Olaszy, G.: The most important prosody patterns of Hungarian. Acta Linguistica Hungarica, Vol. 49 (3-4), 2002. 277-306.

The most important prosody patterns of Hungarian

Gábor Olaszy
Kempelen Farkas Speech Research Laboratory, Hungarian Academy of Sciences, Budapest, Hungary
olaszy@nytud.hu

Abstract

Prosody is a general term for the following features in speech: pitch and intonation, stress, articulation speed, sound intensity and time structure (rhythm and breaks). During verbal communication various prosody forms contribute to the expression of the content of the message (the information built in 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 relative unified scale in which no physical values, but relative distances in pitch values and intensity are used to characterise the melody forms and the intensity levels. This description makes possible the representation of these two prosody elements independently of the personal features (mean Fo value, the range of the Fo of the speaker etc.). The representation makes possible to express the crossfunctions among the melody forms of different expressions. This means that complete prosody patterns can be predicted for any text withouth acoustic analysis .

Introduction

Examination of prosody 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 the information technology. In these applications better and better speech quality is needed in text reading (continuous news reader, e-mail reading, different talking services like book reviews, weather forecast, prose reading etc.), and also in such services where automatic dialogues are realized between the machine and the client. Different models have been constructed in the last decade to describe the inherent structure of intonation for the given language like: for Dutch (Collier, 1990; Terken and Collier, 1990) for German (Möbius 1997); for Japanese (Fujisaki 1992); for English (Silverman et al. 1992; Taylor 1998 and 2000). Also, the research on "emotional synthesis" seems to be increasingly important in constructing life-like verbal situations between human and machine (Rank and Pirker, 1998; Montero et al., 1999). The detailed, phonetic level modeling of the prosody of Hungarian verified by speech synthesis experiments has been done also in the latest years (Olaszy 2000).

Concerning the earlier works for Hungarian, mainly melody patterns have been examined. 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 the intonation from various points of view. Olaszy gave the first results verified by TTS synthesis for statements and for questions. 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 F0 value of the speaker. The first perceptual measurements on the melody forms of statements, questions, commands and exclamations have been done by Gósy (1992). She used special audio material in which only the fundamental frequency of the real speech was present, the higher frequency components were eliminated. These speech stimuli were produced by a special F0 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 played utterance, only the melody form).

The goal of the present research was to define the most important components of Hungarian prosody. Another goal was to construct a generalised manner of description. A relative unified Fo and intensity scale have been defined in which no physical values, but distance values are used to characterise the melody forms and the intensity levels. This form makes possible the representation of the prosody elements independently of the personal features (mean Fo value, the range of the Fo of the speaker etc.). Moreover it is possible to express the crossfunctions among the melody forms of different expression types. This means that complete prosody patterns can be predicted for any text withouth acoustical analysis.

Material and method

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

As to the method of melody- and intensity curve representation a generalised 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 the socalled 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 (for e.g. accent on a prominent word) overlaid or superimposed on global shapes (for e.g. the distinction between a statement and question). In Fujisaki's model a low reference basic Fo 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) Fo value of the speaker is relatively small, about 3% (Möbius 1997). Ladd (1996) faces this model to the so called 'target and transition' models which are based on the ToBi idea (Silvermann et al. 1992). He thinks, the advantage of a phonological (target-transition) model of intonation is that speaker pitch becomes a relatively low-level realization parameter. In a superpositional model, it is much 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 superpositional model was used. The difference is that the sentence structure was taken into consideration when defining the reference level (Olaszy 2000). Another difference is that the same description philosophy is applied to the Fo and the intensity. As the most frequent expression form is the declaration in speech, the refrence level for the Fo calculation was decided to be the beginning pitch and intensity value of the simple structure declarative sentence (the reference value is 1, ie. 100%). By this solution the relative differences among sentence types in the function of declarative sentence show clearer structure then in the earlier methods. We think that the beginning part of the sentence has the main roll - as to the general shape- in speaking. The modality of the sentence in Hungarian can be predicted already from the beginning part of the sentence. Therefore emphasis was made to define the beginning Fo points of all sentence types in the function of the declaration. The sentence ending parts have been defined in the function of the ending of declaration. Another advantage of this method is that the rules for transformation the Fo patterns from one modality to another one (i.e.to generate a question, or a request, or a command from a statement) show also clearer structure, because the reference value is joined 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 Fo value, the range of the Fo of the speaker etc.). Only the value of the reference level have to be changed and the Fo and intensity pattens move to a higher or lower range. Applying the generalised stylized patterns complete Hungarian prosody patterns (for longer texts, dialogues etc.) can be predicted, if the personal dependent reference Fo is given in Hz.

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

The slope of falling and rising depends on one hand on the duration of the time interval where the contour is present (applied), and on the other hand on the minimal and maximal frequency value 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.

The unified melody and intensity representation

Both the Fo and the intensity structure of the analysed sententence types will be described in the unified scale with stylized lines. For the Fo 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) line as multiplication factors to modulate the main contour (similarly to Fujusaki’s representation). The value of the multiplication factors may vary betwee 0.5 and 1.5. The modulation is calculated from the Fo values of the main pitch contour. Thus the range reduction is realised as well. The description of the intensity structure follows the same philosophy. The stylised contours on the figures below show the main, prase 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 Fo contour.

Statements

The phrase level, main melody form for statements (Figure 1.) is the continuously falling pattern. If the sentence is short the original Fo pattern and the stylized one is closely the same (Figure 1). If the sentence is longer, word and syllable level pitch changes may modulate the falling melody form towards positive or negative directions. The pitch curve of the sample sentence on Figure 2 shows that unaccented words (articles) have lower Fo value as their surroundings, the accented syllables (marked with dots) have pitch peaks. Microintonation also modulates the pitch curve on sound level (marked with up-down arrows on Figure 2), but these segmental level changes are not involved into the present description.

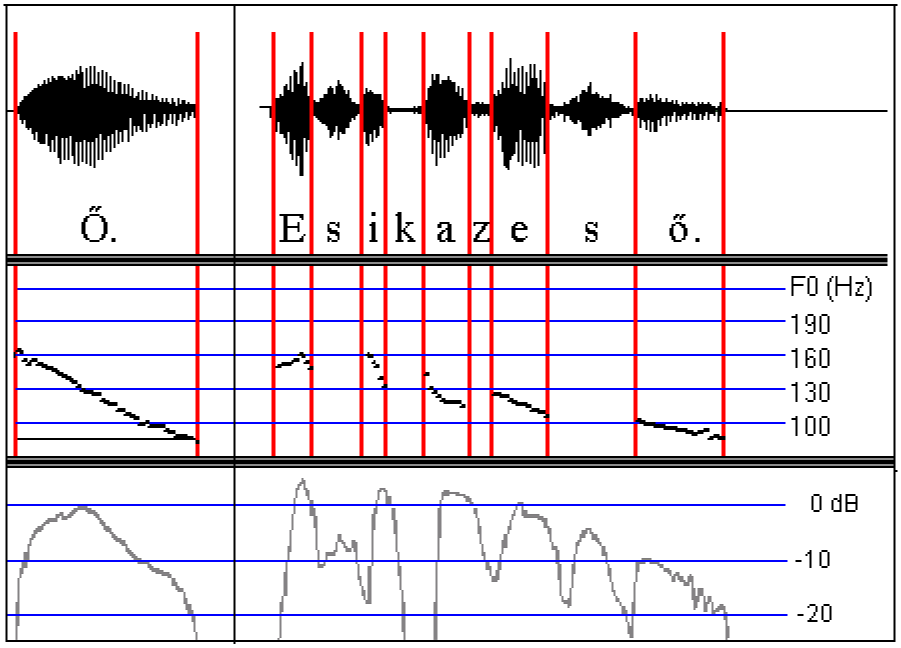

Figure 1.
The declination in pitch and intensity for two simple structure short statements (*Ő.* 'He/She.'; *Esik az eső.* 'It is raining.'). The vertical lines show sound boundaries

In statements the declination in pitch was found to range between 30%-42%. The deepest Fo endpoint was realized if the sentence was pronounced alone, or if it was a very 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).

