



WELLINGTON DA SILVA

**ANÁLISE FONÉTICO-ACÚSTICA DA EXPRESSIVIDADE
DE EMOÇÕES EM DEPOIMENTOS REAIS**

**CAMPINAS,
2014**



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ESTUDOS DA LINGUAGEM**

WELLINGTON DA SILVA

**ANÁLISE FONÉTICO-ACÚSTICA DA EXPRESSIVIDADE DE
EMOÇÕES EM DEPOIMENTOS REAIS**

**Dissertação de mestrado apresentada ao
Instituto de Estudos da Linguagem da
Universidade Estadual de Campinas para
obtenção do título de Mestre em
Linguística.**

Orientador: Prof. Dr. Plínio Almeida Barbosa

**CAMPINAS,
2014**

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Estudos da Linguagem
Crislene Queiroz Custódio - CRB 8/8624

Si38a Silva, Wellington da, 1989-
Análise fonético-acústica da expressividade de emoções em depoimentos reais / Wellington da Silva. – Campinas, SP : [s.n.], 2014.

Orientador: Plínio Almeida Barbosa.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Ciência cognitiva. 2. Emoções. 3. Fonética. 4. Percepção da fala. 5. Comunicação intercultural. I. Barbosa, Plínio Almeida, 1966-. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Acoustic-phonetic analysis of the expression of emotions in real life speech

Palavras-chave em inglês:

Cognitive science

Emotions

Phonetics

Speech perception

Intercultural communication

Área de concentração: Linguística

Titulação: Mestre em Linguística

Banca examinadora:

Plínio Almeida Barbosa [Orientador]

Emílio Gozze Pagotto

Sandra Madureira Fontes

Data de defesa: 04-07-2014

Programa de Pós-Graduação: Linguística

BANCA EXAMINADORA:

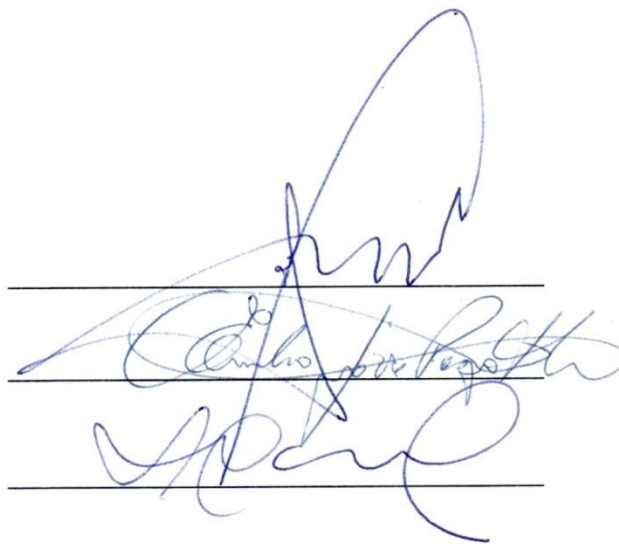
Plínio Almeida Barbosa

Emílio Gozze Pagotto

Sandra Madureira Fontes

Maria Bernadete Marques Abaurre

Pablo Arantes

Handwritten signatures in blue ink over horizontal lines. The first signature is for Plínio Almeida Barbosa, the second for Emílio Gozze Pagotto, and the third for Sandra Madureira Fontes. The signatures are stylized and cursive.

IEL/UNICAMP
2014

À minha mãe, Mariza, por toda ajuda e incentivo.

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por financiar esta pesquisa por meio das bolsas de Mestrado (processo 2012/04254-4) e de Estágio de Pesquisa no Exterior (processo 2013/06082-9). As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do(s) autor(es) e não necessariamente refletem a visão da FAPESP;

Ao meu orientador Prof. Dr. Plínio Almeida Barbosa, por me guiar dedicadamente nesta jornada, pelos valiosos ensinamentos e por me ajudar nos momentos difíceis;

Aos membros das bancas de Defesa e de Qualificação, por aceitarem o convite e contribuir para esta pesquisa;

À Profa. Dra. Sandra Madureira Fontes, por nos indicar o filme documentário “Jogo de Cena”;

A Lilian e Alf, por me hospedarem em Gotemburgo e por fazerem com que essa viagem seja inesquecível para mim;

A Renata Passetti, por me apresentar a Lilian e Alf;

À Profa. Dra. Åsa Abelin, por me orientar na Universidade de Gotemburgo e contribuir para esta pesquisa;

Ao Prof. Dr. Anders Eriksson, por também me ajudar no estágio de pesquisa na Universidade de Gotemburgo;

A Natália Durigon e Beatriz Salgado, por ajudarem no recrutamento de sujeitos para os experimentos de percepção;

A todas as pessoas que participaram dos experimentos de percepção desta pesquisa;

Aos colegas do GRUPO DE ESTUDOS DE PROSÓDIA DA FALA, pelo convívio e pela troca de conhecimentos;

Aos professores do Instituto de Estudos da Linguagem, pela dedicação e conhecimento passado aos alunos;

Aos funcionários da UNICAMP, por garantir o funcionamento da universidade;

A Deus, pela minha vida e por todas as oportunidades.

*Is it not monstrous that this player here,
But in a fiction, in a dream of passion,
Could force his soul so to his own conceit
That from her working all his visage wan'd;
Tears in his eyes, distraction in's aspect,
A broken voice, and his whole function suiting
With forms to his conceit? And all for nothing!
For Hecuba?
What's Hecuba to him, or he to Hecuba,
That he should weep for her? What would he do,
Had he the motive and the cue for passion
That I have? He would drown the stage with tears
And cleave the general ear with horrid speech;
Make mad the guilty, and appal the free;
Confound the ignorant, and amaze, indeed,
The very faculties of eyes and ears.*

(William Shakespeare, *Hamlet*)

RESUMO

Esta dissertação apresenta um estudo intercultural da percepção de emoções expressas pela fala em situações comunicativas reais conduzido com ouvintes brasileiros e suecos. A literatura sobre o tema mostra que sujeitos ouvintes são capazes de reconhecer emoções expressas na fala em experimentos de percepção com um percentual de sucesso muito superior ao que seria obtido ao acaso, mesmo quando avaliando estímulos em uma língua estrangeira e de uma cultura diferente. Os principais objetivos do estudo apresentado foram investigar quais parâmetros fonético-acústicos dentre os extraídos são afetados pelo estado emocional dos falantes, em quais deles os sujeitos ouvintes se guiam para julgar o grau de expressão dessas emoções nos experimentos de percepção, investigar se os sujeitos realizam melhor essa tarefa para dimensões emocionais do que para emoções discretas, se a cultura e a língua materna dos sujeitos influenciam na percepção das emoções e se os sujeitos brasileiros e suecos se guiam pelos mesmos parâmetros acústicos para fazer esse julgamento. Os corpora utilizados consistem de enunciados de mulheres falantes do português brasileiro e do sueco extraídos de um documentário e de programas de televisão e de rádio. Os enunciados passaram por uma análise acústica, na qual as seguintes classes de parâmetros acústicos foram medidas automaticamente com um *script* para o programa PRAAT: frequência fundamental e sua primeira derivada, intensidade global, inclinação espectral e espectro médio de longo-termo (LTAS). Dessas classes foram calculados vários descritores estatísticos, totalizando doze parâmetros acústicos. Conduzimos dois experimentos de percepção com sujeitos brasileiros e suecos. No primeiro deles, os sujeitos avaliaram, em escalas graduadas de 0 a 4, o grau de expressão de emoções discretas descritas por oito adjetivos (alegre, comovido, surpreso, triste, contente, angustiado, aflito e entusiasmado) para os enunciados em português brasileiro. No segundo experimento, os sujeitos julgaram o grau de expressão de cinco dimensões emocionais (ativação, justiça, valência, motivação e envolvimento) para os enunciados em sueco. As respostas dos sujeitos nos experimentos foram submetidas a análises de PCA. Para o experimento I, essas análises revelaram que as emoções descritas pelos oito adjetivos foram avaliadas pelos sujeitos de ambas as nacionalidades conjuntamente por duas grandes dimensões

emocionais: felicidade e neutralidade. Para o experimento II, os dois componentes da PCA revelaram que os sujeitos de ambas as nacionalidades julgaram as cinco dimensões emocionais distinguindo entre um estado de “calma” de um estado de maior agitação emocional. Os componentes da PCA foram correlacionados com os parâmetros acústicos por meio de análises de regressões lineares, que mostraram que os sujeitos de ambas as nacionalidades se guiaram pelos mesmos parâmetros acústicos para realizar os julgamentos nos dois experimentos. Esses parâmetros também foram muito robustos em classificar os trechos de fala nas análises de LDA realizadas. Os parâmetros que se mostraram mais relevantes foram o LTAS, a mediana e a semiamplitude entre quartis da frequência fundamental e a média e o desvio-padrão da inclinação espectral. É possível concluir que a cultura e a experiência emocional dos sujeitos brasileiros e suecos não influenciaram na percepção das emoções expressas nos dois corpora.

Palavras-chave: Ciência cognitiva, Emoções, Fonética, Percepção da fala, Comunicação intercultural.

ABSTRACT

This dissertation presents a cross-cultural study on the perception of real emotions expressed in speech conducted with Swedish and Brazilian listeners. The literature on the subject shows that listeners are capable of recognizing emotions expressed in speech in perception experiments with much higher percentage of success than that expected by chance, even when evaluating stimuli in a foreign language and of a different culture. The main objectives of the present study were to investigate which acoustic-phonetic parameters among those extracted are affected by the emotional state of the speakers, on which of them the listeners rely to judge the degree of expression of these emotions in the perception experiments, to investigate whether the subjects perform better in this task when evaluating emotional dimensions rather than discrete emotions, whether the listeners' culture and mother language affect the perception of the emotions and whether the Brazilian and the Swedish subjects rely on the same acoustic parameters to make the judgements. The corpora used consist of utterances of female speakers of Brazilian Portuguese and Swedish extracted from a documentary film as well as from radio and TV programs. The following acoustic parameters were measured automatically for these utterances with a script for the software PRAAT: fundamental frequency and its first derivative, global intensity, spectral tilt and Long-Term Average Spectrum (LTAS). A total of twelve statistical descriptors were computed for these parameters. We conducted two perception experiments with Brazilian and Swedish subjects. In the first one they evaluated the degree of expression of the discrete emotions described by eight adjectives (joyful, moved, surprised, sad, contented, anguished, distressed and enthusiastic) for the utterances in Brazilian Portuguese on scales ranging from 0 to 4. In the second experiment the subjects rated the degree of expression of five emotional dimensions (activation, fairness, valence, motivation and involvement) for the utterances in Swedish. The responses of the subjects in the experiments were analyzed by means of PCA. For experiment I, the PCA revealed that the listeners of both nationalities evaluated the emotions described by the eight adjectives jointly by means of two major emotional dimensions: happiness and neutrality. For experiment II, the two PCA components revealed that the listeners of both nationalities

evaluated the five emotional dimensions distinguishing between a state of “calmness” and a state of higher emotional agitation. The PCA components were correlated with the acoustic parameters by means of linear regression analyses. These analyses showed that the subjects of both nationalities relied on the same acoustic parameters to make the judgements in both experiments. These acoustic parameters were also very robust in classifying the utterances in the LDA performed. The parameters which proved more relevant were the LTAS, fundamental frequency median and semi-interquartile range and spectral tilt mean and standard deviation. It is possible to conclude that the culture and the emotional experience of the Brazilian and the Swedish subjects did not affect the perception of the emotions expressed in both corpora.

Keywords: Cognitive science, Emotions, Phonetics, Speech perception, Intercultural communication.

LISTA DE FIGURAS

Figura 1.1	Distribuição de algumas emoções de acordo com três dimensões emocionais (ativação, valência e potência) de Paescke (2003).	21
Figura 2.1	Ciclo de uma senoide (delimitado pelas barras verticais).	35
Figura 2.2	Forma de onda da vogal [a] da primeira sílaba da palavra “passado”, pronunciada por um falante masculino do português brasileiro.	36
Figura 2.3	Espectro FFT para a vogal [a] da primeira sílaba da palavra “passado” pronunciada por um falante masculino do português brasileiro.	36
Figura 2.4	Espectrograma (eixo vertical: frequência em hertz; eixo horizontal: tempo em segundos) para o trecho “(n)unca andava de” retirado da sentença “Manuel nunca andava de olhos baixos”, pronunciada por uma falante feminina do português brasileiro.	37
Figura 2.5	Espectro FFT das ondas sonoras produzidas pela fonte (a vibração das pregas vocais), para uma f_0 igual a 100 Hz.	39
Figura 2.6	Trecho da forma de onda da vogal [a] de uma falante feminina do português brasileiro de nosso corpus “Jogo de Cena” expressando tristeza.	42
Figura 2.7	Trecho da forma de onda da vogal [a] da mesma falante da figura 2.6 expressando alegria.	43
Figura 2.8	Dois espectros médios de longo-termo de uma falante feminina adulta de nosso corpus sueco.	46
Figura 3.1	Tela do experimento de percepção aplicado por meio da plataforma online “SurveyGizmo”.	58
Figura 3.2	Gráfico de dispersão com os adjetivos do experimento I com sujeitos brasileiros, projetados de acordo com os <i>loadings</i> para o PC1 e para o PC2.	63
Figura 3.3	Regressão linear simples com o parâmetro “inclinação do LTAS” como variável independente e o primeiro fator da PCA como variável dependente.	69
Figura 3.4	Regressão linear simples com o parâmetro “média da inclinação espectral” como variável independente e o primeiro fator da PCA como variável dependente.	69
Figura 3.5	Regressão linear simples com o parâmetro “assimetria de f_0 ” como variável independente e o primeiro fator da PCA como variável dependente.	70
Figura 3.6	Regressão linear simples com o parâmetro “mediana de f_0 ” como variável independente e o segundo fator da PCA como variável dependente.	71

Figura 3.7	Regressão linear simples com o parâmetro “assimetria da inclinação espectral” como variável independente e o segundo fator da PCA como variável dependente.	71
Figura 3.8	Gráfico de dispersão com os adjetivos do experimento I com sujeitos suecos, plotados de acordo com os loadings para o PC1 e para o PC2.	77
Figura 4.1	Página do experimento II na versão apresentada aos sujeitos ouvintes suecos.	91
Figura 4.2	Gráfico de dispersão com as dimensões emocionais avaliadas pelos sujeitos ouvintes suecos no experimento II plotadas de acordo com os <i>loadings</i> para o primeiro fator da PCA (PRONTIDÃO PARA AÇÃO) e para o segundo (INSATISFAÇÃO).	94
Figura 4.3	Regressão linear simples com o parâmetro “inclinação do LTAS” como variável independente e o primeiro fator da PCA com sujeitos ouvintes suecos (PRONTIDÃO PARA AÇÃO) como variável dependente.	100
Figura 4.4	Regressão linear simples com o parâmetro “mediana da frequência fundamental” como variável independente e o primeiro fator da PCA com sujeitos ouvintes suecos (PRONTIDÃO PARA AÇÃO) como variável dependente.	100
Figura 4.5	Gráfico de dispersão com as cinco dimensões emocionais avaliadas no experimento II por sujeitos ouvintes brasileiros, projetadas de acordo com os loadings para o primeiro e o segundo componente da PCA (CALMA e SATISFAÇÃO, respectivamente).	106
Figura 4.6	Regressão linear simples com o parâmetro “inclinação do LTAS” como variável independente e o primeiro fator da PCA com sujeitos ouvintes brasileiros (CALMA) como variável dependente.	112
Figura 4.7	Regressão linear simples com o parâmetro “mediana de f_0 ” como variável independente e o primeiro fator da PCA com sujeitos ouvintes brasileiros (CALMA) como variável dependente.	113
Figura 4.8	Regressão linear simples com o parâmetro “média da inclinação espectral” como variável independente e o primeiro fator da PCA com sujeitos ouvintes brasileiros (CALMA) como variável dependente.	113

LISTA DE TABELAS

Tabela 2.1	Resumo de alterações em alguns parâmetros acústicos para seis emoções. Elaborada com base em Johnstone e Scherer (2000) e Scherer (2003).	48
Tabela 3.1	Adjetivos usados no experimento I com sujeitos brasileiros.	60
Tabela 3.2	Valores de <i>kappa</i> e <i>z</i> para as oito etiquetas avaliadas pelos sujeitos brasileiros no experimento de percepção I.	61
Tabela 3.3	<i>Loadings</i> dos adjetivos do experimento I com sujeitos ouvintes brasileiros para o fator 1 (PC1) e para o fator 2 (PC2).	63
Tabela 3.4	<i>Loadings</i> dos adjetivos do experimento I com sujeitos ouvintes brasileiros, calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 (PC1) e para o fator 2 (PC2).	65
Tabela 3.5	<i>Scores</i> dos trinta trechos de fala do corpus “Jogo de Cena” calculados separadamente por sexo para o primeiro fator (PC1) e para o segundo (PC2) a partir das respostas do experimento I com sujeitos ouvintes brasileiros.	65
Tabela 3.6	Coefficientes de determinação (R^2 ajustado) e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes brasileiros para o experimento I.	68
Tabela 3.7	Significância estatística dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.1.	68
Tabela 3.8	Significância estatística dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.2.	70
Tabela 3.9	Etiquetas atribuídas a cada trecho de fala do corpus “Jogo de Cena” para o primeiro fator da PCA (PC1) e para o segundo (PC2), de acordo com os <i>scores</i> e também com a LDA.	73
Tabela 3.10	Adjetivos usados no experimento I com sujeitos suecos.	75

Tabela 3.11	Valores de <i>kappa</i> e <i>z</i> para as oito etiquetas avaliadas pelos sujeitos suecos no experimento de percepção I.	75
Tabela 3.12	<i>Loadings</i> dos adjetivos do experimento I com sujeitos ouvintes suecos para o fator 1 (PC1) e para o fator 2 (PC2).	77
Tabela 3.13	<i>Loadings</i> dos adjetivos do experimento I com sujeitos ouvintes suecos, calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 (PC1) e para o fator 2 (PC2).	79
Tabela 3.14	<i>Scores</i> dos trinta trechos de fala do corpus “Jogo de Cena” calculados separadamente por sexo para o primeiro fator (PC1) e para o segundo (PC2) a partir das respostas do experimento I com sujeitos ouvintes suecos.	79
Tabela 3.15	Coefficientes de determinação e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes suecos para o experimento I.	82
Tabela 3.16	Significância estatística dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.3.	82
Tabela 3.17	Significância estatística dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.4.	83
Tabela 4.1	As cinco partes do experimento II. Versão em sueco à esquerda e em português à direita.	90
Tabela 4.2	Valores de <i>kappa</i> e significância das cinco dimensões emocionais avaliadas por sujeitos ouvintes suecos no experimento II.	93
Tabela 4.3	<i>Loadings</i> das cinco dimensões avaliadas pelos sujeitos ouvintes suecos no experimento II para o primeiro fator da PCA e para o segundo.	93
Tabela 4.4	<i>Loadings</i> das dimensões emocionais avaliadas no experimento II por sujeitos ouvintes suecos, calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 e para o fator 2 da PCA.	96
Tabela 4.5	<i>Scores</i> (separados por sexo) para o primeiro e segundo componentes atribuídos pela PCA aos 40 trechos de fala do corpus “SweEmo” avaliados por sujeitos ouvintes suecos no experimento II.	96

Tabela 4.6	Coeficientes de determinação e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes suecos para o experimento II.	99
Tabela 4.7	Significância estatística dos coeficientes de interceptação e inclinação da equação 4.1 para $\alpha = 10\%$.	99
Tabela 4.8	Etiquetas atribuídas a cada trecho de fala do corpus “SweEmo” para o primeiro fator da PCA e para o segundo, de acordo com os <i>scores</i> e também com a LDA.	102
Tabela 4.9	Valores de <i>kappa</i> e significância das cinco dimensões emocionais avaliadas por sujeitos ouvintes brasileiros no experimento II.	104
Tabela 4.10	<i>Loadings</i> das cinco dimensões avaliadas pelos sujeitos ouvintes brasileiros no experimento II para o primeiro fator da PCA e para o segundo.	105
Tabela 4.11	<i>Loadings</i> das dimensões emocionais avaliadas no experimento II por sujeitos ouvintes brasileiros calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 e para o fator 2 da PCA.	108
Tabela 4.12	<i>Scores</i> atribuídos pela PCA aos 40 trechos de fala do corpus “SweEmo” analisados separadamente por sexo para o primeiro fator e para o segundo a partir das respostas dos sujeitos ouvintes brasileiros no experimento II.	108
Tabela 4.13	Coeficientes de determinação e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes brasileiros para o experimento II.	111
Tabela 4.14	Significância estatística dos coeficientes de interceptação e inclinação da equação 4.3 para $\alpha = 10\%$.	111

LISTA DE ABREVIATURAS E SIGLAS

Parâmetros Acústicos

assimf0 - assimetria da frequência fundamental

assiminclinespec - assimetria da inclinação espectral

assimint - assimetria da intensidade

assimdf0div10 - assimetria da primeira derivada da frequência fundamental

desvpadinclinespec - desvio-padrão da inclinação espectral

desvpaddf0 - desvio-padrão da primeira derivada da frequência fundamental

f_0 – frequência fundamental

LTAS - *Long-Term Average Spectrum* (Espectro Médio de Longo-Termo)

mednf0 - mediana da frequência fundamental

medinclinespec - média da inclinação espectral

medderivf0 - média da primeira derivada da frequência fundamental

quan995f0 - quantil 99,5% da frequência fundamental

sampquartisf0 - semi-amplitude entre quartis da frequência fundamental

slopeLTAS - inclinação do espectro médio de longo-termo

Análises Estatísticas

LDA – *Linear Discriminant Analysis* (Análise Discriminante Linear)

PCA – *Principal Component Analysis* (Análise de Componentes Principais)

SUMÁRIO

Introdução.....	1
PARTE I: REVISÃO DA LITERATURA.....	5
Capítulo 1 - As emoções: Abordagens teóricas e aspectos metodológicos.....	7
1.1 O que são as emoções?.....	9
1.2 A perspectiva darwinista das emoções.....	11
1.3 A teoria de William James.....	13
1.4 A perspectiva cognitivista.....	15
1.4.1 O modelo de processo componencial de Klaus Scherer.....	17
1.5 A visão dimensional das emoções.....	18
1.6 O papel da cultura na expressão e no reconhecimento de emoções.....	23
1.7 Emoções reais vs. Emoções atuadas.....	28
Capítulo 2 - Expressão de emoções na fala: Conceitos básicos.....	31
2.1 O mecanismo de produção da fala.....	32
2.2 Os aspectos físicos da fala.....	34
2.3 Correlatos fonético-acústicos das emoções.....	40
2.4 Características fonético-acústicas de algumas emoções.....	46
2.5 O script “Expression Evaluator”.....	48
PARTE II: METODOLOGIA, ANÁLISES E RESULTADOS.....	51
Capítulo 3 - Experimento de percepção I.....	55
3.1 O corpus “Jogo de Cena”.....	55
3.2 Metodologia.....	56
3.3 Análise com sujeitos brasileiros.....	59
3.3.1 Concordância entre os sujeitos.....	60
3.3.2 Identificação das variáveis correlacionadas entre si.....	61
3.3.3 Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos....	66
3.3.4 Classificação automática dos trechos de fala.....	72

3.4	Análise com sujeitos suecos.....	74
3.4.1	Concordância entre os sujeitos.....	74
3.4.2	Identificação das variáveis correlacionadas entre si.....	76
3.4.3	Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos....	80
3.4.4	Classificação automática dos trechos de fala.....	83
3.5	Discussão.....	84
Capítulo 4 - Experimento de percepção II.....		87
4.1	O corpus “SweEmo”.....	87
4.2	Metodologia.....	88
4.3	Análise com sujeitos suecos.....	91
4.3.1	Concordância entre os sujeitos.....	92
4.3.2	Identificação das variáveis correlacionadas entre si.....	92
4.3.3	Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos....	97
4.3.4	Classificação automática dos trechos de fala.....	101
4.4	Análise com sujeitos brasileiros.....	103
4.4.1	Concordância entre os sujeitos.....	103
4.4.2	Identificação das variáveis correlacionadas entre si.....	104
4.4.3	Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos....	109
4.4.4	Classificação automática dos trechos de fala.....	114
4.5	Discussão.....	114
Conclusão.....		117
Referências.....		121
ANEXOS.....		129
Anexo I – Script “Expression Evaluator”.....		131
Anexo II – Transcrição dos enunciados do corpus “Jogo de Cena”.....		135
Anexo III – Instruções para o experimento I com sujeitos brasileiros.....		137
Anexo IV - Instruções para o experimento I com sujeitos suecos.....		139
Anexo V - Instruções para o experimento II com sujeitos brasileiros.....		141
Anexo VI - Instruções para o experimento II com sujeitos suecos.....		143

Introdução

Esta dissertação apresenta uma pesquisa experimental cujo tema é a expressão de emoções na fala. Por meio dela, buscamos investigar como a fala é afetada pelas emoções, através de análises acústicas, e como os sujeitos percebem essas alterações em experimentos de percepção. As etapas de desenvolvimento da pesquisa e de interpretação dos resultados foram realizadas à luz da perspectiva cognitivista, que atualmente é a perspectiva teórica sobre emoções mais influente devido, em parte, ao seu grande poder explicativo (Cornelius, 2000). É importante ressaltar que nossa intenção foi estudar como os sujeitos percebem a expressão das emoções e não se eles são capazes de dizer se os falantes estão realmente experienciando a emoção expressa, o que demandaria métodos de investigação diferentes dos utilizados nesta pesquisa.

Além de transmitir informação linguística, o sinal de fala (*i.e.*, o enunciado) também transmite as informações indexicais do falante (sexo, idade, dialeto, entre outras) e informações sobre seu estado afetivo (Laver e Trudgill, 1979). A influência do estado emocional do falante na sua voz ocorre porque as emoções causam, entre outras reações, respostas fisiológicas no indivíduo, tais como alterações no sistema nervoso somático e no sistema nervoso autônomo, que são responsáveis pelo controle voluntário dos músculos e pelo controle da respiração e da circulação sanguínea, respectivamente. Essas alterações, por sua vez, ocasionam variações na respiração, fonação e articulação, processos diretamente relacionados à fala (Scherer, 1986). Além disso, as emoções também podem afetar a atenção e a cognição do falante, o que acaba por dificultar o planejamento da fala e a prosódia (no caso do medo ou da ansiedade, por exemplo) ou facilitá-lo (no caso da alegria ou felicidade, por exemplo) (Johnstone e Scherer, 2000).

Embora a expressão de emoções pela fala seja um fato reconhecido desde a antiguidade pelos retóricos gregos e romanos, que enfatizavam o papel de uma fala expressiva na persuasão e na influência social (Pittam e Scherer, 1993), até algumas décadas atrás a maioria dos pesquisadores das emoções humanas se dedicava somente ao estudo da expressão das emoções na face. Foi somente no século XX, com o

desenvolvimento de meios de comunicação como o rádio e o telefone e de sistemas de síntese de fala e de reconhecimento por voz, que o interesse desses pesquisadores se voltou também para a expressão vocal das emoções (Scherer, 2003) e com isso começaram a ser desenvolvidos estudos experimentais sobre o tema e hipóteses acerca das alterações ocorridas na fala sob a influência das emoções começaram a ser formuladas (Kaiser, 1962; Lieberman e Michaels, 1962; Scherer, 1986). Para investigar essas alterações por meio de análise acústica, os foneticistas e estudiosos da fala em geral usam medidas fonético-acústicas das propriedades físicas da fala (intensidade, duração, frequência fundamental e outras), às quais nos referimos neste trabalho por “parâmetros acústicos”. Um grande número de pesquisadores, mais especificamente aqueles adeptos à teoria evolucionista das emoções, postulam a existência de um conjunto pequeno de emoções universais e muito diferentes entre si, com expressões faciais e alterações fisiológicas específicas a cada uma delas e que recebem rótulos das línguas naturais, como alegria, tristeza, medo, nojo, surpresa, entre outros (ver Ekman, 1992; Fontes, 2014). Seguindo essa linha de raciocínio, a maioria dos pesquisadores que estudam a expressão das emoções na fala assume que as emoções também possuem padrões específicos de mudanças nos parâmetros acústicos e busca então descobrir quais dessas modificações caracterizam cada uma delas. Os parâmetros mais estudados na literatura têm sido aqueles relativos à vibração das pregas vocais (frequência fundamental), ao tempo (taxa de elocução e duração de enunciados e pausas) e à intensidade (quantidade de energia no sinal de fala, responsável pela sensação de volume). Scherer (2003) faz uma crítica a essa limitada variedade de parâmetros estudados e enfatiza a necessidade de se investigar também parâmetros relativos à distribuição da energia no espectro de frequência e às frequências de ressonância do trato vocal (formantes).

Apesar do grande empenho, os pesquisadores ainda não conseguiram diferenciar as emoções satisfatoriamente através de parâmetros acústicos da fala (Johnstone e Scherer, 2000; Xu, Kelly e Smillie, 2013). Por exemplo, no resumo dos padrões de alteração em alguns parâmetros acústicos para seis emoções que Scherer (2003:233) apresenta, tanto a raiva (*anger/rage*) como a alegria (*joy/elation*) apresentam aumento nos valores para a frequência fundamental, intensidade, energia nas altas frequências e taxa de

elocução. De acordo com esse resumo, o único parâmetro que seria capaz de discriminar essas duas emoções seria a curva da frequência fundamental. Porém, esse parâmetro não está especificado para a emoção alegria. O paradoxo é que se por um lado os pesquisadores ainda não tiveram sucesso em encontrar parâmetros acústicos robustos e consistentes para discriminar as emoções, por outro lado juízes mostram grande coerência em reconhecer emoções expressas na fala em experimentos de percepção, com porcentagem de acerto muito superior àquela que seria obtida ao acaso (Scherer, 1981), mesmo quando avaliando estímulos em uma língua estrangeira e de uma cultura diferente (Abelin e Allwood, 2000; Scherer *et al.*, 2001; Menezes *et al.*, 2012).

Os avanços nessa linha de pesquisa contribuem para a área de síntese de fala, pois possibilitam o desenvolvimento de uma fala sintetizada mais expressiva, o que por sua vez permite a elaboração de conversores texto-fala mais sofisticados, que podem ser aplicados, por exemplo, na educação de crianças (Alm e Sproat, 2005). O desenvolvimento de aplicativos que fazem o reconhecimento automático da emoção do falante por meio de combinações de parâmetros acústicos permite o aperfeiçoamento dos sistemas de interação homem-máquina e também oferece outras aplicações, como o monitoramento do estado afetivo de clientes durante ligações para *call centers*, por exemplo, que permite aos atendentes saber quando o cliente está ficando muito insatisfeito com o atendimento e então transferir a ligação para uma pessoa mais experiente (Petrushin, 2000; Vogt, André e Wagner, 2008). As pesquisas sobre fala expressiva são, ainda, muito importantes para a área de fonética (acústica e articulatória), pois nelas são explorados os limites de variação dos parâmetros acústicos da fala, contribuindo, assim, para um maior entendimento sobre o funcionamento da produção e da percepção da fala.

Os fatos expostos acima apontam para uma necessidade de se realizar mais pesquisas sobre o tema. A pesquisa apresentada teve como principal objetivo contribuir para um melhor conhecimento sobre como as emoções afetam a fala e como os sujeitos percebem essas alterações, determinando quais dos parâmetros acústicos investigados melhor discriminam as emoções expressas nos trechos de fala dos nossos corpora, quais deles os sujeitos utilizam para inferir o estado emocional do falante pela sua voz e estudar como alguns itens lexicais e algumas dimensões emocionais são usados pelos sujeitos para

identificar o estado emocional do falante em experimentos de percepção da fala. Além disso, conduzimos, com o apoio da “Bolsa Estágio de Pesquisa no Exterior” da FAPESP (BEPE - Mestrado), um estudo intercultural com ouvintes brasileiros e suecos com a finalidade de investigar se a cultura e a língua materna dos sujeitos influenciam na percepção de emoções expressas pela fala.

No capítulo 1 é feita uma revisão bibliográfica sobre as principais teorias que se propõem a explicar o que são as emoções, quais suas funções e as etapas de um processo emocional. Abordo também o papel desempenhado pela cultura na expressão e no reconhecimento das emoções e as vantagens e desvantagens de dois tipos de corpus muito utilizados atualmente: emoções reais e emoções atuadas.

No capítulo 2 é feita uma breve introdução aos aspectos básicos da produção da fala e à fonética acústica. Em seguida, os principais parâmetros acústicos investigados nas pesquisas sobre fala expressiva são descritos, incluindo aqueles estudados nesta pesquisa. É apresentado, também, um resumo das alterações em alguns parâmetros acústicos encontradas em vários estudos para algumas emoções.

No capítulo 3 descrevo a metodologia, as análises e os resultados referentes ao experimento de percepção I, no qual sujeitos ouvintes brasileiros e suecos avaliaram, em escalas graduadas de 0 a 4, o grau de expressão de **emoções discretas** descritas por oito palavras, para os mesmos trechos de fala, que foram extraídos de depoimentos reais de mulheres falantes do **português brasileiro**.

No capítulo 4 é descrito o experimento de percepção II, no qual sujeitos ouvintes brasileiros e suecos avaliaram, em escalas graduadas de 0 a 4, o grau de expressão de cinco **dimensões emocionais** em enunciados extraídos de situações comunicativas reais de mulheres falantes do **sueco**.

PARTE I

REVISÃO DA LITERATURA

Capítulo 1

As emoções: Abordagens teóricas e aspectos metodológicos

O ser humano possui a característica de ser social e, por isso, está em constante interação com seus coespecíficos. As emoções são fundamentais nesse processo e desempenham um papel importante na formação e na manutenção de laços sociais, tais como amizade e casamento. Isso porque, como bem lembrado por Pinker (2009 [1997]: 392):

Emoção não é só fugir correndo de um urso. Ela pode ser desencadeada pelo processamento de informações mais complexo de que a mente é capaz, como por exemplo quando lemos uma carta da pessoa amada terminando o relacionamento ou chegamos a casa [sic] e deparamos com uma ambulância na porta. E as emoções ajudam a engendrar intrincadas tramas de fuga, vingança, ambição e namoro.

Segundo Ekman (1992), indivíduos que possuem paralisia nos músculos da face (síndrome de Möbius) e que são, por isso, incapazes de produzir expressões faciais, dizem ter grande dificuldade em criar e manter relacionamentos, mesmo que breves. Além disso,

Ross (1981 *apud* Ekman, 1992) reportou que pacientes afásicos que não eram capazes de identificar ou produzir a prosódia característica de certas emoções tinham grave dificuldade nas interações do dia-a-dia.

A interação humana é tão singular que algumas emoções parecem ser exclusivas aos seres humanos, como a vergonha (*shame*), a culpa (*guilt*), o orgulho (*pride*) e o amor (*love*), por exemplo. A razão disso pode estar no fato de que essas emoções surgiram através da grande capacidade cognitiva da espécie humana e também de sua complexa atividade social (Lazarus, 1991). A existência de certas emoções e a ausência de outras depende do ambiente em que determinada espécie habita. Isso se deve, muito provavelmente, à força exercida pela seleção natural ao longo da evolução das espécies (Pinker, 2009 [1997]).

A verdade é que as emoções dão um tom colorido à nossa vida e é muito difícil imaginá-la sem elas. Praticamente todas as nossas atividades envolvem emoções: nossos sonhos, objetivos, viagens, família etc. Além disso, o ser humano aprendeu a utilizar as emoções também como meio de entretenimento. Filmes, esportes, livros, músicas, danças e todos os tipos de competições são exemplos de atividades enraizadas nas emoções que causam ao público.

Neste capítulo, são apresentadas várias teorias que buscam explicar a função das emoções, como elas são expressas e como afetam o organismo. Segundo Cornelius (2000), é importante que os pesquisadores que estudam as emoções conheçam as teorias sobre o tema. Isso permite a esses pesquisadores saber que hipóteses fazer sobre as emoções que pretendem estudar, bem como que conclusões tirar a partir dos resultados obtidos. Veremos então que um episódio emocional é um fenômeno muito complexo que intriga pesquisadores de várias áreas do conhecimento.

1.1 O QUE SÃO AS EMOÇÕES?

É muito difícil definir o que é “emoção”, pois ela é um construto hipotético, ou seja, não é um fenômeno diretamente observável. Sua existência é inferida a partir de várias reações que ela causa no organismo. Não é claro quantos componentes são necessários para definir o que é um estado emocional. Para alguns pesquisadores, por exemplo, a emoção é somente o sentimento que temos sobre a situação vivida, enquanto que para outros ela também abrange as alterações fisiológicas e as expressões corporais a ela associada, entre outros componentes. Além disso, vários aspectos, como as diferentes formas de expressão, a aparente universalidade de algumas emoções, a influência da cultura e das normas sociais ou as diferentes funções desempenhadas pelas emoções, por exemplo, deram origem a muitas teorias para explicar o que são e qual a função delas.

A maioria dos estudiosos concorda que as emoções são **processos** que se encontram em contínua mudança e cuja ocorrência se dá por etapas (Scherer, 1984; Frijda e Mesquita, 1994). Algumas dessas etapas são causas, como a ocorrência de um evento relevante para o organismo, por exemplo, e outras são consequências, como as alterações corporais sofridas pelo organismo (Scherer, 2000a).

Atualmente, a perspectiva dominante é a **visão componencial** (*componential view*), para a qual as emoções são processos formados por vários componentes em interação, sendo o sentimento apenas um deles. Essa posição difere, portanto, da compartilhada pelo senso comum, que vê a emoção apenas como sentimento, ou seja, como sendo a experiência subjetiva do estado de ativação emocional causado por um evento (Scherer, 2000a). Os demais componentes são essenciais para descrever e caracterizar os processos emocionais, que, como já mencionado, ocorrem por etapas. Portanto, os termos “emoção” e “sentimento” **não** devem ser tratados como sinônimos (Scherer, 2005).

A **perspectiva cognitivista**, adotada por este trabalho e a ser apresentada mais adiante neste capítulo, reconhece os seguintes componentes, de acordo com Scherer (2000a): sentimento, padrões de respostas neurofisiológicas, expressões motoras (na face, voz e gestos), tendências à ação e avaliação. Mesquita e Frijda (1992) incluem nessa lista

os eventos antecedentes, categorização do evento e controle. Esses componentes serão descritos mais adiante, quando falarmos da perspectiva cognitivista.

