

## High clarity speech separation using synchroextracting transform

Shahin M. Abdulla<sup>1</sup>, J. Jayakumari<sup>2</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Noorul Islam College of Engineering, Nagercoil, TN, India

<sup>2</sup>Department of Electronics and Communication Engineering, Mar Baselios College of Engineering and Technology, Kerala, India

---

### Article Info

#### Article history:

Received Jul 19, 2020

Revised Dec 24, 2020

Accepted Jan 7, 2021

---

#### Keywords:

Reassignment

Short time Fourier transform

Synchroextracting transform

Synchrosqueezing transform

Time-frequency representation

---

### ABSTRACT

Degenerate unmixing estimation technique (DUET) is the most ideal blind source separation (BSS) method for underdetermined conditions with number of sources exceeds number of mixtures. Estimation of mixing parameters which is the most critical step in the DUET algorithm is developed based on the characteristic feature of sparseness of speech signals in time frequency (TF) domain. Hence, DUET relies on the clarity of time frequency representation (TFR) and even the slightest interference in the TF plane will be detrimental to the unmixing performance. In conventional DUET algorithm, short time Fourier transform (STFT) is utilized for extracting the TFR of speech signals. However, STFT can provide on limited sharpness to the TFR due to its inherent conceptual limitations, which worsens under noise contamination. This paper presents the application of post-processing techniques like synchrosqueezed transform (SST) and synchroextracting transform (SET) to the DUET algorithm, to improve the TF resolution. The performance enhancement is evaluated both qualitatively and quantitatively by visual inspection, Renyi entropy of TFR and objective measures of speech signals. The results show enhancement in TF resolution and high clarity signal reconstruction. The method also provides adequate robustness to noise contamination.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

### Corresponding Author:

Shahin M. Abdulla

Department of Electronics and Communication Engineering

Noorul Islam College of Engineering

Nagercoil, Tamilnadu, India

Email: shahinmabdulla@gmail.com

---

## 1. INTRODUCTION

The under-determined BSS method called sparse component analysis has been widely used in audio source separation in the current scenario. This involves transforming audio mixtures into a sparse domain where manipulation and separation of these mixtures becomes easier. Various techniques have been used for audio source separation in sparse domain. The first approach on this direction was proposed by [1, 2]. These papers demonstrated that mixtures of source signals could be separated without knowledge of underlying source signals or mixing procedure. These methods assumed an instantaneous mixing procedure and a scenario where number of mixtures exceeds number of sources. Speech separation in underdetermined cases is difficult as it does not have a linear solution. The first efforts in this direction are presented in [3, 4]. The first practical algorithm for separation of arbitrary number of speech signals from two anechoic mixtures was initially proposed by [5] and further explored by [6] and is known as DUET algorithm.

DUET algorithm works well for convolutive mixtures. The speech mixtures are first converted to TF domain using STFT where speech is assumed to be sparse. Then it partitions the time-frequency domain into regions corresponding to individual sources. The region for separation depends on the closeness of TF coefficients to the estimated delay and amplitude parameters and each source is then demixed by synthesising the estimated coefficients in the region. This technique mainly relies on the correct estimation of amplitude and delay parameters corresponding to each source which in turn depends on correct estimation of TF coefficients. However the TF resolutions are restricted by Heisenbergs uncertainty principle which limits how accurately time varying information can be captured over short time intervals. This results in ‘blurring’ or ‘smearing out’ of TFR regardless of the analysis tool used which leads to wrong estimation of TF coefficients and hence reducing clarity of separated speech.

Many researches has been carried out in designing high resolution TF techniques, at the same time retain their invertible ability to recover original time series signal. Usually STFT, Vigner-Ville distribution, Wavelet are used to convert speech to TF domain. However the TF resolutions of these transforms are poor.

STFT converts a one dimensional time series signal into two-dimensional TFR where we can see both the time and frequency of the signal. However band limited window function is used in STFT which causes energy blurred spectrogram. Various post processing techniques on STFT have been proposed to improve TF resolutions. This include reassignment method (RM) [7], synchrosqueezing transform (SST) [8-10], parametric time frequency analysis (PTFA) method [11-14] and demodulated time frequency analysis (DTFA) [9, 15]. The ultimate aim of these methods is to improve TF resolution by developing an ideal time frequency analysis (ITFA) method [16] and to obtain an ideal TF representation (ITFR) which is of the form:

$$ITFR(t, \omega) = \sum_{k=1}^n A_k(t) \cdot \delta(\omega - \varphi'_k(t)) e^{i\varphi_k(t)} \quad (1)$$

where  $A_k(t)$  is the time varying amplitude,  $\delta$  is the Dirac distribution function, and  $\varphi(t)$  is the time varying phase of the signal.  $\varphi'_k(t)$  is the derivative of  $\varphi(t)$  and is the instantaneous frequency (IF). We know that an ITFR is the one producing an impulse at IF of signal and elsewhere it is zero. This can be achieved by squeezing or reassigning the TF coefficients, so that signal energy only appears in IF trajectory, which results in good time frequency resolution and anti-noise property.

