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

Aspectos da percepção e do controle entoacional do Português Brasileiro

São Paulo
2012

Marcus Vinícius Moreira Martins

Aspectos da percepção e do controle entoacional do Português Brasileiro

Dissertação 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 tíulo de Mestre em Filologia e Língua Portuguesa.

Orientadora:
Prof^{a}. Dr^{a}. Beatriz Raposo de Medeiros

São Paulo - SP

2012

MARTINS, M.V.M. **Aspectos da percepção e do controle entoacional do Português Brasileiro**. Dissertação apresentada a Faculdade de Filosofia Letras e Ciências Humanas da Universidade de São Paulo para a obtenção do título de Mestre em Filologia e Língua Portuguesa

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

Prof^{a}. Dr^{a}. Beatriz Raposo de Medeiros
Departamento de Letras Clássicas e Vernáculas - USP
Orientadora

Prof. Dr. .....
Instituição: .....
Julgamento: .....

Prof. Dr. .....
Instituição: .....
Julgamento: .....

Aos meus pais Cida e Marcos Martins.
Que me deram a linguagem.

Agradecimentos

Ohana quer dizer *família* e *família* quer dizer *nunca abandonar ou esquecer alguém*, assim já diziam Lilo & Stitch, por esta razão agradeço, em primeiro lugar, à minha família. Meus pais, Marcos e Cida, e meus irmãos, Paulo e Larissa, que foram, são e sempre serão a minha medida de todas as coisas, sem eles as coisas não teriam o mesmo valor, a vida não teria tanta graça. Agradeço ao "clã dos Martins" pela confiança, pelo carinho, apoio e afeto, pois como dizia nosso patriarca *Família é coisa séria* e guardamos isso como obrigação. Agradeço também aos Moreiras pelo carinho e confiança, que mesmo na distância, sempre foram sentidos.

Agradeço à minha orientadora, Profª. Drª. Beatriz Raposo de Medeiros, quem primeiro me apresentou os domínios quase obscuros da fonética (obscuros ao menos para um aluno de graduação) e assumiu a responsabilidade de me orientar quando foi preciso e quando desejei que estes domínios não fossem mais tão obscuros. Seus ensinamentos sobre fonética/fonologia, sobre o método, sobre a honestidade intelectual e a necessidade de clareza do pensamento me foram fundamentais, os guardarei sempre. Leitora atenciosa, os erros teóricos e de escrita que permanecem nessa dissertação são culpa minha e da minha leitura prematura de Wittgenstein.

Agradeço ao Prof. Dr. Waldemar Ferreira Netto, que me mostrou um mundo de fenômenos suprassegmentais incompreendidos, mas inteligíveis com o método correto e a dedicação necessária. Agradeço também pela leitura atenta do meu relatório de qualificação, os erros teóricos que permanecem são culpa minha.

Agradeço ao Prof. Dr. Paulo Chagas pela leitura atenta de meu relatório de qualificação e sugestões.

À CAPES e à USP pelo apoio financeiro e institucional.

Agradeço aos colegas de ExProsodia e do Grupo de Estudos em Fonética. Nossas conversas e discussões foram essenciais para este trabalho. Agradeço, em especial, ao Daniel Peres e a Renata Rosa por terem me guiado no misterioso mundo dos documentos burocráticos e projetos de pesquisa e à Fernanda Consoni, cujo trabalho inspirou o meu. Agradeço também à Aline Vassoler por me confiar os dados de sua

dissertação, essenciais para a compreensão de alguns fenômenos da entoação.

Agradeço ao Rodolfo Souza por me emprestar sua voz para as gravações deste trabalho e aos bixos da graduação que participaram ativamente do processo de aplicação do teste de percepção, sempre empolgados. Sem a ajuda deles, com certeza, este seria um trabalho menor.

Agradeço aos meus amigos de Mogi, da USP e de todos os lugares por me serem tão caros e queridos, cada um me marcou de algum modo, em algum momento durante meu processo (inacabado) de formação pessoal ou intelectual, seria injusto indicar os mais especiais.

Agradeço à Jéssica Costa e à Regiane Matos que sem saber me ajudaram a não desistir do curso de Letras: qualquer título que eu venha a ter, deverei à elas. À Bruna Polachini que me convenceu a ser linguista e a continuar sendo um; do contrário hoje eu seria um engenheiro ou matemático (de verdade). Agradeço ao Raphael Guimarães que me ofereceu sua casa, logo que ingressei no mestrado e no Instituto de Matemática e me vi sem moradia. Eles são os culpados por esse trabalho.

E, por fim, a Deus com quem tenho feito às pazes, graças aquela moça loira que me disse sem dizer,que: "We're the reflection of imperfection, we come from the infinite place of limitation".

