

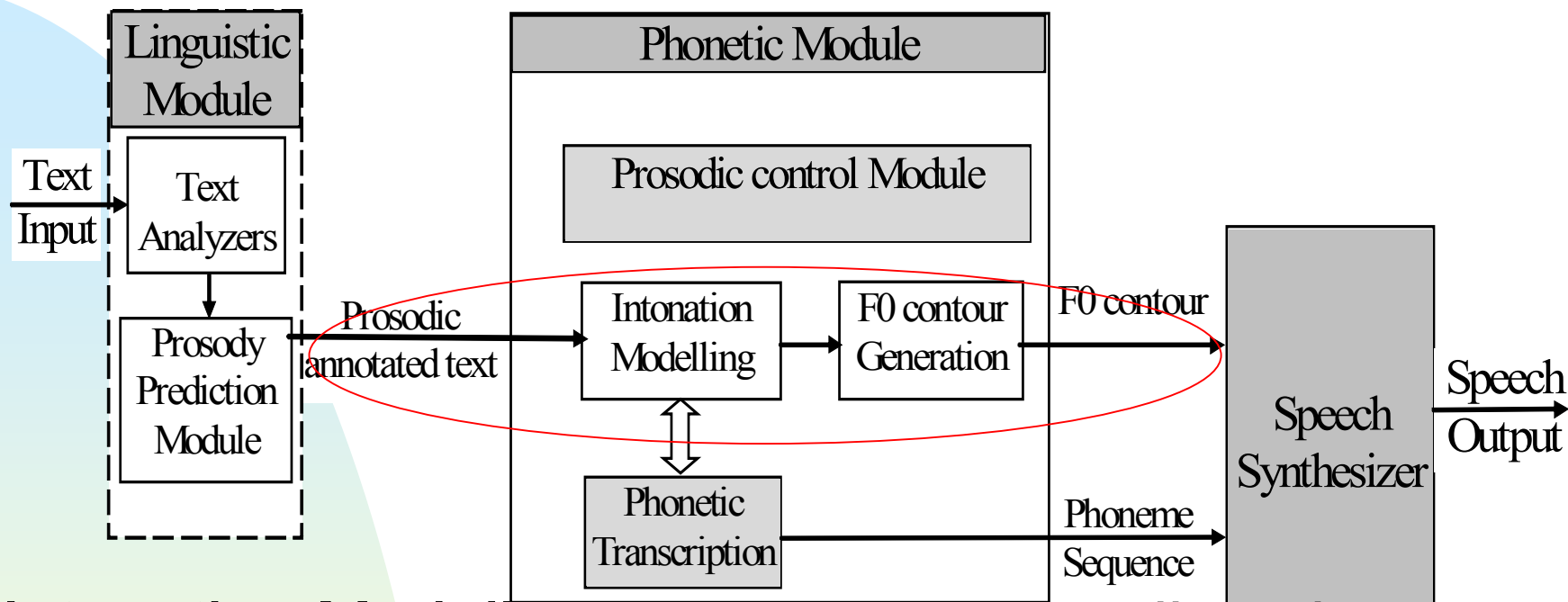
A PROSODIC CONTROL MODULE FOR A ROMANIAN TTS SYSTEM, BASED ON MELODIC CONTOUR DICTIONARIES

Doina Jitcă, Vasile Apopei

Institute of Computer Science, Romanian Academy,
Iasi Branch, Romania



The Romanian Text-to-Speech system



□ *Intonation Modelling* - maps the prosodic units annotated by prosodic labels, into the synthesized utterance F0 scale and selects their F0 patterns.

□ *F0 Contour Generation* - scales the F0 patterns for each prosodic unit and anchor them at the tonal coordinates determined by *Intonation Modelling*.



The Intonation Model

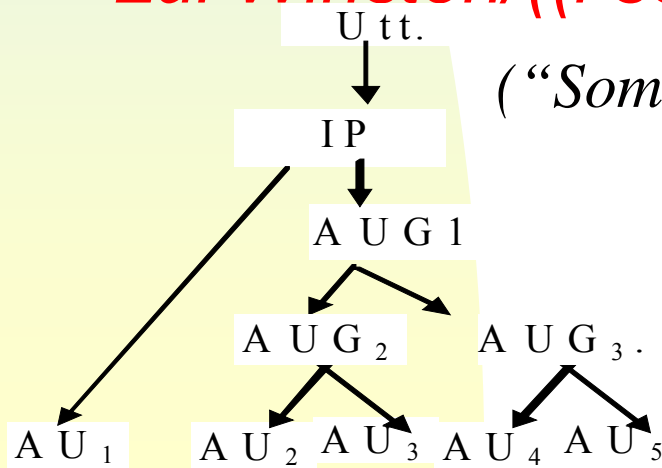
- An intonation model offers:
 - a perspective for understanding the intonational phrase meaning
 - A basis for the prosodic description language
- The perspective: Our intonational modeling relies in concatenating a sequence of F0 contour patterns that correspond to certain elementary melodic contours (EMCs) considered as being functional at communicative act level.
- The description language: is based on a set of prosodic functional labels with attributes for the phonologic and prominence characterization



