Acoust. Sci. & Tech. **42**, 1 (2021)

PAPER

Effects of listening task characteristics on auditory spatial attention in multi-source environment

Ryo Teraoka[1,2,*], Shuichi Sakamoto[1], Zhenglie Cui[1], Yôiti Suzuki[1] and Satoshi Shioiri[1]

[1]*Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, 2–1–1 Katahira, Aoba-ku, Sendai, 980–8577 Japan*

[2]*Research Fellow of the Japan Society for the Promotion of Science, Japan*

(*Received 12 January 2020, Accepted for publication 31 July 2020*)

Abstract: Human listeners can readily extract sounds of interest from distracting sounds by directing their auditory spatial attention. Although the extent to which the auditory spatial attention influences listening performance and its spatial distribution in daily situations is important, the characteristics of this ability remain unclear. To investigate the characteristics of the auditory spatial attention, we measured the word intelligibility (4-mora words) and detection threshold of a target sound (1/12 octave-band noise burst) in the presence of distractor sounds (speech sounds/noises with the same bandwidth but with different center frequencies). In the experiment, we presented a target and multiple distractors simultaneously from loudspeakers surrounding the listeners. Results showed that word intelligibility improved when the target direction was attended compared to when it was not, whereas the detection threshold of the narrow-band noise was not influenced significantly by attention. These findings suggest that we can observe the effect of auditory selective attention when the listeners continuously direct their attention to a specific direction. Moreover, the spatial pattern of word intelligibility showed a peak corresponding to the attended direction. By contrast, the threshold of the narrow-band noise was constant regardless of the presented direction in which the target was presented.

Keywords: Auditory spatial attention, Auditory selective attention, Cocktail-party effect, Auditory scene analysis

1. INTRODUCTION

In daily life, our ears receive a mixture of sounds from multiple directions. Even in an acoustically complex environment, humans can extract a sound of interest from a mixture of various sounds — a well-known phenomenon as the "cocktail-party" effect [1]. This effect is typically observed when a listener tries to understand a specific sound despite interference from multiple sources. Numerous studies have investigated about how humans can perceptually segregate a sound of interest from other sounds [for reviews, see [2–6]].

These studies have demonstrated that several factors are involved in this phenomenon. In particular, a spatial separation of the sound sources should be a strong cue for extracting the target sound from noises. For example, when a listener hears a target sound in the presence of spatially separated masker sounds, the performance of the detection, discrimination, and identification of the target is better than for when the sounds are all from the same direction [7]. This improvement is often referred to as the spatial release from masking (SRM; [8–10]), which is due to the spatial separation of sound sources based mainly on binaural interaction. When the target sound is speech, the speakers' individual voice characteristics (i.e., fundamental frequency (F0), formant frequencies, and accents) are also used as cues for segregation. Studies with voices have suggested that such speech sound characteristics play a role in enhancing the segregation of the target from distractors and, consequently, in improving the hearing performance (e.g., speech intelligibility).

The stimulus-driven auditory processing based on ear inputs plays an important role in extracting the target sound. However, such a processing cannot fully explain the "cocktail-party" effect, with the active auditory process based on goal-directed cognitive processing also playing a significant role. One of the most prominent goal-directed

*e-mail: terar@dc.tohoku.ac.jp
[doi:10.1250/ast.42.12]

processes is auditory selective attention. In particular, auditory selective attention in a spatial domain (hereafter referred to as "auditory spatial attention") is vital to isolating a sound of interest from spatially distribute distractors. However, some of previous studies have demonstrated that the effect of auditory spatial attention is small [7,11,12] or no effect [13]. Ebata *et al.* [7] investigated the auditory spatial attention when listeners focused on a pure tone presented from a specific direction, using the probe-signal method. In their study, they presented a target tone in front of a listener (i.e., 0°) in 90% of the trials and at either 45°, 90°, or 135° in the rest. Under such listening conditions, a listener's auditory spatial attention was expected to be attracted to 0°. Contrary to this prediction, the results showed that the difference in threshold between the attended (0°) and others (45°, 90°, and 135°) was small ($\lesssim$1 dB), and therefore not statistically significant.

On the other hand, there are studies reporting the positive effect of auditory spatial attention [14–16]. Arbogast and Kidd [14] demonstrated that the effects of the auditory spatial attention are observed when the irrelevant signals are similar to the target signal, and the spatial information is therefore critical to isolate the target. Ericson, Brungart, and Simpson [15] experimented with speech sounds and proved that prior information of the speech sound and its location could enhance target detection performance significantly in situations when there are more than one interfering speakers.

As mentioned previously, it remains unclear which situations are critical for auditory spatial attention to be useful, that is, which factors concerning the listening situation contribute to auditory spatial attention. The results of the previous studies imply that auditory spatial attention is observed when identifying the target speech sound in a multi-talker environment [15,16] (or sounds with a complex temporal structure [14]). However, previous studies could not determine the exact conditions in which the auditory spatial attention functions, because experimental conditions such as number of distractors, temporal conditions, experimental setups are different between the experiments that showed attentional modulation and those did not. Therefore, the aim of present study is to investigate the effects of the sound stimulus and listening task on auditory spatial attention. To clarify this problem, the effects on auditory spatial attention were compared two experiments: (1) recognizing the target speech sound in an environment with multiple talkers (i.e., simulating "cocktail-party"), and (2) detecting steady narrow-band noise (NBN) bursts (no semantic information, steady and simple frequency spectrum) as target in an environment with multiple sources (i.e., having a contradictory nature to (1)). Since these two experiments use aligned conditions, e.g., number of distractors, temporal conditions, experimental setups, except for the sound stimulus and listening task, the specific effects of the sound stimulus and listening task could be extracted by comparing results of these experiments.

