Chapter 8
Speech Intelligibility and HRTF

Abstract This chapter describes the effects of the HRTF on speech intelligibility. An HRTF causes an interaural phase difference. Under the presence of a masker, the threshold of the target sound is affected by the relationship in the interaural phase difference between the target sound and the masker. In other words, the threshold of the target sound is changed by the incident directions of the target sound and the masker. This also appears as a difference in speech intelligibility.

8.1 Binaural Masking Level Difference

When the target sound (maskee) and the interfering sound (masker) are presented to both ears through headphones, the masked threshold of the target sound, the sound pressure level at which the target sound can be heard in the presence of disturbing sound, changes according to the interaural phase difference (interaural time difference). In other words, the target sound becomes easier to hear or harder to hear.

Based on the masking threshold when both the masker and maskee are presented to a single ear (N_mS_m), the amount of change in the masking threshold when the masker and maskee are presented to both ears with a phase difference is referred to as the binaural masking level difference (BMLD). Here, the subscript m indicates monaural. The BMLD is known to be 12–15 dB at N_0S_π (interaural phase difference of the masker and the maskee are 0° and 180°, respectively) (Blauert 1996).

Moreover, the amount of change in speech intelligibility, not in the masking threshold, using a voice as maskee is referred to as the binaural intelligibility level difference (BILD).

Experiments were performed in which pink noise as a masker and click trains as a maskee were presented by loudspeakers placed on the horizontal plane or the median plane (Saberi et al. 1991). For a masker presented from the front, when the maskee was presented from just lateral direction (azimuth of ±90°), the masking threshold decreased by approximately 15 dB, as compared with the case in which the maskee was presented from the front. On the other hand, when the maskee was presented from the rear (azimuth of 180°), the masking threshold was approximately the same

K. Iida, *Head-Related Transfer Function and Acoustic Virtual Reality*,
https://doi.org/10.1007/978-981-13-9745-5_8

as that presented from the front. When the maskee was presented from above (vertical angle of 60°–150°), the masking threshold decreased by approximately 8 dB.

8.2 Influence of Incident Direction on Word Intelligibility

As described above, the interaural phase difference caused by the HRTF changes the masking threshold, therefore, it is assumed that the speech intelligibility changes depending on the incident direction of the target sound and the interference sound when the target sound is speech. In order to verify this finding, the following experiment was conducted.

In the anechoic chamber, the preceding sound was presented from the front, and a single echo was presented from either the right horizontal plane (in 30° steps) or the upper median plane (in 30° steps), as shown in Fig. 8.1. The time delay and sound pressure level of the single echo compared with the preceding sound were 1s and 0 dB, respectively.

The sound source signal was a quadruple word in which four-mora words are connected four by one at intervals of 1s, as shown in Fig. 8.2. The first word and the fourth word had a timing such that the words could be heard alone, but the second word and the third word always overlapped with another word in time. In this case, the preceding sound and the single echo can be both the target sound and the interference sound.

The subjects were nine students in their twenties (five males and four females).

The intelligibility for all words from the first word to the fourth word is shown in Fig. 8.3. Figure 8.3(a) shows the case in which a single echo comes from the horizontal plane, and Fig. 8.3(b) shows the case of the median plane.

When single echo comes from the horizontal plane, the intelligibility is high at an azimuth of approximately 90° and is low at the front and rear. This is the same tendency as the masking threshold. The results of a chi-square test showed statistically significant differences between 0° and 60° and between 0° and 90°.

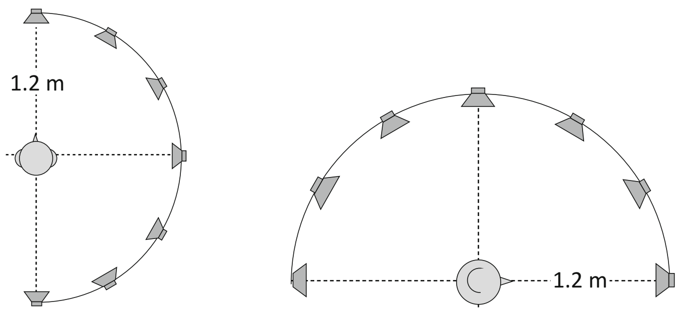

Fig. 8.1 Loudspeaker arrangement for a single echo

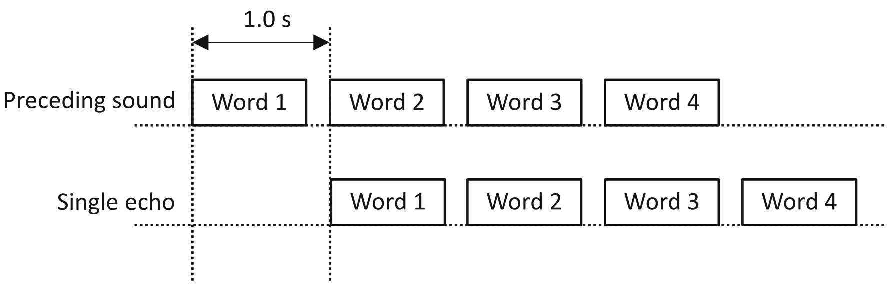

Fig. 8.2 Temporal structure of the speech stimuli

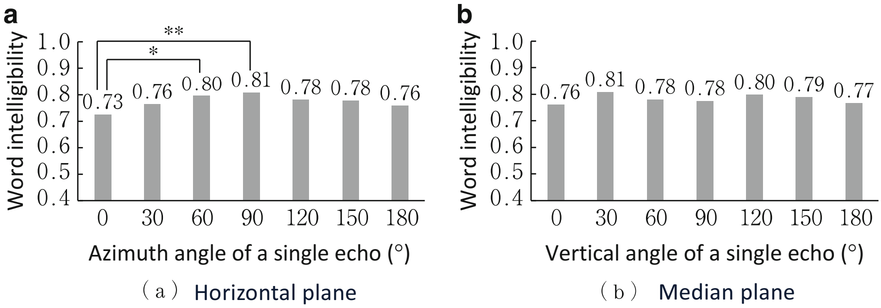

Fig. 8.3 Word intelligibility for each incident angle of a single echo. (**a**) horizontal plane and (**b**) median plane. $p < 0.05$, $p < 0.01$

On the other hand, when a single echo comes from the median plane, the effect of the incident rise angle is hardly noticeable.

