International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 5 (2015), pp. 127-132
http://dx.doi.org/10.14257/ijsip.2015.8.5.13

Speech Signal Analysis Using Concentrated Spectrogram Method

Dr. G. Manmadha Rao and N. Srinivasa Rao

GMR institute of Technology
manmadharao.g@gmail.com
cnu472@gmail.com

Abstract

A new method of energy distribution estimation in the joint time-frequency domain using the Channelized Instantaneous Frequency (CIF) and Local Group Delay (LGD) is proposed. The signal energy distribution is estimated by discarding and displacement of energy parts. The signal energy leads to high concentrated distribution in the time-frequency domain due to the relocation of the CIF and LGD values.

In addition to this, a channelized instantaneous bandwidth and local group duration are used to remove undesired energy part. The channelized instantaneous bandwidth and local group duration express a local stretching of the signal in frequency and time respectively. This method is being used for speech signal analysis.

Keywords: *Non-Stationary Signals, Time-Frequency Representation, Short Time Fourier Transform, Cross-Spectral Method, Channelized Instantaneous Frequency, Local Group Delay*

1. Introduction

Time-frequency methods are well suited to the analysis of non-stationary multicomponent FM signals, such as speech signals. The method is based on group delay and instantaneous frequency surfaces computed from the Short Time Fourier Transform (STFT). In general, there are two characteristic curves in the frequency-time plane: one which gives the instantaneous frequency as a function of time and the other which gives the group delay time as a function of frequency. Many different methods have been proposed for analyzing time-varying signals. Such as Amplitude Maximum of the Spectrum (MS), Amplitude Maximum of the Envelope (ME), Moving Window Method (MWM) and Modified Moving Window Method (MMWM) [1, 2]. In this Modified Moving Window Method is used for calculation of classical spectrogram.

Both CIF and LGD are components of the gradient of the STFT complex phase. Other components are the signed channelized instantaneous bandwidth and signed local group duration. All mentioned components of the gradient of the STFT complex phase are used in the presented approach [1]. The concentrated spectrogram is calculated using CIF and LGD values.

2. Non-stationary Signals

A stationary signal is one whose frequency doesn't change over time; *e.g.*, deterministic and random signals. On the contrary, you have non-stationary signals where frequencies change over time [3]. Non-stationary signals are divided into continuous and transient types. Transient signals are defined as signals which start and end at zero level and last a finite amount of time. The synthetic FM two-mono component signal is

$$y_{s1}(t) = \exp(j(2\pi t f_a + 0.25 f_d \sin(4\pi t)/\pi)) + \exp(j(2\pi t f_b - 0.25 f_d \sin(4\pi t)/\pi))$$

Where f_a =300Hz, f_b =700Hz and f_d =150Hz. In general, the multicomponent complex signal waveform can be represented by the following model:

ISSN: 2005-4254 IJSIP

$$y(t) = \sum_{m=1}^{N} b_m(t)\exp(j\varphi_m(t))$$

Where N is a number of monocomponents, $b_m(t)$ and $\varphi_m(t)$ represents envelope and instantaneous phase of m^{th} monocomponent.

3. Time-frequency Representation

The two classical representations of a signal are the time-domain representation $y(t)$ and the frequency-domain representation $S(f)$. In both forms, the variables t and f are treated as mutually exclusive. Consequently, each classical representation of the signal is non-localized with respect to the excluded variable; that is, the frequency representation is essentially averaged over the values of the time representation at all times, and the time representation is essentially averaged over the values of the frequency representation at all frequencies.