In addition to the main question, we aimed to investigate the spatial shape of the auditory selective attention, which is one of the most fundamental characteristics of selective attention. In the visual domain, when humans direct their spatial selective attention to a specific location, the effects of the spatial selective attention spread over an area around the focal point [17] as the "spotlight" metaphor suggests [18]. A recent electrophysiological study confirmed the spatial spread of visual spatial attention and suggested different shapes of spatial attention corresponding to various stages of visual processing [19]. In the auditory domain, there is evidence that the auditory spatial attention operates in the manner of the spotlight of attention [20–22]. Teder-Sälejärvi and Hillyard [21] and Teder-Sälejärvi, Hillyard, Röder and Neville [22] showed that the listening performance drops off steeply (almost to the level of chance) when the direction of the target position differs by 8° or more from that of the attended direction. In contrast, Arbogast and Kidd [14] showed that the performance decreases gradually as the angular distance from the attended direction increases. These results suggest that the spatial shape of the auditory selective attention changes depending on the listening environment and listening tasks. However, there seem to be no reports describing the spatial spread of auditory selective attention in response to competing speech sounds. To elucidate this problem, we measured the shape of the auditory spatial attention based on speech intelligibility in the presence of multiple speech sounds and compared the results to those obtained for the NBN detection task.

2. EXPERIMENT 1

Experiment 1 aimed to clarify the extent of the influence of the auditory spatial attention on speech recognition and, furthermore, the area covered by the auditory selective attention spotlight with respect to speech sound. To accomplish this purpose, we measured the word intelligibility under two conditions: (1) target speech sound presented at the direction and then not at the direction explicitly indicated by a preceding sound and (2) target speech sound presented at the implicitly expected direction using the probe-signal method.

2.1. Methods

2.1.1. Listeners

Twenty listeners (10 males and 10 females aged between 18–24 years) participated in this experiment. All the listeners were naïve to the purpose of the experiment.

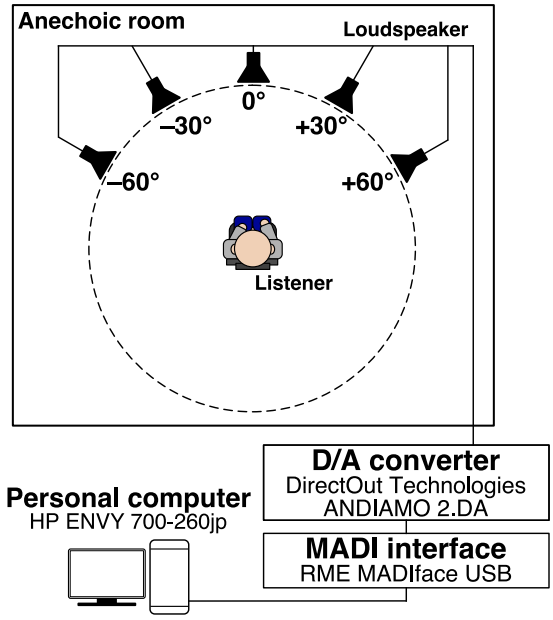

Fig. 1 Schema of the experimental setup. The listener sits on a chair at the center of the loudspeaker array, facing the front (0°).

All the listeners were native Japanese speakers with normal-hearing acuity. We obtained informed consent from each listener before the experiment. The Ethics Committee of the Research Institute of Electrical Communication, Tohoku University approved the procedure.

2.1.2. Apparatus and stimuli

We conducted the experiment in an anechoic room at the Research Institute of Electrical Communication, Tohoku University [23]. Figure 1 shows the experimental setup. The five loudspeakers were placed circularly on a horizontal plane centered in-between two ears in intervals of 30° at the height of 1.3 m. The loudspeaker directions were −60°, −30°, 0°, +30°, and +60° (positive values represent the righthand side of the listener) at 1.6 m from the listener. Sound stimuli were generated using a desktop computer (HP ENVY 700-260jp) through a MADI interface (RME MADIface USB) and a D/A converter (DirectOut Technologies ANDIAMO 2.DA). We used MATLAB (version 2017a) with an open-source audio I/O library (Playrec, http://www.playrec.co.uk/) to control the experiments. Speech sounds comprising of Japanese words of four moras uttered by one male and female speaker were selected from "Familiarity-controlled Word lists 2003" (FW03; [24,25]). This word list is divided into four sets according to word-familiarity rank: high familiarity (7.0–5.5), upper-middle familiarity (5.5–4.0), lower-middle familiarity (4.0–2.5), and low-familiarity (2.5–1.0). In this experiment, we used the 1,000 words with the highest familiarity rank (7.0–5.5) from the total of 4,000 words (i.e., each rank includes 1,000 words). The target speech sounds were the 400 words that are included in "Familiarity-controlled Word lists 2007" (FW07; [26,27]). FW07 is a shrunken version of FW03 for clinical use and comprises 20 lists for each familiarity rank. Each list contains 20 words (i.e., each rank includes 400 words). The other 600 words were used as distractors. In experiment 1, one female (fhi) and one male (mya) speech sound were assigned as a target and a distractor, respectively.

The A-weighted sound pressure level of each target and distractor speech sound was set at 65 dB at the center position of a listener's head. The level was represented by equivalent continuous A-weighted sound pressure level. Here, the time duration for the time averaging was the length of the words recorded in FW03 because the words were used as recorded without any editing. The target sound and four distractors were presented simultaneously from five loudspeakers.

