Analysis and Interpretation of Intonation Contours of Slovene

Aleš Dobnikar

J. Stefan Institute
Jamova 39, 1000 Ljubljana, SLOVENIA
E-mail: Ales.Dobnikar@ijs.si

ABSTRACT

Prosodic characteristics of natural speech, especially intonation, in many cases represent specific feelings of the speaker at the time of the utterance, with relatively vast variations of speaking styles over the same text. We analyzed a collected speech corpus, recorded with ten Slovene speakers. Interpretation of observed intonation contours was done for the purpose of modelling the intonation contour in synthesis process. We devised a scheme for modeling the intonation contour for different types of intonation units based on the results of analyzing intonation contours. The intonation scheme uses a superpositional approach, which defines the intonation contour as the sum of global (intonation unit) and local (accented syllables or syntactic boundaries) components. Near-to-natural intonation contour was obtained by rules, using only the text of the utterance as input.

1. INTRODUCTION

Previous observations and studies of Slovene intonation [1-6] have rarely been made for large speech corpora and for the purpose of synthesis. Such insufficient knowledge led to our improving and redefining the rules for determining intonation contours (henceforth F_0) from phonologically relevant descriptions.

The scheme for modelling F_0 contours is based on results, obtained with the INTSINT system (INternational Transcription System for INTonation), proposed by D. Hirst [7, 8]. For generating an adequate F_0 shape, we need to know the relationship between linguistic units and the structures of an utterance with appropriate F_0 contour. This approach maximally reduces the amount of input prosodic information by applying a set of rules directly to the text. The so-called quantitative model of analysis and interpretation of the F_0 contour was proposed by many authors [9, 10, 11], with a differing number and complexity of the functions which try to simulate natural F_0 contours. In this paper, the so-called superpositional approach to modeling the F_0 contour is defined, which regards the F_0 contour as consisting of two different types of components:

- **global components,** related to the whole intonation unit, and
- **local components,** related to accented syllables or syntactic boundaries.

Global components represent the baseline for the whole F_0 contour for the intonation unit. The local components represent local movements (rise, fall, rise-fall) of the shape at accented syllables or syntactic boundaries. Syntactic boundaries with local ascent often indicate the final F_0 shape at various types of intonation units. The generated F_0 contour is then represented as the sum of both components.

2. SPEECH MATERIAL

In order to generate rules for our synthesis scheme, data was collected by analyzing the readings of ten speakers. All of them are native Slovene speakers, five males and five females. Eight of them (male and female equally) are professional speakers on national radio. The largest part of the speech material consists of declarative sentences, in short stories, monologues, containing sentences of various complexities and types, news, weather reports and commercial announcements. This speech data-base largely contains lexical emphasis and aims to be maximally intelligible and convincing. Other parts of the corpora are interrogative sentences with yes/no and wh-questions and imperative sentences. The first part of the corpora contains 500 declarative sentences, uttered by eight speakers, and the second part 100 questions and 30 imperative clauses uttered by 2 speakers.

3. ANALYSIS OF UTTERANCES

An intonation unit is defined as any connected part of speech between two pauses, longer that 40 msec. Shorter pauses did not represent boundaries between intonation units, because this length is the low-limited value for the duration of Slovene inherent phonemes [12, 13, 14].

The classical points of pauses in speech occur:

- at prefaces, new paragraphs and new topics of readings,
- at the end of clauses,
- at places of prosodic phrases inside clauses,

- at places of rhythmical division of some clauses, and
- at places of increased attention to some word or part of the text.

Depending on orthographic delimiters, four phrase boundaries were introduced:

- boundaries without classical orthographic delimiters:
  - at prefaces, between paragraphs, new topics of readings, ...
  - at rhythmical divisions in the clause - before the Slovene grammatical words *in, ter* (and), *pa* (but), ...
- boundaries with the delimiters '.' '...' '?' '!'
- boundaries with the delimiters ',' ';' ':' '-' '(...)' '"..."'

Pauses have a very important role in the intelligibility of speech. In the normal conversations typically half of the time consists of pauses; in the analyzed readings they represent more than 18 % of the duration. The obtained results show that pause duration is independent of the duration of the intonation unit before the pause, only breathing greatly depends on the duration of the pauses. The durations for various types of pauses are shown in table 1. Taking into account the fact that speakers show relatively vast variations of speaking style, the average values were taken to be the time between the first and third quartile, located around a median, because the mean value greatly depends on extreme values, often added for different reasons (physical and emotional states of the speaker, style, attitude,...).

Phrase boundaries:	first quartile	median	third quartile
at prefaces, between paragraphs, new topics of readings	1355	1658	1977
at rhythmical divisions in the clause	80	306	350
delimiters '.' '...' '?' '!'	651	857	1146
delimiters ',' ';' ':' '-' '(...)' '"..."'	119	273	417

Table 1: Average pause durations independent of breathing for both genders of speakers.

