Low-delay wind noise cancellation with preservation of spatial information for binaural hearing aids

Nobuhiko Hiruma¹,*, Hidetoshi Nakashima²,† and Yoh-ichi Fujisaka¹,‡

¹Rion Co., Ltd., 3–20–41 Higashimotomachi, Kokubunji, 185–8533 Japan
²National Institute of Technology, Kumamoto College, 2659-2 Suya, Koshi, Kumamoto, 861–1102 Japan

(Received 9 April 2018, Accepted for publication 2 August 2018)

Abstract: Wind noise annoys hearing-aid users, and it is hard to attach a windscreen to a hearing-aid microphone, for cosmetic reasons. Some hearing-aid devices reduce the low-frequency components of input signals by using high-pass filters to suppress the wind noise. Although wind noise can be attenuated by this approach, the perceived binaural information of the desired signals will also be degraded simultaneously, resulting in partial information loss. We had previously proposed a short-time fast-Fourier-transform-based (STFT-based) binaural wind noise cancellation algorithm that preserves binaural cues. This algorithm required a frame length of 32 ms to maintain a high frequency resolution. However, it is known that the tolerable group delay for mild hearing loss should be less than approximately 5 ms, in the high-frequency region. In this paper, we propose a low-delay binaural wind noise cancellation algorithm that uses a frequency-warping filter. The processing latency of this algorithm is shorter than the tolerable delay. The objective evaluation results — signal-to-noise ratios and perceptual evaluation of speech quality (PESQ) scores — were improved while maintaining a low latency. Subjective experiments demonstrated that the proposed method produced almost the same score as our previous STFT-based method, in terms of the directionality of output signals.

Keywords: Binaural hearing aids, Wind noise, Low delay, Frequency warping, HRTF

PACS number: 43.72.–q [doi:10.1250/ast.39.397]

1. INTRODUCTION

With the increase in the number of active hearing-aid users, opportunities for speech communication have been increasing, even in outdoor situations. Therefore, a hearing aid that can adapt to various environments, automatically and appropriately, is desirable. According to a survey by MarkeTrak VIII [1], the wind noise generated by the airflow around the head, i.e., around the pinna, past the tragus and helix [2,3], is a major cause of dissatisfaction among hearing-aid users. Generally, windproof equipment such as windscreens are utilized to reduce the noise [4]. However, as it is difficult to attach the windscreen to hearing aids without causing cosmetic issues, reducing the wind noise through signal processing is a more attractive alternative. Several algorithms using signal processing have been proposed to reduce the discomfort of wind noise [5–8]. In the case of single-channel wind noise cancellation algorithms, high-pass filtering has been utilized to reduce the wind noise components that mainly lie in the low-frequency region [9,10]. Although wind noise can be attenuated by the high-pass filter (HPF), the desired signals such as a speaker’s voice are also degraded simultaneously. Furthermore, when hearing aids on both the left and right ears work independently, the perception of the direction of arrival (DOA) and distance to the desired signal position will be uncertain. To preserve spatial information, a binaural wind noise cancellation algorithm using the characteristics of head-related transfer functions (HRTFs) was proposed by the authors [11–13].

In a previous study [13], we demonstrated the advantage of maintaining spatial information in the short-time fast-Fourier-transform–based binaural wind noise cancellation method (STFT-based method), through comparisons with the conventional method, by performing psychoacoustic experiments (externality and directionality tests). However, 512 samples were required for the frequency analysis, to obtain a frequency resolution of 31.25 Hz for a
The sampling rate of 16 kHz. The processing latency, which is the group delay of the system, was 32 ms. In addition, when considering the actual hearing aid system, the delay caused by the ADC and DAC was added to the processing latency, increasing the total delay. Thus, the processed sound would be perceived as an echo, especially, by a user who wears the device in an open-fitted style with a large vent [14]. This style is frequently used in the case of mild hearing loss, to prevent the occlusion effect that arises from the canal obstruction caused by the earpiece or ear mold. However, in this case, in addition to the processed sound from the hearing aid, the direct sound through the vent will also be perceived. Stone and Moore [15] showed that the tolerable group delay of the hearing aid for processed sounds must be less than approximately 5 ms. Therefore, the processing latency of our previous STFT-based system should ideally have been shorter than the tolerable group delay. However, when the number of signal samples was reduced in the Fourier analysis, the frequency resolution was degraded. As a result, effective wind noise suppression was not possible because the analysis frequency bandwidth was larger than that of the fine harmonic structure of speech.

