

Noise Reduction using DUET algorithm for dual-microphone mobile station

Aoulass Nabila¹, 2nd Chakkor Otman,
{ nabilaoulass@gmail.com¹, o.chakkor@gmail.com² }

signal and image processing, university Abdlmalek Esaadi ENSAT laboratory NTT¹, telecommunication systems and networks engineering, university Abdlmalek Esaadi ENSAT²

Abstract. This paper presents a novel technique which can improve the quality of signal used for mobile station using dual-microphone. Our method is based on the degenerate unmixing estimation technique (DUET). This technique can separate any number of sources using only two mixtures. The DUET algorithm is used as a tool to denoise a received signals. Multiple sources are acquired by two sensors. The received signals are called mixtures, which they are processed in order to recover only one source signal, whereas others are considered as noise.

Keywords: Blind Separation of sources, Degenerate Unmixing Estimation Technique, histogram, time-frequency masking.

1 Introduction

Blind source separation (BSS) consists of estimating a set of latent signals from a set of observed signals, called mixtures. Typically, the observations are obtained at the output of a set of sensors, where each sensor receives a different combination of the source signals. The adjective "blind" highlights two facts [1]:

- The source signals are not observed.
- No information is available on the mixing system.

The BSS is a very active research subject because of its many practical applications. In the field of audio signal separation, we find the "cocktail-party" problem, that we encounter when we record several speakers who speak simultaneously, and that we want to isolate a speaker. Radar and sonar when signals from multiple transmitters interfere at the receiver antenna, and in many other areas. The lack of prior knowledge of the source signals and the mixing system is always compensated by assumptions that unknown sources should answer [1].

Some common assumptions are that the sources must be statistically independent. The implementation of ICA algorithm is the Fast Independent Component Analysis (FastICA). The main problem using FastICA is that it assumes instantaneous mixing without time delay in the recording process.

As we know, many techniques based on ICA and among the recently proposed methods are found the DUET (Degenerate Unmixing Technical Estimation) which is one of the demixing

algorithms in the field of BSS. It can separate any number of sources using only two mixtures, such this method is based on W-disjoint orthogonal sources.

There are several algorithms for signal separation that use similar or even completely different concepts.

DUET is a solution for separating for signals using two microphones located in a mobile station (MS). This method is suited for audio source separation. To realize this separation we will implement the DUET solution for speech signals in real environment with a noisy background.

The DUET algorithm is effective for sources with parsimony property in the time-frequency domain, such as the voice signal, that is, the target speech signal in a noisy environment can be effectively recognized. However, in many cases, there are more sources than mixtures, so we refer to a case as degenerate [5]. In the degeneration of the blind source, separation poses a challenge because the mixing matrix is not reversible.

Basically, the traditional method such as the ICA of the demixing by estimating the inverse mixture matrix of which the DUET algorithm operates in the frequency domain, no inverse matrix is calculated and this is one of the reasons why it shows very good performance. Another advantage over ICA is that the number of sources can be greater than the number of mixtures, in fact it can be used for an arbitrary number of sources.

ICA methods have several drawbacks. Often, it is required that the number of source signals is known in advance, which is not our case.

When a mobile user talked in a public place e.g. office, markets, universities, a speech noise is created. Then, the corrupted speech needs to enhance in a minimum quality so that the user in opposite side can understand. Speech enhancement algorithms are, therefore, developed and improved depending on cost effectiveness and maintaining the minimum quality multiple unwanted speakers.

2 Degenerate Unmixing Estimation Technique

2.1 Introduction

The Degenerate Unmixing Estimation Technique (DUET) is an algorithm that allows us to blindly separate an arbitrary number of sources with only two sensors. The received signals are called mixtures that provide the time-frequency representations of the sources. DUET assumes the sources are already separate in the time-frequency plane.

The sources are disjoint they do not overlap too much, which is ideal for speech [5]. This method applied when the sources are W-disjoint orthogonal, when the time-frequency representations of any two signals in the mixtures are disjoint sets.

The principle of this method is to perform a search for amplitude-Delay mix parameters, and simultaneously construct time-frequency bit masks that are used to partition one of the mixtures to recover the original source signals [5].

