Self-Produced Speech Enhancement and Suppression Method using Air- and Body-Conductive Microphones

Moe Takada*, Shogo Seki* and Tomoki Toda†
* Graduate School of Informatics, Nagoya University, Japan
E-mail: {takada.moe, seki.shogo}@g.sp.m.is.nagoya-u.ac.jp
† Information Technology Center, Nagoya University, Japan
E-mail: tomoki@icts.nagoya-u.ac.jp

***Abstract*—This paper presents a self-produced speech enhancement and suppression method for multichannel auditory signals recorded with both air- and body-conductive microphones. In processing auditory sound signals recorded with wearable microphones for memorizing conversation and acoustic scenes and events that each user experienced in daily life, source separation is a promising technique as only a mixing of various sound sources can be observed. To separate the recorded signals into self-produced speech, i.e., user's own speech and ambient environmental sounds, we propose a self-produced speech enhancement and suppression method using not only air-conductive microphones to record multichannel air-conducted signals but also body-conductive microphones to record body-conducted signals. A separation filter applied to the multichannel air-conducted signals is estimated in an unsupervised manner while effectively using the body-conducted signals dominantly including self-produced speech components. We conduct an experimental evaluation to investigate the effectiveness of the proposed method, thereby demonstrating that the proposed method outperforms the conventional method using only the air-conductive microphones.**

I. Introduction

With the development of small audio recording devices and the increase of attention for understanding speech and audio scenes, novel audio applications using the wearable audio recording devices are expected to be developed. For instance, auditory sounds recorded with wearable microphones are expected to be utilized for memorizing conversation and acoustic scenes and events that each person experienced in daily life. On the one hand, the recorded auditory sound signals usually contain various sound source signals, such as user's speech and ambient environmental sounds. On the other hand, target source signals usually depend on individual applications, e.g., it will be ideal that we can obtain only user's speech for an application for transcribing it with speech recognition, and we can obtain only ambient environmental sounds for classifying acoustic scenes and detecting acoustic events. Therefore, a sound source separation technique is expected to be effective for extracting the desired target source signals from the mixed signals.

As a typical source separation technique, Blind Source Separation (BSS) has been actively discussed for many years. In BSS based on Independent Component Analysis (ICA) [1]–[3], a mixed signal is separated using a linear separation filter estimated assuming that source signals are independent of each other. Frequency-Domain ICA (FDICA) has been successfully applied to auditory sound source separation [4], [5], and it has been extended to Independent Vector Analysis (IVA) [6], [7] to solve a permutation problem in FDICA. Recently, Independent Low-Rank Matrix Analysis (ILRMA) [8], [9] has been proposed to achieve high separation performance by further introducing a sound source model based on non-negative matrix factorization (NMF) [10]–[12] into IVA. These approaches are expected to be also available for the mixed signals recorded with the wearable microphones and be further extended for achieving higher separation performance.

As the first step towards the development of BSS for the mixed signals recorded with the wearable microphones, in this paper we focus on enhancement and suppression of a self-produced speech, i.e., user's own speech. As a conventional method related to this task, a speech enhancement method using both air- and bone-conductive microphones has been proposed [13]. The bone-conductive microphone is robust against external noise sounds. Therefore, bone-conducted signals tend to contain self-produced speech components dominantly. Although their sound quality significantly degrades owing to a mechanism of bone-conductive recording, these signals are effectively used for enhancing air-conducted speech signals under noisy conditions.

Inspired by this conventional work, we propose a self-produced speech enhancement and suppression method using not only multiple air-conductive microphones as in the traditional BSS framework but also a body-conductive microphone to dominantly capture self-produced speech components. As one of the high-quality body-conductive microphones, we focus on the Non-Audible Murmur (NAM) microphone [14], which is a special microphone developed to detect non-audible murmur, i.e., very soft whispered voice. NAM microphone is also robust against external noise and can also record normal voices as well as NAM. Moreover, NAM microphone is used by attaching it to the skin surface behind the ear, and therefore, it is straightforward to install it in a wearable device with it, such as a neckband-type device. In our conventional work, we have reported that Semi-Blind Source Separation (Semi-BSS) [15], [16] and Non-negative Tensor Factorization (NTF) [17] can be effectively used to develop a NAM enhancement method under noisy conditions by using both NAM and air-conductive microphones. Although the task handled in this conventional work is different from that in this paper, it is expected that multichannel signal processing using both NAM and air-conductive microphones will be effective for enhancing or suppressing self-produced speech from the mixed signals.

