Jana Volk   UDK 811.163.6'342.9(497.4Slovenska Istra):81'322
University of Primorska*

USING THE TOBI TRANSCRIPTION TO RECORD
THE INTONATION OF SLOVENE

INTRODUCTION

ToBI is an acronym for *Tones and Breaks Indices*.[1] The term can be used in two ways: a) it originally denotes an intonation system developed between 1991 and 1994 designed to annotate intonation and prosody in the database of spoken American English (Mainstream American English); b) this transcription system rapidly spread onto prosodic annotation of other variants of English and many other languages. Terminologically, "MAE_ToBI" is used to denote the original meaning of the term, while "ToBI" is used for the evolving systems in individual languages. The paper presents the process of creating the transcription system, its basic parts and symbolic labels. It also deals with the annotation of Italian and speaks of the first ever application of the ToBI system onto a corpus of spontaneous speech of Slovene Istria.

1 ToBI

The MAE_ToBI was developed in a series of meetings which were attended by representatives of various disciplines: engineers, who wanted to learn about automated voice recognition and to develop a better system for the automatic conversion of written text to speech; psychologists, who wanted to explore the relationship between prosody and the process of human speech; computer experts, who wanted to build better models of dialogue and systems of speech; and phoneticians, who wished to test their theories about the integration of tone sequences and the alignment of tone and text. In the short run, they wanted to create tools that would allow researchers from different fields to work together in the development of a comprehensive prosodically annotated and freely accessible online database with a wide range of use in speech technologies and speech science. In the long term, they wanted to develop a common terminology that researchers from different fields could use for the interpretation of their data and one which they could consequently contribute to with further analyses and enhancement of the basic sets of methods and data.

The transcription system is based on extensive studies of English intonation and the segmentation of the speech signal. Any adjustment of the ToBI system to other

---

Author's address: Znanstveno-raziskovalno središče Koper, Garibaldijeva 1, 6000 Koper, Slovenia. Email: Jana.Volk@zrs.upr.si

[1] Unless otherwise stated, the description and presentation of the method are summarized following Beckman/Elam 1997, Beckman/Hirschberg, online source, ToBI, online source, Beckman et al. 2005.

languages should reflect the typical and theoretically-based understanding of intonation and prosodic grammar in the given language. Every new version of the ToBI-system should therefore be based (at least) on analyses of intonational phonology; ideally, the rules should be based on extensive studies in phonology, dialectology, pragmatics and discourse analysis. If in a given language only segments of the transcribed and analysed phenomenon were explored, the adjustment of the ToBI system could help in drafting relevant topics for further research. The MAE_ToBI group was diverse enough to enable the participants to reach an agreement that met the interests and needs of experts in various fields of interest. The group established a system which covered only those segments of prosody they wished to identify because the system had to be effective and couldn't afford to waste the transcribers' time with incidental tasks, such as the symbolic representation of undistinctive tone sequences which could be automatically extracted from the recording of the speech signal. The MAE_ToBI system was therefore aimed at a broader community of researchers with different interests and theoretical currents. The system allows all other evolving systems to include larger and more diverse groups of users, upon the condition that the system is in constant evolution and is accepted as a broader social standard. Rules should therefore be sufficiently simple so that their use would not be restricted to a handful of experts and trained transcribers.[2] A freely accessible guide for using the system, with many examples of already transcribed utterances from simple to complex structures is provided (http://www.ling.ohio-state.edu/~ToBI/). The rules of the new or adapted system must constantly be checked and updated. Also, multiple testing and evaluation of the transcribers' consistency in using the rules are important in the development process because they assure the researchers that the system is reliable (eg. Yoon et al. 2004). Setting up transcription rules is an ongoing process which requires agreement among all the participants in the process and adjustment to their needs and interests. Any proposed change to the original ToBI-system is based on a review of the speech material. A good ToBI-transcription therefore cannot and should not replace the recording of the speech signal only with a symbolic representation, but should include symbolic comments with data and acoustic recordings on which it is based.

The MAE_ToBI-system is based on the five most important facts about intonation and prosodic structure of language:

a) Prosodic patterns of the utterance can be presented/transcribed separately in individual tiers representing independent structural types. The intonational contour can be represented linearly with a series of tone sequences, while the metrical hierarchy of intonational phrases and minor prosodic units and its parts should be presented hierarchically, for example, with a scale of the perceived break index between two words.

b) Intonational contour is structured in relatively high and relatively low pitch levels. In order to indicate or describe a tone, H tags are used for high tone and L tags for low tone. These levels of labeling are, statistically speaking, in paradigmatic contrast with

---

[2] For further information on the purpose of transcription, cf. Šumenjak 2012.

one another – relatively high means high in the local pitch range of the observed phrase in comparison with the nearest low pitch peak.

c) Local pitch range is determined by various factors, e.g. the so-called downsteps and up-steps. A high tone (H) can therefore in some parts of the phrase be lower than a low tone (L) which means that high and low tones cannot be quantified in an absolute way.

d) Tones in any phrase functionally differ depending on whether they are boundary tones or whether they are included in the lexical pitch accent. The absolute value of the highest or lowest tone therefore depends on its function and its position – the lexical pitch accent is generally aligned with a corresponding accented syllable, while the boundary tone is aligned with the corresponding phrase boundary or the ending of the utterance.

e) The distinction between high and low boundary tones (H- and L- or H% and L%) is shown also at two different levels of intonational phrasing which indicates two different levels of the intensity of boundaries (phonological and intonational phrase).

