On the use of melodic patterns as prosodic correlates of emotion in Spanish

Juan-María Garrido[1], *Yesika Laplaza*[2], *Montserrat Marquina*[2]

[1] Computational Linguistics Group (GLiCom), Department of Translation and Language Sciences, Pompeu Fabra University, Barcelona, Spain
[2]Speech and Language Group, Barcelona Media Centre d'Innovació, Barcelona, spain
juanmaria.garrido@upf.edu, yesika.laplaza@barcelonamedia.org, montse.marquina@barcelonamedia.org

Abstract

This paper presents an acoustic and perceptual analysis of the use of melodic patterns as prosodic correlates of the expression of emotional states in Spanish. The data of the acoustic analysis of a corpus of neutral and emotional declarative sentences in Spanish are presented first, showing that emotional utterances are frequently closed by a 'rise-fall' pattern that is not used in neutral speech. However, the results of the perceptual experiment carried out to test the perceptual relevance of this pattern for the identification of emotions in speech seem to indicate that, although it does contribute to recognize a utterance as emotional, its use is not a sufficient cue by itself.

Index Terms: melody, intonation, emotion, perception, Spanish

1. Introduction

The analysis of the melodic correlates of emotion in Spanish has been mainly focused, as for other languages, on global parameters, such as F0 range or height ([1], [2], [3] among others). However, much less attention has been paid to the use of specific melodic patterns (local F0 movements) as phonetic correlates of emotion. Classical studies such as the one by Navarro [4] suggested that Spanish speakers do not use the same melodic patterns in neutral than in emotional utterances, at least as final (boundary) movements is concerned. More specifically, they reported the use of a 'rise-fall' final pattern as prototypical of emotional speech. Nothing was said about non-final patterns. Since then, no experimental studies have been carried out to validate their hypotheses.

This work presents the results of an acoustic and perceptual study on the use of melodic patterns to express emotional states in Spanish, carried out in the framework of the I3Media project. In the following sections, the results of an acoustic analysis comparing the melodic movements used in both neutral and emotional utterances, using as base material a corpus of emotional and neutral speech recorded during the project by a professional speaker, are presented, and also the results of a perceptual experiment carried out to validate the results of the acoustic analysis.

2. Acoustic analysis

The goal of the acoustic analysis was to determine which are the most frequent melodic patterns used in both neutral and emotional speech, in order to determine if there are differences between both styles.

2.1. Corpus

The material used for this analysis has been extracted from two subcorpora ('neutral' and 'emotional') of a larger corpus of speech recorded by a professional actor for analysis and synthesis purposes. The speaker was selected among six candidates considering, among other criteria, his stills to imitate emotional speech in a realistic manner. The 'neutral' subcorpus contained 1.000 utterances, extracted mainly from newspapers, that the speaker was told to read in a neutral style. The 'emotional' subcorpus included 378 utterances, extracted mainly from literary texts, representing 42 different emotional states (9 items x 42 emotional labels), that the speaker uttered to express the target emotions. The corpus was recorded in a professional studio at Pompeu Fabra University, using a headset microphone to allow the speaker to 'act' as freely as possible.

2.2. Analysis procedure

Both subcorpora were processed using MelAn, the automatic annotation and modelling tool for melodic contours described in [5], which applies the modelling framework of melodic contours proposed at [6, 7]. In this framework, melodic patterns are defined as series of perceptually relevant inflection points occurring in the domain of a stress group (portion of utterance including a stressed syllable and all the unstressed syllables following it). Inflection points are labelled as 'peak' (P, P+) or 'valley' (V, V-) according to the relative height in the F0 range of the speaker, and anchored to specific parts of the container syllable. For example, a pattern labelled as VI0_VI1_PM1 (illustrated in figure 1) is made up of three different inflection points: the first one (VI0) is a 'Valley' (V) point anchored at the vicinity of the beginning (I) of the syllabic nucleus of the stressed syllable (0); the second one is also a 'Valley' (V) point, located near the beginning of the syllabic nucleus of the post-stressed syllable (1); and the last one (PM1) is a 'Peak' (P) point placed around the mid point (M) of the syllabic nucleus of the post-stressed syllable (1).

