Associazione Italiana di Acustica
42° Convegno Nazionale
Firenze, 16-17 luglio 2015

DISCRIMINAZIONE DELLA DISTANZA RELATIVA TRA SORGENTI SONORE VIRTUALI

Erica Tavazzi (1), Simone Spagnol (1), Federico Avanzini (1)

1) Università di Padova, Dipartimento di Ingegneria dell'Informazione

1. Introduzione

Lo sviluppo negli ultimi anni di applicazioni che richiedono la simulazione di ambienti sonori ha favorito la ricerca di nuove e più precise tecniche di elaborazione di segnali audio. Le tecniche binaurali, che permettono la riproduzione di suoni tramite un sistema di cuffie o altoparlanti, sono infatti alla base di applicazioni quali l'immersione in realtà virtuali a scopo ludico, nell'ambito delle telecomunicazioni o nell'assistenza con tecnologia audio di supporto.

Le caratteristiche spaziali tridimensionali dei suoni possono essere ricreate processando l'input acustico attraverso una coppia di filtri (rispettivamente destro e sinistro), ciascuno in grado di riprodurre le trasformazioni lineari che caratterizzano il suono emesso dalla sorgente nel campo libero nel suo tragitto fino al timpano dell'ascoltatore. Tali filtri sono noti in letteratura come HRTF (Head Related Transfer Functions) e rappresentano formalmente il rapporto tra il livello di pressione del suono p_s che una sorgente puntiforme genera sul timpano dell'ascoltatore e la corrispondente pressione p_{ff} al centro della testa in campo libero, ovvero come se l'ascoltatore fosse assente. Tale funzione, che comprende dunque tutti i parametri di riflessione, diffrazione, risonanza ed assorbimento relativi alla testa umana, è misurabile e riproducibile tramite modelli matematici. Si parla quindi di sorgenti virtuali, in opposizione alle reali, nel caso in cui i suoni riprodotti siano stati elaborati tramite l'utilizzo di HRTF. Combinando l'uso di queste funzioni con dispositivi di head-tracking, si rende possibile la creazione di uno scenario audio real-time interattivo, in grado di simulare sorgenti sonore in qualsiasi posizione rispetto all'ascoltatore. Questi display audio virtuali (VAD, Virtual Auditory Display) trovano importanti applicazioni nell'intefaccia uomo-macchina e nella creazione di ambienti virtuali realistici [1].

Tipicamente, nei VAD il parametro più difficile da stimare a livello percettivo è quello legato all'informazione di distanza, in particolare per quanto riguarda tentativi di localizzazione della sorgente in termini assoluti.

Nella determinazione della distanza di un suono emesso da una sorgente in condizioni anecoiche, il fattore percettivo predominante è l'intensità percepita: quanto più l'intensità è debole, tanto più la sorgente viene considerata lontana. Tuttavia, la familiarità col suono udito è un presupposto fondamentale per poter esprimere un giudizio basato sull'informazione di intensità, che diventa significativa, altrimenti, solo su base relativa. La soglia minima di discriminabilità (jnd, just noticeable difference) nella distanza relativa tra due sorgenti isodirezionali può essere direttamente correlata ad una jnd di intensità del 5% [2], fino a raggiungere valori intorno al 50% per sorgenti poste molto vicine all'ascoltatore.

Si è osservato, inoltre, che la distanza assoluta di una sorgente che emette suoni familiari all'ascoltatore risulta meglio stimata quando la sorgente è posta lateralmente al soggetto (specialmente lungo il suo asse interaurale), rispetto a quando la sorgente è posta sul piano mediano, pur emergendo una sottostima della distanza per sorgenti lontane ed una sovrastima per sorgenti vicine. Nel caso di ambienti riverberanti, invece, il fattore principale su cui si basa la percezione assoluta della distanza è il rapporto tra l'energia riflessa e la diretta, definito come *R/D ratio* (per una review più dettagliata sull'argomento si consiglia [3]).

