



**Ana Margarida
Belém Nunes**

Voz e Emoção em Português Europeu

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Ciências e Tecnologias da Saúde, realizada sob a orientação científica do Doutor António Joaquim da Silva Teixeira, Professor Auxiliar do Departamento de Electrónica Telecomunicações e Informática da Universidade de Aveiro e da Doutora Rosa Lídia Coimbra, Professora Auxiliar do Departamento de Línguas e Culturas da Universidade de Aveiro.

Apoio financeiro da FCT, no âmbito do
Programa POCTI, integrado no III
Quadro Comunitário de Apoio.

o júri

presidente

Doutor Luís António Ferreira M. Dias Carlos
Professor Catedrático da Universidade de Aveiro
em representação da Reitora da Universidade de Aveiro

vogais

Doutor Nelson Fernando Pacheco Rocha
Professor Catedrático da Universidade de Aveiro

Doutor Jorge Eduardo de Freitas Spratley
Professor Auxiliar Convidado da Faculdade de Medicina da Universidade do Porto

Doutor António Joaquim da Silva Teixeira
Professor Auxiliar da Universidade de Aveiro (orientador)

Doutora Rosa Lúcia Torres do Couto Coimbra e Silva
Professora Auxiliar da Universidade de Aveiro (co-orientadora)

Doutora Ana Paula de Brito Garcia Mendes
Professora Adjunta da Escola Superior de Saúde, Instituto Politécnico de Setúbal

Doutor António José Monteiro Amaro
Professor Adjunto da Escola Superior de Saúde da Universidade de Aveiro

Doutora Daniela Filipa Macedo Braga Moreira da Silva
Program Manager Lead, MLDC - Microsoft Language Development Center

agradecimentos

Um agradecimento muito especial à minha mãe, pelo apoio incondicional, a força e motivação que me passou a cada momento. Pelo que sempre fez por mim e pelo que fez de mim. Ao Luiz pela companhia, força, carinho e energia transmitidas. À minha família que sempre me acompanhou e acreditou em mim.

Agradeço ao Professor António Teixeira, a forma como me orientou, a sua compreensão, tempo, paciência e entrega. Agradeço ainda o modo como me ajudou a crescer em termos científicos e pessoais pela forma como acreditou em mim, quando me competia representar o nosso trabalho em Conferências Internacionais. À minha co-orientadora, Rosa Lídia Coimbra, pela amizade, compreensão, força e apoio que sempre me ofereceu.

Agradeço também aos amigos verdadeiros aqueles que estiveram sempre presentes a dar o apoio e a força necessários, acreditando e apostando sempre em mim.

Merecem ainda um agradecimento particular a Universidade de Aveiro pela concessão de bolsa para a realização do Doutoramento; a Fundação para a Ciência e Tecnologia pelo apoio para deslocação a Congressos e a Fundação Calouste Gulbenkian pela bolsa concedida para a investigação nos EUA.

palavras-chave

Voz, Emoção, Análise Acústica, Português Europeu

resumo

No trabalho apresentado realiza-se uma primeira descrição de voz e emoção para o Português Europeu.

Estudamos, utilizando como base estudos realizados em diversas línguas (finlandês; inglês; alemão), os parâmetros relacionados com voz e que variam consoante a emoção que expressamos. Analisamos assim os parâmetros relacionados com a frequência Fundamental (F0) com a perturbação (*jitter*) com a amplitude (*shimmer*) e com aspectos relacionados com o ruído (HNR).

Trata-se de um estudo abrangente que estudando voz e a sua relação/variação de acordo com a emoção o faz em três vertentes: patologia de voz de origem psicogénica (carácter emocional); emoção produzida por actores e a análise de emoção espontânea. Conseguindo, como trabalho pioneiro nesta área, valores para todos estes tipos de produção. Salientamos o facto de no nosso trabalho apenas existir a análise de voz sem recurso a expressão facial ou à postura dos indivíduos.

Para que pudéssemos realizar estudos comparativos com os dados que íamos recolhendo em cada *corpus* (patologia; emoção por actor e emoção espontânea), procurámos utilizar sempre os mesmos métodos de análise (Praat; SFS; SPSS, *Hoarseness Diagram* – para a análise de voz com patologia - e o sistema *Feeltrace* - para as emoções espontâneas).

Os estudos e análises relativos à emoção produzida por actores são complementados por testes de percepção aplicados a falantes nativos de Inglês Americano e a falantes de Português Europeu. Este teste, juntamente com a análise da emoção espontânea, permitiu-nos retirar dados particulares relativos à língua portuguesa. Apesar de haver tanto na expressão como na percepção de emoções muitas características consideradas universais, em Português percebe-se algo de peculiar. Os valores para a expressão neutra; tristeza e alegria são todos muito próximos, ao contrário do que acontece noutras línguas. Além disso estas três emoções (de famílias distintas) são as que mais dificuldades causam (aos dois grupos de informantes) em termos de distinção no teste de percepção. Poderá ser esta a particularidade da expressão da emoção no Português Europeu, podendo estar ligada a factores culturais.

Percebe-se ainda, com este trabalho, que a emoção expressa pelo actor se aproxima da emoção espontânea. No entanto, alguns parâmetros apresentam valores diferentes, isto porque o actor tem a tendência de exagerar a emoção.

Com este trabalho foram criados *corpora* originais que serão um recurso importante a disponibilizar para futuras análises numa área que é ainda deficitária, em termos de investigação científica, em Portugal. Tanto os *corpora*, como respectivos resultados obtidos poderão vir a ser úteis em áreas como as Ciências da Fala; Robótica e Docência.

keywords

Voice, Emotion, Acoustic Analysis, European Portuguese

abstract

This research is a first description of voice and emotion for European Portuguese.

Based on studies for several languages (Finish, English, German), we studied voice related parameters varying according to the emotion expressed. The analysed parameters are related to F0, jitter, shimmer and Harmonic Noise Ratio (HNR).

This is an all-embracing study that approaches voice and its relation/variation according to emotion. Three sources are considered: psychogenic voice pathology; emotion produced by actors; and spontaneous emotion analyses. Being a trailblazing research in this area, values were measured for all three types of production. We highlight the fact that our work only considers voice analysis without considering facial expression and body posture.

In order to make comparative studies with the data collected for each corpus (pathology, acted emotion and spontaneous emotion), we used the same analysis methods (Praat, SFS, SPSS, and Hoarseness Diagram - for pathological voice analysis; and Feeltrace system - for the spontaneous emotions).

Studies and analyses related to acted emotion are complemented by perceptual tests with American English and European Portuguese speakers. These tests, as well as spontaneous emotion analysis, allowed the extraction of data for Portuguese. Both emotion expression and perception have many universal characteristics. However, Portuguese language proved to have some particularities. Values obtained for neutral expression, sadness, and joy are very close, contrary to what happens in other languages. Moreover, these three emotions (from distinct families) are those that present the most difficulty (to both informant groups) in the distinction perceptual tests. This is probably the main particularity in emotion expression as far as Portuguese is concerned, maybe due to cultural factors.

This research also shows that acted emotion is close to spontaneous emotion expression. However, some parameters present different values, because the actor tends to somewhat exaggerate the emotion.

This work led to the creation of original corpora that can be an important resource for future analyses in an area still in deficit in terms of scientific research in Portugal. Both corpora and the obtained results may be useful in areas such as Speech Science, Robotics, and Education.

Índice

Capítulo 1 -	Introdução.....	1
1.1.	Motivações	2
1.2.	Objectivos do trabalho	9
1.3.	Estrutura da Tese	10
1.4.	Trabalhos publicados.....	12
Capítulo 2 -	Acerca de Voz e Emoção	15
2.1.	Voz	15
2.2.	Emoção	54
2.3.	Voz e Emoções.....	70
Capítulo 3 -	Estudo de voz de uma patologia de base emocional.....	87
3.1.	Disfonia psicogénica.....	88
3.2.	Trabalhos relacionados	92
3.3.	Metodologia	95
3.4.	Resultados.....	100
3.5.	Discussão	108
3.6.	Comentários Finais.....	109
Capítulo 4 -	Estudo de voz emocional de um actor.....	111
4.1.	Voz emocional produzida por actores.....	113
4.2.	Análise acústica.....	122
4.3.	Teste perceptual	134
4.4.	Discussão	140
4.5.	Comentários Finais.....	143
Capítulo 5 -	Estudo de voz emocional espontânea	145
5.1.	Trabalhos relacionados	146

5.2.	Metodologia	148
5.3.	Anotação das emoções	150
5.4.	Resultados	163
5.5.	Comparação com voz emocional produzida por actor	177
5.6.	Discussão	179
5.7.	Conclusões	181
Capítulo 6 -	Conclusão	183
6.1.	Resumo do trabalho.....	183
6.2.	Principais resultados.....	184
6.3.	Sugestões de continuação.....	185
Capítulo 7 -	Referências Bibliográficas	187

Lista de Figuras

Figura 1 – Anatomia da laringe. Fonte: University of Pittsburgh Voice Center	18
Figura 2 – Todos os elementos envolvidos no processo de produção de voz. Fonte: <i>Tutorials – Voice Production. National Center for Voice and Speech.</i>	19
Figura 3 – Estruturas envolvidas na fala e produção de voz. Fonte: <i>National Institute on Deafness and Other Communication Disorders (NIDCD), 2008.</i>	22
Figura 5 – Processo de vibração das cordas vocais. Fonte: University of Pittsburgh Voice Center.	24
Figura 6 – Ciclo de abertura e fechamento das pregas vocais, ciclo glotal (Titze, 1998).....	25
Figura 7 – Movimento das pregas vocais.	26
Figura 8 – Representação da onda da voz modal Masculina. Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive.....	33
Figura 9 – Representação da onda da qualidade de voz estridente para o homem. Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive...	35
Figura 10 – Representação da onda da qualidade de voz aspirada Masculina. Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive...	36
Figura 11 – Representação da onda da voz falsete masculina. Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive.....	38
Figura 12 – Representação da amígdala, responsável pelo reconhecimento das emoções.	61
Figura 13 – Descrição das emoções proposta por Scherer (2005).....	66
Figura 14 – Círculo bidimensional de amostragem da qualidade das emoções, (GEW), http://www.unige.ch/fapse/emotion/resmaterial/gew.zip	67
Figura 15 – Espectrograma, retirado do Praat, da vogal sustida [a].	98
Figura 16 – Parâmetros do Voice Report do Praat relativos à vogal sustida [a]. ..	99

Figura 18 - Localização de patologias no <i>Hoarseness Diagram</i> (Fröhlich, 1997).	101
Figura 19 - Localização, de acordo com o <i>Hoarseness Diagram</i> , da patologia analisada e descrita: Assinalado pela seta verde temos a localização da fonação normal, pela seta vermelha o local onde se encontra no diagrama a fonação psicogénica.	102
Figura 20 – Parâmetros relativos a F0: média, mínimo, máximo e desvio padrão.	103
Figura 21 – Intervalos de confiança para três parâmetros de <i>Jitter</i> (local, RAP e PPQ).	105
Figura 22 – Intervalos de confiança a 95% para dois parâmetros relativos ao <i>Shimmer</i> (local e APQ3) em função do tipo de voz e tipo de produção.	106
Figura 23 – Intervalo de confiança a 95 % para HNR em função do tipo de voz e tipo de material/produção analisado.	107
Figura 24 - Exemplo de elisão de [ə] – /u mLOr sra tumarS ko~nt6d el@s/.....	124
Figura 25 - Exemplo de crase [6]+[6] - [a] e acrescentamento [j] – /n6~w~ t6Ju ko~ s@rteza vOS djum6 psow6 kiSko~d@ kwalker kojz6/	125
Figura 26 – Efeitos da emoção nos 4 parâmetros de F0 – de cima à esquerda, diagrama de extremos e quartis para: média, desvio padrão, mínimo e máximo.	126
Figura 27 – Correlações entre os valores máximo e mínimo de F0 para as 5 emoções mais a expressão neutra.	128
Figura 28 – Efeitos da emoção no parâmetro <i>jitter</i> PPQ5.....	129
Figura 29 – Efeitos da emoção em <i>Shimmer</i> APQ3.....	130
Figura 30 - Efeitos da emoção em Harmonic Noise Ratio (HNR).	132
Figura 31 – Comparação das 5 emoções e expressão modal baseado nos quatro parâmetros analisados. Para F0 foi escolhido o valor da mediana como representativo dos outros quatro parâmetros que fazem parte do nosso estudo.	133

Figura 32 – Comparação entre as respostas dos dois grupos de informantes. Na legenda PT é usado para Português Europeu e AE para Inglês Americano.	136
Figura 33 – Resultados gerais de acertos (informantes/emoções).....	137
Figura 34 – Percentagens de acerto em função do tipo de frase.	138
Figura 35 – Resultados para cada emoção e falante de acordo com o tipo de frase.....	139
Figura 36 – Exemplo do <i>Feeltrace</i> 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.	152
Figura 37 – Grau de confiança na marcação da emoção ouvida.....	154
Figura 38 – Média do grau de confiança na marcação das emoções ouvidas, por indivíduo.	155
Figura 39 – Média do grau de confiança, por emoção marcada como primeira.	156
Figura 40 – Grau de intensidade atribuído à primeira escolha	157
Figura 41 – Diversidade das escolhas para descrever a emoção ouvida.	158
Figura 42 – Etiquetagem realizada para o primeiro enunciado. Família da emoção da primeira escolha: Negativa.....	159
Figura 43 – Exemplo de resultados <i>Feeltrace</i> . Anotações dos vários ouvintes e média das mesmas.....	161
Figura 44 – Exemplo de resultados <i>Feeltrace</i> , em que existe muita variabilidade. Anotações dos vários ouvintes e média das mesmas.	162
Figura 45 – Valores de média de F0 e desvio padrão para os informantes masculinos e femininos, da esquerda para a direita.....	164
Figura 46 – Valores de <i>jitter</i> local para as diferentes expressões estudadas.	166
Figura 47 – Valores de <i>jitter</i> PPQ5 para as diferentes expressões estudadas. ...	167
Figura 48 – Valores de <i>Shimmer</i> local e APQ3 para as quatro emoções e discurso neutro.....	168

Figura 49 – Valores de HNR para os cinco enunciados estudados.	169
Figura 50 – Valores médios de F0 em função do género e dos eixos do <i>Feeltrace</i>	171
Figura 51 – Desvio padrão de F0 em função do género e eixos <i>Feeltrace</i>	172
Figura 52 – Valores para <i>Jitter</i> local em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).	173
Figura 53 – Valores para <i>Jitter</i> PPQ5 em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).	174
Figura 54 – Valores para <i>shimmer</i> local em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).	175
Figura 55 – Valores para <i>Shimmer</i> APQ3 em função do eixo negativo-positivo (à esquerda) e passivo-activo.	176
Figura 56 – Valores para HNR em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).	177

Lista de Tabelas

Tabela 1 – Parâmetros utilizados na síntese das emoções básicas por Cabral e Oliveira (2006), partindo de estudos em outras línguas, ajustados pelos investigadores à língua Portuguesa.....	5
Tabela 2 – Factores que modificam os parâmetros de fala, relacionados com qualidade de voz.....	5
Tabela 3 – As quatro fases de esforço muscular, no ciclo respiratório. Retirado de: <i>Tutorials – Voice Production. National Center for Voice and Speech.</i>	20
Tabela 4 – Características fonéticas de cada qualidade de voz	32
Tabela 5 – Métodos de observação da actividade da fonte glotal.	43
Tabela 6 – Representação dos subsistemas orgânicos e componentes do processo emotivo (Scherer, 2003).	56
Tabela 7 – Amostragem das diferentes emoções apresentadas como primárias, por vários estudiosos, ao longo do tempo.	62
Tabela 8 – Relação emoção-qualidade de voz (Yanusgevskaia et al, 2008).....	83
Tabela 9 – Valores médios de F0, para o género feminino, tendo em conta a idade e situação clínica.	85
Tabela 10 – valores para duração, F0 e intensidade para as diferentes emoções analisadas nas vogais tónicas produzidas em /'aba/ e /'ava/.....	115
Tabela 11 – Resumo dos resultados das análises de Rodrigues (2007). Retirado de Rodrigues (2007), p. 65.	121
Tabela 12 – Análise estatística dos resultados do teste de percepção. Nota: foram eliminadas algumas linhas e colunas produzidas pelo SPSS cuja informação não se considerou essencial.....	140
Tabela 13 – Resumo da marcação das emoções em todas as frases e respectivo grau de confiança.	160
Tabela 14 – Apresentação e comparação dos resultados, para os parâmetros F0; <i>jitter</i> , <i>shimmer</i> e HNR, na produção de emoções por actor e espontâneas.....	178

Capítulo 1 - Introdução

*Não escondas nas palavras coisa alguma:
Ouvindo-se sabe o que não se lê.
Não finjas na voz que sempre te engana
No timbre da voz é que a gente vê.*

Augusto Deodato Guerreiro – site LerparaVer, 2006

A voz é o nosso meio de comunicação por excelência. Ela transporta múltiplas informações, sendo um importante atributo do Ser Humano. A análise e estudo da voz fornecem indícios sobre os parâmetros físicos, psicológicos, sociais e culturais do indivíduo (Gilvaneia et al. sem data). A produção de voz é o resultado da acção de um sistema versátil e complexo em que se associam diversos órgãos e estruturas como o cérebro, os pulmões, a laringe, as cavidades nasais, a língua e os lábios.

A voz dá a conhecer sobre o indivíduo o grau de afectividade e de interesse ou mesmo o estado emocional em que ele se encontra. Outros aspectos veiculados referem-se a características que permitem identificar o género e a idade (Laver, 1994).

Para além da riqueza e unicidade do timbre, que define em grande parte a individualidade do emissor, há ainda um universo imenso de combinações e variações múltiplas de entoações, projecções, articulações, modulações, durações, pausas, ritmos, intensidades dos sons e dos silêncios que os envolvem. É neste sentido paralinguístico que se enquadra toda a informação da voz associada à **emoção**. Como características extra-linguísticas e paralinguísticas entendemos o que não está incluído no âmbito da linguística e da linguagem.

Apesar de estudos relativos à forma como a emoção se transmite pela voz e como a afecta começarem já a merecer um maior enquadramento científico, continua ser muito difícil conseguir classificar de forma concreta e precisa a informação afectiva. Este aspecto só vem dificultar a utilização da voz associada à expressão de emoção nas novas tecnologias.

No entanto, é graças ao desenvolvimento destas novas tecnologias que vem merecendo um maior destaque o estudo das emoções e da sua expressão pela voz. Exemplo disso é o estudo exaustivo e reconhecido cientificamente de António Damásio e colaboradores, na origem de “*O Erro de Descartes: Emoção, Razão e Cérebro Humano*” (1995). Merece também destaque, nesta área, o estudo alargado do *Geneva Emotions Research Group*, que vem sendo realizado na Universidade de Genebra, sob a orientação do Professor Klaus Scherer.

1.1. Motivações

Vários factores, interdisciplinares (Linguística; Tecnologias da Saúde; Síntese de Voz) e a escassez de material, nesta área específica, para o Português Europeu, motivam o presente trabalho:

1.1.1 Necessidade de contemplar aspectos paralinguísticos

Importa estudar, para além do que reconhecemos como único, o timbre da voz do indivíduo, as diversas variações e parâmetros que permitem verificar como a voz é afectada pelas emoções. Estes aspectos não são normalmente estudados em termos linguísticos, dizendo respeito à dimensão paralinguística. Englobam as

diferentes entoações, formas de articulação, duração, pausa, cadência e intensidade, por exemplo. É neste sentido que se enquadra toda a informação da voz associada à emoção.

1.1.2 Necessidade de contemplar aspectos emocionais na avaliação clínica

É relevante reconhecer como certos aspectos emocionais podem afectar a voz, condicionando os comportamentos sociais e até profissionais do sujeito. Importa que em termos clínicos se tenham em atenção factores psicoemocionais que podem estar na origem de perturbações/patologias de voz e que o trabalho do terapeuta seja acompanhado e complementado, muitas vezes, por um psicólogo. Nos nossos dias muitas perturbações podem advir de pressões sociais, profissionais (sem que haja malformações ou anomalias físicas) o stress acarretado por todas essas condições, acaba por desencadear no indivíduo problemas em termos de fonação.

1.1.3 Necessidade da inclusão de reconhecimento e expressão de emoções pela voz em sistemas de interacção Humano-computador

O reconhecimento fiável das emoções vai permitir que os computadores e todos os sistemas mais automatizados consigam estabelecer um tipo de comunicação mais natural e adequado, o que permitirá, por seu turno, um avanço na implementação de sistemas de apoio, por exemplo em termos de saúde ou de pessoas idosas. No que concerne a estas, os computadores poderiam detectar problemas e contribuir para a reabilitação pela interacção com componente emocional com os idosos (Ren, 2009, pp.48). A investigação pode ainda, segundo este mesmo autor e neste seu recente estudo, ir mais além e pode vir a ser possível diagnosticar de forma automática doenças de foro mental.

É esta premente necessidade de interacção, o mais natural possível, entre homem e máquina e a utilização crescente de sistemas automáticos de diálogo entre humanos e máquinas, que torna da máxima importância a obtenção de informação acerca do estado emocional do interlocutor humano, pela análise da

sua voz. O aumento do conhecimento e capacidades de análise da qualidade de voz terá um impacto directo no desempenho destes sistemas.

1.1.4 Necessidades de síntese de voz expressiva

Também com o aumento da qualidade dos sistemas de síntese, algumas deficiências, anteriormente camufladas, tornam-se notórias. O estudo sobre a variação da qualidade de voz, podendo vir a melhorar aspectos prosódicos, ajuda a uma mais correcta percepção e transmissão de informação extra linguística (emoção). A síntese da expressão de emoções tem vindo a demonstrar que é impossível obter “sons naturais sem se trabalharem e ter em conta os parâmetros relacionados com qualidade de voz” (d'Allessandro, 2003). Uma área importante de aplicação destes sistemas são os jogos de computador e entretenimento.

O EmoVoice (Cabral e Oliveira, 2006) é um sistema de modificação manipulação do discurso, criado para estudar a correlação entre os parâmetros da fala e os estados emocionais. Este sistema foi implementado por investigadores Portugueses, transformando algumas regras acústicas para conseguir simular sete emoções básicas, para além do discurso modal. Este sistema utiliza o *Pitch-synchronous Time-Scaling* (PSTS) do sinal de excitação para poder alterar a prosódia e os parâmetros mais relevantes ligados à fonte glotal e qualidade de voz. Numa avaliação geral, o sistema consegue gerar emoções possíveis de serem reconhecidas, no entanto e para discriminar alguns pares de emoções, precisa ainda de melhoramentos.

No referido trabalho para o Português Europeu as fontes utilizadas foram investigações até agora realizadas para outras línguas. Assim, nesta pesquisa, frequência fundamental, duração e energia foram baseadas nos valores apresentados por Drioli et al. (2003); Zovato et al. (2004) e Whiteside, (1998). Os valores utilizados são apresentados na Tabela 1.

Tabela 1 – Parâmetros utilizados na síntese das emoções básicas por Cabral e Oliveira (2006), partindo de estudos em outras línguas, ajustados pelos investigadores à língua Portuguesa.

<i>Emoção</i>	<i>Média de F0</i>	<i>Gama de F0</i>	<i>Duração</i>	<i>Intensidade</i>
Raiva	15%	30%	-16%	70%
Alegria	18 %	30%	-19%	30%
Tristeza	-16%	-38%	16,5%	-50%
Medo	25%	30%	-10%	0%
Surpresa	15%	40%	-5%	2%
Aborrecimento	-16%	0%	10%	-20%
Aversão	-20%	20%	-10%	-20%

Os parâmetros relacionados com qualidade de voz utilizados neste trabalho foram: ruído; *jitter*; *shimmer*; parâmetros da fonte glotal; quociente de abertura (*Open Quotient*); quociente de velocidade (*Speed Quotient*) e de retorno (*Return Quotient*). O ruído, relacionado com a respiração cria a sensação de voz aspirada enquanto que *jitter* e *shimmer* dão-nos a sensação de uma voz áspera. Mais uma vez utilizando resultados de outras línguas (Whiteside, 1998 e Drioli et al, 2003), realizaram com as suas próprias avaliações perceptuais, as variações dos parâmetros anteriormente referidos e que estão presentes na Tabela 2.

Tabela 2 – Factores que modificam os parâmetros de fala, relacionados com qualidade de voz.

<i>Emoção</i>	<i>Jitter</i>	<i>Shimmer</i>	<i>Ruído de Aspiração (dB)</i>	<i>Quociente de abertura</i>	<i>Quociente de velocidade</i>	<i>Quociente de retorno</i>
Raiva	6	40	0	0,9	0,9	1,5
Alegria	3	0	9	0,8	1,0	0,7
Tristeza	0	0	0	0,8	1,2	0,9
Medo	0	0	0	0,5	1,0	1,0
Surpresa	7	0	7	1,0	1,0	1,0
Aborrecimento	0	0	0	1,0	1,0	1,0
Aversão	0	0	0	1,0	1,0	1,0

Utilizaram uma base de dados de voz emocional para Alemão – contemplando sete emoções: estado neutro; raiva; alegria; tristeza; medo; aborrecimento e aversão - e o sistema Aparat que é um sistema de

parametrização e filtro inverso para voz, criado por Airas et al. (2005), para conseguirem os três últimos parâmetros (OQ, SQ, RQ). Os resultados mostraram que o OQ desce, especialmente para a alegria e medo, o que é consistente com o aumento de energia nas frequências mais altas. O SQ para raiva e alegria desce, o que pode ser explicado pela voz aspirada associada a estas emoções. O RQ é muito mais elevado na expressão de raiva, o que se liga à qualidade de voz relacionada com esta emoção (voz aspirada).

É importante realizar estudos de base sobre as características da voz emocional em Português para que sistemas como o EmoVoice possam evoluir.

1.1.5 Estudos recentes sobre acústica de voz emocional

Recentemente tem vindo a ser constatado que muitos resultados ou conhecimentos que existem sobre qualidade de voz e emoção são na maioria das vezes estereotipados, porque simulados por actores. Estes adoptam propositadamente uma postura de exagero em relação aos padrões de determinada emoção, o que não reflecte, na sua essência, os efeitos incontroláveis de uma emoção espontânea. Este tipo de investigação tem por base *corpora* produzidos por actores. Contudo, de acordo com Cowie et al (2003) este tipo de análise, aproxima-se da realidade, havendo, contudo, características de discurso espontâneo e natural que ainda não se conseguiram extrair.

O estudo da expressão emocional baseado apenas em parâmetros relacionados com qualidade de voz tem sido até agora pouco aprofundado, de acordo com Bänziger e Scherer (2003), Cowie, Schröder (2003). Estes autores referem que os elementos até agora mais analisados são produzidos em monólogo, o que não é uma situação normal de expressão das diferentes emoções. Estas são sempre mais fortes e naturais em contexto interactivo. Nesta situação, é possível analisar a percepção, pelo outro, da diferença de entoação que poderá demonstrar uma alteração em termos emocionais.

Estudando as análises feitas para outras línguas, temos uma visão prévia das alterações que cada um dos diferentes estados emocionais provoca na qualidade de voz, quer em termos prosódicos, quer em termos articulatorios. No

estudo realizado por Campbell (2007), a principal preocupação era, exactamente, a relação existente entre qualidade de voz e a expressão de afectos. Utilizou para isso diversas frases indicando: compreensão; atenção; interesse. Estas expressões eram complementadas com risos e expressões enfáticas. Tiveram-se em conta, na análise, diferentes parâmetros acústicos: F0 (máximo, mínimo e mediana); intensidade; duração. O objectivo máximo não é definir porque variam os parâmetros, mas mostrar que, de facto, eles variam de acordo com o falante e com o interlocutor (grau de familiaridade e proximidade). Demonstrou que os parâmetros F0, energia, duração e qualidade de voz variam de acordo com factores paralinguísticos.

Os estudos existentes ajudam, não só em termos de investigação para emoção produzida por actores, como para patologias de índole psicológica e emoção espontânea, pois permitem identificar importantes relações entre variados parâmetros acústicos e atitudes e emoções (Johnstone e Scherer, 1999). Sabe-se que existem importantes efeitos da emoção na qualidade de voz e que os parâmetros acústicos que mais importa analisar nesta correlação incluem os estudos da Frequência Fundamental, energia e amplitude.

Algumas das emoções apresentam valores que as diferenciam mais das outras, como a alegria e a raiva se falarmos em termos de *jitter*. Uma característica geral será também o facto de as emoções que se caracterizam por ter valores de F0 e energia mais elevados apresentarem um tempo de fechamento da glote mais rápido. Este movimento rápido é sinal de um maior esforço vocal e de uma maior tensão muscular que a expressão dessas emoções implica (alegria, irritação, ansiedade). Vários autores consideram, no entanto, que apesar de todos os estudos que existem é ainda muito importante o estudo da emoção espontânea e dos mecanismos psicológicos nesta envolvidos e que condicionam a nossa qualidade de voz (Johnstone e Scherer, 1999; Bänziger e Scherer, 2003).

1.1.6 Falta de estudos para o PE

Face a escassez de estudos sobre qualidade de voz em Português Europeu, pretendemos que este estudo seja abrangente, contemplando emoção e patologia de índole psíquica; emoção simulada por actores e emoção espontânea. Assim, é importante referir que no nosso trabalho fazemos as análises com *corpora* formulados, seleccionados e recolhidos por nós, todos eles em Português Europeu.

Poucos estudos existem para o Português Europeu na área da emoção e, se tivermos em conta aqueles que existem, usam na maioria das vezes *corpora* em outras línguas, nomeadamente o Alemão. Como referimos anteriormente, o estudo realizado por Cabral e Oliveira, (2006), o EmoVoice, utilizou uma base de dados de emoções em Alemão. Portanto todo o estudo foi baseado na análise de outras línguas, para no final se tentar construir um sistema para o PE.

Um estudo pioneiro na área da caracterização da voz emocional em Português Europeu, foi o trabalho realizado por Américo Rodrigues, na sua tese de Mestrado em 2007. No entanto, apesar de já trazer resultados exclusivamente para o PE, porque o *corpus* é original, o estudo aborda apenas a emoção produzida por actores. Por outro lado ainda não são analisados tantos parâmetros, e os que se analisam estão relacionados com F0; duração, não estudando parâmetros mais relacionados com qualidade de voz como *jitter*, *shimmer* e HNR.

É nesta conjuntura, reconhecendo as limitações do nosso trabalho, que é transversal, e por isso mais abrangente, que compreendemos que a nossa análise seja uma primeira abordagem a várias áreas dentro da qualidade de voz e emoção no PE. Trata-se de um trabalho pioneiro, com muitos aspectos que podem e devem vir a ser evoluídos, até porque como já vimos são importantes em termos de robótica e tecnologia tanto em termos de saúde (em gerontologia, por exemplo) como em termos lúdicos (jogos de computador).

1.2. Objectivos do trabalho

O principal fundamento da nossa pesquisa é conseguir ter dados que nos permitam descrever, em termos quantitativos, de acordo com os parâmetros analisados, as emoções em Português Europeu. Com estes dados pretende-se caracterizar a emoção produzida por actores e a emoção espontânea, diferenciando-as, se os dados o permitirem. Por outro lado ainda, num estudo abrangente da voz e emoção em PE, analisamos um caso de disfonia psicogénica (em que o que está por trás de um problema de voz são razões emocionais). Procuramos verificar como as emoções afectam a nossa voz, utilizando *corpora* em Português, obtendo dados extraídos na e para a nossa língua.

Outra análise a realizar surge do interesse em saber se a emoção produzida por actores, até agora utilizada em tantas investigações e para tão diversas línguas, é ou não um meio de análise válido. Para isso, temos também a análise de emoção espontânea em PE que nos permitirá realizar uma comparação, com o *corpus* referido anteriormente. Procuramos verificar até que ponto convergem ou divergem os valores de cada uma das diferentes produções de emoção.

Conhecemos as limitações e dificuldades para realizar um estudo sobre emoção espontânea, mas acreditamos, no entanto, que ao contemplar, ainda que de forma limitada, um *corpus* de emoção espontânea, este poderá vir a servir de base para outros estudos e *corpora*, fornecendo referências em termos de valores e análises apresentadas. Por outro lado, através do estudo que fazemos poderemos tirar algumas conclusões sobre qual poderá ser a melhor forma para a recolha de *corpora* em termos de emoção espontânea, já que esta pode acarretar dificuldades éticas e até deontológicas.

Outro ponto de destaque no nosso trabalho será perceber até que ponto o conhecimento de uma língua e cultura poderá afectar a percepção das emoções por falantes de outras línguas e outras culturas. Procurar entender o que é que a Língua Portuguesa pode ter de diferente na expressão de emoções, onde se manifesta mais essa diferença e quais os parâmetros que nos ajudam a perceber essa distinção. Procura-se ainda perceber de que forma é que as possíveis diferenças podem caracterizar um Povo.

Outro objectivo é vir a disponibilizar *corpora* à comunidade científica, alargá-los e organizá-los de tal forma que possam vir a ser úteis em várias áreas, por exemplo, para a síntese de voz emocional e desenvolvimento de sistemas de reconhecimento de emoções.

1.3. Estrutura da Tese

O tema subjacente a todo este trabalho é Voz e Emoção em Português Europeu, tal como o indica o título da dissertação. Assim importa começar por descrever o processo de fonação, tudo o que pode interferir neste processo e quais os parâmetros que, normalmente, são tidos em conta na análise de voz.

Procuramos com este trabalho apresentar para o PE resultados que permitam caracterizar a voz emocional, seja esta emoção simulada, forçada por condições de origem psicogénica ou emoção espontânea. Neste âmbito, procuramos interligar todos os estudos que realizamos. Todos eles são desde o início e na sua essência, relacionáveis (todos versam sobre voz e emoção em PE).

Em relação a cada uma das partes do estudo procuramos fazer sempre uma resenha e contextualização, em termos científicos, dos estudos já realizados. Todo este enquadramento, mais teórico, e de fundamento básico para o nosso trabalho, é realizado no **capítulo 2**.

A análise, mais concreta, tem início no **capítulo 3** com a análise de uma patologia de base psicogénica. É, no nosso informante, diagnosticada uma disfonia psicogénica, sendo assim alvo do nosso estudo por se tratar de uma alteração na voz, provocada de forma inconsciente por força de um estado emocional. A patologia não aparece como algo de índole física, não é constatada qualquer anomalia em termos de tracto ou pregas vocais mas, toda a dificuldade que encontra em se expressar, é de foro emocional, daí que analisemos todos os parâmetros para saber o que muda e como muda a voz do nosso informante. Curiosamente este informante permitiu-nos ter dois registos de voz distintos. Estes seus registos (patológico e controlado) são comparados entre si e com o

registo de uma voz normal, para verificar a oscilação de valores nas análises realizadas.

No **capítulo 4** abordamos e estudamos a emoção produzida por actores. Como fundamento, não só tentámos saber se estes de facto se aproximam dos valores das emoções espontâneas, mas também fazer uma primeira caracterização das emoções em PE. Foi o que realizámos, percebendo que o nosso informante, actor do Teatro e Televisão Nacional, se aproxima, embora por vezes “forçando”, dos valores das emoções espontâneas. Temos apenas como informante um actor, o que pode ser algo limitativo. Contudo, a sua experiência em termos de trabalho parece-nos bastante rica e fundamental para a nossa análise. Ser um actor profissional é um factor importante, uma vez que tem experiência em registos diversos que implicam trabalhos de voz diferentes (publicidades; voz off; telenovelas e Teatro). Esta diversidade pareceu-nos benéfica, pois cada registo implicará uma colocação e qualidade de voz diferentes. Por outro lado, o facto ser um actor profissional acaba por dificultar o contacto com este, dada a carga horária do trabalho que realiza e a disponibilidade que era necessária para realizar o teste que lhe foi proposto. Também nem todos os actores se submetem a esta “prova”, sabendo que no final os resultados são por nós avaliados, em termos de qualidade de voz e produção de emoção.

Este capítulo é ainda reforçado com um teste perceptual. Este ajuda a verificar como eram as emoções simuladas pelo actor percebidas pelos ouvintes. Sabendo que há características defendidas como universais, procuramos identificar o que se pode distinguir e diferenciar das demais línguas no Português Europeu. O teste perceptual foi aplicado a dois grupos distintos, um grupo de alunos americanos, da Universidade da Louisiana, Lafayette, que nunca tinham estudado qualquer língua Românica e a um grupo de alunos da Escola de Saúde da Universidade de Aveiro, falantes nativos de PE. Este teste permitiu-nos aferir que, de facto, a cultura e conhecimento de uma língua pode influenciar o reconhecimento de certas emoções.

No seguimento do nosso estudo sobre emoção, importava estudar a emoção espontânea. Este é o tema do **capítulo 5**, a análise da emoção espontânea em PE. As emoções analisadas foram alegria, tristeza, raiva e o estado neutro, retiradas de programas dos 4 canais da televisão Nacional. Com o *corpus* que seleccionámos, conseguimos extrair os valores dos parâmetros utilizados para todas as análises realizadas ao longo desta dissertação, e pudemos obter resultados para o PE sobre emoção espontânea e ainda compará-los com os dados que tínhamos, do capítulo anterior, emoção por actores.

Este capítulo é complementado por um estudo utilizando o sistema *Feeltrace* que permite verificar de forma abrangente se o indivíduo percebe uma emoção como positiva ou negativa; activa ou passiva e, assim, pudemos passar a designar família de emoções como é defendido por Paul Ekman (1992).

À semelhança de cada um dos outros capítulos, terminamos com a análise e comparação de resultados, discussão e conclusão.

Para finalizar, são apresentados os principais resultados desta pesquisa e apontados novos caminhos e novas investigações a realizar neste campo científico, nomeada e principalmente, para o Português Europeu. Toda a pesquisa apresentada permite a criação e disponibilização de *corpora* numa área deficitária na nossa Língua.

1.4. Trabalhos publicados

No decorrer da realização desta dissertação foram várias as apresentações orais e em poster, acompanhando a pesquisa que ia sendo realizada. Obteve-se sempre uma boa resposta e interesse sobre o assunto nomeada e principalmente em conferências internacionais e em publicações. De referir que os estudos e análises estatísticas existentes nos estudos desta dissertação resultam, precisamente, de várias secções desta tese terem sido submetidas para revistas internacionais. Nesta análise estatística houve um ainda maior trabalho e empenho por parte do Orientador.

O primeiro estudo relativo a um caso específico de patologia de voz (patologia de foro emocional) que nos permitiu a aplicação de diferentes ferramentas (*Praat*; SPSS; *Hoarseness Diagram*) para a caracterização do corpus em análise foi apresentado em conferência nacional, Associação Portuguesa de Linguística, em poster, “*Qualidades de voz em disfonia psicogénica: estudo de um caso*” (Ana Margarida Nunes; Rosa Lima; Lurdes Castro Moutinho; Rosa Lúcia Coimbra; António Teixeira, 2005).

A continuação deste estudo, com mais material relativo a voz normal e voz com patologia foi aceite para apresentação oral em conferência internacional, na Croácia, (*International Clinical Phonetics and Linguistics Association*), “*Voice Quality with psychological origin: a case study*”, Ana Margarida Nunes; Rosa Lima; Lurdes Castro Moutinho; Rosa Lúcia Coimbra; António Teixeira, 2006. Este estudo foi aceite e encontra-se já publicado pela revista *Clinical Linguistics and Phonetics* com o título “*Voice Quality with psychological origin: a case study*”, (Nunes et. al. 2008) volume 22, Outubro de 2008, páginas 906-916.

O estudo perceptivo no âmbito da voz e emoção por actores foi apresentado oralmente no Congresso da *International Clinical Phonetics and Linguistics Association*, Istambul, 2008 “*Cross-linguistic effects on the perception of emotions*” Nunes, Ana Margarida; Roussel, Nancye [da University of Louisiana Lafayette; EUA]; Rodrigues, Américo; Coimbra, Rosa Lúcia; Teixeira, António, 2008.

Recebemos, a respeito deste trabalho, o convite do Doutor Krzysztof Izdebski, editor dos livros sobre “*Voice Emotions*”, para publicação. O artigo já foi enviado e aceite. À data da conclusão desta tese, o processo encontra-se parado, mas, a concretizar-se, a publicação fará parte de um dos volumes IV a VI.

A pesquisa realizada para a emoção no Português Europeu foi aceite para apresentação no formato de poster (também na *International Clinical Phonetics and Linguistics Association*, 2008, Istambul): “*Voice Quality of European Portuguese emotional speech*”, Ana Nunes; Américo Rodrigues; Rosa Lúcia Coimbra; António Teixeira, 2008.

Capítulo 2 - **Acerca de Voz e Emoção**

Neste capítulo apresentam-se informações de base sobre produção de voz, emoções, expressão das emoções pela voz e efeitos destas na voz, terminando-se com uma revisão de conhecimentos e estudos na combinação destes dois temas, voz e emoção.

O discurso emocional é um campo multi-disciplinar em termos de investigação com contribuições que podem vir de áreas como a Psicologia; a Acústica; Ciências da Linguagem; Medicina; Linguística; Processamento de Sinal e Engenharia Computacional. A cooperação entre todas estas áreas poderá trazer franca convergência e consolidação em termos de investigação de Voz e Emoção, o que permitirá um melhor entendimento de como as emoções são reveladas através de vários aspectos da voz (Patrik, 2008).

2.1. Voz

A voz pode ser definida como o som produzido pela vibração das pregas vocais, ou globalmente como sinónimo de fala. Detalhes de fonação e articulação, *pitch* e variações de amplitude e ainda parâmetros temporais, todos contribuem

para a diferente voz de cada falante. Todas as definições de voz reflectem este facto (Kreiman et al, 2004).

Podemos definir e explicar a voz, segundo Guimarães (2007), de uma forma clara, como o som audível, resultado de uma interacção entre pressão e velocidade do fluxo do ar que expiramos. Este fenómeno actua sobre a intensidade dos vários movimentos de adução e abdução das pregas vocais. De um ponto de vista mais objectivo, voz é o produto sonoro da actividade laríngea, veiculando muito mais do que mera informação semântica.

A produção de voz é um processo que envolve três partes: pulmões, cordas vocais e tracto vocal. A função dos pulmões é ser a fonte de ar e pressão. Nos sons vozeados as pregas vocais abrem e fecham periodicamente e dividem a passagem do ar vindo dos pulmões, o que funciona como excitação acústica e produção de fala. O tracto vocal é um conjunto de cavidades acima das pregas vocais. Funciona como filtro acústico que permite, através da sua forma, obter os diferentes sons. Finalmente, o som é propagado através dos lábios e das narinas. Para além disso, a voz humana também tem sons não vozeados, tais como as fricativas surdas, durante os quais as pregas vocais não vibram.

Analisando a voz do informante, podemos ter acesso a alguns dados (alguns mais óbvios que outros) como sexo, idade, saúde e estado emocional do indivíduo (Spreckelmeyer et al. 2009). Torna-se uma marca de personalidade, que permite reconhecer cada indivíduo, sendo, por isso, importante em todos os contextos interpessoais e sociais. É relevante saber o que veicula a nossa voz (como é percebida), o que ela diz de nós (da nossa identidade) e como se modifica (emocionalmente, por exemplo). Este tipo de estudos apresenta resultados de relevo não só em termos de conhecimento da fisiologia geral, como em termos linguísticos, na explicação da produção de sons, na terapia da fala, sem deixar de estar intimamente relacionado com a neurologia.

2.1.1 Estruturas anatómicas envolvidas na produção de voz

Toda a geometria do tracto vocal desde a glote (pregas vocais) até aos lábios é um factor determinante de todo o processo de fala do indivíduo, influenciando as características de uma melhor ou pior articulação (Schröder, 1967). Explicar ou medir o tracto vocal, não é uma tarefa fácil, uma vez que a sua forma se modifica ao longo do tempo do discurso.

Aplicando métodos dinâmicos, em que se analisa o tracto vocal, durante a fonação (produção de frases), conseguem-se resultados que em muito vêm ajudar na definição e redefinição dos modelos articulatórios. Estes resultados são facilmente passados para computador, conseguindo-se perceber tudo o que acontece em termos de tracto vocal quando se dá a fonação, resultados que podem ser verificados, num computador, em simples imagens ou em figuras em movimento (Schröder, 1967).

2.1.1.1 *O tracto vocal*

O tracto vocal é constituído pela faringe, boca, cavidades nasais e paranasais. Estas são as verdadeiras cavidades de ressonância da voz e têm forma de F. Têm uma importância capital na qualidade final do som emitido. A laringe, onde se encontram as pregas vocais, apresenta características particulares.

2.1.1.2 A laringe

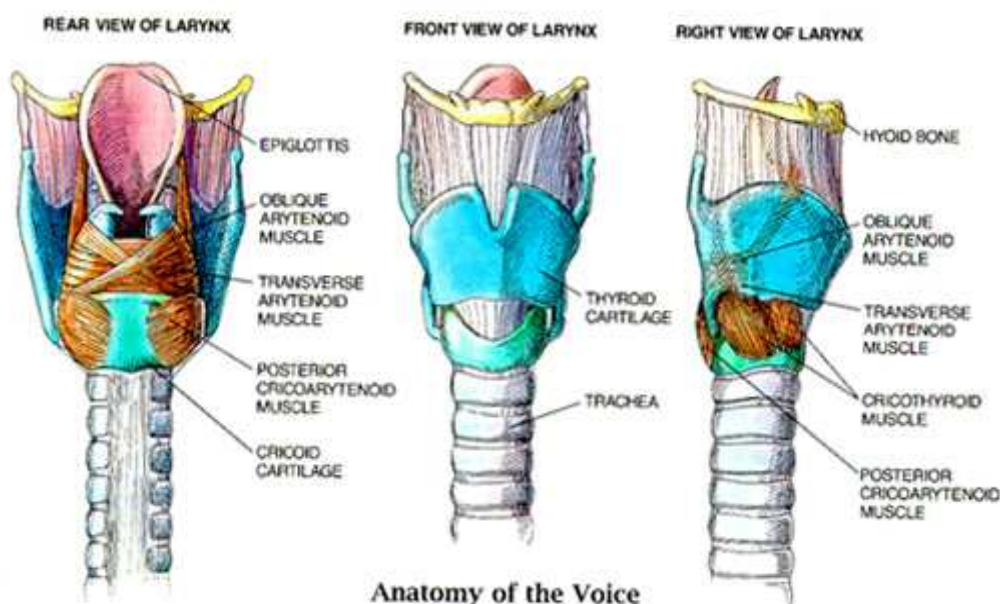


Figura 1 – Anatomia da laringe. Fonte: University of Pittsburgh Voice Center

A laringe (Figura 1) é constituída por quatro componentes anatómicos básicos: cartilagem; músculos intrínsecos e extrínsecos e mucosa. À superfície é do tipo escamoso, bastante fino e transparente. É na zona da cartilagem que encontramos as pregas vocais, constituídas por ligamentos e músculos cobertos por uma membrana mucosa. Estão ligadas à cricóide, cartilagem da tiróide (maçã de Adão nos homens, por exemplo) e às cartilagens aritenóides. Todas estas cartilagens estão ligadas a outras estruturas da cabeça e pescoço, através de músculos extrínsecos. Os músculos intrínsecos da laringe alteram a posição; forma e tensão das pregas vocais.

Durante o processo de respiração as pregas vocais permanecem afastadas, apresentando a forma de um “V”, permitindo a passagem do ar para os pulmões. No processo de fonação as pregas aproximam-se, fechando completamente a passagem do ar. Durante a expiração, a grande quantidade de ar que vem dos pulmões, encontra a pressão causada pelo fechamento das pregas vocais. Esta pressão, quando atinge um certo nível mais elevado, consegue ultrapassar essa resistência e é libertada uma pequena quantidade de ar. Como a pressão

decrece, as pregas vocais fecham novamente e um novo ciclo começa, são estas breves passagens de ar que produzem as vibrações que podemos ouvir como sons, captados pelo ouvinte e interpretados no cérebro.

A Figura 2 mostra todas as componentes e ligações envolvidas no processo de fonação.

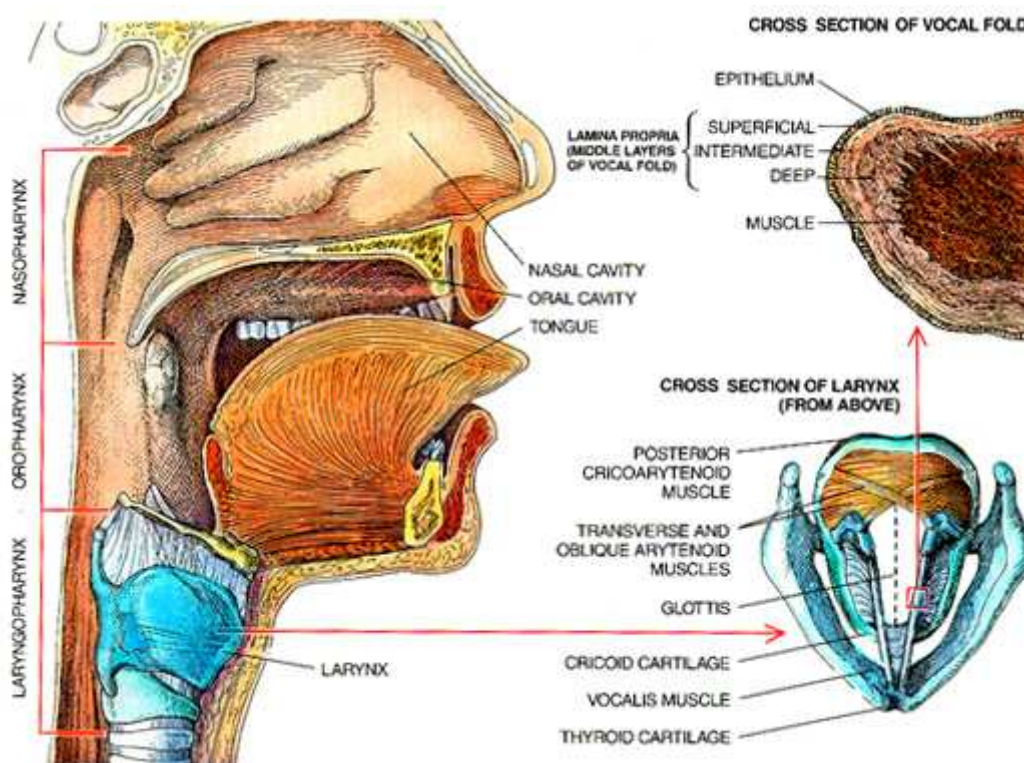


Figura 2 – Todos os elementos envolvidos no processo de produção de voz. Fonte: *Tutorials – Voice Production. National Center for Voice and Speech.*

As cavidades nasais e paranasais, apesar de terem sido, durante muito tempo, desconsideradas como cavidades de ressonância, são agora consideradas a partir de estudos que demonstraram sua eficácia na ressonância, principalmente de sons graves. Numa visão geral, podemos descrever o aparelho fonador como composto por uma fonte geradora de pressão aérea, uma região de modificação de energia, uma região de modulação sonora e uma de articulação do som. A região geradora de pressão aérea é composta pelo tórax, pulmões e pela musculatura abdominal. Este grupo de estruturas tem como função gerar força suficiente para mobilizar o ar contido nos pulmões em direcção à laringe. Há

neste processo dois grupos musculares envolvidos, os intercostais internos (inseridos nas costelas), participantes da expiração, e os intercostais externos, responsáveis pela inspiração (Gilvaneia et al. s.d.).

A Tabela 3 apresenta de forma sumária as fases descritas.

