A MULTIMODAL APPROACH FOR FREQUENCY DOMAIN INDEPENDENT COMPONENT ANALYSIS WITH GEOMETRICALLY-BASED INITIALIZATION

S. M. Naqvi†, Y. Zhang†, T. Tsalaile†, S. Sanei‡ and J. A. Chambers†

†Advanced Signal Processing Group, Department of Electronic and Electrical Engineering
Loughborough University, Loughborough LE11 3TU, UK.
‡Centre of Digital Signal Processing, Cardiff University Cardiff, CF24 3AA, UK.
Email: {s.m.r.naqvi, y.zhang5, t.k.tsalaile, j.a.chambers}@lboro.ac.uk, saneis@cf.ac.uk

ABSTRACT

A novel multimodal approach for independent component analysis (ICA) of complex valued frequency domain signals is presented which utilizes video information to provide geometrical description of both the speakers and the microphones. This geometric information, the visual aspect, is incorporated into the initialization of the complex ICA algorithm for each frequency bin, as such, the method is multimodal since two signal modalities, speech and video, are exploited. The separation results show a significant improvement over traditional frequency domain convolutive blind source separation (BSS) systems. Importantly, the inherent permutation problem in the frequency domain BSS (complex valued signals) with the improvement in the rate of convergence, for static sources, is shown to be solved by simulation results at the level of each frequency bin. We also highlight that certain fixed point algorithms proposed by Hyvärinen et. al., or their constrained versions, are not valid for complex valued signals.

1. INTRODUCTION

Convolutive blind source separation (CBSS) performed in the frequency domain, where the separation of complex valued signals is encountered, has remained as a subject of much research interest due to its potential wide applications for example in acoustic source separation, and the associated challenging technical problems, most important of which is perhaps the permutation problem. Generally, the main objective of BSS is to decompose the measurement signals into their constituent independent components as an estimation of the true sources which are assumed a priori to be independent [1] [2].

CBSS algorithms have been conventionally developed in either the time [3] or frequency [1] [4] [5] [6] domains. Frequency domain convolutive blind source separation (FDCBSS) has however been a more popular approach as the time-domain convolutive mixing is converted into a number of independent complex instantaneous mixing operations. The permutation problem inherent to FDCBSS presents itself when reconstructing the separated sources from the separated outputs of these instantaneous mixtures. It is more severe and destructive than for time-domain schemes as the number of permutations grows geometrically with the number of instantaneous mixtures [7]. In unimodal BSS no priori assumptions are typically made on the source statistics or the mixing system. On the other hand, in a multimodal approach a video system can capture the approximate positions of the speakers and the directions they face [8]. Such video information can thereby help to estimate the unmixing matrices more accurately and ultimately increase the separation performance. Following this idea, the objective of this paper is to use efficiently such information to mitigate the permutation problem. The scaling problem in CBSS is easily solved by matrix normalization [9] [10]. The convolutive mixing system can be described as follows: assume m statistically independent sources as $\mathbf{s}(t) = [s_1(t), \ldots, s_m(t)]^T$ where $[.]^T$ denotes the transpose operation and t the discrete time index. The sources are convolved with a linear model of the physical medium (mixing matrix) which can be represented in the form of a multichannel FIR filter $\mathbf{H}$ with memory length p to produce n sensor signals $\mathbf{x}(t) = [x_1(t), \ldots, x_n(t)]^T$ as

$$\mathbf{x}(t) = \sum_{\tau=0}^{P} \mathbf{H}(\tau)\mathbf{s}(t-\tau) + \mathbf{v}(t) \tag{1}$$

where $\mathbf{v}(t) = [v_1(t), \ldots, v_n(t)]^T$ and $\mathbf{H} = [\mathbf{H}(0), \ldots, \mathbf{H}(P)]$. In common with other researchers we assume $n \geq m$. Using time domain CBSS, the sources can be estimated using a set of unmixing filter matrices $\mathbf{W}(\tau), \tau = 0, .., Q$, such that

$$\mathbf{y}(t) = \sum_{\tau=0}^{Q} \mathbf{W}(\tau)\mathbf{x}(t-\tau) \tag{2}$$

where $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^T$ contains the estimated sources, and Q is the memory of the unmixing filters. In FDBSS the problem is transferred into the frequency domain using the short time frequency transform STFT. Equations (1) and (2) then change respectively to:

