PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Directional selectivity of auditory spatial attention in multi-talker environment

Ryo TERAOKA[1]; Shuichi SAKAMOTO[1]; Zhenglie CUI[1];
Yôiti SUZUKI[1,2]; Satoshi SHIOIRI[1]

[1] Tohoku University, Japan,

[2] National Institute of Information and Communication Technology (NICT), Japan

ABSTRACT

When people direct their auditory selective attention in a specific direction, the information from that direction can be perceived easily, even in a noisy environment. Understanding how this effect is distributed spatially is important. The extent of spatial distribution may depend on the direction of the attention focus, and thus may differ between the frontal and oblique directions. In the present study, we investigated the dependence of the spatial extent of auditory selective attention on the focus azimuthal direction. To examine the spatial extent, speech intelligibility was measured in a multi-talker environment. In the experiment, a target sound and multiple distracting speech sounds were presented simultaneously from loudspeakers surrounding observers. By manipulating the probability of the target presentation directions, the listener's auditory spatial attention was directed to a specific loudspeaker. The attracted direction was one of three directions: –30, 0, or +30 degrees. The results showed that the greatest improvement in intelligibility was observed at the attracted direction, for all attracted directions. The spatial extent of the auditory selective attention seems identical for all attracted directions. This means that the attention spotlight is not modulated by the direction of the attention.

Keywords: auditory spatial attention, cocktail-party effect, auditory scene analysis, psychoacoustics

1. INTRODUCTION

In real-world listening environments, various sounds are mixed acoustically before reaching our ears. Nevertheless, even in the environment, most normal-hearing listeners can extract a sound of interest from the mixed sounds. This remarkable ability, called the "cocktail-party effect" (1, 2), has been studied for over half a century. Despite its essential importance for both communication and orientation in space, a number of important questions remain.

Among the questions regarding the cocktail-party effect for which we are lacking a satisfactory answer is how human listeners highlight one sound (voice) from background noises. Numerous previous studies have addressed what kind of physical characteristics contributed to the sound source segregation in a "cocktail-party" situation (for review, see (3–6)). Bregman (3) proposed a theory known as "auditory scene analysis". In this theory, a stream segregation involves the grouping together of sounds according to frequency, spatial, and temporal aspects, following Gestalt principles (e.g., sounds that start at the same time are more likely to be coming from the same source). Indeed, several studies have demonstrated that such acoustical characteristics can facilitate segregation of the target sound from background noises. For example, Ebata and his colleagues (7) demonstrated that when a listener hears a target sound in the presence of masker sounds that are spatially separated from the target, the detection threshold of the target sound is lower than in a situation when they are co-located.

Although the physical characteristics of the sounds themselves affect the performance of the segregation, it is also necessary for listeners to actively focus on to the target sounds, which is known as selective attention. Indeed, recent studies have reported that auditory attention in the spatial domain (hereafter "auditory spatial attention") plays a critical role in this ability under a complex listening environment (8–10). Kidd and his colleagues (9) demonstrated that the

[1] terar@dc.tohoku.ac.jp

performance of perceiving the target sound is relatively high when the listener knows *a priori* where the target sound will be presented in multi-source listening environments. Furthermore, we also demonstrated that speech intelligibility in a multi-talker environment is improved when the auditory spatial attention is directed to a specific direction where the target sounds are presented (10). These studies suggest that auditory spatial attention contributes to the occurrence of the cocktail-party effect.

When a human listener directs their spatial selective attention to a specific location, the effects of the spatial attention spread over a range around the focal point, in accordance with the "spotlight" metaphor (11–12). This phenomenon has been well-documented in the visual domain. Indeed, a recent electrophysiological study demonstrated the spatial spread of visual spatial attention (13). In the auditory domain, there are a few pieces of evidence suggesting that auditory spatial attention operates in the manner of a spotlight of attention (10, 14). Our study has demonstrated that listening performance (i.e., speech intelligibility) decreases as the angular distance from the center of attentional focus increases (10). However, previous studies have investigated the spatial spread of auditory spatial attention only in regards to the frontal direction. It is still unclear whether the shape of the spatial spread is identical between the frontal and oblique directions.

