

AUDIO 3D E ANCORAGGIO SONORO PER L'ESPLORAZIONE MULTIMODALE DI AMBIENTI VIRTUALI

Michele Geronazzo

Università di Padova
Dip. di Ingegneria dell'Informazione
geronazz@dei.unipd.it

Luca Brayda

Fondazione Istituto Italiano di Tecnologia
Dip. Robotics, Brain and Cognitive Sciences
luca.brayda@iit.it

Alberto Bedin e Federico Avanzini

Università di Padova
Dip. di Ingegneria dell'Informazione
avanzini@dei.unipd.it

SOMMARIO

Questo lavoro presenta un sistema interattivo audio-aptico di ausilio all'orientamento e alla mobilità per soggetti non vedenti, e un esperimento soggettivo volto a studiare i meccanismi cognitivi nella rappresentazione spaziale in assenza di informazione visuale. Si presenta in particolare un esperimento di riconoscimento di oggetti, che investiga il ruolo dell'informazione auditiva spaziale dinamica integrata con feedback aptico, in un semplice ambiente virtuale. Tale informazione è strutturata come una "ancora sonora", erogata in cuffia tramite tecniche di rendering 3D binaurale del suono (in particolare tramite *Head-Related Transfer Functions*, o HRTF, opportunamente personalizzate). I risultati sperimentali relativi al tempo di riconoscimento da parte dei soggetti mostrano una relazione tra la posizione dell'ancora sonora e la forma dell'oggetto riconosciuto. Inoltre, un'analisi qualitativa delle traiettorie di esplorazione suggerisce l'insorgere di modifiche comportamentali tra le condizioni monomodale e multimodale.

1. INTRODUZIONE

La nostra interazione con il mondo è intrinsecamente multimodale. In particolare, le interazioni che coinvolgono la modalità uditiva sono legate all'esplorazione attiva dell'ambiente. La natura multimodale dell'interazione rende difficile comprendere e quantificare la rilevanza dell'informazione uditiva nello svolgimento di attività motorie complesse, poiché essa dipende dall'accuratezza e l'accessibilità dell'informazione associata alle diverse modalità sensoriali [1]. Inoltre diverse modalità si combinano e si integrano [2], anche attraverso corrispondenze crossmodali [3]. La comprensione di questi meccanismi è essenziale per progettare interfacce uomo-macchina in cui vari flussi di informazione unimodale si fondano efficacemente in un singolo display interattivo multimodale.

Molti sistemi interattivi in diversi ambiti applicativi fanno un uso crescente dell'informazione spaziale legata al suono, come dimostrato anche dall'introduzione del termine "auralizzazione" già all'inizio degli anni '90 del secolo scorso [4]. Una panoramica [5] sugli approcci usati in

letteratura nella sonificazione di quantità fisiche, categorizzati per strategie di mapping e la loro relazione con il mondo fisico, mostra che la dimensione spaziale uditiva è usata tipicamente per sonificare quantità cinematiche. Ciò motiva un'indagine sulla rilevanza di tale dimensione in un contesto di interazione continua con un ambiente virtuale.

Il feedback uditivo assieme a quello aptico è stato usato in letteratura per approcci di sostituzione sensoriale, allo scopo di verificare se l'acquisizione di informazione spaziale attraverso modalità non visuali rispecchia rappresentazioni interne costruite attraverso la vista [6]. Un esempio notevole di questo tipo di sistemi di sostituzione sensoriale è costituito da dispositivi per l'orientamento e la mobilità (*Orientation & Mobility*, O&M) per soggetti non-vedenti, i quali includono riconoscimento di oggetti/ostacoli, definizione di punti di riferimento, assistenza nella pianificazione di tragitti, e così via.

Un altro aspetto rilevante e non completamente compreso in letteratura è la relazione tra l'informazione sensoriale (sia quella "idiotetica", associata a comandi motori per il movimento e alla propriocezione, sia quella "allogenetica", associata alla percezione dell'ambiente esterno), e i livelli di attenzione e memoria coinvolti nei processi di apprendimento spaziale [7]. Un dispositivo multimodale per applicazioni in ambito O&M dovrebbe minimizzare i conflitti intermodali, fornire informazioni ecologicamente consistenti con quelle disponibile nel mondo reale, e quindi non appesantire eccessivamente il carico cognitivo richiesto all'utente.

