

UNIVERSIDADE DE SÃO PAULO

Faculdade de Filosofia, Letras e Ciências Humanas

Programa de Pós-Graduação em Filologia e

Língua Portuguesa

DANIEL OLIVEIRA PERES

**A PERCEPÇÃO DA EMOÇÃO NA FALA POR
NATIVOS E NÃO NATIVOS**

(versão corrigida)

São Paulo

2016

UNIVERSIDADE DE SÃO PAULO

Faculdade de Filosofia, Letras e Ciências Humanas

Programa de Pós-Graduação em Filologia e

Língua Portuguesa

DANIEL OLIVEIRA PERES

**A PERCEPÇÃO DA EMOÇÃO NA FALA POR
NATIVOS E NÃO NATIVOS**

(versão corrigida)

Tese apresentada ao Programa de Pós-Graduação em Filologia e Língua Portuguesa do Departamento de Letras Clássicas e Vernáculas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de Doutor em Letras.

Orientador: Prof. Dr. Waldemar Ferreira Netto

São Paulo

2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

P426p Peres, Daniel Oliveira
A Percepção da Emoção na Fala por Nativos e Não
Nativos / Daniel Oliveira Peres ; orientador
Waldemar Ferreira Netto. - São Paulo, 2016.
250 f.

Tese (Doutorado)- Faculdade de Filosofia, Letras
e Ciências Humanas da Universidade de São Paulo.
Departamento de Letras Clássicas e Vernáculas. Área
de concentração: Filologia e Língua Portuguesa.

1. Percepção da Fala. 2. Fonética Acústica. 3.
Entoação. 4. Emoção . 5. Português Brasileiro. I.
Ferreira Netto, Waldemar , orient. II. Título.

*“The advantage of the emotions is that they lead us astray,
and the advantage of science is that it is not emotional.”*

Oscar Wilde

A Waldir Tavares Peres, meu pai.

In memoriam

Agradecimentos

Agradeço a Deus pela força e disposição que me fizeram seguir em frente, mesmo quando os problemas me fizeram fraquejar.

Agradeço à minha família, à minha mãe Iara (sempre paciente), ao meu irmão Eduardo e ao meu pai Waldir pelo apoio de sempre.

Agradeço à Fátima Baia, companheira que me apoia e me ensina desde os idos de 2006. Obrigado pela sua presença de espírito e pelo carinho nessa caminhada.

Ao meu orientador, Prof. Dr. Waldemar Ferreira Netto pela orientação e pela amizade nesses quase dez anos de convivência.

Ao Prof. Dr. Plínio Almeida Barbosa pela leitura minuciosa do relatório de qualificação desta tese, críticas e sugestões.

Ao Prof. Dr. Paulo Chagas pelas sugestões e críticas importantes feitas na qualificação deste trabalho.

Ao Prof. Dominic Watt pelas reuniões e sugestões, além da acolhida durante o meu estágio na University of York, Inglaterra.

Aos professores Paul Kerswill e Carmen Llamas pela discussão sobre minha pesquisa.

A Andrew MacFarlane pela conversa valiosa durante a feitura do terceiro experimento desta tese e pelas sugestões de leitura.

Aos grupos de pesquisa Language Variation and Change e Forensic Speech Science pelas questões levantadas durante as apresentações dos resultados parciais deste estudo, além das sugestões para a sua continuidade.

A todos os membros do grupo de pesquisa ExProsodia (USP) pela parceria.

A todos os participantes que dedicaram seu tempo para participarem dos experimentos de percepção.

Agradeço à Capes pelo apoio a esta pesquisa, tanto no Brasil como na Inglaterra. Sem esse amparo, esta pesquisa não seria possível.

Aprovado em: 24/10/2016

Banca examinadora:

Prof. Dr. Plínio Almeida Barbosa

Instituição: Universidade Estadual de Campinas – UNICAMP

Prof. Dr. Pablo Arantes

Instituição: Universidade Federal de São Carlos – UFSCAR

Prof.^a Dra. Sandra Madureira

Instituição: Pontifícia Universidade Católica de São Paulo – PUC-SP

Prof. Dr. Paulo Chagas de Souza

Instituição: Universidade de São Paulo – USP

Resumo

Esta pesquisa de doutorado tem como intuito investigar a percepção da emoção na fala por nativos e não nativos (falantes nativos do português brasileiro – PB, e falantes de língua inglesa sem conhecimento do PB). Esta tese orienta-se pela visão evolucionista (DARWIN, 1965 [1872]; PLUTCHIK, 1980, 1984; COSMIDES; TOOBY, 2000), que defende a universalidade das emoções; e pela visão social das emoções (AVERILL, 1980, 1993; HARRÉ, 1986; RUSSELL, 1991), que defende as emoções como produto das interações sociais.

Para a análise, foram desenvolvidos 3 experimentos de percepção de fala envolvendo 4 emoções básicas: raiva, medo, tristeza e alegria. O primeiro experimento (piloto) foi baseado na abordagem com emoções básicas e escolha forçada; o segundo na abordagem dimensional das emoções (valência, ativação e dominância); e o terceiro na metodologia *thin slicing*. Ao todo, 110 informantes participaram dos experimentos, sendo 8 no experimento piloto, 36 no experimento com abordagem dimensional (julgando fala normal e fala delexicalizada) e 76 no experimento com *thin slices*. Nos dois primeiros experimentos, foram utilizados 32 trechos de fala emotiva espontânea do português brasileiro (PB). No experimento com *thin slices*, foram selecionados 48 trechos curtos de fala emotiva (PB) com duração de até 1400ms. Os julgamentos dos participantes foram cotejados com os parâmetros acústicos provenientes da análise automática (ExProsodia) e com parâmetros acústicos relacionados à qualidade de voz.

Os resultados dos dois primeiros experimentos foram significativos e demonstraram que, de maneira geral, os participantes nativos e não nativos foram capazes de julgar com êxito as emoções. Entretanto, os participantes não nativos não apresentaram resultado significativo no experimento com abordagem dimensional e fala delexicalizada.

De acordo com a análise dos dados do experimento com *thin slices*, ao contrário do que foi encontrado no experimento com abordagem dimensional (fala delexicalizada), não houve diferença significativa entre o desempenho dos nativos e dos não nativos.

Embora os achados deste trabalho corroborem uma visão universalista das emoções, eles também dão pistas de que há uma vantagem intragrupo, ou seja, de que os nativos possuem uma habilidade maior em reconhecer as emoções do que os não nativos.

Com base nos resultados dos experimentos com limitação de informação dos estímulos (fala delexicalizada e *thin slices*), a hipótese é a de que a percepção da emoção está sujeita à *redundância de informação* contida na fala. Dessa forma, a percepção da emoção na fala é possível mesmo com a escassez de informação do sinal acústico.

Palavras-chave: percepção da fala; fonética acústica; entoação; emoção, português do Brasil

Abstract

This study aims to investigate the perception of emotional speech by natives and non-natives. This study is based on the evolutionary view (DARWIN, 1965 [1872]; PLUTCHIK 1980, 1984; COSMIDES; TOOBY, 2000), which claims that the emotions are universally manifested; and the social vision of emotion (AVERILL 1980, 1993; HARRÉ, 1986; RUSSELL, 1991), which claims that emotions are a product of social interactions.

In order to do so, three perception experiments were developed involving four basic emotions: anger, fear, sadness and joy. The first experiment – pilot – was based on the basic emotion approach (four-alternative forced choice); the second was based on the dimensional approach to emotions (valence, activation and dominance); and the third was based on the *thin slicing* methodology. Altogether, 110 participants performed the experiments, 8 in the pilot experiment, 36 in the experiment with dimensional approach (judging normal and delexicalized speech) and 76 in the experiment with *thin slices*. In the first two experiments, 32 Brazilian Portuguese (BP) excerpts of emotional speech were selected. In the experiment with *thin slices* 48 short excerpts of PB emotional speech were selected with duration up to 1400ms. The judgements of the participants were compared to the acoustic parameters from the automatic analysis (*ExProsodia*) and acoustic parameters related to voice quality.

The results of the first two experiments were significant and showed that, in general, native and non-native participants were able to judge successfully emotions. However, non-native participants showed no significant result in the experiment with dimensional approach and delexicalized speech.

According to the data analysis of the *thin slices* experiment showed that there was no significant difference between the natives and non-natives performances, unlike what was found in the experiment with dimensional approach (delexicalized speech).

Although the results of this study support a universalist perspective of emotions, they also give clues that there is an in-group advantage, namely that the natives have a greater ability to recognize emotions than non-natives.

The assumption is that the perception of emotion is subject to the *redundancy of information* contained in the speech, based on the results of the experiments with restrict content stimuli (delexicalized speech and *thin slices*). Thus, the perception of emotion in speech is possible even with the lack of information of the acoustic signal.

Keywords: speech perception; acoustic phonetics; intonation; emotion; Brazilian Portuguese

ÍNDICE

1. Introdução, 29
2. Estudos sobre a emoção – *Status Quaestionis*, 37
 - 2.1 *Estudos sobre a manifestação das emoções: evolucionismo*, 37
 - 2.2 *A perspectiva social das emoções*, 41
 - 2.3 *A perspectiva cognitivista*, 46
 - 2.4 *Integrando as perspectivas teóricas*, 52
 - 2.5 *Modelos de análise*, 56
 - 2.5.1 *Emoções básicas*, 56
 - 2.5.2 *Análise dimensional das emoções*, 64
 - 2.6 *Por que fala espontânea?*, 70
3. Qualidade de voz, 76
 - 3.1 *H1-H2, Ênfase Espectral e LTAS*, 81
 - 3.2 *Qualidade de voz e emoção*, 86
4. Entoação, 92
 - 4.1 *ExProsodia*, 99
 - 4.2 *Parâmetros acústicos via análise automática*, 103
5. Teste piloto com nativos e não nativos (emoções básicas) – qualidade de voz e entoação, 105
 - 5.1 *Metodologia*, 105
 - 5.2 *Análise e resultados parciais*, 107
6. Teste perceptivo com nativos e não nativos – qualidade de voz e análise dimensional das emoções, 109
 - 6.1 *Metodologia*, 109
 - 6.2 *Análise e resultados parciais*, 112
7. Teste perceptivo com nativos e não nativos – parâmetros acústicos (ExProsodia) e análise dimensional das emoções, 116
 - 7.1 *Metodologia*, 116

7.2	<i>Análise e resultados</i>	117
7.3	<i>Análise das dimensões e emoções básicas</i>	164
8.	Discussão preliminar	170
9.	Experimento com estímulos de curta duração - <i>thin-slicing</i>	180
9.1	<i>Justificativa e background</i>	180
9.1.1	Thin-Slicing	185
9.1.2	<i>Como avaliar os estímulos?</i>	187
9.2	<i>Metodologia</i>	191
9.3	<i>Análise dos dados</i>	194
9.3.1	<i>Resultados</i>	194
9.3.2	<i>Análise com d-prime</i>	209
9.3.3	<i>Resultados</i>	211
10.	Discussão e considerações finais	219
11.	Referências bibliográficas	227
12.	Anexos	247

Lista de Figuras

Figura 1	Esquema que ilustra a maneira como as perspectivas de análise das emoções irão interagir neste trabalho 55
Figura 2	Modelo tridimensional de Wundt 65
Figura 3	Espectro com indicação de H1-H2 – <i>modal</i> . As medidas foram extraídas de vogal tônica [a] produzida por falante masculino 82
Figura 4	Espectro com indicação de H1-H2 – <i>creaky</i> . As medidas foram extraídas de vogal tônica [a] produzida por falante masculino 82
Figura 5	Espectro com indicação de H1-H2 – <i>breathy</i> . As medidas foram extraídas de vogal tônica [a] produzida por falante masculino 83
Figura 6	Medidas de LTAS em fala emotiva: (A) – raiva [-12dB/Hz]; (B) – medo [-10dB/Hz]; (C) – tristeza [-22dB/Hz] e (D) – alegria [-9dB/Hz] 85
Figura 7	Representação da relação entre sustentação (S) e finalização (F) na formação do ritmo tonal 100
Figura 8	Esquema das componentes formadoras da entoação 101
Figura 9	Exemplo de avaliação de estímulo 111
Figura 10	Exemplos de expressões faciais médias das emoções (AKDFE) (LUNDQVIST; LITTON, 1998). No canto superior direito temos a expressão média de 35 atrizes para a emoção raiva. No canto superior direito temos a expressão média de 35 atores para a emoção medo. Abaixo delas, da esquerda para a direita, temos tristeza e alegria 190
Figura 11	Exemplo de estímulo (experimento – <i>thin-slicing</i>) 194

Figura 12: Ilustração do desempenho dos participantes dos experimentos de acordo com o acesso à informação fonético-linguística. O eixo y representa o desempenho dos participantes; quanto mais alta a posição, melhor a performance. O eixo x representa a disponibilidade da informação fonético-linguística..... 222

Lista de Gráficos

Gráfico 1	(Bra – fala normal) Julgamento de ativação e coeficiente de variação (cvTM). Valor de R ² : 0,82 119
Gráfico 2	(Bra – fala normal) Julgamento de ativação e duração das unidades entoacionais (durUE). Valor de R ² : 0,84 120
Gráfico 3	(Bra – fala normal) Julgamento de ativação e tom médio (TM). Valor de R ² : 0,61 120
Gráfico 4	(Ing – fala normal) Julgamento de ativação e coeficiente de variação de TM (cvTM). Valor de R ² : 0,73 121
Gráfico 5	(Ing – fala normal) Julgamento de ativação e duração das unidades entoacionais (durUE). R ² : 0,81 121
Gráfico 6	(Ing – fala normal) Julgamento de ativação e duração entre as unidades entoacionais (durintUE). R ² : 0,64 122
Gráfico 7	(Ing – fala normal) Julgamento de dominância e coeficiente de variação de TM (cvTM). R ² : 0,57 122
Gráfico 8	(Ing – fala normal) Julgamento de dominância e duração das unidades entoacionais (durUE). R ² : 0,62 122
Gráfico 9	(Ing – fala normal) Julgamento de dominância e duração entre as unidades entoacionais (durintUE). R ² : 0,69 123
Gráfico 10	(Bra – fala delexicalizada) Julgamento de dominância e coeficiente de variação de TM (cvTM). R ² : 0,54 125
Gráfico 11	(Bra – fala delexicalizada) Julgamento de dominância e duração das unidades entoacionais (durUE). R ² : 0,53 125
Gráfico 12	(Bra – fala delexicalizada) Julgamento de dominância e a duração dos intervalos entre as unidades entoacionais (durintUE). R ² : 0,66 126

- Gráfico 13** Modelo de visualização dos gráficos de dispersão 3D com a primeira perspectiva (ângulo 1). As letras A, B e C indicam cada um dos eixos que formam o gráfico. Na primeira perspectiva, o eixo A está em primeiro plano. As setas pretas pontilhadas indicam o movimento que irá gerar o gráfico com a segunda perspectiva (ângulo). O eixo B, destacado pelo círculo e as setas vermelhas, será o lado que estará em primeiro plano na segunda perspectiva (ângulo) ilustrada no Gráfico 14 129
- Gráfico 14** Modelo de visualização dos gráficos de dispersão 3D com a segunda perspectiva (ângulo 2). As letras A, B e C indicam cada um dos eixos que formam o gráfico. Neste ângulo, o eixo B está em primeiro plano por conta do giro à direita feito no Gráfico 13 129
- Gráfico 15** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,86 130
- Gráfico 16** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R^2 : 0.53 .. 131
- Gráfico 17** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração das unidades entoacionais – durUE). Valor de R^2 : 0.92 132
- Gráfico 18** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R^2 : 0,85 133
- Gráfico 19** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio

- padrão da duração das unidades entoacionais – dpdurUE). Valor de R^2 : 0,81 134
- Gráfico 20** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração dos intervalos entre as unidades entoacionais - dpdurintUE). Valor de R^2 : 0,82 135
- Gráfico 21** Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,66 137
- Gráfico 22** Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + desvio padrão da duração das unidades entoacionais – dpdurUE). Valor de R^2 : 0,61 138
- Gráfico 23** Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R^2 : 0,73 139
- Gráfico 24** Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + desvio padrão da duração dos intervalos entre as unidades entoacionais - dpdurintUE). Valor de R^2 : 0,66 140
- Gráfico 25** Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R^2 : 0,54 .. 141
- Gráfico 26** Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio

- TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,85 144
- Gráfico 27** Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R^2 : 0,69 .. 145
- Gráfico 28** Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,86 146
- Gráfico 29** Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração das unidades entoacionais – dpdurUE). Valor de R^2 : 0,77 147
- Gráfico 30** Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos das unidades entoacionais – durintUE). Valor de R^2 : 0,87 148
- Gráfico 31** Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração dos intervalos das unidades entoacionais – durintUE). Valor de R^2 : 0,78 .. 149
- Gráfico 32** Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,71 151
- Gráfico 33** Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração do intervalo entre as unidades entoacionais – durintUE). Valor de R^2 : 0,75 152

- Gráfico 34** Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração das unidades entoacionais – durUE). Valor de R²: 0,68 153
- Gráfico 35** Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração das unidades entoacionais – dpdurUE). Valor de R²: 0,61 .. 154
- Gráfico 36** Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R²: 0,79 .. 155
- Gráfico 37** Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração dos intervalos entre as unidades entoacionais – dpdurintUE). Valor de R²: 0,66 156
- Gráfico 38** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R²: 0,66 159
- Gráfico 39** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R²: 0,71 .. 160
- Gráfico 40** Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R²: 0,61 162

Gráfico 41	Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R ² : 0,59 ..	163
Gráfico 42	Interação entre as dimensões no experimento com fala normal – nativos	164
Gráfico 43	Interação entre as dimensões no experimento com fala normal – não nativos	164
Gráfico 44	Interação entre as dimensões no experimento com fala delexicalizada – nativos	165
Gráfico 45	Interação entre as dimensões no experimento com fala delexicalizada – não nativos	165
Gráfico 46	Dispersão das emoções pela análise dimensional (Bra – fala normal)	166
Gráfico 47	Dispersão das emoções pela análise dimensional (Ing – fala normal)	167
Gráfico 48	Dispersão das emoções pela análise dimensional (Bra – fala delexicalizada)	167
Gráfico 49	Dispersão das emoções pela análise dimensional (Ing – fala delexicalizada)	168
Gráfico 50	Curvas de distribuição de frequência para cada uma das emoções julgadas pelos participantes nativos. As curvas superiores, da esquerda para direita, representam raiva e alegria; as inferiores, a partir da esquerda, representam tristeza e medo. As curvas foram feitas com dados padronizados por Z-score	203
Gráfico 51	Curvas de distribuição de frequência para cada uma das emoções julgadas pelos participantes não nativos. As curvas superiores, da esquerda para direita, representam raiva e alegria; as inferiores, a partir da esquerda, representam tristeza e medo. As curvas foram feitas com dados padronizados por Z-score	204
Gráfico 52	Curvas de distribuição das proporções de acertos de nativos e não nativos	205

- Gráfico 53** Curvas de distribuição de acertos de nativos e não nativos para cada uma das emoções (alegria, medo, raiva e tristeza) 206
- Gráfico 54** Curvas de distribuição das proporções de acertos de nativos e não nativos de acordo com o sexo. A curva de cor salmão ilustra as respostas dadas por participantes de mesmo sexo, ou seja, quando homens ouvirem estímulos de voz masculina, e mulheres, de voz feminina. A curva verde ilustra as respostas dadas por homens e mulheres a estímulos diferentes, ou seja, homens ouvindo voz feminina e mulheres ouvindo voz masculina 207
- Gráfico 55** Conjunto de gráficos de barras com as tendências de erros para o grupo de participantes nativos. No canto superior esquerdo está a distribuição dos erros para a emoção raiva, no canto superior direito estão os resultados para a emoção medo. Abaixo deles, da esquerda para a direita, estão os resultados para a tristeza e a alegria 216
- Gráfico 56** Conjunto de gráficos de barras com as tendências de erros para o grupo de participantes não nativos. No canto superior esquerdo está a distribuição dos erros para a emoção raiva, no canto superior direito estão os resultados para a emoção medo. Abaixo deles, da esquerda para a direita, estão os resultados para a tristeza e a alegria 217

Lista de Quadros

Quadro 1	Número de termos relacionados às emoções, sentimentos e atitudes em diferentes línguas (adaptado de Russell (1991)) 42
Quadro 2	Funções adaptativas das emoções 62
Quadro 3	Emoções primárias e suas combinações (díades), segundo a proposta de Plutchik (1980) 63
Quadro 4	Esquematisação dos julgamentos das dimensões – fala normal para nativos e não nativos (“+” = valores altos; “-” = valores baixos; “±” valores medianos; “♣” =aleatório) 177
Quadro 5	Esquematisação dos julgamentos das dimensões – fala delexicalizada para nativos e não nativos (“+” = valores altos; “-” = valores baixos; “±” valores medianos; “♣” = aleatório) 177
Quadro 6	Resumo dos resultados da análise dimensional 178
Quadro 7	Lista de <i>thin slices</i> utilizadas no experimento com duração e transcrição. Na coluna da esquerda estão os nomes dos arquivos. O número é a identificação do arquivo, a primeira letra (M ou F) é referente ao sexo do falante, a última letra diz respeito à inicial de cada emoção (raiva, medo, tristeza e alegria) 192
Quadro 8	Julgamentos dos participantes sobre estímulos (<i>thin slices</i>) de raiva. O painel à esquerda refere-se aos participantes nativos; à direita, aos não nativos. A numeração (coluna em branco) é a identificação de cada participante (76 no total, 38 nativos e 38 não nativos). A identificação de cada estímulo foi escrita verticalmente acima de cada coluna. A soma dos julgamentos corretos e a respectiva porcentagem estão abaixo do quadro 195
Quadro 9	Julgamentos dos participantes sobre estímulos (<i>thin slices</i>) de tristeza 196
Quadro 10	Julgamentos dos participantes sobre estímulos (<i>thin slices</i>) de alegria 197

Quadro 11	Julgamentos dos participantes sobre estímulos (<i>thin slices</i>) de medo	198
------------------	---	-----

Lista de Tabelas

Tabela 1	Resultados do teste Kappa-Fleiss 112
Tabela 2	Regressões lineares entre os julgamentos dos participantes sobre fala normal e medidas acústicas de qualidade de voz. Os resultados não foram significativos 113
Tabela 3	Coefficiente de correlação de Pearson entre as dimensões – fala normal 114
Tabela 4	Coefficiente de correlação de Pearson entre as dimensões – fala delexicalizada 115
Tabela 5	Coefficiente de correlação de Pearson entre mesmas dimensões entre nativos e não nativos 115
Tabela 6	Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala normal) e os parâmetros acústicos – ExProsodia 118
Tabela 7	Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala normal) e os parâmetros acústicos – ExProsodia 118
Tabela 8	Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala delexicalizada) e os parâmetros acústicos – ExProsodia 124
Tabela 9	Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala delexicalizada) e os parâmetros acústicos – ExProsodia 124
Tabela 10	Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala normal) e os parâmetros acústicos combinados – <i>ExProsodia</i> 127

Tabela 11	Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala normal) e os parâmetros acústicos combinados – <i>ExProsodia</i> 127
Tabela 12	Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala delexicalizada) e os parâmetros acústicos combinados – <i>ExProsodia</i> 158
Tabela 13	Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala delexicalizada) e os parâmetros acústicos combinados – <i>ExProsodia</i> 158
Tabela 14	Matriz de confusão das emoções apresentadas como <i>thin slices</i> aos participantes nativos. As linhas representam os estímulos apresentados aos participantes, somando 456 para cada emoção. As colunas representam as respostas dadas a cada uma das emoções. A última coluna à direita representa o total de estímulos ouvidos pelos participantes. A última linha é formada pela soma das colunas referentes às respostas dadas pelos participantes para cada emoção. As respostas corretas foram dispostas diagonalmente nas células preenchidas em cinza. Os erros estão dispostos acima ou abaixo da linha diagonal 201
Tabela 15	Matriz de confusão das emoções apresentadas como <i>thin slices</i> aos participantes não nativos 201
Tabela 16	Estatística descritiva – nativos 202
Tabela 17	Estatística descritiva – não nativos 202
Tabela 18	Resultados dos testes dos pressupostos de normalidade (Shapiro-Wilk) e homogeneidade da variância (Fligner-Killeen). Os testes não foram significativos nos modelos em que as células estão destacadas. Nesses casos, os pressupostos não foram violados 208
Tabela 19	Resultados das análises de variância (ANOVA) e dos testes de Kruskal-Wallis. O modelo 1

	(acertos e emoção) obteve resultado significativo	209
Tabela 20	Valores de d' para nativos e não nativos para cada participante de acordo com cada emoção (raiva, medo, tristeza e alegria). Os índices d' são preenchidos com diferentes tons de acordo com o valor. O preenchimento torna-se mais escuro à medida que os valores d' aumentam	212
Tabela 21	Estatística descritiva de d' – nativos	213
Tabela 22	Estatística descritiva de d' – não nativos	213

Parâmetros relacionados à frequência fundamental – F0

- ❖ *Tom médio (TM)* – medida de tendência central dos valores válidos de F0 (porções de frequência fundamental selecionadas a partir da combinação dos parâmetros de análise estipulados pelo pesquisador) calculada como a média aritmética acumulada no tempo.
- ❖ *Desvio padrão do tom médio (dpTM).*
- ❖ *Coefficiente de variação de (cvTM).*
- ❖ *Assimetria do TM (α TM).*
- ❖ *Diferença entre o tom médio e a menor unidade entoacional (TMmUE)* – A unidade entoacional, como dito acima, é uma porção sonora selecionada pelo *software* a partir de parâmetros predeterminados.
- ❖ *Menor valor de uma unidade entoacional (mUE)* – É o menor valor (Hz) dentre as unidades entoacionais selecionadas pelo *software*.
- ❖ *Coefficiente de variação dos valores positivos de Foco/Ênfase (cvposF/E)* – Os valores positivos da componente Foco/Ênfase são aqueles que ocorreram acima do limite superior do tom médio.
- ❖ *Coefficiente de variação dos valores negativos de Foco/Ênfase (cvnegF/E)* – Os valores negativos da componente Foco/Ênfase são aqueles que ocorreram abaixo do limite inferior do tom médio.

Parâmetros relacionados à duração

- ❖ *Duração das unidades entoacionais ($durUE$)* – Duração em milissegundos das unidades selecionadas pelo *software* a partir da combinação dos valores predeterminados pelo pesquisador.
- ❖ *Desvio padrão da duração das unidades entoacionais ($dpdurUE$)*.
- ❖ *Duração dos intervalos entre as unidades entoacionais ($durintUE$)* – Duração em milissegundos entre as unidades selecionadas pelo *software*.
- ❖ *Desvio padrão da duração dos intervalos entre as unidades entoacionais ($dpdurintUE$)*.

1. INTRODUÇÃO

A manifestação das emoções nos humanos tem sido objeto de estudo com vasta variedade de interesses e com mais de 2000 anos de estudos no ocidente. Por essa razão, esta introdução pretende traçar um breve percurso dos estudos sobre a emoção, desde as abordagens filosóficas até o início das abordagens psicológicas, evolutivas e sociais. Esse panorama não tem intenção de ser uma abordagem exaustiva do ponto de visto histórico, tampouco um relato inequívoco dos passos que levaram os estudos da emoção ao atual estado da arte. O intento desta introdução é mais modesto e tem como finalidade dar uma breve ideia de como os estudos passaram das preocupações filosóficas às abordagens científicas atuais.

Dando início em Platão (428-347 a.C.), no livro IV de *A República* (2001), as emoções estão relacionadas ao espírito e ao apetite, considerando sua visão tripartite da alma. A alma, segundo a abordagem platônica (PLATÃO, 2001), é formada por três partes: a razão, o espírito e o apetite.

Na primeira está o prazer pela contemplação do ser, o amor à verdade, o raciocínio e a atividade intelectual. Na segunda está a conduta irascível, os impulsos, a ira. Na terceira, onde há o prazer dos desejos fisiológicos e paixões. Com essa divisão, Platão atribui às emoções uma estreita relação com as características não racionais e impulsivas da alma.

Aristóteles (384-322 a.c.), em *Retórica* (2005), aborda a emoção como parte do juízo que se quer formar durante o discurso. Assim, a emoção tem grande importância por despertar na plateia, a partir da forma como o discurso é apresentado, determinado estado de espírito com relação ao orador. Ao lado dos fins da retórica, Aristóteles define emoção como “as causas que fazem alterar os seres humanos e introduzem mudanças nos seus juízos, na medida em que comportam dor e prazer: tais são a ira, a compaixão, o medo e suas semelhantes, assim como as contrárias” (ARISTÓTELES, 2005, p. 160).

Para o filósofo, a emoção está ligada a estímulos provenientes do ambiente e da relação social. Por exemplo, a ira é um desejo de vingança acompanhado de dor, mormente provocado por desprezo ou ofensa. Assim, os seres humanos “encolerizam-se se alguém se opuser à sua ação ou se alguém não colaborar com ele, ou se, de alguma forma, alguém o perturbar quando está em tal estado” (ARISTÓTELES, 2005, p. 163). Como contraponto, a calma é o estado contrário da cólera, na medida em que não há confronto com as vontades alheias, tampouco desprezo pelo interlocutor.

Da mesma forma, o medo é causado por situações de perigo ou por um mal, mas o fator determinante para o desencadeamento do medo é a iminência do perigo ou do mal, em outras palavras, a proximidade do que é mal é crucial para a manifestação do medo. A abordagem aristotélica da emoção possui alguma semelhança com relação aos estudos sobre emoção a partir do século XIX. Para Darwin (1965) e demais abordagens evolucionistas advindas de seus estudos, o ambiente é um potencial

precursor da manifestação das emoções. A abordagem darwiniana das emoções será tratada em detalhes adiante.

Ainda sobre as semelhanças entre a abordagem aristotélica e as abordagens modernas, Solomon (2008) afirma que Aristóteles antecipou as teorias modernas da emoção por levar em conta componentes cognitivos, reação física e contextos sociais.

Avançando um pouco mais no tempo, diferentemente da visão aristotélica das emoções, os estoicos (III a.C.) viam na manifestação delas, em especial a raiva, um vício. Sêneca (1900) discorre sobre a manifestação da raiva, mas, diferentemente de Aristóteles, atenta para seu caráter nocivo:

[...] they whom anger possesses are not sane, look at their appearance; for as there are distinct symptoms which mark madmen, such as a bold and menacing air, a gloomy brow, a stern face, a hurried walk, restless hands, changed colour, quick and strongly-drawn breathing [...] their eyes blaze and sparkle, their whole face is a deep red with the blood which boils up from the bottom of their heart, their lips quiver, their teeth are set [...] their whole body is highly-strung and plays those tricks which mark a distraught mind, so as to furnish an ugly and shocking picture of self-perversion and excitement. (SÊNECA, 1900, p. 48)

Sêneca descreve acima algumas alterações físicas que ocorrem em pessoas coléricas, possibilitando novamente uma comparação ao que mais tarde foi descrito exaustiva e detalhadamente por Darwin (1965) do ponto de vista fisiológico e anatômico.

Durante a Idade Média, o estudo da emoção estava ligado à ética e ao Cristianismo, especificamente a estudos sobre aspectos cognitivos e de volição (SOLOMON, 2008, p. 6). Tomás de Aquino (1225-1275) estudou as emoções – *passiones animae* – sob a perspectiva somática, cognitivista e taxonômica.

Segundo Aquino, as *passiones animae* são 11, divididas em dois grupos – concupiscíveis e irascíveis. Em cada grupo, as emoções são

organizadas em pares. As emoções concupiscíveis são: amor e ódio; desejo e aversão; prazer e aflição. As irascíveis são: esperança e desespero; confiança e medo; e raiva. Para esta última não há emoção contrária, diferentemente do que Aristóteles pensava sobre a calma e a ira.¹

Na Idade Moderna, Descartes (1596-1650) investigou as paixões de maneira diferente das visões aristotélica (retórica) e tomista (escolástica). Um trecho de uma carta enviada a um amigo não deixa dúvidas sobre sua intenção “my intention was to explain the passions only as a natural philosopher, and not as a rhetorician or even as a moral philosopher” (DESCARTES, 1985, p. 327). É evidente o rompimento que Descartes pretendeu estabelecer entre o que tinha sido feito nos estudos das emoções até então e a sua abordagem.

Descartes (1985) afirma que, para se entender as paixões da alma, deve-se separar suas funções das do corpo. Dessa forma, o dualismo corpóreo é essencial para entender assim o funcionamento das paixões nos humanos. Descartes propôs uma taxionomia para as paixões. Dentre as várias enumeradas, há seis que o filósofo define como primitivas: admiração, amor, ódio, alegria, tristeza e desejo. Além da definição das paixões, Descartes observa a utilidade e os danos que elas podem causar, como, por exemplo, alterações na circulação do sangue. Como dito acima, esse tipo de abordagem que avalia as modificações no corpo será largamente descrita por Darwin dois séculos mais tarde.²

A partir do século XIX, os estudos sobre as emoções começaram a ter maior riqueza de detalhes fisiológicos e psicológicos. Em 1872, o trabalho pioneiro de Darwin, *The expression of the emotions in man and animals* (1965), é exemplo desse tipo de refinamento descritivo e teórico, além de fomentar, ainda nos dias de hoje, a discussão sobre a manifestação das emoções em outras perspectivas teóricas. Dentre as áreas de interesse,

¹ Para uma abordagem mais detalhada da filosofia tomista das emoções, cf. King (2012).

² Para uma revisão mais extensa sobre os estudos filosóficos da emoção, cf. Solomon (2008).

estão as teorias evolutivas, psicologia, neurologia, estudos relacionados à produção e percepção de fala. Essa última vertente dos estudos sobre a manifestação das emoções se faz mais presente no século XX, no qual teorias evolutivas, cognitivistas e socioculturais tentam explicar pelos seus meios a relação entre as emoções, cognição, língua e contextos sociais.

Dentre os diversos questionamentos que envolvem os estudos de fala expressiva ou emotiva, o papel que a língua e a cultura desempenham na categorização e na percepção das emoções é um dos mais relevantes. Da mesma forma, a busca de características universais na manifestação e percepção das emoções é um fator importante para os estudos na área da ciência afetiva.

Embora a maioria das referências ao trabalho de Darwin (1965) seja relacionada à descrição da manifestação das emoções do ponto de vista fisiológico e de sua função para a sobrevivência das espécies, a variação dessas manifestações entre povos distintos também está presente na obra de Darwin. Uma amostra disso são suas *Queries about expression*,³ datadas de 1867, um questionário enviado por Darwin a pessoas que viviam em contato com diferentes povos nativos ao redor do mundo. As respostas apresentadas serviram de substrato para a sua obra *The expression of the emotions in man and animals* (1965). Como conclusão, Darwin argumenta que as emoções e o seu reconhecimento fazem parte de uma herança biológica e que, portanto, são universalmente manifestas.

Para explicar como as emoções foram herdadas, ele propõe três princípios: *princípio dos hábitos associados úteis*, *princípio da antítese* e *princípio das ações diretas do sistema nervoso*. Esses princípios⁴ são válidos para os animais, dentre eles os humanos, e independentes de regiões, línguas e sociedades distintas. Embora Darwin não ignore as diferentes

³ A relação total das questões está disponível *online* em: http://darwin-online.org.uk/EditorialIntroductions/vanWyhe_queries.html.

⁴ Os princípios propostos por Darwin serão expostos no próximo capítulo.

sociedades, etnias e línguas, o seu foco está no que pode ser compartilhado pelos humanos e pelos animais, sem atentar para detalhes regionais.

Por outro lado, a abordagem social das emoções tem como objeto de pesquisa as variações sociais da manifestação das emoções. O interesse dos pesquisadores sobre o papel da sociedade, da língua e da cultura na categorização das emoções teve início em trabalhos de antropologia (DODDS, 1951; CARSTAIRS, 1967; HARRÉ; FINLAY-JONES, 1986). As motivações são a discrepância entre as palavras que designam as emoções em culturas diferentes e a influência do ambiente social para a manifestação das emoções. Na visão dos autores, a grande variação no número de palavras que designam estados emotivos é um argumento forte a favor da influência da sociedade e da língua na expressão das emoções. A diferença no número de palavras emotivas é expressiva, variando de sete palavras em uma língua (chewong)⁵ a 2000 palavras em outra (inglês). O quadro com os estudos e seus problemas serão expostos adiante. Sustento, desde já, que esses estudos, embora tenha sido importante num primeiro momento, não são criteriosos do ponto de vista metodológico e, por conseguinte, implicam resultados não confiáveis.

Russell (1991), por meio de revisão etnográfica e de estudos comparativos entre culturas, defende que as características sociais exercem grande influência na categorização e expressão das emoções. O autor afirma que a divisão do mundo em categorias é uma característica inerentemente humana e que é estendida à categorização que cada língua faz das emoções. O que está em pauta, nesse tipo de abordagem, não é qual emoção está representada em determinada língua, mas como os falantes de uma mesma língua a percebem.

Uma terceira via às teorias biológicas/evolutivas e sociais seria uma que procurasse tratar dos dois campos de interesse em questão e buscasse saber quais aspectos universais e socioculturais podem ser importantes

⁵ Língua da família austro-asiática falada na Malásia e que possui cerca de 460 falantes.

para o estudo das emoções. Para tal, a abordagem experimental da fala, envolvendo suas características acústicas e perceptivas, é uma alternativa possível para que se possa determinar em que medida “o que é dito” e o “como é dito” interagem para serem reconhecidos como fala emotiva. Sendo assim, este estudo será norteado pelos pressupostos *evolucionistas* (DARWIN, 1965; PLUTCHIK, 1980, 1984, 2001; GRIFFITHS, 1997; COSMIDES; TOOBY, 2000) e *sociais* da expressão das emoções (AVERILL, 1980; HARRÉ, 1986; RUSSELL, 1991). Por se tratar de uma análise com foco na percepção da fala, também será feita uma revisão da abordagem *cognitivista* (ARNOLD, 1960; SCHERER, 1999). Dessa maneira, pretende-se, pois, investigar de maneira quantitativa qual o peso que o conhecimento linguístico e cultural pode ter na percepção das emoções. Com esse intuito, a percepção da fala emotiva por nativos e não nativos⁶ será cotejada com os parâmetros acústicos provenientes da produção da fala, dentre eles aqueles relacionados à qualidade de voz e às características entoacionais que possam explicar os julgamentos dados pelos participantes. Os experimentos foram desenhados a partir de trechos de fala espontânea emotiva do português brasileiro⁷ e foram apresentados a falantes nativos do português e do inglês, por meio de três experimentos de percepção.

Do ponto de vista experimental e metodológico, as emoções envolvidas neste estudo (raiva, medo, tristeza e alegria) serão tratadas de três maneiras: abordagem com emoções discretas (piloto), abordagem dimensional das emoções e abordagem com estímulos curtos (*thin-slicing*).

Este trabalho está dividido em duas partes principais: a primeira apresenta os fundamentos pelos quais esta pesquisa se orienta, a segunda

⁶ Num primeiro momento, os informantes dos experimentos foram chamados de falantes nativos e não nativos (do PB), inclusive em Peres (2013) – *The perception of emotion by native and non-native speakers*. Por conta da ambiguidade presente nessa maneira de nomear os participantes (o termo “falantes não nativos” pode levar à interpretação de que se trata de falantes de português como L2), optou-se por chamá-los apenas de nativos e não nativos, visto que os não nativos não são falantes de português e não possuem qualquer conhecimento da língua.

⁷ Doravante PB.

é composta pelos experimentos de percepção, análise de produção e análise estatística dos dados, discussão e considerações finais.

2. ESTUDOS SOBRE A EMOÇÃO – *Status Quaestionis*

2.1 Estudos sobre a manifestação das emoções: evolucionismo

O panorama dos estudos das emoções pode ser dividido em três grandes áreas de interesse, a saber: evolutiva, social e cognitivista (processos internos). Cornelius (2000), por sua vez, divide-os em quatro grandes perspectivas: darwiniana, jamesiana, cognitivista e socioconstrutivista. Distribuir as teorias em três ou quatro grandes áreas depende do critério adotado pelo pesquisador, de forma que o primeiro denota uma visão mais geral das abordagens darwiniana e jamesiana (e suas variações), visto que elas têm muito em comum e se sobrepõem em muitos de seus pressupostos. Dessa forma, neste trabalho, as visões de Darwin (1965) e James (1890) serão tomadas como sendo variações do mesmo campo teórico – o dos estudos evolucionistas. Segundo Cornelius

(2000), Darwin e James partilham a mesma visão sobre as emoções, tidas como adaptações ao ambiente com funções de preservação. Além disso, eles entendiam a emoção como uma resposta mais ou menos automática a eventos do ambiente, cuja importância maior era a sobrevivência. Por outro lado, os pesquisadores tinham diferentes interesses; Darwin era interessado na expressão das emoções, enquanto James ateve-se à experiência emocional.

Ademais, os estudos evolucionistas sobre a emoção ainda podem ser subdivididos em três tipos de acordo com as hipóteses defendidas por cada uma das teorias sobre como se deu a herança e a evolução das emoções. A primeira delas sustenta que as emoções podem ter sido selecionadas nos hominídeos. Como exemplo, Cosmides e Tooby (2000) citam o medo provocado nos hominídeos por estarem sozinhos durante a noite, correndo o risco de serem atacados por predadores. A segunda argumenta que as emoções foram selecionadas anteriormente aos hominídeos e que são encontradas em várias espécies (PLUTCHIK, 1980, 1984, 2001). A terceira defende que as emoções foram herdadas por meio de um ancestral comum (GRIFFITHS, 1997). Essa herança, na teoria evolutiva, tem como ponto fundamental a sobrevivência da espécie e é considerada universal.

A abordagem evolutiva visa sobretudo à *manifestação* das emoções e seu conseqüente resultado somático. Em outras palavras, o que é manifestado corporalmente ou, por conseqüência, vocalmente configura-se como objeto de estudo. Como exemplo desse tipo de associação, correlato corpóreo, Darwin cita um trecho das anotações⁸ da conferência proferida pelo anatomista francês, Pierre Gratiolet, o qual diz:

Il résulte, de tous les faits que j'ai rappelés, que les sens, l'imagination et la pensée elle-même, si élevée, si abstraite qu'on la suppose, ne peuvent s'exercer sans éveiller un sentiment corrélatif, et que ce sentiment se traduit

⁸ *De la physionomie et des mouvements d'expression*. Disponível em: <<http://archive.org/stream/delaphysionomie00grat#page/n5/mode/2up>>. Acesso em: 5 novembro 2012.

directement, sympathiquement, symboliquement ou métaphoriquement, dans toutes les sphères des organes extérieurs, qui le racontent tous, suivant leur mode d'action propre, comme si chacun d'eux avait été directement affecté. (GRATIOLET, 1865,⁹ apud DARWIN, 1965, p. 6)

Esse excerto é uma tentativa de justificar as alterações corporais que transparecem um determinado estado de espírito, o mesmo sendo válido quando a manifestação da emoção é notada na fala. Assim, um pressuposto importante do ponto de vista evolutivo, segundo Bachorowski e Owren (2000, p. 197), é que as expressões afetivas, dentre elas a vocal, são verídicas. Pode-se considerar como exemplo desse pressuposto as galvanizações feitas por Duchenne (2010), as quais demonstraram que alguns músculos da face só são ativados quando estão sob efeito de estados emotivos verdadeiros.¹⁰ Dessa forma, Darwin (1965) enumera uma série de características que emergem corporalmente quando homens ou animais estão emocionalmente estimulados.

As emoções, segundo Darwin, podem ser explicadas por meio de três princípios: i) *princípio dos hábitos associados úteis* – ações despertadas no organismo que tiveram alguma utilidade em estados de espírito distintos no passado e que, ao menor sinal desses estados no presente, há uma tendência em repeti-las, por associação e hábito, mesmo que não tenham função aparente; ii) *princípio da antítese* – quando um estado de espírito contrário é induzido, há tendência involuntária para a execução de movimentos opostos, mesmo que nunca tenham sido úteis, e.g., a docilidade dos cães; iii) *princípio das ações diretas do sistema nervoso* – ações

⁹ GRATIOLET, P. *De la Physiognomie et des Mouvements d'Expression*. Paris: Hetzel; 1865. Disponível em <<http://gallica.bnf.fr/ark:/12148/bpt6k24626p/f5.image>>. Acesso em 10 novembro 2012.

¹⁰ A técnica de galvanização de músculos faciais feita por Guillaume Duchenne (2010) consistia em aplicar cargas elétricas conduzidas por agulhas em determinados músculos da face. Com isso, o experimentador conseguia obter expressões diferentes como alegria, surpresa, horror, tristeza, riso, etc. Darwin mostrou uma foto a 24 pessoas com a expressão de terror provocada pela galvanização de Duchenne, dentre elas 20 disseram se tratar de “medo intenso”.

reconhecidas como indicadoras de certos estados de espírito são diretamente atribuídas à constituição do sistema nervoso, desde o início independentes do hábito e da vontade. Com esses princípios postos, Darwin explica como as emoções foram herdadas por sucessivas gerações, sem atentar com mais detalhes para o fator linguístico e cultural, por serem estágios mais tardios do ponto de vista evolutivo. Algumas menções sobre alterações da voz em estados emotivos podem ser encontradas como no trecho a seguir:

He (Spencer) clearly shows that the voice alters much under different conditions, in loudness and in quality, that is, in resonance and *timbre*, in pitch and intervals. [...] from the analogy of other animals, I have been led to infer that the progenitors of man probably uttered musical tones, before they had acquired the power of articulate speech; and that consequently, when the voice is used under any strong emotion, it tends to assume, through the principle of association. (DARWIN, 1965, p. 86)

Nota-se que Darwin (1965), baseando-se no estudo de Spencer (1891),¹¹ considera que as emoções também são expressas concomitantemente à fala/voz, não se limitando somente à sua manifestação primária encontrada em outros animais. Os termos “fala” e “voz” podem ser tratados como atividades distintas se definirmos a primeira como uma atividade predominantemente supralaríngea e a segunda como atividade laríngea. Essa divisão, no entanto, não é tão pacífica. Parte dos animais também se expressa por meio da laringe, da voz, mas não produzem fala. Por outro lado, podemos ter fala sem que tenhamos atividade laríngea, *e.g.*, *whispered voice*.

¹¹ cf. SPENCER, H. The origins and function of music. In: SPENCER, H. (ed.), *Essays: Scientific, Political, and Speculative*, v. 2. Library Edition, containing Seven Essays not before republished, and various other Additions. London: Williams and Norgate, 1891.

De maneira oposta a Darwin, para James (1884),¹² a emoção é a percepção de diferentes mudanças corporais, desencadeadas por algum agente, que denotam um estado emotivo. Assim, as emoções são resultado da percepção das reações fisiológicas e comportamentais desencadeadas no corpo por determinado evento externo ou interno.¹³ Como dito anteriormente, assim como Darwin, James leva em consideração a função preservadora da espécie, mas propõe uma ordem inversa no seu mecanismo. Segundo a teoria jamesiana, experienciamos as emoções porque nosso organismo evoluiu para responder de forma automática e adaptativa às características do ambiente importantes para a sobrevivência. Em outras palavras, o corpo responde primeiro e a experiência dessas mudanças corpóreas constituem o que chamamos de emoção (CORNELIUS, 2000, p. 4). Para James, o corpo desencadeia as emoções; para Darwin, o corpo é o lugar onde as emoções são manifestadas.

2.2 A perspectiva social das emoções

Com interesse de outra natureza, os estudos sociais ou socioconstrutivistas sobre a manifestação das emoções explicam as emoções como construções sociais, ou seja, como um produto de culturas distintas e apreendidas pelos integrantes de cada sociedade por meio da experiência.¹⁴ O socioconstrutivismo é o mais novo e diverso campo de estudo a tratar da emoção, caso não levemos em conta os estudos anteriores feitos por antropólogos e sociólogos. A primeira diferença mais saliente entre o socioconstrutivismo e o evolucionismo é que a segunda é mais ampla

¹² De maneira independente, Carl Lange (1885) também partiu do mesmo pressuposto de James (1884), ou seja, de que o corpo é desencadeador de emoções. Por essa razão, esse tipo de abordagem ficou conhecido como *James-Lange Theory of Emotion*.

¹³ Visão que, em parte, é compartilhada por Damásio (2012). A posição do autor é mais complexa, pois faz uma distinção básica entre emoção e sentimentos. As emoções surgem de sentimentos, mas nem todos os sentimentos surgem das emoções.

¹⁴ Sabendo-se que a definição do termo “cultura” pode variar consideravelmente, neste estudo, o termo cultura será entendido, segundo Tylor (1871, p. 1), como sendo um “*complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society*”.

quanto ao escopo das manifestações das emoções, incluindo nele outros seres vivos além dos humanos. Para Averill (1980, p. 306), o uso de conceitos relacionados à emoção em animais é essencialmente metafórico ou derivativo. A segunda diferença, mas não menos importante, é a posição central dada à cultura e à língua. A área de estudos sociais da emoção tornou-se mais claramente identificável a partir dos trabalhos de Averill (1980) e Harré (1986). Por ser recente e, portanto, em fase de consolidação, será necessário retornar um pouco no tempo para expor os pressupostos dessa perspectiva de análise das emoções. Dentre as justificativas, a discrepância entre os nomes dados às emoções em diferentes línguas parece indicar uma visão social e local das emoções. O número de palavras utilizadas pelas línguas para nomear emoções, sentimentos e atitudes varia consideravelmente. No quadro abaixo, há alguns estudos que foram feitos por etnógrafos nas décadas de 1970 e 1980:¹⁵

Quadro 1: Número de termos relacionados às emoções, sentimentos e atitudes em diferentes línguas (adaptado de Russell (1991))¹⁶

Língua	Palavras para emoções	Pesquisadores
<i>Inglês</i>	<i>Mais de 2000</i>	<i>Wallace e Carson (1973)</i>
<i>Holandês</i>	<i>1501</i>	<i>Hoekstra (1986)</i>
<i>Chinês (Taiwan)</i>	<i>750</i>	<i>Boucher (1979)</i>
<i>Malaio</i>	<i>230</i>	<i>Boucher (1979)</i>
<i>Ifalukian¹⁷</i>	<i>58</i>	<i>Lutz (1982)</i>
<i>Chewong</i>	<i>7</i>	<i>Howell (1981)</i>

Embora sejam dados interessantes para atestar a variação na categorização das emoções entre as línguas, gozam de pouca credibilidade

¹⁵ Apesar de os estudos etnográficos não estarem diretamente ligados à pesquisa desenvolvida neste trabalho, a quantidade de termos relacionados às emoções, atitudes e sentimentos foram importantes para a formação dos pressupostos da teoria socioconstrutivista da emoção, razão única pela qual foram citados acima. Como será visto adiante, o léxico terá papel secundário nos experimentos de percepção desenvolvidos neste trabalho.

¹⁶ Em português, numa busca rápida na internet, pode-se encontrar variadas listas com palavras relacionadas a emoções, estados e sentimentos. A variação na quantidade de palavras é grande, partindo de 40 a 403 palavras. De fato, não são estudos científicos sobre o assunto, mas já dão uma ideia razoável sobre como esse tipo de lista pode variar.

¹⁷ Língua falada em Ifaluk, arquipélago localizado nas Ilhas Carolinas, no Oceano Pacífico.

pela variação da metodologia adotada em cada uma das pesquisas. Não são, portanto, estudos paralelos e dificultam a possibilidade de meta-análise.

O conflito de significados entre a língua do entrevistado e do entrevistador pode por si só causar diferenças quanto ao número de palavras relacionadas às emoções e seus significados. Russell (1991), apesar de ser um pesquisador com foco na abordagem socioconstrutivista, levanta alguns problemas relacionados à etnografia aplicada aos estudos da emoção. Segundo o autor, diferenças são encontradas na frequência, na expressão, nas crenças, nas atitudes e na regulação da emoção. Considerando essas variáveis, o que muda com a cultura são os eventos que envolvem/provocam as emoções, sendo a emoção universal. Esse tipo de visão é chamado de *standard view*. Russell (1991) também aponta dois outros problemas inerentes a essa visão; primeiramente não é possível determinar o que seria invariante ou universal nas emoções uma vez que as causas, consequências e expressões são removidas (além do problema metodológico envolvido no isolamento dessas variáveis); em segundo lugar, o pesquisador interpreta a expressão emocional de outra pessoa desde fora, o que não implica necessariamente que essa pessoa compartilhe da mesma interpretação do pesquisador (RUSSELL, 1991, p. 428).

Outro argumento utilizado pelos pesquisadores socioconstrutivistas é que a ocorrência das manifestações emocionais se dá em ambiente social ou, ao menos, interpessoal. Assim, para Parkinson (1996, p. 680), as emoções teriam seus significados muito mais caracterizados de maneira relacional do que pessoal. Ademais, a própria expressão dos estados emotivos é mediada por valores e normas sociais, os quais indicariam qual emoção seria adequada expressar e qual a intensidade dessa manifestação. Averill (1993) descreve algumas regras da manifestação da raiva, dentre elas a proporcionalidade da reação a um estímulo, ou seja, a reação não pode exceder o necessário para corrigir a situação.

Como exemplo extremo, o próprio conceito de emoção, embora tido como universal no estudo de Brandt e Boucher (1986), pode não ter um termo correlato válido universalmente (LUTZ, 1982; HOWELL, 1981).

Russell (1991, p. 444) compartilha dessa visão e afirma que:

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

A partir dessa assunção, Russell (1991) afirma que um dos pontos principais de se estudarem as emoções em diferentes culturas é saber se pessoas de diferentes línguas estabelecem a mesma fronteira imposta pelo termo emoção. Novamente, nota-se o interesse em termos ou palavras ou, de outra maneira, no modo como as emoções são categorizadas.

Um dos estudos mais importantes desse tipo de abordagem é o artigo “A constructivist view of emotion”, de James Averill (1980). Nele, o autor, variadas vezes, enfatiza que a emoção é melhor explicada, desde a sociedade e seus papéis, como em um grande *drama*.

O termo *construtivista* tem dois significados para o autor. O primeiro refere-se às emoções como *construções sociais*, não como dados biológicos; afirmação um tanto peremptória, mas que é desacentuada pelo autor ao longo do texto. Dessa forma, sua teoria coloca a sociedade e os papéis desempenhados pelos seus membros em primeiro plano, sem excluir a contribuição dos fatores biológicos na manifestação das emoções. Assim, emoções são construtos sociais e só podem ser totalmente entendidas por uma análise do ponto de vista social. O segundo significado refere-se às emoções como *improvisações* baseadas em interpretações subjetivas feitas pelos indivíduos.

Averill (1980) entende as emoções como síndromes socialmente constituídas ou como papéis sociais transitórios. Num primeiro momento,

o termo *síndrome* parece inapropriado, uma vez que é fortemente marcado pelo uso em medicina ou psicologia, todavia o seu uso enquadra-se perfeitamente ao que o autor pretende descrever. Segundo Averill (1980, p. 307), “*a syndrome may be defined as a set of responses that covary in a systematic fashion*”.

Os termos *set* [conjunto] e *systematic* [sistemático] são muito importantes para a definição de síndrome proposta por Averill (1980). O primeiro termo implica que síndromes são formadas por uma variedade de diferentes elementos, i.e. a síndrome não é uma reação unitária e invariável; o segundo termo implica sistematicidade e coerência entre os elementos. A definição de síndrome como um conjunto coerente de mecanismos ou regras é muito ampla, pois não define a natureza dos mecanismos que regem as reações afetivas. No caso dos papéis sociais, a natureza das normas relevantes é estipulada por normas sociais e expectativas sobre comportamentos apropriados (AVERILL, 1980, p. 308).

As normas sociais que fazem parte dessas síndromes são representadas psicologicamente como estruturas cognitivistas ou esquemas. Segundo Averill (1980, p. 308), essas estruturas seriam semelhantes à gramática de uma língua, pois fornecem a base para a avaliação de estímulos, a organização das respostas e a vigilância do comportamento; em outras palavras, para a improvisação de papéis emocionais. Averill (1980) afirma que a definição de emoção como síndromes ou papéis sociais transitórios não é suficiente ou muito ampla, pois tal definição abarcaria outros tipos de fenômenos psicológicos não ligados necessariamente às emoções.

Para tornar a definição acima mais precisa, o autor cita outros dois conceitos: *appraisal*¹⁸ e passividade. O primeiro é um conceito central não

¹⁸ Um pequeno parêntese é necessário com relação à dificuldade de tradução do termo *appraisal*, usado largamente em estudos cognitivos que lidam com a emoção. Uma das alternativas é entender *appraisal* como uma *avalição* que o indivíduo faz do ambiente. O problema aqui é que o termo *appraisal*, quando usado nos estudos que lidam com emoção,

só para a abordagem construtivista, mas especialmente para a cognitivista. Averill (1980) refere-se às emoções como sendo direcionadas a algo, a algum objeto no mundo, ou seja, uma pessoa não fica simplesmente com raiva ou com medo, ao contrário, fica “com raiva de alguma coisa” ou “com medo de algo/alguém”. Assim, o objeto ao qual a emoção está relacionada depende da avaliação do sujeito, ou do *emotional appraisal* do indivíduo.

Após afirmar que as emoções são uma classe de papéis sociais transitórios, portanto diferentes das síndromes essencialmente orgânicas, o termo *appraisal* é essencial para o entendimento da abordagem construtivista. Dessa maneira, os papéis sociais são interpretações ativas do indivíduo. Emoções são dependentes de atividade cognitiva, o que as difere não só de outros tipos de síndrome orgânicas, mas também de certos tipos de papéis sociais.

Apesar de sua importância, *appraisal* por si só não basta para diferenciar fenômenos emocionais e não emocionais. Averill (1980) comenta que podemos avaliar uma situação como sendo uma transgressão passível de punição, mas sem sentirmos raiva. Para tal, o segundo conceito – passividade – diz respeito a um termo antigo para emoção: paixão (*pathos*). A emoção não é algo que fazemos, mas que acontece conosco, i.e., algo que experienciamos. Em resumo, segundo a visão socioconstrutivista, a emoção pode ser definida como “*a transitory social role (a socially constituted syndrome) that includes an individual's appraisal of the situation and that is interpreted as a passion rather than as an action*” (AVERILL, 1980, p. 312).

2.3 A perspectiva cognitivista

Para tratar de maneira mais abrangente da área de estudos cognitivos seria necessário recuar aos filósofos gregos clássicos, tarefa que extrapola o intuito deste trabalho e que já foi brevemente exposta na

não se trata de uma *avaliação* distanciada ou, nas palavras de Averill (1980), fria. Muito pelo contrário, *appraisal* em emoção é altamente pessoal.

introdução. Um breve parêntese, porém, será necessário para situar minimamente essa perspectiva de análise. Matlin (2003) traça, de maneira sucinta, o trajeto histórico pelo qual percorreram os estudos cognitivos até os dias de hoje. A autora traça um paralelo entre a proposta aristotélica de que os seres humanos adquirem conhecimento através de experiência e observação e a ênfase que os psicólogos cognitivistas dão à experiência empírica obtida por investigação científica e metodologia cuidadosas. A proposta aristotélica desencadeou séculos de debates filosóficos, mas o seu estudo sistemático do ponto de vista psicológico aconteceu muito tempo depois, com o advento da psicologia como disciplina autônoma no final do século XIX.¹⁹ Especificamente, sua vertente cognitivista teve início somente na segunda metade do século XX.

Nesse tipo de vertente de estudos, o foco está no próprio processo da emoção. De maneira geral, a emoção começaria com a percepção de um estímulo (externo ou interno) e terminaria por ativar uma resposta corporal (mudanças no batimento cardíaco, sudorese, expressão facial). No caso específico da fala, essas mudanças afetariam os parâmetros fonéticos da produção dos sons da fala.

Do ponto de vista cognitivo, a primeira fase do processo emocional – percepção do estímulo – é decisiva para determinar qual emoção será originada. Nota-se aqui que é a percepção do estímulo que está em pauta, não o estímulo *per se*. Para deixar claro, os pesquisadores cognitivistas têm se debruçado sobre dois problemas relacionados aos estímulos e suas respostas. O primeiro está relacionado à elicitación da mesma emoção por eventos totalmente diferentes em sua natureza. Roseman e Smith (2001, p. 4) afirmam que “*it is impossible to list all the elicitors of an emotion, because any emotion may be evoked by an infinite number of events, including events that have never been encountered previously*”. Por

¹⁹ O ano de 1879 é tido como o ano do nascimento da psicologia científica, devido à inauguração do laboratório de Wilhelm Wundt, em Leipzig, Alemanha.

exemplo, a raiva pode ser elicitada por diferentes estímulos, como um desentendimento no trânsito em horário de pico ou alguma forma de trapaça num jogo de azar. O segundo problema seria o inverso do primeiro ou a sua outra face, ou seja, o mesmo estímulo desencadearia diferentes respostas emocionais em pessoas diferentes, ou a mesma pessoa responderia diferentemente ao mesmo estímulo em tempos diferentes.

Para os autores, a variação temporal e individual na resposta emocional é difícil de ser explicada por teorias que afirmam que eventos são as causas diretas das respostas emocionais ou que as emoções são respostas a estímulos determinados evolutivamente ou aprendidos por generalização e associação. Claramente, a crítica aqui destina-se às perspectivas evolutivas.

Num primeiro momento, o efeito da emoção na cognição foi ignorado pelos psicólogos cognitivistas, assegurando que os participantes de experimentos estivessem em estados emocionais neutros. Entretanto, com o passar do tempo, houve um crescimento significativo de estudos sobre cognição e emoção. Dentro do escopo da perspectiva cognitivista há duas questões importantes sobre cognição e emoção. Primeira, como a cognição influencia nossa experiência emocional? Segunda, como a emoção influencia nossos processos cognitivos? (EYSENCK; KEANE, 2015).

Cornelius (2000) e Scherer (1999), embora também atribuam o início das indagações cognitivistas aos filósofos helenistas, reconhecem que o advento da abordagem cognitivista dos estudos da emoção teve início no trabalho de Magda Arnold (1960). Como exposto em Cornelius (2004, p. 5), pensamento e emoção são inseparáveis dentro da perspectiva cognitivista. Especificamente, as emoções são dependentes de um processo chamado *appraisal*, brevemente tratado anteriormente, que consiste num processo em que eventos são avaliados pelas pessoas como bons ou ruins

(CORNELIUS, 2000, p. 3).²⁰ As *appraisal theories* defendem que a interpretação que as pessoas fazem de uma situação é que dá origem à emoção, ou seja, não é a situação em si que desencadeia estados emocionais, mas a avaliação que se faz dela.

No início dos estudos cognitivos da emoção, Arnold (1960) criticou a afirmação de James segundo a qual as mudanças corporais seguem-se diretamente à percepção do fato existente. A sensação causada por essas mudanças no momento mesmo em que ocorrem é a emoção. As questões levantadas por Arnold (1960) são: como a percepção acontece? Que tipo de percepção é essa? E como ela acarreta as mudanças corporais referidas por James? Para Arnold, a resposta está no processo de *appraisal* (*apud* CORNELIUS, 2000).

Arnold foi a primeira a fazer uso do termo *appraisal* para explicar a elicitação de diferentes emoções. Diferentemente da visão unidimensional exposta por Cornelius, Arnold (1960) propôs que eventos são avaliados (*appraised*) com base em três dimensões: benéfico *vs.* prejudicial, presença *vs.* ausência e evitar *vs.* aproximar-se de um objeto. Um dos pressupostos fundamentais das teorias cognitivistas que tratam de *appraisal* é a ideia de que as pessoas apresentam emoções diferentes para a mesma situação, dependendo de como elas a interpretam e avaliam. Assim, é a avaliação de uma situação, não a situação em si, que determina a qualidade e a intensidade de uma resposta emocional (SIEMER; MAUSS; GROSS, 2007).

As teorias que lidam com *appraisal* não formam um todo coeso, pelo contrário, apresentam variações em seus pressupostos. Gross e Barret (2011) dividem-nas em três grupos, deixando bem claro que a divisão não é estanque, i.e., os grupos não são isolados e apresentam alguns pontos de sobreposição.²¹ A primeira perspectiva é a proposta por Arnold (1960), que

²⁰ Veremos adiante que o processo de *appraisal* é mais complexo e envolve um número maior de critérios, não se limitando somente à dimensão “bom-ruim”.

²¹ Os autores categorizaram quatro perspectivas teóricas (emoções básicas, *appraisal*, construção psicológica e socioconstrutivismo) de acordo com uma lista de dez questões

entende *appraisal* como sendo um antecedente cognitivo das emoções que produzem o significado a partir do mundo. Gross e Barret (2011, p. 10) ilustram esse tipo de modelo:

In these models [...] appraisals are like a set of switches, which when configured in certain patterns, trigger biologically basic emotional responses characterized either by stereotyped outputs or by a strong and almost inescapable tendency to interact with the world in a particular way.

