

Perception of emotional prosody: investigating the relation between the discrete and dimensional approaches to emotions

Percepção da prosódia emocional: investigando a relação entre as abordagens discreta e dimensional das emoções

Wellington da Silva

Universidade Estadual de Campinas, Campinas, São Paulo / Brasil

wellingtondasilva@rocketmail.com

Plinio Almeida Barbosa

Universidade Estadual de Campinas, Campinas, São Paulo / Brasil

pabarbosa.unicampbr@gmail.com

Abstract: Emotional phenomena can be described according to various psychological approaches, the most adopted being the discrete (basic) and the dimensional ones. This study aimed at investigating the relation between the perception of some basic emotions and emotional dimensions in speech, as well as determining which acoustic cues are related to their perception. We conducted two perception experiments with utterances selected from a foreign language (Swedish) of which the listeners had no knowledge. In the first one, Brazilian subjects rated on 5-point scales the expressivity of four basic emotions: joy, anger, sadness, and calmness. In the second, a distinct group of Brazilian subjects rated the expressivity of five emotional dimensions: activation, fairness, valence, motivation, and involvement. The perception of the basic emotions and of the emotional dimensions was then compared by means of the Spearman's correlation coefficient. The five emotional dimensions were significantly correlated to some extent with the basic emotions, and these correlations were, in general, consistent with the literature and with the hypotheses that guided this study. We also performed an acoustic analysis,

in which twelve acoustic parameters were automatically computed for the utterances evaluated by the listeners. The parameters which correlated better with the listeners' judgments were fundamental frequency (median, interquartile semi-amplitude, 99.5% quantile), spectral tilt (mean and standard deviation), and LTAS slope. We concluded that it is possible to describe the perception of basic emotions in speech as a combination of emotional dimensions and that emotional dimensions may be better for describing the expression of emotions in speech.

Keywords: emotional prosody; basic emotions; emotional dimensions; perception test.

Resumo: Os fenômenos emocionais podem ser descritos de acordo com várias abordagens psicológicas, sendo a discreta (básica) e a dimensional as mais adotadas. Este estudo teve como objetivo investigar a relação entre a percepção de algumas emoções básicas e dimensões emocionais na fala, bem como determinar quais pistas acústicas estão relacionadas com sua percepção. Conduziram-se dois experimentos de percepção com enunciados selecionados de uma língua estrangeira (sueco) da qual os ouvintes não possuíam nenhum conhecimento. No primeiro, sujeitos brasileiros julgaram em escalas de cinco pontos a expressividade de quatro emoções básicas: alegria, raiva, tristeza e calma. No segundo, um grupo distinto de sujeitos brasileiros avaliou a expressividade de cinco dimensões emocionais: ativação, justiça, valência, motivação e envolvimento. A percepção das emoções básicas e das dimensões emocionais foi então comparada por meio do coeficiente de correlação de Spearman. As cinco dimensões emocionais correlacionaram-se significativamente em algum grau com as emoções básicas e essas correlações foram, no geral, consistentes com a literatura e com as hipóteses que nortearam este estudo. Realizou-se também uma análise acústica, na qual doze parâmetros acústicos foram computados automaticamente para os enunciados avaliados pelos ouvintes. Os parâmetros que melhor se correlacionaram com os julgamentos dos ouvintes foram: frequência fundamental (mediana, semi-amplitude entre quantis, quantil 99,5%), inclinação espectral (média e desvio padrão) e inclinação do LTAS. Concluiu-se que é possível descrever a percepção das emoções básicas na fala como uma combinação de dimensões emocionais e que as dimensões emocionais podem ser melhores para descrever a expressão de emoções na fala.

Palavras-chave: prosódia emocional; emoções básicas; dimensões emocionais; teste de percepção.

Recebido em: 17 de agosto de 2016.

Aprovado em: 16 de novembro de 2016.

1 Introduction

Emotions are complex phenomena which have intrigued researchers from different fields for decades. Many theories have been developed to explain what an emotion is (CORNELIUS, 2000). Currently, most scholars assume a componential view, for which emotions are (brief) processes made up of several components. This position differs, for example, from the one shared by the common sense, which sees emotion only as a feeling, that is, as the subjective experience of the state of emotional arousal caused by an event (SCHERER, 2000). The following components have been postulated (SCHERER, 2000): feeling, neurophysiological response patterns, motor expressions, action readiness, and appraisal. Mesquita & Frijda (1992) include in this list the antecedent events, event coding and regulation.

According to this perspective, an emotional process is triggered by the appraisal of an event in the environment (the antecedent event) as being relevant for the goals and needs of the individual. The appraisal of the event is automatic and unconscious, *i.e.*, it is not dependent on the deliberate action of the individual. However, it is mediated by the event coding, which is the meaning that the antecedent event has to the organism. Such process depends on the knowledge shared by the culture, the history of life, the personality and the personal beliefs of the person. Therefore, the emotion which the person will experience does not depend on the nature of the event itself, but rather on how it is coded by the individual (MESQUITA; FRIJDA, 1992). This is why the same event can trigger different emotions in different cultures and even between different people. Emotions are also subject to the regulation of the individual, who can try to disguise an expression (as is the case with shame) or exaggerate it (when someone tries to fake joy, for example). This process often depends on cultural and social norms, but it does

not inhibit the emotional process (for example, even disguising the expression, the person is still experiencing the emotion).

Researchers have also sought to find the most suitable way of describing the emotional processes and several methods have been developed for this purpose (COWIE; CORNELIUS, 2003). In the present study, we compare the two most influential approaches: the basic emotions approach and the dimensional. One of the big debates in emotion research is whether variations in the emotional processes should be regarded as gradual differences on a set of underlying dimensions or as discrete differences among emotion categories (FONTAINE, 2009).

According to the *basic emotions approach*, there exists a small set of universal emotions (known as basic or discrete), which are qualitatively different from each other and characterized by specific patterns of cognitive appraisal, expression, and physiological changes (EKMAN, 1992). This approach has its roots in the evolutionary theory of emotions (DARWIN, 2009[1872]), which understands that emotions are evolved phenomena which have been selected in the course of the evolution of the species because they help us to cope with relevant events in the environment, such as telling the individual when to attack, defend, flee, reject food, etc. As a result, one expects to find the same patterns of expression (facial and vocal) for a given emotion in all human cultures (CORNELIUS, 2000; MATSUMOTO; EKMAN, 2009). The basic emotions are described by means of labels provided by the natural languages, such as joy, sadness, anger, etc. Each of these labels actually denotes a family of related emotions. For example, the anger family is composed of emotions described by the labels angry, annoyed, irritated, furious, etc. (EKMAN, 1992; MATSUMOTO; EKMAN, 2009). However, there is no absolute agreement between researchers as to which emotions are “basic”. The most known proposal is Paul Ekman’s (EKMAN, 1992), which recognizes six basic emotions: joy, sadness, fear, disgust, anger, and surprise.