2.1.3. Procedure

This experiment consisted of two conditions: cue and probability-control conditions. The order of the two conditions was counterbalanced across listeners. For both conditions, the listeners were instructed to focus on a target speaker (i.e., female speaker) and to write down the uttered words spoken on a response sheet as soon as they heard them. Moreover, the listeners were asked to keep their head stationary and straight ahead at 0° during the whole session. A single target word uttered by the female speaker was presented from one of the five loudspeakers, whilst four distractors uttered by the same male speaker were presented from the other four loudspeakers.

In the cue condition, the loudspeaker from where the target speech sound would be presented was indicated beforehand. To indicate the direction of this loudspeaker, a 500 ms burst of white noise (including a 10 ms rise/fall, A-weighted sound pressure level: 65 dB) was delivered via a loudspeaker 1,000 ms prior to the presentation of the speech sounds. The sound pressure level of this noise burst was defined as the value when the steady part of the sound stimulus was presented continuously. Listeners were asked to direct their attention to the loudspeaker from which the cue sound was presented. Twenty words (i.e., 1 list) were assigned as target speech sounds to one of five loudspeakers, and 80 words, which were selected randomly from the 600 distractor candidates, were assigned separately to the remaining four loudspeakers (i.e., 20 words to each loudspeaker). Target speech sounds were presented from each of the five loudspeakers with equal probability ($p = 0.2$); consequently, 100 (20 words × 5 loudspeakers) and 400 (100 words × 4 remaining loudspeakers) words were assigned as the target and distractors, respectively. This condition consisted of 100 trials. The order of the presented target word and the loudspeaker from where it was presented were randomized.

In the probability-control condition, the direction to be attended was not explicitly indicated to the listeners. By contrast, to attract attention to a 0° loudspeaker implicitly, the probe-signal method was applied [28]. Among the 400 trials, the target speech sound was presented from the 0° loudspeaker in 80% of the trials. In the remaining 20%, the target speech sound was presented from one of the other four loudspeakers chosen randomly with equal probabilities (5% each). That is, 320 words (i.e., 16 lists) were presented as the target speech sound from the 0° direction and 1,280 words (i.e., 320 words × 4 remaining loudspeakers) were assigned separately as distractors from the remaining 4 loudspeakers. Whereas, 80 words (i.e., 4 lists) were presented as the target sound from the other 4 loudspeakers (−60°, −30°, +30°, and +60°) and 320 words (i.e., 20 words × 4 remaining loudspeakers) were presented separately as distractors from the remaining 4 loudspeakers. Consequently, 400 and 1,600 words were assigned as the target and distractors, respectively. Following this procedure, we expected that the listeners would be aware that most of the target speech sounds were presented from the loudspeaker at 0°, and thus direct their attention to 0°.

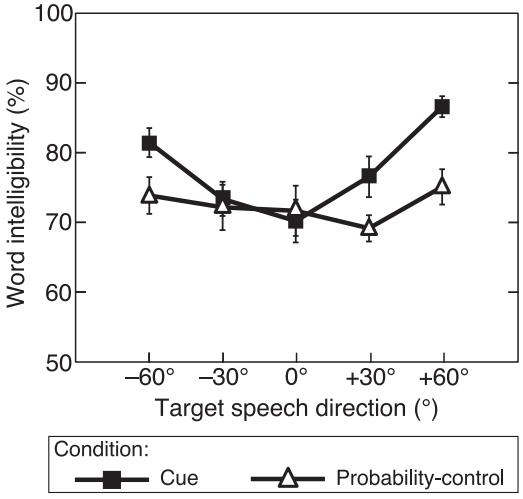

Fig. 2 Word intelligibility score as a function of the direction of target speech sound. Solid squares and open triangles represent the results of the cue and probability-control conditions, respectively. Error bars denote the standard error of the mean.

2.2. Results and Discussion

In experiment 1, we calculated the word intelligibility score as the rate at which target words were recognized correctly for each condition. At 0° in the probability-control condition, the word intelligibility score was calculated using the results of the last 20 trials (from a total of 320) to match the number of trials corresponding to other angles and to select the trial where the attention would be best established in the probe-signal method. Figure 2 shows the word intelligibility scores as a function of target speech sound direction. Data points and error bars represent the mean and standard errors, respectively, averaged across all listeners. Solid squares and open triangles represent the results of the cue and probability-control conditions, respectively.

In the cue condition, the word intelligibility at 0° was lower than that for other directions and the score increased as the angular distance from 0° increased. By contrast, in the probability-control condition, the word intelligibility deviated a little (approximately ±5%). This means that the effect of the direction was not observed clearly. The mean intelligibility scores for the probability-control condition were approximately 12% lower than those for the cue condition. A two-way repeated-measure analysis of variance (ANOVA) was performed on the mean data with the condition (cue/probability-control) and target speech sound directions (−60°, −30°, 0°, +30°, and +60°) as factors. Both the main and the interaction effects were statistically significant (condition: $F_{1,19} = 4.69$, $p = 0.043$, $\eta_G^2 = 0.049$; target speech sound direction: $F_{4,76} = 6.94$, $p < 0.001$, $\eta_G^2 = 0.091$; condition × target speech sound direction: $F_{4,76} = 2.54$, $p = 0.046$, $\eta_G^2 = 0.039$). Further analysis of this interaction revealed that the word intelligibility scores for −60° and +60° in the probability-control condition were significantly lower than those in the cue condition (−60°: $F_{1,19} = 3.95$, $p = 0.050$, $\eta_G^2 = 0.219$, +60°: $F_{1,19} = 9.08$, $p = 0.003$, $\eta_G^2 = 0.503$). Moreover, these scores for the cue condition were significantly different depending on the target speech sound directions ($F_{4,76} = 8.24$, $p < 0.001$, $\eta_G^2 = 0.869$). The *post hoc* test (Ryan's method, $p < 0.05$) revealed significant differences in the direction between +30° and +60°, −30° and +30°, 0° and +60°, and −30° and +60°. The results show that the word intelligibility for the cue condition was significantly higher than that for the probability-control condition. The results of experiment 1 demonstrated that the mean word intelligibility scores in the cue condition are higher (ca. 10%) than those in the probability-control condition, with the scores at ±60° highlighting the significant difference between the conditions.