Tabela 3 – As quatro fases de esforço muscular, no ciclo respiratório. Retirado de: *Tutorials – Voice Production. National Center for Voice and Speech.*

Fase	Músculos envolvidos
Inspiração	Diafragma e/ou intercostais externos
Expiração 1	Nos processos de contracção e distensão, o diafragma é usado para impedir a pressão adicional da inspiração
Expiração 2	Os processos de contracção e distensão intercostais internos, continuam até não haver mais expiração
Expiração 3	Intercostais internos e abdominais

Existem diferenças laríngeas entre homens e mulheres (em termos anatómicos) que se manifestam ao nível da cartilagem tiróide, das pregas vocais e da largura glótica, sendo que a posição da laringe é semelhante em ambos os géneros. A cartilagem tiróide é 20% mais larga nas mulheres. No sexo masculino, por comparação com o feminino, a cartilagem tiróide é 40% superior, as pregas vocais são mais compridas, densas e largas, com uma porção membranosa 60% mais longa. A largura glótica, em repouso, é 5% mais larga sendo a largura máxima 30% maior (Guimarães, 2007).

Todas estas diferenças causam as variedades acústicas a que todos somos sensíveis em termos de voz e qualidade de voz.

2.1.2 Produção

Sabendo já que a voz é de facto um som audível que transporta informação (através não só da fala, mas do canto e de sons ligados a emoções como o riso, por exemplo), importa saber como e onde se dá este processo de geração de som (fonação).

2.1.2.1 *A principal fonte de som*

A laringe constitui a principal fonte de som, na produção da fala, localizando-se na região anterior do pescoço, na extremidade superior da traqueia, abaixo da faringe, boca e nariz (Zemlin, 2000 e Harris, 2007), formando um tubo que na garganta possibilita a passagem do ar para os pulmões e destes para o exterior. É, precisamente, todo este sistema músculo-cartilágíneo e neuromuscular que tem como funções a protecção das vias aéreas inferiores; a respiração, a deglutição e a comunicação oral. A forma e comprimento de toda esta cavidade pode ser transformada por todo este sistema muscular para conseguir produzir diferentes sons. Fazendo um paralelismo que torna este funcionamento mais claro de entender, esta cavidade modifica a ressonância funcionando como o violino ou a guitarra; a diferença é que as pregas vocais não são fixas, e mudam facilmente para a amplificação e produção de uma grande variedade de sons. Os movimentos da língua e palato mole permitem transformar os sons que reconhecemos como vogais ou consoantes no nosso discurso (Harris, 2007). Todo o resultado final da nossa voz acontece na boca, nariz e garganta, onde a língua, palato e lábios também se encontram envolvidos.

A Figura 3 mostra todas as estruturas que se encontram envolvidas no processo de fonação.

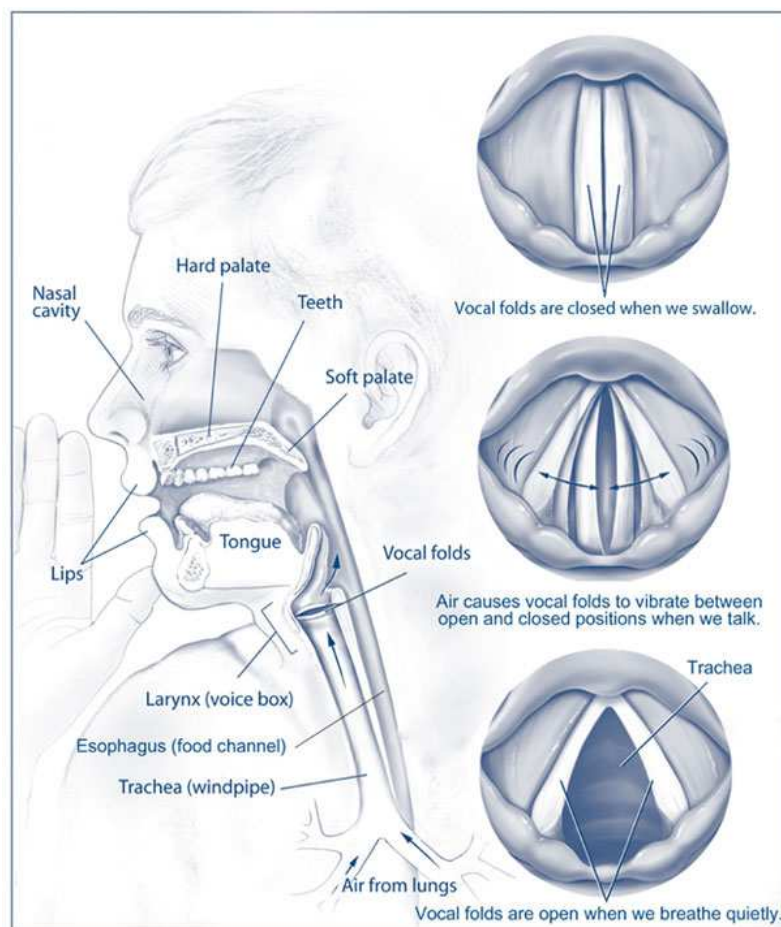


Figura 3 – Estruturas envolvidas na fala e produção de voz. Fonte: *National Institute on Deafness and Other Communication Disorders (NIDCD)*, 2008.

Assim, o ar que vem dos pulmões, passa pelas pregas vocais, possibilitando a sua vibração e vai produzir som que pode ser diferenciado consoante a posição da língua, palato e lábios (Harris, 2007). As pregas vocais, descritas como um tipo de músculo muito flexível, têm então aqui um papel fundamental na produção do som. Ao vibrarem com a passagem do ar, causam vibrações acústicas. Se a forma do tracto vocal controla a prosódia, as cordas vocais controlam o *pitch* da voz (definindo-se este parâmetro como a percepção do som), de acordo com Baken e Orlikoff (2000) é o atributo perceptual do som, apresentado geralmente numa escala que abrange dos sons mais altos aos mais baixos de uma forma contínua. Todas as possíveis alterações contribuem para caracterizar a voz do

indivíduo e as modificações que possam existir no tracto vocal contribuem para as diferentes produções de fones (Sato, 2004).

2.1.2.2 A fonação

Na produção da fala, a fonação inicia-se com um ciclo vibratório, com as cordas vocais na zona mediana, começando ligeiramente a afastar-se no sentido de baixo para cima. Quando uma determinada abertura máxima é atingida, a pressão subglótica diminui e a elasticidade das pregas vocais, conjuntamente com a diminuição da pressão na glote, conduzem ao efeito de sucção das pregas vocais que é compensado pela resistência elástica das mesmas, levando-as a deslocarem-se para a posição normal. A equação que descreve este efeito é conhecida como a lei de Bernoulli – O aumento da velocidade provoca uma diminuição na pressão.

Há uma estimulação do movimento de adução das pregas vocais, no sentido de baixo para cima o que provoca o fechamento da glote (Figura 5). O final da fase de encerramento acontece quando se inicia um novo ciclo vibratório, ou seja, com um novo aumento da pressão subglótica ao ponto de forçar a resistência das pregas vocais (Guimarães, 2007).



Figura 4 – Imagem da abertura e fecho da laringe (à direita e esquerda respectivamente). Fonte: Sara Harris, *Specialist Speech and Language Therapist and Team member, The Voice Clinic*, London, 2007.

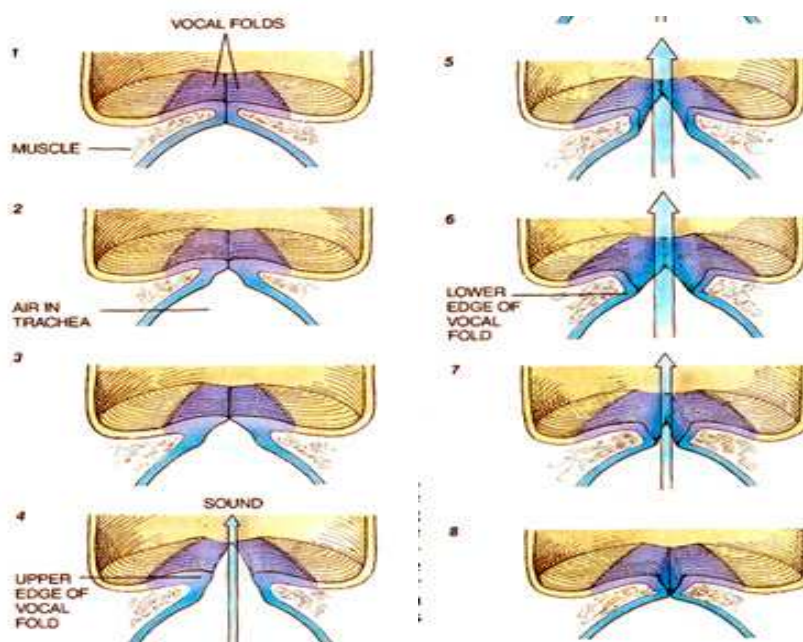


Figura 5 – Processo de vibração das cordas vocais. Fonte: University of Pittsburgh Voice Center.

2.1.2.3 *Controlo da frequência de vibração*

O comprimento, tensão e massa das pregas vocais determina a sua frequência vibratória. Se o comprimento das pregas vocais aumenta, a frequência e a tensão também aumentam. Quando a massa aumenta, a frequência, por sua vez, diminui. Os mecanismos de variação de intensidade podem ser semelhantes aos da variação da frequência, abrangendo a actividade muscular, fluxos e pressões de ar, podendo ainda ter a ajuda das estruturas de ressonância. Saliente-se ainda a influência da resistência glótica no controlo da intensidade: quanto maior for a resistência, maior será a pressão subglótica e, como consequência, maior a intensidade (Colton e Casper, 1996). O movimento lateral das pregas vocais continua até que a sua elasticidade natural as acabe por estabilizar, assim as pregas vocais retornam à sua posição original (fechadas). De seguida, o ciclo repete-se e cada ciclo produz uma pequena passagem de ar. A voz humana nada mais é do que centenas destas pequenas libertações de ar, passando a cada segundo e sendo filtradas pela cavidade vocal (Titze, 1998). Podemos observar um desses ciclos na Figura 6.

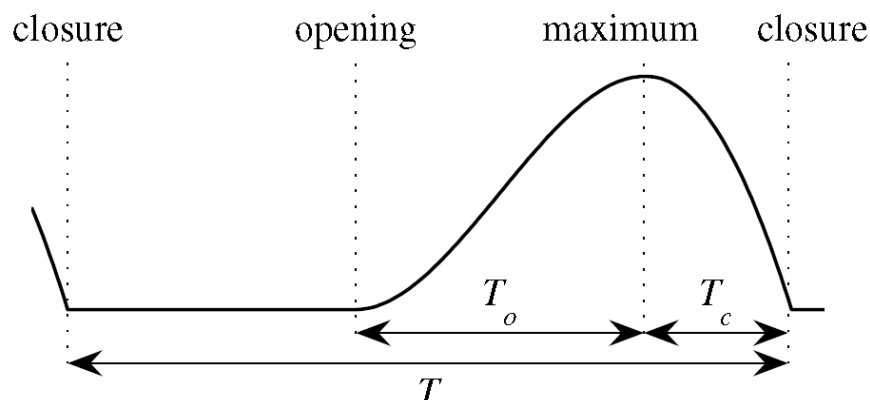


Figura 6 – Ciclo de abertura e fechamento das pregas vocais, ciclo glotal (Titze, 1998).

O tamanho e tensão das pregas vocais podem, também, ser controlados pela acção dos músculos que regulam a frequência da oscilação e a qualidade de voz. O tamanho das pregas vocais que vibram num indivíduo adulto masculino tem cerca de 16 milímetros e perto de 10 milímetros num adulto do sexo feminino. As cordas vocais podem ainda ser esticadas alguns milímetros através dos músculos da laringe (Laukkanen e Leino, 1999, citados por Pullaka, 2005).

A oscilação não é mais do que o movimento de vai e vem que se vai repetindo. São vários os aparelhos que vão oscilar (vibrar) durante um certo período de tempo se os manipularmos (sinos ou, por exemplo, copos de cristal). O que faz as pregas vocais oscilar é desta forma mais curioso, por depender de um fenómeno físico que pode ser sustido durante algum tempo (Ingo Titze, 1998).

A vibração das cordas vocais começou por ser explicada de uma forma muito simples e por isso também incompleta. À medida que cada vez mais se foi podendo estudar o fenómeno ligado à fonação humana, vários outros modelos foram desenvolvidos apresentando maior complexidade. Nas últimas décadas os cientistas procuraram incrementar esta investigação, utilizando todos os benefícios das novas tecnologias: computadores e sistemas de imitar a oscilação das pregas vocais (Titze, 1998).

Todo o movimento das pregas vocais é aerodinâmico, pois depende do movimento do ar e das forças de outros corpos em movimento, nomeadamente no processo de fonação, das pregas vocais. Depende ainda da elasticidade dos músculos que as compõem e sobre os quais temos um controlo bastante activo.

O aumento do uso da estroboscopia na investigação clínica e investigação em laboratórios permitiu observar de variadíssimas formas o movimento das pregas vocais. Foram estas observações que mostraram que as pregas vocais raramente se movem de forma uniforme e em conjunto, podemos ver um exemplo na Figura 7. Elas movem-se num movimento ondulatório de baixo para cima.

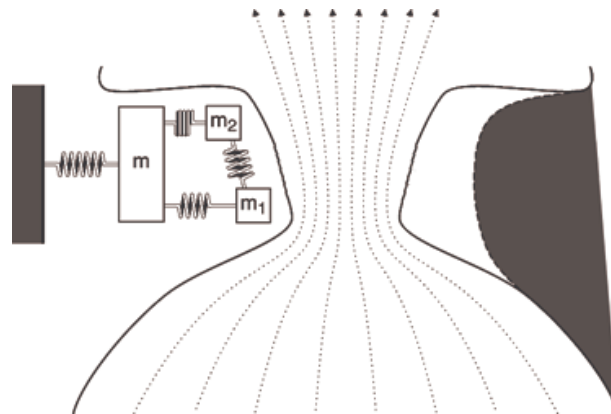


Figura 7 – Movimento das pregas vocais.

As pregas vocais estão mais afastadas na zona inferior. Esta forma é chamada de convergente uma vez que parece que a passagem do ar as faz juntar-se. Por outro lado, quando também o ar diverge mais, será a parte de baixo das pregas vocais que se encontrará mais fechada (forma glotal divergente). A média da passagem de ar tende a ser maior quando existe uma configuração glotal convergente, o que resulta numa diferenciação na passagem do ar, necessária para, por vezes, sustentar essa oscilação (Titze, 1998).

A forma da onda glotal ajuda a determinar a altura do som produzido e também o timbre de voz. Logo, uma onda caracterizada por um Quociente de Abertura (*Open Quotient* - OQ – que é a medida de tempo durante o qual as pregas vocais estão abertas) – de muito curta duração representa uma mudança súbita na passagem do ar, o que vai levar a registos de frequências mais

elevadas. Uma curva mais suave representa uma passagem do ar bastante gradual (período de OQ mais longo) (Titze, 1998).

Para além dos sons normais e desejados que são produzidos pela laringe, são também frequentes algumas produções que apresentam turbulência (em que não há uma passagem do ar de forma suave), outras mais aspiradas, quando se combina a passagem do ar com a vibração das pregas vocais. É de facto, a aspiração um dos factores que contribui para as vozes, aspiradas (*breathy*).

Outros sons que conseguimos produzir e que dizem respeito a *clicks* ou *stops*, são resultado de um maior esforço e de um mecanismo de abertura ou fechamento abrupto das cordas vocais. Para além de tudo isto qualquer presença de outros fluidos ou mucos junto das pregas vocais origina também outros tipos de produção: uma voz mais rouca, “*a rough or 'gurgly' noise*”. Estas diferenças e características de produção/fonação levam-nos a distinguir e falar de qualidade de voz (Titze, 1998).

2.1.3 Qualidade de Voz

2.1.3.1 Definição

Qualidade de voz é, *per si*, um conceito de difícil definição precisamente por incluir questões diversas como timbre, medida do esforço vocal, variedades linguísticas (diatópicas, diafásicas, diastráticas e diacrónicas) e sua relação com factores linguísticos e supra-segmentais, como a prosódia. O conhecimento das diferenças e características de diversas qualidades de voz torna-se necessário ao sabermos que, em algumas línguas, qualidade de voz funciona como um traço distintivo adicional ao nível fonológico. A aspiração e tensão na voz são usadas para veicular informação socialmente e de forma paralinguística, tal como o são as variações de Frequência Fundamental (F0 – corresponde ao número de ciclos que as pregas vocais fazem por segundo (Braga et al. 2009)), amplitude e variações linguísticas.

A qualidade de voz é a característica da voz percebida auditivamente e que diferencia cada uma das vozes de cada um dos indivíduos. Deriva de uma variedade de comportamentos laríngeos e supra-laríngeos. O tom natural e distintivo produzido por uma pessoa gera uma voz única e particular.

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 da voz. Os aspectos impressionistas estão ligados à percepção e audição da voz, o ouvinte é que descreve, de acordo com o que ouve, os diferentes tipos de voz. No aspecto fonético englobam-se todos os aspectos relacionados com a sua produção, anatomia e fisiologia. De facto, estes aspectos são os relevantes na articulação que leva à produção de diferentes qualidades de voz (<http://www.ims.uni-stuttgart.de/phonetik/EGG/page8>).

Na literatura linguística sobre qualidade de voz, esta é geralmente vista sob duas perspectivas: **Fonética** ou **Não-Fonética**.

Qualidade de Voz na perspectiva Fonética tem uma função contrastiva no sistema fonológico da língua. Em muitas línguas um contraste entre certos segmentos realiza-se, precisamente, por uma diferença articulatória. Por exemplo, as vogais aspiradas constituem oposições fonológicas no Gujarati e nas oclusivas em Igbo (Ladefoged e Maddieson, 1996:47, 304).

Na perspectiva **Não-Fonética** a qualidade de voz é vista como tendo uma função interessante, embora ainda pouco estudada, ligando-se à forma como a qualidade de voz é percebida, de forma inconsciente, por cada indivíduo. Independentemente do que uma pessoa diz, isso pode ser percebido como simpático, curioso ou sem interesse e todas estas diferentes percepções de um mesmo enunciado estão baseadas na percepção da qualidade de voz (Helmholtz (1863) citado em <http://www.ims.uni-stuttgart.de/phonetik/EGG/page8>).

Todos seremos mais ou menos capazes de produzir diferentes qualidades de voz, sem que haja qualquer tipo de patologia, consoante a nossa vontade e

esforço. Esta variação poderá relacionar-se também com o contexto social/cultural em que nos encontramos (Campbell, Mokhtari, 2003). O que muitas vezes parece um problema incontrolável, parecendo uma patologia de voz para alguns, é para o outro indivíduo um contraste fonológico necessário na sua língua (Gordon e Ladefoged, 2001).

2.1.3.2 *As várias dimensões da qualidade de voz*

Foi o final do século XX, inícios do século XXI, que trouxe 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.

As diferenças no fechamento da fonte glotal é que vêm fazer a distinção entre as várias qualidades de voz, distinguindo-as entre: modal, aspirada, sussurrada, áspera. A qualidade de voz depende ainda do grau de tensão da laringe e faringe, como já anteriormente referimos.

A 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 considerado 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.

Uma questão 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 cordas/pregas vocais são parâmetros que em termos de qualidade de voz e patologia devem ser tidos sempre em linha de 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 espectograma;

- “*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.

Apresentamos, na Tabela 4, uma parte da descrição das características fonéticas de cada qualidade de voz, de acordo com Laver, 1991, p.227.

Tabela 4 – Características fonéticas de cada qualidade de voz

Ajustes Supralaríngeos	Comportamento da laringe
Eixo longitudinal: Labial Protrusão labial Labiodentalização Laríngeo Aumento da laringe	Tipos de fonação: Voz Modal Falsete Sussurro Voz áspera
Eixo latitudinal: Labial Fechamento arredondado Abertura arredondada Abertura labial Articulação retroflexa Corpo da língua Dentalização Velarização	Tipos compostos: Voz sussurrada Falsete sussurrado Voz estridente Falsete estridente Voz aspirada Voz áspera Falseado áspero Voz áspera sussurrada
Ajustes velofaríngeos: Nasal Não nasal	Comportamento da tensão muscular: Voz tensa Voz relaxada

2.1.3.3 Tipos de Fonação

Qualquer voz se caracteriza pelas mudanças de tensão e compressão que sofrem as pregas vocais. Num dos limites encontra-se a voz tensa, associada a valores elevados de tensão, a uma compressão medial e longitudinal das cordas vocais. É uma qualidade de voz muitas vezes associada a mudanças abruptas, a um aumento da amplitude na frequência dos harmónicos mais altos. A voz mais relaxada, calma é exactamente o oposto e está associada a uma atenuação da frequência mais alta dos harmónicos.

Ladefoged sugere que existe um *continuum* de tipos de fonação, que são definidos em termos de abertura das cartilagens aritenóides, e assim os tipos de voz vão do extremo não vozeado (mais aberto) até ao fechamento glotal (mais fechado). As diferenças nos tipos de fonação podem ser assinaladas por um grande número de propriedades fonéticas de acústica, aerodinâmicas e domínios articulatorios. Estes tipos de qualidade de voz/fonação, mostram que podemos usar a glote de formas totalmente inesperadas sem que isso signifique a presença de qualquer tipo de patologia. Existem, neste contexto, variados tipos de fonação de acordo com a vibração das pregas vocais e seus aspectos aerodinâmicos.

Uma **fonação neutra**, ou ao que passaremos a denominar como voz **modal**, apresenta uma vibração periódica das cordas vocais, com um completo fechamento da glote. Em termos de frequência fundamental (designada por F0) e reportando-nos ao homem uma voz modal apresenta uma média perto dos 120 Hz enquanto que para a mulher se aproxima dos 220 Hz. A principal característica deste tipo de fonação é o contacto das pregas vocais. A vogal modal, por exemplo, encontra-se num nível intermédio entre a estridente e a aspirada. Esta vogal é intermédia, caracterizada por uma amplitude muito idêntica, tanto para a frequência fundamental, como para o primeiro formante. A Figura 8 mostra-nos o sinal sonoro de uma voz modal, masculina.

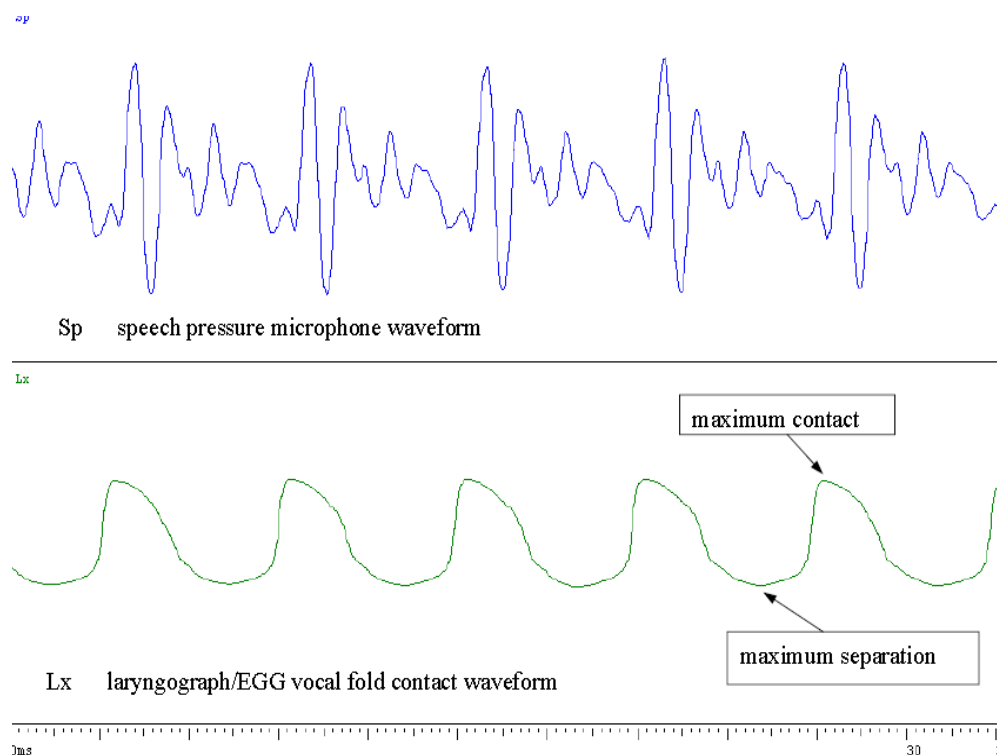


Figura 8 – Representação da onda da voz modal Masculina. Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive.

Uma fonação estridente¹, denominada, de acordo com Laver (1994), por “creak phonation (also called vocal fry)” é utilizada em várias línguas na produção de oclusivas (Maddieson e Ladefoged, 1996: 53), contrastando com a voz modal em muitas línguas. A voz estridente caracteriza-se por uma fonação irregular, com períodos de pitch espaçados e com uma intensidade acústica que vai diminuindo. As vogais estridentes estão associadas a vibrações irregulares e com ruído, sendo os valores do primeiro e segundo formantes mais altos, ficando-se isto a dever à constrição da faringe. Este tipo de fonação está, normalmente, associado a pregas vocais que estão apertadas mas que abrem o suficiente, durante um período de tempo, permitindo a fonação (Ladefoged 1971, Laver 1980, Ní Chasaide and Gobl 1995). Existe vibração irregular e lenta e a pressão subglotal e glotal são mais baixas do que na fonação modal (Hirose, 1996).

Na vogal estridente, a amplitude do segundo harmónico é ligeiramente maior do que a amplitude da frequência fundamental. Nestas vogais, o harmónico mais junto ao primeiro formante tem uma amplitude muito maior do que a da fundamental. A Figura 9 representa um exemplo de sinal sonoro e informação sobre o funcionamento das cordas vocais para uma voz estridente para um homem adulto.

¹ Creaky voice é também traduzida como voz rangida no Português do Brasil, de acordo com Eleanora Albano.

<http://periodicos.ufsc.br/index.php/desterro/view/8937/8282>

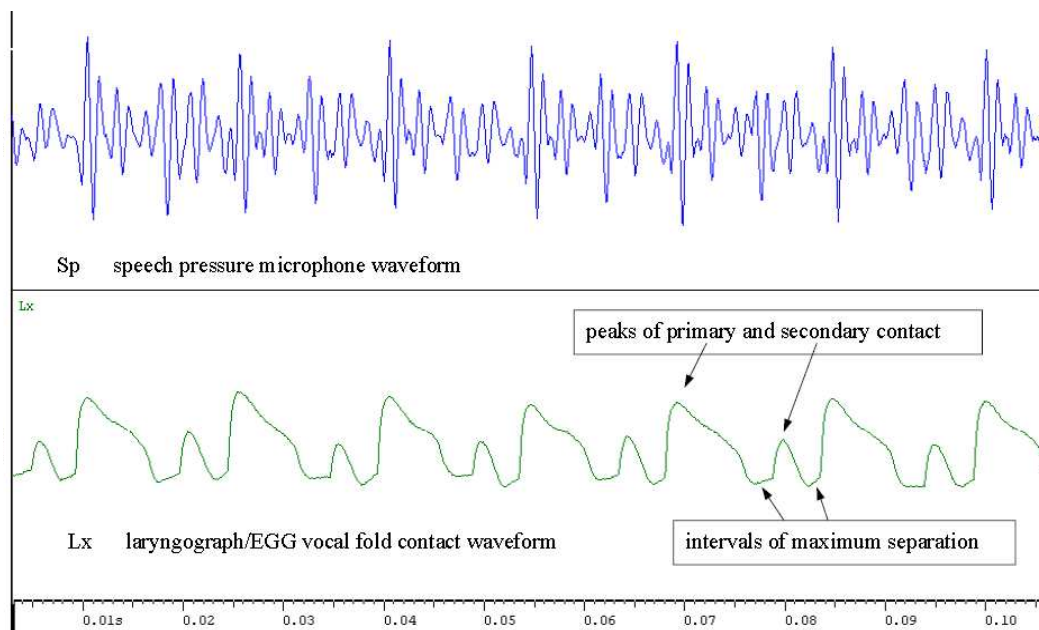


Figura 9 – Representação da onda da qualidade de voz estridente para o homem.
Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive.

Uma **voz aspirada** terá como principais características em termos de fonação a pouca tensão que existe nas cordas vocais e um incompleto fechamento da glote, o que vem permitir uma maior e contínua passagem de ar, uma voz misturada com a respiração, ar. Assim, a voz aspirada é marcada por um aumento do ruído, principalmente nas frequências mais altas. A frequência fundamental encontra-se abaixo dos valores apresentados para a voz modal. Em muitas línguas, este tipo de fonação aparece ligado a uma descida generalizada dos parâmetros acústicos. Nas nasais com voz aspirada existe muito ruído, enquanto que na voz modal este ruído não acontece de forma tão evidente. Algumas línguas fazem esta distinção/contraste mais nas vogais do que nas consoantes.

Outra característica das vogais vozeadas aspiradas é a turbulência na energia o que faz com que seja difícil diferenciar e dividir os *pitch*. Nas vogais vozeadas com voz estridente, esta característica é mais pronunciada a meio da vogal, resultando em frequências fundamentais mais baixas. Segundo os autores, estas diferenças nos três tipos de fonação (modal, aspirada e estridente) variam

de acordo com a pessoa e evidenciam que o gênero feminino utiliza muito mais a voz aspirada.

Tanto as vogais aspiradas como as estridentes caracterizam-se pela diminuição da energia. Nas ondas, apresentam também uma frequência fundamental inferior, relativamente à voz modal. Como frequência, de acordo com Baken e Orlikoff (2000), entende-se a característica física de alguns sinais, um parâmetro físico que diz respeito à vibração, normalmente a repetição de uma mesma forma da onda. Tanto na voz estridente como na aspirada é notória uma redução das características relativas à intensidade, em relação à voz modal, o que pode ser verificado na representação do sinal sonoro e funcionamento das pregas vocais (Figura 10). Um dos parâmetros acústicos que diferencia, em muitas línguas, os tipos de fonação é a inclinação do espectro, ou seja, o grau a que a intensidade desce à medida que a frequência aumenta.

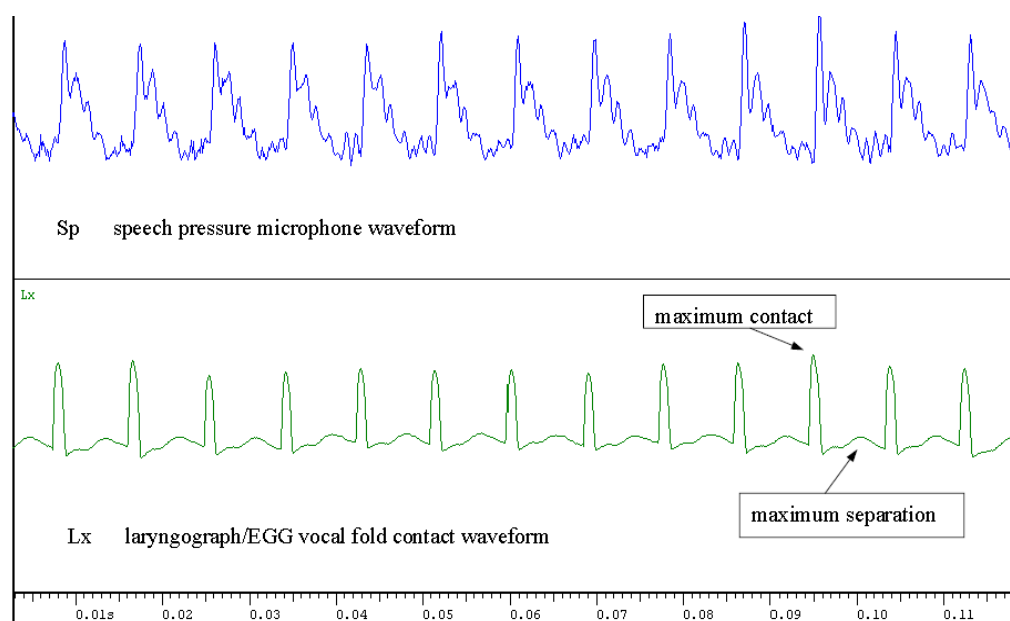


Figura 10 – Representação da onda da qualidade de voz aspirada Masculina. Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive.

Após a descrição e comparação dos tipos de fonação que podemos considerar mais comuns, até porque são traços distintivos importantes em certas línguas, pode-se ainda falar da **voz sussurrada (*whisper*)**. Nesta produção existe uma fraca compressão das pregas vocais e um esforço de vozeamento muito

pequeno. Laver (1980) considera que não existe uma diferença clara entre voz aspirada e sussurrada.

Uma voz mais **áspera**, e que é denominada em termos de qualidade de voz como *harsh* (Cabral, 2006), apresenta uma forte tensão ao nível das pregas vocais, o que resulta numa grande aproximação entre as pregas vocais. Toda a laringe é submetida a uma tensão bastante elevada, “*upper larynx becomes highly constricted with the ventricular folds pressing on the upper surfaces of the vocal folds*” tornando a vibração ineficaz (Hirose, 1996). Este tipo de fonação é bastante irregular e a frequência fundamental apresenta valores abaixo dos 100 Hz.

Falsete, outro tipo de voz, apresenta uma fonação mais elevada do que a voz modal. As pregas vocais encontram-se comprimidas longitudinalmente, tornado-se assim muito finas. A vibração é assim menor e a voz apresenta um tom bastante elevado (Figura 11). A glote permanece ligeiramente aberta, havendo pouca pressão subglotal e existe um ruído constante (Hirose, 1996). Uma voz **tensa** será comparável a uma voz áspera, com um pequeno decréscimo na tensão da laringe. Contudo, o termo *tense voice* também é usado na descrição de uma tensão mais elevada ao longo de todo o tracto vocal, (Ní Chasaide e Gobl, 1997, p.451).

Importa ainda referir que estes tipos de fonação não ocorrem isolados, isto é, muitos deles acontecem em simultâneo ao longo do nosso discurso, trabalham em “conjunto” modificando os tipos de fonação. Apenas voz modal e falsete são incompatíveis, dada a utilização diferente da laringe.

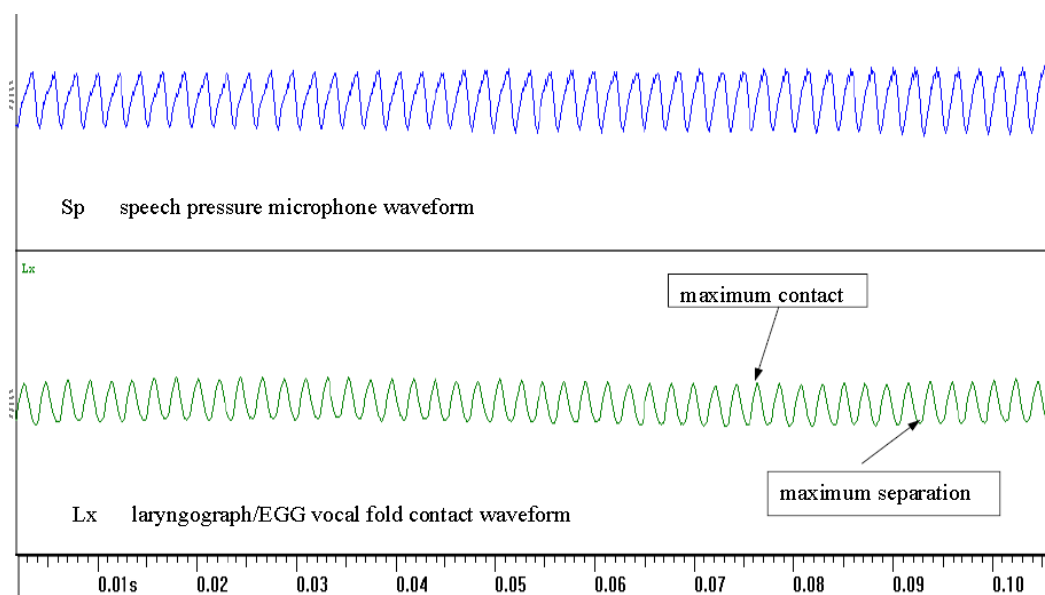


Figura 11 – Representação da onda da voz falsete masculina. Fonte: Apresentação relativa a Drioli et al. (2003), disponível no ISCA Archive.

2.1.3.4 Variação da Qualidade de voz

São todas as variações da voz de acordo com o indivíduo, contexto e até momentos do discurso que, de acordo com Abelin e Allwood (2003) e Cowie (2003), permitem diferenciar e identificar atitudes e emoções (alegria, tristeza, raiva, etc). Isto leva a alterações prosódicas (Campbel e Mokhtari, 2003): as pausas no discurso, a duração e acentuação de sílabas e parâmetros extralinguísticos relacionados com interjeições e alteração da respiração, linguagem corporal e expressão facial (Schröder, 2004). Estas variações da Qualidade de Voz estão essencialmente relacionadas com as alterações dos parâmetros da fonte glotal tais como a variação do F0 e o contorno do *jitter* (Gobl e Chasaide, 2003). Com base nestas características identificam-se (Laver, 1980) várias qualidades de voz: *modal*, *tense*, *harsh*, *creaky*, *breathy* e *whisper*, ou, como passaremos a designar: *modal*, *tensa*, *rouca*, *estridente*, *aspirada* e *sussurrada*, respectivamente. A qualidade de voz deveria ainda ser considerada o quarto parâmetro da prosódia, juntamente com “*pitch; power, and duration*” (Nick Campbell e Parham Mokhtari, 2003).

A qualidade de voz varia ainda de acordo com o **género**. Assim, de acordo com o estudo de Mendonza et al. (1996), existem diferenças significantes na distribuição da energia. Estas diferenças estão aparentemente relacionadas com a presença de maior barulho de passagem de ar na voz das mulheres. Esta diferença faz com que as vozes femininas quando comparadas às vozes masculinas sejam muito mais aspiradas.

Como referimos, anteriormente, a qualidade de voz varia consoante o **contexto** em que falamos, a preocupação com que nos expressamos ou dirigimos a alguém, mesmo que não estejamos/sejamos sensíveis a essa mudança. São possíveis de verificar, inclusive, diferentes entoações consoante a relação que temos com a pessoa à qual nos dirigimos (a qualidade de voz e inerentes aspectos prosódicos serão diferentes se falar para um filho, para um amigo ou para um desconhecido, por exemplo).

Quando falamos com crianças, a nossa voz é mais aspirada do que quando falamos com um amigo ou com um desconhecido, o que significa que consoante o nosso **interlocutor** e relação que temos com este a nossa voz também varia. Usamos ainda uma diferente entoação quando pedimos uma informação. Esta entoação será oposta àquela que empregamos quando damos uma informação (Campbell, Mokhtari, 2003). Em termos práticos de aplicação, esta distinção e diferenciação têm particular interesse em livros falados, filmes para crianças, em que as personagens têm estilos próprios, conseguindo-se reconhecer o vilão ou o herói de acordo com a sua qualidade de voz.

Apesar de muitas vezes de forma incontrolável, a nossa qualidade vocal varia de acordo com a situação, contexto sociocultural, dependendo do nosso interlocutor e mesmo, tendo em conta o efeito que queremos ter no nosso receptor. Compreende-se assim que tanto factores internos como externos interfiram e determinem qualidade de voz, funcionando como variáveis psicológicas inerentes ao processo de comunicação. Fazem, por isso mesmo, variar factores como a frequência fundamental, ritmo, intensidade, velocidade e entoação. (Behlau et al, 2004; Behlau; Ziemer, 1987; Behlau; Pontes, 1990).

2.1.3.5 Utilização nas línguas do mundo

Numa descrição mais detalhada das diferenças entre qualidades de voz, verifica-se que diversas línguas utilizam, ao longo do período de fonação, dois pontos de contraste: sons vozeados e não vozeados. No Inglês, Japonês, Russo e Árabe, as diferentes qualidades de voz surgem, não por existir qualquer anomalia em termos de aparelho fonatório mas, para facilitar a distinção de sons. A maior parte das línguas utiliza diferentes tipos de fonação nas consoantes oclusivas (distinguindo-as por vozeadas vs. não vozeadas); na Islândia esta distinção serve para a fonação de diferentes nasais e nas línguas Ameríndias, Bantu e do Congo este contraste utiliza-se na diferenciação de vogais e no Japonês é comum vogais vozeadas serem desvozeadas em certos contextos tal como em posição final e quando perto de consoantes surdas (Ladefoged e Maddieson, 1996).

Por exemplo, o Gujarati, usa a qualidade aspirada para distinção entre vogais (Ladefoged e Maddieson, 1996: 315; Gordon e Ladefoged, 2001). Nas vogais aspiradas a amplitude do segundo harmónico é consideravelmente menor do que a do F0. O harmónico que se encontra mais próximo do primeiro formante tem uma amplitude muito mais baixa do que a fundamental. É nestas vogais que se nota uma maior descida de energia à medida que aumenta a frequência.

Outras línguas distinguem as suas consoantes com o contraste entre voz estridente e modal (contraste particularmente comum em línguas dos Índios da América). Normalmente, as vozes mais irregulares são as que apresentam perturbações na fonte glotal e estas incluem perturbações em termos de *jitter* e de *shimmer* – mais informação na secção seguinte – associando-se estas a uma instabilidade em termos de vibração das cordas vocais. Assim, uma voz estridente está associada a impulsos aperiódicos da glote, o que se verifica pela medição dos valores de *jitter* que são mais elevados durante a fonação de uma voz estridente, do que em qualquer outro tipo de fonação.

Num estudo levado a cabo por Nick Campbell (2007) salienta-se que para estudarmos e percebermos a nossa voz e sua variação de acordo com diversos e diferentes contextos sociais, teremos sempre que ter em conta os quatro

parâmetros prosódicos “*voice pitch, vocal energy, speech timing and voice quality*”. O nosso estudo, em termos de qualidade de voz, procura verificar, estudar e analisar essas diferenças nos valores dos parâmetros acústicos em tudo o que à expressão de emoção diz respeito.

2.1.3.6 *Parâmetros caracterizadores da qualidade de voz*

Tendo em conta a investigação até agora realizada sobre qualidade de voz, que tentámos sempre seguir na versão mais actualizada possível, apresentamos uma resenha de todos os parâmetros relacionados com a fonação utilizados por outros estudos que nos serviram de guia à nossa pesquisa².

Todos os estudos relacionados com qualidade de voz envolvem diversas medidas, métodos e técnicas, não só relacionados com fonação e articulação, como com prosódia, por exemplo. A entoação é também importante na análise de qualidade de voz, uma vez que é capaz de veicular muita informação sobre a pessoa e mensagem, expressando intenção, atitude e emoção (Ishi et al, 2008). Tendo em conta as mudanças que a nossa voz pode sofrer (de acordo com o estado psíquico, físico ou de saúde) importa ao investigador procurar os estímulos correctos para conseguir uma análise articulatória (da actividade glotal, muito sensível aos aspectos anteriormente focados) o mais controlada possível. Os parâmetros mais relacionados com as características da glote são considerados mais importantes, uma vez que são inerentes aos comportamentos biomecânicos das cordas vocais e porque na detecção de patologias específicas serão necessários outros parâmetros relacionados com medidas de perturbação (medidas do primeiro harmónico e dos formantes) (Gómez et al, 2005). Há, em diferentes estudos, a tentativa de definição e análise dos diferentes parâmetros que se alteram na presença de uma patologia de voz, uma preocupação na análise detalhada de parâmetros acústicos, tais como o *pitch, jitter* e *shimmer*

² Seguimos as descrições e medidas da Universidade de Stuttgart.

(relacionados mais com dispersão) e ruído (relação entre as altas e as baixas frequências).

No entanto, tendo em conta todas essas mutações, as técnicas, o estudo recorrente e aprofundado tem conduzido a resultados importantes. Os estudos sempre se reportaram mais a parâmetros relacionados com F0. Só mais recentemente se começaram a alargar as análises a parâmetros como *jitter*, *shimmer*, HNR e parâmetros relacionados com a fonte glotal (quociente de abertura).

Métodos

Um método comum para estudar o funcionamento glotal é a filtragem inversa (*inverse filtering*), no qual os efeitos do tracto vocal e lábios são retirados (Airas, 2008). Este método foi primeiramente introduzido por Miller (1995), que aplicou filtros digitais, de forma a cancelar os dois formantes mais baixos e os movimentos labiais, de um sinal de voz capturado através de um microfone (Pullaka, 2005).

O que se sabe sobre a fonte glotal permite a existência da teoria necessária para a técnica do “*inverse filtering*”. Se transferir a função do filtro do tracto vocal que é conhecido, pode-se construir uma filtragem inversa. Neste caso, a função glotal pode ser reconstruída passando o sinal de voz através do filtro inverso do tracto vocal.

Aplicando este filtro inverso ao sinal de voz, chega-se a uma estimativa da intensidade do sinal e da velocidade da fonte glotal. Este mecanismo é também conhecido como *flow glottogram* (FGG) (Hertegard et al., 1992; Hertegard e Gauffin, 1995, citados por Pullaka, 2005).

Para se medir a actividade da fonte glotal são utilizados vários métodos de observação durante o acto de fala, resumidos na Tabela 5.

Tabela 5 – Métodos de observação da actividade da fonte glotal.

<i>Técnicas para avaliação do comportamento da laringe</i>	
Métodos auditivos	Testes de percepção
Métodos visuais	Laringoscopia: directa, laser, estroboscópica Fotografia de alta velocidade
Correlacionados com o movimento das cordas vocais	Fotoglografia Electroglotografia Ultrasonografia
Extracção a partir do sinal acústico	Directamente de: microfone de contacto, tubo de Sondhi Filtragem inversa de: sinal de microfone, sinal de pressão obtido com máscara de Rothemberg

A **observação directa** do comportamento das pregas vocais só é possível visualmente. No entanto, os aspectos relacionados com os parâmetros de fonação são medidos de forma indirecta, sendo precisa uma análise alargada, tendo em conta diferentes métodos: a audição; a análise visual e a análise indirecta. A complexidade deste estudo e a franca procura de resultados cada vez mais detalhados levam a que muitas das análises acabem por ser, de certa forma, invasivas (como o uso do electroglotógrafo, EGG).

No nosso estudo e dadas algumas condicionantes (como a fala espontânea) não utilizámos o EGG, uma vez que consideramos que este limita essa espontaneidade pretendida. A forma mais directa de análise é a avaliação através da audição. Esta técnica é principalmente usada ao nível das vozes com patologia (Wendler et al., 1996; Hammarberg, 1997).

Todas as diferenças em qualidade de voz são possíveis de investigar tendo em conta análise da intensidade, inclinação do espectro, frequência fundamental, frequências dos formantes, duração, passagem do ar e periodicidade.

Frequência fundamental – F0

A medida da frequência fundamental³ depende muito do registo periódico do sinal. Este é periódico se caracterizado pela repetição exacta dos contornos da onda, sendo cada repetição da onda um ciclo: “*A series of events is termed periodic if the events cannot be distinguished from one another by shifting time forward or backward by a specific interval*” (Titze, 1994, pp.87-88, 1995, p.8). A frequência fundamental é, em suma, o número de vezes que uma onda se repete numa determinada unidade de tempo, é medida em hertz (Hz), sendo que 1 Hz é igual a um ciclo por segundo

A frequência da vibração glotal determina a frequência fundamental. A média de F0 é aproximadamente, para os homens, 120 Hz, enquanto que para as mulheres é de 220 Hz. A frequência fundamental é ainda mais elevada para as crianças (Karjalainen, 2000, citado por Pullaka, 2005). No entanto, estas medidas não se apresentam como estanques. Por exemplo, são comuns F0 abaixo dos 100 Hz em homens; no entanto, os tenores podem conseguir alcançar os 600 Hz. Para as mulheres, a frequência fundamental mais baixa estará perto dos 150 Hz enquanto que o limite para um soprano pode ultrapassar os 1300 Hz (Titze, 1994, citado por Pullaka, 2005).

As mudanças na frequência fundamental estão, normalmente, relacionadas com uma mudança na qualidade de voz, embora, em trabalhos anteriores, Ní Chasaide e Gobl (1996) tenham considerado que a qualidade de voz e o F0 podiam ser controláveis de formas distintas.

³ *Pitch* e frequência fundamental, apesar de por vezes serem confundidos, referem-se a diferentes medidas. *Pitch* é a sensação de frequência, ou seja, a forma como percebemos um som como mais grave ou mais agudo. A frequência é um atributo físico do sinal: velocidade de repetição de uma onda (Baken e Orlikoff, 2000).

Perturbações na frequência fundamental

A perturbação da frequência é normalmente designada por *jitter*, representando a variação da frequência fundamental de um ciclo para outro (Baken e Orlikoff, 2000, p.90). É uma medida de curto termo (ciclo a ciclo) de variabilidade involuntária do F0, o que permite determinar o grau de estabilidade do sistema fonatório (Baken e Orlikoff, 2000; Guimarães, 2007). Os resultados da perturbação são apresentados em milissegundos ou microsegundos (Baken e Orlikoff, 2000). Correlacionando F0 e *jitter*, pode-se dizer que as frequências mais elevadas apresentam um menor índice de perturbação. Apesar dos vários parâmetros de *jitter* que podem ser analisados⁴, no nosso estudo reportamo-nos, principalmente à análise de *jitter* Local (diferença média absoluta entre dois períodos consecutivos; *jitter RAP* (*Relative Average Perturbation*), ou seja a diferença média absoluta entre um período e a sua média e dos seus dois períodos consecutivos, dividida pela média do período. O *jitter PPQ5* (*Period Perturbation Quotient*) dá-nos a diferença média absoluta entre um período e a sua média e a dos quatro períodos seguintes, dividida pela média do período (Boersma e Weenink, 1992-2005).

Perturbação na amplitude da vibração

Outro parâmetro que temos em conta na nossa análise refere-se à medida da variação da amplitude da vibração. A esta medida chamamos *shimmer* e representa a variação da amplitude de ciclo para ciclo. Se, em termos de fonação, a voz humana fosse absolutamente estável, o *shimmer* seria zero. No entanto, tal como o *jitter*, havendo uma patologia em termos de funcionamento da laringe, por exemplo, os valores serão sempre mais elevados. Parece claro que os valores de

⁴ *Jitter local*; *Jitter* local absoluto; *Jitter rap*; *Jitter* ppq5 e *Jitter* ddp, (Boersma e weenink, 1992-2005; Casper e Colton, 1998).

shimmer tendem a normalizar quando a patologia laríngea apresenta melhoras (Baken e Orlikoff, 2000, p.130). Se uma patologia laríngea for eliminada, os valores de *shimmer* tendem a normalizar (Hirano et al. 1998).

O *shimmer* permite quantificar as alterações mínimas da amplitude do sinal, com base em cada ciclo fonatório (Casper e Colton, 1998; Guimarães, 2007). Em termos de análise de voz, no que diz respeito à presença de patologias com ela relacionadas, as medidas de *jitter* e *shimmer* parecem merecer destaque. Na nossa análise e apesar de, tal como o *jitter*, o *shimmer*⁵ também apresentar várias medidas, aprofundámos mais as medidas de *shimmer* Local e APQ3 (*Amplitude Perturbation Quotient*). A primeira dá-nos a medida absoluta das diferenças entre as amplitudes de períodos consecutivos dividida pela média da amplitude. O segundo parâmetro mede o quociente de perturbação em três pontos (média absoluta da diferença entre a amplitude de um período e a média das amplitudes dos seus vizinhos, dividida pela média da amplitude).

As medidas de *jitter* e *shimmer* apresentam-se como fundamentais na detecção de uma patologia de voz, ou mesmo na diferenciação entre vozes com patologia e vozes saudáveis (McAllister et al., 1996, pp.252-253).

