INTENSITY VECTOR DIRECTION EXPLOITATION FOR EXHAUSTIVE BLIND SOURCE SEPARATION OF CONVOLUTIVE MIXTURES

Banu Günel[*]
I-Lab/CCSR,
University of Surrey,
Guildford, GU2 7XH, UK
b.gunel@surrey.ac.uk

Hüseyin Hacıhabiboğlu [†]
CDSPR,
King's College London,
Strand, London, WC2R 2LS, UK
huseyin.hacihabiboglu@kcl.ac.uk

Ahmet M. Kondoz[*]
I-Lab/CCSR,
University of Surrey,
Guildford, GU2 7XH, UK
a.kondoz@surrey.ac.uk

ABSTRACT

This article presents a technique that uses the intensity vector direction exploitation (IVDE) method for exhaustive separation of convolutive mixtures. While only a four-element compact sensor array is used, multiple channels for all possible source directions are produced by exhaustive separation. Singular value decomposition (SVD) is then applied to determine the signal subspace and the directions of the local maxima of the signal energy. This information is then used to select the channels containing the individual sources. While the original IVDE method requires the prior knowledge of the directions of sources for separation, the present method eliminates this need and achieves fully blind separation. Performing SVD at a post-processing stage also improves the sound quality. The method has been tested for convolutive mixtures of up to four sources and typical separation performances are given.

***Index Terms*—** BSS, coincident array, intensity, deterministic, SVD

1. INTRODUCTION

Blind source separation (BSS) aims to obtain interference-free versions of simultaneously active sound sources without any information about their number and positions, or the acoustic environment itself. Stochastic techniques, such as ICA [1, 2] aim to improve the statistical independence of the signals by exploiting the properties of the signals themselves. Adaptive techniques, such as adaptive beamforming (ABF) [3, 4] optimize a multichannel filter structure according to the properties of the signals and the source geometry. ABF techniques utilize spatial selectivity to suppress the interferences and improve the capture of the target source. Another group of the BSS techniques, which can be called as deterministic, utilise solely the deterministic aspects of the problem, such as the directions of the sources or the multipath characteristics [5, 6, 7]. As these pieces of information are not available for the blind separation problem, the performance of the algorithms are limited by the accuracy of their estimations.

A deterministic technique that is based on intensity vector direction exploitation (IVDE) for acoustic source separation has previously been proposed by the authors [8]. However, the formulation of the solution required the prior knowledge of the source directions and produced as many channels as the number of sources. The technique presented in this paper is an improvement over the IVDE method, eliminating the need for source direction information, thereby achieving fully blind separation.

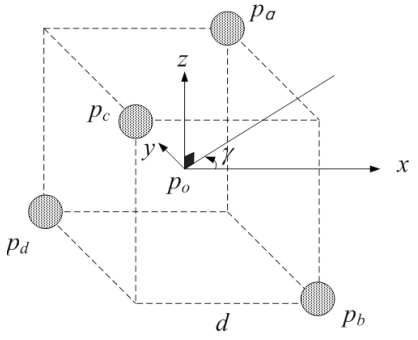

Fig. 1. Microphone array setup with four microphones positioned at the non-adjacent corners of a cube.

The proposed method performs exhaustive separation, i.e., produces multiple channels of output corresponding to all directions, while using only four input channels. The dimension of the matrix formed by the signals for each direction is reduced by singular value decomposition (SVD) after selecting the singular values that correspond to the signal subspace. The directions of the local maxima of the signal energies are selected as the target source channels.

This paper is organized as follows. In Section 2, an overview of the IVDE method is given together with the usage of compact arrays for calculating intensity vector directions. Section 3 describes the theory of the exhaustive source separation method for obtaining multiple channels. Section 4 explains dimensionality reduction and the selection of channels corresponding to the source signals together with its relevance to direction-of-arrival (DOA) estimation. Section 5 describes the experimental test conditions and provides the obtained results. Section 6 concludes the paper.