O conceito de “emoção” torna-se mais claro quando o distinguimos dos demais estados do afeto. Além das emoções, Scherer (2000b) identifica mais quatro estados afetivos, resumidos a seguir:

- **Humor:** É um estado afetivo de duração relativamente longa (podendo durar dias) e que não apresenta uma causa aparente. Tem um baixo impacto no comportamento, baixo nível de sincronização dos subsistemas do indivíduo e baixa intensidade. Exemplos: apatia, melancolia, depressivo, tédio etc.
- **Postura interpessoal:** Postura afetiva adotada por uma pessoa para com seu interactante em uma determinada interação. A duração e a intensidade variam de rápida para média. Exemplos: distante, próximo, apoiando, indiferente etc.
- **Atitudes:** Posição que é adotada intencionalmente pelo falante em uma interação para expressar sua opinião e crenças ao seu interlocutor. A intensidade da expressão varia de nenhuma até média e a duração, de média a longa. Exemplos: sarcasmo, ironia, incredulidade, certeza etc. Para um exemplo de como as atitudes se relacionam com a prosódia da fala, ver Rilliard *et al.* (2012).
- **Personalidade:** Características que determinam como o indivíduo age e pensa. A duração é longa, pois dificilmente um traço de personalidade de uma determinada pessoa se altera. Exemplos: tímido, extrovertido, fechado, cuidadoso etc.

Segundo Scherer (2000b), as **emoções** são mais intensas, mas de curta duração (Ekman, 1992, fala em segundos de duração). São produzidas a partir de uma avaliação cognitiva automática do indivíduo sobre eventos externos e internos como sendo de relevância para seus interesses e necessidades. Também são caracterizadas por um alto grau de sincronização de quase todos os subsistemas do organismo (cognitivo, motivacional,

fisiológico e motor) e se alteram rapidamente. Além disso, trata-se de um fenômeno involuntário (apesar de poder ser disfarçado ou atuado).

1.2 A PERSPECTIVA DARWINISTA DAS EMOÇÕES

Na antiguidade, as emoções eram consideradas pelos filósofos (como Sócrates, Platão e outros) como ameaças à nossa razão, pois nos impediriam de sermos racionais. As melhores decisões deveriam ser tomadas quando não estivéssemos “influenciados” por uma emoção. Por isso, as emoções deveriam ser, sempre que possível, suprimidas ou mantidas em harmonia com a razão (Solomon, 2008).

As emoções começaram a ser vistas de uma forma diferente com o estudo pioneiro de Charles Darwin, que, em seu livro “A expressão das emoções no homem e nos animais” (Darwin, 2009 [1872]) apresentou uma teoria inovadora para explicar as expressões corpóreas causadas por emoções em várias espécies, incluindo a humana, identificando várias expressões semelhantes entre elas. Além disso, a partir de um questionário sobre expressões humanas que enviou a pesquisadores de todo o mundo, concluiu que várias expressões eram expressas de maneira semelhante em todas as raças humanas.

A hipótese de Darwin (2009 [1872]) é a de que as emoções são produtos da evolução e foram selecionadas por ajudarem nossos ancestrais a enfrentar problemas rotineiros. De acordo com essa perspectiva, as emoções têm a função de garantir a sobrevivência do organismo, preparando-o para enfrentar os eventos encontrados no ambiente por meio de alterações corporais, como explicou Nesse (1990: 268 *apud* Johnson, s.d.):

The emotions are specialized modes of operation shaped by natural selection to adjust the physiological, psychological, and behavioral parameters of the organism in ways that increase its capacity and tendency to respond adaptively to the threats and opportunities characteristic of specific kinds of situations.

A perspectiva darwinista deu origem a uma abordagem teórica das emoções conhecida como *discrete emotion theory* (**teoria das emoções discretas**¹). Para essa teoria, os processos emocionais são diferentes entre si e possuem características (eventos antecedentes, avaliação, alterações fisiológicas, características expressivas e comportamentais) **únicas** (Ekman, 1992). Essas emoções são conhecidas na literatura como **básicas, fundamentais** ou ainda **discretas**. A existência delas está limitada a um pequeno número que varia de 6 a 14, dependendo do pesquisador (Scherer, 2000b). As emoções discretas recebem nomes das línguas naturais, tais como “alegria”, “raiva”, “tristeza”, “medo”, “nojo” etc. As demais emoções seriam de alguma forma derivadas das emoções básicas.

Para Ekman (1992), seriam seis as emoções básicas: alegria, tristeza, medo, nojo, raiva e surpresa. Essa lista é bastante conhecida na literatura, sendo às vezes referida como “*The Big Six*” (Cornelius, 2000). Ekman (1992:172) afirma que cada emoção básica não é um único estado afetivo, mas sim uma “família” de estados relacionados, cujos membros compartilham algumas características em comum, como expressões, alterações fisiológicas etc. O autor também apresenta nove critérios para que se considere uma emoção como “básica”. Todas essas características teriam contribuição biológica. São elas:

1. Sinais universais distintos;
2. Presença em outros primatas;
3. Atividades fisiológicas distintas;
4. Eventos antecedentes universais e distintos;
5. Coerência nas respostas emocionais;
6. Início rápido;
7. Duração breve;
8. Avaliação automática (*appraisal*);
9. Ocorrência espontânea.

¹ O adjetivo “discreto” é utilizado aqui como sinônimo de “distinto”, a partir da tradução do termo em inglês “*discrete*”. Um uso semelhante é feito pela matemática e pela física.

Os pesquisadores que adotam essa abordagem buscam identificar padrões de resposta específicos a cada uma das emoções básicas, não apenas na sensação, mas também em respostas fisiológicas e na expressão. Além de expressões faciais diferentes, também são esperados padrões específicos para a voz (Scherer, 2000a), embora ainda não tenham sido bem identificados.

A visão evolucionista das emoções implica a noção de **universalidade**: se as emoções básicas são um construto da evolução, então elas devem se manifestar de forma semelhante em todas as culturas, uma vez que os seres humanos constituem uma única espécie, a do *Homo sapiens*. A universalidade é, portanto, condição *sine qua non* da teoria evolucionista das emoções (Scherer *et al.*, 2001; Ekman, 1992). Entretanto, não é possível afirmar, *a priori*, o inverso: somente o fato de uma emoção ser reconhecida universalmente não corrobora a hipótese de que ela seja produto da evolução. Outra explicação pode estar na aprendizagem de regras sociais, na disseminação global da mídia etc. (Ekman, 1992; Cornelius, 2000).

Para investigar a universalidade das expressões emocionais, a metodologia adotada por esses pesquisadores consiste geralmente em apresentar a sujeitos de diversas culturas fotografias de expressões faciais e pedir-lhes que escolham uma palavra dentre as sugeridas para descrever a emoção expressa em cada face. Esse tipo de metodologia incorre em algumas dificuldades, como a de encontrar uma palavra correta para descrever a emoção expressa e tem recebido críticas devido ao número limitado de alternativas que foram apresentadas aos sujeitos (mais detalhes na seção 1.6).

1.3 A TEORIA DE WILLIAM JAMES

Essa perspectiva tem sua origem nos trabalhos de William James, um dos fundadores da psicologia moderna, particularmente em seu artigo de 1884 intitulado “*What is an Emotion?*”. A teoria de James, também conhecida como “teoria James-Lange” por também ter sido proposta pelo fisiologista Carl Lange, ganhou destaque por ir contra o senso comum. Conforme o próprio James (1884) salienta, o senso comum acredita que as alterações corporais (arrepios, alterações no ritmo dos batimentos cardíacos, sorriso etc.)

ocorrem **após** o sentimento da emoção, ou seja, primeiramente há a consciência do estado emocional e depois as reações fisiológicas. A teoria de James, ao contrário, sugere que “*the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion*” (James, 1884:189). Ou seja, para a perspectiva jamesiana, após o organismo presenciar um evento relevante no ambiente, primeiramente ocorrem as respostas fisiológicas e sua percepção dessas respostas é a emoção. James exemplifica sua ideia dizendo que “*(...) we feel sorry because we cry, angry because we strike, afraid because we tremble (...)*” (James, 1884:190). Para James, seria impossível ter emoções sem as alterações corpóreas:

If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its characteristic bodily symptoms, we find we have nothing left behind, no “mind-stuff” out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains.

(James, 1884: 193)

Segundo Cornelius (2000), desde a publicação do artigo de James (1884), pesquisadores têm buscado evidências para sustentar ou refutar essas hipóteses. Alguns estudos (Hohmann, 1966; Chwalisz, Diener e Gallagher, 1988)² mostraram que sujeitos com lesão na medula espinhal tiveram uma diminuição na sensação de algumas emoções e que essa diminuição era proporcional ao grau da lesão. Entretanto, todos esses sujeitos ainda tiveram resposta da musculatura facial e a diminuição da sensação devido à lesão só ocorreu para algumas emoções, principalmente para aquelas relacionadas à raiva, sendo que para algumas, como o amor e a compaixão, houve até um aumento na sensação.

Recentemente, a perspectiva de James tem sido expandida por Prinz (2004 *apud* Johnson, s. d.) e Damasio (1994).

² *apud* Cornelius (2000).

1.4 A PERSPECTIVA COGNITIVISTA

A perspectiva cognitivista é a que adotamos nesta pesquisa e é a mais difundida atualmente (Cornelius, 2000). Ela não é uma única teoria, mas sim uma “família” delas e surgiu da noção de que a cognição desempenha um papel fundamental na experiência das emoções. A reação de um indivíduo a um evento no ambiente depende de como ele avalia esse evento. Ao encontrar uma ameaça (um ladrão ou um animal, por exemplo), uma pessoa pode ficar com medo ou com raiva, dependendo de como julgar seu potencial de ação, além de outros fatores (Scherer, 2000a). A seguir, apresento os componentes do processo emocional.

De acordo com essa perspectiva, os processos emocionais têm início com a **avaliação** (*appraisal*) de um evento no ambiente (**evento antecedente**) como sendo relevante para as necessidades e interesses do indivíduo. Esse processo é automático e inconsciente, isto é, não depende de ação deliberada para ocorrer. Entretanto, ele é mediado pela **categorização do evento**, isto é, pelo significado que o evento antecedente tem para aquele sujeito. Essa categorização depende do conhecimento partilhado pela cultura, do histórico de vida, da personalidade e das crenças pessoais do indivíduo. Dessa forma, a emoção que o indivíduo irá experienciar não depende da natureza do evento em si, mas sim de como ele é categorizado pelo sujeito (Mesquita e Frijda, 1992). Assim, um mesmo evento pode desencadear emoções diferentes em diferentes culturas e até mesmo entre diferentes pessoas. O processo de avaliação é, portanto, a ligação entre o indivíduo e a situação que desencadeou a emoção (Ellsworth e Scherer, 2003: 574).

Após os processos de avaliação e de categorização do evento, o indivíduo sofre alteração em seu estado de prontidão: ele pode ficar mais alerta e predisposto a agir, querer se proteger de uma ameaça, sentir vontade de chorar, de sorrir ou de se esconder (no caso da vergonha), entre outros. Essas alterações no estado de prontidão do indivíduo constituem o componente conhecido como “**tendências à ação**”. É importante ressaltar que a ação em si (como o chorar, o rir, o correr etc.) não é considerada como componente do processo emocional, mas sim como comportamentos que ocorrem em consequência da emoção (Scherer, 2000a).

Junto com a alteração no estado de prontidão estão as **alterações neurofisiológicas** que ocorrem no organismo e os “sintomas” associados, como batimentos cardíacos acelerados, alteração na respiração e na salivação, dilatação das pupilas, alteração na temperatura da pele, entre outros. Embora algumas dessas alterações indiquem claramente excitação emocional e sejam, por isso, usadas para inferir o estado emocional de uma pessoa (em experimentos de reconhecimento de emoções, por exemplo), a maioria dos psicólogos acredita que a função primária das alterações fisiológicas seja prover energia e preparar o indivíduo para a ação, como é possível observar neste exemplo de Pinker (2009 [1997]: 394):

O medo também aperta um botão que prepara o corpo para a ação, a chamada reação lutar ou fugir. [...] O coração bate forte para enviar sangue aos músculos. O sangue é redirecionado do tubo digestivo e da pele, provocando dor de estômago e suor frio. A respiração rápida absorve oxigênio. A adrenalina libera combustível do fígado e ajuda o sangue a coagular. E dá ao nosso rosto aquela aparência universal de corça sob a luz dos faróis dianteiros. [*sic*]

Outro componente do processo emocional é o **sentimento**, que é a experiência subjetiva que o indivíduo tem da ativação (excitação) emocional. Segundo Scherer (2005: 699), o sentimento tem uma função importante no monitoramento dos estímulos internos e externos ao indivíduo e, portanto, reflete todo o processo de avaliação de um determinado evento.

Por fim, as emoções também estão sujeitas ao **controle** (*regulation*) do indivíduo, que pode tentar disfarçar uma expressão (caso da vergonha, por exemplo) ou exagerá-la (quando alguém tenta fingir a alegria, por exemplo). O processo de controle geralmente depende de normas sociais e culturais e é importante ressaltar que ele não inibe o processo emocional (por exemplo, mesmo disfarçando a expressão, o sujeito ainda está experienciando a emoção).

1.4.1 O modelo de processo componencial de Klaus Scherer

Scherer (1986) desenvolveu um modelo, conhecido como “modelo de processo componencial”, que busca explicar os critérios pelos quais o processo de **avaliação cognitiva** ocorre. Como é enfatizado por Scherer (1986: 146), a emoção é vista nesse modelo como um **processo** e como **resposta adaptativa de vários subsistemas** (componentes) do indivíduo a um evento antecedente relevante para ele. Percebe-se aqui a presença da perspectiva cognitivista no modelo aliada à visão evolucionista das emoções, pois para Scherer as alterações nos subsistemas do indivíduo teriam a função de ajudar esse indivíduo a sobreviver (ver seção 1.2).

A ideia central desse modelo é a de que os diversos subsistemas de processamento de informação do indivíduo (cognitivo, motivacional, fisiológico e motor) continuamente checam estímulos internos e externos através de **critérios definidos**, os **SECs** (*Stimulus Evaluation Checks*), os quais ocorrem em uma ordem fixa. O autor especifica os seguintes SECs:

- 1. Checagem de novidade:** verifica se um evento novo ocorreu, ou seja, se houve alteração no padrão de estimulação interna ou externa;
- 2. Checagem de prazer intrínseco:** avalia se o estímulo é agradável ou não. Se agradável, o estímulo provocará uma tendência à aproximação, caso contrário a tendência será de afastamento. Ainda, este SEC é baseado em associações inatas ou aprendidas.
- 3. Checagem de relevância:** verifica se o estímulo é relevante para as metas ou necessidades do indivíduo, se é consistente ou não com as expectativas que o indivíduo tem da situação, se o evento o facilita ou o impede de alcançar seu objetivo e se é necessária uma resposta comportamental urgente. De acordo com Scherer (1986:152), o resultado deste SEC determina o grau de **envolvimento** do indivíduo com o evento.
- 4. Checagem do potencial de controle:** avalia se o indivíduo tem potencial para mudar ou evitar as consequências do evento e também seu grau de controle sobre o mesmo. Esse SEC pode ser exemplificado pela situação na qual um homem encontra um urso na mata. Se o

homem julgar que pode derrotar o urso, ele lutar. Caso contrário é provável que fuja ou se esconda.

5. Checagem da compatibilidade do evento: checa se o evento está em conformidade com normas sociais e culturais e também se é consistente com a imagem que o indivíduo tem de si mesmo (normas internas).

Os resultados dos SECs terão influência no sistema nervoso somático e no sistema nervoso autônomo do organismo. Este último tem um efeito direto na fala, já que é responsável por controlar a respiração e a salivação. O sistema nervoso somático, por sua vez, está relacionado ao controle voluntário dos músculos.

Para compreender os efeitos dos SECs nos parâmetros acústicos da voz, considere o seguinte exemplo (de Scherer, 1986: 148): o sujeito, ao avaliar um estímulo como perigoso, teria a frequência fundamental da voz aumentada devido à resposta do sistema nervoso somático (que produziria um aumento da tensão muscular e, conseqüentemente, um aumento na taxa de vibração das pregas vocais). Do mesmo modo, a salivação diminuiria devido às alterações no sistema nervoso autônomo, causando alterações nas frequências de ressonância (formantes) do trato vocal. Esses efeitos fariam com que a voz desse indivíduo soasse mais aguda.

Pode-se ver, portanto, que a grande vantagem desse modelo é possibilitar que hipóteses sobre os efeitos dos SECs no corpo humano (e, conseqüentemente, nos parâmetros acústicos da fala) sejam feitas e testadas empiricamente, através das alterações esperadas nos órgãos e mecanismos envolvidos na produção da fala.

1.5 A VISÃO DIMENSIONAL DAS EMOÇÕES

Para alguns pesquisadores, as emoções possuem características em comum (dimensões ou primitivos emocionais), diferentemente do que é proposto pela perspectiva das emoções básicas, e diferem quanto ao nível de cada uma delas. Essas dimensões se manifestam por um **contínuo comportamental**, e não apenas por pólos discretos de máxima ou mínima

intensidade (Schlosberg, 1954). Uma pessoa pode, por exemplo, apresentar um grau “baixo” de raiva, caracterizado por ruminção, mas sem o desejo de agir ou um grau elevado da mesma emoção, em que sente um impulso grande de agir ou até mesmo de quebrar algo. Scherer, que em suas pesquisas tem adotado a perspectiva discreta, também já propôs distinguir entre “*cold anger*” para o primeiro caso e “*hot anger*” para o segundo (Scherer, 1986). Obviamente, esses dois exemplos são extremos do contínuo comportamental da emoção “raiva”.

As principais dimensões (as mais estudadas) são a ativação, o prazer intrínseco e a dominância, mas há estudiosos que chegam a listar dezenas delas (como Frijda *et al.*, 1995, por exemplo). A **ativação** diz respeito ao grau de atividade do organismo, indo de calmo a agitado. Por sua vez, o prazer intrínseco, também conhecido como **valência**, tem a ver com a experiência subjetiva do grau de prazer que o evento oferece, sendo as emoções comumente distinguidas nessa dimensão em positivas ou negativas. Por fim, a **dominância** refere-se à possibilidade do organismo em lidar com a situação, isto é, se ele está submisso ou se pode mudar o curso dos eventos (Kehrein, 2002). Esta última dimensão é bastante semelhante ao quarto mecanismo de checagem proposto por Scherer (1986), “checagem do potencial de controle” (cf. seção 1.4.1). A dimensão de valência seria uma combinação dos mecanismos de checagem de prazer intrínseco (2) com o de checagem da possibilidade de atingir o objetivo, que é parte do mecanismo (3). Por fim, a dimensão de ativação seria uma combinação dos sub-mecanismos da relevância, expectativa, urgência (partes do mecanismo 3) e controle (parte do mecanismo 4) (Scherer, 1986: 157).

A ideia central da perspectiva dimensional é, portanto, a de que todos os estados emocionais podem ser descritos com relação ao seu grau para cada dimensão. Por exemplo, a emoção que descrevemos por “raiva” apresenta prazer intrínseco negativo e um grau de ativação de médio para alto. Por sua vez, a emoção que descrevemos pelo rótulo “alegria” difere de “raiva” por apresentar prazer intrínseco positivo enquanto que a emoção “tristeza” apresenta prazer intrínseco negativo e um grau de ativação baixo. Contudo, o grau das dimensões pode variar para a mesma emoção, dependendo de como o evento é avaliado pelo indivíduo. Sobre a ativação, Schlosberg (1954: 87) escreveu que “*This*

continuum, general level of activation, has its low end in sleep, its middle ranges in alert attention, and its high end in the strong emotions”.

As dimensões de ativação e prazer intrínseco foram as primeiras a serem propostas e são as mais investigadas, tanto com relação às expressões faciais quanto às expressões vocais. Wundt (1905 *apud* Scherer, 2000b), um dos primeiros a adotar uma visão multidimensional das emoções, propôs que a natureza do estado emocional seria determinada por três dimensões independentes: *pleasantness-unpleasantness* (valência), *rest-activation* (ativação) e *relaxation-attention* (atenção). Russell (1980) propôs um modelo circular no qual classifica as emoções com base em dois eixos ortogonais que representam essas duas dimensões (*pleasure-displeasure* e *arousal*). Schlosberg (1941) obteve, a partir da escala de expressões faciais de Woodworth (1938), um círculo cujos eixos ortogonais eram gerados pelas dimensões de prazer intrínseco (*pleasantness - unpleasantness*) e atenção-rejeição (*attention – rejection*). Entretanto, o autor reportou grande dificuldade em explicar essa última dimensão aos sujeitos. De acordo com Barbosa (2008), a oposição entre os pólos “atenção” e “rejeição” pode ser captada pela dimensão de **envolvimento**.

O argumento que é frequentemente usado pelos pesquisadores adeptos da perspectiva discreta contra o uso de dimensões é o de que as dimensões não permitem descrever as emoções de forma completa. De fato, somente as dimensões de prazer intrínseco e ativação não são suficientes para diferenciar, por exemplo, as emoções “medo” e “raiva”, que são ambas negativas com relação à valência e geralmente de um alto grau de ativação (Laukka, 2004). Vê-se, portanto, a necessidade de se utilizar também outras dimensões. A dimensão da dominância, também conhecida por “poder”, “controle” ou “potência”, parece ser importante para diferenciar emoções de prazer intrínseco negativo, como o “medo” e a “raiva” (Smith e Ellsworth, 1985 *apud* Laukka, 2004). Entretanto, essa dimensão parece não ser muito bem expressa pela fala, pois Barbosa (2009), por exemplo, obteve um índice *kappa* de concordância entre juízes muito baixo para a dimensão de dominância. Além disso, Amir *et al.* (2010) obtiveram taxas de reconhecimento automático (por meio de combinações de parâmetros acústicos) muito baixas para as dimensões de dominância e valência, e concluíram então que os parâmetros acústicos são mais úteis em

prever a ativação do que estas outras duas dimensões (veja também um resultado semelhante em Lugger e Yang, 2007).

A figura 1.1 mostra uma classificação de algumas emoções de acordo com as dimensões de ativação, valência e dominância (potência), proposta por Paescke (2003) e Colamarco (2009). Os pares de emoções desespero e depressão, fúria e irritação, exultação e satisfação, representados nessa figura, correspondem a variações de uma mesma experiência emocional para a dimensão de ativação, a saber, tristeza, raiva e alegria, respectivamente. As emoções desespero, fúria e exultação possuem um grau de ativação maior do que depressão, irritação e satisfação.

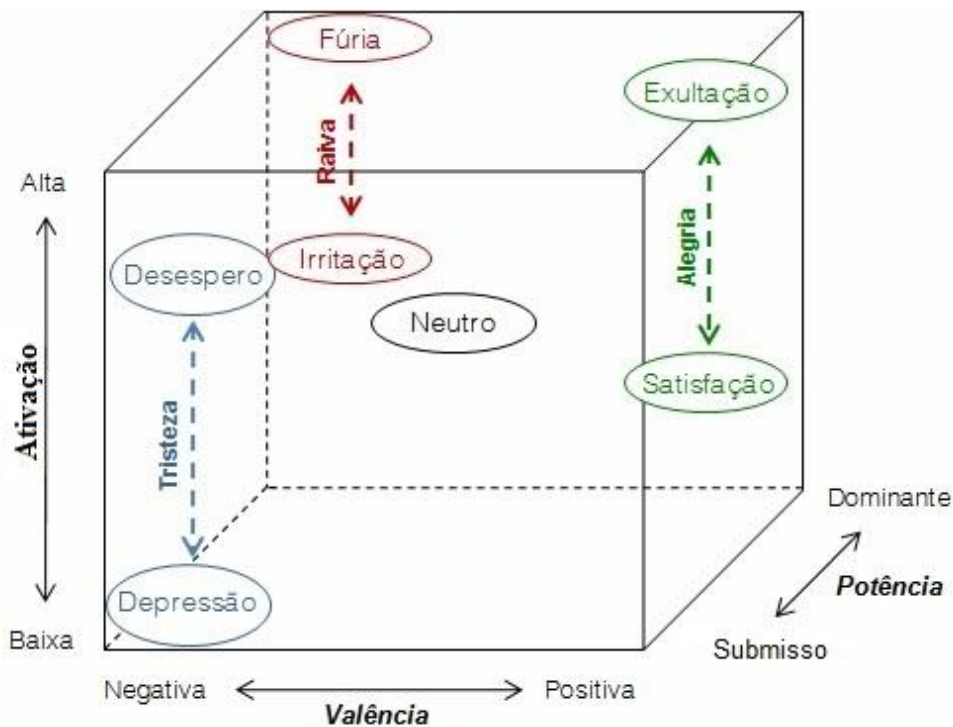


Figura 1.1: Distribuição de algumas emoções de acordo com três dimensões emocionais (ativação, valência e potência) de Paescke (2003). Adaptado de Colamarco (2009).

Os pesquisadores adeptos à abordagem dimensional evitam, portanto, usar os rótulos de emoções discretas, como “alegria”, “tristeza”, “raiva” etc. Entretanto, como bem notado por Mauss e Robinson (2009), é possível conciliar, em parte, as duas perspectivas ao assumir que cada emoção discreta é formada por uma combinação única dessas dimensões, como exemplificado acima. De fato, segundo Paescke (2003 *apud* Colamarco, 2009), muitos autores acreditam que a perspectiva discreta e a dimensional não são mutuamente exclusivas, pois “*todas as emoções discretas estão localizadas em um ou mais pontos do continuum de um modelo dimensional*” (tradução de Colamarco, 2009: 20).

Esses pesquisadores seguem uma metodologia diferente da utilizada pelos pesquisadores afiliados à perspectiva das emoções discretas. Eles não buscam encontrar uma palavra apropriada para descrever uma expressão emocional, como estes últimos fazem, mas sim pedem a juízes que avaliem cada estímulo (expressões faciais, trechos de fala etc.) de acordo com algumas dimensões emocionais, representadas por escalas nas quais os juízes podem escolher, além dos valores máximos e mínimos, valores intermediários (Schlosberg, 1941; Pereira, 2000; Laukka, 2004; Barbosa, 2009; Amir *et al.*, 2010). A partir das respostas coletadas, esses pesquisadores obtêm padrões para as dimensões para cada estado emocional estudado. Uma das vantagens dessa metodologia é a de que não há a necessidade de escolher palavras apropriadas para descrever as emoções, um problema que cria muitas dificuldades, sobretudo em estudos interculturais (Goddard, 2002). Por outro lado, muitos pesquisadores da perspectiva dimensional utilizam antônimos para nomear os pólos máximo e mínimo das escalas, como no caso da escala semântica diferencial proposta por Osgood (1952). Exemplos: calmo – agitado, submisso – sob controle, desagradável - agradável (Barbosa, 2009); calmo – excitado, fraco/submisso – forte/dominante (Kehrein, 2002). Isso é um problema, pois alguns estudos recentes sugerem que nem sempre dois termos antônimos são tratados por juízes como valores opostos de uma escala, mas sim se comportam como dois afetos distintos, como, por exemplo, prazer e desprazer (*pleasure - displeasure*) (Schimmack, 2001). Uma solução para esse problema é utilizar uma palavra apenas, auxiliada pelos advérbios “nada” (para o valor mínimo da escala) e “muito” (para o valor máximo), como fazemos nesta pesquisa (exemplo: nada agitado – muito agitado).

Com relação às pesquisas sobre expressão de emoções na fala, o uso de dimensões emocionais parece incorrer em uma taxa de reconhecimento maior das emoções pelos sujeitos do que as taxas frequentemente obtidas com etiquetas de emoções discretas (Lugger e Yang, 2007). Talvez isso ocorra porque, como alguns estudos já reportaram, emoções com um nível semelhante para algumas dimensões (ativação e prazer intrínseco, por exemplo), compartilham das mesmas características para alguns parâmetros acústicos (relativos à frequência fundamental e intensidade), o que pode gerar confusão ao se tentar discriminar essas emoções (Pereira, 2000).

1.6 O PAPEL DA CULTURA NA EXPRESSÃO E NO RECONHECIMENTO DE EMOÇÕES

Estereótipos sobre diferenças culturais com relação a algumas emoções são muito comuns. Vejamos alguns deles. Falou-se que o povo Utku Inuit não possui uma palavra que designe “raiva” e que, portanto, não experimenta essa emoção (Briggs, 1970 *apud* Mesquita e Frijda (1992). Os taitianos, por sua vez, não sentiriam culpa, tristeza, saudade (*longing*) ou solidão e descrevem o pesar (*grief*) como fadiga, doença ou sofrimento físico. Os japoneses são tomados pelo medo de envergonhar a família. Os alemães seriam os únicos a experimentar a emoção *Schadenfreude* (prazer com o dano alheio), simplesmente porque possuem uma palavra para designá-la.³

Em seu estudo pioneiro sobre a expressão das emoções, Darwin (2009 [1872]) elaborou um questionário referente a expressões humanas causadas por emoções e o enviou a vários observadores de populações aborígenes de diferentes lugares no mundo. Ele solicitou que as questões fossem respondidas com base apenas na observação direta e não na memória. O autor, então, escreveu:

³ Exemplos de Pinker (2009 [1997]).

Conclui-se, a partir das informações assim adquiridas, que um mesmo estado de espírito exprime-se ao redor do mundo com impressionante uniformidade; e este fato é ele mesmo interessante como evidência da grande similaridade da estrutura corporal e da conformação mental de todas as raças humanas.

(Darwin, 2009 [1872]: 23)

Desde então, pesquisadores têm investigado de forma mais elaborada a hipótese da universalidade da expressão de algumas emoções. Essas pesquisas, realizadas sobretudo com expressões faciais, revelaram que essa universalidade existe somente até certo ponto, mas que diferenças na expressão e no reconhecimento de algumas emoções também existem entre culturas.

Ekman *et al.* (1969) mostraram trinta fotografias de expressões faciais de seis emoções (felicidade, medo, nojo, raiva, surpresa e tristeza) a sujeitos dos Estados Unidos, Brasil, Japão, Nova Guiné e da ilha de Bornéu. A hipótese dos pesquisadores era a de que sujeitos de diferentes culturas reconhecem a mesma emoção para as mesmas fotografias. Após observarem a fotografia, os sujeitos tinham que escolher uma palavra de uma lista com as seis emoções. Os autores obtiveram altas porcentagens de reconhecimento para os sujeitos alfabetizados (Estados Unidos, Brasil e Japão). Entretanto, as porcentagens de reconhecimento foram menores para as culturas iletradas, as quais tiveram pouco ou nenhum contato com a mídia (filmes, televisão ou revistas). Mesmo assim, elas apresentaram reconhecimento semelhante para as emoções felicidade, raiva e medo em todas as amostras e para nojo, surpresa e tristeza em duas das três amostras. Os autores concluíram, então, que essas expressões faciais são universais, pois foram reconhecidas por sujeitos de diversas culturas. Porém, alguns problemas metodológicos não podem deixar de ser notados. Os autores tiveram que traduzir os termos emocionais para as respectivas línguas dos sujeitos. Contudo, de acordo com os próprios autores, não havia um termo equivalente para as emoções de nojo e surpresa na língua pidgin neo-melanesiano (de Nova Guiné). Nesses casos, os pesquisadores substituíram os termos por sentenças como “olhando para algo que cheira mal” e “olhando para algo novo”. Essa diferença metodológica pode ter influenciado no reconhecimento das expressões por parte dos

sujeitos de Nova Guiné. Além disso, como apontado por Russell (1991 *apud* Goddard, 2002), o número de alternativas apresentadas aos sujeitos foi limitado. Das seis emoções utilizadas pelos pesquisadores, somente a emoção da felicidade é positiva (os autores não mencionaram se a variante da emoção “surpresa” estudada foi a positiva ou a negativa). Desse modo, os sujeitos podem ter sido “induzidos” a escolher uma palavra por falta de outra opção ou ter aplicado o artifício da exclusão. Esse método de “escolha forçada” pode ter influenciado os resultados de muitos estudos interculturais sobre o reconhecimento de emoções.

A fim de investigar se a cultura influencia nas expressões faciais das emoções, Ekman e Friesen (1975) conduziram um célebre experimento no qual mostraram filmes para elicitare o estresse em estudantes universitários norte-americanos e japoneses. O experimento foi aplicado em duas condições: na primeira, cada sujeito assistia ao filme sozinho e na segunda, o sujeito assistia ao filme com um assistente de pesquisa da mesma cultura que o sujeito. Os pesquisadores filmaram secretamente as expressões faciais dos sujeitos enquanto estes assistiam aos filmes, em ambas as condições. Os dados mostraram que, quando sozinhos, tanto os sujeitos norte-americanos quanto os japoneses exibiram expressões faciais de horror idênticas, mas quando assistiam ao filme acompanhados do assistente, os sujeitos japoneses disfarçaram as expressões faciais de horror muito mais do que os sujeitos norte-americanos. Os autores concluíram, então, que os sujeitos japoneses tendem a atenuar as expressões faciais de emoções quando em público, o que não ocorre (ou ocorre menos intensamente) com os sujeitos norte-americanos. Ainda, de acordo com os autores: “*The universal feature is the distinctive appearance of the face for each of the primary emotions. But people in various cultures differ in what they have been taught about managing or controlling their facial expressions of emotion*” (Ekman e Friesen, 1975: 24).

Ekman e Friesen (1975) interpretaram este último resultado com base no conceito de *display rules* (Ekman e Friesen, 1969), que seriam hábitos adquiridos que restringiriam quais emoções podem ser demonstradas em determinadas situações (exemplo: “homens não choram”) ou a maneira com a qual certas emoções podem ser expressas (disfarçando-as ou intensificando-as). Essas normas seriam adquiridas desde a infância e variam entre classes sociais e grupos étnicos, bem como entre culturas. Da mesma forma,

Scherer (1985) fala em *pull effects*, normas ou expectativas impostas pela cultura e pelo ambiente externo, como, por exemplo, a distância entre os organismos interagentes, que pode requerer um maior esforço vocal e os sinais sociais convencionalizados. Esses dois conceitos, *display rules* e *pull effects*, estão relacionados com o componente do controle (*regulation*), descrito na seção 1.4.

Outro componente do processo emocional que pode ser responsável por diferenças interculturais na experiência emocional é a categorização dos eventos antecedentes. De acordo com Mesquita e Frijda (1992), os eventos antecedentes são categorizados diferentemente por diferentes culturas e isso pode elicitare diferentes emoções para o mesmo evento. Semelhanças interculturais na categorização de eventos são encontradas, por exemplo, na antecipação de perigo, que leva à ansiedade e no flagrante de estar fazendo algo errado, que leva à vergonha (*shame*). Diferenças incluem, por exemplo, a situação de estar sozinho. Segundo Briggs (1970 *apud* Mesquita e Frijda, 1992), para o povo Utku Inuit, estar sozinho representa isolamento social e leva a uma tristeza profunda. Já os taitianos acreditam que estar sozinho pode fazer com que a pessoa seja perturbada por espíritos e esse fato provoca, nesse povo, a emoção do medo.

Semelhanças e diferenças interculturais também são encontradas nos estudos de expressividade de emoções na voz. A metodologia utilizada nesse tipo de pesquisa é muito semelhante àquela utilizada nas pesquisas com expressões faciais. Menezes *et al.* (2012) investigaram a percepção de emoções (tristeza, alegria e fala neutra) por ouvintes norte-americanos, japoneses e sul-coreanos em sentenças em inglês. O *corpus* foi dividido em duas modalidades: semântica, na qual as sentenças eram lidas com voz neutra (*i.e.* sem expressar nenhuma emoção), mas continham palavras relacionadas com as emoções e a modalidade afetiva (prosódica), na qual as sentenças eram lidas com voz afetiva, mas tinham um conteúdo semântico neutro. Como esperado, os ouvintes nativos (norte-americanos) reconheceram melhor as emoções, com uma taxa de acerto de 56% contra 45% dos sul-coreanos e 41% dos japoneses. Os ouvintes norte-americanos e japoneses foram melhores em reconhecer as emoções na modalidade afetiva (78% para norte-americanos e 46% contra 37% para os japoneses) e os sul-coreanos na modalidade linguística (50% contra 34%).

Scherer *et al.* (2001) conduziram um estudo envolvendo nove países (Alemanha, Suíça francófona, Inglaterra, Holanda, Estados Unidos, Itália, França, Espanha e Indonésia) utilizando sentenças sem sentido gravadas por quatro atores alemães, expressando alegria, tristeza, medo, raiva e fala neutra. Após a apresentação de cada estímulo, os ouvintes escolhiam uma (ou duas) das emoções apresentadas em uma lista. Todas as emoções, incluindo a fala neutra, foram reconhecidas por todos os países com taxa superior ao acaso. Entretanto, a emoção alegria foi a menos identificada, sendo frequentemente confundida com fala neutra. Ainda, houve variação entre os países quanto à taxa de reconhecimento. Os ouvintes alemães (os quais ouviram os estímulos em sua língua materna) foram os que acertaram mais, seguidos dos suíços, ingleses, holandeses, norte-americanos, italianos, franceses, espanhóis e, por último, os indonésios (observe aí um afastamento das culturas de língua germânica, passando pelas de língua neolatina para a de língua asiática).

Abelin e Allwood (2000) estudaram a percepção de emoções pela fala por falantes de sueco, inglês, finlandês e espanhol. Os estímulos utilizados nos experimentos de percepção consistiram de gravações de uma sentença de conteúdo semântico neutro feitas por um falante de sueco, expressando as seguintes emoções e atitudes: “alegria”, “tristeza”, “surpresa”, “medo”, “timidez”, “raiva”, “dominância” e “nojo”. As emoções “tristeza”, “raiva” e “surpresa” foram bem reconhecidas por todos os grupos de falantes. Entretanto, as emoções e atitudes “felicidade”, “timidez”, “nojo” e “dominância” apresentaram altas taxas de confusão. Os falantes de espanhol confundiram “felicidade” com “tristeza” e também “surpresa” com “raiva”. “Timidez” foi confundida por todos os grupos com “tristeza” e “medo”. A emoção “nojo” foi a mais problemática. Os falantes de espanhol e de finlandês tenderam a interpretá-la como “tristeza”. Já os falantes de sueco como “surpresa” ou “felicidade”. Após uma análise acústica e semântica, os autores concluíram que as emoções que foram mais confundidas são semelhantes tanto com relação aos parâmetros acústicos como também semanticamente.

