UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS CLÁSSICAS E VERNÁCULAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOLOGIA E LÍNGUA PORTUGUESA

MARCUS VINÍCIUS MOREIRA MARTINS

**Alvos tonais: unidades fonético-fonológicas da
entoação**

São Paulo

2017

Marcus Vinícius Moreira Martins

ALVOS TONAIIS: UNIDADES
FONÉTICO-FONOLÓGICAS DA ENTOAÇÃO

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

Orientador:
Prof^o. Dr^o. Waldemar Ferreira Netto

São Paulo - SP

2017

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

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

M379a Martins, Marcus Vinicius Moreira
 Alvos tonais: unidades fonético-fonológicas da
 entoação / Marcus Vinicius Moreira Martins ;
 orientador Waldemar Ferreira-Netto. - São Paulo,
 2017.
 275 f.

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

 1. Fonética Acústica. 2. Fonologia. 3. Entoação.
 4. Emoções. 5. Português Brasileiro. I. Ferreira-
 Netto, Waldemar, orient. II. Título.

Aprovado em/...../.....

Prof. Dr. Waldemar Ferreira Netto
Departamento de Letras Clássicas e Vernáculas - USP
Orientador

Prof. Dr.
Instituição:
Julgamento:

Prof. Dr.
Instituição:
Julgamento:

Prof. Dr.
Instituição:
Julgamento:

Prof. Dr.
Instituição:
Julgamento:

Aos meus avós, Maria e Francisco.

Agradecimentos

Quando visitei a USP pela primeira vez tinha 9 anos de idade e disse aos meus pais que um dia estudaria aqui. Dezenove anos depois daquele dia, fechando todo um ciclo nessa Universidade, só posso agradecer a eles, Cida e Marcos, pelo apoio incondicional, pelas orientações e pela confiança que recebi ao longo de todo esse tempo. Meus irmãos, Paulo e Larissa, também são partes fundamentais desse processo, a eles fica o agradecimento pelo companheirismo e pela paciência com os meus horários alternativos desde sempre.

Agradeço ao meu orientador, Waldemar, que ao longo desses anos me mostrou o quanto um fato linguístico não é só um fato linguístico, essa visão foi fundamental para que eu desenvolvesse qualquer hipótese ou trabalho sobre a linguagem. Além disso, sua leitura atenta e orientação foram fundamentais para que eu terminasse esse trabalho.

Aos professores André Souza e Rosane de Sá Amado, pela leitura da minha qualificação, suas orientações me ajudaram a começar a formalizar um raciocínio mais claro sobre um tema tão árduo. Os erros que permanecem são, sem dúvidas, meus.

Ao meus colegas do grupo ExProsodia, em especial ao Daniel e à Renata, que sempre me ajudaram a navegar pelas águas turbulentas da Fonética e Fonologia com suas análises e ideias.

Aos meus colegas do Grupo de Estudo de Fonética da USP, pelas discussões interdisciplinares que enriqueceram minha visão sobre o papel da linguagem nas mais diversas esferas. À Beatriz, que foi minha orientadora de mestrado e continua me orientando, demonstrando a importância da escrita orientada ao leitor, principalmente,

quando andamos nas margens da interdisciplinariedade.

Agradeço também à Aline Vassoler, importante interlocutora nesse tema e me ajudou com seus dados e suas análises a ter uma visão mais ampla dos problemas relacionados à entoação. Sem essa interlocução esse trabalho seria algo menor.

Ao pessoal do DLCV que sempre me ajudou a decifrar prazos e regulamentos, que por algum motivo sempre foram complexos para mim.

À USP e à CAPES pelo apoio institucional e financeiro.

Por fim, aos meus amigos de Mogi das Cruzes, da USP e de tantos outros lugares que sempre demonstraram apoio e companheirismo ao longo desses últimos 10 anos. Em especial, ao Raphael, à Mayara e ao Rodolfo pelo carinho e estima.

Stephen Fry: *Oh, of course it is, of course it is, of course it is, of course it is . . .
Language is my mother, my father, my husband, my brother, my sister, my whore,
my mistress, my check-out girl . . . a complimentary moist lemon-scented cleansing
square or handy freshen-up wipette . . . Language is the breath of God.*

A Bit of Fry and Laurie - Flexibility of Language

Resumo

MARTINS, M.V.M. **Alvos tonais: unidades fonético-fonológicas da entoação**. Tese (Doutorado). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2017.

O objetivo deste trabalho é desenvolver nossa hipótese em torno dos alvos tonais, que seriam unidades fonético-fonológicas responsáveis pela mediação entre o nível representacional da entoação e o nível físico da implementação de F_0 . Os alvos tonais foram divididos em duas grandes categorias, topológicos e pontuais. Os primeiros ocorrem em um espaço limitado, denominado por Ferreira-Netto (2008) de tom médio. Os limites desse espaço são definidos pelos limiares de diferenciação tonal (LDT) e estão a +3 e -4 semitons do tom médio. Além e aquém destes limiares encontram-se as faixas de frequência do Foco/Ênfase. Nestas faixas os tons passam a ser eventos relevantes para os ouvintes que podem atribuir significados a eles. Aos tons que ocorrem nessa região damos o nome de pontual, uma vez que são eventos específicos. Para testar essa hipótese aplicamos o teste 1, no qual era solicitado aos participantes que repetissem uma sentença pré-gravada dotada de uma divisão entoacional marcante: na primeira parte tratava-se de uma voz masculina entoando uma frase declarativa, na segunda parte uma voz feminina entoando uma frase interrogativa. De um total de 15 participantes obtivemos 24 amostras, contando repetições intra-sujeitos. A análise foi conduzida em duas etapas, na primeira avaliamos a capacidade do falante detectar o alvo tonal topológico subjacente à primeira parte do estímulo e reproduzi-lo. Na segunda etapa, avaliamos a capacidade do falante em detectar o alvo pontual caracterizado pela interrogativa e implementá-lo em sua fala. A análise da primeira condição foi feita por meio do que denominamos índice de relação (i_r), que media o

grau de correlação entre o estímulo e a repetição do falante. A análise revelou que os participantes demonstraram uma grande acuidade na execução da tarefa, o que sugere que os falantes são capazes de monitorar a implementação da frequência fundamental, a partir da detecção dos alvos topológicos. Já a segunda análise demonstra que a implementação dos alvos pontuais pode ser aleatória em certa medida, uma vez que ela não precisa respeitar um limite específico, apenas um limiar. Na segunda parte do trabalho aplicamos um método semelhante, voltado à análise da fala emotiva atuada em três condições: raiva, tristeza e neutra. A frase consistia de um trecho de um livro de ciências lido nessas três emoções por atrizes profissionais, a análise por meio de testes de hipótese ($n=196$, $p<0,005$) revelou que os alvos topológicos entre as condições eram distintos, o que sugere que o espaço entoacional e a variação de frequência em seu interior podem ser uma pista significativa para a distinção da fala emotiva.

Palavras-chave: fonética acústica; fonologia; entoação; emoções; português brasileiro.

Abstract

MARTINS, M.V.M. **Tonal targets: phonological-phonetic units of intonation.** Tese (Doutorado). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2017.

The main purpose of this work is develop our hypothesis about the tonal targets, which would be phonological-phonetic units responsible for the mediation between the representational level of the intonation and the physical level of the implementation of F_0 . The tonal targets were divided into two major categories, topological and punctual. The first occur in a limited space, and is called by Ferreira-Netto (2008) as midtone. The boundaries of this space are defined by the tonal differentiation thresholds (TDT) and are specified at +3 and -4 semitones from the midtone. Beyond these thresholds are the Focus/Emphasis frequency bands. In these bands the tones become relevant events to the listeners who are able to attribute meanings to them. the tones that occur in this region we gave the name of punctual, since they are specific events. To test this hypothesis, we applied test 1, in which participants were asked to repeat a pre-recorded sentence with a striking intonational division: in the first part it was a male voice spealing a declarative phrase, in the second part a female voice spea-king an interrogative phrase. From 15 participants we obtained 24 samples, counting intra-subject repetitions. The analysis was conducted in two stages, in the first one we evaluated the ability of the speaker to detect the topological tonal target underlying the first part of the stimulus and to reproduce it. In the second step, we evaluated the ability of the speaker to detect the punctual target characterized by the interrogative and implement it in his speech. The analysis of the first condition was done by means of what we call the relationship index (i_r), which measures the degree of correlation

between the stimulus and the repetition of the speaker. The analysis revealed that the participants demonstrated a great acuity in the execution of the task, which suggests that the speakers are able to monitor the fundamental frequency implementation, from the detection of the topological targets. The second analysis, on the other hand, shows that the implementation of specific targets can be random to some extent, since it does not need to respect a specific limit, only a threshold. In the second part of the work we apply a similar method, focused on the analysis of emotional speech, in three conditions: anger, sadness and neutrality. The phrase consisted of an excerpt from a science book, read in these three emotions by professional actresses. The analysis, using hypothesis tests ($n = 196$, $p < 0.005$), revealed that the topological targets between the conditions were distinct, suggesting the intonational space and the variation of frequency in its interior can be a significant clue to the distinction of the emotive speech.

Keywords: acoustic phonetics; phonology; intonation; emotions; brazilian portuguese.

Sumário

Lista de Figuras	p.7
Lista de Tabelas	p. 12
Lista de Abreviaturas	p. 15
1 Introdução	p. 16
1.1 Apresentação	p. 16
1.2 Nossos objetivos	p. 17
1.3 Fala emotiva	p. 22
1.4 Descrição dos capítulos	p. 24
2 Entoação	p. 26
2.1 A produção da entoação	p. 26
2.1.1 A Laringe	p. 28
2.1.2 A glote	p. 30
2.1.3 A produção da frequência fundamental	p. 33
2.2 Abordagens da entoação	p. 38
2.2.1 Definições da entoação	p. 38
2.2.2 Níveis vs Configuração	p. 41

2.2.3	O modelo de Pierrehumbert (1980)	p.45
2.2.4	INSTINT Hirst e Cristo (1998)	p.47
2.3	Papel da entoação na fala	p.49
3	Alvos Tonais	p.55
3.1	Alvos tonais	p.55
3.1.1	Questões de Linguagem e questões de informação	p.55
3.1.2	O modelo de t'Hart, Collier e Cohen (1990)	p.58
3.2	Explosões, Ferreira-Netto (2008)	p.66
3.2.1	Modelamento	p.69
3.2.2	Tom médio, limiares de diferenciação tonal e finalização	p.71
3.3	Discussão: alvos topológicos, alvos pontuais	p.75
4	Fala	p.79
4.1	Fala, comunicação	p.80
4.1.1	Fala	p.80
4.1.2	Comunicação	p.83
4.2	Sistema, fala	p.90
4.3	Contínuo, descontínuo	p.96
4.3.1	Sistemas acoplados	p.103
4.4	Acoplamento	p.107
5	Teste 1	p.112
5.1	O desenho experimental	p.112

5.2	Análise experimental	p.115
5.2.1	Caracterização do estímulo	p.115
5.2.2	Condições experimentais	p.116
5.2.3	Coleta de dados	p.118
5.3	Primeira condição amostral: alvos topológicos	p.119
5.3.1	Caracterização das amostras	p.119
5.3.2	Análise distribucional	p.121
5.3.3	Análise contrastiva	p.131
5.4	Segunda condição: alvos pontuais	p.138
5.4.1	Análise descritiva	p.141
5.4.2	Análise do alvo pontual	p.149
5.4.3	Discussão dos dados	p.157
6	Emoções	p.159
6.1	Introdução	p.159
6.2	Razão e emoção, mente e corpo	p.160
6.2.1	A contribuição grega, estoica e tomista	p.160
6.2.2	Estoicos	p.165
6.2.3	A doutrina escolástica	p.168
6.3	Fundando as bases modernas: Descartes, James e Darwin	p.173
6.3.1	Paixões cartesianas	p.173
6.3.2	Hume: paixões como ação empírica	p.177
6.3.3	Darwin e James: a expressão das emoções	p.178

6.3.4	Cannon-Bard: questões de neurologia	p. 181
6.4	Desdobramentos contemporâneos	p. 184
6.5	A visão evolucionista	p. 186
6.5.1	Ganho evolutivo	p. 186
6.5.2	As Emoções Básicas	p. 190
6.6	A questão cognitiva e social	p. 196
6.6.1	A hipótese cognitiva	p. 200
6.6.2	A teoria sócio-construtiva	p. 207
6.7	Discussão	p. 209
7	Fala emotiva	p. 211
7.1	O modelo de Damásio (2012): integrando as perspectivas	p. 211
7.2	Linguagem e emoção	p. 215
7.3	A Expressão vocal da emoção	p. 217
7.4	Teste 2: um estudo de caso das emoções	p. 220
7.4.1	Natureza dos dados	p. 222
7.5	Análise dos dados	p. 223
7.5.1	Valores para o LDT Superior: o alvo pontual	p. 230
7.6	Discussão dos dados	p. 236
8	Conclusões	p. 237
	Anexos	p. 244
	Anexo I	p. 244

Anexo II	p.245
Anexo III	p.246
Referências Bibliográficas	p.247

Lista de Figuras

- 1.1 Comparação das unidades entoacionais na fala com raiva e neutra. Retirado de Martins e Vassoler (2013). p. 18
- 2.1 Modelo massa-mola esquematizado. k é o coeficiente de elasticidade da mola, isto é, o quanto ela é capaz de deformar e voltar ao seu estado original. p. 27
- 2.2 Representação esquemática da glote utilizando o modelo de duas massas. p. 28
- 2.3 Desenho esquemático da laringe e seus componentes. Retirado de The Web site of the National Cancer Institute p. 29
- 2.4 Desenho anatômico da *rima glottidis* vista de cima, a partir da entrada da laringe. Retirado de Gray (1918) p. 31
- 2.5 Vista frontal da tráquea, destaque para os músculos *cricotireoideo* e *tireoaritenoideo*, responsáveis pelo grau de tensão das pregas vocais. . . p. 32
- 2.6 Vista superior das etapas de abertura da glote, os triângulos negros representam as cartilagens aritenoides. p. 34
- 2.7 Vista superior das etapas de abertura da glote, os triângulos negros representam as cartilagens aritenoides. p. 35
- 2.8 Movimento de fechamento das pregas vocais, a parte inferior tende a se deslocar antes da parte superior, devido a sua constituição e a força lateral resultante. Retirado de Titze (1988) p. 36

2.9	Movimento de fechamento das pregas vocais. Retirado de Titze (1988) . . .	p. 36
2.10	Esquema de organização sintagmática no modelo de Pierrehumbert (1980)	p. 45
3.1	Representação esquemática de um alvo tonal exemplificado. Retirado de Xu e Wang (2001)	p. 63
3.2	Valores de F0 mínimo e máximo para vozes masculinas. Retirado de Baken e Orlikoff (2000, p.188)	p. 65
3.3	Valores de F0 mínimo e máximo para vozes femininas. Retirado de Baken e Orlikoff (2000, p.188)	p. 66
3.4	As setas diagonais representam a declinação desencadeada em cada mo- mento pelo desejo de supressão do esforço, ao passo que as setas verticais representariam a sustentação desencadeada pela tentativa de retomada da tensão inicial, e o produto resultante desse esforço, o ritmo tonal. Retirado de Ferreira-Netto (2008)	p. 68
3.5	Limiares de Diferenciação tonal e os pontos de proeminência. Retirado de Consoni (2011)	p. 70
3.6	Exemplo de análise se usando o tom médio para a oração: “O Dome- nico entrar aqui...como assim?”.	p. 74
3.7	Esquema que representa a relação entre os alvos topológicos e alvos pontuais	p. 76
3.8	Esquema que representa a relação as valências e a transmissão de significado.	p. 77
4.1	Esquema de comunicação de Fonàgy (1991). O código linguístico sofre uma segunda codificação, o paralinguístico que também seria natural.	p. 85

4.2	Representação esquemática do trato vocal e das forças que operam na produção da frequência fundamental	p. 104
4.3	Corte sagital da glote demonstrando o vestíbulo laríngeo, a região acima glote.	p. 109
5.1	Curva de frequência do estímulo apresentado aos falantes em nosso teste. .	p. 115
5.2	Demonstração da relação entre o valor de α e o tamanho n da amostra. A partir de 35 <i>tokens</i> o resultado do teste se mostra satisfatório estatisticamente	p. 117
5.3	Demonstração da relação entre o tamanho do efeito e o valor n da amostra. A partir de 40 <i>tokens</i> o resultado do teste se mostra satisfatório estatisticamente	p. 118
5.4	Variação de frequência fundamental para a frase-estímulo que utilizamos em nosso teste dos alvos topológicos.	p. 121
5.5	Histograma da média de frequência fundamental para todas as amostras da frase (1). Os valores estão em MIDI.	p. 124
5.6	Histograma da média de frequência fundamental para todas as amostras da frase (1). Os valores estão em Hertz	p. 125
5.7	<i>Box-plot</i> com as duas medidas, Hz e MIDI.	p. 128
5.8	Gráfico quantil-quantil demonstrando a posição do valor 46,9% indicado pelo asterisco.	p. 129
5.9	Histograma dos coeficientes de variação da frase (1). Os valores estão em Hz.	p. 130
5.10	Histograma da diferença entre os coeficientes de variação das amostras da frase (1) e o do estímulo. Os valores estão em MIDI.	p. 134
5.11	Histograma da diferença entre os coeficientes de variação das amostras da frase (1) e o do estímulo. Os valores estão em MIDI.	p. 134
5.12	Gráfico de barra de erros do índice de relação entre estímulo e amostras da frase (1), com as duas condições de análise Hertz e MIDI.	p. 136

5.13	Gráfico de dispersão comparando o índice de relação e coeficiente de variação de cada amostra da frase (1), em Hz	p. 137
5.14	Gráfico de dispersão comparando o índice de relação e coeficiente de variação de cada amostra da frase (1), em Midi	p. 137
5.15	Variação de frequência fundamental para a frase-estímulo que utilizamos em nosso teste dos alvos pontuais.	p. 141
5.16	Histograma de médias da frequência fundamental da frase (2), os valores estão em MIDI.	p. 143
5.17	Histograma de médias da frequência fundamental da frase (2), os valores estão em Hertz.	p. 144
5.18	Barra de erros que determina os valores aceitáveis do coeficiente de variação.	p. 147
5.19	Histograma dos valores médios de F/E.	p. 154
5.20	<i>Box-Plot</i> que demonstra os valores de limite	p. 156
6.1	Comparativo entre os modelos James-Lange e Cannon-Bard. As setas pretas demonstram que, de acordo com James-Lang, a percepção do estímulo leva a uma reação, a qual se converte na experiência emocional. Por outro lado, as setas azuis apresentam a visão de Cannon-Bard, que entendem que a experiência emocional é anterior a manifestação em si.	p. 182
7.1	Histograma da frequência fundamental para voz neutra, medida em MIDI. .	p. 225
7.2	Histograma da frequência fundamental para tristeza, medida em MIDI. . . .	p. 226
7.3	Histograma da frequência fundamental para a raiva, medida em MIDI. . . .	p. 226
7.4	<i>Box-Plot</i> para frequência fundamental de todas as expressões vocais, os valores estão em MIDI.	p. 228
7.5	Barra de Erros para frequência fundamental de todas as expressões vocais, os valores estão em MIDI.	p. 229

7.6	Histograma dos valores de Limiar de Diferenciação Tonal superior para voz neutra.	p. 232
7.7	Histograma dos valores de Limiar de Diferenciação Tonal superior para raiva..p.	232
7.8	Histograma dos valores de Limiar de Diferenciação Tonal superior para raiva..p.	233
7.9	<i>Box-Plot</i> do Limiar de Diferenciação Tonal superior de todas as expressões vocais.	p. 234
8.1	Exemplo de análise se usando o tom médio para a oração: “O Dornico entrar aqui...como assim?”.	p. 244

Lista de Tabelas

- 5.1 Valores de média geral e desvio padrão da Frequência Fundamental das UPEs de cada uma das amostras da frase (1). Os valores estão em MIDI e Hertz p. 122
- 5.2 Valores de assimetria e curtose para as médias de Frequência Fundamental de cada uma das amostras da frase (1). p. 123
- 5.3 Valores do coeficiente de covariação de frequência fundamental de cada uma das amostras da frase (1). p. 127
- 5.4 Tabela com os resultados do teste Kolmogorov-Smirnov para a distribuição dos coeficientes de variação da frase (1). p. 131
- 5.5 Diferença entre o coeficiente de variação de cada uma das amostras e o estímulo da frase (1). p. 132
- 5.6 Cálculo do índice de relação intra-repetições da frase (1). p. 133
- 5.7 Teste-t do índice de relação com a média da frequência fundamental da frase (1). p. 135
- 5.8 Resultado do teste Kolmogorov-Smirnov para a distribuição significativa dos dados. p. 135
- 5.9 Valores de média geral e desvio padrão da frequência fundamental das UPEs de cada uma das amostras da frase (2). Os valores estão em MIDI e Hertz p. 142

5.10	Estatísticas descritivas da frequência fundamental das amostras e do estímulo da segunda condição experimental. $n=25$	p. 143
5.11	Análise descritiva dos coeficientes de variação das amostras e estímulo da frase (2). Os valores podem ser vistos no Anexo II	p. 146
5.12	Análise descritiva dos índices de relação entre amostras e estímulo da frase (2). Os valores podem ser vistos no Anexo III	p. 146
5.13	Valores do índice de relação para as porções inicial e final da frase (2).	p. 148
5.14	Média da frequência fundamental para as duas sentenças. Os valores estão em MIDI	p. 150
5.15	Estatísticas descritivas da frequência fundamental das porções inicial e final da frase (2). $n=25$	p. 150
5.16	Média dos valores de frequência fundamental de Foco/ Ênfase para a porção final da frase (2).	p. 152
5.17	Estatísticas descritivas da frequência fundamental das porções inicial e final da frase (2). $n=23$, o caso 7 não pode ser calculado	p. 153
6.1	Resumo dos conceitos envolvendo emoção compilados no estudo de Kleinginna e Kleinginna (1981).	p. 185
6.2	Categorias de definições descritas por Kleinginna e Kleinginna (1981) comparadas ao sistema de Cornelius (2000).	p. 185
6.3	Grau de concordância entre cientistas sobre quais emoções são universais	p. 194
6.4	Etapas que compõem a manifestação das emoções dentro de uma análise cognitiva	p. 203
6.5	Relação entre subsistemas orgânicos e as funções e componentes da emoção, segundo Scherer (2005)	p. 205
7.1	Características vocais da raiva e da tristeza. Adaptado de Scherer (1986)	p. 222

7.2	Análise descritiva	p.224
7.3	Teste-t para voz neutra, tristeza e raiva	p.229
7.4	Valores dos Limiares de Diferenciação Tonal Superiores em Midi para Voz Neutra, Raiva e Tristeza	p.230
7.5	Teste-t para voz neutra, tristeza e raiva	p.234
7.6	Valores relativos ao Foco Linguístico, em MIDI, para cada uma das emoções	p.235

Lista de Abreviaturas

SNA - Sistema Nervoso Autônomo

SE - Super Ego

FAR - Fonética Articulatória

P_{atm} - Pressão atmosférica

P_s - Pressão sub-glotal

SPE - Sound Patterns of English

TM - Tom Médio

S - Sustentação

F - Finalização

LDTs - Limiares de Diferenciação Tonal

IF0 - Intrinsic F0

CofVar - Coeficiente de Variação

1 Introdução

1.1 Apresentação

Nos últimos 15 anos é notório o grande número de publicações que se utilizaram de dados acústicos da entoação para sustentar hipóteses em torno de fenômenos linguísticos ou de comunicação. Na base de artigos da *Acoustical Society of America* (A.S.A.) nota-se uma variedade de temas e análises abordados, com trabalhos que abordam desde a identificação de padrões até a análise de dados respiratórios combinados a dados acústicos.

É importante observar que foi a obra seminal de Gunnar Fant *Acoustic Theory of Speech Production* (FANT, 1966) que permitiu a abordagem de fenômenos acústicos da fala e, por consequências, da língua serem descritos em termos de modelos matemáticos lineares, nos quais os processos de produção dos sons são explicados em termos de fenômenos encadeados, como sistemas consecutivos, em que um processo precisa ocorrer para que o próximo seja ativado.

Para os autores desta área, estes processos poderiam ser definidos e explicados utilizando a quantificação de dados, baseando-se no princípio de aproximação. O método é simples, após uma extensa análise quantitativa é possível sugerir a existência de um valor padrão, aproximado e comum a diversas medições. Este valor padrão seria o fim de um conjunto de equações, que formariam uma explicação dos mecanismos de produção da fala. Uma das obras mais importantes desta “Fonética Quantitativa” é o trabalho de Ken Stevens *Acoustic Phonetics* (STEVENS, 2000), no qual aparecem os

delineamentos mais teóricos deste modelo largamente empregado ao longo do final da década de 80 e 90.

Em paralelo a esses trabalhos, a Fonologia desenvolveu propostas que não levavam em conta o aspecto quantitativo do fenômeno linguístico, isto é, a interface entre produção da fala e língua foi tratada como um acessório à teoria. Nota-se que o aspecto acústico dos sons fora até mesmo excluído do projeto original de trabalho, como se pode ver em Chomsky e Halle (1968), em que a tentativa de busca de padrões sonoros está diretamente associada a questões de lógica formal e de pressupostos associados à representação mental. Nesta linha destacam-se, por exemplo, as teses de Pierrehumbert (1980) e os trabalhos de Goldsmith (1976, 1990b), cujo princípio organizador está associado, mas não interligado às questões acústicas.

1.2 Nossos objetivos

Com relação a essa tese, sua hipótese surge justamente de uma discussão sobre a relação entre os aspectos fonológicos e as questões de quantificação da fonologia, especificamente no domínio da fala emotiva. Em Martins e Vassoler (2013) analisamos falas atuadas que emulavam a raiva e um estado neutro, entoadas por 3 atrizes em 5 repetições para uma sentença (VASSOLER, 2011; MARTINS; VASSOLER, 2013). Nossa análise revelou alguns aspectos que, naquele momento, nos pareceram bastante pertinentes a uma hipótese teórica mais ampla, a qual pretendemos desenvolver aqui.

Os resultados daquela pesquisa demonstravam que haviam dois elementos concorrentes nas falas atuadas por nós estudadas. O primeiro aspecto seria um padrão entonacional subjacente aos estados analisados, o que chamamos de *forma fonológica*, embora o nome não fosse adequado, pois sugeria que algo, como a fala emotiva, poderia ser não fonológica. O segundo parâmetro seria a variação de F_0 em si, a qual compreendemos como a implementação executada pelo falante a partir desta forma fonológica subjacente.

A forma fonológica, por nós citada, era um produto derivado da análise estatística de unidades menores, chamadas de UBI por Ferreira-Netto et al. (2012) e organizadas na forma de uma série temporal. Naquela análise, levamos em consideração a média dos valores de F_0 , convertidos em semitons, para cada uma dessas unidades, que serão melhor definidas no capítulo 2. Em análises posteriores do mesmo *set* de dados detectamos que esses padrões de UBI mantinham, entre estados emocionais, um alto grau de correlação ($\rho > 0,8$), embora cada um deles estivesse em determinada altura no espectro. Este resultado pode ser visto na figura 1.1 logo abaixo.

Figura 1.1: Comparação das unidades entoacionais na fala com raiva e neutra. Retirado de Martins e Vassoler (2013).

A despeito dessa forte correlação, a análise das curvas de F_0 em si não era capaz de detectar esse tipo de relação, senão apenas a qualidade das variações, ou seja, se F_0 se afastava, se aproximava ou se tomava um determinado comportamento diante da média dos valores de F_0 . O coeficiente de variação, por exemplo, demonstrava que a raiva variava cerca de 3,78% e a voz neutra, cerca de 5,78%. Estes dados demonstravam que a raiva utilizava um *pitch range* mais amplo, o que não permitia um excesso de variação, isto é, o falante teria pouco espaço para variação, ao passo que na voz neutra esse espaço seria mais amplo. No entanto, esse tipo de dado não explicava o porquê da alta correlação entre as curvas que avaliam as unidade entoacionais, ao invés da

variação global.

Análises que avaliam uma tendência universal de um determinado evento, como aquelas executadas por Banziger e Scherer (2005), revelam um comportamento estatístico que demonstra um padrão de variação, mas nunca um padrão de estabilidade, isto é, podemos explicar a variação entre estados a partir de um conjunto de hipóteses associadas às suas produções, mas não poderíamos explicar a correlação entre eles, a não ser por uma hipótese linguística mais ampla. Neste sentido, os resultados por nós encontrados derivaram uma série de questões sobre os aspectos linguísticos associados à emoção.

O primeiro deles remete à suposição de Halle (2002) sobre a relação entre formas fonéticas e a memória. Para o autor, as unidades fônicas somente poderiam estar armazenadas na memória do falante em termos de padrões e nunca como realizações. A justificativa é simples e bastante adequada: não nos lembrariamos de uma determinada elocução pela forma como ela é pronunciada, mas, sim, pelo padrão que emergiria dela quando um falante a pronuncia. Desta forma, não adquirimos a qualidade de uma determinada vogal, mas, sim, os parâmetros que nos permitiriam executá-la e, é claro, percebê-la como um som da fala. Para Liberman e Mattingly (1985) e Fowler (1986) é bastante evidente que esses parâmetros seriam os elementos que produzem a fala em si, ou seja, nossa habilidade de perceber os sons da fala, estaria associada à capacidade de detectar os gestos que os produzem, embora para Ohala (1996) esse argumento não seja válido, uma vez que poderíamos perceber sons semelhantes aos da fala que são produzidos de forma artificial, por exemplo.

Em nosso entender, o pressuposto de Halle (2002) poderia valer para a entoação da fala emotiva (WILCE, 2009). O falante não teria a tendência a produzir uma forma entoacional natural e estabelecer as variações de frequência fundamental a partir dela, mas, antes, estabelecer pontos de referência a partir dos quais ele poderia iniciar a implementação de F_0 . Desta forma, a fala emotiva não seria uma modalização da fala

natural ou espontânea, mas, sim, um parâmetro linguístico estabelecido pela própria língua. Na outra frente, o ouvinte seria capaz de mapear essa variação e decodificar a atitude do falante.

Neste sentido, entendemos que no nível cognitivo existe uma forma fonética não-padronizada, mas referenciada, operando como um parâmetro, que permite codificar e decodificar determinados eventos linguísticos em sua forma física. Nossa hipótese é que essa forma fonética na entoação opere como zonas de convergência, isto é, pontos para os quais os tons convergem durante a produção da fala e que transmitem informações entoacionais. Xu e Wang (2001) denominam essas áreas de convergência de *alvos tonais* e nossa sugestão é que elas funcionam como mecanismos que orientam a implementação fonética da entoação, fazendo uma ponte entre o aspecto cognitivo e o físico.

A continuidade de nossa hipótese é a de que a fala emotiva transformaria esta forma fonética subjacente, transportando os tons da fala para novas zonas de convergência, ou seja, a atitude do falante alteraria os alvos tonais que podem ser implementados de algumas maneiras, graças a uma série de variáveis, como o *pitch range* de um falante, a duração de uma elocução e a intenção comunicativa. Entretanto, estes alvos, em tese, manteriam uma determinada regularidade tanto na percepção, graças ao que Kuhl (1991) chama de *efeito magnético de percepção*, quanto na produção, como aquela encontrada por nós em Martins e Vassoler (2013) e representadas na figura 1.1. É importante destacar que não estamos pensando unicamente na variação de F_0 , mas antes em um elemento protípico, a ser investigado, capaz de acumular a maior quantidade de características destas curvas de F_0 .

Teríamos, então, quatro elementos que compõem a produção da entoação. O primeiro deles seriam os (i) *alvos tonais*, unidades fonético-fonológicas prototípicas que servem de referência cognitiva para o falante implementar a entoação. O segundo elemento seria (ii) a *implementação* em si, o aspecto físico que promoveria as variações

de acordo com as orientações e a “programação” dos alvos. Desta forma, associado aos alvos viriam os movimentos tonais e as unidades portadoras. Os (iii) *movimentos tonais* (T'HART; COLLIER; COHEN, 1990), como o nome sugere, são as mudanças no curso de F_0 , como já estabelecido por Maeda (1976), Pierrehumbert (1980) e t'Hart, Collier e Cohen (1990) e representariam o comportamento de F_0 no curso de uma elocução. A quarta parte seriam (iv) as *unidades portadoras*, a contraparte linguística, que nada mais seria do que as porções de fala que contém F_0 .

Para que pudéssemos analisar a entoação a partir dos alvos, definimos que eles podem ter dois comportamentos. O primeiro, que chamamos de *topológico*, está associado à distribuição das frequências em um espaço determinado, o *espaço entoacional*. Neste caso, nossa hipótese é a de que os falantes saberiam, intuitivamente, em qual região de frequências implementar a entoação, de modo tal que essa implementação em si já teria algum significado. No Teste II reportado nessa tese, notamos que o espaço entoacional da raiva é menor do que o da fala neutra e essa diferença causa uma alteração na distribuição das frequências em um e no outro caso. O segundo tipo denominamos *alvo pontual*, pois não está associado à distribuição em um espaço definido, mas, sim, a execução de uma tarefa que tem um ponto focal como objetivo final, tal como na produção da proeminência tonal no Português Brasileiro. A análise reportada no capítulo 5 sugere que essa diferenciação é possível e, provavelmente, produtiva na análise linguística.

Por fim, a nosso ver, um problema intrínseco à entoação é justamente a natureza e o *status* da frequência fundamental como parte da fala ou da língua. Sendo assim, para uma análise mais delimitada da entoação entendemos que é preciso, antes de mais nada, compreender que a frequência fundamental pode ser utilizada como um recurso “articulatório”, ora atendendo às necessidades expressivas do falante, ora atendendo às necessidades linguísticas. Nesta visão, temos de assumir que, de um ponto de vista fonético, a produção dos segmentos pode ser dissociada da produção da entoação, embora elas sejam mutuamente dependentes. Nossa hipótese, nesse caso, é a de que a

fala possui dois sistemas: um contínuo e outro descontínuo. A entoação seria o sistema contínuo, ao passo que os segmentos seriam o sistema descontínuo. Tratamos desse assunto no capítulo 4.

1.3 Fala emotiva

Pode se notar que derivamos uma hipótese linguística a partir de uma análise prática da fala emotiva. Desta forma, nossa intenção é utilizar a análise das emoções como objeto de estudo de caso, uma vez que nossa hipótese principal surgiu justamente daí. Em geral, assume-se que a emoção pode ser compreendida como a alteração de estados fisiológicos e mentais, a partir de mecanismos cognitivos de avaliação dos estímulos ambientes, de modo que essas alterações fisiológicas, implicariam em mudanças na forma final da fala (SCHERER, 2003). Portanto, as emoções seriam mapeadas pelos ouvintes a partir das informações linguísticas que sofrem algum tipo de alteração, sendo que haveria um padrão para esta variação (SCHERER, 2003, 2005).

A partir da discussão feita sobre o papel da quantificação na determinação de formas prototípicas, é natural compreendermos a fala emotiva como desvios de um modelo de produção relativamente estável, como aqueles definidos por Fant (1966) e Stevens (2000). Por este motivo, usualmente, métodos de análise como os de Williams e Stevens (1972) e Banziger e Scherer (2005) utilizam recursos de comparação de alturas ou análise da dispersão da frequência fundamental para explicar seus dados. Nossa proposta é que a análise da distribuição dos alvos em espaços entoacionais também pode revelar informações importantes sobre como os falantes utilizam as curvas de F_0 para transmitir informações sobre os seus estados emocionais.

Dentro dessa nossa proposta há uma segunda hipótese, baseada em Damásio (2012). Para o autor, as emoções não são necessariamente avaliações cognitivas que derivam em alterações corporais, mas, possivelmente, alterações bio-funcionais, a partir de neuropeptídeos e hormônios, que geram mudanças globais, inclusive na cognição.

Nesse sentido, as reações emocionais não são apenas um reflexo automático disparado a partir da avaliação cognitiva de um estímulo e, sim, um estado de alerta gerado a partir desta resposta reflexa, que seria nada mais, nada menos do que uma estimulação bio-química.

A nosso ver, a manifestação da fala emotiva parece se enquadrar neste aspecto, pois passado o efeito do estímulo que levou ao estado emocional, ainda é possível notar na fala marcas que transmitem alguma informação expressiva do sujeito. Ainda nesta perspectiva, para Damásio (2012) nunca estamos em um “estado neutro”, pois somos bombardeados por estímulos ambientais constantemente, assim ao experienciarmos uma emoção, a partir do mecanismo reflexo acima citado, passaríamos a modular nossa experiência e interações a partir dos estados mentais despertados por ela.

O modelo de Damásio (2012) parece-nos o mais adequado quando tratamos da fala, uma vez que não simplifica sua manifestação ao fisiologismo, bem como torna possível explicar a relação entre emoção, linguagem e cognição. Entendemos, dentro de nossa proposta, que a emoção altera o aspecto cognitivo dos alvos tonais, orientando os falantes a implementá-los em certos espaços entoacionais que podem se manter ao longo de toda uma elocução e não apenas no momento de experimentação da emoção. No capítulo 7 explicamos melhor esse modelo e sua relação com a linguagem.

Para o estudo de caso reportado no 7 utilizaremos a raiva, a tristeza e a hipotética fala neutra em falas atuadas como material de análise. Escolhemos essas variáveis, pois entendemos que as diferenças acústicas delas são bastante proeminentes, o que pode nos ajudar a melhor explicar nossa hipótese. Em uma segunda etapa pretendemos analisar aquelas emoções que não demonstram diferenças claras, como é o caso da alegria e do medo.

1.4 Descrição dos capítulos

Ao final desse trabalho notamos que ele se divide em duas grandes partes. Na primeira delas explicamos o alvo tonal como uma unidade fonético-fonológica, aplicável à análise linguística da fala. Na segunda parte, nos dedicamos a uma análise mais aplicada, buscando relacionar a hipótese teórica aos problemas práticos encontrados na fala emotiva.

No Capítulo 2 apresentamos os modelos teóricos desenvolvidos ao longo dos anos para explicar o papel da frequência fundamental na língua e na fala. Além disso, apresentamos um esquema de como a frequência fundamental é produzida. No capítulo 3 apresentamos a nossa hipótese dos *alvos tonais* e como eles se relacionam com as teorias fonológicas.

O leitor notará que o capítulo 4 parece ligeiramente desconectado desta primeira parte de nossa tese. Nossa intenção é justificar porque a implementação de F_0 pode ser compreendida como um processo à parte, dividindo a fala em dois sistemas: um contínuo e um descontínuo. O primeiro é o da entoação, ao passo que o segundo é o dos segmentos. Nossa proposta é que a entoação não funcionaria como mero recurso paralinguístico, mas desempenharia um papel linguístico importante para a transmissão de conteúdos expressivos e comunicativos, como os emocionais.

O capítulo 5 apresenta os resultados dos testes que fizemos para avaliar a possibilidade dos alvos tonais operarem em dois níveis, como fenômenos topológicos e pontuais. Os alvos topológicos serão avaliados pela capacidade do falante notar a distribuição dos tons dentro do espaço tonal e implementar a variação de F_0 nesta região virtual, enquanto os pontuais serão medidos pela capacidade do falante executar uma tarefa. Esse capítulo encerra a primeira parte da tese, mais voltada para questões linguísticas do que para os problemas associados às emoções.

Para iniciar a segunda parte da tese, legamos ao capítulo 6 a tarefa de explicar o

estado de arte das teorias e modelos que buscam compreender o papel das emoções seja na evolução do homem, seja em sua relação com o meio ou a sociedade. O leitor notará que esse capítulo parece não pertencer ao conjunto do trabalho. Explicamos. Como estamos lidando na área da interdisciplinariedade, nos pareceu adequado fazer uma discussão teórica mais aprofundada sobre o tema das emoções, a fim de evitar que não tratemos algumas questões teóricas da segunda área com uma superficialidade exagerada, comum quando falamos de áreas correlacionadas. Por fim, no capítulo ?? apresentamos o modelo teórico das emoções que adotamos nessa tese, bem reportamos o resultado de nosso estudo de caso sobre as emoções.

2 Entoação

Neste capítulo vamos tratar dos aspectos associados à produção de F_0 , os mecanismos fisiológicos ligados à glote e o produto resultante. Em nosso entender, compreender a produção da frequência fundamental é essencial para que possamos compreender o seu papel na linguagem. Por este motivo, também analisaremos o tratamento dado pelas teorias fonológicas sobre o tema e como elas se correlacionam, além das possíveis funções da entoação na linguagem.

2.1 A produção da entoação

Seja na literatura ou nos cursos introdutórios, é comum vermos a separação dos conceitos de frequência fundamental (F_0) e *pitch*. No geral, trata-se do primeiro como a manifestação física e mensurável de um som, ao passo que o segundo seria a unidade linguística e perceptiva da entoação (T'HART; COLLIER; COHEN, 1990). Não faremos, necessariamente, essa distinção, pois entendemos que a entoação é o resultado das variações de frequência fundamental, modalizadas por funções linguísticas específicas.

A noção de "variação da frequência fundamental" está apoiada em dois conceitos contíguos. O primeiro deles é o de que um corpo é excitado por uma fonte, o que gera uma transferência de energia e movimento, o segundo conceito é o de que essa excitação se repete de forma periódica, o que gera a frequência. No caso da entoação, o corpo excitado são as pregas vocais; a fonte de energia são os pulmões ou o fluxo de ar que irrompe deles para a atmosfera e a repetição desses movimentos é determinada

pelas necessidades expressivas ou comunicativas de um falante.

Desta forma, estes dois conceitos sugerem a existência de dois domínios interligados. O primeiro deles é o fisiológico, que busca uma explicação de como os órgãos dos sistemas respiratório e digestivo são utilizados para a produção da fala. O segundo domínio é o físico, que lida com a relação entre estes órgãos, a dinâmica dos fluxos aéreos e o som produzido. Entendemos que a relação entre estes domínios é direta, no entanto, para efeito de análise, faremos uma separação. Neste capítulo analisaremos como se dá a produção da frequência fundamental por estes órgãos, ao passo que no capítulo 4 vamos analisar a questão física e suas implicações na fala com mais profundidade.

Para explicar a produção da frequência fundamental pelo domínio da fisiologia, utilizaremos o modelo de apresentado por Stevens (2000), em que o autor se vale do princípio de "oscilador massa-mola", um recurso matemático utilizado para definir o comportamento de sistemas oscilatórios (MARION, 2008) e para demonstrar a produção da frequência fundamental. A figura 2.1 apresenta um esquema deste tipo sistema.

Figura 2.1: Modelo massa-mola esquematizado. k é o coeficiente de elasticidade da mola, isto é, o quanto ela é capaz de deformar e voltar ao seu estado original.

Em modelamentos que utilizam este princípio, os sistemas são analisadas a partir de um ponto de equilíbrio (quando a força da mola sobre a massa é nula $F = 0$) e que se deslocam para um estado S_n (quando a força da mola sobre a massa é diferente de zero $F \neq 0$). A força aplicada leva a uma deformação da mola (coeficiente k) e com isso a massa se distancia um valor x do seu ponto do repouso. O decréscimo da energia aplicada, graças a resistência r da própria mola, reduz paulatinamente a sua deformidade e, por consequência, a distância x , de modo que cada ciclo de impulso

será menor do que o anterior, até que a massa retorne ao estado de equilíbrio, quando $F = 0$ e $x = 0$

No caso das pregas vocais, os tecidos que a compõem têm diferenças morfológicas e anatômicas que alteram a sua forma, principalmente, a rigidez e espessura ao longo de sua extensão. Embora não sejamos sensíveis a essas diferenças, elas afetam a forma como as pregas vocais vibram. Por isso, o modelo massa-mola na fala é considerado um sistema de duas massas, uma vez que essa diferença estrutural impacta na produção da frequência fundamental. A figura 2.2 representa este modelo:

Figura 2.2: Representação esquemática da glote utilizando o modelo de duas massas.

Como se pode ver na figura 2.2 a própria extensão da prega vocal imprime uma resistência a sua vibração e, por consequência, à variação da frequência fundamental. Em nosso entender, para que possamos explicar melhor o fenômeno da produção da fala a partir desse modelo é importante que façamos uma descrição anatômica e morfológica dos órgãos da respiração envolvidos.

2.1.1 A Laringe

A primeira porção anatômica a ser considerada é a laringe, uma vez que é nela que a produção da frequência fundamental ocorre. Anatomicamente, a laringe estabelece a ligação entre a faringe/raiz da língua até a traqueia, localiza-se na parte anterior do pescoço e se estende da quarta à sétima vértebra. É composta acima pela epiglote;

no centro pelo ventrículo, onde se encontram as falsas pregas vocais; as pregas vocais/glote e abaixo pela tráquea.

A figura 2.3 demonstra a posição das pregas vocais e do ventrículo da laringe, em relação à epiglote e a tráquea:

Figura 2.3: Desenho esquemático da laringe e seus componentes. Retirado de The Web site of the National Cancer Institute

Stevens (2000) sugere que a laringe é apenas um sistema suporte para as pregas vocais; Baer (1975) citando Negus (1948) assume que a principal função da laringe, de um posto de vista evolutivo, é o de fechamento do sistema respiratório. Entende-se que esse processo sirva para que corpos estranhos não atinjam o sistema respiratório, o que poderia causar sufocamento. Além dessa proteção, a laringe também seria responsável por outros processos, como o de deglutição e de regulação da pressão torácica.

Com relação à fala, evidências anatômicas sugerem que a laringe possui uma extensão maior do que a necessária para a produção da fala, isto é, a laringe não tem um papel ativo nesse processo. Deste modo, é improvável que ela tenha se adaptado

unicamente para contribuir com a produção da fala (BAER, 1975), Negus (apud BAER, 1975) sugere que essa evidência é significativa, quando comparadas a laringe humana com a de outros animais. Para o autor, essa diferença não é de ordem estrutural e está localizada na habilidade adquirida pelo homem de controlar a laringe, bem como a glote. Esta hipótese confronta a teoria que toma a laringe como "órgão da fala", abordagem adotada por anatomistas como Gray (1918) e outros, colocando a glote como principal responsável pela produção da fala, com um grau de independência funcional da laringe.

2.1.2 A glote

Como se pode ver na seção anterior, a laringe funciona basicamente como um suporte ou estrutura de sustentação para os órgãos da fala em si. A segunda parte no processo de fonação corresponde à glote que é basicamente a abertura (*rima glottidis*) entre as pregas vocais (*plioe vocales*) e as pregas vocais em si. Anatomicamente, as pregas vocais são compostas por duas camadas de ligamentos, chamados ligamentos vocais, os quais são ligados na parte anterior à cartilagem tireoide que se liga a raiz da epiglote e na parte posterior às cartilagens aritenoides, sendo estas estruturas as responsáveis pelo movimento da glote e que estão ligadas à cartilagem cricoide. Na lateral, em paralelo aos ligamentos, encontram-se os músculos vocais *vocalis muscle*.

A figura 2.4 representa com fidelidade a estrutura anatômica das pregas vocais acima descrita, nela pode se ver a estrutura muscular e cartilaginosa da glote, começando pelos ligamentos vocais (com a anotação *chorda vocalis*), os músculos tireoari-tenoideos (em vermelho à esquerda), a ligação abaixo em ângulo com a cartilagem tireoide e a ligação acima com a cartilagem aritenóide, responsáveis pelos movimentos da glote. A figura 2.5 representa as mesmas estruturas, mas em uma visão frontal.

Figura 2.4: Desenho anatômico da *rima glottidis* vista de cima, a partir da entrada da laringe. Retirado de Gray (1918)

De acordo com Baer (1975), os músculos que compõem a laringe podem ser analisados de duas formas: (1) os que abrem e fecham a glote e (2) aqueles que controlam o grau de tensão das pregas vocais. No processo de fonação, os músculos cricoaritenoides laterais rotacionam as cartilagens aritenoides para dentro, de modo que elas se aproximem e fechem a *rima glottidis*. Esse processo fricciona a passagem de ar e gera a turbulência necessária para a geração da frequência fundamental. Para a respiração, os músculos cricoaritenoides posteriores rotacionam as cartilagens aritenoides para fora em torno de um eixo vertical, passando-as através das juntas cricoaritenoides, na parte lateral da glote em vermelho na figura 2.4.

Já os músculos que tomam parte no grau de tensão das pregas vocais são, respectivamente, o *cricotireoideo* e o *tireoaritenoides*. O primeiro deles está fora da glote, localizado na parte anterior do pescoço, como demonstra a figura 2.5, por isso é capaz de alongar ou tensionar as pregas vocais, por meio da elevação da ângulo cartilagem

tireoide. O segundo, por sua vez, está ligado diretamente às pregas vocais e controla sua elasticidade e direção, rotacionando a cartilagem aritenoide e deixando a *rima glottidis* mais estreita ou mais larga.

Figura 2.5: Vista frontal da tráquea, destaque para os músculos *cricotireoideo* e o *tireoaritenoideo*, responsáveis pelo grau de tensão das pregas vocais.

De acordo com Stevens (2000), devido a morfologia das pregas vocais é natural que as diversas camadas vibrem em modos diferentes. Este fenômeno decorre da rigidez própria dos músculos e tecidos que compõem a glote. Para explicar como se dá essas vibrações em multicamadas, Stevens (2000) faz referência ao "Módulo de Young", uma equação matemática utilizada para o cálculo da rigidez de materiais elásticos e aplicada para cada tecido das pregas vocais. O "Módulo de Young" é representado pela fórmula (STEVENS, 2000):

$$\gamma = \frac{\Delta F / A}{\Delta x / x} \quad (2.1)$$

Onde ΔF é a força aplicada a A uma área, de modo que a direção em que a força aplicada é igual a x e a mudança causada pela força é representada por Δx . A unidade de medida é a de força por área Newtons por metro quadrado (SI) ou dynes/cm²

(CGS). Segundo Kakita, Hirano e Ohmaru (1981), cada um dos tecidos possui um módulo distinto e relacionado com o alongamento do tecido. Por exemplo, os músculos *vocalis* e a *lamina propria* precisam de um baixo valor do módulo de Young (cerca de 10^4 dynes/cm²) para que se alonguem cerca de 10% do seu tamanho original. Por outro lado, a mucosa necessitaria de cerca de 10^6 dynes/cm² para sofrer o mesmo alongamento.

Os valores aqui apresentados têm um papel relevante na determinação do *pitch range* de cada falante. As pregas vocais podem ser entendidas, nestes termos, como um mecanismo de resistência e distensão, como em um modelo massa-mola. A massa é determinada pelo módulo Young geral das pregas vocais, calculado a partir das medidas de espessura e altura, bem como deslocamento horizontal. A partir do módulo de Young, da densidade dos tecidos e da espessura das pregas, pode-se chegar ao valor das massas e, por consequência, da frequência natural para uma força contínua. Para pregas vocais femininas, com cerca de 2,5 mm tem se frequências naturais, por volta de 200 Hz, enquanto para vozes masculinas com pregas vocais de 4 mm tem se frequências naturais de cerca de 120 Hz (STEVENS, 2000).

2.1.3 A produção da frequência fundamental

Em geral, assume-se que apenas dois fatores sejam necessários para o início da fonação. O primeiro deles, é a diferença entre a pressão subglotal (P_s) e a pressão acima da glote (P_{atm}), a qual é a pressão atmosférica. O segundo fator necessário é a existência de um fluxo de ar (ϕ) entre os pulmões e o meio externo. Assume-se que a pressão subglotal necessite ser da ordem de 0,05 atm para o início da fonação (LADEFOGED; MCKINNEY, 1963; TITZE, 1989; STEVENS, 2000), enquanto a pressão acima da glote é da ordem de 1 atm, ou seja, o valor da pressão atmosférica.

A figura 2.6 é um esquema simplificado do processo de fechamento da glote e do início de vibração das pregas vocais.

Figura 2.6: Vista superior das etapas de abertura da glote, os triângulos negros representam as cartilagens aritenoides.

No painel A nota-se a glote em sua posição natural quando respiramos, a parte deste processo inicia-se o movimento de rotação das cartilagens aritenoides, causadas pelos músculos cricoaritenoides laterais, com isso as pregas vocais sofrem uma adução em direção uma a outra, como se pode ver nos painéis B e C da figura 2.6. No painel D vemos o movimento de adução completo acarretando o fechamento das pregas, como a tensão que as une é uma força lateral, a força $\vec{F}$ inferior proporcionada pela pressão do fluxo de ar vindo dos pulmões pode não ocasionar o movimento de abdução completo e, sim, apenas uma abertura parcial, como representado no painel E.

O processo entre os painéis D e E se inicia quando a pressão sub-glotal atinge um determinado limiar de pressão (TITZE, 1988, 1989; LUCERO et al., 2009). Enquanto os valores de pressão sub-glotal estiverem acima do limiar de pressão, é natural que o fechamento representado pelo painel F não ocorra. A figura 2.7 ilustra o que ocorre durante esse processo e como a vibração das pregas vocais está diretamente condicionada a sua morfologia, como apresentado por Stevens (2000).

Enquanto o fluxo de ar entre as pregas vocais se mantiver contínuo e a pressão subglotal se mantiver em um valor que supere o valor necessário para que as forças laterais proporcionadas pelos músculos cricoaritenoides se anulem e voltem ao seu estado repouso, teremos a frequência fundamental. Com a diminuição da passagem do fluxo de ar pela glote, a pressão sub-glotal torna-se próxima de zero ou negativa, de modo que a força exercida na parte inferior da prega vocal se torna maior que a força exercida na parte superior (efeito Bernouilli). A diferença de forças traz para próximo de si a parte inferior das pregas vocais, levando as cartilagens aritenoides a

movimentar-se em sentido longitudinal, ocasionando a abertura da glote.

Figura 2.7: Vista superior das etapas de abertura da glote, os triângulos negros representam as cartilagens aritenoides.

Com relação aos movimentos representados na 2.7 Titze (1988) estabelece como princípio que há uma relação entre a frequência fundamental e o fluxo de ar. Segundo o autor, há uma transferência contínua de energia entre o fluxo de ar e o tecido das pregas vocais, de modo que ocorrerão perdas em forma de fricção no contato entre estas partes. As perdas ocasionam a oscilação do tecido e, por consequência, a vibração das pregas vocais. Além disso, Titze (1988) propõe que exista um diferencial de tempo de reação entre as parte inferior e superior das pregas vocais. Devido a sua morfologia, a parte inferior seria mais suscetível a mudança da forma original com a aplicação de uma determinada quantidade de força lateral dependente da velocidade do fluxo de ar e da pressão exercida nas pregas vocais, enquanto a parte superior levaria mais tempo para se submeter a esta força.

A figura 2.8 demonstra esquematicamente o mecanismo de oscilação proposto por Titze (1988).

Figura 2.8: Movimento de fechamento das pregas vocais, a parte inferior tende a se deslocar antes da parte superior, devido a sua constituição e a força lateral resultante. Retirado de Titze (1988)

No caso da abertura da glote, na figura 2.9 nota-se a existência de uma força lateral com velocidade $\dot{\xi}$ e com direção lateral interna, sendo que é essa força que mantém as pregas vocais aproximadas. Observa-se que a parte inferior sofre alterações, enquanto a parte superior mantém-se inalterada. No instante em que a parte superior mudar sua forma, devido as trocas de energia entre as duas partes, a porção inferior terá retornado ao seu estado natural, o que irá gerar uma alternância entre estas partes. A duração entre essas alternâncias irá determinar o valor da frequência fundamental.

Na figura 2.9 vemos que a troca de energia entre o fluxo de ar e o ambiente, seguido da diminuição da pressão sub-glotal ocasionam uma mudança de sentido na velocidade $\dot{\xi}$, a qual passa a se mover em sentido contrário, levando a abertura da glote. Novamente, a parte inferior sofre uma alteração antes da superior, ver figura 2.9.

Figura 2.9: Movimento de fechamento das pregas vocais. Retirado de Titze (1988)

A figura 2.7 demonstra essa movimentação das massas da glote. Os painéis de 1 a 3 ilustram a movimentação da parte inferior e a transferência de energia para a parte

superior das pregas vocais no processo de abertura da glote. Enquanto os painéis de 8 a 10 ilustram o movimento de fechamento da glote, observa-se a aproximação da parte inferior antes da parte superior, como sugerido por Titze (1988). Os painéis de 4 a 5 demonstram o processo intermediário entre a abertura e o fechamento, ao passo que os painéis 6 e 7 demonstram a tendência de aproximação das partes inferiores, ocasionada pelo efeito Bernouille.

Em seu trabalho posterior, Titze (1989) propõe mais duas variáveis para a variação da frequência fundamental. A primeira delas seria o grau de expansão da prega vocal, em relação ao seu estado neutro; a segunda seria o comprimento da prega vocal em estado de repouso. Em um segundo momento o autor relaciona estas duas variáveis a pressão sub-glotal e o comprimento da prega vocal em um modelo matemático unificado. De acordo com a modelagem do autor, a frequência fundamental é menos sensível ao comprimento quando as pregas vocais são curtas do que quando elas são mais longas.