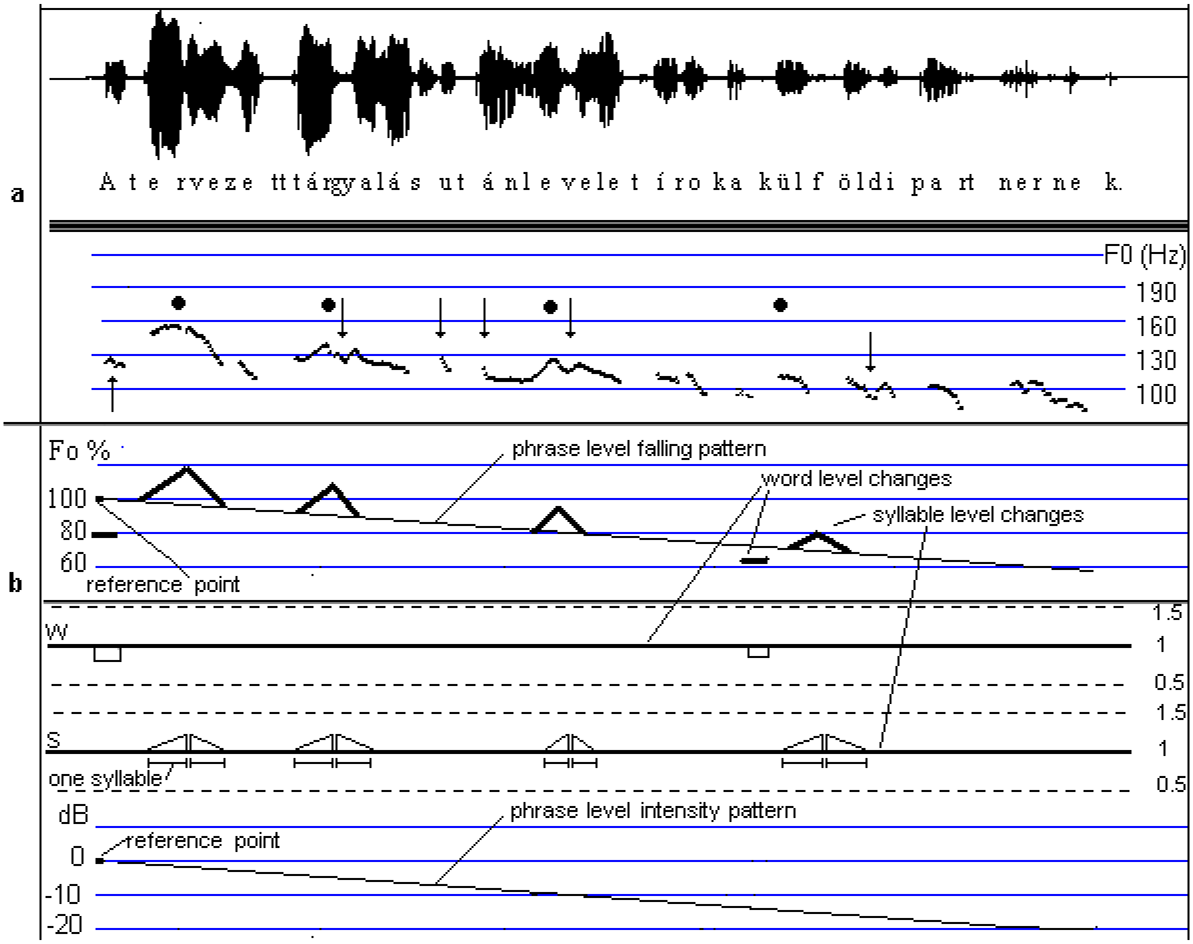

Figure 2.
The Fo structure (a) and the stylized representation (b) of a statement.
(*A tervezett tárgyalás után levelet írok a külföldi partnernek.* 'After the planned discussion I will write a letter to the foreign partner.')

The lower Fo value in unaccented words depends on the place of them within the sentence. The greatest difference compared to the main pattern can be measured if the sentence begins with such element (see on Figure 2. marked with down-up arrow). In this case the negative modification may reach the factor 0.8.

In case of complex statements more intonation phrases (falling patterns) build up the whole sentence. The beginning Fo value is at the reference point, the sentence final one is on the same value as it was in simple statements. The intermediate falling patterns show a tooth wave like structure which itself also has a sligh declination. A sample example is shown on Figure 3. Commas separate the sentence into three main falling patterns. The comma effect is expressed both on word and on syllable level, i.e. the equalisation of the falling Fo into a level one is expressed by the word level modification, the final rise may be expressed with the syllable level one. The result in the word before the comma will be similar what can be seen in the natural Fo pattern. In the second falling pattern two word-accents and one comma effect forms additionally the main Fo contour. The third, final falling pattern ends on 58% and contains two syllable level changes. The negative 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 word (*azt; Péter; levelezik; három; külföldön*).

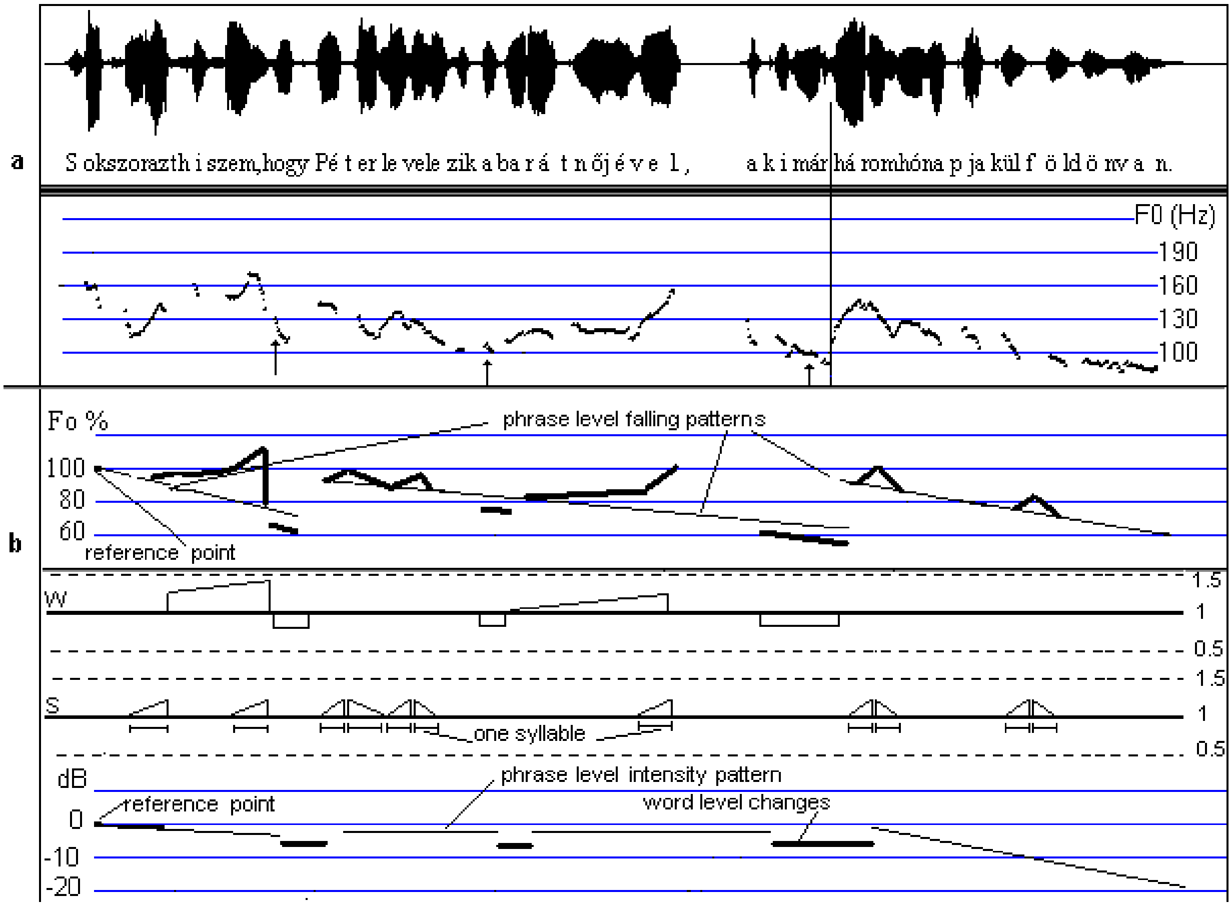

Figure 3.
The Fo structure (a) of the complex statement and its stylized (b) representation.
Sokszor azt hiszem, hogy Péter levelezik a barátnőjével, aki már három hónapja külföldön van.
'Usually I think that Peter is corresponding with his girl friend, who has been staying abroad already for three months.'

As to the intensity structure of complex declarative sentences (Figure 3 down.) the negative accented parts (marked with arrow) have lower intensity than their surrounding. The intensity level is close to 0 dB in the first two phrases, a declination to –20 dB is present only in the last subsentence.

Summing up the main Fo 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 Fo structure (each). The effect of comma represents a syllable level change into a rise or level form in the final part of the word before the comma. In very long sentences the slope of the declination in one intonational phrase can be so small, that practically the Fo structure shows a level form. The unaccented elements have mostly deaper Fo value than their surrounding the accented syllables have higher one.

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

Questions

The melody patterns in Hungarian interrogative sentences vary in a great measure, depending on different 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 as in statements, and in certain questions the sound durations are strongly lengthened.

Wh-questions beginning with Q word

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

Mikor *indultok?* (**When** will you start?) The main Fo structure for the Wh question is a falling pattern, which starts from a lower value (about 80%) and ends on a similar point as it was in the statement, i.e the slope of the falling pattern is smaller in these questions as in statements. This form is realised independently of the length of the question.

Kivel *fogtok most találkozni?* (**With whom** will you meet now?)

