MODELLING OF RUSSIAN INTONATION: A "CONTOUR INTERACTION" BASED ALGORITHM

E. Meister

Institute of Cybernetics, Tallinn, Estonia

ABSTRACT

The paper describes the algorithm of modelling of Russian intonation based on the theory of "Contour Interaction". As the basic units the algorithm uses four intonation patterns of syntagma and one pattern of word.

1. INTRODUCTION

There are two opposite theories to characterize the structure of intonation contours: the theory of Tonal Sequence (TS) and the theory of Contour Interaction (CI) /1/. The theory of TS sees intonation contours as being composed of an inventory of abstract tonal elements. According to the theory of CI an intonation contour can be viewed as the composite result of a set of hierarchial patterns. The classical approach to the Russian prosodic system / 2, 3/ is just based on the theory of CI. The basic category of prosodic system in Russian is "syntagma". It is defined as "the phonetic whole expressing one unit of meaning" /3/. The minimal unit of intonation is intonation contour of syntagma which could still be divided into the three functional parts: precentre, centre and postcentre. A special role in a syntagma is played by a centre because the changes of the pitch in the centre are the most important feature in distinguishing different intonation types. The inventory of basic intonation types according to /2/ includes seven different intonation patterns of syntagma.

The other essential category of Russian prosodic system is a word stress. In the algorithm introduced in this paper both categories are used as two interacting levels: intonation pattern of syntagma as a higher level and word stress as a lower level. Definitely, the number of interacting levels should be greater (for example the levels of sentence and phoneme should be added) but in the first stage the realization of this simple algorithm has fulfilled two main goals of this work: first, the theoretical goal - to test the validity of intonation patterns of syntagma described in /2,3/ and to find suggestions for further work, second, the practical goal - to get the real programm of intonation modelling for theRussian text-to-speech synthesizer.

2. THE ALGORITHM

2.1. Basic Patterns

In the algorithm four intonation patterns of syntagma described in /2/ have been realized: the declarative, the interrogative, the nonterminal and the exclamatory patterns.

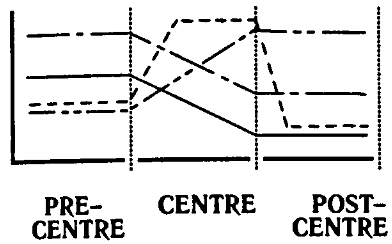

Fig.1. The intonation patterns of a syntagma:
——— a declaration,
- - - - - - an interrogation,
— - — a nonterminal,
—— - —— an exclamation.

To model the word accents one shape of pattern is implemented which is characterized by pitch level before accented phoneme, by rising pitch during the accented phoneme, by falling pitch during the next phoneme and by level during the rest of a word. The peak value of a word contour is approximately 10 % higher of the value of pitch level at the beginning of a word.

Fig.2. The word pattern.

2.2. Input Text

It is assumed that the input text is manually supplied with the marks of word-stress and main-stress because in Russian it is not possible to find correctly the location of the word-stresses without semantic parsing. The mark of main-stress (") is used once per syntagma and it is located on the most important word of a syntagma (on a centre of a syntagma). All other words are marked with a mark of word-stress ('). In order to distinguish between different intonation patterns the following punctuation marks are used at the end of syntagma: [.] - a declaration, [,] - a nonterminal, [?] - an interrogation, [!] - an exclamation.

These punctuation marks and also conjunctives are used as cues in dividing the input text into syntagmas.

2.3. Contour Generation

The generation of an intonation contour is the third step of the whole algorithm of speech synthesis. It is preceded by the grapheme-to-phoneme transformation and by the speech timing model. The minimal unit processed by the intonation algorithm is a syntagma. The input specification for the intonation algorithm contains the phoneme durations generated by the speech timing model and the string of stress- and punctuation marks. The algorithm works in three steps:

- determination of the intonation pattern of a syntagma,
- determination of the overall contour of a syntagma according to the durations of the precentre, centre and postcentre,
- superimposing of word accents into the overall contour.

The intonation contours are generated within the range from 80 up to 200 Hz.

3. TESTING

In order to estimate the validity of the intonation patterns used in the algorithm two methods were used:
- auditory estimation,
- comparison of the fundamental frequency contours derived from natural and from synthetic speech.

3.1. Auditory estimation

A set of short sentences consisting of one and two syntagmas with four types of intonation patterns were synthesized using the expert system /4/. The listeners were asked to recognize the type of intonation and location of the main stress. In most cases the type of intonation and location of the main stress were distinguished correctly. But in several cases the exclamatory contours were recognized as the declaration. It is due to the similarity of the contour shapes of the declaration and the exclamation.