Quando mais curtas, a pressão afeta de maneira mais direta o controle dos valores de frequência fundamental, assim temos uma maior variação de F_0 , conforme eleva-se a pressão sub-glotal, ao passo que em pregas vocais mais longas não se vê o mesmo efeito. (TITZE, 1989a) demonstra que pregas vocais masculinas, em média, são cerca de 1,7 vezes maiores que as femininas¹. De modo que vozes masculinas têm a frequência fundamental em torno de 120 Hz e vozes femininas em torno de 210 Hz, ou seja, cerca de 1,75 vezes maiores que as masculinas; confirmando, em partes, a hipótese do autor.

Neste sentido, nota-se que a produção da frequência fundamental é resultado de uma equação que envolve a pressão subglotal, a pressão atmosférica, as trocas de energia potenciais e cinéticas, que serão melhor exemplificados no capítulo 4. Além disso, de um ponto de vista fisiológico vimos que a morfologia das pregas vocais são determinantes para o estabelecimento de F_0 e para o *pitch range* de um falante.

¹O comprimento das pregas vocais masculinas é de 17 mm em média, enquanto as pregas vocais femininas tem cerca de 10mm.

Fica claro, também, que é a combinação destes aspectos físicos, com os aspectos fisiológicos que geram as oscilações que irão determinar a frequência fundamental e, por consequência, a entoação de uma língua. Desta maneira, a entoação é a contraparte linguística do fenômeno de variação de F_0 decorrente dos elementos explicitados nesta seção e no capítulo 2.

2.2 Abordagens da entoação

2.2.1 Definições da entoação

Na seção 2.1 fica-nos claro que a definição do objeto de estudo da entoação parece ser bastante evidente: trata-se das variações melódicas ou da variação frequência fundamental (F_0) que compõe a fala. Ainda que grande parte dos estudos de entoação tenham partido deste princípio, não há uma completa consistência quanto as teorias, objeto e objetivo dos estudos sobre entoação, como observa Vaissière (2004):

“Furthermore, there is no broad consensus as to the object and aim of intonational studies. Pierrehumbert suggests “that it is just the grammatical intonation distinctions which are properly of interest for linguists” (1980, p. 60), whereas other researchers emphasize that “the grammatical functions of intonation are secondary to the emotional one” (Bolinger, 1986, p. 260; see also Fonagy, 1983). Intonation is “a symptom of how we feel about what we say and how you feel when you say it” (Bolinger, 1989, p. 1)(VAISSIÈRE, 2004, p. 238).

Para a autora, as abordagens da entoação podem se diferenciar de forma significativa, ainda que estejamos tratando do mesmo objeto. Isso se dá, pois a forma como se define a finalidade da entoação dentro dos usos linguísticos pode variar, levando a uma reinterpretação do método. De um lado a análise dos papéis gramaticais que a entoação pode exercer (HALLIDAY, 1967; PIERREHUMBERT, 2000), isto é, seu papel dentro da língua como elemento que a organiza. Do outro lado, a visão de que a entoação contribui no aspecto semântico ou pragmático, atribuindo novos significados aos elementos linguísticos (PIKE, 1945; BOLINGER, 1989; FONÀGY, 1991).

Ao analisar esse tipo de distinção, Ladd (1996) sugere que seria necessário criar

uma Fonologia da Entoação, a qual encampasse estas definições e também se apoiasse em alguns princípios da Fonologia Segmental. Para a autor, a análise segmental poderia fomentar a discussão sobre a capacidade do falante recuperar elementos distintivos categorialmente (fonemas, traços, articulações) do *continuum* variável que constitui a língua e a partir daí construir uma teoria capaz de compreender os fenômenos entoacionais, seja eles gramaticais, sejam eles comunicativos.

Desta maneira, o estudo da entoação deveria categorizar elementos mínimos, os quais pudessem servir de referência para uma análise da língua e não somente de fenômenos locais. Destaca-se que, tal como Troubetzkoy (1976), o autor faz tais considerações, buscando elucidar que é necessário fazer alguma distinção entre *entoação* e *tom*, de modo que o segundo seriam aquelas variações tonais intrínsecas aos item lexical, como ocorrem no Thai (LEBEN, 1973) e no Mandarim (PROMON; XU; THIPAKON, 2009), ou ainda a variação funcional, como o *pitch accent* no inglês (HALLIDAY, 1967; LADD, 1996).

Nesse ponto, podemos retomar uma ideia de Lévi-Strauss (2010), que define a existência, na história das ideias, de uma "sonologia" anterior a uma "fonologia". Por este motivo, as teorias linguísticas seriam construídas a partir de uma fonética, a *sonologia*, mesmo que rudimentar, das línguas. O SPE (CHOMSKY; HALLE, 1968) e da segunda parte do CLG (SAUSSURE, 2012) seriam exemplos de como se catalogar estes sons e verificar sua função em cada uma delas. A partir deste processo surgiria uma Fonologia, que seria a etapa, em que esses sons passariam por um filtro de significação e poderiam fazer parte de um novo ciclo, amplamente baseado no valor adquirido por eles. Processos fonológicos seriam derivados dessa fonética e, com isso, muitos fenômenos poderiam ser excluídos ou não compreendidos em sua totalidade, caso ficassem no nível da sonologia, isto é, da variação qualitativa apenas.

A entoação encontra-se justamente nesse cruzamento. Sua existência sem a cadeia segmental é vista como a melodia, a organização temporal de elementos rítmicos, na

definição de Helmholtz (1895). Neste caso, a variação tonal é um objeto da sonologia, onde a frequência fundamental pode ser compreendida como a base harmônica da variação, numa existência *quasi-musical* (LIST, 1963). Por outro lado, o estudo ligando-a à cadeia segmental pode ser interpretado como um estudo do tom fonológico, ou seja, da fonologia de um elemento, a despeito do seu aspecto global, o aspecto entoacional (BOLINGER, 1951).

Com relação ao segundo item, Leben (1973) pode formular uma análise mais complexa do tom fonológico em seu modelo suprasegmental. Para o autor, as variações tonais, pensadas como elementos fonológicos, não poderiam ser compreendidas como partes do segmento exclusivamente, mas antes elementos autossegmentais, que se associariam aos segmentos em um nível acima dentro do processo fonológico, por isso de forma independente deles. Nesse sentido, os tons se associariam e sofreriam mudanças de qualidade (como o espriamento), sem necessariamente alterar o aspecto segmental ou entoacional. Liberman e Prince (1977) chegariam a uma conclusão parecida, ao propor que o acento (ou qualquer proeminência) não é uma propriedade do segmento ou da sílaba, mas antes o resultado de uma estrutura hierárquica que organiza a sílaba, as palavras e as sentenças.

O corpo teórico que se formaria em torno destes estudos, os elementos acima do segmento e seus níveis hierárquicos, foi chamado por Ladd (1996) de *Autossegmental / Metrical Phonology* (LEBEN, 1973; GOLDSMITH, 1976; LIBERMAN; PRINCE, 1977; PIERREHUMBERT, 1980) e assumiria, no bojo de sua concepção, que os processos entoacionais estariam um nível hierárquico acima do léxico ou no nível da sentença, ou seja, seriam pós-lexicais e transmitiriam informações como:

"(...)tipo de sentença ou ato de fala, ou foco e estrutura da informação. Por esta definição, a entoação exclui características como acento, *accent* e tom as quais são determinadas pelo léxico e que servem para distinguir uma palavra da outra"(LADD, 1996, p.7).

Nesta abordagem, portanto, a questão seria ponto pacífico, uma vez que os processos entoacionais estariam em um nível acima dos processos fonológicos do tom e,

por força da lógica, um nível abaixo dos processos sintáticos, mas associados a eles (GOLDSMITH, 1990a). Além disso, essa restrição permite afirmar que a entoação não pode ser intrínseca ao segmento. Desta maneira, entendemos que essa distinção é a mais coerente e melhor aceita no âmbito dos estudos linguísticos, uma vez que nos permite analisar a entoação localmente, pelo entorno de um item lexical, como o foco ou ainda pela variação de F_0 numa sentença ou em um conjunto delas. Contudo, a entoação de um único item lexical, isolado e descontextualizado parece não revelar o uso linguístico da variação de F_0 .

A partir deste ponto teríamos uma segunda questão, como os tons do nível entoacional se organizariam nesse eixo sintagmático: seriam eles organizados por níveis ou configurações? Reformulando, a capacidade do falante decodificar a entoação se dá a partir da diferença de alturas entre os tons da fala ou graças a forma como eles se organizam? Pike (1945), Trager e Smith (1957) ativeram-se a ideia de níveis tonais, estabelecendo quatro deles, chamados de *tonemas*: *E (Extra High)*, *H (high)*, *M (mid)* e *L (low)* e demonstrando como eles aparecem ou se organizam na língua, principalmente em pontos de proeminência. Em outra vertente, Bolinger (1951) se preocupou com as questões de como se configuram os padrões tonais em um eixo sintagmático e foi capaz de demonstrar que os níveis tonais não são suficientes para explicar a complexidade da variação tonal (BOLINGER, 1964, 1986).

2.2.2 Níveis vs Configuração

Pike (1945) assumia que as variações de F_0 estavam condicionadas à variações do estado do falante e/ou do contexto em que a frase era produzida. Desta forma, o *pitch* teria um significado relativo, construído a partir da relação de um de padrão de *pitch* com outro. Por este motivo, o autor precisou construir um arcabouço analítico, capaz de descrever os padrões de *pitch* (o modelo de 4 tons, numerados como 1,2,3 e 4, sendo o 1 para Extra High e assim por diante), a relação deles com outros elementos prosódicos, como o ritmo e a pausa, descritas pelos sinais (/) para pausa curtas e (//)

para pausas longas e a tonicidade (^o), embora, para o autor, a entoação tivesse certa independência destes elementos.

Além dessa análise, em Pike (1964) o autor propõe uma gramática da entoação, sugerindo que a entoação seja vista de 3 formas complementares, como *partícula*, *onda* ou *campo*, uma aproximação com as análise de física. Se a entoação é uma *partícula* da língua, pode ser ela vista como elemento estático. Porém, não é clara a posição do autor com relação a esta definição, podemos sugerir que essas partículas entoacionais seriam como segmentos da entoação, limitados pela sua própria qualidade, uma espécie de abordagem fonêmica de eventos suprasegmentais. Como *onda*, a entoação seria relativizada pela sua própria dinâmica e os significados transmitidos se dariam pelas mudanças de tom, como proposto por Bolinger (1951).

No domínio do *campo*, Pike (1964) traz um conceito da Topologia (MUNKRES, 2000) e estabelece que a entoação pode ser analisada por uma rede de contraste existente entre (i) as categorias da forma e as categorias do significado; (ii) na intersecção do léxico com a fonologia e a gramática (iii) para o *output* resultante das categorias dinâmicas e por fim (iv) para a relação entre qualidade de voz e categorias segmentais e suprasegmentais. Dentro desta perspectiva, a análise da entoação nunca estaria isolada ou fora de contexto, mas sempre correlacionada a um destes elementos, porém é preciso destacar que ela sempre estaria condicionada a variação entre os níveis pressupostos.

Para Bolinger (1951) a definição da entoação como uma coleção de tons alocados em determinados níveis não seria suficiente para análise linguística. Para o autor, uma tríade 123, seria semelhante a sequência 234, uma vez que não haveria uma referência clara para que se estabelecessem esses níveis, senão a ausência de *pitch*. Desta forma, a natureza dos tons não seria de todo considerada por aqueles que usavam os níveis em sua análise. Para resolver este problema, Bolinger (1951) sugere que o estudo da entoação perpassasse pela "configuração" tonal, que é basicamente tendência ou

característica ascendente/descendente atribuída a uma determinada frase entoacional.

Não entraremos nos pormenores deste debate que durou cerca de 30 anos e parece não ter encontrado um ponto pacífico (LADD, 1996). Como se pode ver, pelo viés dos níveis, havia uma tendência a compreender os tons como parte da fonêmica, analisando suas relações com outros elementos prosódicos e as mudanças daí derivadas. Já pelo viés da configuração, a análise se daria pela tendência ascendente ou descendente que existe naturalmente em uma determinada elocução. Nota-se, também, que a distinção entre o conceito de *tom fonológico* e *entoação* é bastante fluído, de modo tal que a relação entre tom e o núcleo da sílaba é debatida por Pike (1945).

O problema derivado destas abordagens é que, se existe um relativismo nos tons estabelecidos por Pike (1945) e Trager e Smith (1957), existe também uma necessidade epistemológica de se considerar as variações de F_0 como níveis quando se faz uma análise de sua configuração (LADD, 1996), isto é, de alguma maneira é preciso considerar a variação vertical da frequência fundamental. Neste sentido, o problema toma outra dimensão, pois a noção de configuração teria de vir *a posteriori*, em tese, já que a implementação dos tons, como alvos, seria o preceito fundamental da entoação. Alguns trabalhos, como os de Halliday (1967), Crystal (1969) e Maeda (1976) passaram a incorporar as duas visões.

Halliday (1967) adota um sistema de 5 tons simples e 2 compostos, de modo que ele compreende que estes sons se espraíam por ao longo de um trecho da elocução e tem seu núcleo na primeira sílaba do pé silábico. Desta maneira, a tonalidade é relacionada aos conjuntos tonais presentes em uma elocução. Para Crystal (1969) a definição de *pitch* passava por questões mais complexas e estava relacionada à prosódia como um todo. Desta forma, a entoação seria a relação entre o tom, a variação da frequência fundamental até chegar esse tom e pela direção do *pitch*. O tom teria um núcleo de proeminência, estabelecido pela variação de F_0 ou pela sustentação de F_0 na sílaba tônica do item lexical mais proeminente da sentença. Para o autor, os tons

nucleares seriam divididos em dois tipos: de queda e de ascensão, se considerado os *pitch movements* a partir do núcleo.

Maeda (1976) se apoia no sistema de 4 tons de Pike (1945, 1964) para a análise da entoação do inglês americano. No entanto, para o autor a entoação deveria ser analisada apenas em termos de variação de F_0 , sem considerar outros elementos que constituem a prosódia da língua, desta forma a entoação teria 5 atributos: BL (*Baseline*), R (*Rise*), P (*Peak*), L (*Lowering*) e R1 (*Rise Plateau*). Além disso, para Maeda (1976) a variação de F_0 estaria associada as duas componentes: (a) a componente não-localizada e (b) a componente localizada.

A componente não-localizada seria *baseline*, também chamada de declinação frasal, definida pela queda gradual nos valores de F_0 ao longo de uma frase (BOLINGER, 1958; COHEN; T'HART, 1967). Desta forma, a *baseline* seria a referência tonal para os movimentos das componentes não-localizadas. A hipótese de Maeda (1976) se sustentava na proposta de Cohen e t'Hart (1967), o qual sugeria que os movimentos tonais não poderiam ser analisados sem que se levasse em consideração a declinação frasal, chamada por Maeda (1976) de *baseline*.

Sendo assim, as componentes localizadas dependeriam da variação da *baseline* ao longo do tempo e seriam definidas não pela sua qualidade, mas, sim, pelo seus movimentos, no caso específico, o aumento e a queda pontual, representados pelos atributos (R) e (L) e teriam a função de marcar acentos lexicais em palavras específicas. Junto destes dois atributos básicos, haveria mais dois, o *peak* (P) e *rise plateau* (RP), sendo que estes dois últimos representariam a forma do contorno. A entoação, então, seria definida pelas relações destes atributos em uma elocução e pela forma como eles se associam para gerar determinados significados e não apenas pelas suas variações estabelecidas por atributos prosódicos, como o tonicidade ou pela configuração global.

Entretanto, apenas com a tese de Pierrehumbert (1980) que se pode ter uma visão mais integral de como a interface entre a hipótese dos níveis tonais e a hipótese das

configurações entoacionais poderia explicar a competência do falante em compreender e produzir a entoação. De um ponto-de-vista teórico, Pierrehumbert (1980) assumiu que a entoação seria a variação de F_0 em um domínio mais amplo do que o silábico, tendo em mente que é a sílaba a portadora da entoação. De um ponto-de-vista metodológico, sua tese aperfeiçoou o sistema de 4 tons de Pike (1945), reduzindo-o para apenas 2 tons [H] e [L] e incorporou os conceitos introduzidos por Maeda (1976), estabelecendo uma relação entre estes tons com um referencial frasal, a *baseline*, as variações decorrentes dos movimentos tonais e, por fim, a forma destes movimentos.

2.2.3 O modelo de Pierrehumbert (1980)

Lieberman (1975) e Maeda (1976) usaram-se da análise instrumental para construir o corpo de análise em suas teses, um importante aspecto adotado na metodologia de Pierrehumbert (1980) e bastante distante das análise impressionísticas feitas pelos opositorista "nível-configuração"(PIKE, 1945; BOLINGER, 1951). Após definir os níveis tonais e a metodologia, o próximo passo da autora foi o de estabelecer as restrições sintagmáticas, as quais os tons estariam sujeitos. A figura 2.10 demonstra como seria a distribuição possível dos tons em uma cadeia sintagmática.

Figura 2.10: Esquema de organização sintagmática no modelo de Pierrehumbert (1980)

Na figura 2.10 os tons que vão de 0 a 1 são chamados de tons de fronteira (*Boundary*

tones), marcados pelo símbolo (%), ao passo que os tons que vão de 1 a 2 são os acentos tonais (*pitch accents*), os de 2 a 3 são os acentos frasais (*Phrase accent*) e aqueles que vão de 3 a 4 são os tons de fronteira (*Boundary tones*). Desta forma, as proeminências tonais podem assumir diversas funções gramaticais. Os *pitch accents*, por exemplo, estariam alinhados com a sílaba tônica, ao passo que os tons de fronteira, marcariam as fronteiras lexicais independentemente da sílaba tônica. Os *pitch accents* poderiam ter um ou dois tons, o asterisco (*) é o diacrítico que marca o tom que recai sobre a sílaba tônica. Em um acento bitonal, o tom não marcado cai nas imediações do tom marcado, seja na mesma sílaba, seja em uma das sílabas próximas.

Pierrehumbert (1980, p.18) assume duas hipóteses básicas. A primeira é a de que os tons, organizados como níveis [H] e [L], representariam alvos tonais, a segunda é que a variação segue um padrão, no caso o movimento tonal, representado pela *baseline*, tal qual Maeda (1976). No entanto, Maeda (1976) compreendia esse valor como um decaimento linear que ocorreria ao longo da elocução e que estabeleceria os valores de referência dos tons, já Pierrehumbert (1980) entende que a *baseline*, além de representar esse decaimento natural, é também o valor mais baixo de frequência fundamental ao qual um falante pode chegar em um dado enunciado.

Desta maneira, a *baseline* para Pierrehumbert (1980, p.135) não é um valor referencial para o estabelecimento dos demais tons, já que essa referência é estabelecida pela diferença entre as proeminências e a especificação tonal em si e não pela *baseline*, como supõe Maeda (1976). Essa definição parte do pressuposto de que todos os falantes possuem uma base para a F_0 , a qual se mantém constante, mesmo que haja uma variação de *pitch* ou que comprimento da elocução variassem, uma vez que todo falante apresentaria um declínio na *baseline* característico de sua própria voz.

O modelo de (PIERREHUMBERT, 1980) é bastante competente no que se refere a descrição de padrões tonais, de tal maneira que foi a base do modelo ToBI (BECKMAN; PIERREHUMBERT, 1986; SILVERMAN et al., 1992), o qual tem por objetivo, justamente,

contribuir para a descrição dos padrões entoacionais nas mais variadas línguas (JUN, 2006). É preciso ter em mente que o modelo de Pierrehumbert (1980) é amplamente baseado na gramática e tem como objetivo principal a descrição da implementação dos contornos de *pitch* que definem os *pitch accents*, antes de uma análise da distinção fonológica destes tons ou de uma análise semântica destas curvas (PIERREHUMBERT, 1980, p.59).

Retomando a questão dos níveis de proeminência, proposto por Liberman e Prince (1977), Beckman e Pierrehumbert (1986) compreendem que determinados processos fonológicos/ entoacionais, como a proeminência frasal, pode estar em nível acima do léxico e um nível abaixo da sentença, enquanto outros processos, como os tons de fronteira, descritos por Pierrehumbert (1980) estariam somente no nível frasal. Os problemas associados a essas hierarquias de proeminência e sua relações com a língua seriam melhor explorados dentro do modelo de Fonologia Prosódica, criado por Nespor e Vogel (2007).

2.2.4 INSTINT Hirst e Cristo (1998)

Outros autores que se propuseram a criação de um modelo universal de análise da entoação foram Hirst e Cristo (1998), os quais destacam a ambiguidade existente decorrente do próprio termo, segundo os autores:

The first ambiguity depends on whether or not intonation is defined in a broad sense, that is as including factors such as word-stress, tone and quantity which can be an essential part of the lexical identity of words, or in a narrow sense, as excluding such factors. (HIRST; CRISTO, 1998, p.4)

A oposição entre fatos lexicais (tom, acento e qualidade) e fatos não-lexicais (entoação) derivam em uma série de novas distinções. Por exemplo, o aspecto *físico*, que seria a análise do fenômeno da variação de F_0 , junto de aspectos fonéticos que interfiram na variação de *pitch*, como a duração segmental e a redução vocálica, seria diferente do aspecto *formal*, que seria um nível abstrato e estaria relacionado ao aspecto

linguístico. No entanto, para os autores dicotomias como essas sempre caem em uma espécie de falácia metodológica, uma vez que elas restringem a possibilidade de se compreender a entoação como um fato linguístico que estabelece uma relação entre o falante e o ouvinte, por exemplo.

Nesta perspectiva, é necessário uma interface entre estas dimensões, nos termos do que propõe Ohala (1990), isto é, compreender a correlação necessária e direta existente entre um nível físico e um nível cognitivo. Para explicar essa distinção e, também, a interface, Hirst e Cristo (1998) estabelecem uma diferença entre (i) sistemas prosódicos e (ii) parâmetros prosódicos. Os sistemas prosódicos seriam o aspecto cognitivo e abstrato, associados a linguagem, além disso, seriam divididos em dois sistemas, o lexical e o não-lexical, ao passo que os parâmetros prosódicos representariam a emergência ao nível físico, coordenado pelos aspectos acústicos e fisiológicos da entoação. Seguindo esta distinção, Hirst e Cristo (1998) propõem uma redefinição dos conceitos apresentados e sugerem que a entoação seja vista como um mecanismo de implementação dos parâmetros acústico-prosódicos, associados aos objetos pertencentes aos sistemas lexical e não-lexical.

O modelo de Hirst e Cristo (1998) derivou um sistema de notação, chamado INTSINT (**I**Nternational **T**ranscription **S**ystem for **I**NTonation). Para os autores, o INTSINT teria, nos estudos da linguagem, a mesma função que o IPA (International Phonetic Alphabet), porém devotado à anotação prosódica, uma vez que foi proposto a partir do estudo de um conjunto de 20 línguas (HIRST; CRISTO, 1998a). Ao invés de definir os movimentos tonais, os autores partem do pressuposto topológico, de modo tal que é preciso definir a posição do tom, ou o alvo tonal. Sendo assim, as variações podem ser (HIRST; CRISTO, 1998, p.15):

- (a) Em relação ao ponto imediatamente anterior: H (*Higher*), S (*Same*) e L (*Lower*);
- (b) Em relação ao movimento tonal feito, a partir do tom anterior: < (*Upstep*) e > (*Downstep*);

- (c) Em relação ao *pitch range*² de um falante: T (*Top*) ou B (*Bottom*). A notação M (*mid*) serviria para registrar o tom inicial.

Ao curso das análises feitas em Hirst e Cristo (1998a) observa-se que novas notações são sugeridas, em geral, elas derivam da necessidade metodológica de se explicar determinados fenômenos. Um exemplo seria o *high-fall* do Português Europeu, descrito por Cruz-Ferreira (1998), e que representa uma diminuição significativa de F_0 , de modo tal que o tom estaria além do limite inferior B, descrito por Hirst e Cristo (1998). Em nosso entender, o modelo de Hirst e Cristo (1998) se mostra competente na descrição dos fenômenos entoacionais das mais variadas línguas, uma vez que dentro de sua concepção metodológica, ele é capaz de colocar em voga a questão dos problemas conceituais sobre a entoação, tendo em vista que incorpora as questões da variação de tom locais e globais.

2.3 Papel da entoação na fala

Existem alguns pontos relativamente importantes dentro da discussão do papel da entoação na fala e que estão relacionados ao modo que se compreende a própria entoação. No entanto, não discutiremos esses problemas, uma vez que não nos parece pertinente neste momento, embora uma revisão teórica seja válida.

Os modelos apresentados na seção anterior e que analisaram em entoação dentro de um programa linguístico, como de Pike (1945), Bolinger (1951) e Pierrehumbert (1980) têm valorizado o aspecto estático de possíveis realizações, ou seja o *status semântico* e representacional de determinados tons com relação a outras instâncias linguísticas, ao mesmo tempo que têm desvalorizado o aspecto dinâmico da própria variação de F_0 , o aspecto da *forma*. Por outro lado, modelos baseados em questões cognitivas t'Hart, Collier e Cohen (1990), o qual vamos analisar no capítulo 3, de modelamento (MAEDA, 1976) ou associadas à fonética (LADD, 1983) têm seguido o

²Tessitura entoacional (CAGLIARI, 1983).

caminho inverso e valorizado a forma, a despeito do *status* semântico.

Essa questão parece limitar o modo como compreendemos o significado e o papel da entoação na fala. Se olharmos pelo prisma do *status* semântico de um determinado tom, temos de construir um sistema capaz de deduzir os padrões fonológicos a partir das variações de F_0 , tal como faz a Fonologia Segmental. No entanto, a entoação é cercada de variações locais e globais que afetam a sua forma, de tal modo que a variação parece ser o padrão. Na tentativa de encaixar padrões que demonstram uma representação subjacente a essa variação, estes sistemas constroem mecanismos que são abstrações impressionísticas, como o sistema de tons de Pike (1945) ou construções que relativizam o valor de um tom, como o de Pierrehumbert (1980).

Por este motivo, dentro desta dicotomia, um conjunto significativo de fenômenos relacionados à forma da entoação foram colocados no guarda-chuva das idiossincrasias, estilo de comunicação e necessidades expressivas do falante, pois não possuem uma representação mental capaz de ser mapeada de forma representacional. Sendo assim, estas manifestações não fazem parte da língua e, sim, da fala, uma vez que semelhança de forma não representa semelhança de conteúdos (PIERREHUMBERT, 1980, p.60), mas padrão de forma representa função linguística (SAUSSURE, 2012, p.72). Vamos tratar desse assunto com mais profundidade no capítulo 4.

Na outra face desta dicotomia, quando olhamos o significado, está a questão mais ampla: qual o real *status* linguístico da entoação? Alguns fenômenos demonstram uma grande regularidade quando analisados pelo aspecto representacional hipotetizado por Pierrehumbert (1980), porém não parece o suficiente para se dizer que eles são universais. Tendo em vista que os dois principais programas de pesquisa linguística do século XX se pautaram pela busca ou explicação de fenômenos universais, é natural que o estudo da entoação tenha sido restrito à análise desse conjunto de fenômenos. Para Ladd (1996, p.114), é possível mapear alguns destes aspectos, embora seja difícil considerá-los universais:

1. A tendência de que os valores F_0 decaiam paulatinamente ao longo de uma elocução, gerando o decaimento tonal, que parece ser um elemento determinante para a percepção de finalização;
2. A utilização da elevação de *pitch* em final de frase para demarcar uma interrogativa do tipo sim ou não;
3. Proeminências tonais com a finalidade de marcar itens lexicais de especial interesse em uma elocução, o foco.

Adicionaríamos mais um item a essa lista:

4. Transmitir informações relacionadas ao contexto e ao falante, por meio de estruturas entoacionais.

Observando a lista de Ladd (1996) podemos sugerir que a entoação tem um papel pragmático (LINDSEY, 1981; WARD; HIRSCHBERG, 1985) bastante significativo ou exerce funções fono-sintáticas (BECKMAN; PIERREHUMBERT, 1986; NESPOR; VOGEL, 2007). No primeiro caso, ela transfere significados do contexto para o interior da sentença, servindo de apoio semântico no nível frasal, como atestam os itens 2, 3 e 4. Em certo sentido, a variação de frequência fundamental ajuda o ouvinte a determinar quais informações sobre as atitudes do falante e quais conteúdos pragmáticos foram transferidos para a fala.

Se olharmos pelo aspecto da consciência linguística dos falantes de uma língua é possível notar casos em que estes reportam-se a uma determinada sentença dizendo ela não deveria ser pronunciada daquela forma, uma vez que o "tom" foi inadequado. De acordo com Wittgenstein (1953, sec.219), os detalhes de alguns fenômenos, como a linguagem, as emoções e o domínio da consciência, passam despercebidos para os seres humanos, uma vez que estamos excessivamente familiarizados com eles. Neste sentido, pode-se dizer que a consciência dos falantes sobre a papel da entoação revela o quão conectada ela está ao aspecto pragmático e comunicativo.

Como citado mais acima, essa análise que envolve o aspecto pragmático da língua parece muito mais estar associada à forma do que ao conteúdo ou ao domínio representacional, de modo que determinar uma fonologia que esclareça todos esses fenômenos associados à pragmática parece ser improvável, uma vez que o conjunto de contextos é inumerável. Entretanto, acreditamos que seja possível estabelecer uma metodologia linguística capaz de definir alguns padrões utilizados pelos falantes, não como parte de idiossincrasias, mas como elementos pertencentes a língua, justamente por meio do conceito de alvos tonais, apresentados no capítulo 3

Com relação ao aspecto sintático, a consciência linguística dos falantes não parece ser tão evidente. A literatura do tema sugere que a entoação tem o papel de determinar, principalmente, fronteiras de frases, como sugere o item 1 e o estabelecimento de unidades sintáticas (LADD, 1996). Schmerling (1976) sugere que, no inglês, um falante modula sua entoação com a intenção de marcar a diferença entre um argumento sintático e um predicativo. A clivagem e o deslocamento de constituintes parecem exercer algum papel na alteração de F_0 , porém, seria necessário um estudo sistemático para demonstrar esse efeito.

A noção mais importante é a de que a entoação desempenha um papel de segmentação sintática, estabelecendo "unidades sintáticas", por meio da demarcação de limites perceptuais entre constituintes ou sentenças. Porém, de maneira geral, é difícil rastrear um conjunto significativo de pistas entoacionais que tenham funções exclusivamente sintáticas. Modelos como de Nespor e Vogel (2007), que tentaram desenvolver esse tipo de abordagem, também tiveram que lidar com a hipótese de que as unidades sintáticas estão correlacionadas a outros aspectos prosódicos, como o acento ou processos morfo-fonológicos.

Ainda assim, os fenômenos relacionados à sintaxe não parecem ser bem explicados ou não mantêm uma regularidade significativa nas línguas. Como dito acima, a entoação parece ter um papel muito mais crítico na percepção de unidades sintáticas

do que na determinação destas unidades em si. O caso da interface morfologia-fonologia parece ser o mesmo, de modo que é improvável pensar que a entoação desempenhe algum papel morfológico ou lexical, uma vez que isso a colocaria como parte do tom lexical.

A questão da entoação servir como pista perceptual para determinar unidades linguísticas, parece ser um problema comum e que já foi discutido por outros autores, como Pierrehumbert (1980) e t'Hart, Collier e Cohen (1990). Em certo sentido, a tentativa de mapear os padrões representacionais leva a crer que entoação desempenha um papel muito mais perceptual, recorrendo a intuição do falante para determinar unidades linguísticas ou significados, sem necessariamente haver um domínio representacional. De modo que o julgamento dos falantes não parte de aspectos relacionados à linguagem propriamente, isto é, o papel da entoação na linguagem é exercido de forma quase indireta, variações determinam aspectos linguísticos e não o contrário.

Esse parece ser um dos problemas mais complexos quando analisamos a questão do papel da entoação na fala. A literatura demonstra que essa relação não é evidente, se ela é um fenômeno puramente perceptual é possível que a questão da forma seja, de fato, o elemento relevante para a análise linguística. Por outro lado, pode ser que essa forma só tenha significado, pois ela é implementada a partir de categorias representacionais internalizadas profundamente na língua. De toda maneira, existe uma forma entoacional que emerge na fala e pode ser compreendida como categoria representacional, por isso atrelada a funções essencialmente estruturais e linguísticas, ou como simples uma forma, a qual é capaz de transmitir conteúdos que estão contidos no contexto.

A nosso ver, as duas respostas estão corretas. A entoação exerce este duplo papel, operando como elemento estruturador e, por isso, associada a um nível representacional, e como elemento semântico, incorporando os significados do contexto, como sugere Ferreira-Netto (2008). Contudo, para que essa análise possa ser produtiva

precisamos utilizar uma unidade que opere entre os dois campos e seja capaz de implementar as duas funções. No capítulo a seguir, sugerimos o alvo como essa unidade.

3 Alvos Tonais

No capítulo 2 nos dedicamos a explicar a produção da entoação, bem como os modelos teóricos que abordaram o fenômeno da entoação e algumas de suas funções na língua. Neste capítulo nos dedicaremos a apresentar o que chamamos de *alvos tonais* e o que eles implicam em nossa hipótese. Além disso, apresentamos o modelo utilizado por nós nessa tese e desenvolvido por Ferreira-Netto (2008).

3.1 Alvos tonais

3.1.1 Questões de Linguagem e questões de informação

No capítulo anterior focamo-nos naqueles modelos que levam em consideração, pelo menos em algum nível, a correlação existente entre o aspecto físico e o aspecto simbólico. Nota-se que em todos eles, exceto os de Bolinger (1951) e Hirst e Cristo (1998), propõe-se o estabelecimento de uma definição de entoação, que a explique como um recurso prosódico, caracterizado pelas proeminências positivas ou negativas e associado de forma significativa com o acento ou outras características rítmicas, como as pausas ou as fronteiras de frase e palavra.

Entendemos que essa preocupação epistemológica é vital para alguns modelos, uma vez que a definição destas propriedades ajudaria a compreender como falantes e ouvintes organizam e mapeiam as estruturas mentais subjacentes à linguagem (DILLEY; SHATTUCK-HUFNAGEL; OSTENDORF, 1996; SHATTUCK-HUFNAGEL et al., 2012). No entanto, ao mesmo tempo, parece-nos difícil que essas abordagens

consigam quantificar os fenômenos informacionais por elas analisados, o que, por outro lado, parece ser a tônica das pesquisas que envolvem a fala emotiva (SCHERER, 2003; BANZIGER; SCHERER, 2005).

A nosso ver, os modelos descritos no capítulo 2, em sua maioria, não são capazes de descrever fenômenos quantificáveis, senão em termos simbólicos ou da relação entre estes elementos simbólicos e a linguagem. Desta forma, a materialidade da variação de F_0 é descartada em função de uma sistematização perceptual das proeminências, variações e mudanças tonais que ocorrem em uma elocução, relegando o aspecto físico a uma fonética vagamente dissociada do fenômeno. É o caso do modelo de Pierrehumbert (1980), por exemplo, onde um tom H é visto como uma unidade fonológica que especifica uma altura relativa, mas não apresenta um modelo capaz de explicar a sua implementação na fala.

Em certo sentido, o instrumental de análise destas pesquisas, apoia-se no aspecto fonético para demonstrar ou exemplificar uma série de unidades fonológicas prototípicas, tal como é feito na Fonologia Segmental. Mas é preciso lembrar que dentro do estudo dos segmentos é natural que se encontre uma estabilidade significativa entre manifestações, o que é mais e mais improvável na entoação. Por este motivo, análises desenvolvidas no âmbito destes modelos representacionais nos dão acesso a uma variada gama de descrições linguísticas, as quais nos permitem compreender os fenômenos entoacionais em seu uso mais abstrato, quando alocados como objetos de um sistema linguístico tão somente (GODJEVAC, 2000; GUSSENHOVEN, 2006; JUN, 2006; LUCENTE, 2008).

Entretanto, a nosso ver, esse mesmo arcabouço teórico pouco explica sobre os fenômenos comunicativos/ expressivos derivados deste processo de abstração linguística, isto é, eles são capazes de explicar os elementos representacionais e linguísticos utilizados pelos falantes, mas não suas finalidades informacionais. Pierrehumbert (1980), por exemplo, define que função expressiva é uma das propriedades da entoação,

a qual, no entanto, não pertence ao escopo da sua tese, uma vez que seu modelo lida com elementos representacionais e não com o aspecto da variação, daí a ideia de que padrão não representa significado. Podemos entender que para a autora os significados subjacentes da entoação, aparte do aspecto gramatical, são, muitas vezes, construídos de forma local, ou seja, pelo uso da língua em si. Deste modo, os aspectos universais, objetivo de modelos linguísticos, são incapazes de compreender essas variações, as quais pertencem a uma sistema contíguo ao linguístico, no caso, a fala.

O *status* e o papel dessa variação na língua têm sido uma observação constante e bastante importante dentro de alguns modelos teóricos defendidos por autores, como t'Hart, Collier e Cohen (1990) e Xu (2004, 2005), Promon, Xu e Thipakon (2009). Algumas abordagens mais recentes, como a tese de Dilley (2005) e Cho (2010) têm se baseado na ideia de que as relações entre as variações de F_0 e outros aspectos, sejam eles prosódicos ou informacionais, não são óbvias, de modo que as unidades fonológicas previstas em modelos representacionais são incapazes de explicar a entoação como um todo. Sendo assim, é preciso mais do que uma análise hierárquica que pressupõe uma correlação entre o núcleo silábico e o pico entoacional (H), por exemplo, para que se determine a função linguística e informacional deste fenômeno, uma vez que, aparentemente, é o contexto que proporciona a sua ocorrência (DILLEY, 2005).

Para Dilley (2005) essa é a questão central no domínio da entoação. Para a autora, a avaliação que se faz do objeto a ser analisado somado ao método de análise escolhido, necessariamente, incorporam questões que estão além da mera manifestação de F_0 , tais como o contexto ou outros eventos. No entanto, Dilley (2005) entende que a forma como estes dados são correlacionados podem determinar uma visão mais restrita, focada na descrição de unidades fonológicas ou mais ampla, focada na análise da implementação dessas unidades como recursos comunicativos. Sendo que a primeira visão tem sido a corrente nos estudos linguísticos, ao passo que a segunda tem sido incorporada pela área de *speech analysis*.

3.1.2 O modelo de t'Hart, Collier e Cohen (1990)

O modelo de t'Hart, Collier e Cohen (1990) é um exemplo de como o problema de relacionar aspectos comunicativos aos aspectos linguísticos pode ser encarado. Antes de mais nada, é preciso dizer que o sistema de análise desenvolvido pelo IPO (*Instituut voor Perceptie Onderzoek*, em holândes) não deve ser tomado como um modelo estritamente linguístico, tendo em vista que seu objetivo é fazer uma descrição da percepção dos tons, pertençam eles ou não ao domínio da entoação. Por este motivo, os problemas relacionados à comunicação ou ao aspecto informacional contido nos tons é pressuposto antes da estrutura linguística por mecanismos cognitivos ou perceptuais.

Para os autores, a entoação deve ser compreendida como a variação de F_0 no decurso de uma elocução, sendo que certos movimentos tonais (*pitch movements*) são perceptualmente relevantes pelos sujeitos. Estes movimentos tonais no plano fisiológico/cognitivo corresponderiam a comandos para movimentos específicos das pregas vocais, os quais podem ser recuperados no sinal da fala. Estas variações seriam voluntárias, ou seja, comandadas pelo falante, junto delas haveriam as variações involuntárias, que não teriam relevância para o ouvinte e não seriam coordenadas por ele. Por exemplo, a manutenção da variação de F_0 não seria voluntária e, sim, o resultado de características próprias do processo de produção da fala, por isso sujeita a pequenas oscilações involuntárias.

É possível supor que estas oscilações involuntárias, que compõem aquilo que os autores chamam de *microentoação*, não tenham um impacto significativo na percepção do *pitch*, quando comparadas com aquelas variações programadas pelos falantes (T'HART; COLLIER; COHEN, 1990; WHALEN; LEVITT, 1995). Por este motivo, para que uma variação de F_0 seja interpretada como relevante não basta que ela seja associada a um determinado fato linguístico, como um núcleo silábico ou um evento psicofísico, mas que esta variação também esteja relacionada a algum tipo de atitude do falante. Neste sentido, a percepção teria um papel preponderante, pois ela seria

capaz de avaliar se uma determinada proeminência é relevante para um sujeito, desde que essa variação seja fisiologicamente executável e que as regras da língua aceite essa determinada variação.

No Português Brasileiro, o trabalho de Consoni (2011) demonstrou, por meio de testes de percepção em palavras isoladas, que variações da ordem de 3 semitons ascendentes ou 4 semitons descendentes são suficientemente representativas para os falantes, o que chamamos de *limiar de diferenciação tonal* (T'HART, 1981; CONSONI, 2011; MARTINS, 2013). O trabalho de Consoni (2011) também revela o condicionamento fisiológico subjacente, pois é possível sugerir que alterações acima de 3 semitons ascendentes podem ser de difícil execução para os falantes. Martins (2013), replicando os testes de Consoni (2011) no nível da frase, demonstrou que as mesmas regras se aplicam ao âmbito frasal, de modo que estes estudos revelam que a variação sistematicamente programada é perceptível para o falante independente da hierarquia, desde que preencha alguns requisitos perceptuais.

Uma das vantagens de um modelo de análise como o proposto por t'Hart, Collier e Cohen (1990) é a capacidade subjacente de organizar os significados dos elementos que os compõem com base em outros processos cognitivos e supor que eles são atualizados em tempo real na estrutura da língua. Assim, um fenômeno linguístico como o foco lexical, por exemplo, só poderá ser considerado como tal quando o falante, durante o ato de produção da fala, ultrapassar o limiar de diferenciação tonal acima citado, o que tornaria aquele ponto específico dotado de um significado. Desta forma, um processo linguístico é dependente da percepção, que precisa se conectar a estrutura linguística para atribuir significados seja para seu falante, seja para seu ouvinte.

De um ponto-de-vista da percepção esse parece ser um problema menor, pois a própria percepção é capaz de fazer as avaliações necessárias para julgar se um determinado evento é dotado de relevância ou não e qual o grau de relevância deste evento. A outra face deste problema se dá quando precisamos supor que alguns

processos fonéticos se associam em tempo real a estes processos cognitivos, de modo que ajustes de implementação podem ser feitos para compensar uma debilidade ou para supervalorizar uma variação, dentro dos limites supostos por t'Hart, Collier e Cohen (1990).

A FAR tem se debruçado sobre estes pressupostos e para Albano (2009, p.53) a noção de tarefa (SALTZMAN; KELSO, 1987), que subjaz a definição de gesto, tem se mostrado um importante conceito para explicar a correlação entre o físico e o simbólico. Para a autora, uma tarefa não supõe uma implementação puramente mecânica de trajetórias pré-estabelecidas, mas, antes, a noção de uma tarefa a ser cumprida, de modo que essas trajetórias podem ser ajustadas em tempo real, o que permite aceitar determinadas oscilações. Neste sentido, as unidades fonológicas, como elementos simbólicos, podem sofrer alterações no nível da implementação, as quais não representam mudanças em si, mas antes explicam variações encontradas no sinal de fala.

No caso da entoação, essa definição é de suma importância, até mesmo dentro do modelo AutoMetrical Phonology (LADD, 1996). Se considerarmos que os tons H e L representam unidades fonológicas, as quais se referem a picos e vales de F_0 , é necessário considerar um sistema de implementação para se atingir esses pontos. Em Xu e Wang (1997) os autores sugerem que algumas dessas variações F_0 poderiam ser explicadas pela interação entre duas forças independentes: (i) as exigências linguísticas e (ii) as restrições mecânico-fisiológicas. As exigências linguísticas determinariam, por exemplo, variações intencionais de F_0 e poderiam ser implementadas por comandos neurais, ao passo que as restrições mecânico-fisiológicas seriam condicionadas pelas leis físicas e pelas propriedades fisiológicas dos articuladores, tais como as características mecânicas dos articuladores ou a velocidade de reação dos músculos que os compõem, de modo que elas não seriam voluntárias.

Para Xu (2005) as unidades fonológicas não especificam o papel da entoação ou da

melodia da fala em uma língua, uma vez que representam apenas funções gramaticais dos tons e justamente por essa opção não são capazes de explicar os detalhes fonéticos de sua implementação, que podem conter dados informacionais. Para que o estudo da entoação incorpore estas dimensões, Xu (2005, p.221) sugere dois princípios básicos: (i) a melodia da fala transmite determinadas informações comunicativas e (ii) essa transmissão é feita pelo próprio sistema que produz a fala, por isso pode ser articulada.

A primeira ideia é bastante aceitável. Como se sabe, a entoação de uma língua que não utiliza as variações de F_0 em seu nível lexical, como o Português Brasileiro, tende a utilizar essa característica para fins informacionais, como demonstram Moraes (1998), Cagliari e Massini-Cagliari (2001) e Ferreira-Netto (2006), que sugerem que determinadas variações de F_0 têm a função principal de transmitir uma informação e, em segundo plano, contribuir para resolver ambiguidades prosódicas. Alguns trabalhos como os Moraes e Colamarco (2007) e Moraes, Carnaval e Coelho (2010) têm demonstrado, por exemplo, que pistas entoacionais maiores que uma palavra fonológica são pouco utilizadas pelos falantes para compreender frases interrogativas.

Em outra frente, a da finalização, os trabalhos de Ferreira-Netto, Consoni e Peres (2009, 2010) demonstram que os falantes têm a tendência a marcar a finalização de frases, durante a leitura, em uma cadência maior dominante-tônica, enquanto que na fala corrente a mesma tendência se segue, embora de forma mais aleatória. Garcia (2015) demonstra que para o dialeto caipira da região do médio tietê essa tendência vai direção a uma finalização plagal. Baz e Ferreira-Netto (2011) sugerem que apenas variações da ordem de 4 semitons descendentes ou mais são significativos para os ouvintes perceberem uma finalização, ao passo que Rosa (2015) demonstrou que com uma variação da ordem 7 semitons falantes tendem a compreender com absoluta certeza ou média certeza que uma determinada frase está finalizada.

Outros estudos, como o de Bodolay (2009) e Moura (2016) têm revelado que pistas prosódicas utilizadas na entoação são importantes para a detecção da atitude do

falante. No primeiro caso refere-se à distinção contextual feita pela prosódia para demonstrar pedidos e ordens, já o segundo se foca na análise da manifestação do *éthos* discursivo em debates políticos por meio da variação de F_0 . Vassoler (2011) e Martins e Vassoler (2013) demonstraram que a variação global ascendente de F_0 é um forte indício para a compreensão da raiva na fala, ao passo que a variação descendente seria importante para a compreensão da tristeza. Resultados semelhantes foram encontrados em Ferreira-Netto et al. (2014), Ferreira-Netto, Martins e Vieira (2014) e Ferreira-Netto e Martins (2017) fomos capazes de identificar, por meio da análise de *pitch*, um simulacro de entoação neutra, utilizada por pacientes esquizofrênicos para maquiar suas emoções.

Como se pode ver, a entoação ou a prosódia, em sentido amplo, tem um papel significativo para a transmissão de informações complementares em uma grande quantidade de contextos, além, é claro, daqueles lexicais utilizados por línguas tonais. Xu (2005) argumenta que uma grande parte dessa informação transmitida pela entoação é negligenciada quando colocada em um modelo que valoriza as unidades fonológicas, pois não existe uma preocupação com a análise fonética dessa dimensão, para isso o autor sugere que alguns parâmetros acústicos devem ser pensados para estruturar essa passagem do componente fonológico para a fala

A noção ideal seria a de *alvos tonais*, a nosso ver um equivalente entoacional do gesto fonológico (BROWMAN; GOLDSTEIN, 1989). Xu e Wang (2001) definem alvo tonal como: *pitch target are the smallest articulatorially operable units associated with linguistically functional pitch units such as tone and pitch accents.*(XU; WANG, 2001, p.321). D'Imperio (2002) propõe que alvos tonais são posições topológicas, definida em um eixo cartesiano de altura e tempo, marcados pelo início e fim de um *pitch movement* (D'IMPERIO, 2002). Na realidade, estas definições são extensões do mesmo problema, uma vez que para Xu e Wang (2001) o alvo tonal pressupõe a existência de fases de transição do início até o seu fim, como se pode ver na figura abaixo:

Figura 3.1: Representação esquemática de um alvo tonal exemplificado. Retirado de Xu e Wang (2001)

As linhas pontilhadas representam os alvos tonais como sugeridos pelos autores. Alvos podem ser dinâmicos (à esquerda) quando têm uma aceleração pressuposta (Rise) ou então estáticos (à direita), quando se mantêm constantes ou próximo disso. A linha sólida representa a realização fonética destes alvos, ao passo que as linhas verticais representam as fronteiras e pontos de transição destes alvos. Em tese, as unidades fonológicas subjacentes, seriam um tom R (*Rise*) e um tom L (*Low*). No entanto, a variabilidade existente no sinal de fala real é bastante diferente, de modo que o falante tende a uma aproximação de um alvo ideal, utilizando os recursos articulatórios subjacentes.

Por este motivo, a implementação fonética precisa respeitar alguns condicionantes. O primeiro é duracional, um alvo não pode durar mais do que o seu portador, seja uma sílaba, seja uma unidade maior, isto é, ele precisa iniciar em seu *onset* e terminar em seu *offset*. No Português Brasileiro temos encontrado fortes evidências da existência de uma Unidade Portadora de Entoação (UPE). Em Ferreira-Netto et al. (2012) definimos que as condições de existência destas unidades são: frequência fundamental maior que 20Hz e menor que 600Hz, intensidade de ao menos 10% do valor médio da intensidade total e a duração de ao menos 20 ms.

Estes números foram calibrados, a partir de um teste de percepção, no qual foi requisitado a um conjunto de falantes treinados que indicassem quais seriam as sílabas

ou unidades entoacionais de uma determinada frase. A seguir, a mesma frase foi gravada por 11 falantes do sexo feminino e testada na Rotina ExProsodia, na tentativa de detectar de forma automática as unidades percebidas pelos falantes. Em Ferreira-Netto et al. (2014) podemos observar que o tom médio destas unidades pode servir de pistas para diferenciação de falas emotivas, em especial da raiva e tristeza, enquanto a variação de duração intra-unidades pode ser significativa na diferenciação de falas emotivas, quando analisadas intra-gênero.

Xu (2005) sugere que é necessário considerar 3 outros aspectos, relacionados à variação de F_0 . O primeiro deles seria a velocidade de mudança tonal, isto é, o quanto um falante demora para aumentar ou diminuir a resistência da glote, de modo a alterar os valores de *pitch*. Os dados reportados em Xu e Sun (2002) demonstram que para se fazer uma alteração da ordem de um semitom ascendente ou descendente são necessários 100ms (1st/100ms), sendo que o maior tempo está associado à ascensão de um tom. Este é um efeito esperado, uma vez que o aumento da frequência fundamental implica na aplicação de uma resistência maior às pregas vocais (XU; SUN, 2002).

O segundo parâmetro destacado por Xu (2005) é a sincronização entre o tom e a sua unidade portadora. Na análise conduzida por nós em Ferreira-Netto et al. (2012) observamos que no PB a sincronização ocorre principalmente em consoantes sonoras e vogais, de modo que em consoantes surdas ou pausas não notamos valores significativos de F_0 ou então dentro da faixa de intensidade por nós estabelecida. Xu (2005) argumenta que em línguas tonais essa sincronização é mais evidente, pois tem uma importância ainda maior, uma vez que um tom fora de fase pode alterar o significado de um determinado item lexical.

O terceiro parâmetro é o *pitch range* do falante, isto é, o quanto um indivíduo pode aumentar o alcance de sua voz e ainda mantê-la normal. De um ponto de vista fisiológico, Baken e Orlikoff (2000), em uma revisão da literatura, encontraram os valores mínimos de F_0 em torno de 80Hz para vozes masculinas e dos 130Hz para

vozes femininas. Os valores de pico de F_0 têm uma grande variação, pois o avanço da idade causa uma perda natural da tonicidade das cartilagens das pregas vocais, levando necessariamente a um decréscimo dos valores de F_0 .

Sendo assim, para homens com mais de 68 anos, a média gira em torno de 394.2Hz, enquanto que para jovens entre 17 e 25 anos, esse valor é de 763.6 Hz. Os trabalhos analisados por Baken e Orlikoff (2000) demonstravam homens entre 18 e 39 anos encontraram valores de 567.3Hz e para homens entre 35 e 70 anos, a média é da ordem de 260 Hz. Para mulheres com mais de 66 anos, a média é de 570Hz, enquanto estudos que analisaram mulheres entre 18 e 36 anos encontraram médias que variam entre 840Hz e 1100Hz. As figuras 3.2 e 3.3 apresentam os valores integrais:

Table 6-16. Maximal Phonational Frequency Range

Age		No. of Subj	Lowest F_0^*		Highest F_0^*		MPFR*		Sources
Mean	Range		Mean	Range	Mean	Range	Mean	Range	
Men									
20.3	17.9–25.8	157	79.5	62.0–110.0	763.6	292.0–1568.0	864.1		3
			/27.4/	/23.1–33.0/	/66.5/	/55.0–79.0/	/38.8/	/29–54/	3
21.4	18–36	332	80.1	61.7–123.5	674.6	220.0–1567.8	594.5		1
			/27.5/	/23–35/	/64.4/	/45–79/	/37.9/	/13–55/	1
	young adult	14	87	69–110	571	440–698	484		4
			/28.9/	/24.9–33.0/	/61.5/	/57.0–65.0/	/32.4/	/29–36/	4
27.1	23–33	10	84.8	61.7–123.5	752.8	493.8–932.2			7
			/28.5/	/23–35/	/66.3/	/59–70/	/37.8/	/30–45/	7
27.6	18–39	31	77.3		567.3		490		2
			/26.9/		/61.4/		/34.5/		2
76.9	68–89	27	85.3		394.2		308.9		2
			/28.6/		/55.1/		/26.5/		2
56.8†	35.5–75.0	17	80†	40–120	260.0†	190–440			5
			/27.5†/	/15.5–34.5/	/47.9†/	/42.5–57.0/	/22.1†/	/8–36/	5

Figura 3.2: Valores de F_0 mínimo e máximo para vozes masculinas. Retirado de Baken e Orlikoff (2000, p.188)

Women									
21	18-33	44	141.4	93.8-190.1	884.1	334.1-1917.4	743.4	199.6-1816.0	6
			/37.3/	/30.2-45.2/	/85.4/	/52.2-82.5/	/30.3/		6
22.8	18-36	202	140.2	98-196	1121.5	587.3-2092.8	981.3		1
			/37.2/	/31-43/	/73.2/	/62.0-84.0/	/37.0/	/23-50/	1
23.5	18-38	31	134.6		895.3		760.7		1
			/36.5/		/69.3/		/32.8/		2
27.1	23-33	10	127.1	98.0-164.8	1102.2	830.5-1666.1			7
			/35.5/	/31-40/	/72.9/	/68-80/			7
76.9	66-93	36	133.8		570.6		436.8		2
			/36.4/		/61.5/		/25.1/		2

Data have been converted from Hz to ST, or vice versa, as required.
 *Data within slashes (/ /) are in ST, others in Hz.
 †Median

Figura 3.3: Valores de F0 mínimo e máximo para vozes femininas. Retirado de Baken e Orlikoff (2000, p.188)

De um ponto de vista linguístico, os estudos conduzidos por Consoni (2011) e Medeiros e Martins (2014) demonstram que esse alcance não deve ultrapassar um terço de oitava, a partir de um tom de referência. Como citado acima, os valores de +3 e -4 representam os limiões de diferenciação tonal, pontos de mudança, nos quais os falantes podem passar a notar diferenças semânticas em uma sentença. Desta forma, podemos sugerir a existência de dois *pitch range* linguísticos. O primeiro deles seria *operacional*, isto é, uma faixa de frequência, a qual os falantes podem usar sem que necessariamente se transmita informações que se destaquem para os seus ouvintes. O segundo deles seria uma faixa *funcional*, a qual os falantes usam para transmitir novas informações por meio da entoação.

Nossa proposta nessa tese é aproximar estes conceitos do modelo por nós adotado, por isso na seção a seguir iremos apresentar o modelo *Exprosodia* (FERREIRA-NETTO, 2008), buscando demonstrar a relação existente entre estes conceitos e o aspecto teórico sob o qual o modelo foi pensado.

