Acta Linguistica Hungarica, Vol. 49 (3–4), pp. 277–306 (2002)

THE MOST IMPORTANT PROSODIC PATTERNS OF HUNGARIAN*

GÁBOR OLASZY

Abstract

Prosody is a general term for the following features in speech: pitch and intonation, stress, articulation rate, sound intensity and time structure (rhythm and pauses). During verbal communication various prosodic forms contribute to the expression of the content of the message (the information carried by the text, emotional expression, to imitate a situation etc.). So, prosody can be represented as a multivariable function in which the number of variables is rather high. Therefore it is difficult to describe the complex process for all situations, meanings, and emotions. In this paper we try to give a phonetic level characterization of pitch and intonation structure and also the function of intensity in time of the main Hungarian sentence types (using a unified description). The manner of description is new concerning Hungarian. It is based on a unified relative scale in which not physical values but relative distances in pitch values and intensity are used to characterize the melody forms and the intensity levels. This description allows for the representation of these two prosodic elements independently of the personal features (mean F_0 value, the range of the F_0 of the speaker, etc.). The representation makes it possible to express the crossfunctions among the melody forms of different expressions. This means that complete prosodic patterns can be predicted for any text without an acoustic analysis.

1. Introduction

Examination of the prosodic structure (mainly intonation patterns) of continuous speech has become more and more important in the last decade while the fields of applications of automatic speech generation have grown drastically due to the industrialization of information technology. In these applications increasingly better speech quality is needed in text reading (continuous news reader, e-mail reading, various talking services like book reviews, weather forecast, prose reading, etc.), and also in services where automatic dialogues are realized between the machine and the client. A number of models have been constructed in the last decade to describe the inherent structure of intonation—e.g., for Dutch (Collier 1990; Terken–Collier 1990) for German

* This research was partly supported by the Hungarian Research Fund OTKA 030292.

(Möbius 1997); for Japanese (Fujisaki 1992); for English (Silverman et al. 1992; Taylor 1998; 2000). Also, the research on "emotional synthesis" seems to be increasingly important in constructing life-like verbal situations between humans and machines (Montero et al. 1999). The detailed, phonetic level modelling of the prosody of Hungarian, verified by speech synthesis experiments, has been completed recently (Olaszy 2000; Olaszy–Koutny 2001).

In earlier works on Hungarian, mainly melody patterns were studied. The first systematic investigation was performed by Fónagy and Magdics (1967). They examined the melody form of a few hundred sentences by ear, and the description of the melody was presented as a series of musical notes in a five-line system. This description gave only some general information about the melody patterns of Hungarian. Later works (Olaszy 1989; Varga 1993) also examined intonation from various points of view. Varga described a phonological assumption about Hungarian sentence melody forms. He represented the melody forms by schematic lines which were drawn between two theoretical horizontal lines representing the highest and the lowest F_0 values of the speaker. The first perceptual measurements on the melody forms of statements, questions, commands and exclamations were done by Gósy (1992). She used special audio material in which only the fundamental frequency of real speech was present, the higher frequency components were eliminated. These speech stimuli were produced by a special F_0 imitator device for which the input was real speech and the output was the melody in audible form (i.e., test subjects did not hear the content of the recorded utterance, only its melody form).

The goal of the present research was to define the most important components of Hungarian prosody. Another goal was to construct a generalized manner of description. A unified relative F_0 and intensity scale has been defined in which not physical values but distance values are used to characterize the melody forms and the intensity levels.

2. Material and method

The speech material for this research contained 800 sentences, mainly statements, questions, commands, warnings and requests. The sentence structures were also diverse, ranging from simple one-word sentences to longer ones up to complex, long sentences and even short dialogues containing 2–3 sentences. The text material was read by a male speaker (a 58-year-old trained speaker, born in Budapest, speaking everyday Hungarian) digitized with 22kHz, 16 bit, labelled by pitch period markers, as well as sound and word boundary signs, by a semiautomatic Hungarian software (Olaszy et al. 2001). The average articulation rate of the speaker was 13 sounds/s.

As to the method of melody and intensity curve representation, a generalized manner of description was used. The melody and intensity patterns are described with stylized straight lines in a relative scale. The same reference level is defined for all sentence types. By applying a relative scale the definition of a reference level is arbitrary. Most of the earlier authors take the speaker's sentence final pitch value as the low reference. In what is called the superposition model (Fujisaki 1992), the linguistic pitch contour is treated as if it were some sort of complex function which can be decomposed into simpler component functions (e.g., accent on a prominent word) and overlaid or superimposed on global shapes (e.g., the distinction between a statement and a question). In Fujisaki's model a low reference F_0 value (speaker specific) represents the fundamental point for the superposition of the phrase component and the accent component. The pitch values are then expressed by distance functions from the reference level. This approach is based on the experience that in declarative sentences the dispersion of the final (lowest) F_0 values of the speaker is relatively small, about 3% (Möbius 1997). Ladd (1996) compares this model to the 'target and transition' models which are based on the ToBi idea (Silverman et al. 1992). He thinks the advantage of a phonological (target and transition) model of intonation is that speaker pitch becomes a relatively low-level realization parameter. In a superposition model, it is difficult to distinguish language-specific or universal aspects of intonation from speaker specific features of pitch range.

In the present work, basically the idea of the superposition model was used. The difference is that sentence structure was taken into consideration when defining the reference level (Olaszy 2000). Another difference is that the same description philosophy is applied to F_0 and to intensity. Given that, of all sentence types, it is statements that occur the most frequently in speech, the reference level for the F_0 calculation was decided to be the initial pitch and intensity values of the simple declarative sentence (the reference value is 1, i.e., 100%). By this solution the relative differences among sentence types as a function of declarative sentences show a clearer structure than in the earlier methods. We think that the initial part of the sentence has the main role—as to the general shape—in speaking. The modality of the sentence in Hungarian can be predicted already from the initial part of the sentence. Therefore, it is appropriate to define the initial F_0 points of all sentence types as a function of the declarative sentence. The sentence final parts have been defined as a function of the ending of declaratives. Another advantage of this method is that the rules for transforming the F_0 patterns from one modality to another (i.e., to generate a question, or a request, or a command from a statement) also show clearer structure, because the reference value is attached to a real sentence mode. The main melody structure of all sentence types is described in the same scale. The reference for intensity is defined similarly. The reference level (0 dB) is the beginning point of the declarative sentence. The above representation is independent of personal features (mean F_0 value, the range of the F_0 of the speaker, etc.). Applying the generalized stylized patterns, complete Hungarian prosody patterns (for longer texts, dialogues, etc.) can be predicted if the person-dependent reference F_0 is given, e.g., 100% = 125 Hz.

Three levels of pitch changes have been used to describe pitch structure: the phrase level main melody contour as a carrier item and the word and syllable level modifications as local F_0 movements that are superimposed on this main contour. A sentence can be made up of one or more intonation phrases. Local F_0 changes may occur within the intonation phrases mostly in relation to accentuation and boundary marking. Word level modifications represent those text parts in which the F_0 change is characteristic of the whole word. For example, articles and conjunctions are treated as unaccented words in which the F_0 is lower within the whole word than in the main contour. The syllable level F_0 changes represent mostly positive modifications in the main melody form (accented syllables and positive F_0 changes in boundary marking or in questions). In Hungarian, the accent is placed invariably on the first syllable of the word. In this description we use two levels to characterize the status of the syllable: syllable with positive F_0 change (accented or marking boundary, etc.) and neutral. The neutral status belongs to those syllables in which the F_0 is the same as in the phrase level pattern. The pitch changes in general are stylized, with three major contour types: falling, level and rising.

The slope of falling or rising depends on the one hand on the duration of the time interval where the contour is present, and on the other hand on the minimal and maximal frequency values of the frequency band in which the fall or rise movement is realized. The combination of these two factors may determine several exact melody contours as building elements of the final melody.

3. The unified melody and intensity representation

Both the F_0 and the intensity structure of the analyzed sentence types will be described in a unified scale with stylized lines. For the F_0 changes the phrase level main pitch contour is given with its beginning and end points as a carrier element. The word and syllable level additional changes are given under this scale in the word (W) and syllable (S) lines as multiplication factors to modulate the main contour (similarly to Fujusaki's representation). The value of the multiplication factors may vary between 0.5 and 1.5. The modulation is calculated from the F_0 values of the main pitch contour. Thus a range reduction is realized as well. The stylized contours on the figures below show the main, phrase level shape (thin line) and the local changes (thick line). The local changes overwrite the thin lines, i.e., they are valid for the final F_0 contour. The description of intensity structure follows the same philosophy.