No segundo tipo de abordagem, *appraisal* não é definida como causa das emoções, mas como parte constitutiva delas. Por sua vez, as emoções são tidas como uma tendência de resposta vagamente coordenada e determinada pelo contexto. Ademais, nesse tipo de perspectiva as emoções são associadas a tendências de ação, ou seja, disposições para se relacionar com o mundo de maneira particular (SCHERER, 1984; FRIJDA, 1986; LAZARUS, 1991).

Na terceira perspectiva, emoções são formas de experimentar o mundo. Nessa perspectiva, *appraisal* mantém a hipótese de que as emoções são estados funcionais distintos. Nesse tipo de teoria, as emoções são cada vez mais vistas como atos emergentes de criação de significado que têm seu conteúdo descrito por *appraisal*. Dessa forma, estar num estado de tristeza é experienciar a perda; estar num estado de raiva é experienciar ofensa ou transgressão, etc.

Uma das críticas aos modelos cognitivos em geral é a intelectualização exagerada das emoções. De fato, é verdade que muitas vezes o surgimento ou a percepção de estados emocionais ocorre de maneira inconsciente. Cornelius (2000), ao comentar a objeção feita por Zajonc (1980) e LeDoux (1996), baseada na demasiada intelectualização das emoções, afirma que a crítica é fruto de uma interpretação equivocada feita

centrais. De acordo com as respostas dadas a elas, as *appraisal theories* foram divididas em três partes.

pelos pesquisadores sobre o que seria *appraisal*. Arnold (1960) descreve *appraisal* como sendo direto, imediato, não reflexivo, não intelectual e automático. Em outras palavras, *appraisals* não são deliberados (ARNOLD, 1960).

Smith e Kirby (2001) acreditam que o processo de *appraisal* pode ser dividido em dois tipos distintos. O primeiro deles é chamado de processamento associativo, envolvendo *priming* e ativação de memórias e pode ocorrer de forma rápida e automática. O segundo envolve raciocínio, ou seja, é um processo mais controlado, deliberado e flexível, além de ser mais lento. Entender *appraisal* a partir desse ponto de vista parece ser mais adequado, pois de fato a emoção pode se manifestar de maneira imediata ou por meio de um processo mais lento que demanda memória e raciocínio.

Para a perspectiva cognitivista, *appraisals* são associados a todas as emoções. Uma consequência lógica desse tipo de afirmação é que a mudança no *appraisal* acarreta necessariamente uma alteração na emoção. Para testar essa hipótese, estudos empíricos verificaram a relação entre algumas configurações de *appraisal* e a natureza da emoção subsequente. Segundo Scherer (1999), a estratégia mais comum em estudos dessa natureza é pedir aos participantes que se lembrem de experiências emocionais específicas e perguntar sobre os processos de avaliação que antecederam a emoção. Outra maneira é induzir emoções experimentalmente e acessar o processo de *appraisal* por meio de relato. O uso de cenários também é bastante profícuo. Os participantes, após terem contato com um cenário específico relacionado a um determinado *appraisal*, têm de indicar qual reação emocional teriam nessa situação (SCHERER, 1999, p. 641).

A postulação de *appraisal* como sendo a causa das emoções é controverso e levantou muita crítica. Para Reisenzein (1995), há um problema de natureza metodológica, pois *appraisals* só são analisados após

o efeito alegadamente causado por ele: a emoção. De outra maneira, os relatos dos participantes sobre o que se passou antes da emoção em si não são totalmente confiáveis e sujeitos a interferências. Ademais, para o autor, não há evidências que *appraisals* causam emoções, pois seria necessário demonstrar uma relação de causa e consequência que pressupõe o ordenamento temporal entre *appraisal* e emoção.

2.4 Integrando as perspectivas teóricas

Após essa breve visão geral sobre as perspectivas teóricas dos estudos da emoção, duas questões podem ser levantadas: como integrá-las sem que haja contradição entre elas? Como essas teorias servirão como base para este trabalho?

Cornelius (1996) demonstra em seu artigo que a sobreposição de perspectivas é inevitável e que já está em curso nos trabalhos de Ekman (1972) e Levenson (1992), nos quais pode-se notar uma interação entre as perspectivas darwiniana e jamesiana por atestar a concomitância entre as expressões faciais das emoções e as mudanças corporais associadas a elas. Além disso, trabalhos como os de Lazarus (1991), mais ligado à corrente cognitivista, também tratou da universalidade das expressões faciais.

Outro ponto de intersecção que parece estar bem estabelecido ocorre entre as perspectivas socioconstrutivista e cognitivista. Vimos anteriormente que a visão sociocognitivista considera que a emoção é sempre direcionada a algo e que depende de uma avaliação feita pela pessoa diante do objeto *appraised* em questão, somando-se a ela a passividade (emoção como experiência, não como ação). Portanto, do ponto de vista socioconstrutivista, *appraisal* é uma condição necessária, mas não suficiente para que haja emoção.

Neste trabalho, a manifestação das emoções será analisada do ponto de vista da variação entre línguas distintas, fazendo-se, portanto, necessário considerar as emoções desde um ponto de vista sociocultural. É

importante salientar que este estudo não tomará a teoria socioconstrutivista em sua plenitude, isto é, não será uma simples aplicação da teoria para verificação de seus pressupostos. A teoria socioconstrutivista servirá como efeito de contraste aos pressupostos adotados nas teorias evolutivas, ou seja, a análise das emoções como fenômenos universais e adaptações ao ambiente com intuito de preservação. Como se viu anteriormente, para o socioconstrutivismo a expressão da emoção (ou a não expressão) é intermediada por normas sociais e valores que só são definidos em ambiente social. Esses pressupostos serão entendidos, pois, como forças opostas que agem na configuração dos estados e expressões emocionais.

Com vistas a tratar da fala emotiva considerando esses dois vieses de análise, social e evolutivo, Scherer (2006) propõe uma abordagem baseada nos efeitos *push* e *pull*.²² A vantagem desse tipo de análise é abarcar não somente o que é geral/universal na fala emotiva, mas também tratar das características linguísticas locais responsáveis pela variação que essas manifestações possam apresentar em diversas línguas.

O efeito *push* é entendido como a influência da ativação psicofisiológica, ligada fortemente à fala emotiva; ao passo que o efeito *pull* é atribuído às normas culturais de expressão, ligadas à pragmática²³ da língua. Essa distinção implica que o efeito *push*, marcas biológicas, é diretamente exteriorizado na expressão motora, conseqüentemente, na produção da fala; ao passo que o efeito *pull* é baseado em modelos socioculturais que são apreendidos pelos membros de cada sociedade. Com relação às diferenças entre as culturas, é esperado que haja poucas diferenças quando se trata do efeito *push*, e muitas quando se trata do efeito *pull*.

²² Optou-se pelos termos em inglês.

²³ O termo “pragmática” em Scherer (2006) não corresponde exatamente ao nível de análise estudado pela Linguística, pois é tomado de maneira geral, indicando características gramaticais amplas da língua.

Para a análise, Scherer (2006) propõe os princípios de covariação e configuração. O princípio da covariação assume uma relação contínua, mas não necessariamente linear, entre o aspecto emocional e a variação acústica. Se esse princípio é verdadeiro, experimentos perceptuais que tentam buscar alguma relação entre emoções e correlatos acústicos (ou de outra natureza – visuais, por exemplo) devem demonstrar que mudanças em determinados parâmetros podem prever qual será o julgamento dos participantes. Por exemplo, a frequência fundamental (F0), a qualidade de voz e a energia das vocalizações podem assinalar as expressões de medo, raiva, alegria, etc. Esse princípio é definido por Scherer (2006) como um estágio anterior com relação à evolução da linguagem. O princípio de configuração determina que o significado específico transmitido por um enunciado é inferido pelo ouvinte a partir da configuração prosódica como um todo e das escolhas linguísticas. Esse princípio determina as características pragmáticas da fala, como ênfase e tipos de sentenças (SCHERER, 2006, p. 13-14). É um princípio mais recente do ponto de vista evolutivo, ligado à emergência da linguagem humana.

Assim, é possível associar as variáveis contínuas (parâmetros acústicos) com o efeito *push*, ligado à exteriorização de estados emocionais, enquanto as configurações de categorias pragmáticas da fala são melhor relacionadas ao efeito *pull*, ligado aos contornos entoacionais e demais tipos de padrões encontrados nas línguas naturais.

Uma ressalva pode ser feita quanto ao choro, na medida em que essa manifestação vocal é associada ao efeito *push*, mas apresenta padrões de intensidade e declínio de F0 bem característicos, aproximando-se do que se conhece como contorno entoacional. O estudo de Mampe *et al.* (2009) demonstrou que o choro de recém-nascidos é fortemente determinado pelos padrões entoacionais de suas línguas maternas, percebidas ainda antes do parto. Bebês franceses tendem a produzir padrões entoacionais ascendentes (L – H), ao passo que bebês alemães tendem a produzir

padrões descendentes (H – L). Por essa razão, não está claro se a proposta de divisão dos efeitos *pull* e *push* pode ser tomada de maneira estanque, pois pressupõe a ação desses efeitos, mas não a sua interação, sem a qual seria impossível obter os resultados encontrados em Mampe *et al.* (2009).

A Figura 1 é uma esquematização de como as perspectivas teóricas abordam as emoções e como elas podem interagir numa análise perceptiva da fala.

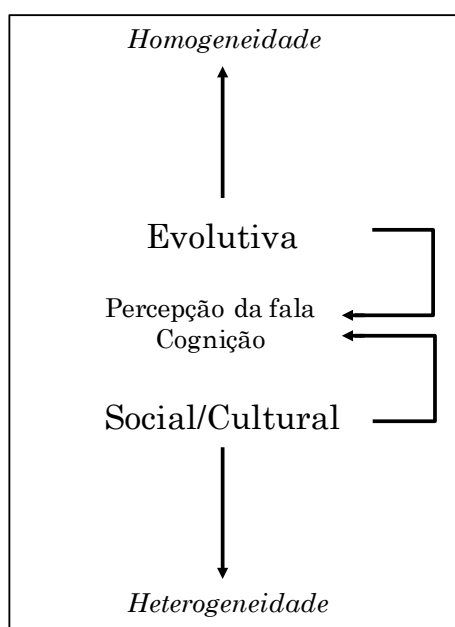


Figura 1: Esquema que ilustra a maneira como as perspectivas de análise das emoções irão interagir neste trabalho

No esquema acima, há também o fator cognitivo na percepção da fala. Dessa forma, a perspectiva cognitivista faz-se necessária por se tratar de um estudo experimental e perceptivo da manifestação das emoções na fala. Novamente, não se trata da utilização fiel e textual das teorias cognitivistas, mas do entendimento de que o processo emocional é baseado em mecanismos que passam pela percepção do estímulo e de sua avaliação pelos sujeitos.²⁴

²⁴ No experimento com abordagem dimensional das emoções, o pressuposto cognitivo da avaliação das emoções por meio de *appraisals* é condição necessária para esse tipo de análise.

2.5 Modelos de análise

2.5.1 Emoções básicas

Neste subitem, serão tratados os dois tipos de análise das emoções que dão suporte a este estudo. A primeira delas é a *abordagem com emoções básicas*,²⁵ também denominadas emoções fundamentais ou primárias. A segunda delas é a *abordagem dimensional das emoções* que, em certa medida, é antagônica à perspectiva de análise com emoções básicas. Após os três primeiros experimentos desenvolvidos neste trabalho, as duas abordagens serão cotejadas na análise dos resultados.

A existência de um grupo pequeno de emoções básicas é uma hipótese muito difundida entre alguns pesquisadores, a saber Ekman (1992), Plutchik (1980, 1984, 2001), dentre outros. A hipótese se sustenta em duas perspectivas, uma biológica e outra psicológica. A primeira apresenta características fisiológicas, neurológicas e anatômicas correspondentes às emoções básicas; a segunda trata as emoções básicas como estruturas fundamentais que, quando combinadas, formam as emoções não básicas, ou mais complexas (ORTONY; TURNER, 1990).

O estudo das emoções como primitivos biológicos propõe que as questões centrais sobre as emoções podem ser analisadas a partir de sua origem evolutiva e de sua importância para a manutenção da espécie. A análise baseia-se em encontrar e examinar os fundamentos biológicos da emoção. Outra característica que pode ser notada é o caráter funcional desse tipo de análise, pois apresenta como objetivo investigar a importância das emoções e suas funções em organismos e suas espécies (ORTONY; TURNER, 1990, p. 371).

O estudo das emoções básicas como primitivos psicológicos parte do pressuposto de que há um grupo limitado de emoções básicas não

²⁵ Algumas das emoções relatadas aqui, a partir de estudos em inglês, terão seus nomes mantidos na língua original tendo em mente a dificuldade de tradução em alguns casos.

reduzíveis, ou seja, não passíveis de serem divididas em emoções primárias. Esse pequeno grupo de emoções básicas é usado como *building blocks* em um modelo combinatório capaz de gerar emoções mais complexas. O modelo proposto por Plutchik (1980, 1984, 2001) é um bom exemplo desse tipo de análise e será tratado com mais detalhes adiante.

Ekman (1992) argumenta a favor das emoções básicas, apoiando-se em seus estudos prévios, que demonstraram que algumas expressões faciais da emoção são universais. O autor, a princípio, propusera que as emoções podiam ser explicadas de duas formas. A primeira delas era a utilização de uma escala que variava de *prazeroso* a *não prazeroso*. Para ele, essa dimensão era suficiente para explicar as diferenças entre as emoções. A segunda é que a relação entre as expressões faciais e seus significados era socialmente aprendida e culturalmente variável. Nas palavras dele:

This basic emotions perspective is in contrast to those who treat emotions as fundamentally similar in most respects, differing only in terms of one or more dimensions, the most common ones being arousal, pleasantness and activity, or those who carve emotions into just a positive and a negative state. (EKMAN, 1992, p. 170)

Nota-se que Ekman deixa de se apoiar em duas perspectivas de análise da emoção, a primeira predominantemente cognitivista e a segunda socioconstrutivista. No entanto, nenhuma delas foi descartada totalmente, mas assumiram uma posição secundária em sua abordagem.

Ekman define o termo *básica* de três maneiras diferentes. No primeiro uso, *básica* refere-se à assunção de que há um número determinado de emoções que são diferentes entre si de maneira considerável. Além disso, o primeiro significado refere-se não somente às diferenças que as emoções possuem em sua manifestação, mas a alterações importantes como a avaliação do ambiente, provável comportamento, fisiologia, etc. No segundo significado, *básica* indica que as emoções foram

desenvolvidas para a sua função adaptativa de lidar com tarefas fundamentais do dia a dia, i. e., a evolução desempenhou uma função importante na configuração da expressão e da função das emoções (primitivos biológicos) (EKMAN, 1992, p. 170). No terceiro significado, o termo é usado para postular que outras emoções, não básicas, são combinações das emoções básicas, como já mencionado anteriormente (primitivos psicológicos).

O estudo de Ekman (1992) baseou-se nos dois primeiros usos, ou seja, de que há emoções discretas fundamentais desenvolvidas para preparar os humanos para a ação. No entanto, não se pode reduzir a função das emoções, pois, além de servirem para lidar com imprevistos entre o organismo e as pessoas, entre animais e pessoas, as emoções podem surgir sem a presença de outros ou na imaginação. Como exemplo, pode-se ter reações emocionais com música (EKMAN, 1992, p. 171).

O número de emoções básicas varia de maneira significativa não só na quantidade, mas também na forma pela qual foram escolhidas por cada pesquisador. Apesar da variação presente entre as propostas, há uma grande sobreposição entre elas. Ekman e Friesen (1975), por exemplo, postularam seis emoções básicas baseadas nas expressões faciais (raiva, medo, tristeza, alegria, nojo e surpresa); Arnold (1960) propôs 11 emoções baseadas na prontidão para ação (raiva, aversão, coragem, abatimento, desejo, desespero, medo, ódio, esperança, amor e tristeza). Recuando um pouco mais no tempo, James (1884) postulou quatro emoções básicas fundamentadas em aspectos corporais (medo, dor, amor e raiva).²⁶

Embora haja uma certa concordância entre as perspectivas de análise com emoções básicas, Ortony e Turner (1990) apresentam alguns problemas entre elas. O primeiro problema é a discordância sobre quais seriam as emoções básicas, o segundo seria sobre a quantidade de emoções básicas assumida pelos pesquisadores. Além das diferenças, pode-se, por

²⁶ Para a tabela completa cf. Ortony e Turner (1990).

outro lado, atentar para a concordância entre as emoções básicas postuladas por diferentes pesquisadores. Entre as propostas, constata-se que as emoções *raiva* e *medo* estão presentes em todas elas.²⁷ Além disso, pode-se afirmar que a diferença entre as emoções básicas não é real, pois muitos termos podem referir à mesma emoção, como por exemplo, em inglês, *rage* e *anger*, *happiness* e *joy* (ORTONY; TURNER, 1990, p. 316). Todavia, os autores entendem que não há evidências para se afirmar por certo que termos diferentes sejam variações da mesma emoção. Outra justificativa para o descrédito desse tipo de argumento reside na imprecisão da língua com relação a estados psicológicos. Ainda que a diferença entre grupos de emoções básicas seja devido a uma inexatidão linguística, assumir tal justificativa apenas reduz o problema, mas não o soluciona. Por essa razão, no último experimento deste trabalho, será utilizada uma metodologia que visa a minimizar o papel de língua na categorização das emoções (cap. 9).

O terceiro e mais importante problema levantado por Ortony e Turner (1990) é o porquê de se adotarem emoções básicas e as consequências do sucesso ou não desses estudos. A principal razão, segundo os autores, é fornecer respostas a observações rotineiras sobre as emoções: i) algumas emoções aparentemente estão presentes em todas as culturas; ii) as expressões faciais das emoções parecem ser universais; iii) algumas emoções básicas são associadas a funções de sobrevivência. Nota-se que o escopo de análise é amplo e apresenta questionamentos que vão desde uma abordagem social da emoção, passando por um viés evolucionista, ou mesmo cognitivista, como em Arnold (1960).

Dessa forma, é importante enfatizar que a admissão da existência de emoções básicas não limita seu estudo à perspectiva evolutiva. Ekman deixa claro que um estudo socioconstrutivista seria possível tomando-se

²⁷ Em Ortony e Turner (1990), a concordância é alta. Em 13 abordagens, *anger* ou *rage* estão presentes em 12 delas.

somente o primeiro significado de *básicas*, ou seja, assumindo que há *a priori* um número de emoções distintas entre si. Outra hipótese levantada por Tooby e Cosmides (1990) é a de que a avaliação (*appraisal*) que fazemos de eventos atuais é influenciada pelo passado, em outras palavras, é também uma herança evolutiva. Na perspectiva dos autores, mesmo mecanismos como a consciência dependem de processos de categorização moldados ancestralmente e dependentes do passado (TOOBY; COSMIDES, 1990, p. 375). Esta última perspectiva é também uma perspectiva evolucionista, mas trata também de cognição.

Ekman e Friesen (1975) propõem que as emoções básicas não configuram somente um estado afetivo, mas uma família de estados. O termo família é entendido por Ekman (1992) como um grupo de coisas relacionadas por semelhança.²⁸ Assim, cada membro de uma família de emoção compartilha algumas características, e. g., semelhanças na expressão, na atividade fisiológica, na natureza dos eventos antecedentes que elicitam a emoção e talvez processos de *appraisal* (EKMAN, 1992, p. 172).

Os autores (EKMAN; FRIESEN, 1975) propuseram a análise de famílias de estados emotivos fundamentada nos resultados de sua pesquisa, na qual mais de 60 variações da expressão facial de raiva foram descritas. De acordo com esse resultado, a relação entre a emoção e sua expressão não é biunívoca, pelo contrário, há um espectro considerável de variação na expressão de uma mesma emoção. Essa variação foi interpretada em termos do controle da emoção (simulação ou espontânea) e das características dos eventos que a desencadearam. A ideia de que haja famílias de estados afetivos, em outras palavras, de que haja variação na expressão de uma mesma emoção, é muito interessante por levantar uma questão: o mesmo princípio poderia ser aplicado à expressão da emoção na

²⁸ O autor deixa claro que o termo *família* não tem o sentido comum encontrado em famílias humanas, ou seja, o de uma estrutura.

fala? A questão não diz respeito somente a um conjunto de parâmetros acústicos já estudado exaustivamente na literatura, mas a interação entre eles e possíveis grupos e combinações, i. e., os parâmetros acústicos não se comportariam de maneira independente, mas seriam variáveis que interagem na produção da expressão da emoção na fala. Da mesma forma que a expressão facial da emoção é resultado da ação de grupos musculares, a fala também está sujeita à ação da musculatura. Em casos específicos, a expressão facial da emoção e a fala compartilham os mesmos grupos musculares. Portanto é razoável pensar numa variação da fala emotiva semelhante à encontrada nas expressões faciais.²⁹

A existência de famílias de estados emotivos torna a tarefa de descrever as emoções e suas expressões menos útil do que trabalhosa, pois, a despeito da grande variedade na expressão, as pessoas não apresentam grandes problemas em reconhecê-las, mesmo em experimentos controlados como os que veremos neste trabalho. Por esse ponto de vista, teríamos, como em Borges (1982), um “mapa do império do tamanho do império”. Para não enveredar pela descrição pura e simples, Ekman (1992) propõe uma metáfora musical para lidar com tamanha variação, a saber, *tema e variação*. O tema é composto por características únicas a uma família, já as variações seriam resultado de outras influências como, por exemplo, diferenças na constituição biológica, nas experiências de aprendizado e nas situações que evocam as emoções. Em resumo, com a noção de tema e variações, Ekman (1992) explica a grande variedade de emoções encontrada por meio da variação na expressão de um grupo reduzido de emoções básicas. Em contrapartida, na teoria proposta por Plutchik (1980, 1984, 2001), a variedade de emoções é explicada pela combinação das

²⁹ Embora uma das características mais salientes da evolução na fala seja o controle independente da frequência fundamental e da intensidade, a interação entre alguns músculos é inevitável.

emoções básicas de maneira que produzam outras categorias, não apenas variações da mesma emoção.

Plutchik (1980, 1984, 2001) propõe oito emoções básicas que, quando combinadas, dão origem a outras emoções não básicas. Para o autor, uma emoção é uma:

[...] complex sequence of reactions to a stimulus and includes cognitive evaluation, subjective changes, autonomic and neural arousal, impulses to action and behavior designed to have an effect upon the stimulus that initiated the complex sequence. (PLUTCHIK, 1980, p. 361)

Sobre as emoções, Plutchik (1980, 1984) afirma que funções adaptativas básicas das emoções ocorrem em todos os animais, mas que a sua realização com graus maiores de complexidade aparece nos animais superiores, dentre eles os humanos. O quadro abaixo apresenta emoções e suas respectivas funções adaptativas.

Quadro 2: Funções adaptativas das emoções (PLUTCHIK, 1980)

<i>Função adaptativa</i>	<i>Emoções correlatas</i>
<i>Protection</i>	<i>Fear</i>
<i>Destruction</i>	<i>Anger</i>
<i>Rejection</i>	<i>Disgust</i>
<i>Exploration</i>	<i>Anticipation</i>
<i>Reproduction</i>	<i>Joy</i>
<i>Affiliation</i>	<i>Trust</i>
<i>Reintegration</i>	<i>Sadness</i>
<i>Orientation</i>	<i>Surprise</i>

A partir das emoções básicas, o pesquisador propôs a formação de díades, subdivididas em primárias, secundárias e terciárias. A hierarquia entre elas deve-se a uma ordem crescente de ocorrência, das mais comuns às mais raras.

Quadro 3: Emoções primárias e suas combinações (díades), segundo a proposta de Plutchik (1980)

Emoções básicas	Emoções básicas	Díades	Resultado das combinações
Anger +	Anticipation	1 ^a	= Aggression
	Joy	2 ^a	= Pride
	Trust	3 ^a	= Dominance
Fear +	Surprise	1 ^a	= Alarm
	Sadness	2 ^a	= Despair
	Disgust	3 ^a	= Shame
Joy +	Trust	1 ^a	= Love
	Fear	2 ^a	= Guilt
	Surprise	3 ^a	= Delight
Sadness +	Disgust	1 ^a	= Remorse
	Anger	2 ^a	= Envy
	Anticipation	3 ^a	= Pessimism
Trust +	Fear	1 ^a	= Submission
	Surprise	2 ^a	= Curiosity
	Sadness	3 ^a	= Sentimentality
Disgust +	Anger	1 ^a	= Contempt
	Anticipation	2 ^a	= Cynism
	Joy	3 ^a	= Morbidity
Anticipation +	Joy	1 ^a	= Optimism
	Trust	2 ^a	= Fatalism
	Fear	3 ^a	= Anxiety
Surprise +	Sadness	1 ^a	= Disappointment
	Disgust	2 ^a	?
	Anger	3 ^a	= Outrage

Além das díades, Plutchik (1980) divide as oito emoções primárias em emoções opostas, uma proposta já vista em Aristóteles. As oposições são: *fear vs anger*; *sadness vs joy*; *trust vs disgust*; *surprise vs anticipation*.

A despeito das limitações supramencionadas, as propostas de análise com emoções básicas são uma maneira profícua de analisar a manifestação das emoções, além de ser uma perspectiva que transita em grandes áreas teóricas como a evolucionista, socioconstrutivista e cognitivista. Para este trabalho, essa abordagem será utilizada no primeiro experimento – piloto –, de forma que os estímulos selecionados possam ser avaliados quanto à concordância das respostas dadas pelos participantes, em outras palavras, servirá para verificar se se trata da mesma emoção, *e. g.*, se o estímulo selecionado pelo experimentador como *alegria* é de fato

percebido como alegria pela maioria dos participantes. Essa concordância abalizará quais emoções estão sendo de fato analisadas neste trabalho.

2.5.2 Análise dimensional das emoções

Como visto anteriormente, os modelos de análise baseados em emoções básicas podem ser definidos como modelos discretos. Nos trabalhos de Ekman e Friesen (1975) e Ekman (1992), a tarefa principal é identificar e descrever as emoções com base nas características que as diferenciam entre si. Todavia, pesquisas na área da psicologia da emoção apresentam uma análise alternativa do fenômeno. De acordo com elas, as emoções podem ser analisadas por meio de pontos num espaço multidimensional.³⁰ Assim, os modelos de análise dimensional diferem e definem as emoções como sendo um produto de um conjunto subjacente de dimensões.

O estudo pioneiro de Wundt (1897) foi uma das primeiras propostas para um sistema multidimensional ao sugerir que a natureza de cada estado emocional é definida por sua posição entre dimensões ortogonais: *pleasantness-unpleasantness, rest-activation e relaxation-attention*. Para melhor entendimento, é necessária uma breve incursão ao modelo proposto pelo autor. Para Wundt, o conteúdo da experiência psíquica é de caráter composto (1897, p. 30). Assim, é necessário que haja uma análise em elementos menores de maneira que se possa estudar quais são seus componentes e suas interações.

Há dois níveis de análise: o nível dos elementos psíquicos e o nível dos compostos psíquicos. O nível básico é formado por *sensações* – resultado do estímulo dos órgãos do sentido, chamado também de *conteúdo objetivo*; e por *sentimentos* – que acompanham as sensações, ou seja, as impressões

³⁰ Uma das vantagens desse tipo de análise é a possibilidade de se analisar a intensidade das emoções.

que as sensações despertam. O sentimento é analisado por meio de um modelo tridimensional: *pleasantness-unpleasantness*, *rest-activation* e *relaxation-attention*.

Para Wundt (1897), o conteúdo da experiência psíquica consiste de numerosas combinações de sensações e sentimentos. Por exemplo, a sensação de calor pode desencadear sentimentos diversos que podem ser descritos pela contribuição de cada dimensão. O esquema abaixo ilustra o modelo tridimensional proposto por Wundt (1897):

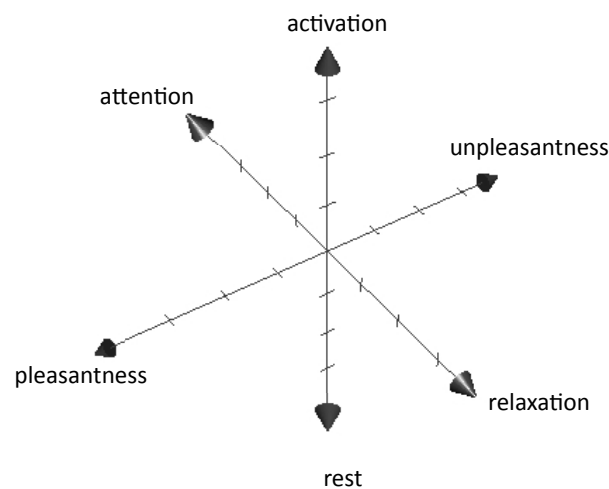


Figura 2: Modelo tridimensional de Wundt (1897)

Os complexos psíquicos propostos por Wundt são combinações de vários elementos psíquicos. A emoção é um tipo de complexo psíquico composto por vários sentimentos simples que agem de forma mais intensa do que um único sentimento simples. A proposta de Wundt (1897) não foi baseada em experimentos controlados, fato que veio a ocorrer mais tarde com o trabalho de Schlosberg (1941). O estudo desenvolvido por ele é muito interessante, sobretudo como exemplo comparativo entre modelos discretos (emoções básicas) e dimensionais. Schlosberg baseou-se no estudo de Woodworth (1938), que propôs uma escala linear para organizar expressões

faciais, e obteve seis graus: 1) *love/happiness/mirth*, 2) *surprise*, 3) *fear/suffering*, 4) *anger/determination*, 5) *disgust*, 6) *contempt*.³¹

Schlosberg (1941) pediu para que os participantes organizassem o mesmo material (fotos)³² usado por Woodworth em caixas separadas e nomeadas com as seis categorias supramencionadas. A pesquisa de Schlosberg propôs uma escala a partir de julgamentos feitos por 45 participantes sobre três conjuntos idênticos de 72 fotografias (216 no total). Schlosberg concluiu, a partir dos resultados obtidos, que as categorias propostas seriam melhor representadas por meio de um espaço bidimensional por acreditar que haveria algumas variáveis subjacentes que regessem a escala resultante de seu experimento. Para o autor, as variáveis subjacentes seriam as dimensões valência e ativação, dispostas num plano circular (modelo circumplexo, mais tarde revisto por Russell (1980)). Esse modelo foi previsto por conta das sobreposições encontradas entre as categorias. O trabalho recebeu críticas por que não seria capaz de distinguir emoções como medo ou raiva, pois ambas tiveram um alto grau de valência, ou seja, não seriam emoções *prazerosas*. Além disso, medo e raiva são tidas como emoções que possuem um alto grau de ativação (LAUKKA et al., 2005, p. 635). Mais tarde, Schlosberg (1954) propôs um outro modelo composto por três dimensões (*pleasantness-unpleasantness*, *attention-rejection* e *activation* [sleep – tension]).

Da mesma forma que as abordagens com emoções básicas variam na escolha de quais emoções seriam primárias, as abordagens dimensionais também apresentam a mesma característica variável. Trnka (2011) demonstra que, em sete modelos dimensionais (WUNDT, 1897; SCHLOSBERG, 1952, 1954; OSGOOD, 1966; DAVITZ, 1969; PARAMEI et

³¹ Optou-se pela não tradução dos graus propostos por Woodworth (1938) por conta de problemas como, por exemplo, na tradução de *mirth* e *happiness*, pois muitas vezes podem ser tidos como sinônimos. O dicionário Cambridge define *mirth* como *laughter, humour or happiness*.

³² O conjunto de fotos de expressões faciais de *Frois Wittman* descrito em Hulin e Katz (1935).

al., 1994; SOKOLOV; BOUCSEIN, 2000), somente a dimensão valência (*prazeroso-não prazeroso*) é comum a todos, seguida pela dimensão ativação (*arousal-calm*), presente em cinco dos sete estudos acima. Osgood (1966) é um bom exemplo de variação não só do número de dimensões, mas também da metodologia aplicada para se obterem as dimensões. Em seu experimento, Osgood (1966) utilizou 50 pares de adjetivos (claro-escuro, azedo-doce, quente-frio, etc.) e pediu para que participantes julgassem alguns conceitos (criança, mãe, trabalho, etc.). Com uma análise fatorial dos julgamentos dos participantes, Osgood (1966) propôs um sistema com três dimensões (*evaluation, potency e activity*).

Church *et al.* (1998) publicaram estudo semelhante, mas com foco na variação das dimensões entre culturas distintas, buscando analisar a universalidade das dimensões no estudo da emoção. Os autores utilizaram uma lista com 256 termos relacionados a emoções em filipino. Após a listagem dos termos, eles foram comparados com termos em inglês utilizando técnicas de *clusters*. Após análise de escalonamento multidimensional, Church *et al.* (1998) encontraram seis dimensões, podendo variar também entre modelos com duas ou três dimensões. As dimensões valência e ativação foram as mais eficazes em agrupar os termos emotivos baseados na similaridade entre eles.

Outros modelos como o PANAS – *Positive and Negative Affect Scales* – (WATSON; CLARK; TELLEGEN, 1988) foram largamente utilizados em psicometria e, atualmente, possuem variações como PANAS-X³³ (com 60 palavras para serem julgadas) ou I-PANAS-SF,³⁴ sendo este último um modelo simplificado. A escala de afetos positivos é composta de palavras que denotam emoções e sentimentos positivos, já a escala de afetos negativos é composta de palavras que denotam emoções e sentimentos

³³ Disponível em: <https://wiki.aalto.fi/download/attachments/50102838/PANAS-X-scale_spec.pdf>. Acesso em: 12 janeiro 2014.

³⁴ *The International Positive and Negative Affect Schedule Short Form.*

negativos. Os participantes devem se autoavaliar de acordo com seu estado emocional, portanto configura-se como uma abordagem com foco na introspecção como ferramenta de análise. Na versão reduzida (PANAS-SF), cada uma dessas dimensões é representada por 10 adjetivos. Segundo os autores, as duas dimensões conseguem explicar de metade a $\frac{3}{4}$ da variância das autoavaliações de afeto.

A diversidade das perspectivas dimensionais pode ser notada também nas diferenças entre os rótulos das extremidades de cada dimensão. Ativação, por exemplo, pode ser um contínuo entre *calm*, *sleep* e *rest* (baixa ativação) e *arousal*, *tension* e *excitation* (alta ativação). Outra questão importante nas abordagens dimensionais é o número necessário de dimensões para análise das emoções, razão pela qual alguns modelos acabam por fazer uso de apenas uma dimensão. As dimensões ativação e valência, cada uma de maneira isolada, são tidas como suficientes para analisar as emoções, como visto em Schlosberg (1941). Por outro lado, Trnka (2011) defende que modelos bidimensionais ou tridimensionais são insuficientes e que, portanto, as emoções não podem ser descritas por meio de poucas e limitadas dimensões. Para justificar, Trnka faz menção ao sentido da visão que, segundo ele, é perceptualmente limitado a objetos com duas ou três dimensões. Em analogia, Trnka (2011) propõe que as emoções seriam como uma experiência auditiva de um concerto, no qual variações de timbres, tempos e intensidade se combinam, ou melhor, como um piano, sendo cada tecla uma dimensão que pode ser acionada de maneira diferente. Sem dúvida é uma proposta interessante por conta da analogia e de uma intuição de que as emoções são estados muito mais complexos do que os modelos dimensionais presumem, mas parece pouco factível que possa ser testada.

A perspectiva dimensional das emoções é ora defendida como suficiente para lidar com os estados emocionais, ora é tida como uma visão limitadora dos estudos da emoção. Lazarus (1991) defende que a restrição

das emoções a poucas dimensões é uma simplificação que reduz a experiência emotiva, deixando de fora alguns significados importantes advindos da relação entre pessoa e ambiente. A riqueza de nuances dos estados afetivos é espelhada no número elevado de termos que se tem para expressar as emoções. Segundo Lazarus (1991, p. 84), “the reduction through factor analysis of the vocabulary of emotion to a few basic dimensions [...] loses important relational meanings inherent in our rich vocabulary of emotions”.

Como se pôde notar, as pesquisas sobre emoções e seus modelos são diversos e de difícil organização.³⁵ Como dito anteriormente, os dois modelos – emoções básicas e abordagem dimensional – serão utilizados nos experimentos de percepção e serão confrontados com os resultados da análise acústica. Bachorowski e Owren (2000, p. 197) afirmam que as duas perspectivas de pesquisa compartilham duas importantes características quando buscam respostas na expressão vocal das emoções: i) ambas pressupõem que haja uma relação entre emoção e mudanças no sinal acústico; ii) ambas, ao menos de forma implícita, adotam um perspectiva representacional, no sentido de que a informação relacionada à emoção é codificada pela voz e decodificada pelos ouvintes para que estes possam avaliar o estado emocional do falante.

Neste trabalho, as emoções utilizadas foram: alegria, raiva, tristeza e medo, por serem definidas mais comumente como emoções básicas, como afirma Plutchik (2001, p. 349):

[...] over the centuries from Descartes to the present, philosophers and psychologist have proposed anywhere from 3 to 11 emotions. All the list include *fear*, *anger* and *sadness*, most include *joy*, *love* and *surprise*.

Sobre a abordagem dimensional, as dimensões utilizadas foram: *valência* – não agradável-agradável; *ativação* – não agitado-agitado;

³⁵ Nas palavras de Schlosberg (1954, p. 81): “I think you will agree that the field (of emotion) is chaotic”.

dominância – submisso-não submisso. Da mesma forma, baseando-se no estudo de Trnka (2011), ativação e valência estão presentes na maioria das propostas, à exceção da *dominância*, que muitas vezes apresenta problemas de interpretação por parte dos participantes. Esta última dimensão não poderia ser descartada *a priori*, pelo contrário, deve ser incluída e verificada juntamente com as demais. No entanto, a despeito de ser uma dimensão menos frequente na literatura, Peres (2014, 2015a) apresentou resultados significativos de *dominância* para os julgamentos de nativos e não nativos sobre a fala emotiva no PB. Portanto, somente após a análise dos resultados, será possível mensurar o papel de cada uma delas na percepção das emoções e sua interação com os parâmetros acústicos utilizados neste trabalho.

2.6 Por que fala espontânea?

A maioria dos estudos que trataram da fala expressiva fez uso de sentenças com fala atuada ou outros tipos de elicitación (SKINNER, 1935; COSTANZO; MERKEL; COSTANZO, 1969; SCHERER et al., 1973; WALLBOTT; SCHERER; 1986; SCHERER, 2000; SCHERER; BANSE; WALLBOTT, 2001). No português brasileiro, estudos com fala elicitada³⁶ foram conduzidos por Colamarco e Moraes (2008) e Vassoler e Martins (2013).

A favor do uso da fala atuada ou elicitada está o controle dos estímulos em sentenças idênticas, pronunciadas nas mais variadas emoções. Essa característica permite ao experimentador um maior controle das variáveis. Além disso, Scherer (1981, p. 204) já apontara os problemas encontrados em gravações de fala espontânea, sem intervenção direta do

³⁶ Em Colamarco e Moraes (2008), uma locutora gravou a frase “Prepara a panqueca” com 16 tipos de entoação, combinando características gramaticais e emoções. Em Vassoler e Martins (2013), três locutores gravaram um trecho de texto retirado de um trabalho sobre imunologia.

experimentador, afirmando que “naturally recorded emotions are by definition singular cases, both in terms of speaker identity, situation context, and verbal content of utterance”. Segundo o autor (SCHERER, 1981), com essas características da fala espontânea, ficaria difícil a separação de quais variáveis estão de fato agindo para configurar a fala expressiva, apresentando um problema quanto à ortogonalidade do experimento.

Por outro lado, Roberts (2011, p. 1694) demonstrou que a fala atuada pode ser fortemente impregnada de estereótipos, afirmando que esse tipo de estímulo “may merely reflect stereotypical behaviours that actors are trained to adopt”. A utilização desse tipo de fala poderia causar não só diferenças na produção, mas, provavelmente, na percepção dos estímulos. Scherer (1981) também alertou para esse problema, apontando para o uso equivocado de emoções discretas no julgamento de estímulos com fala atuada. Neste último caso, poderia haver uma dupla influência nos resultados por conta de estímulos estereotipados e da utilização de rótulos estanques para o reconhecimento das emoções.

Scherer, Banse e Wallbott (2001) utilizaram uma metodologia que tentou dar conta da variabilidade dos estímulos na fala espontânea, utilizando fala atuada elicitada por meio de cenários.³⁷ Embora os estímulos tenham sido gravados por atores profissionais, a diferença principal residiria na maneira como os atores foram instruídos a produzir as sentenças. Em geral, as falas emotivas produzidas por atores são elicitadas pelo experimentador da seguinte forma, por exemplo, “leia o texto como se estivesse triste, ou com raiva, etc.”. Segundo os autores, esse tipo de procedimento apresenta dois problemas: i) diferentes atores podem

³⁷ Os autores utilizaram dois cenários para cada uma das cinco emoções estudadas (alegria, tristeza, medo, raiva e nojo). A situação (cenário) era passada ao locutor, que deveria utilizar as *meaningless multilanguage sentences* para produzir o estado emocional sugerido pelo cenário. Como exemplo de um cenário de alegria: “My uncle arrives for my birthday. He asks me to look out of the window. In front of the house is his gift: the car I had been dreaming of for a long time” (SCHERER et al., 1991, p. 146).

atribuir significados diferentes aos rótulos fornecidos pelo experimentador; ii) os atores podem enviesar situações diferentes enquanto elicitadores das emoções (SCHERER; BANSE; WALLBOTT, 2001, p. 79).

As frases portadoras das emoções elicitadas por meio de cenários foram padronizadas, permitindo, assim, que o experimentador tenha um controle maior sobre as variáveis independentes. As *meaningless multilanguage sentences* (SCHERER; BANSE; WALLBOTT, 2001) foram formadas por duas sílabas de cada uma das seis línguas europeias presentes no estudo, a saber: alemão, inglês, francês, italiano, espanhol e dinamarquês. As sentenças, compostas por sete sílabas cada, passaram por uma avaliação de neutralidade e naturalidade.

Embora a solução tenha sido muito engenhosa, alguns problemas podem aparecer com o uso desse tipo de estímulo. O primeiro deles relaciona-se com o próprio objetivo desta tese. Se o intuito é tentar analisar qual o papel das características socioculturais, dentre elas a língua local, sentenças que não pertencem a nenhuma língua natural podem soar de forma estranha aos ouvidos dos participantes. Mesmo que o falante reconheça alguma estrutura silábica particular de sua língua materna, as demais sílabas da sentença podem não compartilhar as mesmas características. Além disso, não há como medir o seu efeito quanto à informação semântica da sentença.

Se se optasse por sentenças normalmente formadas de cada língua, mas com sentido diverso às emoções, ainda assim apareceriam outros problemas. Segundo Van Lancker e Sidtis (2008,³⁸ apud KREIMAN; SIDTIS, 2011, p. 303), a informação emocional é muito pouco expressada por fonologia, morfologia e sintaxe, dependendo quase que totalmente da escolha semântica das palavras, unida à prosódia. Do ponto de vista neurológico, Kreiman e Sidtis (2011, p. 303) afirmam que

³⁸ VAN LANCKER, D.; SIDTIS, J. J. The identification of affective-prosodic stimuli by left- and right-brain damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, v. 35, p. 963–970, 1992.

[...] processing in the brain follows this rough dichotomy, as language and thought are more cortically than subcortically represented, while speech and emotion are alike in partaking relatively more richly of subcortical nuclei.

Dessa forma, a expectativa do ouvinte é a de que o significado contido nas variações tonais e no léxico seja congruente. A desarmonia entre as partes ou a contrariedade entre os dois componentes linguísticos podem trazer problemas de percepção por parte dos ouvintes. Por outro lado, a incongruência não seria necessariamente um problema, pois poderia provocar outro tipo de percepção nos ouvintes. A ironia, por exemplo, é um uso incongruente da língua que requer habilidade do falante e do ouvinte, por isso não raras vezes esse descompasso requer tempo para ser notado pelo ouvinte. Além disso, a ironia geralmente depende do contexto de fala. O exemplo da mãe que encontra o quarto de seu filho desarrumado e diz “que beleza de quarto” só faz sentido em seu contexto de fala. Assim como a emoção, a ironia está presente desde muito cedo, como demonstra o estudo de Filippova e Astington (2010), ainda que a ironia em crianças seja reduzida se comparada aos adultos. Esse aparte sobre ironia é importante para deixar claro que a incongruência entre o suprasegmental e o segmental não tem como resultado somente o não entendimento, a confusão, mas pode desencadear outros tipos de significado, i.e., a fala incongruente pode ser uma instância comunicativa legítima. Em resumo, a utilização de estímulos com conteúdo lexical fixo, muitas vezes completamente desconectados das emoções, pode afetar os resultados de diversas maneiras não previstas.

Outro exemplo de experimento com estímulos conflitantes é o estudo de Stroop (1935),³⁹ o qual demonstrou a interferência na leitura dos nomes de cores escritos em preto e em cores conflitantes, *e.g.*, o nome da cor

³⁹ O trabalho do autor originou o chamado *Efeito Stroop*, que demonstra os efeitos da atenção seletiva em tarefas de discriminação visual.

amarela escrito em azul. Para os nomes de cores em preto, o tempo médio de leitura de 100 palavras foi de 63,3s, aumentando para 110,3s quando se tratava de nomes e cores desconexos. Os resultados demonstram como a incongruência pode influenciar a percepção.

O estudo de Nygaard e Queen (2008) encontrou resultados semelhantes ao estudar a relação das palavras relacionadas à felicidade e à tristeza e a maneira pela qual elas são produzidas. Por exemplo, quando uma prosódia 'alegre' coincidia com a palavra *lucky*, as respostas eram dadas rapidamente, já quando não havia congruência, as respostas eram mais lentas. Para as autoras, a percepção de palavras parece não ser independente da percepção da prosódia emotiva, agindo como um facilitador na seleção e no processamento das palavras (NYGAARD; QUEEN, 2008, p. 1025).

Os estudos relatados acima demonstraram que uso de fala atuada, de sentenças sem sentido e estímulos incongruentes, embora favoreçam a ortogonalidade na elaboração e análise dos experimentos, podem ser problemáticos por direcionar, de uma forma ou de outra, os resultados. Por essa razão, este estudo defende o uso de fala espontânea no sentido de expressar uma situação real de manifestação emocional e de ter uma menor influência do experimentador. Ainda que a variabilidade dos estímulos seja um problema por apresentar contextos de produção e características pessoais distintas, essas influências podem ser contornadas.

Como será visto na análise deste trabalho, as variáveis não controladas que possam desempenhar alguma influência nos resultados, *e.g.*, a escolha lexical, podem ser neutralizadas por meio de técnicas de delexicalização. A discussão sobre as técnicas de delexicalização será feita na metodologia do último experimento desta tese. Por ora, basta dizer que a delexicalização visa a impossibilitar a capacidade do sujeito de recuperar o léxico das sentenças. Para tal, os métodos de delexicalização (ou *content-masked speech*) removem as informações segmentais relevantes da fala,

tornando-a ininteligível, mas mantendo intactas as características prosódicas. Portanto, entende-se que é preferível intervir nos estímulos espontâneos para diminuir a sua variabilidade a analisar um estímulo com fala elicitada que possa, em alguns casos, estar equivocado desde o momento de sua gravação.

3. Qualidade de voz

A voz é condição necessária para que a fala humana se concretize como evento acústico e perceptual entre os usuários das línguas orais.⁴⁰ Enfatizo aqui que se trata de línguas orais, pois existem línguas que prescindem do som, mas que apresentam as demais estruturas presentes nas línguas naturais, i.e., as línguas de sinais utilizadas pelas comunidades surdas ao redor do mundo. Além de ser a substância pela qual as línguas são veiculadas, há muito na voz humana que não está ligado à língua em si e que traz informações de outra natureza. A voz carrega consigo marcas mais estáveis do falante como idade e sexo, e mais transitórias como estados de humor e emoção. Segundo Laver (1980), a voz

⁴⁰ Doravante, para efeito de concisão, as demais ocorrências do termo língua terão o significado de língua oral.

é um dos meios primários pelo qual os falantes veiculam suas identidades, suas características físicas, psicológicas e sociais. Ademais, a voz ainda veicula características indeléveis do falante (salvo a utilização de disfarces por criminosos), caracterizando-o quase como uma “face auditiva”, como em Belin, Fecteau e Bedard (2004), que propuseram um modelo de percepção de voz análoga ao proposto para a percepção da face humana. Um exemplo cotidiano é que não raras vezes, quando ouvimos vozes no rádio ou no telefone, criamos uma imagem dos falantes a partir da voz, imaginando suas características físicas, humor, idade, sexo, etc.

Em face da pluralidade encontrada no uso da voz humana, pode-se entendê-la como portadora de dois fatores: orgânicos e intencionais.⁴¹ Laver (1980) denomina esses fatores como intrínsecos e extrínsecos, respectivamente. Para Laver (1980, p. 5), o fator orgânico seria o fundo onde se projeta a figura, ou seja, o uso intencional da voz, sendo a fonética responsável por descrever todos os fenômenos padrões e recorrentes que caracterizam a língua falada de uma certa comunidade de fala.

De maneira geral, os termos *voz* e *qualidade de voz* são tomados como sinônimos, embora seja esse tipo de entendimento fruto de confusão. Kreiman e Sidtis (2011) definem a *voz* como portadora de uma realidade *física e fisiológica* que se refere ao sinal acústico, ao passo que a *qualidade de voz* está relacionada à *percepção*, ou seja, à impressão que a voz provoca nos interlocutores. O trecho a seguir de Kreiman e Sidtis (2011, p. 5) define essa divisão:

The term “voice” has a physical and physiological base that refers to the acoustic signal (as generated by the voice production system), while “voice quality” refers to the perceptual impression that occurs as a result of that signal,

⁴¹ Xu e Wang (1997) propuseram modelo semelhante para o estudo da entoação, no qual alguns fatos prosódicos são resultado de restrições orgânicas e outros de intenções expressivas dos falantes. Ferreira Netto (2006) propôs um modelo de análise da entoação do PB baseado na proposta de Xu e Wang (1997), o qual será brevemente exposto no capítulo 4.

analogous to the distinction between “frequency” (a physical property of vibration) and “pitch” (a listener’s sensation).

A divisão no estudo da voz – produção *vs* percepção – é análoga à feita entre F0 e *pitch*, a primeira relacionada ao evento acústico fruto da vibração das pregas vocais (de maneira estrita) e a segunda relacionada à impressão que o parâmetro acústico F0 causa nos ouvintes. Entretanto, a relação entre F0 e sua percepção não é biunívoca. Por conta disso, impressões de *pitch* nem sempre decorrem de variações de F0, mas de outros correlatos como, por exemplo, o padrão formântico das vogais.

Do ponto de vista da produção, a voz é produto do aparelho fonador, que gera o som por meio do fluxo de ar advindo dos pulmões. Num primeiro momento, o fluxo de ar sofre modificações das pregas vocais,⁴² passando pela laringe e pelos demais componentes supraglotais. Na perspectiva psicológica, a qualidade de voz diz respeito aos atributos perceptuais que dela são inferidos. Segundo Kreiman e Sidtis (2011), a distinção entre os parâmetros acústicos e a impressão que eles causam é muitas vezes desconsiderada em escritos informais, mas essa imprecisão pode acarretar uma falsa ideia de que os parâmetros acústicos da qualidade de voz são percebidos sempre da mesma forma pelos ouvintes (biunivocidade entre produção e percepção). Nesta tese, será respeitada a divisão entre os aspectos fisiológico-acústico e perceptual envolvidos na produção da voz e na percepção de sua qualidade. A qualidade de voz será utilizada como sinônimo das características nas quais os participantes dos experimentos possivelmente basearam seus julgamentos. Por voz entender-se-ão os parâmetros acústicos analisados neste trabalho.

⁴² Há algumas teorias que tentam explicar como o som é gerado nas pregas vocais. A teoria mioelástica-aerodinâmica (VAN DEN BERG, 1958) é a mais profícua e descreve a produção da voz por meio da interação entre as pregas vocais e a força aerodinâmica.

Além da divisão produção-percepção, o estudo da voz e da qualidade de voz pode ser feito de maneira estrita, considerando somente as ondas sonoras produzidas pela vibração das cordas vocais, ou seja, somente a *fonte* é considerada, nos termos da *Acoustic Theory of Speech Production* (FANT, 1960). Nessa abordagem, ficariam excluídas, por exemplo, as ressonâncias e as turbulências criadas pelo trato vocal.

Em uma abordagem mais abrangente, as demais características da produção da voz são consideradas como, por exemplo, os movimentos dos lábios, da mandíbula, da língua e suas respectivas constrições, ressonâncias e turbulências. Dessa forma, a *voz* pode ser entendida como sinônimo de *fala*⁴³ (KREIMAN; SIDTIS, 2011, p. 6). A mesma divisão – estrita ou abrangente – é válida para o estudo da qualidade de voz, permitindo analisar uma característica em particular ou suas características gerais.⁴⁴ Laver (1980) já havia proposto essa divisão para o estudo da qualidade de voz. Por ser tratar de uma descrição fonética da qualidade de voz, o autor baseou seu modelo numa visão ampla do fenômeno, ou seja, considerando a influência de componentes supralaríngeos.

Do ponto de vista da produção da voz, estudos que tratam dos correlatos acústicos relacionados à manifestação da emoção na fala consideram a qualidade de voz um componente importante para os estudos sobre fala emotiva (HOLLIEN, 1980; LADD et al., 1985; JOHNSTONE; SCHERER, 1999).

Segundo Kreiman e Sidtis (2011), as medidas acústicas utilizadas na análise da voz podem ser divididas em duas categorias: medidas de

⁴³ Fala - *speech*, como referida pelas autoras, está relacionada mais amplamente à *produção de fala*, do ponto de vista articulatório, sem relação com a plenitude de significação que a *fala* pressupõe.

⁴⁴ Divisão semelhante da entoação foi proposta por Hirst e Di Cristo (1998). A entoação pode ser entendida de forma estrita, ou seja, composta por modulações de F0; ou vista de forma mais abrangente, neste caso englobando os demais parâmetros como duração e amplitude (HIRST; Di CRISTO, 1998).

domínio temporal e medidas de domínio espectral ou de frequência. O primeiro tipo de medida analisa o sinal acústico do ponto de vista temporal, ou a taxa em que eventos recorrentes acontecem. Como exemplo de medida temporal, as variações de F0 estão ligadas intrinsecamente aos movimentos cíclicos das pregas vocais, sendo a taxa de abertura e fechamento responsável pelos valores de F0. Outras medidas ligadas a frequência fundamental, como o *jitter* e o *shimmer*, também são uma maneira de aferir a voz do ponto de vista temporal.

O *jitter* e o *shimmer* são medidas de perturbação de F0. O primeiro diz respeito a pequenas variações encontradas na duração do período que ocorre entre os ciclos glóticos; o segundo indica a diferença de amplitude entre os ciclos glóticos. Em outras palavras, o *jitter* diz respeito às pequenas variações de altura encontradas; o *shimmer*, às variações de amplitude. No caso do *jitter*, a perturbação ocorre por uma falta de controle de vibração das pregas vocais, já o *shimmer* é devido à redução da resistência glótica ou lesões de massa nas pregas vocais, estando correlacionada com a presença de ruído na emissão da fala e com a soprosidade (WERTZNER; SCHREIBER; AMARO, 2005). Por outro lado, essas pequenas perturbações de F0 não são exceções e conferem à produção da voz o seu caráter humano, pois sem essas pequenas variações a voz soaria mecânica e artificial (KREIMAN; SIDTIS, 2011, p. 55). Apesar de importante para o diagnóstico de falas patológicas, o *jitter* e o *shimmer* muitas vezes não são percebidos pelos ouvintes em falas normais e em fala patológica. Por razão da sua baixa influência em resultados de testes perceptivos, o *jitter* e o *shimmer* não serão utilizados neste trabalho (KREIMAN; GERRAT, 2005).

As medidas espectrais são destinadas a quantificar a energia relativa em diferentes partes do espectro. Se comparadas às medidas temporais, as medidas espectrais seriam um tipo de análise vertical da produção da voz, ao passo que as medidas de F0 seriam uma análise horizontal por levar em consideração a temporalidade inerente de sua

constituição. Como exemplo de medida espectral, há a diferença entre o primeiro harmônico (H1) e o segundo (H2), obtida pela fórmula H1-H2, bem como outros parâmetros como a ênfase *espectral* e LTAS (*Long Term Average Spectrum*).⁴⁵

3.1 H1-H2, Ênfase Espectral e LTAS

Neste estudo, optou-se por três medidas indiretas de qualidade de voz, são elas: *H1-H2* – a intensidade do segundo harmônico subtraída da do primeiro, ambas mensuradas em uma mesma vogal; ênfase *espectral* – diferença de energia entre regiões de baixa e alta frequência do espectro; e *LTAS* – inclinação do espectro em trechos de fala contínua.

A medida H1-H2, segundo Shue, Chen e Alwan (2010), é o parâmetro acústico que tem maior correlação com o quociente de abertura das pregas vocais e, conseqüentemente, com os tipos de voz decorrentes delas. Os tipos de fonação estão relacionados ao quociente de abertura das pregas vocais (*OQ – open quotient*), que pode ser definido por $OQ = T_o - T_c / T$, em que T_o é a duração da fase de abertura, T_c é a duração da fase de fechamento e T é o tempo total do ciclo glotal. Entretanto, *OQ* é uma aferição indireta da qualidade da voz que pode ser feita por eletroglotografia (EGG). A sua primeira utilização na produção de fala foi feita por Fabre (1957⁴⁶ apud FOURCIN, 2000). Por meio de eletrodos colocados nas alas da cartilagem tireóidea, a impedância elétrica é mensurada, sendo inversamente proporcional à abertura das pregas vocais. Assim, o aumento da impedância se dá quando as pregas vocais estão afastadas e a diminuição quando se encontram fechadas, tendo como resultado a onda LX. Embora seja um método não invasivo do ponto de vista clínico (quando utilizado para diagnóstico de patologias), torna-se invasivo quando utilizado em

⁴⁵ *Espectro médio de longo termo.*

⁴⁶ FABRE, P. Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation. *Bulletin de l'Academie Nationale de Medecine*, v. 141, p. 66-99, 1957.

estudos com fala emotiva espontânea. A simples colocação do equipamento já teria uma influência considerável nos resultados. Portanto, nesta tese serão feitas menções diretas à abertura e ao fechamento da glote, evitando-se referências a medidas indiretas que não foram utilizadas neste trabalho.

Sendo os tipos de fonação uma consequência do grau de abertura das pregas vocais, a aferição H1-H2 é utilizada, por exemplo, como medida de *breathiness*. Por conta dessa relação, essa medida pode ser um correlato acústico importante para a análise da manifestação de estados emotivos na fala. A seguir, há um exemplo ilustrativo de como se dá a medição de H1-H2. A medição foi feita a partir de gravação de voz masculina e adulta (35 anos). A sentença gravada foi “Eu viajei para a casa de um amigo”. A medição foi feita no [a] tônico de “casa”. As figuras abaixo são exemplos de três tipos de voz: *modal*, *creaky voice* e *breathy voice*.

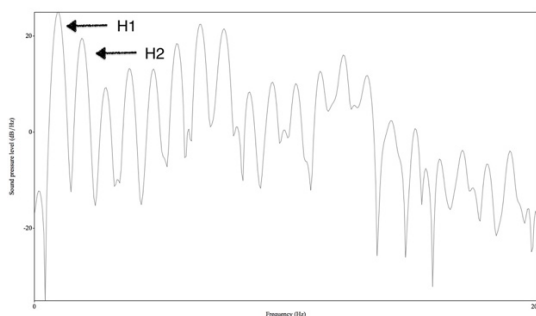


Figura 3: Espectro com indicação de H1-H2 – *modal*. As medidas foram extraídas de vogal tônica [a] produzida por falante masculino

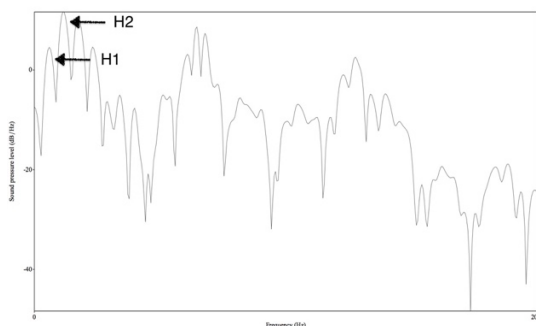


Figura 4: Espectro com indicação de H1-H2 – *creaky*. As medidas foram extraídas de vogal tônica [a] produzida por falante masculino

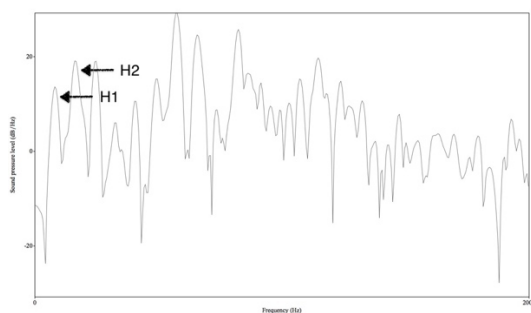


Figura 5: Espectro com indicação de H1-H2 – *breathy*. As medidas foram extraídas de vogal tônica [a] produzida por falante masculino

Nota-se que os valores de H1 e H2 modificam-se em cada um dos tipos de fonação. Nas Figuras 4 (*creaky*) e 5 (*breathy*), pode-se notar que H1 teve uma perda de energia com relação a H2 se compararmos à fala modal (Figura 3). *Creaky voice* geralmente tem baixos valores de H1-H2, porque a glote geralmente está contraída.⁴⁷ Além de H1-H2, *creaky voice* pode ser caracterizada acusticamente pela frequência fundamental baixa e irregular. *Breathy voice*, além de H1-H2, apresenta como característica a adição de ruído por conta da passagem de ar pelas pregas vocais, as quais não se encontram totalmente aduzidas durante a fase fechada do ciclo glotal.

A ênfase espectral é utilizada como correlato acústico de esforço vocal. Eriksson, Thunberg e Traunmüller (2001)⁴⁸ propuseram uma equação para sua medição, a saber: *ênfase espectral (dB) = L – L₀*, em que L é a intensidade de todo o espectro e L₀ é a intensidade obtida até 1,5 F₀.⁴⁹ Essa medida resulta da hipótese de Lehiste e Peterson (1959,⁵⁰ apud ERIKSSON; THUNBERG; TRAUNMÜLLER, 2001) sobre a percepção do

⁴⁷ Keating *et al.* (2015) demonstram que não há somente um tipo de *creaky voice*, há outros tipos. Por exemplo, há um tipo em que o valor de H1-H2 é mais alto se comparado com a voz modal.

⁴⁸ Medida também utilizada por Barbosa, Eriksson e Åkesson (2013) em estudo comparativo entre o acento lexical do português brasileiro e do sueco.

⁴⁹ Valor obtido de F₀ somado ao valor de sua metade. Ex. F₀ = 100Hz + 50Hz = 150Hz (1,5 F₀).

⁵⁰ LEHISTE, I.; PETERSON, G. E. Vowel amplitude and phonemic stress in American English. *The Journal of the Acoustical Society of America*, v. 31, p. 428-435, 1959.

acento. Segundo os autores, a percepção do acento lexical é baseada no esforço fisiológico despendido para a produção das vogais.

