

## **Estudo de emoção espontânea em Português Europeu utilizando Feeltrace – primeiras observações**

*Spontaneous Emotional speech in European Portuguese using Feeltrace system – first approaches*

NUNES, A.<sup>1\*</sup>; TEIXEIRA, A.\*

<sup>1</sup>University of Macau

---

### **Abstract**

*The present work studies spontaneous emotional speech in European Portuguese, comparing the values obtained in measuring acoustic parameters in this study with the ones known for other languages and for simulated. The spontaneous expression of emotions recorded in the street or from television programs may bring very relevant results, but even in those situations it is difficult to determine the individual's true emotion (it is often only possible to be identify through the context) (Patrick, 2008). The main focus of our study is on voice related parameters such as F0, jitter, shimmer and HNR that will be liable to compare with previous studies of emotions produced by an actor. Using the Feeltrace system it is possible to better describe the sensation caused by the stimuli (in any of the emotions in question) and describe in more dynamic manner the emotion. We also realize that, although the emotions produced by actors show some differences, because they are necessarily exaggerated, the parameter values are all very similar, valuing the studies that are usually obtained based on utterances produced by performers. Given the scarcity of studies on voice quality in European Portuguese, it is important to highlight that this work presents an original corpus specifically created for this study. We recognize the limitations of the presented research seeing the results as first approaches that must be deeply studied being important to areas such as robotics and health technology.*

Keywords: European Portuguese; Voice; Emotion; Perceptual Phonetics; Feeltrace

---

## 1. Introdução

O estudo apresentado neste artigo pretende analisar correlatos acústicos da expressão de emoção espontânea em Português Europeu (doravante identificado como PE) e comparar os valores obtidos na medição de parâmetros acústicos do PE com os resultados conhecidos para outras línguas e para a emoção simulada. A expressão espontânea de emoções gravadas em contexto de rua, ou retiradas de programas televisivos, poderá trazer aspectos bastante relevantes, contudo ainda aqui existe a problemática relacionada com o facto de determinar a verdadeira emoção do indivíduo (o que muitas vezes só conseguimos identificar através do contexto) (Patrick et. al 2008). Procuramos a obtenção de dados que permitam caracterizar e descrever a expressão de emoção espontânea no PE. O nosso foco de estudo centra-se em parâmetros como F0, *jitter*; *shimmer* e HNR que serão passíveis de comparar com estudos realizados previamente, por nós, para emoções produzidas por um actor. Utilizando o sistema *Feeltrace* é possível descrever a sensação que cada enunciado (em qualquer das emoções em análise) provoca no receptor e descrever de forma mais dinâmica a emoção. Face a escassez de estudos sobre qualidade de voz em Português Europeu, é importante referir que este trabalho apresenta análises com *corpora* formulados, seleccionados e/ou recolhidos para este estudo especificamente.

É nesta conjuntura, reconhecendo todas as limitações da presente pesquisa, que se compreende esta análise como uma primeira abordagem, um trabalho pioneiro, com muitos aspectos que podem e devem vir a ser evoluídos, podendo vir a ser importantes em áreas como a da robótica e tecnologia tanto em termos de saúde (em gerontologia, por exemplo) como em termos lúdicos (jogos de computador).

## 2. Formas de descrição de emoções

A grande dificuldade para conseguir caracterizar um determinado tipo de voz e correlacioná-lo com uma determinada emoção tem a ver com o facto de nessa expressão haver grande marca pessoal do indivíduo. Muitos estudos vêm referindo *corpora* pouco credíveis, nos quais se correlaciona apenas a elevação do nível de F0 na análise e explicação de cada tipo de voz para cada emoção. Este facto veio trazer várias réplicas dos mesmos estudos em que apenas se via esta subida de F0 como praticamente o único parâmetro possível de analisar porque era o que alterava o padrão de voz “...led some propose that only arousal is coded in voice” (Patrik et al., 2008).

No entanto, existem evidências consideráveis que mostram que as emoções estão para além de uma simples dimensão de ativação activo/passivo (*aroused/sleepy*) e da distinção agradável/desagradável, ou negativo/positivo como será denominado neste trabalho. Para além disto, muitos estudos basearam-se apenas em algumas emoções consideradas básicas, negligenciando outras mais complexas. Muitas teorias sobre emoção não fazem um estudo detalhado da sensação que a

emoção pode causar ao ouvinte. Por exemplo, podem-se esperar diferenças nas dimensões de activação e nas valências (contínuas ou discretas). Estas categorias diferenciam a qualidade vocal. A avaliação destas teorias mostra também que o discurso emocional transmite algumas nuances que refletem a avaliação cognitiva e conseqüentes tendências de acção que estão na base de cada emoção (Patrik et al., 2008).

Uma forma de definir/descrever as emoções é através de uma imagem bidimensional Activo/Passivo Vs. Negativo/Positivo. Ainda que esta definição possa variar de indivíduo para indivíduo pode tornar mais clara a descrição de uma emoção. Na figura 1 apresenta-se a descrição realizada por Scherer (2005, pp.720).

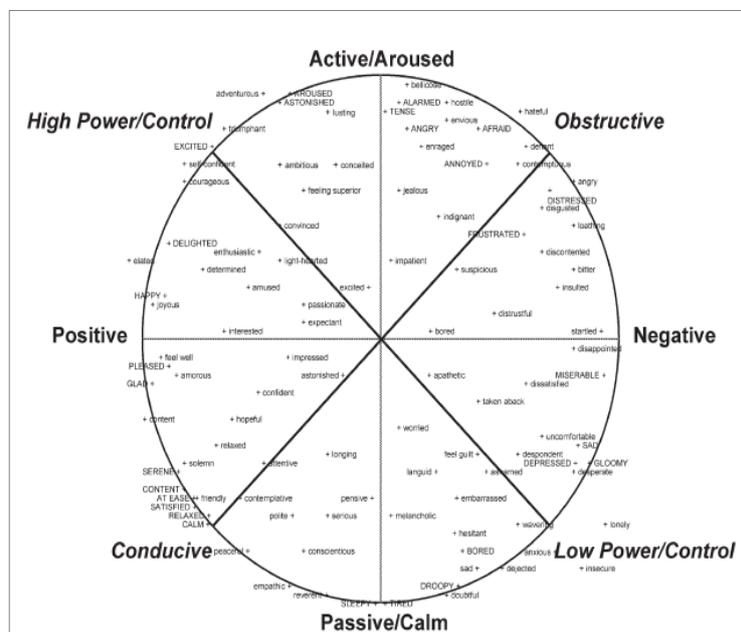


Figura 1: Descrição das emoções proposta por Scherer (2005).

No seu estudo Scherer (2005) apresenta ainda o protótipo de instrumento para classificação de emoções criado na Universidade de Genebra, chamado *Geneva Emotion Wheel* (GEW), Figura 2. No seu estudo, Scherer considera que o conceito emoção se tornou algo banalizado, porque é usado com muita frequência. No entanto, face à pergunta “o que é de facto uma emoção?” não parece haver uma resposta consistente, sendo que diferentes áreas de investigação (Humanidades; Ciências Sociais e do comportamento) raramente conseguem chegar a um acordo. A emoção é, então, um episódio que surge como resposta a um estímulo interno ou externo (Scherer, 2005). Scherer procura uma forma de analisar essas respostas, essas emoções. Na sua primeira versão, o Feeltrace era um instrumento que permitia identificar todas as famílias de emoções através de um parâmetro específico, que se tornava visível através do movimento do rato dentro do círculo. No entanto, esta primeira experiência mostrou

que era difícil medir a intensidade de uma emoção. Assim, uma versão mais recente apresenta já medidas diferentes. Estas não têm tanto a ver com a família de emoções mas com o grau de distância ou aproximação a que se encontram do estado neutro. O GEW parece ser, neste momento, o primeiro instrumento a apresentar uma verdadeira amostragem da qualidade das emoções, num espaço bidimensional (apreço: positivo ou negativo e intensidade: distância do estado neutro).

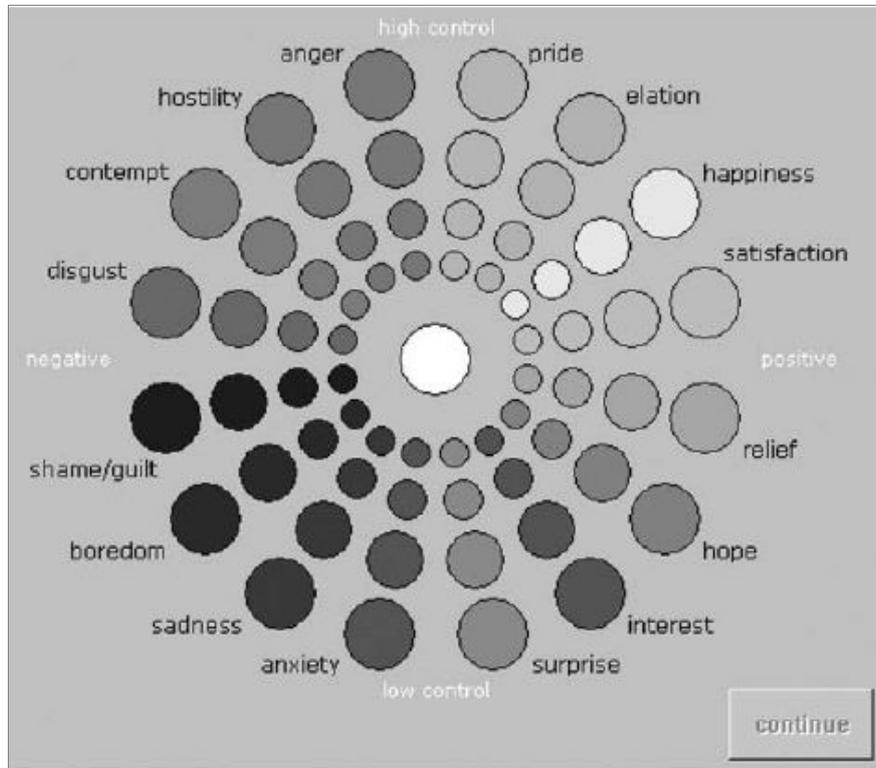


Figura 2: Círculo bidimensional de amostragem da qualidade das emoções, (GEW, Scherer, 2005)

## 2.1. Trabalhos relacionados

A maior parte dos estudos existentes versa, sobretudo, sobre emoções simuladas, induzidas em contexto de laboratório e/ou produzidas por actores profissionais. Os conhecimentos que se têm hoje em dia sobre a expressão de emoção são baseados neste tipo de *corpora*, ainda complementados, muitas vezes, pela expressão corporal e facial associada a cada tipo de emoção (Cowie et al, 2003). Esta carência científica em *corpora* espontâneos levanta um problema na interacção homem/máquina que se pretende que seja cada vez mais desenvolvida. Daí que seja importante reconhecer até que ponto a simulação se aproxima da espontaneidade. As experiências realizadas em laboratório, com manipulação de expressões de afecto, em que se tenta o mais possível a aproximação à emoção espontânea são de certa forma válidas, mas esta indução de emoções pode trazer resultados pouco válidos ou pouco específicos, assim a importância da análise da emoção espontânea para confronto e validação de dados (Patrick et al, 2008). Também o constante desenvolvimento tecnológico em que se

procura uma cada vez melhor interação homem e máquina, levanta a necessidade premente de estudos no campo da emoção espontânea (Pantic e Rothkrantz, 2003, citados por Kostoulas et al., 2008).

As propriedades acústicas das emoções espontâneas em Coreano, foram analisadas tendo em conta os parâmetros relacionados com F0 (mediana, máximo, mínimo e média) *jitter*; *shimmer* e duração (Chung, 2000). O autor apontou como principais resultados da análise acústica o facto de a expressão de alegria levar a um aumento da mediana de F0, enquanto que a tristeza conduz a uma descida deste mesmo valor e ainda do F0 mínimo. Neste aspecto, considera que as medidas de F0 são um bom indicador da intensidade de uma emoção (seja ela positiva ou negativa). As análises de *jitter* e *shimmer* parecem aumentar de acordo com a tensão emocional. No entanto, estas variações, de acordo com os resultados obtidos por Chung (2000) não foram significantes, em termos estatísticos para o caso do Coreano.

A análise do discurso espontâneo, tendo como base os dados da “*Belfast Naturalistic Emotional Database*”, foi realizada Cowie et al. (2003). Esta base de dados contem 298 gravações audiovisuais para as quais usaram 125 informantes. Mais uma vez, neste estudo, se usa como apoio e suporte a expressão corporal e facial. O *corpus* utilizado apresentava o estudo de diversas características comunicativas (físicas/paralinguísticas) e de voz; *pitch*; intensidade; tempo; articulação e qualidade de voz. Neste estudo foi usado o *Feeltrace*, permitindo que os ouvintes descrevessem a emoção que percepcionavam. Observações preliminares sugeriram que os parâmetros de descrição permitiam uma boa diferenciação entre emoções e o estado neutro, importante foi também perceber que existem alguns indicadores vocais que não são percepcionados quando se estuda emoção induzida.

As emoções negativas passivas são as que se caracterizam por haver uma quebra na voz, uma redução de intensidade e ritmo. A alegria intensa, por exemplo, também pode ter momentos de silêncio, tal como a tristeza. A raiva (activa) apresenta um aumento do *pitch* e volume. Com este estudo percebem-se melhor as lacunas que podem existir num estudo de voz induzida, mostrando que há outros parâmetros que poderão ser analisados: repetições; erros e má articulação, diferenças de intensidade e que ocorrem frequentemente em emoções espontâneas negativas activas.

O estudo realizado por Kostoulas et al. (2008) visou a criação de um *corpus*, em Grego, em que a preocupação era fazer com que as emoções fossem espontâneas e não simuladas, uma vez que o resultado seria para vir a ser utilizado no projecto de reconhecimento “*smart-home*” (casas inteligentes). Os informantes foram divididos por idade e sexo e escolhidos entre estudantes e empregados da Universidade de Patras, Grécia.

Nesta análise tiveram em conta duas categorias de emoções: negativo vs positivo, eram considerados negativos os enunciados manifestando confusão e raiva; Positivos os que englobavam alegria, prazer e estado neutro. Apesar de todos os esforços os resultados que obtiveram mostraram

que é muito difícil trabalhar e construir um *corpus* baseado em emoção espontânea, corroborando resultados apresentados por Batliner et al (2007). Consideram que o estudo que realizaram e os seus resultados podem ser o começo para a criação de uma base de dados e para melhoria em termos de metodologias a adoptar, mas que tem muito que ser desenvolvida e melhorada.

### 3. Metodologia

Criámos um pequeno *corpus* constituído por excertos de programas televisivos e do qual faziam parte vários enunciados, que depois seleccionámos e tratámos (de forma a que ficassem descontextualizados). Procedemos à gravação, para DVD, de programas televisivos que tivessem discurso directo e não preparado, seguidamente, realizou-se a extracção do sinal de voz para que pudéssemos trabalhar todo o material no programa Praat.