3.2 *Exprosodia, Ferreira-Netto (2008)*

Na seção 3.1 apresentamos algumas definições que nos permitem compreender o fenômeno da entoação como uma rede complexa de elementos. O primeiro deles é a

necessidade de atualização em tempo real da cognição ou percepção com a variação de F_0 , como proposto por t'Hart, Collier e Cohen (1990). A segunda refere-se a importância de compreender a existência de um nível "gestual", no caso os alvos tonais, que podem ser compreendidos como uma unidade que conecta o aspecto simbólico das unidades fonológicas do modelo de Pierrehumbert (1980) com a implementação fonética de F_0 e, por extensão, as funções comunicativas da entoação, descritas no capítulo 2.

Em nosso entender, esse conjunto de definições nos permite compreender a entoação como a organização temporal de um conjunto de alvos tonais, articulados por um processo de controle da variação de F_0 e coordenados por processos cognitivos e linguísticos, de modo que estas variações de F_0 podem transportar consigo informações semânticas, pragmáticas, atitudinais e discursivas. Como apresentaremos no capítulo 4, entendemos que essas variações fazem parte de um sistema interdependente da língua, de modo tal que o falante pode acioná-las conforme suas necessidades comunicativas.

O conceito de organização temporal foi adotado por Ferreira-Netto (2008), que propôs que a entoação seja compreendida como uma série temporal e organizada em componentes estruturadoras, a Sustentação e a Finalização e componentes semântico-funcionais, o Foco/Ênfase e o Acento Lexical. O elemento de organização destas componentes seria o Tom Médio (TM), um valor de referência que representa o esforço contínuo do falante em manter estáveis as variações de F_0 . De um ponto de vista fonológico, podemos compreender o Tom Médio como um alvo tonal, ou dentro de uma reformulação, um alvo entoacional, isto é, um elemento linguístico-cognitivo utilizado pelos falantes para implementar a entoação e suas variações.

As componentes estruturadoras, por sua vez, estão associadas ao esforço realizado pelo falante para sustentar a frequência fundamental dentro deste alvo. Sendo assim, elas seriam recursos inconscientes e corresponderiam às restrições mecânico-fisiológicas propostas por Xu e Wang (1997). Dentro deste modelo, o esforço do falante em manter

uma determinada frequência em sua fala seria paulatinamente suprimido pela natureza da própria declinação frasal, debatida acima.

A manutenção dos valores de F_0 deriva, então, da supressão, seguida de uma declinação pontual e da tentativa de recuperação da tensão inicial. A esse incremento de tensão, o autor deu o nome de Sustentação (S) e a esse movimento de supressão e tensão, o nome de Ritmo Tonal (RT). Já, a partir, da supressão completa desse movimento teríamos a Finalização (F). De acordo com o autor:

A sustentação é consequência do esforço que se acrescenta a cada um dos momentos da fala, incluindo-se o inicial, para compensar a declinação pontual de finalização. Ritmo tonal é consequência da ação dessas duas tendências que atuam em sentidos opostos, possibilitando a produção da fala. A componente F associa-se ao fato de que se trata do tom-alvo da declinação pontual, estabelecida por um intervalo ideal decrescente de 7 st do tom médio (TM) obtido até o momento Z_t (CONSONI; FERREIRA-NETTO, 2008). (FERREIRA-NETTO, 2008, p.2)

O gráfico da figura 3.4 apresenta esquematicamente a hipótese do autor:

Figura 3.4: As setas diagonais representam a declinação desencadeada em cada momento pelo desejo de supressão do esforço, ao passo que as setas verticais representariam a sustentação desencadeada pela tentativa de retomada da tensão inicial, e o produto resultante desse esforço, o ritmo tonal. Retirado de Ferreira-Netto (2008)

O modelo proposto por Ferreira-Netto (2008) opera com um conjunto de alvos. A seguir apresentamos uma definição destes alvos junto de uma modelo matemático de análise por nós criado.

3.2.1 Modelamento

Em Medeiros e Martins (2014) fizemos um modelamento básico da estrutura entoacional sugerida por Ferreira-Netto (2008). Partimos do princípio que entoação pode ser definida como as variações de F_0 em um intervalo de tempo e analisada como a organização dos elementos entoacionais em uma série temporal θ , de modo que os elementos que a compõem são as medições de frequência z ($z_1, z_2, \dots, z_{n-1}, z_n$, com $n - 1 > 0$, para $z_n = F_0(t)$), tomadas nos instantes t ($t_1, t_2, \dots, t_{n-1}, t_n$, onde $n - 1 > 0$). O que propomos aqui é uma revisão, utilizando os conceitos de alvos toansis.

Em nosso entender, podemos tomar como corolário a hipótese de que cada valor z é um ponto de movimento tonal, o qual representa a implementação parcial de um alvo tonal. Deste modo, um conjunto de A de valores z representaria uma unidade entoacional, com uma estrutura fonológica subjacente e uma estrutura fonética de superfície. A equação (3.1) é a forma matemática desta hipótese e retorna um conjunto de variações de frequência (z), dado um intervalo temporal que vá de t a t_n , estabelecendo uma relação de dependência entre os termos z e t :

$$\theta = z(t) = [z_1(t_1), z_2(t_2), \dots, z_n(t_n)] \quad (3.1)$$

Equação 3.1. Entoação como série temporal.

Em termos matemáticos, a equação (3.1) demonstra que o conjunto das observações z seriam alturas tonais intrínsecas aos instantes. Uma das vantagens analíticas em se compreender as curvas de F_0 como uma série temporal é a capacidade de se poder fazer comparações entre alturas em função do tempo. Além disso, as séries temporais preveem um condicionamento entre os instantes, dessa maneira, ao se analisar o instante z_2 , necessariamente, estaríamos analisando o instante z_1 e sua correlação com todos os instantes anteriores a ele.

A vantagem desse tipo de análise é que ele se mostra capaz de incorporar os

dois aspectos que discutimos na seção 3.1: alvos e movimentos tonais. Como dito, a estrutura de um conjunto de elementos z representaria um alvo, ao passo que a análise temporal destes elementos, avaliando as posições passadas para estabelecer o valor presente e estimar um valor futuro, seriam fundamentais para compreender os *pitch movements*. Neste sentido, a avaliação do comportamento dos elementos $(z_1, z_2 \dots z_n)$ incorpora a noção de movimentos tonais e seria um elemento importante para compreendermos os recursos fonéticos empregados pelos falantes para atingir seus alvos tonais e manifestar F_0 .

A proposta de Ferreira-Netto (2008) incorpora também a noção de limiares de diferenciação tonal, discutidos acima na seção 3.1. Os limiares seriam estabelecidos a partir de um tom de referência, no caso, podemos utilizar o Tom Médio como este valor. Desta forma, os elementos acima ou abaixo destes limiares representariam proeminências, sejam positivas ou negativas. A figura 3.5 apresenta esquematicamente esse conceito.

Figura 3.5: Limiares de Diferenciação tonal e os pontos de proeminência. Retirado de Consoni (2011)

Assim, as proeminências também representam alvos tonais a serem atingidos pelos

falantes no decurso de uma elocução, quando estes necessitarem que determinadas porções de fala sejam significativas para seus falantes. Assim, é natural imaginar que os movimentos tonais em torno destes elementos tenham determinadas distinções, as quais são analisáveis dentro de um método mais amplo. No mais, é possível, portanto, afirmar que a entoação é composta basicamente de três grandes alvos: o tom médio, as proeminências e a finalização. Vamos analisar a descrição de cada um deles a seguir.

3.2.2 Tom médio, limiares de diferenciação tonal e finalização

De acordo com o modelo de Ferreira-Netto (2008), durante o processo inicial de fonação, o falante estabeleceria um valor tonal de referência e, durante o curso de fala, tenderia a manter os valores de frequência em torno desta referência. Sendo assim, os valores de F_0 encadeados nesta sequência traduziriam a tendência natural dos falantes em manter uma determinada estabilidade, ou ainda, tentariam manter estes valores numa média aproximada. Para Ferreira-Netto (2008, p.3), o tom médio é definido como "a tendência central dos valores válidos de F_0 , calculados como a média aritmética acumulada no tempo. Os valores válidos mensurados são os momentos de F_0 que cumprem as restrições de altura, intensidade e duração."

A equação (3.3) define a forma matemática do tom médio ao passo que o sistema (3.4) define como seriam estabelecidos os valores dos limiares de diferenciação tonal (LDT's), a partir do tom médio. Todos os valores são medidos em Hz. É importante ressaltar que a equação (3.3) é uma média aritmética instante-a-instante, isto é, para t_3 , por exemplo, temos $t_3 = \frac{t_1+t_2}{3}$, onde a média em t_3 é a soma das médias dos instantes anteriores, no caso, t_1 e t_2 . Assim,

Para a série temporal θ :

$$z(t) = [z_1(t_1), z_2(t_2), \dots, z_n(t_n)] \quad (3.2)$$

Equação 3.2. Análise da entoação como série temporal.

Desta maneira, temos o tom médio (τ) definido pelas médias da série temporal θ :

$$\tau = \frac{\sum_{i=1}^n z_n(t_n)}{n} \quad (3.3)$$

Equação 3.3. Equação que define o tom médio em Hz.

A partir desta equação podemos pressupor o sistema 3.4, onde os índices a e b são os extremos do intervalo de frequência, os LDTs. Em nossos trabalhos usualmente avaliamos estes limiares em semitons, porém, como nesta análise estamos trabalhando com as unidades em Hz, iremos mostrá-las convertidas:

$$p/[a,b] = \begin{cases} a = \tau * 1,19 \\ b = \tau * 0,78 \end{cases} \quad (3.4)$$

Equação 3.4. Sistema que estabelece os limiares de diferenciação tonal. O *espaço entoacional*.

Os valores de a e b representam os LDTs e são obtidos por meio de uma progressão geométrica, cujo fator q é igual à 6%. Este valor refere-se a diferença entre um semitom e outro (MARTINS; FERREIRA-NETTO, 2010). No caso dos valores ascendentes utilizamos um fator de 1,06 para que os valores sejam crescentes de fato. Para os valores descendentes utilizamos o fator q de 0,06, de modo que os valores fossem descendentes a partir do tom médio. Os valores n para o cálculo da progressão geométrica, foram 3 e 4, ou seja, os valores de LDT estabelecidos por Consoni et al. (2009) e Consoni (2011). Desta maneira temos $1,06^3 = 1,191$ para os tons ascendentes e $x_2 = x_1 - (x_1 * 0,06)$ até x_4 para os semitons descendentes.

Para nós, a região delimitada pelos limiares de diferenciação tonal compõe um *espaço entacional*, no qual as variações internas não têm um papel significativo para ouvinte. Dessa maneira, as “oscilações” que ocorrem em seu interior não são perceptíveis, por isso pensamos que estes alvos são *topológicos*, pois ocorrem uma área

delimitada, ao passo que aqueles que excedem esses limiares são alvos *pontuais*, já que estão condicionados a uma determinada altura. Essa definição é fundamental para a nossa análise, por isso vamos retomá-la na seção 3.3 deste capítulo.

O modelo de Ferreira-Netto (2006, 2008) supõe uma terceira componente, além do intervalo lateral e do tom médio, a saber: a finalização. Dentro deste modelo, a finalização representa o fim do esforço do falante em manter os valores de F_0 dentro dos limites do tom médio, estabelecidos pelos limiares, seguido de um decaimento das frequências em direção à declinação frasal. De acordo com Peres, Consoni e Ferreira-Netto (2011) e Rosa (2015), o término de um ato de fala seria notado no instante em que a frequência fundamental chegasse ao valor de -7st de distância do tom médio estabelecido. Seguindo este princípio, seria necessária a inclusão de um terceiro elemento no sistema (3.4), como demonstrado abaixo.

$$p/[a \rightarrow b, c] = \begin{cases} a = \tau * 1,19 & \text{estabelece o LDT superior.} \\ b = \tau * 0,78 & \text{estabelece o LDT inferior.} \\ c = \tau * 0,65 & \text{estabelece o valor de finalização.} \end{cases} \quad (3.5)$$

Equação 3.5. Sistema que demonstra o processamento de percepção da entoação.

O valor c de 0,65 foi estabelecido de maneira semelhante ao limiar de diferenciação tonal inferior. Contudo, a fórmula estabelecida seguia a série até x_7 , de modo que podemos repensar a fórmula da seguinte maneira: $x_{n+1} = x_n - (x_n * 0,06)$ indo de x_1 até x_7 . Por fim, podemos incluir neste sistema, mais dois elementos: o tom médio e a variação natural da frequência fundamental (F_0), definindo, por fim, o sistema $\epsilon(t)$.

Assim para todo $F_0 > 0$ poderíamos executar a seguinte decomposição:

$$= \epsilon(t) \begin{cases} a = \tau * 1,19 & \text{estabelece o LDT superior.} \\ F_0 = z(t) & \text{variação da frequência fundamental.} \\ \tau(t) & \text{tom médio.} \\ b = \tau * 0,78 & \text{estabelece LDT inferior.} \\ c = \tau * 0,65 & \text{estabelece o valor de finalização.} \end{cases} \quad (3.6)$$

Equação 3.6. Sistema que demonstra o processamento global da entoação

A figura 3.6 apresenta uma análise da oração "O Domenico entrar aqui...como assim?", utilizando-se o algoritmo acima descrito. A figura pode ser melhor vista no Anexo I.

Figura 3.6: Exemplo de análise se usando o tom médio para a oração: "O Domenico entrar aqui...como assim?".

Note-se que a linha central marcada com "x" refere-se ao tom médio, ao passo que as linhas tracejadas que o acompanham são os limiares de diferenciação tonal e a linha preta é a variação de F_0 . Como se vê, em geral, o falante tende a manter os valores de F_0 próximo ao tom médio, tratado por nós como um alvo tonal, por se tratar de uma elocução curta, é natural que essas duas curvas coincidam. Em outro ponto, é notável que o falante passa a alterar o seu movimento tonal se distanciando do tom médio, justamente a partir de quando ele precisa atingir um novo alvo: a proeminência, a qual irá marcar a interrogação.

Assim, observa-se que a curva de F_0 , somente irá ultrapassar as margens -3 e $+4$ ao fim da oração, quando há um aumento paulatino da frequência fundamental com a

intenção de marcar que trata-se de uma interrogativa. Os valores de frequência, no caso analisado, aparecem no eixo das coordenadas e foram tomados as médias de frequência de cada sílaba, sendo estes valores em seguida convertidos para semitons.

3.3 Discussão: alvos topológicos, alvos pontuais

A partir do exposto acima, entendemos que os alvos tonais se organizam em dois grandes grupos. O primeiro deles é o que chamamos de *alvos topológicos* e o segundo são os *alvos pontuais*. Adotamos o nome *topológico*, pois no início desta pesquisa acreditávamos ser possível analisar estas variações como mudanças em um espaço e não apenas variações. Para isso, utilizaríamos justamente os conceitos de Topologia Matemática. No entanto, o recurso se mostrou dispendioso e excessivamente impreciso.

Ainda assim, para nós, esses alvos são topológicos, pois coordenam a implementação de F_0 em um espaço delimitado pelos limiares de diferenciação tonal. Assim, o fator que nos interessa mais de perto é a tendência que os falantes demonstram de manter as variações de F_0 sempre no interior deste espaço entoacional. Hipoteticamente, acreditamos que isso se dá, pois o falante sabe de forma intuitiva que ao ultrapassar esses limiares ele irá modalizar a entoação e passar a transmitir significados.

Portanto, os alvos topológicos seriam uma espécie de “configuração” tonal, que transporta características particulares da elocução e do estado do sujeito, por exemplo. Neste caso, um falante com raiva pode mobilizar toda a sua curva entoacional para um registro tonal mais alto, assim, o seu espaço entoacional terá uma nova configuração, determinada pelas de referências que os alvos topológicos fornecem no momento de implementar a frequência fundamental. De maneira geral, entendemos que os alvos topológicos são esta referência, ou seja, o falante intuitivamente sabe qual altura precisará atingir para implementar suas variação, a partir destes alvos.

Por outro lado, os alvos pontuais são aqueles que necessariamente precisam

ultrapassar esses limiares. Por este motivo, eles se caracterizam pela tentativa do falante de crescer ou decrescer os valores de frequência fundamental, com o intuito de transmitir um conteúdo informacional ao seu ouvinte. A nosso ver, essa variação não é natural e implica em um esforço atípico do falante, de modo que ela necessariamente será pontual. Seja porque ela ocorre em pontos específicos de uma elocução, seja porque ele é um evento significativo dentro da elocução como um todo. A figura 3.7 representa essa relação.

Figura 3.7: Esquema que representa a relação entre os alvos topológicos e alvos pontuais

A partir de uma análise mais profunda, entendemos que o aspecto fonológico da entoação reside justamente na alternância entre alvos topológicos e alvos pontuais. A figura 3.7 demonstra que um falante mantém os valores de F_0 dentro do espaço entoacional, a partir de um decréscimo ou acréscimo desses valores para se atingir um determinado valor pontual, o qual tem alguma relevância semântica, como é o caso da proeminência tonal (CONSONI, 2011). A partir disso, o falante pode manter os valores de F_0 neste novo patamar, estabelecendo um novo alvo topológico e um novo espaço entoacional ou pode retornar aos valores que estava.

Os dois movimentos descritos carregam consigo uma valência, ou seja, um valor intrínseco a eles e perceptível pelo ouvinte. No caso do primeiro, temos chamado essa valência de *distensão*, ao passo que no segundo caso chamamos de *tensão*, utilizamos esse nomenclatura pensando nas questões associada à duração. Na *distensão*, a duração

é estendida, uma vez que o novo tom terá de ser sustentado ao longo do tempo, ou seja, o efeito é permanente, por isso não é concentrado. Já na *tensão* essa duração precisa, necessariamente, ser breve para que se marque sua posição de relevo, o efeito é duracional e concentrado. A figura 3.8 demonstra a relação entre as valências tensão e distensão.

Figura 3.8: Esquema que representa a relação as valências e a transmissão de significado.

De toda forma, a ruptura causada pelo alvo pontual gera, necessariamente, algum tipo de significado para o ouvinte. Como se verá no capítulo 5 a implementação da frequência fundamental dentro do espaço entoacional, segue um padrão relativamente estável, ao passo que a implementação do alvo pontual não segue o mesmo padrão, sendo que a valência parece ser o fato determinante para essa diferença. de modo que encontramos os mesmos resultados para a fala emotiva, discutida no Capítulo 7.

Um dos desafios do ponto de vista teórico, é explicar como o falante pode implementar as variações de frequência fundamental, sem que se afete diretamente o nível segmental. Nossa hipótese é a de que a fala pode ser decomposta em dois sistemas: um contínuo e outro descontínuo. A nosso ver, essa diferenciação seria capaz de explicar a implementação fonética da contraparte física do alvo. No capítulo a seguir

apresentaremos nossa hipótese buscando relacioná-las com os problemas relativos à fala emotiva.

4 Fala

Neste capítulo iremos traçar um paralelo entre o modelo comunicativo de Shannon (1948) e as teorias linguísticas clássicas. Nossa hipótese, de um ponto-de-vista epistemológico, é a de que a linguística, como ciência, entendeu que questões relacionadas a fenômenos prosódicos, principalmente os entoacionais, ou exercem um papel fonológico-funcional, limitado a regras de implementação fonológica (PIKE, 1945; BOLINGER, 1958; MAEDA, 1976; GOLDSMITH, 1976; PIERREHUMBERT, 1980) ou têm função estilística, atribuindo significados complementares (TRAGER, 1958, 1961; FONÀGY, 1991).

Em nosso entender, estas duas funções estão diretamente correlacionadas e representam um único fenômeno. As regras de implementação fazem parte do domínio sistêmico da língua, ao passo que os mecanismos de expressão tem o papel de ajuste destas regras com uma função comunicativa. Dentro de nossa hipótese, o alvo entoacional surge justamente desta relação. De um lado temos estruturas fixas, que têm funções linguísticas, por isso são alvos bem estabelecidos e produtivos em uma língua, de outro estruturas variáveis, os alvos que estabelecem as funções comunicativas da fala, de modo que a implementação de F_0 é orientada por estes alvos.

Para podermos tratar da entoação como esse aspecto variável, temos de considerá-la como um sistema a parte dentro da produção da fala. Nossa sugestão é que a fala seja compreendida como o acoplamento de dois sistemas: um contínuo e um descontínuo. O primeiro seria responsável por implementar as variações de F_0 , ao passo que o segundo seria responsável por implementar os segmentos que compõem

fala.

4.1 Fala, comunicação

4.1.1 Fala

Desde os primeiros estudos linguísticos no início do século XX, assumiu-se que os menores elementos de uma língua seriam os seus segmentos fonéticos ou seus elementos constituintes (SAUSSURE, 2012), os traços do nível merismático (HJELMSLEV, 2006; MARTINET, 1975; BENVENISTE, 1976). Na abordagem gerativista, estes receberam o nome de traços distintivos (CHOMSKY; HALLE, 1968), sendo que proposta semelhante já havia aparecido na análise seminal de Jakobson, Fant e Halle (1952).

Em geral, assume-se que estes elementos mínimos distinguíveis se organizam em unidades maiores, no caso os fones, por meio de regras de combinação e de restrição (TROUBETZKOY, 1976; CHOMSKY; HALLE, 1968) e deste ponto para unidades maiores (CHOMSKY, 1965). Para Chomsky e Halle (1968) estas combinações sempre são tridimensionais, pois operam sobre três variáveis: o ponto de articulação, o modo de articulação e a sonoridade. Obviamente, estas regras são condicionadas pela (i) forma como os sons são produzidos, desta maneira, uma nasal glotal parece ser improvável; ou (ii) pelo próprio sistema da língua, sendo que em tupi a sonoridade em plosivas bilábias não é necessariamente um traço distintivo.

Existem, portanto, dois universais possíveis, o primeiro, é o conjunto dos elementos mínimos, os traços distintivos, enquanto o segundo seria o das regras de combinação, cuja restrição é indefinida, embora pareça seguir um padrão interlínguas quando se trata de gerar significados (CHOMSKY, 1965; CHOMSKY; HALLE, 1968). Os produtos finais, as línguas ou sistemas de linguagem construídos (ECO, 2001), seriam distintos por natureza, uma vez que cada um deles pode selecionar regras e elementos específicos (CHOMSKY, 1959).

Deste modo, tanto estruturalistas, quanto gerativistas entendem que o conjunto infinito de frases de uma língua são gerados a partir de dois princípios universais: seleção e combinação. Embora estes entendam a língua como um fenômeno mental e aqueles como um fenômeno social, ambos concordam que o conjunto de frases produzidas pela língua e colocadas em uso pertenceriam a um novo domínio, o do falante. Saussure (2012) é categórico ao definir a separação entre estas dimensões:

"O estudo da linguagem comporta, portanto, duas partes: uma, essencial, tem por objeto a língua, que é social em sua essência e independente do indivíduo - esse estudo é unicamente psíquico; outra, secundária, tem por objeto a parte individual da linguagem, vale dizer, a fala, inclusive a fonação - é psicofísica."(SAUSSURE, 2012, p.51).

Neste sentido, para Saussure (2012), pequenos ajustes comunicativos fazem parte de escolhas individuais e apenas mudanças significativas teriam valor linguístico. Por exemplo, o aumento de F_0 para atribuir foco em um determinado item lexical seria uma idiossincrasia da comunicação e apenas a "cristalização" desta regra em um determinado item poderia colocá-lo no arcabouço de fatos linguísticos (SAUSSURE, 2012). Sendo assim, as frases 1 e 2 :

- (1) O ↑PEDro mora na casa azul.
- (2) O Pedro mora na casa ↑AZUL.

Seriam elocuições idiossincráticas que não representam fenômenos linguísticos em si, apenas a variação de um objeto da língua. A utilização do foco seria um recurso próprio do falante. Para ele, "é a fala que faz evoluir a língua: são as impressões recebidas ao ouvir os outros que modificam nossos hábitos linguísticos"(SAUSSURE, 2012, p.51). Neste sentido, o foco só poderia ser considerado um fato linguístico se houvesse alguma regularidade na estrutura, como se supostamente o foco sempre recaísse no primeiro sintagma nominal de uma frase, como nos exemplos (3) e (4):

- (3) O ↑PEDro mora na casa azul.
- (4) ↑AZUL é a casa do Pedro.

Chomsky (1965) segue o mesmo princípio criando a distinção *competência/ performance* depois atualizada para *I-language/ E-Language* (CHOMSKY, 1986). Para o linguista norte-americano:

"Todavia é igualmente claro que o uso real observado da linguagem - o *desempenho* real - não reflete apenas as ligações intrínsecas de som-significado estabelecidas pelo sistema de regras linguísticas. O desempenho envolve também muitos outros fatores. (...) O desempenho linguístico é, ademais, governado por princípios da estrutura cognitiva (por exemplo, por restrições de memória) que não são, propriamente falando, aspectos da linguagem". (CHOMSKY, 2009, p.194)

O conceito suficientemente importante aqui refere-se a ideia de que a performance é governada por estruturas cognitivas mais amplas do que somente a linguagem e envolve regras complexas, as quais estão apoiadas na audiência e no contexto, de modo tal que estes não são problemas da linguística em si, pois são elementos observados de forma extensiva à língua e não são parte da linguagem por definição (CHOMSKY, 2009). O problema da variação de F_0 para demarcação do foco pode ser compreendido como um processo acessório a língua, isto é, uma saída linguística governada por outros processos cognitivos, como a percepção.

Em nosso entender, o principal desafio destas perspectivas é compreender que as mudanças que podem ocorrer na fala e gerar um novo significado associado à mensagem são construídas pela própria língua. É importante destacar que mudança aqui não se refere a variação linguística dentro de uma perspectiva Laboviana, isto é aquela que altera regras ou parte delas (LABOV, 1965). O tipo de mudança que nos referimos trata da possibilidade de que a fala sofra mudanças pontuais de acordo com as necessidades comunicativas dos falantes sem que essa nova estrutura se transforme, necessariamente, em uma regra linguística, mas mantenha um *status* de regra de comunicação, uma vez que ela tem um papel semântico ao transmitir significados.

4.1.2 Comunicação

McLuhan (2007) sugere que a fala, como realização fonética, é, por definição, um meio frio e de baixa definição, isto é, pouquíssima informação é transmitida de fato por seus componentes linguísticos, de tal modo que parte da mensagem precisa ser preenchida pelo ouvinte. Neste sentido, é natural assumir que existam mecanismos comunicativos de compensação das variáveis acústicas, os quais compensariam a baixa definição da fala durante o instante de sua produção. Dentro desta perspectiva, Barthes (2012) defendia a existência de um terceiro plano, o qual ele chama de "plano do uso". Para ele, este nível seria um "pré-significante", o qual tem como função principal oferecer suporte para a transmissão do conteúdo:

"...e para completar o par *Língua/Fala* por um terceiro elemento, pré-significante, matéria ou substância e que seria o suporte (necessário) da significação: numa expressão como 'um vestido comprido ou curto' o 'vestido' não é senão o suporte de um variante (*comprido/curto*), que pertence plenamente à língua indumentária." (BARTHES, 2012, p.42)

Entendemos que estes pré-significantes sejam estruturas concebidas pelo uso ou pela língua em si, tal como vemos a questão do alvo entoacional. Sendo assim, a variação de F_0 para indicar um foco linguístico pode ser compreendida como uma destas estruturas, uma vez que ela poderia se associar ao signo e alterar o significado, ainda que a estrutura segmental do significante não tenha sofrido alterações. Estes suportes semânticos podem ter uma variedade de interpretação em sistemas semiológicos (objeto principal de Barthes), no caso da fala eles parecem se restringir as certas propriedades, como a prosódia e a qualidade vocal.

Jakobson (2008) em seu ensaio "Linguística e Poética", bem como Fonagy (1991) e Fonagy (2001) observavam este problema de uma perspectiva do falante. Para estes autores, é possível que o falante manipule a linguagem, fazendo ajustes de expressão para a transmissão de um significado complementar, sem que necessariamente se altere um conjunto de regras ou elementos mínimos de uma língua. Esses ajustes seguiriam regras próprias, de modo que não podem ser considerados simplesmente aleatórias ou

meras idiossincrasias do falante, mas também não fazem parte da língua e, sim, de uma segunda estrutura.

Para Fonàgy (1991) essas regras fazem parte de um código próprio, o qual ele denomina *para-código*, que seria um conjunto de características de variações que se associaria ao código linguístico em um segundo nível durante a produção da fala. Assim, essas variações não fazem parte da fonêmica de uma língua em si, mas complementam os sentidos dela.

Il faudrait admettre, par conséquent, deux actes successifs d'encodage: un encodage linguistique qui transforme un message global, une idée, en une sequence de phonèmes, et un deuxième codage - qui coïncide admirablement avec l'acte de mise en sons des phonèmes - au cours duquel le message *secondaire*, gestuel, est greffé su le message *primaire*". (FONÀGY, 1991, p.14)

A figura 4.1 ilustra esse processo. A estrutura fonética de um determinado item lexical seria construída pelo mecanismo denominado *Cod.1*, ao passo o foco seria associado a este item lexical por meio do *Cod.2*, alterando a entoação para transmitir um conteúdo informacional adicional, como o fato de que a casa do Pedro é azul e não verde, compreendidos pelas estruturas *Dec.1* e *Dec.2*. (FONÀGY, 1991) entende que o fenômeno de decodificação das mensagens no *Dec.2* estão em outro nível mental, de modo que estas estruturas não têm conteúdos semânticos por si só, de modo tal que apenas percebemos os seus efeitos.

Figura 4.1: Esquema de comunicação de Fonagy (1991). O código linguístico sofre uma segunda codificação, o paralinguístico que também seria natural.

Problema semelhante já aparecia nas teorias da comunicação, como a de Shannon (1948). Para o autor, o fenômeno da linguagem opera em dois eixos principais, o sintagmático e o paradigmático. No entanto, a simples seleção dos itens lexicais coordenada por um processo de caráter quase Markoviano não explica a totalidade de fenômenos que ocorrem no ambiente comunicativo, que precisa incorporar o contexto e o seu receptor em seu modelo teórico. No entanto, regras de restrições não permite a adição de um código em um segundo processo, como sugere Fonagy (1991), pois isso fere o princípio binário no qual foi construído o modelo de Shannon (1948) ¹. Desta forma, a segunda estrutura desenvolvida por Fonagy (1991) precisaria estar "dissolvida" entre os componentes da teoria da comunicação.

Em *The Mathematical Theory of Communication*, Shannon e Weaver (1964) definem que a comunicação, seja ela linguística ou não, opera sempre em 3 níveis. O primeiro deles, o nível simbólico, está diretamente relacionado à transmissão dos sinais de

¹O modelo de Shannon (1948) era baseado no pressuposto que seria mais tarde aplicado à Máquina de Turing (TURING, 1950), também conhecido como Modelo Transformacional. Dentro deste tipo de computação, os valores computados sempre são binários, ou seja, o processamento precisa incorporar a maior quantidade de informação em uma única unidade. Atualmente, sabe-se que a computação pode operar em múltiplos níveis, graças aos avanços da computação digital e das linguagens de programação. No entanto, naquele período, em que tudo era analógico e feito por meio de *protoboards*, esse tema ainda era incerto, tal como é hoje a questão da computação quântica, a qual é capaz de operar com um condicional "talvez", por exemplo, a despeito das máquinas atuais que operam apenas com "sim" e "não".

comunicação e as condições de produção e recepção, seja o contexto, seja a qualidade do código. Por exemplo, um falante monolíngue de russo tentando iniciar uma conversa com um falante monolíngue de japonês em uma rua de Nova York, feriria os dois princípios: o do código e o do contexto. No caso de falantes de línguas irmãs, como Espanhol e Português, apenas a qualidade do código seria afetada.

O segundo nível está associado à semântica, e diz respeito ao quão preciso o significado de uma determinada informação foi construído e será interpretado pelo receptor. Em outras palavras, esse nível lida com o grau de disjunção dos conjuntos simbólicos de cada falante e o grau de interpretação destes. No primeiro nível, acima descrito, os elementos dos conjuntos simbólicos são discretos, neste segundo nível, eles continuam limitados pelos seus significados, mas estão em um eixo temporal de execução lógica, como uma espécie de sintaxe que reflete na semântica. Em PB, por exemplo, um verbo transitivo indireto supõe uma preposição como elemento subsequente, do contrário o efeito de sentido pode se perder. O terceiro e último nível refere-se ao quão efetivo foi o efeito da informação, isto é, o quanto o receptor alterou o seu estado (qualquer que seja ele) após ter recebido uma determinada mensagem (SHANNON; WEAVER, 1964, p.5 e ss).

Em nosso entender, é possível dizer que a estrutura da linguagem, somada aos seus signos, está no primeiro nível de Shannon e Weaver (1964), elas seriam os requisitos mínimos para que ocorra qualquer comunicação linguística. Dentro desta perspectiva, as estruturas linguísticas e os signos, por serem elementos discretos de um conjunto finito, somente podem ser alterados pelos mecanismos responsáveis pela mudança linguística, tais como pressões sociolinguísticas (LABOV, 1965, p.110-112) ou vicissitudes da própria língua (CHOMSKY; HALLE, 1968, p.187-235). Entendemos que não é do escopo deste trabalho discutirmos estes aspectos, uma maior explicação pode ser encontrada em McMahon (1994), Lass (1997), Labov (2010, 2011).

O segundo nível seria o local onde as trocas ocorrem e as estruturas linguísticas do

primeiro nível têm seus significados atualizados, no entanto ainda estamos no interior do código e não diante de um segundo código, como sugere Fonàgy (1991). Deste modo, a quantidade de informação transmitida por uma língua está associada aos limites e itens existentes no código, bem como a sua estrutura informacional. Sendo assim, um processo como a elevação do tom em uma sílaba pode não ter um efeito lexical no Português Brasileiro, enquanto em uma língua como Chinês é bastante provável que ele o tenha.

Dentro deste modelo comunicativo, a elevação do *pitch*, por exemplo, está inserida no código de uma língua de qualquer maneira, sendo que a utilização dela pode ter finalidades diferentes entre línguas, ou seja, não se trata de um efeito secundário ou uma idiossincrasia do falante, mas antes um recurso linguístico. Desta maneira, é mais simples imaginar que os elementos fonéticos são universais e pertencem a um único código, sendo que cada língua selecionaria um conjunto para exercer a função de transmissão dos significados, sejam eles fonológicos ou expressivos. Os primeiros seriam intrínsecos à língua, ao passo que o segundo seria dependente do contexto e dos atores da conversação.

Este segundo aspecto tem uma importância significativa para nossa tese. Para Saussure (2012) o contexto é contíguo às operações que definiriam uma sentença, ou seja, a seleção paradigmática e a combinação sintagmática sempre têm o contexto como referência, de modo tal que o significado é dado *a priori*, junto da própria estrutura linguística. Em frases como:

- (5) a. Pedro é um leão.
- i. Existe um leão que se chama Pedro.
- ii. Pedro parece-se, por um algum motivo, com um leão.

Temos a dupla articulação dos eixos saussurianos, onde as operações de seleção paradigmática e combinação sintagmáticas foram satisfeitas. Contudo, nota-se que duas mensagens co-ocorrem, uma vez que o item lexical <leão> pode ter duplo significado.

Ainda que se discuta o grau de contiguidade existente entre estes dois significados, é possível dizer que a interpretação final não depende unicamente da estrutura linguística em si ou das operações dela derivada, mas, também, dos participantes envolvidos no cenário comunicativo, como sugerem Shannon e Weaver (1964).

Em “Dois aspectos da linguagem e dois tipos de afasia” Jakobson (2008) vislumbra essa questão. Para o autor, além da seleção, combinação e contiguidade é necessário levar em consideração mais uma operação, a de concorrência. Uma vez que "os constituintes de qualquer mensagem estão necessariamente ligados ao código [estrutura] por uma relação interna e à mensagem por uma relação externa" e ainda (JAKOBSON, 2008, p.51) "deve haver certa equivalência entre os símbolos utilizados pelo remetente e os que o destinatário conhece e interpreta"(op.cit).

Como se pode ver, a linguística sempre fez uma distinção entre aqueles elementos que são universais e tidos como parte da língua, daqueles que são locais e vistos como idiossincrasias dos falantes. A teoria da comunicação e as teorias semiológicas conseguiram incorporar de maneira mais clara este segundo elemento, entendendo que eles não são fatos aleatórios e que têm uma certa regularidade, não suficiente para ser um fato, mas o suficiente para serem analisados quando entendidos como ajustes semânticos utilizados pelos falantes para gerar significados.

Não pretendemos colocar em evidência a hipótese de que os sentidos são construídos a partir de uma rede de relações que envolvem a língua, o contexto e as condições de produção. Nossa sugestão é a de que a fala utilizaria um único processo para a produção dos sons, entretanto, este processo utiliza dois "recursos", por assim dizer, que são contíguos. O primeiro deles pertence à própria língua e está relacionado às operações fonético-fonológicas que dão conta de criar estruturas dotadas de significados, os quais já estão estabelecidos pela própria língua, como é o caso da finalização interrogativa (MORAES; COLAMARCO, 2007) ou do foco (CONSONI, 2011).

O segundo recurso refere-se aos mecanismos de ajuste da produção, que mani-

pulariam a estrutura gerada pelo falante, criando novos significados para o ouvinte (SCHERER, 2003). No entanto, a nosso ver, esses processos são sincrônicos, isto é, não se trata de um efeito de um código sob outro, mas, sim, a realização de uma estrutura expressiva/comunicativa que se relaciona com a a estrutura que é capaz de gerar a linguagem. Desta forma, uma determinada elocução contém os dois planos em um único.

A nosso ver, os alvos entoacionais ajudam a compreender como a entoação pode ser utilizada tanto no domínio fonológico, quanto no domínio expressivo/comunicativo da linguagem. No primeiro caso, o alvo representaria algumas regras de implementação, como a finalização e o foco, acima citados, que podem ser entendidos como elementos fonológicos, uma vez que seguem determinadas regras linguísticas (LADD, 1996). No segundo caso, os aspectos da implementação estão amplamente conectados aos alvos, os quais ajudariam modular a forma como uma determinada curva será implementada, respeitando as regras fonológicas. Sendo assim, as variações tonais de cunho expressivo, podem exercer uma função linguística, isto é, não são meras modalizações de formas neutras, mas antes, estruturas linguística concebidas a partir de uma finalidade expressiva.

Por exemplo, uma fala com raiva teria um *set* particular de alvos entoacionais, o qual não se confunde com o *set* de uma fala triste, sendo essa a função comunicativa a ser estabelecida pela entoação. Dessa forma, aqueles alvos que desempenham uma função linguística, como a demarcação de uma finalização sofreriam uma alteração de acordo com o tipo função comunicativa que essa elocução estabelece. Ao se ampliar a largura de banda de F_0 na raiva, o efeito seria o de se alterar como os alvos fonológicos-linguísticos se comportariam.

A questão que propomos é que as variações determinadas por necessidades expressivas ou comunicativas não são idiosincrasias do falante, nem tampouco variações de um modelo universal, como tem sido tratado pela Linguística, mas, antes, um

fenômeno regular que tem uma interface comunicação-linguagem, como proposto por Jakobson (2008) e Shannon (1948). Para que possamos desenvolver essa hipótese de forma mais clara, precisamos supor que a entoação pode ser implementada de acordo com a necessidade expressiva do falante, junto das regras de formação linguística supostas por sua língua. Por este motivo, iremos tratar a entoação como um sistema particular da fala. Na seção seguinte, iremos apresentar o conceito de sistema por nós adotado e como essa definição se dá dentro da linguística.

4.2 Sistema, fala

Neste ponto é preciso fazermos um aparte e introduzir o conceito de sistema, ao menos como o compreendemos. A Teoria Geral dos Sistemas (BERTALANFFY, 1975; MACIEL, 1974) foi influenciada de maneira mais direta pelo modelo transformacional de Turing (TURING, 1950) e pelo *Principia Mathematica* de Whitehead e Russell (2005), tendo seus mais importantes desdobramentos no campo da ciência da informação com os trabalhos de Claude Shannon (SHANNON, 1940, 1948) e no campo da computação com o trabalho de Neumann (1993).

Noam Chomsky é citado em Maciel (1974, p. 14) como um dos nomes mais influentes desta teoria, graças aos trabalhos feitos na área da Teoria dos Autômatos e da Gramática Formal, principalmente pelos estudos sobre as Hierarquias de Chomsky (CHOMSKY, 1956, 1959), que explicam a geração de linguagens formais que se complexificam de acordo com o grau de liberdade relativa em cada uma das hierarquia. *Sound Patterns of English* (CHOMSKY; HALLE, 1968) é amplamente baseado neste tipo de análise.

Estes trabalhos têm em comum a hipótese do *input/output* transformacional. Neste modelo, um *input* seria, necessariamente, processado por um centro pressuposto e enviado na forma de um sinal (químico, elétrico, simbólico, etc.) que dispararia um *feedback* em forma de um *output* interrompendo o sistema ou ativando outros módulos

a fim de se continuar o processo global. Desta forma, ao final teríamos duas classes distintas de componentes: periféricos e centrais. Módulos periféricos teriam importante papel na execução de movimentos e na percepção de elementos do mundo real, ao passo que o centro de processamento desenvolveria o papel de mediador de processos.

Desta forma, um conjunto κ de elementos é processado por uma regra σ e retornam um produto Φ :

$$\Phi = (\kappa, \sigma) \quad (4.1)$$

Na representação (4.1) o produto Φ é sempre dependente dos elementos de entrada (κ) e das regras estabelecidas pelo sistema (σ). Funções matemáticas, como $f(x) = ax + b$ são definições deste tipo de sistema, onde os elementos de entrada (*input*), chamado de domínio da função, são alterados de acordo com a regra/lei de composição estabelecida, a função em si, gerando um produto (*output*), a imagem desta função. Nesta linha, as regras de restrição, que formam as leis de composição são baseadas em definições ou condições iniciais. Por exemplo, tomemos a função abaixo como uma regra capaz de gerar n estados em $\mathbb{R}$:

$$f(x) = \frac{x}{0}$$

Nesta fórmula sabemos que não existe um resultado Φ possível, pois a divisão por zero não existe dentro do conjunto de números reais (em $\mathbb{R}$, $\frac{x}{0} = \nexists$), sendo esta uma condição inicial para todo e qualquer sistema que se baseia na álgebra, por exemplo. Por outro lado, em uma função como:

$$\lim_{x \rightarrow 0} \frac{1}{x} = \infty$$

Neste caso em específico, podemos afirmar que existem infinitas respostas Φ , pois a

função de $\frac{x}{0}$ continua a existir enquanto houver algum número maior que 0, por menor que ele seja. Sendo que a função limite estabelece que esta é a condição inicial deste sistema. Essas equações demonstram que a regra de restrição é dependente de dois fatores: (i) de uma definição dada *a priori*, como é o caso da divisão por zero ou (ii) de uma regra que constitui o sistema, como no caso do limite, onde a função define que se deve dividir até o infinito antes de zero.

Como discutido na seção 4.1 para a Linguística do século XX, os objetos linguísticos se originam basicamente por meio de processos iniciais: a seleção e a combinação. Desta forma, os problemas relacionados à Fonética e Fonologia se resolvem de maneira perfeitamente adequada com esta definição parcial de sistema, onde objetos A se tornam objetos B, por meio de uma regra σ , a qual possui mecanismos de restrição associados a ela.

No entanto, de um ponto de vista epistemológico, esse parece ser um problema mais amplo. Saussure (2012), por exemplo, utiliza o termo *sistema* no sentido de uma “organização” dinâmica de elementos estáticos. Neste caso, os fones poderiam ser vistos como os elementos estáticos de um conjunto, enquanto a fonologia seria o sistema em si. Para os gerativistas, a ideia parece ser qualitativamente a mesma, sendo que a hipótese de traços e regras é a base para a composição deste sistema.

A diferença, embora tênue, sugere que para os estruturalistas a língua é composta de diversos sistemas interdependentes, ou seja, um sistema necessita do outro, embora isso não seja uma condição de existência (MARTINET, 1975). Já para os gerativistas a língua é um sistema único, formado por níveis de processamento dependentes, ou seja, o processamento da informação morfológica depende da informação fonológica, etc. (CHOMSKY, 2009).

A nosso ver, nesta definição o sistema é, na realidade, o conjunto de elementos mínimos e a relação estabelecida internamente entre eles. Por este motivo, essa definição não dá conta de tratar da relação estabelecida entre o conjunto inicial, a

regra de alteração e o conjunto derivado. Trata-se da definição de uma parte pelo todo. Nesta perspectiva, quando olhamos para os problemas relacionados com a comunicação, que estão fora do conjunto inicial, é natural crer que não existe outra alternativa senão assumir a hipótese de Fonàgy (1991) e compreender que as alterações por ele chamadas de *estilísticas* (FONÀGY, 1991, p.121) são, de fato, processadas em uma segunda etapa, para além do processamento linguístico.

Isso se dá, pois parte dos estudos linguísticos tende a ver seus objetos como elementos estanques, que sofrem alterações modais, ou seja, a natureza dos objetos é modificada a partir destes processos para novos objetos, ainda que a representação mental deles se mantenha a mesma, ou seja, estamos pensando em um sistema independente, com uma série de relações internas e retroalimentado por estas relações continuamente, sendo que a interferência externa é limitada pela quantidade de uso (SAUSSURE, 2012; LABOV, 1965). Um fenômeno fonético do Português Brasileiro, como a palatalização da oclusivas alveolares diante da vogal alta anterior pode ser usado como exemplo deste tipo alteração modal:

S_1	S_2	Restrição
-------	-------	-----------

$$(6) \quad \begin{bmatrix} +\text{alveolar} \\ +\text{oclusiva} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{alveolar} \\ +\text{oclusiva} \\ +\text{palatal} \end{bmatrix} / - \begin{bmatrix} +\text{frontal} \\ +\text{alta} \end{bmatrix}$$

O estado S_2 representa um estado modal de um objeto simbólico S_1 , de modo que a regra deste sistema define que este elemento deve se tornar S_2 apenas e somente apenas diante do contexto apresentado, dentro desta relação. Para nós, está é uma definição funcional de sistemas, como aquela que pode ser encontrada em Maciel (1974), que os define como "um conjunto de elementos quaisquer ligados entre si por cadeias de relações de modo a constituir um todo organizado"(MACIEL, 1974, p. 13), sendo assim, só fazem do código linguístico os elementos S_1 .

No entanto, para se compreender fenômenos complexos, como a fala emotiva, que

incorpora diversas substâncias, sejam elas cognitivas ou naturais, temos de assumir a hipótese que a mudança "estilística" não é um mero recurso acessório e, sim, estabelecida na concepção da frase, ou seja, os elementos s_2 fazem parte do código linguístico de alguma forma. Como aponta Ashby (1957), é fundamental que se compreenda que a condição de existência de um sistema perpassa a noção de *mudança* de modo que "the most fundamental concept in cybernetics is that of "difference", either that two things are recognisably different or that one thing has changed with time"(ASHBY, 1957, p. 9).

Para Ashby (1957) os elementos que constituem um sistema precisam, necessariamente, ter diferenças entre em si e levar em consideração a natureza evolutiva dele próprio, ainda que essa evolução se dê a partir da relação dele com outros sistemas. A definição que enquadra elementos estáticos em uma estrutura não é suficiente, pois além dele ser um todo organizado em uma série de relações internas baseadas em diferenças, ele precisa internamente evoluir para n estados ($S_1, S_2, S_3, \dots, S_n$).

Deste modo, os elementos que compõem um sistema não podem ser continuamente estáticos, do contrário temos apenas uma constelação de objetos em rotação. A Fonologia Gerativa (CHOMSKY; HALLE, 1968) se debruçou amplamente sobre essa ideia e considerou que os sistemas fonológicos podem ser compreendidos como uma coleção de fatos simbólicos estáticos, os traços fonológicos, restritos a poucos graus de liberdade, principalmente por conta das regras fonológicas e das operações por elas permitidas, as quais tendem a ser universais. Desta maneira, estes traços seriam os objetos iniciais da implementação fonética, esta sim regulada por mecanismos e regras com diversos graus de liberdade, por isso pertencentes a outros domínios.

No campo da entoação, em específico, destacam-se os trabalhos de J. Pierrehumbert (PIERREHUMBERT, 1980, 2000) e o de Nespor e Vogel (NESPOR; VOGEL, 2007). Os trabalhos das autoras apoiaram-se na ideia de "restrições" tonais, quando passaram a analisar as possíveis ocorrência de determinados tons dentro de uma cadeia sintática

estabelecida. A noção de *frase entoacional* (I) em Nespor e Vogel (2007) representa esta visão, uma vez que as autoras a estabelecem como:

The formulation of the basic *I* formation rule is base on the notions that the intonational phrase is the domain of an intonation contour and that the ends of intonational phrases coincide with the positions in which pauses may be introduced in a sentence. (NESPOR; VOGEL, 2007, p. 188)

Nota-se que a definição da frase entoacional (I) se dá no limite em que se tem as pausas como suas limitadoras e, por outro lado, se tenha um mínimo de frases fonológicas. O modelo de Nespor e Vogel baseia-se novamente em restrições com regras mais permissíveis. A primeira delas, a sílaba (σ) possui restrições em diversos níveis, desde o combinatório ao do seu alfabeto de formação (quais elementos podem ocupar as posições de ataque, núcleo e coda). A frase entoacional (I) possui restrições quanto a presença de Frases Fonológicas (ϕ), ao menos uma forte e uma fraca. Por outro lado, a variedade de produções de frases entoacionais é bastante grande e relacionada a tantos outros fatores, como a atitude do falante, de modo que se torna difícil conceber restrições neste nível prosódico.

O que estas abordagens trazem é uma série de análises de restrições mecanicistas de produção, baseadas em regras de reescrita, as quais não explicam, de fato, como a entoação ou a fonologia é "utilizada" pelos falantes de uma língua, apenas como ela não pode ser usada, uma vez que o mau uso gera a agramaticalidade. A questão relacionada ao aspecto expressivo e linguístico é um dos problemas destacados por Chomsky e Halle (1968) no capítulo final de *SPE* quando tratam da noção de marca, onde irão assumir que sua abordagem é excessivamente formal:

The problem is that our approach to features, to rules, and to evaluation has been overly formal.(...) There is nothing in our account of linguistic theory to indicate that the result would be the description of a system that violates certain principles governing human languages. To the extent that this is true, we have failed to formulate the principles of linguistic theory, of universal grammar, in a satisfactory manner. In particular, we have not made any use of the fact that the features have intrinsic content. By taking this intrinsic content into account, we can, so it appears, achieve a deeper and more satisfying solution to some of

the problems of lexical redundancy as well as to many other problems that we have skirted in the exposition. (CHOMSKY; HALLE, 1968, p. 400)

Em certo sentido, Chomsky e Halle (1968) sugerem que as questões da implementação precisariam ser levadas em alguma conta para se analisar a língua como um todo, a questão é que essas necessidades expressivas violariam as regras pressupostas por eles no modelo fonológico. Assim, um sistema que tente englobar o aspecto das regras (padrão/fonológico) ao aspecto da implementação (variável), como sugerimos, precisa levar em conta que a fala não pode ser vista como um sistema que representa uma constelação de elementos no nível simbólico e que são distintos em sua realização.

Por este motivo, parece nos adequado pensar que os fenômenos expressivos são parte do domínio linguístico, ao invés de imaginar que este fenômenos sejam meras variações dos elementos estáticos deste sistema regras, como no trecho citado acima. Dessa forma, adotaremos a noção de sistema como aquele que organiza um conjunto de elementos básicos em processos que os transforma no tempo e não como um conjunto de regras e elementos organizados em um conjunto finito.

4.3 Contínuo, descontínuo

Utilizando o conceito acima exposto, de que sistemas não se tratam de uma constelação de elementos inter-relacionados, mas, sim, de processos de transformações derivados destas relações, sugerimos que a produção da fala possa ser compreendida como um sistema acoplado, variável no tempo, como descrito na seção 4.2, com duas componentes básicas: de um lado um (i) *sistema oscilatório contínuo*, do outro um (ii) *sistema oscilatório descontínuo*.

O primeiro daria conta da produção dos segmentos, sem um impacto direto na entoação de uma frase, uma vez que ele opera no nível das constrições supra-glotaais, como o movimento da língua ou a abertura dos lábios. Seu caráter descontínuo está associado ao fato de que esses movimentos têm uma curta duração, além de estarem em

uma veloz sucessão durante o ato de fala. O segundo sistema seria o responsável pela produção dos sons ligados à entoação, isto é, a produção da frequência fundamental na glote, a qual precisa ser contínua ou semi-contínua.

Esta separação só é possível quando olhamos para a fala como um processo sistêmico e temporal, e não como uma atualização sistemática de elementos estanques, como sugere a Fonologia Gerativa (CHOMSKY; HALLE, 1968; PIERREHUMBERT, 2000). Gunnar Fant (1973) ao fazer uma revisão do seu *Preliminaries to Speech Analysis* (JAKOBSON; FANT; HALLE, 1952) faz uma observação justamente sobre este ponto e sugere que os traços distintivos não são suficientes para análise das rápidas variações que envolvem a produção da fala.

Para Fant (1973) esse fato se torna evidente quando executamos uma análise espectrográfica e notamos uma quantidade maior de sons do que traços distintivos. Esses sons são resultado da combinação por superposição dos traços, as coarticulações, que são resultado de um efeito temporal. Para Fant (1973) o modelo de traço, como pensado por ele e seus colegas em Jakobson, Fant e Halle (1952), não poderia dar conta desta sucessão temporal, que é fundamental para o estudo da fala, justamente por terem considerado os traços como elementos totalmente estáticos. Para o autor, quando falamos da produção da fala temos duas dimensões a serem consideradas:

"The speech spectrogram display a mixture of continuous e discontinuous elements. A succession of 'segments' is to be seen but what appears from the spectrogram aspiration segment or the occlusion segment of a noninitial unvoiced stop. In other instances a piece of speech that stands out from the rest of the sequence as a separate unit, e.g., in virtue of a continuity of voicing, may be associated with several successive phonemes"(FANT, 1973, p.154).

Para Fant (1973) a teoria dos traços distintivos é capaz de explicar os primeiros elementos, tratando deles como categorias estanques. Por exemplo, de um ponto de vista da distinção categórica, o fonema /i/ não teria correlação direta com o fonema /ɨ/, pois trata-se de um som vocálico oral em oposição a um som consonantal nasal. Dentro desta perspectiva, os traços distintivos dão conta de analisar todo o espectro

sonoro de forma teórica. Se tratarmos a realização fonética como um fenômeno sucessivo e altamente estável, essa hipótese seria válida.

No entanto, em um vocábulo como *wing* ou *weeding* do Inglês Americano, o espraçamento da nasalidade pode ocorrer durante o levantamento da ponta da língua em direção ao palato, de modo que alguma quantidade do fluxo aéreo seja direcionado para a cavidade nasal, durante a produção da vogal, numa espécie de antecipação da nasalidade. Sendo assim, embora os traços destes sons sejam significativamente distintos, é natural considerar que eles sejam produzidos em um *continuum*, sendo o /ŋ/ apenas um apêndice a vogal, já nasalizada, sobrepondo os traços.

Esse problema pode ocorrer graças a segunda característica da produção da fala: os fenômenos contínuos. Fant (1973) não trata diretamente desta característica, porém entendemos que exista uma variedade de fenômenos fonéticos que podem ser explicados pelo aspecto contínuo da produção da fala, como o de espraçamento de vozeamento ou nasalidade. A alternância [b] / [p], como em *pudim* e *budim*; *princezinha* e *brincezinha* (LARA; BATTISTI; FILHO, 2013), ou a pronúncia [muiⁿtʊ]. Nestes casos, teríamos um efeito retardado do fechamento das pregas vocais ou do palato o que alteraria a sonorização ou nasalização dos segmentos graças a incompletude do movimento.

Desta forma, os *fenômenos descontínuos* seriam os segmentos, ao passo que os *fenômenos contínuos* seriam processos subjacentes aos segmentos, mas que estão associados a eles de forma direta. Hierarquicamente, o sistema descontínuo é dependente do sistema contínuo, uma vez que a condição de existência do primeiro fenômeno é a manifestação da frequência fundamental, ao menos nas vogais, uma vez que é necessário que sons se articulem e tenham diferenças entre si para que exista a linguagem, como já havia postulado Saussure (2012) com o conceito de valor.

Com relação aos fenômenos contínuos, subjacentes aos significados, Troubetzkoy (1976) já havia atentado para algo parecido. Para o autor, os fatos prosódicos se valeriam desta característica contínua da produção dos sons e se sobreporiam aos

segmentos em extensão. Assim a prosódia poderia ser parte da oração ou da sílaba como um todo (TROUBETZKOY, 1976) e não somente da vogal, como no modelo de Chomsky e Halle (1968).

