PITCH RANGE AND IDENTIFICATION OF EMOTIONS IN SPANISH SPEECH: A PERCEPTUAL STUDY

RANGO TONAL E IDENTIFICACIÓN DE EMOCIONES EN EL HABLA EN ESPAÑOL: UN ESTUDIO PERCEPTIVO

JUAN MARÍA GARRIDO ALMIÑANA
Universidad Nacional de Educación a Distancia
jmgarrido@flog.uned.es

JUAN ANTONIO CHICA SABARIEGO
Universitat Pompeu Fabra
j.a.sabariego@gmail.com

Artículo recibido el día: 05/10/2017
Artículo aceptado definitivamente el día: 09/02/2018
Estudios de Fonética Experimental, ISSN 1575-5533, XXVII, 2018, pp. 13-36

ABSTRACT

This paper describes two perceptual experiments carried out to determine the contribution of pitch range to the identification of emotions in speech. Synthetic manipulated speech was used to observe the effect on emotion perception of two specific acoustic cues: global pitch range along utterances and local excursions at final rising-falling pitch movements. The obtained results suggest that increasing F0 (the main acoustic parameter related to pitch perception) in both cues, global and local pitch range, is associated by subjects to the expression of particular emotions, especially surprise (ESCP hypothesis), rather to a general expression of arousal (ACP hypothesis) or as a general cue for emotional content (EGP hypothesis), at least in Spanish.

Keywords: *prosody, pitch, perception, emotion, Spanish.*

RESUMEN

Este artículo describe dos experimentos de percepción llevados a cabo para determinar la contribución del rango tonal a la identificación de las emociones en el habla. Se utilizó habla sintética manipulada para observar el efecto sobre la percepción de las emociones de dos indicios acústicos específicos: el rango melódico global a lo largo de los enunciados y los movimientos melódicos locales de ascenso-descenso en posición final de enunciado. Los resultados obtenidos sugieren que los sujetos asocian el aumento de la F0 (el principal parámetro acústico relacionado con la percepción de la melodía) en ambos indicios, rango tonal global y local, con la expresión de emociones específicas, especialmente sorpresa (hipótesis ESCP), más que con una expresión genérica del grado de excitación (hipótesis ACP) o como un indicio general de contenidos emocionales (hipótesis EGP), al menos en español.

Palabras clave: *prosodia, melodía, percepción, emoción, español.*

1. INTRODUCTION

It is widely known from previous studies that the vocal expression of emotions in speech is strongly related to the variation of several prosodic cues, such as duration, amplitude, pausing or pitch (Lieberman and Michaels, 1962; Scherer, 1979, among many other; see, for example, Scherer, 2003; Juslin and Laukka, 2003, or Cowie and Cornelius, 2003, for a review of the large literature on the

topic). Studies on emotional speech in Spanish led to conclusions similar to those obtained for other languages (Navarro Tomás, 1944; Rodríguez *et al.*, 1999; Montero *et al.*, 1999; Iriondo *et al.*, 2000; Montero, 2003; Francisco *et al.*, 2005; Martínez and Rojas, 2011; Garrido, 2011, for example). As far as pitch is concerned, the expression of emotional content has been repeatedly associated to several pitch contour features, such as global pitch range (size of pitch excursions along utterances), register (mean pitch level) or pitch patterns. However, the role of these cues in the identification of particular emotions is not fully clear yet. This paper explores the contribution of one of these parameters, pitch range, to the identification of emotional contents in Spanish speech. Two different dimensions of pitch range are considered here for analysis: 'global' range, related to the size of pitch contour excursions along large domains, such as intonation/breath groups or full utterances, and 'local' range, referred to the size of pitch excursions in final boundary pitch patterns. Previous studies have focused their attention on global pitch range, but no specific studies have been devoted to local pitch range in the sense defined here. Both dimensions, global and local pitch range, although related, could have a more or less independent role in the expression of emotional states, and it is expected that make different contributions to their perceptual identification. Three possible roles of these two types of pitch range in the perception of emotions are explored here: as perceptual cues of specific emotional states ('emotional state categorical perception', ESCP hypothesis), as cues for the arousal emotional dimension ('arousal continuous perception', ACP hypothesis), and as global cues of emotion expression ('emotion global perception', EGP hypothesis).

1.1. Global Pitch Range

Pitch range, as a global feature affecting intonation groups or whole utterances, has been repeatedly reported as an acoustic cue for emotion expression in speech (Scherer, 1979; Ladd *et al.*, 1985; Pereira, 2000; Laukka, 2004; Laukka *et al.*, 2005; among many others). Literature on Spanish emotional speech production (Rodríguez *et al.*, 1999; Francisco *et al.*, 2005; Garrido, 2011) reveals that these cues behave in a similar way to other languages, with some emotions (joy, fear, surprise) showing high values of pitch range and other, such as sadness, rather low values; but they revealed also some variability in the use of pitch range when expressing emotions: depending on the study or analysed speaker, joy (Rodríguez *et al.*, 1999; Garrido, 2011), anger (Francisco *et al.*, 2005) and surprise (Garrido, 2011) are the emotions showing the highest mean values for pitch range. These studies could not establish then a clear relation between pitch range levels and

specific emotional categories, that is, that the expression of emotional categories is directly related to a given level of pitch range (ESCP hypothesis).

Based on data from languages different to Spanish, several authors (Scherer *et al.*, 1984, Ladd *et al.*, 1985; Pereira, 2000; Laukka, 2004; Laukka *et al.*, 2005) have hypothesised that pitch range variations are used in production (and interpreted in perception) as a cue for more global, continuous features of emotions, such as arousal. Parametrical classifications of emotions, proposed in classical works such as Russell (1980) or Whissell (1989), suggested that emotional states can be described using a set of continuous dimensions, usually two: arousal/activation, which measures how dynamic an emotional state is, and evaluation, which measures the positive or negative feeling associated with an emotional state. These dimensions define a continuous, multidimensional 'emotional space' in which emotional categories are located. In the original definition of this proposal, every emotional state would be then characterised by an intrinsic arousal level: some emotional states, such as joy or anger, would have by definition a high arousal level, whereas other states, such as sadness, would be characterised by a low arousal level. This hypothesis is referred here as the ACP hypothesis.