2. INTENSITY VECTOR DIRECTION CALCULATION

The compact microphone array used for intensity vector direction calculation is made up of four microphones placed at the four non-adjacent corners of a cube. This geometry forms a tetrahedral microphone array as shown in Fig. 1.

Let us consider a plane wave arriving from the direction $\gamma(\omega, t)$ on the horizontal plane with respect to the center of the cube. If the

[*]The work presented was developed within VISNET II, a European Network of Excellence, funded under the European Commission IST FP6 programme.

[†]The author is funded by the Engineering and Physical Sciences Research Council (EPSRC) Research Grant EP/F001142/1.

pressure at the center due to this plane wave is $p_o(\omega,t)$, then the pressure signals recorded by these four microphones can be written as,

$$p_a(\omega,t) = p_o(\omega,t)e^{jkd\sqrt{2}/2\cos(\pi/4-\gamma(\omega,t))}, \quad (1)$$

$$p_b(\omega,t) = p_o(\omega,t)e^{jkd\sqrt{2}/2\sin(\pi/4-\gamma(\omega,t))}, \quad (2)$$

$$p_c(\omega,t) = p_o(\omega,t)e^{-jkd\sqrt{2}/2\cos(\pi/4-\gamma(\omega,t))}, \quad (3)$$

$$p_d(\omega,t) = p_o(\omega,t)e^{-jkd\sqrt{2}/2\sin(\pi/4-\gamma(\omega,t))}, \quad (4)$$

where k is the wave number related to the wavelength λ as $k = 2\pi/\lambda$, j is the imaginary unit and d is the length of the one side of the cube. Using these four pressure signals, B-format signals [9], p_W, p_X and p_Y can be obtained as $p_W = 0.5(p_a + p_b + p_c + p_d)$, $p_X = p_a + p_b - p_c - p_d$ and $p_Y = p_a - p_b - p_c + p_d$.

If $kd \ll 1$, i.e., when the microphones are positioned close to each other in comparison to the wavelength, it can be shown by using the relations $\cos(kd\cos\gamma) \approx 1$, $\cos(kd\sin\gamma) \approx 1$, $\sin(kd\cos\gamma) \approx kd\cos\gamma$ and $\sin(kd\sin\gamma) \approx kd\sin\gamma$ that,

$$p_W(\omega,t) \simeq 2p_o(\omega,t), \quad (5)$$

$$p_X(\omega,t) \simeq j2p_o(\omega,t)kd\cos(\gamma(\omega,t)), \quad (6)$$

$$p_Y(\omega,t) \simeq j2p_o(\omega,t)kd\sin(\gamma(\omega,t)). \quad (7)$$

The acoustic particle velocity, $\mathbf{v}(\mathbf{r},w,t)$ in two dimensions is defined as [10],

$$\mathbf{v}(\mathbf{r},\omega,t) = \frac{1}{\rho_0 c}\left[p_X(\omega,t)\,\mathbf{u_x} + p_Y(\omega,t)\,\mathbf{u_y}\right], \quad (8)$$

where ρ_0 is the ambient density, c is the speed of sound, $\mathbf{u_x}$ and $\mathbf{u_y}$ are unit vectors in the directions of corresponding axes.

The product of the pressure and the particle velocity gives instantaneous intensity:

$$\mathbf{I}(\omega,t) = \frac{1}{\rho_0 c}\left[Re\{p_W^*(\omega,t)p_X(\omega,t)\}\mathbf{u_x} + \right. \quad (9)$$

$$\left. Re\{p_W^*(\omega,t)p_Y(\omega,t)\}\mathbf{u_y}\right], \quad (10)$$

where $*$ denotes conjugation and $Re\{\bullet\}$ denotes the real part of the argument.