Troubetzkoy (1976) atribui a estes fenômenos o nome de *prosodemas*, e sugere que eles devam ser interpretados como modulações dos segmentos e não como fenômenos estanques, isto é, não como categorias que se associam a outras categorias. Por esta razão estariam no nível sintagmático, uma vez que estabeleceriam sentidos complementares, ou ainda como inerentes a própria palavra, como o acento.

Nous appelons de *prosodème* la plus petite unité prosodique de la langue en question, autrement dit la syllabe dans les langues qui comptent les syllabes et la more dans les langues qui comptent les mores. Les particularités peuvent être divisées en particularités de différenciation et en particularités de mode de liaison. Par les particularités de mode de différenciation les prosodèmes sont distingués les uns des autres, tandis que par les particularités de mode de liaison ce ne sont pas les prosodèmes eux-mêmes, mais seulement la manière dont ils se lient à l'élément phonétique suivant qui est indiquée de façons variées. (TROUBETZKOY, 1976, p. 213)

Nota-se que para o autor elementos prosódicos que tivessem função distintiva, como o acento lexical em certas línguas, não seriam compreendidos como um signo linguístico, ao passo que aqueles usados para diferenciar sentenças seriam compreendidos como signos linguísticos independentes, dado que o significado por eles veiculados estaria contido neles mesmos. Troubetzkoy (1976) exemplifica esta premissa com a entoação de *continuidade* que indica que a frase não foi terminada ainda.

(...) Cette différence de principe consiste en ce que les phonèmes et les particularités prosodiques distinguant des mots ne sont jamais en eux-mêmes des *signes linguistiques*, mais toujours et seulement des *parties de signes linguistiques*. Le phonème *m* n'a par lui-même aucune valeur de signe, il ne désigne, il ne signifie rien, il n'est qu'une partie de différents signes linguistiques (...). Par contre les procédés distinguant des phrases sont des signes linguistiques indépendants: l'intonation "de continuité" *signifie* que la phrase n'est pas encore achevée (...). (TROUBETZKOY, 1976, p. 245).

De um lado, teríamos uma Fonética do segmento, amplamente sustentada pelo sistema descontínuo da produção da fala. Do outro uma Fonética da Entoação,

apoiada sobre o caráter contínuo da fala. Para Troubetzkoy (1976), o estado dos estudos linguísticos em sua época ainda não eram suficientemente avançados para se ter uma *fonologia da frase*, como aquela que seria proposta por Pierrehumbert (1980). Troubetzkoy (1976) demonstra que a questão está em torno da metodologia de coleta dos dados e não dos princípios teóricos em si, ou seja, seriam necessários avanços metodológicos para que observações feitas no nível da palavra pudessem ser replicadas no nível frasal.

Para Troubetzkoy (1976) outro ponto importante seria a necessidade de se avaliar o *status* fonológico do tom em uma determinada língua. Caso se averiguasse que a tonalidade de um item lexical é distintiva, como ocorre nas línguas tonais, a análise entoacional teria de se sobrepor a este primeiro nível, ou seja, o tom fonológico é mais proeminente que a entoação. Agora, se o tom não tiver nenhuma participação na fonologia de uma determinada língua, seria possível analisar a entoação, pensando em uma possível fonologia da frase. Desta maneira, a entoação “seria um meio fonológico para se diferenciar as frases” (TROUBETZKOY, 1976).

Com relação ao *status* do tom, pressuposto parecido é adotado na teoria Autossegmental/Metrical (LADD, 1996, p.15). A nosso ver, essa discussão é bastante elucidativa, de modo que pretendemos retomá-la no capítulo 2, por ora nossa proposta é a de posarmos compreender a prosódia, e nesta seara se enquadra a entoação (HIRST; CRISTO, 1998), como o fenômeno contínuo da produção da fala. Enquanto isso, na outra face, os segmentos fazem parte do sistema descontínuo, ainda que sejam dependentes, em partes, deste fenômeno contínuo.

A noção de sistema, por nós apresentada, tem melhor aplicação justamente quando analisamos a produção da fala. Se considerarmos que as duas etapas de produção (continuidade e descontinuidade) estão forçosamente unidas em um sistema de correlação direta e inequívoca, isto é, a variação na frequência fundamental (acréscimo ou decréscimo de F_0) sempre implicará em uma variação no segmento produzido, então,

poderíamos crer que as variações entoacionais, como elevações de *pitch*, influenciam, de alguma forma, na produção dos segmentos. Parece não haver evidências para esse fenômeno, se não o contrário, como a variação de F_0 que não implica em uma alteração na qualidade dos segmentos.

Leben (1978) demonstra que na língua Mende, a adição de um morfema atonal a um substantivo pode alterar o comportamento do tom deste morfema, graças ao contexto tonal anterior, mas que essa alteração não afeta a qualidade dos segmentos, sejam eles do substantivo base ou do morfema. Em suma, a alteração do tom ocorre sem que, necessariamente, existam alterações na cadeia segmental, como pode se ver em (7):

(7)	a. Tom	Substantivo	+ hu 'em'	+ ma 'no'
	b. /H/	kó - guerra	kó-hú	kó-má
	c. /L/	b̀̀̀̀ - calças	b̀̀̀̀̀-̀̀̀̀	b̀̀̀̀̀-̀̀̀̀
	d. /HL/	mbû - coruja	mbú-̀̀̀̀	mbú-má
	e. /LH/	mbă - arroz	mbà-hú	mbà-má
	f. /LHL/	nyáhâ - mulher	nyàhá-̀̀̀̀	nyàhá-mà

Dados como o de Leben (1978) e o *corpus* robusto da Fonologia Autossegmental (GOLDSMITH, 1990b) e da Fonologia Entoacional (LADD, 1996) nos permitem sugerir que a fala seja vista como a composição de dois sistemas acoplados e interdependentes, de modo que as variações entoacionais, por exemplo, não dependem dos itens lexicais ou segmentais presentes em uma elocução, sendo causadas pelos próprios mecanismos de produção da frequência fundamental e ativados deliberadamente pelos falantes.

Vale dizer que os segmentos também estão condicionados a mesma regra e podem ser alterados por características próprias dos articuladores ativos, independente da entoação específica. Estudos que utilizam a técnica do *bite-block*, em que os articuladores são limitados demonstram que os falantes são capazes de produzir os segmentos de

forma relativamente adequada, seja de um ponto de vista da qualidade ou da duração (FOWLER; TURVEY, 1980; SMITH, 1987; JACKS, 2008).

Entretanto, algumas evidências demonstram que a variação de F_0 pode ser intrínseca a um segmento (Intrinsic F_0 - IF0) e, por este motivo, nossa suposição de não influência entre sistemas poderia estar errada. Ohala (1978), por exemplo, demonstra que a diferença de *pitch* entre vogais altas e baixas pode ser da ordem de 25Hz (OHALA, 1978, p.29) e seriam causadas pelo movimento de alçamento da língua (*pull tongue*) (OHALA; EUKEL, 1987), movimento esse que levaria a um aumento da tensão nas pregas vocais e por consequência um ganho de *pitch*.

Em uma revisão bibliográfica do tema, Whalen e Levitt (1995) demonstram que esse efeito não é causado por um aumento deliberado de F_0 , como o que supomos nessa tese, mas, sim, condicionado pelas características de produção das vogais, como sugerido por Ohala e Eukel (1987). Desta maneira, os valores de IF0 seriam gradientes e relacionados à altura vocálica: vogais mais baixas, menor IF0; vogais mais altas, maior IF0. Além disso, os autores destacam que outros movimentos, como anterioridade e posterioridade da língua ou arredondamento e protuberância dos lábios não são fatores relevantes para o acréscimo de F_0 .

Em línguas tonais sempre houve a pressuposição que em tons baixos (L) a entoação intrínseca seria neutralizada, sendo encontrada apenas em vogais com o tom alto (H) ou então não existiria, uma vez que sua presença poderia afetar a qualidade associada ao tom lexical ou então seria um contra-efeito causado pelo próprio do decréscimo de *pitch* e que poderia ser encontrado em línguas não-tonais (LADD; SILVERMAN, 1984). Connell (2002) demonstra que algumas línguas tonais, de fato, não possuem o IF0 em suas vogais, não importando a alturas delas ou do tom fonológico, a sugestão do autor é que o tom associado ao item lexical não é determinante para a existência ou não de IF0, mas, sim, a modulação de F_0 utilizada para se produzir estes tons.

Outros estudos, como os de Pierrehumbert e Talkin (1992), Pierrehumbert (1995) e

Dilley, Shattuck-Hufnagel e Ostendorf (1996) demonstram que a estrutura prosódica pode ter um papel significativo na alteração da qualidade de alguns segmentos. De acordo com as autores, vogais em início de palavra tendem a ser mais glotalizadas quando estão no início frase entoacional e em maior grau quando esta palavra possuiu *pitch accent*, de modo que o contexto anterior não é significativo e não tem um impacto nesses casos. A sugestão de Dilley, Shattuck-Hufnagel e Ostendorf (1996) é a de que a hierarquia de fenômenos coloca a prosódia em primeiro plano e serve de base para os falantes orientarem-se durante a implementação fonética de uma elocução, sendo que a glotalização operaria como fonte de informação sobre a estrutura prosódica.

Em nosso entender, estes dados demonstram que a estrutura prosódica e entoacional pode servir de orientador para a implementação dos segmentos, enquanto o contrário parece não ser verdadeiro. Neste sentido, as escolhas lexicais parecem não ter um efeito significativo na variação da entoação, de modo que não é o conteúdo lexical e, principalmente, a cadeia segmental de uma determinada frase que irá determinar a forma como a entoação é implementada, sendo que essa tarefa cabe ao falante e as suas escolhas comunicativas/expressivas. Desta forma, podemos pensar que os sistemas contínuo e descontínuo estão associados como sistemas acoplados, descritos abaixo na seção 4.3.1.

4.3.1 Sistemas acoplados

Sistemas acoplados são aqueles que dividem ao menos uma variável independente e pelo menos duas variáveis dependentes (HEDRIH, 2008; GEISER, 2014). Variáveis dependentes, como o próprio nome sugere, são aquelas que dependem da variável independente para terem os seus valores alterados, de modo que a variável independente é o *input* da equação. Na análise da fala, podemos considerar o tempo (t) como uma variável independente, já que a fala é a sucessão de elementos numa elocução, organizados como uma série temporal (MORETTIN, 1999; FERREIRA-NETTO, 2008; EHLERS, 2009). Deste modo, as duas variáveis dependentes (de t) seriam: (i) a *frequên-*

cia fundamental e (ii) *as frequências de ressonância*. (FANT, 1966; FLANAGAN, 1983; STEVENS, 2000)

Esquemáticamente, no modelo fonte-filtro (FANT, 1966; STEVENS, 2000), a produção da fala se dá por um processo de transferência de energia, fundamentalmente baseado na Terceira Lei de Newton, a da ação e reação. Ao se fechar, a glote passa a exercer uma resistência R para o fluxo de ar (ϕ) que vem do pulmão, de modo que o volume de ar expelido pelo pulmão precisa de uma força $\vec{F}$ maior que a resistência mecânica da glote para fazer com que as pregas vocais vibrem. Stevens (2000) determina que este valor $\vec{F}$ precisa ser maior que 50kPa, ou seja, da ordem de 51 N/m² para que se possa vencer a inércia das pregas vocais

Anatomicamente, a produção da fala é dividida em três partes (FANT, 1966; STEVENS, 2000):

- (1) A **região sub-glotal**, onde se encontram os pulmões ou a fonte de energia ($E_p = P_{atm} + P_s$);
- (2) A **glote**, onde estão as pregas vocais, que seria o oscilador estimulado pela fonte de energia ($E_c \rightarrow F_0$);
- (3) A **cavidade oral-nasal**, que seria o ressoador deste oscilador ($F_0 \rightarrow F_{Res}$).

Figura 4.2: Representação esquemática do trato vocal e das forças que operam na produção da frequência fundamental

Como se nota pela figura 4.2 nos pulmões o ar têm uma massa L , com uma energia potencial E_p , essa Energia E_p é convertida em Energia Cinética E_c , que ao passar pela constrição formada pela glote passa a sofrer uma variação angular ω , se tornando a Frequência Fundamental F_0 . A transformação da massa L em E_c seria a primeira etapa deste processo e alimentaria o *sistema oscilatório contínuo* de produção

da fala. A taxa de transferência dessa energia E_c para o ambiente está condicionada às constrições da cavidade oral-nasal, de modo que os ressoadores podem alterá-las de n formas, tornando este parte do *sistema descontínuo*. Iremos explorar mais o processo de produção de F_0 no capítulo 2.

Vamos tratar do sistema contínuo, o qual a nosso ver segue um modelo simples de transformação do fluxo aéreo dentro de um tubo fechado em seu fim, como apresentado na equação 4.2 (STEVENS, 2000):

$$f_n = \frac{c}{4\ell}(2n - 1) \quad (4.2)$$

Equação 4.2: Equação de transferência do volume para o trato vocal.

Onde:

c = velocidade do som, por se tratar de função velocidade, podemos assumir que ela é derivada de uma função aceleração, deste modo, a Força $\vec{F}$, necessária para vencer a resistência mecânica da glote, é pressuposta.

ℓ = comprimento do tubo medido desde sua fonte de energia.

n = Número do harmônico.

A glote pode ser considerada um oscilador estimulado (PIKOVSKY; ROSENBLUM; KURTHS, 2001), de tal modo que seu movimento se iniciará somente e somente se houver algum estímulo $\vec{F}$ maior que sua própria inércia, o qual irá gerar uma frequência angular (ω) (STEVENS, 2000). A frequência angular, representa a taxa de variação angular que um corpo, em um movimento rotacional, se deslocou de sua posição de origem s_0 em um determinado espaço de tempo.

A vibração das pregas vocais tem de ser um movimento cíclico, isto é, ela precisa sair de sua posição de repouso, atingir o seu máximo estado ($\pi/2$) e retornar para sua posição de repouso (s_0) para começar um novo ciclo, sem esse movimento, o som da glote seria um estampido. Desta forma a frequência angular (ω) define a existência de

um ciclo, necessário para estabelecer a frequência de uma onda sonora.

Após o *kick* inicial causado pela força $\vec{F}$, o movimento de vibração continuará existindo, enquanto a fonte puder transferir energia e manter $\omega > 0$. Note-se que pode haver oscilações locais da frequência gerada por este movimento, mas a natureza do processo permanece a mesma, isto é, enquanto a relação entre a força $\vec{F}$ e a variação ω for maior que zero, haverá uma onda sonora.

Sendo assim, a força $\vec{F}$ pode ser definida como a diferença da pressão da atmosfera (P_{atm}) para a pressão sub-glotal (P_s). É necessário que essa força atinja um determinado valor, o suficiente para vencer a inércia da glote e iniciar seu movimento oscilatório, esse valor é chamado de limiar de pressão (LUCERO et al., 2009). A partir disso, a variação ω iniciada se manterá enquanto houver um fluxo de ar (m^3/s) de (P_s) para (P_{atm}). Essa relação é o elemento que estimula o oscilador, no caso a glote, base para o sistema contínuo, e a coloca em movimento.

O sistema acoplado a esse oscilador é o sistema descontínuo. O chamamos de descontínuo, pois seus elementos de saída não são necessariamente dependentes da frequência angular ω . Pelo contrário, algumas constrições orais estão encadeadas em movimentos que não são cíclicos e, sim, lineares, com um coeficiente angular α . As pautas gestuais da Fonologia Articulatória (BROWMAN; GOLDSTEIN, 1989; STUDDERT-KENNEDY; GOLDSTEIN, 2003) têm demonstrado que os movimentos destes articuladores são a soma de diversos elementos lineares coordenados para produzir uma determinada sequência. Por exemplo, a produção de um tepe não pode ser considerada um som contínuo, uma vez que ele não possui uma variação angular para sua produção, embora isso seja óbvio de um ponto de vista da lógica.

Nesse sentido, as consoantes surdas não têm correlação alguma com o sistema contínuo e são totalmente descontínuas. Por outro lado, as consoantes sonoras e as vogais são dependentes deste sistema. No caso das vogais, as equações que definem as frequências de ressonância são variações da equação 4.2 e levam em consideração,

essencialmente, a variação de comprimento do tubo ℓ , ocasionada pelo movimento dos articuladores ativos e o grau de distorção de F_0 , derivada desta variação. Não iremos discutir os pormenores de cada uma delas uma vez que não cabe apresentar um corolário de equações neste ponto. Porém, é preciso novamente destacar que o produto delas é variável e condicionado a estes movimentos, os quais têm como característica a descontinuidade.

Desta forma, podemos considerar a frequência fundamental da fala como o resultado de um sistema oscilatório contínuo, ao passo que consoantes e vogais são partes de um sistema descontínuo. O acoplamento destes sistemas é o que produz a fala e os seus elementos expressivos. Nossa hipótese é que esse sistema contínuo pode ser controlado e ajustado de acordo com as necessidades expressivas do falante, a despeito das estruturas existentes no sistema descontínuo. A nosso ver, esses elementos que compõem o sistema descontínuo são os responsáveis por essa "articulação" tonal.

As evidências para esse tipo de fenômeno são bastante relatadas nos estudos da AutoMetrical Phonology (LADD, 1996), que tendem a compreender a língua como uma estrutura em camadas, onde tons e segmentos estão em camadas diferentes e, por este motivo, as regras fonológicas aplicadas a uma não necessariamente afetam a outra. Na seção 4.4 trataremos deste processo de acoplamento de maneira mais detalhada.

4.4 Acoplamento

Como apresentado na seção anterior, a fala pode ser entendida como a combinação de dois sistemas com finalidades distintas que se acoplam para a produção dos seus sons. Dentro do modelo fonte-filtro, esse acoplamento é sempre linear, isto é, a pulsação glotal é independente do trato vocal, de modo que o resultado final da fala pode ser compreendido como a convolução de duas funções distintas: de um lado a função que estabelece as frequências da fonte, do outro a que determina as frequências do filtro (FANT, 1966; FLANAGAN, 1983; STEVENS, 2000). Desta maneira, de forma

simplificada, o som que ouvimos durante o ato de fala é soma indireta de todos os *inputs* acústicos que o determinam.

Se olharmos o fenômeno por essa perspectiva, o resultado da fala é basicamente um processo de controle da pressão subglotal, por meio da manutenção do fluxo de ar dos pulmões, e um conjunto de ajustes dos músculos da laringe e da cavidade oral (FANT, 1966; OHALA, 1983; STEVENS, 2000). A frequência fundamental seria o resultado direto deste oscilador estimulado e contínuo, uma imagem temporal de sua variação. Já os segmentos seriam a modulação da intensidade em harmônicos específicos, os quais definiriam a qualidade de cada um das vogais e sonorantes, por exemplo.

Titze e Riede (2008) e Titze (2008) sugerem que este modelo linear não é capaz de descrever todos os fenômenos acústicos que envolvem o acoplamento destes dois sistemas. Para os autores, além do modelo linear de Fant (1966) e Stevens (2000), existe, também, um mecanismo de acoplagem não-linear, que seria acionado em algumas situações específicas. Nestes casos, a pressão transglotal, isto é, o diferencial de pressão na glote poderia alterar as frequências de ressonância do filtro, fazendo surgir elementos como sub-harmônicos, que fazem variar a qualidade de um som oral.

Titze e Riede (2008) sugerem que essa alteração se dê por meio da retroalimentação da pressão subglotal, causada por um estreitamento do vestíbulo laríngeo, a região anatômica que compreende do ádito da laringe (a ligação faríngeo-laríngeo) à glote, como pode ser visto na figura 4.3. O estreitamento dessa região pode aumentar a impedância do vestíbulo, alterando a relação da força $\vec{F}$ com o fluxo de ar da glote para cima, o que afetaria as frequências da fonte e indiretamente do filtro.

Figura 4.3: Corte sagital da glote demonstrando o vestibulo laríngeo, a região acima glote.

De acordo com Titze (2008) o aumento de impedância do vestibulo² poderia causar uma desestabilização do sistema, o que levaria a um cruzamento de F_0 com F_1 em alguns pontos, seja por decréscimo de F_1 , seja pelo aumento de F_0 . Esse *crossover* de $F_0 - F_1$ alteraria a vibração das cordas vocais, causando alterações na fonte, ao menos, teoricamente. Além disso, essa diferença entre impedâncias poderia causar o efeito de sub-harmônicos acima citado.

Para Titze e Riede (2008) a acoplagem não-linear pode ocorrer, devido à patologias ou necessidades expressivas do falante. Neubauer, Edgerton e Herzel (2004) demonstram que cantores de improviso, como Demetrios Stratos³ eram capazes de fazer esses acoplamentos não-lineares, com o intuito de fazerem suas vozes de forma polifônica e não apenas centrada em fórmulas melódicas. Herzel et al. (1994) demonstra que

²Supõe-se que a impedância na glote é sempre maior do que no resto do sistema como um todo

³Demetrios Stratos (1945-1979) foi um músico, compositor e foneticista italiano, que fez diversos estudos no campo da musicologia, utilizando a voz e a fonação como recurso musical. Seus trabalhos partiam do pressuposto que a voz poderia ser utilizada tal como um instrumento. Para esse fim, ele utilizava extensões vocais, recortada por modalizações segmentais. Entre seus trabalhos destacam-se *Cantare la voce* (STRATOS, 1978) e a participação no álbum de John Cage, *Nova Musicha N. 1* (CAGE, 1974)

alguns problemas relacionados à saúde vocal, como a *diplofonia* (dois *pitchs*) ou a voz gutural (*vocal fry*) em voz saudáveis podem ser entendidas dentro desta perspectiva como uma forma alternativa de produção da voz e não como desvios.

Mende, Herzel e Wermke (1990) demonstram que o choro de recém-nascidos também é cercado destas variações não-lineares. Fitch, Neubauer e Herzel (2002) sugerem que pequenos primatas, como os macacos Rhesus, também apresentam essas variações não-lineares em suas vocalizações, o que sugere que a não-linearidade vocal pode ser resultado de um processo evolutivo a ser investigado. Zhou, Hansen e Kaiser (2001) demonstram que é possível sugerir uma tipologia da voz a partir dessas análises, bem como Kumar et al. (2016) demonstram que é possível prever e modelar essas variações não-lineares em sistemas de equações.

Em suma, o acoplamento destes dois sistemas pode seguir uma regulação ativa, feita pelo falante de acordo com suas necessidades, características ou limitações expressivas. Deste modo, existem dois problemas a serem considerados. De um lado, a fala segue um processo determinístico, onde os *inputs* não têm uma correlação direta com os *outputs*, mas é possível determinar os resultados de forma totalmente previsível. Do outro, como citado na seção 4.2, os *outputs* podem ser regulados pelos mesmos processos determinísticos e sofrer algum tipo de variação no processo, o que geraria resultados aparentemente aleatórios.

Sistemas que são determinísticos e passam por estados aleatórios em sua constituição tendem a ser caóticos, como descrito no âmbito da Teoria do Caos (ALLIGOOD; SAUER; YORKE, 1996; BADDI; POLITI, 1999; THOMPSON; STEWART, 2002). Lorenz (1963) já observava que sistemas hidrodinâmicos, ou seja, que possuem algum tipo de fluxo de A para B, podem variar abruptamente, saindo de um estado linear, para um estado não-linear e retornar para o primeiro estado. Para Herzel et al. (1994) a fala pode ser compreendida desta mesma maneira, de modo que diversos *outputs* que são considerados atípicos, sejam, na realidade, resultados destes acoplamentos

não-lineares.

De um ponto de vista físico-fisiológico, o que estes dados revelam é a capacidade intrínseca e uma determinada plasticidade na variação de F_0 comandada pelo seu falante e independente do caráter linguístico daquilo que é produzido no sistema descontínuo. Os dados relativos ao acoplamento também sugerem que determinados fenômenos podem ser mapeados de forma inversa, isto é, uma alteração da entoação, pode representar uma mudança de qualidade vocal, por exemplo, mas sem causar uma alteração segmental.

Podemos, então, sugerir que as necessidade expressivas estão inseridas no próprio código da língua e são atualizadas na fala durante o processo de acoplamento destes sistemas. Os ajustes comunicativos feito pelos falantes não se tratam de meras idiossincrasias, como tratava Saussure (2012), e são derivados de um processo muito mais amplo, que envolve o ajustamento das variações contínuas e dos movimentos descontínuos.

5 Teste 1

Neste capítulo vamos reportar os testes experimentais que conduzimos com o intuito de encontrar evidências sobre os alvos tonais. Nossa proposta era a de eliciar nos falantes a consciência de se atingir determinados alvos a partir da escuta de estímulo e da replicação. Os resultados demonstram que os falantes são de fato capazes de armazenar a informação sobre os alvos e replicá-las de forma bastante razoável.

5.1 O desenho experimental

No Capítulo 3 definimos os alvos como unidades que fazem a coordenação da necessidade linguística do falante e a implementação entoacional no domínio físico. Desta forma, em uma face os alvos operam como mecanismos cognitivos, pois mapeiam as informações linguísticas pretendidas pelo sujeito e na outra face coordenam as operações musculares que servem de base para a implementação fonética, como definimos nos Capítulos 2 e 4

Portanto, os alvos não são unidades puramente representacionais, nem mesmo unidades essencialmente fonéticas. No domínio da análise essa definição nos permite trabalhar com um grau significativo de variabilidade, isto é, a relação entre um tom e sua manifestação não é unívoca ou relacionado com um valor fixo (PIERREHUMBERT, 1980) e, sim, estabelecida pelas próprias características que cercam um tom, como sugerido por Xu e Wang (2001).

Deste modo, o design experimental aqui relatado partiu de dois princípios. O primeiro deles é o de que os falantes são capazes de processar os tons da fala de forma produtiva sem depender de outras pistas, ou seja, os falantes são capazes de remontar um conjunto de alvos tonais de uma dada elocução. O segundo é o de que eles são capazes de processar e implementar esse mesmo conjunto de alvos quando estimulados. Neste segundo caso, essa implementação necessariamente respeita as características vocais particulares do falante, de modo que ele pode implementar um alvo a partir de uma referência sem a necessidade de alterar o seu *pitch range*, por exemplo.

Pensando nestes dois princípios estabelecemos um teste de repetição, no qual após ouvir um estímulo o participante precisa repeti-lo, mantendo um determinado grau de correlação com aquilo que ele ouvira. A nosso ver, a melhor forma de analisar o processamento dos tons como alvo era esse, uma vez que o referencial estaria definido pelo estímulo e a tarefa do falante seria “encontrá-los” e reproduzi-los. A literatura sobre este tipo de teste, chamado de *imitação elicitada*, é relativamente vasta, por ser bastante comum em estudos de aquisição de L2 ou L1 (SLOBIN; WELSH, 1973; VINTHER, 2002).

Assim, nosso teste partia de duas hipóteses:

1. Falantes do Português Brasileiro são capazes de inferir um conjunto de alvos tonais presentes em uma elocução.
2. Se isso é verdade, eles são capazes de reproduzir esse conjunto tonais, respeitando os limites da sua voz.

Na montagem do teste piloto nos deparamos com mais uma questão teórica: até que ponto os falantes estariam sustentando suas repetições apenas em sua memória e não em seu processamento linguístico? Caso isso seja um problema, até que ponto memória e processamento linguístico devem ou não estar integrados?. Hamayan,

Saegert e Larudee (1977) argumentam que esse é um problema a ser considerado neste tipo de teste, de modo que a informação final a ser analisada pode estar associada apenas à memória do falante acerca do estímulo e não ao seu conhecimento linguístico.

Entendemos que o processamento deste tipo de informação acontece em conjunto com a memória de trabalho (COWAN, 2005), a qual funciona como um domínio operativo, com capacidade limitada e que é capaz de armazenar informações que serão rapidamente processadas. Okura e Lonsdale (2012) demonstram que não existe uma correlação significativa entre este tipo de memória e o processamento linguístico de um teste de imitação elicitada. A hipótese dos autores era a de que a repetição de palavras *non-sense* seria processada pela memória de trabalho, ao passo que palavras da língua inglesa seriam processadas pelo conhecimento linguístico. Deste modo, seria natural que os primeiros dados tivessem uma grande regularidade em torno da média entre-sujeitos, ao passo que o segundo teria uma variação relativa, graças ao conhecimento de cada falante, o que de fato foi encontrado pelos autores.

Tendo isso em mente estabelecemos dois pressupostos de teste:

1. O estímulo não poderia ter uma extensão significativa, para evitar que o falante tivesse de recorrer a sua memória;
2. O estímulo teria que ter uma variação pontual e saliente além de uma variação *flat*, que não chamasse a atenção dos participantes.

Partindo destes princípios, na seções a seguir apresentaremos com mais detalhes o estímulo, bem como as condições experimentais e também os resultados obtidos por nós até o momento.

5.2 Análise experimental

5.2.1 Caracterização do estímulo

O estímulo desse teste consistiu na repetição de uma frase, retirada de uma cena de novela, a qual, a nosso ver, atendia todos os pressupostos acima delineados. Com cerca de 7 segundos de duração, sem marcas de regionalismo e, principalmente, de algum tipo variação na qualidade vocal. A frase era entoada por uma voz masculina e uma feminina na sequência. Seu conteúdo linguístico era o seguinte.

- (1) Voz masculina: “Eu tava dormindo, de repente, acordei, dei de cara com o Domenico.”
- (2) Voz feminina: “O Domenico entrar aqui ... como assim?”.

A figura 5.1 apresenta o desenho da curva entoacional das frase acima descrita.

Figura 5.1: Curva de frequência do estímulo apresentado aos falantes em nosso teste.

As marcas em azul demonstram os dois pontos de interesse. A primeira marca representa o ponto em que o alvo tonal é a variação, o que temos chamado de *variação topológica*, a segunda marca representa o *locus* em que o alvo é bastante específico,

o que temos chamado de *variação pontual* e representam uma ruptura dos primeiros alvos, os quais serão melhor definidos ao longo deste capítulo. Dessa forma, o primeiro alvo não se trata de um único elemento específico, mas o conjunto total de alvos que o constitui, ou seja, da capacidade do falante implementar a variação de F_0 . Por outro lado, o segundo alvo visa analisar a capacidade de avaliação e percepção específica da proeminência.

Nosso entendimento, é que os alvos não podem ser compreendidos exclusivamente como pontos específicos, como o são as proeminências, de modo que precisam abarcar a variação subjacente a elas. Retomando a discussão entre *níveis* e *configurações* apresentadas no capítulo 2, entendemos que a principal dificuldade teórica e metodológica é justamente fazer este tipo de transposição. Dessa forma, nossa proposta é analisar esse estímulo em duas frentes a primeira delas é a que analisa as variação topológica, aquela sem as marcas de alguma proeminência significativa, ao passo que a segunda pretende analisar a variação pontual. A seguir, trataremos do aspecto experimental, seguido da análise de cada uma destas dimensões.

5.2.2 Condições experimentais

O teste aqui reportado foi aplicado o seguinte roteiro:

- Os participantes ouviam o estímulo uma única vez e caso tivesse alguma dúvida poderiam ouvir novamente;
- A seguir era pedido a eles que reproduzisse o estímulo, mantendo as mesmas características entoacionais;
- A tarefa era proposta de maneira simples pelo aplicador: “agora é só repetir a frase que você ouviu da forma que você ouviu”;
- Os participantes tinham que reproduzir as duas partes de forma contínua (voz masculina + voz feminina) em uma única elocução;
- Para evitarmos o problema da interface entre a repetição e a memória de trabalho

citado acima, os falantes tinha de executar a tarefa duas vezes. Em tese, a segunda execução dissolveria o problema da memória de trabalho;

- As amostras eram gravadas com um gravador portátil, no formato *.wav*, estéreo, com resolução de 32bit e *range* de 48kHz;
- O teste era aplicado em um ambiente de ruído controlado, mas não issonorizado, com um participante por vez .

No total, 15 falantes nativos de Português Brasileiro, sendo 8 mulheres e 7 homens, participaram do teste. Os participantes eram oriundos da região metropolitana de São Paulo ou habitavam a região há mais de 5 anos. Tivemos de excluir o *set* de gravações de dois sujeitos e a segunda repetição de dois deles, por questões técnicas, como a qualidade da gravação ou pelo não atendimento da execução da tarefa. Deste modo, obtivemos um total de 24 *tokens*, divididas em 2 condições amostrais, totalizando 48 *tokens*.

Nosso modelo amostral, construído com base no programa G*Power (FAUL et al., 2007), demonstra que para um teste-t de amostras emparelhadas com médias dependentes, o valor $\alpha < 0,05$ é satisfeito a partir de uma amostra de no mínimo 35 amostras, como se pode ver na figura 5.2

Figura 5.2: Demonstração da relação entre o valor de α e o tamanho n da amostra. A partir de 35 *tokens* o resultado do teste se mostra satisfatório estatisticamente

A figura 5.3 mostra que para obtermos um tamanho de efeito médio, da ordem de 0,4, precisamos de um total de 40 amostras. É importante ressaltar que o tamanho de efeito serve para calcularmos o magnitude de um fenômeno, seja o seu grau de correlação ou a sua probabilidade (COHEN, 1988)

Figura 5.3: Demonstração da relação entre o tamanho do efeito e o valor n da amostra. A partir de 40 *tokens* o resultado do teste se mostra satisfatório estatisticamente

Desta forma, o tamanho da nossa amostra aparenta satisfazer algumas condições experimentais prévias, de modo que julgamos desnecessária a coleta de mais dados por agora. Além disso, uma de nossas suposições é a de que o efeito de gênero, ou seja, a diferença entre homens e mulheres não é significativa, por este motivo a não consideramos em nossa análise.

5.2.3 Coleta de dados

Como gravamos as duas condições experimentais em um único áudio fizemos a separação de cada um delas por meio do Audacity, além disso, fizemos a mixagem dos canais por meio da mesma aplicação, de modo que as amostras analisadas eram todas mono. A frequência fundamental das amostras foi extraída utilizando o *software* Sound Filing System, por meio do processo de autocorrelação. Já a análise das características

prosódicas da entoação foi feita pela Rotina ExProsodia (FERREIRA-NETTO, 2010) e as análises estatísticas apresentadas foram feitas com as ferramentas do MS Excel e do SPSS.

A análise com a Rotina ExProsodia detecta as Unidades Portadoras de Entoação, definidas no Capítulo 3, a partir dos seguintes parâmetros:

1. Frequência fundamental maior que 20Hz e menor que 600Hz;
2. Intensidade de ao menos 10% do valor médio da intensidade total ;
3. Duração mínima: 20 ms.

Desta maneira, pode haver mais UPEs do que sílabas em uma elocução, visto que a Unidade não considera o aspecto segmental como elemento limitador. Calculados esses valores, a rotina os converte para MIDI e os relaciona com os limiares de diferenciação tonal. Como demonstramos em Martins e Ferreira-Netto (2010), a medição com MIDI parece ser menos precisa na análise de variação, pois engloba uma maior quantidade de frequências em único tom, ao mesmo tempo que em medidas relacionais, como comparação de alturas, ela demonstra uma precisão maior que a escala de Hz. Dito isso, destacamos que vamos utilizar estas duas escalas em nossas análises.

5.3 Primeira condição amostral: alvos topológicos

5.3.1 Caracterização das amostras

Em nossa hipótese entendemos que variações topológicas são aquelas as quais os alvos tonais operam em um espaço de frequência delimitado pelos limiares de diferenciação tonal, apresentados no Capítulo 3. Nesta primeira condição amostral nossa intenção é verificar se os falantes são capazes de intuir que (a) as variações de F_0 devem seguir um determinado padrão, estabelecido pelo estímulo, ou seja, os alvos topológicos e (b) notar que não é necessário que se ultrapasse os limiares de

diferenciação tonal de forma acentuada, de modo que a implementação da frequência fundamental se mantenha dentro do espaço entoacional.

Nesta condição experimental, o alvo é a implementação da frequência fundamental dentro do espaço entoacional, no entanto, a tarefa essencial é que os falantes sigam as variações do estímulo nessa implementação. Dessa maneira, são três as hipóteses a serem analisadas:

1. A correlação entre as variações de F_0 do estímulo e as das amostras (gravações dos sujeitos) - *análise populacional*;
2. A significância dessa correlação, ou seja, a execução ou não da tarefa - *análise populacional* ;
3. A acuidade com que os participantes executaram a tarefa - *análise de hipótese*.

Em nosso entender, a frase (1) é a que melhor satisfaz às nossas necessidades experimentais. A figura 5.4 representa a sua curva tonal, obtida a partir da Rotina ExProsodia. Nota-se a existência de três proeminências relativas, porém nossa análise prévia demonstrou que elas não são significativas e parecem ser algum tipo de vicissitude das ferramentas que utilizamos ou do áudio original.

(1) “Eu tava dormindo, de repente, acordei, dei de cara com o Domenico.”

Características vocais:

- Frequência fundamental média: 98,20Hz;
- Duração total: 4,4s;

Figura 5.4: Variação de frequência fundamental para a frase-estímulo que utilizamos em nosso teste dos alvos topológicos.

Para a primeira hipótese, utilizaremos um cálculo derivativo, comparando as médias de cada uma das amostras com a média do estímulo, por isso, usaremos uma análise distribucional, faremos o mesmo para a segunda hipótese, mas utilizaremos o coeficiente de variação nesse caso. Já para a terceira hipóteses, utilizaremos uma análise distribucional combinada a testes de hipótese e normalidade. A subseção 5.3.2 trata dos dois primeiros casos, ao passo que a subseção 5.3.3 trata do terceiro.

5.3.2 Análise distribucional

A análise distribucional se justifica, uma vez que os falantes podem não obedecer a orientação tonal estabelecida pelo estímulo e produzir elocução que de um ponto-de-vista fonético, são totalmente aleatórias. Se isso é um fato, significa dizer que os falantes não são capazes de intuir os alvos a partir de uma elocução. Outro ponto importante nessa discussão refere-se à tendência individual dos falantes ao pronunciar alguns segmentos. Assim, é natural que as extensões de cada elocução e, por conseguinte, a quantidade de UPEs de cada uma delas, seja diferente caso-a-caso.

Consideremos cada UPE um *score* que relaciona F_0 e tempo, para que não tenhamos de passar em revista por todos os valores de UPEs ($n=1819$), a alternativa mais lógica

é calculamos a média e o desvio padrão total de cada uma das amostras e do estímulo como se pode ver na tabela 5.1. A etapa posterior é justamente analisar a relação entre estes valores e a população.

Amostra	Frequência Fundamental			
	Média MIDI	DvP MIDI	Média Hz	DvP Hz
<i>Estímulo</i>	42,3	4,9	98,20	25,66
suj1	56,3	6,3	220,25	52,88
repsuj1	55,6	5,8	211,64	49,62
suj2	55,8	5,4	212,75	51,57
repsuj2	56,1	4,0	213,34	42,66
suj3	44,6	2,5	108,53	15,04
suj4	56,1	7,7	225,54	69,29
repsuj4	56,1	7,7	225,89	69,11
suj5	50,7	6,2	161,18	41,56
repsuj5	48,8	7,2	146,74	48,49
suj6	58,4	1,1	239,74	15,27
repsuj6	56,5	7,0	227,91	73,04
suj7	46,6	2,8	122,05	17,07
repsuj7	49,4	8,4	125,76	21,06
suj8	54,0	4,2	189,06	37,49
repsuj8	55,2	3,7	201,70	35,30
suj9	47,9	4,1	130,67	36,28
repsuj9	46,6	4,8	123,92	27,79
suj10	44,4	3,9	109,54	24,98
repsuj10	44,1	2,7	105,19	20,25
suj11	49,1	9,5	158,68	74,36
repsuj11	54,3	11,4	199,91	56,31
suj12	49,8	6,8	155,22	52,90
repsuj12	49,6	7,1	154,14	52,90
suj13	43,8	4,5	105,89	24,51
Geral	50,9	2,4	166,94	18,7

Tabela 5.1: Valores de média geral e desvio padrão da Frequência Fundamental das UPEs de cada uma das amostras da frase (1). Os valores estão em MIDI e Hertz

Pode se notar que os valores de média aparentam ter uma grande dispersão. Isso se dá, pois como acima dito o *pitch range* de cada um dos falantes é distinto entre sujeitos, o que automaticamente altera as medições das amostras. Por um lado, é natural que exista uma variação das médias de F_0 . Por outro lado, se essa variação é demasiadamente grande, isso pode indicar que os dados não estão corretos, seja por uma quantidade excessiva de *outliers* nas amostras, seja por algum tipo de enviesamento de ordem mais prática, como erro coleta ou na separação de grupos amostrais.

Como não estamos analisamos as UPEs uma-a-uma diretamente, temos de supor que as médias descritas na tabela 5.1 não são aleatórias. Por isso, o primeiro passo é calcular os valores de assimetria e curtose da média da frequência fundamental da população total. O primeiro deles indica o quão concentrado um determinado conjunto de valores está em relação à média. Uma assimetria positiva, indica que a massa de distribuição está concentrada à esquerda, uma assimetria negativa indica que os valores estão concentrados à direita. Por sua vez, a curtose indica o achatamento de uma curva normal. Se maior de 0, essa curva é *leptocúrtica*, de modo que seus valores estão distribuídos por sua cauda. Se maior de 0, indica que essa curva é *platicúrtica* e que seus valores estão distribuídos próximos à média.

Os valores de curtose e assimetria das repetições da frase (1) podem ser visto na tabela 5.2:

Estatísticas descritivas		
	Média Midi	Media Hz
N	25	25
Média	50,9	166,94
Variância	24,79	2269,25
Assimetria	-0,17	0,03
Curtose	-1,40	-1,59
Intervalo	16,10	141,54

Tabela 5.2: Valores de assimetria e curtose para as médias de Frequência Fundamental de cada uma das amostras da frase (1).

A partir destes valores, podemos sugerir que a distribuição dos dados é relativamente normal. Os histogramas apresentados nas figuras 5.5 e 5.6 também indicam essa distribuição. No entanto, é notável a existência de uma ligeira assimetria negativa, ou seja, uma massa de distribuição à direita. Essa variação pode ser explicada por haver um desbalanceamento no total de homens e mulheres, que têm um *pitch* relativo mais alto do que o dos homens. É também importante destacar que o cálculo da assimetria de cada uma das amostras revelaria apenas as características daquele conjunto e não da população, o que nos interessa mais perto.

Figura 5.5: Histograma da média de frequência fundamental para todas as amostras da frase (1). Os valores estão em MIDI.

Figura 5.6: Histograma da média de frequência fundamental para todas as amostras da frase (1). Os valores estão em Hertz

Estas análise demonstram, ao menos em parte, que estes valores não são totalmente aleatórios e que estão em uma distribuição quase normal. Por se tratar de uma análise indireta, testes de normalidade como o de Kolmogorov-Smirnov e de Shapiro-Wilk retornam uma não significância. A explicação é bastante longa, de modo que não vamos entrar nessa discussão por ora. Em nosso entender, as implementações feitas pelos participantes não foram aleatórias quando comparadas ao estímulo, ou seja, não se tratam de vicissitudes dos falantes, de modo que podemos prosseguir com a nossa análise.

Coeficiente de variação

Para avaliarmos se os falantes foram ou não capazes de intuir os alvos tonais a partir do estímulo, utilizamos o Coeficiente de Variação (*CofVar*), que é a razão entre a média de uma amostra e o seu desvio padrão e mostra a extensão da variabilidade

de uma amostra em relação à média (seja da população ou da própria amostra):

$$c_v = \frac{\sigma}{\mu} \quad (5.1)$$

Equação 5.1: Fórmula do Coeficiente de Variação.

Uma das característica do coeficiente de variação é o fato dele ser adimensional e relativo. Desta forma podemos desconsiderar o *pitch range* dos falantes, que hipoteticamente foi um dos fatores de assimetria. Explicamos, como o coeficiente é uma medida relativa, obtida a partir das variações da própria amostra, ele automaticamente dissolve a diferença de alturas entre-falantes, pois normaliza a variação de *pitch* para os seus valores. A tabela 5.3 apresenta os coeficientes de variação para cada uma das amostras, calculados a partir da equação 5.1.

Coeficiente de variação - F_0		
	CofVar - Midi	CofVar - Hz
<i>Estímulo</i>	11,7%	26,1%
su1	11,2%	24,0%
repsu1	10,3%	23,4%
su2	9,6%	24,2%
repsu2	7,1%	20,0%
su3	5,7%	13,9%
su4	13,8%	30,7%
repsu4	13,7%	30,6%
su5	12,2%	25,8%
repsu5	14,7%	33,0%
su6	1,8%	6,4%
repsu6	12,4%	32,0%
su7	6,0%	14,0%
repsu7	17,1%	16,7%
su8	7,9%	19,8%
repsu8	6,8%	17,5%
su9	8,7%	27,8%
repsu9	10,2%	22,4%
su10	8,8%	22,8%
repsu10	6,0%	19,2%
su11	19,3%	46,9%
repsu11	21,1%	28,2%
su12	13,8%	34,1%
repsu12	14,3%	34,3%
su13	10,2%	23,2%
Média	11,0%	24,7%
Assimetria	0,35	0,35
Curtose	0,89	1,08
Variância	0,20%	0,69%

Tabela 5.3: Valores do coeficiente de covariação de frequência fundamental de cada uma das amostras da frase (1).

A tabela 5.3 apresenta as mesmas questões de distribuição discutidas mais acima. Note-se que a assimetria da análise utilizando os Hertz é ligeiramente maior que aquela feita em MIDI. A hipótese que explica essa diferença é o valor de 46,9%, respectiva ao coeficiente de variação do sujeito 11. Embora esse valor seja alto em relação aos demais, uma análise utilizando o *box-plot* e outros testes de detecção de *outliers* demonstrou que se trata de um dado normal, que está no limite do desvio padrão, como pode se ver nas figuras 5.7 e 5.8, logo abaixo.

Figura 5.7: *Box-plot* com as duas medidas, Hz e MIDI.

Figura 5.8: Gráfico quantil-quantil demonstrando a posição do valor 46,9% indicado pelo asterisco.

A figura 5.7 demonstra que os valores mensurados em MIDI são relativamente melhores distribuídos do que aqueles medidos em Hz. Novamente, essa diferença se explica pelo fato do MIDI incorporar um maior número de frequências em um único valor. A análise do histograma apresentado na figura 5.9, sugere que a distribuição dos coeficientes de variação medidos em Hertz não é necessariamente assimétrica, embora tenha uma forte tendência a demonstrar uma assimetria à direita.

Figura 5.9: Histograma dos coeficientes de variação da frase (1). Os valores estão em Hz.

De modo geral, essa tendência de assimetria à direita é positiva, pois demonstra que os valores dos coeficientes de variação tendem a orbitar em torno do valor encontrado no estímulo (26,1%) e não a se dispersarem em torno dos valores de máximo (46,9%) ou mínimo (6,3%), por exemplo. Portanto, a partir destas análises, é possível afirmar com alguma segurança que os dados apresentados não são aleatórios, uma vez que nossa hipótese nula H_0 é a de que não há diferenças entre os coeficientes de variação da população. Assim, o indicativo da distribuição normal é positivo, pois sugere que os participantes foram capazes de executar a tarefa de forma satisfatória.

Para nos certificarmos dessa hipótese aplicamos um teste Kolmogorov-Smirnov para significância da normal, reportado na tabela 5.4. Como se pode ver o teste retornou uma significância positiva, de modo que não se deve rejeitar H_0 .

Testes de Normalidade - Kolmogorov-Smirnov ^a			
	Estatística	df	Sig.
Cof Var - MIDI	0,094	24	,200*
Cof Var - Hz	0,101	24	,200*

*. Este é um limite inferior da significância verdadeira.

a. Correlação de Significância de Lilliefors

Tabela 5.4: Tabela com os resultados do teste Kolmogorov-Smirnov para a distribuição dos coeficientes de variação da frase (1).

5.3.3 Análise contrastiva

Os dados apresentados na seção acima sugerem que os participantes foram capazes de executar a tarefa a contento. Significa dizer que (i) conseguiram de intuir os alvos topológicos e o espaço entoacional do estímulo (ii) implementaram de forma relativamente significativa a variação da frequência fundamental e, por consequência, os alvos. Assim sendo, precisamos agora avaliar o quão acurada foi a execução desta tarefa. Para isso, adotamos a seguinte fórmula do que chamamos de *índice de relação*:

$$i_r = CofVar_{am} - CofVar_{est} \quad (5.2)$$

Equação 5.2 Define o *índice de relação*, que utilizaremos em nossa análise.

Onde: i_r = índice de relação entre o estímulo e a implementação do falante;

$CofVar_{am}$ = Coeficiente de variação da amostra;

$CofVar_{est}$ = Coeficiente de variação do estímulo.

A partir deste cálculo obtivemos os valores apresentados na tabela 5.5.

Índice de relação		
	i_r MIDI	i_r Hz
su1	-0,5%	-2,1%
repsu1	-1,3%	-2,7%
su2	-2,1%	-1,9%
repsu2	-4,5%	-6,1%
su3	-5,9%	-12,3%
su4	2,2%	4,6%
repsu4	2,1%	4,5%
su5	0,5%	-0,3%
repsu5	3,0%	6,9%
su6	-9,8%	-19,8%
repsu6	0,7%	5,9%
su7	-5,7%	-12,1%
repsu7	5,4%	-9,4%
su8	-3,8%	-6,3%
repsu8	-4,9%	-8,6%
su9	-3,0%	1,6%
repsu9	-1,4%	-3,7%
su10	-2,9%	-3,3%
repsu10	-5,6%	-6,9%
su11	7,6%	20,7%
repsu11	9,4%	2,0%
su12	2,1%	7,9%
repsu12	2,7%	8,2%
su13	-1,5%	-3,0%
Média	11,0%	24,7%
Assimetria	0,314	0,310
Curtose	0,205	1,231
Variância	0,20%	0,69%

Tabela 5.5: Diferença entre o coeficiente de variação de cada uma das amostras e o estímulo da frase (1).

Alguns pontos nos chamam a atenção. O primeiro deles é o fato da média geral do índice de relação (i_r) ser da ordem de -0,7% para MIDI e 1,5% para Hertz, indicando que os participantes, de modo geral, cumpriram a tarefa com um alto grau de acuidade. O segundo ponto é tendência a valores negativos: 14 *tokens*, contra 10. Por um lado, esse número pode revelar um problema de ordem prática, que é a tentativa consciente dos falantes em manter suas variações limitadas, a fim de cumprir a tarefa. Por outro lado, isso não é necessariamente um problema, quando notamos que a mesma tendência pode ser encontrada nos dados intra-repetições dos sujeitos (tabela 5.6).

Variação intra sujeito		
	MIDI	Hz
suji1	-1,1%	-4,1%
suji2	0,5%	0,3%
suji3	-	-
suji4	0,0%	0,2%
suji5	-3,9%	-9,8%
suji6	-3,4%	-5,2%
suji7	5,7%	3,0%
suji8	2,2%	6,3%
suji9	-2,8%	-5,4%
suji10	-0,8%	-4,1%
suji11	9,6%	20,6%
suji12	-0,4%	-0,7%
suji13	-	-

Tabela 5.6: Cálculo do índice de relação intra-repetições da frase (1).

A partir desta tabela podemos sugerir que esse é um fato aleatório ou pouco relacionado ao fenômeno que estudamos. A nosso ver, os valores negativo não denotam necessariamente uma tentativa consciente do falante de manter a variação de F_0 limitadas, mas, antes, uma vicissitude natural dos dados, uma vez que aparecem com regularidade em mais de um lugar e não são a regra. Aparte dessa questão, os dados da tabela 5.5 se mostram também em uma distribuição relativamente normal, já que são derivados daqueles primeiros. As figuras 5.10 e 5.11 apresentam o histograma de distribuição do índice de relação.

Figura 5.10: Histograma da diferença entre os coeficientes de variação das amostras da frase (1) e o do estímulo. Os valores estão em MIDI.

Figura 5.11: Histograma da diferença entre os coeficientes de variação das amostras da frase (1) e o do estímulo. Os valores estão em MIDI.

A partir dos dados da tabela 5.5 executamos dois testes de hipótese. O primeiro deles é um *teste-t* de amostras emparelhadas e visa avaliar se os índices de relação são significativos quando comparados com a média de F_0 das amostras. A ideia, neste caso, é verificar se os índices representam, de fato, uma relação com as amostras ou

se são algum tipo de aleatoriedade. Como se pode ver na tabela 5.7, a hipótese nula não pode ser rejeitada para um intervalo de confiança de 95%, ou seja, os índices de variação mantêm uma relação significativa com a variação de F_0 das amostras.

Teste de amostras emparelhadas				
		t	df	Sig.
Par 1	i_r amostras vs Média de F_0 (MIDI)	-34,33	23	.000
Par 2	i_r amostras vs Média de F_0 (HZ)	-17,34	23	.000

Tabela 5.7: Teste-t do índice de relação com a média da frequência fundamental da frase (1).

O segundo teste foi o de Kolmogorov-Smirnov (K-S), que é utilizado para se comparar uma determinada amostra com uma distribuição de probabilidade de referência, em geral, a distribuição normal. A aplicação desse teste visa verificar o grau de significância dos índices de relação, isto é, verificar se eles são, de fato, significativos. A hipótese nula nesse caso é a hipótese geral, que a variação de cada uma das amostras é semelhante a do estímulo. É importante ressaltar que no teste K-S os níveis de significância são determinados a partir do tamanho da amostra, dada por uma tabela de valores críticos. Assim, para uma população de 24 amostras o valor é de 0,2, tal como se pode ver na tabela 5.8.

Teste de normalidade			
Kolmogorov-Smirnov ^a			
	Estatística	df	Sig.
ir - Midi	0,093	24	,200*
ir - Hz	0,101	24	,200*

*. Este é um limite inferior da significância verdadeira.

a. Correlação de Significância de Lilliefors

Tabela 5.8: Resultado do teste Kolmogorov-Smirnov para a distribuição significativa dos dados.

Dessa maneira, temos duas “confirmações”. A primeira é a de que os índices são produtivos para a análise que propomos, ao passo que a segunda é a evidência de que os valores que obtivemos a partir da equação 5.2 são significativos. Estas análises

são gerais, isto é, avaliam o índice de relação de todos os sujeitos. Um dos problemas relacionado à essa medida é justamente a verificação de até qual valor ela é realmente significativa, ou seja, qual é o limiar aceitável do índice de relação.

Não conseguimos desenvolver um cálculo ou análise que resolvesse essa questão de forma satisfatória, de modo que tomamos isso como um próximo passo em nossa pesquisa. Entretanto, para satisfazer essa questão de forma provisória utilizamos dois recursos. O primeiro deles é a utilização de um gráfico de barra de erros, que pode ser visto na figura 5.12. O segundo foram os gráficos de dispersão apresentados nas figuras 5.13 e 5.14.

Figura 5.12: Gráfico de barra de erros do índice de relação entre estímulo e amostras da frase (1), com as duas condições de análise Hertz e MIDI.

Figura 5.13: Gráfico de dispersão comparando o índice de relação e coeficiente de variação de cada amostra da frase (1), em Hz

Figura 5.14: Gráfico de dispersão comparando o índice de relação e coeficiente de variação de cada amostra da frase (1), em Midi

A figura 5.12 demonstra que os limiares para esse grupo de dados, em específico, é da ordem de $\pm 10\%$ para a análise em MIDI e da ordem de $\pm 15\%$ para Hertz. Ao observarmos a tabela 5.5, notamos que cerca de dois ou três itens da amostra passam desses limites, porém como o limiar é previsto para 95% dos casos, este fato os torna

aceitáveis.

Já as figuras 5.13 e 5.14 revelam algo bastante óbvio para esses dados, mas que são fundamentais nessa questão: quanto maior a variação, maior será o índice de relação. Entretanto, os gráficos também demonstram que as amostras tendem a se agrupar de forma mais ou menos coesa na faixa que vai de -5% a +5% para MIDI e de -10% a +10% para Hertz. De modo geral, é possível afirmar que até o momento nossas hipóteses se mostram válidas. Em primeiro lugar, os falantes são capazes de intuir os alvos topológicos e implementá-los com um grau de acuidade da ordem dos 5% ou 10%, a partir do estímulo.

Como estes alvos são distensos, ou seja, não demandam um nível semântico, é necessário supor que a habilidade de captar esses alvos e implementá-los tal como solicitado é baseada em primeiro em alguma habilidade cognitiva do falante. A nosso ver, essa habilidade é justamente a de coordenar um nível simbólico pressuposto com um nível de implementação. Discutiremos esses dados com mais profundidade no fim deste capítulo e na Conclusão. Por ora, nos voltaremos à análise dos alvos pontuais.

5.4 Segunda condição: alvos pontuais

No Capítulo 3 definimos que os alvos pontuais são aqueles que representam um efeito perceptual significativo dentro de uma elocução, ou seja, que transmite algum significado pela sua valência. Partindo do modelo de Ferreira-Netto (2008), o efeito citado só é possível quando estes tons ultrapassam os limiares de diferenciação tonal (MEDEIROS; MARTINS, 2014), estabelecidos como +3 e -4 para o Português Brasileiro e mantêm uma duração breve, ainda não determinada.

