

InMuSIC: un sistema interattivo multimodale per l'improvvisazione elettroacustica

Giacomo Lepri

Centre for digital music - Queen Mary University of London

g.lepri@qmul.ac.uk

ABSTRACT

InMuSIC è un sistema interattivo concepito per l'improvvisazione elettroacustica (clarinetto e live electronics). Attraverso un'analisi multimodale del comportamento del musicista (produzione del suono e movimento del corpo) InMuSIC genera e organizza nuovi materiali sonori. Gli interventi audio originati rappresentano una integrazione musicale in dialogo con le sonorità articolate dallo strumentista. La codifica e interpretazione di un complesso insieme di interazioni musicali e motorie permettono quindi al sistema di bilanciare differenti gradi di adattività e autonomia musicale.

1. INTRODUZIONE

Nel contesto della musica occidentale l'improvvisazione ha spesso avuto un ruolo rilevante. Le prassi vocali del Gregoriano e dell'Ars antiqua, la cadenza libera e le sperimentazioni relative alla musica contemporanea e al free jazz del Novecento sono solo alcuni degli esempi che permettono di considerare l'improvvisazione come una componente autorevole di molte tradizioni musicali europee. A partire dagli anni Sessanta del XX secolo compositori, strumentisti e performers hanno iniziato ad esplorare la pratica dell'improvvisazione anche all'interno dei contesti relativi alla musica elettronica. Il tentativo di sviluppare convincenti e spontanee interazioni musicali attraverso l'uso di nuove tecnologie e di nuovi linguaggi musicali sta tuttora spostando (se non dissolvendo) i confini tra composizione ed improvvisazione nella cultura musicale occidentale.

Con l'avvento delle tecnologie digitali si è sviluppato all'interno del Sound and Music Computing un nuovo campo di studi centrato sul design di sistemi interattivi concepiti per essere utilizzati in contesti improvvisativi. Queste ricerche di carattere compositivo, vengono spesso supportate da discipline extramusicali sia scientifiche che umanistiche (e.g. psicologia cognitiva, neuro scienze e human-computer interaction). I sistemi interattivi musicali offrono infatti uno scenario ideale per condurre studi relativi alla percezione, alla comunicazione/espressione, ai processi creativi e allo sviluppo di nuove tecnologie.

Copyright: ©2018 Giacomo Lepri et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

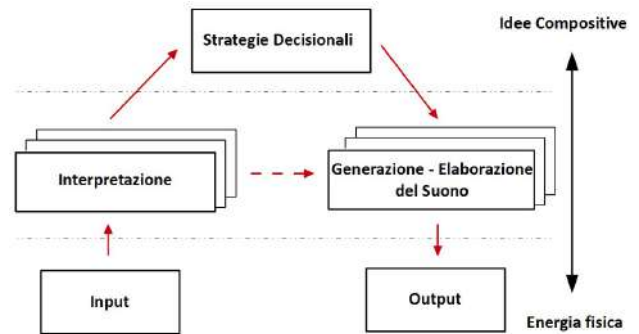


Figure 1. Il modello concettuale del sistema interattivo InMuSIC.

Negli ultimi vent'anni l'interesse al design di sistemi interattivi in grado di dialogare con i musicisti è significativamente cresciuto. I sistemi progettati dal compositore/improvvisatore George Lewis *Voyager* [1] e dal ricercatore informatico François Pachet *Continuator* [2] rappresentano due delle applicazioni musicali più significative recentemente sviluppate.

Gradualmente le esperienze acquisite da compositori, interpreti e ricercatori grazie alle indagini multidisciplinari inerenti alla composizione algoritmica [3] e alla liuteria digitale [4] forniscono una sempre più proficua cognizione del tema. In questo contesto sono evidenti i rapporti tra le strategie di progettazione e realizzazione di una specifica tecnologia/software ed il paradigma compositivo definito per sviluppare le interazioni musicali. Il framework utilizzato per la realizzazione di InMuSIC riflette quindi particolari intuizioni compositive e strategie improvvisative maturate all'interno di questo specifico contesto culturale.