3.1. Statements

The phrase level main melody form for statements (Figure 1, overleaf) is a continuously falling pattern. If the sentence is short, the original F_0 pattern and the stylized one are roughly the same. If the sentence is longer, word and syllable level pitch changes may modulate the falling melody form in positive or negative directions. The pitch curve of the sample sentence in Figure 2 (page 283) shows that unaccented words (articles) have lower F_0 values than their surroundings, whereas the accented syllables (marked with dots) have pitch peaks. Microintonation also modulates the pitch curve on the sound level (marked with downward pointing arrows in Figure 2), but these segmental level changes are not involved in the present description.

In statements the declination in pitch was found to range between 30% and 42%. The lowest F_0 endpoint was realized if the sentence was pronounced in isolation, or if it was the final one in the text. The shape of the intensity structure was similar to that of the pitch change; the range of the declination was between 15 and 20 dB along the whole sentence (Figure 2).

The lower F_0 value in unaccented words depends on their place within the sentence. The greatest difference compared to the main pattern can be measured if the sentence begins with such an element (see in Figure 2, marked with the upward pointing arrow). In this case the negative modification may reach the factor 0.8.

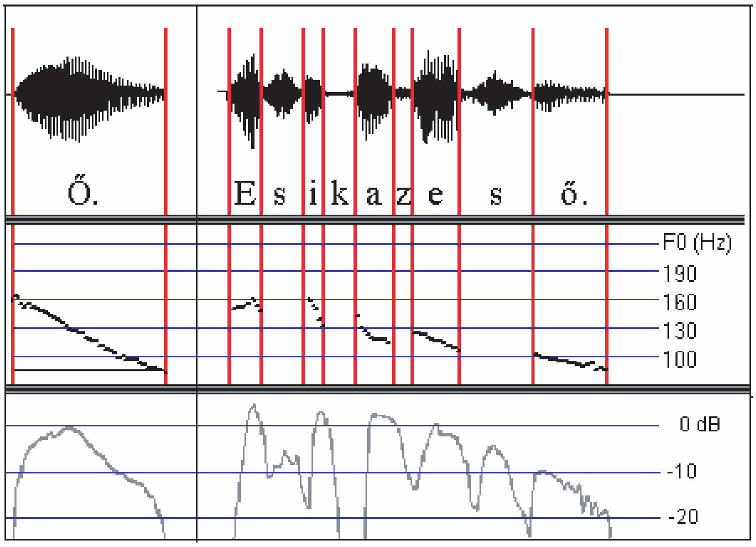

Fig. 1
The declination in pitch and intensity for two simple structure short statements (*Ő.* 'He/She.'; *Esik az eső.* 'It is raining.'). The vertical lines show sound boundaries

In the case of complex statements, several intonation phrases (falling patterns) make up the whole sentence. The initial F_0 value is at the reference point, the sentence final one is on the same value as it did in simple statements. The intermediate falling patterns show a sawtooth structure which itself also has a slight declination. An example is shown in Figure 3 on page 284. Commas separate the sentence into three main falling patterns. The comma effect is expressed both on word and on syllable level, i.e., the equalization of the falling F_0 into a level one is expressed by the word level modification, while the final rise may be expressed by the syllable level one. The result in the word before the comma will be similar to what can be seen in the natural F_0 pattern. In the second falling pattern two word-accents and one comma effect form additionally the main F_0 contour. The third, final falling pattern ends on 58% and contains two syllable level changes. The unaccented parts of the sentence (*hogy*; *a*; *aki már*) are marked with negative word level modifications. The accents show pitch maximums in the first syllable of the accented words (*azt*; *Péter*; *levelezik*; *három*; *külföldön*).

As to the intensity structure of complex declarative sentences (see the lowest part of Figure 3), the unaccented parts (marked with arrows) have

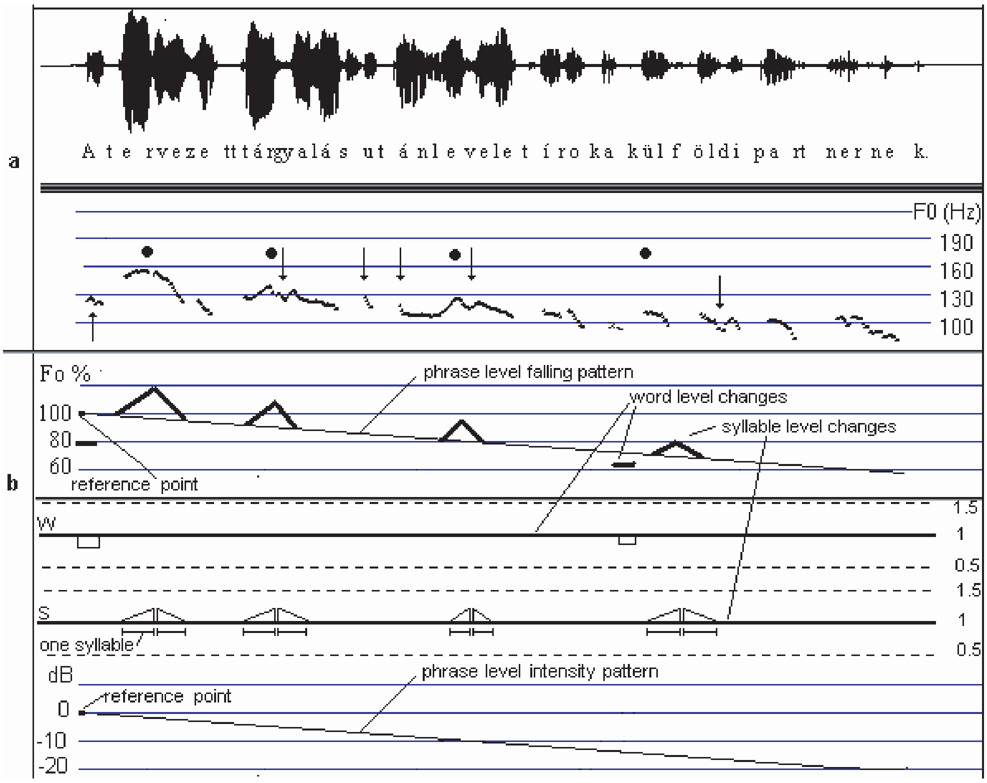

Fig. 2

The F$_0$ structure (a) and the stylized representation (b) of a statement.
(*A tervezett tárgyalás után levelet írok a külföldi partnernek.*
'After the planned discussion, I will write a letter to the foreign partner.')

lower intensity than their surroundings. The intensity level is close to 0 dB in the first two clauses, a declination to −20 dB is present only in the last one.

Summing up the main F$_0$ features of complex statements in Hungarian, it was found that in general the range of declination is about 40%, independently of the length of the sentence. The internal phrase level intonational parts have also falling F$_0$ structure (each). In very long sentences the slope of the declination in one intonation phrase can be so small that practically the F$_0$ structure shows a level form. The effect of comma represents a syllable level change into a rise or level form in the final part of the word before the

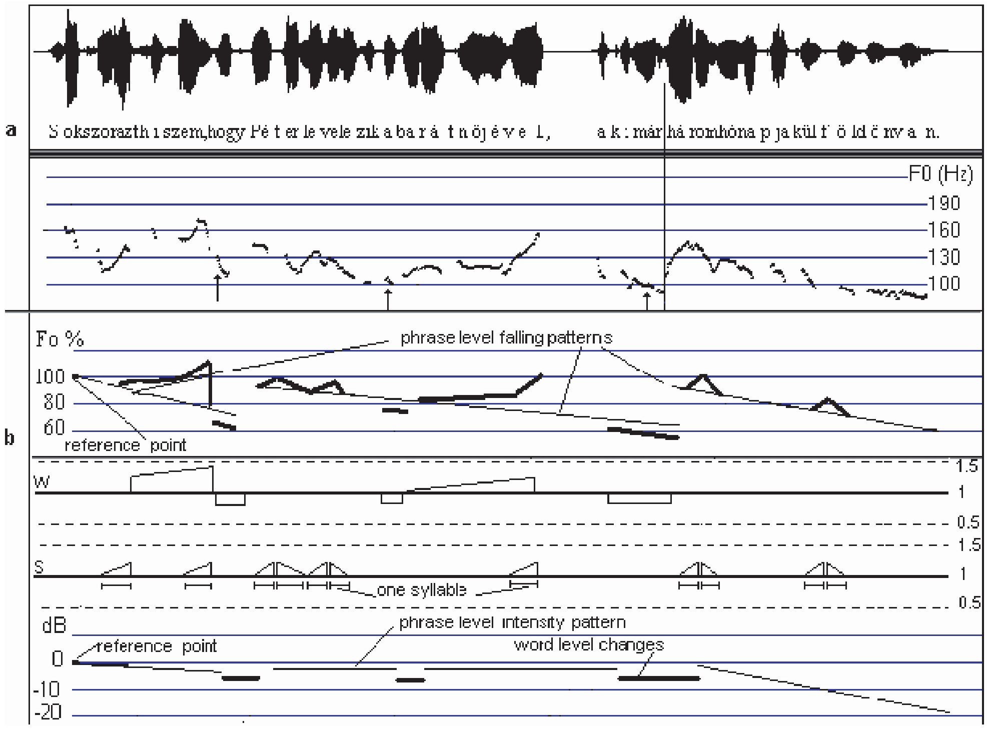

Fig. 3
The F_0 structure (a) of a complex statement and its stylized (b) representation.
Sokszor azt hiszem, hogy Péter levelezik a barátnőjével, aki már három hónapja külföldön van. 'I often think that Peter is corresponding with his girlfriend, who has been staying abroad for three months now.'

comma. The unaccented elements have mostly lower F_0 values than their surroundings, whereas the accented syllables have higher values.

