SPEECH SIGNALS VISUALIZATION

Vladimír Sedlák, Daniela Ďuračková

Department of Microelectronics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Ilkovičova 3, 812 19 Bratislava, Slovakia

E-mail: vladimir.sedlak@stuba.sk

Received 30 April 2011; accepted 15 May 2011.

1. Introduction

This paper presents the possibility of visualization of non-stationary signals mainly through mathematical transformations (Short Time Fourier Transform, Wavelet Transform). The result is a comparison of methods. We take into account their advantages and disadvantages for the visualization but also for signal processing. We used the speech signals as an example of a non-stationary signal.

Speech signals are composed of a sequence of sounds. These sounds and the transitions between them serve as a symbolic representation of information. There are three general categories of the source for speech sounds: periodic, noisy, impulses, but combinations of these sources are often present.

2. Short Time Fourier Transform (STFT)

The Fourier transform (FT) in signal processing is used to transform signals from time domain to frequency domain. It is an expression of time-dependent signal by harmonic signals, i.e. sine and cosine. Fourier transform is not suitable if the signal has time varying frequency, i.e. signal is non-stationary. In other words, whether the frequency component “f” appears at time t1 or t2, it will have the same effect on transformation. If only the signal has the same frequency component “f” at all time, the result obtained by the Fourier transform is appropriate [1, 2].

Because speech signals have non-stationary character the use of the Fourier transformation is not correct. Although we obtain information about the spectrum of signal, we lose information about the time in which the spectral component occurs. The disadvantage of this can be resolved by Short Time Fourier transform.

There is only one difference between STFT and FT. In STFT, the signal is divided into enough small segments, where these segments of the signal can be assumed to be stationary. For this purpose, a window function “w” is chosen. The width of this window must be equal to the segment of the signal where it is stationary. This window function is first located to very beginning of the signal. That is, the window function is located at t=0 and width of the window at T. At this time, the window function will overlap with the first T/2 seconds. The window function and the signal are the multiplied. Then we apply Fourier transform on this product. The result of this transformation is the FT of the first T/2 seconds of signal. The next step would be shifting this window to a new location, multiplying with the signal and taking the FT of the product. The STFT is expressed using the formula 1.

$$STFT(\tau,\omega)=\int_t [x(t)\times w(t-\tau)]\times e^{-j\omega t}\,dt \tag{1}$$

The problem with the STFT is in the width of the window function that is used. If we use a window of infinite length, we get FT, which gives perfect frequency resolution, but no time information. Furthermore, in the order to obtain the stationarity, we have to have a short window, in which the signal is stationary. If the window is narrower we get better the time resolution, but poorer the frequency resolution. In fig. 1 and fig. 2 we can see STFT of input signal. In the first case we used speech of an adult male who said “My mobile phone doesn’t work” and the other one “The sky that morning was clean and quite blue”. For computing these figures was used Gaussian function as window function.

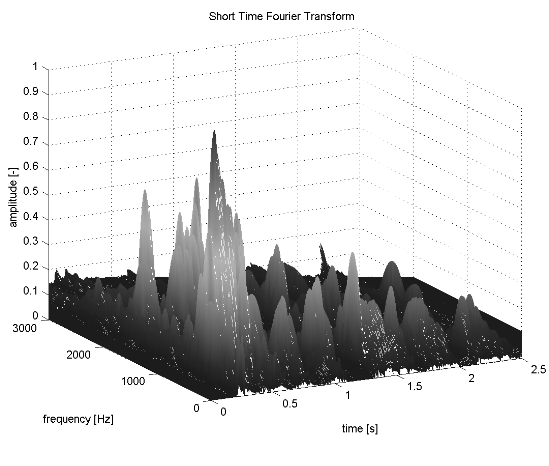

Fig. 1: *STFT of input signal 1, window length = 5ms*

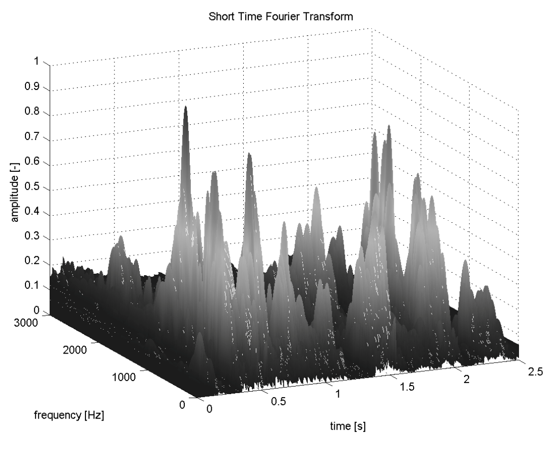

Fig.2: *STFT of input signal 2, window length = 5ms*

3. Wavelet Transform (WT)

Like a Fourier transform, the continuous wavelet transform uses inner products to measure the similarity between a signal and an analyzing function. The wavelet transform was developed as alternative approach to the short time Fourier transforms to overcome the resolution problem. The wavelet analysis is done in a similar way to the STFT analysis, in the sense that the signal is multiplied with a function, similar to the window function in the STFT, and the transform is computed separately for different segments of the time-domain signal. The most significant difference between STFT and the WT is change of window width. The continuous wavelet transform is defined by formula 2. [1, 2]

$$CWT = \frac{1}{\sqrt{|s|}} \times \int_t x(t) \times \psi\left(\frac{t-\tau}{s}\right) dt \tag{2}$$

The transformed signal is a function of a two variables, translation (τ) and scale (s). Psi ψ is the transforming function and it is called mother wavelet. There are about 400 mother wavelets that are more or less suitable for different tasks. The most used are Mexican hat, Morlet wavelet and Meyer wavelet. The parameter scale in the wavelet analysis is like the scale used in map. High scales represent a non-detailed global view and low scales represent a detailed view.

