

Intonation and Emotion: Influence of Pitch Levels and Contour Type on Creating Emotions

Emma Rodero, *Barcelona, Spain*

Summary: Intonation is a vehicle for communication, which sometimes contributes greater meaning than the semantic content of speech itself. This prosodic element lends the message linguistic and paralinguistic meaning, which carries a highly significant communicative value when conveying emotional states. For this reason, this article analyses the use of intonation as an instrument for arousing various sensations in the listener. The aim was to verify which elements of intonation are more decisive to generate a specific sensation. Experimental research is conducted, in which certain pitch patterns (pitch levels and contour type) are assigned different emotions (joy, anxiety, sadness, and calmness) and are then listened to and assessed using a questionnaire with a bipolar scale of opposed pairs, by a sample audience comprising 100 individuals. The main conclusion drawn is that, although both the variables analyzed—pitch level and contour type—are representative of expressing emotions, contour type is more decisive. In all the models analyzed, contour type has been highly significant and constitutes the variable that has been determined as the final component for recognizing various emotions.

Key Words: Emotion–Intonation–Pitch level–Contour type.

INTRODUCTION

Intonation constitutes a decisive prosodic element in arousing sensations in the listener. It conveys aesthetic information to us, in many cases, concerning unconscious perception, which is the key to recognize the emotions being communicated.

Probably, the chief impetus to this long-standing effort stems from the conviction that speech melody has a particular communicative value. Indeed, individual speech sounds do not carry an intrinsic meaning of their own, whereas intonation and other prosodic features seemingly add something to the content of a message that is not already expressed in the semantics of its individual words nor in their syntactic relations.¹

As an interdisciplinary field, its study may have a range of applications in clinical, psychological, pedagogical, or communicative research. Accordingly, a host of researches has been carried out in relation to intonation and emotions, where the first challenge is the concept of the term emotion that often includes attitudes, feelings, or intentions. In our case, emotion is understood to be more related to attitude than to feeling or intention. In this study, sensation and emotion will be taken to be synonymous.² Another important element is the selection of these emotions and whether they are basic or combinations of basic elements. In this respect, a range of studies have been conducted showing highly diverse perspectives. For the purpose of this research, sensations that are considered to be basic in literature relating to emotions have been chosen.^{3,4} These emotions illustrate opposed states and levels of heightened or little activity and positive

and negative values.⁵ The emotions chosen are joy, anxiety, sadness, and calmness.

The second and most important challenge in studying intonation is the very definition of the parameters surrounding this prosodic element. Again, varied conceptions exist, depending on the author. For this research, it is understood that the emotional load conveyed by intonation is sustained by movements in the intonation curve and by pitch levels. Therefore, the emotional meaning is formed by means of the selection of the mean pitch level and the contour of intonation.^{6–8} It is evident that these two elements on their own do not establish the entire burden of expressive meaning, because speech is more complex than this and includes other prosodic characteristics, such as accent; speech rate; and other aspects related to vocal qualities, namely, duration, intensity, or timbre. However, this study is solely restricted to these two variables, because these are considered extremely significant in arousing sensations in the listener, as shown in other researches.^{9–11}

One of the determining factors in forming intonation is the main pitch level at which it is located, in other words, the average level of pitch extracted that forms the prosodic contour. Studies relating to intonation and emotions that use this parameter (pitch level) to analyze the relationship between these two variables are highly numerous.^{2,12,13} These studies have found that pitch level is quite significant when determining the type of emotion that is conveyed. In a study applied to radio news, the selection of an inadequate pitch level entailed a greater error in intonation for the sample.¹⁴ In addition, Gussenhoven¹⁵ concludes that “the Frequency Code is widely used for the expression of affective meanings,” whereas for Mozziconacci,² “the pitch level and excursion size of the pitch movements would be more related to the speaker’s emotional state.”

A link established by many studies on prosody attributes the high pitch levels to drawing the attention of the listener, because it establishes a contrast, and to emotions, such as joy, anxiety, or fear. Medium pitch levels account for more neutral attitudes. However, low pitch levels are related to sober emotions:

Accepted for publication February 16, 2010.

From the Department of Communication, Pompeu Fabra University (UPF), Barcelona, Spain.

Address correspondence and reprint requests to Emma Rodero, Department of Communication, Pompeu Fabra University (UPF), C/Roc Boronat, 138, 08018 Barcelona, Spain. E-mail: emma.rodero@upf.edu

Journal of Voice, Vol. 25, No. 1, pp. e25–e34
0892-1997/\$36.00

© 2011 The Voice Foundation
doi:10.1016/j.jvoice.2010.02.002

sadness, calmness, or security. Thus, emotions carrying a high level of activity, such as joy or fear, tend to be situated in the top end of the frequency spectrum of the speaker. Conversely, sadness and desire tend overall to be formed at the lower end.^{5,16,17} In this respect, the results of Mozziconacci's² research are interesting. It shows that joy obtained a suitable percentage of identification for the sample of 62%, whereas for sadness, the result was 97%. Therefore, in studies by this author, for the German language, joy was less recognizable than sadness. In the case of the sensations analyzed in this research, the high levels will apply to anxiety and joy, whereas, in a logical way, the lowest levels will be associated with sadness and calmness.

Nevertheless, analyses relating to pitch level have, in many cases, been declared as insufficient to determine a specific emotion. For that reason, Mozziconacci² considers that "indeed, emotions are conveyed by pitch, pitch level and range, and also conveyed by type of contour. Additionally, it has been shown that using both features, pitch and contour type yields better results." Thus, after pitch level, the other significant element is the curve for intonation or contour type. For some authors, such as Cantero,¹⁸ contour type is the fundamental element when an attempt is made to emphasize an expression, that is, when a sensation is sought to be aroused: "again, the relevant phonetic characteristic to mark the characteristic 'emphasis' is the contour structure." Therefore, once the speaker selects the pitch level, the contour type must be altered, raising and lowering the levels, based on the emotion that is sought to be provoked, thereby creating different inflexions in the speech. In this case, the contour with the greatest inflexions is associated with emotions showing greater excitement: joy, anxiety, or fear. On the contrary, those that maintain fewer variations are related to more sober states: sadness, calmness, or security. In this regard, Sosa¹⁹ concludes that "greater movement, greater excitation perceived; smaller movement, minor degree of emotional implication, although the configuration does not change." According to the studies of Boula de Mareüil et al.,²⁰ applied in Spanish, although sadness is pronounced clearly with few variations in contour, and anxiety or fear have pronounced inflexions, in the case of joy, "pitch variations are replaced by straight lines between peaks and valleys." On the other hand, Cantero¹⁸ characterizes the emphatic contours illustrated as "inconstant slope of the declination." For this author, the practical elimination of the decline is related to sadness or discouragement. Other authors back this description of the parameters defining these emotions in their studies.²¹⁻²³

All these data reveal that, although inflexions in the most intense sensations are necessary, the form these diversions take on turns out to be a factor to take into consideration. Therefore, with these prior considerations, the sensations of joy and anxiety have been constructed in this study with greater inflexions to a different extent, whereas the sensations of sadness and calmness have maintained the contour with less pitch variations also at different levels.