The Prosodic Units

- The Functional EMCs: are functional prosodic units of Accentual Unit (AUs) type (or Stress Unit).
- The AUs may be grouped into Accentual Unit Group(AUGs) and generates non-EMCs
- Our functional perspective decomposes the Intonational Phrase (IP) into a hierarchy of prosodic units (AUs and AUGs)

“Lui Winston/((i se treziră/niște)/(vagi /amintiri))”

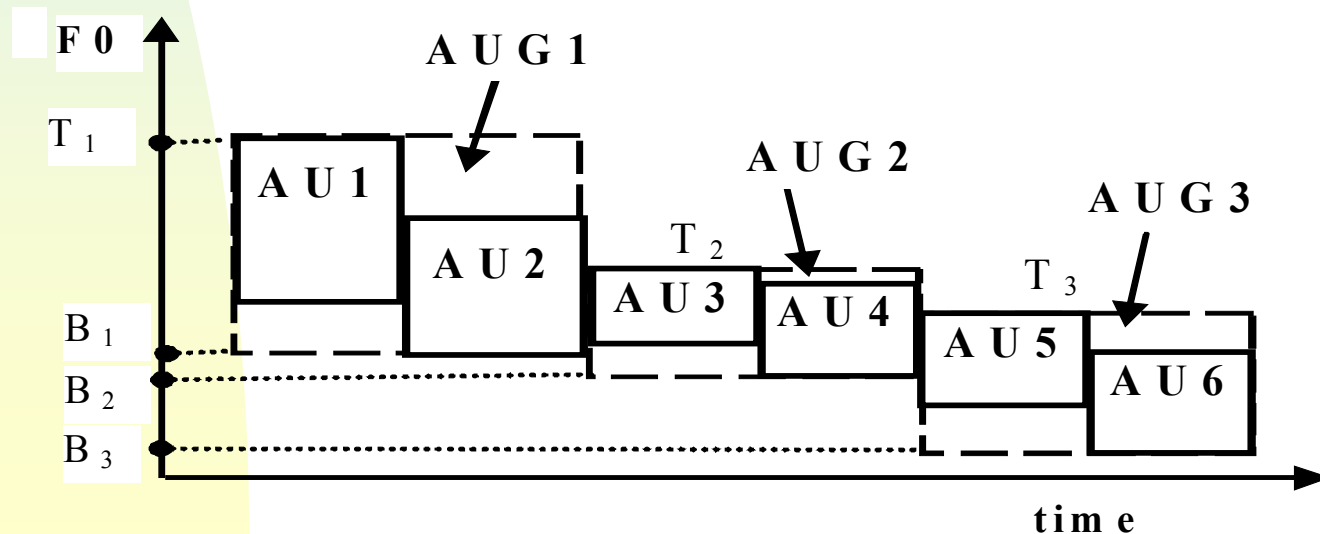


(“Some vague memories arised to Winston.”)



The Prosodic Unit Function → → The Prosodic Unit Mapping

- After the utterance tree mapping into the utterance F0 scale, a sequence of AUs results. Their F0 pattern concatenation generates the utterance F0 contour.
- The reasons for AU grouping is to put them into a tonal relation in order to generate local meanings at AUG levels and global meanings at IP levels



The Basic Prosodic Functions

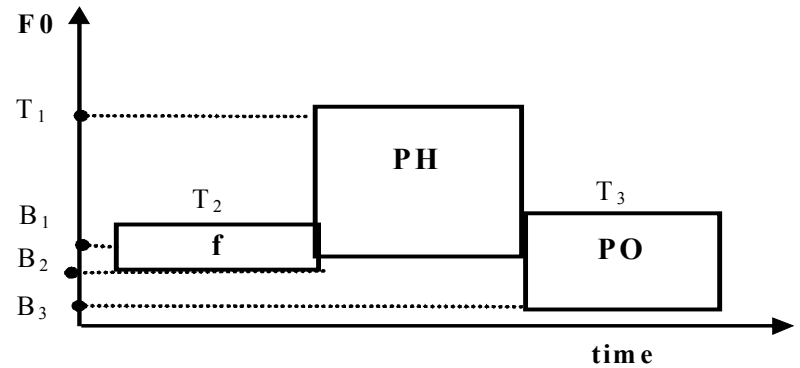
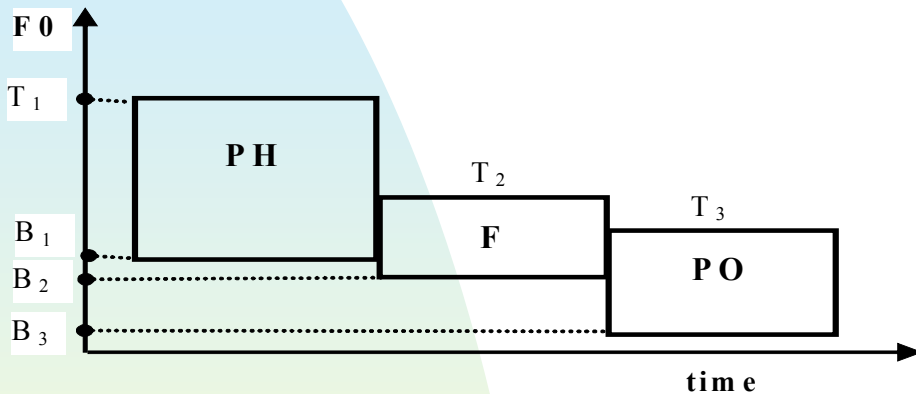
□ The **PUSH** function units contribute to pushing forward the communication act by creating high tension segments of speech with one target tone up to the top level of the Group F0 scale, reached usually by means of large pitch movements, during the accented syllable and even on the following unaccented one.

