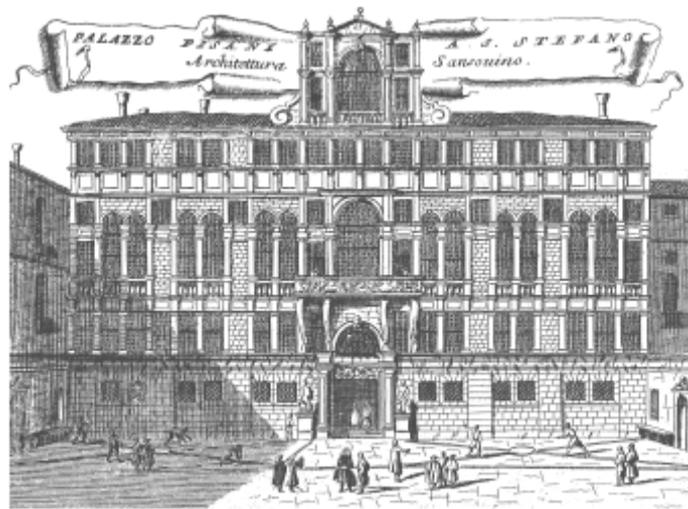


XVII Colloquio di Informatica Musicale



Il Palazzo Pisani di Santo Stefano sede del Conservatorio di Musica di Venezia in una incisione del Cornelli, sec. XVII

**Sala Concerti del Conservatorio “Benedetto Marcello”
Palazzo Pisani
15 - 17 ottobre 2008**



Associazione di Informatica
Musicale Italiana



Conservatorio “B. Marcello”
di Venezia



Ex Novo Ensemble

I
-
U
-
A
-
V

Facoltà di Design e Arti
dell'Università IUAV di
Venezia



La Biennale di Venezia

Indice

Pag.	Contenuto <i>Autore</i>
3.	Programma
7.	“NoaNoa” for Flute and Electronics: problemi e strategie di esecuzione della parte elettronica. <i>Simone Conforti</i>
17.	Il ruolo del movimento corporeo nell’esecuzione musicale in presenza di feedback uditivo ritardato <i>Marco Matteazzi</i>
23.	Dalla Computer Music al Sound and Music Computing: Contesti culturali e tecnologici <i>Nicola Bernardini, Alvise Vidolin</i>
29.	Voice Painter: un’interfaccia multimodale per dipingere con la voce <i>Amalia de Götzen, Riccardo Marogna, Federico Avanzini</i>
35.	Un modello per la generazione dinamica di paesaggi sonori <i>Mattia Schirosa, Andrea Valle, Vincenzo Lombardo</i>
43.	Del suonare l’elettronica <i>Gabriel Bacalov, Massimo Marchi, Francesco Vitale</i>
47.	Principi elementari per una sonificazione del gesto <i>Maurizio Goina, Pietro Polotti</i>
53.	Meta-Morfologie Musicali: un ambiente di composizione elettroacustica <i>Riccardo Dapelo, Roberto Doati, Emilio Pozzolini, Alessandro Quaranta, Martino Sarolli, Corrado Canepa, Barbara Mazzarino</i>
59.	Sostituzione Incrociata della Fase nel Dominio di Scala <i>Antonio De Sena</i>
65.	La Computer Vision per l’annotazione automatica di documenti audio <i>Lauro Snidaro, Sergio Canazza</i>
69.	Analisi del comportamento di fruitori di sistemi di ascolto attivo: il caso dell’Esploratore d’Orchestra <i>Francesca Cavallero, Antonio Camurri, Corrado Canepa, Gualtiero Volpe</i>
77.	Modelli VCF discreti a retroazione istantanea <i>Marco Civolani, Federico Fontana</i>
85.	La ricostruzione dello Studio di Fonologia della Rai di Milano <i>Maddalena Novati</i>
89.	Un approccio alle interfacce acustiche tangibili basato sulla trasformata wavelet di onde dispersive <i>Antonio Camurri, Corrado Canepa, Chiara Erra, Alberto Massari, Gualtiero Volpe, Luciano Fava</i>