2. COGNIZIONE SPAZIALE

Nella nostra società, la visione è la modalità sensoriale principale nell'espletamento di innumerevoli attività quotidiane, così come nella comunicazione. In particolare, una semplice attività di esplorazione e riconoscimento della forma di un oggetto si complica notevolmente in assenza di informazione visiva. Lederman e Klatzky [8] hanno studiato il legame tra movimenti della mano ed esplorazione aptica di oggetti, con soggetti non-vedenti. I loro risultati mostrano che l'informazione aptica è utile nel riconoscimento di un singolo oggetto, ma non sufficiente in un task più complesso di riconoscimento di layout e strutture spaziali (ad esempio un ambiente 2D con dati di altezza [9]). In maniera complementare, Yu et al. [10] hanno mostrato che la percezione spaziale e la propriocezione sono invece elementi necessari all'esplorazione non-visuale.

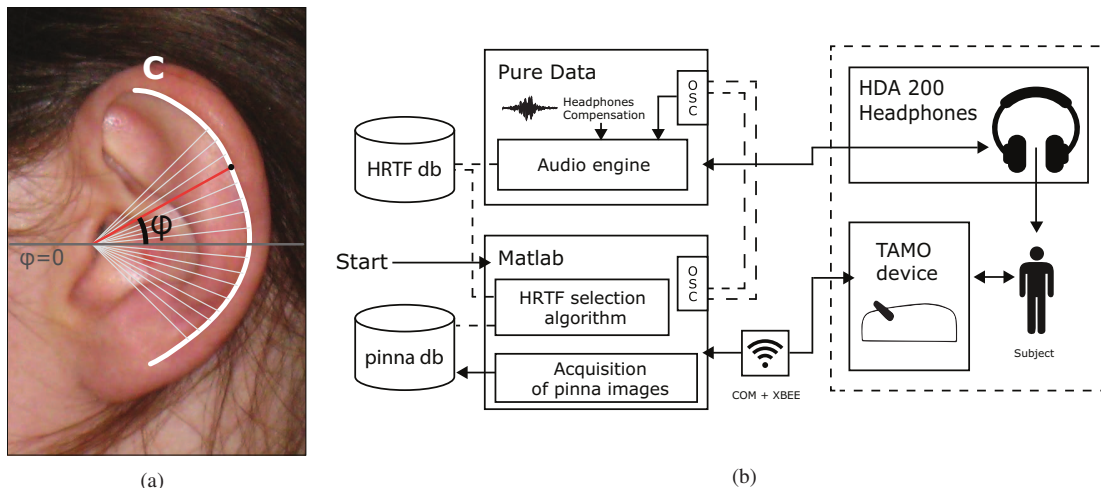


Figura 1. (a) Immagine di lato e contorni dell'orecchio esterno di un soggetto. (b) Uno schema semplificato della configurazione sperimentale.

Nello specifico ambito degli ausili per O&M, la letteratura presenta molti sistemi aptici di sostituzione sensoriale. Le principali limitazioni riguardano gli ingombri, come nel caso del OPTical to TActile CONverter (Optacon) [11], o conflitti intersensoriali, specialmente a danno della modalità uditiva quando presente [12]. Tra i dispositivi proposti, il sistema *HOMERE* [13] usa la metafora del “bastone bianco” per l'esplorazione lungo tragitti predefiniti, e richiede un ingombrante sistema di audio multicanale. Lahav et al. [14] hanno svolto studi qualitativi sull'esplorazione di mappe da parte di soggetti non-vedenti, mostrando una riduzione dei tempi di esplorazione di un ambiente reale, nel caso di soggetti sottoposti a training in ambiente virtuale.

Un ambiente desktop di realtà virtuale più completo e complesso è *BlindAid* [15], il quale usa un Phantom[®] 1 per il rendering aptico. Azioni e comandi sono tipici di molti contesti di interazione uomo-macchina (come zooming, scrolling, undoing, ecc.). Lo scopo è accelerare il processo di apprendimento spaziale. Il suono è reso spazialmente attraverso *Head-Related Transfer Functions* (HRTF), ossia quei filtri peculiari di ogni ascoltatore che catturano gli effetti acustici della testa umana e delle orecchie [16]. Tuttavia questo sistema fa uso di HRTF generiche misurate su un manichino KEMAR², invece che HRTF individuali, pregiudicando l'accuratezza dell'informazione uditiva.