In the analysis of F_0 contours, the following parameters were studied:

- durations of different types of intonation units,
- average frequency of intonation units,
- onset frequency of intonation units,
- offset frequency of intonation units,
- details of the final F_0 part,
- frequency and number of syllable of the main accent and
- frequency and number of syllables of secondary accents.

The global component of the F_0 contour depends on the duration of the intonation unit, its average, onset and offset frequency; other parameters model the local components of the intonation contour. An accent was assumed to be any rise in frequency which differs more than 10 % in Hz from its vicinity. Every intonation unit has one main accent, which represent the highest value of pitch, and zero or more secondary accents.

Previous studies of Slovene macroprosody differ in determining the places of the main accents [2, 6, 15]. Results in table 2 show the percentage of intonation units where main accents appear at the beginning (first three syllables), at the end (last three syllables), at the other places and the

percentage of intonation units where frequency did not differ anywhere more than 10 % from its vicinity.

Position of main accents in the intonation unit:	Percentage:
At first three syllables	62
At last three syllables	19
At other places	13
Without main accents	6

Table 2: Percentage of intonation units with main accents at different places in the unit.

4. INTERPRETATION OF THE F_0 CONTOUR

After determination of word types, accent positions and phoneme durations in the sentence, interpretation of the F_0 contour consists in two main stages:

- text is divided into intonation units with punctuation marks and syntactic analysis,
- main and secondary accents are determined by rule, considering the type of words and the type of prosodic phrase (declarative statement, interrogative statement, yes/no question, non-terminal, ...).

The duration of pauses, determined from the type of the intonation unit, is in the range between the first and third quartile, as shown in table 1. Values closer to the median have a greater probability. This stochastic variance in the range of pause durations prevents the synthetic, discrete nature of pauses in synthetic speech. Many functions were tested for global (phrase) and local (accent) components (linear, power, transfer, decay, exponential), for the best approximation of the natural F_0 contour. In the presented system, an exponential function for the phrase component $P_c(t)$ [11, 16, 17] was adopted and a cosinusoidal function for accents and final boundary contours $A_c(t)$. The F_0 contour is thus defined by the following equation:

$$F0(t) = P_c(t) + A_c(t)$$

$P_c(t)$ and $A_c(t)$ are defined as:

$$P_c(t) = F_a \; e^{A_p \alpha \; t} e^{-\alpha t}$$

$$A_c(t) = A_a \; (1 + \cos\frac{T_a - t}{d})$$

where the expression $(T_a - t)$ must be in the range $(-\pi, \pi)$, otherwise $A_c(t) = 0$.

The symbols in these equations denote:

F_a: the asymptotic value of F_0
A_p: global F_0 maximum
α: parameter for F_0 shape control
T_a: time of accent
A_a: accent magnitude
d: local accent shape duration

The parameters are changed during the synthesis process according to the analysis results of the F_0 contour, word types, accent positions and phoneme durations in the sentence. Figure 1 illustrates the obtained results. The reference natural-speech sentence is uttered by a female speaker. The parameters for the synthesized F_0 are the same for the whole sentence. The panels display (top to bottom) the speech wave, the beginnings of the orthographic notation for every intonation unit, the original F_0 contour noted with the INTSINT system, indicated by squares, and the synthesized F_0 contour, generated with the equations presented, indicated by circles.

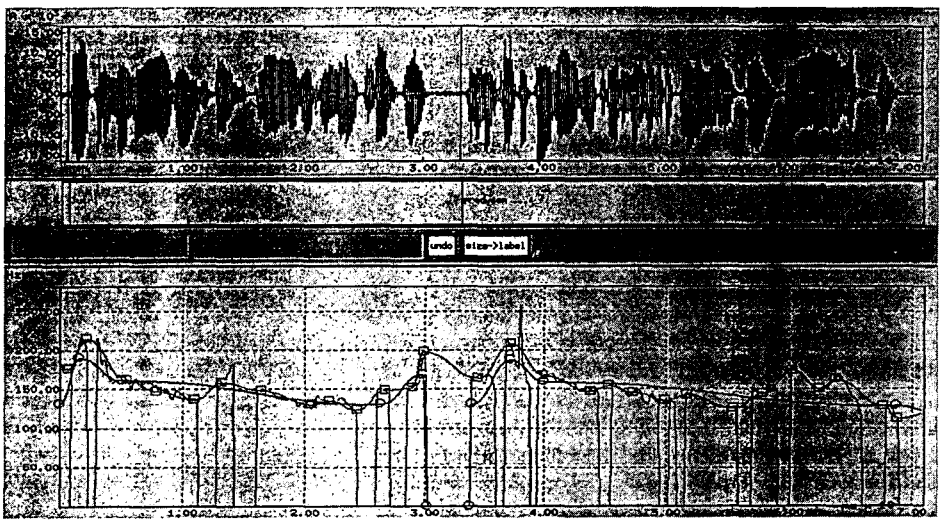

Figure 1: Example result of F_0 contour modelling with the INTSINT system for the Slovene sentence "Jutri bo jasno s spremenljivo oblačnostjo, predvsem popoldne in zvečer bodo še krajevne plohe in nevihte." Engl.: "Tomorrow will be clear, periodically cloudy, especially in the afternoon and in the evening local showers and storms could still come".

