

SONIFICAZIONE DI OSTACOLI COME AUSILIO ALLA DEAMBULAZIONE DI NON VEDENTI

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

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

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

SOMMARIO

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

1. INTRODUZIONE

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

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

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

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

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

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

2. SONIFICAZIONE TRAMITE SUONI D'IMPATTO

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

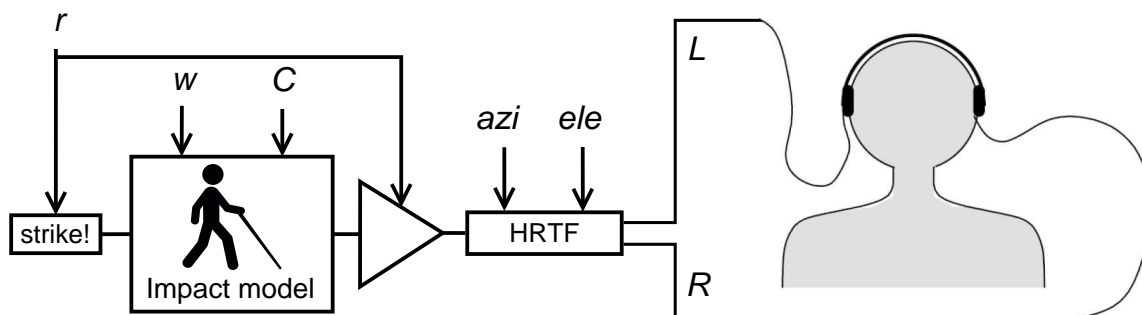


Figura 1. Lo schema di sonificazione proposto. I parametri in ingresso sono azimuth (azi), elevazione (ele), distanza (r), larghezza (w) e categoria (C) dell'oggetto.

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

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

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

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

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

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

t_d , è

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

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

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

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

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

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

¹ <http://soundobject.org/SDT/>

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

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

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

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

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

3. METODOLOGIA SPERIMENTALE

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

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

3.1 Partecipanti

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

3.2 L'approccio psicofisiologico

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

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

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

³ <http://www.ledalab.de>

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

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

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

3.3 Apparato sperimentale

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

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

3.4 Stimoli

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

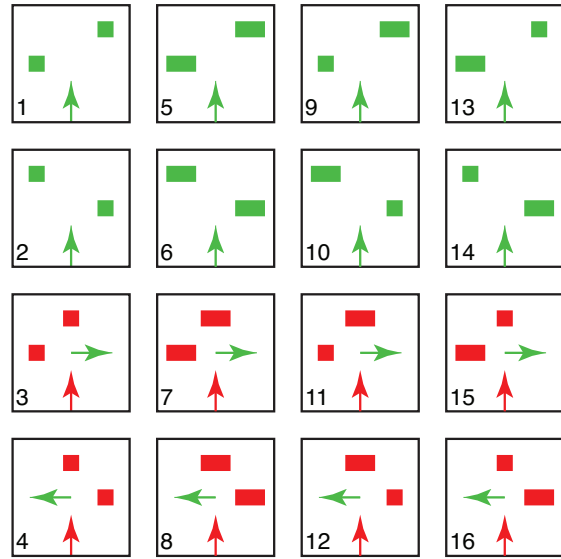


Figura 2. Disposizione degli oggetti nelle scene virtuali usate per l'esperimento. Le frecce verdi in ogni scena indicano la direzione relativa all'uscita corretta, mentre le frecce rosse indicano le situazioni in cui il partecipante deve girare a destra o sinistra.

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

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

3.5 Protocollo sperimentale

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

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

⁵ <https://earhero.com>

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

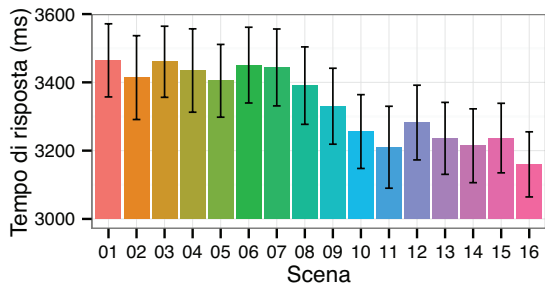


Figura 3. Tempo medio di risposta per scena. Le barre di errore rappresentano l'errore standard della media.