Medidas de Harmonicidade

Outra medida que tivemos em conta no nosso trabalho, foi *Harmonic Noise Ratio* (HNR). Trata-se de uma medida que relaciona a componente harmónica com a componente de ruído da onda. Se o ruído espectral entre harmónicos aumenta, a energia da frequência dos harmónicos diminui (Guimarães, 2007). *Harmonic noise ratio*, HNR, é um dos parâmetros acústicos que representa a medida de ruído presente no sinal de voz (Michaelis et al., 1998, p.1628). Esta

⁵ *Shimmer* Local dB mede a diferença absoluta do logaritmo de base 10 da diferença entre as amplitudes de períodos consecutivos, multiplicado por 20. *Shimmer* APQ5 mede o quociente de perturbação da amplitude em cinco pontos (média absoluta da diferença entre a amplitude de um período e a média das amplitudes dos quatro vizinhos, dividida pela média da amplitude) (Helena Vilarinho, dissertação de mestrado, 2008).

medida é expressa em dB: se 99% da energia do sinal for periódica e 1% representar ruído, o HNR é $10 * \log_{10}(99/1) = 20\text{dB}$. Um indivíduo sem qualquer patologia de voz pode produzir as vogais sustentadas [a] e [i] com valores de HNR perto dos 20 dB, e a vogal [u] perto dos 40 dB. Valores inferiores a 7 dB podem indicar a presença de uma patologia (Behlau, 1997).

2.1.4 Patologias

A percepção de qualidade de voz não é universal e depende tanto dos traços e características culturais como do sistema fonológico inerente à língua nativa do receptor. A descrição de patologias de voz tenta, contudo, ser universal e baseia-se em aspectos mais abstractos do funcionamento laríngeo (Wendler et al., 1996; Hammarberg, 1997)⁶.

Torna-se pertinente, no Português Europeu, dada a escassez de pesquisa nesta área, apresentar uma análise relativa às diferenças que podem existir em determinados parâmetros passíveis de ser indicadores de uma patologia. O presente estudo, através da análise de patologias de voz e da descrição da variação dos seus parâmetros, torna-se uma ferramenta de trabalho e suporte.

O estudo de características acústicas (através de testes de percepção) na descrição da qualidade de voz patológica tem sido implementado em vários contextos e com objectivos diversos. No entanto, os resultados aparecem, na maioria das vezes, como ambíguos e muitas vezes contraditórios. Várias pesquisas limitaram-se a avaliar a correlação entre as medidas de perturbação, que são acusticamente perceptíveis e a capacidade de definição do ouvinte em caracterizá-la como aspirada e, por isso mesmo, com características de ruído ou como áspera ou irregular e assim com características aperiódicas. As vozes com patologia têm, geralmente, valores mais elevados em parâmetros como *jitter* e *pitch*. Os parâmetros mais relacionados com as características da glote, são

⁶ www.ims.uni-stuttgart.de/phonetik

considerados mais importantes, uma vez que são inerentes aos comportamentos biomecânicos das pregas vocais. Na detecção de patologias específicas, outros parâmetros relacionados com medidas de perturbação são também as medidas do primeiro harmónico e dos formantes (Gómez et al, 2005).

O ruído é a primeira característica de uma patologia de voz. Este pode ir de um limiar bastante moderado até à total perda de voz. Estas patologias existem quando a qualidade, *pitch*, intensidade ou flexibilidade diferem das vozes de outros indivíduos da mesma idade, sexo e grupo cultural (Aronson, 1990). Um grande número de patologias da fala está relacionado com problemas de funcionamento da laringe (*disorder laryngeal function*), relacionando-se com qualidade de voz (vozes aspiradas ou roucas) e problemas no controlo do *pitch*, por exemplo, a doença de Parkinson caracteriza-se pela rigidez das cordas vocais: estas resistem ao movimento voltando, rapidamente, à posição central (Weismer, 1997, pp.207-209). Entre as várias formas de descrever patologias de voz, as mais comuns estão relacionadas com o grau de ruído (Hirano, 1981; Nawka e Anders, 1996). O ruído, em termos perceptivos, explica uma anomalia vocal mais do que qualquer anomalia que possa ter origem na configuração do tracto vocal. Esta percepção de ruído pode obedecer a uma escala. Hirano (1981) propõe exactamente este tipo de descrição que inclui aspectos e dimensões perceptuais de voz. Os factores envolvidos encontram-se numa escala que pode ir de 0 a 3 e são os seguintes: O grau de ruído presente no som produzido (G); O grau de rouquidão, tendo em conta a irregularidade da frequência fundamental, (R); O grau de passagem de ar (voz aspirada), (B); Astenia, ou seja, a percepção de uma voz sem “força”, (A); Tensão muscular, causando, por isso, uma voz tensa, (S).

Este sistema é conhecido por GRBAS, utilizado nos Estados Unidos da América e no Japão (Isshiki e Takeuchi, 1970; Hirano, 1981 e 1989). No entanto, este sistema tem sido contestado na Europa por se considerar que o aspecto astenia está muito relacionado com a passagem de ar e voz aspirada e que, por outro lado, os julgamentos do que se considera uma voz tensa, também divergem. Assim, passou a usar-se um sistema mais simplificado, utilizando apenas três dimensões perceptuais (rouquidão, grau de passagem de ar e grau

de ruído), chamando-se RBH (*roughness; breathiness and hoarseness*) (Wendler et al, 1986; Nawka e Anders, 1996)⁷. Embora os dois sistemas coexistam.

As lesões na laringe podem causar disfonia devido a vários factores: interferência da lesão no contacto das pregas vocais durante a fase de encerramento do ciclo glótico; escape de ar em consequência de uma lesão; alterações do tecido nas camadas da lâmina própria que inibem ou suprimem a onda da mucosa (Harris, 1998, citado por Behlau, Madazio e Pontes, 2001).

Clinicamente mudanças involuntárias podem ser causadas por outros aspectos, como fadiga, depois de falar durante longos períodos de tempo (Boone e Macfarlane, 1994, citados por Li et al., 2006). Este fenómeno também acontece no processo de mudança da fala dos adolescentes. A terapia vocal tem-se mostrado efectiva na melhoria da qualidade de voz e por isso na melhoria da qualidade de vida (saúde mental). No entanto, não afecta a patofisiologia laríngea ou reduz os altos níveis de stress que caracterizam indivíduos com disfonia. Muitos terapeutas utilizam estratégias psicológicas e encaminham os pacientes para um psicólogo clínico (MacKenzie et al, 2001).

Quer seja no meio clínico quer na literatura, os problemas de voz que não estão directamente relacionados com patologias orgânicas da laringe são definidos como funcionais, disfonias de tensão muscular; disfonias histéricas, ou ainda disfonias psicogénicas. Este tipo de disfonia é a manifestação de um desequilíbrio psicológico, e pode estar relacionada com medos inconscientes que se sobrepõem aos sentimentos racionais, deixando perceber grandes cargas emocionais: Há um aumento de manifestações de desequilíbrios como ansiedade, depressão, desordens de personalidade ou reacções de conversão, para além disso o controlo normal da fonação é perdido. (Janet Baker, 1998).

No nosso trabalho analisámos patologias não orgânicas uma vez que estas se relacionam com factores não controláveis e, por isso, na sua maioria

⁷ www.ims.uni-stuttgart.de/phonetik

emocionais. Convém pois fazer a distinção entre **patologias orgânicas e não orgânicas**.

A abordagem tradicional da classificação das disfonias utiliza a dicotomia funcional versus orgânica (Freeman et al, 2004; Titze, 1994). As perturbações orgânicas da voz são aquelas para as quais uma lesão específica pode ser identificada ao nível de um órgão do corpo (alterações congénitas; estruturais, inflamações; tumores). As perturbações funcionais não apresentam lesão, embora seja evidente a perturbação vocal (Titze, 1994).

Existe uma dificuldade em separar os dois tipos de patologia, uma vez que as desordens psicogénicas são consideradas, na sua globalidade, manifestações de um desequilíbrio psicológico. No entanto, estas patologias conduzem a uma tensão muscular que leva a um esforço maior e descontrolado das pregas vocais, podendo, por isso, confundir-se com uma patologia física. Assim, deve-se procurar que o diagnóstico não fique restringido a exames que apenas tenham em conta parâmetros relacionados com a qualidade vocal, mas que englobe todos os factores de stress e transtornos emocionais ao longo da vida (Seifert, 2006). As patologias de índole não orgânica, muitas vezes também denominadas funcionais ou psicogénicas, caracterizam-se por uma reduzida ou acrescida dificuldade na produção de voz, sem que haja qualquer tipo de “deformação” física.

As **patologias orgânicas** podem resultar de malformações da laringe, de inflamações crónicas, traumas, tumores benignos ou malignos e mesmo paralisia das pregas vocais. As patologias de origem orgânica incluem as lesões de massa e afectam a zona membranosa das pregas vocais, em forma de nódulos, pólipos, quistos, edemas, papilomas; leucoplasias, neoplasias, granulomas, úlceras de contacto, paralisia da prega vocal, entre outras (Guimarães, 2007).

A hiperfunção vocal está também na causa de muitas disfonias. Esta pode levar a alterações do tecido das pregas vocais, causando nódulos, pólipos, úlceras ou edemas. Estas são mudanças orgânicas das pregas vocais que vão reflectir-se na qualidade vocal, uma vez que há um aumento da massa e uma redução da elasticidade. Todos estes aspectos restringem a normal vibração das pregas vocais, o que afecta a fonação e o registo vocal.

É por vezes contestada a análise da qualidade vocal apenas pela produção de vogais sustentadas. Sapienza e Stathopoulos (1995) verificaram que não há diferenças significativas em termos de medidas do quociente de abertura nem de frequência fundamental. Comparam as medidas acústicas e aerodinâmicas da laringe entre mulheres sem qualquer problema de voz ou audição com mulheres sofrendo de nódulos vocais. Foram realizadas três tarefas distintas: produção de vogais sustentadas, repetição de sílabas e leitura. Este estudo veio corroborar diversas análises anteriores e, uma vez mais, se percebeu que o prolongamento de uma vogal representa uma medida de F0 mais elevada em vozes com patologia, uma vez que há também uma maior tensão das pregas vocais. Para além disso, os autores consideram que esta tarefa se revela de simples execução, em termos motores, uma vez que não há qualquer problema de co-articulação. Assim, os valores de F0 não são os mesmos em discurso normal ou em vogais sustentadas. Contudo, do ponto de vista clínico, estas diferenças (que são muito precisas na distinção de certos sons 0,1%-0,2%) não são perceptivelmente tão relevantes.

Debruçando-nos um pouco mais sobre as possíveis causas para o aparecimento de nódulos aparecem-nos como factores predominantes o stress, desidratação; refluxo farolaríngeo; inflamações e infecções (Biase e Pontes, 2002). Acrescentam-se ainda as variações hormonais, o consumo de medicamentos e drogas, tabaco, consumo de álcool, condições ambientais, abuso vocal e mau uso vocal (Guimarães, 2007). Há ainda factores anatomo-funcionais predisponentes a causar certos tipos de patologia (proporção glótica; assimetria laríngea e ângulo de abertura das pregas vocais) (Behlau, 2001),

Os nódulos vocais resultam da colisão repetida das pregas vocais durante a vibração. A frequência da colisão e a amplitude da vibração exercem um papel de relevo no aparecimento destes nódulos (Titze, 1994). Um nódulo vocal pode exercer um efeito bem pronunciado sobre o mecanismo de vibração, o que vai levar a uma vibração mais aperiódica, maior perturbação de frequência e uma maior rouquidão (Colton e Casper, 1996). As vozes com elevada frequência são mais susceptíveis ao aparecimento de nódulos. Assim, indivíduos que utilizam mais vezes uma maior intensidade vocal estão mais sujeitos a vir a ter nódulos

vocais, no meio da prega membranosa, onde a amplitude de vibração é maior. Os nódulos vocais são, por isso, mais recorrentes em mulheres e crianças do género masculino (Behlau et al. 2001; Guimarães, 2007; Morente et al., 2001).

Quando a voz se deteora, seja por resultado de um excessivo esforço vocal ou por doença, toda a personalidade do indivíduo é afectada, surgindo sentimentos de ansiedade e insegurança, principalmente naqueles que têm como instrumento de trabalho a própria voz (professores, cantores, actores, locutores), acabando este tipo de patologia por conduzir também a problemas profissionais e económicos (Greene, 1989).

Quando efeitos/causas psicológicas actuam na forma de comunicação do indivíduo, nomeadamente em problemas de voz, às alterações percebidas e que estão vinculadas a aspectos psico-emocionais é atribuída a designação de **disfonia psicogénica**. Uma desordem funcional, uma vez que o processo que conduziu ao aparecimento e instalação de voz disfónica, apresenta um simbolismo directo com a função fonatória da laringe. Indivíduos com disfonia têm normalmente dificuldades em variar a intensidade da voz independentemente da frequência (Heylen et al., 1998, citado por Li et al, 2006). Estes pacientes desenvolvem com frequência uma forma abusiva do uso da voz, falam alto diariamente.

Estes distúrbios podem ser explicados como emocionais ou de inadaptação; a voz é influenciada pelos sentimentos e atitudes do próprio indivíduo mas também se modifica de acordo com a reacção que o falante percebe no receptor (Behlau et al, 2004). Este tipo de perturbações da fala em que há uma relação directa com distúrbios psicológicos denomina-se como disfonia psicogénica. Os estados emocionais são elementos comuns em problemas que afectam a tensão músculo-esquelética, levando a comportamentos aberrantes da voz. No entanto, segundo os autores, convém salientar que nem todos estes tipos de disfonia têm origem apenas em factores psicológicos. O termo psicogénica aparece da mesma forma ligado a denominações como **funcional ou não-orgânico**. A última designação é preferida em relação à primeira, dada a ambiguidade que lhe pode estar subjacente (Rubin e Greenberg, 2002).

A disфонia é, normalmente, o primeiro e mais óbvio sintoma de uma patologia laríngea. Esta também pode advir de variados distúrbios, em que a disfunção vocal é uma característica secundária, resultante de situações anómalas ao nível dos sistemas respiratório, endócrino, nervoso ou psicológico e, em muitos casos, sem que haja qualquer componente orgânica presente (Fawcus, 2001). Estas últimas serão as patologias privilegiadas neste nosso estudo, uma vez que se trata de patologias do foro psíquico e emocional, como veremos no capítulo 3.

Existe um comportamento padrão que parece ligar todos os casos de disфонia. De acordo com Baker (2002), na população diagnosticada com disфонia existe uma inibição em expressar pensamentos e sentimentos. Para a psicologia comportamental, a disфонia surge como uma resposta/escape em casos de maior ansiedade, torna-se uma resposta habitual para a ansiedade “long after the original conflict may have first become evident, forgotten or even resolved” (Baker, 2002). As diferentes terminologias para este tipo de desordem reflectem a complexidade e dificuldade em encontrar uma definição exacta e apropriada para estes casos onde não existe qualquer problema estrutural ou patologia neurológica que possa ser identificada como causadora de uma disфонia (Baker, 2002). Existe ainda a disфонia espasmódica de adução que é marcada por espasmos dos músculos da laringe durante a fonação. Este processo leva à produção de uma voz entrecortada, tensa, forçada. Esta disфонia, de acordo com Imamura (2006) é classificada como distonia focal de etiologia ainda desconhecida, sendo uma das disfonias mais difíceis de serem tratadas. Os sintomas decorrem da contracção involuntária dos músculos durante a fonação, levando a que as pregas vocais se encontrem extremamente tensas, pressionadas uma contra a outra, havendo também um aumento da resistência glótica (Imamura, 2006).

2.2. Emoção

Se uma emoção é um conjunto das alterações no estado do corpo associadas a certas imagens mentais que activam um sistema cerebral específico, a essência do sentir de uma emoção é a experiência dessas alterações em justaposição com as imagens mentais que iniciaram o ciclo.

António Damásio, "O Erro de Descartes: Emoção, Razão e Cérebro Humano", 1996, p.159

As emoções são reacções naturais ao Ser Humano e surgem também como uma defesa natural para a sobrevivência (Charles Darwin, 1872). Sob esta perspectiva as emoções são vistas como comuns aos Homens, pelo que todos nós deveremos ter mais ou menos as mesmas emoções e, segundo Darwin, algumas das emoções do Homem são comuns às de outros mamíferos.

Os estudiosos que têm vindo a estudar a voz continuam a debater-se com os aspectos verbais e não verbais que se podem separar para conseguir fazer uma análise de qualidade de voz. O grau de independência que se consegue perceber entre estes dois itens liga-se ao facto das pessoas conseguirem perceber várias mensagens mesmo que misturadas. Logo, as palavras fornecem só por si uma informação, mas os aspectos não verbais fornecem tipos de informação bastante diferentes e que podem influenciar a interpretação das mensagens orais (Patrik N. Justin and Klaus R. Scherer 2008).

Analisar o discurso emotivo implica a utilização de vários métodos, uma vez que é preciso estudar o comportamento vocal, que é um dos factores que marcam a emoção (emoções, estados de espírito e stress) e é preciso também centrar o estudo em aspectos não verbais. A maior parte dos afectos e emoções envolvem

reações fisiológicas (alterações autônomas e do sistema nervoso somático), e tudo isto contribui também para modificar diferentes aspectos no processo de fonação. A excitação associada à raiva produz mudanças também na respiração e aumenta a tensão muscular, o que por sua vez influencia a vibração das pregas vocais e a forma do tracto vocal. Combinando todos estes aspectos as características acústicas da voz são afectadas (Scherer, 1986).

A forma como as emoções são expressas pela voz podem ser analisadas em três níveis diferentes:

- A Nível Psicológico (descrevendo os impulsos nervosos e intervenção dos músculos);
- A Nível Fonatório e articulatório (descrevendo a posição e movimentos das principais estruturas das pregas vocais);
- A Nível Acústico (descrevendo as características da onda sonora, no momento em que é produzida) (Patrik et al. 2008).

A Tabela 6 apresenta a síntese proposta por Scherer (2003), explicando a relação entre os subsistemas orgânicos e as componentes do processo emotivo.

Tabela 6 – Representação dos subsistemas orgânicos e componentes do processo emotivo (Scherer, 2003).

Função da emoção	Subsistema orgânico	Componentes da emoção
Avaliação dos objectos e situações	Processamento de informação (Sistema Nervoso Central) – (SNC)	Componente cognitiva/avaliação
Sistema de regulação	Suporte: (SNC); Sistema neuro-endócrino (SNE); Sistema nervoso autónomo (SNA)	Componente neurofisiológica (sintomas corporais)
Preparação e direcção da acção	Processo de execução Sistema nervoso somático (SNS)	Componente motivacional (tendência de acção)
Comunicação da reacção e intenção comportamental	Acção (SNC)	Componente das expressões motoras (expressões faciais e vocais)
Monitorização do sistema interno e interacção do organismo com o meio envolvente.	Monitorização (SNC)	Componente de sentimentos subjectivos (experiência emocional)

Uma outra dicotomia por resolver é a confusão entre emoção e sensação. Podemos adoptar desde já uma possível distinção entre as duas, considerando que a sensação é gerada “na nossa estrutura física”. Tanto as emoções como as sensações podem, contudo, ser consideradas sentimentos de acordo com Herbert Spencer (Essays, 1863, p.138, citado por Charles Darwin, 2006, p.33).

A emoção é uma mistura entre respostas psicológicas, cognitivas, fenómenos (biológicos, físicos e culturais) e comportamentos complexos (Ortony e Turnet, 1990). Na nossa vida, socialmente, percebemos que as nossas emoções só são desencadeadas após um processo mental de avaliação que é voluntário e não automático.

Desde que as emoções têm vindo a ser estudadas que existem grandes debates sobre o que é de facto uma emoção ou afecto e como é que estes podem ser descritos e analisados o mais objectivamente possível (Boehner et al, 2007). Este objecto de estudo torna-se tanto mais complicado quanto mais rico é (atentemos por exemplo na análise de emoções espontâneas). As emoções ocorrem de forma dinâmica, não são estáticas e por isso se tornam mais difíceis de medir e/ou analisar.

Em 2005, Scherer vem tentar clarificar um pouco toda a controvérsia que existe na definição e estudo de emoção. Considera também que o conceito se tornou algo banalizado, porque 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 poderá então ser um episódio que surge como resposta a um estímulo interno ou externo (Scherer, 2005).

Em termos neurobiológicos quando o Ser Humano experiencia uma emoção, esta experiência implica uma mudança em variados parâmetros relacionados com o funcionamento das vísceras (coração, pulmões, intestinos, pele), alterações da musculatura esquelética e glândulas endócrinas (pituitária e as supra-renais). O cérebro liberta moduladores péptidos na corrente sanguínea, o sistema imunológico altera-se; o ritmo de actividade dos músculos pode aumentar e causar contracção ou estreitamento dos vasos sanguíneos (causando palidez, por exemplo, associada a uma emoção triste); e se os músculos relaxarem e os vasos dilatarem, resulta no rubor da face. Em suma, todas estas alterações acontecem como desvios a um estado médio de equilíbrio funcional do organismo. Quando experienciamos diferentes emoções, muitas partes do nosso corpo atingem um novo estado, apresentando mudanças significativas.

Adoptaremos a explicação de António Damásio quanto à diferenciação entre emoções primárias (ou iniciais) e emoções secundárias (adultas)⁸.

⁸ Damásio (1996) refere que “as emoções que experienciamos na infância, para as quais «um mecanismo pré-organizado» de tipo jamesiano seria suficiente e as emoções que experienciamos em adultos, as quais se foram gradualmente alicerçando sobre as fundações daquelas emoções «iniciais».

2.2.1 Emoções Primárias (inatas ou pré-organizadas)

Segundo António Damásio, é forçoso que o Ser Humano e mesmo os animais tenham como características inatas o **medo**, por exemplo. Assim, de forma inata estamos “programados” para responder de determinada forma a determinados estímulos. Estas características inatas são processadas e depois detectadas por uma componente do sistema límbico do cérebro (a amígdala) cujos núcleos “neurais possuem uma representação de disposições que desencadeia a activação de um estado do corpo, característico da emoção medo”, alterando o processamento cognitivo. É graças a esta estrutura da amígdala que o cérebro recebe os sinais que desencadeiam os comportamentos/reacções ligadas a medo (fugir; exhibir raiva). As emoções primárias dependem de toda uma rede de circuitos do sistema límbico, sendo a amígdala e o cíngulo as principais estruturas. Esta relação entre emoções e amígdala aparece já no trabalho de Heinrich Klüver e Paul Bucy (1980) que demonstraram, através de uma ressecção cirúrgica da parte do lóbulo temporal (onde se encontra a amígdala criativa) que de entre uma série de outros sintomas, existe a indiferença afectiva (Damásio, 2006).

A razão mais comum para que sejam propostas emoções básicas prende-se com facto de tentarmos explicar e compreender da melhor forma certas rotinas que se prendem com as nossas emoções. Ou seja, certas emoções parecem existir em todas as culturas enquanto que algumas se podem considerar Universais quando associadas a certas expressões faciais, outras parecem existir como algo que nos identifica e diferencia biologicamente, ligando-se à nossa sobrevivência e necessidades básicas de cada indivíduo ou de cada espécie (Damásio, 2006).

As emoções primárias constituem o mecanismo básico. Em termos de desenvolvimento, cada indivíduo vai desenvolvendo emoções secundárias que acontecem quando “começamos a ter sentimentos e a formar ligações sistemáticas entre categorias de objectos e situações, por um lado, e emoções primárias, por outro.” Assim considera o autor que as estruturas do sistema límbico não são suficientes para sustentar também as emoções secundárias. “A

rede tem de ser alargada e isso requer a intervenção dos córtices pré-frontal e somatossensorial.”

Importa perguntar e tentar esclarecer se existem e, existindo, quais são as **emoções básicas**. Neste aspecto, seguimos o estudo de Ortony e Turnet (1990), segundo o qual algumas emoções parecem ter um papel privilegiado em termos de literatura e de investigação, sendo denominadas como emoções básicas, primárias ou mesmo fundamentais. No entanto, existe sempre uma divergência ao longo dos tempos e de acordo com cada autor. Se alguns autores identificam apenas duas emoções como básicas (por exemplo a dor e o prazer, como é sugerido por Mowrer (1960, citado por Ortony e Turnet, 1990), já outros chegam a propor 18 emoções (Frijda, 1986).

Nos Seres Humanos o processo, sendo similar ao dos animais irracionais, é, no entanto, mais complexo. Após as alterações corporais de resposta a dado estímulo, passa-se à “sensação da emoção em relação ao objecto que a desencadeou”. Isto porque há um processo de consciencialização, ou seja, esta tomada de consciência vai proporcionar uma maior protecção. Vamos estar preparados para reagir de determinada forma mediante o “perigo” que se aproxima ou que identificamos. Damásio defende que “sentir os estados emocionais (...) equivale a afirmar que se tem consciência das emoções”. Este processo dá-nos a possibilidade de resposta a estímulos com base nas nossas experiências anteriores e da nossa relação com o meio envolvente.

Cada emoção básica não contempla apenas um sentimento, mas uma família ou estados com ela relacionados. As emoções que pertencem a uma mesma família partilham de uma mesma actividade em termos psicológicos; cognitivos; natureza dos acontecimentos e forma de se manifestar, é esta variedade de características que as distingue e agrupa enquanto família (Ekman Paul 2003).

Para a raiva estão identificadas 60 expressões distintas e todas elas partilham características, por exemplo, em termos de contracção muscular e semblante mais fechado se falarmos em termos de descrição física.

No nosso trabalho cingimo-nos a um agrupamento de emoções pelas suas característica em termos de qualidade de voz. Cada emoção tem um tema base que é característico de família para família e dentro desse aspecto apresenta ainda variações que têm a ver com a idiosincrasia, integração social e características pessoais de cada indivíduo. As subdivisões de cada emoção básica, são exactamente um reflexo do desenvolvimento e aprendizagem do Ser Humano que se preocupou em alargar as suas capacidades e formas de expressão (Ekman, 2003).

A grande dificuldade reside em reconhecer que algumas emoções apresentam apenas variações, mas pertencem à mesma família. Continuam-se, por outro lado, a confundir fenómenos como estado de espírito; disposição e atitudes com emoção, Cowie et al (2003).

2.2.2 Emoções Secundárias (Adultas⁹)

Numa emoção considerada secundária, o estímulo ainda pode actuar directamente na amígdala (como vimos nas emoções inatas). No entanto, agora é analisado no processo de pensamento, podendo assim activar os córtices frontais¹⁰, e estes sim actuam através da amígdala. As disposições pré-frontais adquiridas e necessárias para as emoções secundárias são diferentes das que são necessárias para as emoções primárias. Todos estes mecanismos são autónomos e acontecem de forma inconsciente, automática e involuntária. A resposta das disposições pré-frontais é assinalada à amígdala (Figura 12) e ao

⁹ Estas emoções recebem também a denominação de emoções adultas pelo facto de serem já um resultado, uma resposta a disposições adquiridas e não inatas (como as primárias)

¹⁰ O lobo frontal, inclui o córtex motor e pré-motor e o córtex pré-frontal, está envolvido no planeamento de acções e movimento, assim como no pensamento abstracto. A parte da frente do lobo frontal, o córtex pré-frontal, liga-se à tomada de decisões de que forma as sequenciar e avaliar o seu resultado. As suas funções incluem o pensamento abstracto e criativo, a fluência do pensamento e da linguagem, respostas afectivas e capacidade para ligações emocionais.

cíngulo anterior, enviando sinais ao corpo através dos nervos periféricos e sistema motor, de tal forma que as alterações na musculatura esquelética são a expressão física e visível da emoção (expressões faciais e posturas corporais) (Damásio, 1996, p.152).

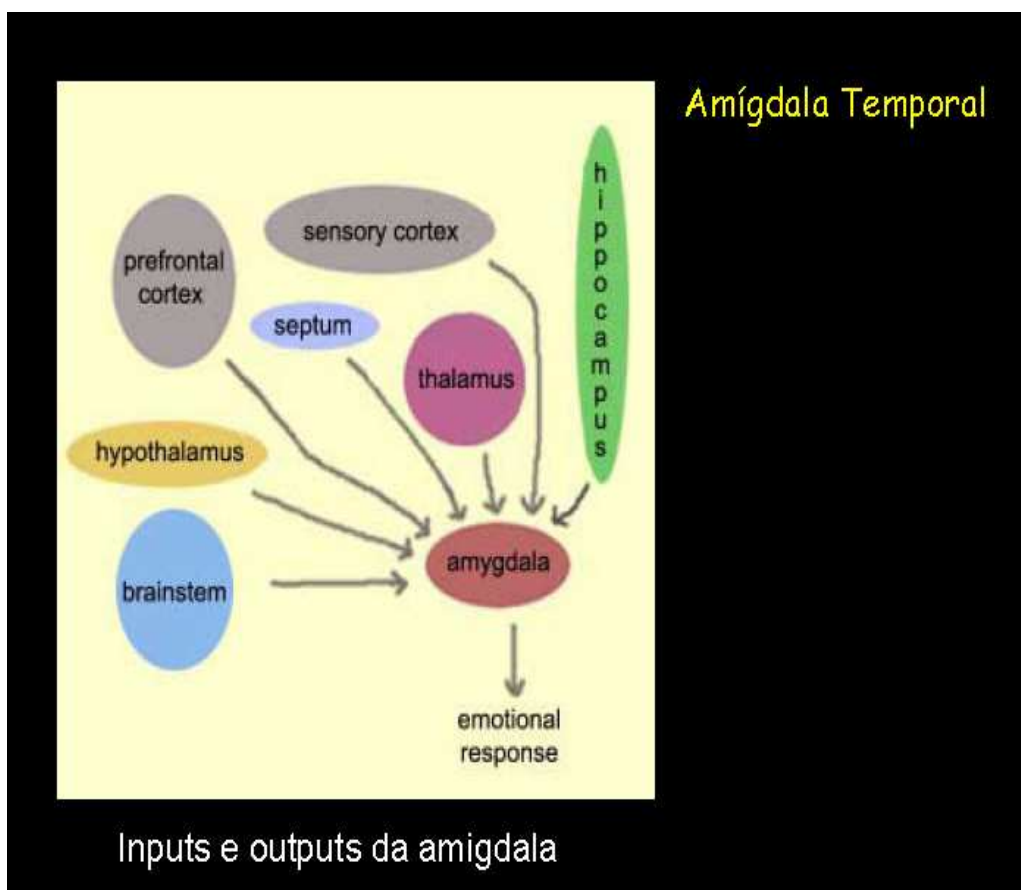


Figura 12 – Representação da amígdala, responsável pelo reconhecimento das emoções.

A emoção é uma combinação de “um processo avaliatório mental, simples ou complexo, com respostas disposicionais e esse processo, na sua maioria dirigidas ao corpo (...) resultando num estado emocional do corpo (...) dirigidas ao próprio cérebro (núcleos neurotransmissores no tronco cerebral), resultando em alterações mentais adicionais” (Damásio, 1996, p.153).

A essência da emoção engloba todas as alterações no estado do corpo que são induzidas a toda uma infinidade de órgãos através das “terminações das células nervosas sob o controle de um sistema cerebral dedicado” que vai

responder a pensamentos relacionados com uma determinada situação ou entidade.

A Tabela 7, apresentada seguidamente, é baseada no quadro apresentado na *Psychological Review*, (1990, Vol. 97, No. 3, 315-331) mostrando uma resenha das diferentes emoções que são referidas como primárias, por vários estudiosos ao longo do tempo.

Tabela 7 – Amostragem das diferentes emoções apresentadas como primárias, por vários estudiosos, ao longo do tempo.

Autor	Emoções Básicas	Respostas do organismo
Arnold, 1960	Raiva; aversão; coragem; desejo; desespero; medo; ódio; esperança; amor e tristeza.	A relação com as tendências das nossas acções.
Ekman, Friesen e Ellsworth, 1982	Raiva; aversão; medo; alegria; tristeza e surpresa.	Expressões faciais Universais
Frijda, 1986	Desejo; alegria; interesse; surpresa; imaginação e arrependimento.	A forma rápida como a elas respondemos.
Gray, 1982	Raiva; terror; ansiedade e alegria.	Bastante fortes
Izard, 1971	Raiva; contentamento; aversão; aflição; medo; culpa; interesse; alegria; vergonha e culpa.	Bastante fortes
James, 1884	Medo; amor; raiva; sofrimento	Envolvimento corporal
McDougall, 1926	Raiva; aversão; carinho; imaginação; exaltação.	Relação com os instintos
Mowrer, 1960	Dor e prazer	Estados emocionais que não se podem aprender
Oatley e Johnson Laird, 1987	Raiva; aversão; ansiedade; alegria e tristeza.	Não necessitam de um propósito consciente
Panksepp, 1982	Expectativa; medo; raiva; pânico	Bastante fortes
Plutchik, 1980	Aceitação; raiva; antecipação; aversão; alegria; medo; tristeza e surpresa.	Relacionados com o processo biológico adaptativo
Tomkins, 1984	Raiva; interesse; contentamento; aversão; aflição; medo; alegria; vergonha e surpresa.	Densidade, poder neurológico
Watson, 1930	Medo; amor e raiva	Bastante fortes
Weiner e Graham, 1984	Alegria e tristeza	Atribuições independentes

Tentando encontrar um consenso, Turner e Ortony (1992) centram-se nas emoções que são visadas mais vezes por mais estudiosos. Assim, poderemos ter como emoções básicas (mais comuns) a raiva, alegria, tristeza e medo. Concordam também que o facto de não se chegar a um verdadeiro consenso tem a ver com a forma como a palavra pode ser vaga em termos linguísticos, especialmente no que se refere a estados psicológicos.

2.2.3 Famílias de emoções

Outro aspecto que salientamos do estudo Ortony e Turner (1990), é a divisão das emoções em classes (ou famílias como falaremos mais adiante neste trabalho). Assim, temos emoções positivas, negativas ou neutras¹¹.

O que melhor consegue explicar o que queremos dizer quando falamos de emoções básicas ou primárias tem a ver com a comparação entre emoções e linguagem. Existem inúmeras linguagens humanas e muitas mais são ainda possíveis. No entanto, os linguistas não procuram explicá-las postulando um pequeno grupo como línguas básicas a partir das quais todas as outras advêm, o mesmo acontecendo, portanto, com as emoções (Ortony e Turner, 1990).

2.2.4 Medição/Caracterização

Alegria, tristeza, esperança, medo, raiva, curiosidade, são emoções tratadas como se cada uma delas por si mesma fosse uma espécie de entidade “that enters full-made upon the scene” algo que pode demorar muito ou pouco tempo, mas cuja duração, amadurecimento e tempo são relevantes para a sua natureza. **As emoções são qualidades** significativas são uma complexa experiência que se vai transformando (Dewey, 1934, McCarthy e Wright, 2004, p.83, citados por Boehner et al, 2007).

Para Scherer (2005), a emoção é um **episódio** em que existem variadas mudanças que se encontram todas elas interligadas e que são a resposta do nosso organismo a um estímulo interno ou externo. Definindo e explicando o que é emoção, Scherer recorre à descrição de simples actos do quotidiano. Assim, as

¹¹ “compare the cases of being surprised about winning a huge prize in a lottery (positive), being surprised about failure of one’s brand new car to start one morning (negative), and being surprised by some highly improbable but personally irrelevant fact such as that all the members of some committee by chance share the same birthday (neutral)”, (Ortony e Turner, 1990, pp. 317).

peças não mostram qualquer tipo de resposta emocional face um indivíduo que desconheciam, logo a emoção é uma resposta a estímulos que são já tidos pela pessoa como relevantes e reconhecidos, pelo que há uma resposta apropriada a cada situação.

Uma emoção não é estanque e muda rapidamente de acordo com estímulos e respostas, pelo que outro aspecto a ter em conta na análise das emoções é a **instabilidade da emoção**. Por outro lado, muitas vezes as nossas emoções são controladas, dado o impacto que certas respostas emocionais podem causar socialmente (daí que sejam também condicionadas pelo contexto).

Outro aspecto importante é a **intensidade**, que pode ajudar a distinguir emoções de estados de espírito, “emotions from moods” (Scherer, 2005).

O último ponto desta caracterização é a **duração**. A duração de uma emoção, geralmente como uma resposta a uma acção exterior é, normalmente, rápida, uma vez que o organismo procura sempre adequar o seu comportamento ao contexto envolvente. A duração tem que ser relativamente curta na medida em que não limita o organismo e permite uma flexibilidade de comportamentos. Em contraste disposições consideradas de baixa intensidade, que têm pouco impacto em termos de comportamento, podem ser mantidas por períodos de tempo muito mais longos (Scherer, 2005, p.702).

Nesta perspectiva de constante mudança torna-se complicado “medir” e caracterizar as diferentes emoções, tanto mais que sabemos que elas são dinâmicas e que englobam características da própria estrutura social do indivíduo.

Para se poder medir exactamente uma emoção seria necessária uma análise contínua de vários factores (Scherer, 2005):

- Mudanças que acontecem ao nível do sistema nervoso central;
- Medição dos padrões de resposta neuroendócrinos, autónomos e sistema nervoso somático;
- As motivações das mudanças (as respostas);

- O padrão das expressões faciais; expressão da voz e postura corporal;
- A natureza da subjectividade que está inerente a qualquer um destes tipos de comportamento e que reflecte cada uma das respostas.

Estas medidas nunca foram realizadas. No entanto, muito se tem já feito para compreender o mecanismo das nossas emoções. Têm vindo a ser realizados estudos medindo componentes individuais ligadas a apreço, por exemplo; o mecanismo cerebral; padrões de resposta fisiológica; comportamento em termos de expressão. Enquanto que comportamentos verbais e não verbais podem ser utilizados como indicadores fisiológicos para perceber determinada emoção, o mesmo não acontece com a subjectividade inerente a cada sujeito na sua experiência de determinada emoção.

2.2.5 Formas de descrição

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. Há muita variação, o que complica a investigação sobre voz: as diferenças individuais de cada um; a interacção entre discurso espontâneo e expressões estratégicas (meios sociais). Muitos estudos vêm referindo *corpora* pouco credíveis, onde 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 activação activo/passivo (*aroused/sleepy*) e da distinção agradável/desagradável, ou negativo/positivo como denominaremos no nosso 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 reflectem a avaliação cognitiva e consequentes 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 14 apresenta-se a descrição realizada por Scherer (2005, pp.720).

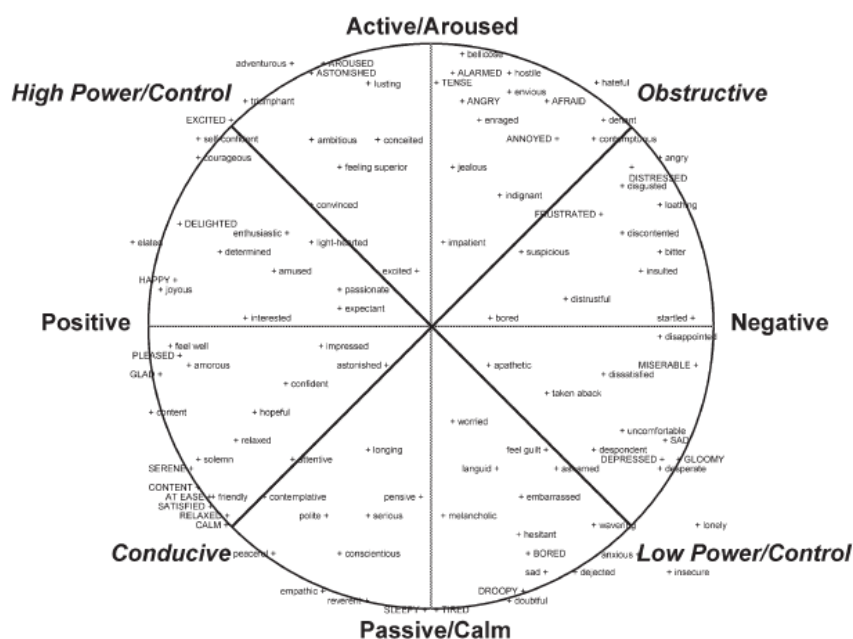


Figura 13 – Descrição das emoções proposta por Scherer (2005).

No seu estudo Scherer (2005) apresenta ainda o protótipo de um instrumento para classificação de emoções criado na Universidade de Genebra, chamado *Geneva Emotion Wheel* (GEW), Figura 15. Na sua primeira versão este instrumento 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, de acordo com o ouvinte, 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).

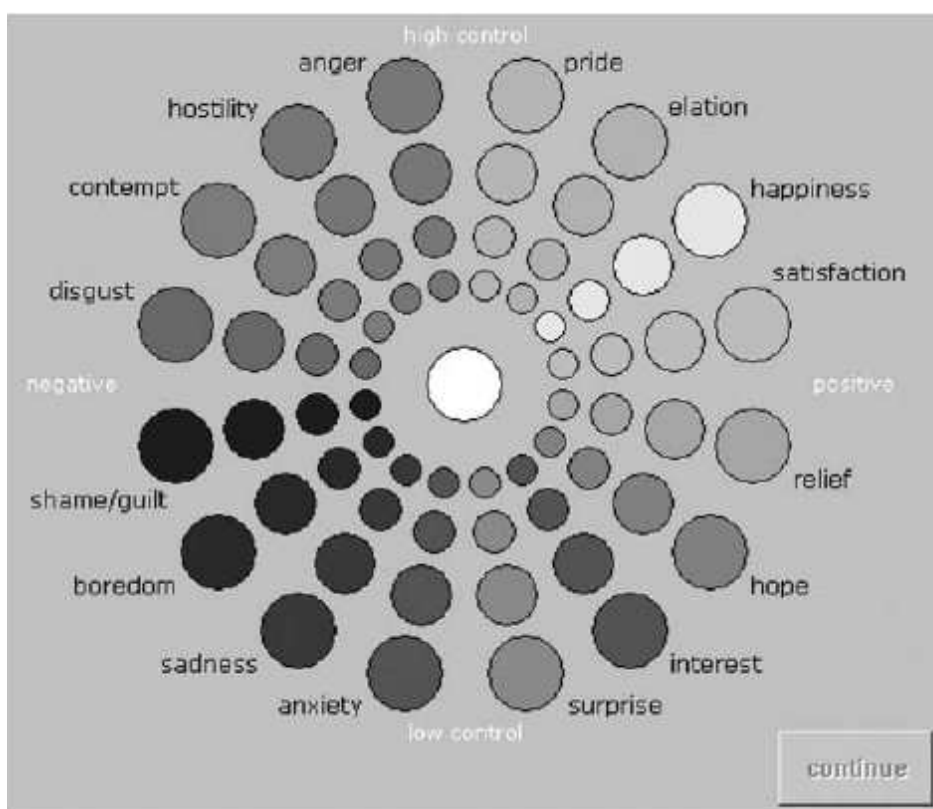


Figura 14 – Círculo bidimensional de amostragem da qualidade das emoções, (GEW), <http://www.unige.ch/fapse/emotion/resmaterial/gew.zip>.

2.2.6 Expressão das emoções

O Homem pode recorrer a várias formas de manifestar e dar a conhecer as suas emoções, de forma mais ou menos discreta: a voz; a expressão corporal; a expressão facial, o riso e o choro.

É muito provável que o ser humano antes de possuir a faculdade de uma linguagem articulada se tenha comunicado por sons, por exemplo, por sons musicais, pelo que quando a nossa voz é associada a alguma emoção em particular tende a assumir, através de um princípio de associação, um carácter musical (Darwin, 2006).

A **voz** está sem dúvida associada à expressão de sentimento. Se um indivíduo se queixa de algo de mal que lhe aconteceu ou de alguma dor, exprime-o quase sempre num tom agudo; já no medo, ansiedade ou nervosismo é possível perceber uma voz trémula e rouca.

Também a **postura corporal** ou as sensações que percorrem o nosso corpo podem ser manifestações de diferentes e diversas emoções. “Se juntamente com a surpresa, existir uma sensação de dor” a reacção do nosso corpo manifesta-se numa contracção muscular, incluindo os músculos do rosto. O medo, por exemplo, tal como o nervosismo, produz tremuras musculares (Darwin, 2006). Em termos de reacções corporais, temos ainda, por exemplo, associada à vergonha a ruborização da face; a transpiração é um outro resultado involuntário de resposta às nossas emoções, à raiva e ao medo, por exemplo.

A nossa **expressão facial** compromete muitas vezes a tentativa de ocultar as nossas emoções, precisamente porque as contracções musculares são involuntárias e são nada mais do que respostas a estímulos (agradáveis ou desagradáveis). As partes mais expressivas do rosto humano têm a ver com a zona dos lábios, testa e olhos (daí que também se refira que um sorriso verdadeiro implica um ligeiro franzir de olhos). Quando erguemos as sobrancelhas, por exemplo, franzimos também a testa (que se deve à contracção dos músculos supraciliares). Este movimento pode evidenciar dúvida, atenção, preocupação ou mesmo desaprovação, raiva. Quando falamos de algum tipo de tristeza é possível observar uma descida das pálpebras, lábios, bochechas e maxilar inferior. Por outro lado, a tristeza faz também com que a respiração se torne mais lenta, muitas vezes interrompida por profundos suspiros. Darwin (2006) chega ainda a referir como “músculos de pesar” os músculos das pálpebras (que quando estamos tristes, de certa forma arqueiam), a pele por

baixo das pálpebras encontra-se também descaída (nos cantos exteriores dos olhos), parecendo enrugada. A prega nasolabial fica bastante marcada, dada a sua contracção. Quando estamos alegres e rimos também o nosso rosto se altera. São os músculos dos olhos que, mais uma vez, assinalam a nossa emoção (exteriormente) bem como as rugas que se formam na testa e a curvatura da prega nasolabial.

A forma de manifestar, muitas vezes, as suas emoções e a primeira forma de comunicação numa criança é o **choro**. E também este pode ser interpretado. Quando as crianças choram e abrem muito a boca, procuram desde logo chamar a atenção e projectar o som (Darwin, 2006). Quando sente qualquer espécie de desconforto a criança “emite berros fortes e prolongados” Darwin (2006). Parecendo um paradoxo, também o adulto expressa tanto a emoção de tristeza como a de extrema alegria através do choro.

O **riso**, quando autêntico, pode também ser agudo ou grave. Neste aspecto, o autor refere Haller (de acordo com a citação de Gratiolet, *De la Physionomie*, 1865, p.115) que defendia que nos homens adultos o som do riso se identifica com as vogais [O] e [A] (na sua pronúncia alemã), enquanto que o riso das crianças e mulheres se identifica com o som das vogais [E] e [I], que têm um tom mais agudo do que as primeiras. No entanto, é incontornável que se trata de expressão de alegria e boa disposição tanto nuns como noutros. O riso parece ser a expressão genuína da alegria ou da felicidade

Todos estes nossos movimentos voluntários ou involuntários são manifestações das nossas emoções. Contudo, neste trabalho cingimo-nos ao estudo e análise da emoção pela voz. Pretende-se aferir como varia a nossa voz de acordo com o que sentimos e que parâmetros podem ser analisados para percebermos as diferenças entre as expressões de diferentes emoções. Até agora, muitos estudos, e alguns que vamos referindo ao longo deste trabalho, baseiam as suas análises na audição de enunciados juntamente com a expressão facial o que pode, em muito, facilitar a identificação de uma emoção. Uma vez mais, salientamos que a nossa pesquisa versa unicamente sobre voz e qualidade de voz e emoção no Português Europeu.

2.3. Voz e Emoções

Se atentarmos nas considerações de Charles Darwin (edição de 2006), verificamos que esta atenção à voz e suas modulações é importante em diversos contextos interpessoais, em suma, importante na socialização. Segundo o autor, a voz humana já apresenta diferentes expressividades desde a nossa tenra idade. Darwin defende que o hábito de produzir “sons musicais foi inicialmente desenvolvido, entre os nossos antepassados, como meio de fazer a corte, tendo assim ficado associado às mais fortes emoções de que é capaz a nossa espécie (...) amor ardente, a rivalidade e o sentimento de triunfo”. O tom da nossa voz está então claramente relacionado com os nossos sentimentos e emoções, “Se uma pessoa comenta num tom queixoso os maus tratos que sofreu, ou se queixa de uma dor leve, quase sempre se exprime num registo agudo” (Darwin, 2006: 85).

Os estados emocionais têm uma influência directa em estados fisiológicos, os quais têm efeitos previsíveis na voz (são acusticamente perceptíveis), especialmente na frequência fundamental e aspectos temporais, relacionando-se, deste modo, com parâmetros de definição de qualidade de voz e prosódia (Cahn: 1990, Oudeyer, 2002).

Neste âmbito, falando de voz e emoção, ou na alteração da primeira recorrente de uma mudança na segunda, parece-nos claro o elo de ligação. Quando emocionalmente nos sentimos mal há reflexo no nosso estado físico. Se tivermos em linha de conta as mudanças físicas que ocorrem no indivíduo com baixa auto-estima, as alterações são visíveis ao nível da postura corporal: apresentando ombros caídos, cabeça baixa, postura contida, com muita tensão acumulada nos ombros e, concomitantemente, no pescoço e maxilares; ao nível fisiológico são perceptíveis alterações na respiração e ritmo cardíaco, que se torna mais fraco.

Variáveis psicológicas actuam de forma determinante na formação dos padrões de comunicação do indivíduo, podendo traduzir-se em causa, co-ocorrência ou consequência de um distúrbio de comunicação no qual a relação entre corpo e voz pode traduzir a “mutilação” do primeiro sobre a segunda. A este

conjunto de alterações da voz, vinculada a aspectos psico-emocionais, é atribuída a designação de disfonia psicogénica. Trata-se de uma desordem funcional, uma vez que o processo que conduz ao aparecimento e instalação de voz disfónica apresenta um simbolismo directo com a função fonatória da laringe. Alguns desvios em qualidade de voz (como frequência; intensidade ou entoação) podem ser explicados como distúrbios emocionais ou mesmo inaptações. Há aspectos da voz que são influenciados pelos sentimentos e atitudes do próprio indivíduo mas que também se modificam consoante a reacção que o falante percebe no receptor (Behlau et al, 2004). O stress é, por exemplo, um factor emocional que não podemos controlar. De acordo com Kiecolt-Glaser e Yehuda (2005) e Endler e Kocovski (2001), a ansiedade (emoção associada ao stress) acrescenta significado ao stress que acaba por conduzir a alterações físicas (tensão muscular; cefaleias) e emocionais (mal-estar, ansiedade, falta de sono) que se repercutem então na nossa qualidade de voz.

Pode surgir, em termos de anotação de emoções, se a análise se centrar apenas nos parâmetros acústicos, a dificuldade em distinguir entre o que é a expressão característica de um determinado estado emocional e o que é uma entoação prosódica, que pode ser característica da língua ou do próprio falante. Para contornar esta dificuldade, importa, segundo Olga Dioubina (2003) que se analisem vários parâmetros prosódicos que podem diferenciar emoções. Um destes prende-se com a acentuação (identificando quais as sílabas mais prolongadas e onde é dada maior ênfase na frase, por exemplo). Existem valores mais elevados de *pitch* e intensidade em enunciados expressando alegria e raiva.

2.3.1 Universalidade

Alguns dos parâmetros de análise das emoções aparecem com resultados idênticos em várias (ou mesmo todas) as línguas e outros parecem fazer parte das especificidades de uma língua, ou mesmo de um falante.

Já desde os anos 80 que a ciência tem estudado a relação existente entre os parâmetros acústicos em termos de fala e as emoções e expressões do indivíduo (Johnstone e Scherer, 1999). A investigação realizada até ao momento

tem mostrado que as emoções não são reconhecidas de forma igual por todas as línguas e culturas (Zovato et al 2004). Estes autores, utilizando testes perceptuais, perceberam que existem problemas de identificação, ou seja, dificuldades na diferenciação entre registos neutros vs. tristeza e alegria vs. raiva. Sawamura, num estudo publicado em 2007, reconheceu que existia quando da aplicação de um teste de percepção, muitas similaridades entre desgosto e raiva; surpresa e alegria e que medo é muitas vezes confundido com tristeza.

Shahid et al. (2008) numa pesquisa realizada na qual se pretendia estudar emoção verdadeira e a emoção simulada, compararam 50 falantes caucasianos (holandeses) e 45 paquistaneses agindo de acordo com determinada emoção. Atestaram que as emoções “representadas” ou “interpretadas” eram, nas duas culturas, percebidas como mais fortes. A pesquisa revela ainda que, em termos de observação da expressão corporal, os resultados podem estar muito próximos da realidade e são muito semelhantes entre as duas culturas. Mas já no que diz respeito à análise apenas pela voz, expressando diferentes emoções, verificou-se que os falantes holandeses eram mais sensíveis às diferenças, sendo que as emoções negativas eram mais facilmente percebidas por este grupo, enquanto que os indivíduos paquistaneses eram mais sensíveis e percebiam melhor as emoções de carácter positivo.