$$\mathbf{x}(\omega, t) \approx \mathbf{H}(\omega)\mathbf{s}(\omega, t) + \mathbf{v}(\omega, t) \tag{3}$$

$$\mathbf{y}(\omega, t) \approx \mathbf{W}(\omega)\mathbf{x}(\omega, t) \tag{4}$$

where ω denotes discrete normalized frequency. An inverse STFT is then used to find the estimated sources $\hat{\mathbf{s}}(t) = \mathbf{y}(t)$; however, this will be certainly affected by the permutation effect due to the variation of $\mathbf{W}(\omega_i)$ with frequency bin ω_i. In the following section we present a fast fixed-point algorithm for ICA of these complex valued signals, carefully motivate the choice of contrast function, and mention the local consistency of the algorithm. In Sec. 3 we examine the use of spatial information indicating the positions and directions of the sources using "data" acquired by a number of video cameras. In Sec. 4 we use this geometric information to initialize the fixed-point frequency domain ICA algorithm. In Sec. 5 the simulation results for real world data confirm the usefulness of the algorithm. Finally, conclusions are drawn.

2. A FAST FIXED-POINT ALGORITHM FOR ICA

Recently, ICA has become one of the central tools for BSS [1], [2]. In ICA, a set of original source signals $\mathbf{s}(t)$ in (1) are retrieved from their mixtures based on the assumption of their mutual statistical independence. Hyvärinen and Oja [2] [11] presented a fast fixed point algorithm (FastICA) for the separation of linearly mixed independent source signals. Unfortunately, these algorithms are not suitable for complex valued signals. The use of algorithm [12] in this paper is due to four main reasons, its suitability for complex

Work supported by the Engineering and Physical Sciences Research Council (EPSRC) of the UK.

signals, the proof of the local consistency of the estimator, more robustness against outliers and capability of deflationary separation of the independent component signals. In deflationary separation the components tend to separate in the order of decreasing non-Gaussianity. In [12] the basic concept of complex random variables is also provided and the fixed point algorithm for one unit is derived, and for ease of derivation the algorithm updates the real and imaginary parts of $\mathbf{w}$ separately. Note for convenience explicit use of the discrete time index is dropped and $\mathbf{w}$ represents one row of $\mathbf{W}$ used to extract a single source. Since the source signals are assumed zero mean, unit variance and with uncorrelated real and imaginary parts of equal variances, the optima of $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$ under the constraint $E\{|\mathbf{w}^H\mathbf{x}|^2\} = \|\mathbf{w}\|^2 = 1$, where $E\{.\}$ denotes the statistical expectation, $(.)^H$ Hermitian transpose, $\|.\|$ Euclidian norm, $|.|$ absolute function; and $G(.)$ is a nonlinear contrast function, according to the Khun-Tucker conditions satisfy

$$\nabla E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta \nabla E\{|\mathbf{w}^H\mathbf{x}|^2\} = 0 \tag{5}$$

where the gradient denoted by ∇, is computed with respect to the real and imaginary parts of ω separately. The Newton method is used to solve this equation for which the total approximative Jacobian [12] is

$$J = 2(E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 \acute{g}(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta)I \tag{6}$$

which is diagonal and therefore easily invertible, where I denotes the identity matrix and $g(.)$ and $\acute{g}(.)$ denote the first and second derivative of the contrast function. Bingham and Hyv*ä*rinen obtained the following approximative Newton iteration:

$$\begin{aligned} \mathbf{w}^+ &= \mathbf{w} - \frac{E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\}}{E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 \acute{g}(|\mathbf{w}^H\mathbf{x}|^2)\} - \beta} \\ \mathbf{w} &= \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|} \end{aligned} \tag{7}$$

where $(.)^*$ denotes the complex conjugate. In the experiments the statistical expectation is realized as a sample average.

2.1 Robustness of Contrast Function

A good contrast function is one for which the estimator given by the contrast function is more robust to outliers in the sample values. The function used in our experiments is $G(y) = \log(b+y)$ and its derivative is $g(y) = 1/(b+y)$, where b is an arbitrary positive constant, empirically $b \approx 0.1$ is a reasonable value. The robustness of the estimator is captured in the slow growth of G, as its argument increases [12].

