SUBBAND BEAMFORMER COMBINED WITH TIME-FREQUENCY ICA FOR EXTRACTION OF TARGET SOURCE UNDER REVERBERANT ENVIRONMENTS

Masahito Togami, Yohei Kawaguchi, Hiroaki Kokubo, and Yasunari Obuchi

Central Research Laboratory, Hitachi Ltd.
1-280, Higashi-koigakubo Kokubunji-shi, 185-8601, Tokyo, Japan
phone: +81-42-323-1111, fax: +81-42-327-7823,
email: { masahito.togami.fe, yohei.kawaguchi.xk, hiroaki.kokubo.dz, yasunari.obuchi.jx }@hitachi.com

ABSTRACT

We propose a novel beamforming method that is complementarily combined with the time-frequency domain independent component analysis (TF-ICA). Under reverberant environments, performance of TF-ICA is known to be limited. On the other hand, when time period in which there are only noise sources (noise-only period) is given, beamforming based on multichannel spatial prediction (MSP-BF) are shown to achieve good noise reduction performance. The proposed method applies noise-period identification based on the output of TF-ICA to obtain a noise-only period. MSP-BF using the obtained noise-only period is performed. Furthermore, to avoid performance degradation, the proposed method selects the better output signal obtained by TF-ICA and MSP-BF. Experimental results under a reverberant environment indicate that the proposed method outperforms TF-ICA. Furthermore, thanks to output-signal selector, even when there are no noise-only periods, performance degradation can be avoided.

1. INTRODUCTION

Multichannel noise-reduction techniques have been widely studied. In particular, for conference-recording systems or video-conferencing systems, noise-reduction techniques that work well under reverberant environments are strongly required. Minimum variance beamformer [1] and generalized sidelobe canceller [2] are frequently used. As for these beamforming techniques, the impulse response between the target source and a microphone array is one important system parameter. When the impulse response is given, a completely-noise-reduced signal can be achieved by conventional methods [3]. However, obtaining the impulse response of a target source from mixed signals of the target source and the noise sources is difficult and not realistic. Recently, blind source separation (BSS) techniques, such as independent component analysis (ICA), have been studied [4][5]. ICA can reduce noise without pre-knowledge of the impulse response between the target source and a microphone array. It only requires that the target source and noise sources are mutually independent. Accordingly, ICA is commonly used as an alternative noise-reduction method to conventional beamforming techniques. For example, ICA performed in the time-frequency domain (TF-ICA) is a popular approach. However, Araki, et al. [6] found that the performance of TF-ICA is limited under reverberant environments. To overcome this limitation of TF-ICA, ICA in the subband domain [7] or a combination of TF-ICA with ICA in the time domain [4] have been studied. However, even if subband ICA or time-domain ICA are used, it is very hard that the separation filter of ICA converges under reverberant environments.

On the other hand, aiming to improve noise-reduction performance of beamforming techniques under reverberant environments, the authors have proposed a novel beamforming technique based on multichannel spatial prediction (MSP-BF) [8]. MSP-BF does not require the impulse response of the target source, and it can reduce non-stationary noise under reverberant environments. Moreover, it requires a time period in which there are only noise sources (noise-only period). In this paper, we propose a novel beamforming method that is complementarily combined with TF-ICA. Using the output signals of TF-ICA, *Noise period identification* detects a noise-only period. The noise-reduction filter of MSP-BF is adapted according to the obtained noise-only period, and the target source component is extracted from the noisy input signal by MSP-BF. To avoid performance degradation when there are no noise-only period, the proposed method selects the better output signal of either TF-ICA or MSP-BF according to the non-linear correlation between the output target source estimation and the noise-source estimation. Experimental results in the case of a reverberant environment (reverberation time is about 300 ms) indicate that the noise reduction performance of the proposed method is higher than that of TF-ICA. Furthermore, thanks to an output-signal selector, even when there is no noise-only period, performance degradation can be avoided.

