The influence of visual cues on auditory distance perception

Jens Cubick[1], Sébastien Santurette[1], Søren Laugesen[2], Torsten Dau[1]

[1] *Centre for Applied Hearing Research, DTU (DK), Email: jecu@elektro.dtu.dk*
[2] *Eriksholm Research Centre (DK)*

Introduction

In many situations, human listeners depend on their sense of hearing to orient themselves in their surroundings, which requires the localization of a multitude of sound sources. Correct localization of a sound source in space not only requires an estimate of the angle, but also of the distance to the sound source. In natural listening situations, listeners usually perceive the auditory event roughly at the position of the corresponding sound source, at a distance and outside their head, i.e., externalized. In the literature, some conflicting results on externalization were reported related to the effect of the frequency content of the stimuli. In an early study [1], the term externalization was used synonymously with distance perception and the authors found that trains of short high-pass filtered noise bursts were perceived closer than the broadband version, whereas lowpass-filtered stimuli were perceived at a greater distance. Later studies also described reduced high-frequency content of a sound as one of the acoustic cues for longer distance to the sound source (e.g., in [2]). Two recent studies that focused explicitly on externalization, however, found that lowpass-filtering actually *decreased* the amount of externalization [3, 4].

Another recent study [5] investigated if the reduction of externalization reported by [3] for lowpass-filtered signals could also be found in a distance perception experiment. The results, however, showed no dependence of the perceived distance on the stimulus bandwidth. Furthermore, the measured distance curves were also very different from those typically reported in the literature (e.g., [6]), even though the measurement technique for the binaural room impulse responses (BRIRs) was very similar and the measured distances were identical. The main difference between [5] and [6] was that the experiment in [6] was conducted in a listening booth, whereas it was conducted in the same room where the individual BRIRs had been recorded in [5]. This and some evidence from [7], where similar differences in distance judgements were found between an experiment conducted in the dark and the same experiment with a distance scale provided by four pairs of LEDs at 2, 4, 6, and 8 m, suggested that the *visual* impression might have a major effect on auditory distance perception.

In the present study, the experiments from [5] were repeated in a double-walled listening booth with 7 of the 10 listeners who had participated in the original experiment from [5] to test the hypothesis, that the difference in distance perception found in [5] as compared to the behaviour typically reported in the literature (e.g.,

[6]) might be explained by the visual impression of the room in which the experiment was conducted. In the following, the methods and main findings from [5] will be briefly summarized and supplemented with the data from the current experiments in the listening booth. The results will be compared and discussed with regard to three questions. 1) Does the bandwidth of the stimulus have an influence on the perceived distance of a sound? 2) Does the environment in which the experiment was conducted have an influence on the perceived distance? 3) How are the distance and externalization percepts related?

Methods

In this study, the same individual BRIRs were used that had been measured for [5] at nine log-spaced egocentric distances (0.43, 0.61, 0.86, 1.22, 1.72, 2.44, 3.45, 4.88 and 6.9 m) at an angle of 25° in a workshop of about 12.65 x 6.75 x 3.10 m with an acoustic ceiling and an average reverberation time T_{30} of about 0.6 s (see Fig. 1 for a photograph of the room). During the measurement, the listeners were blindfolded, seated in a listening chair and provided a small headrest to help keeping the position of the head fixed. The BRIRs were measured at the entrance of the open ear canal with DPA 4060 lapel microphones attached to the pinna with a wire hook, using six repetitions of a 5 s logarithmic sine sweep and a deconvolution method according to [8]. The headphones used in the experiment were equalized with an inverse filter generated from the headphone impulse response measured with 10 repetitions of a 2 s sine sweep.

Figure 1: Photograph of the listening experiment in the workshop room with visual markers at 2, 4, 6, and 8 m.

