Spatially target adaptive speech masking
- A pilot study on masking effect and annoyance -

C. T. Justine Hui[a], Moeto Ikuta[b], Mochinobu Obata[b], Yusuke Hioka[a,*], Takayuki Arai[b]

[a] *Acoustics Research Centre, Department of Mechanical Engineering, University of Auckland, Auckland 1142 New Zealand*
[b] *Faculty of Science and Technology, Sophia University, Tokyo 102-8554, Japan*

Abstract

The present paper proposes a speech masking system framework that *adapts* to the characteristics of the target speech. The characteristics are observed in three signal domains: temporal, spectral (frequency), and spatial, and are integrated into the design of the masker via how the masker is projected from loudspeakers. The effectiveness of the prototype design as a proof of concept was examined using a subjective listening test. Results show the proposed framework has potential to reduce the level of annoyance caused by the addition of masker to the environment while maintaining the effect of masking the target speech.

Keywords: Speech masking, masking effect, annoyance, spatially adaptive, microphone array

1. Introduction

Speech privacy concerns the protection of sensitive conversation from being overheard by third party audience. Lack of speech privacy has repeatedly been highlighted as the main cause of complaints in building acoustics where people share spaces with others, such as in open plan offices [1, 2, 3]. The problem exacerbates in spaces such as pharmacies and hospitals since it then becomes an ethical issue to keep conversation private to protect patients' confidentiality.

While various attempts have been made to protect speech privacy, a commonly used cost effective solution is utilising a speech masking system, where the masker (jammer sound) played via loudspeakers is used to "cover" the target speech (speech to be concealed), making the target speech unintelligible by triggering auditory masking. This technique has been implemented in some practical setups including scenario of open plan offices as well as pharmacies [4, 5, 6]. Because speech masking is based on auditory masking, the effectiveness of hiding conversation (masking effect) is improved when the masker's sound level is increased, resulting in a decrease in target to masker energy ratio (TMR). However, increasing the sound level of the masker causes annoyance [7, 8] both to the talkers involved in the confidential conversation as well as the unintended listeners due to the added noise to the environment by the masker itself. Hence, the relationship between masking effect and annoyance is deemed a trade-off, where the trade-off can be often compromised in studies on speech masking systems. A potential approach to address this challenge is by improving masking effect of the system while minimising the masker's sound level.

Some recent studies have proposed utilising a masker generated from the recordings of the target speech [9, 10, 11]. The idea behind this approach is to reflect the characteristics of the target speech to the masker design [6] so that the generated masker will be more effective, i.e. it will not require the masker being played at higher sound level to achieve the same effectiveness. However, this approach poses a potential risk that the confidential contents in the target speech may still be intelligible from the generated masker.

Speech is dynamic and unique to individuals, hence its characteristics fluctuate. For example, the sound level of speech varies continuously in time while the spectral shape of speech is dependent on the talkers. Such changes of characteristics have been reflected in masker design by e.g. temporally modulating [10] and spectrally shaping [6, 12, 13] the masker. In addition, when it is conversational speech, the dominant talker alternates resulting in variations in the spatial position of the dominant talker. Listeners are "released" from

*Corresponding author
Email address: yusuke.hioka@ieee.org (Yusuke Hioka)

masking when the target talker and masker are located in different angles, widely known as *spatial release from masking* [14]. This inspires us to hypothesise that rendering the masker from the same angle as that of the target talker would help avoid listeners to benefit from the spatial release from masking, which would subsequently contribute to increase the masking effect without increasing the sound level (i.e. annoyance). However, no attempt has been made to incorporate such spatial characteristics of target speech into speech masking systems.

The current study proposes an alternative design of speech masking system that *adapts* to the characteristics of target speech. Of the characteristics discussed, a particular focus is given to the use of spatial characteristics, which is enabled by acquiring the target speech using a microphone array. In order to minimise the effect of spatial release from masking, the angle of dominant target talker is monitored and used to control the position from which the generated masker is rendered. To avoid risking the confidential contents in the target speech, the proposed system generates a new masker from an arbitrary sound (i.e. not a clone of the target speech) by integrating the characteristics of the target speech. As a proof of concept study, a prototype of the proposed speech masking system was developed and verified through a subjective listening test.

