

# VALUTAZIONE PARAMETRICA DI UN MODELLO STRUTTURALE DI ORECCHIO ESTERNO PER IL RENDERING BINAURALE DEL SUONO

Sandro Scaiella, Simone Spagnol, Michele Geronazzo, Federico Avanzini

Università di Padova, Dipartimento di Ingegneria dell'Informazione

avanzini@dei.unipd.it

## SOMMARIO

Gli effetti acustici cui è sottoposto il suono nel tragitto dalla sorgente alle orecchie dell'ascoltatore sono riassunti in una specifica funzione di trasferimento (Head-Related Transfer Function, HRTF), di cui la PRTF (Pinna-Related Transfer Function) rappresenta il solo contributo dato dal padiglione auricolare. La PRTF di un singolo ascoltatore può essere approssimata con un apposito modello sintetico che ne approssima i principali picchi e notch, i parametri dei quali sono solo in parte relazionati all'antropometria dell'ascoltatore stesso. Partendo da tale modello, l'articolo descrive un esperimento psicoacustico approntato con l'obiettivo di testare l'accuratezza (in termini di localizzazione verticale) del modello per sorgenti sonore sul piano mediano e per diverse combinazioni di parametri non relazionati all'antropometria. Si analizzano quindi i risultati relativi ai soggetti sperimentali e si ottiene la combinazione di parametri che meglio si adatta a una buona resa dell'elevazione della sorgente, ottenendo così utili indicazioni per la progettazione di modelli strutturali per la sintesi binaurale.

## 1. INTRODUZIONE

L'illusione che un dato suono sia posizionato in una determinata locazione virtuale rappresenta l'obiettivo principale di qualsiasi sistema di rendering di audio 3D. Tale obiettivo può essere raggiunto sia per mezzo di altoparlanti sia di cuffie. L'idea di base, sottesa al citato sistema, è quella di presentare ai due timpani dell'ascoltatore gli stessi segnali che produrrebbe una sorgente reale posizionata in quel particolare punto dello spazio di cui si vuol dare la percezione.

Il presente lavoro è dedicato allo studio della percezione dell'elevazione di sorgenti sonore localizzate sul piano mediano, ovvero equidistanti dalle due orecchie dell'ascoltatore. L'idea di fondo è quella di studiare e modellare un sistema che permetta, una volta ricavate alcune caratteristiche antropometriche relative all'ascoltatore, di ricostruire un paio di HRTF (*Head-Related Transfer Function*) individuali. Esse rappresentano dei filtri che modellano gli

effetti sulla propagazione del suono dovuti alla testa, all'orecchio esterno, al busto e alle spalle. Questo lavoro rientra all'interno di tale ambito di studi, soffermandosi esclusivamente sul contributo dato dal padiglione auricolare (*Pinna-Related Transfer Function*, PRTF) i cui effetti, come vedremo nel corso della trattazione, permettono di discriminare l'elevazione di una sorgente sonora [1].

Al fine di valutare i parametri di un modello di PRTF sintetico non relazionati alle caratteristiche antropometriche dell'orecchio dell'ascoltatore, si è resa necessaria la progettazione e l'implementazione di un esperimento psicoacustico nel quale tali parametri potessero assumere diversi possibili valori. Tale sperimentazione permette un'analisi comparativa di suddetti parametri al fine di determinare la combinazione degli stessi in grado di rendere il più fedelmente possibile l'elevazione desiderata.

## 2. IL MODELLO STRUTTURALE

Secondo la cosiddetta *modellazione strutturale* [2], la HRTF misurata all'ingresso del canale uditivo può essere scomposta in contributi tra loro indipendenti, dovuti agli effetti di testa, padiglione auricolare (pinna) e busto/spalle sull'onda sonora. Possiamo quindi considerare la pinna come un oggetto indipendente dalle altre parti anatomiche e modellarne il contributo separatamente.

Tale risposta assume una particolare rilevanza in quanto tipicamente le differenze più marcate tra le HRTF di più soggetti sono principalmente imputabili alle diverse caratteristiche della pinna (forma, grandezza e orientazione), talmente varie da rendere la pinna oggetto di recenti studi di identificazione biometrica [3]. La pinna ha un ruolo fondamentale nel determinare la forma spettrale delle HRTF grazie a due fenomeni acustici principali: riflessioni (sui bordi principali) e risonanze (nelle cavità). Di conseguenza, la corrispondente PRTF presenta nella sua ampiezza una sequenza di amplificazioni (picchi) in corrispondenza delle frequenze di risonanza e di nette attenuazioni (*notch*) in corrispondenza delle frequenze in cui si ha massima interferenza distruttiva tra onde dirette e onde riflesse. È proprio la locazione spettrale di tali picchi e notch a costituire un indicatore fondamentale per la caratterizzazione della posizione spaziale della sorgente sonora, e in particolare della sua elevazione [4].