□ The **POP** function units corresponds to the relax or low tension speech segments within the communication act by targeting tones near the bottom of the AUG/IP F0 scale. The name “POP” suggest that a POP unit is related to a PUSH unit by the means of a tonal contrast.

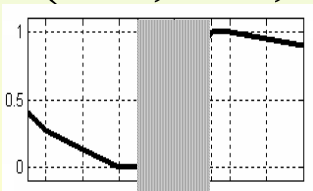
□ The *F/f* (“Focus”) function units having the role only in word neutral focusing (F-more prominent/f-less prominent). Their acoustic characteristics consist of F0 variations around a tonal level, after the focusing tonal level is reached (by large/small pitch range variations).

The Basic Prosodic Functions

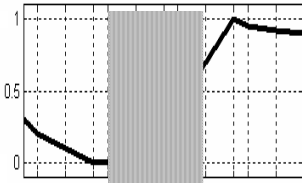
- the function of a prosodic unit → :
 - the position and dynamic into the IP/AUG F0 scale



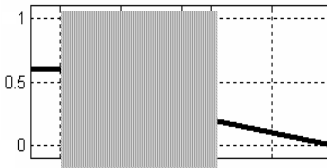
- the pattern – controlled by attributes: pitch accent type - (H*, L*, L+H*, H+!H*) or tonal contrast size (top-bottom)



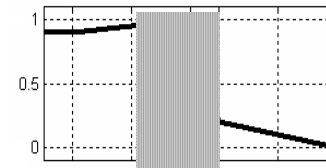
PH - H*



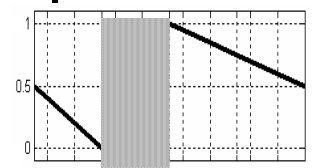
PH - L+H*



PO - L*



PO - H+!H*



F - H*



The Derived Prosodic Functions

PD - PUSH-DOWN - performs a PUSH event and then it brings the tone at a lower level.

PU-POP-UP - performs a POP event and then brings the tone at higher level in order to suggest the continuation of the melodic phrase

PH+f - performs a PUSH event (Acc syl) and a minor focus

PH+F- performs a PUSH event (Acc syl) and a major focus

PD+f- performs a PUSH event (NAcc syl) and a minor focus

PD+F- performs a PUSH event (NAcc syl) and a major focus

PO+f- performs a POP event and a minor focus

PO+F- performs a POP event and a major focus



The Prosodic XML Schema- for prosodic text annotation

Tag	Attribute	Values	Unit type
<Utt>			Utterance
<IP>	<i>Function</i>	PH, PO, F	Intonational Phrase
	<i>TonalContrast</i>	s(trong),w(eak)	
<AUG>	<i>Function</i>	PH, PO, f, F, PU, PD, PH+f, PH+F, PO+F, PO+f, PD+F, PD+f	Accentual Unit Group
	<i>TonalContrast</i>	s(trong),w(eak)	
<AU>	<i>Function</i>	PH, PO, f, F, PU, PD, PH+f, PH+F, PO+F, PO+f, PD+F, PD+f	Accentual Unit
	<i>Accent</i>	H*, L*, L+H*, H+!H*, L*+H	
	<i>TonalContrast</i>	s(trong),w(eak)	
<Syl>	<i>Length</i>	Small, Medium, Large	Syllable
	<i>Energy</i>	Small, Medium, Large	



