Ana Margarida Belém Nunes[1]

UDC 811.134.3'36

# PARTICULARITIES OF THE (NOT SO) EMOTIONAL SPEECH IN EUROPEAN PORTUGUESE: ACTED AND SPONTANEOUS DATA ANALYSIS

**Abstract:** *The present article is a symbiosis of two previous studies made by the author on European Portuguese Emotional Speech. It is known that nonverbal vocal expressions, such as laughter, vocalizations and, for instance, screams are an important source of emotional cues in social contexts (Lima et al., 2013). In social contexts we get information's about others emotional states also by facial and corporal expressions, touch and voice cues, (Lima et al., 2013 & Cowie et al, 2003). Nevertheless most of the existent research on emotion is based on simulated emotions that are induced in laboratory and/or produced by professional actors.*

*In this study in particular, it is proposed to explore how much and in which voice related parameters spontaneous and acted speech diverge. On the other hand, this study will help to obtain data on emotional speech and to describe the expression of emotions, by voice alone, for the first time for European Portuguese. Analyses are mainly focused on parameters that are generally accepted as more directly related with voice quality like F0; jitter; shimmer and HNR (Lima et all, 2013; Tiovanen et al, 2006; Drioli et all, 2003).*

*Given the scarcity of studies on voice quality in European Portuguese, it is important to highlight that this work presents original corpora specifically created for the presented research: a small corpus for spontaneous emotional speech and Feeltrace system to provide the necessary annotation and interpretation of emotions; a second corpus for acted emotions produced by a professional actor.*

*It is particularly important to highlight that was found that European Portuguese presents some specificities on the values obtained for neutral expression, sadness and joy, that do not occur in other languages.*

**Keywords**: *European Portuguese; Emotion; Voice analysis; spontaneous speech; acted speech; Feeltrace*

## 1. Introduction

There are, clearly, different ways of thinking and talking about feelings according to all different languages, cultures, societies, epochs and religions in the exactly same appraise that there are different ways to express and perceive those emotions.

[1] Associate Professor of Linguistics and Associate Head of Department of Portuguese at the University of Macau, China.

Nevertheless there are no doubts about the existence of shared characteristics and qualities being then possible to talk about emotional universals as stated by Wierzbicka, 2005.

Emotions influence physiological state, with important effects on speech production and especially on the phonation process. These effects are reflected in varied and complex voice quality related parameters, such as fundamental frequency (F0) and jitter. While some of the parameters are language independent, others are part of specificities of a Language, Culture or even Speaker. Many emotion theorists defend that emotions are mostly learned and affected by social environment. As a result, emotions are conjectured to vary considerably across cultures, (Sauter, Eisner, Ekman and Scott, 2010 and Adolphs, 2003 cited by Sauter and Eimer, 2009).

Speakers vary in their capacity to express, recognize and interpret attitudes or emotions. According to Scherer & Scherer 2011, cited by Lima et al., 2013, "Understanding the behavioral, socio-cognitive, and neural underpinnings of emotion perception in different channels is thus a topic of central importance". Research has shown that emotions are not equally recognized across genders, Cultures and languages. Zovato et al, 2003 using perceptual tests, demonstrated identification problems between the pair's neutral/sadness and joy/anger; Sawamura et al, 2007 found that, disgust and anger are similar, surprise and joy similar as well, and fear is often confused with sadness.

"Increased emotional arousal is accompanied by greater laryngeal tension and increased sub glottal pressure which increases a speaker's vocal intensity". For example, Darwin observed that angry utterances sound harsh and unpleasant because they are meant to strike terror into an enemy (Darwin, 1872).

Anger is usually associated with an increase in mean F0 and energy, it also includes 'increases in high frequency energy and downward-directed F0 contours. The increase of F0 mean and range is also a characteristic of fear, also with high frequency energy; sadness shows a decrease in mean F0, F0 range and mean energy; joy a positive emotion (one of the few that are usually studied), has an increase in mean F0, F0 range, F0 variability, mean energy, and an increase in high frequency energy (Banse & Scherer, 1996). 'Understanding a vocal emotional message requires the analysis and integration of a variety of acoustic cues'

(Schirmera & Kotz, 2006), those acoustic signals should be related to temporal aspects, intensity, fundamental frequency and, of course, voice quality related parameters, (Lima, Castro and Scott, 2013).

Johnstone and Scherer, 1999, have studies in which emotional vocal recordings were made using a computer emotion induction task. Voice quality acoustic parameters included F0 minimum, F0 range, jitter and spectral energy distribution. The emotions studied were: tense, neutral, irritated, happy, depressed, bored and anxious. The authors report that: 'values for jitter are correlated with F0 floor, thus indicating that period to period F0 variation tends to be larger with higher F0. This tendency is absent for anxious and tense speech though, which is in agreement with previous findings of a reduction of jitter for speakers under stress. Happy speech presents significantly higher values of jitter than all other emotions. Also as expected, F0 floor was found to be lowest for the emotions bored and depressed, and highest for happy and anxious speech.

It was also important to find out how well voice quality conveys emotional information to be perceived by humans and computers, Toivanen et al., 2006, carried out a study which had as informants nine professional actors producing data to be studied, simulating: neutral; sadness; joy; anger and tenderness states, in which they extract only a vowel from the entire running speech (approximately one minute). They considered vowel [a]. The samples were presented to 50 listeners to recognize the emotion and classify it using automatic methods. Human were better than the machine at recognizing anger, since humans are probably more 'tuned into anger' than the computer which ponders all emotions on a neutral basis.

Drioli, Tisato, Cosi and Tesser (2003) analyzed F0, duration, intensity, jitter, shimmer, HNR and other voice quality indexes such as Hammarberg Index. The authors used Praat voice report. Regarding irregularities, and for stressed vowels, they report a high shimmer value for anger; higher jitter values for joy and surprise (with anger in third place). The HNR is lower for anger and joy.

Chung explored acoustical properties of Korean emotional speech. The author measured: F0 parameters (mean, maximum, minimum, mean of the 20% lowest values, range), jitter, shimmer, speaking rate and spectral distribution. The analysis

showed that joy increases F0 mean, whereas sadness enhances the decrease of F0 minimum. The increase of F0 maximum and of F0 range was found to be 'a good indicator of the general emotional arousal'. 'The jitter and the shimmer values seem to increase under the emotional tension (...). However, these variations (...) were not statistically significant in the case of Korean data' (Chung, 2000).

As conclusion it should be said that Voice quality aspects are very often described qualitatively. In quantitative studies, the most considered parameters are the F0 related. More recently, the list of examined parameters expanded to include jitter, shimmer, HNR, glottal source parameters, among others.

## 2. Studying vocal emotion

*The Expression of the Emotions in Man and Animals* by Darwin (1872) is probably the start of descriptive and theoretical studies about expression of emotions. Since Darwin's work that several discussions, perspectives and theoretical works have been published, also other areas (like psychology, neurology, cognitive sciences, sociolinguistics) started to be interested on the analyzes of production and perception of emotions. Among different questions that are important in this field, the function of language and culture turn out to be very relevant and considered. According to Darwin, Culture does not have an important role on the expression of emotions once that for the author the recognition of emotions makes part of a biological heritage, therefore universally recognizable. Nevertheless, Darwin recognized that there were different societies, ethnicities, languages and even different cultures and social environment that could influence the expression of emotions, however the important matter of study was what Humans had in common with animals in what concerns to emotions.

The first problem one faces in starting the investigation is choosing valid *corpora*. These can be divided into three main categories: spontaneous speech, acted speech and elicited speech. All three present pros and cons (for further information: (K. R. Scherer, 1989; K. R. Scherer, 2003). Spontaneous speech was the option for the Belfast Naturalistic Emotional Database (Cowie, Cowie and Schroeder, 2003), which consists of 298 audiovisual clips from 125 speakers. The *corpus* was described according to several tiers of descriptors: impaired communication, pitch and

volume, timing, paralanguage, voice quality, articulation. Despite the controversies, many studies on voice quality and emotions use actors for *corpus* creation.

As an example of acted speech, Vogt, André and Bee, 2008, recorded 10 professional actors (5 men and 5 women) acting 10 utterances with 6 different emotions (anger, joy, sadness, fear, disgust and boredom) as well as a neutral emotional state. The sentences were semantically neutral. Consistent with other work, the authors found acted emotions to be more easily recognized than realistic emotions.