2. Spatially target adaptive speech masking

2.1. Problem definition

Figure 1 shows a scenario considered in this study, where there are two talkers, talker A and B, having a conversation with highly confidential and/or private content. An unintended listener who should not understand the content of the conversation is located some distances away from the talkers. An array of microphones is located in the proximity of the talkers in order to capture the speech that is used to acquire their characteristics. Two loudspeakers projecting the masker are placed on a straight line connecting each talker to the listener. It is assumed that the talkers are located sufficiently apart from each other in terms of the angle with respect to the microphone array as well as the listener, and the angles of the talkers from the microphone array are known *a priori*. Real-time implementation of the system is left out of scope of the study.

2.2. Masker generation process

Figure 2 shows the masker generation process used in the proposed masking system. The process consists of three signal processing blocks to make the generated masker adapt to the characteristics of the target speech, namely spatial, spectral, and temporal adaptations. An arbitrary audio file is given as the *seed* of the generated masker (masker seed), which is processed in the spectral and temporal adaptation processes and routed to one of the loudspeakers, all of which are informed by the spatial adaptation process.

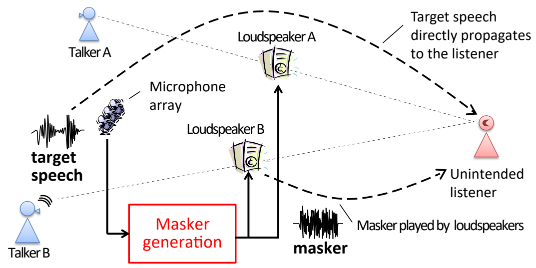

Figure 1: Proposed speech masking system and its problem setup.

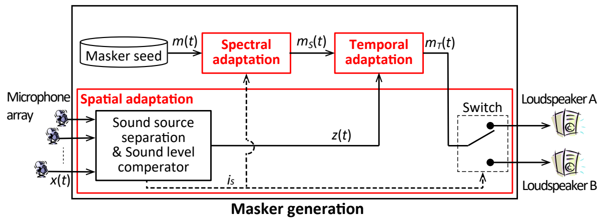

Figure 2: Masker generation process.

2.2.1. Spatial adaptation

In the proposed speech masking system, the spatial adaptation plays the key role in controlling the overall behaviour of the system. The sound signals received by the microphone array are first processed by a sound source separation algorithm that extracts the speech of the talker positioned in the angles known *a priori* as assumed in Section 2.1. Given that each of the separated signals that extracted the speech of talker A and B are denoted as $z_A(t)$ and $z_B(t)$, respectively, the sound level of $z_A(t)$ and $z_B(t)$ are compared and the louder signal is passed to the temporal adaptation process, as in (1):

$$z(t) = \begin{cases} z_A(t) & \phi_A(t) > \phi_B(t) \\ z_B(t) & \phi_A(t) < \phi_B(t) \end{cases}, \tag{1}$$

where $\phi_i(t)$ is the sound level of the talker i ($\in \{A, B\}$) measured at the position of the microphone array and t denotes the sample index of discrete time signals. The index of the selected talker i_S is passed to the spectral adaptation and the switch; the latter routes the output of the temporal adaptation process to the loudspeaker located in the direction of the dominant talker.

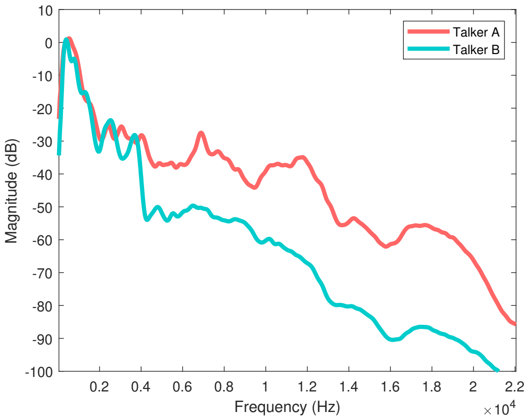

Figure 3: Normalised amplitude response of the FIR filters that follow the long-term average spectrum of talkers A and B.

2.2.2. *Spectral adaptation*

Generally, the effect of auditory masking increases when the masker has its spectral shape similar to that of the target speech [6]. Spectral adaptation manipulates the spectral shape of the masker seed in such a way that the masker's power spectral density (PSD) fits the PSD of the target speech in order to maximise the effect of auditory masking while minimising its sound level. A filter whose amplitude response resembles the spectral shape of the dominant speech i_S given by the spatial adaptation is designed and applied to the masker seed $m(t)$, providing the output $m_S(t)$.