5. CONCLUSION

The paper describes an attempt to interpret the F_0 contour by rules for Slovene intonation units. A few authors mention the use of the paragraph as intonation unit for speech synthesis [18, 19]. However, these authors conclude that this feature, although significant, is as yet fairly vague. The reason is the insufficient knowledge of intonation organization of the paragraph and the fact that no account is taken of the effect of the text level on the intonation structure [7]. Analysis revealed that every speaker has a peculiar speaking style. One of the ways to define general intonation parameters could be to take the average values of large sets of utterances. The results of synthesized F_0 contour, based on average parameters, confirmed that the here presented model could roughly, but realistically simulate the natural F_0 contour. With additional information in the given text (especially levels and durations of local accents), the similarity of natural and synthesized F_0 contours was essentially improved. On the other hand, the analyzed speech corpus was limited, so that all aspects of the original speech could not be covered. Work towards further prosodic and linguistic analysis is still in progress.

6. REFERENCES

1. Toporišič, J., *Slovenska stavčna intonacija*, V. seminar slovenskega jezika, literature in kulture, Faculty of Philosophy, University of Ljubljana, 1969.
2. Toporišič, J., *Slovenska slovnica*, Založba obzorja, Maribor, 1984.
3. Toporišič, J. et al., *Slovenski pravopis 1 - pravila*, Slovenska akademija znanosti in umetnosti, DZS, Ljubljana, 1994.
4. Šuštaršič, R., *Kontrastivna analiza angleške in slovenske stavčne intonacije*, Ph.D. thesis, Faculty of Arts, University of Ljubljana, 1994.
5. Rakar, A., *Modul makroprozodike za oblikovanje stavčne intonacije v okviru sinteze slovenskega govora*, B.Sc. thesis, Faculty of Electrical and Computer Engineering, University of Ljubljana, 1995.
6. Vitez, P. and Aubergé, V. "Intonation Gesture of Slovene: First Indications", Proc. EUROSPEECH 95, Vol. 3, pp. 2073-2075, Madrid, 1995.
7. Nicolas, P. and Hirst, D.J. "Symbolic Coding of Higher-Level Characteristics of Fundamental Frequency Curves", Proc. EUROSPEECH 95, Vol. 3, pp. 2065-2068, Madrid, 1995.
8. Hirst, D. and Espesser, R. "Automatic Modelling of Fundamental Frequency", *MULTEXT*, LRE PROJECT 62-050, Task 2.6 Prosody Tools, Deliverable 2.6.1, Version B, Centre National de la Recherche Scientifique, 1995.
9. Moore, C.A. et al. "Quantitative description and differentiation of fundamental frequency contours", Computer Speech and Language, Vol. 8, Num. 4, pp. 385-404, 1994.
10. Taylor, P. "The rise/fall/connection model of intonation", Speech Communication, Vol. 15, Num. 1-2, pp. 169-186, 1994.
11. Fujisaki, H. and Ohno, S. "Analysis and modeling of fundamental frequency contours of English utterances", Proc. EUROSPEECH 95, Vol. 2, pp. 985-988, Madrid, 1995.
12. Srebot-Rejec, T., *Word Accent and Vowel Duration in Standard Slovene*, Slavistiche Beiträge, Band 226, Verlag Otto Sagner, München, 1988.
13. Vračar, V., *Modul za nastavljanje mikro prozodike vezanega besedila v okviru sinteze govora slovenskega jezika*, B.Sc. thesis, Faculty of Electrical and Computer Engineering, University of Ljubljana, 1995.
14. Bakran, J., *Model vremenske organizacije hrvatskoga standardnog govora*, Ph.D. thesis, Faculty of Arts, University of Zagreb, 1984.
15. Aubergé, V. and Bailly, G. "Generation of intonation: a global approach", Proc. EUROSPEECH 95, Vol. 3, pp. 2065-2068, Madrid, 1995.
16. Mixdorf, H. and Fujisaki, H. "A scheme for a model-based synthesis by rule of F_0 contours of German utterances", Proc. EUROSPEECH 95, Vol. 3, pp. 1823-1826, Madrid, 1995.
17. Hirai, T., Higuchi, N. and Sagisaka, Y. "Automatic detection of major phrase boundaries using statistical properties of superpositional F_0 control model parameters", Proc. EUROSPEECH 95, Vol. 2, pp. 1341-1344, Madrid, 1995.
18. Sluijter, A.M.C., Terken, J.M.B. "The development and perceptive evaluation of a model for paragraph intonation in Dutch", Proc. ICSLP 92, Vol. 1, pp. 353-356, Banff, 1992.
19. Terken, J.M.B., Collier, R. "Automatic synthesis of natural-sounding intonation for text-to-speech conversion in Dutch", Proc. EUROSPEECH 89, Vol. 1, pp. 357-359, Edinbourgh, 1989.