Nella letteratura dedicata troviamo diverse proposte modellistiche per rendere sinteticamente le PRTF [5, 6, 7]. Tuttavia, tali modelli soffrono di evidenti limiti: la presenza della sola componente riflettente [5], la validità in regioni spaziali eccessivamente ristrette [6], e/o l'assenza di

Copyright: ©2014 Sandro Scaiella, Simone Spagnol, Michele Geronazzo, Federico Avanzini. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

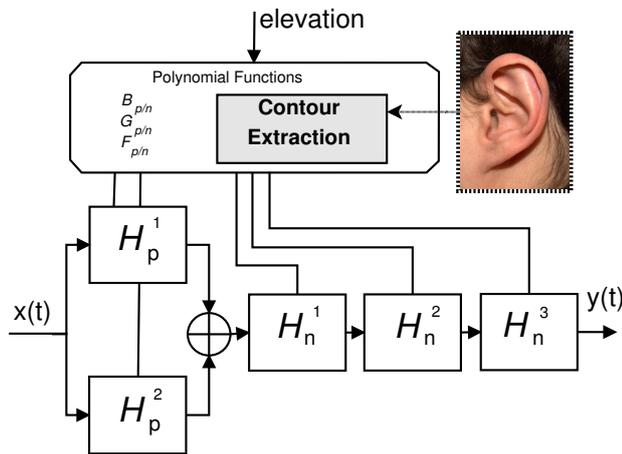


Figura 1. Il modello strutturale di pinna.

una parametrizzazione esplicita sull'antropometria dell'ascoltatore [7]. In un lavoro precedente [8] gli autori hanno proposto un modello strutturale di pinna costituito da due blocchi di filtri, il blocco *risonante* e il blocco *riflettente*. Il blocco risonante è costituito da due filtri *peak* del secondo ordine posti in parallelo; le uscite di tale blocco vengono sommate per essere inviate all'ingresso del blocco riflettente, costituito dalla cascata di 3 filtri *notch* del secondo ordine. Il modello è riportato schematicamente in Figura 1.

Nell'ambito dello stesso lavoro, gli autori hanno studiato la relazione tra le frequenze centrali dei notch presenti nelle PRTF e la geometria della pinna. A tal fine è stata utilizzata una procedura di *ray-tracing* su immagini 2D della pinna per mappare i punti di riflessione a una certa distanza dal punto di riferimento del canale uditivo, univocamente determinata dalla frequenza centrale di ogni notch. Mentre i modelli di riflessione precedenti tipicamente assumevano coefficienti di riflessione positivi, gli autori hanno verificato che l'utilizzo di coefficienti di riflessione negativi risulta un fattore chiave nella determinazione della frequenza centrale stessa. Assumendo dunque coefficienti negativi, la relazione tra frequenza del notch e distanza del punto di riflessione dal canale uditivo è completamente riassunta nella semplice equazione

$$d_i(\phi) = \frac{c}{2f_i(\phi)}, \quad (1)$$

dove la costante  $c$  rappresenta la velocità del suono,  $\phi$  è l'elevazione di cui si considera la PRTF,  $f_i$  è la frequenza centrale dell' $i$ -esimo notch in tale PRTF e  $d_i$  è la distanza tra il corrispondente punto di riflessione e l'ingresso del canale uditivo. Dai punti così trovati e mappati su immagini di pinne di un *pool* di soggetti sperimentali, gli autori hanno notato l'ottima corrispondenza tra i punti di riflessione e i tre contorni principali della pinna, ovvero bordo dell'elice, antielice/parete interna della conca, e bordo esterno della conca.

Partendo da tali risultati, gli autori in [9] hanno realizzato la procedura inversa: a partire dall'immagine della pinna sono stati tracciati i tre sopracitati contorni e, attraverso semplici calcoli trigonometrici, trasformati in coppie di coordinate polari  $(d_i(\phi), \phi)$  rispetto al canale uditivo.

Dall'Equazione 1 si ricavano quindi le frequenze centrali dei notch ad ogni elevazione  $\phi$  desiderata e per ognuno dei tre contorni. L'unico parametro indipendente usato dal modello è infatti l'elevazione della sorgente sonora virtuale, alla quale sono associate tre funzioni polinomiali che interpolano le frequenze centrali ricavate dai tre contorni per un certo passo di campionamento di  $\phi$  (vedi Figura 1).