Nonostante le notevoli difficoltà incontrate nello stimare la localizzazione assoluta nel caso di VAD [4][5], sono stati condotti esperimenti sulla percezione della distanza relativa tra sorgenti virtuali [6] che hanno confermato la presenza di fattori legati alla percezione di distanza anche nel momento in cui l'informazione di intensità del suono emesso dalla sorgente viene eliminata. Tale studio ha evidenziato, inoltre, una differenza nella percezione della distanza relativa di una coppia di stimoli a seconda dell'ordine di presentazione degli stessi. Ci si è dunque proposti di approfondire e discutere quest'ultimo effetto tramite un esperimento psicoacustico *ad hoc*, basato sull'analisi delle soglie individuali percentuali di discriminazione di coppie di stimoli creati con il metodo DVF che si presenta di seguito.

2. Spherical Transfer Function e metodo DVF

Considerata la testa in prima approssimazione di forma sferica e di raggio a, presa una sorgente ad un angolo di incidenza α dall'ascoltatore (intendendo con questo l'angolo che si viene a formare tra i segmenti che congiungono il centro della testa rispettivamente con la sorgente e il punto di osservazione sulla superficie della sfera) e distante r dal centro della testa, l'HRTF nel punto di osservazione è ben descritta dalla seguente funzione di trasferimento, definita come STF (Spherical Transfer Function):

$$STF(\mu,\alpha,\rho) = -\frac{\rho}{\mu} e^{-i\mu\rho} \sum_{m=0}^{\infty} (2m+1) P_m(\cos\alpha) \frac{h_m(\mu\rho)}{h'_m(\mu)} \tag{1}$$

dove:

ρ è la distanza normalizzata, definita come $\rho = r/a$;
P_m è il polinomio di Legendre di grado m;
h_m è la funzione sferica di Hankel di ordine m-simo;
h'_m è la derivata prima di h_m rispetto al suo argomento;
μ è la frequenza normalizzata, definita come:

$$\mu = f\frac{2\pi a}{c} \tag{2}$$

dove:

f è la frequenza [Hz];

a è il raggio della sfera [m];

c è la velocità di propagazione del suono nel mezzo, fissata a c = 343.2 [m/s].

In un precedente lavoro [7], gli autori hanno analizzato l'andamento della STF utilizzando l'analisi alle componenti principali (PCA), facendo emergere che tale funzione è fortemente dipendente dall'angolo di incidenza e, solo in misura più attenuata, dalla distanza. Tuttavia, la dipendenza dalla distanza emerge chiaramente se si considera il rapporto tra i moduli di una STF nel campo vicino (r < 1m) ed una STF nel campo lontano (r > 1m). Questo concetto sta alla base del metodo DFV introdotto da Kan *et al.* [8], in cui viene definita la Distance Variation Function come:

$$DVF(\mu,\alpha,\rho_n,\rho_f) = \frac{p_S(\mu,\alpha,\rho_n)}{p_S(\mu,\alpha,\rho_f)} \tag{3}$$

dove:

p_s é il livello di pressione del suono generato dalla sorgente sul punto di osservazione;

ρ_n e ρ_f sono le distanze normalizzate rispettivamente di campo vicino (near field) e campo lontano (far field)

La DVF così definita può essere applicata come fattore moltiplicativo (frequenza per frequenza) ad una HRTF in una data direzione nel campo lontano, così da ottenere la risposta nel campo vicino alla distanza normalizzata ρ_n e nella direzione fissata dalla HRTF.

3. Design sperimentale

Per investigare la soglia di discriminazione della distanza relativa tra sorgenti sonore virtuali rese binauralmente nel campo vicino tramite il metodo DVF illustrato sopra, è stato condotto un esperimento psicoacustico sul modello di quelli proposti in [6].

Come stimolo sono state usate coppie di sorgenti virtuali isodirezionali a due differenti distanze, presentate in sequenza. Al soggetto è stato assegnato il compito di determinare quale dei due suoni uditi nella stessa sequenza fosse più vicino.

Nella creazione del VAD si è scelto di utilizzare come riferimento HRTF misurate su un manichino KEMAR (Knowles Electronics Manikin for Acoustic Research) con sorgenti poste nel campo lontano (distanza r = 1.6 m dal centro della testa del manichino) dal database PKU&IOA [4]. La scelta di utilizzare HRTF non individuali nel campo lontano, già effettuata in lavori precedenti [5], è giustificata dalla prospettiva di riuscire a simulare uno scenario virtuale per applicazioni pratiche in cui le HRTF individuali non siano disponibili. Nonostante sia noto dalla letteratura come l'uso di HRTF non individuali sia fonte di errori di localizzazione, tra cui l'inversione della percezione azimutale (front/back reversal) [9], la falsa percezione dell'angolo di elevazione [10] e la sensazione che la sorgente sia posta in un punto interno alla testa dell'ascoltatore [11], non sono state evidenziate differenze significative per quanto riguarda la percezione della distanza nel confronto con l'uso di HRTF individuali [12]. Il database PKU&IOA è stato scelto per consistenza con gli esperimenti precedenti.