Though RM gives sharper TFR for speech mixtures it is based on absolute TFR which leads to loss of signal reconstruction ability. PTFA and DTFA are not suitable for processing signals containing multiple components with distinct frequency modulation laws continuously. Here we use SET and SST as post processing technique of STFT which are considered to be promising TFR method, as it enhances TF resolution at the same time allows for perfect signal reconstruction particularly in the case of noisy speech mixtures.

The paper is organized as follows. Section 2 describes the background of the proposed method of SET and SST DUET algorithm. In Section 3 we present the SET and SST enhanced DUET algorithm. Section 4 includes results and discussion and Section 5 gives summary. Expansion of major acronyms used in the following text are-TFR: time frequency representation, TF: time frequency, DUET: degenerate unmixing estimation technique, STFT: short time Fourier transform, SET: synchroextracting transform, SST: synchrosqueezed transform.

## 2. PROPOSED METHOD

DUET is a well-established method for multichannel source separation and localization. It is used in various applications like separation of EEG and ECG signals from medical sensors, separation of radio signals in telecommunication, in audio applications as in hearing aids, for demixing stereo recordings etc. This technique is not bound to any particular type of signals but it performs extremely well when used for separating speech signals due to its various latent properties.

### 2.1. Duet algorithm

Given an audio mixture recorded using two omnidirectional microphones in an anechoic room we know that if a source  $j$  has a distinct spatial position then that source possesses a distinct magnitude parameter  $\alpha_k$  and phase delay  $\delta_k$  which is unique to that particular source [6]. Provided the audio source signals have scanty and disjoint time-frequency characteristics the mixture can be partitioned based on these spatial characteristics. But DUET algorithm fails to provide exact partitioning for separation in real world situation. One reason for this is not able to characterize the mixing parameters exactly in TFR and its effect becomes more prominent in presence of noise. Though STFT is used to convert speech mixtures to TF domain it gives blurred TFR due to various limitations imposed by Heisenberg’s uncertainty principle. This

leads to wrong partitioning of sources based on mixing parameters. Various techniques have been proposed to improve clarity of separated speech [17]. SET based on adopting the reassignment approach of SST and theory of ITFA has been proposed in this paper for sharper TFR which helps in accurate estimation of mixing parameters belonging to each source and hence gives nearly perfect speech separation especially in noisy environments.

## 2.2. Problem formulation

In real world scenarios a time domain mixing model is depicted as

$$X_i(t) = \sum_{j=1}^n a_{ij} s_j(t) + n(t), i = 1, \dots, m, t = 1, \dots, k \quad (2)$$

where  $a_{ij}$  are the mixing coefficients,  $s_j(t)$  are source signals,  $n(t)$  the noise and  $X_i(t)$  are the resultant audio mixtures obtained from the system shown in Figure 1. The main aim is to use a better TFR which helps in recovering original sources from their mixtures with utmost clarity.

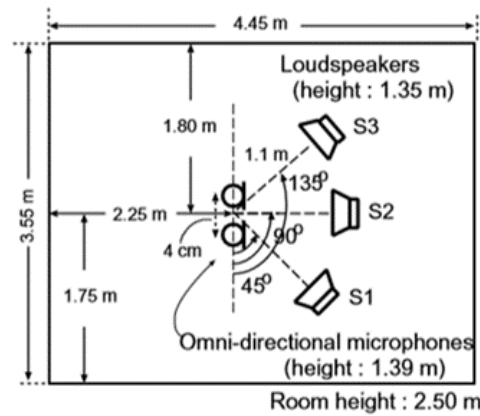


Figure 1. Two channel microphone arrangement with multiple sources

The mixing model consist of a room with 2 microphones and  $k$  sources which is actually the number of speakers who speak in the room and the position of speakers and the microphones are as shown in Figure 1. The speakers are assumed to be stationary and the speakers were randomly assigned to one of the position as shown in Figure 1 and 50 different recordings are made. The performance evaluation for the given arrangement can be done with more number of sources provided minimum angle between two consecutive microphones was  $30^\circ$ .

