Auditory distance perception for sounds around the listener

Gabriela Andrejková and Norbert Kopčo
P. J. Šafárik University in Košice, Faculty of Science
Jesenná 5, 04001 Košice
gabriela.andrejkova@upjs.sk, norbert.kopco@upjs.sk

Abstract

The accuracy of auditory distance perception is very different in the region near the listener (≤1 meter) from the distal region (>1 meter). Santarelli et al. (1999) performed an experiment in a reverberant room in which listeners were asked to point to the perceived 3-dimensional (3-D) position of sound sources presented from a random location in the right hemifield within 1 m of the listener's head.

Here, a new analysis focuses on the distance localization performance in these data. It evaluates the mean absolute error (MAE) and standard deviation (SD) in responses as a function of source location varying simultaneously in all three dimensions. Results provide a characterization of 3-D localization performance for nearby sources in a reverberant room (is characterized by high energy density of reflected sound), showing that MAE is significantly different for stimuli (sound sources) directly above the listener compared to stimuli originating in front, behind, or below the listener. SD is highest for stimuli bellow the listener, and it is dependent on distance, polar angle, and lateral angle. Distance judgments are less accurate for medial stimuli than for lateral stimuli at larger distances. The analysis of temporal profile during the whole experiment showed a significant decrease in MAE but not in SD. The decrease in MAE occurred mainly in the second half of the responses, indicating that spontaneous learning has occurred on a slow time on hours.

1 Introduction

The ability to determine the distance of sounds is important for a person's orientation in the environment and for interpretation of these sounds. There are many various acoustical and non-acoustical factors thought to contribute to source distance perception. Zahorik et al. (2005) demonstrated that people use multiple cues to estimate distance, especially the Direct to Reverberant energy Ratio (DRR), systematically underestimate distant sources and these abilities emerge in early childhood.

For sounds near to the listener, Kopčo and Shinn-Cunningham (2011) showed that distance judgments were dominated by a reverberation related cue (commonly interpreted as DRR), which listeners mapped to distance in a direction independent but frequency dependent manner. Brungart and Durlach (1999) evaluated 3-D localization of sound sources near the head in anechoic room (walls absorb sound). They found that in an anechoic space distance localization performance in the proximal field is better than at larger distances, but at the same time it is strongly dependent on azimuth - that is, on whether the source is in a lateral position or in the medial plane.

Santarelli et al. (1999) implemented nearly identical study in a reverberant room and he showed that the presence of reverberation does indeed have a substantial effect on localization ability. The important finding of this work is that different cues underlie the distance perception of nearby sounds in anechoic and reverberant space.

Both studies focused on response analysis separately in each space dimension. Sound localization is typically examined separately in the three spatial dimensions (azimuth, elevation, distance), or for combinations of two dimensions (e.g., azimuth and elevation Best et al. (2011), or azimuth and distance, Ihlefeld and Shinn-Cunningham (1999). Estimating multiple dimensions simultaneously can be more challenging than for fewer dimensions. On the other hand, when location covaries in two dimensions, perception can improve over 1-dimensional baseline (example of visual distance perception Loomis et al. (1998).

Very few studies looked at 3-D localization. Kopčo and Andrejková (2025) examined biases in distance perception for nearby sources varying location in polar and lateral angles in reverberant room. For sources in the horizontal plane the results don't vary dramatically, with an overall underestimation (approximately -10% of presented stimuli distance) that tends to increase for nearby lateral sources (approximately -20 %) and appears to be stronger in front than behind the listeners.

Here, we evaluate distance localization performance of Santarelli et al. (1999) data in a perceptually based coordinate system, in which the space was divided into 9 directional bins based on lateral angles (bin centered at 15°, 45°, 75°), polar angles (bins centered at 0°, 90°, 180°, 270°). Each directional bin is divided into 5 log-distance bins (centered at 20, 28, 41, 58, 83 cm). We use MAE as the main accuracy metric and SD as the reliability metric. MAE assesses the overall localization error, while SD expresses the variability of the responses.

Together, the two metrics provide a robust assessment of accuracy and consistency in the near field, where acoustic parameters cause nonlinear and significant variations.

2 Material and methodology

2.1 Original data

The experiment was done by Santarelli et al. (1999) in the room (9x5x3-m (length x width x height) located within the Department of Cognitive and Neural Systems at Boston University. The testing apparatus, shown in Fig. 1, was prepared according to description of Brungart and Durlach (1999). Seven listeners participated in the experiment. The listener sat on a wooden chair placed in the center of the room. The chair was equipped with an adjustable headrest to help prevent the listener from making head movements during data collection. The experimenter stood to the right of the listener with the testing apparatus. The transmitter for the electromagnetic tracker was placed to the rear of the listener, slightly to the left, at mid-back level.