In Fig. 2, the results of the cue condition depend on the attended direction and the direction from where the target speech sound was presented, although a listener's attention was expected to be directed equally between all the target directions. As the distance from 0° increased, the observed score (i.e., word intelligibility) also increased. One possible reason of such result is the effect of SRM which has been reported elsewhere [8,29]. Peissig and Kollmeier [29] have indicated that the speech recognition threshold (SRT) is decreased by the effect of SRM when the angular distance between a target sound and distractor is increased up to

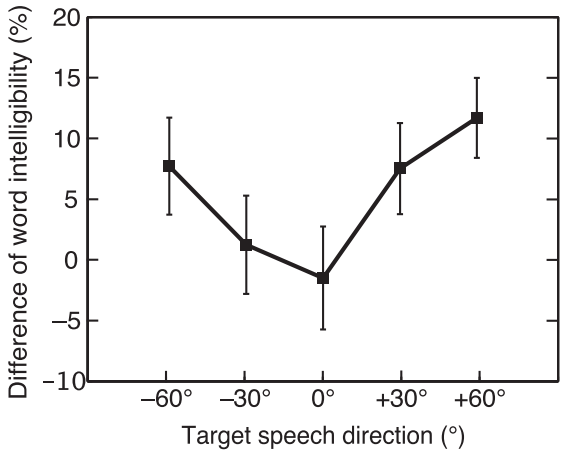

Fig. 3 Difference in word intelligibility between conditions (i.e., cue and probability-control) as a function of target speech sound direction. Error bars denote the standard error of the mean.

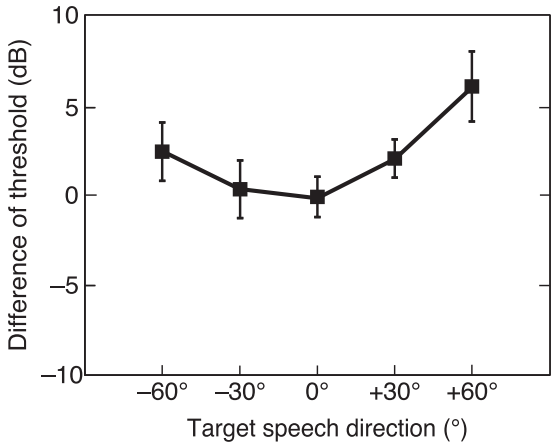

Fig. 4 Estimation of the speech recognition threshold for target speech sound as a function of the target sound direction. Error bars denote the standard error of the mean.

approximately 90°; after that, the SRM increases as this distance approaches 180°. Therefore, when the target sound is presented from peripheral directions (i.e., ±30° and ±60°), the word intelligibility score increases. The results of the cue condition presented in experiment 1 of this study indicate a U-shaped function due to SRM. By contrast, in addition to SRM, the results of the probability-control condition may be affected by spatial attention, resulting in a relatively flat shape.

To negate the impact of common factors that were encountered for both conditions, such as SRM, we subtracted the scores of the probability-control condition from those of the cue condition and regarded the difference as the valued proportional to the effect of spatial attention. Figure 3 illustrates the result as a function of the target speech direction. This figure shows clearly that the difference at 0° is lower than that for other directions. Moreover, this score monotonically increases with the increase of the direction difference from 0°. A one-way repeated-measure ANOVA was performed on the subtracted results with the target speech sound directions (−60°, −30°, 0°, +30°, and +60°) as a factor. The main effect is significant ($F_{4,76} = 2.54$, $p = 0.046$, $\eta_G^2 = 0.074$). The *post hoc* test (Ryan's method, $p < 0.05$) revealed a significant difference due to the direction changing from 0° to +60°. The results showed that the word intelligibility scores decreased significantly as the angular distance increased.

Since previous studies examined the spatial extent of auditory selective attention, measuring a detection threshold in dB, it is preferable to convert our results to a threshold in dB in order to compare with previous studies. Therefore, we calculated the SRT based on the results of a previous study [30]. To convert the word intelligibility score into the SRT, the relationship between the word intelligibility and the signal-to-noise ratio for the words recorded in FW03 (see Fig. 1(d) in the Amano *et al.* [30]) was used. For example, since the mean score of word intelligibility at 0° in cue condition is 70%, the 70% point of the function (−8 dB) is defined as the SRT. In the present study, the word intelligibility score was subtracted between the two conditions after the word intelligibility was converted to the SRT.

The study by Amano *et al.* examined the relationship between the word intelligibility score and signal-to-noise ratio for FW03, although some of listening conditions differed from the present study; the authors measured the word intelligibility score for a monaural listening situation using headphones, with the speech spectral shape noise was used as the masker. Therefore, their results do not include in the masking effects such as SRM. Similarly, the subtracted result (cue − probability-control) in the present study does not contain the effect of masking, as we considered it broadly irrelevant with respect to our investigation.

Figure 4 shows the estimated SRT as a function of the target speech direction, with the results forming a U-shaped function.

