

ESPERIENZE DI SONIFICAZIONE NELLA COMPUTER MUSIC E NELLA RIABILITAZIONE ALL'ISTI -CNR

Graziano Bertini, Leonello Tarabella , Massimo Magrini

ISTI- Istituto di Scienza e Tecnologie dell'Informazione,
Area della Ricerca del CNR, Pisa

1. Introduzione

La sonificazione consiste nella generazione di suoni non verbali in funzione di dati di diversa natura utile a interpretarne certe proprietà. E' oramai impiegata in molti settori ed è in continuo sviluppo di pari passo con il progresso tecnologico. In questo lavoro vengono richiamati i principali elementi che la contraddistinguono e alcune tipiche applicazioni. Nel laboratorio di segnali e immagini (SILab) dell'ISTI-CNR di Pisa speciali meccanismi di sonificazione del gesto sono stati applicati nell'ambito della musica elettronica dal vivo. Ricordiamo qui che fin dagli anni '70, su iniziativa del M° Pietro Grossi (Conservatorio Musicale di Firenze), era nata una collaborazione tra la sezione Musicologica del CNUCE-CNR e il reparto di Ingegneria Biomedica dell'IEI-CNR di Pisa, con lo scopo principale di ottenere l'esecuzione di computer music in tempo reale, fino allora non consentita neanche con l'impiego di potenti computer. Per inciso, nel 1975 terminò la costruzione di un apparato elettronico, il Terminale Audio TAU2 [1], un sintetizzatore di suoni polifonico-politimbrico, controllato da un computer delle famiglie time-sharing IBM, tramite l'apposito sw musicale TAUMUS. A seguito dell'attività di produzione divulgativa/concertistica di musica elettronica portata avanti, un componente del gruppo di ricerca di Pisa (Leonello Tarabella) ha poi sviluppato l'idea di far dipendere la generazione dei suoni dalla gestualità dei musicisti, arricchendo le musiche con la dimensione performativa. I sistemi via via sviluppati sono basati principalmente su sensoristica a infrarossi e sull'analisi in tempo reale di segnali video acquisiti da telecamere.

Seppure nella maggior parte dei progetti realizzati la tipologia dei suoni generati è di valenza quasi esclusivamente musicale, le tecniche utilizzate e i sistemi sviluppati si possono inquadrare nell'ambito della *sonificazione*. Le competenze acquisite hanno permesso in seguito di lavorare allo sviluppo di sistemi più attinenti, in senso più stretto, al campo della sonificazione. Infatti, cambiando le sonorità di uscita e le strategie di elaborazione dei movimenti dei soggetti, i sistemi sono stati usati anche nell'ambito della medicina riabilitativa: in particolare l'*auditory feedback* del movimento ha permesso

di creare ambienti interattivi per il trattamento di disturbi del comportamento, nel quale l'interazione con l'ambiente circostante è compromessa, come i disturbi dello spettro autistico. Nel seguito daremo un cenno ad alcune problematiche coinvolte in generale nella sonificazione, a cui seguirà una breve descrizione di alcuni dei sistemi sviluppati, dei risultati ottenuti e una riflessione sulla funzione del mapping.

2. Cenni ad alcune applicazioni e problematiche della sonificazione

Abbiamo tutti presente l'esempio in natura del tuono come la sonificazione del fulmine: con la misura del tempo di arrivo del suono in un luogo si può stimare la distanza dalla scarica elettrica, non possibile con la sola vista dell'evento. Generalmente per sonificazione si è intesa la modalità di far corrispondere l'emissione di particolari sonorità in funzione di eventi di diversa natura, con l'uso di sistemi appositamente realizzati, allo scopo di consentire e/o facilitarne l'interpretazione o comunque estrarre delle informazioni. In tal senso si possono citare due datate applicazioni:

- il monitoraggio della presenza di radiazioni in un luogo tramite l'uso di appositi rivelatori, quali contatori Geiger..
- l'uso di ecoscandagli basati su sonar per la rilevazione della profondità di navigazione o ostacoli.