The final prosody realizations for statements may be influenced by the content of the text and also by the intention of the message. Different styles are used, for example, in news reading or in prose interpretation. A traffic information announcement has its special style as well. If names and addresses are read in an information system, their prosody also has special elements. All this means that in speech technology applications the exact prosodic structure of statements can be determined after a detailed study of the diverse texts and purposes of the application.

3.2. Questions

The melody patterns in Hungarian interrogative sentences vary to a large extent, depending on various features. Besides the two main categories (yes/no and *wh*-questions) there are other question types and subtypes with individual melody patterns. The melody forms in questions may also depend on the length of the question (one, two or more syllables), on the internal structure of the sentence and on the intention and emotion of the speaker. The intensity structure of certain questions shows different characteristics from those in statements, and in certain questions the sound durations are strongly lengthened.

3.2.1. *Wh*-questions beginning with a Q-word

The minimal structure of this type of question is: Q-word + one word, e.g.:

(1) **Mikor** indultok?
‘**When** will you start?’

The main F_0 structure for *wh*-questions is a falling pattern, which starts from a lower value (about 80%) and ends on a similar point as it did in statements, i.e., the slope of the falling pattern is flatter in these questions than in statements. This form is realized independently of the length of the question.

(2) (a) **Kivel** fogtok most találkozni?
‘**Who** will you meet now?’

(b) **Mikor** írod meg a levelet az édesanyádnak?
‘**When** will you write the letter to your mother?’

A syllable level F_0 modification in the Q-word realizes the question intonation; word level modifications do not occur. The syllable level high-low modification is as follows: the F_0 value is high in the first syllable (the peak may reach 130%) and is reduced in the second (Figure 4). The right proportion between the peak and the slope of the main falling pattern determines the proper intonation of the whole question. The higher the peak value and the lower the starting point of the main falling pattern the more characteristic the question will be. Other syllable level modifications (word accents) do not appear in the descending part.

There exists another variant for the pronunciation of these questions (Gósy 1993). The difference between the standard rendering (described above) and the variant is in the pronunciation of the final part, i.e., people may raise the F_0 in the last syllable. This rise is about 10% compared to the F_0 value of the last but one syllable. Another difference is that the main F_0 pattern is not falling but it shows a level character. It begins with a slightly lower F_0 frequency than in the standard version and this level is kept until the last syllable. This difference can be explained by the fact that the human prediction mechanism for F_0 generation decides the ending form of the question already after the pronunciation of the question word. If the decision is low ending (standard version), a descending part will be produced after the question word. If the decision is to rise up at the end, the same part will be changed into level form to prepare the way for the rise at the end.

The intensity structure of *wh*-questions shows very similar structure to what it was like in statements.

3.2.2. *Wh*-questions with a topic

If a *wh*-question has a topic part before the actual question, the melody structure can be represented by two phrase level patterns. The topic has a slightly rising form, whereas the question part is the same as described for simple *wh*-questions. The topic part before the question begins with a lower F_0 value (about 80%) and has a slowly rising (to 85–90%) character which prepares the way for the question (Figure 5, page 288).

3.2.3. Complex *wh*-questions

In complex forms the *wh*-question is followed by another clause.

(3) **Mikor** mész az üzletbe és veszed meg a kávét?
‘**When** will you go to the shop and buy the coffee?’

In these cases the question part has similar characteristics as in the simple *wh*-questions, but the descending part will end higher. This higher ending can be explained by the fact that the sentence has not been finished at this point, it will be continued. In the additional part the descending F_0 change is continued until the very end of the complex sentence. The very final F_0 value is close to that in simple statements (60%). Word accents may occur in the additional clause.

If the complex *wh*-question contains more than one question word, one falling pattern is present over the complex question and the syllable level peaks in the Q-words will have consequently lower and lower F_0 values along the sentence, realizing the range reduction.

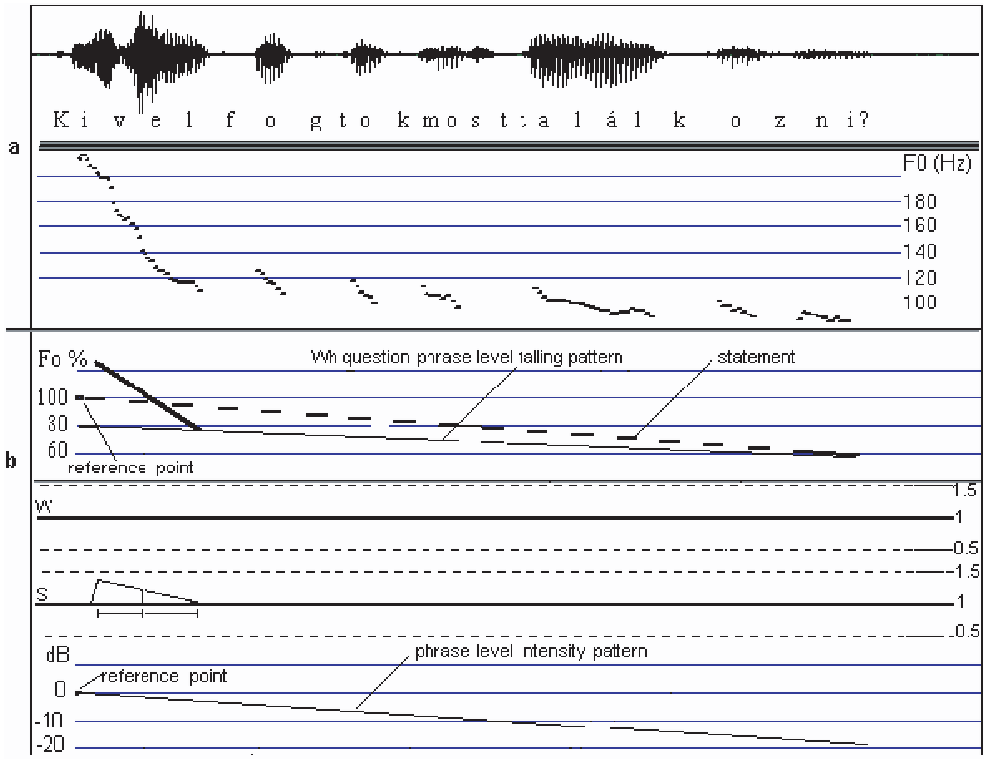

Fig. 4
The typical F_0 pattern of a *wh*-question longer than three syllables (a) and its stylized (b) representation

(4) **Mikor** fejezed be a munkát és **mikor** jössz haza?
‘**When** will you finish work and **when** will you come home?’

3.2.4. Yes/no questions and their environment

The main intonation pattern of yes/no questions can be of a rise-fall or a level-fall form (Figure 6, page 289). If rise-fall is realized, the starting point is lower (80%) than in statements and the end of the rising part is about 100%. This rising structure prepares the way for the F_0 peak of the questioning part, which is placed at the beginning of the last but one syllable. In the second version, the level pattern starts from 110%. The falling part ends in both versions close to the same value as in statements (about 60%). The question intonation is realized by the sharp pitch jump and fall in the last but one

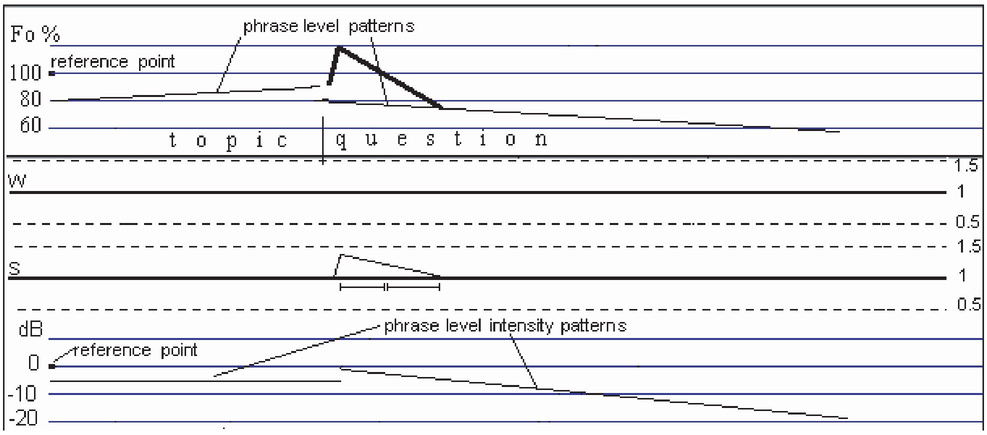

Fig. 5
The stylized patterns of a *wh*-question with a topic
Ezt a témát illetőleg ***mikor*** *válaszoltok a kérdéseimre*?
'Concerning this topic, when will you answer my questions?'

syllable. The jump is realized at the beginning of the nucleus of this syllable and the fall ends at the end of this syllable. The peak in this syllable may reach 120–130%.

