Analysis of Multicomponent Speech-like Signals by Continuous Wavelet Transform-based Technique

Pradip Sircar, Keshava Prasad, Bandra Harshavardhan
Department of Electrical Engineering
Indian Institute of Technology Kanpur
KANPUR 208016, India

***Abstract*-A novel technique based on the continuous wavelet transform (CWT) is presented for analysis of multicomponent amplitude and frequency modulated signals. In the process, the multicomponent signal is decomposed in its constituents on the time-frequency plane and the analysis is carried out on individual components. It is demonstrated that the separation of components for analysis brings many advantages in the proposed method, viz., simplicity in procedure, and noise immunity. The developed technique is employed for analysis-synthesis of speech phonemes.**

I. Introduction

A variety of signals encountered in real life are multicomponent signals, where each component follows some characteristic amplitude and frequency (or phase) modulation laws [1]. The analysis of such signals involves estimation of the amplitude and frequency functions of each component of the signal.

Among the nonparametric approaches, there are several methods based on the energy operator, the Hilbert transform [2], or the Wigner distribution [3], which can estimate the time-variant amplitude and frequency functions of a monocomponent signal. However, the methods often fail for a multicomponent signal due to presence of cross terms. The parametric approach of analysis of a multicomponent signal often requires optimization of a multivariate cost function [1], and the analysis may become inaccurate in presence of noise [4].

In this paper, we present an intelligent method based on the continuous wavelet transform (CWT) for analysis of a multicomponent signal [5]. The underlying principle of the proposed method is as follows: The time-frequency representation (TFR) of a multicomponent signal as obtained by the CWT shows that the signal energy tends to concentrate in several regions in the time-frequency plane [6]. Each such region of energy concentration can be treated as a component of the composite signal, and each component is characterized by instantaneous frequency (or phase) and amplitude functions. It is the purpose of this paper to develop a procedure to analyze each component of the composite signal independently, and to extract its instantaneous frequency and amplitude functions. In the following sections, we present a method based on the CWT for analysis and synthesis of multicomponent signals, where the analysis is carried out by first separating each component on the time-frequency / time-scale plane.

For illustration the proposed method, we consider speech signals which can be represented by the sum of complex amplitude modulated (AM) and frequency modulated (FM) sinusoids corresponding to voiced and unvoiced phonemes respectively [7-9]. The nonlinear problem of estimation of the frequency parameters of the multicomponent signal turns out to be a simple problem when each component of the signal is separated on the time-frequency plane. The approach of separation of components of a multicomponent signals brings many advantages in the process of analysis as explained in the sequel.

II. Continuous Wavelet Transform

Let the analytic signal $s(t)$ be expressed as

$$s(t) = A_s(t)\exp(j\phi_s(t)) \tag{1}$$

such that, $A_s(t) \geq 0$ and $\phi_s(t) \in [0, 2\pi)$. It is assumed that the signal is asymptotic with the condition

$$\left|\frac{d\phi_s}{dt}\right| >> \left|\frac{1}{A_s}\frac{dA_s}{dt}\right| , \tag{2}$$

which essentially means that the signal is oscillating due to the phase term [6]. The instantaneous angular frequency $\omega_s(t)$ is defined by

$$\omega_s(t) = \frac{d}{dt}(\phi_s(t)) \tag{3}$$

It is assumed that the signal is locally monochromatic, i.e., $\omega_s(t)$ varies slowly with time. One can estimate the instantaneous angular frequency (and analytic amplitude) of the signal when the above assumptions are valid [10].

The CWT of the signal $s(t)$ having finite energy is defined as follows:

$$T_s(b,a) = \int_{-\infty}^{\infty} s(t)\, \psi_{b,a}^*(t)dt \tag{4}$$

where $\psi_{b,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$ is a dilated and shifted version of $\psi(t)$, a is the scale parameter and b is the time parameter. It is assumed that the wavelet satisfies the admissibility condition, that is,

$$c_\psi = \int_{-\infty}^{\infty} \left|\Psi(a\omega)\right|^2 \frac{da}{a} \tag{5}$$

is a nonzero finite constant [6,11].

