PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Effects of target speech distance on auditory spatial attention in noisy environments

Shuichi Sakamoto[(1)], Florent Monasterolo[(1)], César Salvador[(1)], Zhenglie Cui[(1)], and Yôiti Suzuki[(1)]

[(1)]Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, Japan,
{saka@ais.riec, fmonaste@dc, salvador@ais.riec, sai@ais.riec, yoh@riec}.tohoku.ac.jp

Abstract

When people perceive auditory information in noisy environments, auditory spatial attention plays a major role in extracting target information from multiple sound sources. For auditory spatial attention, the effects of distance have been unclear. In this study, we examined the relationship between auditory spatial attention and the distance to the target speech sound. In an experiment, background multi-talker noise was presented 1 m from the listeners. The target sound or distracting speech sound was presented at one of four distances: 0.13, 0.25, 0.5, or 1 m. By manipulating the probability of the target presentation distance, the listener's attention was implicitly focused at a specific distance. Listeners were asked to respond as soon as the target speech sound was perceived. Experimental results clearly revealed that the listeners showed faster responses to the target speech sound at the focused distance than at other distances. Moreover, the shape of spatial selectivity of auditory spatial attention varied based on the focus distance. These results are expected to reflect the effect of peripersonal space.

Keywords: Auditory spatial attention, Distance perception, Head-related transfer functions

1 INTRODUCTION

We can extract a target sound from various mixed sounds even when we are in noisy environments. This remarkable ability is called the "cocktail-party effect [1, 2]." Many researchers have investigated the factors that contribute to the extraction of a target sound. One of the major mechanisms is known as auditory selective attention. Recent studies have reported that auditory selective attention in the spatial domain ("auditory spatial attention") plays a critical role in this phenomenon under complex noisy conditions [3, 4]

Results from many studies have highlighted the benefits of auditory spatial attention such as the directional separation of sound sources and enhancement of sound intelligibility for a particular direction (e.g. [4, 5]). However, how the distance separation contributes to auditory selective attention remains unclear. When the distance between listeners and a sound source is more than approximately 1 m, nearly all physical parameters of the sound except the intensity do not change. This means that people hardly judge the sound source distance that is more than approximately 1 m from them. By contrast, for sounds within a space of less than 1 m, people can judge the distance accurately using various physical parameters [6, 7, 8, 9]. In addition, the space within the 1 m distance corresponds to the peripersonal space (PPS), within which various sensory processes change [10]. Therefore, the contributions of auditory spatial attention vary based on the function of distance. Indeed, by working with virtual sound sources presented from within a 1 m distance, Shinn-Cunningham *et al.* [11] and Brungart & Simpson [12] reported the benefits of the distance separation of sound sources on the speech reception threshold and the importance of interaural differences in these benefits. This suggests the effects of spatial unmasking for source separation along near distances within PPS.

In this study, we investigated the relationship between auditory spatial attention and the distance to the target speech sound. To generate proximal distance sound virtually, head-related transfer functions (HRTF) for various distances were synthesized using distance-varying filters (DVF) [13]. Under the condition when listeners focus on the sound source from a specific distance, the reaction time of the target sound was measured to analyze the effects of auditory spatial attention.

2 METHOD

2.1 Listeners

In the experiment, seven young males (aged 23–24 years, mean age 23.7) with normal hearing participated. All were recruited from the Graduate School of Information Sciences, Tohoku University. Informed consent was obtained from each listener before the experiment. The procedure was approved by the Ethics Committee of the Research Institute of Electrical Communication of Tohoku University.

2.2 Apparatus

The experiment was conducted in a sound-proof room. The sound stimuli were presented through headphones (Sennheiser HDA-200) binaurally. The transfer functions of the headphones were compensated by convolving a 2048-point inverse filter calculated from the headphones' impulse responses.

2.3 Stimuli

All stimuli were virtually spatialized by convolving listeners' near-distance HRTFs. These HRTFs were synthesized by applying distance-varying filters (DVF) [13] to the individual HRTFs measured at 1.5 m with 5° angular resolution. Here, a 512-point DVF was applied to listeners' HRTFs. From the results of the calculation, HRTFs at two directions (front (0°) and left (−90°)), and four distances chosen on a logarithmic scale (0.13 m, 0.25 m, 0.50 m, and 1.00 m) were generated.

