Speech Enhancement using DUET & Weiner Filter Technique

[1]Aiswarya Venugopal [2]Anjana M R [3]Ashna Hasif [4]Delna Varghese [5]Prajeesh P A
[1,2,3,4]Student [5]Assistant Professor
[1,2,3,4,5]Department of Electronics and Communication Engineering
[1,2,3,4,5]Adi Shankara Institute of Engineering and Technology, Kalady, India

Abstract

Speech enhancement in degenerate mixtures is a challenging task for the signal processing engineers. Degenerate Unmixing Estimation Technique (DUET) is a commonly used algorithm in speech signal separation. This approach of DUET can be used in auditory training for individuals with hearing loss use their residual hearing maximally and can be implemented in home studio for music recording. The method is valid when sources are W-disjoint orthogonal, that is, when the supports of the windowed Fourier transform of the signals in the mixture are disjoint. For anechoic mixtures of attenuated and delayed sources, the method allows one to estimate the mixing parameters by clustering relative attenuation-delay pairs extracted from the ratios of the time–frequency representations of the mixtures. The estimates of the mixing parameters are then used to partition the time–frequency representation of one mixture to recover the original sources. The technique is valid even in the case when the number of sources is larger than the number of mixtures. The algorithm is coded and implemented using matlab software. The wiener filter is used to enhance the desired speech signal from the original signal in which the noises are present. Weiner filter minimizes the mean square error between the estimated random process and the desired process.
Keyword- Weiner Filter, DUET Blind Source Separation, Anechoic Mixtures and Matlab

I. INTRODUCTION

Speech Enhancement aims to improve speech quality by using various algorithms. The objective of enhancement is improvement in intelligibility or overall perceptual quality of degraded speech signal. Noise estimation is the major component in speech enhancement techniques, because better noise estimation gives a high quality of speech extraction. In this paper we propose a technique based on Weiner filtering under uncertainty of signal presence in the noisy observation and DUET is a blind source separation technique capable of the separation of N sources from 2 mixtures [3]. The Wiener filter is to compute a statistical estimate of an unknown signal using a related signal as an input and filtering that known signal to produce the estimate as an output. The known signal might consist of an unknown signal of interest that has been corrupted by additive noise.

The Wiener filter can be used to filter out the noise from the corrupted signal to provide an estimate of the underlying signal of interest. The Wiener filter is based on a statistical approach, and a more statistical account of the theory is given in the minimum mean square error (MMSE) estimator article.

In the field of blind source separation (BSS), assumptions on the statistical properties of the sources usually provide a basis for the demixing algorithm. Some common assumptions are that the sources are statistically independent are statistically orthogonal are non-stationary or can be generated by finite dimensional model spaces. The independence and orthogonality assumptions can be verified experimentally for speech signals. Some of these methods work well for instantaneous demixing, but fail if propagation delays are present [2]. One of the problems that DUET suffers is introduction of musical noise in separated speech signal due to T-F binaural masking or discontinuous zero padding in estimated signals and it produces nonlinear speech distortion, hence produces low quality speech. This problem can be solved by using soft mask or relatively smaller frame shifts [1].

Thus, the method presented in this paper utilizes an anechoic mixing representation and also employs the Fast Fourier Transform (FFT).DUET technique which is used in this study is based on BSS and confirms that number of mixtures is less than the number of sources. This study uses a version of DUET which can be executed and is implemented in Matlab. As speech enhancement requires the preservation of shape and characteristics of the original signal, linear correlation coefficient as statistical measurement and histograms are used. All these measurements quantify the objective speech quality by reducing artifacts in the enhanced speech. Assumptions for DUET technique are described at the following subsections.