Word accents are not present in the slowly rising part. This can be explained by the structure of this question type. The first part only prepares the way for the peak at the end which expresses the main information. The second form of this question with the level-fall intonation is pronounced in the case of expressing impatience or anger.

3.2.5. Yes/no questions with topic or focus

In this case the sentence is divided into two phrase level F_0 patterns. The sentence begins with a slightly falling structure (from 100% to 80%) until the end of the topic or the word in the focus position. This is followed by the second pattern which is similar to that shown in Figure 6B.

(5) (a) Holnap délután **elmentek végre moziba**?
'Tomorrow afternoon **will you go finally to the cinema**?'
(The question part is marked by bold letters.)

(b) A tegnap kiadott **szakácskönyvet** vetted meg a barátnődnek?
'Did you buy the **cook book** published yesterday for your girlfriend?'
(The word in focus is marked by bold letters.)

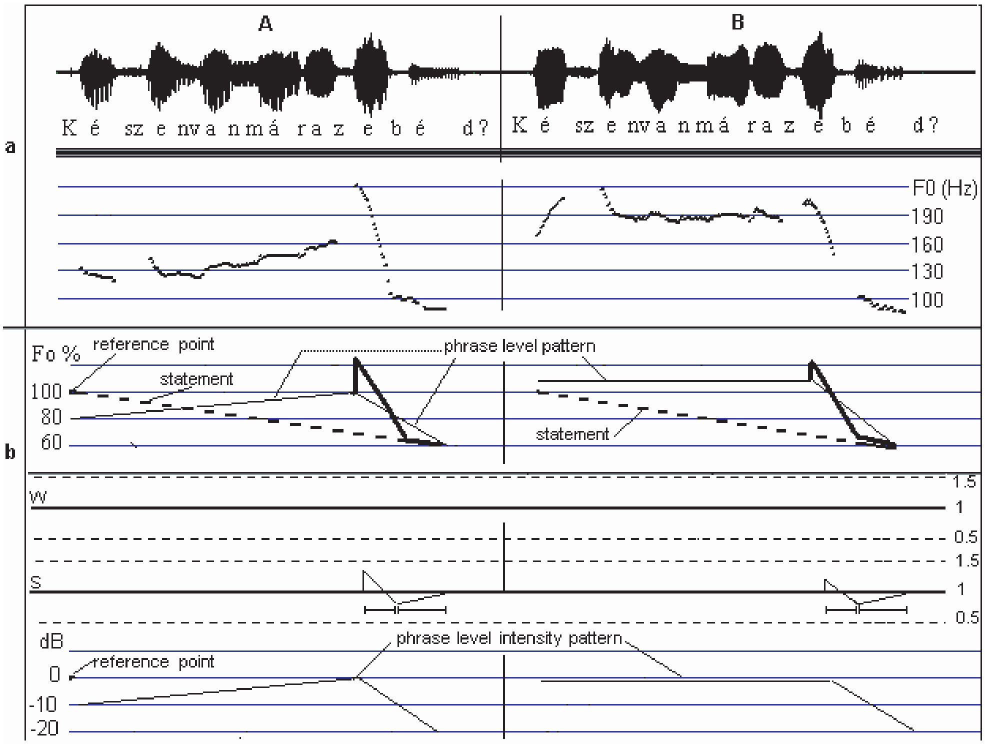

Fig. 6

Two realization forms (a) of a yes/no question and their stylized structures (b). *Készen van már az ebéd*? 'Is lunch ready now?'

It is important to mention that in yes/no questions the place of the peak on the last but one syllable is independent from word structure. Thus the peak can even be realized on an article if the last word of the question has one syllable.

(6) Elvetted **a** sót?
'Did you take **the** salt?'

The intensity curve of the standard yes/no question (Figure 6A) can be characterized by the following general structure: slowly rising until the last but one syllable (the range of the rise is 10 dB), the highest point takes place in that syllable. In the last syllable of the question the intensity falls to the level of −20 dB. In the variant (Figure 6B) the intensity is constantly high until the last but one syllable and the fall is realized from this point until the end of the sentence.

3.2.6. One and two-syllable yes/no questions

One-syllable yes/no questions (*Jó?* 'Good?', *Én?* 'Me?') have basically a rising F_0 contour (Figure 7A). The two-syllable ones (*Elég?* 'Enough?', *Ő volt?* 'Was it she/he?) can be characterized basically by a rise-fall.

If the one or two-syllable yes/no question has a topic-like preceding part, the intonation of the question part will remain the same, the topic will have a slowly falling structure preparing the way for the question part. This slowly falling part will start at 90% and will end at 70–80%. The point where the topic meets the question has the lowest F_0 value in the sentence.

(7) (a) Ennyi már **jó**?
'So many will already be **good**?'

(b) Ennyi már **elég**?
'So many will already be **enough**?'

In both cases, the rise starts definitely lower (60–80%) than a statement does. The end of the highest F_0 value is on 100–120% depending on the situation and emotion. The great distance in F_0 between the start and the highest point forms the question intonation. In one-syllable versions the rise itself is not linear. In the first part of the syllable the F_0 changes slowly, in the second, it changes abruptly. The duration of the vowel is much longer than in sentence internal position. In the case of two-syllable questions (Figure 7B, C) the rise-fall movement is realized mainly in the second syllable. If we want to fit these special cases into the unified description format, we have to define special syllable level modification forms.

3.2.7. Complex yes/no questions

The F_0 and intensity structure of these questions can be concatenated from the stylized patterns discussed earlier (Figure 8, page 292). For example, if the complex yes/no question contains two or more subquestions, the whole F_0 structure will contain two complete questions' phrase elements.

(8) Befejezed a **munkát** és megnézed a **filmet**?
'Will you finish the **work** and watch the **film**?'

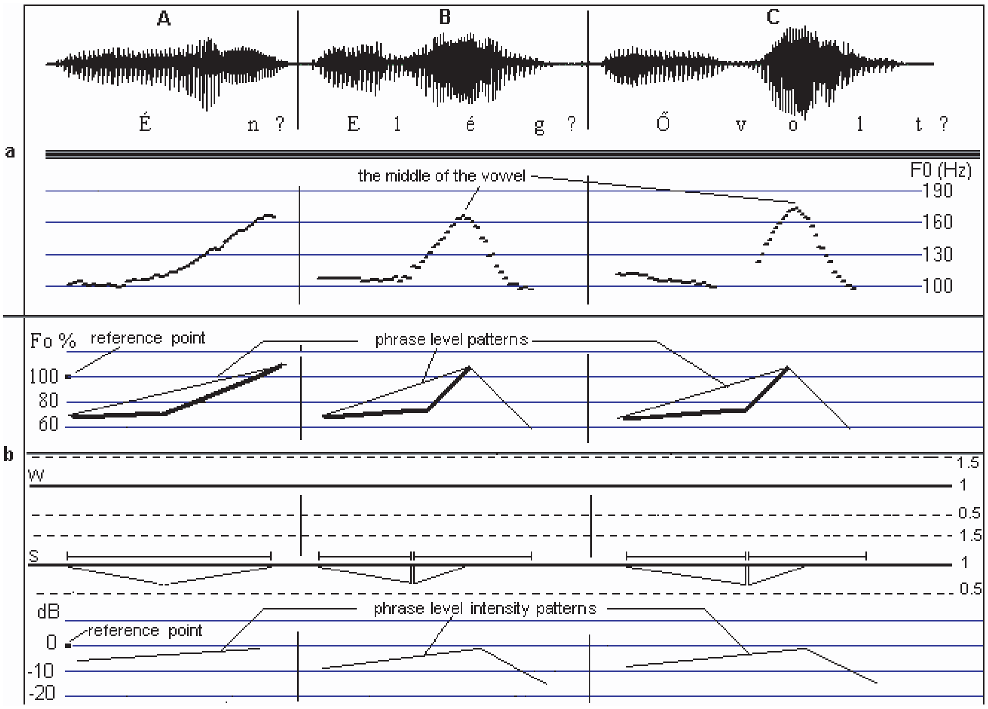

Fig. 7
The F_0 structure of the one and two-syllable yes/no questions (a)
and their stylized representations (b)

As there are two questions in the sentence, the end of the falling parts at the phrase boundaries shows a general falling structure, the lowest F_0 value is at the very end of the question.

