APPROXIMATION OF INTONATION STRUCTURE OF SPEECH

TAMARA BROVCHENKO

Lab. of experimental phonetics
Odessa State University
Odessa, Ukraine, USSR 270021

VLADIMIR VOLOSHIN

Lab. of experimental phonetics
Odessa State Univeristy
Odessa, Ukraine, USSR 270021

ABSTRACT

The approximated intonation contours allow one to vizualise the most typical features of the melody and energy structure of the utterance in the form, directly appliable in automatic recognition and synthesis of speech prosody.

In a series of experiments discussed in the present paper typical intonation contours of various communicative types of phrases in Russian and English expressive conversation (as compared to the monotonous one) have been determined.

The most adequate methods of approximation of intonation contours have been analysed. Analytical expression which offers opportunity for presenting each intonation contour as a mathematical model has been suggested.

INTRODUCTION

In studing the intonation structure of speech a number of problems arise. Alongside with the problem of determining the physical nature of the phenomenon under study and defining typical intonation contours it is extremely important to elaborate the form of presentation of the intonation countours which should be precise and easy to apply.

The purpose of this paper is to compare the intonation structure of phrases read with expression to those read monotonously and to make an attempt to elaborate an analytical expression of typical intonation contours of expressive speech.

INTONATION CONTOURS OF EXPRESSIVE SPEECH

In our studies, five adult male speakers of British English and five speakers of Russian recorded a set of English and Russian written dialogues read with expression, lively and animatedly and then a set of the same dialogues, read monotonously, without expression. 20 statements, 20 questions (yes, no) and 20 request were picked out of these dialogues (a total of 600 utterances) and used for this experiment. The acoustic characteristics (fundamental frequency, duration and intensity) were measured for the two sets of the data. The problem was not simple that of describing the acoustic characteristics, but it was just as important to determine which of these characteristics are significant in discriminating expressive utterance and those read monotonously.

It has been commonly assumed that any speech realization is a random process which is described in terms of a functional dependence of the variable in time, whose parameter value can be presented with the help of the parametric equation:

$$X[t] = f[A_i B_i D_i C_i] \quad (1)$$

where A_i – constant parameters, unchangeable in all realisations;

B_i – interfering factors, varying from one realisation to another by some unknown law of distribution;

C_i – occasional interference, varying in separate elements of the utterance and describable by normal distribution;

D_i – the unknown parameters being sought, which determine the realisation as belonging to a given linguistic phenomenon.

In case occasional interferences are minimized, they will slightly influence the characteristics of the phenomenon under study, and the parametric model may be presented as a model with additive interference:

$$X[t] = E[D_i A_i B_i] + C_i \quad (2)$$

where $E[D_i A_i B_i]$ – range of parameters, describing the realisation being formed with no interferences present.

With various values of the parameters defined, function $E[D_i A_i B_i]$ gives a set of specific realisations as an ensemble, presenting phenomenon analysed.

The parametric model is described in the present paper in terms of discrete values of the fundamental frequency and intensity. These are associated with a definite number of points within each structural element of the utterance: 3 measurements within the initial unstressed syllables; 7 measurements within the head of the utterance (the first stressed syllable and all the stressed and unstressed syllables preceding the nucleus, 4 measurements within the nucleus and 2 within the tail. In total 16 measurements within each utterance. As a result the so called dynamic or temporal series was obtained.

Occasional interferences were reduced by the requirements of the procedure being kept fairly equal through the whole experiment. The realiability of the characteristics obtained in this experiment were ensured by statistically reliable number of speakers and amount of the experimental data. In addition only those utterances which were accurately identified by not les that 95% of the listeners were selected for further electroacoustic analysis.

The average values of the fundamental frequency and intensity were taken as a basis for a generalized intonation contour which reflects the main regularities of the phenomenon under study.

The results of these experiments have shown that the acoustic peoularities of expressive speech find vivid reflection in the dynamic series of the fundamental frequency, i.e. in the melody contour of various communicative types of phrases. It will be noted that the quality of speech (expressive or monotonous) determines the frequency level of the utterance, the speed of the fundamental frequency within the head and the nucleus, the location of the melodical peak of the utterance. These cues have been found typical of both English and Russian and it may be suggested that they are typological.

Figures 1 and 2 represent melody contour (dynamic series) of utterances in expressive and monotonous speech. The solid curve represents the average fundamental frequencies of statements, the dashed curve of questions, the dotted curve of requests. Structural elements of the utterance (P/h-initial unstressed syllables; h - head; n - nucleus, t - tail) were plotted as abscissae. Average normalized fundamental frequency as ordinates.

