A ROBUST AND PRECISE METHOD FOR SOLVING THE PERMUTATION PROBLEM OF FREQUENCY-DOMAIN BLIND SOURCE SEPARATION

Hiroshi Sawada   *Ryo Mukai*   *Shoko Araki*   *Shoji Makino*

NTT Communication Science Laboratories, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
{sawada,ryo,shoko,maki}@cslab.kecl.ntt.co.jp

ABSTRACT

This paper presents a robust and precise method for solving the permutation problem of frequency-domain blind source separation. It is based on two previous approaches: the direction of arrival estimation and the inter-frequency correlation. We discuss the advantages and disadvantages of the two approaches, and integrate them to exploit their respective advantages. We also present a closed form formula to estimate the directions of source signals from a separating matrix obtained by ICA. Experimental results show that our method solved permutation problems almost perfectly for a situation that two sources were mixed in a room whose reverberation time was 300 ms.

1. INTRODUCTION

Blind source separation (BSS) is a technique for estimating original source signals using only sensor observations, which consist of mixtures of the original signals. If the mixture is instantaneous, we can directly employ independent component analysis (ICA) [1, 2] to separate the mixed signals. In a real room environment, however, signals are mixed in a convolutive manner with reverberations. This makes the BSS problem difficult since we need a set of filters, not just scalars, to separate the signals. One of the major methods to obtain such separating filters is frequency-domain BSS [3–10], where a convolutive mixture in the time domain is converted into multiple instantaneous mixtures in the frequency domain. Thus, we can apply ICA to instantaneous mixtures in every frequency bin.

The problem with frequency-domain BSS is the indeterminacy of permutation that is inherent to ICA. We need to map a separated signal at each frequency to a target source signal so that we properly reconstruct the separated signal in the time domain. Various approaches have been proposed to the permutation problem. Making separating matrices smooth in the frequency domain is one solution. This has been realized by averaging separating matrices with adjacent frequencies [3], limiting the filter length in the time domain [4], or considering the coherency of separating matrices at adjacent frequencies [5]. Another approach is based on direction of arrival (DOA) estimation in array signal processing. By analyzing the directivity patterns formed by a separating matrix, source directions can be estimated and therefore permutations can be aligned [6, 7]. If source signals are speech, we can employ the inter-frequency correlations of signal envelopes to align permutations [8, 9]. Each of these approaches has different characteristics. They may perform well under certain specific conditions but not others. Therefore, we believe that integrating some of these approaches is one way of obtaining better performance.

In this paper, we propose a new method for solving the permutation problem, which incorporates two of the previous approaches. The first is the DOA approach, which is described in Sec. 3. The second is based on inter-frequency correlations, and is discussed in Sec. 4. Our new method is proposed in Sec. 5. The experimental results reported in Sec. 6 are very promising.

As another contribution, we propose a new way of estimating the direction of sources in Sec. 3.2. Unlike conventional methods [6, 7], it does not require the calculation of directivity patterns. Instead, it calculates the directions of target signals directly from an estimated mixing matrix, which is the inverse of a separating matrix obtained by ICA.

2. FREQUENCY-DOMAIN BSS

Suppose that P source signals $s_p(t)$ are mixed and observed at Q sensors $x_q(t) = \sum_{p=1}^{P} \sum_k h_{qp}(k) s_p(t - k)$, where $h_{qp}(k)$ represents the impulse response from source p to sensor q. The goal of BSS is to obtain separated signals $y_1(t), \ldots, y_P(t)$ that are estimates of the source signals $s_1(t), \ldots, s_P(t)$. The separating system typically consists of a set of FIR filters $w_{rq}(k)$ that produces separated signals $y_r(t) = \sum_{q=1}^{Q} \sum_k w_{rq}(k) x_q(t - k)$.