The *dimensional approach* understands that emotions are best represented by positions within continuous scales specified by underlying *emotional dimensions*. This approach, thus, emphasizes the gradual nature of the emotional phenomenon, which can occur with varying levels of intensity, and not only with two discrete poles of minimum and maximum intensity (FONTAINE, 2009; SCHLOSBERG, 1954). The most known emotional dimensions are *activation* and *valence*. Activation

(also labeled as arousal) refers to the degree of arousal of the individual (increase in behavior or physiological activity) and varies from calm to agitated (FOWLES, 2009; SCHLOSBERG, 1954). Valence (also known as pleasantness) corresponds to the subjective feeling of the degree of pleasantness caused by the antecedent event and emotions are commonly distinguished within this dimension as either positive (pleasant) or negative (unpleasant). Examples of positive emotions are joy and pride, and examples of negative emotions are anger, fear, shame (BROSCH; MOORS, 2009). It is possible to integrate the basic emotions approach with the dimensional by considering that each basic emotion can be represented by a set of underlying dimensions (MAUSS; ROBINSON, 2009; SMITH; ELLSWORTH, 1985). For example, the emotion labeled by the discrete approach as “anger” has negative valence and a degree of activation that varies from medium to high. The emotion “joy”, in turn, has positive valence, whereas “sadness” has negative valence and a low degree of activation.

In addition to valence and activation, many more emotional dimensions have been proposed in the literature (see, for example, FRIJDA *et al.*, 1995; SMITH; ELLSWORTH, 1985). Barbosa (2009) used the dimension of *involvement*, which is related to the degree of involvement of the individual with the event and is analogous to the opposition between attention – rejection used by Schlosberg (1941). A dimension that is also used is *dominance*, which refers to the capability of the individual to handle the situation, that is, whether he/she is in the control of the event or controlled by it (KEHREIN, 2002). Dominance has been regarded as an important dimension, as in some cases it is the only dimension that distinguishes pairs of emotions, such as anger from fear (SMITH; ELLSWORTH, 1985). However, some studies have found that dominance may not be well inferred from speech. Barbosa (2009) obtained very low inter-rater agreement for this dimension. Amir *et al.* (2010) obtained very low rates of automatic recognition (by means of combination of acoustic parameters) for the dimensions of dominance and valence, concluding, thus, that the acoustic parameters were better at predicting the dimension of activation than these other two dimensions. A similar result was found by Lugger & Yang (2007). Therefore, we decided here to replace dominance by other dimensions, in order to assess whether they can also be inferred from speech.

Following Frijda *et al.* (1995), we investigate in this study the perception of the dimensions of *fairness* and *motivation*. Fairness is related to the appraisal of the eliciting event by the individual, *i.e.*, whether the individual considered what happened fair or unfair. Motivation is a dimension related to action readiness, *i.e.*, whether the eliciting event enhanced or diminished the individual's disposition to act on the event. It has been suggested that the appraisal of an event as unfair can trigger and increase the intensity of various emotions, especially anger (ELLSWORTH; SCHERER, 2003, p. 581). In addition, the emotion the individual is expressing may also signal his/her behavioral intentions, such as approaching, avoidance, or touching, as well as his/her disposition to change his/her behavior (FRIJDA *et al.*, 1995).

Emotions can be expressed through different modalities, such as facial, body posture and gesture, and vocal (SCHERER; ELLGRING, 2007). This work addresses the expression of emotions in speech. The physiological responses triggered by emotions cause variations in respiration, phonation, and articulation, which are processes directly related to speech (SCHERER, 1986). In addition, the cognition of the speaker can also be affected by the emotional episodes, and this can have an impact on the temporal characteristics of speech (JOHNSTONE; SCHERER, 2000). Research on speech and emotion has focused on investigating the emotion-related changes in speech and whether listeners are able to recognize the intended emotion based only on speech samples (SCHERER, 2003). A common approach in this area is to measure acoustic properties of speech, such as amplitude, duration, and fundamental frequency, which we refer to in this work as "acoustic parameters". It is assumed that different emotions are characterized by specific patterns of changes in the acoustic parameters (see JOHNSTONE; SCHERER, 2000; and SCHERER, 2003, for a review of empirical findings on acoustic patterns for some basic emotions). The most studied parameters in the literature have been those related to vocal fold vibration (fundamental frequency), time (speech rate, duration of utterances and pauses) and intensity (perceived amount of energy in the speech signal). Because *prosody*, a branch of linguistics, is the area which studies these properties (BARBOSA, 2012), the term "emotional prosody" is often used to refer to utterances which convey emotional information.

With regard to the perception of emotional prosody, many studies which followed the basic emotion perspective have shown that listeners

can satisfactorily recognize a number of basic emotions through speech, even in intercultural contexts (*e.g.* BANSE; SCHERER, 1996; PITTAM; SCHERER, 1993; SCHERER; BANSE; WALLBOTT, 2001). Recent studies have demonstrated that some emotional dimensions can also be inferred from speech (LAUKKA; JUSLIN; BRESIN, 2005; PEREIRA, 2000; BARBOSA, 2009). Laukka, Juslin, & Bresin (2005) had listeners rate utterances conveying five acted emotions (anger, happiness, sadness, fear, and disgust) with respect to four emotional dimensions (activation, valence, potency, and intensity). The authors found that these emotions were associated with different patterns of judgments for these emotional dimensions, which suggests that these basic emotions can be described and distinguished by means of dimensions. In addition, some acoustic correlates were found in the utterances for all four dimensions. Laukka & Elfenbein (2012) have found that emotional dimensions related to the appraisal of the emotion-eliciting events (*e.g.* valence, novelty, urgency, goal conduciveness, etc.) can also be inferred reliably from vocal expressions, which suggests that speech can also signal information about the cognitive representation of events.

In spite of the great interest of speech researchers in the discrete and dimensional approaches, it is still unclear which of the two perspectives is more suitable to describe the expression of emotions in speech and there is a lack of studies which have directly compared these approaches in order to shed light on this question. Some authors have suggested that emotional dimensions may be better for distinguishing and describing the vocal expression of emotions than labels of discrete emotions, since some dimensions (*e.g.* activation) directly reflect the emotion-related changes in physiology (COWIE; CORNELIUS, 2003; SCHERER, 1986). In addition, emotions which have a similar pattern for some dimensions (*e.g.* activation and valence) share the same changes for some acoustic parameters (*e.g.* fundamental frequency and intensity) and this may cause confusion when trying to describe these emotions by means of discrete labels (LUGGER; YANG, 2007; PEREIRA, 2000). Emotional dimensions may be useful for research on speech synthesis and automatic recognition of emotions from speech, since their continuous quality allows the modeling of weaker emotional expressions, as well as gradual changes in speech expressiveness over time (GRIMM *et al.*, 2007; SCHRÖDER *et al.*, 2001; BARBOSA, 2009).

