SELFIE ACUSTICHE CON IL PROGETTO SELFEAR: UN'APPLICAZIONE MOBILE PER L'ACQUISIZIONE A BASSO COSTO DI PINNA-RELATED TRANSFER FUNCTION

Michele Geronazzo
Università degli Studi di Verona
Dip. di Neuroscienze, Biomedicina e Movimento
michele.geronazzo@univr.it

Jacopo Fantin, Giacomo Sorato, Guido Baldovino, Federico Avanzini
Università degli Studi di Padova
Dip. di Ingegneria dell'Informazione
corrispondenza a avanzini@dei.unipd.it

ABSTRACT

Le esperienze di realtà virtuale e aumentata stanno riscontrando una grande diffusione e le tecnologie per la spazializzazione del suono in cuffia saranno fondamentali per la diffusione di scenari applicativi immersivi in supporti mobile. Questo articolo affronta le problematiche legate alla acquisizione di *head-related tranfer function* (HRTF) con dispositivi a basso costo, accessibili a chiunque, in qualsiasi luogo e che forniscano delle misurazioni fruibili in tempi brevi. In particolare la soluzione proposta denominata "'the SelfEar project"' si focalizza sull'acquisizione delle trasformazioni spettrali ad opera dell'orecchio esterno contenute nella *pinna-related transfer function* (PRTF); l'utente viene guidato nella misurazione di HRTF in ambiente non anecoico attraverso una procedura auto-regolabile. Le informazioni acustiche sono infatti acquisite tramite un headset per la realtà acustica aumentata che include un set di microfoni posizionati in prossimitá dei canali uditivi dell'ascoltatore. Proponiamo una sessione di misurazione con l'obiettivo di acquisire le caratteristiche spettrali della PRTF di un manichino KEMAR, confrontandoli con i risultati che si otterrebbero con una procedura in ambiente anecoico. In entrambi i casi i risultati dipendono fortemente dalla posizione dei microfoni, senza considerare in questo scenario il problema legato ai movimenti di un eventuale soggetto umano. Considerando la qualità generale e la variabilità dei risultati, così come le risorse totali necessarie, il progetto SelfEar propone una promettente soluzione per una procedura a basso costo di acquisizione di PRTF, e piú in generale di HRTF.

1. INTRODUZIONE

L'obiettivo delle tecnologie binaurali è riprodurre suoni che risultino i più realistici e naturali possibili, illudendo l'ascoltatore di essere circondato da sorgenti sonore reali a lui esterne. Tale tecnologia risale alla fine del 20° secolo [1]; consiste nella registrazione di suoni attraverso una testa artificiale che simuli le caratteristiche di quella dell'ascoltatore, incorporando due capsule microfoniche nei condotti uditivi, modellando le membrane del timpano [2]. Questa tecnologia ottiene la massima efficienza attraverso la riproduzione in cuffia, che mantiene intatte le caratteristiche del segnale, senza riflessioni e riverberi dell'ambiente.

La resa di ambienti acustici virtuali coinvolge le cosiddette *binaural room impulse responses* (BRIR) nelle quali è possibile riconoscere due principali componenti: la prima è connessa con le caratteristiche ambientali, descritte nella *room impulse response* (RIR), l'altra è correlata agli aspetti antropometrici dell'ascoltatore, contenuti nella *head-related impulse response* (HRIR) [2]. Tutte queste risposte all'impulso hanno la loro controparte nel dominio della frequenza, ovvero nella loro trasformata di Fourier: *binaural room transfer funcrtion* (BRTF), *room transfer function* (RTF), e *head-related transfer function* (HRTF) rispettivamente. In particolare, le HRTF descrivono un filtro lineare tempo-invariante che definisce il filtraggio acustico al quale contribuiscono testa, torso e orecchio esterno dell'ascoltatore.

Il processo di acquisizione di HRTF individuali è attualmente molto costoso in termine di attrezzature e tempo. La misurazione acustica in camera anecoica offre una risposta all'impulso di controllo che è di alta qualità e precisione; richiede però un attrezzatura molto costosa e difficilmente reperibile per applicazioni reali. È quindi inevitabile sacrificare alcune caratteristiche individuali del soggetto per ottenere una rappresentazione della HRTF più economica ma che dia comunque un'informazione psico-acustica accurata [3]. Il processo di acquisizione di HRTF in ambiente domestico è una questione complessa; le tendenze più recenti sono supportate da dispositivi a basso costo per l'acquisizione di mesh 3D [4]) e da algoritmi per la modellazione e personalizzazione delle HRTF [5]. Purtroppo queste soluzioni mancano di informazioni robuste ed individuali per l'acustica dell'orecchio esterno (*pinna*) a causa di una dettagliata struttura antropometrica. Queste informazioni sono contenute nella cosiddetta *pinna-related transfer function* (PRTF) [6] che risulta molto difficile da modellare computazionalmente anche nel contesto di simulazioni numeriche [7]. Le PRTF contengono indizi salienti sulla localizzazione per quanto riguarda la percezione dell'elevazione (vedere [8] per una bibliografia più approfondita), perciò è necessaria una rappresentazione accurata per fornire la percezione della dimensione verticale nelle tecnologie di audio binaurale.

