

Ján Sabol
Július Zimmermann
Renáta Gregová

FROM SONAGRAM TO SCALOGRAM IN EXPERIMENTAL PHONETICS

Abstract

The study compares the basic principles of the Fourier transform and the wavelet transform of acoustic speech in discrete form. In order to show the higher time resolution of scalogram, the sonagram and the scalogram of one glottal pulse of the vowel /e/ were used as an example.

Key words: sonagram, scalogram, transformation of acoustic speech, experimental phonetics

1. INTRODUCTION

In the last decades, experimental phonetics, as well as the other (linguistic) disciplines, has gone through many changes resulting from rapidly developing field of computer science and from new discoveries in acoustics and mathematics. Among others, it is the application of the Fourier transform and the wavelet transform on speech signal processing that enables more precise and thus even more reliable analysis of speech sounds. These two (originally mathematical) tools are a great asset to all scientific branches dealing with human speech.

2. THE FOURIER TRANSFORM

When in 1822 the French mathematician Jean Baptiste Joseph Fourier (1768 – 1830) published his work *Théorie analytique de la chaleur* in which he dealt with the analysis of heat propagation in two-dimensional objects by differential calculus, he could not suppose what a strong device he had given into the "hands" of, except the other sciences, theoretical and experimental phonetics. Of course, we refer to the harmonic analysis of periodic processes by means of the infinite series of sine and cosine functions that was later revised by Dirichlet, Riemann and Lebesgue. The continuous formulation of the exponential form of the Fourier series is as follows:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt$$

This, the so-called Fourier transform, is the transform of time function $x(t)$ of speech signal from time domain t into spectrum $X(\omega)$ in continuous frequency range ω .

Based on the works of Gabor, Cooley and Tukey made harmonic analysis algorithmic in 1965. The discrete form of the so-called fast Fourier transform is defined by the formula:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j \frac{2\pi nk}{N}}$$

The formula expresses transform of discrete samples $x(n)$ of speech signal from time domain into discrete spectrum $X(k)$, while $k = 0, 1, 2, \dots, N-1$ is the order number of spectral line, $n = 0, 1, 2, \dots, N-1$ is the order number of a sample and N is the number of samples.

The principle of harmonic analysis is ingeniously simple, similarly as the handicraft of experts and craftsmen during the centuries. It is based on the use of templates (patterns, masters, etalons) for the verification of the shape of a drawing or substance. If the pattern of drawing (graph) that captures the time changes of an acoustic pressure of speech signal (oscillogram) is connected to the time changes of sine (harmonic) function, the correspondence of both drawings can be expressed in terms of a numeral coefficient. In the Fourier transform, this comparison is done in different frequencies of sine function in order to find out the coefficients of correspondence for all audible frequencies of sine function. Consequently, the obtained coefficients are represented graphically as the spectrum of speech signal. In the spectrum, there are formants, the key parameters of vowels and sonorous consonants. The analysis of the continuous changes of a coherent speech spectrum enables to detect the overall acoustic features of signal, to delimit the boundaries of sounds and – to a certain extent – to discover the voice modulations at the level of suprasegments.

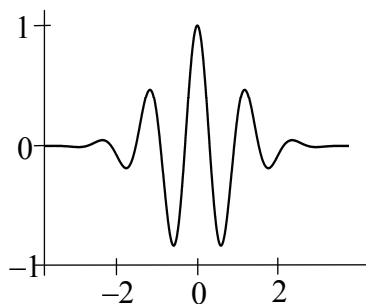
The drawback of the Fourier transform is the template itself – sine or cosine function since its time axis is divided from $-\infty$ to $+\infty$. If the comparison of infinitely long template to short graph is to be rational, the template and the signal accommodate. The shorter time interval with "rounded" edges taken from shorter time events can be cut off from a sine function by a weight function, e. g. after cutting out the shorter time events from a plosive, the repetition of the remains forms a sequence of the same plosive. At the end the ration between the lasting of template and the signal is indisputable. Despite the mentioned amendments, the high frequency resolution in spectrum can be achieved, but the time resolution is not able to depict events shorter than one glottal pulse. Since the vowel in isolation is only the sequence of glottal pulses with amplitude and temporal lability of the speaker's

voice, it is necessary to realize that in one glottal pulse there is a lot of information which cannot be shown by the Fourier transform.

3. THE WAVELET TRANSFORM

The problem of low time resolution can be eliminated by the wavelet transform. In comparison with the Fourier transform that develops signal into harmonic segments with various frequencies, the wavelet transform decomposes the signal into a modified version of the mother wavelet. The first reference to wavelets appeared in the appendix of the thesis published by the Hungarian mathematician A. Haar (1910). Haar's wavelet was rectangular and discrete, which made its use difficult to a certain extent. At the beginning of 1980s, the French geophysicists J. Morlet and A. Grossmann (Morlet et al., 1982 a, b; Grossmann & Morlet, 1984) were the first to mention the general definition of wavelets in the context of quantum physics. The French mathematician S. Mallat accommodated wavelets to signal processing. In 1989, he published the theory of multi-resolution in speech signal decomposition into wavelets (Mallat, 1989). In 1990, the Frenchman Y. Meyer, inspired by works of S. Mallat, constructed the first non-trivial continuously derivable wavelet (published in translation; Meyer, 1995). In 1987, I. Daubechies constructed the category of orthogonal wavelets that became the keystone of the cotemporary applications of the wavelet transform (Daubechies, 1988).