1.7 EMOÇÕES REAIS VS. EMOÇÕES ATUADAS

Um grande problema metodológico que envolve a investigação da expressão de emoções na fala consiste na coleta do corpus a ser utilizado. O problema se dá pela dificuldade de se obter amostras de fala expressando emoções autênticas e intensas e também pelas questões éticas que emergem nesse tipo de coleta. Nesta seção discorrerei brevemente sobre os métodos mais empregados para a coleta de corpora nas pesquisas sobre expressão de emoções na fala.

Um dos métodos para obter amostras reais de fala emotiva é o método da **indução**, que consiste em induzir as emoções desejadas nos sujeitos por meio de procedimentos adequados, tais como a atribuição de tarefas a serem completadas sob pressão, jogos de *videogame* (Johnstone *et al.*, 2005), apresentação de filmes ou imagens, leitura de sentenças com palavras de forte conteúdo emocional (Wilting, Kraemer e Swerts, 2006) e outros. Entretanto, essa prática não é simples, pois a indução de emoções nos sujeitos envolve questões éticas. Obviamente, esses métodos só podem ser abordados por pesquisadores capacitados para lidar com o comportamento humano, uma vez que emoções verdadeiras estão sendo induzidas nos sujeitos, o que requer uma atenção maior durante e após o experimento. Esse método, porém, é pouco adotado devido também a outra desvantagem: as emoções obtidas geralmente são de baixa intensidade, o que dificulta a identificação de alterações acústicas na fala. Além disso, não há garantia de que todos os sujeitos experienciarão a mesma emoção para o método de indução adotado, pois, como já discutido, a experiência de uma emoção depende da avaliação do indivíduo sobre o evento, que é particular a cada pessoa.

O tipo de corpus mais utilizado é o de **emoções simuladas** (atuadas). Ele é obtido pedindo-se aos informantes (atores ou não) que produzam amostras de fala simulando uma determinada emoção. Em alguns casos, um roteiro correspondendo a uma determinada situação é dado aos sujeitos de modo a garantir que eles sigam o mesmo procedimento para simular as emoções. A simulação garante emoções muito mais intensas do que a indução. Além disso, ela permite ao pesquisador controlar o conteúdo semântico dos enunciados a serem gravados, de forma que não contenham pistas lexicais da emoção

expressa. Isso geralmente é feito através do uso de pseudopalavras, isto é, palavras sem nenhum sentido na língua dos sujeitos que participarão da pesquisa. Porém, a validade desse método é alvo de um grande debate, pois não se sabe até que ponto as emoções simuladas representam as emoções “reais” (Scherer, 2003). Ao simular uma emoção, o sujeito pode estar dando ênfase a características estereotipadas (expectativas socioculturais) e deixando de lado outras mais sutis. Além disso, muitas vezes as reações fisiológicas que causam as alterações corporais não estão presentes na simulação. As emoções simuladas são muito bem reconhecidas em testes de percepção (vide Banse e Scherer, 1996 para um exemplo), mas podem não ser eficazes se o objetivo da pesquisa for identificar parâmetros acústicos que sejam afetados pelas alterações fisiológicas causadas pelas emoções.

Em um dos poucos estudos que compararam experimentalmente a percepção de emoções reais e atuadas, Wilting, Kraemer e Swerts (2006) pediram a um grupo de sujeitos que lessem sentenças com palavras de forte conteúdo emocional (positivo e negativo) e a outro grupo que simulasse emoções positivas e negativas. Em seguida, os autores aplicaram um experimento de percepção com ambas as emoções simuladas e as induzidas no qual os sujeitos tinham que avaliar o grau de valência da emoção expressa em uma escala graduada de 1 a 7. Os autores constataram que as emoções atuadas foram avaliadas com um grau de valência maior do que as emoções induzidas (reais). Isso, segundo os autores, indica que os atores podem exagerar ao simular as emoções, em comparação com as emoções induzidas.

Diante dessas questões, adotamos nesta pesquisa um terceiro método, que consiste em obter as amostras de fala de **situações reais** nas quais os participantes expressaram emoções. As possibilidades incluem programas de televisão (entrevistas, *talk shows*, *reality shows*, *game shows*, debates etc.), trotes telefônicos (Barbosa, 2009), diálogos entre pilotos de avião, ligações telefônicas de emergência para a polícia (Demenko e Jastrzębska, 2012) e para *call centers*, psicoterapias e muitas outras. Essa forma de obter o corpus garante emoções expressas em situações reais e bastante intensas. Uma desvantagem é que muitas vezes as gravações não são realizadas em um ambiente ideal e com isso podem apresentar um alto grau de ruído, o que pode comprometer as análises acústicas. Por esse motivo, a qualidade das gravações é um fator que merece atenção e deve ser verificada *a priori*.

Capítulo 2

Expressão de emoções na fala: Conceitos básicos

Como vimos no capítulo 1, a avaliação de um evento como sendo de relevância para o indivíduo causa, entre outras reações, respostas fisiológicas em seu corpo, tais como alterações no sistema nervoso somático e no sistema nervoso autônomo, que são responsáveis pelo controle voluntário dos músculos e pelo controle da respiração e da circulação sanguínea, respectivamente. Essas alterações, por sua vez, afetam a respiração, fonação e articulação, que são processos diretamente relacionados à fala (Scherer, 1986). Portanto, para compreender como as emoções afetam a fala, é importante primeiramente entender como esta é produzida pelo falante. É com esse objetivo que este capítulo apresenta alguns conceitos básicos, porém importantes, sobre fonética acústica e articulatória, como os órgãos envolvidos na produção da fala e as propriedades físicas do sinal de fala que são de interesse para as pesquisas sobre a expressão vocal de emoções.

2.1 O MECANISMO DE PRODUÇÃO DA FALA

O ramo da fonética, a ciência que estuda os sons das línguas, responsável por estudar como os sons são produzidos pelos órgãos do aparelho fonador e os mecanismos fisiológicos dessa produção é a fonética articulatória. Neste trabalho não entraremos em detalhes como a classificação dos sons da fala, por exemplo. Estaremos focados nos aspectos teóricos que permitirão ao leitor entender o fundamento por trás das medidas acústicas realizadas na pesquisa.

Os órgãos envolvidos na produção da fala são divididos em três sistemas, conforme Abercrombie (1967; ver também Kent, 1997): sistema respiratório, que agrupa os pulmões, os músculos pulmonares, os brônquios e a traqueia; sistema fonatório, onde se localiza a laringe e o sistema articulatório, que agrupa o nariz, a língua, os dentes e os lábios.

Os sons das línguas são normalmente produzidos com corrente de ar egressiva (aquela em que o ar é expelido dos pulmões), entretanto existem línguas que usam também sons ingressivos (realizados pelo processo de inspiração). Na expiração, o ar é expelido para fora dos pulmões pela ação do diafragma, um músculo em forma de cúpula localizado logo abaixo dos pulmões, pela diminuição da caixa torácica e pelo relaxamento dos músculos intercostais. Ele passa pela traqueia e chega à laringe, onde se localizam as pregas vocais, que são músculos estriados com a capacidade de adução, movimento que fecha a glote, bloqueando a passagem do ar. A glote é o espaço localizado entre as pregas vocais. Quando a pressão abaixo da glote (pressão subglótica) for maior do que a pressão acima dela (supraglótica), as pregas vocais se separam (abdução) e uma camada de ar segue à faringe. Logo após a abdução das pregas vocais, a pressão subglótica cai (efeito de Bernoulli), e as pregas se fecham novamente, de forma mais rápida do que na sua abertura, gerando a onda glotal. A pressão subglótica afeta diretamente o tom e o volume da voz (Marasek, 1997). Desse modo, a laringe é usada para gerar sons audíveis, transformando a energia potencial do ar comprimido abaixo da glote em energia cinética (a corrente de ar). Esse processo é conhecido como fonação. Os sons vozeados da fala (produzidos com vibração das pregas vocais) consistem desses pulsos glotais. A região acima da glote

(região supra-laríngea) é chamada de trato vocal e compreende a faringe, a cavidade oral e a cavidade nasal.

A faringe se bifurca, dando origem à cavidade oral e acima, à cavidade nasal. A abertura da cavidade nasal é controlada pela posição do palato mole. Quando a parte posterior do palato mole estiver levantada, ela estará fechando a passagem para a cavidade nasal (passagem velofaríngea), impedindo o acesso do ar e gerando um som oral. Portanto, os sons nasais são produzidos com o palato mole abaixado. Segundo Stevens (1998), a área da seção transversal da passagem velofaríngea pode chegar a 1,0 cm², mas durante a produção de sons nasais ela fica entre 0,2 e 0,8 cm².

Os diferentes sons possíveis nas línguas naturais são produzidos por diferentes configurações de articulação no trato vocal, determinadas, sobretudo, pelas posições dos lábios, língua, úvula e também pela condição das pregas vocais (com ou sem vibração).

A frequência de vibração das pregas vocais é um parâmetro muito importante, pois determina a frequência fundamental da voz, que é percebida como tom (*pitch*). Essa frequência é inversamente proporcional à massa das pregas vocais e diretamente proporcional à tensão delas (Marasek, 1997). A massa das pregas vocais das mulheres é, geralmente, menor do que a dos homens. É por isso que elas possuem uma voz mais aguda (frequência fundamental maior) do que os homens.

O aumento da tensão das pregas, que leva ao aumento da frequência de vibração, é alcançado através da contração do músculo cricotireoideo e também possivelmente através da contração do músculo cricoaritenoidio posterior (Stevens, 1998). A situação oposta, de decréscimo da frequência de vibração devido à redução da tensão das pregas vocais, é causada pela contração do músculo vocalis, que leva ao encurtamento e espessamento das pregas. O encurtamento das pregas vocais e, com isso, a redução da tensão nelas, pode também ser alcançado pelo abaixamento da laringe (Stevens, 1998).

2.2 OS ASPECTOS FÍSICOS DA FALA

A área da fonética que se ocupa do estudo das propriedades físicas que envolvem a fala é conhecida como fonética acústica.

A fala é transmitida através de ondas sonoras que se propagam em um meio elástico, como o ar. Uma onda sonora consiste em flutuações de pressão que fazem com que as moléculas de ar (assumindo que o meio de propagação é o ar) vibrem. Quando um objeto vibra (as pregas vocais, por exemplo), ele empurra as moléculas de ar que se encontram perto dele. Essas moléculas por sua vez empurram outras moléculas, e assim sucessivamente. Quando as moléculas condensam (colidem) há um aumento de pressão, e quando se separam (rarefação), há um decréscimo da pressão. Portanto, é a flutuação da pressão que se propaga, e não as moléculas de ar (Johnson, 1997). Quando essa flutuação de pressão alcança o tímpano, este vibra, e essas vibrações são traduzidas pelo sistema auditivo em pulsos neurais, que percebemos como som.

Uma onda periódica simples (o tipo mais básico de uma onda na qual um padrão, o ciclo, se repete) é caracterizada por três variáveis: amplitude, frequência e tempo. A amplitude, indicada no eixo vertical da figura 2.1, reflete a perturbação da pressão do meio elástico em que a onda se propaga. Quanto maior a pressão, maior a amplitude da onda. A frequência corresponde ao número de ciclos completos que a onda realiza por segundo e sua unidade de medida é o Hertz (Hz). O tempo que a onda leva para realizar um ciclo é o período (T) da onda, medido em segundos (s).

Ao plotarmos as séries de condensações e rarefações do ar em um gráfico, o resultado é uma senoide, esquematizada na figura 2.1. Porém, o sinal de fala é composto por muitas senoides, o que o caracteriza como uma onda complexa. No caso de sons vozeados (produzidos com vibração das pregas vocais), essa onda é uma complexa quase periódica. A frequência fundamental também corresponde à taxa de repetição do ciclo da onda complexa periódica por segundo.

Um teorema matemático, chamado de teorema de Fourier, afirma que qualquer onda periódica consiste em uma soma de senoides. Portanto, é possível assumir que a onda periódica complexa dos sons vozeados é constituída de senoides. A figura 2.2 mostra a

forma de onda da vogal [a] da primeira sílaba da palavra “passado”, pronunciada por um falante masculino do português brasileiro. Nessa figura, podemos observar 10 ciclos. Graças ao teorema de Fourier, é possível decompor uma onda periódica complexa nas diversas ondas senoidais simples que a compõem (processo conhecido como análise de Fourier). Se plotarmos essas componentes com relação às suas respectivas frequências e amplitudes (nível de pressão sonora em decibéis/Hz) em um gráfico, obtemos o chamado espectro FFT (*Fast Fourier transform*), como na figura 2.3.

Outra ferramenta muito útil para a análise acústica da fala é o espectrograma, criado no final da década de 1940 para representar visualmente três características da fala simultaneamente: tempo, frequência e amplitude (Fujimura e Erickson, 1999). Nesse gráfico, o eixo horizontal representa o tempo, o vertical a frequência e a amplitude do sinal para um determinado ponto (de uma dada frequência em um determinado período tempo) é representada pelo grau de cinza naquele ponto. A figura 2.4 mostra um espectrograma para o trecho “(n)unca andava de” retirado da sentença “Manuel nunca andava de olhos baixos”, pronunciada por uma falante feminina do português brasileiro. O trecho está segmentado em unidades V-V, que é uma unidade do tamanho da sílaba, composta pelos segmentos delimitados por dois *onsets* de vogais consecutivos, incluindo a vogal à esquerda.

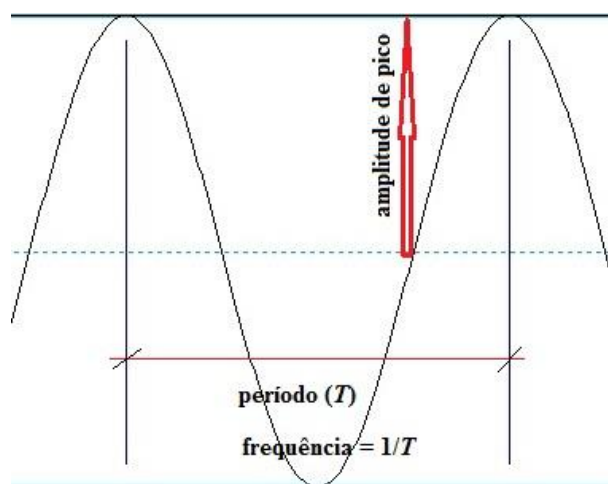


Figura 2.1: Ciclo de uma senoide (delimitado pelas barras verticais).

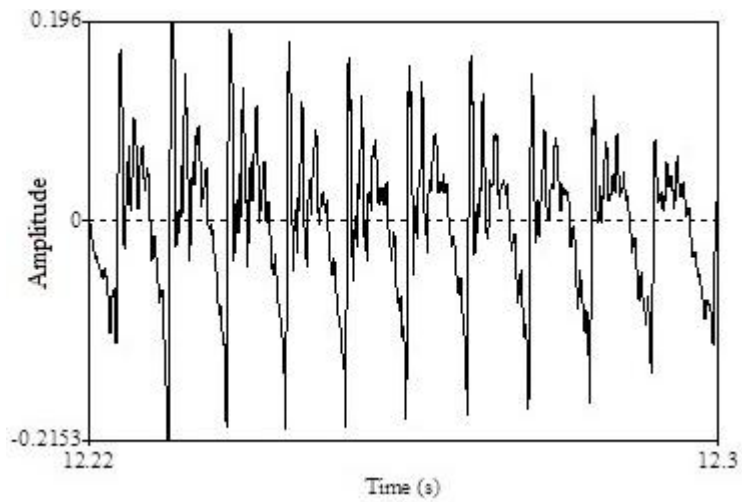


Figura 2.2: Forma de onda da vogal [a] da primeira sílaba da palavra “passado”, pronunciada por um falante masculino do português brasileiro. O eixo horizontal corresponde ao tempo (em segundos) e o vertical à amplitude da onda (pressão sonora).

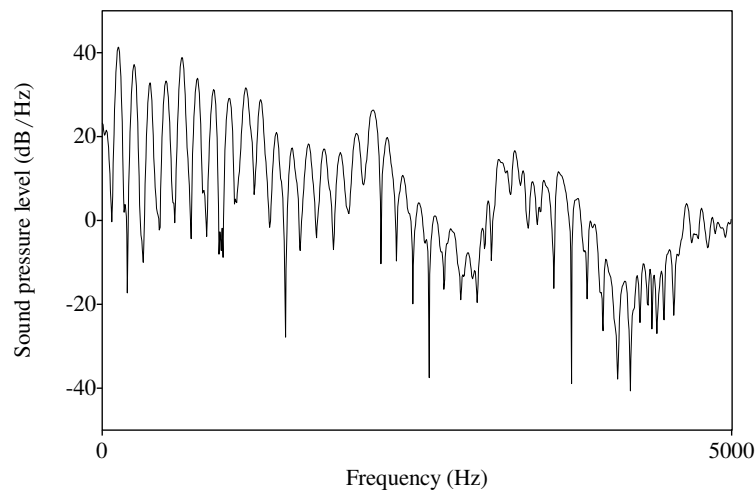


Figura 2.3: Espectro FFT para a vogal [a] da primeira sílaba da palavra “passado” pronunciada por um falante masculino do português brasileiro. O eixo horizontal corresponde às frequências (em Hz) e o vertical ao nível de pressão sonora (em dB/Hz).

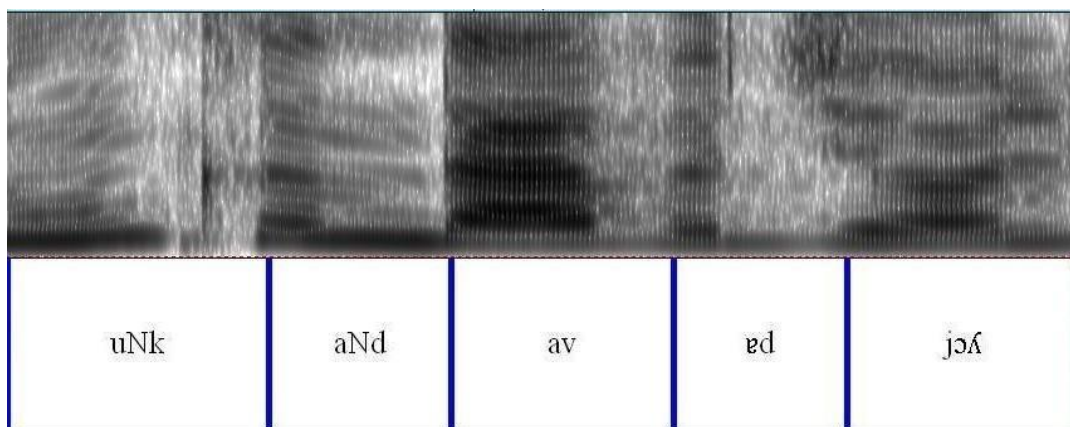


Figura 2.4: Espectrograma (eixo vertical: frequência em hertz; eixo horizontal: tempo em segundos) para o trecho “(n)unca andava de” retirado da sentença “Manuel nunca andava de olhos baixos”, pronunciada por uma falante feminina do português brasileiro. A parte inferior da figura mostra a segmentação do trecho em unidades V-V (vide texto para explicação).

Da perspectiva fonético-acústica, a produção da fala pode ser entendida como o resultado da passagem do som gerado por uma (ou várias) fonte sonora por um filtro. Essa abordagem é conhecida como **teoria fonte-filtro da produção da fala** (Fant, 1960). O filtro do qual trata essa teoria é o trato vocal. A fonte sonora corresponde às pregas vocais no caso das vogais e constrictões no trato vocal no caso de consoantes, como as fricativas, por exemplo.

Assim como todos os corpos materiais, o trato vocal possui frequências naturais de vibração, que dependem, principalmente, de seu comprimento. As ondas sonoras geradas pela fonte cujas frequências coincidirem com as frequências naturais de vibração de uma determinada configuração articulatória do trato vocal entrarão em ressonância e terão sua amplitude amplificada, gerando os **formantes**. A língua é utilizada para produzir constrictões no trato vocal e, juntamente com a configuração dos lábios (o arredondamento dos lábios abaixa a frequência de todos os formantes) e a posição da laringe, determina diferentes comprimentos para o trato vocal. Como as frequências naturais de vibração do trato vocal dependem do comprimento deste, as diferentes configurações articulatórias do trato vocal determinam formantes em frequências diferentes. São, portanto, essas

configurações formânticas que distinguem as vogais. Por serem regiões onde a energia é relativamente maior, os formantes podem ser identificados visualmente no espectrograma como “faixas” mais escuras. Teoricamente, existem infinitos formantes, mas os mais importantes são os três primeiros (aqueles de frequências menores), indicados respectivamente por F1, F2 e F3. Os dois primeiros formantes (F1 e F2) são correlatos da altura da língua e do avanço horizontal dela no trato vocal, respectivamente. Por isso, quanto mais alta uma vogal for, menor será seu F1 e quanto mais anterior a vogal, maior seu F2.

Os formantes de frequências maiores (F3, F4 etc.) não variam significativamente com a qualidade vocálica. Eles parecem estar mais relacionados com as características particulares do trato vocal dos falantes. Portanto, esses formantes são potenciais correlatos do estado afetivo do falante (Pittam e Scherer, 1993). De fato, no estudo conduzido por Laukkanen *et al.* (1997), a percepção pelos ouvintes da dimensão de valência em sentenças emotivas sem conteúdo semântico e com a curva melódica (curva de f_0) normalizada correlacionou-se com o primeiro e quarto formantes (F1 e F4). De acordo com esses autores, um F1 muito baixo ou alto foi percebido como valência negativa, enquanto que F3 e F4 altos contribuíram para a percepção de valência positiva.

Salvo em condições especiais, o filtro é independente da fonte. Isso significa que as mudanças na taxa de vibração das pregas vocais não influenciam demasiado as frequências dos formantes, pelo menos num modelo de primeira ordem. É por isso que até mesmo a fala produzida com uma voz sussurrada é inteligível e que os falantes podem produzir as vogais com diferentes tons sem perder as características perceptuais das mesmas (Kent e Read, 1992).

A onda complexa produzida pelas pregas vocais em vibração corresponde física e matematicamente a infinitas ondas sonoras simples cujas frequências são múltiplos inteiros da frequência fundamental. Por exemplo, para uma mulher adulta com uma frequência fundamental de 230 Hz, essas ondas terão frequências de 230 Hz (f_0), 460 Hz, 690 Hz, 920 Hz e assim sucessivamente. Essas ondas são chamadas de harmônicos e a energia delas decai com o aumento da frequência, a uma taxa de 12 dB por oitava, isto é, quando a frequência dobra, a energia decresce em 12 dB. Esse espectro é ilustrado pela

figura 2.5. Portanto, a maior parte da energia dos sons vozeados concentra-se nas frequências mais baixas do espectro. Sabe-se que o aumento do esforço vocal resulta em um aumento da energia nas altas frequências do espectro do falante (Traunmüller e Eriksson, 2000).

Segundo Laukkanen *et al.* (1997: 161; tradução minha), “o esforço vocal refere-se à sensação ou impressão subjetiva da quantidade de energia usada para a produção da voz. Um esforço vocal alto está relacionado à uma alta pressão subglótica e, normalmente, a uma forte adução das pregas vocais.”⁴

O fato de as mulheres e as crianças terem uma frequência fundamental média relativamente elevada incorre em um maior espaçamento entre os harmônicos e como consequência há poucos harmônicos para certa banda de frequências. Isso dificulta a mensuração de alguns parâmetros, como os formantes. Por isso, a maioria dos estudos em fonética acústica é conduzida apenas com falantes masculinos adultos (Kent e Read, 1992; Scherer, 1982).

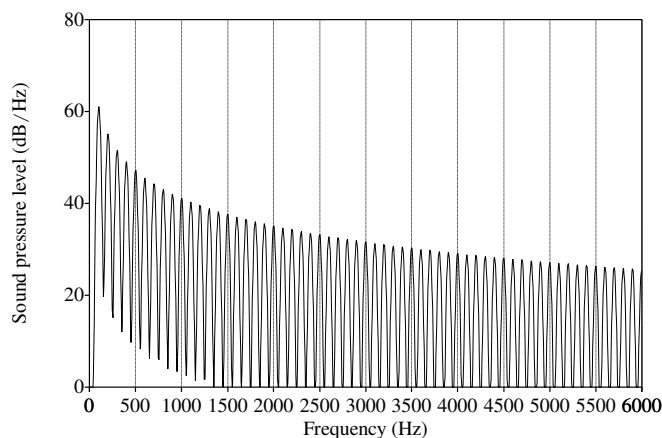


Figura 2.5: Espectro FFT das ondas sonoras produzidas pela fonte (a vibração das pregas vocais), para uma f_0 igual a 100 Hz. Explicação no texto.

⁴ “Vocal effort refers to the subjective sensation or impression of the amount of energy used for voice production. High effort level is related to high subglottic pressure and usually to a strong adduction of the vocal folds.”

2.3 CORRELATOS FONÉTICO-ACÚSTICOS DAS EMOÇÕES

Para investigar as alterações que ocorrem na fala devido à influência das emoções através de análise acústica, usam-se medidas acústicas das propriedades físicas da fala, como a intensidade, duração, frequência fundamental, entre outras. Neste trabalho, nos referimos a essas medidas como “parâmetros acústicos”. Por ser uma análise indireta, essa metodologia oferece a vantagem de ser totalmente não invasiva, já que nenhum equipamento precisa ser acoplado ao falante. Os principais parâmetros acústicos são descritos a seguir.

- **Frequência fundamental**

Como já mencionado anteriormente neste capítulo, a frequência fundamental (f_0) é um correlato acústico da taxa de vibração das pregas vocais e é percebida como o tom (*pitch*) da voz. Ela pode ser identificada no sinal acústico da fala como o número de ciclos da onda complexa por segundo ou como o inverso do período (T) da onda ($1/T$) e é medida em Hertz (Hz).

Ela é um parâmetro prosódico muito importante para a fonética, pois permite aos falantes distinguir uma asserção de uma interrogação, assinalar proeminências, produzir e identificar diversas atitudes (ironia, sarcasmo, incredulidade, certeza etc.), marcar fronteiras entoacionais (Barbosa, 2012), distinguir palavras em línguas tonais (como o mandarim e o cantonês) etc. Além disso, a frequência fundamental tem se mostrado muito importante para a expressividade de emoções na fala, tendo se comportado de forma consistente para características da personalidade e variações no estado emocional do falante (Scherer, 1982).

Os descritores estatísticos mais frequentemente computados para a frequência fundamental são: *baseline* (menor valor no enunciado ou trecho), pico (valor máximo atingido), média ou mediana, variabilidade (desvio-padrão, variância), assimetria e a gama ou amplitude (f_0 máximo – f_0 mínimo ou amplitude entre quartis).

O uso da mediana e da (semi-) amplitude entre quartis é preferível ao uso da média e da diferença entre f_0 máximo e f_0 mínimo, pois estas últimas medidas são mais

sensíveis a erros de medida, já que levam em consideração os extremos dos dados. Entretanto, elas ainda são mais comuns na literatura. Ao se dividir os dados em grupos de $\frac{1}{4}$ do total, cada um com 25% dos dados, o último valor do primeiro desses grupos é o primeiro quartil (Q_1), ou seja, o valor que divide $\frac{1}{4}$ dos dados à esquerda e o restante à direita. O segundo quartil é o valor que divide os dados em duas metades, chamado de mediana. O terceiro quartil (Q_3) é o último valor do terceiro grupo, que deixa $\frac{3}{4}$ dos dados para trás e, por fim, o quarto quartil corresponde ao maior valor da lista. A amplitude entre quartis é, portanto, a diferença entre o terceiro e o primeiro quartis ($Q_3 - Q_1$). A semi-amplitude entre quartis é a metade da amplitude entre quartis.

A assimetria é um descritor estatístico que permite distinguir uma distribuição simétrica (aquela cuja média é igual à mediana e à moda, para uma distribuição unimodal) de uma assimétrica. Se assimétrica (assimetria diferente de zero), a distribuição pode ter assimetria negativa (quando a cauda do lado esquerdo da distribuição for mais longa do que a do lado direito) ou assimetria positiva (quando a cauda do lado direito for mais longa do que a do lado esquerdo). Desse modo, a assimetria nos permite saber se, no caso de uma distribuição assimétrica, o desvio da média é positivo ou negativo.

Outra medida relacionada à frequência fundamental é o *jitter*, que mede variações na periodicidade da onda, ciclo-a-ciclo, revelando possíveis irregularidades na vibração das pregas vocais. Segundo Scherer (1982), é possível que exista uma relação entre *jitter* e ativação emocional.

É possível ainda computar variações ciclo-a-ciclo na amplitude da forma de onda, por meio de uma medida denominada *shimmer*.

As figuras 2.6 e 2.7 correspondem a um trecho da forma de onda para a vogal [a] (em palavras diferentes) pronunciada por uma falante feminina adulta do português brasileiro de nosso corpus “Jogo de Cena” expressando tristeza (Fig. 2.6) e alegria (Fig. 2.7). Pode-se observar que na fig. 2.6 existem muito mais variações no período dos ciclos e na amplitude da onda do que na fig. 2.7. As medidas de *jitter* e *shimmer* computadas para esses trechos por meio do programa Praat (Boersma e Weenink, 2011) revelam isso: no trecho da fig. 2.6 foram medidos 2,134% para Jitter (rap) e 10,214% para Shimmer (apq3), enquanto que no trecho da fig. 2.7 foram medidos 0,244% para Jitter (rap) e 1,753% para

Shimmer (apq3). As siglas “rap” e “apq3” referem-se ao modo como esses parâmetros foram calculados: rap = “*Relative Average Perturbation*” e apq3 = “*three-point Amplitude Perturbation Quotient*”. Apesar de no enunciado correspondente à figura 2.6 a falante estar expressando uma emoção com menor grau de ativação (a tristeza), os valores de *jitter* e *shimmer* foram consideravelmente maiores nesse enunciado do que no enunciado da figura 2.7. Isso se deve muito provavelmente à qualidade de voz da falante nesse enunciado, pois ela estava quase chorando, o que pode ter ocasionado as elevadas perturbações na periodicidade e na amplitude da onda mostrada na figura 2.6.

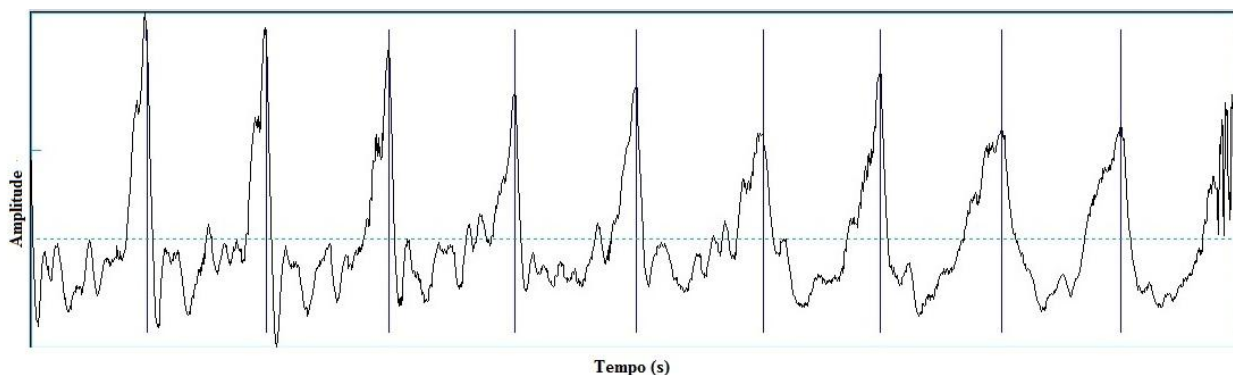


Figura 2.6: Trecho da forma de onda da vogal [a] de uma falante feminina do português brasileiro de nosso corpus “Jogo de Cena” expressando tristeza (enunciado 10). Jitter (rap): 2,134%; Shimmer (apq3): 10,214%. Os ciclos da onda estão separados por barras verticais.

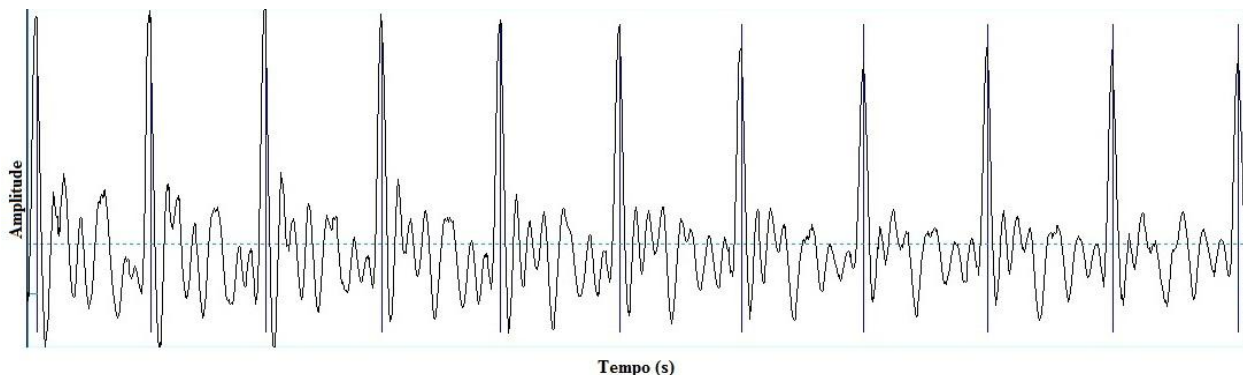


Figura 2.7: Trecho da forma de onda da vogal [a] da mesma falante da figura 2.6 expressando alegria (enunciado 28). Jitter (rap): 0,244%; Shimmer (apq3): 1,753%. Os ciclos da onda estão separados por barras verticais.

- **Intensidade**

A intensidade sonora reflete melhor nossa sensação de volume (*loudness*) do que a amplitude da onda (Johnson, 1997) e mede a quantidade de energia no sinal de fala. A intensidade pode ser obtida pelo cálculo da raiz quadrada média (*root mean square – RMS*) da amplitude de pico da onda, isto é, multiplicando a amplitude de pico por 0,707. Esse parâmetro é normalmente medido em decibéis (dB), que corresponde a uma escala logarítmica.

Os descritores estatísticos mais comumente usados para a intensidade (*I*) são: valor máximo de intensidade no som, média e assimetria.

Outro parâmetro muito importante que envolve a intensidade é a **inclinação espectral**, que mede a distribuição da energia no espectro de frequências. Como mencionado anteriormente, a maior parte da energia dos sons vozeados se concentra nas frequências mais baixas. A inclinação espectral mede o grau de decaimento da energia com o aumento da frequência e pode ser calculada pela diferença entre a amplitude da frequência fundamental (o primeiro harmônico) e a dos harmônicos de frequências maiores (Gordon e Ladefoged, 2001). Essa medida é, portanto, um correlato acústico do esforço vocal, pois o aumento do esforço vocal aumenta a energia nas altas frequências fazendo com que a diferença entre a amplitude da frequência fundamental e dos harmônicos de

frequências maiores seja menor. Além disso, a inclinação espectral também é útil para diferenciar diferentes tipos de fonação. Gordon e Ladefoged (2001), por exemplo, constataram que na língua Zapoteca de San Lucas Quiaviní (no México) a amplitude do segundo harmônico para a vogal [a] é maior do que a amplitude da frequência fundamental para a voz crepitante, enquanto que para a voz soprosa essa amplitude é consideravelmente menor (menor ainda do que seria esperado para a voz modal).

A **ênfase espectral** também é uma medida da distribuição da energia no espectro. Porém, ao invés de computar diferenças na amplitude entre harmônicos, ela calcula diferenças na intensidade entre bandas de frequência, como, por exemplo, a diferença da intensidade média entre toda a banda de frequências da amostra de fala e uma banda baixa (de frequência máxima igual a 1,5 vezes a média de f_0) (Traunmüller e Eriksson, 2000).

- **Parâmetros temporais**

O número de segmentos que o falante produz em um determinado período de tempo também pode ser um correlato de seu estado emocional. Normalmente essa medida é feita em termos de sílabas fonéticas por segundo e caracteriza a **taxa de elocução** do falante. Ela consiste da razão entre o número de sílabas fonéticas em um enunciado (ou trecho) e a duração total em segundos do enunciado (ou do trecho). Há evidências de que o medo, a raiva e a alegria, por exemplo, são frequentemente expressos com uma taxa de elocução maior do que a habitual enquanto que a tristeza e o tédio são expressos com uma taxa de elocução mais baixa (Scherer, 2003).

Outro parâmetro relacionado ao tempo, que pode ser útil para caracterizar o estado emocional do falante, diz respeito às **pausas**, que correspondem a períodos no enunciado em que há interrupção da fala, podendo ser silenciosas ou não (Oh, 2010). Emoções como a tristeza, caracterizadas por baixo grau de ativação e prazer intrínseco negativo, podem apresentar pausas relativamente longas, bem como grande quantidade de hesitações na fala. Isso também ocorre com emoções que prejudicam o planejamento da fala, como a ansiedade (Johnstone e Scherer, 2000). Esse parâmetro pode ser quantificado

através da proporção de período de pausa em relação à duração total do enunciado ou trecho.

- **Espectro médio de longo-termo**

O espectro médio de longo-termo (LTAS - *Long-Term Average Spectrum*) é um tipo de espectro como o da figura 2.3, mas obtido a partir da média de vários espectros extraídos do trecho de fala para uma determinada gama de frequências. Esses espectros são divididos em um número igual de pontos equidistantes na gama de frequências escolhida e então a média da amplitude em cada um desses pontos é calculada, gerando o LTAS (Pittam e Scherer, 1993).

A utilidade do LTAS está no fato de que ele representa as características mais estáveis da voz do falante em uma determinada amostra de fala, pois as variações no espectro relacionadas a segmentos linguísticos individuais (fones) são “diluídas” quando a média de vários espectros é computada (Scherer, 1982).

A figura 2.8 mostra dois espectros médios de longo-termo de uma falante feminina adulta de nosso corpus sueco. No trecho correspondente ao LTAS representado em vermelho a falante expressa um maior grau de raiva do que no trecho correspondente ao LTAS traçado em preto. Pode-se observar um aumento na energia das frequências mais elevadas do LTAS vermelho, especialmente em torno dos 3000 Hz e 8000 Hz (indicados na figura por uma seta). Isso indica um maior esforço vocal e uma possível alteração na qualidade de voz dessa falante no trecho de fala correspondente ao LTAS representado em vermelho.