In this paper, we propose a binaural wind noise cancellation algorithm that uses a frequency-warped filter (FWF) in the analysis/synthesis part instead of the STFT, in order to ensure that the latency is less than the tolerable group delay and to achieve the required frequency resolution.

2. BINAURAL WIND NOISE CANCELLATION SYSTEM

The frequency response from a sound source to both ears can be described by HRTFs. Interaural phase differences (IPDs) and interaural level differences (ILDs) are derived from these HRTFs [16], and are utilized in our proposed system shown in Figs. 1(a) and 1(b), where the horizontal axis shows the DOA of a sound source. It is apparent that the IPDs and ILDs vary depending on the DOA and frequency of sound. These characteristics are stored in the storage media of the hearing-aid system to create a database [17,18] for the detection of wind noise and determination of each gain, as described in subsections 2.4 and 2.5. Accordingly, the DOA can be estimated by finding similarities between the values from the database and the characteristics derived from the observed binaural signals. The database of each frequency band is \( \theta_{DB}(k, \psi) \) for IPDs and \( \xi_{DB}(k, \psi) \) for ILDs, where \( \psi \) represents the azimuth angle of a sound source.

2.1. Definition of Input Signal

Figure 2 shows a block diagram of the proposed system. In this system, the use of binaural hearing aids is assumed, for which microphones are placed at each ear position. The observed signals at the left and right sides, \( l(n) \) and \( r(n) \), at a discrete time \( n \), are defined as

\[
l(n) = s(n) * h_l(\phi, n) + w_{nl}(n),
\]

\[
r(n) = s(n) * h_r(\phi, n) + w_{nr}(n),
\]

where \( h_l(\phi, n) \) and \( h_r(\phi, n) \) provided the HRTFs from the DOA of a sound source \( \phi \), and \( w_{nl}(n) \) and \( w_{nr}(n) \) denote the wind noise signals in the discrete time domain. The asterisk symbol “*” represents the convolution operator. \( w_{nl}(n) \) and \( w_{nr}(n) \) are assumed to be uncorrelated so that the IPD and ILD for a wind noise produced from a turbulence around the head will be random.

2.2. Analysis and Synthesis

A block diagram of the analysis and synthesis using the FWF is shown in Fig. 3, where \( M \) represents the number of all-pass filters (APFs). The signal sequence \( p_l(0, n) \) to \( p_l(M, n) \) in Fig. 3 is a warped one, whose frequency components are warped from the low- to high-frequency region.

The all-pass filter is given by
where $\gamma$ is the warping parameter [19,20]. The frequency-warped sequence of the left channel is given by

$$ p_l(0,n) = l(n), $$

$$ p_l(m,n) = \gamma [p_l(m,n-1) - p_l(m-1,n)] + p_l(m-1,n-1), $$

where $m = 0, 1 \ldots M$, and that of the right channel, $p_r(m,n)$, is obtained by the same calculation. $p_l(m,n)$ and $p_r(m,n)$ are transformed from the time domain into the frequency domain using a discrete Fourier transform (DFT) to obtain the spectra $P_l(k,n)$ and $P_r(k,n)$, where $k$ is the index number of the frequency band.

Figure 4 shows the characteristics of the frequency warping from the original frequency to the warped frequency when $\gamma = 0.8$ is applied. It is observed that the frequency regions below 1000 Hz of the original frequency are warped to the high-frequency region.