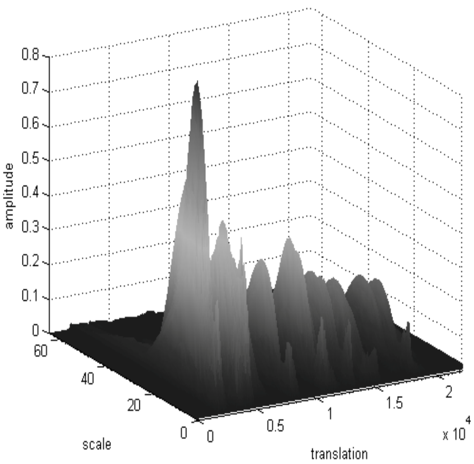

Fig. 3: *CWT of input signal 1, scale = 64*

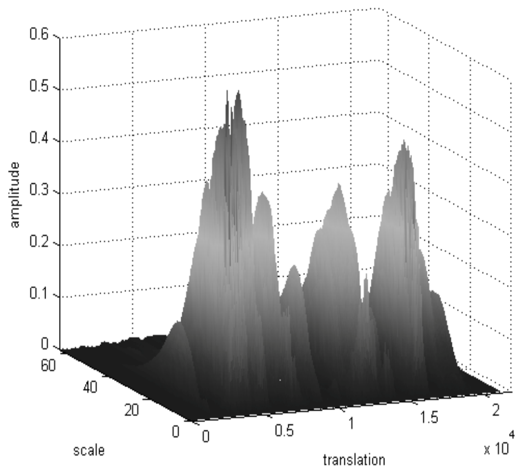

Fig. 4: *CWT of input signal 2, scale =64*

In fig. 3 we can see CWT of input signal 1 (“My mobile phone doesn’t work”) and fig. 4 shows CWT of input signal 2 (“The sky that morning was clean and quite blue”). For computing these figures was used mother wavelet from family SYMLETS, concretely SYM8 wavelet.

4. Wigner distribution

The Wigner distribution has been developed for application in quantum mechanics but later it has found application in signal processing as a tool for time-frequency analysis of signals. It can be interpreted as a distribution of signal energy in time and frequency. [3]

The definition of Wigner (or Wigner Ville) Distribution (WD) of signal s(t), with analytic associate x(t) shows formula 3. Fig. 5 shows example of this distribution.

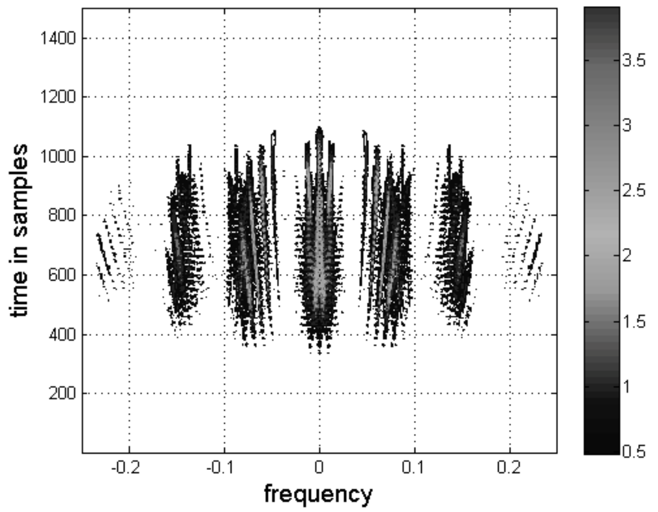

Fig. 5: *Wigner distribution of first 1500 samples of signal number 1.*

$$WD(t,\omega) = \int x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i\omega\tau} d\tau \tag{3}$$

5. Conclusions

In our work we have presented the possibilities of transformation and visualization of non-stationary signals. First we introduce STFT. This transformation is an improvement FT, because it provides us the time information. Disadvantage of this transformation is dependence of the result on the width of window and constant window width during whole process. Next we presented WT. A big advantage of this transformation is change of window width during process. Last we presented WD, this distribution we can also be used for time-frequency representation of non-stationary signals.

We have created scripts in program MATLAB that compute discrete variants of above mentioned transformations of input signal in WAV format. All figures in this paper were created by these scripts.

References:

[1] T.F. Quatieri: Discrete-Time Speech Signals Processing, Prentice Hall PTR (2002).

[2] R.M. Warner: Auditory Perception, Cambridge University Press, Cambridge (2008).

[3] H.P. Meana: Advances in Audio and Speech Signal Processing, Idea Group Publishing (2007).