The Second International Conference "Problems of Cybernetics and Informatics" September 10-12, 2008, Baku, Azerbaijan. Section #1 "Information and Communication Technologies" www.pci2008.science.az/1/28.pdf

ALGORITHMS OF PHONEMES CLASSIFICATION IN FIELD OF COMPILATIVE SPEECH SYNTHESIS SYSTEMS REALIZATION

Yedilkhan Amirgaliyev[1], Rustam Musabaev[2]

Kazakh National Technical University, Almaty, Kazakhstan
[1]*amir_ed@mail.ru*, [2]*rmusab@gmail.com*

By estimations of the majority of experts modern information technologies have reached that level of quantity indicators when the further growth of quality indicators will be carried out basically due to an intensification of application of intellectual algorithms and technologies at their realization. Speech technologies concern to one of such intellectual directions also [1].

The phoneme is one of base units for construction of the majority of modern speech systems. The phonetic model of language is applied at the decision of following problems of cybernetics: artificial speech synthesis; automatic speech recognition; voice-activated control systems; voice biometrics and identification of the person.

It is possible to tell, that the phoneme represents a certain elementary atom of language of which there is all language variety with all distinctions of forms and shades. Actually having the limited set of phonetic components of language, we can receive infinite set of its various more complex forms: syllables, words, phrases and the whole texts. For this purpose it is necessary to change and combine phonemes, to connect them by various rules, to build in certain sequence with the purpose of reception of the necessary form of text representation. For this purpose it is necessary to change and combine phonemes, to connect them by various rules, to build in certain sequence with the purpose of reception of the necessary form of text representation. The optimal decision of the given problem in modern conditions is attraction of means of computer facilities and realization of the given problem in the form of program algorithms presented by means of any of high-level languages. One of the basic problems at realization of similar model is definition of optimum sufficient set of parameters for realization of a necessary classification level of phonemes within the limits of projected program model. Three variants of construction of similar model in which classification parameters will be are actually possible: insufficient; superfluous; optimum.

In case of insufficient quantity of parameters of classification it is inconvenient to realize to the full model which would correspond to all requirements on quality. At their superfluous quantity superfluous computing resources will be spent for calculation of not used parameters. One of the definition variants of an optimum set of classification parameters is the approach for which at a stage of model construction the superfluous set of parameters is used, and at a stage of final realization it is reduced down to minimally necessary.

The phoneme acts in a role of classified unit. Under a class of one phoneme sets of its sound realizations with greater distinctions in their acoustic characteristics can be united. Actually the phoneme represents desirable for pronouncing by the person, and sounds - actually obtained results during the moment of pronouncing. For various languages allocate various sets of phonemes. Thus the part of phonemes of one language can be identical to a part of phonemes of another.

As concrete sound realization of a phoneme for the certain case acts allophone. Usually a allophone describe influence of the certain set of factors on a choice of concrete sound realization of a phoneme. Let's result the major factors influencing formation of allophone base of language:

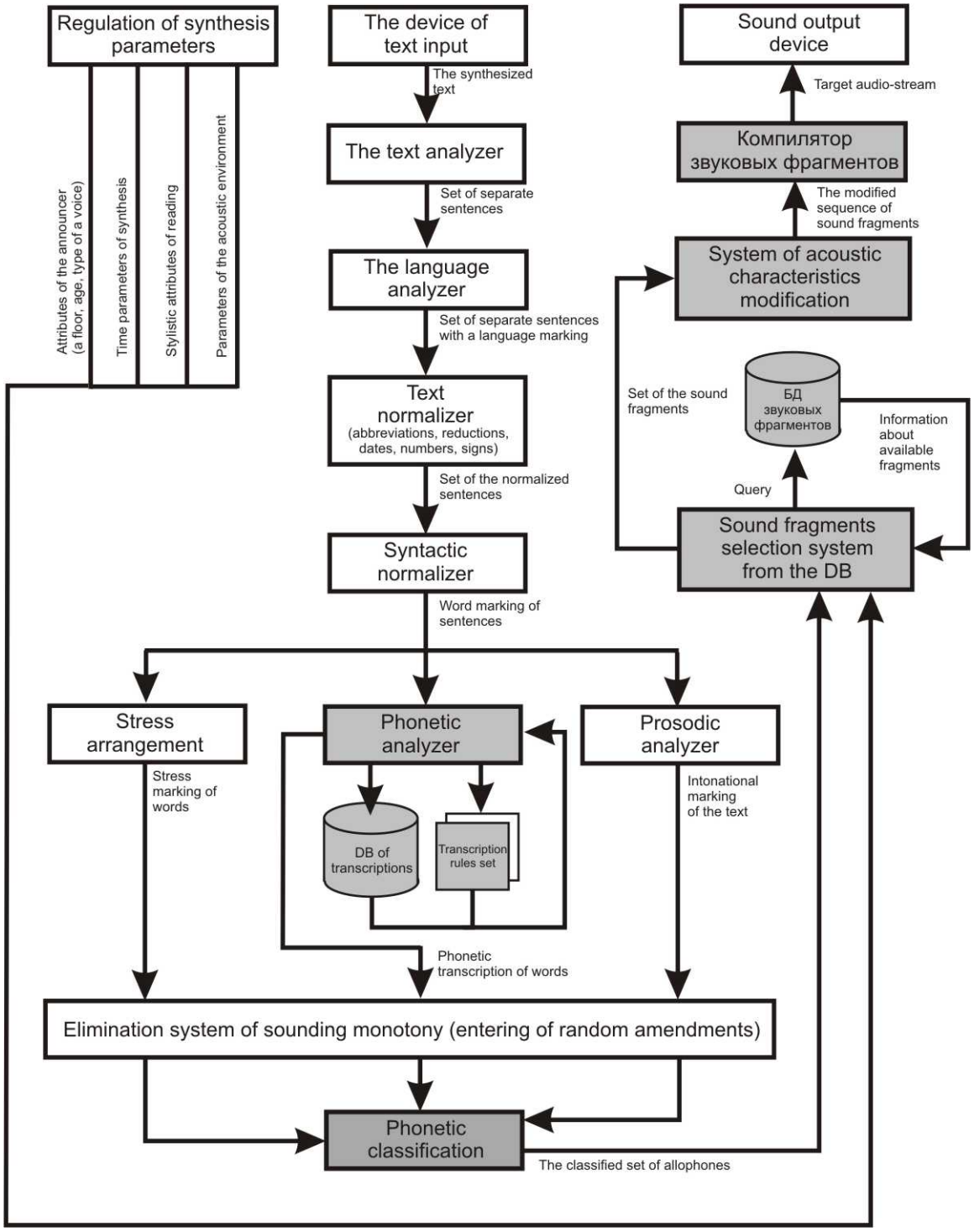

Fig. 1 – The general scheme of complex interaction of a phonetic level (grey color) with other subsystems of speech synthesis.

1) Total of phonemes count for concrete language;
2) Rules of mutual combinations of phonemes;
3) Position of a phoneme concerning other phonemes in a word and in the sentence (initial, median, final, isolated);
4) Type of a phoneme (vowel, concordant);
5) Types of the previous and subsequent phonemes (vowel, concordant);

6) Phoneme stress (primary, secondary, without an accent);
7) Neighbouring phonemes stress;
8) Presence of an explosive or clicking components;
9) Presence of a noisy component;
10) Presence of a tonal-periodic component;
11) Attribute of a complex vowel phoneme (consisting of several more simple sounds);
12) Initial variability (variability of a phoneme in the beginning of a word);
13) Median variability;
14) Final variability;
15) Communication with vowel in the beginning of a word (the adjunction, imposing, isolated);
16) Communication with wovel in the middle of a word;
17) Communication with vowel in the end of a word;
18) Rules of phoneme duration change;
19) Rules of base tone frequency change;
20) Rules of sound energy change.

With application of the listed attributes it is possible to build the mechanism of classification for a phonetic level. For construction of audiophonetic model of language according to attributes it is necessary to carry out classification of all phonemes. The given classification model will be "ideal" model in the certain degree. At a stage of direct synthesis it is necessary to carry out search in a DB most proper sound fragments of phonemes close to an ideal. Thus sound fragments should be preliminary proclassified. Dimension of sound fragments depends on the chosen approach for construction of system. Most often used dimensions:

1) Phoneme - a sound signal dimension in the whole phoneme;
2) Diphone - half two adjacent phonemes from the middle of one up to the middle another;Слог;
3) Sequence of several syllables;
4) Word.