In a complex yes/no question like (9), the real question appears in the main clause (*Megnézed*), but the characteristic question pattern with the peak on the last but one syllable is at the very end of the sentence (*besz**él**tél*). In this type of sentences the F_0 structure is similar to that shown in Figure 6B.

(9) Megnézed azt a filmet, amiről a múlt héten **beszéltél**?
‘Will you watch that film about which you **spoke** last week?’

If the first part of the complex yes/no question functions as a topic, it will have a slowly descending F_0 pattern starting from 100% and ending on 80–85% and the question part will have its structure as shown in Figure 6B.

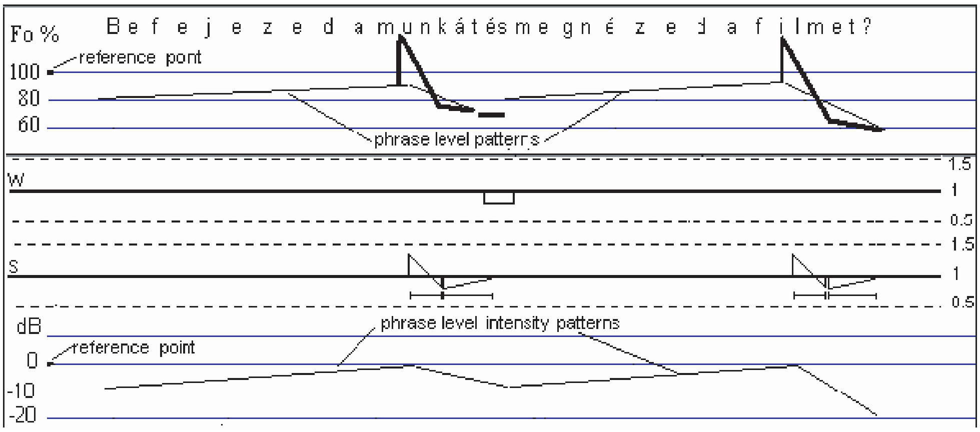

Fig. 8
The stylized F_0 and intensity structure of a complex yes/no question which comprises two questions

(10) Ha megnyernéd a főnyereményt, megvennéd a **házat**?
'If you won the jackpot, would you buy the **house**?'

3.2.8. Alternative questions

Alternative questions consist of two parts which are separated by the word *vagy* 'or'.

(11) (a) Az **első** vagy a **második** lehetőséget választod?
'Do you choose the **first** or the **second** possibility?'

(b) **Enni** akarsz vagy **inni**?
'Do you want to **eat** or to **drink**?'

(c) **Én** vagy **ő**?
'**Me** or **him/her**?'

The two parts can be treated as two phrases. In the first phrase the main F_0 pattern is basically rising (from 90% to 120%), in the second one falling (from 120% to 60%). Syllable level changes define the final, detailed F_0 curve as it is shown in Figure 9A. The rising takes place mainly in the second and third syllables of the first phrase (from 90% to 120%). The F_0 remains on 120% if this phrase has more than three syllables. The fall in the second phrase belongs mainly to the second syllable. Here the F_0 changes from 120% to 60–80%. The place of the endpoint depends on the length of this phrase.

If it has one or two syllables, the endpoint will be on 60%. If it is longer, the fall will be realized in two parts, i.e., from 120% to 80% and from 80% to 60%. The second fall begins in the third syllable and lasts till the end of the sentence independently of the length of this phrase (Figure 10). If the sentence consists of only three syllables, the rise will be shifted to the first syllable, the fall to the last (Figure 9B).

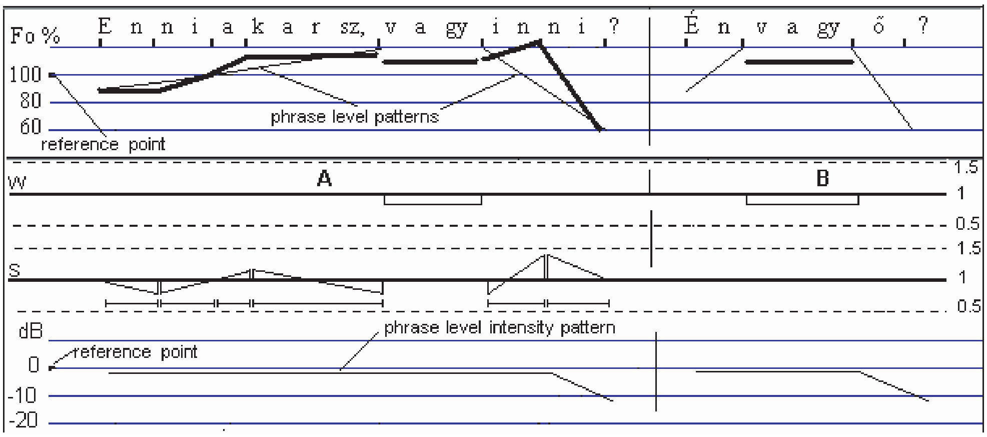

Fig. 9

The stylized melody and intensity structure of alternative questions having different numbers of syllables. Syllables are marked with short thick vertical lines below the text

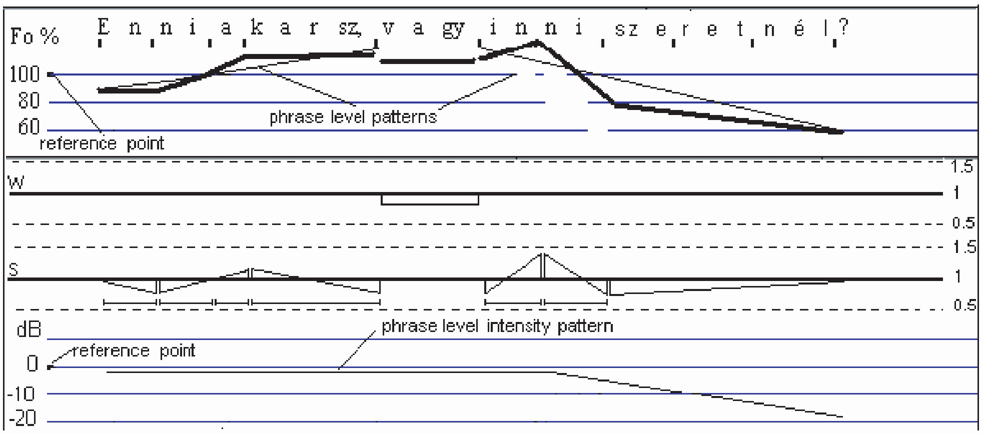

Fig. 10

The stylized melody and intensity structure of the alternative question *Enni akarsz vagy inni szeretnél*? 'Do you want to eat or you would like to drink?'

3.2.9. Elliptic questions

Unfinished questions have basically a rising character (from 80–90% to 120–130%). This pattern is fixed to the last syllables of the last word (Figure 11). If this word has a single syllable, the F_0 change will be realized on this syllable. In the case of two syllables, the rise is divided into two parts: in the first syllable a moderate rise will be produced (from 80–90% to 100%), in the second a sharper one from 100% to 120–130%. In the case of three or more syllables, the rise is divided into three parts along the last three syllables (Figure 11A).

(12) (a) És **ő**?
'And **he**/**she**?'

(b) És **Mari**?
'And **Mary**?'

(c) A **fizetésem**?
'My **salary**?'

Word accents may occur in the part preceding the last word.

(13) És a múlt havi fizetésem?
'And my salary from the preceding month?'