O LTAS é utilizado em trechos de fala contínua para se obter a inclinação média da energia do enunciado. Segundo Nordenberg e Sundberg (2003, p. 93), “LTAS displays the average sound level in different frequency bands, and reflects both glottal and vocal tract characteristics”. As regiões de maior ou menor energia do LTAS estão relacionadas à percepção da qualidade de voz. Quanto maior a duração dos enunciados, mais neutralizadas serão as influências espectrais dos segmentos produzidos nas sentenças. Os trechos de fala deste estudo não ultrapassam a duração de 20s, por isso tem-se consciência dos problemas que podem ocorrer por conta da menor variedade dos segmentos utilizados nas frases. Outro problema, de acordo com Nordenberg e Sundberg (2003), é a influência da intensidade da produção da voz no uso de LTAS. O aumento da intensidade vocal provoca um maior incremento de energia na região de 3kHz do que em 0,5 kHz, i.e., o aumento do nível LTAS não é uniforme ao longo da escala de frequências. Isso causa alguns problemas quando se trata da mesma voz gravada em diferentes ocasiões, como antes e após terapia. Para esta tese, entretanto, essa característica de LTAS não é um problema, pois não serão analisadas instâncias diferentes de um mesmo falante pronunciando frases idênticas (como justificado no subcapítulo 2.5). Entretanto, apesar dessa limitação, foram obtidos resultados interessantes no estudo piloto que será detalhado adiante. Abaixo, há quatro ilustrações de medições de LTAS para cada sentença de cada emoção estudada neste trabalho (raiva, medo, tristeza e alegria):

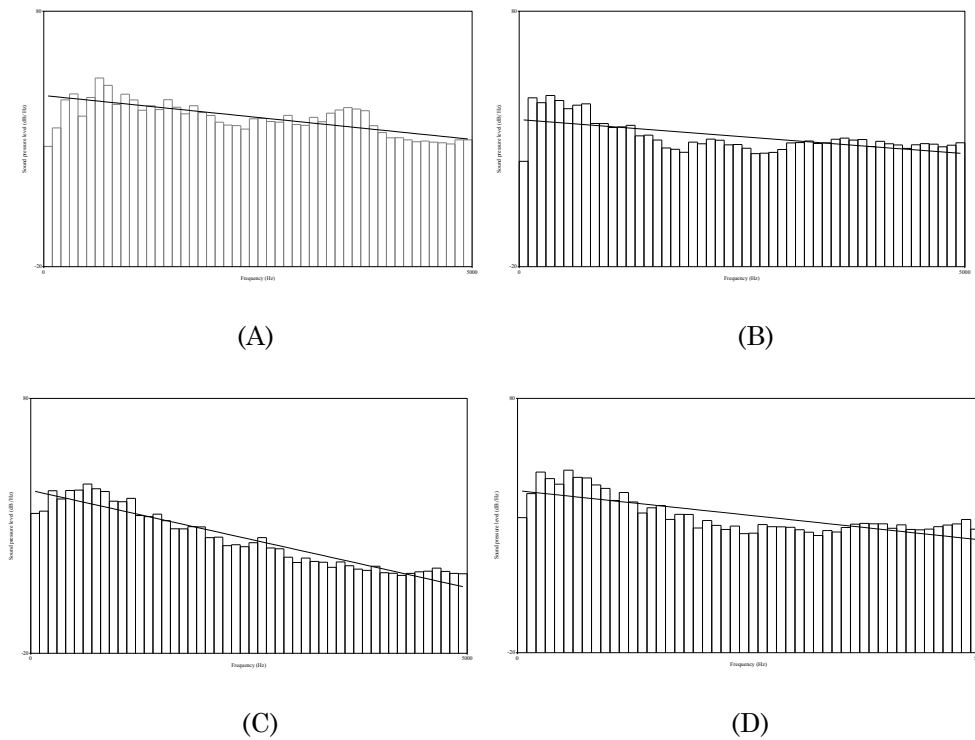


Figura 6: Medidas de LTAS em fala emotiva: (A) – raiva [-12dB/Hz]; (B) – medo [-10dB/Hz]; (C) – tristeza [-22dB/Hz] e (D) – alegria [-9dB/Hz]

Quanto maior o esforço empregado na produção da voz, maior será a energia global e em altas frequências, como se pode notar no caso da alegria (D). Por consequência, a alegria apresentou uma linha de tendência mais horizontalizada. Da mesma forma, raiva (A) e medo (B) tiveram um aumento de energia em frequências altas, enquanto a raiva apresentou uma energia global ligeiramente elevada se comparada com as demais. A tristeza (C) teve menos energia global e em altas frequências, resultando numa linha de tendência mais acentuada.

Após a exposição das medidas acústicas relacionadas à qualidade de voz, é necessário que se faça um breve paralelo entre elas e a manifestação da emoção na fala. Para os fins deste estudo, o conjunto de análises espectrais (H1-H2, ênfase espectral e LTAS) será aplicado nas amostras de fala espontânea que compõem os dois primeiros experimentos

desenvolvidos nesta tese: teste piloto com emoções básicas e análise dimensional das emoções.

3.2 Qualidade de voz e emoção

A expressão da emoção pelos falantes está relacionada principalmente à maneira como os sons da fala são produzidos pelo aparelho fonador, ou seja, é um objeto de estudo que está ligado muito mais à sua realização fonética do que a níveis mais elevados de análise linguística, como a fonologia, a morfologia e a sintaxe. Entretanto, a semântica e a escolha lexical também têm um papel importante na análise da fala emotiva. Por essa razão, o papel do léxico foi considerado nos dois primeiros experimentos desta tese. É importante dizer que não se tratou de uma análise baseada na escolha do léxico, mas no interesse de atestar o papel da língua na percepção da emoção. Tem-se, portanto, uma dicotomia entre o que é dito e como é dito. O “como é dito” é entendido como um comportamento não verbal que ocorre concomitantemente ao uso linguístico da voz. Do ponto de vista fonético, esse comportamento é o foco da análise quando se trata de fala emotiva.

Os parâmetros utilizados para a análise são diversos e abarcam os dois domínios mencionados anteriormente: o temporal e o espectral (ou de frequência). Em estudo perceptual, Costanzo, Merkel e Costanzo (1969) demonstraram que as variações vocais e, por consequência, os seus correlatos acústicos são relacionados às variações na expressão da emoção. Van Lacker e Sidtis (1992) analisaram o papel dos parâmetros acústicos na classificação das emoções e encontrou uma alta correlação entre a média e as variações de F0 e amplitude. Scherer (1991) também analisou a fala de atores pronunciando uma sentença padrão. A leitura das sentenças foi baseada em cenários (ver subcapítulo 2.5) que representavam quatro emoções (alegria, tristeza, raiva e medo) e estado neutro. Após a análise acústica, a variação e a média de intensidade, F0, energia em altas

frequências do espectro e taxa de elocução variaram de acordo com cada emoção elicitada.

Hammerschmidt e Jürgens (2007) analisaram seis variações emocionais na pronúncia da palavra “Anna”. O estudo apresentou 94 parâmetros acústicos, que foram reduzidos a 15 por meio de análise de correlação. Os autores julgaram que uma análise mais detalhada do ponto de vista acústico teria um poder maior de descrição, o que *a priori* não é necessariamente verdadeiro. Os resultados demonstraram que cada estado emocional pode ser caracterizado por um perfil acústico específico. Com apenas seis parâmetros, a taxa de acerto foi de 77,4%, valor elevado se comparado a outros estudos como, por exemplo, o de Banse e Scherer (1996), o qual obteve apenas 48% de acerto. Além disso, o estudo de Hammerschmidt e Jürgens (2007) teve êxito em dividir as emoções em dois grupos: *hedonistic* e *aversive emotions*. Os parâmetros acústicos importantes para essa divisão são os picos de frequência, porcentagem de tempo de segmentos não harmônicos, a variação de frequência de cada segmento e o tempo máximo do pico de frequência em cada palavra.

Willians e Stevens (1972) analisaram três emoções (tristeza, raiva e medo) e contraste com estado neutro na fala de atores. De acordo com os resultados, a raiva apresentou valores elevados de F1, provavelmente devido ao abaixamento acentuado do maxilar. A análise de LTAS demonstrou que a raiva tem intensidade elevada em frequências maiores que 1kHz, ao passo que a tristeza apresenta valores mais baixos. Segundo os autores, as emoções estudadas apresentam as seguintes características acústicas: a) raiva: F0 elevado em quase uma oitava com relação à fala neutra, a variação de F0 é maior, maior energia em altas frequências do espectro; b) medo: F0 mais baixo se comparado à fala neutra e articulação mais precisa dos segmentos; c) tristeza: pouca variação de F0, articulação mais lenta e com pausas longas, menor energia em altas frequências do

espectro e irregularidade da voz (WILLIAMS; STEVENS, 1972, p. 1248-1249).

Kreiman e Sidtis (2001, p. 321) elencaram um conjunto de parâmetros acústicos e como eles são geralmente associados às emoções. No caso do F0 médio, há um aumento dos valores para o medo e alegria, mas os valores decaem quando se trata de tristeza. A dispersão de F0 é menor na tristeza, entretanto é maior no caso da raiva e da alegria. A curva espectral acentuada, i.e., menos energia nas frequências altas do espectro, é associada à tristeza, ao passo que a raiva e a alegria apresentam um aumento de energia nessa região do espectro. O medo apresenta duas possibilidades, podendo haver aumento ou diminuição de energia em alta frequência.

Outros estudos trataram exaustivamente da relação e adequação dos parâmetros acústicos utilizados nas análises e, em sua maioria, lidaram com a relação entre os correlatos acústicos e a percepção decorrente deles. Por outro lado, não há muitas tentativas de discutir as três perspectivas que estão intimamente ligadas à fala emotiva e à fonética de maneira geral: articulação dos sons, correlatos acústicos e percepção. Apesar de não ser o escopo desta tese, tem-se ciência de que as mudanças corporais provocadas por estados emocionais podem interferir na produção e, por consequência, na acústica e na percepção da fala. Até mesmo alterações fisiológicas que, num primeiro momento, parecem não ter relação com a fala podem exercer influência, como, por exemplo, os efeitos da frequência dos batimentos cardíacos sobre F0 (ORLIKOFF; BAKEN, 1989). Segundo Sundberg (1982), o padrão respiratório pode ser influenciado pelas emoções, que por sua vez terá resultado na pressão subglotal. O aumento dessa pressão tem consequências na intensidade e até na frequência fundamental dos sons da fala. De acordo com o autor, a raiva pode ser associada hipoteticamente ao aumento de pressão subglotal e ao aumento da adução dos músculos. Se isso de fato acontece, a glote

fecharia mais rapidamente em cada ciclo e os harmônicos do espectro se tornariam mais intensos. No caso da tristeza, a diminuição da atividade muscular deve ser a razão do efeito contrário causado na voz (SUNDBERG, 1982, p. 140). Sundberg (1982) também menciona a influência que a hipossalivação (sintoma presente em alguns estados emotivos) exerce na articulação dos sons. Considerando essas mudanças corporais inerentes a estados emocionais, é importante que um modelo de análise da qualidade de voz trate os diferentes modos de fonação como sendo variações de um padrão neutro (LAVER, 1980; FÓNAGY, 1976).⁵¹

Laver (1980) propôs um modelo fonético de descrição da qualidade de voz, analisando desde as características articulatórias e passando pelos atributos acústicos e perceptuais. Para tal, Laver formulou um modelo componencial de análise da qualidade de voz baseado em configurações (*settings*) controladas independentemente. As configurações são divididas em supralaríngeas e fonatórias. A primeira é referente às modificações que o trato vocal pode sofrer no eixo longitudinal, latitudinal, além de modificações velofaríngeas; a segunda diz respeito aos modos de fonação, dentre eles *modal*, *falsetto*, *whisper*, *creak*, *harshness* e *breathiness*. As configurações fonatórias podem ser combinadas, como, por exemplo, *whispery falsetto*, *whispery creaky falsetto*, *harsh whispery creaky falsetto*.

Ao deixar a análise holística de lado, Laver (1980) defende que seu modelo deve ser científico e não se basear numa descrição impressionística, a qual seria idiossincrática por conta da influência individual do pesquisador. O sistema de análise, entretanto, é baseado em análise oitiva, o que parece ser uma contradição num primeiro momento. Entretanto, segundo o autor, as análises impressionísticas podem ser confrontadas com outros três níveis de análise: articulatorio, fisiológico e acústico. Além

⁵¹ De acordo com Laver (1980, p. 14), a configuração neutra não pode ser confundida com uma suposta configuração 'normal' ou posição de repouso do aparelho fonador. Pelo contrário, deve ser entendida como a maneira habitual com que o falante se expressa, sem a influência de estados emocionais.

disso, Laver deixa claro que a análise descritiva de uma voz deve ser passível de ser comunicada por meio da escrita (1980, p. 7), o que parece novamente contraditório se levarmos em conta a afirmação do autor de confrontar dados impressionísticos com sua realização acústica.

O modelo de Laver (1980) é bastante profícuo,⁵² mas tem limitações quanto à sua aplicação em estudos de fala emotiva, pois não foi desenvolvido para esse fim. Embora reconheça que as configurações propostas por ele possam ser aplicadas em análises paralinguísticas, muito pouca atenção é dada às influências que estados emotivos podem causar na qualidade de voz. Algumas poucas associações entre qualidade de voz e afetos foram feitas por Laver (1980), como, por exemplo, *breathy voice* e intimidade, *whispery voice* e confidencialidade, *harsh voice* e raiva e *creaky voice* e tédio.

Fónagy (1962,⁵³ apud Sundberg, 1982) desenvolveu estudo sobre o funcionamento glotal na produção de fala emotiva, utilizando dois métodos de medida: raio-x e a utilização de espelho para visualizar a glote. Foram analisados três falantes articulando a vogal [i] diversas vezes em sete estados emocionais diferentes. Fónagy (1962) descreveu os tipos de vozes que foram produzidas da seguinte forma: 1) *soft voiced*; 2) *soft unvoiced*; 3) *hard and unvoiced, spiteful whisper*; 4) *hard and hateful, voiced, pressed phonation, creaky voice*; 5) *contemptuous growl*. Em estudo posterior, Fónagy (1976) analisou os padrões de movimentos da língua e abertura da boca na produção de sentenças em húngaro pronunciadas em seis estados emocionais (*anger, hate, sadness, joy, tenderness* e *irony*). De maneira geral, a raiva foi associada à movimentação abrupta da musculatura entre posições articulatórias extremas, ao passo que o oposto foi observado em *tenderness*.

Como se pôde notar, os estudos sobre qualidade de voz dependem de

⁵² O resultado do seu modelo é, além de uma descrição apurada das configurações da qualidade de voz, uma lista de símbolos que podem ser usados em transcrições fonéticas.

⁵³ FÓNAGY, I. Mimik auf glottaler Ebene. *Phonetica*, v. 8, p. 209-219, 1962.

uma rotulação impressionística, mesmo quando baseados nas configurações do aparelho fonador, afinal de contas, trata-se do que foi percebido pelos ouvintes. De acordo com Gobl e Ní Chasaide (2003, p. 192), o problema de rótulos impressionísticos é que eles podem significar coisas diferentes para diferentes pesquisadores. Dessa forma, tem-se um rótulo, *e.g.*, *creak*, que pode referir-se a fenômenos distintos, por outro lado pode-se ter o problema inverso, ou seja, o de haver diferentes rótulos para descrever fenômenos muito semelhantes. A confirmação desses problemas está no estudo de Keating *et al.* (2015) sobre diferentes tipos de *creaky voice* (nota de rodapé 47, pág. 83).

Neste estudo, na medida do possível, serão evitados os rótulos supracitados sobre as diferentes qualidades de voz (*modal, creak, harsh, etc.*). Não há, entretanto, restrições ao uso dos nomes tradicionais encontrados na literatura, por ser provada a sua utilidade ao longo da literatura sobre o assunto. Na realidade, o que se busca nesta tese não é o atrelamento entre as qualidades de voz e as emoções, embora haja trabalhos que façam referência aos rótulos propostos por Laver (1980) e emoções correlatas. O problema neste trabalho é que não há uma descrição impressionística dos estímulos utilizados nos experimentos que siga essa nomenclatura. Assim, na análise dos dados, será feita a confrontação entre as emoções contidas nos estímulos e o seus correlatos acústicos, que podem ser eventualmente correlatos das qualidades de voz supracitadas. Entretanto, não decorre daí uma ligação entre a emoção e a presença de determinada qualidade de voz. Em outras palavras, o que se analisará será o papel dos três correlatos acústicos presentes na voz (H1-H2, ênfase espectral e LTAS) que hipoteticamente sejam importantes para a percepção da emoção na fala.

4. ENTOAÇÃO

O termo entoação vem do latim *intonāre*, que tem como significado primeiro trovejar, atroar. Em sentido figurado, o termo tem o significado de retumbar e ressoar. Outro sentido, esse mais próximo do uso atual, seria o de falar com força, gritar de forma ruidosa. Para o português, a tradução mais próxima seria *entoar*, que tem como primeira acepção o significado de “emitir uma melodia com a voz”. Dentre outras acepções, também há a de “falar em voz alta”. Esses dois significados fazem bastante sentido quando estudamos a trajetória dos estudos entoacionais ao longo dos séculos, dada a relação estreita entre fala e música.

Antes de traçar um breve histórico da evolução dos estudos sobre a entoação, é necessário um parêntese sobre como será entendida a entoação nesta tese. É frequente o uso do termo entoação como sinônimo de

prosódia,⁵⁴ possivelmente por conta da ambiguidade no próprio uso do termo entoação. Por um lado, o termo abarcaria, além das características suprasegmentais ou pós-lexicais, fatores como o acento lexical, o tom e a quantidade, partes essenciais do léxico. Esse significado mais amplo de entoação pode ser entendido como a própria prosódia, reservando-se ao termo entoação somente o seu significado mais estrito, ou seja, somente as características suprasegmentais (HIRST; DI CRISTO, 1998, p. 4). Uma outra ambiguidade reside na distinção entre o nível físico (parâmetros físicos observáveis e mensuráveis) e o nível formal (representação abstrata). Do ponto de vista físico, a entoação se refere à variação de parâmetros acústicos que a compõem, dentre eles F0, entendido como um parâmetro primário na formação da entoação. Entretanto, a intensidade e a duração podem fazer parte dos parâmetros acústicos da entoação. O nível formal ou abstrato diz respeito à formulação de modelos que possam explicar e descrever os dados entoacionais (HIRST; DI CRISTO, 1998, p. 4). Nesta tese, a entoação será analisada no sentido estrito e físico, ou seja, somente serão levadas em conta características suprasegmentais e físicas da entoação, somando-se a elas a análise perceptiva dos dados.

Os estudos sobre entoação datam de mais de 2000 anos e apresentam como característica o paralelismo entre fala e música, buscando o que há de comum e diferente entre elas. Em 350 a.C., Aristóxenes de Tarento diferenciou a fala da música ao caracterizar a primeira como movimento tonal contínuo e a segunda como movimento tonal estacionário (MACRAN, 1902,⁵⁵ apud ROOT, 1930). Essa divisão se tornou muito influente nos estudos posteriores.

Dionísio de Halicarnasso (I a.C.) associou música e fala, afirmando

⁵⁴ Do grego *προσῳδία* - *προσ-* junto / *ᾠδία*- canto. A estreita relação com a música parece ser um fator determinante para tornar os termos entoação e prosódia intercambiáveis em alguns estudos.

⁵⁵ MACRAN, H. S. *The Harmonies of Aristoxenus*. Oxford: Clarendon Press, 1902.

que elas diferem apenas em grau e não em tipo (ROBERTS, 1910,⁵⁶ apud ROOT, 1930). Segundo Dionísio, a melodia da língua falada pode ser analisada em intervalo de quinta.⁵⁷ No século XVIII (1775), Steele (1969), também baseado na análise musical, apontara as semelhanças entre as variações tonais da fala e os contornos melódicos encontrados na música. Em sua proposta, Steele utilizou o quarto de tom para marcar as variações tonais da fala. Avançando ao século XIX, Weber (1987) analisou os diversos sons da natureza, dentre eles a fala humana, a partir do ponto de vista musical. Para ele, um ouvido cuidadoso pode notar que tais sons são melódiosos e harmônicos e correspondem a intervalos e acordes. Uma passagem de seu texto é interessante menos pelo seu rigor metodológico do que pela capacidade ilustrativa:

The human voice in speaking uses also these intervals foremost, but it moves also over most of the other intervals in melodious and harmonious combinations. We speak in melodies and harmonies, improvising them by the impulse of our thoughts and feelings over an extent or compass of one and a half to two octaves; as every plant grows with a certain color, so every sentence is spoken in some melody which rises in sympathy with the sense and sentiment of the words, giving character to the whole sentence; and from the quality and accent of this musical investment, the truth and sincerity of the words may be felt, and the character of the speaker be traced. (WEBER, 1887, p. 5)

Já no século XX, Wilson (1927, p. 356), seguindo a mesma ideia de Dionísio de Halicarnasso, refere-se à variação tonal da música como portadora de delimitações, i.e., as variações tonais possuem níveis estanques, ao passo que na fala não obedecemos a esses limites. Para ela, o tom de voz move-se continuamente enquanto falamos.

Diferentemente dessas abordagens impressionísticas baseadas nos fundamentos musicais, os estudos sobre entoação começaram a ter caráter

⁵⁶ ROBERTS, W. R. *Dionysius of Halicarnassus*. London: Macmillan, 1910.

⁵⁷ Dionísio de Halicarnasso também tratou da melodia de uma única palavra que, segundo ele, apresenta mudanças constantes com movimentos de subida e descida. Parece aqui que há uma sugestão do que seria o acento lexical.

experimental no século XX, sobretudo o estudo de Weaver (1924a, 1924b, 1924c). O pesquisador foi pioneiro ao estabelecer medidas objetivas para descrever as variações da fala. Em seu estudo, Weaver (1924a, p. 47) elenca os parâmetros obtidos para análise da fala a partir de gravações de leitura expressiva. São eles: *pitch span* – número de vibrações entre o mais baixo e o mais alto valor de *pitch* de uma melodia; *total deviation* – soma dos desvios com relação à média; *total fluctuation* – a soma das variações tonais; *curve area* – área formada abaixo da curva entoacional. Esses parâmetros somados descrevem a quantidade de inflexão na voz. Por fim, Weaver propôs uma *average normal pitch* para analisar o tom médio dos falantes. Os parâmetros acima serviram para distinguir vozes masculinas e femininas.

Outros estudos como o de Skinner (1935) analisaram a fala do ponto de vista da sua variação tonal e o efeito ocasionado por estados emocionais. Como resultado, Skinner encontrou relação entre valores mais baixos de F0 e tristeza, e valores mais altos e alegria. A mesma característica foi encontrada para a intensidade, apresentado valores menores para tristeza e maiores para alegria.

Os estudos experimentais proporcionaram ferramentas importantes para análise da entoação, no entanto ainda pouco se sabia sobre a interação entre a realidade física da entoação e sua contraparte psicológica. Com esse propósito, Cohen e 't Hart (1967) propuseram um modelo que pudesse refletir o que de fato é importante na entoação, baseando-se na convicção de que a redução da entoação ao estudo do *pitch* seria mais prolífica. O modelo propõe um método perceptual que estabelece os movimentos de *pitch* relevantes para os ouvintes, além de ser possível sistematizar um conjunto limitado de padrões os contornos entoacionais a partir de regras subjacentes passíveis de verificação experimental. Nota-se, nessa abordagem, o interesse no que há de recorrente na entoação. Como resultado, estudos posteriores propuseram formas de estilização de F0, que

são basicamente métodos que possibilitam reduzir a curva melódica da fala em uma série de eventos relevantes para a percepção da entoação, como os modelos de 't Hart, Collier e Cohen (1990) e Hirst e Spesser (1993). Ambos os modelos assumem que as variações de F0 são de duas naturezas distintas: involuntárias e voluntárias. A variação involuntária é chamada de microentoação, a qual depende mormente de características fisiológicas do falante e eventuais influências dos segmentos do enunciado. A contraparte voluntária é chamada de macroentoação e é formada movimentos intencionais e perceptualmente relevantes para a configuração da entoação. Apesar de muito importante para os estudos entoacionais do ponto de vista linguístico, essa perspectiva inviabiliza, ou pelo menos limita, a análise da fala emotiva, pois, eliminar a influência da microentoação pode acarretar uma queda significativa na percepção de estados emocionais.

Lieberman e Michaels (1962) demonstraram em seu estudo que a estilização de F0 traz consequências à percepção das emoções. Os pesquisadores apresentaram a um grupo de ouvintes oito sentenças contendo oito modos emocionais⁵⁸ (*emotional modes*). O índice de acerto foi de 85% para as sentenças sem manipulação. O segundo passo foi delexicalizar as frases, extraindo as curvas de F0 e sintetizando as sentenças com uma vogal fixa. O resultado foi uma queda do índice de acerto para 44%. O índice continuou caindo quando os pesquisadores estilizaram a curva de F0 utilizando uma janela de 40ms e 100ms. Os índices caíram para 38% e 25%, respectivamente. Os resultados foram

⁵⁸ Bänziger e Scherer (2005, p. 255) entendem que os modos emocionais utilizados por Lieberman e Michaels (1962) seriam mais atitudes interpessoais do que emoções (*fullblown emotions*). O termo *fullblown* diz respeito às emoções em sua plenitude, ou seja, que são totalmente desenvolvidas, em oposição a *underlying emotions*. Os modos emocionais são: (1) *bored statement*, (2) *confidential communication*, (3) *question expressing disbelief or doubt*, (4) *message expressing fear*, (5) *message expressing happiness*, (6) *objective question*, (7) *objective statement* e (8) *pompous statement* (LIEBERMAN; MICHAELS, 1962, p. 922).

ainda menores em frases monotonizadas: somente 14% de acerto. Pode-se tirar duas conclusões desse experimento: i) as variações F0 são importantes para o reconhecimento das emoções, mesmo quando isoladas; ii) a microentoação, ou microperturbação, parece ser importante para o índice de reconhecimento. Não obstante, modelos entoacionais para análise das emoções foram utilizados por Mozziconacci (1998, 2000), especificamente o modelo proposto por 't Hart, Collier e Cohen (1990), a saber o modelo IPO (Institute for Perception Research). O modelo IPO assume a divisão da entoação em movimentos voluntários e involuntários de F0. Segundo Mozziconacci (1998, 2000), o uso do modelo entoacional aumenta o grau de controle dos parâmetros para análise e torna viável o modelamento da expressão da emoção na fala, além de aferir a sua validade.

Segundo estudo de Bänziger e Scherer (2005), trabalhos anteriores analisaram a entoação da fala emotiva em busca de padrões entoacionais específicos que pudessem ser atribuídos às emoções. Esse tipo de abordagem é baseado em modelos linguísticos de análise que, embora sejam profícuos para o estudo dos padrões entoacionais das línguas, ainda carecem de maiores evidências no campo da expressão da emoção. Em contraponto à análise da curva entoacional, há estudos empíricos sobre o papel da entoação na expressão das emoções. São estudos quantitativos que descrevem, em sua maioria, a influência dos descritores globais de F0, como a frequência fundamental média, sua variação e intensidade. O objetivo do estudo de Bänziger e Scherer (2005) é analisar os contornos entoacionais específicos de cada emoção. Embora os autores sejam críticos quanto ao uso de modelos entoacionais na análise das emoções, defendem uma análise quantitativa dos contornos entoacionais estilizados. Sentenças sem palavras reais (*meaningless sentences*) foram estilizadas a partir de 10 pontos demarcados na curva original que correspondem aos valores mínimos e máximos de F0. O modelo assemelha-se ao IPO ('t HART; COLLIER; COHEN, 1990) por utilizar segmentos de reta para unir os

pontos da curva de F0.⁵⁹ As emoções elicitadas foram combinadas com níveis de *arousal*, e.g., raiva + *high arousal* (*HA anger*), raiva + *low arousal* (*LA anger*), etc. A partir da análise de 144 sentenças, os autores chegaram à conclusão de que os níveis de *arousal* subjacentes às emoções elicitadas afetam valores globais e a variação das curvas entoacionais. Entretanto, as características contidas na curva entoacional apresentaram diferenças no *accent* das frases. O *accent* tende a ser maior em emoções como *anger* (*HA* ou *LA*) e menores em *sadness* (*LA*) e alegria (*HA*). A forma da curva entoacional foi muito pouco afetada pelas emoções, além disso deve-se ressaltar que as sentenças não continham informações sintáticas e semânticas, dando liberdade para os atores escolherem os contornos que julgassem mais apropriados para cada emoção. Segundo os autores, a falta de sistematicidade na produção de curvas entoacionais específicas para as emoções parece indicar que emoções consideradas fora do contexto linguístico não fornecem outros parâmetros senão alterações globais de F0.

O trabalho de Pakosz (1983), baseado na revisão de estudos psicolinguísticos, examina a aparente contradição entre a falta de correspondência direta entre entonação e emoção, e o reconhecimento significativo das emoções baseado em pistas prosódicas. Uma vez que o reconhecimento das emoções ocorre mesmo quando as informações de F0 são reduzidas ou ausentes do sinal de voz, pode-se supor que a entonação é irrelevante e secundária para o reconhecimento dos estados emocionais. Por outro lado, os trabalhos de Uldall (1960) e Fónagy (1978) demonstraram que os falantes reconhecem de forma consistente estímulos baseados apenas na informação entoacional (PAKOSZ, 1982, p. 322). De fato, como veremos nos resultados desta tese, a entonação é um fator

⁵⁹ O modelo de estilização da curva de F0 – Momel (*Modelling melody*) –, proposto por Hirst e Spesser (1993), segue o mesmo pressuposto básico do modelo IPO, o de que a entonação é formada por microvariações e macrovariações de F0. A diferença principal entre os modelos é a forma como os pontos-alvo são interligados. No modelo IPO são segmentos de reta; no Momel são parábolas.

relevante, mesmo quando isolada das demais características do sinal da fala. Entretanto, não se pode negar a melhora no desempenho dos participantes quando outras informações estão disponíveis, como, por exemplo, a qualidade de voz e o léxico. A discussão sobre o papel de cada um desses parâmetros será feita em detalhe nos resultados dos experimentos.

Após essa breve exposição sobre a entoação, suas características e trabalhos relevantes na análise da emoção, segue a descrição do modelo de análise a ser utilizado neste trabalho, a saber, o ExProsodia. Dentre as justificativas da escolha desse modelo, está a da possibilidade de replicar a análise com exatamente a mesma metodologia, pois se trata de uma análise automática da entoação. Outra motivação reside no número de parâmetros e na riqueza de detalhes que o modelo fornece como resultado. Trabalhos como Ferreira Netto *et al.* (2013), Ferreira Netto, Martins e Vieira (2014), Ferreira Netto *et al.* (2014) e Peres (2013, 2014, 2015a, 2015b) já obtiveram resultados significativos na análise da fala emotiva do PB.

4.1 ExProsodia

A proposta de análise de F0 realizada pelo aplicativo ExProsodia® (FERREIRA NETTO, 2010) considera a entoação como uma sequência de tons, iguais ou diferentes, produzidos pela voz durante a fala. A análise baseia-se na hipótese de Ferreira Netto (2006, 2008) de que a entoação pode ser entendida como uma série temporal, ou seja, um conjunto de observações sequenciadas e interdependentes, sendo o resultado da observação feita no momento $t+1$ condicionada ao momento t . Em outras palavras, o valor observado a cada momento de medição depende do valor do momento imediatamente anterior, não permitindo inversão dos valores observados. Assim, a ordem de ocorrência de cada momento deve ser mantida como uma informação inerente ao valor obtido.

A análise com ExProsodia está baseada na proposta de Xu e Wang

(1997) de que alguns fatos prosódicos são decorrências de restrições *mecânico-fisiológicas* (componentes estruturadoras) e outros configuram as *intenções expressivas* dos falantes (componentes semântico-funcionais). As intenções expressivas seriam as variações intencionais de F0 com fins linguísticos, i.e., alterações na onda sonora que configuram a intenção de imprimir significados diversos à fala. As restrições mecânico-fisiológicas seriam condicionadas pelas características fisiológicas do falante. As componentes estruturadoras da frequência fundamental são decorrentes do esforço exigido para a produção da fala e para a sustentação da voz com uma frequência relativamente estável, definida por Ferreira Netto (2006, 2008) como tom médio ideal (TM) de F0. Na proposta de Ferreira Netto, as restrições mecânico-fisiológicas são chamadas de *ritmo tonal*.

O ritmo tonal é formado por duas componentes, a saber: finalização (F) e sustentação (S). A sustentação (S) é o esforço despendido durante a produção da fala para compensar a declinação pontual de finalização (F), que é definida pela tendência natural da perda de esforço durante a produção da fala, que exige a retomada da tensão inicial.

Assim, o ritmo tonal é consequência da ação dessas duas tendências que atuam em sentidos opostos, possibilitando a produção da fala. A figura abaixo ilustra a relação entre as componentes – F e S – formadoras do ritmo tonal:

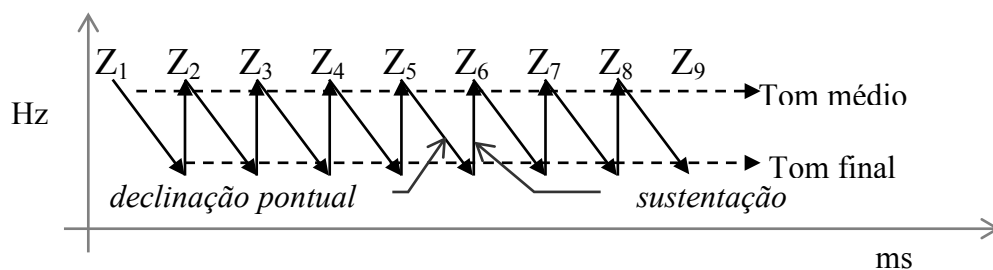


Figura 7: Representação da relação entre sustentação (S) e finalização (F) na formação do ritmo tonal

A entoação tida como significativa é formada pela componente semântico-funcional, formada por duas componentes, a saber: foco/ênfase (E) e acento lexical (A). Para o português do Brasil, desconsiderou-se a componente A, por conta do papel secundário que F0 desempenha na marcação do acento lexical. Embora a variação de F0 em torno do acento lexical seja um parâmetro presente nas análises, como visto em Eriksson, Thunberg e Traunmüller (2001) e Barbosa, Eriksson e Åkesson (2013), é necessário pensar quão importante F0 é para o acento lexical. Se analisarmos outros parâmetros como duração e ênfase espectral, é razoável afirmar que F0 desempenha um papel menor na marcação do acento. Os trabalhos de Llisterri *et al.* (2003) e Consoni (2006) demonstraram que F0 tem uma importância menor na marcação do acento quando isolado. Por essa razão, por ora, o modelo de análise em questão deixará de lado a componente A, ao menos para o PB.

Entende-se por entoação significativa as variações de F0 que extrapolam os limites do ritmo tonal, em outras palavras, os movimentos de F0 que são empreendidos pelo falante além do esforço natural requerido para a produção da fala. Em suma, a subdivisão das componentes é dada da seguinte maneira: *ritmo tonal* — finalização (F) e sustentação (S) —, *semântico-funcional* — foco/ênfase (E) e acento lexical (A). A figura a seguir ilustra essa divisão:

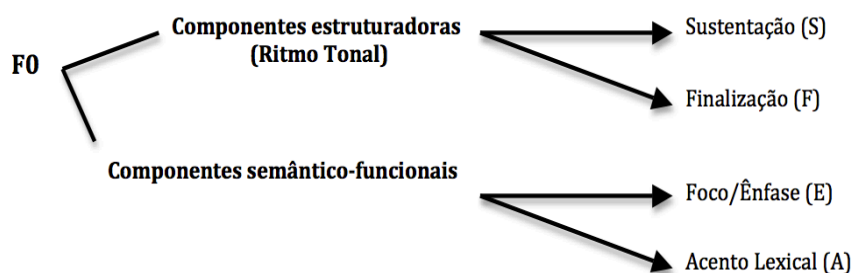


Figura 8: Esquema das componentes formadoras da entoação

Para poder obter cada uma das componentes, é necessário que a curva de F0, tida como uma série temporal, seja decomposta: a série temporal proposta constitui-se pela adição $Z(t)=S(t)+F(t)+E(t)+A(t)$. Assim, é possível decompor os momentos $Z(t)$ a partir de $TM(t-1)$ e definindo-se $F(t)$ como $TM(t-1-7)$ st. Para a definição de $E(t)$ com $Z(t)$ maior do que $TM(t-1)$ extrai-se $TM(t-1+3)$ st de $Z(t-1)$; para a definição de $E(t)$ com $Z(t)$ menor do que $TM(t-1)$, extrai-se $F(t-1-4)$ st de $Z(t-1)$. Para a definição de $S(t)$, ou extraem-se $F(t)$ e $E(t)$ de $Z(t)$, ou extrai-se 7st de $TM(t-1)$. Os limites estabelecidos aqui são fixos e frutos de resultados de experimentos de percepção desenvolvidos nos trabalhos de Consoni (2011), Peres, Consoni e Ferreira Netto (2011), Consoni e Ferreira Netto (2008, 2016) e Lassak (2016). Os limites estabelecidos se provaram adequados para análise do português brasileiro.⁶⁰

A seleção das unidades $Z(t)$ — ou *unidades entoacionais* — é feita pelo *software* ExProsodia, que faz a análise automática de trechos de fala a partir de dados de F0 extraídos por autocorrelação pelo *software* *Speech Filling System* (HUCKVALE, 2008). A mensuração de F0 feita pelo ExProsodia é tomada a cada cinco milissegundos, levando em consideração informações predeterminadas pelo pesquisador. As informações requeridas são medidas de F0 (Hz) e intensidade (RMS). O pesquisador também pode colocar limites de duração (ms). A combinação desses parâmetros formará uma *unidade entoacional*, a qual será considerada a porção sonora a ser analisada. Abaixo estão os parâmetros utilizados para as análises feitas neste trabalho:

- *Limiar inferior de frequência fundamental (F0): 50Hz.*
- *Limiar superior de frequência fundamental (F0): 350Hz.*
- *Limiar inferior de duração: 20ms.*
- *Limiar superior de duração: 300ms.*

⁶⁰ Para uma visão mais abrangente do funcionamento e dos resultados do ExProsodia, ver Ferreira Netto (2016).

- *Limiar de intensidade: 2000RMS.*

A análise automática tem como resultado parâmetros acústicos que podem ser utilizados na análise entoacional da fala. Estudos prévios como os de Vassoler e Martins (2013), Ferreira Netto *et al.* (2013) e Peres (2013, 2014, 2015a) utilizaram alguns desses parâmetros com o intuito de analisar a emoção da fala. No caso deste estudo, outros parâmetros foram selecionados e serão apresentados na próxima seção.

4.2 Parâmetros acústicos via análise automática

Os parâmetros acústicos obtidos por meio da análise automática – ExProsodia – referem-se a diferentes medidas de F0 e duração.

Os parâmetros relacionados a F0 são:

- ❖ *Tom médio (TM)* – Medida de tendência central dos valores válidos de F0 (porções de frequência fundamental selecionadas a partir da combinação dos parâmetros de análise estipulados pelo pesquisador) calculada como a média aritmética acumulada no tempo.
- ❖ *Desvio padrão do tom médio (dpTM).*
- ❖ *Coefficiente de variação de (cvTM).*
- ❖ *Assimetria do TM (α TM).*
- ❖ *Diferença entre o tom médio e a menor unidade entoacional (TMmUE)* – A unidade entoacional, como dito acima, é uma porção sonora selecionada pelo *software* a partir de parâmetros predeterminados.
- ❖ *Menor valor de uma unidade entoacional (mUE)* – É o menor valor (Hz) dentre as unidades entoacionais selecionadas pelo *software*.

- ❖ *Coefficiente de variação dos valores positivos de Foco/Ênfase ($cv_{pos}F/E$)* – Os valores positivos da componente Foco/Ênfase são aqueles que ocorreram acima do limite superior do tom médio.
- ❖ *Coefficiente de variação dos valores negativos de Foco/Ênfase ($cv_{neg}F/E$)* – Os valores negativos da componente Foco/Ênfase são aqueles que ocorreram abaixo do limite inferior do tom médio.

Os parâmetros relacionados à duração são:

- *Duração das unidades entoacionais ($durUE$)* – Duração em milissegundos das unidades selecionadas pelo *software* a partir da combinação dos valores predeterminados pelo pesquisador.
- *Desvio padrão da duração das unidades entoacionais ($dpdurUE$)*.
- *Duração dos intervalos entre as unidades entoacionais ($durintUE$)* – Duração em milissegundos entre as unidades selecionadas pelo *software*.
- *Desvio padrão da duração dos intervalos entre as unidades entoacionais ($dpdurintUE$)*.

5. Teste piloto com nativos e não nativos (emoções básicas) – qualidade de voz e entoação

5.1 Metodologia

Os estímulos utilizados neste estudo foram coletados da internet, no *site* de vídeos www.youtube.com. Foram escolhidos 32 trechos de fala espontânea do português brasileiro considerados como fala emotiva pelo experimentador. O grande intuito desse experimento é certificar se as respostas aos estímulos dadas pelos participantes correspondem às dadas pelo experimentador. Essa será, pois, a confirmação de que os estímulos correspondem às emoções analisadas. Para a utilização dos estímulos nos experimentos de percepção e na análise de produção foi necessária a conversão dos vídeos em arquivos de áudio formato *mp3* (320 kbps).⁶¹

⁶¹ Conversões feitas *online* por meio do *site* <http://www.youtube-mp3.org/pt>.

Com a extração dos 32 arquivos, os estímulos foram apresentados aleatoriamente a falantes do português brasileiros (São Paulo, capital) e a falantes do inglês (Inglaterra). É importante salientar que os falantes ingleses não possuem conhecimento de português. Esse cuidado foi tomado para que informações de ordem linguística não pudessem influenciar o resultado da tarefa. O fato de falantes paulistanos saberem ou não inglês não foi controlado por ora, haja vista que todos os estímulos utilizados nesta tese são trechos do PB. Os falantes foram igualmente divididos entre femininos e masculinos e possuem idade entre 20 e 45 anos.

Os estímulos foram intercalados e numerados com o número correspondente de cada sentença (a numeração foi dada em português e em inglês) e separados por um bipe de 0,5s. Foi pedido aos informantes (4 brasileiros e 4 ingleses) para que preenchessem uma planilha, colocando um “X” nas emoções que eles julgavam representar melhor cada um dos estímulos.⁶² As emoções básicas utilizadas foram: *alegria – joy*, *tristeza – sadness*, *raiva – anger* e *medo – fear*.

Os parâmetros acústicos utilizados na análise foram:

- Qualidade de voz – *H1-H2*, *ênfase espectral* e *LTAS*.
- Entoação (Exprosodia) – *TM* (*tom médio*), *cvTM* (*coeficiente de variação do médio*), *dpTM* (*desvio padrão do tom médio*), *aTM* (*assimetria do tom médio*) e *cvF0* (*coeficiente variação de F0*).⁶³

⁶² A íntegra da instrução desse experimento está disponível nos anexos.

⁶³ O *cvF0* é o único parâmetro que não foi processado pelo ExProsodia, tendo seus valores mensurados a partir de dados brutos de F0 extraídos pelo SFS (*Speech Filing System*) (HUCKVALE, 2008).

5.2 Análise e resultados parciais

Os informantes brasileiros, como esperado, tiveram 90% de concordância com relação ao julgamento dado pelo experimentador, ao passo que os informantes ingleses tiveram um desempenho inferior (66%). Um teste de *inter-rater reliability* (Kappa-Fleiss) foi realizado a fim de avaliar a concordância entre participantes. O grau de concordância entre sujeitos brasileiros foi de razoável a bom - $K = 0,73$, $p \sim 0$. No caso dos participantes ingleses, a concordância foi baixa - $K = 0,35$, $p \sim 0$ (FLEISS, 1981).

Foi feito um teste de variância (ANOVA – 1 fator) para todos os sujeitos, envolvendo as emoções básicas (raiva, medo, alegria e tristeza) e os parâmetros entoacionais (TM, cvTM, dpTM, aTM e cvF0). Os resultados do teste para os parâmetros cvTM, dpTM e aTM não foram significativos: $F(3,28) = 0.26$, $p > 0,05$; $F(3,28) = 1.84$, $p > 0,05$ e $F(3,28) = 1.97$, $p > 0,05$, respectivamente. Por outro lado, TM e cvF0 apresentaram resultados significativos: $F(3,28) = 7,97$, $p < 0,05$ e $F(3,28) = 5,55$, $p < 0,05$, respectivamente.

O mesmo procedimento foi adotado para a análise das medidas de qualidade de voz. Somente o parâmetro H1-H2 não foi significativo: $F(3,38) = 1.32$, $p > 0,05$. Os demais parâmetros – LTAS e ênfase espectral – apresentaram resultados significativos: $F(3,38) = 4.92$, $p < 0,05$ e $F(3,38) = 4.25$, $p < 0,05$, respectivamente. Um teste *post hoc* de Tukey (HSD) foi feito para determinar quais diferenças entre as médias foram encontradas no teste de variância. As diferenças significativas encontradas foram: TM – alegria e medo, tristeza e raiva; cvF0 – tristeza e raiva; ênfase espectral – tristeza e raiva; LTAS – tristeza e raiva.

A diferenciação feita pelos ouvintes foi semelhante àquela presente em Darwin (1965) – *exciting and depressing states of the mind*, i.e., as emoções raiva e alegria foram diferenciadas da tristeza e do medo. Parece

razoável que o grau de agitação (*arousal*) dos excertos teve um papel importante, corroborado pelas variações de TM médio, ênfase espectral e LTAS. O aumento de TM correspondeu, na maioria das vezes, a graus de excitação maiores. Em uma análise multidimensional das emoções, poder-se-ia dizer que a distinção entre as emoções foi feita considerando-se os níveis de ativação (não agitado – agitado). Além disso, uma análise baseada somente na ativação seria insuficiente, pois não seria capaz de determinar diferenças entre alegria – raiva e tristeza – medo. O mesmo é válido para os valores de ênfase espectral e LTAS, ou seja, quanto menor a diferença entre as regiões de baixa e alta frequência maior o grau de agitação percebido pelos ouvintes.

Os resultados do estudo piloto são relevantes, pois demonstram que alguns dos parâmetros acústicos escolhidos para a análise podem servir como indicativos de estados emocionais. O próximo passo será a abordagem multidimensional das emoções juntamente com os demais parâmetros acústicos da entoação via análise automática e qualidade de voz.

6. Teste perceptivo com nativos e não nativos – qualidade de voz e análise dimensional das emoções

6.1 Metodologia

O experimento com abordagem multidimensional das emoções e parâmetros de qualidade de voz está subdividido em duas partes: a primeira tem como estímulo excertos de fala normal; a segunda, excertos de fala delexicalizada.⁶⁴ Ambos os experimentos foram realizados por 18 informantes ingleses sem conhecimento de português e 18 informantes brasileiros. Dentre os participantes, três foram descartados por não avaliarem uma das dimensões ou por não completarem a tarefa. Os participantes ouviram, em ambos os experimentos – original e delexicalizado (SONNTAG; PORTELE, 1998)⁶⁵ – os mesmos 32 trechos de fala emotiva utilizados no teste piloto.

⁶⁴ As instruções fornecidas aos participantes estão disponíveis nos anexos.

⁶⁵ Os métodos de delexicalização são, *grosso modo*, maneiras de impossibilitar o sujeito de recuperar o léxico

No método de delexicalização proposto por Sonntag e Portele (1998), os autores demonstraram em experimento de reconhecimento rítmico que o método foi muito bem avaliado. Os participantes foram capazes de relacionar frases escritas de mesma estrutura sintática com estímulos gerados pelo método PURR, ainda que a parte segmental das sentenças fosse diferente das apresentadas nas frases escritas. Apesar de preservar o ritmo e as variações de frequência fundamental, o método não preserva a qualidade de voz. O método PURR determina marcas de *pitch* ao longo da sentença nas quais três ondas senoidais são sobrepostas: a primeira refletindo a amplitude e a frequência originais do som captado; a segunda com o dobro da frequência e 1/4 da amplitude; e a terceira com o triplo da frequência e 1/16 da amplitude.⁶⁶

Para cada um dos estímulos, os participantes tiveram de dar uma nota para cada uma das dimensões utilizadas: *valência* – de não agradável a agradável; *ativação* – de não agitado a agitado; *dominância* – de submisso a não submisso. Para avaliar cada uma das dimensões, os participantes tiveram de utilizar um botão deslizante, que permitiu maior liberdade e maior gama de variação nos julgamentos dos estímulos. Em um primeiro teste, feito com dois informantes brasileiros, houve muita dificuldade em situar a posição do botão que melhor espelharia a avaliação do participante. Para amenizar esse tipo de problema, optou-se por um modelo híbrido, ou seja, que faz uso do botão deslizante (maior liberdade de julgamento) e de valores numéricos de 0 a 100 para cada uma das dimensões. Os valores apareceram no interior do botão simultaneamente ao seu deslizamento. A Figura 9 ilustra a maneira como os participantes julgaram os estímulos durante os experimentos.

⁶⁶ Outros métodos de delexicalização serão discutidos no quarto e último experimento desta tese.



Figura 9: Exemplo de avaliação de estímulo

Os estímulos utilizados nos experimentos de percepção foram analisados através de medidas de qualidade de voz previamente abordadas neste trabalho: *H1-H2*, *ênfase espectral* e *LTAS*. As medidas de ênfase espectral foram extraídas das vogais tônicas [a] que estavam na região mais próxima possível do centro da sentença. Essa decisão foi tomada por três motivos: 1) A escolha do [a] justifica-se por ser uma vogal baixa e por estar menos sujeita aos efeitos de ensurdecimento; 2) A vogal [a] tem o F1 e o F2 mais afastados; caso fosse a vogal [i], F1 estaria na mesma região de H1 e H2, o que tornaria difícil saber se os valores extraídos são efeito da laringe ou da região supralaríngea; 3) Para evitar possíveis influências das bordas da sentença e para se obter um contexto de medida equânime para todas as vogais.

As medidas de ênfase espectral foram extraídas de acordo com os trabalhos de Eriksson, Thunberg e Traunmüller (2001) e Barbosa, Eriksson e Åkesson (2013), *i.e.*, $L - L_0$ em que L é a intensidade do espectro todo e L_0 é a intensidade medida até $1,5F_0$. As medidas de *H1-H2* foram extraídas das mesmas vogais utilizadas nas medições de ênfase espectral. As medidas de *LTAS* foram extraídas das sentenças como um todo.

6.2 Análise e resultados parciais

Para análise dos resultados, foram feitos 12 testes Kappa-Fleiss envolvendo cada uma das dimensões julgadas por nativos e não nativos (Bra. e Ing.) nas duas modalidades de estímulo – normal e delexicalizado. A tabela abaixo apresenta os resultados:

Tabela 1: Resultados do teste Kappa-Fleiss

<i>Fala Normal</i>		
Dimensões	Bra.	Ing.
<i>Valência</i>	<i>0,39⁶⁷</i>	<i>0,15</i>
<i>Ativação</i>	<i>0,59</i>	<i>0,09</i>
<i>Dominância</i>	<i>0,53</i>	<i>0,06</i>
<i>Fala Delexicalizada</i>		
Dimensões	Bra.	Ing.
<i>Valência</i>	<i>0,26</i>	<i>0,19</i>
<i>Ativação</i>	<i>0,41</i>	<i>0,22</i>
<i>Dominância</i>	<i>0,29</i>	<i>0,23</i>

De acordo com os resultados, os participantes brasileiros apresentaram boa concordância nos julgamentos em fala normal, exceto para os julgamentos da dimensão dominância. Por outro lado, os ingleses tiveram uma baixa concordância nos estímulos normais nas três dimensões. No experimento com fala delexicalizada, brasileiros e ingleses se equipararam nas dimensões dominância e valência, apresentando uma baixa concordância.

Na dimensão ativação, os brasileiros apresentaram uma boa concordância, ao passo que os ingleses tiveram uma baixa concordância. Pela ausência do léxico, os brasileiros tiveram a percepção da valência comprometida, o que aponta para um papel importante do léxico nessa dimensão. No caso dos participantes ingleses, ocorreu exatamente o contrário, ou seja, a valência não variou muito com relação aos dados de fala normal, mas ativação e dominância, sim.

No caso da percepção dos não nativos, os segmentos do português

⁶⁷ Todos os resultados apresentaram $P \sim 0$.

poderiam exercer alguma influência? De acordo com os resultados obtidos em Peres, Consoni e Ferreira Netto (2011), variações tonais maiores são exigidas para a percepção da ênfase quando a cadeia segmental está presente. Nesse caso, então, a ausência do léxico seria favorável a um melhor desempenho dos participantes não nativos. Teremos uma resposta plausível quando os julgamentos dos participantes forem cruzados com os parâmetros acústicos. Para a análise via inferência acústica, foram feitas regressões lineares⁶⁸ envolvendo as medidas de qualidade de voz (H1-H2, ênfase espectral e LTAS) e os julgamentos dos participantes. A tabela 2 apresenta os resultados obtidos para cada regressão:

Tabela 2: Regressões lineares entre os julgamentos dos participantes sobre fala normal e medidas acústicas de qualidade de voz. Os resultados não foram significativos

<i>Fala normal</i>			
	Bra.		Ing.
Dimensões	Percepção vs acústico⁶⁹	Dimensões	Percepção vs acústico
Valência	Percep./H1-H2 - R ² = 00,1 Percep./LTAS - R ² = 0,05 Percep./EE - R ² = 0,01	Valência	Percep./H1-H2 - R ² = 0,08 Percep./ LTAS - R ² = 0,18 Percep./EE - R ² = 0,03
Ativação	Percep./H1-H2 - R ² = 0,14 Percep./LTAS - R ² = 0,24 Percep./EE - R ² = 0,18	Ativação	Percep./H1-H2 - R ² = 0,19 Percep./LTAS - R ² = 0,04 Percep./EE - R ² = 0,06
Dominância	Percep./H1-H2 - R ² = 0,30 Percep./ LTAS - R ² = 0,01 Percep./EE - R ² = 0,02	Dominância	Percep./H1-H2 - R ² = 0,02 Percep./ LTAS - R ² = 0,01 Percep./EE - R ² = 0,03

Como demonstrado acima, não houve resultado significativo nas regressões lineares entre os julgamentos dos participantes e as medidas acústicas de qualidade de voz. Com o intuito de buscar alguma relação entre as dimensões estudadas, foram feitos testes de correlação de Pearson envolvendo os julgamentos que os participantes nativos e não nativos deram para cada dimensão quando participaram do experimento com fala normal. Os resultados não foram significativos.

⁶⁸ A regressão linear é uma equação para estimar o valor de uma variável y (valor esperado), dados os valores de uma ou mais variáveis x.

⁶⁹ EE – Ênfase espectral.

Tabela 3: Coeficiente de correlação de Pearson entre as dimensões – fala normal

<i>Fala normal</i>	
Bra.	
<i>Valência – Ativação</i>	R = 0,04
<i>Valência – Dominância</i>	R = 0,31
<i>Ativação – Dominância</i>	R = 0,37
Ing.	
<i>Valência – Ativação</i>	R = 0,14
<i>Valência – Dominância</i>	R = 0,02
<i>Ativação – Dominância</i>	R = 0,10

O mesmo foi feito para o experimento com fala delexicalizada, obtendo, da mesma forma, resultados não significativos.

Tabela 4: Coeficiente de correlação de Pearson entre as dimensões – fala delexicalizada

<i>Fala delexicalizada</i>	
Bra.	
<i>Valência – Ativação</i>	R = -0.05
<i>Valência – Dominância</i>	R = 0.03
<i>Ativação – Dominância</i>	R = 0.28
Ing.	
<i>Valência – Ativação</i>	R = 0.18
<i>Valência – Dominância</i>	R = 0.24
<i>Ativação – Dominância</i>	R = 0.31

Por outro lado, os testes de correlação entre julgamentos de mesma dimensão entre falantes de línguas diferentes apresentaram resultados significativos:

Tabela 5: Coeficiente de correlação de Pearson entre mesmas dimensões entre nativos e não nativos

<i>Bra. x Ing.</i>		
Dimensões	Normal	Delexicalizada
Valência	R = 0.54	R = 0.62
Ativação	R = 0.71	R = 0.89
Dominância	R = 0.57	R = 0.64

Os resultados apontam para uma regularidade no julgamento dado pelos ouvintes a cada uma das dimensões. Os graus de ativação, seguidos pelos graus de valência e dominância, parecem ser uma característica saliente para ambos os participantes (nativos ou não nativos), com estímulos normais ou delexicalizados. As medidas acústicas da qualidade

de voz utilizadas na análise de inferência acústica parecem não explicar o julgamento dos juízes nativos, da mesma forma que parecem não explicar o menor valor K obtido pelos falantes não nativos.

7. Teste perceptivo com nativos e não nativos – parâmetros acústicos (ExProsodia) e análise dimensional das emoções

7.1 Metodologia

A metodologia desse experimento é rigorosamente a mesma do anterior por se tratar da mesma aplicação do teste de percepção. A única diferença está na análise de produção, a qual é baseada nos parâmetros acústicos da entoação extraídos por meio da análise automática – ExProsodia. Assim, os mesmos julgamentos dos 36 participantes (18 brasileiros e 18 ingleses) do experimento anterior (com fala normal e delexicalizada) serão utilizados no cotejo com os parâmetros acústicos da entoação, a saber: tom médio (TM), desvio padrão do tom médio (dpTM), coeficiente de variação do tom médio (cvTM), assimetria do TM (aTM), diferença entre o tom médio e a menor unidade entoacional (TMmUE),

menor valor de uma unidade entoacional (mUE), coeficiente de variação dos valores positivos de Foco/Ênfase (cvposF/E), coeficiente de variação dos valores negativos de Foco/Ênfase (cvnegF/E), duração das unidades entoacionais (durUE), desvio padrão da duração das unidade entoacionais (dpdurUE), duração dos intervalos entre as unidades entoacionais (durintUE), desvio padrão da duração dos intervalos entre as unidades entoacionais (dpdurintUE).

7.2 Análise e resultados parciais

Para a análise dos dados, foram feitas duas análises de regressão linear: uma simples e outra múltipla. A primeira delas envolveu cada uma das dimensões – valência, ativação e dominância – e os parâmetros acústicos; a segunda envolveu as mesmas três dimensões e pares dos parâmetros acústicos utilizados nas regressões lineares simples. No total, foram feitas 240 regressões, divididas em 144 simples e 96 múltiplas. Para iniciar a análise, as tabelas 6 e 7 (a seguir) apresentam os resultados das regressões lineares simples,⁷⁰ envolvendo os julgamentos dos informantes brasileiros e ingleses sobre os estímulos com fala normal e parâmetros acústicos obtidos via análise ExProsodia. Os resultados em cinza escuro indicam valores maiores que $R^2 = 0,50$. A relevância estatística dos valores acima de $R^2 = 0,40$ não pode ser descartada *a priori*. Por essa razão esses valores estão destacados em cinza claro.⁷¹

⁷⁰ Para todas as regressões lineares foi tomado o valor de R ajustado.

⁷¹ Cohen (1988, apud FIGUEIREDO FILHO; SILVA JUNIOR, 2009) classifica valores entre 0,10 e 0,29 como baixos, entre 0,30 e 0,49 como médios, e entre 0,50 e 1 como altos.

Resultado das Regressões lineares simples (valores de r^2 ajustado)

Parâmetros acústicos - ExProsodia

Bra. – fala normal

	TM	dpTM	cvTM	aTM	mUE	TMmUE	cvposF/E	cvnegF/E	durUE	dpdurUE	durintUE	dpdurintUE
<i>Valência</i>	0,01	0,11	0,16	0,08	0,00	0,06	0,17	0,03	0,10	0,02	0,04	0,00
<i>Ativação</i>	0,40	0,14	0,82	0,02	0,02	0,00	0,11	0,01	0,84	0,01	0,44	0,10
<i>Dominância</i>	0,61	0,08	0,31	0,11	0,05	0,29	0,03	0,01	0,38	0,01	0,41	0,18

Tabela 6: Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala normal) e os parâmetros acústicos – ExProsodia.

Ing. – fala normal

	TM	dpTM	cvTM	aTM	mUE	TMmUE	cvposF/E	cvnegF/E	durUE	dpdurUE	durintUE	dpdurintUE
<i>Valência</i>	0,04	0,11	0,17	0,07	0,00	0,05	0,13	0,03	0,01	0,01	0,06	0,00
<i>Ativação</i>	0,45	0,21	0,73	0,02	0,01	0,07	0,05	0,006	0,81	0,06	0,64	0,17
<i>Dominância</i>	0,48	0,23	0,57	0,02	0,06	0,16	0,02	0,01	0,62	0,11	0,69	0,23

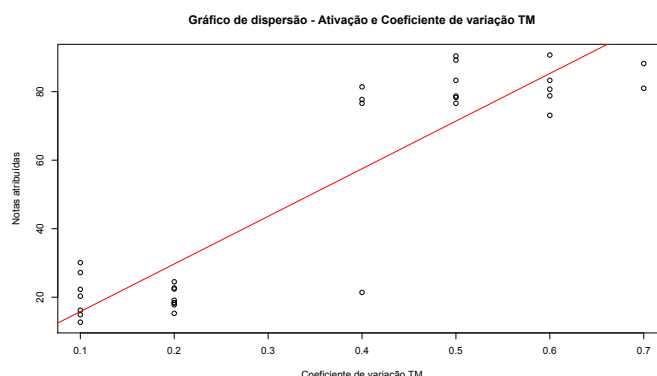
Tabela 7: Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala normal) e os parâmetros acústicos – ExProsodia.

(TM) - tom médio ; (dpTM) - desvio padrão do tom médio; (cvTM) - coeficiente de variação do tom médio; (aTM) assimetria do TM; (mUE) - menor valor de uma unidade entoacional; (TMmUE) - diferença entre o tom médio e a menor unidade entoacional; (cvposF/E) - coeficiente de variação dos valores positivos de foco/ênfase; (cvnegF/E) - coeficiente de variação dos valores negativos de foco/ênfase; (durUE) - duração das unidades entoacionais; (dpdurUE) - desvio padrão da duração das unidades entoacionais; (durintUE) - duração dos intervalos entre as unidades entoacionais; (dpdurintUE) - desvio padrão da duração dos intervalos entre as unidades entoacionais

A valência não apresentou resultados significativos nas regressões lineares com parâmetros acústicos do ExProsodia. Esse é um problema comum na dimensão valência, pois é uma dimensão que trata do que é ou não prazeroso e que, portanto, pode ser muito variável e dependente do que é dito. Os julgamentos dados a essa dimensão foram muito diferentes entre si, o que foi demonstrado pelo teste de Kappa-Fleiss. Não só a dimensão valência apresenta problemas dessa natureza, pois os rótulos de qualquer dimensão podem influenciar as respostas. Adiante, será discutido o caso da ativação com relação aos não nativos.

Nas tabelas com fala normal, houve alguns resultados significativos. No caso dos brasileiros, os parâmetros acústicos cvTM e durUE explicaram mais de 80% da variação dos julgamentos de ativação dados por esses informantes (0,82 e 0,84, respectivamente). A variação dos julgamentos da dimensão dominância pôde ser explicada pelos parâmetros TM (0,61). Os gráficos de dispersão^{72, 73} abaixo ilustram esses resultados:

Gráfico 1: (Bra – fala normal) Julgamento de ativação e coeficiente de variação (cvTM).
Valor de R²: 0,82



⁷² Foram feitos somente gráficos de dispersão com valores de R² maiores que 0,50.

⁷³ O gráfico de dispersão representa pontos no espaço cartesiano XY que representam ao mesmo tempo valores de duas variáveis quantitativas. A linha de tendência indica em que grau se estabelece a interação entre os dados. A reta pode ter um sentido crescente [↗] ou decrescente [↘]. Quanto mais perto a reta está de 45° ou 135°, maior será o coeficiente de determinação (R²).

Gráfico 2: (Bra – fala normal) Julgamento de ativação e duração das unidades entoacionais (durUE). Valor de R^2 : 0,84

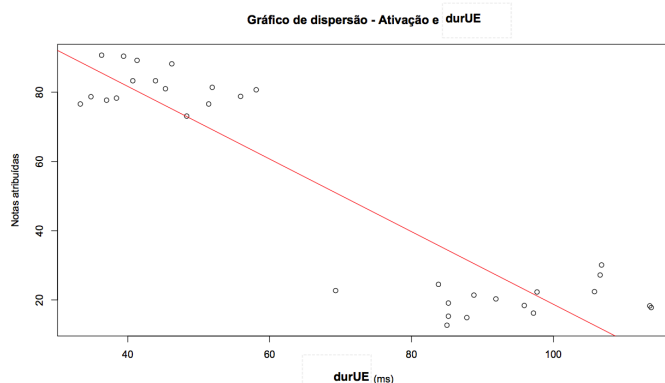
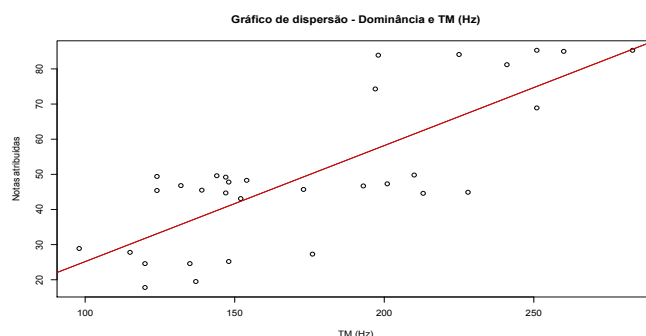


Gráfico 3: (Bra – fala normal) Julgamento de ativação e tom médio (TM). Valor de R^2 : 0,61



O Gráfico 1 apresentou uma linha ascendente, i.e., à medida que o cvTM aumenta, os valores atribuídos à ativação também aumentam. O Gráfico 2 exibiu a dispersão dos dados envolvendo as variáveis ativação e duração das unidades entoacionais (durUE) e apresentou uma linha de tendência descendente. Assim, no caso de ativação e durUE, parte das notas mais altas pode ser explicada pela brevidade das unidades entoacionais (UE), da mesma forma que o seu inverso, ou seja, menores notas e maiores valores de durUE.