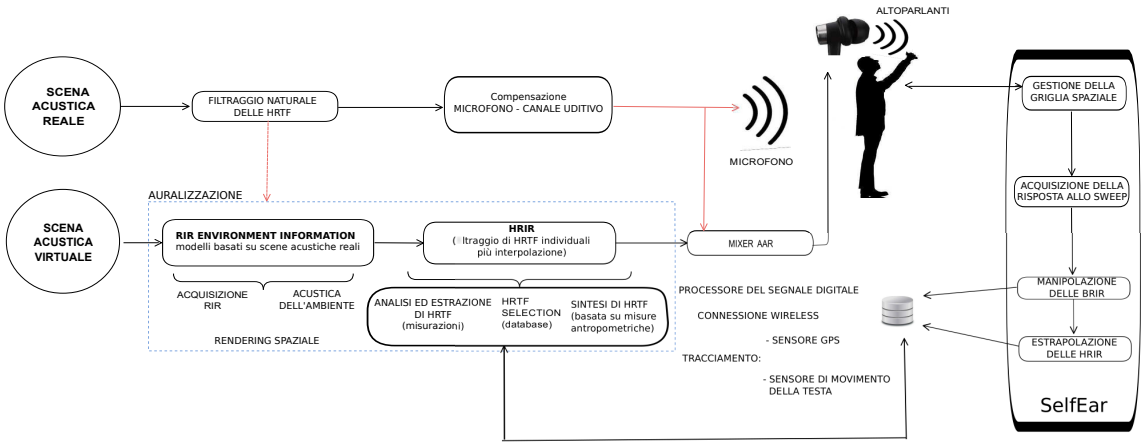


Figure 1: Vista schematica del progetto SelfEar in un contesto di realtá audio aumentata.

Questo articolo affronta il problema della riduzione dei costi nel processo di acquisizione/misurazione di HRTF, con particolare attenzione all'estrapolazione di PRTF per sistemi mobile di realtà audio aumentata (*mobile audio augmented reality* - mAAR). Questi sistemi sono composti di cuffie provviste di microfoni esterni integrati per l'acquisizione di flussi audio multi-canale dall'ambiente, oltre ad algoritmi per la riproduzione di audio binaurale. Una possibilitá per poter ottenere ovunque le proprie HRTF utilizza i microfoni integrati a partire da stimoli sonori riprodotti dagli altoparlanti del dispositivo mobile; il progetto SelfEar ha lo scopo di sviluppare e implementare gli algoritmi di elaborazione del segnale e l'interazione col dispositivo per ottenere una procedura auto-regolabile dall'utente.

Sono stati condotti solamente pochi studi che mirino alla consistenza delle HRTF acquisite in ambiente non anecoico e che considerino il contributo acustico del piano mediano-sagittale [9], che è di rilevante importanza per il contenuto spettrale individuale introdotto nelle PRTF. Il compromesso tra costi e portabilità conduce a due problematiche principali. La prima riguarda il fatto che il processo di acquisizione tramite dispositivo mobile implica l'influenza dell'ambiente circostante, con colorazioni in frequenza e interferenze da ritardi. In secondo luogo, impiegare gli speaker di uno smartphone o tablet come sorgente sonora e microfoni binaurali di mercato *consumer* per l'acquisizione comporta registrazioni meno accurate rispetto a quelle ottenute con attrezzatura professionale.

In questo articolo presentiamo una serie di misurazioni condotte in un ambiente controllato su un manichino KEMAR [10]. Il nostro obiettivo finale è confrontare risposte ottenute grazie a SelfEar con quelle ottenute con attrezzatura professionale; di seguito la struttura dell'articolo: Sez. 2 contiene la descrizione di un sistema di realtà audio aumentata mobile e i criteri per l'esternalizzazione del suono virtuale; nella Sez. 3 viene presentato il progetto SelfEar. La Sez. 4 descrive gli esperimenti acustici su di una testa artificiale in ambiente non anecoico. All'interno della Sez. 6 vengono discussi i risultati ottenuti e la Sez. 7 conclude la valutazione preliminare esposta con l'esposizione delle piú promettenti direzioni di ricerca per il futuro.

2. REALTÀ AUDIO AUMENTATA IN CONTESTO MOBILE

In un sistema di tipo mAAR, l'ascoltatore può usufruire di un mix di sorgenti sonore reali e virtuali. Le prime vengono riprese dai microfoni dell'headset previo filtraggio naturale del soggetto con conseguente applicazione di una compensazione in frequenza-fase che tenga conto dell'errore introdotto dalla differenza di posizione dei microfoni rispetto al punto d'ingresso del canale uditivo [11]. Le sorgenti virtuali necessitano invece di un processo di auralizzazione dinamico per ottenere una sovrapposizione ideale con la realtà. L'auralizzazione comprende un rendering tramite BRIR,risultante nel filtraggio tramite RIR e HRIR che devono essere personalizzate in accordo con l'ambiente circostante [12] e con l'ascoltatore [3]. Gli algoritmi di signal processing (DSP) implementano filtri correttivi che compensano i microfoni, gli speaker e le loro interazioni, prendendo in considerazione gli effetti psicoacustici e di colorazioni causati dall'indossare la cuffia rispetto alle normali condizioni di ascolto senza di essa [13].

