



HAL
open science

How Much is the Noise Level be Reduced? – Speech Recognition Threshold in Noise Environments Using a Parametric Speaker –

Noko Kuratomo, Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani, Keiichi Zempo

► To cite this version:

Noko Kuratomo, Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani, Keiichi Zempo. How Much is the Noise Level be Reduced? – Speech Recognition Threshold in Noise Environments Using a Parametric Speaker –. 18th IFIP Conference on Human-Computer Interaction (INTERACT), Aug 2021, Bari, Italy. pp.542-550, 10.1007/978-3-030-85616-8_31 . hal-04196877

HAL Id: hal-04196877

<https://inria.hal.science/hal-04196877v1>

Submitted on 5 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

How much is the noise level be reduced? – Speech Recognition Threshold in Noise Environments Using a Parametric Speaker –

Noko Kuratomo^{1,2}[0000–0002–1204–6677], Tadashi Ebihara³[0000–0001–5457–6783],
Naoto Wakatsuki³, Koichi Mizutani³[0000–0001–5789–5392], and
Keiichi Zempo³[0000–0003–2339–5298]

¹ Graduate School of Science and Technology, University of Tsukuba, Japan

² Research Fellow of Japan Society for the Promotion of Science, Japan

³ Faculty of Engineering, Information and Systems, University of Tsukuba, Japan
so.noko.sj@alumni.tsukuba.ac.jp, zempo@iit.tsukuba.ac.jp

Abstract. As a technology that allows sound to be heard only in specific areas in public spaces, the directional sound of parametric speakers has been attracting attention. The parametric speaker is also expected to have a wide range of uses, in settings ranging from shopping malls to museums for presenting information sound. The purpose of this study is to measure evaluate the Speech Recognition Threshold (SRT) of parametric speakers and compare with that of loudspeakers at various noise levels through experiments. It can also reveal how much the parametric speaker can reduce the noise heard by non-target users, which is unavoidable with loudspeakers.

Through experiments, the volume required to transmit an informative sound to a particular listener is about 12.61 dBA less than the level necessary for the loudspeaker. Moreover, the SRT of the surrounding people is about 2.48 dBA higher than that for the loudspeaker. Based on the results, the volume received by people in the surrounding area, including environmental noise, was calculated, and the effect of noise reduction is discussed. This study provides important insights for the flexibility in the design of the sound space.

Keywords: Soundscape · Speech Recognition · Parametric Speaker

1 Introduction

The primary objective of soundscape design for public places is to create a comfortable space. Sound is an essential factor in the spatial comfort [7, 17, 19], which can thus be increased by reducing noise levels. One of the methods often used is masking noise with comforting sounds. For example, there are studies have examined the effects of playing sounds of streams and birds in traffic noise [6, 15], and masking noise with the sound of fountains [14, 18]. Since outdoor noise is uncontrollable, masking it increases the comfort level. In contrast, sounds with

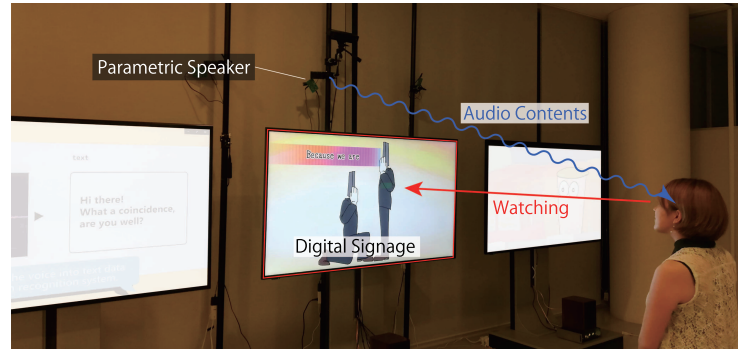


Fig. 1. One of the applications of the parametric speaker [9] : A system in which only the person viewing the advertisement can hear the sound by using parametric speakers improves the effectiveness of advertisements and spatial comfort.

controllable volume, such as announcements of exhibits in a museum or advertisements on digital signage, are noisy to people who are not interested in them. In other words, unlike traffic noise, controllable sound is perceived by both people who need to hear it and those who do not, and both types of listeners are in the same space. When sound is emitted by a loudspeaker, it is perceived as noise by those who do not need it, and their comfort level is decreased. If the sound is masked with a comforting sound as a countermeasure, the people who need to hear it will be unable to do so.