Outra característica é a aparente relação entre as variações de TM e de cvTM, ou seja, parece que cvTM é uma consequência da variação de TM. Entretanto, a dispersão dos valores de TM e cvTM acontece de maneira diferente. No caso de TM (Gráfico 3), a distribuição dos valores ao longo do eixo X é mais distribuída, ao passo que os valores de cvTM (Gráfico

1) apresentam um vazio no centro do eixo X. Note-se que esse mesmo vazio também está presente nos valores de durUE (Gráfico 2).

Em dominância e TM, as notas mais elevadas estão relacionadas a valores elevados de TM, sendo o inverso verdadeiro. No caso dos ingleses, ativação e dominância tiveram resultados significativos com relação aos parâmetros cvTM, durUE e durintUE. Os resultados para ativação foram: cvTM (0,73), durUE (0,81), durintUE (0,64); para dominância foram: cvTM (0,57), durUE (0,62), durintUE (0,69). Os gráficos a seguir ilustram os resultados:

Gráfico 4: (Ing – fala normal) Julgamento de ativação e coeficiente de variação de TM (cvTM). Valor de R²: 0,73

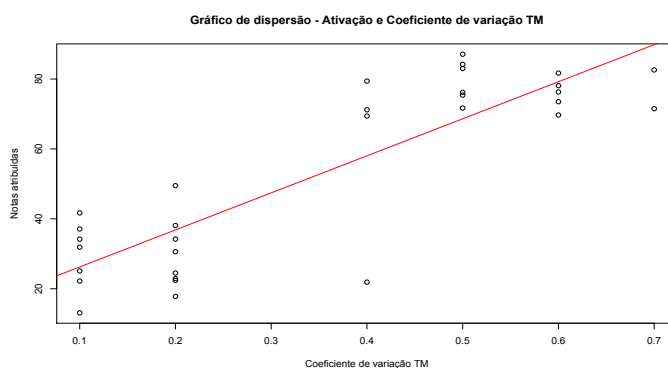


Gráfico 5: (Ing – fala normal) Julgamento de ativação e duração das unidades entoacionais (durUE). R²: 0,81

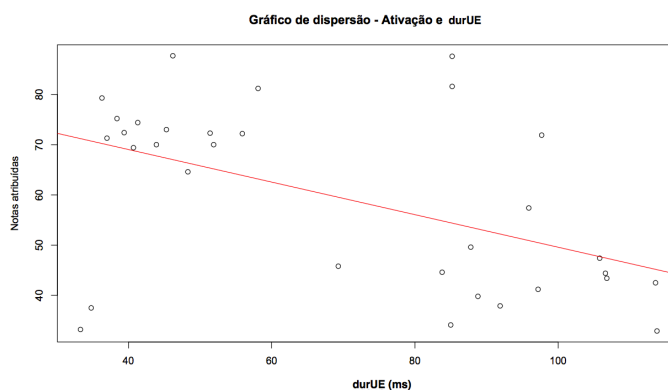


Gráfico 6: (Ing – fala normal) Julgamento de ativação e duração entre as unidades entoacionais (durintUE). $R^2: 0,64$

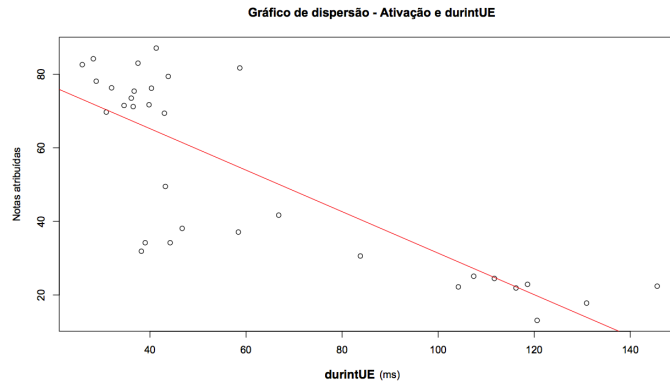


Gráfico 7: (Ing – fala normal) Julgamento de dominância e coeficiente de variação de TM (cvTM). $R^2: 0,57$

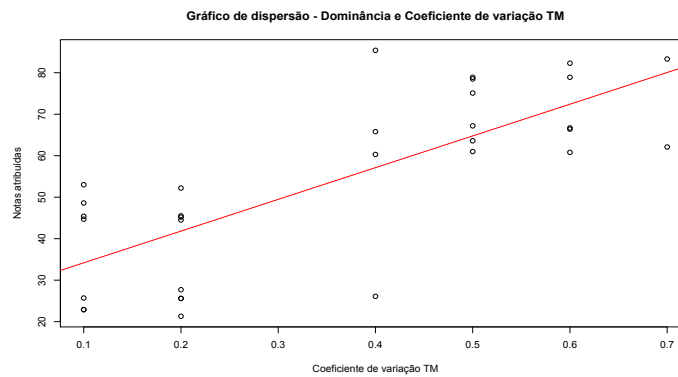


Gráfico 8: (Ing – fala normal) Julgamento de dominância e duração das unidades entoacionais (durUE). $R^2: 0,62$

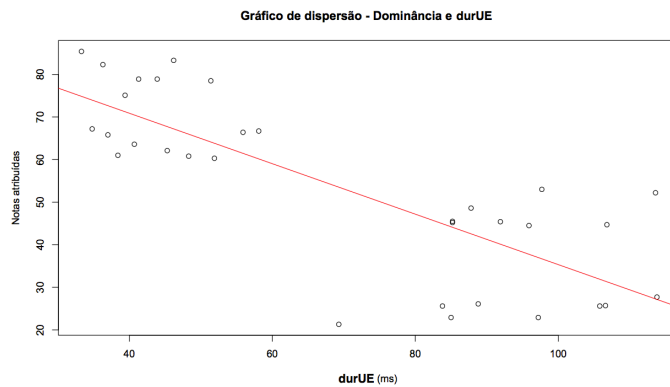
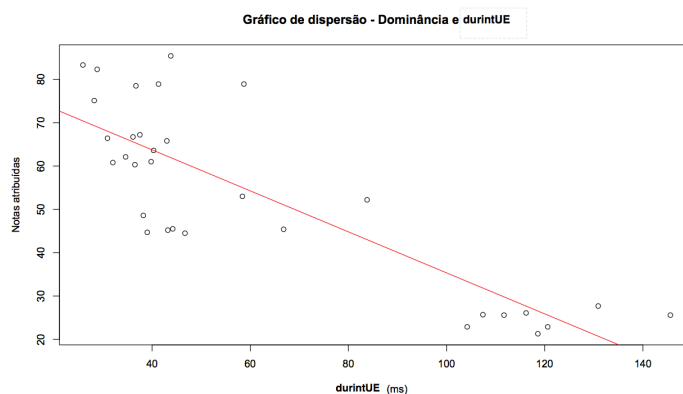


Gráfico 9: (Ing – fala normal) Julgamento de dominância e duração entre as unidades entoacionais (durintUE). $R^2: 0,69$



Igualmente aos resultados dos informantes brasileiros, os julgamentos dados pelos ingleses cresceram à medida que cvTM aumentava, nas duas dimensões, ativação e dominância. As durações das unidades entoacionais (durUE) tiveram uma relação inversa, do mesmo modo que a duração dos intervalos entre elas (durintUE), em ambas as dimensões (ativação e dominância). Em outras palavras, as notas de ativação e dominância cresceram à medida que valores de durUE e durintUE diminuíram. Apesar de os valores de R^2 terem sido elevados, em alguns casos é possível perceber que há alguns valores espúrios que estão afastados da linha de tendência. Essa característica pode ser notada no Gráfico 8.

Se compararmos os gráficos de 1 a 3 (brasileiros) com os gráficos de 4 a 9 (ingleses), percebe-se que há um detalhamento perceptivo maior por parte dos ingleses, i.e., aparentemente eles se detêm numa quantidade maior de parâmetros para avaliar o estímulo. No caso dos brasileiros, três parâmetros apresentaram valores de R^2 significativos, ao passo que, para os ingleses, foram seis.

As tabelas 8 e 9 (a seguir) apresentam os resultados das regressões lineares envolvendo os julgamentos dos brasileiros e dos ingleses sobre os estímulos com fala delexicalizada e os parâmetros acústicos do ExProsodia.

Resultado das Regressões lineares simples (valores de r^2 ajustado) Parâmetros acústicos - ExProsodia

Bra. – fala delexicalizada

	TM	dpTM	cvTM	aTM	mUE	TMmUE	cvposF/E	cvnegF/E	durUE	dpdurUE	durintUE	dpdurintUE
<i>Valência</i>	0,02	0,12	0,14	0,20	0,03	0,16	0,11	0,03	0,10	0,03	0,01	0,03
<i>Ativação</i>	0,29	0,34	0,39	0,01	0,07	0,32	0,03	0,03	0,28	0,11	0,66	0,23
<i>Dominância</i>	0,19	0,24	0,54	0,00	0,03	0,03	0,08	0,03	0,53	0,05	0,42	0,10

Tabela 8: Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala delexicalizada) e os parâmetros acústicos – ExProsodia.

Ing. – fala delexicalizada

	TM	dpTM	cvTM	aTM	mUE	TMmUE	cvposF/E	cvnegF/E	durUE	dpdurUE	durintUE	dpdurintUE
<i>Valência</i>	0,19	0,18	0,48	0,03	0,01	0,01	0,06	0,02	0,43	0,01	0,25	0,06
<i>Ativação</i>	0,10	0,24	0,32	0,00	0,01	0,12	0,01	0,01	0,23	0,07	0,43	0,17
<i>Dominância</i>	0,00	0,32	0,34	0,04	0,03	0,03	0,11	0,02	0,30	0,04	0,42	0,09

Tabela 9 : Resultado das regressões lineares simples envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala delexicalizada) e os parâmetros acústicos – ExProsodia.

(TM) - tom médio ; (dpTM) - desvio padrão do tom médio; (cvTM) - coeficiente de variação do tom médio; (aTM) assimetria do TM; (mUE) - menor valor de uma unidade entoacional; (TMmUE) - diferença entre o tom médio e a menor unidade entoacional; (cvposF/E) - coeficiente de variação dos valores positivos de foco/ênfase; (cvnegF/E) - coeficiente de variação dos valores negativos de foco/ênfase; (durUE) - duração das unidades entoacionais; (dpdurUE) - desvio padrão da duração das unidades entoacionais; (durintUE) - duração dos intervalos entre as unidades entoacionais; (dpdurintUE) - desvio padrão dos intervalos entre as unidades entoacionais

Houve resultados significativos entre os julgamentos dos informantes brasileiros e alguns dos parâmetros acústicos. No caso dos informantes ingleses, não houve resultado significativo. Entretanto, houve valores de R^2 entre 0,40 e 0,49 nas dimensões ativação e dominância para os parâmetros $durUE$ e $durintUE$ (0,43 e 0,42, respectivamente). No caso da ativação, o valor foi de 0,48 para o parâmetro $cvTM$, número que não pode ser ignorado, pois explica quase 50% da variação dos dados da regressão linear em questão. Para os julgamentos dados pelos brasileiros, na dimensão dominância, os parâmetros acústicos relevantes foram: $cvTM$, $durUE$, com valores de 0,54 e 0,53; na ativação somente $durintUE$ apresentou resultado significativo (0,66).

Gráfico 10: (Bra – fala delexicalizada) Julgamento de dominância e coeficiente de variação de TM ($cvTM$). R^2 : 0,54

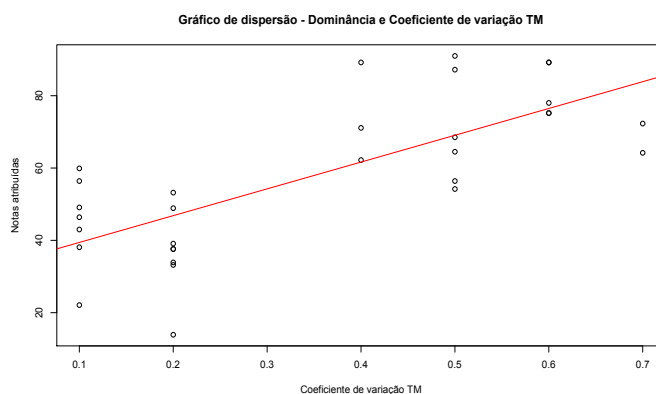


Gráfico11: (Bra – fala delexicalizada) Julgamento de dominância e duração das unidades entoacionais ($durUE$). R^2 : 0,53

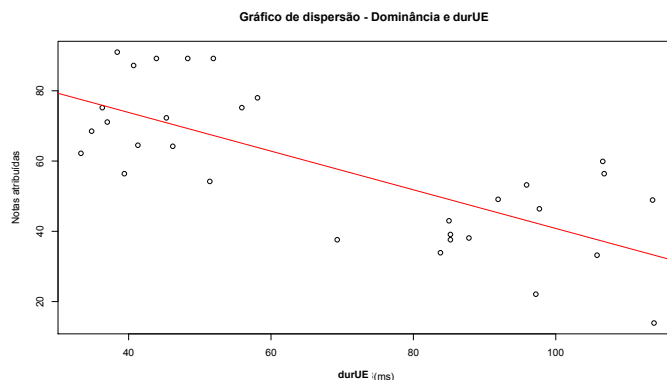
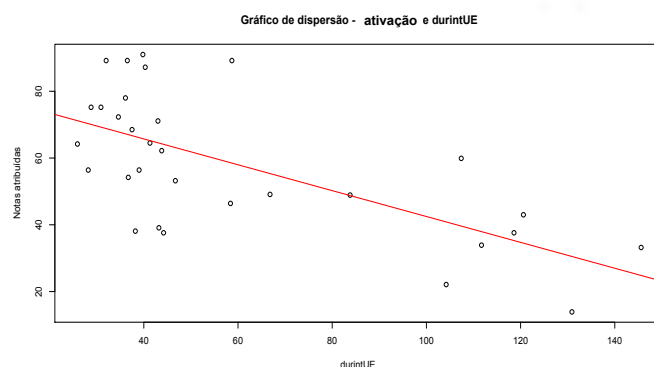


Gráfico12: (Bra – fala delexicalizada) Julgamento de ativação e a duração dos intervalos entre as unidades entoacionais (durintUE).

$R^2: 0,66$



Pode-se notar, nas tabelas 8 e 9, que a ausência do léxico influenciou significativamente os julgamentos de brasileiros e ingleses e, por conseguinte, os resultados das regressões lineares. Os valores de R^2 foram inferiores aos encontrados no experimento com fala normal.

O próximo passo da análise será o resultado das regressões lineares múltiplas, envolvendo os mesmos julgamentos dos informantes brasileiros e ingleses sobre estímulos com fala normal e combinações de alguns parâmetros acústicos. As regressões lineares múltiplas (tabelas 10 e 11 a seguir) apresentaram resultados significativos entre os julgamentos dos brasileiros sobre fala normal e as combinações dos parâmetros acústicos. Para ativação, as combinações relevantes foram: TM + durUE (0,86), TM + durintUE (0,53), cvTM + durUE (0,92), cvTM + dpdurUE (0,81), cvTM + durintUE (0,85) e cvTM + dpdurintUE (0,82).

**Resultado das Regressões lineares múltiplas
(valores de r² ajustado)
Parâmetros acústicos - ExProsodia**

Bra. – fala normal

	TM + dpdurUE	TM + durintUE	TM + dpdurintUE	cvTM + durUE	cvTM + dpdurUE	cvTM + durintUE	cvTM + dpdurintUE
<i>Valência</i>	0,38	0,18	0,02	0,13	0,11	0,10	0,12
<i>Ativação</i>	0,86	0,53	0,38	0,92	0,81	0,85	0,82
<i>Dominância</i>	0,66	0,73	0,66	0,44	0,33	0,54	0,40

Tabela 10: Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala normal) e os parâmetros acústicos combinados – ExProsodia.

Ing. – fala normal

	TM + dpdurUE	TM + durintUE	TM + dpdurintUE	cvTM + durUE	cvTM + dpdurUE	cvTM + durintUE	cvTM + dpdurintUE
<i>Valência</i>	0,38	0,21	0,03	0,15	0,13	0,11	0,13
<i>Ativação</i>	0,85	0,40	0,48	0,86	0,77	0,87	0,78
<i>Dominância</i>	0,71	0,48	0,47	0,68	0,61	0,79	0,66

Tabela 11: Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala normal) e os parâmetros acústicos combinados – ExProsodia.

(TM) - tom médio; (cvTM) - coeficiente de variação do tom médio; (durUE) - duração das unidades entoacionais; (dpdurUE) - desvio padrão da duração das unidades entoacionais; (durintUE) - duração dos intervalos entre as unidades entoacionais; (dpdurintUE) - desvio padrão da duração dos intervalos entre as unidades entoacionais

O aumento dos valores dos coeficientes de determinação (R^2), quando se tem a combinação de parâmetros nas regressões lineares múltiplas, é bastante expressivo. As combinações possuem parâmetros de natureza diferente, um de natureza tonal e outro de natureza rítmica.

A seguir, os gráficos tridimensionais de dispersão ilustram os resultados para ativação – participantes brasileiros. Cada um dos gráficos foi feito em dois ângulos diferentes, como se o gráfico fosse girado para a direita, para que se possa visualizar melhor os *clusters* resultantes das combinações dos parâmetros acústicos e dos julgamentos das dimensões.

Com as duas perspectivas disponíveis (ângulos 1 e 2), consegue-se notar características na dispersão dos dados que não seriam visíveis se o gráfico fosse visto somente por um ângulo. No caso dos gráficos 13 e 14, os *clusters* são bem formados e não há sobreposição, mas os dados nem sempre são tão bem organizados assim, como será visto nos dados do experimento.⁷⁴ Por conta disso, as duas perspectivas serão úteis, pois somente com elas será possível ter uma ideia mais clara sobre como os dados foram agrupados. Após os exemplos (Gráficos 13 e 14), seguem os gráficos da análise dos dados.

As regressões lineares múltiplas apresentaram resultados significativos entre os julgamentos dos brasileiros sobre fala normal e as combinações dos parâmetros acústicos. Para ativação, as combinações relevantes foram: TM + durUE (0,86), TM + durintUE (0,53), cvTM + durUE (0,92), cvTM + durintUE (0,85), cvTM + dpdurUE (0,81) e cvTM + dpdurintUE (0,82). O nome de cada emoção contido no gráfico foi atribuído pelos participantes do experimento piloto relatado no capítulo 5.

⁷⁴ Os dados utilizados nos gráficos 13 e 14 são provenientes do conjunto de dados *Car: Companion to Applied Regression* para R.

Gráfico 13: Modelo de visualização dos gráficos de dispersão 3D com a primeira perspectiva (ângulo 1). As letras A, B e C indicam cada um dos eixos que formam o gráfico. Na primeira perspectiva, o eixo A está em primeiro plano. As setas pretas pontilhadas indicam o movimento que irá gerar o gráfico com a segunda perspectiva (ângulo 2). O eixo B, destacado pelo círculo e as setas vermelhas, será o lado que estará em primeiro plano na segunda perspectiva (ângulo 2) ilustrada no Gráfico 14

Exemplo - ângulo 1

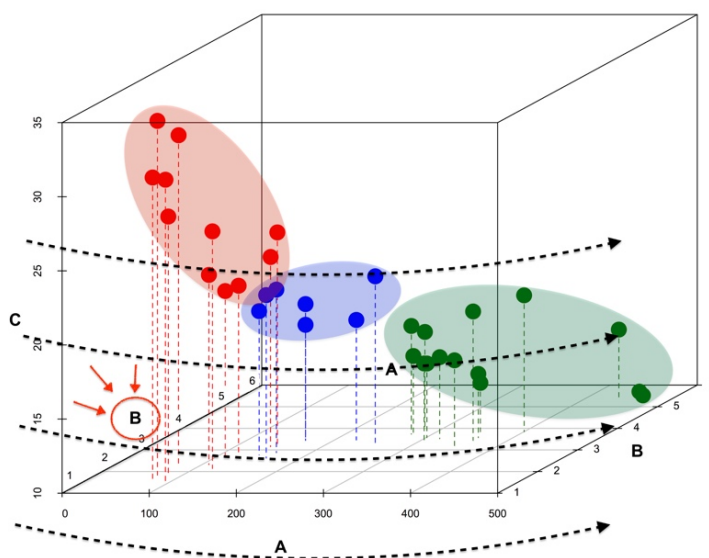


Gráfico 14: Modelo de visualização dos gráficos de dispersão 3D com a segunda perspectiva (ângulo 2). As letras A, B e C indicam cada um dos eixos que formam o gráfico. Neste ângulo, o eixo B está em primeiro plano por conta do giro à direita feito no Gráfico 13

Exemplo - ângulo 2

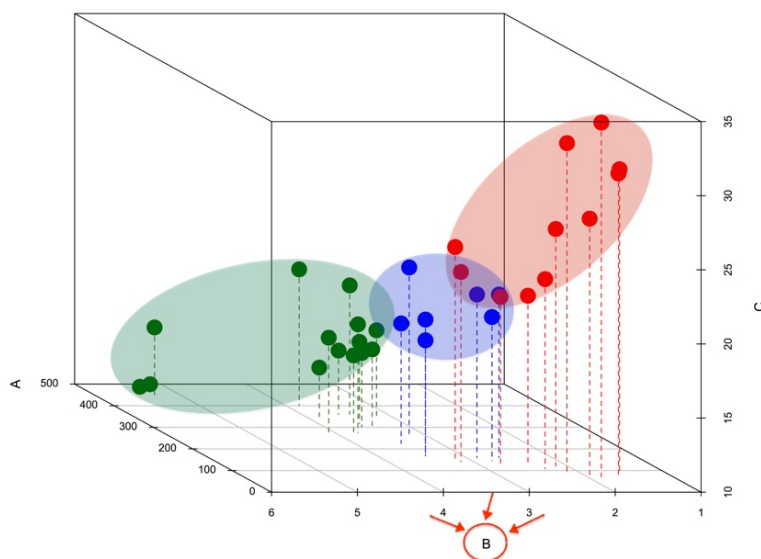


Gráfico 15: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,86

Gráficos de dispersão 3D - Ativação | TM e durUE

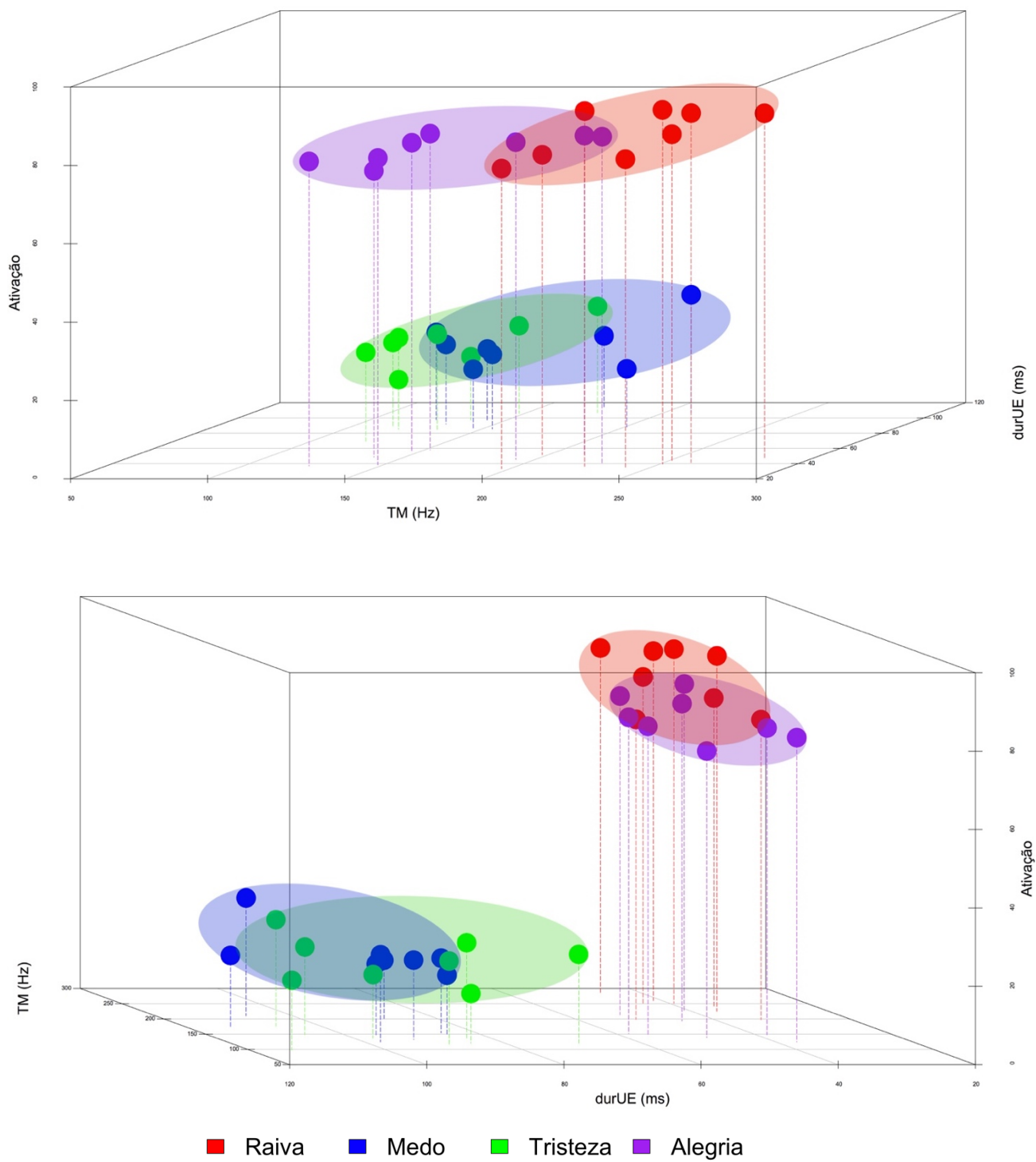


Gráfico 16: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração dos intervalos entre as unidades entoacionais - durintUE). Valor de R^2 : 0.53

Gráficos de dispersão 3D - Ativação | TM e durintUE

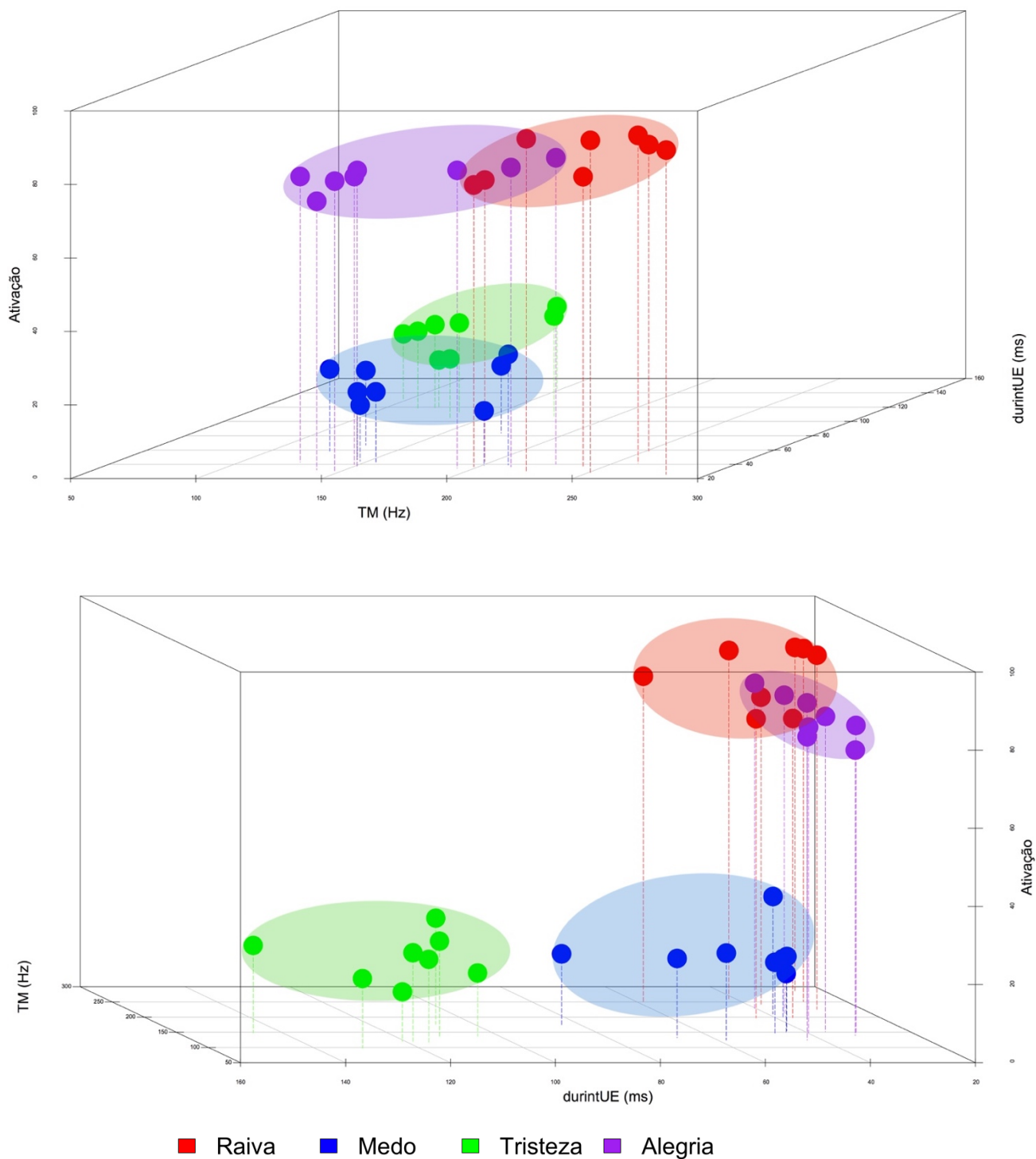


Gráfico 17: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração das unidades entoacionais – durUE). Valor de R²: 0.92

Gráficos de dispersão 3D - Ativação | cvTM e durUE

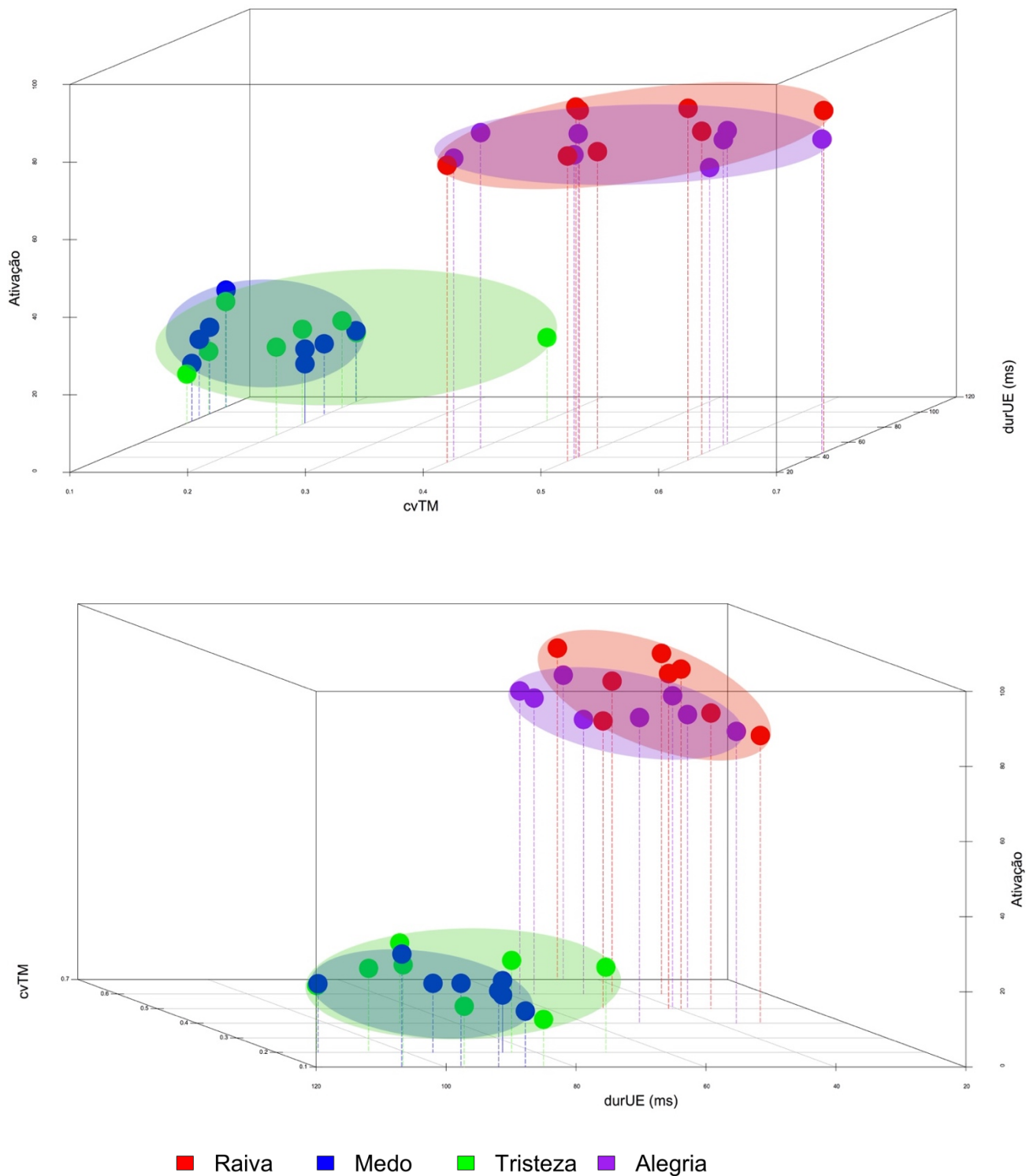


Gráfico 18: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos entre as unidades entoacionais – durintUE).
 Valor de R^2 : 0,85

Gráficos de dispersão 3D - Ativação | cvTM e durintUE

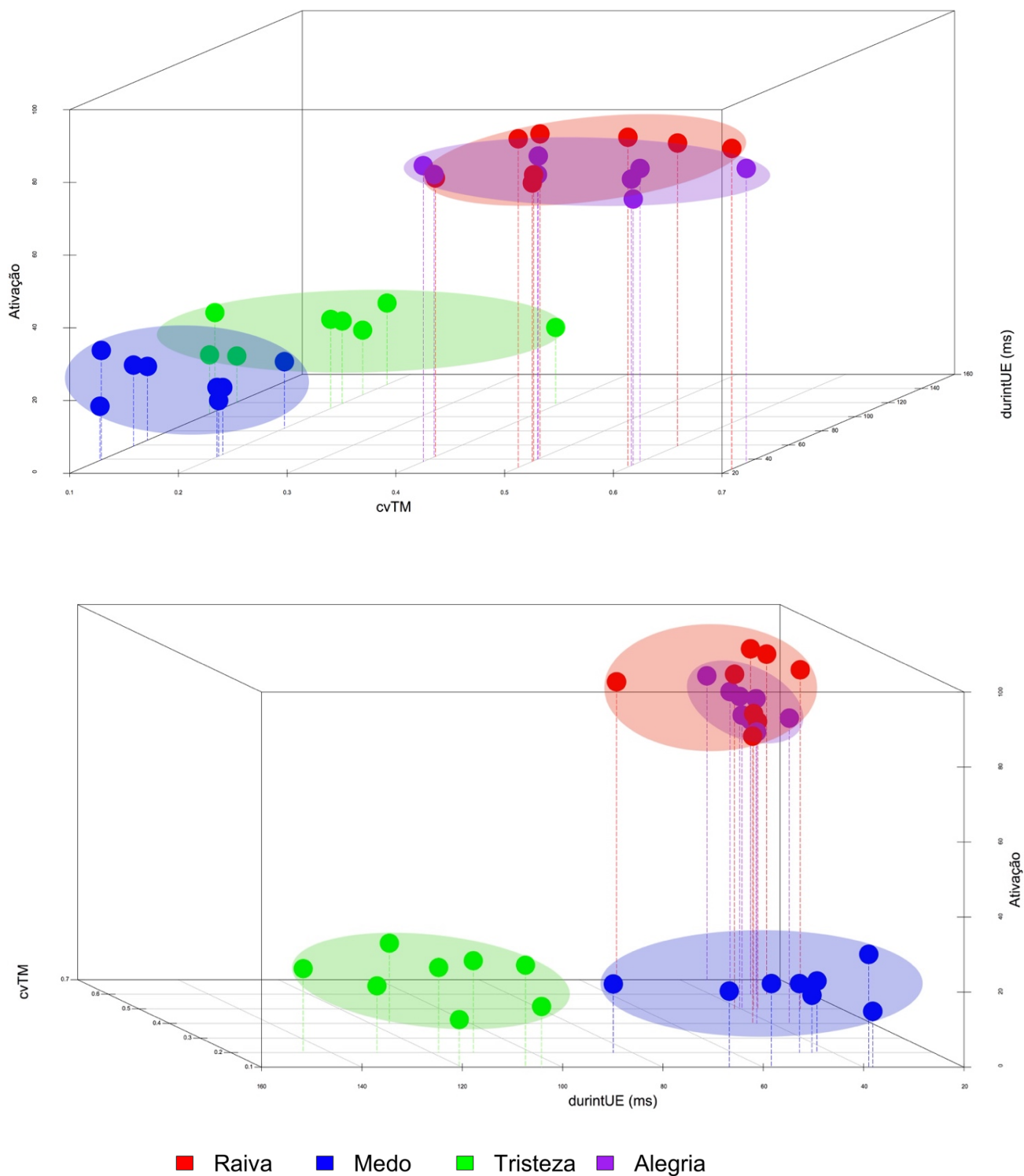


Gráfico 19: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração das unidades entoacionais – dpdurUE). Valor de R^2 : 0,81

Gráficos de dispersão 3D - Ativação | cvTM e dpdurUE

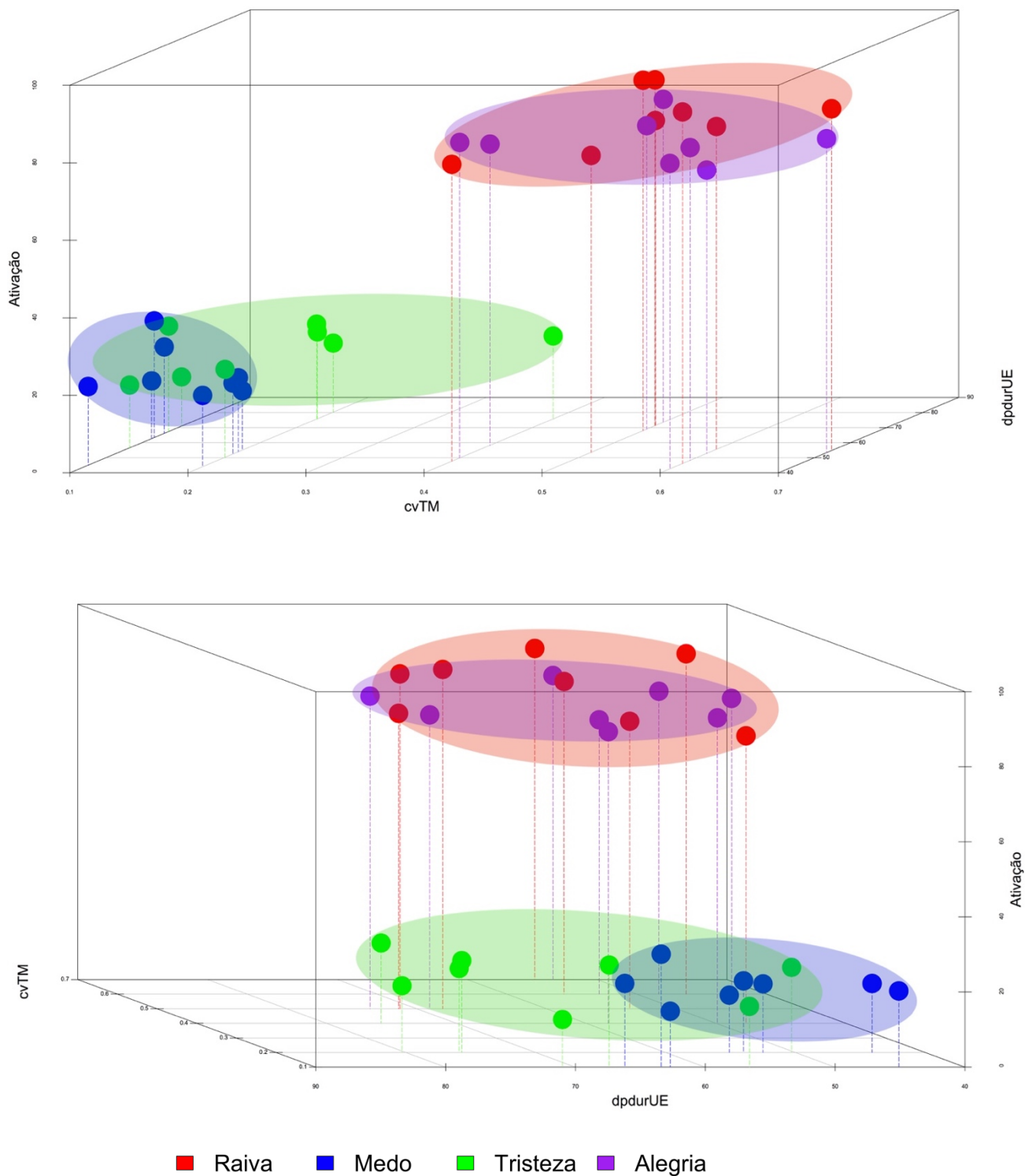
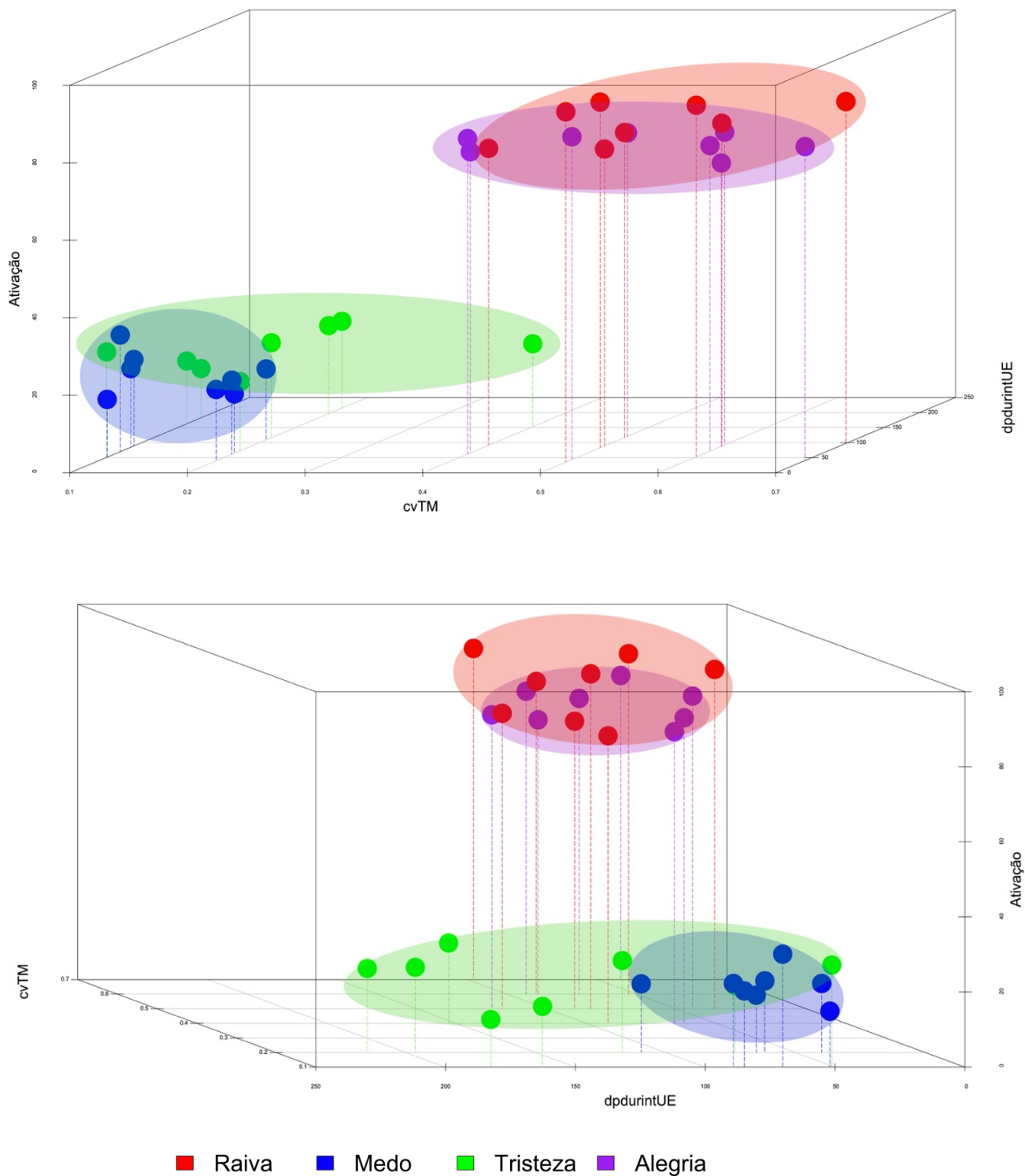


Gráfico 20: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração dos intervalos entre as unidades entoacionais – dpdurintUE). Valor de R²: 0,82

Gráficos de dispersão 3D - Ativação | cvTM e dpdurintUE



A relação entre a duração das unidades entoacionais ($durUE$) é inversamente proporcional às notas dadas pelos participantes brasileiros. Por outro lado, a relação entre TM e $cvTM$ é diretamente proporcional, ou seja, as notas aumentam à medida que esses parâmetros têm valores elevados.

Os gráficos de dispersão tridimensionais ilustram com detalhes como acontece a interação entre os parâmetros acústicos e os julgamentos dos participantes. Um padrão geral patente é que certas emoções se agrupam de maneira regular quando certos parâmetros acústicos são analisados, o que deixa transparecer uma ligação estreita entre a percepção dos estímulos (por conseguinte a avaliação dos participantes) e a variação dos parâmetros acústicos.

As características relacionadas com a altura tonal (TM), ou com sua variação ($cvTM$), estão estreitamente ligadas à diferenciação das emoções, especificamente ao agrupamento delas. Os valores elevados de ativação estão ligados às emoções raiva e alegria. Os *clusters* estão bem ilustrados nos gráficos de 15 a 20 (p. 130 – 135). O mesmo se aplica às emoções com baixos valores de ativação como tristeza e medo. Os parâmetros de natureza rítmica como $dpdurUE$, $durintUE$ e $dpdurintUE$ estão ligados às emoções de baixa ativação e são responsáveis pela distinção entre elas, como no caso do medo e da tristeza (gráficos 16, 18, 19 e 20).

Para dominância, as combinações relevantes foram: $TM + durUE$ (0,66), $TM + dpdurUE$ (0,61), $TM + durintUE$ (0,73), $TM + dpdurintUE$ (0,66) e $durintUE$ (0,54). Os gráficos tridimensionais de dispersão com os resultados para dominância – participantes brasileiros – estão a seguir.

Gráfico 21: Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,66

Gráficos de dispersão 3D - Dominância | TM e durUE

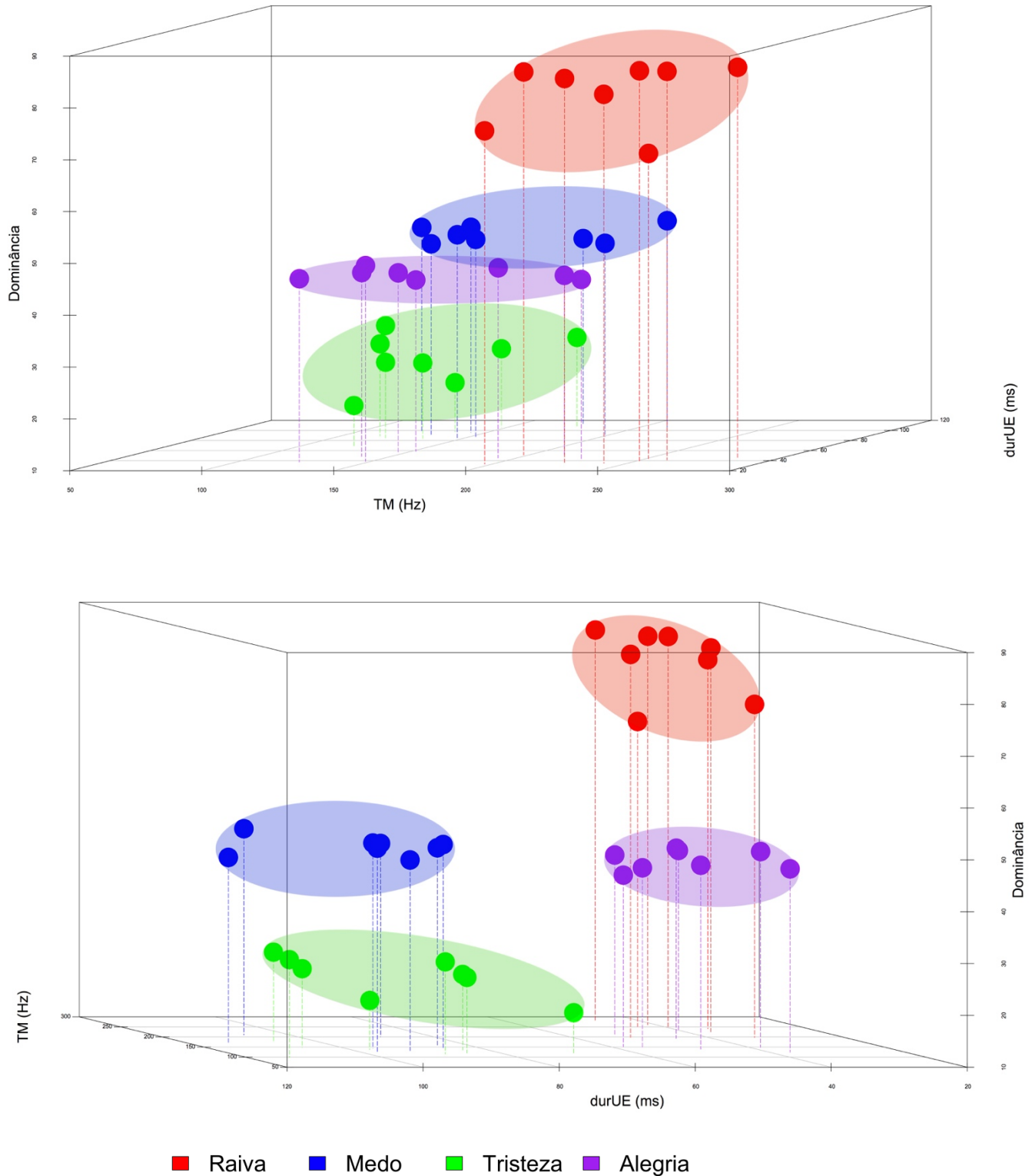


Gráfico 22: Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + desvio padrão da duração das unidades entoacionais – dpdurUE). Valor de R²: 0,61

Gráficos de dispersão 3D - Dominância | TM e dpdurUE

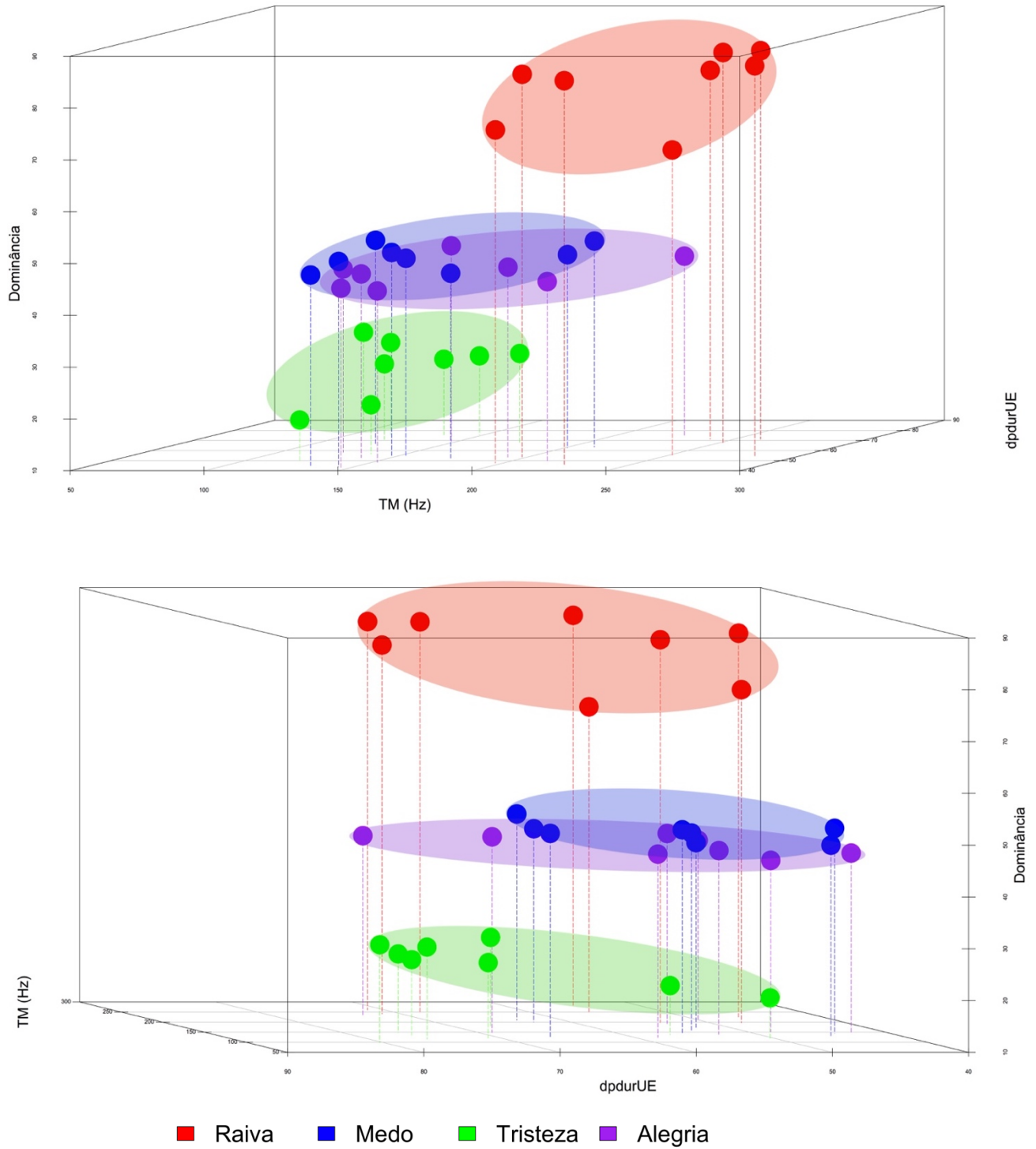


Gráfico 23: Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R^2 : 0,73

Gráficos de dispersão 3D - Dominância | TM e durintUE

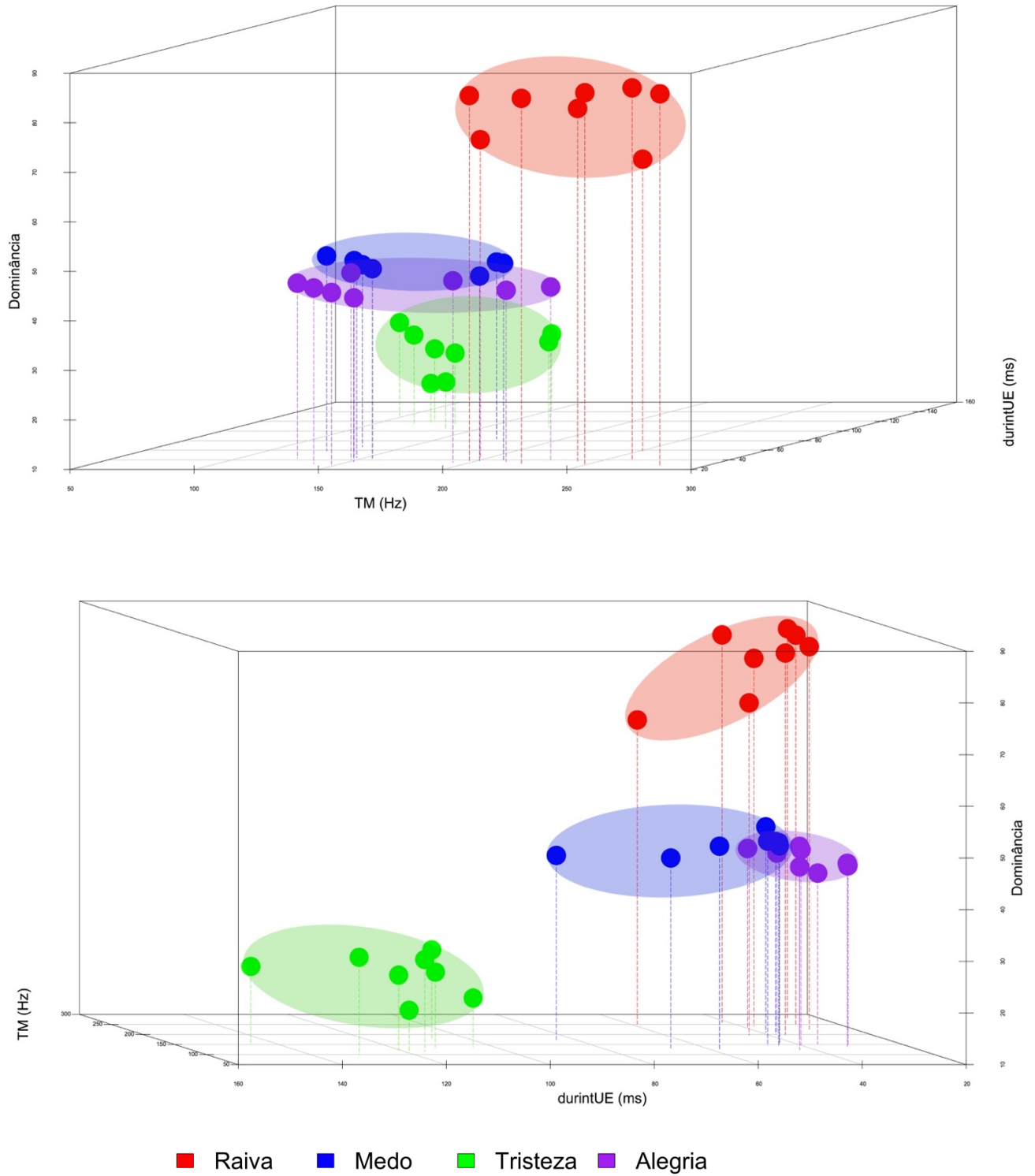


Gráfico 24: Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + desvio padrão da duração dos intervalos entre as unidades entoacionais - dpdurintUE). Valor de R²: 0,66

Gráficos de dispersão 3D - Dominância | TM e dpdurintUE

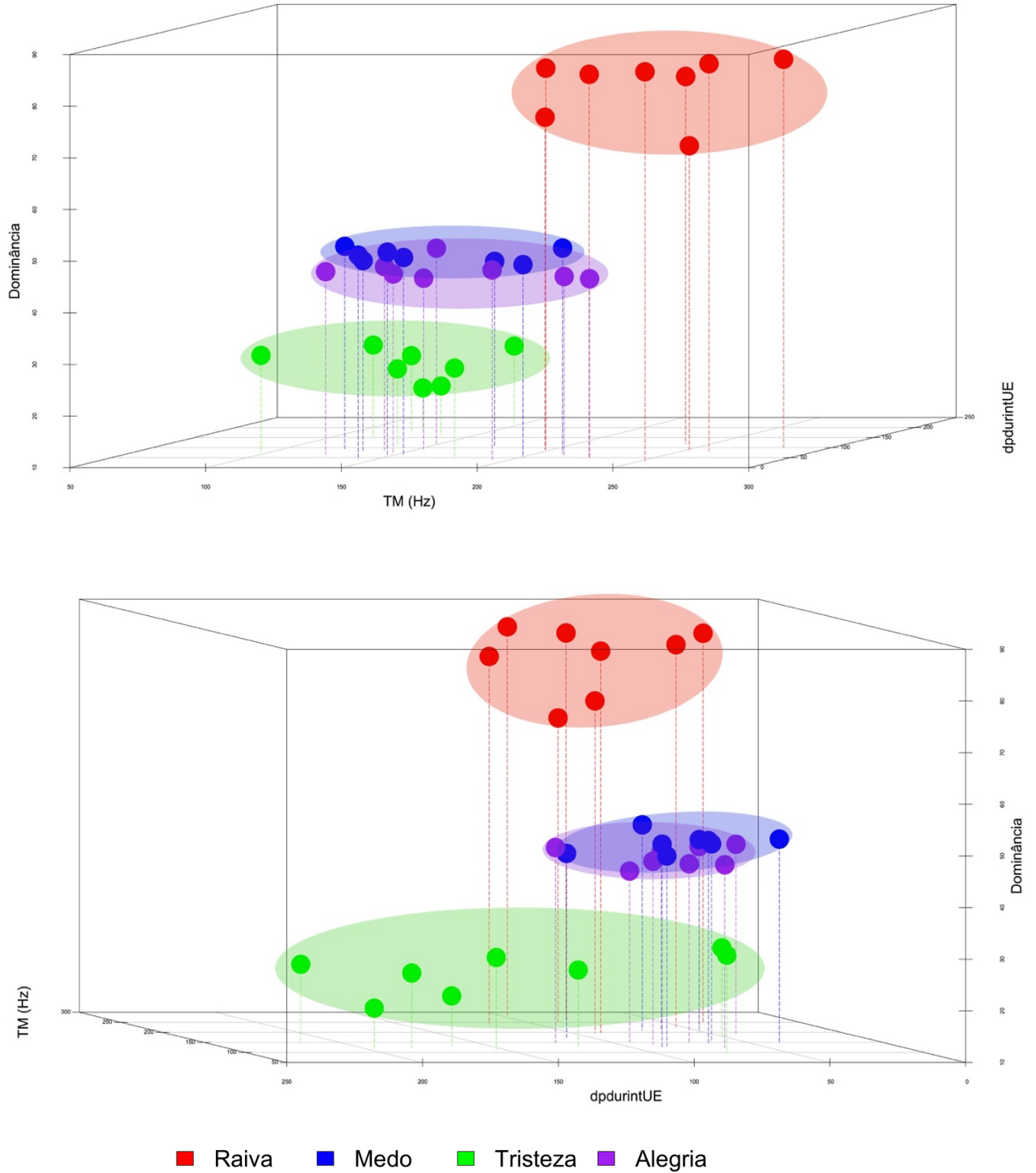
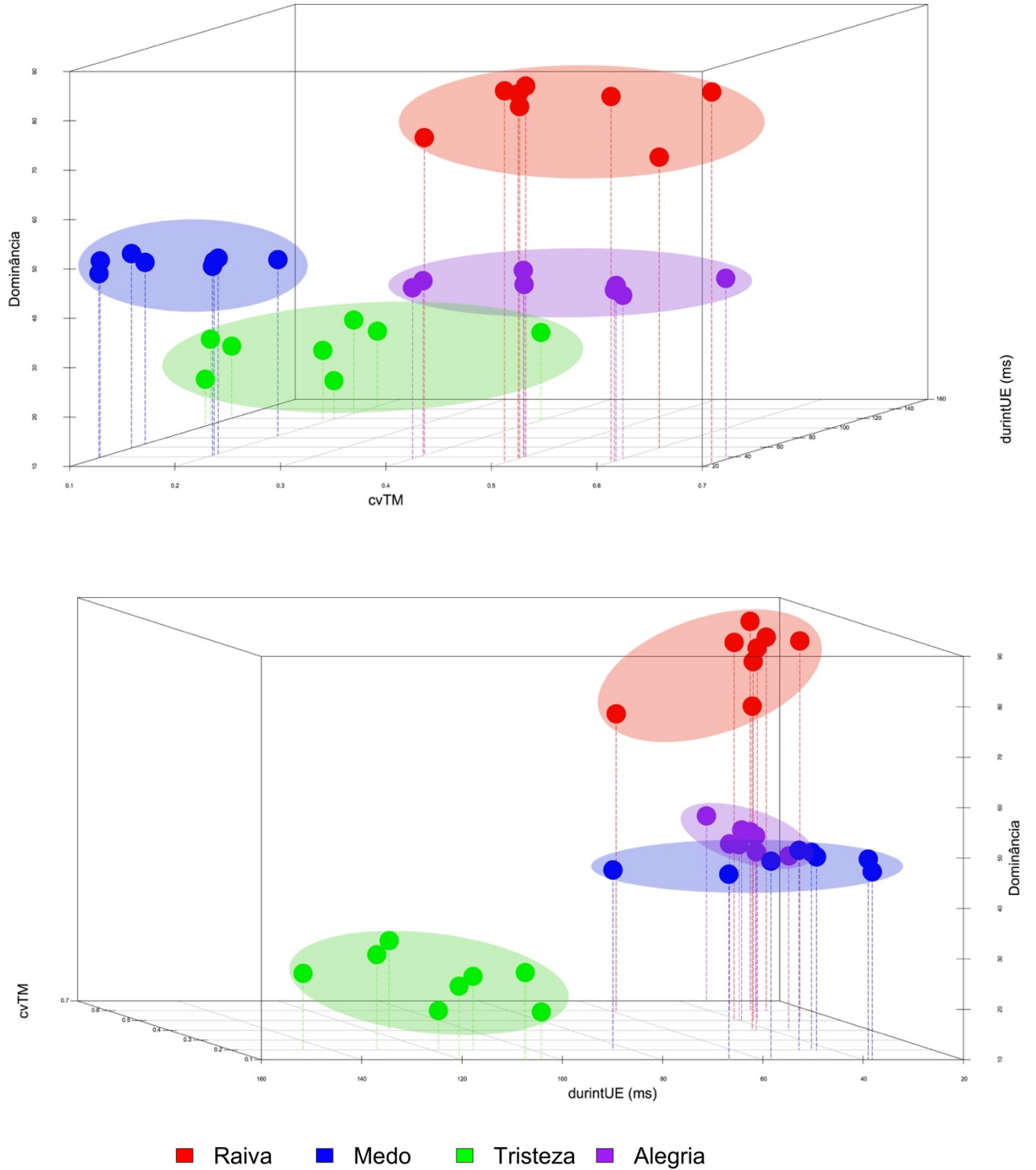


Gráfico 25: Dois ângulos diferentes da distribuição de julgamentos de dominância (Bra – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos entre as unidades entoacionais – durintUE).
 Valor de R^2 : 0,54

Gráficos de dispersão 3D - Dominância | cvTM e durintUE



Os julgamentos de ativação e dominância pelos nativos mostraram-se diferentes quanto à sua dispersão. A ativação, de modo geral, teve seus valores divididos em dois grupos antagônicos, i.e., regiões de alta ativação (raiva e alegria) e de baixa (medo e tristeza). No caso da dominância, a formação dos *clusters* foi dividida em três regiões. Na região de baixa dominância, a emoção correlata foi a tristeza; na região média, medo e alegria; na alta, raiva. Os gráficos de 21 a 25 ilustram bem essa divisão (p. 137 – 141).