2.3. IPD and ILD Calculation

The IPD of the input signals, $\theta_{pl}(k)$, is calculated as

$$ \theta_{pl}(k,n) = \tan^{-1} \left[ \frac{\text{Im}(C_{pl}(k,n))}{\text{Re}(C_{pl}(k,n))} \right], $$

where the cross spectrum of each warped input signal, $C_{pl}(k,n)$, is represented as

$$ C_{pl}(k,n) = P_l(k,n)P_r(k,n)^*, $$

where * indicates the complex-conjugate operator. Similar to the IPD calculation, the ILD of the input signals, $\xi_{pl}(k,n)$, is calculated as

$$ \xi_{pl}(k,n) = 20 \log \left| \frac{C_{pl}(k,n)}{C_{lr}(k,n)} \right|, $$

where $C_{lr}(k)$ denotes the power spectrum of the warped input signal $p_l(m,n)$.

2.4. Wind Noise Detection

Wind noise detection is performed by comparing
where \( \phi \) denotes the target source direction for the actual situation, and is selected by the hearing-aid user. The minimum value of \( \Delta \theta(k) \) and \( \Delta \xi(k) \) is 0, whereas the maximum value varies depending on the frequency, so that normalized differences are introduced as follows.

\[
D(k, n) = \begin{cases} 
\Delta \theta(k, n)P(k)^{-1} & \text{(for low freq. region)} \\
\Delta \xi(k, n)L(k)^{-1} & \text{(for high freq. region)},
\end{cases}
\]

where \( P(k) \) is the maximum difference of \( \theta_{DB}(k, \psi) \) when \( \psi \) varies, and is defined as

\[
P(k) = \max[\theta_{DB}(k, \psi)] - \min[\theta_{DB}(k, \psi)].
\]

The maximum difference of \( \xi_{DB}(k, \psi) \) is defined as

\[
L(k) = \max[\xi_{DB}(k, \psi)] - \min[\xi_{DB}(k, \psi)].
\]

It is assumed that wind noise exists in the frequency region \( k \) if \( D(k, n) > Tw(k) \), where \( Tw(k) \) is a predetermined wind detection threshold. Here, \( c_w(k, n) \) is a detection flag for wind noise in each frequency band and is described by

\[
c_w(k, n) = \begin{cases} 
1 & \text{(if } D(k, n) > Tw(k)) \\
0 & \text{(otherwise)}.
\end{cases}
\]

The wind counter \( \hat{c}_w(n) \), which determines the amount of cancellation for wind noise, is calculated as a summation of \( c_w(k, n) \) across time and frequency, as follows

\[
\hat{c}_w(n) = \sum_k \sum_i c_w(k, n - i).
\]

### 2.5. Gain Calculation and Filtering

The instantaneous binaural wind noise cancellation (BWNC) filter gain \( W_i(k, n) \) is defined as follows.

\[
W_i(k, n) = \frac{1}{\exp(\beta(k, n)D(k, n))},
\]

where \( \beta(k, n) \) is a gain-control parameter that varies depending on the wind counter \( \hat{c}_w(n) \) and is determined appropriately.

\[
\beta(k, n) \propto \hat{c}_w(n).
\]

Thus, \( \hat{c}_w(n) \) and \( \beta(k, n) \) will be large under windy conditions, and \( W_i(k, n) \) will approach 0. On the contrary, \( W_i(k, n) \) approaches 1 under calm conditions. To avoid sudden changes in the filter gain, the actual BWNC filter gain \( W(k, n) \) will be defined as

\[
W(k, n) = \lambda W(k, n - 1) + (1 - \lambda)W_i(k, n),
\]

where \( \lambda \) is a forgetting factor. The impulse response form in the time domain is obtained by the inverse Fourier transform of the filter coefficient,

\[
w(m, n) = \text{IDFT}[W(k, n)].
\]

The processed output signals,

\[
l(n) = \sum_{m=0}^{M} w(m, n)p_i(m, n),
\]

and

\[
l'(n) = \sum_{m=0}^{M} w(m, n)p_i(m, n),
\]

are obtained by convoluting both warped signals and the filter coefficient \( w(m, n) \).