According to the original agreement, the MAE_ToBI-transcription consists of six parts (cf. Table 1) – the audio recording, the graphical representation of the sequence of basic frequency, the symbolic representation of intonational contour, the ortho-graphic transcription of the utterance, the quantitative record of the degree or strength of disjuncture between words, as well as the transcriber's comments during the tran-scription of the utterance.

Audio	audio recording of the utterance in selected format
F0	electronic and/or manual representation of the sequence of fundamental frequency F0
Tones	transcription of the intonational contour and other tone-related data
Words	ortographic transcription of every word of the utterance aligned with the ending of the word
Break-indices	numerical index of the perceived degree of disjuncture between ortographically transcribed words
Misc	tags for lack of smoothness, comments, other data

Table 1: Basic elements of the MAE_ToBI transcription (Beckman et al. 2005).

The symbolic representation of prosodic elements is usually arranged in four tempo-rally aligned tiers and is also aligned with the corresponding contour of fundamental fre-quency F0 and the representation of the sound wave. The four parallel aligned tiers are:

- *tone tier* – the tier for transcribing tone sequences,
- *ortographic tier* – the tier for the ortographic transcription of the utterance,
- *break-index tier* – the tier for transcribing breaks between words,
- *miscellaneous tier* – the tier for noting down additional observations.

The basic transcription can be complemented with two additional tiers:

- *alternative tier* – the tier for alternative annotation in cases of ambiguity,
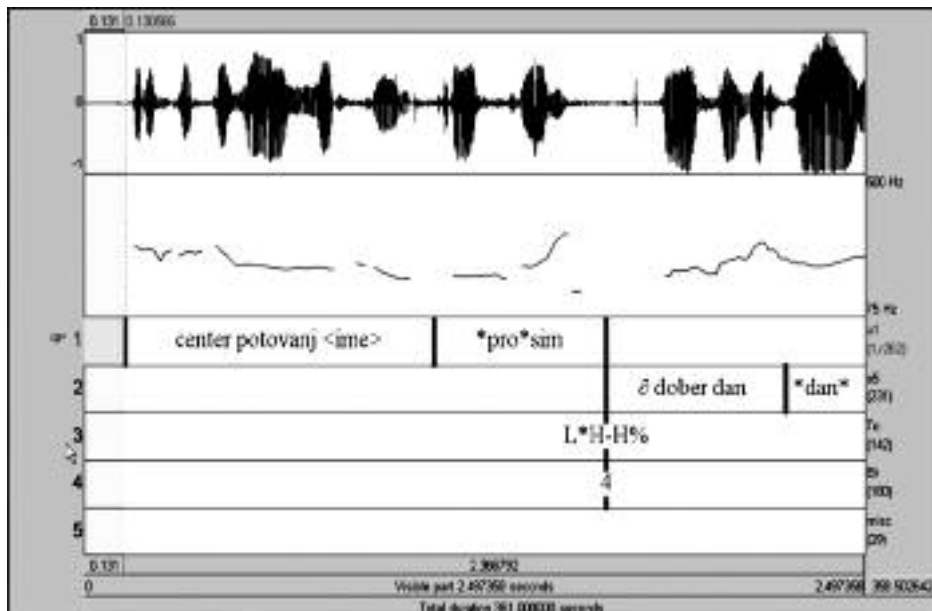- *discussion tier* – the tier for recording data which refer only to a part of the research results.

Figure 1: The use of the ToBI system in Praat.
(English explication: Speaker 1: travel agency <name> / how can I help you
                         Speaker 2: hello / hello)

1.1 Tone tier

The tone tier is that part of the transcription which corresponds most approximately to the phonological analyses of the intonational pattern of each utterance. It consists of labels for distinguishing pitch peaks or distinctive pitch events which are transcribed as a sequence of high (H) and low (L) tones. A tone may represent a pitch accent (H* or L*) or a part of it (H+L* or L+H*); in these cases we speak about bitonal pitch accent, or a phrasal tone which indicates the endings of two types of intonationally annotated prosodic units or phrases, high (H-) or low tone (L-) in a phonological phrase (phrase accent), and high (H%) or low tone (L%) in an intonational phrase (boundary tone). The tone is therefore related to the intonational boundaries or to the (boundary tone) after the accented syllable in each intonational phrase.

Phrasal tones
Phrasal tones are annotated for every *phonological* and *intonational phrase*:

172

a) labels L- and H- indicate a tone which occurs at the end of a phonological phrase; it is marked with number 3 in the break-index tier, which represents a similar annotation to the one used in the dissertation by J. Pierrehumbert (1980);

b) labels L% and H% indicate the tone which occurs at the end of the intonational phrase and is marked with number 4 in the break-index tier;

c) label %H denotes a high initial tone that starts relatively high in the speaker's voice register. Transcribers use %H only in cases where a high initial tone sequence cannot be associated with a high pitch accent on the first or second syllable of the utterance (when the first word is not accented or when the accented syllable is placed too close to the end of the word and therefore cannot be considered an initial accent) and when the utterance contrasts with a possible interpretation with low tonality.

An intonational phrase can consist of one or more phonological phrases, and can therefore exhibit one or two phrase accents and a boundary tone; the symbolic transcription of the sequence of phrasal tones can therefore be as follows:

a) L-L%, indicates an intonational phrase with a low phrase accent (L-), the last phonological (and therefore intonational) phrase ends in a boundary tone (L%) which falls to a lower point in the speaker's voice register. In American English, this sequence is typical of declarative sentences;

b) L-H%, indicates an intonational phrase with a low phrase accent (L) ending in a final high boundary tone (H%), which may indicate a continuation of the utterance;

c) H-H%, indicates an intonational phrase with a high (H-) phrase accent growing to a high boundary tone (H%). In English, this is typical of yes/no questions;

d) H-L%, indicates an intonational phrase with a high phrase accent (H-), followed by a low boundary tone (L%) which comes close to the middle of the speaker's voice register; this generates the final plateau.