3. EXPERIMENT 2

In comparison with speech sounds, experiment 2 used more. In this experiment, we examined the extent of the influence of the auditory spatial attention on NBN detection and, in addition, the area covered by the auditory selective attention spotlight with respect to NBN. To accomplish this, we measured the detection threshold of NBN under identical conditions as experiment 1; namely, the aforementioned cue and probability-control conditions. In order to compare the results of experiment 2 with those

of experiment 1 meaningfully, we minimized the differences between the methods of both experiments, e.g., number of distractors, temporal conditions, experimental setups, with the exception of the stimuli themselves.

3.1. Methods

3.1.1. Listeners

Twenty listeners (3 females, 17 males aged between 20–24 years) participated in this experiment. All the listeners were naïve to the purpose of the experiment and none of them participated in experiment 1. All the listeners had normal-hearing acuity. We obtained informed consent from each listener before the experiment. The Ethics Committee of the Research Institute of Electrical Communication, Tohoku University approved the procedure.

3.1.2. Apparatus and stimuli

The apparatus was the same as experiment 1, one exception. To collect a listener's response by a PC mouse, we used MATLAB (version 2017a) with the Psychtoolbox 3.0 [31,32]. The sound stimulus was a burst of 1/12-octave-band noise with center frequencies of 125, 200, 350, 500, and 1,000 Hz. The duration of this NBN burst was 750 ms (including a 5-ms rise and fall) according to the average duration of the speech sounds that were used in experiment 1. In this experiment, an NBN centered at 1,000 Hz was assigned as the target. Other NBNs were assigned as distractors and each of them was presented at one of the four non-target locations.

3.1.3. Procedure

In experiment 2, we measured the detection threshold of the target NBN in the presence of other competing NBNs. To measure the threshold of the target NBN, the sound pressure level of the target NBN was varied adaptively using a one-up/one-down method. The listeners were divided into two groups depending on the sound pressure level of the NBN at the beginning of the test. For half of the listeners, at the beginning of the test, the A-weighted sound pressure level of the target NBN was initially set to 80 dB. This level was defined as the value when the steady part of the sound stimulus was presented continuously. For the remaining 10 listeners, the NBN was initially set to 55 dB. The sound pressure level of the distractors was set at a constant level of 65 dB for the duration of the session. The listeners were asked to click the left mouse button when they could hear the target NBN and the right mouse button when they could not. The level of the target NBN was decreased by 2 dB when the left button was pressed, whereas the sound pressure level was increased by 2 dB when the right button was pressed. The listeners were instructed to focus on the target NBN (i.e., the NBN with a center frequency of 1,000 Hz). Moreover, the listeners were asked to keep their head stationary and facing straight ahead at 0° during the whole session. A single target NBN was presented from one of the five loudspeakers and the four distractors were presented from the other four loudspeakers.

This experiment consisted of two conditions: the cue and probability-control conditions. The order of the two conditions was counterbalanced across listeners. In the cue condition, the loudspeaker where the next target NBN would be presented was indicated beforehand by a 500 ms burst of white noise from the loudspeaker 1,000 ms prior to the presentation of the target NBN burst. The listeners were asked to direct their attention to the loudspeaker from which the cue was presented. The experiment in this condition consisted of a set of blocks in which there were 100 trials (20 trials per target direction). After reaching 10 reverse points for all the five directions at the end of the block, the experiment in progress was finished. If this criterion was not satisfied, the next block was started. The threshold for each trial was computed by averaging the midpoint values over the last five reversals of a stimulus level. In addition, the order of direction was randomized.

In the probability-control condition, as in experiment 1, a listener's attention was implicitly attracted to the front (0°) using the probe-signal method. The experiment in this condition consisted of a set of blocks with 400 trials. The target NBN was presented from 0° in 80% of the trials (i.e., 320 trials). In the remaining 20% of the trials (i.e., 80 trials), the target NBN was presented from one of the other four loudspeakers chosen at random. If the number of reverse points reached 10 times at −60°, −30°, +30°, and +60° and 160 times at 0° at the end of the block, the experiment in this condition was finished. If this criterion was not satisfied, the next block was started. Regarding the cue condition, the threshold for each trial was computed by averaging the midpoint intensities over the last five reversals of a stimulus level. Once again, the order of direction was randomized. By following this procedure, we expected that the listeners would be aware that most target speech sounds were presented from the loudspeaker at 0°, resulting in their attention being directed to 0°.

3.2. Results and Discussion

Figure 5 shows the detection threshold for the NBN target. The thresholds are plotted as a function of the loudspeaker direction from which the target NBN burst was presented. The data points and error bars represent the mean and standard error, respectively, averaged across all listeners. The solid squares and open triangles denote the results of the cue and probability-control conditions, respectively.

For both conditions, the detection thresholds showed a small amount of deviation (approximately ±3%) and there was no discernable effect due to the direction of the target noise. We performed a two-way repeated-measure

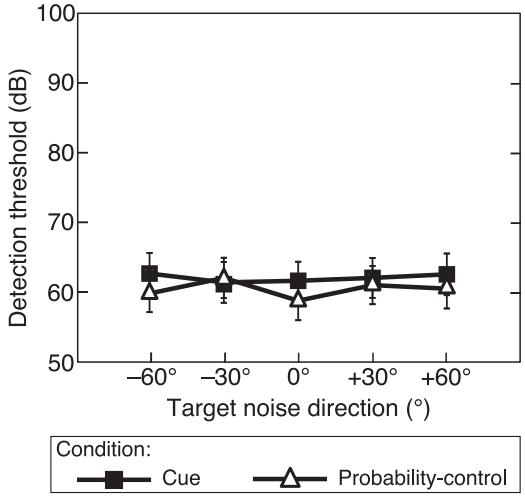

Fig. 5 Detection of threshold score for the narrow-band noise burst as a function of the direction of target noise. Solid squares and open triangles represent the results of the cue and probability-control conditions, respectively. Error bars denote the standard error of the mean.