Quanto mais penso na linguagem...
mais me surpreendo com o fato de as pessoas
conseguirem entender umas às outras
Kurt Gödel

Resumo

O objetivo deste trabalho é determinar os limiares de diferenciação tonal (LDTs) do Português Brasileiro, no que se refere à entoação. LDTs são definidos como valores a partir dos quais a percepção de uma determinada grandeza passa a ser perceptualmente relevante. Consoni (2011) determinou que tais valores seriam de +3 e -4 semitons para palavras manipuladas em uma frase. Nosso objetivo é rever tais valores para o nível frasal. Nossa hipótese é a de que a extensão temporal afeta a percepção, graças ao efeito *tau* (SHIGENO, 1986). Para este fim foram aplicados dois experimento com 13 pares de uma mesma oração falada por uma voz masculina não-marcada. Os pares consistiam de uma sentença neutra (com tom zero) pareada à outras doze com F0 manipulados (6 semitons ascendentes e 6 semitons descendentes, a partir do zero), mais o par zero-zero. O primeiro teste possuía apenas duas alternativas "sim" e "não"e foi chamado de teste 2AFC, ao passo que no segundo havia três opções: "sim", "não" e "talvez", chamado de 3MFC. Aos participantes foi requisitado que respondessem se notavam ou não qualquer diferença entre os pares apresentados. Os testes foram aplicado em 16 pessoas. O teste3 MFC teve seus resultados desconsiderados, seus dados foram utilizados como um padrão qualitativo para o teste de nossa hipótese. A seguir foi aplicada uma ANOVA entre as variáveis do teste 2AFC a fim de se determinar se havia algum efeito, obtivemos um $p < 0,001$, para $n = 16$, $Fo > Fe$. Para determinar os limiares de diferenciação tonal foi aplicada uma análise de componentes principais, a qual retornou como limiares os valores de -3 e +3 semitons

Palavras-chave: fonética; entoação; percepção; fonologia; prosódia.

Abstract

The aim of this study is to determine the differential treshold of pitch (DTP's) for Brazilian Portuguese, with regard to intonation. DTP's are defined as values ??from which the perception of a certain greatness becomes perceptually relevant. Consoni et (2011) determined that such values ??would be of +3 and -4 semitones for manipulated words in a sentence. Our goal is to review these values ??from the phrasal level. Our hypothesis is that the temporal extension affects perception, due to the effect *tau* (SHIGENO, 1986). For this purpose two experiments were applied using 13 pairs of the same sentence, spoken by a male voice unmarked. The pairs consisted of a neutral setence (with zero tone) paired with the others twelve sentences with manipulated F0 (6 semitones ascending semitones and 6 descendants, from zero), plus the zero-zero pair. The first test had only two alternatives "yes" and "no" and was named 2AFC test, while the second one had three options: "yes", "no" and "maybe" and was named 3MFC. Participants were asked to answer whether or not noticed any difference between the presented pairs. The tests were applied to 16 people. The results from 3MFC test were disregarded, the 3MFC data were used as a qualitative standard for testing our hypothesis. Following, ANOVA was applied in the 2AFC test in order to determine whether there was any effect, we obtained a $p < 0.001$ to $n = 16$, $Fo > Fe$. To determine the differential treshold of pitch was applied principal component analysis, which returned as threshold values ??of -3 and +3 semitones.

Keywords: phonetics; intonation; perception; phonology; prosody.

Sumário

Lista de Abreviaturas p. 5

Lista de Figuras p. 6

Lista de Tabelas p. 9

1 Introdução p. 10

1.1 Apresentação do tema . . . . . p. 10

1.2 Objetivos . . . . . p. 11

1.3 Visão Geral . . . . . p. 13

2 Um modelo para a entoação p. 14

2.1 Sistemas . . . . . p. 14

2.1.1 Sistemas e Linguística . . . . . p. 14

2.1.2 Sistemas, intervalos e a proposta de Brooks (1997) . . . . . p. 20

2.1.3 Um esboço de modelo para a entoação . . . . . p. 27

2.1.3.1 O Tom Médio . . . . . p. 30

2.2 Visão do Capítulo . . . . . p. 34

3 Entoação p. 35

3.1 A Fonologia da entoação . . . . . p. 36

3.1.1 Os primeiros passos . . . . . p. 36

3.2 Fonologia e Entoação ou Fonologia da entoação? . . . . . p. 39

3.3 Os estudos de entoação. . . . . p. 42

3.4 A Fonética da entoação . . . . . p. 44
3.4.1 A Laringe . . . . . p. 44
3.4.2 A glote . . . . . p. 47
3.4.3 A produção da frequência fundamental . . . . . p. 51
3.5 A Percepção da entoação . . . . . p. 55
3.5.1 Propriedades da percepção . . . . . p. 55
3.5.2 Jack LaMotta e a percepção . . . . . p. 57
3.5.3 Percepção e língua . . . . . p. 59
3.5.4 A proposta de t´Hart, Collier e Cohen (1990) . . . . . p. 62
3.6 Discussão . . . . . p. 63