95. Un modello per la simulazione in tempo reale di un riverbero a piastra
Stefano Zambon
103. Un campionatore per “Musica su due dimensioni” di Bruno Maderna
Marco Gasperini
109. Sviluppo di un’interfaccia audio-aptica basata sulla profondità spaziale
Stefano Delle Monache, Delphine Devallez, Pietro Polotti, Davide Rocchesso
115. Omaggio a Stockhausen - Technical set-up digitale per una performance di SOLO [Nr.19] für melodieninstrument mit rückkopplung
Enrico Francioni
125. Comporre (con) oggetti audio-visivi: un ambiente di lavoro in via di sviluppo
Alessio Santini
131. Generazione di suoni e musica mediante sistemi caotici
Costantino Rizzuti
139. DepThrow: uno strumento di indagine sulla percezione uditiva della distanza in forma di gioco audio
Stefano Papetti, Delphine Devallez, Federico Fontana
145. CHORALE - Compositional Hyper ORganized ALgorithmic Environment
Gianni Della Vittoria
149. Suoni scolpiti e sculture sonore: alcuni esempi di installazioni d’arte elettroacustica
Silvia Lanzalone
157. Oggetti sonori nella rete
Dante Tanzi
161. Rendering acustico di liquidi in movimento rappresentati con modelli a particelle
Carlo Drioli, Davide Rocchesso
165. Echi tra le Volte: un progetto di sound design rivolto agli edifici di culto
Andrea Taroppi
171. Photos of GHOSTS (Photos of Grooves and HOles, Supporting Tracks Separation)
Conservazione attiva di dischi fonografici per mezzo di immagini digitali
Sergio Canazza, Giovanni Ferrin, Lauro Snidaro
177. Atomi distratti di Mario Garuti (2005) - Analisi musicale e tecnologica ed elaborazione di un algoritmo di sintesi in ambiente Max/MSP per l’esecuzione della parte di tastiera elettronica
Marco Marinoni
191. AudioDrome
Alessandro Petrolati
199. Score Following in Tempo Reale basato su Hidden Markov Models
Nicola Montecchio

Voice Painter: un'interfaccia multimodale per dipingere con la voce

Amalia de Götzen, Riccardo Marogna, Federico Avanzini
Dipartimento di Ingegneria dell'Informazione
Università di Padova
{degotzen, marognar, avanzini}@dei.unipd.it

SOMMARIO

Il presente lavoro descrive un sistema - denominato *voice painter* - che permette di dipingere su una tela virtuale attraverso l'uso della propria voce e del movimento del corpo. Tale sistema, trasforma lo spettatore della performance multimodale in un performer/en-attore, proponendo così una nuova metafora artistica.

Parole Chiave

Enazione; Informatica Musicale; Sound and Music Computing; Performing Arts.

INTRODUZIONE

L'argomento di questo lavoro è inerentemente multidisciplinare: per scoprire ed enfatizzare l'approccio enattivo nelle produzioni artistiche (danza, musica, pittura, scultura, ecc.) usando le nuove tecnologie, è necessario colmare il divario tra tecnologia e arte prendendo in considerazione i suggerimenti e le richieste degli artisti e le limitazioni e possibilità proposte dai tecnici.

Le arti e l'enattività possono essere considerati strettamente correlati anche nei lavori del passato, dove le tecnologie possono essere, semplicemente, un pennello e qualche colore, o il materiale di una scultura e gli strumenti utilizzati dall'artista per creare il proprio lavoro. In questi casi il ruolo dell'enazione, per coloro che esploreranno l'oggetto artistico, è limitato al muoversi intorno a quest'ultimo per coglierne i vari aspetti del suo significato. Due esempi molto noti ed appropriati sono i lavori di Close nella pittura [7] e quelli di Serra nella scultura [14].

Tuttavia, le nuove tecnologie di espressione artistica che contemplano l'interazione multimodale danno all'artista nuovi strumenti e nuovi modi di pensare al proprio lavoro, coinvolgendo gli spettatori in un'interazione enattiva nella fruizione di una specifica opera d'arte.

