NRC Publications Archive Archives des publications du CNRC

An Improved Method of Permutation Correction in Convolutional Blind Source Separation

Wang, Lin; Ding, Heping; Yin, Fuliang

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.2478/v10168-010-0038-9>

Archives of Acoustics, 35, 4, pp. 493-504, 2010-12-24

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=ebb4e9bd-d733-4aa5-bb24-4dae8bfeded5>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=ebb4e9bd-d733-4aa5-bb24-4dae8bfeded5>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

P O L I S H A C A D E M Y O F S C I E N C E S
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH
COMMITTEE ON ACOUSTICS • POLISH ACOUSTICAL SOCIETY

ARCHIVES of ACOUSTICS

QUARTERLY

Vol. 35, No. 4

2010

WARSZAWA

Editorial Board

Editor: B. KOSTEK (Gdańsk University of Technology, Gdańsk)

Associate Editors

D. AUGUSTYŃSKA (Central Institute for Labour Protection, Warszawa)
A. CIARKOWSKI (Warsaw University of Life Sciences, Warszawa)
E. DANICKI (Institute of Fundamental Technological Research PAN, Warszawa)
G. GRELOWSKA (Polish Naval Academy, Gdynia)
R. GUBRYNOWICZ (Polish-Japanese Institute of Information Technology, Warszawa)
T. GUDRA (Wrocław University of Technology, Wrocław)
T. KUJAWSKA (Institute of Fundamental Technological Research PAN, Warszawa)
R. MAKAREWICZ (Adam Mickiewicz University, Poznań)
A. MIŚKIEWICZ (The Fryderyk Chopin University of Music, Warszawa)
W.P. RDZANEK (University of Rzeszów, Rzeszów)

Secretary: J. ŻYCHOWICZ–POKULNIEWICZ

Advisory Editorial Board

L. BJØRNØ (Technical University of Denmark, Lyngby)
J. BLAUERT (Ruhr University, Bochum)
D. BRADLEY (The Pennsylvania State University, Pennsylvania)
M. CROCKER (Auburn University)
G. DEMENKO (Adam Mickiewicz University, Poznań)
A. DOBRUCKI (Wrocław University of Technology, Wrocław)
S. J. ELLIOTT (University of Southampton, Southampton)
Z. ENGEL (AGH University of Science and Technology, Kraków)
C. HANSEN (University of Adelaide, Adelaide)
W. HESS (University of Bonn, Bonn)
E. KOZACZKA (Naval University of Gdynia, Gdynia)
T. G. LEIGHTON (University of Southampton, Southampton)
P. A. LEWIN (Drexel University, Philadelphia)
A. NOWICKI (Institute of Fundamental Technological Research PAN, Warszawa)
T. PUSTELNY (Silesian University of Technology, Gliwice)
A. RAKOWSKI (The Fryderyk Chopin University of Music, Warszawa)
J. SADOWSKI (Building Research Institute, Warszawa)
J. SUNDBERG (Royal Institute of Technology, Stockholm)
A. ŚLIWIŃSKI (University of Gdańsk, Gdańsk)
B. R. TITTMANN (The Pennsylvania State University, Pennsylvania)
P. TORTOLI (University of Florence, Florence)
M. VORLÄNDER (Institute of Technical Acoustics, RWTH Aachen)
Z. WESOŁOWSKI (Kielce University of Technology, Kielce)
J. J. ZWISLOCKI (Syracuse University, Syracuse)

Editorial Board Office: Pawińskiego 5B, 02-106 Warszawa (Poland)
fax (48-22) 826 98 15, phone (48-22) 826 12 81 ext. 206
e-mail: akustyka@ippt.gov.pl <http://acoustics.ippt.gov.pl>

PRINTED IN POLAND

Indexed and abstracted in Science Citation Index Expanded (also known as SciSearch®)
and Journal Citation Reports/Science Edition

Abstracted in Acoustics Abstracts, in Applied Mechanics Reviews
and indexed in Journal of the Acoustical Society of America

Recognised by the International Institute of Acoustics and Vibration, I IAV

Edition co-sponsored by Ministry of Science and Higher Education

Arkuszy wydawniczych 22,5 Arkuszy drukarskich 18,00
Skład i łamanie w systemie T_EX: Dział Wydawnictw IPPT PAN – K. JEZIEWSKA
Druk i oprawa: Drukarnia Braci Grodzickich
ul. Geodetów 47A, Piaseczno