Todavia, a interação entre dominância e parâmetros acústicos aconteceu de maneira semelhante à encontrada na ativação. Em outras palavras, os gráficos entre dominância e combinações de parâmetros acústicos apresentaram também uma relação positiva entre as notas dadas pelos participantes e os valores de TM. Assim, tristeza e medo ficaram em regiões de baixa dominância e de baixo valor de TM. O inverso ocorreu com os valores de duração das unidades entoacionais, bem como os intervalos entre elas. Menores valores de *durUE*, *dpdurUE*, *durintUE* e *dpdurintUE* estão ligados a valores médios e altos de dominância.

O tom médio (TM + *durUE* e TM + *dpdurUE*) e os julgamentos de dominância apresentaram, em regra, valores sobrepostos para as emoções (sobretudo no eixo referente ao TM), à medida que os valores de dominância foram aumentando. Os gráficos 21 e 22 (p. 137 e 138) ilustram bem essa relação. Por outro lado, quando TM foi combinado com *durintUE* e *dpdurintUE* (gráficos 23 e 24), a emoção raiva destacou-se um pouco das demais (p. 139 – 140).

Com relação aos parâmetros rítmicos, a interação com os julgamentos de dominância apresentou três configurações. A primeira dividiu as emoções em medo e tristeza, de um lado; raiva e alegria, de outro (Gráfico 21, ângulo 2 com *durUE* representado no eixo X); a segunda deixou a tristeza isolada das demais emoções, com valores elevados de *durintUE*

(gráficos 23 e 25); a terceira, com o parâmetro $dpdurUE$, formou *clusters* quase na mesma região (Gráfico 22).

Passando agora para os resultados dos não nativos, as regressões lineares múltiplas tiveram alguns resultados significativos. Para ativação, as combinações relevantes foram: $TM + durUE$ (0,85), $TM + durintUE$ (0,69), $cvTM + durUE$ (0,86), $cvTM + dpdurUE$ (0,77), $cvTM + durintUE$ (0,87) e $cvTM + dpdurintUE$ (0,78). Abaixo, os gráficos tridimensionais de dispersão ilustram os resultados para ativação – participantes não nativos:

Gráfico 26: Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,85

Gráficos de dispersão 3D - Ativação | TM e durUE

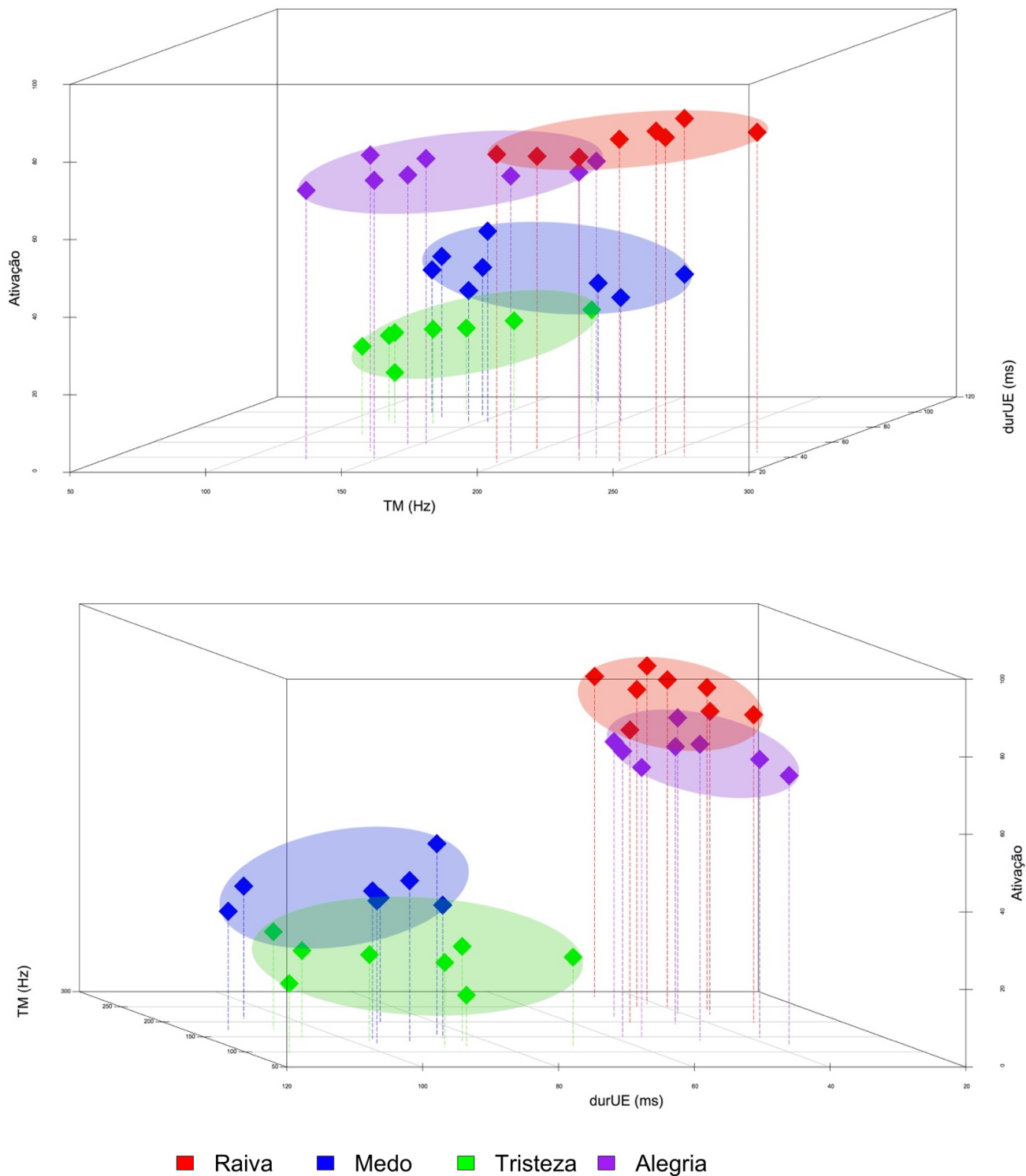


Gráfico 27: Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE). Valor de R^2 : 0,69

Gráficos de dispersão 3D - Ativação | TM e durintUE

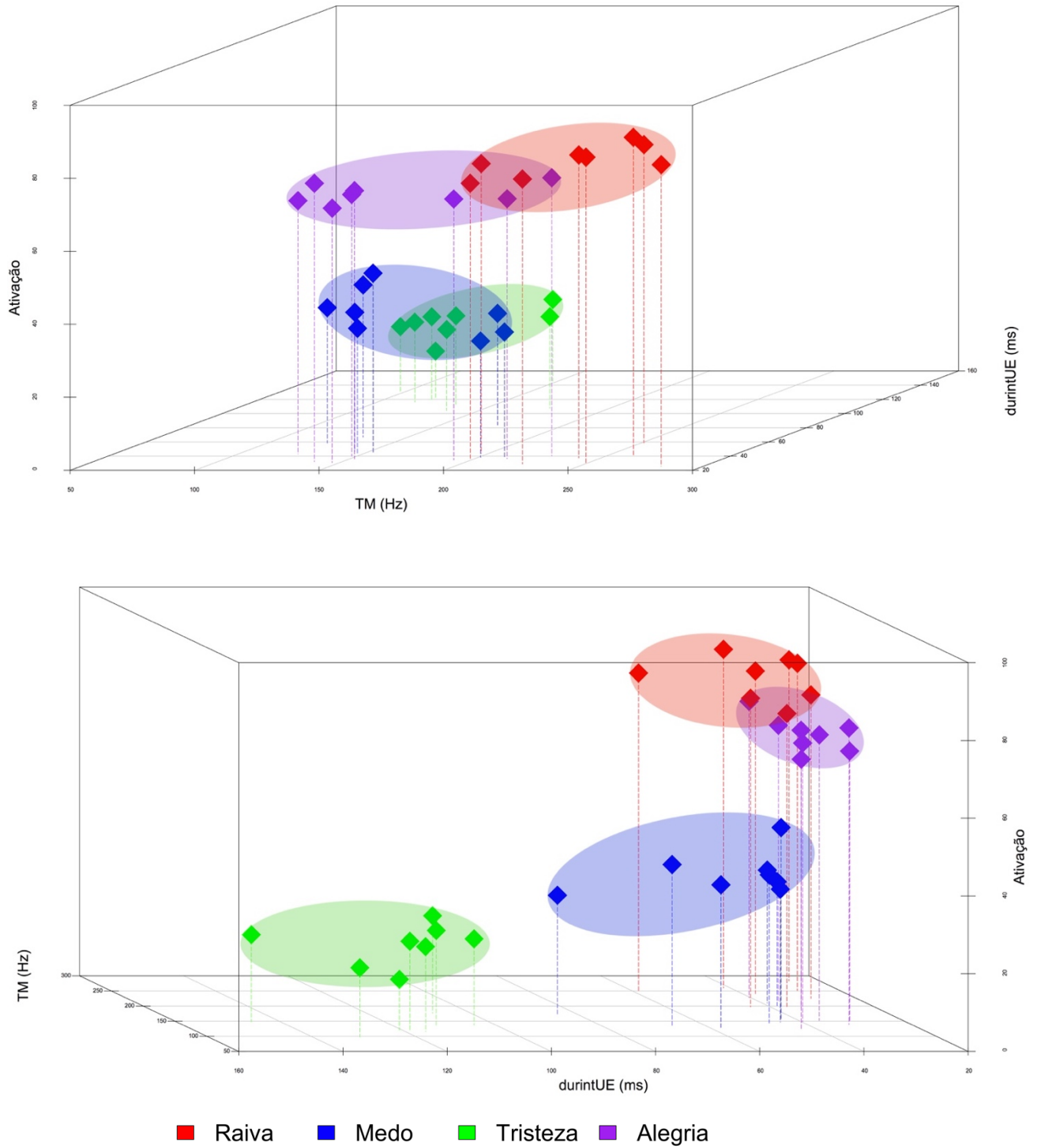


Gráfico 28: Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração das unidades entoacionais – durUE). Valor de R²: 0,86

Gráficos de dispersão 3D - Ativação | cvTM e durUE

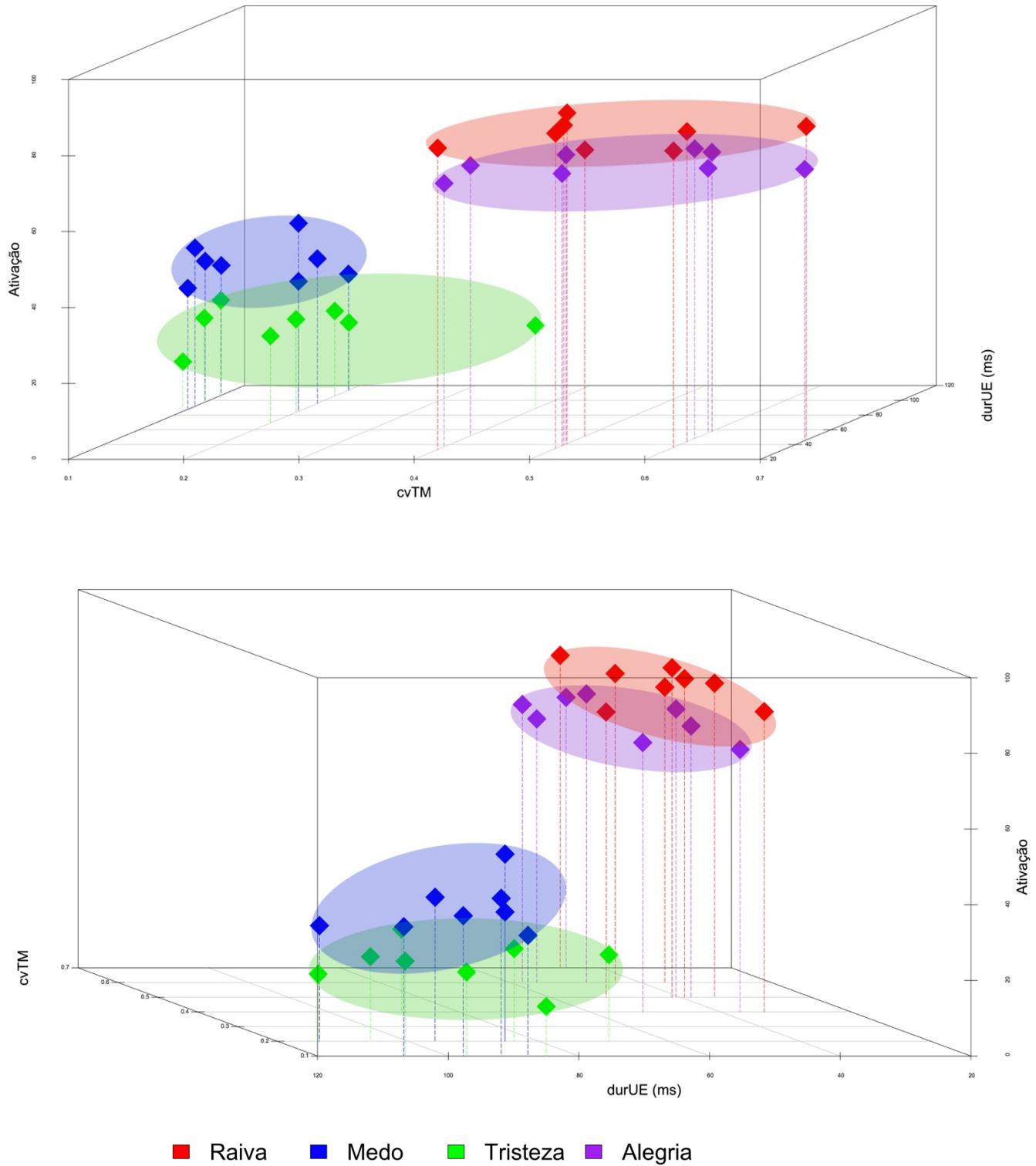


Gráfico 29: Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração das unidades entoacionais – dpdurUE). Valor de R^2 : 0,77

Gráficos de dispersão 3D - Ativação | cvTM e dpdurUE

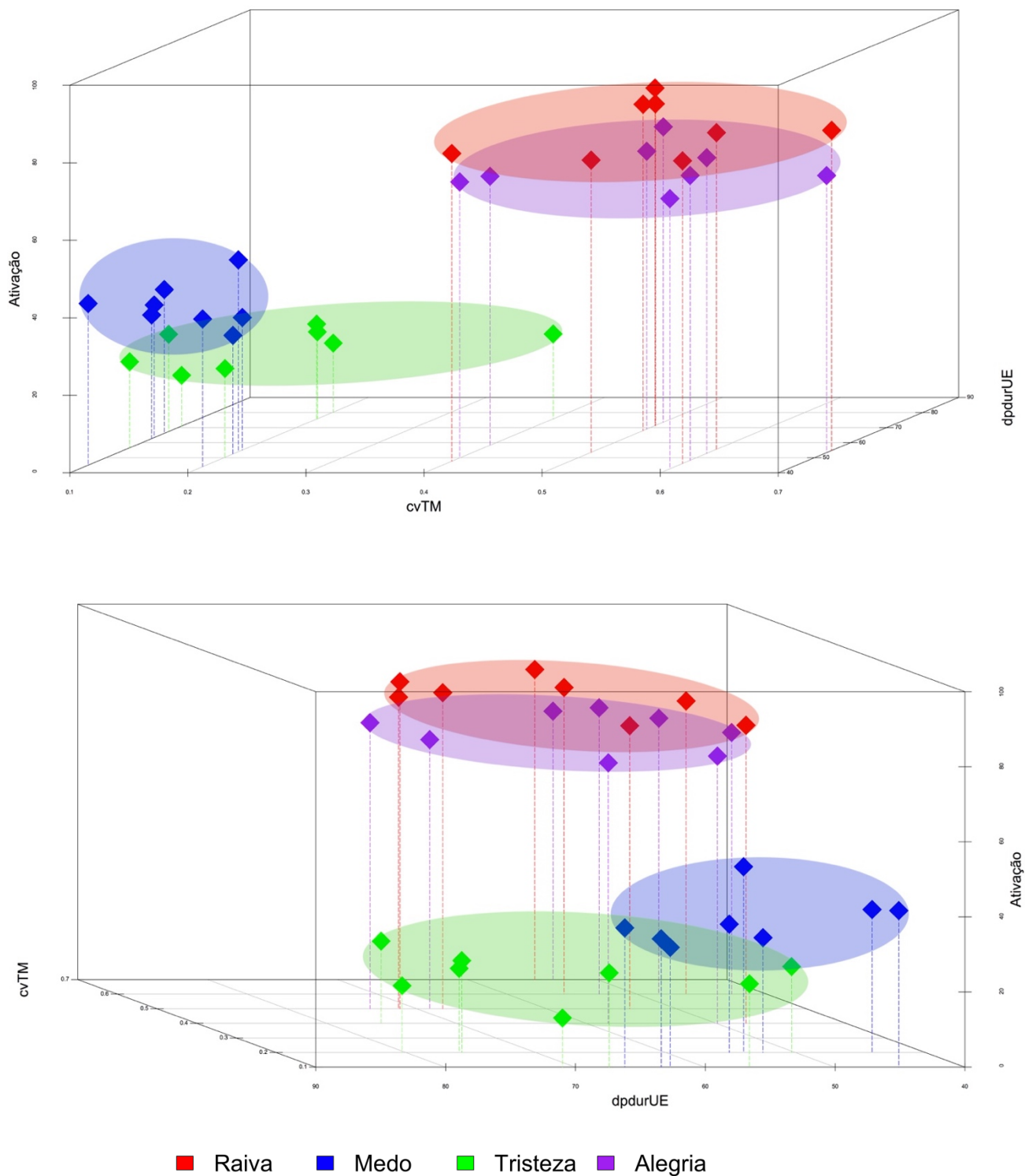


Gráfico 30: Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos das unidades entoacionais – durintUE). Valor de $R^2: 0,87$

Gráficos de dispersão 3D - Ativação | cvTM e durintUE

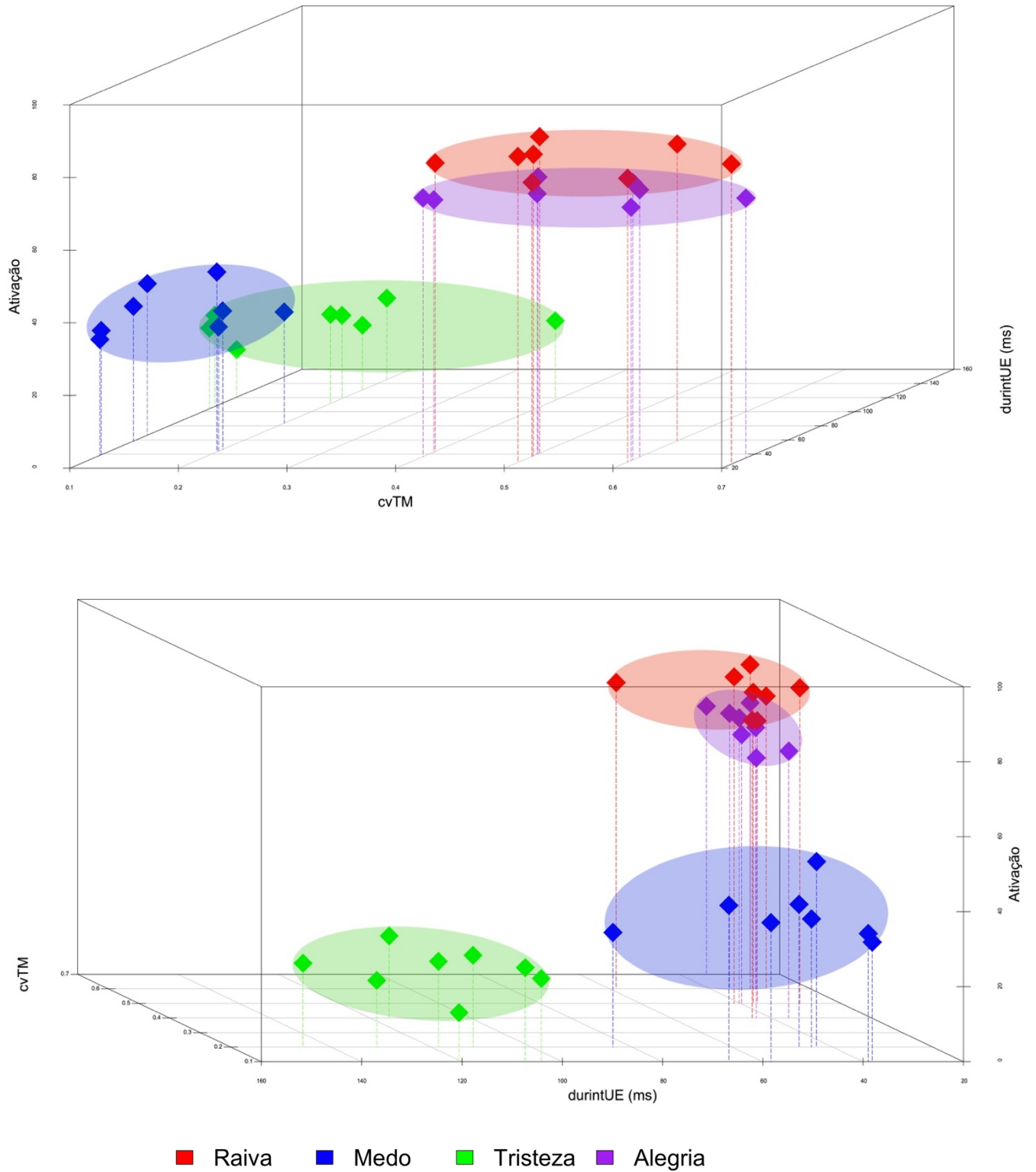
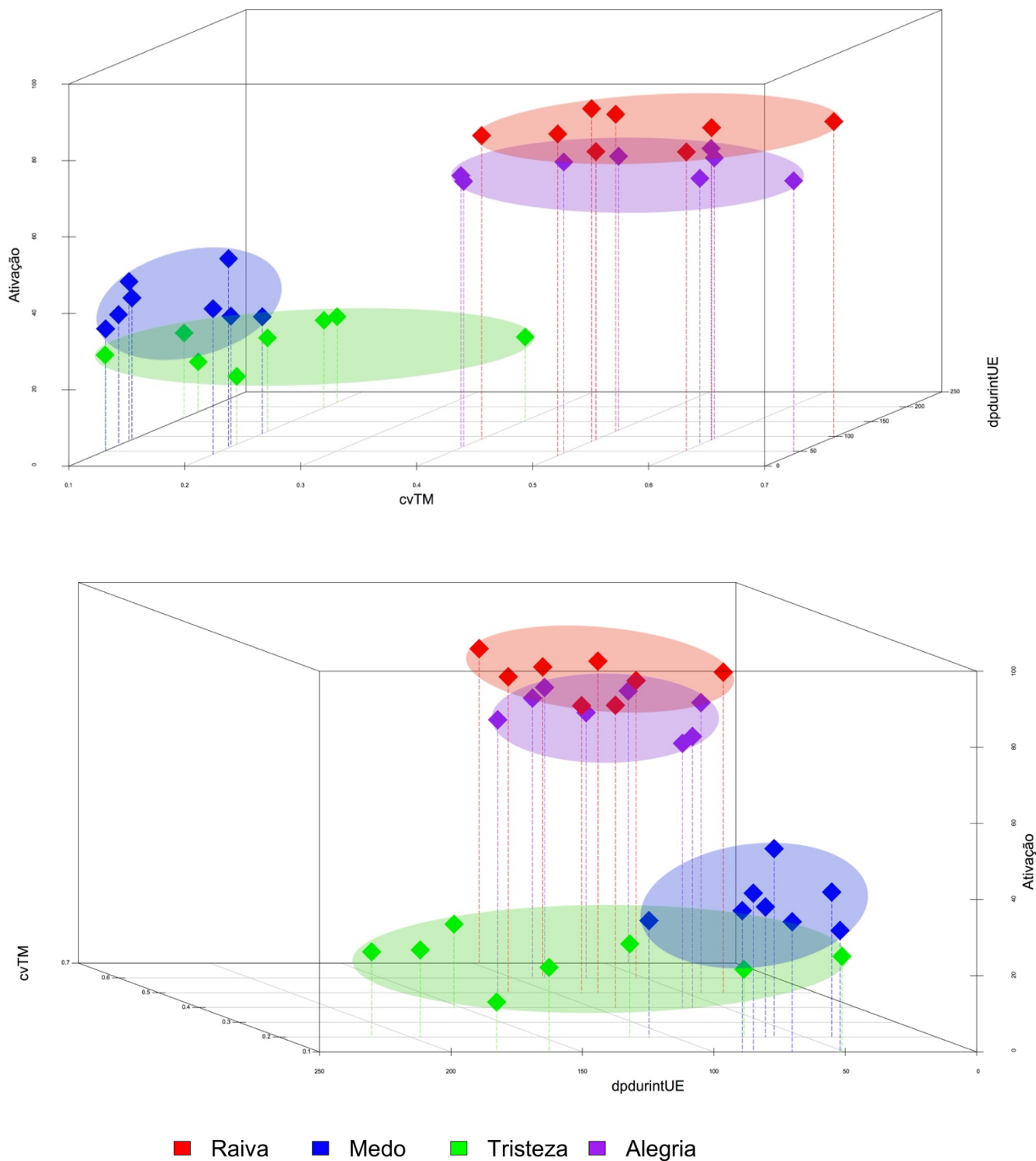


Gráfico 31: Dois ângulos diferentes da distribuição de julgamentos de ativação (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração dos intervalos das unidades entoacionais – dpdurintUE). Valor de R²: 0,78

Gráficos de dispersão 3D - Ativação | cvTM e dpdurintUE



Embora os valores de R^2 tenham sido significativos, os participantes ingleses apresentaram algumas diferenças quanto à distribuição de notas na dimensão ativação. Uma característica é uma distinção não tão clara de *clusters*, como no caso dos falantes nativos. A ativação, de modo geral, apresentou a mesma configuração encontrada nos nativos, ou seja, as emoções foram divididas em dois grupos, medo e tristeza, raiva e alegria. Os valores de $TM + durUE$ também apresentaram pouca definição, principalmente no eixo dos valores de TM (Gráfico 26, p. 144). O parâmetro $cvTM$ apresentou resultados semelhantes aos dos nativos, separando medo e tristeza de raiva e alegria (gráficos de 28 a 31, p.146 – 149). Já $durintUE$ isolou a tristeza das demais emoções (gráficos 27 e 30, p. 145 – 148).

Para dominância, as combinações relevantes foram: $TM + durUE$ (0,71), $TM + durintUE$ (0,75), $cvTM + durUE$ (0,68), $cvTM + dpdurUE$ (0,61), $cvTM + durintUE$ (0,79) e $cvTM + dpdurintUE$ (0,66). Abaixo, os gráficos tridimensionais de dispersão ilustram os resultados para dominância – participantes não nativos:

Gráfico 32: Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,71

Gráficos de dispersão 3D - Dominância | TM e durUE

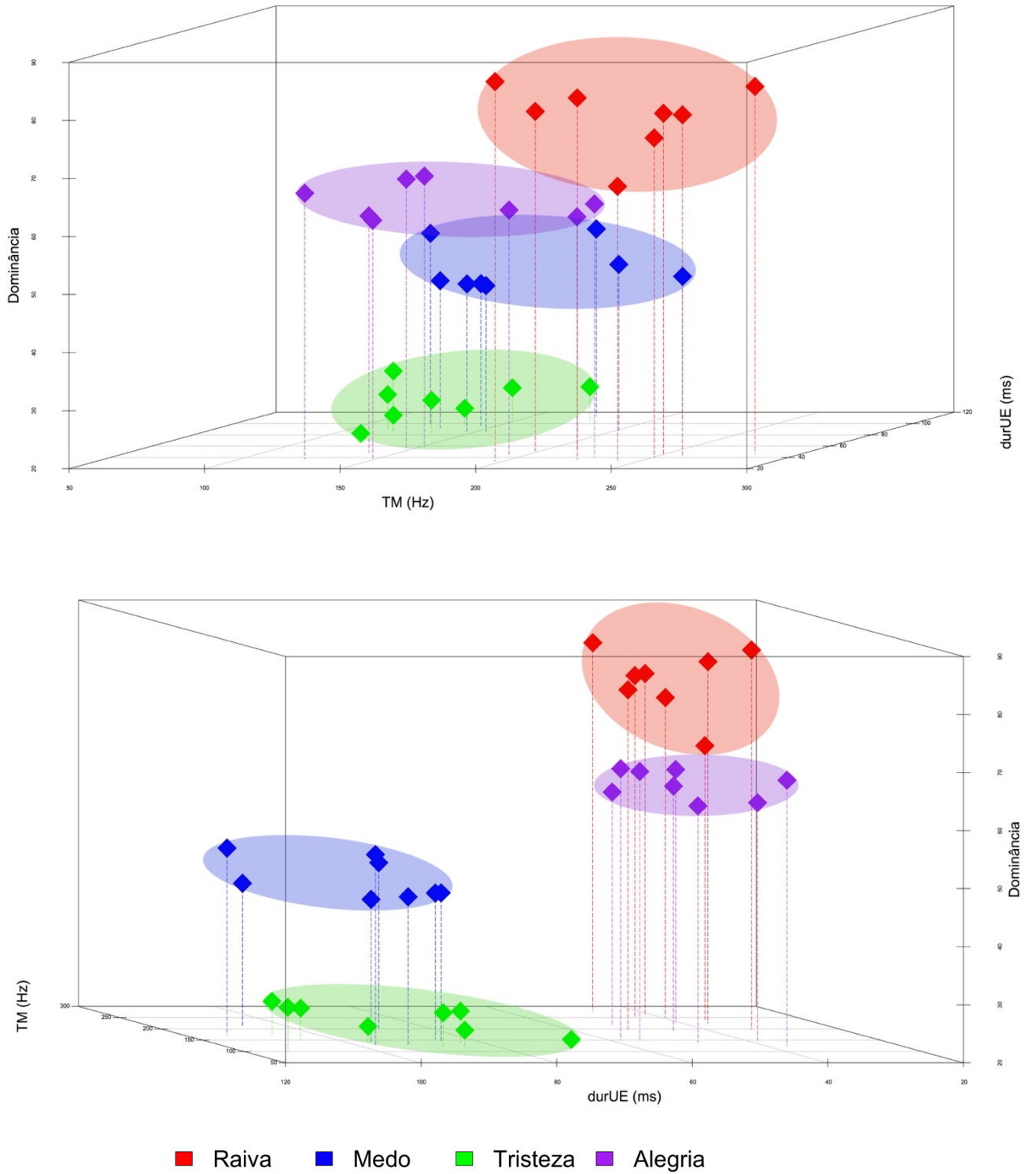


Gráfico 33: Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (tom médio – TM + duração do intervalo entre as unidades entoacionais – durintUE). Valor de R²: 0,75

Gráficos de dispersão 3D - Dominância | TM e durintUE

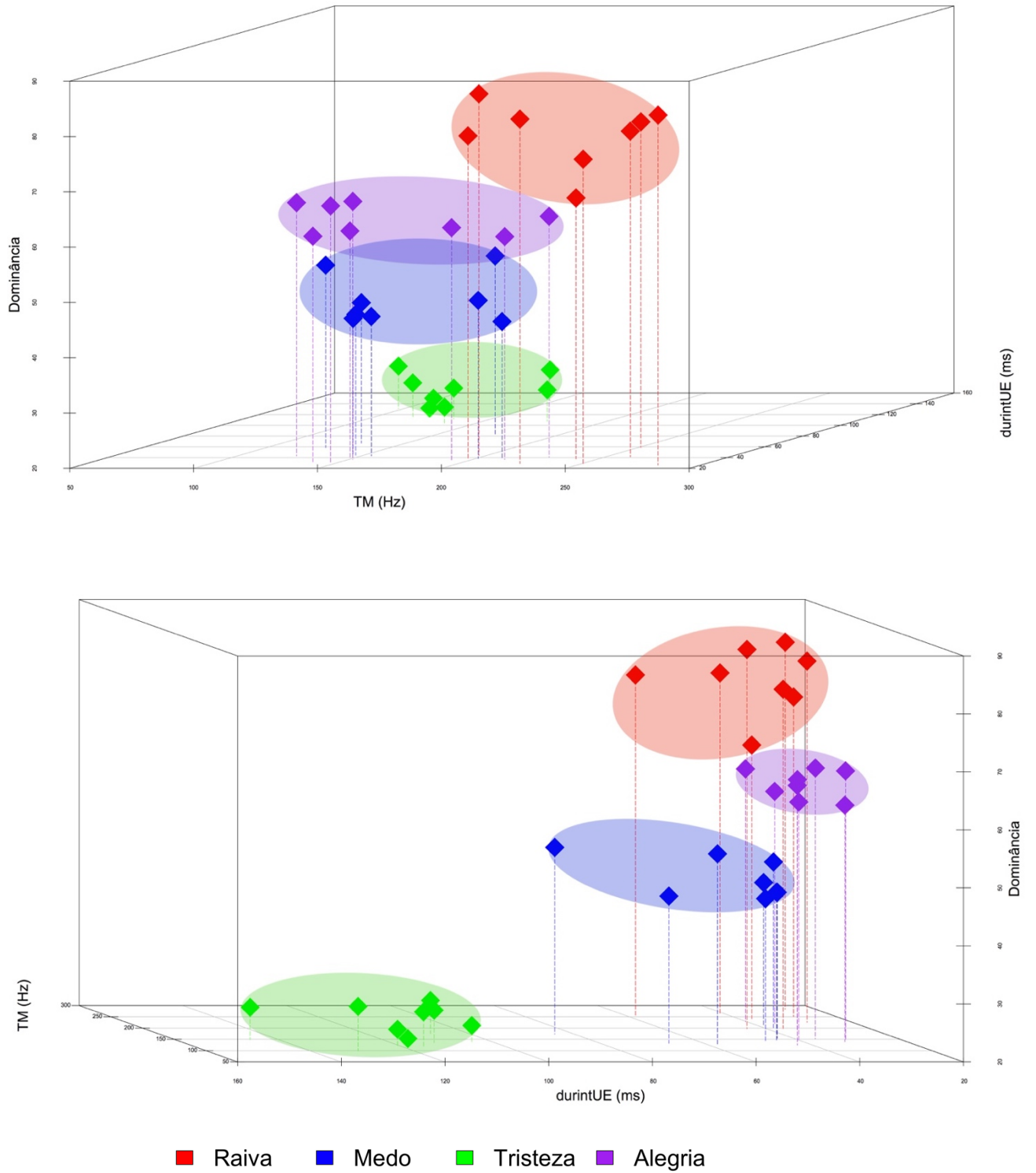


Gráfico 34: Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração das unidades entoacionais – durUE). Valor de R²: 0,68

Gráficos de dispersão 3D - Dominância | cvTM e durUE

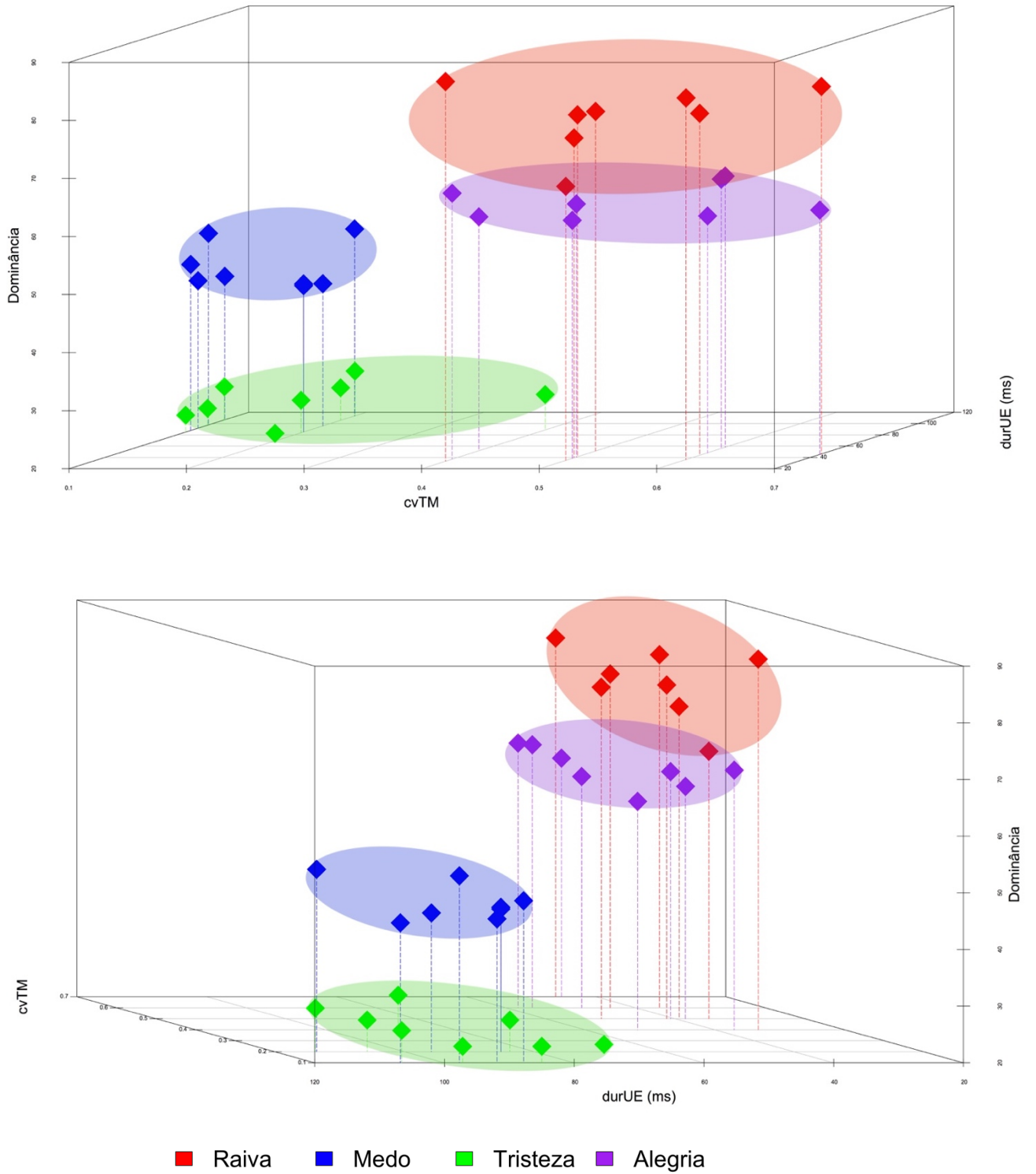


Gráfico 35: Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração das unidades entoacionais – dpdurUE). Valor de R^2 : 0,61

Gráficos de dispersão 3D - Dominância | cvTM e dpdurUE

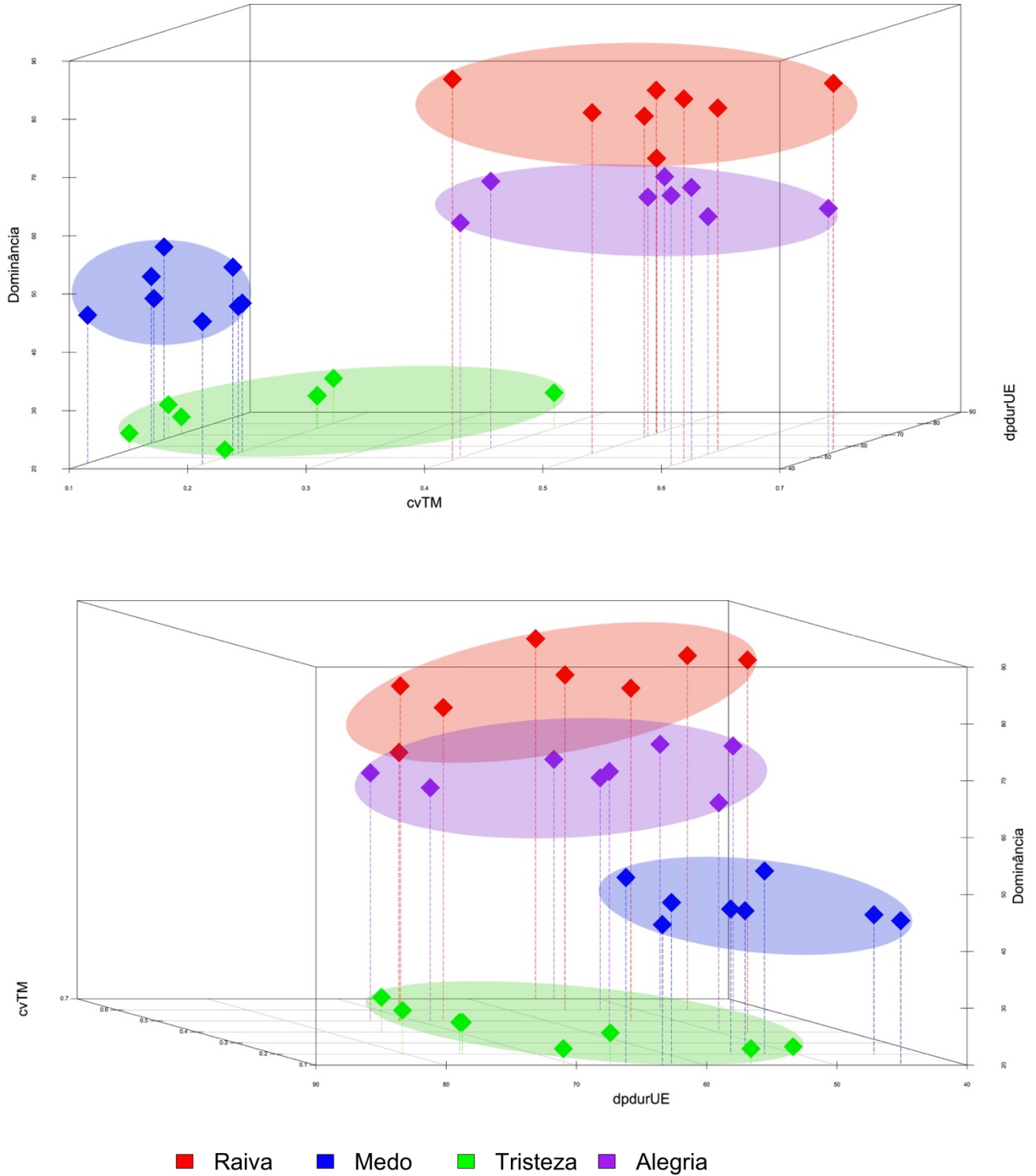


Gráfico 36: Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + duração dos intervalos entre as unidades entoacionais – durintUE).
 Valor de R²: 0,79

Gráficos de dispersão 3D - Dominância | cvTM e durintUE

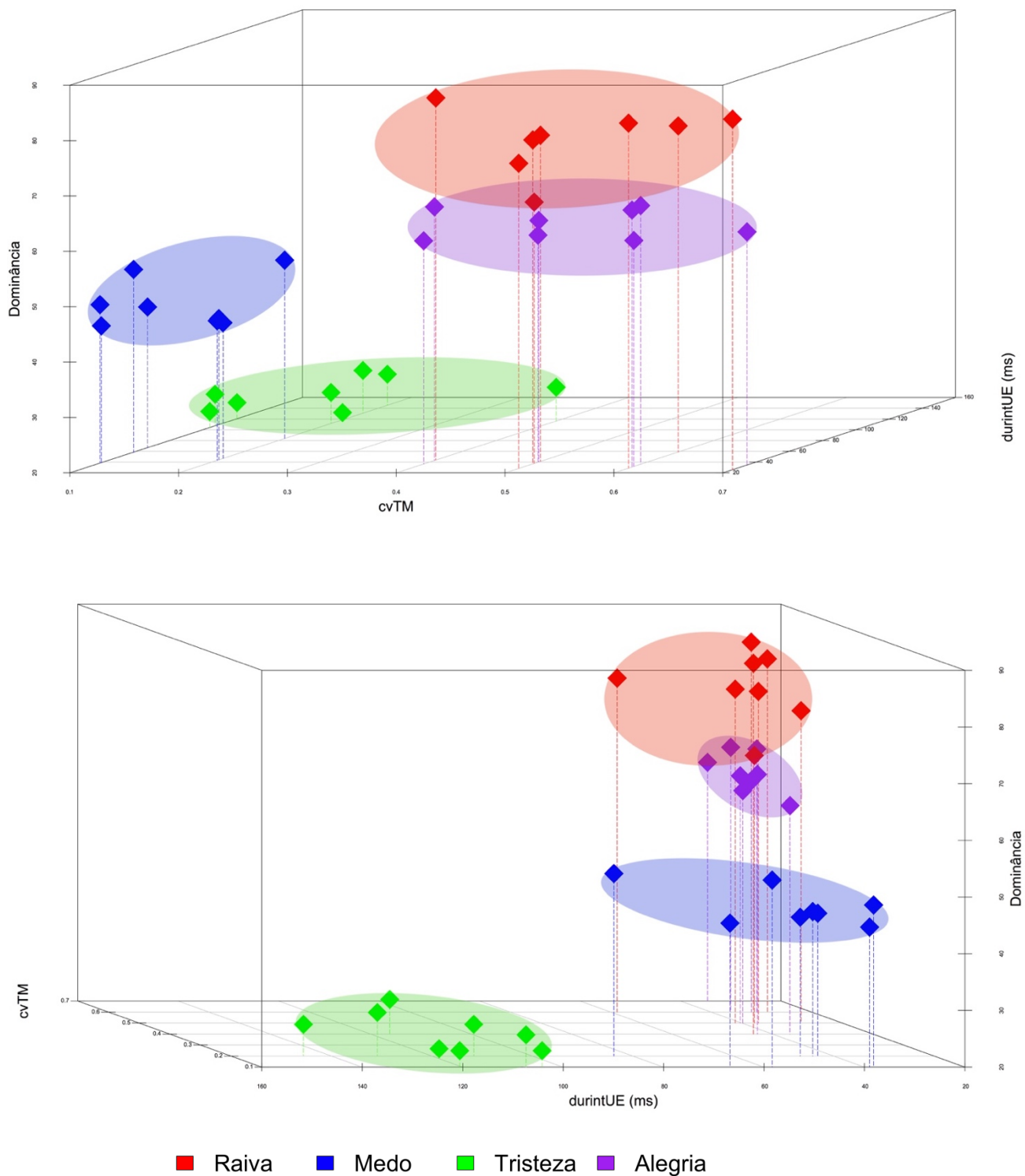
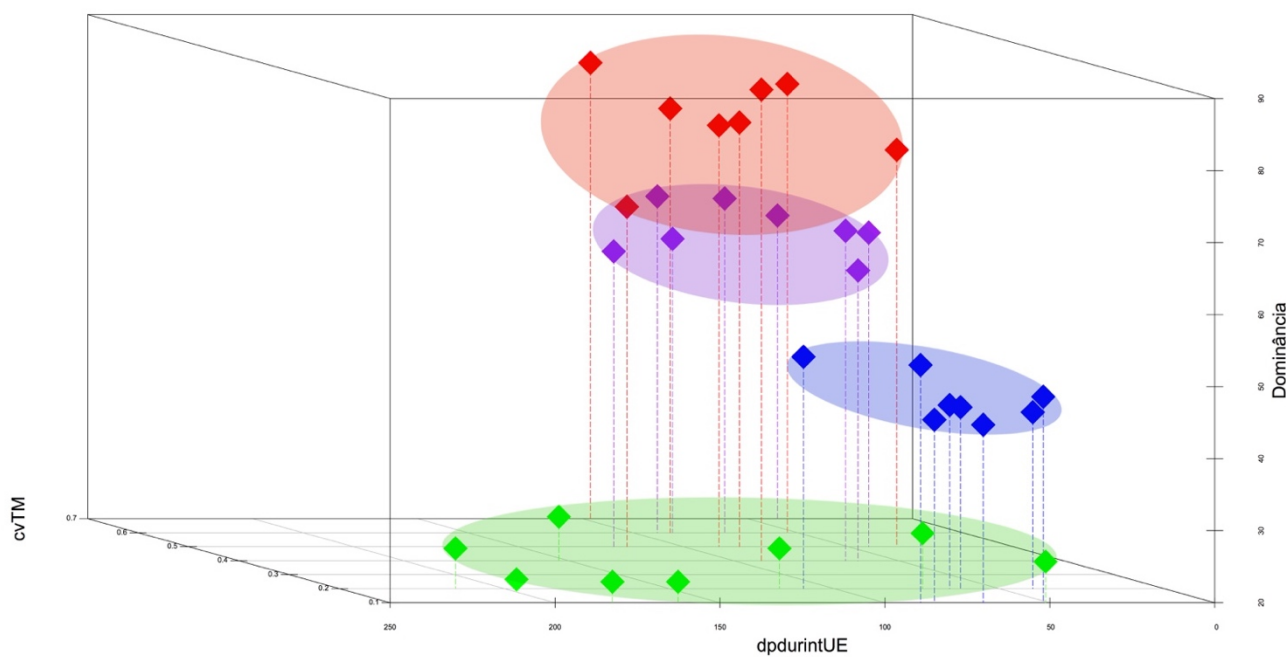
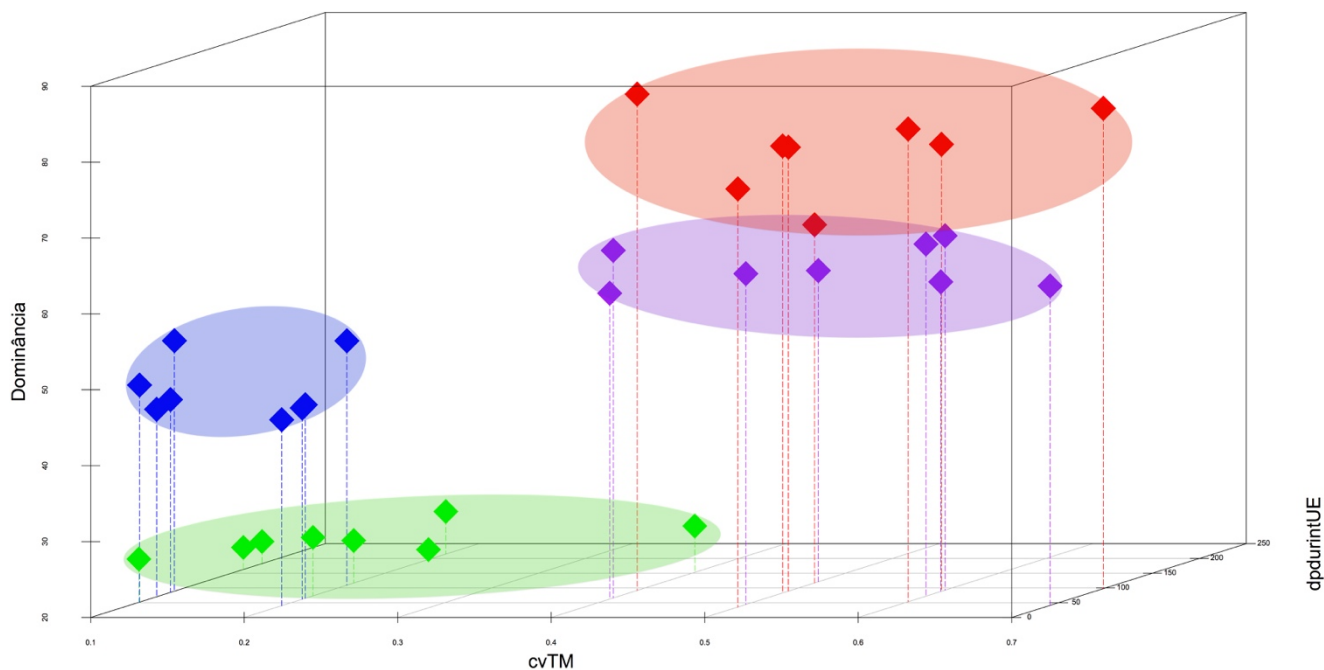


Gráfico 37: Dois ângulos diferentes da distribuição de julgamentos de dominância (Ing – fala normal) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – cvTM + desvio padrão da duração dos intervalos entre as unidades entoacionais – dpdurintUE). Valor de R²: 0,66

Gráficos de dispersão 3D - Dominância | cvTM e dpdurintUE



■ Raiva ■ Medo ■ Tristeza ■ Alegria

Da mesma forma como ocorrido com ativação, valores elevados de dominância atribuídos pelos não nativos foram associados a altos valores de TM. No caso das emoções, a alegria e a raiva formaram dois *clusters* distintos quando relacionadas ao TM (gráficos 32 e 33, p. 151 – 152). O Gráfico 34, envolvendo cvTM e durUE, apresentou divisões mais claras entre as emoções. Altos valores de durUE, relacionados a menor cvTM e menor dominância, agruparam a emoção tristeza. O medo teve uma configuração muito parecida, exceto na dominância, que apresentou valores medianos. O inverso dessa combinação configura-se como raiva, isto é, valores elevados de dominância foram associados a baixos valores de durUE e valores elevados de cvTM. Igualmente ao medo e à tristeza, a alegria formou um *cluster* à parte da raiva, obtendo valores medianos de dominância, mas ainda assim foram mais elevados que o medo. A formação mais clara de *clusters* deve-se em parte pelos níveis de dominância atribuídos pelos participantes não nativos. Nota-se que, nos gráficos de 32 a 37 (p. 151 – 156), os níveis de dominância podem ser divididos em quatro partes, cada um deles relacionado com uma emoção em particular. Essa configuração dos julgamentos de dominância dados pelos não nativos pode se dizer previsível, haja vista a variação encontrada nos dados com falantes nativos, na qual houve diferentes tipos de agrupamentos de acordo com os julgamentos de dominância e parâmetros de duração (ver pág. 142).

O próximo passo da análise são os resultados dos experimentos com fala delexicalizada. As tabelas 12 e 13 (a seguir) apresentam os resultados das regressões lineares múltiplas envolvendo os julgamentos dos informantes brasileiros e ingleses sobre os estímulos com fala delexicalizada. As regressões lineares múltiplas apresentaram poucos resultados significativos para os brasileiros. Para ativação, as combinações relevantes foram: TM + durintUE (0,66) e cvTM + durintUE (0,71). No caso dos ingleses, não houve resultado significativo para nenhuma das dimensões e parâmetros acústicos.

Resultado das Regressões lineares múltiplas (valores de r² ajustado) Parâmetros acústicos - ExProsodia

Bra. – fala delexicalizada

	TM + durUE	TM + durintUE	TM + dpdurintUE	cvTM + durUE	cvTM + dpdurUE	cvTM + durintUE	cvTM + dpdurintUE
<i>Valência</i>	0,36	0,08	0,02	0,15	0,12	0,09	0,09
<i>Ativação</i>	0,34	0,66	0,38	0,36	0,54	0,71	0,49
<i>Dominância</i>	0,61	0,47	0,18	0,57	0,54	0,59	0,55

Tabela 12: Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes brasileiros para cada dimensão (fala delexicalizada) e os parâmetros acústicos combinados – ExProsodia.

Ing. – fala delexicalizada

	TM + durUE	TM + durintUE	TM + dpdurintUE	cvTM + durUE	cvTM + dpdurUE	cvTM + durintUE	cvTM + dpdurintUE
<i>Valência</i>	0,42	0,28	0,15	0,51	0,48	0,47	0,47
<i>Ativação</i>	0,20	0,39	0,16	0,27	0,31	0,46	0,38
<i>Dominância</i>	0,41	0,44	0,04	0,40	0,33	0,46	0,37

Tabela 13: Resultado das regressões lineares múltiplas envolvendo os julgamentos dados pelos participantes ingleses para cada dimensão (fala delexicalizada) e os parâmetros acústicos combinados – ExProsodia.

(TM) - tom médio; (cvTM) - coeficiente de variação do tom médio; (durUE) - duração das unidades entoacionais; (dpdurUE) - desvio padrão da duração das unidades entoacionais; (durintUE) - duração dos intervalos entre as unidades entoacionais; (dpdurintUE) - desvio padrão da duração dos intervalos entre as unidades entoacionais

Gráfico 38: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (tom médio –TM + duração dos intervalos entre as unidades entoacionais –durintUE). Valor de R^2 : 0,66

Gráficos de dispersão 3D - Ativação | TM e durintUE

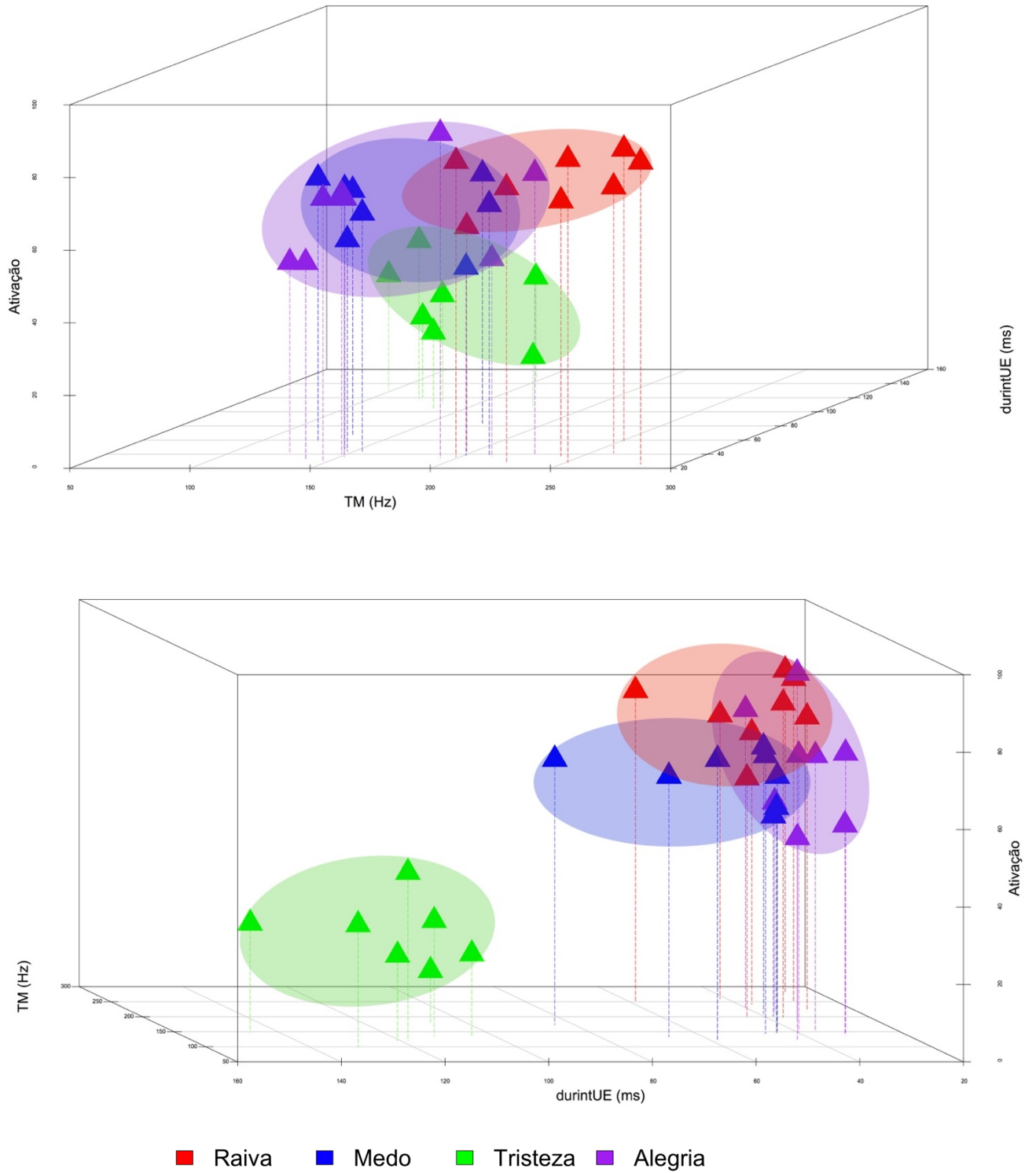
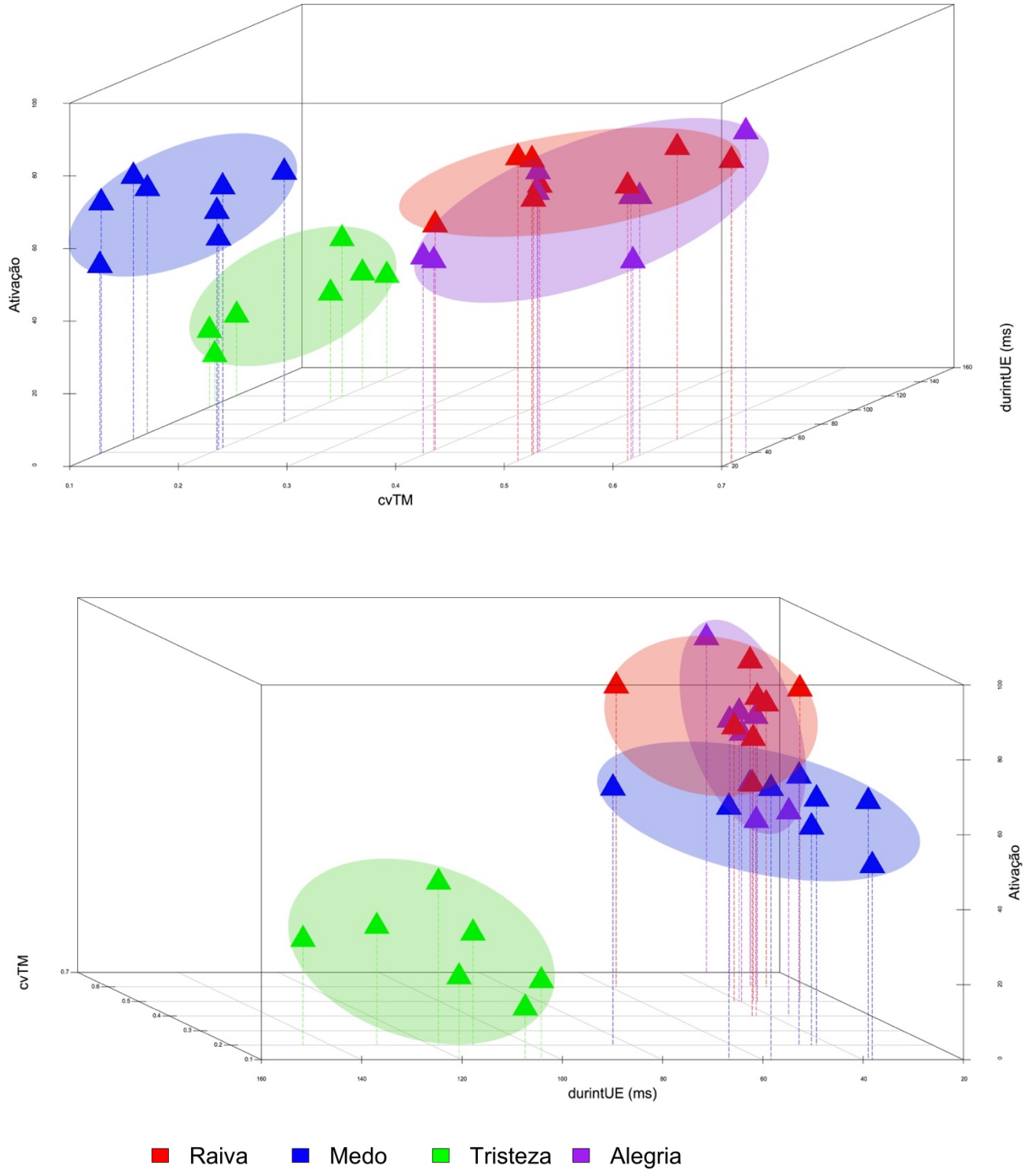


Gráfico 39: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio –cvTM + duração dos intervalos entre as unidades entoacionais –durintUE).
 Valor de R²: 0,71

Gráficos de dispersão 3D - Ativação | cvTM e durintUE



As emoções, segundo o julgamento de ativação dos participantes nativos para fala delexicalizada, puderam ser divididas em dois grupos, entretanto diferentes daqueles encontrados nos resultados de fala normal para os mesmos participantes. Em fala normal, medo e tristeza ficaram delimitados a um mesmo *cluster*, raiva e alegria a um outro (e.g., Gráfico 15, p. 130). No caso dos julgamentos de ativação sobre fala delexicalizada, os *clusters* tiveram uma formação diferente. No Gráfico 38, de acordo com os valores de TM, os agrupamentos das emoções apresentam muitas sobreposições, mas, quando se observa do ponto de vista do parâmetro acústicos durintUE, nota-se de maneira inequívoca a formação de um *cluster* de tristeza (segunda figura do Gráfico 38), ao passo que as demais emoções ficaram reunidas num mesmo *cluster*. No Gráfico 39, cvTM obteve melhor resultado que TM e dividiu as emoções em dois grupos tal qual encontrado nos resultados de ativação com fala normal (como supracitado). O último passo da análise deste experimento é o resultado dos julgamentos de dominância e as combinações relevantes dos parâmetros acústicos. Foram eles: TM + durUE (0,61) e cvTM + durintUE (0,59).