La realizzazione del sistema *voice painter* prende spunto dall'idea di esplorare ed enfatizzare l'approccio enattivo in ambito artistico attraverso l'utilizzo di tecnologie appropriate. Nelle forme artistiche classiche molto spesso lo spettatore è forzato dall'opera stessa ad esplorare, a muoversi e ad usare il proprio corpo per mettersi in relazione con l'opera d'arte. La mediazione della tecnologia, può enfatizzare questo tipo di relazione ed il sistema che viene descritto nei paragrafi che seguono ne è un esempio. Noi pensiamo che questo

approccio sia genuinamente enattivo e che suggerisca una "terza via" tra arti allografiche ed arti autografiche.

Questo saggio è organizzato in tre sezioni. La prima sezione riassume lo stato dell'arte e i principali punti di vista riguardanti il come le esperienze enattive informano le rappresentazioni artistiche. In questo contesto, le differenze tra arti autografiche e arti allografiche verranno analizzate nei termini delle loro proprietà enattive. La seconda sezione presenterà e descriverà i dettagli tecnici del sistema *voice painter*: verranno delineate le componenti *hardware* e *software* e verranno motivate le strategie principali di design dell'interazione. La terza sezione è dedicata alle possibili applicazioni del sistema, dalla discussione di uno scenario reale nel quale il sistema è già stato sperimentato da numerosi utenti sino ai miglioramenti futuri e all'analisi di nuovi scenari applicativi.

ARTI AUTOGRAFICHE E ARTI ALLOGRAFICHE

Una delle principali categorizzazioni tra le diverse forme d'arte è quella di Goodman [10] il quale definisce le arti 'autografiche' e le arti 'allografiche':

the former cannot be noted and do not contemplate performance, while the latter can be translated into conventional notation, and the resulting 'score' can be performed with a certain freedom of variation.

(le prime non possono essere trascritte e non contemplano l'esecuzione, mentre le seconde possono essere trascritte in notazione convenzionale, e la 'partitura' risultante può essere eseguita con una certa libertà di variazione). La pittura e la musica sono le due espressioni artistiche generalmente utilizzate come esempi di questa distinzione. È difficile determinare le regole che abbiano generato un dato dipinto, e non c'è notazione o scrittura che possa aiutare qualcun altro a produrre una replica esatta dell'opera d'arte originale: è persino possibile definire ogni copia un 'falso'. In musica il punto di vista è completamente diverso: ciascuna copia/esecuzione di un brano è una possibile interpretazione. La scrittura permette a tanti musicisti diversi di suonare un dato lavoro musicale: i segnali musicali 'discreti' vengono dapprima scritti dal compositore e poi interpretati dai musicisti. Si può dire che mentre le arti autografiche sono arti ad uno stadio, le arti allografiche sono arti a due stadi. La distanza tra queste due forme d'arte può ridursi drasticamente nelle esecuzioni moderne nelle quali, per esempio, un dipinto può essere il risultato di una performance dal vivo: un/a ballerino/a che dipinge con il proprio corpo o un musicista

che controlli una tecnologia multimodale che produce una parte video mentre suona.

L'Enattività nelle Arti

Uno dei principali risultati del progetto ENACTIVE¹ è stata la profonda e fruttuosa riflessione sul ruolo dell'enattività nel processo creativo artistico [3]. Questo argomento è particolarmente difficile da affrontare perché collega insieme concetti astratti che sfuggono alle definizioni (enazione, creazione): la conferenza Enactive/07, tenutasi a Grenoble, ha messo insieme numerosi contributi che possono essere analizzati per tracciare una sorta di "filo rosso" che colleghi diverse espressioni artistiche con l'intento comune di esplorare il processo creativo enattivo.

La teoria della percezione enattiva dichiara che non è possibile disaccoppiare schematicamente la percezione e l'azione, dal momento che ogni tipo di percezione è intrinsecamente attiva e premeditata. La "concezione enattiva" dell'esperienza è qualcosa che un animale *enagisce* nell'esplorazione del proprio ambiente [15, 20]. Da questo punto di vista, il soggetto degli stati mentali è tutto il corpo dell'animale (*embodied*) situato nel suo ambiente. La conoscenza enattiva viene poi immagazzinata in forma di risposte motorie acquisite attraverso l'atto del "fare" [12].