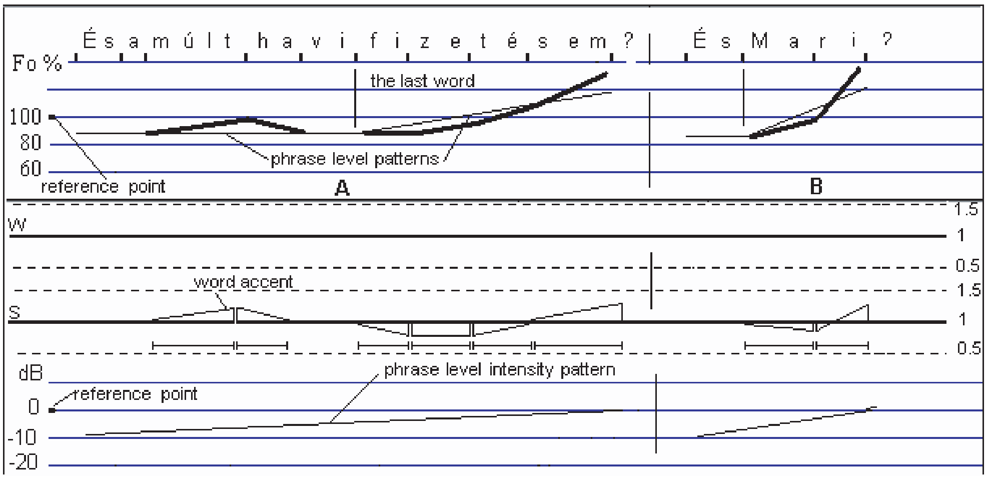

Fig. 11
The stylized melody and intensity structure of elliptic questions. The short thick vertical lines below the text mark syllable boundaries

3.2.10. Control questions

A control question occurs when we want, within a dialogue, to verify the information heard (shown by boldface in the example).

(14) A: Mikor indul a repülő?
‘When does the plane start?’
B: 12 órakor.
‘At 12 o’clock.’
A: **Mikor?**
‘**When?**’

The number of syllables defines the structure of the pitch contour in control questions. In the case of one syllable, the same rising contour is generated as in one-syllable yes/no questions (Figure 7A). In the case of two syllables, the pattern is the same as shown in Figure 7B. If the control question has more than two syllables, the pitch contour will be the same as in simple yes/no questions (Figure 6).

If the control question concerns a whole statement, the F_0 structure may become complicated.

(15) A: Mikor mentél haza?
‘When did you go home?’ (normal question)
B: Azt kérdezted, mikor mentem haza?
‘Did you ask when I went home?’ (control question)

In the example, the first part of the control question (*azt kérdezted*) is realized as a yes/no question. The intonation in the second part may be different depending on the intention of the speaker. If the time is the questioned element (*mikor* ‘when’), the second part will have the F_0 pattern of a yes/no question starting with low F_0 value (Figure 6A). If, however, the place is the questioned element (*haza* ‘home’), the sound sequence *mikor mentem* (‘when did I go’) will have a similar F_0 pattern as it was in the *wh*-questions and the last word, *haza*, will have a rise-fall in the last syllable as shown in Figure 7B for two-syllable yes/no questions.

3.2.11. Morphologically marked questions

Although in most cases it is intonation that differentiates between statements and questions, Hungarian has the possibility to signal a question also with morphemes. The morpheme *-e* attached to the verb means that the sentence is a question, the F_0 pattern of which is similar to that of statements.

(16) **Elkészíted-e** holnapra a cikket?
‘**Will you make** the article for tomorrow?’

The same case occurs when the particle *ugye* introduces the question.

(17) **Ugye** elmész külföldre?
‘You travel abroad, **don’t you**?

In this case two phrase level patterns characterize the question: the first is rising, the second is falling (Figure 12A). The beginning of the rise is around 80%, the end is on 100%. The fall has similar structure as a *wh*-question. If the particle *ugye* closes the question (Figure 12B), the two-syllable control question intonation is manifested in it, the essential part of the question, the first phrase, will have similar structure to that of a *wh*-question, and the second one will be realized as a two-syllable yes/no question.

(18) Elmész külföldre, **ugye**?
‘You travel abroad, **don’t you**?

3.3. Sentences ending with an exclamation mark

3.3.1. Requests

Of the many different forms of requests, we analyzed the one in which the intonation carries the fact of request and the tone of voice expresses a kind request coloured by a slight impatience.

(19) Adja már meg az érkezés időpontját!
‘Would you give me the time of the departure?’

The results of the analysis are the following: the phrase level F_0 pattern is a rise-fall. The starting point of the rise is lower (80%) than in a declarative sentence, the end point is close to 100%. The fall ends at the 70% value (higher than in statements). The final, detailed F_0 curve is formed by syllable level modifications in the first three syllables of the sentence. Word accents do not occur in these requests. The intensity structure of these sentences begins with a lower value (-6 dB) than in a statement. The highest intensity value can be found in the second syllable, the remaining part will have a descending intensity value down to -15–20 dB. The stylized F_0 and intensity patterns are shown in Figure 13 (page 298).

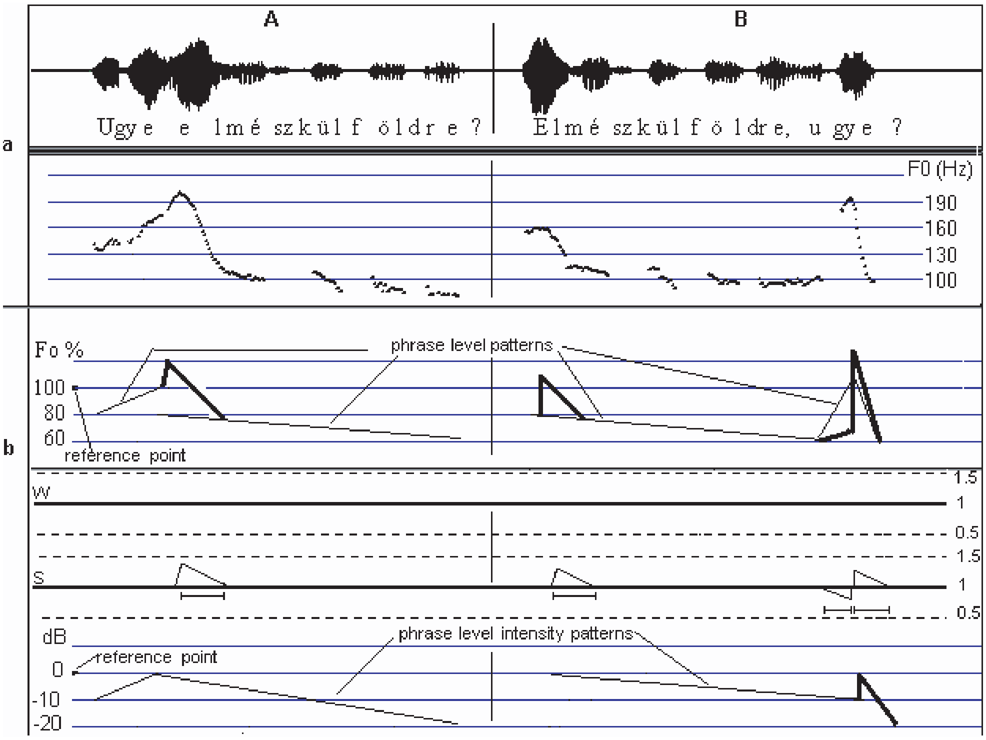

Fig. 12
The F_0 structures (a) and the stylized forms (b) of questions beginning or ending with the word *ugye*

3.3.2. Warnings

Warnings have many representation forms, depending on the situation in which they occur. In the present study we analyzed those warnings in which the listener's attention was drawn to a mistake.

(20) Rosszul csinálod!
 'You do it wrong!'

The phrase level F_0 pattern is falling. Both the beginning and end points are higher than in a statement. A slight modification on this falling pattern is made in the first two syllables. The intensity is generally higher by 5–10 dB in comparison with statements.

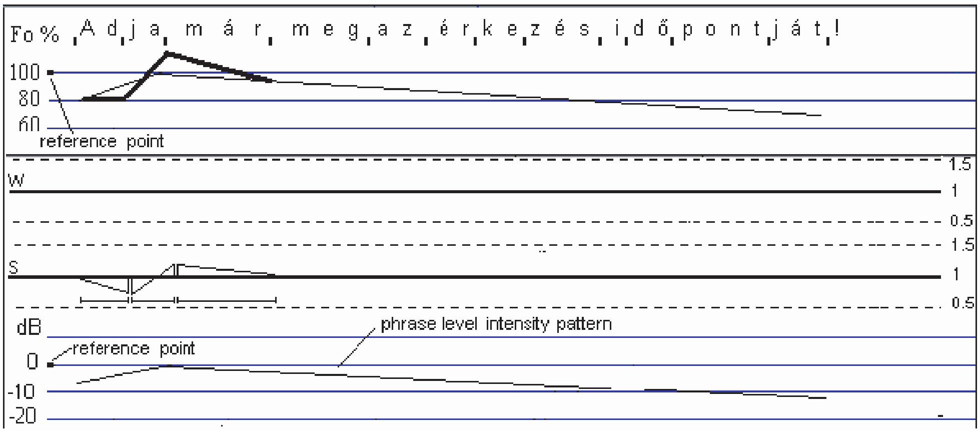

Fig. 13
The stylized F_0 and intensity structure of a request

The stylized F_0 and intensity representation of this type of warning is shown in Figure 14.

