Minhas and Gaydecki *EURASIP Journal on Advances in Signal Processing* 2014, **2014**:92
http://asp.eurasipjournals.com/content/2014/1/92

EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

A hybrid algorithm for blind source separation of a convolutive mixture of three speech sources

Shahab Faiz Minhas* and Patrick Gaydecki

Abstract

In this paper we present a novel hybrid algorithm for blind source separation of three speech signals in a real room environment. The algorithm in addition to using second-order statistics also exploits an information-theoretic approach, based on higher order statistics, to achieve source separation and is well suited for real-time implementation due to its fast adaptive methodology. It does not require any prior information or parameter estimation. The algorithm also uses a novel post-separation speech harmonic alignment that results in an improved performance. Experimental results in simulated and real environments verify the effectiveness of the proposed method, and analysis demonstrates that the algorithm is computationally efficient.

Keywords: Blind source separation; Convolutive BSS; Second-order statistics; Pitch detection

1 Introduction

The blind source separation (BSS) of speech signals also known as convolutive BSS is a very challenging problem in real room environments. It can be broadly divided into two categories, those that use an information-theoretic approach and those based on de-correlation. Some of the most widely applied information-theoretic approaches include independent component analysis (ICA) [1], maximum likelihood [2], information maximisation [3] and Kurtosis maximisation [4]. Based on these information-theoretic approaches, the neural network-based algorithms presented in [5-8] are unsuitable for implementation of BSS in real time. The reason is massive complexity, i.e. the calculation of thousands of adaptive filter coefficients and also the temporal whitening problem (pp. 340–345 in [9]).

The frequency domain implementation decomposes this convolutive mixture problem into multiple instantaneous mixing problems; however, this in turn leads to scaling and permutation alignment problems (pp. 352–353 in [9]). To solve this permutation problem, many algorithms have been proposed, such as in [10] and [11], that exploit the direction of arrival (DOA) and also speech harmonics. These DOA-based algorithms are more semi-blind in nature than blind itself since they are dependent on certain geometrical arrangement. Another way to resolve the permutation alignment issue is to exploit the correlation property between separated signals at adjacent frequency bands [12]. The reliability of this and other similar techniques is based on the amount of correlation and that surely varies case by case. In [13], a different approach is developed, based on de-correlation in the frequency domain; the algorithm avoids permutation with its very slowly converging diagonalisation procedure, but this slow convergence makes it less suitable for real-time implementation. Apart from the permutation problem, there are some other frequency-based limitations as discussed in detail in [14]. In [15], a secondary algorithm is proposed, based on time-frequency masking, which improves the signal-to-interference ratio (SIR) of separated streams. Such techniques are completely dependent on the BSS of the primary algorithm, and if the primary fails, then so does the secondary.

The BSS of more than two sources is a more complicated and computationally intense problem. The fact can be seen in [16]: a detailed survey reveals that, of 400 publications that employed convolutive source separation, only 2% of the publications dealt with more than two sources. Even for two sources, it was concluded that only 10% of them really worked with varying degrees of SIR from 5 to 18 dB using real room impulse responses. The results are still questionable since it is very difficult to analyse or compare different algorithms due to a lack of unified test bench methods for performance measure [17].

* Correspondence: shahab.faiz@manchester.ac.uk
School of Electrical and Electronic Engineering, University of Manchester, F-45, Sackville Street Building, Sackville Street, Manchester M13 9PL, UK

In this paper a novel algorithm for the BSS of three speech sources in real room environments is proposed. It uses both the information-theoretic and de-correlation approaches to achieve superior source separation with fast convergence. The algorithm has low complexity and is optimised for real-time implementation. In addition, it does not require any prior parameter estimation; furthermore, a harmonic alignment methodology, presented in this paper, improves the quality of separated speech in a real room environment.

The paper is organised as follows: In the following section, the motivation behind the new hybrid algorithm will be discussed. In Section 3, the hybrid algorithm will be presented followed by a discussion of its constituent parts. In Section 4, a novel harmonic alignment method will be presented. In Section 5, the performance of the algorithm will be analysed based on a simulated room environment. The results from the real room experiment will be shown in Section 6, followed by discussion (showing computational load) and conclusion in Sections 7 and 8, respectively.

The notation that will be used in this paper will be small letter x for scalar quantities, small and bold letter $\boldsymbol{x}$ for vector quantities or first-order tensors and bold and capital letter $\mathbf{X}$ for two-dimensional matrices or second-order tensors, and for three-dimensional matrices or third-order tensors, it will be similar to two-dimensional ones but with a double bar on top $\overline{\overline{\mathbf{X}}}$.

2 Motivation

The motivation behind the hybrid algorithm will become evident as we progress in this section. The DOA-based algorithms presented in [18-21] have considered source separation cases for more than two sources in real room environments. However, for this, a single large microphone that consists of an array of microphones (within it) is used, which has limitations. The limitations are not only in the placement of sources in a geometrical arrangement, but also the performance is dependent on the distance of the microphone from sources. The source separation for a speaker behind a speaker or a speaker whose face is towards the wall (rather than the microphone) also cannot be achieved through DOA.

Most of the practical scenarios require cases that do require the placement of arbitrary microphones to pick up the stronger source signal and cancel the other weaker interfering signals. For example, in the case of musical instruments in concerts, acoustics in theatre performances, meetings in conference rooms, discussion in parliament houses, etc. All of these cases do require a BSS algorithm for real-time separation. This research will show the potential of working with only three speech sources with an equivalent number of sensors (microphones), i.e. a critically determined BSS case in a real room environment.

The case of three statistically independent speech sources (loudspeakers) and three sensors (microphones) is considered first without any background noise that can separately be dealt within a supervised way (explained later). The mathematical way of expressing this linear time invariant (LTI) system is shown as

$$x_p(n) = \sum_{q=1}^{S}\sum_{k=0}^{K-1} \mathrm{h}_{\mathrm{pq}}(k)\, s_q(n-k) \tag{1}$$

where $\boldsymbol{s}_q$ is the speech source that is convolved with the FIR filter containing the impulse response (channel response) given by $\mathbf{h}_{\mathrm{pq}}$ between the source and the sensor and then added at the sensor to give the final convolutive mixture represented by $\boldsymbol{x}_p$. In the above, K represents the length of the filters, S represents the total number of sources, i.e. three in our case, and n represents the sample number. Equation 1 represents speech signals passing through a (third-order tensor or three-dimensional) mixing matrix $\overline{\overline{\mathbf{H}}}_{\mathbf{m}}$ given by

$$\overline{\overline{\mathbf{H}}}_{\mathbf{m}} = \begin{bmatrix} \boldsymbol{h}_{11} & \boldsymbol{h}_{12} & \boldsymbol{h}_{13} \\ \boldsymbol{h}_{21} & \boldsymbol{h}_{22} & \boldsymbol{h}_{23} \\ \boldsymbol{h}_{31} & \boldsymbol{h}_{32} & \boldsymbol{h}_{33} \end{bmatrix} \tag{2}$$

To obtain original speech signals $\boldsymbol{s}_1$, $\boldsymbol{s}_2$ and $\boldsymbol{s}_3$, the de-mixing matrix $\overline{\overline{\mathbf{W}}}_{\mathbf{d}}$ needs to be calculated. Most of the algorithms only use simulated room environments for mixing matrix instead of real room as shown in [22,23]. Apart from this, the temporal whitening caused by the equalisation filters $\mathbf{w}_{11}$, $\mathbf{w}_{22}$ and $\mathbf{w}_{33}$ will render the output useless. To address this problem, in [24], a linear predictive codec-based solution is proposed, but that is not suitable in all cases. In [25], it is stated that the main difficulty is that audio source separation problems are usually mathematically ill-posed and to succeed it is necessary to incorporate additional knowledge about the mixing process and/or the source signals. However, by definition, blindness implies an absence of prior information.

This research has exploited the fusion of two different criteria, i.e. one based on de-correlation and the other based on information theory. The former requires the implementation in the frequency domain, and the latter requires that in the time-frequency domain using neural networks. This fusion used in the hybrid algorithm improves the SIR performance compared to each technique if used individually (independently). It obviates the requirement for semi-blind array processing methodologies to resolve the permutation problem. It also does not have any temporal whitening problem and is suitable for real-time digital signal processing (DSP) board implementation based on its low computational load shown later on.

3 Hybrid algorithm

The hybrid algorithm is a frequency domain multiple conditioned integrated approach for the solution of BSS problems respecting speech signals in real room environments. Its adaptive methodology not only converges faster but is computationally efficient for real-time hardware implementation. In blind signal processing, neither a reference signal nor any prior information regarding the channel is provided, so the algorithms proposed in this field use separation criteria that are actually mere conditions imposed on the output streams. These include changing the probability density function from Gaussian to super-Gaussian [3] or by de-correlation of the output streams [13,26].

The hybrid algorithm proposed here uses two different conditions instead of one. The conditions and implementation mechanism are chosen in such a way that they actually mitigate each other's flaws and work complimentarily by improving the signal-to-interference ratio at the output (discussed in the following sections). The block diagram of the hybrid algorithm is shown in Figure 1.