In general, the spatial acuity of hearing is particularly good when the sound source is located near 0 degrees azimuth. This performance generally deteriorates as the angular distance moves off the front, and then, this performance is the worst at 90 degrees. Mills (15) measured the resolution of an auditory location with broadband or low-frequency sources by the threshold angle for discriminating the position, called the minimum audible angle (MAA). This study showed that the MAA is always more than 40 degrees at an azimuth of 90 degrees, although the MAA is approximately 1 degree at the front of the listener. These studies suggest that the auditory spatial resolution is not constant, and differs depending on the angle. This tendency may also influence the spatial extent of auditory selective attention, because a human listener must first localize the sound of interest to then direct auditory spatial attention to the specific direction. However, it has still unclear whether the shape of the auditory selective attention is different between the frontal and oblique directions. To investigate this question, we examined the respective widths the spatial extent of auditory selective attention when the listener's attention is attracted to –30, 0, or +30 degrees, and compared these spatial extents.

2. METHODS

2.1 Listeners

Twenty listeners (fifteen males and five females, aged 20–24 years, mean age 21.6) participated in the experiment. All listeners were naïve to purpose of the experiment. All listeners were native Japanese speakers with normal hearing acuity. 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 and stimuli

The experiment was conducted in an anechoic room at the Research Institute of Electrical Communication, Tohoku University. Figure 1 shows the experimental setup. In the anechoic room, five loudspeakers were set circularly on a horizontal plane centered on the listener, in intervals of 30 degrees: – 60, – 30, 0, + 30 and + 60 degrees (a positive value represents the right-hand side of the listener). The radius of the loudspeaker array was 1.6 m.

The speech sounds, comprised of four Japanese moras uttered by male and female speakers, were selected from the "Familiarity-controlled Word lists 2003" (FW03 (16–17)). The target words were selected from the one thousand entries of the FW03 ranked as the highest familiarity. These words were entries of the "Familiarity-controlled Word lists 2007" (FW07 (18–19)), which is a shrunken version of FW03 for clinical use. The total number of the target words was 400 (20 lists, with 20 words per list). The other 600 words were used as distractors. In this study, one female (fhi) and one male (mya) voice were assigned as the target and distractor, respectively.

The A-weighted sound pressure level (L_{Aeq}) of each target and distractor speech was 65 dB at the position of the center of the listener's head. The target and four distractors were simultaneously presented from five loudspeakers.

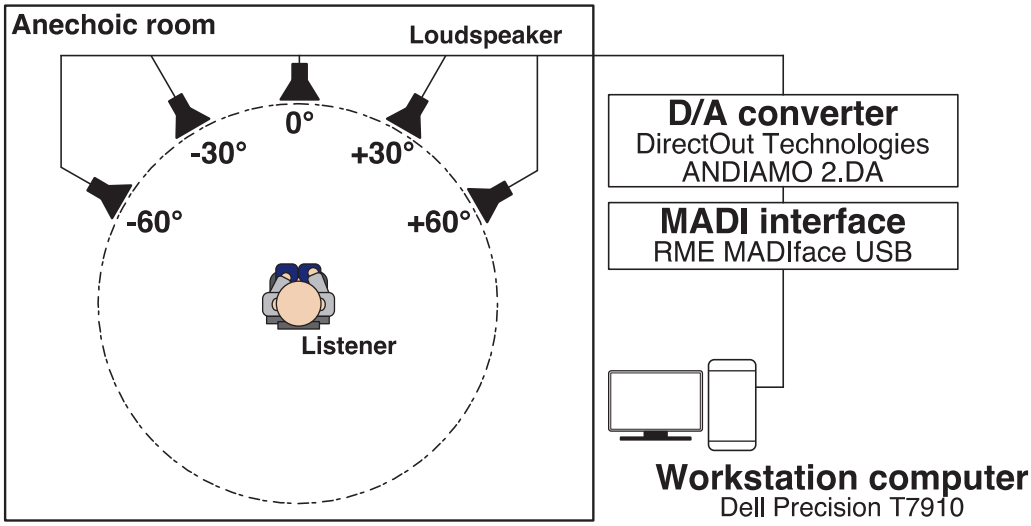

Figure 1 – Schema of the experimental setup. The listener sits on a chair at the center of the loudspeaker array, facing the front (0 degrees).

2.3 Procedure

In the experiment, data was collected for two conditions: cue and probability-control. The order of the two conditions was counterbalanced across the listeners. For both conditions, the listeners were instructed to focus on the target speaker (i.e., female speaker), and to write down the uttered words on a response sheet as soon as they heard the words. Moreover, the listeners were asked to keep their head stationary and straight ahead at 0 degrees during the entire session. One target word, uttered by the female speaker, was presented from one of the five loudspeakers, while four distractors, uttered by the male speaker, were presented from the other four loudspeakers.