In this paper, we apply the ILRMA-based BSS framework to the self-produced speech enhancement and suppression processing using multiple air-conducive microphones and a NAM microphone. A separation filter applied to the multichannel air-conducted signals is estimated by effectively using not only the multichannel air-conducted signals but also the body-conducted signal as a clue to estimate self-produced speech components. Our experimental result will demonstrate that the proposed method outperforms the conventional method using only the air-conducted signals.

II. Conventional source separation method

BSS estimates individual source signals from mixed signals in an unsupervised manner. When the number of sound sources N is less than or equal to that of observed channels M, the observed multichannel signal is usually well separated by using a linear separation filter. Let us denote frequency components (i.e., complex values) of the source signal $\boldsymbol{s}_{ij}$, the observed signal $\boldsymbol{x}_{ij}$ and the separated signal $\boldsymbol{y}_{ij}$ as follows:

$$\boldsymbol{s}_{ij} = (s_{ij,1}, \ldots, s_{ij,N})^\top, \tag{1}$$
$$\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,M})^\top, \tag{2}$$
$$\boldsymbol{y}_{ij} = (y_{ij,1}, \ldots, y_{ij,N})^\top, \tag{3}$$

where $i = 1, \ldots, I$ is a frequency index and $j = 1, \ldots, J$ is a time index. We assume that the mixing process is modeled as follows:

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}, \tag{4}$$

where $\boldsymbol{A}_i$ is the time-invariant mixing matrix. Then, the separated signal is given by

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}, \tag{5}$$

where $\boldsymbol{W}_i$ is the separation matrix to be estimated assuming that the source signals are statistically independent of each other in ICA, IVA and ILRMA.

A. Independent Component Analysis (ICA)

We assume that source signals are statistically independent of each other and are generated from non-Gaussian distribution. The joint distribution function of $\boldsymbol{y}$ is denoted as

$$p(\boldsymbol{y}) = p(y_1, \ldots, y_N), \tag{6}$$

where the notation i, j are omitted for simplicity. Each component y_n of $\boldsymbol{y}$ is independent of each other if the separation matrix $\boldsymbol{W}$ can separate $\boldsymbol{x}$ into independent components. In such a case, the joint distribution function is given by

$$p(\boldsymbol{y}) = \prod_{n=1}^{N} p(y_n), \tag{7}$$

where $p(y_n)$ is the marginal distribution function of y_n. In ICA, $\boldsymbol{W}$ is estimated so that Eq. (7) holds.

Kullback-Leibler divergence [18] is used to evaluate Eq. (7), which is defined as follows:

$$\begin{aligned}\mathrm{KL}(\boldsymbol{W}) &= \int p(\boldsymbol{y}) \log \frac{p(\boldsymbol{y})}{\prod_{n=1}^{N} p(y_n)} d\boldsymbol{y} \\ &= -H(\boldsymbol{y}; \boldsymbol{W}) + \sum_{n=1}^{N} H(y_n; \boldsymbol{W}),\end{aligned} \tag{8}$$

where $H(\boldsymbol{y}; \boldsymbol{W})$ is a joint entropy and $H(y_n; \boldsymbol{W})$ is a marginal entropy. Denoting $|\boldsymbol{W}|$ as the determinant of $\boldsymbol{W}$, the following formula holds,

$$\mathrm{KL}(\boldsymbol{W}) \propto -\log |\boldsymbol{W}| - \sum_{n=1}^{N} E_{\boldsymbol{x}}[\log p(y_n)], \tag{9}$$