Here $x_1(n)$, $x_2(n)$ and $x_3(n)$ are three convolved mixed streams of data coming from the sensors. The output of the algorithm $u_1(n)$, $u_2(n)$ and $u_3(n)$ are three separated signals. The hybrid algorithm fuses two approaches based on two conditions in a sequential manner. The first approach uses frequency domain diagonalisation based on a de-correlation condition; the second approach is neural network feedback based on a statistical independence condition using information maximisation [3]. The reason for choosing each condition with its relevant approach will be discussed in the following subsections. The implementation mechanism for both of these approaches is novel. Each structure of the hybrid algorithm, i.e. controlled frequency domain diagonalisation (CFDD) and frequency domain adaptive feedback separation (FDAFS), will be discussed in the following two subsections.

3.1 Controlled frequency domain diagonalisation

Frequency domain diagonalisation is applied here through a controlled mechanism in order to avoid the permutation problem similar to that shown by Schobben and Sommen in [13] for two sources in a real room environment. Joint diagonalisation of correlation matrices based on the Jacobi method [27] could also be implemented in the frequency domain for convolutive mixture problems, but the adaptive controlled diagonalisation mechanism proposed here is more robust.

The CFDD starts by converting the time domain BSS problem into the frequency domain. This simplifies the time domain (multi-dimensional) matrix inversion problem to bin-by-bin separation in the frequency domain. The time to frequency domain conversion process is performed by using the overlap and save method; a Hanning window is applied. This is also known as the short-time Fourier transform (STFT). The length of the fast Fourier transform is N, the length of the filter in the

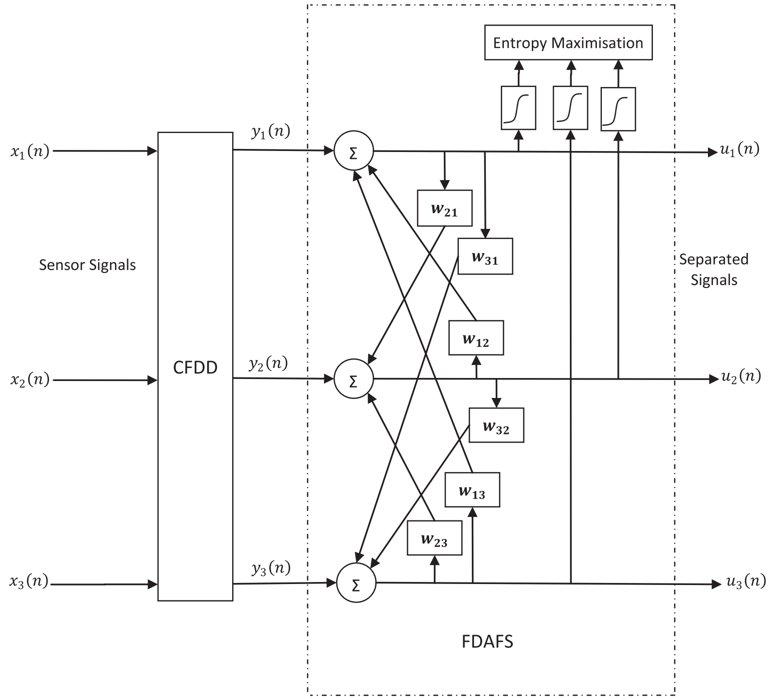

Figure 1 Block diagram of the hybrid algorithm showing CFDD and FDAFS blocks.

time domain is K and the size of the speech signal block taken is B.

The frequency domain conversion using the fast Fourier transform (FFT) of the convolved mixed streams of data shown in Equation 1 is

$$\mathrm{X}_i(\mathrm{k}) = \mathcal{F}\{x_i(\mathrm{n})\}, i = \{1, 2, 3\} \tag{3}$$

$$\mathbf{R}_{kb} = \begin{bmatrix} \mathrm{X}_1(\mathrm{k})\mathrm{X}_1^*(\mathrm{k}) & \mathrm{X}_1(\mathrm{k})\mathrm{X}_2^*(\mathrm{k}) & \mathrm{X}_1(\mathrm{k})\mathrm{X}_3^*(\mathrm{k}) \\ \mathrm{X}_2(\mathrm{k})\mathrm{X}_1^*(\mathrm{k}) & \mathrm{X}_2(\mathrm{k})\mathrm{X}_2^*(\mathrm{k}) & \mathrm{X}_2(\mathrm{k})\mathrm{X}_3^*(\mathrm{k}) \\ \mathrm{X}_3(\mathrm{k})\mathrm{X}_1^*(\mathrm{k}) & \mathrm{X}_3(\mathrm{k})\mathrm{X}_2^*(\mathrm{k}) & \mathrm{X}_3(\mathrm{k})\mathrm{X}_3^*(\mathrm{k}) \end{bmatrix}_b \tag{4}$$

where $\mathbf{R}_{kb}$ is a frequency domain correlation matrix where k denotes the bin number and b denotes the block number and the asterisk indicates the conjugate value. In order to obtain the de-mixing (inverse) system adaptively, a strong correlation should exist over multiple blocks. However, this is not the case in speech that is stationary only over 10 to 30 ms, and apart from that, it is non-stationary. So, the first step of the algorithm is the block-based correlation constraint, realised as

$$\overline{\mathbf{R}}_{kb} = \alpha\overline{\mathbf{R}}_{k\,(b-1)} + (1-\alpha)\mathbf{R}_{kb} \tag{5}$$

In the above, α is the weighting factor that can take any value from $0 \to 1$. The value recommended for non-stationary signals like speech is above 0.9. This step can also be referred as the intersection of solution sets as in [28]. The initial correlation matrix from which Equation 5 starts is an identity matrix. Now, taking the square root inverse of the constrained correlation matrix and apply normalisation,

$$\mathbf{W}_{kb} = \overline{\mathbf{R}}_{kb}^{-1/2} \tag{6}$$

$$\mathbf{W}_{kb} = \frac{\mathbf{W}_{kb}}{\|\mathbf{W}_{kb}\|} \tag{7}$$

The purpose of the normalisation is to avoid whitening or scaling problems. Equation 6 removes only cross-correlation elements of the constrained correlation matrix $\left(\mathbf{W}_{kb}\overline{\mathbf{R}}_{\mathrm{kb}}\mathbf{W}_{kb}^{\mathrm{H}} = \mathbf{I}\right)$. However, the whitening forces the diagonalisation matrix to be an identity matrix; this results in a variation of power in each bin that results in spectral distortion. In order to avoid this, the normalisation shown in Equation 7 is applied to the de-mixing matrix either by dividing the de-mixing matrix by its maximum eigenvalue or by the Frobenius norm given by the following equation:

$$\|\mathbf{W}^{\mathrm{L,M}}\| = \sqrt{\sum_{i=1}^{\mathrm{L}}\sum_{j=1}^{\mathrm{M}}\left(\mathbf{W}_{(i,j)}\right)^2}$$

where L and M are the number of rows and columns of the matrix, respectively. Prior to this step, all the steps followed in the algorithm are similar to those proposed by Schobben and Sommen in their ECoBLISS algorithm [13]. Unlike the ECoBLISS that uses a hard unitary matrix condition for the update of the de-mixing matrix $\mathbf{W}_{kb}$ from the previous blocks, the CFDD uses a stochastic-based approach for updating. The stochastic-based approach is similar to that shown for the instantaneous case of BSS based on the Frobenius norm in [29], but here it is applied to the convolutive case. The previous block de-mixing matrix is made unitary by minimisation of the following cost function and running it in a least mean square (LMS) manner, i.e.

$$\mathcal{J}_1 \triangleq \left\|\mathbf{W}_{k(b-1)}\mathbf{W}_{k(b-1)}^{H} - \mathbf{I}\right\|_{\mathrm{F}}^2 \tag{8}$$

The gradient of the above cost function is

$$\nabla\mathbf{w}\,\mathcal{J}_1 = 4\left(\mathbf{W}_{k(b-1)}\mathbf{W}_{k(b-1)}^{H} - \mathbf{I}\right)\mathbf{W}_{k(b-1)} \tag{9}$$

Finally, the de-mixing matrix is updated as shown in the following equation:

$$\mathbf{W}_{kb} = \mathbf{W}_{k(b-1)} \times \left\{ \overline{\mathbf{R}}_{kb}^{-1/2} \Big/ \left\|\overline{\mathbf{R}}_{kb}^{-1/2}\right\| \right\} \tag{10}$$

The above steps are calculated for $N/2$ bins since the other half is the conjugate mirror of it. Also, $\mathbf{W}_{kb}$ needs to be adjusted to avoid circular convolution and perform linear convolution, a step that can be seen in the next section too. The original signals s_1, s_2 and s_3 can be recovered by multiplying the de-mixing matrix $\mathbf{W}_{kb}$ with the mixed streams bin $\mathrm{X}_1(k)$, $\mathrm{X}_2(k)$ and $\mathrm{X}_3(k)$ and then taking the inverse Fourier transform (IFFT) of the signal to convert it back to the time domain. The final step before the inverse STFT is

$$\begin{bmatrix} \mathrm{Y}_1(k) \\ \mathrm{Y}_2(k) \\ \mathrm{Y}_3(k) \end{bmatrix}_b = \mathbf{W}_{kb} \times \begin{bmatrix} \mathrm{X}_1(k) \\ \mathrm{X}_2(k) \\ \mathrm{X}_3(k) \end{bmatrix}_b \tag{11}$$

The permutation problem in the algorithm is resolved by the linear convolution constraint that results in the population of zeros in the time domain that links the otherwise independent frequencies, similar to Parra and Spence in [30]. However, the length of the filter K versus frequency resolution constrains the length of the filter to be less than the typical impulse response of the room, approximately 200 to 300 ms. This CFDD algorithm presented here has a more flexible approach based on LMS and also avoids (three-dimensional) matrix inversion. However, the process of convergence is deliberately slowed (discussed later) through over-damping to achieve a robust SIR for all cases. However, this drawback and short filter length are mitigated with the help of the second structure, FDAFS, in the hybrid algorithm.