On this area not much research was conducted on the subject of emotional speech for European Portuguese. There is no *corpus*, big or small, of EP emotional speech available. Even the work on emotional speech synthesis of Portuguese (Cabral, 2006; Cabral and Oliveira, 2006) was based on information published for other languages, complemented by extraction of glottal parameters (such as open quotient) from a German database. The present work was also based on the studies that were made in other languages so that it would be possible to understand and know better what parameters should be analyzed.

## 2.1. Cross linguistic studies on Emotional Speech

Scientific studies have been carried out crossing speakers and listeners of several origins. In the following paragraphs some studies are mentioned to serve as background for these aspects.

According to Zinken, Knoll, Panksepp, 2010, the languages and cultures studied so far are not actually very diverse. Moreover, only a few specific emotions have been studied systematically, usually 'basic' emotions even if it can be universally accepted that anger is the better perceived emotion, even in a foreign language and the other emotions are better perceived in our native language and culture. Those aspects were comported and prove in this available study, showing that there is, in fact, an impact of linguistic and cultural aspects on recognizing emotions.

Abelin, 2004, made an experiment in cross-cultural multimodal interpretation of emotional expressions. The aim of the study was to investigate how speakers of Spanish and Swedish interpret emotions in each other's languages. The emotions studied were: sad and tired, angry, sad, skeptical, happy, afraid, depressed, very happy. Results show that Spanish

listeners were better at interpreting the Swedish speakers. Certain emotions, such as happiness and fear were more difficult to interpret only from prosodic information, by both groups.

Sawamura et al., 2007, in a study with Japanese, American and Chinese speakers, showed that there are some common factors, independent of language and culture that determine emotion perception in speech sounds. It was found that multiple emotional components were perceived in most speech materials, even when a single emotion was intended. Anger, joy and sadness seem to be the three basic emotions, while the other emotions converge to them.

Sauter et al., 2010 studied the differences between westerns and some isolated and remote Namibian Villages, finding out that the vocalizations of basic emotions were "bidirectionally recognized" while some other emotions were only perceived within but "not across, cultural boundaries." It is also interesting to notice that "a number of primarily negative emotions have vocalizations that can be recognized across cultures" while most of the positive ones were only understood using "culture-specific signals".

It is known that from the production point of view speech acts and expressive pattern are independent categories, once the emotions do not disfigure the melodic contours which are typical of the different speech acts. This is confirmed by the fact that the normally "proposed phonological representation for a neutral utterance is also applied to expressive utterances". In Brazilian Portuguese (from now on BP), according to a study carried out by Colamarco and Moraes, 2008, emotional patterns do not always affect the different speech acts in the same way. It was verified by the author that the relation between the emotional patterns is different in every speech act. It is possible then to assume that there it seems to exist a general tendency: neutral utterances and those expressing sadness present lower values for pitch level, average intensity and higher value for duration; in an utterances expressing joy and anger pitch level is higher and duration has a lower value. The expression of Anger and Joy also present similar values to what is generally described for other languages: an increase of pitch and average intensity of melodic contours, even if these emotions affect the F0 in BP in very similar ways, they were not confused in perception tests.

For sadness also BP presents different values, just like EP (European Portuguese), when compared to results for other languages. In BP sadness, when compared to neutral utterances, does not present a decrease on pitch and intensity but an increase. Values described for sadness in BP are closer to the ones reported for despair.

In another study reported for BP carried out by Peres, 2014, it was found that BP intonation parameters play an important part on the prediction, perception and distinction of emotional states. In this reported study F0 related parameters along with duration parameters were analyzed for BP. The 32 excerpts of spontaneous emotional speech stimuli were collected from a website. First, two Brazilians and two non-Latin speakers classified the stimuli (presented randomly) according to basic emotions: happiness, sadness, fear and anger. After this first analyzes 18 BP native speakers and 18 English speakers participated on the experiment. They had to indicate the valence (pleasant/unpleasant); activation (non-agitated/agitated); dominance (submissive/non-submissive). Results showed that the perception degree of activation could be predicted by some acoustic parameters of intonation. Regarding the degree of dominance - middle tone had significant results for BP; and the coefficient variation of medium tone and duration (intonation) had significant better results for English speakers, as example.

According to the author, there is an important difference between the two groups of participants: Brazilians were better differentiating each dimension (valence, activation and dominance) while English speakers were more confused. The author states that analysis showed that evaluation of non-native speakers could be explained by acoustic information, without the influence of lexicon. According to Peres, 2014, it is still necessary to find more acoustic parameters that could help to explain the differences between judgment made by BP and English participants. However it seems to be a linguistic component related to the perception of emotion in addition to the acoustic parameters (co variation principle) that may explain the performance of native speakers. But in the case of non-native speakers, the lack of linguistic knowledge of BP could explain their performance.

## 2.3. Describing emotions

The great difficulty on characterizing a certain type of voice and correlate it with a particular emotion has to do with the fact that there is a great personal and unique expression that varies according to each individual.

There are considerable evidences to prove that emotions are beyond a simple activation dimension of active/passive (aroused/sleepy) and pleasant/unpleasant (negative/positive). In addition, many studies were based only on few emotions that are often described as basic, therefore neglecting others more complex. Many theories of emotion do not present a detailed study of the feeling that emotion can cause to the listener. For example, one can expect differences on the activation dimensions and valences (continuous or discrete). These categories may well differentiate voice quality.

The evaluation of these theories also shows that emotional speech conveys some nuances that reflect the cognitive assessment and subsequent action tendencies that underlie each emotion (Patrik et al., 2008).

One way to define/describe emotions is through a bi dimensional image Active/Passive vs. Negative/Positive. Although this definition may vary from individual to individual it may clarify a better description of an emotion. Figure 1 represents Scherer (2005) description of emotions.

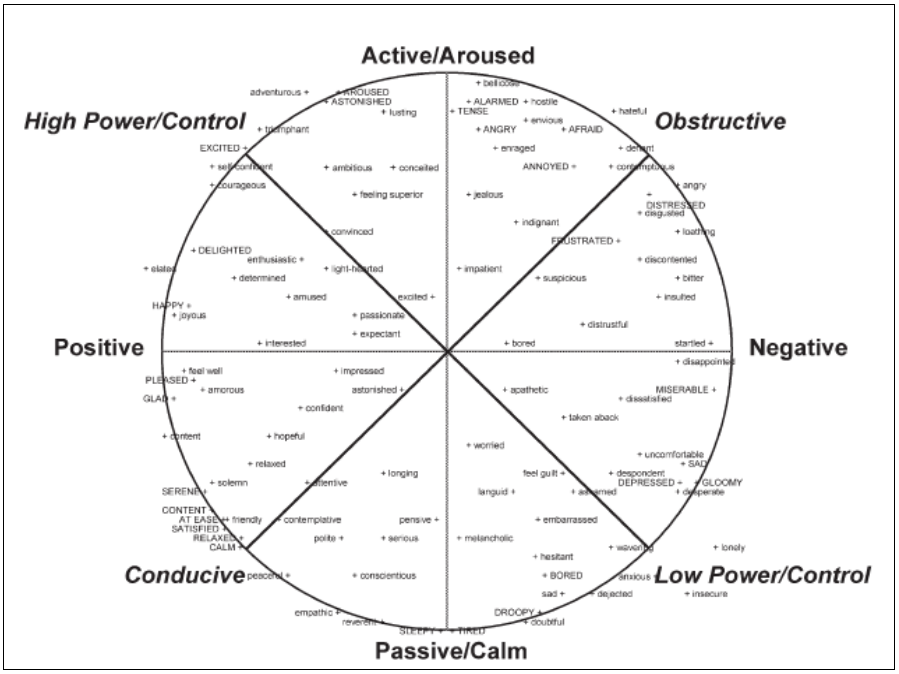


Figure 1: Description of Emotions presented by Scherer (2005)

In his study Scherer (2005) also presents a prototype tool for classification of emotions (created at the Geneva's University) called Geneva Emotion Wheel (GEW), Figure 2.

The author believes that the emotion concept became trivialized because is used very often. Nevertheless, facing the question "what is in fact an emotion?" there seems to be a consistent answer, and different research areas (like Humanities, Social Sciences and behavior) rarely reach an agreement. Emotion is then an episode that is a response to an internal or external stimulus (Scherer, 2005). In the first version, the Feeltrace was a tool that allowed identifying all the families of emotions through a specific parameter, which became visible by moving the mouse inside the circle. However, this experiment indicated that it was problematic to measure the intensity of emotions. Thus, a newer improved version could provide different measures and values. The new analyzes were related with the degree of distance or approach that all the emotions were from the neutral state and didn't have much association with emotions family. The GEW looks to be the first instrument to provide a true quality sampling of emotions in a bi-dimensional

space (present: positive or negative, intensity: distance from the neutral state).