Il sapere enattivo può essere acquisito anche esplorando un dipinto o una scultura se il fruitore è immerso in questo anello di azione-percezione. Il tipico esempio di un'arte enattiva è la musica: un violinista ha bisogno di sentire e di ascoltare il suono per correggere ed affinare l'esecuzione. In questa prospettiva molte applicazioni artistiche enattive create con il supporto di tecnologie diverse esplorano strumenti virtuali attraverso diversi tipi di gesti e di posture. Nello stesso tempo, queste applicazioni devono prendere in considerazione il feedback specifico ricevuto, ad es., esplorando una superficie o utilizzando un arco su una corda [5]: noi percepiamo attraverso le nostre mani e le nostre dita specifiche sensazioni aptiche che stimolano il fruitore/esecutore a reagire per capire. Gli strumenti musicali virtuali devono quindi essere corredati da tecnologie aptiche che possano restituire le superfici e le forze coinvolte nel suonare uno strumento reale. Allo stesso modo, molte creazioni artistiche sono concentrate sull'uso della retroazione aptica per esplorare l'ambiente e, per esempio, un oggetto/scultura sul quale potremmo volgere il nostro sguardo.

Dipingere con la voce

L'uso della voce come strumento enattivo è un argomento piuttosto inesplorato, in particolare nel contesto di possibili applicazioni artistiche. La voce è un mezzo di comunicazione umana universale, e viene utilizzato anche per trasmettere elementi non-verbali e paralinguistici: emozioni, prosodia, stress. In più, la voce e la parola sono sempre accompagnate da altri canali di comunicazione non-verbale, tra i quali l'espressione facciale, il gesto e il linguaggio del corpo. Una visione influente nella psicologia cognitiva e nella linguistica stabilisce che il gesto e il parlato formano un

sistema di comunicazione unico, e che il gesto è una componente integrale della comunicazione anziché un semplice accompagnamento o ornamento [13].

Queste osservazioni forniscono la motivazione di base per lo sviluppo di un'interfaccia che utilizzi l'espressione vocale come mezzo per creare segni visivi. L'idea centrale è di esplorare le caratteristiche più rilevanti dell'espressione vocale (incluso anche caratteristiche prosodiche come l'altezza, l'intensità, ecc.) e di correlarle a caratteristiche grafiche, creando così uno strumento semplice e versatile che possa essere utilizzato da un esecutore esperto ma anche da un utilizzatore alle prime armi.

L'esecutore potrà dipingere su uno schermo nero utilizzando la propria voce, e potrà scoprire, nel contempo, le correlazioni tra il suono prodotto ed il dipinto che apparirà in tempo reale. La voce potrà essere considerata come un pennello: per disegnare sull'intera superficie dello schermo l'esecutore dovrà muoversi, coinvolgendo così tutto il corpo e non la sola voce. L'anello chiuso azione-percezione verrà così ricreato con l'aiuto di un sistema che verrà descritto nella sezione che segue.

DESCRIZIONE DEL SISTEMA

Il sistema è costituito da un array di 8 microfoni i cui segnali sono processati in tempo reale per la localizzazione ed il tracking della posizione dell'utente. Si suppone che l'utente si muova in un'area ad una determinata distanza dallo schermo, definendo un piano di interesse. Il rendering grafico è retro-proiettato sullo schermo al fine di ottenere una perfetta corrispondenza tra la posizione della sorgente vocale e la posizione del rendering grafico da questa prodotto. In questo modo il feedback visivo è temporalmente e spazialmente correlato con l'evento sonoro. Dal segnale processato vengono inoltre estratti una serie di parametri indicativi di alcune caratteristiche notevoli del segnale vocale, che vengono poi mappati in altrettante caratteristiche del tratto grafico.

Localizzazione e tracking in tempo reale della sorgente vocale

Il sistema audio realizza la localizzazione ed il tracking del performer attraverso un algoritmo organizzato in 2 fasi. Dapprima, per ciascuna coppia di sensori viene stimato il ritardo di arrivo (TDOAE, *Time Delay Of Arrival Estimation*). Il vettore di ritardi così stimati viene processato attraverso un algoritmo di stima ai minimi quadrati al fine di ottenere la coppia di coordinate (x,y) nel piano di interesse. Una stima del rapporto tra le ATF (*Acoustical Transfer Functions*) [9]. Il vettore di ritardi risultante da questa prima fase viene confrontato con gli elementi di una matrice pre-calcolata contenente un set di vettori ottenuti tramite discretizzazione del piano di interesse ed il successivo calcolo, per ciascuna posizione così ottenuta, del corrispondente vettore di ritardi. La procedura di *fitting* ai minimi quadrati del vettore stimato con gli elementi della matrice delle posizioni realizzabili fornisce quindi una stima della posizione (x,y) della sorgente vocale. Questa fase della procedura di localizzazione può essere molto pesante dal punto di vista computazionale se si desidera una certa accuratezza nella stima; essa può es-