Furthermore, word intelligibility was obtained for each order of word presentation, as shown in Figs. 8.4 and 8.5. The intelligibility of the first and fourth words, which can be heard alone, was high in both the horizontal and median planes, and the difference due to the incident direction was small.

In the horizontal plane, the intelligibility of the second and third words was higher at the side, as compared to the front. A significant difference was found for the second word.

In the median plane, there is a high intelligibility for an upward angle, as compared to the front. However, there is no commonality between the second and third words.

The above results suggest that the word intelligibility is influenced by the incident direction in the presence of a masker, the influence of which is qualitatively consistent with the characteristics of the masking threshold.

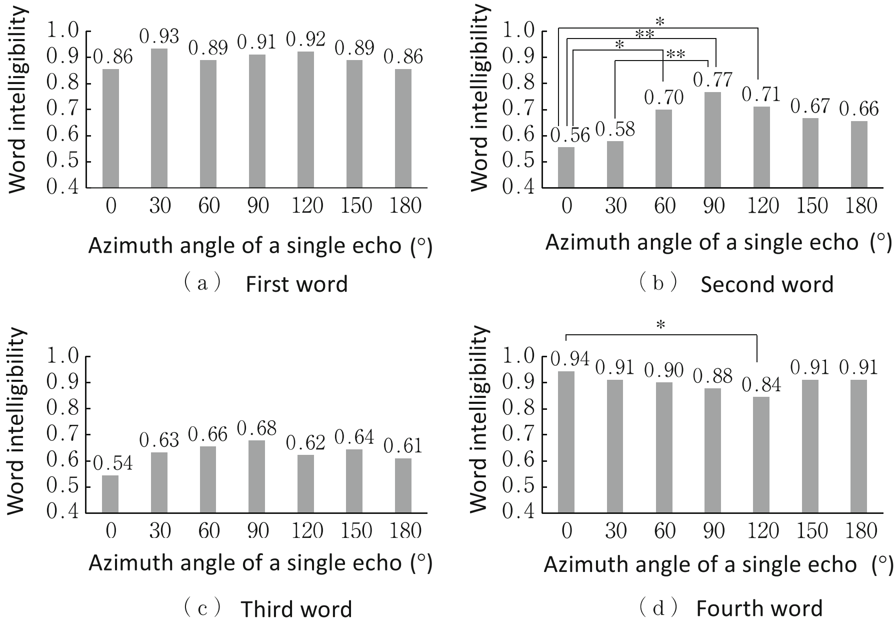

Fig. 8.4 Word intelligibility for each azimuth angle of a single echo. $*$ $p < 0.05$, $**$ $p < 0.01$

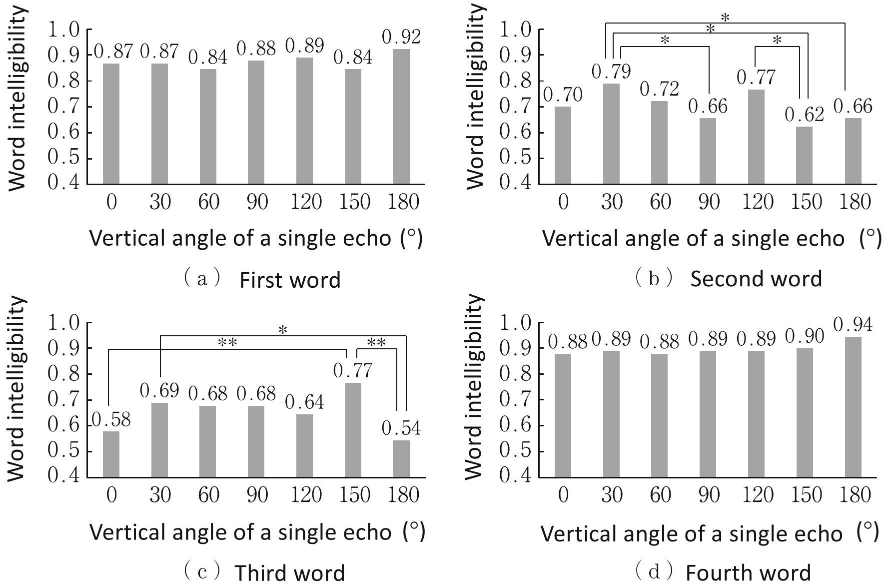

Fig. 8.5 Word intelligibility for each vertical angle of a single echo. $*$ $p < 0.05$, $**$ $p < 0.01$

References

Blauert J (1996) Spatial hearing–the psychology of human sound localization, Rev edn. The MIT Press, Cambridge, MA, pp 257–271

Saberi K, Dostal L, Sadralodabai T, Bull V, Perrott DR (1991) Free–field release from masking. J Acoust Soc Am 90:1355–1370