CONTENTS

Z. Tao, H. Zhao, J. Wu, J. Gu, Y. Xu, D. Wu , A lifting wavelet domain audio watermarking algorithm based on the statistical characteristics of sub-band coefficients	481
L. Wang, H. Ding, F. Yin , An improved method of permutation correction in convolutive blind source separation	493
A. Alpkocak, M.K. Sis , Computing impulse response of room acoustics using the ray-tracing method in time domain	505
D. Augustyńska, A. Kaczmarska, W. Mikulski, J. Radosz , Assessment of teachers' exposure to noise in selected primary schools	521
W. Batko, B. Przysucha , Determination of the probability distribution of the mean sound level	543
J. Kotus, M. Szczodrak, A. Czyżewski, B. Kostek , Long-term comparative evaluation of acoustic climate in selected schools before and after acoustic treatment	551
M.R. Monazzam, M. Naderzadeh, P. Nassiri, S.M.B. Fard , Performance of environmental T-shape noise barriers covered with primitive root diffusers	565
M. Pawlaczyk–Łuszczczyńska, A. Dudarewicz, M. Zamojska, M. Śliwińska–Kowalska , Hearing ability in orchestral musicians	579
E. Walerian, R. Janczur, M. Czechowicz, Y. Smyrnova , Possible improvement of acoustical climate. Part II: Possible solutions	595
A. Perelomova, P. Wojda , Studies of nonlinear sound dynamics in fluids based on the caloric equation of state	619
P. Karwat, Z. Klimonda, M. Sęklewski, M. Lewandowski, A. Nowicki , Data reduction method for synthetic transmit aperture algorithm	635
Z. Klimonda, J. Litniewski, A. Nowicki , Tissue attenuation estimation from backscattered ultrasound using spatial compounding technique – preliminary results	643
W. Secomski, A. Nowicki, J. Wójcik, M. Lewandowski, M. Walczak, R. Tymkiewicz , Annular array transducer and matched amplifier for therapeutic ultrasound	653
Y. Tasinkevych, E. Danicki , Full-wave analysis of periodic baffle system in beam-forming applications	661
Y. Tasinkevych , Wave generation by a finite baffle array in application to beam-forming analysis	677
I. Trots, A. Nowicki, M. Lewandowski, Y. Tasinkevych , Multi-element synthetic transmit aperture in medical ultrasound imaging	687
J. Wójcik, J. Litniewski, A. Nowicki , Example of structure modeling and analysis of ultrasound scattering for trabecular bone	701
U. Kaatze , Acoustical spectroscopy of carbohydrate aqueous solutions: saccharides; alkyl glycosides; cyclodextrins. Part I. Conformer variations	715
U. Kaatze , Acoustical spectroscopy of carbohydrate aqueous solutions: saccharides; alkyl glycosides; cyclodextrins. Part II. Association and Complexation	739

An Improved Method of Permutation Correction in Convolutive Blind Source Separation

Lin WANG^{(1),(2)}, Heping DING⁽²⁾, Fuliang YIN⁽¹⁾

⁽¹⁾*Dalian University of Technology*
School of Electronic and Information Engineering
Dalian, 116023, P.R. China
e-mail: wanglin_2k@sina.com
heping.ding@nrc-cnrc.gc.ca
flyin@dlut.edu.cn

⁽²⁾*Institute for Microstructural Sciences*
National Research Council Canada
1200 Montreal Road, Ottawa, Ontario, K1A 0R6, Canada

(received June 7, 2010; accepted September 1, 2010)

This paper proposes an improved method of solving the permutation problem inherent in frequency-domain of convolutive blind source separation (BSS). It combines a novel inter-frequency dependence measure: the power ratio of separated signals, and a simple but effective bin-wise permutation alignment scheme. The proposed method is easy to implement and surpasses the conventional ones. Simulations have shown that it can provide an almost ideal solution of the permutation problem for a case where two or three sources were mixed in a room with a reverberation time of 130 ms.

Keywords: blind source separation, cocktail party, convolutive mixing, frequency domain, permutation problem.

1. Introduction

A typical problem in array processing and data analysis is to recover the source signals from their mixtures. Blind source separation (BSS) deals with this problem (HYVARIEN *et al.*, 2001). In recent years, BSS has received considerable attention and became a hotspot of modern signal processing. It has many potential applications, ranging from wireless communication, speech processing, image processing, biomedicine and radar technology, even to financial data analysis. An interesting application is the separation of audio sources that have been mixed and recorded by multiple microphones in a room, i.e. the so-called cocktail party

problem. A major challenge is that the mixing process is convolutive, i.e. the observations are combinations of filtered versions of the sources. This requires many channel parameters to be estimated.

