# Low-Latency Real-Time Blind Source Separation with Binaural Directional Hearing Aids

*Masahiro Sunohara*[1], *Chiho Haruta*[1], *Nobutaka Ono*[2]

[1]RION Co., Ltd., Japan
[2]National Institute of Informatics, Japan

suno@rion.co.jp, c-haruta@rion.co.jp, onono@nii.ac.jp

## Abstract

Understanding desired speech in noisy environments is one of the important issues for hearing aid systems, which require a strong real-time property. Recently, the authors have proposed a low-latency algorithm for real-time blind source separation (BSS) based on online auxiliary-function-based independent vector analysis (AuxIVA) by the truncation of non-causal components of time-domain demixing impulse responses, and we evaluated the separation performance using omnidirectional binaural microphones. On the other hand, directional microphones have been widely used for hearing aids to improve the signal-to-noise ratio. In this paper, the effects of the truncation of demixing impulse responses is investigated when using the proposed algorithm with binaural directional microphones. By experimental evaluation using a head and torso simulator in a real environment, the performance of the proposed algorithm with directional microphones in the case of 10 ms latency is 9.0 dB in terms of the signal-to-interference ratio (SIR), which is only a 2.2 dB performance loss from the case of 128 ms latency. Moreover, the performance with directional microphones is about 1.0 dB better than that with omnidirectional microphones.

**Index Terms**: hearing aids, directional microphone, blind source separation, independent vector analysis, low latency

## 1. Introduction

Hearing-impaired people have difficulties communicating with others even if they wear hearing aids, especially in a noisy environment such as a party venue or a crowded restaurant. Improving speech communication in such difficult situations is one of the most challenging issues to be solved for hearing aids. As a technique for solving these problems, blind source separation (BSS) may be applicable [1, 2, 3]. BSS is a signal processing method that can extract a desired sound source from a mixture by using multiple microphones without requiring information on the source signals. In the frequency-domain approach for convolutive BSS, independent vector analysis (IVA) has been proposed as a technique that does not require the solution of a permutation ambiguity problem [4, 5, 6]. Furthermore, auxiliary-function-based IVA (AuxIVA) has been proposed as a state-of-the-art approach with rapid convergence and a low calculation cost [7, 8, 9].

To apply BSS as an application of binaural hearing aids, which are real-time systems, it is important to reduce the latency from the input to the output of the system [10, 11]. In addition to computational complexity, an algorithm may require an inherent delay, which is referred to as an algorithmic delay. In the case of frequency-domain BSS, a delay of at least one frame length is necessary for frame analysis [12]. Although several real-time implementations of IVA have been proposed [13, 14], this delay is unavoidable. Such a large delay causes various problems in a hearing aid system such as difficulty in speaking owing to the delayed auditory feedback effect or a sense of discomfort due to the loss of lip synchronization [15].

Recently, the authors have proposed a low-latency algorithm for real-time BSS based on online AuxIVA for hearing aids [16]. This proposed algorithm can significantly shorten the algorithmic delay by the time-domain implementation of demixing matrices as FIR filters and the truncation of part of their non-causal components. Generally, the truncation of the non-causal components should degrade the separation performance. However, if all the non-causal components of the demixing impulse response are originally zero, these components can be truncated without performance degradation. In our previous work [16], we confirmed that the proposed system with an algorithmic delay of within 10 ms worked with little performance degradation by experimental evaluation using binaural behind-the-ear (BTE)-type hearing aids consisting of omnidirectional microphones.

However, bilateral directional microphones have been widely used in actual hearing aids to improve the signal-to-noise ratio of front speech signals in a noisy background [17]. Thus, in this paper, the separation performance of binaural BSS based on the low-latency online AuxIVA algorithm with directional microphones was investigated as a more practical verification.

## 2. Low-latency real-time BSS

### 2.1. Overview of online AuxIVA

We assume that $K$ sources are observed by $K$ microphones and that their short-time Fourier transform (STFT) representations are known. Let $\boldsymbol{x}(\omega,\tau) = [x_1(\omega,\tau) \cdots x_K(\omega,\tau)]^t$ be the vector representations of the observation signal in the $(\omega,\tau)$th time-frequency bin, where $^t$ denotes the vector transpose. In the frequency domain, the sources are estimated by the following linear demixing process:

$$\boldsymbol{y}(\omega,\tau) = W(\omega;\tau)\boldsymbol{x}(\omega,\tau), \tag{1}$$

where $W(\omega;\tau) = (\boldsymbol{w}_1(\omega;\tau) \cdots \boldsymbol{w}_K(\omega;\tau))^h$ is a demixing matrix, $^h$ denotes the Hermitian transpose, and $\boldsymbol{y}(\omega,\tau) = [y_1(\omega,\tau) \cdots y_K(\omega,\tau)]^t$ represents the estimated sources.