Pitch accents

Pitch accents are annotated in every accented syllable. The MAE_ToBI transcription distinguishes five different annotations of pitch accents:

1. H* – *peak accent*; tone target in the accented syllable which is in the high part of the speaker's voice register in the pronounced intonational phrase. This label can also be used for tones in the middle voice register, but it excludes a very low level of F0;
2. L* – *low accent*; tone target in the accented syllable which is in the lowest part of the speaker's voice register;
3. L* + H – *scooped accent;* low tone target in the accented syllable, directly followed by a sharp rise into the upper part of the speaker's voice register;
4. L + H* – *rising peak accent*; high tone target in the accented syllable which follows relatively high from the lowest part of the speaker's voice register;

5. H + !H* – downstep to the accented syllable from a pitch, which in itself cannot be counted as a high boundary tone (H-) at the end of the preceding phonological phrase, or if in the same phrase the previous accent is the H-accent. It is used only when the pitch before the label is high and unaccented in the speech signal (otherwise it is marked by !H*).

As can be seen, pitch accents can be simple or monotonal, or complex or bitonal, which means that two tones (high and low) are combined in the same accented syllable. In the symbolic representation of pitch accents L*+H and L+H*, the speaker forms the same sequence of high and low tones – L H L H, but the difference in tone sequences is marked with the position of the asterisk (*). Both tone sequences contain a bitonal pitch accent in the same syllable, followed by a low boundary tone (L-) and a high ending (H%); they have identical numerical break indices and words. The difference is in the alignment of the pitch accent with the accented syllable. In the symbolic representation L*+H the initial tone, which is aligned with the accented syllable, is low (L) and growing towards the end of the syllable. In the symbolic representation L+H*, the syllable starts with a low tone which is growing throughout the accented syllable and reaches the top at its end.
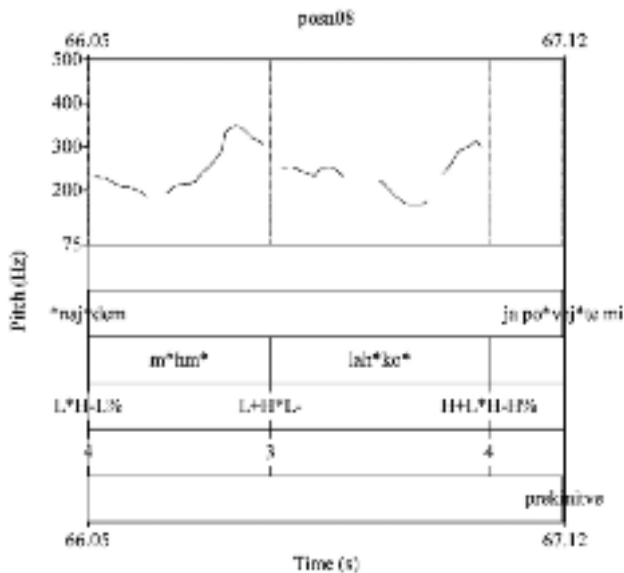

Figure 2: L + H* pitch accent in the word "mhm".
(English explication: Speaker 1: (let me) find / yes tell me
                          Speaker 2: aha / may I

1.2 Break-index tier

This tier indicates how strongly linked every word of the analysed segment is with the following word. A numerical scale from 0 to 4 is used for indicating different types of breaks. Index 0 indicates a low boundary between words, for example, in the blend-

ing of voices (e.g. the Slovene numeral "šest sto" (six hundred) which is pronounced "šesto" in the corpus) or in clitics. Index 1 is used in most word boundaries which occur between words in the middle of phrases (phrase medial word boundary). Index 2 is used for stronger boundaries with pauses or seeming pauses, but without tone breaks (the tone continues beyond this boundary) or for boundaries which are weaker than expected, as in cases with a clear intermediate or full intonational boundary between phrases. Index 3 is used for boundaries within intonational phrases (e.g., it only indicates the phrasal tone which affects the part of the phrase from the last accent to the end of the last phrase, i. e. the tone in boundary tones). Index 4 is used for the boundary at the end of intonational phrases or for final boundary tone.

Uncertainties regarding the annotation of break indices are marked with a minus (-) on the right side of the numerical index. Ambiguities (e.g. unclear or abrupt termination or prolongation of sound or word) are marked with a "p" on the right side of the numerical index. Ambiguity labels are only used with numerical indices 1, 2 and 3, where "1p" stands for unclear terminations and "2p" and "3p" stand for prolongation; more precisely, "3p" stands for delay after the pitch peak in a phonological phrase.

1.3 Ortographic tier

The ortographic tier is used for the transcription of the words of the analysed utterance. The transcription of the word is aligned with its position in the speech signal.

1.4 Miscellaneous tier

The miscellaneous tier is used for noting down comments; coughing, laughter, long silence and other non-verbal events are recorded in square brackets. Like the ortographic tier, this tier can include types of events which are not essential to prosodic analysis in itself, but can contribute to the interpretation of the tone tier and break-index tier analyses, since they interrupt the rate of speech and the annotated and presented intonational contours. Labels in this tier usually appear in pairs – they indicate the beginning and the ending of a given event; with the exception of the label *disfl*, which indicates lack of smoothness.