Per ciò che riguarda i parametri di larghezza di banda e profondità dei notch stessi, non è ancora stata trovata una relazione con la forma della pinna. Gli autori hanno precedentemente utilizzato per questi due parametri valori medi di una popolazione di soggetti sperimentali [9], così come per i parametri caratteristici delle risonanze. Si rende quindi necessario uno studio approfondito dell'impatto psicoacustico di tali parametri nella percezione dell'elevazione di una sorgente sonora.

### 3. CONDIZIONI SPERIMENTALI

Poon e Brugge [10] descrivono i risultati di esperimenti su gatti nei quali vengono registrate le risposte di singole fibre del nervo uditivo a rumori a banda larga. All'interno di tali rumori viene introdotto un notch sintetico di frequenza centrale variabile tra 2.5 e 40 kHz. La tesi degli autori è che l'informazione contenuta nella frequenza e nell'intensità di rilevanti notch nelle HRTF venga trasmessa al sistema nervoso centrale attraverso il tasso di scarica delle fibre uditive. Dai risultati si ricava che le fibre del nervo uditivo sono molto sensibili alla presenza di notch con frequenza centrale prossima alla frequenza caratteristica delle fibre. In particolare, si nota un'ampia riduzione nel tasso di scarica delle fibre attorno alla frequenza caratteristica delle stesse quando la frequenza del notch è vicina, tasso che scende occasionalmente anche al di sotto dell'emissione spontanea che si avrebbe in assenza di stimolo acustico. Gli autori hanno altresì stimato il livello dello stimolo, sopra la soglia della fibra, al quale il notch produce la maggiore riduzione del tasso di scarica, ricavandone un valore compreso tra 15 e 35 dB. Infine, i due autori hanno trovato un'ottima corrispondenza tra un notch di 35 dB di profondità e la curva di sintonizzazione della fibra uditiva di frequenza caratteristica più vicina.

In [11], Alves-Pinto e Lopez-Poveda mostrano come l'abilità di rilevare notch ad alta frequenza dipenda dalla larghezza di banda del notch così come dal livello dello stimolo. In generale, i risultati mostrano come la soglia di profondità del notch affinché risulti rappresentabile nella risposta del nervo uditivo (in termini di variazione nel tasso di scarica) incrementi al diminuire della larghezza di banda del notch stesso. Da questo lavoro si evince infatti che mentre per notch con larghezza di banda tra 2 e 4 kHz la soglia resta al di sotto dei 10 dB, il rilevamento di notch con larghezza di banda minore (1 kHz) risulta problematico anche per notch molto profondi.

Risultati simili erano precedentemente stati trovati da Moore [12]. Nel citato studio l'autore analizza l'abilità di un gruppo di soggetti di rilevare e discriminare picchi e notch con frequenze centrali di 1 e 8 kHz inseriti in stimoli di rumore a banda larga. Le larghezze di banda prese in esame per i notch sono di  $0.125f_c$ ,  $0.25f_c$  e  $0.5f_c$ , tutte

relative alla frequenza centrale  $f_c$  del notch. Considerando la soglia per la frequenza di 8 kHz - tipica di uno dei maggiori indicatori per il rilevamento dell'elevazione nel piano mediano nelle HRTF, ovverosia un notch di larghezza di banda di  $0.7f_c$  [4] - osserviamo che i notch con larghezza di banda pari a  $0.125f_c$  risultano di difficile discriminazione, mentre tutti i soggetti esaminati riescono a rilevare notch con larghezze di banda di  $0.25f_c$  e  $0.5f_c$ . Le soglie di profondità variano tra 10 e 20 dB per larghezze di banda di  $0.25f_c$  e diminuiscono sensibilmente per larghezze di banda di  $0.5f_c$ . In base a tali risultati, il sopracitato indicatore di elevazione dovrebbe essere sempre rilevato.

In [13], Iida *et al.* hanno effettuato test di localizzazione nel piano mediano utilizzando un modello parametrico di HRTF. Tale modello è costituito solo da alcuni dei picchi e notch estratti dalle HRTF originali. Gli autori hanno trovato che per la HRTF parametrica (*p-HRTF*) composta da tutti i picchi e notch rilevanti, l'elevazione percepita è accurata tanto quanto quella percepita utilizzando le HRTF misurate sul singolo individuo. Al contrario, si hanno scarsi risultati considerando *p-HRTF* costituite da un solo notch o da un solo picco. Per le *p-HRTF* costituite solo dai primi due notch rilevanti (*N1* e *N2*), estratti dalle HRTF a partire dalla frequenza di 4 kHz, i risultati mostrano un'accuratezza nella percezione dell'elevazione paragonabile a quella che si avrebbe con le HRTF originali. Questo implica che *N1* e *N2* assumono un ruolo importantissimo nella percezione dell'elevazione di un suono nel piano mediano. Dai risultati è evidente anche il ruolo del secondo picco nella percezione di sorgenti sopra la testa. Per quel che riguarda il primo picco (*P1*), pur non essendo legato alla percezione dell'elevazione, può essere visto come un importante riferimento per il sistema uditivo umano nell'analisi dei primi due notch.