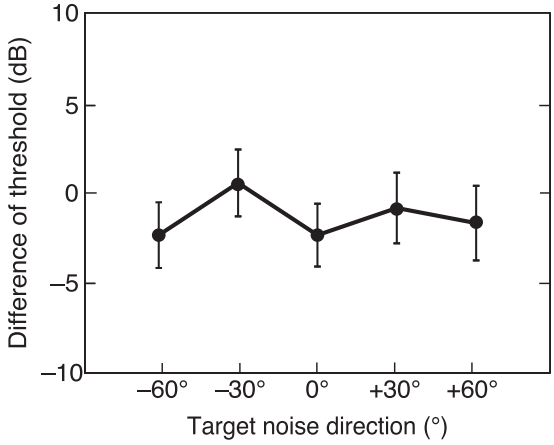

Fig. 6 Difference in the detection threshold for the narrow-band noise between the conditions (i.e., cue and probability-control) as a function of the target noise direction. Error bars denote the standard error of the mean.

ANOVA on the mean data with the condition (cue/probability-control) and target NBN directions (-60°, -30°, 0°, $+30^\circ$, and $+60^\circ$) as factors. The main or interaction effect was insignificant (condition: $F_{1,19} = 0.67$, $p = 0.421$, $\eta_G^2 = 0.004$; target NBN direction: $F_{4,76} = 0.72$, $p = 0.584$, $\eta_G^2 = 0.002$; condition $\times$ target NBN direction: $F_{4,76} = 1.33$, $p = 0.266$, $\eta_G^2 = 0.003$), indicating that there was no difference in the detection threshold between the two conditions.

To examine the effects of auditory spatial attention, we estimated the effects of attention by subtracting the scores in the probability-control condition from those in the cue condition. Figure 6 shows the result as a function of the target NBN direction. This figure shows that the difference in the detection threshold for the two conditions is flat. A one-way repeated-measure ANOVA was performed on the subtracted results with the target sound directions (-60°, -30°, 0°, $+30^\circ$, and $+60^\circ$) as a factors. The main effect is not significant ($F_{4,76} = 1.33$, $p = 0.266$, $\eta_G^2 = 0.017$). These findings suggest that the auditory spatial attention does not affect the simple detection performance for an NBN.

4. GENERAL DISCUSSION

Through these two experiments, we investigated the extent of the effect of auditory spatial attention on listening performance and whether the effect manifests differently depending on the sound stimulus and listening task within the multi-source environment. To differentiate the effects of sound stimulus and listening task, the same experimental conditions were used for both experiments, with the exception of the stimuli and schemes of the listening task (speech identification/noise detection). Both the experiments were performed in a multi-source environment using two conditions: (1) target sound presented from an expected direction (the cue condition) and (2) target sound presented usually from an implicitly expected direction with a certain rate for which the target sound was presented from an unexpected direction (i.e., the probability-control condition). In experiment 1, the word intelligibility score increased approximately 12% when the target sound was presented from an expected direction ($\pm 30^\circ$ and $\pm 60^\circ$). By contrast, in experiment 2, no difference was found between the two conditions. These results show clearly that the effect of auditory spatial attention is observed for experiment 1 only, indicating that the effect of the auditory spatial attention may depend on the sound stimulus and listening task. However, the present study cannot discriminate the effects of the type of sound stimulus (speech sounds/NBNs) and the listening task (recognizing/detecting) because these two factors were different between experiments 1 and 2. Further refinement is therefore required to elucidate the effect of auditory spatial attention more completely in future study. Incidentally, in experiment 1, a spectral overlap occurs between the target and the distractor and, also, between distractors, whereas there is no overlap of stimuli in experiment 2. Best, Thompson, Mason and Kidd [33] demonstrated that the spectral overlap has little bearing on the effect of SRM, regardless of the type of sound stimulus (speech sound/narrow band noise). Therefore, this difference is not considered an influence on the results presented herein.

As discussed, we estimated the effects of the auditory spatial attention by subtracting the scores for the probability-control condition from those for the cue condition.

The corresponding results showed that in experiment 1, the difference between the two conditions increased gradually depending on the angle increase, whilst in experiment 2, the difference was little and seemed flat. This spatial pattern is regarded as the spatial spread of the auditory selective attention. A number of other studies have observed, as with the results of experiment 2, that the effects of the auditory spatial attention are small or nonexistent [7,11–13]. Ebata and his colleagues [7] measured the effects of the auditory spatial attention for a sound detection task, showing that the difference in threshold between the attended (0°) direction and others (45°, 90° and 135°) was small ($\lesssim$1 dB). By contrast, more significant effects of auditory spatial attention are reported in other studies [14,15,21,22]. Teder-Sälejärvi and Hillyard [21] investigated whether the detection performance for the target stimulus was affected by the effects of the auditory spatial attention. Their results showed that detection accuracy at all the attended directions were approximately 50% higher than the unattended directions. Arbogast and Kidd [14] examined whether the identification performance for the target frequency pattern was affected by the effect of auditory spatial attention in the presence of several distractors. Their results showed that the mean accuracy for the unattended directions was 5.6% lower than that for the attended directions.