## 2.3. From STFT to SST and SET

The STFT of a multicomponent signal  $s(t)$  with  $k$  modes is given by

$$X_e(t, \omega) = \sum_{k=1}^n A_k(t) \hat{w}(\omega - \varphi'_k) e^{i\varphi_k(t)} \quad (3)$$

where  $A_k, \varphi_k, \varphi'_k$  denote the  $k$ th mode instantaneous amplitude, instantaneous phase and instantaneous frequency (IF) respectively and  $\hat{w}(\cdot)$  denote the Fourier transform of Hamming window function. STFT gives smeared time frequency energy [18] hence it is impossible to identify time-varying feature accurately. From (3) the instantaneous frequency is given by;

$$\varphi'(t, \omega) = \sum_{k=1}^n \varphi'_k(t, \omega) = -i \frac{\partial_t X_e(t, \omega)}{X_e(t, \omega)} \quad (4)$$

According to [4] SST congregates the STFT coefficients with identical frequency and location given by (5).

$$S_s(t, \eta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_e(t, \omega) \cdot \delta(\eta - \varphi'(t, \omega)) d\omega \quad (5)$$

where  $\int_{-\infty}^{\infty} \delta(\eta - \varphi'(t, \omega)) d\omega$  is the synchrosqueezing operator (SSO). Here TF coefficients are squeezed into IF region  $\eta = \omega_0$  resulting in a new TF plane  $S_s(t, \eta)$  instead of original TF plane  $S(t, \omega)$ . Here reassignment of TF coefficients takes only in frequency direction ie from  $(t, \omega)$  to  $(t, \eta)$

However SET removes trivial interference and smeared time frequency energy and keeps the TF information most associated with TF attributes of target signal by synchroextracting operator (SEO) given by (6).

$$\delta(\omega - \varphi'(t, \omega)) = \begin{cases} 1, & \varphi'(t, \omega), \\ 0 & \text{elsewhere} \end{cases} \quad (6)$$

Hence SET is formulated as (7).

$$S_e(t, \omega) = X_e(t, \omega) \delta(\omega - \varphi'(t, \omega)) \quad (7)$$

Thus it is clear from equation (5) that SST reassigns the coefficients around IF trajectory while SET extracts the TF coefficients in IF trajectory with SEO and the rest of TF coefficients are removed. Thus we find that SET is more energy concentrated than SST. Hence by taking only the TF coefficients that are more energy concentrated we can remove the most smeared time frequency energy and get high clarity TF representation in case of SET whereas SST is actually reassignment of instantaneous frequency of the smeared energy coefficients to a new point in TF plane. Also though SST and SET needs to know the instantaneous frequency trajectories for reconstruction SST needs to have additional information about the integration regions [8]. Compared to SST, SET provide a sharpened and focused representation of coefficients in TF plane especially in noisy environments and our aim is to find out which technique is better when applied to DUET.

### 3. RESEARCH METHOD

In BSS techniques DUET has been accepted as one of the most effective way for signal separation especially in cases when number of sources is greater than number of sensors. However TFR which is the core part of DUET fail to localize and separate individual signals when STFT is used due to fixed spectral resolution caused by predetermined window width. Various potential methods like SET and SST has been used to increase the sharpness of TFR and is used in various application like identifying power quality disturbances [19], fault diagnosis in rolling bearing [20], hydrocarbon detection [21], model based deep learning [22-23], seismic time-frequency analysis [24] etc. Hence SST and SET is an ideal method for improving TF resolution in DUET which ultimately results in high quality speech separation especially in noisy conditions.

#### 3.1. SST & SET enhanced duet algorithm

SET –DUET Algorithm

- Obtain speech mixtures  $X_1(t)$  and  $X_2(t)$ ,
  - Find gradient of window function. Let window be Hamming window
  - Find STFT's,  $X_{e11}(t, \omega)$  and  $X_{e12}(t, \omega)$  for mixture  $X_1(t)$ , where  $X_{e11}(t, \omega)$  is STFT with original Hamming window and  $X_{e12}(t, \omega)$  is STFT with derivative of window function. Similarly find  $X_{e21}(t, \omega)$  and  $X_{e22}(t, \omega)$  for  $X_2(t)$ .
- ```

- for i=1:numfreq/2
    for j=1:numtime
        if abs(real(Xe12(t, ω)/Xe11(t, ω))) > λ
            IF(i, j)=1
        end
    end
end

```
- $T_{ef1} = X_{e11}(t, \omega) * IF(i, j)$ ;
  - Repeat steps 4 to 5 for mixture  $X_2(t)$  and obtain  $T_{ef2}$
  - For each TF points given by  $T_{ef1}$  and  $T_{ef2}$ , find mixing parameters  $(\alpha(t, \omega), \delta(t, \omega))$ , where  $\alpha(t, \omega)$  and  $\delta(t, \omega)$  are the instantaneous estimates of the relative attenuation and delay of sources [6] respectively as observed by  $X_1(t)$  and  $X_2(t)$

