Effects of stimulus order on auditory distance discrimination of virtual nearby sound sources

Simone Spagnol[a)]
Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, 107 Reykjavík, Iceland
spagnols@hi.is

Rebekka Hoffmann and Árni Kristjánsson
Faculty of Psychology, University of Iceland, 107 Reykjavík, Iceland
rebekkah@hi.is, ak@hi.is

Federico Avanzini
Department of Information Engineering, University of Padova, 35131 Padova, Italy
avanzini@dei.unipd.it

Abstract: Stimulus order has been reported to affect perceived loudness. This letter investigates how temporal order affects distance discrimination of receding and approaching pairs of sound sources rendered binaurally in the anechoic near-field. Individual discrimination thresholds for different virtual locations were measured through an adaptive procedure. The threshold values show a bias toward approaching stimuli for closer reference distances (≤50 cm) and toward receding stimuli for farther reference distances (100 cm), but only when absolute intensity cues are available. The results show how an illusion of loudness can translate into an illusion of perceived relative distance.

[QJF]
Date Received: September 23, 2016 **Date Accepted:** March 23, 2017

1. Introduction

There is considerable evidence for an asymmetry in hearing, leading human listeners to overestimate increasing compared to equivalent decreasing sound intensity. Humans may have evolved to perceive increasing sound intensity as changing more than equivalent decreasing intensity, to selectively prepare for the sound source's arrival.[1] Accordingly, neurobiological studies of auditory looming demonstrate anisotropic neural processing of acoustic intensity change.[2] Increasing, but not decreasing, sound intensity activates a distributed cortical network, concerned with space perception and the allocation of sensory attentional resources, which is likely to provide an adaptive advantage.[3]

However, Olsen and Stevens[4] found that when the absolute reference intensity level is low (50–70 dB) perceived loudness change in pairs of discrete sound stimuli is significantly higher when the pair is presented in decreasing rather than increasing order of intensity, as opposed to high reference intensity levels (70–90 dB) where perceived loudness change is—in accordance with the auditory looming principle—higher for increasing intensity pairs. Given the importance of sound intensity as a distance cue,[5] this may affect distance perception, leading to higher sensitivity to receding sound sources at a reasonable distance from the listener as opposed to approaching sound sources.

Accordingly, in this study we look for asymmetries in relative distance perception of virtual stimuli receding or approaching in the listener's anechoic near-field at six reference locations, given by the combination of 3 distances (25, 50, 100 cm) and 2 directions (lateral or medial). The aim is to assess the salience of the intensity cue as opposed to other near-field cues such as the monaural low-pass effect (available at closer distances only) and the binaural frequency-dependent Interaural Level Difference (ILD, available for lateral sources only and increasing with decreasing distance).[6] Our hypothesis is that listeners rely on absolute intensity as a major distance cue—rather than other near-field cues—to selectively tune in on approaching close sources rather than receding close sources, and on receding far sources rather than approaching far sources.

[a)]Author to whom correspondence should be addressed.

2. Methods

2.1 Sample

Forty participants (14 F, 26 M), evenly split into two groups (7 F and 13 M per group), participated on a voluntary basis. Ages ranged from 21 to 60 (mean = 29.9, standard deviation = 8.5). All participants had normal hearing defined as thresholds no higher than 23 dB hearing level (HL) in the range of 125 Hz to 8 kHz according to a custom audiometric screening.

2.2 Stimuli

Even though there are notable challenges in reproducing spatial auditory cues in the near-field,[7] virtual sound sources can be reliably rendered by applying an ILD correction to a set of far-field head-related transfer functions (HRTFs). Here we use the Distance Variation Function (DVF) method,[8] which codes ILD into a function that takes into account the pressure ratio between a near-field and the corresponding isodirectional far-field sound source observed on the surface of a rigid sphere. Experimental results suggest that the DVF method is effective for conveying both absolute[8] and relative[9] distance information even when intensity cues are unavailable, especially for close ranges (<40 cm).

The reference far-field virtual auditory display came from the PKU&IOA HRTF database,[10] which includes KEMAR HRTFs measured in the far-field (distance $r_{ff} = 1.6$ m from the center of the manikin's head). As in some previous studies,[11,12] non-individual HRTFs were chosen as the far-field display in order to simulate a feasible scenario for practical applications where individual HRTFs are typically not available. Although non-individual HRTFs can cause localization errors, it has been shown that distance estimation does not significantly change when switching from an individual HRTF to a non-individual HRTF.[13]

The sounds were 400-ms white noise bursts with 30-ms onset and offset linear ramps. Spatial sounds were then created by filtering the sound source signal through a pair of near-field HRTFs obtained with the DVF method. The head radius parameter was fixed to the standard value of 8.75 cm.