Foi necessário verificar todo o material e retirar os que apresentavam maiores problemas em termos de gravação (ruído) e que iriam dificultar a anotação. De um total de seis horas de gravação seleccionaram-se e analisaram-se 20 gravações diferentes (pequenos excertos, descontextualizados) que foram etiquetadas pela autora de acordo com o que era percebido (até porque existia o acesso a todo o contexto, incluindo expressão facial e corporal). Estes 20 enunciados incluíam, para além do discurso neutro, tristeza, raiva e alegria.

Tendo em conta a selecção que efectuámos, fizemos seguidamente uma etiquetagem de acordo com a emoção que era percebida, utilizando para isso o sistema *Feeltrace* complementado pela atribuição de um conjunto de etiquetas relativas à emoção, intensidade da emoção e grau de confiança do ouvinte nessa atribuição.

A análise acústica teve em conta, os parâmetros F0; *shimmer*; *jitter* e HNR, sendo que para a análise dos valores de F0, e porque estes são dependentes do género do indivíduo apresentamos os valores obtidos para as mulheres e homens. O nosso estudo recai na análise destes parâmetros tendo em conta o que é normalmente analisado no estudo de voz e qualidade de voz.

#### 3.1. Escolha dos parâmetros analisados (base propedêutica)

Como qualquer descrição fonética, a descrição do que se entende por voz também precisa de ser apropriada e transparente nos seus diferentes níveis. Laver (1991) caracterizava a qualidade de voz distinguindo aspectos impressionistas e fonéticos. Os aspectos impressionistas estão ligados à percepção e audição da voz, o ouvinte é que descreve, de acordo com o que ouve, os diferentes tipo de voz. No aspecto fonético englobam-se todos os aspectos relacionados com a sua produção, anatomia e fisiologia. Foi só no final do século XX, início do século XXI, que existiram novas e avançadas pesquisas e métodos para análise de qualidade de voz. Baseados no algoritmo proposto por Alku e

Vilkman (1996) para trabalhar em discurso fluente, é agora possível estudar as várias qualidades de voz, como por exemplo a medida de aspiração sempre presente na voz (Campbell e Mokhtari, 2003). Tanto a voz aspirada como a voz mais tensa são utilizadas de forma a transmitir diferentes informações de acordo com o contexto em que nos encontramos (seja social, seja da relação com o interlocutor).

A qualidade de voz está dependente de aspectos da fala como: formato da cavidade vocal, *pitch*, formantes, intensidade, nasalidade e ajustes/características de voz. O formato da cavidade glotal é por si um dos elementos mais importantes da fala, na maneira como pode afectar a percepção da qualidade de voz. Assim, qualidade de voz deve ser medida e analisada de uma forma contínua, desde uma voz tensa a uma voz aspirada, utilizando para isso o Quociente de Amplitude Normalizado (*Normalized Amplitude Quotient*), designado por NAQ, funcionando como independente das medidas de frequência fundamental, para assim se obterem medições relacionadas com itens paralinguísticos. O NAQ varia de acordo com a forma de falar e com o próprio indivíduo (característica paralinguística). Assim, Campbell e Mokhtari (2003) chegam a considerar que a qualidade de voz deveria ser considerada o quarto parâmetro da prosódia. No seu estudo verificaram que a nossa voz varia consoante o nosso grau de parentesco, preocupação ou afecto para com a pessoa a quem nos dirigimos. No entanto, estas características que têm algum relevo prosódico não podem ser separadas de dimensões como frequência fundamental, intensidade, duração e amplitude.

Para o nosso estudo a questão mais pertinente é entender quais os parâmetros de voz que estão mais relacionados com a transmissão e facilitação da percepção da emoção. Uma das principais dimensões a ter em conta quando falamos em qualidade de voz diz respeito à energia (altura do som que produzimos), uma vez que este é um parâmetro que, aumentando, faz variar a qualidade de voz (Bele, 2007). Na sua experiência Audibert et al (2005) utilizaram um *corpus* em que tinham estímulos de emoções gravadas por actores e emoções espontâneas, expressando ansiedade, desapontamento, nojo, alegria, resignação, satisfação e tristeza. Utilizaram todos estes estímulos num teste de percepção. Os resultados vieram precisar que nenhum parâmetro por si só veicula a informação necessária sobre uma emoção. Os contornos de F0 revelaram trazer mais informação quando se trata de emoções positivas. Qualidade de voz e duração, pelo contrário, fornecem mais informação quando se trata de emoções negativas. Intensidade, por sua vez, não traz qualquer informação significativa quando utilizada de forma isolada (Audibert et al 2005).

Outras preocupações a ter em conta na descrição de qualidade de voz, ligam-se à possível existência de patologias. Dentro desta área, um vasto leque de parâmetros merece especial atenção. Para se poder descrever uma voz como rouca (*“roughness”*), o que a caracteriza é, por exemplo, a flutuação de amplitude em termos temporais, ou seja, em curtos espaços de tempo (Bele, 2007). Também as perturbações da frequência fundamental e da amplitude de vibração das pregas vocais são parâmetros que, em termos de qualidade de voz e patologia, devem ser tidos em conta.

Os níveis de ruído têm um grande efeito em termos de percepção de qualidade de voz, o que ficou conhecido como “*voice hoarseness*”, a rouquidão da voz (Bele, 2007). Durante vários anos, várias técnicas se foram desenvolvendo para medir este tipo de ruído (rouquidão na voz, um dos principais e primeiros indícios de uma possível patologia vocal), Bele (2007) cita como exemplos:

- “*The long-time averaged spectrum (LTAS)*” (Frokjaer-Jensen e Prytz, 1976; Gauffin e Sundberg, 1977e 1989), capaz de distinguir entre certos tipos de voz de acordo com as frequências;

- “*The spectral flatness of the residue signal*” (Markel e Gray – 1978), que revela a dependência da qualidade de voz do ruído que depois é representado em termos de espectrograma;

- “*Teager Energy Operator*” (Gavidia-Ceballos et al., 1996; Cairns et al., 1996), estudo onde já se conseguem analisar os níveis de ruído gerados pela voz (componente harmónica) e a turbulência (fricção e passagem do ar);

- “*Harmonic-to-noise ratio (HNR)*” (Davis, 1978), para caracterizar as vozes patológicas.

Como resultado de toda esta complexidade e multiplicidade de estudos e teorias, várias vezes acontece existirem falhas na análise de parâmetros de certas patologias. Os métodos acima mencionados ou são invasivos ou têm alguma falta de robustez. Não permitem a descrição do *Open* e do *Speed Quotient* da forma da onda que facilitaria a correlação entre vários tipos de qualidade de voz.

Todos estes componentes devem, portanto, ser utilizados de forma complementar com os parâmetros descritivos de qualidade de voz, para que se possa vir a conseguir uma cada vez melhor e mais precisa descrição de qualidade de voz.

### **3.1.1. Corpus**

O *corpus* recolhido para este estudo foi retirado de programas, em directo, dos vários canais da televisão Nacional: Serviços informativos (Notícias) em que há, muitas vezes, reportagens realizadas no exterior e os denominados *Talk Shows* em que há conversas em directo. Tentámos sempre eliminar alguns tipos de ruído e evitar entrevistas a políticos, actores e comentadores, cujo discurso é sempre mais comedido e estudado, para além de haver uma diferente colocação de voz.

Os enunciados seleccionados são, na sua globalidade, de pessoas convidadas para contar histórias das suas vidas em programas televisivos ou indivíduos entrevistados na via pública. Uma vez que os indivíduos em qualquer um destes formatos televisivos se encontram a falar de si mesmos, de factos marcantes das suas vidas, pela positiva, ou negativa, acabam por ser mais espontâneos na expressão de

sentimentos. O *corpus*, na sua globalidade é constituído por 20 enunciados, manifestando alegria, tristeza e raiva, para além do discurso neutro, com a indicação do sexo do informante.

As emoções que analisámos são as que seriam mais fáceis de obter com os meios disponíveis, estão englobadas todas as famílias de emoções e contempladas as emoções que apresentaram algumas diferenças no teste de percepção apresentado no capítulo anterior (principalmente a alegria e a tristeza).

Apesar deste *corpus* ser necessariamente limitado, traz-nos bastante informação nova e original em termos científicos para esta área. Sabemos de todas as limitações que um *corpus* de discurso espontâneo apresenta, é fácil despoletar diversos sentimentos num indivíduo, mas também reconhecemos que razões, princípios de ética e direito à privacidade de cada um nos impedem de o fazer.

Foi esta a forma mais ética de se conseguir emoção espontânea. As gravações apresentam alguns problemas em termos de ruído e qualidade de som, no entanto, não impossibilitam, na generalidade, a análise e extracção de importantes resultados.

### **3.1.1.1 Anotação fonética e extracção de parâmetros acústicos**

Todos os segmentos seleccionados foram anotados no sistema Praat usando SAMPA. Realizou-se, seguidamente, a extracção de parâmetros através do Praat, passando todos os valores necessários e identificação de tempo e de cada frase para Excel. Durante todo o processo não se anotaram segmentos em que o Praat não conseguia ter uma estimativa precisa de F0, o que nos fez, por vezes, substituir alguns enunciados.

As análises estatísticas foram posteriormente realizadas, bem como as comparações entre os vários parâmetros e valores que para eles obtemos.

## **3.2 Anotação das emoções**

Depois da análise e tratamento destes dados, realizámos um teste de anotação de emoções no sistema *Feeltrace* que permite ao ouvinte marcar a forma como percebe uma emoção, marcando-a ao longo do tempo (enquanto a ouve) tendo em conta parâmetros de anotação/marcação explícitos no sistema (positiva/negativa; activa/passiva).

Nesta secção descreveremos o processo seguido na obtenção de “etiquetas” emocionais para os segmentos seleccionados para análise, começando pela descrição breve do *Feeltrace*, usado nesse processo.

Os enunciados foram, inicialmente, etiquetados para servirem de orientação no estudo posterior, não querendo dizer que a nossa denominação fosse a correcta. Esta etiquetagem nunca foi dada a conhecer aos ouvintes que, no decorrer das nossas análises, vieram a validar a maioria das nossas escolhas.

### 3.2.1. Feeltrace

#### O *Feeltrace* (exemplo na

Figura ) é um instrumento desenvolvido para permitir que ouvintes registem a emoção que estão a perceber, e as suas mudanças ao longo do enunciado, ou seja, de uma forma dinâmica. É baseado num espaço de activação/avaliação uma representação que deriva da psicologia. A dimensão de activação mostra como uma emoção pode ser dinâmica e a avaliação a forma como ela pode ser manifestada: positiva ou negativa (Cowie et al., 2000).

O *Feeltrace* pode não ser um instrumento perfeito, há algumas distinções em que falha, como por exemplo entre raiva e medo (Cowie et al, 2000). No entanto, para muitos estados emocionais a análise através deste sistema é um importante ponto de partida. Parâmetros relacionados com a duração são certamente os mais difíceis de analisar na expressão da emoção, tendo em conta apenas a voz.

Uma das maiores dificuldades em termos de investigação, quando falamos de emoção, nomeadamente espontânea e quando a análise recai sobre amostras de discurso fluente (por oposição a vogais sustentadas), é perceber a sua gradação, a forma como varia ao longo do tempo. Nas nossas análises para a emoção espontânea, concretamente para o teste de *Feeltrace*, esse é um aspecto que merece atenção, assim torna-se mais eficiente e eficaz em termos descritivos.

Seguindo Ekman (2003) e Cowie et al. (2003), agrupamos as emoções em três grupos distintos: Negativas (raiva; tristeza); Positivas (alegria) e Neutra. Fazemos ainda, de acordo com o teste *Feeltrace* a divisão Activas (por exemplo, raiva será uma emoção negativa/activa); Passivas (por exemplo, sendo uma emoção positiva a euforia é, marcadamente, mais activa do que a alegria).

Optámos por usar, o *Feeltrace*, na sua versão original e conforme aparece na figura 2 (em Língua Inglesa) uma vez que verificámos que este nos reportaria a uma descrição mais detalhada e precisa das emoções espontâneas em análise. É um instrumento desenvolvido de forma a permitir que os ouvintes descrevam determinado estímulo de um modo contínuo (no tempo) e dinâmico. É um instrumento baseado na ideia da representação do espaço de activação e avaliação defendido pela psicologia. É de fácil utilização e ao mesmo tempo permite resultados fiáveis em termos de investigação científica, segundo os autores.

A dimensão de activação mede o grau de dinamismo de uma emoção (activa/passiva); a dimensão de avaliação permite distinguir entre um sentimento positivo ou negativo associado a um estímulo.

A investigação sugere que o espaço é naturalmente circular, ou seja, as emoções que são as mais fortes (no seu limite de intensidade) definem um círculo sendo que, por isso mesmo, o estado neutro se encontra no centro.

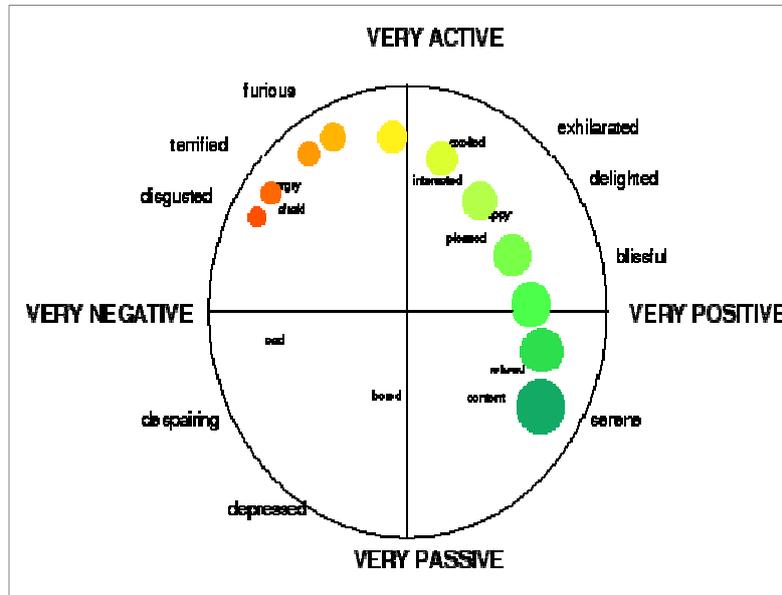


Figura 3: Exemplo do *Feeltrace* durante uma sessão de gravação. O cursor muda de cor de vermelho para laranja (à esquerda) e de amarelo para verde, de acordo com o eixo em que se encontra.

### 3.2.1.1 Procedimento

A preocupação inicial foi dar a conhecer o material com o qual os ouvintes iriam trabalhar, para que não houvesse qualquer tipo de interferência por parte de quem estava a aplicar o teste.