- Construct high resolution histogram and smooth it
- Locate peaks in histogram, there will be N peaks (one for each source) with peak location approximately equal to the true mixing parameter pairs
- For N pairs of mixing parameter pairs construct TF masks using ML partitioning [6] and apply these masks to one of the mixtures to get estimate of TF representation of original sources.
- Find inverse SET to convert each source back to time domain

where numfreq is frequency components per time point and numtime is time components per frequency point and  $T_{ef1}$  is SET TF points. Here  $\lambda=10^{-8}$  in noise free conditions and  $\pi = \sqrt{2 \log_2 N} * \mu$  where N is signal length and;

$$\mu = \text{median}(|X_{e11}(\eta, t) - \text{median}(X_{e11}(\eta, t))|) / 0.6745 \text{ [25]}$$

#### SST-DUET Algorithm

All the steps of SST-DUET algorithm is the same as SET-DUET algorithm except in steps 4. In SST the IF region is found out and all the TF coefficients are squeezed into that region along frequency direction. We find the IF trajectory  $\omega(i,j)$  and the step 4 of algorithm becomes;

```
for i=1:numtime
  for j=numfreq
    if abs(Xe11(i,j)) > lambda
      eta = omega(i,j)
      Ts(i,eta) = Ts(i,eta) + Xe11(i,j)
    end
  end
end
```

Here the coefficients are squeezed into IF trajectory to obtain a new TFR given by  $Ts(i,\eta)$ .

## 4. RESULTS AND DISCUSSION

Here we analyze the influence of SET & SST on the characteristics of signal mixtures in their TF domain. For the purpose of accessing the performance of SET & SST, STFT is considered as reference tool. Also performances of TF sharpening tools in DUET algorithm are numerically validated using known mixture of five sinusoidal signals under both noisy and noise free conditions it is further experimentally evaluated using speech mixtures. General evaluation of clarity of TF domain is validated using Renyi Entropy. In speech mixtures quality of demixing, noise robustness and ability of signal reconstruction is evaluated using BSS-Eval tool box [25].

### 4.1. Evaluation in TF domain

We evaluate the performance of the proposed method on two different situations: synthetic mixture of five sinusoidal signals and real speech mixtures. The synthetic mixtures of sinusoidal signals are modeled as;

$$S_1(t) = \sin(2\pi 6t) + 0.8 \sin(2\pi 10t) + 0.3 \sin(2\pi 15t) + 0.5 \sin(2\pi 20t) + 0.7 \sin(2\pi 25t)$$

where the five signals have frequencies of 6 Hz, 10 Hz, 15 Hz, 20 Hz and 25 Hz and sampling frequency is 100 Hz. The TF representations of mixtures using three different techniques of SET, SST and STFT for Hamming window are shown in Figure 2. Figures 2(a) and 2(b) represents these TF representations for noise free and noisy conditions respectively. The efficiency of SET in separating the component frequencies to regions with well-defined boundaries is clearly evidenced from these figures. In case of STFT though TF representation brings out all the component frequencies the boundaries are much blurred, while in case of SET boundaries are much sharper. Of the three transforms investigated here SET gives a clear TFR compared to other two.

The TFR using these three transforms in real speech mixtures is shown in Figure 3. In general for speech signals, there are possibilities of multiple frequencies to be present which in most cases will be highly overlapped. In this situation, TFR of SET shows much clear separation between component frequencies where in overlapping is low and hence less blurring compared to other two transforms. Further we investigate the performance of these transforms in terms of histogram obtained for estimation of mixing parameters in DUET. Figure 4(a) and Figure 4(b) shows histogram obtained using SET TFR and STFT TFR. Figure 4(a) shows concentration of peaks than the histogram in Figure 4(b) that uses STFT-TFR. Figure 4(a) produces concentrated clusters because on estimating IF trajectory and considering only TF points that are in IF removes unwanted TF points from interfering when forming clusters to estimate mixing parameters. This results in better reconstruction of individual sources and hence high clarity speech separation especially in

case of noisy speech mixtures. Explicit differences by relative comparison of SET, SST and STFT in Figure 3, Figure 4 and Figure 5 by visual inspection are not clearly evidenced. Hence we further investigate the efficiency of SET in terms of quantitative analysis.