¹<http://www.enactive.org>

sere altresì alleggerita con l'introduzione di un algoritmo di tracking della sorgente. Tale algoritmo di stima della posizione è stato implementato in linguaggio C ed integrato nella piattaforma per l'elaborazione audio in tempo reale Pure Data [17]. I segnali audio vengono acquisiti ad una frequenza di campionamento pari a 32 kHz (un compromesso tra accuratezza e velocità).

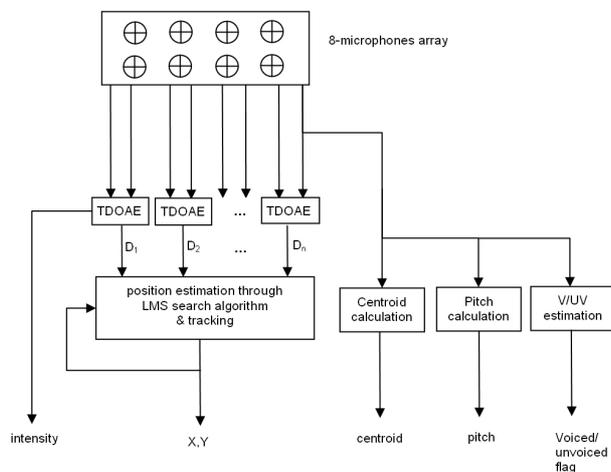


Figura 1. Schema a blocchi del sistema di localizzazione ed estrazione delle features.

Mappatura dei parametri vocali e rendering grafico

La voce può essere definita da numerose caratteristiche che gli esseri umani sono in grado di controllare tramite la loro emissione vocale.

L'interfaccia multimodale è stata progettata al fine di realizzare una mappatura delle caratteristiche rilevanti della voce in qualità grafiche ben riconoscibili (forma, dimensione, colore) seguendo un approccio il più possibile naturale.

Tale mappatura è illustrata in Fig.2. Tra i diversi parametri che possono caratterizzare la voce umana ne sono stati identificati alcuni notevoli che potessero essere controllati facilmente dal performer e stimati in tempo reale con un certo grado di robustezza:

- *Intensità*, calcolata come RMS il quadrato della pressione del segnale;
- *Centroide*, il centro di gravità della ampiezza dello spettro calcolata su un frame audio;
- *Suono vocalizzato/non vocalizzato*, caratteristica che dipende dalla presenza o meno di vibrazioni pseudo periodiche
- *Altezza*, l'attributo soggettivo della frequenza di un suono;

Sebbene per le prime due caratteristiche sia stata utilizzata la definizione comune e siano facili da estrarre, le altre due non sono altrettanto banalmente stimabili. L'attributo relativo al suono vocalizzato indica la presenza di periodicità dovuta alle vibrazioni delle corde vocali. Questo tipo di segnali possono essere ricavati tramite diversi approcci. Qui è stata im-

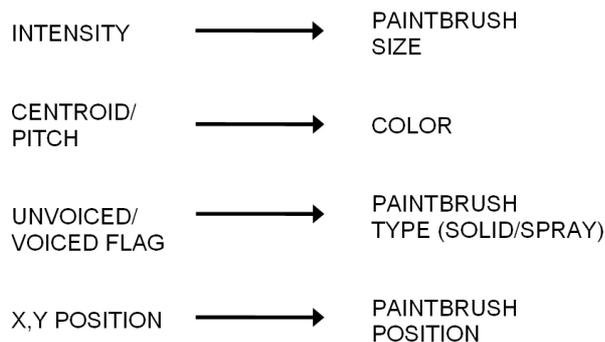


Figura 2. Mappatura delle caratteristiche vocali nelle caratteristiche grafiche.

plementata la tecnica che combina il rilevamento dello zero-crossing e l'estrazione mediante cepstrum [1]. L'estrazione dell'altezza è in questo caso facilitata dalla struttura relativamente semplice dei segnali vocalizzati, così che il problema si riduce alla stima della frequenza fondamentale. Questa è stimata usando un algoritmo che estrae e correla le componenti spettrali armoniche su frames successive del segnale vocale [16].