This paper employs a frequency-domain approach where frequency responses $W_{rq}(f)$ of the separating filter $w_{rq}(k)$ are first calculated. By L-point short time DFT, time-domain signals $x_q(t)$ are converted into frequency-domain time-series signals $X_q(f, m)$, where $f = 0, f_s/L, \ldots, f_s(L-1)/L$ (f_s: sampling frequency), and m is the frame index. As-

sume that $\mathbf{X}(f,m)$ is a Q-dimensional vector $\mathbf{X}(f,m) = [X_1(f,m), \ldots, X_Q(f,m)]^T$. To obtain frequency responses $W_{rq}(f)$, we solve the ICA problem

$$\mathbf{Y}(f,m) = \mathbf{W}(f)\mathbf{X}(f,m),$$

where $\mathbf{Y}(f,m) = [Y_1(f,m), \ldots, Y_P(f,m)]^T$ and $\mathbf{W}(f)$ is a $P \times Q$ matrix whose elements are $W_{rq}(f)$. $Y_r(f,m)$ is a frequency-domain representation of $y_r(t)$.

The ICA algorithm we use is the information maximization approach [1] combined with the natural gradient [2]. A separating matrix $\mathbf{W}$ is improved by the learning rule

$$\Delta\mathbf{W} = \mu\left[\mathbf{I} - \langle\Phi(\mathbf{Y})\mathbf{Y}^H\rangle\right]\mathbf{W},$$

where μ is a step-size parameter, $\langle\cdot\rangle$ denotes the averaging operator, and $\Phi(\cdot)$ is a nonlinear function for a complex signal $Y_r = |Y_r|\,e^{j\cdot angle(Y_r)}$. We use

$$\Phi(Y_r) = -\frac{\partial}{\partial|Y_r|}\log p(|Y_r|)\ e^{j\cdot angle(Y_r)}$$

as a nonlinear function assuming that the density $p(Y_r)$ is independent of the angle [10].

An ICA solution has permutation ambiguity: if we permute the rows of $\mathbf{W}(f)$, it is still a solution. Thus, we have to align the rows of $\mathbf{W}(f)$ so that $Y_r(f,m)$ at all frequencies correspond to the same source $s_p(t)$. After solving this permutation problem, we obtain separating filters $w_{rq}(k)$ by applying inverse DFT to $W_{rq}(f)$.

3. THE DIRECTION OF ARRIVAL APPROACH

In this section, we discuss how to estimate the directions of source signals, and align permutations based on them. If half the minimum wavelength of the source signals is longer than the sensor spacing (e.g., conditions in Table 1), there is no spatial aliasing. In most of such cases, each row of $\mathbf{W}(f)$ forms spatial nulls in the directions of jammer signals and extracts a target signal in another direction. Once we estimate the directions $\Theta(f) = [\theta_1(f), \ldots, \theta_P(f)]^T$ of target signals extracted by every row of $\mathbf{W}(f)$, we can align the permutations according to $\Theta(f)$.

3.1. Directivity patterns and null directions

We first review the method [6, 7] that solves the permutation problem by plotting the directivity pattern of each output $Y_r(f,m)$. Let d_q be the position of sensor q (we assume linearly arranged array sensors), and θ_p be the direction of source s_p (the direction orthogonal to the array is $90°$). In beamforming theory [11], the frequency response of an impulse response $h_{qp}(t)$ is approximated as

$$H_{qp}(f) = e^{j2\pi fc^{-1}d_q\cos\theta_p}, \tag{3.1}$$

where c is the propagation velocity. In this approximation, we assume a plane wavefront and no reverberation. The frequency response $B_{rp}(f)$ from a source s_p to a separated signal y_r can be expressed as $B_{rp}(f) = \sum_{q=1}^{Q} W_{rq}(f) \cdot$
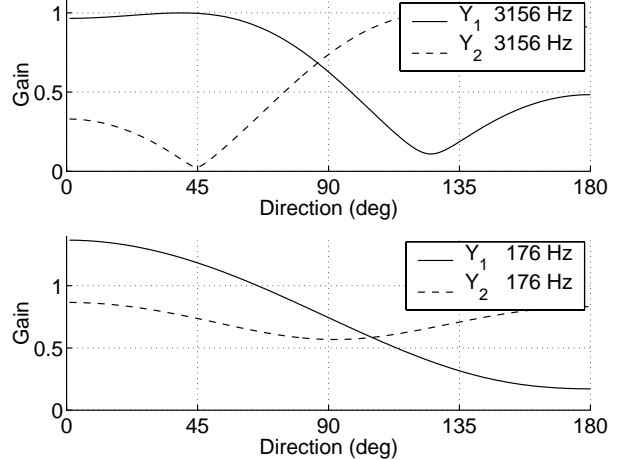