Produrre scenari acustici virtuali e aumentati in cuffia con attenzione alle caratteristiche di ambiente e di ascoltatore rimane una delle sfide più importanti in questo dominio di ricerca, a causa delle forti interconnessioni presenti tra i vari componenti del sistema mAAR che concorrono ad una corretta "'esternalizzazione"' della scena acustica. [1] I criteri per una corretta auralizzazione ed esternalizzazione possono essere riassunti quindi in quattro categorie:

- *sistema ergonomico*: le cuffie ideali dovrebbero essere acusticamente trasparenti: l'ascoltatore non si dovrebbe rendere conto della loro presenza. A tale scopo sono essenziali cuffie poco invasive nel peso e nelle dimensioni;

- *tracciamento*: i movimenti della testa nell'ascolto binaurale producono segnali interaurali dinamici [14];

[1] Con esternalizzazione indichiamo la percezione di un suono to fuori dalla testa, piuttosto che al suo interno.

tracciare la posizione dell'utente nell'ambiente permette il riconoscimento dell'interazione acustica e una rappresentazione spaziale coerente tra scena reale e simulazione virtuale;

- *conoscenza dell'ambiente*: la percezione spaziale della scena acustica richiede la conoscenza di riflessioni e riverberazioni tipiche dell'ambiente stesso; queste informazioni risultano essenziali per una impressione realistica dello spazio in cuffia [15];

- *individualizzazione*: testa e orecchio del soggetto filtrano in maniera individuale i suoni; va presa in considerazione una correzione individuale per l'accoppiamento acustico tra cuffie e orecchio esterno [16].

3. IL PROGETTO SELFEAR

3.1 Panoramica del sistema

SelfEar è un'applicazione progettata per piattaforma Android con lo scopo di ottenere le HRIR personali dell'utente a partire da uno stimolo sonoro riprodotto dal dispositivo mobile utilizzato. Il telefono/tablet deve essere tenuto in mano con il braccio teso e spostato lungo il piano mediano del soggetto fermandosi a specifiche elevazioni. I microfoni in-ear catturano il segnale acustico proveniente dagli altoparlanti del dispositivo, registrando in questo modo le BRIR specifiche per posizione, ascoltatore e ambiente.

I dati collezionati tramite smartphone possono essere successivamente impiegati per ottenere una HRIR customizzata. Dopo le procedure di post-processing che compensano gli effetti acustici delle condizioni di acquisizione e riproduzione, le HRIR individuali possono direttamente essere usate per la resa spaziale del suono. A seconda della complessità degli scenari virtuali, la sintesi in tempo reale delle HRIR è oggigiorno possibile in ambiente mobile. Una possibile soluzione impiega la selezione di HRTF misurate in camera anecoica (ad esempio il database CIPIC [17]) [2] pre-esistenti in basi di dati pubbliche, basandosi sui parametri acustici estratti con SelfEar: la procedura seleziona la migliore approssimazione per l'HRTF dell'utente SelfEar basandosi su metriche psicoacustiche e di similaritá antropometriche [18].

3.2 Gestione della sorgente

Il sistema di gestione della griglia spaziale di SelfEar guida l'utente attraverso il processo di acquisizione delle BRIR definendo una procedura autoregolata descritta in Fig. 2. Nel paragrafo seguente descriviamo ogni passo, partendo dal lancio dell'applicazione fino al termine della sessione, che ha come risultato la restituzione di un insieme di BRIR individuali.

Nella schermata iniziale di SelfEar, l'utente seleziona la posizione degli speaker del dispositivo: questi possono trovarsi sul lato superiore, frontale, inferiore o sul retro. Questa scelta determinerà l'orientamento del dispositivo

[2] Una collezione di diverse misurazioni acustiche condotte su 50 soggetti distinti (più di 1200 misurazioni ciascuno), che comprendono anche informazioni antropometriche.

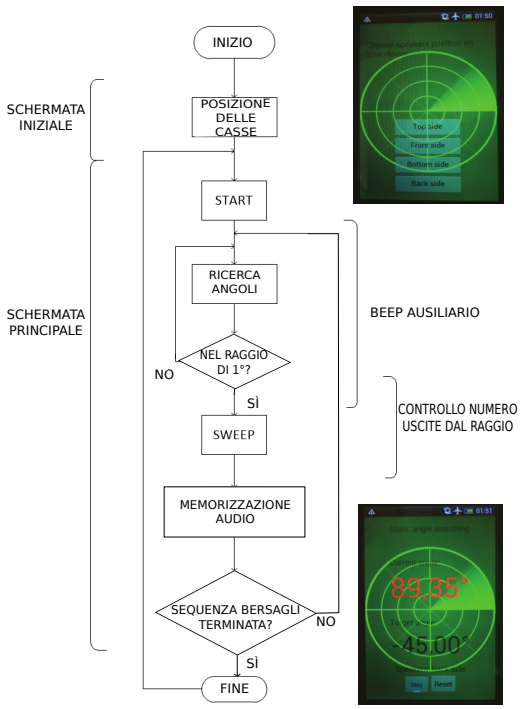


Figure 2: Diagramma a blocchi della procedura SelfEar per l'acquisizione di BRIR nel piano mediano. Vengono riportati anche degli screenshot delle due schermate dell'applicazione.

durante la riproduzione dello stimolo sonoro per massimizzare il rendimento degli altoparlanti del dispositivo a seconda della caratteristica di direzionalità. L'utente può in seguito premere il pulsante "'Start"' per iniziare la procedura di acquisizione delle BRIR. Prima di raggiungere la prima elevazione target, si istruisce l'utente nel portare il dispositivo (sostenuto dal braccio teso) a livello degli occhi, permettendo così la creazione di un riferimento propriocettivo/motorio per l'estrazione di tutti gli altri angoli; in seguito i passi seguono questa sequenza logica:

1. *Raggiungimento del bersaglio*: l'attuale elevazione del dispositivo rispetto all'orizzonte, ϕ_i, compare sopra all'elevazione bersaglio da raggiungere. SelfEar elabora i dati provenienti dall'accelerometro lungo i tre assi cartesiani, $a_{x,y,z}$, del dispositivo per calcolare ϕ_i tramite la formula seguente:

$$\phi_i = \arctan\left(\frac{\pm a_y}{|a_z|}\right)$$

nel caso gli altoparlanti siano posizionati sul lato superiore o inferiore; mentre con la formula:

$$\phi_i = \arctan\left(\frac{\pm a_z}{|a_y|}\right)$$

nel caso si trovassero invece sul lato frontale o sul retro. Il segno del numeratore è:

+ per altoparlanti sul lato inferiore o sul retro;
- per altoparlanti sul lato superiore o frontale.

