

Perception of emotional speech in Brazilian Portuguese : an intonational and multidimensional approach

Daniel Oliveira Peres

Departamento de Letras Clássicas e Vernáculas

University of Sao Paulo/Capes

<danielperes@usp.br>

Résumé

L'objectif de cet article est d'analyser la perception que des sujets parlants brésiliens et anglais (sans connaissances du portugais) ont de la parole émotionnelle en portugais brésilien (PB), en utilisant un modèle tridimensionnel de l'émotion. Les résultats des jugements faits par les participants ont été comparés aux paramètres acoustiques de l'intonation, extraits par le logiciel ExProsodia. (Ferreira Netto, 2008, 2010). Des régressions linéaires simples et multiples ont été réalisées, ayant pour objectif de savoir comment le jugement des participants et les paramètres acoustiques de l'intonation peuvent être reliés. Les résultats ont montré que les paramètres intonationnels peuvent prédire les degrés de perception d'activation et de dominance de la parole émotionnelle en PB.

Keywords : *phonetic, intonation, emotion*

1. Introduction

In the West, the expression of emotions in humans has been an object of study with wide range of interests for at least 2000 years now. For the purpose of this work, emotion will be analysed in a nineteenth-century approach, namely evolutionary. Furthermore, a social approach will be considered to deal with the different perceptions of emotional speech by native and non-native speakers. Since the nineteenth century, studies on emotions have begun to have a greater wealth of physiological and psychological details. The pioneering work of Darwin (1965 [1872]) *The Expression of the Emotions in Man and Animals* is an example of this type of descriptive and theoretical refinement. Discussions about the expression of emotions in other theoretical perspectives have progressed since the publication of Darwin's work, in subjects such as psychology, neurology and other areas related to the production and speech perception. In the late nineteenth century, cognitive and sociocultural theories began to be concerned with the expression of emotions and their relationship to cognitive, linguistic and social contexts. Among the many questions

that involve the study of emotional speech, the role played by language and culture in the perception and categorization of emotions is highly relevant. Likewise, the search for universal features in the expression and perception of emotion is crucially important for studies in affective science.

This study will be guided by the following two approaches in order to investigate the role of linguistic and cultural knowledge¹ on the perception of emotions. Culture is not central to Darwin's argumentation, the author argues that emotions and their recognition are part of a biological inheritance and are therefore universally manifest. Nowadays, alongside the Darwinian approach, researchers adopt the social approach of emotions. Although Darwin did not ignore the different societies, ethnic groups and languages, his focus is on what can be shared by humans and animals, with no regard to regional details. The interest of researchers in the role of society, language and culture in the categorization of emotions began with anthropological studies (Dodds 1951, Turnbull 1961, Carstairs 1967, Wallace, Carson 1973, Harre and finlay Jones 1986, as cited in Russell 1991). The motivations of such studies were the discrepancy between words that designate emotions in different cultures and the influence of the social environment in the expression of emotions.

Russell (1991) states that the division of the world into categories is a human characteristic that is extended to the categorization of emotions by natural languages. The central concern in this approach is not what emotion is represented in a given language, but how speakers of the same language perceive it. An alternative approach to biological/evolutionary and social theory would be a theory that seeks to address both of them, trying to establish the importance of universal and sociocultural aspects for the study of emotions. To this end, the experimental approach of speech involving acoustic characteristics and their perception is necessary so that we can determine the extent to which « what is said » and « how it is said » can contribute to the perception of emotional speech.

1.1. Emotion : Status Quaestionis

The study of emotion can be divided into three major areas of interest, namely evolutionary, social and cognitive (internal processes). The approaches that I will take in this research are the evolutionary and

¹ As in Tylor : [...] « a complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society » (Tylor, 1871, 1).

social. Darwin attempted to explain how emotions were inherited by successive generations, but he did not take language and culture into account because he considered that these would be later stages from the evolutionary point of view. Some mentions of voice changes in emotional states can be found as in the following :

« He (Spencer) clearly shows that the voice alters much under different conditions, in loudness and in quality, that is, in resonance and timbre, in pitch and intervals. [...] from the analogy of other animals, I have been led to infer that the progenitors of man probably uttered musical tones, before they had acquired the power of articulate speech ; and that consequently, when the voice is used under any strong emotion, it tends to assume, through the principle of association » (Darwin 1965 [1872], 87).

