APPROXIMATION OF INTONATION STRUCTURE OF SPEECH

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

ABSTRACT

The approximated intonation contours allow one to visualise the most typical features of the melody and energy structure of the utterance in the form, directly applicable 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 studying 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 contours 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.

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

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:

\[ X(t) = f(A_i, B_i, D_i, C_i) \]  

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:

\[ X(t) = E[D, A_i, B_i] + C_i \]  

where \( E[D, A_i, B_i] \) - range of parameters, describing the realisation being formed with no interferences present.

With various values of the parameters defined, function \( E[D, A_i, B_i] \) 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 equ-
The experiments suggest that it is possible to establish the melody contours typical of expressive speech.

As to the values of intensity, the analysis revealed relatively distinct differences between expressive and monotonous speech, the latter being usually higher in expressive speech.

On the other hand, the form of the intensity curve has shown several variants of 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 different unit of speech: a syllable, a monosyllable, a line, a syllable, etc. are characterised by a similar envelope of the intensity. It is possible to conclude that the form of the intensity curve is of paramount importance in organising units of speech.

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, the acoustic characterisation of the whole text might be expressed as an expression of the fundamental frequency changes and intensity measurements, obtained during the speech. These acoustic characters show a strong correlation between the fundamental frequency and intensity, as well as between the configuration of the curve and account for the occasional interferences and the parameters of the model sought for. The analytical expression describing the trajectory of the fundamental frequency changes have been developed experimentally and calculated by the formula:

$$y(t) = A \sin(\omega t + \phi) + \frac{e}{\epsilon}$$

where $F_0$, $F_{\text{min}}$, $A$, $\omega$, $\phi$, $\epsilon$ - parameters of the value at the beginning and the end of the speech samples; $A$ - amplitude of the segment coefficients, selected for each realisation in terms of the intonation contour.

The analytical expression describing the trajectory of the fundamental frequency changes is possible to express that function as follows:

$$y(t) = A \sin(\omega t + \phi) + \frac{e}{\epsilon}$$

$$\text{Appropriated intonation contours}
$$

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 modal 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 intensity, as well as the configuration of the curve are fixed in the mind of the speaker, and the utterance, facing its configuration value and the form of the configuration of the curve, etc. are characterised by the mental image of the curve.

The method of analytical approximation includes:

1. Establishing the character of the dependence and determination of corresponding equations (2) evaluating trajectory deviations of the analytical expression from natural speech contour (3) evaluating the coefficient values 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:

$$y(t) = A \sin(\omega t + \phi) + \frac{e}{\epsilon}$$

where $F_0$, $F_{\text{min}}$, $A$, $\omega$, $\phi$, $\epsilon$ are parameters of the value at the beginning and the end of the speech samples; $A$ - amplitude of the segment coefficients, selected for each realisation in terms of the intonation contour.

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

$$y(t) = A \sin(\omega t + \phi) + \frac{e}{\epsilon}$$

$$\text{Appropriated intonation contours}
$$

In cases of complicated curves (those having zero points of the curve), the approximation is calculated by formula (4), with the beginning and the end of each structural element taken for the value of $F_{\text{min}}$.

Coefficients $A$, $\omega$, $\phi$, $\epsilon$ determine the profile of the curve and amount for the interferences and the parameters of the model sought for. Coefficients $A$, $\omega$, $\phi$, $\epsilon$ vary in the range $[0.1, 1.5]; \omega$, $\phi$, $\epsilon$ - in the range $[1, 5.24]$.
Table 1. Analytical expression of intonation contours approximation

<table>
<thead>
<tr>
<th>Communication Language</th>
<th>Analytical expression of approximation of the phrase</th>
<th>Analytical expression of approximation of larger speech units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statements</td>
<td>$Y(t) = 100e^{-0.01t^2 + 0.02t + 0.03}$</td>
<td>$Y(t) = 100e^{-0.01t^2 + 0.02t + 0.03}$</td>
</tr>
<tr>
<td></td>
<td>$Y(t) = 106e^{-0.05t^2 + 0.04t + 0.05}$</td>
<td>$Y(t) = 106e^{-0.05t^2 + 0.04t + 0.05}$</td>
</tr>
<tr>
<td></td>
<td>$Y(t) = 175e^{-0.05t^2 + 0.06t + 0.07}$</td>
<td>$Y(t) = 175e^{-0.05t^2 + 0.06t + 0.07}$</td>
</tr>
<tr>
<td></td>
<td>$Y(t) = 125e^{-0.03t^2 + 0.04t}$</td>
<td>$Y(t) = 125e^{-0.03t^2 + 0.04t}$</td>
</tr>
<tr>
<td></td>
<td>$Y(t) = 100e^{-0.02t^2 + 0.03t}$</td>
<td>$Y(t) = 100e^{-0.02t^2 + 0.03t}$</td>
</tr>
<tr>
<td></td>
<td>$Y(t) = 150e^{-0.06t^2 + 0.07t}$</td>
<td>$Y(t) = 150e^{-0.06t^2 + 0.07t}$</td>
</tr>
<tr>
<td></td>
<td>$Y(t) = 100e^{-0.01t^2 + 0.02t}$</td>
<td>$Y(t) = 100e^{-0.01t^2 + 0.02t}$</td>
</tr>
</tbody>
</table>

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

Fig. 7. Melody contours 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 effective.

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

REFERENCES