3. ESPLORAZIONE AUDIO-TATTILE

La soluzione tecnologica proposta in questo articolo segue l'idea del “tablet virtuale audio-tattile” per l'esplorazione attiva non visuale di una mappa. L'informazione “locale” viene resa tramite feedback aptico (ad es. urti con oggetti incontrati durante l'esplorazione), mentre un'informazione “globale” viene fornita tramite feedback uditivo (in parti-

colare un punto di ancoraggio sonoro, fisso nell'ambiente virtuale).

I punti di ancoraggio sono collegati ad informazioni tattili locali in relazione biunivoca con ogni punto disponibile nell'area di lavoro; è stato utilizzato un rendering audio basato su HRTF personalizzate, che permettono una simulazione fedele del segnale audio che arriva all'ingresso del canale uditivo in funzione della posizione della sorgente sonora rispetto all'ascoltatore. Le tecnologie binaurali sono infatti fortemente legate alle caratteristiche antropometriche e percettive individuali e hanno un ruolo fondamentale nella caratterizzazione e modellazione di HRTF [17].

3.1 Spazializzazione con audio binaurale

Il nostro sistema utilizza un'innovativa procedura di selezione di HRTF generiche. Per ogni soggetto si seleziona un set di HRTF da un database utilizzando la procedura descritta in [18], con l'obiettivo di ottimizzare automaticamente l'impressione spaziale individuale. L'acquisizione di un'immagine dell'orecchio esterno permette di calcolare una distanza *ad hoc* definita tra contorni dell'orecchio esterno e le caratteristiche a livello di segnale delle HRTF presenti nel database.

Le frequenze centrali dei principali notch di una determinata HRTF nel piano mediano possono essere estratte con ragionevole precisione a partire da alcune caratteristiche peculiari della forma del padiglione auricolare (pinna); viene calcolata la distanza tra un punto all'ingresso del canale uditivo (il punto di *focus*) e ogni punto che si trova sul bordo dell'elice, responsabile della principale riflessione dell'orecchio esterno, Fig. 1(a). Ad elevazione ϕ , la distanza tra il punto di riflessione e il *focus* è di $d(\phi) = ct(\phi)$ dove $t(\phi)$ è il ritardo temporale tra raggi diretti e riflessi, e c è la velocità del suono. Supponendo che ogni coefficiente di riflessione sia negativo e responsabile per un solo notch, la frequenza centrale corrispondente, $f_0(\phi)$, è

$$f_0(\phi) = \frac{c}{2d_c(\phi)}. \quad (1)$$

¹ Un dispositivo aptico con feedback di forza controllato tramite uno stilo: <http://www.sensable.com>

² <http://kemar.us/>

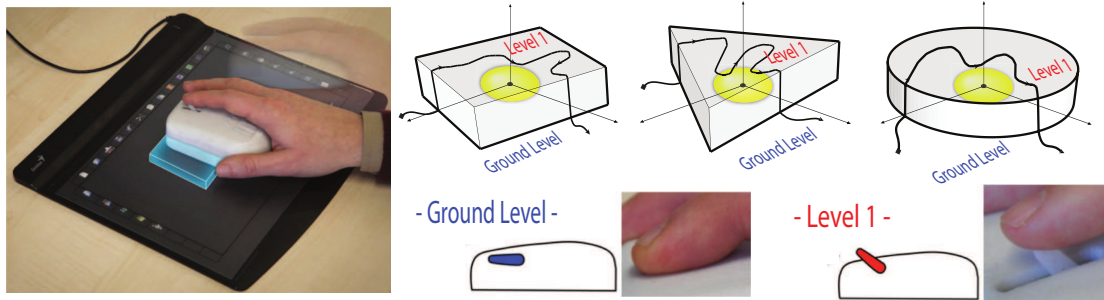


Figura 2. Il dispositivo TAMO in un singolo frame di esplorazione di un oggetto virtuale, raffigurato in azzurro (a sinistra). Le traiettorie di esplorazione (linee nere continue) di tre oggetti virtuali con un'ancora sonora abbozzata in giallo (in alto). Due stati della leva del TAMO (in basso): quando il puntatore è sopra l'oggetto virtuale (Level 1) la leva sale; quando il puntatore non tocca l'oggetto virtuale (Ground Level), la leva si abbassa.