Many approaches have been proposed to do convolutive BSS (PEDERSEN *et al.*, 2007). Among them, frequency-domain approaches are considered to be promising with faster convergence and lower complexity (SMARAGDIS, 1998; SAWADA *et al.*, 2007a). In frequency-domain BSS, the observed time-domain signals are converted into the frequency-domain, e.g. using the short-time Fourier transform (STFT), and then instantaneous BSS is applied to each frequency bin, after which the separated signals of all frequency bins are combined and re-transformed to the time-domain. An issue with frequency-domain BSS is that the permutation ambiguity may become serious. Even if satisfactory separation in each frequency bin is achieved, combining of the separated frequency bins to recover the original sources is difficult because of the unknown permutations associated with individual frequency bins.

Considerable work has been done to tackle the permutation problem. A popular strategy is to exploit mutual dependence of bin-wise separated signals across the frequencies, which tends to be high if the components originate from the same source (MURATA *et al.* 2001; SAWADA, 2007b). Another strategy is based on the direction-of-arrival estimations. By estimating the arriving delays of sources or analyzing the directivity pattern formed by a separation matrix, source direction can be estimated and permutations aligned (SAWADA *et al.*, 2004; IKRAM, MORGAN, 2005). The advantage of the first strategy over the second one is that it is less affected by adverse mixing conditions such as reverberation and sources being closely located. Thus, this paper focuses on solving the permutation problem based on inter-frequency dependence. We will consider two issues: a metric to measure the inter-frequency dependence, and a scheme to align the permutation across the frequencies.

A classical method, proposed by MURATA (2001), aligns permutation based on an inter-frequency dependence measure: separated signal envelopes. The bin-wise permutation alignment scheme in (MURATA *et al.*, 2001) is straightforward. However, since the envelope dependence can only be clearly observed in a small percentage of all frequencies, the method in (MURATA *et al.*, 2001) is limited when dealing with many sources. Recently, a method proposed by SAWADA (2007b) tries to solve the permutation problem based on another measure: separated signal power ratio, with which the inter-frequency dependence is more clearly exhibited. However, the alignment scheme in (SAWADA *et al.*, 2007b) involves a K-mean clustering step, which is unsupervised and it is difficult to control it. In addition, the performance of (SAWADA *et al.*, 2007b) also degrades with increasing number of sources.

Considering the advantages of both methods mentioned above, we propose an improved solution which combines the inter-frequency dependence of power ratio in (SAWADA *et al.*, 2007b) and the permutation alignment scheme in (MURATA

et al., 2001). Comparing with them, the proposed method is more effective and easier to implement.

The rest of the paper is organized as follows. The principle of frequency-domain blind source separation is introduced in Sec. 2. The proposed permutation method is described in detail in Sec. 3. Experimental results are presented in Sec. 4. Finally, Sec. 5 concludes the paper.

2. Frequency-domain blind source separation

Suppose that N source signals $s(n) = [s_1(n), \dots, s_N(n)]^T$ are mixed and recorded by M sensors, the observed signals $x(n) = [x_1(n), \dots, x_M(n)]^T$ are given by

$$x_j(n) = \sum_{i=1}^N \sum_{p=1}^P h_{ji}(p) s_i(n-p+1), \quad (j = 1, \dots, M), \quad (1)$$

where h_{ji} is a P -point impulse response from source i to microphone j . With blind source separation, the estimated source signals $y(n) = [y_1(n), \dots, y_N(n)]^T$ are obtained:

$$y_j(n) = \sum_{i=1}^M \sum_{q=1}^Q w_{ji}(q) x_i(n-q+1), \quad (j = 1, \dots, N), \quad (2)$$

where w_{ji} is a Q -point unmixing filter.

The unmixing filters can be calculated using the frequency-domain method. Figure 1 gives the system structure of frequency-domain BSS. First, $x(n)$ is converted into a time-frequency series $X(m, f)$ by a blockwise L -point short-time Fourier transform (STFT). Thus, the convolution in (1) becomes a multiplication:

$$X(m, f) = H(f)S(m, f), \quad (3)$$

where m is the frame index, $f \in [f_0, \dots, f_{L/2}]$ is the frequency.

$$X(m, f) = [X_1(m, f), \dots, X_M(m, f)]^T$$

and

$$S(m, f) = [S_1(m, f), \dots, S_N(m, f)]^T$$

are the STFT of $x(n)$ and $s(n)$, respectively; $H(f)$ is the $M \times N$ mixing matrix at frequency f .