The present study was conducted to address these questions and thus contribute to the literature on emotional prosody. We carried out two perception experiments with utterances selected from spontaneous speech of a foreign language (Swedish). The use of utterances of an unknown foreign language is important to guarantee that listeners rely only on the prosody of the utterances to make their judgments (this is discussed in section 2.1). In the first experiment, Brazilian subjects rated on 5-point scales the expressivity of four basic emotions: joy, anger, sadness, and calmness. In the second, a distinct group of Brazilian subjects rated the expressivity of five emotional dimensions: activation, fairness, valence, motivation, and involvement. Our objectives were to investigate the relation between the perception of these basic emotions and emotional dimensions in speech, and to determine which acoustic cues among those investigated here are related to their perception. Based on the studies reported above, we hypothesize that each of these basic emotions is associated with a distinct pattern of perceptual judgments for these emotional dimensions and that the ratings for the discrete emotions and emotional dimensions correlate to some extent with some acoustic parameters extracted from the utterances.

2 Materials and methods

2.1 Authentic emotional speech

The corpus used in the present study consists of 40 speech samples from 5 Swedish female speakers,¹ each one with duration between 1 and 6 seconds and with acceptable quality for performing acoustic analysis. It was set up for a cross-cultural study on the perception of emotions through spontaneous speech which was carried out with Brazilian and Swedish listeners and showed no difference in the perception of the emotions between the subjects of both cultures, that is, the Brazilian listeners' perception of the emotions expressed by the speakers of this corpus was very similar to the native speakers' (SILVA; BARBOSA;

¹ The programs from which the utterances were extracted did not provide information regarding the age of the participants. According to our perception, the speakers were aged between 30 and 60 at the time of the recording.

ABELIN, 2016).² The reason for using this corpus is that the subjects who took part in the perception experiments reported here had no knowledge of the Swedish language. The use of utterances of an unknown foreign language prevents listeners from using words referring to emotions as clues to determine the emotional state of the speaker. Therefore, the listeners could rely only on the prosody and voice quality of the utterances to make their judgments.

The utterances of this corpus were extracted from spontaneous speech (talk shows and interviews) of the Swedish television and of one Swedish interview program which was freely available over the internet as podcasts. They were saved on the hard drive into wave sound format with a sampling frequency of 44.1 kHz (Mono).

2.2 Perception experiment I: basic emotions

2.2.1 Participants

13 subjects completed this experiment (7 women and 6 men). All of them were born and have lived most part of their lives in Brazil and have Portuguese as their mother language. They were either undergraduate or graduate students and reported having no hearing impairment. The average age of the judges was 28 years, ranging from 21 to 50 years. They also reported having no knowledge of the Swedish language.

2.2.2 Procedure

In this experiment, subjects were asked to rate on 5-point scales ranging from 0, “not at all adjective”, to 4, “very adjective”, the degree with which the speaker in each stimulus was expressing the discrete

² Studies comparing the perception of emotions through speech between speakers of Portuguese and of other languages are rare, but we can also mention the work by Peres (2014), which compared the perception of emotions expressed in Brazilian Portuguese utterances between Brazilian and English subjects. In this study, both groups recognized the emotions with better-than-chance rate, but the native listeners (Brazilian) performed better (90% against 66%).

emotions described by four adjectives: *joyful*, *angry*, *sad*, and *calm*.³ Therefore, the experiment consisted of 4 parts, which were completed in a single session. In each part the listeners evaluated one emotional adjective for all 40 speech samples. The stimuli were presented in a random order, but the adjectives were evaluated by all listeners in the order presented above. The experiment was developed in Portuguese and carried out over the internet through the “Survey Gizmo” online software (<http://www.surveygizmo.com/>). The link for accessing it was sent by email to the subjects who were interested in taking part. They were asked to use earphones and to do the experiment in a quiet room. One speech sample was presented on each screen along with its corresponding scale and it was reproduced automatically as the page was finished loading. After listening to the utterance, the subjects had to mark their response on the scale by clicking on the desired value and then click on the “next” button at the bottom of the page to proceed to the next stimulus. It was not possible to return to the previous page or to proceed to the next one without having marked the response on the scale.

2.3 Perception experiment II: emotional dimensions

2.3.1 Participants

This experiment was carried out with a group of judges different from that of experiment I. 20 subjects completed the experiment (7 men and 13 women). Their average age was 25 years, ranging from 18 to 34 years. They were either graduate or undergraduate students, were born and have lived most part of the life in Brazil and have Portuguese as their mother language. They reported having no knowledge of the Swedish language at all and no hearing impairment.

2.3.2 Procedure

In this experiment, judges were asked to rate on 5-point scales ranging from 0, “not at all adjective”, to 4, “very adjective”, the degree

³ The adjective “calm” is used here to balance the emotional space under investigation with an item that corresponds to an emotion of positive valence and low level of arousal. Thus, we have a representative label for each quadrant of the arousal-valence space (see, for example, James Russell’s circumplex model of affect (RUSSELL, 1980).

with which the speaker in each stimulus was expressing the emotional state described by emotional dimensions. The experiment consisted of 5 parts, which were carried out in a single session. In each part the listeners evaluated one emotional dimension for all 40 speech samples. The dimensions investigated in this experiment were: *activation* (“How agitated was the speaker?”), *fairness* (“How fair did the speaker consider what happened?”), *valence* (“How pleasant for the speaker was the situation he/she was in?”), *motivation* (“How motivated to act on the situation was the speaker?”), and *involvement* (“How involved is the speaker with the situation he/she was in?”). The stimuli were presented randomly, but the emotional dimensions were evaluated by all listeners in this order. After listening to each utterance, the subjects had to judge the degree of expressivity of the emotional dimension specific to that part of the experiment by clicking on the desired value of the 5-point scale and then click on the “next” button at the bottom of the page to listen to the next stimulus. The experiment was developed and run in Portuguese through the “Survey Gizmo” online software. The questions related to the dimensions (shown above) were presented along with the scales to guide the judges. The remaining of the procedure is the same as that followed in experiment I.

3 Analyses and results

To ensure comparable magnitudes with the normalized values of the acoustic parameters, the judges’ responses were linearly converted to a scale ranging from 0 to 1 (0, 0.25, 0.50, 0.75, and 1). Then, the 5-level responses were transformed into three categories: low (0 and 0.25), medium (0.50), and high (0.75 and 1). This was done to avoid the influence of outliers on the mean ratings for each utterance to be used in the following analyses, thus guaranteeing more reliable judgments. The statistical analyses reported in this paper were performed with the software R in its 3.1.2 version (R CORE TEAM, 2014).

3.1 Inter-rater reliability

The reliability of the listeners’ responses in both experiments was verified by means of the Fleiss’ kappa index (FLEISS, 1971), which is an index bounded between 0 and 1 (the closer to 1, the greater the agreement). The index was calculated separately for each adjective, considering the three categories of responses.

Table 1 shows the kappa figures for the discrete emotional labels and emotional dimensions as well as their corresponding z figure (this test is statistically significant for $\alpha = 0.001$ when $z > 3.09$). The agreement between the listeners' responses was statistically different from 0 for all adjectives ($p < 0.001$) and similar to other studies on the perception of emotions through speech (ALM; SPROAT, 2005; DEVILLERS *et al.*, 2006; BARBOSA, 2009).