Dato un soggetto per cui le HRTF individuali non siano disponibili, la precedente relazione permette di selezionare da un database una HRTF generica che abbia il minimo *mismatch* m tra le frequenze estratte dai contorni personali dell'orecchio esterno, f_0 , e le frequenze centrali dei notch, F_0 , del piano mediano per la HRTF disponibile:

$$m = \frac{1}{|\phi|} \sum_{\phi} \frac{|f_0(\phi) - F_0(\phi)|}{F_0(\phi)}, \quad (2)$$

dove l'elevazione ϕ spazia tutti gli angoli frontali disponibili tra -45° e 45° , guidando il confronto tra le frequenze centrali dei notch delle HRTF prese in considerazione con i corrispondenti punti di riflessione ai bordi dell'elice.

3.2 Mouse tattile

Abbiamo scelto un feedback tattile minimalista, il dispositivo TActile Mouser (*TAMO*) [19] che fornisce uno stimolo tattile minimale con una metafora di interazione tipo mouse (v. Fig. 2). Una leva mobile posta sopra il dispositivo al posto della comune rotellina crea un contatto tattile con il polpastrello del soggetto: quando il puntatore raggiunge un oggetto virtuale sulla mappa, la leva segnala un bordo virtuale di una certa altezza. Il feedback tattile corrispondente allo spostamento minimo in altezza viene chiamato *taxel*: le informazioni relative ad un *taxel* vengono memorizzate e rese in analogia con il pixel in *computer graphics* [20]. Il TAMO genera un singolo *taxel* per ogni corrispondente pixel dell'area di lavoro (210×297 -mm), come in una rappresentazione a bassorilievo. La leva si sposta da una posizione orizzontale corrispondente ad altezza terra, $\varphi_0 \approx 0^\circ$, ad una quasi verticale, corrispondente ad una rotazione di $\varphi_{max} \approx 80^\circ$.

L'utilizzo di tale dispositivo è motivato dalle seguenti considerazioni:

- la combinazione di tatto e di esplorazione attiva è adeguata per il riconoscimento e la ricostruzione di geometrie semplici;
- la risposta minimalista del dispositivo garantisce una curva di apprendimento rapido con importanti implicazioni pratiche per gli utenti finali;
- la maggior parte dei dispositivi tattili, es. il Phantom, sono difficilmente accessibili in termini di costi e per questo motivo non è mai stato riscontrato un loro uso diffuso; contrariamente il TAMO fornisce una soluzione a basso costo.

4. L'ESPERIMENTO

Alcune indagini recenti [21, 19] hanno studiato l'influenza nella qualità della costruzione di mappe tattili della combinazione di fattori quali carico cognitivo, tasso di acquisizione delle informazioni e strategia casuale/identificabile. Seguendo un approccio simile, l'esperimento qui proposto vuole indagare le strategie di esplorazione in un'attività di riconoscimento di oggetti utilizzando degli stimoli audio-tattili. Ai soggetti coinvolti è stato chiesto di riconoscere il più rapidamente possibile semplici oggetti virtuali con forme geometriche di base. Tali oggetti sono stati collocati al centro della zona di lavoro virtuale, ossia lo spazio limitato dai bordi di un tablet (Fig. 2). A seconda delle condizioni di feedback, un suono spazializzato di ancoraggio viene sintetizzato al centro del tablet come indicatore di orientamento globale.

Viene adottata una visuale egocentrica della mappa virtuale in cui il puntatore corrisponde alla testa dell'ascoltatore; spostando il TAMO dal centro verso il lato destro della tavoletta, il suono di ancoraggio viene spazializzato verso sinistra, come se la testa dell'ascoltatore fosse posizionata sul puntatore; tale metafora, chiamata *ears in hand*, è stata proposta da Magnusson et al. [22].

4.1 Partecipanti e apparato

Hanno preso parte all'esperimento nove soggetti (6 maschi e 3 femmine) di età compresa tra i 21 a 40 anni (media 29, DS 5, 7). Tutti normo-udenti e con diversi livelli di esperienza relativi alla partecipazione ad esperimenti psicofisici.

L'esperimento è stato eseguito in una cabina silente. Le cuffie Sennheiser HDA 200³ sono state collegate a una scheda audio esterna Roland Edirol AudioCapture UA-101 ad una frequenza di campionamento di 44,1 kHz. La Figura 1(a) propone una rappresentazione schematica del nuovo prototipo di interfaccia per il rendering aptico che utilizza il TAMO.