Basic emotions (anger, disgust, fear, joy, sadness, surprise; Ekman *et al.*, 1982) would be then distributed in the arousal/evaluation space as shown in Figure 1 (Pereira, 2000; Laukka, 2004; Laukka *et al.*, 2005). Some authors (Scherer *et al.*, 1984, for example) have proposed that the arousal parameter is some kind of non-intrinsic, additional feature which refers to the strength or intensity of the emotion, in the line of some other proposals (e.g. Schröder *et al.*, 2001; Garrido *et al.*, 2012b) which add a third power dimension, different from the arousal feature, to take into account the level of strength of a given emotion.

Anyway, irrespective of the fact that the arousal feature is intrinsic or additional, global pitch range would be interpreted then in a 'continuous' way, by placing the perceived utterance in a point of the arousal axis, rather than 'categorical', by assigning the utterance to a particular emotional category (Scherer *et al.*, 1984). According to this hypothesis, differences in pitch range would be interpreted by listeners in terms of degree of arousal when expressing emotions, rather than as a cue for a specific emotional category: high values of pitch range would be then perceptually associated by listeners to high degrees of arousal, whereas low values would be related to low arousal states.

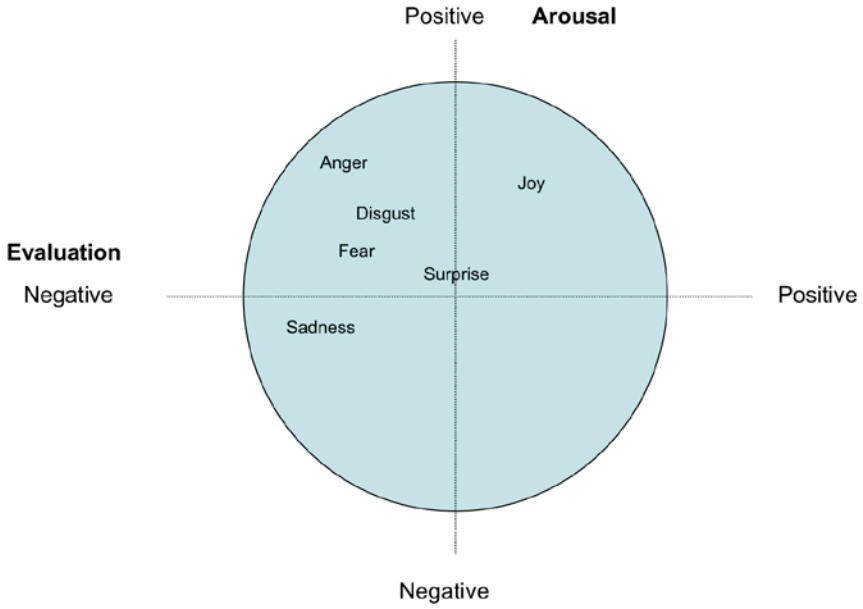

Figure 1. *Distribution of basic emotions along the evaluation-arousal space.*

Finally, some production data for Spanish (Garrido, 2011) suggest a third hypothesis: pitch range would be used as a general cue for emotional expression (EGP hypothesis). Basic emotions analysed in (Garrido, 2011) present higher mean pitch range values than neutral, non-emotional speech, independently of their positive or negative theoretical arousal level. This cue would be used then as a generic way to indicate that the speaker is trying to transmit emotional content, irrespective of the specific emotional category.

This study aims then at giving some light to the question of how listeners identify variations in global pitch range, at least in Spanish, as the perceptual effect of this cue in emotional speech has not been systematically analysed yet in this language.

1.2. Local Pitch Range

The role of intonation patterns in the expression of emotion in speech has been less studied than global pitch cues, but the few available studies show that intonation do play a role in the expression of emotions in speech (Ladd *et al.*, 1985) and that this role is related, among other cues, to the use of specific boundary pitch movements (Mozziconacci, 1995, 1998; Mozziconacci and Hermes, 1999). More specifically, Mozziconacci's production studies report the use in Dutch of a 'rise-fall' (RF) boundary movement ('1A' movement, using IPO conventions) when expressing several emotional states. RF patterns are also used frequently in Spanish emotional speech to express some emotions (surprise, joy and fear, mainly), in alternation with other types of falling and rising boundary patterns, but it is rarely used in the expression of other, such as sadness (Navarro Tomás, 1944; Garrido, 2011; Garrido *et al.*, 2012a). The shape of this RF pattern (a rising pitch movement starting at the beginning of the last stressed syllable of intonation groups, followed by a fast falling movement, still on the same stressed syllable) is illustrated in Figure 2.

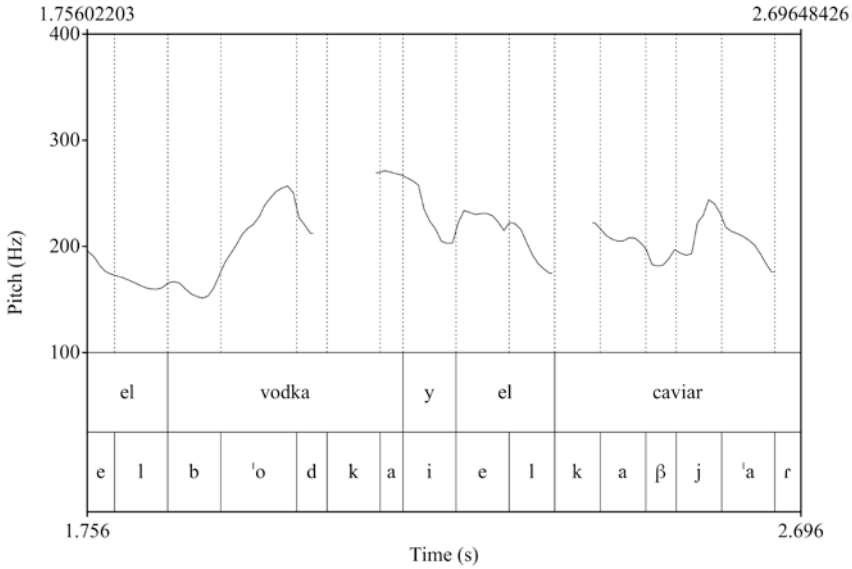

Figure 2. *F0 contour and annotation of the Spanish utterance* "el vodka y el caviar"*, uttered by a male speaker to express joy. The RF pattern appears in the last syllable of the utterance.*