Dalle osservazioni tratte dai lavori sopracitati abbiamo ricavato utili indicazioni per la scelta dei nostri parametri sperimentali. Consideriamo i 3 parametri associati ad un singolo notch:

- $f_c$ , frequenza centrale del notch;
- $G$ , guadagno (profondità) del notch;
- $BW$ : larghezza di banda del notch calcolata a -3 dB.

Come detto in precedenza, l'unico parametro associato alle misurazioni antropometriche del soggetto è  $f_c$ . Per la larghezza di banda consideriamo invece tre possibili valori:  $BW = 2$  kHz,  $BW = 0.15f_c$ ,  $BW = 0.25f_c$ . La scelta di larghezze di banda di  $0.25f_c$  e fissa a 2 kHz è dovuta alla constatazione che tali valori corrispondono alle minime larghezze di banda, relative e fisse, alle quali la maggior parte dei soggetti riesce a rilevare un notch alle frequenze di nostro interesse, come mostrato rispettivamente in [12] e in [11]. Anche se la larghezza di  $0.15f_c$  risulta di difficile percezione in base alla letteratura [12], la decisione di testare tale larghezza di banda nasce dalla considerazione che a tale valore si ha meno sovrapposizione tra notch sintetici affiancati, ottenendo quindi una PRTF più verosimile.

Per quanto concerne la profondità del notch, abbiamo deciso di utilizzare 2 diversi valori:

Condizione	Guadagno	Banda	Nr. notch
$C_1$	-10 dB	$0.15f_c$	3
$C_2$	-10 dB	$0.15f_c$	2
$C_3$	-10 dB	$0.25f_c$	3
$C_4$	-10 dB	$0.25f_c$	2
$C_5$	-10 dB	2 kHz	3
$C_6$	-10 dB	2 kHz	2
$C_7$	-30 dB	$0.15f_c$	3
$C_8$	-30 dB	$0.15f_c$	2
$C_9$	-30 dB	$0.25f_c$	3
$C_{10}$	-30 dB	$0.25f_c$	2
$C_{11}$	-30 dB	2 kHz	3
$C_{12}$	-30 dB	2 kHz	2

Tabella 1. Condizioni sperimentali.

1.  $G = -10$  dB: la minima soglia alla quale, con larghezza di banda di  $0.25f_c$ , gran parte dei soggetti riesce a rilevare il notch, come mostrato in [12], e, allo stesso tempo, la soglia sotto la quale, con larghezze di banda fisse a 2 e 4 kHz, si distribuiscono le soglie psicoacustiche di profondità dei notch, come mostrato in [11];
2.  $G = -30$  dB: deriva dalla corrispondenza tra un notch con profondità di 35 dB, di poco superiore al valore da noi scelto, e le curve di sintonizzazione delle fibre uditive, come mostrato in [10].

Infine, l'ultimo parametro variabile delle nostre PRTF sintetiche è il numero di notch presenti. Abbiamo testato PRTF costituite da 2 picchi e 3 notch e PRTF composte da 2 picchi e 2 notch, eliminando quello centrale. Tale decisione è in accordo con i risultati ottenuti in [13], dove, come esposto poc'anzi, la presenza di due soli notch è risultata sufficiente per una localizzazione accurata nel piano mediano. In Tabella 1 sono riepilogate le 12 condizioni sperimentali testate, date da tutte le combinazioni di larghezza di banda, profondità e presenza/assenza del secondo notch.

#### 4. SETUP SPERIMENTALE

Al fine di valutare le prestazioni soggettive di ognuna delle 12 condizioni sperimentali applicata al modello strutturale di PRTF, abbiamo effettuato un esperimento di localizzazione su 12 soggetti (9M/3F, età 23-41). Soltanto 1/3 dei soggetti aveva precedentemente avuto esperienze con test di localizzazione, e nessuno è stato sottoposto a sessioni di training. Una procedura automatizzata per la rilevazione della soglia uditiva [14] ha certificato che nessun soggetto presentasse perdite rilevanti fino agli 8 kHz binauralmente.

