice*, Springer Verlag, New York, 1981.
- [13] ASHA [American Speech-Language-Hearing Association], *Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)*. [Online, Available at: http://www.asha.org/uploadedFiles/ASHA/SIG03/CAPE-V-procedures.pdf?_ga=1.93405981.1485057542.1473103950#search=%22Consensus%22 - Accessed 31.07.2016], 2006.
- [14] G. B. Kempster, B. Gerratt, K. Verdolini Abbott, J. Barkmeier-Kraemer and R. Hillman, "Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol," *American Journal of Speech-Language Pathology*, Vol. 18, No. 2, pp. 124-132, 2008.
- [15] P. N. Carding, E. Carlson, R. Epstein, L. Mathieson and C. Shewell, "Formal perceptual evaluation of voice quality in the United Kingdom", *Logopedics & Phoniatrics Vocology*, Vol. 25, No. 3, pp. 133-138, 2000.
- [16] A. Lomax and V. Grauer, "The Cantometric Coding Book," in *Folk Song Style and Culture*, Transaction Publishers, New Brunswick-London, pp. 34-74, 1968.
- [17] M. Garnier, N. Henrich, M. Castellengo, D. Sotiropoulos and D. Dubois, "Characterisation of Voice Quality in Western Lyrical Singing: from Teachers' Judgements to Acoustic Descriptions," *Journal of interdisciplinary music studies*, Vol. 1, No. 2, pp. 62-91, 2007.
- [18] N. Henrich, P. Bezard, R. Expert, M. Garnier, C. Guerin, C. Pillot, S. Quattrochi, B. Roubeau and B. Terk, "Perception and verbalisation of voice quality in western lyrical singing: Contribution of a multidisciplinary research group," *3rd Conference on Interdisciplinary Musicology, CIM07, Aug 2007, Tallinn, Estonia*. 2007, Tallinn - ESTONIA, 2007.
- [19] S. Morange, D. Dubois and J.-M. Fontaine, "Voices of Caruso: Cognitive evaluation and acoustic analysis of reedited editions," *Proceedings of the Third Conference on Interdisciplinary Musicology (CIM07)*, 2007.
- [20] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. [Online Available at: <http://www.fon.hum.uva.nl/praat/> - Accessed 15.04.2016], 2016.
- [21] M. K. Christmann, A. R. Brancalion, C. Ramos de Freitas, D. Zacarias Vargas, M. Keske-Soares, C. Lisbôa Mezzomo and H. Bolli Mota, "Use of the program MDVP in different contexts: a literature review," *CEFAC*, Vol. 17, No. 4, pp. 1341-1349, 2015.
- [22] M. Beghelli, "Voci e cantanti," in *Enciclopedia della musica. Vol. V: L'unità della musica*, Einaudi, Torino, pp. 814-841, 2005.
- [23] P. Arcangeli, G. Pietrucci and P. Bravi, "Ricordando la figura di Lina Marinuzzi Lattanzi," *Quaderni Musicali Marchigiani*, [in press].

Software di analisi etnomusicologica: modelli di pensiero

Lorenzo Vanelli

Università di Bologna, Dipartimento di Arti
Visive, Performative, Mediali
lorenzo.vanelli3@unibo.it

ABSTRACT

Oggi esistono diversi programmi in grado di svolgere un'analisi approfondita di una fonte audio, producendo su di essa dati di varia natura. In questo articolo si discute su quali livelli, epistemologico, metodologico e qualitativo, l'uso di questi software può portare valore aggiunto al processo di ricerca etnomusicologica. La questione verrà discussa prima di tutto partendo da uno sguardo generale, riassumendo le caratteristiche di alcuni strumenti storicamente e recentemente utilizzati per svolgere queste operazioni. In seguito si discuterà la questione in particolare, proponendo come caso di studio l'analisi di alcune peculiarità degli *holler* afroamericani, canti di interesse etnomusicologico sui quali l'autore sta svolgendo ricerche di approfondimento ancora in corso. Si proverà a ragionare quindi sul grado di oggettività o soggettività dell'*output* di un software e sul tipo di lavoro che con esso può essere svolto. Infine, si tenterà di focalizzare il discorso soprattutto sulle caratteristiche di quell'apporto che l'uso di software specializzati aggiunge ai metodi di ricerca. L'obiettivo non è quindi quello di discutere il programma o il caso di studio in sé, ma le peculiarità del rapporto interattivo generato dal confronto tra il ricercatore, impegnato nel percorso di approfondimento e interpretazione dei materiali, e lo strumento di analisi di cui usufruisce.

1. INTRODUZIONE

Il lavoro dell'etnomusicologo è parzialmente cambiato in concomitanza con l'invenzione di software informativi in grado di analizzare una traccia audio.

Non si tratta di un'invenzione: metodi di traduzione analogica di una fonte audio in elementi grafici sono stati utilizzati fin dai primi del Novecento. In realtà il cambiamento è più profondo e radicale, ed ha a che fare con la natura stessa dello strumento in questione. L'uso di un software comporta, rispetto al passato, delle nette differenze sul piano dei dati prodotti e su quello dell'approccio mentale, da parte del ricercatore, al processo di analisi stesso.

Copyright: © 2016 Lorenzo Vanelli. 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.

I dati estratti da una traccia audio, rispetto al passato, sono di più, e di qualità migliore. I software permettono di rallentare, sezionare, misurare i brani e le loro componenti ritmiche, melodiche, persino timbriche. Su ogni piano di analisi il programma è in grado di produrre grandi moli di dati, successivamente passabili di agevole confronto, schematizzazione, discussione.

I software per analisi di tracce audio sono molti. Per questo breve intervento farò riferimento alle proprietà di Sonic Visualiser, un programma prodotto dall'Università Queen Mary di Londra;¹ tale programma condivide con gli altri una natura fortemente modulare. Ma quale relazione intercorre tra la modularità di questi software e il tipo di pensiero che sostiene e guida il ricercatore durante l'analisi di un brano?

2. CORPO, STRUMENTO, MENTE

2.1 Limiti Fisici

L'orecchio umano è uno strumento di indagine potente, ma imperfetto. Esistono alcuni limiti fisici alla nostra capacità di ascolto: tra i più noti, le barriere delle frequenze di udibilità e l'incapacità di riconoscere la presenza di una frequenza ad ampiezza ridotta se molto vicina ad un'altra ad ampiezza più elevata. L'orecchio umano ha inoltre alcune difficoltà a rilevare persino il volume reale di un suono: al variare della frequenza fondamentale di quel suono, l'orecchio tende a percepire un grado di pressione sonora differente.²

L'orecchio umano è inoltre facilmente portato all'errore anche a causa dell'insieme delle nostre abitudini uditive. È stato dimostrato³ con esperimenti di psicologia che in condizioni normali fatichiamo a riconoscere pattern differenti rispetto a quelli a cui siamo abituati. Anzi: tendiamo quando possibile a ricondurre il diverso al già noto: è il motivo per cui saremmo normalmente più disposti a pensare che un cantante sia stonato piuttosto che immaginare che sia intonato su un sistema differente di rapporti scalari rispetto al nostro sistema temperato.

Certamente seguire ciò che ci consiglia l'orecchio, nella vita di tutti i giorni, è essenziale: ci aiuta a mu-

¹ Per maggiori informazioni sul programma vedi [4].

² Per una introduzione ai problemi di acustica musicale, e per maggiori riferimenti bibliografici, consiglio [26].

³ Per una introduzione generale ai più recenti sviluppi della psicologia della mente, e per maggiori riferimenti bibliografici, consiglio di partire da [16].

verci con consapevolezza nel paesaggio sonoro che ci circonda, a interpretarlo e dargli significato, quindi ad interagire con esso. Fidarsi delle proprie abitudini uditive non è però ammissibile quando bisogna cimentarsi nell'interpretazione di musiche basate su sistemi anche solo parzialmente diversi dal proprio: in tal caso gli schemi abitudinari diventano un intralcio, e la complessità del reale si fa ancora più criptica.

Interpretare sistemi musicali diversi dal proprio richiede quindi un'operazione complessa: si tratta di riconoscere le proprie abitudini uditive, lottare con esse e cercare di ampliarle a nuove prospettive; capire quali sono gli elementi e gli schemi su cui si articola il discorso musicale altrui e cercare di farli propri. Questo procedimento può essere semplificato e, anzi, coadiuvato dall'uso consapevole di alcuni strumenti informatici.

2.2 Strumenti di Analisi

Seguire il flusso di un evento musicale richiede di investire l'interessa del proprio corpo nell'atto. Non si ascolta solo con le orecchie, ma ci si coordina con l'atto sonoro anche con il corpo: annuendo con il ritmo, battendo le mani, muovendo gli arti, cambiando respirazione. L'ascolto è quindi a pieno titolo una tecnica corporea.

Che l'apporto strumentale comporti un ampliamento delle tecniche corporali non è certo una novità. L'antropologia delle culture materiali ha già da tempo ragionato sul modo in cui gli oggetti possono diventare, attraverso l'uso, parte integrante del nostro corpo. Vi sono, in proposito, due utili esempi.⁴ Uno è quello del non vedente, la cui percezione si lega indissolubilmente al suo bastone, prolungamento sensibile del braccio. L'altro è quello dell'autista alla guida di un'automobile. Warnier [27] lo riassume così: “un buon automobilista [...] è qualcuno che ha dilatato la sua sintesi corporale all'insieme della vettura o, ancora, che ha incorporato la dinamica della vettura alle sue proprie condotte motrici.”

La notazione musicale è, in musicologia, uno strumento. Attraverso lo studio delle partiture è possibile “dilatare la propria sintesi corporale” al contenuto di un'opera musicale. Questa operazione, nel caso delle musiche afferenti alla tradizione occidentale scritta, trova fondamento nel fatto che sono state esse stesse concepite dagli autori all'interno di un sistema di pensiero che fa della notazione su pentagramma e del lavoro attraverso la partitura un luogo privilegiato di composizione. In tal senso, musicologi e compositori, nel riferirsi alla partitura, parlano la stessa lingua, si esprimono e ragionano attraverso lo stesso schema concettuale, le cui peculiarità sono il frutto di una sedimentazione storica di convenzioni.

Uno dei nuclei fondamentali di questo schema ne prova le peculiarità. Si tratta del concetto, pratico tanto quanto analitico, di “nota”, ovvero un suono ad una certa altezza che si sviluppa per una certa durata, calco-

late secondo procedimenti di semplificazione matematica del reale: la prima corrisponde ad una posizione in hertz rispetto ad una scala temperata, la seconda ad una suddivisione in frazioni semplici del tempo unitario di una battuta. Al netto delle acciaccature, trilli, vibrati, glissati, e così via, comunque intesi come elementi aggiunti ad essa, la nota ideale rimane sensibilmente stabile ad una certa altezza per un certo tempo.

Un concetto come quello appena descritto è chiaramente funzionale ad alcuni scopi analitici, ma non a tutti. Essere consapevoli delle potenzialità e dei limiti degli strumenti che si usano è il primo passo per usarli al meglio.

2.3 Sensibilità e Approssimazione

Esiste un concetto fondamentale nel campo delle scienze naturali, ed è quello di “sensibilità dello strumento”. Per raccogliere dati è necessario usare lo strumento giusto per l'oggetto che vuol esser misurato, sia come grandezza misurata (un metro da sarta non è adatto per misurare una corrente elettrica), sia come sensibilità e portata (un metro da sarta non è adatto per misurare le dimensioni dei componenti di un microprocessore).

Usare strumenti di misurazione significa lavorare ad una modellizzazione del reale, ovvero alla riduzione della sua esacerbante complessità a strutture più semplici, afferrabili, comunicabili. Nessuno si sognerebbe di dire alla sarta che per misurare la stoffa ci vuole per forza un calibro: il suo strumento è perfettamente adatto allo scopo. L'importante è tenere a mente le peculiarità dell'operazione di scelta dello strumento stesso. In questo senso, trascrivere su un pentagramma un brano appartenente ad una tradizione non scritta può avere le sue utilità: permette di ridurre un brano complesso ad alcuni dei suoi aspetti, in particolare permette di ragionare in modo rapido e efficace sulle proprietà generali delle linee melodiche, sui rapporti armonici tra le voci o gli strumenti. Al tempo stesso, però, la sensibilità limitata dello strumento di trascrizione obbliga a tralasciare tanti piccoli aspetti apparentemente minori di un evento acustico complesso.

Una partitura non può dar conto direttamente del timbro, che è lasciato all'immaginazione. Inoltre, seppure esistano segni grafici che cerchino di sopperire a questo tipo di limitazioni, una partitura non può dare del tutto conto delle caratteristiche di intonazione dei suoni, non sempre in perfetto accordo con la nostra scala temperata. Infine, anche sfruttando una scrittura mensurale, è facile che qualcosa delle peculiarità ritmiche del brano originale vada a perdere.

2.4 Phonophotography

Nel 1908 Dayton Miller [11] produsse il primo prototipo di un macchinario in grado di convertire analogicamente una fonte audio in un grafico, sfruttando la lettura delle oscillazioni di una luce riflessa su uno specchio a sua volta messo in vibrazione da onde sonore. Seppure si trattasse ancora di una procedura piuttosto costosa e laboriosa, iniziò quasi subito a dare dei risultati di rilevante interesse accademico. Una delle prime pubblica-

⁴ Entrambi riassunti nell'introduzione a [27], a cui tra l'altro rimando qualora qualcuno fosse interessato ad espandere sull'argomento, assieme ai volumi [9, 10].

zioni che facevano ampio uso di tale sistema di trascrizione visuale delle tracce audio fu *Phonophotography in Folk Music* di Milton Metfessel e Carl Seashore [14]. Da allora il metodo è stato utilizzato in innumerevoli pubblicazioni scientifiche, la cui disamina richiederebbe altri spazi ed altri tempi.

E' però solo con l'arrivo della rivoluzione informatica che il processo di traduzione in veste grafica di tracce audio è diventato più immediato e approfondito. Prendiamo ad esempio il programma Sonic Visualiser: le potenzialità e funzioni del programma sono numerose, perciò mi limiterò ad indicare quelle inerenti al nostro caso di studio.⁵

Una volta caricato un file audio, un sistema a *layers* permette di aggiungere in sovrapposizione nuovi livelli di lavoro, automaticamente sincronizzati lungo l'asse orizzontale del tempo. Tra questi *layers* il più utile ai fini del nostro caso di studio è il *melodic range spectrogram*: tale spettrogramma è composto da un grafico con tempo t (lineare) sull'asse delle ascisse e frequenza f (lineare o logaritmica, a scelta)⁶ sull'asse delle ordinate. Un diverso livello di colore segnala l'intensità della frequenza componente nel punto (t,f) corrispondente della registrazione.

I parametri di scala sono regolabili a piacere. Si noti che il programma permette anche alcune manipolazioni dell'immagine per superare eventuali difficoltà operative. Un problema tipico, lavorando con documenti sonori datati, è legato al pessimo stato di conservazione di certe registrazioni della prima metà del '900: in esse tutta la zona delle frequenze fondamentali dei suoni è completamente coperta dal rumore di fondo dovuto all'usura del disco di supporto originale. Se la frequenza fondamentale non è più visibile (o lo è solo in parte), è in genere però possibile ricorrere alle tracce degli armonici superiori, il cui profilo è esattamente lo stesso del suono principale, permettendo così lo stesso di svolgere considerazioni analitiche.

I valori dei parametri risultanti dalle analisi possono essere esportati in una serie di tipi di *output*. È possibile salvare una sezione dell'immagine su cui si sta lavorando in un file *png* la cui altezza in *pixel* è fissa, permettendo eventualmente un rapido accostamento di immagini differenti. Inoltre, tutti i *layers* contenenti oggetti possono essere esportati in un file dati ed utilizzati su altri programmi di elaborazione. Per lavorare sulle immagini prodotte dal programma ho prodotto, assieme ad Alberto Malagoli, ingegnere informatico dell'Università di Modena, un programma in grado di aggiungere una griglia personalizzabile alle immagini *png*,⁷ grazie alla quale ragionare più agevolmente sulle altezze dei suoni

⁵ Tra le altre funzionalità, segnalo anche un layer che permette di posizionare dei marcatori in istanti di tempo notevoli nell'incedere della registrazione, e un utile strumento di modifica della velocità di riproduzione, che permette un accurato ascolto dei più minimi dettagli ritmici.

⁶ Ai fini di uno studio musicologico, una visualizzazione con scala logaritmica sull'asse y dà un risultato visivamente più intuitivo, eliminando l'effetto di distorsione dato dall'aumento esponenziale della frequenza con il crescere del suono.

⁷ Il programma è gratuitamente utilizzabile all'indirizzo [29].

rispetto alla tipica scala temperata, e ottenere informazioni sulle distanze tra essi in *cent*.

E' evidente che le funzionalità del programma sono molteplici. E' un buon strumento di indagine? Misuriamolo su un caso esemplare.

3. HOLLERS AL MICROSCOPIO

3.1 La Documentazione

Gli *hollers* sono dei canti a voce sola, a tema primariamente profano, in uso tra cantori strettamente afroamericani nel Sud degli Stati Uniti fino alla metà del secolo scorso. Erano canti utilizzati durante le pause dal lavoro o per comunicare a distanza nei campi, ed erano mezzi di espressione e condivisione dei pensieri, delle sensazioni, dei timori esistenziali, in generale della situazione del cantore stesso. Probabilmente di largo uso nel mondo relativamente chiuso delle comunità afroamericane impiegate nei lavori agricoli prima dell'abolizione della schiavitù,⁸ scomparirono assieme al contesto economico in cui venivano usati a causa delle migrazioni verso nord, della meccanizzazione dei lavori agricoli e dei profondi cambiamenti sociali avvenuti nel periodo delle due guerre mondiali. Rimasero però in uso, non a caso, in ambienti che per certi versi riproducevano la situazione ottocentesca: nelle State Farms⁹ e nei Levee Camp,¹⁰ dove continuaron a trovare un contesto, per così dire, favorevole al loro utilizzo fin verso gli anni '60. Questi canti vennero registrati per la prima volta all'inizio del secolo scorso da H. W. Odum,¹¹ su dischi andati perduti. La maggior parte della documentazione su questi canti viene dai viaggi di ricerca di John e Alan Lomax, che li registrarono a più riprese, dagli anni 30' ai primi anni '60.¹²

⁸ La documentazione antecedente all'abolizione della schiavitù riguardante le pratiche private e profane delle comunità afroamericane è a dir poco parca di informazioni. Alcune informazioni frammentarie si possono derivare dalle interviste agli ex-schiavi del *Federal Writers' Project*, in cui effettivamente si fa riferimento all'uso di "hollers" all'interno delle comunità. Tutta la documentazione del *Federal Writers' Project* è depositata alla Library of Congress di Washington, e consultabile online all'indirizzo [31].

⁹ Le State Farms erano prigioni di Stato dove i carcerati erano costretti a lavorare ampi appezzamenti terrieri in circostanze non dissimili da quelle delle piantagioni del secolo precedente. La più famosa tra esse era denominata Parchman Farm. Per approfondire l'argomento, vedi gli articoli di Tannenbaum reperibili gratuitamente all'indirizzo [30].

¹⁰ I Levee Camp erano campi di lavoro per la costruzione degli argini dei fiumi, in particolare quelli del Mississippi. In questi luoghi la manodopera nera veniva asservita economicamente attraverso un sistema di indebitamento nei confronti dei proprietari bianchi che gestivano la manutenzione dell'argine. Per una descrizione di prima mano dell'ambiente dei Levee Camp si veda l'intervista a Broonzy, Chatman, Williamson [2]. Per un'analisi molto approfondita e ben documentata dell'ambiente dei Levee Camp, attraverso documenti scritti e orali, consiglio l'ottimo articolo [8].

¹¹ Il risultato delle ricerche di Odum, svolte assieme ad H.W. Johnson, è stato pubblicato nei volumi [17, 18].

¹² Alcuni documenti contenenti *holler* provengono dalla spedizione di Herbert Halpert del 1939 nel Sud degli Stati Uniti, quella di Harry Olster del 1959 alla Angola State Prison, e quelle di Harold Courlan-

3.2 Canti Liberi

In ambito accademico gli *holler* sono parzialmente sotto osservazione da più di un secolo.

Senza discutere nel dettaglio le teorie proposte a riguardo, mi preme soffermarmi più che altro sulle osservazioni strettamente tecniche storicamente svolte su questi canti. Diversi autori hanno infatti sostenuto che questi canti siano: liberamente intonati,¹³ a tempo libero,¹⁴ privi di struttura o di contenuto testuale stringente.¹⁵ Nessuna di queste ipotesi fondamentali è stata confermata o smentita su base documentaria, nemmeno nei lavori più recenti sull'argomento.¹⁶ Insomma, tutte descrizioni che, muovendosi per assoluti e ragionando per negazione, in realtà non dicono nulla dell'oggetto a cui fanno riferimento.

Gli *holler* hanno un'intonazione effettivamente estremamente cangiante, con linee melodiche molto mobili. Ma siamo così sicuri che siano intonati liberamente e a tempo libero? Osserviamo due spettrogrammi di due moduli in un *holler*: si tratta del modulo introduttivo alle strofe usato da Henry Ratcliff in *Louisiana* [22]:

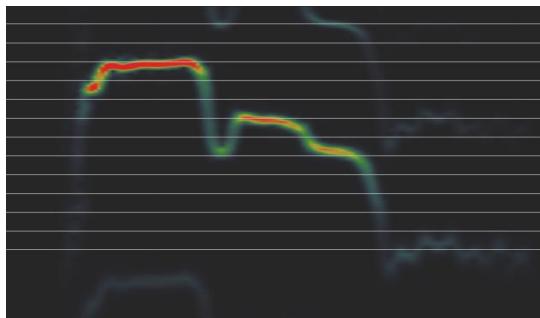


Figura 1. Modulo melodico di introduzione, da 00:13.6

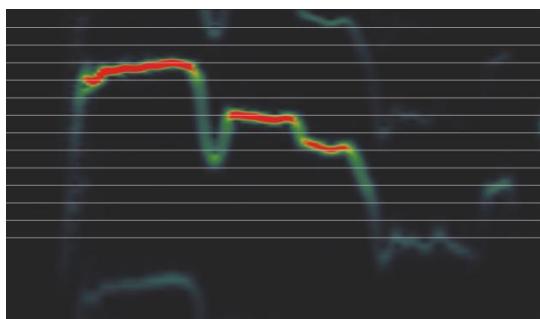


Figura 2. Modulo melodico di introduzione, da 00:55.5

Ho volutamente lasciato le immagini spoglie di annotazioni o elementi grafici aggiuntivi, fatto salvo per la

der nelle medesime zone nei primi anni '60. Tutti questi materiali sono depositati alla Library of Congress di Washington.

¹³ [7, 19, 20, 23, 25].

¹⁴ [12, 21].

¹⁵ [1, 6, 20].

¹⁶ [3, 5, 24].

griglia a barre orizzontali che segnala le distanze in semitonni dall'altezza fondamentale su cui si adagia la *finalis* del modulo, per lasciare che l'immagine parli da sé. Se dovessimo discutere del grado di libertà ritmica su cui fa affidamento il cantante, guardando queste immagini di due riproduzioni dello stesso modello melodico dovremmo concludere che essa è pari a zero. Le due immagini, al netto di microvariazioni, sono ritmicamente sovrapponibili. Certo sarebbe complesso spiegare su che tipo di ritmo si basino, intendendo per ritmo una suddivisione ciclica di pulsazioni forti e deboli. Sarà uno schema ritmico complesso, aperto a variabilità, stratificato quanto si vuole, ma che uno schema esista è difficile metterlo in dubbio.

Un discorso simile andrebbe fatto per le altezze raggiunte dalla voce durante il canto. Il cantante è davvero libero di intonare a piacimento il suo canto? O non sta forse seguendo uno schema anche su questo piano? Le due immagini appaiono sovrapponibili anche rispetto a questo aspetto. Certo, vi sono alcune differenze minori, come l'imperfezione della cuspide verso il basso che spezza centralmente il modulo, o l'attacco vagamente diverso alla prima arcata, o le microscopiche variazioni nella gestione del vibrato sulla *finalis*, ma immagino si possa concordare che, data anche la rapidità dell'evento acustico (l'immagine mostra circa 3.2 secondi di registrazione), e tenendo conto del fatto che Henry Ratcliff non era un cantante professionista, ma un carcerato in una prigione dove viveva in condizioni prossime a quelle di schiavitù, qualche leggera sfasatura nella gestione del percorso melodico può anche essere considerata come ininfluente.

Che il cantante si stia affidando a degli schemi fissi non è per nulla sorprendente:¹⁷ gli *holler* appartengono all'ambito delle musiche di tradizione orale, nelle quali lo sfruttamento di strutture formulaiche è perfettamente funzionale all'esposizione del contenuto musicale.¹⁸ Ma la forza con cui queste immagini ci mostrano la presenza di queste formule non sarebbe raggiungibile in modo efficace con nessun tipo di descrizione verbale o trascrizione su pentagramma. Il senso di chiarezza che si ottiene accostando queste immagini deriva dal modo in cui le immagini stesse ci parlano dell'oggetto musicale, ovvero dalle peculiarità dello stesso strumento di analisi informatico.

3.3 Blue Notes

Una delle descrizioni (solo apparentemente) più tecniche e stringenti date di questi canti è che utilizzino scale

¹⁷ Chiaramente questi schemi o formule contengono certo grado di variabilità applicativa, tale da permetter loro di essere sfruttati in più situazioni, con testi differenti, o magari riprodotti cambiando qualche piccolo elemento. Spesso nelle formule stesse sono in qualche modo prescritti i limiti stessi di questa variabilità, rilevabili attraverso un confronto minuzioso di un numero sufficientemente grande di casi. Questo tipo di discorso esula dall'argomento qui in esame.

¹⁸ Mi riferisco qui al significato di "formula" come è ampiamente accettato e utilizzato in campo etnomusicologico (e non solo) fin dal pionieristico lavoro di Lord [13].

melodiche fortemente pentatoniche, con largo uso di “blue notes”.¹⁹

Concetto dalla tortuosa e non sempre limpida storia,²⁰ indicherebbe note “ad altezza variabile”, all’interno di un *range* in cui la libertà di selezione è, per l’infinita divisibilità dello spazio, infinita essa stessa. Si noti inoltre come nel concetto di “blue note” sia sottinteso quello di “note”: l’unione dei due concetti produce uno strumento epistemologico tutt’altro che intuitivo. Per usarlo dovremmo postulare che all’interno di uno spazio indefinito ma dai confini definiti vi sia un oggetto musicale indefinito la cui definizione matematica dovrebbe essere obbligatoriamente negoziata volta per volta in maniera del tutto soggettiva dal cantante.

Sarebbe del tutto fuori dall’argomento qui in discussione provare a dirimere la questione se il concetto di “blue note” sia o meno epistemologicamente efficace nel descrivere alcunché. Soffermiamoci però sul caso specifico degli *holler*.

Come esempio si potrebbero prendere alcuni frammenti dal brano *Levee Camp Holler* di Johnny Lee Moore [15]. Il modulo in questione è quello usato dal cantante per chiudere la prima sezione.²¹

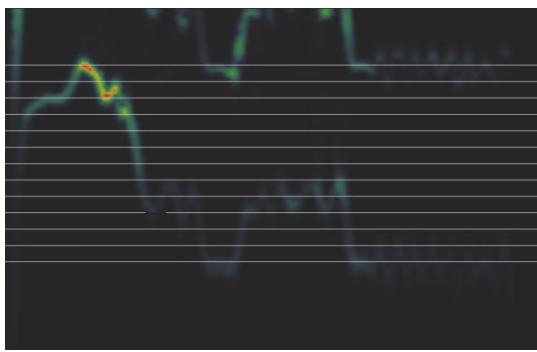


Figura 3. Modulo melodico di chiusura, da 1:32:8

Com’è possibile vedere, questo modulo si compone di due brevi arcate complesse; la prima sale all’altezza di una settima minore dalla fondamentale, scatta quindi con una rapida cuspide all’ottava, torna alla settima minore, quindi scende verso la fondamentale, fermandosi per un attimo in un rapido vibrato. L’altezza di questo vibrato ritorna anche nella seconda breve arcata: dal grafico è osservabile come questi vibrati si muovano sostanzialmente tra le altezze di terza minore e quarta, ovvero con un baricentro su una terza maggiore. Lo stesso scarto di semitono sopra e sotto al baricentro per il vibrato torna d’altronde anche a colorare la *finalis*, muovendosi attorno all’altezza del tratto che separa le due arcate.

A mio parere in questa immagine non esiste quasi nulla che possa assomigliare al concetto di “nota”: al massimo il concetto potrebbe essere applicato alla lunga

¹⁹ [7, 20, 25].

²⁰ Sull’argomento, suggerisco lo stimolante articolo [28].

²¹ Per avere un’idea delle durate, l’immagine riprende 5 secondi e 4 decimi del brano.

stasi di chiusura. La voce del cantante, che pure, come si può evincere all’ascolto, come Henry Ratcliff non era chiaramente un professionista e non aveva particolari abilità di intonazione o gestione timbrica, si muove fluidamente tra le altezze creando arcate i cui punti fondamentali (picchi, avvallamenti, più o meno brevi stasi) coincidono, con ragionevole margine di errore, con quelle tipiche di una scala temperata. Persino il rapidissimo vibrato del cantante può essere ragionevolmente ricondotto ad un’oscillazione attorno ad un punto centrale “intonato” con la tradizionale scala temperata. Risulta perciò difficile sostenere che il cantante usi delle note ad altezza variabile.

Il problema è che queste non sono note intonate, ma punti di passaggio, picchi, centri di vibrazione. In realtà la voce del cantante non si sofferma veramente su nessuna delle altezze descritte: le raggiunge, le sottintende appoggiandovi la sillaba (visibile tralaltro per il virare verso il rosso all’aumento dell’ampiezza di segnale proprio in corrispondenza di queste altezze), solo per abbandonarle immediatamente dopo.

Per quanto riguarda la variabilità di scelta delle altezze di questi punti particolari, confrontiamo tra loro tre versioni, dallo stesso brano, della prima arcata del modulo, tratte da tre ripetizioni della stessa identica formula melodica usate in posizioni funzionali identiche rispetto al testo:

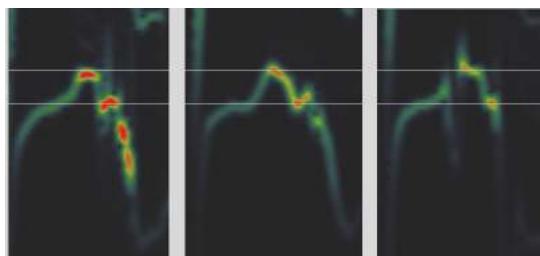


Figura 4. Primo elemento ad arcata dei moduli melodici di chiusura (da sinistra) a 1:23.4, a 1:32.8, ovvero quella della fig. 3, e a 2:27.5

Ho ritagliato le immagini alla sola parte dell’arcata e lasciato solo le linee che segnalano le altezze di settima minore e ottava dalla fondamentale. I profili melodici delle tre strutture ad arcata, per quanto complessi e al netto di minori variazioni comunque legate anche alle necessità di enunciazione del testo, sono sostanzialmente identici. Diventa allora ragionevole sostenere che il cantante non stia pensando ad un movimento suddivisibile in singoli oggetti ad altezze definite, come potrebbe suggerire una riduzione di questo passaggio ad una trascrizione su pentagramma con tre note sul settimo, ottavo e settimo grado di una scala, ma stia invece ragionando in base ad un’unica struttura formolare, affrontata sempre con lo stesso metodo. Vorrei sottolineare tra l’altro che questa particolare forma di arcata torna all’interno del brano solo per introdurre moduli della stessa forma e in posizioni linguistiche identiche rispetto

al flusso testuale. Ciò, se verificato su ampia scala,²² sarebbe segnale di un forte grado di connessione tra il materiale musicale utilizzato e il contenuto verbale esposto, a dimostrazione che questi canti non sono, come alcuni sostengono, "privi di struttura".²³

3.4 Visualizzare la Formulaicità

La particolare forma ad arcata della figura 4 potrebbe perciò esser pensata come unico elemento melodico sostanzialmente formulaico, che come tale può quindi trovare anche altri impieghi. Ad esempio, il cantante utilizza nell'arco dell'intero brano diversi altri elementi simili a quello che introduce questo modulo. Confrontiamone alcuni:

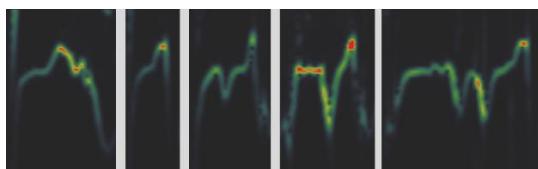


Figura 4. Elementi ad arcata dei moduli melodici (da sinistra) che iniziano a 1:32.8 (di nuovo quello di prima, come confronto), a 1:15, a 1:26, a 2:22, a 2:53,7

Ho semplificato ulteriormente la visualizzazione rimuovendo la griglia perché vorrei che quest'immagine venisse osservata non tanto sul dettaglio delle altezze raggiunte dal canto, quanto sul piano della morfologia dell'elemento interessato.

Tutti questi elementi, ognuno singolarmente riutilizzato più volte nel brano, fanno riferimento allo stesso tipo di movimento melodico, alla stessa formula, diversamente coniugata in forme più o meno complesse. Accorgersi della questione apre ovviamente una serie interminabile di domande: in che modo queste tipologie sono in relazione con un modello fondamentale? E' il modello a prevedere una variabilità tale da potersi coniugare in diverse forme rimanendo lo stesso, oppure queste forme condividono con il modello solo il profilo, e vanno intese in realtà come oggetti funzionalmente diversi? Per rispondere a queste domande non dovremo anche capire che relazione intercorre tra i testi enunciati in questi punti e la loro posizione rispetto alla forma generale del discorso?

Molte di queste domande dovranno attendere il completamento di studi più approfonditi a riguardo, e non è nemmeno certo che sia possibile rispondere a tutte. Quel che importa, per la discussione qui in essere, è che queste osservazioni, nonché soprattutto queste domande, le quali aprono la strada a nuovi possibili gradi di interpretazione dei materiali in analisi, sono emerse grazie soprattutto al tipo di manipolazione della traccia audio permessa dal software.

4. TRADUZIONE E OGGETTIVITÀ

4.1 Attenzione al Dettaglio

Questi sono solo alcuni esempi di come possano svolgersi percorsi di approfondimento su materiali sonori quando il ricercatore si affida a software di analisi, e che relazione le forme di questi percorsi abbiano con i risultati, anche parziali, ottenuti. Come è stato possibile vedere, il tipo di analisi permesso da questi strumenti produce una quantità di dati limitabile solo dalla pazienza del ricercatore. Attraverso questi software è davvero possibile sezionare, misurare, analizzare dettagli minimi degli avvenimenti sonori, potendone così avere un quadro più completo. D'altra parte, questo tipo di lavoro non sempre è necessario, e non sempre dà ottimi risultati. Lo strumento, come dicevamo all'inizio, dev'essere commisurato allo scopo del ricercatore, e gli scopi di un'analisi etnomusicologica non si limitano alla vivisezione di una registrazione. Questo benché ne costituiscano una parte rilevante, dato che sulla base delle osservazioni contenutistiche avvengono poi deduzioni più ampie e si fondano determinate interpretazioni, come abbiam visto nel caso degli *holler*, la cui fondatezza è legata a doppio filo alla qualità dell'indagine analitica da cui inferisce.

D'altro canto, il linguaggio dei numeri, dei grafici, delle variabili, di cui spandono a piene mani gli *output* di questi software, si rivela essere qualcosa che è tutto sommato vicino alla nostra sensibilità. E questo è certamente un punto a favore di questi *software*: se l'etnomusicologo è convinto di aver correttamente interpretato, attraverso un lungo e complesso percorso, il significato del contenuto sonoro di un determinato evento, gli rimarrà comunque il dubbio di come trasmetterlo in modo efficace ai suoi colleghi e lettori. Essi certo non possono, per arrivare alle stesse interpretazioni, riprodurre nuovamente e per intero lo stesso percorso svolto dal ricercatore. Inoltre, parlare di musica è sempre complicato. La difficoltà è sempre la stessa: quella di trovare le giuste parole che traducano i contenuti, che rendano esplicite le intuizioni del ricercatore. Il tipo di riduzione a grandezze quantificabili, osservabili, parametrabili operato dai software di analisi musicale sulle registrazioni è un supporto sicuramente efficace alla traduzione di quegli elementi, quei parametri, quelle differenze che il ricercatore ha incontrato e vuol provare a comunicare. Per certi versi, un software musicale non sta solo misurando dei parametri: sta misurando delle distanze, sta rendendo esplicito numericamente il lavoro di ragionamento del ricercatore sul suo materiale d'indagine. E la matematica, si sa, è un linguaggio esatto e universale. Oppure no?

4.2 Oggettivare Soggettività

Il linguaggio dei numeri è allettante: quando qualcosa viene ridotto a percentuali, parametri, variabili e grafici, ci sembra sempre automaticamente ragionevole pensare di averne dato una descrizione oggettiva. Non vanno però dimenticate alcune questioni.

²² Per averne la certezza dovranno attendere l'esito degli studi di ricerca avviati.

²³ Vedi nota 16.

I software di analisi non possono dire nulla del contesto di provenienza della registrazione, né della situazione che ha dato origine alla registrazione stessa o del rapporto che intercorreva tra ricercatore e soggetto. Inoltre, è il ricercatore che sceglie di ricorrere ad un certo software, per dar rilievo a una componente particolare della sua analisi piuttosto che a un'altra. Usandolo, è il ricercatore che sceglie di volta in volta come sistemare i parametri al fine di interrogare il segnale sonoro: è lui che posiziona i marcatori manualmente, o assegna le variabili. E' sempre lui che sceglie un certo *plugin* o un altro, o decide di usare una certa visualizzazione o un'altra. Infine, i numeri sono puri: è il ricercatore che dona un significato ai dati di *output*, mai il contrario. Queste osservazioni sono in genere ben presenti a chi si occupa di ricerca in campo di scienze naturali, e riguardano il concetto stesso di grandezza, misurazione, e soprattutto di errore associato ad esse.

L'analisi prodotta attraverso un software non va mai intesa come dato oggettivo su una realtà, ma come parte del processo di negoziazione del ricercatore con quella realtà stessa e con la comunità con cui deve dialogare e a cui vuole riportare i risultati delle sue ricerche. Un altro ricercatore che svolga le stesse analisi sugli stessi materiali con gli stessi strumenti potrebbe insomma giungere a risultati diversi, e dal confronto tra questi risultati potrebbe emergere una nuova lettura del dato fenomenico, che andrà ancora una volta intesa, necessariamente, come parziale, fino a quella successiva.

Ma se un'analisi basata su software non è in grado di dare risultati oggettivi, esattamente come i vecchi metodi, e anzi richiede competenze differenti e un lavoro più prolungato sulle registrazioni, perché affidarsi ad essa?

4.3 Differenti Modi per Pensare

Uno dei punti di forza dell'uso di un software per affrontare i dettagli dell'analisi di un brano è indubbiamente il cambio di prospettiva. Se un ricercatore impara, parafrasando la citazione da Warnier, a dilatare la propria sintesi corporale all'insieme stratificato di informazioni proveniente da questi programmi, se non altro avrà come minimo arricchito la sua capacità d'ascolto con le dinamiche di un senso, per così dire, informatico, con differenti limiti e differenti potenzialità rispetto al semplice orecchio umano.

Non solo: questi software, o almeno quelli più avanzati, sono anche stati pensati in modo modulare. Vi sono tanti diversi modi in cui, attraverso di essi, si può interrogare un segnale audio. Sfruttando questi strumenti, i ricercatori hanno quindi accesso a un'ampia gamma di possibili metodi di analisi e confronto dello stesso materiale, che possono concorrere all'abbattimento delle eventuali limitazioni legate al proprio punto di vista e alla propria preparazione. Abbiamo visto come l'analisi di alcuni elementi di una registrazione possa far sorgere diverse domande, su cui svolgere e far avanzare la ricerca in nuove o più approfondite direzioni.

Certo, a volte l'accumularsi di diversi punti di vista sugli stessi materiali rischia di produrre un apparentemente caotico quadro cubista più che una nitida e per-

fettamente rassicurante statua michelangiolesca. Ma questo non è a causa del software in sé, o dell'imperizia del ricercatore, quanto per la natura complessa e stratificata dell'evento sonoro che viene indagato.

Solo cercando di prendere in considerazione tutti gli aspetti dell'evento sonoro, e solo assegnando a ciascuno di essi il giusto strumento metodologico, si può sperare di arrivare ad una discussione sufficientemente accurata dello stesso. Su questo lato i software di analisi trovano la loro massima efficacia: non tanto per la loro più o meno esatta capacità di produrre dati oggettivi sui materiali indagati, ma anche e soprattutto grazie alla loro capacità di offrire al ricercatore un numero ben più ampio di metodi per interrogare ed interrogarsi sull'oggetto delle proprie ricerche.

5. RIFERIMENTI BIBLIOGRAFICI

- [1] M. Ariodante, *Hollers e Cries nella tradizione musicale dei neri nord-americani*, tesi di Laurea in Etnomusicologia, facoltà di Lettere e Filosofia dell'Università La Sapienza di Roma, A.A. 2001-2002.
- [2] W. L. C. Broonzy, P. Chatman, J. L. Williamson, *Conversation with Big Bill Broonzy, Memphis Slim, and Sonny Boy Williamson about levee camps and Southern prisons / Levee camp song frgments*, Archivio Association for Cultural Equity, TD93R02, 1947.
- [3] J. H. Brunvard, *American Folklore, an encyclopedia*, Garland Publishing, New York & London, 1996.
- [4] C. Cannam, C. Landone, and M. Sandler: "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files," *Proceedings of the ACM Multimedia 2010 International Conference*.
- [5] G. Chiriacò, "Filling the Space: Field Hollers and the Social Role of Singers in African-American Communities," *American Musicologica Society/Society for Ethnomusicology/Society for Music Theory 2012 Annual Meeting Abstract*, Vol. 1, novembre, p. 34-46, 2012.
- [6] J. Cogdell Djedje, "African American cries, calls and hollers", *The Garland Encyclopedia of World Music*, Garland Publishing, New York & London, Vol. III, pp. 81-86, 2000.
- [7] H. Courlander, *Negro Folk Music U.S.A.*, Columbia University Press, New York, 1963.
- [8] J. Cowley, "Shack Bullies and Levee Contractors: Bluesmen as Ethnographers", *Journal of Folklore Research*, Vol. 28 No. 2/3, pp. 135-162, 1991.
- [9] N. Crossley, *Reflexive Embodiment in Contemporary Society*, Open University Press, Maidenhead, 2006.
- [10] N. Crossley, *The Social Body: habit, identity and desire*, SAGE Publications, London, 2001.
- [11] M. Dayton, *The science of musical sounds*, The Macmillan Company, New York, 1916.

- [12] A. Lomax, fascicolo “The music of the area” contenuto in The Library of Congress and Fisk University Mississippi Delta collection (AFC 1941/002), Archive of Folk Culture, American Folklife Center, Library of Congress.
- [13] A. Lord, *The Singer of Tales*, Harvard University Press, Cambridge, 2000.
- [14] M. Metfessel, C. Seashore, *Phonophotography in folk music*, University of North Carolina Press, Chapel Hill, 1928.
- [15] J. L. Moore, *Levee camp holler (Downtown money waster)*, Archivio Association for Cultural Equity, T886R03, 1959.
- [16] I. Neath, A. Surprenant: *La memoria umana: Apprendimento e organizzazione delle conoscenze*, Idelson-Gnocchi, Napoli, 2010.
- [17] H. W. Odum, G. B. Johnson, *The negro and his songs: a study of typical negro songs in the South*, University of North Carolina Press, Chapel Hill, 1925.
- [18] H. W. Odum, G. B. Johnson, *Negro workaday songs*, University of North Carolina Press, Chapel Hill, 1926.
- [19] P. Oliver, *Savannah syncopators*, Studio Vista, Londra, 1970.
- [20] P. Oliver, *The story of the Blues*, Barrie & Rockliff, Londra, 1969.
- [21] C. Peabody, “Notes on Negro Music,” *Journal of American Folk-lore*, 16 luglio 1903, pp. 148-152.
- [22] H. Ratcliff, (*Look for me in*) *Louisiana*, Archivio Association for Cultural Equity, T883R08.
- [23] J. S. Roberts, *Black music of two worlds*, Allen Lane, London, 1972.
- [24] J. Shepherd, *Continuum Encyclopedia of Popular Music of the World*, Bloomsbury Academic, London, 2003.
- [25] M. Stearns, *Negro Blues and Hollers*, Archivio Library of Congress AFS L59, 1962.
- [26] L. Tronchin, *Elementi di Fisica Tecnica e Controllo Ambientale*, Ed. Esculapio, Bologna, 2008.
- [27] J. P. Warnier, *La cultura materiale*, Meltemi, Roma, 2005.
- [28] H. Weisethaunet, “Is there such a thing as the ‘blue note’?”, *Popular Music*, Vol. 20 No. 1, pp. 99-116, 2001.
- [29] Pagina html programmata per fornire un ulteriore metodo di analisi delle immagini prodotte dal programma Sonic Visualiser:
<http://alpert.altervista.org/LorenzosTesi/>
- [30] Sito contenente tutti gli articoli di F. Tannenbaum riguardanti la situazione delle prigioni nel Sud degli Stati Uniti:
<http://www.unz.org/Author/TannenbaumFrank?view=Overview&Title=prison>
- [31] Sito da cui è possibile accedere all’intera documentazione prodotta dal *Federal Writer’s Project*:
<https://www.loc.gov/collections/federal-writers-project/>

IL CASO DELL'ARCHIVIO STORICO DEL TEATRO REGIO DI PARMA: APPLICAZIONE DI UNA METODOLOGIA SCIENTIFICA PER LA CONSERVAZIONE ATTIVA DI DOCUMENTI SONORI

Alessandro Russo

Dip. di Ing. dell'Informazione
Università di Padova
russoale@dei.unipd.it

Valentina Burini

Dip. di Ing. dell'Informazione
Università di Padova
v.burini@hotmail.it

Sergio Canazza

Dip. di Ing. dell'Informazione
Università di Padova
canazza@dei.unipd.it

Antonio Rodà

Dip. di Ing. dell'Informazione
Università di Padova
roda@dei.unipd.it

ABSTRACT

In questo articolo sono presentati i risultati del progetto di ricerca finalizzato alla conservazione, al restauro e alla digitalizzazione di 200 documenti sonori provenienti dal Fondo dell'Archivio Storico del Teatro Regio di Parma. Il progetto è stato affidato nel 2015 ad Audio Innova Srl, spin off dell'Università degli Studi di Padova, presso i laboratori del Centro di Sonologia Computazionale (CSC) del Dipartimento di Ingegneria dell'Informazione. Le bobine riversate sono relative alla stagione concertistica in programmazione al Teatro Regio nel biennio 1991-1992. I nastri, registrati con un magnetofono Revox PR99, contengono materiale inedito di grande valore artistico-culturale. Il progetto è passato attraverso diverse fasi di lavoro, dall'ottimizzazione e il restauro del supporto alla conversione A/D dell'informazione sonora. In particolare, grande attenzione viene rivolta alla raccolta dei metadati, applicando una metodologia che permetta di ottenere una copia conservativa tale da rispecchiare il documento sonoro nella sua interezza con la minor perdita di informazione e non si limitasse al solo segnale audio.

1. INTRODUZIONE

Quando si parla di *bene culturale* è di fondamentale importanza chiedersi che cosa si intenda con questo termine. Una prima definizione venne proposta dalla Commissione Franceschini nel 1964 [1]. Essa sottolinea

Copyright: © 2016 Russo 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.

come a questa categoria appartengano i «*beni d'interesse storico, archeologico, artistico, ambientale, archivistico, librario, nonché – più in generale – «qualsiasi altro bene che costituisca testimonianza materiale avente valore di civiltà»*». Una importante categoria di beni culturali è costituita dai cosiddetti *beni culturali audiovisivi*. Si tratta di una categoria relativamente “giovane”, ma estremamente vasta per contenuti e informazioni, di giorno in giorno in continua espansione ed evoluzione. Nel 1877, infatti, Thomas Edison inventò il primo apparecchio in grado di registrare suoni, il cilindro fonografico, brevettato poi nel 1878. Da quel momento in poi si sono succedute numerose invenzioni e nuove tecniche di registrazione, fino ad arrivare, nel 1935 in Germania, al primo magnetofono [2], strumento in grado di registrare e riprodurre informazioni sonore aventi come supporto un nastro magnetico. Sebbene nel corso degli anni, soprattutto grazie all'introduzione del formato digitale, il progresso scientifico abbia portato allo sviluppo di nuovi formati, tecnologie e tecniche di registrazione, ancora oggi la qualità delle registrazioni su nastro è considerata da molti audiofili insuperabile. Che si tratti di composizioni o esecuzioni musicali, registrazioni di carattere politico, antropologico o scientifico-divulgativo, esiste un'enorme eredità culturale su nastro che deve essere conservata e preservata dal trascorrere del tempo. Malgrado i nastri magneticci siano stati prodotti con diversi materiali nel corso degli anni, hanno mantenuto la caratteristica di essere supporti deperibili, che possono andare incontro a fenomeni di degrado molto rapidi. Non bisogna dimenticare come numerosi fenomeni di degrado fisico/chimico possano portare, nei casi più gravi, all'impossibilità di lettura della stessa informazione audio, oltre che alla perdita del supporto. Negli ultimi anni si

sente molto parlare della cosiddetta *corsa alla digitalizzazione* [3]: si tratta senza dubbio di un processo di grande importanza culturale, in quanto permette di estrarre e preservare informazioni da supporti ormai obsoleti ed archiviarle in formati moderni. Il termine “corsa”, però, ne denota una connotazione negativa in quanto spesso porta al fenomeno della “digitalizzazione massiva”, noncurante di molti aspetti fondamentali per una corretta conservazione dei documenti sonori. Tra questi aspetti vi è, per esempio, la tendenza a considerare importanti solamente il segnale audio, trascurando i supporti. Da un punto di vista culturale non bisogna dimenticare che essi costituiscono, infatti, una testimonianza importantissima dello stato del progresso scientifico-tecnologico del tempo di interesse. Il documento sonoro va infatti considerato nella sua interezza, così come un manoscritto è costituito da pergamena e inchiostro e, come tale, deve essere preservato. A questo proposito, va ritenuta di fondamentale importanza tutta una serie di metadati che completa la copia conservativa e permette di archiviare, oltre i file audio, informazioni importanti per la ricreazione del documento sonoro virtuale. Nel corso degli anni '80-'90, la digitalizzazione era considerata una pratica finalizzata principalmente all'accesso a materiale particolarmente raro e/o distante e non alla conservazione [2]. Smith riprende nuovamente questo concetto nel 1999, ribadendo come la pratica non sia finalizzata alla conservazione dei documenti - “at least not yet” [4]. La digitalizzazione è una pratica che viene ormai accettata dalla comunità internazionale come parziale soluzione al problema del degrado dei supporti. Una volta ottenuta una copia conservativa digitale, infatti, è possibile ascoltare l'informazione estratta senza sottoporre ogni volta il documento sonoro a stress meccanici, che rischerebbero di aggravare ulteriormente situazioni spesso già critiche. Oltre a preservarne il contenuto, la digitalizzazione permette la fruibilità dell'informazione audio in larga scala consentendo, ad esempio, l'accesso a documenti materiali ora archiviati presso piccole realtà locali, rendendoli virtualmente raggiungibili da un sempre maggior numero di persone grazie alla rete web. La quasi totale mancanza di caratterizzazione delle differenti tipologie e marche di nastro, così come di linee guida condivise per la loro conservazione, è un punto di interesse verso cui guardano le comunità archivistiche internazionali. Nell'articolo vengono presentati il progetto e lo scenario in cui si colloca (2), la strumentazione e la metodologia adottata (3), lo stato di conservazione dei nastri magnetici (4) e la creazione delle copie conservative (5).

2. IL PROGETTO

Il progetto di restauro e digitalizzazione dell'Archivio storico del Teatro Regio ha visto una prima fase portata avanti dal Laboratorio “Mirage” dell'Università di Udine a partire dall'anno 2014, grazie alla quale sono stati riversati circa 300 nastri da 1/4" relativi al periodo di tempo compreso tra il 1961 e il 1989. Una seconda fase ha

visto affidare il progetto di digitalizzazione e restauro del secondo lotto di nastri magneticci da Audio Innova Srl., spin off dell'Università degli Studi di Padova. Il lavoro è stato svolto all'interno dei laboratori dello storico Centro di Sonologia Computazionale (CSC), attivo dal 1979 e dal 2009 parte del Dipartimento di Ingegneria dell'Informazione (DEI) dell'Università di Padova con il nome CSC - Sound and Music Computing Group. Il progetto interessa 200 nastri relativi alle rappresentazioni messe in scena nel Teatro Regio nei primi anni '90. Di ogni opera i tecnici del Regio tennero memoria attraverso la registrazione su nastro magnetico anche delle prove e delle successive repliche, documenti utili da un punto di vista musicologico nell'individuazione di elementi significativi delle diverse rappresentazioni. I nastri denotano una direzione artistica della stagione concertistica registratavi fortemente legata alla tradizione operistica parmense, come il caso della *Luisa Miller* di Verdi, l'opera che ha segnato per il compositore di Busseto il passaggio ad uno stile compositivo più maturo, ma attenta anche all'allora scenario cantautorale proponendo inoltre, tra i tanti, concerti di Giorgio Gaber e Ivano Fossati.

2.1 L'Archivio del Teatro Regio di Parma

Il fondo del Teatro Regio di Parma è diviso in due sezioni, una relativa agli anni 1816 – 2001, custodita presso il Servizio Casa della Musica di Parma, e l'altra dal 2001 ad oggi, presso il Teatro Regio. Oltre alle registrazioni audio/video delle rappresentazioni messe in scena, il corpus del materiale archiviato comprende anche ritratti e fotografie degli artisti, manifesti, figurini, bozzetti, progetti tecnici, spartiti musicali a stampa e manoscritti, stampe e fotografie di interpreti e compositori, libretti, periodici, costumi e gioielli di scena provenienti da donazioni, lasciti e depositi.

2.2 Lo Scenario

L'obsolescenza e il degrado dei supporti sono alcuni dei maggiori problemi con cui gli archivi di tutto il mondo si devono confrontare, di conseguenza le metodologie di conservazioni sono da considerarsi valide per la grande maggioranza dei casi. Tuttavia bisogna sottolineare come vi siano in realtà due principali tipologie di archivi. L'Archivio del Teatro Regio di Parma è un esempio rappresentativo del tipo di archivio maggiormente diffuso in Europa, che si contrappone alla tipologia caratteristica che è possibile ritrovare in Stati relativamente più giovani quali gli Stati Uniti, il Canada e l'Australia. La principale differenza risiede nel fatto che, mentre in Europa sono presenti numerosissimi archivi di piccola o media dimensione, negli Stati sopracitati vi è una generale tendenza alla centralizzazione, con pochi archivi di grande dimensione. A livello europeo costituiscono esempi di eccezioni la *British Library* di Londra e l'*Institut National Audiovisuel* di Parigi [5]. Se le ragioni di questa divisione sono prettamente di carattere storico, il quadro generale in

ambito europeo delinea una generale frammentazione del patrimonio archivistico culturale. In questo scenario, la digitalizzazione permette di ovviare al problema, in quanto, a partire dai file archiviati, è possibile creare un network di condivisione delle risorse che permetta la ricreazione virtuale di collezioni in realtà frammentate, nonché una maggiore facilità di accesso.

3. STRUMENTAZIONE E METODOLOGIA

Il metodo di lavoro messo a punto e consolidato all'interno del Centro di Sonologia Computazionale può essere riassunto nelle seguenti fasi [6]:

1. Preparazione del supporto
 - 1.1 Documentazione fisica
 - 1.1.1 Fotografie
 - 1.1.2. Scansioni
 - 1.1.3 Analisi dati
 - 1.2 Ispezione visiva
 - 1.3 Ottimizzazione del supporto
2. Riversamento del segnale (Figura 1)
 - 2.1 Analisi dei formati/parametri di registrazione
 - 2.2 Setup del sistema
 - 2.2.1 Strumentazione per la lettura (es. registratore a bobine)
 - 2.2.2 Strumentazione per il riversamento A/D (es. convertitori, software di acquisizione, sistema di monitoraggio, ...)
 - 2.3 Monitoraggio
 - 2.4 Analisi dati
 - 2.5 Archiviazione del supporto di origine
3. Elaborazione dei dati e archiviazione
 - 3.1 Estrazione dei metadati
 - 3.2 Completamento della copia conservativa

Con questa metodologia sono stati avviati diversi progetti di digitalizzazione e restauro di documenti sonori, sia a livello nazionale, come ad esempio il progetto REVIVAL (REStoration of the Vlcentini archive in Verona) della Fondazione Arena di Verona [7], l'intero Archivio Luigi Nono di Venezia (in corso) che internazionale, come il progetto relativo ai nastri inediti di Luciano Berio [8], finanziato dalla Paul Sacher Stiftung di Basilea (in corso). Oltre ai documenti sonori di carattere compositivo sono stati completati progetti relativi agli ambiti dell'etnomusicologia e della pedagogia, come la digitalizzazione del materiale raccolto nei suoi studi da Johannella Tafuri. Per la lettura delle bobine sono stati utilizzati due differenti modelli di magnetofono, entrambi di qualità professionale, uno Studer A810 e un Revox PR99 MkII. Le letture sono state effettuate con curva di equalizzazione CCIR in quanto standard maggiormente diffuso in Europa, nonché l'unica presente nel magnetofono Revox PR99 utilizzato dai tecnici del Teatro Regio per le registrazioni. Prima di procedere con il riversamento, essendo le bobine dell'intero lotto riavvolte in testa, i nastri sono stati svolti e riavvolti a velocità controllata, così da verificarne lo stato di conservazione e la capacità di resistere allo stress meccanico. La conversione A/D è stata effettuata attraverso due interfacce audio professionali commercializzate da Prism Sound, in particolare il convertitore λύρα, interfacciato con un iMac Intel Core i5 3470, 2,9 GHz, 8 GB RAM DDR3 con sistema operativo OS X 10.9.5 e il convertitore Ορφέως interfacciato via Firewire ad un iMac Intel Core i5 4570R, 2,7 GHz, 8 GB RAM DDR3 con sistema operativo OS X 10.9.5. Come software per l'acquisizione è stato scelto Adobe Audition CS6. Il riversamento è stato effettuato in maniera consona agli standard internazionali [2], con una frequenza di campionamento di 96 KHz e una profondità di 24 bit e con livello di acquisizione 0 dBu = -18dBFS.

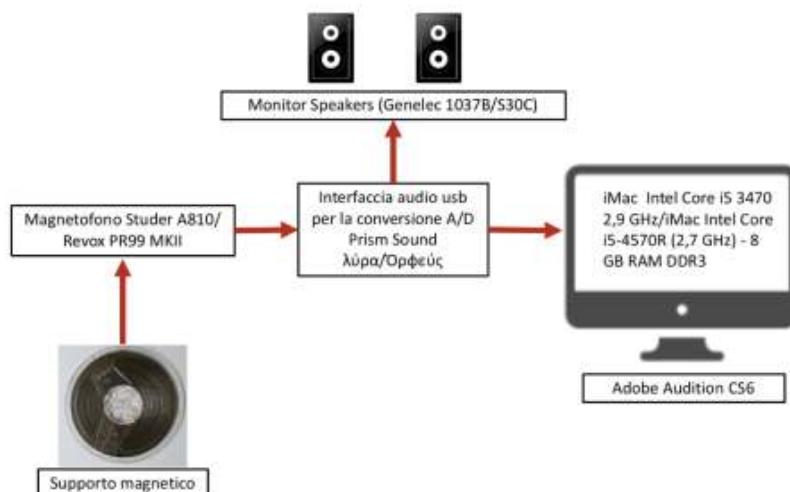


Figura 1. Schema di riversamento dei supporti magnetici.

La principale funzione di una copia conservativa è quella di sostituire il documento di origine e permettere l'accesso alle informazioni qualora venisse perduto o ne risultasse impossibile la lettura. Di conseguenza è di fondamentale importanza che esse non si limitino ai singoli file audio, bensì permettano di ottenere un quadro il quanto più possibile completo di quello che era il documento sonoro originale nella sua interezza. Una particolare attenzione è stata infatti rivolta alla raccolta dei metadati, affinché le copie conservative realizzate contenessero la maggior quantità di informazioni possibili. Per la creazione delle copie di accesso e la raccolta dei metadati è stato utilizzato il software PSkit [9, 10]. In una copia conservativa è importante distinguere il concetto di metadato da quello di informazione contestuale [3]: mentre il primo termine indica le informazioni che è possibile estrarre automaticamente dal segnale acquisito, il secondo indica informazioni addizionali indipendenti dal file audio. Un'importante categoria all'interno delle informazioni contestuali è costituita dalla documentazione fotografica raccolta durante ogni fase del progetto, finalizzata a testimoniare lo stato di conservazione delle bobine prima del loro riversamento e a non perdere tutte le informazioni aggiuntive che spesso è in grado di fornire la stessa custodia del documento sonoro, quali annotazioni a mano e/o stampe che ne indicano, ad esempio, il contenuto. A seconda dello stato di conservazione dei nastri magnetici, in alcuni casi è stato necessario procedere con piccoli interventi di restauro e di ottimizzazione dei supporti, in modo tale da permetterne una corretta lettura e minimizzare il rischio di eventuali danni successivi allo stress meccanico in fase di riversamento. Durante il di digitalizzazione il riversamento è stato monitorato, così da poter annotare tutte le anomalie, sia quelle relative alle condizioni del nastro, sia i disturbi eventualmente introdotti dal sistema in fase di conversione A/D. Il monitoraggio è stato effettuato con una coppia di speakers Genelec 1037B e una coppia di Genelec S30C.

I principali disturbi monitorabili durante il playback possono essere [11]:

- (i) Rumore locale: click, pop, attenuazioni del segnale dovute a giunte o a fenomeni di degrado del nastro.
- (ii) Rumore globale: fenomeni di hum, rumore di fondo o distorsioni
- (iii) Alterazioni prodotte durante la fase di registrazione: rumori elettrici, distorsioni microfoniche, feedback microfonici.
- (iv) Degrado del segnale dovuto al malfunzionamento del sistema di registrazione (ad es. parziale cancellazione delle tracce audio).

4. STATO DI CONSERVAZIONE E OTTIMIZZAZIONE DEI SUPPORTI

Dato il buono stato di conservazione delle bobine, l'intero lotto non ha presentato grosse problematiche e l'assenza di nastri in condizioni particolarmente critiche non ha permesso di riconoscere sintomi di fenomeni di degrado fisico/chimico in stadio avanzato. Relativamente al buono stato di conservazione ha sicuramente influito la giovane età dei documenti sonori in questione, in quanto relativi a rappresentazioni teatrali del biennio '91-'92. In alcuni casi, tuttavia, si sono osservati fenomeni di *windowing* [10] e/o di totale distacco del nastro magnetico dalla flangia centrale (Figure 2 e 3). Questi fenomeni sono particolarmente diffusi per bobine la cui modalità di archiviazione non rispetta le condizioni ottimali indicate dalle linee guida internazionali [12] (ad es. disposte sul piano orizzontale e non verticalmente).



Figura 2. Esempio di distacco del nastro magnetico dalla flangia centrale.



Figura 3. Nastro magnetico soggetto a fenomeni di windowing. Si osservi come il centro della flangia sia completamente fuori asse rispetto al piano di avvolgimento.

Un altro dei problemi più ricorrenti riguarda la cattiva condizione in cui spesso si presentavano le giunte. Lo svolgimento preliminare dei nastri a velocità controllata e un attento monitoraggio hanno permesso di riconoscere quali fossero i punti di giunta e di poterne testare la resistenza allo stress meccanico.



Figura 4. Nastro soggetto a distacco della pasta magnetica.

Le cattive condizioni in cui si trovavano alcune di esse hanno comportato la rottura del nastro durante lo svolgimento preliminare (Figura 5). In molti casi, probabilmente a causa dell'utilizzo di nastri adesivi non specifici, la giunta si era disfatta del tutto, lasciando sui nastri solo la componente collosa che è stata poi rimossa durante la fase di ottimizzazione del supporto. Nei casi sopraccitati si è proceduto con il rifacimento delle giunte utilizzando lo specifico nastro adesivo. In alcuni casi sono state osservate zone in cui era presente un distacco della pasta magnetica (Figura 4). Tuttavia questo fenomeno è stato riscontrato solo in pochi casi isolati e si limita ad alcune aree ristrette. La ragione del distacco è probabilmente di tipo meccanico e non riconducibile a

fenomeni chimici in corso. Per l'intero lotto non sono stati osservati sintomi riconducibili alla *Sticky-shed Syndrome* [10] e, di conseguenza, non è stato necessario sottoporre i nastri magnetici a trattamenti termici controllati.



Figura 5. Punto di giunta soggetto a rottura. Sulla superficie del nastro si osserva la presenza di residui di colla.

Oltre il rifacimento delle giunte, per numerosi campioni è stato necessario procedere con l'applicazione e/o sostituzione del nastro leader in coda e in testa, in quanto mancante o in cattive condizioni. La fase di monitoraggio ha inoltre permesso di individuare la presenza di alcuni nastri registrati con doppia velocità di lettura (7,5 ips e 15 ips), probabilmente per esigenze tecniche da parte degli operatori. Per questi nastri si è proceduto con un doppio riversamento, uno per ogni velocità di scorrimento. Una volta riversate, le bobine sono state archiviate con il nastro riavvolto in coda, come suggerito dalle linee guida per la conservazione dei documenti sonori [12].

5. LE COPIE CONSERVATIVE

Il modello di copia conservativa realizzata è un insieme organizzato di dati che descrive in modo completo l'informazione primaria costituita dal contenuto sonoro, le informazioni che è possibile estrarre dal segnale audio e le informazioni contestuali (documentazione fotografica, documentazione video, note scritte a mano ecc.) in modo da rendere minima la perdita di informazioni. Su richiesta dell'Archivio del Teatro Regio di Parma, la consegna delle copie conservative è stata effettuata in due modalità differenti. Oltre alla consegna dei file su NAS, infatti, sono stati creati supporti ottici, tre per ogni bobina. I CD/DVD così realizzati contengono i file in formato non compresso Broadcast Wave Format a 24 bit/96 khz e i file compressi .mp3 a 320 e 192 kbps a 24 bit/48 khz. Le copie conservative comprendono anche la documentazione fotografica dell'intero lotto in formato JPEG, file di checksum .md5 e una scheda contenente i metadati relativi

allo stato di conservazione, tipologia di supporto, velocità di lettura e annotazioni varie riguardo l'intero processo di riversamento delle bobine. Generalmente, oltre alla documentazione fotografica, il metodo messo a punto e consolidato dal CSC prevede la realizzazione di video ad alta risoluzione durante il riversamento dei nastri magnetici, così da permettere l'analisi di ogni singolo frame. Questo è utile per il monitoraggio di numerosi fattori, quali l'insorgere di fenomeni di degrado, la presenza di punti di giunta, l'eventuale presenza sulla superficie del nastro di annotazioni del compositore e/o personale qualificato. A questo proposito, partendo sia dalla forma d'onda digitale, sia dall'acquisizione video, presso il CSC sono portati avanti progetti di ricerca finalizzati all'individuazione automatica di discontinuità sulla superficie del nastro. Non è stato effettuato alcun processo di restauro dell'informazione audio digitalizzata, in quanto lo scopo del lavoro era di creare una copia virtuale fedele al documento originale. Per questo motivo l'eventuale presenza di fenomeni di disturbo e/o rumori analogici non attribuibili al sistema di riversamento è stata solo annotata in fase di monitoraggio e inclusa nei metadati. La scelta di non apportare modifiche all'informazione audio dopo la conversione A/D è stata fatta anche per quelle bobine che includono parti di silenzio, in alcuni casi lunghe anche decine di minuti, che sono state mantenute nelle copie conservative. I nastri registrati a due differenti velocità di lettura hanno comportato un doppio riversamento, nonché la creazione di file .wav e .mp3 sia per i 7,5 ips, sia per i 15 ips. In presenza di doppia velocità di lettura, inoltre, i nastri sono stati riversati nella loro interezza e non limitandosi alle parti con variazioni, così da non dover intervenire sulla forma d'onda digitale con un processo di editing.

6. CONCLUSIONI

L'Archivio del Teatro Regio di Parma costituisce un esempio importante del grande patrimonio artistico culturale presente sul territorio nazionale. Il progetto di digitalizzazione ha permesso di creare copie conservative delle bobine riversate, permettendo una maggiore valorizzazione del materiale inedito e proponendosi come parziale soluzione a problemi quali l'obsolescenza dei supporti originali. Il metodo utilizzato rispecchia un'idea molto chiara di quello che si deve intendere per documento sonoro, che va considerato nella sua interezza e non come semplice informazione audio, per quanto essa ne costituisca il principale centro di interesse. Di conseguenza particolare attenzione è stata riservata all'ottimizzazione e al restauro dei supporti originali, sia per il valore degli stessi in quanto tali, sia per contenere rischi derivanti dal sottoporre i nastri a stress meccanici durante la riproduzione. La raccolta dei metadati e delle informazioni contestuali è stata un punto fondamentale in ogni fase del lavoro, così da creare una copia conservativa che rispecchiasse il più possibile il documento sonoro nella sua totalità. Ogni informazione relativa al tipo di nastro utilizzato, marca, annotazioni sulla confezione e/o stampe

può infatti fornire importanti indicazioni al musicologo per completare l'analisi del documento sonoro. La messa a punto di una metodologia standardizzata condivisa a livello internazionale è un obiettivo verso cui devono tendere le comunità archivistiche di tutto il mondo, così da creare un network di risorse che permetta la condivisione e faciliti l'accesso al grande patrimonio artistico culturale registrato sui supporti audio.

7. BIBLIOGRAFIA

- [1] Commissione Franceschini, "Dichiarazione XXXIX", *Riv. trim. Dir. pubbl.*, (1966)
- [2] Orio N., Snidaro L., Canazza S., Foresti G. L., "Methodologies and tools for audio digital archives", *International Journal on Digital Library* 10 pp. 201–220, (2009)
- [3] Edmondson, R., "Memory of the Word: General Guidelines to Safeguard Documentary Heritage", *Parigi UNESCO*, pp. 7-9, (2002)
- [4] Smith, A., "Why digitize?", *Council on Library and Information Resources*, Washington D.C., USA, p. 13, (1999)
- [5] Bressan, F., Canazza, S., Bertani R., "Metodologie e nuovi strumenti informatici per la tutela delle memorie sonore: il caso dell'archivio storico della Fondazione Arena di Verona", *Bonfiglio-Dosio, G., Archivi VIII/2*, pp. 61-121, (2013)
- [6] Bressan F., Canazza S., "A Systemic Approach to the Preservation of Audio Documents: Methodology and Software Tools", *Journal of Electrical and Computer Engineering*, (2013)
- [7] Bressan F., Canazza S., Rodà A., Bertani R., Fontana F., "Pavarotti Sings Again: A Multidisciplinary Approach to the Active Preservation of the Audio Collection at the Arena di Verona", *Journal of New Music Research*, 42:4, pp. 364-380, (2013)
- [8] Bressan, F., Canazza S., "Digital Philology in Audio Long-term Preservation: A Multidisciplinary Project on Experimental Music", *Ginters, E. et Schumann, M., ed., Procedia Computer Science vol. 38*, pp. 48-51, (2014)
- [9] Bressan, F. "Un ambiente informatico per la gestione dei processi relativi alla conservazione attiva dei documenti sonori", *Proceedings of the 19th CIM*, pp. 34-39, (2012)
- [10] Bressan, F., "The preservation of sound archives: A computer science based approach to quality control", *PhD Thesis*, Università degli Studi di Verona (2013)
- [11] Brock-Nannestad, G., "What are the sources of the

noises we remove?", *Pichler, H., Vajda Z., ed., Proceedings of AES 20th International Conference*, New York: Audio Engineering Society, pp. 175-182, (2001)

[12] Audio Engineering Society Inc., "AES recommended practice for audio preservation and restoration, Storage and Handling, Storage of polyester-base magnetic tape", *AES22*, (1997)

RILEVAMENTI Sperimentali per la conservazione attiva dei documenti sonori su nastro magnetico: Individuazione delle curve di equalizzazione

Valentina Burini

Università degli Studi di Padova
burini@dei.unipd.it

Federico Altieri

Università degli Studi di Padova
altieri@dei.unipd.it

Sergio Canazza

Università degli Studi di Padova
canazza@dei.unipd.it

ABSTRACT

Nel contesto delle *digital music libraries*, il seguente articolo si propone di andare ad approfondire la fase di riversamento nel processo di conservazione attiva dei documenti sonori. Nell'esperienza comune, infatti, spesse volte, anche su documenti provenienti da archivi importanti, manca l'annotazione della curva di equalizzazione utilizzata in registrazione. Per garantire un riversamento consono ai dettami della recente dottrina della filologia digitale, è indispensabile, qualora questo parametro non sia indicato, avere la capacità di poterlo dedurre con una ragionevole sicurezza. A tal scopo, all'interno del Centro di Sonologia Computazionale di Padova è stato portato avanti un esperimento sulla percezione delle curve di equalizzazione applicate al segnale audio. Seguendo la metodologia MUSHRA, è stata dimostrata la differente percezione di campioni registrati e riprodotti con curva di equalizzazione corretta ed errata, offrendo diverse possibilità di apportare maggior conoscenza alla definizione del documento sonoro.