2. IL MODELLO COMPOSITIVO

La definizione di un modello concettuale di riferimento rappresenta uno degli aspetti fondamentali nello sviluppo di sistemi interattivi musicali. L'esplicitazione del modello teorico permette infatti sia di chiarire la struttura implementativa nel suo insieme sia di organizzare le singole unità del sistema a seconda delle loro specifiche funzioni. Comporre interazioni musicali convincenti ed efficaci è una delle principali sfide che questo tipo di ricerca pone. Adottare un solido e ponderato modello compositivo agevola l'adempimento di questo complesso onere e facilita le varie fasi implementative. L'atto compositivo deve quindi tenere in considerazione ogni singola fase del processo interattivo (dalla cattura ed organizzazione delle informa-

zioni in ingresso fino alla articolazione dei materiali sonori generati durante l'improvvisazione).

In questa sezione viene presentato il modello concettuale di riferimento per la realizzazione di InMuSIC. La struttura del sistema si basa sul modello proposto da Leman e Camurri [5] per lo sviluppo di piattaforme interattive multimodali rivolte all'analisi dell'espressività musicale. La struttura è composta da cinque moduli collocati su tre piani concettuali (figura 1). Ciascun piano è associato ad un differente grado di astrazione. È possibile quindi tracciare un continuum che partendo dalla rappresentazione del mondo fisico si estende verso una dimensione musicale associata ad idee ed intuizioni compositive.

Il livello più basso è quindi attinente a quelle unità del sistema che svolgono compiti connessi al dominio fisico (e.g. rilevamento dei movimenti e delle sonorità prodotte dal musicista). Il livello più alto è invece relativo alle componenti più astratte del sistema, responsabili delle scelte compositive che concretizzano le interazioni musicali in tempo reale. Le funzioni dei singoli moduli sono quindi associate al trattamento dei dati in ingresso (Input e Interpretazione), in uscita (Generazione - Elaborazione del Suono) e alle dinamiche che definiscono le relazioni tra i vari processi di input ed output. Questa rappresentazione esplicita un ciclo interattivo distribuito su diversi piani concettuali e offre la possibilità di comporre e sviluppare quello che sarà il comportamento musicale del sistema.

Il modello presentato potrebbe essere interpretato come una possibile strategia di mapping, tuttavia, in questo contesto, si preferisce non fare strettamente riferimento al termine mapping. Esso sembra infatti perdere efficacia in quelle situazioni in cui processi algoritmici complessi vengono implementati per influenzare il comportamento del sistema e quindi la generazione del materiale sonoro [6]. Le funzioni dei singoli moduli e le loro relazioni vengono, di seguito, brevemente descritte.

- *Input*: il modulo si distingue per due funzioni principali: rilevare i movimenti e le sonorità prodotte dal musicista e convertire questa energia (cinetica e sonora) in informazione digitale. InMuSIC prevede l'utilizzo di due sensori: il suono dello strumento viene captato usando un microfono a condensatore, il movimento del performer viene catturato tramite un sensore di movimento 3D (Microsoft Kinect 2).
- *Interpretazione*: all'interno di questa unità le informazioni audio e gestuali in ingresso vengono analizzate ed interpretate attraverso diversi processi paralleli. In questa fase vengono estratte specifiche informazioni di carattere musicale ed espressivo significative del comportamento del musicista. Durante questa analisi la quantità di informazione diminuisce notevolmente poichè questo modulo filtra dati specifici di carattere prettamente musicale ed espressivo.
- *Strategie decisionali*: il modulo è collocato al livello più alto di astrazione musicale all'interno del modello compositivo. La sua funzione principale è relativa alla organizzazione temporale dei processi

adibiti alla generazione e manipolazione del suono. Le strategie decisionali implementate si basano principalmente sulla continua correlazione delle analisi di suono e movimento contestualizzate in differenti finestre temporali.

- *Generazione*: questa unità è composta da diversi algoritmi adibiti all'elaborazione digitale del suono: i materiali sonori proposti dal sistema vengono qui concretamente generati ed elaborati. Il sistema può assegnare, al fine di stabilire interazioni dirette con ciò che il musicista propone, il controllo di alcuni dei parametri degli algoritmi per la generazione del suono ai dati estratti in tempo reale dai moduli relativi all'analisi del comportamento sonoro e motorio dell'esecutore.
- *Output*: il modulo è adibito al trasferimento nel dominio fisico delle informazioni sonore generate dalle unità più astratte. I processi principali coinvolti in quest'ultima fase sono relativi all'amplificazione e compressione del segnale generato, alla conversione del segnale da digitale ad analogico e alla proiezione del suono nello spazio performativo.