The target sound was a 4-mora word chosen from a familiarity-controlled word list (FW07) [14]. As the background sound, six streams of meaningless word sequences were overlapped. In each stream, words were randomly selected from FW03 [15] and sequentially connected. The total length of the background sound was more than 3 min.

2.4 Experimental procedure

Experiment was conducted under the following two conditions: no-focus and focus. In both conditions, the distance from which the target sound was presented was selected as 0.13 m, 0.25 m, 0.50 m, and 1.00 m, whereas the distance to the background sound was fixed at 1.00 m. The intensity cue for both sounds for the distances was eliminated. Both the target and background sounds were always presented from the same direction: 0° or −90°. Between the duration of the two target-sound presentation, one, two, three, or no distracter sounds were presented from one of the four distances. Inter-stimulus interval of the two consecutive sounds was randomly determined at a range of between 750 ms and 1250 ms.

In the no-focus condition, no prior knowledge of the distance from the target sound was provided to the listeners. Moreover, the target sound was presented with equal probability from each distance. The presentation of the target sound from each distance was repeated 12 times. Thus, the total number of target sound presentations was 48: 12 presentations × 4 distances (0.13 m, 0.25 m, 0.50 m, and 1.00 m) × 2 directions (0° and −90°). These two directions were used in separate sessions.

In the focus condition, the listener's auditory attention was implicitly directed to a particular distance using the probe-signal method [16]. As the focus distance, 0.13 m and 1.00 m were selected. These focus distances and directions were used in separate sessions. In each session, 80% of the target sound distance was set at the focus distance, whereas 20% of the distance was selected from three other distances with equal probability (6.66%). In the condition, the target sound was presented 144 times from the focus distance and 12 times from the three other distances based on the probe-signal method. Thus, the total number of target sound presentations was 720: (144 presentations from focus distance + 12 presentations × 3 distances) × 2 focus distances (0.13 m and 1.00 m) × 2 directions (0° and −90°). These distances and directions were selected for separate sessions.

Listeners were instructed to press a response button as soon as they heard the target word.

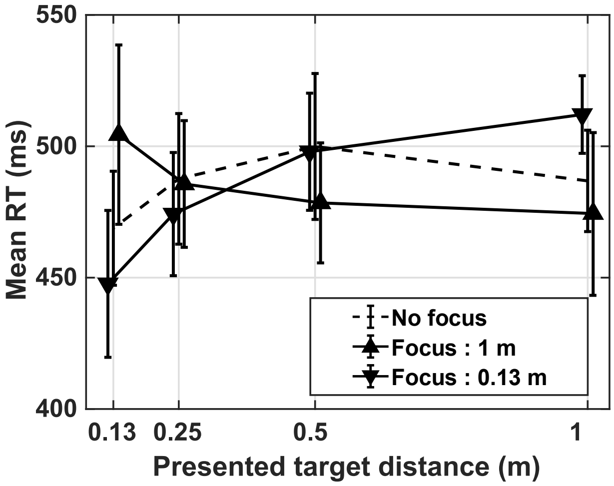

Figure 1. Mean RTs on all azimuths and listeners for conditions with and without focus on the target distance.

3 RESULTS AND DISCUSSION

The average reaction time (RT) for all individual listeners was calculated only for responses when listeners answered correctly until the presentation of the next target or distractor sound. In the focus condition, the target sound was presented from each focus distance 144 times, whereas it was presented only 12 times for the three other distances. To compare average RTs calculated based on the same amount of data, the average RTs for the focus distance were calculated using the last 12 correct responses.