Fig. 1. Directivity patterns

$H_{qp}(f) = \sum_{q=1}^{Q} W_{rq}(f) \cdot e^{j2\pi fc^{-1}d_q\cos\theta_p}$. If we regard θ_p as a variable θ, the formula is expressed as

$$B_r(f,\theta) = \sum_{q=1}^{Q} W_{rq}(f) \cdot e^{j2\pi fc^{-1}d_q\cos\theta}. \tag{3.2}$$

It changes according to the direction θ, and thus is called a directivity pattern.

Figure 1 shows directivity patterns for two sources. The upper part (3156 Hz) shows that output Y_1 extracts a source signal originating from around $45°$ and suppresses the other signal coming from around $125°$, which is called a null direction. With a similar consideration regarding Y_2, we estimate the directions $\Theta(3156) = [45, 125]^T$ of the target signals. A simple way to solve the permutation problem is to permute $\mathbf{W}(f)$ at each frequency so that the elements of $\Theta(f)$ are sorted. However, not every frequency bin gives us such an ideal directivity pattern. The lower part of Fig. 1 is the pattern at a low frequency (176 Hz). We see that the null is not well formed for Y_1 and the null of Y_2 is in an obscure direction. In fact, we cannot estimate $\Theta(176)$ or decide a permutation for this frequency with confidence.

Now we state three problems with this method: 1) directions of arrival cannot be well estimated at some frequencies, especially at low frequencies where the phase difference caused by a sensor spacing is very small, 2) the calculation of null directions by plotting directivity patterns is time consuming, and 3) estimating DOAs from null directions is difficult when there are more than two sources. The first problem will be solved in Sec. 5, and the other two problems will be solved in the next two subsections.

3.2. Novel way of estimating DOAs from $\mathbf{W}^{-1}$

Instead of plotting directivity patterns and searching for the minimum as a null direction, we propose a new way of estimating the directions $\Theta(f)$ of source signals directly from the inverse $\mathbf{W}^{-1}$ of separating matrix $\mathbf{W}$.

After ICA is solved, we can estimate the frequency response of the mixing system by $\mathbf{H}(f) = \mathbf{W}^{-1}(f)$. An element $H_{qp}(f)$ of the matrix $\mathbf{H}(f)$ may have an arbitrary amplitude. Since the approximation (3.1) of the mixing system does not suit this situation, we remodel the mixing system with attenuation A_{qp} (real-valued) and phase modulation $e^{j\varphi_p}$ at the origin:

$$H_{qp}(f) = A_{qp} \, e^{j\varphi_p} e^{j2\pi f c^{-1} d_q \cos \theta_p}.$$

Here, we should take account of the scaling and permutation ambiguities of ICA: the columns of $\mathbf{H}(f)$ can have arbitrary scales, and be permuted arbitrarily. The scaling ambiguity can be eliminated by taking the ratio between two elements $H_{qp}(f)$ and $H_{q'p}(f)$ corresponding to the same source p:

$$H_{qp}/H_{q'p} = A_{qp}/A_{q'p} \, e^{j2\pi f c^{-1}(d_q - d_{q'}) \cos \theta_p}.$$

Then, taking the angle yields a formula for estimating θ_p :

$$\theta_p = \cos^{-1} \frac{angle(H_{qp}/H_{q'p})}{2\pi f c^{-1}(d_q - d_{q'})}. \tag{3.3}$$

Because of the permutation ambiguity, θ_p may not correspond to s_p but to another source signal. This formula does not incur the computational cost to search a directivity pattern for the minimum. Moreover, it can be applied even when there are more than two sources.