The blind separation of the audio signals using the DUET algorithm requires following a set of successive steps in order to estimate the sources:

2.2 Creation of mixing model

Firstly we need to establish a model to describe the mixed signals. In our case, we have two microphones that can capture the k sources that are represented by a number of people who speak in the scene, are supposed to be on different locations.

The mixed signals x_1 and x_2 can be described as follows:

$$x_1(n) = \sum_{j=1}^k s_j(n) \quad (1)$$

$$x_2(n) = \sum_{j=1}^k a_j(n) \cdot s_j(n - \delta_j) \quad (2)$$

where a_j is a relative attenuation factor, and δ_j is the arrival delay between the sensors.

2.3 Hamming window and FFT

In order to avoid spectral leakage, a Hamming window can be applied to a signal in the frequency domain[7]. The Hamming window has relatively low values, close to zero, at the ends and values up to one, in the center of the window.

When the Hamming window is applied to a signal, the amplitudes of the frequencies in the highest and lowest part of the frequency spectrum are attenuated[7]. The amplitude of the frequencies in the middle of the spectrum remains essentially the same[7].

The x_1 and x_2 mixtures are sampled and divided into blocks of length N with overlap. These sample blocks are multiplied with the Hamming W windowing function and then the Fast Fourier transform is applied.

$$x_{1,2}^\theta(n) = W(n)x_{1,2}(n) \quad (3)$$

$$X_{1,2}(f) = \sum_{n=0}^{N-1} x_{1,2}^\theta(n) \cdot e^{-\frac{i2\pi n f}{N}} \quad (4)$$

2.4 Estimation of amplitude-delay parameter

So if we will refer to the Fourier transform of sources $S_j(f)$ For a given source j we can describe their FFT as follows:

$$\begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} = \begin{bmatrix} 1 \\ a_j e^{-i2\pi f \delta_j} \end{bmatrix} S_j(f) \quad (5)$$

The DUET algorithm is based on the property of W-disjoint orthogonality between the sources, and each source has a sparse frequency spectrum for a given time compared to the other such that[8]:

$$S_j(f) \cdot S_i(f) = 0, \quad \forall i, j \quad i \neq j \quad (6)$$

They then assume that the sources are orthogonal W-disjoint in time-frequency plane associated with the transform considered with the window W[8]. This strong hypothesis means that the sources have no overlap in the time frequency domain. The associated amplitude and delay to a source are then given by:

Where:

$$a_j, \delta_j = \left(\left\| \frac{X_2(f)}{X_1(f)} \right\|, \text{Im}(\log(\left\| \frac{X_2(f)}{X_1(f)} \right\|)) / 2\pi f \right) \quad (7)$$

$$\frac{X_2(f)}{X_1(f)} = a_j e^{-i2\pi f \delta_j} \quad (8)$$

After we added the time dimension, and then realize a time-frequency study, leading to: Using the hypothesis that the sources are Wdisjoint-orthogonal, and performs an amplitude/delay histogram of all the values of $a_j; \delta_j$ encountered of which this histogram has peaks around the values corresponding to the parameters of the mixing matrix seen in the time-frequency plane.

2.5 Separation via binary masking and IFFT

This step consists in generating a bit mask for each pair of parameters (a_j, δ_j) of mixture corresponding to the time-frequency points which produce particular mixing parameters then the sources are separated by multiplying each mask by a mixtures[8].

The mask is defined as:

$$M_j(t, f) = \begin{cases} 1 & \text{si } J(t, f) = j \\ 0 & \text{elsewhere} \end{cases} \quad (9)$$

As J is the objective function obtained by maximum likelihood ML.

Essentially, we assign each time-frequency point to the pair of mix parameters, which is the best explains the mixes at that particular time-frequency point. We denoise the signals via masking and the ML combination:

$$J(t, f) = \text{ArgMin}_j \frac{|a_j e^{-i2\pi f \delta_j} X_1(t, f) - X_2(t, f)|^2}{1 + a_j^2} \quad (10)$$

We can then reconstruct the sources from their time-frequency representations by converting them back to the time domain using the inverse Fast Fourier transform (IFFT).

3. Simulation

We implement blind signal reduction using DUET algorithm. We use two mixed signals x_1 and x_2 from five sources. It means that the mobile station use two microphones, a first one is used for a short acoustic path between mouth, which ensures less reverberation and high direct path power. A second microphone is placed on the back of the mobile station in order to capture the speech signal with a lower sound pressure level.