Agli stimoli non si è aggiunta informazione legata al riverbero, così da poter esercitare maggior controllo sui fattori anecoici che permettono la determinazione della distanza.

3.1. Soggetti e apparato sperimentale

All'esperimento hanno preso parte 20 soggetti (7 femmine e 13 maschi) su base volontaria. L'età dei soggetti è compresa tra i 22 e i 49 anni (media = 27.2, SD = 6.8). Tutti i soggetti rientrano nella definizione di normo-udenza, con soglia uditiva inferiore ai 25 dB nell'intervallo tra i 125 Hz e gli 8 kHz, verificata tramite uno screening audiometrico.

L'esperimento è stato condotto all'interno di una cabina insonorizzata Sound Station Pro 45. Il soggetto sperimentale è stato fatto sedere di fronte a un tavolino, sul quale era posta una tastiera con le frecce di direzione superiore ed inferiore colorate rispettivamente di blu e di rosso.

Al soggetto sono state fatte indossare una paio di cuffie Sennheiser HDA 200 (risposta in frequenza 20 - 20k Hz, impedenza 40Ω), connesse ad una scheda audio esterna Roland Edirol AudioCapture AU-101, con frequenza di campionamento 48 kHz. Per compensare la risposta delle cuffie è stato utilizzato il filtro di compensazione proposto da Lindau e Brinkmann [13].

Tastiera e cuffie sono state entrambe collegate al PC utilizzato per l'esperimento, il cui schermo era posto di fronte al soggetto ed oscurato durante le sessioni di acquisizione, per non risultare di distrazione. Al soggetto è stata lasciata la libertà di accenderlo nelle pause tra una sessione e l'altra, così da poter visualizzare una schermata di countdown alla sessione successiva. Prima di iniziare, tutti i soggetti sono stati sottoposti ad una breve sessione di training, svolta in modo analogo alla sessione sperimentale vera e propria.

Tutto il codice dell'esperimento è stato implementato in ambiente MATLAB.

3.2 Stimoli

Lo stimolo utilizzato come segnale emesso dalla sorgente è un rumore bianco uniformemente distribuito, della durata di 400 ms, sagomato con rampe lineari di 30 ms sia in capo che in coda. Studi condotti [14] hanno dimostrato come suoni aventi la stessa energia in tutte le bande di frequenza, come il rumore bianco, siano più facilmente localizzabili rispetto a suoni aventi un contenuto spettrale sparso. Inoltre, questa scelta permette un rapido confronto coi risultati ottenuti in lavori precedenti [2] sulla localizzazione della distanza. L'ampiezza media del segnale all'ingresso del canale uditivo è di 60 dB.

I suoni spazializzati sono stati creati filtrando lo stimolo emesso dalla sorgente attraverso la coppia di HRTF nel campo lontano ed applicando il metodo DVF, con raggio della testa a fissato al valore standard 8.75 cm.

La sorgente sonora virtuale è stata simulata sul piano orizzontale a due valori azimutali, espressi di seguito in un sistema di coordinate polari: L (laterale), ottenuto come randomizzazione bilanciata di stimoli presentati lungo le direzioni laterali destra (θ = 90°) e sinistra (θ = 270°) e M (mediano) con θ = 180°. La scelta di considerare solo il piano orizzontale è stata fatta in seguito alle osservazioni secondo cui la stima della distanza varia in modo più significativo con l'azimut che con l'elevazione [15], mentre il fatto di considerare solo la direzione mediana posteriore è giustificata dal numero potenzialmente elevato di front/back reversal dovuta all'uso di HRTF non individuali e per evitare possibili associazioni con riferimenti visivi.

Per ogni direzione si sono scelti 3 valori di riferimento di distanza: 25, 50 e 100 cm (rispettivamente near field (N), half-way (H) e soglia per il far field (F)).

