How can a functional perspective be used in intonation modelling

Doina Jitca[1], Vasile Apopei[1], Magdalena Jitca[2]

[1] Institute for Computer Science, Romanian Academy, Iasi Branch, Romania
[2] University "Alexandru Ioan Cuza", Iasi, Romania

jdoina@iit.tuiasi.ro

Abstract

This paper presents some reasons to take a functional perspective in intonation modelling into account. We distinguish two functional prosodic unit types that are functionally equivalent: an elementary one - the accentual units (AUs) - and other, nonelementary- the AU groups (AUGs). Their functions at the communicative act level give meanings to corresponding F0 contour segments. A functional category includes different F0 contour prototypes. We illustrate a manner of dealing with a functional model by means of a set of functional prosodic unit categories generated by a Romanian intonation modelling. We consider the accentual units, their groups and their functional labels are proper terms for describing the meanings of the intonational contours.

1. Introduction

Our intonation model belongs to the perceptual model category and it has many aspects in common with the IPO theory of intonation structure. It was motivated by the search for a model of Romanian intonation in order to be used in speech synthesis. In our view the prosody is a means by which the speaker communicates to the listener the content of his discourse units: words, word groups. Consequently, the intonation melodic flow may be thought as consisting of small melodic unit sequences, corresponding to the above mentioned textual units. The intonation production process can be represented in a systemic view by a module that applies different "prosodic accents" which are perceived as different "melodies" applied to the input words and/or word groups (Figure 1).
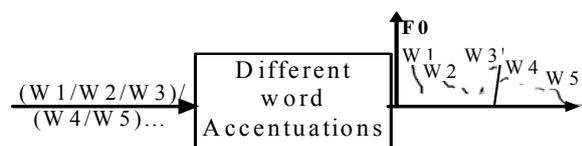

Figure 1: The intonation production perspective as "accent" sequence generation, applied to related textual unit sequences

For Romanian intonation, we consider it is important to keep together, in the same pattern, the tones implied by pitch accents with those from the prosodic unit beginning and ending. The same pitch accent type gives different meanings to the unit in different beginning/ending tonal ratio cases. Furthermore, we have tried to give meaning to every prosodic unit even if it contains no significant pitch movement, because it is possible the related word carries the sentence emphasis.

Our model, in the same manner as the IPO theory, idealizes the F0 contour as a sequence of predefined pitch unit forms corresponding to the elementary intonation units [1]. Unlike the IPO model that applies no meaning (function) to the pitch movement, we relate the predefined F0 contour patterns to a set of functional categories at communicative act level. Xu claims that the communicative meanings are not only important for prosody but they are the essence of it [2].

The functional descriptions apply invariant meanings to the prosodic units with different F0 contour forms and even different scopes.

Kohler also distinguishes between several types of word accentuations at semantic level, for German, but they can also be related to certain communicative functions, corresponding to certain pitch configurations [3].

The pitch movement is the smallest descriptive unit of IPO model. In an equivalent manner, our model works with accentual units (AUs) as elementary prosodic units (the smallest units of perceptual analysis). An AU corresponds to one accented word, but may also include other unaccented words [4].

Analogue to the combinations of pitch movements into "configurations" from IPO model, our model defines the AU groupings as nonelementary prosodic units (AUGs). They may be equivalently related to the functional labels used for AUs annotation. In his *functional semantic model* - the FSP theory- Firbas also generalized the semantic functions defined at elementary unit level in order to be used in group unit annotation [5]. Finally, the F0 contour is generated by a sequence of elementary patterns, but their mapping in the space (time, F0) is modulated by the utterance tree hierarchy.

The analysis of the F0 melodic contours within the Romanian natural AUs drives us in building a set of functional categories, each of them consisting of several F0 pattern variants. After that, a set of functional labels was built related to these categories [4], [6].

In our functional modelling the ToBI pitch accent labels are used as phonologic attribute values assigned to one AU category, in order to discriminate between different AU melodic contour variants. For example, a prosodic unit that both plays a role in pushing forward the communicative act in a group or a sentence, and focuses the related word, at local group level or at sentence level should be assigned the function "PUSH" and an L+H* value for its pitch accent parameter. In this manner a reduced number of functions can be assigned to several AU melodic contour prototypes.