Basic tones	
Phrase accents	H-, (!H), L- (mandatory with BI 3 or higher)
Boundary tones	H%, L% (mandatory with BI 4) %H (marginal, at the beginning of certain intonational phrases, after the pause)
Pitch accents	L*, H*, (!H*), L+H*, (L+!H*), L*+H (L*+!H), H+!H*
Other labels	
Downstep	! (indicates the beginning of a falling tone)
Uncertainty	*?, -?, %? (indicating uncertainty regarding the presence of *, - or %) X*?, X-?, X%? (indicating uncertainty regarding the type of tone)
Phonetic events, transcribed and precisely annotated	< (delayed peak) HiF0 (highest F0, associated with accent in an intermediate phrase)
Restart	%r

Table 2: Inventory of tags of the MAE_ToBI system in the tone tier (Beckman et al. 2005).

With the aid of these suprasegmental aspects of spoken language, spoken texts can be interpreted semantically, syntactically and even morphologically. For this reason, researchers who investigate speech are keenly interested in the transcription of prosodic structure.

The ToBI-transcription system has been adopted in a number of languages and dialects, for example in German, Greek, Dutch, Serbo-Croatian, Japanese, Korean, Pan-Mandarin, Chinese, the Chinese dialect Cantonese, Native American language Chickasaw, Bininj Gun-Wok in Australia, Swedish, French, Italian, Portuguese, as well as in English and Irish dialects. Several new ToBI systems, adapted to a specific language, were created: C-ToBI for Chinese, J-ToBI for Japanese, K-ToBI for Korean, GR-ToBI for Greek, G-ToBI for German, SP-ToBI for Spanish, Cantonese-ToBI for Chinese dialect Cantonese, Pan Mandarin-ToBI for Pan-Mandarin, P-ToBI for Portuguese.

2 TOBIT

The reason why this paper presents the use of ToBIt, the ToBI system for Italian, is that the research and transcription of intonation in spontaneous speech of Slovene Istria was compared with studies of intonation in Northern Italian dialects.

The first attempt to adapt the system for the annotation of prosodic events proposed by Pierrehumbert (1980) was presented for Italian by Avesani in 1995. The author proposes an inventory of pitch accents and boundary tones for standard Italian which is based on data from the Tuscan dialect and the speaking style of a professional speaker. Avesani (1995) lists four types of pitch accents for Italian – two simple tones, H* and L*, and two composite tones, H+L* and L+H*, while the possible tone se-

quences for boundary tones are: L-L%, which is typical for completed declarative, imperative and exclamative contours and a possible wh- question contour, the H-H% and L-H%, which are typical for yes-no question contour or for indicating that the interpretation of the utterance rests on the utterance following it (continuation); and H-L%, which is phonetically realized as a plateau in the middle speech register and denotes the so-called calling tune ("richiamo") – it is used to name/call/attract attention of a person who is spatially distant from the speaker. The use of the adapted ToBI system for Italian therefore gives the possibility of distinguishing intonational contours which share the same phonological structure in the post-nuclear position (they discriminate the same boundary tones) based on different types of pitch accents and intonational contours which have the same type of pitch accent.

The system introduced by Avesani for the study of intonation of standard Italian began to be used for the study of intonation in other regional variants of Italian. Most research was done for Southern Italian variants: Bari, Naples, Palermo, Florence, Rome, Perugia, Treviso, Parma, Milan, Genoa, for Venetian Italian and for the nine provincial variants. All research dealing with Northern dialects shares in common the observation that these variants display a more pronounced tonal variation compared to the Central or Southern Italian dialects. Canepari (1980) finds that in the variants of Friuli-Venezia Giulia transitions between tonal planes are frequent. Using the ToBI method, tone sequence in completed utterances can be transcribed as L*H-L% (in standard Italian H*L-L%), in incomplete utterances as (H+L)*H-L% (in standard Italian H*H-L%), and in questions as (L+H)*L-H% (in standard Italian L*L-H%). According to the study of Payne and Folli (2006), the patterns of tone sequences in questions and statements in the Northern Italian town of Treviso are identical – falling (H+L*), rising (L*+H) or rising-falling pitch accent (L*+H or L+H*, followed by a phrase accent L-) in declarative sentences; all three types of accents may also occur in yes-no questions (the question form is determined from the syntax). Pitch accents in the studied materials are always bitonal (H*+L, H+L*, L*+H or L+H*). Payne (2005) transcribes the typical tone sequence of completed declarative utterances as a sequence of high and low tones – H L * H L, but does not specify a more accurate tonal structure of the analyzed data. Typical structure of incomplete utterances is L*H- and (H+L*)H-. For wh-questions, possible tone sequences are L*H-H%, and L*H-L% for questions associated with emotions, or H*L-L% in the case of a change in word order. In yes-no questions, the most frequent structure is H*L-L%, with a high final accented syllable, boundary tone(s) (are) mid to high. If yes-no questions are associated with emotions, the most frequent structure is L*H-L%.

3 TOBI IN SLOVENE MATERIAL

The described transcription method was first used in a corpus of spontaneous speech of speakers from Slovene Istria. The corpus consists of 12 audio recordings of telephone conversations between employees in travel agencies[3] in Slovene Istria and their (potential) customers. A total of 15 speakers participated, 5 travel agents (marked A1, A2, A3, A4 and A5 in the corpus) and 10 customers (marked with the letter S + numbers 1 through 10 in the corpus). The overall length of the recordings is 36 minutes and 17 seconds. The speakers were classified according to age, gender, place of residence, education, occupation, ethnicity and language (cf. Table 3).

speakers / variables		15
gender	female	11
	male	4
age	15–29 years	4
	30–44 years	10
	45–59 years	1
residence	coastal towns (Koper, Izola, Piran)	9
	hinterland of Slovene Istria	6
education	completed university or higher education	11
	uncompleted university or higher education	3
	university student	1
profession	independent trader/employer	2
	employee with upper secondary education	4
	employee with university or higher education	8
	other	1
nationality	Slovene	14
	Istrian	1
language	monolingual speaker of Slovene	3
	bilingual speaker of Slovene and Italian[4]	12

Table 3: Speakers according to variables.