Os resultados de dominância para fala delexicalizada apresentaram um agrupamento semelhante ao encontrado nos julgamentos de ativação e fala normal (participantes nativos – Bra), com tristeza e medo de um lado, raiva e alegria, de outro. Esses resultados estão relacionados às variações dos parâmetros de duração, durUE e durintUE (Gráficos 40 e 41). Os valores de TM não apresentaram agrupamentos definidos, já os valores de cvTM apresentaram a mesma divisão supracitada.

Gráfico 40: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (tom médio – TM + duração das unidades entoacionais – durUE). Valor de R^2 : 0,61

Gráficos de dispersão 3D - Dominância | TM e durUE

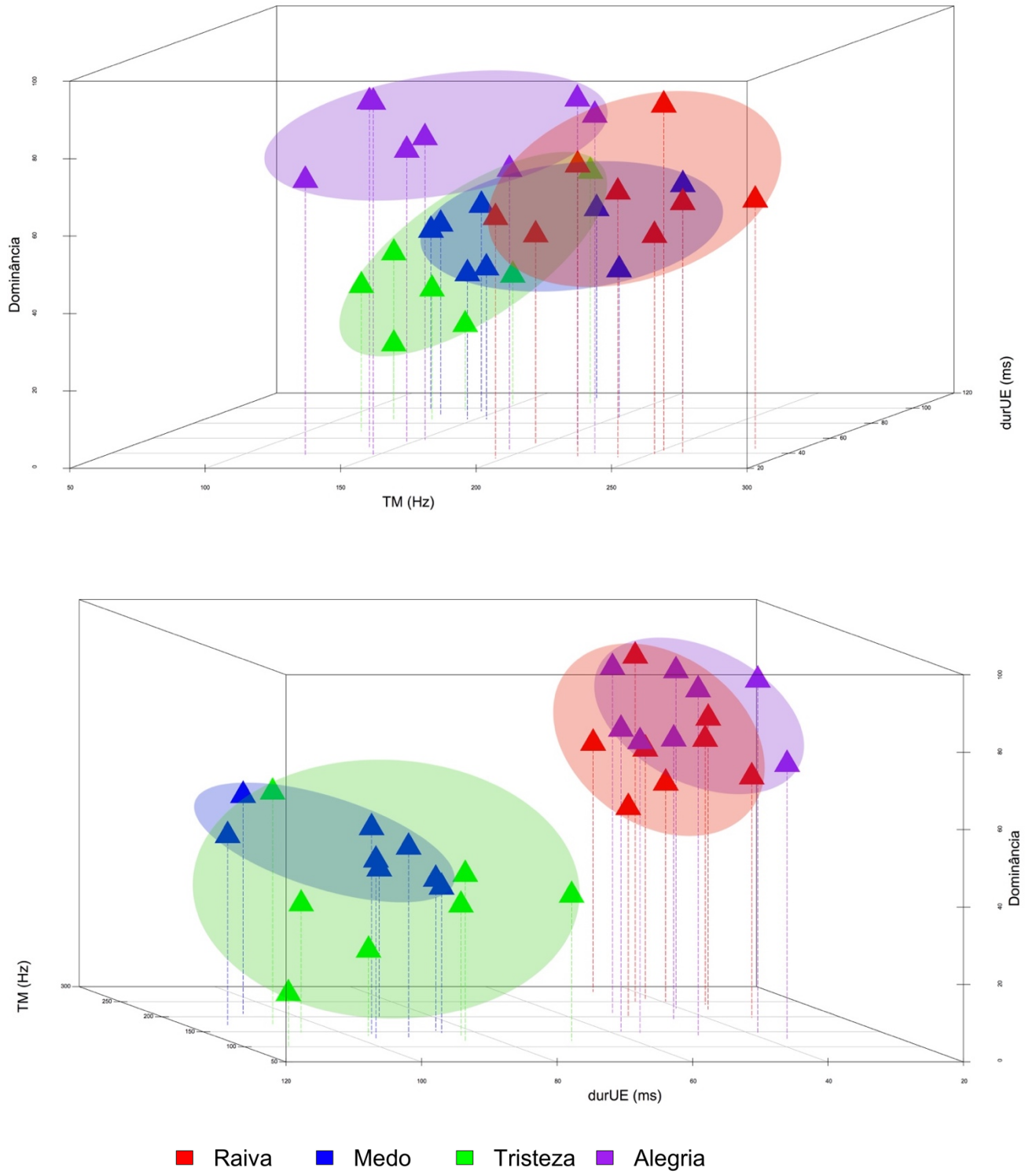
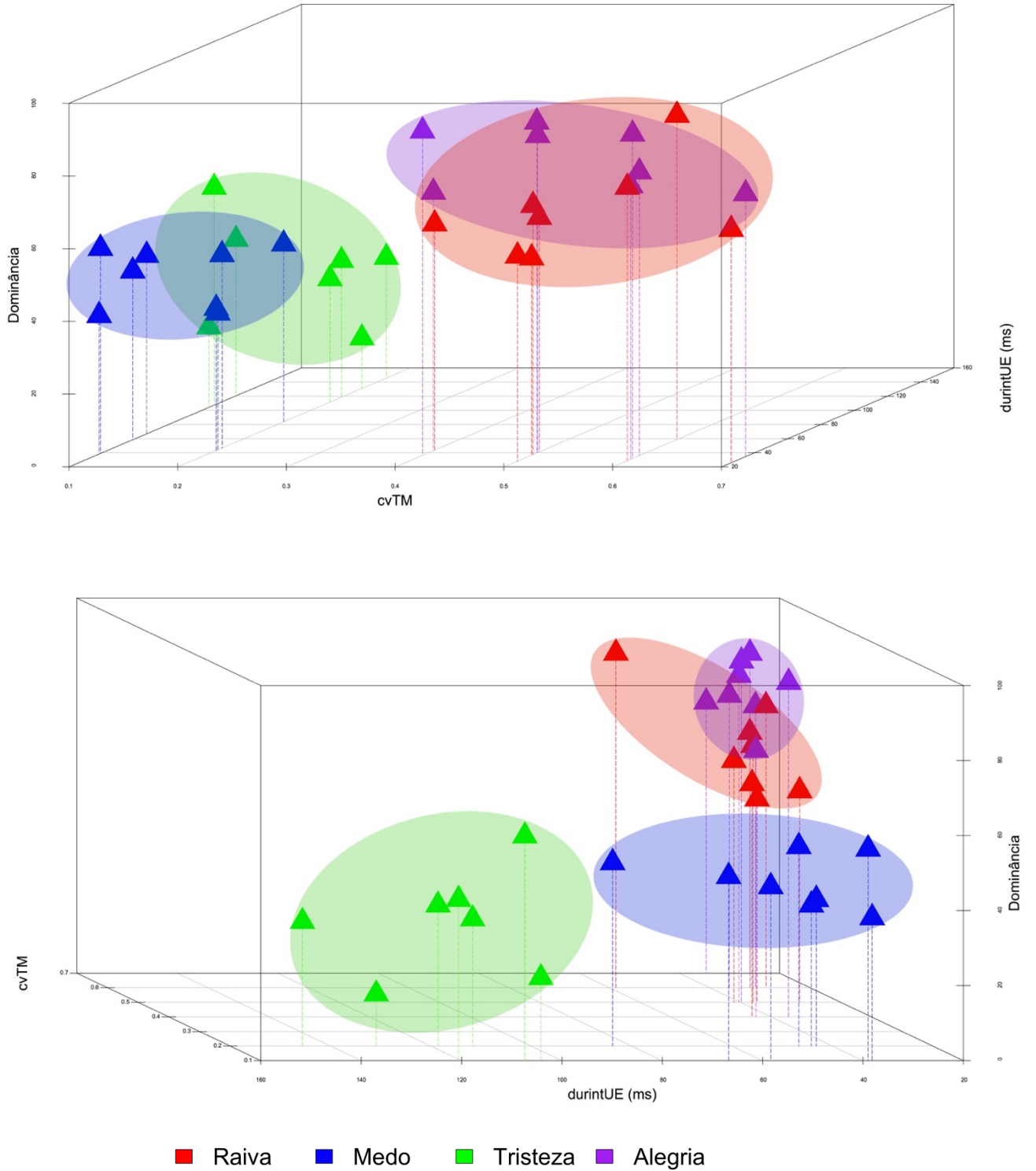


Gráfico 41: Dois ângulos diferentes da distribuição de julgamentos de ativação (Bra – fala delexicalizada) e a combinação de parâmetros acústicos (coeficiente de variação do tom médio – TM + duração dos intervalos entre as unidades entoacionais – durintUE).
 Valor de R^2 : 0,59

Gráficos de dispersão 3D - Dominância | cvTM e durintUE



7.3 Análise das dimensões e emoções básicas

Após a análise de inferência acústica e das dimensões, este subitem visa a analisar a distribuição da média das notas dadas pelos participantes para cada uma das dimensões. Os gráficos de interação abaixo ilustram as dimensões no experimento com fala normal para ambos os participantes, brasileiros e ingleses, lembrando que as emoções foram nomeadas pelos participantes do estudo piloto.

Gráfico 42: Interação entre as dimensões no experimento com fala normal – nativos

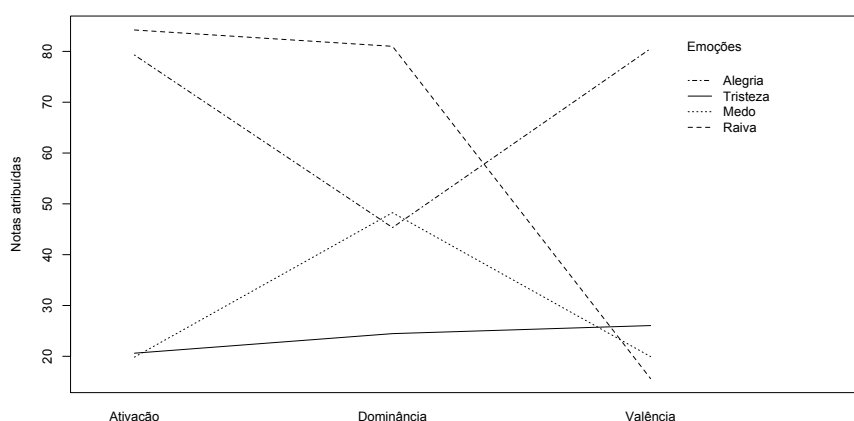
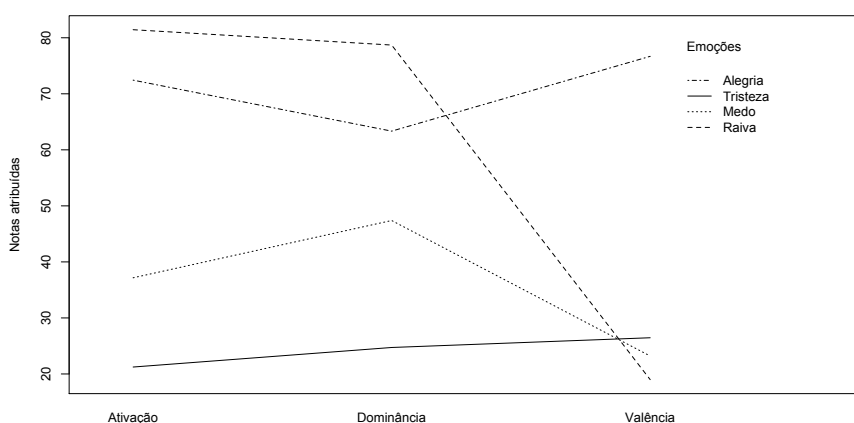


Gráfico 43: Interação entre as dimensões no experimento com fala normal – não nativo



O desempenho dos participantes brasileiros e ingleses no experimento com fala normal foi semelhante, mas algumas diferenças podem ser notadas; por exemplo, os brasileiros tenderam a dar notas

similares para dominância quando julgavam alegria e medo. Essa tendência foi também descrita nos gráficos de dispersão tridimensionais, quando a tristeza ficou isolada das demais emoções, restando num só grupo alegria, medo e raiva (e.g., gráficos 23 e 25, p. 139 – 141).

Os ingleses tenderam a dar notas semelhantes para dominância nas emoções alegria e raiva. Os gráficos de interação abaixo referem-se à distribuição da média das notas dadas pelos participantes em cada dimensão no experimento com fala delexicalizada.

Gráfico 44: Interação entre as dimensões no experimento com fala delexicalizada – nativos

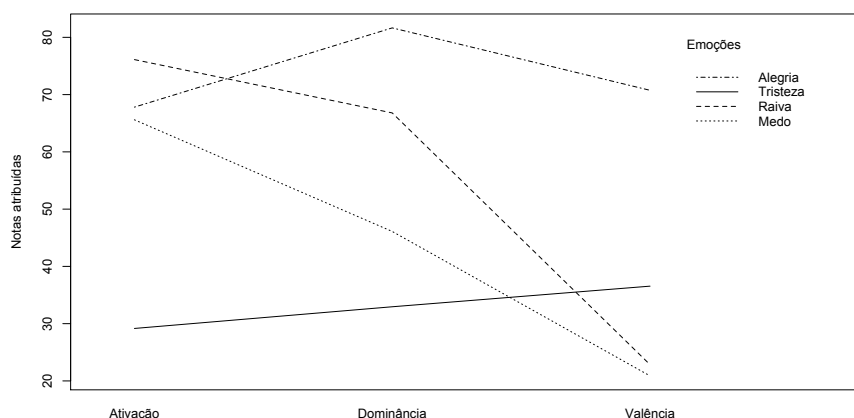
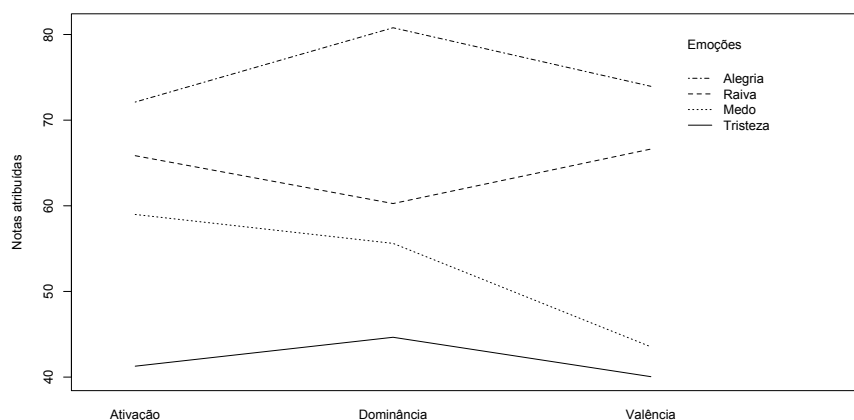


Gráfico 45: Interação entre as dimensões no experimento com fala delexicalizada – não nativos



Para os participantes brasileiros, não houve muita diferença para os julgamentos dados para tristeza ao longo das três dimensões, por conta da pouca inclinação do segmento de reta que representa seus resultados. As emoções raiva e medo tenderam ao mesmo julgamento dado à fala normal, ou seja, alta ativação e baixa valência. Para os ingleses, a raiva apresentou notas altas na dimensão valência, característica oposta às notas dadas no experimento com léxico. Outra diferença encontrada foi o maior valor de ativação para medo, tendência contrária à encontrada no experimento com fala normal. Para a visualização da disposição dos julgamentos dos participantes (brasileiros e ingleses), em ambos os experimentos (fala normal e delexicalizada), as emoções foram dispostas em um espaço tridimensional. Com esse tipo de ilustração, pode-se perceber como as emoções são agrupadas segundo os julgamentos das dimensões pelos participantes. Além disso, pode-se notar a influência do conhecimento da língua e do tipo de estímulo (normal e delexicalizado) na percepção das emoções.

Gráfico 46: Dispersão das emoções pela análise dimensional (Bra – fala normal)

Gráfico de dispersão 3D (Bra - fala normal)

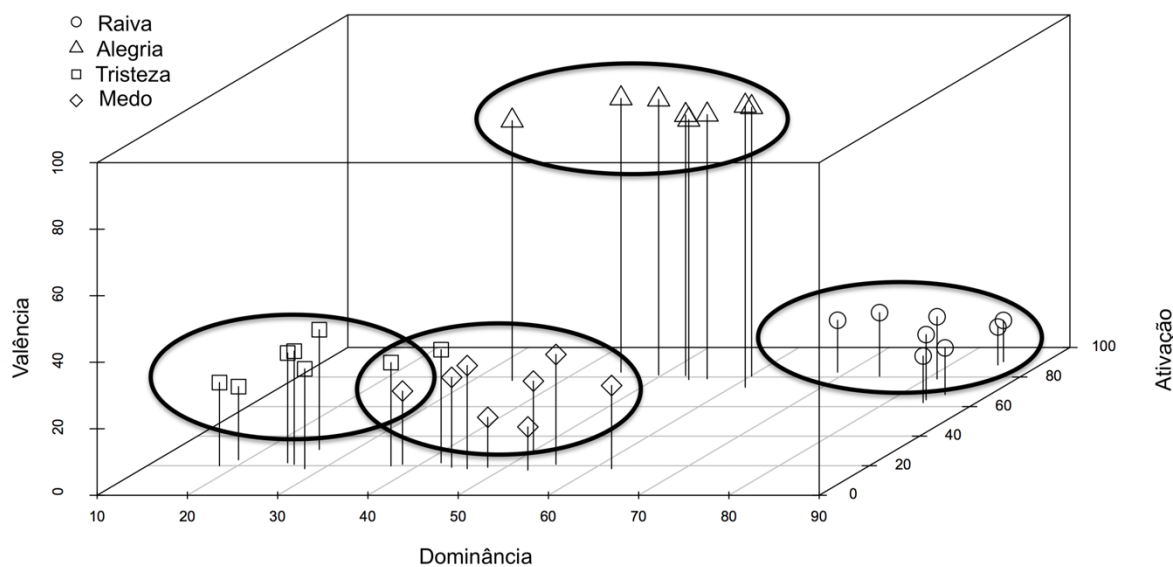


Gráfico 47: Dispersão das emoções pela análise dimensional (Ing – fala normal)

Gráfico de dispersão 3D (Ing - fala normal)

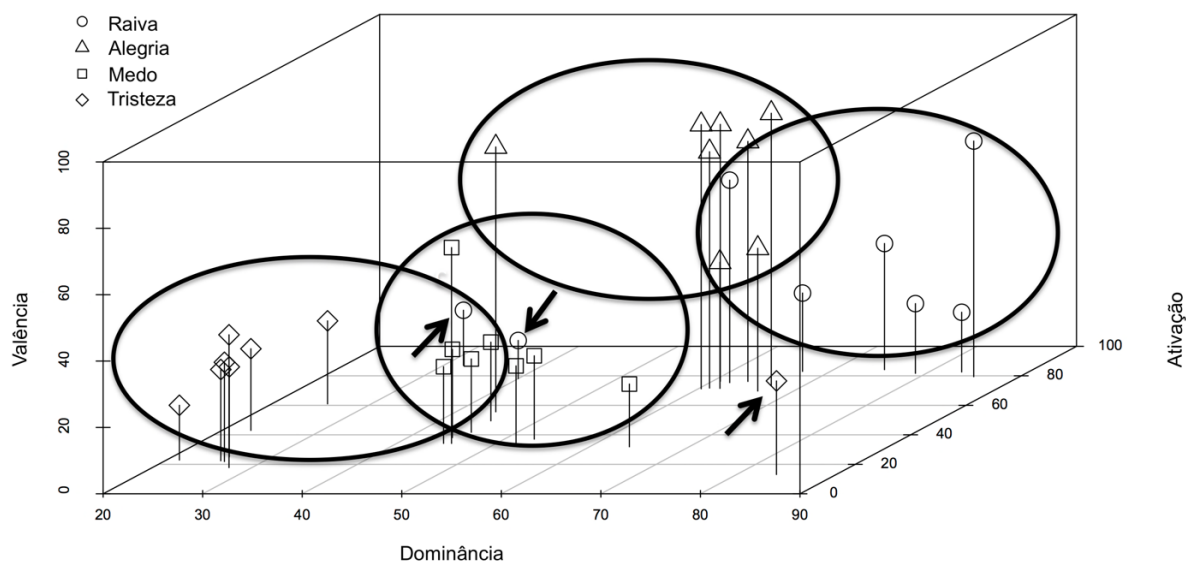


Gráfico 48: Dispersão das emoções pela análise dimensional (Bra – fala delexicalizada)

Gráfico de dispersão 3D (Bra - fala delexicalizada)

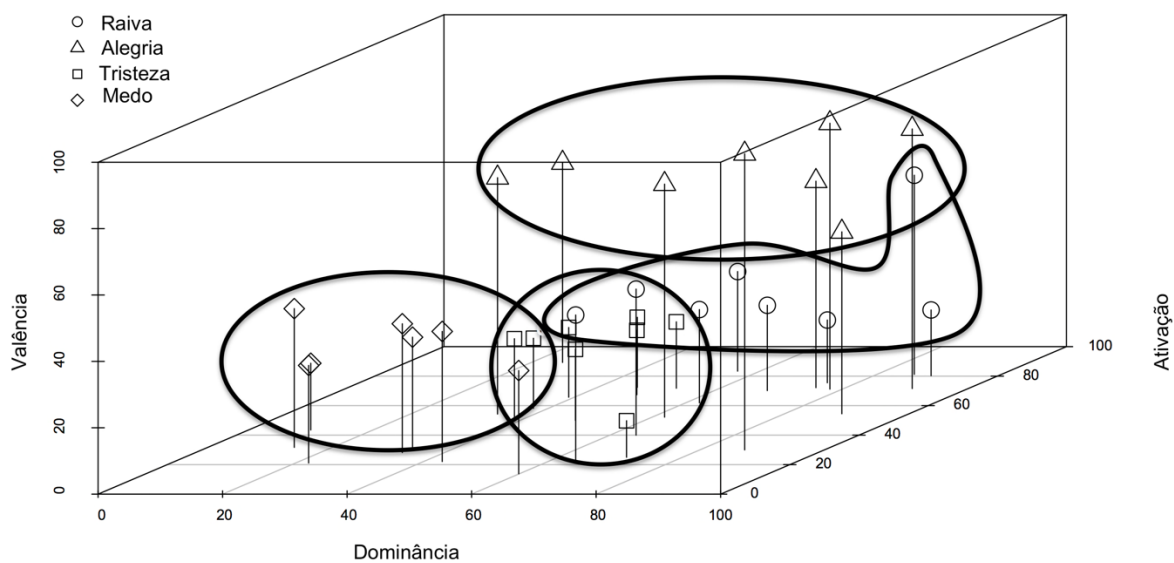
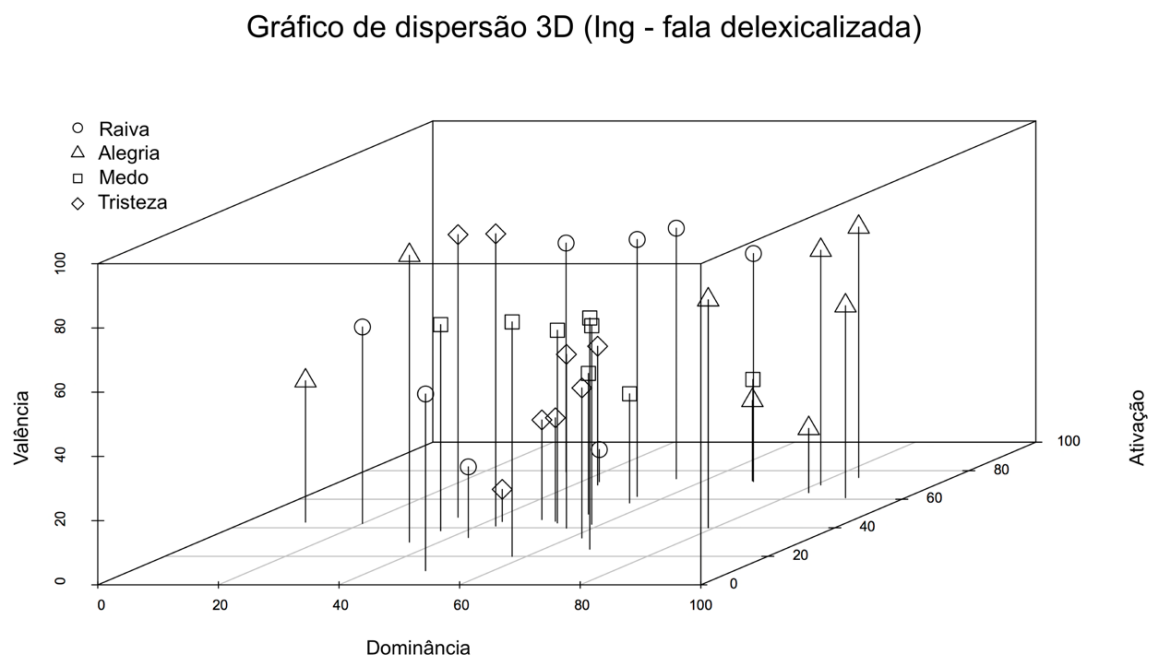


Gráfico 49: Dispersão das emoções pela análise dimensional (Ing – fala delexicalizada)



No experimento com fala normal, os julgamentos das dimensões dados pelos participantes brasileiros formaram *clusters* bem definidos, ou seja, houve distinção quase que inequívoca das emoções. Os ingleses também tiveram êxito razoável, pois as notas atribuídas por eles acabaram por formar *clusters* de emoções, mas não com a mesma precisão dos brasileiros. Pode-se notar que houve sobreposição de emoções, como, por exemplo, entre raiva e alegria.

Os gráficos do experimento com fala delexicalizada demonstram um resultado bem diferente do obtido com fala normal, pois ocorreram muitas sobreposições entre as emoções. Apesar da aparente aleatoriedade, é possível notar alguns agrupamentos nos julgamentos dos brasileiros. Por exemplo, tristeza e medo estão razoavelmente agrupados, sendo diferenciados principalmente pelas dimensões dominância e ativação. Os julgamentos de valência, no caso dessas duas emoções, foram bastante imprecisos. Quando comparados aos resultados com fala normal, pode-se

notar que a diferença entre tristeza e medo também ocorreu na dominância.

Para alegria, os julgamentos dos brasileiros envolvendo fala normal e delexicalizada não apresentaram diferenças importantes, pois os valores de dominância do experimento com fala normal foram medianos, tal qual os encontrados no experimento com fala delexicalizada. A única diferença é que o *cluster* das médias dos julgamentos foi mais concentrado nos dados de fala delexicalizada. Outra diferença ocorreu nos valores de ativação, elevados no experimento com fala normal, mas bastante imprecisos na fala delexicalizada. No caso dos participantes ingleses, a ausência de léxico causou grande dificuldade nos julgamentos, como se pôde observar na sobreposição quase total das emoções, excetuando-se a alegria e o medo, que apresentaram alguns pontos agrupados.

8. Discussão preliminar

Antes de iniciar a discussão preliminar, é necessário que se faça uma recapitulação resumida dos resultados por razão da extensão da análise dos experimentos. Esse breve resumo será útil para se ter uma ideia geral dos resultados, além de encontrar problemas decorrentes da metodologia e esboçar algumas conclusões.

1) Teste piloto com nativos e não nativos (emoções básicas) – qualidade de voz e entoação:

- ❖ Os informantes brasileiros tiveram 90% de concordância com relação ao julgamento dado pelo experimentador; nos ingleses a concordância foi de 66%. As variações de TM e os valores de

ênfase espectral e LTAS apresentaram resultados significativos ($p < 0,05$). No teste *post hoc* de Tukey (HSD), as diferenças significativas encontradas foram: TM – alegria e medo, tristeza e raiva; cvF0 – tristeza e raiva; ênfase espectral – tristeza e raiva; LTAS – tristeza e raiva. O resultado encontrado foi semelhante à proposta de Darwin (1965), que dividiu as emoções em dois grupos gerais – *exciting and depressing states of the mind*.

- ❖ Os resultados do teste piloto serviram para rotular com mais propriedade cada uma das emoções. Os rótulos *raiva, medo, tristeza e alegria* foram utilizados nos demais experimentos.

2) Teste perceptivo com nativos e não nativos – qualidade de voz e análise dimensional das emoções:

- ❖ No experimento com fala normal os parâmetros de qualidade de voz não apresentaram resultados significativos quando cotejados com as dimensões.
- ❖ Apesar disso, houve boa concordância entre os julgamentos das dimensões dados pelos brasileiros. Os ingleses apresentaram baixa concordância.

3) Teste perceptivo com nativos e não nativos – parâmetros acústicos (ExProsodia) e análise dimensional das emoções:

- ❖ Somente as dimensões ativação e dominância apresentaram resultados significativos quando cotejadas com os parâmetros acústicos.
- ❖ *Fala normal* – o resultado das regressões lineares simples entre os julgamentos das dimensões (nativos e não nativos) foi significativo envolvendo TM, cvTM, durUE, durintUE.

- ❖ *Fala delexicalizada* – o resultado das regressões lineares simples entre os julgamentos (nativos) das dimensões ativação e dominância foi significativo envolvendo cvTM, durUE, durintUE.
- ❖ A relação entre os parâmetros acústicos relacionados à frequência fundamental e as dimensões foi diretamente proporcional, ou seja, quando os valores de TM e cvTM crescem, maiores são os valores de ativação e dominância.
- ❖ Por outro lado, a relação entre os parâmetros acústicos relacionados à duração e as dimensões foi inversamente proporcional, ou seja, quando os valores de durUE, durintUE decrescem, maiores são os valores de ativação e dominância.
- ❖ A ausência do léxico influenciou significativamente os julgamentos de brasileiros e ingleses e os resultados das regressões lineares. Ambos os desempenhos decaíram, sobretudo o desempenho dos ingleses, o qual não apresentou resultado significativo.
- ❖ A combinação dos parâmetros acústicos nas regressões lineares múltiplas elevou de forma ostensiva os valores de R^2 .
- ❖ As combinações mais significativas foram TM + durUE, TM + durintUE, cvTM + durUE, cvTM + durintUE, cvTM + dpdurUE e cvTM + dpdurintUE. Os parâmetros foram os mesmos das regressões lineares simples, mas o poder de explicação dos modelos de regressão linear múltipla aumentou por conta da combinação de parâmetros acústicos.
- ❖ As dimensões apresentaram maneiras diferentes de agrupamento de acordo com o parâmetro acústico. Em fala normal percebida por nativos e não nativos, a ativação apresentou, de forma geral, duas regiões de formação de *cluster*, um nas regiões de alta ativação, outro nas de baixa. Os parâmetros envolvidos foram TM, cvTM, durUE e durintUE.

- ❖ A dominância apresentou mais variação e, portanto, diferentes maneiras de agrupamento das emoções. Para os nativos, houve três regiões de *clusters*. Na região de baixa dominância, a emoção correlata foi a tristeza; na região mediana, medo e alegria; na alta, raiva.
- ❖ A interação entre os julgamentos de dominância e os parâmetros de duração apresentou três configurações. A primeira dividiu as emoções em medo e tristeza de um lado, raiva e alegria de outro; a segunda deixou a tristeza isolada das demais emoções, com valores elevados de dur_{intUE} , na terceira não houve formação definida de *clusters*, quando o parâmetro dp_{durUE} foi analisado.
- ❖ Para os não nativos, a dominância pode ser dividida em quatro partes, cada uma delas remetendo a uma emoção em particular. Da região mais baixa de dominância à mais alta, a ordem das emoções foi tristeza, medo, alegria e raiva.
- ❖ O experimento com fala delexicalizada apresentou resultados inferiores aos obtidos no experimento com fala normal. Para os nativos, somente três parâmetros foram relevantes na comparação com o julgamento das dimensões: $TM + dur_{intUE}$, $cvTM + dur_{intUE}$ (com ativação); $TM + durUE$ (com dominância).
- ❖ Os não nativos não apresentaram resultado significativo no experimento com fala delexicalizada.

Algumas questões podem ser levantadas a partir dos resultados obtidos nos experimentos. Os dois modelos de análise (emoções básicas e abordagem dimensional) apresentaram pontos positivos e negativos. No modelo com emoções básicas, tem-se um ponto de partida necessário para uma primeira divisão das emoções, ou para guiar a maneira como o pesquisador deve selecionar o material de estudo. A adoção de um modelo baseado em emoções básicas parece ser apriorística num primeiro

momento, mas que se prova necessária não só pelo motivo expresso acima, mas também por ser uma maneira de avaliar a eficácia de modelos dimensionais. A simples combinação das dimensões não faz sentido sem um ponto de chegada, ou seja, uma emoção que possa ser formada pela intersecção de dimensões ortogonais. Afinal de contas, mesmo com o poder de detalhamento que o modelo dimensional dispõe, pessoas são capazes de reconhecer e nomear emoções de forma eficiente no cotidiano, como demonstrado no experimento piloto desta tese.

O experimento com qualidade de voz teve um resultado interessante, tendo em vista as diversas pesquisas que encontraram, em parâmetros acústicos de qualidade de voz, respostas para a manifestação das emoções na fala. O teste piloto com emoções básicas obteve resultado positivo, mas o modelo dimensional não obteve o mesmo êxito. Um ponto que necessita ser explorado mais a fundo é a grande influência que a ausência da qualidade de voz provocou na percepção dos brasileiros (além da ausência do léxico) e, principalmente, na dos ingleses.

Apesar de os julgamentos das dimensões e da qualidade de voz não terem apresentado resultados significativos, a perda das características vocais, por meio da delexicalização das sentenças, dificultou bastante a tarefa dos ingleses, que, antes mesmo da delexicalização, não tiveram acesso ao léxico por razões de desconhecimento da língua portuguesa. Assim, a fala delexicalizada dificultou a percepção dos ingleses menos pela ausência de léxico do que pela perda das características da qualidade de voz proporcionada pelo processo de delexicalização. Ademais, as dimensões, de maneira geral, tiveram julgamentos bastante diversos daqueles atribuídos no experimento com fala normal. No caso dos brasileiros, as dimensões parecem estar intimamente ligadas ao conteúdo linguístico das sentenças, resultado que sugere um papel importante da língua na percepção das emoções.

O desempenho dos ingleses com fala normal indica que há algumas características acústicas da fala que prescindem o dado linguístico para a percepção das emoções. Por exemplo, algumas regularidades como valores baixos dos parâmetros acústicos de F0 (TM, cvTM) foram relacionados a medo e tristeza, e valores elevados a raiva e alegria. Essas características podem ser pistas acústicas com as quais o ouvinte não nativo pode se orientar quando exposto a trechos de fala emotiva de língua estrangeira. Outras características recorrentes foram os valores dos parâmetros acústicos de duração, os quais foram baixos para raiva e alegria e altos para medo e tristeza. Esses resultados também estão presentes na *performance* dos nativos, mesmo quando o léxico não estava disponível, reforçando a hipótese de que as variações de altura e duração são fatores determinantes para a percepção da fala emotiva.

Tomemos então a divisão das emoções proposta por Darwin (1965), na qual as emoções foram divididas em *exciting and depressing states of the mind*. Os resultados, principalmente entre julgamentos de ativação e os parâmetros acústicos – ExProsodia –, apontaram para uma divisão semelhante. Nesses casos, a sensação de agitação percebida pelos participantes nativos e não nativos ocorreu invariavelmente quando houve maiores variações de F0 (cvTM), ou quando seus valores gerais (TM) foram elevados. Juntamente com essas variações, as unidades entoacionais apresentaram valores menores (durUE, durintUE, etc.). Em outras palavras, a agitação percebida dependeu em grande parte da combinação desses parâmetros. Essa configuração parece ser intuitiva na medida em que emoções como raiva e alegria apresentam graus elevados de ativação, característica que imprime à fala uma taxa de elocução elevada e maior valor médio de F0, além de maior variação (DAVITZ, 1964; MURRAY; ARNOTT, 1993). Pode-se concluir que trechos de fala mais variáveis do ponto de vista tonal e com um *tempo* mais acelerado são entendidos como portadores de alta ativação e, portanto, passíveis de serem julgados como

alegria ou raiva. No caso dos nativos, a confusão entre as duas emoções foi provavelmente dirimida por conta do conhecimento linguístico, no caso dos não nativos houve sobreposição entre as duas por conta da ausência da informação linguística. A partir desse raciocínio, pode-se entender por que a formação de *clusters* nos nativos foi mais definida.

A mesma argumentação é válida para as emoções de baixa ativação, medo e tristeza, mas com uma configuração inversa de parâmetros acústicos. Em resumo, trechos de fala menos variáveis do ponto de vista tonal (baixos valores de F0 médio e com pouca variação) e com um *tempo* mais desacelerado (altos valores de duração como *durUE*, *durintUE*, etc.) são entendidos como portadores de baixa ativação e, portanto, sujeitos a serem julgados como medo e tristeza.

Como primeiros passos desta pesquisa, os experimentos contidos neste estudo abrem caminho para uma série de questionamentos sobre a pertinência de cada modelo de análise, sobre a quantidade de dimensões necessárias para esse tipo de pesquisa e sobre a busca e aperfeiçoamento de um modelo híbrido para a análise da emoção na fala. Esse modelo deve ser capaz de lidar com a dupla face das emoções, uma marcadamente cultural e linguística, outra universal e presente em diferentes culturas.

Quanto ao número de dimensões, ativação e dominância se mostraram suficientes para explicar a maioria dos dados e agrupá-los de acordo com os resultados do teste piloto com emoções básicas. Esses resultados são interessantes se lembrarmos que o estudo de Trnka (2011) demonstrou que a valência é a única dimensão presente em todos os estudos analisados pelo autor. Até mesmo Ekman (1992) propôs análise das emoções baseando-se nessa dimensão. Entretanto, apesar de a valência não ter apresentado resultado positivo quando analisada juntamente com os parâmetros acústicos, ela foi determinante para separar a alegria (alta valência) das demais emoções nos gráficos de dispersão (gráficos 39 e 40). Em outras palavras, as intersecções dessas dimensões refletem a percepção

dos informantes sobre as emoções básicas, sendo eles nativos ou não, mesmo com a variação entre os desempenhos dos dois grupos. Como vimos, os primeiros conseguiram uma *performance* melhor provavelmente por conta do conhecimento linguístico. Se a relação entre as dimensões fosse formalizada, teríamos algo semelhante aos quadros abaixo:

Quadro 4: Esquematização dos julgamentos das dimensões – fala normal para nativos e não nativos (“+” = valores altos; “-” = valores baixos; “±” valores medianos; “♣” = aleatório)

Fala normal

	Raiva		Medo		Tristeza		Alegria	
	Nativo	Não nativo	Nativo	Não nativo	Nativo	Não nativo	Nativo	Não nativo
<i>Valência</i>	-	±	-	-	-	-	+	+
<i>Ativação</i>	+	+	-	±	-	±	+	+
<i>Dominância</i>	+	+	±	±	-	-	±	±

Quadro 5: Esquematização dos julgamentos das dimensões – fala delexicalizada para nativos e não nativos (“+” = valores altos; “-” = valores baixos; “±” valores medianos; “♣” = aleatório)

Fala delexicalizada

	Raiva		Medo		Tristeza		Alegria	
	Nativo	Não nativo	Nativo	Não nativo	Nativo	Não nativo	Nativo	Não nativo
<i>Valência</i>	±	♣	-	♣	-	♣	±	♣
<i>Ativação</i>	±	♣	-	♣	-	♣	±	♣
<i>Dominância</i>	±	♣	-	♣	±	♣	±	♣

De acordo com os quadros, a precisão do julgamento sobre as dimensões decai significativamente quando o léxico e a qualidade de voz são retirados do sinal acústico. No caso dos brasileiros, o desempenho decaiu, mas ainda houve uma certa sistematicidade nos julgamentos; mas, no caso dos ingleses, o problema foi ainda maior, pois parece que os não nativos ficaram totalmente sem referência no momento de julgar os estímulos delexicalizados. Para os falantes nativos, é bastante razoável

afirmar que a ausência do léxico ocasionou a queda no desempenho, mas o que dizer dos não nativos? Certamente, o mesmo argumento da ausência de léxico não pode ser considerado no caso dos ingleses. Então, se a língua não pode desempenhar nenhum tipo de papel no reconhecimento das emoções por não nativos, por razão de desconhecimento completo, é razoável pensar que haja outra informação na fala que os não nativos utilizaram para reconhecer emoções. Os resultados, de forma resumida, sugerem isso:

Quadro 6: Resumo dos resultados da análise dimensional

	<i>Fala normal</i>	<i>Fala delexicalizada</i>
<i>Bra</i>	Desempenho muito bom Valores de R ² por volta de 0,80 <i>Clusters</i> bem definidos	Bom desempenho Valores de R ² por volta de 0,60 <i>Clusters</i> com sobreposição
<i>Ing</i>	Bom desempenho Valores de R ² por volta de 0,70 <i>Clusters</i> com sobreposição	Baixo desempenho Valores de R ² por volta de 0,50 Sem formação de cluster

Com o quadro acima, fica mais fácil perceber que o desempenho entre nativos (fala delexicalizada) e não nativos (fala normal) foi muito semelhante. A característica comum a esses dois resultados é a ausência do entendimento do que estava sendo dito nas sentenças. Se o léxico, na fala normal, não desempenhou nenhum papel na *performance* dos ingleses, era de se esperar que o mesmo ocorresse com estímulos delexicalizados, entretanto o desempenho foi ainda mais baixo.

Neste caso, somos levados novamente a cogitar a qualidade de voz como veiculadora da informação sobre estados emotivos na fala, a despeito dos resultados negativos obtidos na análise dimensional. É importante lembrar, entretanto, que a qualidade de voz apresentou resultados significativos no teste piloto com emoções básicas.

De alguma forma, parece que a metodologia do experimento dimensional pode ter influenciado os resultados como, por exemplo, os

termos utilizados nos dois polos de cada dimensão, ou os parâmetros acústicos analisados. A dimensão ativação, em especial, foi problemática para os ingleses por conta dos rótulos *agitated* e *non agitated*, referência direta à versão brasileira do experimento, “agitado” e “não agitado”. Em inglês, *agitated* não tem exatamente o mesmo significado do termo em português, ou seja, em inglês o termo tem uma conotação mais negativa do que em português. Estar agitado em português não transparece a natureza da agitação ou sua motivação, podendo ser ocasionada por razões não exatamente negativas. Por outro lado, em inglês, o significado tem o sentido mais direcionado à agitação por nervosismo ou descontentamento. Em suma, há dois problemas que precisam ser diminuídos: a influência da língua na metodologia do experimento e a quantidade de informação disponível nos estímulos dados aos participantes no momento da pesquisa.

Para o primeiro problema, é necessário pensar numa metodologia que prescindia o uso da língua. Para o segundo, é preciso conseguir isolar estímulos que preservem a qualidade de voz, i.e., um método de delexicalização que mantenha a qualidade de voz. Esse método resolveria o problema em parte, pois ainda assim restariam as informações entoacionais, o que seria problemático no momento de isolar qual variável é responsável pelas respostas dos participantes. Para tentar responder por que os resultados entre fala normal e fala delexicalizada foram tão díspares para os não nativos, é necessário outro experimento que considere os dois problemas expostos acima. No capítulo a seguir, serão expostos as justificativas e os pressupostos teóricos do próximo experimento.

9. Experimento com estímulos de curta duração – *thin-slicing*

9.1 Justificativa e *background*

A justificativa para um novo experimento reside nos dois problemas decorrentes dos resultados do experimento anterior: a disparidade entre os desempenhos dos não nativos em ambos os experimentos (fala normal e delexicalizada) e a possível influência dos rótulos das dimensões. A solução parece óbvia, mas não simples: i) fornecer um tipo de estímulo que mantenha as informações acústicas da fala intactas, mas que retire a informação linguística; ii) fornecer uma ferramenta de julgamento para os informantes que prescindam o dado linguístico, rótulos, etc.

Para a primeira solução, há outros métodos de delexicalização além do filtro de passa-baixa e o PURR, de Sonntag e Portele (1998), este já

utilizado nesta tese. O problema desses dois métodos é que a informação referente à voz é descartada, mantendo-se somente informações de intensidade, variações de F0 e estrutura rítmica. Como os próprios autores afirmam, a técnica de delexicalização PURR reduz a fala ao seu conteúdo prosódico, sem interferência de outros componentes do sistema. O problema da disparidade entre fala normal e delexicalizada para os não nativos provavelmente está em algum componente isolado pelo método PURR.

O método proposto por Vainio *et al.* (2009) seria ideal para o que se pretende fazer neste experimento, pois mantém a qualidade de voz do sinal da fala. O método consiste numa filtragem glotal inversa na qual a fala passa por um procedimento de cancelamento dos efeitos causados pelo trato vocal no sinal produzido pela glote. Há, entretanto, dois problemas. O primeiro diz respeito à dificuldade de uso desse método, pois os arquivos têm de ser enviados para processamento pelos próprios autores. Infelizmente, os autores não estão mais utilizando esse método.

O método chamado *randomized splicing* também poderia ser uma solução (SCHERER, 1971). A técnica consiste em recortar trechos de uma sentença e reordená-los de modo que o conteúdo veiculado por ela fique inacessível ao ouvinte. Na época da proposta, o método era aplicado de maneira manual, isto é, os trechos eram recortados literalmente de fitas de gravação e colados de maneira aleatória. Alguns cuidados eram tomados como isolar as pausas e trechos silenciosos. Segundo o autor, a grande vantagem desse método é a manutenção de todo o espectro da voz (SCHERER, 1971). Entretanto, um de seus pontos negativos é a falta de naturalidade do estímulo, além de provocar profundas alterações no contorno entoacional, tanto do ponto de vista da altura (F0) quanto do ritmo. Alguém poderia argumentar que o foco desse método é manter a qualidade vocal não importando suas consequências na entoação. De fato, é um argumento válido, mas encerra no experimentador a responsabilidade

de orientar os participantes para que julguem somente a qualidade de voz e ignorem as características entoacionais. Entretanto, esse tipo de instrução é problemático para ouvintes leigos e requereria treinamento prévio.

Outro método utilizado é o chamado *reverse speech*, que consiste em tocar trechos de fala de trás para frente (MUNRO; DERWING; BURGESS, 2003, 2010). O método cumpre o que promete do ponto de vista da ininteligibilidade do léxico, mas apresenta problemas como a completa descaracterização da curva de F0, além de afetar as consoantes e as transições de formantes.

Outra maneira de deixar a informação de qualidade de voz disponível aos ouvintes seria utilizar vogais isoladas como estímulo, ou as chamadas *steady-state vowels* (HILLENBRAND; GAYVERT, 1993). Os falantes têm de ler uma lista de palavras que contemple o inventário da língua em questão, em seguida o experimentador deve isolar as vogais para o experimento de percepção. Esse tipo de método estaria fora do escopo desta tese por não se tratar de uma instância espontânea de fala. Ademais, vogais isoladas, mesmo que retiradas de fala espontânea, são estímulos um tanto artificiais para serem julgados do ponto de vista emocional.

Apesar dos métodos acima serem profícuos, sua aplicação não seria adequada aos propósitos deste experimento. Dessa forma, uma saída seria restringir o acesso ao léxico ao menor nível possível sem que haja manipulações invasivas ou que resultem em estímulos pouco naturais. Para tal, trechos curtos extraídos dos experimentos anteriores (menos de 1s) serão apresentados aos ouvintes de maneira que o pouco acesso ao léxico não influencie o julgamento dos nativos e, certamente, dos não nativos.

Entretanto, os trechos devem obedecer a algumas limitações. A primeira delas é que a(s) palavra(s) contida(s) nos trechos não pode(m) ter qualquer ligação semântica com as emoções. Para isso, palavras gramaticais como preposições, artigos, conjunções, substantivos que sejam

neutros do ponto de vista emocional, ou até mesmo nomes próprios, são candidatos a serem utilizados como estímulo.

Antes de prosseguir, há uma questão central quanto a esse método: é possível que uma informação tão restrita possa veicular o conteúdo necessário aos ouvintes? Em outras palavras, trechos tão curtos como, por exemplo, 400ms podem ser percebidos de maneira adequada? Essa pergunta pode ser respondida com base em evidências acústico-perceptuais, cognitivas, psicológicas e neurológicas.

Argumentos acústicos-perceptuais:

- ❖ Habilidade em perceber mudanças rápidas e mínimas no som ao longo do tempo.
- ❖ Experimentos sobre detecção de pausas/lacunas no som – *gap detection experiment* (LESHOWITZ, 1971). O limiar de detecção para ruído branco foi de 3ms (PENNER, 1977).
- ❖ A redundância da fala:
 - A fala pode resistir a consideráveis perdas de informação e ainda continuar inteligível.
 - A fala tem mais informação do que o necessário (WARREN et al., 1995).
 - A fala pode ser completa e corretamente entendida sem a estrutura harmônica ou pistas de vozeamento (REMEZ et al., 1981).
 - *Spectral smearing*⁷⁵ – simula os efeitos da redução de detalhes de frequência (perda auditiva coclear) (BAER; MOORE, 1993).

⁷⁵ Consiste na perda de detalhes espectrais. Os picos dos formantes são “borrados” para que sejam menos proeminentes. Mesmo com filtros seis vezes maiores que o normal, ouvintes ainda foram capazes de perceber palavras-chave nas sentenças (PLACK, 2013).

- Limiares de diferença para estímulos sonoros isolados com duração entre 40ms a 250ms (JOOS, 1948).

Argumentos cognitivos, psicológicos e neurológicos:

- ❖ O processamento das características prosódicas da fala emotiva é um processo altamente rápido e complexo.
- ❖ Sentenças produzidas com emoções básicas podem ser diferenciadas de sentenças neutras antes de 200ms após o início da fala (P200) (PAULMANN et al., 2009; PAULMANN; BLEICHNER; KOTZ, 2013).⁷⁶
- ❖ Percepção de sílabas que subitamente tornam-se emocionais quando comparadas com sílabas neutras (SCHIRMER et al., 2008).
- ❖ *Thin-slicing* – refere-se à habilidade inconsciente para encontrar padrões em situações e comportamentos baseados em informação restrita (*narrow slices of experience*) (AMBADY; BERNIERI; RICHESON, 2000; AMBADY; ROSENTHAL, 1992, 1993; AMBADY; KRABBENHOFT; HOGAN, 2006; CARRERE, GOTTMAN, 1999).

Após os argumentos supramencionados, fica claro que os ouvintes são capazes de extrair informação e detectar padrões baseados em estímulos limitados em diferentes aspectos, como, por exemplo, a duração.

Dentre os argumentos acima, o método *thin-slicing* é adequado ao que se pretende analisar neste experimento, pois se trata da limitação da informação contida nos estímulos. A limitação é feita no âmbito temporal, ou seja, na quantidade de informação (do ponto de vista da duração)

⁷⁶ Também conhecido como P2, é um componente de onda relacionado a ERP (*Event-related potential*), que tem picos por volta de 200ms após o aparecimento de um estímulo externo.

disponível para os participantes. Dessa forma, é importante detalhar e definir o método *thin-slicing*, com seus pressupostos, justificativas e resultados prévios.

9. 1. 1 *Thin-Slicing*

As pesquisas sobre a precisão na percepção têm demonstrado que os sujeitos são capazes de apreenderem dados da realidade de forma bastante consistente mesmo sob circunstâncias limitadoras. Nesse sentido, estudos no âmbito da psicologia social comprovam que conseguimos julgar de maneira satisfatória situações baseadas nas primeiras impressões. O trabalho de Ambady e Rosenthal (1993) foi o primeiro a utilizar o termo *thin slice*, que é definido como um excerto breve de um comportamento expressivo retirado de um *continuum* comportamental. Ambady, Bernieri e Richeson (2000) o definem como qualquer excerto com informação dinâmica com menos de cinco minutos de duração. Segundo os autores, as *thin slices* retêm muita informação, senão a maior parte dela, contida na comunicação, reduzindo ou às vezes eliminando: i) a informação codificada dentro do fluxo verbal; ii) o contexto histórico; iii) o contexto global no qual o comportamento está ocorrendo (AMBADY; BERNIERI; RICHESON, 2000, p. 204). Em outras palavras, é uma maneira de isolar os estímulos das circunstâncias gerais nas quais eles foram produzidos. Dessa forma, essa delimitação obriga o participante a se concentrar em pistas não verbais, no nosso caso, sem influência do contexto de fala e da informação verbal.

Essa metodologia tem sido aplicada em estudos com os mais variados interesses. Ambady e Rosenthal (1993) analisaram a acurácia no julgamento de personalidade feito por participantes de um experimento de percepção. A primeira tarefa foi avaliar professores universitários e de ensino médio por meio de trechos de vídeos silenciosos com duração de 30s (três trechos de 10s). Numa escala de 15 dimensões (profissional, otimista,

ativo, confiante, honesto, etc.), os participantes foram capazes de prever, de maneira significativa, a avaliação feita pelos alunos ao final do semestre. Mesmo quando os estímulos foram ainda mais curtos (de 2s a 5s), os participantes ainda foram capazes de avaliar de maneira coerente. O resultado demonstrou que não houve diferença significativa no julgamento dos participantes enquanto julgavam diferentes durações de estímulos: 2s, 5s e 10s.

Resultados semelhantes foram encontrados por Ambady, Krabbenhoft e Hogan (2006). Participantes de um experimento de percepção de fala foram capazes de julgar a eficácia das vendas por meio de trechos de 20s de fala normal e delexicalizada retirados de entrevistas com gerentes de vendas. Curhan e Pentland (2007) também demonstraram que é possível prever o resultado de entrevista de emprego baseando-se apenas nos primeiros cinco minutos. Outro estudo de Ambady *et al.* (2002) analisou a relação entre o tom de voz dos médicos⁷⁷ (dois trechos de 10s de fala delexicalizada) e o histórico de reclamações de mau atendimento. Os participantes julgaram os estímulos baseando-se numa escala de sete pontos para cada uma das 10 dimensões (competente, ansioso, hostil, etc.). Os resultados apontaram para uma associação entre o tom de voz e o histórico de reclamações de mau atendimento dos médicos.

Quanto à duração dos estímulos, qual seria a medida? Por intuição, tendemos a pensar que quanto maior o estímulo, mais apurado será o julgamento dado pelo participante. Na verdade, a meta-análise feita por Ambady e Rosenthal (1992) mostrou que o aumento da duração dos estímulos não resultou numa melhora no desempenho dos participantes. Então, qual seria a duração ideal do estímulo para este experimento? Como dito anteriormente, a adoção da metodologia *thin-slicing* é uma maneira de

⁷⁷ Os autores referem-se ao “tom de voz”, mas não há nenhum tipo de medição objetiva do ponto de vista acústico, somente a avaliação dos participantes baseada numa escala dimensional. Seria mais adequado referir-se à “impressão sobre o tom de voz”.

limitar o acesso dos participantes à informação lexical relevante para o julgamento da fala emotiva, além de ser uma maneira de descontextualizar os trechos de fala sem causar alterações no sinal acústico. Para o experimento dessa tese, trechos de 200ms a 1400ms foram extraídos de trechos maiores com duração de até cinco minutos. O critério de recorte obedeceu a dois critérios: 1) limitação do acesso ao contexto e ao léxico; 2) restrição dos trechos à duração mais breve possível. Os trechos de fala selecionados serão listados na metodologia deste experimento.

9. 1. 2 Como avaliar os estímulos?

Nos experimentos anteriores, os participantes julgaram os estímulos por meio de rótulos com emoções básicas (raiva, medo, tristeza e alegria), no caso do experimento piloto, ou por meio de um botão deslizante quando julgaram as emoções no experimento com abordagem dimensional (valência, ativação e dominância). Uma das objeções feitas diz respeito à influência que os rótulos podem exercer na maneira como os estímulos são julgados. No caso das emoções básicas, os participantes foram forçados a escolher dentre as alternativas disponíveis (4AFC – *four alternative forced-choice*), mesmo quando possivelmente eles tiveram dúvida sobre qual rótulo era mais adequado à emoção percebida. No caso da abordagem tridimensional, o problema foram os rótulos dos polos opostos de cada dimensão.

Como visto no subcapítulo 2.5.2, os rótulos de cada polo podem variar bastante de um estudo para o outro. No caso do experimento desenvolvido nesta tese, houve outro problema: o da versão dos termos em duas línguas diferentes. Durante comunicação oral sobre o desenvolvimento desta pesquisa,⁷⁸ o principal problema apontado por falantes nativos do inglês foi a versão inglesa dos rótulos utilizados para os polos da dimensão ativação.

⁷⁸ *The Perception of Emotional Speech by Native and Non-native speakers*. The University of York, Reino Unido, 05 de março de 2015.

No caso do português, os termos utilizados para ativação foram *agitado vs. não agitado*, na versão inglesa os termos foram *agitated vs. non-agitated*. Apesar de os resultados dos julgamentos não variarem tanto entre os nativos e os não nativos, o termo *agitated* não tem o mesmo sentido encontrado em português. O verbo *to agitate*, segundo o *Cambridge Dictionary* (2013), é definido como *to make someone feel anxious or angry*. Nota-se que o sentido do termo em inglês remete à raiva e à ansiedade, o que pôde ser constatado nas críticas feitas pelos nativos. Já em português, o termo *agitado* não apresenta somente esse sentido, mas incorpora outras acepções não necessariamente negativas, ou seja, pessoas podem estar agitadas em razão de algo positivo. Talvez os termos *arousal* ou *energy* pudessem ser mais adequados.

Em razão dessa disparidade, este experimento busca uma maneira de julgar os estímulos sem recorrer ao uso da língua. A primeira solução aventada foi a utilização de cenários, como proposto em Scherer (1991). Neste experimento, a aplicação dos cenários seria diferente do que usualmente é feito, pois em estudos prévios essa metodologia foi usada para elicitare as emoções.

Entretanto, neste experimento, os cenários seriam usados de maneira oposta, ou seja, os cenários seriam lidos pelos participantes e depois relacionados aos estímulos. Embora seja uma alternativa interessante porque deixa de fora rótulos estanques sobre as emoções, pode apresentar uma variação muito grande e difícil de ser controlada por conta do histórico de cada participante e da experiência de cada sujeito com cada um dos cenários. A nota de rodapé 37 (pág. 71) apresenta um exemplo de cenário que possivelmente tenha apelos diferentes para cada pessoa.

Por conta da dificuldade exposta acima, a utilização de imagens que representam cada uma das emoções é uma saída possível. Como primeira alternativa, optou-se pelo banco de fotos de expressões faciais das emoções do *Karolinska Institutet*, as KDEF – *The Karolinska Directed Emotional*

Faces. As KDEFE são um conjunto de 4900 imagens de expressões faciais das emoções desenvolvido por Lundqvist, Flykt e Öhman (1998) e originalmente concebido para estudos de percepção, atenção, emoção, memória, etc. O conjunto contém 70 indivíduos (35 atores e 35 atrizes), cada um exibindo sete expressões emocionais diferentes (neutra, alegria, raiva, medo, nojo, tristeza e surpresa), fotografadas duas vezes em cinco ângulos diferentes. Como há muita variação de *performance* entre os atores, uma solução mais adequada e menos variável é as AKDFE – *The Averaged Karolinska Directed Emotional Faces* (LUNDQVIST; LITTON, 1998).

Esse material foi feito a partir do material completo das KDEFE. As imagens das KDEFE foram convertidas de RGB de 32 bits para 8 bits em tons de cinza. Após a conversão, as imagens foram “empilhadas” por tipo de emoção, por exemplo, duas séries de 35 indivíduos masculinos felizes foram colocadas na mesma “pilha”. A partir dessas pilhas, a média das imagens foi calculada, mantendo preservadas as características comuns presentes em cada um dos atores para cada emoção. É uma solução engenhosa que diminui a influência individual de cada ator no momento da representação das emoções.



Figura 10: Exemplos de expressões faciais médias das emoções (AKDFE) (LUNDQVIST; LITTON, 1998); no canto superior esquerdo temos a expressão média de 35 atrizes para a emoção raiva; no canto superior direito temos a expressão média de 35 atores para a emoção medo; abaixo delas, da esquerda para a direita, temos tristeza e alegria

É importante ressaltar que não se trata de um experimento de reconhecimento facial das emoções, mas de um experimento de percepção de fala que utiliza expressões faciais como modo de categorização do que está sendo percebido. Em outras palavras, os participantes tiveram de relacionar o estímulo acústico (*thin slices*) à figura que melhor o representou.

Em resumo, os dois problemas relacionados ao experimento anterior, ou seja, a disparidade dos resultados dos não nativos em ambos os experimentos (fala normal e delexicalizada) e a possível influência dos

rótulos necessitavam de duas soluções: i) estímulos que mantivessem as informações acústicas da fala intactas e inviabilizassem o entendimento linguístico; ii) ferramenta de julgamento para os informantes que prescindisse de rótulos linguísticos. A solução encontrada para “i” foi o método *thin-slicing*; para “ii”, o conjunto de expressões faciais médias das emoções – AKDEF.

