

## Increase the Accuracy of Speech Signal Categories in High Noise Environments

Mehdi Morovati <sup>1\*</sup>

<sup>1</sup>Mechanical Engineering and Engineering Science, University of North Carolina at Charlotte, Charlotte, USA

Correspondence to: Morovati M. (E-mail: mmorovat@uncc.edu)

### Abstract

this paper, in order to improve the quality and reconstruction of the noisy speech signal, an algorithm based on clustering in the spectrographic image texture is proposed and evaluated. In this algorithm, we first determine the cluster centers by the method of Glass Compressed Transformation (GCT), based on image tissue extraction and Weighted K-mean algorithm in the field of GCT. After determining the centers of the clusters and their variance, by using the Gaussian Distribution Function (GDF), we determine the mask for each cluster in the GCT space and apply a signal to the GCT domain. In the continuation of this research, this method can be used to separate the two speakers that have been combined and it also can be used in separating speakers in noisy situations. The results of the method evaluation in the presence of a variety of conventional acoustic noises show the advantages of this method in improving the quality of speech.

**Keywords:** Weighted K-mean, Gaussian Distribution Function, Clustering, Mask

**Received:** 18 July 2020, **Accepted:** 24 August 2020

**DOI:** 10.22034/JBR.2020.232801.1023

### 1. Introduction

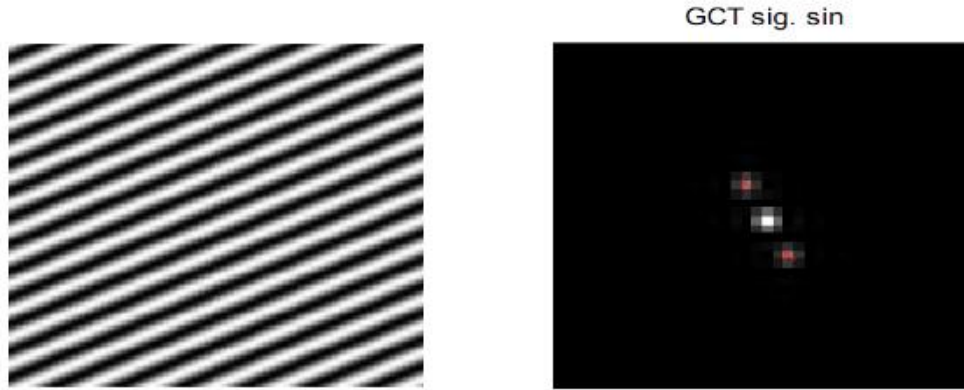
Different signal processing methods have been used in order to process different signals and remove noise from them such as Butterworth low pass digital, Fast Fourier Transform and Two-dimensional Fourier Transform [1]. Hilbert-Huang Transform can be used to determine the dynamic specification of a system, Kelareh et al. determined the dynamic characteristics of an eight-story structure with acquisition noise and different loads through the Hilbert-Huang Transform An adaptive extended Kalman filter is applied in proposed algorithm to reduce the noises [2]. Hamed et al. showed that using a state-feedback controller with extended Kalman filter can decrease the noise of signal significantly [3]. Mohamed et al. implemented Fast Fourier Transformation for

analyzing the signals extracted from PZT in the structural health monitoring application [4,5].

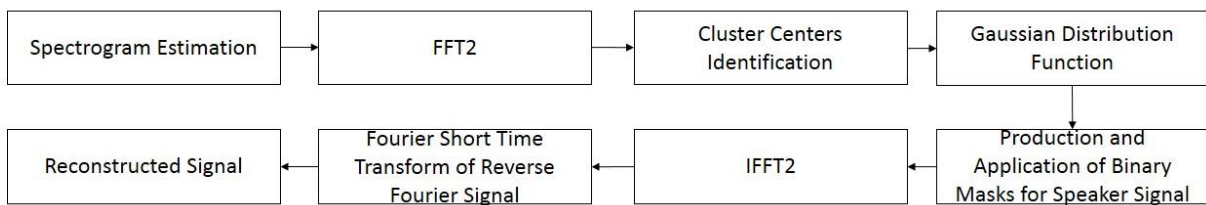
Two-dimensional Fourier Transform field, widely used different process such as noise removal in image processing, , postprocessing of tomographic Particle Image Velocimeters [6,7] and many features and image descriptors are extracted in this area. Conversely, in speech processing, valuable information for speech processing is usually extracted from the amplitude of the short signal spectrum time, which is displayed as a spectrogram image. Although the image is anywhere, the value of the spectrum amplitude in the frame of speech and it is displayed at a known frequency, it cannot completely reconstruct the input time signal, but the reconstructed signal from an auditory point of view has the same quality as the input signal. In some other applications, like system identification and FDIR methods, FFT is a convenient tool to extract the features of the



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



**Figure 1:** STFT from signal (left) and GCT from left picture



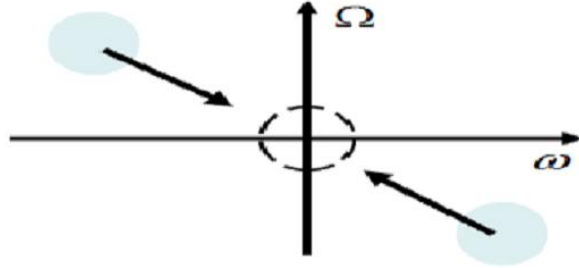
**Figure 2:** Diagram block of the general process of the proposed method

System [8,9]. The extracted features are the key indicator for different classified faults and errors. Also, the combination of FFT and a multi-layer neural network is used to detect electromechanical faults in RW [9]. The time for generating biomedical samples is a vital factor when we consider ambulatory devices, with the fact that information should be sent to the physician as soon as possible. In addition, there are some wearable ECG recorders that have limited power, and may only be capable of doing simple algorithms. In these cases also using signal characteristics in bio-medical applications can be based in DFT [10].