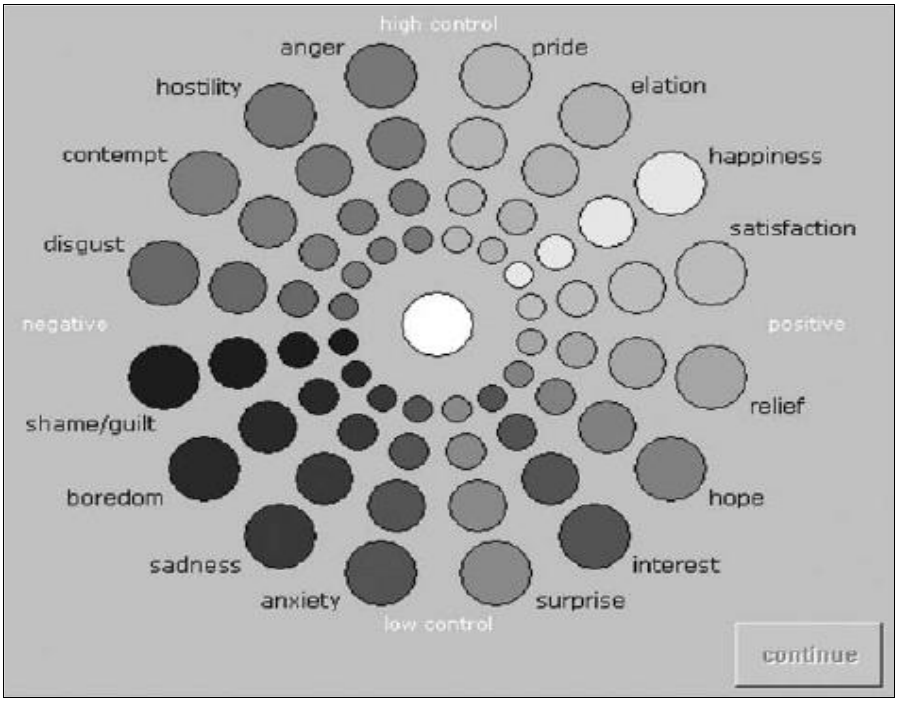



Figure 2: Geneva Emotion Wheel (GEW), Scherer (2005)

## 3. European Portuguese Spontaneous Speech

A *corpus* was created and developed for this particular study. It was recorded from live different Talk Shows of Portuguese National TV channels. The first task was to clean noise from street and talk shows interviews (street noise, people talking, etc.). The *corpus* consists of 20 statements, representing joy, sadness and anger, plus the neutral speech. The emotions that were analyzed are those that one thought to be easier to reach with the resources available, are encompassed all families of emotions. Although this *corpus* is necessarily limited, brings us quite new and original information in scientific terms for this area.

Knowing the limitations that a *corpus* of spontaneous speech features, and even recognizing it is easy to trigger different feelings in an individual, it is also important to follow the ethical principles and the right to privacy of all individuals. The recordings have some problems in terms of noise and sound

quality, however, does not impede, in general, analysis and extraction of important results.

### 3.1. Phonetic annotation and acoustic parameters extraction

All the selected segments were annotated in Praat system using SAMPA. Subsequently a feature extraction using Praat was held copying all required values and time information of each sentence to Excel. Throughout the process Praat was not able to provide accurate results of some segments, in those cases it was possible to replace the utterance for another clear one. Statistical analyzes were then performed, as well as comparisons between the various parameters and values that were obtained.

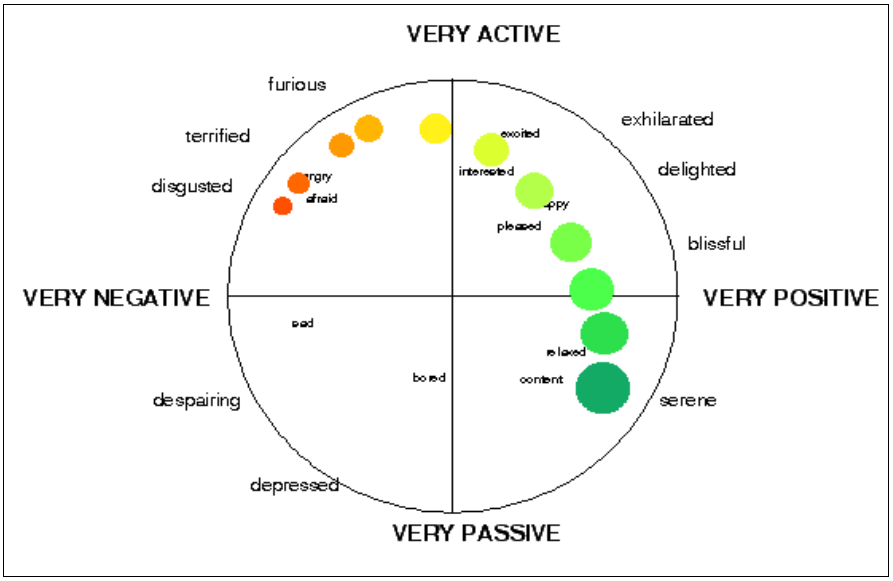
## 4. Feeltrace

Feeltrace (Figure 3) is a tool designed to allow the listeners registration of the emotion they are perceiving, and their changes over the utterance, ie, in a dynamic way. It is based on the activation space/assessment that derives a representation of psychology. The extent of activation shows how an emotion can be dynamic and assessing how it can be manifested: positively or negatively (Cowie et al., 2000).

Feeltrace has some distinctions where it fails, for example between anger and fear (Cowie et al, 2000). However, for many emotional states through the analysis of this system it is possible to have a very important starting point. Duration parameters are certainly the most difficult to analyze taking into account only the voice. A major difficulty in research, when it comes to emotion, particularly spontaneous and when the analysis is on fluent speech samples (as opposed to sustained vowels), is to realize its gradation, how varies over time.

Freeltrace is an instrument developed to allow listeners to describe certain stimulus in a continuous (in time) and dynamic mode. It is based on the idea of representation space of activation and evaluation advocated by psychology. It is easy to use and at the same time enables reliable results in terms of scientific research. The activation dimension measures the degree of dynamism of an emotion (active / passive); the dimension of evaluation allows the distinction between positive and negative feelings associated to a stimuli. Research also suggests that space is naturally circular, i.e., the strongest emotions (in terms of

intensity limit) do form a circle and, therefore, the neutral one is presented at the center.



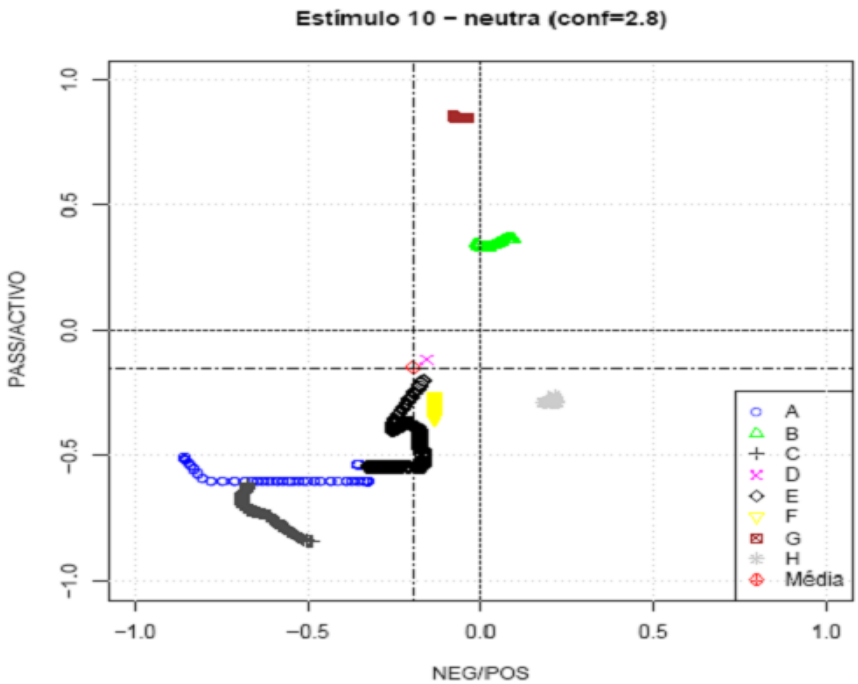




Figure 3: *Feeltrace* during a record session of a neutral utterance. It is perfectly clear the informant's confusion on identifying the emotion. The stimuli felt much more like sadness than neutral.