3.3. Equivalence between θ_p and a null direction

For a two-source case, we prove that θ_p calculated by (3.3) is the same as a null direction that is the minimum of a directivity pattern. When $|B_r(f, \theta)|$ is minimized, θ corresponds to a null direction. Let $\alpha_q = 2\pi f c^{-1} d_q$ and f be omitted in (3.2). The value to be minimized is

$$
\begin{aligned}
J(\theta) &= B_r(\theta) \cdot B_r(\theta)^* \\
&= (W_{r1} e^{j\alpha_1 \cos \theta} + W_{r2} e^{j\alpha_2 \cos \theta}) \cdot \\
&\quad (W_{r1}^* e^{-j\alpha_1 \cos \theta} + W_{r2}^* e^{-j\alpha_2 \cos \theta}).
\end{aligned}
$$

Let $\alpha = \alpha_2 - \alpha_1$. The first and second derivatives are

$$
\begin{aligned}
\frac{dJ}{d\theta} &= -\alpha \sin \theta \cdot 2 \, im(W_{r1} W_{r2}^* e^{-j\alpha \cos \theta}), \\
\frac{d^2 J}{d\theta^2} &= -\alpha \cos \theta \cdot 2 \, im(W_{r1} W_{r2}^* e^{-j\alpha \cos \theta}) \\
&\quad -\alpha^2 \sin^2 \theta \cdot 2 \, re(W_{r1} W_{r2}^* e^{-j\alpha \cos \theta})
\end{aligned}
$$

where *re* and *im* extract the real and imaginary part of a complex, respectively. If $angle(W_{r1} W_{r2}^* e^{-j\alpha \cos \theta}) = \pi$, $\frac{dJ}{d\theta}$ is zero and $\frac{d^2 J}{d\theta^2}$ is positive, and $J(\theta)$ is minimized. Thus, the null direction formed by the r-th row of $\mathbf{W}$ is given by

$$angle(-W_{r1} W_{r2}^*) = \alpha \cos \theta_r^{null} \quad \Leftrightarrow$$

$$\theta_r^{null} = \cos^{-1} \frac{angle(-W_{r1}/W_{r2})}{2\pi f c^{-1}(d_2 - d_1)}.$$

Considering $H_{21} = -W_{21}/det(\mathbf{W})$ and $H_{11} = W_{22}/det(\mathbf{W})$, we see that θ_1 and θ_2^{null} are the same:

$$\theta_1 = \cos^{-1} \frac{angle(H_{21}/H_{11})}{2\pi f c^{-1}(d_2 - d_1)} = \cos^{-1} \frac{angle(-W_{21}/W_{22})}{2\pi f c^{-1}(d_2 - d_1)} = \theta_2^{null}.$$

The derivation of θ_r^{null} is based on derivatives. We have another derivation of θ_r^{null} based on the graphical interpretation of a directivity pattern [12].
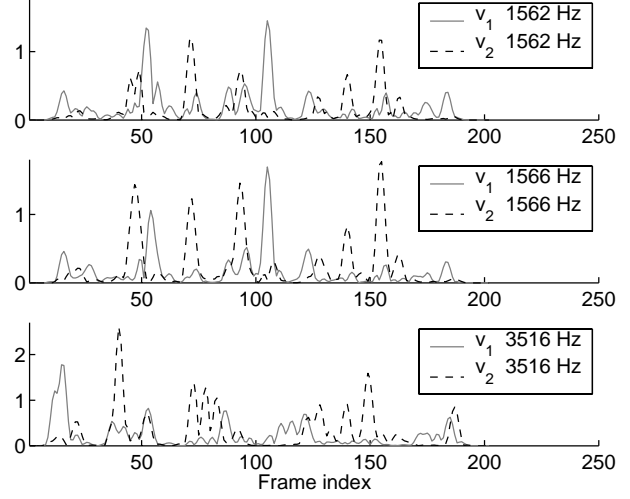