Again, as in the case of global pitch range, previous findings about the use of this RF pitch pattern in the expression of emotions are not conclusive, neither in Spanish nor other languages. The results of a perception experiment described in Garrido *et al.* (2012a), for example, show that synthetic stimuli in Spanish generated with this type of RF pattern at the end of short utterances tend to be more identified as emotional than those generated with a falling pattern, but the results are not conclusive about if they can be associated to the expression of particular emotions: some emotions (such as fear, joy or surprise) are more identified by listeners when this pattern is used, but no one-to-one relation between emotional categories and pitch patterns could be established. These findings are similar to the ones reported in Mozziconacci (1995) for the perception of 1A pattern in Dutch. As suggested by Scherer *et al.* (1984), the perceptual interpretation of the use of this pattern should be categorical rather than continuous, as the presence or absence of a given pitch pattern, as a linguistic sign, only allow binary contrasts. This fact would favour the idea that pitch patterns should be considered general cues of emotion expression (EGP hypothesis).

Some studies (Mozziconacci, 1995; Borràs-Comes *et al.*, 2014) have suggested, however, that this RF pattern can show also continuous variation, and that this variation could have an effect on the perceived emotion. Mozziconacci (1995) gives acoustic evidence of differences in the F0 range of the rising-falling movement of RF patterns in Dutch depending on the expressed emotion: emotions with higher theoretical arousal, such as anger or joy, showed higher mean F0 ranges for this parameter than emotions with low theoretical arousal, such as sadness. This variation in range of the RF pitch movement could be interpreted then as a parameter to express emotional content. But this hypothesis, to our knowledge, has not been explored yet in a systematic way through perception experiments. The data presented in Mozziconacci (1995), as well as the informal observation of F0 contours of emotional speech in Spanish, suggest that this pitch range variation works independently on the global (intonation group) pitch range, as the size of the local excursion at boundary pitch 'rising-falling' patterns may be clearly higher than the global pitch range of the container intonation group. Also, the perception data reported in Borràs-Comes *et al.* (2014) for Catalan reveal that systematic range variations of this kind of pattern in synthetic stimuli are interpreted by listeners in terms of differences in the pragmatic meaning of the utterance. However, nothing is known about the role of these variations when expressing emotions. Considering this fact, the perceptual role of the variation of pitch range in this pitch pattern will be analysed in the second perception experiment presented here.

2. GOALS AND HYPOTHESES

This paper focuses then on the role of pitch range in the perceptual identification of emotions in Spanish. The effect of the systematic modification of these two dimensions (global pitch range along the whole sentence and size of rise-fall excursion at 'rising-falling' final patterns) on the identification of emotions in Spanish is analysed separately in two different perception experiments. The ultimate goal of both experiments is to determine which of the three hypothesis presented in the previous section (ESCP, ACP or EGP) fits better the obtained data on the perception by listeners of systematic changes of global and local pitch range in Spanish utterances, and to see if both types of pitch range are used in Spanish as unique or separate perceptual cues.

For practical reasons, both experiments have been limited to the analysis of basic emotions (anger, disgust, fear, joy, sadness, surprise; Ekman *et al.*, 1982), and assume, following the findings described in previous studies (Pereira, 2000; Schröder *et al.*, 2001), that these basic emotions are distributed in the arousal space along an axis as the one shown in Figure 3.

Three possible scenarios are expected, depending on the hypothesis: in the case of ACP hypothesis, those emotions which higher levels of arousal (anger, joy) should systematically be related by listeners to higher levels of global and local pitch range than those located in the negative side of the axis (sadness), and neutral, non-emotional sentences to pitch range values in between of those of positive and negative arousal; in the case of EGP hypothesis, all the emotional states, irrespective of its positive or negative arousal, should be related to higher levels of pitch range than in the case of neutral utterances; and finally, in order to validate ESCP hypothesis, only specific emotions should be associated to specific levels of pitch range, not following necessarily a given ordering.

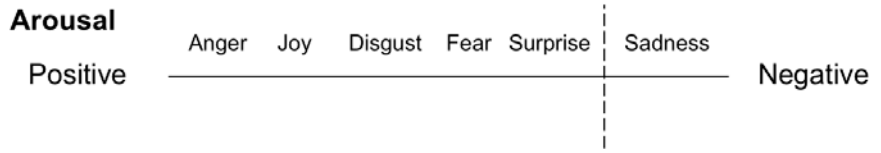

Figure 3. *Distribution of basic emotions along the arousal scale. The vertical line represents the "neutral" state.*

3. GENERAL MATERIALS AND METHODS

3.1. Base Material

The base material for both perception experiments was a set of four neutral natural utterances (two different sentences, '*Tiene un nombre polaco con connotaciones judías*' and '*En Léijar han desratizado tres viviendas*', uttered by two speakers, one male and one female) extracted from the INTERFACE corpus (Hozjan *et al.*, 2002), the same corpus used for the acoustic study described in Garrido (2011). The INTERFACE corpus is a corpus of emotional acted speech, in which two actors read aloud several times a set of utterances with a theoretically neutral lexical content.

Each utterance was read by the speakers several times, expressing the six basic emotions, plus a neutral condition (no emotion at all). The corpus has then the limitations of any corpus of emotional acted speech (lack of naturalness, exaggerated expression of emotions in some cases), but allows a direct comparison of the vocal expression of emotions by keeping lexical content unchanged. The utterances used for both experiments were selected from the neutral subset, and the criteria used for selection was length (rather short, 13-17 syllables), and lexical content (particular care was taken to not induce any particular emotional interpretation).

3.2. Generation of the Synthetic Stimuli

Synthetic speech as a research methodology to evaluate the perception of emotions has been proven to be useful in previous studies for Spanish (Garrido et al., 2012a) and other languages (Scherer *et al.*, 1984; Ladd *et al.*, 1985; Banziger and Scherer, 2005; Borràs-Comes *et al.*, 2014), as it allows to control systematically the shape of F0 contours and to analyse the contribution of the modified parameter to the perception of emotional states. It has the disadvantage, however, that unnatural utterances may be obtained if unrealistic acoustic values are used to generate the stimuli.