In the present paper we explain how the F0 contour of an utterance can be described by functional label structures related to the utterance tree structures. In the next section we explain why we need the recursive prosodic unit hierarchy in our functional modelling. The third section contains a short presentation of the function set the model works with. In the fourth section we exemplify the use of the functional labels in annotating Romanian F0 contours.

2. The prosodic domain hierarchy

Our intonation modelling defines three domain types into the prosodic domain hierarchy, above the syllable level: "accentual unit (AU)", "accentual unit group (AUG)" and intonational phrase (IP). What is particular to our prosodic domain hierarchy refers to the exclusion of the "Strict Layer Hypothesis" constraint. Consequently, an AUG may consist of AUs but, in general cases, it may also have constituents of AUG type (Figure 2).
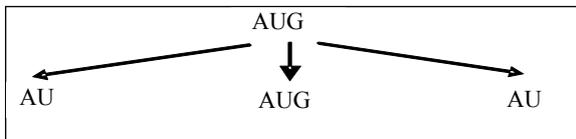

Figure 2: The AUG as a compound prosodic domain

A common neutral Romanian intonation for an intonational phrase or just for an AUG consists of two prosodic units that generate the tonal contrast of the unit. Usually, they are in the first and in the last position, respectively. This tonal contrast they generate between a relative high tone vs. a relative low one is the essence of one unit. The AU groups introduce local focuses at different levels of the utterance tree.

Between the tonal contrasted units a special shaped unit (AU or AUG) focuses the word or word group in the middle position (if such a position exists). The next unit after the middle focused group must be linked with the first unit on the same level of the utterance tree, because they are in a tonal contrast relation.

That explains the option for the recursive structures in our intonation modelling. The acoustical evidences for the middle focused group consist in the tonal return in the end of the group at the same tone level as the beginning.

D.R. Ladd introduces in [7] the term of compound prosodic domain (CPD) in order to define a prosodic domain X whose immediate constituents are themselves of type X. We make this definition more general by accepting that the constituents of the prosodic domain AUG may be both of the same type AUG and AU type. Despite the functional equivalence between the AU and AUG domains we consider them distinct types because the prosodic units of the first type are *elementary prosodic units* and those of the second type are *non-elementary prosodic units*.

Figure 3 illustrates an example of utterance tree that consists in both recursive and non-recursive prosodic structures.
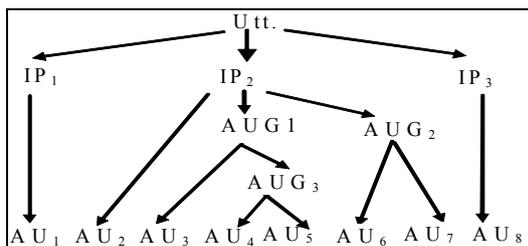

Figure 3: An example of utterance tree

The recursive prosodic hierarchies in our intonation model are a consequence of changing the non-elementary unit perspective from the weak/strong partitioning into a more general one, as a sequence of functional components.

The elementary AUs are annotated by a functional label to which the model assigns either implicit or explicit (selected by parameters) *F0 contour patterns*.

The AUG content is described by a functional label sequence related to their component units. These are functional descriptions for non-elementary prosodic units. To these functional label sequences, the model assigns corresponding *envelope patterns* for the F0 contour. The AUG F0 contours are generated by concatenating the patterns of their AU constituents, modulated by the branch hierarchy.

3. The functional prosodic categories

In [6] we propose the following four basic functional units in order to describe the neutral Romanian intonation at AUG or IP level:

- PH ("PUSH") units contribute to pushing forward the communication act by creating high tension segments of speech at its parent unit level.
- PO ("POP") units correspond to the low tension speech segments within the communication act at its parent unit level. The name "POP" suggest that a *POP* unit is related to a *PUSH* unit by a tonal contrast.
- L ("Link") units generate tonal linking between units of the previous types, without generating contrastive focus at the related word level. It occurs in certain word deaccentuation cases.
- F/f ("Focus") − units having the role only in word neutral focusing (F-more prominent/f-less prominent).