##### 4.1 Stimoli

Lo stimolo è costituito da un treno di 3 impulsi di rumore bianco uniformemente distribuito, ognuno della durata di 300 ms. In assenza di filtraggio con PRTF, esso è presentato a un livello di 60 dBA (misurati sulle cuffie, al livello della pinna). Gli impulsi sono separati da pause di 250 ms e sagomati con finestre di Hann in modo da ottenere

fade-in e fade-out. Lo stimolo originale viene fatto passare attraverso un filtro per la compensazione delle cuffie utilizzate per l'esperimento [15] e quindi convoluto con una delle PRTF sintetiche modellata a seconda della posizione spaziale della sorgente virtuale e della condizione sperimentale. La componente risonante del modello strutturale è fissata a priori per tutti i soggetti e riceve in ingresso parametri mediati su una popolazione di HRTF, i quali variano con l'elevazione [9].

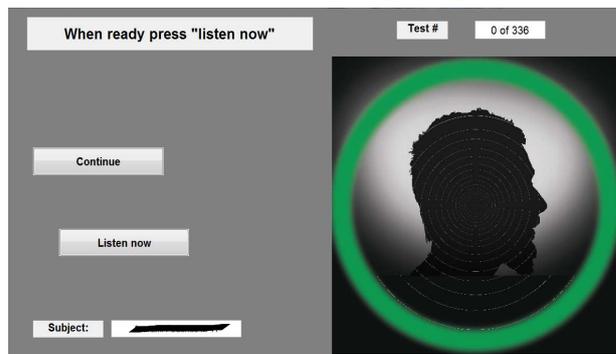
È importante a questo punto chiarire due aspetti fondamentali. Innanzitutto, ciò che viene presentato all'ascoltatore è uno stimolo filtrato con una PRTF e non con una HRTF completa. Tuttavia, l'unica differenza sostanziale tra PRTF e HRTF nel piano mediano è rappresentata dalla presenza del contributo di busto e spalle nella HRTF. Dal momento che tale contributo non rappresenta che un debole indicatore di elevazione alle basse frequenze, possiamo assumere  $PRTF(\phi) \approx HRTF(\phi)$  nel piano mediano. In secondo luogo, mentre i parametri delle risonanze e le frequenze centrali dei notch variano con l'elevazione, i parametri banda e profondità dei notch restano fissi lungo tutto il range a parità di condizione sperimentale. In realtà, secondo uno studio precedentemente svolto su un database di HRTF, all'aumentare dell'angolo di elevazione la profondità dei notch tende mediamente a decrescere e la banda ad aumentare [9]. Ciononostante, tale scelta sperimentale è stata dettata dalla difficoltà nel rilevare un andamento univoco dei due parametri all'interno dei set di HRTF analizzati.

## 4.2 Protocollo

Come primo step acquisiamo un'immagine di profilo del soggetto, in modo tale da avere una rappresentazione della sua pinna sinistra. In una seconda fase, l'immagine viene prima ruotata per allineare la punta del naso del soggetto con il trago dell'orecchio; quindi, il punto di massima protuberanza del trago viene scelto come punto d'ingresso del canale uditivo. I tre principali contorni dell'orecchio vengono tracciati manualmente e usati per ricavare le frequenze centrali di ogni notch come esposto in Sezione 2. Il soggetto entra quindi in una cabina silente Sound Station Pro 45 e indossa un paio di cuffie Sennheiser HDA 200 cablate a una scheda Roland Edirol AudioCapture UA-101 impostata a una frequenza di campionamento di 44.1 kHz.

Le elevazioni testate sono  $\phi \in [-45^\circ, -30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ, 45^\circ]$ , tutte sul piano frontale e mediano, dove un'elevazione di  $0^\circ$  corrisponde esattamente alla direzione frontale. Elevazioni maggiori di  $45^\circ$  sono state scartate a causa della tipica mancanza di notch nelle corrispondenti HRTF [16], mentre per elevazioni minori di  $-45^\circ$  interverrebbero questioni di postura del soggetto a complicare la percezione delle elevazioni stesse.

Nel corso del test di localizzazione ogni stimolo viene presentato, in modalità diotica, 4 diverse volte per ogni elevazione e condizione. Tale scelta permette di evitare che i risultati finali siano influenzati da possibili errori di attenzione e di vagliare un possibile effetto di *learning*. Considerando dunque le 7 elevazioni, le 12 condizioni e le 4 ripetizioni, otteniamo un totale di  $7 \times 12 \times 4 = 336$



**Figura 2.** La GUI utilizzata per l'esperimento di localizzazione.

stimoli, presentati al soggetto in 4 blocchi da 84 stimoli, ciascuno associato ad una singola ripetizione. L'ordine degli stimoli all'interno di ogni blocco è una permutazione pseudocasuale degli stessi.