Al soggetto sono state presentate coppie di suoni composte da uno stimolo ad una distanza di riferimento (es. 50 cm) ed uno a distanza inferiore (es. 40 cm), proposti in due ordini: avvicinamento (es. 50-40 cm) indicato di seguito con A (approaching), ed allontanamento (es. 40-50), indicato con R (receding), con 500 ms di pausa tra i due suoni.

3.3 Protocollo

La combinazione di 3 distanze di riferimento (N, H, F), 2 azimut (L, M) e 2 ordini di presentazione (A, R) definisce un totale di 12 condizioni. Per ogni condizione, ci si è posti l'obiettivo di determinare la soglia individuale percentuale di discriminazione dei due stimoli. L'esperimento è stato suddiviso in blocchi da un massimo di 200 trial ciascuno, in cui ogni coppia di stimoli è stata determinata scegliendo ad ogni trial una tra le 12 sequenze attive in maniera pseudocasuale. Ogni blocco è stato separato dal successivo da una pausa di 3 minuti.

L'esperimento si è svolto fornendo in cuffia al soggetto le indicazioni necessarie attraverso una guida vocale creata con un software Text-To-Speech (TTS). Un processo a risposta forzata richiedeva al soggetto di inserire tramite tastiera una risposta sull'ordine di presentazione della coppia di stimoli: la freccia rossa verso il basso per indicare che il secondo suono era stato percepito più vicino rispetto al primo (avvicinamento, A), la freccia blu verso l'alto se il secondo suono era stato percepito più lontano rispetto al primo (allontanamento, R). La guida vocale segnalava inoltre l'inizio e la fine di ogni blocco, invitando il soggetto a concedersi una breve pausa. La durata media totale dell'esperimento è stata intorno ai 45 minuti.

Nella creazione della coppia di stimoli, la distanza inferiore tra le due è stata ottenuta decrementando percentualmente la distanza di riferimento corrispondente. In particolare, per ogni condizione la prima coppia di stimoli è stata ottenuta con un decremento del 20% della distanza di riferimento. I trial seguenti sono stati creati muovendo adattivamente la sorgente virtuale di distanza inferiore a passi dell'1% dalla distanza di riferimento, seguendo un algoritmo 1-down 1-up fino al quinto errore di risposta del soggetto, 2-down 1-up per i trial successivi [16]. Per esempio, fissata la distanza di riferimento 25 cm, il secondo stimolo parte da 20 cm e si sposta a passi di 0.25 cm, avvicinandosi al riferimento se la riposta del soggetto sull'ordine di somministrazione degli stimoli ("la sorgente virtuale si avvicina o si allontana?") è esatta, allontanandosi altrimenti, fino al quinto errore. Dal quinto errore in poi, lo stimolo si avvicina al riferimento solo se la riposta viene data correttamente due volte di seguito, continuando invece ad allontanarsi ad ogni singolo errore. Il test su ogni condizione continua fino alla 20ma inversione di ordine nella risposta. Le soglie di discriminazione individuali sono state poi calcolate mediando le differenze percentuali corrispondenti alle inversioni 6-20. Nel caso in cui la distanza adattativa abbia raggiunto quella di riferimento (0% di differenza), il task si ferma. La soglia viene posta uguale a zero solo se il soggetto ha precedentemente fatto meno di 6 errori, altrimenti è ancora una volta la differenza mediata tra il 6° errore e l'ultimo.

4. Risultati

In figura 1 sono mostrate media e SD delle soglie percentuali calcolate a partire dai risultati dei soggetti.

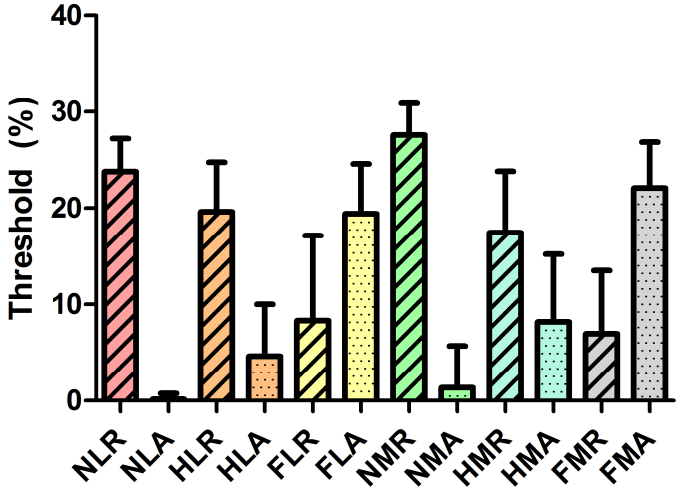

Figura 1 – Media e SD (%) delle soglie di discriminazione della distanza relativa