The mean RTs calculated over all listeners are shown in Fig. 1 for the no-focus, focus 1 m, and focus 0.13 m conditions. In the figure, the RTs of two directions are averaged. The error bar gives the standard error. When the target sound was presented from the focus distance, the observed RT was decreased as compared with RT at the same distance in the no-focus condition. By contrast, when the targets were presented from other distances, RTs were longer than when the targets were presented from the focus distance. In the focus 0.13 m condition, the difference between RT at the focus distance (0.13 m) and that at the same distance in the no-focus condition was as much as approximately 20 ms. The increase in RT for targets presented from the farthest distance was approximately 40 ms. For the focus 1 m condition, the decrease in RT was as much as 10 ms, whereas the increase was approximately 60 ms.

A three-way analysis of variance (ANOVA) was performed on mean RT as factors of distances (four target distances), azimuths ($-90°$ and $0°$), and conditions (no-focus, focus 1 m, and focus 0.13 m). The results revealed that the main effects of azimuths ($F(1,6) = 6.15, p < .05$) and distances ($F(3,18) = 4.02, p < .05$) were statistically significant, whereas no conditions ($F(2,12) = 0.10, p = .90$). Interaction between conditions and distances was statistically significant ($F(6,36) = 5.0, p < .001$), but was not between azimuths and conditions ($F(2,12) = 1.32, p = .31$), azimuths and distances ($F(3,18) = 0.46, p = .71$), and three-way interaction ($F(6,36) = 0.44, p = .85$).

The simple main effect of conditions was statistically significant at both focus distances (1 m: $F(2,48) = 3.9, p < .05$, 0.13 m: $F(2,48) = 8.75, p < .001$). The *post-hoc* test (Ryan's method, $p < .05$) revealed that for these two focus distances, the difference in RTs between in the focus 0.13 m condition and in the focus 1 m condition was statistically significant. By contrast, the simple main effect of distances was statistically significant only in the focus 0.13 m condition ($F(3,54) = 10.0, p < .001$), whereas the effect was not statistically significant in other conditions (focus 1 m condition: $F(3,54) = 2.2, p < .1$; no-focus condition: $F(3,54) = 2.06, p = .12$). The results of multiple comparison (Ryan's method, $p < .05$) revealed that for focus 0.13 m condition, the differences between 1 m and 0.13 m, between 1 m and 0.25 m and between 0.5 m and 0.13 m were statistically significant

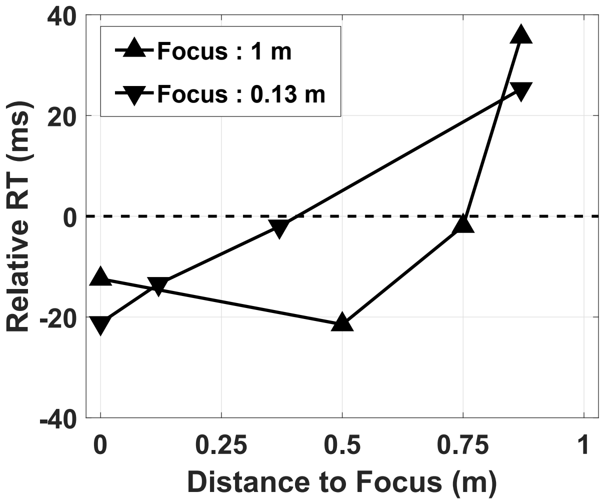

Figure 2. Relative RT as a function of distance from a focus point.

($p < .01$).

The results indicate that auditory spatial attention on a particular distance affected the RTs to the target sound. When the listener's attention was directed to a particular distance, the response to the sounds coming from the directed distance was enhanced. By contrast, the response to sounds coming from a different distance slowed. A similar tendency was observed in studies of auditory spatial attention on direction expressed by the RTs [4, 5, 17, 18].

Figure 2 represents the effect of attention on RT as a function of distance from a focus point. As the target source was presented farther away from the focus distance, the RT increased. However, the shape of the auditory spatial attention for focus 1 m and 0.13 m appears different in Fig. 2. When the focus distance was 1 m, the observed size of the attention was broader than that observed when the focus distance was 0.13 m. This may mean that when listeners directed their attention to their peripersonal space, the sensitivity of the attention at the focus point increased. When a sound was presented at the near-field in the peripersonal space, auditory processes were enhanced. The results suggest that an enhancement of the resolution of the auditory spatial attention might occur for sounds in the peripersonal space.