2. PROBLEM STATEMENT

2.1 Input Signal Model

M is defined as the number of microphones. N is defined as the sum of the number of target signals N_s and the number of noise signals N_n. The received sound signal at the m-th microphone element is given as $x_m(t)$. t is the sampling number of an A/D converter. The original source signal of the i-th target signal is described as $s_i(t)$, and that of the i-th noise signal is described as $n_i(t)$. $h_{i,m}$ is the impulse response of the i-th target signal between the i-th source position and the m-th microphone. $g_{i,m}$ is the impulse response of the i-th noise signal. $x_m(t)$ is defined as $x_m(t) = \sum_{i=0}^{N_s-1}\left(h_{i,m} * s_i(t)\right) + \sum_{i=0}^{N_n-1}\left(g_{i,m} * n_i(t)\right)$, where $*$ is the operator of convolution. In this paper, the noise reduction is defined as extraction of the target source component $\sum_{i=0}^{N_s-1}\left(h_{i,c} * s_i(t)\right)$ at the c-th microphone from M noisy input signals $[x_1(t),\ldots,x_M(t)]$. c is defined as the *target microphone* index.

2.2 Subband Beamformer with Multichannel Prediction [8]

Subband beamfomer with multichannel spatial prediction (MSP-BF) has been shown to achieve good noise reduction under reverberant environments [8]. MSP-BF does not require the impulse response of the target source; it requires a time period in which there are only noise sources. For easy description, the target source component of the recorded signal in the m-th microphone is replaced as $y_m(t) = \sum_{i=0}^{N_s-1}\left(h_{i,m} * s_i(t)\right)$, and the noise source component is replaced as $v_m(t) = \sum_{i=0}^{N_n-1}\left(g_i(t) * n_i(t)\right)$. Therefore, $x_m(t) = y_m(t) + v_m(t)$. $x_m(t)$ is transformed into subband domain signal $x_m(k,t)$ by DFT filter bank [9], k is the subband index, and $x_m(k,t)$ is the k-th subband signal at the m-th microphone. MSP-BF for each subband is composed of two processes; noise reduction and distortion-restoration.

2.2.1 Noise Reduction Process

$v_m(k,t)$ is removed from $x_m(k,t)$ by multichannel spatial prediction with no constraint on the target source as follows:

$$\hat{y}_m(k,t) = x_m(k,t) - \boldsymbol{a}_m \boldsymbol{x}_m^e(k,t), \quad (1)$$

where $\boldsymbol{x}^e_m(k,t)$ is the noisy input signal that do not include the noisy signal at the m-th microphone, $\boldsymbol{x}^e_m(k,t) = [\boldsymbol{x}_1(k,t)^T,\ldots,\boldsymbol{x}_{m-1}(k,t)^T,\boldsymbol{x}_{m+1}(k,t)^T,\ldots]^T$, $\boldsymbol{x}_m(k,t) = [x_m(k,t),x_m(k,t-1),\ldots,x_m(k,t-L+1)]^T$, and $(M-1)L$ is the length of the noise reduction filter. $\boldsymbol{a}_m$ is the spatial prediction coefficient of the noise in the m-th microphone, and it is adapted in the time period when there is only one noise source (noise-only period) so as to remove noise component $v_m(k,t)$ as follows:

$$\boldsymbol{a}_m \leftarrow \underset{\boldsymbol{a}_m}{\operatorname{argmin}} \sum_{t\in T_n} ||v_m(k,t) - \boldsymbol{a}_m \boldsymbol{v}^e_m(k,t)||^2,$$
$$\boldsymbol{a}_m = \sum_{t\in T_n} v_m(k,t)\boldsymbol{v}^e_m(k,t)^H \Big(\sum_{t\in T_n} \boldsymbol{v}^e_m(k,t)\boldsymbol{v}^e_m(k,t)^H\Big)^{-1}, \quad (2)$$

where H is the operator for conjugate transpose, and T_n is the noise-only period. In MSP-BF, the noise-only period is assumed to be pre-defined. In Eq. 1, $\boldsymbol{a}_m\boldsymbol{x}^e_m(k,t)$ is the estimate of the noise component in the m-th microphone, and the noise component in the m-th microphone can be filtered out by the subtraction. The noise reduction process can be executed without pre knowledge of the impulse response of the target source. The output signal $\hat{y}_m(k,t)$ is noise-less, but the target source component in $\hat{y}_m(k,t)$ is greatly distorted. The target source component in the output signal of noise reduction process is $y_m(k,t) - \boldsymbol{a}_m\boldsymbol{y}^e_m(k,t)$. The second term is not always 0-valued and causes distortion of the target source component. The distortion is restored described as follows.