For the analysis of a real signal, we use an analytic wavelet $\psi(t)$ of the form

$$\psi(t) = g(t)\exp(j\xi t) \tag{6}$$

where $g(t)$ is a real symmetric window function [11].

The choice of an analytic wavelet allows us to work essentially with the analytic signal associated with the real signal [10]. The Fourier transform of the wavelet $\psi(t)$ is related to Fourier transform of the window $g(t)$ as shown

$$\psi(\omega) = G(\omega - \xi) \tag{7}$$

We choose $G(\omega)$ to be band-limited by the angular frequency ξ, which ensures that $\psi(t)$ is an analytic wavelet. Assuming that the signal $s(t) = A_s(t)\exp(j\phi_s(t))$ is asymptotic and locally monochromatic, the CWT $T_s(b,a)$ can be derived as given by

$$T_s(b,a) = \sqrt{a}A_s(b)\exp(j\phi_s(b))G[a(\omega - \omega_s(b))] \tag{8}$$

where $\omega = \frac{\xi}{a}$ and $\omega_s = \frac{d\phi_s}{dt}$ [11]. The normalized TFR is expressed as

$$\frac{1}{a}\left|T_s\left(b,\frac{\xi}{\omega}\right)\right|^2 = A_s^{\,2}(b)\left|G\left[\xi\left(1-\frac{\omega_s(b)}{\omega}\right)\right]\right|^2 \tag{9}$$

which attains its maximum value at the angular frequency

$$\omega(b) = \frac{\xi}{a(b)} = \omega_s(b), \tag{10}$$

assuming $G(\omega)$ to be maximum at $\omega = 0$. The ridge of the transform where the TFR attains its maximum is defined by the time-angular frequency points $(b, \omega(b))$. Thus, the instantaneous angular frequency of the signal, $\omega_s(b)$ is given by the angular frequency on the ridge. The analytic amplitude is given by

$$A_s(b) = \frac{\frac{1}{\sqrt{a}}\left|T_s(b,a)\right|}{\left|G(0)\right|} \tag{11}$$

computed on the ridge. The above equation follows directly from (9).

For a multicomponent signal $x(t)$ expressed as

$$\begin{aligned} x(t) &= \sum_{i=1}^{M} s_i(t) \\ &= \sum_{i=1}^{M} A_i(t)\exp(j\phi_i(t)) \end{aligned} \tag{12}$$

with $A_i(t) \geq 0$ and $\phi_i(t) \in [0, 2\pi)$, the wavelet transform is given by

$$T_x(b,a) = \sqrt{a}\sum_{i=1}^{M} A_i(b)\exp(j\phi_i(b))\Psi[a\omega_i(b)] \tag{13}$$

where (7) and (8) are utilized.

Since $\Psi(\omega)$ is maximum at $\omega = \xi$, the wavelet transform modulus is localized near the M ridges with

$$\frac{\xi}{a_i(b)} = \omega_i(b) = \frac{d}{dt}(\phi_i(t)) \tag{14}$$

where $a_i(b)$ is the scale parameter on the i th ridge.

III. Analysis of Complex AM and FM Sinusoidal Signals

The complex sequence $x[n]$ consisting of M single tone AM signals is represented by

$$x[n] = \sum_{i=1}^{M} A_{ci}[1 + \mu_i \exp(j\nu_i n)]\exp(j\omega_i n) \tag{15}$$

where $A_{ci} = A_i\exp(j\phi_i)$, A_i are the carrier amplitudes, ϕ_i are the carrier phases, μ_i are the modulation indices, ω_i are the carrier angular frequencies, and ν_i are the modulating angular frequencies [7,8].

We determine the carrier and modulating angular frequencies of the modeled signal (15) from the time-frequency plot by employing the CWT-based technique as presented in the previous section. The carrier amplitude and phases, and modulation indices are estimated next as follows:
By utilizing the sequence $\{x[n]; n = 0,1,...,N-1\}$ and substituting $\xi_i = \exp(j\omega_i)$, $\varsigma_i = \exp(j\nu_i)$, we form the matrix equation,