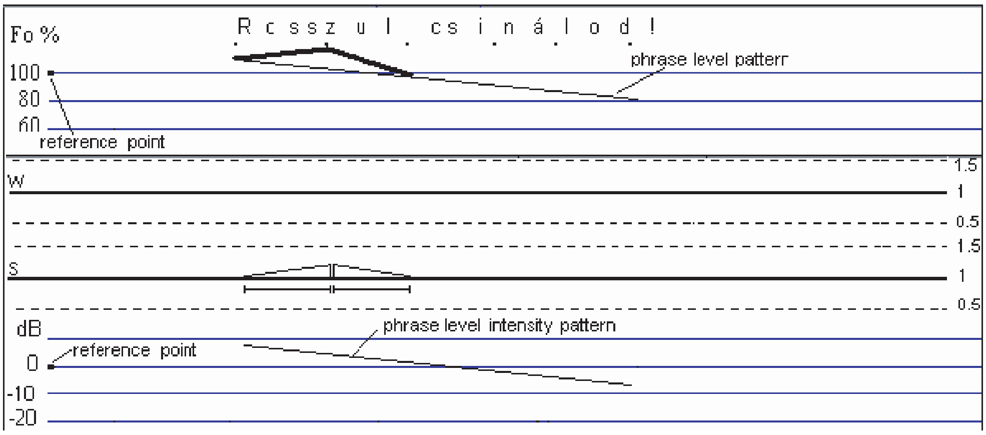

Fig. 14
The stylized F_0 and intensity structure of a warning. The short thick vertical lines under the text mark syllable boundaries

3.3.3. Commands

Various degrees of temperament have been found among the commands analyzed. The increase of temperament was realized mainly by increasing the intensity level and also the value of F_0. The results of the analysis are as follows. The phrase level F_0 pattern is similar to that of a *wh*-question (from 80% to 60%). This pattern is modified in the first syllable as shown in Figure 15. The intensity structure is similar to that in warnings.

(21) Ne menjetek oda!
'Do not go there!'

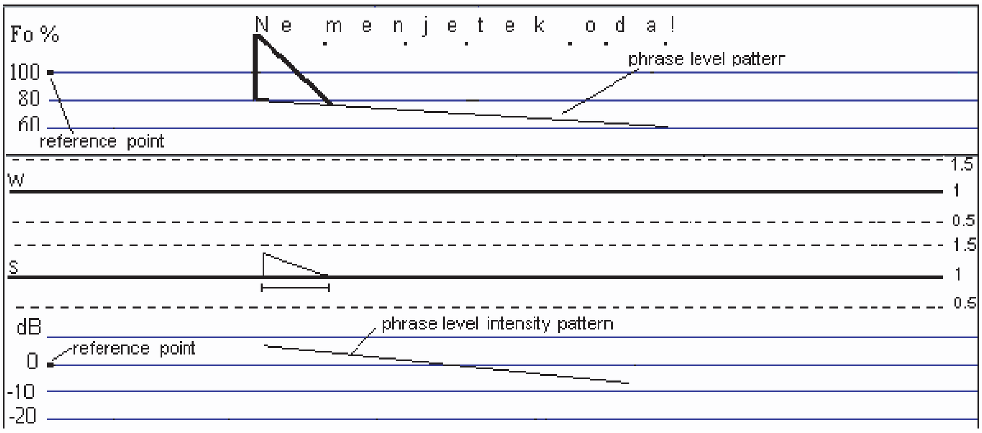

Fig. 15
The stylized F_0 and intensity structure of a command

3.3.4. Sentences expressing desire

Mainly sentences beginning with the interjection *Bárcsak...* 'If only...' have been analyzed.

(22) Bárcsak eljönne a barátom!
'If only my friend would come!'

The phrase level F_0 pattern is falling. The F_0 begins on a slightly lower frequency (90%) than in statements and ends on 80%. The desire is expressed by a syllble level pitch peak (120–130%) in the first syllable. The height of the peak depends on the emotional level of the speaker. The stronger the desire the higher the peak. The stylized representation is shown in Figure 16 (overleaf).

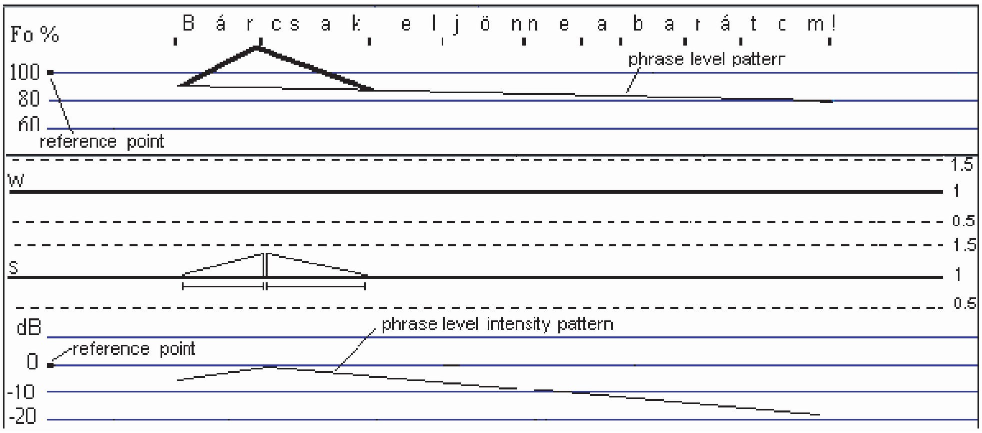

Fig. 16
The stylized F_0 and intensity structure of a sentence expressing desire

4. Verification of the stylized patterns

The F_0 and intensity patterns defined for the most important sentence types in a unified form have been verified in two manners. First the stylized F_0 and intensity patterns have been superimposed on natural sentences by the PDS Prosody Composer tool (Olaszy et al. 2001). This tool enables the researcher—among other things—to change the original F_0 pattern of a natural sentence to a predefined one. Thus the original and the processed sentence differ only in one parameter, F_0 structure. Listening to the processed sentence one can evaluate how the modelled melody sounds in comparison with the original one. This check makes it possible to find the weak points of the modelled patterns and the model can be adjusted more precisely by listening. Such tests and corrections have been carried out by a trained phonetician. After this work, a series of listening tests was organized for general evaluation.

4.1. Listening tests

Two listening test have been carried out. The aim of the first was to compare the natural and synthetically generated F_0 and intensity patterns, in the second one the prosody of generated dialogues was tested.

4.1.1. Test 1

The test material consisted of ten sentence pairs. In each pair, two sentences were put one after the other separated by a pause of three seconds. The first sentence was natural and served as a carrier sentence for the second one. In the second sentence the predefined F_0 pattern (according to the data of the unified F_0 scale) was superimposed on the body of the carrier sentence. Thus, the two sentences in each pair were identical except for the realization of their F_0 structure. Ten such sentence pairs (three *wh*-questions, two yes/no questions, three commands, one request and one statement) were prepared and used in the test. 20 subjects (eight female and twelve male persons, aged from 25 to 55) had to mark in a scale how close the simplified and modelled F_0 pattern was to the natural one. The task was: Compare the melody of the two sentences and evaluate them according to the following scale: they are the same, very similar, similar, less similar, different.

4.1.2. Results

The distribution of the responses is shown in Figure 17 (overleaf). Summarizing the results of the first three columns, 86.5% of the responses found the modelled F_0 structure similar to the original one (or better). This high score allows us to declare that the description of the phrase level F_0 patterns and the word and syllable level local modifications on it represent the structure of Hungarian F_0 patterns at the sentence level tolerably well. The sentences receiving the “less similar” (or worse) evaluation were examined once more concerning the modelled F_0 structure. It became clear that the basis of these negative judgements was not only the slight difference between the natural and modelled F_0 structures: in some cases they were rather due to a slight difference in the general fundamental frequency level of the two sentences (for example, the natural sentence sounded slightly higher than the modelled one, but the form of the F_0 structure was very similar). This latter case was due to the fact that during the whole procedure the reference point was given the same value. In natural speech the general F_0 level may change by 2–8 Hz from sentence to sentence. Thus, in some sentences the modelled F_0 structure sounded slightly different in terms of general F_0 height. Some subjects found this difference enough to give a response “less similar” or “different”.

