MELODIC ANALYSIS OF SPEECH METHOD (MAS) APPLIED TO SPANISH AND CATALAN

Dolors Font-Rotchés, Francisco José Cantero Serena
Applied Phonetics Laboratory, Universitat de Barcelona
dolorsfont@ub.edu & cantero@ub.edu

Abstract

In this paper, we present the *Melodic Analysis of Speech* method (MAS) that enables us to carry out complete and objective descriptions of a language's intonation, from a phonetic (melodic) point of view as well as from a phonological point of view. It is based on the acoustic-perceptive method by Cantero (2002), which has already been used in research on prosody in different languages. In this case, we present the results of its application in Spanish and Catalan.

KEY WORDS: melodic analysis, spontaneous speech, intonation, phonetics, Spanish, Catalan

Resum

En aquest article presentem el mètode *Anàlisi Melòdica de la Parla*, el qual permet fer descripcions completes i objectives de l'entonació d'una llengua, tant des del punt de vista fonètic (melòdic) com fonològic. Està basat en el mètode de base acusticoperceptiva de Cantero (2002), el qual ja s'ha utilitzat en investigacions sobre la prosòdia de diferents llengües. En aquest cas, presentem els resultats de la seva aplicació al castellà i al català.

PARAULES CLAU: anàlisi melòdica, parla espontània, entonació, fonètica, castellà, català

Resumen

En este artículo presentamos el método *Análisis Melódico del Habla*, que permite hacer descripciones completas y objetivas de la entonación de una lengua, tanto desde el punto de vista fonético (melódico) como fonológico. Está basado en el método de base acústico-perceptiva de Cantero (2002), que ya se ha utilizado en investigaciones sobre la prosodia de diversas lenguas. En este caso, presentamos los resultados de su aplicación al castellano y catalán.

PALABRAS CLAVE: análisis melódico, habla espontánea, entonación, fonética, castellano, catalán

1. Introduction

The acoustic-perceptive analysis method, Melodic Analysis of Speech (MAS), that we aim to describe, expounded in Cantero (2002), has been experimented with positive results in several papers on Spanish intonation (Cantero, 2007; Cantero *et al*., 2002, 2005) and Catalan intonation (Font-Rotchés 2005, 2007, being printed; Font-Rotchés *et al*. 2002), as well as in studies on the acquisition of Spanish by Chinese people, speakers of a tonal language (Cortés, 2000, 2004; Liu *et al.*, 2002; Liu, 2005) and the acquisition of Catalan by Hungarian learners (Pálvölgyi, being printed). We believe that it is a valid intonation analysis method, for two main reasons:

- It offers an exclusively phonic criterion in speech melody segmentation, independent of any other level of analysis, which can be used to analyse even spontaneous and genuine speech analysis.

- It presents an acoustic data processing system that enables us to obtain the relative values that form the melodies in order to compare and classify them, to reproduce them exactly, experiment with them by using voice synthesis, subject them to perception analysis and make linguistic generalisations.

This enables us to use many speakers and a large number of contours with the purpose of establishing language intonation patterns with precise values to be applied in several areas, such as language teaching, voice retraining, speech disorders, voice synthesis, speech recognition and in dialogue systems among others.

The method we present has been elaborated from Cantero's proposal (2002) with some considerations that have emerged from Font-Rotchés's research. We have used it for describing intonation, firstly, in Spanish and Catalan spontaneous speech and, secondly, in television advertisements (Font-Rotchés, 2009; Font-Rotchés & Machuca, being printed) in order to apply it to language teaching (for teachers, narrators, presenters, etc.) and identify the prosodic advertising techniques devised to convince the audience.

2. Identification of melodic units in a corpus of spontaneous speech

Intonation studies are often made from a corpus of "laboratory speech" (prepared sentences to be read or, at most, induced by the researcher) in order to identify and separate the utterances whose intonation is going to be described. This dependence on other analysis levels (syntactic or pragmatic) comes from theoretical patterns that have not defined the speech phonic units clearly enough, but are limited to describing the phonic properties of the grammatical units (such as, syntactically well-formed sentences).

All this leads to a type of analysis that doesn't describe the linguistic reality itself, but a linguistic reality created solely in the laboratory, which obviously is not representative of the complete language.

From our papers we have developed a theoretical pattern that allows us to identify exactly the speech phonic units, regardless of any other analysis level. This enables us to describe the intonation of a speech corpus that is neither prepared nor induced and is completely removed from the researcher's intervention, with anonymous speakers and with genuine spontaneous speech. With such corpus, we cannot control either the utterances or their shape in any way; therefore, our intervention is restricted to "choosing" the utterances, with no need to influence them.

Our theoretical model is based on the concept of *phonic hierarchy*, according to which speech is made up of linked phonic units with a very hierarchical structured: the syllable, the rhythmic group (or phonic word) and the phonic group.

The intonation analysis unit is the phonic group (with a particular melody); the rhythm analysis unit is the rhythmic group (whose melody is also relevant within the contour); and the melody analysis unit is the *tonal segment* (that is, the relative tonal value of the syllabic core: the vowel). Each vowel constitutes a tonal segment, except for the tonic vowels, which can constitute a *tonal inflection* of two (or more) tonal segments: as occurs with the sentence accent (or *syntagmatic accent*), which is the core of the phonic group and therefore the core of the melody: the final inflection (FI) of the contour.

In order to segment spontaneous speech into melodic units we use the formal criterion as a starting point: the presence of a tonal inflection that defines the phonic group.

