The corporation Visionics offers a means of recognition of under the name FaceIt PC [4], which is intended for amplification (strengthening) protection of independent PCs equipped OC Windows 95. The company DCS (Dialog Communication Systems) Inc. represents technology BioID, on the basis of which the identification of the person is made on three biometric parameters: the person, vote, movement of lips [5]. The testing of our system has shown results not making a concession of systems TrueFace ID and BioID. However, taking into account an orientation of new technology – n-GN on mass parallelism, the hardware realization of system will give significant advantages above systems TrueFace ID and BioID

References:

[1] V.A. Yashchenko Neural-like growing networks - new class of the neural networks // Proceedings of the International Conference on Neural Networks and Brain Proceeding, pages 455 -458, Beijing, China, Oct. 27-30' 98.
[2] V.A. Yashchenko Receptor-effector neural-like growing network - an efficient tool for building intelligence systems // Proceedings of the second internatiional conference on information fusion, July 6-8, 1999, Sunnyvale Hilton Inn, Sunnyvale, California, USA, Vol.11, pp. 1113-1118
[3] Technology TmeFace ID of the company Miros. - http: // www.miros.com, 2000
[4] Technology FaceIt of the company VISIONICS. - http: // www.faceit.com, 2000
[5] Technology BioID of the company DCS Inc.- http: // www.bioid.com, 2000
[6] Method main компонент.-http:// ww.mechmath. psu.ru, 2000
[7] P. Duda, item Харт Pattern recognition and analysis сцен.- Пер. With engl., under ed. V.L. Stefanuk m. The world, 1976E. Hall Computer Image Processing and Recognition - N.Y.: Academic Press, 1979
[8] K.S. Fu, J. Mu Pattern Recognition, 1981,v.I3.№I

Author information

Vitaliy Yashchenko - Institute of Mathematics of Machine and Systems; Kiev, Ukraine;
e-mail: mis@immsp.kiev.ua

SEGMENTATION OF A SPEECH SIGNAL WITH APPLICATION OF FAST WAVELET TRANSFORMATION

T. Yermolenko

Abstract: the article describes the method of preliminary segmentation of a speech signal with wavelet transformation use, consisting of two stages. At the first stage there is an allocation of sibilants and pauses, at the second – the further segmentation of the rest signal parts.

Key words: wavelet transformation, wavelet coefficients, approximation, segmentation.

Introduction

As known, the speech signal will consist of quasi-stationary parts corresponding to voice and sibilant phonemes, alternated by parts with rather fast changes of signal spectral characteristics (interphoneme transitions, explosive and occlusive phonemes, interword transitions speech - pause). It is possible to say, that the speech signal is characterized by nonlinear fluctuations of various scales. Therefore multireslution analysis and wavelet – transformation is considered to be rather effective for the analysis of a speech signal.
Segmentation of a speech signal (SS) means allocation of signal parts corresponding to separate structural units of SS. Considering phonemes as such units the task of segmentation is reduced to detection of interphoneme transitions. Within the framework of traditional approaches the decision of this task is rather problematic. However WT allows to solve this problem at least for the phonemes corresponding to rather extensive quasi-stationary SS parts. The matter is that on interphoneme transitions the signal undergoes significant changes at once on many research scales, and, accordingly, is characterized by increase of wavelet coefficients for many levels of decomposition while on stationary parts of phonemes wavelet coefficients appear grouped near to the certain scales. Thus, search of interphoneme borders can be reduced to search of moments of wavelet coefficients increase for a significant amount of levels of resolution.

Algorithm of sibilants and pauses border detection

The n-level wavelet decomposition of a discrete signal $f(t)$ is defined as

$$f(t) = \sum_{k=0}^{\frac{N}{2^n}-1} s_{nk}\varphi_{nk} + \sum_{j=1}^{n}\sum_{k=0}^{\frac{N}{2^j}-1} d_{jk}\psi_{jk} \quad (1)$$

thus N is a amount of samles of signal, s_{jk}, d_{jk} - the wavelet coefficients, named average and differences accordingly (in this work we shall call them as coefficients of approximation and detail) $\varphi_{jk} = 2^{j/2}\varphi(2^j t - k),\ j,k \in Z$, φ - scaling function or scale function $\psi_{jk} = 2^{j/2}\psi(2^j t - k),\ j,k \in Z$, ψ - basic or "mother" wavelet.