The analysed material[5] shows the following pitch accents: monotonal H* and L*, bitonal (H+L)* and (L+H)* and tritonal H+L+H* and L+H+L*.

---

[3] We would like to thank Center potovanj and Partner travel agencies from Koper, and Laguna from Izola for their permission to record telephone conversations. Conversations were recorded for the basic research project Multilingual corpus of tourist texts: an information resource and analytical database of Slovene natural and cultural heritage (available online at http://jt.upr.si/turk).

[4] Nine bilingual speakers of Slovene and Italian learnt Italian from the environment prior to any formal education in primary school (e.g. through the Italian television and radio programs, Italian music etc.), while three respondents started learning Italian in primary school.

[5] For a more precise analysis of the material, cf. Volk 2011.
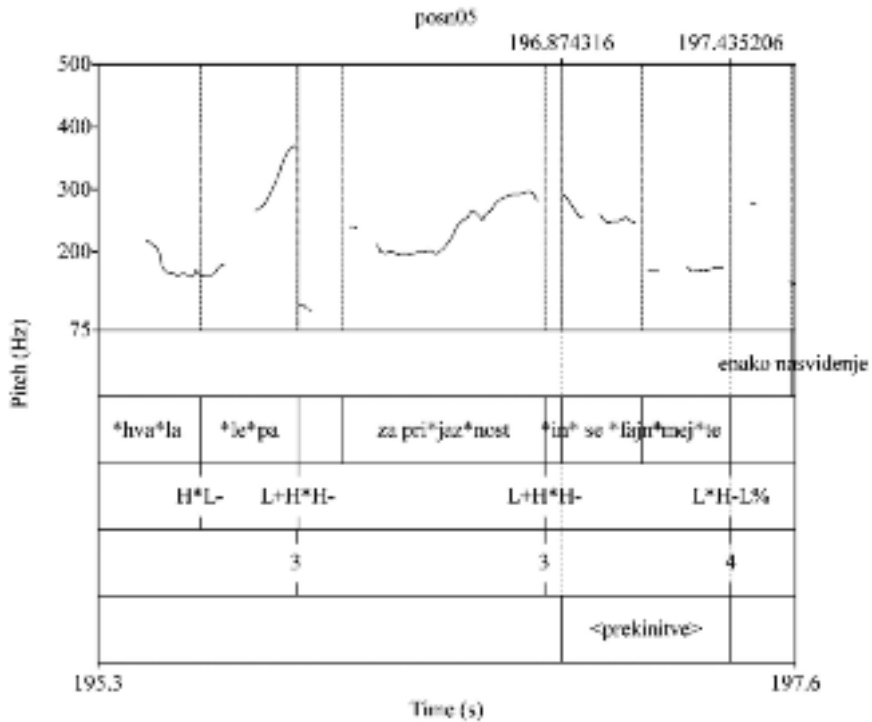
Figure 3: H* pitch accent in the word "hvala".
(English explication: Speaker 1: you too goodbye
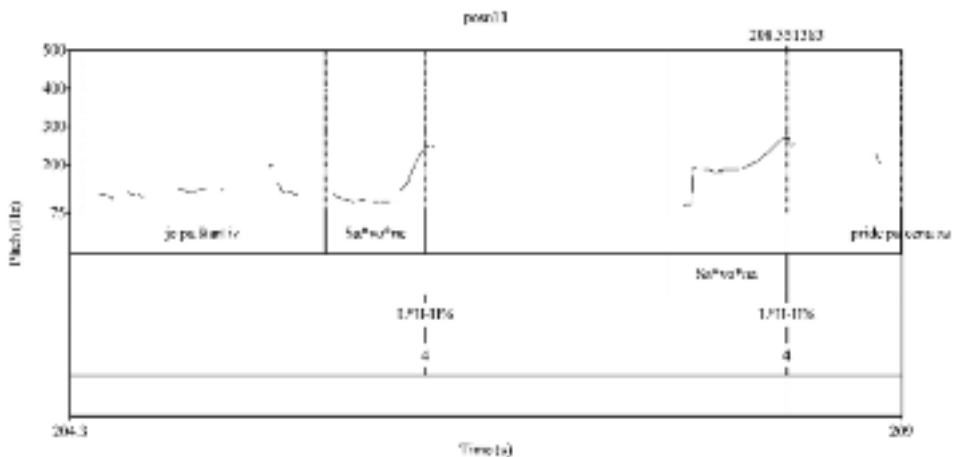   Speaker 2: thank you / for your kindness / and have a great time

Figure 4: L* pitch accent in the word "Savone".
(English explication: Speaker 1: it starts from Savona / and it costs
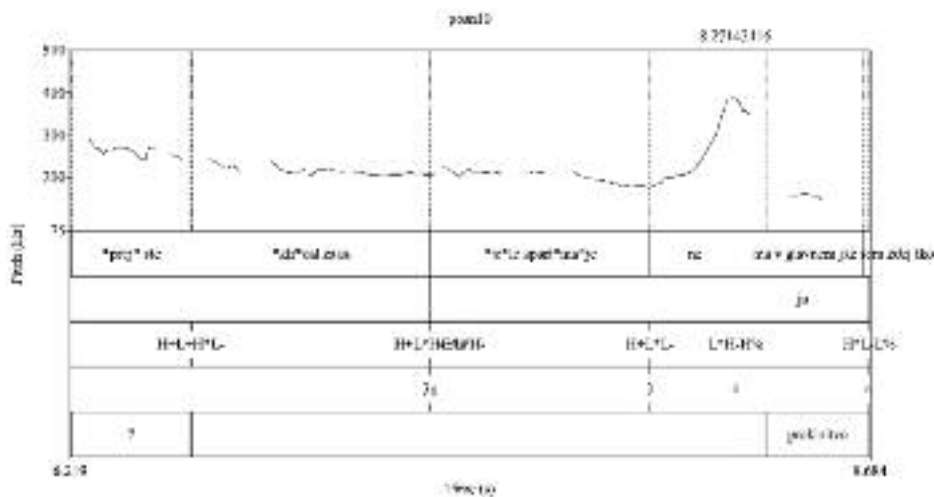   Spaekar 2: Savona