With a spontaneous speech corpus, initially choosing the utterances that coincide with a turn to speak in the dialogue (normally short and easy to identify) is advisable, until the researcher has acquired enough confidence in identifying phonic groups and in defining speech melodies.

Once the melodic units have been identified (phonic groups), they must be treated as autonomous utterances (although always clearly identifying the tonal context in which each has been produced, in order to relativise it later, if necessary). For example, from an original recording (in DAT or in video) they can be digitalised as independent files. This will make the analysis task much easier, as we will be able to deal with each contour independently from the others, as an autonomous speech unit.

Each contour will sometimes have more or less defined syntagms and sometimes more or less grammatically complete sentences (as we are working with spontaneous oral language): but under no circumstances do we depend on these grammatical units for the analysis, because the units that we analyse are phonic units (whether they contain whole grammatical units or not). We believe that the grammatical units are the ones that locate and adapt to a melodic container, the true organiser of spoken language: what we call *prelinguistic intonation*.

3. Acoustic phase: determination and standardisation of relevant frequency values

The first step in the melodic analysis of an utterance consists in differentiating the relevant frequency values (the tonal segments) from the irrelevant values (the F_0 of the sonorous consonants and of the glides); once the utterances have been separated we establish the F_0 value of each vowel, of each tonal segment.

Using a reliable acoustic analysis instrument (Praat analysis and synthesis software), we identify the vowels (for example, guided by a sonogram) and we establish their central value (the average of the F_0 values of the vowel if it is stable enough).

When a stressed vowel contains a tonal inflection, we have to establish a value of two tonal segments constituting the inflection (or three segments if it is a circumflex inflection). These values are calculated from the initial or final stable values, or from extreme values of the inflection (if there is no tonal stability). See the Spanish example of *Entra**mos**?* "Shall we go in?" and the Catalan example of *el Gau**dí**?* (a family name):

Utterance	en	tra	**mos**	**mos***
Hertz	133	158	**186**	**295**

Utterance	el	Gau	**dí**	**í***
Hertz	243	202	**221**	**329**

On other occasions, the tonal inflection begins with a stressed vowel and ends with an unstressed vowel, for example, in the final inflections that coincide with a word with the stress on the penultimate syllable. See the Spanish example *¿Tienes **hijos**?* "Do you have children?" and the Catalan example *Has comprat **oli**?* "Have you bought olive oil?":

Utterance	Tie	nes	**hi**	**jos**
Hertz	265	292	**369**	**481**

Utterance	Has com	prat	**o**	**li**
Hertz	243	202	**221**	**329**

When the final inflection ends with a tonal vowel followed by a nasal or a lateral consonant, this consonant usually constitutes the last total segment of the inflection. See the Spanish example *¿Te parece **bien**?* "Is that OK?" and the Catalan example *Fa molts **anys**?* "Was it many years ago?":

Utterance	Te	pa	re	ce	**bie**	**n**
Hertz	94	98	113	141	**137**	**228**

Utterance	Fa	molts	**a**	**nys**
Hertz	198	182	**239**	**464**

The absolute values obtained in this first analysis phase (extraction and determination of F_0 of the tonal segments, measured in Hertz) do not yet constitute the contour melody because they are only rough data, which needs to be suitably processed.

The second step in the melodic analysis is, therefore, the standardisation of the frequency data.

The contour melody is not only the simple succession of frequency values, but their relationship. In other words, the melody is not a succession of absolute values, but a succession of relative values: a succession of intervals.

Therefore, data in Hertz should be made relative in order to describe the contour melody. The melody that constitutes the succession of values 100Hz–200Hz is not the same as one of values 200Hz-300Hz, although the difference is exactly the same: 100Hz in both cases. In absolute terms, the difference is the same (the same number of Hz), but not in relative terms: the tonal interval is different. In the first case, there is a difference of 100% but in the second case, there is a difference of 50%. Therefore, between 100Hz and 200Hz we find the same interval as between 200 Hz and 400 Hz (100%), and between 200Hz and 300Hz we find the same interval as between 300Hz and 450Hz (50%).

Some authors express this interval using the semitone (st) as a unit, which has the advantage of being a logarithmic unit used to measure musical intervals. In speech melodies, however, using semitones means working with a somewhat complex standardisation formula and with intervals usually expressed in decimals (which does not occur in music, where the intervals are always the same and the minimum unit is the semitone).

The advantage of using percentages is that they are much more intuitive, because they allow us to express a logarithmic phenomenon in a linear way: we calculate (with an easy rule of three) the percentage of the variation of each absolute value with regard to the previous value, expressing the rise as a positive percentage and the fall as a negative percentage:

Utterance	Te	pa	re	ce	bie	n
Hertz	94	98	113	141	137	228
Percentage	**100%**	**4.3%**	**15.3%**	**24.8%**	**-2.8%**	**66.4%**

Utterance	Fa	molts	a	nys
Hertz	198	182	239	464
Percentage	**100%**	**-8.1%**	**31.3%**	**94.1%**

With this calculation we obtain the relative values of each contour that make up the algorithm: 100% 4.3% 15.3% 24.8% -2.8% 66.4% (Spanish utterance) and 100% -8.1% +31.3% +94.1% (Catalan utterance). This algorithm is the expression of the melody.