4 Metodologia p. 66
4.1 A questão experimental nos estudos linguísticos . . . . . p. 66
4.2 A preparação experimental . . . . . p. 69
4.2.1 Gravação dos estímulos . . . . . p. 70
4.2.2 Análise das gravações . . . . . p. 71
4.2.2.1 Análise da sentença *O João mora na casa verde* . . . . . p. 72
4.2.2.2 Análise da sentença *O gato está naquele muro* . . . . . p. 75
4.2.2.3 Hz *versus* semitons: uma escolha metodológica . . . . p. 78
4.2.3 Manipulação da frequência fundamental . . . . . p. 81
4.3 O delineamento experimental . . . . . p. 83
4.3.1 A montagem do experimento . . . . . p. 84
4.3.2 O experimento . . . . . p. 87
4.3.3 O tamanho da amostra . . . . . p. 92
4.4 Aplicação dos testes . . . . . p. 94

5 Análise p. 96

5.1 Tratamento dos dados . . . . . p. 96
5.2 Primeiras análises . . . . . p. 97
5.3 Análise de dados: testes de hipóteses . . . . . p. 100
5.4 Análise fatorial . . . . . p. 105
5.5 Síntese dos testes aplicados . . . . . p. 109
5.6 Análise dos dados do experimento 3Multiple Forced Choice . . . . . p. 110

6 Considerações Finais p. 113
6.1 Um breve resumo . . . . . p. 113
6.2 Os limiares de diferenciação tonal . . . . . p. 114
6.3 Palavras finais . . . . . p. 122

Referências p. 124

Anexos p. 133
Anexo I . . . . . p. 133
Anexo II . . . . . p. 134
Anexo III . . . . . p. 135
Anexo IV . . . . . p. 136
Anexo V . . . . . p. 137
Anexo VI . . . . . p. 139
Anexo VII . . . . . p. 140
Anexo VII . . . . . p. 142
Anexo VIII . . . . . p. 143

Lista de Abreviaturas

F_0	Frequência Fundamental
Hz	Hertz
LDT	Limiar de Diferenciação Tonal
TM	Tom Médio
F	Finalização
st	semitons
ms	milissegundos
s	segundos
min	minutos
mm	milímetros
VD	Variável dependente
2AFC	*Two Alternative Forced Choice*
3MFC	*Three Multiple Forced Choice*

Lista de Figuras

1 Componentes para o sistema de decomposição e análise da entoação. Retirado de Consoni (2011) . . . . . p. 23

2 Análise usando os LDTs e o tom médio como parâmetros de análise. . . . . p. 29

3 Exemplo de análise se usando o tom médio para a oração: "O Domenico entrar aqui...como assim?". . . . . p. 33

4 Esquema de organização sintagmática no modelo de Pierrehumbert (1980) . . . . . p. 41

5 Desenho anatômico da laringe, vista em um corte coronal. Retirado de Gray (1918) . . . . . p. 45

6 Desenho anatômico da laringe, vista em um corte sagital. Retirado de Gray (1918) . . . . . p. 46

7 Desenho anatômico da glote, vista de cima em um corte transversal. Retirado de Gray (1918) . . . . . p. 48

8 Desenho anatômico da *rima glottidis* vista de cima, a partir da entrada da laringe. Retirado de Gray (1918) . . . . . p. 49

9 Vista lateral dos músculos da laringe. Retirado de Gray (1918). As setas indicam a posição dos músculos *tireoaritenoideo* e *cricotireoideo*, respectivamente . . . . . p. 50

10 Vista superior das etapas de abertura da glote, os triângulos negros representam as cartilagens aritenoides. Retirado da Internet . . . . . p. 52

11 Retirado de Titze (1988) . . . . . p. 53

12 Retirado de Titze (1988) . . . . . p. 54

13 Vista lateral das etapas de abertura da glote, baseado em Baer (1975) . p. 54

14 Cena externa do filme "Touro indomável" . . . . . p. 58

15 Cena *indoor* do filme "Touro indomável" . . . . . p. 58

16 Curva de pitch para a oração "O João mora na casa verde" . . . . . p. 72

17 Curva de pitch e espectrograma para a oração "O João mora na casa verde" . . . . . p. 73

18 Curva de média silábica, em st, para a oração "O João mora na casa verde", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010) . . . p. 74

19 Curva de tom médio, em st, para a oração "O João mora na casa verde", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010) . . . p. 75

20 Curva de pitch para a oração "O gato está naquele muro' . . . . . p. 75

21 Curva de pitch e espectrograma para a oração "O gato está naquele muro" . . . . . p. 76

22 Curva de média silábica, em st, para a oração "O gato está naquele muro", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010) . . . p. 77