In the experiments presented here, the F0 contour of the selected utterances was modified systematically using ModProso, a Praat-based tool for F0 modification and synthesis (Garrido, 2013), to obtain different versions of the original contour but with seven different levels of global or local pitch range, depending on the experiment. ModProso applies the intonation description model described in Garrido (1996, 2001) to manipulate F0 contours. This framework, inspired by the

IPO model (t'Hart *et al.*, 1990), proposes that F0 contours are the result of the interaction of two types of F0 patterns: a global component, controlling the relative height and the range of contours along intonation groups; and a local component, defining the actual shape of F0 contours as the result of the concatenation of local F0 patterns. Local patterns are modelled as a series of relevant inflection points in the F0 contour anchored to specific positions within its container stress group (Figure 4), whereas global patterns are defined in the form of reference lines, defining the top and bottom margins for F0 excursions (figure 5).

Both global and local patterns for the four selected utterances had been previously obtained using MelAn, another Praat-based tool for automatic F0 contour stylisation, annotation and modelling (Garrido, 2010) also based on the Garrido (1996, 2001) model. ModProso allowed then manipulating separately these two global and local components for each utterance, a fact that led to a maximum control of the parameters considered for this study during the generation process.

To avoid unnatural results, the F0 values used for each version were calculated taking into account the real F0 range of the speakers who read the original sentences (two professional actors), calculated from the production data presented in Garrido (2011).

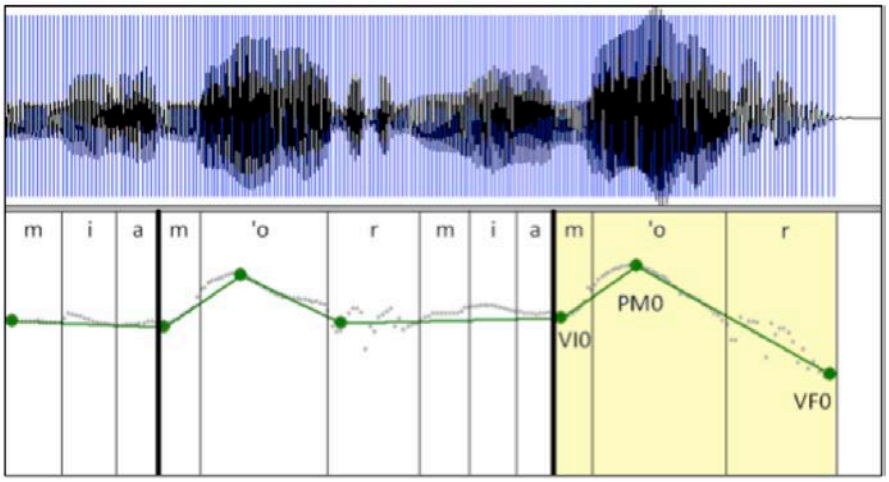

Figure 4. *Example of RF sentence-final pattern at the Spanish utterance* "mi amor, mi amor"*, uttered by a male speaker. Dots and straight lines represent the stylised contour obtained with MelAn.*

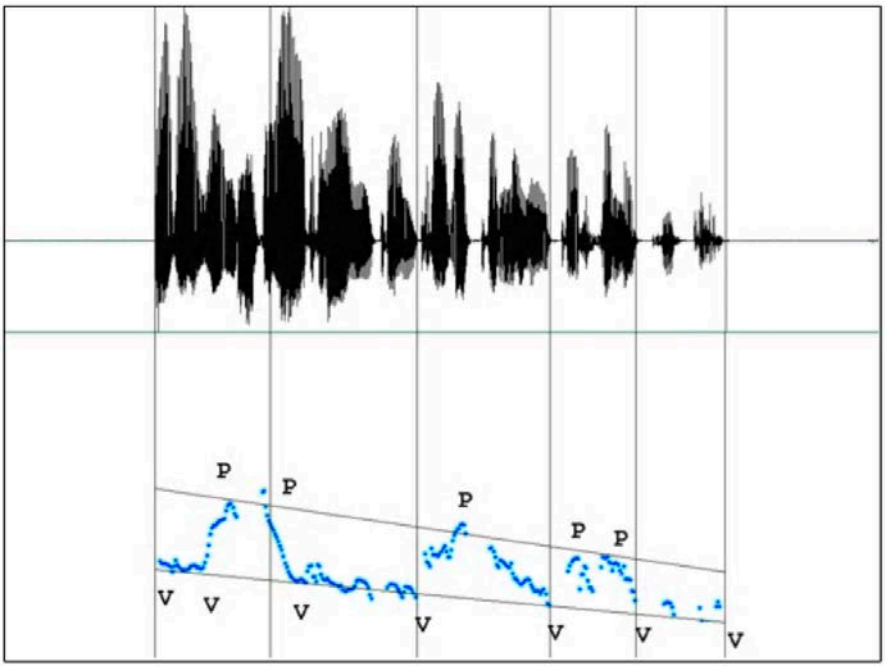

Figure 5. *Waveform and F0 contour of the utterance* "Aragón se ha reencontrado como motor del equipo"*, uttered by a Spanish female speaker. Vertical solid lines represent SG boundaries.*

3.3. Participants

Ten subjects (two men and eight women) participated in the experiments, all the same for both. They were all students at Pompeu Fabra University, with an age range of 23-30, native monolingual speakers of Northern Peninsular Spanish (Basque Country, Asturias, Galicia and Valladolid) and with no hearing nor visual impairments.

3.4. Procedure

Subjects ran the test alone, in a single session with a short break between both experiments, supervised by the person in charge of the experiment (one of the authors). Individual sessions were held in a quiet room at Pompeu Fabra university premises, and participants used headphones to listen to the stimuli. The stimuli

were presented in both experiments in a fixed random order, so all participants listened to the stimuli in the same order. They could listen to the stimuli as many times as they wanted, and had no time limitations to complete the test. They have to choose the emotional label that, according to them, fitted best the stimulus, among a closed set of seven ('anger', 'disgust', 'fear', 'joy', 'sadness', 'surprise' and 'neutral').