A. DUET Blind source separation

DUET, the Degenerate Unmixing Estimation Technique, solves the degenerate demixing problem in an efficient and robust manner. It is possible to blindly separate an arbitrary number of sources given just two anechoic mixtures provided the time–frequency representations of the sources do not overlap too much, which is true for speech. The way that DUET separates

All rights reserved by www.grdjournals.com

degenerate mixtures is by partitioning the time–frequency representation of one of the mixtures. In other words, DUET assumes the sources are already 'separate' in that, in the time–frequency plane, the sources are disjoint [1]. The 'demixing' process is then simply a partitioning of the time–frequency plane. Although the assumption of disjointness may seem unreasonable for simultaneous speech, it is approximately true. By approximately, we mean that the time–frequency points which contain significant contributions to the average energy of the mixture are very likely to be dominated by a contribution from only one source. Stated another way, two people rarely excite the same frequency at the same time.

1) Anechoic Mixing

First, anechoic mixing of speech signal and two types of noise signals (speech babble and competing speech) are considered. The noise signals mix with the target speech signals separately. It is considered that mixtures of N speech source signals, sj(t), j = 1, . . . , N, being received at a pair of microphones where only the direct path is present. In this case, without loss of generality, it can absorb the attenuation and delay parameters of the first mixture, x1(t), into the definition of the sources. The two anechoic mixtures can therefore be expressed as in

$$x_1(t) = \sum_{j=1}^{N} s_j(t),$$

$$x_2(t) = \sum_{j=1}^{N} a_j s_j(t - \delta_j),$$

where, N is the number of sources, δj is the arrival delay between the sensors, and aj is a relative attenuation factor corresponding to the ratio of the attenuations of the paths between sources and sensors[1].The DUET method, which is based on the anechoic model, has proven to be quite robust even when applied to echoic mixtures

B. Weiner Filter

The idea behind Wiener filtering is to emphasize the frequencies where speech signal is dominant over the noise signal, and to attenuate the frequencies where the speech signal is weak relative to noise. Here, we propose a technique based on a Wiener filtering under uncertainty of signal presence in the noisy observation. Two different estimators of the a priori signal-to-noise ratio are tested and compared. The main interest of this approach comes from its low complexity. Weiner filter is used for Speech enhancement / Noise cancellation and suppression by a convex combination of two DD approaches with Minimum mean squared error (MMSE) to estimate desired speech signal [4].

The Wiener filter is based on a statistical approach, and a more statistical account of the theory is given in the minimum mean square error (MMSE) estimator article.

The mean square error (MSE) may be rewritten as:

$$\begin{aligned} E\left[e^2[n]\right] &= E\left[(x[n] - s[n])^2\right] \\ &= E\left[x^2[n]\right] + E\left[s^2[n]\right] - 2E[x[n]s[n]] \\ &= E\left[\left(\sum_{i=0}^{N} a_i w[n-i]\right)^2\right] + E\left[s^2[n]\right] - 2E\left[\sum_{i=0}^{N} a_i w[n-i]s[n]\right] \end{aligned}$$