23 Curva de tom médio, em st, para a oração "O gato está naquele muro", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010) . . . . . p. 77

24 Gráfico de relação entre Hz e semitons plotados em MIDI. . . . . p. 80

25 Exemplo de manipulação com mais 6st para a frase "O João mora na casa verde". . . . . p. 82

26 Interface usada para o experimento de percepção . . . . . p. 85

27 Texto inicial, como visto pelos participantes do teste. . . . . p. 90

28 Interface do teste com três alternativas. . . . . p. 91

29 Relação entre o valor de α e o tamanho n da amostra. . . . . p. 93

30 Número de repostas "não" obtidas no teste 2AFC. . . . . p. 99

31 Número de repostas "sim" obtidas no teste 2AFC para cada variável . p. 100

32 Intervalos de confiança para o grupo variável de variáveis. . . . . p. 103

33 Diagrama de Venn ilustrando as categorias de cada uma das variáveis p. 105

34 Número de repostas "não" e "talvez" obtidas no teste 3MFC. . . . . p. 110

35 Número de repostas "sim" e "talvez" obtidas no teste 3MFC. . . . . p. 111

36 Intervalos de confiança para as variáveis mais4, mais3 e mais2. As linhas tracejadas indicam a intersecção entre as barras de erros. . . . p. 118

37 Gráfico que demonstra o cruzamento das curvas de reconhecimento e de não-reconhecimento das diferenças entre os pares das manipulações ascendentes. . . . . p. 119

38 Gráfico que demonstra o cruzamento das curvas de reconhecimento e de não-reconhecimento das diferenças entre os pares das manipulações descendentes. . . . . p. 119

39 Esquema representando os limiares de diferenciação tonal para o Português Brasileiro, como relação a um tom natural . . . . . p. 122

Lista de Tabelas

1 Tabela dos valores de Hz médio para cada sílaba da frase "A Maria gosta de quindim", demonstrada na Figura 2 . . . . . p. 30

2 Tabela dos valores de pitch para a frase "O João mora na casa verde" p. 73

3 Tabela dos valores de F_0 para a frase "O gato está naquele muro" . . . p. 76

4 Tabela dos valores com os parâmetros utilizados para a manipulação. p. 82

5 Tabela contrastiva dos valores de pitch, após manipulação com o Praat p. 83

6 Tabela com os valores de tempo para cada etapa do teste . . . . . p. 91

7 Tabela com os valores de correlação entre as respostas para o teste com duas e três respostas . . . . . p. 98

8 Valores de variância para as 13 variáveis em análise . . . . . p. 101

9 Agrupamento de variáveis com base na variância . . . . . p. 101

10 Valores de significância do teste-t comparando-se o grupo variável ao fator fixo. . . . . p. 102

11 Tabela de correlação entre as variáveis do grupo com variância maior que zero . . . . . p. 107

12 Matriz rotacionada e cargas fatorais . . . . . p. 108

|

1 Introdução

1.1 Apresentação do tema

O estudo da prosódia vem ocupando uma posição cada vez mais privilegiada dentro dos estudos linguísticos. Uma pequena busca em um banco de dados popular, como o *Google Scholar*, revela-nos uma variedade de abordagens em torno do tema, principalmente nos últimos vinte anos. O domínio da entoação segue padrão parecido, embora as pesquisas mais antigas datem de cerca de trinta anos. Maciel e Rothe-Neves (2006) demonstram que no período de 1976 a 2006 houveram cerca de 89 estudos experimentais sobre a entoação do Português Brasileiro, sendo que vinte e sete deles foram publicados até o ano de 1997 e os demais em um período de oito anos, de 1998 a 2006. Tais números demonstram um aumento significativo da investigação sobre a entoação, apresentando um panorama propício para a discussão do tema.

Os primeiros estudos experimentais da entoação, como aqueles desenvolvidos por Pike (1945), dentre outros, esbarravam na falta de tecnologia de gravação e processamento de sinais. Estes trabalhos, em sua maioria, eram feitos com análises de oitiva e um sistema de anotação criado pelos próprios pesquisadores, não havendo um consenso geral sobre este mecanismo de anotação. Steele (1775) já sugeria, em meados do séc. XVIII, um sistema de anotação das variações de frequência fundamental, baseado na escala *cromático-diatônica* e no conceito de *slides*, o que para o autor, representava o movimento melódico da fala.

De fato, apenas em meados do século XX se veria um estudo experimental da questão da entoação com um aparato metodológico/tecnológico melhor refinado.