Figure 5: Tritonal pitch accent in the word "prej".
(English explication: Speaker 1: you called before fooor / these appartments /
right / so basicly now I
Speaker 2: yes

The phrasal tones found in the analysed material are: H- and L- (for incomplete utterances), and H-L%, L-L%, H-H%, L-H% for completed utterances, wh-questions and yes-no questions.

The analyses of boundary tones and phrase accents in different types of utterances show the expected tone sequences: L% for completed utterances and H- for incomplete utterances, L% for wh-questions and H% for yes-no questions, but the same boundary tone can appear in more than one type of utterance, and each type of utterance may have several possible tone endings. Incomplete utterances can therefore end in an unexpected low final tone (L), these are mostly utterances including discourse markers and/or speaker signals and utterances with a prolonged final sound; completed utterances may contain an unexpected high boundary tone (H%), especially utterances with speaker signals and/or discourse markers.
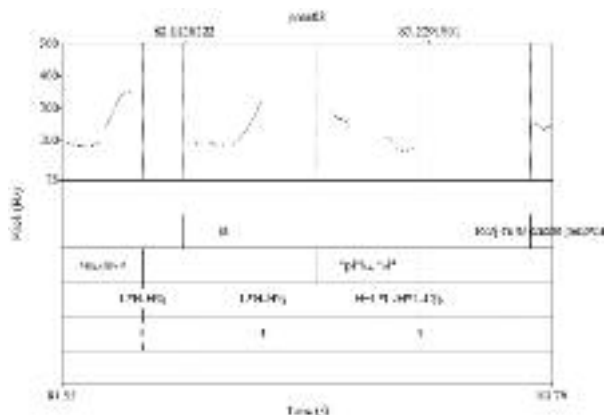
Figure 6: Expected L% tone boundary in completed utterance "pika si".
(English explication: Speaker 1: yes let me repeat it again
                     Speaker 2:<adress> / dot s i
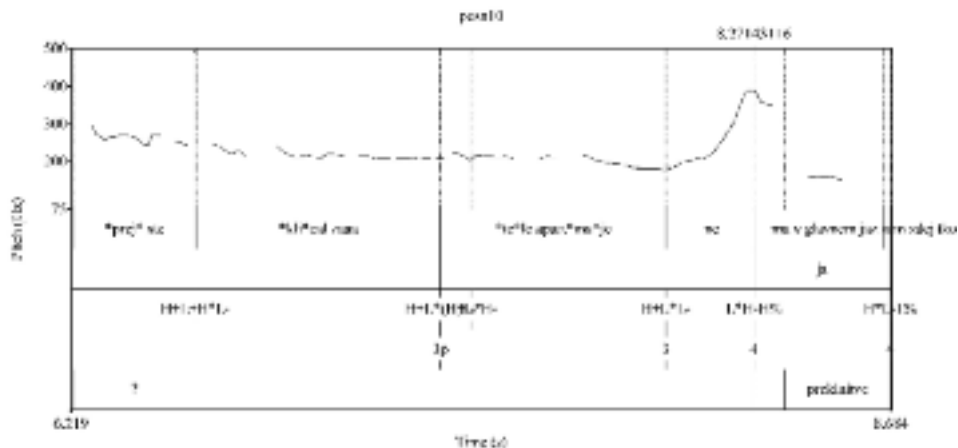
Figure 7: Unexcepted H% tone boundary in completed utterance "prej ste klical zaaa tele
                apartmaje ne".
(English explication: Speaker 1: you called before for / the appartments / right /
                     so basicly now I
                     Speaker 2: yes