### 4.1.1. F0 values

Analyzing the figures for F0 (Figure 4), it is possible to make the distinction between men and women, since this parameter varies significantly according to gender. The recordings that were used allowed analyzing for women all the emotions. However, for men it was only taken into account: sadness, neutral and joy.

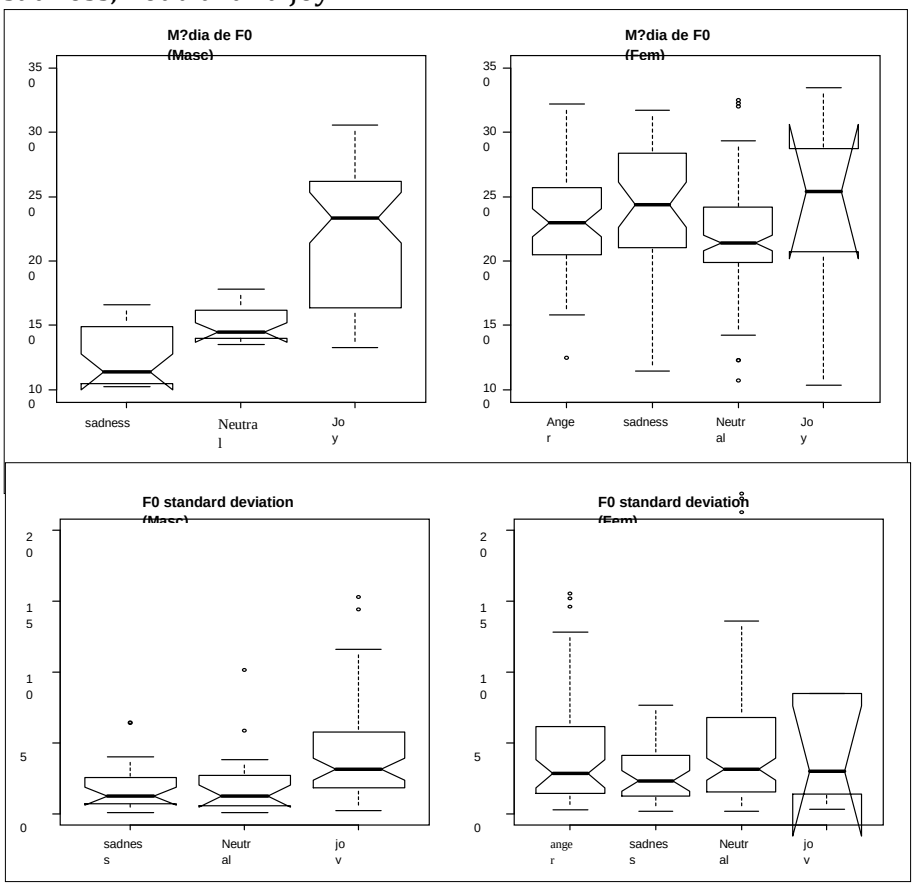



Figure 4: F0 values for men and women, in that order.

In a global analyzes of the results one realizes that in all the informants (either men or women) F0 is always higher when it comes to the manifestation of joy.

On average F0 for male is perceived an evident high register when expressing joy, with a value close to 230 Hz. For neutral speech it is observed that values for F0 are around 150 Hz and with lower values sadness with about 120 Hz. In terms of standard deviation it is clear that for men this is greater in

utterances that express joy, understanding that between the neutral speech and sadness expression there is no significant difference.

In analyzing the results for women, it is possible to have a wider perspective, since data allows to analyze all the emotions under study. Therefore, joy and sadness in terms of average F0 for PE are very close. There is a very slight difference on the expression of sadness with 230 Hz and happiness, which is close to 250 Hz. Thus appear two completely opposite families of emotions (positive/ negative) with similar results in terms F0 average values.

Anger appears in-between position near to 225 Hz, and finally neutral speech with about 215 Hz. Particularly there is in terms of females F0 average very small differences between anger expression and neutral state.

Thus, in general terms, observing the two groups it is possible to state that the differences are more prominent in men; that joy is the emotion which stands for both groups, with F0 higher values and, the lowest values, are related to sadness.

### 4.1.2 Jitter

In what concerns to jitter examination (Figure 5) it is possible to more directly compare each emotion with the neutral speech. Becomes easier to observe and highlight the possible similarities and / or differences in relation to each other and all in relation to the average level, which will be neutral.

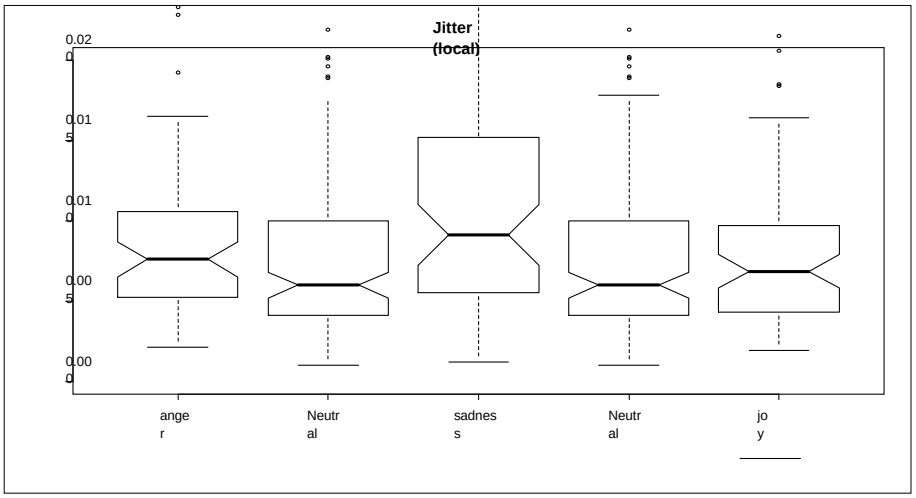



Figure 5: J*itter* Local values for the different studied emotions

The graph in Figure 5, referring to the Jitter Local, allows to observe a clear difference in the values for sadness, which are higher in relation to all other expressions in study.

Joy and anger are very similar and in closer positions in relation to the neutral speech than in relation to sadness. The neutral sentences are the ones presenting jitter lower values.

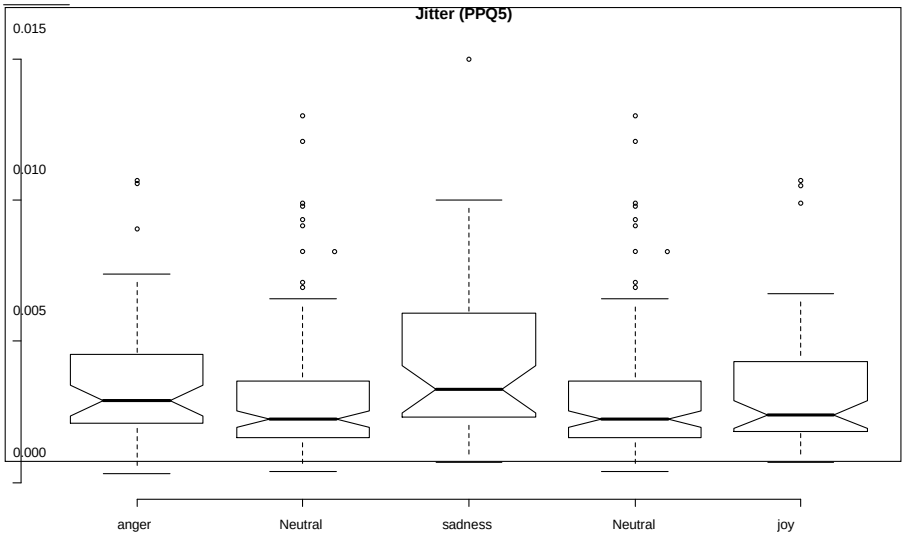



Figure 6. J*itter* PPQ5 values for all the studies emotions

In this graph it is possible to observe, that sadness stands out, with the highest value for jitter PPQ5. It is noteworthy that for jitter PPQ5, joy and neutral speech have the lowest values and all similar, with no significant difference between these productions with regard to this parameter analysis. Anger seems to be closer to the values presented for sadness than the ones for neutral. This nearness comes from the fact that in this case the analyzed segment it was very close to suppressed rage (cold anger), since the recordings analyzed were taken from a television program, not as denoting high values in any of the parameters analyzed.

### 4.1.3 Shimmer

A detailed study of the figures presented for Shimmer Local and APQ3 shows a comparison between emotions and the average value reported for neutral speech (Figure 7).