1. INTRODUZIONE

Prendendo in esame il contesto italiano nell'ultimo biennio, un riguardo al patrimonio audiovisivo si è concretizzato nella sua descrizione e valorizzazione sotto forma di Open Data, andando a definire un'ontologia caratterizzante gli istituti, i luoghi e gli eventi della cultura peninsulare, e attraverso la creazione di una serie di portali tematici [24]. Sempre in [24], si evidenzia però la mancanza ancora di un significato univoco da assegnare all'espressione "documento sonoro". Di conseguenza, compositori, musicologi, archivisti e bibliotecari sono invitati a riconsiderare il bene musicale non esclusivamente dal proprio punto di vista, ma in

Copyright: © 2016 Valentina Burini 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.

un'ottica ampia e globale che richiede una stretta collaborazione interdisciplinare all'interno di una visione condivisa e coordinata.

Di fatto, le soprintendenze archivistiche italiane si attivarono concretamente a contrastare la degradazione fisica dei supporti solo verso la metà degli anni '90 [17]. Il tentativo odierno di implementare piattaforme idonee non solo all'accesso, ma più che altro al contenimento di tutte le informazioni che il documento porta con sé, attribuisce al processo di digitalizzazione un ruolo cardine nella successiva valorizzazione del materiale archivistico, come nell'intenzione comunitaria europea verso la creazione di *digital music libraries* [5]. Concetti quali "multidisciplinarietà" e "unità documentaria" sono principi base nei progetti di digitalizzazione [7]. Diventa quindi imperativo pensare alla conservazione come lo step preliminare a qualsiasi altra attività degli istituti.

La riflessione ripropone così un modo di guardare al documento come insieme di due costituenti: il segnale audio e il supporto su cui questo risiede. Ancor meglio è distinguere tra informazione audio primaria e informazioni secondarie [10]. Non legate solo al supporto, quest'ultime sono indispensabili, in egual misura all'informazione primaria, per la trasmissione della memoria del documento stesso. Nel riversamento da un supporto a un altro è allora fondamentale mantenere l'accesso ad entrambe le informazioni, abbinate sotto forma di metadati alla copia conservativa [8]. I metadati rappresentano quindi un set di informazioni necessarie per garantire quell'unità documentaria cui prima si è accennato.

Il lavoro di riversamento richiede inoltre una completa integrazione tra una conoscenza storico-critica dell'oggetto e il saper esaminare e analizzare il supporto e il segnale in esso contenuto [10]. Da un punto di vista sperimentale e nell'ottemperanza di un processo attivo di conservazione, è dovere allora della comunità scientifica di riferimento soffermarsi ancora a riflettere su quali siano le caratteristiche che effettivamente rappresentino il documento e su come reperirle.

Giunto nel 2005 alla stesura della sua terza versione, il documento dell'International Association of Sound and audiovisual Archives (IASA), noto come IASA-TC 03

[10], insieme allo IASA TC-04 [9], resta tuttora uno dei principali testi di riferimento sulla salvaguardia del patrimonio sonoro.

Quale contributo nel reperire le informazioni caratterizzanti la futura copia conservativa, la nascente disciplina della filologia digitale, introdotta in Sezione 2.1, si aggiunge oggi alla metodologia riportata all'interno dei testi sovra citati. In Sezione 2, sono presentati infatti gli aspetti della conservazione attiva di interesse per il seguente articolo, incentrato sui documenti sonori su supporto magnetico analogico. Cercando di adoperare una delle tipiche attività di analisi filologica a sostegno delle direttive standard in letteratura, questo lavoro si focalizza su un particolare aspetto del processo di riversamento: l'individuazione delle curve di equalizzazione, di cui in Sezione 2.2.

A tal fine è stato realizzato presso il Centro di Sonologia Computazionale (CSC) di Padova un rilevamento sperimentale, presentato in Sezione 3, utilizzando il metodo di valutazione soggettiva MUSHRA [12]. Quale studio di una delle informazioni secondarie caratterizzanti il documento sonoro, l'esperimento, i cui risultati sono stati discussi in Sezione 3.5 e 3.6, conta inoltre di apportare maggiore consapevolezza alla comunità umanistica nell'adempiere gli atti di descrizione e valorizzazione del patrimonio sonoro.

2. ASPETTI DELLA CONSERVAZIONE ATTIVA DEI DOCUMENTI SONORI

2.1 Filologia digitale applicata ai documenti sonori

Nella prospettiva che aumenti il numero di casi in cui non vi sia più un supporto fisico a sostegno della relativa copia conservativa digitale, risulta ancora più urgente la necessità che quest'ultima risponda ai requisiti di autorevolezza, autenticità e completezza. Per sopprimere a tali esigenze, nasce la filologia digitale quale combinazione di due discipline: la filologia e l'informatica [2]. Quest'ultima contribuisce sia con la sua tecnologia, sia con l'apporto di un metodo e di principi tipici delle scienze. D'altra parte, la filologia offre l'opportunità di assorbire questi elementi e accettare la trasformazione che dall'incontro conseguiranno le sue tipiche attività d'analisi. Tra tutte, si pone in evidenza la comparazione che ha come principale obiettivo la ricostruzione di un documento originale a partire dal confronto tra le varianti copie dello stesso [8].

Nel momento in cui si crei inoltre un dialogo tra l'operatore e il curatore scientifico dell'archivio, può rivelarsi utile un approccio comparativo tra documenti quale quello definito in filologia. Nel caso delle audio bobine sono indispensabili per il riversamento indicazioni quali la velocità di riproduzione, la larghezza

del nastro rispetto alla dimensione delle testine, l'eventuale presenza di uno o più sistemi di *noise reduction* e l'equalizzazione applicata in fase di registrazione. In mancanza di alcuni di questi dati, l'intervento del curatore scientifico dell'archivio può esser d'aiuto nel reperirli, specialmente confrontando documenti sonori appartenenti a una stessa collezione e quindi con peculiarità affini quali ad esempio l'esser stati prodotti con la stessa macchina di registrazione. A partire dalla raccolta sotto forma di metadati delle suddette informazioni secondarie, la loro correlazione potrebbe condurre, qualora ve ne fossero i presupposti, alla deduzione di informazioni anche per quei documenti ove queste non siano rintracciabili.

2.2 Le curve di equalizzazione

Come conseguenza alle peculiarità fisiche e meccaniche del processo stesso di registrazione magnetica del suono, il flusso magnetico impresso sul supporto sotto forma di segnale audio non è lineare. Per tali motivi, in registrazione è applicata al segnale un'equalizzazione affinché la densità di flusso impressa sul supporto decresca in maniera controllata all'incrementare della frequenza [16]. La velocità di scorrimento del nastro è una delle principali variabili che influiscono sulla risposta in frequenza del segnale audio. Per questo furono designate una serie di equalizzazioni standard, ciascuna per ogni velocità prescelta. Di seguito, affinché il segnale in uscita risulti lineare, in riproduzione è necessario intervenire in maniera complementare alla fase di registrazione facendo sì che la risposta del segnale audio incrementi proporzionalmente all'aumentare della frequenza [3]. È per tali motivi che un'errata applicazione delle curve di equalizzazione in riproduzione comporta un'alterazione del segnale riversato incompatibile filologicamente con le linee guida della conservazione attiva. È da notare inoltre che gli standard di equalizzazione hanno subito modifiche nel corso degli anni. Per questo, durante la fase di riversamento, è necessario individuare sia il nome assegnato alla curva di equalizzazione sia l'anno cui si riferisce. In [9] sono riportati gli standard di equalizzazione più comuni per la riproduzione di nastri magnetici analogici ed i relativi anni di adozione.

Al fine di frazionare il problema di ricerca, si è scelto di testare i due standard di equalizzazione principalmente usati allo scorrimento di 15 ips, velocità di maggior uso comune insieme ai 7,5 ips [9]. Per tal motivo, è stata utilizzata la macchina ad impiego professionale Studer A810, le cui frequenze di transizione sono riportate in Tabella 1.

L'azione delle curve sul flusso magnetico è rappresentata dall'equazione

$$L_\Phi(f) = 10 \log \frac{1 + (\frac{F_{low}}{f})^2}{1 + (\frac{F_{hi}}{f})^2} [dB] \quad (1)$$

TAPE SPEED	TRANSITION FREQUENCIES, LOW ; HIGH (TRANSITION TIME COSTANT)	
	IEC - 1968	NAB - 1965
3,75 ips 9,53 cm/s	50 Hz ; 1800 Hz (3180 µs ; 90 µs)	50 Hz ; 1800 Hz (3180 µs ; 90 µs)
7,5 ips 19,05 cm/s	0 Hz ; 2240 Hz (∞ ; 70)	50 Hz ; 3150 Hz (3180 µs ; 50 µs)
15 ips 38,10 cm/s	0 Hz ; 4500 Hz (∞ ; 35 µs)	50 Hz ; 3150 Hz (3180 µs ; 50 µs)
30 ips 76,2 cm/s	0 Hz ; 9000 Hz (∞ ; 17,5 µs)	AES 1971 0 Hz ; 9000 Hz (∞ ; 17,5 µs)

Tabella 1. Frequenze di transizione per le varie curve e velocità dello Studer A810 [23].

ove ogni curva è caratterizzata dalle relative frequenze di transizione o, trasformando il tempo in frequenza, dalle relative costanti temporali.

L'intervento applicato in registrazione è dunque similare al comportamento di un filtro passa-basso, presentando inoltre a volte un incremento sulle basse frequenze. Anche in questo caso, verrà applicata un'azione complementare in fase di riproduzione. Alle alte frequenze, i due standard presi in considerazione presentano frequenza di taglio F_{hi} a 4500 Hz per la curva IEC del 1968 e a 3150 Hz per la NAB del 1965. Quest'ultima presenta inoltre un'amplificazione dei bassi in corrispondenza della frequenza di taglio F_{low} a 50 Hz. Per l'esperimento, si è reputato idoneo rinominare la curva IEC con la denominazione dell'originale tipo di equalizzazione (CCIR) a cui sostituì il nome all'inizio degli anni '70. Ad oggi, le curve CCIR e NAB sono conosciute con i rispettivi nomi di IEC1 e IEC2 [16].

3. L'ESPERIMENTO

3.1 Motivazioni

Per evitare eventuali perdite di dati, per necessità di fedeltà filologica e per il continuo evolversi delle tecnologie stesse di trasferimento, è innanzitutto essenziale che l'intera procedura sia documentata e riproducibile. Da questo discende l'imperativo di evitare alterazioni soggettive, limitandole alle sole copie d'accesso [20]. È essenziale altresì che durante il trasferimento non si verifichino perdite in dinamica e in risposta in frequenza, riportando tutti i parametri e le procedure utilizzati. Nel caso specifico delle audio bobine, come discusso in Sezione 2.1, vi sono molteplici caratteristiche legate al supporto e alla macchina utilizzata in fase di registrazione di cui l'operatore deve

disporre, per poi stabilire i corretti parametri di riversamento. Qualora non siano reperibili, IASA richiede all'operatore di prendere autonomamente tali decisioni [9]. Casi come la scelta della corretta curva di equalizzazione o dell'applicazione o meno di un eventuale sistema di riduzione del rumore implicano una valutazione di tipo uditivo; decisione in cui può subentrare la soggettività dell'operatore se non ulteriormente disciplinata.

Al fine di evitare in futuro un simile approccio così arbitrario, all'interno del CSC di Padova è stato progettato e realizzato un esperimento pilota di percezione uditiva su campioni d'ascolto registrati e riversati con i due standard proposti in Sezione 2.2, cercando di introdurre alcuni principi di filologia digitale all'interno della metodologia tracciata in letteratura.

Come passo iniziale, si è cercato di creare diverse versioni di un ipotetico documento sonoro riversandolo prima con l'uso della corretta curva di equalizzazione applicata in fase di registrazione e poi con l'altra a disposizione. La distinta percezione tra riversamenti corretti ed errati e una successiva analisi dei risultati ottenuti potrebbero così essere il fondamento di una metodologia comparativa deduttiva da delinearsi in futuro per la determinazione delle corrette curve di equalizzazione. Tutte le versioni create sono state correlate tra loro, sottoponendole all'ascolto di un pubblico di uditori esperti e casuali al fine di osservare la percezione o meno di differenze statisticamente rilevanti anche da parte di un pubblico non abituato a un ascolto critico, quali possano essere gli addetti alla catalogazione e valorizzazione dei documenti. In tal senso, l'esperimento potrebbe addurre maggiore consapevolezza nel recupero delle tecnicità caratterizzanti la copia conservativa, contribuendo in parte alla definizione dell'espressione "documento sonoro".

3.2 Setup sperimentale

Per l'esperimento si è fatto affidamento a uno dei metodi di valutazione soggettiva raccomandati dall'International Telecommunication Union (ITU) e dall'Audio Engineering Society (AES) [4]: l'ITU-R BS.1534, conosciuto anche come *Multi Stimulus test with Hidden Reference and Anchors* (d'ora in poi, MUSHRA) reputato, per le sue peculiarità, idoneo alle esigenze di questo tipo di esperimento [13]. Presupponendo l'eventualità che tra i vari sistemi di equalizzazione in esame vi siano piccole disuguaglianze, queste, secondo il metodo MUSHRA, verranno valutate quali medie e/o grandi disparità in rapporto tra di loro e con il riferimento dato, potendo così allenare in futuro l'orecchio dell'operatore a distinguere con più facilità. Secondo la metodologia [13], ai soggetti di test è stato chiesto di giudicare la somiglianza qualitativa dei campioni proposti rispetto a un riferimento dato e di

riportare i loro giudizi su una scala di valutazione denominata *Continuous Quality Scale, CQS* [12].

Trattandosi di un test multi-stimolo, il partecipante ha potuto scegliere a suo piacimento di alternare l'ascolto tra il segnale di riferimento e ognuno degli altri frammenti proposti nel test.

Ogni test ha avuto una durata media di circa 40 minuti, complessiva anche di una fase di *pre-test*, durata circa 20 minuti, in cui i partecipanti hanno potuto ascoltare i campioni di tutte le sessioni proposte come prescritto dal MUSHRA [12]. L'ascoltatore è stato guidato in un vero e proprio *training*, senza conoscere le peculiarità di ogni singolo campione. I frammenti sono stati presentati all'ascoltatore come versioni più o meno processate del riferimento dato, se non identiche. In ogni sessione, ripetendo l'ascolto per non più di due volte, il partecipante è stato invitato a confrontare il riferimento con il set proposto, cercando di aiutare l'apprendimento tramite confronti di difficoltà crescente.

Conclusa questa parte, dopo un opportuno momento di pausa, all'ascoltatore sono stati lasciati altri 20 minuti per lo svolgimento in autonomia del test vero e proprio.

I dati dell'esperimento sono stati raccolti ed elaborati utilizzando il software MatLab, per individuare i dati statistici di interesse e realizzare i grafici.

3.3 Partecipanti

Il testo [13] raccomanda la scelta di soggetti abituati a un ascolto critico, quali possano essere ad esempio degli studenti di musica o di ingegneria del suono. È interessante notare come esso dia nota di risultati attendibili tratti da ascoltatori inesperti a fronte di un buon *training*.

Trattandosi di un'applicazione desueta per gli ambiti d'intervento entro cui il metodo MUSHRA fu concepito, il test è da considerarsi quale esperimento pilota con due separati e conseguenti obiettivi. Da un lato vi è la necessità di verificare l'effettiva e distinta percezione dei frammenti ad equalizzazione corretta ed errata e dall'altro il poter dimostrare l'ipotesi che chiunque, anche un ascoltatore non esperto o un musicologo abituato più a uno studio delle partiture musicali, possa essere in grado di riuscirci dopo un buon *training*. Per questo si è scelto di invitare all'esperimento sia 15 ascoltatori esperti, sia 15 partecipanti casuali e di non scartare nessuno nella fase di *pre-screening* (vedi sezione 3.2 di [12]). In qualità di esperti, per le caratteristiche del percorso di studi, sono stati invitati gli studenti del Dipartimento di Nuove Tecnologie e Linguaggi Musicali del Conservatorio "Cesare Pollini" di Padova.

3.4 Materiale

Per quanto riguarda la scelta del materiale per il test, ci si è attenuti a quanto riportato ai punti 5, 5.1 e 7 della

specifica raccomandazione [12]. La metodologia prevede che tali oggetti provengano da registrazioni di puro parlato, di un misto di musica e voce e di sola musica.

Di seguito si è provveduto alla registrazione dei suddetti frammenti su nastro magnetico RMGi SM-911 da 1/4" utilizzando le curve di equalizzazione riportate in Tabella 1 come IEC e NAB per la velocità di 15 Ips. Ogni frammento è stato nuovamente riversato prima con la corretta curva applicata precedentemente in registrazione e di seguito con l'altro standard in esame. In tal modo sono state ottenute 4 versioni dello stesso campione: CCIR_on_CCIR, CCIR_on_NAB, NAB_on_NAB e NAB_on_CCIR. I set per il test sono stati completati da una copia nascosta del riferimento e da una versione chiamata *anchor*, realizzata come prescritto dalla procedura [12].

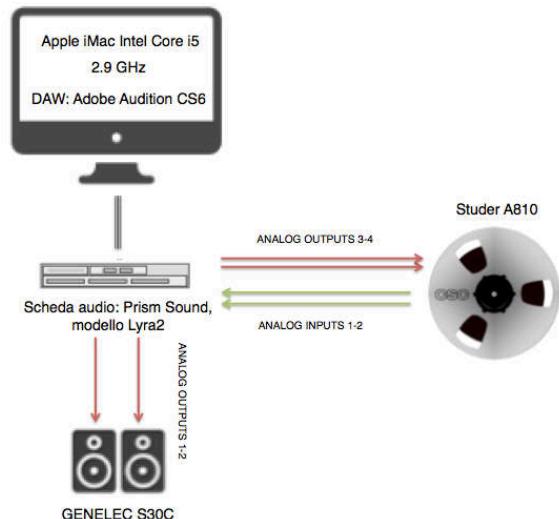


Figura 1. Signal flow della postazione utilizzata per la preparazione dei campioni d'esperimento.

I frammenti prescelti provengono da registrazioni effettuate nel biennio 2014-2015 su DAW, salvate in file WAVE stereo interleaved, con risoluzione 24 bit 48 kHz e normalizzate su un RMS medio intorno ai -18 dB rispettando così le linee guida metodologiche proposte dal documento [12]. Poiché secondo [12] il numero di campioni estratti deve essere maggiore o uguale a 5 e almeno pari a 1,5 volte il numero di sistemi sotto test (in questo caso, le due curve CCIR e NAB), si è scelto di selezionare due frammenti da circa 10 secondi ciascuno per ognuna delle tre tipologie di registrazione proposte. Per aiutare l'ascoltatore nel *training* durante la fase di *pre-test*, i frammenti proposti sono stati scelti per le seguenti peculiarità: presenza di note d'insieme tenute nei frammenti musicali, punti di enfasi recitativa nei campioni di musica e voce e presenza di sibilanti nei frammenti vocali.

La preparazione dei campioni del test è stata anticipata dalla taratura dello Studer A810 e della

strumentazione riportata in Figura 1. Per la taratura, prendendo a riferimento [1] e [5], è stato utilizzato per ogni curva un nastro MRL con ugual velocità di scorrimento e livello di flusso pari a 355 nWb/m. L'insieme dei valori esadecimali tratti dalla calibrazione è riportato in Tabella 2.

	15 ips – CCIR EQ		15 ips – NAB EQ	
REPRO	L	R	L	R
1 kHz	3A	3D	36	3E
10 kHz	E1	63	EE	6C
100 Hz	93	92	97	96
RECORD	L	R	L	R
1 kHz	2A	29	29	28
10 kHz	47	46	43	40
BIAS	2C	28	26	27

Tabella 2. Valori esadecimali di taratura delle testine di riproduzione, di registrazione e del Bias.

È stato reputato idoneo svolgere l'esperimento in una cabina silente, modello PRO 45, utilizzando cuffie audiometriche Sennheiser HDA 200 che hanno la peculiarità di essere equalizzate per una risposta in frequenza su campo diffuso come richiesto al punto 7.3 della raccomandazione [11]. Il livello del sistema di amplificazione è infatti stato controllato per restituire a -18 dB un livello di pressione sonora pari a circa 82 dBA come indicato in [11], controllando il livello di volume in uscita sia del computer, sia della scheda audio EDIROL UA-101 utilizzati.

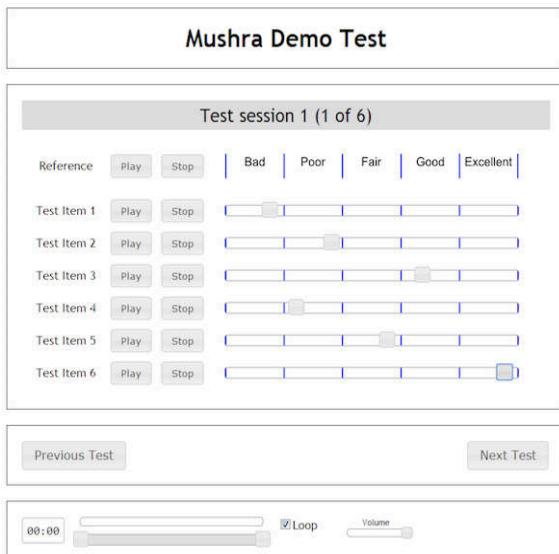


Figura 2. Presentazione esempio dell'interfaccia dell'esperimento.

L'interfaccia di test (Figura 2) è stata realizzata mediante una versione modificata dell'applicazione BeagleJS [14] per i test EBU Mushra, a cui è stata aggiunta una sezione per la realizzazione dei *pre-test* di *training*. Come

riportato alla Sezione 8 di [12], è permesso all'ascoltatore di poter aggiustare il livello di volume in un *range* di ± 4 dB rispetto a un livello di riferimento definito al punto 8.4 di [10]. L'operazione è però concessa solo nel passaggio da una sessione all'altra, poiché non è invece consentito modificare il volume d'ascolto tra un *item* e un altro all'interno di una stessa sessione.

3.5 Risultati

Si riportano di seguito le analisi dei dati effettuate su un numero totale di 30 test. Al fine di valutare la diversa percezione dei due tipi di equalizzazione proposti e delle relative combinazioni, in un primo momento è stato analizzato l'insieme dei dati raccolti dai test di tutti i partecipanti. Per quanto riguarda invece la valutazione della capacità di soggetti casuali di poter dare dei risultati attendibili a seguito di un opportuno *pre-test*, i dati sono stati rielaborati dividendo le due categorie di partecipanti tra 15 *unskilled* ('non esperti') e 15 *skilled* ('esperti') [12].

Il primo passo dell'analisi dei risultati è stato calcolare il punteggio medio \bar{u}_{jk} . I dati sono stati analizzati con il metodo *boxplot* di MatLab (vedere ad esempio Figura 3), in modo tale da evidenziare la posizione dei quartili, della mediana e degli *outlier* nelle valutazioni dei campioni di ascolto. A questi, con asterisco, è stata aggiunta la media. Sono stati così ottenuti 6 grafici, corrispondenti alle 6 sessioni del test per i due frammenti musicali, i due frammenti di musica con voce recitante e i due frammenti vocali.

Per ogni grafico, in ascisse sono riportate le sei tipologie di campione, mentre sulle ordinate è riprodotta la scala di valori da 0 a 100. Le valutazioni sono state confrontate tra di loro al fine di constatare eventuali differenze statisticamente rilevanti. I confronti sono stati realizzati mediante *t-test* per due distribuzioni indipendenti, con differenza statisticamente significativa tra le distribuzioni per $p < 0,05$: laddove $p < 0,05$, la distanza tra le distribuzioni è stata misurata mediante il calcolo del *d* di *Cohen*, nella variante del *g* di *Hegde*, interpretato in modo che per $g \approx \pm 0,2$ si parla di "distanza piccola", per $g \approx \pm 0,5$ di "distanza media", di "distanza grande" per $g < -0,8$ o $g > 0,8$.

A prima vista, osservando i valori medi per tutti e 6 i grafici, è stato osservato un andamento omogeneo in tutte e 6 le sessioni di test, con medie che tendono ai medesimi valori per gli utenti *skilled*, *unskilled* e indistinti in tutte le casistiche proposte per ogni set. Al fine di riscontrare un ascolto effettivamente critico da parte dei partecipanti, si sono osservate le medie relative alle valutazioni delle *anchor* e degli *hidden reference* per tutti i set proposti: ad esclusione dei frammenti puramente musicali, tutti i possibili raggruppamenti di utenti hanno posizionato le valutazioni rispettivamente alle estremità bassa e alta della *CQS*.

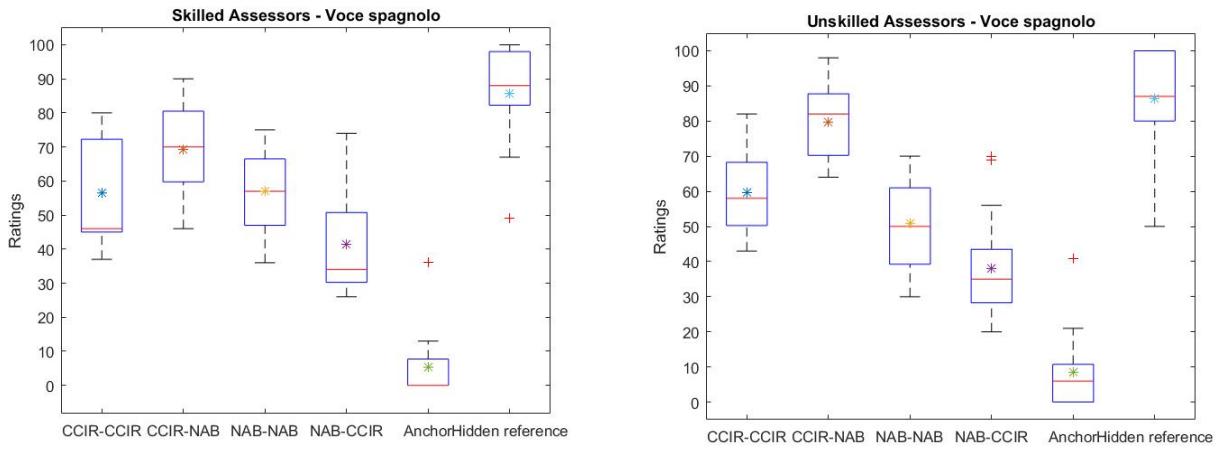


Figura 3. Grafici risultanti per la sessione “Voce spagnolo” divisi tra *skilled* e *unskilled*.

Confrontando i *rating* delle 4 varianti delle curve di equalizzazione tramite *t-test* sono stati verificati i seguenti fatti: i campioni CCIR_on_CCIR e NAB_on_NAB hanno valutazioni simili in tutti i casi per tutti i raggruppamenti di utenti ($p > 0,05$); gli utenti *unskilled* hanno evidenziato meno differenze tra le 4 combinazioni di EQ; osservando tutti gli utenti, le valutazioni sono migliori rispetto al caso *unskilled*, seppur in un caso (batt_6-11) 3 volte su 4 non si è rilevata distinzione significativa rispetto alla valutazione dell’*anchor* (Tabella 3).

battuta_6-11	CCIR_on_CCIR	CCIR_on_NAB	NAB_on_NAB	NAB_on_CCIR	Anchor
CCIR_on_CCIR	1				0,0703
CCIR_on_NAB	0,0493	1			0,0030
NAB_on_NAB	0,7816	0,0273	1		0,1012
NAB_on_CCIR	0,1330	0,0013	0,2158	1	0,4136
Anchor	0,0703	0,0030	0,1012	0,4136	1

Tabella 3. Valori incrociati di p per la sessione “batt_6-11” relativa all’analisi di tutti gli utenti, ove sono evidenziate in grassetto le differenze statisticamente significative.

L’ultimo confronto realizzato è stato tra le valutazioni dei medesimi frammenti tra le due categorie di ascoltatori. Il *t-test* produce un $p < 0,05$ solo nel caso della valutazione del frammento CCIR_on_NAB della sessione “Voce spagnolo” (con un *Hedge’s g* di 0,83).

3.6 Discussione

Dopo aver constatato nei risultati una tendenza all’omogeneità nella distribuzione delle medie delle valutazioni, create per ciascuna “classe” di ascoltatori, si pone una riflessione su eventuali relazioni con il posizionamento in frequenza dei rispettivi campioni. Osservando l’analisi spettrale, è lecito attendersi un’uguaglianza tra i campioni registrati e riversati con le

stesse curve di equalizzazione. Per le versioni spaiate è invece attesa una maggior presenza in bassa frequenza per il frammento NAB_on_CCIR, non essendo enfatizzato in riversamento il *boost* sulle basse frequenze, e una situazione inversa per CCIR_on_NAB: questo giustificherebbe alcune particolari risposte degli utenti, che hanno tendenzialmente premiato la seconda (anche rispetto ai casi di equalizzazione corretta), spesso giudicata come “più brillante”, e penalizzato la prima, reputata “più scura”. Tramite l’uso del software Audition (FFT a 16384 campioni con finestra Blackmann-Harris), si è ancora presa ad esempio la sessione “Voce spagnolo” e la si è analizzata in frequenza come riportato in Figura 4.

Si osserva subito la sovrapposizione dei campioni CCIR_on_CCIR e NAB_on_NAB, come quelli di HIDDEN_REFERENCE con ANCHOR, finché per quest’ultimo non sopraggiunge il punto d’azione del filtro passa-basso [12]. È interessante osservare l’andamento nella zona tra i 20 e i 200 Hz e tra i 3 e i 10 kHz dei campioni NAB_on_CCIR e CCIR_on_NAB, a giustificazione delle valutazioni espresse dagli utenti, e come quest’ultimo si avvicini all’HIDDEN_REFERENCE sulle alte frequenze. Il dislivello creatosi nei grafici tra questi ultimi due campioni rispetto al riferimento è stato allora giustificato dall’azione predominante del campione CCIR_on_NAB in una zona di frequenze che non solo si sovrappone a quelle del riferimento, ma comprende anche l’area di maggior sensibilità uditiva [18]. Questa analisi permette di rileggere i fatti emersi nella sezione precedente. La similitudine tra gli spettri dei campioni equalizzati in registrazione e riproduzione con la corretta curva è confermata dalla non significativa distanza delle distribuzioni statistiche delle rispettive valutazioni.



Figura 4. Analisi in frequenza (Hz in asse x; dB in asse y) del frammento “Voce spagnolo”. In corrispondenza dei 20 Hz, dal basso verso l’alto: in blu, campione CCIR_on_NAB; in rosso, campione NAB_on_NAB e sovrapposto in verde, campione CCIR_on_CCIR; in arancione, campione ANCHOR e sovrapposto in azzurro chiaro, campione HIDDEN_REFERENCE; in azzurro scuro, campione NAB_on_CCIR.

Per quanto riguarda le differenze riscontrate tra gli altri casi, gli utenti hanno rilevato differenze più nette nelle sessioni vocali rispetto a quelle strumentali, specialmente nella sessione in spagnolo, come visibile in Figura 3, e come dimostrato dagli andamenti dei valori di p del t -test (tendenzialmente oltre la soglia di 0,05 nei casi musicali).

Ciò è stato motivato dall’abbondanza di sibilanti presenti nel frammento che hanno permesso di distinguere con più facilità l’alterazione riportata dall’azione delle curve sul segnale originale nel range di frequenze tra i 3 e i 5 kHz, comprendente inoltre la zona di maggior sensibilità uditiva [18]. Infine, non si è riscontrata una differenza statisticamente significativa (ad eccezione di un caso) tra le valutazioni date dagli utenti *skilled* e *unskilled* ai vari frammenti: allo stato attuale, l’unica osservazione che si può fare è una differenza nell’affidabilità del dato, essendo le varianze delle valutazioni date dagli *unskilled* maggiori di quelle date dagli utenti *skilled*. Questo dato può essere indice di un’effettiva possibilità di poter ottenere buoni risultati di corretta classificazione anche da parte di utenti *unskilled* previo consenso addestramento: questa ipotesi è tuttavia da confermare cercando di coinvolgere nel test un campione maggiore di utenti di entrambi le classi, al fine di ottenere un’analisi statistica più solida.

4. CONCLUSIONI

All’interno del complesso progetto europeo di realizzazione di *digital music libraries*, il seguente articolo mira ad approfondire un aspetto critico del processo di conservazione: il riversamento. Analizzando

le procedure standard, è emerso come nella strutturazione del lavoro si crei il pericolo di apportare interventi di natura soggettiva, qualora non siano reperibili le informazioni necessarie all’operatore. Questo avvaluta nuovamente l’importanza di tutte quelle informazioni “secondarie” che completano l’identificazione del documento sonoro. Nello specifico, i risultati riscontrati, prendendo in esame il problema della determinazione della corretta curva di equalizzazione, hanno dimostrato come tali peculiarità leghino ancora di più il segnale audio alle specificità della macchina utilizzata per la sua registrazione, nella definizione dell’identità del documento stesso.

L’esperimento ha confermato l’apporto che diverse figure professionali possono dare al processo di digitalizzazione, non solo in termini operativi, ma anche di presa di coscienza, da parte delle istituzioni, di quali siano le esigenze dell’oggetto di loro tutela. Nello specifico delle equalizzazioni, i dati hanno mostrato una tendenza alla convergenza nei giudizi tra figure professionali nel campo dell’audio e figure non professionali, previo adeguato *training*. Questo lavoro ha quindi presentato la possibilità di addurre nuova conoscenza durante le fasi di lavorazione del processo di conservazione, a partire dall’utilizzo di forme di comparazione uditiva.

5. SVILUPPI FUTURI

L’esperimento presentato in questo lavoro è da considerarsi come pilota. Il primo obiettivo è l’arricchimento del campione di test al fine di ottenere

analisi statisticamente più solide. Queste statistiche, unite alle osservazioni degli utenti, hanno come obiettivo finale l'arricchimento del processo di conservazione mediante la definizione di una procedura standard (comprendiva di adeguato addestramento) per la deduzione delle curve di equalizzazione da applicare al segnale audio in fase di riversamento, anche nei casi dove tale informazione non sia indicata nel documento originale.

6. REFERENZE

- [1] G. M. Ballou: "Magnetic recording and playback", *Handbook for Sound Engineers*, pp. 1039-1098, 2002
- [2] F. Bressan et S. Canazza: "Digital Philology in Audio Long-term Preservation: A Multidisciplinary Project on Experimental Music", *Procedia Computer Science*, vol. 38, pp. 48-51, 2014
- [3] M. Camras: "Equalization standards for analog systems", *Magnetic recording handbook*, pp. 280-287, 1988
- [4] C. Colomes, S. Le Bagousse et M. Paquier: "Categorization of Sound Attributes for Audio Quality Assessment: A Lexical Study", *Journal of Audio Engineering Society*, Vol. 62, N. 11, pp. 736-747, 2014
- [5] Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni: "Un'agenda digitale europea", COM(2010)245 def.
- [6] L. Crane et D. Patterson: "An intro to analog tape deck calibration fun", *Tape Op: The Book about Creative Music Recording*, Vol. II, pp. 60-61, 2007
- [7] R. Edmondson: "Memory of the Word: General Guidelines to Safeguard Documentary Heritage", UNESCO, p. 2-11, 2002
- [8] G. Gigliozzi: "La filologia e l'ecdotica", *Introduzione all'uso del computer negli studi letterari*, pp. 119-127, 2003
- [9] IASA - TECHNICAL COMMITTEE: "Reproduction of Analogue Magnetic Tapes", *Guidelines on the production and preservation of digital audio objects*, 2nd ed., pp. 50-64, 2009
- [10] IASA-TECHNICAL COMMITTEE: "Ethics, principles and presevation strategy", *The safeguarding of the audio heritage*, 3rd ed., pp. 3-13, 2005
- [11] ITU-R BS.1116-1: "Reference sound field condition", *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, pp. 12-18, 1997
- [12] ITU-R BS.1534-1: *Methods for the subjective assessment of intermediate quality level of coding systems*, pp. 3-18, 2003
- [13] F. Kozamernik et G. Stoll: "EBU listening tests on Internet audio codecs", *EBU Technical Review*, pp. 4-13, 2000
- [14] S. Kraft et U. Zölzer: "BeagleJS: HTML5 and JavaScript based framework for the subjective evaluation of audio quality", *Linux Audio Conference*, pp. 2-6, 2014
- [15] E. Laukli et N. C. Stenklev: "Presbyacusis: hearing thresholds and the ISO 7029", *International Journal of Audiology*, Vol. 43, pp. 295-306, 2004
- [16] J. McKnight: "Equalization standard", *Choosing and Using MRL Calibration Tapes for Audio Tape Recorder Standardization*, pp. 2-6, 2014
- [17] R. Leydi: "Documenti sonori e ragioni della ricerca", *Archivi sonori: atti dei seminari di Vercelli (22 gennaio 1993), Bologna (22-23 settembre 1994), Milano (7 marzo 1995)*, pp. 17-20, 1999
- [18] M. Sacco: "Percezione del suono", *Imparare la Tecnica del suono*, pp. 37-41, 2005
- [19] D. Schüller: "The ethics of preservation, restoration, and re-issues of historical sound recordings", *Journal of Audio Engineering Society*, Vol. 39, N. 12, pp. 1014-1017, 1991
- [20] D. Schüller: "Preserving the facts for the future: principles and practices for the transfer of analog audio documents into the digital domain", *Journal of Audio Engineering Society*, Vol. 49, N. 7/8, pp. 618-621, 2001
- [21] A. Smith: "Why digitize?", *Council on Library and Information Resources*, p. 13, 1999
- [22] W. D. Storm: "The Establishment of International Re-recording Standards", *Phonographic Bulletin*, N. 27, pp. 5-12, 1980
- [23] Studer, "Equalizations", *Studer A810: Operating and service instructions*, p. 141, 1982
- [24] M. Tosti Croce: "Gli archivi musicali: una realtà da scoprire", *Recondita armonia: gli archivi della musica, atti del convegno internazionale di studi, Lucca (26-27-28 giugno 2014)*, pp. 55-80, 2014

LA RIDUZIONE DELLO SPAZIO. DI UNA GERARCHIA HEGELIANA NEI PRIMI SCRITTI DI STOCKHAUSEN

Michele Del Prete
Accademia di Belle Arti di Foggia
micheledelprete@virgilio.it

SOMMARIO

Oggetto di questo lavoro è la comprensione del rapporto tra musica e spazio negli scritti del primo Stockhausen. Si è scelto di concentrare l'analisi su alcuni testi degli anni Cinquanta (*Musik im Raum*, del 1958, in modo eminenti), poiché in questo periodo le riflessioni del compositore tedesco sono determinate in modo decisivo dall'esperienza dell'*inizio* della musica elettronica. Esse valgono pertanto come manifesto di un'estetica musicale che si trova a confrontarsi con possibilità del tutto nuove: secondo la percezione e le intenzioni della ricerca di quegli anni si trattava di fare musica come da un grado – se non da *un'ora* – zero.

Tra queste possibilità vi era senza dubbio quella di intervenire microscopicamente sul suono secondo modalità che solamente una certa *tecnologia della musica* permetteva: la possibilità di creare suoni di sintesi in studio permetteva al compositore di affrancarsi dalla datità dei timbri degli strumenti acustici, accomunati da spettri costruiti sulla serie delle armoniche. Per questo motivo Stockhausen considera la composizione musicale non solo come un più efficace esercizio di sovranità rispetto a parametri compositivi già noti (ad esempio come scoperta di nuove combinazioni intervallari e ritmiche di spettri prodotti *naturalmente* da fonti acustiche) ma anche come radicale estensione del dominio di tale sovranità, il cui limite superiore ideale è da pensarsi come la totalità di tutti i parametri musicali (percettivi) possibili.

In questa rifondazione dell'ordine musicale l'idea della subordinazione dello spazio al tempo – eredità, in questa forma, del tutto romantica – non solo non è messa in crisi, ma è anzi da Stockhausen del tutto rilanciata e assunta a sommo principio formale, con la conseguente *riduzione* dello spazio a parametro da modulare in funzione del tempo. Questa ripresa hegeliana è invece del tutto fuori gioco nelle coeve opere elettroacustiche di Maderna.

Copyright: © 2016 Del Prete. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUZIONE

In quella sorta di manifesto del serialismo integrale che è la prefazione al primo numero della rivista *Die Reihe* (1955) scritta assieme da Eimert e Stockhausen si afferma programmaticamente: “La musica seriale estende il controllo razionale a tutti gli elementi musicali.”² È da tale sfondo – nel contesto di un *inizio*³ – che si analizza qui il trattamento del parametro spazio, che Stockhausen intende in costante riferimento a tale totalità (totalità caratterizzata da una precisa gerarchia). Come si vedrà, l'estetica che Stockhausen sviluppa è radicata in un'ontologia ben comprensibile a partire da alcune coordinate hegeliane che definiscono lo spazio e il tempo e la loro relazione. Ai fini del presente studio il testo privilegiato per accedere a tale nucleo teoretico è stato individuato in uno scritto redatto da Stockhausen nell'ottobre 1958 dal titolo *Musica nello spazio (Musik im Raum)*.⁴ In esso si rileva *in primis* come nella musica occidentale la variabile dello spazio, definita dall'autore tramite il conio di *luogo del suono (Tonort)*,⁵ occupi il rango inferiore dell'insieme delle proprietà di un evento acustico considerate rilevanti (o meglio, secondo la prospettiva fortemente storicizzante e storicistica di Stockhausen, fino ad allora considerate rilevanti).⁶ L'ordine gerarchico (*Rangordnung*) degli aspetti di interesse musicale di un dato acustico si disporrebbe

¹ Sono grato a Pierluigi Billone per le molte discussioni di questi anni che si riflettono nel presente scritto.

² Cfr. [3], p. 7: “Die serielle Musik dehnt die rationale Kontrolle auf alle musikalischen Elemente aus.” Il periodico esce poco dopo la celebre diffusione dei *sette pezzi*, avvenuta il 19 ottobre 1954 dal Kölner Funkhaus. Il programma comprendeva queste opere: Karlheinz Stockhausen, *Studie II*; Herbert Eimert, *Glockenspiel*; Karel Goeyvaerts, *Komposition Nr. 5*; Henri Pousseur, *Seismogramme*; Paul Gredinger, *Formanten I und II*; Karlheinz Stockhausen, *Studie I*; Herbert Eimert, *Etüde über Tongemische*.

³ Quando Stockhausen, [11], p. 140, scrive che *la musica elettronica esiste dal 1953* storicizza se stesso, all'epoca attivo presso lo Studio di Colonia come primo collaboratore stabile regolarmente assunto, quale *Urvater* di tale nuovo mondo compositivo. Tutte le traduzioni dal tedesco sono dello scrivente eccetto che per [12].

⁴ Il testo è apparso per la prima volta in *Die Reihe*, nr. 5, Vienna, 1959; poi anche in [13].

⁵ Il composto (*Ton-ort*) indica una posizione spaziale assolutamente definita.

⁶ [13], p. 160.

(*scil.*: si disponeva) in questo modo:⁷ 1) Altezza (Armonia-Melodia); 2) Durata (Metrica-Ritmica); 3) Timbro (Fonetica); 4) Intensità (Dinamica); 5) Luogo del suono (Topologia).

2. UN MANCATO CONTROLLO

L'atteggiamento di profonda sottovalutazione del parametro spaziale da parte della tradizione occidentale si sarebbe conseguentemente tradotta – sempre per Stockhausen – in un'omissione di esercizio di controllo da parte dei compositori; lo spazio musicale non sarebbe mai stato parte attiva (propriamente integrante) del progetto di composizione musicale. Nei termini della pratica del serialismo integrale per come espressa da Stockhausen ciò significa che lo spazio non è mai stato organizzato musicalmente, ad esempio tramite una notazione, e che mai si è fissata con rigore la distanza degli strumenti dall'ascoltatore o la disposizione delle fonti sonore acustiche.

Tale negligenza nei confronti dello spazio (che vale immediatamente come negligenza compositiva) diventa per la prima volta manifesta e allo stesso tempo superabile grazie alle ricerche della musica elettroacustica, ovvero esattamente grazie a quella musica *Elektronisch-seriell* che consente – secondo le parole di Eimert – l'espansione della tecnica dodecafonica fino agli elementi più estremi del suono.⁸ Stockhausen minimizza polemicamente la portata delle esperienze storiche d'organizzazione spaziale per la pratica musicale a lui contemporanea⁹ e rileva come la crescente tendenza della sua epoca, in forza

della quale *tutte le idee musicali si spazializzano*,¹⁰ derivi dal generalizzarsi di una soluzione tecnica sorta nel campo della musica elettronica che è stata poi estesa al campo delle composizioni senza suoni elettronici. La disposizione spaziale degli eventi sonori (la loro provenienza da fonti diversamente dislocate) consente infatti di articolare strutture puntuali *contemporanee* – dunque temporalmente identiche – permettendo di percepire come distinte, superando in questo modo un'omogeneità (omogeneità che Stockhausen intende come indistinzione degli eventi e dunque come controllo approssimativo, e dunque nullo) tanto del materiale sonoro quanto della forma delle prime composizioni di musica elettroacustica.¹¹ La differenza tra materiale sonoro e forma vale del resto per Stockhausen unicamente come distinzione analitica: “Crediamo che la concezione che sta alla base del nostro lavoro personale potrà divenire l'idea centrale della composizione elettronica. La continueremo a sostenere e a confermare nonostante le obiezioni e le critiche: struttura dell'opera e struttura del materiale sono un'unica cosa.”¹²

È in questa prospettiva che il coinvolgimento del parametro *luogo del suono* diventa fondamentale per un processo compositivo che persegue la completa autonomia da ogni dato empirico non controllabile. Stockhausen distingue due aspetti del parametro *luogo del suono*: A) la distanza; B) il *Richtungshören* (percezione uditiva della provenienza e del moto spaziale, cioè della direzionalità del suono rispetto all'ascoltatore).

La distanza, prima proprietà del *luogo del suono*, non può propriamente essere considerata un parametro influenzabile nel caso di fonti acustiche, mentre lo è, anche se limitatamente, in quello della produzione elettroacustica, ovvero nel caso di suoni ancora *ignoti*, per i quali la stima della distanza diventa quasi impossibile¹³ e per i quali è concesso un intervento che suscita l'*illusione*¹⁴ di diversi echi.

⁷ [13], p. 160. “Die Rangordnung der Toneingeschaf ten in der bisherigen abendländischen Musik ist also die folgende: 1 Tonhöhe (Harmonik-Melodik), 2 Tondauer (Metrik-Rhythmik), 3 Tonfarbe (Phonetik), 4 Tonlautheit (Dynamik), 5 Tonort (Topik).” Si noti come, linguisticamente, tutti questi *composita* insistano sull'elemento *Ton-*, tratto che conferisce unitarietà grammatica all'insieme; la relazione con i termini greci è quella di una vera e propria germanizzazione-appropriazione (*Verdeutschung*) di categorie di pensiero dell'antichità ellenica. L'associazione di *Tonfarbe* (timbro) e *Phonetik* (fonetica) si spiega certamente anche per gli studi svolti da Stockhausen con Meyer-Eppler presso l'Institut für Phonetik und Kommunikationswissenschaft dell'Università di Bonn tra il 1954 e il 1956.

⁸ Cfr. [3], p. 11. “La musica dodecafonica conosce il rapporto di altezza (*Ton*) e serie (*Reihe*) solo come principio della costellazione fissa. Da un punto di vista elettronico-seriali invece la permutazione della serie si spinge fino all'ultimo elemento del singolo suono (*Einzelklang*). [...]” Sul timbro degli strumenti come *eine Art uralter Erscheinung* (una specie di fenomeno essenzialmente tonale) cfr. [4], p. 36; sulla questione si rimanda anche a [9], p. 23-34.

⁹ [13], p. 152: “Recentemente a proposito di alcune prime esecuzioni sono stati discussi alcuni esempi della storia musicale dell'occidente nei quali anche la distribuzione (*Verteilung*) di cori o gruppi strumentali nello spazio aveva avuto un certo ruolo. Se si esaminano con attenzione gli esempi citati si vede quanto poco questi abbiano in comune con l'odierna situazione musicale.” Gli esempi citati sono quelli della musica a San Marco dalla metà del Cinquecento a Giovanni Gabrieli – stile ricondotto agli impulsi portati dal Nord da Willaert nel 1527 e allo stile di De Rore –, la *Serenata Notturna* di Mozart (K 239) e il *Requiem* di Berlioz.

¹⁰ [13], p. 155: “[...] wir bemerken mehr und mehr, wie sich alle musikalischen Vorstellungen in zunehmendem Maße verräumlichen.”

¹¹ [13], p. 154.

¹² [12], p. 119; cfr. anche [11], p. 47, sulla “[...] contraddizione tra la natura fisica dei suoni strumentali usati fino ad allora [*scil.* fino ai primi esperimenti di musica elettronica] e le nuove concezioni formali della musica [...]” e sull'osservazione per cui “[...] struttura del materiale e struttura dell'opera devono essere identiche.” Dalla stessa esigenza di autonomia (*scil.* autodeterminazione in senso idealistico del materiale compositivo) deriva la scelta di rifiutare in questa fase suoni già pronti come quelli del Melochord o del Trautonium per la composizione della *Studie I*; cfr. [16], p. 52.

¹³ La stima della distanza richiede infatti l'essere in grado di riconoscere la deformazione spettrale di un dato suono causata dalle riflessioni; cfr. [13], pp. 164-166; [14], p. 386.

¹⁴ [13], pp. 166-7: “A partire da questa esperienza si può quindi suscitare l'illusione di qualsivoglia distanza del suono (*Schallentfernung*) tramite processi di riverberazione (*Schallereignisse*) prodotti elettronicamente.”

Il *Richtungshören* (tanto per la musica strumentale¹⁵ quanto per quella elettronica) è invece una grandezza sulla quale può intervenire direttamente la pianificazione del compositore. Stockhausen rimanda all'esperienza delle *abitudini dell'orientamento*¹⁶ riferendosi con questo ad un orientamento *naturale* che procede da sinistra a destra e dal basso verso l'alto. Tale orientamento *naturale* deve valere come punto di partenza per gli interventi di controllo dello spazio. Lo stesso autore introduce quindi il concetto di *Raum-Melodie* per designare un'organizzazione delle fonti sonore nello spazio producibile analogamente a quella che riguarda i parametri di altezza, durata, timbro ed intensità e che con questi parametri è continuamente in relazione.¹⁷ Quale esempio pratico di un tale intervento sul parametro spaziale in una composizione elettronica l'autore rinvia alla propria opera per voce di fanciullo (incisa) e suoni elettronici *Gesang der Jünglinge* (la cui prima assoluta avvenne a Colonia il 30 maggio 1956), lavoro acusmatico che prevedeva l'impiego di cinque gruppi di diffusori.¹⁸

3. ESERCIZIO DEL DOMINIO DELLO SPAZIO

Al fine di esercitare un controllo sempre più efficace sul parametro del *Richtungshören* Stockhausen propone un ulteriore tipo di intervento: il parametro *luogo del suono* (*Ton-ort*) diventa maggiormente influenzabile se lo *spazio* (*Raum*) in cui uno specifico *luogo del suono* è inserito viene *preventivamente*¹⁹ sottoposto ad un'operazione di messa a norma. Si noti come già da queste caratterizzazioni dello spazio come ordine di contenuti e contenitori traspaia una profondissima affinità con la trattazione dello spazio svolta da Aristotele nel quarto libro della *Fisica*²⁰ (e di rimando una fondamentale estraneità nei confronti dell'indescrivibile complessità della *chôra* platonica di cui si legge nel *Timeo*: uno spazio che trascende ogni programma, ogni *misura*

¹⁵ A proposito della distribuzione nello spazio di gruppi orchestrali distinti Stockhausen fa i nomi di Boulez, Poussier e Berio, [13], p. 158.

¹⁶ [13], p. 169: anche all'originale *Gewohnheiten der Orientierung* è sotteso il rapporto con lo spazio dell'abitare (*Wohnen*).

¹⁷ [13], p. 170. È superfluo ricordare che tale conio richiama le considerazioni svolte da Schönberg nella parte finale della sua *Harmonielehre* (1911) circa l'organizzazione dei timbri secondo una logica simile a quella vigente per le altezze (cfr. [17], p. 503 *sqq.*). Più tardi Stockhausen parlerà anche di *Raummelodik* e *Raumharmonik*, [14], p. 381.

¹⁸ [13], p. 153. Stockhausen richiama l'attenzione sulla disposizione delle fonti di quest'opera anche in [12], p. 119.

¹⁹ Sulla categoria del preventivo quale livello di intervento per il controllo del manifestarsi di un evento si rimanda a E. Severino, [10], pp. 13-64 (in particolare alla sezione XV, p. 46 *sqq.*). Per una lettura politica del tema e il suo rapporto con lo spazio – a partire da W. Benjamin – sia concesso rimandare a M. Del Prete, *Über den Ausnahmezustand des Raums im Passagenwerk*, intervento tenuto al congresso *Topographie der Erinnerung*, Auditorium dell'Institut National d'Histoire de l'Art, Galerie Vivienne, Parigi, il 15 giugno 2005.

²⁰ Cfr. la definizione di *topos* a partire da [18], 212a19-20.

di contenzione e che rimanda ai registri del dono e dell'ospitalità).

L'individuazione o la costruzione di sale da concerto adatte ad una *musica spaziale* (*Raum-musik*) vera e propria costituisce un problema decisivo: dalle premesse esaminate appare infatti contraddittorio servirsi di sale da concerto tradizionali o di altri ambienti preesistenti – dunque extrasistemici – dai tempi di assorbimento e di riverberazione del tutto casuali. Stockhausen propone allora spazi trasformabili (*umwandelbare Räume*) che presenta quale modello supremo di sala da concerto.²¹ Le proprietà più importanti di tali spazi sono elencate in dettaglio come segue:²²

1. Spazio circolare o quadrato per consentire disposizioni dell'orchestra (*Orchester-aufstellungen*) in qualsiasi posto (*Stelle*) si desideri attorno e/o in mezzo agli ascoltatori.
2. Nessun podio fisso; di contro un gran numero di podi mobili (praticabili).
3. Pavimento piano.
4. Disposizione dei posti a sedere modificabili a piacere; nessuna struttura fissa di posti a sedere.
5. Prese per altoparlanti e microfoni ovunque alle pareti e al soffitto.
6. Nicchie nei muri e/o balconi a diverse altezze per piccoli gruppi strumentali.
7. Porte che non impediscono la disposizione circolare di gruppi orchestrali lungo le pareti (possibilmente molte porte distribuite a raggiere all'interno dello spazio circolare).
8. Condizioni di riflessione controllate elettricamente in modo da poter essere adattate alle specifiche condizioni di esecuzione.
9. Cabina di regia (*Schallstudio*) esterna alla sala per la diffusione del suono (*Einspielung*) tramite altoparlanti o per registrazione.
10. Illuminazione indipendente della sala e delle lampade da leggio trasportabili.
11. Non poltroncine di velluto ma sedie di legno.

²¹ In questo modo il modello sociologico e politico della sala da concerto non solo non viene minimamente criticato ma vale piuttosto come presupposto indiscutibile dell'analisi di Stockhausen. Il *confine della trasformabilità* di questo spazio è un confine del tutto interno – interno ad uno spazio contenitore perfettamente aristotelico cosmologicamente finito (e antropologicamente blindato).

²² [13], p. 157-158.

In seguito vengono discussi in dettaglio due modelli di pianificazione dello spazio che mirano a definire la posizione migliore per l'ascolto.²³ Migliore significa qui: funzionale alla programmazione spaziale del compositore, ovvero idealmente *neutral-equidistante* rispetto alla collocazione delle fonti sonore.

Dapprima Stockhausen considera il caso particolare della disposizione di tre fonti sonore disposte presso i vertici di un triangolo equilatero: l'ascoltatore è pensato come collocato all'interno di tale triangolo in prossimità del punto medio della base.

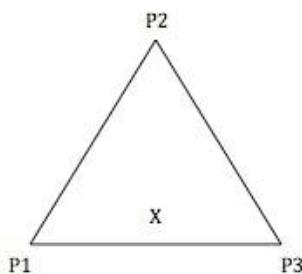


Figura 1. Esempio originale in [13], p. 172.

Questa proposta ha tuttavia un valore del tutto provvisorio, come si può osservare anche da un'imprecisione che Stockhausen commette a questo punto della sua trattazione, sostenendo che nel caso di emissioni successive di un suono dalle posizioni discrete P1, P2 e P3 la variazione P2-P3 sarebbe identica a quella di P3-P1. Questo però – poiché il discorso riguarda qui la percezione e non la geometria – sarebbe vero solo se l'ascoltatore si trovasse nel punto perfettamente centrale (baricentro=ortocentro) del triangolo dato, cosa non contemplata dall'ipotesi. Stockhausen procede inoltre qui da due assunti impliciti: che l'ascoltatore sia rivolto verso il vertice superiore del triangolo (il che introduce una componente di orientamento ben distinta) e che le tre fonti siano collocate alla medesima altezza (e nulla è infine specificato circa il rapporto tra detta altezza e l'altezza della posizione dell'ascoltatore).²⁴

Appare pertanto del tutto prevedibile che Stockhausen proponga quindi il *cerchio* (un *continuum*) quale forma ideale di sala da concerto e ne elegga il centro a miglior posizione possibile per l'ascoltatore. Stockhausen ricorre poi nuovamente all'idea di una disposizione naturale di sinistra e destra, divide il cerchio in due semicerchi e adotta la misurazione in gradi per definire le posizioni dei punti sulla circonferenza che sono poi assimilate a quelle delle diverse sorgenti sonore. Stockhausen procede implicitamente anche in questo caso dall'assunto per cui l'ascoltatore sia rivolto verso 0°

²³ [13], p. 170 *sqq.*

²⁴ [13], p. 170.

e consideri di conseguenza la posizione diametralmente opposta di 180° come angolo morto della percezione della sinistra e della destra.²⁵ Tale partizione dello spazio deve spingersi a identificare la soglia percettiva in forza della quale un ascoltatore è ancora in grado di distinguere due fonti sonore molto vicine come distinte. Il compositore tedesco conclude il suo articolo con un accenno alla possibilità di coinvolgere anche i parametri del *sopra* e del *sotto*, eventualità che però viene considerata di difficile attuazione poiché – come egli stesso scrive – *non è stato possibile fare alcuna esperienza in ambito musicale entro queste dimensioni*.²⁶

Stockhausen formula questa affermazione in senso assoluto, e pare trascurare (negare) il valore di una serie di esempi storici che attestano tali esperienze: gli organi (uno o più) posti almeno dal XIV secolo ad un livello superiore rispetto a quello a cui si trova l'assemblea, la collocazione degli *Stadtpfeifer* sulle torre cittadine in diverse regioni di lingua tedesca, la disposizione dei cori e degli organi su piani di differente altezza nella penisola iberica (notevoli i casi delle cattedrali di Valencia e di Granada, del monastero di San Lorenzo de El Escorial, del Palácio Nacional di Mafra) e a Venezia²⁷ – e in questo ultimo caso non solo a San Marco: un esempio iconografico significativo citato da Nono, il dipinto *Vestuario di una Nobil Dama Veneta a San Lorenzo* di Gabriele Bella, attesta la disposizione dei musicisti della chiesa di San Lorenzo in doppio coro su tre diversi piani di altezza.²⁸

È dunque lo spazio sferico, che coinvolge la dimensione orizzontale e verticale con la stessa compattezza, a essere proclamato modello di riferimento per ogni spazio adatto a esecuzioni musicali; rivolgendosi ad ascoltatori ideali, Stockhausen ne vanta in questo modo i pregi: “Potrete ascoltare una musica composta per tali *spazi normati* [genormte Räume] da sopra, da sotto e da tutti i punti cardinali”.²⁹

Tale processo di *messa a norma* degli spazi per la musica (*Normung der Musikräume*)³⁰ che si compie tramite la realizzazione di uno spazio sferico è dunque la soluzione per risolvere ciò che Stockhausen, come visto, ritiene un duplice assurdo: non solo il fatto per cui il parametro spaziale è stato trascurato dalla pratica musicale, ma anche quello per cui l'interprete è stato spesso addirittura costretto ad adattarsi allo spazio in cui si trovava ad operare – fatto che ancora una volta attesterebbe l'insufficienza di un pensiero musicale incapace di autodeterminarsi in modo assoluto.

Risulta dunque chiaro che la concezione e la ricerca spaziale del *live electronics* di Nono – per

²⁵ [13], p. 173.

²⁶ [13], p. 175.

²⁷ Questo anche se la pratica musicale a San Marco tra Willaert e Giovanni Gabrieli era stata brevemente citata in [13], cfr. *supra*, nota 9.

²⁸ Cfr. [8], p. 58-59; è superfluo ricordare qui l'esperienza del *Prometeo. Tragedia dell'ascolto*, la cui prima assoluta ha avuto luogo il 25 settembre 1984 proprio a San Lorenzo.

²⁹ [13], p. 153.

³⁰ [13], p. 163.

tacere delle pratiche della *prospettiva ecosistemica* di Di Scipio che si rivolge ad un suono che sorge da spazi idiomatici e contingenti – si pone agli antipodi di quella che Stockhausen espone in modo inequivocabile come segue:

“Allo stesso modo oggi non è più quasi concepibile che un interprete adatti il timbro dello strumento a condizioni di esecuzione mutevoli e per la stessa parte scelga di volta in volta uno strumento diverso della stessa famiglia con un timbro più chiaro o più scuro. La prassi dei Kappelmeister per cui si prepara – *in corrispondenza dello spazio* – un pezzo diminuendo, aumentando, combinando altrimenti o addirittura sostituendo gli strumenti è del tutto passata di moda [ist ja ganz außer Gebrauch gekommen].”³¹

4. LO SPAZIO COME FUNZIONE DEL TEMPO

A questo punto è decisivo stabilire la *natura* della norma dello spazio normato. Sono dati due casi, il cui significato è da discutere secondo il criterio di autonomia come autodeterminazione assoluta del materiale, parametro elementare o forma che sia, che Stockhausen considera essere il cardine della futura pratica compositiva. La norma dello spazio può essere di natura *riflessiva* (e dunque determinata dallo spazio stesso: lo spazio *si norma*, medialmente) oppure di natura *eteronoma* (nel qual caso lo spazio *si riduce* ad un soggetto del tutto passivo, poiché viene normato in funzione di – ovvero *da* – qualcosa di diverso). Per le coordinate fornite da Stockhausen è chiaro che il problema della messa a norma dello spazio deve essere considerato nel quadro della totalità di tutti i parametri compositivi elencati. In termini filosofici esplorare tale insieme di relazioni a queste condizioni (indicandone il senso, ovvero identificando ciò per cui e/o in vista di cui si norma) corrisponde a impostare il problema del *sistema* secondo coordinate idealistiche: corrisponde cioè a interrogarsi sull’idea stessa della libertà *relatione* di termini distinti.

Deduciamo la struttura della gerarchia della totalità in questione (giungendo all’individuazione del principio normante) partendo da un testo di Stockhausen redatto nel 1961 intitolato *L’unità del tempo musicale*. Qui si può leggere: “Già in uno stadio relativamente precoce della composizione elettronica ho riflettuto se non fosse possibile far corrispondere all’unitarietà della percezione un’unitarietà della composizione del suono e della realizzazione del suono [*Klang-Komposition; Klang-Realisation*]. Solamente a partire dai lavori preparatori per la composizione elettronica ‘Kontakte’ ho trovato delle strade per riportare tutte le proprietà [*scil:* del suono] ad un unico denominatore. Sono partito dal fatto per cui le *differenze della percezione acustica* sono in realtà [*doch*] tutte ricondu-

cibili a *differenze della struttura temporale delle oscillazioni*.”³²

Il contesto della riflessione (la composizione di musica elettronica) segnala qui, come nei casi precedenti, una decisa volontà di teorizzazione da parte di Stockhausen, che giunge effettivamente ad enucleare il principio grazie al quale forma e materia musicale possono essere pianificati al meglio (e pianificati in ogni senso): questo principio, presente sia al microlivello fisico di ogni fenomeno acustico sia al macrolivello di una forma musicale, è il tempo. Il tempo è dunque l’orizzonte ultimo del senso musicale, e come tempo è svelabile e semplificabile *ogni* aspetto dei fenomeni acustici e musicali.

A partire da questo dato si ottiene la chiave di lettura per interpretare anche il caso particolare del parametro *spazio* (che per Stockhausen, ancora una volta,³³ non può essere se non in relazione al tempo) per come è caratterizzato in *Musik im Raum*. L’omogeneità-indistinzione, intesa come difetto esiziale del discorso musicale elaborato anche in alcune produzioni della musica elettroacustica – omogeneità che era stata definita da Stockhausen come la simultaneità di eventi indistinguibili –, vale esattamente come fallimento dell’articolazione di *fasi temporali di grandi proporzioni*.³⁴ Ma una *buona* spazializzazione, come visto, consente che una forma musicale (di un lavoro elettroacustico, ma non solo) sia coerente, ovvero si dia come *efficace organizzazione del tempo*. Stockhausen considera *in primis* il problema della discriminazione di un dato acustico da un punto di vista percettivo: tuttavia questa estetica-gnoseologia rimanda ontologicamente al problema dell’identità degli indiscernibili – e le estreme conseguenze di questa posizione sono chiare: discernibili sono solo eventi non contemporanei, il che equivale a dire che due eventi *perfettamente contemporanei* (per *qualità* di inizio, durata e fine) non sono che il *medesimo ente*.

Il controllo dello spazio viene dunque considerato da Stockhausen solo nella sua ancillarità, in quanto utile all’articolazione dell’arciparametro tempo: l’organizzazione dello spazio è sempre compresa (alloggiata) in una strategia temporale che allo spazio è sovraordinata. *Non per lo spazio, ma per il tempo si norma lo spazio: lo spazio è ciò su cui si costruisce il buon accadere del tempo – come l’accadere di eventi temporali distinti*.

Stockhausen riscontra tale rapporto tra spazio e tempo anche a livello di fisica acustica: “Possiamo allora dire che l’intensità è una proprietà spaziale del suono nella misura in cui determina l’ampiezza delle onde sonore (maggiori è la distanza, minore diviene la forza di pressione sonora) ma anche che la nostra percezione del ‘vicino’ e del ‘lontano’ si orienta principalmente in base alla composizione spettrale del suono – ovvero alle sue proprietà temporali – che viene influenzata dalla combinazione di

³² [15], p. 212.

³³ Cfr. [2], *passim*.

³⁴ [13], p. 154.

³¹ [13], p. 163. Sottolineatura dello scrivente.

intensità e condizioni spaziali [*räumliche*] di trasmissione (modulazione, deformazione).”³⁵

5. CONCLUSIONI

I presupposti filosofici di questo atteggiamento di Stockhausen sono di enorme portata. Essi richiamano potentemente alcune figure hegeliane, per cui la verità dello spazio è il tempo. Canonico è a questo proposito l’*Aggiunta* al § 257 dell’*Enzyklopädie*, dalla quale ogni riflessione sul senso teleologico del rapporto (gerarchia) di spazio e tempo deve qui muovere.³⁶ L’*Aggiunta* è inserita nella prima sezione (*La Meccanica*) della seconda parte dell’*Enzyklopädie*, parte che è dedicata alla filosofia della natura. Come è noto, la sezione *La Meccanica* si compone di tre unità: la prima (A) è strutturata a sua volta in tre parti: a) *Spazio e Tempo*, b) *Lo Spazio*, c) *Il Tempo*. L’*Aggiunta* esplicita la relazione tra spazio e tempo come una progressione di verità. Hegel insiste dapprima sul processo di negazione dello spazio, che è un processo imperfetto, perché avviene secondo il ritmo, *in primis* del tutto intrasaziale, di punto, linea e superficie. Lo spazio non esce da se stesso, la sua dialettica è limitata, e per questo Hegel nota: “Lo spazio è l’immediata qualità esistente, nella quale tutto resta mantenuto, perfino il confine ha la natura di un mantenere: questo è il difetto dello spazio. Lo spazio è questa contraddizione, avere in sé una negazione, ma in modo che questa negazione decade in un mantenere indifferente.” La vera vita dello spazio è allora possibile solo grazie al tempo, che garantisce il superamento di tale *negazione paralizzata*, che non esce da se stessa. È nel momento³⁷ stesso in cui questa estasi avviene (ovvero continua ad avvenire) che si ha l’irruzione del tempo come vita e verità dello spazio: “Nello spazio la superficie è certo negazione della negazione: ma secondo la sua verità è diversa dallo spazio. La verità dello spazio è il tempo, così lo spazio diviene tempo [...].”³⁸ La scrittura filosofica di Hegel procede come linea che muove *dallo spazio al tempo*, con un gesto che rileva la *disposizio* del quarto libro della *Fisica* di Aristotele e poi l’esordio (o meglio gli esordi) della *Critica della*

³⁵ [13], p. 166.

³⁶ [5], p. 48: “Der Raum ist die unmittelbare daseiende Qualität, worin alles bestehen bleibt, selbst die Grenze die Weise eines Bestehens hat; das ist der Mangel des Raums. Der Raum ist dieser Widerspruch, Negation an ihm zu haben, aber so, daß diese Negation in gleichgültiges Bestehen zerfällt.”

³⁷ *Momento* è una parola essenziale nella spiegazione del rapporto tra tempo e spazio. In sé *momento* è già termine legato tanto allo spazio (ad esempio per la descrizione di una forza, di una velocità) quanto al tempo. Cfr. [5], p. 48: “Da der Raum also nur diese innere Negation seiner selbst ist, so ist das Sichaufheben seiner Momente seine Wahrheit; die Zeit ist nun eben das Dasein dieses beständigen Sichaufhebens, in der Zeit hat der Punkt also Wirklichkeit.”

³⁸ “Im Raume ist die Fläche zwar Negation der Negation; aber ihrer Wahrheit nach ist sie vom Raum unterschieden. Die Wahrheit des Raumes ist die Zeit, so wird der Raum zur Zeit [...]”. In [5], p. 48.

Ragion Pura di Kant, secondo una *ripresa* il cui senso è stato mostrato con assoluta forza da un fondamentale testo di Derrida.³⁹ Questo breve accenno a tali coordinate ontologiche fondamentali rende evidente quanto *metafisica* sia la posizione di Stockhausen circa il rapporto tra spazio e tempo.