where $E_{\boldsymbol{x}}$ is an expectation over $\boldsymbol{x}$. $\mathrm{KL}(\boldsymbol{W})$ is zero when Eq. (7) holds. Therefore, the separation matrix $\boldsymbol{W}$ is estimated by minimizing $\mathrm{KL}(\boldsymbol{W})$. In this estimation, the gradient method is used to iteratively update the estimate of $\boldsymbol{W}$ as follows:

$$\boldsymbol{W}_{(t+1)} = \boldsymbol{W}_t + \eta(\boldsymbol{I} - \phi(\boldsymbol{y})\boldsymbol{y}^\top)\boldsymbol{W}_t, \tag{10}$$

where η is a positive constant and $\boldsymbol{I}$ is an identity matrix. The vector $\phi(\boldsymbol{y})$ is defined as follows:

$$\phi(\boldsymbol{y}) = -\left(\frac{\partial \log p(y_1)}{\partial y_1}, ..., \frac{\partial \log p(y_N)}{\partial y_N}\right)^\top, \tag{11}$$

where the derivatives are approximated with a nonlinear function assuming a specific probability density function $p(y_n)$.

B. Independent Low-Rank Matrix Analysis (ILRMA)

ILRMA is originated from the frequency-domain ICA and has been proposed as natural expansion of IVA, which handles all frequency components over an entire of frequency bands as one vector $\boldsymbol{y}_{j,n} = (y_{1j,n}, ..., y_{Ij,n})^\top$. On the one hand, IVA uses a time-invariant sound source model. On the other hand, ILRMA uses a time-invariant sound source model based on NMF. We assume the complex Gaussian distribution as a generation model of each source signal.

$$\begin{aligned}p(\boldsymbol{y}_{j,1}, ..., \boldsymbol{y}_{j,N}) &= \prod_{n} p(\boldsymbol{y}_{j,n}) \\ &= \prod_{i,n} \frac{1}{\pi r_{ij,n}} \exp(-\frac{|y_{ij,n}|^2}{r_{ij,n}}),\end{aligned} \tag{12}$$

where $r_{ij,n}$ is a time-variant frequency-dependent variance of each sound source. In ILRMA, it is modeled by NMF as follows:

$$r_{ij,n} = \sum_{k} z_{nk} t_{ik} v_{kj}, \tag{13}$$

where z_{nk} is a hidden variable to assign the k-th basis to the n-th source signal, and the sum of z_{nk} over k is 1 subject to $0 \leq z_{nk} \leq 1$. Moreover, t_{ik} and v_{kj} are a component of a basis matrix $\boldsymbol{T} = [t_{ik}] \in \mathbb{R}_{\geq 0}^{I \times K}$ having K basis vectors corresponding to individual sound sources and a component of an activation matrix $\boldsymbol{V} = [v_{kj}] \in \mathbb{R}_{\geq 0}^{K \times J}$, respectively. The objective function of ILRMA is defined as follows:

$$\begin{aligned} Q_{\mathrm{ILRMA}} &= -2J \sum_i \log |\det \boldsymbol{W}_i| \\ &+ \sum_{i,j,n} \frac{|y_{ij,n}|^2}{r_{ij,n}} + \sum_{i,j,n} \log r_{ij,n}. \end{aligned} \tag{14}$$

In R.H.S. of Eq. (14), the first term and the second term correspond to the objective function of estimating the separating matrix in IVA, and the second term and the third term correspond to the objective function of estimating the sound source model in NMF with the Itakura-Saito divergence. The update formula of the separation matrix $\boldsymbol{W}_i$ is obtained by the iterative projection method [19] and the update formula of the sound source model is obtained from the auxiliary function approach [20].

C. Wiener filter

As a typical BSS framework, speech and noise signals are first estimated by ICA, IVA or ILRMA as described above, and then, Wiener filter designed using the estimated speech and noise signals is applied to the observed signal to extract target speech signal. This framework is effective if the observed signal is not well separated into each source signal with the linear filter, e.g., in such a situation as the number of sources N is larger than the number of observed channels M.