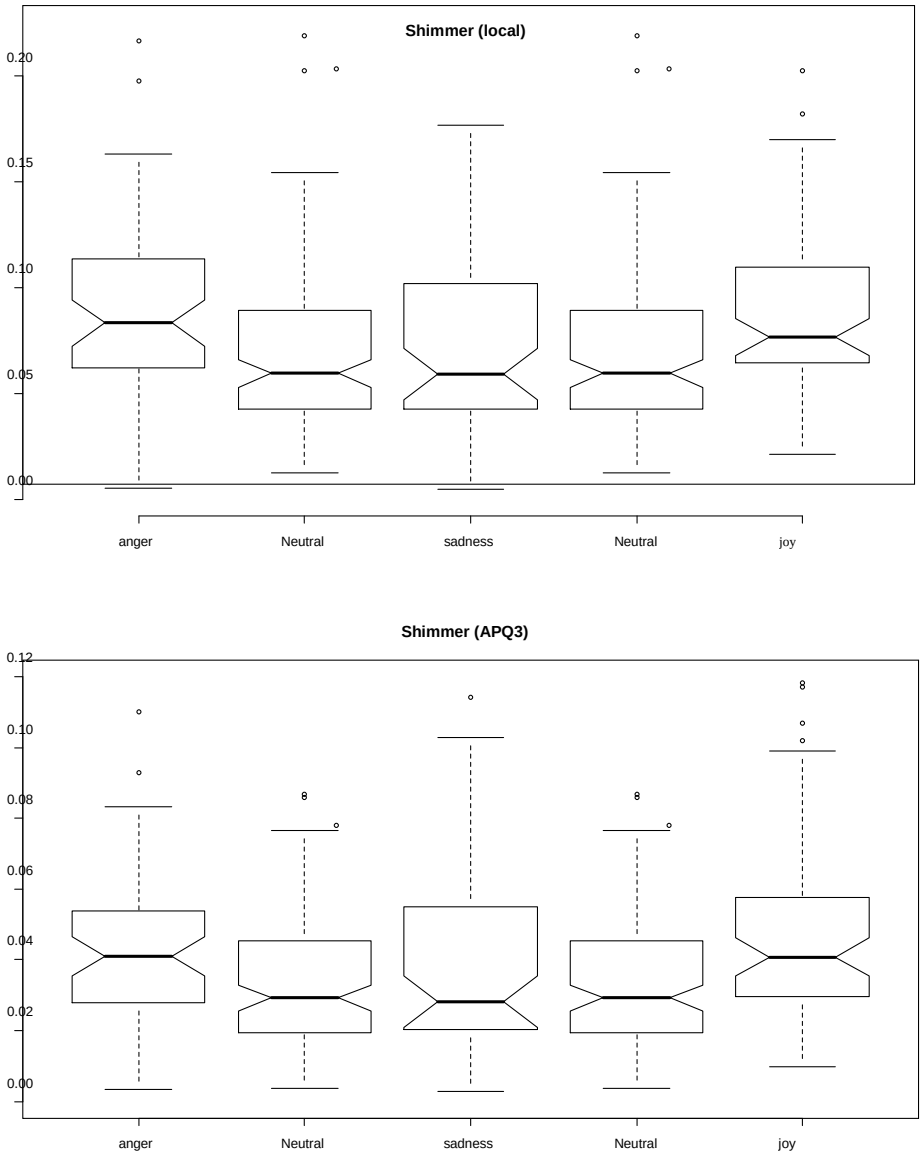



Figure 7: *Shimmer* local and APQ3 values for the four different emotions and neutral state

Observing of Figure 7, it is noticeable that the parameter values are higher for anger and happiness, two emotions interestingly distinct. This may reflect the intensity that was gave to each of them.

On the other hand expression of sadness has lower values, even lower than those reported for neutral state.

### 4.1.4 HNR

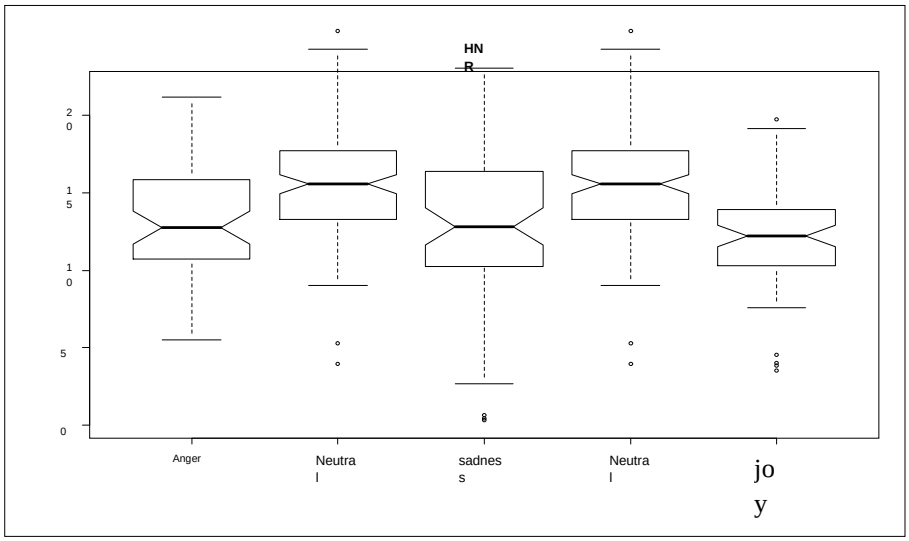



Figure 8: HNR values for the five utterances

Taking as reference neutral speech it is possible to verify that all the 3 analyzed emotions present values below normality.

Anger and sadness have very similar values of 13 dB and the expression of joy, although very close to other emotions, shows the closest reference value for the normal 12 dbs. All of them are, however, at the threshold of normality in terms of noise / harmony.

**4 A) Explaining the "normality" for voice parameters**

In other study for uncontrolled and pathologic voices one also used the Hoarseness Diagram (Fröhlich, Michaelis et al. 2000), a "new approach to the acoustic analysis of pathological voices combining several acoustic measures" (Michaelis, Fröhlich et al. 1998). This diagram separates two independent acoustic measures, the irregularity component (Irreg) and the noise component (Noise). For clinical application the speakers' data are plotted in the diagram as ellipses representing the mean and standard deviation of the two factors. On the previous study only sustained values were analysed according to the program requirements (segments with more than 500 msec duration) made it unsuitable for sentences by Michaelis (1998) for which normal and pathologic values have been studied. Thus it is

possible to know when some parameters of an individual's voice are approaching irregularity consequently close to pathology. Those studied factor for voice quality and emotion was extremely appreciated by actors and TV presenters, once that they would try to be more careful.

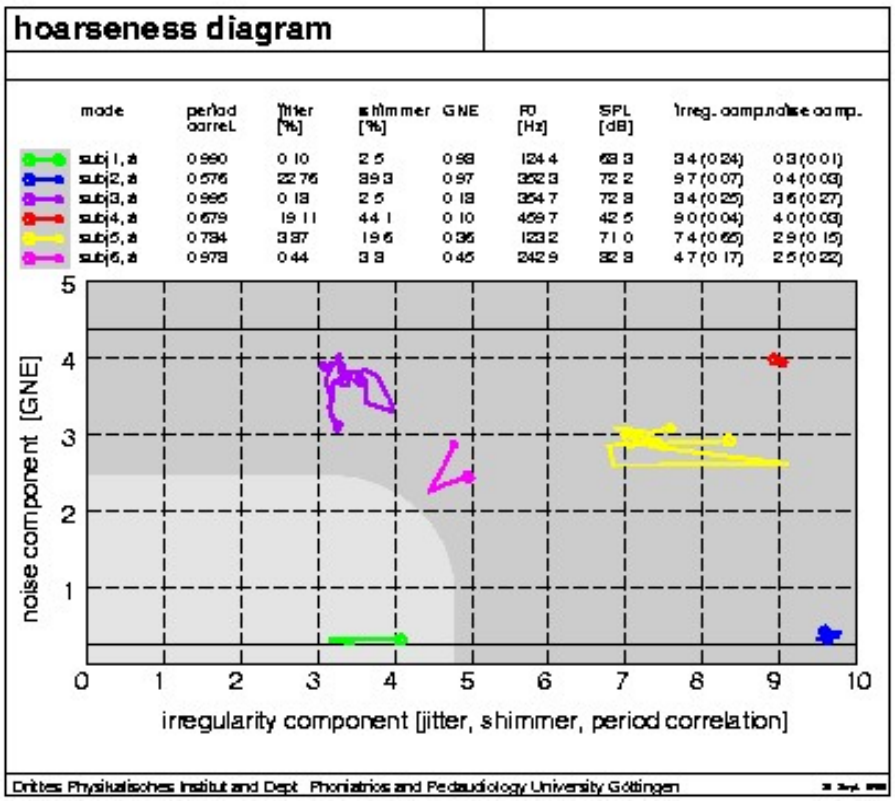



Figure 9: Hoarseness Diagram, (Fröhlich, 1997)