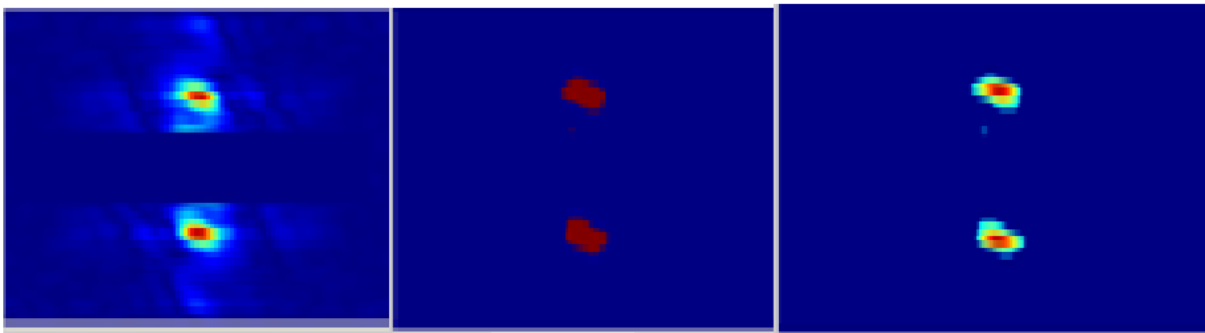
In recent years, there has been a growing interest in spectrogram image processing for spoken information analysis [11]. In a process suggested by Quatiri et al. [11-14], Speech information is searched in the spectrogram image texture using two-dimensional Fourier transform windows of the spectrogram [12]. As the spectrogram image axes represent time (frame number) and frequency,

the axes of this image represent the frequency of changes along with the successive frames (scales) and the new page of the dense window conversion, or GCT, is called the speech signal.

In the one-dimensional Fourier transform of the speech signal, each of the step frequency harmonics can be considered as a sink that appears on the spectrogram plane in parallel or diagonal lines [11-12]. If a two-dimensional Fourier transform of the spectrogram plate multiplied by a local window function such as a square or a Hamming is considered, the resulting rate and scale domain in the convolution of the Fourier transform multiplied window for a window containing parallel lines, according to figure 1, is two clusters which represent the harmonic distance and the slope of the harmonic distance changes over time. This idea is used to extract step frequency, [13,14] and this process is transformed into a noise-resistant algorithm for step extraction by image clustering algorithm [15]. Rafieipour et al. presented a self-organized method of clustering