3.2 Frequency domain adaptive feedback separation

The FDAFS algorithm is based on an information-theoretic approach. The criterion it uses is information maximisation or Infomax [3]. This employs a non-linear function, such as a logistic sigmoid, to exploit the higher order statistics based on super-Gaussian characteristics of the speech signal. The time domain feedback (TD-FB) implementation can be seen in [8] and is expanded to three sources, shown by the following equations:

$$u_i(n) = y_i(n) + \sum_{j \neq i} \sum_{k=1}^{K} \mathrm{w}_{ij}(k)\ u_j(n-k), \quad i,j = \{1,2,3\} \tag{12}$$

In the above equation, the separated output stream is shown by $\boldsymbol{u}_i$. The coefficients of the de-mixing filters $\mathbf{w}_{ij}$ are estimated by

$$\Delta w_{ij}(k) \propto \hat{z}_i(n)\ u_j(n-k), \quad i \neq j \tag{13}$$

where

$$\hat{z}_i(n) = 1-2\left\{1 \Big/ \left(1+e^{-\beta \mathrm{u_i(n)}}\right)\right\}$$

The β in the above equation is the slope parameter; in this algorithm, it is merely assigned the value 1. The purpose of choosing this feedback neural network approach without using equalisation filters $\boldsymbol{w}_{ii}$ is to avoid temporal whitening (also equalisation has nothing to do with separation). It is important to emphasise that the de-mixing filters avoid the inverse of the deterministic mixing matrix but still require the inverse of filters $\boldsymbol{h}_{11}$, $\boldsymbol{h}_{22}$ and $\boldsymbol{h}_{33}$ to estimate the de-mixing filters that also need to be realisable (for details, see the inverse of non-minimum phase systems, pp. 348–349 in [9]). However, it is far less problematic than the inverse of the determinant mixing matrix due to two reasons. Firstly, if the sensors are closer to the speech sources, the unrealisable inverse filtering problem can be avoided all the time [8]. Secondly, the first structure (CFDD) is sufficiently robust and FDAFS works sequentially on already de-correlated speech signals.

The frequency domain implementation structure used in FDAFS is based on fast block-by-block calculation of coefficients instead of sample by sample. For this, the frequency domain block LMS methodology is modified for feedback adaptation and is shown in Figure 2.

This time-frequency domain implementation shown in Figure 2 for the filter $\boldsymbol{w}_{12}$ does not have any permutation problem. The reason is that the separation is not completely in the frequency domain: the error is estimated in the time domain, and therefore, it is not a bin-by-bin separation as in the case of CFDD. The working details of the block LMS can be seen in ([31], pp. 350–353); the above structure is just the interpretation of Equations 12 and 13 whilst using the overlap and save method. Only the power constraint block is integrated into the structure. That is needed to normalise the coefficients of each bin with the corresponding power from each bin of the output signal. Here the output signal is selected for normalisation instead of the input signal due to its superior performance.

4 Harmonic alignment

The purpose of using harmonic alignment (HA) is to exploit the properties of the speech signal to improve the SIR in a real room environment. The DOA techniques [10,11] also use speech properties to align harmonics at lower frequencies where the width of the beam becomes broader. However, in our case, it is applied in a different way after the hybrid algorithm on separated speech.

A speech signal can be broadly divided into two parts: voiced and unvoiced, where the voiced part can be further divided into formants and fricatives (for details, see pp. 121–151 in [32]). The formant part consists of the fundamental frequency (pitch) and the harmonics. It is this part that is exploited to improve the quality of separation. First, the pitch of three output streams of the hybrid algorithm $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ containing separated signals is calculated for small segments of speech called syllables. The size of each syllable is based on the quasi-stationary property of speech and is typically between 10 and 30 ,ms. Any pitch detection algorithm can be used, based on the FFT, and must have a high frequency resolution per bin. For the purpose of completeness, a pitch detection algorithm is shown as below:

$$\mathrm{P}_1 = |\mathrm{U}_1|, \mathrm{U}_1(k) = \mathcal{F}\{u_1(n)\}$$
$$\mathrm{P}_2 = \mathrm{P}_3 = \mathrm{P}_4 = \mathrm{P}_5 = \mathbf{1}$$

Only a few harmonics are needed to calculate the pitch since the maximum energy of the formant is based on initial harmonics. In the above equation, P_2, P_3, P_4 and P_5 are the holding vectors of the second, third, fourth and fifth harmonics, respectively, that are initially populated with a vector of ones. In order to populate them with their required harmonics that are the multiples of the fundamental, the following loop is used:

```
loop: for $q = 1$ to $({}^{N}/_{2}-(a-1))/a$
$\mathrm{P}_a(q) = \{\mathrm{P}_1(q \times a) + \mathrm{P}_1(q \times a + c)\ldots\}/a$,   where $c$
        $= 1, \ldots, (a-1)$,   end
```

In the above, N corresponds to the size of the FFT and a denotes the harmonic number. The pitch is calculated by first taking the product of these harmonic vectors and then obtaining the maximum bin number by applying a periodogram maximiser as shown below:

$$m = \mathrm{argmax}_n\{\mathrm{P}_1 \times \mathrm{P}_2 \times \mathrm{P}_3 \times \mathrm{P}_4 \times \mathrm{P}_5\} \tag{14}$$

The m corresponds to the bin number that has the maximum energy in its fundamental and harmonics. The fundamental frequency is calculated from the bin by

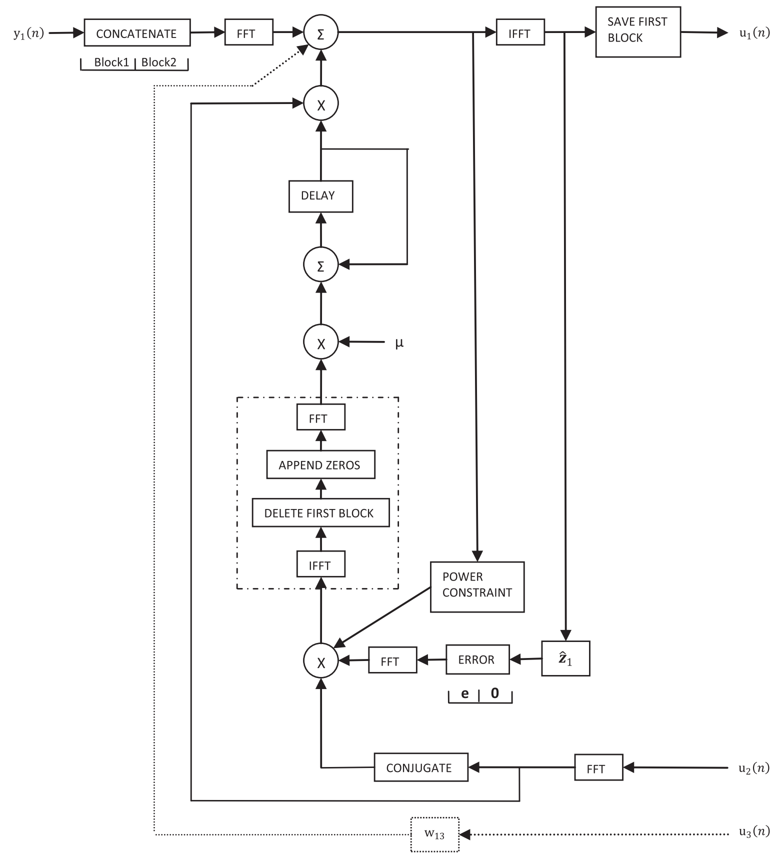

Figure 2 Flow diagram of the FADS de-mixing filter w_{12}.

using $f_f = m \times (N/f_s)$, where f_s is the sampling frequency. Only the fundamental frequency in the range ($50\ \text{Hz} \le f_f \le 450\ \text{Hz}$) is considered a pitch of the formant part of human speech, and anything else is either a non-voiced segment or the noise part (fricatives). Figure 3 shows the formant and fricative sections of a segment of the speech signal from a single speaker. The pitch detection only in formant sections can be seen in Figure 4b. If the harmonic and pitch are removed from this segment, as shown in Figure 4c,d, then the resultant signal will contain only the fricatives and the residual formant (i.e. greatly degraded in strength). The removal of the harmonic and the pitch will be discussed shortly.