Then, the direction of the intensity vector, $\gamma(\omega,t)$ can be obtained by

$$\gamma(\omega,t) = \arctan\left[\frac{Re\{p_W^*(\omega,t)p_Y(\omega,t)\}}{Re\{p_W^*(\omega,t)p_X(\omega,t)\}}\right]. \quad (11)$$

Since the microphones in the array are closely spaced, plane wave assumption can safely be made for incident waves and their directions can be calculated. If simultaneously active sound signals do not overlap in short time-frequency windows, the directions of the intensity vectors correspond to those of the sound sources randomly shifted by major reflections. The von Mises and von Mises mixture distributions, which are the equivalents of Gaussian and Gaussian mixture distributions in circular statistics, have been used before for modelling the distributions of the intensity vector directions for single and multiple sound sources, respectively [8]. The next section explains exhaustive separation by decomposing the sound field into plane waves using intensity vector directions.

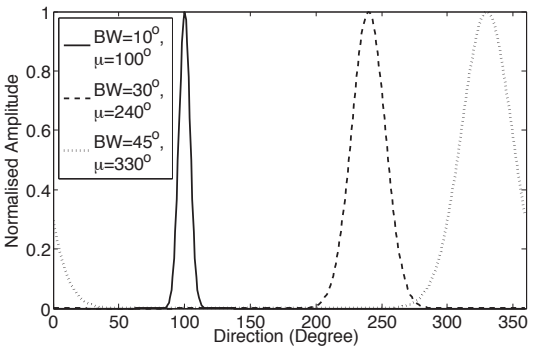

Fig. 2. Three von Mises directional filters with 10 dB, 30 dB and 45 dB beamwidths and 100°, 240° and 330° pointing directions, respectively, normalised to have maximum values of 1.

3. EXHAUSTIVE SEPARATION

In a short time-frequency window, the pressure signal $p_W(\omega,t)$ can be written as the sum of pressure waves arriving from all directions, independent of the number of sound sources. Then, a crude approximation of the plane wave $s(\mu,\omega,t)$ arriving from direction μ can be obtained by spatial filtering $p_W(\omega,t)$ as,

$$\tilde{s}(\mu,\omega,t) = p_W(\omega,t)\,f\big(\gamma(\omega,t);\mu,\kappa\big), \quad (12)$$

where $f\big(\gamma(\omega,t);\mu,\kappa\big)$ is the directional filter defined by the von Mises function, which is the circular equivalent of the Gaussian function. The von Mises function is defined by [11],

$$f(\theta;\mu,\kappa) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)}, \quad (13)$$

where, $0 < \theta \leq 2\pi$, $0 \leq \mu < 2\pi$ is the mean direction, $\kappa > 0$ is the concentration parameter and $I_0(\kappa)$ is the modified Bessel function of order zero. The concentration parameter κ is logarithmically related to the 6 dB beamwidth, θ_{BW} as $\kappa = \ln 2/\left[1-\cos(\theta_{BW}/2)\right]$.

Fig. 2 shows the plot of the three von Mises directional filters with 10 dB, 30 dB and 45 dB beamwidths and 100°, 240° and 330° pointing directions, respectively normalised to have maximum values of 1.

By this directional filtering, the time-frequency samples of the pressure signal p_W are emphasized if the intensity vectors for these samples are on or around the look direction μ; otherwise, they are suppressed.

For exhaustive separation, N directional filters are used with look directions μ varied by $2\pi/N$ intervals. Then, the spatial filtering yields a row vector $\tilde{\mathbf{s}}$ of size N for each time-frequency component:

$$\tilde{\mathbf{s}}(\omega,t) = \begin{bmatrix} f_1(\omega,t) & 0 & \dots & 0 \\ 0 & f_2(\omega,t) & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & f_N(\omega,t) \end{bmatrix} \begin{bmatrix} p_W(\omega,t) \\ p_W(\omega,t) \\ \vdots \\ p_W(\omega,t) \end{bmatrix} \quad (14)$$

where $f_i(\omega,t) = f(\gamma(\omega,t);\mu_i,\kappa)$.

This method implies block-based processing, such as with the overlap-add technique. The recorded signals are windowed and converted into the frequency domain after which each sample is processed as in (14). These are then converted back into the time-

domain, windowed with a matching window function, overlapped and added.