2.2 Local Consistency

It has been shown in [12] that the earlier results for real signals and the exact conditions for convergence of algorithms of the form of (7) to locally consistent separating solutions naturally extend to complex valued random variables. These results substantiate our choice of (7) for frequency domain source separation.

3. THE GEOMETRICAL MODEL

Given the position of the speakers and the microphones, the distances between the ith microphone and the jth speaker d_{ij}, and also their propagation times τ_{ij}, can be calculated (see Figure 1 for a simple two-speaker two-microphone case). Accordingly, in a homogenous medium such as air, the attenuation of the received speech signals is related to the distances via

$$\alpha_{ij} = \frac{\kappa}{d_{ij}^2} \tag{8}$$

where κ is a constant representing the attenuation per unit length in a homogenous medium. Similarly, τ_{ij} in terms of the number of samples, is proportional to the sampling frequency f_s, sound velocity C, and the distance d_{ij} as:

$$\tau_{ij} = \frac{f_s}{C} d_{ij} \tag{9}$$

which is independent of the directionality. However, in practical situations the speaker's direction introduces another variable into the attenuation measurement. In the case of electronic loudspeakers (not humans) the directionality pattern depends on the type of loadspeaker. Here, we approximate this pattern as $\cos(\theta_{ij}/r)$ where $r > 2$, which has a smaller value for highly directional speakers and vice versa (an accurate profile can be easily measured using a sound pressure level (SPL) meter). Therefore, the attenuation parameters become

$$\alpha_{ij} = \frac{\kappa}{d_{ij}^2} \cos(\theta_{ij}/r) \tag{10}$$

If, for simplicity, only the direct path is considered the mixing filter has the form:

$$\hat{H}(t) = \begin{bmatrix} \alpha_{11}\delta(t-\tau_{11}) & \alpha_{12}\delta(t-\tau_{12}) \\ \alpha_{21}\delta(t-\tau_{21}) & \alpha_{22}\delta(t-\tau_{22}) \end{bmatrix} \tag{11}$$

where $\hat{}$ denotes the approximation in this assumption. In the

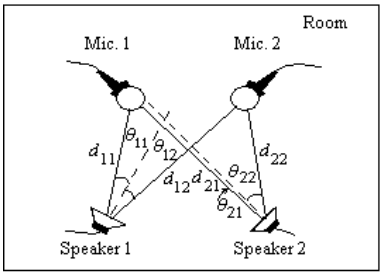

Fig. 1. A two-speaker two-microphone setup for recording within a reverberant (room) environment; only distances and angles between sources and microphones are shown.

frequency domain the above filter has the form

$$\hat{\mathbf{H}}(\omega) = \begin{bmatrix} \alpha_{11}e^{-j\omega\tau_{11}} & \alpha_{12}e^{-j\omega\tau_{12}} \\ \alpha_{21}e^{-j\omega\tau_{21}} & \alpha_{22}e^{-j\omega\tau_{22}} \end{bmatrix} \tag{12}$$

Although the actual mixing matrix includes the reverberation terms related to the reflection of sounds by the obstacles and walls, in such a room environment it will always contain the direct path components as in the above equations. Therefore, we can consider $\hat{\mathbf{H}}(\omega)$ as a crude biased estimate of the frequency domain mixing filter matrix, but one which provides the learning algorithm with a good initialization whilst importantly avoiding the bias in learning when used as a constraint within the FDBSS algorithm as in [9].

4. PROPOSED GEOMETRICALLY-BASED INITIALIZATION ICA ALGORITHM

With the help of the estimate $\hat{\mathbf{H}}(\omega)$, as an initialization of the algorithm in [12], we improve the convergence of the algorithm and also increase the separation performance together with mitigate the permutation problem. Crucially, in the proposed FDCBSS approach, since the algorithm essentially fixes the permutation at each frequency bin, there will be no problem while aligning the estimated sources for reconstruction in the time domain.

As an initial step, it is usual in ICA approaches to sphere or whiten the data i.e. the fist $\mathbf{z}(\omega) = \mathbf{Q}(\omega)\mathbf{x}(\omega)$, where $\mathbf{Q}(\omega)$ is the whitening matrix [2].