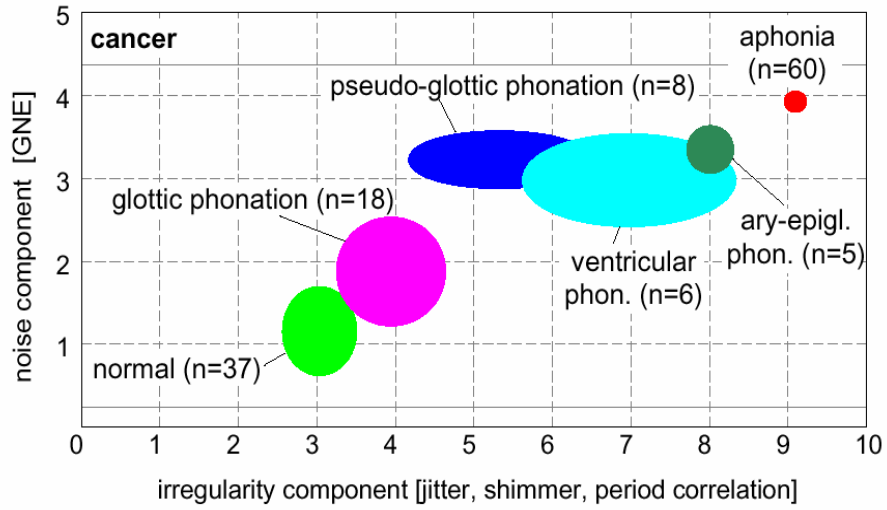



Figure 10 – *Hoarseness Diagram* (Fröhlich, 1997) ellipses reporting voice pathologies

## 5. European Portuguese acted emotional speech

The *corpus* is composed by two sentences, one simple and one complex, both extracted from the Portuguese version of the naturalistic dialogue "The human voice" by the French writer Jean Cocteau (1998). The sentences were recorded one after another not giving the actor time to prepare and concentrate, once that it was thought that it would help on a more spontaneous speech. The complete *corpus* had 42 utterances (once that he repeated 5 times each sentence conveying 5 emotions plus the neutral state). The *corpus* was thus constituted by the simple sentence "O melhor será tomares conta deles" /u m@LOr s@ra tumar@S ko t6 del@S/ (You've better take care of them) and the complex sentence "N˜ao tenho com certeza a voz de uma pessoa que esconde qualquer coisa" [/n6 w t6Ju Ko s@rtez6 6vOS djum6 psow6 kiSko d@ kwalker kojz6/ (I don't really have the voice of a person who hides something). The chosen sentences do not present by themselves any emotional charge or meaning, so the actor may well interpret them according to the intended principles: joy, despair, anger, fear, sadness and the neutral form. The informant was a professional actor, male.

Sentences were first annotated at word and phone levels, using SAMPA *(Speech Assessment Methods Phonetic Alphabet)* transcription in SFS (Speech Filing System). The limits of each

segment were marked and a broad phonetic transcription was made, considering phenomena such as elision, crasis and addition of certain sounds. All data was processed in Praat software (Boersma, 2001), which allowed the extraction of the needed elements using the Praat Voice Report function. Analyses were made in SPSS (v. 16) and R. As part of the parameters from a Normal distribution, non-parametric tests were employed.

The following figures (11 and 12) present both sentences in Phonetic transcription:

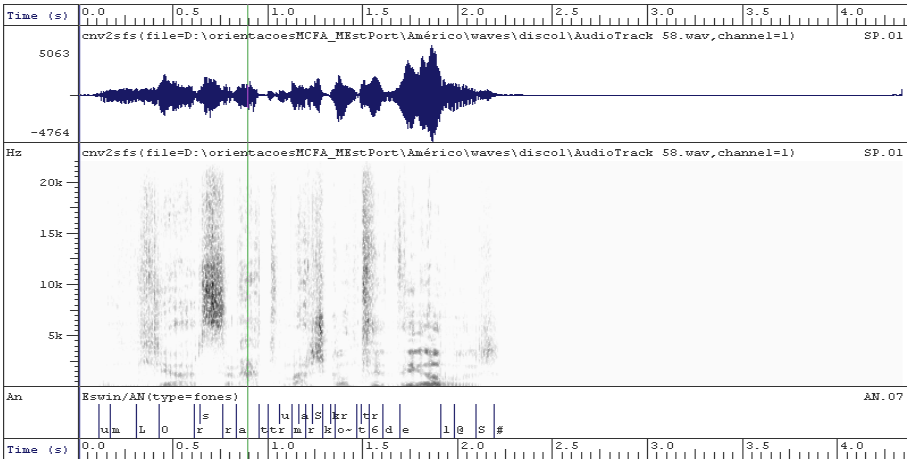



Figure 11 – Simple utterance - /u mLOr sra tumarS ko~nt6d el@s/

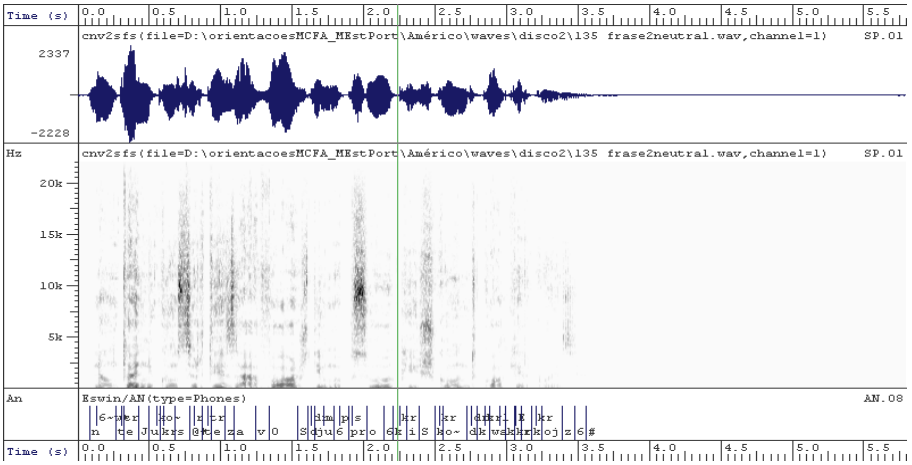
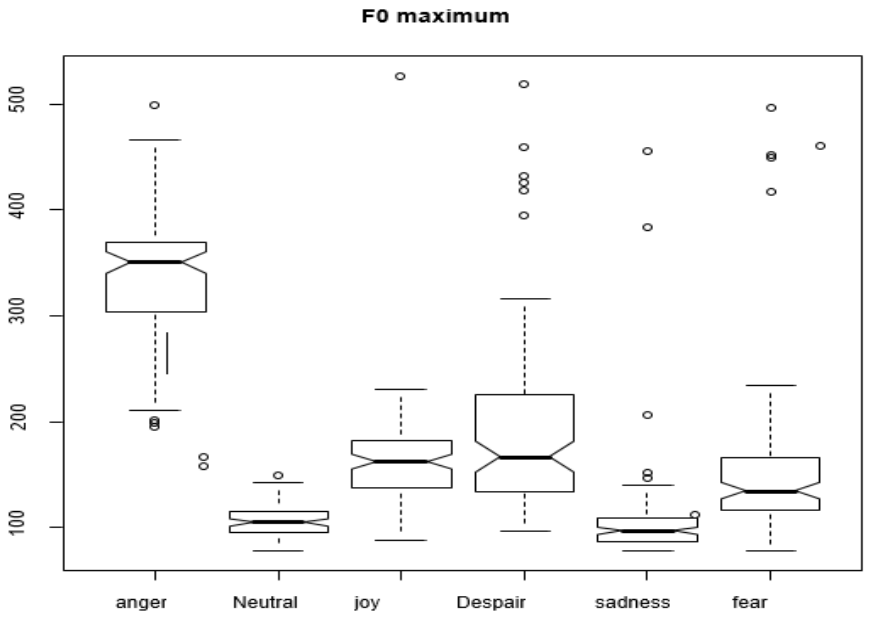



Figure 12 – Complex utterance - /n6~w~ t6Ju Ko~ s@rteza vOS djum6 psow6 kiSko~d@ kwalker kojz6/

On this study four different F0 related parameters were considered; F0 minimum, max, mean and F0 standard deviation.