3. ARCHITETTURA DEL SISTEMA

Da un punto di vista implementativo è possibile suddividere l'architettura del sistema in quattro unità principali: analisi del movimento, analisi del suono, strategie decisionali e generazione del suono (figura 2). Ogni unità è assegnata a componenti software specifiche comunicanti tramite protocollo Open Sound Control (OSC). Il sistema è quindi caratterizzato da un approccio modulare ed è possibile eseguire le varie componenti su diversi calcolatori connessi tra loro. Una descrizione dei quattro moduli e delle loro funzioni viene presentata di seguito.

3.1 Analisi del suono

L'unità cattura e filtra le informazioni provenienti da un microfono condensatore a clip. Differenti tecniche per l'estrazione in tempo reale di specifiche qualità sonore vengono eseguite parallelamente (Max/MSP). Nello specifico questo modulo compie analisi relative al rilevamento dell'ampiezza (loudness), attacco (onset) [7], frequenza fondamentale (pitch variation) [8], rumorosità (noisiness) [9] e baricentro spettrale [7] (spectral centroid).

Durante l'improvvisazione le informazioni audio vengono costantemente interpretate e ridefinite. Tramite una costante comparazione dei dati forniti dalle varie analisi sonore il sistema sceglie l'informazione più attendibile (da un punto di vista percettivo). Un semplice esempio di questo procedimento può essere la strategia utilizzata per effettuare il rilevamento dell'attacco: in questa unità sono implementati infatti tre algoritmi per l'estrazione di informazioni relative all'attacco sonoro. Il modulo seleziona una delle tre informazioni disponibili in relazione alle informazioni provenienti dall'analisi dell'ampiezza dello stesso

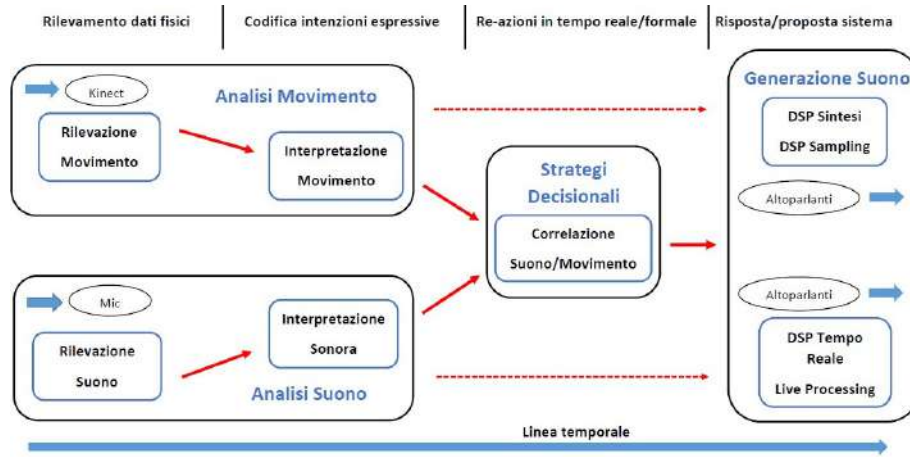


Figure 2. Il flusso dati nella struttura implementativa di InMuSIC.

segnale audio (ogni singolo algoritmo per l'estrazione degli attacchi è più affidabile in relazione a determinati range dinamici).

L'abilità del sistema di scegliere differenti strategie per l'analisi del suono consente quindi una rilevazione delle informazioni musicali di carattere adattivo. La metodologia utilizzata per l'interpretazione dei materiali sonori prodotti dal musicista si ispira alla straordinaria capacità umana di discriminare e selezionare specifiche informazioni musicali all'interno di contesti sonori caratterizzati da un alto grado di complessità. Il tentativo è quindi quello di modellare, con una accettabile approssimazione, questa abilità.