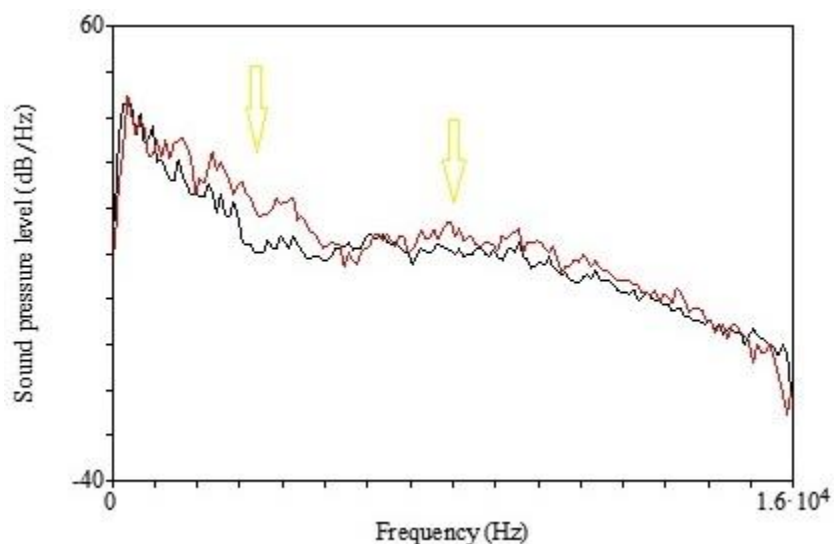


Figura 2.8: Dois espectros médios de longo-termo de uma falante feminina adulta de nosso corpus sueco. No trecho correspondente ao espectro traçado em vermelho a falante expressa um maior grau de raiva do que no trecho correspondente ao espectro traçado em preto. As setas indicam pontos em que houve um aumento considerável de energia no espectro vermelho com relação ao espectro preto.

2.4 CARACTERÍSTICAS FONÉTICO-ACÚSTICAS DE ALGUMAS EMOÇÕES

Esta seção apresenta um resumo, com base em Johnstone e Scherer (2000) e Scherer (2003), das alterações em alguns parâmetros acústicos encontradas em vários estudos para algumas emoções discretas (Tabela 2.1).

O único parâmetro mencionado por Johnstone e Scherer (2000) para a emoção nojo é a média da frequência fundamental (f_0). Ainda assim, os resultados são inconsistentes, pois alguns estudos constataram uma diminuição do valor para esse parâmetro enquanto que outros constataram um aumento para a medida desse mesmo parâmetro. A razão dessa inconsistência, segundo Johnstone e Scherer (2000), pode estar relacionada ao método utilizado nos estudos para obter amostras de fala expressando a

emoção nojo: os estudos que utilizaram emoções atuadas constataram uma diminuição no valor da média de f_0 (e.g. Banse e Scherer, 1996) enquanto que os estudos que induziram o nojo nos sujeitos falantes através de filmes desagradáveis constataram um aumento de valor para esse parâmetro. Além da inconsistência na produção da fala, o nojo também apresentou taxas de reconhecimento muito baixas em muitos estudos, o que indica que essa emoção não é tão bem expressa pela fala como o é através de expressões faciais (Johnstone e Scherer, 2000).

Segundo Scherer (2003), os parâmetros f_0 , energia e taxa de elocução podem estar mais relacionados ao grau da dimensão de ativação, enquanto que o prazer intrínseco à emoção (valência) teria um impacto maior nas características fonatórias e articulatórias. Um estudo conduzido por Johnstone *et al.* (2005), que usou emoções induzidas através de um jogo de computador, corroborou essa hipótese. Os autores desse estudo constataram que as variações no prazer intrínseco de um evento produzem alterações na distribuição espectral da energia: a proporção de energia abaixo de 500 Hz foi significativamente menor para os enunciados que mostraram desagrado do que para os que mostraram agrado, com os primeiros exibindo uma maior concentração de energia nas altas frequências do espectro. Esses resultados indicam que há uma relação entre o prazer intrínseco do evento e a distribuição da energia no espectro de frequências. Além disso, alterações na f_0 e intensidade média se mostraram relacionadas à dimensão de ativação. Ainda, no estudo de Laukkanen *et al.* (1997), a percepção da dimensão de valência pelos sujeitos ouvintes foi baseada nas posições do primeiro e quarto formantes, bem como no tipo de fonação utilizado nas amostras de fala.

Pode-se concluir, pela tabela 2.1, que esses parâmetros acústicos não são suficientes para discriminar satisfatoriamente essas seis emoções. As emoções raiva e alegria, por exemplo, apresentam aumento para todos os parâmetros. Uma explicação para isso pode residir no fato de que essas duas emoções apresentam grau semelhante de ativação (Pereira, 2000). Percebe-se aqui, a necessidade de mais estudos que investiguem outros parâmetros acústicos além daqueles mostrados na tabela 2.1, especialmente parâmetros relacionados a aspectos fonatórios e à qualidade de voz (Scherer, 2003).

Tabela 2.1: Resumo de alterações em alguns parâmetros acústicos para seis emoções. Elaborada com base em Johnstone e Scherer (2000) e Scherer (2003). As células em branco correspondem a resultados não disponíveis.

	Nojo	Raiva	Medo	Tristeza	Alegria	Tédio
Intensidade		↑	↑	↓	↑	
f_0 baseline / média	↓*	↑	↑	↓	↑	
Variabilidade de f_0		↑		↓	↑	↓
Amplitude de f_0		↑	↑*	↓	↑	↓
Energia nas altas frequências		↑	↑	↓	↑*	
Taxa de elocução		↑	↑	↓	↑*	↓

Legenda: ↑ = aumento; ↓ = diminuição; * = resultado inconsistente ou contraditório.

2.5 O SCRIPT “EXPRESSION EVALUATOR”

Os parâmetros acústicos estudados nesta pesquisa foram extraídos automaticamente pelo “*Expression Evaluator*”, um *script* para o *software* Praat (Boersma e Weenink, 2011) implementado por Barbosa (2009; 2013). As classes de parâmetros medidas por esse *script* e usadas aqui são: frequência fundamental (f_0), primeira derivada de f_0 (df_0), intensidade global, inclinação espectral e espectro médio de longo-termo (LTAS). Dessas classes, foram calculados pelo *script* os seguintes descritores estatísticos (totalizando doze parâmetros):

f_0 : mediana, semiamplitude entre quartis, quantil 99,5% e assimetria;

df_0 : média, desvio-padrão e assimetria;

intensidade global: assimetria;

inclinação espectral: média, desvio-padrão e assimetria;

LTAS: inclinação.

Para usar esse *script* é necessário, além do arquivo de áudio, um arquivo no formato “textgrid” que contenha a delimitação do intervalo a ser considerado para a extração dos parâmetros. No nosso caso, esse intervalo correspondeu à duração total do trecho de fala. Esses dois arquivos devem estar no mesmo diretório (pasta). Ao executar o *script*, deve-se informar o nome dos arquivos, bem como o sexo do falante. O *script*, então, cria um arquivo de texto (.txt) com os valores para os parâmetros acústicos. Esse arquivo pode ser facilmente importado no programa Excel.

A curva de f_0 é extraída dentro do limite de 110 a 700 Hz (parâmetros “*left Cut off frequency*” e “*right Cut off frequency*” no Praat, respectivamente, para o caso de falantes femininos) e, em seguida, é suavizada por meio de um filtro LP de 10 Hz.

Os parâmetros relativos à frequência fundamental são normalizados pelo *script* por meio do *z-score*, um cálculo estatístico que mede, em unidades de desvio-padrão, o afastamento do valor medido em relação a uma média do mesmo, de acordo com a equação 2.1, em que x corresponde ao valor medido, μ à média de referência e σ ao desvio-padrão de referência. O *script* adota os seguintes valores de referência de frequência fundamental para as mulheres (falantes adultos): média de 231 Hz e desvio-padrão de 120 Hz. Por exemplo, supondo que o valor medido para o parâmetro “mediana de f_0 ” foi 255 Hz, então o valor considerado para análise será $(255 - 231) / 120 = 0,2$.

$$z - score = \frac{x - \mu}{\sigma} \quad (2.1)$$

Os quantis são valores que dividem os dados. Por exemplo, o quantil 50% divide os dados ao meio e é chamado de mediana. O quantil 25% divide os dados em quatro grupos e é chamado de quartil, como já explicado na seção 2.3. O *script* “*Expression Evaluator*” extrai o quantil 99,5% da frequência fundamental como alternativa ao valor máximo desse parâmetro a fim de evitar um possível erro do algoritmo usado pelo Praat para a extração de f_0 , como também já foi discutido na seção 2.3.

A semi-amplitude entre quartis é calculada aqui pela diferença entre o quantil 95% e o quantil 5%, dividida por 2.

A assimetria da frequência fundamental (*assimetriaF0*) é então calculada pela equação 2.2, em que *semiampquartF0* é a semi-amplitude entre quartis da frequência fundamental.

$$assimetriaF0 = \frac{médiaF0 - medianaF0}{semiampquartF0} \quad (2.2)$$

A primeira derivada de f_0 é um parâmetro que capta mudanças abruptas no contorno melódico (curva de f_0), sendo, por isso, capaz de identificar mudanças na expressividade da fala (Barbosa, 2009).

A assimetria da intensidade (*I*) é calculada pela equação 2.3, em que *desvpadI* é o desvio-padrão da intensidade no trecho de fala.

$$assimetriaI = \frac{médiaI - medianaI}{desvpadI} \quad (2.3)$$

A inclinação espectral é computada como uma diferença de intensidade (em dB) entre as bandas de 0 – 1250 Hz e 1250 – 4000 Hz e normalizada por meio da divisão desse valor pela mediana da intensidade de todo o som.

Por fim, a inclinação do LTAS foi obtida pela diferença de intensidade entre as bandas de 0 – 1000 Hz e 1000 – 4000 Hz e em seguida normalizada pelo *script* pela divisão desse valor por 10, apenas por razões de escala.

PARTE II
METODOLOGIA, ANÁLISES E
RESULTADOS

Os dois experimentos de percepção descritos nesta dissertação foram conduzidos a fim de testar as seguintes hipóteses:

1. Os sujeitos ouvintes são capazes de inferir o estado emocional do falante através da fala;
2. Pelo menos um dos doze parâmetros acústicos computados para cada trecho de fala de nossos dois corpora é capaz de explicar satisfatoriamente as respostas dos sujeitos nos experimentos de percepção;
3. Devido à universalidade das emoções básicas, os sujeitos ouvintes brasileiros e suecos devem inferir as emoções expressas em nossos corpora de maneira semelhante, ainda que tenham mais facilidade em realizar essa tarefa com áudios em sua língua materna. Além disso, os sujeitos de ambas as nacionalidades se guiam pelos mesmos parâmetros acústicos para fazer essa inferência;
4. Ao avaliar as emoções expressas na fala por meio de dimensões emocionais, os sujeitos ouvintes podem ser menos influenciados pelas suas experiências emocionais idiossincráticas que são evocadas diretamente a partir de um item lexical relacionado a uma emoção discreta (Barbosa, 2009). Por isso, espera-se que a concordância entre as respostas dos sujeitos ouvintes nos experimentos de percepção seja maior quando eles avaliam as emoções expressas na fala por meio de dimensões emocionais do que quando o fazem por meio de palavras que descrevem emoções discretas.

As análises estatísticas deste estudo foram realizadas por meio do pacote estatístico R em sua versão 2.11.1 (R Development Core Team, 2010).

Capítulo 3

Experimento de percepção I

No experimento de percepção I, descrito neste capítulo, juízes brasileiros e suecos avaliaram, em escalas graduadas de 0 a 4, o grau de expressão de **emoções discretas** descritas por oito palavras, para os mesmos trechos de fala, que foram extraídos de depoimentos reais de mulheres falantes do **português brasileiro**.

3.1 O CORPUS “JOGO DE CENA”

O corpus com trechos de fala em português brasileiro consiste de trechos de fala expressiva extraídos de depoimentos reais de mulheres, contidos no filme nacional de estilo documentário “Jogo de cena”, que foi dirigido por Eduardo Coutinho e estreou em 2007. Nesse filme, 23 mulheres selecionadas foram filmadas contando sua história de vida em entrevistas com o diretor. Depois, várias atrizes interpretaram, a seu modo, essas histórias. O filme faz um paralelo entre os depoimentos reais e a interpretação das atrizes. No entanto, para este estudo, optamos por utilizar apenas os depoimentos reais e desconsiderar a participação das atrizes.

Foram selecionados 30 trechos de 8 mulheres, cada um com duração variando de 3 a 10 segundos e com qualidade boa o suficiente para garantir análises acústicas adequadas. Achamos que uma duração de 3 a 10 segundos é suficiente para que as emoções expressas pelos falantes sejam reconhecidas pelos juízes, uma vez que, de acordo com a literatura, a maioria das emoções possui curta duração, algo entre meio e quatro segundos (Scherer, 2000b). O documentário não fornece detalhes sobre a idade das mulheres selecionadas nem sobre a origem dialetal delas. Sabe-se, contudo, que elas tinham mais do que 18 anos de idade e moravam na cidade do Rio de Janeiro na época das filmagens. Pode-se perceber, de outiva, que 7 mulheres falavam o dialeto carioca e 1 mulher o dialeto de Campinas, interior do estado de São Paulo. A cidade de origem desta última é confirmada pelo diretor do documentário na faixa de comentário.

A edição dos trechos de fala foi realizada da seguinte maneira: primeiramente, os trechos do filme que continham expressividade foram salvos em formato .avi por meio do programa “SolveigMM AVI Trimmer 2.0”, disponível gratuitamente em <<http://www.baixaki.com.br/>>. Em seguida, o áudio de cada trecho foi extraído em formato .wav com uma taxa de amostragem de 44.100 Hz por meio do programa “Format Factory 3.0.1”, disponível gratuitamente em <<http://www.pcfreetime.com/>>.

Os doze parâmetros acústicos descritos na seção 2.5 foram então medidos automaticamente para cada trecho de fala com o *script* “*Expression Evaluator*”.

3.2 METODOLOGIA

Os experimentos de percepção foram elaborados e aplicados via *internet* por meio da plataforma “Survey Gizmo”, disponível em <<http://www.surveygizmo.com/>>. Os textos do experimento (instruções, questionários e adjetivos) foram apresentados na língua materna dos sujeitos, ou seja, em português para os sujeitos brasileiros e em sueco para os sujeitos suecos. A tradução para o sueco foi realizada a partir de uma versão em inglês pela Profa. Dra. Åsa Abelin, que me orientou no estágio de pesquisa na Universidade de Gotemburgo, Suécia (financiado pela bolsa BEPE – Mestrado da FAPESP).

Nos experimentos, os juízes ouviram os 30 trechos de fala, que lhes eram apresentados **aleatoriamente**, e avaliaram, para cada um deles, o grau com o qual o falante estava expressando a emoção indicada quando produziu o enunciado, em uma escala graduada de 0 a 4, podendo escolher valores intermediários. Para guiar os ouvintes, no extremo esquerdo de cada escala estava escrito “nada X” e no extremo direito, “muito X”, sendo “X” o adjetivo correspondente à emoção a ser julgada naquela parte do teste (esses adjetivos são apresentados abaixo). Por exemplo, para a emoção “alegria”, o valor 0 indicava “nada alegre” e o valor 4, “muito alegre”. O termo “nada” **não** significa, aqui, “fala neutra”. Ele é usado para indicar que o falante não está expressando a emoção indicada em nenhum grau. Portanto, o uso desse termo não exclui a possibilidade de o falante estar expressando outra emoção.

As respostas dos juízes na escala de 0 a 4 foram convertidas em uma escala linear de 0 a 1: 0, 0,25, 0,50, 0,75 e 1, para aproximar esses valores dos valores normalizados por *z-score* dos parâmetros acústicos. Era apresentado na tela um estímulo por vez junto com sua respectiva escala. O áudio era reproduzido automaticamente assim que a página do navegador de *internet* terminava de carregar. O sujeito deveria marcar sua resposta na escala clicando com o *mouse* no valor desejado e então clicar no botão “*next*” localizado na parte inferior da página para seguir para o próximo estímulo. Caso o sujeito não marcasse a sua resposta na escala, não era possível avançar para o próximo estímulo e uma mensagem de erro era exibida. Também não era possível voltar à página anterior e os sujeitos foram informados quanto a isso nas instruções. Eles também foram instruídos a só ouvir novamente os estímulos, caso achassem necessário, no máximo uma vez.

A figura 3.1 mostra a tela do experimento como exibida aos sujeitos ouvintes brasileiros (a versão apresentada aos sujeitos suecos era semelhante, com a única diferença sendo os textos traduzidos para o sueco).



Figura 3.1: Tela do experimento de percepção aplicado por meio da plataforma online “SurveyGizmo”.

O experimento foi composto de 8 partes. Em cada parte, os juízes avaliaram o grau com o qual os falantes dos enunciados expressavam a emoção descrita por um adjetivo, para todos os 30 trechos de fala, sendo um adjetivo diferente para cada parte do experimento. Esses adjetivos estão relacionados aos seguintes substantivos: *alegria*, *comoção*, *surpresa*, *tristeza*, *contentamento*, *angústia*, *aflição* e *entusiasmo*. Escolhemos essas emoções porque eram aquelas disponíveis em nosso corpus “Jogo de Cena”, isto é, aquelas que os falantes expressavam segundo nossa percepção. Além disso, utilizamos palavras com certo grau de sinonímia e antonímia para testar a hipótese de que essas etiquetas formam dois grupos emocionais distintos com relação ao prazer intrínseco da emoção descrita por eles: um positivo (*alegria*, *surpresa*, *contentamento* e *entusiasmo*) e outro negativo (*comoção*, *tristeza*, *angústia* e *aflição*).

Para evitar o cansaço dos juízes, o experimento foi conduzido em duas sessões, realizadas em dias diferentes. Em cada sessão, os sujeitos avaliaram a emoção expressa por quatro itens lexicais diferentes. Na primeira, foram avaliadas as emoções relacionadas aos

substantivos *alegria, comoção, surpresa e tristeza* e na segunda, *contentamento, angústia, aflição e entusiasmo*, exatamente nessa ordem.

Ao utilizar nas escalas os adjetivos correspondentes a esses substantivos (vide tabela 3.1), evitamos um problema metodológico importante, que é o fato de alguns adjetivos apresentarem diferenças de significado ou no emprego em relação ao substantivo correspondente, como, por exemplo, o adjetivo *contente* e o substantivo *contentamento*.

A estrutura do experimento foi rigorosamente a mesma nas duas sessões, alterando-se apenas o adjetivo correspondente à emoção a ser julgada. Na primeira página do experimento, era apresentado um termo de consentimento livre e esclarecido, que informava ao participante a finalidade da pesquisa e que sua participação era voluntária, podendo ele desistir a qualquer momento, bastando para isso fechar a janela do navegador. Foi informado que, em caso de desistência, todas as respostas fornecidas até então seriam descartadas do estudo. Nesse termo também constava o endereço acadêmico e de email para contato com o pesquisador. Na segunda página, eram apresentadas as instruções gerais. Na terceira página, os juízes deveriam informar suas iniciais, seu sexo, idade, email (essa opção era facultativa), cidade em que viveu a maior parte de sua vida, grau de escolaridade máximo e se possuíam algum tipo de problema auditivo. O experimento começava a partir da página seguinte, na qual eram dadas instruções mais específicas quanto àquela parte do experimento.

3.3 ANÁLISE COM SUJEITOS BRASILEIROS

O link para a realização do experimento foi divulgado por email e também através de uma rede social. As pessoas foram informadas que só poderiam participar se fossem brasileiras e foram instruídas a usar fone de ouvido e a fazer o experimento individualmente em um ambiente silencioso.

Somente os sujeitos que completaram as duas sessões do experimento foram considerados para análise. Completaram as duas sessões um total de **26 sujeitos** (9 homens e 17 mulheres), todos brasileiros, de idade superior a 18 anos, universitários (cursando graduação ou pós-graduação) e sem problemas auditivos. A média de idade dos juízes foi

de 23 anos (desvio-padrão: 3,95; idade máxima: 35; mínima: 18). De acordo com os dados coletados antes do início do experimento, 24 sujeitos viveram a maior parte da vida em cidades do estado de São Paulo, 1 sujeito no estado de Minas Gerais (feminino) e 1 sujeito no estado de Pernambuco (feminino).

Tabela 3.1: Adjetivos (em negrito) usados no experimento I com sujeitos brasileiros.

Substantivos	Adjetivos
Alegria	Nada Alegre – Muito Alegre
Comoção	Comovido
Surpresa	Surpreso
Tristeza	Triste
Contentamento	Contente
Angústia	Angustiado
Aflição	Aflito
Entusiasmo	Entusiasmado

3.3.1 Concordância entre os sujeitos

A concordância entre as respostas dos juízes foi verificada pelo cálculo do índice *kappa* de Fleiss (1971), para cada item lexical, com os 5 níveis correspondentes à escala de 5 graus.

O teste proposto por Fleiss (1971) é uma adaptação, para um número n de avaliadores, do teste *kappa* de Cohen (1960), que opera somente com dois avaliadores. Esse teste é significativo para $\alpha = 0,001$ quando $z > 3,09$ e o valor de *kappa* é um número entre 0 e 1, sendo que quanto mais próximo de 1 ele for, maior é a concordância. A tabela 3.2 mostra, em ordem decrescente, os valores de *kappa* para as oito etiquetas, bem como o respectivo valor para z .

O teste *kappa* mostrou boa concordância entre os sujeitos brasileiros quanto às emoções expressas nos trechos de fala de nosso corpus “Jogo de Cena”. A etiqueta *surpreso* teve o menor índice de concordância entre esses sujeitos. O valor máximo de concordância obtido aqui foi praticamente o mesmo que Barbosa (2009) obteve em seu

estudo no qual sujeitos brasileiros avaliaram dimensões emocionais expressas em trotes telefônicos ($k = 0,38$ para a dimensão de ativação).

Tabela 3.2: Valores de $kappa$ e z para as oito etiquetas avaliadas pelos sujeitos brasileiros no experimento de percepção I.

Etiqueta	<i>Kappa</i>	Significância
contente	$k = 0,366$	$z = 69,7$
triste	$k = 0,361$	$z = 66,4$
angustiado	$k = 0,324$	$z = 60,2$
alegre	$k = 0,322$	$z = 61,0$
aflito	$k = 0,253$	$z = 47,6$
entusiasmado	$k = 0,248$	$z = 47,8$
comovido	$k = 0,241$	$z = 46,9$
surpreso	$k = 0,115$	$z = 21,2$

3.3.2 Identificação das variáveis correlacionadas entre si

Realizamos uma **Análise de Componentes Principais (PCA)** a fim de investigar o comportamento dos itens lexicais com relação à percepção dos sujeitos ouvintes brasileiros. A Análise de Componentes Principais é um procedimento estatístico que identifica as variáveis correlacionadas entre si e as agrupa em dimensões (fatores ou componentes) sem correlação umas com as outras, usando o menor número de dimensões necessário para explicar a variância dos dados. Nossa ideia com o uso da PCA é a de que se os itens lexicais forem agrupados em um único fator para determinada emoção, significará que funcionam como sinônimos ou antônimos (quando o agrupamento envolve coeficientes com sinais contrários) para aquela emoção.

Para realizar essa análise, foi utilizada a média das respostas dos juízes no teste de percepção, para cada trecho de fala. A análise com as respostas de todos os 26 sujeitos mostrou que são necessários oito fatores para explicar 100% da variância das respostas. Entretanto, segundo Crawley (2007), a prática comum em Estatística é considerar somente

os fatores da PCA necessários para explicar 90% da variância total dos dados. No nosso caso, dois fatores explicam juntos **96,2%** da variância. O fator 1 (PC1) explica **80,8%** da variância e o fator 2 (PC2), **15,4%**.

A tabela 3.3 mostra os *loadings*, que correspondem a quanto cada variável contribui para cada fator. Por essa tabela, pode-se observar que o fator 1 combina as emoções positivas *alegria*, *contentamento* e *entusiasmo*, que tiveram *loadings* positivos e relativamente altos, em contraste com as emoções negativas *comoção*, *tristeza*, *angústia* e *aflição*, as quais tiveram *loadings* negativos. Já o fator 2 apresentou *loadings* negativos para todas as variáveis, chamando a atenção para a emoção surpresa, que teve um valor muito negativo nesse fator, mas um *loading* positivo e baixo no fator 1. Esses dois grupos, revelados pela PCA, podem ser visualizados mais claramente na figura 3.2, que mostra as oito etiquetas projetadas no plano cartesiano com relação aos seus respectivos *loadings* para os dois fatores da PCA em um gráfico de dispersão. É possível observar também na figura 3.2 que a etiqueta *surpreso* não está incluída em nenhum desses dois grupos, ficando isolada no canto inferior direito do gráfico.

Como as emoções alegria, contentamento e entusiasmo são positivas e estão relacionadas à mesma “sensação”, decidimos chamar o fator 1 de “**FELICIDADE**”. Já para o fator 2, como este é basicamente uma negação de todas as emoções, optamos por nomeá-lo de “**NEUTRALIDADE**”.

Em suma, a PCA mostrou que, como afirmava nossa hipótese formulada na seção 3.2, as emoções *alegria*, *contentamento* e *entusiasmo* formam, de acordo com a avaliação dos sujeitos ouvintes brasileiros, um grupo positivo com relação ao respectivo prazer intrínseco, enquanto que as emoções *comoção*, *tristeza*, *angústia* e *aflição* formam um grupo negativo. A avaliação dos sujeitos ouvintes brasileiros revelou duas grandes dimensões emocionais para nosso corpus “Jogo de Cena”: FELICIDADE e NEUTRALIDADE (a negação de qualquer expressividade).

É importante ressaltar que essas duas palavras (*felicidade* e *neutralidade*) foram escolhidas por nós apenas para nomear os dois fatores da PCA. Além disso, esses resultados não nos permitem afirmar, por exemplo, que *alegria* seja o oposto semântico de *tristeza*, pois o grupo negativo “nega” o grupo positivo como um todo, o que indica apenas

que os sujeitos ouvintes se mostraram capazes de distinguir “felicidade” de “não-felicidade”.

Tabela 3.3: *Loadings* dos adjetivos do experimento I com sujeitos ouvintes brasileiros para o fator 1 (PC1) e para o fator 2 (PC2).

Adjetivo	PC1	PC2
Alegre	0,3783	-0,1310
Comovido	-0,3384	-0,3927
Surpreso	0,1697	-0,7895
Triste	-0,3832	-0,1712
Contente	0,3824	-0,1257
Angustiado	-0,3839	-0,1573
Aflito	-0,3756	-0,2157
Entusiasmado	0,3634	-0,2979

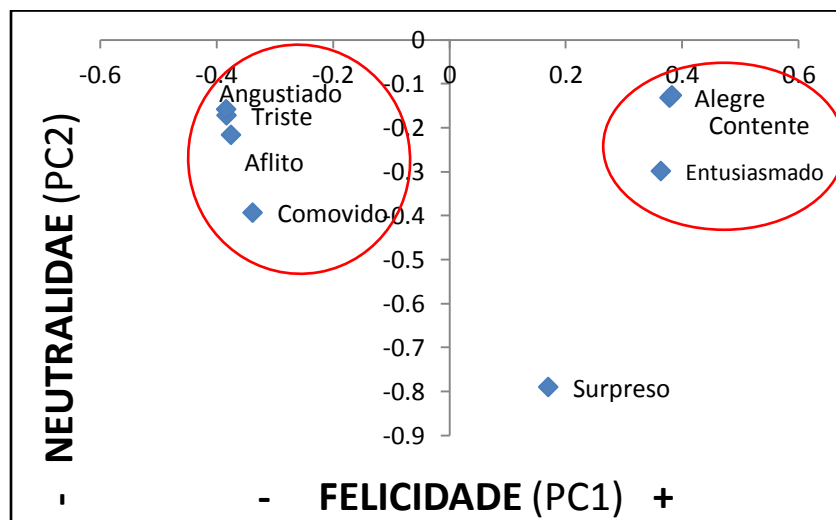


Figura 3.2: Gráfico de dispersão com os adjetivos do experimento I com sujeitos brasileiros, projetados de acordo com os *loadings* para o PC1 e para o PC2.

Em seguida, realizamos a PCA separadamente com as respostas dos sujeitos femininos e masculinos a fim de identificar possíveis diferenças entre os sexos dos ouvintes na percepção das emoções expressas no corpus.

Para as respostas dos sujeitos masculinos, os dois primeiros componentes explicam juntos **93,7%** da variância das respostas, sendo que o primeiro deles, PC1, explica **77,6%** da variância e o segundo, PC2, **16,1%**.

Para as respostas dos sujeitos femininos, os dois primeiros componentes explicam juntos **96,2%** da variância das respostas, sendo que o primeiro, PC1, explica **81,3%** da variância e o segundo, PC2, **14,9%**.

A tabela 3.4 mostra os *loadings* das etiquetas para os dois primeiros componentes calculados separadamente por sexo e a tabela 3.5 mostra os *scores*, também calculados separadamente por sexo. Os *scores* são os novos valores obtidos para os trechos de fala pela PCA, após a dimensionalidade dos dados ter sido reduzida.

Essa análise mostrou que a percepção das emoções expressas no corpus “Jogo de Cena” entre os sujeitos brasileiros masculinos e femininos foi muito semelhante. As etiquetas emocionais tiveram *loadings* com mesmo sinal (positivo ou negativo) para ambos os sexos nos dois fatores (PC1 e PC2). Além disso, os *scores* atribuídos aos estímulos avaliados também foram bastante semelhantes para ambos os sexos, com exceção de alguns trechos que apresentaram valores com sinais contrários (trechos 1 e 3 no primeiro fator e quatro trechos no segundo). Nos enunciados 1 e 3, os falantes aparentam estar comovidos. O enunciado 1 recebeu da PCA *score* positivo no primeiro fator (felicidade) para as respostas dos sujeitos masculinos e foi avaliado por eles com maiores valores para os adjetivos comovido, surpreso e contente, enquanto que para os sujeitos femininos recebeu *score* negativo (não-felicidade) e foi avaliado por eles com valores maiores para os adjetivos comovido, triste, angustiado e aflito. O enunciado 3 recebeu da PCA *score* negativo (não-felicidade) para as respostas dos sujeitos masculinos e foi avaliado por eles com valores maiores para os adjetivos triste e angustiado, enquanto que para os sujeitos femininos recebeu *score* positivo (felicidade) e foi avaliado por eles com valores maiores para os adjetivos comovido, surpreso, contente e entusiasmado.

Tabela 3.4: *Loadings* dos adjetivos do experimento I com sujeitos ouvintes brasileiros, calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 (PC1) e para o fator 2 (PC2).

Adjetivos	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
Alegre	0,3746	0,3795	-0,2034	-0,0967
Comovido	-0,3351	-0,3369	-0,3753	-0,4109
Surpreso	0,0989	0,1961	-0,8236	-0,7690
Triste	-0,3891	-0,3822	-0,1400	-0,1721
Contente	0,3861	0,3810	-0,1603	-0,1094
Angustiado	-0,3901	-0,3822	-0,1437	-0,1586
Aflito	-0,3824	-0,3741	-0,1653	-0,2249
Entusiasmado	0,3724	0,3560	-0,2148	-0,3364

Tabela 3.5: *Scores* dos trinta trechos de fala do corpus “Jogo de Cena” calculados separadamente por sexo para o primeiro fator (PC1) e para o segundo (PC2) a partir das respostas do experimento I com sujeitos ouvintes brasileiros. Valores com sinais opostos em negrito.

Trecho	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
1	0,0111	-1,2378	-1,1212	-0,4636
2	-0,1579	-0,8717	-0,1676	-0,1806
3	-0,3924	0,2425	-0,1197	0,0851
4	2,1191	2,2444	-0,2029	0,7031
5	-4,5302	-4,2045	0,0999	-0,4828
6	1,1500	1,1519	1,6379	0,9758
7	0,3603	0,2467	0,9887	0,0665
8	-0,4781	-1,2204	0,8852	0,4612
9	3,2502	3,5163	-0,7403	-0,9970
10	-4,3519	-4,0456	-0,5667	-1,0056
11	1,8974	2,1166	-1,4944	-1,1462
12	-0,4592	-0,6512	-0,5497	-0,3333
13	-4,2809	-4,0878	-1,0275	-0,8485
14	-4,5000	-4,2088	-0,7824	-0,6949
15	3,6517	3,7702	-2,4759	-2,7518
16	1,6669	1,4553	0,2610	0,8823

17	-3,9129	-3,9774	0,6126	0,5018
18	-3,2900	-3,5371	0,0033	0,6254
19	1,5015	2,6163	0,1163	-0,2231
20	1,5823	2,2072	0,9595	0,0827
21	3,8570	3,8142	-1,1431	-1,0683
22	-2,5000	-2,9928	-1,7212	-1,5042
23	2,2396	2,1983	-0,5978	-0,9927
24	0,9373	0,9617	1,6578	1,8575
25	1,1716	1,4183	1,9417	1,7291
26	0,9091	0,5246	-1,1898	-0,4962
27	-0,0737	-0,1153	1,7229	1,6458
28	2,1778	2,1160	0,6142	0,5704
29	0,4268	0,4355	0,9147	1,3451
30	0,0177	0,1145	1,4845	1,6570

3.3.3 Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos

A fim de determinar quais parâmetros acústicos daqueles extraídos (descritos na seção 2.5) melhor preveem as respostas dos sujeitos ouvintes brasileiros no experimento de percepção I, fizemos uma **regressão linear simples** entre cada um dos doze parâmetros acústicos (variável preditora) e cada fator da PCA (FELICIDADE e NEUTRALIDADE; variável predita). A tabela 3.6 traz o coeficiente de determinação (R^2 ajustado), bem como a significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples de cada parâmetro acústico. O coeficiente de determinação indica o quanto a variância da variável independente (preditora) explica a variância da variável dependente.

Os parâmetros que melhor predizem o fator “FELICIDADE” (PC1) são: assimetria da frequência fundamental, assimetria da intensidade, média e desvio-padrão da inclinação espectral e inclinação do LTAS. Combinados em uma regressão linear múltipla (equação 3.1), os parâmetros assimetria da frequência fundamental, assimetria da intensidade e média e desvio-padrão da inclinação espectral explicam **34%** da variância do primeiro fator (a significância estatística dos coeficientes dessa regressão linear múltipla é mostrada na tabela 3.7). A inclusão do LTAS nessa regressão linear múltipla reduziria o valor do coeficiente de determinação, por isso a exclusão desse parâmetro da regressão.

$$PC1 = 8,68 + 4,59 \textit{assimf}_0 + 2,01 \textit{assimint} - 28,97 \textit{medinclinespec} - 20,44 \textit{desvpadinclinespec} \quad (3.1)$$

A diminuição da energia na banda de frequências superior do espectro (aumento do valor do parâmetro *slopeLTAS*), causada pela diminuição do esforço vocal, tendeu a ser identificada pelos sujeitos ouvintes brasileiros como diminuição do grau do fator “FELICIDADE”, como pode-se observar na figura 3.3. Em alguns desses trechos de fala, os falantes expressavam emoções relacionadas à tristeza. Esse também é o caso para o parâmetro “média da inclinação espectral”, que é o que mais explica esse fator, como se pode observar na figura 3.4. O contrário foi observado para o parâmetro “assimetria de f_0 ” (figura 3.5): o aumento do valor desse parâmetro tendeu a ser interpretado pelos sujeitos ouvintes como aumento do grau de expressão do fator “FELICIDADE”.

Os parâmetros que melhor predizem o fator “NEUTRALIDADE” (PC2) são: mediana, semi-amplitude entre quartis e quantil 99,5% da frequência fundamental, desvio-padrão da primeira derivada de f_0 e assimetria da inclinação espectral. No entanto, a melhor combinação foi a dos parâmetros mediana de f_0 e assimetria da inclinação espectral, que explica **40%** da variância desse fator (equação 3.2; tabela 3.8).

$$PC2 = -0,518 - 2,362 \textit{mednf}_0 - 1,13 \textit{assiminclinespec} \quad (3.2)$$

Para esses cinco parâmetros acústicos, um aumento no valor tendeu a ser interpretado pelos sujeitos ouvintes como diminuição no grau de “NEUTRALIDADE”. Isso pode ser observado nas figuras 3.6 e 3.7, que mostram, respectivamente, as regressões lineares simples para os parâmetros “mediana de f_0 ” e “assimetria da inclinação espectral”. Vê-se que, à medida que os valores desses parâmetros crescem, os *scores* do PC2 tendem a decrescer.

Tabela 3.6: Coeficientes de determinação (R^2 ajustado) e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples (para $\alpha = 10\%$) calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes brasileiros para o experimento I. ns. = não-significativo.

Parâmetros Acústicos	Felicidade (PC1)			Neutralidade (PC2)		
	R^2 ajustado	intercept. sign.	inclin. sign.	R^2 ajustado	intercept. sign.	inclin. sign.
mednf0	-0,0217	ns.	ns.	0,3271	ns.	$p < 0,0006$
sampquartisf0	-0,0339	ns.	ns.	0,1919	$p < 0,02$	$p < 0,009$
quan995f0	0,0475	ns.	ns.	0,1799	$p < 0,04$	$p < 0,02$
assimf0	0,0818	ns.	$p < 0,07$	-0,0054	ns.	ns.
medderivf0	-0,0309	ns.	ns.	-0,0356	ns.	ns.
desvpaddf0	-0,0353	ns.	ns.	0,1028	$p < 0,08$	$p < 0,05$
assimdf0div10	-0,0248	ns.	ns.	0,0007	ns.	ns.
assimint	0,1118	$p < 0,06$	$p < 0,04$	-0,0170	ns.	ns.
medinlinespec	0,2929	$p < 0,002$	$p < 0,002$	-0,0192	ns.	ns.
desvpadinlinespec	0,0979	$p < 0,06$	$p < 0,06$	-0,0327	ns.	ns.
assiminlinespec	0,0125	ns.	ns.	0,0633	ns.	$p < 0,10$
slopeLTAS	0,1232	$p < 0,04$	$p < 0,04$	-0,0023	ns.	ns.

Tabela 3.7: Significância estatística (para $\alpha = 10\%$) dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.1. ns. = não-significativo.

coeficientes	significância
Interceptação	$p < 0,006$
assimf0	ns.
assimint	ns.
medinlinespec	$p < 0,02$
desvpadinlinespec	ns.