Si deve però infine rilevare – senza la necessità di ricorrere a più tarde (e diversissime) ricerche sullo spazio condotte da Nono o da Di Scipio alle quali si è già rinviaiato – che una liberazione dello spazio come *irriducibile* al tempo (alla verità del tempo) avviene già all’epoca a cui risalgono gli scritti di Stockhausen qui considerati. L’ambito è ancora quello degli inizi della musica elettroacustica: è con *Musica su due dimensioni* di Bruno Maderna (opera per nastro e flauto del 1958 inserita in un cammino intrapreso nel 1952), che s’inaugura l’orizzonte di una ricerca musicale sullo spazio radicalmente diversa da quella di Stockhausen. In Maderna la mantenuta differenza *degli spazi* – la distanza di *due* dimensioni fisiche e percettive – eccede (*trascende*) i limiti dell’onnivora contemporaneità *sistemática* di Stockhausen spalancando un *diasistema* che risulta inassimilabile ad ogni forma di *sistema* cronicizzante. Si tratta di una differenza del tutto assente nel *Gesang*, dove le fonti sonore – pur diverse in partenza: suoni elettronici e voce di fanciullo – sono fuse in un nastro che è spazio di un *medesimo* che *si produce* come tempo ben formato.

Ma con *Musica su due dimensioni* si eccede un discorso sullo spazio dall’impianto aristotelico, e si giunge alle soglie di un discorso che è stato sopra solo accennato (e che è forse solo accennabile): quello, complesso e irriducibile *anche* al tempo (al *senso del tempo*)⁴⁰ della spazialità imprendibile, gratuita e donante *del dare luogo* della *chōra* platonica.⁴¹

6. BIBLIOGRAFIA

- [1] J. Derrida, “*Chōra*”, in *Poikilia. Études offertes à Jean Pierre Vernant*, EHESS, Parigi, 1987, pp. 265-296; tr. italiana in J. Derrida, “Il segreto del nome”, a cura di G. Dalmasso e F. Garritano, Milano, Jaca Book, 1997, pp. 41-86.
- [2] J. Derrida, “*Ousia et grammé*”, in *Endurance de la pensée: Pour saluer Jean Beaufret*, Paris, Plon, 1968, pp. 31-78; tr. italiana in J. Derrida, “Margini della filosofia”, a cura di Manlio Iofrida, Torino, Einaudi, 1997, pp. 59-104.
- [3] H. Eimert, “*Die sieben Stücke*”, in *die Reihe. Information über serielle Musik*, a cura di H.

³⁹ [2].

⁴⁰ Per questo Derrida osserva in [1], p. 51: “La *chōra* è *anacronica*, «è» l’anacronia nell’essere, o, meglio, l’anacronia dell’essere. Essa anacronizza l’essere”.

⁴¹ Cfr. [1], pp. 52-57 e *passim*.

- Eimert, K. Stockhausen, Nr. 1, 1955, Wien-Zürich/London, Universal Edition, pp. 8-13.
- [4] P. Gredinger, "Das Serielle", in *die Reihe. Information über serielle Musik*, a cura di H. Eimert, K. Stockhausen, Nr. 1, Wien-Zürich/London, Universal Edition, Nr. 1, 1955, pp. 34-41.
- [5] G.F.W. Hegel, *Enzyklopädie der philosophischen Wissenschaften im Grundrisse I* (riprende l'edizione del 1830), *Theorie Werke Ausgabe IX*, E. Moldenhauer und K. M. Michel, Frankfurt a. M., Suhrkamp, 1970.
- [6] B. Maderna, *Musica su due dimensioni*, Milano, Suvini Zerboni, 1960.
- [7] B. Maderna, *Musica su due dimensioni* (1952), a cura di Nicola Scaldaferrri, Milano, Suvini Zerboni, 2003.
- [8] L. Nono, "Verso Prometeo", conversazione tra Luigi Nono e Massimo Cacciari raccolta da Massimo Bertaggia, in *Verso Prometeo*, Milano, Ricordi, 1984, p. 23-38.
- [9] L. Rognoni, "La musica elettronica e il problema della tecnica", in *Fenomenologia della musica radicale*, Milano, Garzanti, 1974, p. 23-34.
- [10] E. Severino, *Legge e caso*, Milano, Adelphi, 1990³.
- [11] K. Stockhausen, "Elektronische und instrumentale Musik", in K. Stockhausen, *Texte zur elektronischen und instrumentalen Musik*, Köln, DuMont Dokumente, Band 1, 1963, pp. 140-151.
- [12] K. Stockhausen, "Problemi attuali (a proposito di 'Gesang der Jünglinge')", in *La musica elettronica*, a cura di H. Pousser, Milano, Feltrinelli, 1976, pp. 111-120.
- [13] K. Stockhausen, "Musik im Raum", in K. Stockhausen, *Texte zur elektronischen und instrumentalen Musik*, Köln, DuMont Dokumente, Band 1, 1963, pp. 152-175.
- [14] K. Stockhausen, "Vier Kriterien der Elektronischen Musik", in K. Stockhausen, *Texte zur Musik*, 1970-1977, Köln, Du Mont Dokumente, Band 4, 1978, pp. 360-401.
- [15] K. Stockhausen, "Die Einheit der musikalischen Zeit", in K. Stockhausen, *Texte zur elektronischen und instrumentalen Musik*, Köln, DuMont Dokumente, Band 1, 1963, pp. 211-221.
- [16] K. Stockhausen, "Komposition 1953 Nr. 2, Studie I, analisi", in *La musica elettronica*, a cura di H. Pousser, Milano, Feltrinelli, 1976, pp. 51-64.
- [17] A. Schönberg, "Harmonielehre", [1911], Vienna, Universal, 1997³, tr. italiana di G. Manzoni, Milano, Il Saggiatore, 1953.
- [18] Aristotele, "Physica", a cura di W. D. Ross, Oxford, Clarendon Press, 1936; tr. italiana di R. Radice, Milano, Bompiani, 2011.

STIMA DI FEATURE SPETTRALI DI HRTF MEDIANTE MODELLI ANTROPOMETRICI NON LINEARI PER LA RESA DI AUDIO 3D

Simone Spagnol

University of Iceland
Reykjavik, Islanda
spagnols@hi.is

Silvio Galessio, Federico Avanzini

Università di Padova
Padova, Italia
avanzini@dei.unipd.it

SOMMARIO

La relazione tra i parametri antropometrici di un soggetto umano e le feature tipiche delle Head-Related Transfer Function (HRTF), in particolare quelle collegabili al padiglione auricolare (o *pinna*), non è compresa appieno. In questo articolo applichiamo tecniche di elaborazione del segnale per estrarre le frequenze del primo notch dovuto alla pinna (conosciuto come N_1) nella porzione frontale del piano mediano e costruiamo un modello basato su una rete neurale artificiale che relazioni le frequenze stesse a 13 diversi parametri antropometrici della pinna, alcuni dei quali dipendono dall'elevazione della sorgente sonora. I risultati mostrano una corrispondenza incoraggiante tra l'antropometria e le feature spettrali, la quale conferma la possibilità di predire la frequenza centrale del notch a partire da una semplice fotografia dell'orecchio.

1. INTRODUZIONE

La maggior parte delle tecniche per la resa binaurale del suono fa affidamento sull'utilizzo di Head-Related Transfer Function (HRTF), ovvero filtri che catturano gli effetti acustici del corpo umano [1]. Le HRTF permettono una simulazione realistica del segnale che giunge all'ingresso del canale uditivo in funzione della posizione della sorgente sonora nello spazio. La soluzione ideale, ovvero quella che meglio approssima l'ascolto spaziale reale, richiede l'utilizzo di HRTF individuali misurate acusticamente sull'ascoltatore stesso [2]. Tuttavia, la misurazione acustica di HRTF richiede una strumentazione generalmente costosa e procedure di registrazione invasive [3]. Tale è la ragione per cui in pratica viene spesso favorito l'utilizzo di HRTF non individuali, misurate su manichini o su altri individui. Lo svantaggio di tali HRTF è che difficilmente sono relazionabili all'antropometria dell'ascoltatore, in particolare quella della pinna. Ciò si traduce in probabili e frequenti errori di localizzazione, quali inversioni *front/back*, errata percezione dell'elevazione, e localizzazione all'interno della testa [4].

Diverse tecniche per il design di HRTF sintetiche sono state proposte nelle ultime due decadi per affrontare tali problemi. Una delle più promettenti è la modellazio-

ne strutturale [5]. Secondo questo approccio, gli effetti acustici più rilevanti per la percezione spaziale del suono (ritardi ed effetti d'ombra dovuti alla testa, riflessioni sui bordi della pinna e sulle spalle, eccetera) vengono separati e modellati ciascuno con una struttura di filtri digitali. I vantaggi di tale approccio rispetto ad altre tecniche per la resa binaurale del suono sono la possibilità di personalizzare le HRTF sull'ascoltatore, grazie all'acquisizione di quantità antropometriche (raggio della testa, forma della pinna, larghezza delle spalle, e così via), e l'efficienza computazionale, in quanto ogni modello è sottostrutturato in diversi blocchi ciascuno dei quali simula un singolo effetto acustico. Ciononostante, la precedente letteratura riguardante la relazione tra effetti acustici ed antropometria - tra cui applicazioni di metodi di regressione su database di HRTF [6, 7, 8, 9] - hanno prodotto risultati controversi, evidenziando come molte di queste relazioni non siano tuttora comprese appieno.

Possiamo ipotizzare due principali cause del fallimento di questi studi. In primo luogo, nessuna informazione risultante dalla conoscenza pigna della componenti strutturali responsabili degli indicatori di localizzazione è presa in considerazione: molti lavori applicano ciecamente tecniche di apprendimento automatico su lunghi vettori di feature antropometriche, i quali includono molti parametri irrilevanti. In secondo luogo, l'intera HRTF o una versione ridotta dimensionalmente della stessa (ad esempio tramite *Principal Component Analysis*) viene usata come l'insieme di variabili *target*, senza l'applicazione di alcun passo di *preprocessing* che estragga le caratteristiche locali più importanti. In particolare, è risaputo che i minimi (notch) e i massimi (picchi) locali nella funzione di trasferimento sono essenziali per la percezione sonora della dimensione spaziale più "individuale", ossia l'elevazione [10].

L'obiettivo principale di questo articolo è esplorare la relazione tra le frequenze centrali dei notch in un insieme di HRTF misurate nella porzione frontale del piano mediano e un insieme di parametri antropometrici sotto forma sia di misure unidimensionali della pinna (altezza della pinna, larghezza della conca, eccetera) che di misure variabili con l'elevazione della sorgente sonora (ovvero le distanze tra il canale uditivo e i contorni della pinna). Il punto di partenza dell'articolo (riportato in Sezione 2) è un lavoro precedente [11] che suggerisce come le frequenze dei principali minimi spettrali nelle HRTF siano strettamente legate alla forma della pinna, in particolare alle sopracitate distanze. La Sezione 3 descrive la metodologia di estrazione di

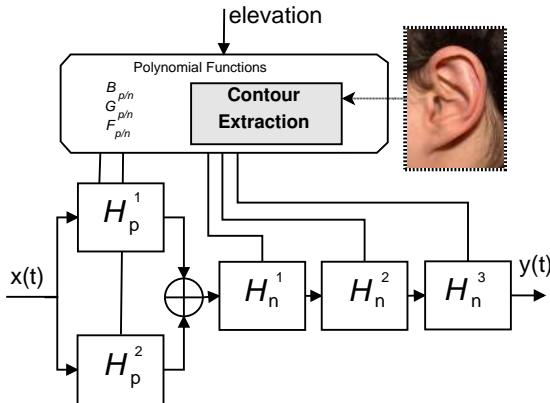


Figura 1. Rappresentazione schematica del modello strutturale della pinna.

feature e il modello di regressione non lineare utilizzato, mentre i risultati sono riportati e discussi nella Sezione 4.

2. UN MODELLO STRUTTURALE DI PINNA

Le differenze più marcate tra le HRTF di due diversi soggetti sono principalmente imputabili alle diverse caratteristiche della pinna (forma, grandezza e orientamento), talmente varie da rendere la pinna oggetto di recenti studi di identificazione biometrica [12]. La pinna ha un ruolo fondamentale nel determinare la forma spettrale delle HRTF grazie a due fenomeni acustici principali: riflessioni (sui bordi principali) e risonanze (nelle cavità). Di conseguenza, l'HRTF presenta nella sua ampiezza una sequenza di amplificazioni (picchi) in corrispondenza delle frequenze di risonanza e di nette attenuazioni (*notch*) in corrispondenza delle frequenze in cui si ha massima interferenza distruttiva tra onde dirette e onde riflesse. È proprio la collocazione spettrale di tali picchi e notch a costituire un indicatore fondamentale per la caratterizzazione della posizione spaziale della sorgente sonora, e in particolare della sua elevazione [10].

Nella letteratura dedicata troviamo diverse proposte modellistiche per rendere sinteticamente la componente della HRTF relativa alla pinna, conosciuta come *Pinna-Related Transfer Function* (PRTF). Tuttavia, tali modelli soffrono di evidenti limiti: la presenza della sola componente riflettente [13], la validità in regioni spaziali eccessivamente ristrette [14], e/o l'assenza di una parametrizzazione esplicita sull'antropometria dell'ascoltatore [15]. In un lavoro precedente [11] gli autori hanno proposto un modello strutturale di pinna costituito da due blocchi di filtri, il blocco *risonante* e il blocco *riflettente*. Il blocco risonante è costituito da due filtri *peak* del secondo ordine posti in parallelo; le uscite di tale blocco vengono sommate per essere inviate all'ingresso del blocco riflettente, costituito dalla cascata di 3 filtri *notch* del secondo ordine. Il modello è riportato schematicamente in Figura 1.

Nell'ambito dello stesso lavoro, gli autori hanno studiato la relazione tra le frequenze centrali dei notch presenti nelle PRTF e la geometria della pinna. A tal fine è

stata utilizzata una procedura di *ray-tracing* su immagini 2D della pinna¹ per mappare i punti di riflessione a una certa distanza dal punto di riferimento del canale uditivo, univocamente determinata dalla frequenza centrale di ogni notch. Gli autori hanno verificato che l'utilizzo di coefficienti di riflessione negativi risulta un fattore chiave nella determinazione della frequenza centrale stessa. La relazione tra frequenza del notch e distanza del punto di riflessione dal canale uditivo è quindi completamente riassunta nella semplice equazione

$$D_i(\phi) = \frac{c}{2F_i(\phi)}, \quad (1)$$

dove la costante c rappresenta la velocità del suono, ϕ è l'elevazione di cui si considera la PRTF, F_i è la frequenza centrale dell' i -esimo notch N_i , e D_i è la distanza tra il corrispondente punto di riflessione e l'ingresso del canale uditivo. Dai punti così trovati e mappati su immagini di pinne di un pool di soggetti sperimentali, gli autori hanno notato l'ottima corrispondenza tra i punti di riflessione e i tre contorni principali della pinna, ovvero bordo dell'elice, antielice/parete interna della conca, e bordo esterno della conca.

Partendo da tali risultati, gli autori in [16] hanno realizzato la procedura inversa: a partire dall'immagine della pinna sono stati tracciati i tre sopraccitati contorni e, attraverso semplici calcoli trigonometrici, trasformati in coppie di coordinate polari $(D_i(\phi), \phi)$ rispetto al canale uditivo. Dall'Equazione 1 si ricavano quindi le frequenze centrali dei notch ad ogni elevazione ϕ desiderata e per ognuno dei tre contorni. L'unico parametro indipendente usato dal modello è infatti l'elevazione della sorgente sonora virtuale, alla quale sono associate tre funzioni polinomiali che interpolano le frequenze centrali ricavate dai tre contorni (vedi Figura 1) per un certo passo di campionamento $\Delta\phi$. Per ciò che riguarda i parametri di larghezza di banda e profondità dei notch stessi, non è stata trovata una relazione con la forma della pinna. Gli autori hanno precedentemente utilizzato per questi due parametri valori medi di una popolazione di soggetti sperimentali [16], così come per i parametri caratteristici delle risonanze.

3. METODOLOGIA

L'insieme di dati di partenza è costituito dalle Head-Related Impulse Response (HRIR) incluse nel database di HRTF CIPIC [17], un database di dominio pubblico di HRIR misurate ad alta risoluzione spaziale in 1250 diverse direzioni per 45 diversi soggetti. Poiché in questo lavoro consideriamo l'antropometria dei soggetti sia sotto forma di dati numerici (i parametri antropometrici inclusi nel database CIPIC) che di una fotografia della pinna sinistra o destra, abbiamo selezionato i 33 soggetti per i quali questi dati sono disponibili per intero.

Prendiamo come riferimento il sistema di coordinate utilizzato nel database CIPIC, ovvero il sistema interaurale

¹ Il ray-tracing viene approssimativamente effettuato sulla proiezione dei contorni della pinna su di un piano parallelo al piano mediano del soggetto sperimentale.

Tabella 1. Coefficienti di correlazione di Pearson tra i parametri antropometrici della pinna.

parametro	descrizione	d_2	d_3	d_4	d_5	d_6	d_7	d_8	θ_1	θ_2
d_1	altezza conca inferiore	-0.06	0.10	0.19	0.51	0.21	0.23	0.24	-0.04	0.20
d_2	altezza conca superiore		-0.02	0.33	0.47	0.13	0.11	0.30	0.02	-0.11
d_3	larghezza conca inferiore			0.03	0.19	0.47	0.59	0.27	0.23	0.02
d_4	altezza fossa triangolare				0.67	0.53	0.03	0.30	-0.06	-0.28
d_5	altezza pinna					0.52	0.22	0.44	-0.11	0.00
d_6	larghezza pinna						0.16	0.45	0.09	-0.24
d_7	larghezza incisura intertragica							0.15	-0.09	0.03
d_8	profondità conca inferiore								0.01	0.14
θ_1	angolo di rotazione									
θ_2	angolo di apertura									-0.12

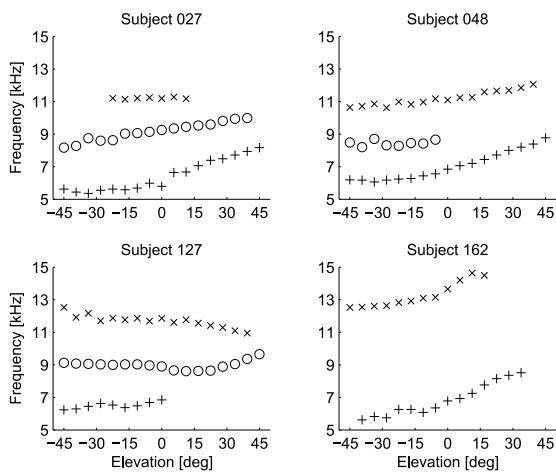


Figura 2. Notch estratti e raggruppati per quattro soggetti CIPIC (N1:+, N2:o, N3:x).

polare, restringendo l'analisi alla porzione frontale del piano mediano (azimut $\theta = 0^\circ$), con l'elevazione ϕ che varia tra $\phi = -45^\circ$ e $\phi = 45^\circ$ a passi di 5,625 gradi (17 HRIR per soggetto). La ragione per considerare il solo piano mediano è dovuta alla constatazione che variazioni relative di azimut fino a $\Delta\theta = 30^\circ$ per una data elevazione producono modeste variazioni spettrali nella PRTF [18], permettendo di generalizzare il modello a un intervallo di valori più ampio per l'azimut. Le elevazioni maggiori di 45° sono state invece scartate a causa della tipica mancanza di notch nelle HRTF corrispondenti [19].

3.1 Estrazione delle frequenze dei notch

Al fine di ottenere le frequenze dei notch più marcati, abbiamo applicato l'algoritmo di elaborazione del segnale di Raykar *et al.* [18] ad ogni HRIR disponibile. In sintesi, l'algoritmo prevede il calcolo della funzione di autocorrelazione del residuo della predizione lineare della HRIR ed estrae le frequenze dei notch come i minimi locali della funzione di *group delay* risultanti al di sotto di una determinata soglia (fissata euristicaamente a -0.5 campioni). Successivamente, i notch appartenenti alla riflessione sullo stesso contorno sono stati raggruppati lungo le diverse

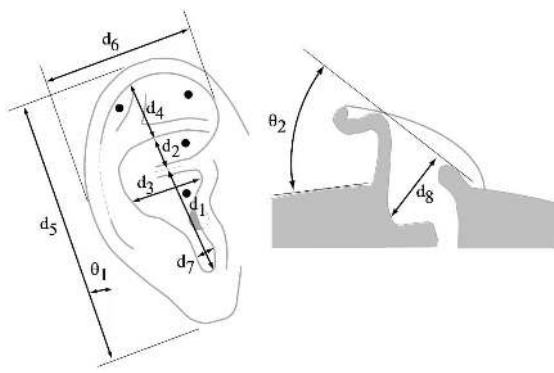


Figura 3. I 10 parametri antropometrici della pinna inclusi nel database CIPIC (figura riprodotta da [17]).

elevazioni ϕ attraverso l'algoritmo di *partial tracking* di McAulay e Quatieri [20]. L'intervallo di *matching* per l'algoritmo è stato fissato a $\Delta = 1$ kHz. La stessa procedura di estrazione e raggruppamento di notch è stata utilizzata in un lavoro precedente [21].

Abbiamo quindi considerato solo i raggruppamenti (*track*) con almeno 3 notch. Nei casi in cui un soggetto presenta più di 3 track che soddisfino tale requisito, abbiamo considerato i 3 track più lunghi ed etichettato ogni frequenza del notch con F_1 , F_2 e F_3 in ordine crescente di frequenza media. Nei casi in cui un soggetto presenta meno di 3 track (9 casi su 33), le etichette sono state assegnate con un criterio di vicinanza alla frequenza media del track calcolata su tutti i soggetti con 3 track. In Figura 2 sono riportati i track di quattro soggetti rappresentativi. Complessivamente, la procedura applicata ai 33 soggetti ha prodotto 367 diversi campioni per F_1 , 401 per F_2 e 303 per F_3 .

3.2 Estrazione di feature antropometriche

Trentasette misure antropometriche unidimensionali per ognuno dei 33 soggetti sono disponibili nel database CIPIC: 17 per la testa e il torso e 10 per ognuna delle due pinne. Poiché il *focus* di questo lavoro è sui notch spettrali, consideriamo soltanto i parametri della pinna in esame, riportati nella Figura 3 e descritti nella Tabella 1. La tabella ri-

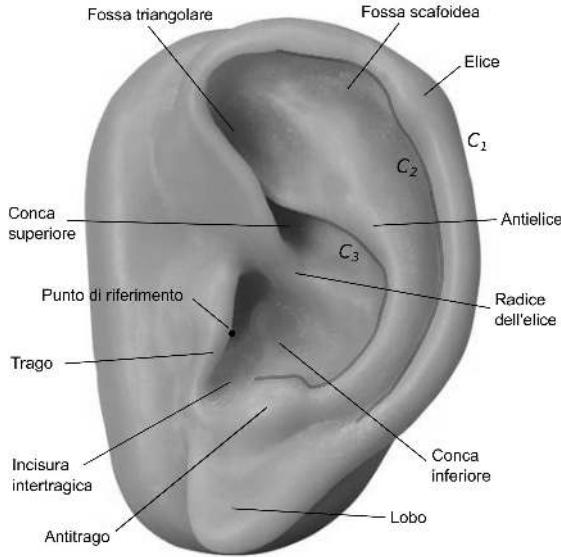


Figura 4. Anatomia della pinna e i tre contorni estratti: C_1 , C_2 e C_3 .

porta inoltre i valori di correlazione tra parametri calcolati sui soggetti CIPIC. Generalmente tali valori sono modesti, tranne quelli per le misure sovrapposte (ad esempio $d_4 - d_5$, $d_3 - d_7$), e ciò denota un grado di ortogonalità accettabile tra i parametri antropometrici.

Inoltre, abbiamo estratto parametri dipendenti dall'elevazione strettamente correlati all'idea di *ray tracing* sviluppata nella precedente Sezione 2. I tre contorni corrispondenti al bordo esterno dell'elice, al bordo interno dell'elice, e al bordo della concava inferiore/antitrago (vedi Figura 4) sono stati dapprima tracciati a mano con l'aiuto di una tavoletta grafica e memorizzati come sequenze di pixel.² Successivamente, il punto di massima protuberanza del trago è stato scelto come punto di riferimento del canale uditivo (vedi ancora Figura 4) per il calcolo dei parametri di distanza. Per ogni angolo di elevazione $\phi \in [-45, 45]$, le distanze in pixel tra il punto di riferimento e il punto che interseca ogni contorno della pinna lungo il raggio che ha origine nel punto di riferimento con pendenza $-\phi$ sono infine convertite in centimetri facendo affidamento sul metro riportato nelle fotografie accanto alla pinna e memorizzate come $r_i(\phi)$, dove $i \in \{1, 2, 3\}$ si riferisce al contorno C_i associato.

A causa della forma approssimativamente ellittica della pinna, i tre parametri r_1 , r_2 e r_3 sono strettamente correlati. Precisamente, il coefficiente di correlazione tra r_1 e r_2 è 0.95, ed entrambi correlano con un valore di 0.75 con r_3 .

² Precisiamo che questi tre contorni non corrispondono univocamente ai tre contorni ipotizzati responsabili delle riflessioni nella precedente Sezione 2. La ragione di tale scelta è pratica: i contorni considerati possono essere estratti automaticamente in maniera robusta attraverso tecniche di elaborazione dell'immagine basate su rilevamento di profondità con dispositivi multi-flash [22]. Tuttavia, in questo lavoro l'estrazione manuale dei contorni si è resa necessaria a causa della disponibilità di singole immagini della pinna e della limitata efficacia di tecniche basate sull'intensità (e.g. Canny, Sobel) su immagini a basso contrasto.

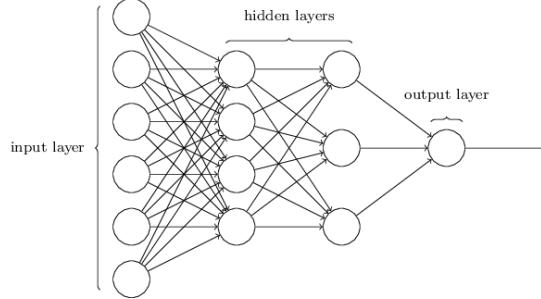


Figura 5. Rappresentazione di un percettrone multistrato.

L'elevata correlazione tra r_1 e r_2 può anche essere spiegata dal fatto che, semplicemente, entrambi i parametri si riferiscono all'elice.

3.3 Modello di regressione

Si vuole ora studiare la relazione tra le frequenze F_1 , F_2 e F_3 e i tredici parametri misurati (d_i , $i = 1 \dots 8$, θ_j , $j = 1 \dots 2$, $r_k(\phi)$, $k = 1 \dots 3$, $\phi \in [-45, 45]$) sui dati sopra descritti. Il problema può essere facilmente configurato come un problema di regressione, dove gli attributi sono i parametri antropometrici e le variabili target sono le frequenze dei notch. In un lavoro precedente [23] gli autori hanno verificato sullo stesso insieme di dati tramite tecniche di regressione lineare che i 13 parametri antropometrici considerati non sono sufficienti per stimare con un errore accettabile i notch N_2 e N_3 . Test preliminari condotti utilizzando il modello non lineare descritto di seguito al posto della regressione lineare hanno evidenziato la medesima difficoltà di apprendimento. Di conseguenza, l'analisi seguente si concentra sulla stima di N_1 .

Il modello non lineare utilizzato per la regressione è il percettrone multistrato (vedi Figura 5), un tipo di rete neurale artificiale elementare ma molto utilizzato la cui struttura è caratterizzata da tre o più strati di neuroni artificiali (nodi) in cascata. Di questi strati il primo è di input, l'ultimo è di output, mentre quelli tra essi compresi sono detti "strati nascosti" (*hidden layers*). Ciascun nodo è connesso con un certo peso sinaptico w_{ij} a ogni nodo dello strato successivo, con il verso della connessione sempre diretto verso l'uscita. Le componenti in ingresso (nel nostro caso i vettori di attributi) vengono trasmesse ai nodi di input, che calcolano la propria funzione di attivazione (non lineare) e ne trasmettono il valore allo strato successivo; la propagazione prosegue fino ai nodi di output (nel nostro caso singolo e relativo a F_1) [24, 25].

Poiché tutti i neuroni di una rete hanno la stessa funzione di attivazione, gli elementi parametrici variabili sono i pesi sinaptici. La fase di apprendimento supervisionato del percettrone, eseguita fornendo un *training set*, è un problema di ricerca del minimo di una funzione di errore tra le stime fornite in uscita a partire dagli ingressi e i valori reali delle variabili in uscita, e si affronta con la tecnica della *discesa del gradiente*. Una volta calcolata l'uscita della rete per una determinata istanza in ingresso, si procede al calcolo della variazione del peso di ciascuna sinapsi, proce-

dendo a ritroso, a partire dal nodo di uscita (*error backpropagation*) [25]. Tale operazione viene svolta ciclicamente, con ogni ciclo chiamato *epoca di apprendimento*.

Mentre il numero di nodi di input e output è fissato, è possibile intervenire sul numero e sulla dimensione degli strati interni: in generale, maggiore è la complessità della rete nascosta e più la rete neurale è efficace nella stima di funzioni articolate e di caratteristiche particolarmente astratte presenti nel dataset (*deep learning*). Oltre un certo livello di complessità, tuttavia, la rete tende all'*overfitting*. Nel nostro caso, avendo un insieme di dati per l'addestramento limitato e non dovendo stimare caratteristiche di alto livello, è stato impiegato un solo strato nascosto. Il numero di neuroni nascosti è stato invece stabilito in base al numero di nodi di input, come vedremo nella sezione successiva.

La motivazione dell'utilizzo del percepitrone multistrato è duplice. In primo luogo, l'acquisizione di HRTF e dei parametri antropometrici è soggetta a errore di misurazione. L'impiego di una rete neurale, se correttamente dimensionata, permette di ridurre l'effetto di tale rumore. In secondo luogo, la scelta del percepitrone multistrato permette di costruire un modello dalla struttura arbitraria e articolata, in grado di stimare eventuali nonlinearità.

4. REGRESSIONE ANTROPOMETRICA

Vengono ora descritti tre diversi modelli di regressione non lineare che si differenziano per gli attributi considerati in ingresso e quindi anche per il numero di nodi negli strati di input e nascosto. Il primo modello M_1 considera i dieci parametri CIPIC, il secondo modello M_2 i parametri di distanza estratti, e il terzo modello M_3 la combinazione dei due insiemi.

Per l'addestramento e la validazione dei modelli vengono usati rispettivamente un *training set* e un *test set* ricavati partizionando le 367 istanze di partenza di N_1 per soggetti. Infatti, poiché il modello deve stimare caratteristiche di un soggetto nuovo, è fondamentale per una corretta validazione che i dati vengano partizionati senza separare le istanze relative allo stesso soggetto. Sono quindi state scelte 20 diverse combinazioni di 7 soggetti in modo tale che ogni test set fosse di circa il 25% delle istanze totali, lasciando il restante 75% nel training set. Il modello viene quindi sottoposto a *cross-validation*, considerando la media delle statistiche risultanti da ogni test set.

La sperimentazione è stata condotta all'interno dell'ambiente WEKA, una piattaforma open source per l'apprendimento automatico caratterizzata da una vasta collezione di strumenti statistici ed algoritmi per l'analisi dei dati [24]. Tali strumenti sono stati utilizzati mediante API Java al fine di automatizzare e raffinare l'analisi.

4.1 Primo modello

I parametri del modello M_1 sono indipendenti dall'elevazione della sorgente sonora, che è pertanto necessario includere come attributo indipendente. I vettori di attributi includono quindi un indice intero di elevazione $n \in \{1, \dots, 17\}$ e i 10 parametri CIPIC $d_i, i = 1 \dots 8, \theta_j, j =$

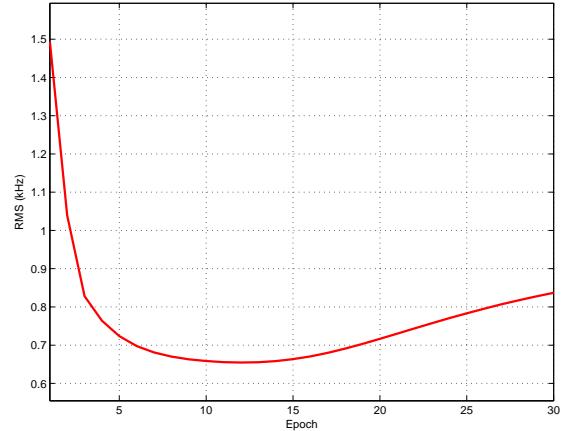


Figura 6. Progresso dell'addestramento di M_1 sul test set che include i soggetti CIPIC 003, 010, 050, 058, 127, 137 e 152.

1 ... 2. Il percepitrone è dunque vincolato ad avere 11 nodi di ingresso e uno di uscita (F_1). È inoltre stato scelto euristicamente uno strato nascosto con 2 soli nodi.

In Figura 6 viene riportato il processo di apprendimento in termini di errore RMS rispetto al numero di epoch di addestramento, con un test set rappresentativo. L'errore diminuisce progressivamente fino a raggiungere un livello ottimale di apprendimento oltre il quale si verifica overfitting. A questo punto il modello perde progressivamente capacità di generalizzazione e quindi l'errore di predizione sul test set torna a crescere.

Le prestazioni medie di M_1 sono: RMSE minimo di 604 Hz e correlazione $r = 0.88$. I valori minimi dell'RMSE sono tuttavia molto diversi tra loro: si va da 425 Hz a 1 kHz con una deviazione standard di 148 Hz nelle 20 istanze considerate. Tutte le istanze considerate raggiungono il minimo valore dell'RMSE entro 40 epoch di apprendimento.

4.2 Secondo modello

I parametri del modello M_2 , le distanze $r_k(\phi), k = 1 \dots 3, \phi \in [-45, 45]$, sono al contrario dipendenti dall'elevazione della sorgente sonora, che non è stata pertanto considerata. In questo caso i nodi di input del percepitrone sono necessariamente 3, uno per ogni distanza. La rete ottimale risulta ancora composta da uno strato di 2 nodi interni nascosti. Poiché in generale la proporzionalità tra le frequenze dei notch e i parametri antropometrici estratti è inversa, prima di applicare l'operazione di regressione è stato considerato il reciproco di questi ultimi. La Figura 7 mostra a titolo esemplificativo il fit polinomiale di quarto grado delle distanze $r_1(\phi)$ dipendentemente dalla frequenza di N_1 .

I valori medi di RMSE e correlazione sono rispettivamente 670 Hz e $r = 0.84$. La differenza tra gli errori minimi è stavolta più contenuta, e spazia tra i 500 e gli 800 Hz con una deviazione standard di 92 Hz. Tuttavia il percepitrone, a causa del numero molto ridotto di nodi, ri-

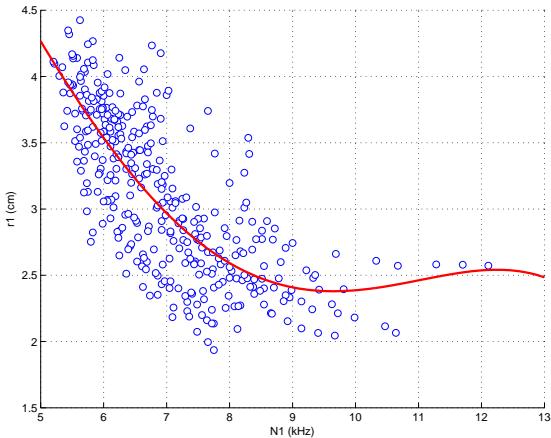


Figura 7. Grafico di dispersione tra F_1 e $r_1(\phi)$. La curva rappresenta il fit polinomiale di quarto grado dei dati.

chiede fino a 400 epochhe di apprendimento per assestarsi sulle prestazioni migliori.

4.3 Terzo modello

Avendo in ingresso tutti i parametri a disposizione meno l'elevazione, il modello M_3 ha 13 nodi di input. La struttura interna scelta è in questo caso composta da 3 nodi nascosti, uno in più rispetto ai precedenti modelli.

Le prestazioni complessive di questo modello risultano (prevedibilmente) migliori rispetto a quelle dei precedenti, in particolare in termini di RMSE, che scende a 554 Hz con una deviazione standard di 84 Hz, mentre la correlazione è $r = 0.88$. I tempi di apprendimento ottimali non superano mai le 70 epochhe. Il valore RMSE risulta inoltre leggermente inferiore al valore ottenuto con la regressione lineare in [23] (590 Hz). Questo risultato suggerisce che l'introduzione di una componente non lineare abbia aumentato la precisione della stima, senza tuttavia migliorarne drasticamente i risultati. Tuttavia, occorre specificare che in questo lavoro è stata utilizzata una differente procedura di validazione dei modelli.

Il modello M_3 introduce errori significativi se utilizzato per predire F_1 , ed è risaputo che lo spostamento in frequenza degli indicatori di elevazione ha un impatto notevole sull'accuratezza della localizzazione. Ad esempio, spostamenti di 1 kHz di N_1 possono corrispondere ad un aumento o una diminuzione dell'angolo di elevazione di 20° od oltre [26]. Tuttavia, la letteratura [27] suggerisce che due notch ad alta frequenza (attorno a 8 kHz) che differiscono soltanto nel valore della frequenza centrale possono essere percepiti come distinti solo se tale differenza è almeno pari al 10% circa del valore di frequenza inferiore.

Se teniamo inoltre in considerazione il fatto che la procedura utilizzata contiene inevitabilmente errori, ad esempio l'estrazione completamente automatica delle frequenze dei notch, o il piazzamento arbitrario del punto di riferimento del canale uditivo necessario per calcolare le distanze r_k , possiamo concludere che la stima di F_1 a partire dai parametri antropometrici considerati è possibile.

5. CONCLUSIONI

In questo articolo abbiamo studiato l'applicazione di un modello di regressione non lineare per stimare la frequenza dei notch nelle HRTF relative alla porzione frontale del piano mediano a partire da dati antropometrici. I risultati per il notch N_1 mostrano una corrispondenza incoraggianente tra l'antropometria e le feature spettrali, e ciò conferma la possibilità di predire la frequenza centrale a partire da una semplice fotografia dell'orecchio.

Tuttavia, occorre precisare che la quantità limitata dei dati sperimentali a disposizione può avere influito sui risultati. In tale situazione, per evitare l'overfitting è necessario ridurre il numero di epochhe di apprendimento. Ne consegue che il modello risultante resta in una certa misura dipendente dai pesi iniziali della rete, stabiliti in modo pseudocasuale. La qualità dei dati sperimentali può inoltre avere avuto un impatto altrettanto importante. I modelli sono stati addestrati e validati sul database CIPIC, che sebbene sia il database di HRTF più utilizzato dalla comunità scientifica soffre di errori di misura (ad esempio evidenti asimmetrie sul piano orizzontale [28]) e mancanza di documentazione (ad esempio non è specificato con esattezza il punto di misurazione delle HRTF).

Per questi motivi, i risultati ottenuti vanno interpretati come un'indicazione sulle potenzialità dell'approccio sperimentale, in prospettiva tanto dell'utilizzo di un database più recente e documentato [19, 29] quanto di avanzamenti nelle tecniche di estrazione dei contorni della pinna.

6. RINGRAZIAMENTI

Questo lavoro è stato finanziato dal progetto di ricerca PADVA (Personal Auditory Displays for Virtual Acoustics), n. CPDA135702 dell'Università di Padova e dal programma di ricerca e innovazione dell'Unione Europea Horizon 2020 in virtù del contratto di sovvenzione n.643636.³

7. BIBLIOGRAFIA

- [1] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, pp. 231–249, April 2001.
- [2] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.*, vol. 49, pp. 904–916, October 2001.
- [3] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*. Plantation, FL, USA: J.Ross Publishing, 2nd ed., June 2013.
- [4] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hamershøi, "Binaural technique: Do we need individual

³ This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636. <http://www.soundofvision.net/>

- recordings?,” *J. Audio Eng. Soc.*, vol. 44, pp. 451–469, June 1996.
- [5] C. P. Brown and R. O. Duda, “A structural model for binaural sound synthesis,” *IEEE Trans. Speech Audio Process.*, vol. 6, pp. 476–488, September 1998.
- [6] L. Li and Q. Huang, “HRTF personalization modeling based on RBF neural network,” in *Proc. 38th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2013)*, (Vancouver, BC, Canada), pp. 3707–3710, May 2013.
- [7] Q. Huang and L. Li, “Modeling individual HRTF tensor using high-order partial least squares,” *EURASIP J. Adv. Signal Process.*, vol. 2014, pp. 1–14, May 2014.
- [8] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, and J. C. Platt, “HRTF magnitude synthesis via sparse representation of anthropometric features,” in *Proc. 39th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2014)*, (Firenze, Italy), pp. 4501–4505, May 2014.
- [9] F. Grijalva, L. Martini, S. Goldenstein, and D. Florencio, “Anthropometric-based customization of head-related transfer functions using Isomap in the horizontal plane,” in *Proc. 39th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2014)*, (Firenze, Italy), pp. 4506–4510, May 2014.
- [10] S. K. Roffler and R. A. Butler, “Factors that influence the localization of sound in the vertical plane,” *J. Acoust. Soc. Am.*, vol. 43, pp. 1255–1259, June 1968.
- [11] S. Spagnol, M. Geronazzo, and F. Avanzini, “On the relation between pinna reflection patterns and head-related transfer function features,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, pp. 508–519, March 2013.
- [12] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, “A survey on ear biometrics,” *ACM Trans. Embedded Computing Systems*, vol. 9, pp. 39:1–39:33, March 2010.
- [13] A. J. Watkins, “Psychoacoustical aspects of synthesized vertical locale cues,” *J. Acoust. Soc. Am.*, vol. 63, pp. 1152–1165, April 1978.
- [14] P. Satarzadeh, R. V. Algazi, and R. O. Duda, “Physical and filter pinna models based on anthropometry,” in *Proc. 122nd Conv. Audio Eng. Soc.*, (Vienna, Austria), pp. 718–737, May 2007.
- [15] K. J. Faller II, A. Barreto, and M. Adjouadi, “Augmented Hankel total least-squares decomposition of head-related transfer functions,” *J. Audio Eng. Soc.*, vol. 58, pp. 3–21, January/February 2010.
- [16] 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, December 2011.
- [17] 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, October 2001.
- [18] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, “Extracting the frequencies of the pinna spectral notches in measured head related impulse responses,” *J. Acoust. Soc. Am.*, vol. 118, pp. 364–374, July 2005.
- [19] 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), pp. 209–212, September 2011.
- [20] R. J. McAulay and T. F. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, pp. 744–754, August 1986.
- [21] S. Spagnol, “On distance dependence of pinna spectral patterns in head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 137, pp. EL58–EL64, January 2015.
- [22] S. Spagnol, D. Rocchesso, M. Geronazzo, and F. Avanzini, “Automatic extraction of pinna edges for binaural audio customization,” in *Proc. IEEE Int. Work. Multi. Signal Process. (MMSP 2013)*, (Pula, Italy), pp. 301–306, October 2013.
- [23] S. Spagnol and F. Avanzini, “Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model,” in *Proc. 18th Int. Conf. Digital Audio Effects (DAFx-15)*, (Trondheim, Norway), pp. 231–236, December 2015.
- [24] R. Rojas, *Neural Networks: A Systematic Introduction*. Berlin: Springer-Verlag, 1996.
- [25] B. Widrow and M. A. Lehr, “30 years of adaptive neural networks: Perceptron, Madaline, and backpropagation,” *Proc. IEEE*, vol. 78, pp. 1415–1442, September 1990.
- [26] 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.
- [27] 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.
- [28] S. Spagnol and F. Avanzini, “Anthropometric tuning of a spherical head model for binaural virtual acoustics based on interaural level differences,” in *Proc. 21st Int. Conf. Auditory Display (ICAD 2015)*, (Graz, Austria), pp. 204–209, July 2015.
- [29] C. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, “Creating the Sydney York morphological and acoustic recordings of ears database,” *IEEE Trans. Multimedia*, vol. 16, pp. 37–46, January 2014.

GOOD OR BAD? POTENZIARE L'ASCOLTO ATTIVO MEDIANTE SPAZI SONORI INTERATTIVI

Federico Altieri

Dip. Ing. dell'Informazione
Università di Padova
altieri@dei.unipd.it

Marcella Mandanici

Dip. Ing. dell'Informazione
Università di Padova
mandanici@dei.unipd.it

Sergio Canazza

Dip. Ing. dell'Informazione
Università di Padova
canazza@dei.unipd.it

Antonio Rodà

Dip. Ing. dell'Informazione
Università di Padova
roda@dei.unipd.it

Emanuele Menegatti

Dip. Ing. dell'Informazione
Università di Padova
menegatti@dei.unipd.it

Matteo Munaro

Dip. Ing. dell'Informazione
Università di Padova
munaro@dei.unipd.it

ABSTRACT

In questo lavoro viene presentata l'applicazione *Good or Bad?* sviluppata per stimolare e migliorare la capacità di ascolto attivo e proposta agli utenti sotto forma di gioco musicale basato sulla tecnologia degli spazi interattivi. L'applicazione è stata provata da oltre 100 utenti in occasione dell'edizione 2015 della 'Notte Europea dei Ricerca'. I dati raccolti mostrano un alto interesse del pubblico e offrono diversi spunti di approfondimento sulle performance educative dell'applicazione.

1. INTRODUZIONE

Con la seconda e la terza rivoluzione industriale, la diffusione capillare di dispositivi di riproduzione audio a basso costo ha moltiplicato a dismisura le occasioni di fruizione musicale, permettendoci e, in diversi casi, quasi "obbligandoci" ad ascoltarne quotidianamente. Questo ha reso necessaria la distinzione dei diversi tipi d'ascolto che il fruitore può sperimentare: forme di ascolto passive, il più delle volte distratte (come l'ascolto della musica di sottofondo di negozi e supermercati) e forme di ascolto attivo, volontarie, dove l'utente è pienamente immerso e concentrato sulla musica in ascolto e compie azioni in funzione di essa, modificando così anche l'oggetto dell'ascolto [1]. L'ascolto è prerogativa fondamentale per la comprensione della musica [2] ed è da considerarsi a tutti gli effetti una abilità, che può essere quindi coltivata e migliorata tramite l'esercizio. *Good or Bad?* si pone l'obiettivo di stimolare e puntare al miglioramento di questa abilità: per farlo in un modo coinvolgente ed efficace, sono state sfruttate le proprietà e le possibilità date dagli spazi interattivi.

1.1 L'ascolto attivo

L'ascolto attivo può essere interpretato in diversi modi. Volpe e Camurri [3] distinguono tra ascolto attivo centrato sull'utente, che dipende dal tipo di interazione e dal grado di eventuale collaborazione tra più utenti, e ascolto

Copyright: © 2016 Federico Altieri 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.

attivo basato sul contenuto, più creativo. *Sync-in-Team* [4] e *MusicSpace* [5] sono esempi rispettivamente per le due categorie appena presentate. In *Sync-in-Team* due o più giocatori devono prestare attenzione alla sincronia dei movimenti per riprodurre uno spezzone musicale: la musica qui gioca un ruolo di feedback per valutare in tempo reale anche l'andamento della performance. In *MusicSpace* l'utente può manipolare il mixaggio di tracce pre-registrate, spostando nello spazio i vari strumenti grazie ad una interfaccia grafica.

È un dato di fatto che, in entrambe queste categorie, alcune operazioni vengono comunque realizzate basandosi sul contenuto e sfruttando diversi tipi di strumentazione e tecniche. È tuttavia possibile distinguere tra applicazioni che sfruttano interfacce utente tradizionali e applicazioni che, come *Good or Bad?*, sfruttano e traggono i movimenti dell'utente.

1.2 Gli spazi interattivi

Negli ultimi anni, la possibilità di includere l'interazione corporea ha portato nuove possibilità per i progettisti di interfacce [6]. Spazi interattivi ed ambienti virtuali sono stati sempre più utilizzati come interfacce di applicazioni per diversi ambiti, come l'intrattenimento (menzioniamo a titolo di esempio [7] e [8]), l'apprendimento (come *Harmonic Walk* [9]) e la riabilitazione (Avanzini et al [10]). Il fattore di engagement, nello specifico, è osservabile, come descritto da Berthouze [11] e da De Kort [12] come la combinazione di due fattori: il coinvolgimento del corpo e l'interazione sociale.

L'interazione a corpo libero è stata approfondita da Berthouze [11]: in questo studio, Berthouze dimostra che medesime applicazioni risultano più divertenti, coinvolgenti ed efficaci nelle performance degli utenti in un contesto di movimento a corpo libero rispetto a forme di interfaccia più tradizionali (come controller o mouse/tastiera), senza dimenticare anche il maggiore benessere fisico dovuto all'attività motoria richiesta.

Per quanto riguarda l'interazione sociale, De Kort et al [12] hanno approfondito la questione, Mettendo in evidenza vantaggi e svantaggi degli spazi di interazione sociale, categoria nella quale rientrano anche i pavimenti interattivi, per via della loro natura "aperta" dove utenti e spettatori interagiscono e condividono un'esperienza

Questa “pubblicità” dell’interazione comporta dei vantaggi, come i sentimenti positivi derivanti da una maggiore socievolezza, ma anche dei potenziali punti critici: difatti la pubblicità delle esibizioni potrebbe comportare con sé una riduzione delle performance dovuta al timore della valutazione.

1.3 Organizzazione del lavoro

Nella prima sezione dell’articolo verrà presentata l’applicazione, con i suoi fondamenti teorici, la logica di gioco (*gameplay*) e la tecnologia impiegata nella realizzazione. Nella seconda si illustreranno i dati raccolti nel contesto dell’edizione 2015 della ‘Notte Europea dei Ricerca’. Nella terza ed ultima parte si menzioneranno le conclusioni, i limiti del setup attuale e si illustreranno i possibili sviluppi futuri.

2. IL GIOCO GOOD OR BAD?

2.1 Fondamenti teorici

Le composizioni musicali possono avere strutture e forme assai diverse in dipendenza delle varie epoche e dei vari stili. Tuttavia, eccezion fatta per generi tipicamente monodici, come il canto Gregoriano, o per composizioni scritte per strumento solo, generalmente un brano di musica può essere visto come la sovrapposizione di diverse parti concordanti fra loro. Queste parti hanno diverse funzioni e possono essere eseguite da strumenti anche timbricamente contrastanti fra loro. L’orchestra sinfonica, per esempio, è caratterizzata da varie sezioni come gli archi, gli ottoni, i legni, le percussioni, le tastiere, etc. a cui vengono affidate le varie parti come il tema principale, il tema o i temi secondari, il basso, varie figurazioni di sottofondo con funzione di riempimento armonico, etc. (Figura 1). Ognuna di queste funzioni costituisce uno “strato” timbrico e musicale che, sommato a tutti gli altri, contribuisce a formare l’insieme della composizione.



Figura 1: Partitura per orchestra d’archi con evidenziate le funzioni di tema principale (in verde), di accompagnamento armonico (in giallo) e di basso (in violetto).

Lo stesso tipo di struttura, anche se con nomi e timbri differenti, si trova anche in molti generi musicali popolari o in composizioni jazzistiche. Proprio su questa caratteristica musicale si basa *Good or Bad?*: l’utente è chiamato ad ascoltare con attenzione i vari frammenti sonori e ad esercitare la sua abilità di ascolto attivo basato sul contenuto per decidere se essi sono compatibili o meno col frammento iniziale.

2.2 Infrastruttura tecnologica e gameplay

Good or Bad? è un’applicazione che poggia sulla tecnologia AILearn [9], qui presentata in una versione più avanzata rispetto al passato. Questa tecnologia comprende un ambiente interattivo per il tracking delle persone in un’area 4x3 metri.

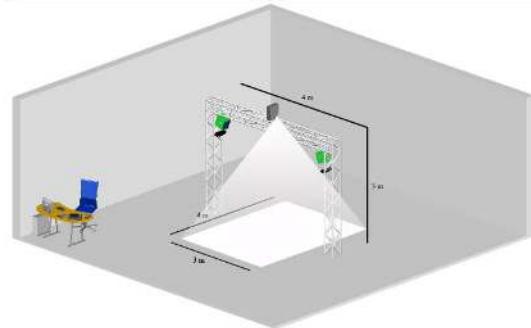


Figura 2: La struttura AILearn: un traliccio di alluminio, due Kinect v2, due altoparlanti e un proiettore.

Il sistema appare come mostrato in Figura 2: un traliccio di alluminio sorregge due sensori Kinect v2¹, due altoparlanti e un proiettore puntato verso il pavimento.

Il sistema di tracking è organizzato e gestito dalla libreria OpenPTrack², sviluppata dal Laboratorio di Intelligent Autonomous Systems (IAS-Lab) del dipartimento di Ingegneria dell’Informazione dell’Università degli Studi di Padova (Munaro et al [13] [14] [15]). OpenPTrack organizza secondo un modello master-slave il lavoro di più elaboratori connessi tra loro in una rete privata: un elaboratore master coordina ed elabora i dati raccolti da altri elaboratori slave, ognuno dei quali collegato ad un dispositivo Kinect v2. Il sistema è in grado di tracciare gli spostamenti di più persone all’interno di un’area, producendo per ognuna di esse una coppia di coordinate su uno spazio cartesiano. Queste coordinate sono utilizzate come input per i nostri applicativi, in modo del tutto simile a dei cursori. Il codice, inclusivo di logica di gioco e controllo delle animazioni, è stato realizzato sfruttando le librerie Processing³: un framework Java per la realizzazione di animazioni 2D/3D, che comunica con una patch realizzata in Max/MSP⁴, che cura la parte audio. La comunicazione tra i due avviene attraverso il protocollo OSC (Open Sound Control [16]).

¹http://www.microsoftstore.com/store/msusa/en_US/pdp/Kinect-for-Windows-Developer-Bundle/productID.314513600

²<http://openptrack.org/>

³<https://processing.org/>

⁴<https://cycling74.com/>

Una partita a GoB può essere giocata in uno o più giocatori e si struttura in queste fasi: per iniziare, ad un utente viene chiesto di camminare su un numero da 1 a 3 per la scelta del repertorio con cui giocare. Successivamente gli viene chiesto di camminare all'interno di uno dei due cerchi comparsi sulla zona sinistra del piano di gioco.

La traccia che si attiva al passaggio nel primo cerchio è la traccia di riferimento: l'obiettivo sarà scoprire le altre tre tracce della stessa canzone senza commettere 3 errori.



Figura 3: una ragazza che gioca una partita a *Good or Bad?*

Passati 10 secondi, la traccia di riferimento si abbassa di volume, in attesa della mossa dell'utente che, camminando su un altro cerchio bianco, scoprirà una seconda traccia: a questo punto si sblocca anche la parte destra dello schermo, che sottintende la domanda che sta alla base del gioco, “l'insieme delle tracce in riproduzione sono in concordanza musicale? Fanno parte della stessa canzone?”. Gli utenti e il pubblico presente sono coinvolti e possono partecipare attivamente alla formulazione della risposta, che viene data camminando su *Good* o su *Bad*. L'errore nel rispondere comporta la perdita di una vita. A questo punto, se la risposta giusta era *Good*, la traccia appena scoperta viene “sovrapposta” al riferimento, altrimenti viene scartata. Tutto questo è riassunto nel diagramma in figura 4.

Come si può evincere dalle figure 3 e 5, l'interfaccia di gioco è stata pensata per una collaborazione tra più giocatori, un giocatore scopre le tracce, mentre il secondo risponde alla domanda posta dal gioco.

3. VALUTAZIONE SPERIMENTALE

3.1 Partecipanti

L'applicazione è stata testata durante l'edizione 2015 della ‘Notte Europea dei Ricercatori’ tenutasi al palazzo del Bo a Padova. Il gioco è stato avviato alle 17:00 e fermato alle 24:00, alla chiusura della manifestazione. I partecipanti al test erano passanti e curiosi di tutte le fasce d'età e ripartiti equamente tra maschi e femmine che si

sono presentati nel nostro settore attratti dallo spazio interattivo, senza alcun tipo di selezione o classificazione. In totale abbiamo avuto 150 persone coinvolte per 120 partite giocate. La valutazione è stata effettuata nell'ambito di una manifestazione pubblica, senza seguire, quindi, un rigido protocollo sperimentale

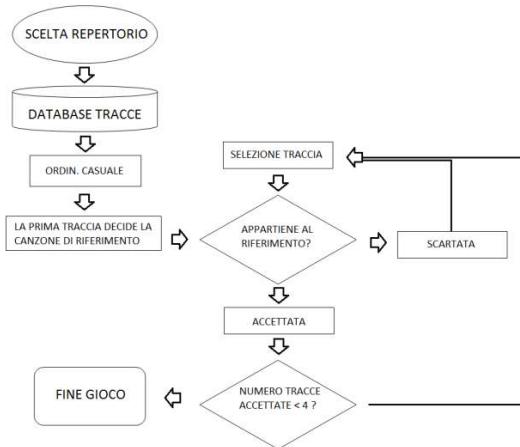


Figura 4: Il flusso di gioco di *Good or Bad?*

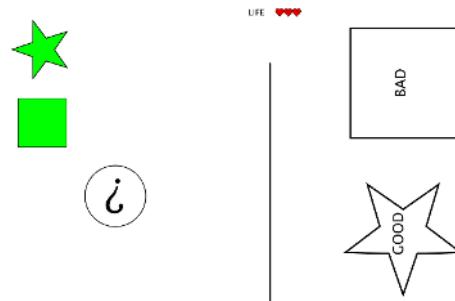


Figura 5: uno screenshot dell'interfaccia di gioco proiettata. Nella metà di sinistra è possibile sbloccare le tracce, mentre a destra esprimere la risposta sui confronti.

3.2 Materiali

La scelta dei pezzi musicali è stata curata in modo tale da bilanciare la difficoltà del gioco, puntando a proporre pezzi famosi, che gran parte del pubblico avrebbe dovuto essere in grado di riconoscere. Per la ‘Notte Europea dei Ricercatori’ sono stati scelti: due canzoni del gruppo svedese Abba, *Dancing Queen* (1976) e *Mamma Mia* (1975); due canzoni di Lucio Battisti, *Acqua Azzurra*, *Acqua Chiara* (1970) e *Fiori rosa, fiori di Pesco* (1970); due pezzi classici presi da *Le Quattro Stagioni* di Antonio Vivaldi (1723): *Concerto No. 1 in Mi maggiore, Op.*

8, RV 269, *La Primavera Allegro in Mi maggiore e Concerto No. 4 in Fa minore, Op. 8, RV 297, L'Inverno Largo in Mi-bemolle maggiore*⁵.

L'organizzazione delle tracce all'interno di ciascun repertorio è spiegata in Tabella 1.

	Battisti	ABBA	Vivaldi
1	Acqua azzurra Acqua chiara - percussionsi	Dancing Queen - percussionsi	L'Inverno - basso
2	Acqua azzurra Acqua chiara - basso	Dancing Queen - basso	L'Inverno - pedale
3	Acqua azzurra Acqua chiara - accompagnamento 1	Dancing Queen - accompagnamento 1	L'Inverno - accompagnamento 1
4	Acqua azzurra Acqua chiara - accompagnamento 2	Dancing Queen - accompagnamento 2	L'Inverno - accompagnamento 2
5	Fiori rosa, fiori di pesco - percussionsi	Mamma Mia! - percussionsi	La Primavera - basso
6	Fiori rosa, fiori di pesco - basso	Mamma Mia! - basso	L'Inverno - pedale
7	Fiori rosa, fiori di pesco - accompagnamento 1	Mamma Mia! - accompagnamento 1	La Primavera - canto
8	Fiori rosa, fiori di pesco - accompagnamento 2	Mamma Mia! - accompagnamento 2	La Primavera - controcanto

Tabella 1. Organizzazione delle tracce del gioco

L'ordine in cui le tracce sono rivelate e disposte nei cerchi è completamente casuale e sorteggiato ad ogni partita.

3.3 Metodo

I partecipanti sono stati raggruppati in gruppi di due: ad ogni gruppo è stata fornita una breve introduzione sulla teoria che sta alla base del gioco e sono state spiegate le regole fondamentali. Successivamente ai giocatori è stata presentata la schermata di selezione del repertorio per iniziare la partita. Al termine, è stata data la possibilità di giocare una nuova partita con un repertorio diverso, possibilità molto gradita dalla maggioranza dei partecipanti.

La registrazione delle performance degli utenti è stata integrata nel codice del gioco: sono stati registrati la sequenza di tracce sblocicate, la correttezza ed i tempi di risposta ad ogni confronto e la correttezza ed i tempi di gioco complessivi.

I dati sono stati raccolti e processati mediante uno script Matlab, con lo scopo di focalizzare l'attenzione su

⁵ Le tracce si possono ascoltare singolarmente qui: <http://www.dei.unipd.it/~altieri/GoB/>

come la scelta di un repertorio piuttosto che di un altro, unita alla sequenza di tracce scoperte durante la partita, possano influenzare i tempi di risposta e l'efficacia dell'azione dell'utente.

3.4 Risultati

Durante esecuzione sono state registrate 120 partite. Tutte le partite si sono svolte alla presenza di almeno uno degli autori del gioco che, in qualità di osservatore, si curava di amministrare ed organizzare le partite, contrassegnando quelle che, per qualche motivo (utilizzo improprio dell'applicazione, fraintendimento del gioco o abbandono volontario della sessione), erano prive di significato: questa selezione ha portato un insieme di 99 partite valide, mentre 21 sono state le partite annullate e non tenute in considerazione per l'analisi.

La Tabella 2 riporta i dati globali sui tre diversi repertori:

Repertorio	Partite totali	Partite Vinte	Partite Perse	Dur media partita	Dur media turno
Battisti	31	22 (71%)	9 (29%)	140 sec.	15 sec.
Abba	37	23 (62%)	14 (38%)	138 sec.	15 sec.
Vivaldi	31	20 (64,5 %)	11 (35,5 %)	133 sec.	14 sec.
Totale	99	65 (65,5 %)	34 (34,5 %)	138 sec.	15 sec.

Tabella 2. Dati generali sulle partite valide della "Notte della Ricerca" 2015

I dati riportati in Tabella 2 denotano una distribuzione abbastanza uniforme del campione sui tre repertori. La durata delle partite e dei turni si è attestata all'incirca sugli stessi valori, mentre il rapporto vittorie/sconfitte varia leggermente, suggerendo una maggiore difficoltà nel secondo repertorio rispetto agli altri.

Volendo andare più in profondità, una delle analisi più interessanti che si possono fare sui dati raccolti è inerente all'esito delle partite osservato in base al repertorio scelto ed alla prima traccia. I dati sono riportati nelle tabelle 3, 4, 5.

I dati suggeriscono come, anche all'interno degli stessi repertori, non tutte le combinazioni abbiano lo stesso rapporto di vittorie e sconfitte.

Battisti	1	2	3	4	5	6	7	8
Sconfitte	0	0	0	1	1	2	2	3
Vittorie	1	1	1	2	8	6	2	1

Tabella 3. Risultati delle partite giocate col repertorio "Battisti" esaminate in base alla prima traccia rivelata

ABBA	1	2	3	4	5	6	7	8
Sconfitte	2	2	1	3	2	2	2	0
Vittorie	0	1	1	4	2	5	5	5

Tabella 4. Risultati delle partite giocate col repertorio “ABBA” esaminate in base alla prima traccia rivelata

Vivaldi	1	2	3	4	5	6	7	8
Sconfitte	1	1	3	2	1	1	1	1
Vittorie	2	1	0	4	1	6	3	3

Tabella 5. Risultati delle partite giocate col repertorio “Vivaldi” esaminate in base alla prima traccia rivelata

In particolare, ci sono alcune combinazioni che si sono rivelate più semplici per gli utenti, come il repertorio ‘Battisti’ con le percussioni della traccia *Fiori rosa, Fiori di Pescio*, o come il repertorio ‘Vivaldi’ ed una delle parti di controcanto della *Primavera*. Queste considerazioni, tuttavia, non sono state avallate dal test del chi quadrato, fatto probabilmente imputabile all’aver un campione troppo scarno.

Approfondendo il ragionamento, sarebbe interessante capire quali confronti tra tracce siano risultati più facili o più difficili per gli utenti. Anche in questo caso, tuttavia, i dati in nostro possesso non sono sufficienti ad avere un campione statistico rilevante. Lo studio di questo aspetto legato alla percezione potrebbe essere interessante per sviluppi ed esperimenti futuri.

I dati qui presentati sono indicativi della difficoltà dei repertori percepita dagli utenti. Il chi-square test applicato ai dati di Tabella 1 ($\rho \leq 0.05$) esclude una differenza statisticamente rilevante dei risultati di uno dei tre repertori rispetto agli altri due. Questo indica che la scelta degli stimoli è stata condotta in maniera da equilibrare fra tutti i repertori la difficoltà del compito richiesto.

L’altro elemento oggetto del nostro interesse è il fattore di engagement, misurato empiricamente osservando il comportamento dei giocatori. Gli osservatori presenti durante la ‘Notte dei Ricercatori’ hanno rilevato un’ottima risposta di pubblico e reazioni soddisfatte ed entusiaste dei giocatori: soltanto 3 utenti hanno abbandonato la sessione di gioco prima di terminare (si trattava di bambini troppo piccoli per capirne le dinamiche ed attratti dal connubio di luci, immagini e suoni). Più del 60% degli utenti ha invece chiesto di poter giocare più partite e circa il 30% ha provato a giocare con tutti e 3 i repertori.

4. CONCLUSIONI

In questo lavoro abbiamo presentato un’applicazione per lo stimolo e il potenziamento dell’abilità di ascolto attivo chiamata *Good or Bad?*, realizzata presso i laboratori di Sound and Music Computing del Dipartimento di Ingegneria dell’Informazione dell’Università degli Studi di Padova. L’applicazione è stata progettata per stimolare l’attenzione degli ascoltatori in merito all’ascolto simultaneo di più tracce musicali, chiedendo loro di determi-

nare se le tracce in questione appartengano o no alla medesima canzone. Abbiamo sperimentato l’uso dell’applicazione presso l’edizione 2015 della ‘Notte Europea dei Ricercatori’. Dai dati raccolti è emersa una diversa performance degli utenti rispetto ai diversi repertori impiegati ed alla traccia utilizzata come riferimento, oltre a un fattore di engagement che risulta essere molto promettente.

I dati raccolti in questo setup sperimentale ecologico hanno fornito risultati incoraggianti sulla robustezza dell’applicazione in ambienti non controllati e sul grado di apprezzamento da parte di un’ampia tipologia di utenti. Tuttavia, restano da approfondire questioni di rilevante interesse dal punto di vista scientifico. Qual è l’efficacia di questo tipo di applicazioni nel potenziare l’ascolto attivo? Che tipo di influenza (in positivo e/o in negativo) esercita l’aspetto sociale delle applicazioni realizzate mediante spazi interattivi? Qual è il contributo dell’interazione a full-body, rispetto ad altre tipologie di interazione più tradizionali, in relazione al coinvolgimento dell’utente e all’efficacia dell’applicazione?

L’indagine di questi aspetti sarà l’obiettivo di una serie di esperimenti che saranno condotti in laboratorio, utilizzando gruppi sperimentali e di controllo.

Il lavoro su *Good or Bad?* continuerà quindi su tre fronti: lo studio dell’applicazione in ambiente controllato per la verifica dell’efficacia degli spazi interattivi, lo sviluppo e lo studio di un’interfaccia web per una fruizione surrogata dell’applicazione e l’inquadramento di questi lavori in un progetto più ampio di sviluppo di una piattaforma integrata per l’apprendimento.

4.1 Studio in ambiente controllato

Un esperimento in ambiente controllato è attualmente in fase esecutiva presso i laboratori di Sound and Music Computing del Dipartimento di Ingegneria dell’Informazione dell’Università degli Studi di Padova. Lo studio ha due obiettivi: dimostrare l’efficacia dell’applicazione nello stimolo dell’abilità dell’ascolto attivo e il maggior engagement di *Good or Bad?* con interfaccia implementata mediante spazio interattivo, rispetto ad una versione surrogata con interfaccia proposta mediante touch screen.

4.1.1 Il setup sperimentale

Gli esperimenti saranno condotti su un gruppo di utenti separati in due insiemi distinti: uno di essi utilizzerà l’applicazione nella sua forma originale, ossia su pavimento interattivo, l’altro utilizzerà una versione con input tramite touch screen.

In entrambi i casi, il setup scientifico prevede un test in 3 fasi.

In una prima fase, il pre-test, gli utenti sono chiamati a rispondere su 18 singoli confronti fra tracce. I confronti saranno svolti su 3 repertori inediti, ma stilisticamente affini a quelli del gioco e saranno, proprio come avviene nel gioco, tra una serie (una o più) tracce di riferimento e una traccia di “sfida”.

Nella seconda fase, il soggetto disputa alcune partite con l'applicazione, dove le tracce non sono proposte in ordine casuale, ma in un ordine stabilito dagli autori e basato sui dati raccolti in occasione della ‘Notte Europea dei Ricercatori’. Durante le partite il soggetto sarà posto nella zona dei pulsanti *Good* e *Bad*, mentre uno degli autori collaborerà camminando sui cerchi interattivi, ma non interferirà in nessun modo nella risposta. I soggetti che utilizzeranno il touch screen giocheranno da soli.

La terza fase è un post test somministrato sullo stesso insieme di repertori e con le stesse modalità del pre-test.

Infine, prima di essere congedati, ai soggetti sarà somministrato un questionario per la valutazione dell'engagement. Il test è stato realizzato sul modello di Moreno et al [17].

4.2 Versione online

Sarà realizzata una versione online di *Good or Bad?*. Lo scopo è rendere disponibile ovunque una versione accessibile del gioco. Non essendo ovviamente trasportabile la struttura di tracking AILearn, l'accesso all'applicazione sarà surrogato in una applicazione web. L'idea è quella di effettuare un porting del gioco su *Unity*, cercando di sfruttare le caratteristiche salienti del popolare engine di gioco, come la gratuità per applicazioni no-profit, l'approccio multiplattaforma e le capacità di gestire ambienti 3D decisamente più gradevoli rispetto all'interfaccia 2D qui proposta.

L'applicazione sarà successivamente soggetta a valutazione ed i dati saranno utilizzati per effettuare un confronto tra le varie versioni dell'applicazione, con lo scopo di raccogliere dati avendo a disposizione un ampio numero potenziale di utenti.

4.3 Una piattaforma integrata

Si pensa infine di realizzare una piattaforma integrata per applicazioni educative di questo tipo. La piattaforma integrata potrebbe essere un *MooC* (acronimo per “Massive Open Online Courses”, in italiano “Corsi aperti online su larga scala”), che rappresenta la galassia di piattaforme didattiche online pensate per l'organizzazione e il miglioramento dell'erogazione della didattica, come, ad esempio, la piattaforma *Moodle*⁶), che consentirebbe l'utilizzo di queste applicazioni in ambito scolastico. Le performance e le esercitazioni sarebbero tracciate e rese monitorabili da supervisori e docenti, che potrebbero assegnare anche compiti personalizzabili.

Si vorrebbero inoltre studiare applicazioni che prevedano la presenza simultanea di più tipi di giocatori con ruoli e metodi d'interazione diversi (come combinazioni di spazio interattivo e controllo con dispositivo mobile).