Steps to proceed for noise reduction are summarized as follows:

- **Calculation of FFT and Hamming window:**

The first step is to apply FFT and then build a time frequency representation of the two mixtures, so we are interested in signals of limited length, and a real signal can only have a limited duration. This can only be done on a finite number of points using the hamming window

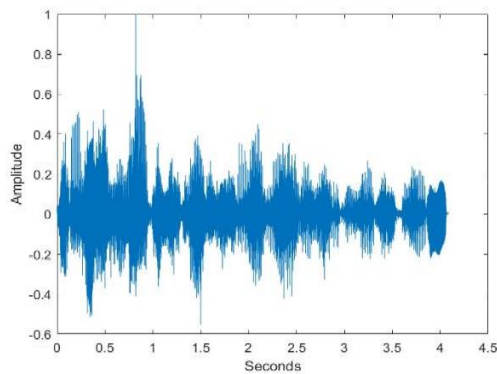


Fig. 1. Ar Mixed signal x_1 .

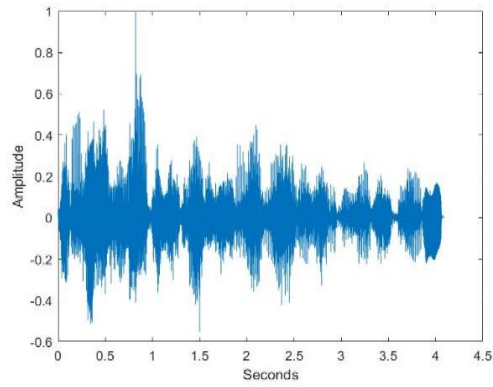


Fig. 2. Mixed signal x_2 .

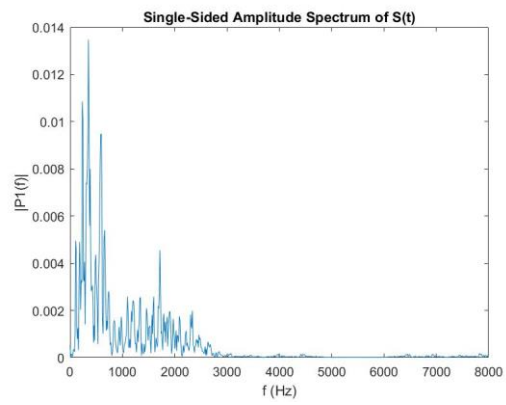


Fig. 3. x_1 after FFT a short term.

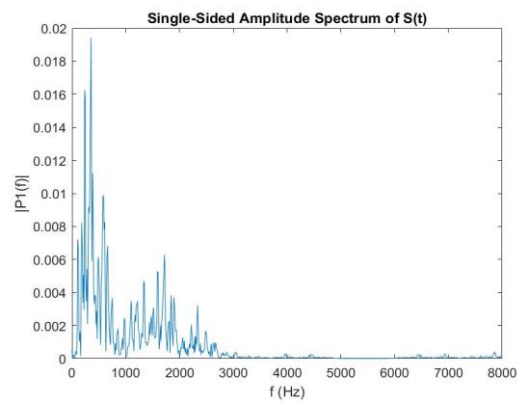


Fig. 4 . x_2 after FFT a short term.

- **Calculation of attenuation parameters and the weighted histogram:**

For each time-frequency point, the phase and amplitude will be calculated, then we build a 2D histogram (one dimension is the phase, the other is the amplitude). This histogram present speaks around the values corresponding to the parameters of the mixing matrix in the time-frequency plane. The parameters p and q represent the weight of the histogram. The histogram in the figure 5 shows five peaks associated with the two mixtures (x_1 and x_2). The peaks are exactly the pairs of mixing parameters for each of the five sources. Using the calculated symmetric attenuation $\alpha_j = a_j - 1/a_j$: It is weighted in terms of two-dimensional cross power ($p = 1, q = 0$), and pairs of two mixtures (x_1 and x_2) of five sources.