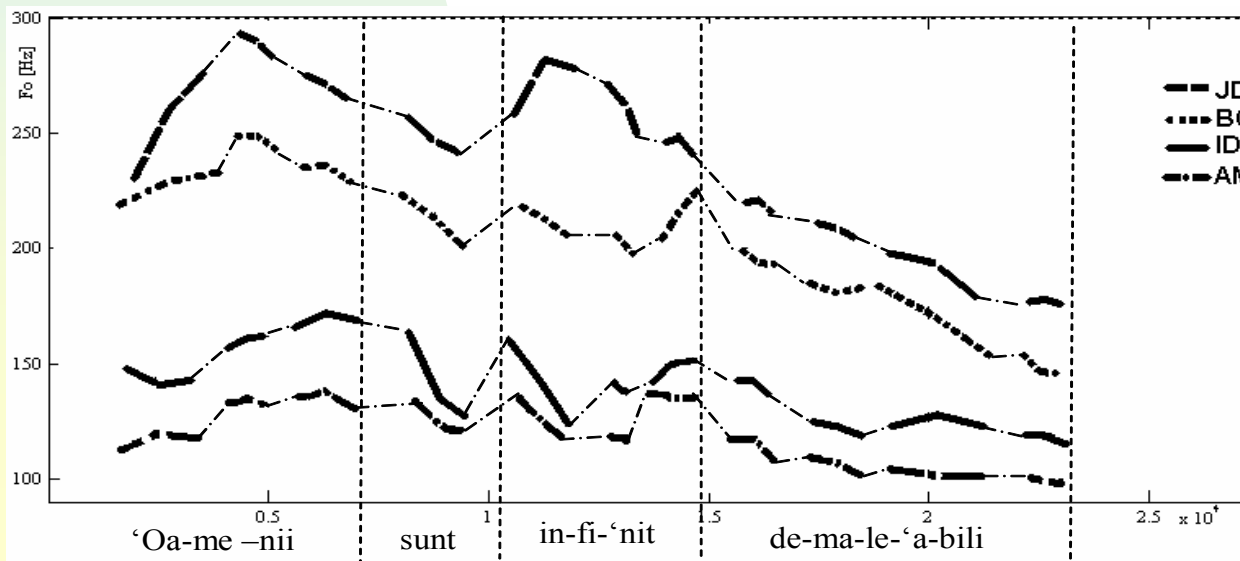
Intonational Variants

“Oamenii sunt infinit de maleabili.”

(People are infinitely flexible)

What they have in common and what it is different:

- The prominence on the subject:
 - **L+H*** pitch accent within the PH pattern
 - PH+f pattern
 - PH+F pattern for the subject+ the auxiliary verb



AM – JD – ID – BC



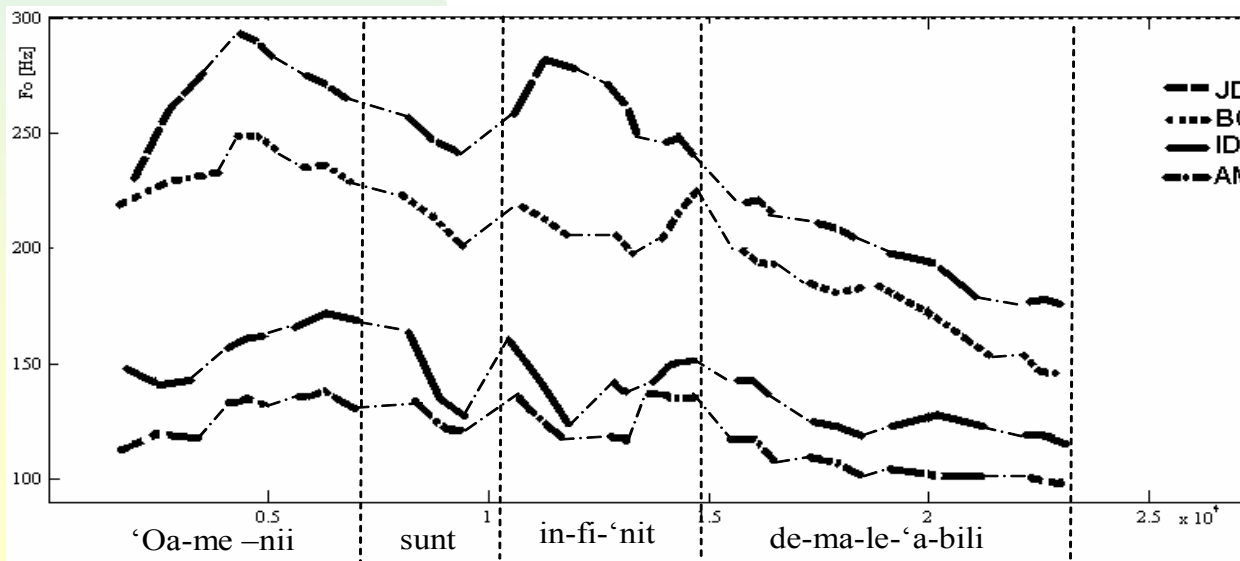
Intonational Variants


“Oamenii sunt infinit de maleabili.”

(People are infinitely flexible)

What they have in common what it is different: :

- The focalization on the predicative subgroup (“*sunt infinit*”):
 - By a minor *f* focus group
 - **By a major *F* focus group**
 - By a PD+f