The wavelet has the form of wave with purposely limited duration and with a zero mean value. It has a tendency to be irregular and asymmetric. Several mother wavelets are defined: Haar's, Meyer's, Morlet's, Daubechies', Gauss', sombrero wavelet, etc. We use Morlet's wavelet as an example:



The mother wavelet – a specific variation of time window – is modified by the change of its time shift and its scale. The time shift is the change of the relative position of the mother wavelet on the time axis of a signal. The scale of mother wavelet has similar meaning as in cartography. The high scale (long, spread wavelet)

is characterized by the complex approach to the signal, the low scale (short, compressed wavelet) depicts the signal in detail. In the wavelet transform, the modified wavelet is the template apposed to the form of a speech signal in order to find out the value of the correspondence coefficient.

The product of the wavelet analysis is not a graph depicting relation time/frequency (spectrogram), but a graph of relation time /scale – scalogram. It is a tridimensional graph with the horizontal axis as the flowing time, the vertical axis is scale and the third coordinate of the graph is the value of correspondence between the wavelet and the signal. For the non-discrete form of the wavelet transform of speech signal $x(t)$ by the mother wavelet $\psi(t)$ with the value of scale s and time shift p the following relation is valid:

$$W(s, p) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-p}{s}\right) dt$$

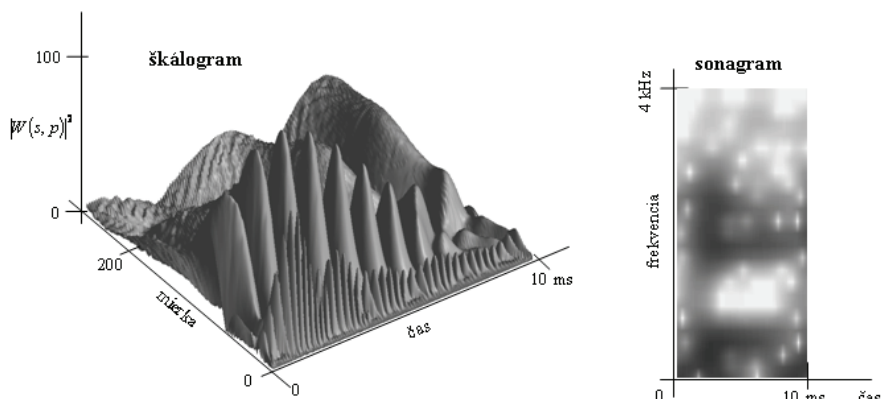
The discrete form of the wavelet transform that "works" with signal samples and wavelet samples is as follows:

$$W(s, n) = \sum_{k=-\frac{(N_{\psi}-1)s}{2}}^{\frac{(N_{\psi}-1)s}{2}} x(n+k) \frac{1}{\sqrt{s}} \psi(k)$$

The value of correspondence coefficient W between k wavelet samples $\psi(k)$ with the scale s and between speech signal samples $x(n)$ results from the notation, while N_{ψ} is the number of the mother wavelet samples.

4. CONCLUSION

In order to compare both mentioned principles of speech signal analysis, we use one glottal pulse of vowel e analysed by the wavelet transform (scalogram), and by the Fourier transform (sonagram).



As it follows from the picture, the increase of the sonogram horizontal dimension by the time magnifying glass does not increase the time resolution – the Fourier transform does not detect the short acoustic phenomena of a speech signal. It follows from what was said above that the wavelet transform is suitable for the analysis of glottal pulse, transients, glottal stops, voice onsets, etc. The latest research has shown that this transform can be used for the analysis of suprasegments. It is expected that it will be successful even in such difficult cases as the stress character of syllables.

REFERENCES

- Cooley, J. W., Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics Computation* **19**, 90, 297-301.
- Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure & Appl. Math* **41**, 7, 909-996.
- Gabor, D. (1946). Theory of communication. *J. IEE*, **93**, 429-444.
- Grossmann, A., Morlet, J. (1984). Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. of Math. Anal.*, **15**, 4, 723-736.
- Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme (Erste Mitteilung). *Math Ann Mathematische Annalen* **69**, 331-371.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **11**, 7, 674-693.
- Meyer, Y. (1995). *Wavelets and operators*. Cambridge University Press, 239 p.
- Morlet et al. (1982a). Wave propagation and sampling theory – part I: Complex signal and scattering in multilayered media. *Geophysics* **47**, 2, 203-221.
- Morlet et al. (1982b). Wave propagation and sampling theory – part II: Sampling theory and complex waves. *Geophysics* **47**, 2, 222-236.

Zimmermann, J. (2002). *Spektrografická a škálografická analýza akustického rečového signálu*. Prešov, Náuka, 176 p.

PUT OD SONOGRAMA DO SKALOGRAMA U EKSPERIMENTALNOJ FONETICI

Sažetak

Rad uspoređuje osnovne principe Fourierove transformacije i wavelet transformacije akustičkog govora u diskretne oblike. Kako bi se pokazala veća vremenska rezolucija skalograma, kao primjer su korišteni sonogram i skalogram glotalnog pulsa vokala /e/.

Ključne riječi: sonogram, skalogram, transformacija akustičkog govora, eksperimentalna fonetika