5. REFERENZE

- [1] F. Pachet: “Music listening: What is in the air?”, *Tech. rep., Sony Computer Science Laboratory Internal Report*, 1999.
- [2] J. A. Sloboda: *The musical mind: The cognitive psychology of music*, Oxford University Press, 1985.
- [3] G. Volpe and A. Camurri: “A system for embodied social active listening to sound and music content”, *Journal on Computing and Cultural Heritage (JOCCH)* Vol. 4, No. 1, 2011.
- [4] M. Leman, M. Demey, M. Lesaffre, L. Van Noorden and D. Moelants: “Concepts, technology, and assessment of the social music games ‘Sync-in-Team’”, *Computational Science and Engineering, 2009. CSE'09 International Conference*, pp. 837–842, 2009.
- [5] F. Pachet, O. Delerue, and P. Hanappe, “Dynamic audio mixing” *Proceedings of ICMC*, 2000.
- [6] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey and J. Zigelbaum: “Reality-based interaction: a framework for post-wimp interfaces” *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 201–210, 2008.
- [7] J. Paradiso, C. Abler, K.-y. Hsiao and M. Reynolds: “The magic carpet: physical sensing for immersive environments” *CHI'97 Extended Abstracts on Human Factors in Computing Systems*, pp. 277–278, 1997.
- [8] F. Morreale, A. De Angeli, R. Masu, P. Rota and N. Conci: “Collaborative creativity: The music room” *Personal and Ubiquitous Computing* Vol. 18, No. 5, pp.1187–1199, 2014.
- [9] M. Mandanici, A. Rodà and S. Canazza: “The harmonic walk: An interactive physical environment to learn tonal melody accompaniment”, *Advances in Multimedia*, 2016.
- [10] F. Avanzini, S. Spagnol, A. De Götzen & A. Rodá: “Designing Interactive Sound for Motor Rehabilitation Tasks”. Chapter 12 in K. Franinovic & S. Serafin: *Sonic Interaction Design*, pp. 273–283, 2013.
- [11] N. Bianchi-Berthouze, W. W. Kim and D. Patel: “Does body movement engage you more in digital game play? and why?” *Affective computing and intelligent interaction* Springer, pp. 102–113, 2007.
- [12] Y. A. De Kort and W. A. Ijsselsteijn: “People, places, and play: player experience in a socio-spatial context”, *Computers in Entertainment (CIE)* Vol. 6 No. 2, 2008.

⁶ <https://moodle.org/?lang=it>

- [13] M. Munaro, F. Basso and E. Menegatti: "OpenPTrack: Open Source Multi-Camera Calibration and People Tracking for RGB-D Camera Networks." *Journal on Robotics and Autonomous Systems*, Vol. 75, part B, pp. 525-538, 2016.
- [14] M. Munaro, A. Horn, R. Illum, J. Burke and R. B. Rusu: "OpenPTrack: People Tracking for Heterogeneous Networks of Color-Depth Cameras" *IAS-13 Workshop Proceedings: 1st Intl. Workshop on 3D Robot Perception with Point Cloud Library*, pp. 235-247, 2014.
- [15] M. Munaro and E. Menegatti: "Fast RGB-D People Tracking for Service Robots" *Journal on Autonomous Robots*, Vol. 37, No. 3, pp. 227-242, Springer, 2014.
- [16] M. Wright, A. Matthew, and Adrian Freed: "Open sound control: A new protocol for communicating with sound synthesizers." *Proceedings of the 1997 International Computer Music Conference*. Vol. 2013. No. 8, 1997.
- [17] A. Moreno, R. van Delden, R. Poppey and D. Reidsma, D. Heylen: "Augmenting Traditional Playgound Games to Enhance Game Experience", *Proceedings of INETAIN conference*, 2015.

SELFIE ACUSTICHE CON IL PROGETTO SELFEAR: UN'APPLICAZIONE MOBILE PER L'ACQUISIZIONE A BASSO COSTO DI PINNA-RELATED TRANSFER FUNCTION

Michele Geronazzo

Università degli Studi di Verona

Dip. di Neuroscienze, Biomedicina e Movimento
michele.geronazzo@univr.it

Jacopo Fantin, Giacomo Sorato, Guido Baldovino, Federico Avanzini

Università degli Studi di Padova

Dip. di Ingegneria dell'Informazione

corrispondenza a avanzini@dei.unipd.it

ABSTRACT

Le esperienze di realtà virtuale e aumentata stanno riscontrando una grande diffusione e le tecnologie per la spazializzazione del suono in cuffia saranno fondamentali per la diffusione di scenari applicativi immersivi in supporti mobile. Questo articolo affronta le problematiche legate alla acquisizione di *head-related transfer function* (HRTF) con dispositivi a basso costo, accessibili a chiunque, in qualsiasi luogo e che forniscano delle misurazioni fruibili in tempi brevi. In particolare la soluzione proposta denominata "the SelfEar project" si focalizza sull'acquisizione delle trasformazioni spettrali ad opera dell'orecchio esterno contenute nella *pinna-related transfer function* (PRTF); l'utente viene guidato nella misurazione di HRTF in ambiente non anecoico attraverso una procedura auto-regolabile. Le informazioni acustiche sono infatti acquisite tramite un headset per la realtà acustica aumentata che include un set di microfoni posizionati in prossimità dei canali uditivi dell'ascoltatore. Proponiamo una sessione di misurazione con l'obiettivo di acquisire le caratteristiche spettrali della PRTF di un manichino KEMAR, confrontandoli con i risultati che si otterrebbero con una procedura in ambiente anecoico. In entrambi i casi i risultati dipendono fortemente dalla posizione dei microfoni, senza considerare in questo scenario il problema legato ai movimenti di un eventuale soggetto umano. Considerando la qualità generale e la variabilità dei risultati, così come le risorse totali necessarie, il progetto SelfEar propone una promettente soluzione per una procedura a basso costo di acquisizione di PRTF, e più in generale di HRTF.

1. INTRODUZIONE

L'obiettivo delle tecnologie binaurali è riprodurre suoni che risultino i più realistici e naturali possibili, illudendo l'ascoltatore di essere circondato da sorgenti sonore reali a lui esterne. Tale tecnologia risale alla fine del 20° secolo [1]; consiste nella registrazione di suoni attraverso una testa artificiale che simuli le caratteristiche di quella dell'ascoltatore.

Copyright: ©2016 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.

incorporando due capsule microfoniche nei condotti uditi, modellando le membrane del timpano [2]. Questa tecnologia ottiene la massima efficienza attraverso la riproduzione in cuffia, che mantiene intatte le caratteristiche del segnale, senza riflessioni e riverberi dell'ambiente.

La resa di ambienti acustici virtuali coinvolge le cosiddette *binaural room impulse responses* (BRIR) nelle quali è possibile riconoscere due principali componenti: la prima è connessa con le caratteristiche ambientali, descritte nella *room impulse response* (RIR), l'altra è correlata agli aspetti antropometrici dell'ascoltatore, contenuti nella *head-related impulse response* (HRIR) [2]. Tutte queste risposte all'impulso hanno la loro controparte nel dominio della frequenza, ovvero nella loro trasformata di Fourier: *binaural room transfer function* (BRTF), *room transfer function* (RTF), e *head-related transfer function* (HRTF) rispettivamente. In particolare, le HRTF descrivono un filtro lineare tempo-invariante che definisce il filtraggio acustico al quale contribuiscono testa, torso e orecchio esterno dell'ascoltatore.

Il processo di acquisizione di HRTF individuali è attualmente molto costoso in termine di attrezzature e tempo. La misurazione acustica in camera anecoica offre una risposta all'impulso di controllo che è di alta qualità e precisione; richiede però un attrezzatura molto costosa e difficilmente reperibile per applicazioni reali. È quindi inevitabile sacrificare alcune caratteristiche individuali del soggetto per ottenere una rappresentazione della HRTF più economica ma che dia comunque un'informazione psico-acustica accurata [3]. Il processo di acquisizione di HRTF in ambiente domestico è una questione complessa; le tendenze più recenti sono supportate da dispositivi a basso costo per l'acquisizione di mesh 3D [4] e da algoritmi per la modellazione e personalizzazione delle HRTF [5]. Purtroppo queste soluzioni mancano di informazioni robuste ed individuali per l'acustica dell'orecchio esterno (*pinna*) a causa di una dettagliata struttura antropometrica. Queste informazioni sono contenute nella cosiddetta *pinna-related transfer function* (PRTF) [6] che risulta molto difficile da modellare computazionalmente anche nel contesto di simulazioni numeriche [7]. Le PRTF contengono indizi salienti sulla localizzazione per quanto riguarda la percezione dell'elevazione (vedere [8] per una bibliografia più approfondita), perciò è necessaria una rappresentazione accurata per fornire la percezione della dimensione verticale nelle tecnologie di audio binaurale.

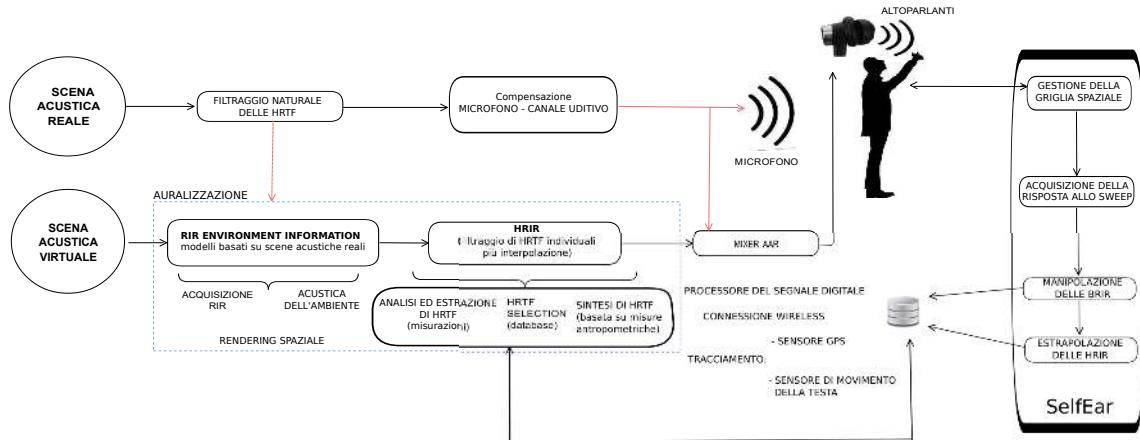


Figure 1: Vista schematica del progetto SelfEar in un contesto di realtà audio aumentata.

Questo articolo affronta il problema della riduzione dei costi nel processo di acquisizione/misurazione di HRTF, con particolare attenzione all'estrapolazione di PRTF per sistemi mobile di realtà audio aumentata (*mobile audio augmented reality* - mAAR). Questi sistemi sono composti di cuffie provviste di microfoni esterni integrati per l'acquisizione di flussi audio multi-canale dall'ambiente, oltre ad algoritmi per la riproduzione di audio binaurale. Una possibilità per poter ottenere ovunque le proprie HRTF utilizza i microfoni integrati a partire da stimoli sonori riprodotti dagli altoparlanti del dispositivo mobile; il progetto Self-Ear ha lo scopo di sviluppare e implementare gli algoritmi di elaborazione del segnale e l'interazione col dispositivo per ottenere una procedura auto-regolabile dall'utente.

Sono stati condotti solamente pochi studi che mirino alla consistenza delle HRTF acquisite in ambiente non anecoico e che considerino il contributo acustico del piano mediano-sagittale [9], che è di rilevante importanza per il contenuto spettrale individuale introdotto nelle PRTF. Il compromesso tra costi e portabilità conduce a due problematiche principali. La prima riguarda il fatto che il processo di acquisizione tramite dispositivo mobile implica l'influenza dell'ambiente circostante, con colorazioni in frequenza e interferenze da ritardi. In secondo luogo, impiegare gli speaker di uno smartphone o tablet come sorgente sonora e microfoni binaurali di mercato *consumer* per l'acquisizione comporta registrazioni meno accurate rispetto a quelle ottenute con attrezzatura professionale.

In questo articolo presentiamo una serie di misurazioni condotte in un ambiente controllato su un manichino KE-MAR [10]. Il nostro obiettivo finale è confrontare risposte ottenute grazie a SelfEar con quelle ottenute con attrezzatura professionale; di seguito la struttura dell'articolo: Sez. 2 contiene la descrizione di un sistema di realtà audio aumentata mobile e i criteri per l'esternalizzazione del suono virtuale; nella Sez. 3 viene presentato il progetto SelfEar. La Sez. 4 descrive gli esperimenti acustici su di una testa artificiale in ambiente non anecoico. All'interno della Sez. 6 vengono discussi i risultati ottenuti e la Sez. 7 conclude la valutazione preliminare esposta con l'esposizione delle più promettenti direzioni di ricerca per il futuro.

2. REALTÀ AUDIO AUMENTATA IN CONTESTO MOBILE

In un sistema di tipo mAAR, l'ascoltatore può usufruire di un mix di sorgenti sonore reali e virtuali. Le prime vengono riprese dai microfoni dell'headset previo filtraggio naturale del soggetto con conseguente applicazione di una compensazione in frequenza-fase che tenga conto dell'errore introdotto dalla differenza di posizione dei microfoni rispetto al punto d'ingresso del canale auditivo [11]. Le sorgenti virtuali necessitano invece di un processo di auralizzazione dinamico per ottenere una sovrapposizione ideale con la realtà. L'auralizzazione comprende un rendering tramite BRIR, risultante nel filtraggio tramite RIR e HRIR che devono essere personalizzate in accordo con l'ambiente circostante [12] e con l'ascoltatore [3]. Gli algoritmi di signal processing (DSP) implementano filtri correttivi che compensano i microfoni, gli speaker e le loro interazioni, prendendo in considerazione gli effetti psicoacustici e di colorazioni causati dall'indossare la cuffia rispetto alle normali condizioni di ascolto senza di essa [13].

Produrre scenari acustici virtuali e aumentati in cuffia con attenzione alle caratteristiche di ambiente e di ascoltatore rimane una delle sfide più importanti in questo dominio di ricerca, a causa delle forti interconnessioni presenti tra i vari componenti del sistema mAAR che concorrono ad una corretta ““esternalizzazione”” della scena acustica.¹ I criteri per una corretta auralizzazione ed esternalizzazione possono essere riassunti quindi in quattro categorie:

- *sistema ergonomico*: le cuffie ideali dovrebbero essere acusticamente trasparenti: l'ascoltatore non si dovrebbe rendere conto della loro presenza. A tale scopo sono essenziali cuffie poco invasive nel peso e nelle dimensioni;
- *tracciamento*: i movimenti della testa nell'ascolto binaurale producono segnali interaurali dinamici [14];

¹ Con esternalizzazione intendiamo la percezione di un suono fuori dalla testa, piuttosto che al suo interno.

tracciare la posizione dell’utente nell’ambiente permette il riconoscimento dell’interazione acustica e una rappresentazione spaziale coerente tra scena reale e simulazione virtuale;

- *conoscenza dell’ambiente*: la percezione spaziale della scena acustica richiede la conoscenza di riflessioni e riverberazioni tipiche dell’ambiente stesso; queste informazioni risultano essenziali per una impressione realistica dello spazio in cuffia [15];
- *individualizzazione*: testa e orecchio del soggetto filtrano in maniera individuale i suoni; va presa in considerazione una correzione individuale per l’accoppiamento acustico tra cuffie e orecchio esterno [16].

3. IL PROGETTO SELFEAR

3.1 Panoramica del sistema

SelfEar è un’applicazione progettata per piattaforma Android con lo scopo di ottenere le HRIR personali dell’utente a partire da uno stimolo sonoro riprodotto dal dispositivo mobile utilizzato. Il telefono/tablet deve essere tenuto in mano con il braccio teso e spostato lungo il piano mediano del soggetto fermandosi a specifiche elevazioni. I microfoni in-ear catturano il segnale acustico proveniente dagli altoparlanti del dispositivo, registrando in questo modo le BRIR specifiche per posizione, ascoltatore e ambiente.

I dati collezionati tramite smartphone possono essere successivamente impiegati per ottenere una HRIR customizzata. Dopo le procedure di post-processing che compensano gli effetti acustici delle condizioni di acquisizione e riproduzione, le HRIR individuali possono direttamente essere usate per la resa spaziale del suono. A seconda della complessità degli scenari virtuali, la sintesi in tempo reale delle HRIR è oggi possibile in ambiente mobile. Una possibile soluzione impiega la selezione di HRTF misurate in camera anechoica (ad esempio il database CIPIC [17])² pre-esistenti in basi di dati pubbliche, basandosi sui parametri acustici estratti con SelfEar: la procedura seleziona la migliore approssimazione per l’HRTF dell’utente SelfEar basandosi su metriche psicoacustiche e di similarità antropometriche [18].

3.2 Gestione della sorgente

Il sistema di gestione della griglia spaziale di SelfEar guida l’utente attraverso il processo di acquisizione delle BRIR definendo una procedura autoregolata descritta in Fig. 2. Nel paragrafo seguente descriviamo ogni passo, partendo dal lancio dell’applicazione fino al termine della sessione, che ha come risultato la restituzione di un insieme di BRIR individuali.

Nella schermata iniziale di SelfEar, l’utente seleziona la posizione degli speaker del dispositivo: questi possono trovarsi sul lato superiore, frontale, inferiore o sul retro. Questa scelta determinerà l’orientamento del dispositivo

² Una collezione di diverse misurazioni acustiche condotte su 50 soggetti distinti (più di 1200 misurazioni ciascuno), che comprendono anche informazioni antropometriche.

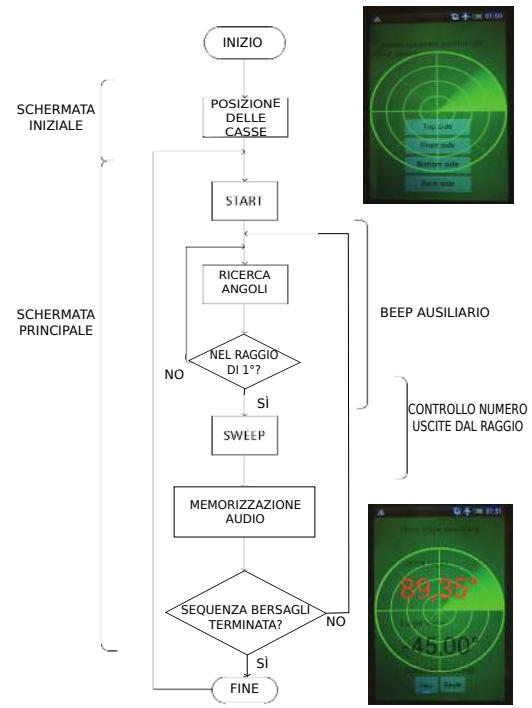


Figure 2: Diagramma a blocchi della procedura SelfEar per l’acquisizione di BRIR nel piano mediano. Vengono riportati anche degli screenshot delle due schermate dell’applicazione.

durante la riproduzione dello stimolo sonoro per massimizzare il rendimento degli altoparlanti del dispositivo a seconda della caratteristica di direzionalità. L’utente può in seguito premere il pulsante “Start” per iniziare la procedura di acquisizione delle BRIR. Prima di raggiungere la prima elevazione target, si istruisce l’utente nel portare il dispositivo (sostenuto dal braccio teso) a livello degli occhi, permettendo così la creazione di un riferimento proprioettivo/motorio per l’estrazione di tutti gli altri angoli; in seguito i passi seguono questa sequenza logica:

1. *Raggiungimento del bersaglio*: l’attuale elevazione del dispositivo rispetto all’orizzonte, ϕ_i , compare sopra all’elevazione bersaglio da raggiungere. SelfEar elabora i dati provenienti dall’accelerometro lungo i tre assi cartesiani, $a_{x,y,z}$, del dispositivo per calcolare ϕ_i tramite la formula seguente:

$$\phi_i = \arctan \left(\frac{\pm a_y}{|a_z|} \right)$$

nel caso gli altoparlanti siano posizionati sul lato superiore o inferiore; mentre con la formula:

$$\phi_i = \arctan \left(\frac{\pm a_z}{|a_y|} \right)$$

nel caso si trovassero invece sul lato frontale o sul retro. Il segno del numeratore è:

- + per altoparlanti sul lato inferiore o sul retro;
- per altoparlanti sul lato superiore o frontale.

La sequenza delle elevazioni bersaglio spazia in ordine crescente tra gli angoli del CIPIC database equispaziati ogni 5.625° in un intervallo di $[-40^\circ, 40^\circ]$. Un beep ausiliario rende acusticamente l'informazione di differenza tra l'attuale elevazione del dispositivo e quella bersaglio, agevolando la procedura di puntamento soprattutto nei casi in cui lo schermo del dispositivo non è visibile a causa della posizione degli speaker (ad es. quando si trovano sul retro). La pausa tra un beep ed il successivo è direttamente proporzionale alla differenza tra ϕ_i , e l'angolo bersaglio, $\hat{\phi}_i$, come mostrato nell'equazione seguente:

$$pause_i = |\phi_i - \hat{\phi}_i| \cdot k$$

dove i è un istante in cui un singolo beep termina la sua riproduzione e k è una costante che rende percepibile la pausa tra due beep consecutivi.³ L'obiettivo di questo step è di avvicinarsi a $\hat{\phi}_i$ con una precisione di $\pm 1^\circ$. Questo passo può venire interrotto e ripreso su richiesta dell'utente.

2. *Controllo della posizione*: una volta che ϕ_i entra nel range di validità di $\pm 2^\circ$ dal bersaglio, scatta un timer di stabilizzazione di 2 secondi; nel caso l'utente uscisse dal range per tre volte prima che il timer finisca, la procedura inizia nuovamente a partire dalla fine del passo 1.
3. *Riproduzione dello sweep*: al termine del timer di stabilizzazione, viene riprodotto lo stimolo sonoro dagli altoparlanti del dispositivo; nel caso l'utente uscisse dal range di validità di $\pm 2^\circ$ solamente una volta durante la riproduzione dello sweep, la procedura di ricerca di ϕ_i viene annullata.
4. *Salvataggio delle BRIR*: quando uno sweep termina con successo, il segnale audio registrato viene salvato localmente in memoria insieme all'angolo di elevazione a cui si riferisce; la procedura ritorna quindi al passo 1 con la successiva elevazione bersaglio in sequenza.
5. *Fine della sessione*: una sessione termina quando tutte le elevazioni bersaglio sono state raggiunte con successo.

4. MISURAZIONI ACUSTICHE

Le sessioni sperimentali sono state eseguite in un ambiente non anechoico usando un manichino KEMAR che ci ha permesso di minimizzare gli errori dovuti al movimento di un eventuale soggetto umano. Ci siamo quindi concentrati sulla direzione con il più alto contenuto di caratteristiche

³ La formula restituisce un valore in millisecondi, che senza una costante moltiplicativa darebbe luogo ad una pausa troppo corta per poter essere percepita. In questa realizzazione è stato scelto un valore arbitrario di $k = 5$.

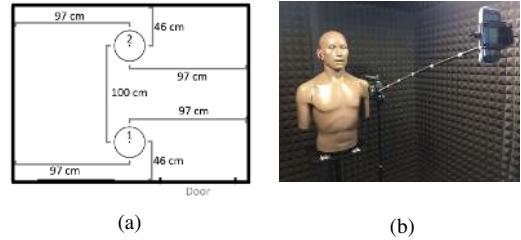


Figure 3: Setup per le misurazioni. (a) Posizioni di sorgente e ricevitore. (b) Setup di misurazione di Selfear, con selfie-stick incorporato.

spettrali, ovvero la direzione frontale [19], $\phi_i = 0$, fornendone un'analisi dettagliata. Infine abbiamo riportato una valutazione qualitativa dell'applicazione SelfEar per una serie di PRIR nel piano sagittale frontale.

4.1 Setup

Struttura e attrezzature - Tutte le misurazioni sono state condotte all'interno di una camera silente di 2×2 m, una Sound Station Pro 45 (SSP), con un isolamento acustico massimo di 45 dB.

La Figura 3a mostra l'interno della SSP dove sono identificabili due posizioni: la posizione #1 per le varie sorgenti acustiche e la posizione #2 per i vari ricevitori.

Negli esperimenti sono stati usati due tipi di dispositivi di riproduzione (con definizione acronimi):

- *L* : un loudspeaker Genelec 8030A calibrato in modo da avere un buon SNR con tono di test a 500 Hz con 94 dB SPL;
- *S* : uno smartphone HTC Desire C sorretto da una struttura autoprodotta che simula un braccio umano, realizzata con un selfie stick;⁴ in questo caso la SPL massima raggiunta con il tono di test è di 51 dB alla frequenza di riferimento di 500 Hz.

E due tipi di ricevitori (con definizione acronimi):

- *H* : un paio di in-ear headphones Roland CS-10EM con due microfoni incorporati;
- *K* : due microfoni professionali G.R.A.S forniti dal simulatore di testa e torso KEMAR; nella configurazione proposta, l'orecchio destro era equipaggiato con il simulatore di canale uditivo, mentre l'orecchio sinistro ne era sprovvisto.

In tutti gli esperimenti, sorgenti e ricevitori, nelle rispettive posizioni, sono sempre stati posizionati alla stessa altezza dal pavimento. Il segnale sorgente utilizzato è un sine sweep logaritmico della durata di 1 -s, il cui contenuto in frequenza spazia uniformemente da 20 Hz a 20 kHz. I segnali acustici sono stati registrati con il software Audacity

⁴ La struttura a selfie stick, di lunghezza 1-m, è più lunga di un braccio medio in estensione; riteniamo questa discrepanza ininfluente nell'estrazione delle caratteristiche spettrali della PRTF per la percezione dell'elevazione in quanto queste si possono assumere invarianti con la variazione della distanza dalla sorgente [20].

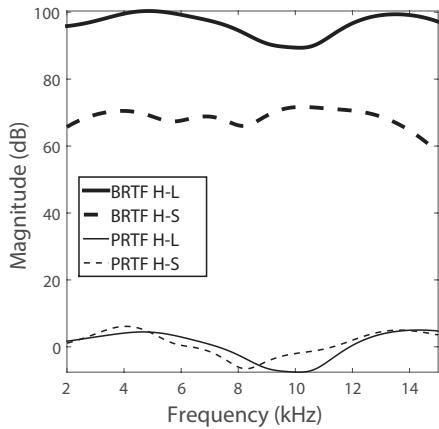


Figure 4: Confronto tra i moduli (in dB SPL) delle BRTF (linee spesse) e le relative PRTF (linee sottili) ottenute usando: come ricevitore - il microfono destro delle cuffie Roland (H); come sorgente - il loudspeaker dello smartphone (S, linee tratteggiate) e il loudspeaker Genelec (L, linee continue).

attraverso un’interfaccia audio Motu 896 mk 3 e sono stati poi processati usando il software Matlab aggiornato alla versione 8.4.

Calibrazione con diffuse-field - Per il calcolo della risposta *diffuse-field* è stata realizzata una struttura apposita composta da due cavi di filo metallico agganciati al soffitto della SSP distanziati di 17.4 cm, ovvero la distanza tra i microfoni della KEMAR. Questa misurazione ci permette di acquisire le specifiche acustiche dell’ambiente e del setup.

Abbiamo acquisito le misurazioni diffuse-field per tutte le combinazioni di sorgente e ricevitore, ottenendo un totale di quattro misurazioni.

4.2 Dati

Prima sessione di misurazione - In questa sessione sono stati posizionati all’interno della SSP, nelle rispettive posizioni (#1 e #2 della Fig. 3a), lo speaker Genelec e la KEMAR. Come primo step sono state misurate le risposte dell’orecchio destro e sinistro. Successivamente è stata inserita nel canale uditivo destro della KEMAR la cuffietta Roland; sono state effettuate dieci misurazioni con questa configurazione riposizionando ogni volta le Roland; in tal modo è stato possibile analizzare la variabilità introdotta nelle misurazioni dalla posizione del microfono.

Seconda sessione di misurazione - In questa sessione è stata posizionata in #1 (vedi Fig. 3a) la struttura di sostegno per lo smartphone; nella posizione #2 di Fig. 3a è invece stata posizionata la KEMAR. Il selfie-stick manteneva lo smartphone ad una distanza di un metro dalla KEMAR e consentiva una regolazione angolare sul piano mediano. Sono state acquisite le misurazioni relative a 15 angoli tra -40° e $+40^\circ$ sul piano mediano riproponendo gli angoli presenti nel CIPIC database. Quindi, al termine della misurazione abbiamo ottenuto due set di 15 misurazioni per

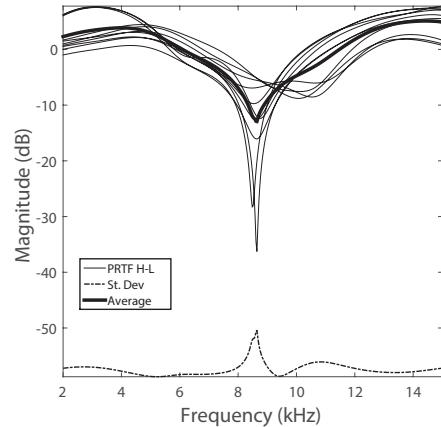


Figure 5: Moduli delle PRTF per i dieci riposizionamenti della cuffia Roland sul manichino KEMAR. La linea spessa rappresenta l’ampiezza media. La deviazione standard è spostata di -60 dB per agevolarne la lettura.

l’orecchio sinistro della KEMAR (senza cuffietta) e per il microfono destro della cuffietta.

5. RISULTATI

Per ogni misurazione è stato calcolato l’onset applicando la funzione di cross-correlazione con il segnale di sweep originale; è stata poi estratta la BRIR deconvolvendo la risposta allo sweep con lo sweep stesso. Sono state successivamente rimosse riflessioni dovute dalla SSP e dalla strumentazione al suo interno, tale procedura è stata effettuata sottraendo alle BRIR le rispettive risposte di tipo *diffuse-field*. Da questa procedura sono state estratte le HRTF. Mentre le PRTF sono state ottenute applicando ad ogni risposta impulsiva una finestra di hanning di 1 ms (48 campioni) centrata temporalmente sul picco massimo e normalizzando rispetto al valore massimo in ampiezza [6]. Tutte le PRTF normalizzate sono state filtrate con un passa banda tra 2 kHz and 15 kHz, mantenendo i picchi ed i notch dovuti alla presenza della pinna.

La Figura 4 mostra il confronto tra i moduli in dB SPL delle BRIR estratte dalle misurazioni usando come sorgente (i) il loudspeaker Genelec (ii) lo smartphone, e come ricevitore il microfono destro delle cuffie Roland inserito nell’orecchio destro della KEMAR. Si può notare una differenza di 30 dB tra la SPL dei due loudspeaker, tale differenza porta ad un SNR minore mentre si usa il loudspeaker dello smartphone. Nella stessa figura sono mostrate le due PRTF normalizzate corrispondenti in modo da poter valutare il contributo introdotto dal *diffuse-field*. Tale contributo è sensibilmente visibile per le misurazioni ottenute con lo smartphone HTC Desire C, dovuto principalmente alla diversità di direzionalità tra lo speaker a basso costo integrato e loudspeaker Genelec.

In Fig. 5, sono riportati i moduli in dB delle PRTF dei dieci riposizionamenti insieme alla loro media. Vi è anche riportata la deviazione standard in modo da poter analizzare la viariabilità introdotta nelle misurazioni dalla po-

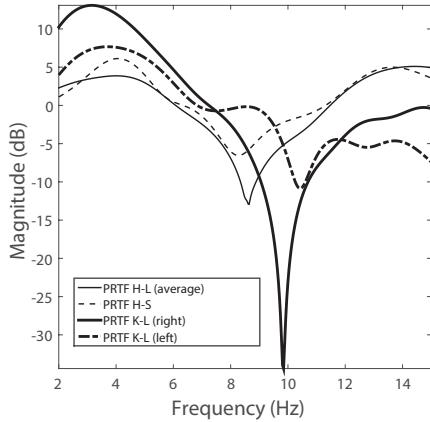


Figure 6: Confronto tra i moduli delle seguenti PRTF: PRTF media della Fig.5 (PRTF H-L); sorgente: smartphone - ricevitore: microfoni delle cuffie (PRTF H-S); sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio destro della KEMAR con canale uditorio (PRTF K-L right); sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio sinistro della KEMAR senza canale uditorio (PRTF K-L left).

sizione delle cuffie/microfono. La variabilità massima la si osserva in prossimità dei notch salienti della PRTF, tra 9 e 11 kHz, che sono molto sensibili ai cambiamenti topologici tra la cuffie e la struttura dell'orecchio [7].

Le principali valutazioni sono state effettuate nella posizione frontale, $\phi = 0$, comparando le PRTF normalizzate in condizioni differenti. La Fig. 6 mostra il confronto tra il modulo delle PRTF acquisite con e senza cuffiette, coinvolgendo sia la Genelec che lo smartphone.

Per le quattro PRTF è stata calcolata la distorsione spettrale media (SD) [8] tra tutte le coppie in un range di frequenze tra $2 \text{ kHz} \leq e \leq 15 \text{ kHz}$ (i valori sono visibili in Tabella 1). Questi confronti permettono diverse considerazioni:

- Acustica della pinna, $K\text{-}L_{right}$ vs. $K\text{-}L_{left}$: le forme diverse delle orecchie e l'acustica del canale uditorio del KEMAR differiscono notevolmente; questa differenza è visibile in tutti i confronti che coinvolgono $K\text{-}L_{left/right}$;
- Loudspeaker, $H\text{-}S_{right}$ vs. $H\text{-}L_{right}$: entrambi i loudspeaker introducono una distorsione spettrale trascurabile ($< 2 \text{ dB}$);
- Procedura SelfEar, $H\text{-}S_{right}$ vs. $K\text{-}L_{left}$: la dif-

PRTF	H-L	H-S	K-L(right)	K-L(left)
H-L	0	1.79	6.92	5.25
H-S		0	7.35	4.64
K-L(right)			0	5.47
K-L(left)				0

Table 1: Distorsione spettrale tra PRTF di Fig. 6. Tutti i valori in dB.

ferenza tra l'acquisizione SelfEar delle PRTF e un sistema di misura tradizionale introducono l'errore in SD più basso tra le coppie analizzate (rimuovendo il confronto di controllo sui loudspeaker);⁵

La Figura 7 permette di confrontare visibilmente i risultati ottenuti con la procedura SelfEar negli angoli considerati (con e senza la compensazione di tipo *diffuse-field*) e le misurazioni CIPIC nello stesso range di angoli per il soggetto 165 (KEMAR). I dati sono stati interpolati per avere una transizione visivamente più fluida.

6. DISCUSSIONE GENERALE

È già noto da Christensen *et al.* [21] che sia la posizione del ricevitore, sia il suo spostamento dal punto di misura dell'HRTF ideale, cioè al centro dell'ingresso del canale uditorio, influenzano i pattern di direttività dell'HRTF per frequenze superiori a 3 – 4 kHz. Il nostro lavoro è in accordo con le loro misurazioni in quanto mostrano uno spostamento della frequenza del notch centrale anche di 2 kHz con una elevata variabilità del modulo nei vari riposizionamenti del microfono (vedere la deviazione standard di Fig. 5) con una differenza massima di 10 dB.

Gli spostamenti delle frequenze centrali di picchi e notch sono osservabili anche dalla Fig. 6 e sono principalmente dovuti a differenze topologiche tra il punto di osservazione, dipendente dalla posizione del microfono, e gli oggetti acustica formati dalla presenza o assenza del canale uditorio e da orecchie destra e sinistra differenti.

Avere a disposizione un ampio intervallo di elevazioni frontalì permette ad ogni sistema di misura di acquisire le caratteristiche rilevanti delle PRTF [19, 6]: ovvero le due principali risonanze (P1: modo omnidirezionale, P2: modo orizzontale) e i tre principali notch (N1-3 corrispondenti alle principali riflessioni della pinna). Le PRTF provenienti dal CIPIC KEMAR (vedere Fig. 7(c)) contengono tali caratteristiche; in particolare P1 ha una frequenza centrale di 4 kHz e P2 di 13 kHz; inoltre N1 si muove tra 6 e 9 kHz, N3 tra 11.5 e 14 kHz con un graduale incremento in elevazione; infine N2 che inizialmente si trova a 10 kHz, progressivamente scompare man mano che si raggiunge la direzione frontale.

L'applicazione SelfEar è in grado di acquisire P1 e N1 efficacemente, considerando sia il caso in cui le PRTF siano state compensate con la *diffuse field*, sia che non lo siano. Dal momento che l'ambiente ha un contributo non trascurabile, il confronto visivo tra Fig. 7(a) e (b) sottolinea l'importanza di poter estrarre con precisione le PRTF dalle BRIR. In particolare dalla Fig. 7(b) si possono identificare anche P2 e, anche se in minima parte, N2. Va quindi menzionato che P1 e N1-2 sono sufficienti per veicolare la medesima accuratezza in localizzazione verticale di una PRTF provvista di tutte le caratteristiche spettrali [22]. Tuttavia, dal punto di vista acustico, N3 è completamente assente, il che suggerisce la presenza di un interferenza introdotta dalle cuffie nella concha della pinna. Seguendo il

⁵ Va notato che questo confronto è stato effettuato su orecchie diverse per ragioni pratiche. Di conseguenza i valori di SD potrebbero avere una differenza ancor più inferiore.

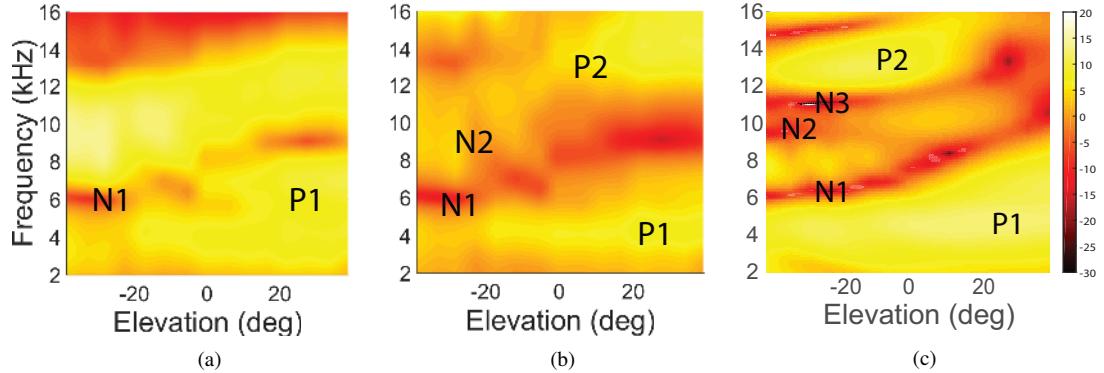


Figure 7: PRTF nel piano mediano. (a) acquisizione con SelfEar senza compensazione; (b) acquisizione con SelfEar compensata con diffuse-field; (c) CIPIC KEMAR, Subject 165 - con compensazione free-field. Vengono visivamente identificati i principali picchi (P1-2) e notch (N1-3), ove presenti.

modello risonanza-più-riflessione per le PRTF [6, 8] possiamo ipotizzare che l'assenza di riflessi nella concha sia dovuta alla presenza della cuffia; inoltre il volume della concha in questa condizione è drasticamente ridotto, producendo così variazioni nei modi di risonanza della struttura della pinna [7]. Inoltre, comparando i valori della distorsione spettrale tra $H - S$ vs. $K - L_{left}$ si ha una differenza di 4.64 dB che suggerisce una buona affidabilità paragonabile al metodo di personalizzazione in [8] (valori di SD tra 4 e 8 dB) e allo stato dell'arte delle simulazioni numeriche in [7] (valori di SD tra 2.5 e 5.5 dB).

Vale la pena notare che i parametri di notch e picchi, cioè la frequenza centrale, il guadagno e la banda, possono essere estratti dalle PRTF disponibili e dati in input a modelli di PRTF sintetiche e/o, seguendo un approccio di modellazione strutturale mista [3]], a procedura di selezione delle HRTF. Infine non vi è nulla che impedisca l'uso diretto di PRTF estratte da SelfEar per il rendering di audio binaurale.

7. CONCLUSIONI E LAVORI FUTURI

L'applicazione SelfEar permette l'acquisizione di HRIR a basso costo nel piano frontale mediano rilevando particolari informazioni spettrali della pinna dell'ascoltatore, ovvero le PRTF. L'applicazione sfrutta una struttura tecnologica di realtà audio aumentata per dispositivi mobile. Dopo essere adeguatamente compensate, le PRTF estratte sono paragonabili in termini di caratteristiche acustiche principali a quelle misurate in camera anechoica.

Il sistema proposto è stato testato seguendo un robusto setup di misurazione senza soggetti umani in una camera silente, ambiente acusticamente trattato. Perciò è tutt'ora da valutare una procedura robusta per ottenere le PRTF in ambiente domestico, stimando statisticamente l'influenza di eventi sonori rumorosi e casuali, nonché i movimenti del soggetto durante l'acquisizione. Per questo motivo la progettazione di algoritmi di elaborazione del segnale per la rilevazione di eventi acustici, l'eliminazione del rumore e il tracciamento dei movimenti è cruciale nella compen-

sazione del segnale e nelle fasi di pre- e post- elaborazione.

Una naturale evoluzione di questa applicazione considererà anche gli altri piani sagittali, ossia i piani verticali passanti per l'ascoltatore con $\text{azimuth} \neq 0$; verranno ottimizzate le procedure per ridurre il numero di posizioni delle sorgenti necessarie e per meglio controllarne la posizione e l'orientamento rispetto ai movimenti dell'utente; SelfEar implementerà algoritmi di computer vision capaci di tracciare la posa della testa dell'ascoltatore in tempo reale attraverso, ad esempio, una fotocamera integrata o sensori di profondità.

Oltre alla funzionalità di acquisizione di HRTF, vi sarà la possibilità di acquisire BRIR strutturate, cioè memorizzando separatamente RIR e HRIR per rappresentare coerentemente scenari di mAAR. I parametri estratti dalle RIR verranno dati in input a modelli computazionali di ambienti acustici dinamici; si potranno utilizzare la modellazione immagine-sorgente e il raybeam-tracing per le prime riflessioni, mentre per le riverberazioni successive si potrà adottare una gestione statistica [12]. Infine, sarà necessaria una valutazione psico-acustica con soggetti umani per confermare l'affidabilità delle tecnologie del progetto SelfEar con l'obiettivo di creare sorgenti sonore virtuali esternalizzate fornendo HRIR efficaci.

8. RINGRAZIAMENTI

Questo lavoro di ricerca è supportato dal progetto di ricerca *Personal Auditory Displays for Virtual Acoustics* (PADVA), Università di Padova, grant no. CPDA135702.

9. BIBLIOGRAFIA

- [1] S. Paul, "Binaural Recording Technology: A Historical Review and Possible Future Developments," *Acta Acustica united with Acustica*, vol. 95, pp. 767–788, Sept. 2009.
- [2] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1983.

- [3] M. Geronazzo, S. Spagnol, and F. Avanzini, "Mixed Structural Modeling of Head-Related Transfer Functions for Customized Binaural Audio Delivery," in *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, (Santorini, Greece), pp. 1–8, July 2013.
- [4] H. Gamper, M. R. P. Thomas, and I. J. Tashev, "Anthropometric parameterisation of a spherical scatterer ITD model with arbitrary ear angles," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 1–5, Oct. 2015.
- [5] M. Geronazzo, S. Spagnol, and F. Avanzini, "A Modular Framework for the Analysis and Synthesis of Head-Related Transfer Functions," in *Proc. 134th Conv. Audio Eng. Society*, (Rome, Italy), May 2013.
- [6] M. Geronazzo, S. Spagnol, and F. Avanzini, "Estimation and Modeling of Pinna-Related Transfer Functions," in *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, (Graz, Austria), pp. 431–438, Sept. 2010.
- [7] S. Prepelită, M. Geronazzo, F. Avanzini, and L. Savioja, "Influence of voxelization on finite difference time domain simulations of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 139, pp. 2489–2504, May 2016.
- [8] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the Relation between Pinna Reflection Patterns and Head-Related Transfer Function Features," *IEEE Trans. Audio, Speech, Language Processing*, vol. 21, pp. 508–519, Mar. 2013.
- [9] A. Ihlefeld and B. Shinn-Cunningham, "Disentangling the effects of spatial cues on selection and formation of auditory objects," *The Journal of the Acoustical Society of America*, vol. 124, no. 4, pp. 2224–2235, 2008.
- [10] W. G. Gardner and K. D. Martin, "HRTF Measurements of a KEMAR," *The Journal of the Acoustical Society of America*, vol. 97, pp. 3907–3908, June 1995.
- [11] V. Valimaki, A. Franck, J. Ramo, H. Gamper, and L. Savioja, "Assisted Listening Using a Headset: Enhancing audio perception in real, augmented, and virtual environments," *IEEE Signal Processing Magazine*, vol. 32, pp. 92–99, Mar. 2015.
- [12] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, pp. 708–730, Aug. 2015.
- [13] B. Boren, M. Geronazzo, F. Brinkmann, and E. Choueiri, "Coloration Metrics for Headphone Equalization," in *Proc. of the 21st Int. Conf. on Auditory Display (ICAD 2015)*, (Graz, Austria), pp. 29–34, July 2015.
- [14] W. O. Brimijoin, A. W. Boyd, and M. A. Akeroyd, "The Contribution of Head Movement to the Externalization and Internalization of Sounds," *PLoS ONE*, vol. 8, p. e83068, Dec. 2013.
- [15] J. S. Bradley and G. A. Soulodre, "Objective measures of listener envelopment," *The Journal of the Acoustical Society of America*, vol. 98, pp. 2590–2597, Nov. 1995.
- [16] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 868–878, 1989.
- [17] 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, Oct. 2001.
- [18] M. Geronazzo, S. Spagnol, A. Bedin, and F. Avanzini, "Enhancing Vertical Localization with Image-guided Selection of Non-individual Head-Related Transfer Functions," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, (Florence, Italy), pp. 4496–4500, May 2014.
- [19] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *The Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 159–168, 1990.
- [20] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, 1999.
- [21] F. Christensen, P. F. Hoffmann, and D. Hammersti, "Measuring Directional Characteristics of In-Ear Recording Devices," in *In Proc. Audio Engineering Society Convention 134*, Audio Engineering Society, May 2013.
- [22] 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," *Applied Acoustics*, vol. 68, no. 8, pp. 835 – 850, 2007.

UN PROTOTIPO DI PIANOFORTE DIGITALE CON FEEDBACK VIBROTATTILE

Federico Fontana

Università di Udine

Dept. of Mathematics and Computer Science
via delle Scienze 206, Udine 33100, Italy
federico.fontana@uniud.it

Federico Avanzini

Università di Padova

Dept. of Information Engineering
Via G. Gradenigo 6/b, Padova 35131, Italy
avanzini@dei.unipd.it

Hanna Järveläinen, Stefano Papetti

Zürcher Hochschule der Künste
Inst. for Computer Music and Sound Technology
Pfingstweidstrasse 96, Zurich 8031, Switzerland
name.surname@zhdk.ch

Giorgio Klauer, Lorenzo Malavolta

Conservatorio di Musica "Cesare Pollini"
Via Eremitani 18, Padova 35121, Italy
name.surname@gmail.com

ABSTRACT

Questo lavoro presenta i risultati di due esperimenti percettivi compiuti su un prototipo di pianoforte digitale aumentato. L'obiettivo di entrambi gli esperimenti era studiare la percezione da parte del pianista di feedback vibrotattile sulla tastiera. Nel primo esperimento i soggetti coinvolti dovevano suonare liberamente al variare del feedback, e fornire un giudizio sulla qualità percepita dello strumento in una griglia di cinque attributi: controllo dinamico, ricchezza, coinvolgimento, naturalezza, e preferenza complessiva. Nel secondo esperimento si è misurata l'accuratezza (in termini di tempo e di controllo dinamico) nell'esecuzione di una scala, al variare del feedback. I risultati mostrano una preferenza per le condizioni in cui è presente feedback vibrotattile, tuttavia per quanto riguarda l'accuratezza dell'esecuzione non sono state osservate differenze significative tra le diverse condizioni.

1. INTRODUZIONE

Lo studio della percezione vibrotattile nell'ambito delle interfacce musicali è un tema di ricerca molto attivo, a partire da strumenti tradizionali come pianoforte e violino [1, 2] fino al design di nuovi strumenti con applicazioni alla interazione musicale e alla didattica [3, 4]. Uno dei principali obiettivi di tale ricerca è quantificare l'effetto delle proprietà tattili di uno strumento sulla qualità percepita e la preferenza soggettiva, e in ultima analisi sull'esperienza e la performance dell'esecutore.

Per quanto riguarda il pianoforte, la simulazione delle proprietà tattili della tastiera ha finora riguardato soprattutto la risposta meccanica dei tasti [5, 6], anche sulla base di noti esperimenti che evidenziano una spiccata sensibilità dei pianisti rispetto alla meccanica della tastiera [7].

Più di recente la ricerca ha cominciato ad analizzare il ruolo percettivo delle vibrazioni dello strumento sulla tastiera [8], anche alla luce di alcuni nuovi prodotti industriali [9]. In particolare si è dimostrato [10] che tali vibrazioni sono chiaramente percepite attraverso le dita e si è ipotizzato che il pianista le avverte anche quando la loro ampiezza è al di sotto delle soglie soggettive tipicamente stimate in psicofisica considerando stimoli sinusoidali [11].

Tali risultati contraddicono solo apparentemente quelli di precedenti esperimenti [1], nei quali non si erano considerati i complessi effetti "additivi" (temporali, spaziali, spettrali) associati al contatto contemporaneo di più dita sulla tastiera né, cosa ancora più importante, il ruolo del tocco attivo rispetto a una semplice stimolazione passiva: in effetti, in tali condizioni le soglie di sensibilità vibrotattile si abbassano notevolmente rispetto a quelle note nella letteratura precedente [12].

Questo lavoro si concentra sull'effetto che diversi tipi di feedback vibrotattile producono sul pianista in termini di qualità percepita [13]. A tal fine, si è messo a punto un setup che utilizza un pianoforte digitale commerciale a cui sono stati applicati due attuatori in grado di creare segnali vibratori alla tastiera. Si sono poi creati due banchi di segnali di vibrazione da utilizzare con gli attuatori, contenenti rispettivamente vibrazioni reali misurate su un pianoforte acustico e vibrazioni sintetiche generate *ad hoc*.

Questo setup è servito per due esperimenti con pianisti, il primo sulla qualità percepita, il secondo sulla accuratezza della performance. I risultati mostrano che il setup aumentato con vibrazioni è stato in generale considerato migliore di quello privo di vibrazioni, con una leggera preferenza per vibrazioni di ampiezza enfatizzata rispetto a quella di un pianoforte acustico reale. D'altro canto non è stato apprezzato alcun effetto significativo delle vibrazioni sulla accuratezza (in termini di tempo e di controllo dinamico) della performance.

2. IL SETUP Sperimentale

Si è utilizzata la tastiera di un pianoforte digitale Viscount Galileo VP-91, separandola dal proprio case metallico e



Figura 1. Il setup sperimentale installato presso il Conservatorio di Padova.



Figura 2. Uno dei due trasduttori utilizzati per erogare vibrazioni alla tastiera.

innestandola su una spessa tavola di legno, come illustrato in Fig. 1. Sotto alla tavola sono stati montati due trasduttori Clark Synthesis TST239 Silver Tactile in corrispondenza dei registri grave e medio, come mostrato in Fig. 2, allo scopo di erogare vibrazioni nelle zone più rilevanti della tastiera [10]. Sul lato inferiore della tastiera equipaggiata in tal modo è stata posizionata della gommapiuma al fine di minimizzare l'interazione meccanica tra la tastiera stessa e lo stand su cui è stata posizionata.

I trasduttori sono stati collegati a un amplificatore Yamaha P2700 in configurazione dual mono, con un segnale di ingresso proveniente da una scheda RME Fireface 800 interfacciata via Firewire a un Apple MacBook Pro, quest'ultimo con il ruolo di generare feedback audio e tattile usando il software Reaper 4¹ configurato con i seguenti plugin: il sintetizzatore a modelli fisici Pianoteq 4.5 per la sintesi di suoni di pianoforte (erogato al pianista tramite cuffie), e il campionatore Kontakt 5² in serie con l'equalizzatore parametrico MeldaProduction MEEqualizer³ per

¹ www.reaper.fm
² www.native-instruments.com
³ www.meldaproduction.com

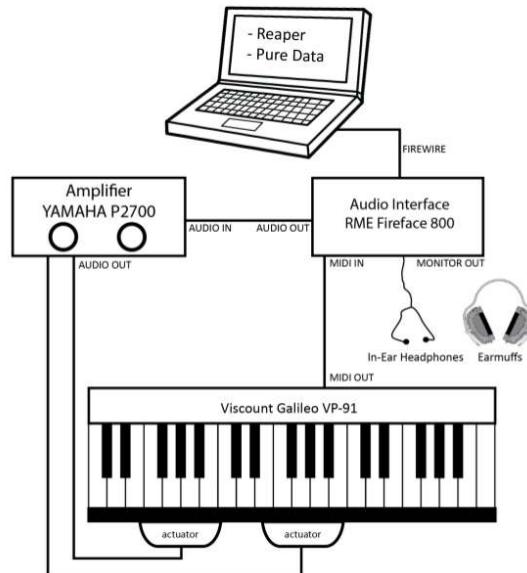


Figura 3. Schema del setup.

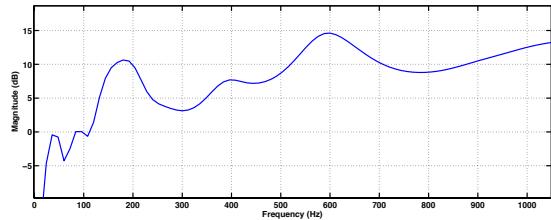


Figura 4. Curva di equalizzazione media nella catena trasmissione-trasduzione.

la resa delle vibrazioni.

La Fig. 3 mostra lo schema di questo setup. Lo stesso computer è stato utilizzato per condurre i test e raccogliere i dati sperimentali, tramite patch sviluppati *ad hoc* per l'ambiente real-time Pure Data,⁴ come descritto più in dettaglio nelle prossime sezioni.

2.1 Equalizzazione spettrale

Una limitazione del setup appena descritto è che la funzione di trasferimento tra i segnali di vibrazione in uscita dalla scheda audio e quelli rilevati sulla tastiera non è piatta, sia a causa della risposta del sistema meccanico tastiera-tavola di legno, sia perché i trasduttori hanno un notch pronunciato intorno ai 300 Hz. Si è quindi misurata la risposta in frequenza dell'intera catena trasmissione-trasduzione in corrispondenza di tutti i tasti LA e poi stimata una risposta in ampiezza media, la cui inversa è stata utilizzata come filtro di equalizzazione: si veda la Fig. 4, notando in particolare che il notch a 300 Hz viene compensato assieme alle risonanze e anti-risonanze del sistema meccanico.

Questa risposta è stata ulteriormente approssimata all'interno del plugin MEEqualizer utilizzando un filtro shel-

⁴ puredata.info

ving con una pendenza di 18 dB nel range [100, 600] Hz, in serie con un filtro risonante del secondo ordine per approssimare il picco a 180 Hz.

2.2 Segnali di vibrazione

Il banco di segnali di vibrazioni reali è stato ottenuto da misure effettuate su un pianoforte acustico Disklavier Yamaha DC3 M4 usando un accelerometro piezoelettrico Wilcoxon Research 736 con Intelligent Transmitter iT100M connesso alla scheda audio. L'accelerometro è stato assicurato a ciascun tasto con del Pongo.⁵ Sfruttando il controllo MIDI del Disklavier si sono potute misurare le vibrazioni generate all'abbassarsi di ciascun tasto con diversi valori di velocity (12, 23, 34, 45, 56, 67, 78, 89, 100, 111).

Il banco di vibrazioni sintetiche è stato costruito in modo da mantenere gli stessi inviluppi di ampiezza dei corrispondenti segnali reali, cambiandone solo il contenuto spettrale. Più precisamente, per ogni nota e ogni valore di velocity si è (1) generato rumore bianco limitato nella banda tattile [11] [20, 500] Hz, (2) passato questo rumore attraverso un filtro risonante del secondo ordine centrato sulla frequenza fondamentale della nota, (3) modulato il segnale risultante con l'inviluppo di ampiezza del corrispondente campione reale (inviluppo stimato dalla “energy decay curve” del campione tramite integrale di Schroeder [14]). Il campione sintetico così ottenuto è stato equalizzato in energia al corrispondente campione reale.

Questi due banchi di vibrazioni reali e sintetiche sono stati forniti al campionatore Kontakt 5, a cui è stata demandata l'interpolazione tra campioni in funzione del valore di velocity MIDI.

2.3 Calibrazione della dinamica

I tasti del Disklavier e del pianoforte digitale Galileo hanno risposte dinamiche molto diverse a causa delle loro diverse meccaniche e masse. Poiché i pianisti adattano il proprio tocco in funzione di tali differenze, è stato necessario calibrare la tastiera digitale secondo una procedura soggettiva per equalizzare la sua dinamica a quella del Disklavier e quindi assicurare che, quando un pianista suona la tastiera digitale a una certa dinamica, viene triggerato il corrispondente campione di vibrazione registrato sul Disklavier.

A tal fine, con l'ausilio di un pianista esperto si è utilizzata la routine di calibrazione soggettiva di Pianoteq, prima sul Disklavier e poi sulla tastiera digitale. Avendo ottenuto – come previsto – due mapping di velocity molto diversi, si è poi utilizzato un filtro MIDI in Reaper per mappare ogni punto della mappa di velocity sulla tastiera digitale nel corrispondente punto sulla mappa del Disklavier. La configurazione così ottenuta è stata successivamente testata da due ulteriori pianisti per un'ulteriore conferma qualitativa della sua accettabilità e affidabilità.

2.4 Calibrazione della loudness

L'ultimo passo di calibrazione del setup è consistito nell'equalizzare la loudness del suono di pianoforte sintetico

Pianoteq, erogato al pianista tramite cuffie, con quella del Disklavier.

A tal fine, si sono registrate le note LA del Disklavier con diversi valori di velocity, usando una testa artificiale KEMAR⁶ posizionata in corrispondenza del pianista [10]. Si sono poi effettuate ulteriori registrazioni facendo indossare alla testa KEMAR le cuffie del setup sperimentale, ovvero degli auricolari Sennheiser CX 300-II a loro volta ricoperti da cuffie passive anti-rumore 3M Peltor X5, e generando le stesse note LA e le stesse velocity con Pianoteq. Infine, comparando questi due set di registrazioni, la loudness del pianoforte sintetico è stata adattata a quella del Disklavier utilizzando la funzionalità di “volume mapping” in Pianoteq, che permette di regolare in maniera indipendente il volume di ciascun tasto.

3. VALIDAZIONE SPERIMENTALE

I due esperimenti descritti di seguito hanno coinvolto 11 partecipanti, di cui 5 donne e 6 uomini, con età media di 26 anni ed esperienza media di 8 anni nello studio del pianoforte a livello di Conservatorio. Due partecipanti erano pianisti jazz, i rimanenti pianisti classici. Tutti hanno firmato un modulo di consenso informato. La durata media di una sessione comprendente entrambi gli esperimenti è stata di circa un'ora.

I partecipanti hanno svolto la sessione sperimentale usando il setup descritto in precedenza (Fig. 3) e indossando auricolari e cuffie anti-rumore allo stesso modo della testa KEMAR durante la procedura di equalizzazione di loudness descritta sopra. Questo è servito a evitare che i partecipanti sentissero il feedback sonoro generato dai trasduttori come effetto collaterale della loro vibrazione, e che percepissero invece solo gli stimoli audio-tattili generati in tempo reale in risposta al loro tocco sulla tastiera, ed erogati rispettivamente in cuffia e sulla tastiera stessa.

3.1 Esperimento 1: qualità percepita

In questo esperimento sono state prese in considerazione quattro diverse condizioni di vibrazione, tutte associate al medesimo feedback uditorio (generato tramite Pianoteq secondo quanto già descritto):

- A: nessuna vibrazione;
- B: vibrazioni reali;
- C: vibrazioni reali, con un boost di 9 dB;
- D: vibrazioni sintetiche.

Ogni partecipante doveva suonare liberamente la tastiera digitale, senza alcuna restrizione di dinamica o range, e valutare la propria esperienza su cinque categorie di giudizio: “controllo dinamico”, “ricchezza”, “coinvolgimento”, “naturalezza”, “preferenza globale”. Queste categorie sono state scelte sulla base di recenti ricerche sulla valutazione della qualità di violini [2].

Ogni partecipante poteva commutare liberamente tra due condizioni, α e β : α corrispondeva sempre a quella senza vibrazioni (A), mentre β era una delle tre condizioni con

⁵ <http://www.fila.it/en/pongo/history/>

⁶ Knowles Electronic Manikin for Acoustic Research, un manichino comunemente utilizzato per registrazioni e misure binaurali.

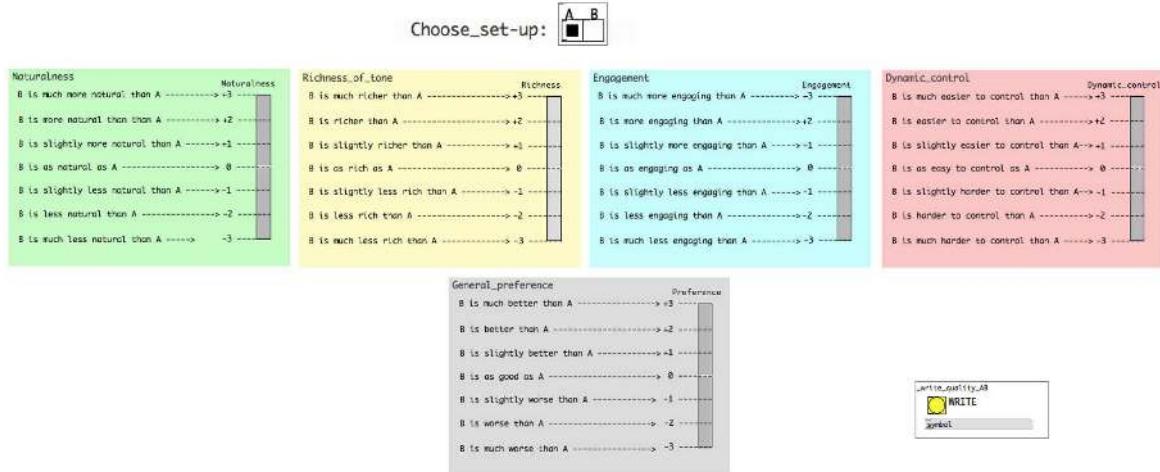


Figura 5. Interfaccia grafica usata dai partecipanti per commutare tra la condizioni (A,B) e per indicare i punteggi sulle cinque categorie di giudizio.

vibrazioni (B, C, D), per un totale quindi di tre possibili coppie (α, β). Tuttavia i partecipanti non erano informati di quali fossero le differenze tra le due condizioni, e in particolare che il feedback uditorio rimaneva sempre invariato.

Il tempo di valutazione per una coppia di condizioni (α, β) era di 10 minuti, durante i quali il partecipante poteva in ogni momento esprimere o modificare i propri punteggi sulle cinque categorie di giudizio, e commutare tra le due condizioni, usando un’interfaccia grafica realizzata in Pure Data e visibile in Fig. 5. Al termine dei 10 minuti si passava alla coppia successiva. Le coppie e condizioni all’interno di esse erano randomizzate tra i partecipanti.

Per i giudizi si è usata una scala continua di tipo “Comparison Category Rating” (CCR) tra -3 e $+3$, tipica per test di qualità soggettiva nelle tecnologie di comunicazione (ITU-T P.800). I partecipanti muovevano uno slider lungo la scala continua per esprimere la propria opinione. La scala riportava le seguenti indicazioni:

- +3: “ β molto meglio di α ”
- +2: “ β meglio di α ”
- +1: “ β leggermente meglio di α ”
- 0: “ β uguale a α ”
- 1: “ β leggermente peggio di α ”
- 2: “ β peggio di α ”
- 3: “ β molto peggio di α ”

3.2 Esperimento 2: accuratezza della performance

L’unica differenza tra il setup di questo esperimento e quello del precedente è la presenza in cuffia di un suono di metronomo a 120 BPM. Inoltre in questo esperimento si sono utilizzate solo le condizioni A e B descritte sopra.

Ogni partecipante doveva suonare una scala ascendente e discendente di RE maggiore, al tempo scandito dal metronomo, a tre diverse dinamiche costanti (*pp, mf, ff*). La scala andava eseguita sulle tre ottave più gravi (quindi con la sola mano sinistra) al fine di massimizzare il feedback tattile. Ogni dinamica veniva ripetuta tre volte, per entram-

be le condizioni A e B, arrivando a un totale di 18 ripetizioni randomizzate. A ogni ripetizione venivano registrati con Pure Data i corrispondenti dati MIDI (note ON, lunghezza della nota, velocity), da utilizzare successivamente per analizzare la performance.

4. RISULTATI Sperimentali

4.1 Qualità percepita

Per quantificare la consistenza intersoggettiva sono stati calcolati i valori di correlazione di concordanza di Lin ρ_c tra tutte le coppie di partecipanti [15]. I valori medi ottenuti sono $\rho_c = 0.018$ per “preferenza globale”, $\rho_c = 0.006$ per “controllo dinamico”, $\rho_c = -0.04$ per “ricchezza”, $\rho_c = -0.02$ per “coinvolgimento”, e $\rho_c = -0.04$ per “naturalezza”. Per tutte le categorie, alcuni soggetti erano in completo accordo o disaccordo, e a causa di questa alta variabilità i valori di ρ_c non si discostano significativamente da 0 per nessuna categoria ($t(54) < 0.77, p > 0.05$). Questi bassi valori di concordanza indicano un alto grado di disaccordo tra i partecipanti.

Le risposte sono correlate positivamente tra tutte le categorie di giudizio. La correlazione più debole è quella tra “ricchezza” e “controllo dinamico” (correlazione di Spearman $\rho_s = 0.18$), mentre quella più forte è quella tra “preferenza globale” e “coinvolgimento” ($\rho_s = 0.75$). Le rimanenti correlazioni tra “preferenza globale” e le altre categorie sono le seguenti: $\rho_s = 0.39$ per “controllo dinamico”, $\rho_s = 0.72$ per “ricchezza”, e $\rho_s = 0.57$ per “naturalezza”.

La Fig. 6 mostra i risultati, mentre i giudizi medi per ogni categoria e condizione di vibrazione sono riportati in Tabella 1. In media tutte le condizioni con vibrazione vengono preferite a quella senza vibrazione, con l'unica eccezione della condizione D per la categoria “naturalezza”. La medesima categoria ha punteggi solo marginalmente positivi per le condizioni B e C. Le preferenze più evidenti sono quelle per “range dinamico” e “coinvolgimento”, mentre quelle per “preferenza globale” e “ricchezza” sono

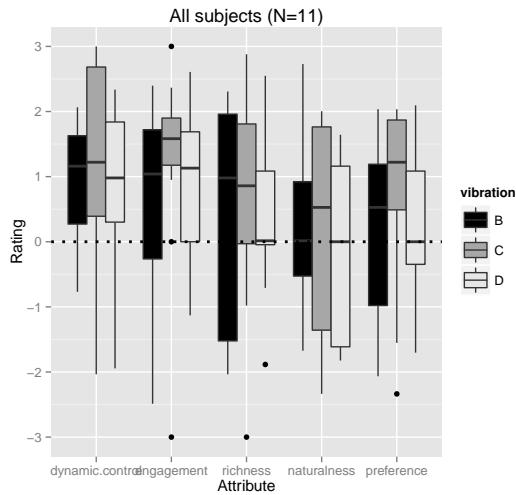


Figura 6. Esperimento 1: box-plot con mediane e quartili per ogni categoria e per ogni condizione con vibrazione.

Condiz.	Contr.	Ricch.	Coinv.	Natur.	Globale
B	0.92	0.30	0.50	0.26	0.24
C	1.28	0.67	1.21	0.17	0.81
D	0.87	0.42	1.00	-0.23	0.29

Tabella 1. Esperimento 1: giudizi medi su tutti i soggetti per ogni categoria e per ogni condizione con vibrazione.

più basse e molto simili tra loro. La condizione C è quella più spesso preferita tra tutte quelle con vibrazione, e riceve i punteggi più alti in quattro categorie su cinque. Si noti però che B viene considerata la condizione più naturale, anche se ha giudizi più bassi su tutte le altre categorie.

Poiché per i dati non sussiste la condizione di normalità ai fini dell'analisi della varianza, si è utilizzato un test non parametrico di Friedman sulle differenze tra misure ripetute, per analizzare i giudizi di preferenza. Il test fornisce un valore Chi-quadrato di 21.9 ($p < 0.05$), suggerendo una differenza significativa tra le condizioni con vibrazioni. D'altro canto, il test di Wilcoxon sull'ipotesi che la mediana sia positiva non ha fornito significatività in nessuna delle condizioni (B: $V = 37.5, p > 0.05$; C: $V = 41, p > 0.05$; D: $V = 28, p > 0.05$).

L'alta variabilità intersoggettiva discussa all'inizio della sezione è associata a una alta eterogeneità nei dati. Utilizzando un algoritmo K-means di clusterizzazione per dividere *a posteriori* i partecipanti in due categorie rispetto ai giudizi di "preferenza globale", otto partecipanti sono stati classificati in un gruppo "positivo" e i restanti tre in un gruppo "negativo". La Fig. 7 mostra i risultati per questi due gruppi. Si osservi l'evidente differenza di giudizi: la media per la condizione preferita C è quasi +2 nel gruppo positivo, e circa -1.5 nel gruppo negativo. La mediana nel gruppo positivo è > 0 in tutti i casi tranne uno ("naturalità", D), mentre in quello negativo è positiva solamente in un caso ("controllo dinamico", B).

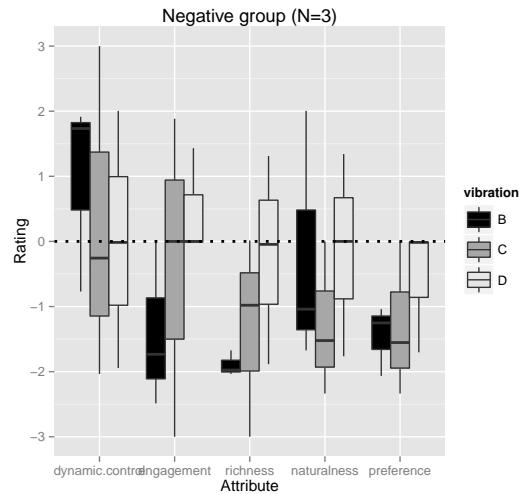
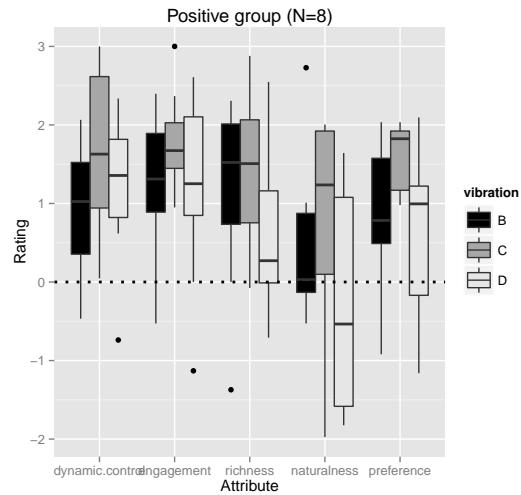


Figura 7. Esperimento 1: gli stessi box-plot di Fig. 6, divisi per i gruppi "positivo" e "negativo".

4.2 Accuratezza della performance

L'ipotesi sperimentale è che, se la presenza di vibrazioni sulla tastiera influisce sulla performance dei partecipanti in termini di tempo e dinamica, allora tali effetti devono essere visibili nei valori medi e nelle varianze delle velocità e degli intervalli inter-onset (inter-onset-intervals, IOI) calcolati dai dati MIDI.