Darwin, relying on the study of Spencer (1858), considered that emotions are also expressed concomitantly with speech and are not limited to their primary expression found in other animals. On the other hand, there are theories that explain emotion as social constructions, i.e., a product of distinct cultures and learned by members of each society through experience. Among other reasons, the discrepancy between the names given to emotions in different languages seems to be due to local and social feature of emotional categorisation. In different languages, the number of words used to name emotions, feelings and attitudes can vary greatly.²

Although it can be a relevant argument in order to demonstrate the variation in the categorisation of emotions, this kind of research has an inherent problem. The methodology can vary widely between surveys and interviews of the same research. Furthermore, the number of words related to emotions and their meanings can be biased by the conflict of meaning between the language used by the interviewer and the interviewee.

Another argument used by social researchers of emotion is that the emotional manifestations occur in a social or at least interpersonal environment. According to Parkinson (1996, 680), the meanings of emotions arise from social interaction. Moreover, the expression of emotional states is mediated by social norms and values, and these indicate which emotion and intensity would be suitable for different situations. Averill (1993) describes some rules for the expression of anger, among which is the proportionality of the response to a stimulus, i.e. the reaction cannot exceed what is necessary to correct the situation.

² For more details see Russell (1991).

As an example, although the concept of emotion is regarded as universal in the study by Brandt & Boucher (1986), there are no universally valid correlative terms (Lutz 1980, Howell 1981, as cited in Russell 1991). Russell (1991) shares this view and states that :

« [...] people of different cultures and speaking different languages categorize the emotions somewhat differently. The boundaries around the domain appear to vary, as do divisions within the domain. Thus, neither the word emotion nor words for even alleged basic emotions, such as anger and sadness are universal » (Russell 1991, 444).

In order to address emotional speech considering these two types of analysis, social and evolutionary, Scherer (2006) proposed an approach based on *push* and *pull effects*. This approach includes what is general/universal in emotional speech as well as local linguistic features. The *push effect* is understood as the influence of psychophysiological activation, strongly linked to emotional speech ; whereas the *pull effect* is related to cultural rules of expression and affects pragmatic prosody. This distinction implies that the *push effect* is externalized directly on the motor response, consequently, in speech production, while the *pull effect* is based on sociocultural models that are learned by the members of society

For the analysis, Scherer (2006) proposed the principles of *covariation* and *configuration*. The covariation principle assumes a continuous relationship between emotional manifestation and acoustic parameters. If this principle is true, perceptual experiments that attempt to describe some relationship between emotion and acoustic correlates (or visual, for example) must demonstrate that changes in certain parameters can predict what the judgment of the participants will be. For example, modulations of fundamental frequency (F0), voice quality and energy of vocalizations can express fear, anger, joy etc. The configuration principle is related to the specific meaning conveyed by the utterance, prosodic configuration and linguistic choices. This principle determines the pragmatic characteristics of speech such as emphasis and sentence types (Scherer 2006, 13-14).

An exception can be made with regard to the infant's cry. This kind of voice expression is associated with the push effect, but it presents patterns of intensity and F0 declination similar to an intonational contour. The study of Mampe et al. (2009) demonstrated that the cries of newborn babies are largely determined by the intonation patterns of their native language, which are perceived even before birth. French babies tend to produce rising intonation patterns (L - H), while German infants tend to produce falling patterns (H - L).

For this reason, it is not clear whether the proposed division of push and pull effects can be accepted in a categorical way. Maybe the interaction of these principles could explain the results presented by Mampe et al. (2009).

1.2. Dimensional approach

The pioneering study of Wundt (1897) was one of the first proposals for a multidimensional system, suggesting that the nature of each emotional state is defined by its position between orthogonal dimensions: *pleasantness - unpleasantness, rest - activation, and relaxation - attention*.

For Wundt, the content of the psychological experience is a compound of smaller components (Wundt 1897, 30). Thus, it is necessary to analyse smaller elements so that one can verify what the components are and how they are related. There are two levels of analysis: the level of psychological elements and the level of psychological compounds. The basic level consists of *sensations* – the result of stimulation of the sense organs, also called objective content; and *feelings* – related to sensations, or impressions that arouse feelings.