Whether the auditory attention effect is found or not might be determined by the time required to maintain a specific direction for a certain period. In Ebata and his colleagues' study [7], a listener could detect the sounds at the moment the sound stimuli were heard, and thus it was not necessary for the listener to keep directing their auditory spatial attention after the detection in order to execute the task. In contrast, in Teder-Sälejärvi and colleagues' study [21], a listener was asked to judge whether the stimulus was the target and whether the presented direction was the attended direction. In this case, the listener was required to maintain their auditory spatial attention in a certain direction for a certain period of time. The target stimulus in Arbogast and Kidd's experiment [14] also lasted 480 ms. Moreover, the frequency of the target stimulus was changed gradually. Consequently, listeners were required to maintain their auditory spatial attention, in the direction where the target sound was presented. Therefore, the listening situation in the experiment in terms of the period that the listeners should keep their attention was the same as for Teder-Sälejärvi and colleagues' study. In experiment 1 of the present study, the target speech sound uttered by the female speaker was a 4-mora word in experiment 1. Further, the listeners had to answer the whole word. Therefore, the listeners had to maintain their auditory spatial attention in the direction where the target speech sound was presented. By contrast, in experiment 2, a listener did not need to keep directing their auditory spatial attention after the detection, because he/she could detect the target NBN burst when the sound stimuli were heard. This factor might explain the difference between the two results for the present study.

In the visual domain, the spatial spread of selective attention has been well-documented. Several studies have suggested that the visual selective attention spreads spatially and has been likened to several metaphors, such as a "spotlight" [18] or a "zoom-lens" [17,34–36]. Posner [18] argued that the "beam" of the visual selective attention is fixed in size and shape, and can be directed at a single area of the visual field. By contrast, Ericksen and St. James [34] proposed that this "beam" can be contracted or expanded as required by a task or instruction, a hypothesis that has supports from some psychophysical studies [17].

In the present study, results show that the spatial spread of auditory selective attention depends on the listening task. These findings suggest that the spatial spread of auditory selective attention could be changed as required by the sound stimulus and/or the listening task, akin to the zoom-lens metaphor.

Interestingly, the attention effect in experiment 1, with broad spatial tuning, is similar to the recent finding regarding the visual spatial attention at the early stage of visual processing [19], in which Shioiri *et al.* also report a narrow-band spatial tuning of the visual attention for a later stage. The similarity of the broad spatial tuning between the auditory and visual attention may or may not indicate a common attention process for audio and visual attention. Although several studies have indicated that the auditory selective attention is affected by other processes such as eye movement control [37,38], there is still insufficient information to understand the relation. Future research should investigate the relation between auditory spatial attention and the processing of selective spatial attention in other modalities such as vision.

The word intelligibility scores on the right side (+30° and +60°) were higher than those on the left side (−60° and −30°) under both conditions (see results for experiment 1). The reason is not clear. Previous studies have reported that speech perception is more accurate in reporting items arriving at the right ear compared to the left ear, when two different sounds are simultaneously and separately delivered in the two ears (i.e., dichotic listening condition) and the listener is asked to repeat the speech sounds spoken by one of the talkers. This phenomenon is known as the "right-ear advantage" [39]. Kimura [39] associated this phenomenon with the specialization of the left hemisphere with respect to language processing. This phenomenon might be closely related to auditory attention because, to accomplish a dichotic listening task, our brain

must focus its attention along the right/left ear. Indeed, several studies have suggested that this phenomenon is susceptible to attentional focus [40,41] in a dichotic listening situation. Consistent with these previous studies, the present findings suggest that the word intelligibility score is affected by a right-ear advantage in a free field listening situation.

5. CONCLUSION

The present study investigated the extent to which auditory spatial attention contributes to the listening performance when competing distractors exist, as in the "cocktail-party" scenario. To clarify this problem, we measured the word intelligibility score in a multi-talker environment with the detection threshold for the target NBN burst in a multi-source environment. To compare the results of experiment 2 with those of experiment 1, we tried to minimize the differences between the experimental methods of both experiments, e.g., number of distractors, temporal conditions, experimental setups, except for the stimuli themselves. Results showed that the effect of the auditory spatial attention appeared in the speech recognition task, resulting in a spatial spread of attention being represented by a U-shaped function with a maximum value at 0°. Furthermore, this difference increases depending on the azimuthal distance between the direction attended and that of the target sound. By contrast, in the noise detection task this effect hardly appeared, resulting in a flat shape. Therefore, the present study suggests that the effect of the auditory spatial attention is more advantageous when maintained in a specific direction while listening to a target sound.

ACKNOWLEDGMENT

This work was partly supported by JSPS KAKENHI grant numbers 19H0111, 17K19990, and 18J13203.