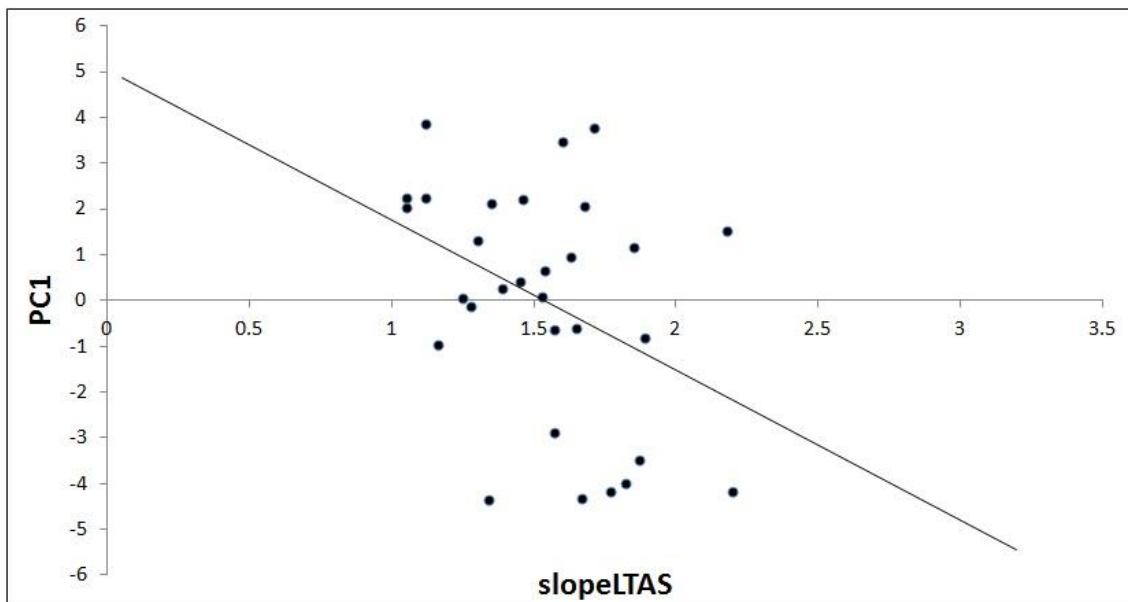


Figura 3.3: Regressão linear simples com o parâmetro “inclinação do LTAS” como variável independente e o primeiro fator da PCA como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 3.6.

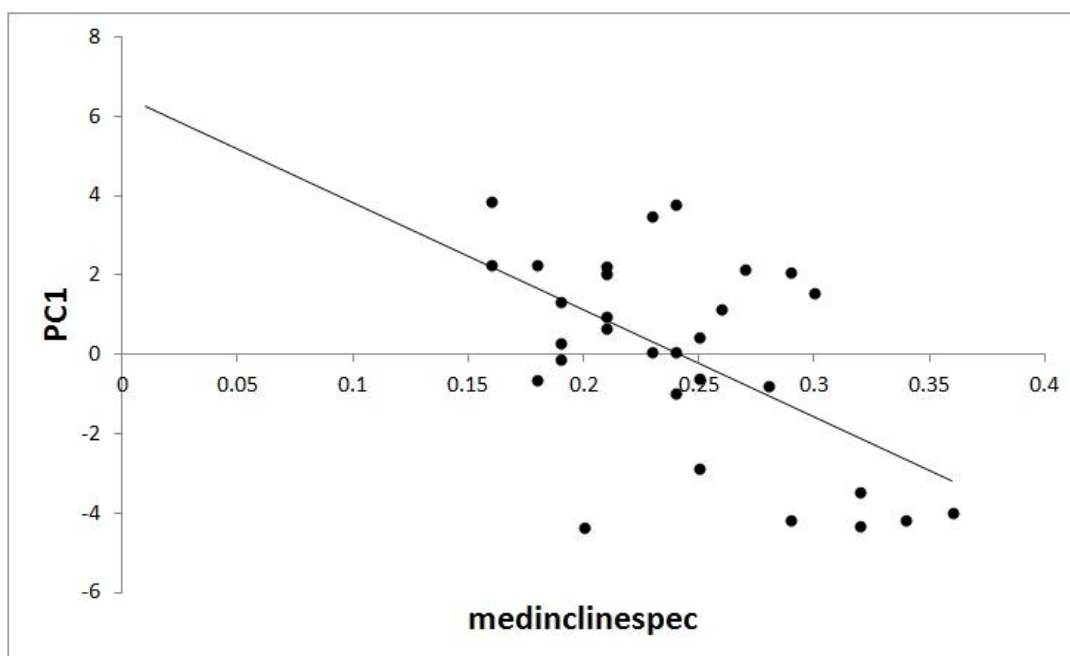


Figura 3.4: Regressão linear simples com o parâmetro “média da inclinação espectral” como variável independente e o primeiro fator da PCA como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 3.6.

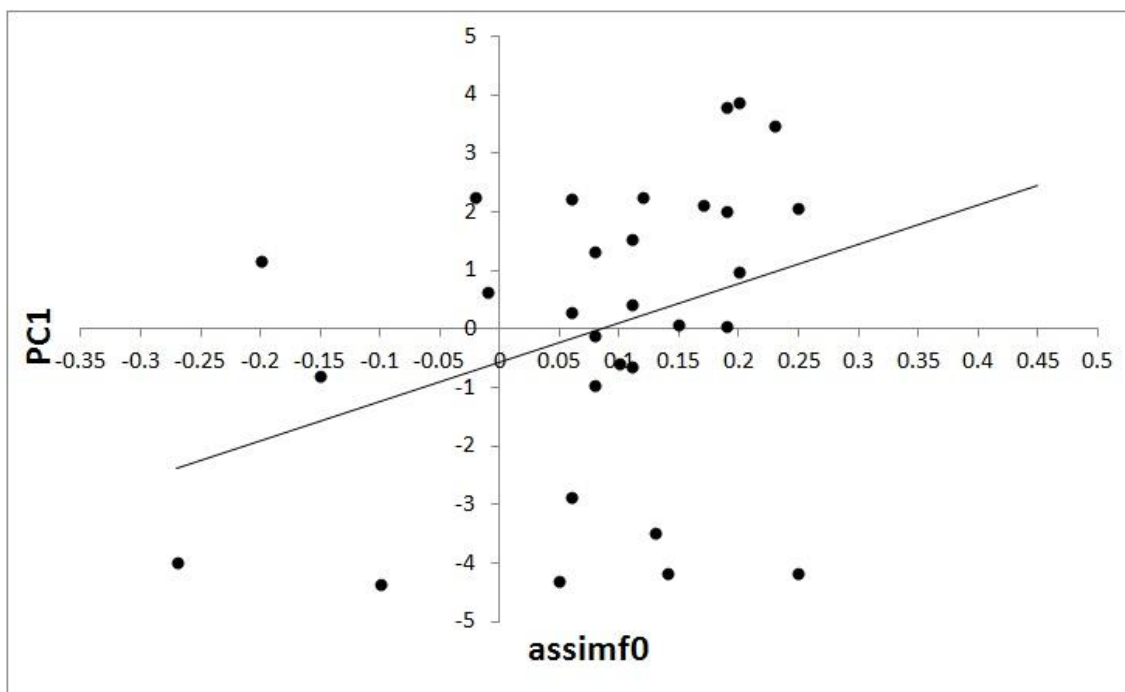


Figura 3.5: Regressão linear simples com o parâmetro “assimetria de f_0 ” como variável independente e o primeiro fator da PCA como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 3.6.

Tabela 3.8: Significância estatística (para $\alpha = 10\%$) dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.2.

coeficientes	significância
Interceptação	$p < 0,04$
mednf0	$p < 0,0004$
assiminclinespec	$p < 0,06$

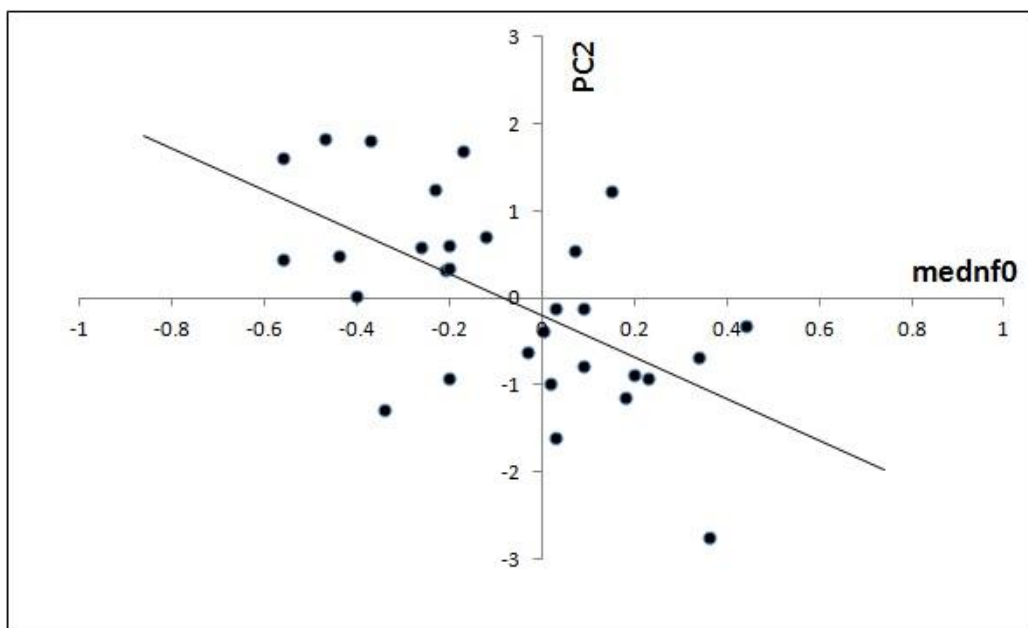


Figura 3.6: Regressão linear simples com o parâmetro “mediana de f_0 ” como variável independente e o segundo fator da PCA como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 3.6.

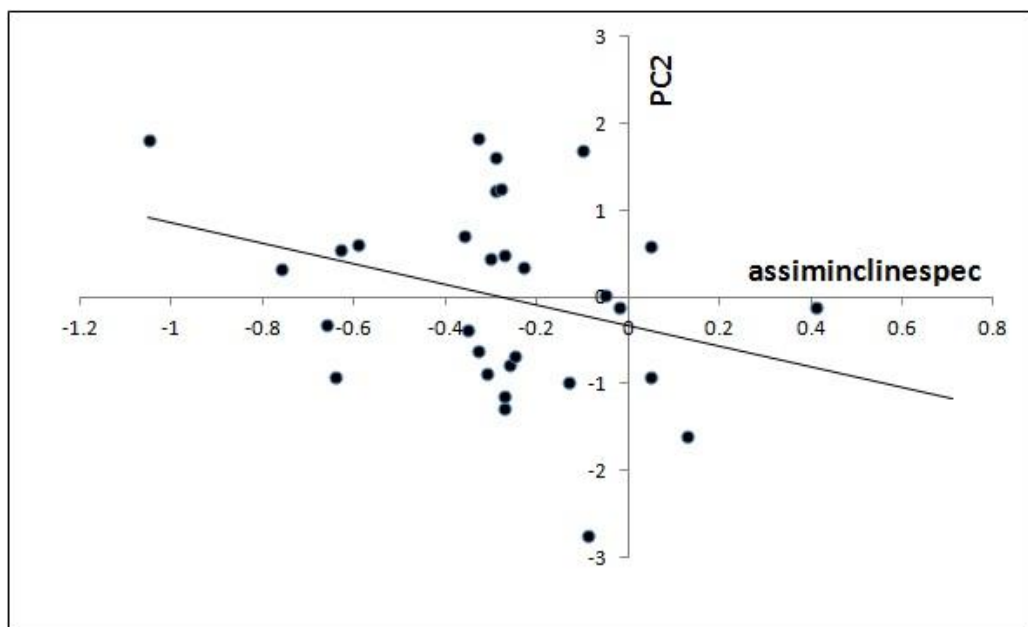


Figura 3.7: Regressão linear simples com o parâmetro “assimetria da inclinação espectral” como variável independente e o segundo fator da PCA como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 3.6.

3.3.4 Classificação automática dos trechos de fala

Por meio da Análise Discriminante Linear (LDA - *Linear Discriminant Analysis*) foi possível classificar automaticamente os trechos de fala do corpus utilizando os parâmetros acústicos. Essa análise opera inferindo as etiquetas de modo que elas sejam mais bem discriminadas estatisticamente, formando grupos com base na distribuição no espaço dos parâmetros acústicos.

Para essa análise, a variável independente foi a combinação de parâmetros acústicos e a variável dependente, os dois componentes da PCA, usados para representar a avaliação dos sujeitos ouvintes brasileiros no experimento I e analisados separadamente. Na LDA, a variável dependente deve ser uma variável categórica (uma palavra, por exemplo). Por esse motivo, atribuímos uma etiqueta a cada trecho de fala usando como referência a média dos *scores* obtidos na Análise de Componentes Principais. Os trechos com *score* acima da média receberam etiqueta positiva (felicidade e neutro) e os trechos com *score* abaixo da média, etiqueta negativa (não-felicidade e não-neutro). Esse procedimento é mostrado na tabela 3.9.

Com o primeiro fator, utilizamos como variáveis independentes os parâmetros acústicos da equação 3.1 (assimetria da frequência fundamental, assimetria da intensidade e média e desvio-padrão da inclinação espectral), correspondentes à combinação que explicou a maior parte da variância desse fator na regressão linear múltipla. A classificação obtida pela LDA foi então comparada com a classificação baseada nos *scores* da PCA. Para esse fator, FELICIDADE, 24 dos 30 trechos de fala do corpus receberam a mesma etiqueta que havia sido obtida com os *scores* da PCA, o que corresponde a **80%** do total.

Para o segundo fator (NEUTRALIDADE), utilizamos como variáveis independentes os dois parâmetros acústicos da equação 3.2: mediana de f_0 e assimetria da inclinação espectral. Dos 30 trechos de fala que compõem o corpus, 27 receberam a mesma etiqueta daquela obtida com os *scores* da PCA, o que corresponde a **90%** do total.

Tabela 3.9: Etiquetas atribuídas a cada trecho de fala do corpus “Jogo de Cena” para o primeiro fator da PCA (PC1) e para o segundo (PC2), de acordo com os *scores* e também com a LDA. nfelicidade = “não-felicidade”; nneutro = “não-neutro”.

Trecho	PC1			PC2		
	Scores	Etiquetas	LDA	Scores	Etiquetas	LDA
1	-0,8066	nfelicidade	nfelicidade	-0,6141	nneutro	nneutro
2	-0,6318	nfelicidade	felicidade	-0,1238	nneutro	nneutro
3	0,0559	felicidade	felicidade	0,0227	neutro	neutro
4	2,2376	felicidade	felicidade	0,4452	neutro	neutro
5	-4,3575	nfelicidade	felicidade	-0,3100	nneutro	nneutro
6	1,1564	felicidade	nfelicidade	1,2271	neutro	nneutro
7	0,2756	felicidade	felicidade	0,3223	neutro	neutro
8	-0,9754	nfelicidade	nfelicidade	0,5984	neutro	neutro
9	3,4763	felicidade	felicidade	-0,9876	nneutro	nneutro
10	-4,1743	nfelicidade	nfelicidade	-0,9226	nneutro	nneutro
11	2,0728	felicidade	felicidade	-1,2917	nneutro	neutro
12	-0,5958	nfelicidade	nfelicidade	-0,3781	nneutro	nneutro
13	-4,1695	nfelicidade	felicidade	-0,9295	nneutro	nneutro
14	-4,3194	nfelicidade	nfelicidade	-0,7723	nneutro	nneutro
15	3,7841	felicidade	felicidade	-2,7415	nneutro	nneutro
16	1,5363	felicidade	felicidade	0,7047	neutro	neutro
17	-4,0017	nfelicidade	nfelicidade	0,5452	neutro	nneutro
18	-3,4860	nfelicidade	nfelicidade	0,4777	neutro	neutro
19	2,2580	felicidade	felicidade	-0,1194	nneutro	nneutro
20	2,0206	felicidade	felicidade	0,3491	neutro	neutro
21	3,8583	felicidade	felicidade	-1,1494	nneutro	nneutro
22	-2,8705	nfelicidade	nfelicidade	-1,6108	nneutro	nneutro
23	2,2274	felicidade	felicidade	-0,8817	nneutro	nneutro
24	0,9678	felicidade	felicidade	1,8187	neutro	neutro
25	1,3160	felicidade	felicidade	1,8389	neutro	neutro
26	0,6423	felicidade	felicidade	-0,6716	nneutro	nneutro
27	-0,1233	nfelicidade	felicidade	1,6906	neutro	neutro
28	2,1273	felicidade	nfelicidade	0,5924	neutro	neutro
29	0,4203	felicidade	felicidade	1,2528	neutro	neutro
30	0,0787	felicidade	felicidade	1,6185	neutro	neutro

3.4 ANÁLISE COM SUJEITOS SUECOS

O experimento foi aplicado em sujeitos suecos seguindo rigorosamente a mesma metodologia adotada com os sujeitos brasileiros, descrita na seção 3.2. O texto do experimento (instruções, questionários e adjetivos) conduzido com sujeitos suecos foi traduzido para o sueco pela Profa. Dra. Åsa Abelin a partir de uma versão em inglês. A tradução dos adjetivos emocionais para o sueco é mostrada na tabela 3.10. Além de traduzir os adjetivos do inglês, usamos um dicionário **sueco-português** para escolher os adjetivos em sueco que apresentassem a maior proximidade semântica dos adjetivos em português possível.

Os sujeitos receberam por email o link para acessar o experimento e foram informados de que somente pessoas nascidas na Suécia e que tivessem o sueco como língua materna poderiam participar do experimento. Além disso, foram instruídos a usar fone de ouvido e a fazer o experimento individualmente e em um ambiente silencioso.

Novamente, somente as respostas daqueles sujeitos que completaram as duas sessões do experimento foram consideradas. Completaram ambas as sessões do experimento **16 sujeitos** (7 homens e 9 mulheres), todos eles tendo vivido a maior parte da vida na Suécia e falantes nativos da língua sueca. Eles também reportaram **não** entender a língua portuguesa nem ter problemas auditivos. Todos eram estudantes de graduação ou de pós-graduação do Departamento de Filosofia, Linguística e Teoria da Ciência da Universidade de Gotemburgo, Suécia. A média de idade desses sujeitos foi de 26 anos (idade mínima: 20 anos; máx. 47 anos; desvio-padrão: 6 anos).

3.4.1 Concordância entre os sujeitos

A concordância entre as respostas dos sujeitos suecos para os oito adjetivos, também verificada pelo teste *kappa* de Fleiss (1971), é mostrada na tabela 3.11.

No geral, a concordância foi satisfatória para todos os adjetivos com exceção do adjetivo “surpreso”, que, apesar de significativo, apresentou um índice *kappa* muito baixo (0,090). O fato de tanto os sujeitos brasileiros quanto os suecos terem apresentado baixa concordância para essa emoção pode indicar que as falantes de nosso corpus “Jogo de

Cena” não expressavam a emoção “surpresa” ou a expressaram com um grau baixo. Isso também pode ter ocorrido por ser a surpresa uma emoção melhor expressa pela face do que pela voz (Ekman, 1999; Fontes, 2014) e pelo fato de a expressão vocal dessa emoção ser facilmente confundida com outras, tais como a alegria ou a raiva (vide Abelin e Allwood, 2000, por exemplo).

O adjetivo “alegre” apresentou, entre os sujeitos suecos, a maior concordância ($k = 0,421$), enquanto que entre os sujeitos brasileiros esse adjetivo apresentou a quarta maior concordância.

Tabela 3.10: Adjetivos (em negrito) usados no experimento I com sujeitos suecos.

Substantivos	Adjetivos
Glädje (alegria)	inte alls (nada) glad – mycket (muito) glad
Rörd (comoção)	rörd
Förvåning (surpresa)	förvånad
Sorg (tristeza)	ledsen
Förnöjsamhet (contentamento)	nöjd
Ångest (angústia)	ångestfylld
Bedrövelse (aflição)	bedrövad
Entusiasm (entusiasmo)	entusiastisk

Tabela 3.11: Valores de *kappa* e *z* para as oito etiquetas avaliadas pelos sujeitos suecos no experimento de percepção I.

Adjetivo	<i>Kappa</i>	Significância
glad (alegre)	0,421	$z = 49,0$
ledsen (triste)	0,373	$z = 41,6$
bedrövad (afrito)	0,329	$z = 37,9$
nöjd (contente)	0,288	$z = 33,4$
ångestfylld (angustiado)	0,241	$z = 28,1$
entusiastisk (entusiasmado)	0,216	$z = 25,3$
rörd (comovido)	0,142	$z = 17,0$
förvånad (surpreso)	0,090	$z = 10,2$

3.4.2 Identificação das variáveis correlacionadas entre si

A Análise de Componentes Principais com a média das respostas dos 16 sujeitos suecos para cada trecho de fala revelou que dois componentes explicam juntos **94,5%** da variância total das respostas. O primeiro desses componentes (PC1) explica **80,3%** da variância e o segundo (PC2), **14,2%**. Os *loadings* de cada adjetivo para ambos os fatores são mostrados na tabela 3.12. A figura 3.8 mostra os adjetivos distribuídos em um gráfico de dispersão de acordo com os respectivos *loadings*.

Pela análise da tabela 3.12, da figura 3.8 e da porcentagem da variância explicada pelos dois componentes, pode-se ver que este resultado é bastante semelhante ao obtido com os sujeitos brasileiros. O primeiro componente da PCA também explica em torno de 80% da variância das respostas dos sujeitos e, novamente, dois grupos emocionais foram revelados por essa análise: um formado por emoções de valência positiva (*nöjd, glad e entusiastisk*) e o outro por emoções de valência negativa (*bedrövad, ledsen, ångestfylld e rörd*). Além disso, os *loadings* desses adjetivos também foram semelhantes aos obtidos para os sujeitos brasileiros. A etiqueta correspondente à emoção da surpresa (*förvånad*) apresentou, novamente, *loading* positivo e relativamente baixo para o primeiro fator e um *loading* negativo e relativamente alto para o segundo.

Devido a essas semelhanças, foi possível nomear ambos os componentes com as mesmas etiquetas aplicadas aos componentes dos sujeitos brasileiros: “FELICIDADE” para o primeiro (PC1) e “NEUTRALIDADE” para o segundo (PC2).

Tabela 3.12: *Loadings* dos adjetivos do experimento I com sujeitos ouvintes suecos para o fator 1 (PC1) e para o fator 2 (PC2).

Adjetivos	PC1	PC2
Glad (alegre)	0,3770	-0,0957
Rörd (comovido)	-0,3064	-0,4713
Förvånad (surpreso)	0,1909	-0,7791
Ledsen (triste)	-0,3865	-0,1465
Nöjd (contente)	0,3784	-0,0629
Ångestfylld (angustiado)	-0,3864	-0,1686
Bedrövad (aflito)	-0,3883	-0,1351
Entusiastisk (entusiasmado)	0,3674	-0,2994

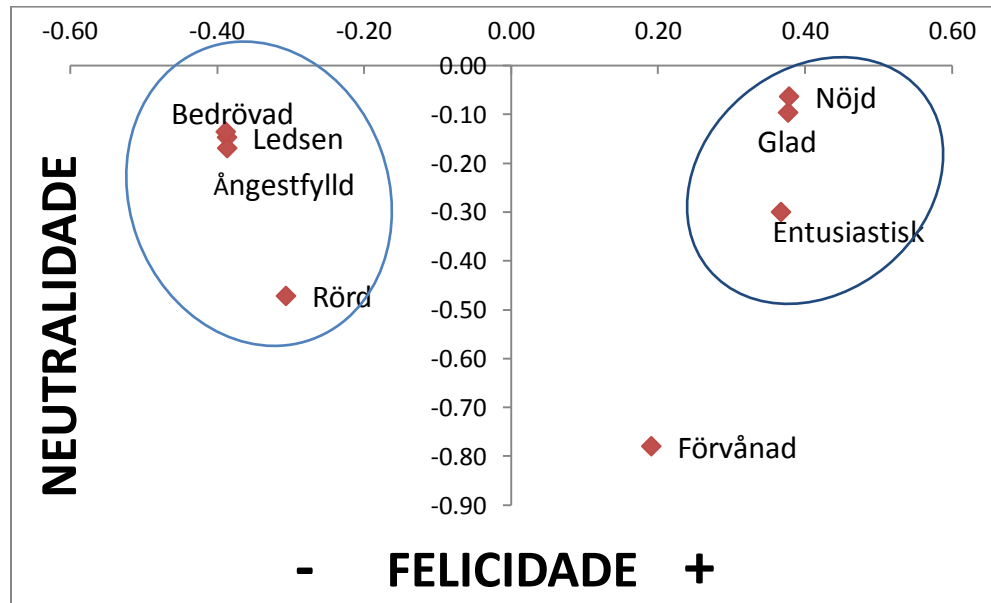


Figura 3.8: Gráfico de dispersão com os adjetivos do experimento I com sujeitos suecos, plotados de acordo com os *loadings* para o PC1 e para o PC2.

Em seguida, também realizamos a PCA separadamente para as respostas dos sujeitos masculinos e femininos. Para ambos os sexos, os dois primeiros componentes explicam mais do que 90% da variância total.

Para os sujeitos ouvintes masculinos, os dois primeiros componentes explicam juntos **92,6%** da variância total, sendo que o primeiro deles explica **73,6%** da variância e o segundo, **19%**.

Para os sujeitos ouvintes femininos, os dois primeiros componentes explicam juntos **93,6%** da variância total, sendo que o primeiro desses componentes explica **83%** dessa variância e o segundo, **10,6%**.

Como também observado entre os sujeitos ouvintes brasileiros, o primeiro componente da PCA explica uma porcentagem da variância maior para os sujeitos suecos femininos do que para os masculinos e a variância total explicada pelos dois componentes também é maior para os sujeitos femininos.

Entretanto, os *loadings* dos adjetivos, calculados separadamente por sexo para os dois componentes da PCA e mostrados na tabela 3.13 e o *scores* de cada trecho de fala para ambos os sexos, mostrados na tabela 3.14, revelam que a percepção das emoções expressas nos trechos de fala do corpus “Jogo de Cena” foi muito semelhante entre os sujeitos suecos masculinos e femininos, como observado também para os sujeitos brasileiros. Os adjetivos apresentaram *loadings* de mesmo sinal (positivo e negativo) para ambos os sexos. Para o primeiro fator (PC1; FELICIDADE), todos os trechos de fala apresentaram *score* de mesmo sinal para ambos os sexos. Já para o segundo fator (PC2; NEUTRALIDADE), alguns trechos apresentaram sinais contrários entre os sexos.

Tabela 3.13: *Loadings* dos adjetivos do experimento I com sujeitos ouvintes suecos, calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 (PC1) e para o fator 2 (PC2).

Adjetivos	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
Glad (alegre)	0,3875	0,3683	-0,0802	-0,1412
Rörd (comovido)	-0,3051	-0,3128	-0,4263	-0,5109
Förvånad (surpreso)	0,0136	0,2545	-0,7885	-0,7280
Ledsen (triste)	-0,4037	-0,3808	-0,0987	-0,1503
Nöjd (contente)	0,3890	0,3694	-0,1020	-0,0846
Ångestfylld (angustiado)	-0,4031	-0,3732	-0,0433	-0,2684
Bedrövad (aflito)	-0,4058	-0,3815	-0,0761	-0,1289
Entusiastisk (entusiasmado)	0,3393	0,3681	-0,4029	-0,2658

Tabela 3.14: *Scores* dos trinta trechos de fala do corpus “Jogo de Cena” calculados separadamente por sexo para o primeiro fator (PC1) e para o segundo (PC2) a partir das respostas do experimento I com sujeitos ouvintes suecos. Valores com sinais opostos em negrito.

Trecho	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
1	-1,5719	-2,0303	-0,0044	0,4946
2	-0,5534	-1,0045	0,1467	0,0807
3	-1,6355	-1,7993	0,1014	0,0492
4	1,8425	1,8052	1,1228	0,5711
5	-4,0111	-3,6250	-1,1519	-0,8147
6	2,2996	0,4802	1,3422	0,7970
7	0,8760	0,4773	0,5822	-0,9780
8	0,6781	0,7146	1,3456	-0,3303
9	2,8618	3,8275	-0,7788	-0,9104
10	-3,8479	-3,8827	-0,4580	-0,8896
11	2,5149	3,1371	-0,8283	-0,8130
12	-1,2066	-0,8053	-0,7547	-0,3855
13	-3,9037	-3,9567	-0,6354	-0,4006
14	-4,0783	-3,8902	-0,6527	-0,7672
15	2,9080	3,0949	-3,9353	-2,0224

16	1,5949	1,7786	0,8840	0,5319
17	-3,6744	-3,9429	0,7793	0,2403
18	-3,2713	-3,4109	-0,1392	0,3579
19	2,6473	2,6299	-1,2230	-0,3677
20	1,4370	1,8649	-0,1737	0,1730
21	3,4130	4,1003	-1,1537	-0,8162
22	-3,0402	-3,7117	-0,7648	-0,7623
23	0,3689	0,1747	1,2811	1,1502
24	0,9956	1,2303	2,0803	1,9326
25	0,7845	0,6568	1,8830	1,6020
26	1,5817	1,2391	-1,5180	-0,8097
27	0,2443	0,4571	0,6240	0,7617
28	2,0519	2,5784	0,0856	-0,2386
29	1,6981	2,0755	0,6694	0,9267
30	-0,0037	-0,2627	1,2443	1,6374

3.4.3 Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos

As respostas dos sujeitos ouvintes suecos (representadas aqui pelos dois componentes da PCA) foram correlacionadas com os doze parâmetros acústicos extraídos por meio de regressões lineares. A tabela 3.15 traz o coeficiente de determinação (R^2 ajustado), bem como a significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples de cada parâmetro acústico.

Os parâmetros que melhor explicam o fator FELICIDADE dos sujeitos suecos são os mesmos que mais preveem esse fator para os sujeitos brasileiros: quantil 99,5% e assimetria da frequência fundamental, assimetria da intensidade, média e desvio-padrão da inclinação espectral e inclinação do LTAS. O quantil 99,5% de f_0 foi o único parâmetro que apresentou um coeficiente de determinação maior do que o apresentado para os sujeitos brasileiros, mas os coeficientes de determinação apresentados pelos outros parâmetros foram muito semelhantes entre as duas nacionalidades.

Uma análise de regressão múltipla revelou que a combinação dos parâmetros “assimetria da frequência fundamental” e “média e desvio-padrão da inclinação espectral” explica **31%** da variância desse fator (equação 3.3; tabela 3.16).

$$PC1 = 7,93 + 4,15assimf_0 - 21,74medinclinespec - 20,54desvpadinclinespec \quad (3.3)$$

Os parâmetros que melhor predizem o fator “NEUTRALIDADE” (PC2) também são os mesmos que mais preveem esse fator para os sujeitos brasileiros: mediana, semi-amplitude entre quartis e quantil 99,5% da frequência fundamental, desvio-padrão da primeira derivada de f_0 e assimetria da inclinação espectral. Entretanto, os parâmetros semi-amplitude entre quartis e quantil 99,5% da frequência fundamental, desvio-padrão da primeira derivada de f_0 e assimetria da inclinação espectral apresentaram coeficientes de determinação maiores com os sujeitos suecos do que com os brasileiros.

A análise de regressão múltipla revelou que a combinação dos parâmetros “mediana e semi-amplitude entre quartis da frequência fundamental” e “assimetria da inclinação espectral” explica **53%** da variância desse fator (equação 3.4; tabela 3.17).

$$PC2 = 1,20 - 1,51mednf_0 - 2,77sampquartisf_0 - 0,88assiminclinespec \quad (3.4)$$

Tabela 3.15: Coeficientes de determinação (R^2 ajustado) e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples (para $\alpha = 10\%$) calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes suecos para o experimento I. ns. = não-significativo.

Parâmetros Acústicos	Felicidade (PC1)			Neutralidade (PC2)		
	R^2 ajustado	intercept. sign.	inclin. sign.	R^2 ajustado	intercept. sign.	inclin. sign.
mednf0	-0,0248	ns.	ns.	0,3157	ns.	$p < 0,0008$
sampquartisf0	-0,0350	ns.	ns.	0,3910	$p < 0,0003$	$p < 0,0002$
quan995f0	0,0734	ns.	$p < 0,09$	0,2422	$p < 0,02$	$p < 0,004$
assimf0	0,0799	ns.	$p < 0,08$	-0,0099	ns.	ns.
medderivf0	-0,0258	ns.	ns.	-0,0357	ns.	ns.
desvpaddf0	-0,0283	ns.	ns.	0,1962	$p < 0,02$	$p < 0,009$
assimdf0div10	-0,0176	ns.	ns.	-0,0089	ns.	ns.
assimint	0,0894	$p < 0,09$	$p < 0,06$	-0,0174	ns.	ns.
medinlinespec	0,2573	$p < 0,004$	$p < 0,003$	-0,0308	ns.	ns.
desvpadinlinespec	0,0937	$p < 0,06$	$p < 0,06$	-0,0356	ns.	ns.
assiminlinespec	-0,0090	ns.	ns.	0,0705	ns.	$p < 0,09$
slopeLTAS	0,1262	$p < 0,04$	$p < 0,04$	-0,0327	ns.	ns.

Tabela 3.16: Significância estatística (para $\alpha = 10\%$) dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.3.

coeficientes	significância
Interceptação	$p < 0,01$
assimf0	ns.
medinlinespec	$p < 0,008$
desvpadinlinespec	ns.

Tabela 3.17: Significância estatística (para $\alpha = 10\%$) dos coeficientes de interceptação e inclinação da regressão linear múltipla da equação 3.4.

coeficientes	significância
Interceptação	$p < 0,06$
mednf0	$p < 0,02$
sampquartisf0	$p < 0,006$
assiminclinespec	$p < 0,08$

3.4.4 Classificação automática dos trechos de fala

Em seguida, realizamos a análise discriminante linear (LDA) com a combinação de parâmetros acústicos como variáveis independentes e os dois componentes da PCA calculados a partir das respostas dos sujeitos ouvintes suecos como variável dependente.

Com o primeiro fator (FELICIDADE), utilizamos como variáveis independentes os parâmetros acústicos da equação 3.3 (assimetria da frequência fundamental e média e desvio-padrão da inclinação espectral), correspondentes à combinação que explicou a maior parte da variância desse fator na regressão linear múltipla. A comparação entre a classificação obtida pela LDA e a classificação baseada nos *scores* da PCA para esse fator mostrou que a classificação pela LDA alcançou **70%** de correspondência com a classificação baseada nos *scores* da PCA, ou seja, 21 dos 30 trechos de fala do corpus receberam da LDA a mesma etiqueta obtida com os *scores* da PCA.

Com o segundo fator (NEUTRALIDADE), utilizamos como variáveis independentes os parâmetros acústicos da equação 3.4 (mediana e semi-amplitude entre quartis da frequência fundamental e assimetria da inclinação espectral). A correspondência entre a classificação obtida pela LDA e a classificação baseada nos *scores* da PCA foi de **73%**, isto é, 22 dos 30 trechos de fala do corpus receberam da LDA a mesma etiqueta obtida com os *scores* da PCA.

3.5 DISCUSSÃO

Nossa hipótese era a de que os sujeitos ouvintes suecos perceberiam as emoções expressas no corpus “Jogo de Cena” de forma semelhante aos sujeitos ouvintes brasileiros devido à universalidade das emoções básicas. As análises com os dados do experimento I corroboraram essa hipótese, pois foi possível perceber que o desempenho dos sujeitos ouvintes de ambas as nacionalidades no experimento de percepção I foi bastante similar.

A concordância entre os sujeitos, verificada por meio do teste *kappa* de Fleiss (1971), foi semelhante entre os sujeitos brasileiros e suecos. A única grande diferença foi a concordância para o adjetivo “alegre”, que apresentou um índice *kappa* maior entre os sujeitos suecos (0,421 contra 0,322). O adjetivo “surpreso” apresentou, com os sujeitos de ambas as nacionalidades, a pior concordância entre os oito adjetivos avaliados nesse experimento.

A Análise de Componentes Principais revelou que a dimensionalidade das respostas dos sujeitos de ambas as nacionalidades para os oito itens lexicais pode ser reduzida a dois componentes: FELICIDADE e NEUTRALIDADE. O primeiro desses componentes, FELICIDADE, explica aproximadamente 80% da variância total para ambas as nacionalidades. De acordo com os *loadings* dessa análise, dois grupos emocionais idênticos foram revelados para as duas nacionalidades: um formado por emoções de prazer intrínseco positivo (alegria, contentamento e entusiasmo) e outro por emoções de prazer intrínseco negativo (comoção, tristeza, angústia e aflição). Esse resultado indica que esses dois componentes da PCA são muito robustos e que os sujeitos ouvintes suecos e brasileiros avaliaram os trechos de fala do corpus “Jogo de Cena” de uma maneira muito semelhante, o que significa que eles tiveram uma percepção semelhante das emoções expressas nesse corpus através dos enunciados.

A correlação dos dois componentes da PCA com os parâmetros acústicos por meio de regressões lineares revelou que os parâmetros acústicos que melhor explicam as respostas dos sujeitos de ambas as nacionalidades são os mesmos: assimetria da frequência fundamental, assimetria da intensidade, média e desvio-padrão da inclinação espectral e espectro médio de longo-termo para o primeiro componente (FELICIDADE) e mediana,

semiamplitude entre quartis e quantil 99,5% da frequência fundamental, desvio-padrão da primeira derivada de f_0 e assimetria da inclinação espectral para o segundo (NEUTRALIDADE). Esse resultado mostra a relevância de parâmetros relacionados à frequência fundamental e intensidade global e relativa (inclinação espectral e LTAS) na expressão e percepção das emoções expressas nos trechos de fala do corpus “Jogo de Cena”. O reforço da energia nas frequências elevadas do espectro, mensurado pelos parâmetros da inclinação espectral e espectro médio de longo-termo, é correlato acústico do esforço vocal e, conseqüentemente, da qualidade de voz. É interessante notar também que parâmetros acústicos diferentes explicam cada um dos dois fatores da PCA. Esses parâmetros acústicos também permitiram classificar automaticamente os trechos de fala por meio da análise discriminante linear com elevadas porcentagens de correspondência com a avaliação dos sujeitos ouvintes. Essa correspondência foi maior para os sujeitos ouvintes brasileiros.