The selection of the time window size is important. If the window size is too short, then low frequencies can not be calculated efficiently. If, however, the window size is too long, both the correlated interference sounds and reflections contaminate the calculated intensity vector directions due to simultaneous arrivals.

It should also be noted that although the processing is done in the frequency domain, the deterministic application of the spatial filter eliminates any permutation problem, which is normally observed in other frequency-domain BSS techniques due to independent application of the separation algorithms in each frequency bin [12].

4. DIMENSIONALITY REDUCTION

Let us assume that the exhaustive separation by block-based processing yields a time-domain signal matrix $\tilde{\mathbf{S}}$ of size $N \times L$, where L is the common length of the signals and typically $N \ll L$. Using (12) and (13), it can be shown that the columnwise sum of $\tilde{\mathbf{S}}$ equals to $p_W(t)$, because, $\int_0^{2\pi} \tilde{s}(\mu, \omega, t) d\mu = p_W(\omega, t)$ due to the fact that $\int_0^{2\pi} f(\theta; \mu, \kappa) d\mu = 1$. Therefore, the exhaustive separation does not introduce additional noise or artifact, which is not present in $p_W(t)$ originally. However, it should be noted that individual separated signals may contain artifacts due to misplaced time-frequency components.

The singular value decomposition (SVD) of the signal matrix $\tilde{\mathbf{S}}$ can be expressed as [13],

$$\tilde{\mathbf{S}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{k=1}^{p} \sigma_k \mathbf{u}_k \mathbf{v}_k^T, \tag{15}$$

where $\mathbf{U} \in \mathbb{R}^{N \times L}$ is an orthonormal matrix of left singular vectors $\mathbf{u}_k$, $\mathbf{V} \in \mathbb{R}^{L \times L}$ is an orthonormal matrix of right singular vectors $\mathbf{v}_k$, $\mathbf{\Sigma} \in \mathbb{R}^{N \times L}$ is a pseudo-diagonal matrix with σ_k values along the diagonals and $p = \min(N, L)$.

The dimension of the data matrix $\tilde{\mathbf{S}}$ can be reduced by only considering a signal subspace of rank m, which is selected according to the relative magnitudes of the singular values as,

$$\breve{\mathbf{S}} = \sum_{k=1}^{m} \sigma_k \mathbf{u}_k \mathbf{v}^T \tag{16}$$

By selecting only the highest m singular values, independent rows of the $\tilde{\mathbf{S}}$ matrix are obtained that correspond to the individual signals of the mixture. Fig. 3 shows the mixture signal $p_W(t)$, the reverberant originals of each mixture signal and separated signals for three speech sounds at directions 30°, 100° and 300° recorded in a room with reverberation time of 0.32 s. The data matrix is of size $N = 360$ and $L = 88200$ samples at 44.1 kHz sampling frequency, calculated using a block window size of 4096 samples. The signal subspace has been decomposed using the highest three singular values. The three rows of the data matrix with highest RMS energy has been plotted.

When, the energies of the signals at each row of the $\breve{\mathbf{S}}$, matrix are calculated and plotted, peaks are observed at some directions. Fig. 4 shows these RMS energies for the previously given separation example.

It should be noted that, the directions of these local maxima do not necessarily correspond to the actual directions of the sound sources. This is due to the fact that highly correlated early reflections of a sound may cause a shift in the calculated intensity vector directions. While the selection of the observed direction, rather than the actual one is preferable to obtain better SIR for the purposes of BSS, for source localisation problems, a correction should be applied. However, source localisation is beyond the scope of this paper and therefore will not be discussed further.