In short, the aim of this study was to verify the effectiveness of these two prosodic elements—pitch level and contour type—to identify an emotion and, thus, determine which of the two was more decisive in generating a certain sensation. The main

hypothesis for this study is that pitch levels will be less significant than contour type in creating these emotions.

METHODS

The experiment was set out to determine the extent to which the variables—pitch level and contour type—affected the arousal of four different sensations in the listener. To do so, a corpus was prepared consisting of 16 sound recordings of the same utterance by four prominent Spanish actors: four recordings for each emotion and one emotion assigned to each actor. Then actors were chosen based on their degree of vocal versatility and their knowledge in intonation, albeit basic. They were all males, to prevent differences in tone interfering in the results. This choice is justified by the need to control both study variables when recording each speaker and, thus, conduct an experimental manipulation of each emotional state, as proposed by authors, such as Scherer.²⁴

Each one of these 16 recordings corresponds to four different sensations in two major categories: joy-sadness and calmness-anxiety. The actors were not given instructions on how the sentence was to be read in terms of intensity, speed, or pauses, but they were given instructions about emotion, whereby they were guided on the intonation patterns they were to follow. For each model, each actor recorded four different versions with the instruction that they had to attain the specific sensation only amending elements relating to intonation with regard to both pitch level and contour pitch. Once all the recordings had been obtained (64), the tests were chosen and were significantly different for each emotion in terms of pitch level and contour type. This analysis was made by using acoustic analysis software *Praat*,²⁵ broadly used for prosodic studies, with the Momel pattern of stylization and the INTSINT transcription system. To analyze the movements, initially, stylization and the transcription comprising the corpus were carried out. Once system detection errors had been corrected, target points were labeled by INTSINT. As a general scheme, the resulting combinations were designed by determining positive frequencies (+) for the two models with the greatest pitch level or inflexions in contour type and negative ones (−) for those with a lower frequency and movement of tone in both cases. Accordingly, for each emotion, the following combinations were drawn: high pitch level-high tone movement (++), high pitch level-low tone movement (+−), low pitch level-high tone movement (−+), and low pitch level-low tone movement (−−). Then, these combinations were assigned a random model number to prevent the presentation order in the questionnaire from being the same for all emotions, and this may affect the results.

Because each actor recorded the four models for a specific emotion, the voice quality parameters were uniformly arranged, and their influence on the results was prevented. As a result, 16 final models were obtained, four for each emotion (M1, M2, M3, and M4), the data for which can be seen in [Table A1](#) along with the intonation curves ([Appendix](#)).

Each of the recordings registered the same utterance to avoid the influence of content, along the lines of Ehrette et al.,²⁶ "... as we used a same sentence, the same segmental content, for all the

TABLE 1.
Descriptives for Each Emotion

Models	Descriptives	Joy	Anxiety	Sadness	Calmness
M1	Mean	3.60	3.20	3.20	2.40
	SD	0.93	0.92	0.98	0.82
	Standard error	0.09	0.09	0.09	0.08
M2	Mean	4.70	1.43	4.00	3.60
	SD	0.54	0.55	0.97	0.98
	Standard error	0.05	0.05	0.09	0.09
M3	Mean	2.40	4.40	2.60	1.10
	SD	0.84	0.76	0.94	0.33
	Standard error	0.08	0.07	0.09	0.03
M4	Mean	1.90	2.06	4.80	4.50
	SD	0.71	0.82	0.44	0.74
	Standard error	0.07	0.08	0.04	0.07
Total	Mean	3.15	2.77	3.65	2.90
	SD	1.33	1.37	1.19	1.48
	Standard error	0.06	0.06	0.06	0.07

utterances in our corpus, it seems that emotion and voice identity is conveyed mostly by prosodic changes.” Moreover, the utterance was chosen considering that the content was not associated with any specific sensation. The sentence construction was designed to be as neutral as possible, in compliance with earlier research.³ The following sentence was used: “It was a morning just like any other. However, the meeting would be a way to talk about life. He did not know if it would be good or bad.”

Subsequently, these recordings were subject to evaluation by a sample comprising 100 communication students ($N = 100$), 46 males and 54 females, chosen randomly and without intonation knowledge. The tests were carried out over four consecutive days in four distinguished sessions: one for each emotion. Consequently, any possible tiredness of the sample because of listening to 16 recordings in one session was avoided, and the listening order was maintained in such a way that it had no effects on the results.

In each session, after listening to each model twice, using a questionnaire with a bipolar scale of opposed pairs, each subject established the order of the recordings, based on a greater or smaller association with the corresponding emotion. Hence, as applied by Pereira²⁷ in one of his studies, each of these dimensions was represented on two scales defined by pairs of adjectives. In the case of joy and sadness, it was more-less cheer, more-less sad, and for calmness and anxiety, it was more-less calm, more-less nervous. The codification was carried out on a scale of 1–5, where 5 represents most cheer, sad, calm, and nervous, and 1 accounts for less cheer, sad, calm, and nervous. An open question was added to the questionnaire where each subject had to justify his or her choice explaining his or her highest score according to the extent to which an association was established with the emotion in question.

RESULTS

The results of the comparison of averages obtained for the variables examined in this study make it possible to establish clear

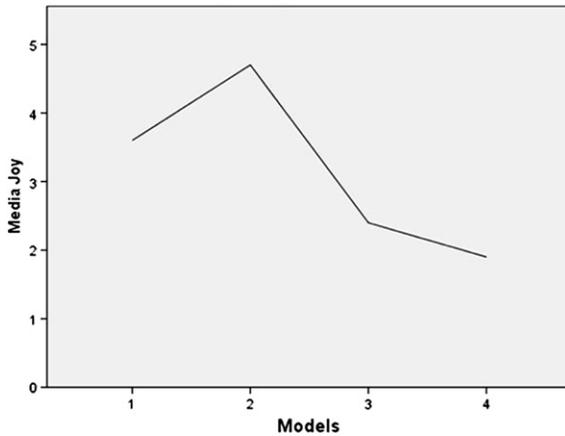
differences between the four models for each emotion, according to the descriptive statistics set out in Table 1.