Moreover, separation is performed in each frequency bin f :

$$Y(m, f) = W(f)X(m, f), \quad (4)$$

where $Y(m, f) = [Y_1(m, f), \dots, Y_N(m, f)]^T$ is the STFT of $y(n)$, $W(f)$ is the $N \times M$ unmixing matrix at f . Complex-valued instantaneous BSS algorithms, such as FastICA (BINGHAM, HYVARIEN, 2000) and Informax (BELL, SEJNOWSKI,

Fig. 1. Workflow of frequency-domain blind source separation.

1995), can be used for the calculation of W . However, even if satisfactory separation may be obtained in each frequency bin, there are inherent scaling and permutation ambiguities. This is expressed as

$$Y(m, f) = W(f)X(m, f) \approx D(f)\Pi(f)S(m, f), \quad (5)$$

where $\Pi(f)$ is a permutation matrix and $D(f)$ is a scaling matrix, all at frequency f . These ambiguities should be corrected before transformation of the frequency-domain signals back into the time domain.

In the third step, the permutation matrix $\Pi(f)$ is determined at each frequency f so that separated frequency components $Y_i(m, f)$ are grouped together for the same source. This issue will be addressed later in Sec. 3.

Next, the scaling ambiguity is corrected by using the Minimal Distortion Principle (MATSUOKA, NAKASHIMA, 2001):

$$W_s(f) = \text{diag}(W_p^{-1}(f)) \cdot W_p(f), \quad (6)$$

where $W_p(f)$ is $W(f)$ after permutation alignment and $W_s(f)$ is the one after scaling correction; $(\cdot)^{-1}$ denotes inversion of a square matrix or pseudo inversion of a rectangular matrix, and $\text{diag}(\cdot)$ retains only the main diagonal components of a matrix.

At the end of the flow, inverse STFT is applied to $W_s(f)$ to get the time-domain unmixing filters $w(n)$, and the estimated source signals $y(n)$ is obtained by the unmixing filtering in (2). The whole procedure described above is depicted in Fig. 1.

3. Permutation alignment

In this section, we describe the proposed permutation alignment method from two aspects: inter-frequency dependence measure and permutation alignment scheme.

3.1. Inter-frequency dependence measure

The inter-frequency dependence of speech sources can be exploited to align the permutations across all frequency bins. An inter-frequency dependence measure

proposed in (SAWADA *et al.*, 2007b), the separated signal power ratio, can exhibit inter-frequency dependence among all frequencies effectively. Here we give the definition of this measure.

The $M \times N$ mixing network at frequency f can be estimated from the separation network by

$$A(f) = W^{-1}(f) = [a_1(f), \dots, a_N(f)], \quad (7)$$

where $a_i(f)$ is the i -th column vector of $A(f)$. The observed signal can be decomposed by

$$X(m, f) = \sum_{i=1}^N a_i(f) Y_i(m, f), \quad (8)$$

where $Y_i(m, f)$ is the i -th component of $Y(m, f)$, i.e.

$$Y(m, f) = [Y_1(m, f), \dots, Y_N(m, f)]^T.$$

A power ratio measure is calculated to represent the activity of the i -th separated signal at frequency f . It is defined as

$$v_i^f(m) = \frac{\|a_i(f) Y_i(m, f)\|^2}{\sum_{k=1}^N \|a_k(f) Y_k(m, f)\|^2}, \quad (9)$$

where the denominator is the total power of the observed signals $X(m, f)$, and the numerator is the power of the i -th separated signal. Being in the range $[0, 1]$, (9) is close to 1 when the i -th separated signal is dominant, and close to 0 when other signals are dominant. Due to the sparseness of speech signals, power ratio can exhibit the signal activity clearly.

The correlation coefficient of signal power ratios can be used for measuring the inter-frequency dependence and aligning of the permutation. The normalized bin-wise correlation coefficient between two power ratio sequences $v_i^{f_1}(m)$ and $v_j^{f_2}(m)$ is defined as

$$\rho(v_i^{f_1}, v_j^{f_2}) = \frac{r_{ij}(f_1, f_2) - \mu_i(f_1)\mu_j(f_2)}{\sigma_i(f_1)\sigma_j(f_2)}, \quad (10)$$

where i and j denote two separated channels, f_1 and f_2 are two frequencies, $r_{ij}(f_1, f_2) = E\{v_i^{f_1} v_j^{f_2}\}$, $\mu_i(f) = E\{v_i^f\}$, $\sigma_i(f) = \sqrt{E\{(v_i^f)^2\} - \mu_i^2(f)}$ are, respectively, the correlation, mean, and standard deviation at m . Note that $E\{\cdot\}$ denotes expectation, where the time index m is omitted for clarity.