Therefore, the directional sound of parametric speakers has been attracting attention as a technology that allows sound to be heard only in specific areas in public spaces. A parametric speaker is a device that arrays ultrasonic waves while transmitting a signal that demodulates them into audible sound by using the nonlinearity of their propagation through the air, resulting in audible sound with a narrow directionality [4, 20]. Whereas conventional loudspeakers diffuse sound, parametric speakers emit sound in a straight line, which has the advantage of limiting the target area and causing less noise. Another feature of this system is that it is easier to localize than loudspeaker stereophonic sound due to reduced crosstalk [5].

One of the applications of the parametric speaker is the delivery of sounds that accompany advertisements on digital signage [9]. This study proposes a system in which only the person viewing the advertisement can hear the sound by using parametric speakers, as shown in Fig. 1. This system has been reported to improve the effectiveness of advertisements compared with those without sound while reducing noise and improving spatial comfort relative to the use of loudspeaker. However, the disadvantage is that it tends to be noisier than silence, and the advertising is registered by fewer people in the surroundings compared with a loudspeaker. That is, there is a trade-off between usage effectiveness and spatial comfort when using parametric speakers.

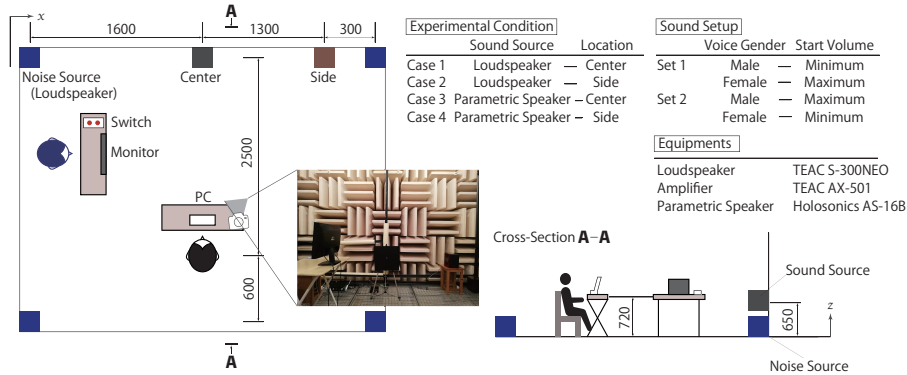


Fig. 2. The experimental environment and conditions.

Sound often contains linguistic information that the listener needs to understand, especially in situations using parametric speakers. The parametric speaker is also expected to have a wide range of uses, in settings ranging from shopping malls to museums. Therefore, it is important for the future sound design of public spaces to clarify how loud the emitted sound must be so that it can be understood in various noisy environments and what the noise level is at that time.

The purpose of this study was to quantitatively evaluate the degree of noise abatement. The Speech Recognition Threshold (SRT) of parametric speakers was measured and compared with that of loudspeakers at various noise levels through experiments. The threshold of understanding was calculated not only for the person receiving the spoken information but also for the people around them who were not meant to receive it. This information can be used to quantitate how much louder a parametric speaker must be compared with a loudspeaker when it is actually used in a public place. It can also reveal how much the parametric speaker can reduce the noise heard by non-target users, which is unavoidable with loudspeakers. This enables soundscape designers to evaluate the degree of contribution to spatial comfort. By examining environments ranging from quiet to noisy, such as museums and shopping malls, a variety of different conditions can be covered.

2 Experimental Environment

The experiment was conducted in an anechoic chamber to measure the SRTs of people who need information and those who do not. The experimental environment is described in Fig. 2. In the experiment, the sound was emitted by a parametric speaker and loudspeaker at various noise levels. For background noise, loudspeakers were placed in the four corners of the chamber to produce sound. The participants sat in chairs and judged whether they could hear the