A parte queste ed altre tipiche applicazioni, risultano evidenti le motivazioni dell'utilità di considerare l'uscita sonora come la modalità più appropriata, quando le informazioni che devono essere tenute presenti ad un osservatore siano molto complesse se espresse nelle usuali forme di visualizzazione grafica, o come altri casi che prevedono avvisi/allarmi richiedenti un'azione immediata. Ad esempio, in ambienti di lavoro pratico l'operatore è spesso impossibilitato di guardare uno (o più) monitor o display; il sistema visivo potrebbe essere occupato con un'altra operazione, o colui che percepisce potrebbe essere non vedente [2], sia fisicamente o come risultato di fattori ambientali come il fumo o la linea di vista. Altre caratteristiche della percezione uditiva sono la capacità di monitorare e di percepire più set di dati acustici (ascolto parallelo), e l'elevata velocità di rilevazione dei suoni, soprattutto in ambienti di stress. Più recentemente grande importanza ha assunto il settore che prevede la creazione del suono a seguito di gesti e movimenti di persone o di altri organismi. In tale ambito intervengono problematiche di interfacce uomo-macchina che richiedono la progettazione di parti HW/SW e soprattutto di un adatto *mapping*, cioè la corrispondenza tra valori dei parametri estratti dagli ingressi e le caratteristiche dei suoni in uscita dal sistema (suoni semplici, complessi, allarmi, suoni musicali particolari, ecc.). Uno schema a blocchi che illustra una architettura generale, di riferimento anche per le nostre applicazioni è riportato in figura 1.

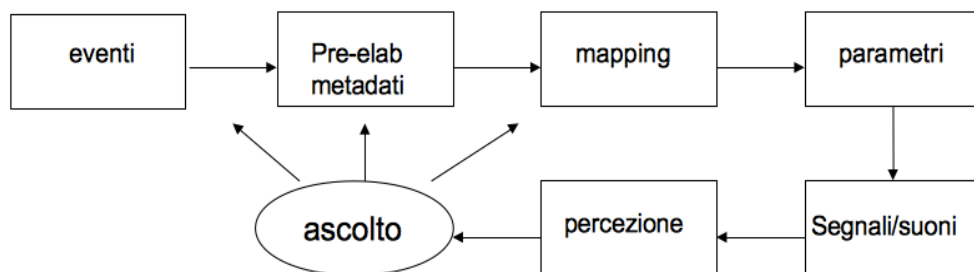


Figura 1 – Architettura e funzioni principali di un sistema di sonificazione di eventi

In effetti la realizzazione dei sistemi di sonificazione risulta alquanto articolata specie nei riguardi degli obiettivi da raggiungere. Si osservi che il tipo di eventi possibili è molto vasto e si riflette direttamente nelle tipologie di rilevazione, nelle interfacce di acquisizione e di pre-elaborazione, a seconda del tipo dei segnali considerati: analogico/discreto, numero di canali, ecc. Il campo inizialmente trattato dal nostro gruppo è quello delle applicazioni interattive di musica elettronica.

3. Alcuni sistemi sviluppati all'ISTI

Il SILab, Laboratorio segnali ed immagini dell'ISTI-CNR, ha competenze consolidate nel settore dell'elaborazione di segnali e immagini e nello sviluppo di innovative interfacce uomo-macchina. Per quanto riguarda la problematica della sonificazione è risultata utile la decennale attività svolta nel campo della bio-ingegneria, nel trattamento di segnali musicali e nei sistemi per il controllo attivo di rumore. Nei seguenti paragrafi illustriamo per sommi capi alcuni dei sistemi sviluppati; in seguito faremo un commento incentrato su uno dei punti critici coinvolti nella sonificazione, cioè il mapping.