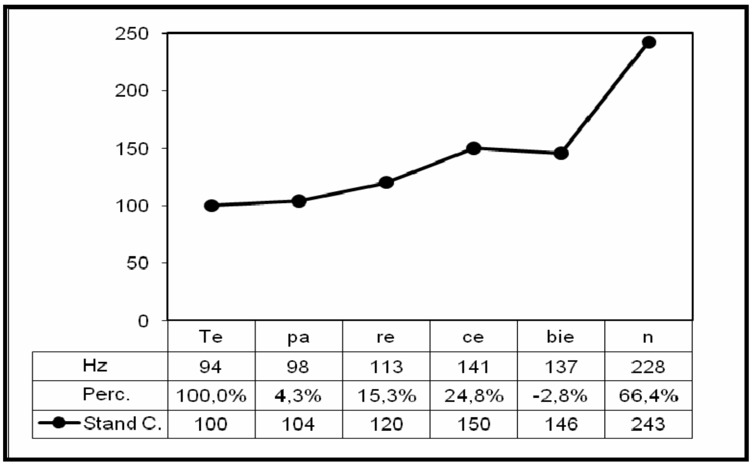
In order to draw the graphic representation of these melodies, we convert the percentages obtained into standard values: starting, for example, with the value 100 (an arbitrary value). By applying (see Catalan utterance) the fall percentage seen in the second segment, 8.1%, we will obtain a second value, 92; to this we apply the rise percentage of the third segment, 31.3% and we will obtain the third value, 121; and so on until the last segment.

From this process, we obtain the following values:

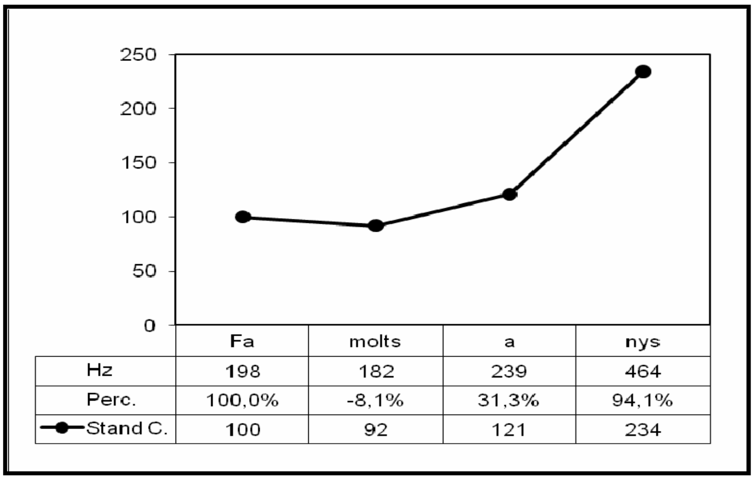
Utterance	Te	pa	re	ce	bie	n
Hertz	94	98	113	141	137	228
Percentage	100%	4.3%	15.3%	24.8%	-2.8%	66.4%
Standard curve	**100**	**104**	**120**	**150**	**146**	**243**

Utterance	Fa	molts	a	nys
Hertz	198	182	239	464
Percentage	100%	-8.1%	31.3%	94.1%
Standard curve	**100**	**92**	**121**	**234**

Using the standard values of each utterance, we can plot the graphic representation of each contour in order to compare them. See the two following graphs of the analysed contours: graph 1 (Spanish utterance) and 2 (Catalan utterance).



Graph 1. *Standard curve of* ¿Te parece bien? "*Is that OK?*".



Graph 2. *Standard curve of* Fa molts anys? "*Was it many years ago?*".

As we will see in the perception phase, the standard values of an intonative contour are validated in the voice synthesizer by ensuring that it has the same melody as the original, but with a different voice tessitura.

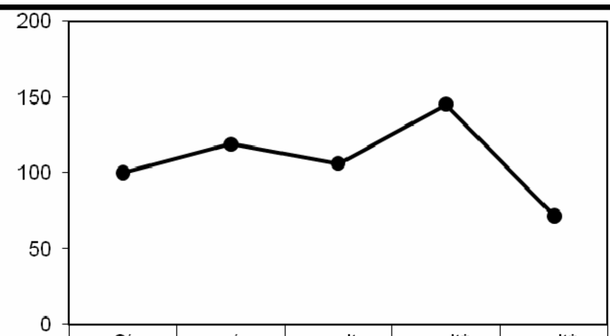
With the application of this first phase of the method, or acoustic phase, we obtain the contour standardisation, which can now be compared and classified, regardless of the age, gender or any other characteristic of the speaker, as all the micromelodic variations have been removed and the values have been standardised (arbitrary value 100).

We do not have to worry about the amount of corpus speakers, because by applying this method all the contours we obtain, whatever they are, are comparable.

This process does not standardise the length of the different contours but isolates and replaces them with the number of tonal segments and their relative value, which is the only relevant data in a melodic analysis. The differences between two contours of different length should be understood from a melodic perspective, regardless of whether the melody contains a word or many words, whether it has a longer or shorter utterance (or whether it has a more or less grammatically correct utterance), etc.

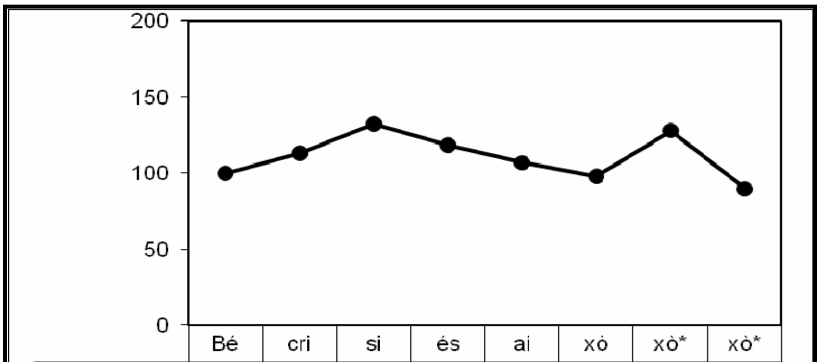
In fact, one of the virtues of our melodic analysis method is that it is exclusively melodic.