Uno schermo posto di fronte al soggetto mostra l'interfaccia grafica riportata in Figura 2. Ad ogni presentazione dello stimolo sonoro il soggetto, utilizzando un comune mouse, indica sull'interfaccia un punto all'interno della corona verde (raffigurante il piano mediano) nella direzione percepita del suono. Pochi decimi di secondo dopo la *click*, un nuovo stimolo viene inviato in cuffia. Alla fine di ogni blocco di 84 stimoli compare una finestra di *countdown* e al soggetto viene concessa una pausa di 3 minuti prima di proseguire con i blocchi successivi. L'intero test di localizzazione dura, considerando le 3 pause inserite, circa 45 minuti in media.

## 5. RISULTATI

Per ognuno dei 12 soggetti e per ognuna delle 12 condizioni sperimentali abbiamo innanzitutto calcolato l'errore medio di localizzazione, inteso come differenza assoluta tra elevazione *target* ed elevazione percepita (calcolata a partire dal punto inserito nella corona circolare della GUI) a seguito di un'eventuale correzione di errori *front/back*,<sup>1</sup> e la percentuale stessa di errori *front/back* sul totale degli stimoli. Poiché non è stato trovato alcun trend significativo riguardo la percentuale di errori *front/back* al variare della condizione sperimentale, non riportiamo i dati associati a questa misura. I valori di errore di localizzazione sono invece riportati in Tabella 2.

Appare innanzitutto netta la differenza in termini di errore di localizzazione assoluto tra i 4 soggetti con esperienza pregressa in test psicoacustici ( $S_3, S_6, S_8, S_{12}$ ) e gli altri. Al fine di non inficiare i risultati dei soggetti migliori nel prosieguo della nostra analisi, abbiamo convenuto di non considerare i soggetti la cui performance fosse paragonabile a una performance completamente casuale in oltre il 60% delle condizioni (8 su 12 o più). Abbiamo perciò calcolato l'errore di localizzazione associato a valori inseriti da un generatore di numeri pseudocasuali distribuiti

<sup>1</sup> La correzione consiste nel mappare, con simmetria assiale, tutti i valori posti nell'emisfero posteriore nell'emisfero frontale. Tale pratica è ormai comune negli esperimenti di localizzazione [17] ed è giustificata dall'elevato tasso di inversioni cui sono soggetti stimoli virtuali.

Sogg. \ Cond.	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$	$C_9$	$C_{10}$	$C_{11}$	$C_{12}$
$S_1$	31.4°	30.7°	32.1°	30.5°	33.6°	26.1°	37.2°	28.7°	37.3°	39.6°	40.8°	29.5°
$S_2$	32.9°	41.3°	40.9°	39.8°	46.2°	41.2°	38.0°	45.0°	50.0°	42.4°	48.3°	35.2°
$S_3$	20.0°	19.6°	12.3°	17.9°	16.6°	20.1°	13.3°	16.4°	14.4°	16.7°	11.5°	17.8°
$S_4$	40.6°	47.5°	44.5°	44.0°	39.1°	43.0°	47.5°	36.3°	47.3°	33.4°	52.6°	45.2°
$S_5$	24.1°	24.8°	24.3°	29.1°	22.7°	28.1°	24.7°	25.5°	25.6°	26.3°	23.1°	24.7°
$S_6$	26.9°	20.3°	22.8°	24.8°	27.0°	17.9°	20.5°	16.8°	26.6°	17.0°	19.4°	23.6°
$S_7$	35.2°	42.4°	47.6°	38.4°	47.2°	40.4°	36.3°	40.7°	43.4°	45.7°	45.5°	46.6°
$S_8$	19.9°	21.2°	20.4°	18.9°	17.5°	23.4°	15.9°	19.3°	19.3°	16.7°	16.6°	17.9°
$S_9$	30.4°	41.0°	36.3°	26.6°	28.6°	25.4°	37.8°	26.1°	35.1°	36.2°	32.4°	39.4°
$S_{10}$	39.6°	41.3°	46.6°	42.1°	32.1°	40.8°	45.5°	38.4°	36.3°	48.6°	45.8°	53.2°
$S_{11}$	35.3°	47.8°	44.0°	49.0°	42.4°	46.1°	39.9°	43.6°	26.6°	46.4°	28.6°	39.4°
$S_{12}$	33.6°	38.7°	31.4°	36.5°	29.0°	45.4°	35.2°	33.7°	34.3°	42.6°	28.7°	41.5°

**Tabella 2.** Errore di localizzazione medio diviso per soggetto e condizione sperimentale. In rosso sono evidenziate le condizioni le cui performance risultano paragonabili a una performance casuale. In verde sono evidenziati i soggetti considerati per la successiva analisi e le migliori condizioni sperimentali associate a ciascuno di essi.