Figure 8: Excepted H- tone sequence in uncompleted utterance "hvala lepa".
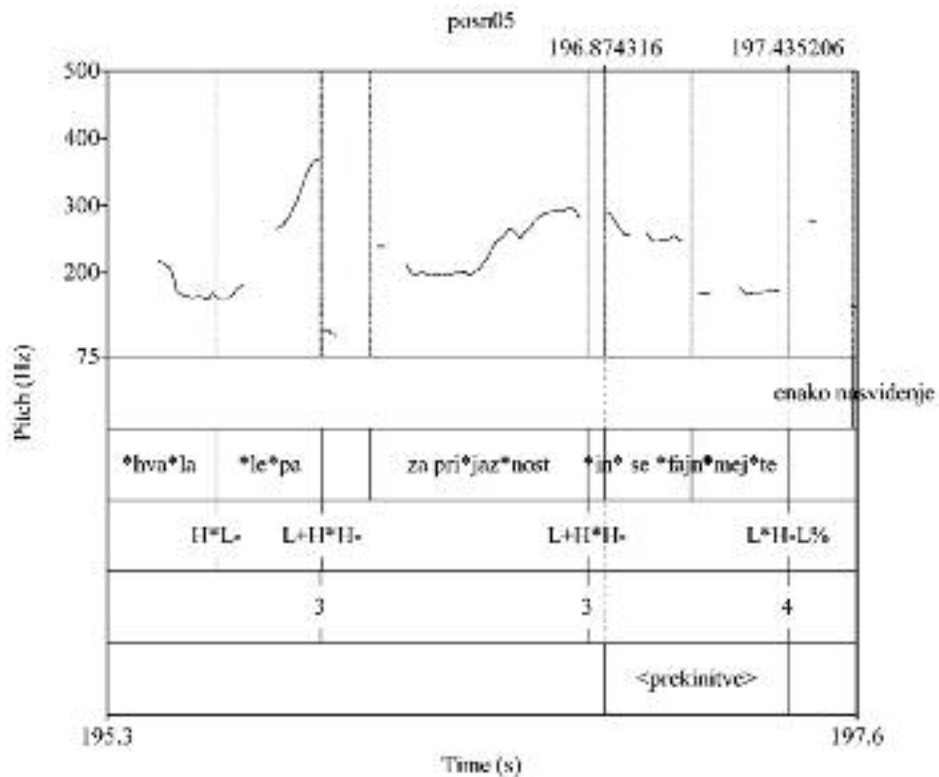(English explication: Speaker 1: you too goodbye
　　　　　　　　　Speaker 2: thank you / for your kindness / and have a great time
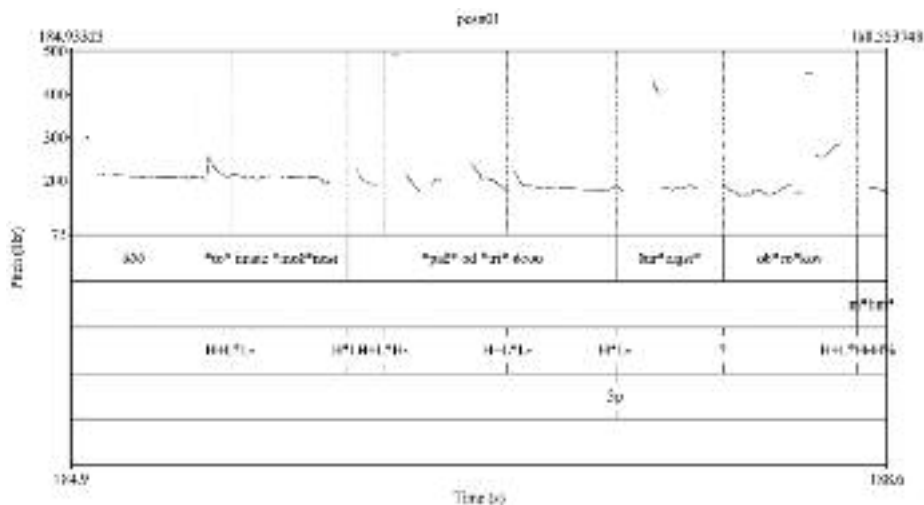
Figure 9: Unexcepted L- tone sequence in uncompleted utterance "(to imate možnost) pač od tri dooo".

(English explication: Speaker 1: er / you have the option / of well three tooo / fourteen installments
Speaker 2: aha

The results of the analysis show that tonal variation is a more important and evident feature of spontaneous speech of Slovene Istria than boundary tones. It appears in all types of utterances and is independent of high or low boundary tones. Related to this is the tritonal pitch accents and the so-called "valley" intonation pattern.[6] The original ToBI transcription system does not annotate tritonal pitch accents, but they are transcribed by researchers of Italian regional dialects, for example, in the town of Treviso in northern Italy, in cases where the tonal variations high-low-high or low-high-low take place in the accented syllable or when tonal variations before or after the bitonal accent cannot be attributed to preceding or subsequent unaccented syllables (e.g., in cases where the accented syllable, which is annotated as pitch accent, is also the first accented syllable of the utterance, or when two pitch accents follow each other).

---

[6] For details on the valley intonation pattern, cf. Volk 2011 and Volk 2011a. This intonation pattern is characterised by a sudden fall in the last accented syllable, followed by rising tones.

Figure 10: "Valley" intonation pattern in an utterance "ma je letalo iz Ljubljane".
(English explication: Speaker 1: Crete / yes yes the plane is from Ljubljana
                            Speaker 2: aha / what is / er / but the plane leaves from Ljubljana / or what

4 CONCLUSION

Since 1992, when it was presented for the prosodic marking of standard American English, the ToBI model has been adjusted to the needs of many languages and researchers have been using it in various fields of research, from linguistics to systemic engineering. There are (at least) two main reasons for the rapid acknowledgment and wide spread of the application: before 1992 there was no widely accepted and used system for the transcription of prosodic events which would include both intonation and voice flow segmentation in units of study; secondly, a pronounced growth in computational methods led to significant progress in speech identification and synthesis (Wightman 2002). Computer technology requires automated analysis of large speech corpora which must be annotated with standardized annotation strings. This need, combined with the need for a vast corpus, led to the creation of the ToBI transcription system. The paper presented the use of the ToBI transcription method in a corpus of spontaneous speech of Slovene Istria, with the aim of investigating whether this transcription system is useful for the prosodic annotation of Slovene. The analyses of tone boundaries show that the speech of Slovene Istria is characterised by a more frequent and more emphasised tonal variation in utterances and by the so-called valley intonation pattern, thus forming a basis for further development of the transcription system for intonation with adjustments, necessary for the particularities of Slovene language and its regional variants.

References

Avesani, Cinzia (1995) "ToBIt. Un sistema di trascrizione per l'intonazione italiana." In: G. Lazzari (ed), *Atti delle V giornate di Studio del Gruppo di Fonetica Sperimentale*. Trento: Tipografia Serigrafia, 85–98.