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

A syllable level Fo modification in the Q-word realises 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 will reduced in the second one (Figure 4). The right proportion among the peak and the slope of the falling pattern determines the proper intonation of the whole question. The higher is the peak value and the smaller is the slope of the main falling pattern (i.e the deeper is the starting point of the main falling pattern) the more characteristic is the question. Other syllable level modifications (word accents) do not appear in the descending part.

There exists another variant for the pronunciation of these questions (Gósy 1994). The difference between the standard one (described above) and the variant is in the pronunciation of the final part, i.e. people raise the F_0 in the last syllable. This rise is about 10% comparing to the Fo value of the last but one syllable. Another difference is that the main F_0 pattern is not falling but it shows rather level characteristic. It begins with a slightly lower Fo frequency as in 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 F0 generation decides the ending form of the question already after the pronunciation of the question word. If the decision is: low ending (standard version), the 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 rise at the end.

The intensity structure of the Wh question shows very similar structure as it was in the statement.

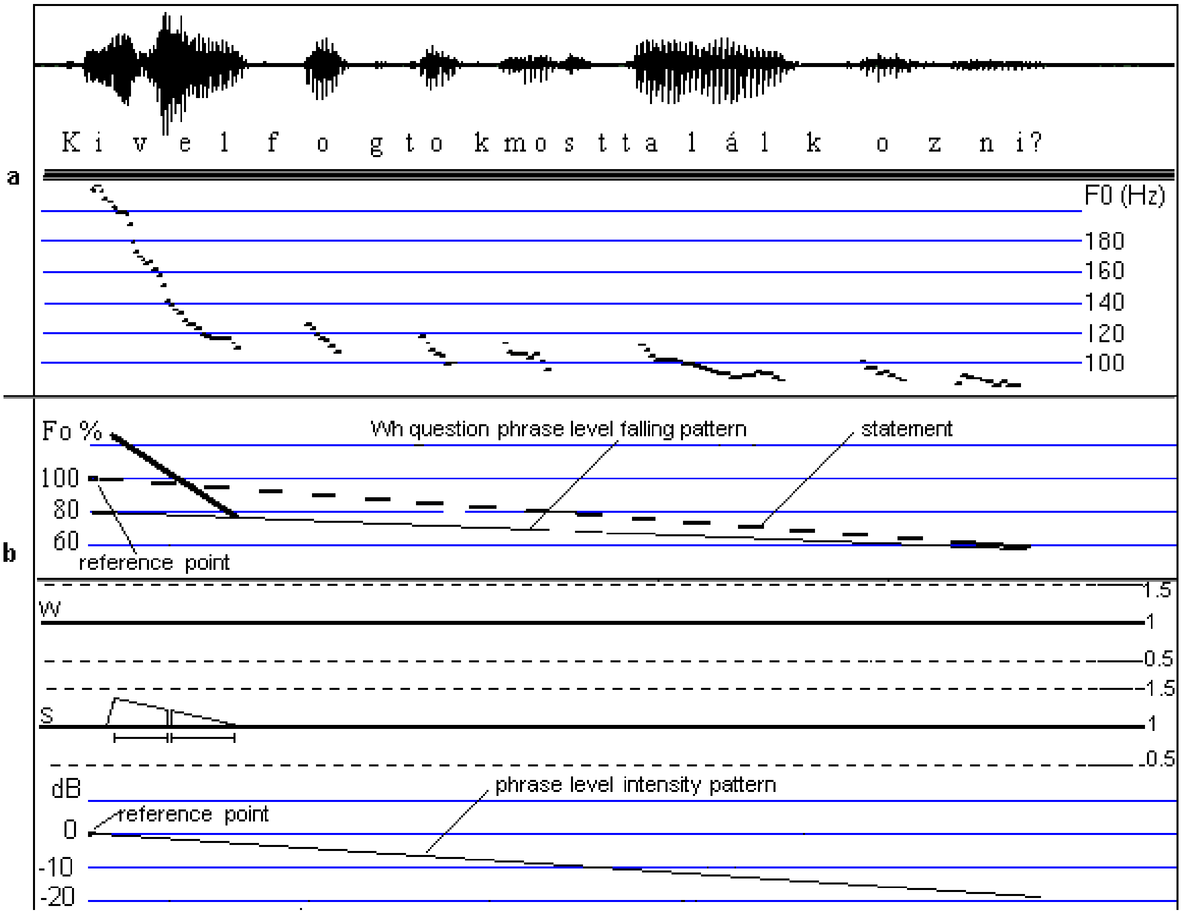

Figure 4.
The typical Fo pattern of a Wh-question longer than three syllables (a) and its stylized (b) representation.

Wh-question with topic

If the Wh-question has a topic part before the question the melody structure can be represented by two phrase level patterns. The topic has a slightly rising form, the question part has the same as described for simple Wh- questions. The topic part before the question begins with lower Fo value (about 80%) and has a slowly rising (to 85-90%) characteristic which prepares the question (Figure 5.).

as described for simple Wh- questions. The topic part before the question begins with lower Fo value (about 80%) and has a slowly rising (to 85-90%) characteristic which prepares the question (Figure 5.).

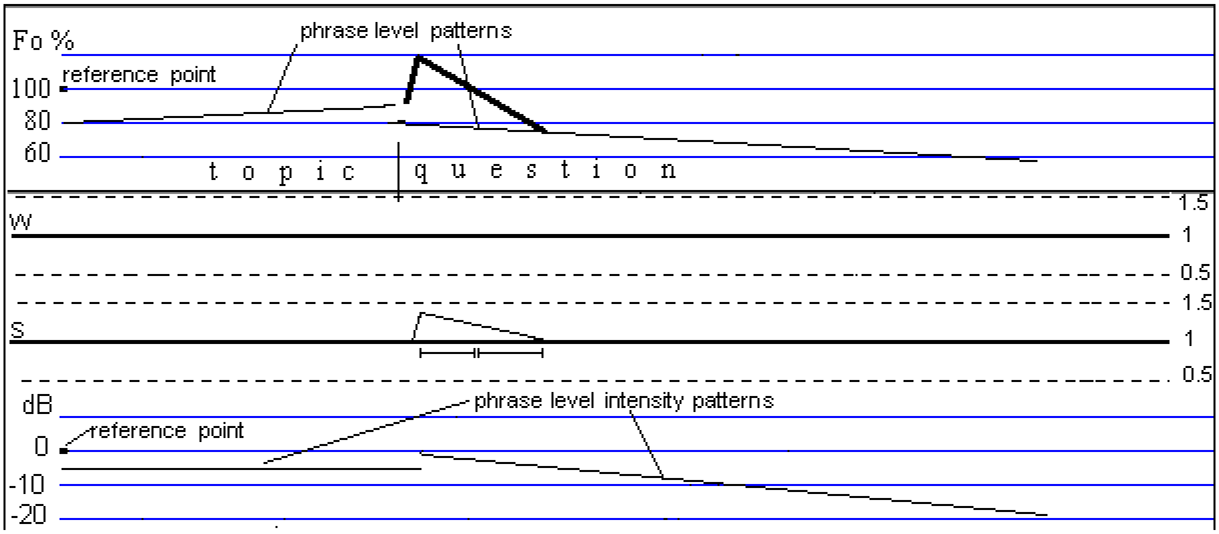

Figure 5.
The stylized patterns of a Wh-question with topic
Ezt a témát illetőleg ***mikor*** *válaszoltok a kérdéseimre?* (’Concerning this topic when will you answer my questions?’)

Complex Wh-questions

In complex forms the Wh question is followed with another subordinated close.
Example:
Mikor *mész az üzletbe és veszed meg a kávét?* (**When** will you go to the shop and buy the coffee?)
In this cases the question part has similar characteristic as in the simple Wh question, but the descending part will end higher. This higher ending can be explained with the fact that the sentence has not been finished at this point, it will be continued. In the subordinated sentence part the descending Fo change is continued until the very end of the complex sentence. The very final Fo value is close to that of in simple statements (60%). Word accents may occur in the subordinated close.
If the complex Wh question contains more than one question, one falling pattern is present over

Complex Wh-questions

In complex forms the Wh question is followed with another subordinated close.
Example:
Mikor *mész az üzletbe és veszed meg a kávét?* (**When** will you go to the shop and buy the coffee?)
In this cases the question part has similar characteristic as in the simple Wh question, but the descending part will end higher. This higher ending can be explained with the fact that the sentence has not been finished at this point, it will be continued. In the subordinated sentence part the descending Fo change is continued until the very end of the complex sentence. The very final Fo value is close to that of in simple statements (60%). Word accents may occur in the subordinated close.

If the complex Wh question contains more than one question, one falling pattern is present over the complex question and the syllable level peaks in the Q-words will have consequently lower and lower Fo value along the sentenc, realising the range reduction.
Example: ***Mikor*** *fejezed be a munkát és* ***mikor*** *jössz haza?* '**When** will you finish the work and **when** will you come home?'