3.1 Short Time Fourier Transform

The Short-Time Fourier Transform (STFT), or alternatively short-term Fourier transform, is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. The short-time Fourier transform is derived in the following manner:

$$Y(t,\omega) = A(t,\omega)\exp(j\varphi(t,\omega)) = \int_{-\infty}^{\infty} y(\tau+t)h^*(-\tau)\exp(-j\omega\tau)d\tau \qquad .....(1)$$

Where complex conjugate is denoted by an asterisk,

$$A(t,\omega) = |Y(t,\omega)| \text{ and } \varphi(t,\omega) = \arg\{Y(t,\omega)\},\ \ A(t,\omega), \varphi(t,\omega) \in R$$

The complex waveform, $y(t)$ should have non zero values and has to be differentiable in every instant, $Y(t,\omega)$ means resultant STFT and $h(t)$ represents an analyzing window function. $A(t,\omega)$ and $\varphi(t,\omega)$ denote accordingly amplitude and phase instantaneous spectra.

3.2 Cross Spectral Method

Cross spectral analysis allows us to determine the relationship between two time series as a function of frequency. The method is on the basis of derivatives with respect to both time and frequency in STFT phase and re-mapping of the STFT surface. The phase derivatives are computed as the arguments of cross-spectral surfaces, which are the product of the STFT of the signal and the complex conjugate of the STFT of the signal delayed in time and/or frequency [10]. While phase derivatives may be estimated by other methods, the primary advantage of the cross-spectral representation is that there is no need to phase unwrap the STFT surface to resolve the discontinuities of the principle-value representation of the argument function.

In the cross-spectral processes, the spectral magnitude is not used directly in the computation of frequency or group delay estimates, but it does play an important role in calculating average or mean frequencies [3, 5]. Spectral amplitude is also used as a criterion for isolation and estimation of the individual FM signal components of multicomponent signals. This criterion is stated as a separability condition under which these individual signal components do not interfere with each other on the STFT surface.

3.3 Channelized Instantaneous Frequency & Local Group Delay

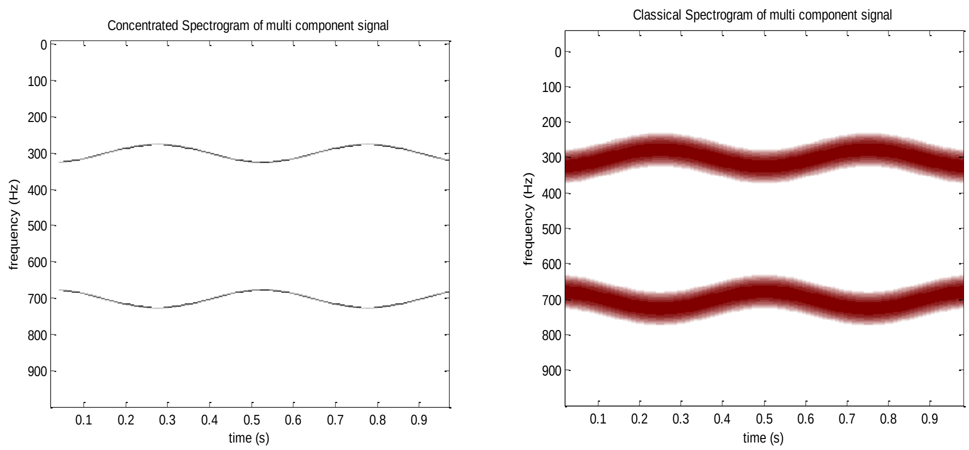
This section relates the widely used concepts of instantaneous frequency and group-delay [9]. Instantaneous frequency is defined as the time derivative of the instantaneous phase (1). Instantaneous Frequency (IF) is one of the basic signal parameter to provide important information of time-varying spectral changes in non-stationary signals [11]. The advantages of using the spectrogram for the TFR in this multicomponent signal IF estimation approach are its simplicity and absence of cross-terms.Group delay is a useful measure of time distortion, and is calculated by differentiating with respect to frequency. The group delay is a measure of the slope of the phase response at any given frequency. Presented in this method, subsequently for each locus of STFT for each locus (t,ω) of STFT corrected localization is estimated in the time-frequency plane by Channelized Instantaneous Frequency (CIF) and Local Group Delay (LGD)[1, 2]. They are expressed respectively:

$$\Omega(t,\omega) = \frac{\partial}{\partial t}\varphi(t,\omega)$$

$$\theta(t,\omega) = -\frac{\partial}{\partial \omega}\varphi(t,\omega) \qquad .....(2)$$