The contents of psychological experience consist of several combinations of *feelings* and *sensations*. Other models such as the ones put forward by Schlosberg (1952, as cited in Schlosberg, 1954) – composed of two dimensions (*pleasantness - unpleasantness* and *attention - rejection*) – and Schlosberg (1954) – composed of three dimensions (*pleasantness - unpleasantness, attention - rejection* and *activation [sleep - tension]*) – addressed the facial expressions of emotions from the combination of dimensions and its levels.

In the present work, the dimensions used were: *valence - unpleasant - pleasant*; *activation - non-agitated - agitated*; *dominance - submissive - non-submissive*. The choice of the dimensions was based on the study of Trnka (2011), who pointed out that *activation* and *valence* are present in most studies of emotion, except *dominance*. This latter dimension could not be ruled out *a priori*; on the contrary, it should be included and verified. After analysing the results it will be possible to measure the role of each dimension in the perception of emotions and its relation with the acoustic parameters used in this work.

1.3. Why spontaneous speech?

Most studies about emotional speech used sentences with acted speech or other types of elicitation (Costanzo et al. 1969, Scherer 2000, Scherer et al. 2001). In Brazilian Portuguese, studies were conducted

with acted speech (Colamarco & Moraes, 2008, Vassoler & Martins 2013). An advantage of acted speech is the possibility to control the stimuli in the same sentences, uttered in the most varied emotions. This feature allows the researcher a better control of the variables. Scherer (1981) pointed out the problems found in recordings of spontaneous speech with no direct intervention of the researcher. He states that « [...] naturally recorded emotions are by definition singular cases, both in terms of speaker identity, situation context, and verbal content of utterance » (Scherer 1981, 204). Because of the characteristics of spontaneous speech, it would be difficult to separate which of the variables are important to distinguish emotions in speech.

On the other hand, Roberts (2011) demonstrated that acted speech can be strongly influenced by stereotypes, stating that this kind of stimulus « merely may reflect stereotypical behaviours that actors are trained to adopt » (Roberts 2011, 1694). The use of this kind of speech could cause differences not only in production, but also probably in the perception of the stimuli. Scherer (1981) has also warned about this problem, pointing out the wrong use of discrete emotions in acted speech experiments. In the latter case, the results could be influenced by the stereotyped stimuli and by the use of discrete emotions.

Scherer et al. (2001) used a methodology that attempted to account for the variability of stimuli in spontaneous speech using acted speech elicited through scenarios.³ This methodology allows the researcher a better control of the independent variables. Scherer and his colleagues used *multilanguage meaningless sentences*, formed by two syllables of each of the six European languages present in their study, namely German, English, French, Italian, Spanish and Danish. Although it was a very inventive solution, some problems can be identified. The first problem is related to the aim of this study. If the principal aim is to try to analyse the role of sociocultural characteristics, e.g. language, sentences that do not belong to any natural language may sound strange to the participants. Even if the speaker recognizes a particular syllabic structure of their mother tongue, the other syllables of the sentence may not share the same features. Moreover, there is no way to measure the role of semantic information such as the lexicon. As it can be noticed, the mismatch between lexical information and intonational variation can be a problem for listeners.

The study of Nygaard & Queen (2008) showed the role of mismatching stimuli in the perception. For example, when the

³ See Scherer et al (2001) for more details about the methodology of scenarios.

« happy prosody » matches a word such as « lucky », or any word associated to happiness, the answers given by the listeners were faster, while the answers for mismatching stimuli were slower. For the authors, « the results suggest that although there is an apparent dissociation between the perception and identification of emotional tone of voice and of linguistic content, they cannot be completely independent or autonomous processes » (Nygaard & Queen 2008, 1025). The studies reported above show that acted speech and meaningless sentences can be a reliable way to carry out experiments on emotion, but also that mismatching stimuli and stereotypical can be a problem. Therefore, the present study supports the use of spontaneous speech in order to express a real situation of emotional expressions and to reduce the influence of the researcher.