2.2.3. *Temporal adaptation*

As the energy of speech fluctuates dynamically in time, the required energy for the masker to achieve sufficient masking effect also varies in time. The temporal adaptation varies the sound level of the spectrally shaped masker $m_S(t)$ to follow the energy of the dominant target speech $z(t)$.

2.3. *Prototype specifications*

A prototype of the proposed speech masking system was developed in order to conduct an initial proof of concept experiment. Some specifics of the prototype are summarised.

In the spatial adaptation, a sound source separation method using beamforming with Wiener postfilter framework was utilised to extract the signals $z_A(t)$ and $z_B(t)$. For the design of the postfilter, PSD estimation in beamspace technique [15] was utilised to estimate the PSD of the talkers. As the metric to evaluate sound level of the talkers, a value similar to the equivalent continuous sound pressure level (Leq) defined by $\phi_i(t) = \sum_{\tau=t-t_S}^{t} |z_i(\tau)|^2$ was used.

An FIR filter was used in the spectral adaptation which was designed by the frequency sampling based method that makes the filter's amplitude response follow the long-term average spectrum (LTAS) of the dominant talker's speech. The LTAS of each talker was acquired beforehand from 600 words spoken by the talkers and the frequency responses of the resultant FIR filters are shown in Figure 3.

For the temporal adaptation, a simple recursive algorithm similar to Welch's method [16] given by (2) is utilised for shaping the masker's waveform:

$$m_T(t) = \alpha m_T(t-1) + (1-\alpha) m_S(t) \frac{\sum_{\tau=t-t_T}^{t} z^2(\tau)}{\sum z^2(\tau)}, \quad (2)$$

where t_T is the number of time samples used for calculating instantaneous energy of a signal and α is a forgetting factor.

3. Experiment

Listening tests were conducted for measuring the masking effect and annoyance of the proposed speech masking system. The test compared five masker design modes as specified with different combinations of the adaptation processes summarised in Table 1. Masker design modes i) and ii) use only the existing adaptation processes (i.e. temporal and spectral adaptations) whereas iii) to v) have the proposed spatial adaptation process. Due to the lack of spatial adaptation process, the spectral and temporal adaptation processes in i) and ii) used the mixture of talker A and B. Namely, the spectral adaptation shaped the masker seed using the LTAS of both speakers collectively, and the temporal adaptation used the recording of one of the microphones in the microphone array for the signal $z(t)$.

3.1. *Stimuli*

For the target speech, Japanese words were taken from the FW03 corpus [17], which consists of 4-mora words ranked by word familiarity. Mora is a phonological unit used in Japanese, similar to a syllable in English. Only words in the highest familiarity list (5.5 - 7.0) were chosen. These target words were then incorporated into a carrier sentence: “kore kara kikoetekuruno ha [target] desu” (The word you are about to hear is [target]). The recordings were carried out by two male Japanese speakers of Tokyo dialect in their 20s specified as talker A and B. Two talkers of the same gender were

Masker design mode	Specifications
i) Temporal	Only temporal adaptation $m_S(t) = m(t)$; $m_T(t)$ routed to both loudspeakers
ii) Temporal + Spectral	Only temporal and spectral adaptation; $m_T(t)$ routed to both loudspeakers
iii) Spatial + Temporal	Proposed design without spectral adaptation
iv) Spatial + Spectral	Proposed design without temporal adaptation
v) Spatial + Temporal + Spectral	Proposed design

Table 1: Masker design modes compared in the listening test.

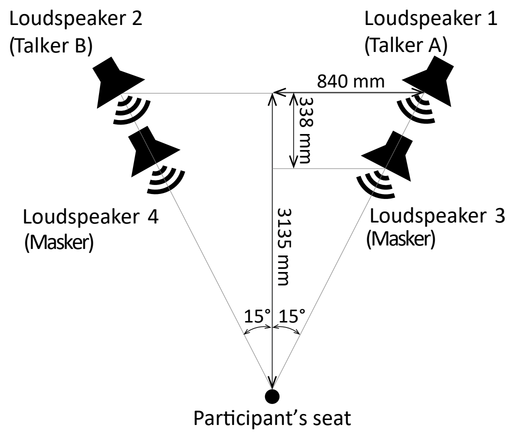

Figure 4: Loudspeakers and participant positions in the experiment.

chosen to eliminate any gender differences in annoyance perception that may have arisen from societal and psychological influences [18, 19, 20] (e.g. listeners finding voices from one gender more annoying or pleasant than the other).