Nelle esecuzioni dal vivo che prevedono l'uso contemporaneo di strumentazione elettronica e musicisti e/o cantanti si sono sempre verificati problemi di sincronizzazione dovuti alle limitate capacità di interazione dei sistemi commerciali. Nella maggior parte dei casi è il musicista che deve adeguarsi alla rigida precisione del sistema di sintesi (computer o apparato MIDI) rendendo difficoltoso l'emergere delle risorse espressive a sua disposizione. Vari gruppi operanti nel settore della computer music hanno sperimentato sistemi per ovviare a questo problema riferendosi a varie situazioni di uso con sintetizzatori ed accompagnamento e/o improvvisazione con strumenti acustici.

Light Baton. Uno dei primi originali sistemi da noi sviluppati e sperimentati negli anni '90 è stato quello che permetteva di controllare l'emissione di computer music da parte di sintetizzatori MIDI, "osservando" i gesti di un performer che impugna la bacchetta del direttore d'orchestra [3]. Il sistema era composto da una speciale bacchetta da direttore e un PC IBM-compatibile, corredato con una scheda di acquisizione di immagini collegata ad una telecamera (n.b. tutti componenti dell'epoca). I movimenti di una piccola lampadina posta sulla punta della bacchetta vengono rilevati dalla telecamera; digitalizzando e ricavando la posizione x-y della sorgente luminosa e inseguendola con opportuni algoritmi, ne viene ricostruita nel PC la traiettoria (figura 2).

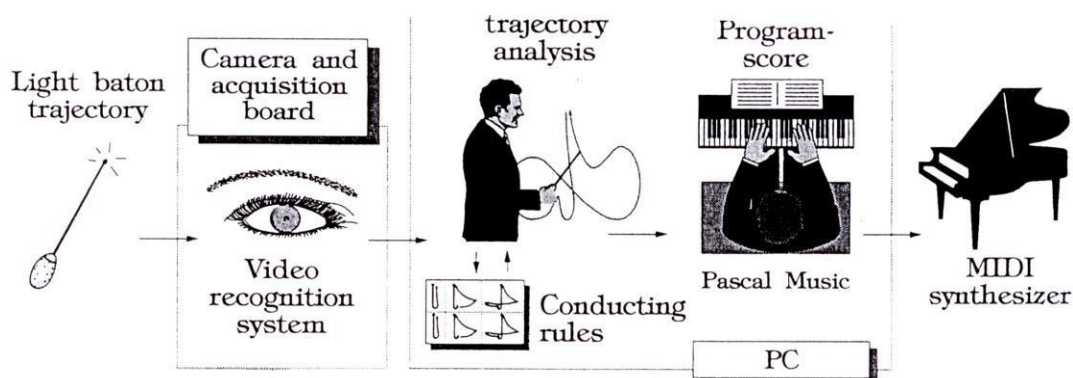


Figura 2 – Funzioni principali del sistema Light Baton

Tramite un metodo di interpolazione della traiettoria si ottengono i dati di interesse: beat point, numero d'ordine delle battute, ampiezza del gesto, fine del movimento, l'as-

senza di punti luce, ecc. I dati vengono applicati come parametri di input di uno speciale ambiente compositivo algoritmico, il Pascal Music [4], dotato di una serie di primitive che danno al performer la possibilità di gestire gli eventi musicali, in precedenza memorizzati, elaborandoli ed eseguendoli in modo interattivo con i movimenti della bacchetta.

Il sistema prevede la visualizzazione di parametri in un articolata interfaccia grafica, contenente le informazioni di interesse del performer e uno spazio virtuale dove viene rappresentata la posizione della bacchetta (tramite un piccolo quadratino); ovviamente quest'ultima deve muoversi in un determinato spazio fisico per non uscire dal controllo della telecamera. Il sistema è stato utilizzato in diversi concerti dal vivo, suscitando notevole interesse. In questo ambito sono state proposte altre soluzioni con sostanziali differenze, nella modalità di cattura dei movimenti del conduttore orchestrale, nel mapping ecc. [5, 6]. Sebbene a livello sperimentale i risultati siano stati suggestivi, a livello pratico concertistico certe sfumature dei gesti, che sono il bagaglio di esperienze e raffinatezze dei vari musicisti sono difficilmente integrabili, e per questo non esistono attualmente versioni commerciali.