Average fundamental frequency values were normalized with the help of the equation:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \cdot 15 \qquad (3)$$

where X_i - selective value of the characteristic;
X_{max}, X_{min} - limit values of the characteristic.

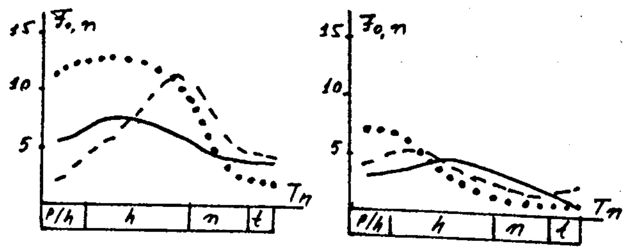

Fig. 1. Average normalized values of the fundamental frequency of utterance in expressive (left) and monotonous (right) Russian speech.

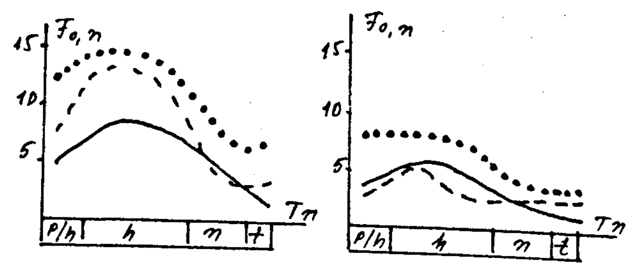

Fig. 2. Average normalized values of the fundamental frequency of utterances in expressive (left) and monotonous (right) English speech.

The experiments suggest that it is possible to establish the melody contours typical of expressive speech.

As to the values of intensity, the analysis reveals a relatively distinct difference between expressive and monotonous speech, the level of intensity both in English and in Russian being considerably higher in expressive speech.

On the other hand, the form of the intensity curve has shown remarkably little variation from utterance to utterance, from speaker to speaker, from one communicative type to another in expressive and monotonous speech. Commonly it has the shape of a gradually descending curve (fig. 3). The fact that different units of speech: a syllable, a sensegroup, a phrase, etc. - are characterized by a similar envelope of the intensity makes it possible to conclude that the form of the intensity curve is of paramouth importance in organizing units of speech.

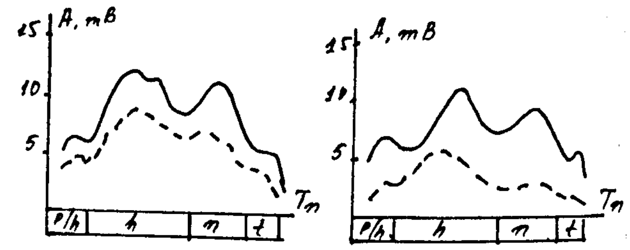

Fig. 3. Average normalized values of intensity of statements in expressive (solid curve) and monotonous (dashed curve) Russian (left) and English (right) speech.

Though the average values of the acoustic characteristics reveal the main regularities of the intonation contour it would not be sufficient to analyze only the average values. One should also study the varieties of acoustic cues in definite speech realizations.

As shown in Fig. 4-6 a number of realizations of utterances of the same communicative type make an ensemble within which it is possible to select from one to four main variants, differing to some extent in frequency, configuration of the curve, etc. In some cases the variants are equivalent and interchangeable in others they are dependent on the degree of expressiveness, on modal and emotional colouring and extralinguistic factors.

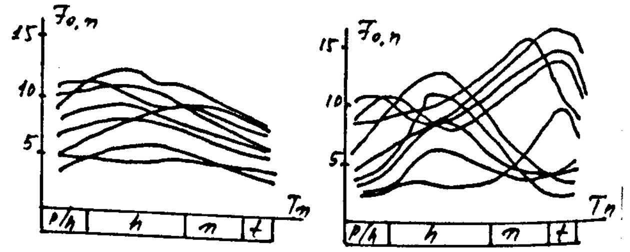

Fig. 4. Ensembles of intonation contours of statements (left) and questions (right) in Russian expressive speech (speaker RM_1).

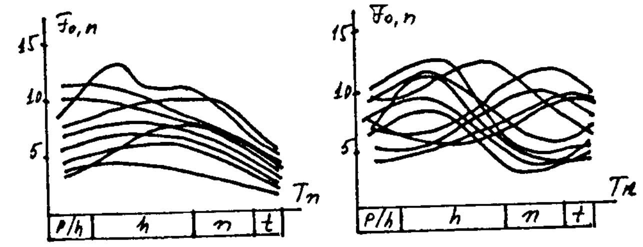

Fig. 5. Ensembles of intonation contours of statements (left) and questions (right) in English expressive speech (speaker EM_1).