Assim, os ouvintes, primeiramente, visualizaram um *PowerPoint* que foi elaborado para explicar todo o funcionamento do teste e para que não se tivesse que intervir enquanto o realizavam (para que se pudessem evitar qualquer tipo de influências). Seguidamente liam a ficha de apoio e de acompanhamento do teste, onde tinham que marcar a emoção que haviam identificado e indicar o grau de certeza com que tinham marcado cada emoção (numa escala de 0 a 5) e qual era a expressão mais forte dessa emoção que tinham de ordenar de 1 a 3, explicitando ainda o grau de intensidade de cada um como Forte; Médio ou Baixo (Cf. Ficha/Questionário e Guião em anexo – Anexo 1).

Realizavam depois, utilizando auscultadores, um treino, onde podiam primeiro ver expressão facial e ouvir o que o indivíduo dizia; ouviam uma música, tendo que acompanhar a emoções que ao longo desta iam mudando; uma parte do teste incluía ainda uma frase do nosso *corpus* anterior (emoção por actores) para que eles a marcassem no *Feeltrace*. Só depois de todos estes procedimentos,

realizavam realmente o teste. Usaram sempre os auscultadores (que permitia também uma melhor insonorização), podiam ouvir as frases até um máximo de três vezes, fazendo as suas marcações com o cursor do programa e, seguidamente, as anotações na ficha de acompanhamento.

O processo demorou cerca de uma hora para cada ouvinte, o que deixa entender que apesar de ser um *corpus* pequeno, o trabalho na sua origem, dependente de outras pessoas para realizarem os testes, é bastante moroso.

### 3.2.1.2 Sujeitos

Neste processo participaram 8 ouvintes, 4 do sexo feminino e 4 do sexo masculino. A preocupação na selecção dos ouvintes esteve, essencialmente, relacionada com os trabalhos que realizam e suas áreas de investigação. Três dos ouvintes são professores Universitários, estando os restantes investigadores a realizar doutoramento, todos eles na Universidade de Aveiro. Todos eles estão ligados a trabalhos relacionados com voz, processamento de voz, sinal, robótica e Linguística.

### 3.2.1.3 Resultados das emoções atribuídas pelos ouvintes

Os participantes tinham, como já referimos, de preencher uma ficha de acompanhamento com a emoção e grau de certeza (0 a 5) com que tinham marcado determinada emoção. Na Figura 4, podemos observar o grau de confiança dos ouvintes na marcação de emoções.

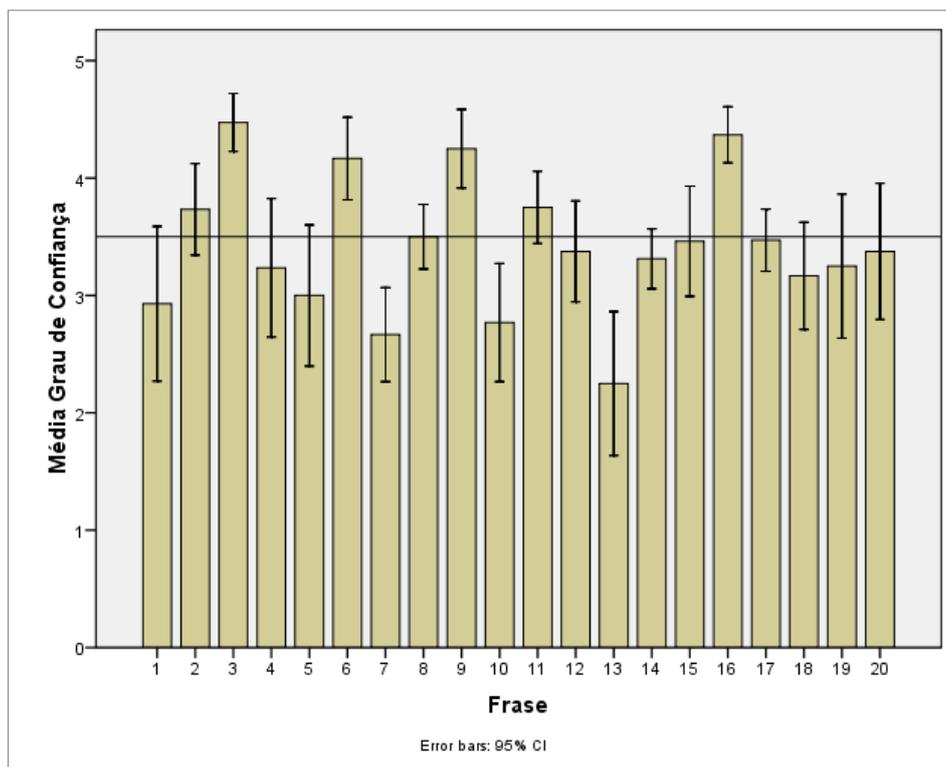


Figura 4: Grau de confiança na marcação da emoção ouvida.

Em termos médios os valores de confiança para os 20 segmentos são perfeitamente aceitáveis. Podemos destacar que os que mereceram um maior grau de confiança foram os enunciados 3, 6, 9, e 16, todos eles referentes a enunciados em que a emoção expressa era a alegria. Estas gravações são marcadamente de euforia, pois dizem respeito a futebol, daí a certeza com que são marcadas.

Seguem-se ainda com um bom nível de certeza os enunciados 2 e 11 curiosamente o primeiro dizendo respeito à expressão de tristeza e o segundo a alegria.

Os enunciados que mais dúvidas suscitaram, daí o menor grau de confiança foram o 7, o 10 e o 13, curiosamente todas elas referentes ao estado neutro.

No gráfico da **Figura 5** faz-se a análise da confiança por ouvinte.

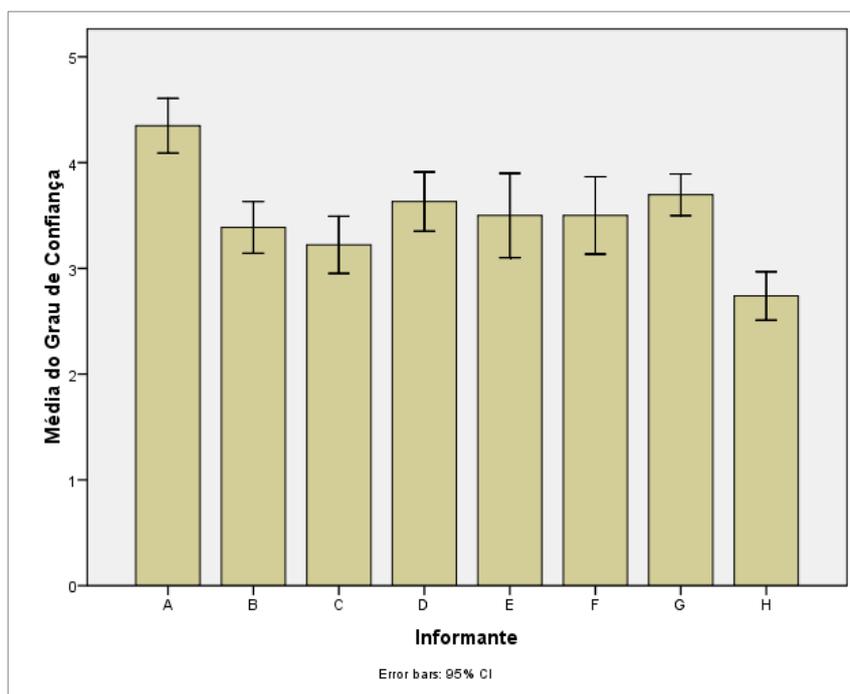


Figura 5: Média do grau de confiança na marcação das emoções ouvidas, por indivíduo.

Observando o gráfico percebe-se claramente uma conformidade entre os oito ouvintes. Embora nenhum tenha uma média de confiança que chegue aos cinco pontos, andam, na sua generalidade, todos muito próximos do 3/3,5. O que mostra uma certa uniformidade na forma como os enunciados eram ouvidos e entendidos.

Na **Figura 6** é apresentada a média do grau de confiança com que os ouvintes marcaram as suas primeiras escolhas em relação ao enunciado que tinham acabado de escutar.

Percebe-se assim claramente e corroborando o que vimos anteriormente que a maior percentagem se centra na marcação de euforia, seguindo-se confiante; bem-disposto e alegre, se atentarmos nestes dados todas as emoções são da mesma família (positivas).

Em termos intermédios aparecem emoções ligadas à família das negativas (preocupado; triste; desapontado). Neutro aparece também a um nível intermédio e se observarmos o estado relaxado (que se pode ligar à família das emoções neutras) foi o que menor grau de confiança suscitou.

Na generalidade há uma relação entre o grau de confiança e a forma como perceberam (e etiquetaram) a emoção.

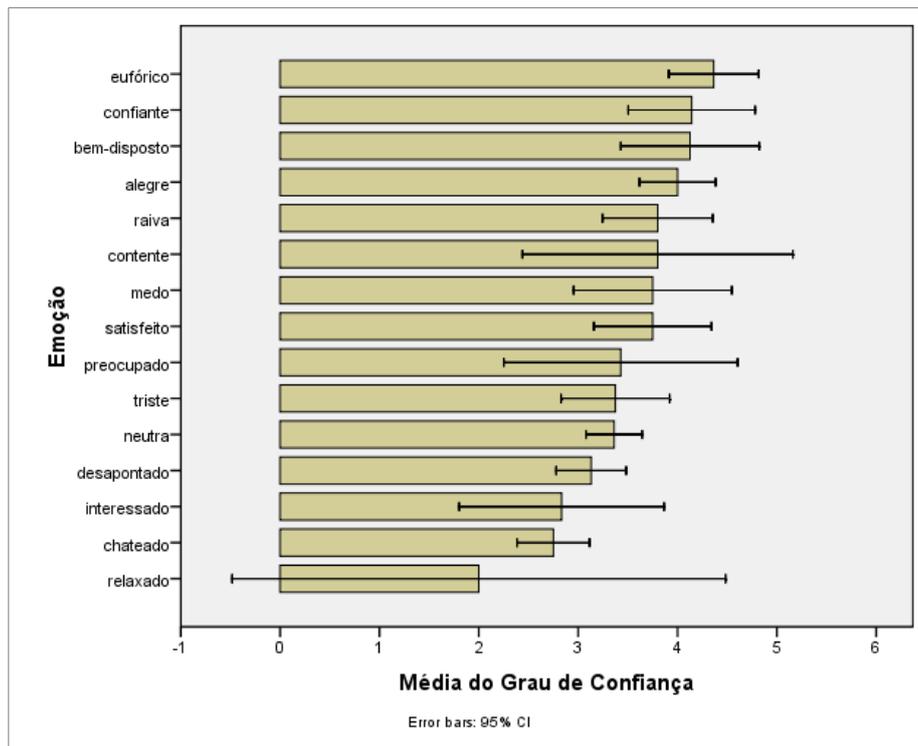


Figura 6: Média do grau de confiança, por emoção marcada como primeira.

Na **Figura 7** encontra-se informação relativa à intensidade com que os ouvintes marcaram as suas primeiras opções. Como se explicou anteriormente podiam escolher três emoções diferentes numerando-as de 1 a 3, com o grau de intensidade que achavam que se fazia sentir. O gráfico seguinte resulta precisamente desses valores e da intensidade que lhes faziam corresponder: Forte; Médio e Baixo.

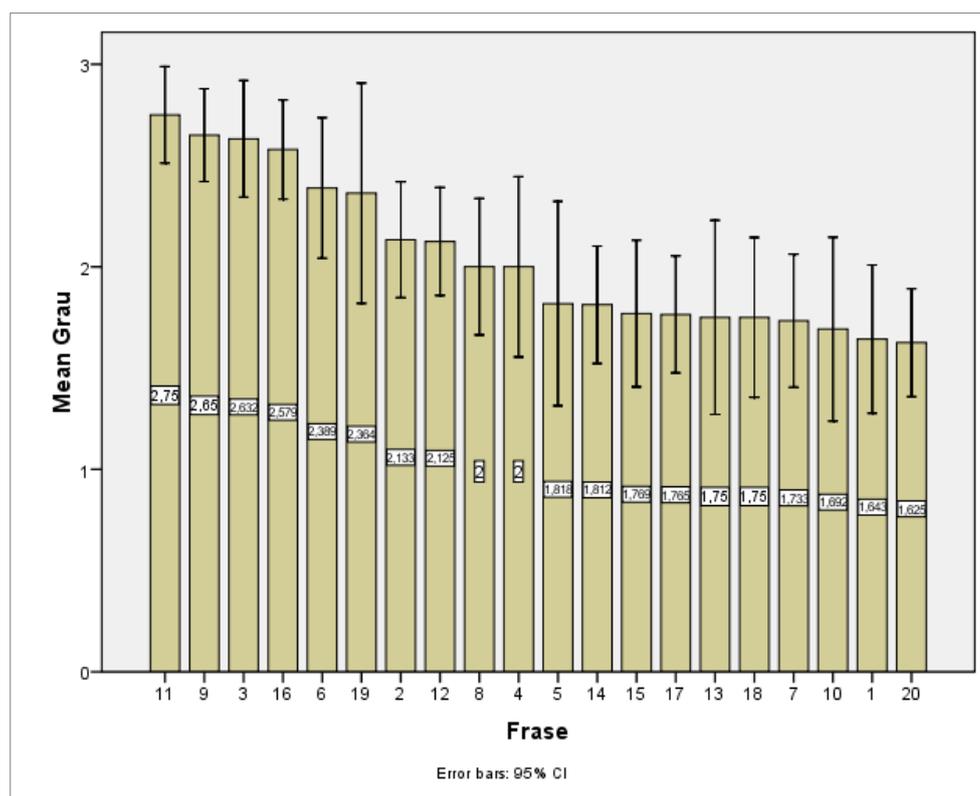


Figura 7: Grau de intensidade atribuído à primeira escolha

### 3.2.1.4 Etiquetagem

O gráfico da **Figura 8**, mostra-nos a diversidade de escolhas. Há grandes discrepâncias, principalmente em termos da expressão neutra que foi a maioria das vezes a primeira escolha seguindo-se o estado desapontado; chateado; triste (estas três últimas da mesma família). Só depois aparecem como primeiras escolhas as emoções positivas como alegre; bem-disposto e eufórico.

As segundas escolhas recaem na sua maioria em emoções negativas (chateado; preocupado; desapontado e triste).