The pitch is calculated (similar to that shown for a single speech) for each syllable of the three output streams. The fundamental of harmonic alignment used here is that each output stream contains a primary pitch of the separated speaker and the secondary pitches of the suppressed interfering speakers. These suppressed speech signals need to be further suppressed by removing the harmonic content from them. Let us suppose that in $u_1(n)$ the primary pitch is that of separated signal $s_1(n)$ and the secondary pitches are of suppressed speakers $s_2(n)$ and $s_3(n)$. The primary pitch f_1^s is calculated from $u_1(n)$ and the secondary pitches f_2^s and f_3^s are calculated from $u_2(n)$ and $u_3(n)$, respectively. The associated amplitudes with these pitches obtained from Equation 14 are a_{f1}^s, a_{f2}^s and a_{f3}^s. The superscript s shows the syllable number. The algorithm is stated as follows:

$$\begin{array}{c}
\text{if } 50\ \text{Hz} \le f_2^s \le 450\ \text{Hz} \\
\text{if } f_2^s = f_1^s \text{ and } a_{f2}^s \le a_{f1}^s \text{ then do nothing} \\
\text{else loop: for } c = 1 \text{ to } d \\
Z_1^s\left\{c * m_2^s \pm v\right\} = \text{zeros} \quad \text{end} \\
\text{end end}
\end{array}$$

In the above, d is the number of harmonics that needs to be removed and v is the width of the comb filter that is needed to remove adjacent frequencies. v can be variable instead of a fixed value, and Z_1^s is the output of HA of the first stream, initialised by $Z_1^s = U_1^s$. Similarly, the pitch frequency from $u_3(n)$ can be removed in the same way. The last step is shown as below:

$$\begin{array}{c}
\text{if } 50\ \text{Hz} \le f_1^s \le 450\ \text{Hz} \\
\text{if } \left(f_1^s = f_2^s \text{ and } a_{f1}^s \le a_{f2}^s\right) \text{ or } \left(f_1^s = f_3^s \text{ and } a_{f1}^s \le a_{f3}^s\right) \text{ then do nothing} \\
\text{else loop: for } c = 1 \text{ to } d \\
Z_1^s\left\{c * m_1^s \pm v\right\} = U_1^s\left\{c * m_1^s \pm v\right\} \quad \text{end} \\
\text{end end}
\end{array}$$

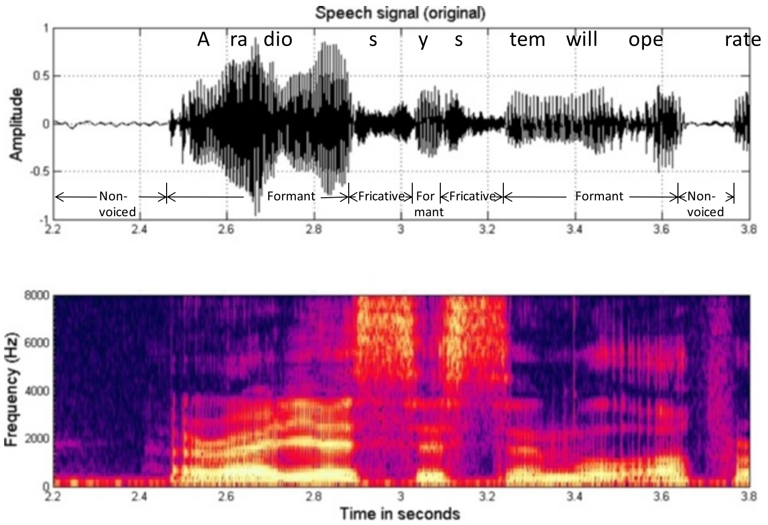

Figure 3 The time domain speech signal and its equivalent spectrogram. The section of the sentence that is uttered by the speaker is 'A radio system will operate'. After the non-voiced section at the end, it is again a formant section 'rate'.

The above step ensures that the quality of the primary speech signal is not affected, and the reason is that the whole HA algorithm is based on the FFT that has an inherent problem of spectral leakage. For this reason, a comb filter is used, but it has a drawback of removing additional adjacent frequencies.

5 Simulated room environment experiment and analysis

The hybrid algorithm has been tested with an artificial room impulse response based on the Stephen room acoustic model [33]. Three speech signals were recorded separately in an anechoic chamber to preserve the super-Gaussian characteristic of a pure speech signal, necessary for the information-theoretic part of the hybrid algorithm. The simulated room environment denoting the placement of the microphones and the speakers is illustrated in Figure 5.

The sampling frequency was 16 kHz, the room reverberation length was 50 ms and the room reflection coefficient value was 0.4. The overlap used in both CFDD and FDAFS was 50%, the recommended optimum value based on complexity. The size of the Fourier transform used in CFDD was 2,048 bins and that in FDAFS was

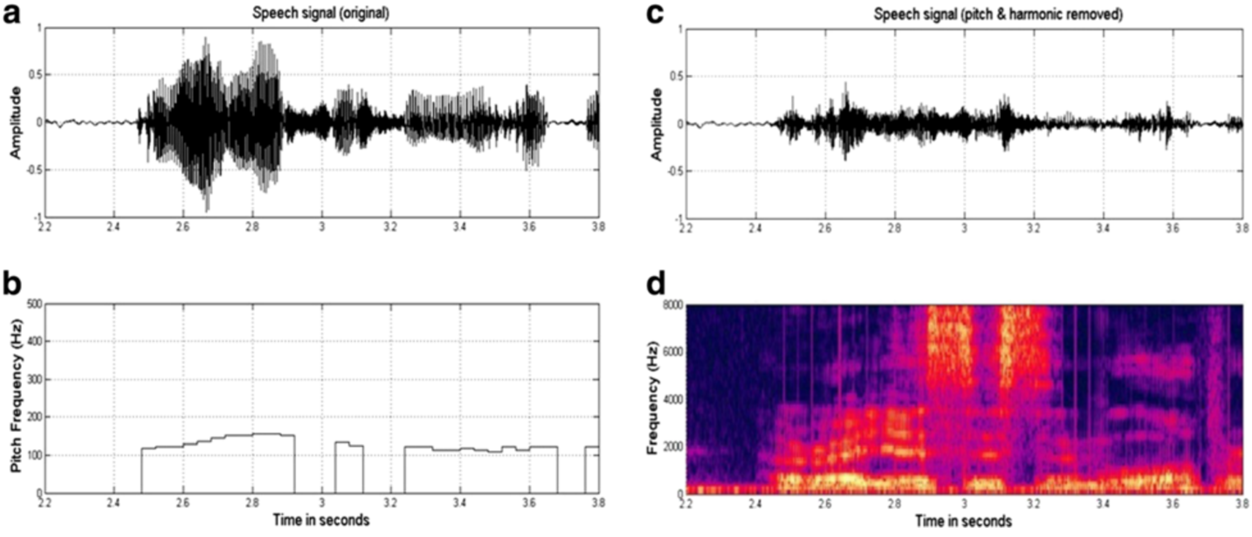

Figure 4 Speech signal segment, detected pitch, speech segment with pitch and harmonic removed, and time domain signal spectrogram. (a) The segment of the speech signal as shown in Figure 3. **(b)** The pitch detected only in the formant sections. **(c)** The same speech segment but pitch and harmonic removed from it (only in formants). **(d)** The spectrogram of the time domain signal in **(c)**.

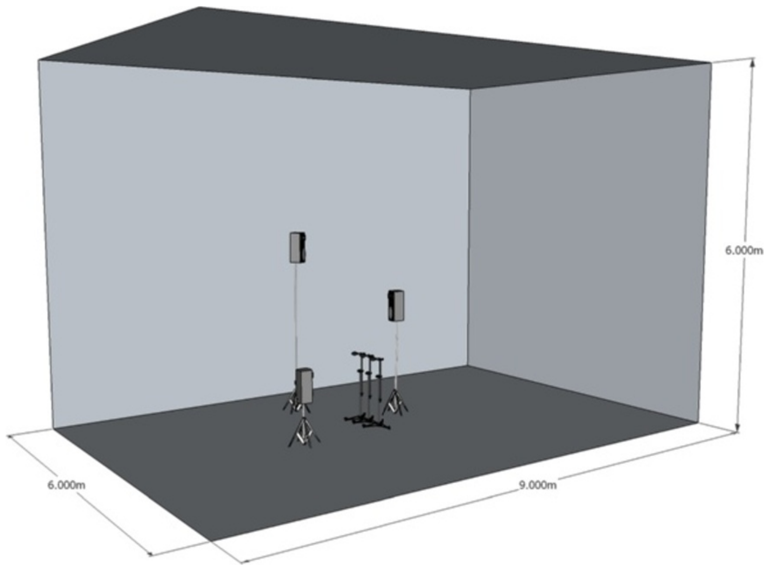

Figure 5 Simulated room environment showing the placement of microphones and loudspeakers. The figure is drawn to scale. The 3D Cartesian coordinates of microphones in metres are (4.52, 2.48, 1.16), (4.52, 2.72, 1.16) and (4.52, 3.00, 1.16) and of loudspeakers are (4.00, 1.30, 2.79), (3.00, 3.20, 0.79) and (5.00, 2.50, 1.79).