**Figure 3:** The process of adding dc to the reconstructed signal using cluster shifts to the source

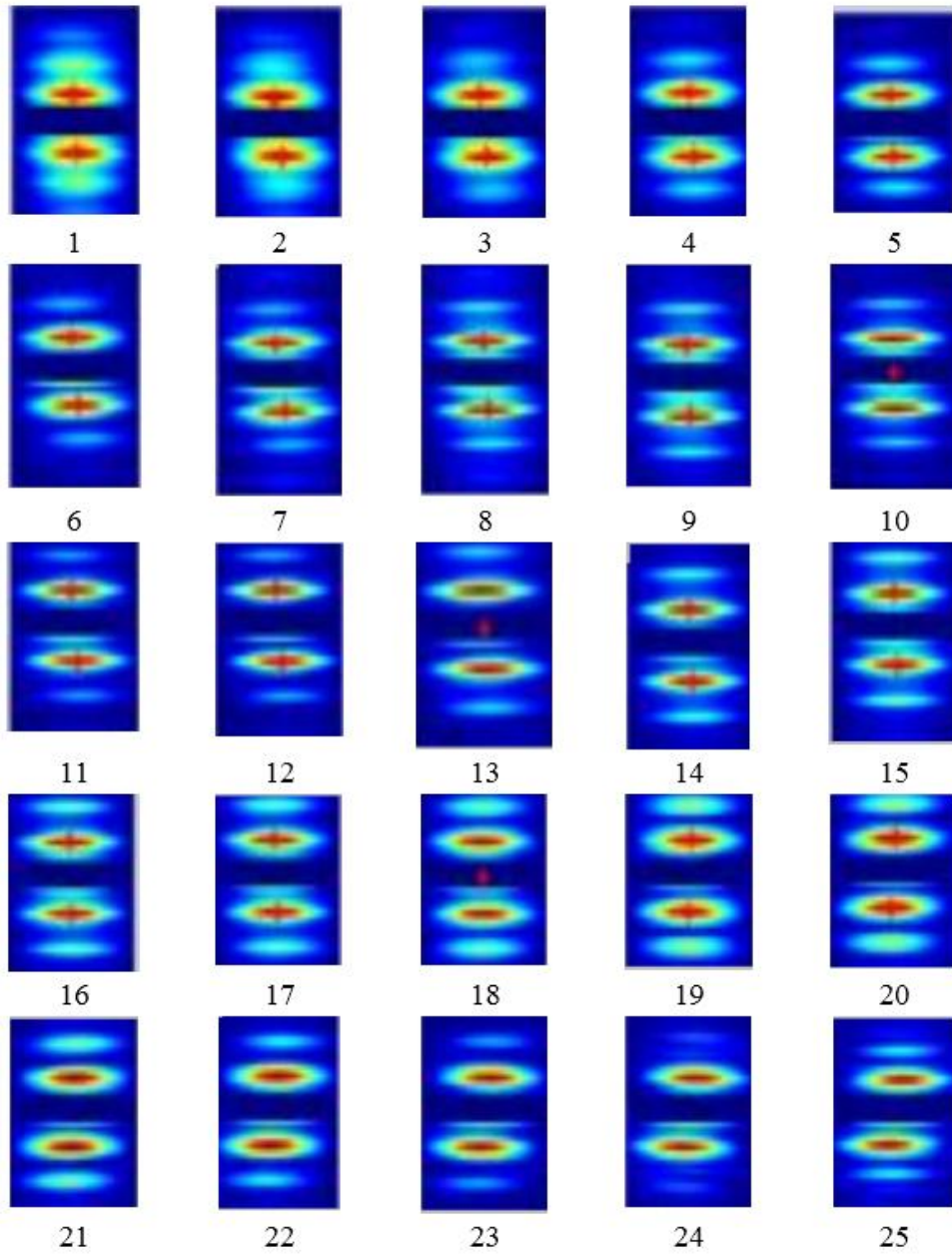


**Figure 4:** GCT from a specific part of the STFT signal and remove dc

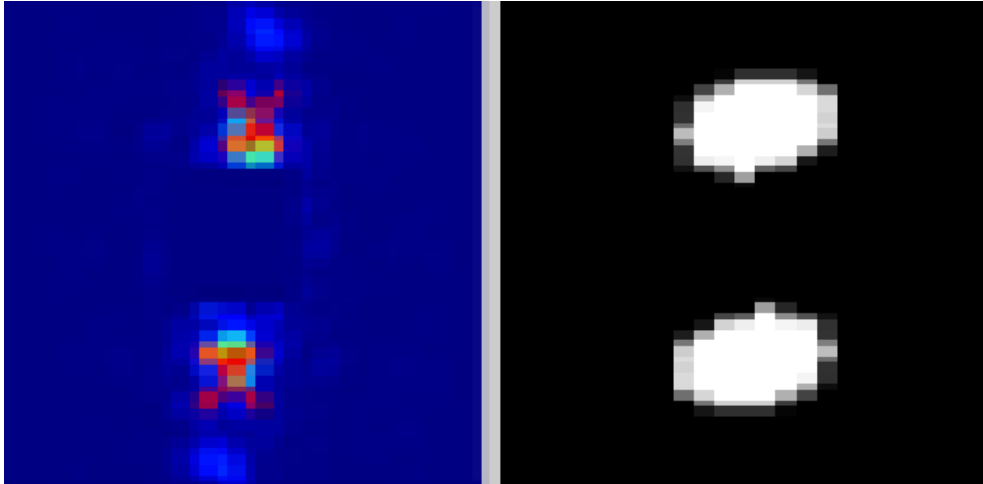
based on Low Energy-Adaptive Clustering Hierarchy (LEACH) algorithm, which considers the frequencies [16]. Wang et al. also proposed a method for extracting formulations from high-frequency steps based on GCT [26]. To separate speech from noise, we use the Weighted K-means clustering method to cluster information in the GCT space, delineate, and extract cluster centers [19]. In this paper, the number of clusters with a random procedure is obtained by means of testing. In this way, the number of clusters is chosen randomly from the samples. Then, the mean and variance of all samples from these centers are calculated, and each sample is computed by comparing with the centers of the cluster and selecting the closest member distance to the cluster. The centers calculate the spectra obtained from the two-dimensional Fourier transform

of the populated spectrogram. Then, the weighted samples are calculated as the centers of the clusters. The WK-means method is used as a precise method for determining the frequency of the speech step. In this paper, we use the above method to determine the centers of the clusters to obtain an accurate mask of the clusters to improve the speech signal in an efficient way by obtaining a proper time-frequency mask using the Gaussian model.

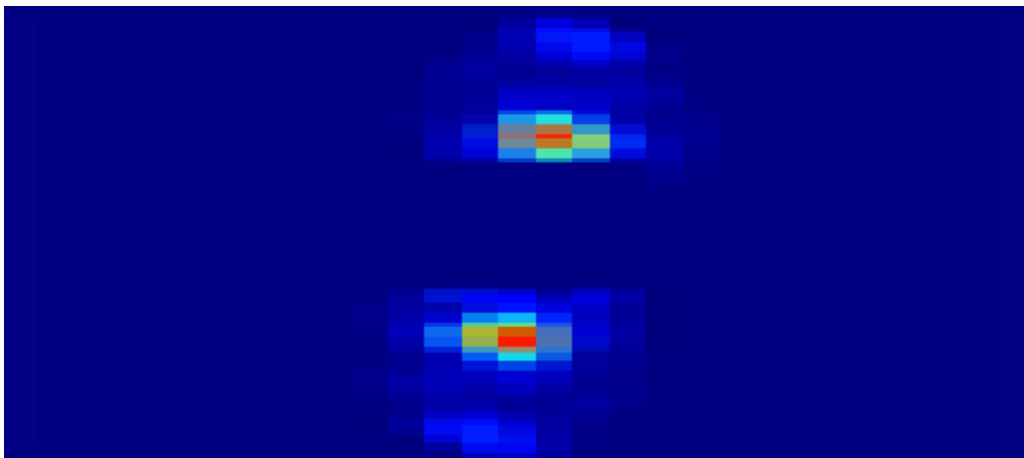
The rest of the article is organized in several sections. Section 2 deals with the GCT formulation and how to analyze and reconstruct the signal from the GCT domain. In Section 3, the proposed method for improving speech signals in GCT domain space is presented. We also describe the experiments and results in Section 4. In this section, the performance of the algorithm against different types of noise with



**Figure 5:** Display HCG and determine the cluster center using the wk-mean algorithm and removal of the dc value for 25 frames for the male speaker



**Figure 6:** Determination of cluster centers (left figure) and Determine the mask using the threshold on the Gaussian function (right figure)



**Figure 7:** Applied Gaussian mask to the GCT range

different SNR is evaluated, and section 5 concludes the article.