In the cue condition, the loudspeaker where the target speech sound would be presented was indicated beforehand. To indicate the direction of this loudspeaker, a burst of 500 ms of white noise (including a 10 ms rise/fall, sound pressure level: 65 dB) was delivered via the loudspeaker, 1000 ms prior to the presentation of the speech sounds. The listeners were asked to direct their attention to the loudspeaker from which the cue was presented. This condition consisted of 100 trials; five lists (i.e., 100 words) were assigned as the target speech sounds, and 400 words were selected from the 600 distractor candidates and assigned as distractors. In one session, the five target lists were separately assigned to five target loudspeakers in the directions of – 60, – 30, 0, + 30, and + 60 degrees. Consequently, 20 target speech sounds were presented from each loudspeaker. The order of the words and the loudspeakers where the target speech sounds were presented were randomized.

In the probability-control condition, the direction was not explicitly indicated to the listeners. In contrast, to implicitly attract attention to a specific loudspeaker, the probe-signal method was applied (20). Among the 400 trials, the target speech sound was presented from the loudspeakers at + 30 degrees (or – 30 degrees) in 80% of the trials (hereafter the “primary angle”). In the remaining 20%, the target speech sound was presented from one of other four loudspeakers, chosen randomly (hereafter the “probe angle”). That is, 16 lists (i.e., 320 words) were presented from the primary angle, whereas four lists (i.e., 80 words) were presented from one of the other four loudspeakers. For half of the listeners, the loudspeaker at – 30 degrees was assigned as the primary angle. For the remaining listeners, the primary angle was + 30 degrees. With this procedure, we could expect that listeners could be aware that most target speech sounds were presented from the loudspeaker at + 30 degrees (or – 30 degrees), resulting in attention directed at the primary angle.

3. RESULTS

Word intelligibility scores are shown in Fig. 2. The scores are plotted as a function of the loudspeaker direction from which the target speech sound is presented. Data points and error bars represent the mean and standard error, respectively, across listeners. Solid squares and open triangles represent the results of the cue and probability-control conditions, respectively. To match

the number of trials in each data point, the results of the probability-control condition at 0 degrees (or ±30 degrees) are the results of the last 20 trials among the 320 trials.

From this figure, it appears that the mean intelligibility scores in the probability-control condition are approximately 15–20% lower than those in the cue condition, in all primary angle conditions. A two-way repeated-measures analysis of variance (ANOVA) was performed on the mean data across all listeners with the condition (cue/probability-control) and target speech sound directions (– 60, –

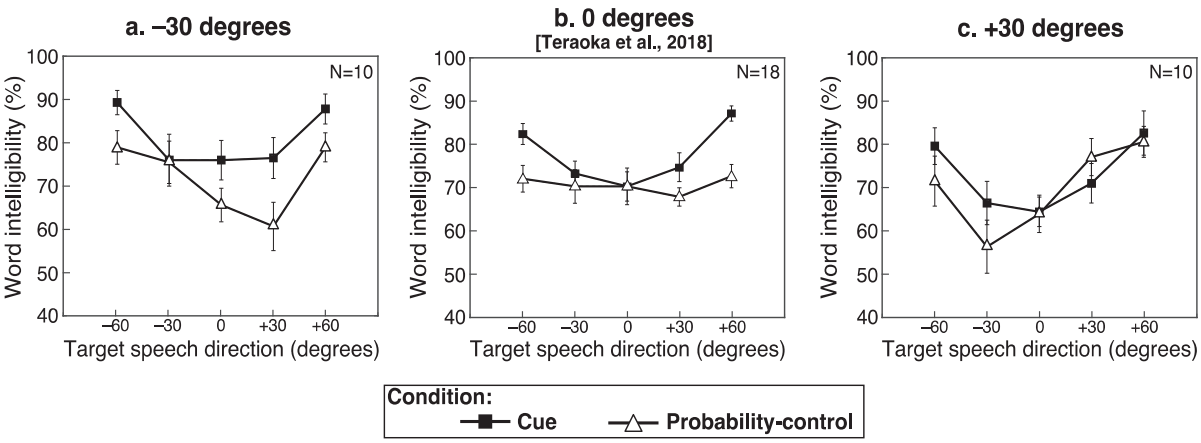

Figure 2 – Word intelligibility score as a function of the direction of target speech sound (**a.** the primary angle is – 30 degrees, **b.** the primary angle is 0 degrees (Adapted from Teraoka et al. (10)), **c.** the primary angle is + 30 degrees). Solid squares and open triangles represent the results of the cue and probability-control conditions, respectively. Error bars denote the standard error of the mean.