And are the obtained new localizations as follows:

$$(t,\omega) \rightarrow (t + \theta(t,\omega)/(2\pi), \Omega(t,\omega)) \qquad .....(3)$$



(A) (B)

Figure 1. Energy Distribution of Test Signal in the Time-frequency Domain: A) Classical Spectrogram, B) Concentrated Spectrogram

Where t and ω mean accordingly time and angular frequency. CIF is denoted by $\Omega(t,\omega)$ and LGD is expressed as $\theta(t,\omega)$. The new distribution of energy is called concentrated spectrogram [7, 8]. In Figure 1 classical and concentrated spectrograms of a synthetic FM two-monocomponent signal are presented.

4. Number of Degrees of Freedom Density

The short-time Fourier transform can be used assign local bandwidth in every instant and in every output channel (1), similarly as the channelized instantaneous frequency [8, 9]. For both continuous time and frequency it is called the channelized instantaneous bandwidth and can be calculated as follows:

$$B(t,\omega) = \frac{1}{2\pi}\left|\frac{\partial}{\partial t}\Lambda(t,\omega)\right| \qquad .....(4)$$

Where $\Lambda(t,\omega) = \ln(A(t,\omega))$ and $\ln()$ is the complex natural logarithmic function. The local group duration can be defined in the time-frequency domain:

$$T(t,\omega) = \frac{1}{2\pi}\left|\frac{\partial}{\partial\omega}\Lambda(t,\omega)\right| \qquad .....(5)$$

The channelized instantaneous bandwidth and the local group duration express a local stretching of the signal respectively in frequency and in time. Then a number of degrees of freedom density (distribution; NDFD) can be estimated by the following formula [12]:

$$\chi(t,\omega) = B(t,\omega)T(t,\omega)$$

Where $\chi(t,\omega)$ is a number of degrees of freedom density distributed in the joint time-frequency domain. In order to distinguish from the global number of degrees of freedom NDFD is given by $\chi(t,\omega)$. NDFD for the test FM chirp signal is presented in Figure 2.

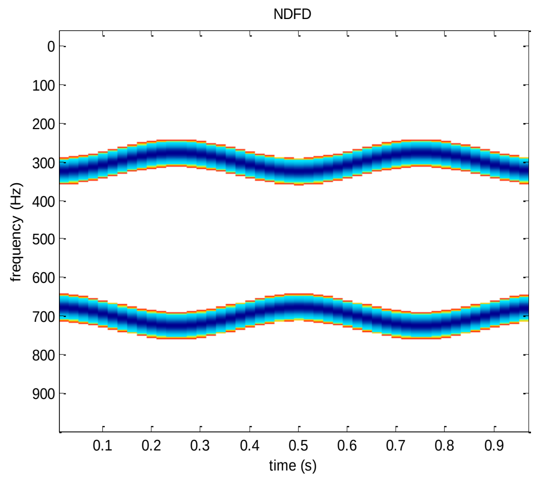
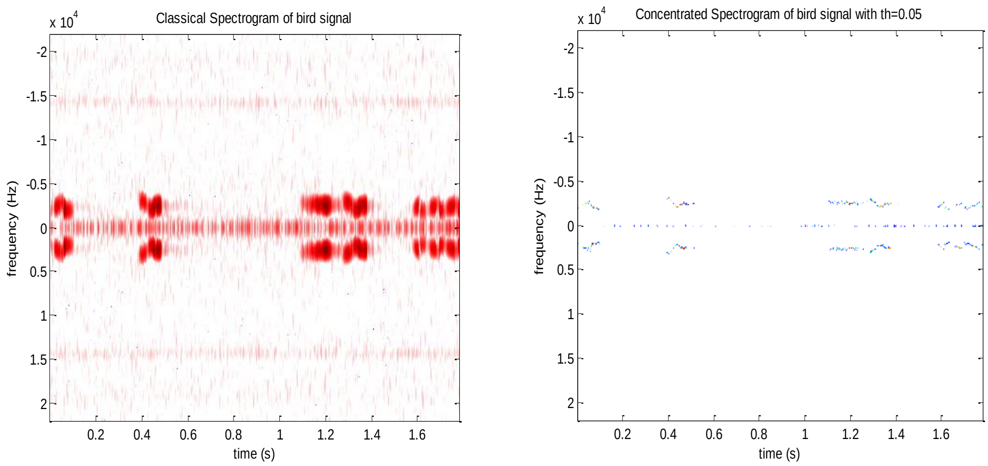