3.2 Analisi del movimento

L'unità esegue l'analisi del movimento per mezzo di vari algoritmi per l'estrazione in tempo reale di specifici gesti espressivi full body [10]. I dati provenienti dal sensore 3D vengono elaborati tramite la piattaforma EyesWeb [11]. Sulla base della definizione di gesto fornita da Kurtenbach e Hultheens [12]: "un movimento del corpo che contiene informazione", il gesto può essere definito espressivo se questa informazione è caratterizzata anche da contenuti di natura espressiva, cioè veicola potenziali messaggi impliciti [13]. Spesso queste informazioni aggiuntive introducono all'interno del processo comunicativo nuove qualità interpretabili, le quali solitamente, pur godendo di una loro autonomia espressiva, vengono giustapposte a specifici significati denotativi.

Sulla base di queste considerazioni, l'approccio adottato utilizza il modello sviluppato da Camurri et al. per l'analisi di gesti espressivi tramite l'impiego di sistemi interattivi multimodali [14]. La nozione di movimento espressivo presa in considerazione implica caratteristiche motorie associate ad un'ampia gamma gestuale non ristretta a tipologie di movimento specifiche. Il modulo non è quindi concepito per riconoscere gesti particolari, la ricerca è invece indirizzata ad una esplorazione del movimento di carattere qualitativo. Certamente ogni strumento (tradizionale o relativo alle nuove tecnologie) è connotato da un specifico apparato di affordances gestuali [15]. Per di più, ogni

musicista sviluppa una personale modalità di interazione fisica con esso.

La sfida consiste proprio nell'implementare algoritmi per l'estrazione di informazioni di carattere astratto e generalizzato, rappresentative di ampie sfere comunicative associate al movimento espressivo. La strategia adottata prevede quindi una semplificazione (minima rappresentazione) del movimento espressivo attraverso un'analisi di tipo qualitativo applicata alla parte superiore del corpo del musicista [15]. Considerando un numero ridotto di informazioni visive (Fig. 4) (posizione, velocità e accelerazione di testa, mani, spalle e gomiti del musicista), vengono effettuate tre analisi del movimento di carattere espressivo: fluidità del movimento (fluidità e continuità vs impulsività), variazione della postura (chiusa vs aperta) ed energia cinetica (variazione della quantità di movimento della parte superiore del corpo) [15].

3.3 Comparazione suono-movimento

In linea con l'approccio di rappresentazione minimale precedentemente introdotto, le informazioni provenienti dagli algoritmi di analisi vengono interpretate in relazione a tre possibili stati: *high*, *low* o *stable*. Prendendo ad esempio la feature audio, è possibile descrivere e associare i vari stati ai seguenti tratti musicali: (i) dinamica high, low o stable (crescendo vs. diminuendo); (ii) high, low o stable rilevazione di attacchi (aumento o diminuzione della densità degli eventi); (iii) high, low o stable deviazione della frequenza fondamentale (espansione vs. riduzione della gamma di frequenza utilizzata). Questa classificazione viene effettuata per tutte le features analizzate (analisi del movimento e del suono). Il modulo è quindi progettato per combinare e confrontare questo tipo informazioni.

Per ogni feature estratta, i vari stati stable vengono ignorati: il rilevamento di uno stato stable non produce alcuna modifica alle condizioni interne del sistema. La figura 4 illustra le combinazioni disponibili in relazione a ciascuno stato high-low. Attraverso un'interfaccia grafica utente (GUI) è possibile selezionare manualmente le combinazioni che il modulo deve rilevare durante la performance. La Figura presenta una possibile selezione delle combinazioni

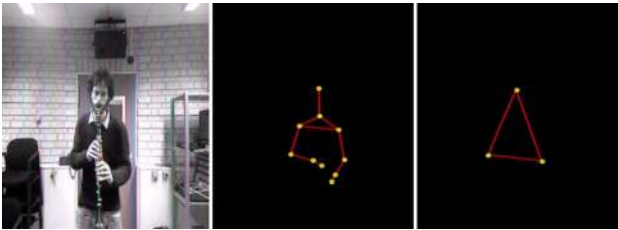


Figure 3. Lo scheletro di un musicista estratto tramite l'unità per l'analisi del movimento. La quantità di informazione viene ridotta fino ad ottenere l'area del triangolo testa/gomiti.

di stati spesso utilizzate dall'autore durante l'esecuzione di InMuSIC.