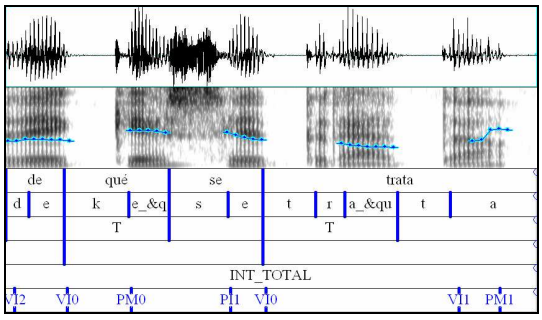

Figure 1. Notation example of two melodic patterns of the utterance '*¿Quiere alguien explicarme de qué se trata?*', uttered by a female speaker.

Using this tool, a list containing the most frequent melodic patterns for each considered condition was automatically obtained. These conditions were defined as the result of the combination of the following variables: style ('neutral' or 'emotional'); number of syllables of the stress group; position of the stressed syllable in the stress group (usually the first one, but not in all cases); position of the pattern in the intonational group ('initial', 'internal', and 'final/non sentence-final' and 'final/sentence-final'); and sentence type ('declarative', 'exclamative', and different type of interrogative sentences). The variable 'emotion' has not been considered, in order to avoid an excessive fragmentation of the data.

2.3. Results

For practical reasons, the results presented here describe only a part of the conditions considered in the full analysis. The following subsections include the data corresponding to two types of patterns, the ones appearing in internal and final (sentence-final) positions, as representative of the patterns related to accent and intonation, respectively. The tendencies observed in these two conditions are similar in the other two analysed conditions ('initial' and 'final/non sentence-final'). Also, only the data obtained from 1-syllable and 2-syllable stress groups in declarative sentences (the most frequent in the corpus) are presented.

2.3.1. Internal patterns

Tables 1 and 2 present the most frequent internal patterns for 1-syllable and 2-syllable stress groups in emotional and neutral speech, respectively.

Syllables	Stressed syllable	Pattern	Appearances
1	1	0	66
		VI0_PM0	28
		PM0	26
		PI0	20
		VI0_PF0	14
2	1	0	46
		VI0_PM0	28
		PI1	25
		VI0_PI1	21
		PM0	18

Table 1. Most frequent internal patterns at stress groups of 1 and 2 syllables in the emotional subcorpus

Syllables	Stresed syllable	Pattern	Appearances
1	1	0	151
		VI0_PF0	73
		VI0_PM0	71
		VF0	52
		VI0	50
2	1	0	119
		VI0_PM0	76
		VI0_PI1	70
		PI1	55
		PM0	41

Table 2. Most frequent internal patterns at stress groups of 1 and 2 syllables in the neutral subcorpus

An analysis of both tables shows that most frequent patterns are similar in neutral and emotional conditions: apart from 0 pattern (absence of relevant inflections along the stress group), which is the most frequent one in 1-syllable and 2-syllable groups, VI0_PM0 pattern (low F0 at the beginning of the group, and F0 peak in the middle of the stressed syllable) is the most common pattern in both conditions. Figure 2 shows an example of this type of pattern in the emotional corpus.

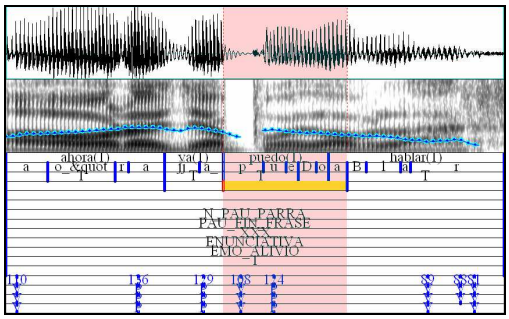

Figure 2. Example of VI0_PM0 internal pattern at the utterance '*Ahora ya puedo hablar*', uttered by a male speaker.