2.2.2 Distortion-Restoration Process

From the m-channel noise-less signal $\hat{y}_m(k,t)$, the target source component in the c-th microphone is restored as follows:

$$\hat{y}_c(k,t) \leftarrow \boldsymbol{w}(k)\hat{\boldsymbol{y}}(k,t), \quad (3)$$

where $\hat{\boldsymbol{y}}(k,t) = [\hat{\boldsymbol{y}}_1(k,t)^T,\ldots,\hat{\boldsymbol{y}}_M(k,t)^T]^T$, ML_2 is the length of the restoration filter, and $\hat{\boldsymbol{y}}_1(k,t) = [\hat{y}_1(k,t),\ldots,\hat{y}_1(k,t-L_2+1)]^T$. The restoration filter $\boldsymbol{w}(k)$ is updated so as to approximate the filtered signal to the target microphone input signal $x_c(k,t)$ as follows:

$$\boldsymbol{w}(k) \leftarrow \underset{\boldsymbol{w}(k)}{\operatorname{argmin}} \quad \Big(\sum_{t\in T_{mix}} ||x_c(k,t) - \boldsymbol{w}(k)\hat{\boldsymbol{y}}(k,t)||^2 + \mu \sum_{t\in T_n} ||\boldsymbol{w}(k)\hat{\boldsymbol{y}}(k,t)||^2\Big), \quad (4)$$

where T_{mix} is the time period when the target sources and the noise sources are mixed, and $\hat{\boldsymbol{y}}(k,t)$ in the second term of the cost function is the residual noise component after the noise reduction process in the noise-only period. The first term of the cost function of Eq. 4 is a penalty for the distortion of the target source, and the second term is a penalty for the residual noise. μ is the weighting coefficient of the second term. When μ is set to 0, residual noise is ignored, and when μ is set to be a large-value, the residual noise is reduced as distortion of the target source increases.

2.3 Time-Frequency Domain ICA (TF-ICA)

The signal representation domain of $x_m(t)$ is converted into the time-frequency domain as $x_m(f,\tau)$, where f is frequency, and τ is the frame index. The i-th source component, $y_i(f,\tau)$ can be extracted by the M channel linear filter $\boldsymbol{w}_i(f)$ as follows:

$$y_i(f,\tau) = \boldsymbol{w}_i(f)\boldsymbol{x}(f,\tau), \quad (5)$$

where $\boldsymbol{x}(f,\tau) = [x_1(f,\tau),\ldots,x_M(f,\tau)]^T$.

The separation filter $\boldsymbol{W}(f) = [\boldsymbol{w}_1(f)^T,\ldots,\boldsymbol{w}_M(f)^T]^T$ at each frequency bin is updated at each iteration t as follows:

$$\boldsymbol{W}(f)_{t+1} = \boldsymbol{W}(f)_t + \eta(\boldsymbol{I} - \langle\varphi(\boldsymbol{Y}(f,\tau))\boldsymbol{Y}(f,\tau)^H\rangle)\boldsymbol{W}(f)_t, \quad (6)$$

where $\langle\cdot\rangle$ is the operator for mathematical expectation, $\boldsymbol{Y}(f,\tau) = [\boldsymbol{y}_1(f,\tau),\ldots,\boldsymbol{y}_M(f,\tau)]$, and φ is the derivative of the probability of sources. In this paper, the derivative of multivariate Laplacian distribution [5] is used. Thanks to multivariate Laplacian distribution, the permutation problem can be solved automatically by Eq. 6. To obtain the i-th source in each microphone, the separation filter is modified to obtain as follows:

$$\boldsymbol{W}_i(f) \leftarrow \boldsymbol{W}^{-1}(f)\boldsymbol{\Lambda}_i\boldsymbol{W}(f), \quad (7)$$

where $\boldsymbol{W}_i(f)$ is the filter that separates the i-th source component from the multichannel input signal, only the (i,i) component of $\boldsymbol{\Lambda}_i$ is 1, and the other components of $\boldsymbol{\Lambda}_i$ are 0. In Eq. 7, ambiguity of the energies of the output signals is solved, because when the separation filter in Eq. 6 is P times larger, the same separation filter in Eq. 7 can be obtained, as $\boldsymbol{W}_i(f) \leftarrow P^{-1}\boldsymbol{W}^{-1}(f)\boldsymbol{\Lambda}_i P\boldsymbol{W}(f) = \boldsymbol{W}^{-1}(f)\boldsymbol{\Lambda}_i\boldsymbol{W}(f)$. The i-th source component at each microphone position is obtained as follows:

$$\boldsymbol{y}_i(f,\tau) = \boldsymbol{W}_i(f)\boldsymbol{x}(f,\tau), \quad (8)$$

where $\boldsymbol{y}_i(f,\tau) = [y_{i,1}(f,\tau),\ldots,y_{i,M}(f,\tau)]^T$, $y_{i,m}(f,\tau)$ is estimation of the i-th source at the m-th microphone.