The virtual sound source was rendered in the horizontal plane either straight to the left or right, balancing among conditions (lateral source), or straight behind the head (medial source). We opted to render sounds only from behind in order to avoid the potentially significant number of front/back reversals that occur with non-individual HRTFs.[11] For each direction, we fixed three reference distance values at 25, 50, and 100 cm from the center of the head, thus ranging from the nearest field to the far-field limit. Having fixed one of the two directions (lateral or medial) and one of the three reference distance values, stimuli corresponding to the reference distance (e.g., 50 cm) and a lower distance (e.g., 40 cm) were presented in sequence to the participant in either approaching order (e.g., 50–40 cm) or receding order (e.g., 40–50 cm), with a delay of 500 ms between the two stimuli.

Stimuli for the two groups of participants differed in that while absolute intensity at the virtual source was fixed, making absolute intensity cues available to one group (FIX group), the reference stimulus level was roved for the other group (ROVE group). In the first case, the measured sound pressure level (SPL) at the earcup was around 60 dB for the reference medial stimulus at 50 cm, and was fixed throughout the whole experimental session. This reference stimulus level was preserved for the ROVE group, but a level rove procedure was implemented in order to ensure that participants could not use absolute intensity information. In order to achieve this, the stimuli in each pair were jointly normalized on the reference distance r according to an inverse $1/r$ distance attenuation rule and subsequently presented at a level chosen from a uniform distribution ranging over ± 10 dB.

Figure 1 shows the spectral effects introduced by the DVF method depending on the angle of incidence between the virtual sound source and the ear and for each reference distance, as well as for distances 20% closer, with overall level effects normalized over distance. Notice that for lateral sources the ILD (considered as the difference between ipsilateral and contralateral ear spectra) decreases non-linearly with distance, suggesting that better sensitivity to distance changes based on ILD is expected for closer reference distances. Also notice the increasing low-pass effect with decreasing distance for medial sources, which suggests that spectral effects may be used as a monaural distance cue in the proximal region even in the absence of ILD variations.[6]

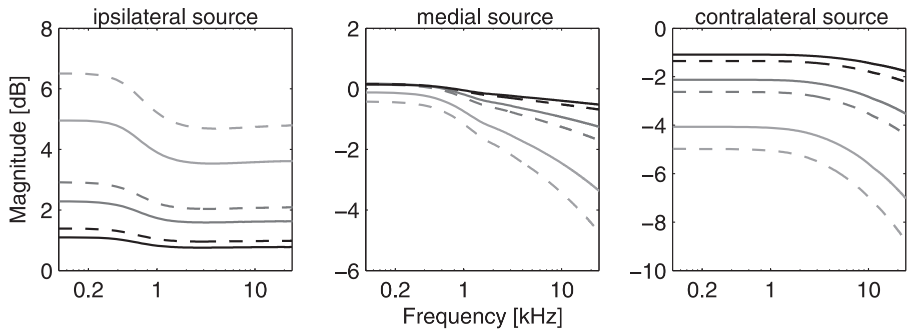

Fig. 1. Spectral modifications introduced by the DVF function for ipsilateral, medial, and contralateral sources, for each reference distance (solid lines) and distances 20% closer (dashed lines). Reference distance values are 25 cm (light gray lines), 50 cm (dark gray lines), and 100 cm (black lines).

2.3 Procedure

The experimental sessions took place inside a silent room. The participant sat on a chair in front of a table that was equipped with a keyboard. The up and down arrow keys were colored blue and red, respectively. The participant wore a pair of Sennheiser HDA 200 headphones (Sennheiser electronic GmbH & Co. KG, Wedemark-Wennebostel, Germany) plugged to a PreSonus AudioBox USB audio card (PreSonus Audio Electronics, Inc., Baton Rouge, LA). A generic digital compensation filter was used to compensate for the headphone response.