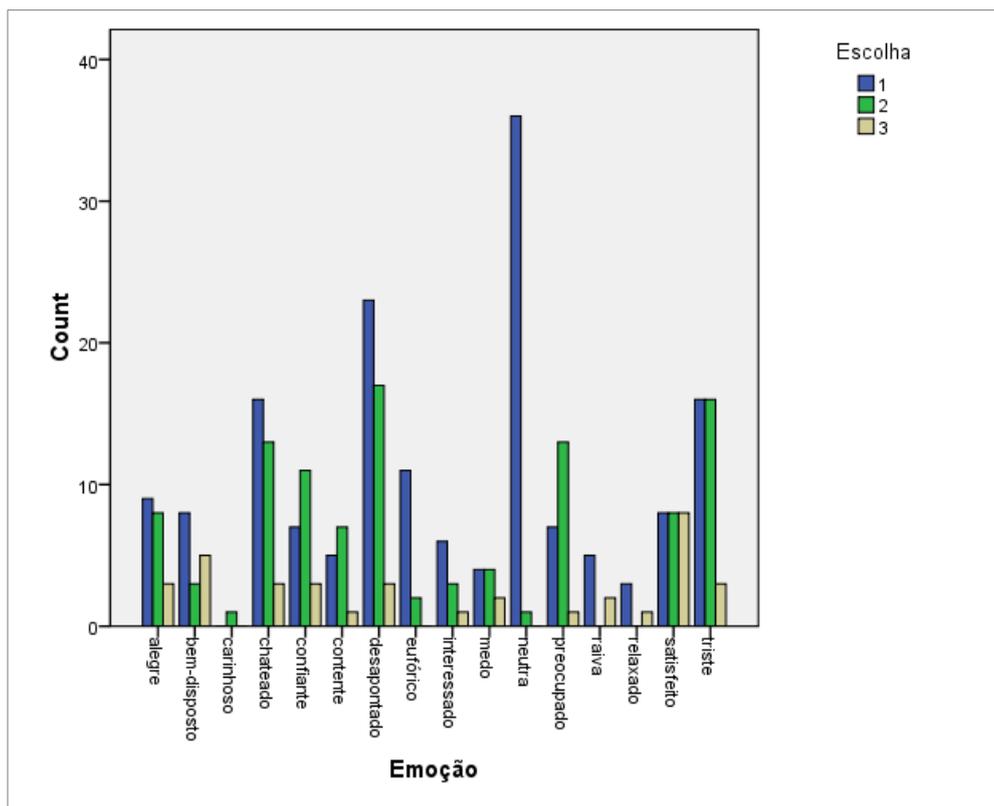


Figura 8: Diversidade das escolhas para descrever a emoção ouvida.

Apresentamos, a título de exemplo, na **Figura 9**, a forma como foi etiquetado o primeiro enunciado. A nossa etiquetagem prévia fazia referência a uma emoção negativa nomeadamente a tristeza.

Se atentarmos nos resultados apresentados na figura percebemos que estes corroboram a nossa escolha. As emoções triste e chateado merecem destaque, seguindo-se desapontado; medo; preocupado, todas elas emoções negativas. É curioso constatar a presença de uma primeira escolha referindo-se a esta emoção como neutra e uma outra, no pólo oposto (emoção positiva), como interessado. No entanto, o resultado global mostra que a emoção era de carácter negativo.

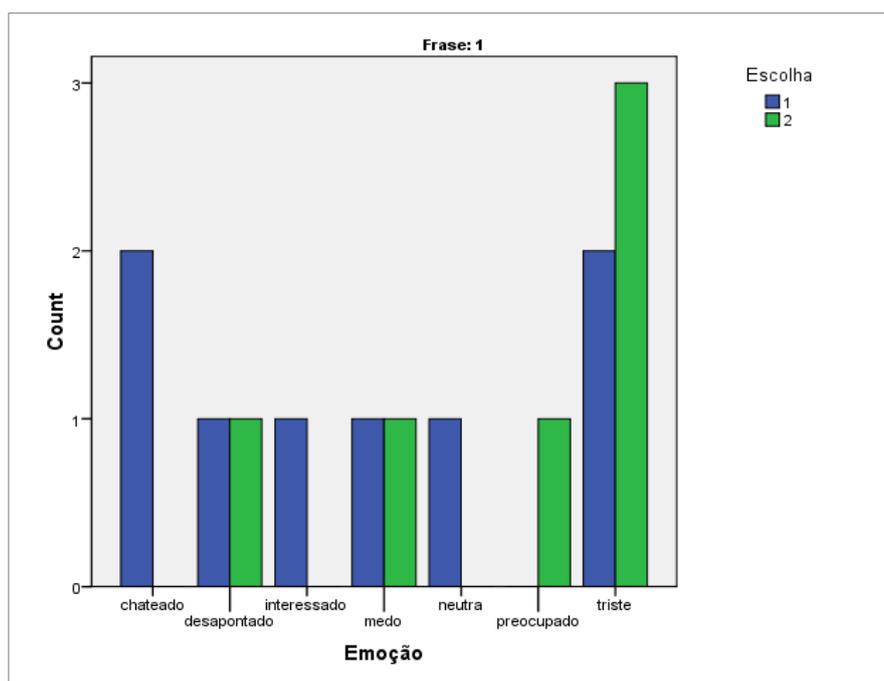


Figura 9: Etiquetação realizada para o primeiro enunciado. Família da emoção da primeira escolha: Negativa.

A **Tabela 1** apresenta-nos um resumo de todos os resultados, em termos de frase; a emoção mais escolhida; o número de vezes que foi escolhida, família de emoções; média de confiança com que marcaram, grau de intensidade da primeira escolha e outras escolhas.

Tabela 1: Resumo da marcação das emoções em todas as frases e respectivo grau de confiança.

<i>Frase</i>	<i>Mais votado escolha 1</i>	<i>Número de votos</i>	<i>Mais votado (várias escolhas)</i>	<i>Família</i>	<i>Média confiança</i>	<i>Grau de intensidade da primeira escolha</i>
1	Chateado Triste	2 2	Triste	Negativa	2,93	1,64
2	Chateado raiva	3 3	Desapontado	Negativa	3,73	2,13
3	Eufórico	5	Eufórico	Positiva	4,47	2,63
4	Triste	3	Triste	Negativa	3,24	2,00
5	Neutra	6	Neutra	Neutra	3,00	1,81
6	Contente	3	Satisfeito	Positiva	4,17	2,38
7	Neutra triste	3 3	Neutra	Neutra	2,67	1,73
8	Desapontado	5	Desapontado	Negativa	3,50	2,00
9	Eufórico	5	Confiante	Positiva	4,25	2,65

10	Neutra	3	Desapontado	Neutra	2,77	1,69
11	Alegre bem-disposto interessado	2 2 2	Bem-disposto	Positiva	3,75	2,75
12	Triste	4	Desapontado	Negativa	3,38	2,12
13	Neutra	4	Neutra	Neutra	2,25	1,75
14	Chateado	4	Chateado	Negativa	3,31	1,81
15	Neutra	6	Neutra	Neutra	3,46	1,76
16	Alegre	4	Alegre	Positiva	4,37	2,57
17	Chateado	3	Chateado	Negativa	3,47	1,76
18	Neutra	2	Chateado	Neutra	3,17	1,75
19	Neutra	7	Neutra	Neutra	3,25	2,36
20	Desapontado	6	Desapontado	Negativa	3,37	1,62

O estudo permitiu-nos verificar que é possível ao ouvinte diferenciar as emoções pelo menos em termos de família (se são negativas; positivas ou neutras). A alegria é a emoção que mais se destaca, e percebe-se que entre as emoções negativas e neutra existe por vezes alguma confusão.

O grau de confiança com que as emoções foram marcadas no teste e depois na ficha de acompanhamento vai de encontro à nossa etiquetagem, o que quer dizer que, de facto, pela voz (mesmo sem grande contexto) conseguimos perceber emoções.

### 3.2.1.5 Resultados Feeltrace

A **Figura 10** e a seguinte dizem respeito ao que foi obtido pelo *Feeltrace*. Podemos observar, pelas diferentes cores (que correspondem aos diferentes ouvintes) as marcações efectuadas pelos sujeitos enquanto ouviam o enunciado.

Na **Figura 10** temos a representação das marcações efectuadas para o estímulo 3, que se tratava de um enunciado exprimindo euforia. Percebe-se claramente que todos os ouvintes estiveram de acordo na marcação desta emoção. O ponto de cruzamento das duas linhas a tracejado indica o valor médio dessas marcações, que se tivermos em conta o gráfico bidimensional do *Feeltrace* corresponde exactamente à expressão de alegria.

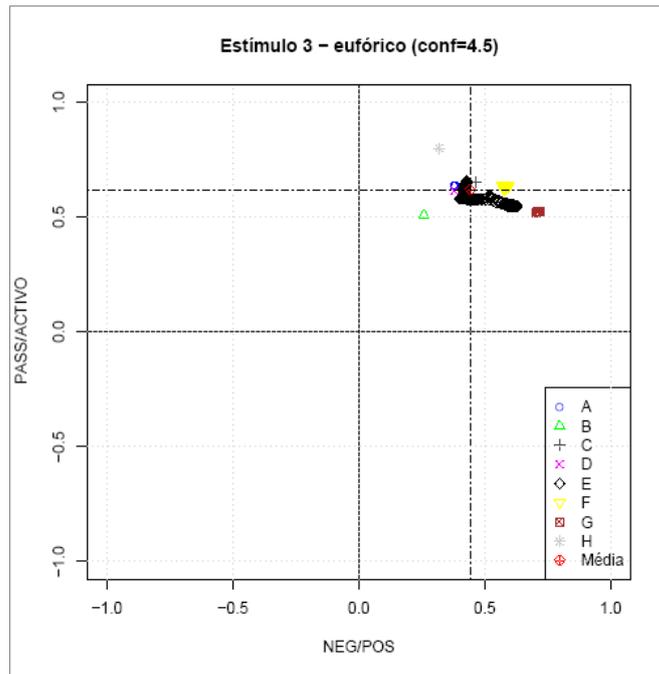


Figura 10: Exemplo de resultados *Feeltrace*. Anotações dos vários ouvintes e média das mesmas.

Neste gráfico não há uma grande diversidade ou variação nas marcações efectuadas pelos ouvintes, o que, certamente, está ligado ao elevado grau de confiança com que anotaram este estímulo.

Contudo, esta confiança, e como já pudemos verificar por gráficos anteriores nem sempre se mantém e há enunciados que geram mais divergência e mais dúvidas. O gráfico apresentado na **Figura 11** demonstra, exactamente, essa variação.

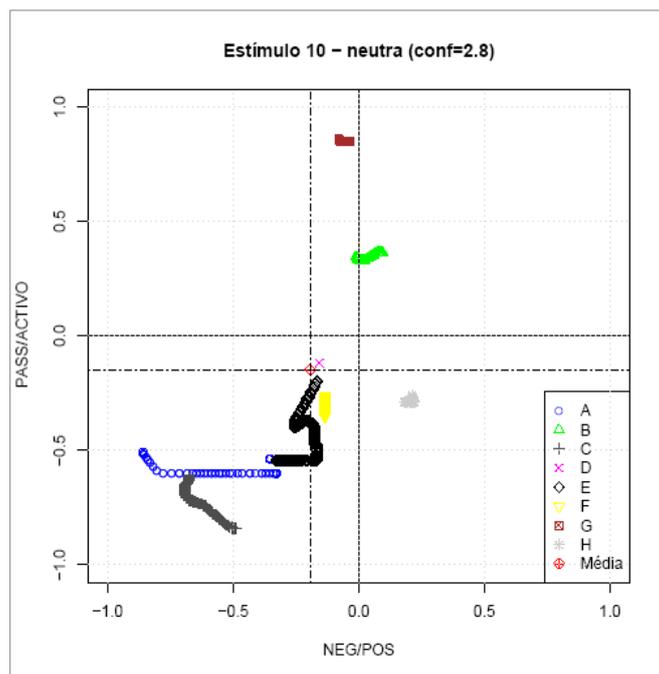


Figura 11: Exemplo de resultados *Feeltrace*, em que existe muita variabilidade. Anotações dos vários ouvintes e média das mesmas.

Para o enunciado neutro houve dificuldade na anotação quer em termos de activação, quer em termos de se perceber como emoção mais negativa ou mais positiva. Apesar de ser mais marcada como negativa/passiva não deixa, mesmo dentro destes parâmetros, de haver algumas dúvidas. Aparecem ainda resultados nos eixos activo/positivo e passivo/positivo.

Verificamos, pela análise dos gráficos, que os ouvintes conseguiram, regra geral, identificar correctamente o eixo em que se encontravam os estímulos que lhes foram dados a ouvir, em termos de positivo ou negativo. A maior dificuldade surge na distinção entre activo/passivo.

Salientamos que a maior convergência na classificação dos estímulos que se reportam a alegria deve-se ao facto destas gravações serem, na sua maioria, entrevistas de rua cujo tema era a vitória de uma conhecida equipa do futebol Nacional. Também pela análise das respostas a este teste se percebe uma confusão entre enunciados neutros e os que expressam tristeza.

## 4. Resultados

A nossa análise dos resultados será feita em duas partes. Primeiro a análise usará as anotações das emoções resultantes da emoção mais votada; numa segunda parte usaremos as médias das marcações *Feeltrace* como informação acerca das emoções.

Atendendo à quantidade, diminuta, de dados e aos vários factores que podem interferir nos valores dos parâmetros e sobre os quais não temos controlo, não serão realizadas análises estatísticas aprofundadas.

### 4.1. Usando as classificações dos ouvintes

#### 4.1.1. Valores de F0

Na análise dos valores respeitantes a F0, na **Figura 12**, fazemos a diferenciação entre homens e mulheres, uma vez que este parâmetro varia significativamente de acordo com o género. As gravações que tínhamos permitiram-nos, para as mulheres, analisar todas as emoções. Contudo, para os homens, apenas pudemos ter em conta: tristeza; neutra e alegria representando, no fundo, as três famílias de emoções presentes no estudo.