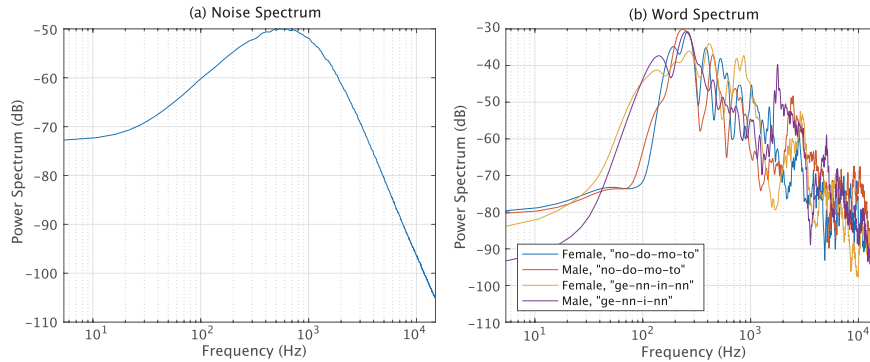


Fig. 3. Spectra – (a) Background noise, (b) Examples of the spoken word

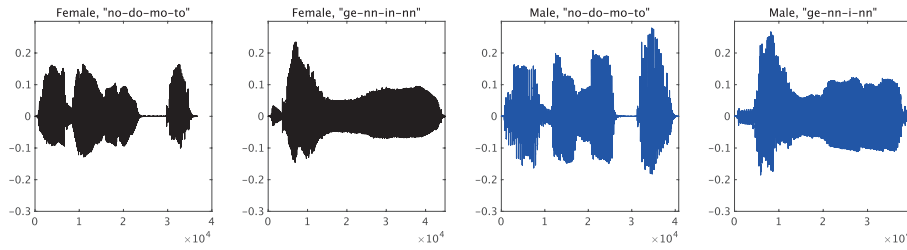


Fig. 4. Examples of the time plot of the spoken words : "nodomoto" means throat, and "genninn" means reason.

target sound. The loudspeaker and parametric speaker that emitted the target sound were to be installed in the center or at the sides. The speakers were placed in front of the participants for the center placement and out of their line of sight for the side condition, as shown in Fig. 2. A list of the equipment used is given in the table at the top right of Fig. 2.

The background noise was white noise with an FIR digital filter (according to ITU-T Rec. G.227). The spectrum of this noise is shown in Fig. 3 (a). The noise volume was set at 20–80 dBA, and was changed by increments of 10 dBA. The volume is defined as the average volume measured by a sound level meter set up where the participants were sitting. For volume calibration, a 1 kHz sine wave was used. The sound was calibrated to a sound pressure equivalent to that of the background noise or the target sound.

SRT measurement experiments need to use native languages with uniform syllable and parenthetical densities. If the target sound is not in the native language [13, 11, 12], or if the number of syllables or the degree of familiarity varies [1, 2], the recognition threshold will decrease and accurate data will not be obtained. Since this experiment was conducted entirely with Japanese participants, the FW07 dataset [10] in Japanese was used for the transmitted sound. This dataset contains words with four syllables and has been previously used in

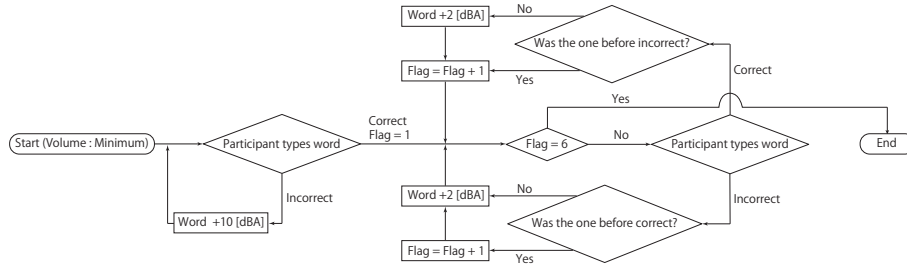


Fig. 5. Experiment procedure of the test started at the minimum volume.

SRT experiments by Japanese researchers [8, 16]. Since familiarity density [3] can also be selected, this experiment used a high familiarity word set to create a language environment close to that of daily life. Loudness normalization (ITU BS.1770-3 standard) was performed for all spoken words speech, and the volume was set to be the same. The time plot of the spoken words is shown in Fig. 4, and the spectrum is shown in Fig. 3.

The participants of this experiment were 14 undergraduate or graduate students (9 males and 5 females) with an average age of 22.64 years. The experiment was conducted after explaining the details of the experiment and obtaining the participants' consent in advance. Before the experiment, we conducted a sound field threshold measurement to confirm that the participants had normal hearing.

This experiment was reviewed and approved by the ethics committee of the University of Tsukuba (Faculty of Engineering, Information and Systems, Permission number: 2020R434).