La soglia percentuale media si pone intorno al 13%. Analizzando la condizione di azimut mediano alla distanza di 1 m, la soglia che si ottiene come media pesata tra le condizioni di allontanamento (FMR) e avvicinamento (FMA) è di 14.51%, con SD = 5.67%. I valori più elevati trovati, rispetto a quelli stabiliti da Ashmead in [2] (media = 5.73%, SD = 2.26%), sono da ricondursi all'uso di sorgenti virtuali, in contrapposizione alle reali.

Si osservano, inoltre, soglie leggermente più alte nelle coppie presentate in allontanamento, se confrontate con le corrispondenti nel campo opposto in avvicinamento (si osservino ad esempio le coppie NLR-FLA, oppure FMR-NMA), come trovato in [6].

Non si osservano differenze particolarmente rilevanti per quanto riguarda le soglie per azimut laterali rispetto al mediano, eccetto che nelle condizioni di campo vicino, in cui gli stimoli laterali hanno soglie più basse (confronto a coppie NLR-NMR, NLA-NMA).

Passando alla domanda di ricerca, fissati la condizione di distanza (N, H o L) e l'azimut della sorgente virtuale (L oppure M) e confrontando le condizioni di avvicinamento/allontanamento, dalla figura si osservano chiaramente trend che confermano l'ipotesi iniziale: le soglie di discriminazione risultano più basse nel campo vicino per coppie in avvicinamento, nel campo lontano per coppie in allontanamento.

L'errore significativamente più alto nel campo vicino per sorgenti in allontanamento conferma quanto osservato da Simpson e Stanton [17] in seguito ad esperimenti con sorgenti reali poste nel campo vicino. In quel caso, gli autori diedero una spiegazione psicoacustica corrispondente alla teoria del looming spaziale, secondo la quale, in condizioni di privazione di indizi ambientali sulla distanza, la variazione della dimensione di uno stimolo (visivo o acustico) percepito vicino all'osservatore crea un'aspettativa di ulteriore avvicinamento. Simpson e Stanton non trovarono tuttavia un trend opposto per quanto riguarda distanze maggiori.

L'asimmetria negli andamenti delle soglie riscontrata tramite questo esperimento trova una probabile spiegazione nella capacità di discriminare due stimoli basandosi sull'informazione di intensità sonora percepita. Come proposto da Olsen e Stevens [18], per coppie discrete di stimoli presentate ad alta intensità sonora (70-90 dB, valori corrispondenti alle nostre condizioni di campo vicino N) il cambiamento di intensità percepita è significativamente più alto quando la coppia è presentata con livello sonoro cre-

scente (corrispondente al nostro avvicinamento A) piuttosto che con livello sonoro decrescente (corrispondente all'allontanamento R). Viceversa, e con andamento opposto, variazioni di intensità nella regione a minore intensità sonora (50-70 dB) vengono meglio percepite se decrescenti (FMR e FLR), piuttosto che se crescenti (FMA e FLA). Finora questi andamenti caratteristici erano stati osservati in suoni naturali, come il suono di un violino [18] o segnali armonici [19], senza trovare corrispondenza in stimoli di rumore bianco.

5. Conclusioni

I risultati dell'esperimento confermano la presenza di trend opposti nella discriminazione della distanza relativa di coppie di suoni creati con il metodo DVF, dipendenti dall'ordine di presentazione degli stimoli e dalla distanza della sorgente virtuale dall'ascoltatore. Tali andamenti sono probabilmente legati non direttamente alla posizione della sorgente nello spazio, quanto all'intensità dello stimolo. Per quanto riguarda i valori delle soglie lievemente più alti per le coppie in allontanamento, confrontate con le corrispondenti nel campo opposto in avvicinamento, si ipotizza anche in questo caso una correlazione col livello di presentazione dello stimolo.

Tale studio conferma i risultati di Spagnol *et al.* [6] ed integra la validazione del metodo DVF di Kan *et al.* [8], basandosi su giudizi di localizzazione relativi (in contrapposizione agli assoluti di Kan) per definire la percezione dell'informazione di distanza nel campo vicino e utilizzando HRTF generiche (piuttosto che individuali) nel campo lontano.

