TEKA. COMMISSION OF MOTORIZATION AND ENERGETICS IN AGRICULTURE – 2013, Vol. 14, No.2, 164-169

Multifractal approach in pattern recognition of an announcer's voice

Victor Solovyov, Yana Byelozorova

Volodymyr Dahl East-Ukrainian National University,
Molodizhny bl., 20a, Lugansk, 91034, Ukraine,
e-mail: mettiltd@mail.ru, bryukhanova.ya@gmail.com

Received May 23.2014: accepted June 05.2014

Summary. The research results on revealing of multi-fractal structures for vowel phonemes are given in this paper. The possibility to "construct" vowel phonemes from "atomic" phoneme structures by means of affixes transformations is shown. Multi-fractal conditions which allow designing voice individual characteristics are considered in this paper.
Key words: self-similar structure, speech fragments, voice recognition, "atomic" components of phoneme fragments, phonemes, multi-fractal.

INTRODUCTION

The researches in the field of automatic recognition of radio announcer's voice during the last ten years resulted in the creation of rather efficient recognition systems. The improvement and evolution of these systems are going on [15, 17]. There are different approaches to the practical realization of recognition systems of an announcer's voice. However, till nowadays there is no precise enough physical model influencing the voice individual characteristics which are the main acoustic parameters of speech formation.

Since Helmholtz times, the father of speech formation acoustic theory, despite a great number of researches the evolution of speech formation physical concepts hasn't been essential.

The appearance of modern computer systems for speech formation researches allowed performing the practical task to recognize speech and identify the voice. However, there is no generally – recognized physical concept of voice individual characteristic formation.

One of the reasons of the concept lack is a great variety of different parameters and factors which are necessary to be considered while researching the voice characteristics [1, 2, 3]. Setting a further developed task, on the one hand allows creating the structural model of speech fragments which gives a full rational description of all main speech formation tasks. On the other hand, this model enables to increase sufficiently the research process of different factors influencing the individual characteristics of announcer's voice.

Any systems and models in the field of speech recognition [8, 18] must use a certain "modes" in generalized sense to recognize "atomic" speech components. Stored modes of words, sentences, phonemes and so forth in the form of sound fragments can be speech components. Just the parameters of mathematical models taking into consideration the pre-history of the process of sound wave alteration into audio file which are compared with some analogue mode-parameters can be speech components too. It is senselessly even to set a task of speech recognition without "modes" in generalized sense solving the task of recognition including modeling on the basis of neuron nets.

The question is arising. Is it possible to design modes which are the result of physical objects describing the definite "atomic" structure with quite definite invariant parameters?

The models of multi-fractal structure of speech sounds and the description of voice characteristics on the basis of these structures have been developed. The results of this development are presented in the paper.

SETTING A TASK FOR THE RESEARCH

Let's consider the task of an announcer's voice recognition within the framework of the context-independent speech with unlimited language vocabulary and voluntary voice characteristics. Designing the main physical concept we shall consider the task at the level of phonemic components of the Russian language speech or the structures equivalent to phonemic components. This research deals not only with vowels of the Russian language – [a], [i], [o], [u], and [je]. Vowel phonemes are considered according to the provisions of Moscow phonological school. As it will be seen further this limitation is not important for the analysis of basic voice individual characteristics [5, 7, 14].

Let's consider the speech fragment in the form of sound wave chart – Figure 1 (here and further programmed complex "Fraktal" is used for charts and drawings design [10]).

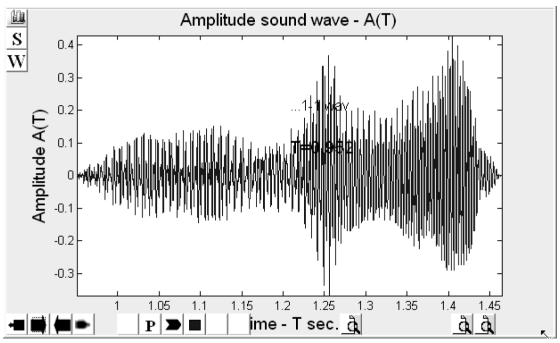

Fig. 1. Speech fragment

The chart view depends on individual characteristics of voice, speech rate, and a number of other factors. The conformity between vowel phoneme of speech fragment and temporary intervals of a sound wave by means of listening to the fragments is set.

The criteria of developed model conformity with objectively existing voice characteristics are the basis of further research and the results of the development. The following research methodic is the basis of methodical solution of this problem.