30, 0, + 30 and + 60 degrees) as the factors to each primary angle condition. In the – 30 degrees condition, the main effect of conditions is close to significant (conditions: $F_{1,9} = 4.13$, $p = .073$, $\eta^2_G = .107$, target speech directions: $F_{4,36} = 5.31$, $p = .002$, $\eta^2_G = .187$, conditions × target speech sound directions: $F_{4,32} = 1.66$, $p = .180$, $\eta^2_G = .034$). The *post hoc* test (Ryan's method, $p < .05$) revealed there are no significant differences. In the 0 degrees condition, only the main effects are significant (conditions: $F_{1,17} = 6.20$, $p = .023$, $\eta^2_G = .070$, target speech directions: $F_{4,68} = 6.03$, $p = .003$, $\eta^2_G = .082$, conditions × target speech sound directions: $F_{4,68} = 2.24$, $p = .073$, $\eta^2_G = .040$). The *post hoc* test (Ryan's method, $p < .05$) revealed significant differences in direction between + 30 and + 60 degrees, – 60 and + 30 degrees, 0 and + 60 degrees, and – 30 and + 60 degrees. In the + 30 degrees condition, the main effect of conditions is insignificant (conditions: $F_{1,9} = 0.91$, $p = .365$, $\eta^2_G = .001$, target speech directions: $F_{4,36} = 7.95$, $p < .001$, $\eta^2_G = .219$, conditions × target speech sound directions: $F_{4,32} = 1.24$, $p = .031$, $\eta^2_G = .040$). The *post hoc* test (Ryan's method, $p < .05$) revealed significant differences in direction between 0 and + 60 degrees, – 30 and + 30 degrees, – 30 and + 60 degrees, and – 60 and 0 degrees. The results show the word intelligibility in the cue condition is higher than that in the probability-control condition, though these differences are not statistically significant.

4. DISCUSSION

In the present study, we investigated whether the spatial extent of auditory selective attention is different between the frontal and oblique directions. In the cue condition, the word intelligibility score was measured when the auditory spatial attention was explicitly directed to the loudspeaker from which the target speech sound was presented. In contrast, in the probability-control condition, the word intelligibility score was measured when attention was implicitly directed to the primary angles (i.e., + 30 or – 30 degrees) using the probe-signal method. Therefore, when the target speech direction was not the primary angle, the auditory attention in the probability-control condition was not directed to the target speech direction. The results show that the word intelligibility scores in the cue condition are higher (mean=15%) than those of the probability-control condition when the target speech direction is not the primary angle. These results reveal that an improvement of speech intelligibility by auditory spatial attention is observed regardless of the angle of attentional focus.

Notably, the results shown in Fig. 2 represent not only the effect of auditory spatial attention, but

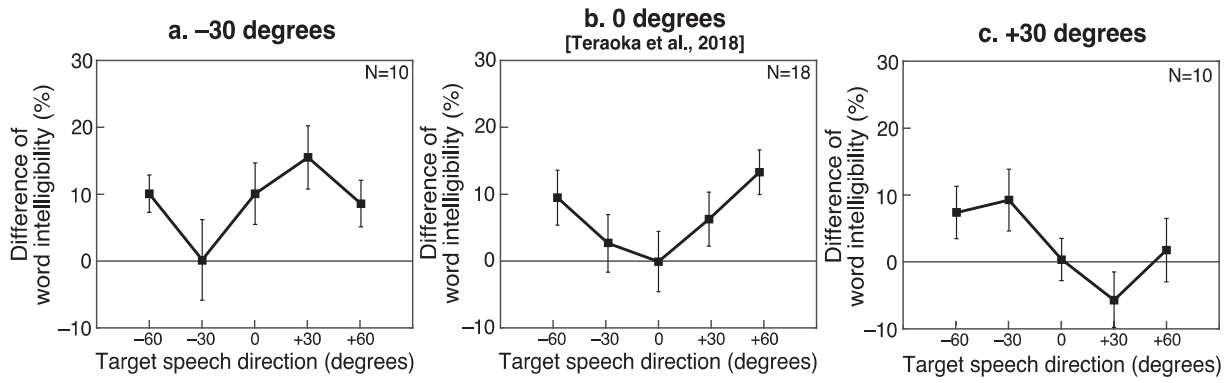

Figure 3 – Difference of word intelligibility score as a function of the direction of target speech sound (**a.** the primary angle = – 30 degrees, **b.** the primary angle = 0 degrees (Adapted from Teraoka et al. (10)), **c.** the primary angle = + 30 degrees). Solid squares represent the results of the difference these conditions. Error bars denote the standard error of the mean.