Una volta scelta una combinazione specifica (ad esempio low energia cinetica e low loudness), il modulo verifica costantemente se i due stati vengono rilevati. Contemporaneamente e a ciascuna combinazione selezionata viene applicata una semplice condizione booleana. In breve, durante la performance, i dati inviati al modulo decisionale definiscono (i) quale condizione selezionata è attualmente vera e (ii) il tempo associato alla persistenza di ciascuna condizione verificata.

Il calcolo dei vari stati high-low consente la raccolta di informazioni relative alla variazione nel tempo delle caratteristiche estratte permettendo di definire se una feature specifica sta aumentando o diminuendo. La combinazione e il confronto degli stati high-low associati alle varie caratteristiche è quindi concepito come un ulteriore livello di astrazione all'interno dell'analisi espressiva del performer.

3.4 Strategie decisionali

Questa unità rappresenta il nucleo del sistema, essa infatti gestisce il livello compositivo più astratto responsabile dei processi che permettono l'interazione del sistema con il performer. La principale funzione di questo modulo è relativa alla generazione ed elaborazione di nuove informazioni musicali che possano essere da stimolo per la continuazione dell'improvvisazione.

Il modulo è realizzato per poter attuare processi decisionali associati a due differenti finestre temporali: 2-4 secondi (re-azioni in tempo reale) e 2-5 minuti (re-azioni in tempo formale). Il paradigma di riferimento si basa su ricerche relative allo studio del funzionamento della memoria uditiva degli esseri umani (memoria a breve e lungo termine) [16]. Da questo punto di vista il tentativo è quindi quello di trasferire nel sistema una sorta di consapevolezza della dimensione temporale. In concreto, sulla base del settaggio effettuato prima che la performance abbia inizio, il modulo effettua una costante correlazione tra le informazioni interpretate dalle unità di analisi audio e di movimento. Le condizioni che occorrono più frequentemente (e.g. simultanea rilevazione di alta presenza di energia cinetica e alta densità di eventi sonori) vengono, durante l'improvvisazione, memorizzate.

Questo database di carattere adattivo è la principale sorgente per lo sviluppo di processi decisionali connessi alla

finestra temporale più ampia. Allo stesso tempo, l'unità osserva costantemente i singoli parametri analizzati: questa procedura permette l'attuazione di un secondo processo parallelo che influenza la generazione ed elaborazione dei dati di controllo associati alle interazioni sonore relative alla dimensione del tempo reale.

In un certo qual modo, le varie analisi audio e di movimento vengono in questa unità software (Max/MSP) interpretate una seconda volta al fine di generare ed organizzare nel tempo i nuovi materiali sonori. Imitazione, variazione, contrasto, continuazione, interruzione, autonomia ed aleatorietà sono alcune delle metafore sulla quale si basano i processi decisionali implementati. In definitiva la procedura fondamentale riguarda la possibilità di generare nuovi elementi musicali osservando i dati in ingresso da diversi punti di vista: distinte finestre temporali e differenti tratti espressivi.

3.5 Generazione del suono

L'unità di generazione audio (piattaforma Max/MSP) è concepita per produrre materiali audio eterogenei. Infatti, in relazione a diversi contesti musicali, la varietà timbrica e di articolazione sonora risulta essere un requisito importante per lo sviluppo di interventi musicali stimolanti ed efficaci. I processi di elaborazione digitale del suono (Digital Sound Processing - DSP) relativi alla generazione delle proposte/risposte del sistema possono essere suddivisi in tre categorie: sintesi (generazione di suoni elettronici ex novo), sampling (elaborazione in tempo reale di suoni pre-registrati) e live processing (elaborazione in tempo reale dei materiali sonori provenienti dal microfono).