Pode-se concluir que a cultura e a experiência emocional dos sujeitos dos dois países não influenciaram significativamente na percepção das emoções expressas por falantes femininos no corpus “Jogo de Cena”.

Capítulo 4

Experimento de percepção II

Com o experimento I, investigamos se os sujeitos ouvintes suecos perceberiam as emoções expressas em enunciados em português brasileiro do mesmo modo que os ouvintes nativos (sujeitos brasileiros). Com o experimento II, relatado neste capítulo, investigamos a condição oposta: a avaliação por sujeitos ouvintes suecos e brasileiros de trechos de fala em sueco. No entanto, neste experimento utilizamos dimensões emocionais ao invés de emoções discretas a fim de investigar se o uso de dimensões emocionais facilita a avaliação dos sujeitos ouvintes (o que seria refletido em uma maior concordância entre os sujeitos).

4.1 O CORPUS “SWEEMO”

O corpus “SweEmo”, utilizado no experimento II, é composto de 40 trechos de fala com conteúdo emocional de 5 falantes femininos do sueco, com duração variando de 1 a 6 segundos e com qualidade boa o suficiente para garantir análises acústicas adequadas. Os enunciados foram extraídos de situações reais (debates e entrevistas) de programas transmitidos por uma emissora de televisão sueca e por um programa de rádio sueco

disponibilizado livremente na *internet* como *podcast*. Os vídeos referentes aos programas de televisão foram fornecidos pela Biblioteca Nacional da Suécia (*Kunliga biblioteket*), a partir de requisição formal da Profa. Dra. Åsa Abelin.

O áudio dos vídeos (em formato .mp4) foi extraído em formato .wav com uma taxa de amostragem de 44.100 Hz com o programa “Format Factory 3.0.1”, disponível gratuitamente em <<http://www.pcfreetime.com/>>. Os trechos de fala foram então selecionados e extraídos por meio do programa Praat (Boersma e Weenink, 2011).

Os doze parâmetros acústicos descritos na seção 2.5 foram então medidos automaticamente no programa Praat para cada trecho de fala, com o *script* “*Expression Evaluator*”.

4.2 METODOLOGIA

Assim como o experimento I, o experimento II também foi elaborado e aplicado via *internet* por meio da plataforma “Survey Gizmo”, disponível em <<http://www.surveygizmo.com/>>. Os textos do experimento (instruções, questionários e adjetivos) foram apresentados na língua materna dos sujeitos (em português para os sujeitos brasileiros e em sueco para os sujeitos suecos). A tradução para o sueco também foi realizada a partir de uma versão em inglês pela Profa. Dra. Åsa Abelin.

Este experimento consistiu de cinco partes, realizadas em sessão única. Em cada uma, os sujeitos ouvintes avaliaram uma dimensão emocional para todos os 40 trechos de fala. As dimensões emocionais avaliadas foram: **ativação**, **justiça**, **valência**, **motivação** e **envolvimento**, nesta ordem. As dimensões de ativação, valência e envolvimento já foram anteriormente estudadas em experimentos sobre percepção de emoções pela fala (vide seção 1.5). As dimensões de justiça e motivação foram propostas por Frijda *et al.* (1995) e utilizadas em questionários do estudo intercultural que esses autores conduziram. A dimensão “justiça” refere-se à avaliação do evento pelo indivíduo como sendo justo ou injusto, enquanto que a dimensão “motivação” está relacionada com o grau de prontidão para ação do indivíduo: se o evento aumentou ou diminuiu a disposição do indivíduo em

agir sobre a situação. Nossa hipótese é a de que o grau dessas duas dimensões emocionais também pode ser inferido por meio da fala de um indivíduo.

Os 40 trechos de fala deste experimento eram apresentados aos sujeitos **aleatoriamente**. Após ouvir cada enunciado, os sujeitos deveriam avaliar o grau da dimensão emocional específica àquela parte do experimento respondendo a uma pergunta (por exemplo: quão agitado estava o falante?). Para realizar essa avaliação, eles deveriam selecionar um valor em uma escala graduada de 0 a 4, podendo escolher valores intermediários. Nos extremos esquerdo e direito das escalas constava um adjetivo referente à dimensão a ser avaliada, precedido pelos advérbios *nada* e *muito*, para guiar os ouvintes. O valor 0 da escala correspondia a “nada *adjetivo*” e o valor 4, a “muito *adjetivo*”. Esses adjetivos e as questões correspondentes são mostrados, em português e em sueco, na tabela 4.1.

Era apresentado na página do navegador de *internet* um estímulo por vez junto com sua respectiva escala. O áudio correspondente a cada trecho de fala era reproduzido automaticamente assim que a página terminava de carregar e os sujeitos foram instruídos a ouvir novamente os estímulos, caso achassem necessário, no máximo uma vez mais. O sujeito deveria marcar sua resposta na escala clicando com o *mouse* no valor desejado e então clicar no botão “next” localizado na parte inferior da página para seguir para o próximo estímulo. Caso o sujeito não marcasse a sua resposta na escala, não era possível avançar para o próximo estímulo e uma mensagem de erro era exibida. Também não era possível retornar à página anterior. A figura 4.1 mostra uma página deste experimento na versão conduzida com sujeitos ouvintes suecos.

Tabela 4.1: As cinco partes do experimento II. Versão em sueco à esquerda e em português à direita. Em cada célula, a segunda linha corresponde à pergunta específica a cada parte e a terceira linha ao adjetivo usado na escala.

<p><i>Del 1 - Aktivering</i> <i>Hur exalterad var talaren?</i> <i>inte alls exalterad - mycket exalterad</i></p>	<p>Parte 1 - Ativação Quão agitado estava o falante? Nada agitado - Muito agitado</p>
<p><i>Del 2 - Rättvisa</i> <i>Hur rättvist ansåg talaren det som hände var?</i> <i>inte alls rättvis - mycket rättvis</i></p>	<p>Parte 2 - Justiça Quão justo o falante considerou o que aconteceu? Nada justo - Muito justo</p>
<p><i>Del 3 - Behaglighet</i> <i>Hur behaglig för talaren var den situation som han/hon befann sig i?</i> <i>inte alls behaglig - mycket behaglig</i></p>	<p>Parte 3 - Valência Quão agradável para o falante foi a situação na qual ele estava? Nada agradável - Muito agradável</p>
<p><i>Del 4 - Motivation</i> <i>Hur motiverad att handla i situationen var talaren?</i> <i>inte alls motiverad - mycket motiverad</i></p>	<p>Parte 4 - Motivação Quão motivado para agir na situação está o falante? Nada motivado - Muito motivado</p>
<p><i>Del 5 - Engagemang</i> <i>Hur engagerad är talaren i den situation som han/hon befann sig i?</i> <i>inte alls engagerad - mycket engagerad</i></p>	<p>Parte 5 - Envolvimento Quão envolvido com a situação o falante está? Nada envolvido - Muito envolvido</p>

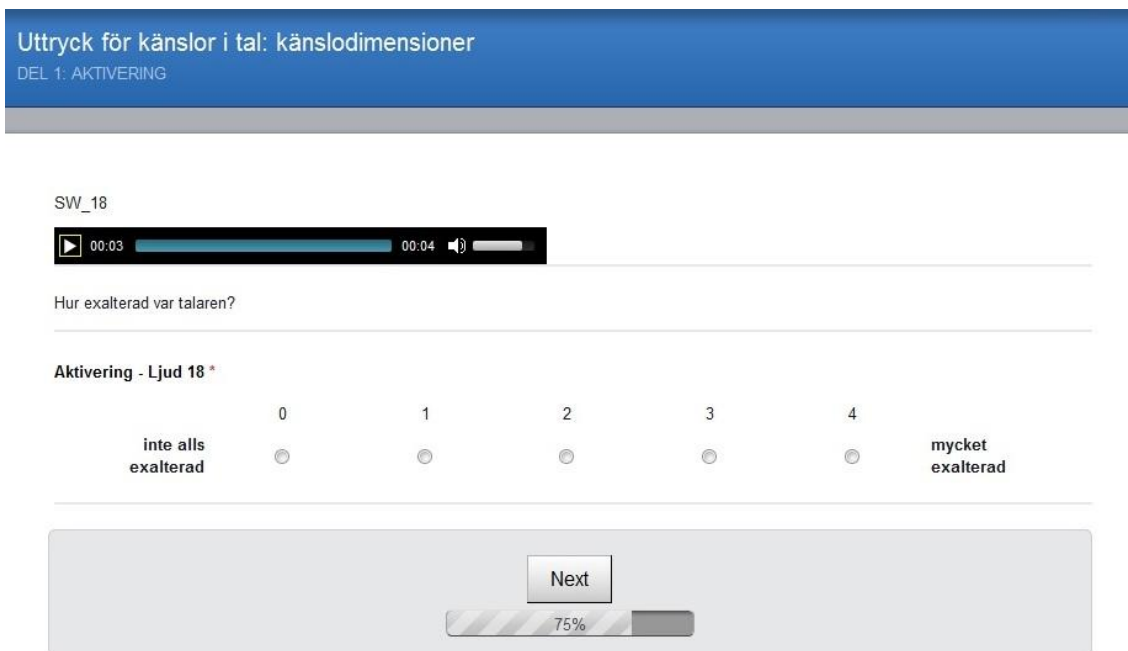


Figura 4.1: Página do experimento II na versão apresentada aos sujeitos ouvintes suecos.

4.3 ANÁLISE COM SUJEITOS SUECOS

Os sujeitos receberam o *link* para acessar o experimento por *email*. Eles foram informados que somente suecos cuja língua materna fosse o sueco poderiam participar do experimento e foram instruídos a utilizar fone de ouvido e a fazer o experimento em um ambiente silencioso. Antes de iniciar o experimento, os sujeitos leram e concordaram com um termo de consentimento livre e esclarecido, que informava que a participação era voluntária e anônima e que eles poderiam desistir dela a qualquer momento se desejassem. Eles também preencheram um questionário com algumas informações como sexo, idade, cidade de origem e outras.

Somente as respostas dos sujeitos que completaram o experimento foram consideradas para análise, o que corresponde a **19 sujeitos** (6 homens). Eles eram estudantes de graduação ou pós-graduação ou funcionários do Departamento de Filosofia, Linguística e Teoria da Ciência da Universidade de Gotemburgo, Suécia. O nível de escolaridade mínimo dos sujeitos era o ensino médio completo (*Gymnasiet*) e a idade média

foi de 30 anos, variando de 21 a 56 anos. Todos eles nasceram e viveram a maior parte da vida na Suécia e possuem o sueco como língua materna. Eles também informaram não ter problemas auditivos.

4.3.1 Concordância entre os sujeitos

A concordância entre as respostas dos sujeitos ouvintes foi verificada pelo cálculo do índice *kappa* de Fleiss (1971), para cada dimensão emocional, com os 5 níveis correspondentes à escala de 5 graus. Os valores de *kappa* para cada dimensão são mostrados em ordem decrescente na tabela 4.2.

Todas as cinco dimensões tiveram um nível de concordância aceitável para as análises posteriores, mas não melhor do que a concordância obtida para o experimento I, como era previsto pela nossa hipótese 4. A dimensão “justiça” apresentou o maior nível de concordância entre os sujeitos, enquanto que a dimensão “motivação” teve o pior. A dimensão “ativação” apresentou somente o quarto maior grau de concordância.

4.3.2 Identificação das variáveis correlacionadas entre si

A Análise de Componentes Principais (PCA) realizada com a média das respostas dos sujeitos ouvintes suecos para cada trecho de fala e com as cinco dimensões emocionais revelou que dois componentes explicam **98,2%** da variância total das respostas. O primeiro desses componentes (PC1) explica **86%** dessa variância, enquanto que o segundo (PC2), **12,2%**.

Os *loadings* das cinco dimensões emocionais para esses dois componentes são mostrados na tabela 4.3 e a figura 4.2 mostra as dimensões distribuídas em um gráfico de dispersão de acordo com esses *loadings*. As dimensões “justiça” e “valência” foram as únicas a apresentar *loading* negativo para o PC1 e todas as cinco dimensões tiveram *loading* negativo para o PC2. Isso indica que os sujeitos ouvintes suecos distinguiram “justiça” e “valência” de “envolvimento”, “motivação” e “ativação”, como se pode ver na figura 4.2.

Tabela 4.2: Valores de *kappa* e significância das cinco dimensões emocionais avaliadas por sujeitos ouvintes suecos no experimento II.

Dimensão	Kappa	Significância
<i>Rättvisa</i> (justiça)	0,254	$z = 39,7$
<i>Behaglighet</i> (valência)	0,241	$z = 39,3$
<i>Engagemang</i> (envolvimento)	0,234	$z = 34,4$
<i>Aktivering</i> (ativação)	0,208	$z = 34,3$
<i>Motivation</i> (motivação)	0,190	$z = 30,0$

Tabela 4.3: *Loadings* das cinco dimensões avaliadas pelos sujeitos ouvintes suecos no experimento II para o primeiro fator da PCA (PC1) e para o segundo (PC2).

Dimensões	PC1	PC2
<i>Aktivering</i> (ativação)	0,4484	-0,4236
<i>Rättvisa</i> (justiça)	-0,4406	-0,4994
<i>Behaglighet</i> (valência)	-0,4207	-0,6133
<i>Motivation</i> (motivação)	0,4585	-0,3313
<i>Engagemang</i> (envolvimento)	0,4664	-0,2920

Os trechos de fala avaliados pelos sujeitos suecos com valores altos para a dimensão “justiça” também foram avaliados com valores altos para a dimensão “valência”, e apresentaram um *score* negativo para o PC1. Por sua vez, os enunciados avaliados pelos sujeitos com valores altos para “ativação”, foram também avaliados com valores altos para as dimensões “motivação” e “envolvimento” (e, portanto, com baixos valores para as dimensões de justiça e valência), apresentando um *score* positivo para o PC1. Ao ouvir os trechos de fala, pode-se perceber que naqueles enunciados com *score* positivo para o PC1, os falantes expressavam a emoção “raiva” com algum grau (na percepção dos sujeitos ouvintes suecos, esses falantes estavam muito agitados, muito envolvidos e muito motivados a esboçar uma ação). Por isso, chamamos o primeiro componente de “**PRONTIDÃO PARA AÇÃO**”.

Como a dimensão de valência foi a que apresentou o *loading* mais negativo para o segundo componente, chamamos o mesmo de “**INSATISFAÇÃO**”.

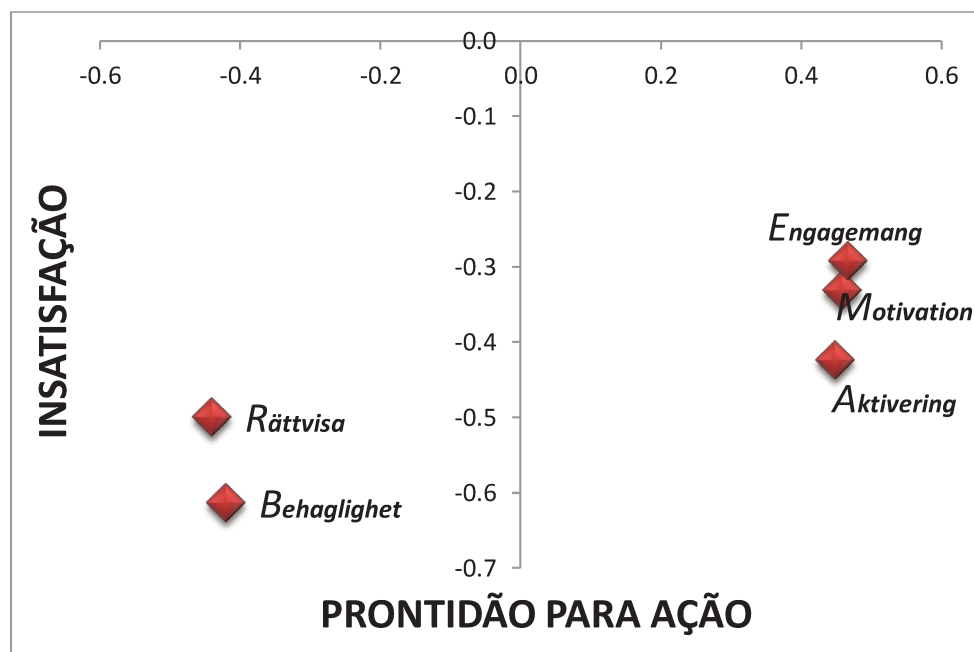


Figura 4.2: Gráfico de dispersão com as dimensões emocionais avaliadas pelos sujeitos ouvintes suecos no experimento II plotadas de acordo com os *loadings* para o primeiro fator da PCA (PRONTIDÃO PARA AÇÃO) e para o segundo (INSATISFAÇÃO).

Em seguida, realizamos a PCA separadamente com as respostas dos sujeitos femininos e masculinos a fim de verificar se houve alguma diferença significativa na percepção das emoções expressas no corpus entre os sexos dos ouvintes.

Para as respostas dos sujeitos masculinos, os dois primeiros componentes explicam juntos **96%** da variância das respostas, sendo que o primeiro deles, PC1, explica **85,7%** da variância e o segundo, PC2, **10,3%**.

Para as respostas dos sujeitos femininos, **97,4%** da variância das respostas é explicada pelos dois primeiros componentes, sendo que o primeiro, PC1, explica **83,8%** da variância e o segundo, PC2, **13,6%**.

A tabela 4.4 mostra os *loadings* das dimensões emocionais calculados separadamente para as respostas dos sujeitos ouvintes masculinos e femininos para os dois componentes da PCA. Para o PC1, as dimensões apresentaram *loadings* semelhantes e de mesmo sinal entre os sexos. Entretanto, para o PC2 todas as cinco dimensões emocionais apresentaram *loadings* com sinais contrários (negativo para as respostas dos homens e positivo para as respostas das mulheres). Os *scores* atribuídos aos enunciados pela PCA para esse componente mostram que houve uma inversão nesse eixo. Assim, no caso de uma avaliação semelhante entre os sujeitos masculinos e femininos para um determinado enunciado, *scores* com sinais opostos foram atribuídos para o PC2 dele. Pela tabela 4.5, que mostra a comparação dos *scores* atribuídos aos enunciados entre os sexos, pode-se ver que para o PC2 houve uma correspondência de avaliação entre os sexos para 31 dos 40 trechos de fala do corpus.

Para o PC1, que explica a maior parte da variância total das respostas dos sujeitos de ambos os sexos, somente o enunciado **35** teve *scores* com sinais opostos entre os sexos (o que, para esse componente, significa julgamentos diferentes). Entre os homens, esse trecho foi avaliado com valores maiores (em média) para as dimensões "ativação" e "valência", enquanto que entre as mulheres esse trecho recebeu valores altos para as dimensões "ativação", "valência", "motivação" e "envolvimento". Essa diferença de sinais dos *scores* atribuídos ao enunciado 35 para os sexos dos ouvintes é, portanto, relacionada à diferença de percepção do grau de expressividade das dimensões "motivação" e "envolvimento" entre os sexos (na percepção das mulheres, a falante desse trecho estava mais motivada a agir e mais envolvida com a situação do que na percepção dos homens).

Portanto, é possível concluir que, com exceção do enunciado 35, a percepção das dimensões emocionais expressas no corpus "SweEmo" foi bastante semelhante entre os sujeitos suecos masculinos e femininos.

Tabela 4.4: *Loadings* das dimensões emocionais avaliadas no experimento II por sujeitos ouvintes suecos, calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 (PC1) e para o fator 2 (PC2) da PCA.

Dimensões Emocionais	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
<i>Aktivering</i> (ativação)	0,4343	0,4572	-0,5680	0,3486
<i>Rättvisa</i> (justiça)	-0,4535	-0,4347	-0,4094	0,5275
<i>Behaglighet</i> (valência)	-0,4237	-0,4204	-0,6379	0,5975
<i>Motivation</i> (motivação)	0,4605	0,4554	-0,2339	0,3663
<i>Engagemang</i> (envolvimento)	0,4628	0,4668	-0,2194	0,3304

Tabela 4.5: *Scores* (separados por sexo) para o primeiro e segundo componentes (PRONTIDÃO PARA AÇÃO e INSATISFAÇÃO) atribuídos pela PCA aos 40 trechos de fala do corpus “SweEmo” avaliados por sujeitos ouvintes suecos no experimento II.

Trechos	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
1	-2,1696	-0,9621	-0,2327	0,9301
2	1,8575	1,8305	0,1481	0,2962
3	-2,0796	-2,3411	1,0375	-0,9092
4	-1,9435	-2,4579	1,0183	-0,7772
5	0,9131	1,2804	0,3927	-0,4308
6	-1,0647	-1,2462	-0,9942	1,1407
7	1,3921	1,1055	0,4240	-0,4986
8	-1,7105	-2,0251	-1,0833	0,7369
9	-3,0143	-2,9110	-0,1808	-0,1845
10	1,6741	1,9163	-0,0101	-0,2003
11	-0,5384	-0,5083	-0,9101	0,7250
12	1,5424	1,4094	0,2853	-0,5857
13	-2,2410	-2,6430	1,0618	-1,6603
14	-1,4574	-2,0811	0,8891	-0,4544
15	0,7443	1,3815	0,4301	-0,3553
16	-1,4542	-0,9166	-1,2037	1,2410
17	2,5068	2,6291	-0,1218	-0,0992
18	-1,9634	-1,9133	-0,1540	0,5750
19	-2,4798	-2,8894	-0,2217	-0,2755
20	2,3055	2,3395	0,2290	-0,3255

21	-1,1596	-1,0774	-0,1374	0,4359
22	1,3624	1,3112	0,4844	-0,5998
23	2,9658	2,7283	-0,1698	0,0719
24	-2,3807	-2,2301	0,4118	0,0183
25	3,3055	3,1624	-0,3230	-0,1325
26	-1,4357	-1,5145	0,8494	-0,9340
27	3,1847	2,8838	-0,1650	-0,1561
28	-1,4318	-1,5438	1,3056	-1,7779
29	-1,9730	-1,3397	-0,7966	1,5048
30	2,8139	2,7411	0,1293	-0,2007
31	2,7591	2,5293	0,0468	0,0092
32	-1,3834	-1,8228	1,0511	-1,1681
33	3,1027	2,7839	0,0354	-0,1077
34	-1,7994	-1,8369	0,8217	-0,7383
35	-1,0250	0,0844	-1,6408	1,6799
36	2,9667	2,8361	0,0517	-0,0128
37	-1,5444	-1,4964	-0,6620	1,0799
38	-1,4845	-1,3785	-1,0665	1,0974
39	2,5562	2,3598	0,0024	-0,2488
40	-0,2186	-0,1774	-1,0319	1,2909

4.3.3 Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos

Dando continuidade à análise, fizemos regressões lineares simples entre cada um dos doze parâmetros acústicos descritos na seção 2.5 (variável preditora) e cada fator da PCA (PRONTIDÃO PARA AÇÃO e INSATISFAÇÃO) para investigar quais desses parâmetros acústicos melhor preveem as respostas dos sujeitos ouvintes suecos no experimento II.

A tabela 4.6 apresenta o coeficiente de determinação (R^2 ajustado) e a significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples de cada parâmetro acústico. O coeficiente de determinação indica o quanto a variância da variável independente (preditora) explica a variância da variável dependente.

De acordo com essa análise, os parâmetros acústicos que mais explicam a variância do primeiro componente, PRONTIDÃO PARA AÇÃO, são: mediana, semi-amplitude entre quartis e quantil 99,5% da frequência fundamental, média e desvio-padrão da inclinação espectral e inclinação do espectro médio de longo-termo. Somente a variância do parâmetro “inclinação do espectro médio de longo-termo” já explica cerca de

70% da variância desse fator. Como se observa na figura 4.3, a diminuição da inclinação do LTAS (causada pelo aumento da intensidade relativa na banda de frequências superior) tendeu a ser interpretada pelos sujeitos ouvintes suecos como aumento no grau de prontidão para ação dos falantes. Já um aumento no parâmetro acústico “mediana da frequência fundamental”, que é o segundo parâmetro que mais explica a variância desse componente, tendeu a ser interpretado por esses sujeitos como aumento do grau de prontidão para ação dos falantes dos trechos de fala do corpus “SweEmo”, como pode ser observado na figura 4.4.

A combinação dos parâmetros mediana e semiamplitude entre quartis da frequência fundamental, média e desvio-padrão da inclinação espectral e inclinação do espectro médio de longo-termo em uma regressão linear múltipla (equação 4.1) prevê **79%** da variância do primeiro componente da PCA (a significância estatística dos coeficientes dessa regressão linear múltipla é mostrada na tabela 4.7).

$$PC1 = 3,26 + 2,11mednf0 + 0,35sampquartisf0 + 10,27medinclinespec + 10,57desvpadinclinespec - 4,78slopeLTAS \quad (4.1)$$

Os doze parâmetros acústicos extraídos não foram capazes de prever a variância do segundo componente da PCA (INSATISFAÇÃO) muito bem. Os parâmetros mediana da frequência fundamental, média da primeira derivada da frequência fundamental e desvio-padrão e assimetria da inclinação espectral explicam uma pequena porcentagem significativa da variância desse componente. A combinação desses parâmetros em uma regressão linear múltipla (equação 4.2) explica **12%** da variância desse componente (coeficientes não significativos estatisticamente).

$$PC2 = -0,55 - 0,38mednf0 + 0,07medderivf0 + 4,93desvpadinclinespec - 0,35assiminclinespec \quad (4.2)$$

Tabela 4.6: Coeficientes de determinação (R^2 ajustado) e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples (para $\alpha = 10\%$) calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes suecos para o experimento II. ns. = não-significativo.

Parâmetros Acústicos	PRONTIDÃO PARA AÇÃO (PC1)			INSATISFAÇÃO (PC2)		
	R^2 ajustado	intercept sign.	inclin. sign.	R^2 ajustado	intercept sign.	inclin. sign.
mednf0	0,4125	ns.	$p < 10^{-5}$	0,0531	ns.	$p < 0,10$
sampquartisf0	0,2324	$p < 0,005$	$p < 0,001$	0,0229	ns.	ns.
quan995f0	0,0542	ns.	$p < 0,10$	-0,0124	ns.	ns.
assimf0	0,0222	ns.	ns.	0,0091	ns.	ns.
medderivf0	-0,0254	ns.	ns.	0,0730	ns.	$p < 0,10$
desvpaddf0	-0,0256	ns.	ns.	-0,0224	ns.	ns.
assimdf0div10	-0,0203	ns.	ns.	-0,0239	ns.	ns.
assimint	-0,0190	ns.	ns.	-0,0146	ns.	ns.
medinlinespec	0,2295	$p < 0,002$	$p < 0,002$	-0,0162	ns.	ns.
desvpadinlinespec	0,1296	$p < 0,02$	$p < 0,02$	0,0533	$p < 0,10$	$p < 0,10$
assiminlinespec	-0,0009	ns.	ns.	0,0481	ns.	$p < 0,10$
slopeLTAS	0,7046	$p < 10^{-10}$	$p < 10^{-11}$	-0,0260	ns.	ns.

Tabela 4.7: Significância estatística dos coeficientes de interceptação e inclinação da equação 4.1 para $\alpha = 10\%$. ns. = não-significativo.

coeficientes	significância
Interceptação	$p < 0,02$
mednf0	$p < 0,001$
sampquartisf0	ns.
medinlinespec	$p < 0,1$
desvpadinlinespec	ns.
slopeLTAS	$p < 10^{-7}$

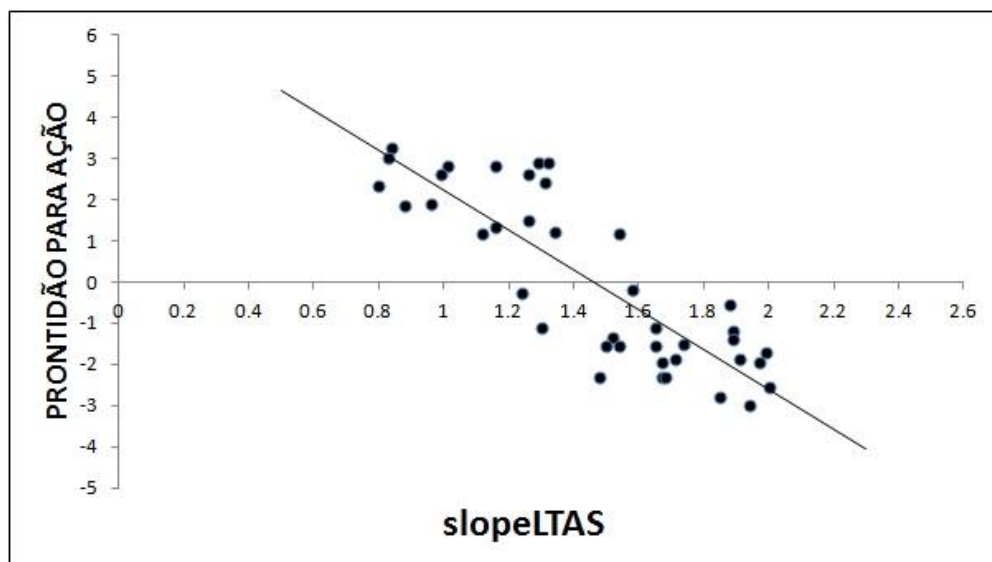


Figura 4.3: Regressão linear simples com o parâmetro “inclinação do LTAS” como variável independente e o primeiro fator da PCA com sujeitos ouvintes suecos (PRONTIDÃO PARA AÇÃO) como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 4.6.

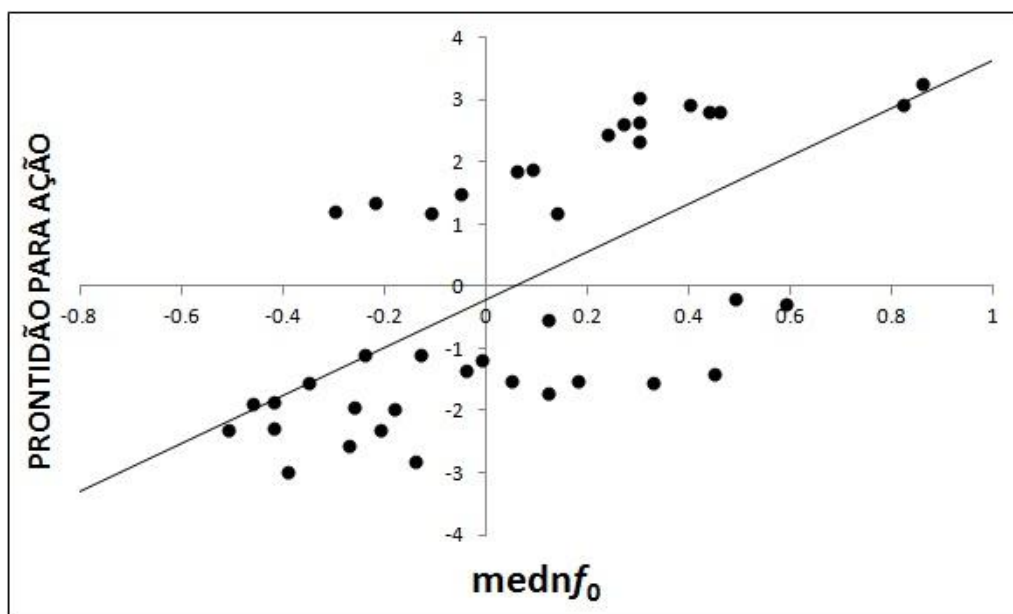


Figura 4.4: Regressão linear simples com o parâmetro “mediana da frequência fundamental” como variável independente e o primeiro fator da PCA com sujeitos ouvintes suecos (PRONTIDÃO PARA AÇÃO) como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 4.6.

4.3.4 Classificação automática dos trechos de fala

A fim de classificar cada trecho de fala de acordo com a distribuição dos parâmetros acústicos, realizamos uma análise discriminante linear (LDA), utilizando os *scores* atribuídos aos trechos de fala para ambos os fatores da PCA para representar a avaliação dos sujeitos ouvintes suecos no experimento II. Como na LDA a variável dependente deve ser uma variável categórica, atribuímos uma etiqueta a cada trecho de fala usando como referência a média dos *scores* atribuídos pela PCA. Os enunciados com *score* maior do que a média dos *scores* receberam etiqueta positiva (prontidão e insatisfação), enquanto que aqueles com *score* menor do que a média dos *scores* receberam etiqueta negativa (não-prontidão e não-insatisfação). Esse procedimento é mostrado na tabela 4.8.

Com o componente PRONTIDÃO PARA AÇÃO, utilizamos como variáveis independentes os parâmetros acústicos da equação 4.1 (mediana e semiamplitude entre quartis da frequência fundamental, média e desvio-padrão da inclinação espectral e inclinação do espectro médio de longo-termo), pois são aqueles que melhor explicam a variância desse componente. Então comparamos essa classificação obtida pela LDA com a classificação baseada nos *scores* da PCA e constatamos que 36 dos 40 trechos de fala do corpus receberam da LDA a mesma etiqueta que havia sido atribuída na classificação baseada nos *scores* da PCA, o que corresponde a **90%** dos trechos de fala do corpus.

Para o componente INSATISFAÇÃO, utilizamos como variáveis independentes os parâmetros acústicos da equação 4.2: mediana da frequência fundamental, média da primeira derivada da frequência fundamental e desvio-padrão e assimetria da inclinação espectral. A comparação dessa classificação pela LDA com a classificação baseada nos *scores* da PCA revelou que 26 dos 40 trechos de fala do corpus receberam da LDA a mesma etiqueta que havia sido atribuída na classificação baseada nos *scores* da PCA, o que corresponde a **65%** dos trechos de fala do corpus “SweEmo”.

Tabela 4.8: Etiquetas atribuídas a cada trecho de fala do corpus “SweEmo” para o primeiro fator da PCA (PC1) e para o segundo (PC2), de acordo com os *scores* e também com a LDA.

Trecho	PC1			PC2		
	<i>scores</i>	etiquetas	LDA	<i>scores</i>	etiquetas	LDA
1	-1,3533	NãoProntidão	NãoProntidão	-0,7126	NãoInsatisfação	NãoInsatisfação
2	1,8832	Prontidão	Prontidão	-0,1605	NãoInsatisfação	Insatisfação
3	-2,3041	NãoProntidão	NãoProntidão	0,9483	Insatisfação	Insatisfação
4	-2,3154	NãoProntidão	NãoProntidão	0,8816	Insatisfação	Insatisfação
5	1,1720	Prontidão	NãoProntidão	0,3854	Insatisfação	Insatisfação
6	-1,1822	NãoProntidão	NãoProntidão	-1,0753	NãoInsatisfação	Insatisfação
7	1,2033	Prontidão	NãoProntidão	0,4739	Insatisfação	Insatisfação
8	-1,9628	NãoProntidão	NãoProntidão	-0,8660	NãoInsatisfação	Insatisfação
9	-2,9873	NãoProntidão	NãoProntidão	0,0362	Insatisfação	Insatisfação
10	1,8606	Prontidão	Prontidão	0,1585	Insatisfação	Insatisfação
11	-0,5497	NãoProntidão	NãoProntidão	-0,7959	NãoInsatisfação	Insatisfação
12	1,4801	Prontidão	Prontidão	0,4865	Insatisfação	NãoInsatisfação
13	-2,5531	NãoProntidão	NãoProntidão	1,4732	Insatisfação	Insatisfação
14	-1,8859	NãoProntidão	NãoProntidão	0,6137	Insatisfação	Insatisfação
15	1,1891	Prontidão	Prontidão	0,4011	Insatisfação	NãoInsatisfação
16	-1,1057	NãoProntidão	NãoProntidão	-1,2180	NãoInsatisfação	Insatisfação
17	2,6199	Prontidão	Prontidão	0,0205	Insatisfação	Insatisfação
18	-1,9575	NãoProntidão	NãoProntidão	-0,4132	NãoInsatisfação	Insatisfação
19	-2,8073	NãoProntidão	NãoProntidão	0,1162	Insatisfação	Insatisfação
20	2,3468	Prontidão	Prontidão	0,3007	Insatisfação	NãoInsatisfação
21	-1,1039	NãoProntidão	NãoProntidão	-0,3102	NãoInsatisfação	Insatisfação
22	1,3504	Prontidão	Prontidão	0,5852	Insatisfação	Insatisfação
23	2,8247	Prontidão	Prontidão	-0,1119	NãoInsatisfação	NãoInsatisfação
24	-2,2931	NãoProntidão	NãoProntidão	0,1471	Insatisfação	Insatisfação
25	3,2567	Prontidão	Prontidão	-0,0366	NãoInsatisfação	NãoInsatisfação
26	-1,5325	NãoProntidão	NãoProntidão	0,9133	Insatisfação	Insatisfação
27	3,0271	Prontidão	Prontidão	0,0450	Insatisfação	NãoInsatisfação
28	-1,5231	NãoProntidão	NãoProntidão	1,6263	Insatisfação	Insatisfação
29	-1,5472	NãoProntidão	NãoProntidão	-1,2636	NãoInsatisfação	Insatisfação
30	2,8102	Prontidão	Prontidão	0,1706	Insatisfação	Insatisfação
31	2,6382	Prontidão	Prontidão	0,0243	Insatisfação	Insatisfação
32	-1,7054	NãoProntidão	NãoProntidão	1,1186	Insatisfação	Insatisfação
33	2,9173	Prontidão	Prontidão	0,0870	Insatisfação	Insatisfação
34	-1,8524	NãoProntidão	NãoProntidão	0,7687	Insatisfação	Insatisfação
35	-0,2725	NãoProntidão	Prontidão	-1,6740	NãoInsatisfação	NãoInsatisfação

36	2,9162	Prontidão	Prontidão	-0,0047	NãoInsatisfação	Insatisfação
37	-1,5422	NãoProntidão	NãoProntidão	-0,9403	NãoInsatisfação	NãoInsatisfação
38	-1,4054	NãoProntidão	NãoProntidão	-1,1075	NãoInsatisfação	Insatisfação
39	2,4415	Prontidão	NãoProntidão	0,1432	Insatisfação	Insatisfação
40	-0,1953	NãoProntidão	NãoProntidão	-1,2347	NãoInsatisfação	NãoInsatisfação

4.4 ANÁLISE COM SUJEITOS BRASILEIROS

O experimento II também foi conduzido com sujeitos ouvintes brasileiros seguindo a mesma metodologia adotada com os sujeitos suecos, que é descrita na seção 4.2. Eles avaliaram os mesmos 40 trechos de fala para cada uma das cinco dimensões emocionais listadas na tabela 4.1. Os textos do experimento (instruções, questionários e adjetivos) foram apresentados aos sujeitos em português brasileiro. O *link* para acessar o experimento pela *internet* foi divulgado em uma rede social junto com a informação de que somente brasileiros nativos poderiam participar. Os sujeitos foram instruídos a fazer o experimento individualmente e em um ambiente silencioso e a usar fone de ouvido.