3. PROPOSED METHOD

TF-ICA can separate sources blindly, but under reverberant environments, the separation performance of TF-ICA is limited. On the other hand, MSP-BF can reduce noise under reverberant environments, but a noise-only period is required. Therefore, to reduce noise under reverberant environments by MSP-BF, the most important task is to detect the noise-only period. To detect the noise-only period, voice activity detection technique (VAD) is frequently used. When the noise sources are stationary and the sound volume of the noise sources are less than that of the target sources, noise-only period can be easily obtained by VAD. However, when the noise sources are non-stationary and their sound volume is equivalent to or more than that of the target sources, noise-only period is difficult to detect. To overcome this problem, a detection technique for noise-only period based on TF-ICA is proposed. In the proposed method, TF-ICA is used for only estimation of the ratio between the sound volume of the target sources and that of the noise sources (target power ratio) at each frame. It can be assumed that estimation of target power ratio is easier than estimation of the waveform of the sound sources. Therefore, even if TF-ICA cannot separate sources completely, target power ratio can be estimated. In TF-ICA, from the view point of computational cost, STFT (short term Fourier transform) is adopted. In the STFT domain, the convolution in the time domain can be approximated as an instantaneous mixture. In MSP-BF, to reduce noise correctly, the signal is modeled as a convolutive mixture in the frequency domain. However, In the STFT domain, it is not possible to obtain an accurate convolutive mixture in each frequency bin. The approximation error of the convolution causes degradation of the noise-reduction performance of MSP-BF. To achieve an accurate convolutive mixture in the frequency domain, DFT filter bank is used in MSP-BF.

The block diagram of the proposed method is shown in Fig. 1. At first, TF-ICA is performed, and the output streams of TF-ICA are divided into either target streams or noise streams by direction of arrival (DOA) estimation. When the estimated direction of the i-th stream is within the target source direction area, Ω_s, the i-th stream is recognized as the target stream. Otherwise, the i-th stream is recognized as the noise stream. Ω_s needs to be pre-defined. After extraction of target stream and noise stream, target power ratio at each frame is estimated. Time period when there are only noise sources (noise-only period) and time period when there is the target source (target-period) are identified by target-power-ratio estimation. The noise reduction filter of MSP-BF is adapted according to the identified noise-only period, and the restoration filter of MSP-BF is adapted according to the identified target period. The target

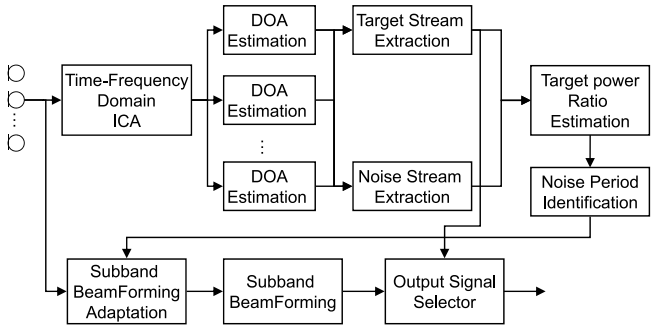


Fig. 1. Block diagram of proposed method

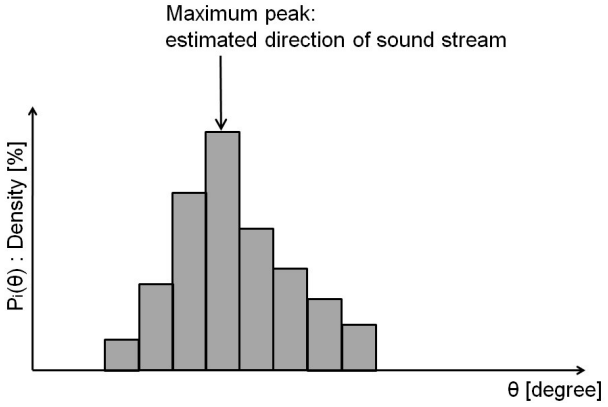


Fig. 2. DOA estimation of each stream

source component in the noisy input signals is extracted by MSP-BF. To avoid performance degradation when there are no noise-only periods, the proposed method selects the better output signal of either TF-ICA or MSP-BF according to the non-linear correlation, *power envelope correlation*, between the output-target-source estimation and the noise-source estimation