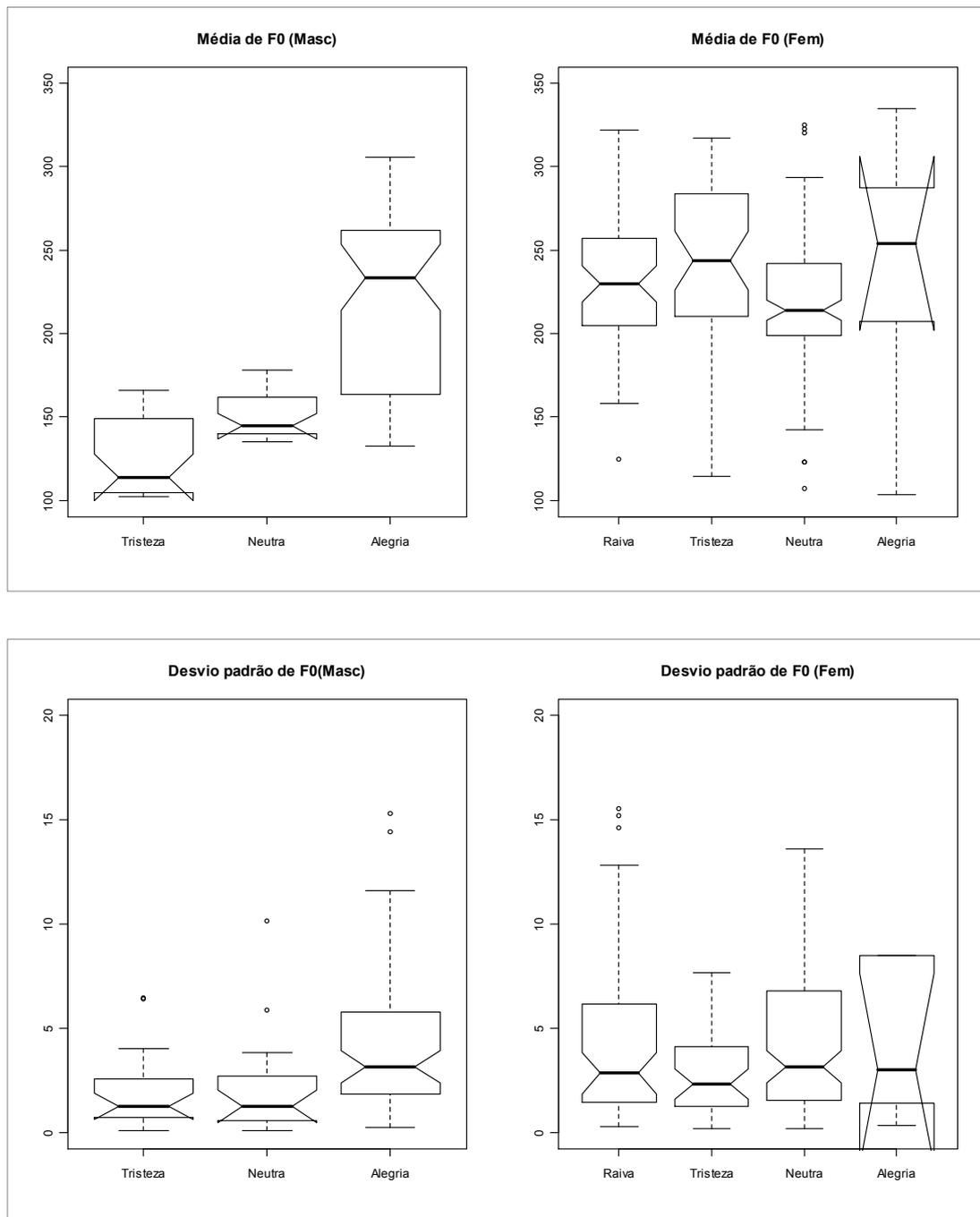


Figura 12: Valores de média de F0 e desvio padrão para os informantes masculinos e femininos, da esquerda para a direita.

Fazendo uma análise global percebe-se que quer os informantes sejam homens ou mulheres o F0 é sempre mais elevado quando se trata da manifestação de alegria.

Na média de F0 masculina percebe-se como evidente um registo elevado quando falamos da expressão de alegria, encontrando-se com um valor perto dos 230 Hz. Quanto à ausência de emoção, discurso neutro, verificamos uma média de F0 junto dos 150 Hz e, por último, com valores mais

baixos a tristeza com cerca de 120 Hz. Em termos de desvio padrão, para os homens, verifica-se claramente que este é maior nos enunciados que expressam alegria, percebendo-se que entre o discurso neutro e a tristeza não existe diferença significativa.

Na análise dos resultados para as mulheres, podemos ter já uma perspectiva mais abrangente, uma vez que temos dados que nos permitem analisar todas as emoções em estudo. Verifica-se então que a alegria e a tristeza se encontram, em termos de média de F0 para o PE, muito próximas. Perto dos 230 Hz encontra-se a tristeza, havendo uma diferença muito ténue na expressão de alegria, que se encontra perto dos 250 Hz. Aparecem-nos, assim, duas emoções de famílias completamente opostas (positiva/negativa) com resultados similares em termos de média de F0.

A raiva aparece numa posição intermédia, junto aos 225 Hz e, por último ausência de emoção com cerca de 215 Hz. Consta-se assim que, em termos de média de F0, para o sexo feminino, as diferenças entre expressão de raiva e estado neutro é muito pequena.

No que concerne aos resultados referentes ao desvio padrão é visível a proximidade de valores existente entre todos os enunciados analisados. Raiva, neutra e alegria apresentam um desvio ligeiramente menor do que o discurso neutro.

Assim, em termos gerais, juntando os dois grupos podemos dizer que as diferenças são mais acentuadas nos homens, que a alegria é a emoção que se destaca em ambos os grupos por ter valores mais elevados de F0 e que a tristeza apresenta os valores mais baixos.

#### **4.1.1.1. Jitter**

Na análise de *jitter* (**Figura 13** e seguinte) comparamos, mais directamente, cada uma das expressões com o discurso neutro. Torna-se mais fácil observar e destacar as possíveis semelhanças e/ou diferenças de uns em relação aos outros e de todos em relação ao nível médio, que será o do neutro.

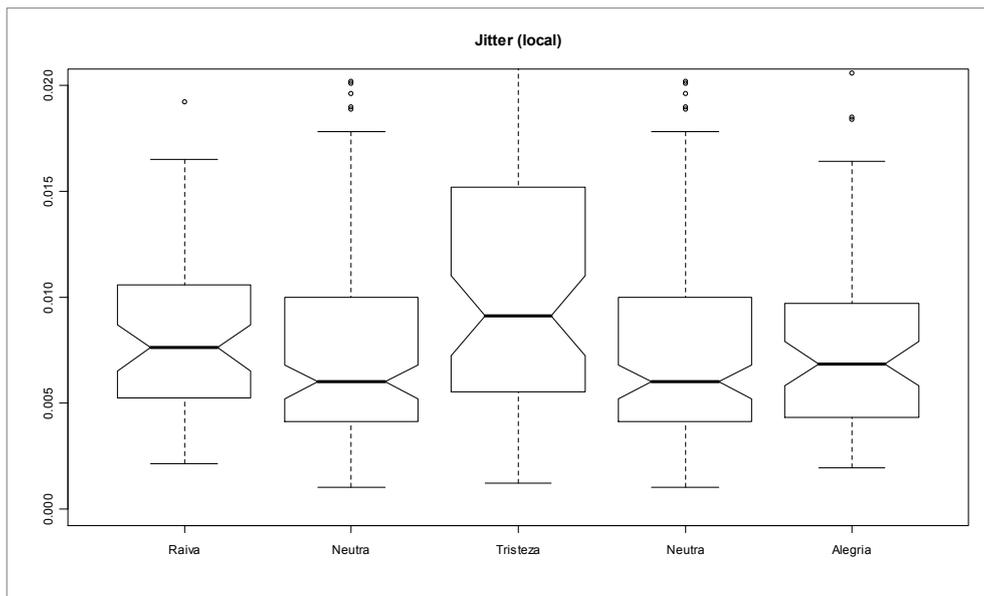


Figura 13: Valores de *jitter* local para as diferentes expressões estudadas.

O gráfico da **Figura 13**, referente ao *jitter* local, permite-nos verificar uma clara diferença nos valores da tristeza, que são mais elevados, em relação a todas as outras expressões em análise.

A alegria e a raiva, encontram-se em termos de *jitter* local em posições muito idênticas e mais próximas em relação ao discurso neutro do que em relação à tristeza. As frases neutras são, como podemos observar, as que apresentam um *jitter* mais baixo.

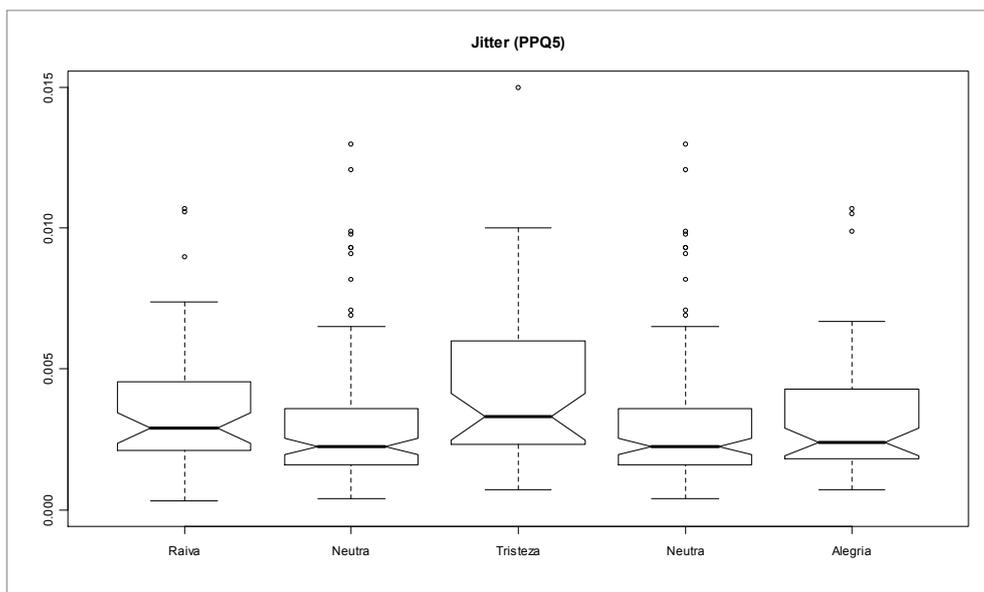
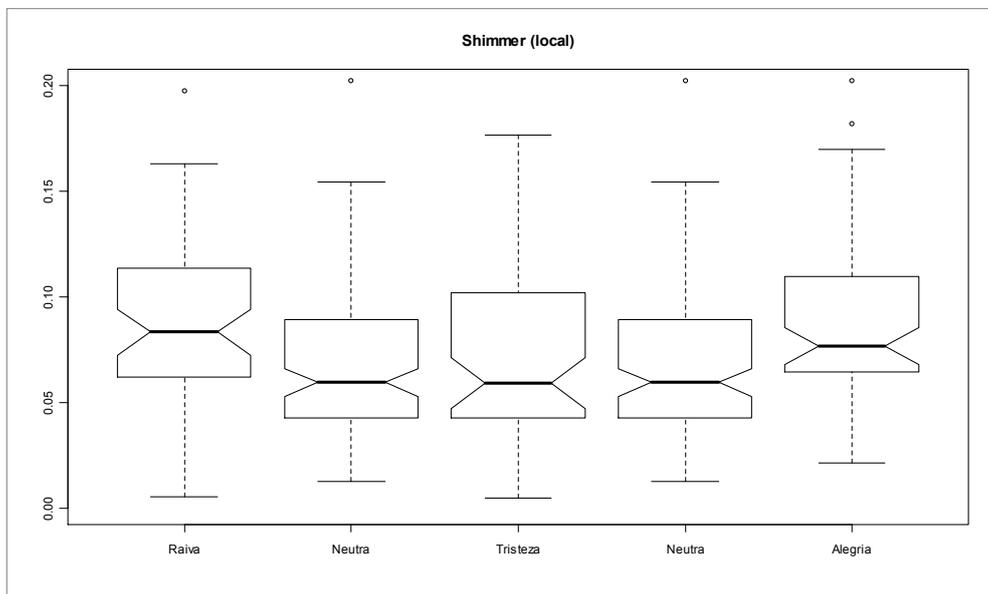


Figura 14: Valores de *jitter* PPQ5 para as diferentes expressões estudadas.

Analisando o gráfico que contempla os valores de *jitter* PPQ5 para as diferentes expressões, **Figura** , observa-se, uma vez mais, que a tristeza se destaca, apresentando o valor mais elevado para *jitter* PPQ5. Destaca-se, no entanto, que para *jitter* PPQ5, a alegria e o discurso neutro apresentam os valores mais baixos e em tudo similares, não havendo uma diferença significativa entre estas produções no que se refere a este parâmetro de análise. A raiva parece aproximar-se mais dos valores apresentados para a tristeza do que para neutra. Esta proximidade advém do facto de estarmos perante expressão de raiva contida (*cold anger*), uma vez que as gravações analisadas foram retiradas de um programa de televisivo, não se denotando assim valores tão elevados em nenhum dos parâmetros em análise. Esta contenção, em termos de expressão da raiva, é defendida por vários autores como uma atitude consciente do indivíduo que procura seguir padrões socialmente estipulados para assim ser aceite: “Anger regulates interpersonal behavior and comes to be regulated in na interpersonal context via socialization (...). Problems in regulating and appropriately expressing anger have implications for social functioning...”, (Lewis et al., 2008).

#### 4.1.1.2. Shimmer

Na análise detalhada dos valores apresentados para *shimmer local* e APQ3, faz-se, à semelhança do estudo de *jitter*, a comparação com o valor médio apresentado para o discurso neutro (**Figura 15**).



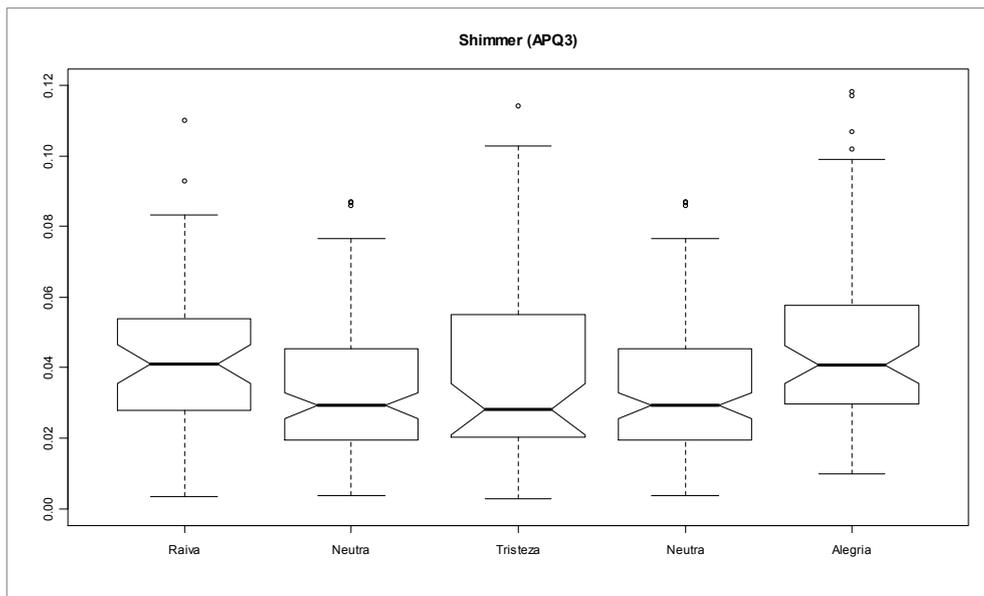


Figura 15: Valores de *Shimmer* local e APQ3 para as quatro emoções e discurso neutro.