### 3. NUMERICAL VALIDATION

#### 3.1. Numerical Configuration

A numerical test was performed to confirm the ability of wind noise detection and control of filter gain. The experimental conditions are summarized in Table 1. The number of APFs and the warping parameter were set to \( M = 31 \) and \( \gamma = 0.8 \) to satisfy the tolerable group delay of processing and higher frequency resolution in the low-frequency region. Therefore, the analysis frame length is totally 32 points, which consist of APF output the data of 31 samples and pass through data of 1 sample. The other parameters regarding the proposed algorithm were optimized by numerical experiments and trial listening.

Figure 5 shows an example of the group delay characteristics of the FWF, compared with the tolerable
group delay. The solid black line represents the tolerable
group delay derived from Ref. [15], as a function of
frequency. The solid red line represents the group delay
of the FWF. The FWF group delay is shorter than the
tolerable group delay.

The filter bank derived from the FWF in this simulation
is shown in Fig. 6. Here, 17 bands from the DC-band to
the Nyquist frequency are plotted. The frequency resolu-
tion in the low-frequency region, obtained by the FWF was
equivalent to that obtained by a 128-point FFT. Thus, FWF
realized fine resolution equivalent to that of 32-point FFT,
from the frame-length standpoint.

Figure 7 shows an image of the relative location from
among a group of possible locations for a listener, target speech source, and wind
noise. The target speech source is located at -60
degrees, while the wind noise flows from 0 degrees.
The signal-to-noise ratio (SNR) is set to 0 dB. Here, 0
degrees implies the front side of the listener.

noise sounds prerecorded using the binaural microphones
of the dummy head were used in our numerical validations.
The velocity of wind noise was approximately 4 m/s in the
recorded situation. There were two silent intervals in the
wind noise test sound, which were set to time periods of
2–3 and 5–6 s. The signal-to-noise ratio (SNR) of the input
signal was set to -10 dB for the observed signal of the left
channel when the speech source was located at 0 degrees,
i.e., in front of the hearing-aid listener. Note that we
assumed that the target speech direction $\phi$ was known.

3.2. Results of Validation

Figure 8 shows the results of the simulation. The
waveforms of the input signal (black solid line) and the
clean speech signal (red solid line) in the left channel (Lch)
are shown in Fig. 8(a). The output signal from the system
in Lch is shown in Fig. 8(b). The waveforms of the input
signal (black solid line) and the clean speech signal (red
solid line) in the right channel (Rch) are shown in Fig. 8(c). The
Rch signal output from the system is shown in Fig. 8.
Here, the SNR of the Rch was less than that of the Lch
because the speech signal arrived from -60 degrees,
relative to the left-hand side. The wind noise component in
the output signal was eliminated according to the wind
noise condition in Figs. 8(b) and 8(d).
Figure 9(a) shows the wind counter value $c^w$. Figure 9(b) shows the BWNC filter gains $W(k, n)$ for the 3rd, 9th, and 13th bands. It is clearly seen that the filter gain $W(k, n)$ for each frequency band was close to 1, when the wind noise disappeared; i.e., the BWNC features are disabled under calm conditions. On the other hand, the filter gain was controlled to reduce the noise gradually when wind noise appeared. These results revealed that the gain controller worked precisely, according to the wind noise conditions.

4. EVALUATIONS

4.1. Objective Evaluations

Two types of objective evaluations were performed to verify the effectiveness of the proposed method. The first one measured the amount of SNR improvement, and the second one measured the perceptual evaluation of speech quality (PESQ) using the wide-band version recommended in ITU-T P.862.2 [21].

First, the SNRs between speech and wind noise levels for the left channel were set to 0 dB and $-10$ dB under the conditions that wind flow arrived from the front side of the dummy head. In this evaluation, the ITU-T Recommendation P.501 test signal [22] was utilized as a speech source.