3.1 Target Stream and Noise Stream Extraction by DOA Estimation

DOA of the i-th ICA output stream is estimated at each time-frequency point as follows:

$$\theta_i(f,\tau) = \underset{\theta}{\operatorname{argmax}} |\boldsymbol{a}_\theta(f)^H \boldsymbol{y}_i(f,\tau)|, \tag{9}$$

where $\boldsymbol{a}_\theta(f)$ is the steering vector of direction θ, which is calculated by the microphone-array alignment, and $|\boldsymbol{a}_\theta(f)|^2$ is normalized to 1. When DOA of the i-th stream is ϕ, $\boldsymbol{y}_i(f,\tau) = s_i(f,\tau)\boldsymbol{a}_\phi(f)$, and $|\boldsymbol{a}_\theta(f)^H \boldsymbol{a}_\phi(f)|$ has maximum value when $\theta = \phi$. In this case, DOA estimation result $\theta_i(f,\tau)$ is ϕ. And $\boldsymbol{y}_i(f,\tau)$ can be replaced by $\boldsymbol{y}_i(f,\tau) = \boldsymbol{W}_i(f)\boldsymbol{x}(f,\tau)$, and $\boldsymbol{W}_i(f)$ is replaced by Eq. 7. Therefore,

$$\begin{aligned}
\theta_i(f,\tau) &= \underset{\theta}{\operatorname{argmax}} |\boldsymbol{a}_\theta(f)^H \boldsymbol{W}^{-1}(f)\boldsymbol{\Lambda}_i \boldsymbol{W}(f)\boldsymbol{x}(f,\tau)|, \\
\theta_i(f,\tau) &= \underset{\theta}{\operatorname{argmax}} |\boldsymbol{a}_\theta(f)^H \boldsymbol{W}^{-1}(f)\mathbf{1}_i s_i(f,\tau)|, then \\
\theta_i(f,\tau) &= \theta_i(f) = \underset{\theta}{\operatorname{argmax}} |\boldsymbol{a}_\theta(f)^H \boldsymbol{W}^{-1}(f)_i|
\end{aligned}$$

where $\mathbf{1}_i$ is a vector (only i-th element is 1 value, other elements are 0 value), $s_i(f,\tau)$ is the i-th element of $\boldsymbol{W}(f)\boldsymbol{x}(f,\tau)$, and $\boldsymbol{W}^{-1}(f)_i$ is the i-th column of $\boldsymbol{W}^{-1}(f)$. $\theta_i(f,\tau)$ histogram is obtained as follows:

$$P_i(\theta) = \sum_f \begin{cases} \sum_\tau \max(\log|y_i(f,\tau)|^2, 0) & \text{when } \theta = \theta_i(f) \\ 0 & \text{otherwise} \end{cases}, \tag{10}$$

where $\delta(x)$ is 1 value only when x is true, otherwise, $\delta(x)$ is 0. DOA of the i-th stream θ_i is estimated by peak-searching for $P_i(\theta)$ depicted in Fig. 2 DOA of the target source (the target source area Ω_s) is assumed to be pre-defined as "from -30 degrees to +30 degrees". When DOA of the i-th source is within Ω_s, this source is recognized as the target source. Otherwise, this source is recognized as the noise source.

3.2 Noise Period Identification Based on Target Power Ratio Estimation

Target power ratio at each time point, $R(\tau)$, is calculated from TF-ICA output signals as follows:

$$R(\tau) = \frac{P_s(\tau)}{P_s(\tau) + P_n(\tau)}, \tag{11}$$

where $P_s(\tau)$ and $P_n(\tau)$ are power estimates of the target source and the noise source and are defined as $P_s(\tau) = \sqrt{\sum_f \sum_{\theta_i \in \Omega_s} |\boldsymbol{y}_i(f,\tau)|^2}$, $P_n(\tau) = \sqrt{\sum_f \sum_{\theta_i \notin \Omega_s} |\boldsymbol{y}_i(f,\tau)|^2}$. Noise-only period is identified by using $R(\tau)$ in the following procedure.