After determining the centers of the clusters, the mask corresponding to each cluster is determined, after which the Gaussian mask transform function is obtained. Figure 6 shows the images of a speech signal block.

## 2. Signal analysis and reconstruction using GCT method

For GCT extraction, the speech signal spectrogram is first extracted. Equation 1 is related to the Fourier transform of a short time  $X[n, \omega]$  [14,15,37,38].

$$X[n, \omega] = \sum_{m=0}^{L-1} x[n+m]w[m]\exp(-j2\pi\omega m)$$

(1)

Where  $x[n]$  is the signal value, and  $w[n]$  is the window  $0 \leq n \leq L-1$  distance, and the window is chosen according to the short time required. In this paper, the shape of the selected Hamming window is selected.

If we show the time and frequency axes in  $m$  and  $n$  respectively in the spectrogram plane, the modulated components can be represented by a static two-dimensional sinusoidal model  $C[m, n]$ , which contains the spatial frequency and  $k$  is the value of the  $dc$  in signal [1]. In Equation 2,  $\theta$  represents the angle of the spectral lines on the spectrogram. In practice, the modulated components are analyzed in small sections of the spectrogram plane separated by multiplied windows. The signal analysis is now continued by multiplying the area specified in the window [12-3].

$$w[n, m]c[n, m] = w[n, m](k + \cos(\omega\Phi[n, m]))$$

$$\Phi[n, m] = n \cos \theta + m \sin \theta$$

(2)

After receiving the two-dimensional Fourier transform of area  $c[n, m]$ , we will have:

$$c[m, \Omega] = 2k\delta(\omega, \Omega) + \delta(\omega - \omega_0 \cos \theta, \Omega + \omega_0 \sin \theta) + \delta(\omega + \omega_0 \cos \theta, \Omega - \omega_0 \sin \theta)$$

(3)

In equation 3, the  $\omega, \Omega$  is the two-dimensional Fourier transform of the STFT signal, and sinusoidal harmonic lines are used to determine step frequency.<sup>[13,14]</sup> In equation 4,  $f_s$  is the sampling frequency, NSTFT, the number of DFT points, and the  $\omega_s \cos \theta$  is the vertical peak distance from the origin of the GCT domain, respectively. The GCT parameters are related to the step frequency determination.

$$f_0 = \frac{2\pi f_s}{NSTFT \omega_s \cos \theta}$$

(4)

the STFT representation of a specific segment of the signal with the GCT is shown in Figure 1. In the figure below, the distance between the two STFT lines is  $\frac{2\pi}{\omega_s}$  [9,11,22,23].

At this stage, after GCT determination, the dc value is removed and demodulated to reconstruct the STFT domain and its phase, and after adding the amplitude and phase, the  $dc$  value is added to the reconstructed signal.[18-29]

### 3. Proposed method for improving signal quality in GCT space using zero DC value

GCT is a two-dimensional analysis of speech signals which is effective in estimating the frequency of the step taken to improve speech and blended speech. Finding the exact position of the frequency spectrum centers is particularly important in calculating the speech step frequency [20]. For this purpose, the Weighted K-means algorithm is used to accurately determine the speech step frequency [19]. After determining the speech step frequency, we obtain the exact centers of the clusters that is an automated and unsupervised way of finding cluster centers from training samples. So that the number of clusters is randomly selected from the samples and then the mean and variance of all samples are selected from these centers and each sample is selected through comparison with centers. To precisely determine the centers of the clusters and to simplify the calculations, the extra samples and dc values are removed. After specifying the centers of the clusters and removing the dc value, it uses the energy of the clusters to determine the mask for each cluster. Due to the symmetry of the GCT axis, we zero the spectrum by selecting the appropriate part of the time-frequency spectrum of the DC part. Then, we determine the mask for each cluster. Using the time-frequency binary mask, which