La sequenza delle elevazioni bersaglio spazia in ordine crescente tra gli angoli del CIPIC database equispaziati ogni $5.625°$ in un intervallo di $[-40°, 40°]$. Un beep ausiliario rende acusticamente l'informazione di differenza tra l'attuale elevazione del dispositivo e quella bersaglio, agevolando la procedura di puntamento soprattutto nei casi in cui lo schermo del dispositivo non è visibile a causa della posizione degli speaker (ad es. quando si trovano sul retro). La pausa tra un beep ed il successivo è direttamente proporzionale alla differenza tra ϕ_i, e l'angolo bersaglio, $\widehat{\phi_i}$, come mostrato nell'equazione seguente:

$$pause_i = \left|\phi_i - \widehat{\phi_i}\right| \cdot k$$

dove *i* è un istante in cui un singolo beep termina la sua riproduzione e *k* è una costante che rende percepibile la pausa tra due beep consecutivi. [3] L'obiettivo di questo step è di avvicinarsi a $\widehat{\phi_i}$ con una precisione di $\pm 1°$. Questo passo può venire interrotto e ripreso su richiesta dell'utente.

2. *Controllo della posizione*: una volta che ϕ_i entra nel range di validá di $\pm 2°$ dal bersaglio, scatta un timer di stabilizzazione di 2 secondi; nel caso l'utente uscisse dal range per tre volte prima che il timer finisca, la procedura inizia nuovamente a partire dalla fine del passo 1.

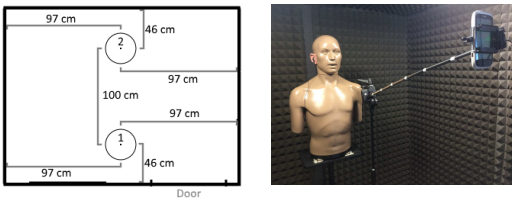
3. *Riproduzione dello sweep*: al termine del timer di stabilizzazione, viene riprodotto lo stimolo sonoro dagli altoparlanti del dispositivo; nel caso l'utente uscisse dal range di validá di $\pm 2°$ solamente una volta durante la riproduzione dello sweep, la procedura di ricerca di $\widehat{\phi_i}$ viene annulata.

4. *Salvataggio delle BRIR*: quando uno sweep termina con successo, il segnale audio registrato viene salvato localmente in memoria insieme all'angolo di elevazione a cui si riferisce; la procedura ritorna quindi al passo 1 con la successiva elevazione bersaglio in sequenza.

5. *Fine della sessione*: una sessione termina quando tutte le elevazioni bersaglio sono state raggiunte con successo.

4. MISURAZIONI ACUSTICHE

Le sessioni sperimentali sono state eseguite in un ambiente non anecoico usando un manichino KEMAR che ci ha permesso di minimizzare gli errori dovuti al movimento di un eventuale soggetto umano. Ci siamo quindi concentrati sulla direzione con il più alto contenuto di caratteristiche spettrali, ovvero la direzione frontale [19], $phi = 0$, fornendone un'analisi dettagliata. Infine abbiamo riportato una valutazione qualitativa dell'applicazione SelfEar per una serie di PRIR nel piano sagittale frontale.



(a) (b)

Figure 3: Setup per le misurazioni. (a) Posizioni di sorgente e ricevitore. (b) Setup di misurazione di Selfear, con selfie-stick incorporato.

4.1 Setup

Struttura e attrezzature - Tutte le misurazioni sono state condotte all'interno di una camera silente di 2×2 m, una Sound Station Pro 45 (SSP), con un isolamento acustico massimo di 45 dB.

La Figura 3a mostra l'interno della SSP dove sono identificabili due posizioni: la posizione #1 per le varie sorgenti acustiche e la posizione #2 per i vari ricevitori.

Negli esperimenti sono stati usati due tipi di dispositivi di riproduzione (con definizione acronimi):

- *L* : un loudspeaker Genelec 8030A calibrato in modo da avere un buon SNR con tono di test a 500 Hz con 94 dB SPL;
- *S* : uno smartphone HTC Desire C sorretto da una struttura autoprodotta che simula un braccio umano, realizzata con un selfie stick; [4] in questo caso la SPL massima raggiunta con il tono di test è di 51 dB alla frequenza di riferimento di 500 Hz.

E due tipi di ricevitori (con definizione acronimi):

- *H* : un paio di in-ear headphones Roland CS-10EM con due microfoni incorporati;
- *K* : due microfoni professionali G.R.A.S forniti dal simulatore di testa e torso KEMAR; nella configurazione proposta, l'orecchio destro era equipaggiato con il simulatore di canale uditivo, mentre l'orecchio sinistro ne era sprovvisto.