Other patterns, such as VI0_PF0 (low F0 at the beginning of the group, and F0 peak at the end of the stressed syllable) in 1-syllable groups or VI0_PI1 (low F0 at the beginning of the group, and F0 peak at the beginning of the post-stressed syllable) in 2-syllable groups, are also frequent in both conditions. The presence of these two patterns among the most frequent patterns in both conditions suggests that the phenomenon of peak delay [8, 9, 10] occurs in a similar manner both in neutral and emotional speech.

2.3.2. Sentence-final patterns

Tables 3 and 4 present the most frequent sentence-final patterns for the emotional and neutral subocorpora, respectively.

Syllables	Stressed syllable	Pattern	Appearances
1	1	VI0_PM0_VF0	23
		VF0	21
		VI0_PF0	8
		VI0_PM0_PF0	8
		PF0	6
2	1	VI0_PM0_VF1	21
		PI1_VF1	15
		PM0_VF1	13
		VI0_PM0_PI1_VF1	11
		VI0_PI1_VF1	10

Table 3. Most frequent sentence-final patterns at stress groups of 1 and 2 syllables in the emotional subcorpus

Syllables	Stressed syllable	Pattern	Appearances
1	1	VF0	45
		PI0_VM0_VF0	22
		PI0_VF0	18
		VM0_VF0	17
		VI0_VF0	14
2	1	PI0_VI1_VF1	45
		VI1_VF1	40
		PI0_VF0_VF1	18
		VM0_VF1	18
		VI0_VF1	17

Table 4. Most frequent sentence-final patterns at stress groups of 1 and 2 syllables in the neutral subcorpus

In this case, the observation of both tables reveals some differences between the two conditions: in the neutral subcorpus, all the most frequent patterns, both in 1-syllable

and 2-syllable groups, are falling (VF0, PI0_VM0_VF0, in the case of 1-syllable groups; PI0_VI1_VF1 or VI1_VF1, for example, in 2-syllable groups); in the case of emotional utterances, falling patterns (VF0, in 1-syllable groups; PI1_VF1 in 2-syllable groups) are also among the most frequent ones, but rise-fall patterns seem to be more frequent than falling ones: VI0_PM0_VF0 (low F0 at the beginning of the group, F0 peak in the middle of the stressed syllable, and low F0 again at the end of the stressed syllable) is the most frequent pattern in 1-syllable groups, and VI0_PM0_VF1 (low F0 at the beginning of the group, F0 peak in the middle of the stressed syllable, and low F0 at the end of the post-stressed syllable) in the case of 2-syllable groups. Figures 3 and 4 show two examples of these patterns. Rising patterns, such as VI0_PF0 (low F0 at the beginning of the group, and F0 peak at the end of the stressed syllable), VI0_PM0_PF0 (low F0 at the beginning of the group, and high F0 level from the middle to the end of the stressed syllable) or PF0 (high F0 level at the end of the stressed syllable) are also among the most frequent ones in the case of 1-syllable groups.

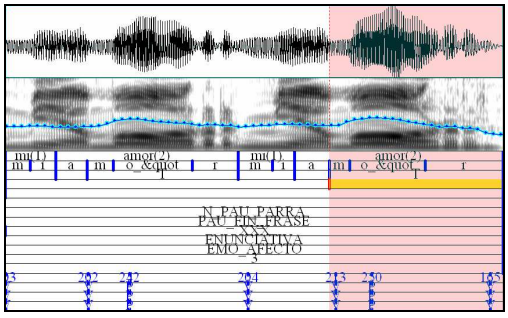

Figure 3. Example of VI0_PM0_VF0 sentence-final pattern at the utterance '*mi amor, mi amor*', uttered by a male speaker.

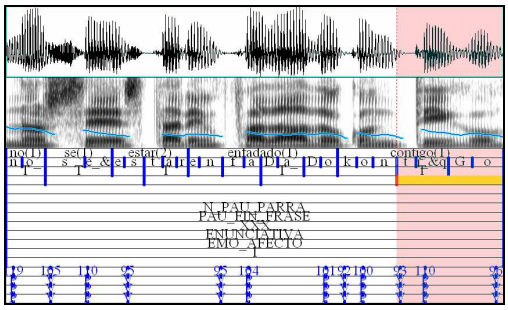

Figure 4. Example of VI0_PM0_VF1 sentence-final pattern at the utterance '*no sé estar enfadado contigo*', uttered by a male speaker.