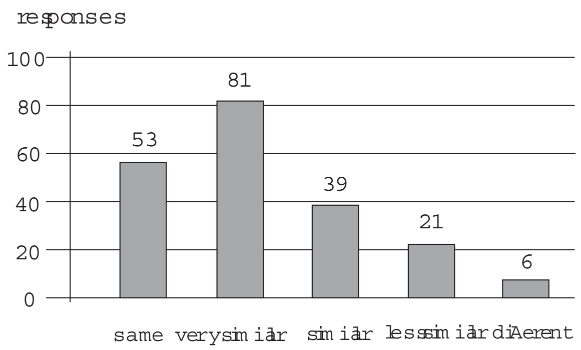

Fig. 17
Evaluation results of the comparison of natural and predefined F_0 structures of Hungarian sentences

4.1.3. Test 2

The goal of this test was to find out whether concatenated unified melody forms—meant to characterize the complex melody of a dialogue—actually give the impression of dialogue. Dialogue elements (two or three sentences concatenated one after the other) have been constructed according to the modelled F_0 structures using natural carrier sentences. Various transformations have also been made concerning their F_0 structure (for example: statement, control question and final statement).

(23) (a) A tervezett tárgyalás után levelet írok a külföldi partnernek.
'After the planned discussion, I will write a letter to the foreign partner'
(basic carrier sentence)

(b) A tervezett tárgyalás után?
(control question, generated from the first part of the carrier sentence)

(c) A tervezett tárgyalás után.
(final strengthening statement, generated from the control question)

In the transformed sentences the time structure of the sound sequences was not changed only the F_0 structure and intensity structure were set according to the previously defined values. The Hz value of the reference point was the same in all sentences. Four dialogues were constructed. The question for the subjects (the same persons as in the first test) was: How do you evaluate the melody pattern of the whole dialogue? They could make a choice from the following scale: very good, good, acceptable, poor. The results are shown in Figure 18.

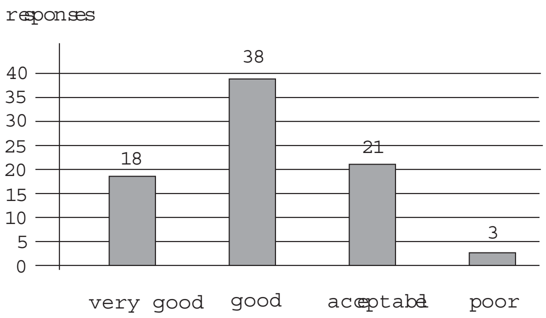

Fig. 18
Evaluation results of the general F_0 structure of the four dialogues

Summarizing the results of the first three columns, 96.25% of the responses found the modelled F_0 structure of the dialogues acceptable or better. This high score shows that the intersentence melody structure defined in the unified F_0 scale gives good F_0 patterns for dialogues as well. Thus the melody pattern of dialogues can be predicted directly from the text.

5. Conclusions

This research concentrated on the systematic description of the intonation and intensity structures of the most frequent Hungarian sentence types (statements, questions, warnings, requests, commands and sentences expressing desire). The description of the melody and the intensity is given in a unified scale in which the beginning point of a statement is fixed as a reference (100% or 0 dB). Thus the patterns building up different sentences can be compared directly with each other and can be transformed from one to the other. The unified scale helps to express the mapping among the melody forms of the sentences. The description of the F_0 and intensity patterns is based on three data structures: the phrase level function (with stylized straight lines), the word level functions and the syllable level modifications (with stylized contours). The word and syllable level functions are expressed by linear changes of multiplication factors in the range 0.5–1.5. The final function is calculated by multiplicating the phrase level function value with the word and syllable level ones. Using this model, the prosody of any text can be predicted without an acoustic analysis if the following information is available: the sentence type, the sentence structure, the phrase boundaries and the accent distribution.

In the prosody of Hungarian, the falling phrase level pattern is characteristic of the majority of sentence types (statements, *wh*-questions, requests, warnings, commands and sentences expressing desire). The beginning and end points of the patterns are sentence type dependent. These differences constitute the basis of the intonation of the given sentence. The rising phrase level pattern is characteristic only in yes/no questions and in control and elliptic ones. The syllable and word level local changes—modulating the phrase contour—have an important role in forming the adequate, final melody pattern of the sentence. The range of pitch movements (taking into account the local changes as well) is between 140% and 60%.

The intensity structure of the analyzed sentences can be summarized as follows. The intensity level is high if the F_0 is high and vice versa. The range of intensity changes was not more than 30 dB.

In some cases, rules could be formulated about the relation of sentence structure and melody. Topic-focus organization has both structural cues and intonational consequences in Hungarian. The intonation of the topic depends on the intonation of the main part. We found that a falling intonation of the main part—as in *wh*-questions and alternative questions—requires a rising pitch contour for the topic part. However, a rising melody contour in the main part—as in yes/no questions—is preceded by a descending one in the topic part. As to the transformation possibilities among different modalities, the realization of the proper intensity contour may be as important as the realization of the proper F_0 curve. This is the case mostly when questions having a rising contour are formed from statements.

Experiments have been carried out to predict and synthesize the prosody of dialogues (using the stylized patterns). The synthesized sentences expressed the internal meaning of the dialogue and the situation quite well.

This study showed that a well-determined F_0 and intensity pattern set can be defined to characterize the prosodic elements of the most important Hungarian sentence types. The pattern set can be used for prosody prediction on the text level. The general results can be used in speech synthesis, speech recognition, language learning programs and in general speech research as well.

References

Collier, René 1990. Multi-lingual intonation synthesis: principles and applications. In: Proceedings of the ESCA Workshop on Speech Synthesis, Autrans, France, 273–6.

Fónagy, Iván – Klára Magdics 1967. A magyar beszéd dallama [The melody of Hungarian speech]. Akadémiai Kiadó, Budapest.

Fujisaki, Hiroya 1992. Modeling the process of fundamental frequency contour generation. In: Y. Tohkura – Vatikoits E. Bateson – Y. Sagisaka (eds) Speech perception, production and linguistc structure, 314–26. IOS Press, Tokyo.

Gósy, Mária 1992. Speech perception. Forum Phoneticum 50. Hector, Frankfurt.

Gósy, Mária 1993. A kiegészítendő kérdés dallamváltozása [Additional melody change in *wh*-questions]. In: Magyar Nyelvőr 117 : 443–57.

Ladd, Robert 1996. Intonational phonology. Cambridge University Press, Cambridge.

Möbius, Bernd 1997. Synthesizing German intonation contours. In: Jan P.H. van Santen – Richard W. Sproat – Joseph P. Olive – Julia Hirschberg (eds) Progress in speech synthesis, 401–15. Springer, Berlin.

Montero, J.M. – J. Gutiérrez-Arriola – J. Colás – J. Macias – E. Enriquez – J.M. Pardo 1999. Development of an emotional speech synthesiser in Spanish. In: Proceedings of the 6th European Conference on Speech Communication and Technology, 2099–102. Budapest.

Olaszy, Gábor 1989. Elektronikus beszédelőállítás [Electronic speech generation]. Műszaki Kiadó, Budapest.

Olaszy, Gábor 2000. The prosody structure of dialogue components in Hungarian. In: International Journal of Speech Technology 3 : 165–76.

Olaszy, Gábor – Ilona Koutny 2001. Intonation of Hungarian questions and their prediction from text. In: Puppel – Demenko (2001, 179–96).

Olaszy, Gábor – Géza Németh – Géza Kiss 2001. Hungarian audiovisual prosody composer and TTS development tool. In: Puppel – Demenko (2001, 167–78).

Puppel, Stanisław – Grażina Demenko (eds) 2001. Prosody 2000. Faculty of Modern Languages and Literature, Adam Mickiewicz University, Poznań.

Silverman, Kim – Mary Beckman – John Pitrelli – Mari Ostendorf – Colin Wightman – Patti Price – Janet Pierrehumbert – Julia Hirschberg 1992. ToBI: a standard for labelling English prosody. In: John J. Ohala – Terrance M. Neary – Bruce L. Darwing – Megan M. Hodge – Grace E. Wiebe (eds) Proceedings of the 1992 International Conference on Spoken Language Processing, 867–70. University of Alberta, Edmonton.

Taylor, Paul 1998. The Tilt Intonation Model. In: Proceedings of the 1998 International Conference on Spoken Language Processing, 1243–7. Sydney.

Taylor, Paul 2000. Analysis and synthesis of intonation using the Tilt Model. In: Journal of the Acoustical Society of America 107 : 1697–714.

Terken, Jacques – René Collier 1990. Designing algorithms for intonation in synthetic speech. In: Proceedings of the ESCA Workshop on Speech Synthesis, Autrans, France, 205–8.

Varga, László 1993. A magyar beszéddallamok fonológiai, szemantikai és szintaktikai vonatkozásai [The phonological, semantic and syntactic aspects of Hungarian speech melodies]. Nyelvtudományi Értekezések 135. Akadémiai Kiadó, Budapest.

Address of the author: Gábor Olaszy
Research Institute for Linguistics
Hungarian Academy of Sciences
Benczúr u. 33.
H–1068 Budapest
olaszy@nytud.hu