Fig. 2. Envelopes at different frequencies

4. THE CORRELATION APPROACH

This section discusses an approach to permutation alignment based on inter-frequency correlations of speech signals [8, 9]. We use the envelope

$$v_r^f(m) = |Y_r(f, m)|$$

of a separated signal $Y_r(f, m)$ to measure correlations. Let us define the correlation of two signals $x(m)$ and $y(m)$ as

$$cor(x, y) = [\langle x \cdot y \rangle - \langle x \rangle \cdot \langle y \rangle]/(\sigma_x \cdot \sigma_y),$$

where $\langle \cdot \rangle$ is an averaging operator and σ is a standard deviation. Based on this definition, $cor(x, x) = 1$, and $cor(x, y) = 0$ if x and y are uncorrelated. Envelopes have high correlations at neighboring frequencies if separated signals correspond to the same speech signal. Figure 2 shows an example. Two envelopes v_1^{1562} and v_1^{1566}, as well as v_2^{1562} and v_2^{1566}, are highly correlated. Thus, calculating such correlations helps us to align permutations.

Henceforth let π denote a permutation: $\{1, \dots, P\} \to \{1, \dots, P\}$. A simple criterion for deciding the permutation π_f of frequency f is to maximize the sum of the correlations between neighboring frequencies within distance D:

$$\pi_f = \text{argmax}_\pi \sum_{|g-f| \le D} \sum_p cor(v_{\pi(p)}^f, v_{\pi_g(p)}^g), \quad (4.1)$$

where π_g is the permutation at frequency g. This criterion is based on local information and has a drawback in that mistakes in a narrow range of frequencies may lead to the complete misalignment of the frequencies beyond that point. To avoid this problem, the method in [9] does not limit the frequency range in which correlations are calculated. It decides permutations one by one based on the criterion:

$$\pi_f = \text{argmax}_\pi \sum_p cor(v_{\pi(p)}^f, \sum_{g \in F} v_{\pi_g(p)}^g), \quad (4.2)$$

where F is a set of frequencies in which the permutation is decided. This method assumes high correlations of envelopes even between frequencies that are not close neigh-

```
/* Fix permutations by the DOA approach */
for ( ∀f ) {
    Θ(f) = DOA(f, W(f))
    Θ_s(f) = sort(Θ(f))
}
m_Θ = ⟨Θ_s(f)⟩_f    /* averaged directions */
F = ∅    /* the set of fixed frequencies */
for ( ∀f ) {
    if ( confident(Θ(f), m_Θ, W(f)) ) {
        π_f = getPermutation(Θ(f))
        F = F ∪ {f}
    }
}
/* Fix permutations by neighboring correlations */
while ( ∃f ∉ F ) {
    for ( ∀f ∉ F ) {
        c_f = max_π ∑_{|g-f|≤D, g∈F} ∑_p cor(v^f_{π(p)}, v^g_{π_g(p)})
        π_f = argmax_π ∑_{|g-f|≤D, g∈F} ∑_p cor(v^f_{π(p)}, v^g_{π_g(p)})
    }
    i = argmax_f c_f
    F = F ∪ {i}
    c_i = 0
}
```

Fig. 3. Pseudo-code for the integrated method

```
/* Fix permutations by harmonic structure */
for ( ∀f ∉ F ) {
    Ha = setOfHarmonics(f);
    c_f = max_π ∑_{g∈Ha∩F} ∑_p cor(v^f_{π(p)}, v^g_{π_g(p)})
    if ( c_f ≥ threshold ) {
        π_f = argmax_π ∑_{g∈Ha∩F} ∑_p cor(v^f_{π(p)}, v^g_{π_g(p)})
        F = F ∪ {f}
    }
}
```

Fig. 4. Pseudo-code for the harmonic method

Our method benefits from both advantages: the robustness of the DOA approach and the preciseness of the correlation approach. Figure 3 shows the pseudo-code.