4.2 Stimoli multimodali

Una sorgente sonora virtuale viene posta al centro della mappa (Fig. 2.). Il suono spazializzato viene riprodotto in

³ Queste cuffie dinamiche circumaurali chiuse offrono un'attenuazione passiva che è efficace contro il rumore ambientale, permettendo di mascherare i suoni meccanici del TAMO.

cuffia in accordo con la posizione relativa tra il puntatore TAMO e l'ancora sonora. Lo stimolo uditivo presentato è un treno continuo di impulsi di rumore gaussiano della durata di 40-ms intervallato da un silenzio di durata pari a 30 ms [23]. L'ampiezza massima all'ingresso del canale uditivo viene fissata a 60 dB(A) dando la possibilità ai soggetti di regolarla a piacimento per ottenere un livello confortevole di ascolto.

Viene acquisita l'immagine dell'orecchio esterno di ogni soggetto, ed avviata la procedura di selezione dell'HRTF descritta nella Sez. 3.1. Il rendering della distanza segue la legge dell'inverso del quadrato sull'attenuazione del livello sonoro ed ha il suo massimo quando il puntatore ricopre la posizione della sorgente sonora, e il suo minimo sulla posizione più lontana che sia raggiungibile nell'area di lavoro del tablet. Un cerchio di raggio 25 px (≈ 12 mm) attorno al suono di ancoraggio definisce l'area in cui il feedback uditivo rimane costante.

Gli stimoli uditivi vengono filtrati attraverso il set di HRTF selezionato, e successivamente con un filtro di compensazione per le cuffie ottenuto con l'algoritmo presentato da [24]. Tale compensazione, misurata su un manichino KEMAR senza orecchie, garantisce un'effettiva equalizzazione fino a 8 – 10 kHz, e replica uno scenario applicativo reale dove non è possibile ottenere dei filtri di compensazione personali.

L'ambiente multimodale da noi proposto è in grado di simulare apticamente la presenza di oggetti virtuali sulla tavoletta come descritto nella Sez. 3.2. Tutte le mappe sono circondate da pareti virtuali sintetizzate con φ_{max} e quando il dispositivo viene spostato al di fuori della zona di lavoro, la leva si muove in maniera alternata da φ_{max} a $\varphi_{max} - 26^\circ$ a frequenza di *refresh* per segnalare il superamento dei confini delimitati dalle pareti virtuali.

4.2.1 Integrazione multimodale

Il sistema proposto sfrutta diversi componenti software / hardware (ad esempio porte COM, adattatori X-Bee®, mouse, interpolazione di HRTF, ecc.). Per questo motivo è fondamentale verificare che i vincoli di tempo reale siano soddisfatti, dando particolare attenzione alla sincronizzazione dello stimolo multimodale all'interno di una finestra temporale di integrazione capace di mantenerne la coerenza percettiva. Dato che i segnali provenienti da diverse modalità sensoriali posseggono diversi tempi di arrivo ed elaborazione all'interno del cervello, una finestra temporale di circa 200 ms garantisce l'integrazione multisensoriale [25].

Per scegliere la frequenza di aggiornamento del rendering multimodale è stata misurata la latenza degli stimoli audio e tattili posizionando due microfoni a condensatore collegati ad un Tascam 680 con frequenza di campionamento di 192 kHz: uno vicino alle cuffie e l'altro in prossimità della leva del TAMO. L'intero sistema introduce un ritardo medio di 68 ms (< 200 ms), guidando la nostra scelta del periodo di *refresh* del rendering multimodale di 80 ms.

4.3 Procedura

Una breve sessione di esercitazione agevola un'introduzione all'esperimento. I soggetti vengono verbalmente informati del dover identificare da bendati un solido geometrico sconosciuto in una mappa virtuale attraverso l'uso di udito (cuffie) e tatto (mouse tattile), utilizzando una precisa metafora di esplorazione; vengono inoltre istruiti nel tenere la testa con la stessa posa per ottenere coerente il rendering spaziale del suono. Lo sperimentatore guida i soggetti bendati verso la cabina silente e successivamente li conduce ad ogni *trial* verso una posizione di partenza rappresentata dal dispositivo TAMO al centro del bordo inferiore dell'area di lavoro.