In all cases, there is a model that attains a high assessment (between 4 and 5), whereby it can be concluded that a high identification is achieved with respect to the emotion studied. This supports the fact that there is a clear influence on the part of the two variables examined when generating emotion. Moreover, the total averages for each emotion make it possible to infer that the most recognizable emotion is sadness (mean [M] = 3.65, standard deviation [SD] = 1.19), followed by joy ($M = 3.15$, $SD = 1.33$), calmness ($M = 2.90$, $SD = 1.48$), and anxiety ($M = 2.77$, $SD = 1.37$). These results, in all cases, surpass the average assessment of 2.5, whereby, generally speaking, it can be considered that the various models attained a high degree of overall identification. The best identified model is M4 on sadness, whereas the least identified model is M3 on calmness.

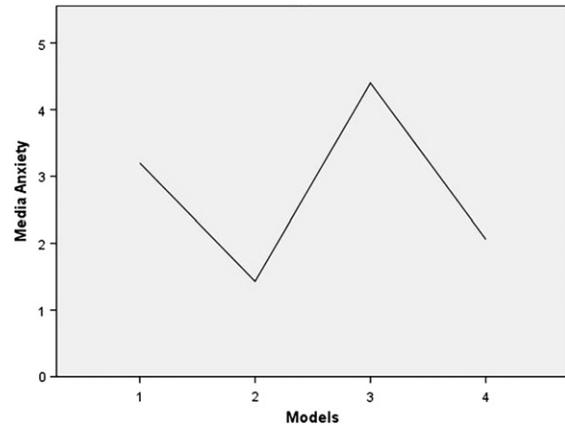
The analysis of these data by means of the analysis of variance (ANOVA) also leads to the conclusion that the differences obtained for the averages for each model for each emotion are highly significant—joy: $F(3, 399) = 264.55$, $P < 0.000$; anxiety: $F(3, 399) = 282.60$, $P < 0.000$; sadness: $F(3, 399) = 121.81$, $P < 0.000$; calmness: $F(3, 399) = 375.33$, $P < 0.000$. The tests conducted post hoc (Tukey and T3 Dunnet) make it possible to confirm this initial conclusion, because, for all the multiple comparisons, the significance level has been $P < 0.000$.

In relation to the analysis of the first emotion, most of the subjects surveyed consider M2 as being closest to joy, followed by M1, M3, and finally, M4. The differences among all models have been significant in all cases. The results can be seen in Graph 1.

The greatest sensation of joy is perceived in M2 ($M = 4.70$, $SD = 0.54$), which means that the sample audience actually perceived it as cheer, achieving a value that is much closer to the highest score of 5. Accordingly, the model identified this emotion to a large extent. This applies to the recording with a higher pitch level and with the greatest variations in contour



GRAPH 1. Plot of the mean values for joy.



GRAPH 2. Plot of the mean values for anxiety.

pitch (++)). For the representation of the intonation curve, it is clearly possible to observe that there are many inflections. Therefore, it is patent that, for joy, high pitch level and many variations in contour type are the decisive elements in creating the sensation. Nevertheless, M1 (—+), which has a lower pitch (102 Hz) than M3 (+—) (125 Hz), albeit with more variations in contour type (127 Hz compared with 110 Hz), is considered to be in second position. In this case, it seems that contour type is more decisive than the pitch level. The sample considers M1 to be closest to joy, because “the actor is realizing more movements.” As a result, reason is sustained in contour type with a greater number of tone inflections on the intonation curve. The last position is for M4 (—), with the lowest pitch level (101 Hz) and variation (74 Hz). The sample affirms that this actor sounds a “little too monotonous” to create joy.

In conclusion, pitch level and contour type are important elements in generating joy. The best manner is shaped by a higher pitch level with many variations. Nevertheless, one element that is not only necessary but also decisive is contour type, as demonstrated in the preference of M1 over M3. Likewise, it is possible to observe that the final order of election of the models established by the subjects (M2-M1-M3-M4) precisely matches the order for contour type from much to little variation (M2-M1-M3-M4), whereas the same is not true in the case of pitch level (M2-M3-M1-M4), which gives another indication regarding the decisive influence of contour type.

Second, for anxiety, the sample considers M3 to be the closest, followed by M1, M4, and M2. In this case also, the differences between each model were significant in all cases. The results are shown in Graph 2.

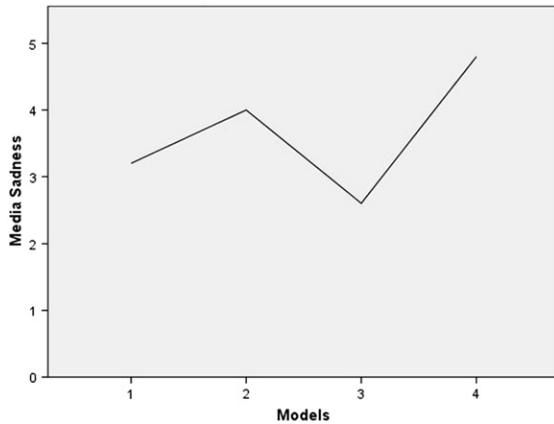
In this case, M3 ($M = 4.40$, $SD = 0.76$) achieved the best results, which, for the sample, sounded “highly nervous,” attaining close to 5, the maximum score. As with joy, it is the model with the highest pitch (155 Hz) and most variations in contour type (144 Hz) (++)). This model has many inflections. Therefore, again, both variables are representative in expressing this emotion. M1 (—+) is next when generating anxiety. It is the second recording with the greatest variations (115 Hz), but once again, it is not the second in terms of mean pitch (83 Hz) compared with M2 with a mean pitch of 85 Hz. Although the difference

is minimal in terms of frequency, the answers from the sample reveal that the most important variable is contour pitch and not pitch level. They consider that M1 is more nervous, because he “raises and lowers his voice more,” in other words, he makes many inflections. The third position is held by M4 (—), with the lower pitch (74 Hz) but with more inflections than M2 (105 Hz compared with 100 Hz). Again, the data demonstrate the importance of contour pitch. This conclusion is reinforced with the data regarding M2 (+—), which was considered to be less nervous because of fewer variations (100 Hz). The sample considers this a “very linear expression for it to be considered nervous.”

In conclusion, the results demonstrate that pitch level is important when it is accompanied by contour type. But the most decisive variable is actually the contour type with more regard to anxiety than to joy, because the model with the lowest pitch is in the third position. Moreover, as is the case with joy, the election order of each model (M3-M1-M4-M2) supports this conclusion, because it matches with the decreasing variation of contour type exactly (M3-M1-M4-M2) but not with pitch level (M3-M2-M1-M4).