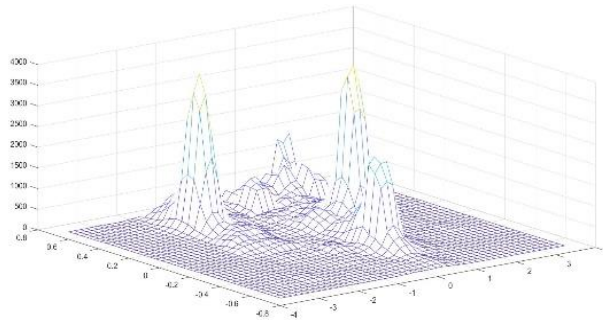


Fig. 5. calculation of histogram

- **Separation of sources:**

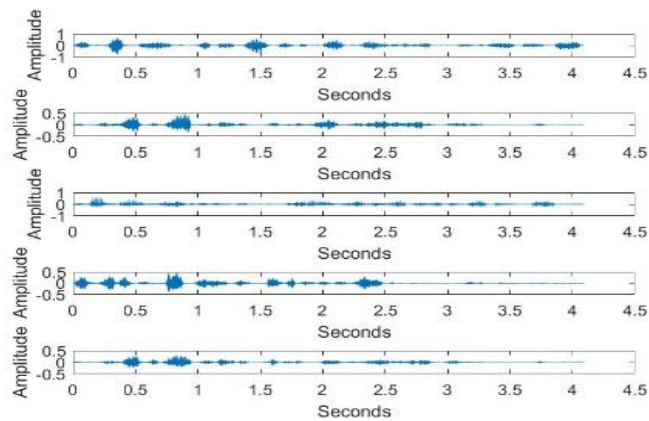


Fig. 6. estimated sources

We generate a binary mask for each pair of mixing parameters at time-frequency points, which produce special mixing parameters, and then separate the sources by multiplying each mask by one of the mixtures, then we apply IFFT to represent the estimated sources in the time domain, figure 6.

3. Conclusion

We propose a solution for noise reduction for speech enhancement in dual-microphone mobile phones. The solution is based on the DUET blind signal separation. The principle of DUET is based on the W-disjoint orthogonal assumption of source signals, which is verified for mobile station conversations. The simulations with recorded signals shows a good performance with low complexity of the system, making it possible for an easy integration in to mobile stations.

References

- [1] K. Itoh and M. Mizushima. Environmental noise reduction based on speech/non-speech identification for hearing aids. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 1, pp. 419-422, 1997.
- [2] Zbynek Koldovsky, Petr Tichavsky, David Botka. Noise reduction in dual-microphone mobile phones using a bank of pre-measured target-cancellation filters. Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 26-31 May 2013.
- [3] Alexander A S Gunawana, Albert Stevelino, Heri Ngarianto, Widodo Budiharto, Rini Wongsob. Implementation of Blind Speech Separation for Intelligent Humanoid Robot using DUET Method. Procedia Computer Science, Volume 116, 2017, Pages 87-98.
- [4] BELL, A. J., SEJNOWSKI, T. J. An information maximization approach to blind separation and blind deconvolution. Neural Computation, 1995, pp. 1129-1159.
- [5] CARDOSO, J.-F. Blind signal separation: statistical principles. Proceedings of the IEEE, 1998, vol. 86, no. 10, pp. 2009-2025.

- [6] M. Billette Analyse en composantes indépendantes avec une matrice de mélange sparse. Université de Montréal des études supérieures en vue de l'obtention du grade de Maîtrise en sciences (M.Sc.) en Statistique, 20 juin 2013.
- [7] Yang Wang, Örgür Yılmaz, Zhengfang Zhou. Phase aliasing correction for robust blind source separation using DUET. IEEE transaction on signal processing.
- [8] Scott Rickard, S. Makino et al. The DUET blind source separation algorithm in Blind speech separation, Springer, pp. 217-241, 2007.
- [9] Alexander Jourjini, Scott Rickard, Örgür Yılmaz, Blind source separation of disjoint orthogonal signals: Demixing N sources from two mixtures". Proc ICASSP2000, vol. 15, pp. 2985-2988, June 5-9, 2000.
- [10] Scott Rickard, The DUET Blind Source Separation Algorithm. University College Dublin 2007 Springer.
- [11] N. Yousefian and P. C. Loizou, A dual-microphone algorithm that can cope with competing-talker scenarios, IEEE Trans. on Audio, Speech, and Language Processing, vol. 21, no. 1, pp. 143-153, 2013.