Ogni *trial* si conclude quando i soggetti hanno identificato verbalmente l'oggetto o dopo un tempo massimo (impostato a 150 s); non vi è limite alle ipotesi che un soggetto può effettuare all'interno del tempo a disposizione, ma non viene fornita alcuna indicazione in merito alle ipotesi non corrette. Se i soggetti non sono in grado di identificare l'oggetto, il *trial* si conclude con esito negativo.

Gli oggetti virtuali di base, sconosciuti ai soggetti, sono: un prisma regolare a base triangolare, un prisma regolare a base quadrata, e un cilindro (vedi Fig. 2). Le dimensioni degli oggetti vengono fissate con riferimento a uno studio precedente [21] ed hanno tutti uguali altezza virtuale ($\varphi_0 + 18^\circ$), ne risultano così movimenti del TAMO che spaziano da ≈ 100 mm a 150 mm sopra agli oggetti virtuali. Con questi vincoli, gli oggetti posseggono un'area maggiore del 50% dello spazio di lavoro disponibile. Ogni oggetto viene così presentato in due condizioni: (i) TAMO, condizione aptica unimodale, e (ii) TAMO + audio 2D, condizione bimodale. Ne risulta un totale di sei stimoli. Le sequenze di presentazione vengono ordinatamente disposte a quadrato latino rispetto alla forma dell'oggetto. Le condizioni di feedback sensoriale, unimodale e bimodale, vengono presentate in maniera alternata per minimizzare gli effetti di apprendimento.

5. RISULTATI E DISCUSSIONE

Un'analisi preliminare ha rivelato un tempo medio di riconoscimento degli oggetti pari a 72,25 s nella condizione TAMO e di 62,87 s nella condizione bimodale, con una deviazione standard di 29,8 s e 35,9 s, rispettivamente. Inoltre, i fallimenti nel riconoscimento sono stati sporadici (10 *trial*).

In generale, i prismi regolari sono stati gli oggetti più facili da riconoscere, mentre il cilindro è stato quello più impegnativo (Fig. 3.(A)). Il suono spazializzato di ancoraggio ha avuto inoltre un effetto limitato sul tempo medio di riconoscimento, tranne nel caso di esplorazione di un prisma regolare a base quadrata dando origine ad un aumento del 30% sulle prestazioni tra TAMO e TAMO + Audio 2D con deviazioni standard di 46,62 s e 23 s, rispettivamente.

In aggiunta all'analisi quantitativa dei tempi di riconoscimento, un'analisi qualitativa delle strategie di esplorazione dei soggetti può essere in grado di rivelare le differenze tra le condizioni unimodali e bimodali. La discussione che segue è incentrata sull'esplorazione di prismi re-

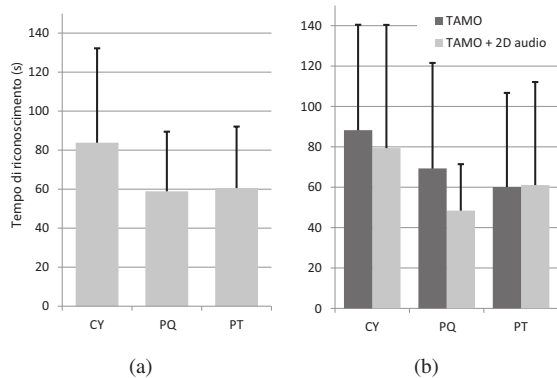


Figura 3. Media e deviazione standard del tempo di riconoscimento per 9 soggetti (a) per ogni oggetto virtuale e (b) raffinato sulle diverse condizioni di stimolo. Gli oggetti virtuali sono identificati come: CY - Cilindro, PQ - prisma regolare a base quadrata, e PT - prisma regolare a base triangolare.

golari con base quadrata per i quali sono state riscontrate differenze marcate. Sono state identificate le stesse due strategie principali riportate in [21]:

1. *Grid-scan*: esplorazione della mappa su una griglia con movimenti lungo linee verticali e orizzontali (Fig. 4(c));
2. *Z-scan*: inseguimento dei profili dell'oggetto tentando di attraversare i bordi con movimenti ortogonali o diagonali (Fig. 4(e)).