AM – JD – ID – BC



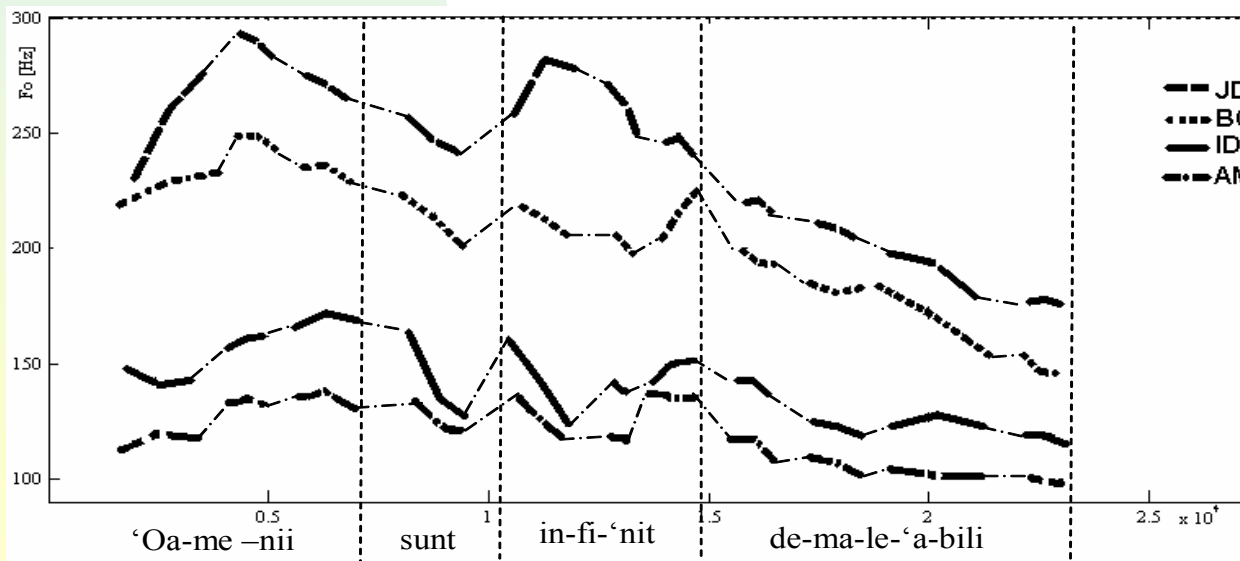
Intonational Variants

“Oamenii sunt infinit de maleabili.”

(People are infinitely flexible)

What they have in common what it is different:

- The last word has:
 - An PO pattern with an L* pitch accent and a continue decreasing down to the last syllable
 - An PO+f pattern with a large slope F0 decreasing during the first syllable and a small range variation after that



AM – JD – ID - BC



1st Intonational Variant

<IP>

<AU Function= " PH" Accent ="L+H*">

<Syl>'oa</Syl>

<Syl>me</Syl>

<Syl>ni</Syl> </AU>

<AUG Function="f" >

<AU Function="PO Accent ="L

<Syl>sunt</Syl></AU>

<AU Function="PH">

<Syl>in</Syl>

<Syl>fi</Syl>

<Syl>'nit</Syl></AU>

</AUG>

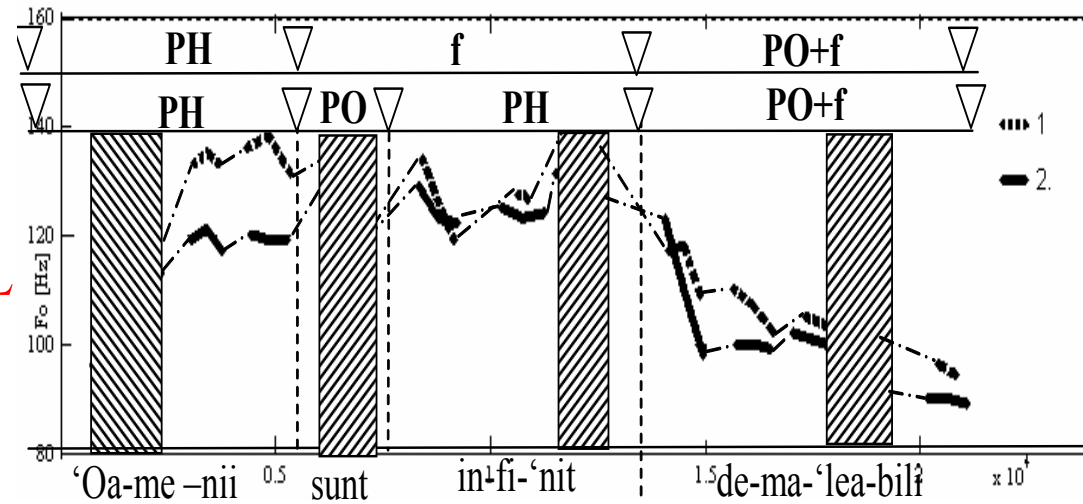
<AU function="PO+f" Accent ="L*">

<Syl>de</Syl>

<Syl>ma</Syl>

<Syl>'lea</Syl>

<Syl>bili</Syl></AU> </IP>



Melodic contours:

- PH/f/PO+f – IP level
- PO/PH –AUG level



2nd Intonational Variant

<IP>

<AUG Function="PH+F" >

<AU Function="PH" >

<Syl>'oa</Syl>

<Syl>me</Syl>

<Syl>ni</Syl> </AU>

<AU Function="PO" Accent="L*" >

<Syl>sunt</Syl>

</AU>

</AUG>

<AU Function="PD+f" Accent="L*" >

<Syl>in</Syl>

<Syl>fi</Syl>

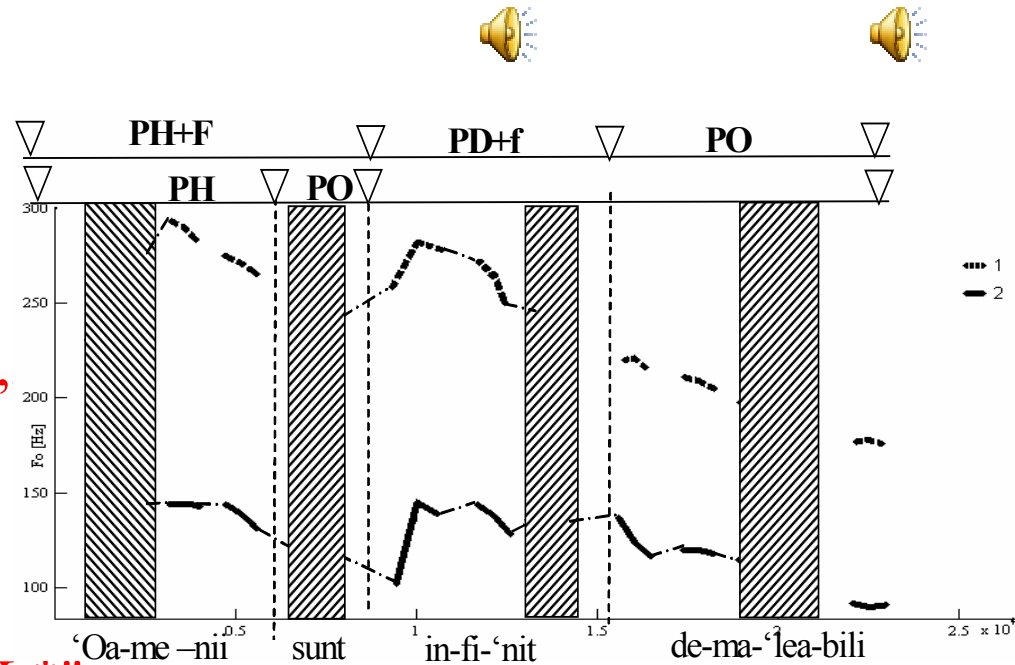
<Syl>'nit</Syl> </AU>

<AU Function="PO" Accent="L*" >

<Syl>de</Syl>

<Syl>ma</Syl>

<Syl>'lea</Syl><Syl>bili</Syl> </AU> </IP>



Melodic contours:

- PH+F/PD+f/PO – IP level
- PH/PO – AUG level



3rd Intonational Variant

<Utt><IP>

<AU Function="PH" Accent="L+H*">

<Syl>'oa</Syl>

<Syl>me</Syl>

<Syl>ni</Syl> </AU>

<AUG Function="F" TonalContrast="s">

<AU Function="PO" Accent="L*">

<Syl>'sunt</Syl></AU>

<AU Function="PH">

<Syl>in</Syl>

<Syl>fi</Syl>

<Syl>'nit</Syl></AU>

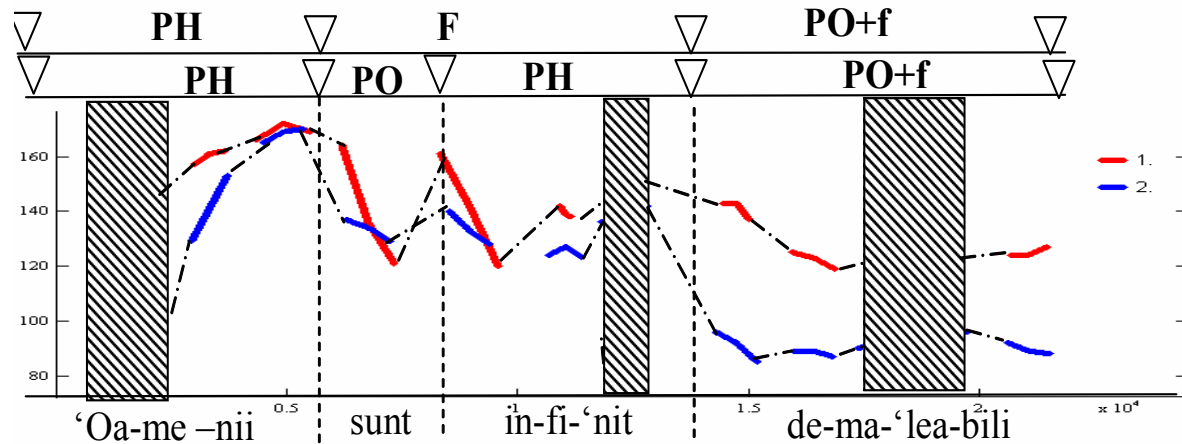
</AUG>

<AU Function="PO+f" Accent="H*">

<Syl>de</Syl>

<Syl>ma</Syl> <Syl>le</Syl>

<Syl>'a</Syl> <Syl>bili</Syl></AU></IP></Utt>



MELODIC CONTOURS:

- PH/F/PO+f – IP level
- PO/PH –AUG level



4th Intonational Variant

<IP>

<AU Function="PH+f" >

<Syl>'oa</Syl>

<Syl>me</Syl>

<Syl>ni</Syl> </AU>

<AUG Function="F" TonalContrast="s">

<AU Function="PO" Accent="L*">

<Syl>'sunt</Syl></AU>

<AU Function="PH">

<Syl>in</Syl>

<Syl>fi</Syl>

<Syl>'nit</Syl></AU>

</AUG>

<AU Function="PO" Accent="L*">

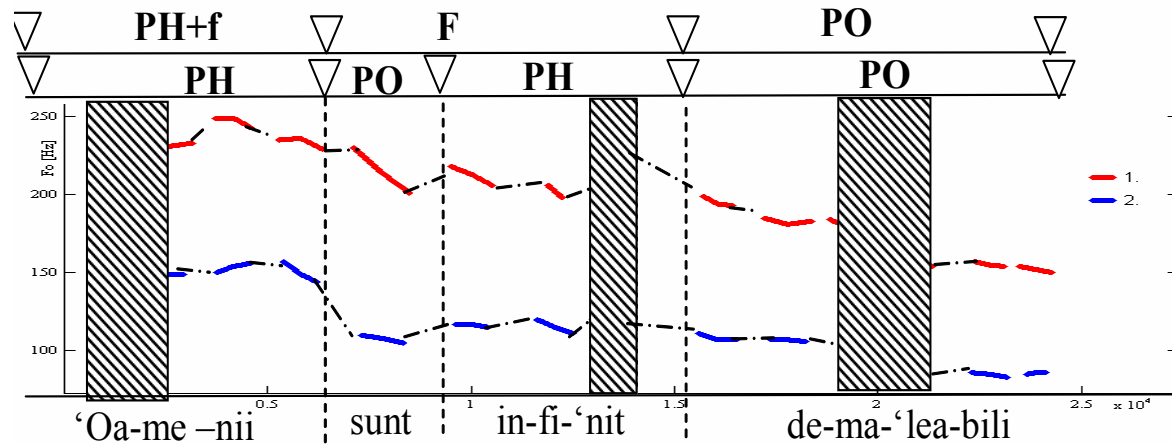
<Syl>de</Syl>

<Syl>ma</Syl><Syl>'lea</Syl><Syl>bili</Syl></AU></IP>

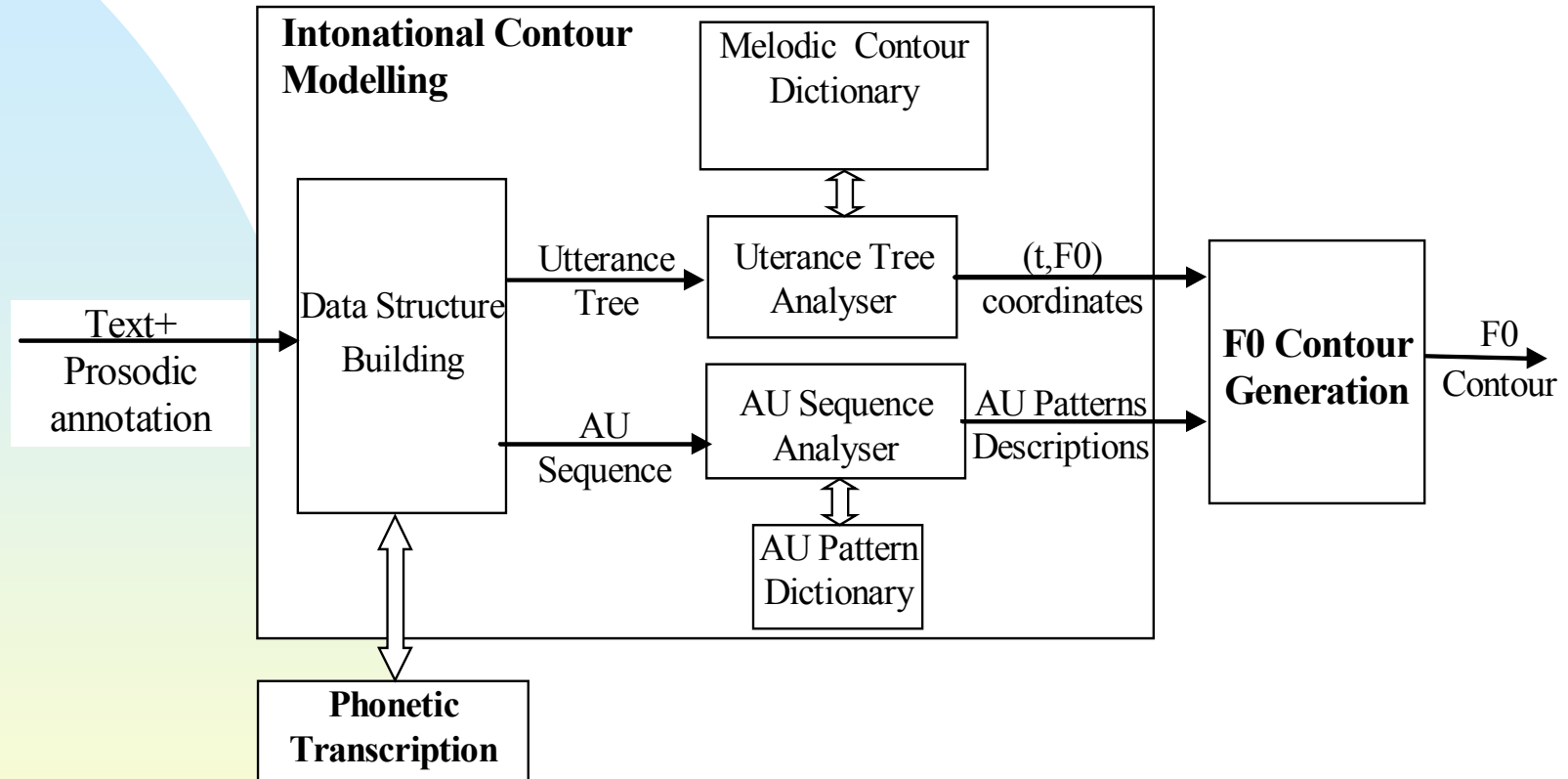


Melodic contours:

- PH+f / F/PO – IP level
- PO/PH –AUG level



The Prosodic Control Module



The AU Pattern Dictionary

{ "6201PH" , 0, 3, 0,0,0,3,0.0, 1,1,0,3,1.0, 2,5,0,1,0.8 } ,
{ "6301PH" , 0, 3, 0,1,0,3,0.0, 2,2,0,3,1.0, 3,5,0,1,0.8 } ,
{ "6401PH" , 0, 3, 0,2,0,3,0.0, 3,3,0,3,1.0 ,4,5,0,3,0.9 } ,
{ "5311PH" , 0.5, 4, 0,1,0,3,0.0, 2,2,0,3,0.7, 3,3,0,1,1.0, 3,5,0,3,0.9} ,
{ "5111PH" , 0.5, 4, 0,0,0,3,0.5, 1,1,0,3,0.7, 2,2,0,3,1.0, 3,5,0,3,0.9} ,
{ "5211PH" , 0.5, 4, 0,0,0,3,0.0, 1,1,0,3,0.7, 2,2,0,1,1.0, 3,5,0,3,0.9} ,
{ "5411PH" , 0.5, 3, 0,4,0,3,0.2, 5,5,1,3,0.7, 5,5,0,3,1.0 } ,

{ "1101PO" , 0, 1, 0,0,0,3,1.0 } ,
{ "1102PO" , 0, 2, 0,0,0,1,0.0, 0,0,0,3,1.0 } ,
{ "1103PO" , 0, 2, 0,0,0,1,0.0, 0,0,0,3,1.0,} ,
{ "1112PO" , 0, 3, 0,0,0,1,0.1, 0,0,1,3,0.3, 0,0,0,3,1.0} ,
{ "1113PO" , 0, 3, 0,0,0,1,0.0, 0,0,1,3,0.3, 0,0,0,3,1.0 } ,
{ "1121PO" , 0, 1, 0, 0,0,3,0.0 } ,
{ "1122PO" , 0, 1, 0,0,0,3,0.0 } ,
{ "1123PO" , 0.5, 1, 0,0,0,3,0.0 } ,



The Melodic Contour Dictionary

{"PH+F/PO/f",1100,1500,850,1000,900,1100. },

{"PH+f/PO/f",1400,1500,900,1200,850,1000. },

{"PH/F/PO",900,1600,1000,1300,850,1100. } ,

{"PH/F/PO+f",900,1600,1000,1300,900,1100. } ,

{"PH/F/PU",900,1600,1000,1300,850,1300. } ,

{"PH/F/PD",900,1500,1000,1300,850,1500. } ,

{"PH/F/PD+F",900,1600,1000,1300, 850,1500. } ,

{"PH+f/F/PO",1500,1600,1000,1300,850,1100. } ,

{"f/PH/F/PO+f",900,1000,900,1600,1000,1300,850,1050. } ,



Conclusions and Future Work

- Using these patterns and their combinations defined as melodic contours, we can drive the prosodic control module in order to reproduce in speech synthesis the natural intonational variants.
- The prosodic macro indications of accentuation and deaccentuation may be transformed into microprosodic descriptions by functional label + attribute sequences. The microprosodic descriptions can generate different prominence and different position of the focus events.
- The main problem for building a prosodic predictor in terms of our proposed functions, is to identify different morphologic unit sequences that correspond to prosodic units with certain melodic contours, in certain morfo-syntactic contexts.



Romanian Speech Synthesis Examples

O să mă întorc **marțea** viitoare



Tocăm **bine** ceapa/ **Tocăm bine** ceapa,



Care pare **a fi** problema?



Lui John nu-i place **mașina**.



Nu-l place **acum** pe John?



Exact **pe John** nu-l place **de loc**.



Lui Winston **i se trezără niște** vagi amintiri.



Helen, **îl cunoaște** pe John?



Pe John nu-l place **de loc**.



Cu asemenea privire cucerești **pe oricine**



Romanian Speech Synthesis Examples

Am luat **ora de întâlnire când** se întunecă afară.



Până primești **scrisoarea** situația **se va fi limpezit.**



Săptămâna asta trebuie să **rezolvăm** problema.



Tu cu mine n-o să mai ai vreodată probleme.



Tu o să mai ai probleme./Tu o să mai ai probleme.



Îl place pe John?



Exact pe John nu-l place de loc.



Vă mulțumesc, **domnule președinte** pentru atenția pe care **ați acordat-o** problemelor pe care **vi le-am prezentat.**



Romanian Speech Synthesis Examples

Mama cântă.



Mama **cântă**.



Mama cântă.