3.5. Analysis

The obtained responses were submitted in both experiments to a set of statistical analyses. First, a Fleiss' Kappa analysis was applied to test the consistency of the responses of the participants. Pearson's Chi-square tests were also applied to check the effect of the several variables ('speaker', 'utterance' and 'range level') on the subjects' responses. Finally, a Pearson product-moment correlation test was used to evaluate the relationship between the range level of the stimulus and the theoretical arousal level of the obtained response. All these analysis were carried out using the R statistical package. Next sections describe the specific procedures and results of both experiments in more detail.

4. EXPERIMENT 1

Experiment 1 was designed to evaluate the effect of systematic changes in global pitch range of Spanish utterances on the perception of emotions by listeners. Synthetic stimuli showing different levels of global range where presented to listeners, who had to link them to a given emotional category.

4.1. Stimuli

The four base utterances described in section 3.1 were used for the preparation of the final stimuli, 28 in total. For each base utterance, seven different modified versions were generated, each one with the same F0 shape (the one of the original base stimulus) but different global F0 range, as illustrated in figure 6. These modified versions were obtained using ModProso by systematically increasing, at regular F0 intervals, the F0 values of the P inflection points in the MelAn representation of the original F0 contour and generating a synthesised version of the utterance with this modified contour. Minimum ('level 0') and maximum ('level 6') F0 ranges were determined using actual F0 data of these speakers collected during the acoustic analysis described in Garrido (2011). The rest of levels was defined by establishing five intermediate F0 ranges between these two.

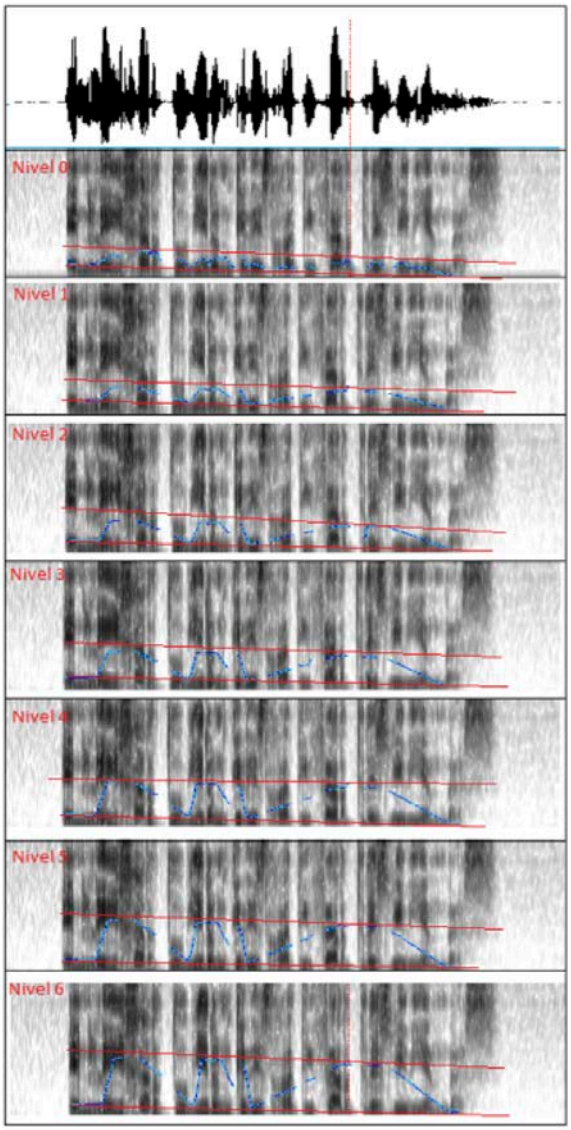

Figure 6. *Stylised F0 contours for the seven stimuli generated for experiment 1 from the base utterance* 'En Léijar han desratizado tres viviendas'*, uttered by a male speaker.*

4.2. Results

Table 1 summarizes the obtained results as a function of stimulus range level and emotion. They show, first of all, a strong tendency to associate intermediate levels of F0 range with the 'neutral' label, with is the preferred one in four of the seven conditions (from level 1 to 4). There is a tendency also to associate highest F0 range levels to 'surprise', with its maximum in the case of level 6 (maximum F0 range) and 5, and to 'joy' to a lesser extent. Inversely, there is a tendency of the number of 'sadness' responses to increase as lowest levels are considered, with its maximum at the 0 level. Finally, it is important to note that 'anger', 'fear' and 'disgust' did not have a remarkable number of responses in any level.

The correlation test carried out to evaluate the relation between F0 level and selected emotion revealed a rather low level of correlation (0.25) but statistically significant (df=248, p=0.00002051).

	anger	joy	disgust	fear	surprise	neutral	sadness
level 0	8	0	3	1	0	12	16
level 1	3	0	4	0	1	22	10
level 2	1	0	1	1	2	29	6
level 3	1	1	0	2	4	32	0
level 4	1	5	2	2	11	14	5
level 5	3	8	4	1	10	8	6
level 6	3	11	0	3	16	3	4

Table 1. *Obtained responses in the perception experiment for the seven considered global F0 range levels, as a function of the proposed emotion tags.*

Table 2 presents the results of the different Chi-square tests carried out to evaluate the effect of the variables 'F0 level', 'utterance' and 'speaker' on the obtained responses. It can be observed that 'F0 Level/Response' and 'Speaker/Response' pairs show p-values below the 0.05 significance level, indicating that the effect of these variables was statistically significant, whereas the 'Utterance/Response' pair did not show statistically significant results (p=0.205).

Finally, the result of the Kappa test showed a low agreement between subjects (0.194), which can be interpreted in the sense that there is some inter-subject variation in the interpretation of this acoustic cue.