1. Execute K-means clustering of $R(\tau)$.
2. Sort clusters by the center value of each cluster in ascending order. Let r_i is the center value of the i-th cluster, and p_i is the occupancy of the i-th cluster, which is defined as $p_i = \frac{N_i}{\sum_j N_j}$, where N_j is the number of the elements connected with the i-th cluster.
3. Sum p_i until the summed value $p_{sum}$is smaller than the given threshold. Let $p_{sum} = \sum_{i=1}^{i_{th}} p_i$. i_{th} is the index of the cluster whose center value is the biggest among the summed clusters.
4. Regard clusters whose index is less than i_{th} are as noise clusters, and other clusters as target clusters.
5. Set the average value R_{th} between the i_{th}-th cluster and the $(i_{th}+1)$-th cluster as the threshold for identification of noise-only period and target period.
6. Perform voice-activity-detection based on $R(\tau)$ and the threshold R_{th}. Short pause of the target source is permitted. The frames at which $R(\tau)$ transiently exceeds threshold are also regarded as a target-period.
7. Regard continuous frames in which there is no target source as noise-only period.
8. When a noise-only period is not detected, the threshold of the occupancy of the noise source becomes lower, and return to step 2. Similarly, when the target-period is not detected, threshold of the occupancy becomes higher, and return to step 2. When both noise-only period and target-period are detected, stop the procedure.

When the target source is mixed into a detected noise-only period, the noise reduction filter of MSP-BF is adapted so as to cancel the target source. On the other hand, a false detection of an noise-only period as the target period is not critical. The proposed method is designed not to mistakenly detect a target period as an noise-only period.

3.3 Output Signal Selector

Let $\hat{y}_c(f,\tau)$ be the target source estimate in the time-frequency domain, and $x_c(f,\tau)-\hat{y}_c(f,\tau)$ be the noise-source estimate. The coefficient of correlation between the target source and the noise source is one candidate for the criteria of output-signal selector. However, primarily, the noise-source estimate and the target source estimate of both MSP-BF and TF-ICA are uncorrelated. Therefore, the simple correlation between the noise source estimate and the target source estimate is not suitable for the criteria of the output signal selector. In this paper, a power-envelope correlation, which is not the criterion of the nonlinear correlation in TF-ICA adaptation, is set as the criterion of the output signal selector. Power-envelope correlation at frequency f, $r(f)$, is defined as the correlation between $|\hat{y}_c(f,\tau)|^2$ and $|x_c(f,\tau)-\hat{y}_c(f,\tau)|^2$. The output signal selector selects the better signal $y_{out}(f,\tau)$ at every frequency bin as follows:

$$y_{out}(f,\tau)=\begin{cases} y_{ica}(f,\tau) & \text{if } r_{msp}(f)\geq r_{ica}(f) \\ y_{msp}(f,\tau) & \text{else} \end{cases}, \qquad (12)$$

where $y_{ica}(f,\tau)$ is the target-source estimation of TF-ICA at each time-frequency point, $y_{msp}(f,\tau)=\mathrm{STFT}(\mathrm{Synth}(y_c(k,t)))$, $y_c(k,t)$ is the output signal of MSP-BF obtained from Eq. 3, $\mathrm{STFT}(\cdot)$ is the operation of STFT, $\mathrm{Synth}(\cdot)$ is the operation of the synthesis operation from subband domain to time domain, and r is the correlation coefficient of power envelelopes of the estimated target source and the estimated noise source.

4. EXPERIMENT

Noise reduction performance of the proposed method under a reverberant environment (RT_{60} is about 300 ms) was evaluated. The noisy input signal was made by convolution of the impulse responses that were recorded in the experimental environment. The noise sources were human speech, office noise. The target sources were set to be human speech. The speech sources were obtained from ATR speech database. The number of the microphones in the microphone array was 12. The length of the microphone array was 0.70 m. The number of target sources was 1, and the number of noise sources was set to 1 or 2. The location of the sound sources are shown in Fig .3. SNR of the noisy input signal was set to 0 dB. The wavelength of the noisy input signal was set to about 10 seconds. The sampling rate was set to be 8000 Hz. The evaluation measures are SIR (signal interference rate), and PESQ [10]. SIR is defined as $SIR=10\log_{10}\frac{\sum_{t=0}^{L_{wave}}(s(t))^2}{\sum_{t=0}^{L_{wave}}(\hat{n}(t))^2}$, $s(t)$ is target source signal, $\hat{n}(t)$ is the residual noise signal after noise reduction, $\hat{n}(t)$ is calculated by the output signal after noise reduction with noise only input signal, and L_{wave} is the length of waves. When SIR and PESQ are high, performance of the noise reduction system is also high.

The proposed method was evaluated in two forms. One form is the combination of TF-ICA and MSP-BF without the output selector ("PROPOSED"). The another is the combination method of TF-ICA, MSP-BF, and the output selector ("SELECT"). These methods are compared with TF-ICA and MSP-BF using the given noise-only period ("ORACLE"). The upper performance limit of the proposed method is equivalent to "ORACLE".