Pela observação dos gráficos da **Figura 15**, constatamos que os valores do parâmetro são maiores na expressão de raiva e de alegria, curiosamente duas emoções de famílias distintas e tão contrárias. Este facto talvez reflecta a intensidade com que cada uma delas é proferida.

A expressão de tristeza, por sua vez, apresenta os valores mais baixos, menores ainda do que os do estado neutro.

#### 4.1.1.3. HNR

Os valores obtidos para HNR encontram-se na **Figura 16**.

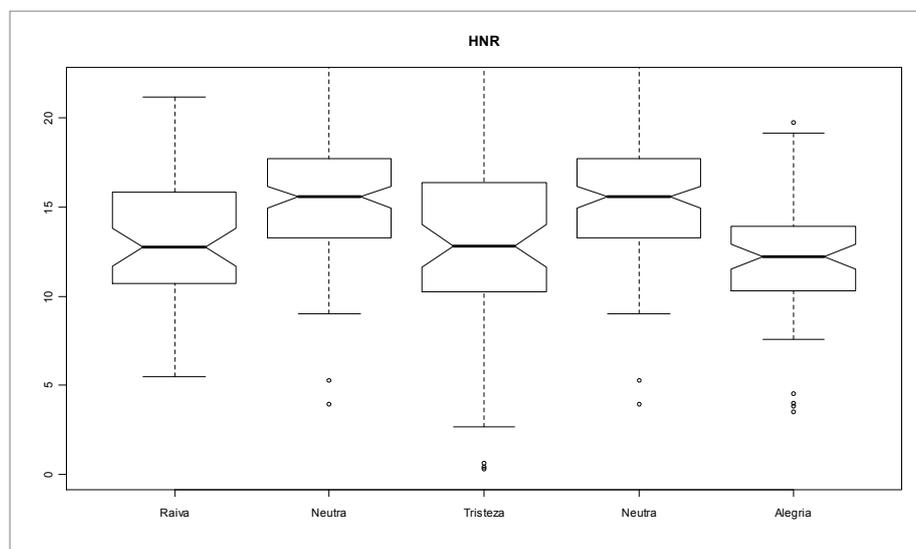


Figura 16: Valores de HNR para os cinco enunciados estudados.

Uma vez mais tendo sempre como ponto de referência, para comparação, o discurso neutro verificamos que as 3 emoções analisadas se encontram, todas elas, abaixo dos valores designados para a normalidade.

Raiva e tristeza, apresentam valores muito próximos dos 13 dB e a expressão de alegria, por sua vez, embora se encontre muito próxima das outras emoções, é a que apresenta o valor mais próximo da referência para a normalidade 12 dB. Todas elas se encontram, todavia, no limiar da normalidade em termos de relação ruído/harmonia, sendo a que mais foge desse equilíbrio é o discurso neutro.

#### **4.1.2. Usando as marcações Feeltrace**

Nesta secção analisamos os mesmos parâmetros acústicos, utilizando agora os valores médios atribuídos pelos ouvintes nos eixos presentes no teste *Feeltrace* que realizaram: passivo/activo e negativo/positivo.

Estas marcações vão permitir verificar se a emoção percebida pelo ouvinte está, por exemplo, de acordo com a nossa etiquetagem prévia, permitem ainda identificar quais as emoções mais difíceis de marcar.

Para melhor se compreender a legenda dos gráficos que se seguem, referentes a este estudo, importa saber que cada um dos eixos do *Feeltrace* foi dividido de -1 a 1, em cinco zonas diferentes. À volta da zona do 0 foi criada uma zona de valores médios para cada um dos lados, seguida de valores mais altos (2), (0) valores em torno do zero e (-2) valores negativos. O -2 corresponde a valores inferiores a -0.6, sendo que o -1 corresponde a valores entre -0.6 e -0.2. O 0 delimita valores entre -0.2 e 0.2; o 1 engloba os valores compreendidos entre 0.2 e 0.6, logo o 2 corresponde a valores superiores a 0.6.

##### **4.1.2.1. F0**

Nos gráficos das figuras que se seguem (**Figura 17** e **Figura 18**) podemos observar os resultados para a média e para o desvio padrão de F0. Tendo em conta os diferentes valores apresentados para homens e para mulheres. Analisamos a variação dos dois parâmetros de F0 com a posição média que foi atribuída, pelos ouvintes, nos eixos horizontal e vertical do *Feeltrace*.

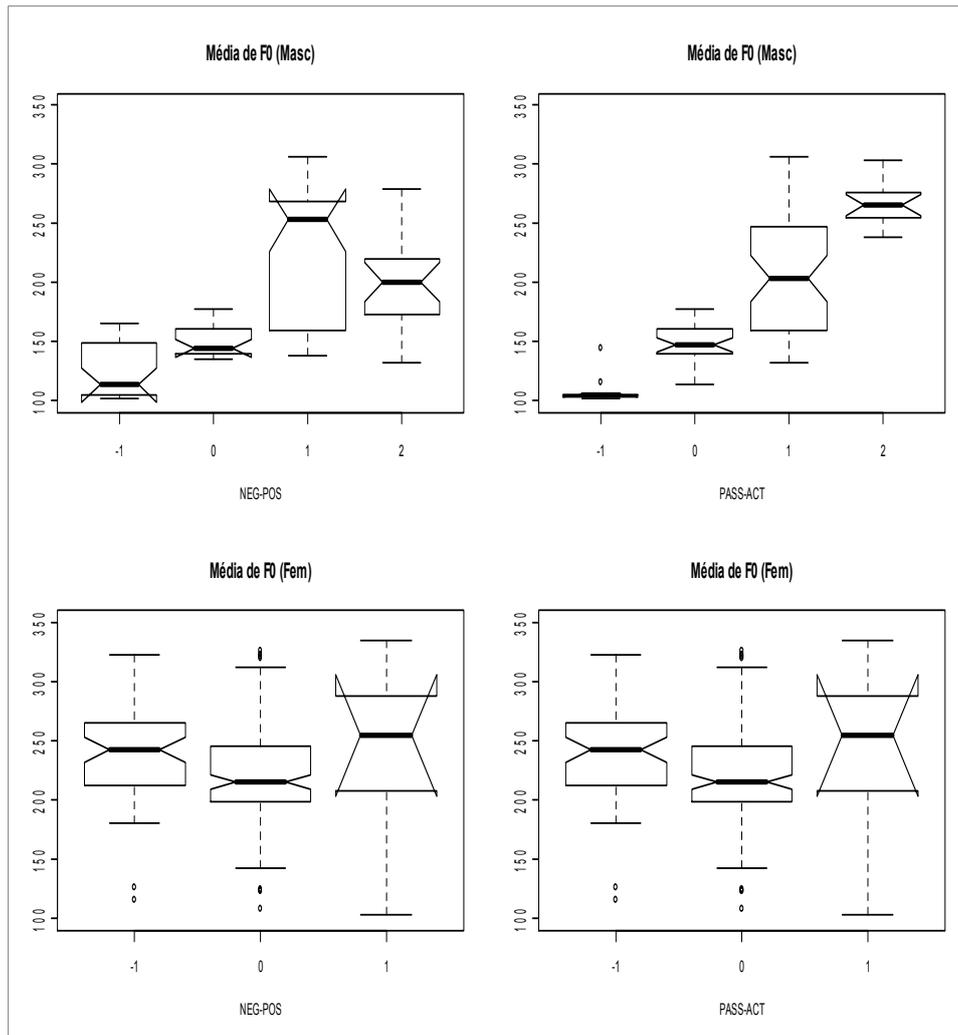


Figura 17: Valores médios de F0 em função do género e dos eixos do *Feeltrace*.

Observamos que os valores médios de F0, para as vozes masculinas, apresentam uma clara variação em ambos os eixos. No eixo passivo-activo é notório o aumento do valor médio para emoções que sejam marcadas por uma maior actividade. O facto dos intervalos de confiança (representados pelas reentrâncias triangulares) não se sobreporem indica significância estatística dessas diferenças. No eixo negativo-positivo observamos também alterações, mas não existe uma tendência clara. Estes resultados devem, contudo, ser alvo de uma cuidada interpretação devido a todos os indivíduos terem valores de F0 diferentes e ao facto de termos presentes todas as emoções para todas as pessoas o que pode interferir nos resultados. Podemos ter um aumento de F0 médio não devido à emoção mas devido ao valor médio do indivíduo ser superior. Infelizmente, por não termos outro material em emoção neutra para todos os informantes, fomos impedidos de descontar esse efeito.

No caso das vozes femininas, a maior variabilidade e valor médio associado ao género (F0 é sempre mais elevado) impede-nos de retirar grandes ilações. Também dificulta o retirar de conclusões o facto de apenas termos três posições em ambos os eixos.

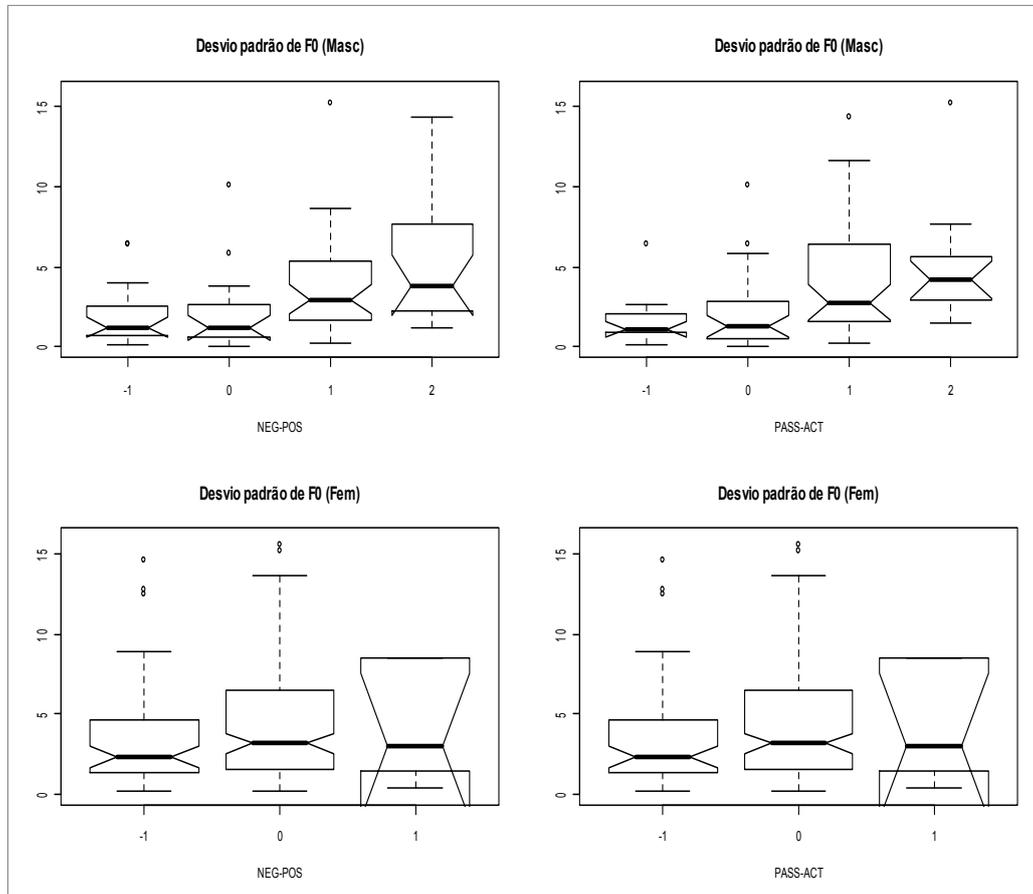


Figura 18: Desvio padrão de F0 em função do género e eixos *Feeltrace*.

No que respeita à análise dos valores de desvio padrão de F0, confirma-se o que foi observado e descrito para o valor médio de F0. Para os homens existe um aumento tanto para maior actividade como para as emoções positivas.

Para as vozes femininas verificamos que as posições são todas elas bastante similares em qualquer um dos eixos. O que não nos permite, uma vez mais, tirar conclusões quanto a estes valores.

#### 4.1.2.2. Jitter

A mesma análise em termos de variação nos dois eixos do *Feeltrace* foi realizada para os parâmetros *jitter* local e *jitter* PPQ5 encontrando-se os resultados nos gráficos na **Figura 19** e **Figura 20**. Nesta análise da variação não dividimos as marcações para homens e para mulheres.

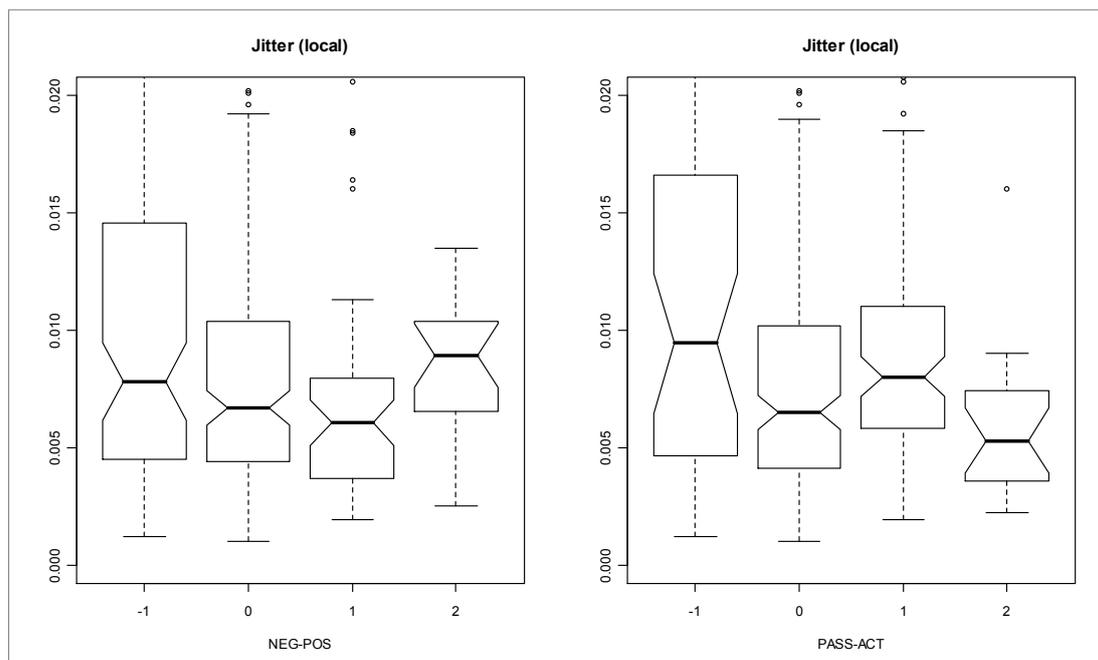


Figura 19: Valores para *Jitter* local em função do eixo negativo-positivo (à esquerda) e passivo-ativo (à direita).