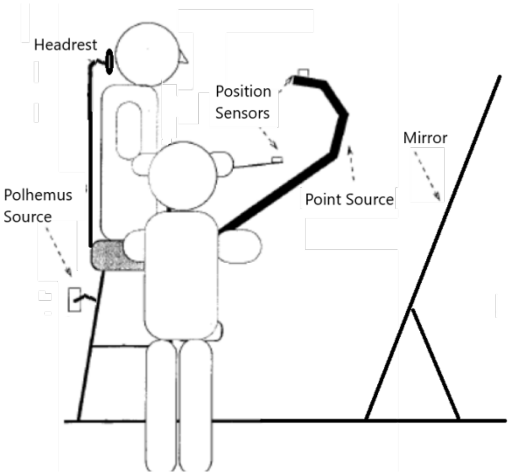

Fig. 1: Modified setup according to (Brungart and Durlach (1999))

2.2 Procedure

At the beginning of each trial, a computer-generated voice instructed the listener to ”listen,” providing the listener with a cue to close his eyes. The computer then randomly chose a spatial location in the listener’s right hemisphere by specifying the appropriate azimuth, elevation, and distance. Next, the experimenter pressed a switch on the source-wand, triggering the presentation of the stimulus. The experimenter promptly returned the source to the starting position and pressed the switch again. The listener opened his eyes and used the response wand to indicate the position at which he perceived the source. With a final press of the switch, this position was recorded, marking the end of the trial.

Stimuli consisted of five pink-noise bursts (150 ms duration, 30 ms inter-stimulus interval, or ISI). During data collection, the trials were partitioned into blocks that consisted of either 25 or 50 trials. Each listener was given a 5–10-minute rest interval between blocks of trials. A typical session lasted roughly 1 to 1.5 hours and included a total of 100 to 150 trials. Each listener was required to complete a total of 1200 trials (the first 200 of which were practice trials), which translates to approximately 12 hours of testing.

2.3 Methodology

We evaluated the results using the interaural polar coordinate system which allowed us to track the dependence of the distance responses simultaneously on the lateral angle (LA) and polar angles (PA), as well as on source distance. The data were binned into 9 directional bins by dividing them into five distances. The directional bins were combinations of lateral angle (3 regular intervals centered at LA = [15°, 45°, 75°]) and polar angle (one bin for LA $>$ 75°, 4 bins centered at PA = [0°, 90°, 180°, 270°] for LA in range of [15°, 45°]). We evaluated Mean Absolute Error (MAE) and Standard Deviations (SD) using a log-log scale (log10(response distance) – log10(stimulus distance)). Log–log scales are used in hearing research because both sound and human perception operate on ratios and power laws, not linear differences. Repeated measures ANOVA were used to analyze the data with Box-Geissler-Greenhouse epsilon used to correct for potential violations of the sphericity assumption.

3 Results

3.1 MAE and SD

We averaged data within each directional and distance bin for each listener and then collapsed across listeners. In the upper panels of Fig. 2, the MAE is shown as a function of log-distance, the polar angle is distinguished by color and a colored icon. The largest deviations are observable in stimuli in bins with a centered polar angle of 270° (magenta), the stimuli below listeners. For LA=15°, near the medial plane, MAE increases with distance for stimuli above the listener. In other directions (front, back, and below) MAE is affected by lateral angle only. Listeners make larger errors in smaller lateral angles, and with increasing lateral angle these errors don’t increase.

A three-way ANOVA with factors polar angle PA = [0°, 90°, 180°, 270°], lateral angle LA = [15°, 45°]

and distance D = [20, 28, 41, 58, 83] cm showed a significant main effect LA, F(1,6) = 35.32, p = 0.001, indicating that MAE for LA = 15° has higher values than for LA = 45°. At the same time, a significant interaction PA × D was shown, F(12,72) = 3.49, p = 0.0004. Subsequent simple effects analyses revealed that distance D is mostly increasing with distance at PA = 90° (red line).

In evaluation SD, 3-way ANOVA showed main effects PA, F(3,18) = 12.47, $p<0.001$, LA, F(1, 6) = 17.36, p = 0.003, and D, F(4,24) = 5.52, p = 0.003. SD is affected by distance, polar and lateral angle independently of each other.