Wiener filter [21] is designed by minimizing a mean square error between an estimated target signal and a true target signal. It is well known that Wiener filter works reasonably well in noise suppression or speech enhancement.

Suppose that the observed signal x_{ij} is given by superposition of a speech signal $x_{ij}^{(\mathrm{s})}$ and a noise signal $x_{ij}^{(\mathrm{n})}$ as follows:

$$x_{ij} = x_{ij}^{(\mathrm{s})} + x_{ij}^{(\mathrm{n})}. \tag{15}$$

Then, Wiener filter is given by

$$G_{ij} = \frac{P(x_{ij}^{(\mathrm{s})})}{P(x_{ij}^{(\mathrm{s})}) + P(x_{ij}^{(\mathrm{n})})}, \tag{16}$$

where $P(\cdot)$ shows power spectrum. The target signal is obtained by applying this filter to the observed signal.

III. Proposed Separation Method for Self-Produced Speech Enhancement and Suppression

A. Multi-channel recording with air- and body-conductive microphones

The left of Figure 1 shows a neckband-type wearable recording device with multiple air-conductive microphones and a NAM microphone used in our proposed framework.

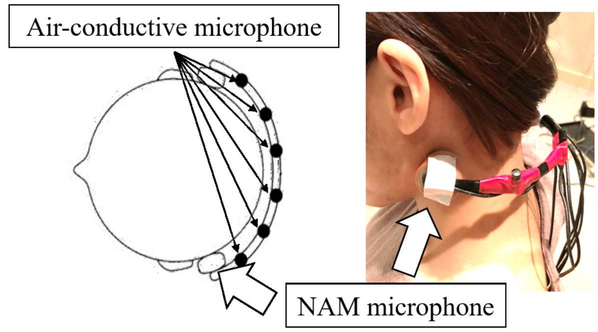

Fig. 1: Air- and body-conductive microphones (left: recording environment, right: microphone position)

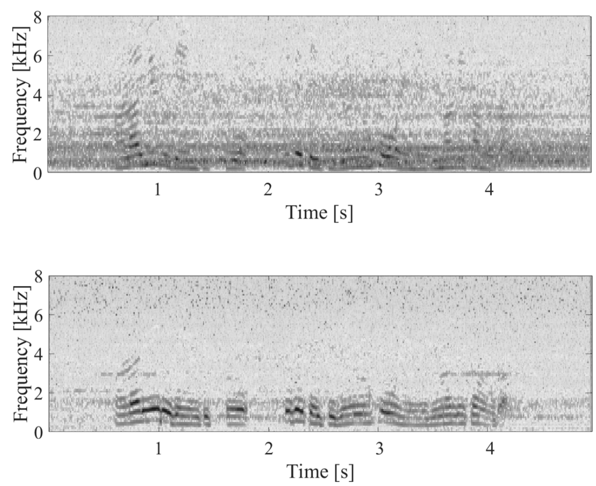

Fig. 2: Noisy spectrograms (top: air-conducted signals, bottom: body-conducted signals)

The air-conductive microphones are installed on the device at equal intervals, which is set to the back of speaker's neck. NAM microphone is also installed on the device and attached to the position shown in the right of Figure 1. Figure 2 shows an example of spectrograms of air- and body-conducted signals recorded in the air-conductive and NAM microphones under 70 dBA of noisy condition. Although air-conducted microphone can record the self-produced speech components in a wide range of frequency, it is easily contaminated by external sound signals. On the other hand, as NAM microphone can record the self-produced speech signal dominantly while suppressing external sound signals, the self-produced speech components are well observed in the body-conducted signal. However, their high frequency components are not observed as they are significantly suppressed by essential mechanisms of body conduction, such as lack of radiation characteristics from lips and effect of low-pass characteristics of the soft tissues, and therefore, the sound quality of the body-conducted speech signal is degraded significantly.