Condizione	Errore	Inversioni $U/D$	$slope+r^2$
$C_1$	27.7°	17.0%	0.52
$C_2$	30.5°	16.5%	0.47
$C_3$	27.9°	17.0%	0.65
$C_4$	29.1°	16.1%	0.51
$C_5$	27.2°	16.1%	0.65
$C_6$	29.1°	16.1%	0.48
$C_7$	28.1°	19.6%	0.59
$C_8$	26.3°	12.5%	0.69
$C_9$	27.4°	12.5%	0.70
$C_{10}$	30.2°	17.9%	0.48
$C_{11}$	25.2°	12.9%	0.82
$C_{12}$	29.2°	16.5%	0.42

**Tabella 3.** Risultati mediati sugli 8 soggetti considerati buoni localizzatori. In rosso sono evidenziate le 3 condizioni migliori per risultato.

uniformemente nell'intervallo  $[-90^\circ, 90^\circ]$ . In questo caso l'errore medio di elevazione converge a un valore di  $49^\circ$  e il quinto percentile a un valore intorno ai  $40^\circ$ . Quest'ultimo valore è stato considerato come soglia dell'errore medio di localizzazione oltre la quale la performance risulta non attendibile, almeno per quella determinata condizione sperimentale. Di conseguenza, sono stati scartati dalle successive analisi i soggetti  $S_2, S_4, S_7$  e  $S_{10}$ .

Nella successiva Tabella 3 sono quindi riportati i risultati medi per condizione sugli 8 soggetti restanti. In particolare, la tabella riporta

- il valore medio dell'errore di localizzazione, considerando anche i valori oltre soglia (errore  $> 40^\circ$ );
- la percentuale di inversioni *up/down*, calcolata con una tolleranza di  $15^\circ$  attorno al piano orizzontale per tutte le elevazioni *target* tranne  $\phi = 0^\circ$ ;
- la somma dei parametri *slope* (pendenza) e  $r^2$  (coefficiente di determinazione) della retta di regressione calcolata tra elevazioni *target* e percepite.

Confrontando i dati delle 3 colonne si evince che la condizione  $C_{11}$  risulta essere la migliore. A supporto di tale

considerazione si noti in Tabella 2 come  $C_{11}$  dia, per 5 degli 8 soggetti selezionati, il migliore o il secondo migliore errore di localizzazione, e che in essa soltanto il Soggetto  $S_1$  fallisce nel localizzare efficacemente la sorgente virtuale. Notiamo inoltre che le condizioni dispari (3 notch) danno risultati migliori rispetto alle condizioni pari (2 notch) associate, ad eccezione di  $C_7$  e  $C_8$ ; e che le condizioni globalmente migliori ( $C_8, C_9, C_{11}$ ) hanno in comune una profondità di notch  $G = -30$  dB.

A prima vista, la larghezza di banda dei notch sembra invece non avere un impatto significativo. Tuttavia, considerando i risultati dell'errore di localizzazione associati a ognuno degli 8 soggetti selezionati, notiamo che per per i soggetti che prediligono una condizione compresa tra  $C_1$  e  $C_6$  ( $S_1, S_5, S_9$ ) l'elemento comune tra tali condizioni è la larghezza di banda, fissa a  $BW = 2$  kHz. Al contrario, tutte e 3 le larghezze di banda appaiono tra le condizioni predilette comprese tra  $C_7$  e  $C_{12}$  (soggetti  $S_3, S_6, S_8, S_{11}, S_{12}$ ), tutte con  $G = -30$  dB. Un'ispezione delle risposte in ampiezza dei filtri associati alle condizioni  $C_7, C_9, C_{11}$  (non riportate in questo articolo) rivela come queste siano tra loro molto simili a prescindere dalla larghezza di banda specificata.

Dal nostro esperimento risultano quindi avere una notevole importanza per una buona percezione dell'elevazione:

- una profondità di notch superiore a 10 dB, risultato in accordo con Moore [12] e Poon e Brugge [10];
- la presenza del secondo notch in una banda di frequenza tra 7 e 10 kHz, in accordo con Iida *et al.* [13];
- una banda sufficientemente larga affinché il notch venga rilevato, ad esempio fissa a 2 kHz, in accordo con Alves-Pinto e Lopez-Poveda [11].