2.1. Automatic analysis

In this study, the automatic analysis was done by the ExProsodia software. Following the hypothesis proposed by Ferreira Netto (2006, 2008), intonation can be understood as a time series. His approach of intonation is based on the proposal of Xu & Wang (1997), which claims that intonation can be divided into mechanic-physiological and expressive components. The first component is related to physiological features of the speaker and the second is the intentional F0 variation produced by speaker. The mechanical and physiological constraints would be conditioned by the physiological characteristics of the speaker, and the expressive intentions would be the intentional variations of F0 with linguistic purposes.

The ExProsodia selects units of the speech that contain relevant prosodic information. The selection of intonation units (IU) is done by the software, which performs automatic analysis of the speech excerpts from F0 data extracted by *Speech Filing System* (Huckvale 2008). The F0 values are measured at intervals of 5 milliseconds based on information predetermined by the researcher. The information required is measurement of F0 (Hz) and intensity (RMS). The researcher may also set the duration (ms). The intonation units (IU) are the result of the combination of these parameters. Below are the parameters used for the analyses in this study :

- Lower threshold of fundamental frequency (F0) : 50 Hz
- Upper threshold of fundamental frequency (F0) : 350 Hz
- Lower threshold of duration : 20ms
- Upper threshold of duration : 300ms
- Threshold of intensity : 2000RMS

The automatic analysis results in acoustic parameters that can be used in the analysis of intonation. Previous studies such as those carried out by Vassoler & Martins (2013), Ferreira Netto et al. (2013, 2014) and Peres (2013) present analyses of emotional speech using some parameters obtained by automatic analysis. Next section shows the parameters used in the course of this study.

2.1.1. *Acoustic parameters*

The acoustic parameters obtained by automatic analysis - ExProsodia - refer to different measures of F0 and duration. For this study, there are nine of them :

Five related to F0 :

- medium tone of sentences (MT)
- standard deviation of medium tone (sdMT)
- skewness of medium tone (sMT)
- coefficient of variation of medium tone (cvMT)
- lower value (Hz) of intonation unit (lvIU)

Four related to duration :

- duration (ms) of intonation unit (IU)
- standard deviation of intonation unit (sdIU)
- interval duration of intonation unit (idIU)
- standard deviation of interval of intonation unit (sdiIU).

3. Methodology

3.1. *Experimental Design*

The stimuli used in this study were collected from the website www.youtube.com. Thirty-two excerpts of spontaneous emotional speech were chosen and converted into .mp3 audio files (320 kbps). In a pre-test four participants (two Brazilians and two non-lusophone English) classified the stimuli into four basic emotions (happiness, sadness, fear and anger). This procedure was necessary for a better graphic analysis of the dimensions and their intersections. It will be displayed in data analysis session (3.2.). The stimuli were presented randomly to speakers of Brazilian Portuguese (São Paulo) and English speakers (England). It is important to note that these English speakers have no knowledge of Portuguese. This precaution was taken so that linguistic information could not influence the outcome of the task. Thirty-six participants (18 English and 18 Brazilian) carried out the experiment. For each stimulus, the participants had to give a score for each of the dimensions : *valence* - unpleasant - pleasant ; *activation* - non-agitated - agitated ; *dominance* - submissive - non submissive. In

order to evaluate each dimension, the participants had to use a slider button in a digital form, which allows a wide range of variation in the judgment. In a first test, made with two Brazilian informants, there was great difficulty in locating the position of the button for each dimension. Hence a hybrid model was made up through the use of the slider button and numerical values above it. The range was from 0 to 100 for each dimension. The values inside the button appeared at the same time as the participants were sliding the button.

3.2. Data analysis

Simple and multiple linear regressions were done in order to measure how the judgement of participants and the acoustic parameters of intonation could be related. The simple linear regression between the judgment of participants and acoustic parameters of intonation presented significant results as shown in table 1.

Dimension	Intonational Parameters	Brazilian	English
Activation	MT	—	—
	cvMT	$R^2 = 0.82$	$R^2 = 0.73$
	IU	$R^2 = 0.84$	$R^2 = 0.81$
Dominance	MT	$R^2 = 0.61$	—
	cvMT	—	$R^2 = 0.57$
	IU	—	$R^2 = 0.62$

Table 1 : Results of simple linear regressions

The results of simple linear regression showed that the perception of degrees of activation could be predicted by some acoustic parameters of intonation. The higher cvMT values are, the greater the activation. On the other hand, the lower IU values are, the greater the activation. Regarding the perception of degrees of dominance, MT had significant results related to Brazilian participants whereas cvMT and IU had significant results related to English participants.