REFERENCES

[1] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.*, **25**, 975–979 (1953).
[2] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, 1990).
[3] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acust. united Ac.*, **86**, 117–128 (2000).
[4] A. W. Bronkhorst, "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," *Atten. Percept. Psychophys.*, **77**, 1465–1487 (2015).
[5] C. J. Darwin, "Listening to speech in the presence of other sounds," *Philos. Trans. R. Soc. Lond. B*, **363**, 1011–1021 (2008).
[6] M. Ebata, "Spatial unmasking and attention related to the cocktail party problem," *Acoust. Sci. & Tech.*, **24**, 208–219 (2003).
[7] M. Ebata, T. Sone and T. Nimura, "Improvement of hearing ability by directional information," *J. Acoust. Soc. Am.*, **43**, 289–297 (1968).
[8] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.*, **83**, 1508–1516 (1988).
[9] D. D. Dirks and R. H. Wilson, "The effect of spatially separated sound sources on speech intelligibility," *J. Speech Hear. Res.*, **12**, 5–38 (1969).
[10] R. Y. Litovsky, "Spatial release from masking," *Acoust. Today*, **8**, 18–25 (2012).
[11] P. T. Quinlan and P. J. Bailey, "An examination of attentional control in the auditory modality: Further evidence for auditory orienting," *Percept. Psychophys.*, **57**, 614–628 (1995).
[12] C. J. Spence and J. Driver, "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," *J. Exp. Psychol. Hum. Percept. Perform.*, **20**, 555–574 (1994).
[13] B. Scharf, S. Quigley, C. Aoki, N. Peachey and A. Reeves, "Focused auditory attention and frequency selectivity," *Percept. Psychophys.*, **42**, 215–223 (1987).
[14] T. L. Arbogast and G. Kidd, "Evidence for spatial tuning in informational masking using the probe-signal method," *J. Acoust. Soc. Am.*, **108**, 1803–1810 (2000).
[15] M. A. Ericson, D. S. Brungart and B. D. Simpson, "Factor that influence intelligibility in multitalker speech display," *Int. J. Aviat. Psychol.*, **14**, 313–334 (2009).
[16] G. Kidd, T. L. Arbogast, C. R. Mason and F. J. Gallun, "The advantage of knowing where to listen," *J. Acoust. Soc. Am.*, **118**, 3804–3815 (2005).
[17] S. Shioiri, K. Yamamoto, K. Oshida, K. Matsubara and H. Yaguchi, "Measuring attention using flash-lag effect," *J. Vis.*, **10**, 10 (2010).
[18] M. I. Posner, "Orienting of attention," *Q. J. Exp. Psychol.*, **32**, 3–25 (1980).
[19] S. Shioiri, H. Honjyo, Y. Kashiwase, K. Matsumiya and I. Kuriki, "Visual attention spreads broadly but selects information locally," *Sci. Rep.*, **6**, 33513 (2016).
[20] V. Best, F. J. Gallun, A. Ihlefeld and B. G. Shinn-Cunningham, "The influence of spatial separation on divided listening," *J. Acoust. Soc. Am.*, **120**, 1506–1516 (2006).
[21] W. A. Teder-Sälejärvi and S. A. Hillyard, "The gradient of spatial auditory attention in free field: An event-related potential study," *Percept. Psychophys.*, **60**, 1228–1242 (1998).
[22] W. A. Teder-Sälejärvi, S. A. Hillyard, B. Röder and H. J. Neville, "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," *Cogn. Brain Res.*, **8**, 213–227 (1999).
[23] S. Sakamoto, F. Saito, Y. Suzuki, M. Kakinuma, H. Ohyama, H. Matsuo and K. Takashima, "Construction of two anechoic rooms with a new experimental floor structure," *Proc. Inter-Noise 2016*, pp. 4952–4960 (2006).
[24] S. Sakamoto, Y. Suzuki, S. Amano, K. Ozawa, T. Kondo and T. Sone, "New lists for word intelligibility test based on word familiarity and phonetic balance," *J. Acoust. Soc. Jpn.* (*J*), **54**, 842–849 (1998) (in Japanese).
[25] Speech Resources Consortium, "Familiarity-controlled word lists 2003 (FW03)," http://research.nii.ac.jp/src/FW03.html (accessed 10 Jan. 2020).
[26] T. Kondo, S. Amano, S. Sakamoto and Y. Suzuki, "Development of familiarity-controlled word-lists (FW07)," *IEICE Tech. Rep.*, **107**, 43–48 (2008) (in Japanese).
[27] Speech Resources Consortium, "Familiarity-controlled word lists 2007 (FW07)," http://research.nii.ac.jp/src/FW03.html (accessed 10 Jan. 2020).
[28] G. Z. Greenberg and W. D. Larkin, "Frequency-response characteristic of auditory observers detecting signals of a

single frequency in noise: The probe-signal method," *J. Acoust. Soc. Am.*, **44**, 1513–1523 (1968).
[29] J. Peissig and B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.*, **101**, 1660–1670 (1997).
[30] Y. Amano, S. Sakamoto, T. Kondo and Y. Suzuki, "Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese," *Speech Commun.*, **51**, 76–82 (2009).
[31] D. H. Brainard, "The psychophysics toolbox," *Spat. Vis.*, **10**, 433–436 (1997).
[32] D. G. Pelli, "The VideoToolbox software for visual psychophysics: Transforming numbers into movies," *Spat. Vis.*, **10**, 437–442 (1997).
[33] V. Best, E. R. Thompson, C. R. Mason and G. Kidd, "Spatial release from masking as a function of the spectral overlap of competing talker (L)," *J. Acoust. Soc. Am.*, **133**, 3677–3680 (2013).
[34] C. W. Ericksen and J. D. St. James, "Visual attention within and around the field of focal attention: A zoom lens model," *Percept. Psychophys.*, **40**, 225–240 (1986).
[35] K. Matsubara, S. Shioiri and H. Yaguchi, "Spatial spread of visual attention while tracking a moving object," *Opt. Rev.*, **14**, 57–63 (2007).
[36] G. L. Shulman and J. Wilson, "Spatial frequency and selective attention to spatial location," *Perception*, **16**, 103–111 (1987).
[37] R. M. Braga, R. Z. Fu, B. M. Seemungal, R. J. S. Wise and R. Leech, "Eye movements during auditory attention predict individual differences in dorsal attention network activity," *Front. Hum. Neurosci.*, **10**, 164 (2016).
[38] M. J. Schut, N. Van der Stoep and S. Van der Stigchel, "Auditory spatial attention is encoded in a retinotopic reference frame across eye-movements," *PLoS ONE*, **13**, e0202414 (2018).
[39] D. Kimura, "Functional asymmetry of the brain in dichotic listening," *Cortex*, **3**, 163–178 (1967).
[40] A. E. Asbjørnsen and M. P. Bryden, "Biased attention and the fused dichotic words test," *Neuropsychologia*, **34**, 407–411 (1996).
[41] A. E. Asbjørnsen and K. Hugdahl, "Attentional effects in dichotic listening," *Brain Lang.*, **49**, 189–201 (1995).