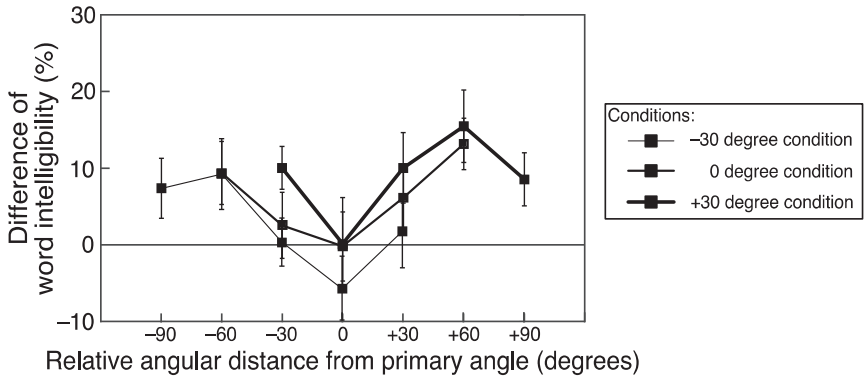

Figure 4 – Difference of word intelligibility score as a function of the relative angular distance from the primary angle. Thickest line, mid-thickness line, and thinnest line represent – 30, 0, and + 30 degrees of the primary angle condition, respectively. Solid squares represent the results of the difference these conditions. Error bars denote the standard error of the mean.

also the effect of other factors, such as the release from masking. Therefore, to analyze the effect of the auditory spatial attention, we extracted the effect of the attention by subtracting the scores of the probability-control condition from those of the cue condition. Figure 3 shows the results as a function of the target speech direction. Figure 4 shows the relative angular distance from the primary angle when all three primary angles are aligned at 0 degrees. Figure 3 shows that the difference at the primary angles is lower than those at any other probe angles. Furthermore, this score monotonically increases as the direction deviates from the primary angle, up to ± 60 degrees. The spatial pattern may represent a spatial spread of auditory selective attention. A one-way repeated-measures ANOVA was performed on the mean data across all listeners, with the target speech directions as a factor to each primary angle condition. In the – 30 degrees condition, there is no significance (conditions: $F_{4,36} = 1.66$, $p = .180$, $\eta^2_G = .072$). In the 0 degrees condition, the main effect is significant (conditions: $F_{4,36} = 1.24$, $p = .312$, $\eta^2_G = .092$). The *post hoc* test (Ryan's method, $p < .05$) revealed a significant difference in direction between 0 and + 60 degrees. In the + 30 degrees condition, there is no significance (conditions: $F_{4,36} = 1.19$, $p = .330$, $\eta^2_G = .090$). These results indicate that the spatial spread of the auditory selective attention is flat in all three primary angle conditions except for the 0 degrees condition, as shown in Fig.4. However, these results seem to suggest that there is a peak of the function at each primary angle. Thus, if the experiment were conducted with a larger number of listeners, we could expect that a significant notch might appear at the primary angle, as in the 0 degrees condition.

In Fig. 2, as the distance from the target speech direction increases from 0 degrees, the observed score also increases in all three primary angle conditions, in the two presentation conditions. This

result is inconsistent with the results of the previous study (8). The reason for the difference is the effect of a spatial release from masking (4, 6, 21–23). In this experiment, the target speech sound and distractors were presented by five loudspeakers (– 60, – 30, 0, + 30, and + 60 degrees). When the target speech sound was presented from 0 degrees, the numbers of distractors to left and right of the listener were the same. In contrast, when the target was presented from other directions, the numbers of distractors on the listener's left and right were different. For example, when the target sound was presented from the peripheral direction (e.g., ± 60 degrees), compared with to present from the center ones, the distance between the target and distractor is relatively longer. Then, due to the effect of the spatial release from masking, the effect of masking is reduced as the distance from distractor to target sound. The results of the present study may reflect the above factor.