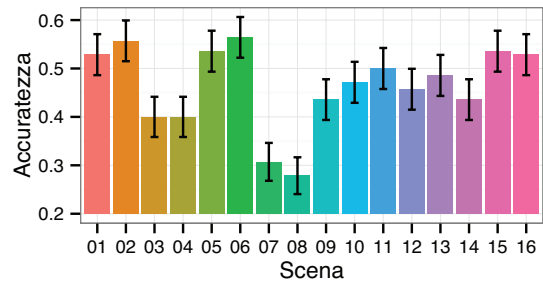


Figura 5. Accuratezza media per scena. Le barre di errore rappresentano l'errore standard della media.

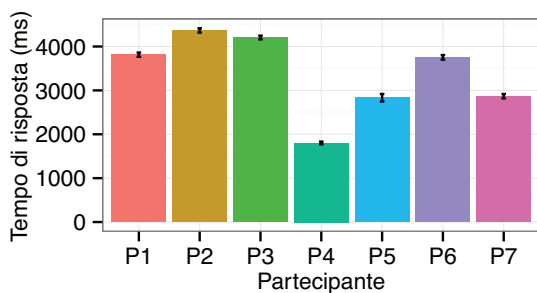


Figura 4. Tempo medio di risposta per partecipante. Le barre di errore rappresentano l'errore standard della media.

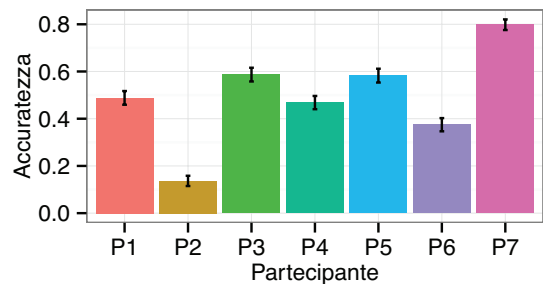


Figura 6. Accuratezza media per partecipante. Le barre di errore rappresentano l'errore standard della media.

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

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

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

4. RISULTATI

4.1 Tempo di risposta e accuratezza

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

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

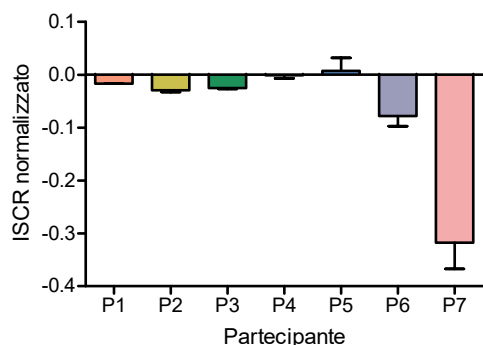


Figura 7. ISCR normalizzato medio per partecipante. Le barre di errore rappresentano l'errore standard della media.

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

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

4.2 Risposta elettrodermica

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

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

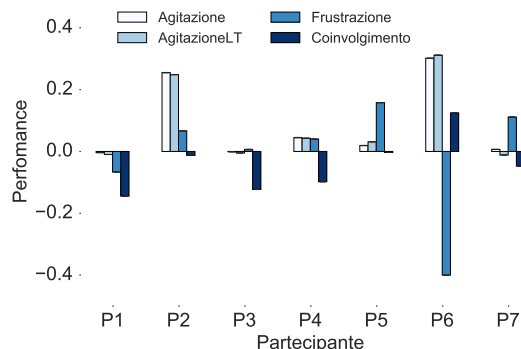


Figura 8. Metriche di performance PMDS medie per partecipante.

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

4.3 Metriche PMDS

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

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

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

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

5. CONCLUSIONI E SVILUPPI

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

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

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

6. RINGRAZIAMENTI

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

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

7. BIBLIOGRAFIA

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

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