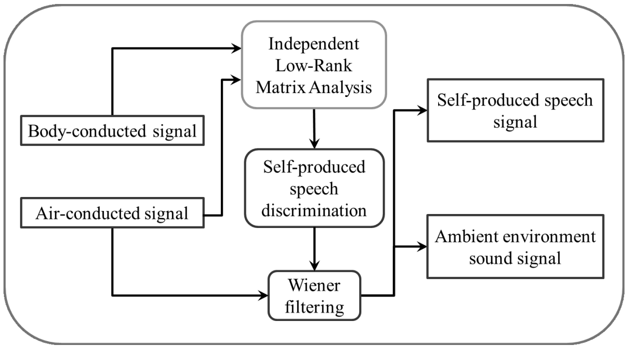

Fig. 3: Overview of the proposed method

B. Self-produced speech suppression and enhancement

In the proposed method, an observed signal is given by

$$\boldsymbol{x}'_{ij} = (\boldsymbol{x}_{ij}^\top, x_{ij}^{(b)})^\top, \tag{17}$$

where $x_{ij}^{(b)}$ is an observed body-conducted signal. We assume that the observed signal is generated by the following mixing process,

$$\boldsymbol{x}'_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{18}$$

and the separated signal is relatively well determined by using the time-invariant linear filter as in the conventional separation method, which is shown as

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}'_{ij}. \tag{19}$$

Figure 3 shows an overview of the proposed method. First, the linear separation filter in Eq. (19) is estimated using ILRMA. The number of source signals N is set equal to the number of observed channels M. The hidden variables $z_{n,k}$, the basis matrix and the activation matrix are randomly initialized. The N-by-N separation matrix $\boldsymbol{W}$ is initialized as the identity matrix. After estimating $\boldsymbol{W}$ with ILRMA, the multichannel separated signals $\hat{\boldsymbol{y}}_{ij,n} = (\hat{y}_{ij,n1}, \ldots, \hat{y}_{ij,nM}, \hat{y}_{ij,n}^{(b)})^\top$ are generated by using projection back [22] as follows:

$$\hat{\boldsymbol{y}}_{ij,n} = \boldsymbol{W}_i^{-1} \boldsymbol{M}_n \boldsymbol{y}_{ij}, \tag{20}$$

where $\boldsymbol{M}_n$ is a diagonal matrix to extract only the n-th component of $\boldsymbol{y}_{ij}$ by masking the other components. Then, we automatically select separated signals corresponding to the self-produced speech signal from the multichannel separated signals by grouping them into a self-produced speech signal group $N^{(\mathrm{s})}$ and an environmental noisy signal group $N^{(\mathrm{n})}$. Under the condition that the number of separated signals in $N^{(\mathrm{s})}$ is given, the separated signals with the largest waveform power calculated in only the channel corresponding to the body-conducted signal (i.e., $\hat{y}_{ij,n}^{(b)}$) are assigned to $N^{(\mathrm{s})}$. This grouping process works reasonably well thanks to an inherent property of the body-conducted signal, i.e., dominantly containing the self-produced speech components. In this paper, we set the number of separated signals in $N^{(\mathrm{s})}$ to 1 or 2.

TABLE I: Experimental condition

Evaluation data	18 sentences
Sampling frequency	48 kHz
Frame size	11.6 ms (512 pt)
Shift size	5.8 ms (256 pt)
Iteration times in ILRMA	100
Number of basis vectors K	200
Number of Channels M	5 ch

After that, we finally use Wiener filter to estimate the self-produced speech signals for the self-produced speech enhancement or the environmental noise signals for the self-produced suppression. The separated signals assigned to $N^{(\mathrm{s})}$ are superimposed to generate the separated signal $\hat{\boldsymbol{y}}_{ij}^{(\mathrm{s})}$ corresponding to the self-produced speech signal, and those signals assigned to $N^{(\mathrm{n})}$ are superimposed to generate $\hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})}$ corresponding to the environment noise signal as follows:

$$\hat{\boldsymbol{y}}_{ij}^{(\mathrm{s})} = \sum_{n \in N^{(\mathrm{s})}} \hat{\boldsymbol{y}}_{ij,n}, \tag{21}$$

$$\hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})} = \sum_{n \in N^{(\mathrm{n})}} \hat{\boldsymbol{y}}_{ij,n}. \tag{22}$$