	X-squared	df	p-value
level/response	159.175	36	< 2.2e-16
utterance/response	8.4805	6	0.205
speaker/response	33.6978	6	7.694e-06

Table 2. *Results of the Pearson's Chi-square tests for the variable pairs 'Range level/Response', 'Utterance/Response' and 'Speaker/Response' in Experiment 1.*

4.3. Discussion

The results obtained in this experiment partially confirm the ACP hypothesis, although they are not fully consistent with it, and provide also evidence to support the ESCP hypothesis. Both correlation and Chi-square tests show a statistically significant relation between global F0 range and arousal level, according to the arousal scale assumed for these experiments, despite the important inter-subject variation observed in the obtained responses. Indeed, highest F0 range values are mainly associated to emotional labels with positive arousal, mainly 'surprise' and 'joy' to a lesser extent, mid F0 range levels are mostly associated to neutral activation, and lowest F0 ranges are linked to negative activations ('sadness') in an important number of cases. However, some of the emotional labels related to positive levels of activation ('anger', 'fear' and 'disgust') are not related by listeners to high F0 range levels, two facts not in accordance with the ACP hypothesis. Finally, the EGP hypothesis does not seem to be validated, as listeners tend to associate the 'sadness' label to pitch range levels lower to those of neutral.

5. EXPERIMENT 2

Experiment 2 was designed to evaluate the effect on the perception of emotions by listeners of systematic changes in the size of the F0 excursion of on an artificially created final 'rise-fall' boundary pattern. Synthetic stimuli showing different levels of local range where presented to listeners, who had to link them to a given emotional category.

5.1. Stimuli

As in experiment 1, the four previously selected base utterances were used for the preparation of the final stimuli, 28 in total. For each base utterance, seven different modified versions were generated, each one showing a final RF pattern at the end with a different local F0 range, as shown in figure 7. Local F0 range was measured here as the difference between F0 values at both V inflection points of the pattern

(both had the same assigned F0 value) and the F0 value at the mid P inflection point, which defined the peak of the rise-fall pattern.

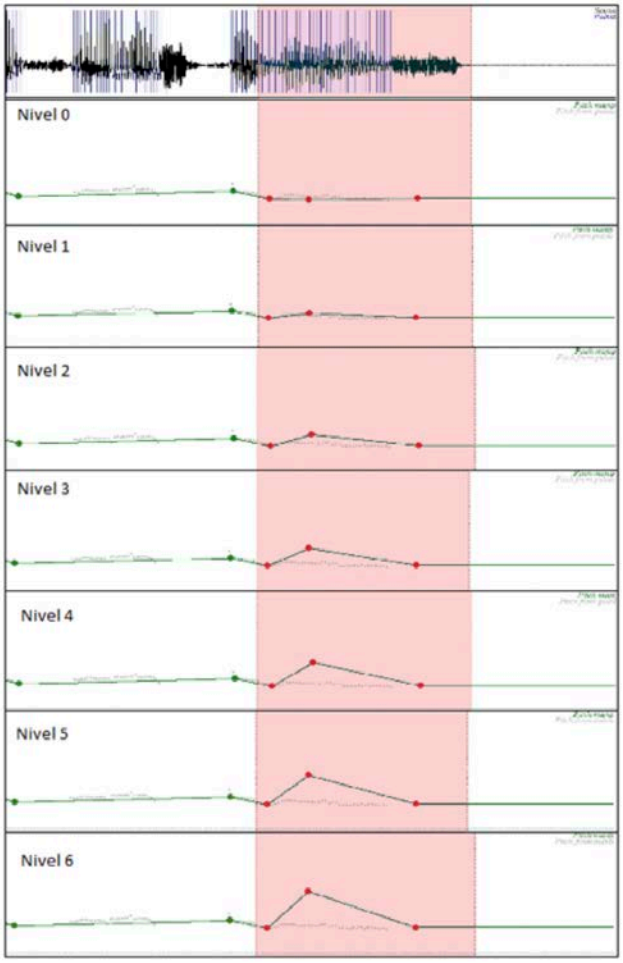

Figure 7. *Stylised F0 contours for the final part of the seven stimuli generated for experiment 1 from the base utterance "*Tiene un nombre polaco con connotaciones judías*", uttered by a female speaker. The highlighted area corresponds to the last stress group of the utterance, where the modified pitch boundary is anchored.*

The procedure to define the specific F0 range values for each level (ranging for 0 to 6, as in experiment 1) was similar to the one applied in the previous experiment: first minimum and maximum F0 was established first taking into account the F0 data collected for both speakers in the acoustic analysis of the INTERFACE corpus (Garrido, 2011), but in this case an F0 range of 0 Hz was associated to level 0, (that is, a flat pattern, with no rise-fall movement at all); then F0 range values for levels 1-5 were set by increasing the F0 range at regular intervals, different for each speaker, from 0 to the maximum F0 range established for each speaker (level 6). Again, ModProso was used to generate the stylised synthetic versions of the original utterances, but in this case a further step of manual modification using Praat was necessary to adapt the F0 range of the final boundary pattern.

5.2. Results

Table 3 summarizes the obtained results for this second experiment as a function of stimuli level and emotion. A clear polarization of the data can be observed: highest levels of F0 range (from 3 to 6) are mainly associated to 'surprise'; lowest levels are mainly identified as neutral, emotionless utterances; and finally, intermediate levels (2-4 show also a tendency to be interpreted as 'anger', but also as 'neutral' or 'surprise'.

	anger	joy	disgust	fear	surprise	neutral	sadness
level 0	2	0	1	0	0	34	3
level 1	1	0	3	1	0	30	5
level 2	11	1	5	0	9	11	3
level 3	11	3	2	1	15	8	0
level 4	9	4	2	0	20	4	1
level 5	6	2	0	1	25	4	2
level 6	4	4	1	2	26	2	1

Table 3. *Responses obtained in the perception experiment for the seven considered local F0 range levels, as a function of the proposed emotion tags.*

The correlation test carried out to evaluate the relation between F0 level and theoretical arousal degree revealed a rather low level of correlation, similar to the one obtained in experiment 1 (0.27), but again statistically significant (df=248, p=0.000004457).

Table 4 presents the results of the different Chi-square tests carried out to evaluate the effect of the variables 'F0 level', 'utterance' and 'speaker' on the obtained

responses. It can be observed that 'Level/Response', 'Utterance/Response' and 'Speaker/Response' pairs show p-values below the 0.05 significance level, indicating that the effect of these variables was statistically significant.

Again, the result of the Kappa test showed a low agreement between subjects, but higher than in the previous experiment (0.246).