TABLE 1
Kappa index for the discrete emotional labels and emotional dimensions and their corresponding z value

Experiment I		
Adjective	Kappa	z
<i>enraivecido</i> (angry)	$k = 0.465$	$z = 34.0$
<i>calmo</i> (calm)	$k = 0.408$	$z = 30.4$
<i>triste</i> (sad)	$k = 0.281$	$z = 21.0$
<i>alegre</i> (joyful)	$k = 0.231$	$z = 17.7$
Experiment II		
Adjective	Kappa	z
<i>agitado</i> (agitated)	$k = 0.359$	$z = 43.9$
<i>agradável</i> (pleasant)	$k = 0.280$	$z = 34.0$
<i>motivado</i> (motivated)	$k = 0.258$	$z = 31.1$
<i>justo</i> (fair)	$k = 0.199$	$z = 24.1$
<i>envolvido</i> (involved)	$k = 0.186$	$z = 22.4$

3.2 Comparing the perception of basic emotions and of emotional dimensions

In order to evaluate the relation between the perceptual judgments for the discrete emotions and emotional dimensions, correlation coefficients were computed between the judges' mean ratings for each utterance for the discrete emotions and for the emotional dimensions. The correlation coefficient measures the degree of association between two variables (DOWDY; WEARDEN; CHILKO, 2004).

Because some of the variables did not meet the normality condition for their distribution (according to the Shapiro–Wilk normality test), we used the Spearman's rank correlation (ρ), which is a nonparametric

alternative to the Pearson product-moment correlation. To perform this test the variables were transformed into ranks. Results are shown in Table 2. The Spearman's rank correlation coefficient is bounded between -1 and 1 , with 1 indicating perfect positive correlation (*i.e.* a given observation of one variable has the same rank order as the corresponding observation of the other variable) and -1 expressing perfect negative correlation (*i.e.* when a given observation of one variable has a high rank order, the corresponding observation of the other variable has a low rank order). The correlations were statistically significant (*i.e.* different from 0) for almost all discrete emotion – emotional dimension pairs.

The ratings for *joy* were positively correlated with fairness, valence, and motivation, which is in accordance with theoretical predictions which state that joy is an emotion of positive valence (ELLSWORTH; SCHERER, 2003; LAUKKA; ELFENBEIN, 2012; SCHERER, 1986). This indicates that according to the listeners' perception the speakers who were joyful were motivated to act on the situation and considered the situation pleasant and fair.

Anger was negatively correlated with fairness and valence and positively correlated with activation, motivation, and involvement. This means that there was a tendency for the utterances which were evaluated by the listeners with high level of anger to be also rated with high level for the dimensions of activation, motivation, and involvement (and thus with low level for the dimensions of fairness and valence). According to the listeners' perception, these speakers were very agitated, very involved and very motivated to act on the situation and did not consider the situation neither pleasant nor fair. This result is also in accord with the theoretical predictions and evidence presented in the literature for the relation of this basic emotion to valence, fairness, and activation (ELLSWORTH; SCHERER, 2003; LAUKKA; ELFENBEIN, 2012; SCHERER, 1986).

As expected, *sadness* was negatively correlated with activation, motivation, and involvement. Based on the literature, we also expected a negative correlation of this basic emotion with fairness and valence (since one expects that a person who is sad must have considered the event unfair and unpleasant). Although this was the case with the latter, the correlation with these two dimensions was not statistically significant.

Calmness was also negatively correlated with activation, motivation, and involvement. It differed from sadness for being positively correlated with fairness and valence, which was expected. In addition,

calmness was exactly the opposite of anger with regard to its correlation with the five emotional dimensions.

TABLE 2
Spearman's rank correlations between the listeners' judgments
(mean ratings for each utterance) for the discrete emotions
and for the emotional dimensions

	Activation	Fairness	Valence	Motivation	Involvement
Joy	0.17 ns	0.36*	0.56***	0.44**	0.23 ns
Anger	0.88***	-0.81***	-0.70***	0.69***	0.79***
Sadness	-0.49**	0.03 ns	-0.09 ns	-0.62***	-0.37*
Calmness	-0.94***	0.82***	0.69***	-0.72***	-0.82***

* p < 0.05 ** p < 0.01 *** p < 0.001 ns = not significant

It is also useful to analyze how the emotional dimensions of fairness, motivation, and involvement relate to the well-known dimensions of activation and valence according to the listeners' perception. The correlations between the basic emotions and the emotional dimensions presented in Table 2 suggest that fairness is more related to valence (*i.e.* they tended to be positively correlated with each other) whereas motivation and involvement seem to be more linked to activation. To better investigate this, the intercorrelations among the listeners' mean ratings for the emotional dimensions were also computed through the Spearman's rank correlation coefficient. Results are shown in Table 3. As expected, fairness presented a high and positive correlation coefficient with valence, whereas motivation and involvement presented high and positive correlations with activation and with each other.

This result indicates that the listeners' perception indeed separated fairness and valence from involvement, motivation, and activation. Put differently, there was a tendency for the utterances which were evaluated by the listeners with high values of the rating scales for the dimension of fairness to be also rated with high values for the dimension of valence. Conversely, the speech samples which were evaluated by the listeners with high values for activation tended to be also rated with high values for the dimensions of motivation and involvement.

TABLE 3
Intercorrelations (Spearman's rho) among the emotional dimensions assessed by the subjects in experiment II

	activation	fairness	valence	motivation	involvement
activation	—				
fairness	-0.77***	—			
valence	-0.58**	0.88***	—		
motivation	0.82***	-0.45*	-0.23 ns	—	
involvement	0.88***	-0.65***	-0.50**	0.82***	—

* $p < 0.01$ ** $p < 0.001$ *** $p < 10^{-5}$ ns = not significant

To examine to what extent the *linear combination* of the five emotional dimensions explains the variance of the listeners' judgments for the basic emotions, we performed a series of multiple linear regression models with basic emotion as the response variable and the listeners' judgments for the emotional dimensions as the predictor variables. However, because the emotional dimensions are highly intercorrelated (as shown in table 3), we applied principal component analysis on the emotional dimensions to eliminate covariate variables. This analysis showed that the first (PC1) and the second (PC2) principal components taken together account for **95%** of the total variance of the listeners' judgments for the five emotional dimensions (PC1 = 73%; PC2 = 22%). Therefore, the scores of the utterances in these principal components were taken as the predictor variables of the multiple linear regression models (representing the listeners' judgments for the five emotional dimensions). To correct for violations of normality and/or of constant variance (statistical assumptions for multiple linear regression analysis), we applied a log transformation to the response variable when necessary (for joy, calmness, and anger).