As mentioned in the introduction, the MAA broadens as the angular distance from the front (i.e., 0 degrees) increases. This means that the auditory spatial resolution is not constant, and differs depending on the angle. Thus, sound source localization is necessary when the auditory spatial attention is directed to a specific location. We expected that the spatial extent of the auditory selective attention is affected by such factors (e.g., the spatial extent becomes broader as the distance from 0 degrees increases). However, Fig. 3 and Fig. 4 indicate that the spatial extent of the auditory selective attention is similar in all three primary angle conditions. The results suggest that there is no directional selectivity in the spatial extent of auditory selective attention. Thus, the present result is not consistent with our expectations. This seems to us to lead to an interesting future study.

5. CONCLUSIONS

In the present study, we investigated whether the spatial extent of auditory selective attention is different between the frontal and oblique directions. The results show that the spatial extent of auditory selective attention is similar in all three primary angle conditions, which suggests that the same factor determines spatial extent independently of attention directions.

ACKNOWLEDGEMENTS

This research was partly supported by JSPS KAKENHI Grant-in-Aid for Challenging Exploratory Research (No. 17K19990) and Grant-in-Aid for JSPS Research Fellow (No.18J13203).

REFERENCES

1. Cherry EC. Some experiments on the recognition of speech with one and two ears, J Acoust Soc Am. 1953;25(5):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. 1954;26(4):554–559.
3. Bregman AS. Auditory scene analysis: the perceptual organization of sound. Cambridge, Massachusetts, USA: MIT Press; 1990.
4. Ebata M. Spatial unmasking and attention related to the cocktail party problem, Acoust. Sci. & Tech. 2003;24(5):208–219.
5. Yost W. The cocktail party problem: Four years later, In: Gilkey RH, Anderson TR editors. Binaural and spatial hearing in real and virtual environments. New York, USA: Taylor & Fransis; 1997. p. 329–347.
6. Bronkhorst AW. The cocktail-party phenomenon: A review of research on speech intelligibility in multiple-talker condition, Acoust. 2000;86(1):117–128.
7. Ebata M, Sone T, Nimura T. Improvement of hearing ability by directional information, J Acoust Soc Am. 1968;43(2):289–297.
8. Arbogast TL, Kidd G. Evidence for spatial tuning in informational masking using the probe-signal method, J Acoust Soc Am. 2000;108(4):1803–1810.
9. Kidd G, Arbogast TL, Mason CR, Gallun FJ. The advantage of knowing where to listen, J Acoust Soc Am. 2005;118(6):3804–3815.
10. Teraoka R, Sakamoto S, Cui Z, Suzuki Y, Shioiri S. Influence of auditory selective attention on word intelligibility, Proc 2018 Joint Conference - Acoustics; 11-14 September 2018; Ustka, Poland 2018, p. 307–10.
11. Posner MI. Orienting of attention, Q J Exp Psychol. 1980;32(1):3–25.
12. Treisman AM, Gelade G. A feature-integration theory of attention, Cogn. Psychol. 1980;12(1):97–136.
13. Shioiri S, Honjo H, Kashiwase Y, Matsumiya K, Kuriki I. Visual attention spreads broadly but selects

information locally, Sci Rep. 2016;6(1):35513.
14. Best V, Thompson ER, Mason CR, Kidd G. Spatial release from masking as a function of the spectral overlap of competing talker (L), J Acoust Soc Am. 2013;133(6):3677–3680.
15. Mills AW. On the minimum audible angle, J Acoust Soc Am. 1957;30(4):237–247.
16. Speech Resources Consortium, <http://research.nii.ac.jp/src/FW03.html>
17. Amano S, Sakamoto S, Kondo T, Suzuki Y. Development of familiarity-controlled word lists 2003 (FW03) to assess spoken word intelligibility, Speech Commun. 2009;51(1):76–82.
18. Speech Resources Consortium, <http://research.nii.ac.jp/src/FW07.html>
19. Sakamoto S, Suzuki Y, Amano S, Ozawa K, Kondo T, Sone T. New lists for word intelligibility test based on word familiarity and phonetic balance, J Acoust Soc Jpn. 1998;54(12):842–849 [in Japanese].
20. Greenberg GZ, Larkin WD. Frequency-response characteristics of auditory observers detecting signals of a single frequency in noise: The probe-signal method, J Acoust Soc Am. 1968;44(6):1513–1523.
21. Dirks DD, Wilson RH. The effect of spatially separated sound sources on speech intelligibility, Hear Res. 1969;12(1):5–38.
22. Bronkhorst A, Plomp R. The effect of head-induced interaural time and level differences on speech intelligibility in noise, J Acoust Soc Am. 1988;83(4):1508–1516.
23. Litovsky RY. Spatial release from masking, Acoustic today 2012;8(2):18–25.