Then, the direction of speech stimulus was varied from $-90$ to $+90$ degrees in steps of 30 degrees. At this time, the wind noise level was maintained at the initial setting level. Therefore, the SNRs of each channel varied according to the DOA of the target speech. In our evaluation, by using the above configuration, the proposed FWF-based method was compared with the previously proposed STFT-based method with $32$ ms (STFT32 method) and $4$ ms (STFT4 method) frame lengths. Note that the FWF is equivalent to the STFT4 from the frame-length standpoint. It is assumed that the target speech direction was known.

4.1.1. Evaluation by SNR

The evaluation results for SNR improvement are shown in Fig. 10. In this figure, the horizontal and vertical axes indicate the DOA for the target speech and the amount of SNR improvement, respectively. Figure 10(a) and (b) show the results of the FWF method (red solid line), STFT4 method (blue solid line), and STFT32 method (black solid line) for both conditions: $0$ dB and $-10$ dB SNR. It is apparent that the SNR of the FWF is improved when comparing with STFT4, which is equivalent to the FWF from the frame-length standpoint.

4.1.2. PESQ scores

The measurement results for the PESQ scores are shown in Figs. 11(a) and 11(b), as a function of DOA of the target speech in each SNR condition. The scores of the proposed FWF method were higher than those of the STFT4 method in all DOA, in spite of the fact that the FWF method has the same frame-length as the STFT4 method. On the other hand, the improvements of the PESQ score by not only the proposed method but also by STFT4 were smaller compared with the SNR improvements shown in Figs. 10(a) and 10(b). This tendency is similar to results obtained by the other speech enhancement methods [23,24]. This implies that the PESQ score is not improved drastically by the speech enhancement method. Note that
PESQ score by the proposed method is higher than that of STFT4 and is the result of SNR improvement. This is because it can be considered that the fine frequency resolution of the low-frequency region is greatly affected for the noise estimation and gain reduction processing in wind noise.

4.2. Subjective Evaluation

To confirm the preservation abilities of the information of binaural cues, two types of subjective evaluations — “directionality” and “externality” tests — were performed for our proposed FWF-based method. These “directionality” and “externality” tests were carried out to evaluate the influence on the perception of the arrival direction of sound and retention of spatial information, respectively. The voice signals with directional information, which were the target signals in the directionality test, were generated by applying HRTFs from $-90$ to $+90$ degree in steps of 30 degrees, i.e., evaluations were performed in seven directions. In addition, the wind noise arriving from the direction of 0 degrees, which was prerecorded by the KEMAR, was added to the target signals to reproduce windy conditions. Such a signal is defined as windy condition voice in this section. The test was carried out in an anechoic chamber, and the intended stimuli were presented through headphones (Sennheiser HD-600). For these subjective evaluations, a stimulus was created using these signals.

4.2.1. Directionality test

Two types of signals, which were of the 4-mora Japanese word corpus [25] type, were utilized to create a stimulus set in this evaluation. The first one was the unprocessed target signal and the second one was the processed signal derived from the windy condition voice by applying the FWF-based method. These audio data were applied to each stimulus set. The stimuli in the set were presented to subjects at random in this subjective evaluation unit that included nine trials per angle. Three normal hearing subjects between 19 and 21 years of age were recruited for this experiment.

The subjects were asked to answer the perceived directions of the 4-mora words. Although the stimuli were presented from seven directions between $-90$ and $+90$ degrees, the answers for their perceived directions were selected on a scale of 10 degree steps.

Figure 12 shows the results of the directionality test for (a) target signal and (b) FWF-based method. It indicated that applying the FWF-based method did not lead to any negative effect on the perceived direction, similar to the previously proposed STFT-based method [13], as can be seen from the Figs. 12(a) and 12(b).