Maeda (1976) em sua tese de doutoramento utilizou-se de dados físicos, obtidos por métodos que não os de oitiva, para propor sua escala de cinco componente: a BL (*Baseline*), RP (*Rise Plateau)* R (*Rise*), L (*Lowering*) e P (*Peak*). Sendo estes um dos primeiros trabalhos a reunir o aspecto fonético da entoação ao aspecto descritivo, por meio de um método mais robusto de coleta de dados.

Os trabalhos desenvolvidos pelos pesquisadores do IPO (Institute for Perception Research) na Holanda seguiam uma abordagem semelhante a de Maeda (1976). No entanto, voltam-se para os princípios teóricos propostos por Bolinger (1961), os quais versavam sobre a possibilidade de se analisar a entoação em termos de configurações/padrões, ao invés de níveis, como feito na tradição americana. A abordagem dos holandeses pressupunham uma contraparte perceptiva, ou seja, um modelo baseado na interpretação do ouvinte e não na produção do falante, unicamente (T´HART; COLLIER; COHEN, 1990).

Neste trabalho usaremos destas duas abordagens: a metodologia de análise e coleta proposta por Maeda (1976) e o modelo baseado no ouvinte, proposto por t´Hart, Collier e Cohen (1990).

1.2 Objetivos

O objeto de pesquisa deste trabalho são os *limiares de diferenciação tonal* (LDT's) do Português Brasileiro. O limiar de diferenciação tonal pode ser definido como um valor z a partir do qual o individuo é capaz de notar qualquer diferença entre dois tons seguidos. t´Hart, Collier e Cohen (1990) apresentam como exemplo os estudos de Ritsma (1965), em que os ouvintes não eram capazes de identificar um tom isolado de 1KHz, porém eram capazes de diferenciar dois tons seguidos, tendo um 1KHz e outro 1,005 KHz. Desta maneira, a sensibilidade tonal dos ouvintes era da ordem de 0,5% para frequências acima de 1KHz. Os autores justificam a definição dos limiares de diferenciação tonal com base nas *just noticeable difference* (j.n.d.). A j.n.d. é uma medida psicofísica proporcional e construída sobre a lei de Weber: dado um determinado

estímulo y fixado, deve se variar seu par associado (y_1) até o ponto em que o sujeito seja capaz de notá-los como distintos. A divisão $\frac{y}{y_1}$ retorna um valor proporcional capaz de expressar o quão acurada é a percepção de um sujeito dado um conjunto de estímulos. Roederer (2002) demonstra que a j.n.d.'s possuem uma alta resolução, em torno dos 3% na faixa dos 500 Hz aos 1000 Hz, sendo que tal resolução diminui com o acréscimo de frequência tornando-se da ordem de 0,5% para frequências entre 1KHz e 2KHz.

t´Hart (1981) aplicou estas propostas aos estudos de fala a fim de determinar os limiares de diferenciação tonal no que se refere a entoação do holandês. Para o autor, o falante poderia produzir modalizações nas curvas entoacionais, como uma marcação de foco, controlando a glote a fim de que os valores de frequência fundamental ultrapassassem os limiares de diferenciação tonal. Em seu estudo sílabas de uma lista de palavras isoladas foram manipuladas dentro de um espectro de 1 a 6 semitons ascendentes e descendentes. Os testes indicaram que para um falante de holandês notar qualquer diferença tonal entre dois estímulos é necessário que se varie ao menos 3 semitons ascendentes ou 3 semitons descendentes.

Consoni (2011)aplicou os mesmos princípios experimentais de t'Hart (1981) para o português brasileiro. A autora aplicou dois testes distintos, porém complementares a fim de determinar os limiares de diferenciação tonal para o Português Brasileiro utilizando palavras isoladas. No primeiro teste foram utilizados1 0 pares de palavras trissílabas, todas tiveram suas frequências fundamentais variadas em 1 a 4 semitons ascendentes e 1 a 6 semitons descendentes. No segundo teste os pares foram agrupados em dois conjuntos: (i) cuja sílaba tônica havia sido manipulada; (ii) cuja palavra toda havia sido manipulada. Em sua análise geral Consoni (2011) chegou aos valores de 3 semitons ascendentes e 4 semitons descendentes.

Em ambos os experimentos as janelas temporais de análise foram da ordem de uma palavra trissílaba ou maior, por volta dos 60-100 milissegundos. Nossa proposta é a de se avaliar os mesmos parâmetros, contudo tendo como parâmetro de análise a frase.

Nosso objetivo é verificar se os limiares de diferenciação tonal sofrem alguma influência direta da duração. Nossa hipótese se sustenta no efeito tau, segundo os experimentos de Shigeno (1986) a distância temporal entre dois estímulos afeta a percepção tonal: quanto menor o tempo entre dois tons, menor a frequência do segundo tom é percebida pelos sujeitos. A fim de verificar estas questões aplicamos um teste de percepção, seguindo os princípios de t'Hart (1981) e Consoni (2011).