Possiamo sorprendentemente notare che emerge una terza strategia in presenza dell'ancora acustica: i soggetti seguono i bordi dell'oggetto controllando la distanza dal centro dell'oggetto stesso, corrispondente alla posizione del suono di ancoraggio. Questo comportamento è illustrato nella Figura 4 (seconda colonna) per tre soggetti esemplificativi. Questa nuova strategia promuove la riduzione della quantità di tempo trascorso al di fuori dell'oggetto, guidando i soggetti attraverso un'esplorazione ottimizzata. Un'analisi dei micro-movimenti (piccoli movimenti precisi) e dei macro-movimenti (rapidi cambiamenti macroscopici sulla mappa) potrebbe fornire, in una successiva valutazione sperimentale, un'ulteriore comprensione circa l'uso dell'ancoraggio sonoro in task di esplorazione che combinano attività in campo lontano e vicino. Si può ipotizzare che l'esplorazione di ambienti di grandi dimensioni possa essere supportato in modo più efficace utilizzando un suono di faro fisso. Nel caso in cui l'area di lavoro abbia dimensioni paragonabile a quelle del corpo umano, gli utenti possono modificare la posizione dell'ancora sonora utilizzando i riferimenti spaziali uditivi che man mano vanno a crearsi per confermare le diverse viste dell'oggetto.

6. CONCLUSIONI

Il sistema multimodale proposto in questo articolo fornisce un banco di prova per l'utilizzo di audio spazializzato al servizio della navigazione in spazi virtuali. Un task di

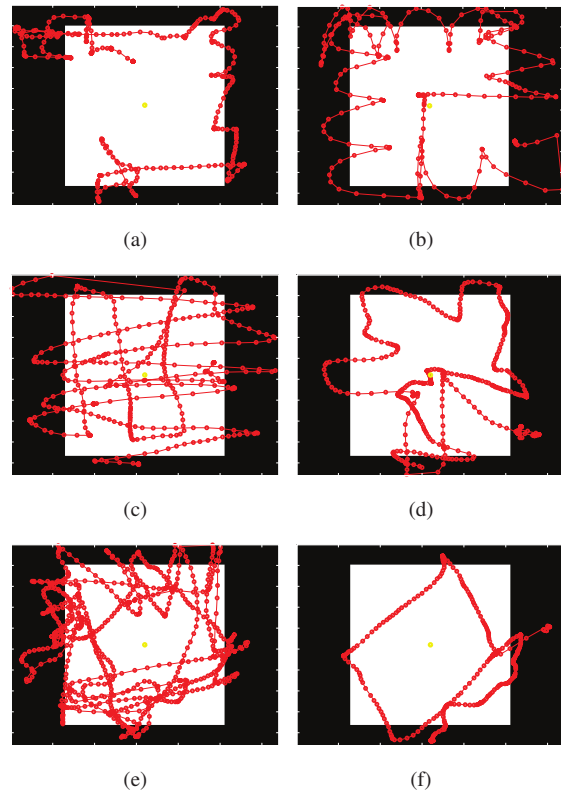


Figura 4. Esplorazione di un prisma regolare a base quadrata da parte di tre soggetti (righe), in condizione unimodale e bimodale (colonne). Le traiettorie sono mostrate in rosso.

riconoscimento di oggetti con soggetti bendati ha fornito dei primi incoraggianti risultati in merito a come i suoni di ancoraggio possano migliorare le prestazioni e modificare le strategie di esplorazione.

Una futura analisi dovrà correlare le caratteristiche dei movimenti esplorativi ai cambiamenti di feedback e strategie, nonché rivelare alcune criticità riguardanti la rappresentazione delle HRTF, la difficoltà di riconoscimento di oggetti complessi, l'efficacia di diversi posizionamenti per i punti di ancoraggio sonoro. Le modifiche temporali e spettrali dell'ancora sonora influenzeranno il carico cognitivo, e di conseguenza le prestazioni nell'attività di navigazione.

7. RINGRAZIAMENTI

Questo lavoro è stato finanziato dal progetto di ricerca PADVA (Personal Auditory Displays for Virtual Acoustics), n. CPDA135702 dell'Università di Padova.

8. BIBLIOGRAFIA

- [1] E. J. Gibson and A. D. Pick, *An Ecological Approach to Perceptual Learning and Development*. Oxford University Press, Aug. 2000.
- [2] M. O. Ernst and H. H. Bühlhoff, "Merging the senses into a robust percept," *Trends in Cognitive Sciences*, vol. 8, pp. 162–169, Apr. 2004.