Figure 4 illustrates the mapping of the functional sequence PH/F/PO corresponding to a Romanian affirmative intonation, in neural manner. The mapping applies tonal limits (Top/Bottom pair sequence T1-B1/T2-B2/T3-B3) for the F0 contour patterns related to the sequence components, into the (T1, B3) F0 scale of the utterance.
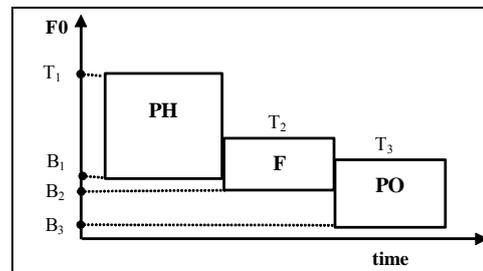

Figure 4: The *envelope pattern* definition for the F0 contour described by the functional sequence PH/F/PO

Our model assigns corresponding Top/Bottom coordinate sequences to each functional sequence in order to use them in mapping the component F0 patterns into the relative F0 scale related to their parent AUG/IP. In the particular case of the PH/F/PO sequence, the F0 contour downsteping tendency (declination phenomena) is a consequence of the decrease of the Bottom level from its mapping description, from PH to PO unit.

Other two variants of PUSH and POP functional units are defined as PUSH-Down and respectively, POP-Up. The PUSH-Down unit performs a PUSH event during the accented syllable and then brings the tone at lower level. The POP-Up unit performs the POP event and then brings the tone at higher level. Based on the basic functional unit categories we have defined other derived functional categories. In Table 1 the functional prosodic categories are characterized into the space (T_{target}, $\Delta F0_{beg/end}$, R) in order

to prove that each function has at least one predominant feature not overlapped by other function [2]. The feature space (T_{target}, $\Delta F0_{beg/end}$, R) has the following definition:

$$T_{target} = \text{Target tone level on the stressed syllable}$$
$$\Delta F0_{beg/end}=F0_{begin}-F0_{end}$$
$$R=F0_{max}- F0_{min}$$

Table 1. The prosodic category characterization into the space (T_{target}, $\Delta F0_{beg/end}$, R)

PH	T_{target}= near the Top R=Large, medium values $0>\Delta F0_{beg/end} >= -R$
PO	T_{target}= near the Bottom R=Large, medium values $0<\Delta F0_{beg/end} <= R$
PD	T_{target}= near the Top R=Large, medium values $0<\Delta F0_{beg/end} <= R$
PU	T_{target}= near the Bottom $0>\Delta F0_{beg/end} >= -R$ R=Large, medium values
L	Bottom$<T_{target}<$ Top R=Large, medium, small values $|\Delta F0_{beg/end}|\sim R$
f	Top-ΔT$>T_{target}>$Bottom+ΔT -dF0$< \Delta F0_{beg/end}<$dF0 (|dF0|~0) R= small values
F	Top-ΔT$>T_{target}>$Bottom+ΔT -dF0 $< \Delta F0_{beg/end}<$dF0 (|dF0|~0) R= medium values
PH+f	T_{target}= near the Top -dF0$< \Delta F0_{beg/end}<$dF0 (|dF0|~0) R= small values
PH+F	T_{target}= near the Top -dF0$< \Delta F0_{beg/end}<$dF0 (|dF0|~0) R= large, medium values
PO+f	T_{target}= near the Bottom

	-dF0$< \Delta F0_{beg/end}<$dF0 (|dF0|~0) R= small values
PO+F	T_{target}= near the Bottom -dF0$< \Delta F0_{beg/end}<$dF0 (|dF0|~0) R= large, medium values
PH+PO	T_{target}= near the Bottom $0<\Delta F0_{beg/end} <= R$ R= Top-Bottom.
PO+PH	T_{target}= near the Top $0>\Delta F0_{beg/end} >= -R$ R= Top-Bottom.

Table 1 illustrates the "mutual-exclusivity" principle formulated by Xu. We also take into account the following principles summarized in [2]:
- "specificity": by defining the prosodic components in terms of different communicative functions (in the beginning of this section);
- "audibility": by limiting the number of AU grouping levels at maximum three in order to generate an audible F0 range for each unit of the utterance tree.

The functional prosodic categories were implemented in the Romanian speech synthesis module by defining prototypes for the elementary F0 contours and for the F0 contour envelopes for the IPs/AUGs.