1,600 bins with a time domain filter length of 800 taps equal to that of room reverberation. The performance of the hybrid algorithm is based on SIR calculated for each output stream using the following equation:

$$\mathrm{SIR}_q = 10\log\frac{\sum_n\left|\sum_{p=1}^{3} w_{qp}(n) * h_{pq}(n) * s_q(n)\right|^2}{\sum_n\left|u_q(n) - \sum_{p=1}^{3} w_{qp}(n) * h_{pq}(n) * s_q(n)\right|^2} \qquad (15)$$

Here q denotes the stream number. The SIR was calculated over the entire range of speech signals using a sliding window syllable of size 20 ms that was equivalent to 320 samples at a sampling frequency of 16 kHz. The performance of the hybrid algorithm is summarised in Table 1. The whole speech length was divided into 4-s segments, and from each segment, the SIR is shown based on a small section of speech (syllable). Also, the best performance given by FDAFS if used independently for separation is not more than 4 to 5 dB which is very poor and, for this reason, not shown separately. The same is true for its time domain equivalent TD-FB [8].

Table 1 shows the performance of the hybrid algorithm for all the three sources. The table also shows the average signal-to-artefact ratio (SAR) [34] of all the streams of the hybrid algorithm. Also, the performance of one of the

Table 1 SIR of hybrid algorithm for the experiment performed in the simulated room environment shown in Figure 5

		Time (s)										
		0 to 4	**4 to 8**	**8 to 12**	**12 to 16**	**16 to 20**	**20 to 24**	**24 to 28**	**28 to 32**	**32 to 36**	**36 to 40**	**40 to 44**
SIR (dB)	Input (stream 1)	4.7	0.8	−9.8	4.6	−1.4	0.8	3.2	−4.4	−1.3	−18.3	−3.1
	CFDD	5.0	1.8	−4.2	15.8	0.2	4.8	10.4	5.4	12.2	−12.5	7.6
	Hybrid	7.5	4.2	−0.7	23.1	10.4	7.7	14.4	9.2	12.7	−9.3	11.7
	Input (stream 2)	4.1	−4.1	0.4	−5.4	−5.6	−3.1	−10.3	−8.8	−6.5	0.5	−6.4
	CFDD	7.0	−1.9	5.4	0.6	0.5	1.1	2.1	6.0	6.0	11.2	4.0
	Hybrid	11.1	2.4	14.7	5.3	10.7	8.4	7.1	10.2	15.9	15.4	8.2
	Input (stream 3)	−5.2	−1.8	−5.1	−2.2	−0.3	1.9	−21.4	7.5	−1.9	3.4	0.7
	CFDD	−3.2	1.4	3.3	5.7	3.6	8.5	−12.8	21.3	12.0	13.5	14.2
	Hybrid	1.0	6.3	10.4	9.8	10.0	13.3	−10.2	25.6	14.5	16.3	15.5
SAR	Average SAR hybrid	32.72	31.24	27.92	26.16	21.13	19.28	18.12	15.71	15.9	16.87	18.24

The whole speech length has been divided into 4-s segments, and SIR and SAR are calculated on 20-ms syllable within that. The room reverberation value is taken to be 0.4 with a room impulse response length of 50 ms.

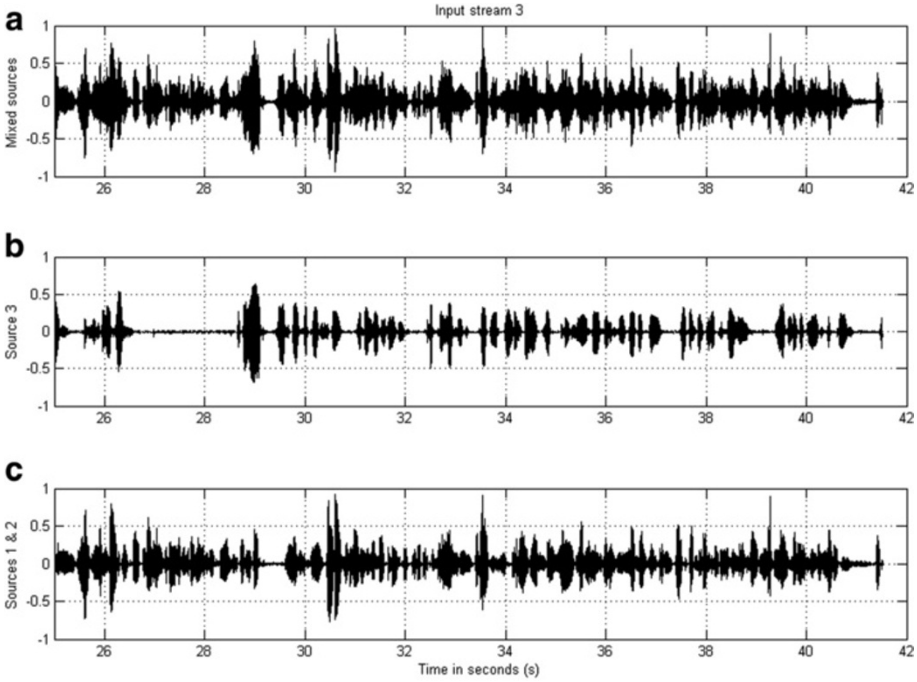

Figure 6 Input stream 3, speech signal from source 3 and interfering speech signals from sources 1 and 2. **(a)** Input stream 3 at microphone 3 containing a speech signal from source 3 and interfering speech signals from sources 1 and 2. **(b)** The speech signal from source 3 separately. **(c)** The interfering speech signals from sources 1 and 2 separately.

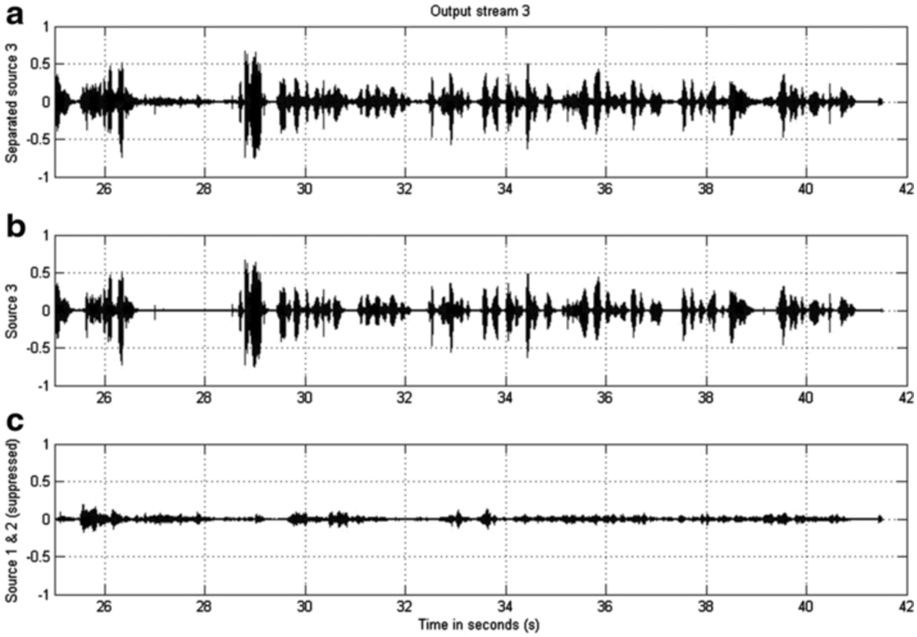

Figure 7 **Output stream 3, speech signal from source 3 and interfering speech signals from sources 1 and 2. (a)** Output stream 3 of the hybrid algorithm containing a separated speech signal from source 3 and suppressed interfering signals from sources 1 and 2. **(b)** The separated speech signal from source 3 separately. **(c)** The suppressed interfering speech signals from sources 1 and 2 separately. (Output gain normalised).