Third, for sadness, the sample considered M4 as the closest, followed by M2, M1, and finally, M3. As in previous cases, the differences between each model were significant in all cases. The results are shown in Graph 3.

The model with the best results for sadness was M4 ($M = 4.80$, $SD = 0.44$). It is important to emphasize that this model has achieved the best results in all study areas, scoring closest to the highest score. This means that it has been the model that has best defined emotion. The selected model (—) has the lowest pitch (64 Hz) and the fewest variations in contour pitch (54 Hz). It is the most linear or monotonous model. The second position is held by M2 (+—), with a pitch level of 81 Hz, compared with the following selected model, M1 (—+), with 69 Hz, but with fewer inflections (77 Hz compared with 110 Hz). This means that the data for sadness once again show that the relevance of contour type is greater than that of pitch level. The last position is held by M3 (++)), with a mean pitch of 83 Hz and a variation of 97 Hz. Therefore, it is the model with the most inflections.

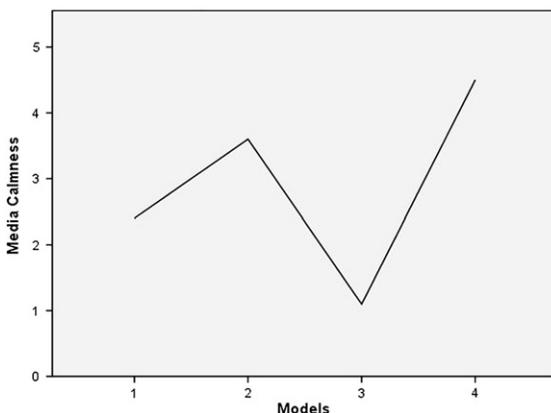


GRAPH 3. Plot of the mean values for sadness.

As with the remaining emotions, the results demonstrate the decisive influence of contour pitch, though with less variation in this case, to create a sensation of sadness. Here, the models were also chosen in the same order (M1-M3-M2-M4) as the decreasing variation for contour type (M1-M3-M2-M4), but not for pitch level (M3-M2-M1-M4), which once again shows the influence of the first variable on the results.

In the fourth instance, for calmness, the sample considered M4 to be the closest to this emotion, followed by M2, M1, and finally, M3. In this latter case also, the differences between each model have been significant. The results are shown in Graph 4.

M4 ($M = 4.50$, $SD = 0.74$) achieved the best results for calmness. As is the case with other emotions, this has been highly recognized, because it is closest to the best score of 5. This recording is characterized by a mean pitch of 70 Hz and a variation of 57 Hz (—). Therefore, it is not the model with the lowest pitch; this is in fact, M1 in third position. However, it is the model with the least variation. The second position is held by M2 (+—) with a mean pitch of 90 Hz and a variation of 85 Hz. It is the third in terms of the lowest pitch level and the second in terms of the fewest inflexions. M1 (—+) is in third position, with a low pitch of 68 Hz and a variation of 98 Hz.



GRAPH 4. Plot of the mean values for calmness.

Lastly, the lowest results recorded are held by M3 (++) , with 94 Hz, and especially, with many inflexions in creating this sensation (138 Hz). The sample considers M3 to be “too cheerful to generate calmness.”

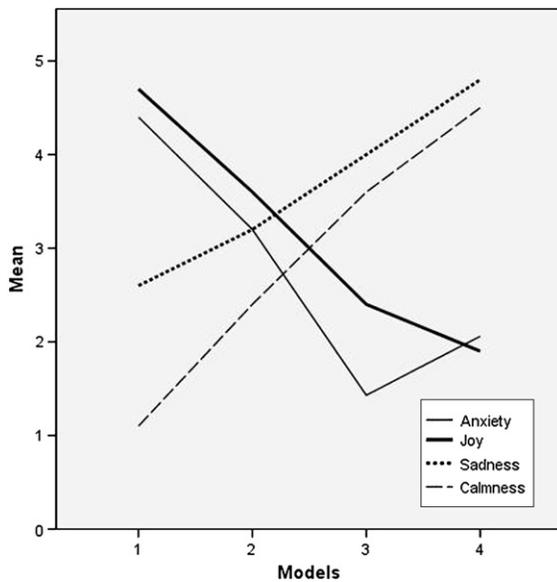
The data analyzed for this emotion show that the most decisive element for creating an emotion is contour type more than pitch level, which, in this case, did not show too many variations, as many inflexions can indicate joy rather than calmness. As in the other cases, the final order chosen by the subjects (M3-M1-M2-M4) was the same as for the variation of contour type according to a decreasing order (M3-M1-M2-M4), though there is no match for pitch level (M3-M2-M4-M1).

Overall, the data lead us to conclude that, for all emotions, the choice follows the same parameters, which also coincide with those envisaged regarding the manner in which they should all be produced in terms of pitch level and contour type. For this reason, for all emotions, the models best identified have maintained the combination ++ (joy and anxiety) and — (sadness and calmness). Also, for three emotions, the worst values were given for models with contradictory combinations — (joy) and ++ (sadness and calmness). The only exception to this was anxiety, the worst combination of which was +—. The differences among intermediate models have also shown a common trend. In all cases, contour type has been more decisive than pitch level. In the case of the models appearing in second position, they all maintain the same order: —+ (joy and anxiety) and +— (sadness and calmness). There is only a difference in the third position. Although joy (+—) and sadness and calmness (—+) maintain the same order in reverse, the only exception to the rule is anxiety, the third-position model that has a combination ——. For all emotions, the final election order of the sample matched the variation order for contour type and, in no instance, matched in the case of pitch level, thereby backing the hypothesis made in the beginning of this study. The re-coding of the models, arranging them uniformly for each emotion as M1 ++, M2 —+, M3 +—, M4 —, makes it possible to get a better overview of the overall results in Graph 5.

DISCUSSION

The analysis of the data from this research using an ANOVA indicates that the differences obtained in the averages for each emotion were highly significant, thereby affirming the statistical validity of the results. These data firstly reveal that, in all cases, a model that was closely associated with the represented emotion existed for values that were close to the maximum of 5: specifically between 4.5 and 4.8. Accordingly, this result shows the validity of the two variables examined—pitch level and contour type—to define these emotions, as certain previous studies have already determined.^{9–11}

Second, the general data reveal that the most recognized emotion has been sadness, followed by joy, calmness, and anxiety, with significant differences. For all four emotions, the average possible score has been surpassed; consequently, it is possible to conclude that the various models achieved a high level of overall identification. These results partly reinforce the data from Mozziconacci,² in which joy was less



GRAPH 5. Plot of the mean values for all emotions, namely, joy, anxiety, sadness, and calmness.

recognizable than sadness, as is the case with other researches in which this emotion obtained the best ratings for recognition.^{6,11,24} It is logical in a certain way to think that sadness is the most identifiable emotion, because, in the first place, it represents a basic sensation with a low level of activity sharing the fewest acoustic properties with other expressions of emotion, and its contour type does not have many combinations. Indeed, it is rather linear and, accordingly, easier to characterize and distinguish in acoustic terms.⁶ Another explanation for this clear recognition offered by authors, such as Waaramaa et al,⁵ is based on the fact that “humans have had to be aware of a possible threat and be sensitively attuned to negative signals especially.”