After averaging across polar angles, 2-way ANOVA with factors LA = [15°, 45°, 75°] and D = [20, 28, 41, 58, 83] cm showed main effect LA, F(2,12) = 64.16, $p<0.001$.

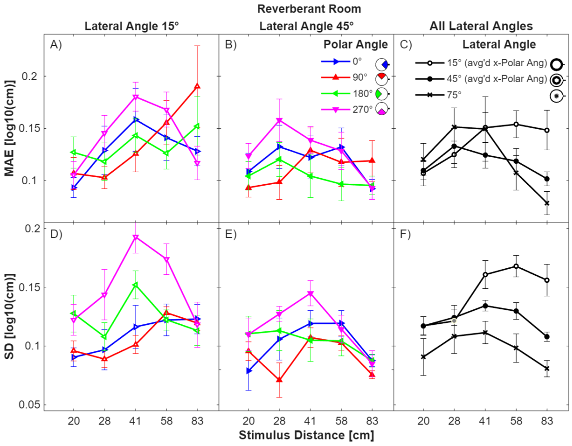

Fig. 2: Across-listener mean of Absolute Error (MAE) and Standard Deviations (SD) in five distance and nine directional bins. The upper panels A) - B) show MAE in two lateral angels and four polar angles (recognized by colors and icons). Panel C) shows the MAE for all three lateral angles after averaging over polar angles. In the lower panels, SDs are drawn in a similar way.

Evaluation MAE and SD in panels C) and F) shows the largest differences for lateral angle LA = 15°. This means the presence of outliers and smaller stability of responses. Listeners' responses to stimuli in this area are the most imprecise and variable.

3.2 Temporal profile

Each listener performed approximately 1200 trials in the same reverberant room. The first 200 trials are considered training stimuli, and they weren't used in an evaluation of temporal profile. In rooms, reverberation provides distance information and candidate cue is Direct-to-Reverberant energy ratio, Bronkhorst and Houtgast (1999). Shinn-Cunningham (2000) showed that in rooms, there exists a learning effect: distance performance is improved with experience.

We analyzed learning process followed by MAE and SD through all evaluated trials. All data of each listener was divided into 5 temporal blocks per 200 trials in chronological order. The sequence of these blocks follows time sequence of trials. For each listener and each trial, the MAE and SD was calculated and averaged across stimuli in each temporal block. The time course of the averaged values also represents the course of adaptation to the stimuli.

In Fig. 3, panel A), the courses of all listeners are shown (colored dotted lines), as well as their average across listeners (black solid line). We assumed some trend in decreasing MAE, a linear approximation of MAE is depicted in panel B), for all listeners and mean across listeners (black solid line). SDs are shown in panels D) and E). One-way ANOVA with factor the first and the last values of approximating line V (V1, V2) showed main effect V F(1,6) = 6.05, p = 0.049. This means that the MAE is decreasing and the decrease in values is significant.

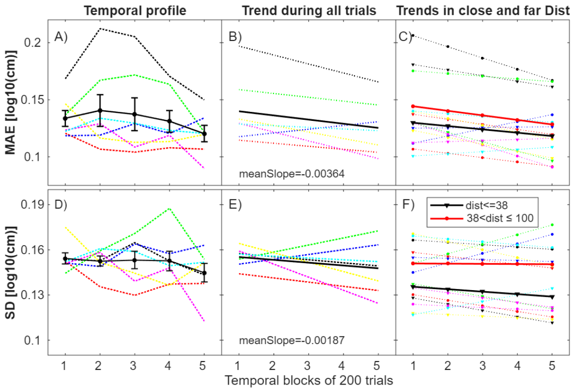

Fig. 3: Temporal Profile (TP) of MAE and SD. The first column of panels A) and D) shows profiles of MAE and SD of all listeners (colored dotted lines) and the mean across the listeners (black solid line) in 5 temporal blocks of 200 averaged trials. The panels B) and D) show linear regressions of temporal profiles. In the panels C) and F), trends of listeners (dotted lines) and averaged across listeners in two log-distances are shown.

Further analysis of a temporal profile in each of two log-distances intervals, division in 38 cm, is shown in panel C). Averaged values across stimuli and across listeners in each log-distance interval are drawn in five temporal blocks. It is shown that MAE doesn't increase in any log-distance interval, in most of them MAE decreases. Two-way ANOVA with factors distance D ($\leq$ 38, $>$ 38) and begin and end values of approximating lines V (V1, V2) showed main effect V, F(1,6) = 6.11, p = 0.048. It means that the MAE is decreasing and the decreasing in values is significant at V, but there is no significant effect at D. Taken together, these results indicate that listeners continue to improve on the task in both distances even after 200 practice trials.