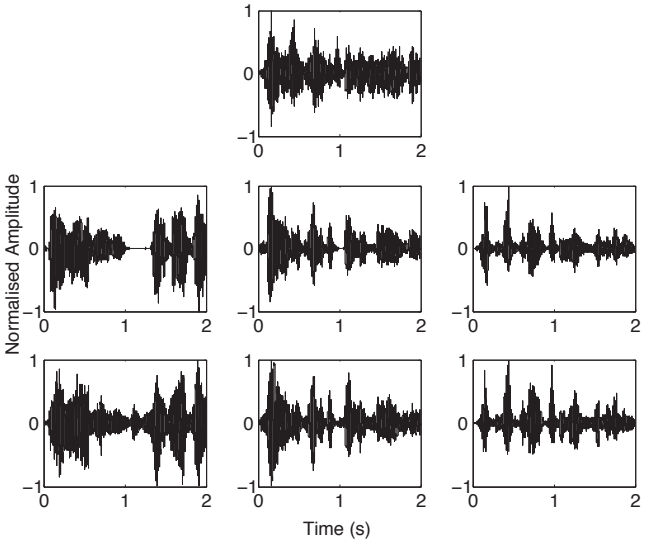

Fig. 3. The mixture signal $p_W(t)$ (top), reverberant originals of three signals (middle row) and separated signals (bottom row) for a mixture of three speech signals at directions 100°, 240° and 330° recorded in a room with reverberation time of 0.32 s.

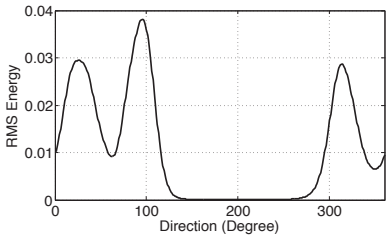

Fig. 4. The RMS energies of the signals at each row of the $\breve{\mathbf{S}}$ matrix calculated for the mixture of three speech signals whose results are given in Fig. 3.

5. RESULTS

The algorithm has been tested with 2-, 3- and 4-source mixtures of 2-second long sound signals consisting of male speech (M), female speech (F), cello (C) and trumpet (T) music of equal energy recorded in a room of size ($L = 8$ m; $W = 5.5$ m; $H = 3$ m) with a reverberation time of 0.32 s. The 2-source mixture contained MF sounds where the first source direction was fixed at 0° and the second source direction was varied from 30° to 330° with 30° intervals. Therefore, the angular interval between the sources was varied and 11 different mixtures were obtained. The 3-source mixture contained MFC sounds, where the direction of M was varied from 0° to 90°, direction of F was varied from 120° to 210° and direction of C was varied from 240° to 330° with 30° intervals. Therefore, 4 different mixtures were obtained while the angular separation between the sources were fixed at 120°. The 4-source mixture contained MFCT sounds, where the direction of M was varied from 0° to 60°, direction of F

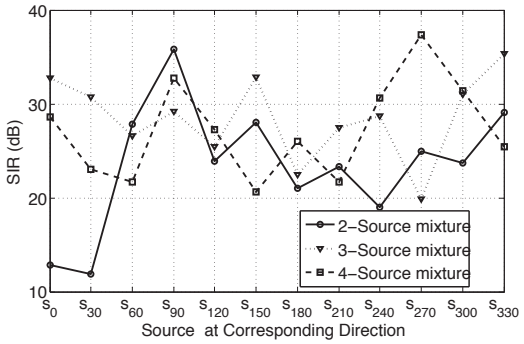

Fig. 5. SIR values in dB for each separated source for the 2-, 3- and 4-source mixtures. Angular interval between the sources increase with 30° intervals for the 2-source mixtures. For the 3-source and 4-source mixtures, the angular interval is fixed at 120° and 90°, respectively.

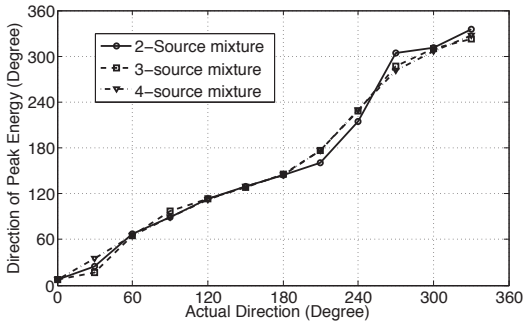

Fig. 6. Actual source directions and directions of the RMS energy peaks in the reduced dimension data matrices calculated for the 2-, 3- and 4-source mixtures.

was varied from 90° to 150°, direction of C was varied from 180° to 240° and direction of T was varied from 270° to $330^\circ$with 30° intervals. Therefore, 3 different mixtures were obtained while the angular separation between the sources were fixed at 90°. Processing was done with a block size of 4096 and a beamwidth of 10° for creating a data matrix of size 360×88200 with a sampling frequency of 44.1 kHz. Dimension reduction was carried out using only the highest six singular values.