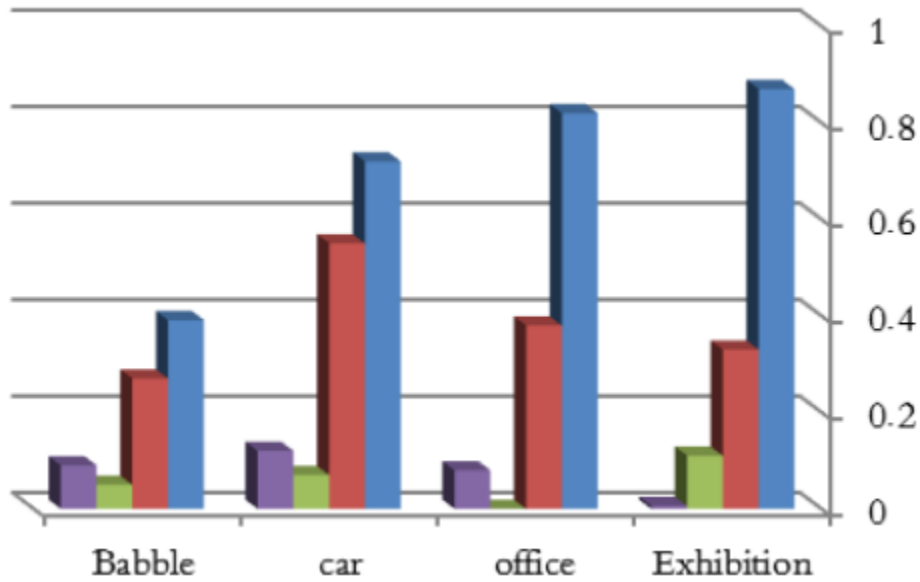


**Table 1:** PESQ results for the above two experiments for both male and female speakers

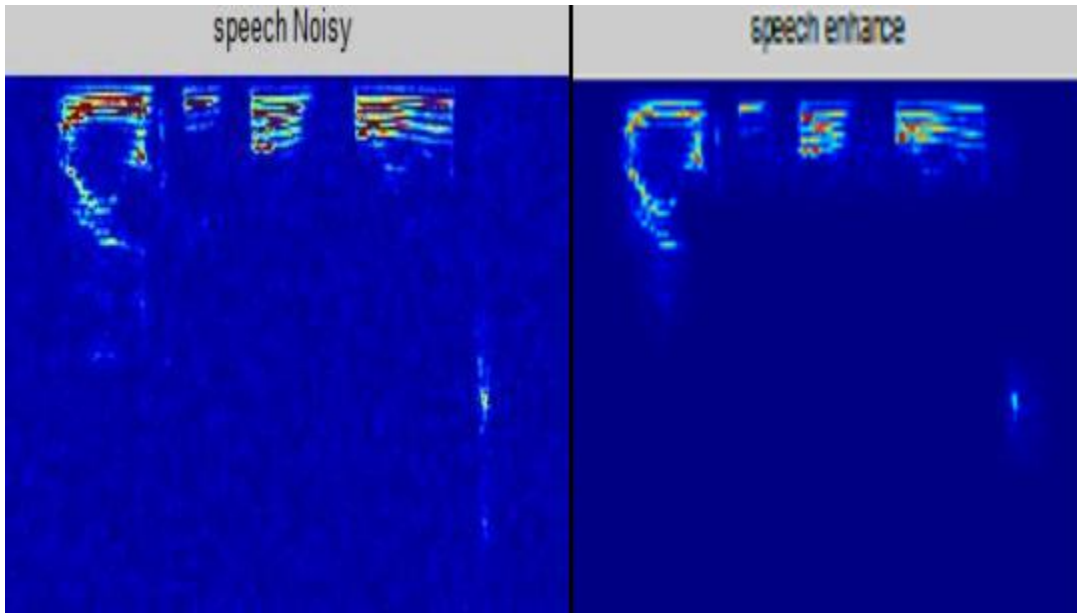
The type of speaker	Experiment 1: Use a binary mask	Experiment 2: Use the Gaussian distribution function
Male	PESQ=4.21	PESQ=4.39
Female	PESQ=3.91	PESQ=4.006

**Table 2:** Comparison between two experiments using white noise with different SNRs. The first test is to remove and add dc using the appropriate threshold. The second experiment is to shift the clusters to the center to get the dc value for the male speaker and the female speaker

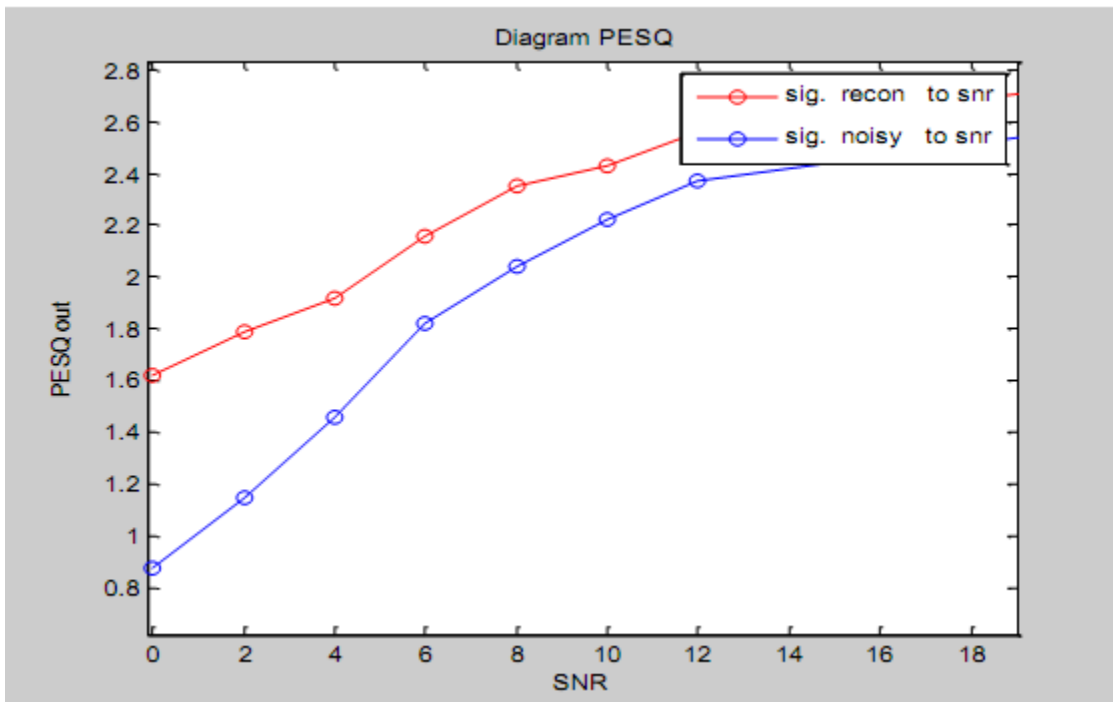
SNR	0db	2db	4db	6db	10db	Clean signal
PESQ_ex2(M)	1.79	1.92	2.16	2.35	2.43	4.39
PESQ_ex1(M)	1.55	1.71	1.92	2.05	2.33	4.26
PESQ_ex2(F)	1.71	1.87	1.95	2.23	2.41	4.28
PESQ_ex1(F)	1.53	1.68	1.90	2.01	2.31	4.17



**Figure 8:** Quality improvement chart for male and female speakers for two proposed and spectral subtraction methods



**Figure 9:** Spectrogram of the speech signal with SNR = 10db (left) Reconstructing the speaker signal in the spectrogram space (right)



**Figure 10:** PESQ Input and Output Results for a Speaker Signal, Reconstructed Signal to SNR (Red) Noise Signal to SNR (Blue) Signal



is described in the following relations, the mask for each cluster is determined. The  $Mb$  binary mask is obtained from the following formula:

$$Mb(f) = \begin{cases} 1, & |y_{i1}(f)| < |y_{21}(f)| \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Where  $y_{i1}(f)$  and  $y_{21}(f)$  are predicted outputs that include different sources from the signal spectrum and  $f = 1, 2, \dots, F$  are shown for different frequencies. Then, using the Gaussian distribution function for each cluster, its Gaussian model is obtained. After determining the Gaussian distribution function and applying it to the GCT domain, the  $DC$  value is added to it, and the reconstructed signal is obtained. The block diagram of Figure 2 shows the general trend of the proposed method.