The combination of 3 reference distances, 2 directions, and 2 orders resulted in 12 different experimental conditions per participant. The lateral direction (left or right) was fixed within each condition and participant, with 3 conditions left and 3 conditions right randomly assigned. The goal was to determine the individual discrimination threshold of the two stimuli in each condition within 12 independent sequences. The adaptive procedure, based on the algorithm proposed by Ashmead *et al.*,[5] ran as follows. Having fixed one of the 12 conditions, the initial adaptive (lower) distance in the first stimulus pair of the sequence was chosen by reducing the reference distance by 20%. The stimulus pair was played and the participant reported whether the second stimulus was perceived nearer or farther than the first, by pressing the red or blue key, respectively. The following pairs in the sequence were determined by moving the adaptive distance point in 1% steps with respect to the reference distance according to a *1-down, 1-up* algorithm up to the fifth reversal (i.e., incorrect answer), and a *2-down, 1-up* algorithm for the following trials. For instance, if the reference distance was 50 cm and the order was approaching, the second stimulus in the pair was set at 40 cm first and subsequently moved in 0.5 cm steps, reducing the separation if the participant perceived the correct order of presentation (i.e., if the red key was pressed) and increasing the separation otherwise (i.e., if the blue key was pressed), up to the fifth reversal. From then onwards, the separation was reduced if two correct answers were given in a row, but kept increasing at each single reversal. Each sequence ended either at the 20th reversal or when the adaptive distance reached the reference distance (0% difference).

In order to minimize recency effects, the order of presentation of stimulus pairs to the participant followed random permutations of the still active conditions. More precisely, the first 12 trials were the first stimulus pairs of the 12 sequences in random order, then the following 12 trials were the second stimulus pairs in a different random order, and so on. When one of the sequences ended, the following permutation did not include the associated condition. However, if less than three conditions remained, dummy trials were randomly interleaved. A mandatory break of 3 min followed every 200 trials. The average total duration of the experiment was roughly 45 min, and the average number of trials per participant was 630. No feedback on answer accuracy was provided.

2.4 Data analysis

Individual discrimination thresholds were computed by averaging the differences (expressed as the percentage of the reference distance) between the two distances corresponding to the last ten reversals. If the sequence ended because the adaptive distance reached the reference distance, the threshold was set to zero.

Since the hypothesis of homogeneity of variance between the two groups of participants was rejected (according to Levene's test), we ran two separate factorial analyses of variance (ANOVA) on the discrimination threshold data, one for the FIX group and one for the ROVE group. Data normality and sphericity were verified using

Table 1. Results of the factorial 3-way analyses of variance. Significant main factors and interactions at the $p < 0.01$ level are reported in bold.

ANOVA	Factor(s)	*F*-value	*p*-value	partial η^2
FIX group	**Order**	$F(1, 19) = 49.06$	$p < 0.001$	$\eta_p^2 = 0.72$
	Direction	$F(1, 19) = 4.32$	$p = 0.051$	$\eta_p^2 = 0.18$
	Reference distance	$F(2, 38) = 3.97$	$p = 0.027$	$\eta_p^2 = 0.17$
	Order*Direction	$F(1, 19) = 5.47$	$p = 0.030$	$\eta_p^2 = 0.22$
	Order*Reference distance	$F(2, 38) = 105.2$	$p < 0.001$	$\eta_p^2 = 0.85$
	Direction*Reference distance	$F(2, 38) = 4.62$	$p = 0.016$	$\eta_p^2 = 0.20$
	Order*Direction*Ref distance	$F(2, 38) = 12.15$	$p < 0.001$	$\eta_p^2 = 0.39$
ROVE group	Order	$F(1, 19) = 3.01$	$p = 0.099$	$\eta_p^2 = 0.14$
	Direction	$F(1, 19) = 10.05$	$p = 0.005$	$\eta_p^2 = 0.35$
	Reference distance	$F(1, 25) = 0.89$	$p = 0.383$	$\eta_p^2 = 0.04$
	Order*Direction	$F(1, 19) = 18.22$	$p < 0.001$	$\eta_p^2 = 0.49$
	Order*Reference distance	$F(2, 38) = 3.56$	$p = 0.038$	$\eta_p^2 = 0.16$
	Direction*Reference distance	$F(2, 38) = 14.93$	$p < 0.001$	$\eta_p^2 = 0.44$
	Order*Direction*Ref distance	$F(2, 38) = 11.08$	$p < 0.001$	$\eta_p^2 = 0.37$

the Shapiro-Wilk test and Mauchly's test, respectively. In the single case where the sphericity assumption was violated (reference distance in ROVE group), degrees of freedom were adjusted using a Greenhouse-Geisser epsilon correction. The significance level for all data analyses was set to $p < 0.01$.