The model with *joy* as the response variable yielded an adjusted R^2 of **67%** ($F[2, 36] = 39.96, p < 10^{-09}$), which indicates that the linear combination of PC1 and PC2 explains a significant and relative high proportion of variance of the listeners' judgments for joy. There was no significant interaction between PC1 and PC2 in this model. The model with *calmness* as the response variable and with a significant interaction between PC1 and PC2 yielded an adjusted R^2 of **89%** ($F[3, 36] = 109.5, p < 10^{-15}$). The model with *anger* as the response variable

yielded an adjusted R^2 of **69%** ($F[2, 37] = 43.71, p < 10^{-09}$), and there was no significant interaction between PC1 and PC2. Finally, the model with *sadness* as the response variable and with a significant interaction between PC1 and PC2 yielded an adjusted R^2 of **68%** ($F[3, 36] = 28.16, p < 10^{-08}$).

3.3 Acoustic analysis

To investigate which acoustic parameters of speech correlate better with the listeners' judgments for the basic emotions and emotional dimensions and thus contribute to advance the knowledge of how these parameters vary as a function of the speaker's emotions, a set of acoustic features was automatically extracted from the utterances evaluated by the judges in the experiments by means of the script "Expression Evaluator", implemented for the software Praat (BOERSMA; WEENINK, 2011) by Barbosa (2009).⁴

This script analyzes the following acoustic features: fundamental frequency (f0), fundamental frequency first derivative (df0), global intensity, spectral tilt and Long-Term Average Spectrum (LTAS). All of these parameters have been reported in the literature as potential correlates of the vocal expression of emotions, since they may be affected by the psychophysiological responses triggered by the emotional processes (BARBOSA, 2009; FRICK, 1985; PITTAM; SCHERER, 1993; SCHERER, 1986). Fundamental frequency is an acoustic correlate of the rate of vocal fold vibration and is perceived as the pitch of the voice. Sound intensity, measured in decibels (dB), corresponds to the variations in the air pressure of a sound wave and it is the major contributor to the sensation of loudness of a sound. Spectral tilt, considered here as the difference of intensity between the 0 – 1250 Hz and 1250 – 4000 Hz frequency bands computed every ten points, measures the degree of the drop in intensity as the frequencies of the spectrum increase. LTAS is an intensity spectrum obtained from the average of several spectra extracted from the speech sample for a given frequency range. Various authors argue that the LTAS reduces the effect of individual linguistic segments on the spectral structure of speech, thus providing a spectral representation of the speaker's voice as a whole (PITTAM; GALLOIS;

⁴ This Praat script is available from the authors upon request.

CALLAN, 1990; SCHERER, 1982). The f_0 first derivative, computed as the difference in Hz between successive odd-numbered f_0 values, is used as a means of revealing abrupt changes in the intonation contour (BARBOSA, 2009).

In addition, spectral tilt and LTAS are acoustic correlates of vocal effort and voice quality, since the increase of vocal effort enhances the energy in the harmonics of high frequencies due to changes in subglottal pressure and in the characteristics of vocal fold vibration (LAUKKANEN *et al.*, 1997; TRAUNMÜLLER; ERIKSSON, 2000).

These acoustic parameters are computed by the script in terms of the following statistical descriptors (yielding a total of twelve parameters):

f_0 : median, interquartile semi-amplitude, skewness, and 99.5% quantile;

df_0 : mean, standard deviation, and skewness;

global intensity: skewness;

spectral tilt: mean, standard deviation, and skewness;

LTAS: slope (difference of mean intensity in dB between the bands 0 – 1000 Hz and 1000 – 4000 Hz of the LTAS).

The statistical descriptors related to f_0 and df_0 were normalized for inter-speaker variability through the z-score technique⁵ by using the following reference values (mean, standard deviation) of f_0 for adult females: (231 Hz, 120 Hz). Spectral tilt was normalized by dividing its value by the complete-band intensity median. The f_0 interquartile semi-amplitude is calculated as the difference between the 95% and 5% quantiles of f_0 , divided by two. It is, therefore, a measure analogous to f_0 range (f_0 maximum - f_0 minimum), but less sensitive to measurement errors, since it does not take into account the extreme values of the data. Similarly, the f_0 99.5% quantile is a measure analogous to f_0 maximum. Skewness indicates whether the distribution of the random variable (*i.e.* the measured values) is symmetric or asymmetric about its mean (and, in the latter case, whether the larger concentration of the values is on the left or on the right of the mean). The f_0 skewness is taken as the difference between f_0 mean and f_0 median, divided by the f_0 interquartile semi-amplitude.

⁵ The z-score is a statistical procedure that expresses the distance in terms of standard deviation units of an observation from its mean (here, the reference values).

The acoustic data were then correlated with the listeners' mean ratings for each utterance for the basic emotions (Table 4) and emotional dimensions (Table 5) by computing the Spearman's rank correlation. The parameters which correlated better with the listeners' judgments were f0 median, interquantile semi-amplitude, and 99.5% quantile, spectral tilt mean and standard deviation, and LTAS slope. In general, these parameters correlated better with the emotional dimensions. The parameters f0 median, f0 interquantile semi-amplitude, spectral tilt mean, and LTAS slope, for example, were significantly correlated with all five emotional dimensions, but only with some of the basic emotions. In addition, f0 skewness was significantly correlated only with fairness, motivation, and involvement.

Joy was not significantly correlated with any of the acoustic features. This may have occurred because of a possible lack of exemplars of this emotion in our corpus, or perhaps because the speakers of our corpus may not have expressed this emotion consistently through speech. The sign of the correlation coefficients (positive or negative) reveals that the perceived increase of *calmness* in the speakers' speech was associated with a decrease of f0 median, f0 interquantile semi-amplitude, f0 99.5% quantile, and of the energy concentrated in the harmonics of higher frequencies (indicated by spectral tilt mean and standard deviation, and LTAS slope).⁶ Higher ratings of *anger* were linked to an increase in f0 median, f0 interquantile semi-amplitude, f0 99.5% quantile, and in high-frequency energy. *Sadness* was associated with a decrease of f0 interquantile semi-amplitude and in high-frequency energy (indicated by spectral tilt standard deviation and LTAS slope) across the utterances.

⁶ Because spectral tilt and LTAS slope are estimated by the difference of intensity between the lower and higher frequency bands and the intensity drops as the frequencies of the spectrum increase, an increase in these parameters means *less* energy concentrated in the harmonics of higher frequencies due to a lower vocal effort used in the production of the utterance (TRAUNMÜLLER; ERIKSSON, 2000).