In tutti gli esperimenti, sorgenti e ricevitori, nelle rispettive posizioni, sono sempre stati posizionati alla stessa altezza dal pavimento. Il segnale sorgente utilizzato è un sine sweep logaritmico della durata di 1 -s, il cui contenuto in frequenza spazia uniformemente da 20 Hz a 20 kHz. I segnali acustici sono stati registrati con il software Audacity

[3] La formula restituisce un valore in millisecondi, che senza una costante moltiplicativa darebbe luogo ad una pausa troppo corta per poter essere percepita. In questa realizzazione è stato scelto un valore arbitrario di $k = 5$.

[4] La struttare a selfie stick, di lunghezza 1-m, è più lunga di un braccio medio in estensione; riteniamo questa discrepanza ininfluente nell'estrazione delle caratteristiche spettrali della PRTF per la percezione dell'elevazione in quanto queste si possono assumere invarianti con la variazione della distanza dalla sorgente [20].

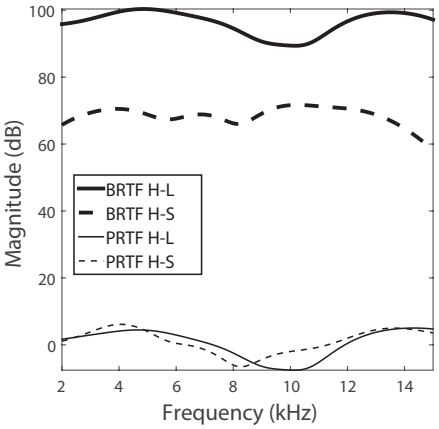


Figure 4: Confronto tra i moduli (in dB SPL) delle BRTF (linee spesse) e le relative PRTF (linee sottili) ottenute usando: come ricevitore - il microfono destro delle cuffie Roland (H); come sorgente - il loudspeaker dello smartphone (S, linee tratteggiate) e il loudspeaker Genelec (L, linee continue).

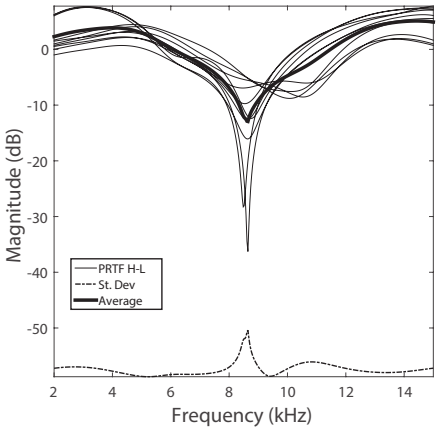


Figure 5: Moduli delle PRTF per i dieci riposizionamenti della cuffia Roland sul manichino KEMAR. La linea spessa rappresenta l'ampiezza media. La deviazione standard è spostata di -60 dB per agevolarne la lettura.

attraverso un'interfaccia audio Motu 896 mk 3 e sono stati poi processati usando il software Matlab aggiornato alla versione 8.4.

Calibrazione con diffuse-field - Per il calcolo della risposta *diffuse-field* è stata realizzata una struttura apposita composta da due cavi di filo metallico agganciati al soffitto della SSP distanziati di 17.4 cm, ovvero la distanza tra i microfoni della KEMAR. Questa misurazione ci permette di acquisire le specifiche acustiche dell'ambiente e del setup.

Abbiamo acquisito le misurazioni diffuse-field per tutte le combinazioni di sorgente e ricevitore, ottenendo un totale di quattro misurazioni.

4.2 Dati

Prima sessione di misurazione - In questa sessione sono stati posizionati all'interno della SSP, nelle rispettive posizioni (#1 e #2 della Fig. 3a), lo speaker Genelec e la KEMAR. Come primo step sono state misurate le risposte dell'orecchio destro e sinistro. Successivamente è stata inserita nel canale uditivo destro della KEMAR la cuffietta Roland; sono state effettuate dieci misurazioni con questa configurazione riposizionando ogni volta le Roland; in tal modo è stato possibile analizzare la variabilità introdotta nelle misurazioni dalla posizione del microfono.

Seconda sessione di misurazione - In questa sessione è stata posizionata in #1(vedi Fig. 3a)la struttura di sostegno per lo smartphone; nella posizione #2 di Fig. 3a è invece stata posizionata la KEMAR. Il selfie-stick manteneva lo smartphone ad una distanza di un metro dalla KEMAR e consentiva una regolazione angolare sul piano mediano. Sono state acquisite le misurazioni relative a 15 angoli tra -40° e $+40^\circ$ sul piano mediano riproponendo gli angoli presenti nel CIPIC database. Quindi, al termine della misurazione abbiamo ottenuto due set di 15 misurazioni per l'orecchio sinistro della KEMAR (senza cuffietta) e per il microfono destro della cuffietta.

5. RISULTATI

Per ogni misurazione è stato calcolato l'onset applicando la funziona di cross-correlazione con il segnale di sweep originale; è stata poi estratta la BRIR deconvolvendo la risposta allo sweep con lo sweep stesso. Sono state successivamente rimosse riflessioni dovute dalla SSP e dalla strumentazione al suo interno, tale procedura è stata effettuata sottraendo alle BRIR le rispettive risposte di tipo *diffuse-field*. Da questa procedura sono state estratte le HRTF. Mentre le PRTF sono state ottenute applicando ad ogni risposta impulsiva una finestra di hanning di 1 -ms (48 campioni) centrata temporalmente sul picco massimo e normalizzando rispetto al valore massimo in ampiezza [6]. Tutte le PRTF normalizzate sono state filtrate con un passa banda tra 2 kHz and 15 kHz, mantenendo i picchi ed i notch dovuti alla presenza della pinna.