In a daily life the person perceives phonemes as elementary individual sounds. However by more detailed consideration of phonemes it is possible to notice, that they can consist of several more simple sounds. For example, English phonemes /NG, AE, OW/ are complex phonemes. Thus if to allocate really elementary sounds them it is possible to classificate as:

1) Simple voice-frequency sounds (a basis - low frequencies) ;
2) Simple noisy sounds (a basis - high frequencies);
3) Impulse like clicking sounds and explosive sounds (short-term impulses of high or low frequency);
4) The combined sounds (can consist of any previous components).

Thus simple voice-frequency and noisy sounds can be lingering in time, i.e. the person can say them any way long. Whereas clicking and explosive sounds have restrictions on time duration. These characteristics of sound components of phonemes can be used for construction of more flexible synthesis systems. At stage of construction of similar systems it is necessary to use various algorithmic approaches for dynamic modification of their acoustic parameters. For example, if we wish to increase duration of the complex phoneme consisting of voice-frequency and clicking components it is necessary to separate them by means of a filtration. Then the voice-frequency component shares for the elementary periods and increases by means of their cloning. Thus usually there is no necessity for cloning of clicking components. After

transformations the separated signal mixs up in single one. The principle of separate modification of frequency components of phonemes on the basis of their classification has shown good results during the spent experiments on speech synthesis.

Also at sound synthesis of phonemes it is possible to use more elementary units of synthesis: a wave duration during one period of the basic tone, the long modulated noisy signal, a short-term clicking noisy signal similar to an impulse.

The system of compilative synthesis of speech is the multilevel complex system consisting of set of subsystems. All subsystems within the limits of uniform system solve the general problem of getting of the synthesized audio-stream. Each of subsystems solves an own problem and has a set of entrance and target parameters. Target parameters of subsystems of higher level are passed to entrance parameters of subsystems of lower level.

The phonetic level of sound synthesis is a base base on which any complex system of compilative synthesis of speech is under construction. In a Fig. 1 the general scheme of complex interaction of a phonetic level (grey color) with other subsystems of synthesis of speech is presented. It is possible to say, that the phonetic level of sound synthesis is a kernel of system. Quality of all system depends on its realization as a whole. If its realization does not satisfy to necessary conditions of quality it can reduce to a minimum all the quality indicators received at higher levels. This level also is the most labour-consuming at realization in the technological plan. It is necessary to consider specificity and rules of acoustic representation of separate phonemes, their mutual influence against each other, to consider effects of an coarticulation and imposing of several sounds, rules of variability of sound representation of phonemes at various prosodic conditions [2]. Also it is necessary to have functional dependence of frequency of the basic tone, energy and duration of phonetic units from current prosodic schemes [3]. At sound synthesis of phonemes it is necessary to consider features of psyhoacoustics which describes nonlinear properties of human hearing [4]. Concepts of psyhoacoustics are used at regulation of such parameters as:

1) Loudness of a speech signal;
2) Pitch of sound tone;
3) Timbre;
4) Duration of sounding.

Practice of the spent experiments shows, that quality of the synthesized speech in many respects depends on correctness of the chosen model of classification of phonemes, from its use in a complex with the optimized methods of updating of acoustic parameters and from sufficiency of an acoustic DB. Only having qualitatively realized level of phonetic synthesis of speech it is possible to start realization of other subsystems of speech synthesis.

References

1. E.N.Amirgaliyev, R.R.Musabaev. Some directions and problems of processing of linguistic data // Bulletin of Kazntu, 2007, # 6 (63), pp. 182-187. (In Russian)
2. E.A.Buraja, I.E.Galochkina, T.I.Shevchenko. Phonetics of modern English language. Theoretical kours.-M.: the Publishing center "Academy", 2008. (In Russian)
3. A.V.Babkin. Automatic synthesis of speech - problems and methods of generation of a speech signal // Works of the International seminar on computer linguistics and its appendices. Dialogue ' 98, 1998. (In Russian)
4. I. Aldoshina. Bases of psychological acoustics // Sound, 1999, #6. (In Russian)