	X-squared	df	p-value
level/response	177.3925	36	< 2.2e-16
utterance/response	23.0204	6	0.0007897
speaker/response	20.5512	6	0.002208

Table 4. *Results of the Pearson's Chi-square tests for the variable pairs 'Range level/Response', 'Utterance/Response' and 'Speaker/Response' in Experiment 2.*

5.3. Discussion

As in experiment 1, the obtained results are in partial agreement with the ACP hypothesis, although they are not fully consistent with it: the results of the statistical tests indicate again that it exists a degree of relation between level of the F0 excursion in the final RF pattern and arousal level. However, as in the previous experiment, this correlation is far from being perfect: high and mid levels of F0 range are clearly related to emotions showing positive arousal ('surprise' again, as in experiment 1, and 'anger' as a second choice in this case), and lowest F0 range values are related to neutral levels, but other emotions with positive arousal, such as joy, disgust and fear, are not related to mid or high levels of F0 excursions, nor sadness to the lowest values of F0 excursions, as it could be expected. It seems, then, that the use of 'rising-falling' patterns with high F0 excursions is more related to the expression of specific emotions (ESCP hypothesis), 'surprise' in the case of highest levels and 'anger' in the case of mid-level stimuli, than to the general expression of the arousal feature.

6. GENERAL DISCUSSION AND CONCLUSIONS

The results of the two described experiments seem to suggest that both types of pitch range are interpreted by listeners in a different way, so it can be suggested that they behave as separate perceptual cues in the identification of emotions, at least in Spanish.

In the case of global pitch range, the results seem to support weakly the ACP hypothesis, as listeners tend to identify stimuli with higher range to positive arousal emotions, stimuli with low range to negative arousal emotions, and mid-level range to neutral condition. But they provide also evidence for the other two hypotheses: global pitch range is mainly associated by listeners to specific emotions (ESCP hypothesis), as 'surprise' and 'joy', two emotions related to positive arousal but not to its highest values. However, 'anger', the emotional label associated to the highest arousal level in this paper, is not associated to maximum pitch levels. And 'disgust' and 'fear', associated usually to mid levels of arousal, are not linked neither by listeners to high levels of global F0 range. These facts are in agreement with the production data presented in Garrido (2011), in which highest levels of F0 range are observed at surprise and joy utterances, but anger utterances show only mid levels of F0 range. Finally, EGP hypothesis can not be fully rejected, as although 'sadness' levels are in general associated to pitch levels lower to the ones associated to neutral condition, the fact that some subjects related high levels of range to sadness indicates that, at least for them, high pitch range is a cue for emotion in general, and not only of emotions with positive arousal. This fact could be interpreted in the sense that some subjects interpret global pitch range as a cue of arousal dimension (ACP hypothesis), and other as a cue of emotion in general (EGP hypothesis), and that it would exist some kind of inter-subject variation on this aspect.

In the case of local pitch range (final RF pattern), the obtained results complete and refine previous production and perception studies on the use of this kind of pattern to express emotions in Spanish (Garrido, 2011; Garrido *et al.*, 2012a), in the sense that they show that the RF pattern seems to be used as an acoustic cue to identify several (but not all) the basic emotions in Spanish, mainly surprise and anger. But the results presented here also show that the continuous modification of pitch range in this kind of pitch patterns do play a role in the identification of emotions in Spanish, and that its role is not exactly the same as the one of global range. As in the case of global range, changes in pitch range tend to be interpreted as different intended emotions, giving support to the ESCP hypothesis, but the interpretation of these changes is not exactly the same as in the case of global range: in this case lowest levels of pitch range (that is, flat or falling patterns) are identified mainly as neutral (and, to a much lower extent, to sadness), intermediate levels to anger and highest levels to surprise.

In summary, the experiments presented in this paper have provided some evidence for the use of global and local pitch range as independent parameters to express emotions in Spanish, and they favour the idea that their variation tends to be

interpreted mainly as a cue of specific emotions (ESCP hypothesis), rather than as a cue for arousal dimension or as a global emotion cue. However, the results are not conclusive at all, and have to be handled with caution, considering also the small amount of listeners who participated in both experiments. So more research should be needed to validate these findings with more subjects and to investigate, for example, the responses of listeners when exposed to the same stimuli and asked about different levels of power instead of different emotional categories.

7. REFERENCES

BANZIGER, T. and SCHERER, K. R. (2005): «The role of intonation in emotional expressions», *Speech Communication*, 46, pp. 252-267.

BORRÀS-COMES, J.; VANRELL, M. M. and PRIETO, P. (2014): «The role of pitch range in establishing intonational contrasts», *Journal of the International Phonetic Association*, 44, pp. 1-20.

COWIE, R. and CORNELIUS, R. R. (2003): «Describing the emotional states that are expressed in speech», *Speech Communication*, 40 (1-2), pp. 5-32.

EKMAN, P.; FRIESEN, W. V. and ELLSWORTH, P. (1982): «What emotion categories or dimensions can observers judge from facial behavior?», in P. Ekman (ed.): *Emotion in the human face*, New York, Cambridge University Press, pp. 39-55.

FRANCISCO, V.; GERVÁS, P. and HERVÁS, R. (2005): «Análisis y síntesis de expresión emocional en cuentos leídos en voz alta», *Procesamiento del Lenguaje Natural*, 35, pp. 293-300.

GARRIDO, J. M. (1996): *Modelling Spanish intonation for text-to-speech applications*, tesis doctoral, Universitat Autònoma de Barcelona.

GARRIDO, J. M. (2001): «La estructura de las curvas melódicas del español: propuesta de modelización», *Lingüística Española Actual*, XXIII(2), pp. 173-209.

GARRIDO, J. M. (2010): «A tool for automatic F0 stylisation, annotation and modelling of large corpora», *Speech Prosody 2010, Chicago, May 2010*. http://speechprosody2010.illinois.edu/papers/100041.pdf [03/10/2017].

GARRIDO, J. M. (2011): «Análisis de las curvas melódicas del español en habla emotiva simulada», *Estudios de Fonética Experimental*, XX, pp. 205-255.