An online AuxIVA algorithm is an effective means of estimating the demixing matrices $W(\omega;\tau)$ in the $(\omega,\tau)$th time-frequency bin under dynamic conditions [14]. The algorithm consists of alternate update rules, which update the weighted covariance and the demixing matrix. In this paper, we focus on the case of $K = 2$ for application to hearing aids.

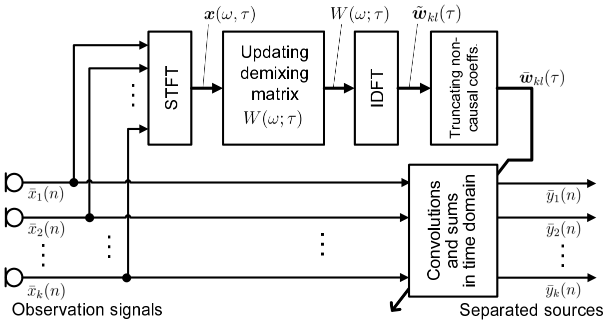


Figure 1: *Signal block diagram of low-latency real-time online independent vector analysis.*

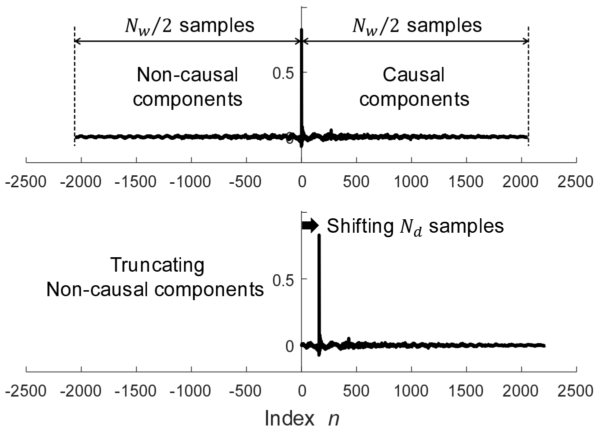


Figure 2: *Time-domain demixing impulse responses. Upper: original response $\tilde{w}_{kl}(n;\tau)$. Lower: shifted and truncated response $\bar{w}_{kl}(n;\tau)$.*

### 2.2. Realization as quasi-causal FIR filter

Figure 1 shows a signal block diagram of a low-latency version of the online AuxIVA algorithm [16]. A means of shortening the delay is to form two paths, one for updating the demixing matrices in the frequency domain and the other for separating the sources using FIR filters in the time domain. After applying back-projection [18], the frequency-domain demixing matrix $W(\omega;\tau)$ is converted to coefficients of multiple time-domain FIR filters $\tilde{w}_{kl}(n;\tau)$ using the inverse discrete Fourier transform. This structure can shorten the algorithmic delay to half of the frame length ($N_\omega/2$ samples). To further shorten the algorithmic delay, the coefficients of only $N_d$ non-causal components are shifted and the other components are truncated as shown in Fig. 2. After that, the algorithmic delay of the system becomes only $N_d$ samples.

If all the non-causal components of $\tilde{w}_{kl}(\tau)$ are originally zero, the algorithmic delay of the system can theoretically be zero without performance degradation. For the simple model consisting of two sound sources and two observations shown in Fig. 3, a theoretical sufficient condition for the ideal separation filters to be causal is obtained as the following inequality [16]:

$$[\log a(\theta_2) - \log a(\theta_1)] \cdot [\tau(\theta_2) - \tau(\theta_1)] < 0, \tag{2}$$

where $a(\theta_k)$ and $\tau(\theta_k)$ are respectively the amplitude ratio and the time difference of the second channel relative to the first channel for source $k$ with direction $\theta_k$.