3 Procedure

In the experiment, the words were played in an environment where background noise was always present. Participants typed the words that they heard on a PC. The researcher watched the input text by mirroring it to determine if it matched the actual words played and then used a switch to increase or decrease the volume. If the test started at the minimum volume, the sound was increased by 10 dBA every time a wrong answer was given. After that, the volume was decreased by 2 dBA when a correct answer was given and increased by 2 dBA after another wrong answer. Repeat these steps, and end the audio presentation if it turns into a wrong answer for the third time. These steps were repeated, and the audio presentation ended after a third wrong answer. This procedure was repeated but starting with the maximum volume, and the sound presentation was ended after the third correct answer. The volume at which the percentage of correct answers was 50% was set as the SRT, which was taken from the midpoint of the volume when the answer reversed from incorrect to correct or vice versa. The above experimental procedure is shown in Fig. 5.

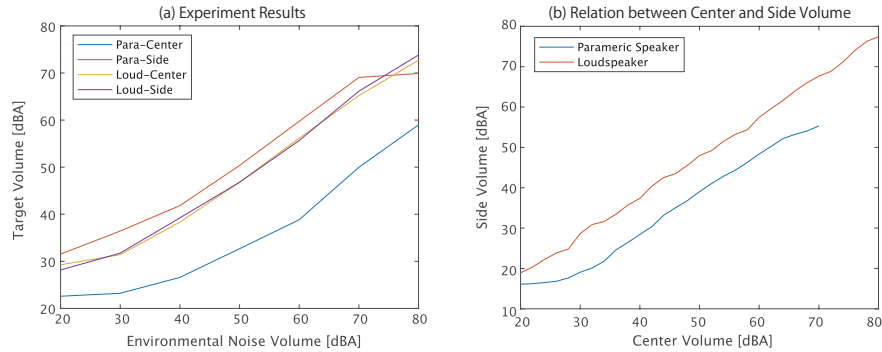


Fig. 6. (a) The SRT results for each speaker and location. High SRT is better for the Side because it is harder to hear; low SRT is better for the Center because it requires less volume.

(b) The volume flowing to the side when the center receives the sound. The smaller the Side is than the Center, the less sound will flow into the Side.

There were four experimental conditions: a loudspeaker or parametric speaker was used; the speaker was placed in a central or lateral location; the test started at the minimum or maximum volume; and the words were read out loud in a male or female voice. The experimental conditions and sound settings are shown in the right corner of Figure 2. The volume of the target sound is defined as the average volume measured by a sound level meter placed where the participants were sitting with the speaker placed in the center of the room. For example, a target volume of 40 dBA means that this is the volume that the participant heard when the speaker was in the center, and the speaker output the same volume when it was at the side. The volume range was 20–80 dBA for the loudspeaker and 20–70 dBA for the parametric speaker on the basis of equipment and ear protection.

4 Results and Discussion

The SRT results for each speaker and location are shown in the Fig. 6 (a). For the loudspeaker, the SRTs are almost the same between the center and side placements at all background volumes. The SNRs, which is the difference between the background noise volume and the SRT, and their average were calculated. The calculated results are about -1.44 dBA for the Center and -1.21 dBA for the Side. In other words, this implies that when the listener can understand the sound, the people around them can also hear the same content. This feature is appropriate for wide announcement applications. However, it is not suitable if sound is intended for only a specific individual or when the surrounding environment needs to remain quiet.

With the parametric speaker, the SRT is different between the center and side placements. The average SNR of the SRT is about -13.88 dBA for the Center and

Table 1. The calculated volume received by people in the center and surrounding area, including environmental noise, and the differences between loudspeaker and parametric speaker.

Background	Loudspeaker			Parametric Speaker			Diff
	SRT	Side	Total	SRT	Side	Total	
20	29.27	27.27	28.02	22.59	14.10	20.99	7.02
30	31.41	29.38	32.71	23.21	14.65	30.12	2.59
40	38.34	36.23	41.52	26.59	17.66	40.03	1.50
50	46.80	44.58	51.10	32.68	23.09	50.01	1.09
60	56.16	53.83	60.94	38.84	28.59	60.00	0.94
70	65.20	62.75	70.75	49.98	38.52	70.00	0.75
80	72.71	70.17	80.43	58.98	46.54	80.00	0.43

unit:[dBA]