$$\begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix} = \begin{bmatrix} 1 & \cdots 1 & 1 & \cdots & 1 \\ \xi_1 & \cdots \xi_M & \xi_1\varsigma_1 & \cdots & \xi_M\varsigma_M \\ \xi_1^{\,2} & \cdots \xi_M^{\,2} & (\xi_1\varsigma_1)^2 & \cdots & (\xi_M\varsigma_M)^2 \\ \vdots & \vdots & \vdots & & \vdots \\ \xi_1^{\,N-1} & \cdots \xi_M^{\,N-1} & (\xi_1\varsigma_1)^{N-1} & \cdots & (\xi_M\varsigma_M)^{N-1} \end{bmatrix} \begin{bmatrix} A_{c1} \\ \vdots \\ A_{cM} \\ A_{c1}\mu_1 \\ \vdots \\ A_{cM}\mu_M \end{bmatrix} \tag{16}$$

which can be solved in the least squares sense to find the complex carrier amplitude and modulation index parameters.

The complex sequence $x[n]$ consisting of M single tone FM signal is represented by

$$x[n] = \sum_{i=1}^{M} A_{ci}\exp\{j[\omega_i n + \beta_i\cos(\nu_i n)]\} \tag{17}$$

where $A_{ci} = A_i\exp(j\phi_i)$, A_i are the carrier amplitudes, ϕ_i are the carrier phases, β_i are the modulation indices, are the carrier angular frequencies, and υ_i are the modulating angular frequencies [7,9].

The carrier and modulating angular frequencies of the modeled signal (17), together with the modulation indices are determined from the time-frequency plot by applying the CWT-based technique. The carrier amplitudes and phases are estimated next by utilizing the sequence $\{x[n]; n = 0,1,...,N-1\}$. We form the following matrix equation by substituting $W_i^n = \exp\{j[\omega_i n + \beta_i\cos(\nu_i n)]\}$,

$$\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} = \begin{bmatrix} W_1^{\,0} & W_2^{\,0} & \cdots & W_M^{\,0} \\ W_1^{\,1} & W_2^{\,1} & \cdots & W_M^{\,1} \\ \vdots & \vdots & & \vdots \\ W_1^{\,N-1} & W_2^{\,N-1} & \cdots & W_M^{\,N-1} \end{bmatrix} \begin{bmatrix} A_{c1} \\ A_{c2} \\ \vdots \\ A_{cM} \end{bmatrix} \tag{18}$$

which can be solved in least squares sense to obtain the complex carrier amplitudes.

IV. Simulation Study

For simulation we use the Morlet wavelet as given by $\psi(t) = \exp(-t^2/2\sigma^2)\exp(j\xi t)$, with $\sigma^2 = 1$ and $\xi = 7\pi/3$ [6, 11]. The choices of σ^2 and ξ ensures that the wavelet is approximately analytic and it is an admissible wavelet. The CWT is approximated by computing the transform for the scales chosen on a discrete grid. We select 32 intermediate scales over each octave $[2^k, 2^{k+1}]$, k be an integer [5]. Implementing the CWT as a combination of the fast Fourier transform and the chirp z-transform, a substantial reduction of computations is achieved [12].

Example 1

We consider the complex AM signal $x[n]$ as given by (15) with the following set of parameters: M=3,

$$A_1 = 1.0, \phi_1 = \pi/4, \omega_1 = 0.35, \mu_1 = 0.3, \upsilon_1 = 0.005,$$
$$A_2 = 1.5, \phi_2 = \pi/6, \omega_2 = 0.25, \mu_2 = 0.5, \upsilon_2 = 0.015,$$
$$A_3 = 1.0, \phi_3 = \pi/8, \omega_3 = 0.15, \mu_3 = 0.0, \upsilon_3 = 0.000.$$

The TFR of the signal as obtained by the CWT is shown in Fig. 1. The carrier and modulating angular frequencies of each component of the signal can be estimated directly from the TFR. When zero mean white Gaussian (ZMWG) noise is added to the signal setting the signal-to-noise ratio (SNR) at 10 dB and 0 dB, the TFR of the signal gets corrupted as shown in Fig. 2. It can be inferred from the figure that the estimation of frequency parameters is robust to noise, whereas the estimation of amplitude parameters can be degraded in presence of noise.