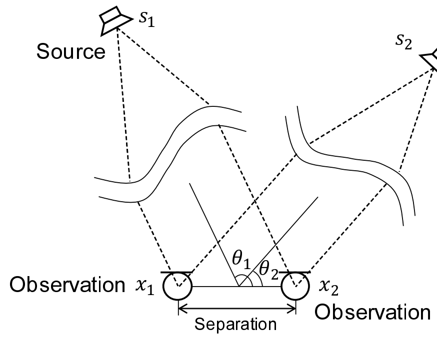


Figure 3: *Simple model consisting of two sound sources and two observations.*

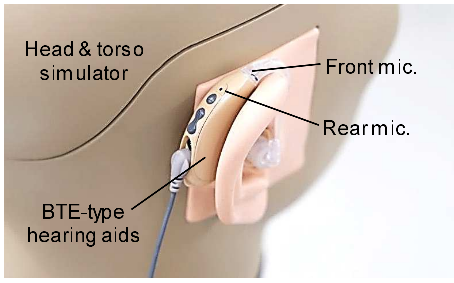


Figure 4: *Locations of the microphones for BTE-type hearing aid with KEMAR dummy head.*

## 3. Directional microphone in hearing aids

Generally, the directivity of the microphone in a hearing aid is produced by a pair of omnidirectional microphone signals. Fig. 4 shows the locations of omnidirectional microphones in a BTE-type hearing aid mounted on an auricle of a KEMAR dummy head. The separation between the front and rear omnidirectional microphones was 1.08 cm in this case. Fig. 5 shows a signal block diagram of the directional microphone as a spatial signal-processing system. The output of the directional microphone system can be expressed in terms of the microphone separation $d$, the angle of arrival $\theta$, the rear microphone time delay $\tau_r$, and the gain $b$. Then, the directional microphone response in the frequency domain $X_d(\omega,\theta)$ when $b = 1$ can be approximated by the following equation [17]:

$$|X_d(\omega,\theta)| \approx \omega \left( \frac{d}{c} \cos\theta + \tau_r \right), \tag{3}$$

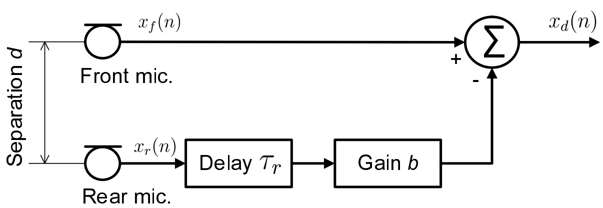


Figure 5: *Block diagram of directional microphone as a spatial signal-processing system.*

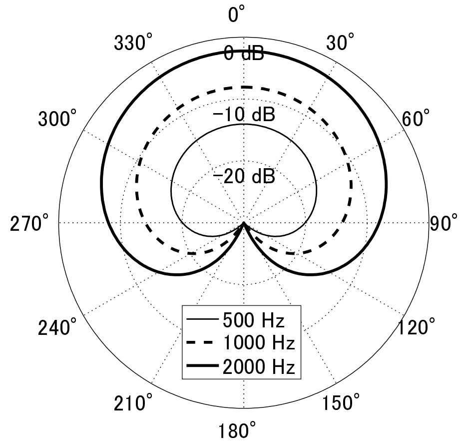


Figure 6: *Microphone directional patterns ($\tau_r = d/c$).*

where $c$ denotes the sound velocity. Fig. 6 shows microphone directional patterns when $\tau_r = d/c$. From the figure, it is found that the sensitivity of the response at lower frequencies is attenuated by 6 dB per octave. These spatial directional responses may affect the extent to which Eq. (2) is satisfied. Therefore, the purpose of this paper is to experimentally evaluate how the proposed low-latency BSS works in the binaural directional system.

## 4. Evaluation

### 4.1. Setup

To evaluate the performance of the low-latency online IVA with binaural directional microphones for hearing aids, a PC simulation was carried out using real mixtures of two speeches recorded by four microphones in binaural BTE-type hearing aids with a head and torso simulator (G.R.A.S.: KEMAR type 45BB) in a meeting room. Fig. 7 shows the setup of the loudspeakers and microphones in the evaluation. Two omnidirectional electret condenser microphones (front and rear) were installed in one BTE-type hearing aid. The hearing aids were attached to each ear of the head and torso simulator. The direction of one of the two sources was fixed at 0° and that of the other source was varied from 30° to 180° in steps of 30°. We selected ten speech sources for each direction from the RWCP Japanese News Speech Corpus [19]. The other experimental conditions are summarized in Table 1.