For the masker seed $m(t)$, a water-based sound (stream) was chosen because previous studies have suggested water-based sound to be the most effective in terms of masking effect and annoyance compared to other commonly used maskers such as pink noise and music [21, 22, 23].

The target speech and maskers were played back using four loudspeakers (Genelec 8020A) via an audio interface (Roland OCTA-CAPTURE) in a sound-treated room, where two loudspeakers (1 and 2) were used to replicate the talkers A and B while two other loudspeakers (3 and 4) were used for playing the maskers as shown in Figure 4. The target speech and masker were sampled at 16 kHz and were presented at five TMRs: 0, -3, -6, -9, and -12 dB. Both t_T and t_S were set to 32000 (i.e. 2 seconds).

3.2. Test procedure

Fifteen self-reported normal-hearing university students (mean age 21.5, 10 female and 5 male listeners) were recruited at Sophia University, Tokyo Japan, to participate in the perception test. The participants were asked to listen to the sentences and write down the target word they heard within the carrier phrase. Each of the 4-mora target word was marked on a mora-basis, where a word would have a maximum of 4 scores if the participant recognised all 4 morae within a word correctly. The participants were also asked to rate the annoyance of listening to sentences under speech masking per trial using a scale ranging from 1 to 4, where 1: not annoying, 2: slightly annoying, 3: annoying and 4: very annoying. For each trial, participants listened to two sentences, one each spoken by talker A and B, respectively.

In total, the participants listened to 25 conditions (5 TMRs $\times$ 5 masker design modes), where each condition was tested using 4 target sentences (two from each talker). Within one masker design mode, the TMR varied in the order from 0 dB to -12 dB. The order of the masker design mode was randomised between participants. The target speech was played back consistently at 50 ± 1 dBA and the masker at 47 - 62 dBA, both at the participant's seat. The participants familiarised themselves with the test format and the sound played by the system through a practice round. The main test only proceeded once they felt comfortable with the test format. The test procedure was approved by the ethics committee of Sophia University.

3.3. Statistical analysis

The speech intelligibility results in terms of the percentage correct were analysed using a linear mixed effect model (LME) with the R package [24] *lme4* [25] and model fitting was carried out using the step function from *lmerTest* [26]. Interactions between two and more factors were included when it improved the fitness of the model. Significance in fixed effects was determined using a likelihood ratio test by comparing between a model with the effect in question and a model without the effect.

For speech intelligibility, `TMR` was treated as a discrete factor, `masker design mode` and `talker` (talker A and B) were included as fixed effects. A random slope of `talker` over `word` was included. `Participant` as a random factor was taken out of the model according to results from step function of *lmerTest*. For annoyance, `TMR`, `masker design mode` and `talker` were included as fixed effects and `participant` was included as a random effect. Post-hoc pairwise comparisons of the models were carried out using the *emmeans* package [27] with p-values adjusted using the Tukey method.

4. Results and Discussion

4.1. Masking effect

A masker is deemed effective if the listener could not understand the target words, therefore masking effect was assessed using the correct word recognition scores from the speech intelligibility test. Figure 5 displays the linear predictions of the intelligibility scores separated by the talkers, illustrating the differences between how listeners perceived the two talkers' target words under the different masker design modes. We found significant 2-way interactions between the `masker design mode` and `TMR` ($\chi^2(16) = 54.64$, $p < 0.0001$), between `masker design mode` and `talker` ($\chi^2(4) = 46.01$, $p < 0.0001$), and between `TMR` and `talker` ($\chi^2(4) = 21.86$, $p < 0.0001$). The two way interactions suggest that masking effect of the masker design mode was dependent on the talker who produced the target speech, while the effect of the TMR on speech intelligibility was also different depending on the talker.

From Figure 5, ii) was the least effective in masking either of the talkers, where listeners could understand the target words at 75% (3 out of 4 morae) even at -12 dB TMR, the lowest TMR level in the current study. The masker design mode iii) performed the best in terms of masker effectiveness and consistency with respect to talkers, i.e. masking both talkers' speech similarly. For example, iii) could achieve almost 50% speech intelligibility at 0 dB TMR. All other masker design modes required lower TMR to achieve the same level of masking effect or never achieved the same masking effect even at -12 dB TMR.