Al fine di indagare più a fondo gli effetti percettivi evidenziati, sarebbe opportuno ripetere l'esperimento presentando gli stimoli a diverse intensità di riferimento. Si potrebbe inoltre riproporre la procedura utilizzando sorgenti reali, per determinare eventuali differenze dovute alla virtualizzazione.

6. Ringraziamenti

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

7. Bibliografia

[1] Brungart D. S., Simpson B. D., *Auditory localization of nearby sources in a virtual audio display*, in Proc. IEEE Work. Appl. Signal Process., Audio, Acoust., New Paltz, New York, USA, October 2001, pp. 107–110

[2] Ashmead D. H., LeRoy D., Odom R. D., *Perception of the relative distances of nearby sound sources*, Percept. Psychophys., vol. **47**, no. 4, April 1990, pp. 326–331

[3] Zahorik P., Brungart D. S., Bronkhorst A. W., *Auditory distance perception in humans: a summary of past and present research*, Acta Acustica united with Acustica, vol. **91**, no. 3, May/June 2005, pp. 409–420

[4] Qu T., Xiao Z., Gong M., Huang Y., Li X., Wu X., *Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap*, IEEE Trans. Audio, Speech, Lang. Process., vol. **17**, no. 6, August 2009, pp. 1124–1132

[5] Parseihian G., Jouffrais C., Katz B. F. G., *Reaching nearby sources: Comparison between real and virtual sound and visual targets*, Front. Neurosci., vol. **8**, September 2014, pp. 1–13

[6] Spagnol S., Avanzini F., *Distance rendering and perception of nearby virtual*

sound sources with a near-field filter model, submitted for publication, 2015

[7] Spagnol S., Avanzini F., *Real-time binaural audio rendering in the near field*, in Proc. 6th Int. Conf. Sound and Music Computing (SMC09), Porto, Portugal, July 2009, pp. 201–206

[8] Kan A., Jin C., Van Schaik A., *A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function*, J. Acoust. Soc. Am., vol. **125**, no. 4 , April 2009, pp. 2233–2242

[9] Wenzel E. M., Arruda M., Kistler D. J., Wightman F. L., *Localization using non-individualized head-related transfer functions*, J. Acoust. Soc. Am., vol. **94**, no. 1, July 1993, pp. 111–123

[10] Møller H., Sørensen M. F., Jensen C. B., Hammershøi D., *Binaural technique: Do we need individual recordings?*, J. Audio Eng. Soc., vol. 44, no. **6**, June 1996, pp. 451–469

[11] Plenge G., *On the differences between localization and lateralization*, J. Acoust. Soc. Am., vol. **56**, no. 3, September 1974, pp. 944–951

[12] Zahorik P., *Distance localization using nonindividualized head-related transfer functions*, J. Acoust. Soc. Am., vol. **108**, no. 5, November 2000, p. 2597

[13] Lindau A., Brinkmann F., *Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings*, J. Audio Eng. Soc., vol. 60, no. 1/2, January 2012, pp. 54–62

[14] Doukhan D., Sédès A., *CW_binaural~: A binaural synthesis external for Pure Data*, in Proc. 3rd Puredata Int. Conv. (PdCon09), São Paulo, Brazil, July 2009

[15] Brungart D. S., Durlach N. I., Rabinowitz W. M., *Auditory localization of nearby sources. II. Localization of a broadband source*, J. Acoust. Soc. Am., vol. **106**, no. 4, October 1999, pp. 1956–1968

[16] Levitt H., *Transformed up-down methods in psychoacoustics*, Journal of the Acoustical Society of America, **49** (1971), pp. 467-477

[17] Simpson W. E., Stanton L. D., *Head movement does not facilitate perception of the distance of a source of sound*, Am. J. Psych., vol. **86**, no. 1, March 1973, pp. 151–159

[18] Olsen K. N., Stevens C. J., *Perceptual overestimation of rising intensity: is stimulus continuity necessary?* Perception, vol. **39**, no. 5, May 2010, pp. 695–704

[19] Neuhoff J. G., *Perceptual bias for rising tones*, Nature, vol. **395**, no. 6698, 1998, pp. 123-124