Next the position and direction information obtained from the video cameras equipped with a speaker tracking algorithm is automatically passed to (9) and (10) to estimate the $\hat{\mathbf{H}}(\omega)$ and then the first column of $\hat{\mathbf{H}}(\omega)$ is used to initialize the fixed point algorithm [12] for each frequency bin.

$$\mathbf{w}_1(\omega) = \mathbf{Q}(\omega)\hat{\mathbf{h}}_1(\omega) \tag{13}$$

The equivalence between frequency domain blind source separation and frequency domain adaptive beamforming is already confirmed in [13].

Multiplying both sides of (7) by $\beta - E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 \acute{g}(|\mathbf{w}^H\mathbf{x}|^2)\}$ we have the following update equation for each frequency bin.

$$\begin{aligned}\mathbf{w}_1^+(\omega) = E\{\mathbf{z}(\omega)(\mathbf{w}_1(\omega)^H\mathbf{z}(\omega))^* g(|\mathbf{w}_1(\omega)^H\mathbf{z}(\omega)|^2)\} \\ -E\{g(|\mathbf{w}_1(\omega)^H\mathbf{z}(\omega)|^2) + |\mathbf{w}_1(\omega)^H\mathbf{z}(\omega)|^2 \\ \acute{g}(|\mathbf{w}_1(\omega)^H\mathbf{z}(\omega)|^2)\}\mathbf{w}_1(\omega)\end{aligned} \tag{14}$$

$$\mathbf{w}_1(\omega) = \frac{\mathbf{w}_1^+(\omega)}{\|\mathbf{w}_1^+(\omega)\|} \tag{15}$$

which importantly eliminates the need to calculate β.

Since we have m independent components, the other separating vectors, i.e. $\mathbf{w}_i(\omega)$, $i = 2, \cdots, m$, are calculated in a similar manner and than decorrelated in a deflationary orthogonalization scheme. The deflationary orthogonalization for the m-th separating vector [2] takes the form

$$\mathbf{w}_m(\omega) \leftarrow \mathbf{w}_m(\omega) - \sum_{j=1}^{m-1}\{\mathbf{w}_m^H(\omega)\mathbf{w}_j(\omega)\}\mathbf{w}_j(\omega) \tag{16}$$

Finally, we formulate $\mathbf{W}(\omega) = [\mathbf{w}_1(\omega), \cdots, \mathbf{w}_m(\omega)]$ after separating all vectors of each frequency bin.

Before starting the update process $\hat{\mathbf{H}}(\omega)$ is normalized once using $\hat{\mathbf{H}}(\omega) \leftarrow \hat{\mathbf{H}}(\omega)/\|\hat{\mathbf{H}}(\omega)\|_F$ where $\|.\|_F$ denotes the Frobenius norm.

The algorithm convergence depends on the estimate of $\hat{\mathbf{H}}(\omega)$, to improve accuracy. In the case of a reverberant environment, $\hat{\mathbf{H}}(\omega)$ should ideally be the sum of all echo paths, but this is not available in practice. As will be shown by later simulations, an estimate of $\hat{\mathbf{H}}(\omega)$ obtained from (12) can result in a good performance for the proposed algorithm in a moderate reverberant environment.

5. EXPERIMENTAL RESULTS

In our experiments which correspond to the environment in Figure 2, the Bingham and Hyvärinen [12] algorithm and the proposed algorithm were tested for real room recordings. The other important variables were selected as: FFT length $T = 1024$ and filter length $Q = 512$ half of T, $r = 4$, the sampling frequency for the recordings was 8KHz and the room impulse duration was 130ms. In our proposed algorithm we select $G(y) = \log(b+y)$, with $b = 0.1$.