During the listening the fragments of sound wave equivalent to vowels of the Russian language sounds are purposefully modified according to the developed physical model (in time and frequency fields). After modification these fragments are listened to by a group of tested people with the aim to reveal identity both phonemic sound of the first fragment and modified one, and the voice individual characteristics. The identification of phoneme sounding and voice characteristics before and after modification is determined in the case of 95% of recognition by the group of tested people. This methodic has definite limitations and a certain share of subjectivism. But nowadays man's organs of hearing is practically the only "standard device" which can be used.

The researches have been done both with separate pronounced vowel sounds and with sounds within the framework of different texts and voices characteristics.

The research results with a following set task are given. This set task includes: to reveal stable structures with definite parameters which completely characterize the definite phoneme, voice individual characteristics and they are context independent.

MULTI-FRACTALS, VOWEL PHONEMS AND INDIVIDUAL CHARACTERISTICS

Speech fragments in audio-data as a discrete time row of sound wave amplitude are considered in the researches. It is necessary to reveal the self-similar structures in speech fragment of time row on the basis of which it is possible to form phonemes and voice characteristics. Self-similarity is considered as self-similarity of multi-fractal structures in accordance with Mandelbrot's concept [9, 11, 12]. It is an approximate geometrical analogue, visually watched sound wave fragments. The fragment of sound wave of equivalent phoneme [i] (a part of phoneme fragment) is presented in Fig. 2.

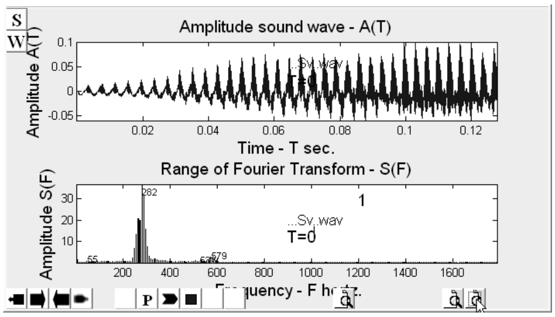

Fig. 2. Fragment of acoustic wave (phoneme [i])

At the bottom of Fig. 2 there is Furrier's spectrum of excreted phoneme fragment. The spectrums of separately excreted definite vowels have some geometrical similarity between them. This approximate similarity is watched while seeing visually; it doesn't depend on the voice characteristics and the nature of phoneme pronunciation. A great number of works dealing with the finding of conformity between a spectrum and particular vowel phoneme indicate at a definite degree of interconnection. However an exact enough scientific grounded model to reveal conformity between vowel phonemes and spectrums doesn't exist.

Let's examine the similarity of amplitude fragments of sound wave phoneme [i] during small time intervals (approximately 5-20 ms) Fig.3.

Periodically repeating self-similar structures are watched in the time field of sound wave. (Fig.3). Fragment Furrier's spectrum during small time intervals is a little bit different from the spectrum on fig. 2 due to the position of local extremes in a frequency field as well as due to the extremes amplitude.

The phoneme fragment in the field of T=0.015-0.02 s. – fig. 4 singled out.

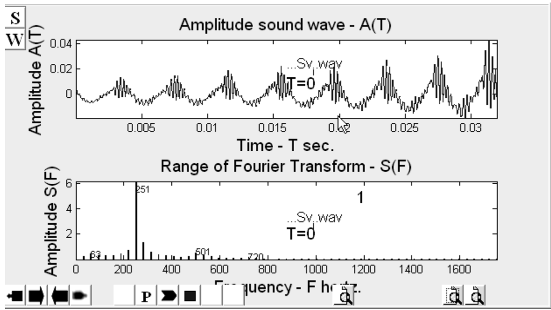

Fig. 3. Fragment of acoustic wave (phoneme [i])

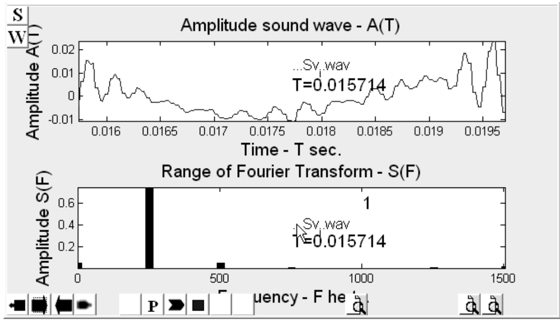

Fig. 4. Phoneme fragment [i]

The spectrum of given small fragment is essentially different from the spectrums with large time intervals for the phoneme [i]. It also differs in frequency resolution due to the shortage of time intervals. The task of phoneme [i] "designing" from a small "atomic" phoneme fragment by means of affix transformations of widening and shortening is considered. From fragment given on fig.4 the acoustic fragment consisting of several tens completely similar fragments different from each other in an amplitude strain-grasp is designed. Fig. 5