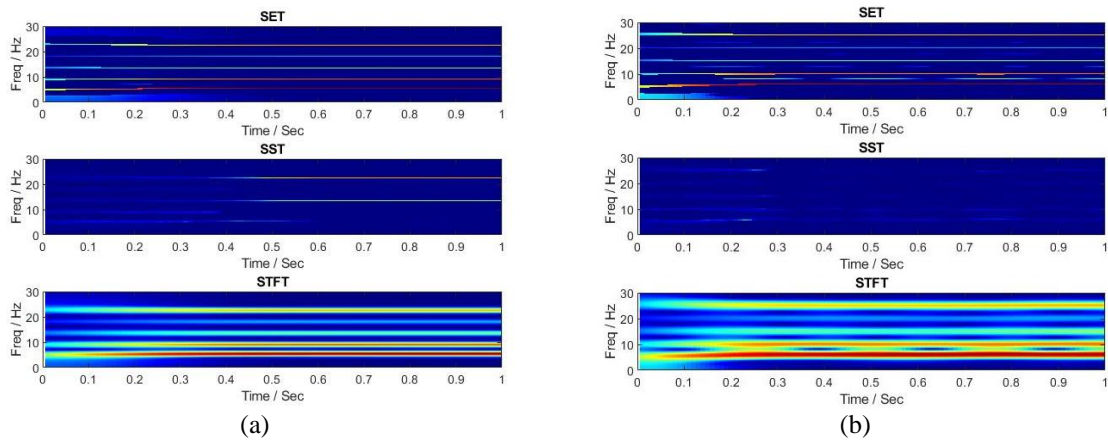


Figure 2. TFR's of a sinusoidal mixture; (a) noise free, (b) noisy

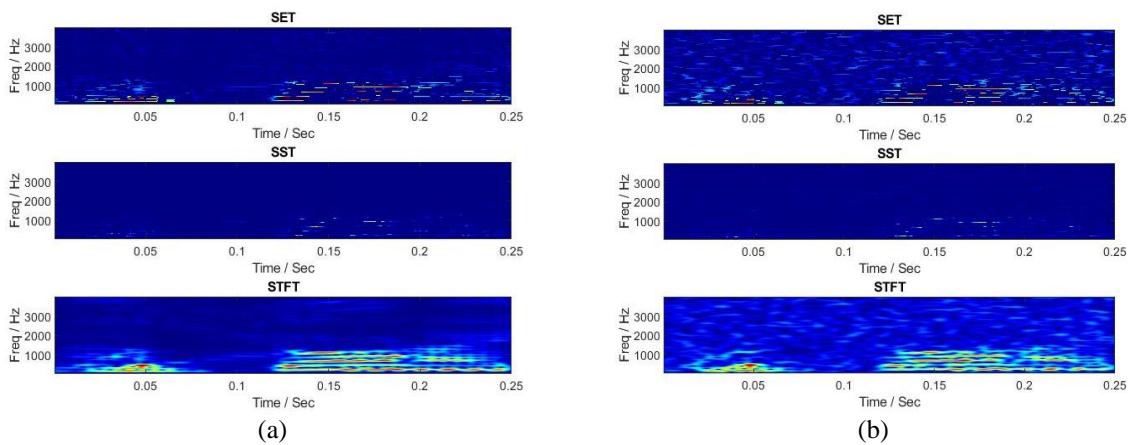


Figure 3. TFR's of a portion of speech mixtures; (a) noise free, (b) noisy

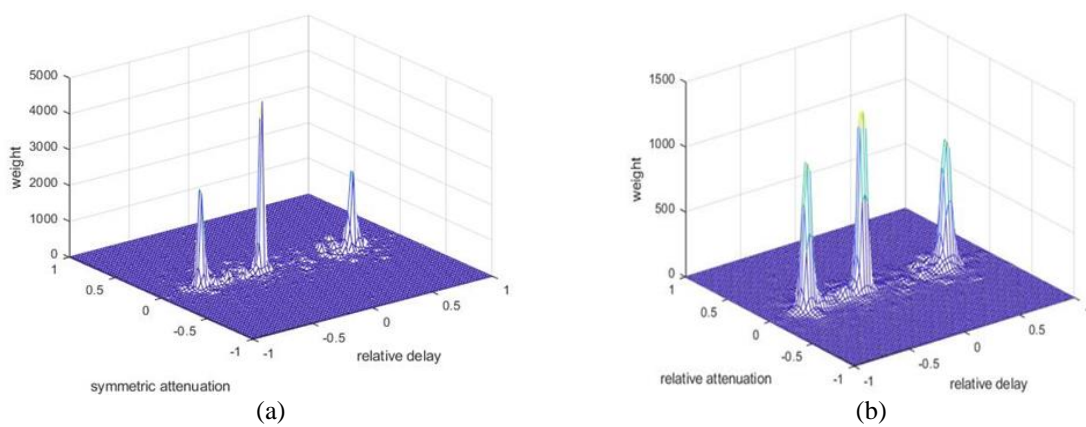


Figure 4. Histogram of 3 source mixture of DUET; (a) SET-TFR, (b) STFT TFR

**4.2. Quantitative analysis**

**4.2.1. Renyi entropy**

In order to evaluate efficiency of proposed method in providing clearly separated speech sources from their mixtures and further comparing it with that of conventional methods we carry out quantitative evaluation using Renyi entropy. In any case of demixing the correct choice of window length is a crucial parameter which influences the clarity of separated sources. To investigate how the choice of window length influences the correct estimation we evaluate the Renyi entropy [18, 19] of TFR of three transforms for varying window length. Figure 5 shows variation of Renyi entropy values for various window lengths in noise free and noisy conditions. From Figure 5(a) it can be inferred that irrespective of window length SET always provides lower Renyi entropy than SST and STFT. However this observation is valid for short window length wherein Renyi entropy decreases as window length decreases. For longer window the performances of all three transforms tend to be similar indicated by similar Renyi entropy in Figure 5(a).