Fig. 5 show the signal-to-interference ratios (SIR) for each separated source at the corresponding directions. The separation performance is not affected by the number of sources in the mixture as long as the angular separation between them is large enough. Fig. 6 shows the actual directions of the sources and the directions of the RMS energy peaks in the reduced dimension data matrix. As explained before, the discrepancies result from the early reflection in the environment, rather than the number of mixtures or their content.

In order to quantify the quality of the separated signals, the signal-to-distortion ratios (SDR) have also been calculated as explained in [14]. For each separated source, the reverberant $p_W(t)$ signal recorded when only that source is active at the corresponding direction was used as the original source with no distortion for comparison. The mean SDRs for the 2-, 3-, and 4-source mixtures were found as 6.46 dB, 5.98 dB, 5.59 dB, respectively. It should also be noted that this comparison based SDR calculation penalises dereverberation or other suppression of reflections, because the resulting changes on the signal are also considered as artifacts. Therefore, the actual SDRs are expected to be higher.

6. CONCLUSIONS

A deterministic BSS technique has been presented based on exhaustive separation and dimensionality reduction that uses the IVDE method [8]. Intensity vector directions were calculated using a compact array and then used for spatial filtering. SVD was applied to reduce the dimension of the data matrix, which was also used for determining the directions of the peak energies. Results have been presented for 2-, 3- and 4-source convolutive mixtures of speech and instrument sounds and good separation has been achieved. SDR and detected peak energy directions have also been displayed.

7. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[2] J-F. Cardoso, "Blind source separation: statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.

[3] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrainted adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, January 1982.

[4] L. C. Parra and C. V. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 6, pp. 352–362, September 2002.

[5] A-J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proc. IEEE*, vol. 86, no. 10, pp. 1987–2008, October 1998.

[6] H. Shindo and Y. Hirai, "Blind source separation by a geometrical method," in *Proc. 2002 IEEE Int. Joint Conf. on Neural Networks*, Honolulu, Hawaii, USA, May 2002, vol. 2, pp. 1109–1114.

[7] J. Yamashita, S. Tatsuta, and Y. Hirai, "Estimation of propagation delays using orientation histograms for anechoic blind source separation," in *Proc. 2004 IEEE Int. Joint Conf. on Neural Networks*, Budapest, Hungary, July 2004, vol. 3, pp. 2175–2180.

[8] B. Günel, H. Hacıhabiboğlu, and A. M. Kondoz, "Acoustic source separation of convolutive mixtures based on intensity vector statistics," *IEEE Trans. Audio, Speech Language Process.*, vol. 16, no. 4, pp. 748–756, May 2008.

[9] P. G. Craven and M. A. Gerzon, "Coincident microphone simulation covering three dimensional space and yielding various directional outputs," US Patent 4,042,779, 1977.

[10] F. J. Fahy, *Sound Intensity*, E&FN SPON, London, 2nd edition, 1995.

[11] K. V. Mardia and P. E. Jupp, *Directional Statistics*, Wiley, London and New York, 1999.

[12] N. Mitianoudis and M. E. Davies, "Audio source separation: Solutions and problems," *Int. J. Adapt. Control and Signal Process.*, vol. 18, no. 3, pp. 299–314, April 2004.

[13] G. Golub and C. Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, 3rd edition, 1996.

[14] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," in *Proc. Intl. Workshop on Independent Component Analysis and Blind Signal Separation*, Aussois, France, Jan 1999, pp. 261–266.