I valori medi di velocity per ogni partecipante sono stati calcolati come le medie tra le tre ripetizioni della scala all'interno della stessa condizione e della stessa dinamica. La Fig. 8 mostra i risultati. Per la dinamica *ff* i valori sono leggermente più alti in presenza di vibrazioni, mentre per la dinamica *mf* la situazione si capovolge. Tuttavia, un'analisi tramite ANOVA a misure ripetute mostra che non ci sono differenze significative tra condizioni di vibrazione ($F(1, 2826) = 2.27, p > 0.05$), e che non c'è interazione tra condizioni di vibrazione e dinamica ($F(2, 2826) = 0.83, p > 0.05$). La non significatività permane anche se si restringe l'analisi all'otta-

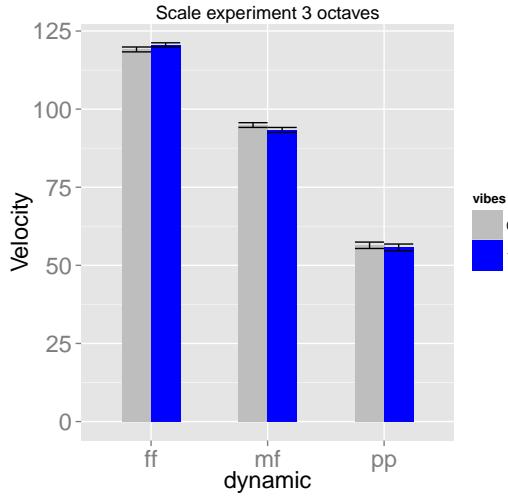


Figura 8. Esperimento 2: valori medi di velocity e intervalli di confidenza al 95% [16].

va più grave, dove la percezione delle vibrazioni è massima. Allo stesso modo, non vi è una differenza significativa delle deviazioni standard (standard deviations, sd_A e sd_B) tra le due condizioni: gli intervalli di confidenza al 95% stimati da t-test a campioni dipendenti ($df=10$) includono $\mu(sd_B) - \mu(sd_A) = 0$ per tutte le dinamiche.

Gli IOI risultano molto stabili tra le due condizioni di vibrazione. Si osserva una maggiore variabilità per la dinamica *pp*, ma senza alcun effetto della vibrazione (si veda la Fig. 9). Allo stesso modo le durate delle note sono stabili, suggerendo che non vi siano differenze significative tra le due condizioni in termini di articolazione.

5. DISCUSSIONE E CONCLUSIONI

I risultati sperimentali mostrano che la presenza di vibrazioni sulla tastiera aumenta la qualità percepita di un pianoforte digitale. Inoltre, una resa amplificata (+9 dB) delle vibrazioni reali sembra essere complessivamente preferita a un'altra più fedele alla realtà, sebbene quest'ultima venga percepita come più naturale. Un secondo risultato interessante è che l'unico caso in cui la condizione senza vibrazione viene preferita a un'altra si ha per la categoria “naturalezza” e per le vibrazioni sintetiche. Questo suggerisce che i pianisti siano sensibili alla coerenza inter-sensoriale tra il feedback uditorio e quello tattile.

Le categorie di giudizio maggiormente correlate alla “preferenza globale” sono “coinvolgimento” ($\rho_s = 0.75$) e “ricchezza” ($\rho_s = 0.72$). Questo risultato è comparabile a quanto mostrato in un recente studio sulla valutazione della qualità di violini, nel quale si è trovata una forte correlazione tra le categorie “ricchezza” e “preferenza globale” [2].

La bassa concordanza tra i partecipanti suggerisce che la consistenza intra- e inter-soggettiva sia un elemento di fondamentale importanza da considerare in esperimenti di valutazione di qualità di strumenti musicali. In questo studio non è stato possibile valutare il grado di consisten-

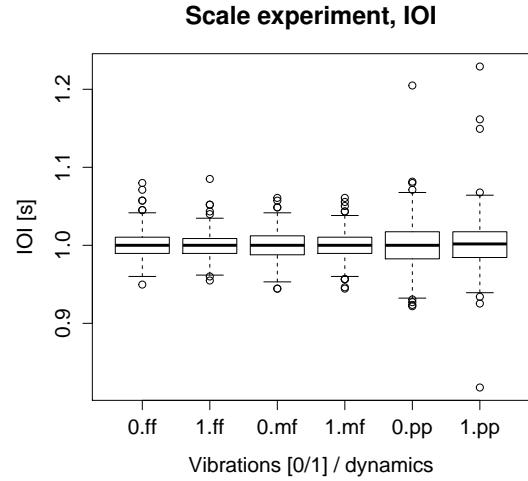


Figura 9. Esperimento 2: box-plot con mediane e quartili degli IOI.

za intra-soggettiva, dal momento che ogni partecipante ha fornito un unico giudizio per ogni condizione e categoria di giudizio. Questo tipo di valutazione verrà incluso in esperimenti futuri.

Va tuttavia sottolineato che l'eterogeneità nei dati sperimentali è presente in egual modo in tutte le categorie di giudizio e in tutte le condizioni: risulta quindi difficile credere che tale eterogeneità provenga unicamente da inconsistenza nelle risposte. All'incirca due terzi dei partecipanti ha espresso una chiara preferenza per le condizioni con vibrazioni (e, tra queste, un apprezzamento lievemente più basso per le vibrazioni sintetiche), mentre il rimanente terzo ha espresso un'opinione opposta.

A questo proposito, è interessante notare che i due pianisti jazz, probabilmente più abituati dei pianisti classici a suonare su pianoforti digitali, rientrano entrambi nel gruppo “negativo”. Questa circostanza potrebbe suggerire che essi percepiscono una tastiera vibrante come meno piacevole di una neutra coerentemente con la propria esperienza pregressa. Alla luce di queste considerazioni, futuri esperimenti dovranno *a priori* considerare pianisti classici e non come appartenenti a gruppi sperimentali distinti, e dovranno inoltre includere dettagliate interviste post-sperimentali per raccogliere l'esperienza pregressa e le osservazioni dei partecipanti.

Il secondo esperimento non ha evidenziato differenze nella performance dei partecipanti in funzione della presenza o meno di vibrazioni. Probabilmente il task scelto (tre ottave di scala di RE maggiore a tempo relativamente basso) era troppo facile da svolgere accuratamente in qualsiasi condizione. E d'altra parte alcuni recenti studi mostrano che i pianisti usano informazione vibrotattile come mezzo per regolare l'accuratezza temporale [17, 18], anche se il ruolo delle vibrazioni sulla tastiera rimane non chiaro. Altri studi mostrano che la presenza di feedback vibrotattile migliora l'accuratezza in task di pressione con le dita, in termini di forza erogata [19, 20].

Il secondo esperimento non ha quindi chiarito se la presenza di vibrazioni sul tasto premuto dal pianista possa aiutare la pianificazione motoria dell'azione sul tasto successivo. Tuttavia il fatto che la categoria "controllo dinamico" abbia ricevuto giudizi molto positivi per tutte le condizioni con vibrazioni lascia supporre che la presenza di feedback tattile non venga apprezzata solo per un effetto di maggiore coinvolgimento, ma che venga anche percepita come di aiuto nella performance.

Futuri esperimenti dovranno utilizzare task performativi diversi, in cui l'informazione fornita dal feedback tattile sia più significativa che non in quello considerato qui. Ad esempio, un task più appropriato potrebbe essere una serie di ripetizioni di accordi di durata sufficientemente lunga, in cui il pianista debba mantenere una dinamica costante: in questo caso le vibrazioni e i loro inviluppi temporali sarebbero percepiti più chiaramente e potrebbero aiutare maggiormente il controllo e la pianificazione motoria.

Ringraziamenti

Questa ricerca è stata parzialmente supportata dal progetto AHMI (Audio-Haptic modalities in Musical Interfaces), finanziato dalla Swiss National Science Foundation (SNSF).

6. REFERENCES

- [1] A. Askenfelt and E. V. Jansson, "On vibration and finger touch in stringed instrument playing," *Music Perception*, vol. 9, no. 3, pp. 311–350, 1992.
- [2] C. Saitis, B. L. Giordano, C. Fritz, and G. P. Scavone, "Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 4002–4012, 2012.
- [3] M. Marshall and M. Wanderley, "Examining the effects of embedded vibrotactile feedback on the feel of a digital musical instrument," in *Proc. Int. Conf. on New Interfaces for Musical Expression (NIME)*, (Oslo, Norway), pp. 399–404, May 30 - June 1 2011.
- [4] M. Giordano and M. M. Wanderley, "Perceptual and technological issues in the design of vibrotactile-augmented interfaces for music technology and media," in *Haptic and Audio Interaction Design* (I. Oakley and S. Brewster, eds.), vol. 7989 of *Lecture Notes in Computer Science*, pp. 89–98, Springer Berlin Heidelberg, 2013.
- [5] C. Cadoz, L. Lisowski, and J.-L. Florens, "A Modular Feedback Keyboard Design," *Comput. Music J.*, vol. 14, no. 2, pp. 47–51, 1990.
- [6] R. Oboe and G. De Poli, "A Multi-Instrument, Force-Feedback Keyboard," *Comput. Music J.*, vol. 30, pp. 38–52, Sept. 2006.
- [7] A. Galemba and A. Askenfelt, "Quality assessment of musical instruments - Effects of multimodality," in *Proc. 5th Triennial Conf. of the European Society for the Cognitive Sciences of Music (ESCOM5)*, (Hannover, Germany), Sep. 8-13 2003.
- [8] F. Fontana, S. Papetti, M. Civolani, V. dal Bello, and B. Bank, "An exploration on the influence of vibrotactile cues during digital piano playing," in *Proc. Int. Conf. on Sound Music Computing (SMC2011)*, (Padua, Italy), pp. 273–278, 2011.
- [9] E. Guizzo, "Keyboard maestro," *IEEE Spectrum*, vol. 47, pp. 32–33, Feb. 2010.
- [10] F. Fontana, F. Avanzini, H. Järveläinen, S. Papetti, F. Zanini, and V. Zanini, "Perception of interactive vibrotactile cues on the acoustic grand and upright piano," in *Proc. Joint ICMC/SMC Conf.*, 2014.
- [11] T. Verrillo, "Vibrotactile thresholds measured at the finger," *Perception and Psychophysics*, vol. 9, no. 4, pp. 329–330, 1971.
- [12] S. Papetti, H. Järveläinen, and G.-M. Schmid, "Vibrotactile sensitivity in active finger pressing," in *Proc. IEEE World Haptics Conf. (WHC2015)*, (Evanston, IL), pp. 457–462, July 2015.
- [13] F. Fontana, F. Avanzini, H. Järveläinen, S. Papetti, G. Klauer, and L. Malavolta, "Rendering and subjective evaluation of real vs. synthetic vibrotactile cues on a digital piano keyboard," in *Proc. Int. Conf. Sound and Music Computing (SMC2015)*, (Maynooth), pp. 161–167, July 2015.
- [14] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 1187–1188, June 1965.
- [15] L. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, vol. 45, pp. 255–268, 1989.
- [16] R. Morey, "Confidence intervals from normalized data: A correction to Cousineau (2005)," *Tutorial in Quantitative Methods for Psychology*, vol. 4, no. 2, pp. 61–64, 2008.
- [17] W. Goebel and C. Palmer, "Tactile feedback and timing accuracy in piano performance," *Exp. Brain. Res.*, vol. 186, pp. 471–479, 2008.
- [18] W. Goebel and C. Palmer, "Finger motion in piano performance: Touch and tempo," in *International symposium for performance science*, 2009.
- [19] H. Järveläinen, S. Papetti, S. Schiesser, and T. Grosshauser, "Audio-tactile feedback in musical gesture primitives: Finger pressing," in *Proceedings of the Sound and Music Computing Conference (SMC2013)*, (Stockholm), 2013.
- [20] T. Ahmaniemi, "Effect of dynamic vibrotactile feedback on the control of isometric finger force," *IEEE Trans. on Haptics*, 2012.

Posters/Demo

INTERPOLAZIONI EASING: IMPLEMENTAZIONE E APPLICAZIONI AUDIO

Renato Messina

ISSM Vincenzo Bellini, Catania
messinarenato@gmail.com

SOMMARIO

L'estensione delle tecniche della computer animation ad ambiti seriali affini, come quello della computer grafica o del web design, ha determinato un'ampia diffusione delle funzioni easing di interpolazione. La stilizzazione prodotta da queste curve di transizione, definite sulla base di modelli cinematici caratteristici, suggerisce in ambito audio strutture simboliche ed iconiche efficaci nella progettazione di segnali acustici e nella comunicazione di significati espressivi correlabili a schemi isomorfici gestuali. Nell'articolo viene descritta l'implementazione Java di un external multipiattaforma per il proxy mxj [1] di Max¹ e PD², dotato di una libreria di 30 curve e rivolto allo sviluppo di architetture cross-modali e di display uditorio.

1. INTRODUZIONE

Nella computer animation le interpolazioni easing forniscono un'interfaccia di programmazione ad alto livello con cui assegnare una curva di accelerazione ad un oggetto in movimento. Le variazioni di velocità che in tal modo ne definiscono lo spostamento rendono le transizioni più fluide e realistiche attraverso la riproduzione di tipologie cinematiche proprie del moto dei corpi puntiformi o della biomeccanica. Per estensione, tali interpolazioni, chiamate anche tweening, possono applicarsi, oltre che alla posizione (motion tweening), anche a parametri descrittivi dell'oggetto (shape tweening), come dimensione e forma, indirettamente correlabili ad altre grandezze cinematiche fondamentali, quali massa e peso.

In via generale, se le easing esprimono una relazione in cui il dominio della funzione è il tempo e il codominio è l'insieme dei valori assunti dall'argomento che si vuole animare, la presenza di una linea temporale di accelerazione comune garantisce potenzialmente un'integrazione coerente automatica ad un numero indefinito di codomini

Copyright: © 2016 Renato Messina. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ www.cycling74.com.

² <https://puredata.info>.

³ L'applicazione delle curve all'articolazione melodica evidenzia una correlazione tra sistemi discreti e sistemi continui in grado di fornire

rappresentativi del set di avars (animation variables) impiegato. Operazioni di inversione e rovesciamento applicate alle rampe di singoli parametri riescono, così, a creare effetti complessi di interazione dinamica tra le proprietà strutturali interne, la traiettoria e la posizione dell'oggetto animato. Un'analogia ricchezza di relazioni, in ambito musicale, si può evidenziare nel raggruppamento di vincoli metrici, agogici e dinamici presente nelle tradizionali tecniche del 'legato' e nelle prassi esecutive di appoggiate, glissandi e portamenti³.

2. NOMENCLATURA E CLASSI

La nomenclatura di una classe easing individua attraverso un prefisso e una radice i due elementi formali che la caratterizzano. Il prefisso specifica una delle tre versioni, *In*, *Out*, *InOut*, del profilo di accelerazione esponenziale, logaritmico o ibrido della funzione richiamata. La caratteristica delle prime due versioni, *In* e *Out*, è l'angolo di discontinuità presente nella fase conclusiva, o iniziale, della curva, vedasi figura 1. I contenuti espressivi e mimici che tale peculiarità riesce a veicolare sono valutabili, in termini di efficacia, in rapporto al contesto di applicazione. Ad esempio, da un punto di vista cinematico, gli involucri logaritmici prodotti dalla versione *Out* possono descrivere forze di tipo esplosivo, con attacco percussivo, mentre gli involucri esponenziali della versione *In*, sono più adatti a descrivere forze reattive, con un'accelerazione progressiva che culmina in un arresto immediato, come avviene nell'esecuzione di salti o di lanci. La versione *In*, d'altro canto, se considerata da un punto di vista pliometrico, può risultare innaturale per l'assenza di vibrazioni residue e per il suo coefficiente di restituzione nullo; e, viceversa, può essere consistente nel caso di rappresentazioni metaforeiche che simulino sistemi elettronici a controllo numerico. Per il senso di precisione che comunica nella sua fase di arresto, si potrebbe anche dire che la versione *In* esprima in modo paradigmatico l'effetto di applicazione di uno schema di controllo feedforward (anticipativo) delle forze necessarie al sistema per l'esecuzione di un determinato piano motorio [2]. L'ultima delle tre versioni, identificata dal prefisso *InOut*, specifica un andamento *In* nella prima

un'interessante descrizione della linea melodica come successione di curve strutturali di transizione ancor prima che come sequenza di intervalli non interpolati.

metà della curva è *Out* nella seconda metà, dunque con attacco e release morbidi. La cosiddetta funzione S-curve, una cubica che implementa questa forma di andamento, è tra le più frequentemente usate, sia in contesti grafici che meccatronici, proprio per la fluidità che conferisce al movimento. È interessante rilevare come la diffusione della versione *InOut* trovi un riscontro biomeccanico in diversi modelli in cui il movimento è scomposto nelle sue due fasi di accelerazione e decelerazione. Ad esempio, il modello a due componenti di Woodworth [3], o il profilo di velocità a campana [4], o il modello 'minimum-jerk', secondo cui uno degli obiettivi del sistema nervoso è proprio la massimizzazione della 'dolcezza' (smoothness) del movimento [5].

Se il prefisso del nome della classe indica l'andamento esponenziale o logaritmico dell'accelerazione, la radice ne specifica la funzione. Le librerie ne implementano circa 30, riassumibili in 3 categorie.

- 1) Esponenziali, dove il nome della classe indica l'esponente assegnato per calcolare la pendenza della funzione.
- 2) Sinusoidale e Circolare.
- 3) Back, Elastic e Bounce, le quali, essendo funzioni non monotone, posseggono un andamento più articolato e creano effetti di anticipazione o rimbalzo con una verosimiglianza sufficientemente realistica. Vedasi in Tabella 1 il riassunto delle classi.

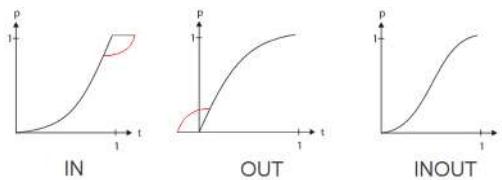


Figura 1. Angolo di discontinuità delle versioni *In* e *Out*. In ascissa è rappresentato il tempo, in ordinata la posizione del parametro associato.

Versioni			Funzione
<i>In</i>	<i>Out</i>	<i>InOut</i>	
<i>In</i>	<i>Out</i>	<i>InOut</i>	Sine
<i>In</i>	<i>Out</i>	<i>InOut</i>	Quadratic
<i>In</i>	<i>Out</i>	<i>InOut</i>	Cubic
<i>In</i>	<i>Out</i>	<i>InOut</i>	Quartic
<i>In</i>	<i>Out</i>	<i>InOut</i>	Quintic
<i>In</i>	<i>Out</i>	<i>InOut</i>	Exponential
<i>In</i>	<i>Out</i>	<i>InOut</i>	Circular
<i>In</i>	<i>Out</i>	<i>InOut</i>	Back
<i>In</i>	<i>Out</i>	<i>InOut</i>	Elastic
<i>In</i>	<i>Out</i>	<i>InOut</i>	Bounce

Tabella 1. Nomenclatura delle classi easing elencate in ordine di pendenza e complessità delle funzioni.

3. FUNZIONI

3.1 Esponenziali

Le curve esponenziali sono rese con le relative funzioni distinte in base al numero dell'esponente:

$$\text{OutLinear} \quad f(t) = t^1 \quad (1)$$

$$\text{OutQuadratic} \quad f(t) = t^2 \quad (2)$$

$$\text{OutCubic} \quad f(t) = t^3 \quad (3)$$

$$\text{OutQuartic} \quad f(t) = t^4 \quad (4)$$

$$\text{OutQuintic} \quad f(t) = t^5 \quad (5)$$

$$\text{OutExponential} \quad f(t) = t^{10} \quad (6)$$

Le versioni *In* sono ottenute tramite flip orizzontale e verticale delle versioni *Out*. Ad esempio la funzione *InCubic* assumerà la forma: $f(t) = (t - 1)^3 + 1$.

Le versioni *InOut*, come già detto, sono ricavate attraverso un metodo che applica la versione *In* alla prima metà della rampa e quella *Out* alla seconda. Si veda ad esempio una codifica della curva *InOutCubic*, in Java:

```
public double easeInOutCubic (double value) {
    double v = 2.0d * value;
    if (v < 1.0d) {
        return 0.5d * Math.pow(v, 3.0d);
    }
    v -= 2.0d;
    return 0.5d * (Math.pow(v, 3.0d) + 2.0d);
}
```

Oppure, in Max (figura 2), la stessa curva, generata all'interno di un buffer di 512 campioni, con frequenza di campionamento 44.1 kHz.

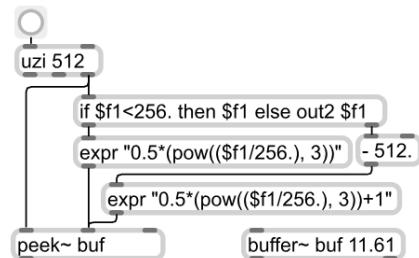


Figura 2. Curva *InOutCubic* (S-curve) in Max.

Nelle funzioni esponenziali, l'incremento dell'ordine di potenza produce un effetto di intensificazione della drammaticità, o reattività, della rampa, in contrapposizione ad un effetto di moto inerziale tipico delle rampe con esponenti più bassi.

3.2 Sinusoidale e circolare

La curva sinusoidale è calcolata con $\pi/2$ radianti di una cosinusoide (*In*) o di una sinusoide (*Out*):

$$\text{SineOut} \quad f(t) = \sin(t * \pi/2) \quad (7)$$

$$\text{SineIn} \quad f(t) = -1 * \cos(t * \pi/2) + 1 \quad (8)$$

Esibisce dunque un andamento molto simile alla quadratica ma con una pendenza leggermente inferiore.

La circolare è calcolata con l'equazione della circonferenza goniometrica: $x^2 + y^2 = 1$, che per $\pi/2$ assume la forma:

$$y = \sqrt{1 - (x - 1)^2}$$

Le prime due versioni, *In* e *Out* sono:

$$\text{easeInCirc} \quad f(t) = -1(\sqrt{1 - t^2} - 1) \quad (9)$$

$$\text{easeOutCirc} \quad f(t) = \sqrt{1 - (t - 1)^2} \quad (10)$$

Presenta un livello di pendenza collocabile tra la Quintica e l'Esponenziale. È la funzione che rende l'effetto di rallentando o accelerando più omogeneo per la distribuzione regolare delle variazioni di velocità sull'intera fase.

3.3 Back

La Back deriva il suo nome dalla versione *In*, dove si osserva un superamento backwards molto morbido della soglia iniziale. È realizzata con una cubica con un overshooting del 10% circa gestito dal fattore 'g' dell'equazione seguente [6]:

$$\text{InBack} \quad f(t) = (1 + g)t^3 - gt^2 \quad (11)$$

$$\text{OutBack} \quad f(t) = 1 - \text{backIn}(1 - t) \quad (12)$$

3.4 Elastic

La classe Elastic si basa su 2 funzioni distinte; la prima produce 3 cicli sinusoidali, la seconda traccia la curva decrescente che crea l'effetto di smorzamento.

$$\text{OutElastic} \quad f(t) = \sin(6\pi) 2^{10}(t - 1) \quad (13)$$

La *In*, viceversa, simula l'effetto di un progressivo accumulo di energia cinetica seguito da una rapida fase di attacco. Può dunque esemplificare fenomeni di discontinuità, o turbolenza, precedenti alla stabilizzazione di un dato livello.

3.5 Bounce

La classe Bounce produce l'effetto di un rimbalzo. Esegue 4 differenti funzioni in rapporto alla progressione della fase [6]. A titolo esemplificativo, si riporta la versione Out codificata in Java⁴:

```
if (p < 1 / 2.75) { // p < 36%
    return 7.5625 * p * p;
}
else if (p < 2 / 2.75) { // p < 72%
    return 7.5625 * (p -= 1.5 / 2.75) * p +
    0.75;
// curva esponenziale con inversione al 50%
}
else if (p < 2.5 / 2.75) { // p < 90%
    return 7.5625 * (p -= 2.25 / 2.75) * p +
    0.9375;
// curva esponenziale con inversione al 81%
```

⁴ <https://github.com/greensock/GreenSock-JS/blob/master/src/uncompressed/easing/EasePack.js>

```
}
return 7.5625 * (p -= 2.625 / 2.75) * p +
+ 0.984375;
// curva esponenziale con inversione al 95%
```

4. PARAMETRI

In molte librerie easing, ad esempio in quelle presenti nei programmi di grafica vettoriale Adobe, per il calcolo delle funzioni vengono adoperati 4 parametri: t (time), b (begin), c (change), d (duration) [7]. Il parametro 'time' indica la suddivisione temporale con cui la durata complessiva 'duration' della rampa viene misurata. Il parametro 'change' indica il valore di destinazione da raggiungere partendo dal valore offset 'begin'. La formula dell'interpolazione lineare ivi applicata,

$$i = c * t/d + b \quad (14)$$

distribuisce la rampa (*c*) lungo l'arco di una durata prestabilita. Il rapporto time/duration (*t/d*) è la fase normalizzata della rampa eseguita alla frequenza controllata dalla suddivisione temporale adottata ('time'). Le classi effettuano dunque una scansione temporale del parametro 'change' a partire dal valore 'begin'. La funzione con cui viene calcolata la curva si applica alla fase suddetta (time/duration).

Per creare una rampa lineare si userà dunque la funzione:

```
Math.linearTween = function (t, b, c, d) {
    return c*t/d + b;
}.
```

Esempio di valore restituito da una funzione easeInQuadratic:

```
Math.easeInQuad = function (t, b, c, d) {
    return c*(t/d)*t + b;
}.
```

Una semplificazione presente in molte librerie consiste nell'omettere i 2 parametri 'change' e 'begin' (che vengono calcolati successivamente sui valori restituiti dalla funzione) e di gestire la rampa come semplice coefficiente di fase.

5. IMPLEMENTAZIONE DELL'EXTERNAL "EASING"

Per una maggiore usabilità delle interpolazioni easing in ambito audio, è stato programmato l'external Java "easing", eseguibile in ambiente PD e Max tramite il proxy mxj. La versione attuale della classe si basa su un blocco di funzioni predefinite, riassunto nella tabella 1, scritto da Michael Heuer [8] e tratto da una delle librerie più diffuse, sviluppata da Robert Penner [6]. L'implementazione si basa, inoltre, su una ricodifica dell'opcode

“line”⁵ di Max finalizzata ad un uso dell’external più immediato ed efficiente.

Il clock interno dell’external, nella versione attuale, è impostato in modo costante a 1 ms e fornisce il parametro ‘time’ delle interpolazioni, ovvero la suddivisione temporale interna con cui la ‘duration’ viene misurata. La formula applicata per la ricodifica di “line” è:

$$result = begin + (time / ((1/(change - begin))) * duration) \quad (15)$$

La caratteristica principale di “line” è l’assegnazione ‘begin = result’ che viene aggiornata ad ogni nuovo avvio di rampa, congiungendo tra loro i segmenti di interpolazioni successive. Questa particolarità consente variazioni continue in runtime all’interno dell’inviluppo in esecuzione e rende l’algoritmo adatto ad un uso in tempo reale. In casi limite, quando i valori in ingresso si succedono con una rapidità superiore al tempo assegnato al fattore ‘duration’, i valori in uscita tendono alla successione stessa in ingresso. La rampa segue, dunque, l’andamento della funzione solo per la durata relativa all’assenza di nuovi dati in ingresso. Ad esempio, per una funzione *InOut* impostata con una durata di 1000 ms, se i valori in ingresso si susseguono a intervalli non superiori ai 500 ms, verrà letta solo la porzione *In*.

5.1 Algoritmo

Per la realizzazione delle curve, l’external implementa il seguente algoritmo:

0) Quando l’external riceve un valore in ingresso, esegue un metodo che preleva il valore corrente in uscita (*outNow*): *outStart* = *outNow*.

1) Calcola la differenza (*diff*) tra il valore in ingresso e il valore in uscita prelevato al punto precedente: *diff* = *change* - *outNow*.

2) Avvia la rampa e ne calcola la fase corrente, iterativamente per l’intero svolgimento della durata assegnata, come differenza tra i valori in uscita (*outFinal*) e il valore precedentemente prelevato al punto 0: *running* = *outFinal* - *diff*.

3) Converte la fase corrente in valori normalizzati: *running normalized* = *running*/*diff*.

4) La fase corrente normalizzata invoca la classe di interpolazione scelta dall’utente: *f(out)* = *running normalized*.

La selezione delle funzioni avviene in runtime tramite le API Reflection di Java. Il metodo che richiama la classe di interpolazione viene quindi lanciato solo quando se ne modifica l’istanza con un conseguente risparmio di risorse.

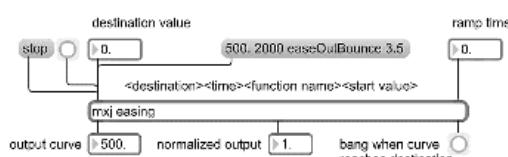


Figura 3. Interfaccia dell’external.

⁵ <https://docs.cycling74.com/max5/refpages/max-ref/line.html>.

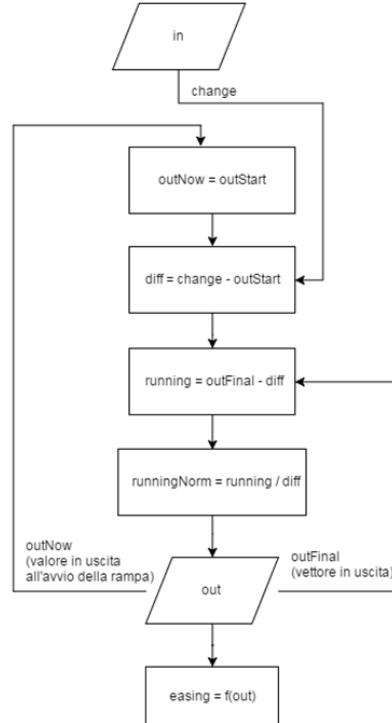


Figura 4. Diagramma di flusso dell’algoritmo.

6. INTERFACCIA

L’external possiede 2 inlet e 3 outlet (figura 3). Gli argomenti accettati sono: ‘nome della funzione’ e ‘change’ (inlet sinistro), ‘duration’ (inlet destro).

Per il settaggio degli argomenti tramite messaggi composti, la sintassi è:

<nome funzione> <change> <duration> <start point (opzionale)>.

Le proprietà implementate richiamabili tramite messaggi inviati all’inlet sinistro sono:

- la selezione della funzione che si desidera applicare <nome funzione>.
- l’interruzione dell’esecuzione della rampa sul suo valore corrente <stop>.
- l’esecuzione della sola rampa normalizzata, elaborata sulla base della durata impostata <bang>.

Gli outlet restituiscono in uscita:

- la rampa con i valori pesati (outlet sinistra);
- la rampa normalizzata (outlet centrale);
- un bang a conclusione della rampa (outlet destro).

I metodi che gestiscono le variabili numeriche in ingresso accettano, in overloading, sia tipi numerici interi (int) che in virgola mobile (float).

7. APPLICAZIONI

La sperimentazione qui condotta sull'uso delle interpolazioni easing in ambito audio ha analizzato principalmente l'applicazione delle curve a inviluppi di ampiezza e di frequenza nel dominio del tempo. In questo contesto le easing dimostrano rilevanza soprattutto nel trattamento di spettri estesi ed inarmonici, atti a simulare fenomeni cinetici prodotti da reazioni vincolari, come attrito e sfregamento, o azioni fluidodinamiche [9], in cui l'oggetto sonoro risultante tende a riflettere più l'intenzione, e la rapidità, con cui è stato generato che il significato 'melodico' correlato. L'aspetto di maggior interesse delle easing sembra risiedere, in tal senso, nella funzione isomorfica di stilizzazione e di rappresentazione iconica che esse svolgono nei confronti di codici espressivi figurativi [10]. Per l'immediatezza con cui le primitive di movimento, e il clustering dei parametri che esse organizzano, risultano assimilabili a stereotipi cinematici e gestuali riconducibili a categorie mimiche e prossemiche [11].

7.1 Relazioni cross-modali

Nell'esempio proposto⁶ per testare alcune delle possibili applicazioni musicali dell'external, si è usato come sorgente audio un generatore di rumore rosa filtrato con un passa banda e un passa alto controllati da funzioni easing; le medesime curve di interpolazione sono state applicate contemporaneamente sia al fattore Q che alla frequenza centrale e di taglio del filtro [12]. Per la scrittura delle versioni *InOut* la rampa è stata scomposta in 2 parti; nella seconda parte l'andamento è invertito al fine di compensare la discontinuità dell'angolo di decadimento illustrato in figura 1.

In ambito cinematico, l'ordinata della funzione controlla rampe d'ampiezza e rampe di traiettoria: le prime sono applicabili alle grandezze di massa e peso (afferenti alla pressione esercitata nei fenomeni di frizione), le seconde allo spostamento spaziale del corpo. Parallelamente in ambito audio, le rampe di frizione definiranno l'inviluppo d'ampiezza, così determinandosi un incremento figurato di massa proporzionale alla velocità delle fasi ADSR; le rampe di traiettoria definiranno invece le curve di portamento, con l'accelerazione dello spostamento spaziale proporzionale allo shifting in frequenza dello spettro audio.

8. BIBLIOGRAFIA

- [1] Topher La Fata: *Writing Max Externals in Java*, <https://pcm.peabody.jhu.edu/~gwright/stdmp/docs/writingmaxexternalsinjava.pdf>.
- [2] P. Morasso e V. Sanguineti: *Paradigmi di Controllo Motorio*, Dipartimento di Informatica, Sistemistica, Telematica (DIST), Università di Genova.
- [3] R.S. Woodworth: *The accuracy of voluntary movement*. Psychological Review Monograph Supplements, 3, no. 3, 1899.
- [4] P. Morasso: "Spatial control of arm movements," *Experimental Brain Research*, 42, pp. 223-227, 1981.
- [5] T. Flash and N. Hogan: "The coordination of arm movements: an experimentally confirmed mathematical model," *Journal of Neuroscience*, 7, pp. 1688-1703, 1985.
- [6] Rober Penner: *Robert Penner's Programming Macromedia Flash MX*, McGraw-Hill / OsborneMedia, 2002.
- [7] Oswald Campesato: *jQuery, CSS3, and HTML5 for Mobile and Desktop Devices*, Mercury Learning & Information, 2013.
- [8] Michael Heuer: *Interface Easing Function*, <http://www.dishevelled.org/interpolate/apidocs/org/dishevelled/interpolate/EasingFunction.html>.
- [9] F. Avanzini, S. Serafin, and D. Rocchesso: "Interactive simulation of rigid body interaction with friction-induced sound generation," *IEEE Trans. Speech and Audio Processing*, Vol. 13, No. 5, pp. 1073-1081, 2005.
- [10] M.M. Blattner, D.A. Sumikawa, and R.M. Greenberg: "Earcons and icons: their structure and common design principles," *Human-Computer Interaction*, Vol. 4, No. 1, pp. 11-44, 1989.
- [11] R.L. Birdwhistell: *Kinesics and Context: Essays on Body Motion Communication*, University of Pennsylvania Press, 1970.
- [12] A. Cipriani e M. Giri: *Musica elettronica e sound design*, Vol. 2, pp. 435-507, ConTempoNet, 2013.

⁶ Esempi di applicazioni audio e download dell'external Java: <http://www.renatomessina.it/?p=328>.

Implementazione su Max MSP di Solo[Nr.19] - K. Stockhausen.

Paolo Pastorino

Conservatorio Pierluigi da Palestrina – Cagliari

paolopastorino83@gmail.com

SOMMARIO

Il lavoro presentato in questo articolo è il risultato di uno studio, effettuato sull'opera Solo[Nr.19] di K. Stockhausen (1966), condotto con l'intento di creare un'applicazione software capace di semplificare notevolmente il lavoro necessario per l'esecuzione di Solo e di rendere autonomo il performer, attraverso l'interfaccia grafica, nell'utilizzo di tale applicazione.

La ricerca è iniziata durante il corso di prassi esecutiva tenutosi al Conservatorio di Sassari nel corso del terzo anno accademico (2014/2015) del triennio di musica elettronica, tale studio ha portato alla creazione di un'applicazione sviluppata in ambiente Max MSP¹.

In origine si è trattato di uno strumento abbastanza essenziale e rudimentale, è stato poi sviluppato ulteriormente e migliorato per il corso di prassi esecutiva e repertorio elettroacustico tenutosi al Conservatorio di Cagliari durante il primo anno accademico (2015/2016) del biennio di specializzazione in musica elettronica.

La collaborazione col M° Roberto Migoni - percussionista e docente di batteria Jazz presso il Conservatorio di Cagliari, il quale ha preparato una versione per vibrafono - si è resa assolutamente necessaria per comprendere meglio le problematiche relative all'esecuzione.

1. INTRODUZIONE

Trattandosi di un'opera realizzata nella seconda metà degli anni sessanta del secolo scorso risulta immediatamente chiaro che l'aspetto tecnologico prevedeva l'impiego di apparecchiature analogiche. Infatti, per l'esecuzione, Stockhausen ha previsto l'utilizzo di sei testine, disposte in un ampio spazio, per la lettura del nastro magnetico nel quale veniva incisa in tempo reale l'esecuzione del performer.

Copyright: © 2016 Paolo Pastorino. 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.

Le testine sono state disposte secondo delle distanze ben precise tali da garantire ciascuna un ritardo nella lettura del contenuto.

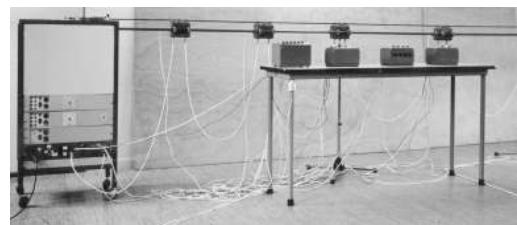


Figura 1. Set-up analogico per l'esecuzione.

La regia del suono prevedeva invece il coinvolgimento di tre assistenti, ciascuno era responsabile di gestire l'apertura e la chiusura dei potenziometri relativi al segnale in ingresso al sistema di registrazione, dei potenziometri relativi al feedback delle linee di ritardo ed infine dei potenziometri riguardanti l'output del segnale audio su due coppie di canali stereo.

Nel corso di questo articolo mi limiterò ad illustrare in modo sintetico il contenuto della partitura e descriverò più nel dettaglio gli aspetti tecnici che hanno interessato l'implementazione su Max.

2. PANORAMICA SULLA PARTITURA

Nella partitura sono riportate le istruzioni dettagliate sull'allestimento (strumentazione impiegata e schema a blocchi del sistema riguardante l'elettronica), sono presenti i sistemi con la notazione musicale destinati al performer, gli schemi per i tre assistenti, tutte le informazioni utili per la preparazione di una versione e per la sua esecuzione.

Gli schemi per gli assistenti (Formschema) sono sei come le pagine che riportano la notazione per il performer, queste devono essere associate ad uno solo dei sei schemi. Ciascuno schema è ripartito in sei cicli, ognuno di questi è suddiviso a sua volta in periodi caratterizzati da tempi di ritardo differenti per ogni ciclo; ciascuno dei

¹ <https://cycling74.com/>

cicli deve essere associato ad una pagina contenente la notazione per il performer, la correlazione tra queste è del tutto libera.

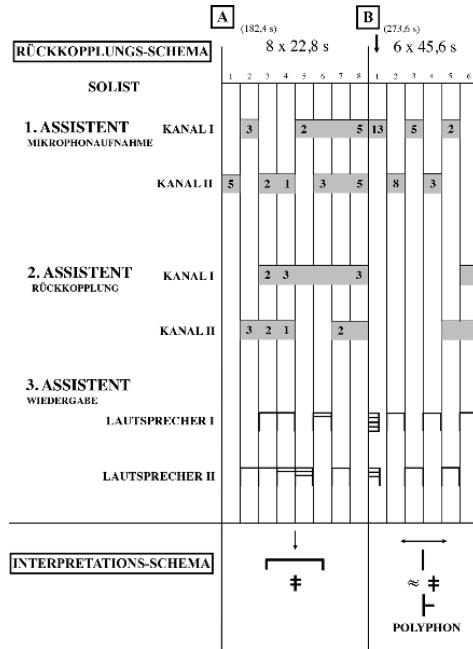


Figura 2. Estratto dello schema V.

Nello schema presente in Fig.2 sono riportati due dei sei cicli della versione V.

In alto, sopra ogni ciclo, viene specificato il numero dei periodi e la loro durata (es. 8x22.8s indica che il primo ciclo è costituito da 8 periodi della durata di 22.8 secondi ciascuno).

Le caselle grigie relative ai canali del primo e del secondo assistente indicano quando i potenziometri sono aperti, i valori presenti al loro interno indicano invece quante volte i potenziometri devono essere aperti e chiusi in modo quasi istantaneo durante ciascun periodo, nella partitura questi passaggi di stato repentini vengono chiamati perforazioni. La parte relativa al terzo assistente riporta una notazione differente rispetto a quella dei primi due, le linee indicate quando i potenziometri devono essere aperti mentre il loro numero indica la sovrapposizione del segnale riprodotto (layers).

La sezione in basso allo schema presenta dei simboli che riguardano l'interpretazione del performer, il loro significato è ampiamente spiegato nella partitura.



Figura 3. Una delle sei pagine per il performer.

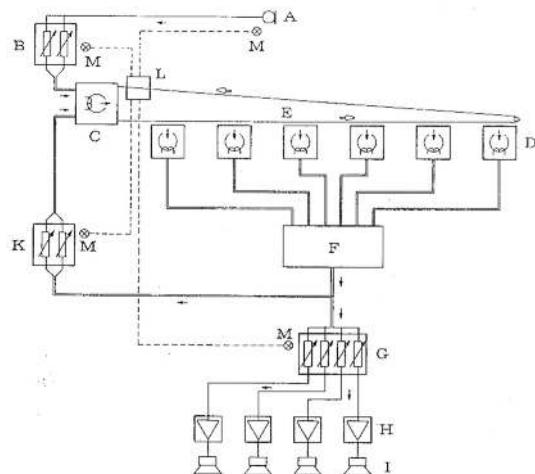


Figura 4. Schema tecnico per l'allestimento.

- A: Microfono
- B: Potenziometri per il controllo del segnale in ingresso al sistema
- C: Testina per la scrittura
- D: Testine per la lettura
- E: Nastro magnetico
- F: Switches per la selezione delle testine
- G: Potenziometri per il controllo del segnale in play back
- H: Amplificatori di Potenza
- I: Altoparlanti
- L: Metronomo ottico
- M: Metronomo ottico per gli assistenti

3. IMPLEMENTAZIONE SU MAX MSP

3.1 L'interfaccia grafica

L'idea di semplificare i processi di esecuzione è stata concretizzata automatizzando la gestione dei potenziometri che controllano il segnale in ingresso, il feedback ed il segnale in uscita, inoltre è stato implementato un sistema capace di eseguire le perforazioni stabilite nell'arco temporale di ciascun periodo.

Tutte le informazioni presenti negli schemi relative alla gestione dei potenziometri sono state memorizzate come file di preset, i dati corrispondenti le perforazioni sono stati invece memorizzati all'interno di sei file .txt ovvero uno per ogni schema. All'avvio della patch viene caricata una delle sei versioni e dunque vengono richiamati tutti i parametri suddetti.

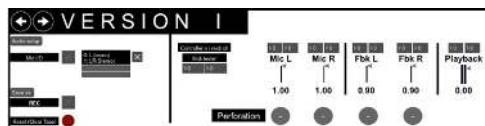


Figura 5. Controlli interfaccia.

Le frecce in alto a sinistra (fig.5) permettono di selezionare la versione da eseguire, ogni qualvolta ne viene scelta una i parametri ad essa collegati vengono richiamati cambiando così la configurazione della griglia e i ritardi per ogni "testina" di lettura.

Nel box a sinistra sono presenti i controlli per attivare l'audio in ingresso, il controllo per registrare la performance, un bottone per attivare o disattivare il microfono o i microfoni nel caso in cui si decida di utilizzarne due, ed infine un bottone che permette un reset generale della patch.

Nel box adiacente a quello appena descritto sono presenti i faders (potenziometri), gli indicatori delle perforazioni eseguite su ogni canale e dei number box grazie ai quali è possibile impostare il canale midi ed il numero del controllo nel caso in cui si decida di pilotare i faders attraverso un controller midi.

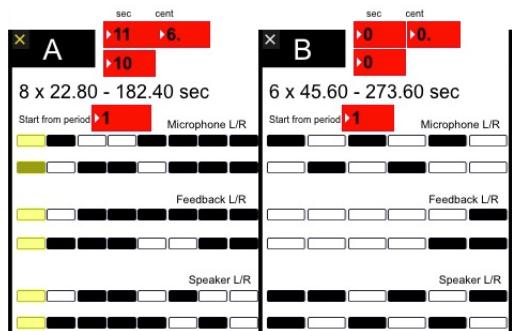


Figura 6. Estratto dello schema V dall'interfaccia grafica.

Per facilitare le prove al performer è stato implementato un sistema che permette di avviare l'esecuzione della patch da un periodo qualsiasi di qualunque ciclo, è sufficiente selezionare dal number box accanto alla scritta "start from period" il periodo dal quale si intende partire, dopodiché si avvia la patch cliccando il pulsante alla sinistra della lettera relativa al ciclo scelto.

Le caselle bianche e nere della griglia presenti nell'interfaccia indicano gli stati dei potenziometri, quelle nere corrispondono al valore massimo - di contro - quelle bianche rappresentano il valore nullo. Il marker giallo verticale indica la lettura del periodo corrente.

La griglia in Fig.6 ha come scopo quello di dare un riferimento grafico della timeline simile a quello in partitura.

3.2 Struttura della patch

Il motore del sistema è costituito da sei linee di ritardo per due canali (Left – Right), queste ricevono dei messaggi relativi ai tempi di ritardo quando la patch viene caricata e ogni qualvolta viene selezionata una versione.

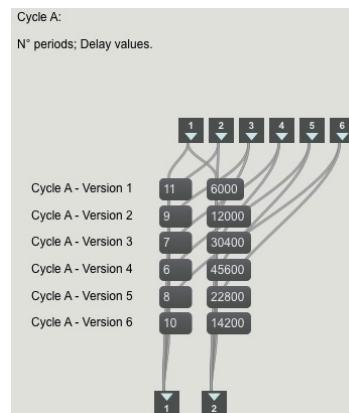


Figura 7. Setup griglia schema e impostazione dei ritardi.

La selezione di una versione mette in moto una serie di processi atti alla configurazione dello schema presente nell'interfaccia grafica ed al richiamo di tutti i parametri riguardanti le automazioni. Il sistema rappresentato in fig. 7 determina, a seconda della versione dello schema che viene selezionata, il numero di periodi (caselle della griglia) per ciclo ("A" nel caso in figura 7) e il relativo tempo di ritardo.

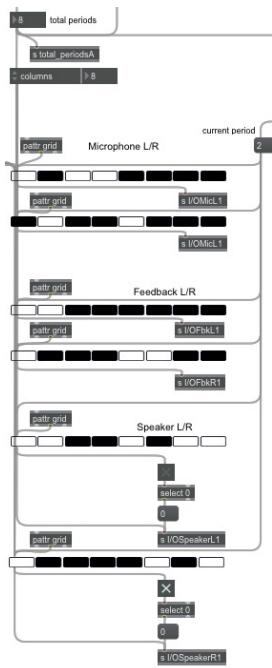


Figura 8. Griglia schema.

I valori relativi ai ritardi vengono inviati in contemporanea alle linee di ritardo e ai contatori che si occupano della lettura della griglia. Ogni ciclo ha un suo contatore, nel caso del ciclo A presente in fig.6 il contatore incrementerà il suo valore di 1 ogni 22.8 secondi, tale valore, che va da 0 a n periodi, indica quale colonna della griglia deve essere letta. Ogni contatore viene avviato all'inizio del relativo ciclo. Tutti i valori prodotti dalla lettura di ciascuna riga vengono inviati alle linee di ritardo.

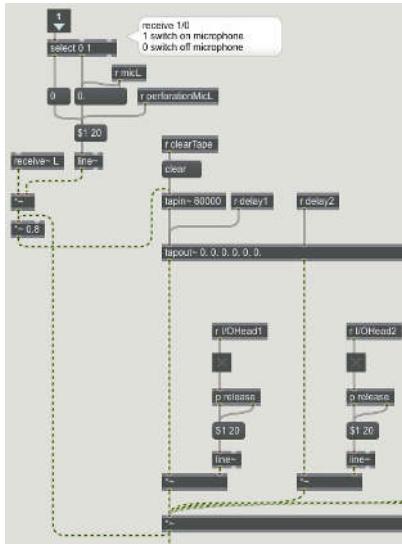


Figura 9. Subpatch linee di ritardo (estratto1).

Tutte le linee di ritardo ricevono il segnale in contemporanea dal modulo *receive~ L*, nel caso del canale sinistro mentre per il canale destro il modulo si chiama *receive~ R*, soltanto quella relativa al ciclo in funzione indirizza il segnale audio verso l'output. Il segnale esce dal modulo *tapout~* e viene moltiplicato per 1 sulla "testina" attiva e per 0 su quelle non attive, a valle di ogni moltiplicatore è presente un moltiplicatore unico che permette la reiterazione del segnale e quindi rende possibile l'effetto feedback.

I moduli *r delay* determinano il ritardo in millisecondi per ciascuna linea.

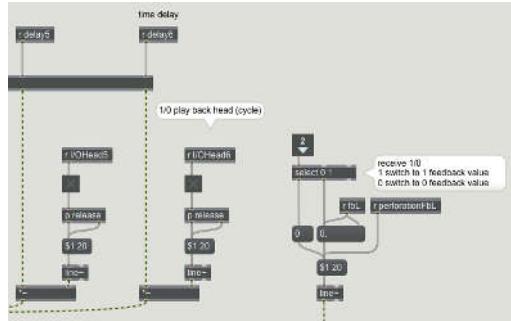


Figura 10. Subpatch linee di ritardo (estratto2).

Il blocco dei moduli presenti sulla destra in fig.10 agisce sul moltiplicatore a valle delle linee di ritardo, questo riceve i valori 1/0 relativi al feedback che vengono letti dalla griglia dello schema.

Lo stesso meccanismo di inviluppo viene applicato ai moltiplicatori che indirizzano il segnale audio verso il DAC (digital analog converter).

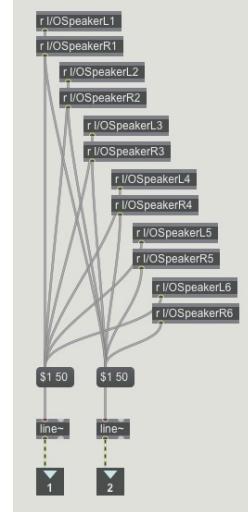


Figura 11. Subpatch che gestisce gli inviluppi per il controllo degli speakers.

3.3 Automazione delle perforazioni

Uno degli algoritmi più complessi della patch riguarda l'automazione delle perforazioni.

Il sistema implementato legge le informazioni, relative al numero di perforazioni da attuare per ogni periodo, da un file .txt contenente 4 colonne (mic left, mic right, feedback left, feedback right) e 51 righe (totale periodi per ogni schema), ogni schema fa riferimento ad un file diverso.

	1	2	3	0	0	0
1	1,	3	0	0	0	;
2	2,	0	2	2	0	;
3	3,	0	1	0	1	;
4	4,	2	0	1	0	;
5	5,	0	0	1	0	;
6	6,	0	1	0	1	;
7	7,	0	0	0	1	;
8	8,	2	0	1	0	;
9	9,	3	1	1	0	;
10	10,	0	2	0	1	;

Figura 12. Estratto del file riguardante lo schema I.

La prima colonna a sinistra riporta gli indici che identificano le righe, questi vanno da 1 a 51. È stato necessario prima di tutto assegnare ad ogni casella dello schema un flag identificativo. Essendo variabile il numero dei periodi per ciclo a seconda dello schema caricato non è stato possibile creare delle semplici costanti correlate a ciascuna casella, si è dovuto quindi procedere con delle variabili che vengono aggiornate sommando il totale dei periodi relativi al ciclo precedente a quello attivo con l'indice del primo periodo relativo al ciclo attivo. Es. se il ciclo A è costituito da n periodi allora la variabile relativa al periodo 1 del ciclo B avrà come valore $n_A + 1$, successivamente la variabile relativa al periodo 1 del ciclo C avrà come valore $n_A + n_B + 1$ e così via. Il risultato di ciascuna somma viene indirizzato verso il modulo *coll* richiamando così la riga che ha come indice lo stesso valore. La lista restituita da *coll* avrà i 4 valori destinati ad un'altra parte del sistema che esegue le perforazioni.

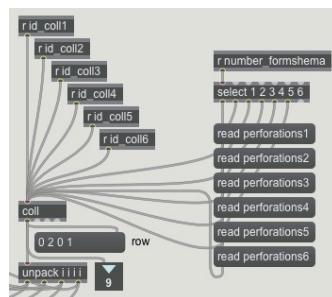


Figura 13. Subpatch perforazioni.

La seconda parte del sistema distribuisce il numero di perforazioni previste per ogni periodo nell'arco della durata relativa al periodo attivo. Es. se la durata dei periodi relativi al ciclo A è di 6 secondi e per il periodo $n = 1$ sono previste 4 perforazioni allora il sistema divide i 6

secondi per 4; al tempo parziale viene aggiunto o sottratto in modo casuale un valore compreso tra -400 ms e 400 ms , per ogni canale (mic left, mic right, feedback left, feedback right) tale intervallo cresce di 400 ms . Questo serve per distanziare le perforazioni tra loro.

Il sistema inoltre ricalcola il tempo residuo al seguito di ogni perforazione e procede alla reiterazione dei processi appena descritti. Per evitare che il risultato della somma di un valore casuale al tempo residuo successivo alla penultima perforazione ecceda oltre la durata del periodo il risultato dell'operazione viene riscalato per evitare l'eccedenza.

3.4 Alterazioni timbriche

Per l'esecuzione Stockhausen ha previsto delle alterazioni timbriche, queste sono segnalate in partitura con i simboli **N**, **I**, **II**, **III** per i cambi timbrici e con le indicazioni **Geräuschaft** (più o meno rumoroso), **Etwas geräuschaft** (rumoroso) e **Sehr geräuschaft** (molto rumoroso) per le gradazioni di rumore.

Tali alterazioni, come riportato nella documentazione dell'opera, possono essere realizzate sia fisicamente preparando lo strumento e/o mediante un sistema di elaborazione del suono.

Nell'articolo presentato da Enrico Francioni e pubblicato nell'edizione del CIM XVII (2008)² sono riportati i commenti relativi ad alcune versioni storiche eseguite da eccellenti interpreti. Le soluzioni attinenti al trattamento timbrico adottate in queste versioni sono diverse: le alterazioni timbriche eseguite attraverso l'elettronica riguardano principalmente l'impiego della Ring modulation e di vocoder; per quanto riguarda invece altre alterazioni si fa riferimento all'uso di sordine nella versione per trombone eseguita da Barry Anderson, al cambio di strumento nella versione per flauto eseguita da Dietmar Wiesner e all'utilizzo di differenti tecniche eseguite con l'arco nella versione preparata da Enrico Francioni.

Nella patch presentata in questo articolo il problema riguardante le gradazioni di rumore viene invece affrontato con l'utilizzo di un pitch shifter basato su linee di ritardo. Il sistema elabora il segnale su tre rami indipendenti ognuno dei quali traspone il suono in una zona frequenziale differente.

La trasposizione viene controllata da un pedale di espressione la cui escursione è suddivisa in tre steps, i valori che vanno da 0.0 a 0.33 interessano la trasposizione di tre ottave sopra il segnale originale, lo step intermedio che va da 0.33 a 0.66 applica una trasposizione di due ottave sotto, lo step finale traspone il segnale di 10 ottave sopra. I risultati di ogni trasposizione vengono sommati e ciascuno ha un peso differente in termini di ampiezza in modo da garantire un equilibrio tra le parti trasposte ed il segnale originale.

Per dare un risultato timbrico più interessante ho scelto di sommare al tempo di ritardo di ciascuna linea dei valori pseudocasuali compresi tra -1 e 1 .

² http://smc.dei.unipd.it/cim_proceedings/2008_CIM_XVII_Atti.pdf

Ogni step da un carattere diverso allo strumento e la variazione applicata al tempo di ritardo introduce una componente di rumore che aggiunge “rugosità” al suono dolce del vibrafono.

Per quanto riguarda invece i cambi timbrici segnalati con i simboli N, I, II, III ho deciso di affidare la scelta allo strumentista.

4. CONCLUSIONI

L'ultima revisione apportata alla patch ha permesso di raggiungere l'obiettivo posto in origine ovvero rendere più agile l'esecuzione di un'opera che presenta una certa complessità sia dal punto di vista tecnico-tecnologico che esecutivo.

La scelta di Max come ambiente di sviluppo è stata dettata da un'esigenza pratica ovvero dall'agilità che questo offre nello sviluppo e nella modifica dei lavori realizzati all'interno del suo ambiente.

Recentemente, lo studio di altri due linguaggi di programmazione, abbastanza simili tra loro, come Processing e p5.js³, ha orientato la mia idea verso un ulteriore sviluppo dal punto di vista grafico sia dell'opera discussa in questo articolo ma anche di altre opere le quali presentano un forte collegamento tra la parte grafica presente in partitura e la modalità di esecuzione. Questi due linguaggi sono uno standard in ambito grafico/artistico ed offrono grandi possibilità anche in campo audio grazie a librerie sviluppate da terzi come Minim⁴.

Ritengo molto interessante, ed anche importante dal punto di vista storico, realizzare una sorta di digitalizzazione di un'opera che interessa non solo l'aspetto sonoro ma anche quello grafico e l'interazione tra questi e l'utente.

Il lavoro resta dunque aperto a nuove evoluzioni.

5. BIBLIOGRAFIA

- [1] K. Stockhausen: *Solo für Melodieinstrument mit Rückkopplung Nr. 19*. Vienna: Universal Edition, 1969.
- [2] Mark Neremberg: ‘Structure Formation’: An Analysis of Electronic Superimpositions in Stockhausen’s Solo. 2012.
- [3] Enrico Francioni: *Omaggio a Stockhausen technical set-up digitale per una performance di SOLO [Nr.19]*. CIM XVII 2008.

³ <https://processing.org/>
<http://p5js.org/>

⁴ <http://code.compartmental.net/minim/>

SISTEMI INTERATTIVI IN AMBITO ARTISTICO PERFORMATIVO Mechanical Track Project

Marco Amistadi

Conservatorio G. Verdi di Torino
mameesto@gmail.com

Andrea Agostini

Conservatorio G. Verdi di Torino
andrea.agostini@conservatoriotorino.eu

ABSTRACT

Mechanical Track Project (MTP) è un percorso artistico tra espressione performativa e tecnologia, focalizzato sull'interazione tra interpreti e apparati tecnologici, e basato sulla convinzione che i sistemi interattivi offrano a un performer possibilità espressive ulteriori. Secondo tale ipotesi un sistema interattivo, oltre ad aumentare le possibilità espressive, facilita la creazione di elementi originali traghettando una performance verso declinazioni altrimenti irraggiungibili.

Il percorso progettuale di *MTP* poggia le fondamenta su un'indagine storica onde determinare punti focali attorno ai quali concentrare gli sforzi produttivi, e su un'analisi formale dei sistemi interattivi con l'obiettivo di esplorarne l'utilizzo in ambito performativo.

Dal punto di vista pratico *MTP* si concretizza attraverso la produzione di opere performative basate su sistemi interattivi. L'obiettivo estetico di tali produzioni è quello di individuare un equilibrio sempre più efficace nella fusione tra performance e tecnologia.

Lo scritto qui presente si propone di illustrare molto brevemente il percorso analitico di formalizzazione e esporre l'applicazione nelle produzioni performative sinora realizzate.

1. PRESUPPOSTI

1.1 Ricerca storica

Il percorso di *MTP* nasce dalla ricerca storica di antecendenti, referenti e fonti di ispirazione individuati in artisti, ingegneri, inventori e scenografi del presente e del passato. Tra questi Filippo Brunelleschi (1377-1446), Giacomo Torelli (1604-1678), Johann Nepomuk Mälzel (1772-1838), Edward Gordon Craig (1872-1966), Merce Cunningham (1919-2009), Giorgio Barberio Corsetti (1951-), Ariella Vidach (1956-) e Claudio Prati

Copyright: © 2016 Marco Amistadi, Andrea Agostini. 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.

(1954-), Antonio Camurri (1959-) Mark Coniglio

(1961-), Studio Azzurro (collettivo attivo dall'inizio degli anni '80), OpenEndedGroup (collettivo attivo dalla fine degli anni '90) e altri. Tale percorso, di cui non scriviamo nel dettaglio in questa dissertazione, ha permesso di determinare modelli attorno ai quali organizzare un'analisi formale delle interazioni [1][4][8][9][10]. Conseguente a tale analisi è stata costruita un'ipotesi di formalizzazione, esposta brevemente nel successivo paragrafo, orientata all'ambito artistico performativo [3][5][6][7][11]. La schematizzazione così formulata è servita come base per le produzioni interattive sinora realizzate in *MTP*.

1.2 Formalizzazione

Un qualsiasi sistema informatico si compone di almeno una periferica d'ingresso, un'unità di elaborazione e una periferica d'uscita.

Nel formulare una nostra definizione di sistema interattivo in tempo reale (vedi figura 1) prenderemo dunque in considerazione i tre blocchi di un sistema informatico più l'utente (performer).

Secondo questo assunto di partenza, il primo blocco di un sistema interattivo (il corrispettivo di una periferica d'ingresso in un generico sistema informatico) sarebbe rivolto all'acquisizione dei dati (attraverso sensori come mouse, webcam e altri).

Il secondo blocco (il corrispettivo dell'unità di elaborazione in un generico sistema informatico) sarebbe orientato alla gestione dei dati raccolti (tracciatura, mapping, data mining e altre tipologie di elaborazione).

Il terzo blocco (il corrispettivo di una periferica d'uscita in un generico sistema informatico) riguarderebbe la produzione finale in output dal sistema (per esempio una produzione audio o video).

Un quarto blocco, infine, consisterebbe nell'utente e nell'interazione di questi con il sistema (feedback interaction).

Bisogna poi pensare che l'obiettivo di un'interazione performativa è tipicamente quello di arricchire e potenziare una performance, la quale ha spesso ragion d'essere nel rapporto con il pubblico. Quest'ultimo normalmente non interagisce direttamente con il sistema (come in un'installazione) ma vi si correla attraverso la fruizio-

ne dell'intera opera (ad esempio osservando uno spettacolo di danza o teatro).

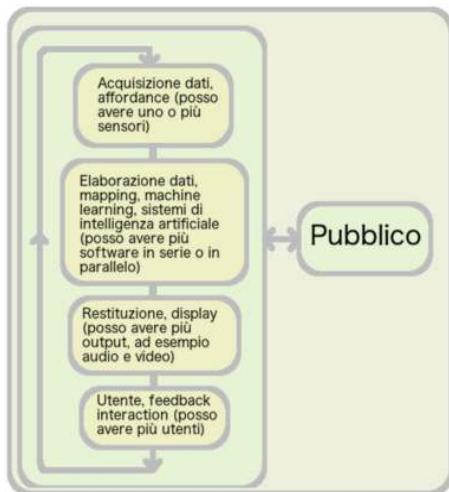


Figure 1. *MTP*, ipotesi di formalizzazione di un sistema interattivo ad uso artistico performativo.

La relazione tra performance e pubblico è un argomento aperto e delicato, molto discusso nella arti performative, e non assimilabile, se non in parte, al rapporto tra un servizio dato e un utente (come invece accade nel gaming). Questa particolare condizione rende piuttosto complessa la valutazione di un sistema interattivo orientato alla produzione di performance artistiche. Infatti, l'utente ultimo non è chi interagisce ma un'entità terza, esterna al sistema, che non ricerca gratificazione direttamente dal successo dell'interazione ma da una qualche modalità di fruizione dell'opera d'arte.

2. LE PRODUZIONI

MTP si configura come un percorso basato su un presupposto artistico che, attraverso l'interazione, coinvolge arte performativa e tecnologia. La concretizzazione ultima di *MTP* è la produzione di opere performative interattive alla ricerca di quello spazio in cui arte e tecnologia si incontrano in comuni obiettivi espressivi, estetici, comunicativi. Il miglioramento dell'equilibrio tra volontà artistica e tecnologia attraverso l'interazione costituisce l'elemento alla cui ricerca si rivolgono le produzioni di seguito descritte.

2.1 7.50

7.50¹ è uno spettacolo di teatro sperimentale, produzione della compagnia Refrain, orientato alla prosa e alla musica elettronica e interattiva. Testo, musica, regia e audio sono di Marco Amistadi, che figura anche come attore e performer. La produzione risale al 2012.

2.1.1 Scheda

- Spettacolo di teatro sperimentale
- Produzione: Refrain 2012

¹ Estratti video: <https://vimeo.com/67659111>

- Durata: 50 minuti
- Testo, Musica: M. Amistadi
- In scena: M. Amistadi (attore e performer)
- Sistema interattivo: Tavola lignea quadrata con 4 microfoni a contatto piezoelettrici applicati agli angoli

2.1.2 Interazione

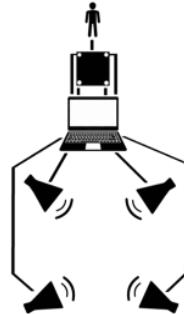


Figure 2. 7.50, schema del sistema interattivo.

In 7.50 vi sono diverse scene interattive. In una di queste (vedi figura 2), una tavoletta quadrata di legno è stata dotata di 4 microfoni a contatto piezoelettrici. Muovendo sulla superficie lignea della tavola il modellino in metallo di un vecchio pulmino, l'attore ne amplifica il suono. Ogni microfono è collegato a un diffusore di un sistema quadrifonico e questo dona alla performance un automatico metodo di spazializzazione. I microfoni si trovano agli angoli della tavola e, quando l'interprete dirige il modellino verso uno di essi, questo amplifica il suono con più vigore, convogliando l'amplificazione nel diffusore a cui è collegato.

2.1.3 Descrizione



Figure 3. 7.50, interazione sonora con modellino metallico e microfoni piezoelettrici su tavola di legno.

7.50 è uno spettacolo teatrale basato sulla narrazione di un abuso sessuale. Fondamentali sono le descrizioni sensoriali del protagonista. Le esperienze riferite ruotano attorno a viaggi pendolari su una corriera. I suoni creati dal modellino, dunque, si intersecano con il testo e producono un'esperienza acustica immersiva che proietta efficacemente il pubblico all'interno dell'ambiente narrativo.

2.1.4 Considerazioni

Nella sua semplicità il sistema interattivo descritto sopra si è dimostrato efficace. I punti di forza sono stati la facilità di utilizzo e il legame calzante dei suoni, e del gesto performativo, con quanto esposto nel testo recitato. Il suono ha conquistato in questo modo un valore narrativo poiché coerente con la materia testuale e con l'espressività emotiva in essa racchiusa.

2.2 MusicRoom

La performance *MusicRoom*², coproduzione Refrain e Tecnologia Filosofica³, risale al 2013. Si tratta di un lavoro di teatro danza per due performer (Elena Valente e Riccardo Maffiotti) e motion tracking pensato per spazi non convenzionali. Testo, musiche e audio sono di Marco Amistadi (ad eccezione di una citazione dalle Variazioni Goldberg di J.S.Bach).

2.2.1 Scheda

- Performance breve di teatrodanza e parola
- Produzione: Refrain 2013
- Durata: 10 minuti circa
- Testo, Musica: Marco Amistadi
- In scena: E. Valente, R. Maffiotti (attori, performer)
- Sistema interattivo: Sistema camera tracking a infrarossi; elaborazione tramite Max e Jitter di Cycling 74; output audio tramite diffusori.

2.2.2 Interazione



Figure 4. *MusicRoom*, schema del sistema interattivo.

La performance si avvale di un sistema interattivo camera tracking per la produzione audio (vedi figura 4). Il sistema consta di acquisizione video tramite camera a infrarossi, elaborazione tramite Max di Cyclig 74, output audio tramite una coppia di diffusori.

Grazie a un'operazione di frame-difference vengono individuate le differenze, per ogni pixel, tra frame successivi. Riscontrando quali pixel cambiano si ottiene una rappresentazione numerica del movimento. In output viene diffusa una produzione audio correlata a tale dato ottenuta attraverso processi di sintesi additiva, sottratti-

² Estratti video: <https://vimeo.com/79143817>

³ <http://www.tecnologiafilosofica.it>

va e di composizione algoritmica (sequenze MIDI suonate da strumenti virtuali) tramite due diffusori.

2.2.3 Considerazioni



Figure 5. *MusicRoom*, una scena della performance.

In *MusicRoom* l'apporto dato dal suono interattivo è integrato nella performance e introduce il pubblico in un territorio astratto che emerge dall'unione tra un testo assurdo, una sonorizzazione surreale e ambienti insoliti (come corridoi e piccole sale) dove è possibile percepire ogni minima variazione delle discipline espressive in gioco. L'interazione, che propone la relazione tra gesto fisico in scena e una conseguente reazione sonora, si è dimostrata un buon mezzo per traghettare il pubblico verso un ambiente narrativo surrealmente sospeso.

2.3 Alice

*Alice*⁴ è uno spettacolo per bambini ispirato a *Alice's Adventures in Wonderland* di Lewis Carroll e basato su un sistema interattivo molto semplice. Musiche, testi, regia, live electronics sono di Marco Amistadi; performer sono Elena Valente (danza e teatro) e Thuline Andreoni (teatro d'oggetto e di parola); la produzione è di Refrain e Tecnologia Filosofica. Lo spettacolo è stato commissionato dall'Unione Musicale di Torino e ha debuttato nel 2016.

2.3.1 Scheda

- Spettacolo di danza e parola per bambini
- Produzione: Refrain/Tecnologia Filosofica/Unione Musicale Torino 2016
- Durata: 40 minuti
- Testo, Musica: Marco Amistadi (da L. Carroll)
- In scena: E. Valente (attrice, danzatrice), T. Andreoni (attrice, manipolazione oggetti)

⁴ Estratti video: <https://vimeo.com/160942482>

- Sistema interattivo: Una videocamera riprende un piano di lavoro sul quale un'interprete sta agendo manipolando oggetti. La ripresa viene proiettata in tempo reale.

2.3.2 Interazione

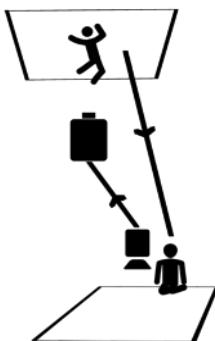


Figure 6. *Alice*, schema del sistema interattivo.

Il sistema consta (vedi figura 6) di una videocamera che riprende un piano di lavoro sul quale un'interprete (Thuline Andreoni) agisce manipolando oggetti. Questa manipolazione viene proiettata senza filtri su uno schermo con il quale interagisce un secondo interprete (Elena Valente). Quanto esegue il secondo interprete a sua volta condiziona l'opera del primo. Non si tratta di un'interazione uomo macchina, ma di un'interazione tra due interpreti di cui il primo fa le veci della macchina attraverso un impianto tecnologico (videocamera e proiezione) dedicato al video.

2.3.3 Considerazioni



Figure 7. *Alice*, interazione con carte da gioco, scacchi e altri oggetti.

L'interazione si è rivelata di grande efficacia. La presenza di un interprete a manipolare gli oggetti con mani a vista è stata importante per simulare l'idea di un gioco infantile di fantasia. Grazie all'uso di oggetti quotidiani, come mollette da bucato per interpretare pesci, vere carte da gioco per l'esercito della regina di cuori, batuffoli di cotone a imitare il fumo del bruco, si è mantenuto un contatto tra reale e fantastico. Il tutto in una dimensione metateatrale che interpreta l'uso immaginoso di elementi concreti come gioco operato da una bambina.

Il sistema interattivo descritto, che è in contatto con la realtà ma la ritorce fantasticamente su se stessa, ha il

doppio pregio di mantenersi chiaro agli occhi degli spettatori ma contemporaneamente fantasioso, imprevedibile, non lineare.

2.4 Mechanical Track 3

*Mechanical Track 3*⁵ è una performance breve per motion tracking, un danzatore (Aldo Torta) e Disklavier. Musiche e live electronics sono di Marco Amistadi.