A nosso ver, o alvo pontual não corresponde a um espaço específico, como é o alvo topológico e, sim, a um ponto, o qual também não é exato como se espera pela definição matemática do termo. Para nós, esse ponto se divide em dois. O primeiro deles é um ponto específico, determinado pelo limiar de diferenciação tonal, de modo

que esse valor seria o alvo pontual *default*. Assim, em qualquer elocução que se pretenda transmitir um conteúdo semântico por meio da entoação, o falante precisa ao menos atingir esse limiar.

O segundo ponto é um conceito mais genérico, pois não vemos a possibilidade de estabelecer um limite máximo após o limiar, se não pela capacidade fisiológica subjacente à produção e percepção da fala. No entanto, ele não deixa de ser pontual, pois tem uma duração intrínseca que é necessariamente breve. Dessa maneira, o alvo pontual, para esta análise, dever ser um valor além dos limiares ou exatamente no limiar.

Nesta questão destacamos um problema de ordem metodológica. Uma das maiores dificuldades práticas é encontrar proeminências que estejam aquém do limiar inferior, ou seja, a cerca de -4 semitons da média geral. Em Martins (2009), demonstramos que os falantes utilizam-se desse recurso em narrativas orais, a fim de se dar ênfase a um ponto específico do enredo. Em geral, trata-se de uma sentença ou de uma unidade mais longa que uma palavra ou locução substantiva. No caso de unidades menores, como sílabas, não fomos capazes de encontrar este tipo de proeminência negativa.

Sendo assim, nesta tese optamos por analisar unicamente a ênfase positiva, aquela que ultrapassa o limiar superior de diferenciação. Nossa opção se justifica pela razão prática acima descrita e por ser ela a mais produtiva no PB, de modo que pretendemos retomar essa questão em trabalho futuro. Com relação à escolha da frase, optamos usar uma frase interrogativa, pois entendemos que esse tipo de proeminência é a mais funcional no PB. Além disso, em algumas análise preliminares observamos que a ênfase em uma sílaba pode induzir a implementação fonética do falante, de modo que optamos por deixar esse experimento para uma segunda fase de nossa análise.

Por fim, nesta segunda condição esperamos que o participante note a variação e a reproduza em sua elocução. As evidências que sugerem que o falante é capaz de executar essa tarefa perceptual são recorrentes na literatura (T'HART, 1981; T'HART;

COLLIER; COHEN, 1990; CONSONI, 2011; MARTINS, 2013; MEDEIROS; MARTINS, 2014; ROSA, 2015; GARCIA, 2015). Desta forma, nosso segundo objetivo é analisar como os falantes implementam os alvos pontuais a partir da percepção de uma proeminência. Para esta finalidade utilizaremos a frase (2), apresentada no início deste capítulo. A seguir, faremos uma breve descrição das características acústicas mais importantes.

(2)[O Domenico entrar aqui ...] p_i [como assim?] p_f

Características vocais:

- Frequência fundamental média: 193,37Hz;
- Duração total: 2,19s;
- Frequência média da porção inicial (p_i): 153,33 Hz;
- Duração da porção inicial (p_i): 1,56;
- Frequência média da porção final (p_f): 266,5 Hz;
- Duração da porção inicial (p_i): 630 ms;
- Distância, em semitons, da virada tonal ao fim (ponto 4): 6st

A figura 5.15 apresenta a curva tonal para a frase 2. É notório o aumento da frequência fundamental na porção final, em especial, no trecho “como assim?”. Calculamos um acréscimo de 5 semitons em um intervalo de 630ms, o que representa um aumento de cerca de 8 semitons por segundo. Números semelhantes àqueles encontrados por Xu e Sun (2002), que previam o aumento de 1st para cada 100ms.

Figura 5.15: Variação de frequência fundamental para a frase-estímulo que utilizamos em nosso teste dos alvos pontuais.

5.4.1 Análise descritiva

Tal como fizemos na demonstração anterior, entendemos que seja importante apresentar os dados e a tendência subjacente a eles, uma vez que estamos lidando com medições indiretas. A tabela 5.9 apresenta os valores de média e o desvio padrão da frequência fundamental de todas amostras e do estímulo da frase (2), já na tabela 5.10 se encontram as estatísticas descritivas desses dados.

Nota-se, como na primeira condição, uma leve dispersão da média entre sujeitos. Observando os valores de assimetria (positiva) e curtose (negativa) notamos que a massa de distribuição dos dados está ligeiramente deslocada para esquerda e mais concentrada em torno da média do que dispersa pelas caudas. Na primeira condição havíamos notado o efeito contrário, causado pelo desbalanceamento entre vozes masculinas e feminina.

A distribuição dos dados pode ser vista com mais detalhes nas figuras 5.16 e 5.17. .

Valores de Frequência Fundamental				
	Média - MIDI	Média - Hz	DvP - MIDI	DvP - HZ
<i>Estímulo</i>	53,9	193,37	5,6	58,20
suji	59,2	259,06	5,8	78,17
repsuji	54,3	215,88	10,5	95,74
suji2	59,2	251,87	2,0	30,10
repsuji2	60,9	282,06	3,0	58,96
suji3	49,4	149,49	5,5	43,56
suji4	59,3	255,52	3,0	46,17
repsuji4	57,1	260,71	11,1	128,10
suji5	49,2	165,14	10,6	92,25
repsuji5	52,2	181,50	7,4	67,16
suji6	62,2	302,29	3,6	61,92
repsuji6	64,1	341,02	3,8	73,76
suji7	51,1	175,59	7,7	111,80
repsuji7	49,2	142,80	3,8	31,92
suji8	58,0	247,98	6,0	90,10
repsuji8	54,9	205,77	6,2	68,62
suji9	46,6	131,52	7,5	53,87
repsuji9	53,0	204,81	10,1	117,33
suji10	46,7	123,75	4,1	27,10
repsuji10	47,9	137,67	5,7	42,92
suji11	51,0	180,23	10,3	83,61
repsuji11	56,9	231,24	6,5	61,21
suji12	52,8	187,54	7,8	72,54
repsuji12	54,7	208,68	8,0	73,25
suji13	46,8	128,50	6,0	38,85
Geral	54,0	206,6	2,6	27,1

Tabela 5.9: Valores de média geral e desvio padrão da frequência fundamental das UPEs de cada uma das amostras da frase (2). Os valores estão em MIDI e Hertz

Figura 5.16: Histograma de médias da frequência fundamental da frase (2), os valores estão em MIDI.

Estatísticas Descritivas		
	F_0 - MIDI	F_0 - Hz
Média	54,02	206,56
Mediana	53,94	204,81
Desvio-Padrão	5,08	57,98
Variância	25,79	3361,84
Assimetria	,216	,455
Curtose	-,918	-,369
Mínimo	46,63	123,75
Máximo	64,06	341,02

Tabela 5.10: Estatísticas descritivas da frequência fundamental das amostras e do estímulo da segunda condição experimental. $n=25$

Figura 5.17: Histograma de médias da frequência fundamental da frase (2), os valores estão em Hertz.

A nosso ver, a dispersão se explica pelas próprias características da produção da voz. Um maior conjunto de dados em torno da média deve-se, em partes, a dificuldade natural de se produzir tons mais altos. Dessa maneira, é natural que os falantes implementem esses alvos em uma faixa de banda relativamente estável. Além disso, a tendência em torno da média pode justificar outra tendência, a saber: que os falantes produzem as proeminências interrogativas de uma forma semelhante, ou seja, a uma altura determinada, independentemente do contexto. Com relação à segunda hipótese precisamos avaliá-la em outros testes.

Já a assimetria se explica pelo dispêndio de energia necessário para a produção de tons agudos. Como citado no capítulo 3, quanto mais alto o tom, maior será o esforço envolvido pelos falantes para implementá-lo. Entendemos que possa haver duas hipóteses explicativas. A primeira delas é a de que alguns falantes reduzem o seu tom médio anterior à proeminência para que essa se destaque e o esforço seja menor. Desta maneira, é natural que os registros de tons mais altos se encontrem próximo ao limiar de diferenciação tonal e numa faixa de frequência mais à esquerda no histograma. A segunda hipótese é a de que alguns falantes marcam de forma considerável suas proeminências, enquanto a tendência natural é de pronunciá-las apenas na medida do necessário. Veremos mais a frente que a segunda explicação parece ser a mais razoável.

O segundo passo de nossa análise foi analisar os coeficientes de variação e o índice de relação, mesmo procedimento adotado na primeira condição. Novamente, a nossa proposta era avaliar se a tarefa foi executada a contento e, se sim, qual o grau de acuidade. Tendo em vista que a mesma análise foi explicada neste capítulo, vamos nos ater a demonstrar o resumo dos valores (Tabelas 5.11 e 5.12), uma vez que o processo de análise foi o mesmo. As tabelas com os valores dos coeficientes e dos índices de relação podem ser consultadas nos Anexos II e III.

Resumo - Coeficientes de Variação		
	CofVar F_0 - MIDI	CofVar F_0 - Hz
Média	12,20	34,11
Mediana	11,39	31,18
Desvio-Padrão	5,22	13,13
Variância	27,25	172,42
Assimetria	,121	,587
Curtose	-,873	-,163
Mínimo	3,32	11,95
Máximo	21,57	63,67

Tabela 5.11: Análise descritiva dos coeficientes de variação das amostras e estímulo da frase (2). Os valores podem ser vistos no Anexo II

Resumo - i_r		
	i_r - MIDI	i_r - Hz
Média	1,92	4,18
Mediana	1,30	2,17
Desvio-Padrão	5,32	13,39
Variância	28,28	179,17
Assimetria	,074	,541
Curtose	-,953	-,292
Mínimo	-7,04	-18,15
Máximo	11,21	33,57

Tabela 5.12: Análise descritiva dos índices de relação entre amostras e estímulo da frase (2). Os valores podem ser vistos no Anexo III

É importante destacar que na segunda condição o coeficiente de variação do estímulo era de 10,4% para MIDI e 30,1% para Hz, ao passo que a média geral ficou em torno dos 12% e 34,11%, o que nos permite dizer que os participantes, ao menos de modo geral, foram capazes de cumprir a tarefa como esperávamos. Destaca-se o caso dos sujeitos 1 e 4 (19,4% MIDI; 44,3% Hz), 5 (21,6%;55,9), 9 (19,1%;57,3%) e 11 (20,2%;46,4%), que tiveram um coeficiente de variação da ordem de 20%.

Para evitarmos qualquer tipo de divergência utilizamos um gráfico de barra de erros 5.18 para verificar os limites verticais aceitáveis. O resultado sugere que se utilizarmos 1,96 Desvio-Padrão como referência, estes dados podem ser aceitos, embora o façamos com ressalvas, uma vez que ainda não calculamos o limite aceitável. O restante dos valores manteve-se na faixa esperada para um intervalo de confiança de 95%, entre 5% e 15% para MIDI e entre 20% e 35% para Hz.

Figura 5.18: Barra de erros que determina os valores aceitáveis do coeficiente de variação.

Ainda com relação à tabela 5.10, ao observar os valores de assimetria e curtose notamos que a distribuição é ligeiramente mais assimétrica na análise em Hertz do que em MIDI. A explicação se dá pelo do efeito de relação unívoca apresentado neste

capítulo, ou seja, da quantidade de informação contida em cada uma das unidades. Ainda assim, essa ligeira assimetria não sugere aleatoriedade na execução da tarefa, de modo que os falantes foram capazes de deduzir os alvos desta condição e implementá-los de forma produtiva, ao menos teoricamente.

No questão da acuidade, medida pelo índice de relação (i_r), podemos notar pelo valor da média apresentado na tabela 5.12 que os participantes não tiveram a mesma performance do que aquela encontrada no teste anterior (0,7%). A nosso ver, a principal fonte de divergência é a porção final da frase (2) e, por consequência, o problema da massa de distribuição ligeiramente deslocado à esquerda. Para verificar essa hipótese comparamos o índice de relação da porção inicial (P_i) com o índice da porção final (P_f) da frase (2). Os valores podem ser vistos na tabela 5.13:

Índice de relação			
$P_i - I_r$ MIDI	$P_f - I_r$ MIDI	$P_i - I_r$ Hz	$P_f - I_r$ Hz
1,29%	8,93%	5,68%	23,16%
6,00%	8,2%	13,30%	20,7%

Tabela 5.13: Valores do índice de relação para as porções inicial e final da frase (2).

É notório que os valores do i_r na segunda parte da sentença (6,0%) são sensivelmente maiores do que na primeira parte (1,29%). Explicamos. O limiar de diferenciação tonal pressupõe um valor de virada a partir do qual o sujeito pode produzir e perceber diferenças semânticas. Em nosso entender, esse valor não possui um limite específico, isto é, a partir deste ponto de virada o falante pode usar um conjunto finito de valores de frequência para marcar as proeminências, mas sem uma especificidade bem clara, o que afeta diretamente a acuidade na execução dessa tarefa.

Obviamente a altura desta proeminência pode transmitir um conteúdo ou outro, de modo que o falante precisa de algum tipo de monitoramento para que o significado pretendido seja, de fato, transmitido. Entretanto, parece nos que o monitoramento

pressuposto no nível topológico é maior quando comparado ao dessa condição, uma vez que o espaço entoacional é relativamente limitado, o que explica a diferença entre os valores encontrados nas duas partes da sentença. É importante ressaltar que não compreendemos esse monitoramento como uma atividade consciente e, sim, uma atividade linguístico-cognitiva dentro do que Ferreira-Netto (2008) define como “tom médio” em seu modelo.

Como se pode ver, de maneira geral, os falantes foram capazes de executar a tarefa proposta com algum grau de acuidade. O que pretendemos demonstrar, a seguir, é a relação entre os alvos pontuais e topológicos, utilizando a implementação dos primeiros como referência para a nossa análise.

5.4.2 Análise do alvo pontual

Para esta etapa da análise utilizaremos unicamente os valores em MIDI. Como estamos analisando alturas relativas, a frequência em medida Hertz retorna uma diferença adimensional, isto é, não somos capazes de identificar o quão relativo é aquela diferença quando comparamos esses valores em Hertz. Por exemplo, uma medida de 15hz não é tão específica como uma comparação de 3 st. Dito isso, a tabela 5.14 apresenta os valores de média da porção inicial e da porção final da frase (2) e tabela 5.15 as suas estatísticas descritivas.

Média da frequência fundamental			
	Média Início	Média Final	Diferença
<i>Estímulo</i>	50	60	11
suj1	59	59	0
repsuj1	54	51	-2
suj2	59	60	0
repsuj2	59	59	0
suj3	49	54	6
suj4	60	56	-4
repsuj4	56	56	0
suj5	46	60	14
repsuj5	50	53	3
suj6	60	65	6
repsuj6	63	66	3
suj7	48	52	3
repsuj7	48	50	2
suj8	53	65	11
repsuj8	54	56	2
suj9	44	57	13
repsuj9	47	58	11
suj10	47	50	2
repsuj10	49	52	3
suj11	53	53	-1
repsuj11	57	59	2
suj12	51	56	5
repsuj12	54	50	-3
suj13	46	49	3

Tabela 5.14: Média da frequência fundamental para as duas sentenças. Os valores estão em MIDI

Estatísticas Descritivas			
	Média P_i	Média P_f	Diferença
Média	52,64	56,24	3,68
Mediana	53	56	3
Desvio Padrão	5,35	4,90	4,91
Variância	28,57	24,02	24,14
Assimetria	0,26	0,39	0,73
Curtose	-1,08	-0,58	-0,20
Intervalo	19	17	18
Mínimo	44	49	-4
Máximo	63	66	14

Tabela 5.15: Estatísticas descritivas da frequência fundamental das porções inicial e final da frase (2). $n=25$.

O primeiro ponto fundamental é o fato da segunda porção estar, em média, 4 st acima da inicial e 5 semitons acima da frase (1), tal como sugerido pelo modelo de Consoni (2011). Nota-se que a condição P_f tem uma ligeira assimetria à esquerda, explicada pela improbabilidade de que alguns tons à direita sejam entoados pelos falantes, pelas limitações fisiológicas. Outro ponto de atenção é fato de em alguns casos a diferença ter sido negativa ou igual a zero. Como as implementações têm algum grau de oscilação, é natural que a média dos trechos analisados seja diferente em um determinado nível, uma vez que a frase não é entoada totalmente no mesmo tom.

Para confirmar a hipótese que a porção final é distinta da inicial, executamos um *teste-t* com essas duas condições, o qual retornou uma significância de $p = 0,001$, ou seja, os falantes necessariamente marcam a diferença tonal entre essas duas condições. Para avaliar se os valores de diferença são corretos, executamos um teste não paramétrico de *Kolmogorov-Smirnov*, o qual sugere que a distribuição é normal e significativa nos dois casos ($p=0,200$) e normal, mas não significativa no caso da “Diferença”. Para avaliarmos qual valor poderia ser discrepante, executamos um teste de *outliers*. Utilizando *bootstrapping* ($n = 5000$), detectamos o valor de 13,7 como limite máximo aceitável, caracterizando o caso 9 como um *outlier*.

No entanto, de modo geral é possível dizer que a porção da sentença que contem o alvo em si é diferente da primeira parte com uma distribuição normal. Dessa forma, para termos uma noção mais exata da implementação dos alvos é importante que consideremos apenas os tons que tenham ultrapassado e/ou sejam igual ao limiar de diferenciação tonal, o que Ferreira-Netto (2008) denomina *Foco/Ênfase* e não a diferença entre a parte inicial e final da frase (2).

Para esse cálculo utilizamos a seguinte fórmula:

$$A_p = ldt(z) - t(z) \quad (5.3)$$

Onde: A_p - Alvo tonal pontual - Foco/Ênfase;

ldt - valor limiar de diferenciação tonal superior no instante;

$t(z)$ - tom no instante (z)

A partir desse cálculo obtivemos os valores apresentados na tabela 5.16. As estatísticas descritivas podem ser vistas na tabela 5.17:

Frequência Fundamental de F/E	
Foco/Ênfase	
<i>Estímulo</i>	
su1	3
repsu1	5
su2	3
repsu2	3
su3	6
su4	3
repsu4	3
su5	7
repsu5	4
su6	4
repsu6	2
su7	11
repsu7	-
su8	3
repsu8	8
su9	6
repsu9	7
su10	10
repsu10	6
su11	5
repsu11	2
su12	5
repsu12	6
su13	3
Média	5

Tabela 5.16: Média dos valores de frequência fundamental de Foco/ Ênfase para a porção final da frase (2).

Utilizando um teste-t de amostra única definimos que o conjunto de dados descrito tem significância estatística, isto é, ele não é totalmente aleatório para um IC de 95% ($p < 0,05$). A partir disso, duas discussões parecem ser fundamentais nesse ponto.

Estatísticas	
Foco/Ênfase	
N	24
Ausente	1
Média	5
Mediana	4,5
Desvio Padrão	2,41
Variância	5,82
Assimetria	1,02
Curtose	0,61
Intervalo	9
Mínimo	2
Máximo	11

Tabela 5.17: Estatísticas descritivas da frequência fundamental das porções inicial e final da frase (2). $n=23$, o caso 7 não pode ser calculado

A primeira delas é a distribuição dos dados, um teste de *Kolmogorov-Smirnov* revela que embora eles não sejam aleatórios, eles não estão distribuídos em uma normal ($p = 0,012$), o que vale dizer indiretamente que eles podem ser aleatórios. O valor de assimetria da tabela 5.17 também sugere uma concentração à esquerda, ou seja, mais próximo ao limiar de distinção tonal. Nossa interpretação, nesse caso, é a de que os falantes tendem a atingir o alvo *default*, o LDT, e como não se trata de uma aproximação exata, eles tendem a o fazer na medida do aceitável.

Para explicar essa possibilidade executamos uma análise de variância, a qual demonstra que cerca de 58,1% dos dados encontram-se alocados até a média acima do limiar (4,9 semitons), esse resultado por ser visto no histograma da figura 5.19. Como a rotina *ExProsodia* desconsidera aqueles dados que estão diretamente sob o limiar ou abaixo dele (para $\forall A_{p,se} \leq 0 = \#$) precisamos supor que os valores iguais a zero não afetam a média geral, a qual poderia ser menor. Essa é questão secundária no momento que envolve algum debate em torno das propriedades matemáticas da distribuição, de modo que não nos dedicaremos a essa discussão neste momento.

A segunda discussão refere-se à aleatoriedade dos dados. A noção de aleatoriedade envolve duas dimensões. Em um delas os dados necessariamente não se encaixam

Figura 5.19: Histograma dos valores médios de F/E.

em nenhuma distribuição e não podem ser previsto por um modelo, por exemplo, qual será o próximo resultado de um lançamento de moedas? É possível prever as chances de se dar cara ou coroa (50% para cada uma delas), mas é bastante improvável que se defina com exatidão qual será o resultado. Na segunda definição trata-se um de elemento ou de um conjunto de elementos que não pertence àquela distribuição. Por exemplo, ao se lançar um dado e após obter uma sequência de números pares, se obtém um ímpar, este último representa uma aleatoriedade, que diferente da primeira tem uma ocorrência previsível.

Os dois fenômenos representam eventos aleatórios, um orientado à probabilidade de ocorrência e outro orientado à distribuição. No primeiro caso, o evento aleatório é randomizado pelas características do sorteio em que ele está inserido, no segundo sua aleatoriedade é definida pelo conjunto em que ele se encontra. No caso da língua é normal considerarmos que todos os eventos são orientados à distribuição pela simples razão de que um fato linguístico não pode ser desgarrado do seu contexto. Assim, após um artigo definido masculino é bastante provável que o próximo termo sintático

seja um substantivo também masculino.

No caso dos valores de F/E essa discussão é importante, pois é possível prever pela média e pelo desvio-padrão qual a tendência falantes seguirão para produzir os tons na faixa do Foco/Ênfase. Pela média, nota-se que em grande parte das vezes, esse valor é da ordem de 4 st, um cálculo mais apurado dos intervalos de confiança para 95% sugere que produções aquém de 3 st ou além de 6st na faixa do F/E podem ser consideradas aleatórias. No entanto, valores aquém não podem ser considerados aleatórios, uma vez que eles representam eventos significativos dentro do aspecto linguístico, sendo assim, vamos nos ater à questão do limite superior.

Para uma correção do valor superior recorreremos a um *bootstrapping* de $n = 1000$, que retornou um ajuste de cerca de 10% para cima, ou seja, o limite superior de confiança estaria em torno de 7 st. Pelo cálculo de variância explicada em ao menos de 17% dos casos teríamos um efeito de aleatoriedade, em que os falantes implementariam seus alvos pontuais de forma randômica e totalmente imprevisível, ao passo que 83%. esses valores flutuariam um intervalo que vai de > 0 a 7st. Assim, em uma população com 100 amostras existe a tendência de que cerca de 17 casos sejam totalmente aleatórios, uma vez que não é possível determinar se estão nesse intervalo ou além dele.

O *box-plot* da figura 5.20 demonstra o comportamento dos dados nessa faixa, além de apresentar um *outlier*, sugerindo que, de fato, os valores acima dos 6 ou 7 são aleatórios.

O que o teste de *Kolmogorov-Smirnov* combinado com o *teste-t* nos sugerem é que a implementação dos alvos pontuais na faixa do F/E segue um padrão randômico explicável. Como citado, entendemos que os valores em torno da média geral representam o que chamamos de *implementação na medida do aceitável*, ao passo que as demais, aquelas que estão além desse limite, consideramos *implementações tonais aleatórias* e outros são simplesmente *outliers*. Neste ponto, a discussão sobre a aleatoriedade se torna importante, pois as implementações tonais aleatórias são aleatórias quando ana-

Figura 5.20: *Box-Plot* que demonstra os valores de limite

lisadas pela distribuição, mas nunca são aleatórias quando observadas como eventos linguísticos, uma vez que são esperadas pelos falantes.

5.4.3 Discussão dos dados

Como fizemos várias notas ao longo do capítulo para explicar os valores estatísticos que fomos encontrando, optamos por restringir essa seção a uma explicação mais geral e a um detalhamento mais teórico ao fim dessa tese. De modo geral, nossas duas hipóteses se confirmam. A condição 1 e 2 demonstram que o falante é capaz de intuir os alvos tonais, sejam eles topológicos ou pontuais e são capazes de reproduzi-los de forma produtiva na fala.

Pela análise da distribuição nota-se que os falantes também são capazes de intuir o espaço entoacional, limitado pelos limiares de diferenciação, bem como são capazes de implementar suas variações tonais dentro deste espaço de maneira significativa e não aleatória. No mais, o inverso também é verdadeiro, pois os dados sugerem que os participantes foram capazes de detectar os alvos pontuais presentes além do espaço entoacional e implementá-los de forma produtiva, embora ligeiramente aleatória.

Essa diferença sugere que os falantes possuem uma acuidade específica para a manifestação da frequência fundamental, de modo que aquelas implementações que se pretendem distintas não são necessariamente desvios de um estado neutro, causado por um evento fisiológico e, sim, podem ser pensadas como realizações de estruturas virtuais e cognitivas adquiridas como um aspecto da linguagem. Neste sentido, estes resultados são de suma importância para nós, pois justificam a distinção entre sistemas contínuos e descontínuos que propomos no 4 e que nos servirão de base para explicar os dados de fala emotiva.

Com relação às questões de estatísticas, pode se observar que os dados gerais, isto é, o das elocuções dos falantes sem o cálculo do coeficiente de variação, mantêm uma forte correlação com o estímulo, ao menos uma relação de forma significativa, no que tange a distribuição. Obviamente, essa é uma análise derivativa, supostas pelos valores da média. Entendemos que o próximo passo seja uma análise ponto-a-ponto, utilizando cálculos relacionados à correlação, no entanto, teremos de adotar outro

design experimental.

Como dito acima, as questões de distribuição são fundamentais, uma vez que nossa sugestão é a de que os alvos topológicos são tendências a uma determinada implementação em um espaço de variação e não à execução exata de uma curva. Dessa forma, a maior preocupação na análise não é a relação um para um e, sim, a análise da implementação de algumas formas que consideramos prototípicas, como é o caso da fala emotiva. Nesse sentido, nossa proposta é poder utilizar estes métodos para detectar essas formas entoacionais prototípicas na fala emotiva. Para isso, a seguir apresentamos a segunda parte de nossa tese que envolve o estudo de caso das emoções na fala atuada.

6 Emoções

6.1 Introdução

Há um conceito sempre presente na cultura ocidental quando se trata das emoções: elas são a conexões do homem com um estado primitivo da natureza. À parte de se ser evolucionista, criacionista ou qualquer outra denominação, a hipótese de que as emoções mantêm uma relação forçosa da racionalidade com um estágio evolutivo anterior ou com um espírito animalesco criado por algum *design* inteligente não só é aceito, como foi estudado ou demonstrado ao longo dos tempos. No mais, é possível afirmar que parte da cultura ocidental se edificou sobre esta interpretação, compreendendo as emoções como mera manifestações fisiológicas, respostas do corpo a um determinado estímulo externo, sem conexões com a cognição. Em suma, as emoções foram lidas e interpretadas como uma parte nebulosa do pensamento humano, ou mesmo externa a elas, pouco dotadas de sentido e responsáveis por cegar a razão. A nosso ver, essa interpretação não é totalmente correta e o que iremos apresentar a seguir é um breve histórico desta questão, levando em conta que ela pode vir a iluminar nossa compreensão sobre um tema tão vasto.

Neste ponto se faz necessário um aparte, como se verá a seguir, utilizaremos ora o termo *paixões*, ora o termo *emoções*. Explica-se: grande parte dos filósofos adotavam os termos **paixão** ou **afeição** para explicar o que entendemos hoje por emoções. Dixon (2003) argumenta que a mudança na terminologia teve um direcionamento ideológico: para se afastar da visão clássica e, principalmente, da visão escolástica, os estudiosos

do séc. XIX, baseados em universidades seculares, optaram por adotar o termo *emoção*. Para pensadores como Humes, Spencer, Bain e Darwin o termo *emoção* eliminava do seu interior a ideia de alma ou espírito e limitava o objeto a ser analisado, além de permitir a eles a criação de uma estrutura de análise. Em certo sentido, tratava-se de uma mudança de paradigma, ao mesmo tempo que representava uma separação dos ideais cristãos dos ideais da ciência do século XIX. Utilizaremos os termos em sinonímia, tentando preservar a distinção acima descrita.

6.2 Razão e emoção, mente e corpo

6.2.1 A contribuição grega, estoica e tomista

As raízes da separação entre emoção e razão são tão difusas quanto o próprio pensamento ocidental, porém, é um fato aceitável que as primeiras impressões sobre esta divisão foram feitas por Sócrates e reproduzidas por Platão. Não apenas isso, esta tradição também insere a questão: seriam as emoções produtos da mente ou do corpo? Se são da mente, morada da razão, significa que essa não é preponderante, se são do corpo significa que podem sobrepujar a razão e apagá-la. Nesta perspectiva, as paixões passam a fazer parte de um sistema complexo, que a dialética socrática se vê incapaz de resolver senão pela exclusão das paixões do seu sistema de análise e a inclusão de suas manifestações em um sistema ético. Obviamente, esta leitura será retomada ao longo dos séculos e tornará a oposição *razão* e *emoção* um valor cultural.

Em *Fedro* (*Fedro*, 237b e ss.)¹ Sócrates argumenta que as paixões não existem e são, na realidade, a manifestação de um desejo (*epithumia* corporal). Na visão socrática, estes desejos seriam de duas formas: (i) inatos ou de (ii) opinião. Como se pode imaginar no primeiro caso trata-se de uma busca inata pela satisfação, ao passo que o segundo é um desejo por aquilo que é melhor, isto é, não se trata de um desejo em si, mas unicamente do desejo pelo que se considera melhor. Quando se

¹Ver Platão (2010)

controla a opinião por meio da razão e se atinge o melhor temos um estado ideal, chamado de autocontrole, mas se nos deixamos levar pela opinião rumo aos prazeres inócuos, estamos em excesso (*hubris*). "Se o desejo por alimento prevalece sobre a razão conducente ao melhor e sobre os demais desejos, é chamado de glotonice e confere ao seu detentor o nome de glutão"(Fedro, 238a)² Sendo assim, quanto mais se vive em excesso, mais distante se está da razão e mais se está condicionado às paixões, um estado que deve se evitar dentro do sistema de pensamento socrático.

Como se nota a relação entre paixões e a razão está forçosamente ligada ao conceito de *desejo*, de tal modo que as paixões seriam ativadas por esse elemento. É importante observar que a ideia de desejo trazia (como ainda traz) por si só a conotação de algo ligado aos impulsos naturais e fortemente relacionado ao corpo, a despeito da mente. Em sua República, Platão vislumbra um sistema ético/moral fortemente influenciado pela oposição lógica entre mente e corpo. Para explicar seu processo, o filósofo dividiu a alma humana em três partes: (i) a parte concupiscente (*epithymía*), aquela que age por necessidade, (ii) a parte irascível (*thymós*), que age pela ira e por impulso e (iii) a parte racional (*lógos*), aquela orientada pela lógica (*República*, IV, 435a e ss.)³.

A primeira delas seria relacionada diretamente à cobiça, aos desejos carnis, as necessidades vitais de um modo geral e a tudo que é relacionado com o instinto de sobrevivência e, por este motivo, é irracional, quase animalesca. A razão seria responsável por controlar esses sentimentos primitivos e canalizá-los para a ordem. Ainda assim, um homem estaria disposto a sobrepujar a razão e enfrentar seus desejos primitivos se fosse acometido pela cólera, evocando sua parte irascível. É preciso destacar que Platão via a separação corpo e alma (ou mente e corpo) como uma separação lógica, isto é, como um fato dado, cuja a demonstração era sua própria existência. Nota-se que no sistema platônico a alma seria composta por duas partes associadas às emoções e vistas como elementos da natureza e uma terceira, a razão,

²*Idem*

³Ver Platão (1965)

que era vista como uma dádiva dos deuses ao homens, de tal modo que era ela a responsável pelas virtudes⁴. Dentro deste sistema um estado ótimo seria em aquele que o indivíduo é capaz de utilizar a razão para reger as demais partes da alma e manifestar cada uma delas em sua justa medida ou ainda a sua completa aniquilação, como apresentado na Alegoria da Carruagem descrita em *Fedro* (PLATÃO, 2010).

Platão compreendia que alguns estados de alma (ou paixões) seriam elevados, pois seriam a demonstração do controle da razão sobre as paixões (*República*, IV, 440a e ss.). A coragem, por exemplo, seria resultado do domínio da razão sobre a cólera, bem como o conhecimento que seria o domínio da razão pela própria razão. Contudo, é preciso destacar que para Platão esse sistema tripartite não era uma tentativa de se explicar a relação mente/corpo ou descrever as paixões e, sim, a tentativa de se desenvolver um sistema moral a partir da separação alma-corpo, onde a razão seria parte da alma e as paixões seriam uma tentativa de domínio do corpo sobre a alma. Dentro deste sistema o homem virtuoso seria aquele que ascende ao ponto onde todas as suas atitudes são reguladas pela razão, ao passo que o homem com vícios é aquele regido pelas outras duas dimensões emocionais, ou seja, por impulsos corporais que dominariam sua alma. Neste caso, o primeiro é digno de governar e fazer parte da República, posto que está acostumado a justiça, valor central no sistema platônico, ao segundo cabe um papel secundário na sociedade, visto que trata-se de um pária.

Ligeiramente distanciado da hipótese de seu mestre, Aristóteles conseguirá sugerir um modelo que agrupe as paixões como parte do corpo e parte da alma, em um modelo contínuo. Ao assumir que as emoções são inerentes ao ser humano e expressas pelo corpo, ele foi capaz de estabelecer um corolário de modalidades de ação a partir da inferência de que sempre existiriam condições internas ao indivíduo (*páthos*) capazes de levá-lo a agir (*práxis*), ou seja, estados mentais que mobilizariam o corpo e

⁴É preciso destacar que a divisão da alma em três partes tinha uma correlação com as suas três categorias de cidadãos: (i) a classe dominante, formada por sábios e filósofos, (ii) a classe média formada por militares e (iii) a terceira classe formada por agricultores e negociantes. Os primeiros deveriam ser norteados pelo espírito, o segundo pela espírito irascível, ao passo que o terceiro deveria ser guiado pela concupiscência

modalidades do corpo que mobilizaram estados mentais (*Retórica*, 1378a19-21)⁵. Em certo sentido, à parte da descrição do *humour* de Hipócrates, Aristóteles foi um dos primeiros pensadores a compreender a emoção como um mecanismo de ativação da ação, uma avaliação cognitiva de um determinado objeto que leva a manifestação de uma determinada emoção em última instância.

A diferença central entre Platão e Aristóteles está no fato de que o primeiro entende que as paixões são totalmente vícios à alma, enquanto o segundo compreende que um grau de sensibilidade a elas, permite-nos compreender melhor o universo à volta. Em *Ética a Nicômano* lê-se:

Ora, nem as virtudes nem os vícios são paixões, porque ninguém nos chama bons ou maus devido às nossas paixões, e sim devido às nossas virtudes ou vícios, e porque não somos louvados nem censurados por causa de nossas paixões (o homem que sente medo ou cólera não é louvado, nem é censurado o que simplesmente se encoleriza, mas sim o que se encoleriza de certo modo); mas pelas nossas virtudes e vícios somos efetivamente louvados e censurados.

(*Ética a Nicômano*, II, 5 e ss.)⁶

Para Aristóteles corpo e alma eram um *continuum* antes de serem categorias distintas e inequívocas, como as via Platão. Ainda que fossem categorias lógicas, a doutrina aristotélica era capaz de explicar a distinção, por meio dos princípios de *forma* e *matéria* e do pressuposto do *hilemorfismo*. A partir deste dualismo, definiu-se que todas as coisas naturais teriam essas propriedades, a matéria seria aquilo que é comum a todos os corpos, ao passo que a forma é aquilo que dá identidade a essa massa disforme, deste modo não existe matéria sem forma, nem forma sem matéria e essa a união inequívoca dá-se o nome de hilemorfismo (*Metafísica*, I, 983a e ss.)⁷. Para Aristóteles a alma é a forma, aquilo que oferece vida ao corpo que a possui apenas potencialmente, por ser matéria. As paixões da alma estão profundamente enraizadas no corpo, isto é, as paixões são exclusivamente do domínio corporal e apenas os

⁵Ver Aristóteles (2005)

⁶Ver Aristóteles (1991)

⁷Ver Aristóteles (1984)

reflexos delas são sensíveis à alma. A alma em si, a parte racional, é incapaz de senti-las e só o faz porque é inseparável do corpo, de tal modo que elas se manifestam no corpo: sudorese, aceleração dos batimentos e etc.

No excerto acima citado nota-se que Aristóteles também coloca a manifestação das paixões como parte principal de seu sistema ético, enquanto Platão colocava as paixões *per se* como este elemento. Vernant e Vidal-Naquet (2005) destacam, por exemplo, o conceito aristotélico de *proaíresis*, um estado máximo de consciência, onde a execução de uma determinada ação é deliberadamente feita pelo sujeito a partir da razão, algo improvável para os animais ou crianças, uma vez que eles não são dotadas desta faculdade. Neste caso, a vontade, ou seja, a opção deliberada de se executar uma ação, a despeito da razão e a favor do prazer das paixões é em última análise, um sentimento primevo e distante da ética aristotélica.

Neste sentido, fica evidente, com algumas diferenças particulares de conceptualização, que as emoções tomam uma parte central no desenvolvimento da ética e moral Socrático-Aristotélica. Se torna também evidente que as paixões foram relegadas, de uma maneira geral, para categorias menores deste sistema, opostas à razão e compreendidas como uma conexão primitiva do pensamento com a natureza, sendo que sua manifestação "crua" é uma representação natural, despida do filtro social e irracional. Ainda que estas definições tenham um importante valor dentro da mentalidade que se formou posteriormente em torno do pensamento grego, principalmente aquela que foi retomado durante a Idade Média tardia. Ainda assim, a nosso ver, a definição aristotélica mais importante no interior desta seara foi a distinção feita entre *páthos* e *lógos*, apresentada em sua Retórica (ARISTÓTELES, 2005).

Para Aristóteles, o *lógos* é a racionalização, a capacidade de conhecer e explicar um tema baseado no pensamento lógico, enquanto o *páthos* é a forma como o público ou um indivíduo reage a partir de uma determinada informação apresentada a ele, sendo que um bom retórico sabe como motivar sua audiência e consegue elicitar essas

emoções. Dentro de sua Retórica elas ocupam uma posição central, de tal modo que Aristóteles se propôs a analisar e descrever 14 paixões, levando em conta a ativação e a manifestação de cada uma delas nos sujeitos. Como aponta Fortenbaugh (2002) esta análise feita por Aristóteles é uma das primeiras análises sistemáticas, organizadas em um método e com um finalidade específica, pois busca compreender os princípios, as condições e os objetos das emoções, ainda que sua descrição seja muitas vezes vista como a simples enumeração de caracteres baixos, uma vez que ela se confunde com os pressupostos de sua ética.

6.2.2 Estoicos

Se Aristóteles, em certa medida, rompe com a tradição platônica e passa a compreender a importância e necessidade da manifestação das paixões, desde que moderada, a tradição estoica descarta totalmente essa hipótese e funda um novo princípio baseado ainda em Sócrates, a saber: que paixões e virtude não podem conviver. Mesmo os textos iniciais da doutrina, estabelecidos por Zenão de Cítio e Crisipo de Solis, tratavam o *páthos* como uma perturbação, uma rejeição à razão. Nota-se que nos textos fundamentais utilizam a terminologia "perturbações" para as paixões, que desviam um homem da conduta "correta", essa terminologia nitidamente destaca as paixões de um sistema complexo e as coloca dentro de um sistema moral, bastante restrito (GONZÁLES, 2005a).

Aristóteles entendia que as emoções eram causadas pela avaliação de um dado objeto, estoicos, por outro lado, entendiam que as paixões são as avaliações em si. Um cão que rosna desperta o medo para os peripatéticos, para os estoicos a avaliação de que o cachorro pode ser perigoso é o que desperta o medo. Essa suposição permitiu aos estoicos, principalmente Sêneca (SÊNECA, 2014), desenvolverem um processo primitivo de análise das emoções, sem levar em consideração a descrição de cada uma delas. Para eles em um primeiro momento o indivíduo tem uma *opinião* a partir de um determinado evento ocorrido, essa opinião, ou primeiro impulso, era considerado

involuntário e o leva a um estado de alma (*krísis*), o qual é um segundo impulso, em que ele avalia a necessidade de reação diante do ato e da opinião involuntária. Somadas essas etapas o indivíduo tem um *impulso reativo*, que é a manifestação da emoção em si. Controlar esse último impulso demonstra uma elevação de alma e é uma valorização da razão, de modo tal que o homem virtuoso seria capaz de o fazer (SÊNECA, 2014, p.46).

Uma obra fundamental dentro desta perspectiva são as *Tusculanae Disputationes* de Cícero (2005), na qual o Cônsul romano, logo após a perda de sua filha, desenvolverá uma teorização das paixões. A hipótese central de Cícero é justamente a de que as paixões são "perturbações da alma", uma derivação da análise de Zenão e o objetivo primeiro de seu trabalho é demonstrar como um homem sábio pode se ver livre delas (*Tusculanae Disputationes*, IV, 2)⁸. Sua metodologia para esse fim é dividir as emoções em gêneros e subgêneros, dentro de um sistema que as combina conforme uma orientação moralista e temporal (*Tusculanae Disputationes*, IV, 9 e ss.).

Neste sistema, as paixões são frutos da nossa opinião, sentimos conforme avaliamos uma determinada situação, mas há também um componente temporal nesta combinação, isto é, a paixão a ser manifestada está condicionada a sensação de presente ou futuro. Neste sentido, o medo seria uma emoção derivada de uma avaliação de um risco presente, já a aflição seria uma paixão associada a um determinado evento futuro (*Tusculanae Disputationes*, IV, 13 e ss.). Nesse cenário, a conclusão possível é a de que as paixões derivam todas de uma mesma paixão, a *intemperança*, que é o abandono da razão, ou seja, sente-se algo única e exclusivamente quando se está desprovido dela já que ela é "propulsora" das demais emoções (*Tusculanae Disputationes*, IV, 22)

Cícero vai além em sua análise e sugere que as perturbações da alma, as paixões, podem ser compreendidas tal como as enfermidades do corpo, sendo que cada indivíduo seria mais propenso a uma determinada doença do que a outra. Sua exposição sustenta que as doenças da alma seriam vícios, perenes no sujeito, ao passo que as

⁸Ver Cícero (2005)

paixões são transitórias e afirma que apenas a alegria, a cautela e a temperança são as disposições que cabem ao homem sábio. Embora não fique claro, entendemos que Cícero assume que as paixões são entendidas como doenças, que podem afetar o corpo, uma ideia que perdura ao longo dos tempos na cultura ocidental e teve um desenvolvimento nas culturas chinesas e árabe (SUCHY, 2011). Essa hipótese é vital para que ele proponha um "sistema de cura" das paixões, uma espécie de proto-terapia (*Tusculanae Disputationes*, IV, 62). Além disso, o método proposto nitidamente visa refutar os pressupostos da escola aristotélica, uma vez que parte pressuposto que sentir emoções em qualquer nível é negativo.

Sêneca em *Sobre a Ira* também busca refutar os pressupostos da escola peripatética, mas de maneira mais evidente, como se lê em:

Afirma-se: "a paixão é útil, se moderada". Não, ela é útil, se baseada na natureza. Mas se não tolera o controle da razão, com a moderação não se conseguirá mais do que isto: quanto menor ela for, menos há de se prejudicar. Portanto, uma paixão moderada não é outra coisa que um mal moderado.

Sobre a Ira, I, 10, 4. (SÊNECA, 2014)

Sêneca também irá se dedicar a combater os epicuristas, que viam as paixões tal qual os peripatéticos. Para eles, as paixões podem ser divididas em duas formas: (i) natural e (ii) vã. A primeira é uma necessidade natural, fonte de prazer e, por isso não compromete a tranquilidade alma, já a segunda por depender de fatores sociais e não ter limites, causa dor e é nociva. A raiva teria duas faces: uma natural e positiva a outra vã e negativa, de tal modo que é necessário o concurso da razão para saber o limite entre uma face e outra, ou seja, Epicuro condena os excessos (SÊNECA, 2014). Embora Sêneca esteja tratando unicamente da Ira, sua argumentação se vira para todas as paixões, exceto aquelas consideradas virtuosas, como o amor e a caridade. Em outro ponto, ao traçar um paralelo entre o uso da razão e das emoções com armas em uma guerra afirma:

Não precisamos de outros instrumentos; a natureza nos proveu suficientemente a razão. Deu-nos esse dardo robusto, duradouro, obediente

que não tem dois gumes nem é passível de ricochetear contra seu dono. Não apenas para nos precavermos, mas para agirmos, a razão é ela própria suficiente por si.

Sobre a ira, I,17,2 (SÊNECA, 2014).

Como se nota, para os pensadores estoicos, somente a total abstração do *páthos* diante do *lógos*, para utilizarmos a terminologia aristotélica, seria capaz de conduzi-los para uma moral "pura". Desta forma, a leitura moral feita por aristotélicos e socráticos era retomada, porém o processo de intermediação destas forças estabelecido por Aristóteles era retirado (*Sobre a Ira*, II, 1, 3). Em outras palavras, no sistema moral aristotélico a manifestação das emoções era vista como um princípio moral baixo, que deveria ser evitada, mas, ainda assim, existente, ao passo que os estoicos consideravam sua simples existência um ato imoral.

6.2.3 A doutrina escolástica

Como se sabe o estoicismo terá braços fortes na fundação do Cristianismo Medieval e no pensamento de Santo Agostinho, um importante nome na construção das doutrinas católicas que sedimentaram e formataram grande parte da cultura ocidental (DAWSON, 2009). Santo Agostinho foi um ávido leitor de Cícero (AGOSTINHO, 1996) e grande parte de seu pensamento foi influenciado pelas ideias do cônsul romano. Em *A Cidade de Deus* Santo Agostinho retoma a discussão feita em *Tusculanae Disputationes* e concorda com a ideia de que as paixões são perturbações da alma e afetam a razão, porém discorda dos estoicos quando estes dizem que elas são vícios e cegam as virtudes, a seu ver, é necessário um grau de *insensibilidade* às emoções, mas, ao mesmo tempo, é necessário senti-las, seja raiva pelos pecadores ou a benevolência pelos inimigos.

As ideias de Santo Agostinho estão assentadas em diversos princípios cristãos, principalmente aqueles debatidos nas epístolas de Paulo aos Coríntios, de modo tal que suas formulações visam antes elucidar um sincretismo filosófico da tradição grega e latina com o cristianismo primitivo, sendo assim, seu objetivo não é analisar as

emoções em si, mas colocá-las no bojo de um sistema filosófico-religioso que começava a se construir na Europa do século VIII. No livro XIV de *A Cidade de Deus* Santo Agostinho assume que as emoções são punições pelo pecado capital, uma vez que elas perturbam a razão e a alma (14, 9)⁹ e são uma lembrança de que habitamos o paraíso e agora estamos condenados a vida carnal.

Em um certo sentido, sua leitura é a de que no Paraíso haveriam emoções, mas estaríamos todos no controle delas. Além disso, ele acredita que não existam emoções elevadas, como defendiam os estoicos, ainda que na tentativa de explicá-las, use-se da divisão das paixões em 4 fundamentais (14,6): o medo, o desejo, a aflição e o prazer. Neste sistema elas eram divididas pelo seus objetos, isto é, se orientadas para algo bom (o prazer e o desejo) ou ruim (a aflição e o medo), além de organizar as paixões em um eixo temporal: desejo e medo seriam categorias de futuro, prazer e aflição de presente. Nota-se que Santo Agostinho assume o mesmo pressuposto de Cícero e demais estoicos, de que as paixões são orientadas pela avaliação de seu objeto e não pelo objeto em si.

Embora essa leitura seja baseada no estoicismo, ela também é profundamente influenciada pelas epístolas de Paulo ao Coríntios e aos Gálatas que tratam de questões que fundamentam a moral Cristã e princípios ritualísticos que estão na base da construção da Igreja Católica. Com relação às paixões Paulo sugere que elas estão em conexão com um eixo temporal (Ct 4,16-17): "as sensações presentes levam a uma eterna glória e as aflições com o futuro são vazias e nos leva a corrupção". Estas epístolas inserem a questão do "ausentes do corpo e presentes com o Senhor", uma doutrina religiosa que vê como divino elementos que não estão em conexão direta com as necessidades corporais, tal como o desejo, e valoriza sentimentos considerados nobres. Esse entendimento causa um impacto profundo na concepção da importância das paixões feita por teólogos posteriores, pois se são do domínio corporal, necessariamente elas não pertencem aos propósitos divinos.

⁹Para todas as citações ver Agostinho (1887)

A análise de Santo Agostinho termina por tentar alocar a luxúria e a raiva neste sistema (14, 12). Seu modelo parte do princípio de que as emoções estão associadas a avaliação de seus objetos e a decisão de avaliá-los como bons ou maus é um ato de vontade¹⁰, que é a distinção principal da teoria Augustiniana para a o estoicos, sendo que a manifestação da emoção depende do seu agente, ou melhor das vontade deste sujeito. Por exemplo, a luxúria é um produto máximo da vontade, originada pelo pecado original que nos levou a vida carnal, portanto é um desejo da carne pela carne, já a raiva é um "desejo carnal pela vingança". (14,15). Neste caso, as emoções não estão em um plano moral em si e o que ocuparia essa posição é a vontade do agente.

Santo Agostinho, assim como Platão e os estoicos, compreendia que éramos divididos em corpo e alma, sendo que as emoções eram partes do corpo e justifica sua argumentação, através da explicação do pecado capital, uma doutrina central em sua filosofia. Em *A Cidade de Deus*, algumas passagens deixam evidente que foi este momento que nos condenou a uma vida de vícios da carne, dentre elas as emoções, embora elas não estejam no centro de um sistema moral, como viam os demais, uma vez que a moral medieval era apoiada nos preceitos da bíblia e não em um sistema filosófico. A doutrina Agostiana norteou grande parte da filosofia medieval e teve seu declínio com as leituras aristotélicas feitas no século XIII, que influenciaram o pensamento de Tomás de Aquino e promoveram uma revisão de objetivos na Filosofia Escolástica.

Se a distinção entre corpo e alma é fundamental na filosofia platônica e agostiniana, no tomismo ela será repensada. Tomás de Aquino procura explicar a doutrina cristã e alia-la aos pressupostos de filosofia natural criados por Aristóteles. Podemos

¹⁰Dentro das filosofias de fundo teológico, a ideia de vontade é concebida como um conceito divino, isto é a vontade é a vontade de Deus, ou de alguma entidade superior. Neste sentido, um dos objetivos de uma vida plena é compreender esta vontade e segui-la como norte de vida. Em um certo sentido, é possível afirmar que Santo Agostinho coloca esta "condição" na manifestação das emoções, uma vez que a vontade tem de necessariamente ser a vontade divina. Além disso, essa postura o ajuda a se distanciar de alguns pressupostos dos maniqueístas, que viam as ações como intrinsecamente boas ou más, sem condicioná-las ao livre arbítrio. Sendo assim, vontade aqui retoma estas duas dimensões: a vontade divina e o livre arbítrio de se adotá-la.

dizer que nesse momento surge uma psicologia embrionária, que será desenvolvida posteriormente por Willian James e que coloca as emoções como um problema natural e as distancia de um sistema ético ou teológico. Como descrito anteriormente, havia uma insistência na tentativa de se conceber as paixões como elementos de um sistema moral, de tal modo que as manifestações não eram analisadas como processos, mas, sim, como finalidades (KING, 2011).

Dentro desta perspectiva, o principal problema encontrado foi o de traçar uma linha em que se pudesse colocar a caráter caótico das emoções no sistema rígido das ideias aristotélicas. A solução simples foi entendê-las como havia feito o próprio Aristóteles, isto é, como elementos que induzem o indivíduo à ação. Mais do que isso, Tomás de Aquino irá assumir o pressuposto do *hilemorfismo*, entendendo que mente e corpo formam uma relação inequívoca. Sendo assim, sua hipótese é a de que as emoções possuem duas dimensões: (i) aquela que a motiva, chamada de *appetite* (ii) e sua manifestação, de um lado um aspecto formal e do outro material (ST, I, q75a4)¹¹. Dissociar essas duas dimensões seria necessariamente desconstruir a finalidade das emoções ou sua definição *per se*.

Platão via a razão como função única e final da mente, sem relação com o corpo. Aristóteles defendia um princípio semelhante e seu inverso, todas as funções são dependentes da mente e também do corpo, porém a razão é a única que depende somente da mente. Tomás de Aquino assume um pressuposto parecido, mas entende que de alguma forma a mente está conectada com o corpo, ou do contrário não há cognição, uma capacidade puramente humana. A questão se as paixões estão ou não nesse domínio é uma discussão que se prolonga entre os mais variados leitores de São Tomás (GOREVAN, 2000). A resposta não é absoluta, pois a definição Tomista é a de que as emoções são movimentos acionados pelo *sensory appetite* (ST, I, q80a2)¹², que movem o corpo em *direção a* ou *para além de* um objeto, baseado na percepção resultante

¹¹Ver Aquino (2002)

¹²Idem.

deste objeto, se ela positiva ou negativa.

A discussão principal refere-se ao fato de que as emoções não podem ser simplesmente explicadas como movimentos de instintos da percepção, elas devem ter algum grau de correlação entre a natureza deste objeto e sua manifestação (ROBERTS, 1992), de tal modo que a este evento pode ser intencional em certa medida. Ao passo que outros veem em sua teoria uma *percepção cognitiva*, isto é, a correlação entre a natureza do objeto e a manifestação da emoção está garantida uma vez que ela é pressuposta, a emoção a ser manifestada é previsível a partir do objeto. Os sentimentos, portanto, podem estar ligados a uma emoção ou suscitar a emoção. É preciso dizer que Aristóteles via esta questão como resolvida, uma vez que eram os estados mentais que despertavam os sentimentos de uma emoção.

À parte desta discussão, a questão central é que Tomás de Aquino é um dos primeiros pensadores a ver a relação entre a manifestação das emoções, o seu aspecto material e a sua relação com um estado mental, o *aspecto formal*. Para ele, essa relação era evidente, visto que a reação física era a réplica de um estado mental, sendo assim para se compreender as emoções era necessário olhar para as duas faces. A definição de emoção, portanto, recai neste princípio, *algo somente é uma paixão se e somente se ele promove uma alteração corporal*. Porém, é preciso levar em conta que lidar com estas sensações não é lidar com as emoções em si. A distinção mente/corpo no que diz respeito às emoções se encerra justamente neste ponto: não é possível limitar o que são categorias da mente e o que são manifestações corporais, já que elas estão em uma relação contínua, como já sugeria Aristóteles, influencia mais marcante nos escritos de São Tomás de Aquino.

Uma série de restrições podem ser feitas ao pensamento tomista que sempre foi taxado como reificante, circular e excessivamente hipostático. No entanto, suas ideias tiveram um impacto significativo na forma como as ideias aristotélicas chegaram até nós e como eram interpretadas, sendo que é difícil estabelecer uma fronteira entre

o seu pensamento e o aristotélico. De modo geral, pode-se notar que os pensadores do período clássico e do período medieval colocaram as emoções como elementos à parte da natureza humana, a exceção é talvez Aristóteles. Para platônicos e estoicos as paixões representavam um ruído para a razão pura, enquanto que para agostinianos elas representavam um risco desnecessário à moral pura.

Dessa forma, a única alternativa para as emoções era o seu expurgo. Como veremos, a seguir, o pensamento cartesiano representou uma grande evolução nesse assunto, ao sugerir que as emoções poderiam ser contempladas dentro de um método. No entanto, ao separar mente e corpo essa hipótese foi rapidamente tragada pelo pensamento platônico e escolástico, que viu na formalização desse raciocínio uma forma de justificar a separação por eles já aventada.

6.3 Fundando as bases modernas: Descartes, James e Darwin

6.3.1 Paixões cartesianas

Quando olhamos em revista o que foi discutido na seção 6.2.1 parece estarmos diante de um emaranhado de conceitos poucos didáticos. A distinção corpo e alma, a localização das emoções e a pressuposição das emoções estarem no bojo de um sistema moral antes de um sistema psicológico parece ser a tônica destes pensadores, sem que a formalização das ideias ou a sistematização de algum processo de análise tenha sido alcançada. Simplificadamente, nota-se que, com pequenos ajustes ou alterações, a noção platônica do *páthos* como uma parte inferior à condição humana e do *lógos* como uma parte superior se manteve inalterada ao longo da formação dos impérios, da Idade Média, principalmente na fase de formação e consolidação do Cristianismo, bem como nas primeiras etapas do Renascimento e ainda se mantem em certos círculos de pensamento.