No que diz respeito à análise da frequência fundamental, Makarova e Petrushin (2003), verificaram que o F0 é mais alto nos enunciados de alegria, raiva, medo e surpresa do que nos de tristeza e neutros. Numa análise mais detalhada, verificaram que alegria apresenta os valores mais altos de F0 (próximo dos 160 Hz), sendo por isso a que se encontra mais distante da expressão neutra, é também aquela em que há uma maior oscilação entre os valores máximos e mínimos de F0. A expressão de tristeza, por outro lado, tem os valores de F0 mais baixos (100 Hz), sendo a que apresenta uma menor variação de contorno. Todas as outras emoções apresentam valores idênticos (próximos dos 120 Hz).

Os resultados conseguidos, de acordo com outras pesquisas, são idênticos em várias línguas, ou seja, será verdade que a alegria está associada a um aumento de intensidade e a um maior *pitch* e que a expressão de tristeza

apresenta os valores mais baixos para estes parâmetros. No entanto, parece-lhes certo que cada estado emocional se relaciona mais com determinado tipo de frase: por exemplo, a surpresa com frases do tipo interrogativo e o medo e a tristeza com proposições declarativas. Parece pois, que certos estados emocionais estão mais fortemente ligados a certos padrões e tipos de frase (Makarova e Petrushin, 2003).

2.3.2 Patologias de voz de base emocional

Parece-nos claro o elo de ligação que existe entre voz e emoção. Se não, não seríamos sensíveis à forma como a emoção que estamos a vivenciar altera a nossa voz, tornando-a perceptivelmente diferente.

Ao conjunto de alterações da voz, vinculada a aspectos psico-emocionais, é atribuída a designação de disfonia psicogénica. Uma desordem funcional, uma vez que o processo que conduziu ao aparecimento e instalação de voz disfónica, apresenta um simbolismo directo com a função fonatória da laringe. De acordo com Behlau et al (2004), alguns desvios em qualidade de voz (como frequência; intensidade ou entoação) podem ser explicados como distúrbios emocionais ou mesmo inaptações. Há aspectos da voz que são influenciados pelos sentimentos e atitudes do próprio indivíduo, mas que também se modificam consoante a reacção que o falante percebe no receptor. O stress é, por exemplo, um factor emocional que não podemos controlar.

A disfonia psicogénica apresenta diversas manifestações que podem ir desde a completa afonia até uma disfonia com diferentes características sonoras. É por isso fundamental estudar, nestes casos, o simbolismo, ou o que representa aquele sintoma vocal, conhecer a sua verdadeira história. Poderá ser necessário um seguimento por um psicólogo que trabalhará em conjunto com o terapeuta da fala na procura de restabelecer a voz e muitas vezes melhorar a auto-estima do paciente.

2.3.3 Estudos acerca da voz emocional

Os vários estudos que têm sido realizados para a emoção encontram-se maioritariamente baseados na relação entre: parâmetros acústicos, a atitude do indivíduo e emoções (Scherer, 1999). Em termos de emoções, sabemos que nem todas são fisiológicas mas, muitas delas, são controladas por pressões sociais (o sorriso social, por exemplo). Neste âmbito, de acordo com um estudo realizado por Baken e Orlikoff (2000), a alegria espontânea e o sorriso espontâneo, não alteram a qualidade de voz só pela quantidade de movimentos faciais (como pode acontecer com o sorriso ou alegria forçados), a alegria e o sorriso espontâneos fazem variar o F0 e a prosódia, provocam uma variação em termos de qualidade de voz, até mesmo porque um sorriso genuíno e um sorriso mecânico não são processados na mesma zona do cérebro e são possíveis de distinguir acusticamente. O sorriso espontâneo/genuíno não tem unicamente a ver com os movimentos faciais. Baken e Orlikoff (2000), não excluem, contudo, a existência de uma possível equivalência entre a expressão da face e a expressão da voz.

Sem se pretender ser exaustivo apresentam-se, de seguida, exemplos significativos de estudos na área da voz e emoção.

2.3.3.1 *Johnstone e Scherer (1999)*

Tom Johnstone e Klaus Scherer (1999), em dois estudos efectuados fazem uma análise, utilizando o electroglotógrafo (sendo precisamente aqui que se começa a implementar a análise detalhada dos momentos de abertura e fecho glotal na emoção, utilizando o electroglotógrafo (EGG)), o electrocardiograma e medindo a tensão muscular facial e a respiração (como factores articulatórios). Na análise acústica tiveram em conta o F0 e sua variação, o *jitter* e a energia. Conforme concluíram, a medição directa dos parâmetros acústicos do discurso, de acordo com a fisiologia do falante, através do uso do EGG, traz resultados mais fidedignos e imediatos. A simples análise das características de abertura e fecho da glote provou ser fundamental na interpretação das particularidades de cada uma das emoções, de acordo com a análise da onda acústica. Este estudo,

direccionado à análise de emoções induzidas, atesta que o uso de jogos de computador, para indução de emoções, parece ser um dos caminhos promissores e viáveis no processo de recolha/tratamento de *corpora* para a emoção.

O tempo de abertura da glote não mostrou (em percentagem) diferenças significativas entre as variadas emoções, o mesmo não se observando no que respeita ao tempo de fecho. Assim, emoções caracterizadas por terem um F0 elevado têm também um fecho mais rápido da glote. Este movimento acelerado é sinal de um aumento do esforço vocal e/ou, segundo os autores, de uma maior tensão muscular da laringe (Tom Johnstone e Klaus Scherer, 1999).

A informação sobre a amplitude (dados que podemos encontrar na onda) é a energia da vibração das cordas vocais, relacionando-se com a força destas. Essa informação é usada para controlar o “stress” colocado numa palavra e a ênfase que usamos quando falamos (Baken; Orlikoff, 2000).

A estrutura temporal corresponde ao trabalho das cordas vocais. Ela é responsável pela formação do ritmo no discurso e criação das pausas, sendo importante na detecção das palavras mais importantes no contexto de determinado enunciado (Sato, 2004). Se aumentarmos o *pitch* em determinado enunciado, manipulando-o, aproximamo-nos claramente de uma voz de criança, “*childish*”, continuando a aumentar o *pitch* e mantendo a amplitude, pode-se, subitamente, produzir uma voz pouco natural tal como a do “*spaceman*”.

2.3.3.2 *Efeitos do stress, Cowie et al (2000)*

O stress, não sendo uma emoção, é contudo algo que influencia o nosso discurso, o nosso estado de espírito, o nosso modo de sentir e experienciar o que se vai reflectir na voz. Num estudo realizado cujo o enfoque eram os **efeitos do stress na voz de um indivíduo** evidenciou-se que será claramente vantajoso o reconhecimento e detecção, por parte de um sistema automático, dos níveis de stress na voz de um piloto de aviões, antes mesmo que os passageiros o possam constatar (Cowie et al, 2000). Isto porque analisando gravações de momentos antes de alguns acidentes é perfeitamente notório o stress em que o indivíduo se

encontra, apenas pela análise da sua voz. Quando o indivíduo se encontra sob o efeito do stress e analisando os parâmetros relacionados com F0, *jitter* e fonte glotal, percebe-se que existe um aumento do F0 que é simultaneamente mais irregular, havendo uma descida dos valores de *jitter*. O sujeito fala muito mais depressa e quanto mais depressa fala mais aumenta a tensão. Em termos de amplitude é perceptível uma maior intensidade bem como uma perturbação da amplitude (*shimmer*) que se torna mais alta. Existe uma maior concentração de energia nas frequências mais altas, com modificações várias e complexas na frequência de cada formante. O stress afecta ainda o modo de articulação, uma vez que se fala de forma muito mais rápida e tensa; a tendência é ser menos preciso no que se diz; alguns sons acabam por ser produzidos de forma incorrecta e desvozeada; há uma tendência para gaguejar e elidem-se algumas palavras. Em termos linguísticos gerais, podemos falar em estruturação de frases muito longas, mais pausas e mais hesitações.

2.3.3.3 *Baken e Orlikoff (2000)*

Num estudo realizado para o Finlandês por Baken e Orlikoff (2000), tentando clarificar se é possível distinguir manifestações de alegria e sorrisos sociais de expressões genuínas e espontâneas, percebeu-se que acusticamente a diferenciação é possível. No entanto, a imagem pode causar confusão, mostrando que esta é mais forte. Esta conclusão advém do facto de terem, inicialmente, gravado as várias expressões genuínas e mecânicas com a respectiva imagem e, trocando as imagens e o som, apresentando uma imagem de expressão de alegria com a audição de uma expressão forçada, denotou-se uma maior dificuldade na identificação precisa de cada uma das representações.

Salientam, no entanto, que é possível identificar uma manifestação de alegria só pela acústica (apesar da expressão facial ser bastante óbvia). A alegria revela sempre uma maior intensidade (seja mecânica ou genuína) do que o discurso neutro em que a frequência é mais baixa, sendo sempre mais marcada do que o sorriso. O que se torna interessante é a análise em termos de acústica em manifestações em que a expressão física é tão forte.

Por outro lado, referem que em discurso a expressão de alegria espontânea, em exteriorizações como riso e gargalhada, vem normalmente no final não o interrompendo, afirmam mesmo que o sorriso durante o discurso pode significar mais do que uma gargalhada (Baken; Orlikoff, 2000).

2.3.3.4 *Gobl, Bennett e Chasaide (2002)*

Como é que a qualidade de voz pode fazer variar o F0 de acordo com a expressão de emoção e atitude, ou seja, qual é o papel da qualidade de voz em aspectos linguísticos e paralinguísticos da comunicação, foi o estudo realizado por Christer Gobl; Eva Bennett e Ní Chasaide (2002). Verificaram a variação da qualidade de voz através da manipulação de F0. Fizeram corresponder a cada emoção uma qualidade de voz diferente que consideravam mais adequada, e realizaram ainda um teste de percepção a pessoas alheias ao estudo. O contorno do F0 para momentos de indignação foi combinado com uma voz rouca; enunciados de raiva e alegria ficaram relacionados com uma voz tensa; medo com um tipo de voz sussurrada; aborrecimento com voz “*lax-creaky*” (estridente mais relaxada) e o sentimento de tristeza com uma voz aspirada. A ausência de sentimento foi combinada com o contorno de F0 da voz modal. Verificaram que a combinação de qualidade de voz mais F0 (QV+F0) foi a que trouxe resultados mais satisfatórios. O simples estímulo só alterou o F0 e não a qualidade de voz. Com a alteração dos dois parâmetros a percepção, por parte dos ouvintes, foi bastante mais clara, conseguindo identificar melhor os estados emocionais.

Assim, a qualidade de voz aparece como sendo realmente um factor crucial na classificação/diferenciação de estados emocionais, sabendo desde já que a simples manipulação de F0 é relativamente ineficaz na identificação destes parâmetros. Há uma dimensão prosódica na qualidade de voz que é pouco entendida mas que existe paralelamente à entoação (Gobl; Bennett; Chasaide, 2002).

2.3.3.5 Makarova e Petrushin (2003)

No estudo realizado sobre a Fonética da Emoção na Língua Russa por Makarova e Petrushin (2003) percebeu-se a necessidade das características acústicas na expressão de emoções e afectos como indispensáveis na interacção Homem-máquina, Robótica, sistemas de tradução e de entretenimento e mesmo no ensino de uma língua estrangeira. Neste caso, revela-se como aspecto fundamental o conhecimento, por parte dos professores, das idiossincrasias e especificidades acústicas da língua que ensinam e da língua dos alunos, facilitando a comunicação e entendimento entre ambas as partes. A máquina deveria ser capaz de perceber as emoções Humanas e poder, dessa forma, responder-lhes adequadamente (é dado como exemplo o cão robot da Sony, para interagir com crianças e o Robot enfermeiro da Mitsubishi que se espera que possa ajudar a pessoa em qualquer situação).

Neste estudo, constituiu-se um *corpus* de 10 frases com diferentes estruturas sintácticas e tipos de discurso. Estas foram lidas por 61 falantes nativos de Russo, tentando expressar seis estados emocionais diferentes: neutro, surpresa, alegria, raiva, tristeza e medo (Makarova e Petrushin, 2003).

Examinando o contorno do *pitch* das frases percebeu-se que o contorno deste é relativamente próximo nas frases que expressam alegria e nas que revelam raiva, sendo também estes tipos de enunciado os que apresentam maior tempo de vozeamento. É nestes casos que a duração da elocução é mais lenta do que para as restantes emoções. Os enunciados manifestando tristeza e medo aproximam-se em estruturas prosódicas apresentando, por exemplo, os mesmos valores em termos de vozeamento. Ressalvam ainda que a duração na expressão de medo apresenta valores altos, podendo significar que pode haver várias formas de o expressar. Os dois tipos de enunciado (medo e tristeza) aproximam-se do neutro, se tivermos em conta a análise da intensidade. No entanto, revelam que o discurso neutro e a surpresa se distanciam dos outros estados afectivos/emocionais, apresentando características específicas.

2.3.3.6 Campbell e Mokhtari (2003)

Num estudo realizado por Nick Campbell e Parham Mokhtari (2003), agruparam-se as diferentes formas de falar do seguinte modo: Educada (“*polite*”); amigável (“*friendly*”) e casual para cada uma das classes: família; amigos e outros, tentando verificar se, de acordo com o grau de afectividade que temos com a pessoa ou a intenção do nosso discurso, a nossa qualidade de voz varia. Foram produzidos 24 actos de fala de diferentes categorias, dos quais foram analisados 5: dar informações; exclamações; pedir informação; murmurar e pedir repetições. Verificou-se que a voz é mais aspirada quando nos dirigimos aos outros (de forma mais educada), seguindo-se a voz que usamos para falar com crianças (de forma suave). Um discurso dirigido a nós mesmos (um monólogo) revela um menor valor de voz aspirada ou suave. Falar com a família exhibe também um nível elevado em termos de aspiração, mais elevado do que quando se fala com os amigos. O F0 é mais elevado quando se fala directamente para uma criança e mais baixo quando se fala para membros da família, excluindo-se uma vez mais as crianças. Torna-se claro que o F0 e a forma mais ou menos aspirada da voz pode ser controlada pelo falante, de acordo com o contexto e interlocutores com que interage.

Campbell e Mokhtari (2003) conseguiram perceber diferenças no tratamento dos diferentes membros da família em termos da nossa qualidade de voz. Assim, partindo dos níveis mais elevados de voz aspirada e F0 para os níveis menores verificaram que: a voz mais suave e com um concomitante F0 mais elevado se verifica quando a interlocutora fala com a filha, seguindo-se o sobrinho, depois o pai, a mãe que se iguala em termos de qualidade de voz à irmã mais velha; tia e, por fim, o marido. Isto parece indicar o grau de cuidado que é tido com cada elemento da família. De acordo com grau de preocupação vs. mudança de voz percebeu-se também que é quando falamos para nós mesmos que existem os menores valores em termos de F0 e voz suave. No que diz respeito aos actos de fala analisados, verificou-se que existe um cuidado em utilizar uma voz mais suave quando se “pede uma informação” do que quando se “dá uma informação”. Por outro lado, percebe-se que murmúrio (“*muttering*”) tem

níveis de F0 e uma qualidade de voz dentro de outros parâmetros, pois tem um F0 significativamente baixo e uma suavidade muito mais nítida. Com esta pesquisa, foi possível verificar que o grau de preocupação, relação com o outro e o contexto social fazem variar a qualidade de voz, a qual consegue ser controlada, mudando o significado de uma frase, pela alteração de marcas paralinguísticas.

2.3.3.7 *Lee et al. (2005)*

Para além das características acústicas que caracterizam e diferenciam cada emoção existem também **particularidades articulatórias** inerentes e próprias a cada estado emocional. Na realidade, existem poucos estudos sobre a expressão da emoção no discurso, particularmente, no que diz respeito à articulação. É importante também verificar as alterações físicas (para além das prosódicas e acústicas) visíveis nos movimentos da língua, do maxilar inferior e na posição dos lábios, consoante a emoção que expressamos. Estas particularidades foram tidas em conta numa descrição apresentada no estudo de Lee et al. (2005), em que se centralizou a análise na produção das vogais. Gravaram 14 frases, (repetidas 5 vezes, expressando quatro emoções diferentes: neutro, raiva, tristeza e alegria) por um falante nativo do inglês americano sem qualquer tipo de formação em teatro ou em áreas de estudo/educação de voz. De acordo com os resultados obtidos neste estudo, e descrevendo apenas as características articulatórias, quando expressamos tristeza denota-se um avanço e elevação da ponta da língua e avanço do maxilar, sendo estes movimentos lentos, quando comparados a um discurso em que a emoção predominante é a raiva. Nesta, os movimentos da língua, particularmente no sentido vertical, são muito mais rápidos, bem como a abertura do maxilar. A alegria pode-se comparar ao discurso neutro sendo, no entanto, aquela que apresenta uma maior variação do *pitch*. Descrevem assim, sumariamente, a raiva como hiper-articulação, por oposição à hipo-articulação que caracteriza a tristeza. Os autores chamam ainda a atenção para o facto de muitas das variações articulatórias poderem ser características inerentes ao falante.

A lacuna em estudos anteriores é a falta da descrição dos padrões de entoação presentes em cada uma das emoções, uma vez mais o grande relevo é dado à medida de *pitch* (médio, máximo, variação).

2.3.3.8 Vroomen et al. (1993)

O trabalho realizado para a língua holandesa por Vroomen et al (1993) procurava analisar os contornos de *pitch* de frases com emoções diversas e classificá-las de acordo com padrões de entoação, para conseguirem um conjunto de regras de entoação e duração de forma a que fosse permitida a criação de actos de fala expressando emoção. O estudo contou com a colaboração de um actor que gravou duas frases diferentes, simulando sete emoções distintas (neutra, alegria, aborrecimento, raiva, tristeza, medo e indignação) que foram analisadas acusticamente. Constituiu-se uma base de dados com enunciados de diferentes emoções, para verificar se factores como duração e entoação são elementos relevantes. Na segunda tarefa, formularam-se regras sobre como a entoação e a duração podem gerar um enunciado com emoção partindo de um discurso neutro.

Na primeira experiência efectuada, com o objectivo de aferir se parâmetros como *pitch* e/ou duração seriam suficientes para classificar e diferenciar as várias emoções, copiaram o contorno de *pitch* e, simultaneamente, a duração dos enunciados em que eram expressas as diferentes emoções para um enunciado monótono. Obtiveram três alvos de análise distintos: introduzindo-se só a duração; só o *pitch*, ou alteração de *pitch* e duração. Com esta experiência verificaram que introduzindo no discurso monótono a estrutura de tempo conseguem-se enunciados análogos à expressão de aborrecimento, raiva e discurso neutro. Aplicando apenas os parâmetros relacionados com entoação das distintas emoções são possíveis distinguir diferenças na frase monótona que levam a identificar o aborrecimento, a tristeza, o medo e a indignação. Com a introdução dos dois parâmetros os resultados são claramente mais satisfatórios. Verificou-se que é, aparentemente, possível expressar várias emoções, em

termos de sintetizador, controlando apenas o tempo (duração) e o *pitch*, sem que seja necessária uma alteração da composição do espectro dos enunciados.

No segundo ponto do trabalho, baseando-se em diversas análises acústicas e perceptuais, os autores verificaram que todos os enunciados que expressavam uma ou outra emoção obedeciam às regras gramaticais de entoação do Holandês. A partir daqui foi possível substituir os contornos de *pitch*, dos enunciados com emoção, por um contorno padrão gerado a partir das regras dessa gramática. O contorno original poderia ser substituído por um contorno artificial sem afectar profundamente o conteúdo emocional da mensagem. Aplicando a um enunciado neutro os métodos de compressão e expansão do PSOLA (*Pitch-Synchronous Overlap and Add*), (um método de manipulação do *pitch* e da duração de um sinal acústico da fala), foram alterando os contornos de *pitch*, os padrões de entoação e duração. Com esta experiência, puderam verificar que os casos particulares de emoções como tristeza e medo foram os menos reconhecidos em relação a todas as outras emoções.

A terceira experiência realizada nesta investigação procurava aferir se estes resultados se mantinham quando em discurso sintetizado com o *LPC-coded diphone*. Os enunciados foram sintetizados, os fonemas foram convertidos em *LPC-coded diphones* e as regras de duração foram aplicadas a esses difones, procurando atingir um enunciado que se aproximasse da ausência de emoção (frase neutra). Seguidamente, este enunciado foi expandido ou comprimido (através do PSOLA); as sílabas acentuadas foram marcadas manualmente e a entoação adequada foi computadorizada. Assim, os difones foram sintetizados e a duração e entoação foram manipuladas. De todos os enunciados 63 por cento foram correctamente identificados pelos 12 ouvintes, o que demonstrou ser um bom resultado, concluindo que as características prosódicas das emoções podem ser controladas e sintetizadas. Os resultados mostraram que as emoções podem ser expressas de forma bem clara manipulando *pitch* e duração, “Emotions are thus signalled by an ensemble of prosodic variables which can be controlled in synthesized speech” (Vroomen et al, 1993).

2.3.3.9 Yanusgevsckaya et al (2008)

Para mostrar como a qualidade de voz varia de acordo com a expressão de afecto e com a intensidade com que falamos, Yanusgevsckaya et al. (2008) desenvolveram um projecto onde apresentaram duas séries de estímulos: uma em que variava a qualidade de voz com variações de altura intrínsecas e outra série onde existia uma qualidade de voz constante mas fazia-se variar o volume com que os enunciados eram produzidos. Concluíram que apesar de não ser a intensidade com que falamos um factor imprescindível para identificar e perceber diferentes afectos, não deixa de ser contudo relevante na medida em que é uma forma de activação (descritiva) de certos afectos.

Questões de prosódia como *pitch*, qualidade de voz, intensidade e tempo têm sem dúvida um papel fundamental na expressão da emoção e um papel fundamental na comunicação. Quando o falante varia a qualidade de voz há em concomitância uma variação de intensidade. O estudo destes autores partiu da manipulação de estímulos em termos dos parâmetros da fonte glotal. Por exemplo, de acordo com Yanusgevsckaya et al. (2008), uma voz “*lax-creaky*” apresenta valores muito altos em emoções como cansaço; tristeza e relaxamento, enquanto que uma voz tensa apresenta estas características na expressão de raiva; alegria e stress. Acrescentam ainda que podemos produzir uma voz modal com diferentes graus de intensidade.

Seguidamente, apresentamos na Tabela 8 a descrição dos resultados deste estudo em relação a diferentes afectos ou emoções.

Tabela 8 – Relação emoção-qualidade de voz (Yanusgevsckaya et al, 2008).

Emoção	Qualidade de voz	Intensidade
Relaxado, triste; aborrecido; intimidado	Estridente (mais relaxada) (<i>Lax-creaky</i>)	Sussurro
Envergonhado	Aspirada	Sussurro
Assustado	Sussurrada	Sussurro
Indignado	Áspera	Tenso
Stressado; formal; interessado; alegre; destemido	Tensa	Tenso

Pode verificar-se de que forma cada emoção está associada a diferentes qualidades de voz e como se verifica esta na intensidade que imprimimos ao nosso discurso.

2.3.3.10 *Trabalhos para o Português (Europeu)*

Para o Português Europeu, existe um estudo realizado por Guimarães e Abberton (2004) para a obtenção dos valores de F0 dos falantes portugueses. Utilizaram para o efeito 52 indivíduos disfónicos e 57 normofalantes para assim verificarem as diferenças no que diz respeito aos valores da frequência fundamental. Nesta tarefa os informantes dos dois grupos produziram de forma sustentada as vogais [a], [i] e [u] e realizaram a leitura do texto “A história do rato Artur”. Para além disso realizaram também discurso espontâneo na resposta a perguntas como: “O que fez ontem?” e “Como chegou até ao hospital?”. Para a recolha de dados realizaram a aplicação de um questionário clínico, exame estroboscópico, avaliação perceptiva, acústica e electroglotográfica de voz. Os resultados deste estudo demonstraram que os falantes disfónicos apresentam valores de F0 mais baixos do que o outro grupo em todas as tarefas propostas, mostrando-se estatisticamente significativa quando falamos em termos de género e tarefa. O género feminino apresenta valores de média de F0 mais altos. Na tarefa de discurso a vogal [a] apresentou os valores mais baixos (em comparação com as outras vogais). Os valores de F0 são mais elevados nas vogais sustidas do que nas tarefas de conversação ou leitura. Ainda assim, esta última apresenta valores mais elevados do que conversação. Apresentam-se na Tabela 9 os dados retirados deste estudo para os valores médios de F0 do género feminino, de acordo com a idade, para o Português Europeu.

Tabela 9 – Valores médios de F0, para o género feminino, tendo em conta a idade e situação clínica.

<i>Idade</i>	<i>Corpus</i>	<i>Situação clínica</i>	<i>F0 (Hz)</i>	<i>Situação Clínica</i>	<i>F0 (Hz)</i>
19-40	[a]	Sem patologia	212.5	Disfonia	190.4
	[i]		225.5		208.2
	[u]		230.1		203.4
	Leitura		177.5		175.7
	Conversação		180.7		168.3
41-67	[a]	Sem patologia	210.2	Disfonia	207.0
	[i]		216.5		217.5
	[u]		222.6		222.3
	Leitura		177.5		191.1
	Conversação		180.7		190.0

Numa perspectiva completamente diferente, a da síntese, foi realizado recentemente um trabalho de Mestrado no INESC-ID/IST onde se procuraram implementar regras acústicas para poder simular sete emoções básicas (raiva; alegria; tristeza; surpresa; aborrecimento e desagrado), partindo do discurso neutro. Criaram assim um sistema denominado EmoVoice. A tentativa que se continua a fazer é a de desenvolver ao máximo a investigação para ter cada vez melhores resultados que facilitem a comunicação/compreensão homem/robot. Nas suas pesquisas, análises e simulações, face a escassez de estudos para o Português Europeu, utilizaram um corpus em língua alemã, sendo que a bibliografia consultada foi também ela de trabalhos realizados dentro do âmbito do estudo de voz e emoção para outras línguas. O sistema EmoVoice gerou sete emoções, sendo que as que foram depois melhor reconhecidas foram a expressão de raiva; alegria; tristeza e medo. Aborrecimento foi confundido com tristeza e surpresa confundida com raiva; alegria e medo. Aversão não foi reconhecida por nenhum ouvinte. Espera-se agora que este estudo venha a ser alargado com a transformação de outras características acústicas (pausas; articulação e parâmetros espectrais) e novos testes terão de ser conduzidos para que para avaliar o desempenho do sistema na produção de emoções básicas em termos de fala sintetizada (Cabral e Oliveira, 2006).

Não havendo muitos estudos sobre este tema para o Português, pelo menos a que tenhamos tido acesso, referimos o trabalho realizado no âmbito do seu Mestrado por Américo Rodrigues na criação de um pequeno *corpus* de voz

emocional produzida por um actor (*corpus* que utilizamos no nosso estudo e no qual colaborámos na anotação). Os parâmetros analisados não incluem os que estudamos no nosso trabalho, à excepção do F0, uma vez que estuda mais as características acústicas. Assim o estudo de emoções apresentado por Américo Rodrigues inclui os resultados para as medidas de duração, F0 e articulação no Português Europeu, como se pode verificar mais à frente no capítulo 4 deste trabalho.

Capítulo 3 - Estudo de voz de uma patologia de base emocional

Se os olhos são o espelho da alma, então a voz é o barómetro das emoções humanas.

Kerry Erie, M.Cl.Sc¹²

A voz e as suas modificações, nomeadamente as que estão relacionadas com emoção, são o enfoque deste nosso trabalho. Assim, analisamos alterações de qualidade de voz que acontecem de forma incontrolável no indivíduo sem que este apresente qualquer tipo de anomalia anatomo-fisiológica. Nesse sentido encontramos neste terceiro capítulo

¹² Master in Clinical Science.

o estudo de qualidade de voz e patologia, quando a diferença na qualidade de voz é motivada por factores de índole psicológica.

Apresentamos uma descrição de patologias psicogénicas e estudos realizados por diversos autores, numa perspectiva de enquadramento teórico e (re)conhecimento de problemas relacionados com a voz. Efectuámos pesquisa, recolha e análise de casos, criámos *corpora*, aplicámos ferramentas de estudo e extracção de parâmetros que nos permitiu fundamentar o estudo sobre qualidade de voz como indissociável dos problemas patológicos que com ela se relacionam.

Adoptamos a designação de disfonia psicogénica, para definir este tipo de patologia vocal. Esta revela-se como uma manifestação de um desequilíbrio psicológico como ansiedade, depressão, desordens de personalidade ou reacções de conversão de tal forma que fazem com que o normal controlo de fonação seja perdido (Baker, 2002 e Rubin e Greenberg, 2002). O termo disfonia inclui todos estes fenómenos, menos a “rouquidão” que se liga à forma como a voz nos soa (Seifert, 2006).

Com este estudo, pretende-se apresentar uma primeira análise de uma disfonia psicogénica, no caso particular de um falante que é capaz de produzir dois registos distintos: um claramente disfónico e outro mais controlado. Procurámos perceber e identificar as diferenças entre este caso, com patologia diagnosticada, e um falante com uma voz saudável. Procurou-se aferir quais os parâmetros mais passíveis de controlar, comparando ainda o registo mais controlado com as vozes disfónica e a normal.

3.1. Disfonia psicogénica

Numa disfonia psicogénica podem ser observadas configurações e posições anormais durante o esforço de fonação, tendo assim uma vocalização em falsete, por exemplo. As pregas vocais não demonstram ter movimento de adução; podem apresentar uma hiper-adução com o envolvimento das bandas ventriculares, o que vai comprimir as estruturas supraglóticas.

A maioria das descrições dentro deste tipo de estudo, de características laringoscópicas de uma disfonia psicogénica, foi realizada entre os anos de 1870 e 1960 por laringologistas como Oliver (1870); Ingals (1890); Smurthwaite (1919); Sokolowsky (1944) e Aronson (1960), citados por Baker (2003).

O termo psicogenia indica a etimologia da palavra reportando-nos para um fenómeno de natureza psicológica mas isto não pode ser visto como uma regra, uma vez que a tensão muscular tem um importante papel na descrição da disfonia. Os músculos intrínsecos e extrínsecos da laringe são extremamente sensíveis ao stress emocional e a sua hiper-contracção é o denominador comum de todas as disfonias psicogénicas (Aronson, 1990 citado por Baker, 2003). Estudos recentes confirmam que as disfonias psicogénicas são mais recorrentes nas mulheres, geralmente numa razão de 8 para 1 (mulheres/homens) (Baker, 2003).

As **disfonias psicogénicas** encontram-se divididas em dois grupos de acordo com o seu grau de recorrência e manifestações. As mais comuns apresentam normalmente os seguintes sintomas:

a) **Disfonia por tensão muscular**, associada a stress emocional. Esta disfonia está associada à tensão muscular excessiva da laringe que leva a uma hiperadução das pregas vocais e a uma constrição da laringe (Imamura et al. 2006).

b) **Reacções por conversão** que existem apesar da estrutura e funcionamento neurofisiológico se apresentarem normais. São, usualmente, despoletadas por stress, depressões, ansiedade ou conflitos interpessoais. Este tipo de disfonia funciona como um mecanismo de defesa que aparece “postvirus or trauma”, podendo ainda estar relacionada com um abuso das cordas vocais. Os pacientes podem, muitas vezes, apresentar uma total afonia, mutismo, falsete com quebras de voz, ou uma qualidade de voz em falsete *continuum*. Permite que o sujeito permaneça parcialmente inconsciente dos conflitos, stress ou qualquer outro tipo de instabilidade, funciona como uma tentativa de se libertar emocionalmente. Normalmente estes pacientes estão convencidos da natureza física da sua patologia (Ziegler, 1970, citado por Baker, 2003)

Ainda em termos de reacções de conversão, Freud falava em dois tipos de reacção: a que se dava por existir uma situação insuportável/inultrapassável (e que se baseava num trauma do passado) e o outro tipo de reacção teria que ver com impulsos inaceitáveis (Freud, 1920, citado por Baker, 2003). Curiosamente, nunca juntou os dois tipos de reacção e a psiquiatria e psicoanálise seguem a sua teoria de realidades psicológicas e de impulsos inconscientes que têm de ser reprimidos, como base da perturbação psicológica. Dentro deste tema, Baker (2003), no seu estudo, percebeu que o problema principal é a saúde mental e que, por isso, é necessário estar em contacto, conhecer a história do paciente e simultaneamente acompanhar a sua vida presente. Segundo a autora, quando as pessoas experienciam memórias muito fortes, as suas mentes podem-se tornar incapazes de saber reagir a essas situações. Observou também que, quando as memórias de determinado trauma não são verbalizadas ou exteriorizadas de alguma forma, ficarão inevitavelmente fixas enquanto não forem encaradas e expostas.

c) **Voz de falsete** predominante em jovens aquando da normal mudança de voz, a passagem da voz mais imatura para uma voz mais forte, apresentando um *pitch* mais baixo próprio da maturidade (Baker, 2002).

No plano das disfonias psicogénicas menos comuns existe alguma controvérsia quanto ao seu enquadramento no campo de patologias verdadeiramente relacionadas com a voz (Aronson, 1990). Estas são diagnosticadas, tendo em conta os comportamentos seguintes:

a) **Adultos que apresentam ainda uma voz não formada**, imatura. Os indivíduos que utilizam uma voz em falsete com um recuo da língua com uma concomitante expressão facial que sugere desejo de aprovação (Baker, 2002).

b) **Indivíduos que falam muito baixo**, provocando no receptor um esforço acrescido para ouvir;

c) **Indivíduos que falam excessivamente alto**, provocando o afastamento do seu ouvinte. De acordo com Aronson (1990), estas manifestações prendem-se mais com uma escolha intencional do falante do que propriamente com uma verdadeira disfonia. Trata-se de uma escolha intencional, um estilo de

comunicação, muito mais do que um verdadeiro caso de disфонia. Estas escolhas têm bases psicológicas muito fortes e em termos de relacionamentos humanos esta quebra na comunicação pode ser tão significativa como uma disфонia psicogénica. Ainda como exemplo de disfonias menos comuns existem: **Casos crónicos** de tosse, pigarreio e espirro na ausência de qualquer doença respiratória.

A qualidade de voz, quando o indivíduo sofre de uma disфонia psicogénica pode apresentar grandes variações de um momento para o outro e pode mesmo confundir-se com disfonias orgânicas como vozes roucas; aspiradas; contraídas ou sussurradas (Baker, 1998). Nestes casos, para a análise da qualidade vocal, deve existir complementarmente uma análise de otorrinolaringologia e um estudo perceptivo da voz, tendo em conta os aspectos acústicos da mesma, como sejam: frequência fundamental, *jitter*, *shimmer* e ruído (*noise to harmonic ratio* – NHR) (Seifert, 2006).

Mesmo quando existe uma inflamação média ou pequenas mudanças estruturais como nódulos vocais, a extensão da visibilidade da patologia é muito menor ou ausente quando comparada com o grau de severidade de uma voz “anormal” (Aronson, 1990, citado por Baker, 2002).

Para se perceber o desenvolvimento de uma patologia de voz não-orgânica, é essencial que o especialista não se guie apenas pelas impressões acústicas e perceptivas ou por algumas lesões das pregas vocais. O estudo deve ser muito mais aprofundado, procurando factores de origem psicossomática que possam estar na origem dessa disfunção. Importa conhecer o contexto em que vive o indivíduo e possíveis factores pessoais, culturais, sociais que tenham feito desenvolver o seu problema vocal.

Os terapeutas fazem as suas análises pedindo aos pacientes que produzam vogais sustentadas e a partir daqui realizam estudos e diagnósticos, tendo em conta a estabilidade, amplitude e qualidade de voz. Este método faz parte do protocolo clínico, uma vez que se trata de uma tarefa de avaliação de voz facilmente observável para o especialista, sem grandes dificuldades de produção para o indivíduo com patologia, tornando-se assim eficaz e eficiente. Tendo em

consideração que o contexto em que ocorrem as vogais influencia as funções da laringe, dados os fenómenos de co-articulação, importa de facto saber se os resultados de cada parâmetro, conseguidos com as análises clínicas, apresentam semelhanças ao discurso oral normal, para que se possa generalizar um diagnóstico e daí partir para a caracterização de diferentes patologias.

É importante fazer notar que neste momento existem ainda poucas informações e estudos sobre informação acústica de vozes psicogénicas, pelo que a sua caracterização se torna mais difícil, neste aspecto.

3.2. Trabalhos relacionados

Procurámos contextualizar o nosso trabalho, tendo em conta os trabalhos já existentes, mesmo que para outras línguas, como é o caso, e que nos vão servindo de guia. Como pouco é conhecido sobre características de voz psicogénica, os estudos que aqui apresentamos reflectem sobre outras patologias vocais e servem-nos aqui para enquadramento da nossa pesquisa.

É difícil ao ouvinte ter a sensibilidade necessária para distinguir as variações das perturbações de frequência e amplitude. Estes parâmetros não permitem a distinção entre patologias médias ou severas, uma vez que a oscilação dos valores de *jitter* se encontra entre os 0.1 dB para uma patologia vocal média e os 2.57 dB numa patologia severa. O *shimmer* apresenta uma variação entre os 0.28 dB e os 1.23 dB. Estes valores contrastam significativamente com os de ruído, tornando estes bastante mais perceptíveis na descrição e distinção de patologias. Estes variam entre os 40.4 dB numa patologia considerada média até aos 6.93 dB numa patologia mais severa. Existe uma grande relevância na tentativa de identificação das diferentes características acústicas inerentes ao sinal de voz. De acordo com Kreiman e Gerratt (2003) tais abordagens revelam-se úteis mesmo quando aplicadas simultaneamente a múltiplas facetas de estímulos multidimensionais.

Importa saber quais as características que permitem um melhor acesso e, por isso, análise da irregularidade e ruído adicional inerente às vozes com patologia. Existem muitas e diferentes categorias como “características

aperiódicas” que foram usadas para descrever a percepção de voz áspera e a periodicidade da vibração das cordas vocais (citando para estes estudos Hirano et al., 1998; Hillenbrand, 1988; Arends et al., 1990; Dejonckere, 1995) e “características de ruído” que foram referidas por alguns investigadores como indicadores de voz aspirada (Michaelis et al, 1998 citam Hammarberg et al., 1981; Klatt and Klatt, 1990; Hillenbrand et al., 1994; Dejonckere, 1995; Hillenbrand and Houde, 1996). As definições de ruído podem ainda estar relacionadas com certas condições psicológicas durante o ciclo fonatório como a abertura glotal (*glottal gap*) ou fechamento suave.

Com o aumento da aperiodicidade do sinal as medidas perdem o seu verdadeiro significado e valor enquanto indicadores dos desvios da periodicidade (Titze and Liang, 1992). Assim, *jitter* e *shimmer* não podem ser determinados de forma exacta em vozes com grandes distúrbios (Michaelis et al, 1998).

Considerando que a escolha e interpretação das medidas acústicas apropriadas é um problema ainda a resolver, Kreiman e Gerratt (2003) realizaram duas experiências. Na primeira fase 20 indivíduos (10 homens e 10 mulheres) gravaram a vogal /a/ com uma duração de 1 segundo. Os informantes apresentavam diferentes patologias de voz, variando entre disfonias médias a severas, subdivididos de acordo com os diagnósticos médicos. Posteriormente, 17 indivíduos com audição normal ouviram as gravações de uma voz natural e uma cópia sintetizada desta sem *jitter*, *shimmer* ou ruído. Tinham que ajustar os valores desses parâmetros, manipulando-os até que as características da voz sintética coincidissem com as da voz original.

As respostas dos indivíduos apresentavam grande diversidade no que concerne ao *jitter* e *shimmer*. Estes parâmetros foram menos ajustados e utilizados, contrastando com a manipulação dos valores de ruído. Quanto pior era a voz (no que diz respeito à severidade da patologia diagnosticada) mais fácil se tornava fazer a estimativa de ruído, que aparece como factor essencial, em termos perceptuais na distinção e definição de qualidade de voz. No entanto, não se verificou nas respostas esta mesma correlação entre a gravidade da patologia e a variabilidade dos parâmetros de *jitter* e *shimmer*.

São apontadas duas possíveis explicações para estas divergências: os indivíduos podem ser insensíveis às diferenças em termos de aumento/total de *jitter* e *shimmer* ou, por outro lado, até podem ser sensíveis a essas alterações mas incapazes de determinar qual o nível correcto uma vez que não conseguem, em termos de percepção, separar estes dois parâmetros do ruído (Kreiman e Gerratt 2003).

Num estudo realizado para o Português Europeu (Constantino e Guimarães, 2005), realizou-se a avaliação vocal de 18 pacientes do sexo feminino, sendo que 12 sofriam de disfonia há menos de um ano e 6 há mais de um ano. Os autores analisaram os parâmetros acústicos de regularidade e irregularidade (F0); ruído e o nível de impacto psicossocial da voz. Tinham como grandes objectivos verificar se existem diferenças significativas no F0, *jitter*; *shimmer*; NNE e impacto psicossocial de acordo com a duração da disfonia e o seu diagnóstico (nódulos *versus* quistos). Os resultados mostraram que não existe uma diferença que possa ser estatisticamente significativa entre os dois grupos para todos os parâmetros acústicos (exceptuado NNE na vogal [u]). Verificou-se um menor impacto psicossocial da voz nas mulheres com disfonia há menos de um ano. Por outro lado, as mulheres com disfonia há mais de um ano apresentam um F0 mais grave em vogais sustentadas e mais elevado para fala encadeada. Os valores de *jitter* revelaram-se mais elevados nas mulheres com nódulos há menos de um ano (para as vogais [a] e [i]); os valores de *shimmer* mostraram ser mais elevados nas mulheres com disfonia há menos de uma ano, com a excepção para a vogal [a]. Em ambas as patologias esta é a vogal que apresenta valores mais indiciadores de patologia em termos de *shimmer*, o que está de acordo com López et al. (2000) e Gouveia (2004) citados por Constantino et al (2005). O domínio físico é mais elevado, seguido do funcional e do emocional em ambos os grupos. Pode-se dizer que existem diferenças significativas entre os domínios físicos e os funcionais e entre os físicos e os emocionais.

3.3. Metodologia

3.3.1 Corpus

O *corpus* utilizado neste estudo foi definido após uma prévia avaliação realizada ao informante (pela terapeuta), uma vez que era importante saber quais as suas limitações. Tínhamos que ter em conta as suas capacidades e facilidades para a realização das frases/sons que foram propostos e apresentados. Após este processo de conhecimento das características do informante, foi definido o *corpus* a gravar.

Foi gravado um *corpus* em duas partes distintas. A primeira contemplava as vogais cardinais, produzidas de forma sustida, [a]; [i] e [u]; a segunda parte, a frase: “O Toneca toca no pateta” proferida nos tipos Declarativo, Exclamativo e Imperativo, sendo cada frase repetida três vezes em cada registo de voz.

No que concerne às vogais analisadas: [a]; [i]; [u], tivemos em consideração as que são normalmente analisadas em trabalhos do género (Guimarães e Abberton, 2004). Estas ocupam os pontos mais afastados no triângulo acústico, e, por outro lado, representam uma maior estabilidade em termos de fonação o que, obviamente, contribui para uma melhor avaliação da estabilidade da laringe e das cordas vocais. A maioria dos estudos que relacionam a perturbação do F0 com qualidade de voz, tem-se baseado sempre em vogais sustidas, como já foi referido e suportado no capítulo anterior deste trabalho. No entanto, pode-se referir que, no estudo realizado com crianças por McAllister et al. (1998), foram gravados enunciados completos uma vez que as vogais pareciam algo (ainda mais) estranhas para as crianças, o que podia causar comportamentos fonatórios atípicos. No entanto, também estes autores chegam à conclusão que o facto de terem utilizado discurso corrido lhes limitou as medidas de selecção, nomeadamente, e tal como no nosso trabalho, a extracção dos valores de *Harmonic Noise Ratio* (HNR).

Na segunda parte do *corpus*, constituído por uma frase, também se levou em linha de conta as observações que se queriam fazer, em relação ao uso das cordas vocais e, respectiva diferenciação, entre os dois tipos de voz (que no caso eram produzidos pelo mesmo falante). Na frase, “O Toneca toca no pateta”, as vogais ocorrem sempre entre consoantes surdas oclusivas [p]; [t]; [k], que não fazendo vibrar as cordas vocais, facilitam o processo de segmentação do sinal.

3.3.2 Informantes

Gravaram-se as produções de dois falantes nativos do Português Europeu, sexo masculino, com uma média de 35 anos de idade e formação superior. O primeiro indivíduo apresenta uma patologia diagnosticada como de origem psicológica que o leva, de forma habitual, a utilizar as falsas cordas vocais na fonação (voz de bandas ventriculares) com marcada hiperfunção ou hipercinésia laríngea. Este aumento da tensão músculo-esquelética induz a um comportamento aberrante da voz, requerendo, por isso mesmo, um maior esforço na sua produção. De acordo com os estudos e análises levados a cabo, a patologia de voz aqui descrita enquadra-se na disfonia por conversão. Assim, o mesmo *corpus* foi gravado, nas suas duas partes, por dois informantes, um sem qualquer tipo de patologia de voz e outro com patologia diagnosticada.

A causa psicogénica desta patologia, sem um problema diagnosticado em termos anatomo-fisiológicos (exames médicos não detectaram qualquer anomalia), deixa a possibilidade ao falante de produzir também um registo (mais) normal, usando as pregas vocais. Neste segundo registo, é notável o esforço e concentração acrescidos que o falante tem que realizar. A possibilidade de haver dois registos, produzidos por um mesmo indivíduo, torna este um caso de particular interesse para análise.

O nosso segundo informante, que não tem nenhuma patologia de voz diagnosticada, repetiu e produziu o mesmo material.

3.3.3 Gravação

As produções do informante com disfonia psicogénica foram recolhidas em ambiente hospitalar, em sala não insonorizada e na presença do terapeuta que o acompanha. Os registos de voz do segundo falante foram realizados no laboratório de fonética da Universidade de Aveiro.

Em ambos os casos foi recolhido o sinal de voz, utilizando o Electroglotógrafo (EGG), para facilitar a análise do funcionamento das cordas vocais. Os registos de voz foram gravados directamente para o computador, usando o Kay Elemetrics CSL 4400 e o *Speech Filing System* (SFS). Tanto o SFS como o Praat¹³ (Boersma e Weenink, 2006; Boersma, 2001) foram utilizados na anotação e análise de voz.

3.3.4 Anotação

Para a anotação e análise de dados, utilizámos os programas SFS e Praat. Demonstrou ser particularmente importante o uso da funcionalidade do Praat *Voice Report*, com o qual podemos extrair todos os dados (como é explicado na secção seguinte). Para análise estatística foi usado o SPSS (***Statistical Package for Social Sciences***). Extraímos do programa de anotação do Praat, um exemplo, representado na Figura 15, que se refere ao espectrograma da vogal sustida [a], produzida pelo informante com disfonia psicogénica.

Podemos ver neste segmento a informação sobre *pitch* a azul; intensidade marcada pelo contorno amarelo e as formantes evidenciadas a vermelho.

¹³ *Praat* é um programa para a análise e síntese de voz criado por Paul Boersma e David Weenink do *Department of Phonetics of the University of Amsterdam*.

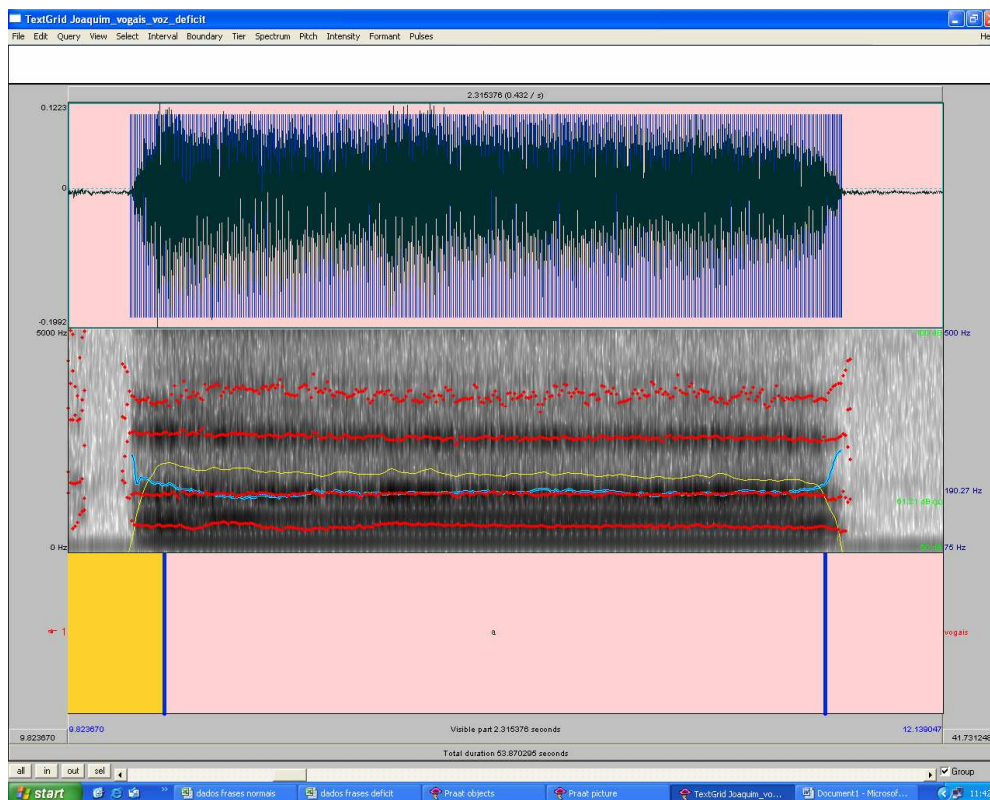


Figura 15 – Espectrograma, retirado do Praat, da vogal sustida [a].

3.3.5 Extracção e análise dos parâmetros

A funcionalidade *Voice Report* do Praat foi particularmente útil. Esta funcionalidade, como se pode verificar pela Figura 16 permite a extracção dos parâmetros mais relevantes e necessários para uma análise mais precisa de voz, tais como medidas de F0, *jitter*, *shimmer* e valores relacionados com a relação harmonia/ruído HNR.