Then, a single channel Wiener filter is estimated separately in each channel using $\hat{y}_{ij,m}^{(\mathrm{s})}$, $\hat{y}_{ij,m}^{(\mathrm{n})}$ as follows:

$$G_{ij,m}^{(s)} = \frac{P(\hat{y}_{ij,m}^{(\mathrm{s})})}{P(\hat{y}_{ij,m}^{(\mathrm{s})}) + P(\hat{y}_{ij,m}^{(\mathrm{n})})}. \tag{23}$$

The enhanced self-produced speech signal is obtained by applying this filter to the observed signal in each channel. Note that only channels corresponding to the air-conducted signals are used in this filtering process. The self-produced speech suppression is also achieved by building Wiener filter to suppress $\hat{y}_{ij,m}^{(\mathrm{s})}$ in a similar manner as the self-produced speech enhancement.

IV. Experimental evaluation

A. Experimental conditions

An experimental evaluation for the self-produced speech enhancement and suppression was conducted. TABLE I shows the experimental condition. Speech and environmental sound were separately recorded and superimposed to generate the mixed sound. Eighteen sentences uttered by one Japanese female speaker were used as evaluation data. Crowd noise with 70 dBA of the sound pressure level was used as the environmental sound. The six noise sources were arranged at intervals of 60 degrees around the speaker, where the front of the speaker was defined as zero degree. In the proposed framework, air-conducted signals recorded with four air-conductive microphones of the wearable recording device and another single-channel body-conducted signal recorded with the NAM microphone were used as five-channel signals. On the other hand, air-conducted signals recorded with five air-conductive microphones of the wearable recording device were used as five-channel signals in the conventional method.

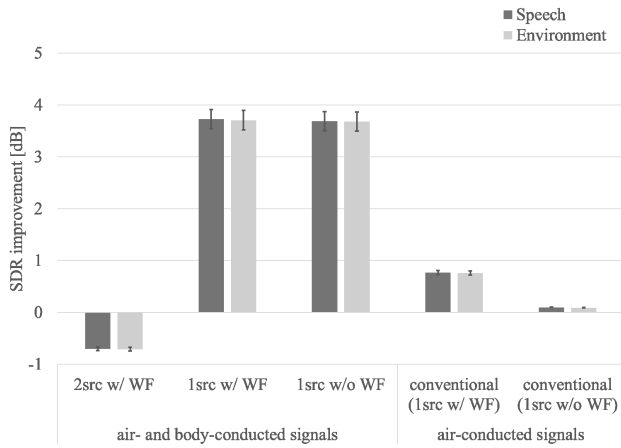

Fig. 4: Self-produced speech enhancement/suppression result

Namely, the number of observed channels was set to be smaller than the number of sound sources as in practical conditions.

To investigate the effect of the number of self-produced speech source signals assumed in the proposed separation process, i.e., the number of separated signals assigned to $N^{(\mathrm{s})}$, and the effect of using Wiener filter in the proposed method, the following settings were evaluated:

- 2src w/ WF: the use of 2 self-produced speech source signals and Wiener filter,
- 1src w/ WF: the use of 1 self-produced speech source signal and Wiener filter,
- 1src w/o WF: the use of 1 self-produced speech source signal and only the linear separation filter.

In the conventional method, 1 self-produced speech source signal was assumed in the separation process and we investigate the effect of using Wiener filter. We calculated the Signal-to-Distortion Ratios (SDRs) [23] at each of the air-conducted speech channels and compared their averaged values in each setting.