Il rendering grafico è stato realizzato utilizzando la libreria GEM (Graphical environment for Multimedia [21]), che ben si integra nella piattaforma Pure Data. Sono state utilizzate caratteristiche grafiche ben riconoscibili (dimensione, colore, geometria) con l'intento di seguire un approccio il più possibile naturale. Questo è un problema non triviale dato che il termine "naturale" non può essere definito a priori, ma può essere solo definito sulla base delle proprie esperienze personali o su scelte arbitrarie (questo problema può essere confrontato con il suo duale, la sonificazione). La naturalezza di un dato mapping dovrebbe quindi essere testata su diversi soggetti in modo da essere giustificata ed eventualmente modificata. Questo risulta particolarmente vero in alcune delle scelte fatte per il voice painter: la corrispondenza tra la frequenza di un suono con la frequenza della luce, anche se piuttosto logica scientificamente, potrebbe non necessariamente essere di facile (e.g. naturale) comprensione per l'utente. Un'altra possibile mappatura potrebbe essere quella sfruttando l'insieme RGB dei colori, associando suoni gravi a colori "caldi" e suoni acuti a colori "freddi". L'effetto visivo risultante è una sorta di schizzo astratto che può contenere elementi geometrici ben definiti (corrispondenti a brevi tratti vocalizzati) e tratti nebulari dovuti a segnali non vocalizzati, entrambi scalati a diverse dimensioni e con diversi colori (Fig.3).

DISCUSSIONE

L'interfaccia è stata sottoposta ad un test preliminare in occasione della Conferenza ENACTIVE 07 svoltasi a Grenoble. Estratti video di alcune sessioni di test possono essere consultati al seguente link: www.smc.dei.unipd.it/voice_painter.html. Ciascun utente poteva testare il sistema interagendo con la tela virtuale senza alcuna spie-

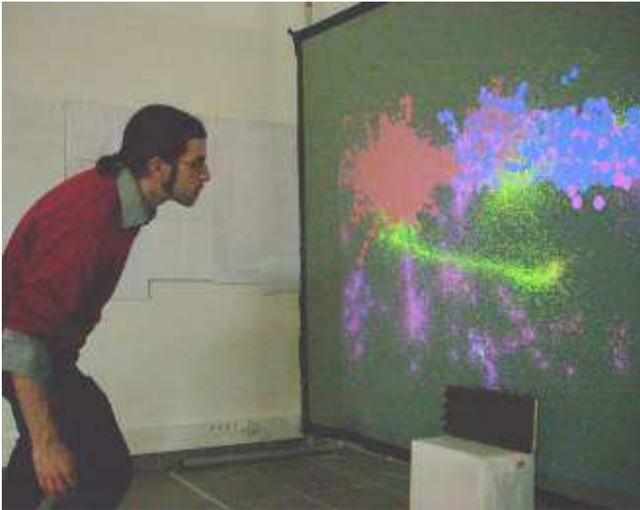


Figura 3. Il voice painter in azione.

gazione preliminare relativa al mapping e senza alcuno specifico task da compiere. L'unica indicazione data era "Usa la tua voce per dipingere". L'obiettivo del test era di valutare

- il modo in cui l'utente si relazionava allo schermo e come esplorava le diverse caratteristiche vocali;
- quali caratteristiche vocali l'utente identificava nel mapping proposto;
- se l'utente era in grado di controllare la/le caratteristica/e identificate in modo da ottenere l'effetto grafico desiderato.