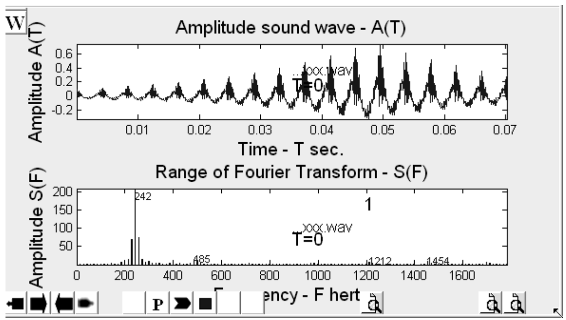

Fig. 5. Fragment of modified amplitude of sound wave (phoneme [i])

Listening to a modified fragment the tested people have to identify clearly the phoneme [i]. The voice characteristics and time periods of “atomic” structure (from 5 till 15ms) varied during the experiment. The length of designed phoneme sounding has been changing in a wide range. Artificially designed phoneme [i] is quite clearly identified after transformations during the listening. Thus individual characteristics of initial voice are either absent or expressed very weakly.

Received results indirectly appear in many researches. However these results are only initial preconditions to design models of a multi-fractal structure of speech sounds and voice individual characteristics in this work.

Thus, fragments given above as research results show that vowel [i] can be presented (“designed”) on the basis of definite “atomic” fragments of the same phoneme by means of a number of affix strain-grasp transformations by the “atomic” fragment amplitude.

Completely the same experiments for “a” phoneme also show the possibility to design this phoneme from “atomic” fragment with the phoneme size from 5 till 15 ms. Vowel phoneme of the Russian language – [o], [u], [je], as researches show have also a multi-fractal structure. These phonemes can be designed from “atomic” fragments of every phoneme.

At the same time it is necessary to note that in a number of cases not only one atomic structure is required to “design” the phoneme. The results of such researches are not the subject of this paper. They will be presented in other publications.

THE DESIGNING OF VOWEL PHONEMES WITH VOICE INDIVIDUAL CHARACTERISTICS ON THE BASIS OF MULTI-FRACTAL STRUCTURES

Properly from the second section the Russian language vowel phoneme characteristics can be studied and determined by characteristics of “atomic” multi-fractal components. However individual characteristics of initial voice are disappeared in the majority of voluntary variations of the phoneme designing. The task to “design” vowel phoneme from multi-fractal components saving voice individual characteristics is examined.

Voice individual characteristics for vowel phoneme studying the modification of sound wave fragments both in a time field and in a frequency one are investigated. Firstly, today it is connected with famous neuro-physical regularities of sound information processing by a man [4, 6, 13].

As a result of these researches which are given in section 2 it is necessary to note that sounding characteristics of “designed” phoneme depend on the sizes of “atomic” structures as well as on the structural phoneme geometry during visual examination in a time field. This fact is obvious. But individual characteristics of initial voice at such modeling are completely absent. It is only possible to identify man and woman’s voices.

Let’s examine the task of more complex affix transformations of “atomic” structures which take into consideration a number of important speech fragments.

Examine the phoneme fragment [i] on a small time scales – Fig. 6

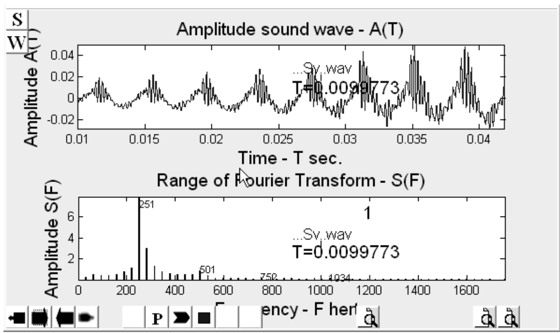

Fig. 6. [i] Phoneme fragment

It is possible to design the phoneme [i] from any chosen “atomic” phoneme fragment within the interval of 5-15 ms on the basis of multi-fractal model given in section 2. However individual characteristics of initial voice are disappearing. We shall “design” atomic structure at time interval equivalent to the frequency of primary tone for the phoneme, Fig. 6. Here we shall examine two basic “atomic” structures for a given phoneme. One structure is with 0.019-0.024 мs time interval. Another one is with 0.031-0.036 мs interval.

Phoneme [i] is designed from two consecutive components on the basis of affix strain-grasp transformations. The first phoneme part is on the basis of the first atomic structure, the second part is on the basis of the second atomic structure. The share of each atomic structure in a designed phoneme is proportionally to corresponding time intervals, Fig.6. Received multi-fractal design is presented in Fig. 7.