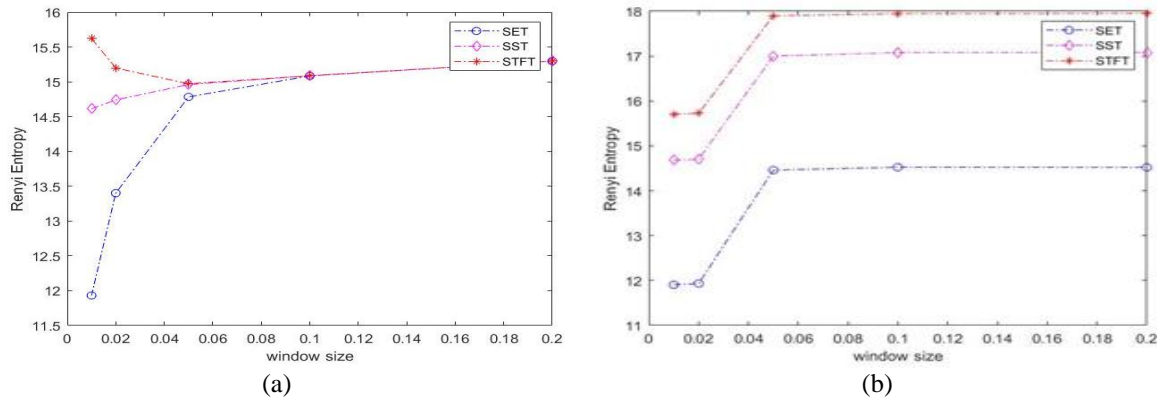


Figure 5. Evaluation of Renyi entropy for speech mixture for varying window length calculated by STFT, SST, and SET; (a) noise free, (b) noisy

The performances of SET and SST in noisy condition for varying window length is investigated using the Renyi entropy of STFT, SST and SET TFR's in the speech mixtures which is given in Figure 5(b). It is clear from the above figure that Renyi entropy of SET is much lower than SST and STFT in noisy conditions for shorter as well as longer window. Unlike SST, which gathers all coefficients to the corresponding IF trajectories, SET gathers only those coefficients in the IF trajectories which has maximum energy [17, 18]. Thus SET generates a novel TFR where in the effect of noise is minimized there by resulting in lower entropy. The above results clearly prove the efficiency of SET over SST in speech separation under noisy conditions.

Table 1 shows the time required for computation of STFT, SET and SST for speech mixtures. On comparing them we find that SET and SST requires almost twice the computational time of STFT. As SET requires evaluation of every IF trajectory in addition to TF point it is natural that computation time is higher than that of STFT. Slightly higher computation time demanded by such transforms is worthified considering the sharper TFR they can provide which in turn contribute to highly efficient demixing. Among these two transforms SET proves to be an ideal choice on considering the increased sharpness of TFR and slightly lower computation time.

| TFR     | STFT  | SET   | SST   |
|---------|-------|-------|-------|
| Time(s) | 0.022 | 0.039 | 0.055 |

**4.3. Objective measures**

We compare the proposed method with existing DUET techniques for various objective measures such as signal to distortion ratio (SDR), signal to artifact ratio (SAR) and signal to interference ratio (SIR) given by equations (8, 9, 10). The signal is decomposed into a source part  $S_{target}$ , along with error terms such as interference  $e_{inter}$  and algorithmic artifacts  $e_{artifact}$  [25]. Also intelligibility of estimated sources is evaluated using correlation between reconstructed sources with its original sources. These objective measures are

evaluated using BSS evaluation tool box [25]. All computations are done using MATLAB version R2017a under Windows 10 with Intel Core i7-6500 central processing unit at 3.3 GHz and 8 GB random access memory.

$$SDR = 10 \log_{10} \frac{\|s_{target}\|^2}{\|e_{inter} + e_{artifact}\|^2} \quad (8)$$

$$SIR = 10 \log_{10} \frac{\|s_{target}\|^2}{\|e_{inter}\|^2} \quad (9)$$

$$SAR = 10 \log_{10} \frac{\|s_{target} + e_{interf}\|^2}{\|e_{artif}\|^2} \quad (10)$$

Table 2 shows values of SDR, SIR and SAR for DUET algorithm enhanced with SET, SST and conventional DUET algorithm which uses STFT TFR's. A high value of SIR is the basic requirement for efficient speech separation algorithm while it can be allowed that the other two measures are at relatively moderate level. From Table 2 it can be inferred that both SST and SET are efficient in enhancing the speech separation performance of conventional DUET algorithm. Reconstruction ability indicated by correlation values of extracted speech signals also indicate efficiency of SET and SST compared to that of traditional DUET algorithm. A comparison of performance of SET and SST in terms of objective measures and correlation value clearly indicate relatively better efficiency of SET.