We first fix permutations at some frequencies where the confidence of the DOA approach is sufficiently high. The procedure *confident* decides whether the confidence is high enough. Our criteria for the decision are: 1) the number of estimated directions is the same as the number of sources, 2) the directions $\Theta(f)$ do not differ greatly from the averaged directions m_Θ, 3) the SNR calculated by the frequency responses $B_r(f, \theta_p)$ for each direction is sufficiently large.

Then, we decide the permutations for the remaining frequencies based on neighboring correlations without changing the permutations fixed by the DOA approach. The permutations are decided in order of the sum of correlations with fixed frequencies $g \in F$ within distance $|g - f| \leq D$.

This method does not result in a large misalignment as long as the permutations fixed by the DOA approach are correct. Moreover, the correlation part compensates for the lack of preciseness of the DOA approach.

5.2. Exploiting the harmonic structure of signals

The method proposed above works very well in many cases. However, there is a case where the DOA approach does not provide any fixed permutation with confidence in a certain range of frequencies. This occurs particularly at low frequencies where it is hard to estimate DOAs as discussed in Sec. 3. In such a case, the proposed method has to align permutations for the range solely through the use of neighboring correlations, and may yield consecutive misalignments.

To cope with this problem, we exploit the harmonic structure of a speech signal. As alluded to in Sec. 4, there are strong correlations among the envelopes of a fundamental frequency f and its harmonics $2f$, $3f$ and so forth. Suppose that the permutation is not fixed at frequency f but fixed at frequencies $2f$ and $3f$. If the correlation

$$c_f = \max_\pi \sum_{g=2f,3f} \sum_p \text{cor}(v^f_{\pi(p)}, v^g_{\pi_g(p)})$$

is sufficiently large, we can fix the permutation at frequency

bors. This assumption is not satisfied for all pairs of frequencies, although a high correlation can be assumed for a fundamental frequency and its harmonics. As shown in Fig. 2, v_r^{1566} and v_r^{3516} do not have a high correlation. Therefore, this method still has a drawback in that permutations may be misaligned at many frequencies.

5. NEW ROBUST AND PRECISE METHOD

In this section, we consider integrating the two approaches discussed above to solve the permutation problem robustly and precisely.

5.1. Integration of the two approaches

Let us begin by reviewing the characteristics of the two approaches.

robustness The direction of arrival (DOA) approach is robust since a misalignment at a frequency does not affect other frequencies. The correlation approach is not robust as discussed in Sec. 4.

preciseness The DOA approach is not precise since the evaluation is based on the approximation of a mixing system as explained in Sec. 3. The correlation approach is precise as long as signals are well separated by ICA since the measurement is based on separated signals.

Table 1. Experimental conditions

Length of source signal	6 sec
Direction of sources	$120°$ and $50°$ (2 sources)
Distance between 2 sensors	$d = 4$ cm
Reverberation time	$T_R = 300$ ms
Sampling rate	$f_s = 8$ kHz
Frame size of DFT	$L = 2048$
Frequency resolution	$\Delta f = f_s/L = 3.90625$ Hz
Distance to take correlation	$D = 3 \cdot \Delta f$
Nonlinear function	$\Phi(Y_r) = e^{j \cdot angle(Y_r)}$

Fig. 5. Overall results

f with confidence. Figure 4 shows the pseudo-code for the harmonic method. The procedure *setOfHarmonics* provides a set of harmonic frequencies of f, and *threshold* controls the confidence.

To incorporate the above idea, the final version of our method fixes all permutations with four steps:

1. By the DOA approach (the upper part of Fig. 3)

2. By neighboring correlations (the lower part of Fig. 3) with the exception that the while loop terminates if the maximum of c_f is too small.

3. By the harmonic method (Fig. 4)

4. By neighboring correlations (the lower part of Fig. 3) again without the exception.

There are two important points to note as regards the final version. The first is that the method becomes more robust because of the exception in step 2. We do not fix the permutations for consecutive frequencies without high confidence. The second point is that step 3 works well only if most of the other permutations are fixed. This means that the harmonic method alone does not work well and we need steps 1 and 2 to fix most of the permutations.