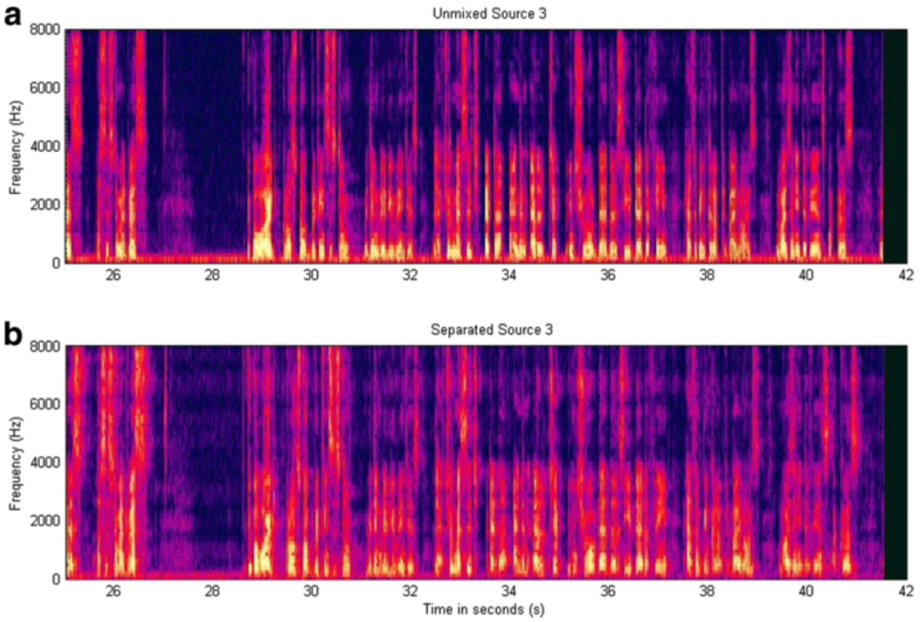

Figure 8 Spectrograms of the unmixed and separated source signals. (a) The spectrogram of the unmixed source signal as shown in Figure 6b. **(b)** The spectrogram of the separated source signal as shown in Figure 7b.

sources (third) can separately be seen in Figures 6, 7, 8. These figures show the performance of the method over the last section of speech, i.e. in the range of 25 to 42 s. This is due to two reasons: the first is to show the smooth suppression of interference and the second is to show that the algorithm retains its minimum point.

It is evident from Table 1 that the hybrid algorithm gives superior performance to CFDD and FDAFS if used independently; after 20 s, it improves the SIR of the input by 12 dB. It is very important to emphasise that the CFDD coefficients are not updated with each consecutive block but is updated after the fifth block with $\alpha = 0.9$. Therefore, it is more immune to the non-stationary behaviour of the speech signal, manifests reduced computational load and converges to a true minimum. This over-damped criterion is necessary but results in a very slow convergence that is compensated by the FDAFS (second stage). However, it has the benefit of reduced complexity since estimation and update is performed only once in five blocks.

In Table 1, it can be seen that for input (stream 1) during the segment from 16 to 20 s, the SIR improvement shown by the CFDD on a small section of speech is only 1.6 dB. It is due to an anomaly (permutation misalignments): certain harmonics that need to be suppressed more are not suppressed at all. These anomalies do happen due to the slow learning process of the CFDD but reduce in amplitude as time progresses. However, these limitations are addressed in the FDAFS (second stage) as can be seen in Table 1. It is pertinent to mention that this slow learning process is to avoid local solution (associated with the non-stationary nature of speech signals). If the weight factor is increased in the algorithm (CFDD), the separation can be achieved in 2 s, but this separation will be local. The coefficients of this separation when applied to the next segment of the convolutive mixture of speech signals will not do any separation. So, either another local solution (permuted) is obtained (that is useless) or true separation filters are calculated.

Table 2 Average SIR of hybrid algorithm for the experiment performed in the simulated room environment shown in Figure 5

		Time (s)										
		0 to 4	**4 to 8**	**8 to 12**	**12 to 16**	**16 to 20**	**20 to 24**	**24 to 28**	**28 to 32**	**32 to 36**	**36 to 40**	**40 to 44**
Average SIR (dB)	Input (streams)	−0.02	−2.68	−2.69	−3.79	−3.72	−0.15	−2.22	−3.90	−2.24	−2.57	−4.34
	CFDD	2.44	0.33	3.06	0.270	1.85	3.44	3.87	5.848	5.489	7.001	5.081
	ECoBLISS	0.04	−1.80	−0.96	−1.01	0.618	2.038	2.752	6.42	4.477	8.32	4.158
	Hybrid	4.13	3.52	7.66	4.30	4.93	8.209	6.505	9.213	7.903	10.2	7.889

The whole speech length has been divided into 4-s segments, and SIR is calculated on 20-ms syllable within that. The room reverberation value is taken to be 0.6 with a room impulse response length of 100 ms.

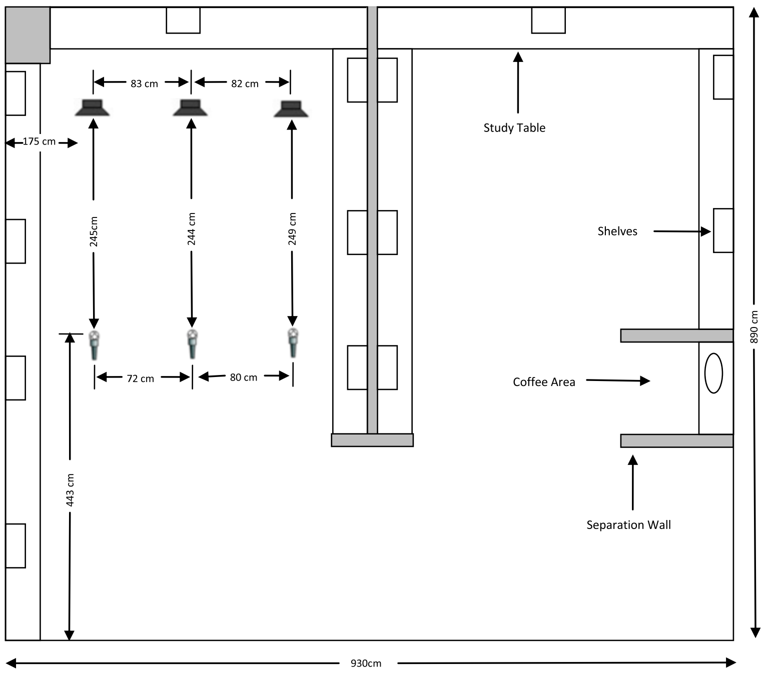

Figure 9 Schematic view of the experiment conducted in the laboratory. It shows the room dimension and the placement of the microphones and loudspeakers. The heights of the loudspeakers from left to right are 118, 118 and 120 cm, and those of the microphones from left to right are 102, 98 and 97 cm. The figure is not drawn to scale.

For this reason, longer speech results (40 s) have been shown to verify the convergence of the algorithm to its true minimum and thus obtaining the true separation filters.

In Table 2, the average SIR performance of the algorithms are shown with completely the same arrangement as that in Figure 5 but with the room reverberation length of 100 ms and the room reflection coefficient value of 0.6. The results show that the hybrid algorithm gives improvement in SIR of only above 10 dB due to increased room reverberation (reflection coefficient). For a comparison, the performance of ECoBLISS has also been shown.

6 Real room experimental results

For real room separation, the same set of speech signals used in the previous section was played out through loudspeakers for the LTI system. The microphones used were cardioids, and the arrangement used for the experiment is shown in Figure 9. The sampling frequency was 16 kHz, with a filter comprising 2,048 taps and an FFT size of 4,096 for the CFDD and 4,096 taps and an FFT size of 8,192 for the FDAFS to accommodate a room reverberation time of 250 ms.

For the SIR calculation, the impulse responses for the real room experiment were obtained using the swept frequency method [35]. The SIR given by the hybrid algorithm for the experimental arrangement shown in Figure 9 on all the three output streams was between 7 and 8 dB approximately and is summarised in Table 3. It was less than the performance in the simulated environment discussed above. The reason was because of the reverberation the real room environment lasted several hundred milliseconds; additionally, the distance of the

Table 3 Average SIR of hybrid algorithm for the experiment performed in the real room environment shown in Figure 9

		Time (s)									
		0 to 4	**4 to 8**	**8 to 12**	**12 to 16**	**16 to 20**	**20 to 24**	**24 to 28**	**28 to 32**	**32 to 36**	**36 to 40**
Average SIR (dB)	Input (streams)	−0.38	0.55	−2.40	−2.44	−1.15	−0.98	4.70	−7.01	0.59	−2.09
	CFDD	0.49	1.77	−0.72	−0.25	4.04	3.96	9.40	1.62	8.01	3.06
	Hybrid	1.43	3.60	1.82	2.42	6.01	5.94	12.55	3.7	9.34	5.74
	Improvement CFDD	0.87	1.21	1.67	2.19	5.19	4.94	4.70	8.63	7.41	5.15
	Improvement hybrid	1.81	3.04	4.22	4.87	7.16	6.92	7.85	10.71	8.74	7.83

The whole speech length has been divided into 4-s segments, and SIR is calculated on 20-ms syllable within that.