- *PalmDriver*. Si tratta di una speciale interfaccia uomo - macchina preposta alla gestione di apparecchiature di sintesi audio mediante il movimento e la posizione assunta dalle mani dell'operatore in uno spazio ben definito; l'acquisizione dei gesti del performer, ad opera di transceivers all'infrarosso, permette di conferire nuova carica interpretativa ed emotiva a composizioni di musica elettronica più vicine a quelle della musica della tradizione. Inoltre, per la peculiarità dell'interazione, il dispositivo si differenzia profondamente da altri dispositivi come tastiere o batterie elettroniche: il controllo, infatti, è effettuato con continuità semplicemente muovendo le mani, traslandole o ruotandole in modo indipendente in determinati punti dello spazio di lavoro, creando nello spettatore l'impressione che l'artista stia suonando qualche tipo di strumento invisibile. L'effetto finale si deve principalmente all'efficace interpretazione (mapping) dei movimenti del performer, demandata ad un calcolatore su cui è in esecuzione un software specifico, mentre l'interfaccia in questione rappresenta l'anello intermedio tra i sensori e il sistema di sintesi del suono (figura 3).

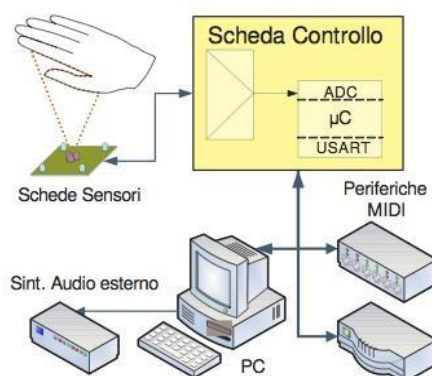


Figura 3 – Componenti di una delle prime versioni del Palm Driver

I suoni da emettere possono appartenere all'ambito della computer music e in tal caso il dispositivo viene utilizzato nei concerti. Negli esemplari realizzati come dimostrativi (versione Exhibit [7]) i suoni sono stati selezionati da una palette opportunamente predisposta: si consente così una sperimentazione percettiva accattivante di variazioni dei parametri del suono, quali altezza, timbriche ecc. Un dispositivo più semplice e da-

tato, che trasforma il movimento delle mani in semplici suoni basandosi sul principio delle onde radio, è il Theremin [8], tutt'oggi molto utilizzato.

-*Handel system*. Handel è un sistema di riconoscimento gestuale basato sull'analisi in tempo reale di immagini catturate da una telecamera (ed opportunamente digitalizzate) relative alla gestualità delle mani di un performer. Nell'area di ripresa della telecamera, è possibile definire zone attive (definite *panes*) all'interno delle quali porre e muovere le mani che vengono riconosciute in base al colore della pelle in contrasto rispetto allo sfondo che di solito è uniformemente bianco oppure nero [9].



Figura 4 – Il sistema Handel

Handel riconosce la posizione x-y di una mano all'interno di una zona attiva opportunamente predefinita; è in grado, inoltre, di riconoscere le dimensioni x-y del rettangolo (*frame*) che delimita la forma della mano. Il sistema produce perciò 4 informazioni per ciascuna delle due mani (a sinistra in figura 4). La qualità del video grabber utilizzato permette una velocità di acquisizione che assicura un controllo fluido degli algoritmi di sintesi sonora e delle strutture melodico/ritmiche opportunamente pre-programmate. La latenza tra gesto e generazione di informazione è inferiore a 5 centesimi di secondo che assicura “contemporaneità” tra visione e suono. Nelle recenti versioni del sistema le esecuzioni sonore sono corredate da grafica interattiva arricchendo con suggestioni visive le performance multimediali (a destra in figura 4).