La Figura 4 mostra il confronto tra i moduli in dB SPL delle BRIR estratte dalle misurazioni usando come sorgente (i) il loudspeaker Genelec (ii) lo smartphone, e come ricevitore il microfono destro delle cuffie Roland inserito nell'orecchio destro della KEMAR. Si puó notare una differenza di 30 dB tra la SPL dei due loudspeaker, tale differenza porta ad un SNR minore mentre si usa il loudspeaker dello smartphone. Nella stessa figura sono mostrate le due PRTF normalizzate corrispondenti in modo da poter valutare il contributo introdotto dal diffuse-field. Tale contributo è sensibilmente visibile per le misurazioni ottenute con lo smartphone HTC Desire C, dovuto principalmente alla diversità di direzionalità tra lo speaker a basso costo integrato e loudspeaker Genelec.

In Fig. 5, sono riportati i moduli in dB delle PRTF dei dieci riposizionamenti insieme alla loro media. Vi è anche riportata la deviazione standard in modo da poter analizzare la viaribilità introdotta nelle misurazioni dalla po-

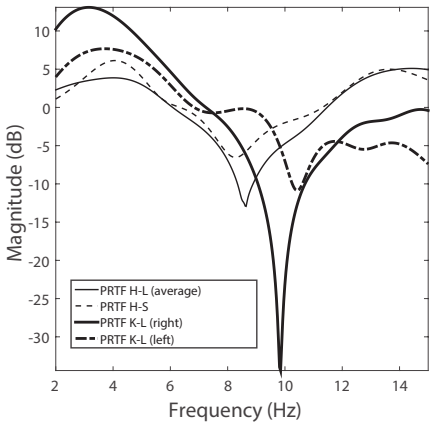


Figure 6: Confronto tra i moduli delle seguenti PRTF: PRTF media della Fig.5 (PRTF H-L); sorgente: smartphone - ricevitore: microfoni delle cuffie (PRTF H-S); sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio destro della KEMAR con canale uditivo (PRTF K-L right); sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio sinistro della KEMAR senza canale uditivo (PRTF K-L left).

sizione delle cuffie/microfono. La variabilità massima la si osserva in prossimità dei notch salienti della PRTF, tra 9 e 11 kHz , che sono molto sensibili ai cambiamenti topologici tra la cuffie e la struttura dell'orecchio [7].

Le principali valutazioni sono state effettuate nella posizione frontale, $\phi = 0$, comparando le PRTF normalizzate in condizioni differenti. La Fig. 6 mostra il confronto tra il modulo delle PRTF acquisite con e senza cuffiette, coinvolgendo sia la Genelec che lo smartphone.

Per le quattro PRTF è stata calcolata la distorsione spettrale media (SD) [8] tra tutte le coppie in un range di frequenze tra 2 kHz $\leq$ e $\leq$ 15 kHz (i valori sono visibili in Tabella 1). Questi confronti permetto diverse considerazioni:

- Acustica della pinna, K-L$_{right}$ vs. K-L$_{left}$: le forme diverse delle orecchie e l'acustica del canale uditivo del KEMAR differiscono notevolmente; questa differenza è visibile in tutti i confronti che coinvolgono K-L$_{left/right}$;
- Loudspeaker, H-S$_{right}$ vs. H-L$_{right}$: entrambi i loudspeaker introducono una distorsione spettrale trascurabile ($<$ 2 dB);
- Procedura SelfEar, H-S$_{right}$ vs. K-L$_{left}$: la differenza tra l'acquisizione SelfEar delle PRTF e un sistema di misura tradizionale introducono l'errore in SD più basso tra le coppie analizzate (rimuovendo il confronto di controllo sui loudspeaker); [5]

PRTF	H-L	H-S	K-L(right)	K-L(left)
H-L	0	1.79	6.92	5.25
H-S		0	7.35	4.64
K-L(right)			0	5.47
K-L(left)				0

Table 1: Distorsione spettrale tra PRTF di Fig. 6. Tutti i valori in dB.

La Figura 7 permette di confrontare visibilmente i risultati ottenuti con la procedura SelfEar negli angoli considerati (con e senza la compensazione di tipo *diffuse-field*) e le misurazioni CIPIC nello stesso range di angoli per il soggetto 165 (KEMAR). I dati sono stati interpolati per avere una transizione visivamente più fluida.

6. DISCUSSIONE GENERALE

È già noto da Christensen *et al.* [21] che sia la posizione del ricevitore, sia il suo spostamento dal punto di misura dell'HRTF ideale, cioè al centro dell'ingresso del canale uditivo, influenzano i pattern di direttività dell'HRTF per frequenze superiori a $3 - 4$ kHz. Il nostro lavoro è in accordo con le loro misurazioni in quanto mostrano uno spostamento della frequenza del notch centrale anche di 2 kHz con una elevata variabilità del modulo nei vari riposizionamenti del microfono (vedere la deviazione standard di Fig. 5) con una differenza massima di 10 dB.

Gli spostamenti delle frequenze centrali di picchi e notch sono osservabili anche dalla Fig. 6 e sono principalmente dovuti a differenze topologiche tra il punto di osservazione, dipendente dalla posizione del microfono, e gli oggetti acustica formati dalla presenza o assenza del canale uditivo e da orecchie destra e sinistra differenti.