1.3 Visão Geral

O capítulo 2 apresenta um pequeno modelo matemático por nós desenvolvido ao longo desta pesquisa. Nosso objetivo é formalizar matematicamente o modelo de análise por decomposição proposto por Ferreira-Netto (2006, 2008) e a pesquisa posterior de Consoni (2011). Na primeira parte do capítulo tentamos introduzir a noção de sistemas por meio de uma análise crítica das principais correntes de estudos linguísticos, bem como elas se inserem no campo das ideias.

No capítulo 3 trataremos dos aspectos fonético-fonológico e perceptuais da entoação. A primeira parte tratará de maneira resumida das teorias que estudaram o caráter fonológico da entoação. Esta parte terá por objetivo demonstrar as diversas teorias e propostas metodológicas que trataram do tema ao longo do século XX. A segunda parte lidará com as características de produção, tais como, a anatomia da glote e manifestações acústicas de F_0 , ao final faremos uma breve apreciação sobre a percepção e os modelo por nós escolhidos

O capítulo 4 apresentará a metodologia por detrás do desenvolvimento e aplicação do testes de percepção aplicado com o intuito de analisar os limiares de diferenciação tonal do Português Brasileiro. O capítulo 5 traz a análise dos dados obtidos e uma breve discussão sobre alguns resultados obtidos, ao passo que o capítulo 6 traz as conclusões de nosso trabalho.

2 Um modelo para a entoação

Deus não faz matemática barata...

Sir Michael F. Atiyah

No capítulo a seguir apresentaremos a forma matemática, por nós desenvolvida, do modelo de Ferreira-Netto (2006, 2008). Nas duas primeiras seções faremos breves comentários sobre a teoria dos sistemas e suas relações com a linguística. Nossa hipótese epistemológica é a de que uma visão sistêmica dos objetos pode nos ajudar a relacionar a produção da fala ao seu uso. Modelos como a Fonologia Articulatória (FAR) (BROWMAN; GOLDSTEIN, 1989, 1992) têm se norteado por esta perspectiva de análise, tomando a produção como um primeiro passo para a compreensão do fenômeno cognitivo.

2.1 Sistemas

2.1.1 Sistemas e Linguística

Skyttner (2001) argumenta em favor de uma visão sistêmica dos objetos de estudo. Para o autor, a história da ciência é contada em termos de (i) verdades absolutas e (ii) oposições tidas como categóricas. Os primeiros escolásticos apoiaram seus estudos na dicotomia *natura-spiritus*, em que as organizações naturais seriam baseadas na vontade de Deus e reflexos de Seu poder. Assumia-se que este poder absoluto e divino seria inteligível e somente por meio de um pensamento racional poder-se-ia tangenciá-lo e passar a compreendê-lo (MARRONE, 2003), pois como afirma Tomás de Aquino em sua

Summa Theologica "Para a salvação do homem, é necessária uma doutrina conforme à revelação divina, além das filosóficas, pesquisadas pela razão humana" (AQUINO, 2007, p. 1).

O Renascimento rompeu com esta forma de análise levando a ciência para um nível primário de explicação e outros tantos níveis dedutíveis. Este nível primário é a raiz do *determinismo*, em que os menores elementos de diversos objetos são elencados (determinados) e, posteriormente, pelo processo de dedução são rearranjados e capazes de explicar as propriedades de qualquer objeto construído com base nestes menores elementos (SKYTTNER, 2001). Podemos tomar como exemplo as obras de Johannes Kepler e Isaac Newton (NEWTON, 2010), uma vez que as descobertas e estudos de ambos foram construídos com base em axiomas lógicos, tomados como leis universais, como a lei da gravitação universal ou as três Leis de Kepler. Mais tarde, em inícios do séc. XX, Kurt Gödel argumentaria em sua Teoria da Incompletude (GöDEL, 1930; KENNEDY, 2011) que nenhum sistema axiomático e universal seria consistente, uma vez que sempre há em uma teoria *consistente* proposições verdadeiras que não podem ser demonstradas nem negadas (KLEENE, 2002).

Em linguística podemos observar as duas doutrinas bem demarcadas. A escolástica está presente em estudos como o de Dante Alighieri em *De vulgari eloquentia* (EWERTY, 1940). Nesta obra o escritor italiano discute o uso do vernáculo *versus* o Latim, suas anotações partem da observação de que as línguas mudam e que há uma diversidade delas. A constatação, embora óbvia, parte do princípio de que houve uma língua divina, aquela com que Deus falou a Adão, sendo esta a raiz de todas as demais antes da *confusio linguarum* de Babel (ECO, 2001; VIVAI, 2009). Ligado aos estudos mecanicistas e deterministas podemos citar linguistas como os irmãos Grimm, Wilhelm von Humboldt e os neogramáticos, como William D. Whitney. A "Lei de Grimm" é um exemplo de lógica epistêmica do período, posto que toma o fonema como unidade mínima e inequívoca para a análise de fenômenos linguísticos universais.