It is interesting to note that the results from the post-hoc analysis show the intelligibility scores of iii) to be significantly lower than those of i) at all TMR except -12 dB TMR ($p < 0.01$). This suggests the introduction of spatial adaptation in iii) contributes to improving masking effect. However, in contrast, the intelligibility scores of iv) and v), both of which also include spatial adaptation, are significantly higher than that of mode iii) at all TMR ($p < 0.01$ except the pair iii) - v) at TMR -3 dB being $p < 0.05$). It is notable that the difference in iv)

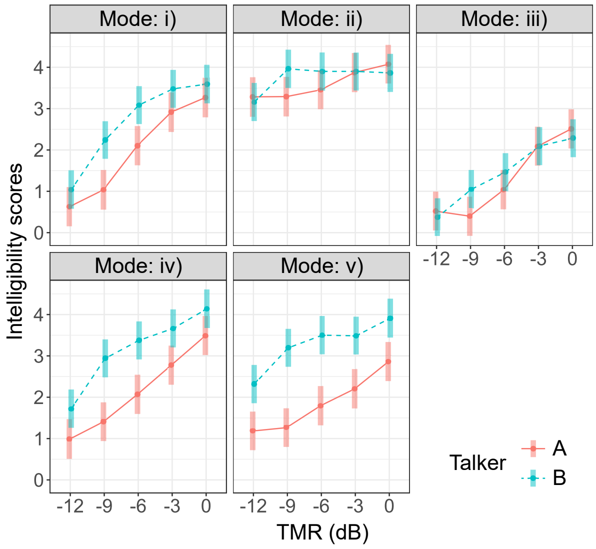

Figure 5: Linear predictions of intelligibility scores from the linear mixed model in terms of masker design modes. Error bars show the 95% confidence interval.

and v) compared to iii) is the fact that spectral adaptation is introduced. Previous studies imply spectral adaptation may become a "double edged sword" in terms of masking effect. For example, it is known that a speech-like masker (a masker that sounds like a speech rather than something else) is significantly more effective than other types of masker in terms of masking effect. The masker becomes more effective when it incorporates the characteristics of the target speech (e.g. [6]). However, care needs to be taken in its design process as some previous studies also suggest that the masking effect is reduced when the masker resembles the target speech too much (e.g. [9, 11]), as the masker starts to sound similar to the target speech. A previous study has looked in particular how resemblance of spectral shape and fundamental frequency (pitch) between masker and target speech affect masking effect [28]. The study found that the masking effect was higher when the spectral shape of masker was farther from that of the target speech while the pitch contour was more similar to the target speech. Although this study did not use the target speech as the masker seed, a similar detrimental effect would have been caused by the spectral adaptation.

4.2. Annoyance

Figure 6 shows the annoyance ratings for the TMR levels in terms of the five masker design modes with responses. As `talker` was not found to be a significant factor in the linear mixed effect model for the annoyance ratings, the combined results from talker A and B are shown and analysed together. Masker design mode differences can be observed, largely between ii) and the other modes, where listeners found ii) to be the least annoying, a trade-off with its masking effect.

As iii) was the most effective masker measured in speech intelligibility, we focus on the analysis of iii) in terms of its annoyance ratings. In Figure 6, iii) starts off as the most annoying masking design mode at 0 dB TMR compared to the other modes. Post-hoc analysis found iii) at 0 dB TMR to be significantly higher than ii) ($t(1114) = -7.41$, $p < 0.0001$) and iv) ($t(1114) = -4.04$, $p = 0.01$). The annoyance level for iii) however increased more gradually compared to the other masker design modes as TMR level decreased towards -12 dB TMR, where at -12 dB, iii) was only significantly different from ii) ($t(1114) = -4.72$, $p = 0.0007$).

We did not find any differences in annoyance ratings in terms of `talker`. As annoyance was rated for the masker and not the target talker, it is understandable that there were no differences between the talkers in terms of the annoyance ratings. More talkers in both genders should be included in future work to generalise the results in terms of both masking effect and annoyance of the proposed speech masking system.

It is worth pointing out that the result indicates the annoyance level was increased by the introduction of spatial adaptation as opposed to the hypothesis stated in Section 1. A plausible explanation for this trend would be the effect of other factors related to the sound source such as tonality and impulsivity [29]. While this finding alone is disappointing, as discussed in Section 1, the overall performance of speech masking system should be evaluated by taking into account the trade-off between masking effect and annoyance. This point will be discussed in the following Section 4.3.