Example 2

In this example, we consider the voiced speech phonemes for analysis and synthesis. The TFRs of the voiced speech phonemes /u/ and /a/ are shown in Figs. 3 and 5 respectively. The time-frequency plot of the corresponding phoneme reveals that /u/ has one AM sinusoidal component, whereas /a/ has two AM carriers and one unmodulated carrier.

We separate each region in the time-frequency plane where a component of the signal is supported, and process each region masking the rest of the transform [5]. A threshold of magnitude is used in the TFR to suppress the effects of noise and cross terms [5]. Once a region is separated corresponding to a component of the signal, the extraction of the ridge and estimation of time-variant frequency and amplitude functions on the ridge can be carried out in a straight forward manner.

The original and reconstructed signals for the voiced speech phonemes /u/ and /a/ are plotted in the Figs. 4 and 6 respectively. A comparison of two plots reveals that the developed technique together with the complex AM signal model can implement analysis –synthesis of voiced speech signals with reasonable accuracy.

Example 3

We consider next the complex FM signal $x[n]$ as given by (17) with the following set of parameters: $M = 2$,

$$A_1 = 1.0, \phi_1 = \pi/4, \omega_1 = 0.40, \beta_1 = 0.05, \nu_1 = 0.02,$$
$$A_2 = 2.0, \phi_2 = \pi/6, \omega_2 = 0.15, \beta_2 = 0.05, \nu_2 = 0.01.$$

The TFR of the signal is shown in Fig. 7. The carrier and modulating angular frequencies and the modulation indices can be estimated directly from the TFR. When the ZMWG noise is mixed with the signal setting the SNR-levels at 10 dB and 0 dB, the TFR gets corrupted as shown in the Fig. 8. It can be observed from the figure that the frequency parameters can still be estimated with good accuracy. However, the estimation of amplitude parameters can be less accurate in presence of noise.

Example 4

In this example, we consider the unvoiced speech phonemes for analysis and synthesis. The TFRs of the unvoiced speech phonemes /q/ and /v/ are shown in Figs. 9 and 11 respectively. The time-frequency plot of the corresponding phoneme shows that /q/ has two FM carriers and three unmodulated carriers, whereas /v/ has three FM carriers and three unmodulated carriers.

We estimate the time-variant frequency and amplitude functions on the ridges of the TFR. Once a region is separated corresponding to a component of the signal, the estimation becomes a simple problem. A comparison of original and reconstructed signals for unvoiced speech phonemes /q/ and /v/ as shown in Figs 10 and 12 respectively shows that the CWT-based technique together with the complex FM model can effectively implement analysis-synthesis of unvoiced speech signals.

V. Conclusion

Since the TFR of a multicomponent signal shows energy concentration in different regions of the time-frequency plane, the approach based of the CWT as presented in this paper, is found to be a natural way to decompose a signal into its constituent components. The separation of component of a composite signal leads to simple procedure for estimation of characteristic functions of modeled signal.

Utilizing the approach developed here, we demonstrated the analysis-synthesis of speech signals which are modeled as the complex AM sinusoidal signal and the FM sinusoidal signal corresponding to voiced and unvoiced speech phonemes respectively.

References

[1] B. Friedlander, J.M. Francos, "Estimation of amplitude and phase parameters of multicomponent signals," *IEEE Trans. on Signal Processing,* Vol. 43, No. 4, pp. 917-926, Apr. 1995.

[2] A. Potamianos, P. Maragos, "A comparison of the energy operator and Hilbert transform approach to signal and speech demodulation," *Signal Processing,* Vol. 37, No. 1, pp. 95-120, Jan. 1994.

[3] F. Hlawatsch, W. Kozek, "Time-Frequency projection filters and time-frequency signal expansions," *IEEE Trans. on Signal Processing,* Vol. 42, No. 12, pp. 3321-3334, Dec. 1994.

[4] S. Mukhopadhyay, P. Sircar, "Parametric modeling of non-stationary signals: A unified approach," *Signal Processing,* Vol. 60, No. 2, pp. 135-152, Apr. 1997.

[5] K. Prasad, P. Sircar, "Analysis of multicomponent non-stationary signals by continuous wavelet transform method," *Proc. IEEE Int. Symp. on Intelligent Signal Processing,* Faro, Portugal, Sept 1-3, 2005.