B. Experimental result

Figure 4 shows the experimental result. The error bars show the 95% confidence intervals. It is shown that the proposed method using both the air-conducted signals and the body-conducted signals can outperform the conventional method using only air-conducted signals. We can see that although Wiener filter improves the enhancement and suppression performance in the conventional methods, it is not helpful in the proposed method. We can also see that the use of multiple source signals for modeling the self-produced speech is not useful, and therefore, it is enough to select only one separated signal corresponding to the self-produced speech signal in the proposed method.

V. Conclusion

In this paper, we have proposed the self-produced speech enhancement and suppression method using a body-conducted signal as well as usual multichannel air-conducted signals recorded using a wearable device with a NAM microphone and multiple air-conductive microphones. In the proposed method, an efficient enhancement and suppression processing has been achieved by effectively using the body-conducted signals dominantly including self-produced speech components. The result of an experimental evaluation has showed that 1) the use of body-conducted signals is effective for the self-produced speech enhancement and suppression, 2) the use of a combination of linear separation based on ILRMA and Wiener filtering is not helpful in the proposed method.

We plan to develop more suitable mixing process to accurately model the observed air- and body-conducted signals.

VI. Acknowledgment

Part of this work was supported by JSPS KAKENHI Grant Numbers 15K12064 and 17H01763.

References

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.
[2] S. Makino, S. Araki, R. Mukai and H. Sawada, "Audio Source Separation based on Independent Component Analysis," *IEEE Proc. ISCAS*, pp. 668–671, 2004.
[3] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa and K. Shikano, "Blind source separation combining independent component analysis and beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1135–1146, 2003.
[4] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1–3, pp. 21–34, 1998.
[5] H. Saruwatari, T. kawamura and K. Shikano, "Blind source separation for speech based on fast-convergence algorithm with ICA and beamforming," *Proc. Eurospeech2001*, pp. 2603–2606, 2001.
[6] T. Kim, H. Attias, S.-Y. Lee and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE/ACM Transaction on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 70–79, 2007.
[7] A. Hiroe, "Solution of Permutation Problem in Frequency Domain ICA, Using Multivariate Probability Density Functions," *Proc. ICA 2006*, pp. 601–608, 2006.
[8] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transaction on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.
[9] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," *IEEE Proc. ICASSP*, pp. 276–280, 2015.
[10] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Proc. NIPS 2000*, pp. 556–562, 2001
[11] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," *IEEE Proc. Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 177–180, 2003.
[12] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," *Proc. ICA 2007*, pp. 414–421, 2007.
[13] A. Subramanya, Z. Zhang, Z. Liu, and A. Acero, "Multisensory processing for speech enhancement and magnitude-normalized spectra for speech modeling," *Speech Communication*, vol. 50, no. 3, pp. 228–243, 2008.
[14] Y. Nakajima, H. Kashioka, N. Campbell and K. Shikano, "Non-audible murmur (NAM) recognition," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 1, pp. 1–8, 2006.
[15] S. Miyabe, T. Takatani, Y. Mori, H. Saruwatari, K. Shikano and Y. Tatekura, "Double-talk free spoken dialogue interface combining sound field control with semi-blind source separation," *IEEE Proc. ICASSP*, pp. 809–812, 2006.
[16] Y. Tajiri, T. Toda and S. Nakamura, "Noise suppression method for body-conducted soft speech enhancement based on external noise monitoring," *IEEE Proc. ICASSP*, pp. 5935–5939, 2006.

[17] Y. Tajiri, H. Kameoka and T. Toda, "A noise suppression method for body-conducted soft speech based on non-negative tensor factorization of air- and body-conducted signals," *IEEE Proc. ICASSP*, pp. 4960–4964, 2017.
[18] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
[19] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *IEEE Proc. Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 189–192, 2011.
[20] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
[21] Liang Hong, J. Rosca, and R. Balan, "Independent Component analysis based single channel speech enhancement," *IEEE Proc. ISSPIT*, 2003.
[22] 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.
[23] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation," *IEEE/ACM Transaction on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.