3. Perceptual analysis

The results of the acoustical analysis showed then that main differences between F0 patterns used in neutral and emotional speech seem to be located at the end of intonational groups: neutral utterances show mainly a falling pattern at the end of declarative sentences, as expected according to previous descriptions, while emotional utterances show more variety in the type of patterns used, with a strong preference by rise-fall patterns. In internal position, however, similar patterns seem to be used in both conditions.

Considering these results, a perception experiment was designed in order to test the perceptual role of this type of rise-fall patterns in the identification of emotional speech by Spanish listeners.

3.1. Experimental design

The experiment was designed to test the perceptual contribution of three different sentence-final patterns to the identification of basic emotions in Spanish appearing among the most frequent patterns in the acoustical analysis:

- Falling pattern 1: low F0 along all the stress group (VF0 in 1-syllable groups; VI1_VF1 in 2-syllable or more groups)
- Falling pattern 2: high F0 at the beginning of the stress groups, and low F0 at the end (PI0_VM0_VF0 in 1-syllable groups; PM0_VI1_VF1 in 2-syllable groups).

Rise-fall pattern: low F0 at the beginning of the stress group, F0 peak in the middle of the stressed syllable and low F0 level again at the end of the group (VI0_PM0_VF0 for 1-syllable groups; VI0_PM0_VF1 for 2 or more syllable groups).

Table 5 shows a schematic representation of these three patterns.

Type	Scheme	Stress group	Pattern
Falling 1		1-syllable 2-syllable	VF0 VI1_VF1
Falling 2		1-syllable 2-syllable	PI0_VM0_VF0 PM0_VI1_VF1
Rise-fall		1-syllable 2-syllable	VI0_PM0_VF0 VI0_PM0_VF1

Table 5. Schematic representation of the three types of patterns considerer in the perception test

3.1.1. Stimuli

Six utterances from the neutral subcorpus were selected as base material, according to their length (not too long) and their semantic content (not inducing any special expressive interpretation). A base F0 contour (using the most frequent patterns detected in the acoustic analysis) was build for each selected sentence. Specific F0 contours for each condition were built by adding the three defined sentence-final patterns to every base contour, which gave 18 different modelled contours (6 base contours x 3 sentence-final patterns). These modelled contours were used to generate synthetic versions of the original utterances, using the Overlap-add synthesis tool included in Praat [11]. F0 range and height were controlled using reference lines for P and V points obtained from the neutral material. 18 different synthetic stimuli were obtained in this way.

3.1.2. Test

A web site was prepared to allow the audition of the different stimuli to the 8 subjects who run the test. They were asked to associate each stimulus to one of the following labels: 'joy', 'disgust', 'anger', 'fear', 'surprise', 'sadness' or 'neutral' (the six basic emotions plus the neutral condition). Subjects run the test individually, and they were encouraged to use headphones while listening the stimuli, and to run the test in a quiet room.

3.2. Results

Table 6 presents the results of the test, as a function of the pattern type and the labels chosen by the subjects.

	Pattern		
Answers *(%)*	Falling 1	Falling 2	Rise-fall
Neutral	21 *(43.7)*	18 *(37.5)*	13 *(27.0)*
Joy	1 *(2)*	7 *(14.5)*	1 *(2)*
Disgust	2 *(4.1)*	2 *(4.1)*	1 *(2)*
Anger	4 *(8.3)*	5 *(10.4)*	3 *(6.2)*
Fear	3 *(6.2)*	4 *(8.3)*	12 *(25)*
Surprise	3 *(6.2)*	2 *(4.1)*	9 *(18.7)*
Sadness	14 *(29.1)*	10 *(20.8)*	9 *(18.7)*