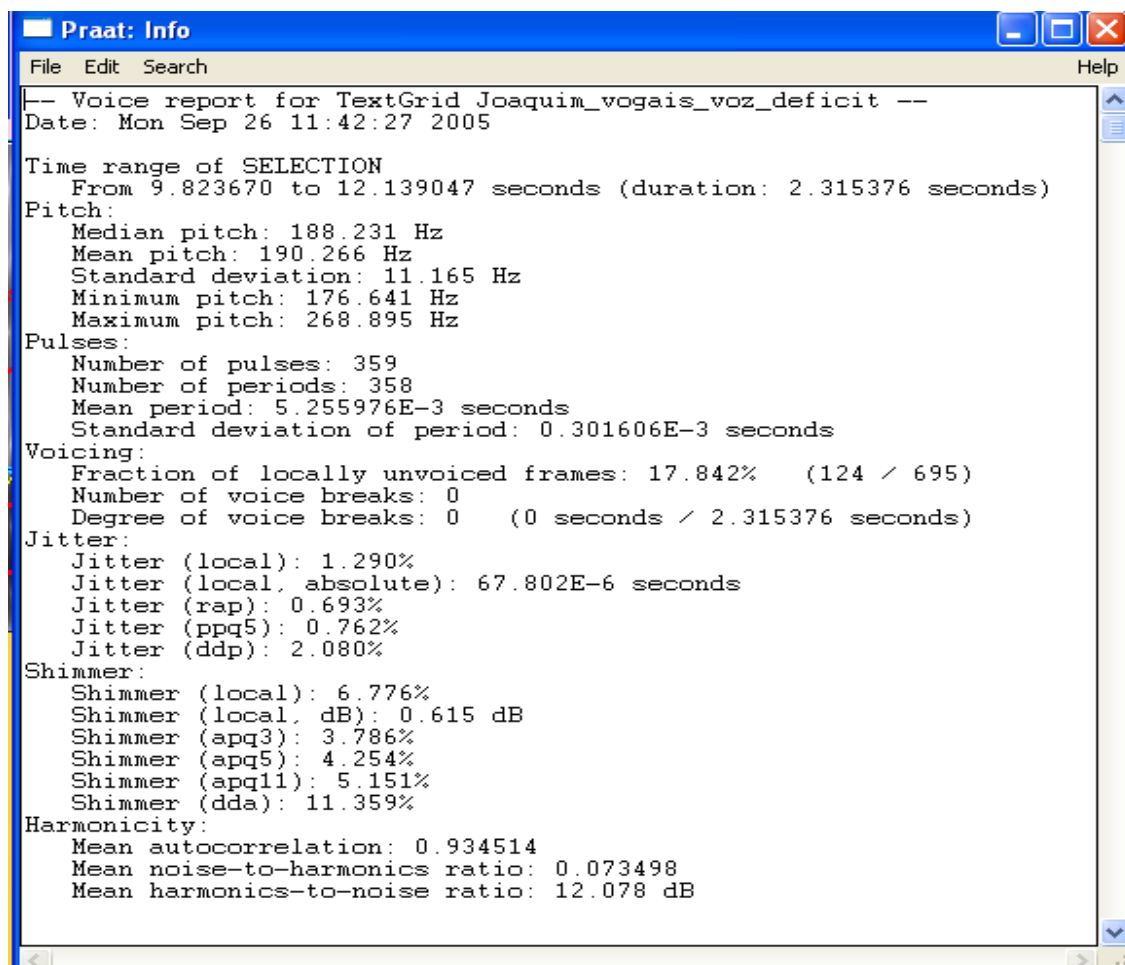


Figura 16 – Parâmetros do Voice Report do Praat relativos à vogal sustida [a].

Utilizámos, para além do Praat, o *Hoarseness Diagram* (Frölich, 1997), um diagrama que permite uma ilustração gráfica, bidimensional, dos parâmetros acústicos da qualidade de voz (turbulência, *jitter* e *shimmer*), Figura 18. No eixo horizontal encontra-se representada a combinação de parâmetros relacionados com a irregularidade (*jitter*, *shimmer* e ACR); no eixo vertical temos a indicação referente à componente de ruído, *glottal to noise excitation ratio* (GNE).

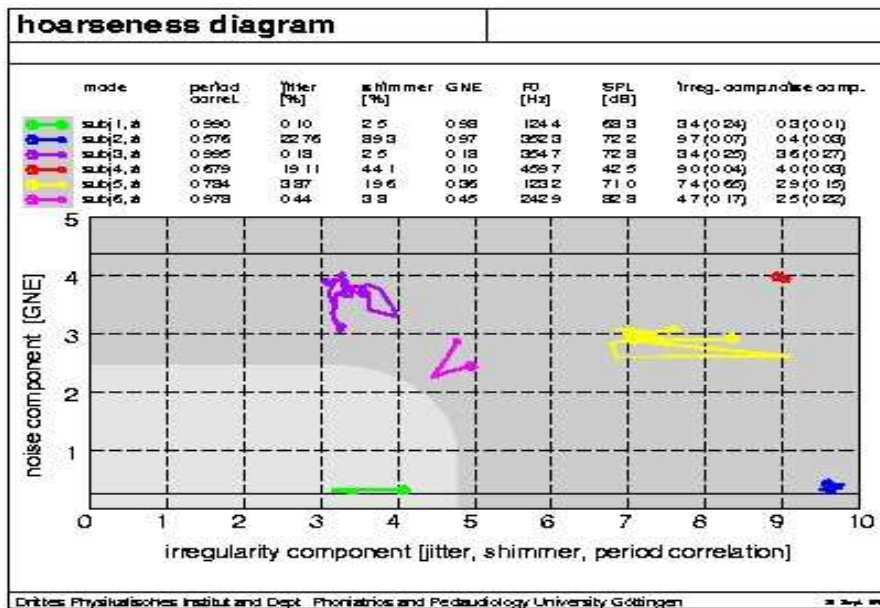


Figura 17 – *Hoarseness Diagram*, (Fröhlich, 1997)

3.4. Resultados

Os nossos resultados, neste estudo específico, são apresentados de acordo com o seu posicionamento no *hoarseness diagram*, tendo em conta os diversos parâmetros relativos à frequência fundamental, irregularidades (*jitter* e *shimmer*) e, para concluir, uma medida relacionada com ruído, o HNR.

3.4.1 Posicionamento no *Hoarseness Diagram*

A nossa descrição baseou-se nos resultados apresentados por Fröhlich et al (2000) “*Acoustic Analysis with the Hoarseness Diagram*”. Nesses resultados podemos ver valores diversos para diferentes mecanismos de fonação, tendo como pontos de referência: problemas fonatórios; fonação normal e afonia.

De acordo com os autores, todos os resultados para vozes normais e com patologia orgânica ou funcional podem ser interpretados utilizando as duas coordenadas do *hoarseness diagram* (representação e informação, a duas dimensões, da irregularidade e ruído): numa coordenada calcula-se o ruído

existente, GNE, na outra apresentam-se as medidas da periodicidade do sinal: *jitter*, *shimmer* e *mean period correlation*.

Devido a requisitos do programa, que pede segmentos com mais de 500 milissegundos, só podemos aplicar o *Hoarseness Diagram* às vogais sustentadas. Assim, esta pesquisa recai apenas nas diferentes produções das três vogais. Os resultados obtidos podem ser observados na Figura 19.

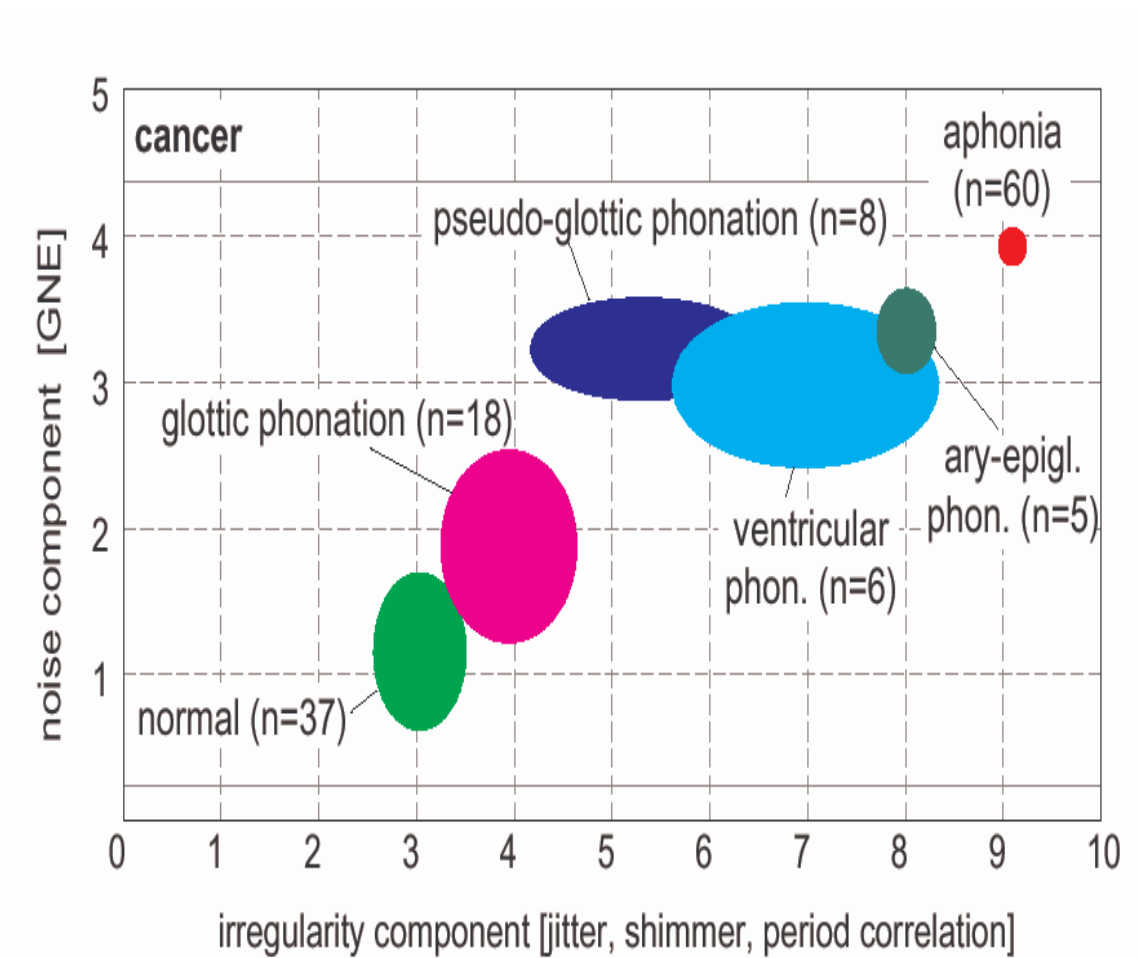


Figura 18 - Localização de patologias no *Hoarseness Diagram* (Fröhlich, 1997).

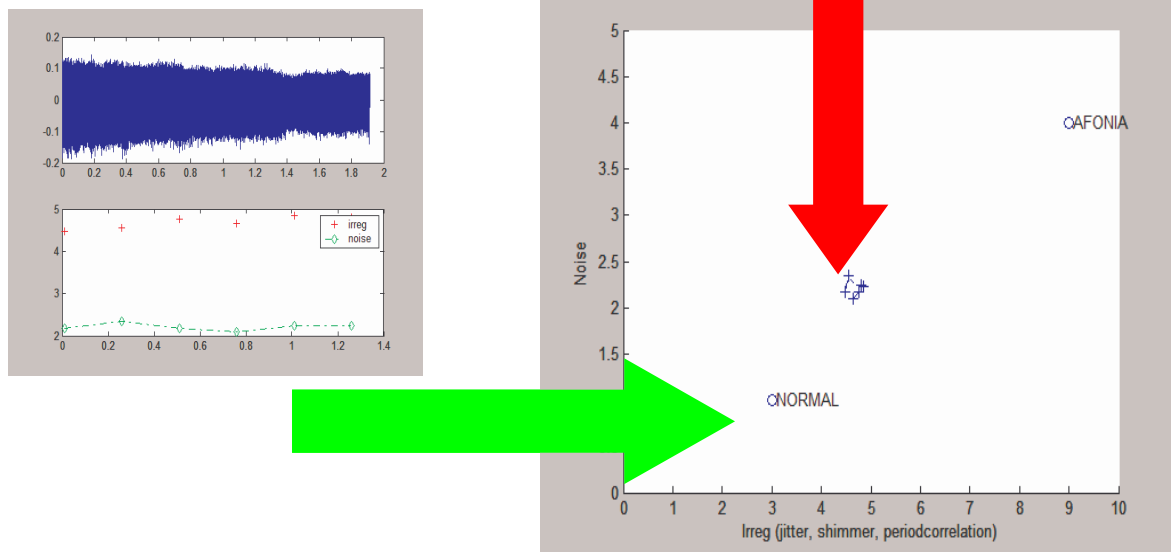


Figura 19 - Localização, de acordo com o *Hoarseness Diagram*, da patologia analisada e descrita: Assinalado pela seta verde temos a localização da fonação normal, pela seta vermelha o local onde se encontra no diagrama a fonação psicogénica.

O caso específico da patologia de voz que estudámos mantém-se numa zona muito próxima de uma fonação pseudo-glótica, como podemos observar na Figura 18 e Figura 19.

Baseámos a nossa análise e extracção de parâmetros naqueles que são considerados os mais relevantes em termos de descrição de qualidade de voz: F0, *jitter*, *shimmer* e HNR. A representação gráfica de todos os resultados é apresentada com intervalos de confiança de 95%, considerando dois factores: tipo de voz (patológica; normal e controlada) e o tipo de discurso/material analisado (frases ou vogais sustentadas).

3.4.2 Frequência fundamental – F0

Analisaram-se os vários parâmetros de F0: Média de F0; máximo e mínimo e F0 standard (Figura 20)

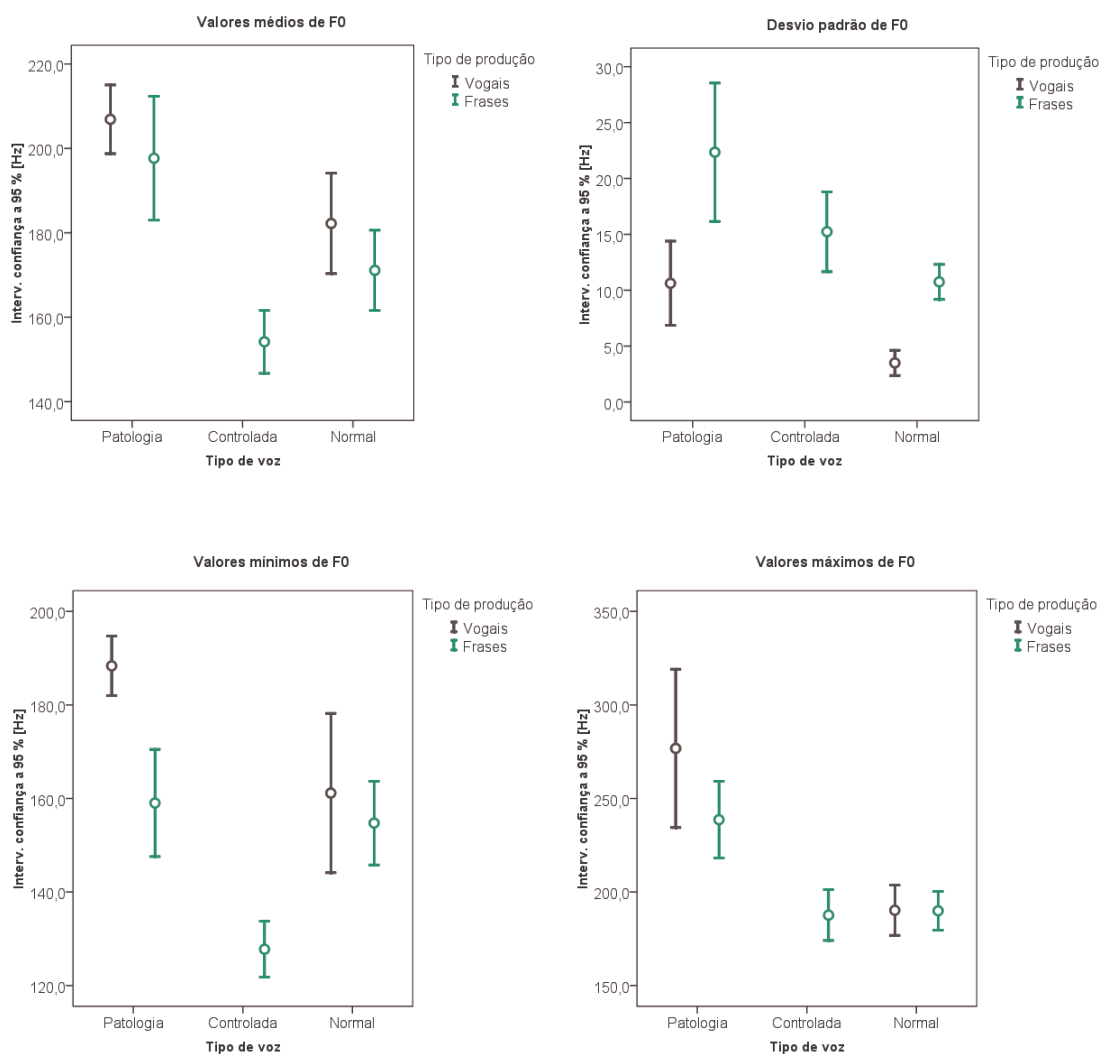


Figura 20 – Parâmetros relativos a F0: média, mínimo, máximo e desvio padrão.

Observando as figuras percebemos que os valores de F0 para a voz controlada do nosso informante se aproximam dos de uma fonação sem patologia nos parâmetros médio e máximo.

Na média de F0 percebe-se uma descida dos quase 200 Hz (da sua voz com patologia) para os 150 Hz no seu registo controlado. O mesmo se verifica no F0 mínimo onde existe uma descida dos 150 para os 130 Hz. Percebe-se também que o informante ao utilizar o seu registo mais controlado, o que implica um esforço e concentração adicionais, atinge valores ainda mais baixos do que os apresentados pelo nosso informante sem patologia de voz diagnosticada.

Conforme podemos ver no gráfico, o nosso indivíduo com patologia de voz, consegue produzir no seu registo controlado, para o F0 máximo, os mesmos valores que o nosso outro falante com uma voz normal, perto dos 180 Hz.

Há uma clara e significativa diferença entre os dois registos do primeiro informante quando analisamos os valores para F0 máximo. A mesma frase é produzida com valores que ascendem os 240 Hz, na sua fonação disfónica e, num discurso controlado, está abaixo dos 200 Hz.

Para o F0 standard podemos verificar que, em termos gerais, o informante com patologia de voz é capaz de se aproximar dos valores de uma fonação normal utilizando o seu registo controlado.

Comparando os 3 tipos de voz, vemos que os valores são muito mais altos nos parâmetros referentes à média de F0, F0 máximo e standard, na produção com patologia diagnosticada. Ao mesmo tempo, é interessante verificar que no registo controlado a diferença não é significativa quando comparada com a fonação normal, aproximando-se bastante desta, principalmente nos valores máximos de F0. Consegue também controlar os valores de F0 (valor médio F0)

3.4.3 Jitter

Neste estudo analisámos 3 parâmetros diferentes de *jitter*: *Local*; *Relative Average perturbation* (RAP) e *Period Perturbation Quotient* (PPQ5). Os resultados podem ser observados na Figura 21.

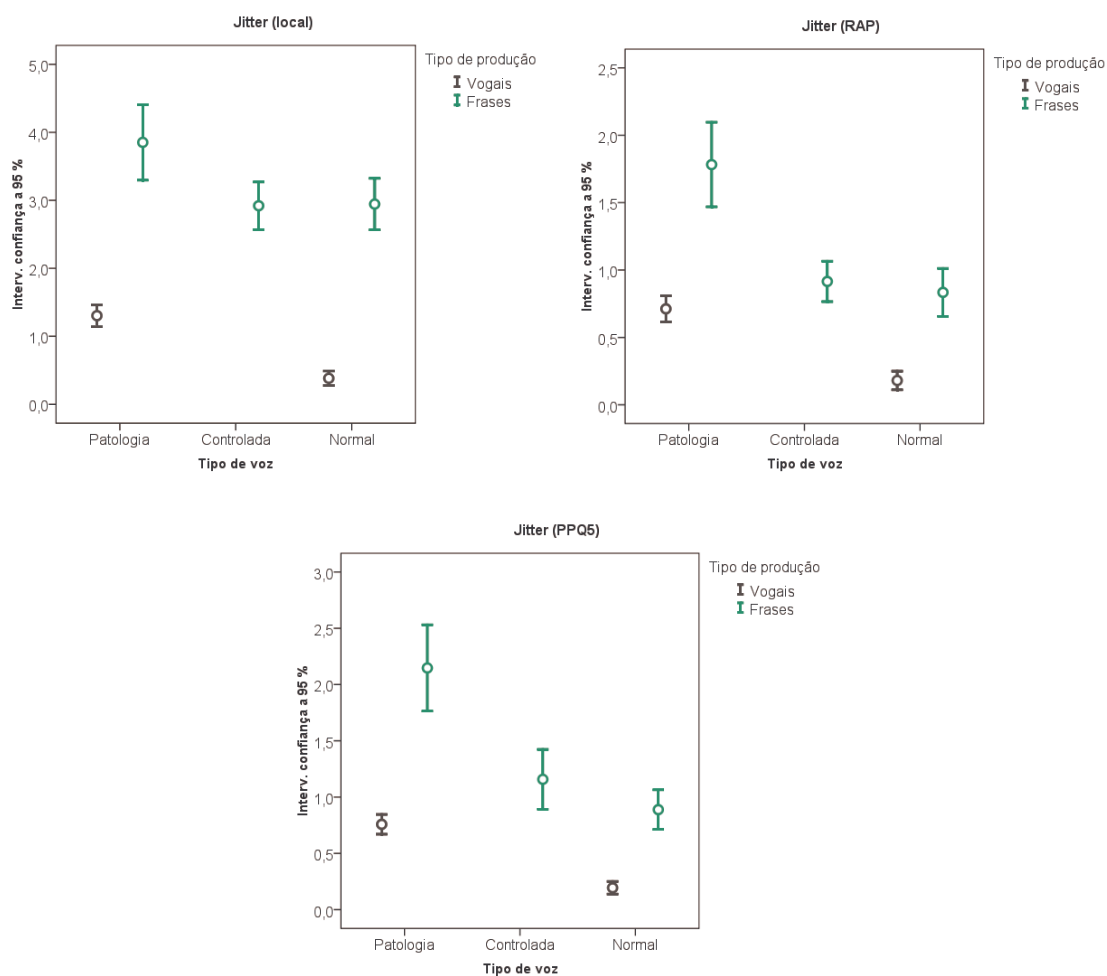


Figura 21 – Intervalos de confiança para três parâmetros de *Jitter* (local, RAP e PPQ).

Verificamos, pela análise da Figura 21, que na produção de frases nos dois tipos (com patologia e normal) os valores apresentam-se muito mais altos do que nas vogais.

Existe uma clara distância entre os valores apresentados para o *Local jitter* entre a voz com patologia e a voz normal, o mesmo não acontecendo em RAP e PPQ5, onde os valores são muito próximos.

Os nossos resultados, para *jitter*, mostram que o informante com patologia de voz é capaz de produzir um registo controlado em que os valores analisados descem para médias normais.

Todos os valores de *jitter* para a voz normal são sempre mais baixos do que em qualquer outra produção.

3.5.1 Shimmer

Os valores limite de *shimmer* são os 3%, este parâmetro serve para quantificar curtos períodos de tempo na instabilidade da amplitude “that does not alter the qualitative features of the vocal waveform”, Titze (1995, p.6). A relação entre perturbação da amplitude com anormalidades específicas do funcionamento glotal ou de outras perturbações mais globais permanece ainda muito pouco clara e explicada, parece, contudo claro, de acordo com (Verdolini-Marston, Sandage, and Titze, 1994; Shaw, Searl, Young and Miner, 1996 citado por Behlau, 2004 pp.130). que os valores de *shimmer* tendem a normalizar à medida que a patologia laríngea melhora.

Na nossa análise, na Figura 22, tivemos em conta os parâmetros: *Local* e *Amplitude Perturbation Quotient (APQ3)* para os valores de *shimmer*.

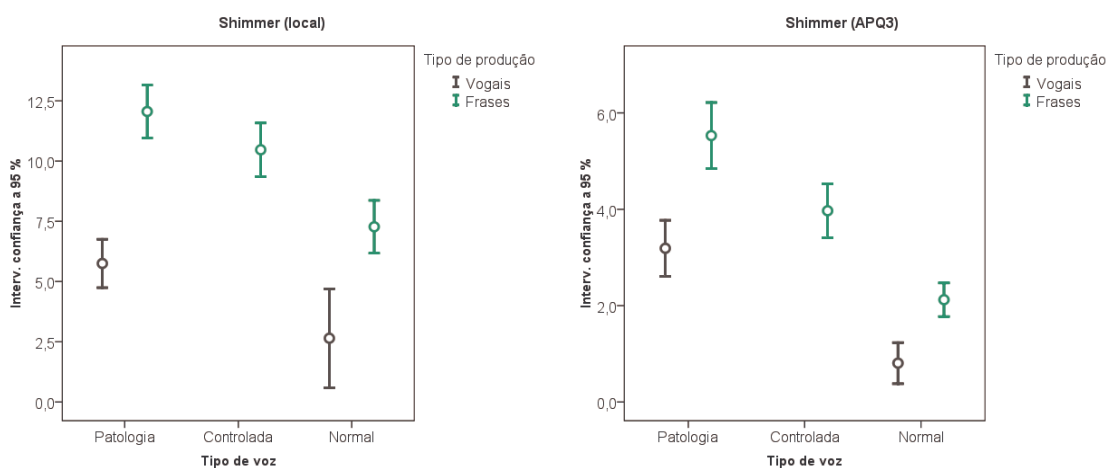


Figura 22 – Intervalos de confiança a 95% para dois parâmetros relativos ao *Shimmer* (local e APQ3) em função do tipo de voz e tipo de produção.

Pelos gráficos da Figura 22, podemos observar que mesmo no seu registo controlado o primeiro informante não consegue aproximar-se dos valores de uma

fonação normal, (próxima dos 1%, nas vogais sustidas e dos 2% nos enunciados), mantendo sempre registos muito mais elevados.

O informante com disfonia mantém valores muito próximos entre a produção com patologia e a controlada para o *shimmer* local e, embora consiga descer esse registo em APQ3 para perto dos 5,5%, está ainda longe dos valores para uma fonação normal, na ordem dos 2%.

3.5.2 Harmonic Noise Ratio (HNR)

HNR é uma medida acústica para obtenção dos valores de ruído. É um parâmetro que mede a componente de ruído no sinal de voz (Michaelis et al, 1998 pp.1628). Muito embora esta seja uma medida importante quando se fala em patologias de voz, importa também referir que este parâmetro tem sido apenas utilizado na análise de vogais sustidas e que as medidas de um discurso fluente são, normalmente, muito mais altas, do que nas vogais (McAllister et al, 1996, pp.252-261).

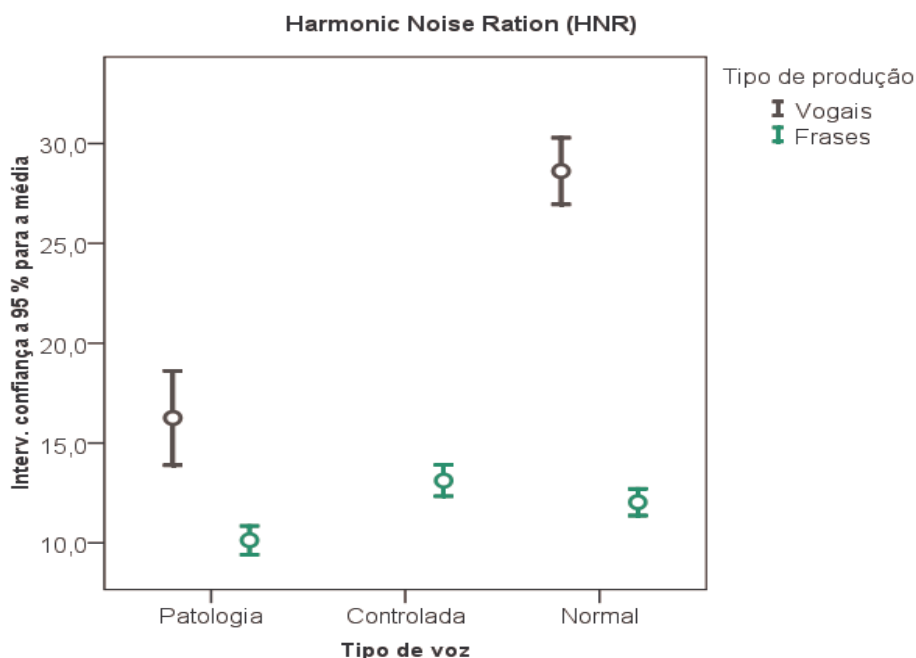


Figura 23 – Intervalo de confiança a 95 % para HNR em função do tipo de voz e tipo de material/produção analisado.

Nos nossos resultados, apresentados na Figura 23, é possível observar que as vogais produzidas com uma fonação normal apresentam valores perto dos 27 dB, por isso mais altos do que as disfônicas, com 17 dB, mas perto do registo controlado. Não temos, no entanto, valores precisos no que concerne à análise das frases, uma vez que são todos muito similares, embora consigamos observar que a voz com patologia apresenta valores baixos, perto dos 9 dB, não é contudo uma diferença significativa.

Encontrámos alguns problemas ao gravar as vogais em registo controlado, uma vez que, com o tipo de patologia apresentado, foi muito difícil ao nosso informante produzir vogais sustidas.

3.5. Discussão

A nossa análise centra-se, principalmente, na descrição/definição da patologia do nosso primeiro informante. Apesar do *corpus* ser limitado, uma vez que, como referimos, tivemos em linha de conta as limitações do informante, pudemos ainda assim verificar uma clara distinção entre as duas qualidades de voz produzidas pelo mesmo falante, uma vez que entre a voz com patologia e a controlada existem diferenças em diversos parâmetros. Numa produção que lhe exigia um maior poder de concentração e esforço, usando as cordas vocais, o informante foi capaz de controlar o valor médio de F0, mas incapaz de corrigir outros como o *shimmer*.

Ao realizarmos uma análise em que tivemos em conta os valores do *Voice Report* do Praat e posterior aplicação do *Hoarseness Diagram* obtivemos uma informação mais detalhada na caracterização da qualidade de voz do falante com patologia diagnosticada. No entanto, salientamos que a distinção, em termos de qualidade de voz, permitida pelo *Hoarseness Diagram* não é aplicável quando falamos em discurso continuado (frases). Parte dos parâmetros apresentam valores muito diferentes entre as medidas resultantes de enunciados mais longos e as das vogais sustidas, o que coloca problemas à sua interpretação, levantando dúvidas quanto à sua utilidade numa análise que vá para além do segmento. Verificamos, ainda assim que, mediante os nossos resultados, no *Hoarseness*

Diagram existe um distanciamento em relação à voz do informante sem patologia, ocupando uma zona próxima à de fonação pseudo-glótica.

Os resultados obtidos para *jitter* mostram que os valores são muito mais elevados na produção de discurso fluente do que em vogais sustentadas o que vai de encontro ao estudo de Baken et al. (2000).

O parâmetro mais relacionado com o ruído HNR mostra que a voz patológica apresenta uma perda de harmonia e níveis elevados de ruído, estando de acordo com o que é referido por McAllister et al. (1996). No entanto, no nosso trabalho, as medidas de HNR não apresentaram resultados importantes, uma vez que não existe uma diferença significativa entre valores nos três registos de voz.

Dada a carência de estudos de medidas acústicas para patologias de voz, este estudo pode talvez vir a ajudar a demonstrar que este tipo de análise é importante, uma vez que permite a verificação e identificação das diferenças entre vários tipos de patologia. Torna-se neste aspecto de particular interesse o uso do *Hoarseness Diagram* e das versatilidades do Praat que nos permitem extrair todos os parâmetros e medidas relacionadas com voz.

3.6. Comentários Finais

Este estudo faz uma amostragem e breve explicação de uma patologia psicogénica mostrando como a emoção pode influenciar a voz, transformando-a.

A nossa análise trata um caso relevante e curioso em termos de disfonia psicogénica. Um indivíduo adulto, que por razões emocionais, desenvolveu uma disfonia mas que, curiosamente, consegue (embora com bastante esforço e capacidade de concentração) ter uma outra qualidade de voz, longe daquela que considerámos, após a nossa análise, de fonação pseudo-glótica.

Como foi exposto neste trabalho, e em comparação com a bibliografia consultada, muitas vezes qualquer trauma psíquico pode vir a afectar inconscientemente a voz do indivíduo. Quando não há uma base física ou estrutural que explique a patologia de voz, torna-se mais difícil explicar a dicotomia funcional vs. orgânica, uma vez que nada explica a desordem vocal

vivenciada. Como fomos verificando pelo nosso trabalho e de acordo com vários autores estudados, Baker (1998) e Seifert (2006), um acompanhamento permanente e em paralelo de um terapeuta e de um psicólogo torna-se importante nestes casos. É necessário conhecer toda a história do paciente para perceber os conflitos subjacentes à disfonia psicogénica.

Capítulo 4 - Estudo de voz emocional de um actor

Vários estudos têm vindo a tentar aferir da universalidade de **produção e percepção de emoções**. Sabendo que há parâmetros como *shimmer*, *jitter*, F0 e HNR que variam de acordo com a emoção e que essas variações são mais significativas em determinadas emoções (raiva e medo, por exemplo, falando no caso específico do Português Europeu - (PE)), torna-se importante saber de que forma são sensíveis a essas alterações indivíduos que desconheçam a língua e sua cultura e mesmo, em teste de percepção, qual é a performance dos falantes nativos do PE, para se poderem então comparar resultados e retirar daí conclusões. Tentámos pois perceber se, de facto, as emoções são percebidas da mesma forma por falantes de línguas diferentes. Pretendemos aferir se podemos, ou não, considerar que um falante de uma língua estrangeira, que nunca tenha estudado nem tido contacto com uma outra língua, consegue, apenas ouvindo frases que veiculam determinadas emoções, identificá-las. Isto pode permitir também saber, pela primeira vez para o PE, se aspectos linguístico-culturais, seu conhecimento ou desconhecimento, interferem na identificação/percepção de emoções.

Neste momento, existem grandes *corpora* para estudo da emoção no Finlandês (*Media Team Speech Corpus*) e para as consideradas “línguas maiores”: Inglês; Francês e Alemão. Para o Português Europeu, pouco ou nada existe neste campo. Espera-se com este trabalho contribuir para suprir essa lacuna em termos de dados, análises e conhecimento numa área que tem cada vez mais importância. A percepção da voz é importante para entender o estado emocional e intenções do nosso interlocutor. Ao contrário da expressão facial, a voz é relativamente independente e percebe-se a diferença mesmo à distância. Perceber toda a mensagem e linguagem emocional exige a análise e integração de uma variedade de pistas acústicas (Schirmer e Kotz, 2006).

Neste capítulo, apresentamos com um *corpus* seleccionado e gravado por um actor, uma análise acústica dos parâmetros relacionados com qualidade de voz, no Português Europeu e um teste perceptivo recaído sobre essas mesmas emoções.

O nosso objectivo, com o estudo apresentado neste capítulo, é perceber e explicar (pela extracção de parâmetros através do *Speech Filing System*¹⁴ e Praat) as diferenças e semelhanças que existem em termos de qualidade de voz em cada emoção analisada, comparando depois cada uma ao estado neutro. Pretende-se analisar o que varia, em termos de produção, na nossa voz e o que faz o nosso interlocutor perceber diferenças no nosso estado anímico, exclusivamente pela voz. O ponto fulcral é obter, de acordo com o nosso conhecimento, pela primeira vez para o PE, os valores para alguns parâmetros acústicos que caracterizam as emoções. Só assim poderemos comparar com os resultados obtidos para as outras línguas e aferir quais os parâmetros que

¹⁴ SFS é um material grátis, para utilização em investigações relacionadas com voz. Realiza operações padrão como aquisição; audição, marcação, apresentação de espectogramas e análises das formantes, da frequência fundamental. “It comprises software tools, file and data formats, subroutine libraries, graphics, special programming languages and tutorial documentation”. Department of Speech, Hearing and Phonetic Sciences, UCL Division of Psychology and Language Sciences.

seguem a tendência geral (se seguem) e quais são características intrínsecas à nossa língua. O ponto seguinte deste capítulo é a **aplicação de um teste perceptual** a dois grupos distintos (portugueses e americanos) para verificar como são percebidas as emoções por uns, falantes da língua, e por outros que nunca tiveram contacto com a mesma.

4.1. Voz emocional produzida por actores

Vários estudos na análise da qualidade de voz e emoção têm vindo a ser realizados e, apesar das diversas controvérsias e algumas hesitações quanto ao uso de actores na gravação dos corpora, estes continuam a ser o grande suporte da maioria dos estudos conhecidos para várias línguas. Pouco continua a ser sabido sobre a relação entre emoções “simuladas” e emoções espontâneas, de que modo diferem e em que escala é essa diferença significativa (Wilting et al, 2005). A experiência efectuada por Wilting et al (2005) vem demonstrar que o discurso emocional proferido por actores leva a melhores resultados em testes perceptuais, principalmente em expressões negativas (embora este estudo se baseie mais na expressão facial do que na voz)¹⁵.

Com o estudo que apresentamos neste capítulo, pretendemos obter para o Português Europeu os valores dos parâmetros que até agora temos vindo a analisar em termos de qualidade de voz. Para além de podermos conseguir dados já de referência, o estudo será complementado com um teste de percepção (verificando se o actor consegue expressar e fazer passar as emoções que lhe foram sugeridas) e, posteriormente, poderemos comparar com os valores que

¹⁵ Um dos pontos fulcrais desta investigação era fazer com que os informantes, mediante expressões faciais que íam observando no computador, as julgassem como positivas ou negativas, “ Seeing speakers producing acted emotional speech leads to more extreme perceived emotions scores than seing speakers produce real emotional speech”. A diferença entre simulação e produção real é particularmente mais forte para as emoções negativas (Wilting et al, 2005).

apresentaremos no capítulo 5 deste trabalho e que se referem à expressão espontânea de emoções.

A grande maioria dos estudos utiliza como *corpus* de análise gravações realizadas por actores, muito embora saibamos que estas têm muito do cunho pessoal do indivíduo que tem uma perspectiva artística, enfatizando as emoções para melhor se expressar. Contudo, sabemos da dificuldade e questões de ética que se podem levantar ao recolher registos de emoção espontânea. A gravação de profissionais que trabalham com a voz e com a emoção é pois o que mais se aproxima da realidade, fornecendo-nos valores que podem ser trabalhados e reconhecidos. Neste seguimento na nossa pesquisa realizámos testes de percepção, para aferir da “espontaneidade” e reconhecimento das emoções produzidas.

O estudo da forma como se podem conseguir algumas emoções induzidas (como stress, excitação ou depressão) é aprofundado e explicado por Scherer (2003). Este autor explica que através de drogas é possível induzir um aumento de certas emoções e cita o exemplo de Helfrich et al. (1984) que utilizou antidepressivos e foi analisando diversos parâmetros na produção da fala, durante o período de actuação destes. O stress também se induz apresentando ao informante uma tarefa que seja bastante difícil, num curto espaço de tempo, por exemplo (Scherer, 2003).

Contudo, a forma mais utilizada e preferida para obter amostras de expressão emocional é o recurso a actores, não havendo, no entanto, muitas dúvidas de que estas emoções apresentarão em alguns parâmetros valores muito mais elevados do que nas emoções induzidas ou espontâneas. Por outro lado, Scherer (2003) acredita que os actores são influenciados por estereótipos convencionais sobre a expressão de emoções¹⁶.

¹⁶ Mais estudos sobre o que é uma emoção, como se exorime e percepçiona uma emoção, podem-se encontrar num estudo alargado realizado por Scherer em 2005 “ *What are emotions? And how can they be measured?*”.

Drioli et al (2003) gravaram um corpus cujo objectivo seria através dele sintetizar emoções e ver quais os parâmetros que sendo alterados, podem levar a uma modificação não só da qualidade de voz mas também da emoção percebida. Utilizaram um informante adulto, do sexo masculino, que trabalhava recitação. Para a análise e extracção de parâmetros recorreram ao Praat Voice Report, analisaram parâmetros como F0, duração, intensidade, *jitter*, *shimmer*, HNR, etc. Os autores concluíram que nas vogais acentuadas, quanto à irregularidade, o valor mais alto de *shimmer* encontra-se na expressão de raiva e o mais alto de *jitter* em alegria e surpresa. Por sua vez, os valores de *Harmonic Noise Ratio* eram, precisamente, mais baixos para as emoções: raiva e alegria.

Na Tabela 10 podemos verificar os resultados em termos de duração, F0 e intensidade para as emoções (A) raiva; (D) aversão; (N) neutra; (J) alegria; (F) medo; (SU) surpresa e (SA) tristeza, de acordo com o estudo realizado por Drioli et al. (2003).

Tabela 10 – valores para duração, F0 e intensidade para as diferentes emoções analisadas nas vogais tónicas produzidas em /'aba/ e /'ava/.

Emoções	Duração (s)	F0 (Hz)	F0 média (Hz)	Intensidade (dB)
Raiva	0,195	177,744	18,276	76,735
Aversão	0,293	138,993	14,935	72,297
Neutra	0,231	126,428	14,588	70,819
Alegria	0,188	260,999	67,873	74,928
Medo	0,211	288,737	26,889	70,829
Surpresa	0,179	265,867	89,857	72,527
Tristeza	0,269	209,032	56,582	70,415

Vogt e André (2003) procuraram, através da sua pesquisa, comparar alguns parâmetros entre a expressão de emoções por actores e emoções espontâneas. Dez actores profissionais (5 homens e 5 mulheres) gravaram, representando 10 enunciados com diferentes emoções: raiva; alegria; tristeza; medo; desgosto e aborrecimento, para além do discurso modal. As frases eram, à partida, frases sem peso semântico, neutras. Realizaram várias conjugações de emoções (em termos de avaliação do grau da emoção; emocional vs. não emocional). As emoções foram simuladas (pelos actores) e foram mais facilmente reconhecidas quanto maior era a extensão da frase. Em termos de emoções mais espontâneas,

utilizaram como elementos extractos do *corpus*, já existente, do Feiticeiro de Oz, sobre o qual existem já resultados muito positivos. Nesta tarefa, os actores acabam por se comportar de forma muito natural, sem seguir um guião, interagindo com um sistema de diálogo, sem saberem que os seus estados emocionais estão a ser observados e analisados. Embora muitas emoções possam, de facto, ser quase realistas, uma grande parte do *corpus* que se acaba por recolher diz respeito a registos de fala neutros. Outro facto que torna esta análise diferente é que neste caso não temos só a voz como veículo de informação emocional mas também a expressão facial.

4.1.1 Estudos perceptivos

“Studies using electromechanical methods of analysing speech surely would be an important step toward defining the vocal cues of feeling. But in addition to such studies, it would be also imperative to investigate the auditory cues which can be discriminated by listeners, rather than by electronic devices, for in the final analysis, the cues heard by listeners must carry the emotional meanings involved in interpersonal, vocal communication.

Davitz, 1964, p.26, citado por Bänziger e Scherer (2003).

Os testes perceptuais são uma forma de investigação que permite qualificar um fenómeno, neste caso linguístico, uma vez que procuram melhor fundamentar as experiências e entender as visões aprofundadas que um pequeno número de indivíduos partilha acerca desse fenómeno (Susca e Healey, 2002: 137).

Várias pesquisas científicas têm vindo a ser feitas tentando fazer estudos comparativos de como os ouvintes de diferentes nacionalidades percebem emoções numa língua que não a sua. De acordo com Zinken et al. (2007), as línguas e culturas até agora estudadas não apresentam resultados muito díspares. Ainda assim, tem de se referir que foram poucas e muito específicas as emoções que até agora foram alvo de interesse para este tipo de estudo.

Começando em 1872 com Darwin, sabe-se que a expressão facial é mais universal do que a expressão verbal, embora diversos estudos possam convergir

em certos aspectos. Assim, tanto estudos de prosódia como de aspectos não-verbais revelam que a raiva é razoavelmente percebida, o mesmo não acontecendo com a expressão de alegria. Normalmente os ouvintes distinguem e percebem melhor as emoções veiculadas nas suas próprias línguas.

Muitas teorias defendem que as emoções são aprendidas e são afectadas pelo ambiente social em que o sujeito está inserido. Assim, pode-se conjecturar que as emoções variam, consideravelmente, de cultura para cultura (Matsumoto et al., 2002). Este autor defende que a universalidade em conjunto com as regras específicas de cada cultura faz com que alguns sentimentos sejam expressos em contextos sociais da mesma forma. Isto deve-se ao facto das pessoas se adaptarem, nas suas expressões faciais, às expectativas das outras culturas, em termos de intensidade das expressões (Matsumoto e Ekman, 1998).

Nos anos 80 a ciência estudava a relação existente entre os parâmetros acústicos em termos de fala e as emoções e expressões do indivíduo (Johnstone e Scherer, 1999). A investigação realizada até ao momento tem mostrado que as emoções não são reconhecidas de forma igual por todas as línguas e culturas (Zovato et al., 2004). Este mesmo autor, utilizando testes perceptuais, percebeu que existem problemas de identificação, ou dificuldades na diferenciação entre registos neutros vs. tristeza e alegria vs. raiva.

Num estudo publicado por Sawamura (2007), reconheceu-se que existem muitas similaridades entre desgosto e raiva; surpresa e alegria e que medo é muitas vezes confundido com tristeza, aquando da aplicação de um teste de percepção. Levou a cabo uma investigação utilizando informantes japoneses, americanos e chineses, demonstrando que existem factores comuns, independentes da língua e cultura, que determinam a forma como uma emoção é percebida. Raiva, alegria e tristeza parecem ser de facto, segundo o autor, as emoções básicas para as quais as outras convergem.

Com o objectivo de verificar se os falantes de Castelhana e Sueco conseguiam interpretar as emoções uns dos outros, Abelin (2004) desenvolveu um estudo para o qual utilizou as expressões de tristeza, cansaço, raiva, cepticismo, alegria, medo, depressão e euforia, em que os informantes teriam de

identificar as emoções expressas. Os resultados mostram que os ouvintes castelhanos são melhores nesta tarefa de interpretação. Algumas emoções, como alegria e medo, foram as mais difíceis de interpretar pelos dois grupos de informantes, quando usada apenas a informação da expressão vocal.

Os estados emocionais podem ser reconhecidos pelo sinal de voz, pela expressão facial e expressão corporal. Em vários estudos, as duas primeiras formas enunciadas são as mais utilizadas e são-no em concomitância tendo em vista a análise e reconhecimento de emoção. Estas duas formas de análise são bastante utilizadas em estudos de percepção de emoção interculturais (Shahid et al., 2008). Na pesquisa realizada por estes autores, utilizando 50 falantes caucasianos (holandeses) e 45 paquistaneses aos quais era pedido que agissem de acordo com determinada emoção, atestaram que as emoções “representadas” ou “interpretadas” eram, nas duas culturas, percebidas como mais fortes. Interessante referir que enquanto que para os informantes holandeses as emoções negativas eram mais facilmente percebidas e entendidas como mais fortes, os paquistaneses eram mais sensíveis e percepcionavam melhor isto aconteciam as emoções de índole positiva.

Reconhecer as emoções expressas em outras línguas poderá ser sempre mais fácil se tivermos acesso à postura corporal e expressão facial. No caso em que só falamos de qualidade de voz, o estudo torna-se mais meticuloso e os resultados podem tornar-se mais interessantes. A expressão facial é quase que assumidamente reconhecida de forma universal, sem que aqui intervenha a cultura, individualidade própria ou facilidade linguística (Pell et al., 2009), No seu estudo Pell et al. (2009) tentaram verificar se realmente apenas pela expressão, mesmo que não sendo na nossa língua materna, conseguimos percepcionar as emoções veiculadas. A um grupo de indivíduos que apenas falavam Castelhana da Argentina, foi-lhes pedido que tentassem reconhecer em três línguas distintas, Inglês, Alemão e Árabe, para além da sua língua nativa as emoções: alegria, tristeza, raiva, medo e aversão. As emoções eram expressas em frases desprovidas de qualquer valor semântico. Os resultados mostraram que a expressão oral destas emoções pode ser percepcionada com um grau satisfatório. No entanto, os indivíduos argentinos percepcionaram e identificaram

com um grau de certeza muito mais elevado as emoções expressas na sua língua materna. A alegria foi melhor identificada quando expressa em Castelhana (89%), seguindo-se o Árabe (59%), o Alemão com (57%) e por último com um grau de acerto de (32%) o Inglês. A raiva teve bons resultados tanto em Castelhana como em Alemão (81% e 77% respectivamente), quando comparada com o inglês e o Árabe (67% e 66%). A expressão de aversão que foi a menos reconhecida de todas as emoções, foi melhor percebida em Inglês (52%) e Árabe (45%) do que na língua materna (43%), já no Alemão as dificuldades foram notoriamente maiores (28%). A tristeza apresenta também resultados curiosos, uma vez que apresenta uma percentagem de acerto, para a língua Árabe de 77%, para o Inglês 74% e para o Alemão 65%, sendo que em Castelhana teve uma percentagem de acerto na ordem dos 51%. A expressão de medo não mostrou diferenças significativas de reconhecimento: Inglês, 61%; Castelhana, 57%; Árabe, (53%) e Alemão com 51% (Pell et al. 2009). Os autores fizeram ainda um estudo paralelo para ver se o sexo dos indivíduos influenciava esta interpretação, os resultados não mostraram que este possa ser um factor diferenciador. Os autores concordam que o reconhecimento de uma emoção expressa na língua nativa do falante é bem mais simples, uma vez que estão enraizados todos os aspectos culturais e comportamentais.

Referem ainda que independentemente da língua as emoções que apresentam uma maior tendência de reconhecimento são a raiva e a tristeza e a que é mais difícil de perceber/identificar é a expressão de aversão (Pell et al. 2009).

4.1.2 Estudos para o Português Europeu

Não existe muita investigação relacionada com o estudo da voz emocional para o Português Europeu, não existindo, por isso, qualquer tipo de *corpus* (extenso ou não) que pudéssemos consultar. Por isso, e para todos os estudos que levámos a cabo para esta dissertação, definimos e criámos os *corpora*.

O estudo realizado por Cabral (2006); Cabral e Oliveira (2006) que diz respeito à síntese da emoção em Português Europeu baseia-se em informações

para outras línguas complementadas com a extracção dos parâmetros da fonte glotal (como o *Open Quotient*) usando uma base de dados existente para o Alemão.

O *corpus* usado no nosso trabalho foi, parcialmente, utilizado por Rodrigues (2007), na sua tese de mestrado. Na pesquisa que realizou, analisou parâmetros relacionados com F0 e articulação, não contemplando parâmetros como *jitter* e *shimmer*. Rodrigues (2007)¹⁷ realizou o estudo das características acústicas (medidas de duração; F0 e articulação) da emoção no Português Europeu. A Tabela 11 é a síntese descritiva da análise acústica para as diferentes emoções no PE.

¹⁷ Na sua tese de mestrado, utiliza-se o mesmo corpus e informante, trabalhado também por nós. Foi um trabalho realizado em colaboração, uma vez que um dos orientadores é comum, onde se obtiveram resultados para outros aspectos realizados para emoção/actores.

Tabela 11 – Resumo dos resultados das análises de Rodrigues (2007). Retirado de Rodrigues (2007), p. 65.

	<i>Medo</i>	<i>Raiva</i>	<i>Desespero</i>	<i>Alegria</i>	<i>Tristeza</i>
Duração	A maior na duração total. A menor no nº de sílabas por segundo, com a raiva e a tristeza. Duração superior à média nas fricativas. As laterais duram mais. Duração elevada nas consoantes nasais. A maior duração nas vibrantes	Igual ao desespero e alegria na duração total. Maior nº de pausas, igual ao desespero. A menor nº de sílabas por segundo, com o medo e a tristeza. Duração inferior à média nas oclusivas. As laterais duram mais. Duração baixa nas consoantes nasais. A mais elevada nas vogais orais e nasais.	Igual à raiva e à tristeza na duração total. Maior nº de pausas, igual à raiva.	A menor na duração total. Menor no número de pausas. A maior nº de sílabas por segundo.	Igual à raiva e desespero na duração total. 2ª maior nº de pausas. A menor nº de sílabas por segundo, com medo e raiva. Muita variabilidade nas vibrantes.
F0	Muita variação na média dos valores máximos.	Valores mais elevados. A maior gama. Valores mais altos de desvio padrão.	Muita variação na média dos valores máximos.	Muita variação na média dos valores máximos.	
Articulação		Maiores valores de F1.			