To generate the stimuli, a random sentence from the Danish HINT speech test corpus [9] was convolved with all nine measured BRIRs and the inverse of the headphone response for each experimental run. The resulting aural-

ized signals were band-limited between 50 and 15000 Hz with 6th order Butterworth filters (Broadband condition). Apart from the broadband condition, two conditions were tested with lowpass-filtered stimuli, either with a cutoff-frequency of 6 kHz to simulate the limited bandwidth of a hearing aid or with a cutoff-frequency of 2 kHz to simulate a typical age-related high-frequency hearing loss. Both lowpass-filters were realized as 32 tap Hamming-window based FIR filters. Furthermore, a diotic condition was tested in which the broadband signal for the right ear was presented to both ears. Therefore, no binaural cues were available in this condition, and even though there was some spectral shaping by the BRIR, the spatial information was inconsistent, because the same IR was applied to both ears. This condition was added to test if the listeners actually externalized the sounds. In that case, it was expected that they would judge the diotic stimuli as being perceived inside the head. Each condition (Broadband, LP 2 kHz, LP 6 kHz, Diotic) was repeated six times, resulting in 24 experimental runs with 9 stimuli each. All signals were played back through a RME Babyface USB sound card and Sennheiser HD 800 headphones.

In the experiment, the listeners responded via a modified MUSHRA (ITU-R BS.1534-1) user interface on a computer screen. Stimuli were generated for each of the nine measured distances and randomly assigned to a playback button and the corresponding slider. Listeners were instructed to listen to all stimuli and judge the distance at which they perceived the auditory event with the slider on an absolute scale in metres. All stimuli that they perceived *inside* the head should be rated as zero (as in [6]). In the workshop room, the scale in the user interface corresponded to visual markers provided in the room at distances of 2, 4, 6, and 8 m (cf. Fig. 1). In the listening booth, no reference scale was given. Once all distance judgements were found, the listener confirmed and the next set of stimuli was loaded. All stimuli within one run had the same bandwidth. The listeners were trained with four runs including all four conditions. Completing the whole experiment took about one hour. A photograph of the experimental setup inside the listening booth is shown in Fig. 2. Both the stimulus generation and the instruction of the listeners were identical to the experiment run in the workshop.

Results

Fig. 3 shows the average distance rating and standard deviation of the perceived distance of 10 listeners measured in the workshop as presented in [5] in the broadband condition (right-pointing triangle), and for the stimuli that were lowpass-filtered at 2 kHz (left-pointing triangle) and 6 kHz (upward-pointing triangle). In contrast to expectations based on the previous literature mentioned above, the different stimulus bandwidths did not result in any systematic changes of the perceived distance. In general, the nearest auralized distances up to about 1 m were perceived closer than the veridical value (indicated by the light-grey, dash-dotted line). At intermediate dis-

Figure 2: The listening experiments were repeated in the listening booth. The presented stimuli were identical to the ones used in the previous experiment in the workshop room.

tances, the distances were generally overestimated, while the farthest distance (6.9 m) was frequently underestimated again. These results are clearly in contrast to the average results from [6], where listeners typically overestimated the distance of nearby sources and progressively underestimated distances beyond about 1.5 m (indicated by the grey, dashed line).
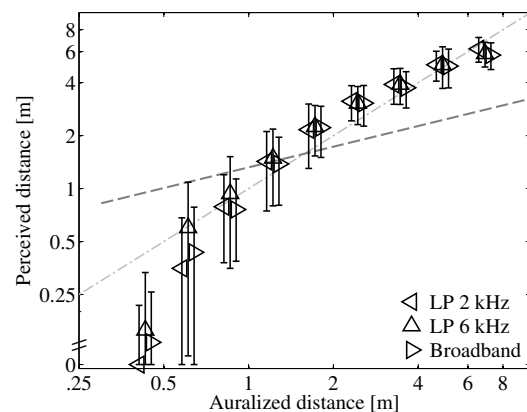
Figure 3: Average value and standard deviation of the perceived distances measured with 10 normal-hearing listeners in the workshop room. The grey dashed line indicates the average perceived distance found in [6].

Fig. 4 shows the average perceived distances for 7 listeners measured in the listening booth (see photograph in Fig. 2). Again, the three different bandwidths of the stimuli did not result in systematically different distance judgements. The closest auralized distances were, on average, perceived farther away compared to the results obtained in the workshop room, the farthest distances were perceived slightly closer. The distance functions thus are more compressive, i.e., they have a shallower slope in the log-log representation than the ones measured in the workshop. This brings them somewhat closer to the results presented in [6] (indicated by the grey dashed line), even though the perceived distance range in the present study was still much wider, and the 'auditory horizon' reported in other studies was not found here. Only for the two farthest distances, the distance functions start