In the researches we used fast discrete wavele transformation of Daubechies, which was perfomed on 6 levels, believing $s_{0,k}$ equal to readout of an initial signal. SS, digital with frequency of digitization of 22050 Hz is broken into overlapped windows in size 20 ms with half overlap windows.

Apparently from figure 1, 2, 3 for allocation of pauses and sibilants the signal has enough information on behavior of detaile coefficients on the 6-th level as for them the small amplitude in comparison with other signal parts is typical.

We build numerical sequence $\{a_{i6}\}_{i=1}^{N/256}$:

$$a_{i6} = \sum_{k=0}^{n_6-1} d_{6,i+k}^2,$$

where i - number of a sliding window, $n_6 = \frac{n}{2^6}$ - the size of a sliding window on the 6-th level, n - the size of a window in an initial signal (512 samples).

The prospective beginning of sibilant (pause) is placed in the beginning of i windows for which the condition $a_{i-1,6} \geq 1000, a_{i,6} < 1000$ is implemented. It is obvious, that at the end boarder of sibilant (pause) this condition is carried out just the other way. The threshold has been received experimentally and is independent of announcer.

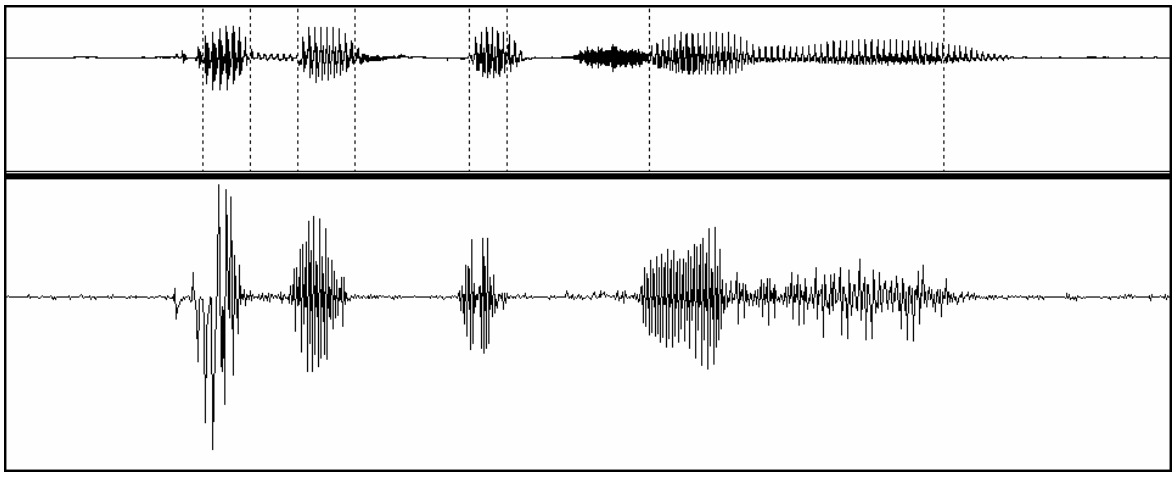

Figure 1. Segmentation of a word *обеспечение* made by the announcer - man. Above – amplitude-time representation of a signal, below – detail coefficients of the 6-th level.

Results of work of algorithm with the word *обеспечение* made by the announcer - man (figure 1), and the announcer - woman (figure 2), and also with a word *кошачий*, pronounced by the announcer - woman (figure 3) are shown below. Last case shows the work of a method in conditions of a significant noise (relation signal / noise makes 15 decibel).

For this algorithm simplicity of realization, independence of announcer, low sensitivity for noise is characteristic. One of properties of this algorithm is reference to a pause of the parts of SS corresponding to mute vowel.