- [3] C. Spence, "Crossmodal correspondences: A tutorial review," *Atten. Percept. Psychophys.*, vol. 73, pp. 971–995, Jan. 2011.
- [4] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer Publishing Company, Incorporated, 1st ed., 2007.
- [5] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PLoS ONE*, vol. 8, p. e82491, Dec. 2013.
- [6] Z. Cattaneo, T. Vecchi, C. Cornoldi, I. Mammarella, D. Bonino, E. Ricciardi, and P. Pietrini, "Image-ery and spatial processes in blindness and visual impairment," *Neuroscience & Biobehavioral Reviews*, vol. 32, pp. 1346–1360, Oct. 2008.
- [7] A. Afonso, A. Blum, B. Katz, P. Tarroux, G. Borst, and M. Denis, "Structural properties of spatial representations in blind people: Scanning images constructed from haptic exploration or from locomotion in a 3-d audio virtual environment," *Memory & Cognition*, vol. 38, no. 5, pp. 591–604, 2010.
- [8] S. J. Lederman and R. L. Klatzky, "Hand movements: A window into haptic object recognition," *Cognitive Psychology*, vol. 19, no. 3, pp. 342–368, 1987.
- [9] L. E. Magee and J. M. Kennedy, "Exploring pictures tactually," *Nature*, vol. 283, pp. 287–288, Jan. 1980.
- [10] W. Yu and S. Brewster, "Comparing two haptic interfaces for multimodal graph rendering," in *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, p. 3–9, 2002.
- [11] W. Schiff and E. Foulke, *Tactual Perception: A Sourcebook*. Cambridge University Press, Mar. 1982.
- [12] R. Farcy, R. Leroux, A. Jucha, R. Damaschini, C. Grégoire, and A. Zogaghi, "Electronic travel aids and electronic orientation aids for blind people: technical, rehabilitation and everyday life points of view," in *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion*, p. 12, 2006.
- [13] A. Lecuyer, P. Mobuchon, C. Megard, J. Perret, C. Andriot, and J.-P. Colinot, "HOMERE: a multimodal system for visually impaired people to explore virtual environments," in *Virtual Reality, 2003. Proceedings. IEEE*, pp. 251–258, Mar. 2003.
- [14] O. Lahav and D. Mioduser, "Exploration of unknown spaces by people who are blind using a multi-sensory virtual environment," *Journal of Special Education Technology*, vol. 19, no. 3, p. 15–23, 2004.
- [15] D. Schloerb, O. Lahav, J. Desloge, and M. Srinivasan, "BlindAid: virtual environment system for self-reliant trip planning and orientation and mobility training," in *Haptics Symposium, 2010 IEEE*, pp. 363–370, Mar. 2010.
- [16] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, p. 476–488, 1998.
- [17] 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, p. 508–519, Mar. 2013.
- [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 *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, (Florence, Italy), pp. 4496–4500, May 2014.
- [19] L. Brayda, C. Campus, and M. Gori, "Predicting successful tactile mapping of virtual objects," *IEEE Transactions on Haptics*, vol. 6, pp. 473–483, Oct. 2013.
- [20] M. Siegel, "Tactile display development: the driving-force for tactile sensor development," in *Haptic Virtual Environments and Their Applications, IEEE International Workshop 2002 HAVE*, pp. 115–118, 2002.
- [21] L. Brayda, C. Campus, R. Chellali, G. Rodriguez, and C. Martinoli, "An investigation of search behaviour in a tactile exploration task for sighted and non-sighted adults," in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, CHI EA '11*, (Vancouver, BC, Canada), p. 2317–2322, ACM, 2011.
- [22] C. Magnusson, H. Danielsson, and K. Rasmus-Gröhn, "Non visual haptic audio tools for virtual environments," in *Haptic and Audio Interaction Design* (D. McGookin and S. Brewster, eds.), no. 4129 in Lecture Notes in Computer Science, pp. 111–120, Springer Berlin Heidelberg, Jan. 2006.
- [23] B. F. G. Katz, S. Kammoun, G. Parsehian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, and C. Jouffrais, "NAVIG: augmented reality guidance system for the visually impaired: Combining object localization, GNSS, and spatial audio," *Virtual Reality*, vol. 16, pp. 253–269, June 2012.
- [24] A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *J. Audio Eng. Soc.*, vol. 60, no. 1/2, p. 54–62, 2012.
- [25] N. P. Holmes and C. Spence, "Multisensory integration: Space, time, & superadditivity," *Curr Biol*, vol. 15, pp. R762–R764, Sept. 2005.