O Renascimento marca também a passagem do pensamento filosófico para o

método científico, quando questões de descrição e percepção, passam a ter que ser demonstradas, comprovadas e passíveis de replicação. René Descartes escreve em seu "As paixões da Alma" que todas ideias sobre emoções anteriores a ele recaem no erro de serem percepções individuais, isto é, não são aplicadas em um método e, sim, a descrição de uma experiência daquele que escrevia sobre elas. Nesse sentido, uma decepção amorosa pode enviesar a compreensão desse sentimento, bem como o êxtase pode confundir a alegria, etc. Descartes, assume, então, que para que as paixões façam parte do programa de pesquisa Cartesiano é necessário que se recrie todo o instrumental de análise, imaginando que as respostas encontradas podem ser colocadas em modelos de previsão da filosofia natural (DESCARTES, 2012, p.31).

Em certo sentido, Descartes coloca o problema do observador, retomado diversas vezes na história da ciência, no bojo da análise das emoções, ao entender que as paixões são *percepções* do indivíduo. Essa, sem dúvidas, é a grande contribuição cartesiana no estudo das emoções, ainda que se veja com certa desconfiança a solução por ele encontrada (DAMÁSIO, 2012; CRICK, 1995; FREEMAN, 2000; DAMÁSIO, 2015). Descartes parte do pressuposto platônico que corpo e alma são unidades distintas, sendo o corpo responsável pelo movimento e pelo calor, ao passo que a alma é onde acontecem as ações e paixões da alma. As ações da alma são, basicamente, ações do desejo, nomeadas de *vontades*, que levam o sujeito à ação, já as paixões da alma são *percepções*, causada pela alma ou pelo corpo, como a percepção de uma ação da alma (DESCARTES, 2012, p.42 e ss.), sendo que algumas delas são causadas pelo corpo e outras pela alma.

Descartes entende que os elementos disparadores das emoções, como faziam estoicos e peripatéticos, não fazem parte das emoções, uma vez que elas são produtos de uma avaliação da alma. Por exemplo, o amor é a percepção da alma de que algo é inofensivo e agradável e em sua anatomia rudimentar, Descartes sugere que a responsável por essa percepção é a glândula pineal que irá fazer mover os músculos e coração com base na avaliação do objeto a sua frente (DESCARTES, 2012, p.50-51).

Em última análise, a manifestação de uma emoção é um fenômeno de emergência de algo totalmente internalizado, um movimento de "espíritos" despertados pela alma, que ocorrem no corpo, mas que estão a serviço da alma. Neste caso, somente o reverso destas sensações, um objeto cristalizado, pode ser considerado a emoção em si.

A questão principal, levantada por Descartes e amplamente discutida por cientistas como Damásio (2012) e Crick (1995), é a hipótese de que as emoções são, portanto, unidades irreduzíveis e manifestadas como base em uma avaliação caso-a-caso. Pressuposto que será retomado posteriormente por Locke (1689) no que é chamado de Tradição Perceptual Atomista das emoções (SCARANTINO, 2016). O problema teórico central é que esta hipótese nos permite considerar que as paixões estão alocadas em um espaço virtual, separado do corpo, criando o dualismo "mente-corpo", onde as facetas apenas convergem em pontos específicos e nunca em seu todo.

Descartes não pressupôs essa correlação a partir das emoções, mas, sim, tentou alocar as paixões nesse modelo, já apresentado em suas *Meditações* (DESCARTES, 1983). O impacto desse dualismo é a consideração de que as emoções são forçosamente resultantes de processos mentais abstratos (as percepções), sem dimensões e amorfos, de modo tal que o corpo não tem ou não deve ter nenhum papel, sendo ele unicamente o vetor da sua manifestação, já que possui variações e dimensões cabíveis de se analisar. Como o corpo é uma unidade extensa ele pode ser medido e avaliado dentro do método cartesiano, ao passo que as emoções, unidades adimensionais, não podem.

Além disso, por estarem alocadas em um espaço virtual de realização e introjectadas, no caso na alma, as emoções não poderiam ser consideradas elementos constituintes da natureza humana, mas, somente parte integrante dela, uma vez que não se pode estabelecer uma relação causal entre elas e os objetos externos que a motivam. Dessa forma, é natural que elas sejam sentidas na alma e por isso estejam na alma. Esta noção é confusa, mas traz uma dimensão relativa, pois sugere que a subtração das emoções não representa um efeito imediato na constituição do Homem

¹³. Nesse sentido, podemos nos equivocar sobre o que causa as paixões, mas não pode negar a existência delas, uma vez que sua realidade é manifesta pelo corpo, o que nos leva a ter consciência delas.

Por este motivo, as emoções foram desprezadas de uma análise sistemática, ao passo que os elementos que constituem sua manifestação foram tratados como os objetos de análise, a serem elencados e analisados dentro do método cartesiano, para posteriormente serem categorizados. Dentro desta proposta, as emoções são apenas aquilo que o corpo sente. Portanto, a psicologia cartesiana compreendia a mente e, por consequência, as emoções como apenas aquilo o que temos de consciência sobre ela por sentirmos no nosso corpo e nunca algo capaz de ser estudado metodicamente. Hipótese que Freud, com o método da cura pela fala, demonstrou estar errada (FREUD, 2015a).

A questão fundamental no pensamento cartesiano é o de que as emoções são a percepção e o julgamento que fazemos dos objetos que as motivam, seguido da avaliação que fazemos de suas manifestações, de modo que não existe um aspecto cognitivo a ser considerado. A razão não toma parte nesse processo, pois as paixões são uma constatação factual, elas simplesmente existem, não é necessário a "acionar" a lógica para definir a sua existência. Daí a separação entre razão e paixão ser retomada.

Além disso, como as emoções dependem da percepção para existir, elas, simplesmente, não podem ser negadas, uma vez que são obrigatoriamente sentidas, embora não possamos compreender o porquê, o que obscurece a razão. Nesta perspectiva, o domínio mental é um receptáculo das sensações físicas provocadas pelas emoções, uma vez que a razão não pode desconsiderar as sensações experimentadas pelos sentidos, nem os sentidos podem deixar de sentir as paixões simplesmente porque elas não são

¹³A ideia de que o homem pode suprimir suas emoções tem sido um tema de debate recorrente na história e, até mesmo, produto cultural, como é o caso do personagem *Spock*, da série *Star Trek* que é incapaz de sentir emoções por não ser humano. Damásio (2012) argumenta que essa hipótese é insensata, pois as emoções funcionam também como reguladores sociais e pessoais, não necessariamente a manifestação delas, mas, sim, sua manifestação.. Mesmo em *Star Trek*, embora Spock não possa manifestar suas emoções ele reiterada vezes se preocupa com seus companheiros ou familiares, por exemplo

claras de um ponto de vista racional.

6.3.2 Hume: paixões como ação empírica

Para Hume (1740) esse dualismo se mostra equivocado. Para ele, as nossas percepções podem ter duas formas: *impressões* e *ideias*. Hume não faz uma distinção objetiva entre essas duas dimensões, a não ser a de que as impressões são mais vivas e nos atingem de forma "mais violenta e forte", e seriam as paixões e sensações enquanto as ideias são "imagens desbotadas" dessas impressões. Hume entende que elas estão em uma relação causal, não existem ideias sem impressões e vice-versa. Dentro desta lógica, as paixões são um "reflexo das impressões", pois são causadas por elas ou pelas ideias, que são motivadas pelas impressões.

Por um viés cartesiano, as emoções são unicamente as suas manifestações, então, evidentemente, não podemos imaginar o que é o medo ou a raiva, a não ser, é claro, que as experimentemos ou os tenhamos em nossa alma. Para Hume, as paixões também não são ideias, mas sempre impressões, as quais estão diretamente conectadas à nossa experiência de mundo e, por isso mesmo, não podem contradizer a verdade ou a razão. Desta forma, a separação entre razão e paixão não é cabível, simplesmente porque as emoções representam um primeiro contato com a realidade. Em certo sentido, no pensamento de Hume (1740) a razão é "dependente" das emoções, já que a razão é parte das ideias e as ideias são dependentes das impressões:

Thus it appears, that the principle, which opposes our passion, cannot be the same with reason, and is only called so in an improper sense. We speak not strictly and philosophically when we talk of the combat of passion and of reason. Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them. As this opinion may appear somewhat extraordinary, it may not be improper to confirm it by some other considerations. (HUME, 1740, T.2.3.3.4)

Para Hume, as emoções são as únicas faculdades psicológicas que nos levam à ação, pois estão em conexão direta com os objetos da realidade e representam uma modificação da existência (a saída de um estado A para um estado B), a partir destes

objetos. Assim, a razão pode estabelecer quais são as melhores formas de se atingir um determinado fim orientado pelas paixões, mas não pode estabelecer um fim por ela mesmo, uma vez que ela não tem relação direta com a realidade, a não ser por meio das ideias que derivam das impressões (HUME, 1740). Noção parecida somente será retomada com a proposta do *marcador somático* desenvolvida por Damásio (2012).

Entretanto, à parte destas distinções entre os autores, existe uma questão que parece não ser resolvida nem por Descartes, nem por Hume, mas que tem uma evidente importância, refere-se ao fato de que as emoções não têm um referencial estabelecido na realidade, elas são o próprio referencial, pois representam uma motivação à ação que seria o seu "significado". Assim, existe uma questão fundamental nos problemas apresentados e profundamente relacionada à semiologia das línguas naturais: porque expressamos as emoções do modo como as expressamos?

Olhando em retrospecto, essa pergunta foi a base para o desenvolvimento da psicologia das emoções no século XIX. Em seu *A Expressão das emoções do homem e dos animais* de 1872, Charles Darwin (1872/2016) expõe uma teoria que visa demonstrar como a manifestação das emoções é similar entre animais e homens, o que sugere uma continuidade entre elas. O pressuposto que justifica essa hipótese é o de que a expressão das emoções possui pouca ou nenhuma variabilidade, seja em uma análise inter-espécies, seja em uma análise intercultural ou linguística. Para ele, as emoções eram estados interiores e a manifestação um ato comunicativo subjacente a esse estados.

6.3.3 Darwin e James: a expressão das emoções

Para explicar sua hipótese, Darwin lança mão de duas ideias que são os pilares de uma possível "ciência das emoções". O primeiro pilar é o de que a expressão das emoções têm um "ancestral" comum e ao comparar centenas de fotografias e imagens de seres humanos e animais em situações emocionais, ele foi capaz de demonstrar as semelhanças entre as mais variadas espécies. Essa proposta, torna a emoção um

elemento de suma importância para a sobrevivência das espécies e um importante elemento da teoria da seleção natural. O segundo pilar, mais polêmico e complexo, é o de que existe um grupo de emoções básicas, invariáveis, seja interculturais ou inter-espécies, assim, um conjunto de emoções, teria sido selecionado pela natureza e contribuído para a sobrevivência das espécies e dos indivíduos.

Neste caso, o desenvolvimento do hábito é um ponto importante desta questão, seguido da ação direta do sistema nervoso no corpo (DARWIN, 1872/2016, p.295 e ss.). Em tese, as emoções são uma resposta a um estímulo externo e a ação de sistemas autônomos, desta forma estas ações se tornariam comportamentais e categorizariam a experiência por meio do hábito: habituar-se a um conjunto de ativações do SNA, ajuda a compreender elementos externos ao corpo e ações necessárias para se minimizar a presença desse estímulo. Sendo assim, as emoções representam uma espécie de elo perdido entre a natureza e a cultura, distanciados e ao mesmo tempo relacionados entre si pela evolução das espécies. Portanto, se há uma inter-relação das emoções, ela se dá pela origem comum que elas não de haver: as emoções nos fazem compreender determinadas situações e qual ação devemos tomar, como resposta reflexa e, por isso, as manifestamos como as manifestamos. (DARWIN, 1872/2016, p.17).

A proposta de Darwin segue o pressuposto do seu programa de pesquisa, mas ao mesmo tempo representa uma ruptura, pois tenta alocar as emoções no princípio natural que rege a vida do homem, ao invés de colocá-la em um sistema criado e desenvolvido pelo próprio homem. Para isso, Darwin define 3 princípios que regem a expressão e a hereditariedade das emoções: (i) *o princípio dos hábitos*; (ii) *o princípio da antítese* e (iii) *o princípio das ações do sistema nervoso autônomo, a despeito de nossas consciência* (DARWIN, 1872/2016, p.33). Sendo assim, a manifestação das emoções, ponto de relação com a linguagem, não é mais que um movimento habitual, sempre iniciado com a intenção de aliviar ou satisfazer alguma sensação, ainda que esse movimento não tenha utilidade prática. Se estamos tristes, então choramos.

Quanto à manifestação das emoções, Willian James vê essa questão por outra perspectiva e, em *Principles of Psychology II*, lê-se:

Our natural way of thinking about these coarser emotions is that the mental perception of some fact excites the mental affection called the emotion, and that this latter state of mind gives rise to the bodily expression. My theory, on the contrary, is that *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, 1890, p.449)

James compreendia as emoções como uma construção feita de partes elementares, isto é, as paixões são elementos compostos (e complexos, por extensão) constituídos de ingredientes mais básicos, ou seja, as emoções são construtos, hipótese semelhante a de Hume (1740). Nesta interpretação, as emoções são produtos secundários que surgem a partir da experiência de determinados estímulos, a saber: as manifestações corporais, que são produtos de um reflexo ou reação a determinados estímulos externos ou internos.

Sendo assim, as emoções são produtos da percepção da alteração de um estado fisiológico: não choramos porque estamos tristes, mas, sim, ficamos tristes porque estamos chorando. Deste modo, as emoções não têm existência fora do corpo, como supunham os cartesianos, não são meros produtos da alma. Essa inversão permite a James compreender mecanismos básicos que as compõem e que não são correlacionados diretamente com elas, sejam eles elementos sociais ou individuais.

Dentro do modelo James-Lange (LANGE; JAMES, 1922) ¹⁴ os sentimentos são as alterações fisiológicas primárias, como a frequência cardíaca, a pressão arterial e o padrão respiratório, essas mudanças formam a base fundamental do que entendemos por emoções, se entendidas como percepções da alteração, é claro. Para os autores se retirada essa manifestação somático-fisiológica, não existe a emoção em si, sem as respostas neurovegetativas causadas por paixões como o medo e a raiva, elas simplesmente não existem. O modelo de Lange e James (1922) coloca como seu

¹⁴Willian James e Carl Lange chegaram as mesmas hipóteses com apenas alguns de diferença, por este motivo, adotou-se utilizar a nomenclatura conjunta para descrever essa teoria

principal programa de pesquisa catalogar e compreender a relação entre os elementos individuais e sociais que estão contidos nas emoções e a associação desses pressupostos a determinados processos fisiológicos.

Tanto Darwin (1872/2016), quanto Lange e James (1922) se apoiam na noção de ganho evolutivo, ou seja, na ideia de que a percepção de um objeto dispara uma série ações corporais que nos levam a uma mudança de estados. A diferença entre os autores é a direção desta ativação. Darwin supõe que a manifestação das emoções é ativada por processos neurológicos, a partir dos estímulos externos, enquanto Lange e James (1922) supõem que as emoções são ativadas a partir da percepção da alteração dos estados, causada pelos estímulos externos. Desta forma, a preocupação de James e Lange é explicar como experienciamos as emoções, enquanto Darwin se preocupa com a análise da expressão das emoções.

A inversão sobre os sentidos das emoções feita por Willian James e Carl Lange representa uma grande revolução na forma como se entende as a relação corpo e emoção, uma vez que podemos deixar de compreendê-las como um simples corolário de ações derivadas de categorias abstratas para compreendê-las como manifestações que categorizam nossa experiência. Sendo assim, o corpo humano evoluiu de tal modo que se tornou capaz de relacionar manifestações/sentimentos aos ambientes experimentados e esta combinação a emoções específicas. Em tese, estas respostas são automáticas e controladas pelos sistemas nervosos autônomos, como demonstrado em Averill (1969), Boudewyns (1976), Ekman, Levenson e Friesen (1983), McHugo et al. (1985), Waters, Bernard e Bucu (1989), Levenson, Ekman e Friesen (1990) e para uma revisão completa recomendamos Kreibig (2010).

6.3.4 Cannon-Bard: questões de neurologia

A principal oposição ao modelo de James e Lange veio dos estudos de Cannon (1927b, 1927a) e Bard e Rioch (1937), base para a Teoria Cannon-Bard. O teste para

demonstrar esta hipótese foi a supressão de alguns nervos, principalmente os aferentes do sistema nervoso autônomo, bem como a remoção do córtex de gatos Bard e Rioch (1937). Nestes casos, os animais estudados apresentavam estados de raiva, chamados de *sham rage*, similares em suas manifestações aos de animais não-decorticalizados. Para os autores esse fenômeno demonstrava que proposta de James-Lange estava errada, uma vez que ela pressupunha o funcionamento saudável destes sistemas. Como proposta os autores sugeriram que há um centro de avaliação das emoções, no caso o tálamo, onde ocorre uma dissociação do estímulo externo em duas faces: (i) uma seria direcionada ao hipotálamo e cuidaria da parte somática das emoções e outra (ii) para o córtex, onde ocorreria a experiência subjetiva das emoções.

A figura 6.1 abaixo representa esquematicamente essa oposição:

Figura 6.1: Comparativo entre os modelos James-Lange e Cannon-Bard. As setas pretas demonstram que, de acordo com James-Lange, a percepção do estímulo leva a uma reação, a qual se converte na experiência emocional. Por outro lado, as setas azuis apresentam a visão de Cannon-Bard, que entendem que a experiência emocional é anterior a manifestação em si.

O modelo de Cannon-Bard foi rapidamente absorvido pela neurociência eminente e junto dos estudos de Papez (1937) e MacLean (1949), contribuíram para formar a base de uma ciência em desenvolvimento, cujo objetivo era investigar as raízes da emoção a partir do estudo do cérebro e por este motivo foi chamada de *Affective Neuroscience* (para uma revisão do assumo recomendamos Panksepp (2004)). Esta teoria, profundamente distanciada dos modelos citados anteriormente traz à tona uma

variedade de discussões e ideias, envolvendo basicamente os mecanismos biológicos que contribuem para ativar as emoções e são ancorados a ideia de que as emoções são resultados de processos cerebrais sistêmicos, isto é, quando diversas partes competem para gerar uma resposta fisiológica a um estímulo.

De modo geral, parece-nos que o percurso da ciência das emoções, desde a Antiguidade até a virada do século XX teve 3 capítulos principais:

- **A Tradição moral:** baseada nas ideias socráticas e estoicas, essa tradição parte do pressuposto de que as emoções devem ser alocadas dentro de um sistema moral, pois sua manifestação é capaz de revelar o real espírito de um homem, diante de um dado fato, de modo tal que apenas sua abstração é capaz de tornar a existência positiva. Ainda que questionável, a ideia de que as emoções revelam um estado de alma é bastante importante para a compreensão posterior que viríamos a ter do papel das paixões no estabelecimento da relação do homem com o seu meio, seja ele natural ou cultural. Em um certo sentido, sua compreensão e método é validado apenas em alguns círculos religiosos ou pseudo-psicológicos. Fazem uma distinção entre *natural* e *cultural*, justamente por representarem uma etapa do pensamento que buscava alocar o homem em uma posição de relevância dentro.
- **A Tradição cartesiana:** nesta etapa, a questão moral não é excluída, mas o fator "observador" é inserido, resultando em um problema metodológico que será resolvido pela distinção entre corpo e alma ou mente e corpo. Essa separação faz surgir outro problema, de natureza distinta, pois subentende que as emoções são parte de um processo realizado pela mente, um conceito amorfo e manifestada pelo corpo, um objeto dimensional. O fato questionável é como se mensurar algo amorfo, por outro lado faz surgir a ideia de que os sentimentos são a base para compreender as emoções. Fazem uma distinção entre *corporal* e *mental*.
- **Os modelos baseados na expressão:** Descartes, Hume e Darwin colocam o fator

observador como principal problema metodológico em seus trabalhos. Darwin é o primeiro que se faz valer de uma quantidade significativa de dados a serem analisados, algumas percepções individuais são, inclusive, postas à prova em suas *Queries About expression* (DARWIN, 1872/2016, p.21 e 22). Alguns dos pressupostos delineados por Darwin e também por James, como a função das emoções, o conceito universal que elas transmitem, bem o que nos leva a senti-las e os comportamentos que elas suscitam formam o eixo teórico mais produtivo do século XXI. Englobam os aspectos naturais e culturais e os entendem como processo que envolvem mente e corpo.

6.4 Desdobramentos contemporâneos

As três abordagens descritas na seção 6.3 geraram uma abertura conceitual na virada do século que permitiu uma variedade de interpretações sobre o conceito de emoção. Kleinginna e Kleinginna (1981) ainda no início da década de 80 distinguiram 101 definições e as organizaram em 11 categorias distintas, usando como referência a definição dada por cada um dos autores estudados. A tabela 6.1 abaixo apresenta um condensado da análise feita por Kleinginna e Kleinginna (1981):

Cornelius (2000) entende que essas categorias devem ser divididas em apenas 4 grandes áreas, a saber: *Darwiniana*, *Jamesiana*, *cognitiva* e *social construtivista*. Para o autor, ainda que exista distinções entre as definições de emoções, elas, necessariamente, devem ser alocadas em uma mesma conceptualização teórica, uma vez que o objeto é o mesmo. É importante destacar que a teoria social construtivista começou a ser pensada partir da década de 80 com os trabalhos de Averill (1980), Lutz (1988) e Thoits (1989). Sendo assim, podemos reorganizar a lista de Kleinginna e Kleinginna (1981) em torno das três linhas teóricas restantes, como se pode ver na tabela 6.2:

Categoria	Característica
Afetiva	As emoções são o resultado de estímulos psicológicos ou prazerosos.
Cognitiva	As emoções se manifestam a partir de avaliações do indivíduo sobre elementos presentes no ambiente.
Estímulo externo	As emoções são ativadas por estímulos externos ao sujeito.
Fisiológica	Buscam explicar as bases fisiológicas das manifestações emocionais.
<i>Behaviorista</i>	Visa avaliar padrões comportamentais que seriam respostas a situações de emoção.
Desruptiva	Entendem as emoções como uma resposta disfuncional diante de um determinado evento.
Adaptiva	As emoções têm a função de equilibrar um determinado estado de ruptura.
Multiaspectual	Diversos aspectos são ativados simultaneamente para que as emoções possam ocorrer.
Restritiva	Compreendem que as emoções são apenas um processo psicológico.
Motivacional	Veem um traço comum entre motivação e emoção.
Céticos	Questionam a necessidade de se definir um conceito tal como emoção.

Tabela 6.1: Resumo dos conceitos envolvendo emoção compilados no estudo de Kleinginna e Kleinginna (1981).

Darwiniana	Jamesiana	Cognitiva	Social-Construtivista
<i>Behaviorista</i>	Afetiva	Cognitiva	Social-Construtivista
Multiaspectual	Fisiológica	Restritiva	
Adaptiva	Desruptiva		
Motivacional	Estímulo externo		

Tabela 6.2: Categorias de definições descritas por Kleinginna e Kleinginna (1981) comparadas ao sistema de Cornelius (2000).

Nas seções apresentaremos cada uma delas de forma sucinta, buscando relacioná-las à questão da manifestação, problema central no estudo da fala emotiva.

6.5 A visão evolucionista

6.5.1 Ganho evolutivo

Como se pode notar grande parte das definições criadas e utilizadas no século XX giraram em torno das ideias de Darwin (1872/2016) e Lange e James (1922), a perspectiva que podemos chamar de evolucionista (CORNELIUS, 2000) e que visa integrar aspectos biológicos, neurais e, em última etapa, condições sociais na mesma abordagem. Nesse sentido, cabe definir que as emoções no século XX foram compreendidas como *sistemas evolutivos de controle da ação, manifestados por respostas fisiológicas automáticas a estímulos externos que contribuem para a sobrevivência*, ou seja, a hipótese central é a de que as emoções mantem estreita relação com o corpo que percebe o ambiente e sente as emoções que este entorno desperta.

No campo da neurociência, uma vasta literatura sobre o assunto tem demonstrado que a área responsável por estas respostas automáticas e pela compreensão das emoções é a amígdala, localizada na camada mais subcortical do cérebro humano e por este motivo uma das mais primitivas, justificando a noção de adaptação (LEDOUX, 1998, 2003; ANDERSON; PHELPS, 2001; TSUCHIYA et al., 2009; PESSOA, 2010). Damásio (2012), por exemplo, argumenta que a correlação entre estes núcleos pertencentes ao sistema límbico, como é o caso da amígdala, também responsável pela regulação bioquímica, e as emoções não é por acaso. Para o autor, foi necessário que nas primeiras fases da evolução, os seres vivos desenvolvessem respostas fisiológicas que gerassem comportamentos de autopreservação e que trouxessem alguma vantagem ao indivíduo: sentir fome ou sono são exemplos dessas autorregulações. Sendo assim, os indivíduos são potencialmente capazes de avaliar algo como positivo ou negativo, simplesmente por sua percepção somada ao comportamento somático que este objeto desperta.

As emoções seriam uma variedade destes mecanismos, com a diferença de que são regulações bioquímicas feitas pelo corpo a partir das informações dadas pelo ambiente. Damásio (2012) cita o hormônio *oxitocina* como exemplo de substância capaz de fazer estes controles emocionais diante de fatos externos aos sujeito. No caso, a *oxitocina* pode ser produzida em uma variedade de contextos, inclusive durante o ato sexual e o parto. Pedersen et al. (1982), Insel (1992), Insel e Shapiro (1992) e Insel e Young (2001) argumentam que determinados comportamentos, como a manutenção ou não de um único parceiro e o cuidado ¹⁵ com a prole, são derivados da liberação de *oxitocina* nos contextos citados. Em resumo, a relação de amor existente entre indivíduos pode ser explicada pela liberação deste hormônio, associados à ativação dos sistemas de reforço positivo decorrentes deste comportamento.

O ganho evolutivo, portanto, concentra-se na possibilidade de se julgar situações possivelmente prejudiciais ou possivelmente positivas, pela simples descarga de neuropeptídeos. Desta maneira, um indivíduo não precisa experimentar todas as possibilidades caso-a-caso, reservando a novas situações o julgamento por meio da experiência. Além disso, a descarga destes neurotransmissores é responsável por um comportamento "padrão", uma resposta reflexa ao estímulo. Por este motivo, é essencial que essas respostas derivadas das liberações dos reguladores químicos sejam transmitidas de geração em geração, como um "conhecimento inato"(DAMÁSIO, 2012, p.119). Ainda nessa linha, é importante frisar que estas respostas não podem ser variadas *ad infinitum*, pois isso traria um problema de significação, isto é, uma variabilidade contínua não permite uma percepção contínua. As manifestações das emoções, objeto primeiro destes estudos, é importante justamente por este motivo.

Sendo assim, as melhores respostas corporais aos estímulos externos foram selecionadas como sendo aquelas dotadas de importância e reproduzidas através do hábito de geração em geração. Por exemplo, a presença de um lobo com os dentes à

¹⁵Insel e Young (2001) utilizam o termo *attachment* para explicar as relações de cuidado estabelecidos pelos pais, em especial a mãe, com a sua prole.

mostra é processada pelo sistema límbico, ativando os mecanismos representacionais "instalados" na amígdala, levando à liberação de neuropeptídeos específicos, que ativam determinados estados do corpo, os quais correspondem ao medo, seguido de uma avaliação cognitiva de todo o cenário, embora essa hipótese tenha sido feita posteriormente. Esses mecanismos representacionais têm de ser, necessariamente, inatos, uma vez que se fossemos depender da experiência, teríamos de aprender todas as ativações emocionais em cada cenário possível no curso de nossa vida.

Se partirmos deste princípio é preciso assumir que existe, então, um sistema pré-conectado ao cérebro humano, capaz de fazer avaliações diante de dados objetos e disparar certas respostas corporais associadas a esses objetos, essa hipótese também nos leva a assumir que algumas emoções são universais (DARWIN, 1872/2016; TOMKINS; MCCARTER, 1964; EKMAN; FRIESEN, 1971; EKMAN, 1992). É possível também afirmar que não haveria, portanto, um estado neutro, pois a todo momento estaríamos sendo excitados por estímulos externos, os quais causam algum tipo de reação. Damásio (2012, p.130) cita, por exemplo, a extensão como um desses disparadores: um vulto, maior que a extensão do nosso corpo, pode disparar manifestações associadas ao medo.

Neste ponto podemos fazer um aparte, pois é natural que passemos a compreender os modelos *behavioristas* como correlatos da teoria evolucionista das emoções, uma vez que eles entendem que estímulos podem corresponder a reforços, os quais geram comportamentos. No entanto, o pressuposto assumido pelos *behavioristas* é o de que o estímulo que elicitava uma resposta emocional positiva, por exemplo, pode elicitar um reforço negativo ou positivo, dependendo do contexto. Assim, o responsável pelo condicionamento instrumental de um comportamento é a resposta ativada pelo reforço. Deste modo, é o reforço que gera um comportamento e não a resposta emocional (STAATS; EIFERT, 1990). O vulto citado acima, pode disparar o medo, mas se ele não se revelar perigoso, pode não elicitar um comportamento como o de correr.

Em nosso entender, a dificuldade teórica neste ponto é a de assumir que as respostas fisiológicas são as únicas responsáveis pelas respostas linguísticas, objeto deste trabalho, pois, sendo assim, teríamos de aceitar a hipótese de que a linguagem pode ser vista, em última análise, como um comportamento adquirido. Sendo assim, as manifestações emocionais baseadas na linguagem também de teriam de ser adquiridas pela experiência e somente pela experiência e estariam limitadas a variações positivas, isto é, não poderia haver interpolações ou ambiguidades, em aspectos mais funcionais, como a Morfologia, é possível assumir essa teoria, porém em aspectos que possuem um arcabouço de recursos limitados, como a Fonética, essa parece ser uma sugestão inviável.

Porém, como se vê, as teorias *behavioristas* compreendem a manifestação da emoção como um segundo nível. Sendo assim, para que as manifestações linguísticas das emoções sejam consideradas comportamentos é necessário que os estímulos levem a reforços, que levem a utilização da linguagem de modo tal que seja reforçada pelo uso entre falantes. Em nosso entender, essa perspectiva viola, justamente, o problema da variabilidade, pois condicionaria o uso da linguagem ao reforço dos falantes, sendo que esse reforço poderia ocorrer intra-grupo e nunca inter-grupo. O que vemos é justamente o oposto, como demonstrado por Paidi, Kadiri e Yegnanarayana (2016), existe uma grande regularidade nas manifestações vocais.

Retomando a discussão. Dentro da visão evolucionista, portanto, as emoções são reações comportamentais de duas dimensões: (i) um *estímulo externo* que gera (ii) *processos neurais*. Pensando na questão do ganho evolutivo, a correlação entre essas duas dimensões precisa inequívoca, isto é, duas reações distintas não podem servir de resposta para um único estímulo, pois violariam o pressuposto da qualidade. Da mesma forma, a universalidade deve ser pressuposta nas duas pontas, pois é preciso supor que determinados estímulos são encontrados de maneira extensiva em variadas culturas, bem como as suas manifestações. Essa hipótese leva invariavelmente, a sugestão de que algumas emoções seriam "básicas", por serem invariáveis, inerentes

aos seres e essenciais para manutenção da vida.

6.5.2 As Emoções Básicas

A ideia de emoções universais, básicas ou primárias está presente em quase toda filosofia ocidental. Aristóteles (1991) já propunha algo semelhante em sua Retórica, o mesmo foi feito por Santo Agostinho, retomado por Darwin (1872/2016) e formalizado como teoria por Ekman (1992). Esta hipótese trata de duas perspectivas essenciais: a primeira lida com os aspectos biológicos e neuronais da manifestação das emoções, enquanto a segunda trata dos aspectos psicológicos, que são as estruturas fundamentais que formam as emoções secundárias (DAMÁSIO, 2012).

Do ponto de vista biológico, a análise é basicamente a que foi descrita no item acima e busca encontrar mecanismos funcionais, tais como hormônios e neurotransmissores, que determinem a forma como manifestamos, lidamos com as emoções e o papel desses mecanismos na sobrevivência das espécies (EKMAN, 1992). Do ponto de vista psicológico, o objetivo é encontrar um grupo limitado de emoções que não podem ser divididas em unidades menores, de modo que essas unidades são os elementos combinatórios para construir emoções mais complexas (PLUTCHIK, 2001).

Darwin (1872/2016) e, posteriormente, Ekman e Friesen (1969) já demonstravam evidências de que as expressões de algumas emoções, bem como o reconhecimento delas tem uma tendência universal, embora algumas características possam sofrer variações culturais. Ekman e Friesen (1971) conduziram um teste, em que apresentavam imagens de ocidentais que representavam emoções, como *happiness*, *anger*, *sadness*, *disgust*, *surprise* e *fear*¹⁶, junto de outras duas imagens distratoras a membros de uma tribo iletrada da Papua Nova Guiné. Estes indivíduos foram divididos em duas categorias, aqueles que tinham pouco ou nenhum contato com a cultura ocidental e aqueles que falavam inglês e tinham algum conhecimento da cultura ocidental.

¹⁶Optamos por não traduzir algumas emoções, a fim de evitar uma conceptualização equivocada

A tarefa era a mesma para os dois grupos e limitava-se a identificar qual figura entre as 3 melhor representava uma determinada história contada pelos pesquisadores. Os resultados desta pesquisa sugerem que não há uma diferença significativa entre-grupos, de modo que determinadas manifestações são associadas a emoções específicas. No entanto, os dados demonstram que o medo e a surpresa possuíam um grau de confusão entre os participantes, Ekman e Friesen (1971) sugerem que esse tipo de variação é causado por fatores culturais, por exemplo, para estes indivíduos um fato surpreendente é um fato que gera medo.

Em Ekman (1992) argumenta que seu modelo anterior a definição do que eram emoções básicas não era totalmente funcional, pois se pautava em dois pressupostos, o primeiro deles era a possibilidade de que as emoções estivessem em uma escala que vai do *prazeroso* ou *não-prazeroso*. Como citado acima, para evolucionistas, a noção de que as emoções têm como função adaptativa nos ajudar a lidar com situações possivelmente prejudiciais ou benéficas é fundamental, pois explica as emoções de forma filogenética, sem a necessidade de recorrer a ontogênese. O segundo pressuposto é o de que o significado das expressões faciais é culturalmente aprendido e socialmente variável, por isso, os participantes do experimento de Ekman e Friesen (1971) podem fazer algum tipo de confusão entre o medo e a surpresa, ou seja, que o significado é atribuído pelo contexto.

A questão da variação na manifestação das emoções é equivalente ao problema da variação linguística: não pode ser encaixada em um modelo universalista sem uma explicação contundente, do contrário representa uma negação da prova. Por este motivo, em Ekman (1992) o autor sugere que a própria definição de *básica* é fluída e comporta ao menos 3 definições. A primeira delas é a de que existe um número de emoções que difere umas das outras de forma significativa, não se pode confundir a raiva com a alegria, não apenas porque suas manifestações são distintas, mas, também, porque os eventos antecedentes que as elicitam são obrigatoriamente diferentes.

A segunda delas é a de que a evolução teve o papel de formar as características únicas e comuns a estas emoções, de modo geral, são os elementos biológicos que compõem todas as emoções e determinam suas funções em nossas vidas. Para Ekman essa função básica é a de intermediar os encontros inter-organismos (humano-humano, humano-animal, por exemplo), embora não se possa reduzir as emoções a essa função exclusivamente, já que podemos imaginar estados emocionais ou auto-elicitar algumas emoções. A terceira é a de que as emoções complexas são compostas dessas emoções básicas (PLUTCHIK, 2001), ou seja, uma emoção secundária, como a saudade, necessariamente, depende da tristeza para existir e por isso não pode ser uma emoção básica. No entanto, para o autor, apenas as duas primeiras definições são essenciais, uma vez que as emoções se desenvolveram para que o humano se preparasse para a ação. (EKMAN, 1992, p.171)

A combinação destas definições é a condição de existência das emoções básicas discretas, bem como é o que as difere dos demais estados afetivos. Junto disso, Ekman (1992) insere o conceito de *família de emoções*, pressuposto de análise que resolve o impasse com relação às variações. A partir desta ideia, as emoções básicas possuem algumas características universais e outras que variam de acordo com a ativação de determinados conjuntos musculares ou de expectativas daquele que sente uma emoção. O autor relata que em uma de suas pesquisas foi capaz de identificar 60 variações da expressão da raiva, porém todas elas mantinham um padrão significativo, como o abaixamento das sobrancelhas. No mais, as evidências de Ekman (1992) sugerem duas emoções nunca excitam os mesmos *padrões musculares* (EKMAN, 1992, p.172), garantindo o pressuposto da qualidade acima discutido.

Junto da hipótese de família de emoções e das 3 definições de básico, o autor sugere que são nove as características necessárias para se distinguir as possíveis emoções básicas dos outros estados afetivos. Seriam elas: (i) sinais distintos e universais, (ii) presença em outros primatas, (iii) fisiologia distintiva (iv) distinção universal em eventos antecedentes, (v) coerência entre respostas emocionais, (vi) rápido *onset*,

(vii) breve duração, (viii) avaliação automática e (ix) ocorrência não opcional.

Não vamos passar em revista por todas estas características, uma vez que isso alongaria a discussão, mas é importante destacar que a evolução teórica e metodológica apresentada em Ekman (1992) restringe significativamente a quantidade de emoções que podem ser consideradas básicas. Em nosso entender, isso ocorre, principalmente, porque existe uma grande dificuldade em atribuir às emoções associadas a situações benéficas um papel relevante na sobrevivência das espécies. A cooperação baseada na demonstração de felicidade existente entre cães e humanos ou as trocas afetivas de relações monogâmicas, podem ser exemplos, mas elas não possuem características distintivas únicas. O sorriso, por exemplo, é uma característica propriamente humana.

Damásio (2012) não parte do mesmo pressuposto e sustenta que a felicidade, a tristeza, a cólera, o medo e o nojo são as emoções mais básicas e mais associadas à sobrevivência das espécies. Ekman e Friesen (2003) sugeriam o número de 6 emoções: raiva, medo, surpresa, alegria, nojo e tristeza, já para em Ekman (1992), o autor sugere que, as evidências apontam, sem dúvidas, para a raiva, o medo, o nojo e a tristeza, com alguma incerteza para emoções positivas, como a alegria e que precisam de mais dados empíricos, para emoções como a surpresa, desprezo, culpa e vergonha, uma vez que elas atendem as 9 características citadas acima. (ARNOLD, 1960) propõe que as emoções são em total de 11, enquanto (JAMES, 1884) sugere um número de 4 (medo, dor, raiva e amor) e Aristóteles (2005) citava o número de 14, ou seja, o critério de definição parece ser relativamente aleatório.

Recentemente, Paul Ekman, lançou uma plataforma *on-line* chama de *Atlas of Emotions*¹⁷. O objetivo é antes comportamental do que acadêmico, a iniciativa tem como objetivo conscientizar as pessoas comuns sobre a importância do uso adequado do "vocabulário das emoções". A proposta é ajuda-las a compreender os sentimentos despertados por um evento, saber nomeá-los e compreender melhor o resultado social da manifestação da emoção. Para estabelecer as famílias de emoções presentes

¹⁷<http://atlasofemotions.org/>. Acessado em jun/17.

na plataforma, Ekman fez uma pesquisa quantitativa (EKMAN, 2016), buscando compreender o grau de concordância entre pesquisadores da área. Os resultados estão compilados na tabela 6.3

Emoção	Grau de Concordância
Raiva	91%
Medo	90%
Nojo	86%
Tristeza	80%
Alegria	76%

Tabela 6.3: Grau de concordância entre cientistas sobre quais emoções são universais

Nota-se pela tabela 6.3 e pelo que foi acima debatido que a raiva e o medo parecem ser as emoções com o maior grau de aderência entre pesquisadores. Isso sugere que suas manifestações e eventos antecedentes são bastante definidos ou que as respostas corporais estão enraizadas profundamente em nosso sistema nervoso e sensorial, de tal modo que pesquisadores encontram uma regularidade significativa entre elas e as suas funções evolutivas. Se retomarmos a hipótese do ganho evolutivo, podemos sugerir que as duas alternativas estão corretas, uma vez que as duas sugestões são contíguas e uma pressupõe a outra, isto é, elas fazem parte da evolução e têm manifestações bem definidas. No entanto, este não pode ser um critério de definição objetivo, uma vez que as variações interculturais ou mesmo inter-espécies podem tomar parte.

Aliás essa é uma das críticas recorrentes à proposta das emoções básicas e discretas, os critérios biológicos que ajudam a definir quais são e, por consequência quantas elas são, variam de acordo com o método de validação ou grau de significação que se dá um ou outro critério. Como demonstramos logo acima, o critério que define as duas respostas é essencialmente autoral, embora o pressuposto essencial seja a do que ela deve estar calcado na suposição biológica ou evolutiva. Ekman (1992), por exemplo,

em suas 9 características se apropria de um conjunto de elementos que envolvem os mecanismos globais de ativação e filogenéticos, enquanto Damásio (2012) sugere que as emoções básicas podem ser definidas a partir da análise de lesões no hipotálamo e no cíngulo basal. Dois critérios biológicos, duas noções.

Em certo sentido, parece-nos que o conceito fundamental dessas pesquisas é a busca pela definição de um elemento que explique automaticamente todas as emoções básicas, enquanto o critério fundamental parece ser na realidade o de regularidade. Isto é, o quanto uma emoção é capaz de desempenhar um papel de regulação da sobrevivência, ao mesmo tempo que atende aos pressupostos associados a sua condição de existência. No caso do estudo da fala emotiva, a definição de regularidade é fundamental, pois ao analisarmos as emoções como manifestações linguísticas, precisamos supor que as suas implementações não serão necessariamente universais já que a natureza das línguas é variável, mas que de alguma forma podem manter uma regularidade entre línguas, como sugerem Scherer (2005) e Paidi, Kadiri e Yegnanarayana (2016).

No mais, a hipótese das emoções básicas tem sido fundamental no estudo da fala emotiva justamente por este aspecto ambíguo: a variabilidade das línguas pode ser neutralizada a partir da suposta invariabilidade das emoções básicas, criando um arcabouço de análise parecido com aquele que analisa as expressões faciais. No entanto, a contribuição entre estas duas áreas parece se esgotar aí, uma vez que os estudos da linguagem se veem incapacitados de transportar para as funções biológicas das emoções os aspectos culturais e sociais, profundamente enraizados nas línguas.

Nesse sentido, acreditamos que uma solução para a compreensão da manifestação das emoções, seja como produtos linguísticos ou como expressão corporal é supor a existência de um nível cognitivo derivado dessa tendência natural a se sentir e demonstrar algumas emoções que funcionam como reguladores da existência. Para isso, precisamos supor que existam mecanismos cognitivos que coordenam uma série de processos que ajudam a manifestar de forma corporal. Podemos questionar, junto

disso, se essa manifestação não é afetada socialmente, se pensarmos que uma parcela dos fenômenos humanos/naturais, como a linguagem e a música, são amplamente condicionadas por fatores locais que podem alterar sua forma final, ainda que se mantenha uma forma subjacente. Na seção a seguir apresentaremos essa questão.

6.6 A questão cognitiva e social

O argumento de que as emoções tiveram um papel evolutivo que contribuiu sobremaneira para a manutenção da vida e, por isso, têm suas origens tem uma explicação filogenética, é bastante adequado e explica uma série de correlações encontradas na manifestação das emoções de homens e animais. Freud, em certo sentido, compreendia que a principal dificuldade não era a explicação da origem das emoções e, sim, a partir de que ponto a relação entre elas, a manifestação e a consciência dessa relação se tornou indispensável para o homem. Dentro de sua teoria, essa intrincada rede de relacionamento terá um importante impacto na definição da diferença entre do sistemas conscientes (*C's*) e inconscientes (*Ic's*) (FREUD, 2015a, p.115).

Nessa etapa de formulação teórica Freud compreendia que determinados eventos psíquicos estavam mergulhados em sistemas inconscientes, incapazes de serem notados pelo indivíduo, de modo tal que a percepção é derivada de mecanismos de censura, os quais permitem que determinados eventos atingissem os sistemas de consciência. Em resumo, o trabalho de análise clínica é desvendar essas relações e analisar os eventos emergentes, os quais são a etapa *física* da mente. O trabalho de análise psicológica seria descrever esses processos.

James (1890) tinha uma visão diferente, pois acreditava que a relação psique/física era quasi-direta, isto é, os eventos psicológicos, como as emoções, eram realizações psíquicas geradas a partir de determinados processos físico-fisiológicos. Sendo assim, não existe um mecanismo de articulação entre estas esferas, o qual mudasse as possíveis saídas. Eventos psíquicos sem relação com os sintomas eram considerados

processos somáticos (FREUD, 2015a, p.102). Em tese, dentro de uma visão puramente evolucionista, as emoções devem se tornar conscientes para um indivíduo, ao menos a manifestação delas, uma vez que é a percepção que determina a mudança de estado do indivíduo, ou seja, as emoções são eventos biológicos sem correlação com qualquer mecanismo psicológico, pois são o evento psíquico em si.

Freud via nisso um problema, pois entendia que existe a possibilidade de que algumas respostas físicas escondam em outros processos mentais a motivação original. Para ele, a emoção precisava ser compreendida como um bloco psíquico à parte, um elemento que constitui a *psique* e ao mesmo tempo mantém uma conexão com outras unidades, como o inconsciente ou as pulsões, de modo tal que diante de um estímulo elas podem emergir ou não. Não há uma definição clara do que são as emoções em Freud (2015a), uma vez que o autor acreditava que o estado de artes tornava impossível uma asserção definitiva sobre o assunto. Por este motivo, o autor assume que "ideias [inconscientes] são investimentos - de traços mnemônicos, no fundo -, enquanto os afetos e sentimentos correspondem a processos de descarga, cujas expressões finais são percebidas como sensações"(FREUD, 2015a, p.117).

As ideias inconscientes mantêm uma forte relação com os elementos externos ao corpo, já que transportam consigo alguns traços destes objetos. Os sonhos, por exemplos, são carregados de imagens "reais"processadas pelo inconsciente. Já as emoções são descargas internas que ativam processos motores para a alteração do estado interno e que levam o sujeito à ação (FREUD, 2015a). A questão é que esta "mobilidade"pode ser freada pelos sistemas que regulam a passagem da inconsciência de um afeto para a sua consciência (JOHNSON-LAIRD; MANCINI; GANGEMI, 2006). Neste sentido, as emoções são um fato psíquico e biológico separados, já que podem ser barrados por mecanismos internos ao indivíduo e não dependem de um estímulo externo(FREUD, 2015a; DAMÁSIO, 2012).

Dentro desta perspectiva, que separa os afetos e a consciência deles, Damásio

(2012) entende que é preciso compreender a emoção como um intrincado processo de composição entre (i) *instintos* e (ii) *impulsos*. Os instintos, de acordo com Freud, estão diretamente relacionados ao *Id*, a parte de nossa mente mais irracional, conectada aos instintos sexuais e ao inconsciente. Os impulsos são as tendências de ação, resultantes das direções do *Id* e são controladas, em grande parte do tempo, pelo Super Ego (SE), a parte da mente responsável pelo controle moral (FREUD, 2016), a fim de que se evite uma manifestação exacerbada dos instintos (DAMÁSIO, 2012, p.123). Freud também usa o termo *descarga* (FREUD, 2015a, p.117), de modo que a exteriorização do afeto é feita a partir da descarga motora (respostas fisiológicas) para alteração do corpo, sem relação (direta) com o exterior.

É importante destacar que Freud não via a possibilidade de que as emoções fossem puramente inconscientes (FREUD, 2015a, p.115). A excitação de um estímulo externo associado à resposta reflexa derivada deste estímulo, mantinha-se a mesma, porém processos como a repressão (FREUD, 2015b) contribuem para que a emergência delas à consciência fosse distinta. Neste caso, a percepção consciente é perturbada por um processo inconsciente, o qual não muda o afeto original, mas muda a direção deste afeto e, por consequência, a sua manifestação. O sentimento de raiva, por exemplo, pode esconder algum tipo de admiração esquizoide, que poderia ser o afeto original. A nosso ver, a possibilidade de que as emoções passem por um "filtro perceptivo", seja ele consciente ou não, parece ser um problema central no modelo que considera que as emoções são exclusivamente excitações corporais e respostas reflexas ou unidades básicas proto-instaladas.

Sendo assim, é possível assumir a hipótese de que as emoções, como instintos, são o resultado de um processo biológico e evolutivo, tal qual como pensado por Darwin (1872/2016), mas que a emergência dos seus significados e a variedade de suas manifestações, o resultado dos impulsos de ação, são frutos de um segundo processo não especificado, mas que tem a função de fazer avaliações do conjunto ambiente-corpo. Podemos pensar em três mecanismos básicos para a emergência

das emoções à consciência e da análise de suas manifestações, os quais são descritos abaixo.

1. O primeiro destes mecanismos é resultante de processos como a repressão (FREUD, 2015b) e o Super Ego (FREUD, 2016). No entanto, parece-nos que o primeiro mecanismo está mais associado às psicopatologias do que as manifestações tradicionais, resultantes da diligência do SuperEgo. Freud Freud (2015a, p.117) argumenta que as emoções são descargas psíquicas e suas expressões finais é que são percebidas como sensações, isto é, o sentir das emoções não pode ser entendido como um processo com direção determinada: $[Emo \rightarrow Sens]$. Assim, quando as sensações sofrem alterações decorrentes da regulação de outros mecanismos, é provável que a resultante seja diferente do afeto inicial. As causas para essa mudança são variadas, mas deve se assumir que elas funcionam como sistemas de autopreservação do Ego.
2. O segundo mecanismo possível está associado a processos cognitivos que levam em conta elementos individuais, como a memória e a personalidade, por exemplo. Em nosso entender, são os pressuposto da teoria cognitiva, demonstrada na tabela 6.2 e colocam a emoção em uma definição restrita. Neste caso, ela é uma parte da mente e está profundamente conectada com outras faculdades, mas não é um processos global, como imaginam alguns neurocientistas que veem na emoção uma função regulatória global (LEDOUX, 2003).
3. O terceiro está relacionado diretamente aos fatores sociais, isto é, a manifestação das emoções está mais associada a elementos coletivos do que a elementos individuais. Neste caso, os componentes cognitivos são partes acessórias que podem ou não entrar em cena na definição das emoções, enquanto os fatos sociais contribuem para a regulação da manifestação e compreensão delas (RUSSELL, 1991).

6.6.1 A hipótese cognitiva

A partir da lista apresentada na seção anterior podemos notar que Freud (2015a) pensava as questões da mente a partir de uma visão totalmente internalizada para uma totalmente externalizada. Quanto mais internalizada, mais próximo estamos do domínio inconsciente e, por isso mesmo, mais distantes de uma percepção objetiva. Do outro lado, quanto mais externalizado, mais próximo estamos do aspecto social e, por isso, mais próximos da percepção e de uma autorregulação baseada em um conjunto de axiomas construídos ao longo de nossas vidas.

Não vamos nos aprofundar no primeiro aspecto (1), pois entendemos que as psicopatologias não são escopo deste trabalho. Destacamos que diversos estudos da manifestação das emoções por um viés clínico têm revelado uma variedade significativa de dados que explicam alguns fenômenos, como é o caso a teoria da Super-Emoção (JOHNSON-LAIRD; MANCINI; GANGEMI, 2006) e das variações de comportamento (EGGER; ANGOLD, 2006; FRIEDMAN et al., 2003; HOLLON; PONNIAH, 2010; STORCH et al., 2012).

Desde a chamada “Revolução Cognitiva” uma das principais hipóteses que se tem assumido é a de que o comportamento não é aprendido apenas a partir dos estímulos externos e dos reforços presentes em nosso entorno, como supõem evolucionistas e, principalmente, *behavioristas*, mas, sim, um produto derivado de um número finito de processos menores, que são universais e estão “instalados” em nossa mente de forma inata e são os responsáveis por gerar uma variedade infinita de comportamentos (PINKER, 2002).

Embora cognitivistas tenham buscado demonstrar que o comportamento, seja ele funcional ou linguístico, não pode ser inteiramente explicado pela ideia de reforço e relação com o ambiente (CHOMSKY, 1959a), dentro da teoria das emoções essa hipótese não foi completamente descartada. Para cognitivistas, como Arnold (1960), as emoções são essencialmente elementos que nos levam à ação a partir do estímulos.

Porém, é a *avaliação* destes estímulos como positivos ou negativos que determina a direção dessa ação. Para Arnold (1960) a emoção é:

the felt tendency toward anything intuitively appraised as good (beneficial), or away from intuitively appraised as bad (harmful). This attraction or aversion is accompanied by a pattern of physiological changes organized toward approach or withdrawal. The patterns differ for different emotions.(ARNOLD, 1960, p.182)

A diferença na definição de Arnold (1960) é inclusão da noção de *appraisal*, o que traduzimos como *avaliação cognitiva*. A autora entende que um problema central no modelo de James (1884) era a falta justamente de uma especificação sobre o que ele chama de *percepção*. Para a autora, esse elemento intermediário era um processo cognitivo e não apenas uma resposta reflexa, o qual ela chamou de *appraisal* e foi definido da seguinte maneira por Mesquita e Frijda (1992):

Events, or events as coded, are appraised with evaluation of events in terms of relevance for the subjects's well being and his or her possibilities for coping with the event. (...) Appraisal processes, in the sense that they are meant here, can be conceived of as a series of checks with respect to a set of dimensions such as positive or negative valence, causation by someone else or the self, blameworthiness, outcome uncertainty, controllability, and modifiability"(MESQUISTA; FRIJDA, 1992, p.180)

Arnold (1960) sugere que as emoções são resultado da avaliação cognitiva dos eventos que desencadeiam as reações associadas às emoções e não respostas reflexas a esses estímulos, ou seja, a manifestação de uma emoção está associada ao significado que se atribui a um determinado evento. A autora entende que esse processo de avaliação é um "sense judgmente", por isso são "direct, imediate, nonreflective, nonintellectual and automatic" (ARNOLD, 1960, p.175). No mais, esse mecanismo de avaliação é limitado, de modo que seria possível mapear seus primitivos.

Além disso, a avaliação está associada a outras dimensões do *self* como a personalidade, o temperamento e o comportamento. Desta forma, *appraisal* refere-se a ativação ou não de padrões de respostas, a partir da decodificação de eventos externos, baseado em uma série de fatores, como a *valência* (o grau de conveniência), *ativação* (urgência) e

poder (coping) (SCHERER, 2001) e podem explicar a variabilidade das emoções interculturais. Sendo assim, a raiva, por exemplo, envolveria a avaliação de uma situação ou de um evento sendo desagradável (unpleseant), como sendo responsabilidade de outra pessoa e de um modo que envolve algum esforço. Já a culpa envolveria a mesma avaliação, mas requisitando apenas um pequeno esforço. (SMITH; ELLSWORTH, 1985)

A crítica mais usual a essa hipótese é a de que essa avaliação cognitiva torna a manifestação das emoções um processo complexo demais. Sendo assim, a resposta para um vulto ameaçador pode ser demasiado longa, a ponto de se perder exatamente a função essencial da emoção, que é a de preparar o corpo para ação (LEDOUX, 1998). A contrarresposta é a de que esses processos são automáticos, estão instalados em nossas mentes de maneira inata, tal como a capacidade de se adquirir a linguagem ou de se localizar espacialmente (ARNOLD, 1960). A tréplica dada por neurocientistas é a de que a hipótese dos processos automáticos, necessariamente os coloca como mecanismos biológicos e não somente cognitivos, o que prova que as emoções são produtos evolutivos e funcionam ou, ao menos teriam funcionado, como respostas reflexas (LEDOUX, 1998). A polêmica explicação filogenética na ontogênese (GOULD, 2010).

A segunda crítica está na definição de *avaliação cognitiva*. Para cognitivas como Lazarus (1993), *appraisal* é “on the one hand, the demands, constraints, and resources of the environment and, on the other, the goal hierarchy and personal beliefs of the individual” (LAZARUS, 1993, p. 5). Neste sentido, é difícil compreender se esse processo é um núcleo capaz de coordenar diversas unidades, como a percepção e o temperamento, ou se estamos lidando com um mecanismo próprio, com características encerradas em si e não diretamente relacionadas com outras.

A primeira resposta parece ser a mais adequada e para explicar as emoções dentro desta hipótese uma série de tentativas teóricas e experimentais têm sido feitas, com o intuito demonstrar que a avaliação não é um componente isolado (WALLBOTT;

SCHERER, 1986; SCHERER, 1999, 2001). A tabela 6.4 resume como a avaliação se relaciona com outros componentes que tem uma participação ativa na elicitação e manifestação das emoções, de acordo com Mesquita e Frijda (1992).