4.3. Trade-off between masking effect and annoyance

An effective speech masking system requires a balanced trade-off between masking effect and annoyance; it should be able to mask targeted speech while keeping the users' annoyance within acceptable limits. We found ii) to be the least effective in terms of masking effect, but also the least annoying, suggesting a typical trade-off between masker effective and annoyance corroborating previous studies [9, 11]. Masker design mode iii) could mask both talkers approximately 50% (i.e. intelligibility score 2 out of 4) of the time at TMR 0 dB, where other maskers could only achieve a comparable performance at TMR -12 dB (i, iv and v) or not at all (ii). Considering the trade-off, the annoyance of iii) at 0 dB TMR was significantly lower than the annoyance ratings at -12 dB TMR of i) ($t(1114) = -6.7$,

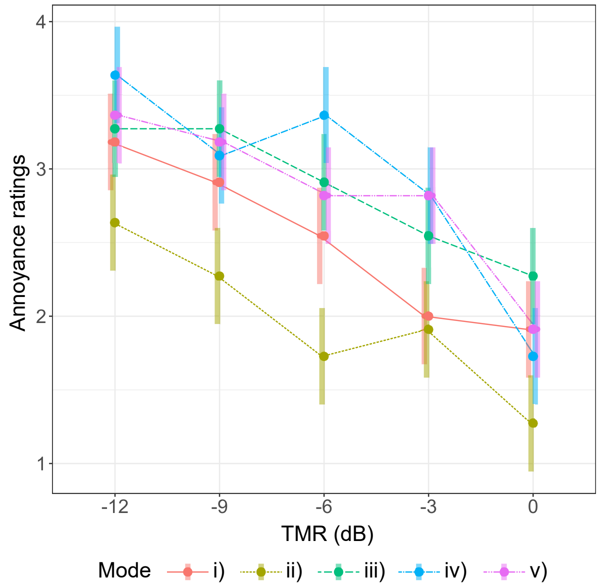

Figure 6: Linear predictions of annoyance ratings from the linear mixed model. Error bars show the 95% confidence interval.

$p < 0.0001$), that of iv) ($t(1114) = -10.1$, $p < 0.0001$), and that of v) ($t(1114) = -8.1$, $p < 0.0001$), and was not significantly different from that of ii) at -12 dB TMR. This shows that iii) achieved a better compromise of the trade-off between masking effect and annoyance. In other words, having the sources both spatially and temporally adapted produces a masker design superior to the other modes.

5. Conclusions

The current study proposed a speech masking system where the masker design adapts to the spatial characteristics of target speech in order to create a more effective masker, while maintaining annoyance at a reasonable level. As a proof of concept, a prototype speech masking system was developed and a subjective listening test was conducted by examining five masker design modes. We found the masker design mode which spatially and temporally adapted to the target talker to be superior when considering trade-off between masking effect and annoyance. Interestingly however, adding spectral adaptations using the talkers' LTAS reduced the masker's effectiveness. Future work involves further modifying the framework to be suitable for real-time processing as well as exploring an alternative approach to incorporate spectral characteristics of the target speech e.g. pitch information into the masker design.

Acknowledgements

We would like to thank our participants for participating in the perception test. This work was partly supported by Faculty Research Development Fund at the University of Auckland.

References

[1] E. Sundstrom, R. K. Herbert, D. W. Brown, Privacy and Communication in an Open-plan office: A case study, Environment and Behavior 14 (3) (1982) 379–392.

[2] W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, B. G. Watters, Speech Privacy in Buildings, The Journal of the Acoustical Society of America 34 (4) (1962) 475–492.

[3] J. Bradley, Acoustical design for open-plan offices, Construction Technology Update 63 (2004) 1–6.

[4] H. Lee, K. Ueno, S. Sakamoto, Subjective experiment on improvement of speech privacy in pharmacy, AIJ Journal of Technology and Design 20 (44) (2014) 165–168.

[5] Y. Kobayashi, K. Kondo, Bootstrap masker generation method for speech masking systems, Inter-noise 249 (7) (2014) 880–887.

[6] M. Akagi, Y. Irie, Privacy protection for speech based on concepts of auditory scene analysis, in: Internoise, no. 485, 2012.