Figure 2. Number of Degrees of Freedom Density of Test Signal

5. Speech Signal Analysis

In speech signal analysis to estimate the classical and concentrated spectrogram of speech signals, such as bird, voice and wind signals [13]. In Fig.3 Classical and concentrated spectrogram of various speech signals presented.



(A) **(B)**

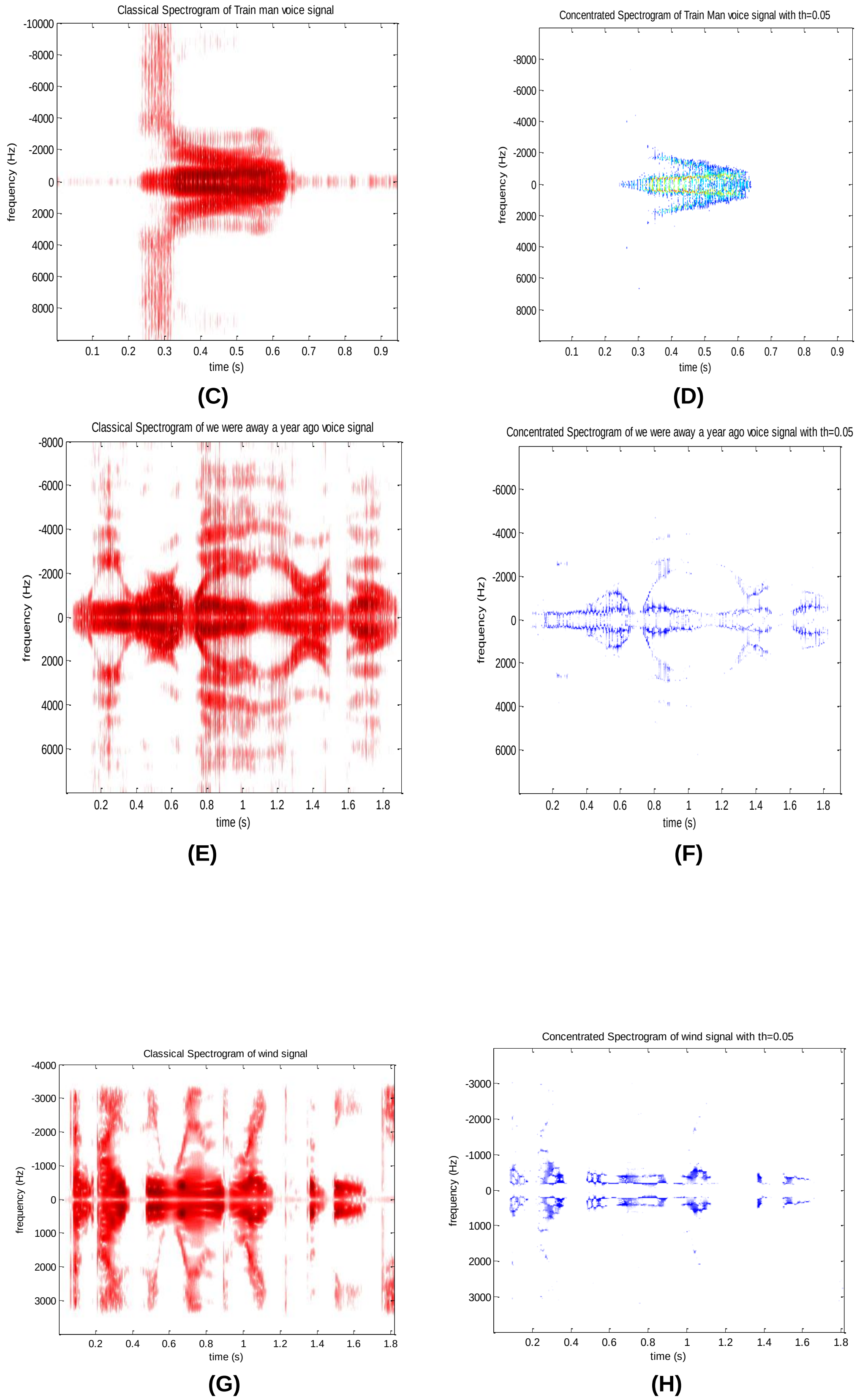


Figure 3. Classical Spectrograms of A (Bird Signal), C (Train Man Signal), E (Voice Signal), G (Wind Signal) and Concentrated Spectrograms of B (Bird Signal), D (Train Signal), F (Voice Signal), H (Wind Signal)

6. Conclusion

A new aspect of this method is the usage of Number of Degree of Freedom Density (NDFD) to allow the distribution of energy in the time frequency domain and for separation of energy into two parts. NDFD distribution is obtained as a product of the channelized instantaneous bandwidth and local group duration. Some part of energy, where NDFD values are small, is referred to as the attractogram. The second part of energy is treated as an irrelevant effect of STFT and it is strongly dependent on Heisenberg Gabor principle. The energy distribution of proposed concentrated spectrogram method is highly concentrated and accurate compared to classical spectrogram for various signals.

References

[1] K. Czarnecki and M. Moszynski, "Usage of concentrated spectrogram for analysis of acoustical signals", Proceedings of the 37th Scandinavian Symposium on Physical Acoustics, **(2014)** February 2-5.
[2] K. Kodera, R. Gendrin and C. D. Villedary, "Analysis of time-varying signals with small BT values", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, no. 1, **(1978)**, pp. 64–76.