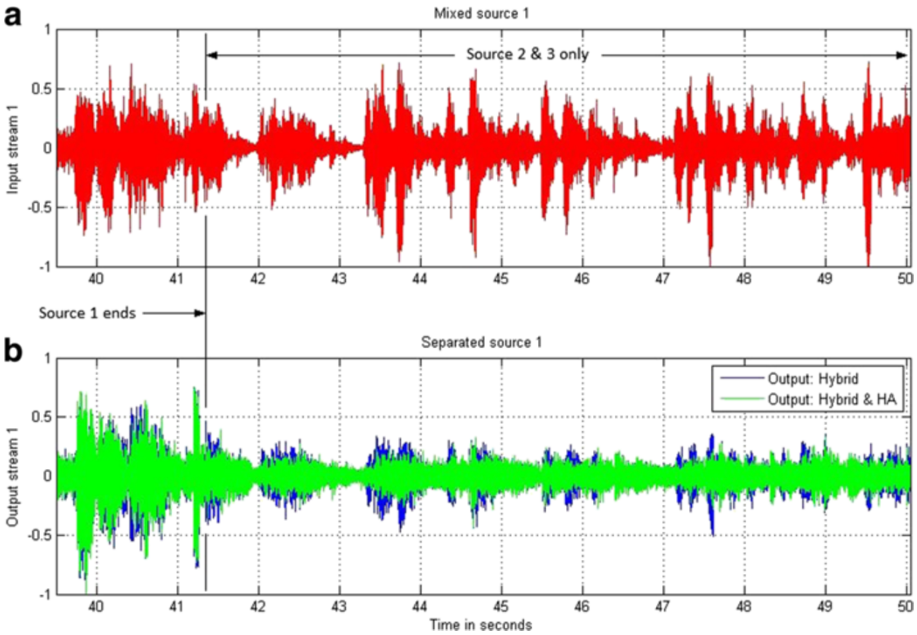

Figure 10 The separation performance of the algorithm in a real room environment as shown in Figure 9. Source 1 is being separated from the interference signal, i.e. sources 2 and 3 in output stream 1. An experimental section is shown in **(a)** in which the speech signal from source 1 ends before 42 s and only the interference signal is left behind. The suppression of the interference signal (sources 2 and 3) by the hybrid and the hybrid and HA can be seen **(b)**.

microphones from the loudspeakers was significant with very close room walls. This made the separation more challenging. The hybrid algorithm suppression of the interference signal can be seen in Figure 10.

On hearing the separated signals, the speech is intelligible and is without any distortion. Application of the harmonic alignment algorithm after the hybrid algorithm further improved the suppression on the formant part (harmonic part) of the interference signal. This improved the SIR of the signal only in the formant areas of the interference signal, as can be seen in Figure 10, thus rendering the interference signal completely unintelligible. This resulted in an intelligible separated signal with a less annoying unintelligible background interfering signal. Apart from

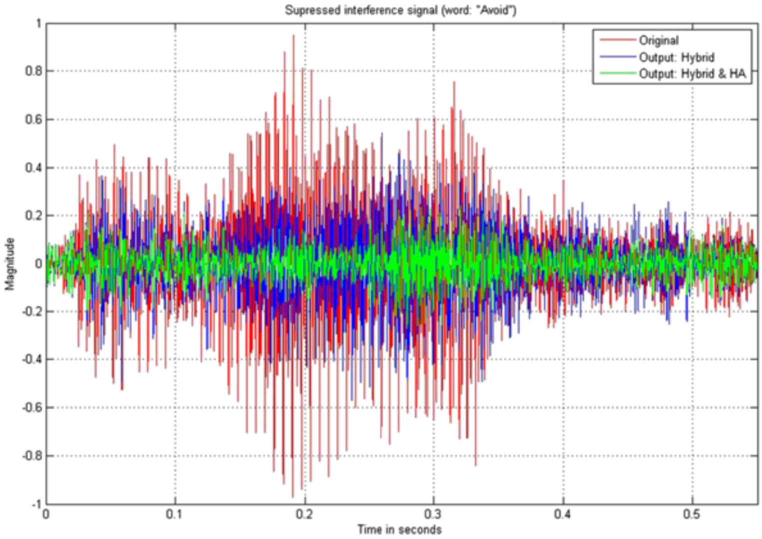

Figure 11 The improved suppression by using harmonic alignment in addition to the hybrid algorithm. The interference signal word 'Avoid' spanning 550 ms from one of the speakers is shown to be suppressed.

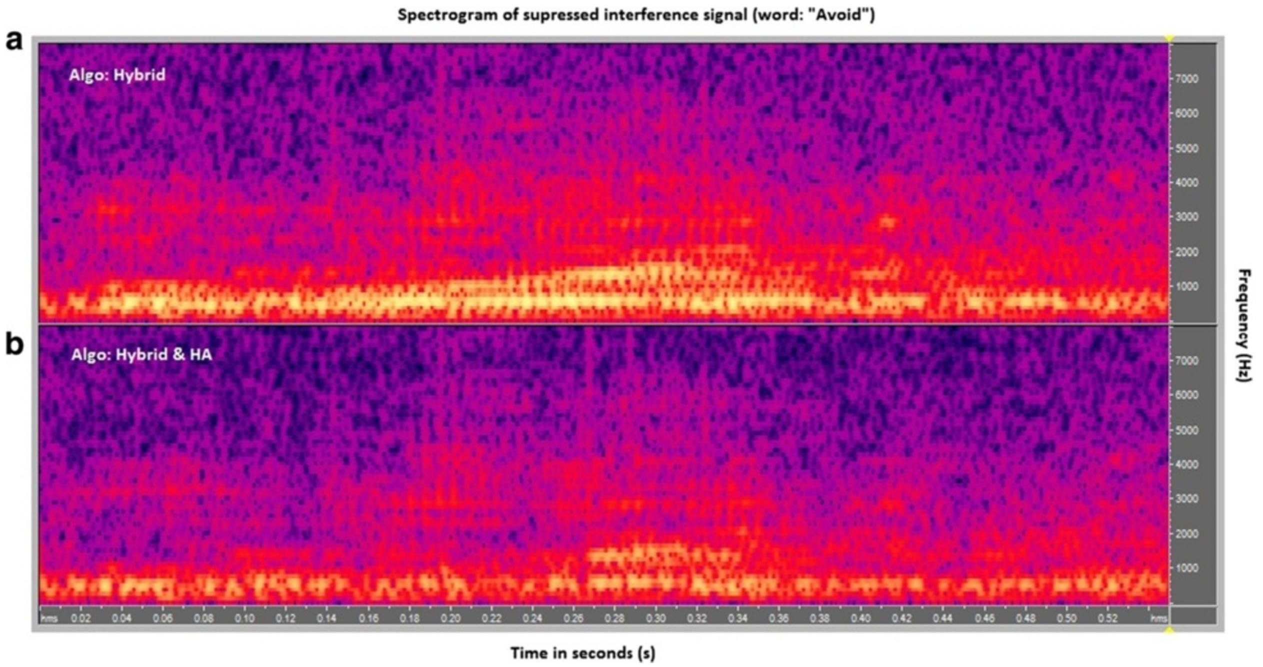

Figure 12 Spectrogram of the time domain signals shown in Figure 11. (a) Hybrid algorithm. **(b)** Hybrid and HA algorithm.

that, the background noise was negligible in the experiment, therefore not removed in a supervised way (explained in the next section).

The harmonic alignment-based improved interference suppression can also be seen in Figures 11 and 12 on a different section of the speech. For harmonic alignment, the overlap and add method was used with a syllable size of 20 ms and FFT size of 4,096. It is pertinent to mention here that the hybrid algorithm along with harmonic alignment only performs blind source separation and does not carry out any equalisation on the separated streams.

7 Discussion

The hybrid algorithm is well suited for real-time implementation due to its block-by-block FFT methodology. The first structure in the hybrid algorithm, the CFDD, updates only once in five blocks, thus reducing computational load. If N is the size of FFT and the CFDD's overlap and add method is running at optimum level with the filter length equivalent to half that of FFT, then the average computational cost over five blocks is $19.2 \times N\log_2 N + 1{,}976.2\,N$ operations. Similarly, the computational cost for the FDAFS structure is $19.2 \times N\log_2 N + 411\,N$ operations whereas its time domain equivalent TD-FB cost is $(18 \times N\log_2 N + 210)f_s$ operations. These computational costs are calculated based on multiplication accumulation (MAC) operation of a typical DSP board. It is important to mention that for complex multiplication, six operations have been considered. Similarly for division or exponential or square root, 20 operations have been considered depending upon if the Taylor series expansion or polynomial fit curve is used by a designer. For the FFT, the number of MAC operation considered

Table 4 Computational load in MMACS for different algorithms at sampling frequency of 16 and 48 kHz

		Sampling frequency						
		16,000 Hz			48,000 Hz			
Impulse length (ms)		64	128	256	21.33	42.66	85.33	170.66
Filter taps (*N*)		1,024	2,048	4,096	1,024	2,048	4,096	8,192
MMACS	CFDD	69.99	70.61	71.22	209.99	211.83	213.67	215.52
	FDAFS	30.04	31.58	33.12	90.14	94.75	99.36	103.96
	ECoBLISS	141.23	142.76	144.30	423.69	428.30	432.91	437.52
	HYBRID	100.03	102.19	104.34	300.13	306.58	313.03	319.48
	TD-FB	298.27	593.18	1183	894.81	1779.6	3549	7088

based on radix 2 implementation is $2 \times N\log_2 N$. The low computational load in terms of million multiplication-accumulations (MMACS) per second of the hybrid algorithm is evident from Table 4.