We first use the performance index PI [1], as a function of the overall system matrix $\mathbf{G} = \mathbf{WH}$, given by

$$\begin{aligned}PI(G) = \Big[\frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{k=1}^{m}\frac{abs(G_{ik})}{max_k abs(G_{ik})} - 1\Big)\Big] \\ + \Big[\frac{1}{m}\sum_{k=1}^{m}\Big(\sum_{i=1}^{n}\frac{abs(G_{ik})}{max_i abs(G_{ik})} - 1\Big)\Big]\end{aligned} \tag{17}$$

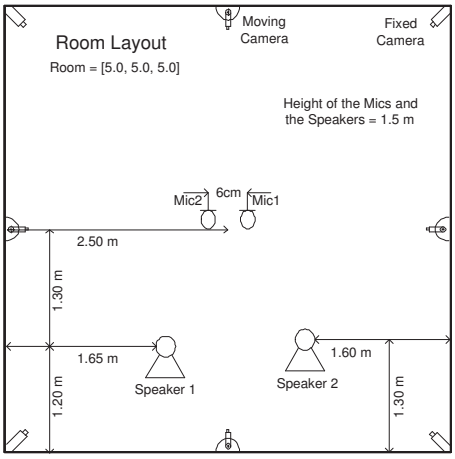

Fig. 2. A two-speaker two-microphone layout for recording within a reverberant (room) environment. Room impulse response length is 130 ms.

where G_{ik} is the ikth element of $\mathbf{G}$, to examine the performance of our algorithm at each frequency bin.

The resulting performance indices are shown in Figure 3 which shows good performance for the proposed algorithm i.e. close to zero across the majority of the frequency bins. This is due to geometrical information used in the initialization. Both algorithms were tested at fixed iteration count of seven, as our proposed algorithm has converged in this number of iterations. The visual modality therefore renders our BSS algorithm semiblind and thereby much improves the resulting performance and rate of convergence.

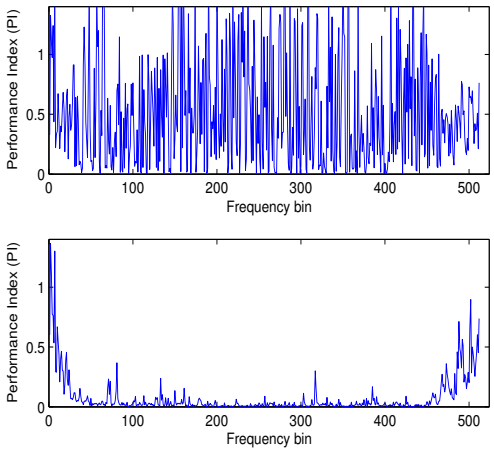

Fig. 3. Performance index at each frequency bin for the Bingham and Hyvärinen algorithm on the top [12] and the proposed algorithm at the bottom, on the recorded signals with fixed iteration count = 7. A lower PI refers to a superior method.

As mentioned in [1] PI is insensitive to permutation. We therefore introduce a criterion for the two sources case which is sensitive to permutation and shown for the real case for convenience, i.e. in the case of no permutation, $H = W = I$ or $H = W = [0,1;1,0]$ then $G = I$ and in the case of permutation if $H = [0,1;1,0]$ then $W = I$ and vice versa therefore, $G = [0,1;1,0]$. Hence for a permutation free FDCBSS $[abs(G_{11}G_{22}) - abs(G_{12}G_{21})] > 0$. We evaluated permutation on the basis of the criterion mentioned above. In

Figure 4 the results confirm that the proposed algorithm automatically mitigates the permutation at each frequency bin.

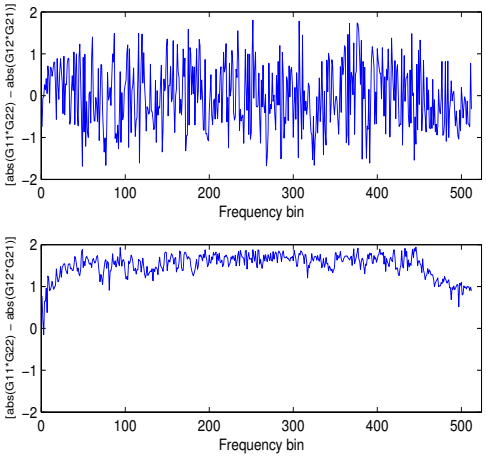

Fig. 4. Evaluation of permutation in each frequency bin for the Bingham and Hyv*ä*rinen algorithm at the top [12] and the proposed algorithm at the bottom, on the recorded signals with fixed iteration count = 7. $[abs(G_{11}G_{22}) - abs(G_{12}G_{21})] > 0$ means no permutation.