2.4.1 Scheda

- Performance per un danzatore, motion tracking e disklavier
- Produzione: Scuola di Musica Elettronica del Conservatorio di Torino 2014
- Durata: 5 minuti
- Musica: M. Amistadi
- In scena: A. Torta (danzatore)
- Sistema interattivo: Acquisizione video tramite Kinect v1; driver OnopenNI NIMate; smoothing, scaling, routing e processi di composizione algoritmica realizzati in Max di Cycling 74; output disklavier Yamaha.

2.4.2 Interazione



Figure 8. *Mechanical Track 3*, schema del sistema interattivo.

Il sistema interattivo di *Mechanical Track 3* (vedi figura 8) agisce come descritto di seguito: Kinect restituisce le posizioni di varie parti del corpo nello spazio; operazioni effettuate in Max eliminano i noise data, veicolano i dati, calcolano la posizione degli arti rispetto al busto e la posizione di questo nello spazio; a seconda dei parametri e dei constraint impostati i dati della tracciatura subiscono differenti elaborazioni e producono un risultato finale diverso; la produzione musicale realizzata in Max di Cycling 74 è affidata a una macchina che prevede una vasta quantità di impostazioni possibili e inviata via MIDI a un disklavier Yamaha.

⁵ Estratti video: <https://vimeo.com/125737759>

2.4.3 Considerazioni



Figure 9. *Mechanical Track 3*, un momento della performance.

La performance è basata su una struttura musicale (e non narrativa come le precedenti produzioni) e si suddivide in tre parti: 1) un breve inizio senza tracciatura mentre il sistema produce sequenze compositive; 2) una sezione con tracciatura nella quale i movimenti del performer controllano il disklavier; 3) una terza parte che si avvale di un algoritmo di tracciatura diverso, modulante la produzione audio in modo più incalzante.

La realizzazione di *Mechanical Track 3* ha messo in evidenza una questione pratica su cui vale la pena soffermarsi. Il sistema usato per questa performance ha un livello di complessità elevato rispetto ai precedenti. Questo significa che ha un ampio spettro di possibilità espressive ma risulta anche più difficile da controllare se paragonato a un sistema semplice. Per questa ragione il raggiungimento di un uso consapevole dell'interazione ha necessitato uno studio da parte dell'interprete simile a quello di un musicista con il proprio strumento.

Emerge dunque che, se da un lato sistemi più complessi hanno spesso più possibilità espressive, dall'altro possono risultare inadatti in mancanza di tempi che consentano un'adeguata preparazione o in assenza della volontà da parte dell'interprete di adattare la propria tecnica di movimento alle esigenze del sistema.

2.5 Wind Interactions

Wind Interactions è un progetto di danza interattiva in via di realizzazione. La produzione terminerà entro febbraio 2017 e si avverrà del sistema di diffusione audio in 8 canali del Conservatorio di Torino.

2.5.1 Scheda

- Progetto di performance per un danzatore, motion tracking e processi di sintesi audio
- Produzione: Scuola Musica Elettronica del Conservatorio di Torino 2017
- Durata: 10-15 minuti
- Musica: M. Amistadi
- In scena: E. Valente (danzatrice)
- Sistema interattivo: Acquisizione video tramite Kinect v2; driver OnopenNI NIMate; gestio-

ne di algoritmi di machine learning tramite Wekinator; smoothing, scaling, routing, processi di composizione algoritmica e di sintesi audio realizzati in Max di Cycling 74; output e spazializzazione gestiti con Spat (8 diffusori).

2.5.2 Considerazioni

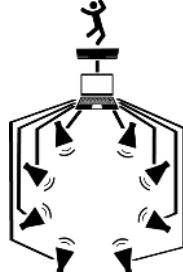


Figure 10. *Wind Interactions*, schema del sistema interattivo.

Wind Interactions vuole essere un'evoluzione di *Mechanical Track 3*. Il sistema (vedi figura 10) verrà ottimizzato e saranno introdotte nuove tecniche e elementi (come machine learning, sintesi audio, spazializzazione). La realizzazione si orienterà verso un sistema sufficientemente sfaccettato da poter supportare 10-15 minuti di performance senza risultare noioso. Si cercherà di costruire un'interazione che si installi sulla tecnica di danza già in possesso del performer per evitare lunghi periodi di training e concentrarsi da subito sullo sviluppo di un rapporto di interazione basato sull'espressività.

3. CONCLUSIONI FINALI

Raffrontando le produzioni artistiche e i modelli presi in analisi, possiamo evidenziare i seguenti punti ottenuti incrociando le riflessioni raccolte dalle diverse équipe attive nelle varie realizzazioni:

- 1) I sistemi più diretti e meno complessi si sono rivelati più usabili e efficaci.
- 2) La praticità e l'affidabilità rimangono elementi primari. In questo ambito hanno funzionato bene i sistemi più semplici e chiari (come quello di *Alice*) ma pure la modularità del sistema di *Mechanical Track 3* si è rivelata vincente, permettendo l'uso anche in altri progetti delle singole parti del sistema.
- 3) L'intelligenza dell'interazione è fondamentale ma lo è anche la fantasia. Trovare il giusto equilibrio tra questi due elementi è un lavoro di assoluta importanza.
- 4) Un punto delicato è la soddisfazione e la disinvoltura con cui gli interpreti usano i sistemi. Sistemi più complessi richiedono livelli di attenzione e training superiori, mentre spesso i performer si trovano a loro agio con sistemi nei quali non debbono cambiare il proprio consueto approccio creativo. Dunque bisogna, da un lato, creare sistemi che

siano il più possibile compatibili con i performer, in modo da non mettere in crisi il loro lavoro; dall'altro, scegliere i performer con cura, tra artisti che abbiano la capacità, la pazienza, l'interesse e l'abilità per lavorare con un sistema interattivo.

- 5) I sistemi interattivi si sono rivelati potenti mezzi quando hanno saputo inserirsi in modo strutturale e integrato nella realizzazione artistica e estetica.
- 6) La partecipazione alla realizzazione estetica, si rivela ancor più necessaria per performance lunghe, nelle quali i sistemi devono entrare a far parte degli equilibri strutturali.
- 7) Nelle performance lunghe i sistemi devono essere in grado di sostenere il loro uso nel tempo, senza risultare mai noiosi. Tra le strategie possibili possiamo elencare:
 - usare più di un sistema interattivo in una stessa performance;
 - creare sistemi con reazioni originali, espressive e sempre varie nel tempo;
 - usare il sistema interattivo solo in alcuni momenti strutturalmente significativi.
- 8) L'ascolto e la disponibilità che l'interprete deve dedicare ai sistemi è importante. Tale pratica deve essere necessariamente più intensa con sistemi sfaccettati e dalle ampie possibilità comunicative.
- 9) In genere, sistemi più complessi possono offrire maggior profondità espressiva, ma l'apporto di caratteristiche tipiche dell'arte (fantasia, comunicatività, originalità, etc.) può rendere espressivo anche un sistema semplice.

In definitiva, gli esperimenti di performance interattive realizzate, ci hanno convinto a:

- 1) Procedere nel progetto cercando sistemi puntuali e intelligibili ma originali, con tipologie differenti di non linearità, e coinvolti direttamente nel processo creativo.
- 2) Trovare un'intesa forte tra il sistema e l'interprete tale da stimolare in quest'ultimo un corretto stato di ascolto e la volontà di imparare a interagire con la macchina.
- 3) Calcolare i tempi di preparazione del sistema e quelli di apprendimento del performer e prevedere a tal fine delle prove dedicate, da collocarsi in una fase antecedente rispetto alla preparazione della performance vera e propria.

Il percorso costruttivo del sistema andrà tracciato a partire dalla mediazione tra le seguenti esigenze:

- 1) Progettare sistemi complessi che, seppur difficili da usare, abbiano più ampie possibilità espressive.
- 2) Rendere i sistemi sempre più agevoli e immediatamente sensibili alle esigenze del performer, quindi meno vari ma facilmente usabili.

Contemporaneamente bisognerà tenere in largo conto i seguenti fattori:

- 1) Affidabilità e praticità del sistema;
- 2) Non lavorare in astratto, ma pensare al processo di costruzione come facente parte della realizzazione creativa, confrontando continuamente i risultati ottenuti con le necessità artistiche della performance.

Riassumendo possiamo individuare una strada per la realizzazione di prossime produzioni in sistemi basati su relazioni uomo-macchina equilibrate, espressive, fantasiose, inserite nel progetto creativo e artistico, ma contemporaneamente sempre più immediate e semplici da usare.

4. REFERENCES

- [1] Buzzi M. L.: "Compagnia AIEP" in *La Scena Digitale – La Scena Italiana*, ed. Marsilio, 2001.
- [2] Brockett O. G.: "Storia del Teatro, dal Dramma Sacro dell'Anico Egitto agli Esperimenti di Fine Novecento", ed. Marsilio, 2000.
- [3] Camurri A.: "L'interazione Uomo Macchina nell'Informatica Musicale" da *Le Arti Multimediali Digitali*, ed. Garzanti 2010.
- [4] Camurri A.: "Il Progetto EyesWeb per Musica, Danza, Teatro", in *La Scena Digitale – Nuovi Media Per la Danza*, ed. Marsilio, 2001.
- [5] Caramiaux B., Françoise J., Schnell N., Bevilacqua F.: "Mapping Through Listening", *Computer Music Journal*, vol 38, n3, Fall 2014.
- [6] Downie M.: "Choreographing the Extended Agent - performance graphics for dance theater", dissertation, Massachusetts Institute of Technology, 2005.
- [7] Fiebrink R.: "Real-time Human Interaction with Supervised Learning Algorithms for Music Composition and Performance" PhD thesis, Princeton University, 2011.
- [8] Mendini E.: "Danza e Tecnologie in Italia: Percorsi" in *La Scena Digitale, Nuovi Media Per la Danza*, ed. Marsilio, 2001.
- [9] Schaffer S.: "Enlightened Automata", in *The Sciences in Enlightened Europe*, Chicago and London, The University of Chicago Press, 1999.
- [10] Vasari G.: "Le vite dei più eccellenti pittori, scultori e architetti - Vita di Filippo Brunelleschi, Scultore e Architetto Fiorentino", "Liber Liber" progetto Manunzio, Edizione edita a Firenze per i tipi della Giunti, 1568.
- [11] Verplank B.: "Interaction Design Sketchbook", articolo per i corsi del CCRMA, <http://www.billverplank.com>, 2009.

AESTHETIC AND PHENOMENOLOGICAL APPROACH TO SOUNDSCAPE

Roberto Zanata

Conservatorio "Monteverdi" di
Bolzano

robertozanata@gmail.com

Federica Pau

Institute Catholique de Toulouse
Equipe de recherche
Métafysique, histoire, concepts, actualité

federpau@gmail.com

ABSTRACT

The modern city is a space filled with signs and sensory stimulations. Therefore if we go beyond the supremacy of the sight sense, a study of aesthetological nature on the sound characteristics of the city, and on our capacities of turning them into experience through hearing, will allow us to perform a new range of analysis clearly rooted in phenomenology. Being meant as invisible and incorporeal cultural good changing due to variable factors, the sounding landscape becomes exposed to the reflections of urban aesthetics, thus helping a new "sensory" approach to city planning in order to carry on a kind of urban exploration aiming at enhancing modes of perception that have been neglected up to now. Although this study should be necessarily compared with the analyses Murray Schafer made in his Sounding City, the guideline of the present research will embrace an aesthetological and phenomenological approach.

1. SOUNDSCAPES AND MIGRANTS: AN AESTHETOLOGICAL POINT OF VIEW

The modern cities a space filled with signs and sensory stimulations. Therefore, if we go beyond the supremacy of the sight sense [1], a study of aesthetological nature on the sound characteristics of the city and on our capacities of turning them into experience through hearing will allow to perform a new range of analysis. Being meant as *invisible and incorporeal cultural goods* changing due to variable factors, the *sounding landscape* becomes exposed to the reflections and researches of urban aesthetics, thus helping a new "sensory" approach to city planning in order to carry on a kind of urban exploration aiming at enhancing modes of perception which have been neglected up to now.

Copyright: © 2016 Zanata R., Pau F. 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.

The scope of the present paper is to focus our attention on acoustic space of the city in relation to the experi-

encing subject, which so comes into relation with the surrounding environment. Such relation has been pointed out also by Walter Benjamin in *The Work of Art in the Age of Mechanical Reproduction* [2], where it is stated that, unlike painting and sculpture, architecture and the city are made for collective reception. As a matter of fact, all the senses are concerned with them, as we enter into them with all our body and do not just glance at them with our eyes. This observation may be referred to the differences which result from comparison between the aesthetical experience of the city and architecture, and that of painting and sculpture. As a matter of fact, on one side the city and architecture form an aesthetical reality that we can observe, as we live and act inside them; on the other side, the aesthetical experience of painting and sculpture confines us to the role of spectators, as our enjoyment of a painting or a statue is always horizontal reception.

Such reflection allows for manifold developments, as the experience itself of architecture and the city may be fairly compared with listening a music piece, which defines an acoustic space and, thus, a form our life is *inside*, a sound environment *tout court*.

On the other hand, the city - just like landscape and architecture - is not only a visual, but also a sound environment. With regard to this, we can't help quoting from Rosario Assunto's *Il paesaggio e l'estetica*: "Listening, being no voluntary act and, for this reason, ever more conditioning, is part of the aesthetical experience of the city in equal measure with and with the same title as sight" [3].

On one side, such reflection opens the door to the relation between hearing perception and synaesthetical perception and, on the other side, makes us realize that the mere side position of ears in humans allows hearing to be one of the most important human sensors of space. So, we can concentrate on hearing by comparing it with sight and focus on the differences between the ear and the eye: The ear has no eyelids, writes Shaffer, and while the eye goes outside, the ear goes inside. It absorbs information [4].

With regard to this, the concept of soundscape defines both an ear-centered perceptual phenomenon, in parallel with the classical definition of landscape, and – according to R. Murray Shafer's definition – any field of acoustic study made up of heard events (instead of seen objects) [5].

Both definitions suggest that any living space defines a soundscape, and so the search for the acoustic identity of places affects the very nature of listening.

The world is teeming with sounds and noises that call for our conscience and impress themselves on our experiences in a profound way. The places we inhabit are therefore marked by a sound identity linked to their morphology, to their specific natural configurations, to the forms of settlement that inhabit it, to the cultural dimension of those who use such resources. According to these different conditions, then, sounds and noises define different soundscapes.

These landscapes represent a real cultural heritage, made up of intangible forms of expression, which maintain a strong connection with the corporeality of places and culture of the people living there.

As it constitutes the basis for elaboration of an analytical study, this kind of analysis can play the basic role of directing the research towards the aesthetical experience of the sounding landscape in urban environment, a subject which has been much neglected or, too often indeed, dealt with in hasty manner. This afore mentioned theoretical framework can develop a real dialogue with the so-called *Neue Ästhetik*, having the concept of atmosphere as its theoretical hub and Gernot Böhme, the author of *Aisthetik und Architektur und Atmosphäre* as its main representative [6].

As we know, reflection on atmospheres involves today different experience fields and heterogeneous disciplines, among which social sciences, psychology, paedagogical sciences, art theoretics (especially architecture and music) and, last but not least, city planning. However, we would like to strongly point out that Böhme is mostly concerned with interpreting atmosphere as a preferential reading key for aesthetical experience. In this regard, since we can consider atmospheres as "interstitial" phenomena between subject and object, or intermediate phenomena located between psychological interiority and environmental exteriority (including not only the "natural", but also the "social" and "communication" environment), we should ask ourselves the question whether it is possible to aesthetically explore a place according to its sounding atmosphere. This is a question Wim Wenders perfectly answers not only from a theoretical and aesthetological general point of view in his *Lisbon Story* [7].

As a matter of fact, this analysis does not want to remain confined to the world of ideas and aims not only at providing soundscapes with a theoretical and aesthetological basis, but also at finding application when the theoretical framework can become a subjective study of every single sounding landscape.

In addition to their peculiarities, the latter can be also examined with reference to the timelapse of the day and the clima and season changes. A good example may be the city nighttime, which is characterised not only by the onset of darkness, but also by different relationships with sounds and noises.

With regard to this, also the experience of silence can be considered as playing a central role [8]. That's why the change of the seasons and the different climatic con-

ditions may give interesting suggestions for analysis, which prove useful where the urban space being investigated shows specific characteristics like the presence of a watercourse or a port, or when the subject of analysis are cities where the phenomenon of fog is common, causing temporary visual attenuation resulting in making hearing more acute [9].

However, we should point out that, if it is true that places grow out of, and develop themselves in relation to its inhabitants, this is even more evidently so with regard to those places in European cities that are inhabited and traversed by the so-called "new urban crossbreed"; people who see the city as a permanent home, migrants in transit, citizens who don't just live in, but live through, the city. With their passages, migrants leave traces behind themselves, which include sounds and noises integrating with an already existing wealth of sounds and noises. In other words, the intangible cultural heritage to which we have referred above is constantly being enriched with new resources. In fact, "Living in a place" means creating an *habitudo*, that is, an acquaintance with the place itself, an attention directed towards its specificity, a behavior informed by the place we live in and that we make our own lifestyle.

"Contemporary nomadic uprising", in fact, "is a complex, polyphonic condition" [10].

Moreover, music itself builds spaces. It expands the place in which we live, filling it with properties defined by the interaction between the harmonic and tonal quality of the sound, and also between the geographical and cultural import of music; the latter being particularly reflected in the works of XX and XXI century composers, whose universality (that is, their ability to speak for everyone, to move all sorts of – minimally cultivated – audiences) paradoxically arises from their being strongly bound to folkloric (and not only strictly musical) elements.

Acoustic experience is therefore in itself a creative experience; the sounds and music generating from distinctive environmental conditions can open new paths through which the listener's imagination flow. Contamination between different cultures multiplies sound-space expansion and produces unpredictable hybrids. The introduction of sampler had the effect of amplifying sound migrations: unpredictable feedbacks between distant musical cultures became possible, prior physical and cultural obstacles being now easily overcome. The same metropolitan experience brings us to recognize the reality of these sound-mixing of cultures. In cities' nodal areas – where different ethnicities and cultures intersect and overlap – new soundscapes, with their mixture of language-, city-, radio-sounds, constantly rise up, each one evoking different places and territories, and so reflecting the multicultural and unstable society in which we live.

In regard to this, one should note the expressive richness of a city neighborhood where people from different ethnic groups live and work, each one talking in their own language and, thus, setting the limits of a diversified acoustic space where a strong local accent and "all

languages of the world" mix [11], thus becoming itself an object of aesthetic experience subject to aesthetic judgement. Think more of the way such landscape changes when the immigrant groups express themselves through their own music, which penetrates into the local, dominant musical culture and creates a "different" sounding landscape enriching the city's immaterial culture heritage.

This analysis cannot but take into account also the inside-outside, private-public space relationship, since even a house, as taught by Lloyd Wright, may have its specific sound and relate to the sounding of the neighbouring area [12].

2. SOUNDSCAPE AND MIGRANTS: FROM THE TECHNICAL POINT OF VIEW

The new forms of economic, social, political, cultural, mobility and information are distributed unevenly in space and in terms of quality and quantity; they maintain numerous and heterogeneous processes of decomposition and territorial and cultural recomposition. Interdependencies of networks and flows call into question the traditional public and private spaces, combine old and new forms of citizenship, draw new landscapes, where even the sound dimension of migrant populations acquires its significance. In fact, by the sounds of the countryside to urban traffic noise sounds alter moods and actions (Schafer, 1977). However, with the advance of technology, the world has become increasingly noisy and citizens are also looking for new and different sounds that can comfort and reassure them (Bull e Back, 2008). In Europe, urban landscapes inherited from history are enriched with new sensory, in which old and new citizens may from time to time to practice their everyday relationships, investigate their roots and projecting their future.

The soundscape is a complex topic, which requires an interdisciplinary approach, already evident at the time of publication of *The tuning of the world* of Raymond Murray Schafer (1977), which remains a fundamental text although by then the international literature on this topic is grown considerably. In *Modernity at Large* the anthropologist Arjun Appadurai defines ethnoscapes as "the landscape of persons who constitute the shifting world in which we live; tourists, immigrants, refugees, exiles, guestworkers and other moving groups and persons constitute an essential feature of the world and appear to affect the politics of (and among) nations to a hitherto unprecedented degree." The research proposes to verify the existence of ethno-soundscapes, soundscapes that are identified in the process of building cultural diversity, as "typical" and characteristic of a particular ethnic group.

The aim is to study the urban places that old and new citizens relate to the soundscapes of their roots, which places projecting their future, and in which they practice their everyday. The research aims to transform the aesthetic experience of the cities (Cagliari, Parma and Milan), and reap the soundscape that relate to the rest of

the world. The analysis of the specificity of soundscapes and linguistic landscape of the three case studies show the complementarity of Cagliari, Parma and Milan. With Milan, Parma shares a central geographical position of belonging to the "heart" of the most attractive immigration, the Centre-North in which lies the vast majority of the foreign population resident in Italy. The differential gap, however, is the limited size scale. It may therefore be interesting to compare the context of a medium-sized city like Parma with that of a metropolitan area as Milan. Compared to Cagliari, open door of Europe on the Mediterranean Sea, instead, Parma has a certain dimensional affinity, which is accompanied, however, with a different geographical location; in this case, therefore, it is useful to compare the role played by the centrality/marginality dialectical in the characterization of tissue migration of the two cities. The possibility of observing certain phenomena in different geographical contexts, characterized by different size and location, it helps to understand the role played by some factors in the evolution of soundscapes related to immigration. For this reason, the comparative size of the study, divided into three locations, it is a necessary and characterizing methodological approach of research. The expression "soundscapes" means, here, took a large of the term, in order to ensure a useful complementary offer from different points of view regarding the size of the sound anthropogenic experiences in an urban context. The research therefore focus:

- screening and analysis of urban soundscapes (the set of background sounds of the many activities that take place daily in urban settings);
- analysis of the artificial "soundtrack" of city that is part of what musicologists have called "muzak", the background music stereodiffusa in a wider range of differentiated and public places (shopping centers, shops, restaurants, waiting rooms, airports, train stations etc..);
- an understanding of the cultural transformations under way: is undoubtedly an innate geographical and anthropological value that redefines the aesthetics of the soundscape through digital archives;
- the need to reduce noise pollution through public awareness.

The sound profile of a society is expressed through the sounds it is "involuntarily" as inevitable effect of the activities and discussions that take place here, both through the production of "voluntary" expressions of music and soundscapes common in urban environments (music for commercial use produced, played in public, or stereodiffusa in public places).

The soundscape of the world is ever changing. It is the most extensive in urban areas. The soundscape is a 'cultural heritage' invisible and intangible, that because of the many different anthropogenic and natural changes. So Human Geography can not afford to neglect this area of research, as it was the case in Mediterranean Europe for many years. The general deterioration of living conditions in many European

cities has contributed to a deterioration in the quality of the landscape in general and of the soundscape in particular. Unlike other phenomena more widely known as the loss of biodiversity, the pollution of the environment, the landscape can deteriorate even to the neglect of those who should keep it alive (Dal Borgo e Gavinelli, 2012).

These conditions are as follows:

- Recording *in situ* of soundscapes from areas of high density of migratory presence: residential presence and occasional episodic and use of public spaces. The aim is to provide a "catalog" of soundscapes from characterize urban areas of migration into the city sample;
- Qualitative interviews of migrants on the perception of soundscapes from the city of arrival. The research questions related to this research activity are: what does the soundscape diffused in the places to immigrants who settle in them? What is the connection between the soundscape of the starting areas of the migratory project with those areas of arrival? What are the main differences that migrants can identify the soundscape of the city of arrival? How are perceived soundscape related to music popular in urban areas (shops, malls, waiting rooms)? Is perceived as the "linguistic landscape" by Italian-speaking people who are not native speakers of Italian-speaking? Are perceived by migrants in a distinct manner soundscapes related to the sphere of the sacred, what Murray Schafer defines "sacred noise" (such as the sound of church bells)?
- Qualitative interviews of Italian residents aimed at detection of the role played by the sound dimension in the perception of the presence of migrants. The research questions related to this phase are: the intercultural relationship between population "native" and immigrant population is also expressed through a comparison between different soundscape? The otherness of migrants is perceived (and thus constructed narrative) through a purely sonic diversity? The construction of the cultural representations of diversity is also based on the dimension of sound and music? As perceived linguistic landscapes of migrants, the sound profiles of languages, unknown to the majority of Italian speakers 'native', but who are increasingly affirming not only in landscapes minutes of advertising billboards, but also in the sound profiles of public places (transport, squares, shopping centers etc.)?

From a methodological point of view, the research will exceed the simple collection of acoustic tracks, with the aim of achieving a mapping of the city, where you can

find sounds more or less known of the newspaper. In fact, research of this type have already been conducted in recent years with various results, although of great interest, such as those dedicated to the city of New York, London and Cologne. The search for cultural geography, in combination with that of audio artists, allow to build a map of the acoustic heterotopias of Cagliari, Parma, Milan.

From the role of the acoustic, geographical investigation will examine sounds, music, voices of Cagliari, Parma and Milan, to redefine through sound the geosemiotics of the city. The expected results are both short and long term.

The research aims to contribute to a conceptual point of view on geographic comparison:

- 1) geosemiologia and polysemy (heterotopias);
- 2) the complexity of European identity (multiculturalism);
- 3) multi-sensory cultural heritage (Ethnoscape, Soundscape and smellscapes Heritage).

From a methodological point of view the aim is to develop an interdisciplinary research space, in which the innovative aspects relate to the heuristic approach to the soundscape. The pursuit of cultural geography and migration, in combination with the artists audio, intends to build a map of heterotopias of the explored cities.

Perhaps more modest scientific interest, but certainly cultural value, the search will produce digital collections of audio tracks of cities, who will find a new space of sound creativity, and links to sound libraries, where you can license the recordings for commercial or institutional.

The survey therefore tends to investigate the existence and characteristics of the "sound component" of immigration. The starting hypothesis of the research is, therefore, the existence of a specific dimension of the auditory intercultural contact. This suggests the next question, linked to the sphere of integration: there is a sound component integration processes that result from the presence intercultural hope in contemporary society? The research would then, in this perspective, to suggest guidelines on the use of sound component in the processes of intercultural mediation adopted in size trans-scale territorial government bodies involved in the management of migration flows and integration processes (Municipalities, Provinces, Regions).

3. CONCLUSION

In general, what we are aiming to accomplish with our methodological and musicological approach is to indicate a possibility of shifting some theoretical considerations, developed through a local research on a specific area, towards a more universal plan regarding the demonstration of "*the way we listen*". The way in which we intent to investigate the sound, by using a microphone, will allow us to emphasize the depth of the space and create a true drama thanks to careful editing that will present the sound object not as an interesting figure standing out from the background, but as the object capable of formulating relations of fusion or

distinction, partially blurring. Since the sounds will be displayed in clean segments and vague segments, masked from the others by complexity and stratification of sound space, the data will be both hi-fi and low-fi. The field recording does not identify single events, but sees the environment as a whole, highlighting the complex network of interactions between distinct sounds and emphasizing its continuity.

In other words, the purpose of the research and the subsequent archiving and cataloguing of the obtained data is not to grasp the original sounds of a certain investigated area, but to depict its modernity together with a principal contribution of old (indigenous) and new (migrants) inhabitants.

4. REFERENCES

- [1] R. Debray: *Vie et mort de l'image. Une histoire du regard en Occident*, Gallimard, Paris, 1992; A. Barbara, *Storie di architetture attraverso i sensi*, Postmedia, Milano, 2011, p. 31.
- [2] W. Benjamin, *The Work of Art in the Age of Mechanical Reproduction. The Work of Art in the Age of Mechanical Reproduction and other writers on media*, tr. Edmund Jephcott, Rodney Lingstone, Howard Eiland and Others, Harvard University Press, Cambridge Massachussets – London England, 2008, p. 40
- [3] R. Assunto, *Il Paesaggio e l'estetica*, Giannini, Napoli, 1974, vol. I, p. 170.
- [4] M. Shaffer, *The Soundscape: Our Sonic Environment and the Tuning of the World*, Destiny Books, Vermont, 1994, pp. 11-12.
- [5] Ivi, p. 7.
- [6] G. Böhme, *Aisthetik, Vorlesungen über Ästhetik als allgemeine Wahrnehmungslehre*, München, Wilhelm Fink Verlag, 2001; G. Böhme, *Architektur und Atmosphäre*, München, Wilhelm Fink Verlag, 2006. See too: T. Griffero, *Atmosferologia. Estetica degli spazi emozionali*, Laterza, Roma-Bari, 2010.
- [7] Wim Wenders, 1994.
- [8] N. Polla-Mattiot, *Pause. Sette oasi di sosta sull'orizzonte del silenzio*, Mimesis, Milano, Mimesis, 2012.
- [9] A. Barbara, *Sensi, tempo e architettura. Spazi possibili per umani e non*, Postmedia, Milano, 2012, p. 32.
- [10] G. Attili, *rappresentare la città dei migranti. Storie di vita e pianificazione urbana*, Jaca Book, Milano, 2008 p. 17.
- [11] "Traîner dans le Panier, c'est sentir le vieux cœur de Marseille palpiter. Un cœur qui parle les langues du monde, les langues de l'exil". J-C. Izzo et Daniel Mordzinski, *Marseille*, Hoëbeke, Paris, 2000, p. 24.
- [12] R. Favaro, *Spazio sonoro. Musica e architettura tra analogie, riflessi, complicità*. Marsilio, Venezia, 2010, pp. 166-188; B. Brooks Pfeiffer, *Wright*, Taschen, Köln, 2006, p. 53.
- [13] J.Blauert, U.Jekosch, *Sound-Quality Evaluation - A Multi-Layered Problem* in ACUSTICA/acta acustica 1997, Vol. 83, No. 5, September/October 1997, pp. 747-753 USA: Wiley; 2006. p. 887-909.
- [14] G.A.Dal Borgo e D.Gavinelli, *Il paesaggio nelle scienze umane. Approcci, prospettive e casi di studio*, Mimesis, Milano, 2012.

Analisi e sintesi in tempo reale mediante riconoscimento timbrico

Marco Matteo Markidis

nonoLab, Parma

mm.markidis@gmail.com

José Miguel Fernández

IRCAM, Parigi

jose.miguel.fernandez@ircam.fr

ABSTRACT

Questo articolo descrive *path~*, un external per *Pure Data* [1] che implementa un sistema di analisi e sintesi concatenativa basata su un *corpus* audio, ossia una collezione di file audio. Il suono è acquisito ed analizzato mediante un insieme di descrittori audio estratti in tempo differito. Il processo di sintesi in tempo reale avviene mediante riconoscimento di similarità tra i grani audio dei *corpora sonora* analizzati e il segnale in tempo reale.

La catena di processo avviene, in tempo differito, segmentando il file e generando i grani, eseguendo successivamente un'analisi di ognuno di questi ed estraendone i descrittori che contribuiscono alla creazione di un albero k -dimensionale e una lista di k -primi vicini per ogni grano generato, dove ogni grano è rappresentato da un insieme di punti. La sorgente entra in *path~*, ne vengono estratti i descrittori in tempo reale ed il grano più simile viene trovato nell'albero. Partendo dall'elemento trovato ed usando la sua lista di primi vicini si rende disponibile un insieme di grani per la sintesi.

1. INTRODUZIONE

Attraverso le tecniche di sintesi del suono si possono produrre innumerevoli timbri. Alcuni di questi non prendono ispirazione da suoni esistenti; altri, attraverso una fase di analisi, possono portare alla ricostruzione timbrica di suoni esistenti.

Storicamente la sintesi granulare compare tra le prime applicazioni di produzione e ricostruzione sonora¹. A differenza di altre tecniche di sintesi, ormai cristallizzate nel tempo, la sintesi granulare continua ad avere un percorso di sviluppo lasciando aperte discussioni su varie tecniche, in particolare la sintesi concatenativa ed i mosaici audio. La differenza principale tra queste tecniche e la sintesi granulare è che nella sintesi concatenativa i grani sono legati fra loro solitamente da un'analisi precedentemente fatta su un file audio. Questa relazione può essere basata su differenti metodi ed è solitamente guidata dall'analisi.

L'idea base della sintesi mediante riconoscimento timbrico è di estrarre descrittori da un file audio, di costruire uno

¹ *Concret PH* di I.Xenakis ne è una delle prime applicazioni musicali.

Copyright: ©2016 Marco Matteo Markidis 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.

spazio timbrico e di introdurre una metrica affinché il concetto di distanza diventi significativo. Tramite l'estrazione di descrittori, le similarità tra i suoni sono sottolineate da una vicinanza geometrica nello spazio dei descrittori. Questo spazio è multidimensionale e la sua dimensionalità dipende dal numero e dal tipo di descrittori usati nell'analisi. Per eseguire un corretto riconoscimento, diverse operazioni vengono eseguite in tempo differito: in particolare, la costruzione del *database* multidimensionale fatto con i dati estratti dall'analisi dei descrittori. In tempo reale, un suono entrante è analizzato mediante lo stesso insieme di descrittori usato per l'analisi in tempo differito e la ricerca del primo vicino è condotta sul database. La fase di sintesi si riferisce alla parte finale dell'algoritmo, in cui partendo dall'elemento più simile un insieme di grani fatto dai k elementi primi vicini è disponibile per essere eseguito in un processo di granulazione.

La motivazione principale di questo lavoro è sviluppare un sistema di analisi e sintesi concatenativa. L'idea è stata di sviluppare questo sistema in ambiente *Pure Data*, in maniera da poter dare al compositore o al regista informatico la possibilità di integrare facilmente questo sistema nei propri programmi dedicati al processamento del suono in tempo reale. Lo scopo del progetto inoltre è stato quello di creare uno strumento *open-source* specifico per la musica contemporanea dato che in altri ambienti di lavoro commerciali progetti e lavori simili erano già presenti².

Il contributo principale del lavoro è di provvedere alle funzionalità richieste dalla sintesi concatenativa in un singolo oggetto, senza l'utilizzo di librerie ausiliarie per l'analisi e la sintesi. In questo modo, *path~* rappresenta un sistema compatto che incorpora internamente tutti i passaggi distinti necessari per la sintesi concatenativa. Inoltre, il design interno è pensato per un suo utilizzo in tempo reale. Oltre all'analisi, altre operazioni sono eseguite in tempo differito, come l'ordinamento del database o il calcolo dei primi vicini, necessario per la parte di sintesi dell'algoritmo. La ricerca su un database grande, ossia su file audio lunghi, è eseguita in un tempo ragionevole; una discussione più dettagliata è esposta in seguito.

L'articolo è organizzato come segue: una sezione descrive i lavori collegati, focalizzando l'attenzione sui progetti presenti in Max/MSP e Pd. La successiva descrive i metodi utilizzati in *path~*, con enfasi particolare sull'algoritmo proposto e alla sua implementazione. I risultati, sia artistici che tecnici, vengono presentati nella quarta sezione. Infine verranno discussi i futuri lavori e ricerche.

² Una simile limitazione è sottolineata anche da William Brent in [2].

2. LAVORI COLLEGATI

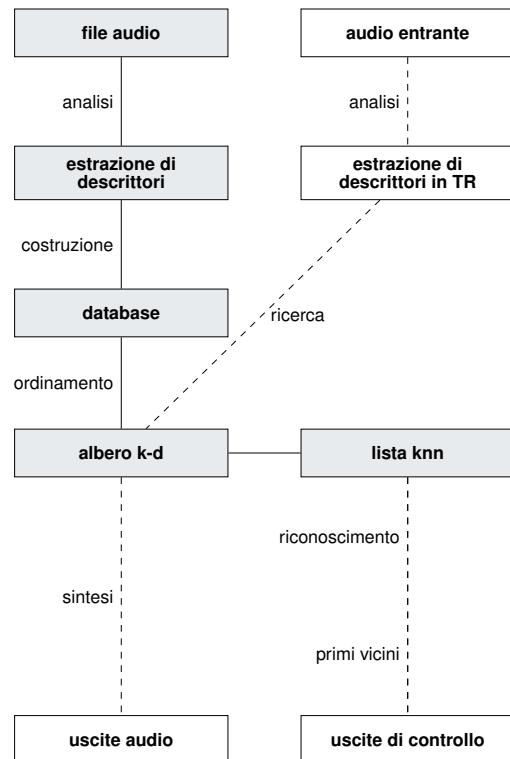
Dopo i lavori pionieristici di Iannis Xenakis, l'attenzione dei compositori e dei ricercatori sulla sintesi granulare e sulla granulazione di suoni campionati è cresciuta al punto che, attualmente, la sintesi granulare è una delle tecniche più usate nella composizione di musica elettronica.

Diversi progetti sono stati sviluppati usando tecniche che involvono sintesi concatenativa a partire dagli anni Duemila. In particolare, la sintesi concatenativa è stata esplorata con *CataRT* [3], un insieme di patch per Max/MSP che utilizza diverse estensioni, come *FTM*, *Gabor* e *MnM*. L'idea è quella di caricare descrittori audio da file SDIF³, o di calcolare descrittori sul momento, e di creare uno spazio dei descrittori che usa grani corti come punti. L'utente può scegliere due descrittori e usando un'interfaccia grafica implementata *CataRT* disegna una proiezione bidimensionale dei grani nello spazio dei descrittori, dove l'utente può navigare ed esplorare lo spazio dei descrittori tramite la posizione del mouse. Attraverso l'interfaccia grafica, la natura espressiva dei grani può essere sottolineata. L'architettura di *CataRT* può essere divisa in cinque parti: analisi, dati, selezione, sintesi ed interfaccia. A parte le utilità grafiche, il sistema utilizza altri strumenti *Max/MSP* come la libreria *Gabor* per la sintesi; inoltre le strutture dati *FTM* vengono utilizzate per la gestione dei dati. Un brano per banjo ed elettronica di Sam Britton è stato composto con *CataRT*.

Un altro progetto *Max/MSP* è stato la realizzazione di Caméléon Kaléidoscope di Marco Antonio Suárez Cifuentes [4] per la *Biennale Musiques en Scène* di Lione nel marzo 2010. Per questa produzione, una combinazione di *FTM*, *Gabor*, *MnM* e librerie *MuBu* è stata implementata. *MuBu* è un contenitore multitraccia che rappresenta uno stream di dati omogenei allineati temporalmente simile allo stream di dati rappresentato dallo standard *SDIF* [5].

Nel mondo *Pd*, alcuni lavori sono stati proposti sul tema dell'analisi timbrica e sulla classificazione sonora. A parte l'oggetto *bonk~* [6], che è un oggetto per la detenzione delle soglie audio, uno dei progetti più importanti è la libreria *timbreID* [2] sviluppata da William Brent. Questa libreria è una collezione di external per l'estrazione di descrittori in tempo reale e in modalità batch. In particolare, oltre alla presenza di descrittori a basso livello, ossia descrittori a singolo valore, alcuni descrittori ad alto livello, ossia una lista di valori, sono stati implementati, come il *mel frequency cepstrum* e il *bark frequency cepstrum*. L'autore mostra che l'utilizzo di descrittori ad alto livello ha un riconoscimento migliore rispetto a quelli a basso livello [7]. Inoltre, mostra anche come l'utilizzo di una combinazione di descrittori, in particolare un descrittore ad alto livello unito a descrittori a basso livello, aumenta la percentuale di riconoscimento a tempo di analisi fissato rispetto all'utilizzo di descrittori singoli. Questa peculiarità ha un grande interesse nella sintesi basata sulla similarità sonora, grazie alla possibilità di un miglior riconoscimento tra il suono

³ Sound Description Interchange Format.



path~:

gli oggetti bianchi e collegati da linee tratteggiate si riferiscono alle operazioni in tempo reale,
gli oggetti grigi e collegati da linee continue si riferiscono a processi in tempo differito.

di input e gli elementi presenti nel database. In aggiunta ai descrittori, un oggetto chiamato *timbreID*, il quale permette una gestione delle informazioni del database, è introdotto.

Infine è stato presentato un recente lavoro audio-video, che implementa un'interfaccia a controllo gestuale [8]. Questo progetto è basato su *Pd*, sulla libreria per computer grafica *GEM* e sulla libreria *timbreID*.

3. METODI

path~ esegue prima l'analisi in tempo differito. Dopo aver caricato i campioni, o in formato di file audio mono dal disco rigido o da un array di *Pd*, *path~* segmenta i corpora audio creando l'insieme di grani. Due tipi di segmentazione sono possibili; il primo è basato su una finestra, ossia data una finestra *path~* taglia i corpora audio ogni data finestra ed esegue l'analisi; tutti i grani avranno quindi la stessa lunghezza. La seconda frammentazione possibile è basata su una soglia, ossia viene eseguita una detenzione degli eventi sui corpora audio; questa frammentazione rispetta la natura gestuale del contenuto del file audio. In questo caso, i grani avranno lunghezze diverse. Successi-

vamente vengono estratti i descrittori da tutti i grani ed i dati raccolti formeranno un database. Abbiamo scelto due strategie per i descrittori.

L'implementazione di default è costituita da un descrittore ad alto livello più un insieme di due descrittori di basso livello, mentre la seconda esclude il descrittore ad alto livello lasciando gli altri. Il descrittore ad alto livello è un mel frequency cepstrum; i descrittori a basso livello sono un centroide spettrale e un descrittore di ampiezza. La computazione dei Coefficienti Mel Frequency Cepstral (MFCC) richiede un banco di filtri passabanda che si sovrappongono equispaziati sulla scala di mel. Un mel è l'unità della scala definita come:

$$Mel(f) = 2595 \times \log_{10}(1 + \frac{f}{700}), \quad (1)$$

dove f è la frequenza in Hz. Infine, per ottenere i MFCC utilizziamo una trasformata discreta del coseno:

$$MFCC_i = \sum_{k=0}^{N-1} X_k \cos[i(k + \frac{1}{2})\frac{\pi}{N}], \quad (2)$$

dove $i = 0, 1, \dots, M - 1$. M è il numero dei coefficienti cepstrali, N il numero dei filtri e X_k è la potenza logaritmica in uscita del k -esimo filtro. L'utente può scegliere la frequenza spaziale di *mel* durante la creazione dell'oggetto. Variando questo parametro, il numero dei filtri risultanti cambierà. Questo significa che la dimensione del nostro spazio timbrico cambierà in accordo a questo parametro. Il valore di default della spaziatura di *mel* è 250 Hz, che risulta in uno spazio timbrico a 14 dimensioni, che è il numero dei MFCC, più 2, che è la dimensionalità risultante del centroide spettrale e del descrittore di ampiezza: ogni grano verrà quindi rappresentato da un insieme di 16 numeri, che saranno le coordinate del grano nello spazio timbrico.

Il centroide spettrale è definito come il centro di massa della magnitudine spettrale:

$$SC = \frac{\sum_{k=0}^{N/2} f(k)|X(k)|}{\sum_{k=0}^{N/2} |X(k)|} \quad (3)$$

con $f(k)$ la frequenza del k -esimo bin.

Il descrittore di ampiezza è di default un'ampiezza in scarso quadratrico medio; può essere cambiato nel picco dell'ampiezza.

L'estrazione dei descrittori è un'operazione che deve essere eseguita su un numero di campioni che corrisponde a una potenza di due. Nel caso della frammentazione basata su finestra, la lunghezza della finestra di analisi è uguale alla finestra di taglio, che è una potenza di due. Nel caso di segmentazione basata sulla soglia, l'estrazione di descrittori avverrà su un numero di campioni arrotondato alla precedente potenza di due rispetto alla lunghezza totale del grano. Questa operazione forma un database, che verrà suc-

cessivamente ordinato in un albero k -dimensionale. Questo ordinamento è utile per l'ottimizzazione della ricerca del primo vicino in tempo reale. Ogni elemento avrà un indice identificativo, che corrisponderà alla posizione del grano nel file audio. Infine *path~* costruisce una lista di k primi vicini, che consiste nei grani disponibili per la sintesi. L'algoritmo per la creazione della lista dei primi vicini è un *Quicksort*.

In tempo reale, *path~* esegue una estrazione di descrittori sulla sorgente di input. Cerca nell'albero k -dimensionale l'elemento che minimizza la distanza euclidea e, accedendo alla sua lista di primi vicini, sintetizza un grano e fornisce le informazioni di controllo. Questa operazione è accurata al campione; infatti l'analisi è fatta sulla richiesta dell'utente ed è indipendente dal *blocksize*. Le informazioni di controllo sono l'indice del grano riconosciuto nel database, la lista con gli indici dei k -primi vicini desiderati, il numero totale di grani e il numero dei grani che stanno suonando. Oltre al contenuto sonoro, le caratteristiche principali dei grani creati sono l'ampiezza, la lunghezza, l'inviluppo e la concatenazione. Queste caratteristiche possono essere controllate direttamente dall'utente tramite specifiche funzioni. Il termine concatenazione si riferisce alla lunghezza del treno di grani, ossia al numero di grani che verranno suonati a ogni richiesta dell'utente. In questo modo, un treno di grani può partire creando tessiture sonore oltre al grano trovato dal riconoscimento dei dati estratti dai descrittori.

```

1 preset @ intro ;
2 threshold = -1.;
3 window = 512;
4 concatenate = 1;
5 amp = 1;
6 hopsize = 100;
7 weight = 1;
8 envelope = 0;
9 # questo e un commento ;
10 endpre
11
12 preset @ sec1 ;
13 window = 2048 ;
14 hopsize = 250 500;
15 concatenate = 2;
16 # envelope = 2 ;
17 endpre
18
19 preset @ sec2 ;
20 threshold = 0.25;
21 concatenate = 3;
22 amp = 0.2 0.4;
23 hopsize = 100;
24 envelope = 2;
25 # questo e un commento ;
26 endpre
27
28 preset @ final ;
29 window = 2048 ;
30 hopsize = 250 500;
31 # envelope = 2 ;
32 endpre
33
34 # token possibili:
35 # window : int potenza di due
36 # threshold: float (0,1)
37 # concatenate: int>0
38 # amp: float>0, un valore = all, due = random ()
39 # hopsize: int>0, un valore = all, due random ()
40 # weight: float (0,1) lista
41 # envelope: int, 0 hann, 1 hamm, 2 blackman, 3 cosine, 4 rectangular (constant envelope)

```

Preset file example.

Inoltre abbiamo implementato un interprete di script pensato per essere un sistema di gestione di preset. In questo modo, l'utente può scrivere uno script con una data sintassi e *path~* lo carica e crea una lista di preset che può essere utilizzata con una funzione dedicata. Opzionalmente, l'utente può specificare per ogni preset un nome identificativo. Oltre all'interprete, un semplice debugger è presente. Inoltre la lista completa dei preset insieme ai nomi identificativi dei preset può essere visualizzata in un editor grafico che può essere richiamato direttamente da *path~*.

path~ supporta una strategia di multithreading. In questo modo, un thread worker esegue l'analisi in tempo differito mentre il thread principale di Pd continua la sua esecuzione. Con questa strategia, l'analisi in tempo differito può essere eseguita senza interrompere il flusso audio.

4. RISULTATI

4.1 Latenza

Nell'audio in tempo reale, uno dei problemi cardine è la latenza, definita come il ritardo temporale tra un input di arrivo e l'output di risposta. Abbiamo quindi studiato il tempo impiegato da *path~* tra l'input dato dall'utente e la risposta dei grani. La nostra stima per il limite di latenza accettabile per il tempo reale è di 6 ms. Sulla macchina di uno degli autori, ossia un Mac BookPro Early 2011 13" 2,7 GHz con Pd Vanilla 0.46.6, su un database di grandezza nell'ordine dei 30K ottenuto usando un file audio lungo dieci minuti con una finestra di analisi di 23.22 ms, la ricerca nel database impiega di media meno di 2 ms, mentre l'analisi in tempo differito impiega di media meno di un minuto.

4.2 Applicazioni nella composizione contemporanea

L'utilizzo di sistemi di analisi in tempo reale per il suono è un concetto chiave nell'interazione tra strumenti acustici ed elettronica. Se negli ultimi anni diversi metodi di analisi del segnale, ad esempio l'estrazione di descrittori audio, sono stati implementati, questi metodi non sono stati esplorati in tutte le loro possibilità.

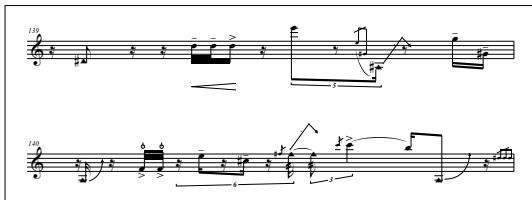


Figure 1: Estratto della partitura di *Dispersion de trajectoires* che abbiamo usato per le nostre applicazioni.

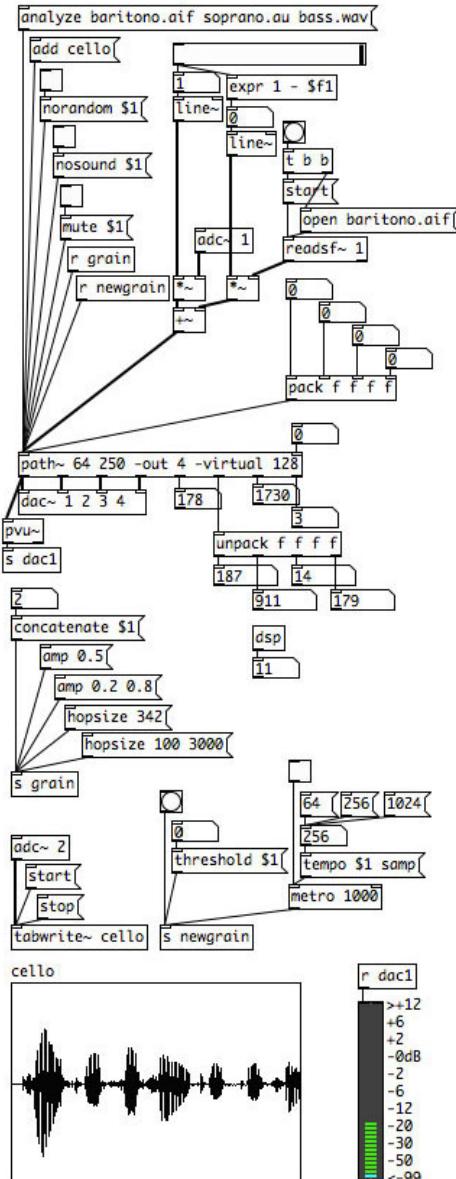


Figure 2: *path~* in un patch d'applicazione.

path~ dà l'opportunità al regista informatico di sperimentare e di creare diversi tipi di materiale elettroacustico in tempo reale o di generare tessiture elettroacustiche. La possibilità di creare ed analizzare diversi file audio creando database come la possibilità di cambiare dinamicamente le caratteristiche dei grani come l'ampiezza, la durata e l'inviluppo attraverso un sistema di preset basato su script dà a *path~* una grande versatilità e flessibilità. Inoltre l'implementazione di una funzione asincrona basata su una soglia, che permette di innescare grani quando sono vicini nello spazio timbrico alla sorgente in ingresso unito alla possibi-

lità di una frammentazione basata sugli eventi permette di creare tessiture concatenate estremamente articolate.

Una grande estensione è offerta dal poter analizzare un array di Pd; in questo modo l'analisi può essere eseguita su materiale appena registrato. Questo permette di registrare suoni da utilizzare per l'analisi in *path~* direttamente durante le prove o i concerti.

Inoltre, *path~* utilizza un sistema di uscite audio vettorializzato. Durante la creazione dell'oggetto, l'utente può specificare il numero di uscite e il numero di canali virtuali, ossia il massimo numero di grani che può essere suonato contemporaneamente. La motivazione principale di questo controllo dinamico è il controllo di configurazioni diverse di diffusori, per esempio durante le sessioni in studio o durante una performance live, e la possibilità di eseguire lo stesso brano musicale su un numero e una configurazione diversa di diffusori.

Infine, *path~* può escludere la parte di sintesi dall'algoritmo lasciando accesso solo alle informazioni di controllo. In questo modo, applicazioni che non fanno uso della sintesi concatenativa ma soltanto della parte di analisi possono essere eseguite con risparmio di CPU.

5. DISCUSSIONE E CONCLUSIONI

Lo scopo primario di questo lavoro è la produzione di brani di musica contemporanea per strumenti acustici ed elettronica. Per raggiungere questo scopo, diverse funzionalità sono pianificate. In particolare, una versione polifonica sarebbe utile per un contesto di ensemble e un controllo maggiore sulla spazializzazione e sulla localizzazione dei singoli grani nello spazio sarebbe desiderabile. Inoltre, un miglioramento dell'editor unito a una maggiore configurazione dei preset è in programma. Un crescente interesse sarà dedicato ad un'implementazione di un'analisi a cluster e sul riconoscimento, cercando di sviluppare alcune strategie specifiche per le tecniche strumentali estese e per utilizzare grani con lunghezze diverse.

Nonostante il lavoro sia ancora in divenire, gli autori stanno usando *path~* per i loro progetti sonori. In particolare, Marco Matteo Markidis sta usando *path~* nel suo duo di improvvisazione radicale Adiabatic Invariants. Attualmente, in *Cattedrali di Sabbia* ne è stato fatto un uso massiccio.

path~ è stato testato su sistemi OS X e Linux usando Pd Vanilla versione stabile 0.46.6. *path~* è un software open source ed è rilasciato sotto licenza GPLv3. È disponibile su Deken, il gestore di externals per Pure Data. L'attuale versione è la 1.5.0 alpha.

6. RINGRAZIAMENTI

Vorremmo ringraziare Stefano Markidis per i suoi consigli e suggerimenti; Martino Traversa e Fondazione Prometeo per il supporto e l'incoraggiamento datoci in questo periodo. Un ringraziamento speciale va infine a Giuseppe Silvi, per il suo aiuto e la sua dedizione, a Daniele Pozzi, senza il quale il nostro interesse in questo lavoro non sarebbe nato, e a Liarss Production, nella persona di Luca Gazzi, che ha dedicato al nostro progetto la sua pazienza e molte ore del suo tempo.

7. BIBLIOGRAFIA

- [1] M. Puckette, "Pure data: another integrated computer music environment," in *Proceedings of the International Computer Music Conference (ICMC)*, 1996.
- [2] W. Brent, "A timbre analysis and classification toolkit for pure data," in *Proceedings of the International Computer Music Conference (ICMC)*, 2009.
- [3] D. Schwarz, G. Beller, B. Verbrugghe, and S. Britton, "Real-time corpus-based concatenative synthesis with catart," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2006.
- [4] N. Schnell, M. A. S. Cifuentes, and J.-P. Lambert, "First steps in relaxed real-time typo-morphological audio analysis/synthesis," in *Proceedings of an International Conference on Sound and Music Computing (SMC)*, 2010.
- [5] N. Schnell, A. Röbel, D. Schwarz, G. Peeters, and R. Borghesi, "Mubu & friends - assembling tools for content based real-time interactive audio processing in max/msp," in *Proceedings of an International Computer Music Conference (ICMC)*, 2009.
- [6] M. Puckette, T. Apel, and D. Zicarelli, "Real-time audio analysis tools for pd and max," in *Proceedings of the International Computer Music Conference (ICMC)*, 1998.
- [7] W. Brent, "Cepstral analysis tools for percussive timbre identification," in *Proceedings of International Pd Conference (PdCon09)*, 2009.
- [8] J. Gossmann and M. Neupert, "Musical interface to audiovisual corpora of arbitrary instruments," in *Proceedings of New Interfaces for Musical Expression (NIME)*, 2014.

La notazione della Dissonanza

Anna Terzaroli

Conservatorio Santa Cecilia - Roma

anna.terzaroli@gmail.com

ABSTRACT

L'articolo prende in esame la realizzazione di uno strumento, nel linguaggio di programmazione Ruby, che, data una partitura o un frammento di partitura, sia in grado di rilevare la dissonanza presente, di quantificarla e di tracciarne una rappresentazione grafica. La dissonanza è qui intesa nell'accezione di Sensory Dissonance. Lo strumento, chiamato d_plot, è costruito guardando all'algoritmo della Sensory Dissonance sviluppato nella libreria open source Essentia 2.0. L'articolo è tratto dalla Tesi di Laurea - Bresson in Discipline Musicali - Musica Elettronica dell'autrice [1].

1. INTRODUZIONE

Il binomio consonanza - dissonanza è sempre stato, nel corso dei secoli, oggetto di discussione e di tentativi di definizione. È stato osservato da diversi punti di vista e considerato da diverse discipline. In aggiunta a ciò, è da sottolineare come la quantità di dissonanza percepita durante l'ascolto non sia determinata, ma cambi in base alle conoscenze e alle preferenze dei fruitori.

Pitagora, nel VI sec. a.C., fu il primo ad interessarsi della dissonanza. Galileo Galilei, nel 1638 circa, condivise e promosse le tesi di Pitagora [2].

Nel XVII secolo, Gottfried Leibniz, il coinventore del calcolo infinitesimale, sosteneva che tra la consonanza e il Bello sussistesse una correlazione e che gli uomini calcolassero inconsciamente i rapporti frequenziali degli intervalli musicali [3]; nel XVIII secolo, Eulero suggeriva che alcuni intervalli musicali comportavano un "senso di benessere".

Nel XIX secolo, Hermann von Helmholtz riprendeva e formalizzava quanto sostenuto da Pitagora e paragonava l'orecchio a un analizzatore di spettro, assumendo che tutti gli ascoltatori giudicavano e quantificavano la dissonanza allo stesso modo [4, 5].

Nel XX secolo, si assiste ad un grande sviluppo della psicoacustica.

Gli studi di Békésy [6] sulla mappa cocleare aprono un nuovo scenario nel quale si muoverà Harvey Fletcher, dimostrando la stretta corrispondenza tra le distanze lungo la membrana basilare e le regioni del mascheramento.

Viene definito il concetto di "banda critica".

Copyright: ©2016 Anna Terzaroli 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.

Donald D. Greenwood [7] collega gli studi di Fletcher [8] a quelli di Békésy. Inoltre, dimostra come gli studi di Mayer siano concordi con la definizione della banda critica.

Mayer, alla fine del 1800, aveva condotto esperimenti e raccolto dati circa la misura del più piccolo intervallo frequenziale, appena prima che gli ascoltatori percepiscano la roughness.

Plomp e Levelt [9] approfondiscono il lavoro di Greenwood, ponendo in relazione la percezione della "Sensory Dissonance" con il concetto di banda critica.

Molti studiosi di psicoacustica hanno sostenuto che l'orecchio giochi il ruolo più importante nella percezione della dissonanza. Essendo questo organo comune a tutti gli uomini, di qualsiasi cultura essi siano, la percezione della dissonanza sarebbe indipendente dalla cultura di appartenenza. Molti etnomusicologi, al contrario, sostengono la dipendenza della percezione della dissonanza dalla cultura degli ascoltatori.

Nel corso del 1900, sono stati realizzati vari esperimenti a tal riguardo, che in parte confermano e in parte smentiscono la correlazione cultura - percezione della dissonanza.

Nel secolo scorso, gli esperimenti sulla determinazione della banda critica sono stati numerosi e hanno portato, tra gli altri risultati, all'adozione di una particolare scala di misurazione della banda critica, chiamata Scala dei Bark [10], che comprende la porzione frequenziale dello spettro udibile dall'uomo.

Sono state proposte molteplici teorie, circa la dissonanza.

Queste possono essere suddivise in alcuni gruppi principali: la teoria acustica, sostenuta da Pitagora; la teoria psicoacustica, che considera il ruolo dell'orecchio e della membrana basilare nella percezione della dissonanza. Inoltre, le teorie cognitive e le teorie che prendono in considerazione il livello di inculturazione degli ascoltatori.

Infine, anche le definizioni di dissonanza sono state e sono ancora ad oggi, molte e diverse tra loro.

"Roughness", traducibile in italiano con rugosità, è quella sensazione di "suono aspro" che è prodotta da due sinusoidi poco differenti tra loro dal punto di vista frequenziale, dopo che già sia stato udito il fenomeno dei battimenti e prima che le due sinusoidi possano essere percepiti come suoni completamente distinti l'uno dall'altro.

Nei contesti in cui la dissonanza viene denominata "dissonance", si guarda alla dissonanza come fenomeno culturale, ovvero a quel tipo di dissonanza che è riconosciuta come tale in alcuni mondi musicali, ma che non lo è in altri.

Quindi, in questo caso, si precinde dal suono inteso come fenomeno fisico e matematico.

La denominazione "tonal dissonance" è contrapposta a "tonal consonance". La "consonanza tonale" è definita come "la singolare natura propria di quegli intervalli di tono, caratterizzati da rapporti frequenziali pari a piccoli numeri interi" [9].

Infine, "sensory dissonance" è un termine coniato da Helmholtz nel 1877, per indicare la dissonanza percepita all'ascolto, connessa ai "sensory factors", che Helmholtz per primo aveva studiato.

2. I MODELLI DI SENSORY DISSONANCE

Al fine di scegliere il modello più adatto di Sensory Dissonance (SD), ovvero il più affine, al nuovo strumento analitico da costruire - d_plot -, da studiare in modo approfondito, sono state ricercate e considerate le diverse implementazioni SD presenti in vari ambienti di lavoro.

E' stata esaminata, ad esempio, la SD di Essentia 2.0, libreria open source scritta in C++, creata da MTG (Music Technology Group), utile all'analisi dell'informazione sonora [11].

La Sensory Dissonance è qui trattata come un descrittore MIR (Music Information Retrieval) e un descrittore dello spettro [12].

L'algoritmo SD sviluppato in questo ambiente di lavoro analizza un segnale audio fornito in ingresso e ne restituisce i picchi spettrali. Quindi viene effettuata una misura della roughness basata sui picchi spettrali già ottenuti, infine la dissonanza totale è stimata come la somma dei valori normalizzati della dissonanza di ogni coppia di picchi spettrali. I valori di dissonanza relativi alle coppie dei picchi sono calcolati in base alla curva della dissonanza ottenuta da Plomp e Levelt.

Vengono passati in rassegna anche altri strumenti analitici, per osservare se e come affrontino il trattamento della SD.

Essendo d_plot scritto nel linguaggio di programmazione Ruby, ci si è dapprima orientati verso i software di analisi del suono realizzati nel medesimo linguaggio o per i quali fosse già in uso il binding in Ruby. E' il caso di MARSYAS, sviluppato da George Tzanetakis [13, 14], tuttavia non è qui previsto un algoritmo relativo alla SD.

Poiché è relativamente facile produrre il binding in Ruby avendo una libreria in Java, in seconda istanza, ci si è orientati verso i software analitici realizzati in Java. E' questo il caso di jAudio/jMIR [15] e libXtract, ma nemmeno in questi strumenti è affrontata la SD.

Sono stati esclusi, come modelli, ai fini di questa ricerca, gli algoritmi di PureData e di MaxMSP, nei quali la dissonanza viene trattata come roughness.

In SuperCollider, invece, è presente l'implementazione di un algoritmo di Sensory Dissonance [16]. Dal confronto tra la funzione per il calcolo della SD utilizzata in SuperCollider e quella utilizzata in Essentia, emerge che le due sono diverse e la prima non è pienamente corretta. Si sceglie pertanto, come modello di riferimento per d_plot, l'algoritmo della Sensory Dissonance così come sviluppato in Essentia.

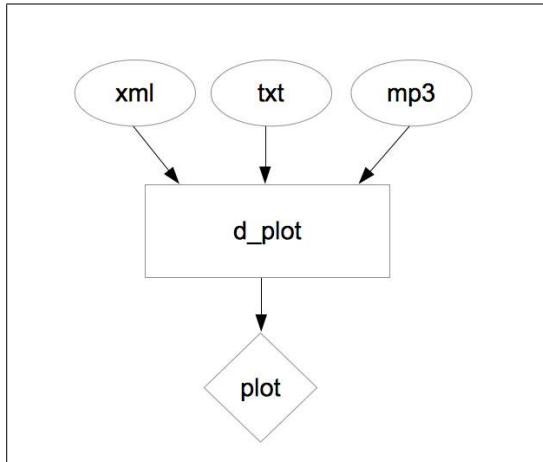


Figure 1. Meccanismo di funzionamento di d_plot.

3. D_PLOT: ARCHITETTURA E IMPLEMENTAZIONE

3.1 Architettura

In Figura 1 è rappresentato il funzionamento dello strumento d_plot, innanzitutto si noti come questo riceva in ingresso tre diversi files: un primo file in formato xml, un secondo file in formato txt e un terzo file in formato mp3 (oppure un file in un altro formato, comunque un file audio).

d_plot restituisce in uscita un file indicato, nello schema, con la parola plot, che indica una rappresentazione, tramite grafico, della dissonanza rilevata dallo strumento stesso, sulla base dei dati di input.

Il formato xml è la sigla di Extensible Markup Language, cioè Linguaggio di Markup Estensibile. Infatti XML permette la rappresentazione strutturale e semantica delle informazioni. I file in questo formato contengono informazioni e dati, scritti in testo ASCII standard. E', inoltre, un linguaggio estensibile poiché i suoi tags (a differenza, ad esempio, di quelli del linguaggio HTML) non sono limitati. XML è, infine, un linguaggio marcatore di tipo descrittivo (e non procedurale): la leggibilità delle informazioni è più importante rispetto alla loro rappresentazione e leggibilità e rappresentazione vengono mantenute separate.

Nel caso specifico, qui esaminato, il linguaggio utilizzato è il MusicXML, lo standard per la condivisione di partiture, che viene riconosciuto da vari software di tipo notazionale e non solo. MusicXML è una codifica di tipo XML con cui è possibile rappresentare la notazione classica occidentale. Sviluppata da Recordare LLC, è rilasciata come formato libero, senza diritto d'autore.

Grazie al file in formato xml, d_plot riceve in input la partitura.

Il file in formato txt fornisce a d_plot un elenco testuale delle battute presenti nella partitura, con la relativa indicazione, per ciascuna di queste, dell'istante di inizio e di termine.

Il file in formato audio permette a d_plot di poter analizzare un segnale audio in tempo reale. Questo punto è importante alla luce del "tipo di dissonanza" che si è scelto di prendere in considerazione, la Sensory Dissonance, che, a differenza della dissonanza intesa come roughness, dissonance, tonal dissonance, si basa principalmente sulla percezione della musica suonata e ascoltata.

Illustrata brevemente la sezione input dello strumento, si passa ora a trattare dell'output.

Questo è costituito da un grafico della Sensory Dissonance rilevata dal segnale audio e rappresentabile, battuta per battuta, grazie al file MusicXML e al file di testo. Considerando una notazione musicale che non preveda l'utilizzo delle battute, si potrà implementare un ulteriore sviluppo di d_plot, per cui la Sensory Dissonance sarà rappresentata riga per riga.

Lo strumento d_plot è costituito da tre lettori (uno per ogni tipo di file in input) e un aggregatore che riunisce e connette tra loro i dati provenienti dall'elaborazione dei dati in ingresso.

3.2 Implementazione

Dall'istante in cui d_plot riceve i files di input, l'implementazione dello strumento compie i seguenti passaggi: viene caricato il file audio e trasformato in mono (se necessario), quindi viene suddiviso in frames e successivamente viene applicata una finestra Hamming. Viene calcolata la FFT di ogni frame e vengono calcolati i picchi spettrali. Queste operazioni restituiscono due vettori: uno con i valori delle frequenze e l'altro con i valori delle magnitudini relative alle frequenze. Questi due vettori confluiscono nell'algoritmo della dissonanza.

L'algoritmo della dissonanza restituisce un solo valore di dissonanza per ogni frame.

Poiché nei files MusicXML la larghezza delle battute è calcolata in pixel (con un fattore di riscalamento), nell'aggregatore viene operata la conversione da pixel a millimetri. Nel file txt le battute sono descritte dalla loro posizione nel tempo, in ogni file audio.

L'algoritmo dell'aggregatore considera ogni battuta e, tra tutti i valori di dissonanza, sceglie solamente quelli propri della battuta di volta in volta presa in esame.

Viene operata, inoltre, una conversione tempo-spazio, ovvero ogni valore di tempo è posizionato in un punto nello spazio appartenente alla larghezza di ogni battuta. L'asse del tempo non è lo stesso per ogni battuta, quindi il rapporto spazio/tempo non è costante.

I valori di dissonanza posizionati nello spazio confluiscono nel software pic, tramite un driver di uscita. Pic disegna le curve nello spazio associato a ogni battuta. Queste curve sono funzioni spline [17], in quanto questo tipo di funzioni, costituite da polinomi collegati tra loro, garantiscono una continuità nel tempo, almeno fino ad un certo ordine di derivata.

Pic è un preprocessore del sistema Troff [18] tramite cui è possibile ottenere immagini, fornendo una descrizione testuale. Il programma pic è stato scritto da Brian Kernighan, come complemento al software Troff di Joseph Ossanna ed è stato successivamente riscritto, con miglio-

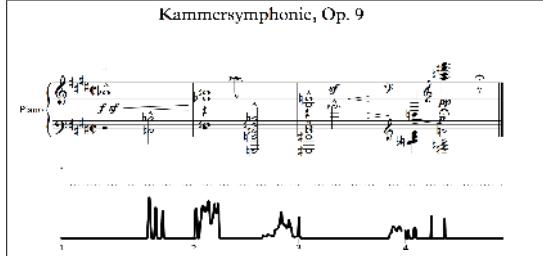


Figure 2. *Chamber Symphony No.1, Op.9* di Arnold Schoenberg.

ramenti sostanziali. Il linguaggio si ispira ai più vecchi linguaggi grafici Ideal e Grap.

Il codice completo utilizzato per l'architettura e per l'implementazione di d_plot è disponibile al seguente indirizzo: www.github.com/aterza/d_plot.

4. D_PLOT: APPLICAZIONI E RISULTATI

In un esempio di funzione grafica prodotta dallo strumento d_plot, ottenuta in base all'audio (e ai file .xml e .txt, a questo correlati) ricevuto in ingresso, i valori delle ascisse del grafico rappresentano i numeri delle battute, i valori delle ordinate indicano il grado di dissonanza. La funzione ottenuta è continua nel tempo. La rappresentazione grafica della dissonanza può essere posta sotto il segnale musicale cui si riferisce, il segnale musicale è scritto secondo il sistema di notazione occidentale, che si avvale, nel caso specifico, dell'uso delle battute.

La curva ottenuta fornisce indicazioni riguardanti non solamente il grado di dissonanza, ma permette una lettura più approfondita dell'intera composizione.

Infatti, nella curva potrebbero notarsi delle ripetizioni dell'intera funzione, che quindi sarebbe periodica, oppure di alcune sue parti. Quest'ultimo caso significherebbe che ci sono delle parti del brano analizzato che sono uguali o molto simili per ciò che riguarda il fattore dissonanza.

Potrebbe essere interessante confrontare le sezioni simili in base al parametro dissonanza e osservare se siano diverse o uguali anche in base ad altri parametri.

Si potrebbe riscontrare, ad esempio, una spiccata similitudine dal punto di vista del contenuto di dissonanza, contrapposta ad una pressocché totale disuguaglianza di ritmo, metro oppure di utilizzo dei registri ecc. Si potrebbe pure riscontrare il caso inverso.

Contrariamente, la similitudine o la non-similitudine del grado di dissonanza potrebbero essere coerenti con i valori degli altri parametri musicali.

La conoscenza di queste infomazioni può, a sua volta, fornire notizie sulla modalità compositiva dell'autore del brano analizzato mediante d_plot.

5. CONCLUSIONI

Il materiale musicale scelto per la prima applicazione pratica dello strumento d_plot è una selezione di battute tratte



Figure 3. *Étude Op.8, No.12* di Aleksandr Skrjabin (parte 1).