The multiple linear regression models were based on the combination of acoustic parameters related to F0 and duration (MT, cvMT, IU, sdIU, idIU and sdiIU) and the judgment made by the participants. The significant results of multiple linear regressions are in table 2.

Dimension	Intonational Parameters	Brazilian	English
Activation	MT + IU	R ² = 0.86	R ² = 0.85
	MT + sdIU	—	—
	MT + idIU	—	R ² = 0.69
	MT + sdiIU	—	—
	cvMT + IU	R ² = 0.92	R ² = 0.86
	cvMT + sdIU	R ² = 0.81	R ² = 0.77
	cvMT + idIU	R ² = 0.85	R ² = 0.87
Dominance	cvMT + sdiIU	R ² = 0.82	R ² = 0.78
	MT + IU	R ² = 0.66	R ² = 0.71
	MT + sdIU	R ² = 0.61	—
	MT + idIU	R ² = 0.73	R ² = 0.75
	MT + sdiIU	R ² = 0.66	—
	cvMT + IU	—	R ² = 0.68
	cvMT + sdIU	—	R ² = 0.61
	cvMT + idIU		R ² = 0.79

Table 2 : Results of multiple linear regressions

The multiple linear regressions showed similar results to simple linear regression, but with more capacity of explanation, i.e. higher R² values. The general results pointed out to a relation between high values of MT and cvMT (which are combined to low values of IU, sdIU, idIU and sdiIU) and high evaluation of activation and dominance. As an illustration, a three-dimensional scatter plots shows the relation between the acoustic parameters (cvMT and IU) and the judgment of activation made by Brazilian (figure 1, below).

As can be seen in figure 1, the relation between the judgement of activation and the acoustic parameters (cvMT and IU) divided the four basic emotions into two separated groups. The higher judgement levels of activation were related to low values of IU (shorter units) and also related to high values of cvMT (greater F0 variations). The cluster highlighted in «A» is made of emotions, namely anger and happiness. The opposite can be seen concerning lower judgement levels of activation and high values of IU (larger units) and low values of cvMT (smaller F0 variations). For this configuration, the emotions were fear and sadness (B).

For English participants, the results of judgement of activation were quite similar when compared with results of Brazilian speakers. For both participants (Brazilian and English) the dimension valence did not present any significant result when compared to the acoustic parameters.

Despite of the significant results of both participants, there is a noticeable difference between them when the judgements of the three dimensions are plotted. The clusters of the Brazilian judgments (figure

2, below) are well marked whereas the clusters of English judgements (figure 3, below) are somewhat fuzzy.

4. Discussion

The analysis showed that evaluation of non-native speakers could be somehow explained by acoustic information, without the influence of lexicon. Also, some acoustic parameters extracted automatically by ExProsodia software could predict the perception of degrees of activation and dominance in emotional speech in BP by native and non-native speakers.

Regarding these results, there was a directly proportional relationship between activation judgement and high values of MT and cvMT. On the other hand, an inversely proportional relationship can be noticed between activation judgement and duration measurements, such as IU and idIU. The same could not be confirmed for sdIU and sdiIU. This relationship is reasonable because smaller IU (understood here as syllable⁴) when uttered in a high tone and with high F0 dispersion could be perceived as an agitated speech manifestation by the listeners. The reverse is also true. The perception of dominance had similar results related to the acoustic parameters related to F0. There was a directly proportional relationship between the judgement and high values of MT and cvMT. The same could not be confirmed for duration parameters. The perception of valence had no significant result when compared to acoustic parameters of intonation.

It is necessary to find more acoustic parameters that could explain the differences between the judgement made by Brazilian and English speakers (figures 7 and 8). There seems to be a linguistic component (pull effect) related to the perception of emotion that in addition to the acoustic parameters (covariation principle) may explain the performance of native speakers. But in the case of non-native speakers, the lack of linguistic knowledge of Brazilian Portuguese could explain their performance.

References

- Averill, J.R. (2013). Illusions of anger. In Folsom, R.B. & Tedeschi, J.T. (Eds.), *Aggression and violence: Social interactionist perspectives* (pp. 171-192) Washington, DC: American Psychological Association.
- Brandt, M.E., & Boucher, J.D. (1986). Concepts of depression in emotion lexicons of eight cultures. *International Journal of Intercultural Relations* 10, 321-346.