Similarly, joy, which is also a basic emotion, though with a high level of activity, is easier to perceive than calmness and anxiety, particularly because, as Aubergé and Cathiard²⁸ have proven, amusement is greatly audible (84%) compared with nonamusing stimuli. These data appear to contradict certain studies⁵ in which joy was the least recognized emotion. One possible explanation may stem from the difference in the choice to analyze emotions of a higher intensity (such as anger) in the case of other studies, allowing for these to be more identifiable compared with joy.

Third, calmness is a low-activity emotion with fewer identification characteristics than sadness or joy, because the level of activity is midway between these two emotions. Even so, there is also a model that has clearly identified with this emotion, because, as shown in other studies, neutral vocal expressions also have recognizable prosodic properties,⁶ albeit to a lesser extent.

Lastly, anxiety is more difficult to recognize, because it is pronounced with greater pitch variations and more combinations. Moreover, it is more difficult to find a standard model for anxiety on account of the various combinations existing. Nonetheless, its value is above the average and is close to calmness.

With respect to the variables analyzed, the study has shown that pitch level is less significant than contour type. Therefore, the main hypothesis is confirmed. In principle, from the data, it is possible to conclude that a higher level of similarity exists in the selection order of the models for all emotions, whereby it is possible to establish a direct link for the pitch level and contour type parameters conforming to the positive emotions (joy and anxiety) and those conforming to the negative ones (sadness and calmness). In this case, the results for each emotion bear the same trend, with the sole variation being in the level of identification: greater for joy compared with anxiety and greater for sadness compared with calmness. The difference between positive and negative emotions is that, this trend, according to the models proposed, is increasing for positive emotions and decreasing for negative emotions, whereby it operates in reverse.

The research data confirm the validity for the combination of the two prosodic elements, pitch level and contour type, when defining a specific emotion. This is shown by the fact that, for all emotions, the best results obtained were for the models in which both characteristics followed the same trend. As a result, the combination of high pitch level and high variation (++) for positive emotions (joy and anxiety) or the reverse—low pitch level and slight variation (--)—for negative emotions (sadness and calmness), obtained the best levels of identification. The same is true of the worst results, which, with the exception of anxiety, maintained the same order in reverse: low pitch level and slight variation (--) for joy and high pitch level and high variation (++) for sadness and calmness. However, the intermediate levels show, with greater accuracy, the influence of contour type on pitch level. All the models analyzed for the second position carried the same order—low pitch level and high variations (-+)—for joy and anxiety and the reverse—high pitch level and few variations (+-)—for sadness and calmness. In the case of the models in the third position, there is only a difference in the case of anxiety, the only emotion that did not follow the general trend. This position incorporates those elements combining high pitch level and few variations (+-) for joy and the reverse—low pitch level and high variation (-+)—for sadness and calmness. Accordingly, in conclusion, these data show that contour pitch has been much more influential than pitch level.

Another additional datum makes it possible to confirm this interpretation. For all the models proposed, in the case of absolutely all emotions, the final order of choice of the models established by the subjects matches the order of variation for contour type for each recording in an increasing manner, in other words, from the greatest to the least variation for positive emotions, joy and anxiety, and in a decreasing order from the least to the greatest variation for negative emotions, sadness and calmness. On the other hand, this order does not coincide for any emotions in terms of the order of frequency variation with regard to pitch level. This is a significant result when interpreting the influence of contour type on pitch level.

The only emotion for which the established model order has varied with respect to the remainder is anxiety. In this specific case, the model with the lowest values did not follow the combination low pitch level and slight variation (--); rather, it

combined high pitch level with the least variation (+−). Despite this difference, compared with the remaining emotions, the result can be again explained by resorting to the influence of contour type, because the model with the lowest values is the one with the lowest variation. Indeed, the sample considers it to be the least nervous, because “he maintains a highly linear expression.” Consequently, the result here is also brought about according to contour type.

Although the results of this study have been conclusive in terms of the goals set out in the scope, they may only be interpreted in the light of a laboratory experiment with all the limitations this entails. In this instance, the first limitation stems from the limited choice of four emotions. To supplement the results, it would be interesting to analyze other different emotions on which previous studies have been conducted (eg, anger or surprise) to verify whether the trends observed are the same for this research. Another interesting aspect to consider as a future line of study would be to analyze these emotions establishing different levels of intensities within each one (eg, extreme or contained joy). This would make it possible to establish different models for each emotion according to the levels and to verify whether the models established in this study, all of which have been identified, show any relationship with such a differing degree of intensity in emotion.

Likewise, it is clear that this study bears a major limitation in the combination of pitch level and contour type, which has been designed for assessment on the part of the sample. The intent has been to ensure that the models are different in terms of the combination of the two variables, but it is clear that the specific values for pitch and contour type can bear other different levels and, accordingly, many other combinations. Additionally, the study has demonstrated how difficult it is to isolate two variables that are inseparable for the spoken sequence to identify an emotion appropriately. Therefore, subsequent research should take care to address these aspects to enhance the results of this study.

CONCLUSIONS

The results of this research have shown that pitch level and contour type are valid prosodic elements that are highly significant when defining each of the emotions analyzed: joy, anxiety, sadness, and calmness. Second, the general data have revealed that the most recognized emotion has been sadness, followed by joy, calmness, and anxiety. For the four emotions, the average possible value has been surpassed, whereby it can be concluded that the various models attained a high overall level of identification. Last, and the most important, conclusion to be drawn is that the results of the study have confirmed the decisive influence of contour type when defining emotions over and above the other variable studied: pitch level. Although the two variables analyzed—pitch level and contour type—are representative of expressing emotions, contour type is more decisive. In all the models analyzed, contour type has been highly significant and is the variable that has been determined as the final order of recognition.