GARRIDO, J. M. (2013): «ModProso: A Praat-based tool for F0 prediction and modification», *Proceedings of TRASP 2013*, pp. 38-41. http://www.lpl-aix.fr/~trasp/Proceedings/19866-trasp2013.pdf [03/10/2017].

GARRIDO, J. M.; LAPLAZA, Y. and MARQUINA, M. (2012a): «On the use of melodic patterns as prosodic correlates of emotion in Spanish», *Speech Prosody 2012, Shanghai*. http://isle.illinois.edu/sprosig/sp2012/uploadfiles/file/sp2012_submission_57.pdf [03/10/2017].

GARRIDO, J. M.; LAPLAZA, Y., MARQUINA, M., PEARMAN, A., ESCALADA, J. G., RODRÍGUEZ, M. A. and ARMENTA, A. (2012b): «The I3MEDIA speech database: a trilingual annotated corpus for the analysis and synthesis of emotional speech», *LREC 2012 Proceedings*, pp. 1197-1202. http://www.lrec-conf.org/proceedings/lrec2012/pdf/865_Paper.pdf [03/10/2017].

HOZJAN, V.; KACIC, Z., MORENO, A., BONAFONTE, A. and NOGUEIRAS, A. (2002): «Interface databases: Design and collection of a multilingual emotional speech database», in *Proceedings of the Third Int. Conference on Language Resources and Evaluation (LREC'02), Las Palmas de Gran Canaria*, pp. 2024-2028.

IRIONDO, I.; GUAUS, R., RODRÍGUEZ, A., LÁZARO, P., MONTOYA, N., BLANCO, J. M., BERNADAS, D., OLIVER, J. M., TENA, D. and LONGUI, L. (2000): «Validation of an acoustical modelling of emotional expression in Spanish using speech synthesis techniques», in *Proceedings of the ISCA Workshop on Speech and Emotion, Newcastle*, pp. 161-166.

JUSLIN, P. and LAUKKA, P. (2003): «Communication of emotions in vocal expression and music performance: Different channels, same code?», *Psychological Bulletin*, 129, pp. 770-814.

LADD, S.; SIVERMAN, K., BERGMANN, G. and SCHERER, K. (1985): «Evidence for independent function of intonation contour type, voice quality, and F0 in signalling speaker affect», *Journal of the Acoustical Society of America*, 78(2), pp. 435-444.

LAUKKA, P. (2004): *Vocal expression of emotion, discrete-emotions and dimensional accounts*, Uppsala, Uppsala University.

LAUKKA, P., JUSLIN, P. and BRESI, R. (2005): «A dimensional approach to vocal expression of emotion», *Cognition and Emotion*, 19(5), pp. 633-653.

LIEBERMAN, P. and MICHAELS, S. D. (1962): «Some aspects of fundamental frequency, envelope amplitude and the emotional content of speech», *Journal of the Acoustical Society of America*, 34, pp. 922-927.

MARTÍNEZ, H. and ROJAS, D. (2011): «Prosodia y emociones: datos acústicos, velocidad de habla y percepción de un corpus actuado», *Lengua y Habla*, 15, pp. 59-72.

MONTERO, J. M. (2003): *Estrategias para la mejora de la naturalidad y la incorporación de variedad emocional a la conversión texto a voz en castellano*, tesis doctoral, Universidad Politécnica de Madrid.

MONTERO, J. M.; GUTIÉRREZ-ARRIOLA, J., COLÁS, J., ENRÍQUEZ, E. and PARDO, J. M. (1999): «Analysis and modelling of emotional speech in Spanish», in *Proceedings of the 14th International Conference of Phonetics, San Francisco*, pp. 957-960.

MOZZICONACCI, S. J. (1995): «Pitch variations and emotions in speech», in *Proceedings of the 13th International Congress of Phonetic Sciences, ICPhS-95, Stockholm, August 13-19, 1995*, Vol. 1, pp. 178-181.

MOZZICONACCI, S. J. (1998): *Speech Variability and Emotion: Production and Perception*, tesis doctoral, Technische Universiteit Eindhoven.

MOZZICONACCI, S. J. and HERMES, D. J. (1999): «Role of intonation patterns in conveying emotion in speech», in *Proceedings ICPhS99, San Francisco*, pp. 2001-2004.

NAVARRO TOMÁS, T. (1944): *Manual de entonación española*, New York, Hispanic Institute on the United States.

PEREIRA, C. (2000): «Dimensions of emotional meaning in speech», in *Proceedings of the ISCA ITRW on Speech and Emotion, Newcastle, 5-7 September 2000*, pp. 25-28.

RODRÍGUEZ, A.; LAZARO, P., MONTOYA, N., BLANCO, J. M., BERNADAS, D., OLIVER, J. M. and LONGHI, L. (1999): «Modelización acústica de la expresión emocional en el español», *Procesamiento del Lenguaje Natural*, 25, pp. 159-166.

RUSSELL, J. A. (1980): «A circumplex model of affect», *Journal of Personality and Social Psychology*, 39(6), pp. 1161-1178.

SCHERER, K. R. (1979): «Personality markers in speech», in K. R. Scherer and H. Giles (eds.), *Social Markers in Speech*, Cambridge, Cambridge University Press, pp. 147-210.

SCHERER, K. R. (2003): «Vocal communication of emotion: A review of research paradigms», *Speech Communication*, 40, pp. 227-256.

SCHERER, K. R.; LADD, D. R. AND SILVERMAN, K. (1984): «Vocal cues to speaker affect: Testing two models», *Journal of the Acoustical Society of America*, 76, pp. 1346-1356.

SCHRÖDER, M.; COWIE, R., DOUGLAS-COWIE, E., WESTERDIJK, M. and GIELEN, S. (2001): «Acoustic correlates of emotion dimensions in view of speech synthesis», in *Proceedings of Eurospeech 2001, Aalborg*, pp. 87-90.

'T HART, J.; COLLIER, R. and COHEN, A. (1990): *A Perceptual Study of Intonation. An Experimental-phonetic Approach to Speech Melody*, Cambridge, Cambridge University Press.

WHISSELL, C. (1989): «The dictionary of affect in language», in R. Plutchik and H. Kellerman (eds.): *Emotion: Theory, Research and Experience, 4*, New York, Academic Press, pp. 113-131.