We expect the correlation coefficient $\rho(v_i^{f_1}, v_j^{f_2})$ of two sequences $v_i^{f_1}(m)$ and $v_j^{f_2}(m)$ to be high if they originate from the same source. The principle behind this is that the active time of bin-wise separated signals are likely to coincide among frequencies for the same source. This property will be used for aligning the permutation.

3.2. Proposed permutation alignment scheme

The permutation alignment scheme employed in (MURATA *et al.*, 2001) was developed especially for the dependence measure of signal envelope. We made some modifications to it to come up with a permutation alignment scheme based on the signal power ratio measure. It is described in the 5 steps presented below.

Step 1. Calculate the power ratio $v_i^f(m)$ for all $L/2 + 1$ frequency bins and all N separated signals by (9).

Step 2. Re-arrange the frequencies: $\{f_0, \dots, f_{L/2}\} \rightarrow \{g_0, \dots, g_{L/2}\}$, in ascending order of similarity between individual components, which is defined by

$$\text{sim}(g) = \sum_{i,j (i \neq j)} \rho(v_i^g, v_j^g); \quad (11)$$

therefore,

$$\text{sim}(g_0) \leq \text{sim}(g_1) \leq \dots \leq \text{sim}(g_{L/2}). \quad (12)$$

It is noticed that $\text{sim}(g) = \rho(v_1^g, v_2^g) = 0$ holds for all frequencies when there are only two sources. In this case, we use the envelope measure $v_i^g = |Y_i(g)|$ instead of (9) to calculate the similarity.

Step 3. For g_0 , keep the permutation as it is, and set $k=1$.

Step 4. For g_k , find a permutation Π_g that maximizes the correlation between the power ratio of g_k and the aggregated power ratio sequence from g_0 through g_{k-1} . This is achieved by maximizing the sum of correlation coefficients

$$\Pi_g \leftarrow \arg \max_{\Pi} \sum_{l=1}^N \rho(v_i^g, c_l) \big|_{i=\Pi_g(l), i'=\Pi_c(l)} \quad (13)$$

and c is the centroid of the power ratio sequence from g_0 through g_{k-1}

$$c_l(m) = \frac{1}{k} \sum_{g=g_0}^{g_{k-1}} v_i^g(m) \big|_{i=\Pi_g(l)}, \quad l = 1, \dots, N. \quad (14)$$

Step 5. Set $k = k+1$ and go to step 4 until $k = L/2$.

It is believed that the more different are the independent components at one frequency, the easier it is to get a correct permutation result. First, the method in (MURATA *et al.*, 2001) sorts frequency bins in increasing order of similarity among independent components; then align the permutation of the sorted frequency bins one by one. However, the method assumes high correlations even between frequencies that are far apart; this assumption is not always correct in all frequencies, especially with an envelope measure. With the power ratio measure, the inter-frequency becomes clearer and the dependence assumption may be satisfied in most frequencies. Thus the proposed method performs better than the one in (MURATA *et al.*, 2001). Furthermore, as it can be seen from the steps above, the permutation alignment scheme is rather straightforward and easy to implement.

4. Experiment results

The performance of the proposed method has been evaluated under light and medium reverberant conditions. Generally, the number of microphones should be larger than or equal to that of the sources, to ensure a complete solution (JOHO *et al.*, 2000). Here we only consider identical number of sources and microphones for convenience and simplicity. Data were obtained by simulated impulse responses of a rectangular room, based on the image model method (ALLEN, BERKLEY, 1979). The simulated environment is shown in Fig. 2. All sources and microphones are placed 1.5 m high. The reverberation time was controlled by varying the absorption coefficient of the wall. To generate the microphone signals, we used 8-seconds long speech signals sampled at 8kHz, and they were convolved with the impulse responses. The proposed method was compared with the methods in (MURATA *et al.*, 2001) and (SAWADA *et al.*, 2007b). (For clarity, we call them the Murata method and the Sawada method, respectively). In addition, results of a “Benchmark” method are also included, which corrects permutation ambiguities by using the known mixing filters; therefore, it represents an ideal method (IKRAM, MORGAN, 2000).