Alla fine del test all'utente era richiesto di giudicare la naturalezza dell'esperienza vissuta. Questo test preliminare ha rivelato che la maggior parte delle caratteristiche vocali utilizzate erano state individuate, con l'unica eccezione del mapping altezza/centroide-colore. Sebbene il test sia stato realizzato in maniera piuttosto informale, la presenza di molte persone che assistevano alle "performances" molto spesso inibiva l'utente, lasciandolo meno libero e spontaneo nell'esplorazione del sistema. Al tempo stesso il feedback visivo è risultato utile per incoraggiare l'utente ad usare la voce nei modi più disparati per ottenere effetti grafici interessanti e particolari. Il prossimo passo di questo lavoro riguarderà la realizzazione di test quantitativi con sistematica raccolta dei commenti degli utenti.

Localizzazione e tracking del parlatore

Uno degli aspetti che si sta attualmente sviluppando riguarda il miglioramento della localizzazione e del sistema di tracking, con particolare riguardo alla robustezza degli algoritmi in contesti rumorosi e quindi più vicini a scenari reali. La stima tramite TDOA e l'algoritmo LMS descritti nella seconda parte di questo articolo, considerano solo sottofondi con rumori impulsivi o stazionari e la performance peggiora in presenza di bassi valori di SNR, specialmente quando il rumore è non stazionario. Per aumentare la robustezza di tale sistema, il primo miglioramento nella procedura di

tracking riguarda l'inclusione di un algoritmo più generale di rilevamento di attività vocale (VAD), che impiega statistiche più accurate che permettono il rilevamento della presenza o meno del parlato in qualsiasi regione dello spettro audio [19]. È stato sviluppato anche un sistema bimodale nel quale la localizzazione ed il tracking sono realizzati non solo sfruttando il segnale vocale ma anche tramite quello video. Una implementazione preliminare è stata sviluppata usando la piattaforma software EyesWeb per l'analisi gestuale [6].

Un ulteriore futuro sviluppo riguarda il problema della localizzazione simultanea di più sorgenti acustiche, che richiede l'utilizzo di più avanzati metodi statistici, come il *particle filtering* [2].

Riabilitazione vocale

Oltre alle applicazioni in ambiti artistici e di intrattenimento, il sistema che abbiamo proposto apre possibili strade di ricerca nella riabilitazione vocale. Le tecniche derivate dall'HCI sono sempre più utilizzate come metodi per insegnare e rafforzare le capacità vocali in diversi contesti, con particolare riferimento al processo educativo di bambini che abbiano alcuni deficit di comunicazione [4, 8, 11, 18]. Possibili aree di applicazione includono sia disfunzioni della parola dovuti a disabilità fisiche o a problemi nel coordinamento (dysarthria, dyspraxia) per i quali gli esercizi di articolazione, effettuati molto spesso durante attività ludiche, sono usati per mostrare ai bambini come pronunciare correttamente i suoni e le sillabe. Precedenti lavori hanno mostrato che sistemi di training del parlato basati su tecniche informatiche possono essere molto utili per suggerire il miglioramento dell'articolazione [4] e per aiutare l'apprendimento della corretta pronuncia di suoni specifici [18].

Una seconda possibile area di applicazione riguarda i problemi di ritardo del linguaggio (i.e. situazioni in cui lo sviluppo del linguaggio non segue il pattern usuale, ma ha un andamento più lento del normale) [8] ed infine la terza area è relativa a problemi di comunicazione, legati allo spettro dei disordini di tipo autistico (ASD) [11]. Il voice painter presentato in questo lavoro incoraggia il comportamento ludico dell'utente, maggiormente a suo agio nell'interazione con una macchina che non lo intimidisce, ma che anzi riduce o addirittura elimina il disagio causato dall'interazione con altre persone.

CONCLUSIONI

Questo articolo presenta un'interfaccia che apre numerose vie di esplorazione. Il voice painter è nato pensando ad applicazioni artistiche: creare uno strumento che potesse mediare tra le arti autografiche ed allografiche, dando vita a performances sofisticate basate su una notazione musicale (usata poi per dipingere), ad improvvisazioni o a semplici occasioni di intrattenimento. Lo strumento è stato mostrato alla Conferenza Enactive/07 ed ha suscitato particolare interesse tra artisti visivi. Diversi scenari applicativi sono quindi stati proposti, suggerendo futuri sviluppi: oltre alle applicazioni artistiche, risultano particolarmente promettenti quelle relative a terapie vocali e a problemi di comunicazione.