Yes/No questions and their environment
The main intonation pattern of Yes/No questions can be a rise-fall or a level-fall form. If rise-fall is realised, the starting point is lower (80%) as in statements and the end of the rising part is about 100%. This rising structure prepares the Fo peak of the questioning part, which is placed on 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 realised by the sharp pitch fall in the last two syllables. The question intonation may be increased by a syllable level jump-fall pattern in the last but one syllable. The jump is realised 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 the 120-130%.
Word accents are not present in the slowly rising part. This can be explained by the structure of this question. The first part only prepares the peak at the end

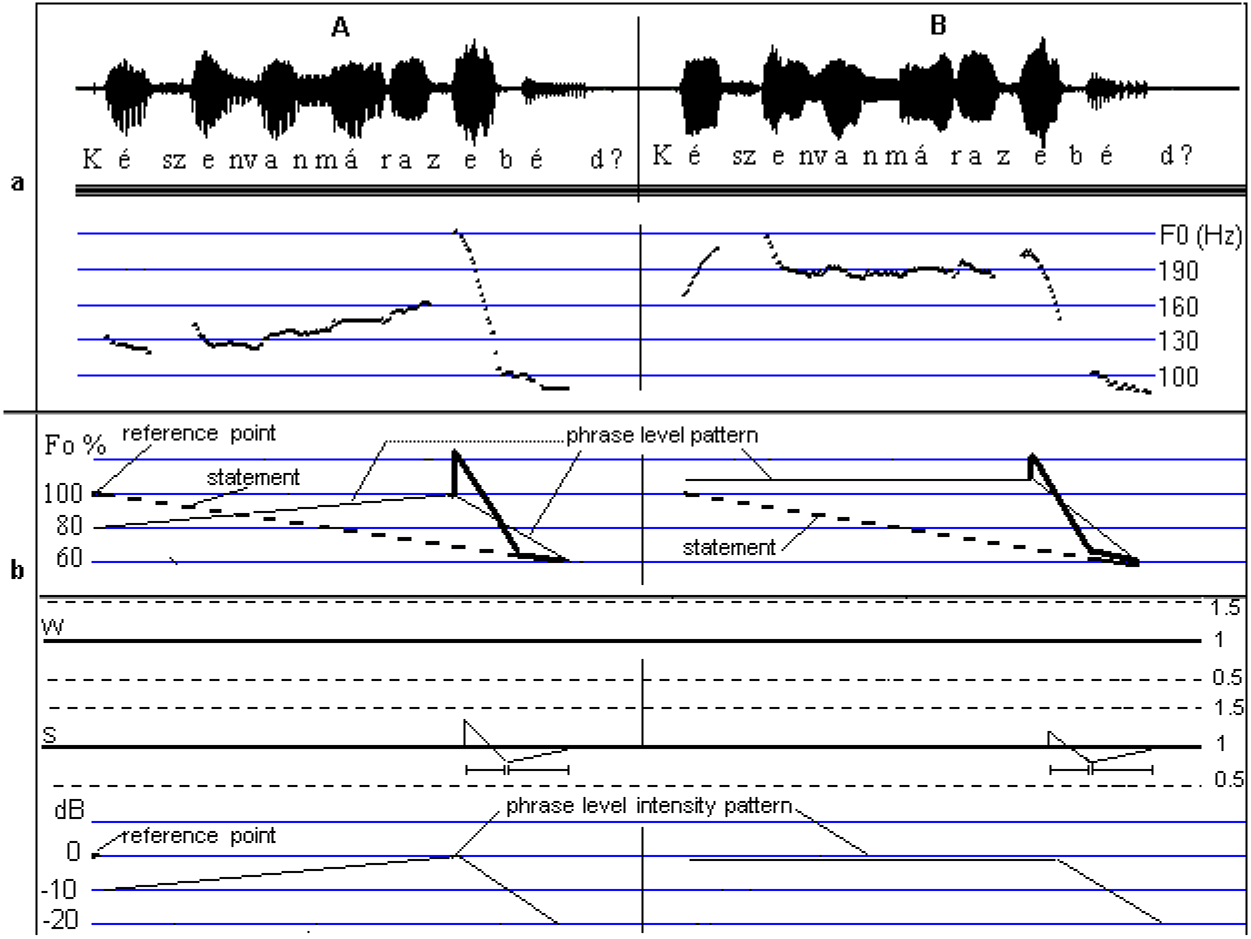

Figure 6.
Two realization forms (a) of the yes-no question and their stylized structures (b).
Készen van már az ebéd? 'Is the lunch already ready?'

which intends to express the main information. The second form of this question with the level-fall intonation is pronounced in the case of expressing unpatiency, or anger.

Yes/no question with topic or focus
In this case the sentence is devided into two phrase level Fo patterns. The sentence begins with a slightly falling structure (from 100% to 80%) until the end of the topic or the word being in the focus position. This is followed with the second pattern which is similar to that of shown on Figure 6 (b).
Examples: *Holnap délután **elmentek vérge moziba**?* Tomorrow afternoon **will you go finally to the cinema**?). (The question part is marked with fat letters.)
*A tegnap kiadott **szakácskönyvet** vetted meg a barátnődnek?* 'Did you buy the yesterday issued **cook book** for your girlfriend?'). (The word in focus is marked with fat letters.)
It is important to mention that in Yes/No questions the place of the peak on the last but one syllable is independent from the word structure. Thus the peak can be realised on an article as well (*Elvetted **a** sót?* 'Did you take **the** salt?) if the last word of the question has one syllable.
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 is falling to the level of –20 dB. In the variant (Figure 6b) the intensity is constantly high until the last but one syllable, the final part of the question ant the fall is realised from this point until the end of the sentence.

One and two syllable Yes/No questions
The one syllable Yes/No questions (*Jó?* 'Good?') have basically a rising Fo contour (Figure 7a).
The two syllable ones (*Elég?* 'Enough?' Ő volt? 'Was it she/he?) can be characterised basically with 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 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 Fo value in the sentence.
Example: *Ennyi már **jó**?* (So many will already be **good**?) ; *Ennyi már **elég**?* (So many will already be **enough**?)
In both case the rise starts definitely lower (60-80%) than the statement. The end of the highest Fo value is on 100-120% depending on the situation and emotion . The great distance in Fo between the start and the highest points forms the questioning intonation. By one syllable versions the rise itself is not linear. In the first part of the syllable the Fo is changing slowly, in the second one sharply. The duration of the vowel is much longer, than for example in sentence internal position. In the case of two syllable questions (Figure 7b, c) the rise-fall movement is realised mainly in the second syllable. If we want to fit these special cases into the unified description form we have to define special syllable level mondification forms.