6. EXPERIMENTAL RESULTS

We performed experiments to separate speech signals in a reverberant environment whose conditions are summarized in Table 1. We generated mixed signals by convolving a speech signal $s_p(t)$ and an impulse response $h_{qp}(t)$ so that we could calculate signal-to-noise ratios (SNRs) by

$$10 \log[\sum_{r=p} \sum_t y_{rp}(t)^2] - 10 \log[\sum_{r \neq p} \sum_t y_{rp}(t)^2],$$

where $y_{rp}(t) = \sum_{q=1}^{Q} \sum_k w_{rq}(k) x_{qp}(t-k)$ and $x_{qp}(t) = \sum_k h_{qp}(k) s_p(t-k)$. Figure 5 shows the overall results of the experiments. We separated 12 combinations of speech signals with 6 different methods for the permutation problem: the DOA approach "D", the correlation approach based on (4.1) "C1", the correlation approach based on (4.2) "C2", the first version of the integrated method (Fig. 3) "D+C1", the final version of the integrated method (Figs. 3 and 4) "D+C1+Ha", and the optimal permutation "Optimal" based on maximizing the SNR at each frequency. Although "Optimal" is not a realistic solution, we used it to estimate the
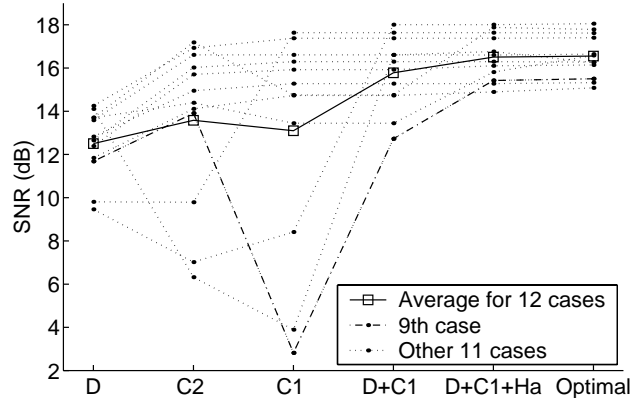
upper bounds of performance.

The performance with "D" is stable, but not sufficient. The results with "C1" and "C2" are not stable and sometimes very poor, although most of the time they are very good. Both integrated methods "D+C1" and "D+C1+Ha" offer stable and good results. Especially, the method exploiting the harmonic structure "D+C1+Ha" offers almost the same results as "Optimal".

Now we examine the effectiveness of the integrated methods by looking at the 9th combination of speech signals in detail. Figure 6 shows the SNRs at each frequency for "C1", "D+C1" and "D+C1+Ha". We see a large region (from 450 Hz to 1400 Hz) of permutation misalignments for the "C1" case, where permutations were decided only with neighboring correlations. We see in Fig. 7 that the correlations around 1400 Hz do not have a large absolute value, and the criterion based on (4.1) does not provide a clear-cut decision. Therefore, there was a decent possibility that the permutations were misaligned around 1400 Hz in this case.

With the "D+C1" method, the misalignments of the region (from 450 Hz to 1400 Hz) were corrected. This is because the DOA approach in the first step provided correct permutations for some frequencies in the region. Figure 8 shows the DOA estimations for each frequency with confidence. We see many estimations from 450 Hz to 1400 Hz. However, there was no DOA estimation with confidence at frequencies lower than 250 Hz. This is why consecutive misalignments occurred even for "D+C1". As shown at the bottom of Fig. 6, the misalignments were corrected with the "D+C1+Ha" method. This shows the effectiveness of exploiting the harmonic structure for low frequencies.