### 3.1. Improve signal quality in GCT space using cluster shifts to obtain dc value

In this method, the steps of the test process are the same as the previous test. However, the only difference is that after the step of adding GCT to the range of spectra resulting from the Gaussian mask, and reconstructed clusters are obtained by using clusters to add  $dc$  in a way that after shifting the clusters to the center of the time-frequency axis, we apply their sum to the dc part of the signal. In this case, unlike the previous method, all parts of the spectrum will be reconstructed and of good quality. Figure 3 shows the process of adding clusters to the center.

## 4. Experiments and Results

We will select the experiments from the 16-bit and 16 kHz TIMIT speech signals. These experiments have been performed for a large number of male and female speakers. Here are some of the results of the speakers for different experiments. Each experiment uses different speakers with different sentences.

In the first experiment after GCT and dc removal, by selecting a suitable threshold and determining the exact centers of the clusters, the binary mask related to the clusters is obtained using the energy of the clusters. In energetic places, a mask for each cluster appears. After determining the mask, the range of GCT is applied to it, and significant clusters are identified. Figure 4 shows the process of this experiment for a speech signal.

In the second experiment, after determining the GCT using the WK-means algorithm, after accurately determining the step frequency, we determine the centers of the frequency-time spectrum clusters [19]. After determining the centers using the Gaussian conversion function, we obtain the probability distribution function for clusters. Then, using the appropriate threshold, the mask corresponding to each cluster is observed. The relation of the function of two Gaussian variables is expressed in Equation 6.

$$p(x, y) = \frac{1}{2\pi \left| \sum \right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\nu - \mu\theta)^T \cdot \sum^{-1}(\nu - \mu\theta)}$$

(6)

Where in the above equation:

$$\nu = \begin{pmatrix} x \\ y \end{pmatrix}, \mu\theta = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \sum = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

(7)

In other words, the covariance matrix is estimated to be weighted based on the member elements of the clusters.

$$\sum = \frac{\sum_{i=1}^N w_i (x_i - \mu)(x_i - \mu)^T}{\sum_{i=1}^N w_i}$$

(8)

Where  $w_i$  is the weight of points in GCT. The frequency of the step using Equation 4 and the cluster centers was created using the wk-means algorithm and deleting the dc value in order to speed up and facilitate the calculation of the algorithm and are determined to improve the clarity of the peaks by selecting a suitable threshold. In Figure 5, the GCT analysis for all blocks of a speech signal related to the male speaker over a short period of time multiplied by a 20ms Hemingway window is shown.

After determining the centers of the clusters, the mask corresponding to each cluster is determined, after which the Gaussian mask transform function is obtained. Figure 6 shows the images of a speech signal block.

After determining the mask corresponding to the Gaussian transform function, each mask is applied to the GCT range, and the result of that is shown in Figure 7.

At this stage, we apply the phase to the mask resulting from the Gaussian function. The following equations represent the signal phase relationships.

$$\begin{aligned}\varphi &= \text{angle}(S(\omega_s \sin \theta, \omega_s)) \\ \psi &= \exp(\sqrt{j\varphi})\end{aligned}\tag{9}$$

In the above equations,  $S(\omega, \Omega)$  is the conversion of SGT from the signal  $s[n, m]$ , and  $\theta$  is the angle of rotation of the clusters from the GCT.

After this step, we will reconstruct the signal. First, using the inverse of the two-dimensional Gaussian transform, the inverse of the GCT is obtained. Then, by adding the spectrogram phase to the inverse of GCT, the inverse short-time fast transform (ISTFT) is taken. The criterion used to compare the results of improved speech quality is the PESQ criterion, and its value is 4.5 for full audio compliance, and its minimum for non-compliance is -0.5, which is in line with the MOS standard. Determining the PESQ score is like the MOS criterion, which is the listening score of the MOS sound quality between 1 and 5, which is

excellent quality to poor quality and poor sound quality.

In Table 1, the PESQ results for the male and female speaker signals are executed under the same conditions. In the first experiment, a binary mask was used to determine the clusters, and in the second experiment, after obtaining the Gaussian distribution function obtained from the cluster centers, we get the corresponding mask. The results of the above two experiments are shown for male and female speakers. Table 1 shows the results of the two experiments.

Table 2 shows the results of quality improvement for a male speaker and a female speaker compared to the two methods described in Section 3. First, in the first experiment, we do the process of improving speech quality by removing and adding dc in such a way that by selecting a suitable threshold, we can delete extra values and dc samples and after performing the algorithm, before reconstructing the signal, we add the deleted dc value. In this experiment, the cluster mask is obtained using the Gaussian distribution function. These tests are performed by adding white noise with different SNRs and in a pure signal mode for female and male speakers.