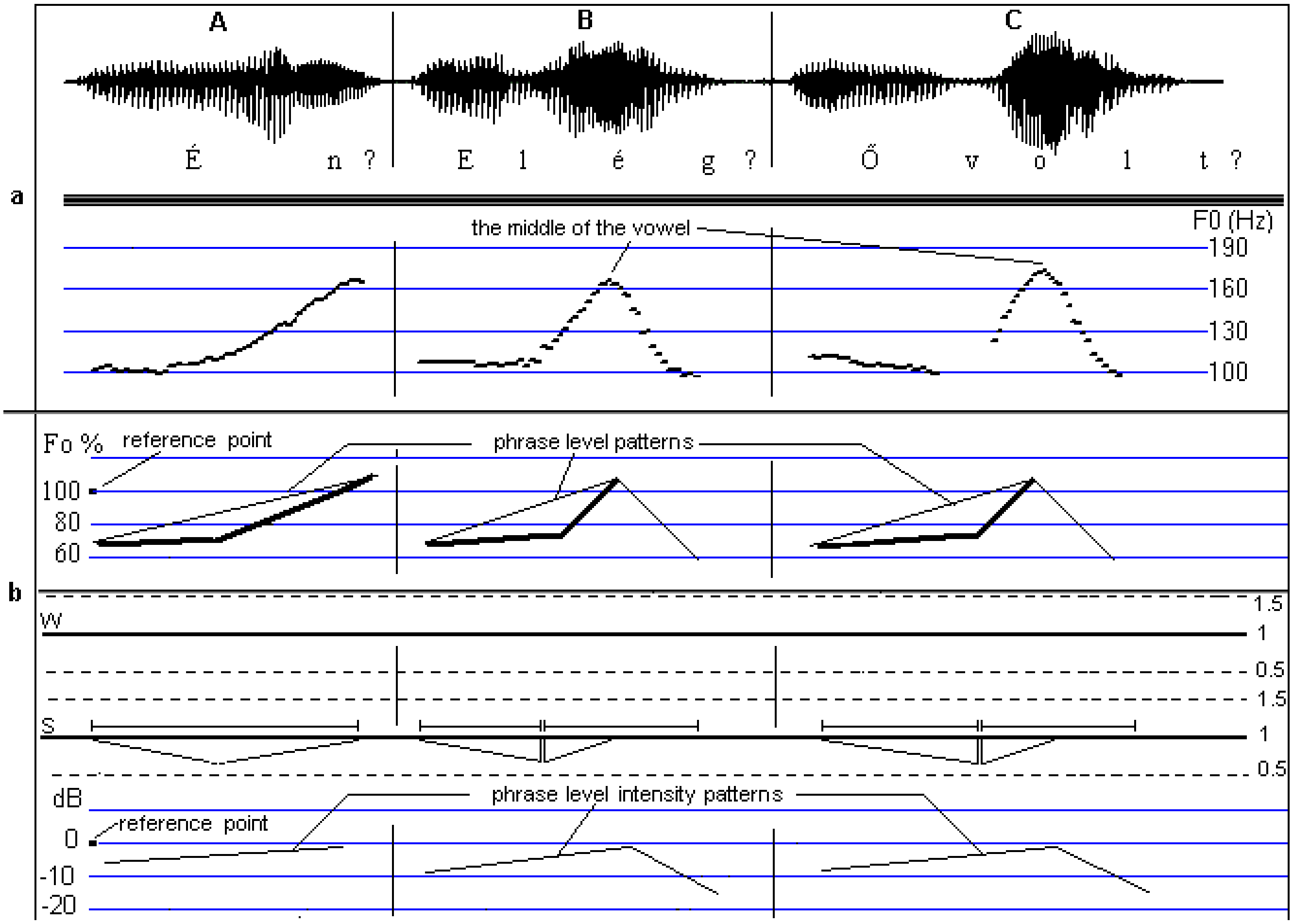

Figure 7
The Fo structure of the one and two syllable Yes/No questions (a) and their stylized representations

Complex Yes/No questions
The Fo and intensity structure of these questions can be concatenated from the earlier discussed stylized patterns (Figure 8). For example if the complex Yes/No question contains two or more subquestions, the whole Fo structure will contain two complete questions phrase elements.
Example: *Befejezed a **munkát** és megnézed a **filmet**?* 'Will you finish the **work** and watch the **film**?'.

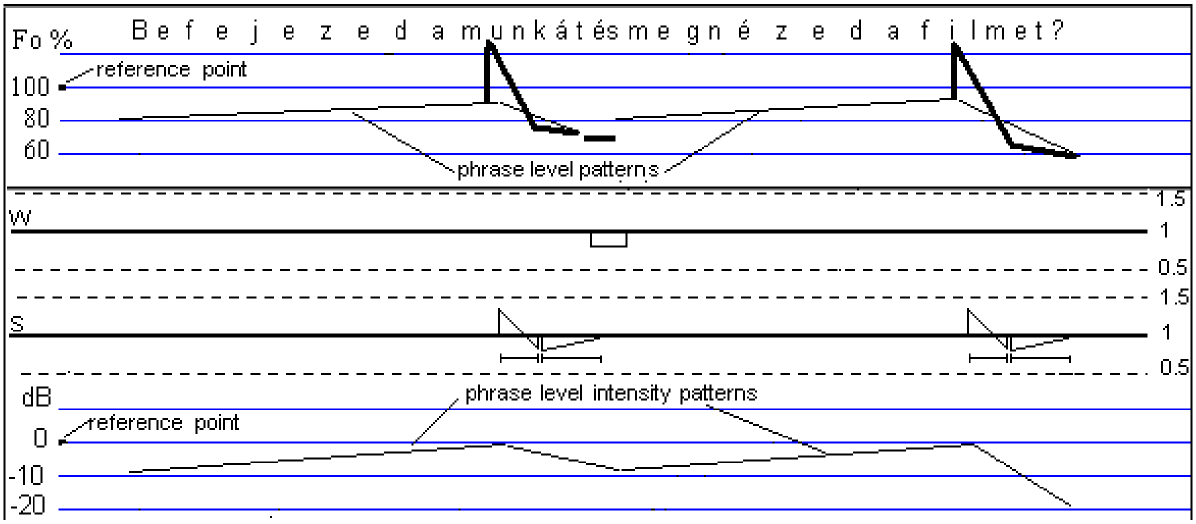

Figure 8.
The stylized Fo and intensity structure of the complex Yes/No question which has two questions

As there are two questions in the sentence the end of the falling parts in the phrase boundaries shows a general falling structure.This phenomena can be explained by the general fact that in Yes/No questions the deepest Fo value is at the very end of the question.

In a subordinated complex Yes/No question like: *Megnézed azt a filmet, amiről a múlt héten **beszéltél**?* 'Will you watch that film about which you **spoke** last week?' the real question appears in the main clause (*Megézed*), but the characteristic question pattern with the peak on the last but one syllable is at the very end of the sentence (*beszé**l**tél*). In this type of sentences the Fo structure is similar to that of shown on Figure 6b. If the first part of the complex Yes/No question functions as a topic, it will have a slowly descending Fo pattern starting from 100% and ending on 80-85% and the question part will have its structure as shown in Figure 6b.

Example: *Ha megnyernéd a főnyereményt, megvennéd a **házat**?* 'If you won the money, would you buy the **house**?'

Alternative questions

The alternative questions consist of two parts which are separated by the word "*vagy*" (or).

Examples:

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

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

Én vagy ő? ('I or she/he?')

The two parts can be treated as two phrases. In the first phrase the main Fo pattern is basically rising (from 90 % to 120%), in the second one falling (from 120% to 60%). Syllable level changes define the final, detailed Fo curve as it is shown on Figure 9a. The rising takes place mainly in the second and third syllable (from 90% to 120%) of the first phrase. The Fo 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 Fo 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 realised 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 lentgth of this phrase. (Figure 10). If the sentence is built up only from three syllables the rise will be shifted to the first syllable, the fall to the last (Figure 9b).

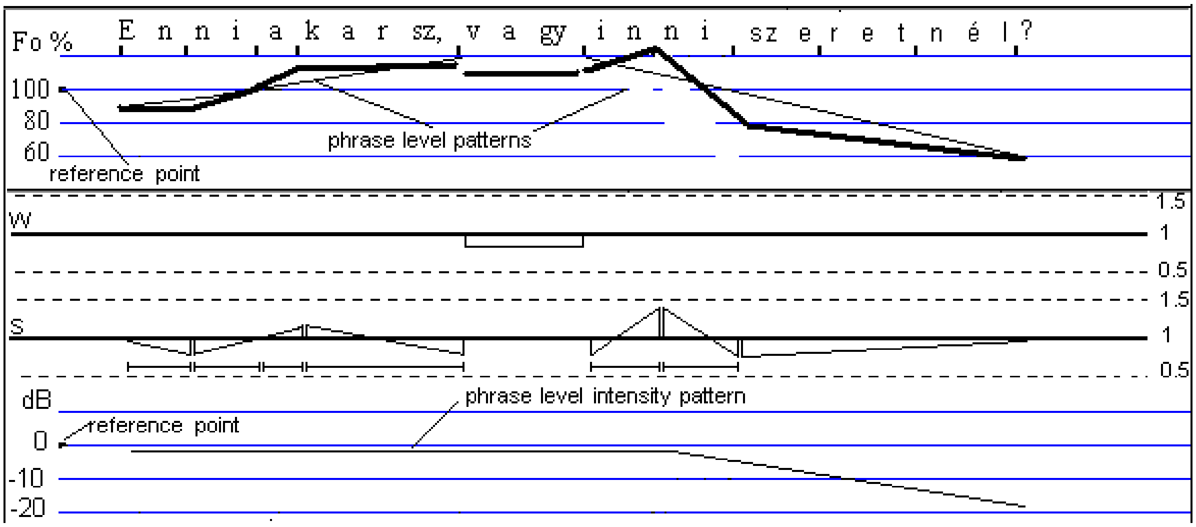

Figure 9.
The stylized melody and intensity structure of alternative questions having different numbers of syllables. Syllables are marked with short thick vertical lines below the text

Elliptic question (not finished)
Not finished questions have basically a rising characteristic (from 80-90% to 120-130%). This pattern is fixed to the last syllables of the last word. If this word has one syllable, the Fo change will be realised in 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 one a sharper from 100% to120-130%. In the case of three or more syllables the rise is divided into three parts along the last three syllables.
Examples:
És Ő? ('And he/she?')
És Mari? (And Mary?)
A fizetésem? (My salary?)
Word accents may occur in the preceding part of the last word.
Example: *És a múlt havi fizetésem?* ('And my salary from the preceding month?')