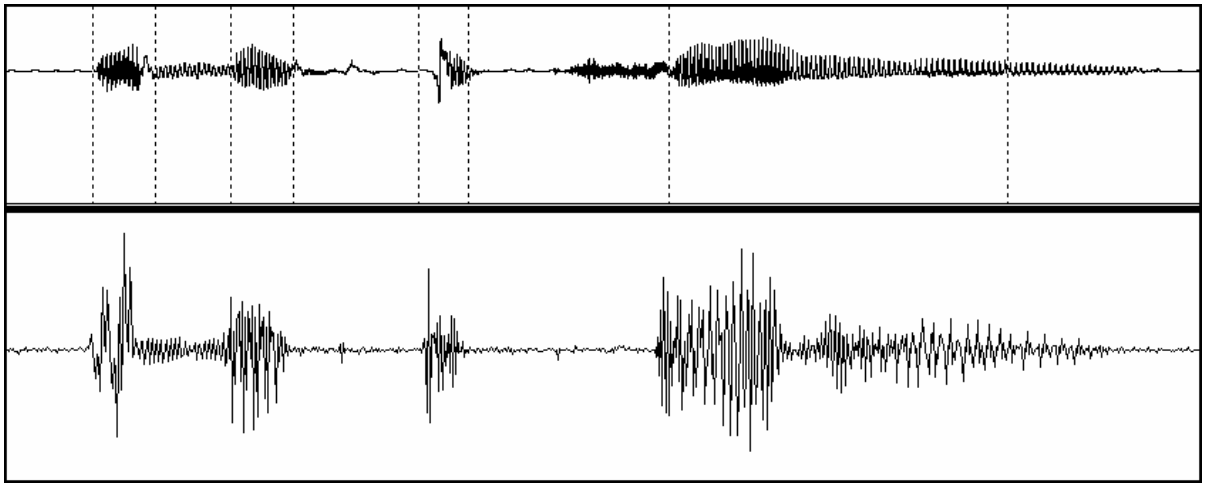

Figure 2. Segmentation of a word *обеспечение* made by the announcer - woman. Above – amplitude-time representation of a signal, below – detail coefficients of the 6-th level.