Data analysis of different F0 parameters shows that anger is clearly differentiated, presenting an average value near 300 Hz and the highest standard deviation and range.

Joy and despair present similar values on the four F0 parameters, with mean around 150 for F0 mean and F0 max. One difference between the two is the higher range of values for despair. Fear has F0 values lower than the previous pair. Standard deviation is also lower. Sadness presents the lower values for those parameters, comparable with neutral. For F0 maximum, minimum and mean, all pairs are significantly different except for despair-fear, despair-joy, fear-joy and neutral-sadness. For F0 standard deviation, also the following pairs were not significantly different: fear-neutral, fear-sadness and joy-sadness. It can be said that some pairs are difficult to differentiate based on F0 parameters only. The standard deviation presents the lowest discrimination power.



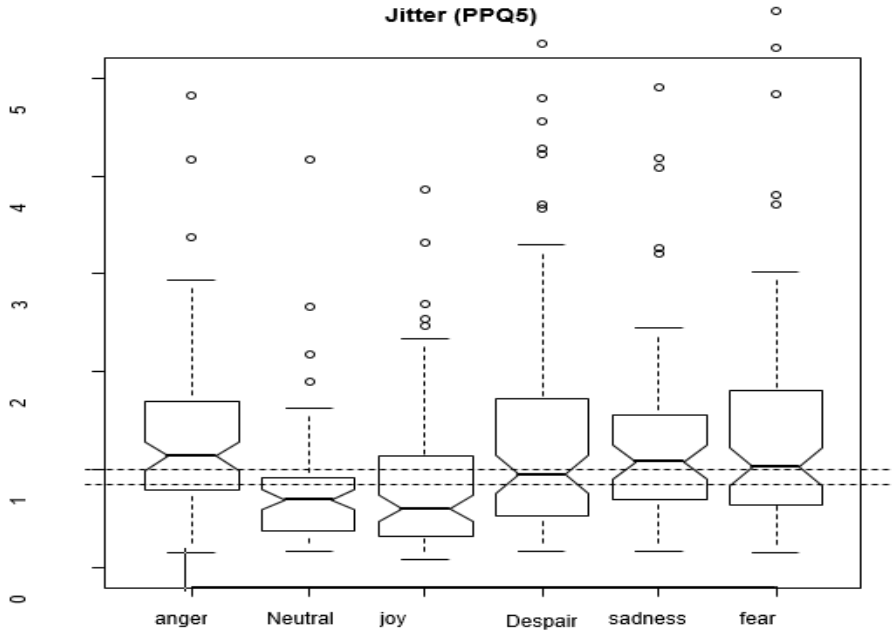



Figure 13 – Maximum F0 and Jitter values for each emotion

Jitter and Shimmer were also studied. For Jitter it was only contemplated PPQ5 parameter. Results showed that higher jitter values are associated with despair, fear, anger and sadness the most negative emotions. The neutral speech and joy present lower or similar values. Regarding jitter values, joy appears clearly lower than three of the other emotions. Jitter seems a relevant factor to detect joy.

For Shimmer parameters it is clear that they are particularly high for anger, followed by the group that combines despair, sadness and fear. Anger only does not present significantly higher shimmer values than sadness and despair. Nevertheless despair also has significantly higher shimmer values than joy and neutral; other emotions present no significant differences. Shimmer only differentiate anger and despair from all the remaining.

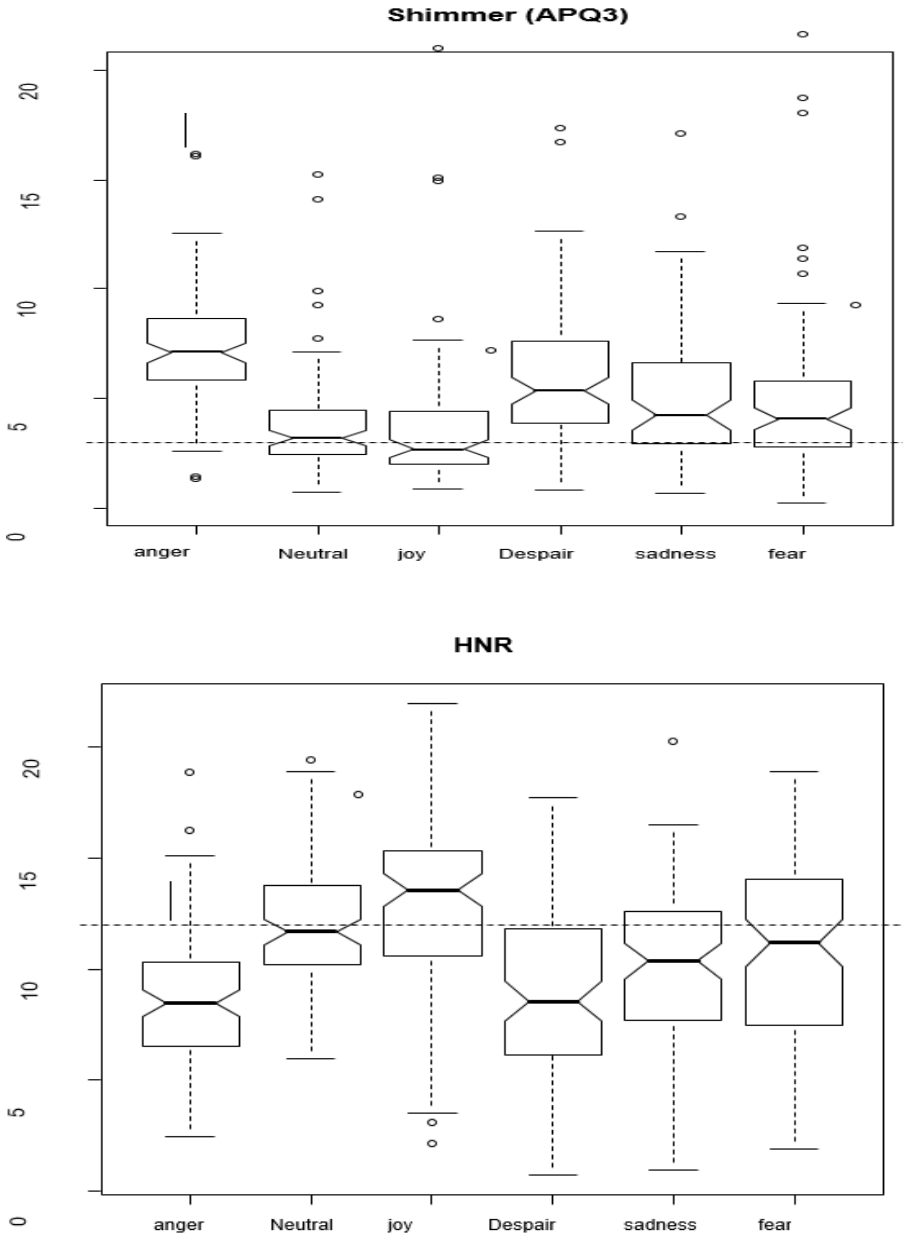



Figure 14 – Shimmer and Harmonic Noise Ratio values for each emotion

## 6. Conclusion

The analyzed parameters are the most common, providing a good starting picture of the effects of emotions on the voice quality, but there is a big margin for improvement. The glottal

source parameters, such as Open Quotient, and the spectral parameters should be also considered.

For spontaneous speech it was possible to observe that F0 maximum and average differentiates anger, sadness and joy as reported by Banse \& Scherer (1996) and Cowie et al. (2003). Anger presents the highest F0; joy an equally high value; sadness the lowest value. Sadness, as reported by Cowie et al. (2003), has values close to neutral. One's measures don't confirm the increase of F0 for fear. As Scherer (cited by Airas \& Alku, 2004) suggested, our F0 measures correlate with activation dimension; high activation relates with higher values of F0.

The comparison of jitter and shimmer values with the literature is more difficult. Firstly, there are few studies reporting such parameters; secondly, there is some uncertainty on the exact parameters report (ex: Local, PPQ5); thirdly, the process of parameter extraction is not necessarily equivalent.

Regarding shimmer results only anger does not present significantly higher shimmer values than sadness and despair, and those are in agreement with published measures such as by Drioli et al. (2003) showing high shimmer values for anger.

In this research concerning to jitter, joy appears clearly lower than three of the other emotions. Jitter seems to be a relevant factor to detect joy. The similarity of jitter for joy and neutral observed for EP is in agreement with the results obtained by Monzo, Alías, Ignasi, Gonzalvo \& Planet (2007) for Spanish happy and neutral speech.