4.2. Análise acústica

4.2.1 Método/Corpus

Criámos um *corpus* baseado no monólogo “A Voz Humana” de Jean Cocteau¹⁸ incluindo a frase simples /u m@LOr s@ra tumar@S ko~t6 del@S/, e a frase complexa /n6~w~ t6Ju ko~ s@rtez6 6 vOS djum6 psow6 kiSko~d@ kwalker kojz6/. Como informante, tivemos um actor do Teatro e Televisão Nacional, que gravou cinco emoções diferentes (utilizando as duas frases distintas – uma simples e uma complexa – semanticamente, sem qualquer carga emocional, simulando: alegria; tristeza; desespero; medo; raiva e o estado modal, em duas sessões diferentes. As frases foram gravadas de forma seguida para que não fosse dado ao actor o tempo necessário para se concentrar ou preparar para a emoção que iria exprimir. Com esta metodologia, pretendíamos conseguir um discurso que se aproximasse o mais possível do espontâneo. Foram primeiro gravadas as frases neutras e, seguidamente, de forma aleatória as outras quatro emoções.

4.2.1.1 Informante

O informante é actor profissional, do sexo masculino, com 42 anos de idade, tendo já um vasto percurso no teatro e televisão nacional. É ainda a ele que pertencem várias das vozes que podemos ouvir em desenhos animados ou como *voz off* de conhecidas publicidades. Quando gravou este corpus foi devidamente informado da natureza e objectivos do trabalho que se estava a realizar.

O actor não costuma realizar exercícios de relaxamento vocal nem treino de voz com regularidade. O que, na sua profissão, se pode vir a tornar um problema,

¹⁸ Este monólogo trata-se de uma representação teatral, uma peça num único acto que se centra numa mulher que fala ao telefone com o seu amante. Toda a acção se desenvolve a partir da conversa entre os dois (Publicada por Engenho e Arte).

podendo até vir a ser contraproducente, uma vez que, tal como para os professores (por exemplo), o seu instrumento de trabalho é precisamente a voz.

O actor nunca teve “rouquidão, falta de ar, dores nos ombros, voz tremida, laringites nem faringites”, salientando ainda que “raramente sente a garganta seca”. Embora estes sejam aspectos positivos, existem outros que demonstram exactamente o esforço vocal a que esta profissão obriga e que com o passar dos anos se podem tornar preocupantes: “*Algumas vezes por ano sente fadiga vocal, desconforto relacionado com a voz; perda de alcance da voz, dificuldades em projectar a voz*” e alguma tensão ao falar e mesmo dores no pescoço (Rodrigues, 2007).

4.2.1.2 Gravação

As gravações foram realizadas numa cabine insonorizada do Teatro Municipal da Guarda, todas no mesmo dia, com intervalo de uma hora entre cada sessão (das duas sessões existentes), havendo entre cada frase apenas as pausas respiratórias.

Utilizou-se, neste procedimento, um microfone AKG C 451 B e um gravador de DAT Tascam DA-P1, estando presente um técnico de som experiente. Na gravação das duas frases, utilizou-se sempre o mesmo procedimento.

Considerámos, desta forma, que a produção poderia aproximar-se mais da emoção espontânea, uma vez que não era dado ao actor o tempo necessário para se concentrar, tornado, por isso, o registo menos representado/interpretado.

4.2.1.3 Anotação e extracção de parâmetros

As frases foram anotadas ao nível de palavra e de fone, utilizando para o efeito a transcrição SAMPA¹⁹ em SFS (*Speech Filing System*)²⁰ anotando-se os

¹⁹ SAMPA (*Speech Assessment Methods Phonetic Alphabet*).

limites de cada enunciado (princípio e fim). Efectuou-se uma transcrição fonética larga, anotando os fenómenos de elisão (Figura 24), crase (Figura 25) e acrescentamento de determinados sons.

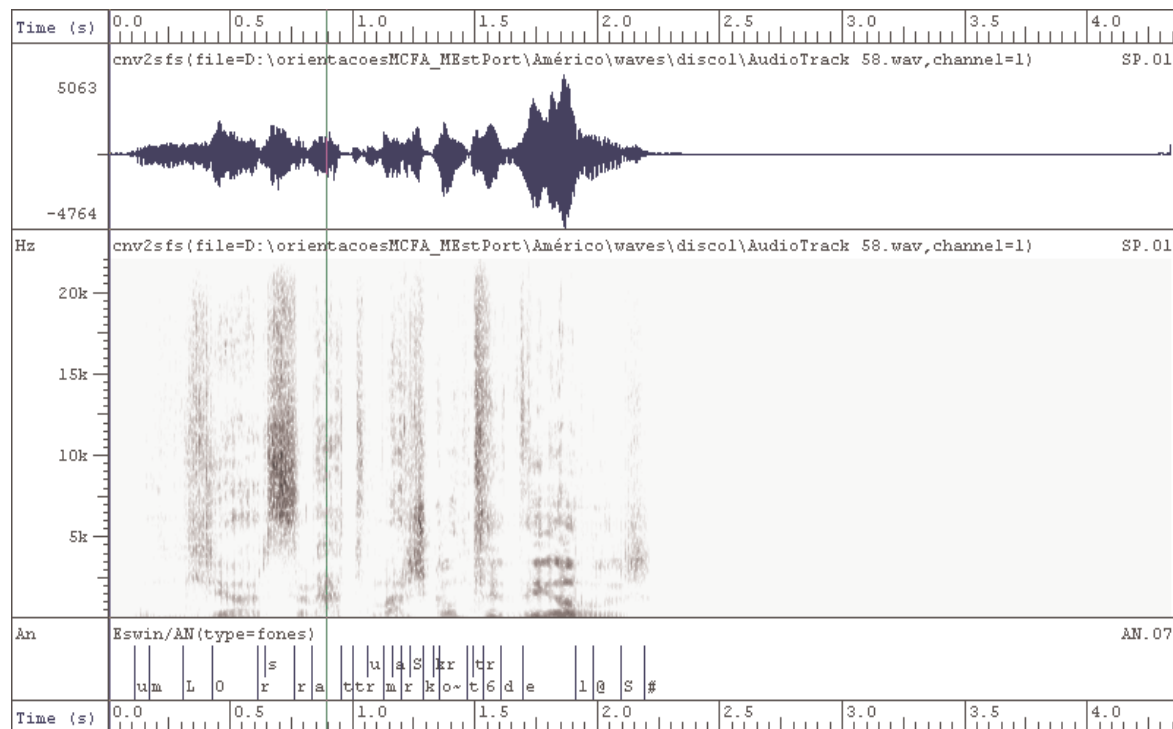


Figura 24 - Exemplo de elisão de [a] – /u mLOr sra tumarS ko~nt6d el@s/.

²⁰ Disponível em <http://www.phon.ucl.ac.uk/resource/sfs/>

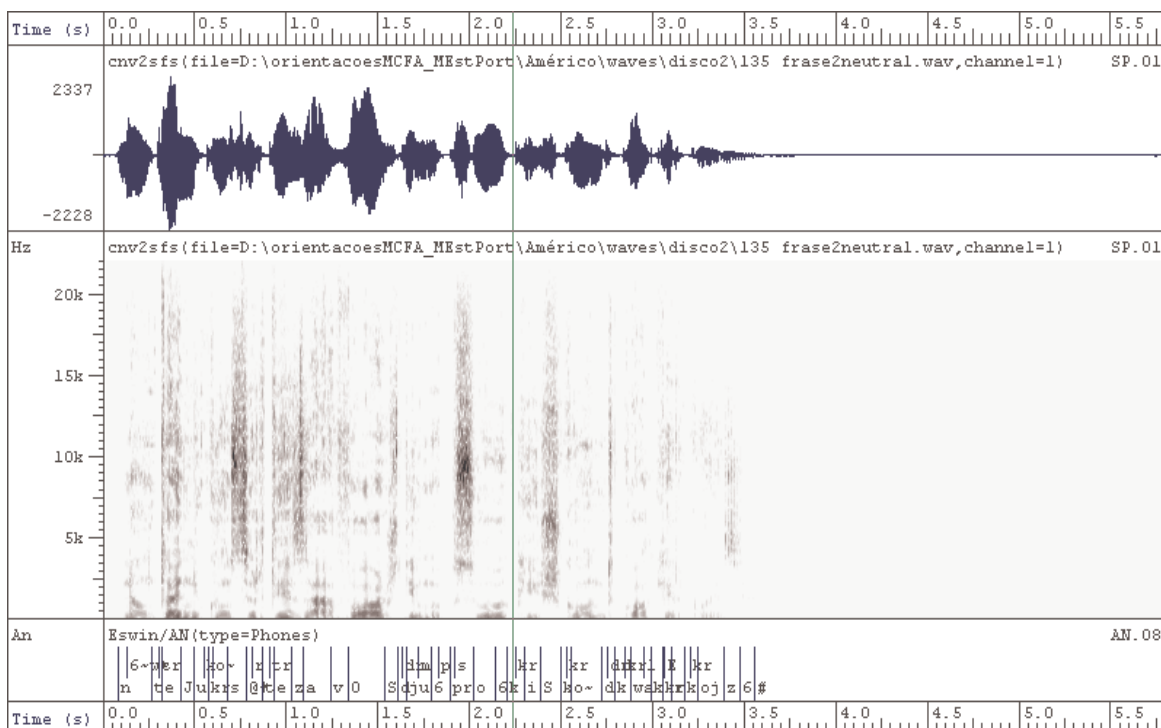


Figura 25 - Exemplo de crase [6]+[6] - [a] e acrescentamento [j] – /n6~w~ t6Ju ko~ s@rteza vOS djum6 psow6 kiSko~d@ kwalker kojz6/

Todos os dados foram passados para o sistema Praat de onde foi possível extrair através da função *Praat Voice Report* todos os elementos de que necessitávamos. Estes, foram seguidamente compilados para Excel e através do programa SPSS elaboraram-se gráficos elucidativos, mostrando mais precisa e directamente as diferenças e semelhanças encontradas na produção de cada enunciado consoante a emoção que se expressava.

Fizemos a extracção, tratamento, análise e comparação entre os diferentes valores de F0, *jitter*, *shimmer*, HNR para todas as emoções gravadas.

4.2.2 Resultados

Apresentamos, seguidamente, os resultados das análises relativas aos parâmetros tidos como mais relevantes na caracterização da qualidade de voz: F0, *jitter*, *shimmer*, *harmonic noise ratio* (HNR)

4.2.2.1 *Frequência fundamental - F0*

Analísamos quatro parâmetros diferentes relacionados com o F0.

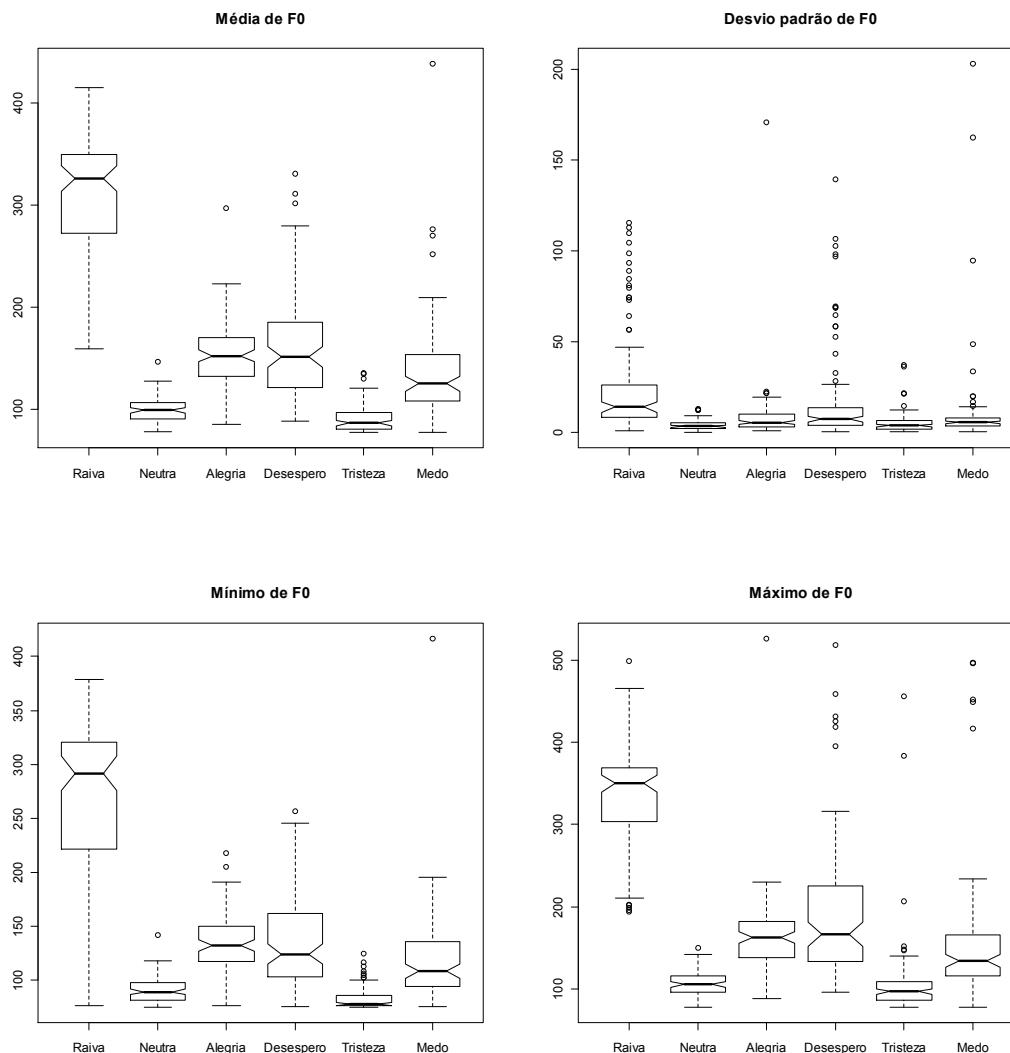


Figura 26 – Efeitos da emoção nos 4 parâmetros de F0 – de cima à esquerda, diagrama de extremos e quartis para: média, desvio padrão, mínimo e máximo.

A média de F0 deveria encontrar-se, para as mulheres, em padrões ditos normais entre os 150 e os 250 Hz e para os homens entre os 80 Hz e os 150 Hz (Behlau, 2001, citada por Susana Dias, 2008).

Os resultados para F0 mínimo, máximo, médio e desvio padrão são apresentados na Figura 26, de acordo com cada emoção em estudo. Podemos verificar os valores de F0 para cada emoção.

As análises destes diferentes parâmetros mostram que a raiva é a emoção que claramente se diferencia de todas as outras, apresentando uma média de valores perto dos 300 Hz, bem como um maior desvio em relação ao standard.

A expressão de alegria e desespero apresentam valores bastante similares para os 4 parâmetros de F0 analisados, encontrando-se para F0 máximo e mediano perto dos 150 Hz. A diferença reside na maior dispersão que se encontra para os valores de desespero.

Medo apresenta valores de F0 ligeiramente mais baixos que alegria e desespero, o mesmo acontecendo em relação ao desvio standard.

A tristeza, em termos de F0 apresenta valores sempre mais baixos que todas as outras emoções analisadas, aproximando-se dos resultados obtidos para a expressão neutra

Aplicando o teste Kruskal-Wallis (KW), ($p=0.01$) confirma-se que o factor emoção é significativo em todos os 4 parâmetros: [$\chi^2(5)= 338.65$, $p<0.001$] para o mínimo; [$\chi^2(5)= 370.72$, $p<0.001$] para o máximo, [$\chi^2(5)= 408.46$, $p<0.001$] para a média; e [$\chi^2(5)= 140.11$, $p<0.001$] para o desvio padrão.

Para F0 máximo, mínimo e mediana os testes *post-hoc* mostram como significativa a diferença entre todos os pares de emoções, à excepção de desespero-medo; desespero-alegria; medo-alegria e neutro-tristeza.

Para o desvio padrão de F0 as diferenças são significantes para os pares medo-neutro; medo-tristeza e alegria-tristeza. Alguns pares são difíceis de diferenciar se nos basearmos apenas nos parâmetros relacionados com a frequência fundamental. O desvio padrão é o que apresenta um menor poder diferenciador, sendo que os outros 3 apresentam um poder similar

É precisamente neste âmbito que se torna também interessante estudar a correlação entre todos os parâmetros de F0.

Na Figura 27 apresentamos os resultados da correlação entre os valores máximos e mínimos de F0. Somos então confrontados com duas situações: existe uma maior correlação para medo, neutra e alegria, indicador de uma maior relação entre os dois parâmetros. No grupo em que os valores se apresentam mais baixos (raiva, tristeza e desespero) os dois parâmetros F0, máximo e mínimo, podem considerar-se mais independentes.

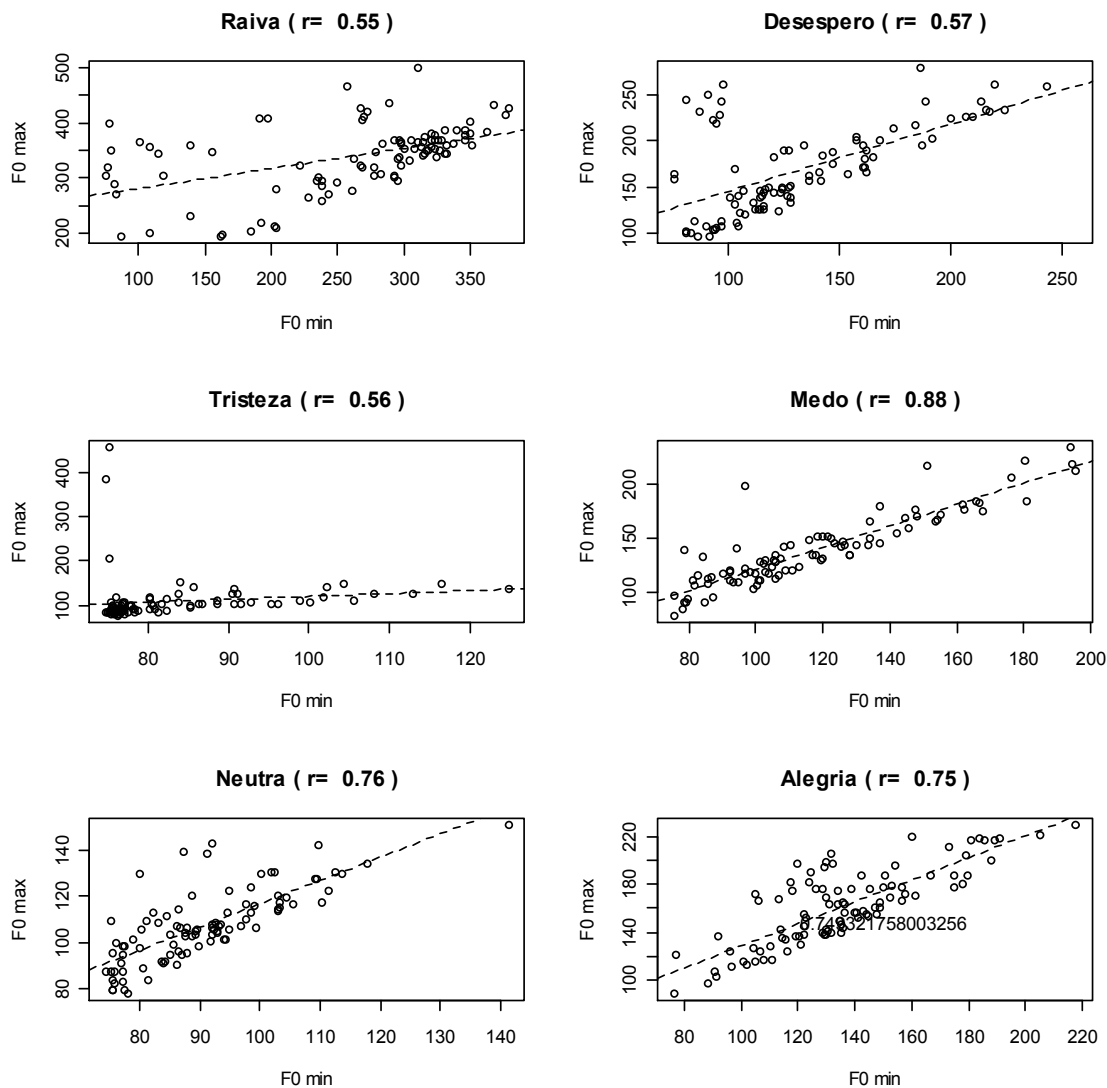


Figura 27 – Correlações entre os valores máximo e mínimo de F0 para as 5 emoções mais a expressão neutra.

Uma vez que utilizamos discurso corrente e não vogais sustidas, mais comuns nestas análises, analisamos para *jitter* apenas o parâmetro PPQ5. Para

as mulheres, o valor considerado como limite da normalidade é 0,5%, de acordo com Behlau (2001), citada por Susana Dias (2008).

Verificamos, pela observação da Figura 28, como é afectado o valor de PPQ5 de acordo com cada uma das emoções em análise.

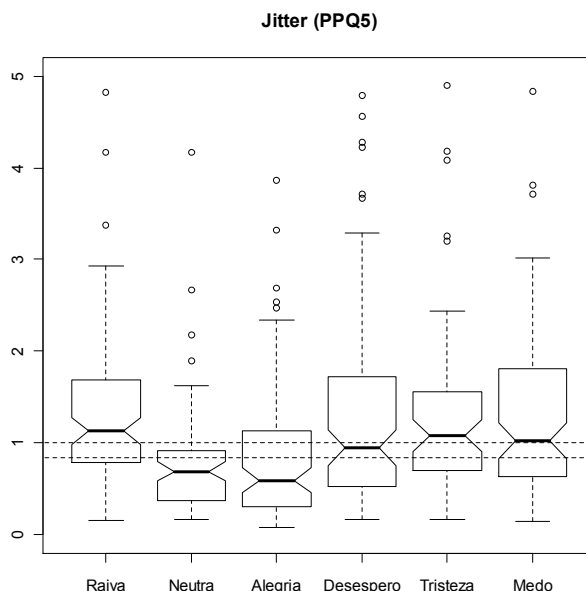


Figura 28 – Efeitos da emoção no parâmetro *jitter* PPQ5.

Comparando com os resultados obtidos para os parâmetros relacionados com F0, as diferenças em *jitter* não são tão evidentes. Os nossos resultados mostram que aos valores mais altos de *jitter* estão associadas emoções como desespero, medo, raiva e tristeza, que podemos denominar como emoções mais negativas. O discurso neutro e a expressão de alegria apresentam valores mais baixos e são similares entre si.

A alegria é a emoção que apresenta valores mais baixos para PPQ5 do que os valores tidos como referência para a normalidade em MDVP²¹. As outras

²¹ Trata-se de uma marca registada “*proprietary term or trade mark*”.

emoções não apresentam valores significativamente mais baixos ou altos do que os que são tidos como padrão.

Mais uma vez, aplicando o teste KW, pode-se confirmar como significativo o factor emoção [$\chi^2(5) = 53.9012$, $p < 0.001$]. Testes *post-hoc* mostram como diferentes os pares: alegria-raiva; alegria-medo; alegria-tristeza e neutra-tristeza. De acordo com os resultados obtidos na análise de *jitter*, a alegria destaca-se pelos seus valores mais baixos. O *jitter* revela-se como um parâmetro importante a ter em conta para a descrição da expressão de alegria.

4.2.2.2 *Shimmer*

Pela análise dos parâmetros de *shimmer*²², apresentados no gráfico da Figura 29 é evidente que são particularmente elevados para raiva, seguindo-se um grupo que agrega as emoções desespero, tristeza e medo.

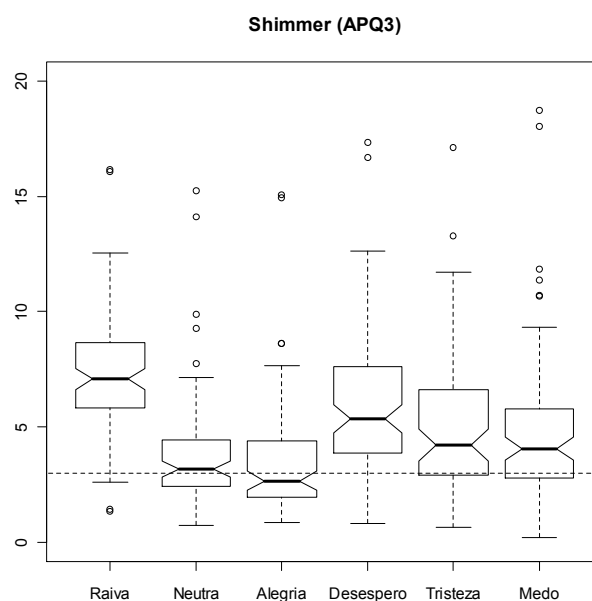


Figura 29 – Efeitos da emoção em *Shimmer* APQ3.

²² Shimmer APQ3 tem como limite da normalidade os 3%, Behlau, (2001).

Tendo em conta os valores de referência para a normalidade, apenas a expressão de alegria e a neutra apresentam valores significativamente acima dessa mesma referência. Os valores de *shimmer* para raiva e desespero encontram-se na região que é, usualmente, considerada como descritiva para patologias relacionadas com a voz.

Através do teste KW confirma-se como significativo o factor emoção [$\chi^2(5)=123.99$, $p < 0.001$]. Múltiplas comparações após a aplicação do teste KW mostram como significativas as diferenças entre: raiva-medo; raiva-alegria; raiva-neutra, desespero-alegria e desespero-neutra. A expressão de raiva só não apresenta valores de *shimmer* significativos mais altos em relação à tristeza e ao desespero.

Desespero, por sua vez, apresenta valores mais altos de *shimmer* que são significantes em relação à alegria e ao estado neutro. As outras emoções não apresentam diferenças significativas.

A análise de *shimmer* permite verificar que este apenas diferencia, em relação a todas as outras emoções aqui analisadas, a raiva e o desespero.

4.2.2.3 HNR

Na Figura 30 são apresentados os valores de HNR. A linha horizontal, apresentada no gráfico, representa a zona de referência²³. para a normalidade, os 12 dB.

Enquanto que a expressão de alegria apresenta valores mais altos do que 12 dB a maioria das emoções apresenta valores perto da média. Os valores de raiva e desespero destacam-se como significativamente mais baixos.

²³ Não existe, quanto aos valores de normalidade um consenso. Os 12 dB são geralmente tidos como um dos possíveis valores para a normalidade, em termos de HNR.

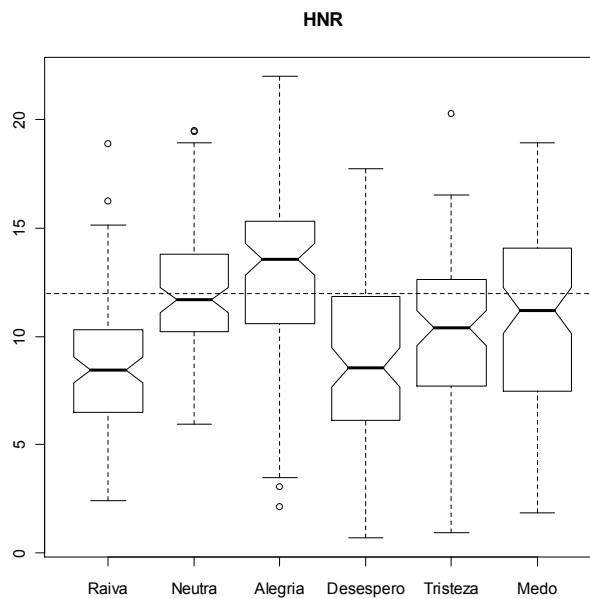


Figura 30 - Efeitos da emoção em Harmonic Noise Ratio (HNR).

Através dos testes ANOVA confirma-se, uma vez mais, como significativo o facto [$\chi^2(5) = 84.23$, $p < 0.001$]. Múltiplas comparações após o teste KW mostram como significativas as diferenças entre: raiva-medo; raiva-alegria; raiva-neutra; desespero-alegria; desespero-neutra e alegria-tristeza. Apresenta resultados bastante similares aos que são apresentados para *shimmer*, revelando novas diferenças significativas entre alegria e tristeza.

Na análise de HNR existe uma diferença significativa entre raiva e desespero (apresentando agora os valores mais baixos). Em termos da emoção positiva, constatamos que alegria apresenta os valores mais altos de HNR. Os valores para raiva e desespero estão localizados numa zona, potencialmente, classificável como patológica.

4.2.2.4 Combinação de todos os parâmetros

Tendo em conta análises multidimensionais dos parâmetros de voz (por exemplo: Kay Elemetrics MDVP e *Hoarseness Diagram*), na Figura 31 combina-se F0 (apenas a mediana, uma vez que não foram obtidas diferenças

significativas para os outros parâmetros de F0), *jitter*, *shimmer* e HNR, apenas num gráfico. Para resultados mais fiéis, os valores de cada parâmetro foram divididos pela sua mediana.

Todas as comparações são possíveis de verificar na Figura 31 e são feitas tendo como referência de comparação com o discurso neutro.

Tendo em consideração os resultados dos testes estatísticos apresentados anteriormente, percebemos que raiva é diferente da neutra nos quatro parâmetros em análise, em *jitter*, *shimmer* e F0 os valores sofrem um aumento enquanto que em HNR desce.

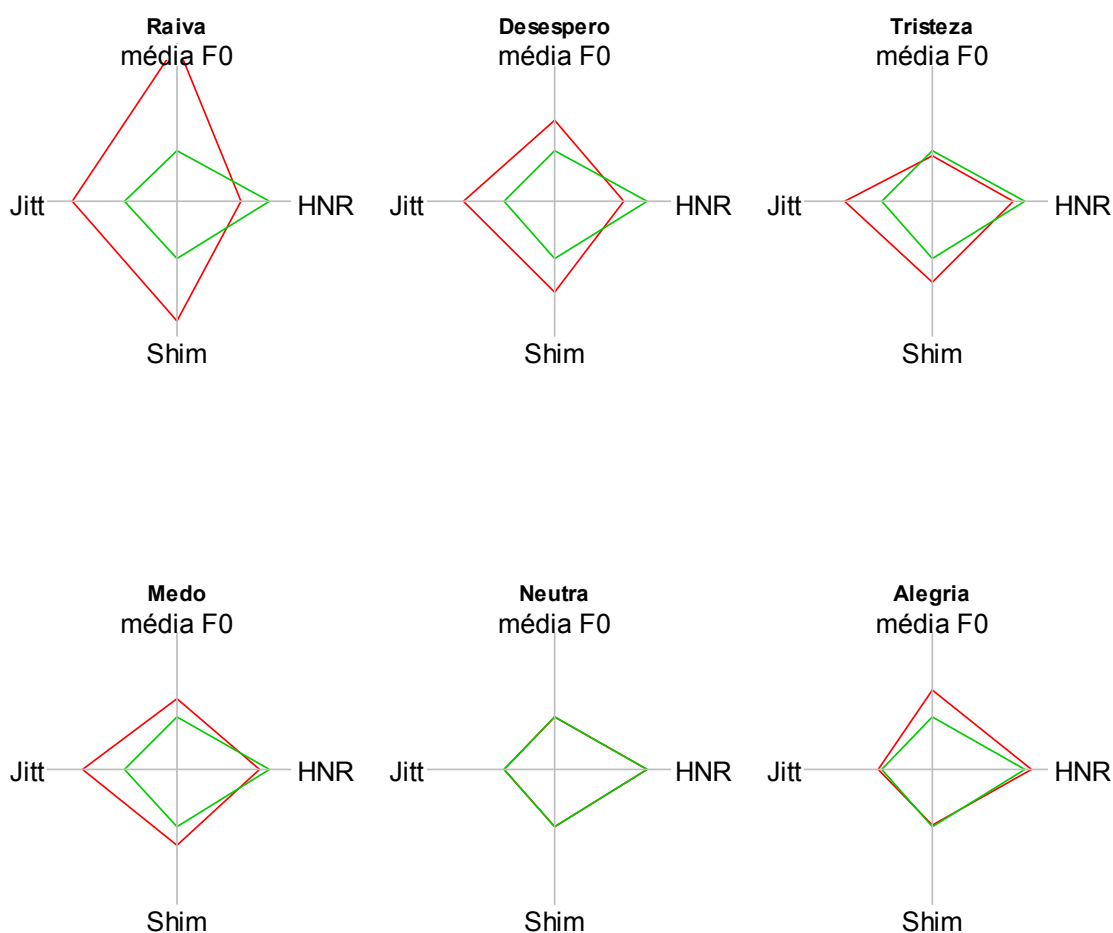


Figura 31 – Comparação das 5 emoções e expressão modal baseado nos quatro parâmetros analisados. Para F0 foi escolhido o valor da mediana como representativo dos outros quatro parâmetros que fazem parte do nosso estudo.

Desespero também difere nos quatro parâmetros, embora com diferenças menores do que as apresentadas para raiva. A expressão de tristeza apenas difere nas medidas de irregularidade (*jitter* e *shimmer*). Medo é similar ao desespero, com valores de *Harmonic Noise Ratio* idênticos ao do discurso neutro e apenas com pequenas diferenças.

A emoção mais positiva que analisamos, alegria, apresenta apenas valores diferentes em relação ao F0.

4.3. Teste perceptual

Todos os indivíduos variam nas suas capacidades de expressar as suas emoções verbalmente, e também parece haver diferenças no reconhecimento e interpretação das atitudes e emoções dos outros. Muitos parâmetros podem ser idênticos em diversas línguas, mas há outros que fazem parte da idiosincrasia de uma língua e cultura, ou mesmo de um indivíduo. Actualmente, vários estudos confirmam que emoções como alegria, tristeza e raiva são as emoções mais facilmente identificáveis em diferentes línguas (sobre estes aspectos de reconhecimento de emoções ver, por exemplo, os estudos de Sawamura (2007) e Pell et al., (2009)).

A percepção ou não de emoções por outras culturas e falantes de outras línguas tem vindo a ser um tema que cada vez mais vem despertando o interesse de áreas como a da psicossociologia, ciências da comunicação e diversos campos científicos (Kehrein, 2002).

4.3.1 Metodologia

Tendo em conta o nosso objectivo, ou seja, a investigação da influência que o conhecimento de uma língua e respectiva cultura pode ter na percepção de emoções, aplicámos um teste de percepção com o mesmo *corpus* descrito para a análise de emoção por actores (início do capítulo 4). As frases foram misturadas e baralhadas de forma a que não houvesse qualquer indício ou suposição, por parte

dos ouvintes de que havia algum tipo de sequência na amostragem das emoções a ouvir.

4.3.1.1 *Sujeitos*

Dois grupos diferentes, com uma média de idades próxima dos 22 anos, participaram do nosso estudo num total de 72 informantes (37 americanos, da Universidade de Lafayette, Louisiana) e 35 portugueses, da Escola de Saúde da Universidade de Aveiro).

O primeiro grupo (universitários americanos) limitou-se a ouvir as frases sem que estas tivessem, para eles, qualquer valor semântico (só no fim, por curiosidade, as frases foram traduzidas). Estes informantes nunca tinham tido qualquer contacto prévio com uma língua românica. O segundo grupo era, tal como o outro, constituído por alunos do quarto ano de linguística e eram todos eles falantes nativos de Português Europeu.

4.3.1.2 *Procedimento*

A mesma bateria de testes foi aplicada a cada um dos grupos, utilizando o mesmo teste e metodologia. Ouviram, em contexto de sala de aula, duas frases distintas, uma simples e uma complexa (num total de 42 enunciados). Todos os informantes ouviam as frases ao mesmo tempo usando o *Windows Media Player* e ligando a saída de áudio do computador às colunas da sala. Os enunciados ouvidos foram produzidos por um actor profissional português, simulando: tristeza, desespero, raiva, alegria, medo e o discurso neutro ou modal.

4.3.2 Resultados

Comparámos os resultados de identificação tendo em conta factores como: língua e emoção e efeitos linguísticos como a duração da frase, para saber de que forma é que esta interfere na identificação da emoção veiculada.

4.3.2.1 Resultados por comparação ouvinte/emoção

Começando pela média dos acertos, representada pelo intervalo de confiança a 95% para a média, na Figura 32, em função das emoções e língua nativa dos ouvintes.

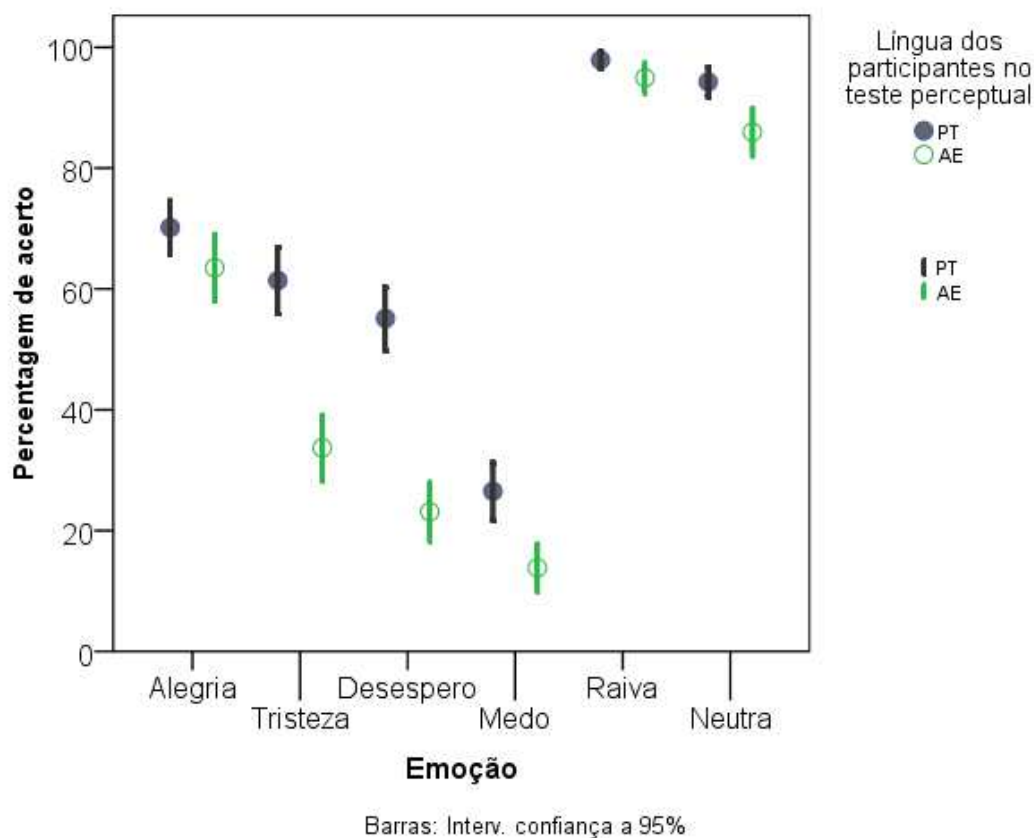


Figura 32 – Comparação entre as respostas dos dois grupos de informantes. Na legenda PT é usado para Português Europeu e AE para Inglês Americano.

Pela observação da Figura 32, é notório que a raiva é a emoção mais facilmente identificada por ambos os grupos seguida pelo estado neutro.

Nas emoções tristeza e desespero percebe-se uma maior discrepância na identificação comparando os dois grupos. Contudo, os testes permitem verificar que estas emoções são mais fáceis de perceberem pelos falantes da própria língua do que pelos informantes americanos.

O medo é a expressão que maiores dificuldades de identificação suscitaram, em ambos os grupos.

O gráfico da Figura 33 permite melhor visualizar o nível de acertos, (tendo em conta todos os parâmetros: emoções; frases e ouvintes). Verifica-se a existência de uma diferença significativa no desempenho dos informantes portugueses. Estes conseguem identificar, com um grau mais elevado de acertos, um maior número de emoções do que os falantes americanos

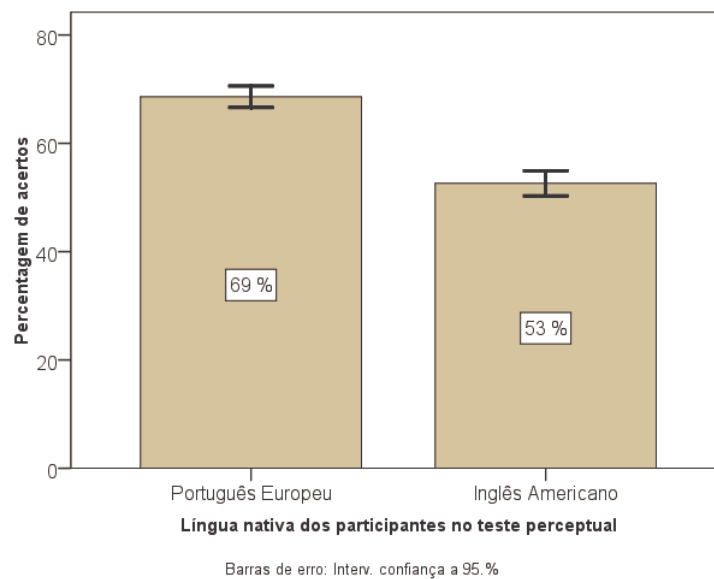


Figura 33 – Resultados gerais de acertos (informantes/emoções).

A possível influência da frase – mais simples ou mais complexa – foi também investigada. Os resultados encontram-se na Figura 34.

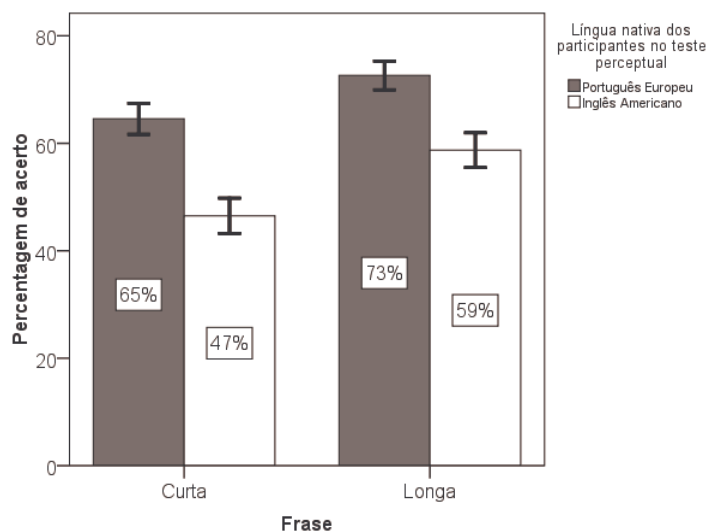


Figura 34 – Percentagens de acerto em função do tipo de frase.

Pela observação do gráfico (Figura 34) percebe-se que quanta mais informação houver em termos de extensão do enunciado, mais fácil se torna a sua interpretação e a concomitante percepção da emoção veiculada. Este factor afecta, de igual forma, os resultados dos dois grupos de participantes.

Se ao factor anterior (extensão do enunciado) adicionarmos informação sobre emoção e língua nativa dos ouvintes obtemos os resultados detalhados possíveis de observar na Figura 35.

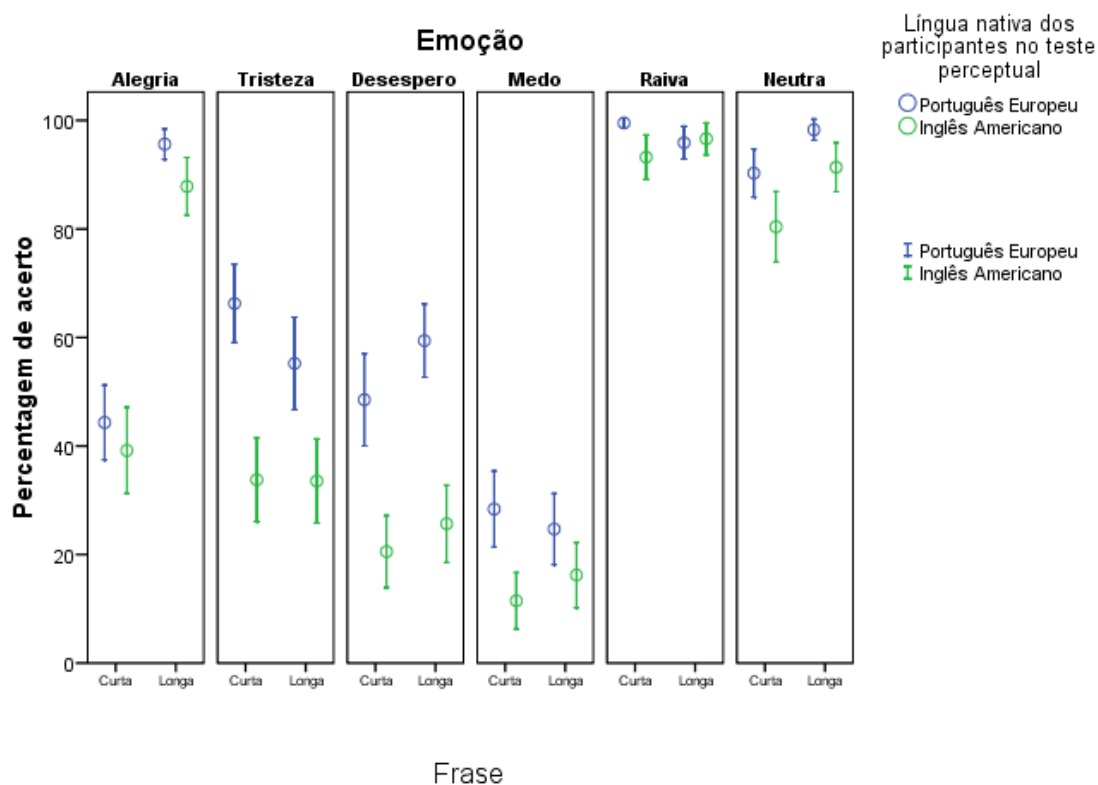


Figura 35 – Resultados para cada emoção e falante de acordo com o tipo de frase.

Os dados presentes na Figura 35 apresentam um intervalo de confiança a 95%, o que nos permite perceber que há algumas oscilações quando comparamos os três factores em análise, principalmente se falar na percepção da alegria. Nesta, a diferença na duração da frase e estrutura melódica conferida por mais informação trouxe diferentes resultados em ambos os grupos; a frase longa tornou muito mais fácil a identificação da emoção veiculada.

Verificando os resultados para tristeza e desespero, percebemos que a duração é uma ajuda mais significativa para o PE. Na tristeza, a duração da frase não parece ter qualquer influência para os AE. Quanto às restantes emoções, nos dois grupos, a duração do enunciado não demonstrou ter grande interferência na percepção das emoções.

A análise estatística, na Tabela 12, confirma como significativos os efeitos (que destacamos o valor de *p* a negrito) da língua, frase e emoção e as interacções mais particulares língua/emoção e frase/emoção.

Tabela 12 – Análise estatística dos resultados do teste de percepção. Nota: foram eliminadas algumas linhas e colunas produzidas pelo SPSS cuja informação não se considerou essencial.

<i>Factor</i>	<i>df</i>	<i>F</i>	<i>p</i>
Língua nativa dos Sujeitos	1	143,699	,000
Frase	1	71,432	,000
Emoção	5	397,423	,000
Língua * Frase	1	1,997	,158
Língua * Emoção	5	14,618	,000
Frase * Emoção	5	47,376	,000
Língua*Frase*Emoção	5	1,166	,323
Erro	3869		

4.4. Discussão

São até hoje poucas as línguas e culturas que apresentam este tipo de estudo de emoção, tendo em conta única e exclusivamente a análise da voz, sem recorrer à expressão facial.

Desde logo é notório que os falantes nativos do PE conseguiram melhores resultados na percepção de todas as emoções, especialmente na tristeza. Podemos afirmar que todas as emoções, à excepção do medo mostraram resultados estatisticamente diferentes entre os dois grupos.

A alegria é, no nosso estudo, muitas vezes confundida com o discurso neutro (como vimos no estudo perceptivo: tanto pelos falantes nativos do PE como pelos americanos). Tendo em consideração as limitações do nosso corpus, não podemos, generalizar. No entanto, podem-se questionar diferenças na expressão da alegria ou uma maior dificuldade, por parte do actor, em conseguir aproximar-se desta emoção

A raiva apresenta o valor mais alto de F0, seguindo-se-lhe a alegria e com o valor mais baixo encontramos para o PE a tristeza, estando estes nossos valores de acordo com os apresentados por Cowie et al. (2003). Ainda segundo os autores e, havendo resultados similares no nosso estudo, a tristeza aproxima-se no F0 da expressão neutra. As nossas medidas não corroboram o aumento de F0 na expressão de medo. Scherer (citado por Airas e Alku, 2004) defendia que as medidas de F0 estão relacionadas com a dimensão mais ou menos activa da emoção. Assim, uma emoção mais activa apresenta valores de F0 mais altos.

A comparação entre *jitter* e *shimmer* com a literatura torna-se algo mais difícil, uma vez que existem poucos estudos que incluam estes parâmetros e também não há uma convergência quanto ao parâmetro concreto a analisar (local ou PPQ5). Seguimos, na nossa pesquisa, uma comparação com Drioli et al. (2003), uma vez que estes usam também o Praat nas suas análises.

Para *jitter* e *shimmer* os nossos resultados são idênticos aos apresentados em estudos de outras línguas latinas como o Espanhol 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 apresentados 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, por nós observadas, 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, afastando esta emoção dos valores de medo.

Para algumas das emoções parâmetros como HNR e *jitter* apresentam valores na zona da patologia. Mostra este resultado que, de facto, a nossa voz é sensível às nossas emoções, havendo uma interligação. Deve por isso, o terapeuta da fala ou qualquer indivíduo que trabalhe com voz ter em atenção o estado emocional do paciente, uma vez que, como podemos demonstrar, este muda a qualidade da nossa voz. Tanto a tristeza como a alegria provocam efeitos na voz que a levam a sair dos parâmetros da normalidade.

Salientamos os resultados obtidos, em termos de comparação ouvinte/emoção que existe uma discrepância maior na identificação das emoções tristeza e desespero. Ainda assim, é mais facilmente percebida esta diferença pelos falantes da própria língua do que pelos informantes americanos. Em termos gerais, medo é a emoção que, para ambos os grupos, maiores dificuldades de identificação trouxe

Quanto à melhor percepção da tristeza, por parte dos falantes nativos do Português Europeu, podemos conjecturar sobre a importância do conhecimento da língua e cultura. Este é de facto um resultado particularmente interessante, uma vez que pode ter possíveis ligações intrínsecas ao conhecimento e integração numa cultura própria. Acresce a esta ideia o facto de o povo português ser, genericamente, caracterizado como nostálgico e saudosista. Uma cultura que se espelha e expressa muito pelo “Fado” e também na “saudade”.

Outra conclusão a retirar é o facto de que em ambos os grupos a emoção que mais se destacou em termos de acertos foi a raiva, o que vem de encontro à bibliografia por nós consultada e já anteriormente referida em que se tomava esta como a emoção universalmente mais reconhecida (Pell et al. 2009).

Percebe-se ainda que a identificação de cada uma das emoções aumenta, nos dois grupos, consoante a duração da frase, sendo que os resultados são melhores quando é analisado o enunciado mais longo.

Em suma, os resultados obtidos para o PE estão, no geral, de acordo com os estudos consultados. Não deixa pois de haver alguma dependência entre a universalidade de uma emoção e a língua em que é expressa, estes factores inter-relacionam-se e influenciam-se.

4.5. Comentários Finais

De acordo com as análises que realizámos, tendo em consideração que o *corpus* analisado não é muito extenso e, sabendo que o nosso informante é um actor profissional, por isso habituado a trabalhar com a voz e emoção, não encontramos características muito claras ou evidentes que possam distinguir o Português Europeu de outras línguas.

Em termos de aspectos interculturais na interpretação de emoções é curioso referir um estudo que advoga que qualidade de vida e satisfação dos indivíduos em termos familiares sociais e culturais (o que também tem relação com a política e economia do país), influencia a capacidade de reconhecer melhor as emoções negativas ou positivas (Inglehart e Klingemann, 2000). Se atentarmos nos resultados gerais obtidos, observamos que a percepção da alegria é muito próxima nos nossos dois grupos de participantes. No entanto, a tristeza e o desespero mostram diferenças significativas; os alunos portugueses conseguem identificar estas emoções muito melhor do que os americanos.