4.2.2. Externality test

From the results of the objective evaluations in Sect. 3, the proposed FWF-based method was superior to the STFT4 method under all conditions. In addition, the STFT32 method was not applicable to hearing aids from the point of view of processing latency. On the other hand, the HPF method [3,6] has been proposed to realize a low-computational cost by using a simple method as far as possible to reduce the discomfort in user’s ears from wind noise; however, there are some concerns regarding this. For example, despite the knowledge that the phase information of the low-frequency region is efficient for the perception of externalization [26], low-frequency components of the desired signal will be significantly degraded when using the HPF method. From the above-mentioned observations, the purpose of this externality test is to evaluate how much to retain the spatial information for the proposed FWF method compared with the HPF method as a competitor.

In this experiment, four Japanese sentences [27] with wind noise were utilized, and then, these windy condition sentences were processed by in each FWF-based method and a conventional HPF method, which is a competitor. These processed signals were used as stimuli and were presented to subjects randomly, similar to the previous
experiment using the pair-comparison method. Five normal hearing subjects between 19 and 21 years of age were recruited for the experiment. The question to the subjects was ‘Which one is more distant from around your head?’ The subjects had to select an answer from the ‘first one’, ‘second one’, or ‘no difference’. The number of trials was one per angle for one sentence. Thus, the total amount of trials was 28 in this experiment.

Figure 13 shows the results of the externality test. The horizontal and vertical axes show the presented angle of the stimulus and the proportion of the selected answers, respectively. The results of all directions in the externalization of the sound image implied that the proposed FWF-based method was superior to the conventional HPF method.

5. CONCLUSION

In this paper, we proposed a low-delay wind noise cancellation algorithm for binaural hearing aids. It was difficult to implement the previously proposed algorithm [13], which required long analysis frame lengths to ensure high resolutions in the low-frequency region, in actual hearing aids from the viewpoint of processing latency for sound transmission from the input to the output. Applying the FWF to the analysis and synthesis stages of our system cleared the way for a trade-off relationship between processing latency and frequency resolution. In the results of numerical evaluations, the proposed system could reduce the wind noise components effectively while maintaining the processing latency of the system shorter than the tolerable group delay. The system had a positive effect on the SNRs and PESQ scores for various DOAs, compared to the STFT-based method with a similar analysis frame length, in the objective evaluations. In addition, in the subjective evaluation tests for sound localization, it was determined that binaural information was preserved in our system, which was superior to the performance of the conventional HPF method. In conclusion, the processing latency of the proposed BWNC system could be made shorter than the tolerable group delay, while maintaining the localization cue of the target signal by using an FWF.

In this paper, the HRTFs provided by MIT were utilized for creating the database and directional target speech source. In a future work, we plan to build a prototype of the binaural hearing-aid device and evaluate the variance of the gain and accuracy of wind noise cancellation for individual users.

REFERENCES

[1] S. Kochkin, “MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing,” Hear. J., 63, 19–32 (2010).
[2] H. Dillon, I. Roe and R. Katsch, “The sources of wind noise in hearing aids,” in Int. Hear. Aid Res. Conf. (2000).
[3] J. M. Kates, “Wind noise,” in Digital Hearing Aids (Plural Publishing Inc., San Diego, 2008), pp. 147–173.
N. HIRUMA et al.: LOW-DELAY BINAURAL WIND NOISE CANCELLATION

[4] G. W. Elko and J. Meyer, “An electronic windscreen for microphones,” J. Acoust. Soc. Am., 119, 3445 (2006).

[5] G. W. Elko, “Reducing noise in audio systems,” US Patent 7 171 008 (2007).

[6] J. M. Kates, “Hearing aid with suppression of wind noise,” US Patent 8 019 103 (2007).

[7] K. S. Petersen, G. Bogason, U. Kjems and T. B. Elmedyb, “Device and method for detecting wind noise,” US Patent 7 340 068 (2008).

[8] J. A. Zakis and C. M. Tan, “Robust wind noise detection,” Proc. IEEE Int. Conf. Acoust. Speech Signal Process., pp. 3683–3687 (2014).