3.1 Il progetto SIREMI

Da alcuni anni il SiLab ha utilizzato le tecnologie di interazione uomo-macchina e di sonificazione nella costruzione di innovativi sistemi per il trattamento di bambini interessati da disturbi dello spettro autistico (ASD). L'approccio seguito trae ispirazione dalla tecnica del FloorTime e dal modello DIR (developmental individual difference relationship, di Stanley Greenspan e Serena Wieder), declinandola in un contesto in cui la tecnologia la valorizzi e ne amplifichi l'efficacia. Il meccanismo di base di SIREMI (Sistema di Rieducazione Espressiva di Movimento e Interazione) è quello di correlare in modo stretto la dinamica corporea agli stimoli sensoriali. Tra tutti i vari canali sensoriali, quello uditivo è quello che richiede un maggior livello di astrazione, risultando quindi problematico per i soggetti interessati da DPS. SIREMI apre un ponte tra questo canale sensoriale e il movimento: la relazione che si viene a creare contestualizza in modo preciso gli stimoli uditivi percepiti, conferendo loro forma e contenuto informativo concreto. Il sistema informatico sviluppato al SiLab consiste in un insieme di dispositivi di rilevamento gestuale, di un calcolatore e di un sistema di amplificazione sonora. Il software (figura 5) sviluppato opportunamente, interpreta i dati provenienti dai dispositivi di rilevamento e produce suoni di vario genere ad essi correlati.

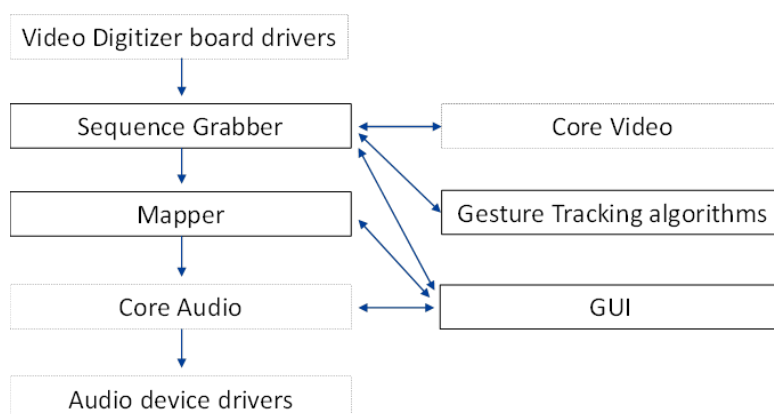


Figura 5 – Architettura del SW di SIREMI

Tramite interfaccia utente l'operatore è in grado di mappare a piacimento le caratteristiche del suono generato (tipologia, timbro, volume, altezza etc.) alle diverse grandezze caratteristiche della postura e del movimento. Il prototipo del sistema è stato installato presso l'Istituto Comprensivo Massarosa 1, dove è stato utilizzato in due stagioni scolastiche consecutive per una serie di sedute di trattamento rieducativo su un gruppo di soggetti della scuola media primaria, con interessanti risultati [10].

4. Il ruolo del Mapping nella sonificazione

Nei primi anni della computer music i ricercatori si sono concentrati sulla realizzazione di strategie e/o apparati speciali per la produzione del suono, facendo grandi sforzi per sviluppare specifici linguaggi musicali in grado di affrontare l'attività compositiva dal micro livello del suono (timbro, effetti, movimenti spaziali, ecc) al macro livello degli eventi musicali. Questo approccio alla composizione è stato in qualche modo ereditato da quella precedentemente in uso in studi elettronici analogici dove il nastro audio è stato considerato il prodotto finale. Anche i primi strumenti commerciali si limitavano alla semplice parte di sintesi audio, unita ad un'interfaccia tangibile di controllo (es. una tastiera musicale con slider etc.), una configurazione quindi riassumibile nella semplice catena *performer* → *controller* → *sintetizzatore*.