4. The functional intonation description

We exemplify how the functional labels can be used in the Romanian intonation description.
We choose four natural F0 contours related to four speakers' utterances (two males and two females) of the Romanian text *Lumina era slabă dar o recunoscu fără greutate* (*The light was weak but he recognized her without difficulty*). The utterance trees, related to the four speakers' intonation variants, are presented in Table 2 by spanning the sentence at prosodic level, into the AUs and AUGs annotated by functional labels (the AUs are delimited by slash).

Table 2. The four variant of spanning the text into the functional annotated prosodic units

AM	(*Lu'mina era* $_{PH:L+H*}$ /*'slabă* $_{PU}$)$_{PH}$ /*dar o recunos'cu* $_F$ /*'fără* $_f$ / *greu'tate* $_{PO}$
ID	(*Lumina e'ra* $_{PH}$ / *'slabă* $_{PU}$)$_{PH+F}$ /(*dar o recunos'cu* $_{PH}$ /*'fără* $_f$ / *greu'tate* $_{PO}$)$_{PO}$
BC	(*Lu'mina* $_{PH}$ / *e'ra* $_L$/ *'slabă* $_{PU}$)$_{PH+F}$ /(*dar o recunos'cu* $_{PH}$ /*fără greu'tate* $_{PO}$)$_{PO}$
JD	(*Lu'mina* $_{PH}$ /*e'ra* $_L$ / *'slabă* $_{PO+PH}$)$_{PH}$ /*dar o recunos'cu* $_F$ /*'fără greutate* $_{PO}$)$_{PO}$

One observes the auxiliary verb (*era*), the conjunction (*dar*) and the preposition (*fără*) may generate or not separate AUs. The round parenthesis encloses the text of the AUGs. The functional labels are edited in subscript manner after each text unit. One functional label is followed by a ToBI label if the pitch accent must be explicitly annotated. These intonational descriptions should be correlated with the F0 contours illustrated in Figures 4 and 5, corresponding to the male (AM, ID) and respectively, female (BC, JD) speakers' utterances [8]. In each figure the F0 contours are synchronized at vowel segment level in order to be easily compared.
At semantic level, the sentence contains a contrast between two statements: the first, *Lumina era slabă* and the second one, *dar o recunoscu fără greutate*. A major difference between the four intonational variants refers to which of the statements is focused. The AM and JD speakers focus on the verb from the second statement *(dar o recunoscu)* by keeping the ending tone and the beginning tone at the same level. At ID and BC speakers one observes that: the verb *recunoscu* is deaccentuated (PH unit - the ending tone is higher then the beginning tone) while the first group *Lumina era slabă* is focused within a group that also generates the PUSH event of the

IP (PH+F group unit - the ending tone equals the beginning tone). The functional melodic contour descriptions at highest level of the AM and JD utterance trees are: PH/F/f/PO and PH/F/PO, respectively. The weak focus ("f "type) is often related to functional words such as prepositions (i.e. *fără*). The deaccentuation of the first statement (PH group) is generated by raising the final tone of the first group above the beginning tone level of the group.
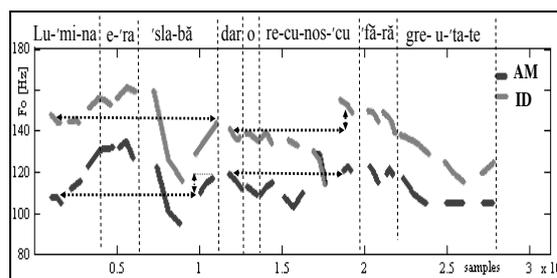

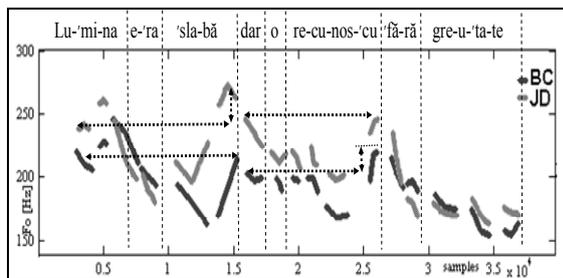**Figure 4**: The F0 contours extracted from AM and ID speaker utterances

Figure 5: The F0 contours extracted from BC and JD speaker utterances

In the BC and ID cases the melodic contours at the highest level are described by PH+F/PO and, at lower PO group level by: PH/PO and respectively, PH/f/PO. The AU of PO type has in the implicit case an L* pitch accent.