⁴ See Ferreira Netto et al. (2012) for more details.

- Colamarco, M., & Moraes, J.A. (2008). Emotion expression in speech acts in Brazilian Portuguese : production and perception. *Proceedings of Speech Prosody 2008* (pp. 717-720). Campinas : Unicamp,.
- Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 91-115). New York : Guilford Press.
- Costanzo, F.S., Merkel, N.N., & Costanzo P.R. (1969). Voice Quality Profile and Perceived Emotion. *Journal of Counseling Psychology* 16/3, 267-270.
- Darwin, C. (1965, [1872]). *The expression of the emotions in man and animals*. Chicago, IL : University of Chicago Press.
- Ferreira Netto, W. (2006). *Variação de frequência e constituição da prosódia da língua portuguesa*. (Tese de Livre-Docência na Área de Filologia e Língua Portuguesa). USP, São Paulo.
- Ferreira Netto, W. (2008). Decomposição da entoação frasal em componentes estruturadoras e semântico-funcionais. *X Congresso nacional de Fonética e Fonologia/IV Congresso Internacional de Fonética e Fonologia* (pp. 26-27). Universidade Federal Fluminense, Niterói, RJ.
- Ferreira Netto, W. (2010). ExProsodia. *Revista da Propriedade Industrial – RPI* 2038, p. 167. Rio de Janeiro.
- Ferreira Netto, W., Martins, M.V.M., Peres, D.O., Rosa, R.C.M., & Medeiros, J. (2012). Brazilian Portuguese intonation : A comparison between automatic and perceptual analyses. *The Journal of the Acoustical Society of America* 132, p. 2004. Kansas City, MO.
- Ferreira Netto, W., Peres, D.O., Martins, M.V.M., & Vieira, M.F. (2013). Análise automática da entoação emotiva (colérica, triste e neutra) pelo aplicativo ExProsodia. *IX Brazilian Symposium in Information and Human Language Technology and Collocated Events. Proceedings-BRACIS 2013* (pp. 1-8). Fortaleza, Brazil.
- Ferreira Netto, W., Peres, D.O., Martins, M.V.M., & Vieira, M.F. (2014). Automatic analysis of emotional intonation in Brazilian Portuguese. *The Journal of the Acoustical Society of America* 135, 2197.
- Griffiths, P.E. (1997). *What Emotions Really Are The Problem of Psychological Categories*. Chicago, London : The University of Chicago Press.
- Huckvale, M.A. (2008). Speech Filing System v.4.7/Windows SFSWin. Version 1.7.
- Mampe, B., Friederici, A.D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current biology* 19/23, 1994-1997.
- Nyggard, L.C., & Queen, J.S. (2008). Communicating emotion : Linking affective prosody and word meaning. *Journal of Experimental Psychology : Human Perception and Performance*, 34, 1017-1030.
- Parkinson, B. (1996). Emotions are social. *British Journal of Psychology*, 87, 663-683.