20 sujeitos completaram o experimento (7 homens), com média de idade de 25 anos (idade mínima = 18 anos; máxima = 34 anos; desvio-padrão = 4,25 anos). Todos eram estudantes de graduação ou de pós-graduação e nasceram e viveram grande parte da vida no Brasil. De acordo com os dados coletados antes do início do experimento, 19 desses sujeitos viveram a maior parte da vida em cidades do estado de São Paulo e 1 sujeito feminino no estado da Bahia. Todos os sujeitos informaram não ter nenhum conhecimento da língua sueca e nenhum problema auditivo.

4.4.1 Concordância entre os sujeitos

A concordância entre as respostas dos sujeitos ouvintes para cada dimensão emocional foi verificada pelo cálculo do índice *kappa* de Fleiss (1971), com os 5 níveis correspondentes à escala de 5 graus. Os valores de *kappa* para cada dimensão são mostrados em ordem decrescente na tabela 4.9.

No geral, os sujeitos ouvintes brasileiros apresentaram um nível de concordância para as dimensões emocionais um pouco menor do que os sujeitos suecos. A única exceção foi o nível de concordância apresentado por eles para a dimensão “ativação” ($k = 0,209$), que foi semelhante ao nível de concordância que os sujeitos suecos apresentaram para essa dimensão ($k = 0,208$). A dimensão “justiça”, que entre os sujeitos suecos teve o melhor grau de concordância entre as cinco dimensões ($k = 0,25$), apresentou entre os sujeitos brasileiros o índice *kappa* mais baixo ($k = 0,11$) entre todas as dimensões.

Tabela 4.9: Valores de *kappa* e significância das cinco dimensões emocionais avaliadas por sujeitos ouvintes brasileiros no experimento II.

Dimensões	<i>Kappa</i>	Significância
Ativação	0,209	$z = 35,8$
Valência	0,196	$z = 33,3$
Motivação	0,172	$z = 29,2$
Envolvimento	0,118	$z = 19,3$
Justiça	0,110	$z = 18,9$

4.4.2 Identificação das variáveis correlacionadas entre si

A Análise de Componentes Principais (PCA) realizada com a média das respostas dos sujeitos ouvintes brasileiros para cada trecho de fala e com as cinco dimensões emocionais revelou que dois componentes explicam **95,4%** da variância total das respostas. O primeiro desses componentes (PC1) é responsável por **76%** dessa variância, enquanto que o segundo (PC2), é responsável por **19,4%** da variância.

Os *loadings* das cinco dimensões emocionais para esses dois componentes são mostrados na tabela 4.10 e a figura 4.5 mostra as dimensões projetadas em um gráfico de dispersão de acordo com esses *loadings*. Contrariamente ao resultado obtido com os sujeitos suecos, as dimensões "justiça" e "valência" apresentaram *loading* positivo para o PC1 enquanto que "ativação", "motivação" e "envolvimento" apresentaram *loading* negativo para esse componente. Para o PC2, todas as cinco dimensões avaliadas

apresentaram *loading* positivo. Essa análise revelou que os sujeitos ouvintes brasileiros distinguíram nesse experimento de percepção "justiça" e "valência" de "ativação", "motivação" e "envolvimento", fato também observado anteriormente com a análise das respostas dos sujeitos suecos (apesar da diferença de sinal dos *loadings* entre as duas nacionalidades).

Os enunciados que apresentaram um *score* positivo para o primeiro componente da PCA foram avaliados pelos sujeitos ouvintes brasileiros com valores altos para as dimensões “justiça” e “valência”, enquanto que aqueles que apresentaram um *score* negativo para esse componente foram avaliados por esses sujeitos com valores altos para as dimensões "ativação", "motivação" e "envolvimento". Nesses últimos enunciados, os falantes expressavam a emoção “raiva” com algum grau, diferentemente dos trechos de fala que receberam *score* positivo para o PC1. Por esse motivo, julgamos que seria apropriado nomear o primeiro componente da PCA de “CALMA”.

As dimensões que apresentaram os maiores *loadings* para o segundo componente foram “justiça”, “valência” e “motivação”. Por isso, chamamos esse componente de “SATISFAÇÃO”.

Tabela 4.10: *Loadings* das cinco dimensões avaliadas pelos sujeitos ouvintes brasileiros no experimento II para o primeiro fator da PCA (PC1) e para o segundo (PC2).

Dimensões	PC1	PC2
Ativação	-0,4936	0,1941
Justiça	0,4600	0,4123
Valência	0,3926	0,6334
Motivação	-0,4080	0,5861
Envolvimento	-0,4735	0,2184

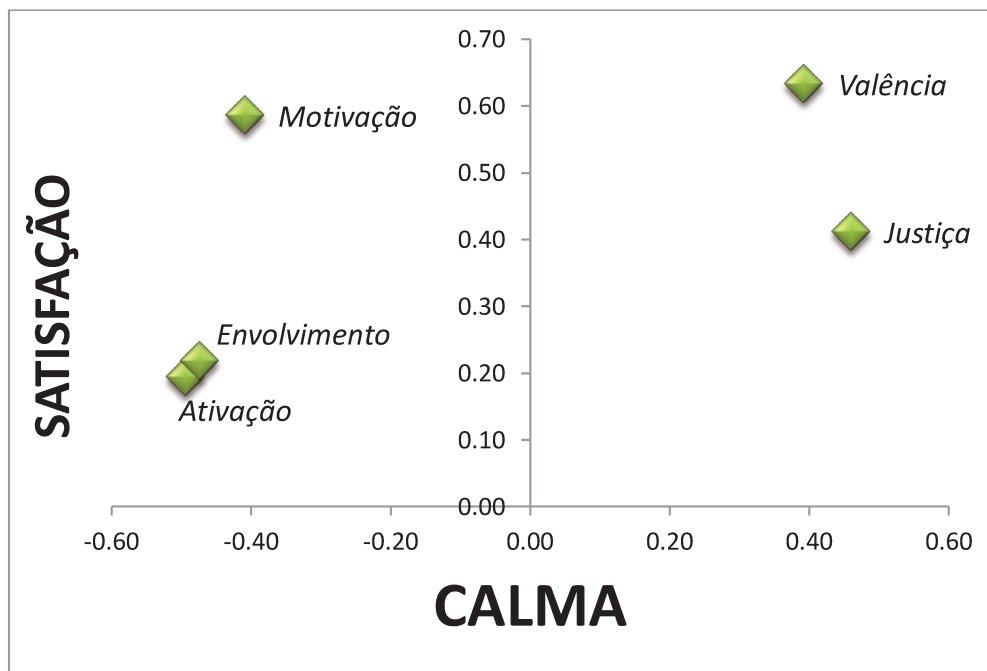


Figura 4.5: Gráfico de dispersão com as cinco dimensões emocionais avaliadas no experimento II por sujeitos ouvintes brasileiros, projetadas de acordo com os *loadings* para o primeiro e o segundo componente da PCA (CALMA e SATISFAÇÃO, respectivamente).

A PCA também foi realizada separadamente para as respostas dos sujeitos masculinos e femininos.

Para as respostas dos sujeitos masculinos, os dois primeiros componentes explicam juntos **92,2%** da variância total das respostas. O primeiro deles é responsável por **73,5%** do total da variância explicada, enquanto que o segundo é responsável por **18,7%** dessa variância.

Para as respostas dos sujeitos femininos, **93,8%** da variância total é explicada pelos dois primeiros componentes da PCA, sendo que o primeiro deles explica **72,9%** dessa variância e o segundo, **20,9%**.

A tabela 4.11 mostra os *loadings* das dimensões emocionais para os dois componentes da PCA, calculados separadamente para as respostas dos sujeitos ouvintes masculinos e femininos. Todas as cinco dimensões emocionais apresentaram *loadings* com

o mesmo sinal para ambos os sexos dos ouvintes, tanto para o primeiro componente da PCA como para o segundo, o que indica que a percepção dessas dimensões emocionais por ambos os sexos para nosso corpus com enunciados em sueco foi, no geral, muito semelhante. O resultado foi o mesmo daquele mostrado na tabela 4.10, em que analisamos as respostas de ambos os sexos conjuntamente. Para o PC1, as dimensões “justiça” e “valência” apresentaram *loadings* positivos enquanto que as outras três apresentaram *loadings* negativos. Para o PC2, as cinco dimensões apresentaram *loadings* positivos para ambos os sexos.

A tabela 4.12 mostra os *scores* atribuídos pela PCA aos 40 trechos de fala do corpus “SweEmo” para as respostas dos sujeitos ouvintes masculinos e femininos. Para o primeiro componente (PC1), somente três trechos de fala tiveram *scores* com sinais opostos entre os sexos (trechos 1, 7 e 35). O enunciado 1 recebeu dos sujeitos masculinos valores maiores para as dimensões "Valência", "Ativação" e "Envolvimento", enquanto que com os sujeitos femininos os valores atribuídos foram maiores para as dimensões "Valência", "Justiça" e "Motivação". Os sujeitos masculinos atribuíram ao enunciado 7 graus maiores para as dimensões "Ativação" e "Motivação", enquanto que os sujeitos femininos atribuíram a esse enunciado graus maiores para as dimensões "Ativação", "Motivação" e "Envolvimento". O trecho de fala 35 recebeu dos sujeitos masculinos valores maiores para as dimensões "Valência", "Motivação" e "Envolvimento" e dos sujeitos femininos graus maiores somente para as dimensões "Valência" e "Motivação". Para o segundo componente (PC2), houve diferença de sinal entre os sexos dos ouvintes para sete enunciados.

Tabela 4.11: *Loadings* das dimensões emocionais avaliadas no experimento II por sujeitos ouvintes brasileiros calculados separadamente para as respostas dos homens e das mulheres, para o fator 1 (PC1) e para o fator 2 (PC2) da PCA.

Dimensões	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
Ativação	-0,5024	-0,4971	0,1659	0,1888
Justiça	0,4023	0,4803	0,5652	0,3429
Valência	0,4021	0,3838	0,5662	0,6450
Motivação	-0,4553	-0,3814	0,4579	0,6371
Envolvimento	-0,4656	-0,4790	0,3504	0,1575

Tabela 4.12: *Scores* atribuídos pela PCA aos 40 trechos de fala do corpus “SweEmo” analisados separadamente por sexo para o primeiro fator (PC1) e para o segundo (PC2) a partir das respostas dos sujeitos ouvintes brasileiros no experimento II. Valores com sinais opostos em negrito.

Trechos	PC1		PC2	
	Homens	Mulheres	Homens	Mulheres
1	-0,0568	0,7856	1,2062	1,2374
2	-0,7460	-1,0268	-0,2415	-0,1652
3	2,0515	2,3050	-0,7316	-0,7228
4	1,2504	1,5460	-1,0076	-1,4986
5	-0,1377	-0,8006	-0,4427	0,1785
6	1,1176	1,4452	0,6381	0,8728
7	0,2115	-0,4408	-1,2495	-0,5892
8	1,3150	1,5073	2,7697	1,9383
9	3,0054	2,1925	-0,7714	-0,9523
10	-1,3321	-1,8395	0,1774	-0,0019
11	0,3700	1,1204	0,4777	0,1781
12	-0,5515	-0,1062	-0,3169	-0,3709
13	1,9689	1,7550	-1,5798	-1,9597
14	2,2926	2,4014	0,2317	0,4166
15	-0,7807	-0,8622	0,0496	0,8689
16	0,6907	1,0609	2,0302	1,5552
17	-2,3658	-1,9780	0,1163	-0,6473
18	0,6884	0,7599	1,5413	1,6685
19	1,8749	2,2730	-0,9870	-1,5671

20	-2,8564	-2,5374	-0,6502	-0,0991
21	0,8280	0,9365	1,0193	0,3291
22	1,5709	0,8440	-0,9400	-1,1560
23	-3,3127	-2,9767	-0,7038	-0,4983
24	2,3125	1,1806	-0,9626	-1,5269
25	-3,8405	-3,8583	-0,3366	-0,5793
26	1,3180	0,7933	-0,7210	-1,5471
27	-3,1172	-3,4171	-0,1177	-0,0118
28	1,9792	2,9423	-0,6708	0,0365
29	1,4621	1,6139	1,1091	1,1376
30	-1,9938	-2,6307	-0,0276	-0,7499
31	-2,6896	-2,9822	0,0187	-0,2265
32	2,0800	2,2153	-0,8081	-0,4108
33	-2,2407	-2,3984	-0,2092	1,0690
34	1,5760	1,7001	-0,3346	-0,5071
35	-0,3172	0,2014	1,3904	2,1281
36	-3,5878	-2,7740	-0,7387	-0,6820
37	0,6282	0,3397	0,8030	0,6627
38	1,8546	1,0311	0,2888	1,0768
39	-1,2557	-1,1757	-0,7318	0,2137
40	-1,2641	-1,1457	1,4133	0,9022

4.4.3 Correlação das respostas dos sujeitos ouvintes com os parâmetros acústicos

A fim de investigar quais parâmetros acústicos melhor preveem as respostas dos sujeitos ouvintes brasileiros no experimento II, realizamos análises de regressão linear simples entre cada um dos doze parâmetros acústicos extraídos (variável independente), descritos na seção 2.5, e cada um dos dois fatores da PCA (CALMA e SATISFAÇÃO). O coeficiente de determinação (R^2 ajustado) e a significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples para cada parâmetro acústico são mostrados na tabela 4.13.

Os parâmetros acústicos que mais explicam a variância do primeiro componente da PCA, CALMA, são: mediana, semiamplitude entre quartis, quantil 99,5% e assimetria da frequência fundamental, média e desvio-padrão da inclinação espectral e inclinação do espectro médio de longo-termo. Com exceção da assimetria de f_0 , esses são os

mesmos parâmetros que mais explicam a variância do primeiro componente da PCA com as respostas dos sujeitos ouvintes suecos (PRONTIDÃO PARA AÇÃO). Contudo, os parâmetros mediana e quantil 99,5% de f_0 e média e desvio-padrão da inclinação espectral explicam uma proporção da variância do primeiro componente maior para os sujeitos brasileiros do que para os suecos.

Das várias regressões múltiplas realizadas, a que explicou a maior porcentagem da variância do componente CALMA (R^2 ajustado = **82%**) foi aquela que combinou os parâmetros mediana de f_0 e inclinação do espectro médio de longo-termo (equação 4.3, tabela 4.14).

$$PC1 = -4,71 - 2,37mednf0 + 3,33slopeLTAS \quad (4.3)$$

Quanto à SATISFAÇÃO, somente o parâmetro média da primeira derivada de f_0 explicou uma parte da variância desse componente (R^2 ajustado = **8%**; equação 4.4).

$$PC2 = -0,269 - 0,163medderivf0 \quad (4.4)$$

Tabela 4.13: Coeficientes de determinação (R^2 ajustado) e significância estatística dos coeficientes de interceptação e inclinação das retas de regressão linear simples (para $\alpha = 10\%$) calculadas separadamente para o primeiro e segundo fator da PCA com sujeitos ouvintes brasileiros para o experimento II. ns. = não-significativo.

Parâmetros Acústicos	CALMA (PC1)			SATISFAÇÃO (PC2)		
	R^2 Ajustado	intercept. sign.	inclin. sign.	R^2 Ajustado	intercept. sign.	inclin. sign.
mednf0	0,5239	ns.	$p < 10^{-7}$	0,0064	ns.	ns.
sampquartisf0	0,1643	$p < 0,02$	$p < 0,006$	-0,0176	ns.	ns.
quan995f0	0,0814	$p < 0,10$	$p < 0,05$	-0,0259	ns.	ns.
assimf0	0,0646	ns.	$p < 0,07$	-0,0222	ns.	ns.
medderivf0	-0,0186	ns.	ns.	0,0805	ns.	$p < 0,05$
desvpaddf0	-0,0263	ns.	ns.	-0,0203	ns.	ns.
assimdf0div10	-0,0243	ns.	ns.	-0,0263	ns.	ns.
assimint	-0,0222	ns.	ns.	-0,0051	ns.	ns.
medinlinespec	0,3379	$p < 10^{-4}$	$p < 10^{-4}$	-0,0147	ns.	ns.
desvpadinlinespec	0,1793	$p < 0,005$	$p < 0,004$	0,0388	ns.	ns.
assiminlinespec	-0,0234	ns.	ns.	-0,0039	ns.	ns.
slopeLTAS	0,6785	$p < 10^{-10}$	$p < 10^{-10}$	-0,0231	ns.	ns.

Tabela 4.14: Significância estatística dos coeficientes de interceptação e inclinação da equação 4.3 para $\alpha = 10\%$.

coeficientes	significância
Interceptação	$p < 10^{-7}$
mednf0	$p < 10^{-5}$
slopeLTAS	$p < 10^{-8}$

Como se observa nas figuras 4.6 e 4.8, o aumento da intensidade relativa nas frequências mais altas do espectro dos enunciados de nosso corpus “SweEmo”, que refletiu na diminuição da inclinação do LTAS e da inclinação espectral desses enunciados, está relacionada à diminuição da percepção dos sujeitos ouvintes do grau de CALMA desses falantes. Por outro lado, como mostra a figura 4.7, o aumento do valor da mediana da frequência fundamental tendeu a ser percebido como diminuição do grau de CALMA dos falantes dos enunciados.

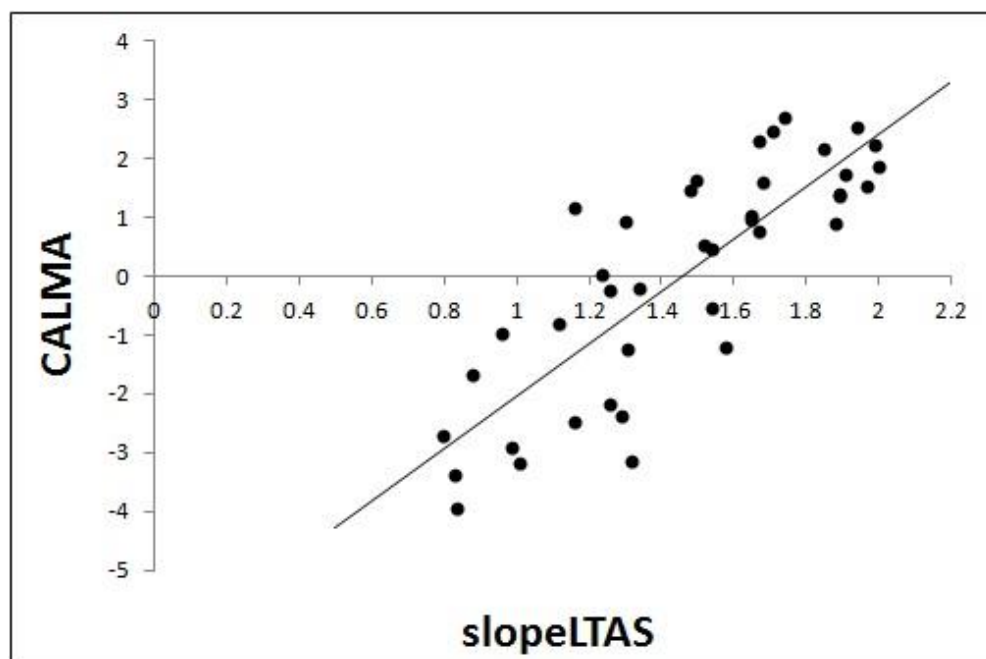


Figura 4.6: Regressão linear simples com o parâmetro “inclinação do LTAS” como variável independente e o primeiro fator da PCA com sujeitos ouvintes brasileiros (CALMA) como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 4.13.

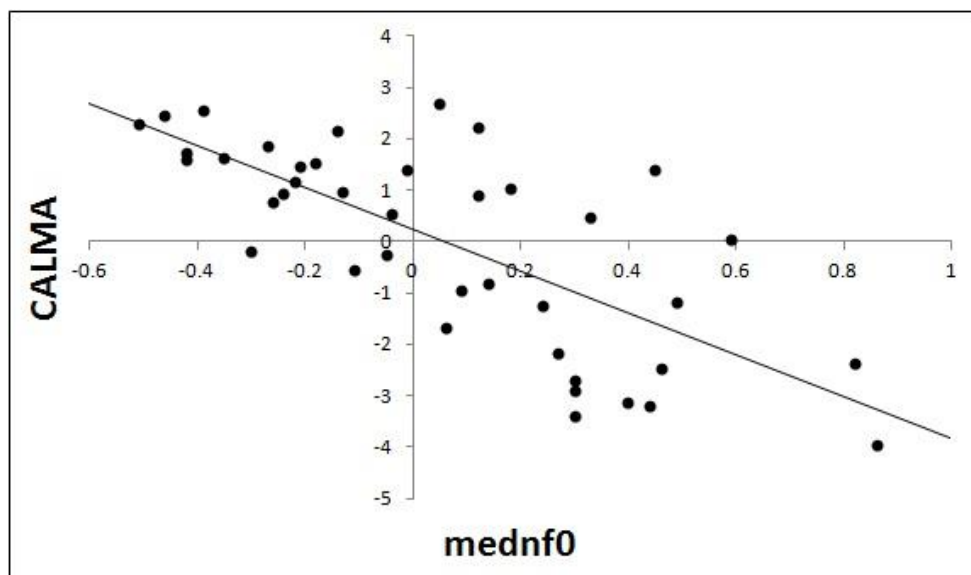


Figura 4.7: Regressão linear simples com o parâmetro “mediana de f_0 ” como variável independente e o primeiro fator da PCA com sujeitos ouvintes brasileiros (CALMA) como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 4.13.

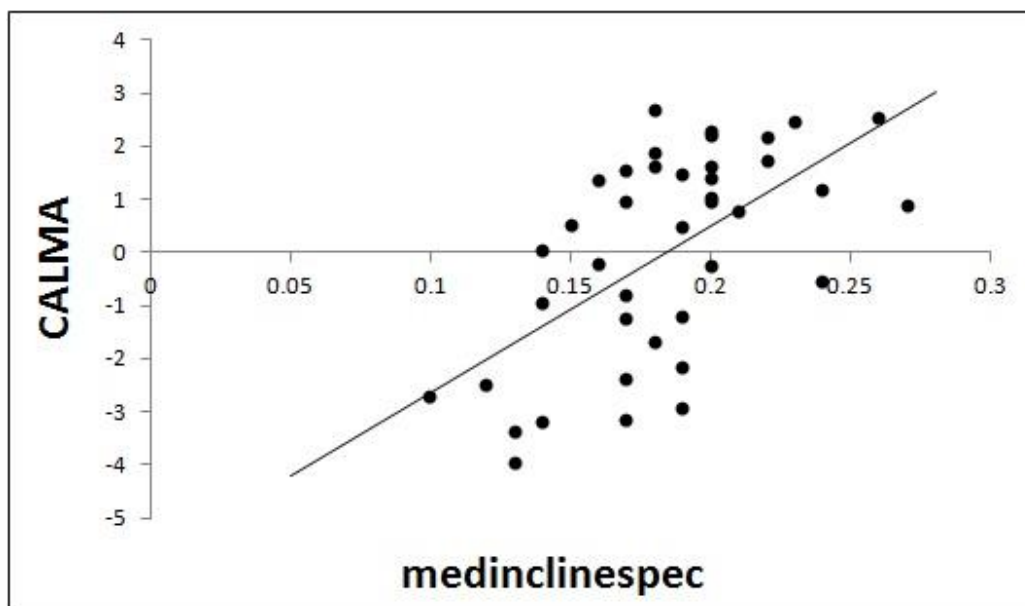


Figura 4.8: Regressão linear simples com o parâmetro “média da inclinação espectral” como variável independente e o primeiro fator da PCA com sujeitos ouvintes brasileiros (CALMA) como variável dependente. Coeficiente de determinação e significância estatística: vide tabela 4.13.

4.4.4 Classificação automática dos trechos de fala

A LDA foi empregada para classificar automaticamente os enunciados com base na combinação dos parâmetros acústicos que melhor preveem as respostas dos sujeitos ouvintes brasileiros no experimento II. A variável dependente desse teste foi composta pelas etiquetas atribuídas a cada trecho de fala de acordo com o *score* de cada um (atribuído pela PCA) para os dois primeiros componentes da PCA. Os enunciados com *score* maior do que a média dos *scores* receberam etiqueta positiva (“calma” e “satisfação”), enquanto que aqueles com *score* menor do que a média dos *scores* receberam etiqueta negativa (“não-calma” e “não-satisfação”).

Para o primeiro componente da PCA (CALMA) utilizamos como variáveis independentes os parâmetros da equação 4.3 (mediana de f_0 e inclinação do espectro médio de longo-termo). A classificação obtida pela LDA foi então comparada com a classificação baseada nos *scores* da PCA. Dos 40 trechos de fala do corpus, 35 receberam da LDA a mesma etiqueta que havia sido atribuída com base nos *scores* da PCA, o que corresponde a **87,5%** do total.

Para o segundo componente (SATISFAÇÃO) utilizamos como variável independente o parâmetro “média da primeira derivada de f_0 ” (equação 4.4). Dos 40 trechos de fala do corpus, 28 receberam da LDA a mesma etiqueta que havia sido atribuída com base nos *scores* da PCA, o que corresponde a **70%** do total.

4.5 DISCUSSÃO

No experimento relatado neste capítulo comparamos a percepção de cinco dimensões emocionais expressas em enunciados em sueco por sujeitos ouvintes suecos (falantes nativos) e brasileiros (sem conhecimento dessa língua).

A concordância entre os sujeitos ouvintes, avaliada pelo teste *kappa* de Fleiss (1971), mostra um consenso entre os ouvintes quanto ao grau de expressão dessas dimensões emocionais e indica que todas elas puderam ser inferidas a partir da fala das mulheres do corpus “SweEmo”. Entretanto, os sujeitos ouvintes brasileiros apresentaram um nível de concordância para as dimensões emocionais um pouco menor do que os

sujeitos suecos. A única exceção foi o nível de concordância apresentado por eles para a dimensão “ativação” ($k = 0,209$), que foi semelhante ao nível de concordância que os sujeitos suecos apresentaram para essa dimensão ($k = 0,208$). A dimensão “justiça”, que entre os sujeitos suecos teve o melhor grau de concordância entre as cinco dimensões ($k = 0,25$), apresentou entre os sujeitos brasileiros o índice *kappa* mais baixo ($k = 0,11$) entre todas as dimensões. Esse resultado corrobora parte de nossa terceira hipótese, que previa um melhor desempenho nos experimentos de percepção dos sujeitos avaliando trechos de fala em sua língua materna. Porém, nossa quarta hipótese, que previa uma maior concordância entre os sujeitos ouvintes quando estes avaliam as emoções expressas na fala por meio de dimensões emocionais, foi refutada, visto que tanto os sujeitos ouvintes brasileiros quanto os suecos apresentaram valores de *kappa* maiores no experimento I, no qual avaliaram as emoções descritas por adjetivos relacionados a emoções discretas. É preciso ressaltar, contudo, que essa diferença pode ter ocorrido em razão dos diferentes corpora utilizados em ambos os experimentos.

A PCA realizada com a média das respostas dos sujeitos ouvintes para as cinco dimensões revelou que os sujeitos de ambas as nacionalidades distinguiram nos enunciados um estado de “calma” (avaliado com valores maiores para as dimensões “valência” e “justiça”) de um estado de agitação emocional (avaliado pelos juízes com valores maiores para as dimensões “ativação”, “motivação” e “envolvimento”). Esse resultado é condizente com aquele obtido por Barbosa (2009). O autor conduziu um experimento em que 12 sujeitos ouvintes brasileiros julgaram o grau de expressão de quatro dimensões emocionais em 130 enunciados extraídos de um programa de rádio no qual os participantes recebiam um trote telefônico sobre uma situação específica. As dimensões investigadas foram a ativação (calmo – agitado), valência (desagradável – agradável), envolvimento (não envolvido – envolvido) e dominância (submisso – sob controle). Essas dimensões foram avaliadas pelos juízes em escalas semânticas diferenciais graduadas de 1 a 7. A dimensão de dominância foi posteriormente descartada devido à baixa coerência entre os juízes. Uma análise PCA mostrou que, assim como neste estudo, dois fatores explicaram uma grande proporção da variância total das respostas dos juízes (97%). O primeiro desses fatores, chamado pelo autor de “PRONTIDÃO” e responsável por 90% da variância total, foi uma

combinação das dimensões de ativação e envolvimento em oposição à dimensão da valência. Observa-se que esse resultado foi o mesmo que o obtido aqui, ou seja, um grau alto de agitação associado a um grau alto de envolvimento e a um grau baixo de valência, e vice-versa. O autor também extraiu desses enunciados os mesmos doze parâmetros acústicos que investigamos nesta pesquisa. Uma análise de regressão múltipla revelou que os parâmetros acústicos que melhor explicaram a variância do fator “PRONTIDÃO”, com uma correlação de 67%, foram a média da primeira derivada da frequência fundamental e a média e o desvio-padrão da inclinação espectral.

As regressões lineares realizadas com os dois primeiros componentes da PCA e os parâmetros acústicos medidos para os trechos de fala mostraram que praticamente os mesmos parâmetros explicam uma parte da variância do primeiro componente (PC1) das respostas dos sujeitos ouvintes de ambas as nacionalidades, corroborando a outra parte de nossa terceira hipótese. São eles: mediana, semiamplitude entre quartis e quantil 99,5% da frequência fundamental, média e desvio-padrão da inclinação espectral e inclinação do espectro médio de longo-termo. Os parâmetros mediana, quantil 99,5% e assimetria de f_0 e média e desvio-padrão da inclinação espectral explicam um percentual maior da variância de PC1 para as respostas dos sujeitos brasileiros do que para as respostas dos sujeitos suecos. A inclinação do espectro médio de longo-termo explica uma grande parte da variância desse componente tanto para os sujeitos brasileiros como para os suecos (cerca de 70%). A média da primeira derivada de f_0 explica parte da variância do segundo componente (PC2) para os sujeitos das duas nacionalidades, mas para os sujeitos suecos a mediana de f_0 e o desvio-padrão e a assimetria da inclinação espectral também explicam parte da variância de PC2.

Pode-se concluir, portanto, que a percepção das emoções expressas em nosso corpus com enunciados em sueco por meio de dimensões emocionais foi muito semelhante entre os sujeitos ouvintes brasileiros e suecos, o que indica que a língua materna dos sujeitos ouvintes, sua cultura e sua experiência emocional não influenciaram significativamente na percepção das emoções expressas por falantes femininos em nosso corpus “SweEmo” por meio de dimensões emocionais.

Conclusão

O estudo experimental apresentado nesta dissertação contribui para a área de pesquisa da expressividade da fala, oferecendo novos achados sobre quais parâmetros fonético-acústicos podem ser mais relevantes para identificar as emoções expressas na fala em situações de comunicação reais, bem como em quais deles os sujeitos ouvintes se guiam para inferir o grau de expressão dessas emoções. O estudo intercultural que conduzimos com sujeitos ouvintes brasileiros e suecos expandem os resultados para uma condição na qual os sujeitos ouvintes avaliaram estímulos em uma língua estrangeira da qual não possuíam nenhum conhecimento. Poucos estudos interculturais existentes na literatura compararam a percepção de emoções pela fala em uma língua estrangeira e também na língua materna dos sujeitos ouvintes.

As análises dos dois experimentos de percepção conduzidos nesta pesquisa revelaram que os julgamentos perceptivos dos sujeitos de ambas as nacionalidades foram muito semelhantes nas duas condições (com estímulos em português brasileiro e em sueco) e que não houve diferença de julgamento entre os sexos dos ouvintes. A PCA mostrou-se muito robusta em organizar ortogonalmente os estados emocionais independentes que subjazem os julgamentos dos sujeitos para as emoções expressas por etiquetas de emoções discretas e por dimensões emocionais.

No experimento I, no qual os sujeitos avaliaram o grau de expressão de emoções discretas descritas por oito adjetivos em enunciados em português brasileiro, dois componentes da PCA explicam mais do que 94% da variância total das respostas dos sujeitos das duas nacionalidades, sendo que o primeiro deles é responsável pela maior parte dessa variância (cerca de 80%). Esse componente revela dois grupos emocionais para as emoções descritas por esses adjetivos, um de prazer intrínseco positivo e outro de prazer intrínseco negativo, e que as emoções descritas pelos oito adjetivos foram avaliadas pelos sujeitos de ambas as nacionalidades conjuntamente por duas grandes dimensões emocionais: felicidade e neutralidade (a negação de qualquer expressividade). Essa divisão das oito etiquetas emocionais em dois grupos (um positivo e outro negativo) pela PCA mostra que os sujeitos ouvintes de ambas as nacionalidades avaliaram as etiquetas alegre,

contente e entusiasmado, de um lado, e comovido, triste, angustiado e aflito, do outro, de uma mesma maneira, apesar das eventuais diferenças semânticas entre elas. Isso mostra que a PCA pode evitar o problema da nomeação das emoções por meio de etiquetas discretas que surge devido às possíveis diferenças semânticas entre as línguas (ver, por exemplo, Goddard, 2002), uma vez que a(s) emoção(ões) expressa(s) nos trechos de fala que apresentaram *score* positivo para o primeiro componente (FELICIDADE), por exemplo, poderia(m) ser nomeada(s) com qualquer um dos adjetivos desse grupo positivo (alegre, contente ou entusiasmado), de acordo com a percepção dos sujeitos ouvintes. Além disso, nesse experimento, o grau de concordância foi semelhante entre os sujeitos brasileiros e suecos.

Para o experimento II, no qual os sujeitos avaliaram o grau de expressão de cinco dimensões emocionais em trechos de fala em sueco, os dois primeiros componentes da PCA explicam mais do que 95% da variância total tanto das respostas dos sujeitos suecos como das respostas dos sujeitos brasileiros. Esses dois componentes revelam que os sujeitos de ambas as nacionalidades distinguiram nos enunciados um estado de “calma” (avaliado com valores maiores para as dimensões “valência” e “justiça”) de um estado de maior agitação emocional (avaliado por eles com valores maiores para as dimensões “ativação”, “motivação” e “envolvimento”). De fato, alguns estudos mostraram que a avaliação de um evento como injusto provoca várias emoções, especialmente a raiva (Ellsworth e Scherer, 2003: 581). Os resultados desse experimento mostram que além das dimensões clássicas de ativação e valência, outras dimensões emocionais, relacionadas aos componentes emocionais de avaliação e de tendência à ação, também podem ser inferidas pela fala. Além disso, nesse experimento, os sujeitos ouvintes brasileiros apresentaram um nível de concordância para as dimensões emocionais um pouco menor do que os sujeitos suecos, com exceção da dimensão de ativação.

De um modo geral, os parâmetros acústicos mais relevantes para explicar as respostas dos sujeitos são os mesmos para as duas nacionalidades nos dois experimentos. Os resultados das análises de regressão linear entre os parâmetros acústicos e os componentes da PCA são condizentes com aqueles encontrados na literatura, segundo os quais um aumento no estado de agitação emocional do falante está relacionado a um

aumento da frequência fundamental da voz (causado pelo aumento da taxa de vibração das pregas vocais) e da energia concentrada nos harmônicos de frequências elevadas. Esses resultados mostram também que não se pode assumir que um determinado estado afetivo esteja relacionado somente a um único parâmetro acústico e que houve uma correlação maior entre os valores dos parâmetros acústicos e o primeiro componente da PCA (aquele que explica o maior percentual da variância total das respostas dos juízes das duas nacionalidades nos dois experimentos) para as dimensões emocionais. Por isso, pode-se concluir que os parâmetros acústicos foram melhores em prever os julgamentos dos sujeitos ouvintes para as dimensões emocionais do que para as etiquetas de emoções discretas.

Outro resultado a ser observado é o fato de a frequência fundamental, que é usada de maneira diferente pelo português brasileiro e pelo sueco no plano linguístico, ter sido interpretada de forma igual pelos sujeitos ouvintes brasileiros e suecos no julgamento da expressão das emoções e dimensões emocionais (plano extralinguístico) nos experimentos de percepção. Isso é mais facilmente observado pela mediana desse parâmetro, que apresentou uma correlação com os componentes da PCA relativamente elevada nos dois experimentos.

Os parâmetros acústicos mais relevantes para explicar as respostas dos sujeitos nos experimentos de percepção também foram muito robustos em classificar os trechos de fala na LDA, atingindo um elevado percentual de correspondência com a percepção dos sujeitos ouvintes para esses enunciados. Isso indica que esses parâmetros acústicos podem ser úteis em programas de reconhecimento automático de emoções e também em sistemas de síntese de fala expressiva.

Dado esses resultados, somente uma das quatro hipóteses levantadas inicialmente foi refutada. Essa hipótese corresponde àquela de número 4, que previa uma maior concordância entre os sujeitos ouvintes nos experimentos de percepção quando julgando as emoções expressas nos trechos de fala por meio de dimensões emocionais do que através de palavras que descrevem emoções discretas. A maior coerência apresentada pelos sujeitos de ambas as nacionalidades quando julgando os estímulos por meio de etiquetas relacionadas com emoções discretas pode ter ocorrido pelo fato de esses sujeitos

estarem mais familiarizados com emoções discretas do que com dimensões emocionais. Mas é possível também que esse resultado tenha relação com os diferentes corpora utilizados nos dois experimentos.

É possível concluir, portanto, que a cultura, a língua materna e a experiência emocional dos sujeitos brasileiros e suecos não influenciaram na percepção das emoções expressas nos dois corpora. Acreditamos que a razão para isso seja a seguinte: as variações que os falantes dos dois corpora apresentaram para os parâmetros acústicos investigados nesta pesquisa são provavelmente consequência das alterações neurofisiológicas que eles sofreram devido às mudanças em seu estado emocional, desencadeadas pela situação na qual se encontravam (relato de experiências traumáticas ou alegres vividas, debate sobre um tema polêmico etc.). Como essas alterações são parte do aparato biológico que a espécie humana (e também outras espécies) desenvolveu no curso de sua evolução e, por isso, universais, os sujeitos ouvintes brasileiros e suecos interpretaram as variações nos parâmetros acústicos de uma mesma maneira, apesar dos diferentes usos linguísticos que alguns desses parâmetros (como a frequência fundamental) têm no português brasileiro e no sueco, o que se refletiu em julgamentos muito semelhantes nos experimentos de percepção. Esses resultados indicam que a percepção da expressão de emoções básicas e de primitivos emocionais pela fala não depende da língua materna ou da cultura do ouvinte, pois as alterações nos parâmetros acústicos causadas pelas emoções são interpretadas universalmente de uma mesma maneira, como é o caso das expressões faciais das emoções. Estudos com outras línguas de estrutura prosódica diferente entre si devem ser conduzidos para reforçar essa conclusão.