9.2 Metodologia

Este experimento terá como estímulos 48 *thin slices* de fala emotiva espontânea retiradas das gravações com maior duração baixadas do *site* www.youtube.com. Do total de 48 *thin slices*, 32 foram retiradas das gravações utilizadas no experimento com abordagem dimensional, o restante foi retirado de novas gravações coletadas da internet. Os 48 trechos foram divididos em quatro emoções básicas, as mesmas utilizadas nos experimentos anteriores: raiva, medo, tristeza e alegria.

Os 16 trechos juntados aos demais 32 foram previamente avaliados em experimento piloto para verificar a concordância entre o experimentador⁷⁹ e ouvintes leigos. O resultado foi similar ao encontrado no experimento piloto descrito no capítulo 5 desta tese. O grau de concordância para os sujeitos brasileiros foi de razoável a bom – $K = 0,77$, $p \sim 0$, para os não nativos a concordância foi razoável – $K = 0,43$, $p \sim 0$ (FLEISS, 1981). A duração dos trechos variou de 200ms a 1400ms. O quadro abaixo apresenta os trechos utilizados no experimento com seus respectivos conteúdos lexicais e duração.

⁷⁹ O experimentador foi quem selecionou os trechos da internet. Por essa razão, foi necessário novamente um teste piloto para aferir a validade das categorizações das emoções. A concordância entres os juízes é uma maneira de fazê-lo.

Quadro 7: Lista de *thin slices* utilizadas no experimento com duração e transcrição. Na coluna da esquerda estão os nomes dos arquivos. O número é a identificação do arquivo, a primeira letra (M ou F) é referente ao sexo do falante, a última letra diz respeito à inicial de cada emoção (raiva, medo, tristeza e alegria)

<i>Thin slices</i>	<i>Duração (ms)</i>	<i>Transcrição</i>
1M_R	921	<i>cinco reais</i>
2M_R	1231	<i>quarenta e duas</i>
3M_R	923	<i>capacidade</i>
4M_R	947	<i>pronunciamento</i>
5M_R	491	<i>da mesa</i>
6M_R	739	<i>desejar</i>
7M_R	531	<i>costas</i>
8M_R	637	<i>valoriza</i>
9M_R	380	<i>toda</i>
10F_R	685	<i>computador</i>
11F_R	769	<i>telefone</i>
12F_R	559	<i>pode levar</i>
1F_M	876	<i>meu cunhado</i>
2F_M	1369	<i>Araçatuba</i>
3F_M	860	<i>vai fazer o</i>
4F_M	386	<i>não</i>
5F_M	549	<i>alô</i>
6F_M	367	<i>sair</i>
7F_M	804	<i>completamente</i>
8M_M	657	<i>nossa frente</i>
9M_M	1161	<i>Paulo aí ó</i>
10M_M	380	<i>céu</i>
11M_M	638	<i>Arnaldo</i>
12M_M	220	<i>cara</i>
1F_T	968	<i>três horas</i>
2F_T	579	<i>família</i>
3F_T	899	<i>dedicação</i>
4F_T	773	<i>profissão</i>
5F_T	495	<i>vida</i>
6F_T	712	<i>por ele</i>
7F_T	1251	<i>e visto a camisa</i>
8M_T	809	<i>pessoas</i>
9M_T	960	<i>conquistas</i>
10M_T	516	<i>estamos</i>
11M_T	628	<i>índole</i>
12M_T	505	<i>brasileiro</i>
1M_A	822	<i>tava no quarto</i>
2M_A	595	<i>é verdade</i>
3M_A	650	<i>programa também</i>
4M_A	837	<i>eu comia tudo</i>
5M_A	547	<i>vou te falar uma</i>
6M_A	628	<i>pode ter</i>
7M_A	439	<i>Manuela</i>
8M_A	442	<i>toda</i>
9M_A	633	<i>consegui</i>
10F_A	985	<i>sabonete</i>
11F_A	866	<i>tomava banho</i>
12F_A	664	<i>cozinhando</i>
Média	714	
Desvio Padrão	246	

Os estímulos foram julgados por 76 participantes, 38 nativos e 38 não nativos, somando um total de 3648 respostas. Os participantes nativos estão divididos em 21 mulheres e 17 homens; os não nativos estão divididos em 23 mulheres e 15 homens. Não foi possível dividir os participantes em grupos iguais de acordo com o sexo por ser um experimento *online* e sujeito à participação voluntária, o que torna difícil controlar esse tipo de variável. De qualquer forma, a variável sexo será tratada nos resultados deste experimento.

A tarefa neste experimento foi ouvir os estímulos e relacioná-los às figuras (AKDEF) presentes na tela. Os ouvintes puderam ouvir os trechos quantas vezes julgaram necessárias, entretanto esse número foi tomado como variável na análise dos dados. Antes de iniciar, os participantes julgaram dois exemplos para se familiarizarem com o experimento. Cada tela do experimento apresentou um *player* com a *thin slice* a ser relacionada com uma das figuras AKDEF. É importante dizer que os participantes não puderam avançar sem antes clicar numa das figuras. Além das figuras, a alternativa “em dúvida/*not sure*” estava disponível em caso de os participantes não terem certeza sobre qual figura melhor representava o estímulo que estavam ouvindo. Os estímulos acústicos foram apresentados de forma aleatória, bem como a disposição das figuras na tela, com exceção da alternativa “em dúvida/*not sure*”, que permaneceu localizada sempre abaixo das demais alternativas com expressões faciais. A figura abaixo ilustra um exemplo de estímulo na tela do computador.

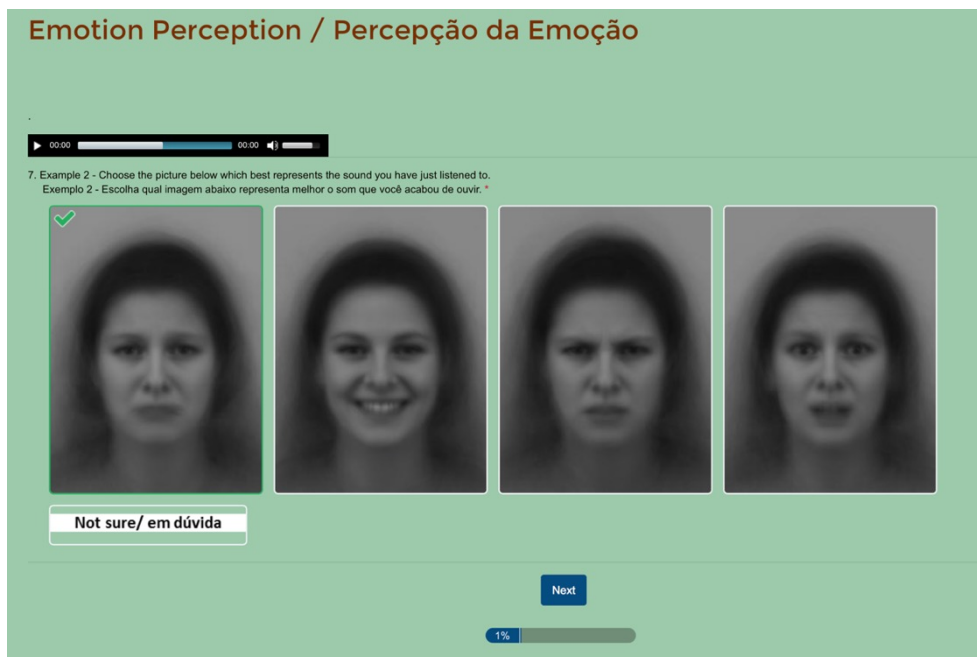


Figura 11: Exemplo de estímulo (experimento – *thin-slicing*)

Diferentemente dos experimentos anteriores, os participantes não são somente brasileiros e ingleses; além deles, há participantes de Portugal, Estados Unidos e Canadá.

9.3 Análise dos dados

9.3.1 Resultados

O primeiro passo da análise é descrever os dados para que se tenha uma ideia mais ampla do que ocorreu nos experimentos. Após essa etapa, os dados serão analisados do ponto de vista inferencial. Para se ter uma visão geral sobre como os dados foram distribuídos, cada uma das emoções está disposta em um quadro separado. Em cada um deles será possível ver quantos foram os acertos e quantos e quais foram os erros. Os quatro quadros a seguir ilustram o desempenho dos participantes em cada uma das emoções.

Quadro 8: Julgamentos dos participantes sobre estímulos (*thin slices*) de raiva. O painel à esquerda refere-se aos participantes nativos; à direita, aos não nativos. A numeração (coluna em branco) é a identificação de cada participante (76 no total, 38 nativos e 38 não nativos); a identificação de cada estímulo foi escrita verticalmente acima de cada coluna; a soma dos julgamentos corretos e a respectiva porcentagem estão abaixo do quadro

	Nativos													Não nativos											
	1F_F	2F_F	3F_F	4F_F	5F_F	6F_F	7F_F	8M_F	9M_F	10M_F	11M_F	12M_F		1M_A	2M_A	3M_A	4M_A	5M_A	6M_A	7M_A	8M_A	9M_A	10F_A	11F_A	12F_A
1	R	R	R	R	R	R	M	D	R	R	R	R	1	R	R	R	A	R	M	M	M	R	M	R	
2	M	M	M	M	M	M	R	M	R	M	M	M	2	R	R	R	R	R	M	R	R	R	R	T	
3	R	R	R	R	R	R	T	R	R	R	R	R	3	T	R	R	M	R	R	R	R	R	R	R	
4	R	R	R	R	R	R	R	R	R	R	R	R	4	D	R	R	R	M	D	M	A	M	D	A	M
5	R	R	R	R	R	R	D	R	A	D	R	R	5	R	R	R	R	R	R	R	R	R	R	R	R
6	M	R	D	R	M	M	M	R	M	R	D	M	6	R	R	R	R	R	R	D	R	R	R	R	R
7	M	R	R	R	R	R	R	R	R	M	R	R	7	R	D	R	R	R	R	R	D	R	M	R	R
8	R	R	R	R	R	A	M	A	R	R	M	R	8	R	R	R	R	R	R	M	R	R	R	R	R
9	M	M	R	R	M	M	R	M	M	M	M	M	9	R	R	R	R	R	A	A	A	R	R	A	A
10	R	R	D	R	M	R	R	A	R	R	A	M	10	R	R	R	R	M	A	R	R	R	M	R	M
11	D	A	R	R	A	A	M	R	R	D	R	R	11	R	R	R	R	A	R	R	R	R	R	A	R
12	R	R	M	R	R	R	R	D	R	R	M	R	12	R	R	R	R	R	R	R	R	R	R	R	R
13	R	A	R	R	R	R	M	A	R	R	R	R	13	R	A	R	R	R	R	R	R	R	R	R	R
14	M	T	R	R	M	M	R	M	A	A	A	T	14	R	R	R	R	R	R	R	R	R	R	R	R
15	M	A	R	D	R	A	M	R	R	D	R	R	15	R	R	R	R	M	R	R	R	M	R	R	R
16	R	R	R	R	R	R	R	R	R	R	R	R	16	M	R	R	R	R	A	R	A	A	R	D	R
17	R	R	R	R	R	T	R	D	R	R	R	R	17	A	R	R	R	R	R	R	R	R	A	R	R
18	R	R	M	R	M	R	R	M	R	R	R	R	18	M	R	R	R	A	M	R	R	R	A	R	R
19	R	R	R	R	R	R	R	R	R	R	R	R	19	R	R	R	R	M	R	R	R	R	R	R	R
20	R	R	R	R	R	R	R	R	R	R	D	R	20	A	R	R	R	R	R	R	A	R	R	R	R
21	M	D	R	R	R	R	R	R	R	M	R	R	21	R	R	M	A	M	R	A	M	M	M	R	R
22	R	R	R	D	R	R	R	R	R	R	D	R	22	M	R	R	R	R	R	A	R	M	R	R	M
23	T	R	R	R	M	M	R	M	R	R	T	T	23	R	R	R	R	R	R	M	R	R	R	R	M
24	R	M	R	R	R	R	R	M	R	R	R	R	24	A	R	R	M	M	R	M	A	T	A	A	D
25	R	R	R	R	D	R	R	M	R	M	R	R	25	R	R	R	R	R	T	R	M	R	R	A	R
26	R	R	R	R	M	A	A	A	R	R	M	R	26	R	R	R	R	R	M	A	R	D	D	M	
27	R	R	R	R	A	D	D	A	R	R	R	R	27	D	R	R	R	R	R	R	R	R	R	R	R
28	R	R	R	D	R	R	M	M	R	M	M	R	28	R	R	R	M	R	M	R	M	R	R	A	M
29	A	R	R	R	R	R	R	R	R	R	R	R	29	R	R	R	R	R	A	R	R	R	R	R	R
30	T	R	R	R	R	R	M	R	R	R	T	R	30	R	R	R	R	M	R	R	R	R	R	R	R
31	R	R	R	R	R	R	M	R	D	R	D	R	31	R	R	R	R	R	R	R	M	R	R	R	R
32	T	A	R	R	R	R	R	R	R	R	R	R	32	T	R	R	R	R	R	M	R	R	R	R	R
33	R	R	R	R	R	R	M	R	R	R	R	R	33	R	R	R	M	R	R	R	M	R	R	R	R
34	A	D	R	R	D	R	R	R	R	R	R	R	34	R	R	R	R	R	R	T	R	R	R	M	
35	R	R	R	R	A	T	A	T	R	R	R	R	35	R	R	R	R	R	R	R	R	R	M	R	R
36	R	R	R	R	M	R	A	M	R	R	M	R	36	R	R	R	M	R	R	M	R	R	R	A	R
37	R	R	R	R	R	R	R	R	R	R	R	R	37	R	R	R	R	M	R	R	R	R	R	R	R
38	R	M	R	R	R	R	T	R	M	M	R	R	38	R	R	R	R	R	R	R	R	R	R	R	R
Total	25	27	33	34	24	26	22	19	33	27	22	32	Total	28	36	37	31	28	29	25	24	31	29	28	28
%	66	71	87	89	63	68	58	50	87	71	58	84	%	74	95	97	82	74	76	66	63	82	76	74	74

Raiva Alegria Medo Tristeza Dúvida

Quadro 9: Julgamentos dos participantes sobre estímulos (*thin slices*) de tristeza

	Nativos												NÃO Nativos	NÃO Nativos											
	1F_T	2F_T	3F_T	4F_T	5F_T	6F_T	7F_T	8M_T	9M_T	10M_T	11M_T	12M_T		1F_T	2F_T	3F_T	4F_T	5F_T	6F_T	7F_T	8M_T	9M_T	10M_T	11M_T	12M_T
1	T	T	T	T	T	T	T	T	T	M	T	T	1	T	T	A	A	A	T	M	R	T	T	A	A
2	R	T	M	M	T	T	M	T	A	T	R	M	2	T	T	T	T	T	T	M	R	T	T	T	T
3	T	T	T	T	T	T	M	A	T	M	T	3	T	T	M	T	T	T	T	M	T	T	D	M	
4	T	T	T	T	T	T	R	T	T	R	T	4	T	T	A	D	T	T	T	A	T	T	T	A	
5	T	T	T	A	M	T	T	T	R	D	A	T	5	T	T	D	T	A	T	T	T	T	D	M	T
6	R	T	T	T	T	T	R	R	A	T	R	T	6	T	T	T	T	T	T	T	T	T	A	T	T
7	T	T	T	T	T	T	M	T	M	T	T	T	7	T	A	T	D	A	T	D	T	A	T	T	T
8	M	T	T	T	T	T	T	T	M	T	A	8	T	T	T	T	T	T	T	T	T	T	A	M	T
9	M	T	M	T	M	M	T	R	R	T	R	A	9	T	T	T	T	R	T	T	R	T	T	T	T
10	T	R	T	T	R	R	T	T	A	T	R	T	10	T	T	T	T	T	A	T	M	T	T	D	T
11	T	T	T	T	T	T	T	A	T	T	R	T	11	T	T	T	T	T	T	T	T	T	T	T	T
12	T	T	M	A	D	T	T	T	T	D	T	T	12	T	T	T	T	T	T	T	T	T	M	T	A
13	T	T	T	D	T	T	T	T	T	T	A	D	13	T	T	T	T	T	T	T	T	T	T	T	T
14	M	A	T	A	A	M	M	A	T	T	A	A	14	T	T	T	T	T	T	T	T	T	T	T	T
15	T	T	T	T	T	T	D	T	T	T	D	T	15	T	T	T	T	M	T	T	T	T	T	A	T
16	T	T	T	T	T	T	T	T	T	T	T	T	16	D	T	A	M	T	T	T	M	T	T	M	T
17	T	T	T	R	T	T	T	D	M	T	M	T	17	T	T	A	A	M	T	T	T	T	M	M	T
18	T	T	T	M	T	T	A	R	R	M	T	M	18	T	T	T	T	T	T	A	T	T	M	T	T
19	T	R	T	T	T	T	T	M	T	A	D	19	T	T	A	T	T	T	T	T	T	T	D	T	T
20	T	T	D	T	T	T	T	T	R	D	T	M	20	T	T	A	T	T	A	T	T	T	T	T	A
21	T	T	T	T	T	T	D	A	A	D	A	T	21	T	T	A	T	T	T	T	T	T	M	A	T
22	D	T	T	T	T	T	T	D	T	M	T	T	22	M	A	A	T	T	T	R	T	T	R	A	T
23	M	M	A	M	T	T	M	A	R	M	T	A	23	T	T	T	T	T	T	R	T	T	T	T	T
24	M	T	T	T	M	M	M	T	A	M	T	T	24	T	T	T	T	T	T	T	T	T	T	T	T
25	T	T	T	T	T	T	T	M	T	M	T	25	T	T	T	T	T	T	T	T	T	T	A	M	A
26	T	T	A	T	T	T	T	T	A	T	R	T	26	T	T	T	T	T	A	A	T	T	A	A	A
27	T	D	T	M	T	R	T	R	A	D	D	A	27	M	T	T	T	T	T	D	R	T	R	D	T
28	T	T	T	T	T	T	T	A	T	T	T	A	28	T	T	T	T	T	T	T	T	T	M	M	T
29	T	D	T	T	R	T	D	D	T	T	T	A	29	T	T	D	T	T	T	D	T	T	T	D	T
30	T	T	T	T	T	T	T	T	T	T	T	T	30	T	T	R	T	T	T	T	T	T	T	T	T
31	T	T	T	T	T	T	D	D	A	T	R	A	31	T	T	T	T	T	T	T	T	T	T	A	T
32	T	T	T	T	T	T	A	T	A	T	T	T	32	T	T	T	T	T	T	T	R	T	T	T	A
33	R	T	T	T	T	T	A	T	T	M	A	T	33	T	T	A	T	T	M	T	T	T	M	T	T
34	T	T	T	T	T	T	T	R	T	T	T	A	34	T	T	T	T	T	T	T	T	T	M	T	A
35	T	T	M	D	T	T	T	T	T	T	R	A	35	T	T	T	T	T	T	T	T	T	T	M	T
36	T	T	A	T	T	T	A	D	D	M	T	A	36	T	T	A	A	T	T	T	T	T	T	A	T
37	T	T	T	T	T	A	M	T	A	T	A	T	37	T	T	T	T	T	T	A	A	T	M	M	T
38	T	A	T	T	A	T	T	M	A	T	R	D	38	T	T	T	T	T	T	A	T	T	T	T	T
Total	29	31	30	28	30	32	23	20	16	25	16	21	Total	35	36	24	32	32	34	28	28	37	22	19	29
%	76	82	79	74	79	84	61	53	42	66	42	55	%	92	95	63	84	84	89	74	74	97	58	50	76

Raiva Alegria Medo Tristeza Dúvida

Quadro 10: Julgamentos dos participantes sobre estímulos (*thin slices*) de alegria

	Nativos												Não nativos												
	1M_A	2M_A	3M_A	4M_A	5M_A	6M_A	7M_A	8M_A	9M_A	10F_A	11F_A	12F_A	1M_A	2M_A	3M_A	4M_A	5M_A	6M_A	7M_A	8M_A	9M_A	10F_A	11F_A	12F_A	
1	A	A	A	A	A	D	A	A	A	A	A	A	1	A	A	A	A	A	A	A	A	A	T	A	
2	M	A	M	A	A	M	M	T	T	A	A	R	2	T	A	T	A	R	A	A	A	A	T	A	
3	A	A	A	A	T	A	T	A	T	A	A	A	3	A	A	T	T	T	A	A	T	A	A	T	
4	A	A	A	A	A	A	A	A	A	A	T	A	4	R	A	A	A	T	A	D	A	A	A	A	
5	A	A	D	A	A	A	A	A	M	A	A	A	5	A	A	D	A	R	A	A	A	T	A	A	
6	R	R	T	A	D	T	R	T	A	A	A	A	6	R	A	D	A	R	T	D	A	T	A	M	
7	A	A	A	A	M	A	M	A	A	A	A	M	7	M	A	D	M	R	A	D	T	T	A	M	
8	A	A	A	T	A	A	D	A	A	A	T	A	8	R	A	A	T	M	A	A	A	A	T	A	
9	T	R	M	A	R	R	M	M	T	A	A	A	9	D	A	T	T	A	A	D	T	A	T	T	
10	A	A	R	A	M	M	A	R	A	A	R	A	10	A	A	A	A	A	A	A	A	A	A	A	
11	A	A	A	A	A	M	A	A	M	A	A	A	11	M	A	R	A	A	M	A	A	T	A	D	
12	T	A	A	A	A	M	A	A	A	A	M	A	12	T	A	R	A	D	A	A	A	T	A	A	
13	A	A	A	A	T	A	A	A	A	A	A	A	13	A	A	T	A	A	A	A	A	A	A	A	
14	A	M	M	M	A	M	A	M	A	A	A	A	14	D	A	T	A	T	T	T	A	T	A	R	
15	A	A	T	A	T	A	T	A	A	A	T	A	15	A	A	T	T	A	A	A	A	R	A	T	
16	A	A	A	A	A	M	A	A	A	A	M	A	16	M	A	M	A	M	A	M	A	T	A	A	
17	A	D	A	A	A	A	D	A	A	A	A	A	17	A	A	M	A	A	A	A	A	A	T	A	
18	A	A	A	A	A	A	A	A	A	A	A	A	18	R	A	A	A	T	A	A	A	R	A	M	
19	A	T	A	A	A	A	T	A	T	A	A	A	19	D	A	T	D	A	A	A	A	A	T	A	
20	A	A	A	A	A	A	D	A	M	A	A	A	20	A	A	A	A	T	A	D	T	R	A	A	
21	M	D	A	A	A	D	A	A	A	A	A	A	21	A	A	A	A	A	A	A	T	A	A	A	
22	A	A	A	A	A	A	A	A	A	A	A	A	22	A	A	A	A	T	A	A	A	T	A	M	
23	A	A	A	A	A	A	A	A	A	A	A	A	23	R	A	D	M	R	M	A	T	A	A	A	
24	R	R	T	A	A	A	D	A	T	A	A	A	24	A	T	A	A	A	A	A	A	A	T	M	
25	A	A	A	A	A	M	A	A	A	A	A	A	25	A	A	A	A	M	A	A	A	A	T	A	
26	A	R	A	A	A	A	A	A	A	A	A	A	26	T	A	R	A	A	A	A	T	T	A	A	
27	A	A	T	A	A	T	D	A	A	A	A	A	27	D	A	A	T	D	D	A	T	D	A	D	
28	A	A	A	A	A	A	T	A	T	A	T	A	28	R	A	A	T	T	M	A	A	A	A	A	
29	A	A	A	A	A	A	D	A	R	A	T	A	29	R	A	A	A	T	A	D	A	T	A	A	
30	A	A	A	A	A	D	A	M	A	A	A	A	30	A	A	A	A	T	A	T	T	T	A	A	
31	D	A	A	A	A	M	R	A	M	A	A	A	31	A	A	M	A	M	A	T	A	T	A	M	
32	A	A	A	A	A	A	A	A	A	A	M	A	32	A	A	A	A	T	A	A	A	R	A	A	
33	M	R	R	A	A	R	A	A	A	A	A	A	33	A	A	A	A	R	A	R	A	A	M	A	
34	A	A	A	A	D	A	A	A	A	A	A	A	34	R	D	D	A	R	A	R	A	A	A	T	
35	M	A	A	A	A	A	A	T	A	A	A	A	35	A	M	T	T	A	R	R	A	R	A	A	
36	A	A	A	A	A	A	A	A	A	A	A	A	36	R	A	A	T	A	T	A	A	A	T	A	
37	T	A	M	A	A	R	A	A	A	A	A	A	37	A	A	R	A	R	M	A	T	A	T	R	
38	R	A	D	A	A	M	R	A	A	A	A	A	38	A	A	T	T	R	A	A	A	T	A	T	
Total	27	29	26	36	30	21	22	31	27	38	29	36	Total	19	35	17	26	13	29	25	29	18	38	19	30
%	71	76	68	95	79	55	58	82	71	100	76	95	%	50	92	45	68	34	76	66	76	47	100	50	79

Raiva Alegria Medo Tristeza Dúvida

Quadro 11: Julgamentos dos participantes sobre estímulos (*thin slices*) de medo

	Nativos													Não nativos												
	1F_M	2F_M	3F_M	4F_M	5F_M	6F_M	7F_M	8M_M	9M_M	10M_M	11M_M	12M_M		1F_M	2F_M	3F_M	4F_M	5F_M	6F_M	7F_M	8M_M	9M_M	10M_M	11M_M	12M_M	
1	R	M	D	M	R	M	T	M	D	D	D	M	1	M	R	R	T	R	M	T	M	R	A	R	T	
2	M	M	M	M	M	A	M	M	M	M	M	M	2	M	M	R	T	R	M	T	M	M	A	T	D	
3	T	M	R	R	A	R	T	M	R	M	R	R	3	M	A	A	M	D	M	T	M	R	T	D	D	
4	T	M	R	T	R	R	A	M	M	T	R	M	4	M	D	D	M	M	D	M	M	M	D	M	A	
5	D	D	D	D	D	A	M	D	D	A	D	A	5	D	D	D	M	D	D	M	M	R	T	D	D	
6	R	M	M	T	M	M	R	M	R	M	M	M	6	T	R	D	D	R	M	D	M	D	D	D	A	
7	T	A	T	R	M	M	M	T	R	M	R	M	7	D	M	M	T	M	M	M	M	R	T	R	D	
8	R	M	M	T	R	A	M	M	R	M	D	T	8	T	T	R	T	R	A	A	M	R	T	R	M	
9	M	M	M	M	M	A	M	M	M	M	M	M	9	D	D	M	D	M	D	A	M	A	D	A	D	
10	M	M	M	T	M	M	M	R	T	T	M	M	10	T	M	D	R	R	D	A	M	M	M	M	T	
11	R	M	M	T	M	A	A	T	R	R	R	M	11	D	M	M	M	R	M	M	M	D	D	M	D	
12	M	M	M	D	M	D	A	M	R	D	M	D	12	D	M	R	M	R	M	T	M	R	A	M	D	
13	M	M	M	M	R	R	T	M	R	M	M	M	13	T	R	R	M	R	R	M	M	M	M	R	M	
14	M	T	T	R	A	A	A	T	R	M	M	R	14	T	R	R	T	R	R	T	M	R	T	R	D	
15	D	M	M	R	M	A	T	M	R	M	M	M	15	T	M	M	T	M	M	T	M	M	M	M	M	
16	M	R	M	M	R	R	A	M	R	T	R	T	16	M	D	M	D	M	M	T	M	D	A	D	M	
17	R	T	R	D	R	R	D	M	R	R	R	M	17	M	M	M	M	R	M	A	M	R	M	R	M	
18	R	M	M	R	D	M	R	T	A	A	A	R	18	T	R	R	R	M	M	M	M	M	R	A	T	
19	M	M	M	M	M	A	R	T	R	T	M	M	19	D	R	R	M	M	A	D	M	R	A	M	R	
20	D	T	D	R	D	A	M	R	R	T	D	M	20	M	M	M	D	M	M	T	A	M	R	R	D	
21	D	M	M	R	A	R	M	M	R	A	A	D	21	M	M	R	M	M	R	T	M	M	M	A	D	
22	M	R	M	T	M	D	T	R	R	R	D	M	22	M	M	M	M	R	M	A	T	R	R	R	A	
23	M	T	T	R	M	T	A	A	T	M	R	M	23	T	M	M	M	R	M	D	D	R	T	M	M	
24	M	M	D	R	R	M	A	M	R	T	T	D	24	T	M	M	T	M	A	M	T	M	A	M	T	
25	M	D	D	R	R	A	M	T	M	T	D	M	25	T	T	M	T	R	T	A	T	M	A	M	M	
26	T	M	M	T	A	A	R	T	M	T	A	D	26	D	R	M	D	D	A	A	M	M	D	M	D	
27	R	T	M	R	M	R	A	M	M	T	T	T	27	T	M	M	T	R	T	D	M	R	A	M	D	
28	T	M	R	M	A	A	R	M	R	R	A	M	28	T	M	M	M	M	M	M	M	M	R	R	R	A
29	D	M	R	M	D	A	A	M	M	D	R	D	29	D	M	D	M	D	D	M	A	M	D	M	M	
30	M	R	D	T	D	M	A	T	M	R	M	M	30	M	M	R	T	R	M	D	M	M	T	D	D	
31	D	D	M	R	D	R	A	M	D	T	D	M	31	T	M	R	M	R	M	M	T	M	M	R	M	
32	M	T	R	M	R	M	A	M	R	T	M	M	32	D	T	D	T	R	A	R	M	R	D	M	M	
33	R	R	R	R	R	A	R	M	R	M	R	M	33	T	M	M	T	R	M	T	M	M	A	M	T	
34	M	M	M	M	R	A	A	M	M	M	R	M	34	D	R	M	M	R	D	M	M	A	A	R	D	
35	M	A	M	R	A	R	A	T	M	M	M	M	35	T	A	A	M	M	A	A	M	M	T	R	R	
36	T	T	M	M	A	A	A	M	R	M	M	M	36	T	M	M	M	M	A	R	R	T	T	M	M	
37	T	R	M	R	R	M	A	T	R	T	R	A	37	R	M	R	M	R	M	R	M	R	R	M	M	
38	R	M	R	M	M	R	T	A	R	R	M	M	38	R	M	R	M	R	M	T	M	R	T	M	M	
Total	16	21	21	12	13	9	9	22	10	14	14	25	Total	10	21	17	19	13	21	11	30	16	6	16	13	
%	42	55	55	32	34	24	24	58	26	37	37	66	%	26	55	45	50	34	55	29	79	42	16	42	34	

Raiva Alegria Medo Tristeza Dúvida

O total de estímulos julgados para cada emoção foi de 912 *thin slices* (76 participantes x 12 estímulos), subdivididos igualmente entre nativos e não nativos (456 para cada grupo de 38 participantes). Para a emoção raiva, os nativos obtiveram um índice de respostas corretas de 71%, somando um total de 324 acertos; os não nativos obtiveram 78% de acertos, somando 354 respostas corretas. A média total de acerto para raiva, considerando ambos os grupos de participantes, foi de 74%, totalizando 678 respostas corretas. Numa primeira observação, pode-se perceber que nativos e não nativos não variaram tanto com relação aos julgamentos sobre a raiva (Quadro 8, p. 195), ou seja, as respostas foram, de maneira geral, semelhantes entre os grupos. Apesar de dados espúrios, como no caso dos participantes nativos 2, 6, 9 e 14 (com uma variação de 66% a 84% de respostas incorretas), ou dos participantes não nativos 4 e 24 (com 75% de respostas incorretas cada), o índice de acerto foi elevado entre nativos e não nativos.⁸⁰

O Quadro 9 (p. 196) ilustra os resultados para a emoção tristeza. De maneira semelhante ao Quadro 8 (raiva), a emoção tristeza teve resultados com índice de acerto de 66% (301 respostas corretas) para nativos e de 78% (356 respostas corretas) para os não nativos. Novamente, houve diferença de desempenho entre nativos e não nativos, mas dessa vez um pouco maior. A diferença de desempenho entre eles na tristeza foi de 12% em favor dos não nativos, já na raiva os não nativos tiveram uma diferença de 7% com relação aos nativos. Alguns participantes nativos tiveram *performance* abaixo dos demais, como o 9, 14, 23 e 27, os quais tiveram um índice de erro entre 66% e 75%. A média total de acerto para tristeza foi de 72%, totalizando 657 respostas corretas.

Por outro lado, no Quadro 10 (p. 197), referente à emoção alegria, os participantes nativos tiveram melhor *performance* com relação aos não

⁸⁰ Decidiu-se mencionar os participantes que tiveram oito erros (2/3 do total) ou mais num universo de 12 estímulos.

nativos. Os nativos tiveram um índice de 77% de acerto (352 respostas corretas), ao passo que os não nativos obtiveram 65% de acerto (298 respostas corretas). A diferença entre os índices de acerto dos participantes foi a mesma encontrada na emoção tristeza, só que inversa, ou seja, no caso da alegria os nativos tiveram 12% de acertos a mais que os não nativos. Do lado dos nativos, apenas o participante 9 teve desempenho abaixo dos demais, com 66% de erro. A média total de acerto para alegria foi de 71%, totalizando 650 respostas corretas.

Ao contrário do bom desempenho dos participantes encontrado nas emoções raiva, tristeza e alegria, a emoção medo (Quadro 11, p. 198) apresentou um resultado um tanto diverso se comparado com os demais. Os nativos tiveram um índice de acerto de 41%, somando 186 acertos, já os não nativos tiveram 42%, com 193 acertos. A média total de acerto para o medo foi de 41,5%, totalizando 379 respostas corretas. Sem dúvida, é um resultado diverso daqueles encontrados nas demais emoções. O medo apresentou muitos casos em que os participantes cometeram mais de 80% de erros, como, por exemplo, o participante nativo 5, que acertou somente um estímulo dentre 12 possíveis. A primeira hipótese aventada foi a de que os estímulos não eram apropriados, ou seja, não representavam de forma confiável um exemplar de fala portador de características associadas ao medo. Entretanto, no experimento piloto, os mesmos estímulos foram bem avaliados, seguindo a *performance* encontrada nas demais emoções. Outro aspecto que chama a atenção é a quantidade de “R” (raiva) atribuídos ao medo, totalizando 110 ocorrências para os nativos e 88 para os não nativos. Outra característica saliente é o número elevado de dúvidas, 61 casos para os nativos e 70 para os não nativos. Para poder visualizar melhor, duas matrizes de confusão foram feitas, uma para cada grupo de participante.

Tabela 14: Matriz de confusão das emoções apresentadas como *thin slices* aos participantes nativos. As linhas representam os estímulos apresentados aos participantes, somando 456 para cada emoção. As colunas representam as respostas dadas a cada uma das emoções. A última coluna à direita representa o total de estímulos ouvidos pelos participantes. A última linha é formada pela soma das colunas referentes às respostas dadas pelos participantes para cada emoção. As respostas corretas foram dispostas diagonalmente nas células preenchidas em cinza. Os erros estão dispostos acima ou abaixo da linha diagonal

		NATIVOS					
<i>Estímulos</i>	<i>Respostas</i>					Total	
	Raiva	Medo	Tristeza	Alegria	Dúvida		
Raiva	324	69	13	26	24	456	
Medo	110	186	46	53	61	456	
Tristeza	32	46	301	50	27	456	
Alegria	21	35	32	352	16	456	
Total	487	336	392	481	128	1824	

Tabela 15: Matriz de confusão das emoções apresentadas como *thin slices* aos participantes não nativos

		NÃO NATIVOS					
<i>Estímulos</i>	<i>Respostas</i>					Total	
	Raiva	Medo	Tristeza	Alegria	Dúvida		
Raiva	354	52	6	33	11	456	
Medo	88	193	65	40	70	456	
Tristeza	11	28	356	47	14	456	
Alegria	33	25	77	298	23	456	
Total	486	298	504	418	118	1824	

Uma característica importante dessas matrizes é a coluna com a resposta “dúvida”, para a qual não há equivalente nos estímulos. Por essa razão, ela foi posta à margem das demais respostas, somando-se aos resultados para que os números totais, tanto dos estímulos como das respostas, coincidisse, a saber, 1824. Isso posto, algumas características podem ser comentadas com mais clareza. Se compararmos as linhas das duas matrizes referentes à emoção medo, pode-se notar que os valores das células em branco são, em geral, elevados se comparados com as demais linhas.

Outro ponto é a soma das respostas “medo” dada a todas as emoções. Para os nativos, houve 336 respostas divididas entre todas as emoções, sendo 186 corretamente relacionadas ao medo. Já para os não nativos,

houve 298 respostas, sendo 193 corretamente relacionadas ao medo. Esse dado é importante para mostrar uma possível tendência de se evitar a resposta “medo” durante a feitura do experimento. Em outras palavras, mesmo como erro, o medo foi a emoção menos escolhida como sendo a emoção percebida pelos participantes. As demais respostas dadas aos estímulos seguiram, *grosso modo*, uma tendência mais homogênea. Para ambos os grupos de participantes, a raiva obteve praticamente o mesmo número de respostas a serem divididas entre as quatro emoções, 487 e 486 para nativos e não nativos, respectivamente. Entretanto, diferentemente do medo, a maior parte das respostas – 324 (71%) e 354 (78%) – foi correta, como visto no Quadro 8. O mesmo padrão foi encontrado para as emoções tristeza e alegria. Abaixo, encontra-se a estatística descritiva dos dados:

Tabela 16: Estatística descritiva – nativos

	Nativos			
	Raiva	Medo	Tristeza	Alegria
Mediana	26,50	14,00	26,50	29,00
Média	27,00	15,50	25,08	29,33
SE da média	1,44	1,58	1,67	1,54
Desvio padrão	4,99	5,49	5,78	5,33
Coef. variação	0,19	0,36	0,23	0,18
Assimetria	0,07	0,35	-0,37	0,12
Curtose	-1,51	-1,49	-1,53	-1,20

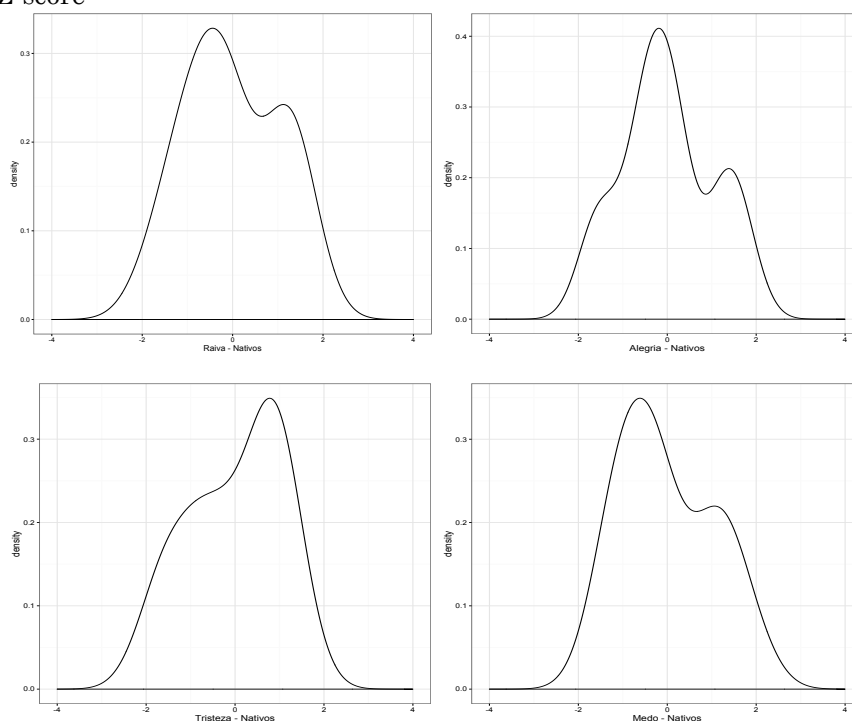
Tabela 17: Estatística descritiva – não nativos

	Não Nativos			
	Raiva	Medo	Tristeza	Alegria
Mediana	28,50	16,00	30,50	25,50
Média	29,50	16,08	29,67	24,83
SE da média	1,11	1,82	1,66	2,23
Desvio padrão	3,85	6,29	5,74	7,72
Coef. variação	0,13	0,39	0,19	0,31
Assimetria	0,63	0,51	-0,42	0,14
Curtose	-0,61	-0,26	-1,22	-1,36

As tabelas acima demonstram que a média e a mediana dos dados para cada emoção possuem valores muito próximos, o que sugere uma

distribuição simétrica dos dados. Os valores de assimetria,⁸¹ embora não sejam iguais a zero, são valores próximos a esse valor. A curtose,⁸² com exceção feita ao medo (não nativos), apresentou valores mais distantes de zero. Além disso, os valores foram todos negativos, apontando para uma forma de distribuição leptocúrtica. O coeficiente de variação teve maior valor na emoção medo, independentemente do grupo de participantes, o que demonstra que os dados para a emoção medo são menos homogêneos que os demais. Abaixo, os conjuntos de gráficos 43 e 44 ilustram a distribuição dos dados padronizados com Z-score:

Gráfico 50: Curvas de distribuição de frequência para cada uma das emoções julgadas pelos participantes nativos. As curvas superiores, da esquerda para direita representam raiva e alegria; as inferiores, a partir da esquerda, representam tristeza e medo. As curvas foram feitas com dados padronizados por Z-score

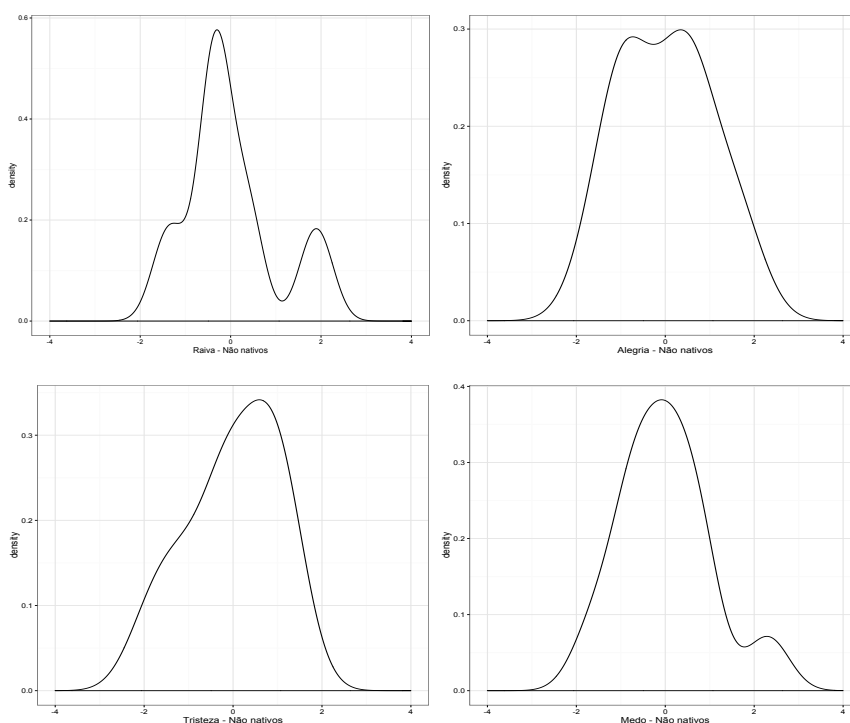


⁸¹ Na distribuição simétrica, a moda, a mediana e a média coincidem: $Mo = \tilde{X} = \bar{X}$. Na assimetria positiva, tem-se $Mo < \tilde{X} < \bar{X}$, com maior ocorrência de valores baixos na distribuição. No caso de assimetria negativa, tem-se $\bar{X} < \tilde{X} < Mo$, com maior ocorrência de valores altos na distribuição.

⁸² Curtose diz respeito ao grau de achatamento da distribuição. Há três tipos de referência: leptocúrtica (mais aguda na parte superior), mesocúrtica (distribuição normal) e platicúrtica (curva de frequência mais aberta e achatada).

As curvas acima demonstram que as distribuições dos dados tenderam a ser bimodais, a mais evidente delas é a alegria (canto superior direito). A assimetria tendeu ligeiramente a ser positiva para as emoções medo e raiva, ao passo que a tristeza apresentou ligeira inclinação negativa. A curtose tendeu a ser leptocúrtica, ou seja, com aparência mais aguda no centro da distribuição. O conjunto de gráficos 51 ilustra como foram as distribuições dos dados para os participantes não nativos:

Gráfico 51: Curvas de distribuição de frequência para cada uma das emoções julgadas pelos participantes não nativos. As curvas superiores, da esquerda para direita, representam raiva e alegria; as inferiores, a partir da esquerda, representam tristeza e medo. As curvas foram feitas com dados padronizados por Z-score

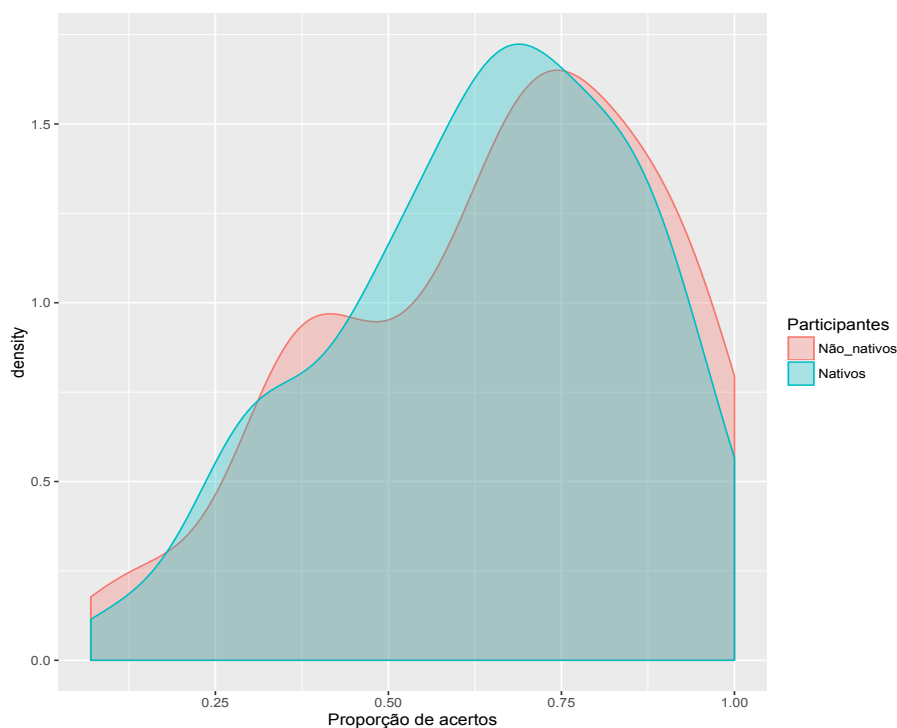


As curvas acima demonstram que as distribuições dos dados tenderam a ser bimodais em três casos (raiva, alegria e medo), exceto para tristeza. As emoções raiva, alegria e medo tiveram assimetria moderada, ou seja, as distribuições tenderam a ser negativas ou positivas de forma

moderada. Essa tendência mostrou-se positiva para as emoções medo e raiva, por outro lado, a tristeza apresentou ligeira inclinação negativa. A curtose novamente tendeu a ser leptocúrtica, mas isso pode ser objeto de discussão no caso da tristeza e da alegria.

O gráfico abaixo foi feito com o intuito de se ter uma visão geral sobre as *performances* de ambos os grupos de participantes (nativos e não nativos). Nele encontram-se as proporções de respostas certas para cada grupo envolvendo todas as emoções julgadas no experimento. Pode-se notar que a sobreposição é quase completa, o que demonstra pouca variação de *performance* entre nativos e não nativos.

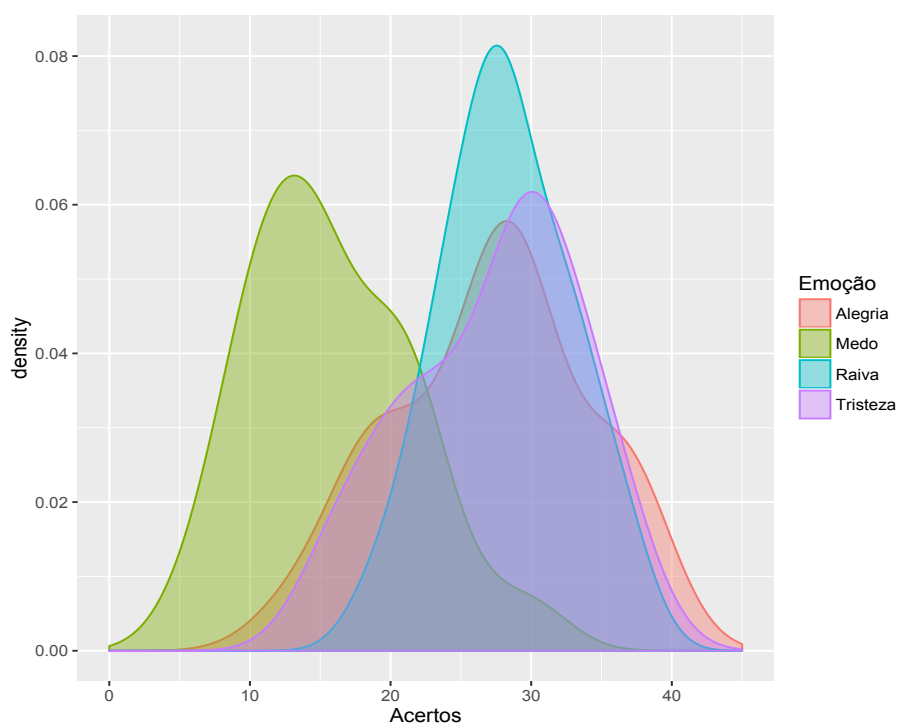
Gráfico 52: Curvas de distribuição das proporções de acertos de nativos e não nativos



O Gráfico 52 ilustra um pico de acertos por volta de 60% a 70%, corroborando os dados apresentados nos quadros de 8 a 11. Apesar de

ilustrar o desempenho parelho entre os participantes, o Gráfico 52 não transparece a *performance* inferior de ambos os grupos (nativos e não nativos) quando julgaram a emoção medo. Pode-se pensar que as respostas para a emoção medo estão concentradas em 0,25 a 0,50 do eixo x. O teste de *inter-rater reliability* de Kappa Fleiss apresentou valor $K = 0,64$ (nível de concordância de razoável a bom), corroborando a hipótese de que as *performances* foram razoavelmente parelhas. Dessa forma, agora que já sabemos que há uma concordância de *performance* entre os grupos, podemos juntar todos os julgamentos num mesmo gráfico, mas separando as emoções. O resultado está no Gráfico 53:

Gráfico 53: Curvas de distribuição de acertos de nativos e não nativos para cada uma das emoções (alegria, medo, raiva e tristeza)

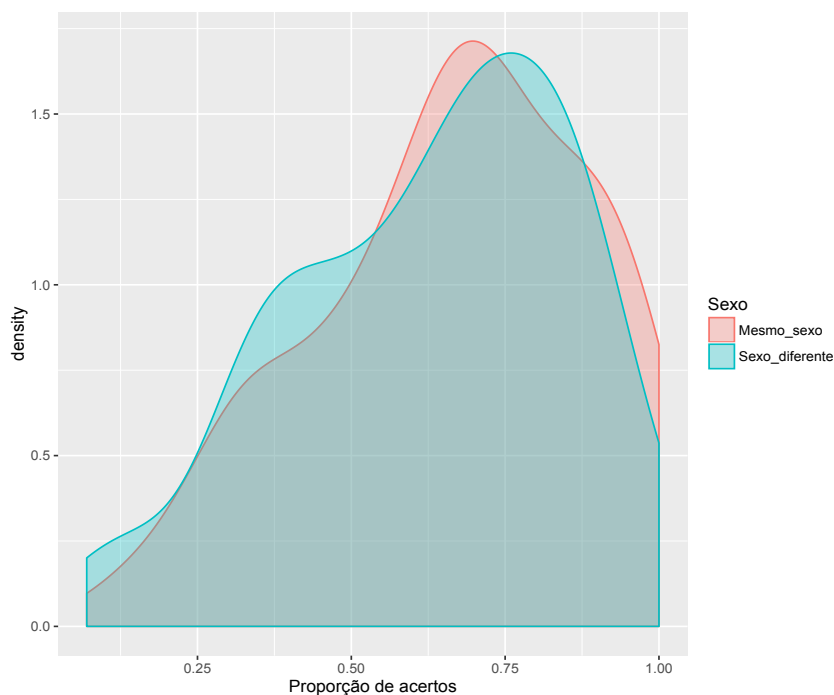


Como esperado, o medo ficou à parte das demais emoções por conta do número inferior de acertos se comparado com as demais emoções. Apesar do baixo número de julgamentos corretos, os resultados para

emoção medo ficaram por volta de 50%, o que não se pode desprezar a princípio.

Outra variável analisada neste experimento foi o sexo dos participantes. O propósito dessa variável é verificar se há influência do sexo no julgamento das emoções. Em outras palavras, seria possível que homem julgando homem, mulher julgando mulher, ou mulher julgando homem e vice-versa traria alguma consequência na maneira como os participantes julgaram os estímulos?

Gráfico 54: Curvas de distribuição das proporções de acertos de nativos e não nativos de acordo com o sexo. A curva de cor salmão ilustra as respostas dadas por participantes de mesmo sexo, ou seja, quando homens ouviram estímulos de voz masculina, e mulheres, de voz feminina. A curva verde ilustra as respostas dadas por homens e mulheres a estímulos diferentes, ou seja, homens ouvindo voz feminina e mulheres ouvindo voz masculina



As curvas de proporção de acertos de acordo com o sexo assemelham-se bastante com a encontrada no Gráfico 52 sobre os grupos de participantes. Parece não haver influência da variável sexo na *performance*

dos participantes, pois os desempenhos não variaram quando os participantes ouviram vozes iguais ou diferentes do seu próprio sexo.

Antes de iniciar a análise probabilística, é necessário testar os pressupostos de normalidade e homogeneidade da variância dos dados. Foram feitas análises prévias de resíduos por meio do teste de Shapiro-Wilk e testes de Fligner-Killeen para cada modelo de análise de variância feito para este experimento. A variável dependente foi o número de acertos, e os fatores foram 1) emoção, 2) duração do estímulo, 3) grupo de participantes (nativos e não nativos), 4) sexo e 5) número de vezes que os participantes ouviram cada estímulo.

Tabela 18: Resultados dos testes dos pressupostos de normalidade (Shapiro-Wilk) e homogeneidade da variância (Fligner-Killeen). Os testes não foram significativos nos modelos em que as células estão destacadas. Nesses casos, os pressupostos não foram violados

<i>Modelos</i>	<i>Shapiro-Wilk</i>	<i>Fligner-Killeen</i>
1) Emoção	(W = 0,973, p = 0,001)	<i>p > 0,05</i>
2) Duração do estímulo	(W = 0,960, p = 0,092)	<i>p > 0,05</i>
3) Grupo de participantes	(W = 0,961, p = 0,001)	<i>p > 0,05</i>
4) Sexo	(W = 0,960, p = 0,095)	<i>p > 0,05</i>
5) Vezes (audição do estímulo)	(W = 0,965, p = 0,001)	p < 0,05

De acordo com os resultados acima, foram realizadas análises de variância (ANOVA de 1 fator) e teste de Kruskal-Wallis. De todos os modelos, os únicos que preencheram os dois pressupostos, o de normalidade e de homogeneidade das variâncias, foram os modelos 2 e 4; todos os demais violaram pelo menos um deles. Por conta disso, uma ANOVA foi feita para o modelo 2 e outra para o modelo 4; os demais foram analisados pelo teste de Kruskal-Wallis (não paramétrico).⁸³

Os resultados dos testes estão abaixo:

⁸³ Os métodos não paramétricos de inferência estatística não necessitam dos requisitos como normalidade e homogeneidade da variância, além de serem apropriados para amostras pequenas.

Tabela 20: Resultados das análises de variância (ANOVA) e dos testes de Kruskal-Wallis. O modelo 1 (acertos e emoção) obteve resultado significativo

<i>Modelos</i>	<i>Teste</i>	<i>Resultado</i>
1) Emoção	Kruskal-Wallis	$p < 0,05$
2) Duração do estímulo	ANOVA	$F(1,180) = 0,57, p = 0,45$
3) Grupo de participantes	Kruskal-Wallis	$p > 0,05$
4) Sexo	ANOVA	$F(1, 180) = 0,10, p = 0,99$
5) Vezes (audição do estímulo)	Kruskal-Wallis	$p > 0,05$

Somente o modelo 1 apresentou resultados significativos. Um teste *post hoc* de Tukey HSD foi feito para o modelo 1 com intuito de verificar quais diferenças foram significativas entre as emoções. O resultado apontou para uma diferença significativa entre medo e as demais emoções, corroborando os resultados expostos do Gráfico 53. A próxima etapa da análise será analisar o desempenho de cada participante utilizando o índice d' (d'prime) proposto pela Teoria da Detecção (MACMILLAN; CREELMAN, 2005).

9.3.2 Análise com d-prime

A Teoria da Detecção – TD – (MACMILLAN; CREELMAN, 2005)⁸⁴ tem como intuito aferir o desempenho de participantes de experimentos de percepção em que o sinal (visual ou auditivo) deve ser detectado na presença de ruído. Em outras palavras, os participantes são avaliados quanto a sua habilidade de extrair um determinado sinal do ruído. No caso deste estudo, a tarefa é detectar qual emoção está sendo veiculada por meio de *thin slices*.

A grande vantagem da TD é o uso do índice d' (*d-prime* ou *d-linha*), que é um quantificador da *sensibilidade* do participante com relação às características associadas ao alvo a detectar, ou seja, d' reflete o grau de separação entre o sinal e o ruído. Para mensurar a sensibilidade do participante, a TD não analisa somente os acertos dos participantes (*hits*),

⁸⁴ A TD é baseada na Teoria da Detecção do Sinal (GREEN; SWETS, 1966).

mas leva em consideração também os erros cometidos (*false alarms*). Com essas duas medidas, acertos (H) e falsos alarmes (FA), consegue-se obter o valor d' , que é calculado subtraindo-se a inversa normal da distribuição dos FA da inversa normal dos H:

$$d' = z(H) - z(FA)$$

Quando os valores de H forem superiores aos de FA, tem-se a possibilidade de avaliar a sensibilidade dos participantes ($d' > 0$), mas se $d' = 0$, as distribuições de H e FA estão sobrepostas e indicam que não há distinção entre o sinal e o ruído. Como exemplo, imagine um experimento com 30 estímulos (15 com sinal presente e 15 com sinal ausente), em que um participante julgou 21 estímulos dos 30 disponíveis como sendo portadores do sinal. Das 21 detecções, 11 foram feitas corretamente (H) e 10 incorretamente (FA). Nessa situação, temos $H \approx FA$ e $d' \approx 0$. Considerando a fórmula acima $d' = z(H) - z(FA)$, temos, com as devidas proporções de H e FA: $d' = z(0,73) - z(0,66)$. O valor de d' é de aproximadamente 0,2, ou seja, valor demasiado baixo (próximo de zero), demonstrando que o participante não foi capaz de diferenciar o sinal do ruído.

Como exemplo contrário, considere um participante que apresente uma alta taxa de acertos e uma baixa taxa de falsos alarmes; então $H > FA$ e $d' > 0$. Considerando o mesmo experimento acima (com 30 estímulos), o participante reconheceu 15 estímulos como sendo sinal. Dos 15 estímulos reconhecidos, 14 foram corretamente detectados (H), mas um estímulo foi detectado incorretamente como sinal (FA). Aplicando novamente a fórmula $d' = z(H) - z(FA)$, temos: $d' = z(0,93) - z(0,06)$. O valor de d' é de aproximadamente 3,00, valor que indica uma separação satisfatória entre sinal e ruído. Os dados a serem analisados com a aplicação da TD são provenientes do experimento com escolha forçada (4MFC). Nesse tipo de experimento, cada tentativa apresenta um sinal (emoção a ser detectada) e um ou mais ruídos (demais emoções ilustradas pelas figuras AKDEF).

9.3.3 Resultados

Os índices d' foram calculados para cada participante de ambos os grupos (nativos e não nativos), julgando cada uma das emoções (raiva, medo, tristeza e alegria). Como dito acima, os valores d' indicam o grau de sensibilidade de cada participante para cada emoção em particular. À medida que os valores se afastam de zero ($d' > 0$), o grau de sensibilidade aumenta e o preenchimento de cada célula da tabela se torna mais escuro. A Tabela 21 (a seguir) ilustra a distribuição desses valores.

De maneira geral, os participantes tiveram um grau de sensibilidade elevado. Se imaginarmos um *continuum* dos tons de cinza utilizados na Tabela 21, nota-se que há uma concentração maior de valores na região mais escura desse espectro. Entretanto, alguns valores baixos foram encontrados. No caso do medo, os valores de d' não foram tão elevados como nas demais emoções, o que era previsível dados os resultados expostos no Quadro 11 (p. 198).

Tabela 20: Valores de d' para nativos e não nativos para cada participante de acordo com cada emoção (raiva, medo, tristeza e alegria). Os índices d' são preenchidos com diferentes tons de acordo com o valor. O preenchimento torna-se mais escuro à medida que os valores d' aumentam

Valores d' - nativos					Valores d' - não nativos				
Participantes	Raiva	Medo	Tristeza	Alegria	Participantes	Raiva	Medo	Tristeza	Alegria
1	2.56	1.38	3.30	3.71	1	0.97	0.29	1.01	2.24
2	0.42	1.38	1.38	1.38	2	2.19	1.70	2.24	2.35
3	2.35	0.92	1.64	2.27	3	2.88	0.79	1.20	1.80
4	3.29	1.90	2.19	3.30	4	1.24	1.43	2.35	1.64
5	2.59	0.21	2.54	1.93	5	3.92	1.24	2.02	2.59
6	0.16	1.40	1.43	1.70	6	2.60	0.95	2.77	1.38
7	2.35	0.65	2.35	2.59	7	2.06	1.22	1.43	0.95
8	1.81	1.01	1.90	1.90	8	2.47	0.42	1.93	1.81
9	0.29	1.52	1.16	1.16	9	1.80	2.10	2.05	0.55
10	0.98	1.65	1.59	1.59	10	2.02	1.01	2.27	3.71
11	1.09	0.95	2.56	1.93	11	2.56	1.80	4.24	1.80
12	2.59	1.09	2.35	2.27	12	3.55	1.70	2.35	2.02
13	2.06	2.35	2.27	2.77	13	2.47	2.77	3.92	3.30
14	0.71	-0.24	0.41	0.64	14	3.19	1.39	3.00	2.12
15	1.59	1.80	2.05	1.81	15	2.88	2.06	1.93	2.35
16	3.41	1.16	3.92	2.88	16	2.54	0.76	1.80	1.30
17	1.73	0.63	2.02	3.29	17	2.35	1.65	2.12	2.05
18	1.64	0.29	1.91	3.55	18	1.40	1.01	2.19	1.43
19	3.71	2.12	1.52	2.27	19	2.60	1.48	2.56	1.81
20	2.60	0.63	2.02	2.88	20	2.35	2.77	2.06	1.40
21	2.06	0.95	2.54	1.64	21	1.38	1.18	2.27	2.47
22	2.19	1.48	2.27	4.65	22	1.40	0.87	1.38	1.64
23	1.38	0.33	0.19	3.41	23	2.05	1.22	2.77	2.33
24	2.05	0.43	1.22	1.80	24	1.65	1.22	3.19	1.64
25	2.27	0.65	2.56	3.30	25	2.59	0.95	1.54	2.05
26	1.59	0.92	1.90	2.15	26	1.80	1.16	1.59	1.20
27	1.52	1.48	0.54	1.76	27	2.60	1.48	1.30	1.70
28	1.43	0.79	2.19	1.76	28	2.12	1.07	2.35	2.02
29	2.60	1.90	2.12	1.90	29	2.77	2.77	2.27	2.02
30	2.27	1.38	3.55	2.88	30	2.77	1.70	2.35	2.76
31	1.90	0.71	2.12	1.81	31	2.77	1.43	2.60	2.12
32	2.35	1.91	2.35	2.60	32	2.05	1.24	1.90	2.56
33	1.81	0.71	2.76	1.81	33	2.35	1.09	1.90	2.27
34	2.06	2.76	3.29	2.60	34	1.93	1.16	2.56	1.38
35	1.81	1.59	1.65	1.93	35	2.47	0.95	2.60	1.22
36	2.35	1.22	1.59	3.19	36	2.06	1.59	1.64	1.52
37	3.29	0.63	1.65	1.76	37	2.15	1.22	2.02	1.59
38	1.20	0.87	1.59	1.65	38	3.41	2.77	2.35	2.12

As colunas com os valores de medo para os dois grupos de participantes presentes na Tabela 21 apresentam o maior número de índices d' com valores abaixo de um. Embora sejam valores inferiores aos encontrados nas demais emoções, isso não significa que possam ser descartados *a priori*. Alguns casos são limítrofes ao valor 1, outros

ultrapassaram esse valor, chegando próximo a 2. Este segundo grupo já apresenta valores d' de sensibilidade similares aos das outras emoções. A sensibilidade para cada uma das emoções variou de acordo com cada emoção. Vejamos abaixo a descrição estatística dos valores de d' para os dois grupos de participantes:

Tabela 21: Estatística descritiva de d' – nativos

	Nativos			
	<i>Raiva</i>	<i>Medo</i>	<i>Tristeza</i>	<i>Alegria</i>
<i>Valor mínimo</i>	0,16	-0,24	0,19	0,64
<i>Valor máximo</i>	3,71	2,76	3,92	4,65
<i>Variação</i>	3,55	3,00	3,73	4,01
<i>Mediana</i>	2,05	1,05	2,04	2,04
<i>Média</i>	1,95	1,15	2,02	2,33
<i>SE da média</i>	0,13	0,10	0,13	0,13
<i>Desvio padrão</i>	0,82	0,64	0,79	0,82
<i>Coef. Variação</i>	0,42	0,56	0,39	0,35
<i>Assimetria</i>	0,16	0,32	0,02	0,65
<i>Curtose</i>	0,13	0,08	0,78	0,46

Tabela 22: Estatística descritiva de d' – não nativos

	Não nativos			
	<i>Raiva</i>	<i>Medo</i>	<i>Tristeza</i>	<i>Alegria</i>
<i>Valor mínimo</i>	0,97	0,29	1,01	0,55
<i>Valor máximo</i>	3,92	2,77	4,24	3,71
<i>Variação</i>	2,95	2,48	3,23	3,16
<i>Mediana</i>	2,35	1,23	2,21	1,92
<i>Média</i>	2,33	1,41	2,21	1,93
<i>SE da média</i>	0,10	0,10	0,11	0,10
<i>Desvio padrão</i>	0,63	0,61	0,66	0,61
<i>Coef. Variação</i>	0,27	0,43	0,30	0,32
<i>Assimetria</i>	-0,17	0,91	0,96	0,57
<i>Curtose</i>	0,36	0,76	2,02	1,35

De maneira geral, os valores mínimos e máximos não variaram muito, inclusive se considerarmos os resultados para o medo. O mesmo é válido para os valores máximos, exceto uma ligeira vantagem para a alegria (4,65) nos nativos e tristeza nos não nativos (4,24). A variação entre os valores mínimos e máximos foi estável, entre 2,48 e 4,01, somando-se os dois grupos. A média da distribuição de d' ficou por volta de 2,00 para ambos os grupos, com exceção do medo, que apresentou médias de 1,15

para os nativos e 1,41 para os não nativos. O desvio padrão também foi uma característica estável ao longo dos dados. Para os nativos, os valores ficaram por volta de 0,60, com exceção da raiva (0,82). No caso dos não nativos, o desvio padrão variou por volta de 0,60 para todas as emoções.