TABLE 4
Spearman's rank correlations between acoustic parameters and listeners' judgments (mean ratings for each utterance) for the basic emotions

Acoustic Parameters	Joy	Calmness	Anger	Sadness
Fundamental frequency (f0)				
median	0.13 ns	-0.65***	0.57***	-0.17 ns
interquantile semi-amplitude	0.13 ns	-0.56***	0.53***	-0.37*
99.5% quantile	0.10 ns	-0.33*	0.31*	-0.15 ns
skewness	-0.06 ns	0.26 ns	-0.26 ns	-0.01 ns
F0 first derivative				
mean	-0.18 ns	0.07 ns	0.03 ns	0.28 ns
standard deviation	0.17 ns	-0.22 ns	0.23 ns	-0.15 ns
skewness	-0.06 ns	0.00 ns	-0.02 ns	0.08 ns
Global intensity				
skewness	-0.18 ns	0.00 ns	0.09 ns	0.17 ns
Spectral tilt				
mean	-0.03 ns	0.54***	-0.54***	0.28 ns
standard deviation	-0.21 ns	0.41**	-0.35*	0.55***
skewness	0.10 ns	0.10 ns	-0.13 ns	-0.01 ns
Long-Term Average Spectrum - LTAS				
slope	0.08 ns	0.81***	-0.79***	0.38*

*p < 0.05 **p < 0.01 ***p < 0.001 ns = not significant

With respect to the emotional dimensions, the perceived increase in the degree of activation, motivation, and involvement was significantly linked to an increase in f0 median, f0 interquantile semi-amplitude and in high-frequency energy (reflected mainly in spectral tilt mean and LTAS slope). Activation was also significantly associated with an increase in f0 99.5% quantile, whereas motivation and involvement with a decrease in f0 skewness. Conversely, the perceived increase in fairness and valence was associated with a decrease in f0 median, f0 interquantile semi-amplitude and in high-frequency energy (also indicated by spectral tilt mean and LTAS slope). In addition, fairness was significantly linked to a decrease in f0 99.5% quantile and an increase in f0 skewness. It can

also be observed that the opposite behavior between the two groups of dimensions (which separate fairness and valence from involvement, motivation, and activation) observed in Table 2 and in Table 3 still holds for the correlations between these emotional dimensions and the acoustic parameters. This means that different patterns of changes in acoustic parameters characterize these groups.

TABLE 5

Spearman's rank correlations between acoustic parameters and listeners' judgments (mean ratings for each utterance) for the emotional dimensions

Acoustic Parameters	Activation	Fairness	Valence	Motivation	Involvement
Fundamental frequency (f0)					
median	0.75***	-0.57***	-0.38*	0.63***	0.72***
interquartile semi-amplitude	0.59***	-0.51***	-0.33*	0.48**	0.48**
99.5% quantile	0.40*	-0.34*	-0.19 ns	0.22 ns	0.29 ns
skewness	-0.29 ns	0.31*	0.25 ns	-0.31*	-0.44**
F0 first derivative					
mean	-0.14 ns	0.00 ns	-0.18 ns	-0.25 ns	-0.00 ns
standard deviation	0.28 ns	-0.20 ns	-0.09 ns	0.16 ns	0.20 ns
skewness	-0.06 ns	0.06 ns	0.03 ns	-0.13 ns	-0.09 ns
Global intensity					
skewness	-0.06 ns	-0.16 ns	-0.29 ns	-0.04 ns	0.09 ns
Spectral tilt					
mean	-0.59***	0.45**	0.38*	-0.54***	-0.63***
standard deviation	-0.47**	0.27 ns	0.13 ns	-0.47**	-0.28 ns
skewness	-0.00 ns	0.05 ns	0.13 ns	-0.02 ns	0.02 ns
Long-Term Average Spectrum - LTAS					
slope	-0.79***	0.72***	0.64***	-0.65***	-0.64***

*p < 0.05 **p < 0.01 ***p < 0.001 ns = not significant

4 Discussion

The present study was conducted to shed light on the relation between the perception of basic emotions and emotional dimensions in speech, as well as to identify some acoustic cues which may guide this process. For this purpose, a group of Brazilian subjects rated the expression of four basic emotions in utterances of a foreign language of which they had no knowledge (Swedish) and a separate group of Brazilian subjects rated the expression of five emotional dimensions for the same utterances. To further advance the knowledge of how emotions are expressed and perceived in everyday interactions, the corpus used in this study was composed of authentic emotional expressions as conveyed in spontaneous speech.

The novelty of this study lies in the direct comparison between the perception of some basic emotions and some emotional dimensions in speech, which provided evidence on how the two perspectives can be related to each other. The use of multiple linear regression analysis to assess this relation is also new, and proved effective. The majority of studies which suggested a relation between basic emotions and emotional dimensions have not done so in the form of a specific relationship, but only presented classes of basic emotions in a dimensional space. In summary, this work has empirically shown that the perception of basic emotions in speech can be described as a combination of emotional dimensions, which, in the study presented in this paper, tended to display a specific pattern for each basic emotion. In addition to contributing to the literature on emotion in general, this finding is also relevant to research on emotion and speech, as it provides researchers with empirical evidence which may help them to choose the best approach for their studies and to better interpret the results obtained with these approaches. Furthermore, three of the emotional dimensions investigated here (fairness, motivation, and involvement), despite being recognized by some theorists, have not been satisfactorily examined in studies on emotional prosody. Thus, this study also contributed by providing knowledge of other emotional dimensions, which can be used in expressive speech applications. The dimensions of motivation and involvement, for example, can be used in automatic processing of meetings, in order to detect heated arguments or periods of high excitement (WREDE; SHRIBERG, 2003) or in call center conversations (together with activation) to monitor the affective state of customers (VOGT; ANDRÉ; WAGNER, 2008).

Our results show that, apart from the classic dimensions of activation and valence, the dimensions of fairness, motivation, and involvement can also be inferred from speech. This is not only ensured by a significant agreement between the listeners' judgments for their expressivity (which was slight to fair but similar to other studies on the subject), but mainly because these dimensions were strongly and significantly correlated with some acoustic parameters, basic emotions and with activation and valence. This finding suggests that other emotional dimensions related to the appraisal of the eliciting event and to action readiness can also be inferred from speech. Investigating other emotional dimensions which could possibly be recognized from speech may advance our understanding of how emotions (and other affective phenomena) are expressed and perceived in speech, as well as of the expressive functions of speech prosody.

All the five emotional dimensions investigated here were significantly correlated to some extent with the basic emotions. In addition, the four basic emotions analyzed in the present study tended to have different patterns of perceptual judgments for these dimensions:⁷ *joy* (positive fairness, positive valence, and high motivation); *anger* (high activation, negative fairness, negative valence, high motivation, and high involvement); *sadness* (low activation, low motivation, and low involvement); *calmness* (low activation, positive fairness, positive valence, low motivation, and low involvement). As can be observed, with regard to activation and valence, these patterns were, in general, consistent with previous findings. In this way, this study contributed to better understand the underlying structure of these discrete emotions with regard to some emotional dimensions. Furthermore, the multiple linear regression analysis performed with each basic emotion and the two principal components corresponding to the listeners' judgments for the five emotional dimensions showed that indeed it is possible to describe the perception of basic emotions in speech as a combination of emotional dimensions, since the linear combination of the two principal components explained a significant and relative high (more than 50%) proportion of variance of the listeners' judgments for all the basic emotions (as revealed by the adjusted R^2 of the models).

⁷ Because its correlations with fairness and valence were not statistically significant, sadness could not be reliably distinguished from calmness in this respect.