Table 1: *Experimental conditions*

| | |
|---|---|
| microphone spacing (interaural) | 18.0 cm |
| microphone distance (front and rear) | 1.08 cm |
| reverberation time | 650 ms at 500 Hz |
| signal length | 30 s × 10 |
| sampling frequency | 16 kHz |
| frame length | 4096 |
| frame shift | 1024 |
| window function | Hanning |
| forgetting factor | 0.98 |

For comparison, we used two different latencies, where the numbers of remaining non-causal components $N_d$ were 160 and 2048 samples, corresponding to algorithmic delays of 10 and

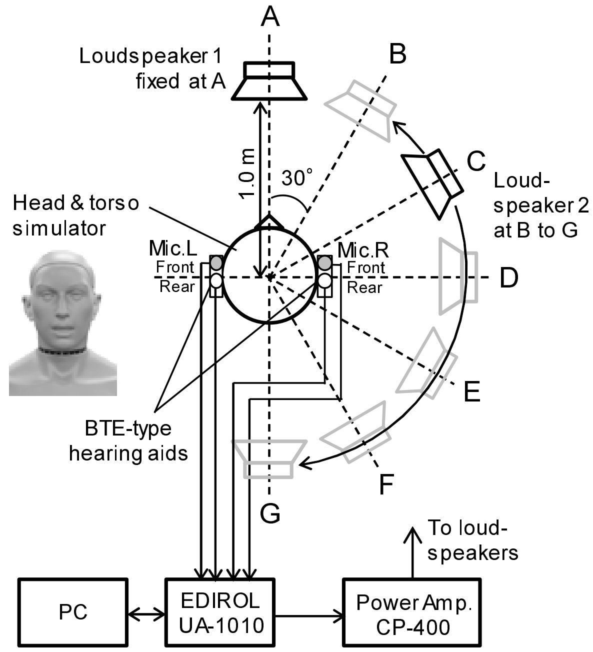


Figure 7: *Setup of loudspeakers and microphones in the evaluation.*

128 ms, respectively. The experiments on the recorded mixtures were performed using MATLAB R2016a on a laptop PC with an Intel Core i7-3770 3.40 GHz processor. The performance was evaluated by the average signal-to-interference ratio (SIR) over all trials with the exception of the first three seconds on each trial, which is defined as the ratio of the signal power of the desired speaker to the signal power from the interfering speaker. The SIR was calculated by bss_eval_images.m in the BSS toolbox [20].

### 4.2. Method

In this experiment, the separated signals were obtained by applying back-projection to the front channel in the omnidirectional case and to the directional channel (the output of Fig. 5). To compare them fairly, it is necessary to compensate the dif-