	Componente	Função
1	Eventos antecedentes	Seriam aqueles eventos, os quais elicitam algum tipo de emoção, seja em um indivíduo, seja em um conjunto.
2	<i>Event coding</i>	É a categorização do evento acima citado, isto é, como o seu significado é reconhecido pelo indivíduo
3	<i>Appraisal</i> - Avaliação	Trata-se da avaliação cognitiva do evento ou do evento codificado. Os danos e ganhos possíveis com o evento, por exemplo
4	Padrão de reação fisiológica	Emoções envolvem respostas fisiológicas, é possível afirmar que algumas delas são padrões
5	Preparação para a ação	Emoções envolvem mudanças de estado, seja na ativação de ações ou na retração do corpo
6	Comportamento emocional	Padrão de comportamento associado a uma emoção
7	Regulação	Regulação das emoções, seja da inibição de controles ou do reforço voluntário de determinadas ações associadas às emoções - seja por aspectos socioculturais ou pessoais

Tabela 6.4: Etapas que compõem a manifestação das emoções dentro de uma análise cognitiva

Em seu artigo, Mesquita e Frijda (1992) demonstram que estes componentes têm tendência universal, isto é, a capacidade de decodificar um determinado evento e

a forma de reagir a ele é bastante regular quando olhamos interculturais, embora algumas diferenças possam ser notadas (MESQUISTA; FRIJDA, 1992, p.180). Assim, de um modo geral, um lobo com os dentes à mostra é decodificado como perigoso e por isso causa medo. Esse dado é positivo, pois confirma a hipótese cognitiva de que certas avaliações são universais, por outro lado, parecem apenas confirmar aquilo que evolucionistas já davam como certo, que essas funções tem origem filogenética. Desta maneira, um dos desafios da teoria cognitiva das emoções é justamente demonstrar a necessidade ou o papel da avaliação dentro da manifestação das emoções.

Seguindo essa linha, Scherer (1987, 2001) sugere que as emoções são mecanismos que evoluíram filogeneticamente e que servem como respostas aos eventos do meio, separando estímulos de repostas e assim criando uma latência que permite uma otimização destas respostas. Por isso, o autor considera que as emoções são:

(...) *an episode of interrelated, synchronized changes in the state of all or most of five organismic subsystem in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism.*
(...) the term “emotion” only for those periods of time during which many organismic subsystems are coupled or synchronized to produce an adaptative reaction to an event that is considered central to the individual’s well-being. (grifos do autor) (SCHERER, 2001, p. 93).

Scherer (2005) entende que os *componentes* das emoções são os respectivos estados dos cinco subsistemas citados acima e especificados na tabela 6.5, ao passo que a coordenação das mudanças ao longo do tempo é o *processo*, por este motivo, tem denominado o seu modelo de componente-processo.

Para Scherer (2001) o primeiro componente, o de avaliação cognitiva, não está somente associado as noções de valência, ativação e poder, como suposto por Arnold (1960), mas, sim, ao que o autor chama de *Stimulus Evaluation Checks*, que são os *objetivos da avaliação* e se ocupa de analisar as informações de um objeto ou evento com a intenção de preparar uma reação apropriada. Esta análise leva em conta 4 aspectos: *relevância, implicações, possíveis danos e significado normativo*, sendo que todos eles estão relacionados com o sujeito e com o sujeito e o ambiente social. Assim, a manifestação

Função da emoção	Componente da emoção	Subsistema orgânico
Avaliação dos objetos e eventos	Componente cognitivo (avaliação)	Processamento da informação (SNC)
Regulação sistêmica	Componente neurofisiológico (alterações corporais)	Suporte (SNC, SNE, SNA)
Preparação e direção da ação	Componente motivacional (tendência de ação)	Executivo (SNC)
Comunicação da reação e intenção comportamental	Componente da expressão motora (facial e vocal)	Ação(SNS)
Monitoramento do estado interno e da interação do organismo com o meio	Componente da sensação subjetiva	Monitor (SNC)

Tabela 6.5: Relação entre subsistemas orgânicos e as funções e componentes da emoção, segundo Scherer (2005)

emocional é condicionada por este sistema de avaliação dos estímulos.

Como se pode ver, o modelo de Scherer (2001) é condicionado a esse sistema de avaliação do estímulo, ou seja, pelo componente cognitivo. Novamente, as críticas que podem ser feitas a esse tipo de hipótese é o excesso de etapas de avaliação que hiper-intelectualizam a sensação e manifestação das emoções. Desta maneira, o significado de vulto seria condicionado à avaliação que o indivíduo faz dele, bem como a manifestação da emoção seria condicionada à avaliação que ele faz da necessidade de manifestação. Um indício de que o aspecto social toma parte na manifestação das emoções.

Damásio (2012) entende que essa etapa cognitiva está mais associada ao quinto componente das emoções do que ao primeiro, pois para o autor a regulação das emoções, depende necessariamente da sensação delas (DAMÁSIO, 2012, p.130). De maneira geral, as evidências sugerem que o conceito de *appraisal* é muito mais dependente de uma sugestão do observador, do que daquele que experiencia uma emoção, junto disso algumas diferenças culturais colocam em xeque esse tipo de hipótese, enquanto

outras parecem reafirmar sua validade (MESQUISTA; FRIJDA, 1992). A nosso ver, o conceito de avaliação cognitiva como proposto por Scherer (2001) parece operar em dois níveis dificilmente integráveis.

O primeiro deles é o nível filogenético, onde as emoções são elicitadas com base em nosso histórico genético. Por exemplo, a ação de lamentar a perda de um ente querido parece ser universal ou de temer a morte, uma vez que a compreensão de vida e morte parece ser um conceito inerente à humanidade. O segundo nível é cultural, em que aspectos sociais ou culturais interveem na forma como se avalia um estímulo. Matsumoto (1989) cita, por exemplo, as diferenças na forma como americanos e japoneses avaliam a responsabilidade pela elicitação de uma determinada emoção. Japoneses entendem que esse não é um critério a ser considerado, enquanto americanos acham que eles mesmo ou outra pessoa pode ser a responsável.

Markus e Kitayama (1991) entendem que isso é uma característica associada a forma como o *self* é construído dentro de uma cultura. Na cultura ocidental o *self* é amplamente baseado na ideia de independência, isto é o foco recai na experiência do indivíduo como elemento essencial, sendo assim sua avaliação é baseada na suas conquistas individuais. Em outras culturas, o indivíduo baseia na interdependência, quando sua experiência é focada na relação inter-grupo, nestes casos, sua avaliação foca-se nos objetos coletivos e na relação interpessoal. Desta maneira, os indivíduos dessa segunda categoria possuem emoções que levam em conta essa avaliação de grupo, como é o caso de *fureai* - sentir se próximo de alguém - e *tukeagari* - sentimento de alguém que se coloca acima do grupo.

Existe uma grande dificuldade dentro da comunidade científica de incorporar o aspecto biológico ao social. Desde o *Contrato Social* de Rousseau (ROUSSEAU, 2011) se intensificou e se teorizou a distância entre essas dimensões, além da oposição categórica entre *natureza* e *sociedade* já estabelecida por Platônicos e posteriormente cartesianos. Assim, a explicação mais plausível para o aspecto social é o de que as

emoções sejam de fato produtos socio-culturais e não apenas resquícios de evolução associada a um processo cognitivo.

6.6.2 A teoria sócio-construtiva

Averill (1980) entende que as emoções não podem ser compreendidas apenas como respostas fisiológicas ao ambiente, pelo contrário, elas são construções sociais e, por este motivo, podem ser compreendidas como tal. Em seu modelo, Averill (1980) faz uma adequação à noção de avaliação, entendendo que ela não elicit um mero padrão de resposta fisiológica, mas, sim, representa a interpretação de um estímulo. Desta maneira, a raiva, por exemplo, é o julgamento de que algum comportamento esperado pela outra parte não foi cumprido de forma intencional, assim a manifestação da raiva é uma forma de contrabalancear essa dinâmica, incluindo uma retribuição na conta.

Para o autor, mesmo outras dimensões, como a percepção que temos da manifestação das emoções, podem ser compreendidas nessa chave. O fato da raiva ser entendida como um momento de descontrole, na realidade é apenas uma interpretação que temos da própria manifestação da raiva e não necessariamente um dado factual. Portanto, as emoções são determinadas por regulações e normas sociais, antes de mecanismos cognitivos ou fisiológicos, sendo que o principal delas é justamente equilibrar as relações e os comportamentos coletivos e interpessoais. O amor romântico, por exemplo, oferece uma relação estável em troca de uma porção de nossa liberdade social e econômica (AVERILL, 1985).

As emoções, portanto, para Averill (1980) são construtos sociais, derivados da nossa relação com o meio e com os actantes em cena. Por este motivo, as emoções não são meros mecanismos fisiológicos, são elementos complexos determinados por uma série de avaliações que operam tanto no nível social, quanto no interpessoal e contribuem para a regulação dessas relações. A cultura nesse processo atua como fator determinante para o estabelecimento da variabilidade das emoções, ou seja, a cultura

é o fator explicativo do porquê uma determinada emoção é vista de duas formas diferentes, como é o caso do medo e da surpresa, citados por Ekman e Friesen (1971).

Mesmo evolucionistas e cognitivistas tendem a concordar que a emoção tem um fator social intrínseco a ela. No entanto, nenhum deles seria capaz de defender que as emoções são unicamente construtos sociais que ajudam a regular as nossas relações. A questão parece mais complexa do que isso. Se cognitivistas são “acusados” de hiper-intelectualizar a forma como lidamos com as emoções, costuma-se entender que social-construtivistas hiper-simplificam a questão.

Atualmente as emoções têm um novo *status* dentro de nossa sociedade e se tornaram produtos culturais. Compreender os eventos que despertam uma emoção, a manifestação delas e o papel que elas desempenham em nossas relações é tema de livros e cursos que oferecem algum tipo de satisfação rápida para questões mentais. A *Inteligência Emocional* compreende, em grande parte, a ideia de que alterar o curso das emoções é fundamental para que possamos nos integrar socialmente de forma salutar, ou seja, que as emoções, em última análise, são produtos gerados e regulados pelas relações sociais única e exclusivamente.

Obviamente, não cabe a nós fazermos uma crítica mais aguda sobre o tema nesse trabalho, mas é importante destacar que estamos diante de um momento em que as emoções são analisadas e colocadas como relações entre estímulos, ações e respostas, o que pode torná-las simplificações acessórias às nossas relações sociais. Para nós, o fato significativo é o de que as emoções desempenham papéis em todas essas dimensões: biológicas, sociais e cognitivas. A ordem de motivação, isto é, o fenômeno biológico antecede o social e assim por diante parece ser uma questão sem resposta clara, embora cada uma das abordagens apresentadas tenha colocado o foco em uma delas. Freud (2015a), citado no começo dessa seção, entende que a questão é ponto pacífico, uma vez que o aparelho mental é dependente de uma base biológica e que o fator social é dependente deste aparelhamento mental.

Em sua teoria psicanalítica o desenvolvimento destas dimensões não é separado, como sugerem muitos teórico e, sim, concorrente, isto é, cada “inovação” em uma destas dimensões tem um respaldo imediato nas demais. Não vamos entrar nesse debate que foge ao escopo desse trabalho, mas entendemos que a compreensão de algumas destas dimensões podem ser fundamentais para um estudo mais aprofundado da relação entre fala e emoção, que abordaremos no próximo capítulo.

6.7 Discussão

Quando tratamos de um tema que margeia mais de uma área, como é o caso da fala emotiva, nos parece salutar apresentar as questões e dilemas envolvidos nestas áreas. Como este trabalho trata da linguística, julgamos por bem fazer uma apresentação mais extensa dos problemas relacionados à emoção antes de apresentarmos um panorama da fala emotiva e do modelo que adotaremos. Além disso, compreendemos que a apresentação pode se alongar, visto que o exposto envolve uma série de detalhes que não fazem parte do nosso domínio teórico.

Entendemos que o papel da emoção é o de regular nossas ações ou nos deixar apto a agir diante dos estímulos do ambiente. Porém, essa definição deixa em aberto duas dimensões: (i) o da *percepção* e (ii) o da *reação* ou *manifestação*. Como se vê, a percepção pode ser entendida como reposta reflexa, quando é conectada diretamente à ação, ou um processo cognitivo, quando outros componentes mentais entram em cena. Com relação à manifestação, ela é essencialmente motora ou comportamental, de modo que seu funcionamento corresponde à ativação de alguns mecanismos de ação.

A nosso ver, as emoções, tal qual a Física, a Linguística e outros campos do saber esteve à mercê da mentalidade de uma determinada época. Desta maneira, sua análise e interpretação esteve condicionada aos ideais e objetivos de cada um desses períodos, de modo que a cada novo paradigma, os ecos anteriores se mantem seja como método, seja como orientação. O aspecto cultural das emoções, isto é o significado delas

na sociedade está conectado de forma significativa com os princípios da formação intelectual ocidental e à tradição escolástica, assim, elas podem ser compreendidas como elementos desnecessários, que cegam a razão. Não à toa ainda vemos a tradição jurídica se utilizar da noção de crimes *passionais*.

No domínio acadêmico, a hipótese evolucionista, desenvolvida por Darwin e James é a mais recorrente. A explicação do papel da emoção na vida da humanidade e dos demais animais, aparentemente, não consegue escapar de uma análise do papel biológico e de adaptação destes seres ao meio. Ainda assim, dentro da área da linguagem, um número expressivos de trabalhos é baseado na hipótese cognitivista, seja porque a manifestação linguística seria inexplicável um modelo puramente biológico, seja porque a fala emotiva ocorre de fato em um segundo nível, como produto cognitivo resultante da percepção da ação da emoção em um indivíduo e não como elemento de avaliação.

A questão social tem um papel fundamental quando vemos as emoções como produtos culturais, isto é, como construtos que exercem uma função social e, por isso, são manipuláveis e passíveis de serem organizados e entendidos como mecanismos de ajustamento. Essa visão parece não ter a devida penetração no campo acadêmico, a não ser na explicação da variabilidade das emoções, um conceito fundamental para evolucionista.

Integrar essas três visões é um dos desafios atuais da ciência das emoções. O modelo de Scherer (2001) parece ser um dos mais competentes no que tange a essa integração, embora possamos criticar a complexa formalização do conceito de *appraisal*. Por outro lado, seu modelo é competente em compreender que as emoções humanas são fundamentais para a regulamentação da vida social, sem que necessariamente dependa dele para existir. Neste trabalho adotaremos um modelo integrado de análise das emoções, desenvolvido por Damásio (2012), o qual entendemos pode nos ajudar a compreender a relação entre as emoções, a fala e a entoação.

7 Fala emotiva

O objetivo deste capítulo é discutir a relação entre a fala e as questões associadas à emoção apresentadas no capítulo anterior. Para isso, vamos tomar como modelo a proposta de Damásio (2012) no que tange à forma como processamos as emoções na fala. Além disso, apresentaremos os resultados do nosso estudo de caso envolvendo falas atuadas.

7.1 O modelo de Damásio (2012): integrando as perspectivas

No capítulo 4 argumentamos que a produção da fala pode ser analisada como dois sistemas interdependentes. Um *contínuo*, responsável pela implementação da entoação e, por consequência, dos aspectos que denominamos comunicativos ou expressivos e outro *descontínuo*, responsável pela implementação dos segmentos. Já no Capítulo 3 sugerimos que a entoação pode ser analisada por meio de alvos tonais, unidades intermediárias que, de um lado, orientam a implementação da frequência fundamental e, por outro, são a contraparte cognitiva, responsável pelo processamento informacional e semântico da entoação. Dessa forma, o componente contínuo da fala está diretamente conectado aos alvos tonais, uma vez que os alvos fazem a coordenação da necessidade expressiva do falante à implementação entoacional no domínio físico.

No caso da fala emotiva, entendemos que a entoação desempenha um papel semântico-informacional, apresentando ao ouvinte (seja ele uma terceira parte ou o

próprio falante) um determinado estado emocional do falante, atraindo para o domínio linguístico, informações que, em tese, fazem parte exclusivamente do contexto. Neste sentido, nossa hipótese é a de que a fala emotiva não seria exclusivamente uma variação do modo como falamos, mas, sim, uma alteração global do sujeito. Dessa forma, os alvos tonais, como elementos que coordenam a implementação por meio da análise cognitiva dos significados pretendidos pelo falante, são alterados e alocados em novas posições topológicas que podem ser mapeadas nas curvas de F_0 .

Neste ponto, a questão central passa a ser o modo como o falante pode acessar as instruções para que essa alteração dos alvos topológicos possa informar o seu estado emocional. Para Damásio (2012), as emoções primárias são, em primeira instância, reações reflexas a um estímulo e representam traços mnemônicos de uma antiga relação que tínhamos com o meio e, por este motivo, são processadas pelas camadas mais inferiores de nosso cérebro. Note-se que para o autor não é importante uma avaliação do estímulo, basta que o cérebro o reconheça e classifique suas características mais fundamentais. Neste caso, a primeira manifestação da emoção não seria um produto da avaliação cognitiva, como pressupõem cognitivistas, e, sim, a detecção destas características, ainda que de forma bastante rudimentar.

A percepção deste estímulo desencadearia uma reação fisiológica, responsável por disparar uma “sopa química” de neuropeptídeos e hormônios, causando uma alteração na homeostase corporal. Damásio (2012) infere que essa alteração é processada pelo corpo, gerando sintomas como a sudorese, o aumento dos batimentos cardíacos ou a fuga rápida de um local. Essa resposta corporal dispara sinais para o cérebro que passa a executar um contingenciamento destas respostas, a partir daí, o indivíduo passa a ter consciência (ao menos o seu cérebro) dos *sentimentos da emoção*.

Nesta perspectiva, as *emoções* necessariamente convergem para esses *sentimentos* que derivam para a percepção e a consciência do objeto que desencadeou o estado emocional. Ao final do processo, são os sentimentos que oferecem respostas otimizadas

para o estímulo com base em uma avaliação do meio. Para o autor:

“...sentir os estados emocionais, o que equivale a afirmar que se tem consciência das emoções, oferece-nos *flexibilidade de resposta com base na história específica de nossas interações com o meio ambiente*. Embora sejam precisos mecanismo inatos para pôr a bola do conhecimento em jogo, os sentimentos oferecem-nos algo extra. (grifos do autor) (DAMÁSIO, 2012, p.131)

Dessa maneira, as emoções fazem parte de um circuito fechado que engloba o cérebro, o corpo e novamente o cérebro. Assim, enquanto as alterações causadas pelo estímulo disparador forem notados pelo cérebro no corpo, a tendência é que exista algum tipo de resposta derivada, seja ela a contração da musculatura ou a aceleração da respiração, por exemplo. Portanto, os sentimentos derivadas das emoções seriam os responsáveis por algumas respostas físico-fisiológicas que daríamos e, em última análise, seriam também responsáveis pelas respostas cognitivas, como a expressão corporal ou a manifestação linguística.

Note-se que para Damásio (2012) não se trata de uma discussão entre resposta inconsciente ou consciente, debate que tem tomado parte no estudo das emoções desde James (1884), pois, para o autor, a resposta do sentimento é sempre consciente, ao menos para o cérebro, que detecta as variações químicas por meio das terminações nervosas. Nesse sentido, a essência de uma emoção é a percepção que temos dessas variações ao longo da duração de um estado emocional, assim uma emoção sempre leva o sujeito a um novo estado de relação com o meio e com seus pares por meio dos sentimentos.

Damásio (2012) ainda sugere outro tipo de processamento das emoções que é diretamente derivado das emoções básicas e chamado de *emoções secundárias*. A principal diferença dessas para aquelas refere-se ao fato de não serem unicamente respostas reflexas disparadas por regiões subcorticais e, sim, processamentos mais altos, muito provavelmente derivados dos córtices e que chegam aos sistemas somatossensoriais, ou seja, um processo de dentro para fora. Nesse caso, é provável que a avaliação cognitiva tome parte, pois é necessário que se avalie internamente a qualidade e a

direção do estímulo. No entanto, o resultado continua sendo os *sentimentos*, de modo que a diferença essencial reside na forma como o estímulo é processado. É importante destacar que neste trabalho vamos lidar apenas com as emoções primárias.

De certa maneira, Damásio (2012) responde a questão *natureza vs social/cognitivo* que tem dominado o campo das emoções nos últimos 50 anos. É provável que as emoções sejam, de fato, mecanismos evolutivos, desenvolvidos ao longo de séculos para nos proteger de ambiente. Entretanto, também é provável que, ao menos nos humanos, ela tenha se desenvolvido em paralelo a outras estruturas, como a linguagem para ajudar na comunicação de estados corporais e não apenas do ambiente. Se isto é um fato, como propõe Damásio (2012), tanto em um caso, quanto no outro, é possível que a manifestação das emoções seja sempre coordenada por estruturas mais altas do processamento mental.

Para sustentar essa hipótese, Damásio (2012) se apoia em evidências biológicas e anatômicas, as quais sugerem que tanto a percepção, quanto a manifestação das emoções são partes de mecanismos neurais específicos. A primeira está associada ao sistema límbico, ao passo que a segunda não é diretamente controlada pelos mecanismos de atos voluntários. Um exemplo desse caso é a diferença notória que existe entre um sorriso *PanAm* e o sorriso verdadeiro. Somos capazes de identificar traços de uma execução forçada no primeiro, ao passo que no segundo não. Isso se dá, pois as partes que coordenam um e outro são diferentes, naquele elas são voluntárias e nesse involuntárias.

A nosso ver, a fala emotiva se enquadra justamente neste aspecto, tanto em emoções secundárias, quanto primárias. É improvável que ela seja totalmente biológica, porém é também improvável que ela seja unicamente cognitiva. Nossa hipótese, é a de que ela opera justamente nessa segunda etapa do processamento, como resposta dos sentimentos. Portanto, elas são dependentes dos estados cognitivos associados aos sentimentos. Na seção a seguir faremos uma apreciação do tema.

7.2 Linguagem e emoção

Em nosso entender, o debate em torno da análise da fala emotiva esbarra na questão de uma ainda seminal “linguagem emotiva”, a qual se volta para a conceptualização da natureza das próprias emoções. No campo da linguagem esta conceitualização é construída em torno de conceitos mais simples do que aqueles apresentados no capítulo anterior, sendo que a discussão gira em torno de dicotomias como: *natural* vs. *social*, *cultural* vs *natural*, ou ainda *mental* vs. *corporal* (BEATTY, 2005; WILCE, 2009). Como se pode imaginar a própria definição de emoção na linguagem é difusa e complexa, pois tenta organizar em blocos questões que são gradientes. Sendo assim, um modelo como o de Brunswik (1956) pode não lidar com a especificidade emocional esperada, mas, sim, com um dado específico da fala.

No bojo das teorias linguísticas, de acordo com Wilce (2009), as emoções deveriam ser compreendidas como um mecanismo de articulação entre aquilo que é natural e aquilo que é socialmente construído, tomando como interface deste processo as necessidades comunicativas do falante e as estruturas linguísticas existentes. Em certo sentido, essa hipótese retoma o modelo sugerido por Damásio (2012), pois entende que a emoção pode acionar um determinado tipo de linguagem. Neste sentido, as emoções, dentro da linguagem, poderiam ser compreendidas como uma contraposição da dicotomia saussureana (SAUSSURE, 2012) de Expressão e Conteúdo, tendo em vista que é a expressão, que determinaria o conteúdo de um determinado enunciado emocional. Por exemplo, um grito de raiva, conteria em sua expressão seu próprio conteúdo.

Dentro desta perspectiva, Wilce (2009) argumenta que os elementos cognitivos e fisiológicos que compõem a manifestação da emoção afetam diretamente a estrutura linguística de uma mensagem, tal como sugere Damásio (2012). Desta forma, o que ocorre é uma alteração global na fala emotiva, isto é, as categorias linguísticas, como a sintaxe, a prosódia e o vocabulário, seriam afetadas pelo estado emotivo do

sujeito, causando alterações nas formas. O autor defende esta possibilidade como um fato linguístico em si, pois entende que as alterações, como propostas por Damásio (2012), não seriam aleatórias e, sim, comporiam um conjunto de variações linguísticas perceptivas, as quais “regulariam” as emoções de um ponto de vista semântico.

Assim, as emoções, que são elementos inerentes à natureza, só possuíram algum significado cultural e social devido à linguagem, que as nomeia e manifesta. Se pensarmos na língua como um elemento social e cultural, o qual permite filtrar e organizar o aspecto natural das emoções e, por consequência, deriva-las em novas partes, seja por necessidades semânticas ou comunicativas, gerando variações intra-culturais. Desta forma, as emoções podem ser consideradas construtos sociais apenas depois que passam pelo filtro da linguagem e se tornam partes de um sistema de valores. Este mesmo argumento é defendido por Berlin e Kay (1969) ao explicar a nomenclatura das cores. Para os autores a existência de universais semânticos de cor é um fato, ao mesmo tempo que estes universais variam de uma cultura para outra, ou seja, trata-se de um fato negável ao mesmo tempo que é explicável pela experiência.

Quando nos voltamos para os estudos linguísticos, a abordagem mais próxima da proposta de Wilce (2009) é a aquela feita pelos estudiosos da Semiótica das Paixões, desenvolvida por Greimas e Fontanille (1993). Para estes autores, haveria um componente *patêmico* presente em todas as ações humanas, alterando a forma como estas ações são conduzidas por meio da internalização deste componente emocional. No entanto, a Semiótica preocupa-se apenas com a questão da discursividade da emoção, tratando-a como um actante da ação, capaz de alterar o estado daquele que recebe a sua modalização (FONTANILLE, 1995).

Deste modo, a abordagem semiótica traduz a dicotomia *saussureana*, ao opor os estados do sujeito da ação para compreender as implicações das emoções na linguagem, entendendo as como parte de um sistema de relações de significado, ou seja, a raiva só existe, porque existem as demais emoções. Há de se destacar que estes autores

(WILCE, 2009; GREIMAS; FONTANILLE, 1993) tendem a assumir as emoções como parte de uma estrutura subjacente a linguagem, ao passo que a linguagem ajuda a organizar e conceitualizar o que é a emoção, tornando-a um elemento social, por meio de suas manifestações.

Na questão da fala, em específico, esse problema já foi abordado numa variedade de temas e *backgrounds*, a seguir apresentamos alguns deles e a nossa análise envolvendo a fala emotiva.

7.3 A Expressão vocal da emoção

A *fala emotiva* ou a fala que transmite alguma informação expressiva passou a ser estudada de maneira mais sistemática, a partir dos avanços técnicos desenvolvidos pela fonética instrumental do início do século XX, principalmente depois das propostas de pesquisadores como a do *Visible Speech* de Alexander Melville Bell (BELL, 1867) e da fisiologia da fala de Hermann Helmholtz (HELMHOLTZ, 1895). No caso da entoação, um dos primeiros autores a tratá-la como o *locus* de fenômenos que podem ser decompostos em unidades menores, foi Weaver (WEAVER, 1924a, 1924b). Sua preocupação metodológica foi, a princípio, compreender a dinâmica das variações entoacionais, no caso entre a voz masculina e feminina e, a partir disso, sugerir um modo de compará-las de forma sistemática, por meio de parâmetros.

A questão metodológica levantada por Weaver se voltaria, anos mais tarde, diretamente para os problemas da análise vocal das emoções, quando tomamos estas como um produto das variações entoacionais. Ainda que tenhamos o mesmo material fonético resultante, as condições de produção foram alteradas, de modo que é necessário se estabelecer ou construir parâmetros de análise, que sejam capazes de restaurar estas condições. Desta forma, apenas a mensuração das curvas de frequência não seria suficiente para qualquer análise mais profunda da relação entre emoção e variação de entoação, por exemplo.

Como demonstra Scherer (1986) diversos estudos no campo da fala emotiva foram feitos a partir deste princípio. Para o autor, ainda que este modelo fosse adequado, as análises feitas a partir dele geraram uma série de “paradoxos” metodológicos, pois ainda que se fosse possível determinar a diferença entre emoções, o método era incapaz de apresentar um conjunto de parâmetros específicos, os quais descrevessem a manifestação de uma emoção. Assim, o que se tinha era a análise de parâmetros diversos, os quais não representavam uma ou outra emoção especificamente. Em última análise, esses trabalhos se dedicavam a descrição de parâmetros e, não, uma análise sistemática das variações entoacionais, como sugerido por Weaver (SCHERER, 1986, 2003)

Scherer (1986) também sugere que este problema tem duas razões básicas: (i) a falta de um estudo mais completo dos processos e mecanismos responsáveis pela manifestação das emoções, sejam eles fisiológicos, psicológicos ou sociais e (ii) a falta de uma definição clara dos processos fonéticos que determinam as variações presentes na fala emotiva. Com relação ao primeiro tópico, entendemos que tenhamos tratado deles de maneira geral no Capítulo 6, por isso, não iremos retomar essa discussão.

Com relação ao segundo tópico, o próprio autor, em outro artigo (SCHERER, 2003), elucida a necessidade de se criar um modelo comunicacional, cuja interface seja, justamente, o mecanismo de produção da voz emotiva. Assim, de um lado estão alocados os recursos que o falante dispõe para codificar e transmitir uma determinada informação emocional. Do outro, estão os parâmetros necessários para que o ouvinte seja capaz de decodificar essa informação.

Brunswik (1956) argumentava que é necessário haver um mecanismo psicológico, capaz de gerar variações motor-sensoriais no falante, as quais seriam a base da especificidade dos parâmetros acústicos das emoções. O mesmo mecanismo seria capaz de perceber essas variações. Desta maneira, temos duas dimensões complementares: as *pistas distais* e as *pistas proximais*. A primeira trata das pistas informacionais que o

falante transmite para que seu estado afetivo seja notado (por este motivo são distais), ao passo que as outras lidam com as pistas percebidas pelo ouvinte (BRUNSWIK, 1956).

O modelo de Brunswik (1956) atende a complexidade pressuposta por Scherer (2003), pois coloca em evidência a necessidade de se compreender a diferença existente entre os aspectos relacionados à produção da voz emotiva em contraste aqueles associados à percepção desta. Em nosso entender, a proposta destes autores é justamente a de compreender a fala emotiva como um processo dinâmico, no qual as variações acústicas alteram os parâmetros que a compõem, ao passo que é a recombinação destes parâmetros por parte do ouvinte que determina a especificidade de cada uma das emoções.

Scherer (2003) demonstra que esse pressuposto foi fundamental para se criar dois paradigmas de análise da fala emotiva. O primeiro deles é o que lida com o domínio da “codificação” da informação emocional, estes estudos supõem que determinados padrões e configurações são utilizados pelos falantes para transmitir estas informações. A determinação destes padrões tem sido discutida desde meados da década de 30 (SKINNER, 1935; FAIRBANKS; PRONOVOST, 1939; FAIRBANKS; HOAGLIN, 1941; BONNER, 1943), teve um grande desenvolvimento nas décadas de 60 e 70 (FONÁGY; MAGDICS, 1963; STEVENS; WILLIAMS, 1969; CONSTANZO; NORMAN; COSTANZO, 1969; WILLIAMS; STEVENS, 1972; SIMONOV; FROLOV, 1973; UTSUKI; OKAMURA, 1976) e com os avanços na área da informática se tornou mais apurado na década de 90 (COWIE; DOUGLAS-COWIE, 1996; BANSE; SCHERER, 1996; KLASMEYER; SENDLMEIER, 1997; KLASMEYER, 2000; GOBL; CHASAIDE, 2003; LAUKKA, 2004; SCHERER, 2005; BANZIGER; SCHERER, 2005).

De maneira geral, estes estudos tratam de três correlatos acústicos: (i) a variação da frequência fundamental, (ii) a qualidade vocal e (iii) o ritmo. . Williams e Stevens (1972) sugere que apenas a frequência fundamental como série temporal é capaz

de especificar uma emoção de forma protípica, de modo que outros correlatos não aparentam definir com a mesma clareza. Por outro lado, Pakosz (1983) argumenta que a entoação transmite pouca informação emocional, a qual precisa ser preenchida pelo contexto ou pelas informações de outros canais, como o visual, transmitido pela expressão facial.

Gobl e Chasaide (2003) demonstram que a qualidade vocal pode operar como um dos indicadores do estado emocional de um indivíduo. No entanto, esse canal opera com baixa definição, uma vez que não é possível mapear uma relação unívoca entre emoções e qualidade, mas antes um conjunto de estados afetivos que se associam de forma menos ou mais significativa com uma qualidade vocal específica. Por sua vez, o ritmo opera como um elemento condicionante, isto é, as variações de F_0 ou de qualidade estão condicionadas à duração ou taxa de elocução para que sejam representativas como pistas para a detecção das emoções (SCHERER, 1986)

O segundo paradigma citado por Scherer (2003) é justamente aquele que estuda como estas pistas vocais são interpretadas pelo ouvinte. Para Scherer (1986) um dos maiores paradoxos no âmbito do estudo da manifestação vocal das emoções é o grau de acuidade com que pessoas são capazes de detectar uma emoção *versus* a falta de evidências que demonstrem quais são as pistas que elas utilizam nessa detecção. Não vamos entrar nos detalhes dessa discussão, pois não nos parece pertinente para este trabalho. No mais, entendemos apresentaremos nossa análise e ao final como ela pode contribuir para o estudo das emoções.

7.4 Teste 2: um estudo de caso das emoções

Retomando a caracterização feita por nós entre *alvos topológicos* e *alvos pontuais* destacamos que em Martins e Vassoler (2013) pudemos demonstrar o comportamento daqueles, de modo que aqui vamos nos ater ao estudo destes. Em Martins e Vassoler (2013) notamos que a variação de frequência fundamental em duas condições, a

raiva e a voz neutra, mantinha uma grande correlação, embora estivessem em alturas espectrais distintas. De um ponto de vista matemático, isso significa dizer que elas eram parecidas na forma como variavam, no entanto, essa variação ocorria em níveis distintos. A raiva nitidamente mantinha-se em um registro superior, ao passo que a voz neutra estava numa zona mais medial de variação.

Supomos a partir daquele trabalho que a alteração dessa zona espectral, alteraria a forma como falantes utilizam a frequência fundamental para expressar suas emoções, em especial, na faixa de Foco/Ênfase ou do que denominamos alvos pontuais. Para compreendermos melhor essa questão propomos, por ora, uma análise das variações de frequência fundamental em duas emoções, a raiva e a tristeza.

Optamos por estas duas emoções seguindo a proposta de Scherer (1986) de que a raiva seguiria um padrão de acréscimo na média de F_0 , em relação a fala neutra, ao passo que a variação tonal seria grande. A tristeza, por sua vez, teria um padrão de decréscimo de F_0 acompanhada de uma pequena variação tonal, ou seja, seriam padrões opostos. Na tabela 7.1 apresentamos um resumo das duas emoções acima citadas. Além disso, como se verá, estas emoções são semelhantes em diversos aspectos e se distinguem em um ponto fundamental, a saber: a *internalização* vs. a *externalização* dos sentimentos.

Scherer (1985) define a raiva como um estado no qual um indivíduo é incapaz eliminar um obstáculo que o impede de atingir um objetivo ou satisfazer uma necessidade, ainda que este indivíduo se sinta capacitado de o fazer, levando a tentativa de se atingir o objetivo proposto de maneira explosiva ou à força. É comum na literatura de língua inglesa se fazer a distinção entre *hot anger* e *cold anger*, tratando a primeira como a cólera e a segunda como a irritação, ou a agressividade. Para este trabalho, usaremos a definição de *hot anger*, por entendermos que a agressividade não se trata de uma emoção e, sim, uma atitude diante de uma determinada situação. A literatura sobre o aspecto vocal da raiva é vasta e os principais achados atestam o aumento da frequên-

cia fundamental, embora não exista uma homogeneidade nos valores F_0 , devido a grande variação, forte intensidade e velocidade de fala rápida (BANSE; SCHERER, 1996; LAUKKA, 2004; SCHERER, 2003, 2005; VASSOLER, 2011; MARTINS; VASSOLER, 2013; FERREIRA-NETTO et al., 2014).

Por outro lado, a tristeza seria um estado em que o sujeito também se vê diante de um obstáculo que o impede de atingir um objetivo ou satisfazer uma necessidade, embora, neste caso, ele saiba que não pode controlar as consequências deste episódio, restando apenas que ele internalize as suas sensações na tentativa de rever seus objetivos e posturas. Por sua vez, a tristeza pode ser compreendida, como passiva ou ativa, a diferença entre elas está no controle dos sistemas responsáveis pela ativação da musculatura da glote e da respiração, no caso, a ativa se definiria pela no controle de produção vocal. Laukka (2004), Vassoler (2011) e Ferreira-Netto et al. (2014) demonstram que a tristeza tem valores de frequência fundamental baixa, com poucas variações e intensidade vocal fraca.

Parâmetros vocais	Raiva	Tristeza
Média de F_0	>	<
Variabilidade de F_0	>	<

Tabela 7.1: Características vocais da raiva e da tristeza. Adaptado de Scherer (1986)

7.4.1 Natureza dos dados

Existe um debate metodológico, amplamente discutido por Scherer (2005), sobre a natureza dos dados de expressão vocal. Deveriam ser eles naturais, retirados de falas espontâneas, ou atuados, gravados em laboratório, utilizando atores? Nesta pesquisa utilizaremos dados de falas atuadas, sendo que em uma etapa posterior pretendemos analisar dados de fala espontânea, a fim de verificar a aplicação destes parâmetros na rotina ExProsodia.

Os dados deste trabalho foram gentilmente cedidos por Aline Vassoler e serviram

de base para sua pesquisa de mestrado (VASSOLER, 2011). Os dados consistem de 3 sentenças retiradas de um texto científico, lido 5 vezes repetidas em 5 emoções diferentes: alegria, medo, neutro, raiva e tristeza. A escolha do texto científico se deu pela dificuldade imposta de se transpor as emoções para a fala, de modo que apenas a atuação poderia fazê-lo de forma coerente. O texto original usado é apresentado abaixo e as frases em negrito representam as sentenças utilizadas por Vassoler (2011):

A reatividade dos linfócitos, as **células do sangue que fabricam anticorpos são individualizadas**. Em cada organismo, **as células do fígado são provavelmente iguais entre si**, as da pele também, mas os linfócitos são diferentes uns dos outros. Cada um difere do seguinte por possuir na membrana diferentes receptores, moléculas que garantem a aderência a certas estruturas (ou a capacidade de fixar certas substâncias). Assim, o linfócito seguinte adere às estruturas diferentes. Para ser mais exato, as diferenças existem entre clones de linfócitos. Quando um determinado linfócito se multiplica e gera duas, quatro, oito milhares de cópias idênticas, **esse conjunto constitui um clone linfocitário**. Dentro de um mesmo clone, os linfócitos são iguais: têm os mesmos receptores de membrana, aderem às mesmas coisas, participam das mesmas interações. (grifos nossos) (VAZ, 1983, p. 33, item 2)

No total obteve-se um total de 75 *tokens* para cada sentença: 3 atrizes x 5 emoções x 5 repetições. Para este trabalho utilizamos apenas a sentença: as células do sangue que fabricam anticorpos são individualizadas, somando um total de 30 *tokens*. Uma análise prévia das sentenças 2 e 3 retornaram resultados parecidos com aqueles, de modo que acreditamos que o cenário ideal seja fazer uma nova rodada de gravações com textos distintos.

7.5 Análise dos dados

Para esta pesquisa analisaremos dois parâmetros básicos: a variação de frequência fundamental e o valor do Limiar de Diferenciação Tonal superior (LDT superior), apresentado no capítulo 3, destacamos que, em termos matemático o LDT superior é um produto da função que estabelece o Tom Médio, sendo assim, seu estudo pode ser compreendido como uma análise indireta deste componente. Contudo, optamos por utilizar o LDT superior por entendermos que ele seja o responsável linguístico

Estatísticas			
	F0 - Neutra	F0 - Raiva	F0 - Tristeza
N	133	135	109
Média	50,4	53,9	49,7
Mediana	51	54	50
Desvio-Padrão	4,25	3,92	4,02
Variância	18,11	15,39	16,17
Assimetria	-0,17	-0,39	0,74
Curtose	-0,774	1,079	3,165
Intervalo	19	25	26
Mínimo	42	39	42
Máximo	61	64	68

Tabela 7.2: Análise descritiva

mais diretamente associado com a marcação do foco linguístico em um determinado vocábulo ou sentença. Em uma próxima etapa, pretendemos analisar os dados relativos ao LDT inferior, à título de análise do Foco Negativo, isto é, a marcação de elementos linguísticos pelo decréscimo de F_0 (CONSONI, 2011; MEDEIROS; MARTINS, 2014). Ressaltamos que nossa análise foi feita utilizando a escala MIDI. Optamos por essa medição, uma vez que ela é referencial e permite uma compreensão melhor dos dados de variação, como apresentado em Martins e Ferreira-Netto (2010)

A tabela 7.2 apresenta a análise descritiva dos valores mensurados de F_0 para as emoções analisadas mais a voz neutra.

A tabela 7.2 corrobora os achados da literatura, demonstrando que a raiva possui um valor médio maior que a voz neutra e a tristeza. De acordo, com os dados acima apresentados a diferença média da raiva para as demais emoções é da ordem de 4 semitons, valor suficiente para a diferenciação tonal (CONSONI, 2011; MARTINS, 2013). Observa-se que os valores de variância, desvio-padrão são relativamente próximos para todas as vozes, indicando que a taxa de variação não é um fator determinante para a caracterização da expressão vocal, embora seja um elemento importante para a diferenciação entre a fala emotiva (raiva e tristeza) para a voz neutra, cerca de 2st de variação. Por outro lado, o intervalo de máximo e mínimo da raiva e tristeza indicam que o falante pode tender a produzir um espectro maior para a vocalização

dessas emoções, embora este valor precise ser melhor analisado, utilizando dados de qualidade da voz.

Seguindo esta análise, a figura 7.1 demonstra que a voz neutra possui uma distribuição próxima da normal (assimetria = - 0,169), com a média em torno de 50 st, ao passo que as figuras ?? e 7.2 demonstram a existência de uma distribuição assimétrica nas demais variáveis. Os dados associados à distribuição demonstram que as emoções acionam configurações específicas dos alvos topológicos, principalmente quando observamos a concentração dos tons. A fala neutra tende a uma concentração central, com *outlier* sendo elementos laterais, ao passo que as emoções raiva e tristeza teriam concentração laterais, com poucos *outliers*. Nota-se que para tristeza estes valores estão próximos à esquerda, indicando que estão mais próximos de zero, ao passo que para raiva estão à direita, indicando um acréscimo em relação à zero.

Figura 7.1: Histograma da frequência fundamental para voz neutra, medida em MIDI.

Figura 7.2: Histograma da frequência fundamental para tristeza, medida em MIDI.

Figura 7.3: Histograma da frequência fundamental para a raiva, medida em MIDI.

Estes dados sugerem duas hipóteses:

- (i) expressão vocal da raiva tende a possuir menos alvos pontuais do que a fala neutra, pois ela ocorre em uma banda de frequências mais alta;
- (ii) o mesmo vale para a tristeza, uma vez que o foco linguístico, neste caso, demandaria um maior esforço para atingir pontos de frequência mais altos.

As figuras ?? e 7.2 também demonstram a existência de uma distribuição leptocúrtica para os valores de frequência fundamental da raiva e da tristeza, indicando que mesmo distante da média pode se obter valores significativos ($k > 0$) e uma distribuição mesocúrtica para a voz neutra, demonstrando uma distribuição próxima da normal e uma concentração de valores significativos menos variável em relação a média ($k \approx 0$). Como acima dito, estes valores também sugerem que a análise da variação de frequência fundamental pode não ser um critério determinante para a avaliação de emoções.

A figura 7.4 corrobora a literatura, uma vez demonstra que raiva e tristeza são emoções distinguíveis, apenas pela variação de sua frequência fundamental, tendo em vista que o valor mediano da segunda é próximo do valor mínimo da raiva, ou seja, os valores de frequência da tristeza estão em um registro mais baixo que aqueles que representam a raiva, como se pode ver no segundo quartil do *box-plot* que representa cada uma das emoções. Contudo, é notório que os valores de F_0 das emoções raiva e tristeza são menos variáveis do que os da voz neutra, indicando que o espaço de variação delas é parecido ou igual, embora em um registro inferior, diferentemente do que prevê o modelo de Scherer (1986), apresentado na tabela 7.1.

Figura 7.4: Box-Plot para frequência fundamental de todas as expressões vocais, os valores estão em MIDI.

Por fim, a figura 7.5 demonstra que os valores médios das duas emoções em questão, raiva e tristeza, estão fora do intervalo de confiança de uma ou outra emoção. Em nosso entender, esse dado demonstra que a variação de F_0 de uma voz com raiva dificilmente seria confundida com a de uma voz triste, uma vez que o intervalo de erros das duas vozes não têm intersecção em nenhum dos pontos, ainda que a questão da variação delas não seja trazido à tona por este cálculo. Esta hipótese foi confirmada por meio do *teste-t* apresentado na tabela 7.3.

Teste - t de amostras emparelhadas				
		t	df	sig.2
Par 1	F_0 -Neutra vs. F_0 -Tristeza	2,84	108	0,005
Par 2	F_0 -Neutra vs. F_0 -Raiva	-6,27	133	0

Tabela 7.3: Teste-t para voz neutra, tristeza e raiva

Figura 7.5: Barra de Erros para frequência fundamental de todas as expressões vocais, os valores estão em MIDI.

Por outro lado, o gráfico demonstra que a voz neutra poderia ser confundida com a voz triste, se analisada somente a variação de F_0 . Contudo, um teste-t, comparando a frequência fundamental da voz neutra com a tristeza, indica que as duas variáveis são distintas ($p = 0,005$) (tabela 7.3). A fim de nos assegurarmos sobre a relação entre a voz neutra e a tristeza, realizamos também um teste não-paramétrico de Wilcoxon de Amostras Relacionadas que retornou a significância de $p < 0$, sugerindo que a H_0 deva ser desconsiderada.

Como se pode ver, as emoções possuem espaços entoacionais distintos, bem como a distribuição dos tons. No entanto, destaca-se o fato da fala neutra ter uma variabilidade intrínseca maior que as demais. A explicação possível é justamente dada pelo fato dos alvos topológicos nas emoções ocorrerem em um espaço restrito: mais baixo na tristeza e mais alto na raiva. Dessa forma, algumas variações poderiam alterar a posição dos alvos topológicos e, por consequência, o aspecto comunicativo pretendido pelo falante.

7.5.1 Valores para o LDT Superior: o alvo pontual

Nesta seção avaliamos os alvos pontuais, em específico os valores os valores do limiar de diferenciação tonal superior para todas as condições. Descartamos de nossa análise as proeminências negativas, uma vez que não as encontramos nos dados analisados. A tabela 7.4 apresenta os valores do limiar de diferenciação tonal.

	Estatísticas Descritivas		
	F0 - Neutra	F0 - Raiva	F0 - Tristeza
N	133	134	109
Média	54,17	57,75	53,22
Mediana	55	58	54
Desvio-Padrão	2,54	2,77	2,84
Variância	6,44	7,66	8,06
Assimetria	-0,25	-1,45	-0,5
Curtose	-1,27	6,62	0,92
Intervalo	9	21	15
Mínimo	49	42	47
Máximo	58	63	62

Tabela 7.4: Valores dos Limiares de Diferenciação Tonal Superiores em Midi para Voz Neutra, Raiva e Tristeza

A tabela 7.4 demonstra que os valores médios de LDT superior, seguem a média dos valores de F_0 , estando a raiva em um registro de cerca de 4 semitons acima das demais emoções. Contudo, essa variação se explica pela Equação 3.5, na qual o LDT superior é um produto derivado da média da frequência fundamental. Destaca-se, neste caso, que o desvio-padrão e a variância são mais baixos do que aqueles apresentados para a frequência fundamental.

Existem duas explicações para a baixa dispersão: (i) o espaço espectral de frequência é reduzido, devido ao aumento do tom médio, levando, necessariamente, a uma variação menor destes valores; (ii) a relação da variação destes valores está associado ao tom médio, o qual possui, como característica geral, uma baixa variabilidade, isso se dá pelo Teorema do Limite Central, no qual um conjunto de média, tende a centralizar em um valor medial e possuir pouca dispersão.

Outro dado relevante para esta análise é o intervalo de frequências do LDT da Raiva, esse valor se estende por 21 semitons, enquanto as demais estão em 9 (neutra) e 15 (tristeza) semitons, ao passo que na análise de F_0 todos os intervalos estava em torno de 20 st. Como dito, na seção anterior este dado é de natureza aritmética, de modo que sua análise poderia ser feita utilizando processos que avaliem a qualidade vocal, por exemplo.

A distribuição das frequências do LDT Superior, como sugerem as figuras 7.6, 7.7 e 7.8. A voz neutra demonstra uma distribuição próxima da normal, ao passo que a raiva e a tristeza tendem a uma assimetria lateral positiva e negativa, respectivamente, fato previsto na seção anterior. Contudo, observa-se que a raiva possui uma grande quantidade de valores significativos, mesmo se distante da média

Figura 7.6: Histograma dos valores de Limiar de Diferenciação Tonal superior para voz neutra.

Figura 7.7: Histograma dos valores de Limiar de Diferenciação Tonal superior para raiva..

Figura 7.8: Histograma dos valores de Limiar de Diferenciação Tonal superior para raiva..

A figura 7.9 demonstra que as falas emotivas têm uma distribuição mais regular dos valores superiores do que a voz neutra, seguindo princípio encontrado para a frequência fundamental. Nota-se também que os valores do LDT superior para a voz neutra ficam mais próximos ou abaixo da média, de modo que a medida do último quartil é relativamente menor, quando comparada ao mesmo quartil na figura 7.4.

Figura 7.9: Box-Plot do Limiar de Diferenciação Tonal superior de todas as expressões vocais.

Por fim, executamos um *teste-t* comparando todas as variáveis em questão (tabela 7.5). Os valores demonstram, novamente, que as variáveis são distintas e um teste não-paramétrico de Wilcoxon de Amostra Relacionadas, confirmou que as amostras possuem diferentes distribuições.

Teste-t de amostras emparelhadas			
Amostras analisadas	t	df	Sig.2
LDT Superior Neutra e LDT Superior Raiva	-9,142	132	.000
LDT Superior Neutra e LDT Superior Tristeza	6,95	108	.000
LDT Superior Raiva e LDT Superior Tristeza	9,63	108	.000

Tabela 7.5: Teste-t para voz neutra, tristeza e raiva

Uma segunda análise feita por nós, utilizando os valores encontrados acima do LDT superior, o que temos chamado de faixa de frequência do Foco/Ênfase. Uma

das características das proeminências é o fato delas terem uma baixa ocorrência, uma vez que ele deve ser utilizado para marcações de elementos linguísticos específicos (FERREIRA-NETTO, 2006, 2008; CONSONI, 2011). Por este motivo, sua incidência seria menor, já que ela passa por um “controle pragmático” do próprio falante.

A tabela 7.6 apresenta dados relacionados aos pontos de foco de cada uma das emoções analisadas. A tabela foi gerada a partir da fórmula:

$$\text{Se } F_0 - LDT_{sup} > 0 = \text{Foco} \quad (7.1)$$

Foco Linguístico			
	Neutra	Raiva	Tristeza
Média dos focos	2,33	2,86	4,2
Desvio Padrão	3,61	3,04	2,87
<i>n</i>	15	7	5

Tabela 7.6: Valores relativos ao Foco Linguístico, em MIDI, para cada uma das emoções

A tabela nos revela dois dados particulares. No primeiro deles, podemos notar que a voz neutra possui o maior número de alvos acima do LDT, ao passo que a raiva e a tristeza possuem menos, respectivamente. Este dado confirma, em um primeiro momento, a hipótese de que devido a assimetria na distribuição dos tons, seria natural haver menos alvos pontuais na expressão vocal das emoções, ao passo que na fala neutra eles seriam mais numerosos, justamente por não estarem condicionados a um espaço restrito de implementação.

Em segundo lugar, a média dos focos, indica que os alvos pontuais da raiva e da voz neutra possuem valores parecidos, cerca de 2 st acima do LDT superior, enquanto a tristeza, está em um registro acima dos 4 st. Estes dado sugere que para a produção

do foco na tristeza, o falante tem de fazer um esforço acima do normal para produzir um fato diferencial, a figura 7.4 demonstra esse fato. Na raiva esse esforço neutralizado pelo fato das frequências seriam naturalmente mais altas, ao passo que na fala neutra os valores corroboram aqueles apresentados no capítulo 5 .

7.6 Discussão dos dados

Destacamos aqui algumas conclusões importantes e derivadas da relação entre alvos pontuais e alvos topológicos. Aparentemente, a análise da variação da frequência fundamental como dispersão não parece ser um critério unívoco de distinção das duas emoções, como sugerido por Scherer (1986). A nosso ver, a distribuição de espaços de frequência pode ajudar a compreender essa dispersão de forma mais significativa. Uma análise da curtose das frequências pode ser um ferramenta importante para esta distinção, como sugerido em Ferreira-Netto et al. (2014).

Os alvos pontuais na faixa de frequência do F/E, embora sejam poucos, parece ser um critério relevante na determinação do estado emocional por meio da entoação, de modo que a análise sistemática destes dados pode relevar elementos importantes para distinção das emoções, dentro do modelo apresentado no capítulo 3. Nessa linha, a incidência deste alvos dar pistas para que possamos diferenciar determinadas emoções de outras.

8 Conclusões

Como fizemos uma série de observações sobre os nossos dados e teste ao longo das análises de cada um deles, vamos tentar nos limitar a uma discussão mais neste capítulo. Como citamos na introdução desse trabalho, o leitor poderia notar que essa tese foi dividida, ainda que sem intenção direta, em duas partes. Em uma delas trataríamos dos aspectos linguísticos mais emergentes e na outra das questões da fala emotiva.

Começemos pelos aspectos linguísticos mais fundamentais da nossa hipótese: o de que a entoação pode ser dividida em unidades menores, chamadas de alvos tonais. A suposição desses alvos, nos levou a deduzir que a produção da fala poderia ser decomposta em dois sistemas: o *contínuo* e o *descontínuo*.

Pode se notar, de maneira geral, que essa distinção visa, antes de mais nada, explicar a implementação da expressividade da fala, não como uma variação de segunda ordem e, sim, como um elemento linguístico, o qual os falantes de uma língua detêm o conhecimento, são capazes de intuir seu valor e também implementar de forma bastante produtiva ou padronizada. Neste sentido, a implementação desses alvos, por nós hipotetizados, está associada à maleabilidade do sistema contínuo, isto é, embora ele seja interdependente do sistema descontínuo, é notório que ele possui uma função linguística que é, muitas vezes, detratada como mero sistema suporte.

A própria noção de alvos que construímos nesse trabalho perpassa essa distinção. Os alvos topológicos supõem uma implementação distribuída em um espaço limitado, de modo que é fundamental que o falante possua em algum nível um controle das

variações de F_0 . Para nós, esse controle é coordenado pela dimensão perceptiva da fala, nela o falante é capaz de monitorar sua implementação e mantê-la num nível ideal ao longo do tempo de uma elocução, o que Ferreira-Netto (2008) denominara *tom médio*. Nesse caso, o tom topológico, possui um certo monitoramento por parte do falante e do ouvinte, de modo que qualquer alteração significativa pode modalizar o discurso.

Nos alvos pontuais, a noção de implementação tem uma significância relativa ainda maior. Se partimos do pressuposto que estes alvos precisam necessariamente ser eventos relevantes em uma elocução, é natural pensar que a sua implementação está associada a mais regras linguísticas e fonológicas do que no primeiro caso. A noção de valência que citamos no capítulo 3, e pouco discutimos ao longo dessa tese, parece ser bastante importante nesta relação entre implementação e fonologia entoacional. Nesse caso, o estudo das durações, bem como da relação entre a duração e um conjunto de tons pode revelar pistas de como os falantes codificam eventos linguísticos ou comunicativos na entoação.

Ao sugerirmos no título que esses alvos podem ser unidades fonético-fonológicas, pretendíamos destacar o seu papel operacional de estabelecer uma relação virtual entre o nível representacional da entoação e o nível físico. Entendemos que o nível representacional é completamente abstrato e hipotético e corresponde, de certa forma, ao processamento linguístico em suas duas formas, a produção e percepção. Neste sentido, caberia a esse nível representacional, receber as informações entoacionais mapeadas pelos alvos e fornecer conteúdos comunicativos ou linguísticos ao ouvinte. Na outra face, o nível representacional estabeleceria quais são as informações linguísticas ou comunicativas a serem transmitidas e os alvos fariam a coordenação da implementação acústica.