## 6. CONCLUSIONI E SVILUPPI

In questo lavoro è stata analizzata, attraverso un esperimento psicoacustico all'uopo approntato, l'influenza della profondità, della larghezza di banda e del numero dei

notch presenti nelle PRTF nella percezione dell'elevazione di sorgenti sonore virtuali poste sul piano mediano. I risultati di tale esperimento hanno sostanzialmente confermato le indicazioni mutuata dalla letteratura precedente riguardanti tali parametri. Inoltre, questi ci hanno offerto incoraggianti riscontri sull'efficacia del modello strutturale nel rendere l'elevazione di una sorgente sonora virtuale, in particolare considerando i risultati dei soggetti cosiddetti "esperti". Una sperimentazione più approfondita sull'impatto psicoacustico del modello individualizzato sulla percezione dell'elevazione, paragonato a condizioni con HRTF individuali e non, rappresenta naturalmente un passo fondamentale verso la validazione dello stesso.

Un ulteriore aspetto da considerare per studi futuri riguardo i parametri caratteristici delle PRTF è la possibile relazione tra la componente risonante della PRTF e la forma delle cavità della pinna. Ricordiamo infatti che nel presente studio si è utilizzato il medesimo blocco risonante per tutti i soggetti. Infine, l'utilizzo di larghezze di banda e di profondità variabili per i notch potrà evidenziare o meno la rilevanza di tali variazioni nel determinare l'elevazione della sorgente sonora virtuale.

Nel corso del lavoro sono emerse alcune criticità dalla cui risoluzione tali successive sperimentazioni non potranno prescindere. In primo luogo, le differenze notate tra soggetti con e senza esperienza pregressa rende necessaria una sessione di *training* preliminare al test psicoacustico. In secondo luogo, le difficoltà riscontrate in alcuni soggetti nell'interpretare il metodo di inserimento dei dati da noi utilizzato pone la necessità di investigare metodi di puntamento alternativi, possibilmente di natura maggiormente ecologica [18].

## 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] S. K. Roffler and R. A. Butler, "Factors that influence the localization of sound in the vertical plane," *J. Acoust. Soc. Am.*, vol. 43, pp. 1255–1259, June 1968.
- [2] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Trans. Speech Audio Process.*, vol. 6, pp. 476–488, September 1998.
- [3] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Trans. Embedded Computing Systems*, vol. 9, pp. 39:1–39:33, March 2010.
- [4] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1829–1834, December 1974.
- [5] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, vol. 63, pp. 1152–1165, April 1978.
- [6] P. Satarzadeh, R. V. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proc. 122nd Conv. Audio Eng. Soc.*, (Vienna, Austria), pp. 718–737, May 2007.
- [7] K. J. Faller II, A. Barreto, and M. Adjouadi, "Augmented Hankel total least-squares decomposition of head-related transfer functions," *J. Audio Eng. Soc.*, vol. 58, pp. 3–21, January/February 2010.
- [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, Lang. Process.*, vol. 21, pp. 508–519, March 2013.
- [9] M. Geronazzo, S. Spagnol, and F. Avanzini, "A head-related transfer function model for real-time customized 3-D sound rendering," in *Proc. INTERPRET Work., SITIS 2011 Conf.*, (Dijon, France), pp. 174–179, November-December 2011.
- [10] P. W. F. Poon and J. F. Brugge, "Sensitivity of auditory nerve fibers to spectral notches," *J. Neurophysiol.*, vol. 70, pp. 655–666, August 1993.
- [11] A. Alves-Pinto and E. A. Lopez-Poveda, "Detection of high-frequency spectral notches as a function of level," *J. Acoust. Soc. Am.*, vol. 118, pp. 2458–2469, October 2005.
- [12] B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Am.*, vol. 85, pp. 820–836, February 1989.
- [13] 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," *Appl. Acoust.*, vol. 68, pp. 835–850, August 2007.
- [14] D. M. Green, "A maximum-likelihood method for estimating thresholds in a yes-no task," *J. Acoust. Soc. Am.*, vol. 93, pp. 2096–2105, April 1993.
- [15] A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *J. Audio Eng. Soc.*, vol. 60, pp. 54–62, January 2012.
- [16] S. Spagnol, M. Hiipakka, and V. Pulkki, "A single-azimuth pinna-related transfer function database," in *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, (Paris, France), pp. 209–212, September 2011.
- [17] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. 85, pp. 868–878, February 1989.
- [18] L. Haber, R. N. Haber, S. Penningroth, K. Novak, and H. Radgowski, "Comparison of nine methods of indicating the direction to objects: Data from blind adults," *Perception*, vol. 22, no. 1, pp. 35–47, 1993.