The basic emotions and the emotional dimensions were also significantly correlated with some acoustic parameters extracted automatically from the utterances used as stimuli in the perception experiments, which indicates that the listeners relied partly on these acoustic features to judge the expressiveness of these emotions and dimensions. The most robust parameters were f_0 median, interquartile semi-amplitude, and 99.5% quantile, spectral tilt mean and standard deviation, and LTAS slope. These parameters confirm the relevance of fundamental frequency and voice quality in the communication of emotions through speech. Among the basic emotions, only anger and calmness correlated significantly with a considerable number of acoustic parameters. Joy was not significantly correlated with any of the acoustic features and sadness correlated moderately with three of them. All five emotional dimensions, on the other hand, correlated significantly with various acoustic parameters. Therefore, it is possible to conclude that, in general, the acoustic parameters correlated better with the emotional dimensions. This result is consistent with some previous studies, which have suggested that emotional dimensions are more suitable to distinguish and describe the vocal expression of emotions than labels of discrete emotions (COWIE; CORNELIUS, 2003; LUGGER; YANG, 2007; PEREIRA, 2000; BARBOSA, 2009). Research on the automatic recognition of emotions from speech and speech synthesis can benefit from this finding, since the use of emotional dimensions may allow the reliable identification and synthesis of more subtle expressions of emotions and changes in speech expressiveness over time (GRIMM, *et al.*, 2007; SCHRÖDER *et al.*, 2001; BARBOSA, 2009).

Acknowledgments

This work was funded by a fellowship from the São Paulo Research Foundation – FAPESP (2012/04254-4 and 2013/06082-9) to the first author. We thank all the subjects who took part in the perception experiments reported here.

References

ALM, C. O.; SPROAT, R. Perceptions of emotions in expressive storytelling. In: INTERSPEECH 2005, 2005, Lisboa. *Proceedings of Interspeech 2005*. Lisboa: ISCA Archive, 2005. p. 533-536. Disponível em: <http://www.isca-speech.org/archive/interspeech_2005>. Acesso em: 14 jul. 2016.

AMIR, N.; MIXDORFF, H.; OFER AMIR, D. R.; DIAMOND, G. M.; PFITZINGER, H. R.; LEVI-ISSERLISH, T.; ABRAMSON, S. Unresolved Anger: Prosodic analysis and classification of speech from a therapeutical setting. In: INTERNATIONAL CONFERENCE ON SPEECH PROSODY 2010, 5., 2010, Chicago. *Proceedings...* Chicago, 2010. Disponível em: <<http://speechprosody2010.illinois.edu/program.php>>. Acesso em: 14 jul. 2016.

BANSE, R.; SCHERER, K. R. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, v. 70, n. 3, p. 614-636, 1996. doi.org/10.1037/0022-3514.70.3.614

BARBOSA, P. A. Detecting changes in speech expressiveness in participants of a radio program. In: INTERSPEECH 2009, 2009, Brighton. *Proceedings of Interspeech 2009 - Speech and Intelligence*. Londres: Causal Productions, 2009. p. 2155-2158. Disponível em: <http://www.isca-speech.org/archive/interspeech_2009/i09_2155.html>. Acesso em: 18 nov. 2016.

BARBOSA, P. A. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 20, n. 1, p. 11-27, 2012. Disponível em: <<http://periodicos.letras.ufmg.br/index.php/relin/article/view/2571>>. Acesso em: 18 nov. 2016.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer (Versão 5.2.25)*. Disponível em: <<http://www.praat.org>>. Acesso em: 20 Mai. 2011.

BROSCH, T; MOORS, A. Valence. In: SANDER, D.; SCHERER, K.R. (Org.). *The Oxford Companion to Emotion and the Affective Sciences*. Oxford: Oxford University Press, 2009. p. 401-402.

CORNELIUS, R. R. Theoretical approaches to emotion. In: ISCA TUTORIAL AND RESEARCH WORKSHOP (ITRW) ON SPEECH AND EMOTION, 2000, Newcastle. *Proceedings...* Disponível em: <http://www.isca-speech.org/archive_open/speech_emotion/>. Acesso em: 14 jul. 2016.

COWIE, R.; CORNELIUS, R. R. Describing the emotional states that are expressed in speech. *Speech communication*, v. 40, n. 1, p. 5-32, 2003. doi.org/10.1016/S0167-6393(02)00071-7

DARWIN, C. *A expressão das emoções no homem e nos animais*. São Paulo: Companhia das Letras, 2009 [1872].

DEVILLERS, L.; COWIE, R.; MARTIN, J.-C.; DOUGLAS-COWIE, E.; ABRILIAN, S.; MCRORIE, M. Real life emotions in French and English TV video clips: An integrated annotation protocol combining continuous and discrete approaches. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, Fifth., 2006, Genoa. *Proceedings...* Disponível em: <<http://www.lrec-conf.org/proceedings/lrec2006/>>. Acesso em: 14 jul. 2016.

DOWDY, S.; WEARDEN, S.; CHILKO, D. *Statistics for research*. 3. ed. Hoboken: John Wiley & Sons, 2004. doi.org/10.1002/0471477435

EKMAN, P. An argument for basic emotions. *Cognition & Emotion*, v. 6, n. 3/4, p. 169-200, 1992. doi.org/10.1080/02699939208411068

ELLSWORTH, P. C.; SCHERER, K. R. Appraisal processes in emotion. In: DAVIDSON, R. J.; SCHERER, K. R.; GOLDSMITH, H. H. (Org.). *Handbook of affective sciences*. New York: Oxford University Press, 2003. p. 572-595.

FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v. 76, n. 5, p. 378-382, 1971. doi.org/10.1037/h0031619

FONTAINE, J. Dimensional Emotion Models. In: SANDER, D.; SCHERER, K.R. (Org.). *The Oxford Companion to Emotion and the Affective Sciences*. Oxford: Oxford University Press, 2009. p. 119-120.

FOWLES, D. C. Arousal. In: SANDER, D.; SCHERER, K. R. (Org.). *The Oxford Companion to Emotion and the Affective Sciences*. Oxford: Oxford University Press, 2009. p. 50-50.

FRICK, R. W. Communicating emotion: The role of prosodic features. *Psychological Bulletin*, v. 97, n. 3, p. 412-429, 1985. doi.org/10.1037/0033-2909.97.3.412

FRIJDA, N. H.; MARKAM, S.; SATO, K.; WIERS, R. Emotions and emotion words. In: RUSSELL, J. A.; FERNANDEZ-DOLS, J. M.; MANSTEAD, A. S. R.; WELLENKAMP, J. C. (Org.) *Everyday Conceptions of Emotion: An Introduction to the Psychology, Anthropology and Linguistics of Emotion*. NATO ASI series D: Behavioural and social sciences. Nova Iorque: Kluwer Academic/Plenum Publishers, 1995. v. 81, p. 121-143. doi.org/10.1007/978-94-015-8484-5_7

GRIMM, M.; KROSCHEL, K.; MOWER, E.; NARAYANAN, S. Primitives-based evaluation and estimation of emotions in speech. *Speech Communication*, v. 49, n. 10, p. 787-800, 2007. doi.org/10.1016/j.specom.2007.01.010

JOHNSTONE, T.; SCHERER, K. R. Vocal communication of emotion. In: LEWIS, M.; HAVILAND, J. M. (Org.) *Handbook of emotions*. 2. ed. Nova Iorque: Guilford, 2000. p. 220-235.