Dessa forma, estes alvos seriam formas prototípicas e não unidades representacionais. Formas prototípicas referem-se a modelos abstratos que possuem um conjunto de atributos capazes de serem mapeados e caracterizados pela nossa cognição. No entanto,

esses atributos não são estanques, tal como os traços fonológicos, mas representam pontos referenciais para os quais nossa percepção converge quando notados, tal como na definição de *efeito magnético de percepção*, desenvolvido por Kuhl (1991). Os alvos não seriam unidades representacionais, como sugere Pierrehumbert (1980), pelo fato de estarem amplamente associados às variações da frequência fundamental.

É dentro desta linha de raciocínio que compreendemos os alvos como unidades fonético-fonológicas. Em certo sentido, eles seriam uma ponte, entre o nível representacional estático e o nível acústico, extremamente variável. As dimensões associadas aos alvos, portanto, seriam capazes de explicar um e outro, assim, ao se “decantar” a variabilidade de F_0 , podemos detectar unidades prototípicas que explicariam porque certa curva de F_0 é compreendida de uma determinada forma e outra não.

No projeto deste trabalho visávamos desenvolver um método que quantificasse essas variações e as organizassem em categorias por grau de semelhança entre os protótipos. Em um primeiro momento, o estudo da inferência *bayesiana* parecia ser promissor, bem como os modelos de ligados à redução de dimensões ou *clusterização*. A questão central era desenvolver um método que realmente quantificasse “feixes de dados” em categorias prototípicas. Os resultados sempre convergiam para uma análise de distribuição *versus* uma análise de dispersão.

No primeiro caso, buscamos identificar a posição dos dados em uma determinada área ou espaço, como em um mapa de calor que demonstra a localização de um jogador de futebol em campo. No segundo caso, a análise de dispersão visa avaliar o quanto este jogador se distanciou de uma determinada região por onde ele mais se deslocou. Essa distinção não parece importante, mas ela é fundamental no momento em que vamos categorizar valores. Se categorizamos pela distribuição, necessariamente, estamos avaliando o comportamento todo, se avaliamos por dispersão, estamos mais focados naquilo que é errático a partir de um determinado valor.

Essa é uma discussão que não cabe neste momento, mas é importante notar que

modelos distribucionais tendem a não reduzir dados a conjuntos de valores, pois visam analisar a relação entre eles e o espaço de probabilidade, ao passo que modelos de dispersão o fazem, pois buscam evidenciar a relação entre variáveis. No teste 1 buscamos utilizar um modelo ainda muito seminal de distribuição que tem por objetivo detectar as variações de F_0 em um espaço determinado.

Duas questões foram fundamentais neste teste. A primeira delas era como *dissolver* a diferença existente entre os *pitch range* dos falantes? Em ambos os modelos descritos, elas necessariamente tomariam parte. A segunda questão era como esse *pitch range* pode afetar a forma prototípica? A solução que utiliza o coeficiente de variação é, para nós, uma solução temporária, pois sabemos que outros recursos matemáticos podem vir a nos ajudar nesse ponto.

Deste modo, o objetivo central deste teste, de um ponto de vista prático, não era criar um modelo e, sim, avaliar se um modelo desse tipo seria possível. O teste de repetição demonstrou em certo sentido, que falantes tendem a executar suas elocuições a partir de um estímulo de forma não aleatória e dentro de valores que podem ser analisados pela distribuição. De todo modo, os resultados que obtivemos sugeriam que essa forma prototípica pode ser definida pela (i) *variação intrínseca* de F_0 em uma elocução, quando ela ocorre no espaço entoacional ou pela (ii) *variação extrínseca*, quando ela está na faixa de frequências F/E.

O primeiro caso (i) refere-se às formas prototípicas associadas à variação da frequência fundamental. Em nosso entender, o *pitch range* do falante não toma parte nesse caso, pois não estamos preocupados com as alturas relativas e, sim, em como um altura se relaciona com a outra em uma série de alturas. É provável que determinados conteúdos informacionais ou linguísticos tenham uma variação intrínseca parecida e estabeleçam um espaço entoacional padronizado, já que ele não parece ser totalmente aleatório, uma vez que os falantes foram capazes de seguir a estrutura do modelo dado. Os mesmos achados foram notados no estudo de caso das emoções, como

pode se notar, essas variações intrínsecas foram fundamentais para se estabelecer uma distinção entre a fala com raiva e a fala com tristeza.

O segundo caso (ii) lida com formas prototípicas associadas à percepção. Em nosso entender, nesse ponto a variação de frequência fundamental é concentrada para que o ouvinte note as proeminências relativas. Como pudemos ver, essa variação parece ser relativamente aleatória, pois ela não possui um limite específico, mas, sim, um limiar a partir do qual ela deve se orientar. A nosso ver, a forma prototípica desse caso é a única que é também fonológica *per se*, pois a alteração das alturas do espaço entoacional necessariamente gerar um evento semântico na elocução.

Portanto, para resumir algumas questões essenciais. O conceito de alvo tonal topológico pode ser avaliado, desde que utilizemos uma análise de distribuição das variações de F_0 para categorização e dissolução do *pitch range*. A noção de alvo tonal pontual está bastante associado a percepção, de modo que sua análise precisa ser mais aprofundada dentro dos conceitos de valência e de movimentos tonais, como o curso de acréscimo de semitons em um alvo desse tipo. De todo modo o sistema contínuo é fundamental nesse processo, já que estas variações e modulações dependem essencialmente dele.

Com relação à segunda parte de nossa tese, relacionada às emoções, é notório que ainda falte um debate mais extenso sobre o papel delas na língua. Ou ainda como as emoções podem afetar categorias linguísticas de diversos níveis, como a sintaxe ou a morfologia. Embora seja notável alguns avanços, o tema das emoções parece ser bastante atrelado às questões da voz.

A análise da literatura nos revela que os estudos da fala emotiva pela entoação, como os de Bänziger e Scherer (2005) e Williams e Stevens (1972) estão sempre associados à ideia de que a emoção representa uma variação de estado, que deriva em uma alteração da fala. A nosso ver, essa hipótese é correta, mas não se trata unicamente de uma alteração de estado fisiológico e, sim, de um estado cognitivo que dispara

certos recursos que nos permite manifestar a emoção na língua ou na fala.

Como pode se ver na análise das emoções, a distribuição da variações intrínsecas sugere que as condições emocionais poderiam ser distinguidas a partir dos alvos topológicos, ao passo que os alvos pontuais parecem apenas fornecer pistas auxiliares para esta distinção. A nosso ver, o espaço entoacional das emoções é fundamental para a implementação da variação e para a distinção de cada uma delas, como pode se ver na análise da raiva e da tristeza. Deste modo, como lidamos com um estudo de caso nesta tese, parece-nos que é apropriado uma análise mais extensiva do tema nessa etapa da pesquisa.

De modo geral, nosso objetivo com a análise das emoções era avaliar, em primeiro lugar, as nossas hipóteses linguísticas e, em segundo lugar, verificar se elas eram aplicáveis ao assunto. A nosso ver, os dois objetivos foram satisfeitos. Nesse sentido, nossas próximas etapas são avaliar um conjunto robusto de falas espontâneas que coletamos e fazer uma análise da fala emotiva em condições mais complexas.

Anexos

Anexo I

Figura 8.1: Exemplo de análise se usando o tom médio para a oração: “O Domenico entrar aqui...como assim?”.

Anexo II

	CofVar - MIDI	CofVar - Hz
<i>Estímulo</i>	10,4%	30,1%
suj1	9,8%	30,2%
repsuj1	19,4%	44,3%
suj2	3,3%	12,0%
repsuj2	4,9%	20,9%
suj3	11,2%	29,1%
suj4	5,0%	18,1%
repsuj4	19,4%	49,1%
suj5	21,6%	55,9%
repsuj5	14,1%	37,0%
suj6	5,7%	20,5%
repsuj6	5,9%	21,6%
suj7	15,1%	63,7%
repsuj7	7,7%	22,4%
suj8	10,4%	36,3%
repsuj8	11,4%	33,3%
suj9	16,1%	41,0%
repsuj9	19,1%	57,3%
suj10	8,9%	21,9%
repsuj10	11,9%	31,2%
suj11	20,2%	46,4%
repsuj11	11,4%	26,5%
suj12	14,8%	38,7%
repsuj12	14,7%	35,1%
suj13	12,7%	30,2%
Média	12,2%	34,1%

Valores do coeficiente de covariação de frequência fundamental de cada uma das amostras da frase 2

Anexo III

	i_r - MIDI	i_r - Hz
su1	-0,6%	0,1%
repsu1	9,0%	14,2%
su2	-7,0%	-18,1%
repsu2	-5,4%	-9,2%
su3	0,8%	-1,0%
su4	-5,4%	-12,0%
repsu4	9,1%	19,0%
su5	11,2%	25,8%
repsu5	3,8%	6,9%
su6	-4,6%	-9,6%
repsu6	-4,4%	-8,5%
su7	4,8%	33,6%
repsu7	-2,7%	-7,7%
su8	0,0%	6,2%
repsu8	1,0%	3,2%
su9	5,8%	10,9%
repsu9	8,7%	27,2%
su10	-1,5%	-8,2%
repsu10	1,6%	1,1%
su11	9,8%	16,3%
repsu11	1,0%	-3,6%
su12	4,5%	8,6%
repsu12	4,4%	5,0%
su13	2,4%	0,1%

Diferença entre o coeficiente de variação de cada uma das amostras e o estímulo.

Referências Bibliográficas

- AGOSTINHO, S. *City of God*. Buffalo: Christian Literature Publishing Co, 1887.
- AGOSTINHO, S. *A cidade de Deus*. 2. ed. Lisboa: Fundação Calouste Gulbenkian, 1996.
- ALBANO, E. C. *O gesto e suas bordas: esboço de Fonologia-Articulatória do Português*. Campinas: Mercado das Letras, 2009.
- ALLIGOOD, K. T.; SAUER, T. D.; YORKE, J. A. *Chaos*. Berlin: Springer, 1996.
- ANDERSON, A. K.; PHELPS, E. A. Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, Nature Publishing Group, v. 411, p. 305–309, maio 2001.
- AQUINO, T. de. *The Treatise on Human Nature: Summa Theologiae 1a 75-89*. Indianapolis: Hackett Publishing, 2002.
- ARISTÓTELES. *Metafísica*. São Paulo: Abril Cultural, 1984.
- ARISTÓTELES. *Ética a Nicômaco e Poética*. 4. ed. São Paulo: Nova Cultural, 1991. (Os Pensadores, v. 2).
- ARISTÓTELES. *Retórica*. 2ª. ed. Lisboa: Imprensa Nacional - Casa da Moeda de Portugal, 2005.
- ARNOLD, M. B. *Emotion and Personality*. New York: Columbia University Press, 1960.
- ASHBY, W. R. *An Introduction to Cybernetics*. London: Chapman & Hall, 1957.
- AVERILL, J. The social construction of emotion: With special reference to love. In: GERGEN, K. J.; DAVIS, K. E. (Ed.). *The social construction of the person*. New York: Springer-Verlag, 1985. p. 89–109.
- AVERILL, J. R. Autonomic response patterns during sadness and mirth. *Psychophysiology*, Wiley Online Library, v. 5, n. 4, p. 399–414, jan. 1969.
- AVERILL, J. R. A constructivist view of emotion. In: PLUTCHIK, R.; KELLERMAN, H. (Ed.). *Emotion: Theory, research and experience*. New York: Academic Press., 1980.
- BADDI, R.; POLITI, A. *Complexity: Hierarchical structures and scaling in physics*. [S.l.]: Cambridge University Press, 1999.
- BAER, T. *Investigation of phonation using excised larynx*. Tese (Doutorado) — M.I.T., Cambridge, 1975.
- BAKEN, R. J.; ORLIKOFF, R. F. *Clinical measurement of speech and voice*. 2. ed. Boston: Cengage Learning, 2000.

- BANSE, R.; SCHERER, K. R. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, American Psychological Association, v. 70, n. 3, p. 614, 1996.
- BANZIGER, T.; SCHERER, K. R. The role of intonation in emotional expressions. *Speech Communication*, v. 46, n. 1, p. 252–267, 2005.
- BÄNZIGER, T.; SCHERER, K. R. The role of intonation in emotional expressions. *Speech communication*, Elsevier, v. 46, n. 3, p. 252–267, 2005.
- BARD, P.; RIOCH, M. D. A study of four cats deprived of neocortex and additional portions of the forebrain. *John Hopkins Med. Journal*, v. 60, 1937.
- BARTHES, R. *Elementos de semiologia*. 19^a. ed. São Paulo: Editora Cultrix, 2012.
- BAZ, D. G. M.; FERREIRA-NETTO, W. Variação tonal na finalização de frases de pb. In: *III Colóquio Brasileiro de Prosódia da Fala*. Belo Horizonte: [s.n.], 2011.
- BEATTY, A. Emotions in the field: What are we talking about? *Journal of the Royal Anthropological Institute*, v. 11, n. 1, p. 17–37, 2005.
- BECKMAN, M.; PIERREHUMBERT, J. Intonational structure in japanese and english. In: EWEN, C.; ANDERSON, J. (Ed.). *Phonology Yearbook*. Cambridge: Cambridge University Press, 1986. v. 3, n. 1, p. 255–309.
- BELL, A. M. *Visible Speech: The Science of Universal Alphabets*. London: Simkin, Marshall & Co, 1867.
- BENVENISTE Émile. *Problemas de Linguística Geral I*. São Paulo: Nacional/Edusp, 1976.
- BERLIN, B.; KAY, P. *Basic color terms: Their universality and evolution*. Bekeley: California University Press, 1969.
- BERTALANFFY, L. von. *Teoria Geral dos Sistemas*. Petrópolis: Ed. Vozes, 1975.
- BODOLAY, A. N. *Pragmática da entonação: a relação prosódia/contexto em atos diretivos no Português*. Tese (Doutorado) — UFMG, Belo Horizonte, 2009.
- BOLINGER, D. L. Intonation: Level vs. configuration. *Word*, v. 7, n. 3, p. 191–210, 1951.
- BOLINGER, D. L. Intonation and grammar. *Language Learning*, VIII, n. 1&2, 1958.
- BOLINGER, D. L. Around the edge of language: Intonation. *Havard Educational Review*, v. 34, n. 2, 1964.
- BOLINGER, D. L. *Intonation and Its Parts: Melody in Spoken English*. Stanford: Stanford University Press, 1986.
- BOLINGER, D. L. *Intonation and Its Use: Melody in Grammar and Discourse*. Stanford: Stanford University Press, 1989.
- BONNER, M. R. Changes in the speech pattern under emotional tension. *The American Journal of Psychology*, JSTOR, v. 56, n. 2, p. 262–273, 1943.

- BOUDEWYNS, P. A. A comparison of the effects of stress vs. relaxation instruction on the finger temperature response. *Behavior Therapy*, Elsevier, v. 7, n. 1, p. 54–67, 1976.
- BROWMAN, C. P.; GOLDSTEIN, L. Articulatory gestures as phonological units. *Phonology*, v. 6, p. 201–251, 1989.
- BRUNSWIK, E. *Perception and the Representative Design of Psychological Experiments*. Berkley: University of California Press, 1956.
- CAGE, J. *Nova Musicha N. 1*. 1974. (CD).
- CAGLIARI, L. C. *Elementos de Fonética do Português Brasileiro*. São Paulo: Editora Paulistana, 1983.
- CAGLIARI, L. C.; MASSINI-CAGLIARI, G. O papel da tessitura dentro da prosódia portuguesa. *Razões e Emoção: Miscelânea de estudos oferecida a Maria Helena Mateus*, p. 67–85, jun. 2001.
- CANNON, W. B. *Bodily Changes In Pain Hunger Fear And Rage*. New York: D. Appleton and Company, 1927.
- CANNON, W. B. The james-lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, v. 39, n. 1, p. 106–124, dez. 1927.
- CHO, H. *A Weighted-Constraint Model of F0 Movements*. Tese (Doutorado) — M.I.T., Cambridge, 2010.
- CHOMSKY, N. Three models for the description of language. *IRE -Transactions on Information Theory*, v. 2, p. 113–124, 1956.
- CHOMSKY, N. On certain formal properties of grammars. *Information and Control*, v. 2, p. 137–167, 1959.
- CHOMSKY, N. A review of bfskinner's verbal behavior. *Language*, v. 35, n. 1, p. 26–58, 1959a.
- CHOMSKY, N. *Aspects of the theory of syntax*. Cambridge: MIT Press, 1965.
- CHOMSKY, N. *Knowledge of Language: Its Nature, Origins and Use*. New York: Praeger, 1986.
- CHOMSKY, N. *Linguagem e Mente*. 3. ed. São Paulo: Editora UNESP & Cambridge University Press, 2009.
- CHOMSKY, N.; HALLE, M. *The sound pattern of the English*. New York: Harper & Row, 1968.
- CÍCERO. *Disputaciones Tusculanas*. Madrid: Editorial Gredos, 2005.
- COHEN, A.; T'HART, J. On the anatomy of intonation. *Lingua*, n. 19, p. 177–192, 1967.
- COHEN, J. *Statistical power analysis for the behavioral sciences*. 2. ed. New York: Psychology Press, 1988.

CONNELL, B. Tone languages and the universality of intrinsic f0: evidence from africa. *Journal of Phonetics*, v. 30, p. 101–129, 2002.

CONSONI, F. *Aspectos da percepção da proeminência tonal em Português Brasileiro*. Tese (Doutorado) — Universidade de São Paulo, 2011.

CONSONI, F.; FERREIRA-NETTO, W. A percepção de variação em semitons ascendentes em palavras isoladas no português brasileiro. In: *Anais do X Congresso Nacional de Fonética e Fonologia / IV Congresso Internacional de Fonética e Fonologia*. [S.l.: s.n.], 2008.

CONSONI, F. et al. The brazilian speaker sensibility of perception in rising semitones variation for isolated words. In: *Book of Abstracts of 40th Poznan Linguistic Meeting*. Póznan: [s.n.], 2009.

CONSTANZO, F. S.; NORMAN, N. N.; COSTANZO, P. R. Voice quality profile and perceived emotion. *Journal of Counseling Psychology*, v. 16, n. 3, 1969.

CORNELIUS, R. R. Theoretical approaches to emotion. In: ISCA. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. Newcastle, 2000.

COWAN, N. *Working memory capacity*. New York: Psychology Press, 2005.

COWIE, R.; DOUGLAS-COWIE, E. Automatic statistical analysis of the signal and prosodic signs of emotion in speech. In: IEEE. *ICSLP 96*. [S.l.], 1996. v. 3, p. 1989–1992.

CRICK, F. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Scribner Book Company, 1995.

CRUZ-FERREIRA, M. Intonation in european portuguese. In: HIRST, D.; CRISTO, A. di (Ed.). *Intonation Systems*. Cambridge: Cambridge University Press, 1998.

CRYSTAL, D. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press, 1969.

DAMÁSIO, A. *O mistério da consciência: do corpo e das emoções ao conhecimento de si*. 2. ed. São Paulo: Cia. das Letras, 2015.

DAMÁSIO, A. R. *O erro de Descartes*. 3. ed. São Paulo: Companhia das Letras, 2012.

DARWIN, C. *A expressão das emoções no homem e nos animais*. São Paulo: Cia. das Letras, 1872/2016.

DAWSON, C. *Religion and the Rise of Western Culture: The Classic Study of Medieval Civilization*. New York: Doubleday, 2009.

DESCARTES, R. *Meditações*. São Paulo: Abril Cultural, 1983. ((Os Pensadores)).

DESCARTES, R. *As Paixões da Alma*. São Paulo: Lafonte, 2012.

DILLEY, L. *The phonetics and phonology of tonal systems*. Tese (Doutorado) — M.I.T., Cambridge, 2005.

DILLEY, L.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, Elsevier, v. 24, n. 4, p. 423–444, 1996.

D'IMPERIO, M. Language-specific and universal constraints on tonal alignment: The nature of targets and "anchors". In: *Speech Prosody 2002, International Conference*. [S.l.: s.n.], 2002.

DIXON, T. *From passions to emotions: the creation of a secular psychological category*. Cambridge: Cambridge University Press, 2003.

ECO, U. *A busca da Língua Perfeita na Cultura Europeia*. Bauru: EDUSC, 2001.

EGGER, H. L.; ANGOLD, A. Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, Wiley Online Library, v. 47, n. 3-4, p. 313–337, 2006.

EHLERS, R. *Análise de séries temporais*. São Paulo, 2009.

EKMAN, P. An argument for basic emotions. *Cognition & emotion*, v. 6, n. 3-4, p. 169–200, 1992.

EKMAN, P. What scientists who study emotion agree about. *Perspectives on Psychological Science*, SAGE Publications Sage CA: Los Angeles, CA, v. 11, n. 1, p. 31–34, 2016.

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

EKMAN, P.; FRIESEN, W. V. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, American Psychological Association, v. 17, n. 2, p. 124, 1971.

EKMAN, P.; FRIESEN, W. V. *Unmasking the face: A guide to recognizing emotions from facial clues*. Cambridge: Malor Books, 2003.

EKMAN, P.; LEVENSON, R. W.; FRIESEN, W. V. Autonomic nervous system activity distinguishes among emotions. *Science*, v. 221, n. 4616, p. 1208–1210, set. 1983.

FAIRBANKS, G.; HOAGLIN, L. W. An experimental study of the durational characteristics of the voice during the expression of emotion. *Communications Monographs*, v. 8, n. 1, p. 85–90, 1941.

FAIRBANKS, G.; PRONOVOST, W. An experimental study of the pitch characteristics of the voice during the expression of emotion. *Communications Monographs*, v. 6, n. 1, p. 87–104, 1939.

FANT, G. *Acoustic Theory of speech production*. Netherlands: Mouton: The Hague, 1966.

FANT, G. *Speech Sound and Features*. Cambridge: MIT Press Classic, 1973.

FAUL, F. et al. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, Springer, v. 39, n. 2, p. 175–191, 2007.

FERREIRA-NETTO, W. *Variação de frequência e constituição da prosódia da língua portuguesa*. Tese (Livre Docência) — Universidade de São Paulo, São Paulo, 2006.

FERREIRA-NETTO, W. Decomposição da entoação frasal em componentes estruturadoras e em componentes semântico-funcionais. In: *X Congresso Nacional de Fonética e Fonologia / IV Congresso Internacional de Fonética e Fonologia*. Niterói: UFF, 2008.

FERREIRA-NETTO, W. Explosodia. processo nº08992-2. *Revista da Propriedade Industrial*, v. 2038, p. 167, 2010.

FERREIRA-NETTO, W.; CONSONI, F.; PERES, D. O. Finalizações de frase em leituras e fala espontânea no pb. In: *57º Seminário do GEL*. Ribeirão Preto: [s.n.], 2009.

FERREIRA-NETTO, W.; CONSONI, F.; PERES, D. O. Finalisation des phrases en lecture et en parole spontanée: le cas du portugais brésilien. In: *XXVIIIèmes Journées d'Étude sur la Parole*. Mons: [s.n.], 2010.

FERREIRA-NETTO, W.; MARTINS, M. V. M. Proposal of description for an intonation pattern: the simulacrum of neutral intonation. To appear at *The Journal of the Acoustical Society of America*. jun. 2017.

FERREIRA-NETTO, W. et al. Brazilian portuguese intonation: A comparison between automatic and perceptual analyses. *The Journal of the Acoustical Society of America*, v. 132, n. 3, p. 2004, set. 2012.

FERREIRA-NETTO, W.; MARTINS, M. V. M.; VIEIRA, M. de F. Automatic analysis of emotional intonation in brazilian portuguese. *The Journal of the Acoustical Society of America*, v. 152, 2014.

FERREIRA-NETTO, W. et al. Análise automática de manifestações emocionais de tristeza e cólera em pb: abordagem pelo programa explosodia. *Revista Leitura*, v. 2, n. 1, p. 1–10, 2014.

FITCH, W. T.; NEUBAUER, J.; HERZEL, H. Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, Elsevier, v. 63, n. 3, p. 407–418, 2002.

FLANAGAN, J. L. *Speech analysis synthesis and perception*. 2.exp.. ed. [S.l.]: Springer Verlag Berlin Heidelberg GmbH, 1983.

FONÁGY, I. *La vive voix: Essais de psycho-phonétique*. Paris: Payot, 1991.

FONÁGY, I. *Languages Within Language: An evolutive approach*. Amsterdam: John Benjamins Publishing Company, 2001. (Foundations of Semiotics).

FONÁGY, I.; MAGDICS, K. Emotional patterns in intonation and music. *STUF-Language Typology and Universals*, v. 16, n. 1-4, p. 293–326, 1963.

- FONTANILLE, J. Le tournant modal en sémiotique. *Organon: Revista do Instituto de Letras da Universidade Federal do Rio Grande do Sul*, v. 23, p. 175–190, 1995.
- FORTENBAUGH, W. W. *Aristotle on Emotion*. 2ª. ed. London: Bristol Classical Press, 2002.
- FOWLER, C. A. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, v. 14, p. 3–28, 1986.
- FOWLER, C. A.; TURVEY, M. T. Immediate compensation in bite-block speech. *Phonetica*, Karger Publishers, v. 37, n. 5-6, p. 306–326, 1980.
- FREEMAN, W. J. *How Brains Make Up Their Minds*. New York: Columbia University Press, 2000.
- FREUD, S. O inconsciente (1915). In: *Obras Completas. Introdução ao Narcisismo e ensaios de Metapsicologia (1914-1916)*. São Paulo: Companhia das Letras, 2015. v. 12, p. 99–150.
- FREUD, S. A repressão (1915). In: *Obras Completas. Introdução ao Narcisismo e ensaios de Metapsicologia (1914-1916)*. São Paulo: Companhia das Letras, 2015. v. 12, p. 82–98.
- FREUD, S. *O mal-estar na civilização (1930)*. São Paulo: Penguin Classics & Companhia das Letras, 2016.
- FRIEDMAN, S. R. et al. Aspects of social and emotional competence in adult attention-deficit/hyperactivity disorder. *Neuropsychology*, American Psychological Association, v. 17, n. 1, p. 50, 2003.
- GARCIA, R. R. Evidências de finalizações plagais no dialeto caipira na região do médio tietê. *Revista Estudos Linguísticos*, v. 44, n. 1, p. 10–23, Jan.-Apr. 2015.
- GEISER, J. *Coupled Systems: Theory, Models, and Applications in Engineering*. Boca Ratón: Chapman & Hall - CRC Press, 2014.
- GOBL, C.; CHASAIDE, A. N. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, v. 40, n. 1-2, p. 189–212, abr. 2003.
- GODJEVAC, S. *Intonation, Word Order and Focus Projection in Serbo-Croatian*. Tese (Doutorado) — Ohio State University, Columbus, 2000.
- GOLDSMITH, J. *Autosegmental Phonology*. Tese (Doutorado) — M.I.T., Cambridge, 1976.
- GOLDSMITH, J. *Autosegmental and metrical phonology*. Cambridge: Blackwell, 1990.
- GOLDSMITH, J. A. *Autosegmental and metrical phonology*. Oxford: Basil Blackwell, 1990.
- GONZÁLES, A. M. Disputaciones tusculanas. In: _____. *Disputaciones Tusculanas*. Madrid: Editorial Gredos, 2005a. cap. Introducción, p. 7–95.
- GOREVAN, P. Aquinas and emotional theory today: mind-body, cognitivism and connaturality. *Acta Philosophica*, v. 9, n. 1, p. 141–151, 2000.
- GOULD, S. J. *Ontogenia y Filogenia*. Barcelona: Crítica, 2010.

- GRAY, H. *Anatomy of the Human Body*. Philadelphia: Lea & Febiger, 1918; Bartleby.com, 2000, 1918.
- GREIMAS, A. J.; FONTANILLE, J. *Semiótica das paixões: dos estados de coisas aos estados de alma*. São Paulo: Ática, 1993.
- GUSSENHOVEN, C. Transcription of dutch intonation. In: JUN, S. ah (Ed.). *Prosodic Typology: The Phonology of Intonation and Phrasing*. [S.l.]: Oxford University Press, 2006. p. 118–145.
- HALLE, M. Speculations about the representations of words in memory. In: HALLE, M. (Ed.). *From memory to speech and back: papers on Phonetics and Phonology 1954-2002*. New York: Mouton de Gruyter, 2002. p. 122–149. First appeared at Phonetic Linguistic Academic Press, Florida, pp.101-114.
- HALLIDAY, M. A. K. *Intonation and grammar in British English*. The Hague: Mouton, 1967.
- HAMAYAN, E.; SAEGERT, J.; LARUDEE, P. Elicited imitation in second language learners. *Language and Speech*, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 86–97, 1977.
- HEDRIH, K. S. Dynamics of coupled systems. *Nonlinear Analysis: Hybrid Systems*, v. 2, n. 2, p. 310–334, jun. 2008.
- HELMHOTZ, H. L. F. *On the sensations of tone as a physiological basis for the theory of music*. London: Longmans, Green and Co., 1895.
- HERZEL, H. et al. Analysis of vocal disorders with methods from nonlinear dynamics. *Journal of Speech, Language, and Hearing Research*, ASHA, v. 37, n. 5, p. 1008–1019, out. 1994.
- HIRST, D.; CRISTO, A. D. A survey of intonation systems. In: HIRST, D.; CRISTO, A. di (Ed.). *Intonation Systems*. Cambridge: Cambridge University Press, 1998.
- HIRST, D.; CRISTO, A. D. *A survey of intonation systems*. Cambridge: Cambridge University Press, 1998a.
- HJELMSLEV, L. *Prolegômenos a uma teoria da linguagem*. São Paulo: Editora Perspectiva, 2006.
- HOLLON, S. D.; PONNIAH, K. A review of empirically supported psychological therapies for mood disorders in adults. *Depression and anxiety*, Wiley Online Library, v. 27, n. 10, p. 891–932, 2010.
- HUME, D. *A Treatise of Human Nature*. [S.l.]: Domínio Público, 1740.
- INSEL, T. R. Oxytocin-a neuropeptide for affiliation: Evidence from behavioral, receptor autoradiographic, and comparative studies. *Psychoneuroendocrinology*, v. 17, n. 1, p. 3–35, 1992.
- INSEL, T. R.; SHAPIRO, L. E. Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 89, n. 13, p. 5981–5985, 1992.

- INSEL, T. R.; YOUNG, L. J. The neurobiology of attachment. *Nature Reviews: Neuroscience*, v. 2, n. 1, fev. 2001.
- JACKS, A. Bite block vowel production in apraxia of speech. *Journal of Speech, Language, and Hearing Research*, ASHA, v. 51, n. 4, p. 898–913, 2008.
- JAKOBSON, R. *Linguística e Comunicação*. 22. ed. São Paulo: Cultrix, 2008.
- JAKOBSON, R.; FANT, G.; HALLE, M. *Preliminaries to Speech Analysis*. Cambridge: MIT Press, 1952.
- JAMES, W. What is an emotion? *Mind*, v. 9, n. 34, p. 188–205, abr. 1884.
- JAMES, W. *The Principles of Psychology*. New York: Henry Holt and Company, 1890.
- JOHNSON-LAIRD, P.; MANCINI, F.; GANGEMI, A. A hyper-emotion theory of psychological illnesses. *Psychological review*, American Psychological Association, v. 113, n. 4, p. 822, 2006.
- JUN, S.-A. *Prosodic typology : the phonology of intonation and phrasing*. Oxford: Oxford University Press, 2006.
- KAKITA, Y.; HIRANO, M.; OHMARU, K. Physical properties of the vocal fold tissue. In: STEVENS, K. N.; HIRANO, M. (Ed.). *Vocal fold physiology*. Tokyo: University of Tokyo Press, 1981.
- KING, P. Emotions. In: DAVIES, B.; STUMP, E. (Ed.). *The Oxford Handbook of Aquinas*. Oxford: Oxford University Press, 2011.
- KLASMEYER, G. An automatic description tool for time-contours and long-term average voice features in large emotional speech databases. In: *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. [S.l.: s.n.], 2000.
- KLASMEYER, G.; SENDLMEIER, W. F. The classification of different phonation types in emotional and neutral speech. *Forensic Linguistics*, Routledge, v. 4, p. 104–124, 1997.
- KLEINGINNA, P. R.; KLEINGINNA, A. M. A categorized list of emotion definitions with suggestions for a consensual definition. *Motivation and Emotion*, v. 5, n. 4, p. 345–379, 1981.
- KREIBIG, S. D. Autonomic nervous system activity in emotion: a review. *Biological Psychology*, v. 84, n. 3, p. 394–421, 2010.
- KUHL, P. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, v. 50, n. 2, p. 93–107, 1991.
- KUMAR, R. K. S. et al. Study of nonlinear properties of vocal tract and its effectiveness in speaker modeling. *Journal of Acoustical Society of India: Vol*, v. 43, n. 2, p. 116–124, 2016.
- LABOV, W. On the mechanism of linguistic change. *Georgetown University Monographs on Language and Linguistics*, v. 18, p. 91–114, 1965.

- LABOV, W. *Padrões Sociolinguísticos*. São Paulo: Parábola Editorial, 2010.
- LABOV, W. *Principles of linguistic change, cognitive and cultural factors*. Oxford: John Wiley & Sons, 2011.
- LADD, D. R. Phonological features of intonational peaks. *Language*, JSTOR, p. 721–759, 1983.
- LADD, D. R. *Intonational Phonology*. Cambridge: Cambridge University Press., 1996.
- LADD, R.; SILVERMAN, K. E. A. Vowel intrinsic pitch in connected speech. *Phonetica*, Karger Publishers, v. 41, n. 1, p. 31–40, 1984.
- LADEFOGED, P.; MCKINNEY, N. P. Loudness, sound pressure, and subglottal pressure in speech. *The Journal of the Acoustical Society of America*, v. 35, n. 4, p. 454–460, abr. 1963.
- LANGE, C. G.; JAMES, W. *The Emotions*. Baltimore: Willians & Wilkins Company, 1922.
- LARA, C. C.; BATTISTI, E.; FILHO, A. A. D. O vozeamento/desvozeamento variável das oclusivas bilabiais em português numa comunidade teuto-brasileira e o relacionamento em rede dos falantes. *Working Papers em Linguística*, v. 14, n. 2, esp. 2013.
- LASS, R. *Historical linguistics and language change*. Cambridge: Cambridge University Press, 1997.
- LAUKKA, P. *Vocal Expression of Emotion*. Tese (Doutorado) — Uppsala University, Uppsala, 2004.
- LAZARUS, R. S. From psychological stress to the emotions: A history of changing outlooks. *Annual review of psychology*, v. 44, n. 1, p. 1–22, 1993.
- LEBEN, W. The representation of tone. In: FROMKIN, V. (Ed.). *Tone: A linguistic survey*. New York: Academic Press, 1978. p. 179–219.
- LEBEN, W. R. *Suprasegmental Theory*. Tese (Doutorado) — M.I.T., Cambridge, 1973.
- LEDOUX, J. *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster, 1998.
- LEDOUX, J. The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology*, v. 23, n. 4/5, p. 727–738, 2003.
- LEVENSON, R. W.; EKMAN, P.; FRIESEN, W. V. Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, Wiley Online Library, v. 27, n. 4, p. 363–384, 1990.
- LÉVI-STRAUSS, C. *Olha escutar ler*. São Paulo: Companhia das Letras, 2010.
- LIBERMAN, A.; MATTINGLY, I. The motor theory of speech perception revised. *Cognition*, v. 21, p. 1–36, 1985.

- LIBERMAN, M. *The intonational system of English*. Tese (Doutorado) — M.I.T., Cambridge, 1975.
- LIBERMAN, M.; PRINCE, A. On stress and linguistic rhythm. *Linguistic Inquiry*, v. 8, p. 249–336, 1977.
- LINDSEY, G. Intonation and pragmatics. *Journal of the International Phonetic Association*, Cambridge University Press, v. 11, n. 1, p. 2–21, 1981.
- LIST, G. The boundaries of speech and song. *Ethnomusicology*, v. 7, n. 1, p. 1–16, 1963.
- LOCKE, J. *An Essay Concerning Humane Understanding*. [S.l.]: HardPress Publishing, 1689.
- LORENZ, E. N. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, v. 20, n. 2, p. 130–141, mar. 1963.
- LUCENTE, L. *DaTo: um sistema de notação entoacional para o português brasileiro baseado em princípios dinâmicos. Ênfase no foco e na fala espontânea*. Dissertação (Mestrado) — UNICAMP, 2008.
- LUCERO, J. C. et al. Validation of theoretical models of phonation threshold pressure with data from a vocal fold mechanical replica. *The Journal of the Acoustical Society of America*, v. 125, n. 2, p. 632–635, jan. 2009.
- LUTZ, C. A. *Unnatural emotions: Everyday sentiments on a Micronesian atoll and their challenge to western theory*. Chicago: University of Chicago Press, 1988.
- MACIEL, J. *Elementos de Teoria Geral dos Sistemas*. Petrópolis: Editora Vozes, 1974.
- MACLEAN, P. D. Psychosomatic disease and the visceral brain; recent developments bearing on the papez theory of emotion. *Psychosom Med.*, v. 11, n. 6, p. 338–53, nov. 1949.
- MAEDA, S. *A Characterization of American English Intonation*. Tese (Doutorado) — M.I.T., Cambridge, 1976.
- MARION, S. T. T. . J. B. *Classical dynamics of particles and systems*. 5. ed. Belmont: Thomsom - Brooks/Cole, 2008.
- MARKUS, H. R.; KITAYAMA, S. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, American Psychological Association, v. 98, n. 2, p. 224, 1991.
- MARTINET, A. *Elementos de lingüística geral*. São Paulo: Martins Fontes, 1975.
- MARTINS, M. V. M. Análise de tom médio em narrativas orais. In: *17º Simpósio Internacional de Iniciação Científica da USP*. São Paulo: [s.n.], 2009. v. 17.
- MARTINS, M. V. M. *Aspectos do controle e da percepção do Português Brasileiro*. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2013.

- MARTINS, M. V. M.; FERREIRA-NETTO, W. Prosódia e escalas de frequência: um estudo em torno da escala de semitons. *ReVEL: Revista Virtual de Estudos da Linguagem*, v. 8, n. 15, p. 286–296, 2010.
- MARTINS, M. V. M.; VASSOLER, A. M. D. O. A entoação em falas teatrais: uma análise da raiva e da fala neutra. *Estudos Linguísticos (São Paulo. 1978)*, GEL, São Paulo, v. 1, p. 9–18, 2013.
- MATSUMOTO, D. Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 92–105, 1989.
- MCHUGO, G. J. et al. Emotional reactions to a political leaders expressive displays. *Journal of Personality and Social Psychology*, American Psychological Association, v. 49, n. 6, p. 1513, 1985.
- MCLUHAN, M. *Os meios de comunicação como extensões do homem*. 1ªed., 15ª reimpr. São Paulo: Cultrix, 2007.
- MCMAHON, A. M. *Understanding language change*. Cambridge: Cambridge University Press, 1994.
- MEDEIROS, B. R. de; MARTINS, M. V. M. Uma proposta de modelamento da percepção da entoação do português brasileiro. *Alfa: Revista de Linguística*, v. 58, p. 195–213, 2014.
- MENDE, W.; HERZEL, H.; WERMKE, K. Bifurcations and chaos in newborn infant cries. *Physics Letters A*, Elsevier, v. 145, n. 8-9, p. 418–424, abr. 1990.
- MESQUISTA, B.; FRIJDA, N. H. Cultural variations in emotions:a review. *Psychological Bulletin*, v. 112, n. 2, p. 179–204, 1992.
- MORAES, J. A. de. Intonation in brazilian portuguese. In: HIRST, D.; CRISTO, A. di (Ed.). *Intonation Systems*. Cambridge: Cambridge University Press, 1998.
- MORAES, J. A. de; CARNAVAL, M.; COELHO, A. B. B. A manifestação prosódica do foco em interrogativas totais no português do brasil e sua percepção. *ReVEL*, Esp., n. 10, 2010.
- MORAES, J. A. de; COLAMARCO, M. Você está pedindo ou perguntando? uma análise entonacional de pedidos e perguntas no português do brasil. *Revista de Estudos da Linguagem*, v. 15, n. 2, abr. 2007.
- MORETTIN, P. A. *Ondas e Ondaletas: Da Análise de Fourier à Análise de Ondaletas*. São Paulo: Edusp, 1999.
- MOURA, L. da S. *O papel da prosódia na expressão de atitudes de ataque ao ethos no discurso político*. Dissertação (Mestrado) — UFOP, Ouro Preto, 2016.
- MUNKRES, J. R. *Topology*. 2. ed. New Jersey: Prentice Hall, 2000.
- NEGUS, V. E. *The comparative anatomy and physiology of the Larynx*. Londres: Heinemann, 1948.

- NESPOR, M.; VOGEL, I. *Prosodic Phonology*. 2. ed. Berlin: Mouton de Gruyter, 2007.
- NEUBAUER, J.; EDGERTON, M.; HERZEL, H. Nonlinear phenomena in contemporary vocal music. *Journal of Voice*, v. 18, n. 1, p. 1–12, jan. 2004.
- NEUMANN, J. von. The first draft report on the edvac. *IEEE Annals of the History of Computing*, v. 15, n. 4, p. 27–75, 1993.
- OHALA, J. There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics*, v. 18, p. 153–171, 1990.
- OHALA, J. J. Production of tone. In: FROMKIN, V. (Ed.). *Tone: A linguistic survey*. [S.l.]: Academic Press New York, 1978. p. 5–39.
- OHALA, J. J. The origin of sound patterns in vocal tract constraints. In: MACNEILAGE, P. F. (Ed.). *The Production of Speech*. New York: Springer, 1983.
- OHALA, J. J. Speech perception is hearing sounds, not tongues. *The Journal of Acoustical Society of America*, v. 99, n. 3, p. 1718–1725, mar. 1996.
- OHALA, J. J.; EUKEL, B. W. Explaining the intrinsic pitch of vowels. In: CHANNON, R.; SHOCKEY, L. (Ed.). *In honor of Ilse Lehiste*. Dordrecht: Foris, 1987, (Netherlands Phonetic Archives, v. 6). p. 207–215.
- OKURA, E.; LONSDALE, D. Working memory's meager involvement in sentence repetition tests. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. [S.l.: s.n.], 2012. p. 2132–2137.
- PAIDI, G.; KADIRI, S. R.; YEGNANARAYANA, B. Analysis of emotional speech-a review. In: ESPOSITO, A.; JAIN, C. L. (Ed.). *Toward Robotic Socially Believable Behaving Systems*. New York: Springer International Publishing, 2016. v. 1, p. 205–238.
- PAKOSZ, M. Attitudinal judgments in intonation: some evidence for a theory. *Journal of Psycholinguistic Research*, Springer, v. 12, n. 3, p. 311–326, 1983.
- PANKSEPP, J. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press, 2004.
- PAPEZ, J. W. A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, v. 38, n. 4, p. 725–743, 1937.
- PEDERSEN, C. et al. Oxytocin induces maternal behavior in virgin female rats. *Science*, American Association for the Advancement of Science, v. 216, n. 4546, p. 648–650, 1982.
- PERES, D. O.; CONSONI, F.; FERREIRA-NETTO, W. A influência da cadeia segmental na percepção de variações tonais. *LLJournal*, v. 6, n. 1, 2011.
- PESSOA, L. Emotion and cognition and the amygdala: From what is it? to what's to be done? *Neuropsychologia*, v. 12, n. 48, p. 3416–3429, 2010.
- PIERREHUMBERT, J. *The Phonology of Phonetic of English Intonation*. Tese (Doutorado) — M.I.T., Cambridge, 1980.

PIERREHUMBERT, J. Prosodic effects on glottal allophones. In: FUJIMURA, O.; HIRANO, M. (Ed.). *Vocal Fold Physiology 8: Voice Quality Control*. San Diego: Singular Press, 1995. p. 39–60.

PIERREHUMBERT, J. Tonal elements and their alignment. In: HORNE, M. (Ed.). *Prosody, Theory and Experiment: studies presented to Gösta Bruce*. Dordrecht: Kluwer Academic Publisher, 2000. p. 11–36.

PIERREHUMBERT, J.; TALKIN, D. Lenition of /h/ and glottal stop. In: DOCHERTY, G. J.; LADD, R. (Ed.). *Papers in laboratory phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press, 1992. p. 90–117.

PIKE, K. L. *The intonation of American English*. Michigan: Ann Arbor: University of Michigan Press, 1945.

PIKE, K. L. On the grammar of intonation. In: *Proceedings of the Fifth International Congress of Phonetic Sciences*. Münster: S. Karge, 1964. p. 105–119.

PIKOVSKY, A.; ROSENBLUM, M.; KURTHS, J. *Synchronization: a universal concept in nonlinear sciences*. New York: Cambridge University Press, 2001. (Cambridge Nonlinear Science Series).

PINKER, S. *Tábula rasa: a negação contemporânea da natureza humana*. São Paulo: Companhia das Letras, 2002.

PLATÃO. *A República*. São Paulo: Difusão Europeia do Livro, 1965.

PLATÃO. *Apologia, Banquete e Fedro*. São Paulo: Levoir, 2010.

PLUTCHIK, R. The nature of human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, Sigma Xi, The Scientific Research Society, v. 89, n. 4, p. 344–350, 2001.

PROMON, S.; XU, Y.; THIPAKON, B. Modeling tone and intonation in mandarin and english as a process of target approximation. *The Journal of the Acoustical Society of America*, v. 125, n. 1, p. 405–424, jan. 2009.

ROBERTS, R. C. Thomas aquinas on the morality of emotions. *History of Philosophy Quarterly*, v. 9, n. 3, p. 287–305, jul. 1992.

ROSA, R. C. de M. *Percepção e ExProsodia: correlação entre análise automática e a finalização de frases assertivas isoladas do Português Brasileiro*. Dissertação (Mestrado) — Universidade de São Paulo, ago. 2015.

ROUSSEAU, J.-J. *Do Contrato Social*. São Paulo: Penguin Classics e Companhia das Letras, 2011.

RUSSELL, J. A. Culture and the categorization of emotions. *Psychological bulletin*, American Psychological Association, v. 110, n. 3, p. 426, 1991.

SALTZMAN, E.; KELSO, J. A. S. Skilled actions: A task-dynamic approach. *Psychological review*, American Psychological Association, v. 94, n. 1, p. 84, 1987.

- SAUSSURE, F. de. *Curso de Linguística Geral*. 28^a. ed. São Paulo: Editora Cultrix, 2012.
- SCARANTINO, A. The philosophy of emotion and its impact on affective science. In: BARRETT, L. F.; LEWIS, M.; JONES, J. M. H. (Ed.). *Handbook of Emotions*. 4. ed. New York: Guilford Press, 2016. p. 37–256.
- SCHERER, K. R. Vocal affect signalling: A comparative approach. *Advances in the study of behavior*, v. 15, p. 189–244, 1985.
- SCHERER, K. R. Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, v. 99, n. 2, p. 143–165, 1986.
- SCHERER, K. R. Towards a dynamic theory of emotion: The component process model of affective states. *Geneva studies in emotion and communication*, v. 1, p. 1–98, 1987.
- SCHERER, K. R. On the sequential nature of appraisal processes: Indirect evidence from a recognition task. *Cognition & Emotion*, Taylor & Francis, v. 13, n. 6, p. 763–793, 1999.
- SCHERER, K. R. Appraisal considered as a process of multilevel sequential checking. In: SCHERER, K. R.; SCHORR, A.; JOHNSTONE, T. (Ed.). *Appraisal processes in emotion: theory, methods, research*. New York: Oxford University Press, 2001. p. 92–120.
- SCHERER, K. R. Vocal communication of emotion: A review of research paradigms. *Speech communication*, Elsevier, v. 40, n. 1, p. 227–256, 2003.
- SCHERER, K. R. What are emotions? and how can they be measured? *Social science information*, Sage Publications, v. 44, n. 4, p. 695–729, 2005.
- SCHMERLING, S. F. *Aspects of English sentence stress*. Austin: University of Texas Press, 1976.
- SÊNECA. *Sobre a Ira. Sobre a Tranquilidade da alma*. São Paulo: Penguin Classics Companhia das Letras, 2014.
- SHANNON, C.; WEAVER, W. *The Mathematical Theory of Communication*. 1^a ed. 10^o printing. Urbana: University of Illinois Press - Urbana, 1964.
- SHANNON, C. E. *An algebra for theoretical genetics*. Tese (Doutorado) — M.I.T., Cambridge, 1940.
- SHANNON, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, v. 27, p. 379–423, Jul-Oct 1948.
- SHATTUCK-HUFNAGEL, S. et al. Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, v. 3, n. 2, p. 337–383, out. 2012.
- SILVERMAN, K. et al. Tobi: a standard for labeling english prosody. In: *Proceedings, Second International Conference on Spoken Language Processing*. Banff, Canada: [s.n.], 1992. v. 2, p. 867–870.
- SIMONOV, P. V.; FROLOV, M. Utilization of human voice for estimation of man's emotional stress and state of attention. *Aerospace Medicine*, v. 44, n. 3, p. 256–258, 1973.

SKINNER, R. E. A calibrated recording and analysis of the pitch, force and quality of vocal tones. *Communications Monographs*, Taylor & Francis, v. 2, n. 1, p. 81–137, 1935.

SLOBIN, D.; WELSH, C. A. Elicited imitation as a research tool in developmental psycholinguistics. In: FERGUSON, C.; SLOBIN, D. (Ed.). *Studies of child language development*. New York: Holt, Rinehart and Winston Inc, 1973.

SMITH, B. L. Effects of bite block speech on intrinsic segment duration. *Phonetica*, Karger Publishers, v. 44, n. 2, p. 65–75, 1987.

SMITH, C. A.; ELLSWORTH, P. C. Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, American Psychological Association, v. 48, n. 4, p. 813, 1985.

STAATS, A. W.; EIFERT, G. H. The paradigmatic behaviorism theory of emotions: Basis for unification. *Clinical Psychology Review*, Elsevier, v. 10, n. 5, p. 539–566, 1990.

STEVENS, K. N. *Acoustic Phonetics*. Cambridge: MIT Press, 2000.

STEVENS, K. N.; WILLIAMS, C. E. On determining the emotional state of pilots during flight: an exploratory study. *Aerospace Medicine*, v. 40, p. 1369–1372, 1969.

STORCH, E. A. et al. Rage attacks in pediatric obsessive-compulsive disorder: phenomenology and clinical correlates. *Journal of the American Academy of Child & Adolescent Psychiatry*, v. 51, n. 6, p. 582–592, 2012.

STRATOS, D. *Cantare la voce*. 1978. CD.

STUDDERT-KENNEDY, M.; GOLDSTEIN, L. Lauching language: The gestural origin of discrete infinity. In: CHRISTIANSEN, M. H.; KIRBY, S. (Ed.). *Language Evolution*. Oxford: Oxford University Press, 2003.

SUCHY, Y. *Clinical Neuropsychology of Emotion*. New York: Guilford Press, 2011.

T'HART, J. Differential sensitivity to pitch distance, particularly in speech. *Journal of Acoustical Society of America*, v. 69, n. 3, p. 811–821, 1981.

T'HART, J.; COLLIER, R.; COHEN, A. *A perceptual study of intonation*. Cambridge: Cambridge University Press., 1990.

THOITS, P. A. The sociology of emotions. *Annual review of sociology*, v. 15, n. 1, p. 317–342, 1989.

THOMPSON, J. M. T.; STEWART, H. B. *Nonlinear dynamics and chaos*. New Jersey: John Wiley & Sons, 2002.

TITZE, I. R. The physics of small-amplitude oscillation of the vocal folds. *The Journal of the Acoustical Society of America*, v. 83, n. 4, p. 1536–1552, abr. 1988.

TITZE, I. R. On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, v. 85, n. 2, p. 901–906, fev. 1989.

- TITZE, I. R. Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, v. 85, n. 4, p. 1699–1707, abr. 1989a.
- TITZE, I. R. Nonlinear source-filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America*, ASA, v. 123, n. 4, p. 2733–2749, maio 2008.
- TITZE, I. R.; RIEDE, T. Nonlinear source-filter coupling in phonation: Vocal exercises. *The Journal of the Acoustical Society of America*, ASA, v. 123, n. 4, p. 1902–1915, abr. 2008.
- TOMKINS, S. S.; MCCARTER, R. What and where are the primary affects? some evidence for a theory. *Perceptual and motor skills*, v. 18, n. 1, p. 119–158, 1964.
- TRAGER, G. L. Paralanguage: A first approximation. *Studies in Linguistics*, v. 13, p. 1–12, 1958.
- TRAGER, G. L. The typology of paralanguage. *Anthropological Linguistics*, JSTOR, p. 17–21, 1961.
- TRAGER, G. L.; SMITH, H. L. *An outline of English structure*. Washington: American Council of Learned Societies, 1957.
- TROUBETZKOY, N. *Principes de Phonologie*. Paris: Klincksieck, 1976.
- TSUCHIYA, N. et al. Intact rapid detection of fearful faces in the absence of the amygdala. *Nature Neuroscience*, n. 12, p. 1224–1225, 2009.
- TURING, A. Computing machinery and intelligence. *Mind*, MIT Press, Cambridge, v. 59, n. 236, p. 433–460, out. 1950.
- UTSUKI, N.; OKAMURA, N. Relationship between emotional state and fundamental frequency of speech. *Reports of the Aeromedical Laboratory, Japan Air Self-Defense Force*, v. 16, p. 179–188, 1976.
- VAISSIÈRE, J. Perception of intonation. In: PISONI, D. B.; REMEZ, R. E. (Ed.). *Handbook of Speech Perception*. Oxford: Blackwell, 2004. p. 236–263.
- VASSOLER, A. M. D. O. *Aspectos acústicos da voz na representação teatral das emoções*. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2011.
- VAZ, N. M. Ideias para uma nova imunologia. *Ciência Hoje*, v. 7, p. 32–38, 1983.
- VERNANT, J.-P.; VIDAL-NAQUET, P. *Mito e Tragédia na Grécia Antiga*. 1ª. ed. São Paulo: Editora Perspectiva, 2005.
- VINTHER, T. Elicited imitation: a brief overview. *International journal of applied linguistics*, Wiley Online Library, v. 12, n. 1, p. 54–73, 2002.
- WALLBOTT, H. G.; SCHERER, K. R. How universal and specific is emotional experience?: Evidence from 27 countries on five continents. *Social Science Information*, Lawrence Erlbaum Associates, Inc, v. 25, n. 4, dez. 1986.
- WARD, G.; HIRSCHBERG, J. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, JSTOR, p. 747–776, 1985.

- WATERS, W. F.; BERNARD, B. A.; BUCO, S. M. The autonomic nervous system response inventory (ansri): Prediction of psychophysiological response. *Journal of Psychosomatic Research*, Elsevier, v. 33, n. 3, p. 347–361, 1989.
- WEAVER, A. T. Experimental studies in vocal expression. *The Journal of Applied Psychology*, v. 8, n. 1, p. 23–51, 1924.
- WEAVER, A. T. Experimental studies in vocal expression. *Quarterly Journal of Speech*, v. 10, n. 3, p. 199–204, 1924.
- WHALEN, D. H.; LEVITT, A. G. The universality of intrinsic f_0 of vowels. *Journal of Phonetics*, v. 23, n. 3, p. 349–366, 1995.
- WHITEHEAD, A. N.; RUSSELL, B. *Principia Mathematica*. Michigan: Ann Arbor: University of Michigan Press, 2005. Disponível em: <<http://name.umdl.umich.edu/AAT3201.0001.001>>.
- WILCE, J. M. *Language and Emotion*. London: Cambridge University Press, 2009.
- WILLIAMS, C. E.; STEVENS, K. N. Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, ASA, v. 52, n. 4, p. 1238–1250, 1972.
- WITTGENSTEIN, L. *Philosophical Investigations*. New York: Oxford University, 1953.
- XU, Y. The penta model of speech melody: Transmitting multiple communicative functions in parallel. *Proceedings of from sound to sense*, v. 50, p. 91–96, 2004.
- XU, Y. Speech melody as articulatorily implemented communicative functions. *Speech Communication*, v. 46, p. 220–251, 2005.
- XU, Y.; SUN, X. Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, v. 111, n. 3, p. 1399–1413, mar. 2002.
- XU, Y.; WANG, Q. E. What can tone studies tell us about intonation? In: BOTINIS, A.; KOUROPETROGLOU, G.; CARAYANNIS, G. (Ed.). *Intonation: Theory, Models and Applications. Proceedings of an ESCA workshop*. Atenas: European Speech Communication Association, 1997. p. 337–340.
- XU, Y.; WANG, Q. E. Pitch targets and their realization: Evidence from mandarin chinese. *Speech Communication*, v. 33, p. 319–337, 2001.
- ZHOU, G.; HANSEN, J. H.; KAISER, J. F. Nonlinear feature based classification of speech under stress. *IEEE Transactions on speech and audio processing*, IEEE, v. 9, n. 3, p. 201216, 2001.