Beckman, Mary E./Gayle Ayers Elam (1997) *Guidelines for ToBI Labelling. Version 3.* Ohio: Ohio State Universisty. May 2012. http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/.

Beckman, Mary E./Julia Hirschberg (1993) *The ToBI Annotation Conventions.* 10 November 2011.

Beckman, Mary E./Julia Hirschberg/Stefanie Shattuck-Hufnagel (2005) "The Original ToBI System and the Evolution of the ToBI Framework." In: S.-A. Jun (ed), *Prosodic Typology. The Phonology of Intonation and Phrasing.* Oxford: University Press, 9–54.

Canepari, Luciano (1980) *Italiano standard e pronunce regionali.* Padova: Cleup.

Payne, Elinor (2005) "Rises and rise-plateau-slumps in Trevigiano." *Cambridge Occasional Papers in Linguistics* 2, 173–186.

Pierrehumbert, Janet (1980) *The phonology and phonetics of English intonation.* Massachusetts: Massachusetts Institute of Technology.

Šumenjak, Klara (2012) "Zasnova dialektološkega korpusa na primeru govora Koprive na Krasu." In: B. Krakar Vogel (ed), *Slavistika v regijah – Koper. Zbornik 23. Slovenskega slavističnega kongresa.* Ljubljana: Zveza društev Slavistično društvo Slovenije, 73–78.

ToBI. 20 June 2012. http://www.ling.ohio-state.edu/˜ToBI. 20 June 2012.

Volk, Jana (2011) *Intonacija v spontanem neformalnem govoru slovenskih govorcev na dvojezičnem področju Slovenske Istre.* Ljubljana: Univerza v Ljubljani.

Volk, Jana (2011a) "Intonacija v spontanem govoru v Slovenski Istri." In: M. Jesenšek (ed), *Globinska moč besede: red. prof. dr. Martini Orožen ob 80-letnici.* Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, 379–395.

Wightman, Colin W. (2002) "ToBI Or Not ToBI?" In: B. Bel/I. Marlien (eds), *Proceedings from Speech Prosody.* Aix-en-Provance: Laboratoire parole et langage, 25–29.

Yoon, Tae-Jin/Sandra Chavarria/Jennifer Cole/Mark Hasegawa-Johnson (2004) "Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation Using ToBI." *Proceedings of the ISCA International Conference on Spoken Language Processing (Interspeech 2004).* May 2012. http://www.isca-speech.org/archive/interspeech_2004/i04_2729.html.

Abstract

USING THE TOBI TRANSCRIPTION TO RECORD
THE INTONATION OF SLOVENE

The paper presents ToBI, a transcription method for prosodic annotation. ToBI is an acronym for Tones and Breaks Indices which first denoted an intonation system developed in the 1990s for annotating intonation and prosody in the database of spoken Mainstream American English. The MAE_ToBI transcription originally consists of six parts – the audio recording of the utterance, the fundamental frequency contour and four parallel tiers for the transcription of tone sequence, ortographic transcription, indication of break indices between words and for additional observations. The core of the transcription, i. e. of the phonological analyses of the intonation pattern, is represented by the tone tier where tonal variation is transcribed by using labels for high tone and low tone where a tone can appear as a *pitch accent*, *phrase accent* and *boundary tone*. Due to its simplicity and flexibility, the system soon began to be used for the prosodic annotation of other variants of English and many other languages, as well as in different non-linguistic fields, leading to the creation of many new ToBI systems adapted to individual languages and dialects. The author is the first to use this method for Slovene, more precisely, for the intonational transcription and analysis of the corpus of spontaneous speech of Slovene Istria, in order to investigate if the ToBi system is useful for the annotation of Slovene and its regional variants.

Keywords: ToBi, intonation, spontaneous speech, Slovene Istria.

Povzetek

UPORABA TRANSKRIPCIJSKE METODE TOBI ZA ZAPIS
INTONACIJE SLOVENSKEGA JEZIKA

Članek predstavlja ToBI, transkripcijsko metodo za zapis prozodičnih dogodkov. ToBI je kratica za *Tones and Breaks Indices*, ki izvirno poimenuje intonacijski sistem, ki je bil razvit v 90-ih letih prejšnjega stoletja in zgrajen za označevanje intonacije in prozodije v podatkovni bazi govorjene ameriške angleščine (Mainstream American English). MAE_ToBI transkripcija po prvotnem dogovoru sestoji iz šestih delov – iz zvočnega posnetka izreka, zapisa poteka osnovne frekvence in štirih vzporedno poravnanih pasov, ki so namenjeni transkripciji tonskega poteka, ortografskemu zapisu izreka, označevanju jakosti mej med besedami ter zapisovanju dodatnih opazovanj. Jedro zapisa oziroma fonoloških analiz intonacijskega vzorca predstavlja tonski pas, v katerem z oznakami za visoki in nizki ton transkribiramo razlikovalna tonska nihanja. Sistem se je zaradi svoje enostavnosti in prilagodljivosti hitro razširil na prozodično označevanje ostalih variant angleščine in mnogih drugih jezikov ter na različna nelingvistična področja, nastali so številnih novih ToBI-sistemi, prilagojeni posameznim jezikov ali narečjem. Metoda je bila prvič uporabljena za zapis in analizo intonacije na korpusu spontanega govora govorcev v Slovenski Istri z namenom preizkusiti, v kolikšni meri je ToBI primeren za opis intonacije slovenskega jezika in njegovih pokrajinskih različic.

Ključne besede: ToBI, intonacija, spontani govor, Slovenska Istra.