As distribuições dos valores de d' para os nativos tiveram uma ligeira assimetria positiva, mas, se atentarmos para a tristeza, o valor foi muito próximo de zero, indicando simetria dos dados. No caso dos não nativos, os valores ficaram próximo de 1 para medo e tristeza, e 0,57 para a alegria, todos indicando uma assimetria positiva. A exceção à regra foi a raiva, com um valor de -0,17, indicando uma assimetria negativa. A curtose das distribuições foi, de maneira geral, leptocúrtica, ou seja, com os dados agrupados majoritariamente em torno da moda.

Outra característica importante é o desempenho desses participantes com relação às outras emoções. Como aparentou anteriormente, a diferença de desempenho dos participantes quando expostos a estímulos da emoção medo pareceu estar ligada aos estímulos da emoção, pois as demais emoções apresentaram resultados parelhos entre si.

Entretanto, com a aplicação da TD, consegue-se mensurar se há uma tendência diferente para cada participante em cada uma das emoções. Para tornar mais claro, tomemos dois exemplos opostos. O primeiro é o participante nativo 14, que apresentou os seguintes valores d' para as emoções: raiva = 0,71, medo = -0,24,⁸⁵ tristeza = 0,41 e alegria = 0,64. Pelos índices obtidos pelo participante 14, fica claro que não se trata de uma variação motivada unicamente pelo tipo de emoção, pois os demais valores d' das outras emoções também foram demasiadamente baixos.

⁸⁵ Note-se que o valor negativo $d' = -0,24$ é caso único dentre todos os participantes do experimento. Quando $d' < 0$, tem-se a proporção de FA maior do que a proporção de H.

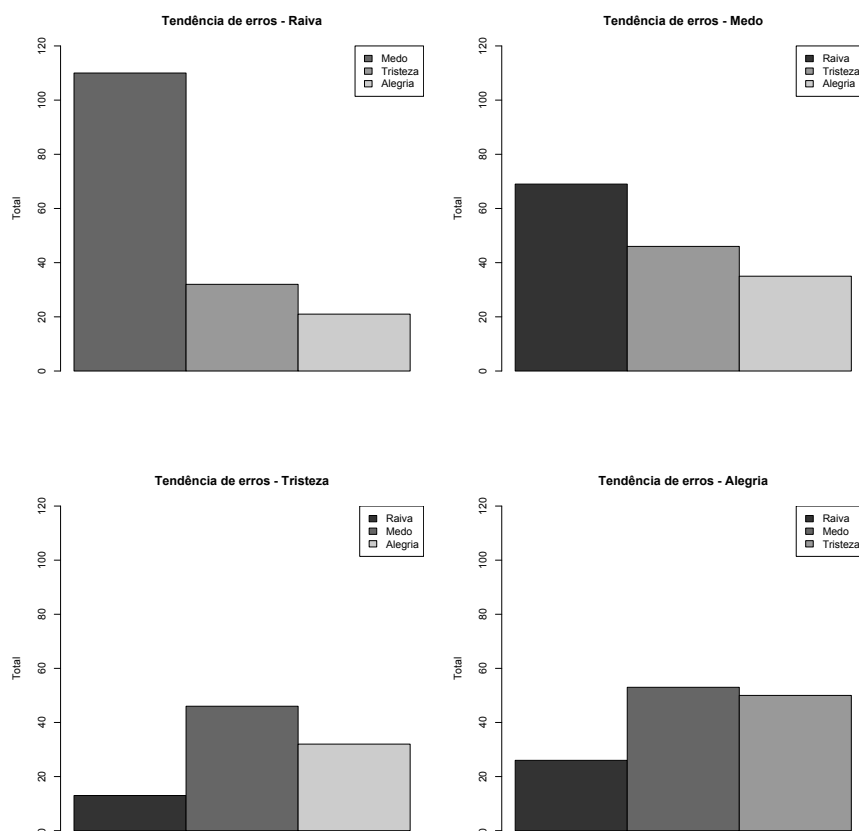
Nesse caso, parece que o participante não obteve êxito em julgar as emoções adequadamente.

O segundo exemplo é o participante 5, que apresentou os seguintes valores d' para as emoções: raiva = 2,59, medo = 0,21, tristeza = 2,54 e alegria = 1,93. No caso do participante 5, somente a emoção medo apresentou valor baixo de d' , ao passo que os outros valores seguiram a tendência geral das demais emoções. A partir desses dois exemplos, fica claro que a diferença de *performance* dos participantes com relação ao medo não pode ser somente atribuída a uma causa intrínseca do estímulo, mas também a fatores extrínsecos como a sensibilidade de cada participante.

Além do desempenho dos participantes com relação aos acertos, é necessário um olhar mais detido aos julgamentos desviantes que ocorreram durante o experimento. Por isso, a tendência dos falsos alarmes (FA) é importante para tentar entender qual a direção dos FA de acordo com cada uma das emoções.

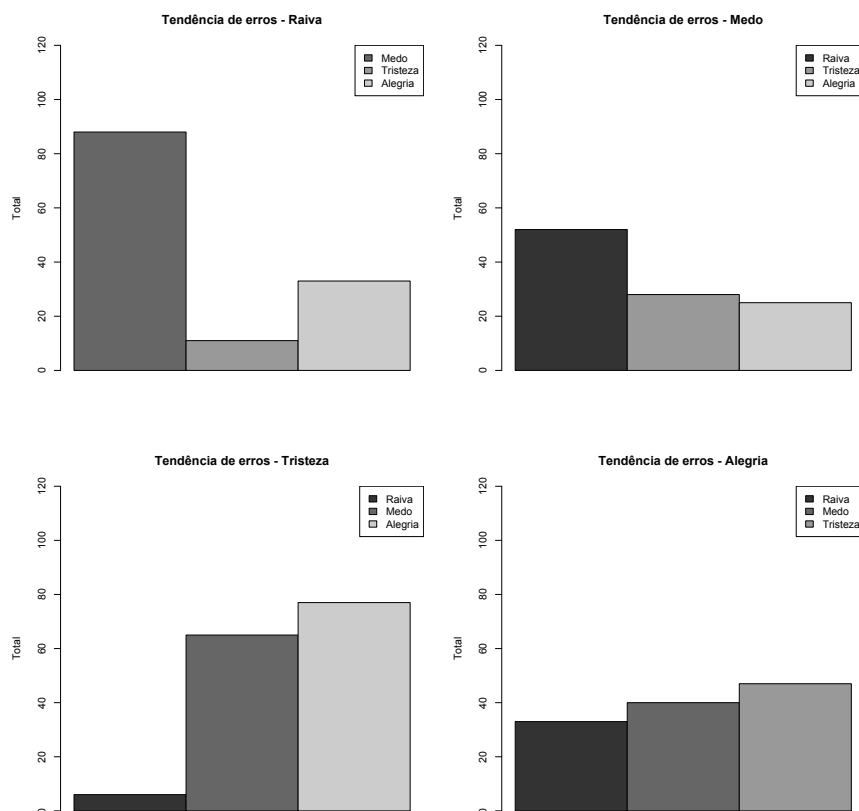
Para os nativos, os erros dos julgamentos da raiva foram divididos em 67,5% para o medo, 19,7 para a tristeza e 12,8 para a alegria. Para a emoção medo, 46% para a raiva, 30,6 para tristeza e 23,4% para a alegria. Para os julgamentos da emoção tristeza, 14,3 para a raiva, 50,5% para o medo e 35,2% para a alegria. Por fim, os erros para a alegria foram 20% para a raiva, 41% para o medo e 39% para a tristeza.

Gráfico 55: Conjunto de gráficos de barras com as tendências de erros para o grupo de participantes nativos. No canto superior esquerdo está a distribuição dos erros para a emoção raiva, no canto superior direito estão os resultados para a emoção medo. Abaixo deles, da esquerda para a direita, estão os resultados para a tristeza e a alegria



As tendências de erros da emoção raiva e medo foram correspondentes, ou seja, a maioria dos erros de julgamentos da emoção raiva foi direcionada ao medo. Da mesma forma, quase metade dos erros de julgamento da emoção medo foi direcionada à raiva. No caso da tristeza e da alegria, os erros foram mais ou menos equivalentes, sendo a raiva a emoção com menos ocorrências, 13 e 26, respectivamente. O gráfico a seguir ilustra as tendências de erros para os participantes não nativos.

Gráfico 56: Conjunto de gráficos de barras com as tendências de erros para o grupo de participantes não nativos. No canto superior esquerdo está a distribuição dos erros para a emoção raiva, no canto superior direito estão os resultados para a emoção medo. Abaixo deles, da esquerda para a direita, estão os resultados para a tristeza e a alegria



Os resultados para as emoções raiva e medo são, de maneira geral, similares aos encontrados no grupo dos nativos, em outras palavras, raiva e medo tiveram erros correspondentes. A única diferença foi o número inferior de julgamentos da emoção raiva em direção à tristeza (canto superior esquerdo do Gráfico 58). No caso da tristeza, a maioria dos erros dos não nativos foi em direção à alegria, diferentemente do encontrado nos nativos, os quais apresentaram maior quantidade de erro em direção ao medo (canto inferior esquerdo do Gráfico 58). No caso da alegria, os resultados foram parelhos entre os grupos (nativos e não nativos), mas com uma ligeira vantagem da tristeza com relação ao medo.

Os resultados até então confirmaram que os participantes

apresentaram uma significativa sensibilidade aos estímulos de curta duração – *thin slices*. A TD mostrou-se uma ferramenta profícua para mensurar o desempenho dos participantes coletiva e individualmente. Os achados deste experimento contribuirão para o debate sobre as estratégias que os sujeitos utilizam para perceber estímulos de natureza emocional, mesmo quando a informação disponível é desprovida de contexto linguístico e pragmático, além de limitada do ponto de vista temporal.

10. Discussão e considerações finais

Os experimentos desenvolvidos nesta tese proporcionaram uma série de resultados, alguns deles nos levam a algumas conclusões interessantes, outros a questionamentos que servirão de proposição para estudos futuros. A primeira questão a se responder, afinal, é como a percepção da emoção acontece entre pessoas falantes da língua em questão – PB –, e pessoas que não conhecem essa língua.

Se considerarmos os resultados do teste piloto com emoções básicas e dos experimentos com a abordagem dimensional, fica evidente que a percepção dos nativos e não nativos guia-se por características que estão além do universo linguístico de cada grupo. O resultado significativo dos não nativos com fala normal é um bom argumento para fundamentar a hipótese de que a emoção se manifesta a um nível que se sobrepõe ao dado

linguístico. O mesmo pode ser dito sobre o resultado dos nativos com fala delexicalizada. Se considerarmos somente esses resultados, ficamos propensos a validar sem ressalvas a hipótese universalista da emoção. No entanto, é importante lembrar que os resultados dos nativos com fala delexicalizada, embora significativos, foram inferiores àqueles obtidos pelo mesmo grupo com fala normal.

Para os não nativos, os resultados foram ainda mais baixos quando tiveram de julgar estímulos delexicalizados. Por conta disso, os resultados significativos dos dois grupos ao longo dos experimentos não invalidam a hipótese de que há uma força de natureza cultural e linguística que exerce influência na percepção dos participantes. Ficou claro nos resultados que o conhecimento da língua teve influência na maneira como os participantes julgaram os estímulos, comparando os dois grupos de participantes (nativos e não nativos) e os dois tipos de estímulos (fala normal e delexicalizada). Neste sentido, a proposta de Scherer (1996) sobre os efeitos *push* e *pull* é bastante esclarecedora na medida em que os experimentos desta tese tentaram isolar, de alguma forma, as ações desses dois efeitos. O efeito *push* representa o que é universal e presente na expressão motora e na fala; ao passo que o efeito *pull* é atribuído às normas culturais de expressão e às características de cada língua. Quanto mais limitado foi o acesso à informação fonético-linguística,⁸⁶ mais difícil se tornou a tarefa. Se colocarmos esses resultados num plano cartesiano, a sua distribuição poderia ser ilustrada, *grosso modo*, como na Figura 12.

⁸⁶ O uso do termo fonética é necessário se pensarmos a partir do ponto de vista do participante não nativo. O participante não nativo não tem conhecimento da língua veiculada no estímulo. Por essa razão, a perda de informação lexical, sintática, etc., não desempenha nenhum papel no julgamento dos não nativos. Há, nesse caso, somente uma limitação fonética. O uso do termo linguística é justificado, no caso dos nativos, pois o apagamento do léxico acarreta perda de informação lexical, morfológica, sintática, etc., características importantes para os nativos.

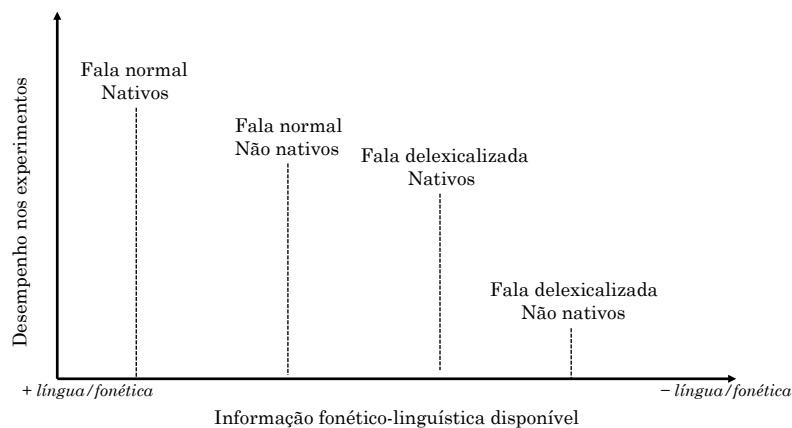


Figura 12: Ilustração do desempenho dos participantes dos experimentos de acordo com o acesso à informação fonético-linguística. O eixo y representa o desempenho dos participantes; quanto mais alta a posição, melhor a performance. O eixo x representa a disponibilidade da informação fonético-linguística.

Com a ilustração, fica mais fácil visualizar a influência do fator fonético-linguístico nos resultados. Ao analisar os dois melhores resultados (fala normal julgada por nativos e não nativos), nota-se que houve um declínio do desempenho quando os participantes não tinham conhecimento linguístico (não nativos). Nos estímulos com fala delexicalizada, o esperado era que os dois grupos de participantes tivessem apresentado resultados similares, mas isso não aconteceu. Como já comentado anteriormente, parece que há outro fator além da língua que justifique essa disparidade de resultados (qualidade de voz ausente), entretanto, essa diferença pode ser interpretada de modo diferente.

O desempenho diverso entre nativos e não nativos ao julgar estímulos com fala delexicalizada pode também ser justificado pelo conhecimento da língua materna. Estudos já demonstraram que falantes nativos de línguas diferentes, ou até mesmo falantes de diferentes variedades de um mesmo idioma, são capazes de reconhecer sua própria língua ou variedade exclusivamente por meio das características entoacionais (OHALA, 1978; BEZOOIJEN; GOOSKENS, 1999; PERES, 2011). Todavia, não estamos falando do uso da língua somente, mas da manifestação das emoções por meio dela. Ao conjecturar que o motivo do melhor desempenho dos nativos seja devido aos aspectos entoacionais

isolados pelo método de delexicalização, então obrigatoriamente devemos depositar na língua uma importância maior e mais profunda quando se trata de percepção da emoção. Ao tomar esse argumento como verdadeiro, o de que a língua conseguiu influenciar os julgamentos dos participantes por meio de informações estritamente entoacionais, devemos necessariamente assumir que o papel da língua está muito além do léxico ou do entendimento do que está sendo dito.

A diferença entre os nativos e os não nativos quando expostos a fala delexicalizada pode ser explicada como uma vantagem intragrupo, ou seja, aspectos da manifestação das emoções que estão intimamente ligados ao grupo linguístico e cultural em questão. Enfenbein e Ambady (2002, p. 205) descrevem essa vantagem como sendo uma correspondência entre a experiência social e cultural dos juízes do experimento e o alvo que deve ser percebido.

Numa perspectiva cognitivista, a diferença poderia ser atribuída à forma de *appraisal*. Essa maneira de avaliar as situações seria uma diretriz para que os participantes dos grupos (nativos e não nativos) julgassem os estímulos. Essa visão, como já visto anteriormente, também é compartilhada pela perspectiva social das emoções, ao entender que a emoção é sempre direcionada a um alvo e dependente de uma avaliação feita pelo sujeito sobre o objeto. Em outras palavras, a avaliação que cada grupo faz de situações potencialmente emocionais pode ser também um fator que explique os resultados distintos dos dois grupos.

De maneira geral, nativos e não nativos tiveram resultados significativos, mesmo com resultados diversos entre os grupos. Isso não significa que os participantes não nativos tenham falhado na tarefa de reconhecimento das emoções, tanto na abordagem com emoções básicas quanto na abordagem dimensional. Pelo contrário, os não nativos tiveram boa performance, exceto na fala delexicalizada. Os resultados desses experimentos deixaram transparecer que os fatores culturais e linguísticos

dão alguma vantagem para os nativos nas tarefas de reconhecimento das emoções.

Com base nos achados desses experimentos, pode-se considerar o termo *dialetos emocionais*, proposto por Enfenbein e Ambady (2002), como uma maneira de ilustrar como a manifestação e a percepção das emoções variam em diferentes culturas. O termo em si pressupõe emoções que sejam universais e comuns entre as culturas, mas que apresentam aspectos diferentes em cada uma delas. Apesar do termo *dialeto* ser metafórico, o emprego do termo faz sentido pela variação local que as emoções apresentam. As diferenças não são salientes de modo que impossibilite o entendimento, mas são sutis o bastante para causar confusão e dúvida. É dentro desse espectro que os resultados dos experimentos com emoções básicas e abordagem dimensional se encaixam, pois apresentaram erros e acertos que podem ser atribuídos às diferenças tênues que as emoções apresentam em culturas e línguas diferentes. Um passo importante, como efeito de contraste, seria aplicar o mesmo experimento, mas invertendo os papéis dos nativos e dos não nativos, ou seja, utilizando a língua inglesa como estímulo aos participantes.

Passando ao último experimento da tese, como explicar os resultados obtidos com estímulos *thin slices*? Ao contrário do que foi encontrado no experimento com abordagem dimensional, não houve diferença significativa entre o desempenho dos nativos e dos não nativos. Vale lembrar que ambos tiveram acesso restrito ao que estava sendo dito, ou seja, o léxico utilizado não tinha nenhuma relação com as emoções em questão. A questão que se impõe é: se o acesso à língua determinou o melhor desempenho em ambos os experimentos anteriores, como explicar os resultados significativos de nativos e não nativos no experimento com *thin slices*. Se os participantes foram capazes de julgar de maneira satisfatória um *input* tão empobrecido de informação lexical e de duração, alguma característica relevante foi veiculada via acústica da fala para que

fosse possível chegar aos resultados obtidos nesse experimento. Considerando que o maior estímulo teve a duração de 1369 ms, é plausível pressupor que as informações necessárias para a percepção da emoção podem estar contidas em finas *slices* de tempo.

Com base nos resultados obtidos nesta tese, a hipótese defendida é a de que a percepção da emoção na fala deve obedecer a um princípio que também está presente na fala como um todo: a *redundância de informação*. Assim, a percepção da emoção na fala é possível mesmo com a escassez de informação do sinal acústico.

Como já dito nas justificativas do experimento com *thin slices*, a fala tem mais informação do que é preciso para percebê-la e decodificá-la. Da mesma forma, quando a fala é o meio pelo qual as emoções são manifestadas, seria contra intuitivo pensar que seria diferente. Pensando em todos os experimentos, tivemos 3 tipos de estímulos: fala normal, delexicalizada e *thin slices*. Com exceção da fala normal, os demais estímulos foram manipulados de maneira que algumas características fossem omitidas. No caso da fala delexicalizada, o léxico e a qualidade de voz foram retirados do sinal acústico; no caso das *thin slices*, a limitação da informação se deu por conta da curta duração dos estímulos. No entanto, os participantes conseguiram julgar os estímulos de forma significativa.

Essa hipótese faz todo sentido se pensarmos na perspectiva evolucionista, segundo a qual as emoções seriam adaptações ao ambiente com funções de preservação. A redundância de informação seria útil para percebermos ou manifestarmos emoções de maneira mais eficiente e inequívoca. Além disso, o reconhecimento significativo dos participantes, baseado em tão pouca informação (*thin slices*), também pode ser um indício de que a rapidez de decisão em situações emocionais seja um fator importante para a preservação da espécie.

Além de considerar a redundância de informação para a percepção da emoção, faz-se necessário entendê-la como um princípio que opera em

dois níveis: um de *curto termo* e outro de *longo termo*. Por *longo termo*, entende-se toda informação que tenha estreita relação com o fator tempo, ou seja, que necessita de intervalos de tempo maiores para se estabelecer. Como exemplo, a entoação, mais precisamente a macroentoação, é um fator de *longo termo* que foi importante para os resultados obtidos nos dois primeiros experimentos desta tese. Por *curto termo*, entende-se as variações que prescindam do tempo ou que, embora transcorram nele, necessitam de um tempo curto para ocorrer. Como exemplo de informação de curto termo, temos a qualidade de voz e a microentoação.

Por que essa subdivisão é necessária? A redundância de informação disponível para os ouvintes deve obedecer a essa divisão para que se possa atingir a finalidade da comunicação, em outras palavras, a emoção deve ser decodificada pelo ouvinte de maneira apropriada mesmo quando características importantes do sinal da fala estão ausentes. Por conta disso, a limitação de informação pode acontecer somente em um dos níveis, – curto ou longo termo –, mas não deve ocorrer em ambos simultaneamente. Pensemos nos experimentos com fala normal e delexicalizada. O método de delexicalização limitou o acesso dos participantes à qualidade de voz, mas deixou disponível as variações de F0 e características rítmicas dos estímulos. Com essa informação, os falantes nativos conseguiram julgar com considerável êxito os estímulos. Como contraponto, o experimento com *thin slices* limitou o acesso dos participantes a variações entoacionais de longo termo. Isso não significa que não houve variações entoacionais dessa natureza, mas que essas variações estavam incompletas ou insuficientemente representadas nos estímulos. Assim, a entoação como um fenômeno de longo termo (macroentoação) pode ser entendida como uma característica suficiente para a percepção da emoção, mas não necessária, i. e., não é necessário que haja informações entoacionais de longo termo para que as emoções sejam corretamente reconhecidas, entretanto em alguns contextos ela é suficiente para tal. Seguindo o mesmo

raciocínio, podemos pensar nas informações de curto termo. Se considerarmos a qualidade de voz, é possível que ela seja responsável pelo bom desempenho de ambos os grupos de participantes (nativos e não nativos) no experimento com *thin slices*.⁸⁷ Neste caso, então, a qualidade de voz poderia ser considerada como fator suficiente porque possibilitou o reconhecimento das emoções em excertos de curta duração; mas não necessário, pois estava ausente nos estímulos com fala delexicalizada.

Para testar a hipótese da subdivisão da redundância de informação, seria necessário que se produzisse estímulos que pudessem ser manipulados de modo que cada um dos níveis – curto e longo termo – fosse levado ao limite, ou seja, que se pudesse testar e mensurar qual seria a quantidade de informação necessária de cada um deles para a percepção da emoção. Tem-se ciência que é um trabalho árduo e de colaboração, mas que poderá ser útil para se ter um panorama mais claro sobre como a emoção é veiculada pela fala, além de contribuir para os estudos da variação da emoção entre as diferentes línguas.

⁸⁷ Fator que será explorado em trabalhos futuros.

11. Referências bibliográficas

AMBADY, N.; BERNIERI, F. J.; RICHESON, J. A. Toward a histology of social behavior: judgmental accuracy from thin slices of the behavioral stream. In: ZANNA, M. P. (Ed.). *Advances in Experimental Social Psychology*, v. 32, p. 201-272, 2000.

AMBADY, N.; KRABBENHOFT, M. A.; HOGAN, D. The 30-sec sale: using thin slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology*, v. 16, p. 4-13, 2006.

AMBADY, N.; ROSENTHAL, R. thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychological Bulletin*, v. 111, n. 2, p. 256-274, 1992.

AMBADY, N.; ROSENTHAL, R. Half a minute: predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, v. 64, n. 3, p. 431-441, 1993.

ARISTÓTELES. *Retórica*. 2. ed. revista. Lisboa: Imprensa Nacional Casa da Moeda, 2005.

ARNOLD, M. B. Emotions and personality. New York: Columbia University, 1960. (Vol. 1. *Psychological Aspects*).

AVERILL, J. R. A constructivist view of emotion. In: PLUTCHIK, R.; KELLERMAN, H. (eds.), *Emotion: theory, research and experience*. New York: Academic Press, 1980. p. 305-339.

AVERILL, J. R. Illusions of anger. In: FOLSON, R. B.; TEDESCHI, J. T. (eds.), *Aggression and violence: social interactionist perspectives*. Washington, DC: American Psychological Association, 1993. p. 171-192.

- BACHOROWSKI, J.; OWREN, M. J. Vocal expression of emotion. In: LEWIS, M.; HAVILAND, J. (Eds.). *Handbook of emotions*. 2. ed. New York: Guilford Press, 2000. p. 196-210.
- BAER, T.; MOORE, B. C. J. Effects of spectral smearing on the intelligibility of sentences in noise. *Journal of Acoustic Society of America*, v. 94, p. 1229-1241, 1993.
- BANSE, R; SCHERER, K. R. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*. v. 70, n. 3, p. 614-636.
- BÄNZIGER, T.; SCHERER, K. R. The role of intonation in emotional expressions. *Speech Communication*. v. 46, p. 252–267, 2005.
- BARBOSA, P. A.; ERIKSSON, A.; ÅKESSON, J. Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese. In: ASU, E. L.; LIPPUS, P. (eds.), *Nordic Prosody. Proceedings of the XIth Conference*, Frankfurt am Main: Peter Lang, Tartu, 2013. p. 97-106.
- BELIN, P.; FECTEAU, S.; BEDARD, C. Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Science*, v. 8, p. 129-135, 2004.
- van BEZOOIJEN, R.; GOOSKENS, C. Identification of language varieties: the contribution of different linguistic levels. *Journal of Language and Social Psychology*, v. 18, p. 31-48, 1999.
- BORGES, J. L. *História Universal da Infâmia*. Tradução de José Bento. Lisboa: Assírio e Alvim, 1982.
- BOUCHER, J. D. Culture and emotion. In: Marsella, A. J.; Tharp, R. G.; Ciborowski, T. V (eds.) *Perspectives on cross-cultural psychology*. San Diego, CA: Academic Press, 1979. p. 159-178.

- BRANDT, M. E.; BOUCHER, J. D. Concepts of depression in emotion lexicons of eight cultures. *International Journal of Intercultural Relations*, v. 10, p. 321-346, 1986.
- CAMBRIDGE ADVANCED LEARNER'S DICTIONARY. 4. ed. Cambridge. Cambridge University Press, 2013.
- CARRERE, S.; GOTTMAN, J. N. Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process*, v. 38, p. 143-158, 1999.
- CARSTAIRS, G. M. *The twice-born: A study of a community of high-caste Hindus*. Bloomington: Indiana University Press, 1967.
- CHURCH, A. T.; KATIGBAK, M. S.; REYES, J. A. S.; JENSEN, S. M. Language and organisation of Filipino emotion concepts: comparing emotion concepts and dimensions across cultures. *Cognition & Emotion*, v. 12, n. 1, p. 63-92, 1998.
- COHEN, A.; 't HART, J. On the anatomy of intonation. *Lingua*, v. 19, p. 177-192, 1967.
- COLAMARCO, M.; MORAES, J. A. Emotion expression in speech acts in Brazilian Portuguese: production and perception. In: *Fourth Conference on Speech Prosody. Campinas. Proceedings of the Speech Prosody 2008*. Campinas: Unicamp. p. 717-720, 2008.
- CONSONI, F. *O acento lexical como pista para o reconhecimento de palavras*. 2006. Dissertação (Mestrado) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8142/tde-22082007-152238/>>. Acesso em: 24 maio 2014.
- CONSONI, F. Aspectos da percepção da proeminência tonal em português brasileiro. Tese (Doutorado) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.

- CONSONI, F.; FERREIRA NETTO, W. A sensibilidade de percepção da variação em semitons ascendentes em palavras isoladas. In: CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA, 4., 2008, Niterói. Caderno de Resumos. Niterói: UFF, 2008. v. 1. p. 14-15.
- CONSONI, F.; FERREIRA NETTO, W. A percepção de variação em semitons ascendentes em palavras isoladas no Português Brasileiro. In: FERREIRA NETTO, W. (Org.). ExProsodia. 1. ed. São Paulo: Paulistana, 2016. p. 19-23.
- CORNELIUS, R. R. The science of emotion: research and tradition in the psychology of emotion. Upper Saddle River, NJ: Prentice-Hall, 1996.
- CORNELIUS, R. R. Theoretical approaches to emotion. Proceedings of the ISCA Workshop on Speech and Emotion. Northern Ireland, UK., p. 3-10, 2000.
- COSMIDES, L.; TOOBY, J. Evolutionary psychology and the emotions. In: LEWIS, M.; HAVILAND-JONES, J. M. (eds.), *Handbook of emotions*. 2. ed. New York: Guilford Press, 2000. p. 91-115.
- COSTANZO, F. S.; MERKEL, N. N.; COSTANZO, P. R. Voice quality profile and perceived emotion. *Journal of Counseling Psychology*, v. 16, n. 3 p. 267-270, 1969.
- CURHAN, J. R.; PENTLAND, A. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, v. 92, n. 3, p. 802-811, 2007.
- DAMÁSIO, A. R. *O erro de Descartes*. Emoção, razão e o cérebro humano. 3. ed. São Paulo: Cia das Letras, 2012. (1ª edição datada de 1994).
- DARWIN, C. *The expression of the emotions in man and animals*. Chicago: University of Chicago Press, 1965. (1ª edição datada de 1872).
- DAVITZ, J. R. Auditory correlates of vocal expression of emotional feeling. In: DAVITZ, J. R. *The communication of emotional meaning*. New

York: McGraw-Hill. 1964. p. 101-112.

DAVITZ, J. R. *The language of emotion*. New York: Academic Press, 1969.

DESCARTES, R. *The philosophical writings of Descartes*. Vol. 1.
Cambridge: Cambridge University Press, 1985.

DODDS, E. R. *The Greeks and the irrational*. Berkeley: University of California Press, 1951.

DUCHENNE, G. B. A. *The Mechanism of Human Facial Expression*.
Studies in Emotion and Social Interaction. CUTHBERTSON, R. A. (ed.).
Cambridge: Cambridge University Press. Disponível em Cambridge
Books Online <<http://dx.doi.org/10.1017/CBO9780511752841>> Acesso em
6 novembro 2012.

EKMAN, P. Universals and cultural differences in facial expressions of
emotion. In: Cole, J. K. (eds.), *Nebraska symposium on motivation*.
Lincoln: University of Nebraska Press, 1972. p. 207-282.

EKMAN, P. An argument for basic emotions. *Cognition and Emotion*, v.
6. p. 169-200, 1992.

EKMAN, P.; FRIESEN, W. V. *Unmasking the face: a guide to recognizing
emotions from facial clues*. Englewood Cliffs, New Jersey: Prentice-Hall,
1975.

ERIKSSON, A.; THUNBERG, G. C.; TRAUNMÜLLER, H. Syllable
prominence: A matter of vocal effort, phonetic distinctness and top down
processing. In P. Dalsgaard, B. Lindberg, H. Benner, & T. Zheng-Hua
(Eds.), *Proceedings of the 7th European Conference on Speech
Communication and Technology (Eurospeech '01)*, p. 399-402, 2001.

EYSENCK, M. W.; KEANE, M. T. *Cognitive Psychology: a student book*.
7.ed. New York: Psychology Press, 2015.

FANT, G. *Acoustic theory of speech production*. The Hague, Netherlands:
Mouton, 1960.

FERREIRA NETTO, W. *Variação de frequência e constituição da prosódia da língua portuguesa*. 2006. Tese (Livre-Docência na Área de Filologia e Língua Portuguesa) – Universidade de São Paulo, São Paulo, 2006.

FERREIRA NETTO, W. *Decomposição da entoação frasal em componentes estruturadoras e semântico-funcionais*. Trabalho apresentado no X Congresso nacional de Fonética e Fonologia/IV Congresso Internacional de Fonética e Fonologia. Universidade Federal Fluminense, Niterói, RJ, 2008.

FERREIRA NETTO, W. ExProsodia. *Revista da Propriedade Industrial – RPII*, Rio de Janeiro, v. 2038, n. 167, p. 167, 2010.

FERREIRA NETTO, W. ExProsodia. Resultados Preliminares. 1. ed. São Paulo: Paulistana, 2016.

FERREIRA NETTO, W.; PERES, D. O.; MARTINS, M. V. M.; VIEIRA, M. F. *Análise automática da entoação emotiva (colérica, triste e neutra) pelo aplicativo ExProsodia*. III Jornada de Descrição do Português. Fortaleza, 2013.

FERREIRA NETTO, W.; MARTINS, M. V. M.; VIEIRA, M. F. Efeitos da entoação e da duração na análise automática das manifestações emocionais. *Estudos Linguísticos*, São Paulo, v. 43, p. 23-32, 2014.

FERREIRA NETTO, W.; PERES, D. O.; MARTINS, M. V. M.; VIEIRA, M. F. Automatic analysis of emotional intonation in Brazilian Portuguese. *The Journal of the Acoustical Society of America*, v. 135, p. 2197-2198, 2014.

FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Desvendando os mistérios do coeficiente de correlação de Pearson (r). *Política Hoje*, v. 18, n. 1, p. 115-46, 2009. Disponível em: <<http://www.ufpe.br/politica/hoje/index.php/politica/article/view/6/6>>. Acesso em: 20 maio 2015.

- FILIPPOVA, E.; ASTINGTON, J. W. Children's Understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony. *Child Development*, 81, p. 913-928, 2010.
- FLEISS, J. *Statistical methods for rates and proportions*. New York: John Wiley & Sons, 1981.
- FÓNAGY, I. La mimique buccale. *Phonetica*, v. 33, p. 31-44, 1976.
- FÓNAGY, I. A new method of investigating the perception of prosodic features. *Language and Speech*, v. 21, n. 1, p. 34-49, 1978.
- FOURCIN, A. Precision Stroboscopy, Voice Quality and Electrolaryngography. In: KENT, R. D.; BALL M. J. (Orgs.) *Voice Quality Measurement*. San Diego: Singular Publishing Group 2000. Disponível em <http://www.laryngograph.com/pdfdocs/paper3.pdf>. Acesso em: 16 julho 2013.
- FRIJDA, N. H. *The emotions*. Cambridge: Cambridge University Press, 1986.
- GOBL, C.; NÍ CHASAIDE, A. The role of voice quality in communicating emotion, mood and attitude, In: *Speech Communication*, v. 40, 2003. p. 189-212.
- GREEN, D. M.; SWETS, J. A. *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.
- GRIFFITHS, P. E. *What emotions really are the problem of psychological categories*. Chicago: The University of Chicago Press, 1997.
- GROSS, J. J., & BARRETT, L. F. Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review*, v. 3, n. 1, p.8-16, 2011.
- HAMMERSCHMIDT, K.; JÜRGENS, U. Acoustical correlates of affective prosody. *Journal of Voice*, v.21, p. 531-540, 2007.

HARRÉ, R. *The social construction of emotions*. New York, NY: Blackwell, 1986.

HARRÉ, R.; FINLAY-JONES, R. Emotion talk across time. In: R. HARRÉ (ed.), *The social construction of emotions* Oxford, England: Basil Blackwell, 1986. p. 220-233.

HILLENBRAND, J.; GAYVERT, R. T. Identification of steady-state vowels synthesized from the Peterson and Barney measurements. *Journal of Acoustic Society of America*, v. 94, p. 668-674, 1993.

HIRST, D.; DI CRISTO, A. (Orgs.) *Intonation Systems*. Cambridge: Cambridge University Press, 1998.

HIRST, D.; ESPESSER, R. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix-em-Provence*, v. 15, p. 71-85, 1993.

HOLLIEN, H. Vocal Indicators of Psychological Stress. *Annals of the New York Academy of Sciences*, v. 347, p. 47-72, 1980.

HOWELL, S. (1981). Rules not words. In: HEELAS, P.; LOCK, A. (eds), *Indigenous psychologies: The anthropology of the self*. San Diego, CA: Academic Press, 1981. p. 133-143.

HUCKVALE, M. A.; BROOKES, D. M.; DWORKIN, L. T.; JOHNSON, M. E.; PEARCE, D. J.; WHITAKER, L. *The SPAR Speech Filing System*. In: EUROPEAN CONFERENCE ON SPEECH TECHNOLOGY, 1987, Edinburgh. Disponível em: <http://www.phon.ucl.ac.uk/home/mark/papers/sparsfs87.pdf>>. Acesso em: 15 set. 2012.

HULIN, W. S.; KATZ, D. The Frois-Wittmann pictures of facial expression *Journal of Experimental Psychology*, v. 18, n. 4, p. 482-498, 1935.

- JAMES, W. What is an emotion? *Mind*, 9, p. 188-205, 1884
- JAMES, W. The Principles of Psychology. Vol. 2. New York: Henry Holt & Company, 1890.
- JOHNSTONE, T.; SCHERER, K. R. The effects of emotions on voice quality. *Proceedings of the 14th International Conference of Phonetic Sciences*. San Francisco, USA. p. 2029-2032, 1999.
- JOOS, M. Acoustic Phonetics. *Language*, v. 24, n. 2, p. 5-136, 1948. Disponível em <http://www.jstor.org/stable/522229> doi:1. Acesso em 11 maio 2015.
- KEATING, P.; GARELLEK, M.; KREIMAN J. Acoustic properties of different kinds of creaky voice. In: *XVIII International Congress of Phonetic Sciences*. Glasgow. Proceedings of the 18th International Congress of Phonetic Sciences, 2015.
- KING, P. *Emotions*. Oxford Handbooks Online. Oxford: Oxford University Press, 2012. Disponível em: <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195326093.001.0001/oxfordhb-9780195326093-e-17>. Acesso em: 5 nov. 2013.
- KREIMAN, J.; GERRATT, B R. Perception of aperiodicity in pathological voice. *Journal of the Acoustical Society of America*, v. 117, n. 4, p. 2201-2211, 2005.
- KREIMAN, J.; SIDTIS, D. *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Wiley-Blackwell, 2011.
- LADD, D. R.; SILVERMAN, K. E. A.; TOLKMITT, F., BERGAMNN, G.; SCHERER, K. R. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, v. 78, p. 435-444, 1985.

- LANGE, C. G. The mechanism of the emotions. Translated by Benjamin Rand. *Classics in the History of Psychology*, 1885. Disponível em <http://psychclassics.yorku.ca/Lange/>. Acesso em 18 jun 2012.
- LASSAK, A. A percepção de variação em semitons descendentes em palavras isoladas no português brasileiro. In: FERREIRA NETTO, W. (Org.). *ExProsodia*. 1. ed. São Paulo: Paulistana, 2016. p. 24-25.
- LAUKKA, P.; JUSLIN, P. N.; BRESIN, R. A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, v. 19, p. 633–654, 2005
- LAVÉ, J. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press, 1980.
- LAZARUS, R. S. *Emotion and adaptation*. New York: Oxford University Press, 1991.
- LEDOUX, J. The emotional brain. New York: Simon & Shuster, 1996.
- LESHOWITZ, B. Measurement of the two-click threshold. *Journal of the Acoustical Society of America*, v. 49, p. 426-466, 1971.
- LEVENSON, R. W. Autonomic nervous system differences among emotions. *Psychological Science*, v. 3, p. 23-27, 1992.
- LIEBERMAN, P.; MICHAELS, S. B. Some aspects of fundamental frequency and envelope amplitude as emotional content of speech. *Journal of the Acoustical Society of America* related to the, v. 34, p. 922–927, 1962.
- LLISTERRI, J.; MACHUCA, M. J.; MOTA, C. de la; RIERA, M.; RÍOS, A. The perception of lexical stress in Spanish. In: SOLÉ, M. J.; RECASENS, D.; ROMERO, J. (Eds.). *Proceedings of the 15th International Congress of Phonetic Sciences*, p. 2023-2026, 2003.
- LUNDQVIST, D., FLYKT, A.; ÖHMAN, A. The Karolinska Directed Emotional Faces – KDEF. CD ROM from Department of Clinical

Neuroscience, Psychology section, Karolinska Institutet, 1998. ISBN 91-630-7164-9.

LUNDQVIST, D.; LITTON, J. E. The Averaged Karolinska Directed Emotional Faces - AKDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998. ISBN 91-630-7164-9.

LUTZ, C. The domain of emotion words in Ifaluk. *American Ethnologist*, v. 9, p. 113-128, 1982.

MACMILLAN, N. A.; CREELMAN, C. D. Detection Theory: a user's guide. 2. ed. New Jersey: Lawrence Erlbaum Associates, 2005.

MAMPE, B.; FRIEDERICI, A. D.; CHRISTOPHE, A.; WERMKE, K. Newborns' Cry Melody Is Shaped by Their Native Language. *Current biology*. Vol. 19 issue 23. p. 1994-1997, 2009.

MATLIN, M. W. *Psicologia Cognitiva*. 5. ed. Rio de Janeiro: LTC, 2003.

MOZZICONACCI, S. J. L. Speech variability and emotion: production and perception. Ph.D. thesis, Eindhoven, the Netherlands, 1998

MOZZICONACCI, S. J. L. The expression of emotion considered in the framework of an intonation model. Keynote paper for ITRW 2000 *Speech and Emotion: A conceptual framework for research*. Newcastle, Northern Ireland, p. 45-52, 2000.

MUNRO, M. J.; DERWING, T. M.; BURGESS, C. S. The detection of foreign accent in backwards speech. *Proceedings of the 15th International Congress of Phonetic Sciences* Barcelona. p. 535-538, 2003

MUNRO, M. J.; DERWING, T. M.; BURGESS, C. S. Detection of nonnative speaker status from content-masked speech. *Speech Communication*, v. 52, n. 7-8, p. 626-637, 2010.

MURRAY, I. R.; ARNOTT, J. L. Toward to simulation of emotion in synthetic speech: a review of the literature on human vocal emotion.

- JASA*, v. 93, n. 2, p. 1097-1108, 1993.
- NORDENBERG, M.; SUNDBERG, J. Effect on LTAS of vocal loudness variation. *TMH-Quarterly Progress and Status Report, Royal Institute of Technology*. v. 45, n.1, p. 87-91, 2003.
- NYGAARD, L. C.; QUEEN, J. S. Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, v. 34, p. 1017-1030, 2008.
- OHALA, J. J.; GILBERT, J. B. Listener's ability to identify languages from their prosody. Report of the Phonology Laboratory, Berkeley, v. 2, p. 126-132, 1978.
- ORLIKOFF, R. F.; BAKEN, R. J. Fundamental frequency modulation of the human voice by the heartbeat: Preliminary results and possible mechanisms, *Journal of the Acoustical Society of America*, v. 85, n. 2, p. 888-893, 1989
- ORTONY, A.; TURNER, T. J. What's is basic about basic emotions? *Psychological Review*, v. 97, n. 3. p. 315-331, 1990.
- OSGOOD, C. E. Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology*, v. 7, p. 1-30, 1966.
- PAKOSZ, M. Intonation and Attitude. *Lingua*, v. 56, p. 153-178, 1982
- PAKOSZ, M. Attitudinal Judgments in Intonation: some evidence for a theory. *Psycholinguistic Research*, v. 12, n. 3, 1983.
- PARAMEI, G. V.; SCHNEIDER, K.; JOSEPHS, I.; SLUSAREK, M. *Identification of emotional meaning in line drawings and faces*. Bochum: Unpublished report of the University of Bochum, 1994.
- PARKINSON, B. Emotions are social. *British Journal of Psychology*, v. 87, p. 663-683, 1996.

PAULMANN, S.; BLEICHNER, M.; KOTZ, S. A. Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, v. 4, article 345, 2013.

PAULMANN, S.; SCHMIDT, P.; PELL, M.; KOTZ, S. A. Rapid processing of emotional and voice information as evidenced by ERPs. *Proceedings of the Speech Prosody 2008*. Campinas, Brasil, p. 205-209, 2009.

PENNER, M. J. Detection of temporal gaps in noise as a measure of the decay of auditory sensation. *Journal of the Acoustical Society of America*, v. 61, p. 552-557, 1977.

PERES, D. O. O papel da prosódia na identificação das variedades regionais do português brasileiro. 2011. Dissertação (Mestrado) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006. Disponível em:

<<http://www.teses.usp.br/teses/disponiveis/8/8142/tde-11042012-154818/es.php>>. Acesso em: 20 julho 2015.

PERES, D. O. The perception of emotion by native and non-native speakers. In: CONSTANTINOU, H.; MAZZARELLA, D.; VARKANITSA, M. (eds.). *First UCL Graduate Conference in Linguistics*. London, 2013. p. 64-65.

PERES, D. O. Perception of emotional speech in Brazilian Portuguese: an intonational and multidimensional approach. *Nouveaux Cahiers de Linguistique Française*, v. 31, p. 153-196, 2014.

PERES, D. O. Intonation as a cue to emotional speech perception: An experiment with normal and delexicalised speech. In: *XVIII International Congress of Phonetic Sciences*. Glasgow. Proceedings of the 18th International Congress of Phonetic Sciences, 2015a.

PERES, D. O. A manifestação da emoção na fala: estudo perceptual com falantes nativos e não nativos. *Estudos Linguísticos*, São Paulo, v. 43, n.

1, p. 10-21, 2015b.

PERES, D. O.; CONSONI, F.; FERREIRA NETTO, W. A influência da cadeia segmental na percepção de variações tonais. *LL Journal*, v. 6, n. 1, 2011.

PLACK, C. J. *The sense of hearing*. Abingdon, Oxon: Psychology Press, 2013.

PLATÃO. *A República*. 9. ed. Tradução de Maria Helena da Rocha Pereira. Lisboa: Fundação Calouste Gulbekian, 2001.

PLUTCHIK, R. *Emotion, a psychoevolutionary synthesis*. New York: Harper & Row, 1980.

PLUTCHIK, R. Emotions: a general psychoevolutionary theory. In: SCHERER, K. R.; EKMAN, P. (Eds.). *Approaches to emotion*. Hillsdale, NJ: Lawrence Erlbaum, 1984. p. 197-219.

PLUTCHIK, R. The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, v. 89, n. 4, p. 344-350, 2001.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 nov. 2012

REINSENZEIN, R. On appraisals as causes of emotions. *Psychological Inquiry*, v. 6, p. 233-237, 1995.

REMEZ, R. E.; RUBIN, P. E.; PISONI, D. B.; CARELL, T. D. Speech perception without traditional speech cues. *Science*, v. 212, p. 947-950, 1981

ROBERTS, L. Acoustics effects of authentic and acted distress on fundamental frequency and vowel quality. Proceedings of The 17th International Congress of Phonetic Sciences, p. 1694-1697, 2011.

ROOT, A. R. The pitch factors in speech – a survey, *Quarterly Journal of*

Speech, v. 16, n. 3, p. 320-343, 1930.

ROSEMAN, I. J.; SMITH, C. A. Appraisal Theory. In: SCHERER, K.; SCHORR, A.; JOHNSTONE, T. (Eds.). *Appraisal processes in emotion: theory, methods, research*. Oxford: Oxford University Press, 2001.

RUSSELL J. A. A circumplex model of affect. *Journal of Personality & Social Psychology*, v. 39, p. 1161-1178, 1980.

RUSSELL, J. A. *Culture and the categorization of emotions*. *Psychological Bulletin*, v. 110, n. 3, p. 426-450, 1991.

SCHERER, K. R. Randomized splicing: a note on a simple technique for masking speech content. *Journal of Experimental Research in Personality*, v. 5, p. 155-159, 1971.

SCHERER, K. R. Speech and emotional states. In: DARBY, J. (ed.) *Speech evaluation in psychiatry*, 1981. p. 189-220.

SCHERER, K. R. Emotion as a multicomponent process: A model and some cross-cultural data. *Review of Personality and Social Psychology*, v. 5, p. 37-63, 1984.

SCHERER, K. R. Appraisal Theory. In: DALGLEISH, T.; POWER, M. (eds.), *Handbook of Cognition and Emotion*. John Wiley & Sons, 1999. p. 638-655.

SCHERER, K. R. A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. *International Conference on Spoken Language Processing Proceedings of ICSLP 2000*, Beijing, China, 2000.

SCHERER, K. R. The affective and pragmatic coding of prosody. In: HUO, Q.; MA, B.; CHNG, E. S.; LI, H. (eds). *ISCSLP 06 Proceedings of the 5th International Conference on Chinese Spoken Language Processing*, 2006. p. 13-14.

- SCHERER, K. R.; BANSE, R.; WALLBOTT, H. G. Inferences from vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, v. 32, p. 76-92, 2001.
- SCHERER, K. R.; BANSE, R.; WALLBOTT, H. G.; GOLDBECK, T. Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, v. 15, p. 123-148, 1991.
- SCHIRMER, A.; ESCOFFIER, N.; ZYSSET S.; KOESTER, D.; STRIANO, T.; FRIEDERICI, A. D. When vocal processing gets emotional: on the role of social orientation in relevance detection by the human amygdala. *Neuroimage*, v. 40, p. 1402-1410, 2008.
- SCHLOSBERG, H. A scale for the judgment of facial expressions. *Journal of Experimental Psychology*, v. 29, p. 497-510, 1941.
- SCHLOSBERG, H. The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology.*, v. 44, p. 229-237, 1952.
- SCHLOSBERG, H. Three dimensions of emotion. *Psychological Review*, v. 61, p. 81-88, 1954.
- SENECA, L. A. *Minor Dialogs Together with the Dialog "On Clemency"*. Translated by Aubrey Stewart, M.A., Late Fellow of Trinity College Cambridge. Bohn's Classical Library Edition; London, George Bell and Sons. Scanned and digitized by Google from a copy maintained by the University of Virginia. p. 48-75, 1900.
- SHUE, Y. L.; CHEN, G.; ALWAN, A. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. *Proceedings of Interspeech Japan*. p. 34-37, 2010.
- SIEMER, M.; MAUSS, I. B.; GROSS, J. J. Same situation — different emotions: How appraisals shape our emotions. *Emotion*, v. 7, n. 3, p. 592-600, 2007.
- SKINNER, E. R. A calibrated recording and analysis of the pitch, force

and quality of vocal tones expressing happiness and sadness; and a determination of the pitch and force of the subjective concepts of ordinary, soft, and loud tones. *Speech Monographs*, v. 2, p. 81-137, 1935.

SMITH, C. A.; KIRBY, L. D. Toward delivering on the promise of appraisal theory. In: SCHERE, K. R.; SCHOOR, A.; JOHNSTONE, T. (eds.), *Appraisal processes in emotion: Theory, methods, research*. Oxford, UK: Oxford University Press, 2001.

SOKOLOV, E. N.; BOUCSEIN, W. A psychophysiological model of emotion space. *Integrative Physiological and Behavioral Science*, v. 35, n. 2, p. 81-119, 2000.

SOLOMON, R. C. The Philosophy of Emotions. In: LEWIS, M.; HAVILAND-JONES, J. M.; BARRET, L. F. *Handbook of Emotions*. 3. ed. New York: The Guilford Press, 2008. p. 3-16.

SONNTAG, G. P.; PORTELE, T. PURR – a method for prosody evaluation and investigation. *Computer Speech and Language, Special Issue on Evaluation*, v. 12, n. 3, p. 437-451, 1998.

STEELE, J. *An essay towards establishing the melody and measure of speech*. Menston, UK: The Scolar Press Limited, 1969 [1775].

STROOP, R. J. Studies on interference in serial verbal reactions. *Journal of Experimental Psychology*, v. 18, n. 6, 1935.

SUNDBERG, J. Speech, song, and Emotions. In: CLYNES, M. (ed.), *Music, mind, and brain*. New York: Springer Science-Business Media, LCC, 1982.

‘t HART, J.; COLLIER, R.; COHEN, A. *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press, 1990.

KING, P. *Emotions*. In: DAVIES, E.; STUMP, E. (eds.), *The Oxford Handbook of Aquinas*. New York: Oxford University Press, 2012. p. 209-226.

TOOBY, J.; COSMIDES, L. The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, 11, 375-424, 1990.

TRNKA, R. How many dimensions does emotional experience have? The theory of multi-dimensional emotional experience. In: TRNKA R.; BALCAR, K.; KUŠKA, M. *Re-constructing emotional spaces: from experience to regulation*. Prague: Prague College of Psychosocial Studies Press, 2011. p. 33-40.

TYLOR, E. B. *Primitive culture: researches into the development of mythology, philosophy, religion, languages, art and customs*. London: John Murray Albermale Street, 1871. vol. 1.

ULDALL, E. Attitudinal meanings conveyed by intonation. *Language and Speech*, 3, p. 223-234, 1960.

VAINIO, M.; SUNI, A.; RAITIO, T.; NURMINEN, J.; JARVIKIVI, J.; ALKU, P. New method for delexicalization and its application to prosodic tagging for text-to-speech synthesis. In: *10th Annual Conference of the International Speech Communication Association - Interspeech*, 2009. p. 1703-1706.

VAN DEN BERG, J. W. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, v. 1, p. 227-244, 1958.

VAN LANCKER, D.; SIDTIS, J. J. The identification of affective-prosodic stimuli by left-and right-brain damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, v. 35, p. 963–970, 1992.

- VASSOLER, A. M. O.; MARTINS, M. V. M. *A entoação em falas teatrais: uma análise da raiva e da fala neutra. Estudos Linguísticos*, v. 42, n. 1, p. 9-18, 2013.
- Wallace, A. E C, & Carson, M. T. (1973). Sharing and diversity in emotion terminology. *Ethos*, 1,1-29.
- WALLBOTT, H. G.; SCHERER, K. R. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, v. 51, n. 4. p. 690-699, 1986.
- WARREN, R. M.; RIENER, K. R.; BASHFORD Jr., J. A.; BRUBAKER, B. S. Spectral redundancy: intelligibility of sentences heard through narrow spectral slits. *Percept Psychophys*, v. 57, p. 175-182, 1995.
- WATSON, D.; CLARK, L. A.; TELLEGEN, A. Development and validation of brief measures of positive and negative affect: the PANAS Scales. *Journal of Personality and Social Psychology*, v. 54, p. 1063-1070, 1988.
- WEAVER, A. T. Experimental studies in vocal expression. *The Quarterly Journal of Speech Education*, v. 10, n. 3, p. 199-204, 1924a.
- WEAVER, A. T. Experimental studies in vocal expression. *The Journal of Applied Psychology*, v. 8, n. 1, p. 23-51, 1924b.
- WEAVER, A. T. Experimental studies in vocal expression. *The Journal of Applied Psychology*, v. 8, n. 2, p. 159-186, 1924c.
- WEBER, F. On Melody in Speech. *Popular Science Monthly*, v. 30, April 1887.
- WERTZNER, H. F.; SCHREIBER, S.; AMARO, L. Análise da frequência fundamental, jitter, shimmer e intensidade vocal em crianças com transtorno fonológico. *Revista Brasileira de Otorrinolaringologia*, São Paulo, v. 71, n. 5, p. 582-588, 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-72992005000500007&lng=en&nrm=iso>. Acesso em: 11 maio 2014.

WILLIAMS, U.; STEVENS, K. N. Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, v. 52, p. 1238-1250, 1972.

WILSON, K. M. Music and Speech. *Music & Letters*, v. 8, n. 3. Oxford University Press, p. 350-365, 1927.

WOODWORTH, R. S. *Experimental psychology*. New York: Henry Holt & Company, p. 250-252, 1938.

WUNDT, W. M. *Outlines of Psychology*. Tradução de Charles Hubbard Judd. Leipzig: W. Engelmann; New York: G.E. Stechert, 1897. p. 342.

XU, Y.; WANG, Q. E. Components of intonation: what are linguistic, what are mechanical/physiological? Presented at International Conference on Voice Physiology and Biomechanics. Evaston, 1997.

ZAJONC, R. B. Feeling and thinking: Preferences need no inferences. *American Psychologist*, v. 35, p. 151-175, 1980.

12. Anexos

***** Experimento piloto sobre percepção de emoção *****

A tarefa dos participantes deste experimento será ouvir os trechos contidos no arquivo de áudio (exp_piloto.mp3) e julgar quais emoções estão sendo manifestas em cada trecho de fala. Cada trecho está separado por um bipe e sua respectiva numeração (de 1 a 32). A marcação de cada emoção percebida será feita na planilha exp.piloto.emocao.xls, como no exemplo abaixo:

	raiva	medo	felicidade	tristeza
track 1				X
track 2			X	
track 3	X			
etc.				

Em caso de dúvida em algum trecho, o participante poderá ouvi-lo novamente.

***** Pilot experiment about perception of emotion *****

The task is to listen to the audio file (exp_piloto.mp3) e judge which emotion best describes each sentence. A beep and its number (from 1 to 32) will play before sentences. After listening and judging the audio, insert a check mark or X into the excel spreadsheet (exp.piloto.emocao.xls). Follow the example:

	anger	fear	happiness	sadness
track 1				X
track 2			X	
track 3	X			

If you have any doubt you can play the audio again.

Experimento com abordagem dimensional da emoção – fala normal

Instructions in English *(instruções em português abaixo)*

The task of this experiment is to listen to sentences and evaluate the degree of pleasantness, agitation and submissiveness through the three slider buttons.

The first slider refers to the degree of pleasantness.
The second slider refers to the degree of agitation.
The third slider refers to the degree of submissiveness.

You have to slide the button according to your judgment.

The audio will start playing automatically as you open the page. If you have any doubt you can play the audio again.
You will have to judge 32 sentences in the experiment, which will not take more than 6 minutes to be performed.

Please, try to find a quiet place to do the experiment and use headphones.

Instruções em português

Neste experimento, a tarefa é ouvir as sentenças e avaliar, por meio de três *botões deslizantes*, o grau de agradabilidade, agitação e submissão das sentenças.

O primeiro botão deslizante refere-se ao grau de agradabilidade.
O segundo botão deslizante refere-se ao grau de agitação.
O terceiro botão deslizante refere-se ao grau de submissão.

Você deve deslizar o botão de acordo com o seu julgamento.
O áudio começará a tocar automaticamente quando você abrir cada uma das páginas. Se houver alguma dúvida, você poderá reproduzir o áudio novamente.
Você terá de julgar 32 sentenças durante o experimento, tarefa que não levará mais do que 6 minutos para ser concluída.

Por favor, procure um lugar silencioso para participar do experimento e use fones de ouvido.

Experimento com abordagem dimensional da emoção – fala delexicalizada

English instructions (*instruções em português abaixo*)

The task of this experiment is to listen to sentences and evaluate through the three slider buttons the degree of pleasantness, agitation and submissiveness.

The first slider refers to the degree of pleasantness.
The second slider refers to the degree of agitation.
The third slider refers to the degree of submissiveness.

You can slide the button according to your judgment.

The sentences were manipulated in order to extract the words of the sentences. Thus, you will only hear the "melody" of the sentences. It sounds like a person speaking behind the wall. Every time you click on "next" button in the bottom of the page, the sentence will start to play automatically. If you have any doubt you can play the audio again. The experiment is made by 32 sentences and does not take more than 6 minutes to be performed.

Please, do the experiment only in quiet places and using headphones.

Instruções em português

Neste experimento, a tarefa é ouvir as sentenças e avaliar, por meio de três *botões deslizantes*, o grau de agradabilidade, agitação e submissão das sentenças.

O primeiro botão deslizante refere-se ao grau de agradabilidade.
O segundo botão deslizante refere-se ao grau de agitação.
O terceiro botão deslizante refere-se ao grau de submissão.

Você pode deslizar o botão de acordo com o seu julgamento.

As sentenças foram manipuladas a fim de extrair as palavras das frases. Assim, você só irá ouvir a "melodia" das frases, como se a pessoa estivesse falando atrás de uma parede.

Toda vez que você clicar no botão "avançar", na parte inferior da página, a sentença começará a tocar automaticamente. Se houver alguma dúvida,

you will be able to reproduce the audio again.

The experiment is formed by 32 sentences and does not take more than 6 minutes to be concluded.

Please, participate in the experiment in quiet places and with use of headphones.

The example 1 illustrates the task / O exemplo 1 ilustra a tarefa:

Experimento com estímulos de curta duração – *thin-slicing*

Instructions

The task of this experiment is to listen to short speech excerpts (from 0.2s to 1.4s) and relate them to the faces on the screen. If you have any doubt you can play the audio again.

The audio will not start playing automatically as you move from one screen to the next. For this reason, you must click on the play button to start the sound.

After the experiment there will be a survey about your experience of performing the experiment. The entire task will not be longer than 7 minutes.

Please, find a quiet place to do the experiment and use headphones.

Instruções em português

Neste experimento, a tarefa é ouvir pequenos trechos de fala (de 0,2s até 1,5s) e relacioná-los às faces que aparecerão na tela. Se você tiver alguma dúvida, você poderá tocar o áudio novamente. O áudio não iniciará automaticamente quando você abrir as páginas. Por isso, você deve clicar no botão de *play* para iniciar o som.

Depois do experimento, haverá uma pesquisa sobre a sua experiência em participar do experimento. Ao todo, a tarefa não durará mais do que 7 minutos.

Por favor, encontre um lugar silencioso para participar do experimento e use fones de ouvido.