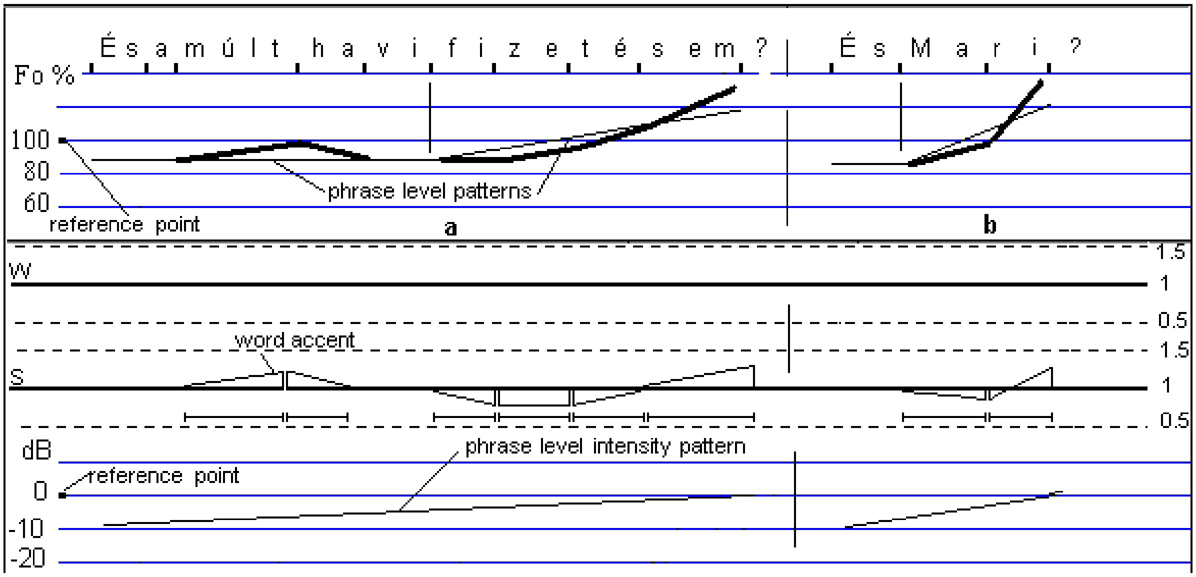

Figure 11.
The stylized melody and intensity structure of the alternative questions.
The short thick vertical lines blow the text mark syllable boundaries

Control questions

Control question occurs when we want within a dialogue to verify the information heard (shown by fat letters in the example).

Example: -*Mikor indul a repülő.* 'When does the plane start?'
-*12 órakor***.** 'At 12 o'clock'
- Mikor? **'When?** '

The number of syllables defines the structure of the pitch contour in the re-questions. In the case of one syllable the same rising contour is generated as in the 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 of it will be the same as in simple Yes/No questions (Figure 6).

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

Example:

- *Mikor mentél haza?* (normal question) 'When did you go home? '
- *Azt kérdezted, mikor mentem haza?* (the control question) 'Did you ask, when did I go home? '

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 (*mikor* 'when') element the second part will have the Fo pattern of a Yes/No question starting with low Fo value (Figure 15.a). If the place is the questioned (*haza* 'home') element the sound sequence *mikor mentem* ('when did I go') will have a similar Fo pattern as it was in the Wh-questions and the last word *haza*, will have a rise-fall in the last syllable as shwn on Figure 7b. for two syllable Yes/No questions.

Morphologically marked questions

Although in most cases the intonation differentiates between statements and questions, Hungarian has the possibility to express the question also with morphemes. The *–e* morpheme put after the verb means a question, the Fo pattern of which is similar to that of the statements.

Example: *Elkészíted-e holnapra a cikket?* 'Will you make the article for tomorrow?'
The same case occurs when the particle *ugye* introduces the question.
Example: ***Ugye*** *elmész külföldre?* 'Do you travel abroad, **don't you**?
In this case two phrase level patterns characterise the question: the first is rising, the second is falling (Figure 12.a). The beginning of the rise around 80%, the end is on 100%. The fall has similar structure as the 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 questio the first phrase will have similar structure to the Wh question, the second one will be realised as a two syllable Yes/No question.
Example: *Elmész külföldre* ***ugye****?* 'Do you travel abroad, **don't you**?

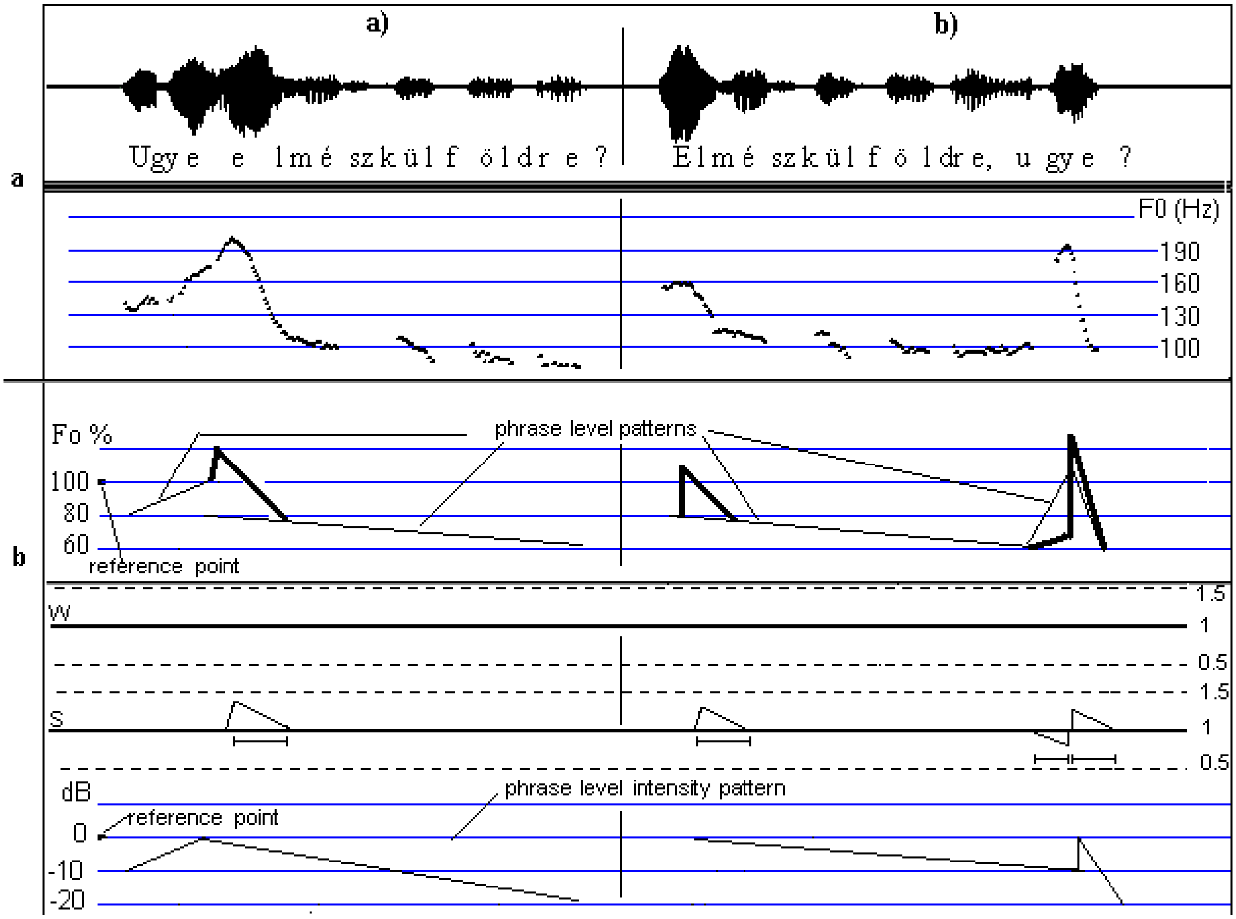

Figure 12.
The stylized structure of questions beginning and ending with the word *ugye.*

Sentences ending with exclamation mark
Request
From different forms of requests we analised the one in which the intonation carries the fact of request and the voice expresses a kind request together with a slight unpatience.
Example: *Adja már meg az érkezés időpontját!* 'Would you be so kind to give the time of the departure!'
The analysis results are the following: the phrase level Fo pattern is a rise-fall. The starting point of the rise is lower (80%) as by the declarative sentence, the end point is close to 100%. The fall

ends at the 70% value (higher than in statements). The final, detailed Fo 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 until –15-20 dB. The stylized Fo and intensity patterns are shown on Figure 13.

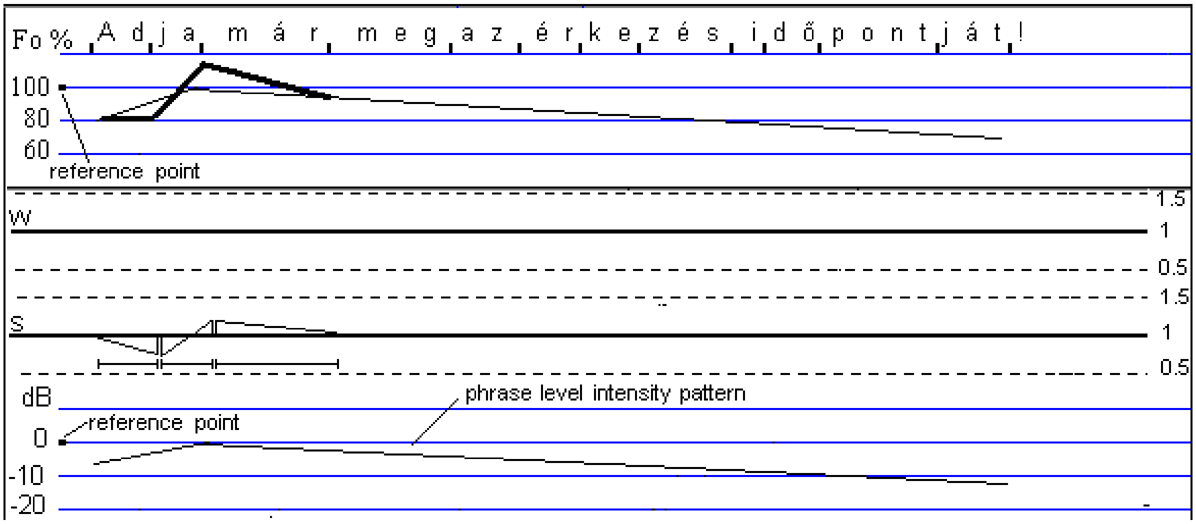

Figure 13.
The stylized Fo and intensity structure of the request

Warnings

Warnings have many representation forms, depending on the situation in which they occur. In the present study we analised those warnings in which the attention was drown to a mistake.