Ad ogni categoria sono associati molteplici algoritmi i quali, durante l'improvvisazione, possono essere eseguiti simultaneamente. Le singole tecniche utilizzate per la generazione ed elaborazione del suono possono essere rappresentate come voci del sistema. Ogni voce si distingue per specifiche qualità spettrali/morfologiche (distribuzione e variazione della energia nello spettro) e gestuali (articolazione e trasformazione dei materiali sonori nel tempo). Parallelamente ogni singolo algoritmo è stato progettato per poter garantire un certo grado di apertura/indeterminatezza in relazione alla generazione dei relativi materiali audio. Al fine di chiarire questo concetto, può essere utile tracciare una analogia con lo sviluppo delle voci (soprano, contralto, tenore e basso) tipico della forma del corale. Nella prassi, tenendo in considerazione i vari registri ed i moti delle singole linee, viene consentito ad ogni voce di variare in relazione a specifici gradi di libertà. Allo stesso modo vengono qui definiti dei limiti entro i quali i parametri dei vari algoritmi possono operare. Generalmente, il controllo dei singoli parametri, all'interno dei confini prestabiliti, viene di volta in volta assegnato a differenti sorgenti (e.g. dati estratti in tempo reale dalle analisi di suono e movimento, processi stocastici, informazioni memorizzate durante la performance).

L'obiettivo è quello di definire algoritmi di generazione audio che, all'interno di contesti sonori predefiniti (e.g. ambiti spettrali, dinamici e temporali) possano sviluppare continue variazioni e manipolazioni dei materiali. In altre

		Loudness		Events Density		Pitch Deviation	
		high	low	high	low	high	low
QOM	low		✓				
	high			✓			
Contraction Index	low			✓			
	high	✓					✓
Smoothness	low						
	high		✓				✓

Figure 4. Le possibili combinazioni relative alle analisi di suono e movimento. Le combinazioni segnate sono quelle usate spesso dall'autore durante le sessioni di improvvisazione con il sistema.

parole, ogni singola voce viene concepita come in grado di esplorare spazi sonori pre-composti e il musicista è invitato a navigare questi spazi timbrici [17] in stretta collaborazione con il sistema. Questo approccio, influenzato dalle procedure tipiche della composizione algoritmica, permette, in parte, di risolvere una dicotomia che emerge nel tentativo di unire le pratiche di composizione (definizione ed organizzazione di strutture musicali - azione svolta in un tempo compositivo) ed improvvisazione (processo che porta allo sviluppo di forme musicali tramite interazioni sonore - azione svolta nel tempo performativo) nelle loro accezioni tradizionali.

4. CONCLUSIONI

InMuSIC è un sistema interattivo multimodale concepito per essere utilizzato in un contesto inerente l'improvvisazione elettroacustica. L'improvvisazione musicale viene considerata come un atto espressivo spontaneo il quale implica abilità cognitive e tecniche veicolate da atteggiamenti motori e comportamenti sonori. Attraverso la correlazione delle analisi relative ai movimenti e ai suoni prodotti dal musicista, il software genera ed organizza in tempo reale materiali sonori eterogenei. InMuSIC è accordato per essere sensibile ad uno specifico vocabolario di sonorità e movimenti relativi alla pratica strumentale del clarinetto. Futuri sviluppi del sistema potrebbero prevedere la possibilità di ampliare questo vocabolario al fine di sperimentare l'utilizzo della piattaforma con differenti strumenti musicali.

Un ulteriore proposito, di carattere ben più arduo e complesso, sarebbe quello di lavorare con piccoli ensemble. In questo caso il livello di complessità concettuale ed implementativa del sistema aumenterebbe esponenzialmente poiché, oltre ad osservare il comportamento del singolo, la piattaforma dovrebbe agire tenendo in considerazione le dinamiche del gruppo. Sarebbe quindi necessario inserire nel modello compositivo presentato un ulteriore livello concettuale al fine di gestire processi decisionali in dialogo con l'atteggiamento espressivo e la condotta musicale dell'ensemble. Questo scenario, sebbene presenti sfide tecniche e compositive considerevoli, rappresenta un percorso di ricerca particolarmente stimolante.