For example, the representation of two very different Catalan contours, with an identical rising–falling final inflection, characterising both contours as belonging to the same melodic class or pattern (graphs 3 and 4). Regardless of their length or the meaning of each utterance, the melody is, from a strictly melodic point of view, the same.



	Sí	sí	molt	molt*	molt*
Hz	281	334	298	408	200
Percent.	100,0%	18,9%	-10,8%	36,9%	-51,0%
Stand. C.	100	119	106	145	71

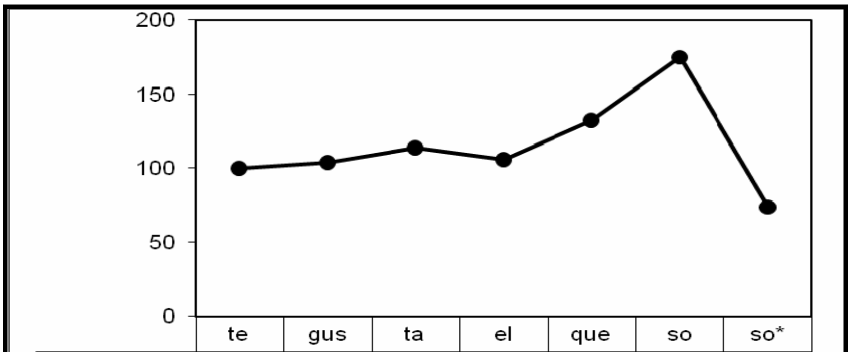
Graph 3. *Standard curve of the utterance* Sí, molt! "*Yes, a lot*".



	Bé	cri	si	és	ai	xò	xò*	xò*
Hz	308	348	408	365	329	301	394	276
Percent.	100,0%	13,0%	17,2%	-10,5%	-9,9%	-8,5%	30,9%	-29,9%
Stand. C.	100	113	132	119	107	98	128	90

Graph 4. *Standard curve of the utterance* Bé, crisi és això! *"Well, crisis is that!"*.

We can also find this rising-falling melody in Spanish. See the next example *¿Te gusta el queso?* "Do you like cheese?":



	te	gus	ta	el	que	so	so*
Hz	205	213	233	217	271	359	151
Percent.	100,0%	3,9%	9,4%	-6,9%	24,9%	32,5%	-57,9%
Stand. C.	100	104	114	106	132	175	74

Graph 5. *Standard curve of the utterance* ¿Te gusta el queso? *"Do you like cheese?"*

4. Perception phase: analysis validation and data interpretation

After the acoustic phase, we can check the validation of our results with a series of perception tests submitted to the listener's judgement from an exact copy (thanks to voice synthesis) of the analysed sentences.

We use the *Praat* programme to obtain this synthesised copy (with the PSOLA method), from which we erase all original data and replace it with our standardised data. This way we can check that the melodic analysis was correct and that it reflects the original melody, without micromelodic variations and with the values standardised.

The following step is suitably interpreting the melody obtained: extracting the relevant melodic data that enable us to interpret the contour, for example, a phonologic interpretation.

In our theoretical pattern, we distinguish between the contours' melodic features (which constitute the intonation's phonetic level) and the phonological features (which allow us to establish the *tonemes* or phonological intonation units).

The phonological features that we study are: /± interrogative/ /± emphatic/ and /± suspended/, whose combination allows us to characterise the language *tonemes* (Cantero, 2002). These phonological features are sufficient for classifying all the contours in a corpus of 136 Spanish utterances produced by 57 people (Cantero *et alii,* 2002, 2005) and another of 580 Catalan utterances produced by 160 people (Font Rotchés, 2005, 2007).

The melodic features are, in turn, the characteristics of the contour's functional elements: *anacrusi*, the *first peak* (*1st p.*), the *body* (or *declination*) and the *final inflection (FI)*. By *anacrusi* we understand the unstressed syllables prior to the first stressed vowel in the contour, called *first peak* and by *body,* the syllables from the first peak to the last stressed vowel in the contour, from which the *final inflection* begins.The description of these elements (especially the description of the *final inflection*) allows us to define the contour melody. We can also establish the typical melodic patterns in our corpus (as "typical contours" of the *tonemes*) and their dispersion margins.

5. Melodic Patterns and its applications

Applying this method we have established twelve melodic patterns in Spanish and eight in Catalan, described in this section.

5.1 Spanish Melodic Patterns

- Neutral Pattern or Pattern I: This is characterised as /-interrogative -emphatic -suspended/ and is representative of all the contours of this language that begin with an optional rise until first peak (first stressed syllable) with a 40% maximum, continuing with a gentle and constant fall until the last stressed syllable and ending with a final inflection, which can have a rise of up to 10%-15% or a fall of up to 30% (see figure 1).