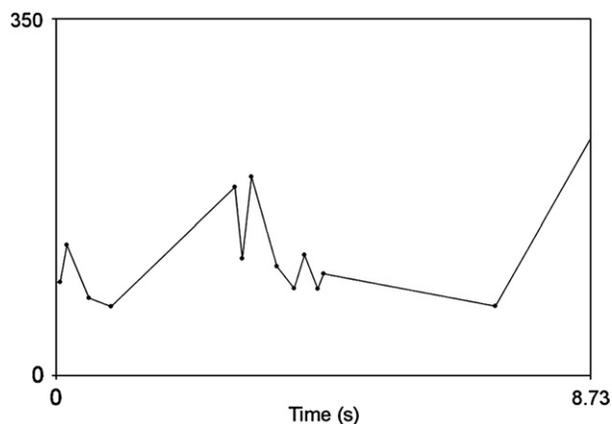
REFERENCES

1. Collier R. Intonation analysis: the perception of speech melody in relation to acoustics and production. *Proceedings of Eurospeech'89, First European Conference on Speech Communication and Technology*, Paris; 1989: 1038–1044.
2. Mozziconacci S. Prosody and emotions. *Speech Prosody 2002*. Aix-en-Provence, France (April 11–13); 2002:1–9.
3. Raithel V, Hielscher-Fastabend M. Emotional and linguistic perception of prosody. *Folia Phoniatr Logop.* 2004;56:7–13.
4. Cowie R, Cornelius R. Describing the emotional states that are expressed in speech. *Speech Commun.* 2003;40:5–32.
5. Waaramaa T, Laukkanen AM, Alku P, Väyrynen E. Monopitched expression of emotions in different vowels. *Folia Phoniatr Logop.* 2008;60:249–255.
6. Pell MD, Paulmann MD, Dara SC, Alasseri A, Kotzb SA. Factors in the recognition of vocally expressed emotions: a comparison of our languages. *J Phon.* 2009;37:417–435.
7. Bachorowski J, Owren MJ. Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context. *Psychol Sci.* 1995;6:219–224.
8. Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *J Acoust Soc Am.* 1972;52:1238–1250.
9. Rodero E. Sensaciones tonales. Empleo y significación de los niveles y tonemas en la entonación [Pitch sensations. Use and meaning of level and contour type in intonation]. *III Congreso de la Sociedad Acústica Forense (Congress of Spanish Acoustical Forensic Society)*. Santiago de Compostela, Spain; 2005:295–324.
10. Mozziconacci S. The expression of emotion considered in the framework of an intonation model. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion 2000*. Newcastle, UK; 2000:45–52.
11. Pell MD. Influence of emotion and focus location on prosody in matched statements and questions. *J Acoust Soc Am.* 2001;109:1668–1680.
12. Higuchi N, Hirai T, Sagisaka Y. Effect of speaking style of parameters of fundamental frequency contour. *J Acoust Soc Am.* 1994;95:2950.
13. Patterson D, Ladd DR. Pitch range modelling: linguistic dimensions of variation. In: *Proceedings of the 14th ICPhS San Francisco*; 1999:1169–1172.
14. Rodero E. *Locución Radiofónica*. [Voice on Radio]. Madrid, Spain: IORTV; 2003.
15. Gussenhoven C. Intonation and interpretation. Phonetics and phonology. *Speech Prosody 2002*. Aix-en-Provence, France (April 11–13); 2002:47–57.
16. Rodríguez A. Propuestas para una modelización del uso expresivo de la voz. [Proposals for a model of the expressive use of voice]. *Zer.* 2002;13:157–175.
17. Johnstone T, Scherer K. The effects of emotions on voice quality. In: *Proceedings of the 14th ICPhS San Francisco*; 1999:2029–2032.
18. Cantero FJ. *Teoría y análisis de la entonación*. [Theory and Analysis of Intonation]. Barcelona, Spain: UAB; 2002.
19. Sosa JM. *La entonación del español*. [Spanish Intonation]. Madrid, Spain: Cátedra; 1999.
20. Boula De Mareüil P, Célérier P, Toen J. Generation of emotions by a morphing technique in English, French and Spanish. *Speech Prosody 2002*. Aix-en-Provence, France (April 11–13, 2002); 2002:187–190.
21. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull.* 2003; 129:770–814.
22. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol.* 1996;70:614–636.
23. Sobin C, Alpert M. Emotion in speech: the acoustic attributes of fear, anger, sadness, and joy. *J Psycholinguist Res.* 1999;23:347–365.
24. Scherer KR. Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 2003;40:227–256.
25. Boersma P, Weenink D. Praat: doing phonetics by computer. Phonetic Sciences, University of Amsterdam. Available at: <http://www.praat.org>. Accessed January 20, 2010.
26. Ehrette T, Château N, D'Alessandro C, Maffioli V. Prosodic parameters of perceived emotions in vocal server voices. *Ergonomics.* 2003;46:1375–1385.
27. Pereira C. Dimensions of emotional meaning in speech. *ITRW Speech Emotion 2000*. Newcastle; 2000:25–28.
28. Aubergé V, Cathiard MA. Can we hear smiles? *Speech Commun.* 2003;40: 87–97.

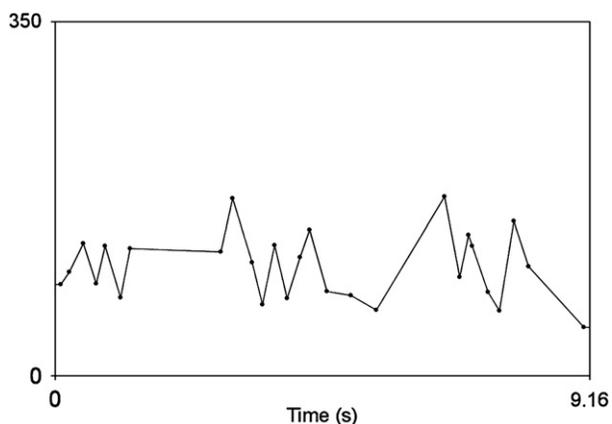
Appendix

TABLE A1.
Data of the Corpus

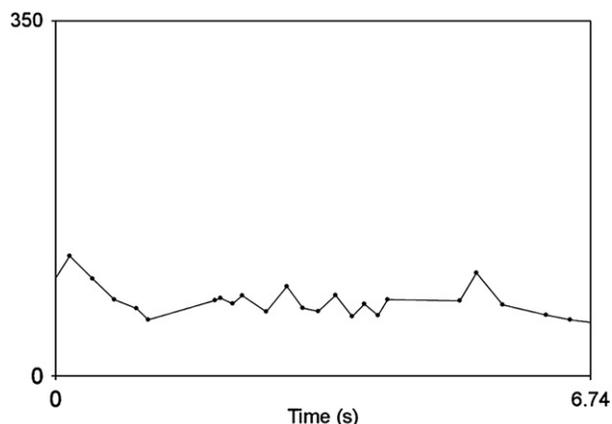
Emotion	Model	Pitch Level (Hz)	Contour Type (Hz)	Combination
Joy	M1	102	127	-+
	M2	128	168	++
	M3	125	110	+ -
	M4	101	74	--
Anxiety	M1	83	115	-+
	M2	85	100	+ -
	M3	155	144	++
	M4	74	105	--
Sadness	M1	69	110	-+
	M2	81	77	+ -
	M3	83	97	++
	M4	64	54	--
Calmness	M1	68	98	-+
	M2	90	85	+ -
	M3	94	138	++
	M4	70	57	--