4 Discussion

The current study examined MAE and SD in distance perception for nearby sources varying location in polar and lateral angles in the reverberant room. For frontal sources close to the horizontal plane (polar angle from -45° to +45°) the results don't vary dramatically (blue lines), the highest value of MAE is in bin centered by 41 cm for frontal sources (panel A). For sources above the listener's head (from 45° to 135°), we observe an increasing MAE with distance for frontal sources. The largest MAE values are for sources below the listener's head at almost all distances. In the regions of lateral angles (0° to 30°), (30° to 60°) and (60° to 90°) we observe an increasing in MAE at small distances, at larger distances the MAE decreases or almost does not change, panel C). The SD is largest for sources below the listener's head and larger for frontal sources, panel D). At larger distances, we observe that SD is almost the same for all polar angles, panels D) and E). Panel F) shows the largest variability for frontal sources.

Although Brungart and Durlach (1999) investigated the localization of very close sources in anechoic space, his results provide a generally valid explanation of the psychophysical and acoustic limitations that affect the accuracy of distance estimation. Brungart shows that for sources up to about 1 m, there are sharp nonlinear changes in binaural and spectral cues, especially the interaural level difference (ILD), while the interaural time difference (ITD) remains relatively constant. These rapid changes in acoustic cues cause impaired distance estimation accuracy and increased response variability already in anechoic environments.

In a reverberant room, these known limitations are compounded by the contribution of DRR, which is a significant distance cue for nearby sources. Santarelli et al. (1999) and later Kopčo and Andrejková (2025) show that reverberation can improve or worsen the availability of distance information depending on the direction of the source, with systematic upward and downward biases being typical of the combination of binaural and reverberant cues.

The resulting MAE and SD values in our data thus represent a natural consequence of these known mechanisms. MAE reflects a combination of the nonlinear changes in ILD, and spectral properties described by Brungart, supplemented by the distribution of reverberation in the room. SD results from the instability of acoustic cues with small changes in source distance and orientation, which are even more pronounced in a reverberant environment. Overall, MAE and SD are therefore consistent with Brungart's theory of local acoustic limitations for very close sounds, as well as with the results of Santarelli et al. (1999) in reverberant conditions.

Our results confirm previous findings Shinn-Cunningham (2000) that room learning occurs in reverberant environments, leading to reduced localization errors. The improvement in MAE and SD during the experiment is therefore an expected and theoretically well-supported consequence of adaptation to the acoustic environment.

Acknowledgement

Work supported by EU HORIZON-MSCA-2022-SE-01 101129903, APVV-23-0054, SK-AT-23-0002.

References

Best, V., Carlile, S., Kopčo, N. a van Schaik, A. (2011). Localization in speech mixtures by listeners with hearing loss. *The Journal of the Acoustical Society of America*, 129(5):210–215.

Bronkhorst, A. W. a Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 39(11):517–520.

Brungart, D. S. a Durlach, N. I. (1999). Auditory localization of nearby sources. ii. localization of a broadband source in the near field. *The Journal of the Acoustical Society of America*, 106(4):1956–1968.

Ihlefeld, A. a Shinn-Cunningham, B. G. (1999). Effect of source spectrum on sound localization in an everyday reverberant room. *The Journal of the Acoustical Society of America*, 130(1):324–333.

Kopčo, N. a Andrejková, G. (2025). Effect of stimulus spectrum on distance perception for nearby sources. *Proc. Forum Acusticum 2025*, str. 3393–3397.

Kopčo, N. a Shinn-Cunningham, B. G. (2011). Effect of stimulus spectrum on distance perception for nearby sources. *The Journal of the Acoustical Society of America*, 130(3):1530–1541.

Loomis, J. M., Klatzky, R. L., Philbeck, J. W. a Golledge, R. G. (1998). Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6):966–980.

Santarelli, S., Kopčo, N., Shinn-Cunningham, B. G. a Brungart, D. S. (1999). Near-field localization in echoic rooms. *The Journal of the Acoustical Society of America*, 105(2):1024.

Shinn-Cunningham, B. G. (2000). Learning reverberation: Implications for spatial auditory displays. *Proc. International Conference on Auditory Displays*, str. 126–134.

Zahorik, P., Brungart, D. S. a Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica United with Acustica,*, str. 409–420.