Avere a disposizione un ampio intervallo di elevazioni frontali permette ad ogni sistema di misura di acquisire le caratteristiche rilevanti delle PRTF [19, 6]: ovvero le due principali risonanze (P1: modo omnidirezionale, P2: modo orizzontale) e i tre principali notch (N1-3 corrispondenti alle principali riflessioni della pinna). Le PRTF provenienti dal CIPIC KEMAR (vedere Fig. 7(c)) contengono tali caratteristiche; in particolare P1 ha una frequenza centrale di 4 kHz e P2 di 13 kHz; inoltre N1 si muove tra 6 e 9 kHz, N3 tra 11.5 e 14 kHz con un graduale incremento in elevazione; infine N2 che inizialmente si trova a 10 kHz, progressivamente scompare man mano che si raggiunge la direzione frontale.

L'applicazione SelfEar è in grado di acquisire P1 e N1 efficacemente, considerando sia il caso in cui le PRTF siano state compensate con la *diffuse field*, sia che non lo siano. Dal momento che l'ambiente ha un contributo non trascurabile, il confronto visivo tra Fig. 7(a) e (b) sottolinea l'importanza di poter estrarre con precisione le PRTF dalle BRIR. In particolare dalla Fig. 7(b) si possono identificare anche P2 e, anche se in minima parte, N2. Va quindi menzionato che P1 e N1-2 sono sufficienti per veicolare la medesima accuratezza in localizzazione verticale di una PRTF provvista di tutte le caratteristiche spettrali [22]. Tuttavia, dal punto di vista acustico, N3 è completamente assente, il che suggerisce la presenza di un interferenza introdotta dalle cuffie nella concha della pinna. Seguendo il

[5] Va notato che questo confronto è stato effettuato su orecchie diverse per ragioni pratiche. Di conseguenza i valori di SD potrebbero avere una differenza ancor piú inferiore.

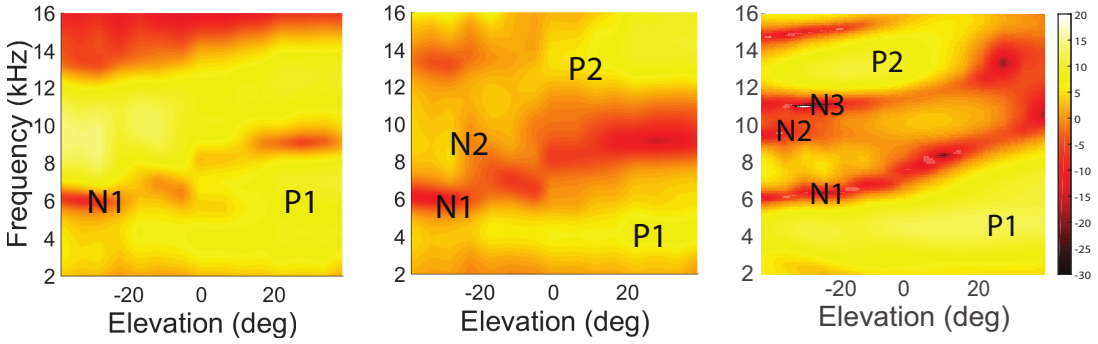


Figure 7: PRTF nel piano mediano. (a) acquisizione con SelfEar senza compensazione; (b) acquisizione con SelfEar compensata con diffuse-field; (c) CIPIC KEMAR, Subject 165 - con compensazione free-field. Vengono visivamente identificati i principali picchi (P1-2) e notch (N1-3), ove presenti.

modello risonanza-più-riflessione per le PRTF [6, 8] possiamo ipotizzare che l'assenza di riflessi nella concha sia dovuta alla presenza della cuffia; inoltre il volume della concha in questa condizione è drasticamente ridotto, producendo cosí variazioni nei modi di risonanza della struttura della pinna [7]. Inoltre, comparando i valori della distorsione spettrale tra $H - S$ vs. $K - L_{left}$ si ha una differenza di 4.64 dB che suggerisce una buona affidabilità paragonabile al metodo di personalizzazione in [8] (valori di SD tra 4 e 8 dB) e allo stato dell'arte delle simulazioni numeriche in [7] (valori di SD tra 2.5 e 5.5 dB).

Vale la pena notare che i parametri di notch e picchi, cioè la frequenza centrale, il guadagno e la banda, possono essere estratti dalle PRTF disponibili e dati in input a modelli di PRTF sintetiche e/o, seguendo un approccio di modellazione strutturale mista [3]], a procedura di selezione delle HRTF. Infine non vi è nulla che impedisca l'uso diretto di PRTF estratte da SelfEar per il rendering di audio binaurale.

7. CONCLUSIONI E LAVORI FUTURI

L'applicazione SelfEar permette l'acquisizione di HRIR a basso costo nel piano frontale mediano rilevando particolari informazioni spettrali della pinna dell'ascoltatore, ovvero le PRTF. L'applicazione sfrutta una struttura tecnologica di realtá audio aumentata per dispositivi mobile. Dopo essere adeguatamente compensate, le PRTF estratte sono paragonabili in termini di caratteristiche acustiche principali a quelle misurate in camera anecoica.