Noise robustness is another crucial factor that affects the reconstruction efficiency of any speech separation algorithm. To evaluate performance of SET and SST under different noise levels white noises with SNR's 1 dB to 80 dB were added to the speech mixture before applying the demixing procedure. Figure 6 shows average correlation values for the speech signals under these noise levels for SET and SST. From the figure it can be observed that for high SNR values SET and SST exhibits similar performance whereas for low SNR values below 40 dB reconstruction ability of SET is much better than SST. Better performance of SET under low SNR values can be attributed to the specific approach of SET which removes the most smeared TF coefficients and thus leading to reduced effect of noise in IF trajectory. Thus SET reconstruction shows the best match between estimated and original sources in low SNR cases. Hence we can conclude that SET reconstruction is more robust to noise than SST in highly noisy speech mixtures.

Table 2. Three source demixing performance of DUET technique against three TFR's

| Source | SDR   |       |       | SIR   |       |       | SAR   |       |       | Corr |      |      |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|------|
|        | SET   | SST   | STFT  | SET   | SST   | STFT  | SET   | SST   | STFT  | SET  | SST  | STFT |
| 1      | 12.52 | 12.39 | 11.95 | 24.67 | 24.03 | 23.97 | 13.85 | 13.73 | 13.36 | 0.97 | 0.92 | 0.90 |
| 2      | 11.41 | 11.11 | 11.02 | 21.54 | 21.32 | 20.84 | 11.47 | 11.21 | 11.09 | 0.95 | 0.90 | 0.88 |
| 3      | 13.67 | 13.23 | 12.96 | 23.17 | 23.15 | 22.99 | 14.54 | 14.42 | 13.39 | 0.98 | 0.94 | 0.91 |

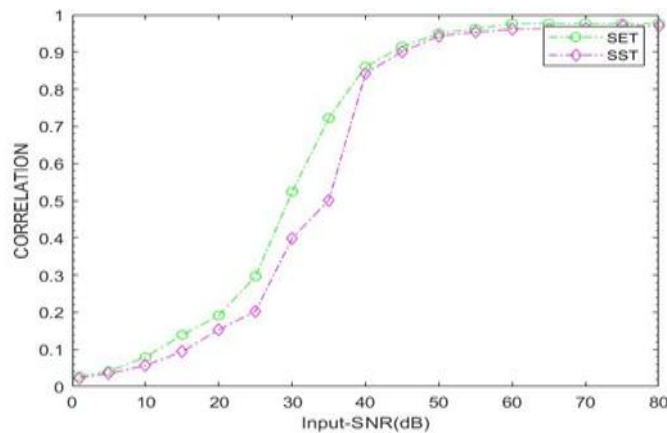


Figure 6. Correlation of reconstructed result of speech mixture under different noise levels



## 5. CONCLUSION

Here we present the results of investigation and the effectiveness of post processing, reassignment techniques of SET and SST in improving TF resolution for the method of DUET under different noise levels. These approaches make use of IF's to further process TF points, for improving readability in both frequency as well as time direction. The performances of these two techniques are addressed qualitatively and quantitatively. Qualitative analysis is carried out by visual inspection of TFR and histogram peaks used for estimation of mixing parameters in conventional DUET algorithm. Quantitative measures are done using Renyi entropy of TFR and objective measures of speech mixtures. Renyi entropy which is used as performance indicator has lower values for SET than SST for different noise levels. Also SDR, SIR and SAR give much better results for SET compared to SST. The efficiency in extraction of original signals are estimated using correlation values between extracted signal and original source signals which is better for SET compared to other two transforms. Thus SET enhances the performance of DUET algorithm in terms of accuracy of source estimation from speech mixtures.

The present results indicate that SET is much better than SST as it requires only fewer parameters for reconstruction of the signal while SST demands information about the regions of integration which is hard to obtain in case of strong FM signals. SST squeezes every TF coefficients into specific IF trajectory which is carried out only in frequency direction whereas SET extracts specific TF coefficients into specific IF trajectory both in time and frequency direction. The objective measures obtained are highly promising and encouraging for the use of SET as an efficient post processing technique of speech signals in underdetermined condition. Thus SET can prove to be an efficient technique in application areas requiring sharp TFR like image processing, speech processing and various other signal processing fields.