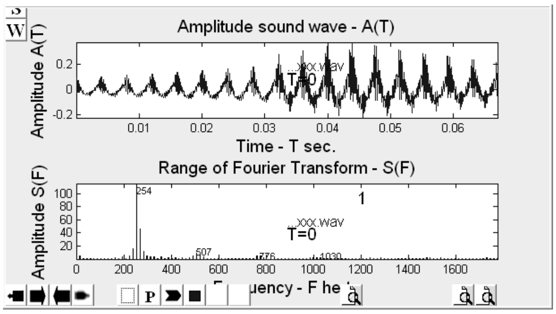

Fig. 7. Phoneme [i] fragment designed of atomic structures

Listening to “designed” phonemes tested people must identify the characteristics of non-modified voice. While perceiving there are still some differences of voice characteristics after modification.

Greater voice characteristics identity can be achieved while listening. It is possible to divide the phoneme into three or four basic “atomic” structures. However in this case it is more difficult to use the developed model in practice.

Thus, given research results show that it is possible to present vowel phonemes of the Russian language in the form of multi-fractal structures. Here the construction of vowel phoneme on the basis of multi-fractal approach permits to design voice individual characteristics.

CONCLUSIONS

1. Experimental researches dealing with frequency – time modification of vowel phoneme of the Russian language have shown obviously vividly-expressed multi-fractal structure of vowel phonemes.

2. The multi-fractal structure of vowel phonemes is revealed that it is possible to “design” vowel phonemes from “atomic” phoneme fragments by means of affix transformations.

3. All these elements of voice model make possible to design voice individual characteristics on the basis of a multi-fractal structure of the Russian language phonemes.

REFERENCES

1. **Agranovskiy A.V., Lednov D.A., 2004.:** Theoretically aspects of processing algorithms and classification of speech signals. – 164.
2. **Aldoshina I.A., 2000.:** Basics of psychoacoustics. – Vol, №6. – 36–40 (In Russian).
3. **Atal B.S., Hanauer S.L., 1971.:** Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. – Vol, 50 (№ 2). – 637-655.
4. **Bextereva N.P., 1988.:** Psychoacoustic aspects of speech perception. Mechanisms of brain activity. — 504 (In Russian).
5. **Fant G., 1970.:** Analysis and synthesis of speech processes. Transl. from English by Lozovskiy. – 297-330.
6. **Feder J., 1991.:** Fractals. Translated from English. – 326 (In Russian).
7. **Flanagan J., 1968.:** Speech analysis synthesis and perception. Translated and edited by Pirogov A.A. – 396 (In Russian).
8. **Hermansky H., 1990.:** Perceptual Linear Predictive (PLP) Analysis for Speech. – Vol, №87. – 1738-1752.
9. **Mandelbrot B., 1969.:** Robustness of the rescaled range R/S in the measurement of non-cycling long-run statistical dependence. – Vol, № 5. – 967-988.
10. **Mandelbrot B., 1972.:** Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis. Annals of Economic Social Measurement.
11. **Mandelbrot B., 1982.:** The Fractal Geometry of Nature. New York: W. H. Freeman.
12. **Mandelbrot B., 1999.:** A Multifractal Walk Down Wall Street. Scientific American.
13. **Pavlov A.N., 2007.:** Multifractal analysis of complex signals. – Vol, №8 (In Russian).
14. **Rybalskiy O.V., Solovyov V.I., 2010.:** Identification system equipment record based on multifractal approach // Herald V. Dahl East-

Ukrainian National University, № 9 (151). – 58–64 (In Russian).
15. **Sieriebriak K., 2012.:** Information technologies in Ukraine: problems and obstacles Teka. Commission of motorization and energetic in agriculture – Vol. 12, № 3, 128-135.
16. **Sorokin V.N., 1985.:** Theory of speech recognition. – 312 (In Russian).
17. **Voronova A., 2011.:** Information technologies in public administration practice. TEKA, Commision of Motorization and Power Industry in agriculture, Lublin University of technology, Volodymyr Dahl East-Ukrainian National University of Lugansk, Volume XD, Lublin, 313-318.
18. **Zwicker Eb., 1971.:** Ear as a receiver of information. Translated from German, edited by Belkin G.I. – 225(In Russian).

МУЛЬТИФРАКТАЛЬНЫЙ ПОДХОД В ЗАДАЧАХ РАСПОЗНАВАНИЯ ГОЛОСА ДИКТОРА

Виктор Соловьев, Яна Белозерова

Аннотация. В статье приведены результаты исследований по выявлению мультифрактальных структур для гласных фонем. Показана возможность "конструирования" гласных фонем из "атомарных" структур фонем путем аффинных преобразований. Рассмотрены условия мультифрактальности, которые позволяют моделировать индивидуальные характеристики голоса.
Ключевые слова: самоподобные структуры, фрагменты речи, распознавание голоса, “атомарные” составляющие фрагментов фонем, фонемы, мультифракталы.