II. METHODOLOGY

The block diagram for the DUET is given in Fig.1

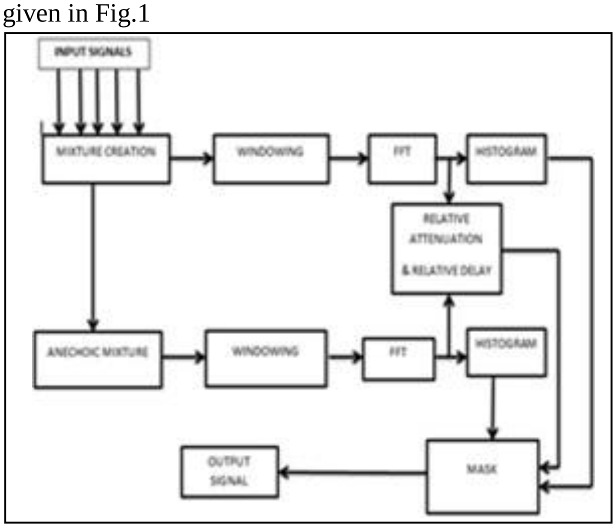

Fig. 1: Block diagram of DUET system

All rights reserved by *www.grdjournals.com*

The first block in Fig. 1 signifies the anechoic mixing which is a legal assumption for DUET technique and already described earlier. The mixed signals are then recorded as sound files. For analysis, Hamming window is used with window length of 1024. Different signals are recorded. First we are taking the DFT of the recorded signals and it is given to mixture. After windowing we are estimating the time delay and attenuation for anechoic environment and Pre-processing is done. Form a 2 dimensional histogram of the amplitude delay estimate. Next step is the Masking. It consists to estimate the masks using the time-frequency representations of the resultant estimated speech signals. DUET two-dimensional cross power weighted histogram of symmetric attenuation and delay estimate pairs from two mixtures of five sources. Each peak corresponds to one source and the peak locations reveal the source mixing parameters. Finally take the inverse fourier transform. Construct time–frequency representations from mixtures. Construct 2D smoothed weighted histogram. Locate peaks and peak centers which determine the mixing parameter estimates. Construct time–frequency binary masks for each peak center. Apply each mask to the appropriately aligned mixtures. Convert each estimated source time–frequency representation back into the time domain.

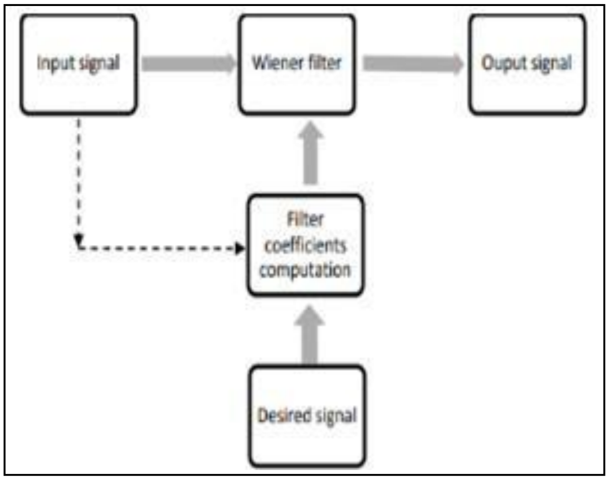

Fig. 2: Noise cancellation system using weiner filter

III. RESULTS AND DISCUSSION

DUET assumes that the source signals are disjoint in the time–frequency domain (W-disjoint orthogonal) and exploits the fact that the ratio of the time–frequency representations of the mixtures can be used to partition the mixtures into the original sources. The key construct in DUET is the two-dimensional smoothed weighted histogram which is used to cluster the mixing parameter estimates. By assuming an anechoic mixing model, all time–frequency points provide input to the histogram as we can eliminate the frequency variable and extract the delay. The fact that both estimation and separation can be done when the number of sources is larger than the number of mixtures without significant computational complexity, as is demonstrated by the Matlab code.

The Wiener filter is a linear estimator and minimizes the mean-squared error between the original and enhanced speech. The algorithm is implemented in the frequency domain and depends on the filter transfer function from sample to sample based on the speech signal statistics; the local mean and the local variance. For the noise estimation, the recursive noise estimation approach is used.