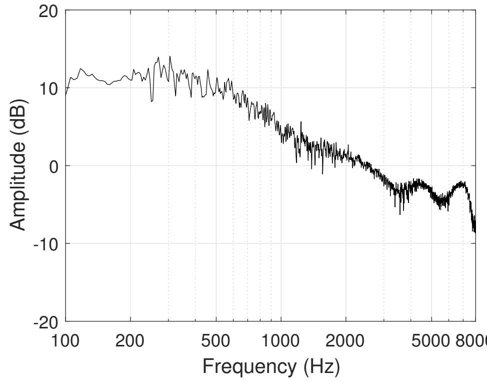


Figure 8: *Example of amplitude response of compensation filter $c_k(\omega)$.*

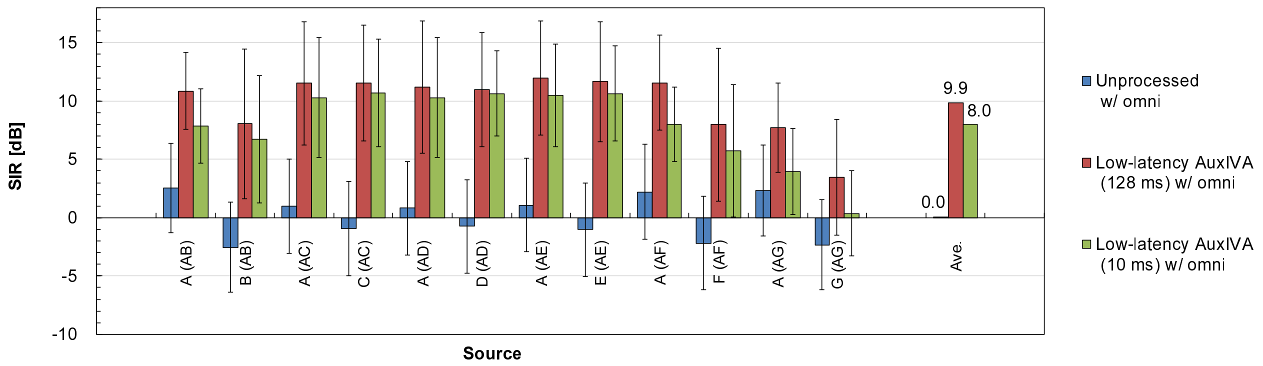


Figure 9: *Separation performance with omnidirectional microphones.*

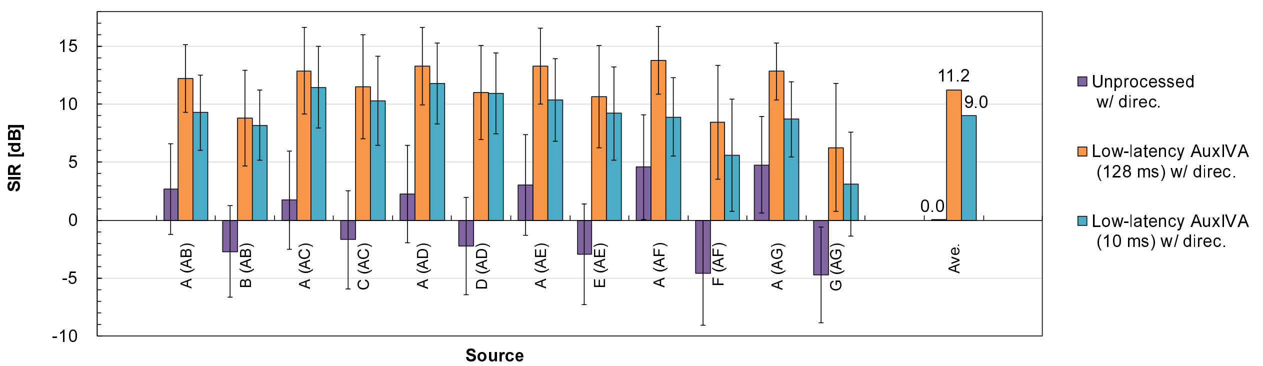


Figure 10: *Separation performance with directional microphones.*

ference in the sensitivity characteristic associated with the directivity. As post-processing for the evaluation, we derive a compensation filter by minimizing the following cost function:

$$J(\boldsymbol{C}(\omega)) = \sum_{\tau} \left| r(\omega,\tau) - \boldsymbol{C}^H(\omega)\boldsymbol{Y}(\omega,\tau) \right|^2, \qquad (4)$$

where $r(\omega,\tau)$ is the STFT representation of the front microphone signal $x_f(n)$, $\boldsymbol{Y}(\omega,\tau) = \begin{bmatrix} y_1(\omega,\tau) \\ y_2(\omega,\tau) \end{bmatrix}$ is a vector comprising of the STFTs of the separated signals $\boldsymbol{y}_k(\omega,\tau)$, and $\boldsymbol{C}(\omega) = \begin{bmatrix} c_1(\omega) \\ c_2(\omega) \end{bmatrix}$ is a vector of the compensation filter $\boldsymbol{c}_k(\omega)$. By differentiating Eq. (4) with respect to $\boldsymbol{C}(\omega)$ and setting the equation to 0, the compensation filter $\boldsymbol{C}(\omega)$ can be calculated as

$$\begin{bmatrix} c_1(\omega) \\ c_2(\omega) \end{bmatrix} = \begin{bmatrix} \sum_{\tau} |y_1(\omega,\tau)|^2 & \sum_{\tau} y_1(\omega,\tau) y_2^*(\omega,\tau) \\ \sum_{\tau} y_1^*(\omega,\tau) y_2(\omega,\tau) & \sum_{\tau} |y_2(\omega,\tau)|^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{\tau} r^*(\omega,\tau) y_1(\omega,\tau) \\ \sum_{\tau} r^*(\omega,\tau) y_2(\omega,\tau) \end{bmatrix}. \qquad (5)$$

Fig. 8 shows an example of the amplitude response of the compensation filter $c_k(\omega)$ for a separated signal with directivity. It was found that the amplitude response was compensated by 6 dB per octave slope.