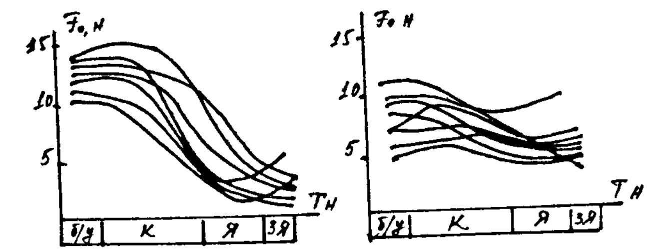

Fig. 6. Ensembles of intonation contours of requests in Russian (left) and English (right) expressive speech (speakers RM_1 and EM_1 correspondingly).

Attention should be drawn to the fact that specific intonation contours represent only one of many possibilities to make speech expressive. A quantitative study of the intonation structure of speech has suggested that besides the above mentioned acoustic cues of separate expressive utterances acoustic characteristics of the whole text might account for the difference between expressive and monotonous speech. Valid data were obtained showing that acoustic characteristics within the text provide effect of expressiveness of speech. Of particular interest in the present study, however, is that the correlation of the fundamental frequency and intensity values at the border of sense groups and phrases constituing the text, the correlation of acoustic measurements of initial and final unstressed syllables of different phrases of the text, etc. indicate whether the text is expressive or monotonous. Besides the alternation of different equivalent variants of intonation contours of one and the same communicative type of the phrase within a speech sample as well as the alternation of phrases with different level of intensity makes speech expressive.

These questions, however, are beyond the scope of the present paper.

APPROXIMATED INTONATION CONTOURS

Our final experiment aimed at the problem of approximation of the intonation contour of the utterance. There is a strong evidence to suggest that the main features of the intonation contour appear to be associated in the mind of the speaker with the communicative type of the utterance, its modality and emotional colouring, the degree of expressiveness and other linguistic and extralinguistic factors. It seems that initial and final values of the fundamental frequency and intencity, as well as the configuration of the curve are direct cues in "planning" the intonation contour of the utterance.

Taking it into consideration the values of the physical characteristics at the beginning and at the the end of the utterance, as well as the configuration of the curve were taken as a basis for approximating the intonation contours of expressive speech.

Variants of the trajectory of fundamental frequency and intensity measurements, obtained in the present study, could be readily approximated as close to the original as possible by analytical expressions, describing the intonation contour with the help of the method of least squares.

The method of analytical approximation includes: (1) establishing the character of the dependence and selection of corresponding equations: (2) minimizing trajectory deviations of the analytical expression from natural speech contour: (3) evaluating the constant coefficients that determine the trajectory of the changes in the parameters under study.

The analytical expression describing the trajectory of the fundamental frequency changes have been developed experimentally and calculated by the formula:

$$y[t] = F_{in}\, e^{-\alpha t^2 + \beta t} + F_{fin}\, e^{-K/t} \qquad (4)$$

where F_{in}, F_{fin} - values of the parameter at the beginning and the end of the speech sample:
t - successive number of time - segment values:
α, β, K - constant coefficients, selected for each realization in terms of the intonation contour.

For the analytical expression describing the trajectory of the intensity changes it is possible to express that function as follows:

$$y[t] = A_{in}\, e^{-\alpha t^2 + \beta t} \qquad (5)$$

where A_{in} - value of the parameter at the beginning of the speech sample.

In case of complicated curves (those having more that two turning points) the approximation is calculated by formula (4), with the beginning and the end of each structural element taken for the values of F's.

Coefficients α, β, K - determine the profile of the curve and account for the occational interferences and the parameters of the model sought for. Coefficient α varies in the range: $.01 \div .3$; β - $.1 \div .5$; K - $2 \div 14$.

Particular values of the coefficients used in approximating each intonation contour are given in Table 1.