Os valores de *jitter* local são, claramente, superiores para marcações mais negativas. No eixo passivo-ativo, são também mais elevados os valores para emoções mais passivas. Assim, os valores mais baixos verificam-se para segmentos marcados como medianamente positivos e como mais activos.

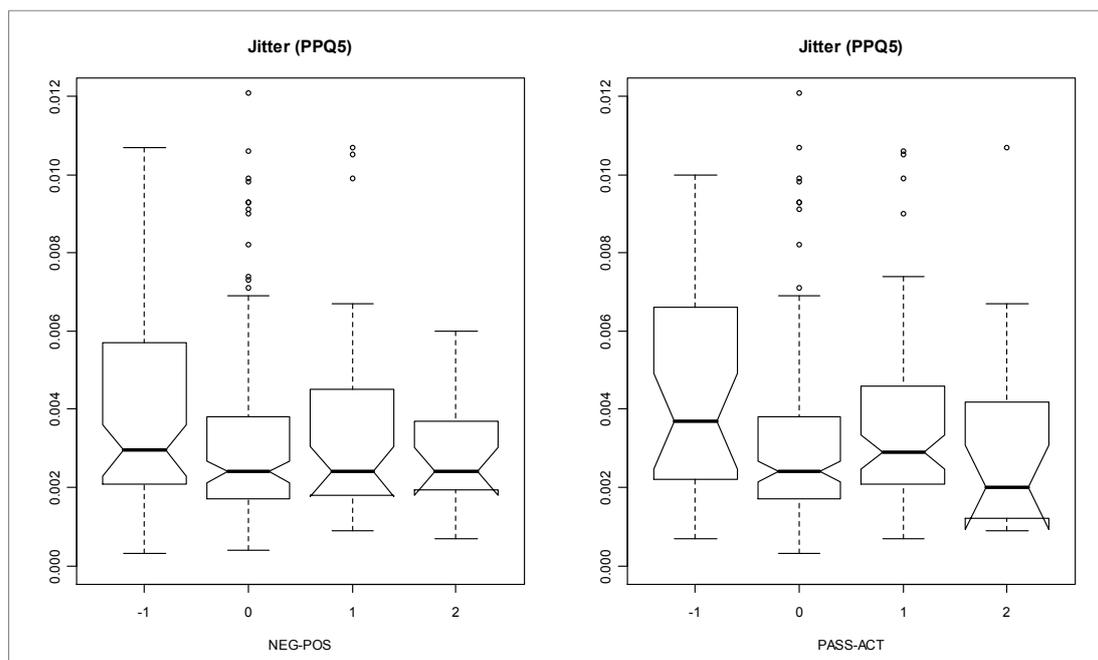


Figura 20: Valores para *Jitter* PPQ5 em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).

O parâmetro *jitter* PPQ5 apresenta um comportamento, em termos dos valores médios, similar ao observado para *jitter* local. No entanto existe agora uma maior sobreposição dos intervalos de confiança, (por exemplo nos parâmetros negativo-positivo), tornando não significativas algumas das diferenças observadas acima.

#### 4.1.2.3. Shimmer

Também para o estudo do parâmetro relacionado com a amplitude, analisámos dois indicadores diferentes: *shimmer* local e *shimmer* PPQ3, os resultados são possíveis de observar na **Figura 21** e seguinte. De referir mais uma vez que aqui também não separámos as análises por género.

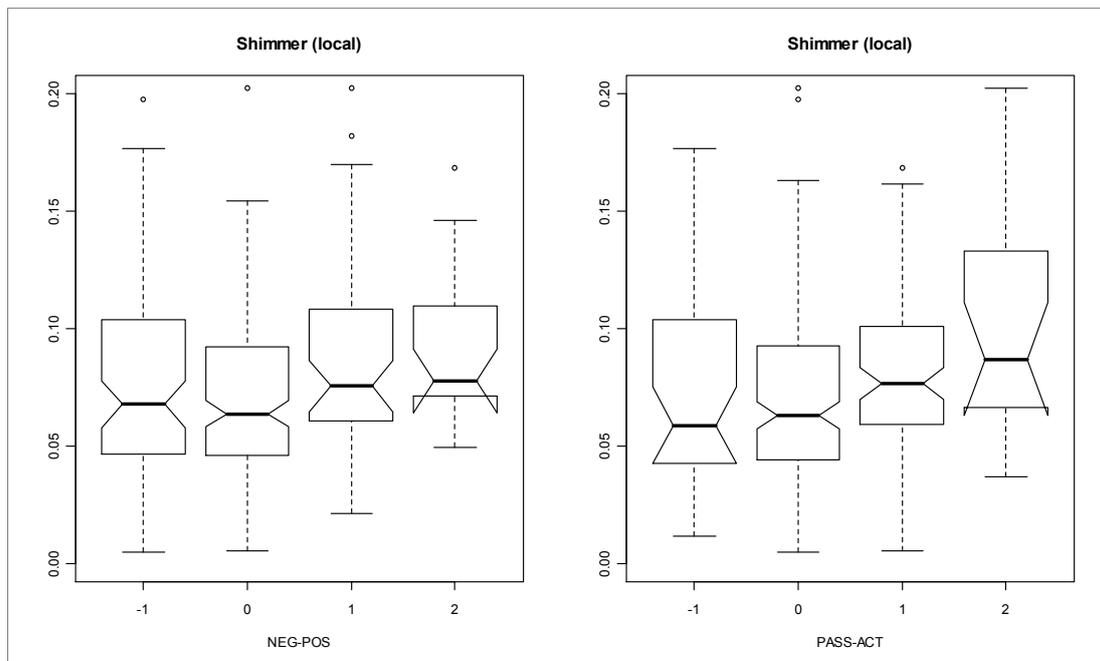


Figura 21: Valores para *shimmer* local em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).

Nos gráficos da **Figura 21** verificamos que no que diz respeito aos valores para *shimmer* local existe alguma similaridade, havendo apenas valores ligeiramente mais elevados no eixo positivo. No entanto notamos a extrema proximidade com o extremo mais negativo (-1). Quanto às diferenças no eixo passivo-activo é evidente o destaque do extremo mais activo (2).

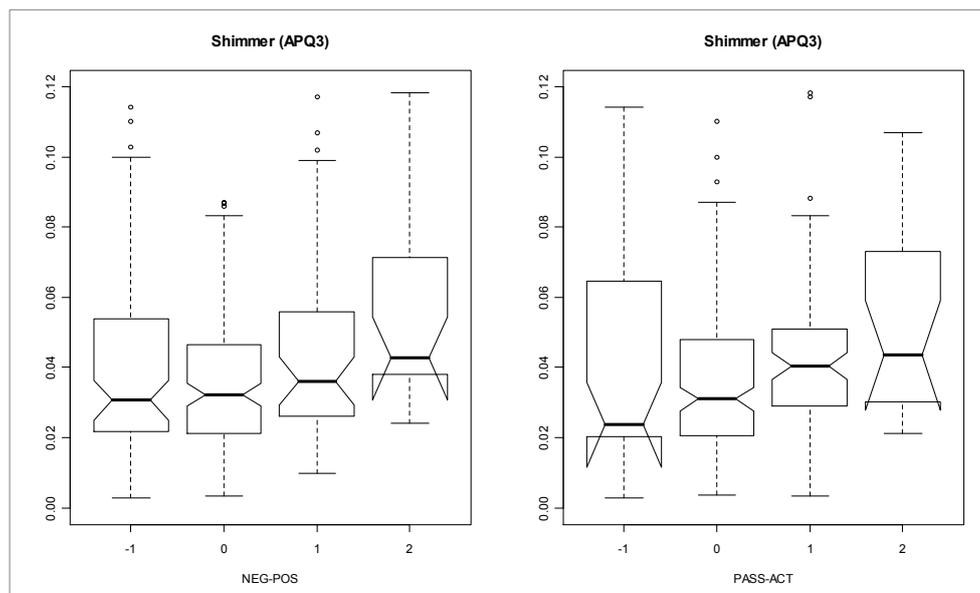


Figura 22: Valores para *Shimmer* APQ3 em função do eixo negativo-positivo (à esquerda) e passivo-activo.

Quanto à análise dos valores de *shimmer* APQ3 percebe-se uma variação menor no que diz respeito aos eixos negativo-positivo, embora seja o extremo mais positivo (2) que apresenta os valores mais elevados, a diferença entre os outros três níveis não é muito grande, principalmente se atentarmos na marcação do eixo mais negativo (-1) e do eixo próximo da neutralidade (1).

No que respeita à análise do outro eixo do *Feeltrace* observa-se também que é o extremo mais activo o que apresenta os valores mais altos, seguido, pelo oposto (-1), sendo que em níveis muito similares se encontram os registos próximos da neutralidade.

#### 4.1.2.4. HNR

Os gráficos da **Figura 23** são referentes aos valores do *Harmonic Noise Ratio*, em função dos eixos do *Feeltrace*. Uma vez mais não há divisão por género, esta só foi efectuada para a análise de F0, uma vez que este é um dos parâmetros que mais diferencia a voz masculina da feminina.

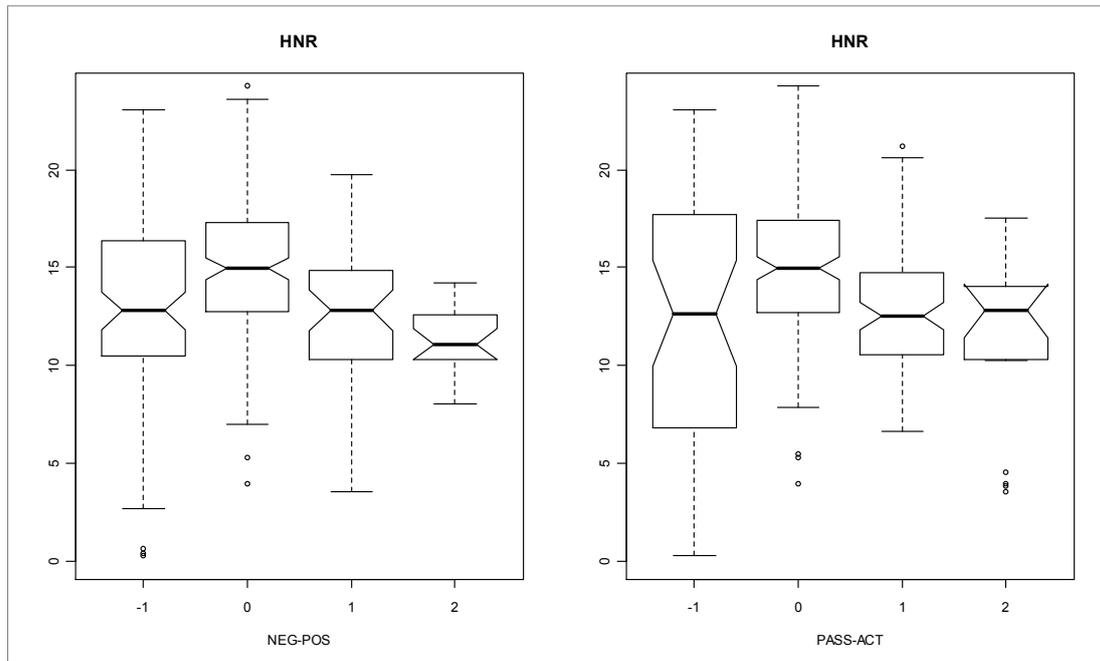


Figura 23: Valores para HNR em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).

Pela observação dos gráficos da figura, percebe-se que existe pouca diferença entre ambos os eixos, todos os valores se encontram muito próximos. Se analisarmos ao pormenor verifica-se a existência de valores ligeiramente mais elevados no eixo negativo, sendo que é ainda o que mais representa a neutralidade (0) que é o mais elevado. No outro eixo há uma pequena diferença, sendo os valores mais altos para os extremos passivos. Observa-se ainda uma grande similaridade entre os dois valores do eixo mais activo (1) e (2).

#### 4.2. Comparação com voz emocional produzida por actor

Terminado o estudo do nosso *corpus* de voz emocional espontânea torna-se importante comparar os resultados obtidos com os anteriormente obtidos para a voz de actor. Nesta comparação procuramos, na **Tabela 2**, mostrar as confirmações ou principais diferenças encontradas. Teremos em conta primordialmente os valores obtidos para homem, uma vez que uma diferenciação em termos de qualidade de voz varia de acordo com o sexo.

Tabela 2: Apresentação e comparação dos resultados, para os parâmetros F0; *jitter*; *shimmer* e HNR, na produção de emoções por actor e espontâneas.

<i>Parâmetro</i>	<i>Resultado com voz de actor</i>	<i>Em voz espontânea</i>
------------------	-----------------------------------	--------------------------

F0	<p>Valor médio perto do 300Hz para raiva</p> <p>Alegria para F0 máximo e médio 150 Hz</p> <p>Tristeza e neutra F0 mais baixo, próximo dos 100 Hz</p>	<p>Os valores conseguidos referentes à emoção expressa por homens<sup>1</sup>, mostram que tristeza se encontra muito próxima dos 120 Hz e a neutra nos 150 Hz, sendo portanto também as mais baixas, como no actor.</p> <p>A Alegria<sup>2</sup> encontra-se muito próxima dos 230 Hz.</p> <p>Na generalidade em termos de F0 os valores entre a voz do actor e a emoção espontânea apresentam ligeiras diferenças.</p>
Jitter (PPQ5)	<p>Alegria e estado neutro apresentam os valores mais baixos (próximo dos 0,7%).</p> <p>Raiva e tristeza têm valores mais elevados (mais de 1%).</p>	<p>Os valores mais elevados encontram-se na expressão de tristeza e de raiva, encontrando-se muito próximos; seguindo-se também com valores bastante similares e mais baixos a expressão neutra e de alegria.</p> <p>Estes valores correspondem aos dados que obtivemos para o actor.</p>
Shimmer (APQ3)	<p>Os valores para a expressão de alegria e estado neutro encontram-se próximos dos 3%.</p> <p>Raiva apresenta o valor mais elevado (próximo de 8%) e tristeza com 4%.</p>	<p>Os valores apresentados para a expressão de alegria e raiva são os mais elevados, na ordem dos 4%.</p> <p>A expressão neutra e a tristeza apresentam resultados muito similares e próximos dos 3%.</p> <p>Aqui os valores divergem do nosso estudo para a emoção por actores, principalmente na expressão da alegria e raiva.</p>
HNR	<p>A alegria apresenta valores na ordem dos 14 dB; a expressão neutra apresenta um valor muito próximo dos 12 dB; Tristeza 11 dB e raiva próxima dos 8 dB.</p>	<p>A alegria apresenta valores próximos dos 12 dB; a expressão neutra encontra-se próxima dos 15 dB.</p> <p>Raiva e tristeza têm resultados similares, 13 dB.</p> <p>Na generalidade os resultados são algo idênticos. No entanto destaca-se a grande diferença que existe na expressão de raiva.</p>

## 5. Discussão

Importa, no final deste estudo, mostrar como se caracteriza a percepção de emoções nos dados de fala do PE neste estudo. Como referimos anteriormente, este trabalho serve não só para essa caracterização e descrição mas, também, para se poder fazer uma primeira e breve comparação com resultados por nós obtidos para voz emocional produzida por um actor, podendo vir a servir de amostragem fiel do que são os parâmetros para a emoção em PE.