3. Results and discussion

The results of the two ANOVAs are reported in Table 1, while Fig. 2 shows mean discrimination thresholds divided by group, direction, reference distance, and order of

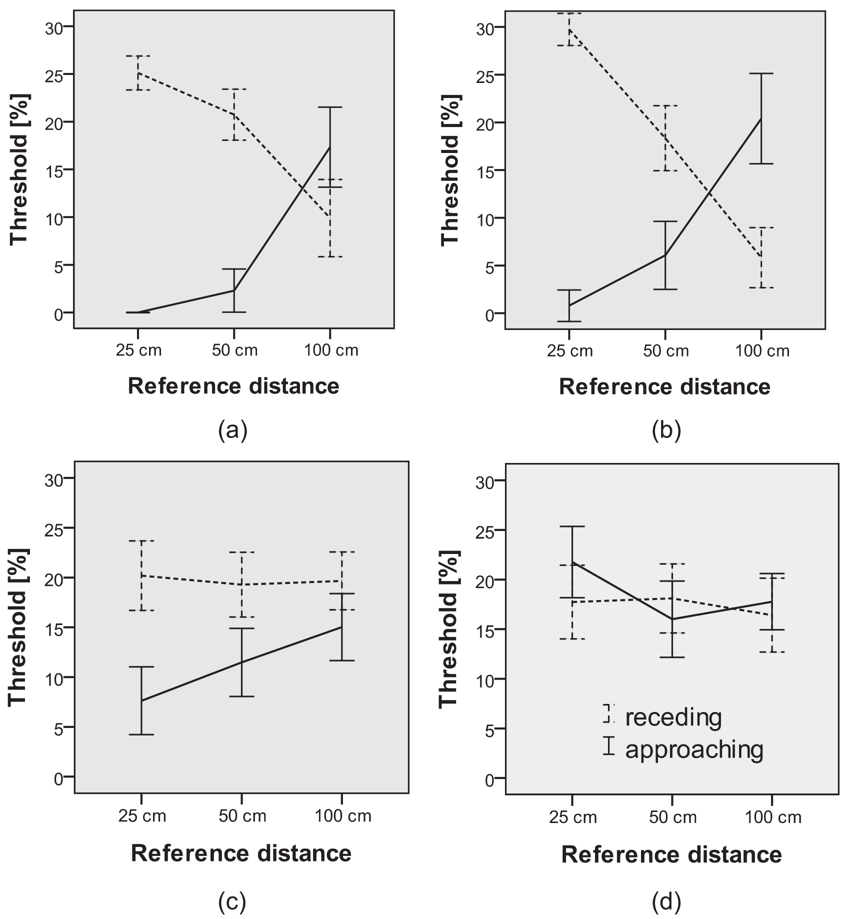

Fig. 2. Mean discrimination thresholds divided by group, direction, reference distance, and order of presentation. Solid lines refer to approaching stimuli, dashed lines to receding stimuli. Error bars represent the within-subjects standard error of the mean. (a) FIX group, lateral direction; (b) FIX group, medial direction; (c) ROVE group, lateral direction; (d) ROVE group, medial direction.

presentation. Apart from a common 3-way interaction, the two groups exhibit very different, almost specular result patterns.

First, there was a highly significant interaction between reference distance and order of presentation for the FIX group ($F(2, 38) = 105.2$, $p < 0.001$, $\eta_p^2 = 0.85$). This result confirms the hypothesis of an asymmetric bias in the perception of receding and approaching stimuli. Figures 2(a) and 2(b) report average discrimination thresholds that are much lower for approaching stimuli than for receding stimuli in the near-field, and lower for receding stimuli than for approaching stimuli toward the far-field. This considerable bias is reflected in the finding that the majority of participants in the FIX group, at the closest reference distance, (1) reached the zero threshold in the approaching conditions and (2) had a threshold well above the starting level of 20% in the receding conditions.

The significantly higher thresholds for receding stimuli at close distances can be interpreted in line with the evolutionary adaptive thesis that the salience of approaching sounds over receding sounds may enlarge the margins of safety for preparatory behaviors to an approaching sound source.[1] On the other hand, the same interaction is not significant for the ROVE group ($F(2, 38) = 3.56$, $p = 0.038$, $\eta_p^2 = 0.16$). This suggests that information about absolute intensity is responsible for the asymmetric bias. This finding is most likely related to the previously mentioned results on perception of loudness change by Olsen and Stevens,[4] and supports our initial hypothesis.