Table 6. Results of the perception test for each considered pattern, as a function of the chosen labels

The observation of the table shows that the 'neutral' label was the preferred one by listeners in all three cases, although the percentage is lower in the case of 'Falling 2' pattern (37.5% of the answers), and even lower in the case of the 'Rise-fall' pattern (27.08%). The 'rise-fall' pattern, however, shows also a significant percentage of answers associated to emotional labels: 'fear' (18.75% of the answers), 'surprise' (18.75% of the answers) and 'sadness' (18.75% of the answers). The label 'sadness' was used in a significant number of cases in the 'Falling 1' (29.16%) and 'Falling 2' (20.83%) conditions.

4. General conclusions

The results of the acoustical and perceptual analyses allow to draw some general conclusions about the role of melodic patterns in the phonetic expression of emotions in Spanish, at least in declarative sentences.

From an acoustical point of view, the results presented here seem to show that there are some differences in the melodic patterns used in neutral and emotional speech. These differences seem to appear only in final (boundary) patterns, and are related to the use of patterns different than falling - rising, but specially rise-fall ones- at this position. Rise-fall patterns arise then as a possible intonational cue for the expression of emotion in Spanish, as already suggested in previous studies. The results of the perceptual test indicate, however, that the role of this type of patterns in the perception of a utterance as emotional are relative: its use do contribute to interpret a utterance as emotional, but it is not a sufficient cue in many cases. When used, this rise-fall is associated to 'fear' or 'surprise' emotions. 'Sadness', however, is generally related to falling patterns.

5. References

[1] Iriondo, I., Guaus, R., Rodríguez, A., Lázaro, P., Montoya, N., Blanco, J. M., Bernadas, D., Oliver, J. M., Tena, D. and Longhi, L., "Validation of an acoustical modelling of emotional expression in Spanish using speech synthesis techniques", *Proceedings of the ISCA Workshop on Speech and Emotion* Northern Ireland, 161–166, 2000.

[2] Montero, J. M., *Estrategias para la mejora de la naturalidad y la incorporación de variedad emocional a la conversión texto a voz en castellano*, Ph. D. thesis, E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid, 2003.

[3] Martínez, H and Rojas, D., "Prosodia y emociones: datos acústicos, velocidad de habla y percepción de un corpus actuado", *Lengua y habla*, 15, 59-72, 2011. Online: http://erevistas.saber.ula.ve/index.php/lenguayhabla/article/viewFile/3356/3257, accessed on 14 Nov 2011.

[4] Navarro, T., *Manual de entonación española*, New York, Hispanic Institute on the United States, 1944.

[5] Garrido, J. M. (2010): "A Tool for Automatic F0 Stylisation, Annotation and Modelling of Large Corpora", Speech Prosody 2010, Chicago, May 2010. Online: http://speechprosody2010.illinois.edu/papers/100041.pdf, accessed on 25 Nov 2011.

[6] Garrido, J. M., *Modelling Spanish Intonation for Text-to-Speech Applications*, Ph D. thesis, Universitat Autònoma de Barcelona, 1996. Online: http://www.tdx.cat/TDX-0428108-155145/, accessed on 25 Nov 2011.

[7] Garrido, J. M., "La estructura de las curvas melódicas del español: propuesta de modelización", *Lingüística Española Actual*, XXIII/2: 173-209, 2001.

[8] Llisterri, J., Marín, R., de la Mota, C. and Ríos, A., "Factors affecting F_0 peak displacement in Spanish", en J. M. Pardo et alii (eds.) *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Technology*. Madrid, 18-21 September 1995, Vol. 3, 2061-2064, 1995.

[9] Prieto, P., van Santen, J. and Hirschberg, J., "Patterns of F0 peak placement in Mexican Spanish", *Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*, 30-34, 1994.

[10] Prieto, P., van Santen, J. and Hirschberg, J., «Tonal Alignment Patterns in Spanish». *Journal of Phonetics*, 23: 429-451, 1995.

[11] Boersma, P., "Praat, a system for doing phonetics by computer", *Glot International*, 5:9/10: 341-345, 2001.