[9] K. Chung, L. Mongeau and N. McKibben, “Wind noise in hearing aids with directional and omnidirectional microphones: polar characteristics of behind-the-ear hearing aids,” J. Acoust. Soc. Am., 125, 2243–2259 (2009).

[10] J. A. Zakis, “Wind noise at microphones within and across hearing aids at wind speeds below and above microphone saturation,” J. Acoust. Soc. Am., 129, 3897–3907 (2011).

[11] N. Hiruma, H. Nakashima and Y. Fujisaka, “Binaural wind noise cancellation algorithm in hearing aids using HRTF cues,” in Int. Hear. Aid Res. Conf. (2014).

[12] H. Nakashima, R. Koyama, N. Hiruma and Y. Fujisaka, “Binaural wind noise detection, cancellation and its evaluation for hearing aids based on HRTF cues,” in Proc. IEEE Int. Conf. IEEE Ind. Electron. Soc., pp. 4896–4899 (2015).

[13] H. Nakashima, R. Koyama, N. Hiruma and Y. Fujisaka, “Binaural wind noise cancellation for hearing aids based on head-related transfer function cues,” Acoust. Sci. & Tech., 39, 40–43 (2018).

[14] H. Dillon, Hearing Aids (Boomerang Press, Turramurra, Australia, 2001), pp. 127–149.

[15] M. A. Stone, B. C. J. Moore, K. Meisenbacher and R. P. Derleth, “Tolerable hearing aid delays. V. estimation of limits for open canal fittings,” Ear Hear., 29, 601–617 (2008).

[16] B. Gardner and K. Martin, “HRTF measurements of a KEMAR dummy-head microphone,” http://sound.media.mit.edu/resources/KEMAR.html (Accessed 2018-03-29).

[17] H. Nakashima, Y. Chisaki, T. Usagawa and M. Ebata, “Frequency domain binaural model based on interaural phase and level differences,” Acoust. Sci. & Tech., 24, 172–178 (2003).

[18] Y. Chisaki, S. Kawano, K. Nagata, K. Matsuo, H. Nakashima and T. Usagawa, “Azimuthal and elevation localization of two sound sources using interaural phase and level differences,” Acoust. Sci. & Tech., 29, 139–148 (2008).

[19] A. Härnäs, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine and J. Houpaniemi, “Frequency-warped signal processing for audio applications,” J. Audio Eng. Soc., pp. 1011–1031 (2000).

[20] J. M. Kates and K. H. Arehart, “Multichannel dynamic-range compression using digital frequency warping,” EURASIP J. Appl. Signal Process., 18, 3003–3014 (2005).

[21] “ITU-T Recommendation P.862.2, Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codes,” ITU (2007).

[22] “ITU-T Recommendation P.501, Test signals for use in telephonometry,” ITU (2012).

[23] C. H. You, S. N. Koh and S. Rahardja, “Signal subspace speech enhancement for audible noise reduction,” Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Vol. 1, pp. 145–148 (2005).

[24] M. Parchami, W.-P. Zhu, B. Champagne and E. Plourde, “Bayesian STSA estimation using masking properties and generalized Gamma prior for speech enhancement,” Eurasip J. Adv. Signal Process., 87, pp. 1–21 (2015).

[25] S. Amano, K. Kondo, Y. Suzuki and S. Sakamoto, “Speech data set for word intelligibility test based on word familiarity (FW03),” NII Speech Resources Consortium, http://research.nii.ac.jp/src/en/FW03.html (Accessed 2018-03-29).

[26] W. M. Hartmann and A. Wittenberg, “On the externalization of sound images,” J. Acoust. Soc. Am., 99, 3678–3688 (1996).

[27] “Speech perception assessment test for cochlear implant users (CI2004),” Escor Co., Ltd. http://escor.co.jp/products/products_item_CI2004.html (Accessed 2018-03-29).