Example: *Rosszul csinálod!* 'You do it wrong!'

The phrase lavel Fo pattern is falling. Both the beginning and end points are higher than in the statement. A slight modification on this falling pattern is made in the first two syllables. The intensity structure of shows also a higher structure than in statements. The intensity is generally higher with 5-10 dB in comparison with statements.

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

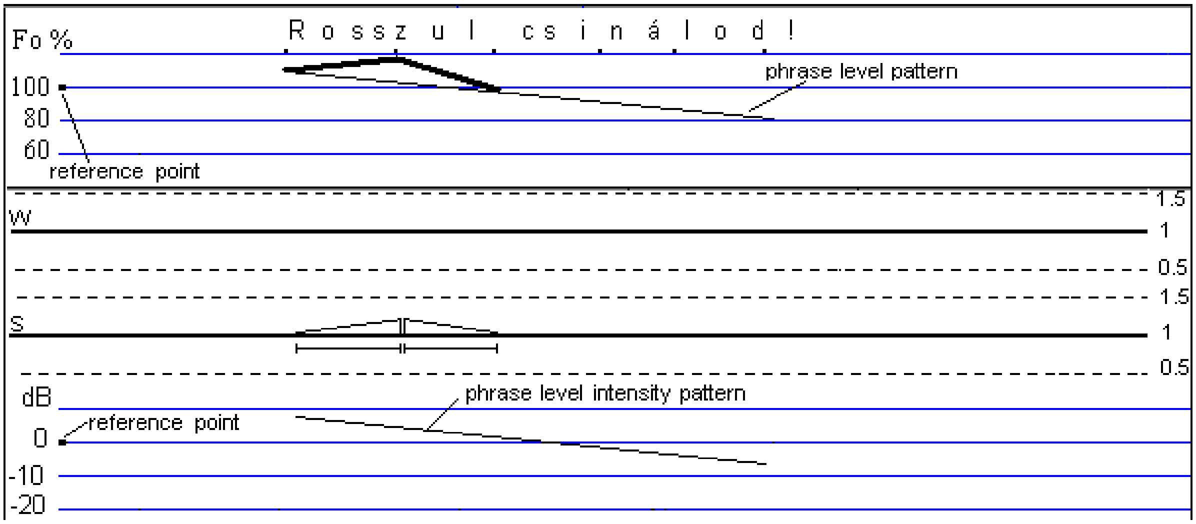

Figure 14.
The stylized Fo and intensity structure of the warning.
The short thick vertical lines under the text mark syllable boundaries

Commands

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

Example: *Ne menjetek oda!* ’Do not go there!’

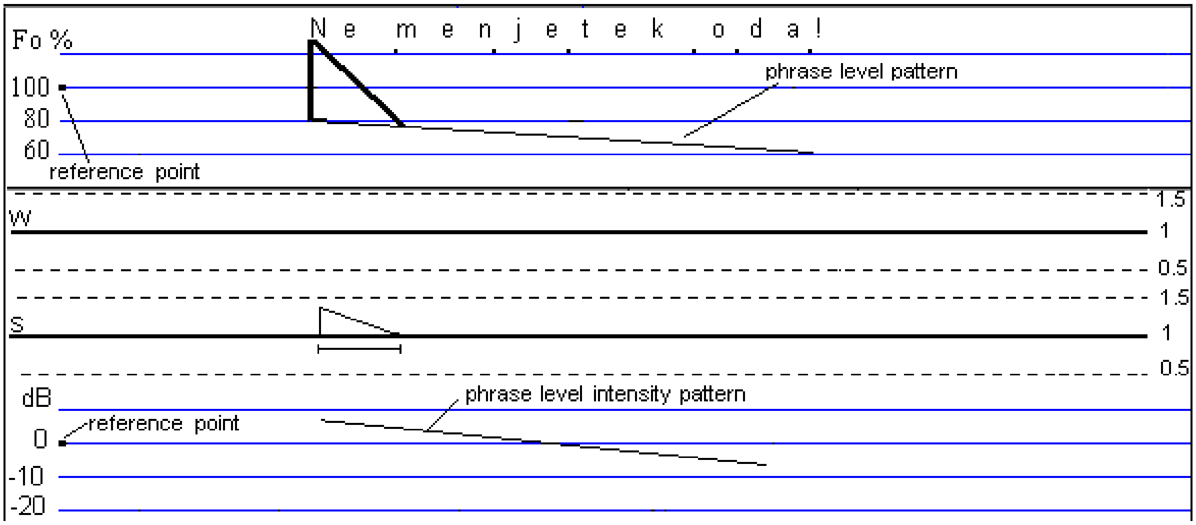

Figure 15.
The stylized Fo and intensity structure of the warning

Sentence expressing desire

Mainly the sentences beginning with the interjection *Bárcsak...*('I would like if only…') have been analised.
Example: *Bárcsak eljönne a barátom!* 'I wish if only my friend would come!'
The phrase level Fo pattern is falling. The Fo begins on a slightly lower frequency (90%) as 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 is the desire the higher is the peak. The stylized representation is shown in Figure 16.

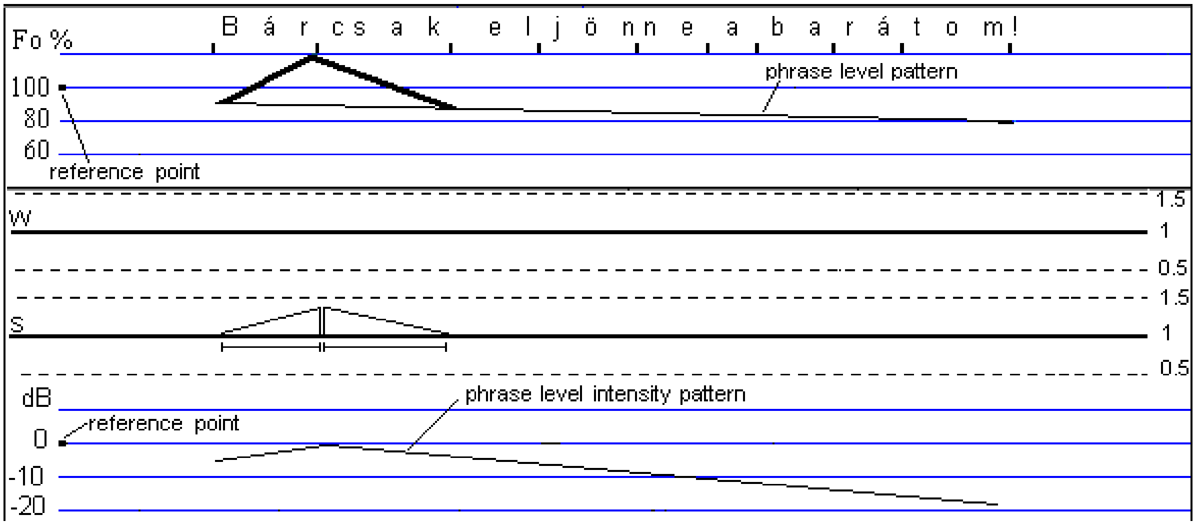

Figure 16.
The stylized Fo and intensity structure of the sentence expressing desire

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

Listening tests
Two listening test have been carried out. The aim of the first was to compare the natural and synthetically generated Fo and intensity patterns, in the second one the prosody of generated dialogues was tested.
Test 1.
The test material consisted of 10 sentence pairs. In one pair two sentences have been put one after the other separated with 3 seconds break. The first sentence was the natural one and served as carrier sentence for the second one. In the second sentence the predefined Fo pattern (according to the data of the unified Fo scale) was superimposed on the body of the carrier one.

Thus the two sentences in one pair had the same segmental speech body, the difference was only in the realisation form of the Fo structure. Ten such sentence pairs (3 Wh and, 2 Y/N questions, 3 commands, one request and one statement) were prepared and used in the test. Twenty test persons (8 female and 12 male persons, age from 25 to 55) had to verify in a scale how close is the simplified and modeled Fo pattern to the natural one. The task was: compare the melody of the two sentences and evaluate according to the following scale: they are the same, very similar, similar, less similar, different.