Fig. 2. Simulated room environment.

The performance is measured by signal-to-interference (SIR) ratios in dB. The input and output SIRs for the J -th channel are defined as, respectively,

$$\text{SIRIN}_J = 10 \log_{10} \frac{\sum_n |\sum_l h_{JJ}(l) s_J(n-l)|^2}{\sum_{k \neq J} \sum_n |\sum_l h_{Jk}(l) s_k(n-l)|^2}, \quad (15)$$

$$\text{SIROUT}_J = 10 \log_{10} \frac{\sum_n |\sum_l g_{Jp(J)}(l) s_{p(J)}(n-l)|^2}{\sum_{k \neq p(J)} \sum_n |\sum_l g_{Jk}(l) s_k(n-l)|^2}, \quad (16)$$

where n is the time index, $J = 1, \dots, N$, and $p(J)$ is the index of the output where the J -th source appears, $h_{Jk}(n)$ is an element of $H(n)$ (see (1)), and $g_{Jk}(n)$ is an element of the overall impulse response matrix $G(n) = W(n) * H(n)$.

The Tukey window is used in short time Fourier transform, with the STFT frame size of 2048 and a shift size of 512. The instantaneous BSS is implemented by means of the Scaled Informax (DOUGLAS, GUPTA, 2007), which can converge to the optimal solution within 100 iterations. In this paper, we set the iteration number as 100. The scaling ambiguity is solved by using Minimum Distortion Principle (6). The smoothing method proposed in (SAWADA *et al.*, 2003) is applied in order to reduce spikes due to the circularity effect of the FFT. The processing bandwidth is between 100 and 3750 Hz (sampling rate being 8 kHz).

4.1. Performance evaluation in light reverberation

The proposed method is applied in a number of conditions. The reverberation time $RT_{60} = 130$ ms. Various 2×2 (2 sources (1F, 1M)⁽¹⁾ and 2 microphones (B, C)), 3×3 (3 sources (1F, 2M) and 3 microphones (A, B, C)) and 4×4 (4 sources (2F, 2M) and 4 microphones) simulation cases are carried out. Different combinations of source locations are tested for each case. The average input SIR is about 0 dB for 2×2 , -2 dB for 3×3 , and -5 dB for 4×4 cases. The separation results with the four methods for the three cases are presented in Figs. 3, 4, and 5, respectively, where the horizontal axis is the source location and the vertical axis is the average output SIR.

Fig. 3. Separation results for 2 speakers and 2 microphones ($RT_{60} = 130$ ms).

⁽¹⁾ Here 'F' means female and 'M' means male.

Fig. 4. Separation results for 3 speakers and 3 microphones ($RT_{60} = 130$ ms).

Fig. 5. Separation results for 4 speakers and 4 microphones ($RT_{60} = 130$ ms).

Compared to the benchmark, all the three “non-ideal” methods show nearly ideal separation in 2×2 cases. In 3×3 cases, the Sawada method performs better than the Murata method, and the proposed method performs even better – with an almost ideal separation. In 4×4 cases, the performance of the three methods all degrades evidently; however, the proposed one is the closest to the benchmark. In a word, combining the power ratio measure and the Murata permutation alignment scheme, the proposed method is superior to the other two under various simulation conditions.

The separation result depends mainly on two factors: instantaneous BSS and permutation alignment. It is easier to separate two sources when they are far apart and hence have different transfer functions to sensors. For example, better separation is observed for the 2×2 cases '1, 9', '2, 7' than for other ones. Similarly, although sources '3, 4' are placed on one straight line, they have different transfer functions to the sensors B and C, thus it is possible to get better separation results than for closely spaced sources '5, 7'.

4.2. Performance evaluation in medium reverberation

The proposed method is evaluated in medium reverberation with $RT_{60} = 300$ ms. The room layout is identical to the one in the previous experiment. We separate 3×3 mixtures with the four methods respectively: the proposed method, the Murata method, the Sawada method, and the Benchmark method. The separation results are depicted in Fig. 6. Comparing Fig. 6 with Fig. 4, it can be seen that the performance of the four methods all degrade evidently due to the longer mixing filters in medium reverberation. Again, the proposed method shows better performance than the Sawada method and the Murata method. In most cases, its performance is close to the ideal result.

Fig. 6. Separation results for 3 speakers and 3 microphones ($RT_{60} = 300$ ms).