In seguito, quando i computer sono diventati sempre più potenti e disponibili a livello personale assieme a flessibili interfacce/controller, i ricercatori si sono resi conto che potevano riprendere il controllo gestuale del risultato sonoro, creando quindi una nuova configurazione degli elementi coinvolti: *performer* → *controller* → *computer* → *sintetizzatore*. Il computer assume così un nuovo ruolo durante la performance, in quanto genera eventi in conformità con la musica/materiale acustico predefinito, in combinazione con informazioni provenienti dai controllori su cui agisce l'artista. La semplice regola del mapping uno a uno, valida per gli strumenti tradizionali, lascia spazio per una serie teoricamente infinita di mapping definibile dal musicista/compositore, specificatamente per un determinato brano e addirittura per ciascuna sezione del brano. Il mapping diventa quindi parte della composizione o, se vogliamo, del meccanismo di sonificazione della gestualità del musicista [11].

Nelle performance elettroacustiche interattive un ruolo importante è giocato dal feedback acustico creato dai movimenti del performer sul controller e di conseguenza il suono generato udito dagli ascoltatori ma anche dal performer stesso. Nel sistema ad anello *performer* → *controller* → *computer* → *sintetizzatore* → *performer* sono insiti alcuni

dei concetti tipici di Teoria dei Sistemi quali: *l'instabilità, la controllabilità, linearità, frequenza di campionamento, quantizzazione, latenza.*

L'instabilità indica che un sistema sottoposto a stimoli finiti produce una risposta infinita e non decrescente. Nel nostro caso può apparire nel caso che l'algoritmo controllato non termini autonomamente al termine dell'azione del performer.

Controllabilità indica la semplicità del controllo del sistema e del suo output variazione i parametri di ingresso. La controllabilità può essere bassa o alta: una bassa manovrabilità in genere è costituito da cattive caratteristiche nel percorso diretto del sistema.

Linearità: molti dei sensori utilizzati nei vari controller non sono esattamente lineari. Come noto la non-linearità è presente anche in strumenti musicali tradizionali. Quando la non linearità rappresenta un problema, possono essere utilizzati metodi adeguati per la linearizzazione sfruttando direttamente il meccanismo di mapping (es. *look up table*). In ogni caso i comportamenti non lineari possono essere anche utilizzati per esigenze artistiche e creative: per esempio, la discontinuità può utilizzata nel mapping quando l'uscita desiderata è un trigger al superamento di una soglia mentre l'input è "continuo".

Frequenza di campionamento: nel nostro caso si deve considerare il campionamento del *gesto* che avviene normalmente a frequenze piuttosto basse, circa una decina Hertz (es. un segnale proveniente da un sensore basato sull'analisi di un segnale video). Per controllare in modo continuo un segnale di sintesi sarà quindi necessario filtrare ed interpolare i parametri del gesto, al fine di evitare discontinuità udibili.

Latenza: assente negli strumenti tradizionali, nel nuovo tipo di strumento elettronico due tipi di latenza sono presenti: uno dovuto al sistema di rilevazione gestuale (es. frame rate del sistema di acquisizione video) ed uno relativo al sistema di generazione audio, dovuto al fatto che di solito la sintesi viene fatta su buffer di campioni (da alcune decine ad alcune centinaia) e non da singoli campioni. Una latenza molto alta rende l'anello performer→controller→computer→sintetizzatore→ performer non strettamente controllabile se non addirittura instabile [12].

5. Conclusioni e prospettive

Oltre alle tipologie di applicazioni da noi citate negli ultimi tempi hanno assunto notevole importanza le ricerche per l'interpretazione delle informazioni insite nei cosiddetti *big data set*. In tale ambito lo scopo è quello di evidenziare efficacemente tramite il suono particolari correlazioni o item di interesse, difficilmente rintracciabili in altro modo, sviluppando concetti di multimodal display, cross-modal, interdisciplinary strategy, ecc.

Nonostante una crescita notevole per l'interesse e la fattibilità delle sonificazione negli ultimi due decenni, il campo nel suo complesso continua ad avere una difficile impostazione teorico/virtuale: le decisioni di progettazione sono spesso realizzate affidandosi alle scelte personali e alle conoscenze del singolo progettista. Anche se il concetto di rappresentazione sonora (auditory display) ha beneficiato enormemente di approcci multidisciplinari nel campo della ricerca e della pratica, le molte situazioni di impiego non facilitano la formazione di una specifica teoria unitaria.