Referências

- ABELIN, Å.; ALLWOOD, J. “Cross Linguistic Interpretation of Emotional Prosody”. In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 2000, Newcastle, Irlanda do Norte: *ISCA Archive*, 2000. p. 110 – 113.
- ABERCROMBIE, D. *Elements of general phonetics*. Edinburgh: Edinburgh University Press, 1967.
- ALM, C. O.; SPROAT, R. “Perceptions of emotions in expressive storytelling”. In: *Proceedings of Interspeech 2005*, Lisboa, Portugal, 2005. p. 533-536.
- AMIR, N. *et al.* “Unresolved Anger: Prosodic analysis and classification of speech from a therapeutical setting”. In: *Proc. of Speech Prosody 2010*, Chicago, E.U.A, 2010. Disponível em: <http://speechprosody2010.illinois.edu/program.php>. Acesso em 10 de junho de 2013.
- BANSE, R.; SCHERER, K. R. “Acoustic profiles in vocal emotion expression”. In: *Journal of Personality and Social Psychology*, v. 70, n. 3, 1996. p. 614 – 636.
- BARBOSA, P. A. “Fala expressiva: elementos básicos para sua avaliação experimental.” Manuscrito não publicado, 2008.
- _____. “Detecting changes in speech expressiveness in participants of a radio program”. In: *Proceedings of Interspeech 2009 - Speech and Intelligence*. Londres: Causal Productions, 2009. p. 2155-2158.
- _____. “Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação”. In: *Revista de Estudos da Linguagem*, v. 20, n. 1, 2012. p. 11 – 27.
- _____. “Semi-automatic and automatic tools for generating prosodic descriptors for prosody research”. In: *Tools and Resources for the Analysis of Speech Prosody Proceedings*. Aix-en-Provence: Laboratoire Parole et Langage, 2013. p. 86-89.
- BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer (Versão 5.2.25)*. Disponível em: <http://www.praat.org>. Acesso em 20 de maio de 2011.
- BRIGGS, J. L. *Never in anger: Portrait of an Eskimo family*. Cambridge, MA: Harvard University Press, 1970. *Apud* Mesquita e Frijda (1992).
- CHWALISZ, K.; DIENER, E.; GALLAGHER, D. “Autonomic arousal feedback and emotional experience: Evidence from the spinal cord injured”. In: *Journal of Personality and Social Psychology*, v. 54, 1988, p. 820-828. *Apud* Cornelius, (2000).

COHEN, J. "A coefficient of agreement for nominal scales". In: *Educational and Psychological Measurement*, v. 20, n. 1, 1960. p. 37-46.

COLAMARCO, M. *A expressão das emoções em atos de fala no Português do Brasil: produção e percepção*. Dissertação de Mestrado - Universidade Federal do Rio de Janeiro, 2009.

CORNELIUS, R. R. "Theoretical approaches to emotion". In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 2000, Newcastle, Irlanda do Norte: *ISCA Archive*, 2000. p. 3-10.

CRAWLEY, M. J. *The R book*. 1. ed. Chichester, Inglaterra: John Wiley & Sons, 2007.

DAMASIO, A. R. *Descartes' error: Emotion, reason, and the human brain*. Nova Iorque: G. P. Putnam, 1994.

DARWIN, C. *A expressão das emoções no homem e nos animais*. São Paulo: Companhia das Letras, 2009[1872].

DEMENKO, G.; JASTRZEBSKA, M. "Analysis of Voice Stress in Call Centers Conversations". In: MA, Q.; DING, H.; HIRST, D. (eds.) *Proceedings of the 6th international conference on speech prosody*. Shanghai: Tongjipress, 2012. p. 183 – 186.

EKMAN, P. "Facial expressions". In: DALGLEISH, T.; POWER, T. (eds.) *The Handbook of Cognition and Emotion*. Sussex, UK: John Wiley & Sons, 1999, p. 301-320.

_____. "An argument for basic emotions". In: *Cognition & Emotion*, v. 6, n. 3/4, 1992, p. 169 – 200.

_____. "Expression and the Nature of Emotion". In: SCHERER, K. R.; EKMAN, P. (eds.) *Approaches to emotion*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc, 1984. p. 319 - 344.

EKMAN, P.; FRIESEN, W. V. "The repertoire of nonverbal behavior: categories, origins, usage and coding". In: *Semiotica*, v. 1, n. 1, 1969, p. 49 – 98.

EKMAN, P.; FRIESEN, W. V. *Unmasking the face: A guide to recognizing emotions from facial expressions*. Nova Jersey: Prentice Hall, 1975.

EKMAN, P.; SORENSON, E. R.; FRIESEN, W. V. "Pan-Cultural Elements in Facial Displays of Emotion". In: *Science, New Series*, V. 164, N. 3875, 1969. p. 86 - 88.

ELLSWORTH, P. C.; SCHERER, K. R. "Appraisal processes in emotion". In: DAVIDSON, R. J.; SCHERER, K. R.; GOLDSMITH, H. H. (Eds.) *Handbook of affective sciences*. New York: Oxford University Press, 2003. p. 572-595.

FANT, G. *Acoustic Theory of Speech Production. With calculations based on X-ray studies of Russian articulations*. Haia, Países Baixos: Mouton, 1960.

FLEISS, J. L. “Measuring nominal scale agreement among many raters”. In: *Psychological Bulletin*, v. 76, n. 5, 1971. p. 378–382.

FONTES, M. A. S. *Gestualidade vocal e visual, expressão de emoções e comunicação falada*. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) - Pontifícia Universidade Católica de São Paulo, 2014.

FRIJDA, N. H.; MARKAM, S.; SATO, K.; WIERS, R. “Emotions and emotion words”. In: RUSSELL, J. A.; FERNANDEZ-DOLS, J. M.; MANSTEAD, A. S. R.; WELLENKAMP, J. C. (eds.) *Everyday Conceptions of Emotion: An Introduction to the Psychology, Anthropology and Linguistics of Emotion*. NATO ASI series D: Behavioural and social sciences. Nova Iorque, NY: Kluwer Academic/Plenum Publishers, v. 81, 1995, p. 121-143.

FRIJDA, N. H.; MESQUITA, B. “The social roles and functions of emotions”. In: MARKUS, H. R.; KITAYAMA, S. (eds.) *Emotion and Culture*. New York: American Psychological Association, 1994. p. 51-87.

FUJIMURA, O.; ERICKSON, D. “Acoustic Phonetics”. In: HARDCASTLE, W. J.; LAVER, J. (eds.) *The Handbook of Phonetic Sciences*. 1. ed. Oxford: Blackwell Publishers Inc., 1999.

GODDARD, C. “Explicating emotions across languages and cultures: A semantic approach”. In: FUSSELL, S. R. (org.) *The verbal communication of emotions: Interdisciplinary perspectives*. Mahwah: Lawrence Erlbaum Associates, 2002. p. 19-53.

GORDON, M.; LADEFOGED, P. “Phonation types: a cross-linguistic overview”. In: *Journal of Phonetics*, v. 29, n. 4, 2001, p. 383 – 406.

HOHMANN, G. W. “Some effects of spinal cord lesions on experienced emotional feelings”. In: *Psychophysiology*, V. 3, 1966, p. 143-156. *Apud* Cornelius, (2000).

JAMES, W. “What is an emotion?” In: *Mind*, v. 9, n. 34, 1884, p. 188-205. Disponível em <<http://www.jstor.org/stable/2246769>>, acesso em 11/02/2014.

JOHNSON, K. *Acoustic and Auditory Phonetics*. Oxford: Blackwell Publishers, 1997.

JOHNSON, G. “Theories of Emotion”. In: *Internet Encyclopedia of Philosophy*. (s. d.) Disponível em: <<http://www.iep.utm.edu/emotion/#H1>>, acesso em: 11/02/2014.

JOHNSTONE, T. *et al.* “Affective speech elicited with a computer game”. In: *Emotion*, v. 5, n. 4, 2005, p. 513 – 518.

- JOHNSTONE, T.; SCHERER, K. R. “Vocal communication of emotion”. In: LEWIS, M.; HAVILAND, J. M. (orgs.) *Handbook of emotions*. 2. ed. Nova Iorque: Guilford, 2000. p. 220-235.
- KAISER, L. “Communication of affects by single vowels”. In: *Synthese* v. 14, n. 4, 1962, p. 300–319.
- KENT, R. *The Speech Sciences*. San Diego: Singular Publishing Group, 1997.
- KENT, R. D.; READ, C. *The acoustic analysis of speech*. 1. ed. San Diego: Singular Publishing Group, 1992.
- KEHREIN, R. “The prosody of authentic emotions”. In: *Proceedings of Speech Prosody 2002*, Aix-em-Provence, França, 2002. p. 423-426.
- LAUKKA, P. “Vocal Expression of Emotion: Discrete-emotions and Dimensional Accounts”. Tese de doutorado, 2004. Uppsala University.
- LAUKKANEN, A-M; VILKMAN, E.; ALKU, P.; OKSANEN, H. “On the perception of emotions in speech: the role of voice quality”. In: *Logopedics Phoniatics Vocology*, v. 22, n. 4, 1997, p. 157-168.
- LAVIER, J.; TRUDGILL, P. “Phonetic and linguistic markers in speech”. In: SCHERER, K. R.; GILES, H. (eds.) *Social markers in speech*. Cambridge: Cambridge University Press, 1979. p. 1-32.
- LAZARUS, R. S. *Emotion and adaptation*. Nova Iorque: Oxford University Press, 1991.
- LIEBERMAN, P., MICHAELS, S. B. “Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech”. In: *Journal of the Acoustical Society of America*, v. 34, n. 7, 1962, p. 922–927.
- LUGGER, M., YANG, B. “An incremental analysis of different feature groups in speaker independent emotion recognition.” In: *Proc. XVIth ICPHs*, Saarbrücken, 2007. p. 2149–2152.
- MARASEK, K. “EGG & Voice quality”. 1997. Disponível em: <<http://www.ims.uni-stuttgart.de/institut/arbeitsgruppen/phonetik/EGG/page1.htm>> Acessado em: 18/02/2014.
- MAUSS, I. B.; ROBINSON, M. D. “Measures of emotion: A review”. In: *Cognition and Emotion*, v. 23, n. 2, 2009. p. 209-237.
- MENEZES, C.; ERICKSON, D.; HAN, J. “Cross-linguistic cross-modality perception of English sad and happy speech”. In: MA, Q.; DING, H.; HIRST, D. (eds.) *Proceedings of the 6th international conference on speech prosody*. Shanghai: Tongjipress, 2012. p. 649 – 652.

- MESQUITA, B.; FRIJDA, N. H. “Cultural Variations in Emotions: A Review”. In: *Psychological Bulletin*, v. 112, n. 2, 1992, p. 179 – 204.
- NESSE, R. “Evolutionary explanations of emotions”. In: *Human Nature*, v. 1, 1990, p. 261–289. *Apud* Johnson, G. (s. d.)
- OH, J. K. *A pausa na locução de poemas de Fernando Pessoa*. Dissertação (mestrado) - Pontifícia Universidade Católica de São Paulo, 2010.
- OSGOOD, C. E. “The nature and measurement of meaning”. In: *Psychological Bulletin*, v. 49, n. 3, 1952, p. 197-237.
- PAESCCKE, A. “Prosodische Analyse emotionaler Sprechweise”. Tese de doutorado - Technischen Universität Berlin, 2003.
- PEREIRA, C. “Dimensions of emotional meaning in speech”. In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 2000, Newcastle, Irlanda do Norte: *ISCA Archive*, 2000. p. 25-28.
- PETRUSHIN, V. A. “Emotion recognition in speech signal: experimental study, development, and application”. In: *Proc. Sixth International Conference on Spoken Language Processing (ICSLP 2000)*. Beijing, China, 2000. p. 222-225.
- PINKER, S. *Como a mente funciona*. 2 ed. São Paulo: Companhia das Letras, 2009 [1997].
- PITTAM, J.; SCHERER, K. R. “Vocal expression and communication of emotion”. In: LEWIS, M.; HAVILAND, J. M. (Eds.) *Handbook of emotions*. 1. ed. New York: Guilford Press, 1993. p. 185 – 198.
- PRINZ, J. J. *Gut reactions: A perceptual theory of emotion*. Nova Iorque: Oxford University Press, 2004. *Apud* Johnson, G. (s. d.)
- R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org>. Acessado em 2010.
- RILLIARD, A. *et al.* “Prosodic analysis of Brazilian Portuguese attitudes”. In: MA, Q.; DING, H.; HIRST, D. (eds.) *Proceedings of the 6th international conference on speech prosody*. Shanghai: Tongjipress, 2012. p. 677 – 680.
- ROSS, E. D. “The aprosodias: functional-anatomical organization of the affective components of language in the right hemisphere”. In: *Archives of Neurology*, v. 38, n. 9, 1981. p. 561-569. *Apud* Ekman (1992).
- RUSELL, J. A. “A circumplex model of affect”. In: *Journal of Personality and Social Psychology*, v. 39, 1980. p. 1161–1178.

_____. "Culture and the categorization of emotions". In: *Psychological Bulletin*, v. 110, n. 3, 1991. p. 426-450. *Apud* Goddard (2002).

SCHERER, K. R. "Speech and emotional states". In: DARBY, J. K. (ed.) *Speech evaluation in psychiatry*. Nova Iorque: Grune & Stratton, 1981. p. 189-220.

_____. "Methods of research on vocal communication: Paradigms and parameters". In: SCHERER, K. R.; EKMAN, P. (Eds.) *Handbook of methods in nonverbal behavior research*. Cambridge: Cambridge University Press, 1982. p 136-198.

_____. "On the nature and function of emotion: A Component Process approach". In: SCHERER, K. R.; EKMAN, P. (Eds.) *Approaches to emotion*. Hillsdale, NJ: Erlbaum, 1984. p. 293-317.

_____. "Vocal affect signaling: a comparative approach". In: *Advances in the Study of Behavior*, v. 15, 1985, p. 189-244.

_____. "Vocal affect expression: a review and a model for future research". In: *Psychological Bulletin*, v. 99, n. 2, 1986, p. 143-165.

_____. "Emotion". In: HEWSTONE, M.; STROEBE, W. (orgs.). *Introduction to Social Psychology: A European perspective*. 3. ed. Oxford, Inglaterra: Blackwell, 2000a. p. 151-191.

_____. "Psychological models of emotion". In: BOROD, J. (org.) *The neuropsychology of emotion*. Oxford: Oxford University Press, 2000b. p. 137-162.

_____. "Vocal communication of emotion: a review of research paradigms". In: *Speech Communication*, v. 40, n. 2, 2003, p. 227-256.

_____. "What are emotions? And how can they be measured?" In: *Social science information*, v. 44, n. 4, 2005, p. 695-729.

SCHERER, K. R.; BANSE, R.; WALLBOTT, H. G. "Emotion Inferences from Vocal Expression Correlate across Languages and Cultures". In: *Journal of Cross-Cultural Psychology*, v. 32, n. 1, 2001, p. 76 – 92.

SCHIMMACK, U. "Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive?" In: *Cognition and Emotion*, v. 15, n. 1, 2001, p. 81-97.

SCHLOSBERG, H. "Three dimensions of emotion". In: *Psychological Review*, v. 61, n. 2, 1954. p. 81-88.

_____. "A scale for the judgement of facial expressions". In: *J. Exp. Psychology*, v. 29, 1941. p. 497-510.

SMITH, C. A.; ELLSWORTH, P. C. "Patterns of cognitive appraisal in emotion". In: *Journal of Personality and Social Psychology*, v. 48, n. 4, 1985. p. 813-838.

SOLOMON, R. C. "The Philosophy of Emotions". In: LEWIS, M.; HAVILAND-JONES, J. M.; BARRETT, L. F. (orgs.) *Handbook of emotions*. 3 ed. Nova Iorque: Guilford, 2008. p. 3-16.

STEVENS, K. N. *Acoustic phonetics*. Cambridge: MIT Press, 1998.

TRAUNMÜLLER, H; ERIKSSON, A. "Acoustic effects of variation in vocal effort by men, women, and children". In: *The Journal of the Acoustical Society of America*, v. 107, n. 6, 2000, p. 3438-3451.

VOGT, T.; ANDRÉ, E.; WAGNER, J. "Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation". In: PETER, C.; BEALE, R (eds.) *Affect and Emotion in Human-Computer Interaction: From Theory to Applications (Lecture Notes in Computer Science)*. Heidelberg, Germany: Springer, v. 4868, 2008, p. 75-91.

XU, Y.; KELLY, A.; SMILLIE, C. "Emotional expressions as communicative signals". In: HANCIL, S.; HIRST, D. (orgs.) *Iconicity in Language and Literature: Prosody and Iconicity*. v. 13. Amsterdam: John Benjamins Publishing Company, 2013, p. 33 – 60.

WILTING, J.; KRAHMER, E.; SWERTS, M. "Real vs. acted emotional speech". In: *Proc. of INTERSPEECH 2006*, Pittsburgh, E.U.A: ISCA Archive: 2006. Disponível em: <http://www.isca-speech.org/archive/interspeech_2006>, acessado em junho de 2011.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938. *Apud* Schlosberg (1941).

WUNDT, W. *Grundzüge der physiologischen Psychologie*. 5 ed. Leipzig: Engelmann, 1905. *Apud* Scherer (2000b).

ANEXOS

ANEXO I – Script “*Expression Evaluator*” de Barbosa (2009)

```
# ExpressionEvaluatorbySG.psc
# Script implemented by Plinio A. Barbosa (IEL/Unicamp) for obtaining
# normalised values of statistical descriptors of acoustic phonetic parameters
# such as f0, df0, spectral tilt, LTAS (skewness de df0 was divided by 10).
# Copyright (C) 2008, 2011 Barbosa, P. A.
# This program is free software; you can redistribute it and/or modify
# it under the terms of the GNU General Public License as published by
# the Free Software Foundation; version 2 of the License.
# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
# GNU General Public License for more details.
# Do not distribute without the author's previous authorisation
# Contact e-mail: pabarbosa.unicampbr@gmail.com
# The following form refers to the name of the TXT output file and edge cut-off
# frequencies for f0 analysis
form File Acquisition
word Fileout_(with_extension) narracaoout.txt
word File_(with_extension) narracao.wav
word Filetextgrid narracao.TextGrid
positive ExtractionTier 4
word Speaker_sex Female
real left_Cut_off_frequency_(Hz) 0 (= auto)
real right_Cut_off_frequency_(Hz) 0 (= auto)
endform
# The following lines assign the mean and sd values for f0 according to gender
if speaker_sex$ == "Female"
  meanfundfreq = 231
  sfundfreq = 120
else
  meanfundfreq = 136
  sfundfreq = 58
endif
filedelete 'fileout$'
fileappend 'fileout$' chunk, mednf0, sampquartisf0, quan995f0, assimf0, medderivf0,
desvpaddf0, assimdf0div10, assimint, medinclinespec, desvpadinclinespec,
assiminclinespec, slopeLTAS 'newline$'
# Defines default values in Hz for f0 search according to gender
male_default_left = 75
male_default_right = 360
female_default_left = 110
female_default_right = 700
```

```

Read from file... 'filetextgrid$'
filenametg$ = selected$("TextGrid")
Read from file... 'file$'
filename$ = selected$("Sound")
plus TextGrid 'filenametg$'
Extract non-empty intervals... 'extractionTier' no
nsounds = numberOfSelected("Sound")
for i to nsounds
  sound'i' = selected("Sound",i)
  label'i'$ = selected$("Sound",i)
endfor
isounds = 1
repeat
# Assumes the default values for f0 normalisation (mean, sd), according to the
automatically defined gender
if left_Cut_off_frequency = 0 ; automatic
  left_Cut_off_frequency = if speaker_sex$ = "Male" then 'male_default_left' else
'female_default_left' fi
endif
if right_Cut_off_frequency = 0 ; automatic
  right_Cut_off_frequency = if speaker_sex$ = "Male" then 'male_default_right' else
'female_default_right' fi
endif
select sound'isounds'
# Computes f0 contour defining boundaries for computation according to gender, then
smooth it using a 10-Hz LP filter
To Pitch... 0.0 'left_Cut_off_frequency' 'right_Cut_off_frequency'
Smooth... 10
quantile995f0 = Get quantile... 0.0 0.0 0.995 Hertz
# Computes the z-scored value of the 99.5% quantile
quantile995f0 = (quantile995f0 - meanfundfreq)/sfundfreq
# Computes z-scored f0 mean, median, quantiles 5%, and 95%, semi-amplitude between
the latter quantiles, and modified Pearson's f0 skewness
meanf0 = Get mean... 0.0 0.0 Hertz
meanf0 = (meanf0 - meanfundfreq)/sfundfreq
medianf0 = Get quantile... 0.0 0.0 0.50 Hertz
medianf0 = (medianf0 - meanfundfreq)/sfundfreq
quantile05f0 = Get quantile... 0.0 0.0 0.05 Hertz
quantile05f0 = (quantile05f0 - meanfundfreq)/sfundfreq
quantile95f0 = Get quantile... 0.0 0.0 0.95 Hertz
quantile95f0 = (quantile95f0 - meanfundfreq)/sfundfreq
sampquantf0 = (quantile95f0 - quantile05f0)/2
skewnessf0 = (meanf0 - medianf0)/sampquantf0
# Starts computation of f0 discrete derivative
Down to PitchTier

```

```

npoints= Get number of points
meandf0 = 0
f0ant = Get value at index... 1
for i from 2 to npoints/2
    j = 2*i - 1
    f0 = Get value at index... 'j'
# Computes f0 discrete derivative, and its cumulative value
    df0'i' = f0 - f0ant
    meandf0 = meandf0 + df0'i'
    f0ant = f0
endfor
nvalues = npoints/2 - 1
meandf0 = meandf0/nvalues
# Computes f0 discrete derivative standard deviation
sdf0 = 0
for i from 2 to npoints/2
    sdf0 = sdf0 + (df0'i' - meandf0)*(df0'i' - meandf0)
endfor
sdf0 = sqrt(sdf0/(nvalues-1))
# Computes f0 discrete derivative skewness
skdf0 = 0
for i from 2 to npoints/2
    skdf0 = skdf0 + ((df0'i' - meandf0)/sdf0)^3
endfor
skdf0 = (nvalues/((nvalues-1)*(nvalues-2)))*skdf0
# Divides skdf0 by 10 just for the sake of graphic visualization
skdf0 = skdf0/10
# Normalises df0 standard deviation value
meandf0 = meandf0/meanfundfreq
sdf0 = sdf0/meanfundfreq
select sound'isounds'
# Computes the intensity of the sound file
To Intensity... 100 0.0 yes
# Computes intensity mean, median, standard deviation, and Pearson's intensity skewness
meanint = Get mean... 0.0 0.0 energy
medint = Get quantile... 0.0 0.0 0.50
sdevint = Get standard deviation... 0.0 0.0
skewint = (meanint - medint)/sdevint
# Low-pass filters sound file under 1,250 Hz (around F1 for [a])
select sound'isounds'
Filter (pass Hann band)... 0 1250 100
Rename... "LB"
To Intensity... 100 0.0 no
# Band-pass filters sound file between 1,250 and 4,000 Hz
select sound'isounds'

```

```

Filter (pass Hann band)... 1250 4000 100
Rename... "HB"
To Intensity... 100 0.0 no
nframes= Get number of frames
meandint = 0
for i from 1 to nframes/10
  j = i*10
  select Intensity _HB_
  highval'i= Get value in frame... 'j'
  select Intensity _LB_
  # Lowval is the estimation of spectral tilt normalised by the median value of intensity
  lowval'i= Get value in frame... 'j'
  lowval'i = (lowval'i' - highval'i')/medint
  meandint = meandint + lowval'i'
endfor
# Computes the mean of normalised spectral tilt
meandint = 10*meandint/nframes
# Computes the standard deviation of normalised spectral tilt
sdint = 0
for i from 1 to nframes/10
  sdint = sdint + (lowval'i' - meandint)*(lowval'i' - meandint)
endfor
sdint = sqrt(sdint/(nframes/10-1))
# Computes the skewness of normalised spectral tilt
skdint = 0
for i from 1 to nframes/10
  skdint = skdint + ((lowval'i' - meandint)/sdint)^3
endfor
vald10 = nframes/10
skdint = (vald10/((vald10-1)*(vald10-2)))*skdint
select sound'isounds'
# Computes the long term spectrum, and gets its standard deviation
Scale intensity... 70
To Ltas... 100
sltas = Get slope... 0 1000 1000 4000 energy
# Normalises the LTAS slope
nsltas = -sltas/10
temp$ = label'isounds'$
meandf0 = meandf0*1000
sdf0 = 10*sdf0
fileappend 'fileout$' 'temp$', 'medianf0:2', 'sampquantf0:2', 'quantile995f0:2',
'skewnessf0:2', 'meandf0:2', 'sdf0:2', 'skdf0:2', 'skewint:2', 'meandint:2', 'sdint:2', 'skdint:2',
'nsltas:2' 'newline$'
isounds = isounds + 1
until isounds > nsounds

```


ANEXO II – Transcrição dos enunciados do corpus “Jogo de Cena”

01 – Desde desse dia que ele me deu, pô... Essa oportunidade, eu tô lá já... Dez anos.

02 – Lá eu aprendi muitas coisas. Lá eu aprendi a ser gente, aprendi a ser mulher, aprendi a ler. A ler um texto, a interpretar.

03 – E aí vai atrás dele e conhece uma louca no caminho, que é fantástica. E a louca, ela perdia a memória instantaneamente. Tinha tipo doença de Alzheimer.

04 – Plasticamente e tal... E aí ele vai brincar de gracinha e pescam ele, carregam ele pra Austrália da costa americana e ele vai parar num aquário na Austrália.

05 – O único objetivo que eu tenho na vida é resgatar isso, nem que seja a última coisa que eu faça.

06 – Mas tipo assim, eu num posso fazê o *beatbox*. Ou eu canto, ou eu faço o *beatbox*.

07 – Porque descobriu que a menina de dezoito anos traía ele com uma meia dúzia de vizinho. Aí ele pediu à minha mãe arrego e minha mãe deixou ele voltar.

08 – Foi tudo de ruim, olha, ele... Ele me deixava... Logo no começo ele me deixava sozinha em casa...

09 – Num sei, porque ela me perguntou e eu fui me embolando... Eu sai assim, caramba, no final num contei nada, fracionei um monte de história...

10 – Horrível, horrível. Mas assim, eu num via saída sabe. Num via saída.

11 – Pra fazê o exame, que era vinte e sete, né... Aí eu fui fazê o exame, aí deu positivo, aí o mundo caiu, né...

12 – Isso é um problema que eu tô agora, um pequeno problema, dentro de um monte de coisa boa. Mas que antes dela, eu num ia tê nem como chegar a esse problema porque eu num teria esse monte de coisa boa que eu tô tendo agora...

13 – Meu filho tinha reagido a um assalto quando ele vinha do serviço e aí mataram ele. Aí foi uma coisa assim muito louca, muito...

14 – Uma coisa doida, meu filho tinha dezenove anos, ele era lindo. A gente tava num momento tão feliz, tão bom, tão...

15 – Caramba! Olha, eu faço parte de um grupo de dança, dessa dança de salão.

16 – Mas ele já me perdoou, já tá tudo bem já...

- 17 – Eu também tô...
- 18 – Meu filho trabalhava de motorista de caminhão. Ele tinha dezenove anos.
- 19 – Veio em busca de uma vida boa, eles viviam... Se a Grécia e a Turquia brigavam os judeus que eram assassinados.
- 20 – Tá difícil. A gente vê um monte, mas a maioria tá comprometido... Ou então já tá com caso, ou então...
- 21 – Que ele chega do trabalho e bota as moedas ali... Ai meu Deus, deve achar até hoje que foi a empregada que roubou.
- 22 – Aí quando eu acordei eu falei pra ela: Paula, eu tive um sonho lindo com meu filho hoje. Então eu contei pra ela, aí ela: Ai mãe, que bom! Ah, eu digo, então, a partir de hoje a gente vai...
- 23 – Eu cheguei em casa, liguei a televisão no Vídeo Show, e tava passando o grupo de teatro “Nós do morro”, e era no Vidigal...
- 24 – Em outro país... Estudar fora... Eu tava esperando, queria esperar fazê dezoito anos, pra podê tentá, né, uma chance fora do Brasil.
- 25 – Converteu ao judaísmo e era uma mulher inteligentíssima. Foi filha de diplomata, né, meu avô serviu na Europa.
- 26 – Eles podem meio que se sentir até mal, mas isso que eu tive que passá é uma coisa que me dá força mesmo pra mim podê sê o que eu sô hoje, entendeu, porque eu pego força nisso, entendeu...
- 27 – A minha mãe fazia costuras, costuras assim... Pra vizinhos. Meu pai num gostava que ela trabalhasse fora, preferia que ela cuidasse da casa.
- 28 – Meu pai já era mais pra esse lado, história da arte... Então juntô tudo, entendeu, lá em casa, assim, daí eu lia de tudo né...
- 29 – Bom, eu gostava do meu pai. Daí um dia eu acabei brigando com ele, eu era muito pequena... Era muito pequena, assim, devia tê uns...
- 30 – Eles acordavam cedo pra trabalhá. Meu filho acordava quatro horas da manhã pra trabalhá. Minha filha acordava mais tarde, que ela pegava mais tarde. E aí, a gente levantava, fazia café, tomava café todo mundo juntos.

ANEXO III – Instruções para o experimento I com sujeitos brasileiros

Instruções Gerais

Você irá participar de um teste sobre a percepção de emoções na fala. Serão apresentados vários sons, correspondentes a trechos de fala. Após escutar cada som, você deverá indicar o grau da emoção que o falante daquele trecho estava sentindo, escolhendo um valor na escala que vai de 0 a 4. É permitida a escolha de valores intermediários nessa escala.

Para cada parte do teste, você deverá avaliar apenas uma palavra, que será dada. Por exemplo: para a parte do teste em que se pede para avaliar a ALEGRIA, avalie o grau de alegria que você acha que o falante estava sentindo. Serão 4 palavras para avaliar.

Atenção: O áudio começará a tocar **automaticamente** assim que a página terminar de carregar. Se precisar ouvi-lo novamente, clique no botão de "play" do Audio Player, mas faça isso somente 1 vez.

A duração do teste é de aproximadamente 25 minutos. Utilize fone de ouvido e faça o teste em um ambiente silencioso. Não tenha pressa! **CONCENTRE-SE NA VOZ DO FALANTE, E NÃO NO QUE ELE ESTÁ DIZENDO**. Faça o teste sozinho (a).

O experimento não permite que você volte na página anterior. Portanto, **não** clique no botão de voltar do seu navegador, caso contrário ocorrerá um erro.

Muito obrigado e bom teste!!! Clique em "Next" para prosseguir.

Parte 1: ALEGRIA – Instruções

Nesta parte do teste, você deverá julgar o grau de **ALEGRIA** que os falantes estavam sentindo quando produziram os enunciados.

O áudio começará a tocar automaticamente assim que a página terminar de carregar. Se precisar ouvi-lo novamente, clique no botão de "play" do Audio Player, mas faça isso somente 1 vez.

Após ouvir o áudio, selecione um valor na escala, que vai de 0 (nada alegre) a 4 (muito alegre). Atenção: você pode também escolher valores intermediários! Lembre-se: concentre-se na voz do falante, e não no que ele está dizendo. Após marcar sua resposta na escala, clique no botão "next".

Agora, clique no botão "next" para começar essa parte do teste.

ANEXO IV - Instruções para o experimento I com sujeitos suecos

Allmänna instruktioner

Du kommer att delta i ett experiment om uppfattningen av känslor i tal. Ett antal ljud som utgör talprover kommer att spelas upp. Efter att ha lyssnat på varje ljud ska du välja hur starkt du tycker att talaren upplevde den känsla som visas i skrift på skärmen. Detta gör du genom att välja ett värde på en skala från 0 till 4. Du kan också välja mellanliggande värden på denna skala.

I varje del av testet ska du bedöma endast ett ord, som kommer att anges. Till exempel: I den del av experimentet som ber dig att bedöma ordet **GLÄDJE**, ska du bedöma graden av glädje som du tycker att talaren uttrycker i ljudprovet. Det kommer att finnas fyra ord att utvärdera.

Obs! Ljudet spelas upp **automatiskt** så fort sidan har laddats klart. Om du behöver höra ljudet igen, klicka på "play"-knappen på spelaren, men gör det bara en gång.

Experimentet tar ca 20 minuter att slutföra. Använd hörlurar och gör experimentet i en lugn miljö. Ta den tid du behöver. **FOKUSERA PÅ TALARENS RÖST OCH INTE PÅ VAD HAN/HON SÄGER.** Gör experimentet ensam.

Systemet tillåter inte att du går tillbaka till föregående ljud. Därför ska du **inte** klicka på "bakåt"-knappen i din webbläsare, eftersom det då uppstår ett fel och alla dina svar försvinner.

Klicka på "next"-knappen nedan för att fortsätta.

Del 1: **GLÄDJE** – Instruktioner

I denna del av testet kommer du att bedöma graden av **GLÄDJE** som talarna uttrycker i yttrandena.

Ljudet spelas upp automatiskt så fort sidan har laddat klart. Om du behöver höra det igen, klicka på "play"-knappen för AudioPlayer, men gör det bara en gång.

Efter att du har hört ljudet, välj ett värde på skalan, som sträcker sig från 0 (inte alls glad) till 4 (mycket glad). Obs: Du kan också välja mellanliggande värden! Kom ihåg: Fokusera på talarens röst, och inte på vad han/hon säger. Efter att ha markerat ditt svar på skalan, klicka på "next"-knappen för att lyssna på nästa ljud.

Klicka på "next"-knappen nedan för att börja denna del av experimentet.

ANEXO V - Instruções para o experimento II com sujeitos brasileiros

Instruções Gerais

Você irá participar de um teste sobre a percepção de emoções na fala, as quais serão avaliadas de forma indireta, através de dimensões emocionais. Serão apresentados vários sons, correspondentes a trechos de fala. Após escutar cada som, pedimos que responda a algumas perguntas relativas ao estado do falante ou à situação na qual o falante estava, selecionando um valor em uma escala graduada de 0 a 4. Você também pode selecionar valores intermediários nessa escala.

O experimento consiste em cinco partes. Para cada uma, você responderá a uma pergunta específica.

Atenção: O áudio será reproduzido **automaticamente** assim que a página terminar de carregar. Se precisar ouvi-lo novamente, clique no botão de “play” do Audio Player, **mas faça isso somente 1 vez.**

A duração do teste é de aproximadamente 30 minutos. Utilize fone de ouvido e faça o teste em um ambiente silencioso. Não tenha pressa! **PROCURE CONCENTRAR-SE NA VOZ DO FALANTE, E NÃO NO QUE ELE ESTÁ DIZENDO.** Faça o teste sozinho (a).

O experimento não permite que você volte na página anterior. Portanto, **não** clique no botão de voltar do seu navegador, caso contrário ocorrerá um erro.

Muito obrigado e bom teste!!! Clique no botão "Next" abaixo para prosseguir.

PARTE 1: ATIVAÇÃO – Instruções

Nesta parte do experimento, pedimos que você responda à seguinte pergunta:

Quão agitado estava o falante?

O áudio será reproduzido **automaticamente** assim que a página terminar de carregar. Se precisar ouvi-lo novamente, clique no botão de “play” do Audio Player, **mas faça isso somente 1 vez.**

Após ouvir o som, selecione um valor na escala graduada de 0 (Nada agitado) a 4 (Muito agitado). **Obs.:** Você também pode selecionar valores intermediários nessa escala!

Lembre-se: concentre-se na voz do falante, e não no que ele está dizendo. Após marcar sua resposta na escala, clique no botão "next" para ouvir o próximo som.

Agora, clique no botão "next" abaixo para começar esta parte do experimento.

ANEXO VI - Instruções para o experimento II com sujeitos suecos

Allmänna instruktioner

Du kommer att ta del i ett experiment om upplevelse av känslor i tal. Du kommer att bedöma känslorna på ett indirekt sätt, genom känslodimensioner. Vissa ljud kommer att motsvara talat material. Efter att du har lyssnat på varje ljud ber vi dig att svara på frågor som gäller talaren eller den situation som talaren befinner sig i. Du svarar genom att välja ett värde på en skala från 0 till 4. Du kan också välja mellanliggande värden på skalan.

Experimentet består av fem delar. För var och en av delarna ska du besvara en specific fråga.

Obs! Ljudet spelas upp automatiskt så fort sidan har laddats klart. Om du behöver höra ljudet igen, klicka på "play"-knappen på spelaren, **men gör det bara en gång.**

Experimentet tar ca 30 minuter att slutföra. Använd hörlurar och gör experimentet i en lugn miljö. Ta den tid du behöver. **FOKUSERA PÅ TALARENS RÖST OCH INTE PÅ VAD HAN/HON SÄGER.** Gör experimentet ensam.

Systemet tillåter inte att du går tillbaka till föregående ljud. Därför ska du inte klicka på "bakåt"-knappen i din webbläsare, eftersom det då uppstår ett fel och alla dina svar försvinner.

Klicka på "next"-knappen nedan för att fortsätta.

DEL 1: AKTIVERING - Instruktioner

I denna del av testet ber vi dig att svara på följande fråga:

Hur exalterad var talaren?

Ljudet spelas upp automatiskt så fort sidan har laddat klart. Om du behöver höra det igen, klicka på "play"-knappen för AudioPlayer, men gör det bara en gång.

Efter att du har hört ljudet, välj ett värde på skalan, som sträcker sig från 0 (inte alls exalterad) till 4 (mycket exalterad). **Obs:** Du kan också välja mellanliggande värden!

Kom ihåg: Fokusera på talarens röst, och inte på vad han/hon säger. Efter att ha markerat ditt svar på skalan, klicka på "next" för att lyssna på nästa ljud.

Klicka på "next"-knappen nedan för att börja denna del av experimentet.