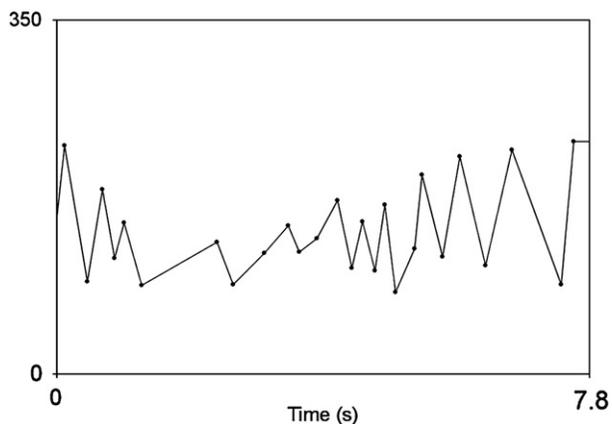
GRAPH A3. Contour type of joy, model 3.



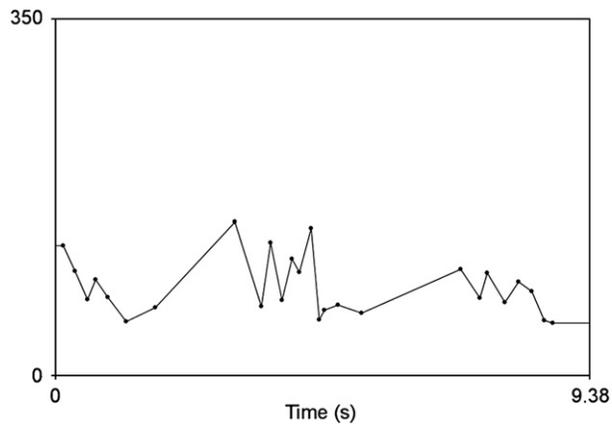
GRAPH A1. Contour type of joy, model 1.



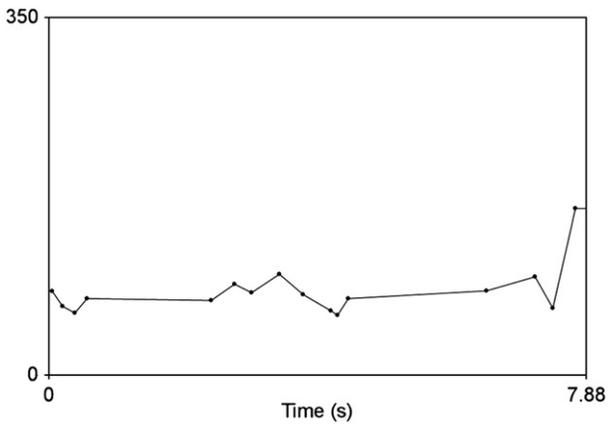
GRAPH A4. Contour type of joy, model 4.



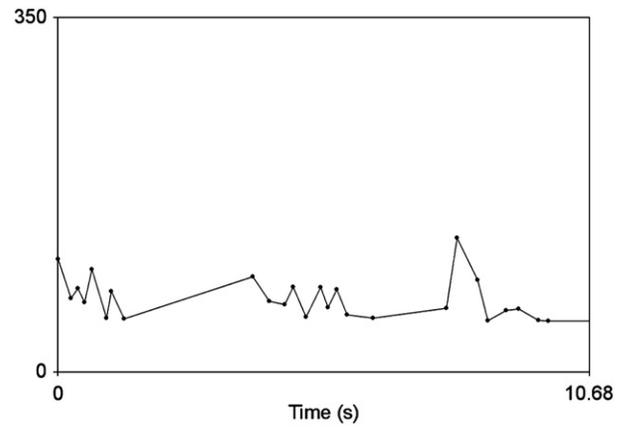
GRAPH A2. Contour type of joy, model 2.



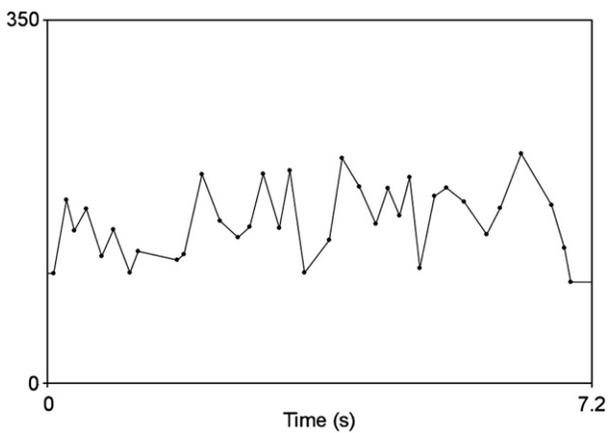
GRAPH A5. Contour type of anxiety, model 1.



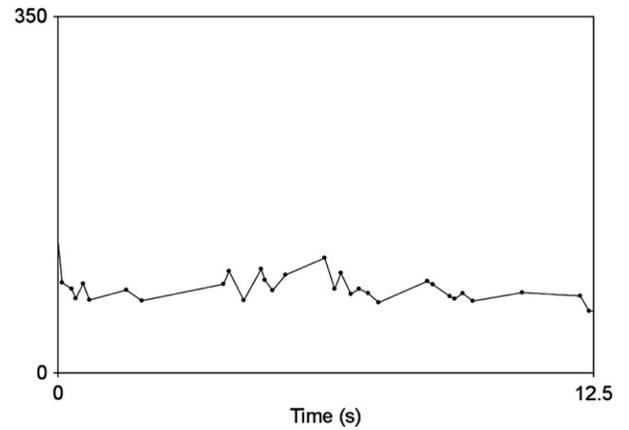
GRAPH A6. Contour type of anxiety, model 2.