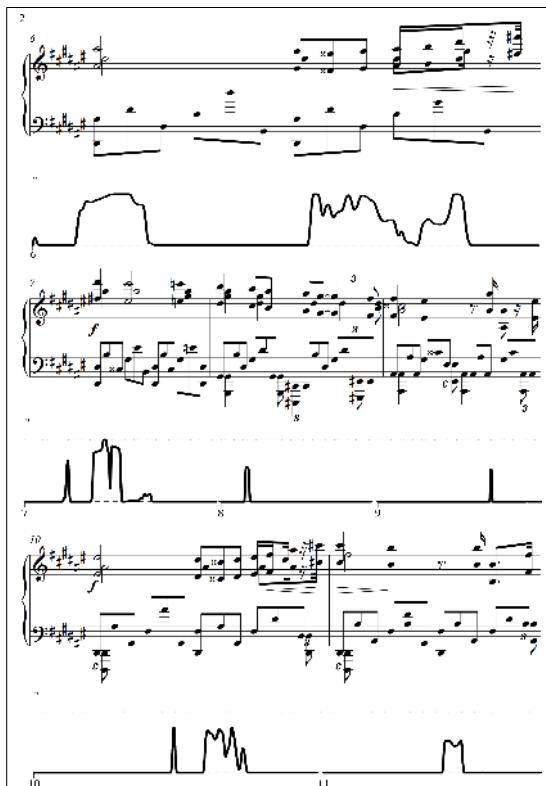


Figure 4. *Étude Op.8, No.12* di Aleksandr Skrjabin (parte 2).

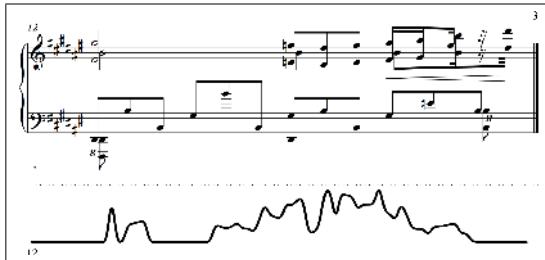


Figure 5. Étude Op.8, No.12 di Aleksandr Skrjabin (parte 3).

dalla *Chamber Symphony* No.1, Op.9 di Arnold Schoenberg.

Eseguito per la prima volta nel 1907, questo lavoro rappresenta un momento di transizione, nella storia della musica. Pur muovendosi ancora entro i vincoli del Romanticismo, contemporaneamente cerca di liberarsi dalla tradizione. Fin dalle prime battute, la tonalità viene attenuata con forza da un movimento melodico privo di semitonni e da una successione di intervalli di quarta.

La Figura 2 mostra l'arrangiamento per pianoforte del brano di A. Schoenberg ed il plot della dissonanza riferito al brano eseguito dall'organico originariamente previsto dal compositore. Si noti un'alternanza abbastanza precisa e puntuale di dissonanze e consonanze, nelle battute di apertura della *Chamber Symphony*. E' interessante notare come lo stesso livello di consonanza sia ottenuto dall'uso di note o accordi che non sono sempre gli stessi, sebbene producano un risultato molto simile. Le stesse affermazioni sono valide per i picchi di dissonanza. Le informazioni fornite da d_plot rappresentano un'aggiunta a quanto riportato dalla partitura tradizionale.

Un'altra selezione di frammenti che verrà sottoposta all'analisi della dissonanza operata da d_plot proviene dall'Étude Op.8, No.12 di Aleksandr Skrjabin. A differenza del materiale schoenberghiano, qui l'armonia tonale è pienamente rispettata.

Nelle Figure 3, 4, 5 è mostrata la partitura dell'opera di A. Skrjabin ed il relativo grafico della dissonanza. Si noti come la dissonanza aumenti nei punti cadenzali: questa è un'informazione coerente con quanto ci si poteva attendere, conoscendo lo stile di riferimento del lavoro musicale. Tuttavia, il plot restituisce le informazioni in modo più diretto e semplice, tramite una rappresentazione visiva. Si noti, inoltre, l'importanza dell'informazione, non fornita dalla partitura tradizionale, relativa alla quantità della dissonanza presente.

6. REFERENCES

- [1] A. Terzaroli, "Dissonance notation. Towards a notion of augmented score," Tesi di Laurea - Biennio Musica Elettronica. 2016.
- [2] G. Galilei, *Discorsi e dimostrazioni intorno a due nuove scienze*. Verona: Cierre - Grafica, 2011 (1638).
- [3] E. G. Bugg, "A criticism of Leibniz's Theory of Consonance," *The Journal of Aesthetics and Art Criticism*, vol. 21, No.4, pp. 467–472, 1963.
- [4] H. L. F. Helmholtz, *On the sensations of tone as a physiological basis for the theory of music*. New York: Dover, 1885.
- [5] H. L. F. Helmholtz, *Die Lehre von den Tonempfindungen, als Physiologische Grundlage für die Theorie der Musik*. Nachdruck: Minerva-Verlag, Frankfurt/Main, 1981 (Vieweg, Braunschweig, 1863).
- [6] G. Békésy, *Experiments in Hearing*. New York: McGraw Hill, 1960.
- [7] D. Greenwood, "Critical bandwidth and the frequency coordinates of the basilar membrane," *Journal of the Acoustical Society of America*, vol. 33, No.10, pp. 1344–1356, 1961b.
- [8] H. Fletcher, *Effects of Filtering and Masking*. Van Nostrand, 1953.
- [9] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *Journal of the Acoustical Society of America*, vol. 38, pp. 548–560, 1965.
- [10] web. https://ccrma.stanford.edu/jos/bbt/Bark_Frequency_Scale.html, last viewed 28-04-16.
- [11] D. Bogdanov, N. Wack, E. Gómez, and oth., "ESSENTIA: an Open-Source Library for Sound and Music Analysis," *Proceedings of the 21st ACM international conference on Multimedia*, pp. 855–858, 2013.
- [12] web. http://essentia.upf.edu/documentation/algorithms_overview.html, last viewed 28-04-16.
- [13] G. Tzanetakis and P. Cook, "Marsyas: A framework for audio analysis," *Organised Sound*, vol. 48, pp. 169 – 175, 2000.
- [14] G. Tzanetakis and P. Cook, "Marsyas3d: A prototype audio-browser editor at large scale using immersion sive visual and audio display," *Proceedings of the international conference on auditory display (ICAD)*. Helsinki, pp. 8–37, 2001.
- [15] web. <http://jmjr.sourceforge.net>, last viewed 28-04-16.
- [16] web. <http://github.com/supercollider/sc3-plugins/blob/93673392968441e36842a6a55796c5ddebd774/source/SCMIRUGens/SensoryDissonance.cpp>, last viewed 28-04-16.
- [17] web. http://www.math.ucla.edu/~baker/149.1.02w/handouts/dd_splines.pdf, last viewed 28-04-16.
- [18] B. W. Kernighan, "Pic - a graphics language for typesetting (revised user manual)," *Bell Labs Computing Science Technical Report*, vol. 116, 1991.

COMPOSING EXPLORATION: A MULTI-AGENT APPROACH TO CORPUS-BASED CONCATENATIVE SYNTHESIS

Daniele Pozzi

Institut für Elektronische Musik und Akustik, Graz
daniele.pozzi@student.kug.ac.at

ABSTRACT

This article discusses an agent-based approach to corpus-based concatenative synthesis and audio mosaicing. A complex composition and performance environment will be described, which makes use of a group of abstract agents of varying complexity in order to explore a two dimensional space of descriptors, populated by sound units. The notion of morphological space, commonly adopted in CBCS, is taken as ground to build a virtual sonic environment, which can be sensed and explored in its variety by a number of software entities, able to interact with one another and with human beings. The result is a virtual ecosystem, characterized by a static part - the spatial organization - and a living part - the agents organization -, capable of generating living-like and emergent sound textures and reacting on external stimuli.

1. INTRODUCTION

In nowadays music-related creative processes and composition sound cataloguing is a well affirmed praxis. Audio descriptors are widely employed in a variety of different tasks, playing the role of essential elements in activities such as computer aided orchestration, audio mosaicing or corpus-based concatenative synthesis. In the most common approach to corpus-based concatenative synthesis, CataRT [1] being the most diffused example, the unit selection takes places as a content-based similarity match based on a weighted Euclidean distance between the audio descriptors of the database units, and the synthesis target. The simplicity of this method explains the relative success of CBCS for interactive descriptor-based granular synthesis [2]: especially when combined with a graphical interface, the interaction with the system is really intuitive and appealing for the user that approaches this particular kind of granular synthesis. The main application of such described systems is an interactive explorative process, which may take place in different ways, but typically finds in the mouse cursor its primary interface for the final user, who is able to define his desired target pointing at a precise area of the screen. The system described in this article seeks for a different approach to content-based environments, aiming to create specific conditions that enable sound morphologies

to emerge as primary and main element of the interactive explorative synthesis process.

2. THE AGENT

In order to pursue this approach, an abstract system, composed of a number of software entities, was formalized. These should exhibit some peculiar features: first of all, the capability of sensing a two dimensional space, as well as all the other entities populating it. Secondly, these should be able to react to such stimuli, according to specific rules, in order to let emerge clear high-level behaviors, programmable by the composer in order to obtain specific - or unpredictable - sound results. This strategy enables to define the qualities of the sonic result both at the micro-level of the single agent and at the macro-level of the overall behavior. The two are strictly connected to each other, resulting in a sort of self-regulating ecosystem capable of adapting to the surrounding virtual ambient and to react to external stimuli, such as sounds or inputs from a human being. This approach is in line with the multi-agent systems paradigm, which recently became quite popular in the field of computer music composition and musical informatics [3, 4], and it's here specifically applied to corpus-based concatenative synthesis.

2.1 The algorithm

Core of the system is the famous Reynold's *Boids* flocking algorithm [5], whereby an agent's position and velocity is dictated by a set of simple rules. Since this approach is well known, it won't be discussed in great detail here. Simply put, Reynolds demonstrated that complex emergent behavior, such as that observed in flocks of birds, schools of fish, swarms of insects, etc., could be simulated using three basic rules: *separation*, or the tendency of individuals (boids using Reynolds terminology) in a flock to avoid their close neighbours; *alignment*, or the tendency of individuals to match the direction of their close neighbours; and *cohesion*, or the tendency of individuals to steer towards the average position of their close neighbours. This model has been extended with other arbitrary rules, for instance goal setting (i.e. attractors), and boundary setting (i.e. keeping the flock within a delimited space) [9]. The correlation between Reynold's algorithm and CBCS is quite immediate, because both share a strong relation with the concept of space.

Copyright: ©2016 Daniele Pozzi et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2.2 Swarms and Music

The adaptation of flocking simulations for the creation and control of algorithmic music prove to possess a great appeal for musicians and numerous example of swarm-based music have been created [10, 11, 12, 13, 14, 15]. Even in their simplest form, dating back to the seminal work by Craig Reynolds [5], these self-organized multi-agent systems can give rise to a range of emergent phenomena, which can be efficiently exploited in musical creation.

Tim Blackwell and Michael Young already applied the swarm simulation paradigm to generative music and computer music improvisation [10], proposing an application of particle swarms to sound synthesis. In [11] they presented a new agent-based approach to granulation, coupling the movements of a virtual swarm to the synthesis parameters in order to produce musically interesting and novel timbres.

Working on the same concepts Tim O'Brien developed ChocK¹, a swarm simulation for algorithmic musical composition. This program, written fully in Chuck, implements Reynolds' *boids* algorithm with many of the musical applications described in [10]. However, instead of being a simulacrum of a musical improviser responding to a human performer, ChocK is built to input a meta-score based on time-varying swarm attractor elements. The output is thus the swarm simulation's rather unpredictable interpretation of the human composer's instructions.

Two important research projects entitled ISO (Interactive Swarm Orchestra) and ISS (Interactive Swarm Space) have been running for several years at the Institute for Computer Music and Sound Technology (ICST) of the Zurich University of the Arts and provided an ideal context to study the issues related to the application of swarm simulations in music. In [16] they stated that the almost unlimited number of possibilities for model customization and transformation into music constitutes some of the main challenges of simulation-based computer music. A significant part of the effort in the creation of a swarm based piece needs to be invested into establishing a meaningful mapping between the simulation's properties and the musical result. This happens because flocking simulations were obviously not designed with any musical application in mind. They suggest that by shifting the focus of the musical creation away from the mapping issue towards the design of the flocking simulation itself, the mapping challenge is diminished and the diversity of the musical possibilities is increased.

This is the exact approach adopted in the development of the system described in this article. The Reynolds' multi-agent system, being elegant in its simplicity, offers an almost unlimited number of possibilities in customisation and transformation, enabling to build different emergent models for the exploration of euclidean spaces. The flocking simulation was indeed designed with precise musical results in mind, such as spectral contractions, drifts, rarefactions, rhythmical patterns.

¹ <http://ccrma.stanford.edu/~tsob/220a/fb.html>

3. TECHNICAL OUTLINE

The system is entirely realized in SuperCollider using a variety of libraries and tools.

- Nick Collins' SCMIR library².
- Dan Stowell's KDTree quark³.
- Fredrik Olofsson's RedUniverse quark⁴.

3.1 Analysis

Following a typical CBCS approach [1, 2, 6], a first part of the process consists in segmenting a sound corpus in small units, which are then analyzed in order to be spectrally disposed within a two-dimensional canvas, according to their perceptual features. This is a preliminary non real-time step, that enables the user to build specific sonic morphologies similar to the one in Figure 1.

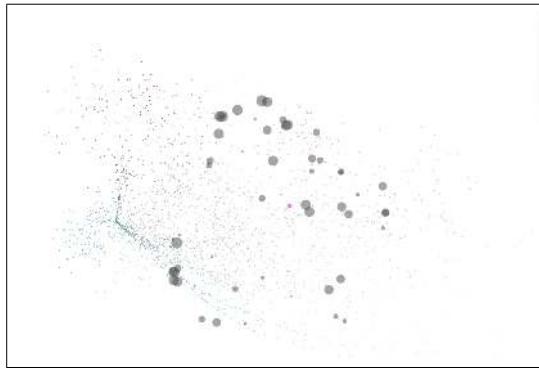


Figure 1: Boids (in grey) exploring a pool of sound units (colored dots) disposed according to their perceptual features.

According to the chosen features and mode of segmentation, it is possible to obtain many different dispositions of samples. The library allows more than twenty audio features to be extracted and automatic, beats-based or onsets-based segmentation [6]. The so-obtained morphological space is taken as ground to build a virtual sonic environment, which can be sensed and explored in its variety by the agents of the flock.

3.2 Flocking

A group of agents (Figure 1, in grey) is then thrown inside this two dimensional environment and flocks according to the defined rules. Through the nearest-neighbors UGen, which searches for the nearest n points in a dataset by Euclidean distance, each agent examines the space, seeking for the n nearest sound units, according to its own position. Then, at a given rate, one single sample is randomly chosen among the nearest to be played back by a granular synthesizer. Each agent is indeed coupled to a

² <http://composerprogrammer.com>

³ <http://www.mclab.co.uk>

⁴ <http://www.fredrikolofsson.com>

particular instance of a granular SynthDef and, according to its personal behavior, controls many parameters of the synthesis in real-time. The specific behavior of the flock can be precisely shaped both at the micro-level of its singularities and at the macro-level of the entire organism, in order to achieve peculiar sound events and musical transitions: considering the entire canvas as a big spectral palette, full filled with all the desired samples, the composer can then craft large scale spectral gestures or static textures.

3.3 Relationships and Interactions

Each boid executes the three tasks of *cohesion*, *separation* and *alignment* according to the position of its neighbors. Neighbors are boids that reside in the perceptual area of a given boid at a given time. In the code example below each agent looks for eventual neighbors within its own circular area, defined by the radius *dist*.

```
boids . do{|o, i|
  boids . do{|obj|
    if (((o . at(0) - obj . at(0)) . squared +
      (o . at(1) - obj . at(1)) . squared) . sqrt
      < dist && (o != obj),
      { neighbors . add(obj)} );
  };
};
```

Different perceptual radius are defined for the three tasks. Three different pool of neighbors are thus defined for each single agent. By modifying the three radii it is possible to deeply influence the behavior of the flock at a high level, inducing different states of contraction, expansion, drifting, aggregation, splitting and so on. A fourth radius defines a further pool, whose size controls the amplitude of the corresponding granulator: the larger the neighbors pool is, the louder the synth will sound. Consequently, the overall resulting sound spectrum will be more prominently connoted by those perceptual features that correspond to the most densely populated areas of the morphological space.

By calculating the average value of the size of this fourth pool over the entire flock, the system can roughly understand whether agents find themselves in a contracted (high value) or in a loose (low value) situation. Confronting the actual pool size of a single boid with the average one, it can estimate whether that specific agent is part of a big or a small group. Consequently, it is possible to program different steering reactions at the level of the single boid, for regulating the distribution of aggregations among the space.

Each agent can also sense the underlying sound units distribution, meaning that it can distinguish between more dense or more vacuum areas and take decisions accordingly. If a boid finds itself in a portion of the canvas containing a lot of samples, it can slow down, to explore and zoom in on this particular region, or speed up to leave it as soon as possible. At a high level, this is a useful tool for the composer, who can choose whether to explore more the singularities and exceptions of his materials or to focus on the samples exhibiting the most common features.

Furthermore, taking advantage of the physical models the RedUniverse quark offers, different mappings and ways of interaction can be built. For instance, by placing an attractive point at the coordinates defined by the spectral analysis of a specific sound input agents will converge toward this portion portion of the space, giving life to a contracting audio mosaicing process.

Even the quality of the sound output can influence the boids' behavior. The system constantly performs a spectral analysis over the sounds produced: when a specific threshold value is reached it modifies the agents' parameters, thus provoking a change in their movement and in the overall timbre. Taking advantage of these simple dynamic reactions, it is possible to build basic musical transitions: processes of bleaching, darkening, embittering, etc., of the spectrum, which shorten the distance between *formalism* and *materism* [8], meaning that the formation of the timbre becomes the form itself.

3.4 Spatial Distribution

The agents' movement can be applied to efficiently distribute granular processes over large scale multichannel systems. Assuming the center of the canvas as the center of the physical room, every granular synthesizer can be assigned to a different speaker configuration, according to the relative position of the corresponding boid at a specific time. Complex particle movements emerge, obtaining a kind of spatial swarm granulation [9], which furthermore reveals the actual disposition of the sound units being explored. The overall result is then a sharp movement of spectral textures in space. A first test has been made in CU-BE at IEM (Graz) employing a 24 channels system with satisfying timbral disposition and source recognition over different areas of the room.

4. COMPOSING EXPLORATION

From an artistic point of view, the aesthetic issues the composer dealing with the system has to face are roughly two: a first part of the work consists in selecting sound materials and organize them in a meaningful temporal order, that is, so to say, composing the materials. This is an important step because, in the context of granular synthesis, the nature of the materials greatly affects the nature of the output [7], while their temporal and timbral disposition ensures smooth and fluid transitions during re-concatenations [1]. But the crucial part of the overall process is composing relations between this pre-composed environment and the virtual agents exploring it.

The choice of a multi agent-based approach entails many aesthetic implications. The most obvious is the renounce to a certain degree of control over the creative process [12]: even if it is possible to minutely craft different reactions at the level of each single boid, the resulting complexity of such a system, whose behavior is defined by the sum of a high number of small interactions among its single elements, inevitably introduces a certain unpredictability and pronounced emergent properties. By creating conditions and shaping the agents' behavior, the composer creates

a micro-level system or procedure such that a higher-level *Gestalt* can emerge and develop in time, following a typical approach of the micro-composition paradigm [7, 8] which leads to the formation of both timbre and form [11].

4.1 Texture Design

This section depicts an example of flocking design that leads to the creation of a fifty seconds long spectral contraction: starting from a punctilistic and chaotic situation, it gradually comes to a stable and muffled granular texture.

The recording of an execution of *Pairs* by Christian Wolff [17] is segmented adopting a window size of 0.242'', analyzed and disposed as shown in Figure 2. Being this version of the piece five minutes and fifteen seconds long, we obtain a pool of 1301 sound units.

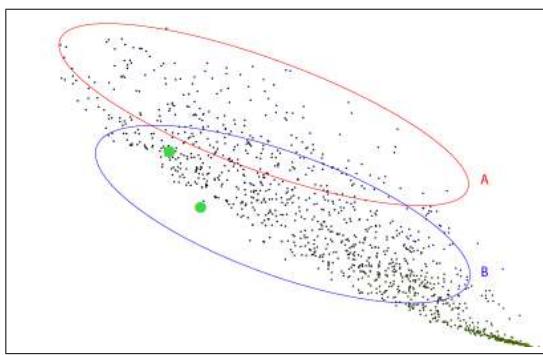


Figure 2: Analysis of *Pairs* by Christian Wolff.
Materials: flute, piano, celesta, percussions.
X axis: spectral centroid, brighter samples on the right side.
Y axis: spectral flatness, noisier samples at the bottom.

This kind of classification typically divides materials into two regions, highlighted in Figure 2: area A mostly contains transients and clear sustained parts of the notes, while B is characterized by more muffled and blurred samples. Units in the right-lower portion are mainly pauses - silence.

Remembering that the more neighbors a boid has, the louder its granulation will be (see Section 3.3), we can sketch a rhapsodic beginning by setting a small radius for the search of neighbors and increase the separation force between boids to avoid local concentrations. At this stage of the process boids find themselves in the situation A in Figure 4. They constantly try to avoid each other, rarely getting close enough to trigger the sound synthesis. This gives life to a punctilistic and chaotic situation, triggering samples from any portion of the canvas every now and then (first fifteen seconds in the spectrogram).

By placing two attractors in the lower-left quarter of the space (green dots in Figure 2) boids slowly start to converge towards this area (situation B, Figure 4). Transients and sustained sounds from the upper portion disappear, while a quite dark and blurred texture starts to emerge, as boids accumulate around the attractors.

Situation C represents the end of the contraction process. Boids find their stability, flocking around the attractors in the blurry portion of the space, but always trying to

keep a minimum distance between each other. This behavior generates a stable yet shimmering texture, visible in the last part of the spectrogram.

By modifying the proportion of forces it is possible to design different kinds of contraction simulations. For example, if we assign a lower mass value to the attractors, these will exhibit a weaker gravitational force and boids will orbit around for a longer period of time before reaching a stable configuration. Changing the amount of *alignment*, *cohesion* and *separation* forces leads to different spatial patterns.

4.2 Controls

In the context of texture composition it is sometimes useful to change these and other parameters in real-time, in order to interactively modify the process of timbre evolution. This is made easy by a control panel (Figure 3) that allows to change the most important variables on the fly. On the left side knobs set the perceptual radius and the amount of force for each of the three tasks of *alignment*, *separation* and *cohesion*. On the right the user can control the overall output amplitude, the attractors' mass, the boids' maximal velocity and the transposition of granulation.

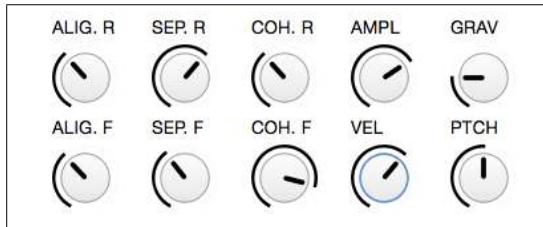


Figure 3: Control panel.

5. MUSICAL CONTEXTS

The author is employing this multi-agent system mainly in three works of different nature: *Breakpoint* - fixed media piece, *In Vitro* - installation, *Cocktail Break* - live performance.

Breakpoint is composed adopting an iterative compositional strategy. First of all, a number of samples is arranged by hand in a simple temporal structure. The resulting short collage is then analyzed and explored through the system. The audio outcome is recorded, edited and musically inserted in the previous collage of samples. The so obtained track is again analyzed and explored, and so on, in a cyclic procedure. At each new cycle the internal timbral movements of the piece are enhanced by adding new textures generated through the exploration process. Moreover, the structures emerging from the exploration lead to new sections, shaping the overall form of the piece in a germinal way.

At the time of writing, for *In Vitro* and *Cocktail Break* a physical interface is under construction (Figure 5): it involves a Rear Diffused Illumination setup [18], through which the audience or the performer could interact in an intuitive

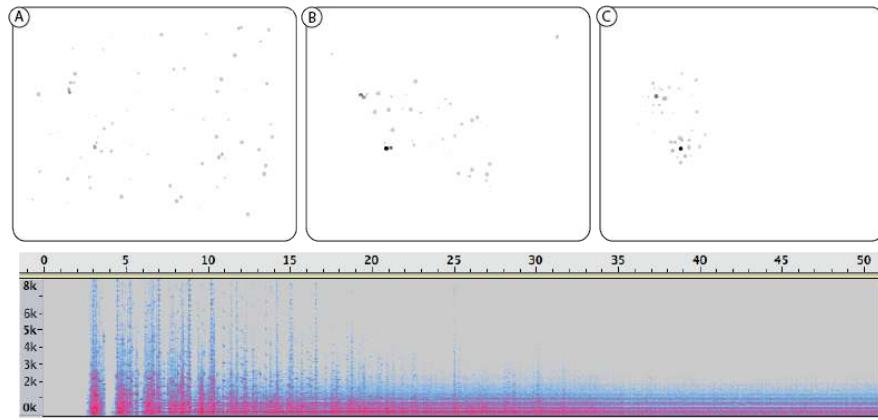


Figure 4: Fifty seconds long spectral contraction process. Swarm movement is represented in the upper part, spectrogram in the lower.

and tangible manner with the swarm. Also, several piezo-electric transducers are attached below the glass projection surface, meaning that the interaction could happen both in a tactile way - multitouch gestures - or through the physical vibrations produced over the screen.

5.1 Cocktail Break

Cocktail Break is a live performance aiming to explore the possibilities these two different ways of interaction can offer. When a gesture is performed over the glass screen, for instance by dragging a wine glass (Figure 5), both a sound and a movement are produced. Core concept of the project is analyzing these two indivisible inputs and using them to establish musical correlations with the boids flocking on the screen. The concatenative synthesis pool is gradually built during the performance by recording, segmenting and analyzing the acoustic sounds produced over the projection layer. Figure 6 displays the schematics of reactions and relations between gestures, boids and sound synthesis.

Each section of the piece features a different physical state that defines the boids' perceptual radii, the proportion of forces in the simulation, the maximal velocity and the viscosity of the virtual world. These parameters affect the swarm's behavior at a high level, designing the way it can react to the performer's inputs. The swarm's movement is then influenced by three kind of attractors:

1. Sounds picked up through piezos are analyzed according to the same features used for the construction of the morphological space. Attractors are placed at the resulting analysis coordinates and their mass is defined based on the signal loudness.
2. When objects or fingers touch the screen surface they become attractors as well, whose mass depends upon the size of the object itself.
3. Attractors are also positioned in the middle of dense areas of the explored morphology. Masses are defined according to density values.

Each boid is coupled to a single granulator and its movement controls some parameters of the synthesis. Its position determines which grain to play and its number of neighbors dictates the grain amplitude. The distance between the boid and its center of mass controls the grain length, while its velocity defines the rate of grain playback. The relative position to the center of mass informs the spatialization unit. At the same time the gestural input controls an IIR low pass filter. The overall output is analyzed by mean of several features and the obtained data are used to further adjust the swarm simulation. For instance, if a certain amplitude threshold is reached the separation force is increased, so that boids divide and volume decreases.

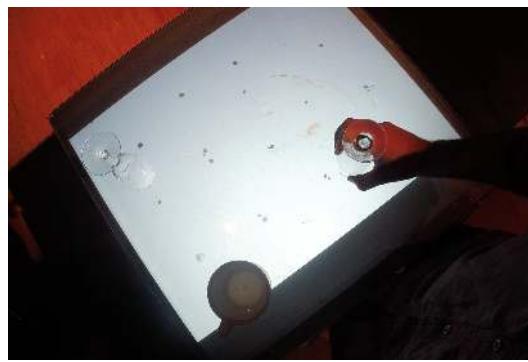


Figure 5: Physical interface for tangible interaction.

Goal of the performance is to musically combine sounds and gestures produced over the glass surface - which are both processed and amplified - with the concatenative processes controlled by the swarm.

6. CONSIDERATIONS

For artistic purposes, the application of a multi agent software in the context of corpus-based concatenative synthesis is an interesting alternative to the standard approach.

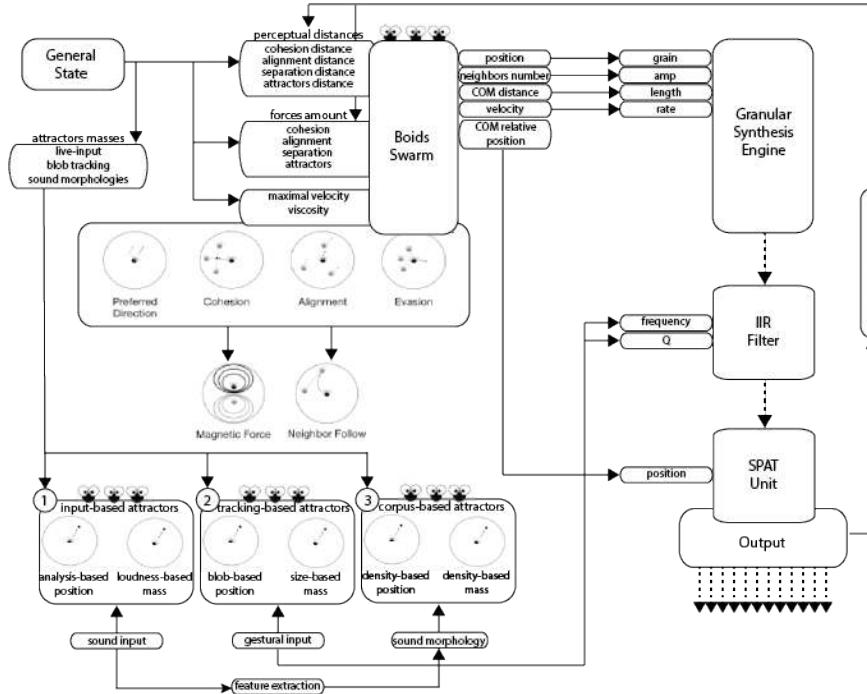


Figure 6: Schematics of the swarm music performance *Cocktail Break*.
The characteristics of the swarm simulation are depicted on the left side.
The musical architecture is shown on the right.

Whereas sound cataloguing is commonly employed to enhance the composer's control over materials, from this perspective it is rather used as ground for building a dynamic system with emergent properties, capable of producing unexpected and surprising results. Through this approach, the materials themselves are enhanced, leading to the formation of both timbre and form in a natural dynamic process. Space-related agents turned out a reasonable and pliable choice for exploring two dimensional sound morphologies: through the alteration of their perceptual parameters it is possible to intuitively balance their behavior, creating more specific timbral movements and changing textures. The boids' capability of sensing underlying morphologies is also a useful tool for the composer, which allows to better define the focus of the explorative process.

7. REFERENCES

- [1] D. Schwarz, G. Beller, B. Verbrugghe, S. Britton, *Real-Time Corpus-Based Concatenative Synthesis with CataRT*, 9th International Conference on Digital Audio Effects (DAFx), Montreal, Canada. pp.279-282, Sep 2006.
- [2] D. Schwarz,, *Distance Mapping for Corpus-Based Concatenative Synthesis*, 8th Sound and Music Computing Conference, Padova, Italy, Sep 2011.
- [3] P. Dahlstedt, P. McBurney, *Musical Agents: Toward Computer-Aided Music Composition Using Autonomous Software Agents*, Leonardo Vol. 39, No. 5, Pages 469-470, MIT Press, Cambridge, Massachusetts, Oct 2006.
- [4] S. Pearse, D. Moore, *An Agent Based Approach to Interaction and Composition*, 42nd International Computer Music Conference, Athens, Greece, Oct 2014.
- [5] C. Reynolds, *Flocks, herds, and schools: A distributed behavioral model*, in computer graphics, SIGGRAPH 87 Conference Proceedings, vol. 21, no. 4, pp. 25-34, 1987.
- [6] N. Collins, *SCMIR: a SuperCollider Music Information Retrieval library*, 41st International Computer Music Conference, University of Huddersfield, UK, Aug 2011.
- [7] C. Roads, *Microsound*, MIT Press, Cambridge, Massachusetts, Aug 2004.
- [8] A. Di Scipio, *Formal Processes of Timbre Composition. Challenging the Dualistic Paradigm of Computer Music*, Proceedings of the 1994 International Computer Music Conference. San Francisco: International Computer Music Association, pp. 202-208, 1994.
- [9] S. Wilson, *Spatial Swarm Granulation*, Proceedings of 2008 International Computer Music Conference, SARC, ICMA, Belfast, Aug 2008.

- [10] T. Blackwell, M. Young, *Self-Organised Music*, in *Organised sound*, 9(2), pp. 123-136, Cambridge University Press, 2004.
- [11] T. Blackwell, M. Young, *Swarm Granulator*, in *Applications of Evolutionary Computing*, pp. 399-408, Springer Berlin Heidelberg, 2004.
- [12] T. Blackwell, *Swarm Music: Improvised Music with Multi-Swarms*, in *Symposium on Artificial Intelligence and Creativity in Art and Science*, pp. 41-49, 2003.
- [13] T. Blackwell, *Swarming and Music*, in *Evolutionary Computer Music*, pp. 194-217, Springer London, 2006.
- [14] T. Unemi, D. Bisig, *Playing Music by Conducting BO-ID Agents*, in *Proceedings of the Ninth International Conference on Artificial Life IX*, pp. 546-550, Boston, 2004.
- [15] Y. Uozumi, *GISMO2: An Application for Agent-Based Composition*, in *Lecture Notes in Computer Science*, Springer Berlin, 2007.
- [16] J. Schacher, D. Bisig, M. Neukom, *Composing with Swarm Algorithms - Creating Interactive Audio-Visual Pieces Using Flocking Behavior*, in *Proceedings of the International Computer Music Conference 2011*, University of Huddersfield, 2011.
- [17] E. Brown, J. Cage, M. Feldman, C. Wolff, *The New York School*, fl. E. Blum, pf. S. Schleiermacher, perc. J. Williams, hat ART, Switzerland, 1994.
- [18] A. Teiche et al., *Multitouch Technologies*, NUI Group, 2009.

Sonic in(tro)spection by vocal sketching

Andrea Cera
Iuav University of Venice

Davide Andrea Mauro
Iuav University of Venice

Davide Rocchesso
Iuav University of Venice
roc@iuav.it

ABSTRACT

How can the art practice of self-representation be ported to sonic arts? In *S'i' fosse suono*, brief sonic self-portraits are arranged in the form of an audiovisual checkerboard. The recorded non-verbal vocal sounds were used as sketches for synthetic renderings, using two seemingly distant sound modeling techniques. Through this piece, the authors elaborate on the ideas of self-portrait, vocal sketching, and sketching in sound design. The artistic exploration gives insights on how vocal utterances may be automatically converted to synthetic sounds, and ultimately how designers may effectively sketch in the domain of sound.

1. VOCAL SKETCHING

Vocal sketching is the act of communicating a sonic concept using the voice. Assuming an idea of sound exists in the mind of a person, sketching is a representational act that uses the most direct means for sound production.

An old jazz adage says “If you can’t sing it, you can’t play it”. Given the constraints and possibilities of the voice organ, an interesting question is whether the way we imagine sounds is affected by our sketching practice and abilities. In the context of modern theories of embodiment, such as the ideomotor theory or the theory of event coding, the internal representations come in the form of perception-action ensembles [1], and we can imagine what we have previously experienced through perception-action loops. Given the effectiveness of sound communication by vocal imitation [2], it seems that our sonic imagery, while transcending the vocal acoustic space, can be effectively and compactly represented via vocal gestures. These vocal sketches can be exploited as proxies for non-utterable sounds or as entry points to vast sonic spaces.

Similarly to drawing, vocal sketching can be instrumental to establishing a positive-feedback loop that let sonic ideas emerge and take form, through a continuous confrontation between internal and external representation. Such a loop can be entirely controlled by the designer or, in a participatory framework, it can be open to other participants, whose sketches can contribute to a reflective process [3].

Copyright: ©2016 Andrea Cera et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2. ON ABSTRACT AND CONCRETE SOUNDS

An underlying axiom of early computer music research and practice was that “there are no theoretical limitations” as “any perceivable sound can be generated” [4]. This view of the computer as a white canvas for sound and music was a promoting factor for abstraction in sound synthesis and composition. On the other hand, when the raw audio material to be painted on canvas was coming from recordings, the limitless view promoted an acousmatic, reduced way of listening [5]. Are these computer-music approaches utopian attitudes or rather feasible approaches to sound design and composition?

Quoting Giorgio Morandi (1890-1964), a painter who used to fill his white canvases with only bottles, vases, and carafes, “nothing is more abstract than reality” and “to achieve understanding it is necessary not to see many things, but to look hard at what you do see”. Which sounds can we imagine? We can imagine, and possibly imitate, those sounds that we can (directly or indirectly) produce, and those of the environment that have the potential to trigger an action.

Action-sound associations are plastic [6] and, as such, can be designed. However, experiments in dimensionality reduction and categorization have highlighted the anchoring role of basic mechanisms of physical sound production, such as gas turbulence, fluid dynamics, impact, friction, rotary machines, etc. [7]. Similarly, the sounds that can be vocally produced can be decomposed and analyzed in terms of fundamental mechanisms such as turbulence, myoelastic vibration, impact, phonation. So, there seems to be a lexicon of basic elements for sound composition, a range of not many things that can be outlined, colored, and arranged on the white canvas in infinitely many ways. In this respect, doing sound design may mean to compose sound using this lexicon of physically-grounded phenomena, thus keeping a direct link to sound imagination.

3. SONIC SELF-PORTRAITS

The self-portrait is a recurrent exercise in visual arts of all times, and across all techniques, styles, and movements. How can we translate self-representation to sonic arts? How can we exploit sound synthesis in perception-action loops? Some attempts have been previously made to devise audio-visual tools that stimulate and enhance self-expression, and to produce visual-sonic self-portraits. For example, Polotti and Goina [8] realized a system based on correspondences between elementary sonic and movement units, and asked participants to express themselves through movement-

generated sound.

The EU Project SkAT-VG aims at developing tools and methods to exploit the innate vocal sketching abilities of humans, in the early stage of the sound design process [9]. In this context, the perception-action loop is including interpretation and synthesis, and tools for effective sonic introspection are going to be available. However, in an exploration stage, sound-designer expertise is sought to envision meaningful translations of vocal utterances into designed, artificial sounds, a sort of inspection into the possibilities of vocal sketching. From a science and technology perspective, the artist/designer acts as an exploratory probe: He inspects the utterances and uses them as proxies to imagined sound spaces. The resulting designed sounds let us foresee what an automatic translation process might be aimed at, so to become an effective sonic extension of the voice.

In *S'i' fosse suono* (lit., if I were sound), we asked sixteen persons (including the second and third authors) to represent themselves with a brief non-verbal vocal sound. They were audio-visually recorded and their performances were used as sketches for two different synthetic renderings, one using physics-based sound modeling, and the other using synthesis by recombination of vocal grains. The resulting materials were used to compose an audiovisual interactive checkerboard, which is a proof of concept of vocal sketching and its exploitation in sound design and sonic arts. By producing a sonic self-portrait, each participant was requested to make imagination audible. The sound designer (Andrea Cera) was expected to understand, interpret, and transform such representational act. Two sets of constraints were given to the sound designer for such transformation, so that two different synthetic sound renderings were produced for each recorded vocal production:

1. use physics-based sound models of fundamental sound-generating mechanisms, as made available by the Sound Design Toolkit (SDT) [10];
2. manipulate the audio buffer containing the recorded vocalization, using granular techniques (MuBu) [11].

4. DESIGN PROCESS

If framed in a situated ontology of design [12], the creation of *S'i' fosse suono* can be described as the sequence:

1. **interpreted** world: The sixteen participants (Figure 1) imagined a sonic self-representation in terms of perception-action associations and concepts;
2. **expected** world: Each participant set a motor program for acting sonically by means of the voice, thus translating imagination into action;
3. **external** world: The utterances were communicated to the sound designer, who interpreted them as blueprints for synthetic sound composition.

Stage 2 is reached from stage 1 via focusing, i.e. taking some aspects of the interpreted world and using them as goals for the expected world. Stage 3 is the effect on the external world achieved via goal-driven action. Such

sequence may be looped, in such a way that the interpreted world (and its perception-action associations) may be modified after new experiences and interpretation of the external world. In this case, however, we present the result of an open-loop process, where the sound designer interprets and affects the external world. The sequence is also indicative of what a tool-mediated sound design process may produce, where the sound designer may be involved in stages 1 and 2, while the translation of vocal blueprints into new sounds of the external world may be performed by a machine. In the proposed artistic installation, instead, human agents with different roles have been involved for stages 1-2 (participants) and for stage 3 (sound designer). The expected world and the external world are made jointly accessible as the audio-visual checkerboard chooses randomly, upon being clicked in one face box, if playing back the vocal utterance or one of its two renderings in synthetic sound. The constraints given to the sound designer in terms of usable sound models, and his use of some automatic feature extractors, make the automation of the external world easier to foresee.

In the framework of embodied music cognition and mediation technology [13], stages 1 and 2 can be associated with a first-person perspective, where “moving sonic forms take the status of actions to which intentionality can be attributed”. Stage 3 is that of a third-person perspective, where phenomena get somehow measured and translated, either by a human observer (in this case, the sound designer) or by a machine. The experience of the installation is that of a second-person perspective, where the observer gets involved “with physical energy in a context of intersubjective communication”.



Figure 1. *S'i' fosse suono*

5. SOUND DESIGN

The sound design-by-transformation process is here presented in some detail by two examples, one based on physical models, and the other based on audio manipulation. The vocal production is that of the participant portrayed in row 4 and column 2 of Figure 1. In articulatory terms, her sonic self-portrait can be described as a train of labial myoelastic pulses superimposed to a steady phonation. The voice-driven sound design process is divided into three stages (physical models):

1. **Analysis.** Two streams are automatically extracted from audio: (A) discrete onsets (Ableton Live) and (B) continuous pitch (SDT pitch extractor).
2. **Synthesis.** Two processes are executed: (X) Stream A drives the SDT bubble sound model (Figure 2, left) and (Y) Stream B drives the gear ratio and the RPM parameters of the SDT dc motor model.
3. **Rendering.** The audio outputs of processes X and Y are layered. Process X is processed through a convolution reverb, following a hand-drawn automation curve.

In the case of audio buffer manipulation, the analysis stage is unchanged, and the second stage is replaced by

2. **Synthesis.** Using MuBu granulator (Figure 2, right): Stream A drives duration and position, in such a way that each onset places the granulator head in proximity of one of the vocal events and then back to a point where there is only the sustained note,

and the third stage is reduced to reverberation.

6. THE PERSPECTIVE OF THE SOUND DESIGNER

When the first author (sound designer) started working on the collection of the sixteen recordings, he immediately realized that he needed a perspective from which to observe and judge the designed sounds and their relations with the original audio-visual counterparts. The driving idea of bad imitation [14] was considered to be a valuable strategy for this case. This paradigm is based on the deliberate production of a discrepancy between one imitated object and one imitating agent, which brings novelty and surprise in the imitation process. It relies on bounces between the internal world and the external world, driven by processes of perception (listening) and action (reproducing). The bad-imitation paradigm can be embedded in the programming of an interactive behavior [15], but it can also be exploited during the preparatory stage of a creative production.

In *S'i' fosse suono*, bad imitations were used to limit the scope of analysis of the original recording by embracing one of two mutually exclusive attitudes, defined by the sound designer as a) acousmatic, or b) concrete:

- a. Acousmatic approach to the recorded utterances. The sound designer was trying to abstract from physical information that could be derived from visual moving images, without searching neither for possible referent physical phenomena nor for articulatory details of the vocal apparatus. The original sound was approached as if it was artificial, trying to imagine which would be the control signals that would have led to its synthetic production. The SDT sound models were selected according to the proximity of their sound output to the hypothetical sound output of such an ideal synthesizer. Once the choice was made, the designer forced himself to stick with it. For example,

participant in row 2 and column 3 (Figure 1) produced a complex and articulated vocal expression, using ingressive air streams and her mouth as a filter. The designed synthetic sounds, both in the SDT and in the MuBu version, bear significant differences from the uttered spectral signatures. However, these discrepancies are offset by the parallelism between the temporal evolutions of the imitated and imitating morphologies;

- b. Concrete approach to the recorded utterances. The sound designer was prompted, by the nature of the reference recording, to consider a (natural or mechanical) everyday sound event. As a consequence, there was no other choice than using the SDT sound models meant to imitate such events (explosions, motors, wind, etc.). In this case, a weaker adherence between the imitated and imitating sounds, both in terms of spectral signatures and of temporal evolution, was accepted due to the physical metaphor overriding the morphological similarities.

The choice between these two attitudes (acousmatic vs. concrete) was also suggested by two ways in which the participants enacted their self-representation: some of them tended to simply explore their voice, while some others seemed to look for specific sounds (wind, far out explosion, motor, etc.), as if to sonically embellish their image. Similar tendencies of self-embellishment are found in the widespread practice of taking visual selfies [16].

It is interesting to notice that what has been described as an acousmatic approach, is indeed referring to a hypothetical synthesizer, an imaginary sound-producing device that just confirms that “there can never be any sound ‘as such’, and that sounds are always events of and in a field of relationships” [17]. Even under a self-imposed discipline, the sound designer could not escape the irreducible nature of listening.

7. RECEPTION AND DEVELOPMENTS

S'i' fosse suono was realized in the form of interactive installation, with a graphic framework (Figure 1), audio-visual recordings, and synthetic soundfiles held together in a Processing program. The piece was first exhibited at the ICT Conference of the European Commission in Lisbon, Portugal, on october 20-22, 2015. Since then, it has been used many times to elicit an intuitive understanding of the concept behind the SkAT-VG project. A multi-touch screen supported a playful engagement with the piece¹. Many visitors, while initially triggering single utterances with circumspection, quickly showed a mixture of surprise and pleasure and started playing with the checkerboard with rapid sequences of finger taps. The overlapping events produce sorts of audio-visual arpeggios, whose actual outcome can never be predicted, as each face gesture is randomly coupled to the original or to one of the two synthetic sounds. In a second-person perspective [13], if the observer succeeds in re-enacting the sound sketch as made of actions

¹ In its web version, and without multi-touch support, *S'i' fosse suono* is available at skatvg.eu/SIFosse/.

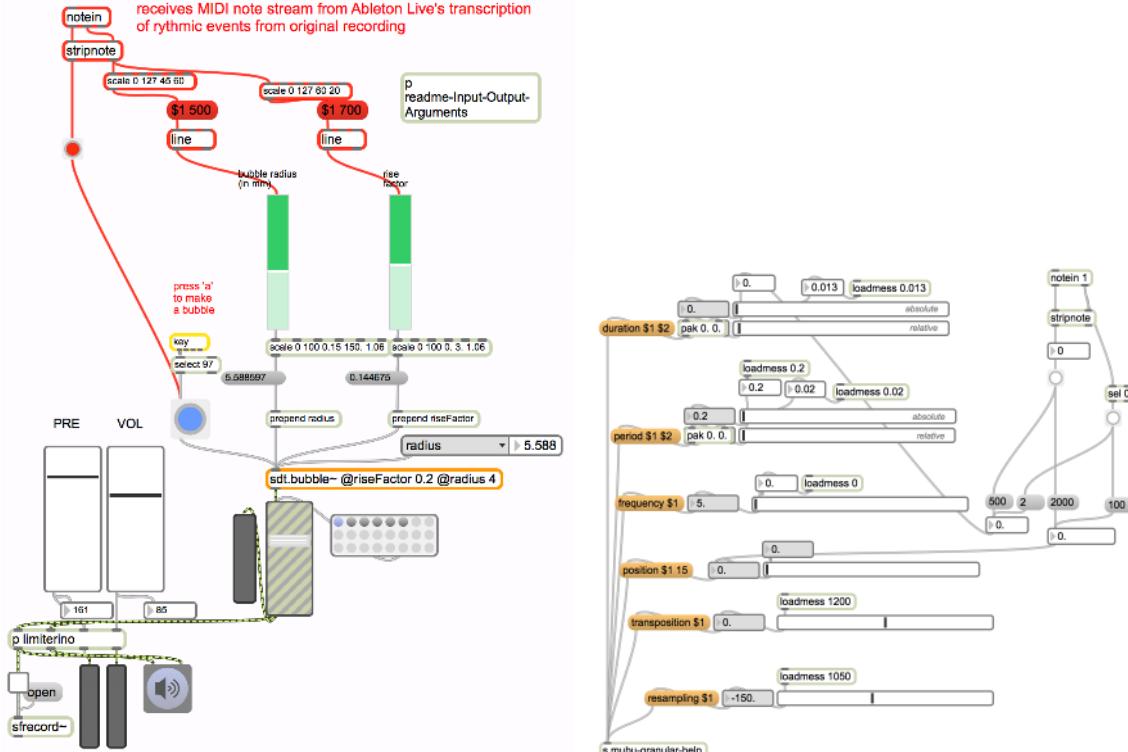


Figure 2. Stream of bubbles with the SDT (left) and granulation with MuBu (right)

with an intention then the communication act is found to be successful. Since this was found to happen for both the vocal sketches and their synthetic translations, it means that plausibility is preserved in the translation stage, and there is evidence of the effectiveness of the voice for sound sketching

“Audio-visual objects are constructs of the mind” [18], the result of a process that produces the most plausible binding of sensory information to objects and events. From the perspective of the observer, the three sound realizations result equally effective at eliciting such binding. The fact that the synthetic sounds are derived from a vocal utterance, which was interpreted as an imitation in acousmatic or concrete terms, makes the association plausible and strong.

This work shows both the concreteness of vocal sound materials and the versatility of sound models that refer to fundamental sound production mechanisms. The role of the sound designer was that of a probe, to explore the vast space of possible interpreted renditions of vocal utterances. By inspecting the sound design process, some indications were derived for future automation of the rendering process, towards a more effective use of the voice as a sound sketching tool. The two embraced attitudes, the acousmatic and the concrete, were largely dependent on the nature of the original vocal production in its recognizability as an

imitation of an everyday sound phenomenon. The leading synthesis technique was tightly following such attitude. Nevertheless, in most cases physical models or audio-buffer manipulation were equally effective in producing consistent and compelling sound realizations. The automation of the voice-to-synth conversion processes, as envisioned by the SkAT-VG project, is encouraging us to explore iteration and plasticity in perception-action loops of sound design processes [19]. It will be interesting to see how new forms of self-representation will emerge.

8. ACKNOWLEDGMENTS

Luca Ludovico (University of Milan) coded the web version of *S'i fosse suono*. Most of the participants who donated their vocal self-portraits are part of the theater group Cantiere Ca’ Foscari, directed by Elisabetta Brusa. We acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open Grant number 618067: SkAT-VG (Sketching Audio Technologies using Vocalizations and Gestures, skatvg.eu).

9. REFERENCES

- [1] B. Hommel, "The theory of event coding (TEC) as embodied-cognition framework," *Frontiers in Psychology*, vol. 6, no. 1318, 2015.
- [2] G. Lemaitre and D. Rocchesso, "On the effectiveness of vocal imitations and verbal descriptions of sounds," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 862–873, 2014.
- [3] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen, "User sketches: A quick, inexpensive, and effective way to elicit more reflective user feedback," in *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles*, NordiCHI '06, (New York, NY, USA), pp. 105–114, ACM, 2006.
- [4] M. V. Mathews, "The digital computer as a musical instrument," *Science*, vol. 142, no. 3592, pp. 553–557, 1963.
- [5] M. Chion, *Audio-Vision: Sound on Screen*. New York: Columbia University Press, 1994.
- [6] G. Lemaitre, L. M. Heller, N. Navolio, and N. Zúñiga-Peñaanda, "Priming gestures with sounds," *PLoS ONE*, vol. 10, no. 11, 2015.
- [7] G. Lemaitre, A. Dessein, P. Susini, and K. Aura, "Vocal imitations and the identification of sound events," *Ecological Psychology*, vol. 23, no. 4, pp. 267–307, 2011.
- [8] P. Polotti and M. Goina, "EGGS in action," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, (Oslo, Norway), 2011.
- [9] D. Rocchesso, G. Lemaitre, P. Susini, S. Ternström, and P. Boussard, "Sketching sound with voice and gesture," *interactions*, vol. 22, pp. 38–41, January 2015.
- [10] S. Delle Monache, P. Polotti, and D. Rocchesso, "A toolkit for explorations in sonic interaction design," in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, AM '10, (New York, NY, USA), pp. 1:1–1:7, ACM, 2010.
- [11] N. Schnell, A. Röbel, D. Schwarz, G. Peeters, and R. Borghesi, "MuBu & friends - assembling tools for content based real-time interactive audio processing in Max/MSP," in *Proceedings of the International Computer Music Conference*, (Montreal, Canada), 2009.
- [12] J. S. Gero and U. Kannengiesser, "The function-behaviour-structure ontology of design," in *An Anthology of Theories and Models of Design: Philosophy, Approaches and Empirical Explorations* (A. Chakrabarti and M. L. T. Blessing, eds.), pp. 263–283, London: Springer London, 2014.
- [13] M. Leman, *Embodied music cognition and mediation technology*. Mit Press, 2008.
- [14] A. Cera, "Écoutes et mauvaises imitations," in *In actu: de l'expérimental dans l'art* (E. During, L. Jeanpierre, C. Kihm, , and D. Zabunyan, eds.), Djon, France: Les presses du réel, 2009.
- [15] A. Cera, "Loops, games and playful things," *Contemporary Music Review*, vol. 32, no. 1, pp. 29–39, 2013.
- [16] F. Souza, D. de Las Casas, V. Flores, S. Youn, M. Cha, D. Quercia, and V. Almeida, "Dawn of the selfie era: The whos, wheres, and hows of selfies on instagram," in *Proceedings of the 2015 ACM on Conference on Online Social Networks*, COSN '15, (New York, NY, USA), pp. 221–231, ACM, 2015.
- [17] A. Di Scipio, "The politics of sound and the biopolitics of music: Weaving together sound-making, irreducible listening, and the physical and cultural environment," *Organised Sound*, vol. 20, pp. 278–289, 12 2015.
- [18] M. Kubovy and M. Schutz, "Audio-visual objects," *Review of Philosophy and Psychology*, vol. 1, no. 1, pp. 41–61, 2010.
- [19] D. Rocchesso, D. A. Mauro, and S. Delle Monache, "mimic: The microphone as a pencil," in *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '16, (New York, NY, USA), pp. 357–364, ACM, 2016.

Installazioni interattive per la valorizzazione di strumenti musicali antichi: il flauto di Pan del Museo di Scienze Archeologiche e d'Arte dell'Università degli Studi di Padova

Edoardo Micheloni

Politecnico di Milano

edoardo.micheloni@gmail.com

Niccolò Pretto, Federico Avanzini,

Sergio Canazza, Antonio Rodà

Università degli Studi di Padova

name.surname@dei.unipd.it

ABSTRACT

Questo articolo descrive un progetto relativo all'implementazione di un flauto di Pan virtuale all'interno di una installazione museale che verrà esposta presso il Museo di Scienze Archeologiche e d'Arte dell'Università degli Studi di Padova. In primo luogo, l'articolo introduce il problema della conservazione attiva e della valorizzazione degli strumenti in ambito museale. A seguire descrive l'installazione multimediale progettata appositamente per valorizzare un antico flauto di Pan ritrovato in Egitto, negli anni Trenta, durante una campagna di scavi archeologici. Il reperto è stato sottoposto a svariate analisi, quali scansioni 3D e Tomografia Computerizzata, da cui sono state estratte le misure interne ed esterne del flauto. Queste sono state indispensabili per ricavare l'intonazione delle canne e quindi creare il modello alla base del flauto virtuale. Infine, l'articolo descrive in maniera approfondita le scelte progettuali, le modalità di interazione e l'implementazione delle due sezioni dell'installazione dedicata al suono. L'utente, infatti, può "suonare" il flauto utilizzando come input sia i comandi touch, sia il "soffio".

1. INTRODUZIONE

Negli ultimi anni, i musei si sono dovuti evolvere da luoghi dove tradizionalmente vengono conservati ed esposti reperti a istituzioni focalizzate sull'educazione e l'intrattenimento dei visitatori [1]. I curatori di musei e mostre, infatti, si sforzano sempre più di progettare e realizzare esposizioni che siano per i visitatori un'esperienza allo stesso tempo piacevole ed educativa [2].

Non di rado, capita di imbattersi in strumenti musicali, antichi o moderni, esposti all'interno di mostre o musei (non necessariamente dedicati alla musica). Questi strumenti possono essere suddivisi in tre principali categorie:

- Strumenti elettrofoni, dai primi fonografi fino ai moderni dispositivi digitali, come quelli esposti presso la Casa della Musica di Parma [3];
- Strumenti acustici tradizionali, ad esempio i violini conservati al Museo del Violino di Cremona;

Copyright: ©2016 Edoardo Micheloni 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. Flauto di Pan esposto presso il Museo delle Scienze Archeologiche e d'Arte (MSA) dell'Università di Padova.

- Strumenti musicali primitivi, come il flauto di Pan conservato presso Museo di Scienze Archeologiche e d'Arte (Figura 1), usato come caso di studio in questo articolo.

Questi oggetti sono molto differenti tra di loro, ma in quanto strumenti musicali hanno in comune la necessità di essere suonati per essere compresi appieno. È evidente, quindi, quanto sia insufficiente presentare uno strumento musicale chiuso in una teca e con una semplice didascalia a supporto. La maggior parte di questi strumenti però non possono essere suonati dal pubblico, sia per salvaguardare il reperto, sia, spesso, a causa di corruttele.

È necessario quindi trovare delle strade alternative per interagire con questi oggetti. La creazione di uno strumento virtuale, in grado di ricreare un suono vicino all'originale, è la soluzione proposta dagli autori, che già in progetti passati si sono dedicati alla conservazione attiva di strumenti musicali [4, 5]. La proposta nasce dalla trasposizione del concetto di copia conservativa che caratterizza la conservazione attiva di documenti sonori storici [6]: la virtualizzazione degli strumenti musicali viene proposta quindi come copia dello strumento, che poi verrà utilizzata e adattata per creare quella che possiamo chiamare copia di accesso.

Mentre per i documenti sonori storici esiste già una ricca letteratura riguardante la digitalizzazione, per la ricostruzione di strumenti musicali mancano ancora metodo-

logie condivise. L'articolo vuole fare un passo in questa direzione proponendo uno caso di studio basato su un antico flauto di Pan conservato presso il Museo delle Scienze Archeologiche e d'Arte (MSA) dell'Università di Padova.

Ritrovato in Egitto negli anni Trenta, il reperto è uno dei tanti oggetti che arrivarono a Padova grazie alle ricerche archeologiche di Carlo Anti, che diresse la Missione Archeologica Italiana in Egitto dal 1928 e condusse i suoi scavi nel borgo antico di Tebtynis nell'oasi del Fayum, nel periodo dal 1930 al 1936, assistito dall'archeologo italo-inglese Gilbert Bagnani. Lo strumento è arrivato in Italia all'interno di una scatola usata per lastre fotografiche. La scatola probabilmente apparteneva a Bagnani, come documentato da una breve nota al suo interno. Il coperchio della scatola riporta invece una frase in francese scritta dalla moglie di Bagnani, che collocerebbe il reperto originale a Saqqara, in una zona limitrofa alla tomba del faraone Pepi II [7]. La datazione precisa verrà fornita dall'esame ^{14}C , i cui risultati definitivi, al momento della scrittura del paper, non sono ancora disponibili.

Il flauto è composto da 14 canne di diversa lunghezza unite da corde e da leganti naturali. In origine, esso era rivestito da uno strato di resina, ora parzialmente mancante. L'eccezionale integrità di questo reperto ha consentito di procedere a uno studio approfondito che ha coinvolto un team multidisciplinare composto da esperti in archeologia, archeometria, archeologia virtuale, ingegneria informatica ed elettronica, musicologia, scienze della terra. Come riportato in [7], sono state effettuate numerose analisi come ad esempio una completa scansione laser 3D e una scansione a raggi X. In aggiunta a quanto riportato nel precedente articolo, alle analisi si è aggiunta una Tomografia Computerizzata (TC), che ha permesso di avere le misure interne del flauto, necessarie a sviluppare il modello.

Tutto il lavoro verrà presentato al pubblico mediante una installazione multimediale presentata nella sezione 2. Questa consentirà ai visitatori non solo di accedere alle informazioni sul flauto, ma anche di "suonare" il flauto virtuale. La sezione 3 presenterà quindi come sono stati ricreati i suoni del flauto, mentre la sezione 4 approfondirà le scelte progettuali e i dettagli implementativi della sezione dedicata all'interazione con il modello virtuale.

2. L'INSTALLAZIONE MUSEALE

Il Museo delle Scienze Archeologiche e d'Arte dell'Università di Padova metterà a disposizione dei visitatori un'installazione multimediale per accedere a tutte le informazioni riguardanti l'antico flauto di Pan. Il design dell'installazione si basa su due differenti sistemi di interazione: esplorativo e creativo [8].

L'interazione esplorativa, realizzata per mezzo di un display touchscreen, permette ai visitatori di muoversi all'interno di uno spazio multimediale attraverso una piattaforma informativa realizzata su misura per l'installazione, che riporta tutte le informazioni iconografiche e didattiche sullo strumento. Il secondo, invece, permette all'utente di accedere alla ricostruzione delle peculiarità sonore originali dello strumento con la modalità di interazione tipica del flauto: "il soffio".



Figure 2. L'installazione museale.



Figure 3. Schermata iniziale dell'applicazione touch con le cinque sezioni.

L'input touch consente ai visitatori di esplorare le 5 sezioni in cui sono raggruppati e presentati i diversi contenuti multimediali (figura 3). La sezione MITO nasce principalmente per soddisfare un pubblico giovane e presenta sotto forma di fumetto il racconto mitologico della nascita del flauto di Pan, tratto dalle Metamorfosi di Ovidio. La parte STORIA presenta le principali fonti letterarie e iconografiche che testimoniano la diffusione del flauto di Pan nel mondo, presentandole in successione mediante una linea temporale. La sezione FLAUTO è composta invece da quattro sottosezioni che presentano rispettivamente nozioni di musica greca e romana in Egitto, la storia del ritrovamento del reperto, una descrizione delle peculiarità dello strumento e i dettagli su come è stata prodotta una ricostruzione del flauto. La parte 3D consentirà di interagire con due modelli tridimensionali: una scansione laser esplorabile con semplici gesture touch, e una Tomografia Computerizzata (TC) renderizzata su tre piani perpendicolari che permettono la visione di sezioni interne del flauto. La sezione SUONO è una rappresentazione stilizzata dello strumento sullo schermo: toccando ogni canna viene riprodotto il suono, ricreato sulla base di analisi e misurazioni dello strumento (sezione 3).

L'interazione creativa si presenta come una rappresentazione stilizzata del flauto di Pan, intagliata sulla superficie del mobile dell'installazione. A differenza della sezione SUONO, appena descritta, l'interazione avviene grazie a dei sensori posizionati su dei fori che rappresentano le imboccature delle canne stilizzate. Il soffio dell'utente è l'input che viene interpretato e tradotto in suono.

3. DALLE MISURE AL MODELLO

Questo flauto di Pan è uno tra i reperti meglio conservati al mondo considerando il rapporto tra età e stato di conservazione, ed ha mantenuto, a distanza di millenni, quasi interamente la sua forma originale. Questo ha permesso di studiare diametro e lunghezza delle canne, al fine di poterne stimare l'intonazione originale. Ciò è possibile grazie alla formula:

$$f = \frac{c}{4(l_{int} + \Delta l)}$$

dove c rappresenta la velocità del suono, l_{int} la lunghezza interna e $\Delta l \sim 0.305d_{int}$ il fattore di correzione all'estremità aperta della canna (proporzionale al diametro interno) [9]. Per ottenere le misure necessarie alle analisi, sono state impiegate due differenti metodologie, basate su due diverse ricostruzioni tridimensionali dello strumento. La prima campagna di misurazioni è stata portata a termine su una scansione laser tridimensionale, realizzata con uno ScanArm v3 con una accuratezza di $\pm 35 \mu\text{m}$. Le misure raccolte però sono solo esterne, e sono state la base per una approssimazione di quelle interne [7]. Il secondo modello tridimensionale utilizzato per raccogliere le misure è stata una Tomografia Computerizzata (TC). Grazie agli strumenti del software medico Horos si è potuto procedere con misurazioni molto più accurate rispetto alle precedenti. Le misure estrapolate risentono di un errore di precisione del macchinario utilizzato per la TC: la risoluzione del pixel volumetrico (voxel) è di 0.625 mm. Nonostante la risoluzione inferiore, le misure interne delle canne sono reali, senza nessuna stima della posizione del nodo.

Nonostante il flauto sia uno dei meglio conservati al mondo, esso presenta alcune deformazioni strutturali che hanno influenzato sia la campagna di misurazione, sia la stima delle misure originarie dello strumento. Tra le principali possiamo enumerare: "ovalizzazione" delle canne, alcune rotture della struttura (in particolar modo nelle canne alle estremità), deterioramento delle imboccature e presenza di materiale organico all'interno delle canne. La procedura adottata durante le misurazioni ha tenuto in considerazione queste problematiche, infatti, le misure sono state raccolte in modo ridondante in differenti punti del flauto e su diversi piani.

Date le dimensioni ridotte dello strumento, piccole variazioni nelle misure (anche di pochi millimetri) possono comportare notevoli variazioni nella frequenza fondamentale risultante. Per ottenere un fattore di correzione più preciso, è stata eseguita una "circolarizzazione" delle canne. Questa operazione mira a migliorare la stima del diametro dello strumento originario: le sezioni delle canne ovalizzate possono essere approssimate a ellissi. Siano a e b diametri ortogonali, il perimetro è calcolabile con

$$p \approx \pi(3(a+b) - \sqrt{(3a+b)(a+3b)})$$

approssimazione dello sviluppo in serie della formula completa. Il risultato è una buona stima del perimetro originale della canna, che permette quindi di ottenere il diametro come

$$d = \frac{p}{\pi}$$

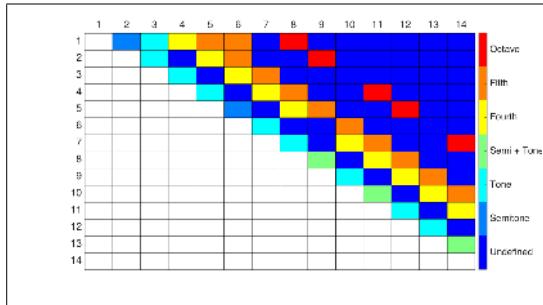


Figure 4. Intervalli di frequenze tra le 14 canne.

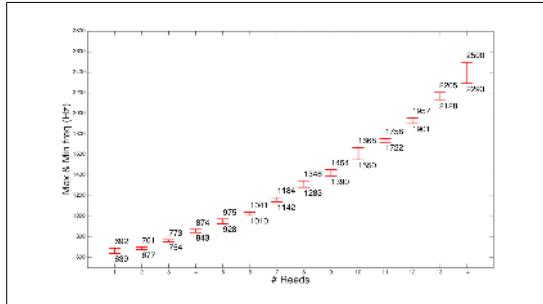


Figure 5. Massima e minima frequenza stimata per ciascuna canna.

Secondo gli studiosi [10], la musica greca antica era basata sul *tetracordo*, ovvero un gruppo di quattro note dove il rapporto tra l'intonazione della prima e quarta nota era uguale 4 : 3. Con le misure raccolte e con quest'ultima informazione, è stato possibile eseguire un studio preliminare che ha permesso di ottenere una corrispondenza di intervalli di frequenze tra le canne. Sono state considerate le misure massime e minime di diametro e di lunghezza, per stimarne la frequenza massima e minima (riportate in Fig. 5) e infine si è verificata la compatibilità dell'intervallo di frequenze stimate dalle misure con i rapporti descritti nei trattati teorici di musica greca antica.. In particolare, sono stati calcolati i rapporti

$$R(n) = \frac{f(n+3)}{f(n)}$$

con $n = 1, 2, \dots, 11$ e $f(n)$ frequenza fondamentale della canna. Verificando quindi che

$$\frac{f(n+3)_{min}}{f(n)_{max}} \leq 4 : 3 \leq \frac{f(n+3)_{max}}{f(n)_{min}}$$

si può avere una prima conferma della presenza di un rapporto di quarta giusta.

Eseguendo questo confronto per tutte le canne e per i principali intervalli, è stato ottenuto il risultato in Fig.(4), dove è possibile notare che gran parte degli intervalli tra gruppi di quattro canne sono compatibili con la definizione di tetracordo. Inoltre, è possibile distinguere intervalli compatibili con il tono (9 : 8) e altri intervalli più pic-

coli compatibili con l'intervallo che nei trattati di teoria è chiamato *diesis* (256 : 234).

Grazie ai risultati ottenuti, è stato possibile identificare le intonazioni di ciascuna canna, ed è stato possibile realizzare i campioni per l'installazione museale, che verranno descritti nelle sezioni successive.

4. IMPLEMENTAZIONE DEL FLAUTO VIRTUALE

I risultati delle analisi presentati nella sezione 3 hanno permesso di realizzare dei campioni audio indispensabili per la ricostruzione del suono originale del flauto di Pan nell'installazione. Questi campioni sono stati utilizzati sia nella rappresentazione del flauto realizzata tramite sensori, sia nella sezione SUONO dell'applicazione touch, descritte nella sezione 2.

4.1 Acquisizione ed elaborazione del segnale

Uno dei principali fattori critici di successo del progetto, in grado di influenzare profondamente l'implementazione dell'installazione, è stata la scelta dei sensori da adoperare nella parte del mobile dedicata all'interazione creativa. Per trovare il sensore adatto all'installazione sono state vagliate diverse alternative. Le prime ipotesi si sono focalizzate su sensori tipicamente usati su anemometri o su strumentazioni per la spirometria. I principali fattori che hanno permesso di scartare l'ipotesi sono stati: le dimensioni non adatte all'installazione, l'impossibilità di fermarsi repentinamente al termine del soffio e in alcuni casi la necessità di dover appoggiare la bocca a dei tubicini, soluzione che non si adatta al design dell'installazione.

La seconda alternativa prevedeva l'utilizzo di sensori piezoelettrici sia nel formato classico "a contatto", sia nella versione sensibile alla flessione della sua struttura (*Piezoelectric Sensor*). Queste tipologie di sensori hanno ottenuto i migliori risultati in termini di interpretazione della durata e dell'intensità del soffio. Nonostante ciò, sono stati scartati a causa delle dimensioni che difficilmente si adattano alle esigenze dell'installazione e la poca resistenza alle vibrazioni causate da eventuali colpi sul mobile.

Nonostante le performance leggermente inferiori, in termini di stabilità del segnale acquisito e resistenza ai rumori ambientali, la scelta finale è ricaduta su capsule microfoniche a condensatore. Le dimensioni contenute e la risposta attendibile in termini di intensità e durata del soffio permettono, con una ridotto apporto di elaborazione del segnale, di ottenere il risultato desiderato.

L'elaborazione è stata implementata con una scheda Arduino Mega 2560. Il software caricato nel microcontrollore si divide in tre principali sezioni: la prima gestisce i dati di input provenienti dai microfoni, attivando le sezioni successive solo se l'ampiezza del segnale supera una soglia preimpostata; la seconda elabora i dati ricevuti ed, infine, la terza gestisce l'invio e la ricezione dei messaggi per la riproduzione del suono del flauto.

Per impostare la soglia della prima sezione, ogni sensore è stato equipaggiato con un circuito di condizionamento

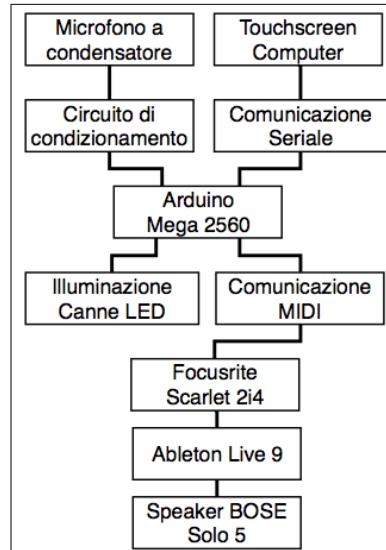


Figure 6. Schema funzionamento.