In contrast, the performance indices and evaluation of permutation by the original FastICA algorithm [12] (MATLAB code available online) with random initialization, on the recorded mixtures are shown in Figure 5. We highlight that thirty iterations are required for the performance level achieved in Figure 5(a) with no solution for permutation as shown in Figure 5(b). The permutation problem in frequency domain BSS degraded the SIR to approximately zero on the recorded mixtures.

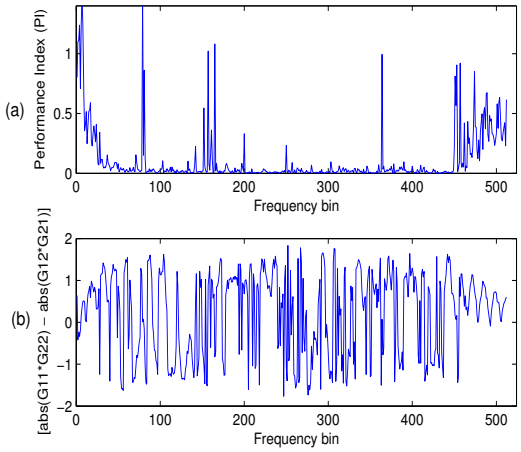

Fig. 5. (a) Performance index at each frequency bin and (b) Evaluation of permutation in each frequency bin for Bingham and Hyv*ä*rinen FastICA algorithm [12], on the recorded signals after 30 iterations. A lower PI refers to a better separation and $[abs(G_{11}G_{22}) - abs(G_{12}G_{21})] > 0$ means no permutation.

Figure 6 confirms the convergence of the underlying cost, i.e. $E\{G(|\mathbf{w}^H\mathbf{x}|^2)\}$, within seven iterations for the proposed algorithm. The results are averaged over all frequency bins. The convergence within seven iterations with solution for permutation confirm that the proposed algorithm is robust and suitable for realtime implementation.

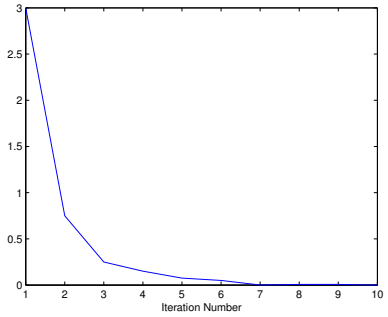

Fig. 6. The convergence graph of the cost function of the proposed algorithm using contrast function $G(y) = \log(b+y)$; the results are averaged over all frequency bins.

The proposed algorithm starts with $\mathbf{W}(\omega) = \mathbf{Q}(\omega)\hat{\mathbf{H}}(\omega)$, if the estimate of $\hat{\mathbf{H}}(\omega)$ is unbiased, then $\mathbf{W}_{opt}(\omega) = \mathbf{Q}(\omega)\hat{\mathbf{H}}(\omega)$. We assumed the estimate of $\hat{\mathbf{H}}(\omega)$ (used in above simulations obtained from (12) the directions of the sources were obtained from video cameras) is unbiased and calculated the performance shown in Figure 7, which confirms the estimate is biased, since the environment is reverberant therefore $\hat{\mathbf{H}}(\omega)$ should include the sum of all echo paths, but practically the directions of these reverberations are not possible to be measured by the video cameras. The convergence of the proposed algorithm within seven iterations including the solution for permutation, with the biased estimate of $\hat{\mathbf{H}}(\omega)$ confirm that the multimodal approach is necessary to solve the cocktail party problem.

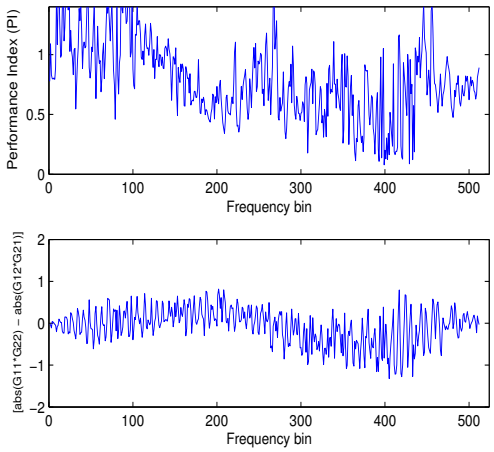

Fig. 7. (a) Performance index at each frequency bin and (b) Evaluation of permutation in each frequency, assumed $\hat{\mathbf{H}}(\omega)$ is correct i.e. $\mathbf{W}_{opt}(\omega) = \mathbf{Q}(\omega)\hat{\mathbf{H}}(\omega)$. A lower PI refers to a better separation and $[abs(G_{11}G_{22}) - abs(G_{12}G_{21})] > 0$ means no permutation.