It can be seen from Table 4 that FDAFS is many times faster than its time domain equivalent TD-FB that is similar to frequency domain LMS as compared to its time domain equivalent (pp. 353 in [31]). The frequency domain implementation will have two types of latencies in the algorithm: the first will be the computational latency calculated from the MMACS of the algorithm divided by the MMACS capacity of the digital signal processor and the second is the time it takes to fill up the block for the FFT. Nothing can be done about the latter issue; however, regarding processor speed, this is continually advancing, with even non-FPGA-type devices routinely available with speeds of a few thousand MMACS.

In this paper we have solely discussed the main source separation algorithm without discussing background noise (not the additive sensor noise as taken in most cases). It has been shown (pp. 397–399 in [36]) that supervised adaptive filtering using a reference microphone to detect the noise source based on the least mean squares (LMS) technique gives the optimum performance in removing background noise. So, supervised adaptive filtering should be implemented prior to the use of an unsupervised hybrid algorithm instead of using acoustic echo cancellation (AEC) as shown in [13].

The convolutive mixture problem is complex, and *in extremis*, the whole source separation scenario becomes mathematically ill-posed (discussed earlier) and thus nothing works. The hybrid algorithm uses adaptive filtering methodology that is also unsupervised. So, in the case of *extremis*, multiple spurious minima can occur that result in either the algorithm taking longer to converge to a true minimum based on step size, or it may not converge at all. In such cases, in order to achieve separation (convergence to a true minimum), it is necessary to incorporate additional knowledge about the mixing process and/or the source signals; however, this would make the separation process semi-blind or supervised. Future investigations will focus on improving the robustness and applicability of the method.

8 Conclusions

In this paper we have presented a novel hybrid algorithm that uses an integrated, multiple conditions approach to solve the convolutive mixture problem of speech sources, instead of relying on only one condition. The performance of the algorithm based on experiments has been shown for simulated and real room environments. The proposed algorithm with its improved SIR using harmonic alignment and efficient computational complexity is suitable for hardware implementation for the real-time blind source separation of speech signals.

Competing interests

The authors declare that they have no competing interests.

Received: 23 August 2013 Accepted: 23 May 2014
Published: 17 June 2014

References

1. P Comon, Independent component analysis, a new concept? Signal Process. **36**, 287–314 (1994)
2. JF Cardoso, Infomax and maximum likelihood for blind source separation. IEEE Signal Process. Letter **4**, 109–111 (1997)
3. AJ Bell, TJ Sejnowiski, An information-maximization approach to blind separation and blind deconvolution. Neural Comput. **7**, 1129–1159 (1995)
4. B Salberg, N Grbic, I Claesson, Online maximization of subband kurtosis for blind adaptive beamforming in realtime speech extraction, in *IEEE, Proc. of the 2007 15 Intl. Conf. on Digital Signal Process* (Bleking Institute Technol, Ronneby, 2007)
5. AK Nandi, *Blind Estimation Using Higher Order Statistics* (Kluwer, London, 1999)
6. S Amari, A Cichocki, Adaptive blind signal processing - neural network approaches. Proc. IEEE **86**(10), 2026–2048 (1998)
7. S Amari, Natural gradient works efficiently in learning. Neural Comput. **10**, 251–276 (1998)
8. RH Lambert, Multichannel blind deconvolution: fir matrix algebra and separation of multipath mixtures, in PhD Thesis (University of Southern California, 1996)
9. S Haykin, *Unsupervised Adaptive Filtering*, vol. 1 (Wiley, New York, 2000)
10. H Sawada, R Mukai, S Araki, S Makino, A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Transon. Speech Audio Process. **12**(5), 530–538 (2004)
11. H Sawada, R Mukai, S Araki, S Makino, Solving the permutation problem of frequency-domain BSS when spatial aliasing occurs with wide sensor spacing, in *IEEE Conference on Acoustic Speech and Signal Processing, ICASSP 2006* (Toulouse, 2006), pp. V77–V80
12. B Peng, W Liu, DP Mandic, Reducing permutation error in subband-based convolutive blind separation. IET Signal Process. **6**(1), 34–44 (2012)
13. DWE Schobben, PCW Sommen, A frequency domain blind signal separation method based on de-correlation. IEEE Trans. Signal Process. **50**, 1855–1865 (2002)
14. S Araki, R Mukai, S Makino, T Nishikawa, H Saruwatari, The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. IEEE Trans. Speech Audio Process. **11**(2), 109–116 (2003)
15. E Hoffman, D Vicente, R Orglmeister, Time frequency masking strategy for blind source separation of acoustic signals based on optimally-modified log-spectral amplitude estimator. LNCS **5441**, 581–588 (2009)
16. MS Pedersen, J Larsen, U Kjems, LC Parra, A survey of convolutive blind source separation methods, in *Springer Handbook on Speech Processing and Speech Communication* (Springer, Berlin, 2007), pp. 1–34
17. SN Jain, C Rai, Blind source separation and ICA techniques: a review. IJEST **4**(4), 1490–1503 (2012)
18. WK Ma, TH Hsieh, CY Chi, DOA estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: a Khatri–Rao subspace approach. IEEE Trans. Signal Process. **58**(4), 2168–2180 (2010)
19. P Pertila, Online blind speech separation using multiple acoustic speaker tracking and time–frequency masking. Elsevier Comput. Speech Lang. **27**, 683–702 (2013)
20. C Blandin, A Ozerov, E Vincent, Multi-source TDOA estimation in reverberant audio using angular spectra and clustering. Elsevier Signal Process. **92**, 1950–1960 (2012)
21. RK Miranda, SK Sahoo, R Zelenovsky, CL Da Costa, Improved frequency domain blind source separation for audio signals via direction of arrival knowledge, in *Society for Design and Process Science, SPDS 2013* (Sao Paulo, 2013), pp. 35–39
22. T Katayama, T Ishibashi, A real-time blind source separation for speech signals based on the orthogonalization of the joint distribution of the observed signals, in *IEEE/SICE International Symposium on System Integration (SII)* (Kyoto, 2011), pp. 920–925
23. Y Na, J Yu, B Chai, Independent vector analysis using subband and subspace nonlinearity. EURASIP J. Adv. Signal Process. **2013**, 1 (2013)
24. K Kokkinakis, V Zarzoso, AK Nandi, Blind separation of acoustic mixtures based on linear prediction analysis, in *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)* (Nara, 2003), pp. 343–348

25. A Ozerov, E Vincent, F Bimbot, A general flexible framework for the handling of prior information in audio source separation. IEEE Trans. Audio Speech Lang. Process. **20**(4), 1118–1133 (2012)
26. R Aichner, H Buchner, F Yan, W Kellermann, Real-time convolutive blind source separation based on a broadband approach, in *Independent Component Analysis and Blind Signal Separation*, ed. by CG Puntonet, A Prieto. Fifth International Conference, ICA 2004, Granada, Spain, September 22–24, 2004. Lecture notes in computer science, vol. 3195 (Springer, Berlin, 2004), pp. 840–848
27. A Belouchrani, K Abed-Meraim, JF Cardoso, E Moulines, A blind source separation technique using second-order statistics. IEEE Trans. Signal Process. **45**(2), 434–444 (1997)
28. DWE Schobben, PCW Sommen, On the indeterminancies of convolutive blind signal separation based on second order statistics, in *ISSPA 99* (Brisbane, 1999), pp. 215–218
29. M Joho, H Mathis, Joint diagonalisation of correlation matrices by using gradient methods with application to blind signal separation, in *SAM 2002* (Rosslyn, 2002), pp. 273–277
30. L Parra, C Spence, Convolutive blind separation of non-stationary sources. IEEE Trans. Signal Process. **8**(3), 320–327 (2000)
31. S Haykin, T Kailath, *Adaptive Filter Theory*, 4th edn. (Pearson Education, Upper Saddle River, 2007)
32. JC Bellamy, *Digital Telephony*, 3rd edn. (Wiley, New York, 2000)
33. SG McGovern, A model for room acoustics. (2004). http://www.sgm-audio.com/research/rir/rir.html. Accessed 21 Jan 2013
34. E Vincent, R Gribonval, C Fevotte, Performance measurement in blind audio source separation. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1462–1469 (2006)
35. M Holters, T Corbach, U Zoelzer, Impulse response measurement techniques and their applicability in the real world, in *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09)* (Como, 2009)
36. P Gaydeck, *A Foundation of Digital Signal Processing: Theory, Algorithms and Hardware Design* (IEE, London, 2004)

doi:10.1186/1687-6180-2014-92
Cite this article as: Minhas and Gaydecki: **A hybrid algorithm for blind source separation of a convolutive mixture of three speech sources.** *EURASIP Journal on Advances in Signal Processing* 2014 **2014**:92.