## REFERENCES

- [1] P. Comon, "Independent component analysis-a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287-314, 1994.
- [2] M. Lewicki, and T. Sejnowski, "Learning overcomplete representation," *Neural Computation*, vol. 12, no. 2, pp. 337-365, 2000.
- [3] T. W. Lee, M. Lewicki, M. Girolami, and T. Sejnowski, "Blind Source Separation of more sources than mixtures over complete representations," *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87-90, 1999.
- [4] P. Bofill and M. Zibulevsky, "Blind separation of more sources than mixtures using sparsity of their short-time Fourier transform," *International Workshop on Independent Component Analysis and Blind Signal Separation*, Helsinki, Finland, pp. 87-92, 2000.
- [5] A. Jourjine, S. Rickard, and Q. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, Istanbul, Turkey, vol. 5, 2000, pp. 2985-2988.
- [6] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830-1847, 2004.
- [7] F. Auger and P. Flandrin, "Improving the readability of time frequency and time scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068-1089, 1995.
- [8] I. Daubechies, J. Lu and H. T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243-261, 2011.
- [9] S. Wang, X. Chen, G. Li, X. Li and Z. He, "Matching demodulation transform with application to feature extraction of rotor rub-impact fault," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1372-1383, 2014.
- [10] T. Oberlin, S. Meignen, and V. Perrier, "Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time frequency representations," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1335-1344, 2015.
- [11] Z. K. Peng, G. Meng, F. L. Chu, Z. Q. Lang, W. M. Zhang, and Y. Yang, "Polynomial chirplet transform with application to instantaneous frequency estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 9, pp. 3222-3229, 2011.
- [12] W. Yang, P. J. Tavner, and W. Tian, "Wind Turbine Condition Monitoring Based on an Improved Spline-Kernelled Chirplet Transform," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6565-6574, 2015.
- [13] Y. Yang, Z. K. Peng, G. Meng and W. M. Zhang, "Spline-kernelled Chirplet transform for the analysis of signals with time-varying frequency and its application," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 3, pp. 1612-1621, 2012.
- [14] Y. Yang, W. Zhang, Z. Peng and G. Meng, "Multicomponent signal analysis based on polynomial chirplet transform," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 9, pp. 3948-3956, 2013.
- [15] S. Wang, X. Chen, G. Li, X. Li and Z. He, "Matching demodulation transform and synchrosqueezing in time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 69-84, 2014.
- [16] L. J. Stankovic, I. Djurovic, S. Stankovic, M. Simeunovic, S. Djukanovic, and M. Dakovic, "Instantaneous frequency in time-frequency analysis: Enhanced concepts and performance of estimation algorithms," *Digital Signal Processing*, vol. 35, pp. 1-13, 2014.
- [17] S. M. Abdulla and J. Jayakumari, "Degenerate Unmixing Estimation Technique of Speech Mixtures in Real Environments using Wavelets," *International Journal of Electronic Letters*, vol. 8, no. 3, pp. 344-353, 2019.

- 
- [18] G. Yu, M. Yu, and C. Xu, "Synchroextracting Transform," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 10, pp. 8042-8054, 2017.
  - [19] Guptaeswar Sahu, Sonali Dash and Birendra Biswal, "Time-Frequency analysis of power quality disturbances using synchroextracting transform," *International Transactions on Electrical Energy Systems*, vol. 30, no. 4, 2020.
  - [20] Zengqiang Ma, Wanying Ruan, Mingyi Chen, and Xiang Li, "An Improved Time-Frequency Analysis Method for Instantaneous Frequency Estimation of Rolling Bearing," *Shock and Vibration*, vol. 3, pp. 1-18, 2018.
  - [21] H. Chen, J. Kang, Y. Chen, D. Xu and Y. Hu, "An improved time-frequency analysis method for hydrocarbon detection based on EWT and SET," *Energies*, vol. 10, no. 8, p. 1090, 2017.
  - [22] G. Aceto, D. Ciuonzo, A. Montieri, A. Pescape, "MIMETIC: Mobile Encrypted Traffic Classification using Multimodal Deep Learning," *Computer Networks*, vol. 165, 2019.
  - [23] G. Aceto, D. Ciuonzo, A. Montieri, A. Pescape, "Mobile Encrypted Traffic Classification Using Deep Learning: Experimental evaluation, lessons learned, and challenges," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 445-458, 2019.
  - [24] Roberto H. Herrera, Jiajun Han and Mirko van der Baan, "Applications of the synchrosqueezing transform in seismic time-frequency analysis," *Geophysics*, vol. 79, no. 3, 2014.
  - [25] E. Vincent, R. Gribonval and C. F'evotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462-1469, 2006.