[6] R. Carmona, W.-L. Hwang, B. Torresani, *Practical Time-Frequency Analysis: Gabor and Wavelet Transforms with an Implementation in S*, Academic Press, San Diego, 1998..

[7] P. Sircar, R.K. Saini, "Parametric modeling of speech," *First Online Symp. for Electronics Engineers*, www.techonline.com , Nov. 21, 2000.

[8] P. Sircar, M.S. Syali, "Complex AM signal model for non-stationary signals," *Signal Processing,* Vol. 53, No. 1, pp. 35-45, Aug. 1996.

[9] P. Sircar, S. Sharma, "Complex FM signal model for non-stationary signals," *Signal Processing,* Vol. 57, No. 3, pp. 283-304, Mar. 1997.

[10] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998..

[11] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martiner, P. Tchamitchian, B. Torresani, "Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies," *IEEE Trans. on Information Theory,* Vol. 38, No. 2, pp. 644-664, Mar. 1992.

[12] D.L. Jones, R.G. Baraniuk, "Efficient approximation of continuous wavelet transforms," *Electronic Letters,* Vol. 27, No. 9, pp. 748-750, Apr. 1991.

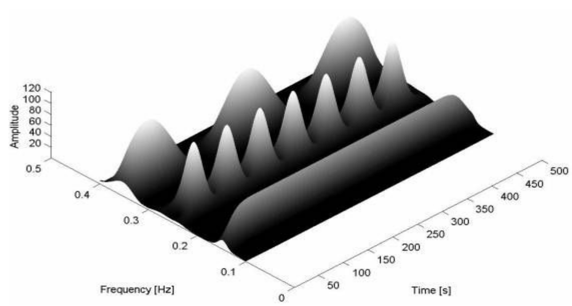

Fig. 1a Time-frequency representation of a 3-component AM signal

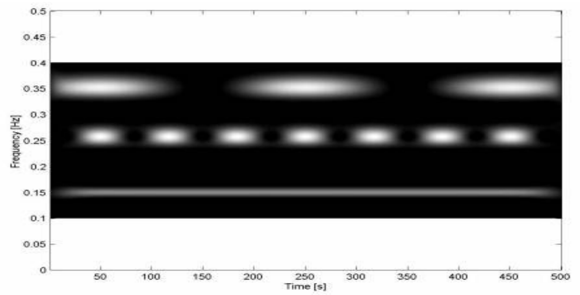

FIG. 1B CWT MAGNITUDE PLOT ON THE TIME-FREQUENCY PLANE

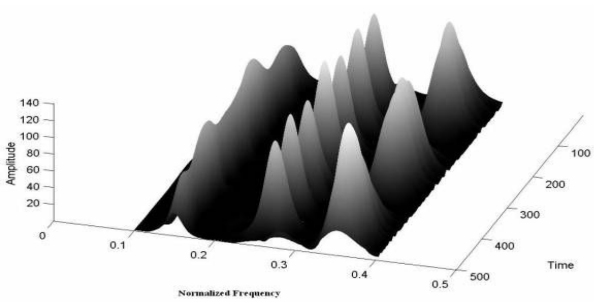

FIG. 2A TIME-FREQUENCY REPRESENTATION AT 10 dB SNR

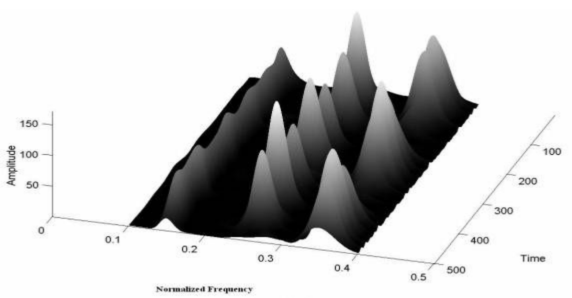

FIG. 2B TIME-FREQUENCY REPRESENTATION AT 0 dB SNR

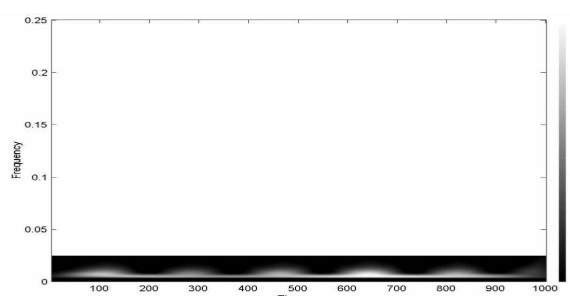

FIG. 3 TFR OF VOICED SPEECH PHONEME /U/

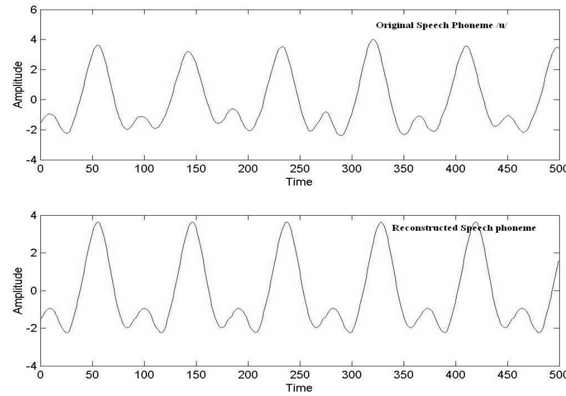

FIG. 4 ORIGINAL AND RECONSTRUCTED SIGNALS