InMuSIC è un sistema costruito intorno ad un complesso insieme di interazioni musicali che, influenzate dal comportamento del musicista, danno luogo ad un dialogo sonoro. Durante la sessione d'improvvisazione, il sistema

bilancia diversi gradi di adattività, autonomia e imitazione/variazione in relazione alle informazioni analizzate. La modalità performativa qui concepita presume quindi lo sviluppo di un intenso scambio musicale tra il sistema e il musicista. Da un punto di vista artistico, l'obiettivo/sfida dell'opera consiste nel riuscire a concretizzare una illusoria percezione di intima e stretta collaborazione tra i due agenti.

InMuSIC potrebbe essere definito come un sistema/macchina che, consentendo interventi esterni, compone/improvvisa musica attraverso una modalità dialogica. In realtà, l'oggetto di questa ricerca è in profonda relazione con quelle dinamiche umane che, all'interno di un determinato contesto culturale, influenzano sia la percezione del mondo esterno che l'espressione della propria identità. Questa ricerca pone domande relative alle modalità di generazione e trasmissione di potenziali contenuti musicali all'interno di un processo creativo spontaneo e condiviso. L'interesse quindi non è tanto indirizzato a comprendere se un computer può essere musicalmente intelligente, quanto ad effettuare una circoscritta esplorazione dell'espressività umana nel contesto musicale dell'improvvisazione elettroacustica.

5. REFERENCES

- [1] G. E. Lewis, "Too many notes: Computers, complexity and culture in voyager," *Leonardo Music Journal*, vol. 10, pp. 33–39, 2000.
- [2] F. Pachet, "The continuator: Musical interaction with style," *Journal of New Music Research*, vol. 32, no. 3, pp. 333–341, 2003.
- [3] D. Van Nort, J. Braasch, and P. Oliveros, "A system for musical improvisation combining sonic gesture recognition and genetic algorithms," in *Proceedings of the 6th Sound and Music Computing Conference*, pp. 131–136, 2009.
- [4] S. Jordà, "Instruments and players: Some thoughts on digital lutherie," *Journal of New Music Research*, vol. 33, no. 3, pp. 321–341, 2004.
- [5] M. Leman and A. Camurri, "Understanding musical expressiveness using interactive multimedia platforms," *Musicae Scientiae*, vol. 10, no. 1 suppl, pp. 209–233, 2006.
- [6] J. Chadabe, "The limitations of mapping as a structural descriptive in electronic instruments," in *Proceedings of the 2002 conference on New interfaces for musical expression*, pp. 1–5, National University of Singapore, 2002.
- [7] M. Malt and E. Jourdan, "Real-time uses of low level sound descriptors as event detection functions using the max/msp zsa. descriptors library," *Proceedings of the 12th Brazilian Smposium on Computer Music*, 2009.

- [8] A. De Cheveigné and H. Kawahara, “Yin, a fundamental frequency estimator for speech and music,” *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.
- [9] T. Jehan and B. Schoner, “An audio-driven perceptually meaningful timbre synthesizer,” *Analysis*, vol. 2, no. 3, p. 4, 2002.
- [10] A. Camurri, G. Volpe, G. D. Poli, and M. Leman, “Communicating expressiveness and affect in multimodal interactive systems,” *Ieee Multimedia*, vol. 12, no. 1, pp. 43–53, 2005.
- [11] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe, “Eyesweb: Toward gesture and affect recognition in interactive dance and music systems,” *Computer Music Journal*, vol. 24, no. 1, pp. 57–69, 2000.
- [12] G. Kurtenbach and E. A. Hulteen, “Gestures in human-computer communication,” *The art of human-computer interface design*, pp. 309–317, 1990.
- [13] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, “Emotional speech: Towards a new generation of databases,” *Speech communication*, vol. 40, no. 1, pp. 33–60, 2003.
- [14] A. Camurri, I. Lagerlöf, and G. Volpe, “Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques,” *International journal of human-computer studies*, vol. 59, no. 1, pp. 213–225, 2003.
- [15] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Morittillaro, and K. Scherer, “Toward a minimal representation of affective gestures,” *Affective Computing, IEEE Transactions on*, vol. 2, no. 2, pp. 106–118, 2011.
- [16] B. Snyder, *Music and memory: An introduction*. MIT press, 2000.
- [17] D. L. Wessel, “Timbre space as a musical control structure,” *Computer music journal*, pp. 45–52, 1979.