Results

The distribution of the responses is shown on Figure 17. Summarising the results of the first three columns 86.5 % of the responses found the modeled Fo structure similar or better to the original one. This high score allows us to declare that the description of the phrase level Fo patterns and the word and syllable level local modifications on it represents tolerably well the structure of Hungarian Fo patterns at the sentence level. The sentences receiving the less similar or worst evaluation were once more examined concerning the modeled Fo structure. It became clear that not only the slight difference between the natural and modeled Fo structure was the basis of these negative judgements, but in some cases the slight difference of the general fundamental frequency level of the two sentences (for example the natural sentence sounded slightly higher than the modeled one, but the general Fo structure was very similar). This latter case was due to the fact that because during the transplantation of the modeled Fo structure the reference points of the structure have to be given too in Hz. This value defines the height of the general Fo structure of the utterance. During the whole procedure this reference point was given the same value. In natural speech the general Fo level may change with 2-8 Hz from sentence to sentence. Thus in some sentences the modeled Fo structure sounded slightly different in general Fo height. Some test persons found this difference enough to give a response „less similar” or „different”.

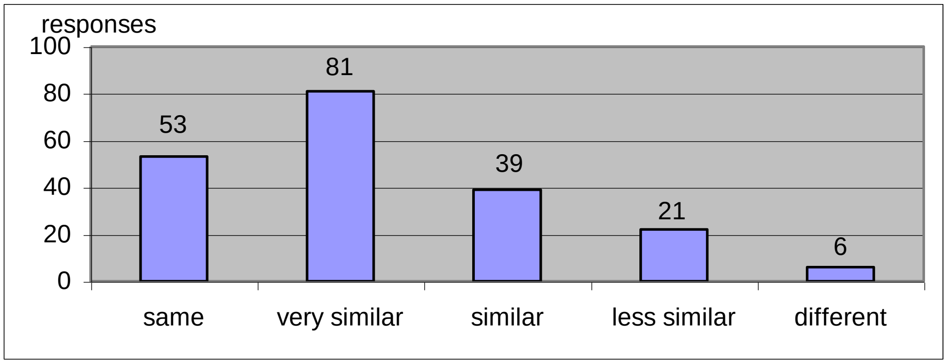

Figure 17.
Evaluation results of comparision of natural and predefined Fo structure of Hungarian sentences

Test 2.

The goal of this test was to know whether the concatenated unified melody forms – to characterise the complex melody of a dialogue – give the impression of the dialogue. Dialogue elements (two-three sentences concatenated one after the other) have been constructed according to the modeled Fo structures using natural carrier sentences. Different transformations have also

been made concerning the Fo structure (For example: statement, control question and final statement)
Example
A tervezett tárgyalás után levelet írok a külföldi partnernek. (basic carrier sentence)
A tervezett tárgyalás után? (controll question, generated from the first part of the carrier sentence)
A tervezett tárgyalás után. (final strengthening statement, generated from the controll question)
In the transformed sentences the time structure of the sound sequences was not changed only the

Fo structure and the intensity structure was set according to the previously defined values. The

Hz value of the reference point was the same in all sentences. Four dialogues have been constructed. The question for the test persons (the same 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 on Figure 18.

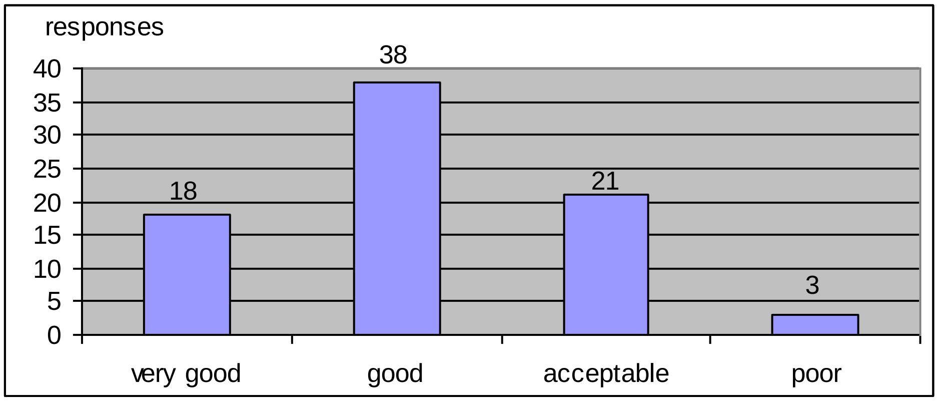

Figure 18.
Evaluation results of the general Fo structure of the four dialogues

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

Conclusions
This research concentrated on the systematic description of the intonation and intensity structures of the most frequent Hungarian sentence types (statements, warnings, requests, commands and 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 form one to the other. The unified scale helps to express the function among the melody forms of the sentences. The description of the Fo and intensity patterns is based on three data structures: the phrase leve 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 change 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 withouth acoustical analysis if the following information is available: the sentence type, the sentence structure, the phrase boundaries and the accent distribution.

In the Hungarian prosody the falling phrase level pattern is characteristic for the majority of sentence and expression 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 form basically the intonation of the given sentence. The rising phrase level pattern is characteristic only in yes/no questions and in controll and elliptic ones. The syllable and word level local changes – modulating the phrase contour - have important roll 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 analysed sentences can be summarised as follows. The intensity level is high if the Fo is high and vica 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 the melody. The topic-focus has 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 the Wh-questions and alternative questions – requires a rising pitch contour for the topic part. However the rising melody contour in the main part – as in yes/no questions – is preceeded by a descending one in the topic part. As to the transformation possibility among different modalities the realisation of the proper intensity contour may be as such important as the realisation of the proper Fo curve. This is the the case mostly when questions having rising contour are formed from statements.

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

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

Acknowledgements

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

REFERENCES

Collier R. (1990). Multi-lingual intonation synthesis: Principles and applications. *Proc. ESCA Workshop on Speech Synthesis, Autrans, France*, pp. 273-276.

Fónagy I., Magdics K. (1969). *A magyar beszéd dallama.* Akadémiai Kiadó, Budapest

Fujisaki, H.(1992): Modeling the Process of Fundamental Frequency Contour Generation. In: Speech perception, production and linguistc structure. (Eds.: Tohkura Y., Vatikoits Bateson E., Sagisaka Y.) IOS Press, Tokyo, 1992. 314-326.

Gósy Mária (1992): *Speech perception*. Frankfurt am Main: Hector

Ladd, D. Robert. *Intonational Phonology*. Cambridge University Press, Cambridge 1996.

Möbius, Bernd (1997): Synthesizing German Intonation Contours. In: Progress in Speech Synthesis. Eds. Jan P.H. van Santen et al. Springer. p 401-415.

Montero J.M., Gutiérrez-Arriola J., Colás J., Macias J., Enriquez E., Pardo J.M. (1999). Development of an emotional speech synthesiser in Spanish. *Proc. of the 6th European Conference on Speech Communication and Technology*, pp. 2099-2102.

Olaszy G. (1989). *Elektronikus beszédelőállítás.*(’Electronic Speech Generation’) Műszaki Kiadó, Budapest

Olaszy G. – Németh, G. (1999): IVR for Banking and Residential Telephone Subscribers Using Stored Messages Combined with a New Number-to-Speech Synthesis Method. In: Human Factors and Voice Interactive Systems. Ed.: Daryle Gardner-Bonneau. Kluwer Academic Publishers, 1999. pp.237-256.

Olaszy G. (2000): The prosody structure of dialogue components in Hungarian. *International Journal of Speech Technology Vol 3/4 . 2000. pp. 165-176.*

Olaszy G. – Németh, G. – Kiss G. (2001): Hungarian audiovisual prosody composer and TTS development tool. In: Prosody 2000. Editors: Puppel Stanislaw, Grazina Demenko. Poznan, 2001. 167-178.

Rank E. – Pirker H. (1998). Generating emotional speech with a concatenative synthesiser. Proc. of ICSLP Sydney, pp. 947--950

Silvermann, K. – Beckman, M. – Pitrelli, J. – Ostendorf, M. – Wightman, C. – Price, P. – Pierrehumbert, J. – Hirschberg, J.: ToBI: a standard for labelling English prosody. In Proc. of ICSLP 92 Vol. 2. 1992. 867-870.

Taylor, P. (1998): The Tilt Intonation Model. Proc. of the ICSLP98 1243-1247.

Taylor, P. (2000): Analysis and Synthesis of Intonation using the Tilt Model. Journal of the Acoustical Society of America. 107/3, 1697-1714.

Terken J. – Collier R. (1990). Designing algorithms for intonation in synthetic speech. *Proc. ESCA Workshop on Speech Synthesis, Autrans, France*, pp. 205-208.

Varga L. (1993). *A magyar beszéddallamok fonológiai, szemantikai és szintaktikai vonatkozásai* . (’The phonological, syntactic and semantic aspects of Hungarian speech melody’) Nyelvtudományi Értekezések 135. Budapest