No capítulo seguinte, apresentamos um estudo, dentro dos mesmos parâmetros que temos vindo até agora a analisar, da emoção espontânea. Esse estudo permitirá uma comparação entre a emoção produzida por um actor e a emoção espontânea, procurando aferir em que parâmetros e valores se aproximam e em quais se afastam. Poderá trazer ainda alguns resultados concretos e específicos para o Português Europeu.

Capítulo 5 - Estudo de voz emocional espontânea

Tendo em conta o estudo realizado no capítulo quatro, torna-se importante verificar se o actor quando simula emoções se aproxima da emoção espontânea nos variados parâmetros relacionados com qualidade de voz, ou se distancia em que parâmetros e em que medida o faz.

O estudo levado a cabo neste capítulo pretende analisar a emoção espontânea em PE, comparar os valores obtidos com a emoção noutras línguas e com a simulada. A expressão espontânea de emoções gravadas ao 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 em contexto) (Patrick, 2008). Procuramos a obtenção de dados que permitam caracterizar e descrever a expressão de emoção espontânea no Português Europeu. Continuamos a manter o nosso foco de estudo nos parâmetros que até agora temos vindo a analisar (F0; *jitter*; *shimmer* e HNR), pois só assim nos será permitida a realização de uma comparação entre emoção simulada e emoção espontânea.

5.1. Trabalhos relacionados

A maior parte dos estudos existentes versa, sobretudo, sobre emoções simuladas, induzidas em contexto de laboratório e/ou produzidas por actores profissionais. Assim, 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, 2008).

O progresso da tecnologia leva à necessidade de cada vez mais estudos, nomeadamente, no campo da emoção espontânea para que haja uma interacção cada vez maior e melhor entre homem e máquina (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 consiste em 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 em 2007 por Batliner et al. 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.

5.2. 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 descontextualizado). 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 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, uma vez mais 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.

5.2.1.1 *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 feitas 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.

5.2.1.2 *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.

5.3. 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.

5.3.1.1 *Feeltrace*

O *Feeltrace* (exemplo na Figura 36) 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 sustidas), 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

No nosso estudo, seguindo Ekman (2003) e Cowie et al. (2003), que citámos no capítulo 2 deste trabalho, 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* 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, 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.

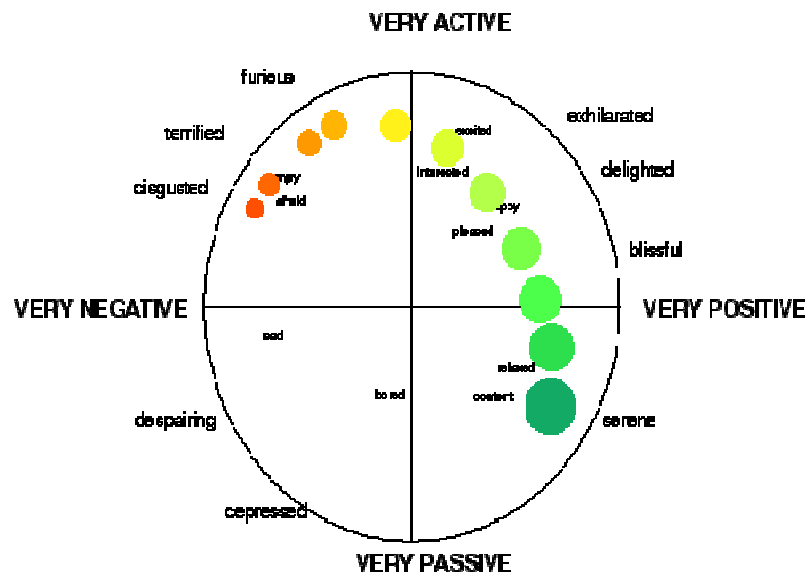


Figura 36 – Exemplo do *Feeltrace* durante uma sessão de gravação. O cursor muda de cor de vermelho para laranja (à esquerda) e de amarelo para verde, de acordo com o eixo em que se encontra.

5.3.1.2 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 estavam realizando (para que se pudessem evitar qualquer tipo de influências). Seguidamente liam a ficha de apoio e de acompanhamento do teste, onde tinham que 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.

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.

5.3.1.3 *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 áreas de investigação. Três dos ouvintes são professores Universitários, sendo os restantes investigadores a realizar doutoramento, todos eles da Universidade de Aveiro. Estão ligados a trabalhos relacionados com voz, processamento de voz, sinal, robótica e Linguística. Três dos ouvintes eram da área da Linguística, os outros mais relacionados com Engenharia (processamento de voz, sinal, robótica)

5.3.1.4 *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 37 podemos observar o grau de confiança dos ouvintes na marcação de emoções.

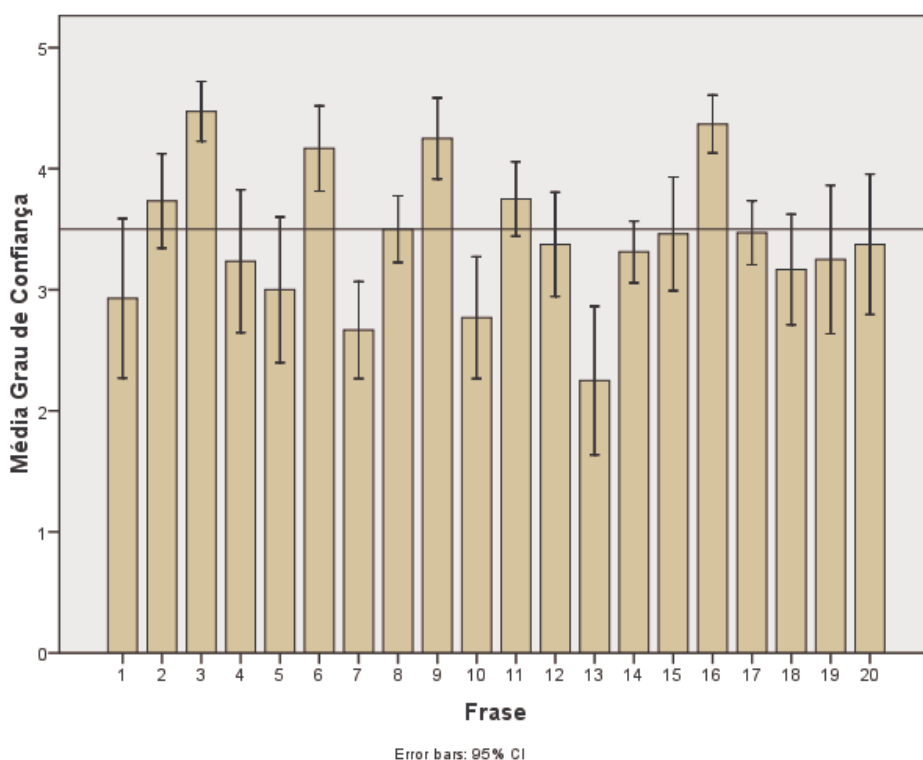


Figura 37 – Grau de confiança na marcação da emoção ouvida.

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 38 faz-se a análise da confiança por ouvinte.

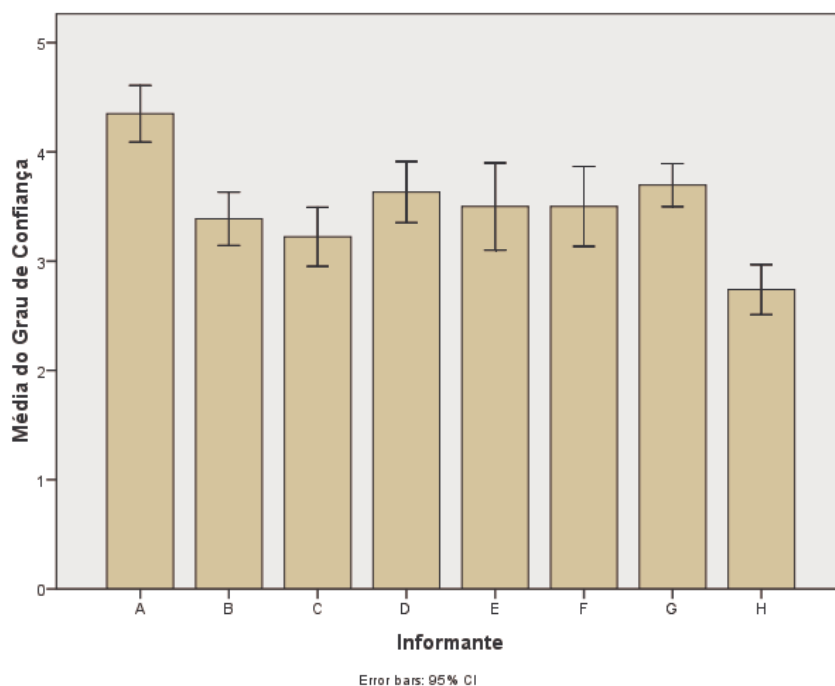


Figura 38 – Média do grau de confiança na marcação das emoções ouvidas, por indivíduo.

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 próximo do 3/3,5. O que mostra uma certa uniformidade na forma como eram ouvidos e entendidos os enunciados.

Na Figura 39 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 elas emoções 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.

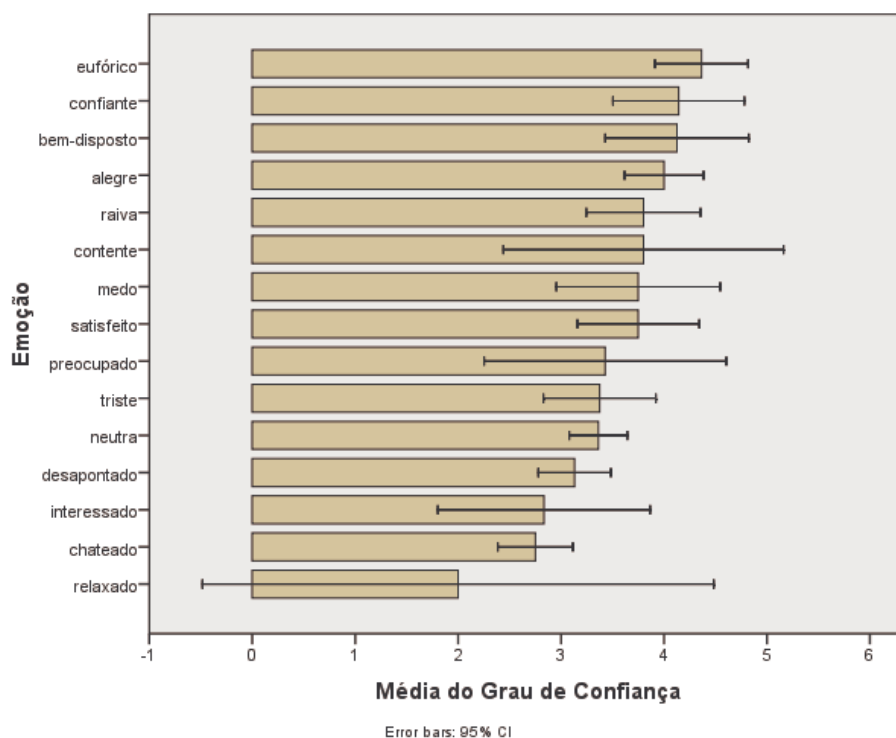


Figura 39 – Média do grau de confiança, por emoção marcada como primeira.

Na Figura 40 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.

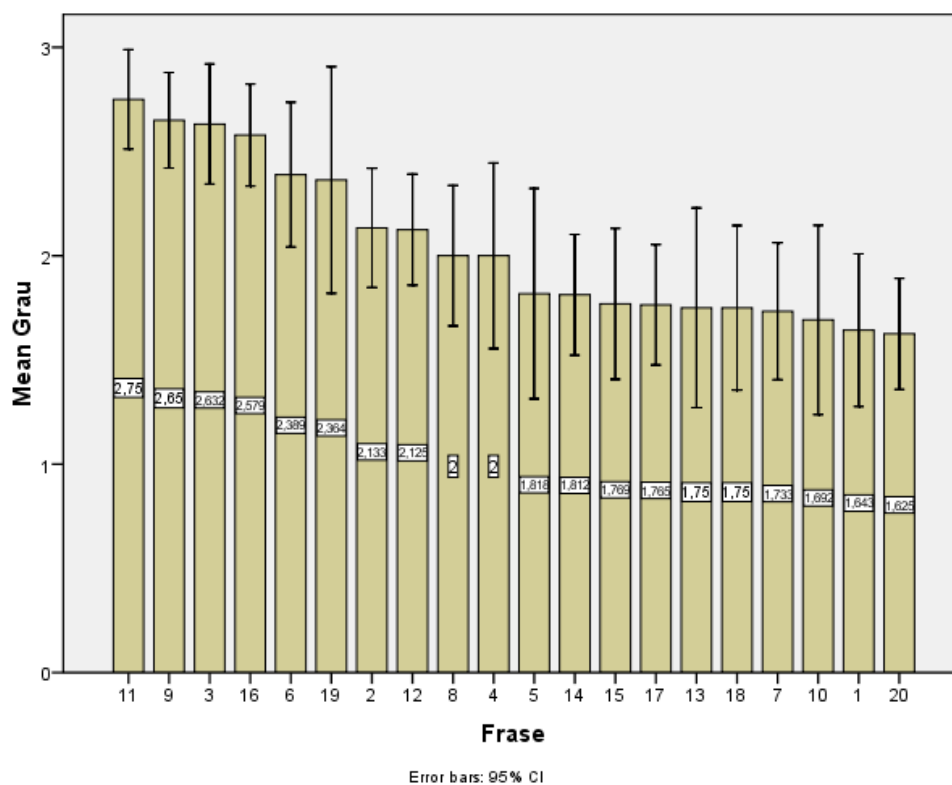


Figura 40 – Grau de intensidade atribuído à primeira escolha

5.3.1.5 Como etiquetaram

Como referimos anteriormente, os ouvintes podiam escolher três emoções diferentes, ordenando-as (da que sentiam ser mais forte para a mais fraca, de 1 a 3).

O gráfico da Figura 41, mostra-nos, precisamente, essa 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).

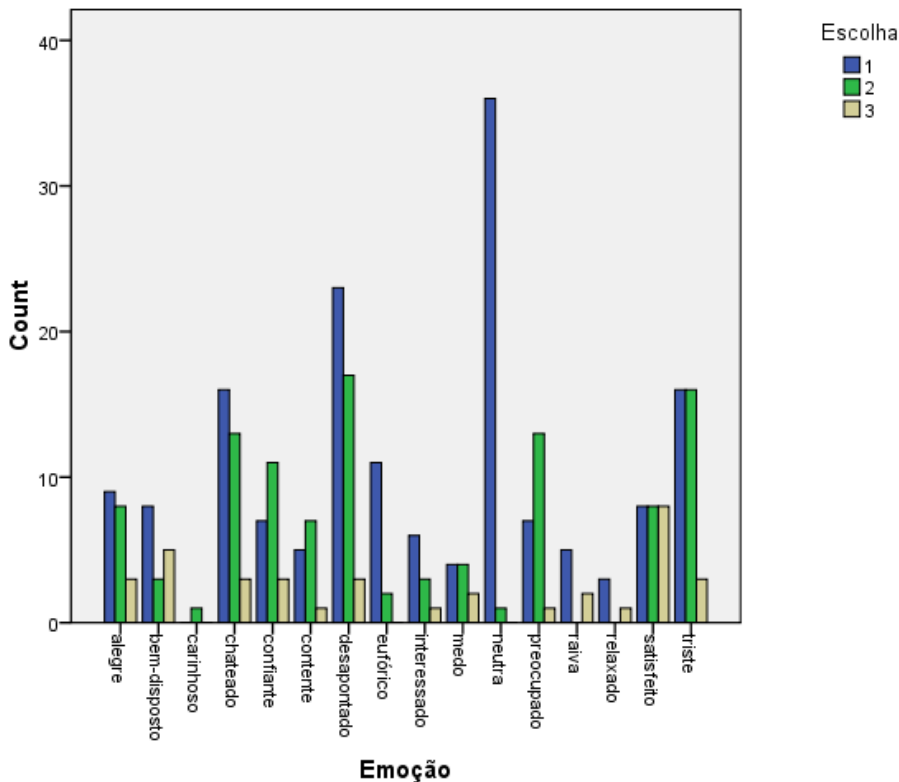


Figura 41 – Diversidade das escolhas para descrever a emoção ouvida.

Apresentamos, a título de exemplo, na Figura 42, 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 é que a emoção era de carácter negativo.

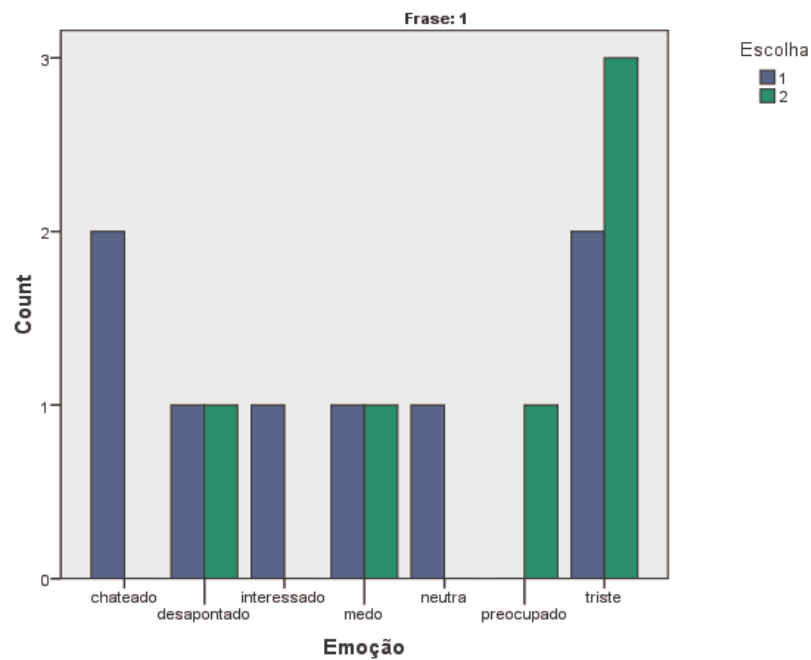


Figura 42 – Etiquetagem realizada para o primeiro enunciado. Família da emoção da primeira escolha: Negativa.

A Tabela 13 apresenta-nos um resumo de todos os resultados, em termos de frase; emoção mais escolhida; número de vezes que foi escolhida, outras escolhas; família de emoções; média de confiança com que marcaram e grau de intensidade da primeira escolha.

Tabela 13 – 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.

5.3.1.6 Resultados *Feeltrace*

A Figura 43 e seguinte dizem respeito ao que foi obtido pelo *Feeltrace*. Podemos observar, pelas diferentes cores (que correspondem aos diferentes ouvintes) as marcações efectuadas enquanto ouviam o enunciado.

Na Figura 43 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.

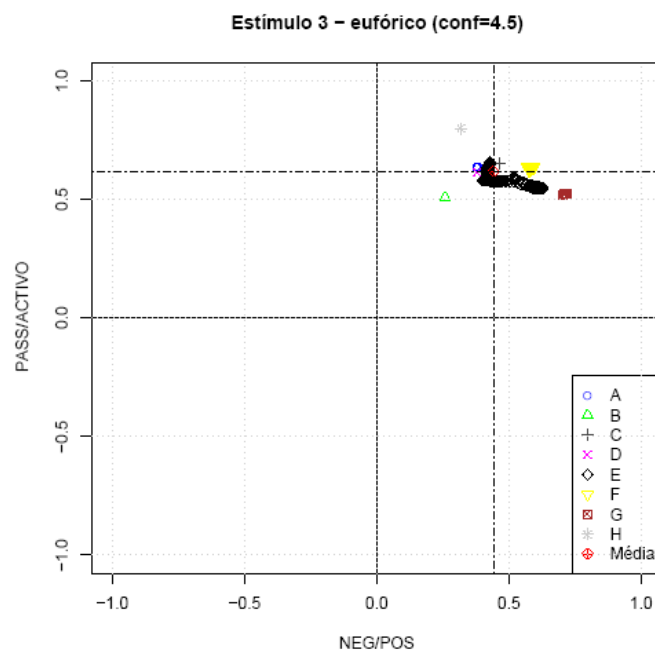


Figura 43 – Exemplo de resultados *Feeltrace*. Anotações dos vários ouvintes e média das mesmas.

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 44 demonstra exactamente essa variação.

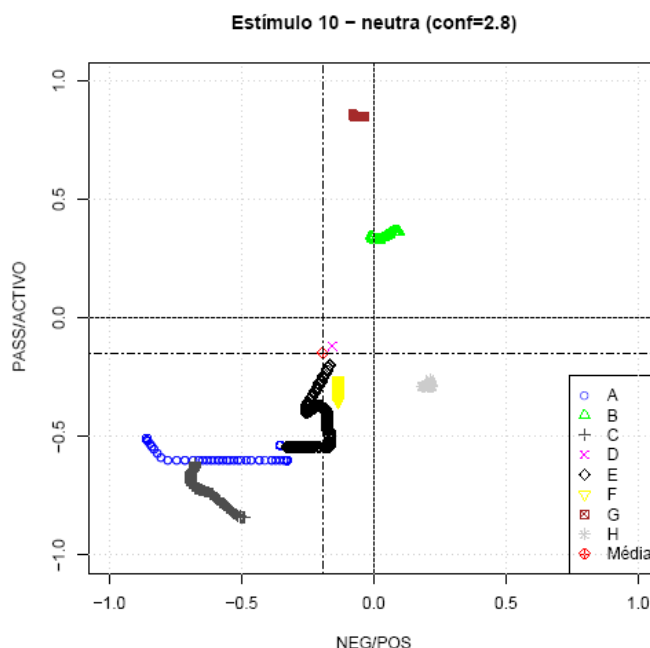


Figura 44 – Exemplo de resultados *Feeltrace*, em que existe muita variabilidade. Anotações dos vários ouvintes e média das mesmas.

Para um enunciado neutro houve de facto uma dificuldade de anotar quer em termos de ser uma emoção activa ou passiva, quer em termos de ser positiva ou negativa. 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.

Todos os gráficos respeitantes ao teste e correspondentes resultados do *Feeltrace* encontram-se no Anexo **A**

5.4. 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.

5.4.1 Usando as classificações dos ouvintes

5.4.1.1 Valores de F0

Na análise dos valores respeitantes a F0, na Figura 45, 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.

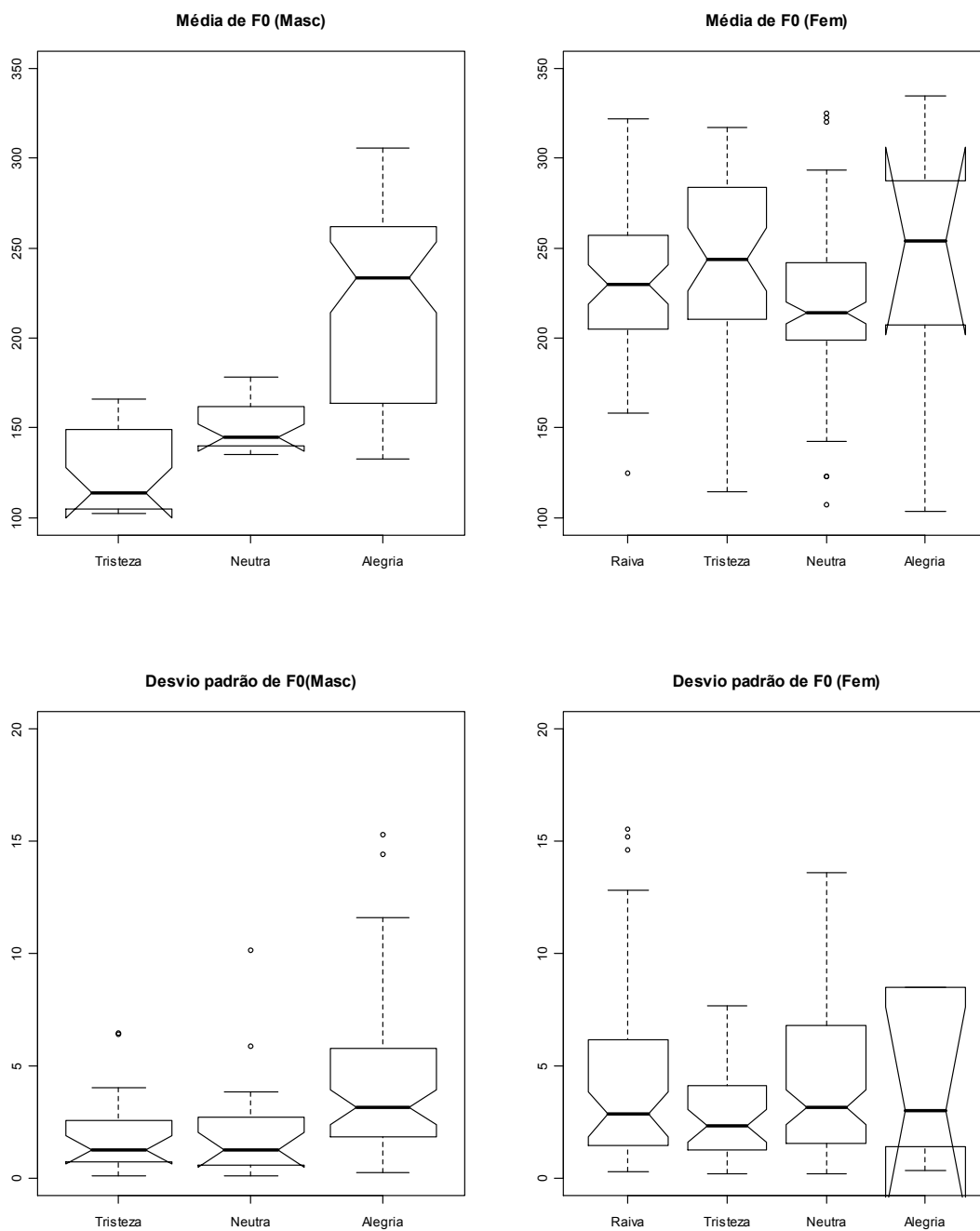


Figura 45 – Valores de média de F0 e desvio padrão para os informantes masculinos e femininos, da esquerda para a direita.

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. Não podemos dizer que haja, em termos de média de F0, para o sexo feminino, diferenças muito grandes.

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.

5.4.1.2 *Jitter*

Na análise de *jitter* (Figura 46 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.

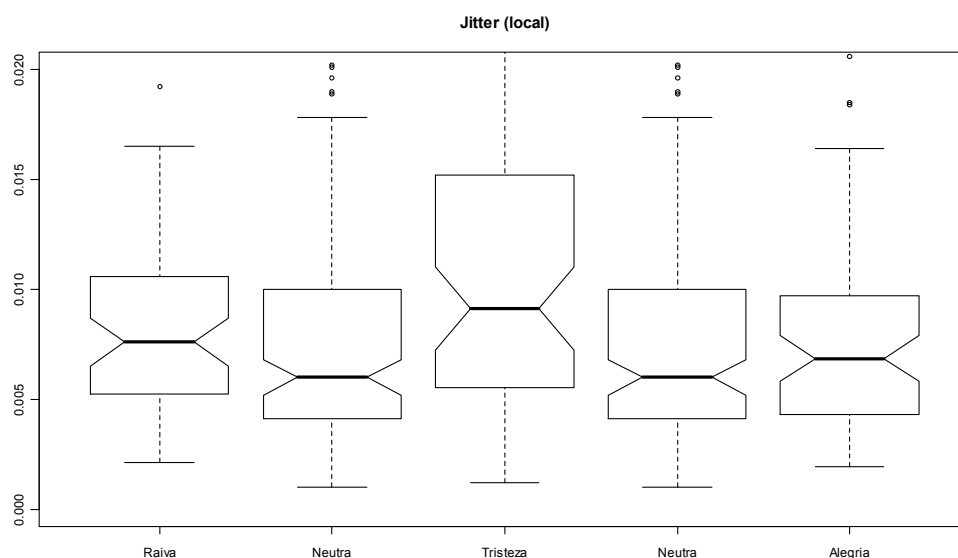


Figura 46 – Valores de *jitter* local para as diferentes expressões estudadas.

O gráfico da Figura 46, 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.

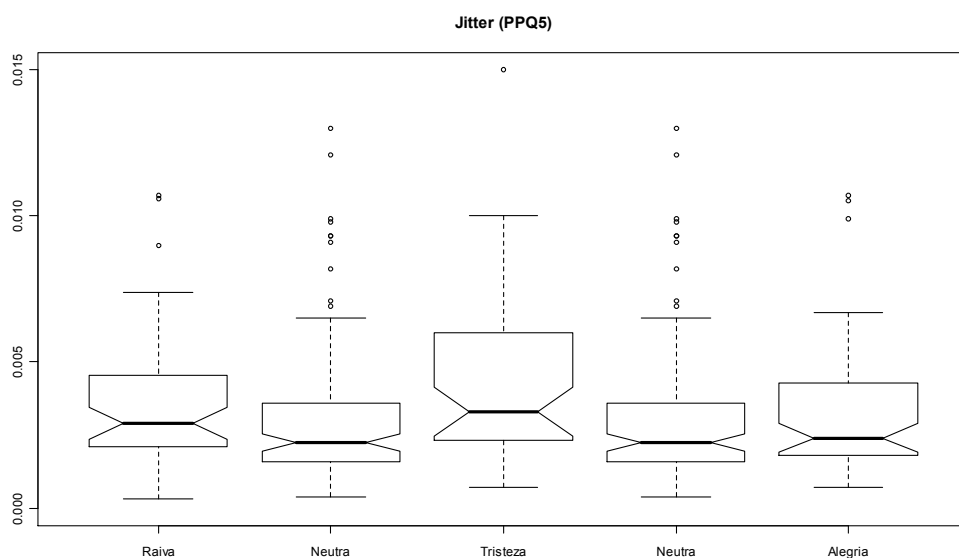


Figura 47 – Valores de *jitter* PPQ5 para as diferentes expressões estudadas.

Analisando o gráfico que contempla os valores de *jitter* PPQ5 para as diferentes expressões, Figura 47, observa-se, uma vez mais, que a tristeza se destaca, apresentando o valor mais elevado para *jitter* PPQ5. Destaca-se, no entanto, que para este parâmetro a alegria e o discurso neutro apresentam os valores mais baixos e em tudo similares, não havendo uma diferença significativa entre estas duas 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.

5.4.1.3 *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 48).

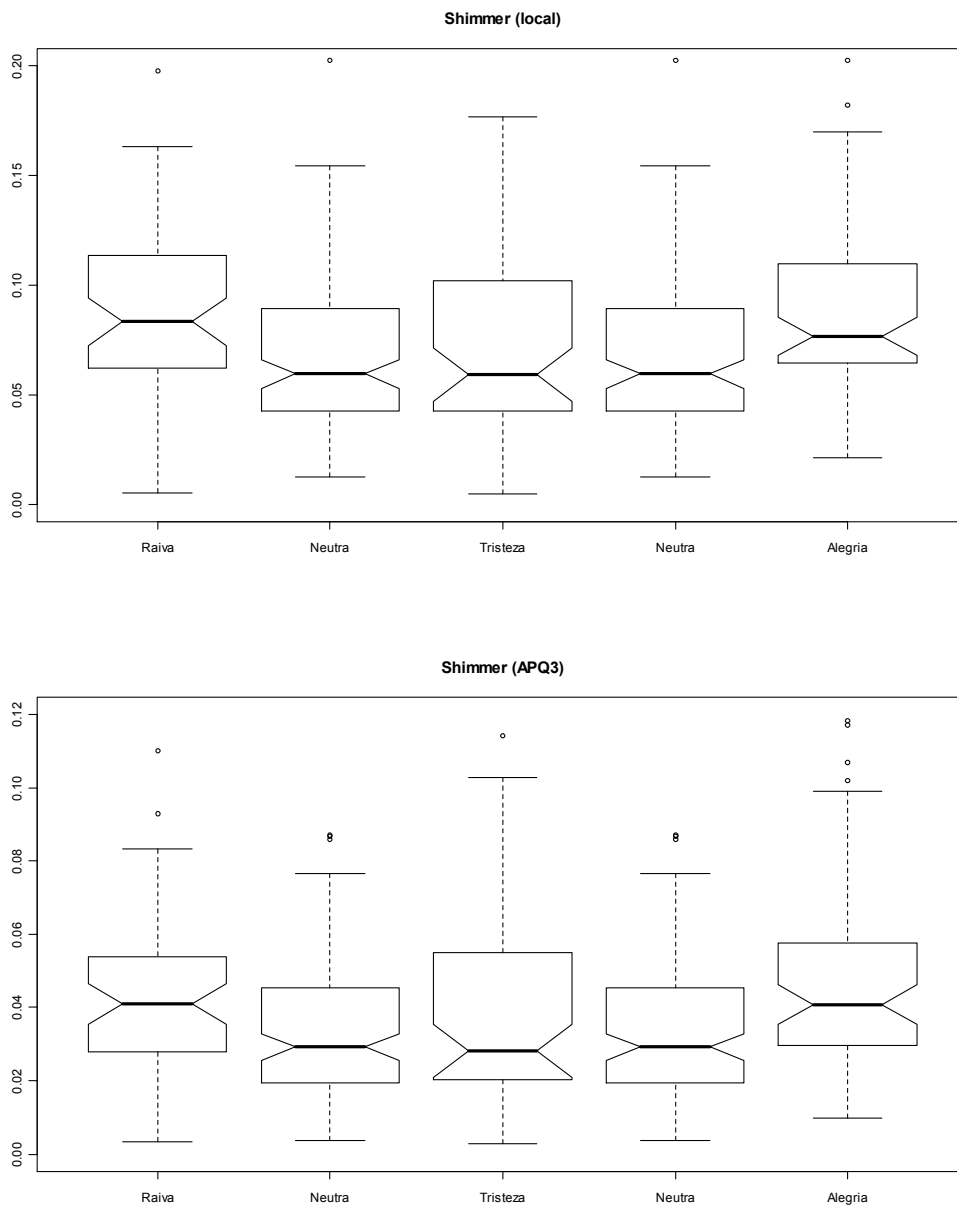


Figura 48 – Valores de *Shimmer* local e APQ3 para as quatro emoções e discurso neutro.

Pela observação dos gráficos da Figura 48, 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.

5.4.1.4 HNR

Os valores obtidos para HNR encontram-se na Figura 49.

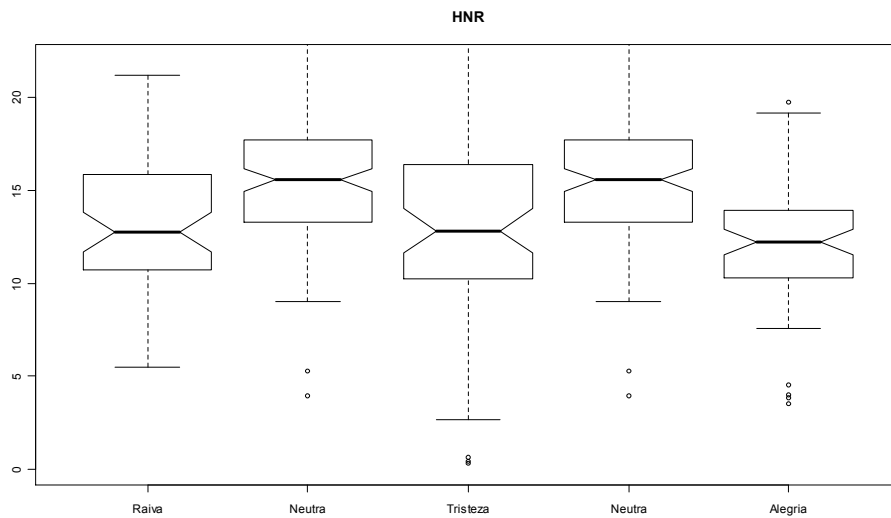


Figura 49 – Valores de HNR para os cinco enunciados estudados.

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.

5.4.2 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, permite 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.

5.4.2.1 *F0*

Nos gráficos das figuras que se seguem (Figura 50 e Figura 51) 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*.

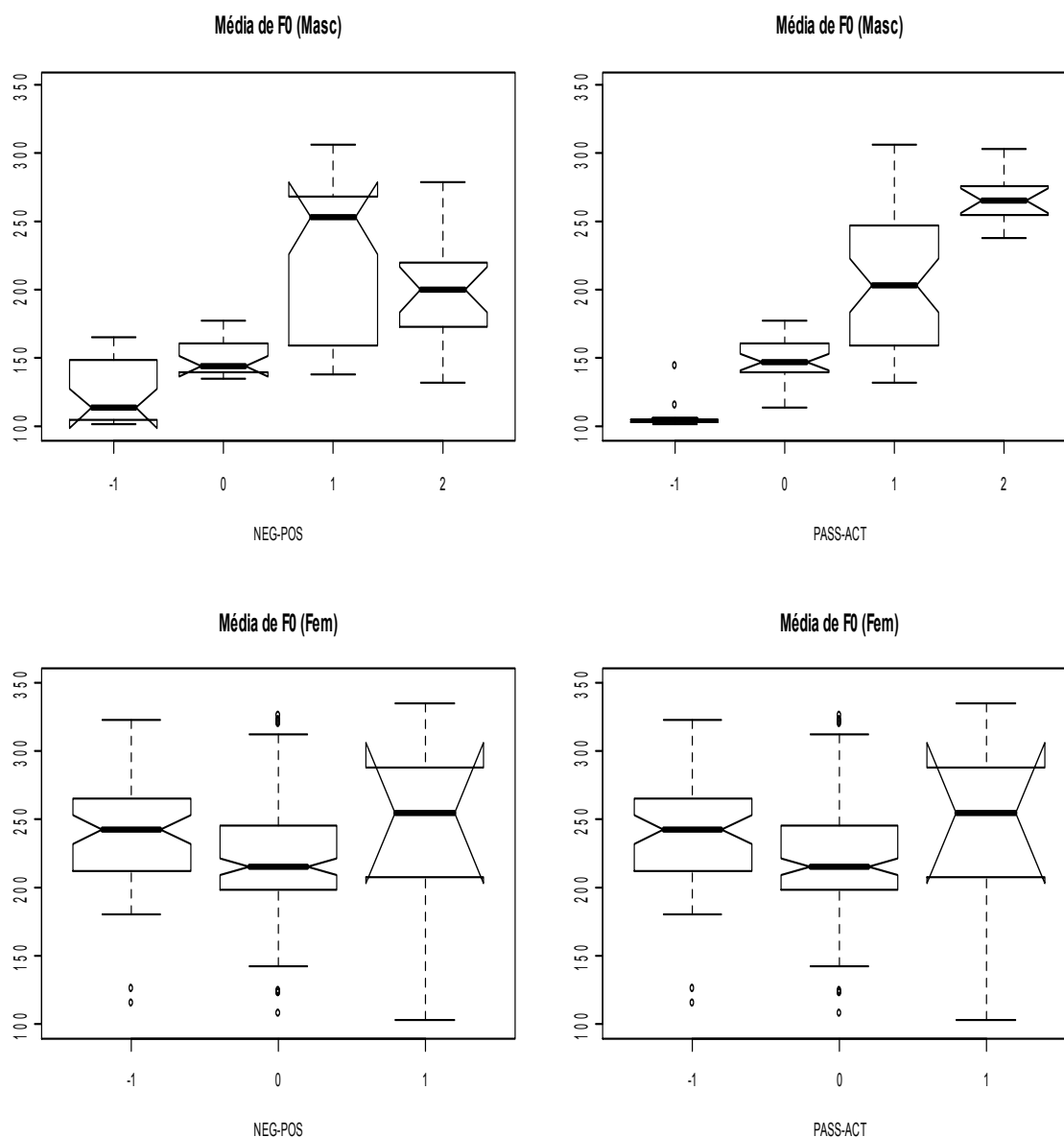


Figura 50 – Valores médios de F0 em função do género e dos eixos do *Feeltrace*.

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.

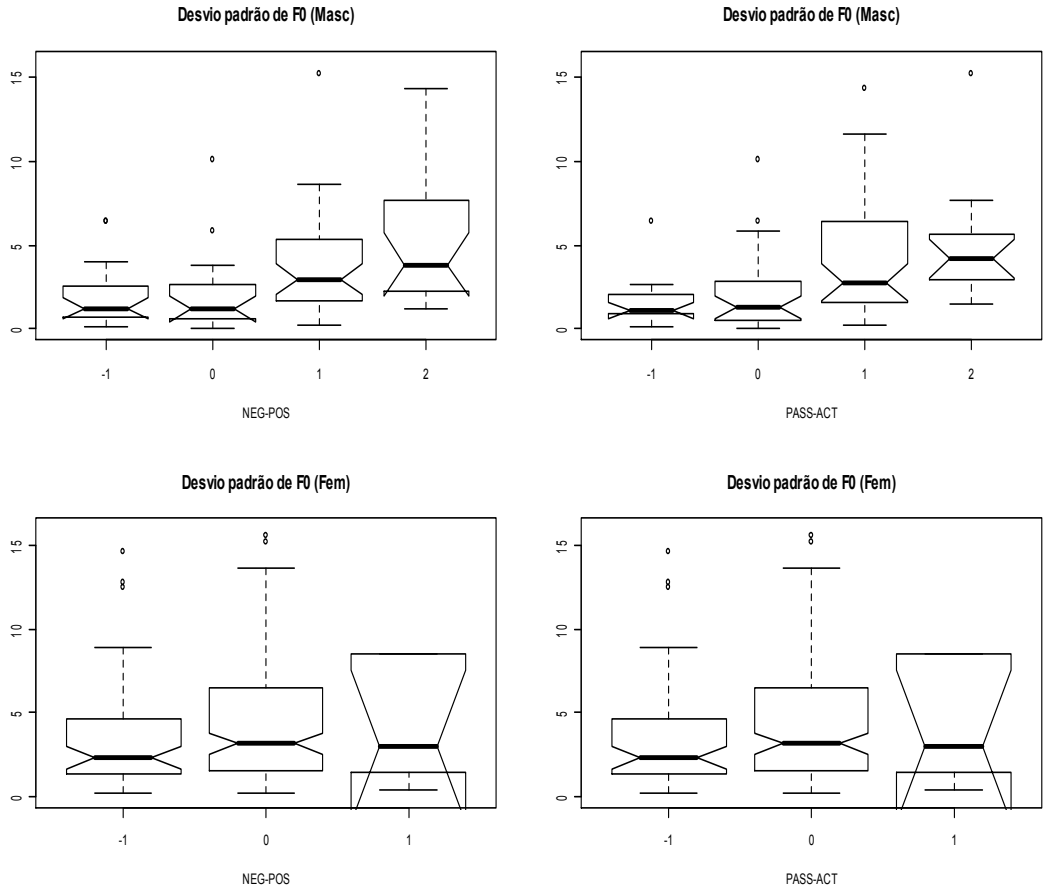


Figura 51 – Desvio padrão de F0 em função do género e eixos *Feeltrace*.

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.

5.4.2.2 *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 52 e Figura 53. Nesta análise da variação não dividimos as marcações para homens e para mulheres.

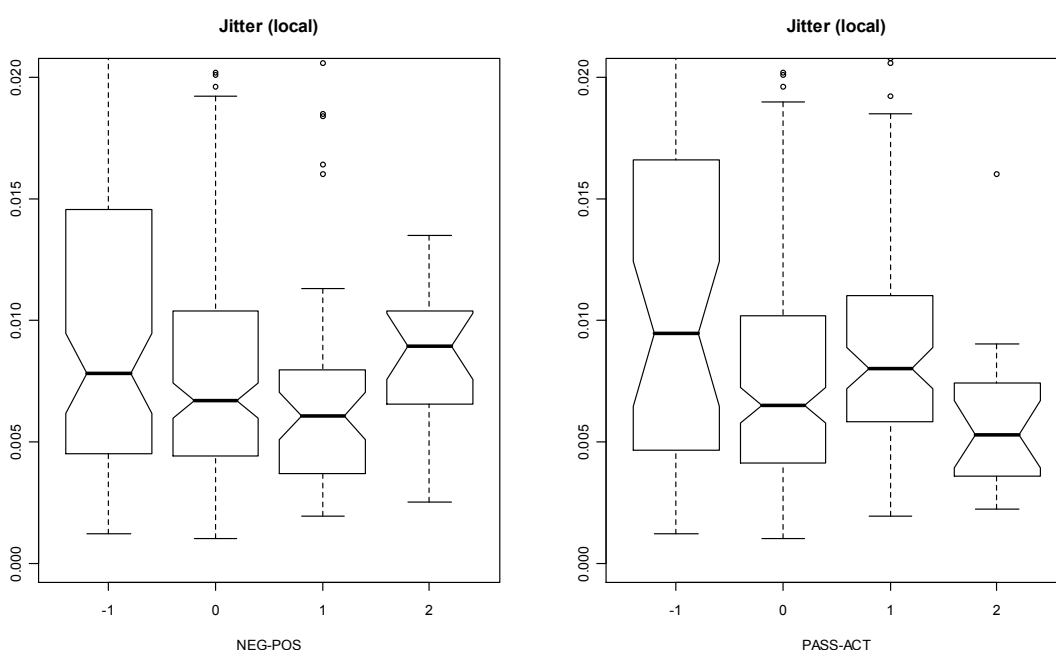


Figura 52 – Valores para *Jitter* local em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).

Os valores de *jitter* local são, claramente, superiores para marcações mais negativas. No eixo passivo-activo, 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.

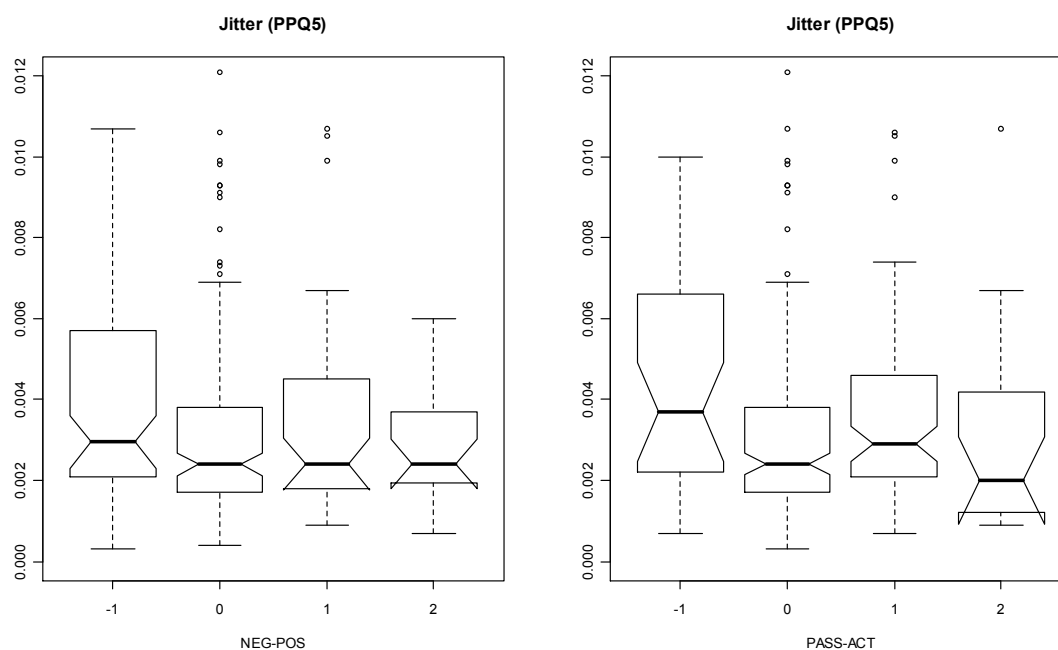


Figura 53 – Valores para *Jitter* PPQ5 em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).

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.

5.4.2.3 *Shimmer*

Também para a estudo do parâmetro relacionado com a amplitude, analisámos dois indicadores diferentes: *shimmer* local e *shimmer* PPQ3, os resultados são possíveis de observar na Figura 54 e seguinte. De referir mais uma vez que aqui também não separámos as análises por género.

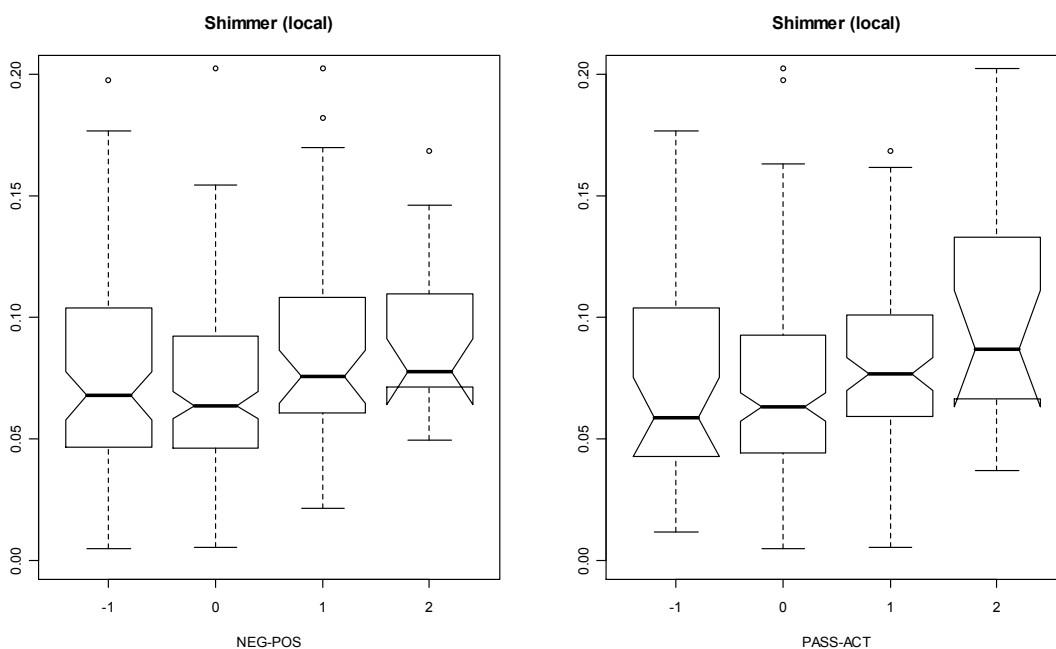


Figura 54 – Valores para *shimmer* local em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).

Nos gráficos da Figura 54 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, percebemos que também os resultados mais altos se encontram na parte referente a activo, sendo evidente o destaque do extremo mais activo (2).

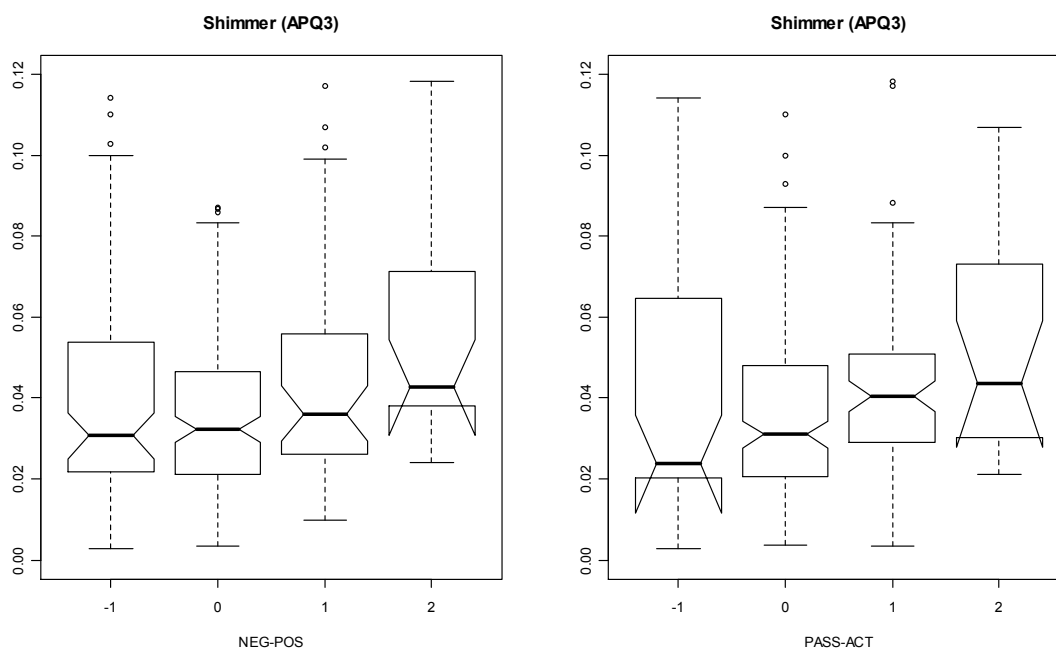


Figura 55 – Valores para *Shimmer* APQ3 em função do eixo negativo-positivo (à esquerda) e passivo-activo.