to saturate and the judgements became smaller than the veridical values. Interestingly, the closest distances in both experiments were frequently rated as being inside the head, whereas no zero-responses were reported in [6].
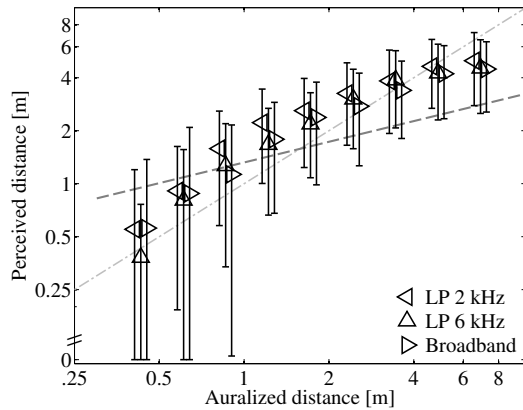

Figure 4: Average value and standard deviation of the perceived distances measured with 7 normal-hearing listeners in the listening booth. The grey dashed line indicates the average perceived distance reported in [6].

In the diotic condition, the listeners could essentially be divided into two groups. Individual data for two extreme examples are shown in Fig. 5. While the overall distance ratings were fairly comparable between the two listeners, they completely disagreed in their judgement of the diotic condition. One listener (top panel) consistently rated the distance of all diotic stimuli as 0 m, i.e., 'inside my head', whereas the presence or absence of binaural information seemed to make no difference for the distance judgement of the other one (bottom panel). The average perceived distance in the diotic condition was similar to all other conditions, except for the closest distance, where it was perceived closer than the binaural stimuli.

The distance ratings were also considerably less reliable in the booth, as indicated by the larger standard deviation in Fig. 4. Comparing the raw data of the two experiments reveals that the larger errors found in the results from the booth are due to both larger intra- and intersubject variability. While some datasets obtained in the booth showed fairly low errors, the data were still more consistent in the workshop experiment in all cases.

Discussion

The expected effect of stimulus bandwidth on distance perception could not be found in either experiment. This might be due to the response method used in the experiment. Since each response screen of the multi-stimulus test only contained stimuli for one condition, the listeners might have used the entire available scale to make a relative comparison within the condition rather than an absolute judgement that would show the "true" perceptual difference across conditions. Using a multi-stimulus test that allows for repeated comparisons between different stimuli probably reduces the variability of the data at the expense of yielding a more relative judgement as compared to other methods with single stimulus presentation
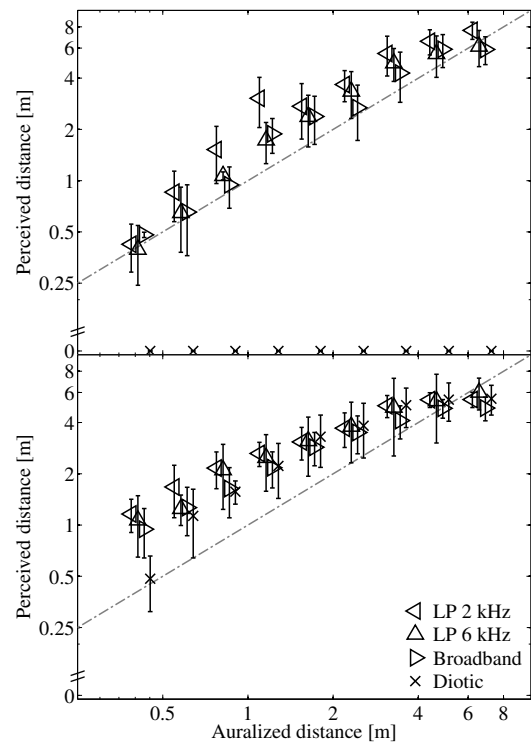
Figure 5: Average individual distance ratings in the listening booth for two listeners. Here, also the diotic condition is shown.

like a direct scaling method. Using an actual MUSHRA test that provides a reference and an anchor might help solve this issue by spanning the range of the percept. However, defining a meaningful reference and anchor in this type of experiment would be difficult. Another approach would be to randomize the conditions over the experimental runs. Presenting stimuli from different conditions within the same run should emphasize the differences between the conditions and thus increase the sensitivity of the test.

With respect to the influence of the visual impression, the comparison between the two experiments might be biased by the fact that all listeners had performed the experiment in the workshop before the booth. This might be the reason why, while closer, the data from the experiment in the booth still look different from those in [6]. Repeating the experiments with naive listeners and counterbalancing the order of the two experiments might show larger differences between the workshop and the listening booth. Even though there was a long time period between the two experiments, some listeners reported that they remembered the setup in the workshop room with the visual distance cues, which they thought helped them making the distance judgement in the booth.