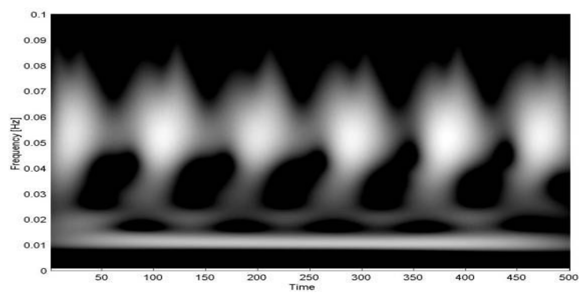

FIG. 5 TFR OF VOICED SPEECH PHONEME /A/

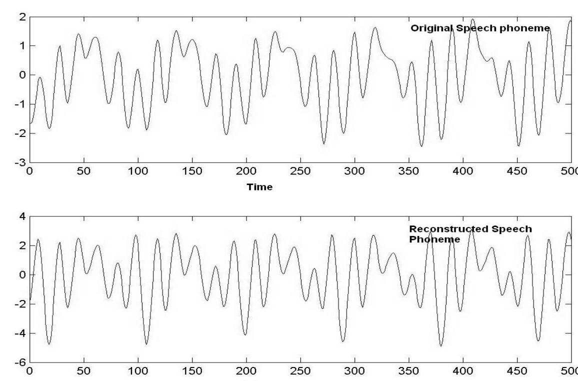

FIG. 6 ORIGINAL AND RECONSTRUCTED SIGNALS

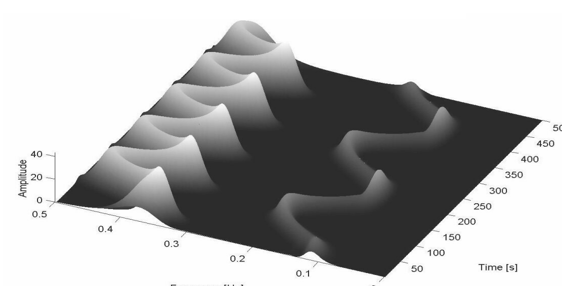

FIG. 7A TIME-FREQUENCY REPRESENTATION OF A 2-COMPONENT FM SIGNAL

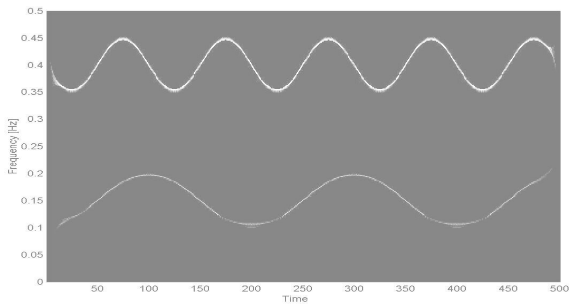

FIG. 7B RIDGES OF THE TRANSFORM

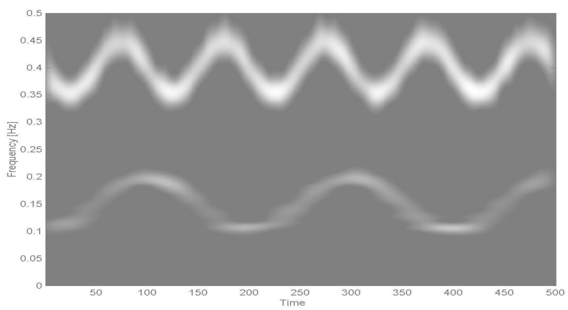

FIG. 8A TFR AT 10 dB SNR

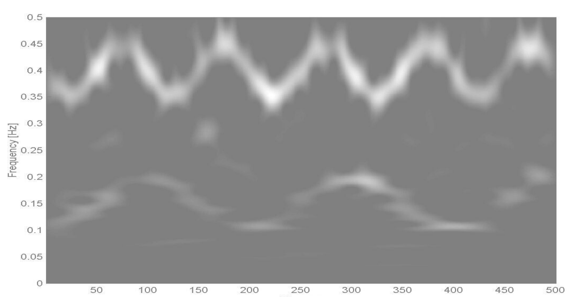

FIG. 8B TFR AT 0 dB SNR

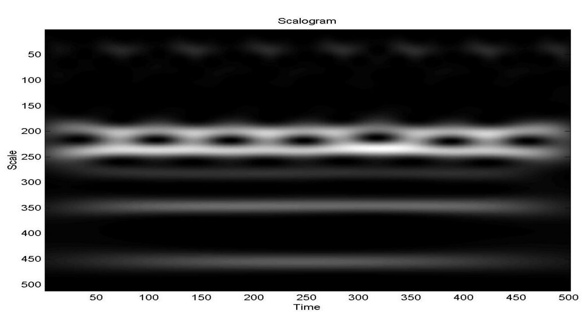

FIG. 9 TFR OF UNVOICED SPEECH PHONEME /Q/

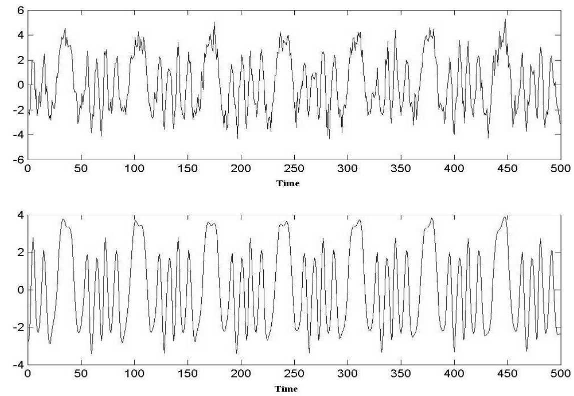

FIG. 10 ORIGINAL AND RECONSTRUCTED SIGNALS

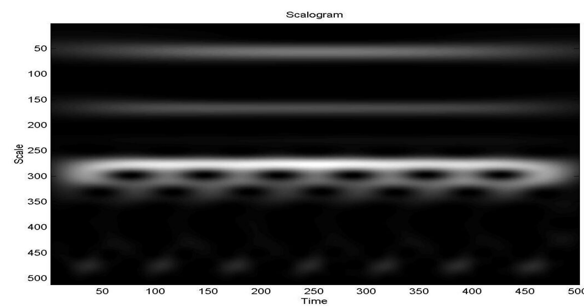

FIG. 11 TFR OF UNVOICED SPEECH PHONEME /V/

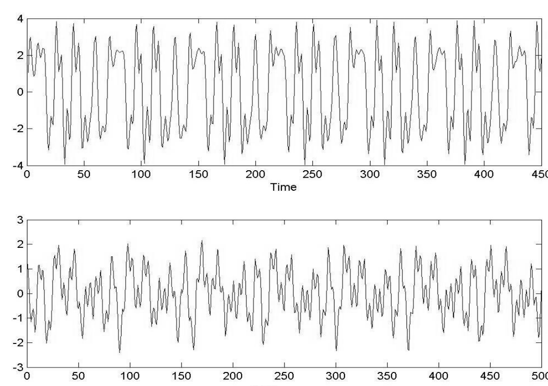

FIG. 12 ORIGINAL AND RECONSTRUCTED SIGNALS