5. Conclusion

Studying the frequency-domain convolutive blind source separation, this paper proposes a new permutation alignment method which employs an inter-frequency measure: the power ratio of separated signals, and the Murata align-

ment scheme. The power ratio measure can exploit the inter-frequency dependence more clearly than the conventional metrics; the permutation alignment scheme is simple but effective. Thus, the proposed method performs better than other ones evaluated. Besides, it is easy to implement. Experimental results showed the effectiveness of the proposed method.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (60772161, 60372082) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (200801410015). This work is also supported by NRC-MOE Research and Post-doctoral Fellowship Program from Ministry of Education of China and National Research Council of Canada.

References

1. ALLEN J.B., BERKLEY D.A. (1979), *Image method for efficiently simulating small room acoustics*, Journal of the Acoustical Society of America, **65**, 943–950.
2. BELL A.J., SEJNOWSKI T.J. (1995), *An information maximization approach to blind separation and blind deconvolution*, Neural Computation, **7**, 6, 1129–1159.
3. BINGHAM E., HYVARIEN A. (2000), *A fast fixed-point algorithm for independent component analysis of complex valued signals*, International Journal of Neural Systems, **10**, 1, 1–8.
4. DOUGLAS S.C., GUPTA M. (2007), *Scaled natural gradient algorithms for instantaneous and convolutional blind source separation*, 2007 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 637–640, Honolulu, USA.
5. HYVARIEN A., KARHUNEN J., OJA E. (2001), *Independent Component Analysis*, John Wiley & Sons, New York.
6. IKRAM M.Z., MORGAN D.R. (2000), *Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment*, 2000 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1041–1044, Istanbul, Turkey.
7. IKRAM M.Z., MORGAN D.R. (2005), *Permutation inconsistency in blind speech separation: investigation and solutions*, IEEE Transactions on Speech and Audio Processing, **13**, 1, 1–13.
8. JOHO M., MATHIS H., LAMBERT R.H. (2000), *Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture*, Independent Component Analysis and Blind Signal Separation ICA 2000, pp. 81–86, Helsinki, Finland.
9. MATSUOKA K., NAKASHIMA S. (2001), *Minimal distortion principle for blind source separation*, 2001 International Workshop on Independent Component, pp. 722–727.
10. MURATA N., IKEDA S., ZIEHE A. (2001), *An approach to blind source separation based on temporal structure of speech signals*, Neurocomputing, **41**, 1–4, 1–24.

11. PEDERSEN M.S., LARSEN J., KJEMS U., PARRA L.C. (2007), *A survey of convolutive blind source separation methods*, [in:] Springer handbook on Speech Processing and Speech Communication, 1–34, Springer.
12. SAWADA H., MUKAI R., KETHULLE S., ARAKI S., MAKINO S. (2003), *Spectral smoothing for frequency-domain blind source separation*, 2003 International Workshop on Acoustic Echo and Noise Control, pp. 311–314, Kyoto, Japan.
13. SAWADA H., MUKAI R., ARAKI S., MAKINO S. (2004), *A robust and precise method for solving the permutation problem of frequency-domain blind source separation*, IEEE Transactions on Speech and Audio Processing, **12**, 5, 530–538.
14. SAWADA H., ARAKI S., MAKINO S. (2007a), *Frequency-domain blind source separation*, [in:] Blind Speech Separation, 47–78, Springer.
15. SAWADA H., ARAKI S., MAKINO S. (2007b), *Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS*, 2007 IEEE International Symposium on Circuits and Systems, pp. 3247–3250, New Orleans, USA.
16. SMARAGDIS P. (1998), *Blind separation of convolved mixtures in the frequency domain*, Neurocomputing, **22**, 1–3, 21–34.

INFORMATION FOR AUTHORS

ARCHIVES OF ACOUSTICS the journal of Polish Academy of Sciences (a quarterly founded in 1976), publishes original papers concerning all fields of acoustics. Basically, the papers should not exceed 40 000 typographic signs (about 20 typewritten pages).

The submitted papers must be written in good English.

Follow this order when typing manuscripts: Title, Authors, their affiliations and e-mails, Abstract, Keywords, Main text, Acknowledgments, Appendix, References, Figure Captions and then Tables.