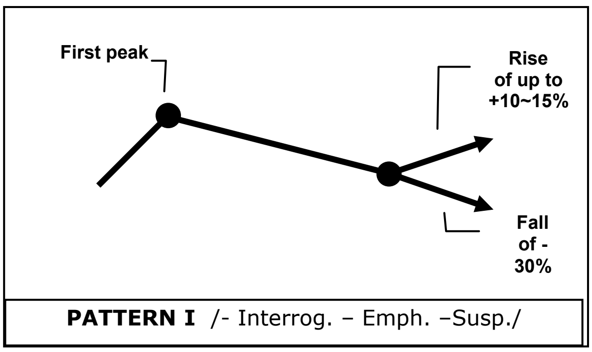


Figure 1. *Spanish neutral pattern according to MAS.*

- Interrogative Patterns: characterised as /+interrogative -emphatic -suspended/. We identified four interrogative patterns: two with a rising final inflection (patterns II and III) and two others with a circumflex rising-falling final inflection (patterns IVa and IVb). See graphs in figure 2:

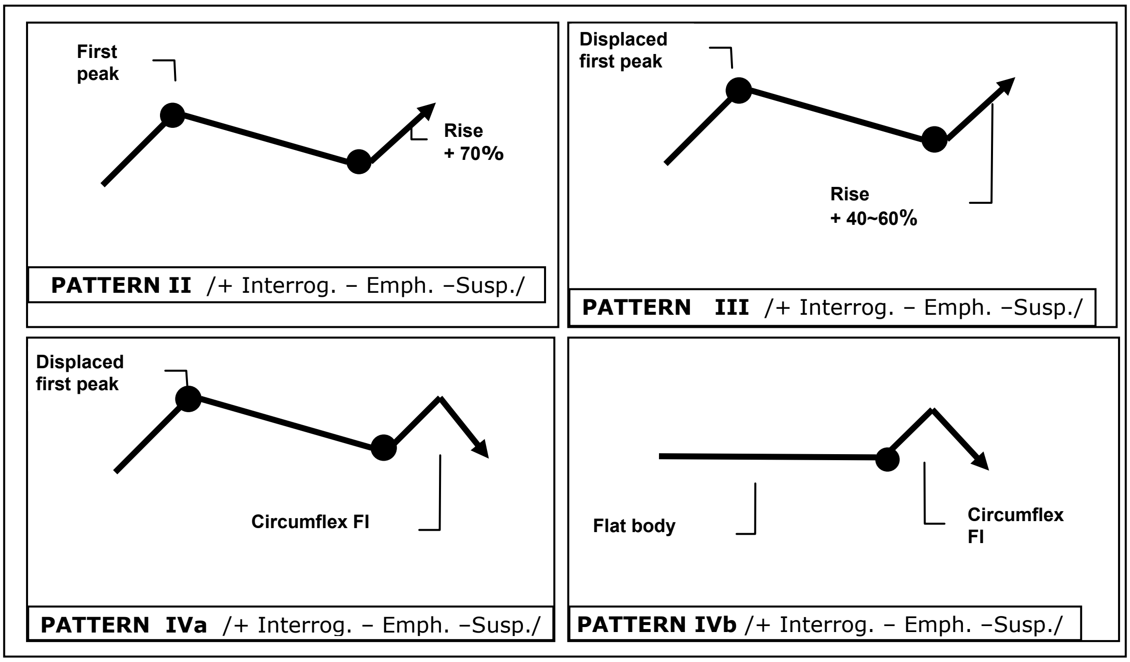


Figure 2. *Spanish interrogative patterns according to MAS.*

- Suspended Patterns: characterised as /-interrogative -emphatic +suspended/. We defined three melodic patterns: two with a rising final inflexion (VIa and VIb) and another with no final inflection (V) (see Figure 3):

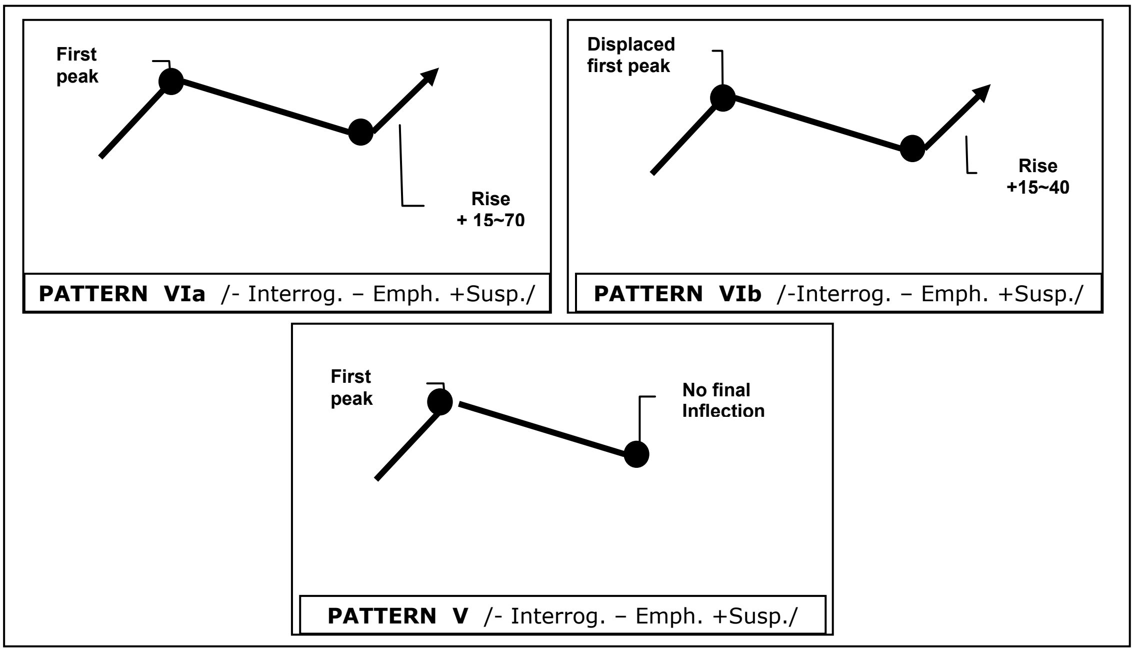


Figure 3. *Spanish suspended patterns according to MAS.*

- Emphatic Patterns: characterised as /-interrogative +emphatic -suspended/. Among other emphatic melodies in our corpus, the most frequent are given in figure 4 (patterns VII, VIII, IX and X).