Finally, the signal-to-interference ratio (SIR) was calculated as in [9] and results are shown in Table 1 for infomax (INFO), FDCBSS, Constrained ICA (CICAu), Para and Spence and Proposed GBFastICA algorithms, where SIR is defined as

$$SIR = \frac{\Sigma_i \Sigma_\omega |H_{ii}(\omega)|^2 \langle |s_i(\omega)|^2 \rangle}{\Sigma_i \Sigma_{i \neq j} \Sigma_\omega |H_{ij}(\omega)|^2 \langle |s_j(\omega)|^2 \rangle} \quad (18)$$

where H_{ii} and H_{ij} represents respectively, the diagonal and off-diagonal elements of the frequency domain mixing filter, and s_i is the frequency domain representation of the source of interest.

The results are summarized in Table 1 and confirm the objective improvement of our algorithm which has been confirmed subjectively by listening tests.

Table 1. Comparison of SIR-Improvement between algorithms and the proposed method for different sets of mixtures.

Algorithms	SIR-Improvement/dB
Parra's Method	6.8
FDCBSS	9.4
INFO	11.1
CICAu	11.6
GBFastICA	18.8

6. CONCLUSIONS

In this research a new multimodal approach for independent component analysis of complex valued frequency domain signals was proposed which exploits visual information to initialize a FastICA algorithm in order to mitigate the permutation problem. The advantage of our proposed algorithm was confirmed in simulations from a real room environment. The location and direction information was obtained using a number of cameras equipped with a speaker tracking algorithm. The outcome of this approach paves the way for establishing a multimodal audio-video system for separation of speech signals, with moving sources.

REFERENCES

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley, 2002.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, New York: Wiley, 2001.

[3] A. S. Bregman, *Auditory scence analysis*, MIT Press, Cambridge, MA, 1990.

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

[5] W. Wang, S. Sanei, and J.A. Chambers, "Penalty function based joint diagonalization approach for convolutive blind separation of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1654–1669, 2005.

[6] S. Makino, H. Sawada, R. Mukai, and S.Araki, "Blind separation of convolved mixtures of speech in frequency domain," *IEICE Trans. Fundamentals*, vol. E88-A, no. 7, pp. 1640–1655, Jul 2005.

[7] W. Wang, S. Sanei, and J. A. Chambers, "A joint diagonalization method for convolutive blind separation of nonstationary sources in the frequency domain," *Proc. ICA, Nara, Japan*, April 2003.

[8] W. Wang, D. Cosker, Y. Hicks, S. Sanei, and J. A. Chambers, "Video assisted speech source separation," *Proc. IEEE ICASSP*, pp. 425–428, 2005.

[9] S. Sanei, S. M. Naqvi, J. A. Chambers, and Y. Hicks, "A geometrically constrained multimodal approach for convolutive blind source separation," *Proc. IEEE ICASSP*, pp. 969–972, 2007.

[10] T. Tsalaile, S. M. Naqvi, K. Nazarpour, S. Sanei, and J. A. Chambers, "Blind source extraction of heart sound signals from lung sound recordings exploiting periodicity of the heart sound," *Proc. IEEE ICASSP, Las Vegas, USA*, 2008.

[11] A. Hyvärinen, "Fast and roubst fixed-point algorithms for independent component analysis," *IIEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.

[12] E. Bingham and A. Hyvärinen, "A fast fixed point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Networks*, vol. 10, no. 1, pp. 1–8, 2000.

[13] S.Araki, S. Makino, Y. Hinamoto, R. Mukai, T.Nishikawa, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures," *EURASIP J. Appl. Signal Process.*, , no. 11, pp. 1157–1166, 2003.