Also in agreement with the revised literature (Drioli et al. (2003)), HNR is lower for anger and for despair. Contrary to some work, were verified significantly higher values of HNR for joy, placing this emotion far from fear relative to HNR has was stated in Drioli study.

The differences observed for joy can be related to the difficulty on identifying the emotion in perceptual tests, in agreement with the observations of Darwin, joy is more difficult to be conveyed by voice alone. Joy was often confused with neutral, by native and non-native EP speakers. The *corpus* is too reduced to generalize, but this question of possible differences in expression of joy or relative ineffectiveness of the actor expressing this emotion, recommend follow-up studies.

For some of the emotions, parameters such as HNR and jitter present values in the 'pathological' ranges usually

considered in voice evaluation. This urges the necessity of controlling the emotional state of the subject to whom a voice evaluation procedure is applied. Being sad or happy has noticeable effects on the 'normal' (Michaelis, Dirk; Frohlich Mathias; Strube, Hans Werner, 1998) range of several parameters.

## 7. Discussion

Comparing the results obtained for spontaneous and acted emotions and taking into account the analysis of the related F0 there is a consistency between the results obtained for the two *corpora*. Thus, in terms of average F0, speaking only in the context of male voice (once that for spontaneous emotion also have recordings of female voice), joy presents the highest values and the difference between the expression of sadness and neutral speech remains minimal. The most active emotions have higher values of F0, according to Scherer (2004), which corroborates the results of this study. Spontaneous expression of joy that makes of this *corpus* is, in most cases, euphoria (since it is linked to the football excitement), thus explaining the higher values than for anger.

In terms of F0, one was unable to study the values for anger in the male discourse. However in this same emotion the female discourse does not have higher values than joy on F0 average. Both present very close values in terms of standard deviation. It is observed that in PE sadness reports lower F0 values than neutral speech, confirming once again the values presented by Cowie et al (2003).

On acted emotions the highest jitter values are associated with anger and sadness (also despair and fear, but those were not considered in the study of spontaneous emotion). Thus, negative emotions have higher values of jitter. Also, for spontaneous emotion sadness has the highest value. Referring to the study of Shimmer related parameters it is seen that with values above the neutral speech it is possible to find only anger and joy. This parameter shows a deviation from the results obtained by the analysis of emotion produced by the actor. While in the study of acted emotions anger is the far above the value of neutral speech and happiness, in turn, is below a threshold.

Finally, considering the values of spontaneous emotion, in what concerns the Harmonic Noise Ratio, it was found that all the

emotions present results close to those considered normal values close to 13 dbs. It is noteworthy that the neutral speech is the one with highest values close to 15 dB, which was the value reported for joy in the study of acted emotions.

In spontaneous emotion HNR gives us very similar values in all emotions and neutral speech, what does not occur on acted speech analysis. This parameter may be one of the most significant in distinguishing emotions regarding acted and spontaneous speech.

In the study of emotion by actor, anger appeared in an area that would be potentially considered pathological.

Regarding the analysis of spontaneous emotions in Feeltrace it is possible to observe that the results for shimmer are the opposite of those obtained for jitter. Shimmer results are more negative and more passive and in jitter more positive and active. Resembling shimmer results to those previously observed for the values of the two F0 parameters analyzed (for men). The results HNR, despite being very similar between them, are more comparable to the results obtained for F0 in which there is an increase in negative and passive axes.

For PE the results to retain are: the proximity of values between the neutral speech and the expression of sadness when in terms of shimmer; the similarity of jitter values between joy, anger and neutral expression and, finally, the close proximity of HNR values of the anger, joy and sadness.

Particularly with regard to values so close between sadness and neutral state, joy and neutral expression and the closeness of values of Harmonic Noise Ratio for joy and sadness, can be indicators of a cultural condition in which the expression of sadness or joy shows no major differentiating values when compared with the neutral speech.

## References

Abelin, A. "Cross-Cultural Multimodal Interpretation of Emotional Expression - An Experimental Study of Spanish and Swedish", *Speech Prosod*y, 2004.

Boersma, P. "Praat, a system for doing Phonetics by computer". *Glot International,* 5(9/10), 341-345, 2001.

Castro, São Luís, and Lima, César F., "Recognizing emotions in spoken language: A validate set of Portuguese sentences and pseudosentences for research on emotional prosody", *Behavior Research Methods*, 42(1), 74-81, 2010.

Colamarco, Manuela, and Moraes, João António de, "Emotion expression in speech acts in Brazilian Portuguese: Production and Perception", *Speech Prosody,* 2008.

Chung, S.-J. "Expression and Perception of emotion extracted from the Spontaneous Speech in Korean and in English", Sorbonne Nouvelle University, Paris, 2000.

Cocteau, J. "A Voz Humana [The Human Voice]: Assirio and Alvim", Portuguese translation, 1989.

Cowie, E. D., and Cowie, R., Schroeder, M. "The description of naturally occurring emotional speech", *Phonetic Sciences (ICPhS)*, 2003.

Darwin, C. ,"The Expression of Emotions in Man and Animals", Portuguese edition, *Relógio D`Água*, 2000.

Drioli, C., et al. "Emotions and Voice Quality: Experiments with Sinusoidal Modeling", *VOQUAL*, 2003

Gobl, C., and Chasaide, A. N. "The role of voice quality in communicating emotion, mood and attitude", *Speech Communication*, 40, 189-212, 2003

Lima, César F., Castro, São Luís, and Scott, Sophie K. "When voices get emotional: A *corpus* of nonverbal vocalizations for research on emotion processing", *Psychonomic Society, Inc*., Springer, 2013.

Johnstone, T., and Scherer, K.R., "The effects of Emotion on Voice Quality", *Phonetic Sciences* (ICPhS), 2003.

Martins, C., Lemos, A.I. de, and Bebbington, P.E. , "A Portuguese/Brazilian study of Expressed Emotion", *Social Psychiatry and Psychiatric Epidemology*, 27: 22-27, Springer, 1992.

Michaelis, Dirk, Frohlich Mathias, and Strube, Hans Werner, "Selection and combination of acoustic features for the description of pathologic voices", *Journal of Acoustical Society of America*, 1998.

Monzo, C., Alías, et al. "Discriminating Expressive Speech Styles By Voice Quality Parameterization", *Phonetic Sciences (ICPhS), 2007.*

*Moraes, João António, et al. "Multimodal perception an*d production of attitudinal meaning in Brazilian Portuguese", *Speech Prosody*, 2010.

Nunes, Ana, Rosa Lídia Coimbra, and António J. S. Teixeira. "Voice Quality of European Portuguese Emotional Speech", *Computational Processing of the Portuguese Language*, Volume, 6001: 142-151, Springer, 2010.

Paeschke, A., and Sendlmeier, W. F., "Prosodic Characteristics of Emotional Speech: Measurements of Fundamental Frequency Movements", Technical University, Berlin, 2003.

Peres, Daniel Oliveira, Intonation as a cue to emotional speech perception: an experiment with normal and delexicalised speech", *Phonetic Sciences (ICPhS),* 2014.

Sauter, Disa Anna, and Eimer, Martin, "Rapid detection of Emotion from Human Vocalizations". *Journal of Cognitive Neurosciences* 22:3, 474-481, 2009.

Sawamura, K., Dang, et al. "Common Factors in Emotion Perception among Different Cultures", *Phonetic Sciences (ICPhS)*, 2007.

Scherer, K. R. "Vocal communication of emotion: A review of research paradigms". *Speech Communication*, 40, 227-256, 2003

Teixeira, António, Nunes, Ana Margarida Belém, et al, "Voice Quality with psychological origin: a case study", *Clinical Linguistics & Phonetics*, 22 (10/11), 906-916, 2008.

Titze, Ingo R., "Principles of Voice Production", Prentice Hall, Englewood Cliffs, 1994.

Toivanen, J., et al. "Emotions in [a]: a perceptual and acoustic study". *Logoped Phoniatr Vocol* 31, 43—48, 2006

Vogt, Thurid, and André, Elisabeth, "Comparing Feature Sets For Acted And Spontaneous Speech in View of Automatic Emotion Recognition", *IEEE*, 2003.

Vogt, T., André, E., and Bee, N., "EmoVoice - A Framework for Online Recognition of Emotions from Voice", 5078, 188-199, Springer, 2008.

Zovato, E., et al. "Towards Emotional Speech Synthesis: A Rule based approach", *Speech Synthesis Workshop*, *(ISCA),* Pittsburgh, 2004.