Una descrizione teorica strutturata della sonificazione richiederebbe l'assimilazione delle conoscenze degli ultimi decenni in tutta una serie di settori. Nell'ambito della sonificazione non si dovrebbe "reinventare la ruota", nei casi in cui la ricerca in altri campi (acustica, psicoacustica, etc.), ha già offerto intuizioni chiare. Ad esempio in [13] è descritto un primo tentativo di mappare sistematicamente la dimensione uditiva alle rappresentazioni visive e quantitative dei dati con variazioni di frequenza nel tempo. In modo analogo si pensa possono essere sfruttate le altre caratteristiche (ampiezza, timbro

etc.) dei segnali sonori. In tale direzione recentemente sono state presentate ricerche riguardanti l'analisi di data base di movimenti di musicisti, con proposte di tools di ausilio per la progettazione di sistemi di sonificazione [14].

6. Bibliografia

- [1] Bertini G., Chimenti M., Denoth F., *TAU2: un terminale audio per esperimenti di computer music*. In: Alta Frequenza, vol. 46 (1977), Ass. Elettrotecnica ed Elettronica Italiana, pp. 600 - 609.
- [2] Revuelta Sanz P., Ruiz-Mezcua B., Sanchez P., Jose M., *A sonification proposal for travels of blind people*, Proceedings of the 18th International Conference on Auditory Display (ICAD2012) Atlanta, Georgia, June 2012 pp. 18-21.
- [3] Bertini G., Carosi P., *Light baton system: a system for conducting computer music performance*. In: Interface, vol. 22 (1993) Swets & Zeitlinger, pp. 243 – 257.
- [4] Tarabella L., *Pascal Music: un sistema di composizione musicale*, Internal note CNUCE-CNR, B4-88-053, 1988.
- [5] Paradiso, J., *Electronic Music: New Way to Play*, IEEE Spectrum Computer Society Press, 1997, pp.18-30.
- [6] R. Michael Winters, Marcelo M. Wanderley., *New Directions for the Sonification of Expressive Movement in Music Performance*, Proceedings of the International Conf. on Auditory Display, Atlanta, Georgia, June 18-22, 2012, pp. 227-228.
- [7] Bertini G., Magrini M., Tarabella L., *An interactive musical exhibit based on infrared sensor* In: Computer Music Modeling and Retrieval, third International symposium CMMR 2005 (Pisa, Italy, Sept. 26-28, 2005). revised papers. R. Kronland-Martinet, T.Voinier, S.Ystad (eds.). (Lecture notes in computer science, vol. 3902). Heidelberg, Springer, 2006, pp. 92-100.
- [8] www.ted.com/.../pamelia_kurstin_plays_the_theremin
- [9] Tarabella L., *Handel, a free-hands gesture recognition system* Proceedings of Computer Music Modeling and Retrieval (CMMR 2004). LNCS Springer, 2005, pp. 139 -148.
- [10] sidoremi.isti.cnr.it
- [11] Tarabella L., Bertini G. *About the role of mapping in gesture-controlled live computer music* , in Computer Music Modelling and retrieval CMMR2003 (Montpellier, France, May 2003). Proceedings,. Uffe Kock Wiil (ed.). (Lecture Notes in Computer Science, vol. 2771). Springer, 2003, pp. 217-224.
- [12] Tarabella L., Boschi G., Bertini G., *Gesture, mapping and audience in live computer music*. Technical report, CNR/IEI-TR 033, 2001.
- [13] Williams, T. M., Aiken, L. S., *Development of pattern classification: Auditory-visual equivalence in the use of prototypes*. Developmental Psychology, 13(3) 1977, pp. 198-204.
- [14] Winters R.M., Alexandre S., Vincent Verfaille V., Wanderley M.M. : *A Sonification Tool for the Analysis of Large Databases of Expressive Gesture*, International Journal of Multimedia and Its Applications, 4(6), 2012 pp. 13-26.