[7] B. Berglund, U. Berglund, T. Lindvall, Scaling loudness, noisiness, and annoyance of community noises, The Journal of the Acoustical Society of America 60 (5) (1976) 1119–1125.

[8] Åsa Skagerstrand, S. Köbler, S. Stenfelt, Loudness and annoyance of disturbing sounds – perception by normal hearing subjects, International Journal of Audiology 56 (10) (2017) 775–783.

[9] Y. Hioka, J. W. Tang, J. Wan, Effect of adding artificial reverberation to speech-like masking sound, Applied Acoustics 114 (2016) 171–178.

[10] A. Krasnov, E. R. Green, B. Engels, B. Corden, Enhanced speech privacy in office spaces, Building Acoustics 26 (1) (2019) 57–66.

[11] Y. Hioka, J. James, C. I. Watson, Masker design for real-time informational masking with mitigated annoyance, Applied Acoustics 159 (2020) 107073.

[12] L. Calandruccio, S. Dhar, A. R. Bradlow, Speech-on-speech masking with variable access to the linguistic content of the masker speech, The Journal of the Acoustical Society of America 128 (2) (2010) 860–869.

[13] J. Donley, C. Ritz, W. B. Kleijn, Multizone Soundfield Reproduction with Privacy- and Quality-Based Speech Masking Filters, IEEE/ACM Transactions on Audio Speech and Language Processing 26 (6) (2018) 1037–1051.

[14] R. Y. Litovsky, Spatial release from masking, Acoustics today 8 (2) (2012) 18–25.

[15] Y. Hioka, K. Furuya, K. Kobayashi, K. Niwa, Y. Haneda, Underdetermined sound source separation using power spectrum density estimated by combination of directivity gain, IEEE Transactions on Audio, Speech, and Language Processing 21 (6) (2013) 1240–1250.

[16] P. Welch, The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics 15 (2) (1967) 70–73.

[17] S. Amano, S. Sakamoto, T. Kondo, Y. Suzuki, Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese, Speech Communication 51 (1) (2009) 76–82.

[18] A. Xu, S. S. Leung, A. Lee, Universal vs. language-specific aspects in human vocal attractiveness: An investigation towards Japanese native listeners' perceptual pattern, Proceedings of Meetings on Acoustics 29 (1) (2016).

[19] B. Borkowska, B. Pawlowski, Female voice frequency in the context of dominance and attractiveness perception, Animal Behaviour 82 (1) (2011) 55–59.

[20] D. A. Laird, K. Coye, Psychological measurements of annoyance as related to pitch and loudness, Journal of the Acoustical Society of America 1 (1) (1929) 158–163.

[21] A. Haapakangas, E. Kankkunen, V. Hongisto, P. Virjonen, D. Oliva, E. Keskinen, Effects of Five Speech Masking Sounds on Performance and Acoustic Satisfaction. Implications for Open-Plan Offices, Acta Acustica united with Acustica 97 (4) (2011) 641–655.

[22] V. Hongisto, J. Varjo, D. Oliva, A. Haapakangas, E. Benway, Perception of water-based masking sounds—long-term experiment in an open-plan office, Frontiers in Psychology 8 (2017) 1177.

[23] J. Cai, J. Liu, N. Yu, B. Liu, Effect of water sound masking on perception of the industrial noise, Applied Acoustics 150 (2019) 307 – 312.

[24] R Core Team, R: A language and environment for statistical computer (2015).
URL `https://www.r-project.org/`

[25] D. Bates, Linear mixed model implementation in lme4 (2007).

[26] A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest Package: Tests in Linear Mixed Effects Models, Journal of Statistical Software 82 (13) (2017).
URL `http://www.jstatsoft.org/v82/i13/`

[27] R. Lenth, emmeans: Estimated Marginal Means, aka Least-Squares Means (2019).
URL `https://cran.r-project.org/package=emmeans`

[28] T. Sannohe, T. Arai, K. Yasu, Sound masking system ni okeru database wo mochi-ita onsei masker sakuseihou no teian [A method of creating a speech masker using a database in a sound masking system], The Journal of the Acoustical Society of Japan 71 (8) (2015) 382–389.

[29] C. Marquis-Favre, E. Premat, D. Aubrée, Noise and its effects–a review on qualitative aspects of sound. part ii: Noise and annoyance, Acta Acustica united with Acustica 91 (4) (2005) 626–642.