Table. 1 (a) lists the experimental results for the case of one human-speech noise. "density" is the ratio of the length of the target-period L_{target} to total wavelength L_{wave}, $\frac{L_{target}}{L_{wave}}$. Table. 1 (b) lists the experimental results for office noise. Table. 1 (c) lists the experimental results for two human-speech noises. When the density of the target sources are from 34% to 74% , SIR and PESQ of "PROPOSED" are much higher than those of TF-ICA. As for comparison of "PROPOSED" with "SELECT", SIR of "SELECT" is slightly less than that of "PROPOSED". However, PESQ of "SELECT" is almost the same as that of "PROPOSED". The difference between the output signals of "SELECT" and "PROPOSED" is considered to be less than the audible level. When the density of the target source is 100%, there is no noise-only period. Therefore, the output signal of "PROPOSED" is distorted. PESQ of "PROPOSED" is less than that of TF-ICA. However, PESQ of "SELECT" is better than that of "PROPOSED". This indicates that thanks to the output signal selector, the distorted-frequency component of MSP-BF is replaced by the frequency component of TF-ICA. The spectrograms of the output signals and the result of noise-only period identification are shown in Fig. 4. Compared with the residual noise signal for TF-ICA, that for "PROPOSED" and "SELECT" is significantly reduced.

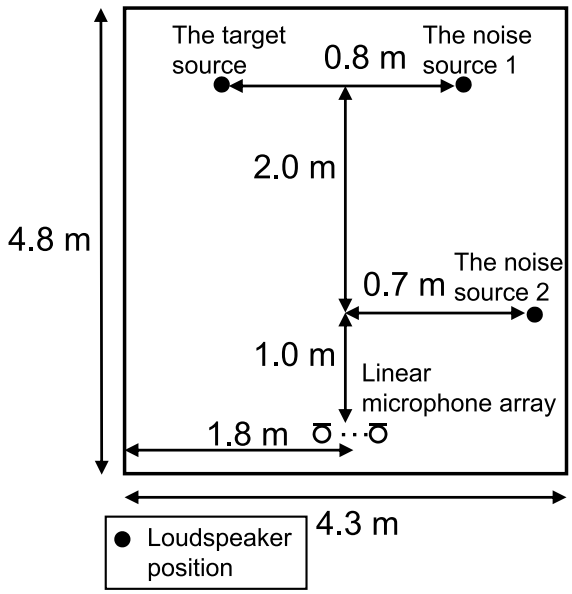


Fig. 3. Location of sound sources

5. CONCLUSION

In this paper, we proposed a novel beamforming method that is complementary combined with time-frequency domain independent component analysis (TF-ICA). Noise-only period can be obtained by noise period identification using the output signals of TF-ICA. A noise reduction filter of MSP-BF (beamforming based on multichannel spatial prediction) is adapted according to the obtained noise-only period. To avoid performance degradation, the proposed method selects the better output signal obtained by TF-ICA and MSP-BF. The cost function used in the output-signal selector is power-envelope correlation between output-target-source estimation and noise-source estimation. Experimental results for a reverberant environment indicate that the proposed method outperforms TF-ICA. Furthermore, thanks to the output signal selector, even when there are no noise-only periods, performance degradation can be avoided.

REFERENCES

[1] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," In *Proc. IEEE*, vol. 60, no. 8, pp. 926-935, 1972.

[2] L. J. Griffith and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. AP*, vol. 30, i. 1, pp. 27-34, 1982.

[3] J. Benesty, et al., "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. ASLP*, vol. 15, pp. 1053-1065, 2007.

[4] T. Nishikawa, et al., "Blind source separation based on multi-stage ICA combining frequency-domain ICA and time-domain ICA", *Proc. ICASSP2002*, vol. I, pp. 2938-2941, 2002.

Table 1. Experimental results: the better result of "PROPOSED" and "SELECT" is highlighted in each column.

(a) The number of noise sources is 1. Noise source is human speech.