As can be seen from Table 2, the second experiment was more successful in the lower SNRs and the accuracy of the second experiment, in which the by means of clusters shift to the origin and the selection of the desired mask using the Gaussian distribution function is obtained, is more. The accuracy of the test has been reduced in terms of PESQ in low SNRs, but it is of good audio quality. In Figure 8, the acoustic noise from the ambient noise is acoustically added to the speech signal of both male and female speakers. In this figure, the blue, red, green and purple bars show the rate of improvement in the proposed method for male and female speakers, the rate of improvement in the differential method for male and female speakers respectively. A comparison of the proposed method algorithm with the spectral subtraction method with four babble noise, car, office, and exhibition, which is taken from the NoiseX database, has been done [36]. The spectral diffraction method is one of the first and most widely used methods in improving speech signal with noise. Noise spectra are

usually detected during a period of silence. Assuming that this estimate is the same for all noise signals and that it is constant and inconsistent with the original speech signal, the noise power spectrum can be reduced from noisy speech signal [33,34].

Figure 8 shows the improvement in speech quality when adding ambient noise to the speaker's speech. The horizontal axis of the diagram indicates the type of ambient noise, and the vertical axis indicates the improvement of male and female speech. Speech Improvement is the difference between the PESQ reconstructed signal and noisy PESQ signal.

According to this figure, it is observed that the proposed method is more successful than the spectral subtraction method for both female and male speakers, it has acceptable quality in terms of reconstruction and hearing. It has a reasonable recovery rate in different noisy environments with SNR of about 5db. However, in the spectral subtraction method in some noisy environments, it has a negative improvement rate. As a result, spectral subtraction has not been successful in low SNRs and has not improved speech reconstruction.

Also, because the female voice is lower and thinner than the male voice, the quality of speech in the female speaker has decreased. The main reason for this is overlapping with environmental noise and its impact on speakers.

Figure 9 shows the STFT signal of a speaker and its reconstruction using the proposed method in a noisy environment. In Figure 9 (left), the STFT signal of a speaker by adding white noise with SNR = 10db, and in Figure 9 (right), the STFT signal of a reconstructed speech signal are shown, respectively. According to the STFT speech signal in Figure 9, it is observed that the information from the noise is well separated in the proposed method.

Figure 10 shows the output and input PESQ diagrams for the proposed method. In this experiment, the dc value of the speech signal is obtained by adding clusters to the center. The output of the diagram is the reconstructed signal to the original signal, and its input is the ratio of the original signal to the noise signal. The figure below shows the process of

improving speech quality in white noise conditions with different SNRs. For noisy signals less than 10db, the PESQ is less than 2.5, but for high SNRs, the PESQ is higher than 3.5.

## 5. Conclusion

In this paper, a new method for improving the quality of speech was presented, in which the step frequency of each cluster was first calculated by the WK-means method. After obtaining the speech step frequency, we identify the exact center of the clusters in the GCT space and then, according to the correct diagnosis of the cluster centers, we attribute the Gaussian distribution to the clusters, and by using the GCT mask in the GCT space, the signal is separated from the noise. In cases where the GCT is not clear, it is possible to go directly to the reconstruction stage and apply the mask only to mosaics in which GCT has significant clustering. In the continuation of this research, this method can be used to separate the two speakers that have been combined and it also can be used in separating speakers in noisy situations.

## Conflict of Interest

The authors declare that there is no financial or commercial conflict of interest.

## References

- [1] Adnani AT., Dokami A., Morovati M., Fault detection in high speed helical gears considering signal processing method in real simulation. *Latin American Journal of Solids and Structures.*, 2016; 13: 2113-21140.
- [2] Kelareh, A. Y., Shahri, P. K., Khoshnevis, S. A., Valikhani, A., & Shindgikar, S. C., Dynamic Specification Determination using System Response Processing and Hilbert-Huang Transform Method. *International Journal of Applied Engineering Research.*, 2019; 14(22), 4188-4193.
- [3] Ahani, H., Familian, M., Ashtari, R., Optimum Design of a Dynamic Positioning Controller for an Offshore Vessel. *Journal of Soft Computing and Decision Support Systems.*, 2020; 7(1), 13-18.
- [4] Mohamed, A. F., Modir, A., Tansel, I. N., Uragun, B., Detection of Compressive Forces Applied to Tubes and