Special care must be attached to the following rules:

1. Text should be submitted with double-lined spacing and preferably in a 12 point font size. The text should be in single-column format.
2. The title of the paper should be as short as possible.
3. Full names and surnames should be given.
4. The full postal address of each affiliation, including the country name should be provided. Affiliations should contain the full postal address, as well as an e-mail address of one author designated as corresponding author.
5. The text should be preceded by a concise abstract (less than 200 words).
6. Keywords should be given. It is also desirable to include a list of notations.
7. The formulae to be numbered are those referred to in the paper, as well as the final formulae. The formula number should be written on the right-hand side of the formula in round brackets.
8. All notations should be written very distinctly.
9. Vectors should be denoted by semi-bold type. Trigonometric functions are denoted by sin, cos, tan and cot, inverse trigonometric functions – by arcsin, arccos, arctan and arccot; hyperbolic functions are denoted by sinh, cosh, tanh and coth.
10. References in the text (author(s) and year of publication) are to be cited between parentheses.
11. All cited works within the text must be included in the References Section, and arranged in alphabetical order by authors' last name. Number them from 1 to n.
12. Acknowledgments should be included in a section before the references.
13. Items appearing in the reference list should include the surname and the initials of the first name of the author, the full title of the paper/book in English followed by the information on the original paper language.
In case of a book, the publisher's name, the place and year of publication should be given.
In case of a periodical, the full title of the periodical, consecutive volume number, current issue number, pages, and year of publication should be given; the annual volume number must be semi-bold so as to distinguish it from the current issue number.
14. For more information on references see <http://acoustics.ippt.gov.pl/Instructions.pdf>

Upon receipt of the paper, the Editorial Office forwards it to the reviewers. Their opinions are the basis for the Editorial Committee to determine whether the paper can be accepted for publication or not.

The Word or $\text{\LaTeX}$ source and pdf files of the final version of articles will be requested after acceptance for publication. High-quality illustrations (format bitmap (gif, jpg) or PostScript, with resolution no lower than 300 dpi) should be sent along with the manuscript.

After publication the number of Archives of Acoustics and CD-ROM containing PDF version of article will be sent to authors.

EDITORIAL BOARD
ARCHIVES OF ACOUSTICS

6. Dissertations in congresses and meetings (unpublished)

Example:

- a. Hardman J.P. (1999), *Rainer Werner Fassbinder's BRD trilogy: a manifesto for social and political reform*, Unpublished undergraduate dissertation, University of Portsmouth, Portsmouth.
- b. Ravazzola C. (1993), *La perspectiva del género en psicoterapia*, Dissertation presented in a meeting in the School of Psychology, Catholic University of Chile, Santiago, Chile.

7. Electronic sources

Author, initials (year), Title, Retrieved month, day, year, from Internet address.

If no date is shown on the document, authors should use n.d. (no date)

If the author is not given, authors should begin the reference with the title of the document.

If a document is part of a large site such as a university or a government department's website, give the name of the parent organization and the relevant department before the web address.

Examples:

- a. Banks I. (n.d.), *The NHS Direct healthcare guide*, Retrieved August 29, 2001, from www.healthcareguide.nhsdirect.nhs.uk.
- b. Deciding your future (2000), Retrieved September 5th, 2001, from University of Portsmouth, Careers Service: www.port.ac.uk/departments/careers/plancareer/deciding-your-future.htm
- c. Alexander J., Tate M.A. (2001), *Evaluating web resources*, Retrieved August 21st, 2001, from Widener University, Wolfgram Memorial Library: www2.widener.edu/Wolfgram-Memorial-Library/webevaluation/webeval.htm

Authors should not write the web address (URL) within the text of the paper, it should appear only in the reference list/bibliography. To cite this source within the text they should use the author's name (if the reference has one) or the first few words of the website title.

8. Audiovisual sources: music

Author, initials (Date of copyright), Title of the song. On Title of the album [medium of recording], Location: Label (Recording date if different from copyright date)

Example:

- a. Puccini G. (1990), *Nessun dorma*, On Carreras Domingo Pavarotti in concert [CD], Decca, London.

Subscription

Subscription orders for "Archives of Acoustics" may be sent to:

Editorial Office "Archives of Acoustics"
Pawińskiego 5B
02-106 Warszawa, Poland
phone: (48-22) 826 12 81 ext. 206
fax (48-22) 826 98 15
e-mail: akustyka@ippt.gov.pl

The journals also are available through:

Foreign Trade Enterprise
ARS POLONA
Obrońców 25
03-933 Warszawa, Poland
tel: (48-22) 509-86-37, 509-86-38

and

RUCH S.A.
Jana Kazimierza 31/33
01-248 Warszawa, Poland
tel.: (48-22) 532 88 16, 532 88 23
fax: (48-22) 532 87 31

Information about subscription are available in <http://acoustics.ippt.gov.pl>