- Peres, D.O. (2013). The perception of emotion by native and non-native speakers. In Constantinou, H., Mazarella, D., & Varkanitsa, M. (Eds.) *First UCL Graduate Conference in Linguistics* (pp. 64-65).
- Plutchik, R. (1980). *Emotion, a psychoevolutionary synthesis*. New York : Harper & Row.
- Plutchik, R. (1984). Emotions : A general psychoevolutionary theory. In Scherer, K.R., & Ekman, P. (Eds.) *Approaches to emotion* (pp. 197-219). Hillsdale, NJ : Lawrence Erlbaum.
- Roberts, L. (2011). Acoustics effects of authentic and acted distress on fundamental frequency and vowel quality. *Proceedings of The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, (pp. 1694-1697).
- Russell, J.A. (1991). Culture and the Categorization of Emotions. *Psychological Bulletin* 110/3, 426-450.
- Scherer, K.R. (1981). Speech and emotional states. In Darby, J. (Ed.) *Speech evaluation in psychiatry*, (pp. 189-220).
- Scherer, K.R., Banse, R., & Wallbott, H.G. (2001). Inferences from vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology* 32, 76-92.
- Scherer, K.R. (2000). A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. *International Conference on Spoken Language Processing. Proceedings of ICSLP 2000*, Beijing, China.
- Scherer, K.R. (2006). The Affective and Pragmatic Coding of Prosody. In Huo, Q., Ma, B., Chng, E. S., & Li, H. (Eds.) *ISCSLP 06 Proceedings of the 5th international conference on Chinese Spoken Language Processing* (pp.13-14).
- Schlosberg, H. (1952). Three dimensions of emotions. *The psychological review* 61/2, 81-88.
- Spencer, H. (2011). The origins and function of music. In *Essays : scientific, political and speculative* Vol. 2, [First published in *Fraser's Magazine* for October 1857]. Accessed from <http://archive.today/7j0B>.
- Trnka, R. (2011). How many dimensions does emotional experience have? The theory of multi-dimensional emotional experience. In Trnka R., Balcar K., Kuska, M. (Eds.) *Re-Constructing Emotional Spaces : From Experience to Regulation* (pp. 33-40). Prague College of Psychosocial Studies Press.
- Tylor, E.B. (1871). *Primitive Culture : Researches Into the Development of Mythology, Philosophy, Religion, Languages, Art and Customs*. Vol 1. London : John Murray Albermale Street.
- Vassoler, A.M.O., & Martins, M.V.M. (2013). A entoação em falas teatrais : uma análise da raiva e da fala neutra. *Revista Estudos Linguísticos*. 42/1, 9-18.
- Wundt, W.M. (1897). *Outlines of Psychology*. Trans. by Charles Hubbard Judd. Leipzig, W. Engelmann ; New York : G.E. Stechert, p. 342.

Xu, Y., & Wang, Q.E. (1997). Component of intonation : what are linguistic, what are mechanical/physiological? Paper presented at *International Conference on Voice Physiology and Biomechanics*. Evanston Illinois.

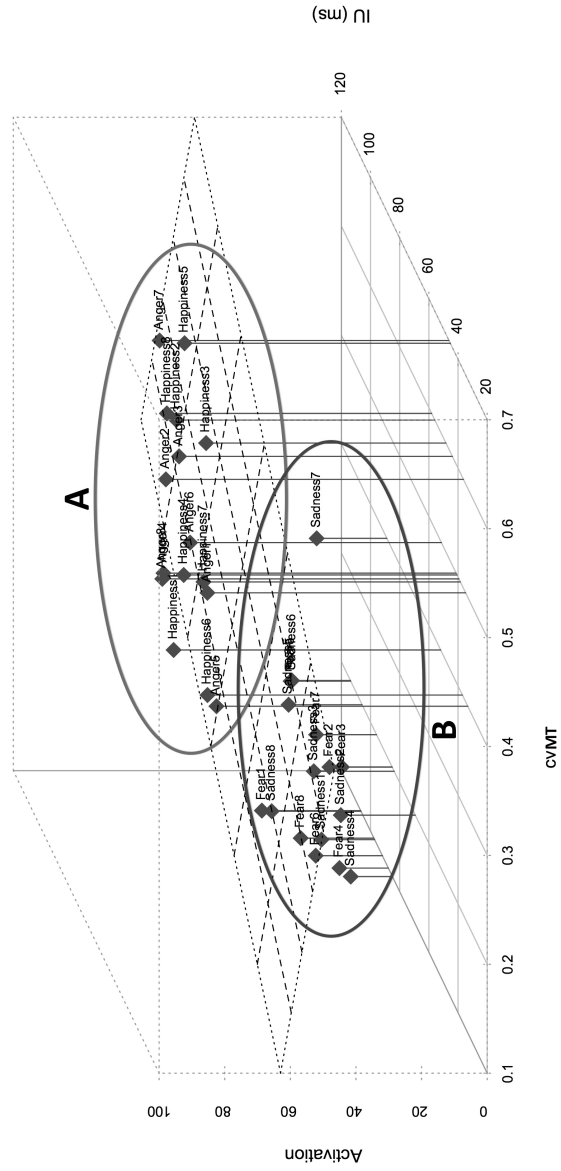


Figure 1 : Scatter plot of activation judgement (Brazilian) and acoustic parameters (cvMT and IU)