KEHREIN, R. The prosody of authentic emotions. In: SPEECH PROSODY 2002, 2002, Aix-en-Provence. *Proceedings of Speech Prosody 2002*. Disponível em: <<http://sprogis.isle.illinois.edu/sp2002/papers.htm>>. Acesso em: 14 jul. 2016.

LAUKKA, P.; ELFENBEIN, H. A. Emotion appraisal dimensions can be inferred from vocal expressions. *Social Psychological and Personality Science*, v. 3, n. 5, p. 529-536, 2012. doi.org/10.1177/1948550611428011

LAUKKA, P.; JUSLIN, P.; BRESIN, R. A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, v. 19, n. 5, p. 633-653, 2005. doi.org/10.1080/02699930441000445

LAUKKANEN, A-M; VILKMAN, E.; ALKU, P.; OKSANEN, H. On the perception of emotions in speech: the role of voice quality. *Logopedics Phoniatrics Vocology*, v. 22, n. 4, p. 157-168, 1997. doi.org/10.3109/14015439709075330

LUGGER, M., YANG, B. An incremental analysis of different feature groups in speaker independent emotion recognition. In: INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, 16., 2007, Saarbrücken. *Proceedings....* Disponível em: <<http://www.icphs2007.de/>>. Acesso em: 14 jul. 2016.

- MATSUMOTO, D.; EKMAN, P. Basic emotions. In: SANDER, D.; SCHERER, K.R. (Org.). *The Oxford Companion to Emotion and the Affective Sciences*. Oxford: Oxford University Press, 2009. p. 69-72.
- MAUSS, I. B.; ROBINSON, M. D. Measures of emotion: A review. *Cognition and Emotion*, v. 23, n. 2, p. 209-237, 2009. doi.org/10.1080/02699930802204677
- MESQUITA, B.; FRIJDA, N. H. Cultural Variations in Emotions: A Review. *Psychological Bulletin*, v. 112, n. 2, p. 179-204, 1992. doi.org/10.1037/0033-2909.112.2.179
- PEREIRA, C. Dimensions of emotional meaning in speech. In: ISCA TUTORIAL AND RESEARCH WORKSHOP (ITRW) ON SPEECH AND EMOTION, 2000, Newcastle. *Proceedings of the ISCA Workshop on Speech and Emotion*. Disponível em: <http://www.isca-speech.org/archive_open/speech_emotion/>. Acesso em: 14 jul. 2016.
- PERES, D. O. A manifestação da emoção na fala: estudo perceptual com falantes nativos e não nativos. *Estudos Linguísticos*, São Paulo, v. 43, n. 1, p. 10-21, 2014.
- PITTAM, J.; GALLOIS, C.; CALLAN, V. The long-term spectrum and perceived emotion. *Speech Communication*, Amsterdam, v. 9, n. 3, p. 177-187, 1990. doi.org/10.1016/0167-6393(90)90055-E
- PITTAM, J.; SCHERER, K. R. Vocal expression and communication of emotion. In: LEWIS, M.; HAVILAND, J. M. (Org.). *Handbook of emotions*. New York: Guilford Press, 1993. p. 185-198.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em: 5 fev. 2014.
- RUSSELL, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology*, v. 39, p. 1161-1178, 1980. doi.org/10.1037/h0077714
- SCHERER, K. R. Methods of research on vocal communication: Paradigms and parameters. In: SCHERER, K. R.; EKMAN, P. (Org.). *Handbook of methods in nonverbal behavior research*. Cambridge: Cambridge University Press, 1982. p. 136-198.

SCHERER, K. R. Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, v. 99, n. 2, p. 143-165, 1986. doi.org/10.1037/0033-2909.99.2.143

SCHERER, K. R. Emotion. In: HEWSTONE, M.; STROEBE, W. (Org.). *Introduction to Social Psychology: A European perspective*. 3. ed. Oxford: Blackwell, 2000. p. 151-191.

SCHERER, K. R. Vocal communication of emotion: a review of research paradigms. *Speech Communication*, v. 40, n. 2, p. 227-256, 2003. doi.org/10.1016/S0167-6393(02)00084-5

SCHERER, K. R.; BANSE, R.; WALLBOTT, H. G. Emotion Inferences from Vocal Expression Correlate across Languages and Cultures. *Journal of Cross-Cultural Psychology*, Sage Publications, v. 32, n. 1, p. 76-92, 2001. doi.org/10.1177/0022022101032001009

SCHERER, K. R.; ELLGRING, H. Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion*, v. 7, n. 1, p. 158-171, 2007. doi.org/10.1037/1528-3542.7.1.158

SCHLOSBERG, H. A scale for the judgement of facial expressions. *J. Exp. Psychology*, Association for the Advancement of Affective Computing., v. 29, p. 497-510, 1941.

SCHLOSBERG, H. Three dimensions of emotion. *Psychological Review*, American Psychological Association, v. 61, n. 2, p. 81-88, 1954. doi.org/10.1037/h0054570

SCHRÖDER, M.; COWIE, R.; DOUGLAS-COWIE, E.; WESTERDIJK, M.; GIELEN, S. C. Acoustic correlates of emotion dimensions in view of speech synthesis. In: INTERSPEECH 2001, 2001, Aalborg. *Proceedings of Interspeech 2001*. Disponível em: <http://www.isca-speech.org/archive/eurospeech_2001/index.html>. Acesso em: 14 jul. 2016.

SILVA, W.; BARBOSA, P. A.; ABELIN, Å. Cross-cultural and cross-linguistic perception of authentic emotions through speech: An acoustic-phonetic study with Brazilian and Swedish listeners. *DELTA*, São Paulo, v. 32, n. 2, p. 449-480, 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502016000200449&lng=en&nrm=iso>. Acesso em: 18 nov. 2016.

SMITH, C. A.; ELLSWORTH, P. C. Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, v. 48, n. 4, p. 813-838, 1985. doi.org/10.1037/0022-3514.48.4.813

TRAUNMÜLLER, H; ERIKSSON, A. Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*, v. 107, n. 6, p. 3438-3451, 2000. doi.org/10.1121/1.429414

VOGT, T.; ANDRÉ, E.; WAGNER, J. Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation. In: PETER, C.; BEALE, R. (Org.). *Affect and Emotion in Human-Computer Interaction: From Theory to Applications (Lecture Notes in Computer Science)*. Heidelberg, Germany: Springer, 2008. v. 4868, p. 75-91. doi.org/10.1007/978-3-540-85099-1_7

WREDE, B.; SHRIBERG, E. Spotting “hot spots” in meetings: human judgments and prosodic cues. In: INTERSPEECH 2003, 2003, Genebra. *Proceedings of Interspeech 2003*. Genebra: ISCA Archive, 2003. p. 2805-2808. Disponível em: <http://www.isca-speech.org/archive/eurospeech_2003/e03_2805.html>. Acesso em: 19 out. 2016.