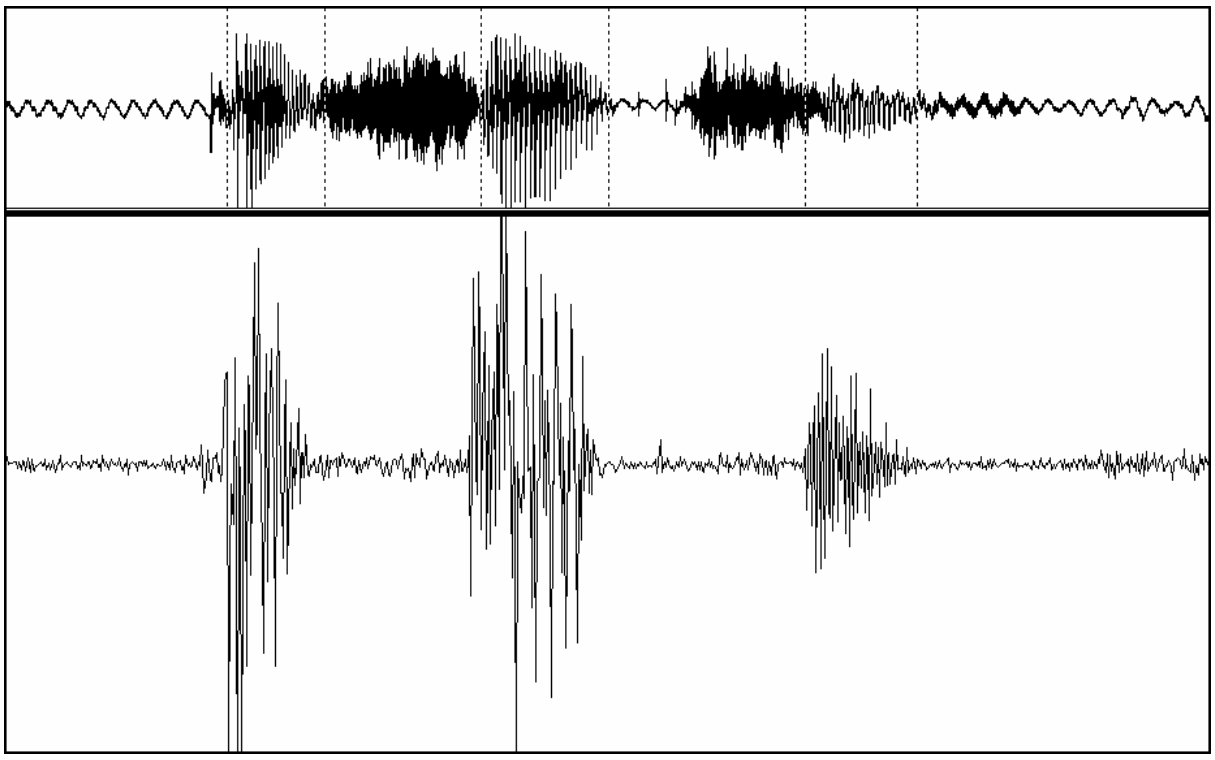

Figure 3. Segmentation of a word *кошачий*, made by the announcer - woman. Above – amplitude-time representation of a signal, below – detail coefficients of the 6-th level.

Segmentation of signal parts containing vowels, sonorous concordants and fricatives

The following criterion for a choice of the most informative level of decomposition has been received by the experimental way. For j level of decomposition, since the 3-rd one, the inequality implementation is checked:

$$\frac{E}{N} < \frac{E_j}{N_j}, \tag{2}$$

where N_j - amount of detail coefficients at j level, more than 0,5;

$$E = \sum_{i=0}^{N-1} s_{0,i}^2 ;$$

$$E_j = \sum_{i=0}^{\frac{N}{2^j}-1} d_{j,i}^2 .$$

The first level for which the condition (2) can be implemented is the most informative.
The behavior of detail coefficients was analyzed in the following way: at the chosen level of decomposition j for a segment, which is not related to a sibilant (pause), the numerical sequence $\{e_{ij}\}_{i=1}^{l}$ was built:

$$e_{ij} = 10 \lg \sum_{k=0}^{n_j-1} d_{j,i+k}^2 ,$$

where i - the number of a sliding window, $n_j = \frac{n}{2^j}$ - the size of a sliding window at j level, n - the size of a window in an initial signal (512 samples), l - amount of windows in an examined segment.
Further averaging sequence on 3 values was carried out.
Borders of prospective segments were put down between windows with numbers i and $i+1$, for which $\left|e_{i+1,j} - e_{i,j}\right| \geq 3.5$.
Lacks of algorithm work are those:

1. In stressed and unstressed *-а-* the superfluous segment can be allocated;
2. In some cases two vowel sounds, one after another, are not distinguished, for example, *-оа-*, *-ио-* in words *коала, миллион*; the characteristic ending *-ия-, for example, квалификация, аппроксимация*, also is not segmented, that is explained, as *-ия-* is graded and sounds, as unstressed *-а-*.
3. Resonant sounds in combination with unstressed vowel are not divided among themselves.

In figure 4 the result of work of algorithm with a word *акселерация* after separation of sibilants and pauses is shown. Apparently from figure, the segment 1 corresponds to unstressed *-а-*. The pause is precisely enough separated from the speech. The segment 2 contains sounds *-к-* and *-с-*. In the 3-d, unstressed *-е-* and *-л-* are combined, that can be explained to some degree: the voice sound is reduced, has short duration and loses its qualities. The second unstressed sound *-е-* is well separated by borders of the 4-th segment; obviously, it is connected with the fact that it is to the first pretonic vowel. The segment 5 corresponds to *-р-*, the 6-th – to stressed *-а-*. The 7-th segment contains sound *-ц-*, the 8-th segment comprises the unstressed ending *-ия-*.