A further difference between the results of the two groups is that the interaction between direction and stimulus order is non-significant for the FIX group ($F(1, 19) = 5.47$, $p = 0.030$, $\eta_p^2 = 0.22$) but highly significant for the ROVE group ($F(1, 19) = 18.22$, $p < 0.001$, $\eta_p^2 = 0.49$). Indeed, Figs. 2(c) and 2(d) show significantly lower thresholds for lateral sources than medial sources in the approaching order. We hypothesize that this effect reflects ambiguous information given by near-field spectral cues in the case of medial sources, where ILDs are weakest. An increase of low frequencies relative to high frequencies can be a signal of both an approaching near source and a receding far source.[14] Thus, when the listener has no absolute references, spectral cues (see again Fig. 1, middle panel) may play a predominant role, causing the intensity-related bias toward approaching sources to disappear.

In addition, as previously suggested,[7] the availability of reliable ILD information as an absolute distance cue in the case of lateral sources may account for the above difference, as well as for the significant main effect of direction ($F(1, 19) = 10.05$, $p = 0.005$, $\eta_p^2 = 0.35$). Also related is the loss of significance of the main effect of order when reference intensity information is not available ($F(1, 19) = 3.01$, $p = 0.099$, $\eta_p^2 = 0.14$). Thresholds for the approaching condition are significantly lower than those for the receding condition in the FIX group, but not the ROVE group.

Taken together, these results both confirm previous findings on auditory distance perception and highlight the salience and importance of the intensity cue in near-field virtual auditory displays. In particular, we show how a previously observed asymmetry in loudness perception can reveal a strong bias in the perceived relative distance of virtual stimuli. Additionally, the observed perceptual asymmetries are striking considering that biases in loudness change of rising level stimuli were found to be much milder with white noise rather than tones or harmonic stimuli.[1,15] It is important to note, however, that even though the participants were instructed to provide relative distance judgments, it is still possible that loudness judgments played some role. Future work will investigate this possibility as well as the case of real sound sources, with the aim of evaluating whether the observed effects are due to limitations of near-field virtual auditory displays or can be generalized to real world settings.

Acknowledgments

The authors would like to thank Professor Massimo Grassi for helpful advices, Erica Tavazzi for her help with the conduction of the experiment, and all the participants. This project was supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 643636.

References and links

[1]J. G. Neuhoff, "An adaptive bias in the perception of looming auditory motion," Ecol. Psychol. **13**, 87–110 (2001).

[2]D. A. Hall and D. R. Moore, "Auditory neuroscience: The salience of looming sounds," Curr. Biol. **13**, R91–R93 (2003).

[3]E. Seifritz, J. G. Neuhoff, D. Bilecen, K. Scheffler, H. Mustovic, H. Schächinger, R. Elefante, and F. Di Salle, "Neural processing of auditory looming in the human brain," Curr. Biol. **12**, 2147–2151 (2002).

[4]K. N. Olsen and C. J. Stevens, "Perceptual overestimation of rising intensity: Is stimulus continuity necessary?," Perception **39**, 695–704 (2010).

[5]D. H. Ashmead, D. LeRoy, and R. D. Odom, "Perception of the relative distances of nearby sound sources," Percept. Psychophys. **47**, 326–331 (1990).
[6]D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479 (1999).
[7]D. S. Brungart, "Near-field virtual audio displays," Presence **11**, 93–106 (2002).
[8]A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," J. Acoust. Soc. Am. **125**, 2233–2242 (2009).
[9]S. Spagnol, E. Tavazzi, and F. Avanzini, "Distance rendering and perception of nearby virtual sound sources with a near-field filter model," Appl. Acoust. **115**, 61–73 (2017).
[10]T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu, "Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap," IEEE Trans. Audio, Speech, Lang. Process. **17**, 1124–1132 (2009).
[11]D. S. Brungart and B. D. Simpson, "Auditory localization of nearby sources in a virtual audio display," in *Proceedings IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics* (New Paltz, New York, 2001), pp. 107–110.
[12]G. Parseihian, C. Jouffrais, and B. F. G. Katz, "Reaching nearby sources: Comparison between real and virtual sound and visual targets," Front. Neurosci. **8**, 269 (2014).
[13]P. Zahorik, "Distance localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **108**, 2597 (2000).
[14]P. D. Coleman, "Dual role of frequency spectrum in determination of auditory distance," J. Acoust. Soc. Am. **44**, 631–632 (1968).
[15]J. G. Neuhoff, "Perceptual bias for rising tones," Nature **395**, 123–124 (1998).