### 4.3. Results

Figure 9 shows the separation performance with omnidirectional microphones for the low-latency online AuxIVA algorithm with algorithmic delays of 128 and 10 ms. The bars and the error bars indicated the averaged SIR and the standard deviation on each ten trials, respectively. On the horizontal axis, A(AB) denotes source A in a mixture of source A and source B. The average SIR with the algorithmic delay of 10 ms is 8.0 dB, which is 1.9 dB less than that for the delay of 128 ms. In particular, the SIR tends to degrade toward the front-back direction such as at $30^\circ$, $150^\circ$, and $180^\circ$.

On the other hand, Fig. 10 shows the separation performance with directional microphones. The average SIR with the algorithmic delay of 10 ms is 9.0 dB, which is 2.2 dB less than that for the delay of 128 ms. The resultant SIRs with the directional microphones show better separation performance, which was on average 1.0 dB greater than that with omnidirectional microphones. In particular, the improvement of the SIR increases toward the front-back direction.

## 5. Conclusion

In this paper, we evaluated the separation performance of low-latency online AuxIVA with directional microphones for binaural hearing aids. When the directivity was processed from an adjacent pair of omnidirectional microphones, the sensitivity of the response at lower frequencies was attenuated by 6 dB per octave compared with the omnidirectional microphones. To compare them fairly, the difference in the sensitivity characteristics was compensated by post-processing. From the evaluation results, the performance of the proposed algorithm with directional microphones in the case of 10 ms latency was 9.0 dB in terms of the SIR, which is only a 2.2 dB performance loss from the case of 128 ms latency. Moreover, the average SIR with directional microphones was about 1.0 dB better than that with omnidirectional microphones. Future work will focus on listening tests to verify the proposed system.

## 6. Acknowledgments

This work was partially supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP16H01735.

## 7. References

[1] Y. Mori, T. Takatani, H. Saruwatari, K. Shikano, T. Hiekata, and T. Morita, "High-presence hearing-aid system using DSP-based real-time blind source separation module," in *Proc. ICASSP*, 2007.

[2] K. Reindl, Y. Zheng, and W. Kellermann, "Speech enhancement for binaural hearing aids based on blind source separation," in *Proc. ISCCSP*, 2010.

[3] R. Aichner, H. Buchner, F. Yan, and W. Kellermann, "A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments," *Signal Processing*, vol. 86, pp. 1260–1277, 2006.

[4] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," in *Proc. ICA*, pp. 601–608, 2006.

[5] T. Kim, T. Eltoft, and T. W. Lee, "Independent vector analysis: an extension of ICA to multivariate components," in *Proc. ICA*, pp. 165–172, 2006.

[6] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASLP*, vol. 15, pp. 70–79, 2007.

[7] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. WASPAA*, pp. 189–192, 2011.

[8] N. Ono, "Fast stereo independent vector analysis and its implementation on mobile phone," in *Proc. IWAENC*, 2012.

[9] N. Ono, "Auxiliary-function based independent vector analysis with power of vector-norm type weighting functions," in *Proc. APSIPA ASC*, 2012.

[10] 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.*, vol. 29, pp. 601–617, 2008.

[11] R. Heurig and J. Chalupper, "Acceptable processing delay in digital hearing aids," *Hear. Rev.*, vol. 17, pp. 28–31, 2010.

[12] D. Mauler and R. Martin, "A low delay, variable resolution, perfect reconstruction spectral analysis-synthesis system for speech enhancement," in *Proc. EUSIPCO*, pp. 222–226, 2007.

[13] T. Kim, "Real-time independent vector analysis for convolutive blind source separation," *IEEE Trans. CAS I*, vol. 57, no. 7, pp. 1431–1438, 2010.

[14] T. Taniguchi, N. Ono, A. Kawamata, and S. Sagayama, "An auxiliary-function approach to online independent vector analysis for real-time blind source separation," in *Proc. HSCMA*, pp. 107–111, 2014.

[15] J. Agnew and J. M. Thornton, "Just noticeable and objectionable group delays in digital hearing aids," *J. Am. Acad. Audiol.*, vol. 11, pp. 330–336, 2000.

[16] M. Sunohara, C. Haruta, and N. Ono, "Low-latency real-time blind source separation for hearing aids based on time-domain implementation of online independent vector analysis with truncation of non-causal components," in *Proc. ICASSP*, 2017.

[17] J. M. Kates, *Digital Hearing Aids*. Plural Publishing, 2008.

[18] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1, pp. 1–24, 2001.

[19] "Real World Computing Project News Speech Corpus," http://research.nii.ac.jp/src/en/RWCP-SP99.html. [Accessed: 11 Sep. 2016].

[20] E. Vincent, C. Fevotte, and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.