RINGRAZIAMENTI

Si ringraziano Barbara Mazzarino e Giovanna Varni del DIST di Genova per i loro preziosi suggerimenti e per il loro aiuto nello sviluppo della prima implementazione del sistema, presentato alla Conferenza Enactive/07 a Grenoble..

RIFERIMENTI

1. S. Ahmadi and A. S. Spanias. Cepstrum-based pitch detection using a new statistical v/uv classification algorithm. *IEEE Trans. on Speech and Audio Processing*, vol. 7, no. 3, May 1999.
2. F. Antonacci, D. Riva, M. Tagliasacchi, and A. Sarti. Efficient localization and tracking of two acoustic sources using particle filters with swarm intelligence. In *Proc. EURASIP European Sig. Process. Conf.*, Poznan, 2007.
3. Association ACROE, editor. *Proc. 4th Int. Conf. on Enactive Interfaces (ENACTIVE/07... Enaction in Arts)*, Grenoble, 2007.
4. O. Bälter, O. Engwall, A.-M. Öster, and H. Kjellström. Wizard-of-oz test of artur: a computer-based speech training system with articulation correction. In *Proc. 7th Int. ACM SIGACCESS Conf. on Computers and accessibility (ASSETS'05)*, pages 36–43, Baltimore, 2005.
5. C. Cadoz. Musical creation process and digital technology. the supra-instrumental gesture. In *4th International Conference on Enactive Interfaces (Enactive'07)*, Grenoble, France, November 2007.
6. A. Camurri, B. Mazzarino, and G. Volpe. Analysis of expressive gesture: The EyesWeb expressive gesture processing library. In A. Camurri and G. Volpe, editors, *Gesture-based Communication in Human-Computer Interaction*. LNAI 2915, Springer Verlag, 2004.
7. C. Close, R. Storr, K. Varnedoe, D. Wye, and G. D. Lowry. *Chuck Close*. The Museum of Modern Art, New York, 2002.
8. H. Fell, C. Cress, J. MacAuslan, and L. Ferrier. visiBabble for reinforcement of early vocalization. In *Proc. 6th Int. ACM SIGACCESS Conf. on Computers and accessibility (ASSETS'04)*, pages 161–168, Atlanta, 2004.
9. S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and non-stationarity with application to speech. *IEEE Trans. Signal Processing*, 49 (8), 2001.
10. N. Goodman. *Languages of Art*. Hackett, 1985.
11. J. Hailpern, K. Karahalios, J. Halle, L. DeThorne, and M.-K. Coletto. Visualizations: speech, language & autistic spectrum disorder. In *Proc. ACM Computer-Human Interaction Conf. (CHI'08)*, pages 3591–3596, Firenze, 2008.
12. A. Luciani, J. Florens, and N. Castagne. From action to sound: a challenging perspective for haptics. In *Workshop on Enactive Interfaces (Enactive'05)*, Pisa, Italy, January 2005.
13. D. McNeill. *Gesture and Thought*. University of Chicago Press, Chicago, 2005.
14. K. McShine, L. Cooke, J. Rajchman, B. Buchloh, and R. Serra. *Richard Serra Sculpture: Forty Years*. The Museum of Modern Art, New York, 2007.
15. A. Noe. *Action in perception*. MIT press, Cambridge, Mass., 2005.
16. M. Puckette and T. Apel. Real-time audio analysis tools for pd and msp. In *Proc Int. Computer Music Conf.*, pages 109–112, San Francisco, 1998.
17. M. S. Puckette. Pure data: another integrated computer music environment. *Proc. of Int. Computer Music Conf.*, pages 37–41, 1996.
18. L. I. Shuster, D. M. Ruscello, and A. R. Toth. The use of visual feedback to elicit correct /r/. *Am. J. Speech-Language Pathology*, 4:37–44, 1995.
19. S. G. Tanyer and H. Ozer. Voice activity detection in nonstationary noise. *IEEE Trans. Speech Audio Process.*, 8(4):478–482, 2000.
20. F. Varela, E. Thompson, and E. Rosch. *The Embodied Mind*. MIT Press, Cambridge, MA, 1991.
21. J. M. Zmólnig. Gem, <http://gem.iem.at>.