The first statement (*Lumina era slabă*) has its local focus positioned on the predicate adjective *slabă (weak)*, generated by means of a PU unit (AM, ID, BC speakers) and a prominent one of PO+PH type (JD speaker). The PO+PH unit has a more prominent L*+H pitch accent than the PU unit (L* pitch accent) and a more prominent boundary tone that reaches the top level. The melodic contours at the first group level are: PH/PU and respectively, PH/PO+PH, both of them corresponding to nonfinal affirmative intonation. The high boundary tone is included in the last AU pattern. They are not explicitly annotated.

Other local differences between speakers related to the first statement group are generated by the auxiliary verb *era*. The female speakers (BC, JD) during the verb *era* bring the tone down to the beginning of the last AU within the first group. In this manner an AU unit of type L is generated during the auxiliary verb, as in most deaccentuation cases of Romanian neutral intonation. In the male speakers cases (ID, AM) the auxiliary verb is included in the local PH event. The speakers raise the tone up to the top in two steps, during the word *lu'mina* and continuing during the auxiliary verb. The ID speaker generate the principal accent on the word *e'ra* while the AM speaker, on the accented syllable of the word *lu'mina*.

The functional intonation model applies melodic contours at different group levels that generate different variants of word accentuation /deaccentuation at semantic level. We consider the elementary melodic contour patterns, their groupings and their functional labels are proper terms for describing the meaning of the intonational contours.

5. Conclusions

Unlike the IPO approach in which the meaning of function plays no role in the intonation analysis [7], we conclude that the intonation meaning must be thought in relation with the textual units (words and word groups) to whom their elementary and non-elementary melodic contours give functions at the communicative act level. The melodic pattern of a prosodic unit must be correlated with those of other units from its neighbourhood in order to understand at what group level of the utterance tree its function works.

The AUGs are considered, in essence, a tonal contrast between the tonal characteristics of two from its component units (a high unit vs. a low one). The groups may contain other functional components and so results a more general partitioning into functional unit sequences, that replaces the well-known weak-strong partitioning. This perspective introduces recursive unit structures into the utterance tree hierarchies. Each group from an utterance tree contains a local focus.

A good understanding of the different types of semantic accentuation and deaccentuation events can be achieved by relating all the prosodic units (elementary and nonelementary) into a functional hierarchy.

In the particular case of focused AUGs, the ending tone perceptually equals the beginning tone (the small tonal differences must be under the perception threshold).

We consider each prosodic unit category applies to a word/word group a type of "accent" at prosodic level that may be significant or not at semantic level. We consider these "prosodic accents" as functions at the communicative act level. Without a special training, the native speakers can't separate an IP melody into its functional components. They unconsciously learn the melodic contours *at intonation phrase level* related both to their enounce type and to the particular message they want to transmit to the listener, but they make no choice of different particular melodic components. But for speakers that learn Romanian as a foreign language, we consider it is useful to first learn the elementary melodic contours related to their communicative functions (PUSH/POP events, prominent/nonprominent focuses, tonal linking) and after that, the rules to combine them into complex intonational contours corresponding to different morphologic /syntactic text structures.

6. References

[1] 't Hart, J., Collier R. and Cohen, A., "A perceptual study of intonation", Cambridge University Press, 2006.

[2] Xu, Y., "Speech prosody as articulated communicative functions", Proceedings of "Speech Prosody", Dresden, Germany, 2006.

[3] Kohler, K. J., "What is Emphasis and how is it Coded?", Proceedings of "Speech Prosody", Dresden, Germany, 2006

[4] Jitca, D., Apopei, V. and Jitca, M., "The F0 Contour Modelling as Functional Accentual Unit Sequences", *International Journal of Speech Technology*, DOI 10.1007/s10772-009-9055-3, Nov, 2009.

[5] Svoboda, A., "An ABC of Functional Sentence Perspective (An e-learning text for the students of English)" , 2005.

[6] Jitca D., Apopei V. and Jitca M., "A description language at the accentual unit level for Romanian intonation", Poster and abstract in the booklet of LangTech2008 Conference, p.45, Rome, www.langtech.it/en/poster/01_JITCA.pdf, 2008.

[7] Ladd, D. R., „Intonational Phonology", Cambridge University Press, 1996.

[8] Download wav files at http://www.academiaromana-is.ro/iit/personal/wav_jd.zip