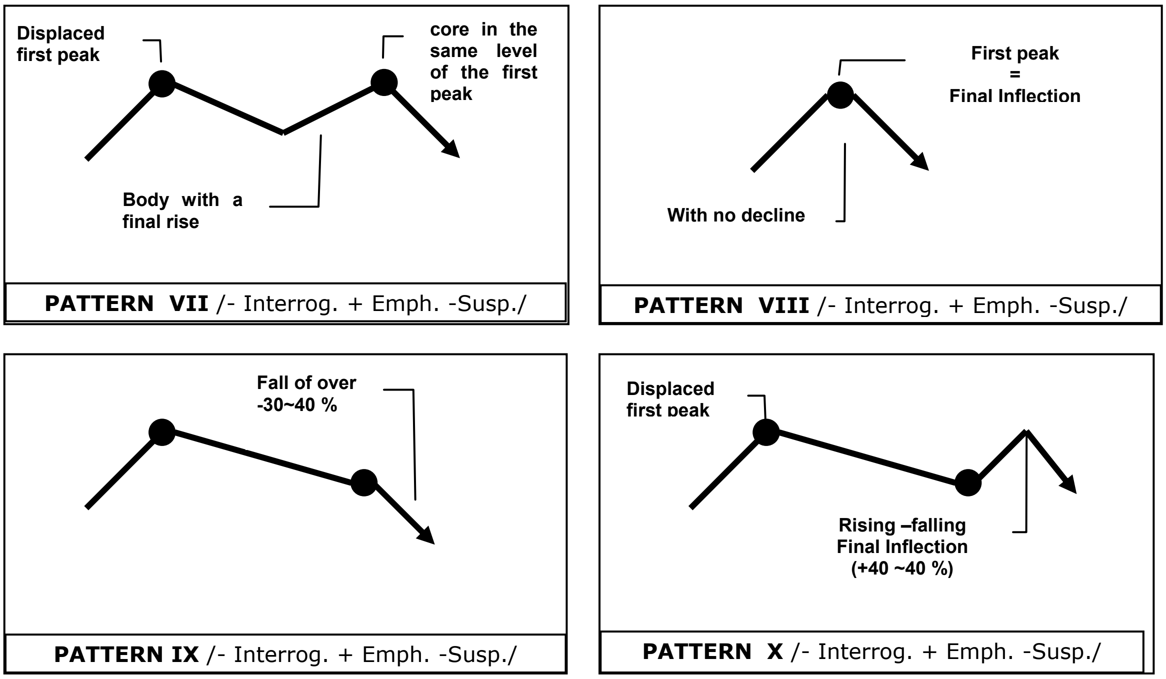


Figure 4. *Spanish emphatic patterns according to MAS.*

5.2 Catalan Melodic Patterns

- Neutral Pattern or Pattern 1: This is characterised as /-interrogative -emphatic -suspended/ and is representative of all the contours of this language that begin with an optional rise until the first peak (first stressed syllable) with a 40% maximum, continuing with a gentle and constant fall until the last stressed syllable and ending with a final inflection, which can have a rise of up to 10% or a fall of up to 40% (see figure 5). This pattern is similar to the Spanish neutral pattern.

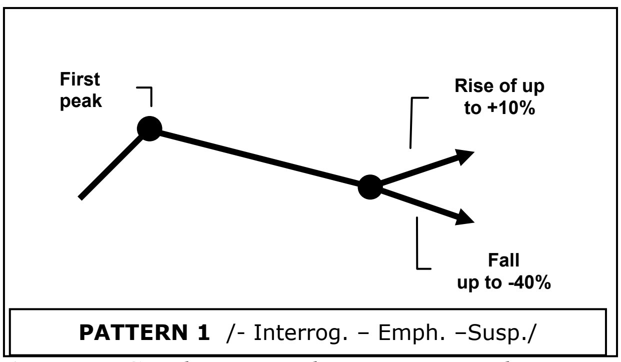


Figure 5. *Catalan neutral pattern according to MAS.*

- Interrogative Patterns: characterised as /+interrogative/. We identified three interrogative patterns: one is /-emphatic/ (pattern 3) and can have a rise of more than 80%; the second (pattern 7) is /+emphatic/ and has the core at the highest point with a rise of over 50% and the third (pattern 8) is also /+emphatic/ but with a circumflex falling-rising final inflection, which should have a rise of over 120% in order for the melody to be identified as /+interrogative/ .