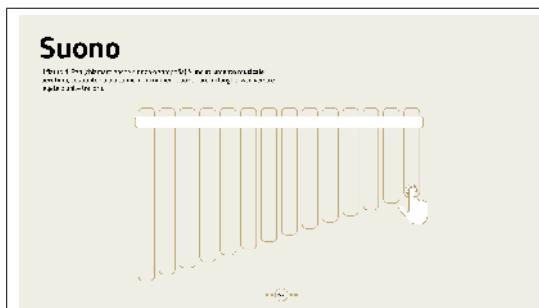


Figure 7. Sezione SUONO.



Figure 8. Sezione "Soffio" dell'installazione.

pensato *ad hoc*, e che si discosta leggermente dai comuni standard di utilizzo. Dato che Arduino può fornire una singola tensione, il segnale è stato infatti amplificato da un operazionale a singola alimentazione (5 V) anziché a doppia (5 V, -5 V). Con l'introduzione di questa soglia, le normali piccole oscillazioni causate da rumori ambientali o elettronici non vengono considerate, mentre l'evento sonoro viene attivato solo al superamento della stessa. Per usufruire anche delle tensioni negative prodotte dalla capsula microfonica (non necessarie per la definizione dei parametri del soffio), sarebbe stato necessario utilizzare un partitore resistivo in ingresso all'operazionale, per introdurre un offset di tensione a 2,5 V anziché a 0 V. Questa implementazione, però, sarebbe risultata meno stabile a causa dell'impossibilità di utilizzare una soglia costante. I segnali che superano la soglia vengono analizzati da Arduino per definire le tre *feature* che caratterizzano il soffio (attacco, intensità e durata [11]), e quindi il relativo suono delle canna virtuale suonata.

Il modello utilizzato dalla sezione SUONO dell'applicazione touch non prevede questo livello di sofisticazione. L'interfaccia, sviluppata con il motore grafico Unity 3D, propone le 14 canne stilizzate (Figura 7). Ognuna di queste è realizzata con un *GameObject* [12] a cui sono associati script, sviluppati in C# che permettono al visitatore di ascoltare il campione relativo alla canna selezionata. A differenza della parte precedente, il campione viene riprodotto sempre con una durata costante. Attraverso la porta seriale viene inviato un messaggio ad Arduino, il quale gestirà la priorità dei messaggi provenienti dai sensori e dal touch. Poiché non è previsto l'utilizzo contemporaneo di entrambi i metodi di interazione e non è possibile "suonare" più canne nello stesso momento, il microcontrollore gestisce l'output con un meccanismo di mutua esclusione, scartando i messaggi che arrivano quando la risorsa è occupata.

4.2 Rendering

I tre parametri individuati nell'elaborazione del "soffio" (attacco, intensità e durata) vengono tradotti in messaggi MIDI da Arduino. I messaggi inviati sono: *NoteOn*, *NoteOff* e *ControlChange*. Il primo viene usato per avviare la riproduzione dei campioni, il secondo per terminarla e il terzo per modulare il suono in base all'intensità del soffio. La comunicazione tra Arduino ed il PC è stata realizzata direttamente con una porta MIDI, e non tramite seriale. La porta in uscita è stata saldata direttamente sul *protoshield* di Arduino Mega, mentre l'altro terminale è stato collegato alla scheda audio del PC alla base dell'installazione.

In questa fase, il progetto prevede di utilizzare dei campioni sonori creati a partire da una libreria commerciale e opportunamente modificati: una volta selezionati i campioni più vicini a ognuna delle frequenze trovate per le 14 canne, questi sono stati opportunamente portati al pitch corretto con il software Melodyne. Per ogni canna sono stati adattati due diversi tipi di campioni: un campione lungo (circa 8 secondi) e uno "flatter". I campioni vengono gestiti con Ableton Live 9, che grazie alla mappatura quasi totale dei propri controlli, il software consente di modifi-

care facilmente i parametri interni grazie ai messaggi ricevuti in input. Il progetto Ableton, creato per la gestione dei suoni del flauto, è costituito da due canali *multirack*, uno per la sezione "soffio" dell'installazione e l'altro per quella touch. In ognuno di questi canali sono state inserite le singole tracce audio, assegnate ognuna ad una differente nota del *Piano Roll*. Per la sezione touch, vengono usati solamente i messaggi fondamentali di *NoteOn* e *NoteOff*. Il *multirack* relativo alla parte sensorizzata dell'installazione, invece, è caratterizzato dalla presenza di 4 *samples* (i primi tre derivanti dal campione "lungo", mentre l'ultimo da quello "flatter") per aggiungere definizione all'attacco, alla quantità d'aria usata e alla saturazione del suono (gestite dai messaggi di *ControlChange*). Questi sono:

Attack - usato per enfatizzare l'attacco del suono, ed esaltarne il suo classico suono "T";

Air - necessario per "arricchire" il suono dopo l'attacco;

Note - il campione originale;

Flutter - per enfatizzare i picchi di intensità del soffio.

II messaggi di *ControlChange*, sono stati mappati su una macro del canale *multirack* che controlla il parametro *Chain* dello stesso. Così facendo, a seconda del parametro ricevuto, *Chain* varierà dando più o meno enfasi alle tracce aggiuntive per interpretare al meglio l'intensità del soffio. Grazie a una scheda audio esterna e a una soundbar Bose Solo5, posizionata in una fessura nella parte alta dell'installazione, il visitatore potrà ascoltare il rendering sonoro della canna desiderata.

Per rendere più coinvolgente l'installazione e per arricchirne l'esperienza interattiva, è stato implementato un feedback di tipo visivo, oltre che sonoro. Ecco perché, all'interno delle canne stilizzate del flauto, sono state aggiunte delle strisce LED di colore bianco che si accendono e variano la loro luminosità a seconda dell'intensità del soffio. L'implementazione sfrutta le porte digitali di Arduino, usabili come modulatori PWM: l'intensità luminosa delle strisce LED varia in maniera inversamente proporzionale alla lunghezza dell'intervallo in cui le porte forniscano tensione alle luci. Per quanto riguarda l'utilizzo della modulazione su un singolo LED, Arduino è perfettamente in grado di fornire sufficiente corrente (40 mA massimi su ciascuna porta). Per strisce LED di lunghezza variabile tra gli 70 cm e i 40 cm, invece, la corrente fornita dalla scheda è insufficiente. Per risolvere il problema è stato aggiunto un trasformatore esterno adeguatamente tarato al consumo energetico delle singole parti. Per poter utilizzare la modulazione PWM con l'alimentatore esterno, è stato utilizzato un BJT che passa dalla zona di "interdizione" a quella di "conduzione", ricevendo tensione alla base dai pin digitali di Arduino, che si occupano della modulazione.

Come spiegato nella sottosezione precedente, Arduino gestisce anche la concorrenza tra le due parti dell'installazione. La riproduzione del suono proveniente da sorgenti diverse o da più di una canna contemporaneamente non è consentita. Per far capire questa regola ai visitatori, al momento della riproduzione di un suono dalla sezione "soffio", verrà colorata la canna equivalente sullo schermo e vi-

ceversa verrà accesa la striscia LED equivalente. La gestione di questi processi è affidata a uno script che si occupa della comunicazione seriale e che gestisce di conseguenza anche i flag necessari ad impedire la contemporanea riproduzione del suono. Anche Arduino ha nel suo codice una sezione delegata alla gestione delle informazioni ricevute ed inviate ad Unity per la gestione della concorrenza.

5. CONCLUSIONI

Il lavoro presentato è parte di un progetto più ampio, che ha coinvolto un team di persone con competenze estremamente eterogenee. La multidisciplinarità è una delle principali peculiarità di questo progetto, che vuole porsi come esempio per progetti simili che mirano alla modellazione e alla valorizzazione di uno strumento musicale. Questo flauto antico è stato uno caso di studio che, per la sua intrinseca complessità, ha permesso di affrontare e risolvere un notevole numero di problematiche in svariati settori scientifici. Gli autori sono certi che le metodologie sviluppate in questo lavoro potranno quindi essere da esempio, o perlomeno un punto di partenza per progetti simili che coinvolgano strumenti musicali di tutte e tre le categorie presentate nella sezione 1. Lo stesso vale per progetti che vogliono affrontare il tema della conservazione attiva di strumenti musicali.

Allo stesso tempo il progetto apre diversi interrogativi che si concretizzeranno in svariati futuri sviluppi. Per affinare la ricostruzione del flauto virtuale è previsto, nel breve termine, uno studio dell'imboccatura del flauto, per comprendere come la sua forma modifichi il suono prodotto. Ulteriori analisi saranno poi effettuate per capire la composizione del materiale depositato nel fondo di alcune canne del flauto. Inoltre, si cercherà di ottenere stime più precise dell'intonazione del flauto, unendo i dati raccolti con elementi di teoria musicale della Grecia antica. Per validare le modalità di interazione con l'installazione musicale è prevista una fase di test di usabilità che coinvolgerà alcuni soggetti sperimentali.

Nel medio termine, è prevista la sintesi del suono del flauto attraverso un modello fisico, che andrà a sostituire l'attuale realizzazione per campionamento. L'obiettivo è rendere più precisa la ricostruzione e aumentare le modalità di interazione e le capacità espressive. Quest'ultimo obiettivo è comune per il futuro progetto che prevede di realizzare una copia del flauto sensorizzata. Una stampa 3D delle stesse dimensioni dell'reperto e l'aggiunta di sensori quali giroscopio e accelerometro, consentirebbe di rilevare nuove *feature*, come ad esempio l'inclinazione del flauto, che influenza la resa sonora ed è alla base di alcune comuni tecniche utilizzate dai musicisti che suonano il flauto di Pan. Quest'ultimo futuro sviluppo, inoltre, entrerebbe di diritto in un ambito di ricerca sempre più di tendenza: la stampa 3D di strumenti musicali [13].

6. RINGRAZIAMENTI

Questo lavoro è parzialmente supportato dal progetto *Archaeology & Virtual Acoustics - A Pan flute from ancient Egypt* sovvenzionato dall'Università di Padova con il finanziamento no. CPDA133925

7. BIBLIOGRAFIA

- [1] J. H. Falk and L. D. Dierking, "School field trips: Assessing their long-term impact," *Curator: The Museum Journal*, vol. 40, no. 3, pp. 211–218, 1997.
- [2] A. Antoniou and G. Lepouras, "Modeling visitors' profiles: A study to investigate adaptation aspects for museum learning technologies," *J. Comput. Cult. Herit.*, vol. 3, pp. 7:1–7:19, Oct. 2010.
- [3] A. Rigolli, P. Russo, and I. L. Casa della musica (Parma, *Il suono riprodotto: storia, tecnica e cultura di una rivoluzione del Novecento : atti del convegno annuale del Laboratorio per la divulgazione musicale : Parma, 10 e 11 novembre 2006*. Giornale della musica, EDT, 2007).
- [4] F. Avanzini and S. Canazza, "Virtual analogue instruments: an approach to active preservation of the studio di fonologia musicale," in *The Studio di Fonologia - A Musical Journey* (M. Novati and J. Dack, eds.), pp. 89–108, Milano: Ricordi (MGB Hal Leonard), June 2012.
- [5] S. Canazza, C. Fantozzi, and N. Pretto, "Accessing tape music documents on mobile devices," *ACM Trans. on Multimedia Comput., Commun. Appl.*, vol. 12, no. 1s, p. 20, 2015.
- [6] F. Bressan and S. Canazza, "A systemic approach to the preservation of audio documents: Methodology and software tools," *Journal of Electrical and Computer Engineering*, vol. 2013, p. 21 pages, 2013.
- [7] F. Avanzini, S. Canazza, G. De Poli, C. Fantozzi, N. Pretto, A. Rodà, I. Angelini, C. Bettineschi, G. Deotto, E. Faresin, A. Menegazzi, G. Molin, G. Sallemi, and P. Zanovello, "Archaeology and virtual acoustics. a pan flute from ancient Egypt," in (*SMC2015*), (Maynooth), pp. 31–36, July 2015.
- [8] N. Simon, *The participatory museum. Museum 2.0*, 2010. Creative Commons.
- [9] N. H. Fletcher and T. D. Rossing, *The physics of musical instruments*. New York: Springer-Verlag, 1991.
- [10] S. Hagel, *Ancient Greek music: a new technical history*. Cambridge University Press, 2009.
- [11] M. Russ, *Sound Synthesis and Sampling*. Music technology series, Elsevier/Focal Press, 2004.
- [12] A. Okita, *Learning C# Programming with Unity 3D*. CRC Press, 2015.
- [13] A. Zoran, "The 3d printed flute: Digital fabrication and design of musical instruments," *Journal of New Music Research*, vol. 40, no. 4, pp. 379–387, 2011.

DYNAUDIOROUTER: SOFTWARE OPEN SOURCE PER IL ROUTING AUDIO DINAMICO IN MAX/MSP

Riccardo Franceschini

Ricercatore indipendente

franceschiniriccardo@gmail.com

Giovanni Sparano

Ricercatore indipendente

info@giovannisparano.it

SOMMARIO

Questo articolo presenta l'implementazione software di un sistema automatizzato di routing audio dinamico. In base a specifiche politiche di gestione, controllabili dall'utente, e al segnale audio in ingresso al sistema, *DynAudioRouter* è in grado di reindirizzare dinamicamente i segnali in ingresso attivi verso le uscite. Nato per un contesto di improvvisazione tra strumenti acustici e *live electronics*, il sistema è scalabile e *open source*, quindi facilmente adattabile a situazioni diverse quali la spazializzazione del suono, le installazioni multimediali e il mixaggio.

1. INTRODUZIONE

Il 28 maggio 2015, nei cortili di Palazzo Pisani di Venezia, si è tenuta la performance del progetto collettivo AQUA, lavoro multidisciplinare di teatro, danza, voci, strumenti e *live electronics*. Evento collaterale dell'esposizione Aquae Venice 2015, è stato realizzato dalle scuole di Composizione e Musica elettronica del Conservatorio "B. Marcello" di Venezia¹. La partitura dell'opera collettiva prevedeva un percorso tra cinque differenti postazioni strumentali e una modalità esecutiva semi-aleatoria, controllata da due direttori d'orchestra secondo la tecnica del *conducting* [1]. Gli autori hanno partecipato ad AQUA in qualità di regista del suono (Sparano) e di esecutore al *live electronics* (Franceschini). In questo contesto, data l'impossibilità di conoscere un preciso attacco strumentale e di conseguenza l'utilizzo o meno di un dato microfono, gli autori hanno realizzato una prima versione di *DynAudioRouter*, software capace di gestire automaticamente e in modo dinamico il *routing* di un numero di ingressi variabile nel tempo verso un numero di uscite predeterminato, garantendo agli esecutori al *live electronics* un flusso sonoro costante da poter elaborare.

Nel contesto originale di sviluppo di *DynAudioRouter* erano previste come sorgenti sonore dei microfoni, mentre le versioni successive del software hanno allargato lo spettro delle possibilità di utilizzo a qualsiasi tipo di sorgente sonora.

¹ Presentazione dello spettacolo all'indirizzo web <http://www.aquae2015.org/cose/eventi-collaterali/aqua-il-progetto-collettivo-del-conservatorio-b-marcello-di-venezia>

Copyright: ©2016 Riccardo Franceschini, Giovanni Sparano. 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. IMPLEMENTAZIONE

Il software di routing variabile è stato implementato grazie all'ausilio di *Max/MSP*, linguaggio di programmazione visuale sviluppato da Cycling '74 [2]. La *patch* realizzata non è compilata e viene distribuita gratuitamente sotto licenza Creative Commons CC BY-NC² su richiesta agli autori, permettendo modifiche personalizzate o l'integrazione in altre *patch*. L'implementazione sfrutta principalmente lo scripting nativo di *Max/MSP* per la generazione dell'ambiente di controllo, per consentire così di preimpostare il sistema secondo le proprie esigenze. L'utente può modificare sia il numero di ingressi, sia il numero di uscite massimi, aggiornando in tempo reale il numero di strisce di canale visibili (in ingresso e uscita) nell'interfaccia grafica.

Le prime tracce di un sistema per il mixaggio automatico risalgono al 1975 [3, 4] e permetteva la correzione automatica del guadagno di una serie di ingressi microfonici in un mixer. Un approccio simile a quello degli autori è riscontrabile in un software sviluppato da François Pachet e Olivier Delerue chiamato *MusicSpace*, che permette all'utente di scegliere tra diversi punti d'ascolto attraverso un'interfaccia grafica modificabile in tempo reale [5].

2.1 Schema generale

DynAudioRouter è costituito da un massimo di N ingressi ed M uscite, entrambi variabili da 1 a 32 canali, modificabili secondo le necessità dell'utente.

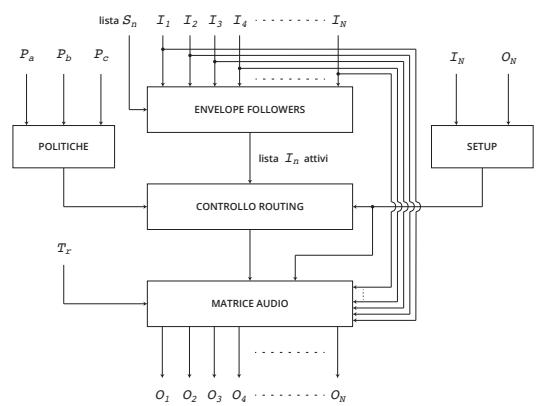


Figura 1. Schema generale

² Descrizione della licenza all'indirizzo <https://creativecommons.org/licenses/by-nc/4.0/> web

A valle degli ingressi è posto un *envelope follower* che determina l'inviluppo dei singoli segnali [6], assegna a ognuno uno stato (*attivo* o *non attivo*) nel caso in cui superino una soglia arbitraria in dB definita dal parametro S_n , diverso per ogni canale in ingresso. Infine, genera una lista binaria di segnali attivi fornita al blocco *CONTROLLO MATRICE*, che si occupa del routing a seconda delle condizioni iniziali definite dall'utente. L'inviluppo dei segnali in ingresso al blocco *ENVELOPE FOLLOWER* viene calcolato determinando il valore quadratico medio in un intervallo di 100 campioni. Questo dato viene successivamente filtrato logaritmicamente secondo la formula:

$$y_n = y_{n-1} + \frac{x_n - y_{n-1}}{\alpha}, \quad (1)$$

dove α è il numero di campioni di smussamento. Per valori di $y_n > y_{n-1}$, il numero di campioni è 100, mentre per $y_n < y_{n-1}$ è pari a 10000. Questi valori, arbitrariamente scelti sulla base di test del dispositivo, impediscono continui cambi di routing dovuti a valori in ingresso prossimi alla soglia (vedi fig. 1).

Il routing verso le uscite dei canali in ingresso attivi varia a seconda della differenza tra canali in ingresso e in uscita, e dal valore dei tre parametri della politica di controllo $\{P_a, P_b, P_c\}$. Il tempo di transizione da una configurazione a un'altra della matrice audio è regolato da un'apposita variabile T_r . Il fading è di tipo lineare e i tempi assegnabili variano nell'intervallo [10, 1000] in millisecondi.

Il blocco denominato *SETUP* si occupa del ridimensionamento della matrice audio, attuato sulla base dei parametri I_N e O_N . Inoltre, fornisce al blocco *CONTROLLO ROUTING* il primo condizionamento del routing sulla base del numero di ingressi e di uscite.

2.2 Routing dinamico

La tabella 1 rappresenta l'albero delle possibilità di routing sulla base dei diversi valori delle variabili di controllo e del numero di canali in ingresso (N) e in uscita (M) e illustra la struttura gerarchica dei parametri della politica di controllo.

$N \leq M$			$N > M$		
P_a		P_a		P_a	
unitario		mix		mix	
P_b	P_b	P_b	P_c	P_c	P_c
cresc.	decresc.	casuale	uni.	non uni.	uni.
					non uni.

Tabella 1. Albero delle possibilità di routing

P_a , P_b e P_c rappresentano un ulteriore livello di condizionamento del routing, come i parametri I_N e O_N del blocco *SETUP*, e possono assumere i seguenti valori:

- P_a - unitario, mix;
- P_b - crescente, decrescente, casuale;
- P_c - uniforme, non uniforme.

Nella condizione in cui $N \leq M$, le possibilità che si presentano sono $P_a = \text{unitario}$ e $P_a = \text{mix}$. Per $P_a = \text{unitario}$, un singolo segnale di ingresso è assegnato a una singola uscita. Per $P_a = \text{mix}$, tutti i segnali in ingresso mixati tra loro sono assegnati alla singola uscita. L'ordine di assegnazione degli ingressi alle uscite e la definizione del tipo di mix sono affidati ai valori assegnati, rispettivamente, ai parametri P_b e P_c . Nella tabella 2 viene illustrato un esempio di routing di 4 canali in ingresso verso 6 uscite per $P_b = \text{casuale}$.

	I_1	I_1	I_3	I_4
O_1	•			
O_2		•		
O_3			•	
O_4				•
O_5	•			
O_6		•		

Tabella 2. Routing con ordinamento crescente

Per $P_b = \text{decrecente}$ l'ordine sarebbe inverso, quindi I_1 sarebbe assegnato a O_6 , I_2 a O_5 , ecc., mentre per $P_b = \text{casuale}$, l'algoritmo assegna casualmente le uscite, esaurendo tutti gli ingressi disponibili. L'assegnazione casuale è senza ripetizioni e, una volta esaurite le possibilità, l'algoritmo riassegna casualmente altri ingressi alle uscite rimanenti. Nella tabella 3 viene illustrato un esempio di routing di 4 canali in ingresso verso 6 uscite per $P_b = \text{crescente}$.

	I_1	I_1	I_3	I_4
O_1			•	
O_2	•			
O_3				•
O_4		•		
O_5			•	
O_6		•		

Tabella 3. Routing con ordinamento casuale

Il parametro P_c rappresenta la tipologia di mix per $P_a = \text{mix}$, sia per $N \leq M$, sia per $N > M$.

Per $P_c = \text{uniforme}$, l'ampiezza dei singoli ingressi viene moltiplicata per un fattore k , pari a $1/N$, dove N rappresenta il numero totale di ingressi, definiti dal parametro I_N dichiarato per il blocco *SETUP*. Questo garantisce che la somma dei fattori di guadagno di tutti gli ingressi che vanno a comporre il mix sia pari a 1.

Per $P_c = \text{non uniforme}$, l'inviluppo dei singoli ingressi viene moltiplicata per un numero reale casuale compreso tra 0 e 1. Definiamo questo numero come fattore α_n , diverso per ogni ingresso. La somma di tutti i fattori di guadagno è sempre pari a 1, secondo la definizione:

$$\sum_{n=1}^N \frac{\alpha_n}{\beta} = 1, \quad (2)$$

dove β è definito come:

$$\beta = \sum_{n=1}^N \frac{\alpha_n}{N}. \quad (3)$$

3. CONCLUSIONI E LAVORI FUTURI

Grazie all'analisi delle sorgenti effettuata con l'*envelope follower* e l'applicazione delle politiche di routing controllate dall'utente, il software è in grado di calcolare i percorsi di instradamento in modo autonomo, dinamico e in tempo reale.

Data la natura *open source* del progetto, il software ben si presta a ulteriori sviluppi e adattamenti a specifiche esigenze, sia tecniche, sia artistiche e i diversi campi di applicazione del software (installazioni multimediali, spazializzazione del suono, ecc.) necessitano configurazioni ed eventuali personalizzazioni del codice ben diverse tra loro.

Gli sviluppi futuri del software da parte degli autori saranno rivolti all'ampliamento dell'autonomia dell'algoritmo nella gestione del routing nell'analisi più accurata e diversificata delle sorgenti in ingresso.

Integrando all'*envelope follower* una serie di descrittori audio [8], sarà possibile definire come attiva una sorgente sonora che presenti caratteristiche precise, non esclusivamente legate alla potenza media del segnale. La gestione della soglia verrà automatizzata grazie a un algoritmo che segue la logica *fuzzy* [9], rendendone dinamica la definizione in base all'ingresso e non più a un valore arbitrariamente scelto a priori. Inoltre, sono previste modifiche alle modalità di assegnazione degli ingressi verso le uscite, consentendo tipologie miste di routing, rendendo possibile la creazione di gruppi di uscite diversificati per contenuto e ordinamento.

4. BIBLIOGRAFIA

- [1] T. T. Stanley, "Butch morris and the art of conduction®," in *PhD Diss. University of Maryland, College Park*, 2009.
- [2] "<https://cycling74.com/>"
- [3] R. E. Gonzalez E. P., "Automatic mixing: live downmixing stereo panner," in *Proceedings of DAFX-07*, (Bordeaux), 2007.
- [4] D. D., "Automatic microphone mixing," in *Proceedings of the Audio Engineering Society AES Convention*, (San Francisco), 1975.
- [5] D. O. Pachet F., "On-the-fly multi track mixing," in *Proceedings of the Audio Engineering Society 109th Convention*, (Los Angeles), 2000.
- [6] M. Puckette, *The Theory and Technique of Electronic Music*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2007.
- [7] J. M. McQuillan, "The birth of link-state routing," *IEEE Ann. Hist. Comput.*, vol. 31, pp. 68–71, Jan. 2009.
- [8] E. M. Malt, Jourdan, "Zsa.descriptors: a library for real-time descriptors analysis," in *5th Sound and Music Computing Conference*, (Berlin), 2008.
- [9] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

BaO: THE BABY ORCHESTRA

Sonia Cenceschi

Politecnico di Milano
sonia.cenceschi@polimi.it

Maria Teresa Trecozzi

Centro di Musica Contemporanea
mariateresa3@hotmail.it

ABSTRACT

The BaO project is an immersive environment for children (up to 7-8 years) born with the purpose to get them interested into the complexity of orchestral music by playing and using, in addition to hearing, touch and sight. The fulcrum of the interaction are many colourful sofa pillows, positioned in a comfortable and protected space, where up to 4-5 children and their parents (if needed) can be hosted. Each pillow activates the playback of a recording when handled or pressed. Recordings consist of musical passages corresponding to individual instruments or orchestral families, whose sound are transmitted with quadraphonic speakers. When the music starts, videos of the corresponding musicians in action are screened. The aim is to stimulate both an active and passive listening and learning (thanks to peripheral attention) of complex musical contents, taking advantage of the natural curiosity and inclination of children for manipulation and use of senses in general.

1. INTRODUCTION

BaO is born from the desire to bring children to the orchestral world. The idea of BaO is not to force the child, but to offer them a chance to play and at the same time also to do so, according to the desires of the moment. Music and video continue to run in the background and offer again and again an opportunity for growth thanks to peripheral attention.

BaO proposes a classical musical tradition (modern orchestra), from the seventeenth century, but plans to extend listening to other music and styles, adapting well to different musical cultures areas, through the replacement of audio-visual materials.

2. PEDAGOGICAL ASSUMPTIONS

With BaO it is assumed that immersion in a complex environment is not a limit, but rather an added value for the musical training at an early age [1]. The child can not access music from the executive point of view, but we think that stimulating listening can help with

Copyright: © 2016 Cenceschi S. 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.

developing an accustomed ear and a familiarity with these forms of art [2]. Classical music is largely studied in music pedagogy [3][4] and the same thing can be said for the use of multimedia and physical interaction (some examples are [5][6][7][8]). BaO passes from theory to practice by providing a modular installation in a suggestive environment. Furthermore, authors who have been music teachers in nursery schools, based on their experience, are in favour of an interactive approach and are acquainted with the awareness of children's ability to absorb from complex environment.

3. HOW IT WORKS

The audio chain starts from the pillow being shaken or pushed by the child. The activated analogical signal reaches the Arduino board using the Wi-Fi transmitter and goes to the computer. Max/MSP software processes information and simultaneously activates the sound corresponding to the pillow and selects a video.

The audio passes to the sound card and to the sound system while the video is sent to the projector. The video-screen is divided into as many squares as the musicians involved in the part: if the pillow relates to a class of instruments, they appear naturally together.

The pillow works essentially as simple switches when pushed or shaken. This is just the first prototype: we plan to add more actions in the future.

BaO can be used in two modalities, according to the typologies of audio you desire to activate (The materials will be those collected in chapter 7):

- Composition mode
- Orchestral mode

3.1 The Composition mode

Each pillow corresponds to one instrument. It plays a predetermined loop in a particular tonality. All pillows correspond to tonalities that fit together.

We can choose whether to play simple or complex melodies (example a continuous C bass or a sequence in the key of C major). It involves each instrument playing a simple melody for a prototype duration of about 1 minute.

Musical tonalities are calculated, making it possible to achieve pleasing overlapping of different instruments and loops.

3.2 The Orchestral mode

It is based on an opera. The main instrument categories can be activated with the 4 main pillows (for example red) and others will add smaller sub-groups of

instruments. It will be possible to use only orchestral groups, or also smaller ones. The opera starts and proceeds in a continuous cycle, but only audio depending on handled pillows are activated.

When a pillow is pressed/shackled a first time the corresponding music start, while when it's handled for the second time, the corresponding instruments are silenced with a short fadeout. The idea is (if enough resources are available) to furnish all single recordings, in order to permit the personalization in Max/MSP of associating each pillow to a single instrument.

4. MAIN PROBLEMS

The main problem is the isolation of sensors, both for the safety of children and to preserve the installation. Sensors and transmitters will be inserted in plastic boxes padded with polystyrene to obtain more resistance and to permit the activation with the weight of the body or the shaking.

All cables are placed close to the computer and Arduino, away from the children's action space. Several measures have also been planned to overcome problems related to the managing of audio-video samples. Some of them are listed above.

In the *Composition mode*:

- When an instrument finishes its part of the instrumentalist video, the image remains stationary for a few seconds (duration to test).
- If a pillow is activated twice in a row, the second selection starts another instrument of the same family, so not to overlap the same part.
- When a loop stops, we plan a simple linear fade out (about 2 seconds). It's not a problem when many children are playing together because all the sounds are combined in a gradable harmony.

In the *Orchestral mode*:

- If an instrument or sub-group is activated after its orchestral group, its sound will be transmitted with a higher volume to make it emerge from the background.
- If children do not touch pillows, the volume goes down after an established period of time and instruments enter one by one until they get to complete the orchestra.

In both cases if a child shakes, for example, a pillow for a long time, it just switches ON once: every movement followed by immobility counts as one command.

5. TESTING

The BaO prototype will be tested on small groups (3 to 4 total) of different ages in order to analyze the best conditions, features and modality for each of them.

It would be better to involve children in the design phase, but it's really difficult to reach a significant number due to their young age. At the same time, it's difficult to get the time and availability of the musicians to make many adjustments and recordings.

6. THE CHILDREN'S EXPERIENCE

6.1 The environment

- BaO aims to become an experimental environment for children. Pillows create the interaction, but also contribute to make the space comfortable: the room must be a place where the child feels at ease. It must also be recognized as a unique place by children, a place where they can feel free to play with instrumental sounds and imagination. It could be better to install BaO in a closed or protected space, lending itself to be easily and tastefully furnished.

6.2 The BaO's sessions

- We have envisioned a long session (e.g. 2 hours) or a long use over time when it's possible (e.g. 1 hour per week): the attention of children is short but BaO is a place where it is possible to relax or to learn passively.

Children should develop a personal taste and associate to the BaO environment a personal meaning: we are not directly interested in the notions knowledge but in the development of a general sensibility in complex music sounds and structures.

6.3 Adult's role

The basilar idea is that adults supervise from outside. Children (a small group of 3 or 4) are left by themselves in the BaO's space. At the beginning children are invited to look around and just to play with pillows. We plan to start with the Orchestral mode, but adults will not explain to them the functioning and _ link of objects with music-video changes.

We know children are very perceptive and –maybe depending on age in different ways– any explanation could be useless because they'll learn by themselves the BaO functioning. The purpose is not to supply a code but to allow each child to discover and to invent their own.

6.4 The interaction

- During the experience the child can sit on, lift, shake, crush or cuddle each pillow and discover new combinations. After the phase of discovering, a second one follows. It includes the adult giving some information (gradually increased and personalized) about instruments, music or handlings. The entire process can be replicated for the *Composition mode* as well.

7. STAFF AND SETTINGS

7.1 Musical materials

As first proposal we plan to start (*Orchestral mode*) by extracts of famous opera orchestral overtures: *Les Folies d'Espagne* by Jean-Baptiste Lully and Symphony number 6 (Pastoral) by Beethoven. They will be

reproduced and recorded by a real orchestra: we'll record both single instrumental parts, and the main groups of family instruments: string instruments, woodwind, brass and percussion.

During the single session-recording we plan also to take samples of instruments to use in the *Composition modality* of BaO. In this context we will start video recordings in order to obtain materials for the visual representation.

In 7.2 and 7.3 we indicate the necessary equipment to set up the space (the recording equipment is not specified, assuming it is already owned).

7.2 Technical Materials

- Software: Max/MSP (Cycling '74) and Jitter
- Laptop + Digital projector + screen
- 4 loudspeakers + cable + digital soundcard
- Arduino MEGA2560 + usb serial adapter
- 10 pressing sensor + 10 gyroscopic sensors and 20 Wireless RS232 modules (the number can be variated on the base of the version)

7.3 setting up of the room

- Carpets
- Toy instruments and other non-hazardous furnishing elements
- 10 or more sofa pillows for the interaction: 4 for main orchestra families and others for other groups and single solo instruments.

8. ACTUAL LEVEL OF DEVELOPMENT

The project has gathered a lot of interest and we are in contact with several nursery schools in the Como and Milan province for possible installations. The response of the authorities in general is positive and we are currently looking for funds to carry out the installation and make it freely available, overcoming the financial problems of the school system and the economic barrier for families.

We are currently actively working on the implementation with plans to have ready by September the connection for software, sensors and video. Afterwards, we can mount and define the physical environment quickly enough once the materials are bought. The Conservatory of Como or other interested institutes are expected to be involved for audio-video recordings. We have already estimated the outlay, planning different costs according to different financial availability and different space size.

9. REFERENCES

- [1] L. Sbattella: *Ti penso, dunque suono. Costrutti cognitive e relazionali del comportamento musicale: un modello di ricerca-azione*, Vita e Pensiero, 2013.
- [2] K. L. Hyde et al.: "Musical Training Shapes Structural Brain Development," *The Journal of Neuroscience*, Vol.29, N.10, pp. 3019-3025, 2009.
- [3] D. Campbell, *The Mozart Effect for Children: Awakening Your Child's Mind, Health, and Creativity with Music*, William Morrow Paperbacks, 2002.
- [4] B. Haines et al., *Leading Young Children to Music*, Pearson, 1999.
- [5] L. Min, "An Exploratory Study of How Pre-Kindergarten Children Use the Interactive Multimedia Technology: Implications for Multimedia Software Design," *Journal of Computing in Childhood Education*, Vol.7, pp. 71-92, 1999.
- [6] E. Lee, "You're the conductor: a realistic interactive conducting system for children," *New Interfaces for Musical Expression Conference*, 2004.
- [7] R. Dillon, "Virtual Orchestra: an immersive computer game for fun and education," *CyberGames*, pp.215-218, 2006.
- [8] B. Bruegge et al, "Pinocchio: conducting a virtual symphony orchestra," *International conference on Advances in Computer Entertainment Technology*, 2008.

A SEA OF STARS FOR KM3NET: A MULTIMODAL VIRTUAL REALITY EXPERIENCE

Alba Francesca Battista

Conservatorio «D. Cimarosa», Avellino - Italy
alba.battista@conservatoriocimarosa.org

Carlos Maximiliano Mollo

INFN Istituto Nazionale di Fisica Nucleare
Naples, Italy
maximil@na.infn.it

ABSTRACT

The collaboration team of KM3NeT (*Cubic Kilometer Neutrino Telescope*) asks us to realize an immersive experience to promote their activities in an uncommon way. In this paper, we describe how we used their scientific data and their incredible sounds recorded by their detectors at 3000 meters deep to create a multimodal virtual reality installation. Users can experience a Detector Unit (DU) deployment in first person by wearing Oculus Rift with Leap Motion and through a 3D audio spatialization. This experience works as a video-game. To realize it, we use SCI – model by Laura Ermí and Frans Mäyrä – and IEZA model to relate the gamer and the game and MAXMSP as audio engine.

1. INTRODUCTION

KM3NeT (*Cubic Kilometer Neutrino Telescope*) is an European research collaboration and future infrastructure that will be located at the bottom of the Mediterranean Sea. It will host the next-generation neutrino telescope in the form of a water Cherenkov detector with an instrumented volume of about five cubic kilometers distributed over three locations in the Mediterranean: KM3NeT-Fr (off-shore Toulon, France), KM3NeT-It (off-shore Portopalo di Capo Passero, Sicily, Italy) and KM3NeT-Gr (off-shore Pylos, Peloponnese, Greece). KM3NeT will search for neutrinos from distant astrophysical sources like supernova remnants, gamma-ray bursts, supernovae or colliding stars and will be a powerful tool in the search for dark matter in the universe. Its prime objective is to detect neutrinos from sources in our galaxy. Arrays of thousands of optical sensor modules will detect the faint light in the deep sea from charged particles originating from collisions of the neutrinos and the water or rock in the vicinity of the detector.

The research infrastructure will also house instrumentation for other sciences like *marine biology*, *oceanography* and *geophysics* for long-term and on-line monitoring of the deep-sea environment and the sea bottom at depth of several kilometers. Once completed, the

Copyright: © 2016 Battista A. F. 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.

KM3NeT research infrastructure will consist of seven large subdetectors. Six of these will indeed form the telescope searching neutrino-sources. The seventh sub-detector will be optimized to measure the properties of the neutrino itself. In that sense, the latter detector is a neutrino particle physics detector. [1]

While it is well-known in the scientific community, this extraordinary setup, that will be the largest neutrino detector of the northern hemisphere, is not so familiar to public as desired. So, the collaboration team of KM3NeT asks us to realize an immersive experience to promote their activities.

We use virtual reality to make the user able to appreciate the gigantic structure of the neutrino telescope. In addition, the user can interact with the detector in order to have information about the components of the detector it self.

2. A SEA OF STARS: KM3NET EXPERIENCE

A sea of stars is an interactive sound installation. It integrates data sonification, interactivity, and sound spatialization.

The data used in the work comes from KM3NeT detectors. This data was collected from two of the detector units (DUs) currently activated. DUs are one kilometer string made up mainly of optical sensors (photomultiplier tube or PMT) and acoustical sensors. PMTs continuously send their electrical states; acoustical sensors, instead, record the marine life at 3000 meters deep. DU structures are very complex devices, but for the purposes of this paper this information can be enough.

In *A sea of stars*, all the sounds come from acoustical sensors of KM3NeT. They form an unbelievable soundscape which is constantly modulated by the *performer* (the user him/herself) through his/her head and hands movements. These mostly act on dynamic and spazialization.

PMTs data are used to manage granularization and spectral spatialization of every original sample, that is activated by some particular gestures of the user.

Virtual reality allows to amplify this sonic experience adding an immersive video dimension.

3. HARDWARE AND SOFTWARE FOR THE VR SCENES

Virtual reality is a powerful mean to extend interaction: it can simulate an environment which creates an illusion that we are present somewhere we are not. This can be achieved by providing realistic inputs of the virtual environment to our senses. Visual and aural senses play the most prominent role in the awareness of the surrounding for human being.

In recent years, prospects of virtual reality have not just become possible but it has become the most anticipated technology of this generation because of advancement in computing power, communication technology and the increasing miniaturization in electronics. Research in signal and image processing is enabling the development of many virtual reality (VR) applications in gaming and entertainment.

The VR devices used to implement the KM3NeT experience are Oculus Rift DK2 [2] and the Leap Motion [3]. The Oculus Rift DK2 is a 3D Headset coupled with an inertial and optical tracking system that was used to provide a three-dimensional, immersive and interactive visualization of the 3D environment. The Leap Motion, instead, is an optical hands tracker that was used to allow the users to interact with the game objects directly by using his hands. The Leap Motion was fixed on the top of the Oculus Rift DK2 and, thanks to its wider field of view (Figure 1), the hands are always tracked wherever the user is looking at.



Figure 1. The comparison between Oculus and Leap Motion field of view [4].

The software application, managing the VR devices, was implemented by using UNITY (Figure 2) [5]. Unity is a game development platform and real-time rendering engine widely used in several fields to develop 3D applications and interactive VR environments.

One of the newest virtual reality technologies was used for this project: the Oculus Rift Development Kit 2 (DK2) integrated with the Leap Motion.

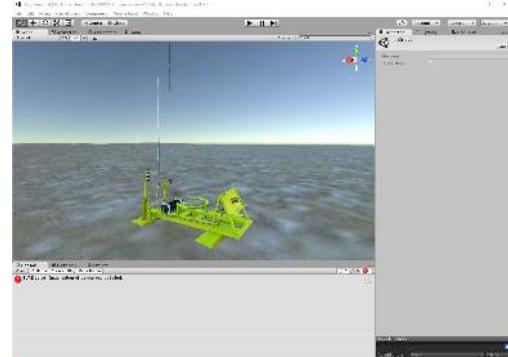


Figure 2. Unity platform.

3.1 Hardware specifications

3.1.1 Oculus Rift DK2

The Oculus Rift DK2 Rift is a virtual reality head-mounted display featuring a resolution of 960×1080 per eye, low-persistence AMOLED display designed to use low-persistence-of-vision to smooth motion, as well as a full positional tracking system [6]. The DK2 uses an optical camera based tracking system which together with the on board IMU provides the ability to track rotational and translational movement using on board IR LEDs which sit behind an IR transparent shell on the front and sides of the DK2. This means that (within the camera's field of view) the system knows where the head is in 3D space. It is possible to move the head up and down, downwards and backwards relative to the camera and a DK2 compatible application can adjust the view appropriately, allowing leaning in and out of a scene or lean left and right.

3.1.2 The Leap Motion

The Leap Motion senses how the hands naturally move and permits to point, wave, reach, grab objects in the virtual environment. The heart of the device consists of two cameras and three infrared LEDs. These track infrared light with a wavelength of 850 nanometers, which is outside the visible light spectrum. Thanks to its wide-angle lenses, the device has a large interaction space of eight cubic feet, which takes the shape of an inverted pyramid – the intersection of the binocular cameras' fields of view. The Leap Motion Controller's viewing range is limited to roughly 60 cm above the device, limited by LED light propagation through space, since it becomes much harder to infer the hands position in 3D beyond a certain distance. LED light intensity is ultimately limited by the maximum current that can be drawn over the USB connection. At this point, the USB device controller reads the sensor data into its own local memory and performs any necessary resolution adjustments. This data is then streamed via USB to the Leap

Motion tracking software. The data takes the form of a gray scale stereo image of the near-infrared light spectrum, separated into the left and right cameras. Typically, the only objects you see are those directly illuminated by the Leap Motion Controller's LEDs.

3.1.2.1 Leap Motion vs Microsoft Kinect

The Microsoft Kinect sensor was developed with hand/arm and full-body gesture recognition in mind. In the paper [7] authors provide a detailed analysis of the accuracy and resolution of the Kinect sensor's depth data for indoor mapping applications. The experimental results show that the random error in depth measurement increases with increasing distance to the sensor, ranging from a few millimeters to approximately 4 cm at the maximum range of the sensor. The quality of the data is also found to be influenced by the low resolution of the depth measurements. The obtained accuracy is, in general, sufficient for detecting arm and full body gestures, but is not sufficient for precise finger gestures such as handwriting. The input device latency and spatial jitter are also important factors [8].

The Leap Motion Controller, instead, presents a milestone in consumer finger/object and gesture tracking input technology. The device itself was made publicly available in summer 2013, and therefore not much scientific work has been published yet. In [9], the authors describe an application of the Leap Motion Controller for the direct manipulation of an industrial robot arm with six axes of freedom. The Leap Motion Controller is used for finger position tracking. To increase the tracking precision, an interpolation of the acquired data is performed using polynomial splines. The aim of the research was to reproduce complex tasks in 3D without constraints on the operator. This goal reflects the importance of gesture-based interfaces that utilize low-cost, consumer-grade input sensor devices for industrial use.

Therefore, both devices show many advantages in their use. However, we chose the Leap Motion because of the dimensions of our setup and the space available for the user. Furthermore the user interacts mainly with hands and his/her movement of the head. So, Leap Motion demonstrated to be our best choice.

3.2 Software

3.2.1 The Leap Motion Application

Programming Interface (API) includes also an effective skeletal tracking model that provides additional information about hands and fingers and improves the overall tracking data. In this way, the user is able to move the object and select the part of interest, even if the two hands are not completely tracked. The Oculus DK2 coupled with the Leap Motion were used to allow users

playing, grabbing, moving and rotating the 3D models in the VR scenario.

3.2.2 3D engine Software

The software application of the VR scenario has been developed by using Unity 5.0, a cross-platform game engine used to develop video games for PC, consoles, mobile devices and web sites. Unity 5.0 was released free on March 3, 2015, adding the real-time global illumination based on the Geometrics Enlighten technology, physically based shaders, HDR skyboxes, reflection probes, a new audio mixer with effects and enhanced animator workflows.

The software application was implemented by importing in Unity the 3D mechanical models from the KM3NeT detector units, 3D models of sea life, etc. In order to give more realism to the objects inside the VR scenario each 3D model was texturized with baked textures using Blender: a freeware 3D modeling software.

4. AUDIO IN GAME EXPERIENCE

4.1 SCI Model. Three Dimensions of Immersion

According to L. Ermi and F. Mäyrä [10], «human experiences in virtual environments and games can be defined as an ensemble made up of the player's sensations, thoughts, feelings, actions and meaning-making in a gameplay setting». It is not a property or a direct cause of certain elements of a game but something that emerges in a unique interaction process between the game and the player. They do not just engage in ready-made gameplay but also actively take part in the construction of these experiences: they bring their desires, anticipations and previous experiences with them, and interpret and reflect the experience in that light.

The SCI model by Ermi and Mäyrä is a heuristic representation of the key elements that make up the experience of playing. Its name stands for the three components which it includes: Sensory, Challenge-based, Imaginative immersion. It is not to be considered as a complete analysis model, but as a guide to the complex dynamics that are created between video games and players [11].

The first dimension, *sensory*, concerns the bond that establishes with aspects of the game world. In games that have their own world, the sensory characteristics are often able to simulate the situation to be *there*, in the sense that the game world becomes a genuine new reality for the player, leaving the real world in second floor. Ermi and Mäyrä emphasize the importance of quality and audiovisual style and say that, of three-dimensional game worlds, feeling of sensorial immersion is due high audiovisual quality that completely surround the player. In our case, the use of Oculus Rift enhances this feeling

because the gamer is completely immersed in the game world.

The second dimension of immersion in SCI model is the *challenge-based*, i.e. the involvement with a competitive process or problem solving, interacting with the game and/or with other players. The challenge-based immersion is closely tied to the gameplay, and is described by as «... the pleasure given by a combination of fast and cognitive engagement pace» [12]. Erni and Mäyrä say that this immersive dimension occurs when players experience a fair balance between challenges and skills, and, taken by the desire to advance and win, find themselves enveloped in a constant state of suspense. It is related to the sensory-motor skills, using the controller and the combination of buttons to get instant feedback in the game world. For KM3NeT experience, we imagine that the best challenge for the player is to be involved into the complex and extremely delicate job of unrolling the Detector Units with a length of almost a kilometer of detectors and hooking it to 4000 meters of depth, including fishes and currents interactions.

The third dimension is *imaginative immersion* itself: the involvement with the imaginative and fantastic world, with the characters and the story of the game. This dimension is linked both to interest into the story to be completely submerged or from an identification with the character of the game. Many times, a player can easily bypass the failures of the gaming experience, from the point of view of gameplay or challenge proposals, if it gives an imaginative immersion.

4.2 IEZA Model for the Audio of the Game: Interface, Effect, Zone, Affect

Audio is one of the main means for transmitting to the user information about the game. The IEZA model joins two conceptual dimensions that describe the communication of meanings with the audio game. IEZA connects two dimensions both regarding what is communicated with the soundscape of the game, with a classification in four different domains (Interface, Effect, Zone and Affect) which respectively belong to two different dimensions: diegetic and non-diegetic one [13]. The main focus of the model is to give a useful categorization of the interactivity audio of the game in the context of the gameplay. There are four different types of instances of audio in a game:

- Audio during interactive gameplay
- Audio during other interactive moments, or when the game is paused or in the menu screen. In these cases, there is still an ongoing interactive process, despite the gameplay is stationary
- Audio as part of the game when the game is in the active state, but the player does not interacts, i.e. cut-scenes and intro
- Audio out of the context of the game. During installation, ambient sounds or background are often repro-

duced. Music and sound are often found in advertising and trailers. Some soundtracks of video games are marketed and there are radio stations that play only music video game. In this case, the audio of the game is used outside of the game world as a reference to the game itself.

It is clear that, for the sense of immersion, the interactivity of the gameplay is the most important part, since the player can fully concentrate on the experience to actively participate. In non-interactive parts of the game, the player's role is more that of a spectator [14].

4.3 Audio in Virtual Reality

Virtual audio scenes can be created using different sound field generation methods based on fundamental properties of the perception of sound. Unique human ear structure allows us to localize sounds and comprehend the spatial information of the surrounding. With deeper understanding of human audio perception, different cues of localization and its psycho acoustic principles, it has become possible to reproduce more realistic virtual audio scenes [15].

A VR system cannot provide an accurate impression of virtual space unless its visual inputs are in perfect synchronization with the audio. A small discrepancy in visuals of the audio source and the corresponding sound pressure at listener's ears can reduce the effectiveness of the experience by large amount. Therefore VR systems typically require a sound system with immersive properties to create realistic sounds associated with visuals. Such sound systems create virtual sound sources and audio scenes to give impression of the realistic audio environment for the listener. Unique hearing system in human beings make possible to naturally sense different cues of sound source such as direction, distance, loudness of sound sources and idea of surrounding space.

HRTF technique is one of the possible solutions to render sound source distance accurately [16]. This technique is founded on Lord Rayleigh's diffraction formula which describes, with a certain approximation, the behavior of a sound wave produced by an infinite point source around the listener's head. A *head-related transfer function* (HRTF) is a more general law descending from diffraction formula by Rayleigh. HRTFs represent the behavior of a sound wave in its path from the source to the eardrum, depending on diffractions and reflections on the head, torso, shoulders and pinnae of the listener [17]. Virtual positioning of sound sources is achieved by filtering the corresponding signals through a pair of HRTFs. As a final result, two different audio signals are created, one for each ear, and transmitted by headphones.

It seems evident that it should be necessary to measure individual HRTF for each listener to make the results of HRTFs really immersive. Due to the great efforts both in time and resources, structural modeling of HRTFs

appears to be a solution. For further details about head-related transfer function model for real-time customized 3D sound rendering, some references are reported in [17], [18] and [19].

For our purposes, we decided to use a MAX/MSP algorithm with fixed HRTFs that allows us to change in real-time the angle information received from Leap Motion. Although this choice obviously represents a handicap for players, it simplifies our system that still needs many improvements. The implementation of a real-time customized 3D audio is one of our future goals.

5. KM3NET EXPERIENCE: SONIFICATION

By using the UDP network protocol, we create an *ad hoc* local network, which allows us to send instances of the game by Unity in MaxMSP. We classify these instances into two categories:

- instances of action
- instances of localization

The first category includes all the variables of boolean type, or all those instances that take on values such as True/False, or 1/0. In our specific case, examples are variations in the type of avatar's movement, objects near the user or events due to user's interactions.

The second category is made up of the position information of specific objects in the virtual space. These parameters assume numerical values, whole or decimal. They could be variable, in the case of dynamic objects, or non-variable, in the case of static objects. We decide to consider the positions of the objects of the game world in terms of distance and angle by the avatar. All this information, real-time generated by the video games, are also relayed by MaxMSP to different algorithms that allow to manage the playback of samples, volume changes and spatialization.

All the samples come from KM3NeT's recorded sounds database; so, they are mainly water sounds, electronic sounds and sounds from marine life which interacts with the experiment. They are used as they were recorded - especially to realize the soundscape - and also with some manipulations, such as granularization and spectral spatialization. All the collected scientific data of KM3NeT are used to manage the generation of the continuous soundscape of the videogame and its constant changes.

5.1 KM3NeT Data

Audio samples are played in response to specific instances of action. MaxMSP allows to play audio files from disk or read buffer. We decided to opt for the first

type of reproduction, considered more suitable for this specific application.

With the exception of the avatar, that always is in the foreground, for other sound sources of the game world it was necessary to create volume changes. In order to make the world game more realistic, these variations have been associated to the distance of the avatar from the sound source and to the angle given by the Leap Motion. In this way, the game world responds directly to the player's actions in the audio domain as well as in the video one.

The game elements, with an active part in the audio domain, have been associated with the four domains proposed by IEZA model. The instances related to the pause menu, or to the gameplay when not active or the case of game over, belong to the *interface domain*.

The game elements classified as AI (Artificial Intelligence) falls into *effect domain*. In our case, the only example of AI is represented from sea life (whales, fishes, etc.).

In addition to the artificial intelligences, there are some objects that react to player actions, with response in the audio domain. It is the case of the teleporting points to change scene. These static sound sources fall in the *zone domain*, as elements that characterize and make the setting more convincing.

The elements that have the function of representing the environment are associated with specific audio samples, consistent with the game world, to recreate a credible virtual space.

The background music belongs to *affect domain*. It consists of a sample that serves as a *theme*, used in situations in which the avatar is out of danger, and an *alarm* version of it, to suggest to the player that something is going wrong. This theme is superimposed to the continuously evolving soundscape made up mainly of KM3NeT recordings. In this way, even in the slow moments of gameplay, the game environment supports and makes the gaming experience more compelling. We opted for a background level, with imposed dynamics, for different reasons: primarily to better accentuate the dynamic point sources of the gaming environment; in addition, using a static background offers the player the opportunity to focus more on the challenges of the game.

6. CONCLUSIONS AND FUTURE WORKS

We officially tested our virtual reality experience to the next KM3NeT convention in Noto (Siracusa, Italia) from the 13th to the 17th of June, 2016 as an interactive installation. It was a first experiment, so we learned more about the gaming feedback after almost all the scientist Collaboration have tried it.

We are already working on improvements that we hope to show at XXI CIM in Cagliari (28th September – 1st October 2016). First of all, we are implementing the possibility of a personalized binaural audio. Nevertheless, this implies some computing problems that we are trying to solve. Then, we are going to combine Microsoft Kinect and Leap Motion to support more gestures and to enhance user experience. Finally, we are now testing the game with the new Oculus Rift device: the CV1 (the commercial version n. 1).



Figure 3. KM3NeT Experience in action.

7. AKNOWLEDGMENTS

We would like to really thank KM3NeT Collaboration for this opportunity, and especially Dr. Giorgio Riccobene, Dr. Edward Berbee, Dr. Tommaso Chiarusi and the KM3NeT spokesperson Marteen de Jong.

8. REFERENCES

- [1] For any other information about KM3NeT Collaboration, visit: www.km3net.org
- [2] <http://www.oculus.com>
- [3] <https://www.leapmotion.com/>
- [4] blog.leapmotion.com
- [5] <http://unity3d.com/>
- [6] <http://digitalchumps.com/gamingnews/318-latest-news/oculus-rift-announces-development-kit-2.html>
- [7] K. Khoshelham and S. Elberink: “Accuracy and resolution of kinect depth data for indoor mapping applications”, in Sensors 2012, 12, 1437–1454
- [8] R. J. Teather, A. Pavlovych, W. Stuerzlinger and I. S. MacKenzie: “Effects of Tracking Technology, Latency, and Spatial Jitter on Object Movement” in Proceedings of the IEEE Symposium on 3D User Interfaces, Lafayette, LA, USA, 14–15 March 2009; pp. 43–50..
- [9] F. Hernoux, R. Béarée, L. Gajny, E. Nyiri, J. Bancalin and O. Gibaru: “Leap Motion pour la capture de mouvement 3D par spline L1. Application à la robotique”, in Journées du Groupe de Travail en Modélisation Géométrique 2013, Marseille
- [10] L. Ermi and F. Mäyrä: “Fundamentals Components of the Gameplay Experience: Analysing Immersion”, in Changing Views: Worlds in Play. Selected Papers of the 2005 Digital Game Research Association’s Second International Conference, pp. 15-27, 2005.
- [11] L. Ermi and F. Mäyrä: “Fundamentals Components of the Gameplay Experience: Analysing Immersion”, in Changing Views: Worlds in Play. Selected Papers of the 2005 Digital Game Research Association’s Second International Conference, pp. 15-27, 2005.
- [12] C. Crawford: “The Art of Computer Game Design”, McGraw-Hill, 1984.
- [13] S. Huiberts and R. Tol: “IEZA: A Framework for Game Audio”, Utrecht University, 2008.
- [14] F. Boccieri: “Il gamer ha anche le orecchie. L’influenza dell’audio sull’esperienza videoludica,” *thesis (supervisor: A. F. Battista)*, Conservatorio “D. Cimarosa” of Avellino, 2015.
- [15] K. Salen and E. Zimmerman: “The Game Design Reader”, The MIT Press, 2006.
- [16] M. Geronazzo, F. Avanzini and F. Fontana: “Use of personalized binaural audio and interactive distance cues in an auditory goal-reaching task”, in Proceedings of 21st International Conference on Auditory Display ICAD 2015
- [17] M. Geronazzo, S. Spagnol and F. Avanzini: “A Head-Related Transfer Function Model for Real-Time customized 3-D Sound Rendering”, in Proceedings of Seventh International Conference on Signal Image Technology and Internet-Based Systems 2011
- [18] S. Spagnol and F. Avanzini: “Real-time binaural audio rendering in the near field”, in proceedings of 6th Sound and Music Computing Conference SMC 2009.
- [19] Y. Tang, Y. Fang and Q. Huang: “Audio personalization using head related transfer function in 3DTV”, In 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011.

Modello riepilogativo delle principali attività organizzative in una composizione elettroacustica

Marco Ferrazza

marcoferrazza@gmail.com

SOMMARIO

Lo scopo di questo articolo è quello di individuare, al di fuori di qualsiasi categoria di studio già sistematizzata, quelle che sono le operazioni basilari per la realizzazione di un'opera di musica elettroacustica¹.

Per semplicità espositiva ci concentreremo soprattutto, secondo una scelta del tutto arbitraria frutto di chi scrive, sulle composizioni acusmatiche e l'improvvisazione elettroacustica dal vivo, il che ci permette di stabilire un connubio tra due forme espressive: una che prevede il controllo minuzioso dei propri materiali sonori e l'altra che si affida maggiormente all'alea come impegno creativo.

Il "modello" qui proposto non pretende né vuole essere in nessun modo un sunto esaustivo e definitivo sul tema della composizione elettroacustica, ma vuole semplicemente proporre un insieme di riflessioni (naturalmente perfettibili) su come affrontare il lavoro compositivo, riflessioni che derivano dalla pratica musicale personale dell'autore.

1. UN MODELLO RIEPILOGATIVO

L'idea che muove questa proposta di studio è in definitiva quella di stabilire una continuità tra ciò che fa un compositore di musica acusmatica al momento di realizzare un proprio lavoro per suoni fissati, e l'improvvisatore che opera nella pratica dal vivo².

¹ La definizione di "musica elettroacustica" comprende un insieme eterogeneo di approcci musicali contemporanei per i quali stabilire una catalogazione precisa e definitiva è compito che il presente articolo non si propone di risolvere. Con tale definizione ci si riferisce allora, in maniera indistinta, a composizioni per suoni fissati, opere per il Live electronics e tutte le forme miste di musica contemporanea che contemplano l'impiego di mezzi elettroacustici per la realizzazione, la manipolazione e l'esecuzione del suono. Per ipotesi definitorie maggiormente articolate si veda [3].

² Fondamentale per questo studio è stato un seminario, tenuto da Roberto Dani nell'ottobre del 2012 al Conservatorio di Cagliari, sul tema dell'improvvisazione, seminario dal quale sono poi derivati molti spunti per il presente articolo.

Copyright: © 2016 Marco Ferrazza. 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.

Spesso il compositore e l'improvvisatore-performer sono la stessa persona, e quindi, a maggior ragione, è avvertita la necessità di voler riassumere le strategie compositive più evidenti in un unico strumento quale potrebbe essere un modello riepilogativo, del quale si vuole proporre qui una prima stesura.

Per semplicità, in questa breve trattazione si userà il termine generale di "musicista" per indicare sia la figura del compositore di musica acusmatica che quella dell'improvvisatore-performer di musica elettroacustica.

Per incominciare, chiediamoci: cosa accade durante una composizione, e, per restringere il campo alla nostra area d'interesse, durante una composizione di musica elettroacustica? Quali sono le pratiche organizzative che il compositore o il performer mettono in atto rispetto al materiale sonoro³ per caratterizzare un'opera e svilupparla nella durata?

Il musicista stabilirà senz'altro degli interventi per dotare il proprio materiale sonoro di una qualche struttura e sviluppare quindi la propria poetica. Tali interventi saranno caratterizzati inevitabilmente da un'escursione ideale all'interno di una gamma di possibilità espressive.

Premettiamo innanzitutto l'esistenza di un Organico, e intendiamoci subito su cosa intendiamo qui con questo termine, dato che non verrà utilizzato in maniera convenzionale: l'Organico è l'insieme dei materiali sonori che il musicista sceglie di utilizzare, vale a dire i vari "strumenti" che egli deciderà di organizzare nella pratica. Gli "strumenti" sono qui nient'altro che l'esito delle elaborazioni elettroacustiche di fonti sonore originarie scelte di volta in volta dal musicista, elaborazioni a loro volta soggette a eventuali ulteriori trattamenti.

Se ipotizziamo che il nostro musicista si incammini nella propria produzione-esecuzione in maniera solitaria, l'Organico sarà semplicemente la totalità dei suoni che egli sceglie di impiegare. Nel caso invece di, ad esempio, un'improvvisazione con più musicisti, al di là del numero di strumenti che ogni singolo musicista utilizzerà, l'Organico è ancora l'insieme dei materiali sonori impiegati e percepiti⁴.

³ Per materiale sonoro si intende qui semplicemente il risultato della produzione fisica dei suoni che il musicista elettroacustico utilizzerà all'interno del proprio lavoro compositivo.

Supponiamo che il musicista scelga di utilizzare un suono tessiturale e un altro suono di carattere gestuale al fine di sviluppare una cosiddetta texture-setting⁵, in questo caso l'Organico è composto soltanto da due elementi, ma esso potrebbe ampiamente coinvolgere un numero impreciso di materiali sonori, partendo da un singolo suono per arrivare a una molteplicità di proposte.

Ecco quindi una prima, globale differenziazione per quel che riguarda l'Organico: a un'estremità troviamo la presenza di un solo suono, all'altra la disponibilità di più elementi:

ORGANICO (1-n)

Stabilito l'Organico, si tratta ora di discutere di ciò che definiamo proprio come gli Interventi del musicista sulla produzione dell'opera.

Essi includono quattro aree principali: Presenza, Azione, Interazione ed Escursione.

L'idea dell'Organico richiama immediatamente il tema della Presenza sonora, vale a dire della sola disponibilità quantitativa⁶ dei materiali sonori nel correre alla formazione dell'opera.

La Presenza sonora si misurerà quindi in base al numero degli eventi sonori percepiti.

In questo caso, la Presenza dei suoni può essere organizzata lungo una linea immaginaria che prevede ad un'estremità il silenzio e all'altra la massima densità dei materiali, includendo ovviamente tra questi due "limiti" varie combinazioni.

Altro coinvolgimento del musicista nella produzione della propria opera riguarda ciò che definiamo semplicemente come Azione, vale a dire, per riprendere ancora il nostro esempio iniziale e compiere un'estrema semplificazione, l'alternanza tra suoni tessiturale e gestuali⁷. Anche in questo caso possiamo immaginare e importare all'interno di questa dualità una certa varietà di comportamenti, dall'uso del legato allo staccato e così via (articolazioni che ampliano le possibilità espressive della schematizzazione duale Texture/Gesto), secondo quello che la cultura musicale e il gusto della ricerca suggeriscono al musicista.

L'Interazione include invece una polarità tra il solo e il segnale. Il solo è ovviamente l'esecuzione sonora effettuata da un unico musicista/esecutore (che può impiegare uno o più strumenti), mentre per segnale intendiamo un chiaro intervento dello strumentista, o una spicata variazione della trama sonora, tale da suggerire

⁴ Per quanto riguarda il tema della percezione, tema che ritornerà anche altrove nel nostro elaborato, si veda [5], [7] e [8].

⁵ Per eventuali chiarimenti sulla terminologia adottata rimandiamo direttamente a [2] e [15, 16].

⁶ Quantitativa e non qualitativa anche per evitare qualsiasi intenzionalità di giudizio estetico che, in questa fase di studio, non fa parte della nostra ricerca.

⁷ Semplificazione che comunque riprende una distinzione, fondamentale per la musica elettronica, teorizzata da Denis Smalley. Si veda quindi [15, 16].

all'Organico un cambio generale nell'esecuzione (ad esempio, un crescendo potrebbe effettivamente preludere a una qualche variazione chiaramente percepibile).

Può certo verificarsi una coincidenza tra solo e segnale ma, nel caso di un Organico formato da più strumentisti, il segnale costituirebbe una chiara disponibilità alla mutazione sonora (disponibilità che, se pensiamo ad esempio all'ambito dell'improvvisazione, può liberamente essere recepita o meno).

Tra queste due estremità, quella del solo e quella del segnale, l'Interazione prevede anche due "tappe" intermedie determinanti per la riuscita dell'opera: l'imitazione e il contrasto.

Imitazione e contrasto richiedono inevitabilmente la presenza contemporanea di almeno due suoni. Uno di essi potrà assumere arbitrariamente un comportamento di imitazione rispetto alla qualità esecutiva dell'altro suono (ad esempio, due elementi dell'Organico potrebbero eseguire entrambi una serie di suoni gestuali) o vi si opporrà secondo un contrasto (un elemento dell'Organico esegue suoni gestuali mentre l'altro suoni tessiturale). Inoltre, imitazione e contrasto possono essere eventi che si verificano sia contemporaneamente che consecutivamente⁸.

Rimane dunque l'Escursione. Essa, molto semplicemente, si riferisce all'escursione della dinamica (per intenderci, la gamma compresa tra ppp e fff) e dell'altezza (quindi, genericamente, tra suoni di frequenze molto basse, gravi, e suoni di frequenze molto alte, acuti).

L'escursione dinamica potrà avere ripercussioni anche sulla varietà timbrica dell'Organico (ad esempio, a seconda all'escursione dinamica, alcuni strumenti potrebbero risultare percettivamente più pregnanti rispetto ad altri, che si disporrebbero sullo sfondo⁹). Tali ripercussioni saranno di volta in volta riscontrabili in base al tipo di materiali sonori utilizzati e al modo in cui verranno impiegati.

In conclusione, riassumiamo la nostra proposta di modello compositivo (susceptibile di ulteriori sviluppi e revisioni) in uno schema riepilogativo semplificato (nel quale, come abbiamo visto, anche la dinamica e l'altezza

⁸ L'interazione richiama quindi la capacità del musicista elettronico di rispondere agli stimoli esterni e modificare a sua volta il contenuto sonoro dell'opera; in questo senso va detto che le possibilità attuali di interagire in tempo reale con il resto dell'Organico (al pari di qualsiasi strumentista "tradizionale") sono oggi in continua espansione. Gli odierni dispositivi tecnologici sembrano attraversare un'evoluzione continua e sono effettivamente capaci di "estendere la sensibilità umana" (come afferma Pietro Montani), così come l'utilizzo di controller sempre più sofisticati che affiancano l'elaboratore elettronico, i tablet o gli altri dispositivi dotati di touch screen (che consente la condensazione di più operazioni su un'unica superficie), favoriscono un'ampia disponibilità di integrazione tra performance corporea e qualità esecutiva, disponibilità che non era per niente scontata fino a qualche tempo fa. Forse il rischio, se di rischio si tratta, legato all'uso di tali dispositivi, è quello di un utilizzo indirizzato degli stessi secondo modalità già predisposte dal loro costruttore, il che limiterebbe in un certo senso la possibilità di incappare in quell'errore creativo da cui sono scaturite molte scoperte originali e innovative. Ma per tali questioni, che esulano dal nostro studio, rimandiamo direttamente a [9].

⁹ Per quanto riguarda le problematiche relative alla percezione figura-sfondo si veda direttamente [8].

za vengono proposte nella loro escursione tra due livelli di massima e minore possibilità espressiva).

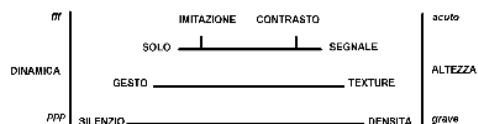


Figura 1. Schema riepilogativo del modello compositivo.

2. BIBLIOGRAFIA

- [1] M. Chion, *L'arte dei suoni fissati o la musica concretamente: opere di musica concreta*, Edizioni Interculturali, Roma 2004, pp. 124.
- [2] A. Cipriani e M. Giri, "L'arte dell'organizzazione del suono: processi di movimento", *Musica elettronica e Sound Design: Teoria e pratica con MAX e MSP. Volume II*, ConTempoNet, Roma, 2013, pp. 433-489.
- [3] F. Delande, "Il paradigma elettroacustico", in Nattiez, Jean-Jacques (a cura di), *Enciclopedia della musica. Volume I. Il Novecento*, Einaudi, Torino 2001, pp. 380-403.
- [4] A. Di Scipio (a cura di), *Teoria e prassi della musica nell'era dell'informatica*, Roma-Bari, Laterza 1995, pp. 195.
- [5] R. Doati, "Analisi dell'opera *Glissandi* di György Ligeti", *Quaderni di Musica/Realtà*, 27, 1991, pp. 162-173.
- [6] S. Emmerson, "La relazione tra linguaggio e materiali nella musica elettroacustica", *Teoria e prassi della musica nell'era dell'informatica*, Laterza, Roma-Bari, 1995, pp. 53-77.
- [7] F. Giomi e M. Ligabue, *Un approccio estesico-cognitivo alla descrizione dell'object sonore*, in R. Dalmonte e M. Baroni (atti a cura di), *Secondo convegno europeo di analisi musicale*, Università degli Studi di Trento, Trento 1992 (pp. 471-481).
- [8] D. Katz, *La psicologia della forma*, Bollati Boringhieri, Torino 1992, pp. 208.
- [9] P. Montani, *Tecnologie della sensibilità. Estetica e immaginazione interattiva*, Raffaello Cortina editore, Milano 2014, pp. 104.
- [10] J. J. Nattiez, *Il discorso musicale: per una semiologia della musica*, Torino, Einaudi 1987, pp. VIII, 197.
- [11] C. Palomba, "Pierre Schaeffer: alla ricerca dell'oggetto sonoro", *Musica/Realtà*, n. 52, 1997, pp. 65-78.
- [12] H. Pousseur, *La musica elettronica*, Feltrinelli, Milano 1976, pp. IX-333.
- [13] F. K. Prieberg, *Musica ex machina*, Einaudi, Torino 1975, pp. 311.
- [14] P. Schaeffer, *Traité des objets musicaux*, Éditions du Seuil, Paris 1996, pp. 701.
- [15] D. Smalley, "La spettromorfologia: una spiegazione delle forme del suono (I)", *Musica/Realtà*, n. 50, 1996, pp. 121-137.
- [16] D. Smalley, "La spettromorfologia: una spiegazione delle forme del suono (II)", *Musica/Realtà*, n. 51, 1996, pp. 87 -110.



Atti del XXI Colloquio di Informatica musicale Proceedings of the XXI Colloquium of Musical Informatics



REGIONE AUTONOMA
DELLA SARDEGNA
Assessorato della Pubblica
Istruzione, Beni Culturali,
Informazione, Spettacolo e Sport



COMUNE DI CAGLIARI
Assessorato alla Cultura,
Pubblica Istruzione,
Sport, Spettacolo e
Politiche Giovanili



Università di Cagliari
Dipartimento di Storia,
Beni culturali e Territorio



CONSERVATORIO
DI MUSICA
G. P. da Palestrina
CAGLIARI

Labimus

X NODI i festival
d'autunno
a Cagliari

ed. DADI - Dip. Arti e Design Industriale, Università IUAV di Venezia, 2016