3.2. Comparison of contours

In order to compare the fundamental frequency contours of natural and synthetic speech the sentence "МАМА МЫЛА МЕНЯ МЫЛОМ" ("The mother washed me with soap") was synthesized and also pronounced by the native speaker as the declaration and as the question with location of the main stress on different words. The fundamental frequency contours were extracted by peak-picker /5/. The visual comparison of natural and synthetic contours exhibits the similarity of the contours (Fig.3, Fig.4).

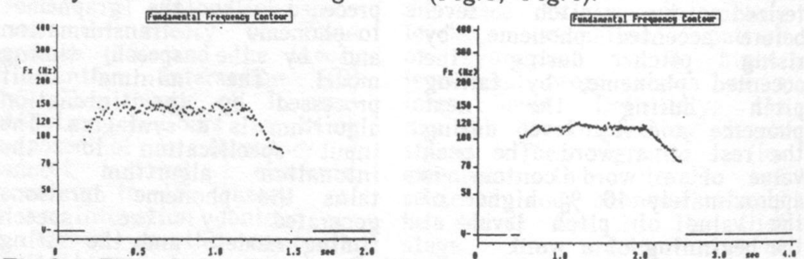

Fig.3. The declarative intonation contours of the sentence: МА'МА МЫ'ЛА МЕНЯ' МЫ"ЛОМ. left - natural, right - synthetic.

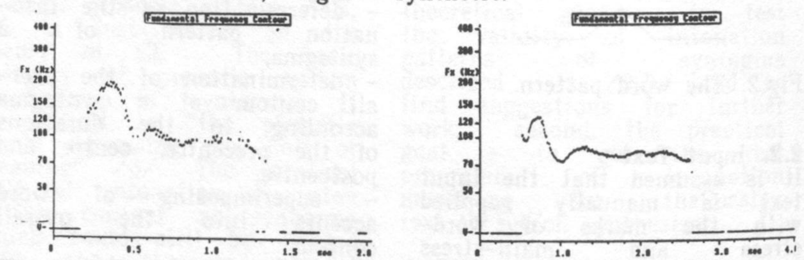

Fig.4. The interrogative intonation contours of the sentence: МА"МА МЫ'ЛА МЕНЯ' МЫ'ЛОМ? left - natural, right - synthetic.

4. DISCUSSION

The experiments with short sentences showed that the patterns of declarative, nonterminal and interrogative contours described in /2/ are valid for the use in synthesis algorithms. In order to express the exclamation with synthetic voice it is not enough to model the intonation contour correctly. The problems occur with longer syntagmas (more than 5 words) where the intonation contour sounds monotonous. On the one hand this may be caused by the fact that the current algorithm is based on the interaction of two hierarchial levels only: the level of syntagma and the level of word. Introducing into the algorithm the level of sentence and segmental perturbations the quality of synthesized speech will certainly improve. On the other hand the problem of monotonous intonation of long syntagmas can be overcome by manually dividing the long syntagmas into smaller ones (by inserting into the input text additional marks of punctuation and main-stress) although this is not always correct nor theoretically motivated. According to the experiments the optimal length of a syntagma is 3-4 words.

5. SUMMMARY

The results obtained in this work on intonation modelling can be formulated as follows:
- some of the intonation patterns described in /2,3/ are valid to practical use in synthesis systems,
- in order to express the exclamation it is not enough to model only the intonation contour correctly, the other prosodic features should be controlled adequately.

In order to better modelling of intonation:
- the levels associated with sentence and phonemes should be introduced,
- the rules dividing the input text into syntagmas with optimal length should be applied,
- the number of intonation patterns of syntagma should be increased.

REFERENCES

1. K.Silverman (1987) The structure and processing of fundamental frequency contours, Ph.D. thesis, University of Cambrige.
2. E.A. Брызгунова (1969) Звуки и интонация русской речи. Москва.
3. N.D.Svetozarova (1975) The Inner Structure of Intonation Contours in Russian, Auditory analysis and perception of speech. London, New York, San Francisco, Academic Press, p.499-510.
4. A. Ott, I. Siil (1987) The synthesis-by-rule development system with expert capabilities, Proceedings XIth ICPhS, vol.3, p.278-281.
5. D.M.Howard (1989) Peak-picking fundamental period estimation for hearing prostheses, JASA. 86(3), p.902-910.