[3] D. J. Nelson, "Cross-spectral methods for processing speech", Journal of the Acoustical Society of America, vol. 110, no. 5, **(2001)**, pp. 2575–2592.
[4] M. H. Ackroyd, "Short time spectra and time-frequency energy distribution," J. Acoust. SOC. Amer., vol. 50, **(1971)**, pp. 1229-1231.
[5] S. A. Fulop and K. R. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications", Journal of the Acoustical Society of America, vol. 119, no. 1, **(2006)**, pp. 360–371.
[6] D. J. Nelson and D. C. Smith, "Linear model for TF distribution of signals", IEEE Transaction on Signal Processing, vol. 54, no. 9, **(2006)**.
[7] K. Czarnecki and M. Moszynski, "Using concentrated spectrogram for analysis of audio acoustic signals Hydroacoustics", vol. 15, **(2012)**, pp. 27–32.
[8] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals", Proceedings of the IEEE, vol. 80, **(1992)**, pp. 520–538.
[9] D. H. Friedman, "Instantaneous-frequency distribution vs. time: An interpretation of the phase structure of speech," in Proceedings of the IEEEICASSP, **(1985)**, pp. 1121–1124.
[10] D. J. Nelson and W. Wysocki, "Cross-spectral methods with an application to speech processing", in Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms, Denver, CO, **(1999)** July.
[11] M. H. Ackroyd, "Instantaneous and time varying spectra-An introduction," J. Radio and Electron. Eng., vol. 39, no. 3, **(1970)**.
[12] D. Gabor, "Theory of communication", J. Inst. Elect. Eng., vol. 93, **(1946)**, pp. 429–457.
[13] www.mathworks.com/file exchange.