A ciência do fim do século XIX e do início do século XX, fundada nos princípios

quânticos, trouxe a questão da relatividade à tona e, por consequência, a relação entre oposições tidas como categóricas. Esta nova visão, impulsionada pelas descobertas de Albert Einstein, contrapôs diversos objetos considerados até então não relacionados. Um exemplo clássico refere-se a proposição do *continuum* tempo-espaço. De acordo com o físico o tempo poderia passar em velocidades diferentes para dois referenciais distintos (um acelerado e outro estacionário, por exemplo), vindo daí o nome "Teoria da Relatividade". O método deste período tende a tornar os estudos verticais, partindo de detalhes particulares até especificidades de ordem especulativas, usando como princípio de organização os sistemas, isto é, organizações que são compostas por outros diversos elementos (SKYTTNER, 2001, p. 32). Portas para um mundo microscópio e ultra-especializado.

A teoria *saussureana* (SAUSSURE, 1977) segue princípios parecidos e se baseia em alguns axiomas lógicos não comprovados e relativizados, como a distinção entre língua e fala, uma vez que o objeto toma feições distintas de acordo com o referencial usado. O mesmo parece valer para as demais dicotomias de sua tese, nas quais suas definições se dão pelo outro e não pelo objeto em si. Podemos, entretanto, citar a questão da arbitrariedade do signo como um eco do determinismo, cuja definição não pode ser provada, apenas especulada. As teorias gerativistas foram além no quesito da relativização ao criar noções como Língua Interna (L_i)e Língua Externa (L_e), demarcando uma relativização entre os objetos materiais e os objetos mentais (CHOMSKY, 1986, p. 21ss.). Em Chomsky e Halle (1968) temos que a realidade material, por eles denominada de *formativos*, está diretamente ligada a uma forma abstrata subjacente (CHOMSKY; HALLE, 1968, p.7), sendo estas formas subjacentes o objeto da Linguística e as demais formas de superfície objetos de outras áreas [1].

A relatividade destas teorias entre objetos particulares e objetos mentais é atuali-

[1]Chomsky (2009) debate a natureza do trabalho de um cientista ou pensador. Sua proposta, em linhas gerais, é a de que um cientista deve propor explicações para o seu objeto de estudos e não ser um catalogador de fatos. Em outro trabalho (CHOMSKY, 2000) o autor demonstra que seu objetivo ao propor o Programa Minimalista e os Princípios e Parâmetros foi o de romper com a profunda separação que se deu dentro da teoria gerativista, no que se refere a "adequação descritiva e adequação explicativa" (CHOMSKY, 2000, p. 8). Sendo que as formas de uso deveriam ser parte do programa de pesquisa, como objeto de descrição.

zada em oposições *quasi-categóricas*, como *fala* e *língua* ou em *competência* e *performance*. Esta atualização tende a eliminar os objetos complexos ou tidos como individuais, criando um sistema de regras e conceitos sistêmicos e universal. Em verdade, o que se tem, de fato, é uma coleção de explicações sintéticas e organizadas com base em uma determinada teoria, sem que ocorra uma "re-montagem" do objeto analisado, a não ser na coleção das explicações. A grande questão que se coloca está relacionada ao método destas vertentes epistemológicas, o qual leva a uma síntese do método em si, não chegando ao ponto em que o objeto é analisado em contraposição a outros fenômenos a ele ligado.

Skyttner (2001) aponta que apenas uma visão sistêmica é capaz de levar as teorias e métodos para este ponto: onde a síntese é a base da explicação e não o ponto final. Nesta abordagem, diversas disciplinas trabalhariam para a explicação de um fenômeno complexo, sendo que a síntese de cada uma delas seria o ponto de partida para uma explicação conjunta. Obviamente, esta proposição demanda um conjunto de especialidades mínimas, as quais estão diretamente ligadas a história da ciência da qual fazem parte e se tornam, por definição, elementos de uma *metaciência*. Esta metaciência se volta para aqueles objetos platônicos, cuja realidade abstrata está para além de sua realização em si, demandando uma nova metalinguagem. Uma das propostas para este problema foi dado por Bertalanffy (1975) em sua "Teoria Geral dos sistemas". Tal teoria seria, em última análise, esta metaciência capaz de reagrupar diversos objetos sintéticos em torno de um objetivo único, a saber: o de resolver questões entremeadas à diversas áreas do conhecimentos.

A Teoria Geral dos Sistemas (BERTALANFFY, 1975; MACIEL, 1974) foi influenciada de maneira mais direta pelo modelo transformacional de Turing (TURING, 1997) 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; SHANNON; WEAVER, 1949) e no campo da computação com o *The First Draft Report on the EDVAC* de John von Neumann (NEUMANN, 1993).