1.27 dBA for the Side. The volume required to transmit an informative sound to a particular listener is about 12.61 dBA less than the level necessary for the loudspeaker. The results also show that people around the particular listener cannot hear much of the information in the sound. The phenomenon described as “whispering in the ear” by the participants of the experiment is thought to be due to the self-demodulating nature of parametric speakers as well as the fact that the SRT is quite small. Moreover, the SRT of the surrounding people is about 2.48 dBA higher than that for the loudspeaker. Furthermore, the volume flowing to the side when the center receives the sound is shown in Fig. 6 (b). The average volume heard by a person at the side is -2.26 dBA lower than that heard by the person at the center for the loudspeaker and -10.92dBA lower for the parametric speaker. There is a difference of about 8.66 dBA between the speakers.

Based on the above data, the volume received by people in the surrounding area, including environmental noise, was calculated. For example, in a 50 dBA environment, the SRT of the target person is 46.80 dBA for the loudspeaker. Based on a linear approximation of the volume relationship shown in Fig. 6 (b), the volume received by the people around the target was calculated to be 44.58 dBA. This means that the people around the target listener are exposed to a combined noise level of $50 \text{ dBA} + 44.58 \text{ dBA} = 51.10 \text{ dBA}$. In contrast, the volume level at which the advertisement sound can be heard in a 50 dBA environment with the parametric speaker is 32.68 dBA. In this case, the surrounding people will receive 23.09 dBA, as indicated in Fig. 6 (b). The overall noise level is $50 \text{ dBA} + 23.09 \text{ dBA} = 50.01 \text{ dBA}$. In other words, when changing from a loudspeaker to a parametric speaker, the people around the target listener hear about 1.09 dBA less sound. The same calculation reveals the differences in noise levels, as shown in Table 1. The smaller the background sound level, the larger the difference in the overall noise level. It can also be shown quantitatively that there is a difference at all background sound levels, as shown in Fig. 7. Compared with the loudspeaker, the parametric speaker is expected to improve the

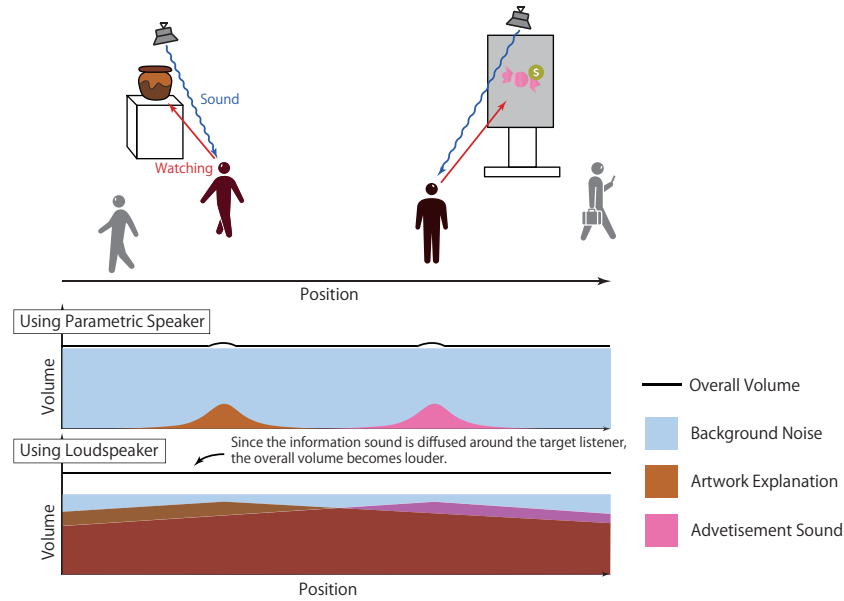


Fig. 7. Compared with the loudspeaker, the parametric speaker is expected to improve the spatial comfort level of non-target individuals, as they receive almost no unnecessary information. The overall volume is also reduced, which allows for more flexibility in the design of the sound space, such as adding other announcements or environmental sounds.

spatial comfort level of non-target individuals, as they receive almost no unnecessary information. The overall volume is also reduced, which allows for more flexibility in the design of the sound space, such as adding other announcements or environmental sounds. In particular, it is quantitatively confirmed that it is possible to create a space design in a quiet environment such as a museum that allows people to listen to explanations without interfering with other viewers.