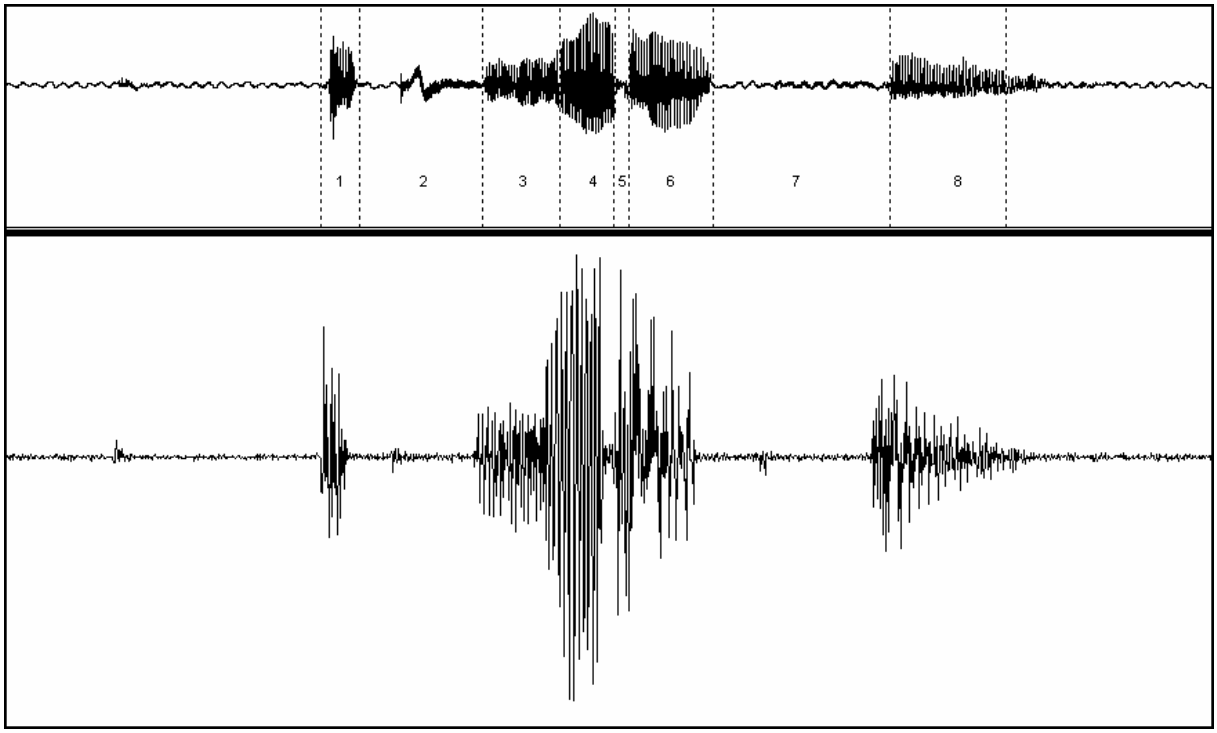

Figure 4. Segmentation of a word *акселерация* after separation of sibilants and pauses. Above – amplitude-time representation of a signal, below – detail coefficients of the 6-th level.

Conclusion

In the work the method with wavelet-transformation use that segments SS in two stages is described. Firstly, sibilants and pauses that are identified among them are allocated. For this stage simplicity of realization, independence of announcer, low sensitivity noise are characteristic. At the second stage the rest parts of a signal are segmented. Results of its performance are less reliable, such mistakes as occurrence of superfluous borders in vowels and non separation of resonant *-л-, -м-, -н-* from *-и-, -е-* in some cases are characteristic. Thus, there is a necessity of more detailed research of behavior of wavelet coefficients at all levels of scaling in view of relative amplitude of a parts of SS, frequency attributes, duration of a segment and similar attributes of the nearest next sites.

Literature

1. Dremin I.M., Ivanov O.V., Nechitajlo V.A. Wavelets and their Use. // Successes of Physical Sciences, v. 171, №5 p. 465-500, 2002.
2. Astafjeva N.M. Wavelet-analysis: Bases of the Theory and Examples of Application. // Successes of Physical Sciences, v 166, №11 p. 1145-1170, 1996.
3. Detection of Change of Properties of Signals and Dynamic Systems. Under M.Bassvil, A.Banvenist's edition. Moscow, "Mir", 1989.
4. Walker J. Fourier Analysis and Wavelet Analysis. // Notices of the AMS, vol. 44№6, p. 658-670, 1997
5. Daubechies I. Ten lectures on wavelets. // Philadelfphia: SIAM, 1991

Authors information

Yermolenko Tatyana – Institute of Artificial Intelligence, B.Hmelnitsky avenue, 84, Donetsk - 83050, Ukraine
e-mail: etv@iai.donetsk.ua