The larger standard deviations measured in the listening booth (cf. Fig. 4) indicate that the listening task is more demanding in the booth than in the workshop room (cf. Fig. 3) and that the listeners are less reliable in their judgement. This suggests that, for a reliable distance judgement, the experiments should preferably be run in the same room where the BRIRs have been measured,

because the incongruence between acoustic and visual impression in the listening booth seems to make the task more difficult. Two listeners also noted that the stimuli sounded more reverberant in the listening booth than in the workshop room, which might be a hint that some unforeseen effects can appear when the stimuli are incongruent with the visual or acoustic impression of the playback room.

Finally, the results for the diotic condition indicate that listeners in the experiment might have utilized different cues for their ratings in the experiment. Some of the listeners clearly rated most diotic stimuli as 'inside the head', which indicates that the percept of externalization, while strongly related to distance perception, seems to require true binaural cues. This is in accordance with [3]. For the other group, the availability of true binaural information did not seem to make a difference in terms of perceived distance. The distance judgement of these listeners might therefore be primarily based on monaural distance cues, like loudness, direct-to reverberant sound ratio, and the stimulus spectrum.

Conclusion

In the presented study, it was found that listeners rated the perceived distance less reliably in a listening booth than in the room where the BRIRs had been measured under otherwise identical experimental conditions. On average, no systematic difference was found for different stimulus bandwidths in neither of the experiments. The range of the distance percept was found to be slightly more compressed in the booth, but still far from what was reported for similar experiments in the literature. This illustrates that further research is required to clarify how to correctly measure auditory distance perception and whether incongruent visual or acoustic cues are the major source of the differences between the distance percepts in different playback rooms.

Furthermore, a clearer distinction is needed between the percepts of distance perception and externalization. To the authors, it seems that externalization is the more "demanding" percept, which breaks down easily when stimuli are modified and cues are missing (cf. [3, 4]). The perceived distance of an auditory event, on the other hand, seems to be much more robust and can be estimated even when the auditory event is not perceived *at* a distance. Even though the diotic condition still contained some spectral information, the fact that some listeners were still able to make consistent and repeatable distance judgements for a mono signal indicates that it would be possible to *estimate* the sound source distance even from a simple recording made with an omnidirectional microphone in a room. Such a recording would, however, most likely not be perceived outside the head. This is consistent with the fact that the three most commonly listed cues for auditory distance perception, loudness, direct-to-reverberant ratio, and the sound spectrum, are essentially monaural cues, whereas externalization seems to require stimuli with true binaural information.

References

[1] Levy, E.T., and Butler, R.A.: Stimulus factors which influence the perceived externalization of sound presented through headphones. Journal of Auditory Research, 18(1) (1978), 41–50.

[2] Zahorik, P., Brungart, D.S., and Bronkhorst, A.W.: Auditory distance perception in humans: A summary of past and present research. Acta Acustica United with Acustica, 91(3) (2005), 409–420.

[3] Boyd, A.W., Whitmer, W.M., Soraghan, J.J. and Akeroyd, M.A.: Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. The Journal of the Acoustical Society of America 131(3) (2012), EL268–EL274

[4] Catic, J., Santurette, S., Buchholz, J.M., Gran, F. and Dau, T.: The effect of interaural-level-difference fluctuations on the externalization of sound. The Journal of the Acoustical Society of America 134(2) (2013), 1232–1241

[5] Cubick, J., Santurette, S., Laugesen, S., and Dau, T.: Influence of High-Frequency Audibility on Distance Perception. Fortschritte der Akustik - DAGA 2014 (2014), 576-577

[6] Zahorik, P.: Assessing auditory distance perception using virtual acoustics. The Journal of the Acoustical Society of America 111(4) (2002), 1832–1846

[7] Calcagno, E.R., Abregu, E.L. and Manuel, C.E.: The role of vision in auditory distance perception. Perception 41(2) (2012), 175–192

[8] Müller, S. and Massarani, P.: Transfer-Function Measurement with Sweeps. Journal of the Audio Engineering Society 49(6) (2001), 443–471

[9] Nielsen, J.B. and Dau, T.: The Danish hearing in noise test. International Journal of Audiology 50(3) (2011), 202–208