- Estimation of Their Locations with the Surface Response to Excitation (SuRE) Method. In 2019 9th International Conference on Recent Advances in Space Technologies., 2019; 83-88. IEEE.
- [5] Mohamed, A. F., Modir, A., Shah, K. Y., Tansel, I., Control of the Building Parameters of Additively Manufactured Polymer Parts for More Effective Implementation of Structural Health Monitoring (SHM) Methods. Structural Health Monitoring., 2019.
- [6] Sheikhsahrokhdehkordi, M., Goudarzi, N., Saffaraval, F., Mousavi sani, S., Tkacik, P., A TomoPIV Flow Field Study of NACA 63-215 Hydrofoil With CFD Comparison. In Fluids Engineering Division Summer Meeting. American Society of Mechanical Engineers., 2019; 59070: V004T04A039.
- [7] Mousavi sani, S., Goudarzi, N., Sheikhsahrokhdehkordi, M., Bisel, T., Dahlberg, J., Tkacik, P., Exploring and Improving the Flow Characteristics of an Empty Water Channel Test Section: The Application of TomoPIV and Flowrate Sensors for Whole-Flow-Field Visualization. In Fluids Engineering Division Summer Meeting. American Society of Mechanical Engineers., 2019; 59070: V004T04A040.
- [8] Izadi, V., Abedi, M., Bolandi, H., Verification of reaction wheel functional model in HIL test-bed. In 2016 4th International Conference on Control, Instrumentation, and Automation (ICCA) IEEE., 2016; 155-160.
- [9] Izadi, V., Abedi, M., Bolandi, H., Supervisory algorithm based on reaction wheel modelling and spectrum analysis for detection and classification of electromechanical faults. IET Science, Measurement & Technology., 2017; 11(8): 1085-93.
- [10] Izadi, V., Shahri, P. K., Ahani, H. A., Compressed-sensing-based compressor for ECG. Biomedical engineering letters., 2020; 6: 1-9.
- [11] Wang TT., F Quatieri T., Towards Interpretive Models for 2-D Processing of Speech. IEEE transactions on audio, speech, and language processing., 2012; 20: 2159-2173.
- [12] Wang TT., F Quatieri T., Two-dimensional speech-signal modeling. IEEE transactions on audio, speech, and language processing., 2012; 20: 1843-1856.
- [13] Wang TT., F Quatieri T., Multi-pitch estimation by a joint 2-D representation of pitch and pitch dynamics. In Eleventh Annual Conference of the International Speech Communication Association., 2010.
- [14] Wang TT., F Quatieri T., High-pitch formant estimation by exploiting temporal change of pitch. IEEE transactions on audio, speech, and language processing., 2009; 18: 171-186.
- [15] Wang TT., Exploiting pitch dynamics for speech spectral estimation using a two-dimensional processing framework. PhD diss., Massachusetts Institute of Technology., 2008.
- [16] Wang TT., Exploiting pitch dynamics for speech spectral estimation using a two-dimensional processing framework. PhD diss., Massachusetts Institute of Technology., 2008.
- [17] Griffin D., Jae L., Signal estimation from modified short-time Fourier transform. IEEE Transactions on Acoustics, Speech, and Signal Processing., 1984; 32: 236-243.
- [18] Abhijith MN., Ghosh PK., Rajgopal K., Multi-pitch tracking using Gaussian mixture model with time varying parameters and grating compression transform. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)., 2014; 1473-1477.
- [19] Jie L., Zhang G., Fu B., Hao Y., Multipitch tracking with continuous correlation feature and hybrid DBNS/HMM model. 11th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)., 2014; 218-221.
- [20] Jilt S., Manoj Kumar PA., Murthy HA., Pitch estimation from speech using grating compression transform on modified group-delay-gram. Twenty First National Conference on Communications (NCC)., 2015; 1-6.
- [21] Tony E., Bouvrie J., Poggio T., AM-FM demodulation of spectrograms using localized 2D max-Gabor analysis. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07., 2007; 4: IV-1061.
- [22] Karimi Shahri, P., Chintamani Shindgikar, S., HomChaudhuri, B., & Ghasemi, A. H. (2019, October). Optimal Lane Management in Heterogeneous Traffic Network. In Dynamic Systems and Control Conference (Vol. 59162, p. V003T18A003). American Society of Mechanical Engineers.
- [23] Shahri, P. K., Ghasemi, A. H., & Izadi, V. (2020). Optimal Lane Management in Heterogeneous Traffic Network Using Extremum Seeking Approach (No. 2020-01-0086). SAE Technical Paper.
- [24] Wang TT., F Quatieri T., Towards co-channel speaker separation by 2-D demodulation of spectrograms. In

- 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics., 2009; 65-68.
- [25] Petros M., F Kaiser J., F Quatieri T., Energy separation in signal modulations with application to speech analysis. *IEEE transactions on signal processing.*, 1993; 41: 3024-3051.
- [26] Rafieipour H, Zadeh AA, Mirzaei M. Distributed Frequent Itemset Mining with Bitwise Method and Using the Gossip-Based Protocol. *Journal of Soft Computing and Decision Support Systems.* 2020 May 7;7(3):32-9.
- [27] [27] Petros M., F Kaiser J., F Quatieri T., On amplitude and frequency demodulation using energy operators. *IEEE Transactions on signal processing.*, 1993; 41: 1532-1550.
- [28] [28] Petros M., F Quatieri T., F Kaiser J., Speech nonlinearities, modulations, and energy operators. In *Proc. ICASSP.*, 1991; 91: 421-424.
- [29] [29] Wang TT., F Quatieri T., High-pitch formant estimation by exploiting temporal change of pitch. *IEEE transactions on audio, speech, and language processing.*, 2009; 18: 171-186.
- [30] [30] Mingyang W., Wang D., J Brown G., A multipitch tracking algorithm for noisy speech. *IEEE Transactions on Speech and Audio Processing.*, 2003; 11: 229-241.
- [31] [31] F Quatieri T., *Discrete-time speech signal processing: principles and practice.* Pearson Education India., 2006.
- [32] [32] Srikanth V., Y Espy-Wilson C., An algorithm for speech segregation of co-channel speech. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing., 2009; 109-112.
- [33] [33] Jae S L., *Two-dimensional signal and image processing.* Ph.D Thesis., 1990.
- [34] [34] Harald G., E Nordholm S., Claesson I., Spectral subtraction using reduced delay convolution and adaptive averaging. *IEEE transactions on speech and audio processing.*, 2001; 9: 799-807.
- [35] [35] Kohei Y., Ogata S., Shimamura T., Spectral subtraction iterated with weighting factors in speech coding. *IEEE Workshop Proceedings.*, 2002; 138-140.
- [36] [36] Varga A., *The NOISEX-92 study on the effect of additive noise on automatic speech recognition.* ical Report, DRA Speech Research Unit., 1992.
- [37] [37] Surakanti, S. R., Khoshnevis, S. A., Ahani, H., & Izadi, V. (2019). Efficient Recovery of Structural Health Monitoring Signal based on Kronecker Compressive Sensing. *International Journal of Applied Engineering Research*, 14(23), 4256-4261.
- [38] [38] Taremi, R. S., Shahri, P. K., & Kalareh, A. Y. (2019). Design a Tracking Control Law for the Nonlinear Continuous Time Fuzzy Polynomial Systems. *Journal of Soft Computing and Decision Support Systems*, 6(6), 21-27.