As a limitation, the background noise in the experiment was filtered white noise for generalization. This is a noise with small amplitude variation, whereas in real life, the amplitude can vary greatly (e.g., road traffic noise and speech announcements). It is necessary to investigate how the SRT changes in that case, especially when the words are delivered by parametric speakers. In addition, the anechoic room is a free sound field with no reverberation. In real life, reverberation is also a factor, so its influence needs to be studied in the future.

References

1. Amano, S.: Effects of lexicon and coarticulation on phoneme perception. *Journal of the Acoustical Society of Japan (E)* **14**(2), 91–97 (1993)

2. Amano, S., Kondo, T.: Estimation of mental lexicon size with word familiarity database. In: Fifth International Conference on Spoken Language Processing (1998)
3. Amano, S., Kondo, T., Kato, K.: Familiarity effect on spoken word recognition in Japanese. In: Proceedings of the 14th ICPhS. vol. 2, pp. 873–876 (1999)
4. Aoki, K., Kamakura, T., Kumamoto, Y.: Parametric loudspeaker—characteristics of acoustic field and suitable modulation of carrier ultrasound. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* **74**(9), 76–82 (1991)
5. Aoki, S., Toba, M., Tsujita, N.: Sound localization of stereo reproduction with parametric loudspeakers. *Applied Acoustics* **73**(12), 1289–1295 (2012)
6. Hong, J.Y., Ong, Z.T., Lam, B., Ooi, K., Gan, W.S., Kang, J., Feng, J., Tan, S.T.: Effects of adding natural sounds to urban noises on the perceived loudness of noise and soundscape quality. *Science of the Total Environment* **711**, 134571 (2020)
7. Jo, H.I., Jeon, J.Y.: Effect of the appropriateness of sound environment on urban soundscape assessment. *Building and Environment* **179**, 106975 (2020)
8. Kobayashi, Y., Kondo, K.: Japanese speech intelligibility estimation and prediction using objective intelligibility indices under noisy and reverberant conditions. *Applied Acoustics* **156**, 327–335 (2019)
9. Kuratomo, N., Yamada, K., Masuko, S., Yamanaka, T., Zempo, K.: Effect of attention adaptive personal audio deliverable system on digital signage. In: SIGGRAPH Asia 2019 Technical Briefs, pp. 118–121 (2019)
10. Sakamoto, S., Yoshikawa, T., Amano, S., Suzuki, Y., Kondo, T.: New 20-word lists for word intelligibility test in Japanese. In: Ninth International Conference on Spoken Language Processing (2006)
11. Scharenborg, O., Coumans, J.M., van Hout, R.: The effect of background noise on the word activation process in nonnative spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **44**(2), 233 (2018)
12. Scharenborg, O., van Os, M.: Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Communication* **108**, 53–64 (2019)
13. Shimizu, T., Makishima, K., Yoshida, M., Yamagishi, H.: Effect of background noise on perception of English speech for Japanese listeners. *Auris Nasus Larynx* **29**(2), 121–125 (2002)
14. Skoda, S., Steffens, J., Becker-Schweitzer, J.: Road traffic noise annoyance in domestic environments can be reduced by water sounds. In: *Forum Acusticum* (2014)
15. Van Renterghem, T., Vanhecke, K., Filipan, K., Sun, K., De Pessemier, T., De Coensel, B., Joseph, W., Botteldooren, D.: Interactive soundscape augmentation by natural sounds in a noise polluted urban park. *Landscape and Urban Planning* **194**, 103705 (2020)
16. Yamamoto, K., Irino, T., Araki, S., Kinoshita, K., Nakatani, T.: Gedi: Gammachirp envelope distortion index for predicting intelligibility of enhanced speech. *Speech Communication* **123**, 43–58 (2020)
17. Yang, W., Kang, J.: Soundscape and sound preferences in urban squares: a case study in Sheffield. *Journal of Urban Design* **10**(1), 61–80 (2005)
18. Yang, W., Moon, H.J.: Effects of indoor water sounds on intrusive noise perception and speech recognition in rooms. *Building Services Engineering Research and Technology* **39**(6), 637–651 (2018)
19. Yang, W., Moon, H.J.: Combined effects of acoustic, thermal, and illumination conditions on the comfort of discrete senses and overall indoor environment. *Building and Environment* **148**, 623–633 (2019)

20. Yoneyama, M., Fujimoto, J.i., Kawamo, Y., Sasabe, S.: The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design. *The Journal of the Acoustical Society of America* **73**(5), 1532–1536 (1983)