Comparando os resultados obtidos para a emoção espontânea e para a emoção simulada verificamos que, tendo em conta a análise dos valores relacionados com o F0, existe uma conformidade entre os resultados obtidos na análise dos dois *corpora*. Assim, em termos de média de

---

<sup>1</sup> Não temos para homens os valores de raiva, mas nas mulheres encontra-se próxima dos 225 Hz.

<sup>2</sup> É de referir que as gravações de alegria que temos retratam muito mais a euforia, daí os elevado valores de F0.

F0, falando apenas no contexto de voz masculina (uma vez que para a emoção espontânea temos também gravações de voz feminina), destaca-se a alegria, com os valores mais elevados, sendo que a diferença entre a expressão de tristeza e o discurso neutro continua a ser mínima. As emoções mais activas apresentam valores de F0 mais elevados, segundo Scherer (2004), o que corrobora os resultados do nosso estudo. Ressalvamos que a expressão espontânea de alegria que compõe os nosso *corpus* é, na maioria das vezes, euforia (uma vez que se encontra ligada à excitação própria dos adeptos de futebol), explicando-se assim os valores mais elevados do que para a raiva.

Em termos de F0, não nos foi possível estudar os valores para raiva, no discurso masculino. No entanto, esta emoção, no discurso feminino não aparece com valores mais elevados do que a alegria para a média de F0, apresentando ambas, em termos de desvio padrão, valores muito próximos.

Os nossos resultados mostram que em PE a tristeza apresenta valores mais baixos de F0 do que o discurso neutro, corroborando, uma vez mais os valores apresentados por Cowie et al (2003).

Na emoção produzida pelo actor os valores mais altos de *jitter* estavam associados à raiva e à tristeza (também desespero e medo, mas estes não contemplados neste estudo de emoção espontânea). Logo, emoções negativas apresentam valores de *jitter* mais altos. Também, para a emoção espontânea temos a tristeza com o valor mais elevado. De acordo com o estudo da emoção produzida por actores encontramos uma convergência, uma vez que os valores de *jitter* para o discurso neutro e a manifestação de alegria são os que apresentam valores mais baixos e similares. A raiva; à semelhança do estudo prévio, também apresenta resultados mais próximos da tristeza em relação a *jitter*.

Reportando-nos ao estudo dos parâmetros relacionados com *shimmer*, verificamos que com valores acima dos do discurso neutro encontramos apenas a raiva e a alegria. Neste parâmetro encontramos, pela primeira vez, um desvio em relação aos resultados obtidos pelas análises da emoção produzida pelo actor. Enquanto que no estudo da emoção simulada a raiva também se encontra bastante acima do valor da expressão neutra, a alegria, por sua vez, encontra-se num patamar abaixo.

Por último, analisando os valores da emoção espontânea, tendo em conta o *Harmonic Noise Ratio*, constatamos que todas as expressões de emoção apresentam resultados próximos dos considerados para a normalidade, valores próximos dos 13 dB. Destaca-se que o discurso neutro é o que apresenta valores mais elevados perto dos 15 dB, que era o valor apresentado para a alegria no estudo de emoções simuladas. Na emoção espontânea HNR dá-nos valores muito semelhantes entre todas as emoções e o discurso neutro, o que não acontecia na análise dos resultados para a emoção simulada. Poderá este parâmetro de análise ser um dos mais significativos na distinção de emoções, quando falamos em simulação e espontaneidade. No estudo da emoção por actor, raiva aparecia numa zona que seria, potencialmente, considerada patológica.

Quanto à análise destas emoções espontâneas no *Feeltrace* pudemos perceber que os resultados obtidos para *shimmer* são os opostos aos obtidos para *jitter*. Se em *shimmer* os resultados são mais

negativos e mais passivos em *jitter* são mais positivos e activos. Assemelhando-se os resultados de *shimmer* aos que foram anteriormente observados para os valores dos dois parâmetros de F0 analisados (para os homens). Os resultados de HNR, apesar de serem muito similares entre si, são mais comparáveis aos resultados obtidos para F0 em que há uma elevação nos eixos negativos e passivos, do que a *shimmer* e *jitter* que são, estes sim comparáveis entre si.

No que diz respeito ao teste do *Feeltrace* realizado pelos oito ouvintes, percebemos, pelas respostas que deram na ficha de acompanhamento que o que se tornou mais difícil para a marcação precisa da emoção era a ausência de contexto. Não tendo achado o teste muito difícil, consideraram contudo que é necessária a concentração suficiente para perceber pelo menos a diferença entre o estado neutro e a tristeza. Notaram também alguma dificuldade na diferenciação activo/passivo.

No entanto, e uma vez mais reforçando a ideia de que não tiveram acesso a todo o contexto nem à prévia etiquetagem, os ouvintes corresponderam à nossa designação (pelo menos em termos de família de emoções, o que é importante) e houve entre eles também bastantes acertos.

## **6. Conclusões**

Percebem-se agora algumas das diferenças que, fundamentalmente, existem em parâmetros como *jitter* e HNR, quando falamos de emoção espontânea ou a que é produzida por um actor. Este exagera, propositadamente, determinadas emoções, pela intensidade, entoação e força com que as procura expressar, daí que a expressão da emoção por actores leva a melhores resultados na percepção de emoções (Wilting et al. 2005). Consegue-se perceber que, ainda que as emoções produzidas por actores apresentem algumas diferenças, por serem necessariamente exageradas, há uma proximidade de valores na maioria dos parâmetros utilizados, valorizando assim os estudos que são realizados com base em enunciados produzidos por profissionais da representação.

Contudo o nosso estudo mostra que há diferenças, em alguns parâmetros, mas que estas não são tão grandes que se tornem inviáveis os estudos da emoção produzida por actores. De outra forma, os nossos resultados poderão também ajudar na formação e aperfeiçoamento de actores, ajudando-os a ter uma orientação em termos de que parâmetros serão mais relevantes na expressão de determinadas emoções, por forma a que saibam quais os aspectos que devem melhorar na sua actuação.

Para o PE os resultados a reter são: a proximidade de valores entre o discurso neutro e a expressão de tristeza, quando falamos em termos de *shimmer*, não havendo uma distinção; a semelhança de valores de *jitter*, entre alegria, raiva e expressão neutra e, por último, a grande proximidade de valores de HNR nas emoções raiva, alegria e tristeza.

Estas observações, principalmente no que diz respeito a valores tão próximos entre tristeza e estado neutro, alegria e expressão neutra e a proximidade de valores de *Harmonic Noise Ratio* para

alegria e tristeza, podem ser indicadores de uma condição cultural, em que a expressão de tristeza ou alegria não apresenta grandes marcas diferenciadores em relação ao discurso neutro.

Para *jitter* e *shimmer* os nossos resultados são idênticos aos apresentados em estudos de outras línguas latinas como o Espanhol e diferem, no entanto, dos que são apresentados em relação ao Coreano, onde o discurso neutro apresenta os valores mais altos de *jitter*, sendo medo o que tem os valores mais baixos de *shimmer* (Mendoza, 2008).

Tendo em conta o *shimmer*, os nossos resultados, mostram que a raiva apresenta os valores mais altos, sendo seguida por desespero e medo o que vem de encontro aos estudos apresentado por Drioli et al. (2003).

De acordo com Drioli et al. (2003), os valores mais elevados de *jitter* dizem respeito a emoções como alegria e surpresa, sendo que raiva se encontra logo no seu seguimento. Os valores mais altos de *jitter* para a alegria eram também já apresentados por Johnstone e Scherer (1999). Os nossos resultados mostram, precisamente o contrário ao apresentado pelo estudo citado: a alegria tem claramente um *jitter* mais baixo do que três das outras emoções, no PE. A análise dos valores de *jitter* parece-nos importante para detectar e descrever a alegria. Os valores elevados de *jitter* para alegria e expressão neutra, estão de acordo com os resultados obtidos por Monzo, Alías, Ignasi, Gonzalvo e Planet (2007) num estudo realizado para o Castelhana.

De acordo com Drioli et al. (2003), os valores de HNR são mais baixos para raiva e desespero, duas emoções negativas. Ao contrário dos trabalhos apresentados por estes autores, a nossa pesquisa mostra que para o PE os valores de HNR para a alegria são significativamente mais altos.

## 7 Referências Bibliográficas

Abelin, Asa, Cross-Cultural Multimodal Interpretation of Emotional Expressions – Experimental Study of Spanish and Swedish, Speech Prosody, 2004.

Airas, Matti; Alko, Paavo, Emotions in short vowel segments: Effects of the glottal flow as reflected by the normalized amplitude quotient. Proceedings of Affective Dialogue Systems Workshop, 2004.

Bänziger, Tanja; Scherer, Klaus R., A study of perceived vocal features in emotional speech, Voqual, 2003.

Batliner, Anton; Fischer, K.; Huber, R.; Spilker, J.; Nöth, E., Desperately seeking emotions on Actors wizards, and Human Beings, ISCA Speech and Emotion, 2000.

Boehner, Kirsten; DePaula, Rogério; Dourish, Paul; Sengers, Phoebe, How emotion is made and measured, Science Direct, Elsevier, 2007.

- Boersma, P.; Weenink, D, Praat – doing phonetics by computer, (<http://www.fon.hum.uva.nl/praat>), 1992-2005.
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B., A data base of German Emotional Speech, InterSpeech, 2005.
- Burkhardt, F.; Audibert, N.; Malatesta, L.; Türk, O.; Arslan, L.; Auberge, V., Emotional Prosody – Does Culture Make a Difference?, Speech Prosody, 2006.
- Cabral, João P.; Oliveira, Luís C., “pitch”\_synchronous time-scaling for Prosodic and Voice Quality transformations, InterSpeech, 2005.
- Cabral, João P.; Oliveira, Luís C., EmoVoice: a System to Generate Emotions in Speech, InterSpeech, 2006
- Cabral, João Paulo Serrasqueiro Robalo, Transforming Prosody and Voice Quality to Generate Emotions in Speech, Dissertação de Mestrado, Universidade Técnica de Lisboa, Instituto Superior Técnico, 2006.
- Campbell, Nick, Accounting for Voice-Quality Variation”, Speech Prosody, 2004.
- Campbell, Nick; Changes in Voice Quality Due To Social Conditions, ICPHS XVI, 2007.
- Cowie, E. Douglas; Cowie, R., Describing the emotional states expressed in Speech, Proceedings of the ISCA Workshop on Speech and Emotion, 2000.
- Cowie, E. Douglas; Cowie, R. and Schroeder, M., The description of naturally occurring emotional speech, 15<sup>th</sup> ICPHS, 2003.
- Cowie, R., Cowie, E. Douglas, Savvidou, Susie, McMahon, Edelle, Sawey, Martin, Schröder, M, ‘*Feeltrace*’: An instrument for recording perceived emotion in real time, ISCA Workshop on Speech and Emotion: A conceptual framework for research, 2003.
- Cowie, Roddy, Describing the emotional States Expressed in Speech, ISCA Workshop on Speech and Emotion: A conceptual framework for research, 2003.
- Damásio, António R., O Erro de Descartes, Emoção, Razão e Cérebro Humano, Fórum da Ciência, Publicações Europa-América, 16<sup>a</sup> edição, 1996.
- Darwin, Charles, A Expressão Das Emoções No Homem e Nos Animais, Setembro 2006, Relógio d’Água editores.
- Ekman, Paul, An Argument for Basic Emotions, Cognition and Emotion, 6, 169-200, 1992.
- Gobl, C., Chasaide, A.N. The role of voice quality in communicating emotion, mood and attitude, Speech Communication 40, 189-212, 2003

- Gobl, Christer; Bennet, Eva; Chasaide, Ailbhe Ní, Amplitude-Based source parameters for measuring Voice Quality,, VOQUAL, 2003.
- Laver J., (eds), *The Handbook of Phonetic Sciences*, Blackwell Publishers Ltd., Oxford, 1996.
- Johnstone, T., Scherer, K.R., *The effects of Emotion on Voice Quality*, ICPHS, 2003.
- Ladefoged, Peter, *Preliminaries to linguistic phonetic*, Chicago: University of Chicago Press, 1971.
- Ladefoged, Peter, *Phonetic Data Analysis*, “An Introduction to fieldwork and instrumental Techniques”, pp.75-103; 169-181, Blackwell, 2003.
- Laver, John, *The gift of speech: papers in the analysis of speech and voice*, Edinburgh, University Press, 1991.
- Lewis, Michael; Haviland-Jones, Jeannette M.; Barrett Lisa Feldman, (edited by) *Handbook of Emotions*, Third Edition, the Guilford Press, 2008.
- Maddieson, Ian; Ladefoged, Peter, *The sounds of the world's languages*. Oxford: Blackwell, 1996.
- Patrik, N. Juslin; Västfjäll, Daniel, *All emotions are not created equal: Reaching beyond the traditional disputes*, Behavioral and Brain Sciences, 2008, Cambridge University Press.
- Pell, Marc D.; Monetta, Laura; Paulmann, Silke; Kotz, Sonja A., *Recognizing Emotions in a Foreign Language*, Springer, 2009.
- Rodrigues, Américo, *As Emoções na Fala*, Tese de Mestrado apresentada à Universidade de Aveiro, 2007.
- Sawamura, Kanae; Dang, Jianwu; Akagi, Masato; Erickson, Donna; Li, Aijun; Sakuraba, Kyoko; Minematsu, Nobuaki; Hirose, Keikichi, *Common Factors in Emotion Perception Among Different Cultures*, ICPHS XVI, 2007.
- Scherer, Klaus R., *Vocal communication of emotions: A review of research paradigms*, *Speech Communication*, 40, Elsevier, 2003, pp. 227-256.
- Scherer, Klaus R., *What are emotions? And How can they be measured?*, *Social Sciences information*, 2005
- Teixeira, António; Nunes, Ana Margarida Belém; Coimbra, Rosa Lídia; Moutinho, Lurdes, *Voice quality with Psychogenic disfunction: a case study.*, *ClinicalLinguistics and Phonetics*, volume [22](#), Outubro de 2008, páginas 906-916.
- Wilting, Janneke; Krahmer, Emiel; Swerts, Marc, *Reals vs Acted Emotional Speech*, *Speech Communication*, 2005