Table 1. Analytical expression of intonation contours approximation

Communicative type of the phrase		Language	Analytical expression of approximation
Fundamental frequency	Statements	Russian	$y[t]=100e^{-.01t^2+.05t}+80e^{-6/t}$
		English	$y[t]=115e^{-.01t^2+.1t}+80e^{-8/t}$
	Questions	Russian	$y[t]=100e^{-.01t^2+.14t}+90e^{-5/t}$
		English	$y[t]=150e^{-.016t^2+.16t}+130e^{-10/t}$
	Requests	Russian	$y[t]=210e^{-.012t^2+.06t}+110e^{-5/t}$
		English	$y[t]=195e^{-.01t^2+.05t}+100e^{-55/t}$
Statements (intensity)		Russian	$y_1[t]=100e^{-.06t^2+.2t}$ $y_2[t]=106e^{-.05t^2+.42t}$ $y_3[t]=175e^{-.05t^2+.21t}$ $y_4[t]=125e^{-.033t^2+.04t}$
		English	$y_1[t]=100e^{-.07t^2+.23t}$ $y_2[t]=100e^{-.03t^2+.3t}$ $y_3[t]=150e^{-.06t^2+.23t}$ $y_4[t]=100e^{-.07t^2+.23t}$

The results of calculations are plotted in Fig. 7 - 10.

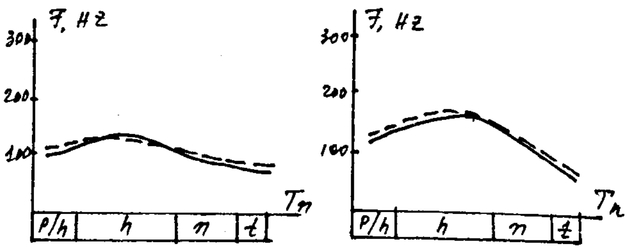

Fig. 7. Melody countours of statements in Russian (left) and English (right) expressive speech (solid curve) and their approximated variants (dashed curve).

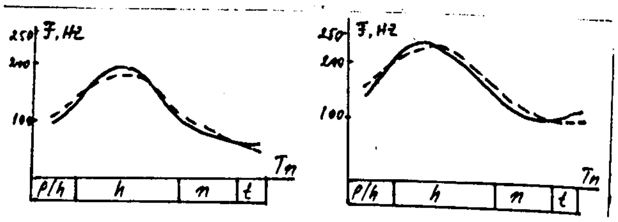

Fig. 8. Melody contours of questions in Russian (left) and English (right) expressive speech (solid curve) and their approximated variants (dashed curve).

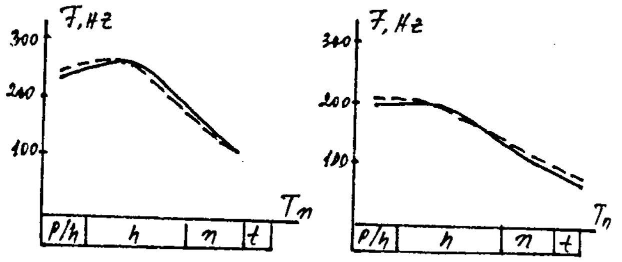

Fig. 9. Melody contours of requests in Russian (left) and English (right) expressive speech (solid curve and their approximated variants (dashed curve).

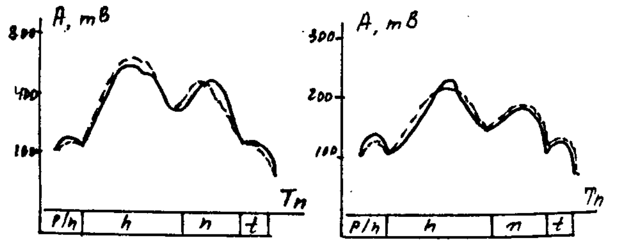

Fig. 10. Intensity contour of statements in Russian (left) and English (right) expressive speech (solid curve) and their approximated variants (dashed curve).

CONCLUSION

It appears from the foregoing analysis of the intonation structure of Russian and English utterances that differences in perception of degree of expressiveness are always associated with respective differences in the characteristics of the intonation contour of the phrase and those of larger speech units.

The analytical expression suggested enables to approximate the intonation contour of various types of expressive utterances close to the original intonation contours, preserving all their main properties. As compared to approximation by polynomial, the present method is more simple and effectual.

The presentation of the intonation contour as a mathematical model makes it possible to use it directly in the synthesis of speech prosody.

REFRENCES

/1/ Bengrande K.D. Text Discourse and Process. - London: Longman, 1980. - 231 p.

/2/ Vazen M. Statisticheskaja approksimatsia. - M. :Mir, 1972. - 295 s.

/3/ Kovalev V.P. Virazitelnije sredstva hudozestvenoj reohi. - Kiev: Rad. Sohkola, 1985.-136 s.

/ 4/ Torsueva I.G. Sovremennaja problematica intonatsionnych issledovanij //Voprosy Yazicoznaniya, 1984. - N 1. - S. 116-126.