7. CONCLUSION

We proposed a robust and precise method for solving the permutation problem. Our method integrates two previous approaches: the DOA approach and the correlation approach. The criterion of the DOA approach is directions
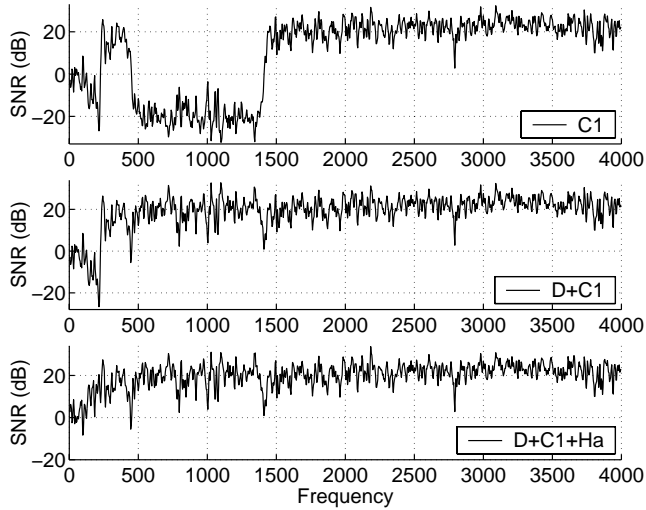
Fig. 6. SNRs at each frequency

Fig. 7. Correlations with neighboring frequencies
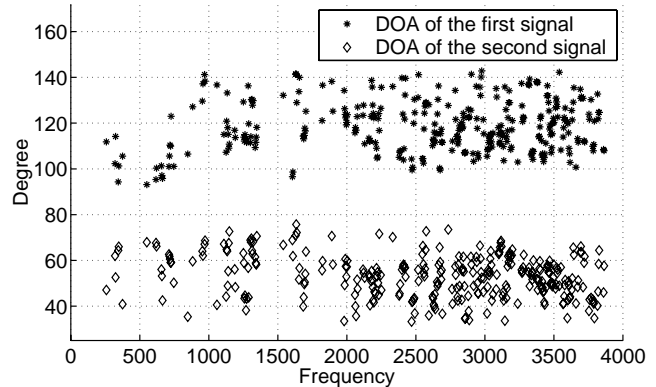
Fig. 8. DOA estimations with confidence

which are absolute. This makes the approach robust. By contrast, the criterion of the correlation approach is calculated from the separated signals themselves. This makes the approach precise. Our proposed method benefits from both advantages. In our experiments, the proposed method solved permutation problems almost perfectly under conditions whereby two sources were mixed in a room where T_R = 300 ms. Future work will focus on experiments for cases with more than two sources.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[2] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.

[3] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[4] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[5] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," in *Proc. ICASSP 2001*, May 2001, pp. 2729–2732.
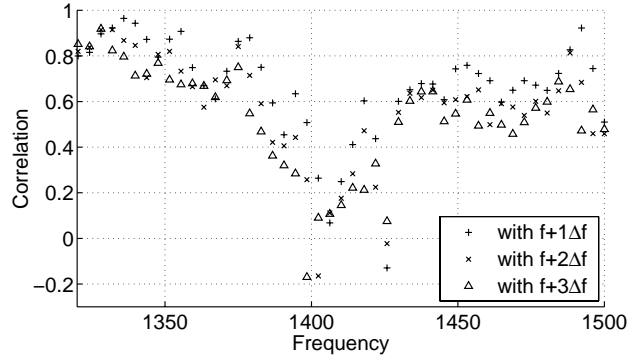
[6] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. ICASSP 2000*, June 2000, pp. 3140–3143.

[7] M. Z. Ikram and D. R. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," in *Proc. ICASSP 2002*, May 2002, pp. 881–884.

[8] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA 2000*, June 2000, pp. 215–220.

[9] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, Oct. 2001.

[10] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," in *Proc. ICASSP 2002*, May 2002, pp. 1001–1004.

[11] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, pp. 2–24, Apr. 1988.

[12] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust approach to the permutation problem of frequency-domain blind source separation," in *Proc. ICASSP 2003*, (accepted).