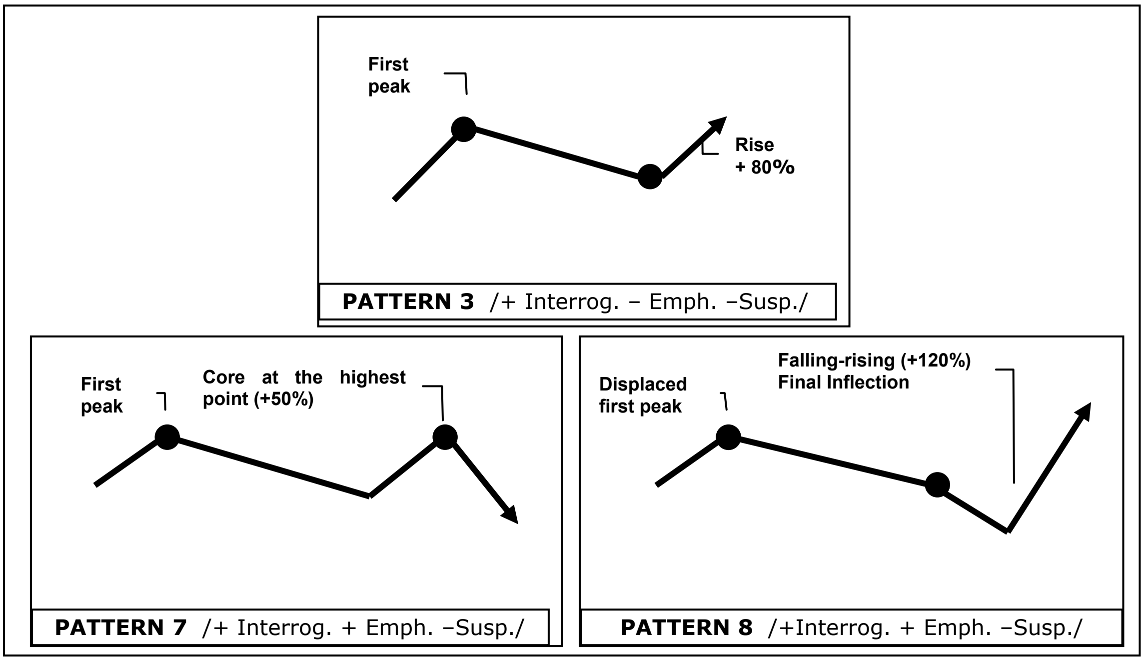


Figure 6. *Catalan interrogative patterns according to MAS.*

- Suspended Patterns: characterised as /-interrogative –emphatic +suspended/. We defined one melodic pattern and one suspended variant which also exist in Spanish, patterns VIa and V, respectively. One has a rising final inflexion of between 10% and 80% (pattern 2) and the other has no final inflection, which we consider to be a suspended variant of the other patterns (see Figure 7).

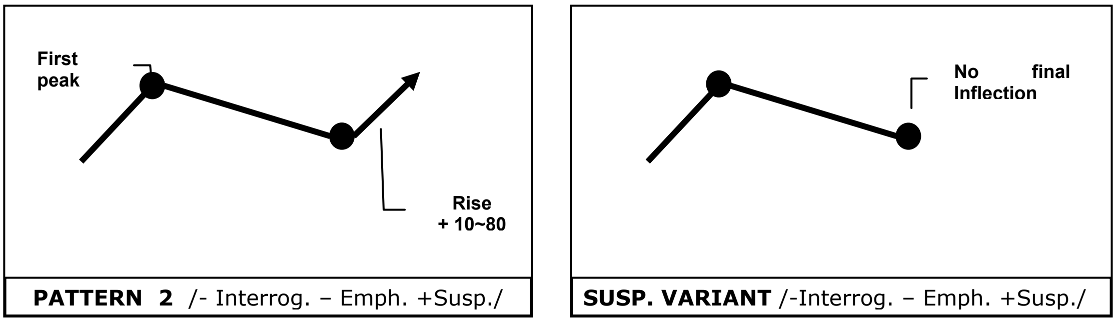


Figure 7. *Catalan suspended pattern and variant according to MAS.*

- Emphatic Patterns: characterised as /-interrogative +emphatic/. We identified three melodic patterns. One, pattern 4, is /-suspended/ and has the core at the highest point between 10% and 50% of rise; the second, pattern 5, is also /-suspended and has a circumflex rising-falling final inflexion. Thirdly is pattern 6 characterised as not only emphatic but also /+suspended/ and with a circumflex falling-rising final inflection which can have a rise of up to 120% (see figure 8).