Il sistema proposto è stato testato seguendo un robusto setup di misurazione senza soggetti umani in una camera silente, ambiente acusticamente trattato. Perciò è tutt'ora da valutare una procedura robusta per ottenere le PRTF in ambiente domestico, stimando statisticamente l'influenza di eventi sonori rumorosi e casuali, nonché i movimenti del soggetto durante l'acquisizione. Per questo motivo la progettazione di algoritmi di elaborazione del segnale per la rilevazione di eventi acustici, l'eliminazione del rumore e il tracciamento dei movimenti è cruciale nella compensazione del segnale e nelle fasi di pre- e post- elaborazione.

Una naturale evoluzione di questa applicazione considererà anche gli altri piani sagittali, ossia i piani verticali passanti per l'ascoltatore con azimuth $\neq 0$;verranno ottimizzate le procedure per ridurre il numero di posizioni delle sorgenti necessario e per meglio controllarne la posizione e l'orientamento rispetto ai movimenti dell'utente; SelfEar implementerà algoritmi di computer vision capaci di tracciare la posa della testa dell'ascoltatore in tempo reale attraverso, ad esempio, una fotocamera integrata o sensori di profondità.

Oltre alla funzionalità di acquisizione di HRTF, vi sarà la possibilità di acquisire BRIR strutturate, cioè memorizzando separatamente RIR e HRIR per rappresentare coerentemente scenari di mAAR. I parametri estratti dalle RIR verranno dati in input a modelli computazionali di ambienti acustici dinamici; si potranno utilizzare la modellazione immagine-sorgente e il raybeam-tracing per le prime riflessioni, mentre per le riverberazioni successive si potrà adottare una gestione statistica [12]. Infine, sarà necessaria una valutazione psico-acustica con soggetti umani per confermare l'affidabilità delle tecnologie del progetto SelfEar con l'obiettivo di creare sorgenti sonore virtuali esternalizzate fornendo HRIR efficaci.

8. RINGRAZIAMENTI

Questo lavoro di ricerca è supportato dal progetto di ricerca *Personal Auditory Displays for Virtual Acoustics* (PADVA), Università di Padova, grant no. CPDA135702.

9. BIBLIOGRAFIA

[1] S. Paul, "Binaural Recording Technology: A Historical Review and Possible Future Developments," *Acta Acustica united with Acustica*, vol. 95, pp. 767–788, Sept. 2009.

[2] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1983.

[3] M. Geronazzo, S. Spagnol, and F. Avanzini, "Mixed Structural Modeling of Head-Related Transfer Functions for Customized Binaural Audio Delivery," in *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, (Santorini, Greece), pp. 1–8, July 2013.

[4] H. Gamper, M. R. P. Thomas, and I. J. Tashev, "Anthropometric parameterisation of a spherical scatterer ITD model with arbitrary ear angles," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 1–5, Oct. 2015.

[5] M. Geronazzo, S. Spagnol, and F. Avanzini, "A Modular Framework for the Analysis and Synthesis of Head-Related Transfer Functions," in *Proc. 134th Conv. Audio Eng. Society*, (Rome, Italy), May 2013.

[6] M. Geronazzo, S. Spagnol, and F. Avanzini, "Estimation and Modeling of Pinna-Related Transfer Functions," in *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, (Graz, Austria), pp. 431–438, Sept. 2010.

[7] S. Prepeliţă, M. Geronazzo, F. Avanzini, and L. Savioja, "Influence of voxelization on finite difference time domain simulations of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 139, pp. 2489–2504, May 2016.

[8] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the Relation between Pinna Reflection Patterns and Head-Related Transfer Function Features," *IEEE Trans. Audio, Speech, Language Processing*, vol. 21, pp. 508–519, Mar. 2013.

[9] A. Ihlefeld and B. Shinn-Cunningham, "Disentangling the effects of spatial cues on selection and formation of auditory objects," *The Journal of the Acoustical Society of America*, vol. 124, no. 4, pp. 2224–2235, 2008.

[10] W. G. Gardner and K. D. Martin, "HRTF Measurements of a KEMAR," *The Journal of the Acoustical Society of America*, vol. 97, pp. 3907–3908, June 1995.

[11] V. Valimaki, A. Franck, J. Ramo, H. Gamper, and L. Savioja, "Assisted Listening Using a Headset: Enhancing audio perception in real, augmented, and virtual environments," *IEEE Signal Processing Magazine*, vol. 32, pp. 92–99, Mar. 2015.

[12] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, pp. 708–730, Aug. 2015.

[13] B. Boren, M. Geronazzo, F. Brinkmann, and E. Choueiri, "Coloration Metrics for Headphone Equalization," in *Proc. of the 21st Int. Conf. on Auditory Display (ICAD 2015)*, (Graz, Austria), pp. 29–34, July 2015.

[14] W. O. Brimijoin, A. W. Boyd, and M. A. Akeroyd, "The Contribution of Head Movement to the Externalization and Internalization of Sounds," *PLoS ONE*, vol. 8, p. e83068, Dec. 2013.

[15] J. S. Bradley and G. A. Soulodre, "Objective measures of listener envelopment," *The Journal of the Acoustical Society of America*, vol. 98, pp. 2590–2597, Nov. 1995.

[16] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 868–878, 1989.

[17] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," in *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, (New Paltz, New York, USA), pp. 1–4, Oct. 2001.

[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 *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, (Florence, Italy), pp. 4496–4500, May 2014.

[19] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *The Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 159–168, 1990.

[20] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, 1999.

[21] F. Christensen, P. F. Hoffmann, and D. Hammershi, "Measuring Directional Characteristics of In-Ear Recording Devices," in *In Proc. Audio Engineering Society Convention 134*, Audio Engineering Society, May 2013.

[22] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835 – 850, 2007.