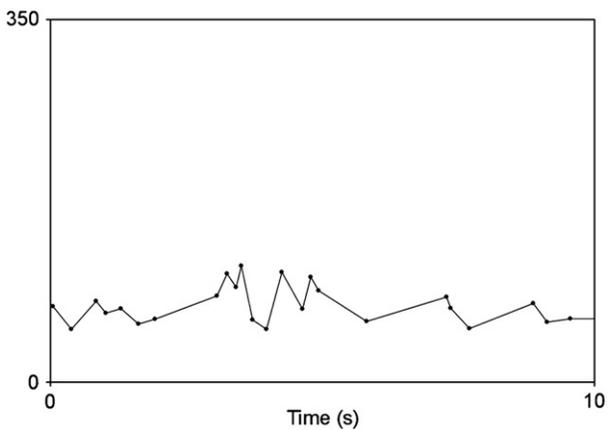
GRAPH A9. Contour type of sadness, model 1.



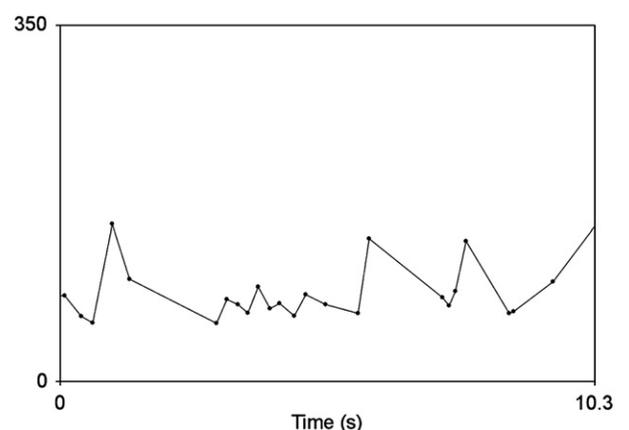
GRAPH A7. Contour type of anxiety, model 3.



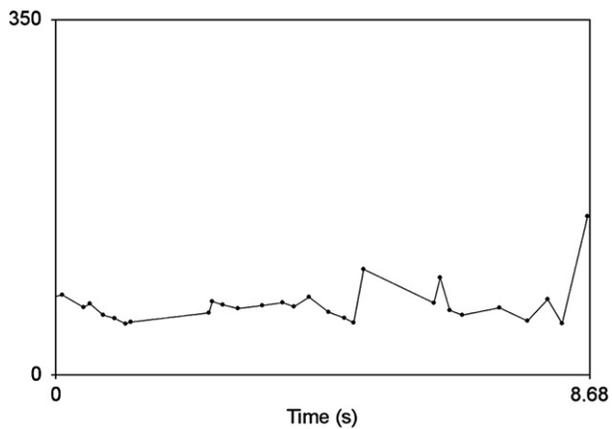
GRAPH A10. Contour type of sadness, model 2.



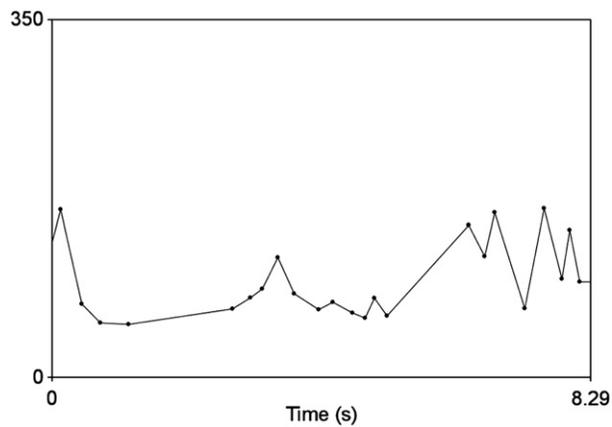
GRAPH A8. Contour type of anxiety, model 4.



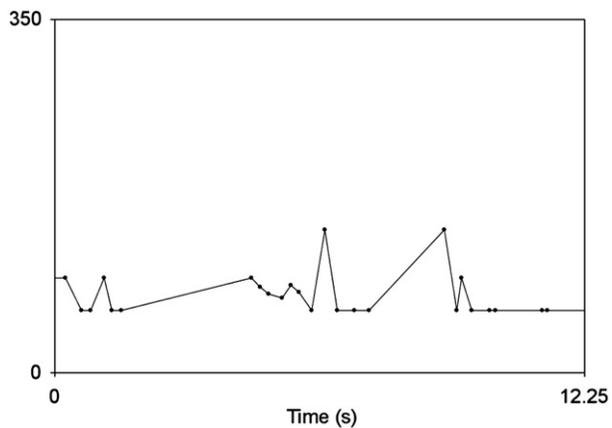
GRAPH A11. Contour type of sadness, model 3.



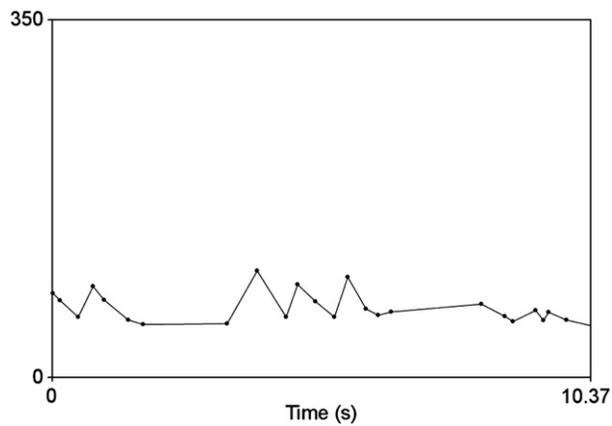
GRAPH A12. Contour type of sadness, model 4.



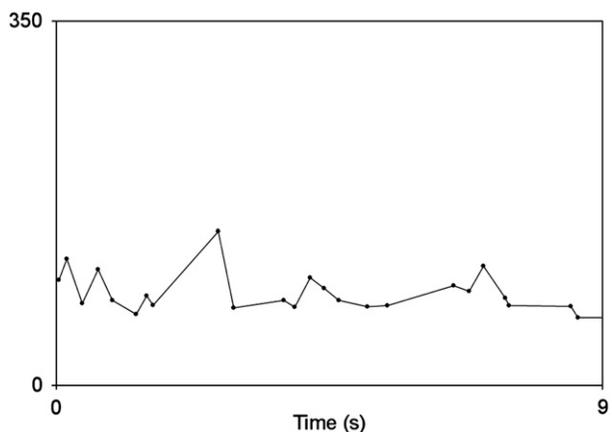
GRAPH A15. Contour type of calmness, model 3.



GRAPH A13. Contour type of calmness, model 1.



GRAPH A16. Contour type of calmness, model 4.



GRAPH A14. Contour type of calmness, model 2.