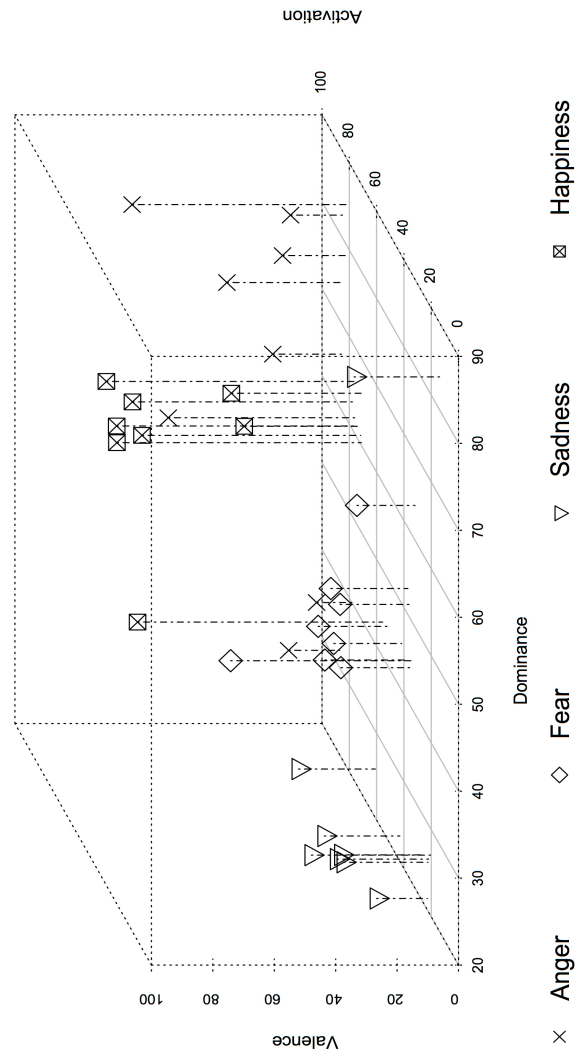


Figure 2 : Scatter plot of three dimensions (judgement made by Brazilians)

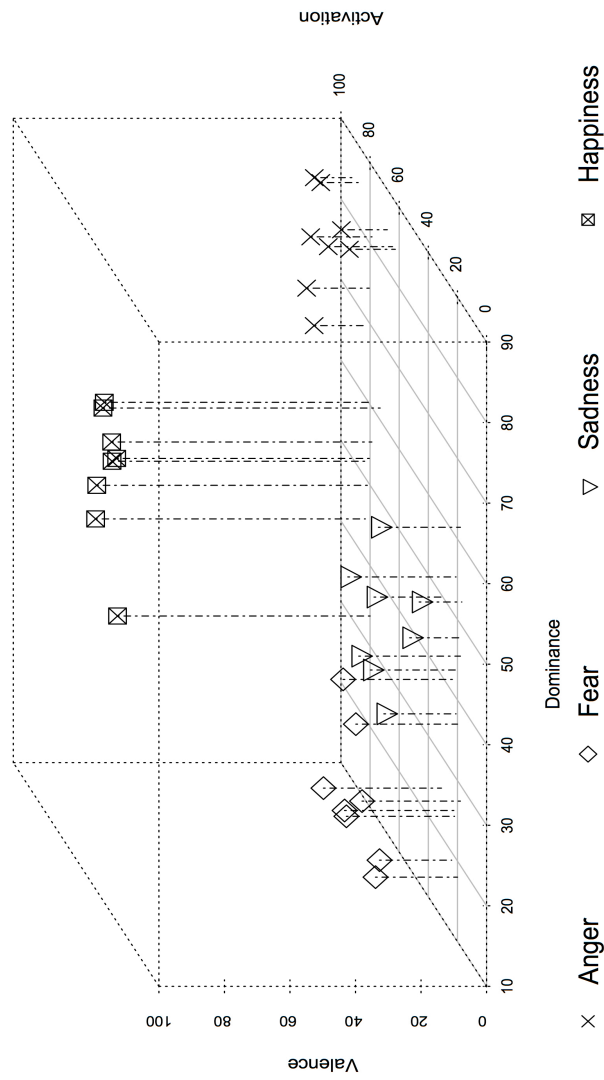


Figure 3 : Scatter plot of three dimensions (judgement made by English)