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.

5.4.2.4 HNR

Os gráficos da Figura 56 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.

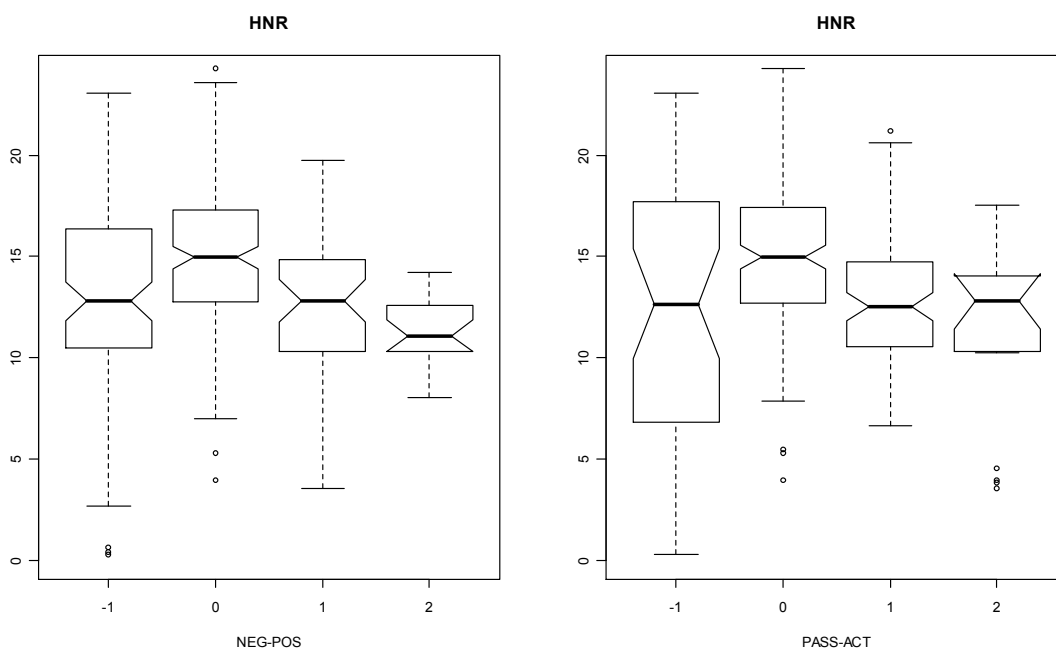


Figura 56 – Valores para HNR em função do eixo negativo-positivo (à esquerda) e passivo-activo (à direita).

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).

5.5. 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 14, 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 14 – Apresentação e comparação dos resultados, para os parâmetros F0; jitter; shimmer e HNR, na produção de emoções por actor e espontâneas.

Parâmetro	Resultado com voz de actor	Em voz espontânea
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 homens²⁴, 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 Alegria²⁵ 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>

²⁴ Não temos para homens os valores de raiva, mas nas mulheres encontra-se próxima dos 225 Hz.

²⁵ É de referir que as gravações de alegria que temos retratam muito mais a euforia, daí os elevados valores de F0.

5.6. Discussão

Importa, no final deste estudo, mostrar como se caracteriza a emoção em PE. Como referimos anteriormente, este capítulo servia não só para essa caracterização e descrição mas, também, para podermos observar se o estudo que efectuámos com voz emocional produzida por um actor pode 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 as análises dos valores relacionados com o F0, existe uma conformidade entre os resultados obtidos na análise dos dois *corpora*. Assim, em termos de média de 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, Scherer (2004) o que corrobora os resultados do nosso estudo. Ressalvamos que a expressão espontânea de alegria que compõe o 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 por actores encontramos, uma vez mais 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 os 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 (ver, por exemplo em anexo, gráficos 3; 7; 8; 15; 16; 19).

5.7. Conclusões

Considerando este capítulo e o anterior, emoção espontânea e emoção por actor, respectivamente, percebemos as diferenças que, fundamentalmente, existem em parâmetros como *jitter* e HNR. O actor "marca", 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).

Contudo o nosso estudo mostra que há algumas 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 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.

Capítulo 6 - Conclusão

6.1. Resumo do trabalho

Começámos, no nosso estudo, por abordar a emoção associada a patologia. Consideraram-se situações em que, não havendo qualquer problema/anomalia física em termos de pregas vocais, existe uma diferente qualidade de voz, ou seja, a diferença existe também de forma involuntária, tal como na expressão de emoção, mas esta advém de problemas do foro psico-emocional. Estudámos e apresentamos dados para a emoção simulada por actores em Português Europeu, análise complementada por um estudo perceptual em que pretendemos perceber de que forma o conhecimento de uma Língua e Cultura permite, ou não, a mais fácil identificação de uma emoção. A nossa questão, no fundo, era apurar até que ponto o conhecimento linguístico e cultural ajudam na percepção das emoções.

O terceiro estudo, o estudo da emoção espontânea, é também complementado, reforçado e validado por um abrangente e original teste, utilizando as funcionalidades do *Feeltrace*. Permitiu-nos aferir os resultados de cada um dos parâmetros relacionados com a voz, agora para o discurso espontâneo. Apesar das limitações deste último estudo, uma vez que é uma

tarefa difícil conseguir material facilmente tratável neste âmbito, conseguimos obter resultados que vêm caracterizar a voz e a emoção em PE.

Esta análise favoreceu ainda a comparação com os dados recolhidos no estudo anterior, para a emoção produzida por actores, mostrando de que forma o profissional consegue, ou não, aproximar-se da expressão de uma emoção e, não o conseguindo, quais são os parâmetros mais difíceis de controlar.

6.2. Principais resultados

Com a realização deste trabalho, acreditamos ter dado a conhecer, não só à comunidade científica portuguesa, mas também e, principalmente, à internacional um pouco mais do que é o Português Europeu. Tivemos esse *feedback* nos encontros internacionais. Outro aspecto que não podemos deixar de sublinhar é o facto deste trabalho ser um estudo original e que apenas utiliza como elementos de análise gravações de voz, não tendo qualquer apoio visual, o que poderia facilitar, em muito, certas descrições. Mas considerámos que esta era, sem dúvida, a melhor forma de estudar a voz e a sua relação com a emoção.

Desde o início confrontámo-nos com o obstáculo, pela dificuldade que é criar um *corpus* de análise para a emoção. No entanto, conseguimos criar *corpora* para emoção de origem psicogénica, analisando e mostrando como os factores psicológicos podem afectar a nossa voz, para que a sua produção se encontre na zona considerada para as patologias. Conseguimos também construir e analisar um *corpus* dedicado ao estudo da produção de emoção por actores, à semelhança do que é feito nas outras línguas e retirar daí importantes resultados. De facto, a nossa voz, mesmo que não nos apercebamos disso é afectada pela emoção e por isso descrevemos os vários parâmetros que caracterizam a qualidade de voz e que mudam consoante a emoção que estamos a sentir e que, evidentemente, se torna perceptível pela voz.

Por outro lado, o *corpus* conseguido para emoção espontânea, com todas as suas implicações que restringem em certa medida todo o trabalho, permitiram-nos verificar que de facto, apesar de grandes semelhanças, a emoção espontânea

pode-se distinguir da simulada, principalmente através de factores relacionados com amplitude e intensidade (*jitter*, *shimmer* e HNR).

O PE apresenta as suas características próprias, sendo a mais marcante a aproximação de emoções de famílias diferentes: tristeza, alegria, e expressão neutra. O que pode mostrar que não somos um povo muito efusivo, ou expansivo na expressão de alegria. Talvez possamos juntar aqui o facto de que a nossa música tradicional seja o Fado, que tão bem retrata o povo Português.

Verificamos ainda que para melhor distinguir emoções numa determinada língua será importante ter algumas noções sobre ela e, também, sobre a cultura do país.

6.3. Sugestões de continuação

Este trabalho não é o fim de um estudo, mas sim o começo de uma nova etapa na análise da emoção para o PE. Esta pesquisa resultou já na aquisição de *corpora* e dados relativos à nossa língua. Verificamos que esta, com tudo o que pode ter de universal, tem também as suas características próprias, apresentando algumas diferenças na expressão de emoção. De salientar que o conhecimento de uma língua e respectiva cultura são factores facilitadores no entendimento intercultural falando, neste caso, da expressão de emoção.

Importa, depois desta abordagem, que se continue a promover a Língua Portuguesa e essa tarefa poderá passar por alargar os *corpora* de análise que aqui iniciámos.

Seria ainda um aspecto relevante o estudo da expressão do Fado cantado, como música referência do nosso país. Seria importante perceber - tendo em conta que falamos aqui de canto, mas também de emoção - usando a mesma análise de parâmetros, se existe alguma relação entre a forma como expressamos alegria ou tristeza quotidianamente e a qualidade de voz e emoção que predominam no Fado.

Por outro lado, poder-se-ão encontrar novas formas de constituição de *corpora* para estudar, mais aprofundadamente, a emoção espontânea, usando também o que já existe, como base, procurando outras fontes, outros meios.

Outro estudo de interesse será estabelecer uma ligação entre a disfonia psicogénica, qualidade de voz e emoção, tentando perceber a que qualidade de voz estão mais ligados estes casos de patologia e, nesses termos, qual será o aspecto emotivo que perpassa.

Outra possibilidade seria procurar obter mais dados sobre a expressão da emoção em PE e procurar fundamentar o estudo da ligação entre o conhecimento de uma língua e sua cultura e o reconhecimento da expressão de emoção. Estudo que pode ser feito alargando os testes perceptivos a outras línguas, utilizando falantes espanhóis e italianos (tão próximos que somos em termos de língua); línguas tonais e eslavas.

Em termos de análise de parâmetros de voz, seria interessante abranger outros parâmetros e obter mais valores. Verificámos que em estudos realizados para outras línguas são também estudados: o *Open Quotient* (OQ); o *Speed Quotient* (SQ) e o *Return Quotient* (RQ).

A continuação deste estudo implicará não só a expansão de *corpora* específicos para o PE, a análise de novos parâmetros mas também a utilização de novos métodos de análise de voz, como o electroglotógrafo (embora este processo possa trazer ainda mais dificuldades na análise da emoção espontânea, por exemplo).

Este trabalho é de facto um estudo transversal e interdisciplinar. Versa todo ele sobre voz e emoção no Português Europeu, em várias áreas e características. Trata-se certamente de um trabalho que pode vir a servir de ferramenta de estudo para Linguística, Engenharia e Ciências da Fala.

Capítulo 7 - Referências

Bibliográficas

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

Abelin, Asa; Allwood, Jens, Cross Linguistic Interpretation of Emotional Prosody, Eurospeech, 2003.

Adell, Jordi; Bonafonte, Antonio; Escudero, David, Analysis of prosodic features: towards modelling of emotional and pragmatic attributes of speech,, Revista de la SEPLN., XXI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, 2005

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.

Airas, Matti; Pulakka, Hannu; Bäckström, Tom; Alku, Paavo, A Toolkit for Voice Inverse Filtering and Parametrisation, InterSpeech, 2005.

Alku, Paavo, Parameterisation methods of the glottal flow estimated by inverse filtering, VOQUAL, 2003.

Alku, Paavo; Airas, Matti; Bäckström, Tom; Pulakka, Hannu, Group delay function as a means to assess quality of Glottal Inverse Filtering, InterSpeech, 2005.

Alter, Kai; Rank, Erhard; Kotz, Sonja A.; Toepel, Ulrike; Besson, Mireille; Schirmer, Annett; Friederici, Angela D.; Accentuation and Emotions – Two Different Systems?, ITRW on Speech and Emotion, 2000.

Amir, Noam, Ziv, Shirley, Cohen, Rachel, Characteristics of authentic anger in Hebrew speech, Eurospeech, 2003.

Amir, Noam; Cohen, Rachel, Characterizing Emotion in the Soundtrack of an Animated Film: Credible or Incredible?, Springer, 2007.

Annibelli, Yara Maria Baggio, Reflexão sobre o Conceito de Disfonia Psicogénica, Monografia apresentada ao CEFAC, 1999.

Aronson, AE, Clinical Voice disorders. Medical Publishers, 1990.

Aubergé, Véronique; Cathiard, Marie, Can we hear the prosody of smile?, Speech Communication, 2003.

Audibert, Nicolas; Aubergé, Véronique; Rilliard, Albert, The Prosodic Dimensions of Emotion in Speech: the Relative weights of Parameters, InterSpeech, 2005.

Audibert, Nicolas; Solange, Rossato; Aubergé, Véronique, Paramétrisation de la qualité de voix: EGG vs filtrage inverse, Institut de la Communication Parlée, 2005.

Baken, R.J.; Orlikoff, Robert F., Clinical Measurement of Speech and Voice, Second Edition, Singular Thomson Learning, 2000.

Baker, Janet, Psychogenic Dysphonia: Peeling Back the Layers, Journal of Voice, Vol.12, No. 4, pp.527-535, 1998.

Baker, Janet, Psychogenic Voice Disorders – Heroes or hysterics? A brief overview with questions and discussion, Log Phon Vocol, 2002.

Baker, Janet, Psychogenic Voice Disorders and Traumatic Stress Experience: A Discussion Paper with Two Cases Report, Journal of Voice, Volume 17, Nr°3, pp.308-318, 2003.

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.

Batliner, Anton, Zeissler, Viktor, Frank, Carmen, Adelhardt, Johann, Shi, Rui, P., Noth, Elmar, "We are not amused-but how do you know? User states in a multi-modal dialogue system", Eurospeech, 2003.

Batliner, Anton; Steide, Stefan; Hacker, Christian; Nöth, Elmar; and Niemann, Heinrich, Private emotions vs. Social Interaction – towards new dimensions in Research on Emotion, Proceedings workshop on Adapting the Interaction Style to Affective Factors, at UM'05.

Batliner, Anton; Huber, Richard, Speaker Characteristics and Emotion Classification, Springer, 2007.

Behlau, M.; Madazio, G.; Feijó, D.; Pontes, P., Disfonias Organofuncionais, Revinter, 2001.

Behlau, Mara (org.), Voz O Livro do Especialista, Volume I, Revinter, 2004.

Behlau, Mara (org.), Voz O Livro do Especialista, Volume II, Revinter, 2005.

Bele, Irene Velsvik, Dimensionality in Voice Quality, Journal of Voice, 2007.

Belin, Pascal; Fecteau, Shirley and Bédard, Catherine, Thinking the Voice: neural correlates of voice perception, Elsevier, 2004.

Björkner, Eva; Sundberg, Johan, MR and Areafunction study of Throaty Voice Quality, VOQUAL, 2003.

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.

Braga, Júnia Novaes; Oliveira, Domingos Sávio Ferreira de; Sampaio, Tania Maria Marinho, Frequência Fundamental da Voz de Crianças, Revista CEFAC, 2009.

Brückl, Markus, Sendlmeier, Walter, Aging female voices: an acoustic perceptive analysis, VOQUAL, 2003.

Bryant, Gregory A.; Tree, Jean E. Fox, Is there an Ironic Tone of Voice?, Language and Speech, 2005, 48, pp. 257-277.

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.

Burkhardt, Felix, Emofilt: the simulation of Emotional Speech by Prosody-Transformation, InterSpeech, 2005.

Burkhardt, Felix, Sendlmeier, Walter F., Verification of Acoustical Correlates of Emotional Speech using Formant-Synthesis. ISCA Workshop on Speech and Emotion: A conceptual framework for research, 2003.

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.

Cahn, Janet E., Generating Expression in Synthesized Speech. Master’s Thesis, Massachusetts Institute of Technology, 1989.

Calvi, Angela, Disfonia Espasmódica, Centro de Especialização em Fonoaudiologia Clínica, Rio de Janeiro, 1999.

Campbell, Nick, Accounting for Voice-Quality Variation”, Speech Prosody, 2004.

Campbell, Nick; Changes in Voice Quality Due To Social Conditions, ICPHS XVI, 2007.

Campbell, Nick; Kashioka, Hideki; Ohara, Ryo, No Laughing Matter, InterSpeech, 2005.

Campbell, Nick; Mokhtari Parham, Voice Quality: 4th Prosodic Dimension, 15th ICPHS, 2003.

Cannito, Michael P.; Buder, Eugene H.; Chorna, Lesya B., Spectral amplitude measures of adductor spasmodic dysphonic speech, Journal of Voice, 2004.

Capková, Sofia Gustafson, Emotions in Speech: Tagset and Acoustic, Speech technology, 2001.

Cardoso, Fabíola Pereira, Disfonia Psicogénica e os mecanismos subjacentes, Monografia apresentada ao Centro de Especialização em Fonoaudiologia Clínica, Porto Alegre, 1999.

Casper, J. K.; Colton, R. H., Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment, Baltimore: Williams and Wilkins Inc., 1996.

Cauldwell, Richard T., Where did the anger go? The role of context in interpreting emotion in speech”, ISCA Workshop on Speech and Emotion: A conceptual framework for research, 2003.

Chung, Soo-Jin, Expression and Perception of Emotion extracted from the Spontaneous Speech in Korean and in English, PhD. Dissertation, Sorbonne, France, 2000.

Constantino, Tânia; Guimarães Isabel, Influência da Duração da Disfonia na Qualidade Vocal e seu Impacto Psicossocial em Mulheres, Re(habilitar), Nrº1, Edições Colibri, pp.2-24, 2005.

Cornellius, Randolph R., Theoretical Approaches to Emotion, ISCA Workshop on Speech and Emotion: A conceptual framework for research, 2003.

Cowie, E. Douglas; Cowie, R.; Romano, A., Changing emotional Tone in Dialogue and its prosodic correlates, ETRW on Dialogue and Prosody, 1999.

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, 15th 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.

D'Allessandro, Voice Quality Modification for Emotional Speech Synthesis, Eurospeech, 2003.

Damásio, António R., O Erro de Descartes, Emoção, Razão e Cérebro Humano, Fórum da Ciência, Publicações Europa-América, 16ª edição, 1996.

Darwin, Charles, A Expressão Das Emoções No Homem e Nos Animais, Setembro 2006, Relógio d'Água editores.

Demenko, Grazyna, Suprasegmental analysis for complex quality assessment in Pathological Voices, Speech Prosody, 2004.

Dias, Susana Patrícia, Fonação em Disartria Atáxica pós TCE – Análise Acústica, Tese de Mestrado, 2008.

Dioubina, Olga I., Annotation of expressive speech, VOQUAL, 2003.

Dolan, R.J.; Morris, J.S.; Gelder, B., Crossmodal binding of fear in voice and face, PNAS, Volume 98, Nrº 17, 2001.

Drahota, Amy; Costall, Alan; Reddy Vasudevi, The Vocal Communication of Different Kinds of Smile, Speech Communication, 2007.

Drioli, Carlo; Tisato, Graziano; Cosi, Piero; Tesser, Fábio, Emotions and Voice Quality: Experiments with Sinusoidal Modeling, VOQUAL, 2003.

Edmondson, Jerold A.; Padayodi, Cécile M.; Hassan, Zeki Majeed; Esling, John H., The Laryngeal Articulator: Source And Resonator, ICPHS XVI, 2007.

Ekman, Paul, An Argument for Basic Emotions, Cognition and Emotion, 6, 169-200, 1992.

Epstein, Melissa Ann, Voice Quality and Prosody in English, PhD, [University of California, Los Angeles](#), 2002.

Erickson, Donna, Expressive speech: Production, perception and application to speech synthesis, Acoustic Science and technology, 2005.

Fawcus, M.; Voice Disorders and their Management, London, 2001.

Fernandez, Raul; Picard, Rosalind W., Modeling Drivers' under stress, ITRW on Speech and Emotion, 2000.

Fredouille, C.; Pouchoulin, G.; Bonastre, J.-F.; Azzarello, M.; Giovanni, A.; Ghio, A., Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia), InterSpeech, 2005.

Frijda, N. H.; The Emotions, Cambridge University Press, 1986.

Frölich, Matthias; Michaelis, Dirk; Strube, Hans Werner, Acoustic "Breathiness measures" in the description of Pathologic voices, 1998.

Frölich, Matthias; Michaelis, Dirk; Strube, Hans Werner; Kruse, Eberhard, Acoustic voice quality description: case studies for different regions of the hoarseness diagram", The Pennsylvania State University CiteSeer Archives, 1997.

Frölich, Matthias; Michaelis, Dirk; Strube, Hans Werner; Kruse, Eberhard, Acoustic Voice Analysis by Means of the Hoarseness Diagram, Journal of Speech, Language and Hearing Research, Vol.43; pp. 706-720, 2000.

Furui, Sadaoki, Recent Advances in spontaneous speech and understanding, ISCA & IEEE Workshop processing and recognition, 2003.

Gilvaneia, Francisca; Saudade, Maria da; Felix, Sydia, A Voz, Universidade Federal do Rio Grande do Norte, sem data.

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.

Gobl, Christer; Bennet, Eva; Chasaide, Ailbhe Ní, Expressive Synthesis: How Crucial is Voice Quality?, IEEE, 2002.

Gómez, Pedro; Díaz, Francisco, Álvarez, Augustin; Martínez, Rafael; Rodellar, Victoria; Baillo, Roberto Fernández; Nieto, Alberto; Fernández, Francisco J., PCA of perturbation in Voice Pathology detection, InterSpeech, 2005.

Gordon, Mathew, Lodefoged, Peter, Phonation types: a cross-linguistic overview, Journal of Phonetics, 29, 383-406, 2001.

Guimarães, Isabel, Abberton, Evelyn, An Electrolaryngographic study of Dysphonic Portuguese Speakers, 15th ICPHS, 2003.

Guimarães, Isabel, Abberton, Evelyn, Fundamental frequency in speakers of Portuguese for different voice samples, Journal of Voice, 2004.

Guimarães, Isabel, A Ciência e a Arte da Voz Humana, Escola Superior de Saúde de Alcoitão, Lisboa, 2007.

Gutiérrez-Arriola, J.M.; Montero, J.M.; Córdoba, R.; Pardo, J.M., Analysis of parameter importance in speaker identity, VOQUAL, 2003

Hammarberg, B., Perceptual evaluation of dysphonic voices, Larynx, 1997.

Harris, Sara, Voice Production and Voice Care, The British Voice Assotiation, 2007.

Henton, Caroline, Challenges and Rewards in using parametric or concatenative Speech Synthesis, International Journal of Speech Technology, 5, pp.117-131, 2002.

Hirano, M.; Kurita, S.; Nakashima, T.; The structure of vocal folds, University of Tokyo Press, Tokyo, 1981.

Hirose, H., Insvestigating the Physiology of Laryngeal Structures, Hardcastle W.J. Laver J., (eds), The Handbook of Phonetic Sciences, Blackwell Publishers Ltd., Oxford, 1996.

Hirschberg, Julia, Communication and Prosody: Functional aspects of Prosody, ETRW on Dialogue and Prosody, 1999.

Hirschberg, Julia; Liscombe, Jackson; Venditti, Jennifer, Experiments in Emotional Speech, ISCA & IEEE Workshop on Spontaneous Speech and Recognition, 2003.

<http://www.phon.ucl.ac.uk/resource/sfs>.

lida, Akemi, Mokhtari, Parham, Campbell, Nick, Acoustica correlates of monosyllabic utterances of Japanese in different speaking Styles, 15th ICPhS, 2003.

Imamura, Rui; Tsuji, Domingos H., Disfonia espasmódica de adução, tremor vocal e disfonia de tensão muscular: é possível fazer o diagnóstico diferencial?, Revista Brasileira de Otorrinolaringologia 72, 2006.

Inglehart, R.; Klingemann, H., Genes, Culture, Democracy, and Happiness, Culture and Subjective Well-being, Cambridge, MA: MIT Press, 2000.

Ito, Mika, The contribution of Voice Quality to politeness in Japanese, VOQUAL, 2003.

Ishi, Carlos, Toshinori; Ishiguro, Hiroshi; Hagita, Norihiro, Automatic extraction of paralinguistic information using prosodic features related to F0, duration and voice quality, Speech Communication, 2008.

Johnstone, T., Scherer, K.R., The effects of Emotion on Voice Quality, Proceedings of the XIVth ICPhS, 1999.

Johnstone, T., Scherer, K.R., The effects of Emotion on Voice Quality, ICPhS, 2003.

Jurjec, Peter, Creaky Voice in Sloven, III Congresso de Fonética Experimental, Santiago de Compostela, 2005

Kawahara, Hideki, Exemplar-based Voice Quality analysis and control using a high quality auditory morphing procedure based on STRAIGHT, VOQUAL, 2003

Kehrein, Roland, The Prosody of authentic Emotions, Speech Prosody 2002. Proceedings of the first International conference on Speech Prosody, pp.423-426, 2002.

Keller, Eric, The analysis in Voice Quality in Speech processing, 2004.

Keller, Eric, Voice characteristics of MARSEC speakers, VOQUAL, 2003.

Khidr, Aliaa, Effects of the “Smith Accent Technique” of voice therapy on the laryngeal functions and voice quality of patients with unilateral vocal fold paralysis, International Congress Series, Elsevier, 2003.

Kiecolt-Glaser, Janice, K.; Yehuda, Rachel, Toward optimal health. Discuss chronic stress in women, *Journal of women's health*, 2005.

Kienast, Miriam, Sendlmeier, Walter F., Acoustical Analysis of Spectral and Temporal Changes in Emotional Speech, *ISCA Workshop on Speech and Emotion: A conceptual framework for research*, 2003.

Klasmeyer, Gudrun, Johnstone, Tom, Banziger, Tanja, Sappok, Christopher, Scherer, Klaus R., Emotional Voice Variability in Speaker Verification,, *ISCA Workshop on Speech and Emotion: A conceptual framework for research*, 2003.

Kostoulas, Theodoros; Ganchev, Todor; Mporas, Iosif; Fakotakis, Nikos, A Real-World Emotional Speech Corpus for Modern Greek, *LREC*, 2008.

Kreiman, Jody; Gerratt, R. Bruce; "jitter", "shimmer", and Noise in Pathological Voice Quality Perception, *VOQUAL 2003*.

Kreiman, Jody; Vanlancker-Sidtis, Diana; Gerratt, R. Bruce, Defining and Measuring Voice Quality, *VOQUAL 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.

Laver, John, The phonetic description of voice quality, Cambridge, University, 1980.

Lee, Sungbok; Yildirim, Serdar; Kazemzadeh, Abe; Narayanan; Shrikanth, An Articulatory study of emotional speech production, *InterSpeech*, 2005.

Leeuw, Irma M. Verdonck-de; Beinum, Florian J. Koopsmans-van, Voice Quality before and after radiotherapy: various acoustical, clinical and perceptual pitch measures, *Journal of Speech and Hearing Research*, 1995.

Li, Nicole Y. K.; Yiu, Edwin M-L, Acoustic and perceptual analysis of modal and falsetto registers in females with dysphonia, *Clinical Linguistics and Phonetics*, 2006.

Lintfert; Britta; Wokurek, Wolfgang, Voice Quality dimensions of “pitch” accents, *InterSpeech*, 2005.

Liscombe, Jackson, Detecting and Responding to Emotion in Speech: Experiments in Three Domains, Ph.D. Thesis, Department of Computer Science Columbia University, 2006

Liscombe, Jackson, Venditti, Jennifer, Hirschberg, Julia, Classifying Subject Ratings of Emotional Speech Using Acoustic Features, *Eurospeech*, 2003.

Luengo, Iker; Navas, Eva; Hernáez, Inmaculada; Sánchez, Jon, Automatic emotion recognition using prosodic parameters, *InterSpeech*, 2005.

Maddieson, Ian; Ladefoged, Peter, *The sounds of the world's languages*. Oxford: Blackwell, 1996.

MacKenzie, Kenneth; Millar, Audrey; Wilson, Janet A.; Sellars, Cameron; Deary Ian J., Is Voice therapy an effective treatment for dysphonia? A randomised controlled trial, *BMJ*, 2001.

Maeda, Shinji, Toda, Martine, Mechanical properties of lip movements: How to characterize different speaking styles?, *15th ICPHS*, 2003.

Makarova, Veronika, Petrushin, Valleri A., Phonetics of Emotion in Russian Speech, *15th ICPHS*, 2003

Makarova, Veronika, Petrushin, Valleri A., The Map Task Corpus of spoken Russian, *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, 2003.

Manfredi, Claudia; Dori, Fabrizio; Iadanza, Ernesto, Improvement in hoarse voice denoising for real-time DSP implementation, *VOQUAL*, 2003.

McAllister, Anita; Sederholm, Elisabeth; Ternström; Sundberg, Johan, Perturbation and Hoarseness: A Pilot Study of Six Children’s Voices, *Journal of Voice*, Vol.10; No. 3, pp. 252-261, 1996.

Meltzner, Geoffrey S.; Hillman, Robert E., Impact of abnormal acoustic properties on the perceived quality of electrolaryngeal Speech, VOQUAL, 2003.

Mendoza, Elvira; Valencia, Nieves; Muñoz, Juana; Trujillo, Humberto, Differences in Voice Quality Between Men and Women: Use of the Long-Term Average Spectrum (LTAS), Journal of Voice, Volume 10, pp.59-66, 1996.

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

Monzo, Carlos; Alías, Francesc; Iriondo, Ignazi; Gonzalvo, Xavier; Planet, Santiago, Discriminating Expressive Speech Styles By Voice Quality Parameterization, ICPHS, 2007.

Murray, I. R.; Arnott, J. L. *Synthesizing emotions in speech: Is it time to get excited?*, Proc. ICSLP, 1996, pp. 1816-1819.

Murray, I.R.; Arnott, John L., Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. Acoustical Society of America, Volume 93, pp.1097-1108, 1993.

Ní, Chasaide, Gobl, Christer, Voice source variation, Hardcastle W.J. Laver J., (eds), The Handbook of Phonetic Sciences, Blackwell Publishers Ltd., Oxford, 1997.

Ogden, Richard, Voice Quality as a resource for the management of turn-taking in Finnish talk-in-interaction, 15th ICPHS, Barcelona, 2003.

Oliveira, Valéria Leal de, A qualidade da voz e o trato vocal nos indivíduos de face curta e face longa, Monografia apresentada ao CEFAC, 1999.

Ortony, Andrew; Turner, J. Terence, What's Basic About Basic Emotions?, Psychological Review, 1990.

Oudeyer, P.-Y. The Synthesis of cartoon Emotional Speech, First International Conference on Speech Prosody (Speech Prosody). pp. 551-554, 2002.

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

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.

Pereira, Cécile, Dimensions of emotional meaning in speech, ISCA Workshop on Speech and Emotion, 2000.

Piot, Olivier, Attitudes and Yes-No Questions in Standard French: Testing Two Hypotheses, ISCA Workshop on Speech and Emotion: A conceptual framework for research, 2003.

Psutka, Josef; Psutka, J. V.; Ircing, Pavel; Hoidekr, Jan, Recognition of spontaneously pronounced TV Ice-Hockey commentary, ISCA & IEEE Workshop on Spontaneous Speech processing and recognition, 2003.

Pulakka, Hannu, Analysis oh Human Voice Production Using Inverse Fitering, High-Speed Imaging, and Electroglottography, Master's Thesis, Helsinki University of Technology, 2005.

Rahurkar, Mandar, A., Hansen, John, H.L., Frequency distribution based weighted sub-band approach for classification of emotional/stressful content in Speech, Eurospeech, 2003.

Ren, Fuji, Affective Information Processing and Recognizing Human Emotion, Elsevier, 2008.

Rigoll, Gerhard, An overview on European projects related to spontaneous speech recognition, ISCA & IEEE Workshop on Spontaneous Speech and Recognition, 2003

Ringeval, Fabien; Chetouani, Mohamed, Exploiting a Vowel Based Approach For Acted Emotion Recognition, Interspeech, 2008.

Roach, Peter, Techniques for the Phonetic Descriptions of Emotional Speech, ISCA Workshop on Speech and Emotion: A conceptual framework for research, 2003.

Rodrigues, Américo, As Emoções na Fala, Tese de Mestrado apresentada à Universidade de Aveiro, 2007

Rosenberg, Andrew, Hirschberg, Julia, Acoustic/Prosodic and lexical correlates of charismatic speech, InterSpeech, 2005.

Rubin, John S., Greenberg, Maurice, Psychogenic Voice Disorders in Performers: A Psychodynamic Model, Journal of Voice, Volume 16, Nrº4. pp.544-548, 2002.

Russel, James A.; Bachorowski, Jo-Anne; Dols, José Miguel Fernandez, Facial and Vocal Expressions of Emotion, Psychology Review, 2003.

Ryan, C., Chasaide, A.N., Gobl, C. Voice Quality and Perception of Affect: Continuous or Categorical, 15th ICPHS, 2003.

Sapienza, Christine M.; Stathopoulos, Elaine T., Speech Task Effects on Acoustic and Aerodynamic measures of Women with Vocal Nodules, Journal of Voice, Volume 9, No.4, pp.413-418, 1995.

Sato, Yuji, Voice Quality Conversion using interactive evolution of Prosodic control, Elsevier, 2004.

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, K.R., Emotion psychology can contribute to psychiatric work on affect disorders: a review. J R Soc Med 82, 545—547, 1989.

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

Schirmera, A., Kotz, S.A. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing, Volume 10, Nr° 1, Janeiro, Cognitive Sciences, Elsevier, 2006.

Schötz, Susanne, Prosody in relation to paralinguistic Phonetics – earlier and recent definitions, distinctions and discussions, Prosody Term paper, 2003.

Schröder, Marc, Determination of the geometry of the human vocal tract by acoustic measurements, Journal of the Acoustic Society of America, 41, 1967.

Schröder, Marc, Can emotions be synthesized without controlling Voice Quality?, Phonus 4, Institute of Phonetics, University of the Saarland, pp.35-50, 1999.

Schröder, Marc, Emotional Speech Synthesis: a Review, EuroSpeech, 2001.

Schröder, Marc, Experimental Study of Affect Bursts, Speech and Emotion, 2000.

Schröder, Marc; Devillers, Laurence; Karpouzis, Kostas; Martin, Jean Claude; Pelachaud, Catherine; Peter, Christian; Pirker, Hannes; Schuller, Björn; Tao, Jianhua; Wilson, Ian, What should a Generic Markup Language Be Able to Represent?, Springer, 2007.

Schröder, Marc; Speech and Emotion research, an overview of research Frameworks and a dimensional approach to emotional speech synthesis, PhD thesis, 2004.

Schröder, Marc; Véronique, Aubergé; Marie-Agnès, Cathiard, Can we hear smile?, ICSLP, 1998.

Seifert, Eberhard; Kollbrunner, Juerg, An update in thinking about nonorganic voice disorders, Arch Otolaryngol Head Neck Surgery, Volume 132, Outubro, 2006.

Seppänen, Tapio; Juhani Toivanen and Vayrynen, Eero, Media Team Speech Corpus: a first large Finnish emotional speech database, 15th ICPHS, 2003.

Seppänen, Tapio; Juhani Toivanen and Vayrynen, Eero, Prosody-based classification of emotions in spoken Finnish, Eurospeech, 2003.

Shahid, Suleman; Krahmer, Emiel; Swerts, Marc, Real vs. Acted Emotional Speech: Comparing South-Asian and Caucasian Speakers and Observers, 2008.

Shriberg, Elizabeth, Spontaneous Speech: How People really talk and why Engineers should care, InterSpeech, 2005.

Soyama, Camila Kayoko; Espassatempo, Cibelle De Longhi; Gregio, Fabiana Nogueira; Camargo, Zuleica, Qualidade vocal na terceira idade: Parâmetros acústicos de longo termo de vozes masculinas e femininas, Revista CEFAC, São Paulo, V.7, n.2; pp.267-279, 2005.

Spreckelmeyer, Katja N.; Kutas, Marta; Urbach, Thomas; Altenmüller, Eckart; Münte, Thomas F., Neural processing of vocal emotion identity, Brain and cognition, 2009.

Strangert, Eva, Prosody in public speech: analyses of a news announcement and a political interview, InterSpeech, 2005.

Strik, H.; Binnenpoorte, D.; Cucchiarini, C., Multiword Expressions in Spontaneous Speech: Do we really speak like that?, InterSpeech, 2005.

Susca, M.; Healey, E. C., Listener perceptions along a fluency-disfluency continuum: A phenomenological analysis, Journal of Fluency disorders, 27, 2002, pp. 135-161.

Takahash, Toru, Fujii, Takeshi; Nishi, Masashi; Banno, Hideki; Irino, Toshio; Kawahara, Hideki, Voice and Emotional Expression transformation based on statistics of Vowel parameters in Emotional Speech Database, InterSpeech, 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.

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

Titze, Ingo R., The Physics of small-amplitude oscillation of the vocal folds, JASA, 1988.

Thompson, J.; Mason, J. S., Effects of Anxiety in Visual and Audio Speech Databases, ESCA-NATO workshop on Speech under Stress, 1995.

- Toivanen, J., Waaramaa, T., Alku, P., Laukkanen, A.-M., Seppänen, T., Väyrynen, E., Airas, M, .Emotions in [a]: a perceptual and acoustic study. *Logoped Phoniatr Vocol* 31, 43—48, 2006
- Toivanen, Juhani; Väyrynen, Eero and Seppänen, Tapio, Gender differences in the ability to discriminate emotional content from speech, *Proceedings, FONETIK 2005*, Department of Linguistics, Göteborg University.
- Tooher, Michelle; McKenna, John G., Variation of glottal LF parameters across F0, vowels, and phonetic environment, *VOQUAL*, 2003.
- Turk, Oytun; Schroder, Marc; Bozkurt, Baris; Arslan, Levent M., Voice Quality Interpolation for Emotional Text-to-Speech Synthesis, *InterSpeech*, 2005.
- Véronique, Aubergé, Ludovic, Lemaître, The Prosody of smile, *ISCA Workshop on Speech and Emotion: A conceptual framework for research*, 2003.
- Vidrascu, Laurence; Devillers, Laurence, Detection of real-life emotions in call centers, *InterSpeech*, 2005
- Vilarinho, Helena, *Estudo de Voz Laringectomizados Falantes de Português Europeu*, dissertação de Mestrado apresentada à Universidade de Aveiro, 2008.
- Vine, D.S.G., Sahandi, R., Synthesizing Emotional Speech by Concatenating Multiple “pitch” Recorded Speech Units, *ISCA Workshop on Speech and Emotion: A conceptual framework for research*, 2003.
- Vogt, Thuriid, André, Elisabeth, Comparing Feature Sets For Acted And Spontaneous Speech in View of Automatic Emotion Recognition, *IEEE*, 2003.
- Vroomen, Jean, Collier, René; Mozziconacci, Sylvie, Duration and intonation in Emotional Speech, *Proceedings of Eurospeech*, pp.577-580, Germany, 1993.
- Wendler, J., Voice disorders. Emphasis in diagnosis and therapy, *Laryngo-rhinotologie*, 1996.
- Weismer, Gary, “Motor Speech Disorders”, *Handbook of Phonetic Sciences*, Chapter 7, Blackwell, 1997.
- Whiteside, Sandra, P.; Marshall, Jeni, Voice Onset Time Patterns in 7-, 9- and 11-Year Old Children, *ISCLP*, 1998.

Wilchmann, Anne, The attitudinal effects of Prosody, and how they relate to emotion, ITRW on Speech and Emotion, 2000.

Wilting, Janneke; Kraemer, Emiel; Swerts, Marc, Reals vs Acted Emotional Speech, Speech Communication, 2005

Wu, Wei; Zheng, Fan; Xu, Ming-Xing; Bao, Huan-Jun, Study on Speaker Verification on Emotional Speech, ISCSLP, 2002.

www.ims.uni-stuttgart.de/phonetik

Yacoub, Sherif, Simske, Steve, Lin, Xiaofan, Burns, John, Recognition of Emotions in Interactive Voice Response Systems, Eurospeech, 2003.

Yanushevskaya, Irena; Gobl, Christer; Chasaide, Ní, Voice Quality and Loudness in Affect Perception, Speech Prosody, 2008.

Yu, Ping; Ouaknine, Maurice; Revis, Joana; Giovanni, Antoine, Objective Voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements, Journal of Voice, Vol.15, 2001.

Zemlin, W, Princípios de Anatomia e Fisiologia em fonoaudiologia, Porto Alegre, 2000.

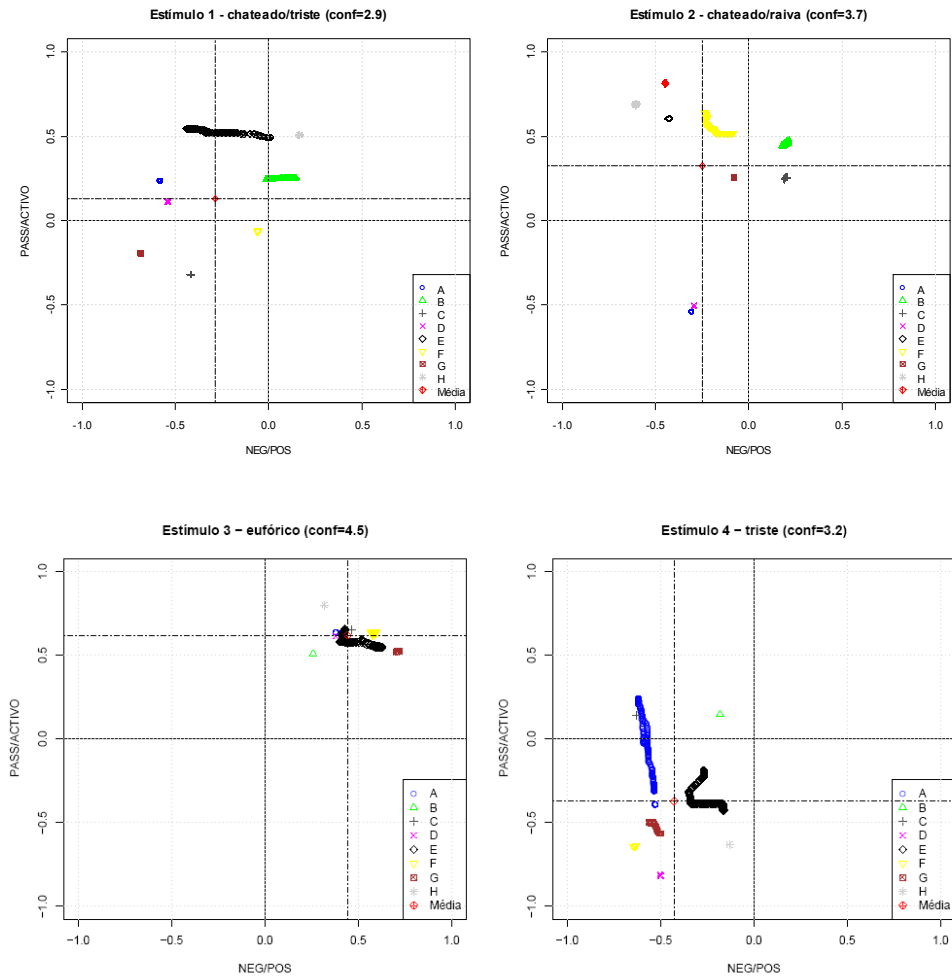
Zetterholm, Elisabeth, The same but different – three impersonators imitate the same target voices, 15th ICPHS, 2003.

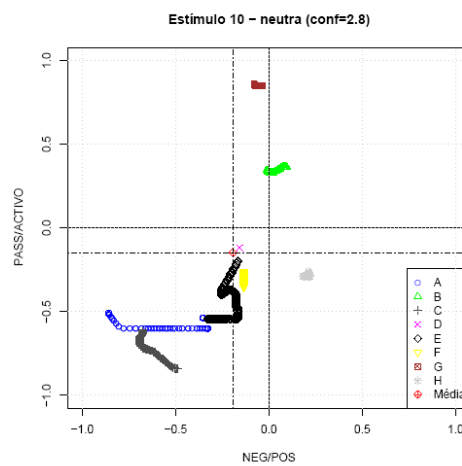
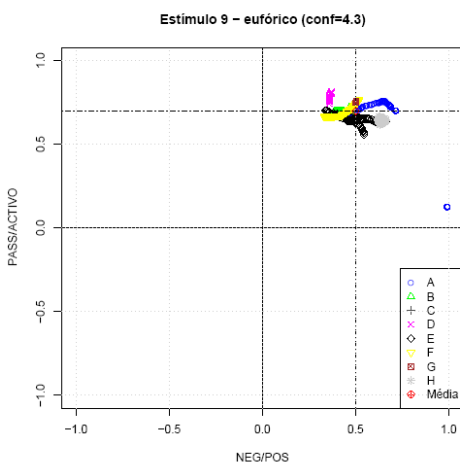
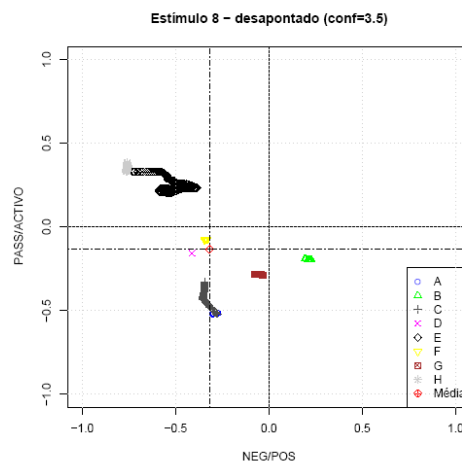
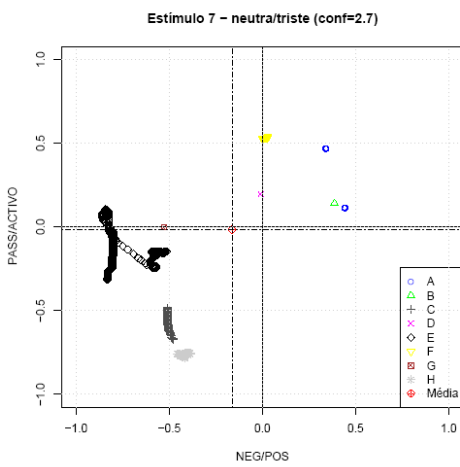
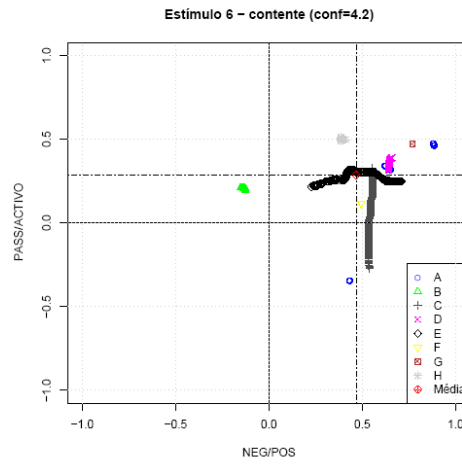
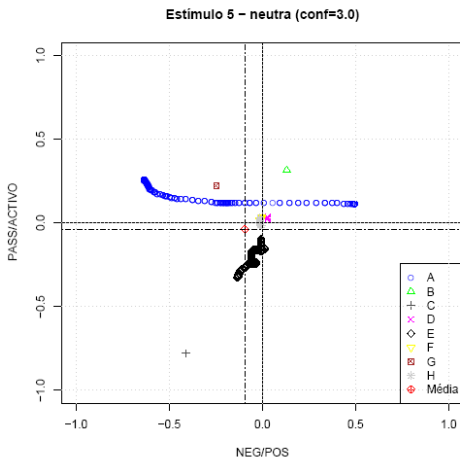
Zinken, J., Knoll, M.A., Panksepp, J. Universality and diversity in the vocalisation of emotion. In: Isdebski, K., (Ed.), Emotions of the human voice Plural Publishing, San Diego, in press.

Zovato, E., Pacchiotti, A., Quazza, S., Sandri, S., Towards Emotional Speech Synthesis: A Rule based approach, 5th ISCA Speech Synthesis Workshop, Pittsburgh, 2004.

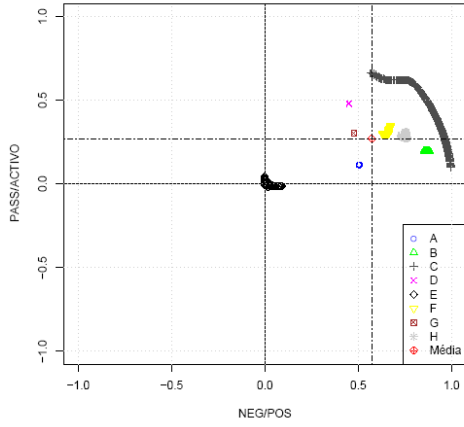
Anexo A

Representação Gráfica das marcações efectuadas em *Feeltrace*

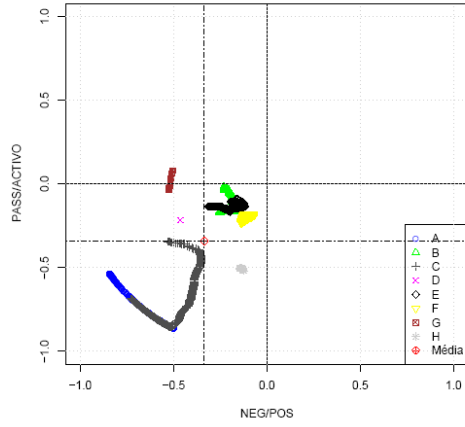




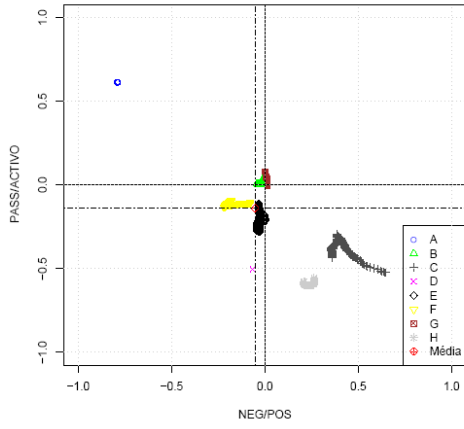
Estímulo 11 – alegre/bem-disposto/interessado (conf=3.8)



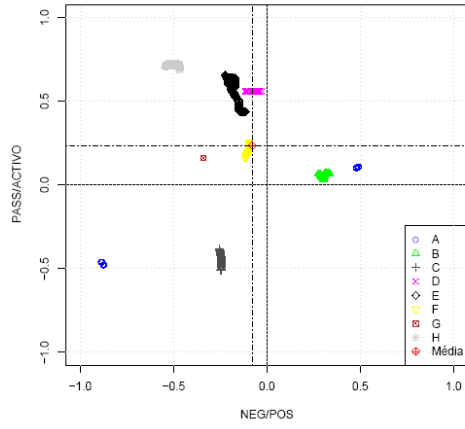
Estímulo 12 – triste (conf=3.4)



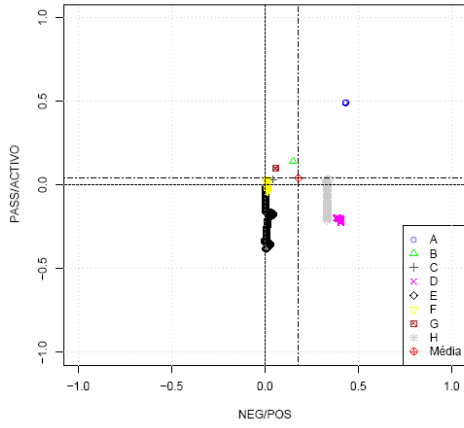
Estímulo 13 – neutra (conf=2.3)



Estímulo 14 – chateado (conf=3.3)



Estímulo 15 – neutra (conf=3.5)



Estímulo 16 – alegre (conf=4.4)