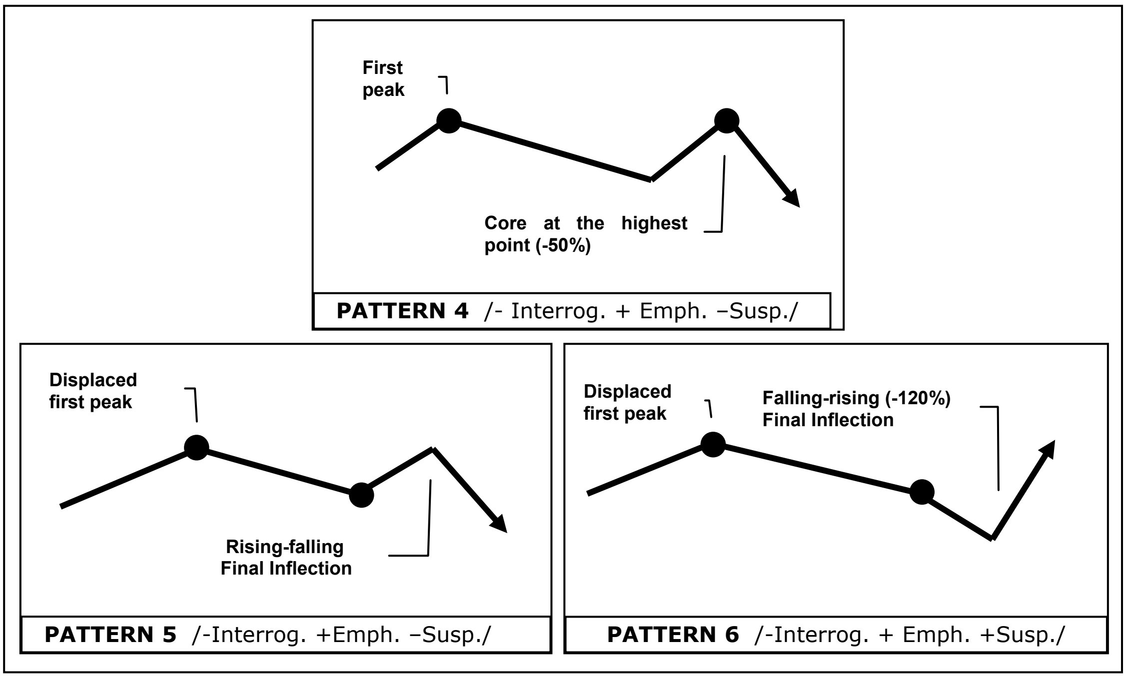


Figure 8. *Catalan emphatic patterns according to MAS.*

According to the method, the melodic pattern is not a mere representation of a line with rises and falls but an abstraction of the real way of speaking, representative of many melodies with the melodic features included in their dispersion margins (clearly defined and objectively quantified).

The pattern dispersion margins allow us to understand the numerous language contours that have a similar final inflection. Speakers use these variations to transmit intentions, feelings and other expressive contents, or to clearly show the social-cultural or dialectal differences (although their phonological value is still the same).

However, the contours can have other melodic features (steep rises or falls in the body, flat declination, lack of final inflection, etc.) that constitute alterations in the typical melodic features of the pattern and involve a phonological sign change. An /-interrogative/ contour can thus become /+interrogative/, an /-emphatic/, /+emphatic/, etc. These features are melodic variations that should be defined by formal comparison and should also be validated perceptively.

If a falling pattern contour characterised as /-interrogative -emphatic -suspended/ has, for example, a very steep declination or a rise of above 30% in any word in the body, its sign changes and therefore becomes /+emphatic/; if it does not have a final inflection because this was cut when the utterance was produced, it also changes its sign: becoming /+suspended/.

Consequently, this method allows us to define the melodic patterns of a language with established dispersion margins within precise boundaries and it also allows us to establish, with the same exactitude, the variations that each pattern can have.

The melodic patterns established in Spanish and Catalan spontaneous speech are the first steps to researching all of the intonation aspects of these languages, such as emphasis, pragmatic functions, differences and similarities between dialects and language varieties, among others. In addition they can be used to describe other languages or interlanguages of speakers learning a foreign language and their prosodic features.

6. References

CANTERO, F. J. (2002): *Teoría y análisis de la entonación*, Barcelona, Ed. Universitat de Barcelona.

CANTERO, F.J. (2007): "Patrones melódicos del español en habla espontánea", *III Congreso Internacional de Fonética Experimental.* Santiago de Compostela: Xunta de Galicia, 181-194.

CANTERO, F.J.; ALFONSO, R.; BARTOLÍ, M.; CORRALES, A.; VIDAL, M. (2005): «Rasgos melódicos de énfasis en español». *Phonica*, *vol.1*. www.ub.edu/lfa

CANTERO, F. J.; DE ARAUJO, M.A.; LIU, Y.H.; ZANATTA, A. (2002): "Patrones melódicos de la entonación interrogativa del español en habla espontánea", *II Congreso de Fonética Experimental.* Sevilla: Universidad de Sevilla, 118-123.

CORTÉS, M. (2000): *Adquisición de la entonación española por parte de hablantes nativos de chino*. Universitat de Barcelona. PhD publicada en microforma.

CORTÉS, M. (2004): "Análisis acústico de la producción de la entonación española por parte de sinohablantes", *Estudios de Fonética Experimental, XIII*.

FONT-ROTCHÉS, D. (2005): *L'entonació del català. Patrons melòdics, tonemes i marges de dispersió*. PhD. Barcelona: Universitat de Barcelona. http://www.tesisenxarxa.net/TDX-0802106-114003

FONT-ROTCHÉS, D. (2007): *L'entonació del català*. Barcelona: Publicacions de l'Abadia de Montserrat.

FONT-ROTCHÉS, D. (2009): "Èmfasis melòdics als eslògans publicitaris", *XIV Col·loqui Internacional de Llengua i Literatura Catalanes*, Budapest 2006. Barcelona: Publicacions de l'Abadia de Montserrat, vol. II, 151-165.

FONT-ROTCHÉS, D. (being printed): "La entonación de las interrogativas absolutas. Publicidad versus habla espontánea", *XXV Congrès International de Linguistique et de Philologie Romanes*, Innsbruck 2007. Berlin: De Gruyter.

FONT-ROTCHÉS, D.; CANALS, A.; ESTER,G.; HERMOSO, A. & F.J. CANTERO (2002): "Patrones melódicos de la entonación interrogativa del catalán en habla espontánea", *Actas del II Congreso de Fonética Experimental.* Sevilla: Universidad de Sevilla, 192-197.

FONT-ROTCHÉS, D.; MACHUCA, M. J. (being printed): "Melodía y eslóganes publicitarios". *Quaderns de Filologia*. València: Universitat de València.

LIU, Y.H. (2005): *La entonación del español hablado por taiwaneses.* Biblioteca Phonica, 2. www.ub.es/lfa

LIU, Y.H. & CANTERO, F.J. (2002): "La entonación prelingüística del español hablado por taiwaneses: establecimiento de un corpus", *Actas del II Congreso de Fonética Experimental.* Sevilla: Universidad de Sevilla, 238-242.

PÁLVÖLGYI, K. (being printed): "Fenòmens tonals conflictius en la llengua catalana i l'hongaresa investigats en la interllengua hongaresa-catalana", *Actes del Catorzè Col·loqui Internacional de Llengua i Literatura Catalanes,* Budapest 2006, Publicacions de l'Abadia de Montserrat.