#	density	SIR (dB)				PESQ				
		TF-ICA	PROPOSED	SELECT	ORACLE	PRE PESQ	TF-ICA	PROPOSED	SELECT	ORACLE
1	34 %	18.69	**46.82**	46.77	53.00	1.99	2.65	3.89	**3.92**	3.99
2	53 %	19.29	**45.84**	44.01	52.15	2.22	2.79	**3.94**	3.93	4.02
3	74 %	17.83	**38.21**	31.66	51.98	2.20	2.77	3.41	**3.49**	4.04
4	100 %	15.91	**20.64**	18.30	51.22	2.05	2.59	1.39	**2.35**	4.04

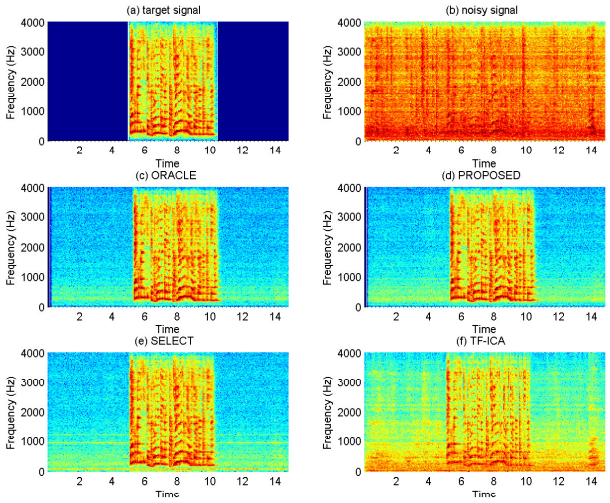
(b) The number of noise sources is 1. Noise source is office noise.

#	density	SIR (dB)				PESQ				
		TF-ICA	PROPOSED	SELECT	ORACLE	PRE PESQ	TF-ICA	PROPOSED	SELECT	ORACLE
1	34 %	18.25	43.25	**43.38**	51.13	1.85	2.82	3.66	**3.72**	3.79
2	53 %	18.38	**42.61**	29.97	50.32	1.85	2.75	3.76	**3.77**	3.99
3	74 %	17.68	**41.32**	39.02	49.50	1.84	2.79	**3.72**	3.71	4.06
4	100 %	13.55	**36.47**	14.99	49.53	1.88	2.59	1.55	**2.58**	4.00

(c) The number of noise sources is 2. Noise sources are human speech.

#	density	SIR (dB)				PESQ				
		TF-ICA	PROPOSED	SELECT	ORACLE	PRE PESQ	TF-ICA	PROPOSED	SELECT	ORACLE
1	34 %	18.79	**40.71**	29.93	47.41	1.85	2.67	**3.37**	3.35	3.60
2	53 %	15.13	**40.61**	31.18	45.85	2.07	2.73	**3.60**	3.57	3.70
3	74 %	17.14	**36.54**	34.39	46.17	2.08	2.79	**3.47**	3.47	3.72
4	100 %	14.53	11.13	**14.96**	45.21	1.80	2.53	1.82	**2.34**	3.72

(a) Spectrograms of the output signals



(b) Result of noise-only period identification

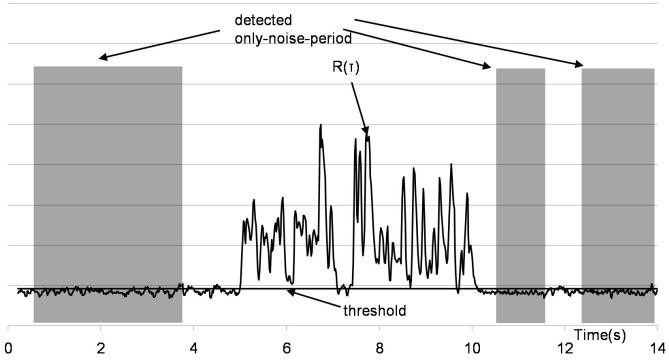


Fig. 4. Spectrograms of the output signals and result of noise-only period identification: The number of noise sources is 1.

[5] T. Kim, et al., "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASAP*, vol. 15, no. 1, pp. 70-79, 2007.

[6] S. Araki, et al., "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. SAP*, vol. 11, no. 2, pp. 109-116, 2003.

[7] S. Araki, et al., "Subband-based blind separation for convolutive mixtures of speech," *IEICE Trans. Fundamentals*, E88-A(12), pp. 3593-3603, 2005.

[8] M. Togami, et al., "Subband nonstationary noise reduction based on multichannel spatial prediction under reverberant environments," *Proc. ICASSP2009*, pp. 133-136, 2009.

[9] R. Crochiere and L. Rabiner, Multirate Digital Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[10] ITU-T Rec. P. 862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," International Telecommunication Union, Geneva, Switzerland, 2001.