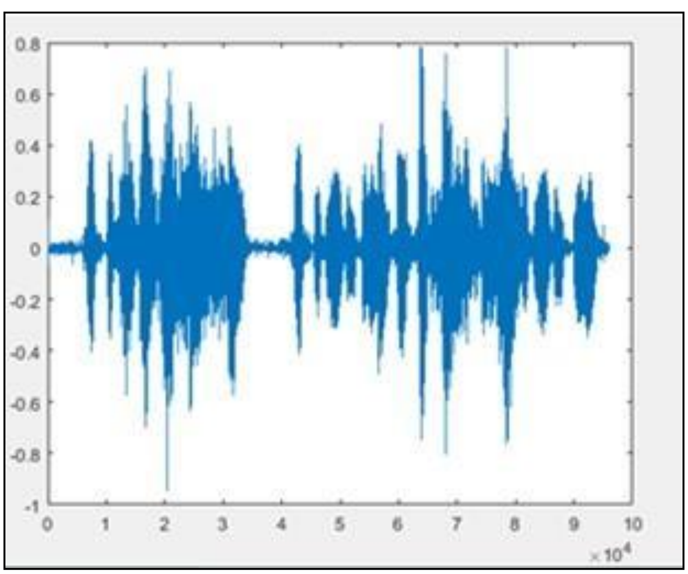

Fig. 3: First mixture output

All rights reserved by www.grdjournals.com

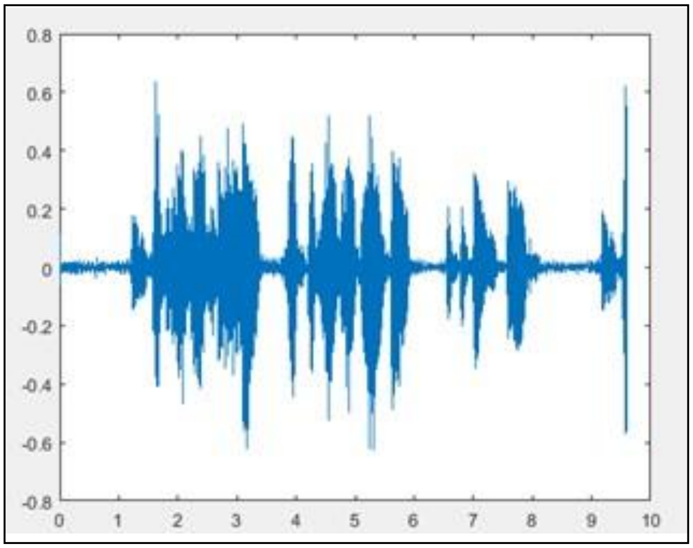

Fig. 4: Anechoic mixture output

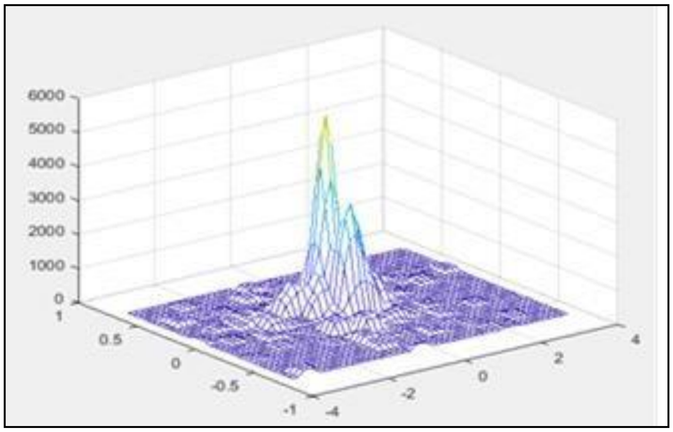

Fig. 5: Histogram analysis

REFERENCES

[1] Abrar Hussain, Kalaivani Chellappan and Siti Zamratol M," Speech Enhancement using Degenerate Unmixing Estimation Technique and Adaptive Noise Cancellation Technique as a Post Signal Processing," IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 978-1-4673-7791-1/16/$31.00 ©2016 IEEE.
[2] Scott Rickard, "The DUET Blind Source Separation Algorithm", S. Makino et al. (eds.), Blind Speech Separation, 217–241. © 2007 Springer.
[3] Niall Hurley,Naomi Harte,Conor Fearon and Scott Rickard," Blind Source Separation of Speech in Hardware", 0-7803-9333-3/05/$20.00 ©2005 IEEE.
[4] Zhang Wenlu and Peng Hua", Modified Wiener Filtering Speech Enhancement Algorithm with Phase Spectrum Compensation", 9th IEEE International Conference on Communication Software and Networks, 978-1-5090-3822-0/17/$31.00 ©2017 IEEE.

All rights reserved by *www.grdjournals.com*