4 CONCLUSION

In this study, we investigated the relationship between auditory spatial attention and the distance to a target speech sound. Because people can barely judge a sound source from a distance of more than approximately 1 m, the focus distance in this study was set within 1 m. This space corresponds to the peripersonal space and various sensory processes were changed in the PPS. Therefore, the effect of the PPS on the auditory spatial attention was also analyzed. Results from measuring RTs revealed that the sounds presented from the focus distance were consistently responded to at a faster rate than those from non-focused distances. Moreover, the shape of spatial selectivity of auditory spatial attention was different based on the focus distance. These results might reflect the effects of peripersonal space.

ACKNOWLEDGMENTS

This research was partly supported by JSPS KAKENHI No. 17K19990 and 15KK0002.

REFERENCES

[1] Cherry, EC. Some experiments on the recognition of speech with one and two ears, J. Acoust. Soc. Am., 25(5), 1953, 975–979.

[2] Cherry, EC.; Taylor, WK. Some further experiments upon the recognition of speech, with one and with two ears, J. Acoust. Soc. Am., 26(4), 1954, 554–559.

[3] Arbogast, TL.; Kidd, G. Evidence for spatial tuning in informational masking using the probe-signal method, J. Acoust. Soc. Am., 108(4), 2000, 1803–1810.

[4] Kidd, G; Arbogast, TL.; Mason, CR.; Gallun, FJ. The advantage of knowing where to listen, J. Acoust. Soc. Am., 118(6), 2005, 3804–3815.

[5] Ebata, M; Sone, T; Nimura, T. Improvement of hearing ability by directional information, J. Acoust. Soc. Am., 43(2), 1968, 289–297.

[6] Brungart, DS.; Rabinowitz WM. Auditory localization of nearby sources, J. Acoust. Soc. Am., 106, 1999, 1465–1479.

[7] Brungart, DS.; Durlach, NI.; Rabinowitz WM. Auditory localization of nearby sources. ii. localization of a broadband source, J. Acoust. Soc. Am., 106(4), 1999, 1956–1968.

[8] Otani, M.; Hirahara, T. Numerical study on source-distance dependency of head-related transfer functions, J. Acoust. Soc. Am., 125, 2009, 3253–3261.

[9] Kim, H.; Suzuki, Y.; Takane, S.; Sone, T. Control of auditory distance perception based on the auditory parallax model, Appl. Acoust., 62, 2001, 245–270.

[10] Graziano, MSA.; Reiss, LAJ.; Gross, CG. A neuronal representation of the location of nearby sounds, Nature, 397, 1999, 428–430.

[11] Shinn-Cunningham BG.; Schickler, J.; Kopčo, N.; Litovsky, R. Spatial unmasking of nearby speech sources in a simulated anechoic environment, J. Acoust. Soc. Am., 110(2), 2001, 1118–1129.

[12] Brungart, DS.; Simpson BD. The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal, J. Acoust. Soc. Am., 112(2), 2002, 664–676.

[13] Salvador CD.; Sakamoto, S.; Trevino, J.; Suzuki, Y. Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values, Acoust. Sci. & Tech., 38(1), 2017, 1–13.

[14] Kondo, T.; Amano, S.; Sakamoto, S.; Suzuki, Y. Development of familiarity-controlled word-lists (FW07), IEICE Soc. Conf. Res. Report, 107, 2008, 43–48.

[15] Amano, S.; Sakamoto, S.; Kondo, T.; Suzuki, Y. Development of familiarity-controlled word-lists 2003 (FW03) to assess spoken-word intelligibility in Japanese, Speech Communication, 51, 2009, 76–82.

[16] Greenberg, GZ.; Larkin, WD. Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method, J. Acoust. Soc. Am., 44(6), 1968, 1513–1523.

[17] Rhodes, G. Auditory attention and the representation of spatial information, Perception & Psychophysics, 42(1), 1987, 1–14.

[18] Teraoka, R.; Sakamoto, S.; Cui, Z.; Suzuki, Y.; Shioiri, S. Influence of auditory selective attention on word intelligibility, Proc. 2018 Joint Conference - Acoustics, 2018, 307–310.