Noam Chomsky é citado em Maciel (1974, p. 14) como um dos nomes mais influentes da Teoria do Sistemas pelos trabalhos feitos na área da Teoria dos Autômatos e da Gramática Formal (CHOMSKY, 1956, 1959). De fato, as *Hierarquias de Chomsky* tem um grande impacto nas ciências formais, como as Ciências da Computação, Cibernética e Lógica, pois operam níveis de geração de linguagens formais que se complexificam de acordo com o grau de liberdade de cada um destes níveis.

Sound Patterns of English (CHOMSKY; HALLE, 1968) possui ecos de tais hierarquias. Em nosso entendimento o modelo fonológico ali apresentado segue o modelo de linguagens formais, sendo estabelecido com base em regras. Em verdade, as hierarquias de Chomsky, se contrastadas aos desenvolvimentos do programa gerativista, possuem uma relação inequívoca com esta, partindo dos níveis mais restritos, o fonológico, ao menos restritos, a composição com frases. As regras de composição fonológica são, em geral, restritas e dependentes do contexto. O termo "Restrição fonológica" atribuído a processos fonológicos intrínsecos a língua nos permite fazer tal inferência. Tomemos como exemplo o Português Brasileiro onde teríamos uma restrição à colocação da consoante fricativa labiodental em posição de coda silábica. Outro caso seria a restrição de qualquer fonema \n\ em posição de núcleo no PB, enquanto em Inglês podemos ver esta ocorrência, de acordo com o contexto. No domínio da frase se permite a construção de uma frase como "Colorless green ideas sleep furiously" (CHOMSKY, 2002, p. 15), embora haja restrições de cunho cognitivo, as chamadas sentenças *agramaticais*.

Este modelo de restrições e combinações lógicas foi seguido pelos estudiosos da teoria. Haja vista os trabalhos de Goldsmith (1976), Liberman e Prince (1977), Pierrehumbert (1980), Hayes (1981), dentre outros. 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 e a posição do foco semântico, de modo que se torna dificil conceber restrições neste nível prosódico.

O que temos de fato nestas abordagens é uma série de análise de restrições mecanicistas de produção, as quais não explicam, de fato, como o a entoação é "utilizada" pelos falantes de uma língua. Nossa proposta é a de analisar a entoação pelo prisma da teoria dos sistemas por entendermos que nesta visão podemos conjugar o uso aos aspectos fonológicos. Nossa intenção é ultrapassar o modelo de restrições corrente, propondo uma forma de análise por decomposição baseada em Ferreira-Netto (2008). Nos últimos 20 anos a Fonologia Articulatória tem se debruçado na Teoria dos Sistemas, em especial na Teoria dos Sistemas Dinâmicos (BROWMAN; GOLDSTEIN, 1989, 1992; STUDDERT-KENNEDY; GOLDSTEIN, 2003). Nos domínios da entoação, o autor que tem se aproximado destas ideias é Yi Xu. Xu (2005) propõe o uso do termo *speech melody* para os estudos que envolvem o tom e/ou a entoação. Nesta abordagem teríamos um estudo unificado em torno da contraparte fonética (por ele denominada "funcional"), a frequência fundamental, sendo possível analisá-la por meio da sua ligação com parte fonológica. No capítulo 3 faremos melhores considerações sobre estes problemas aqui esboçados.

Por fim, podemos dizer que nos estudos Linguísticos a noção de sistema foi bem aceita sem uma definição clara. Até onde pudemos averiguar Saussure (1977) usa o termo *sistema* no sentido de uma "organização" dinâmica de elementos estáticos. Neste caso, o alfabeto poderia ser visto como tais elementos estáticos, enquanto o processo de escrita é o sistema em si. O mesmo vale para a fonologia da língua em que os elementos deste sistema seriam os sons da fala (em uma visão de sons estáticos). Dentro dos domínios da teoria gerativa a ideia de sistema parece ser a mesma, embora não haja uma exemplificação explícita. A formulação da ideia de traços e regras são a base para a composição de um sistema. A diferença, embora tênue, sugere que para os estruturalistas a língua é composta de diversos sistemas interdenpendentes, ou seja, um sistema necessita do outro, embora isso não seja uma conição de . Ao passo que 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, como em um computador. A questão dos sistemas dentro destas teorias, embora pertinente, teria de ser melhor analisada, o que foge do escopo de nossa pesquisa.

A seguir, apresentaremos um pequeno modelo matemático por nós desenvolvidos, o qual se insere na lógica da teoria dos sistemas. Entendemos que a entoação pode ser analisada em termos de um sistema conceitual e aberto. Um sistema conceitual, pois é construído por meio de conceitos e hipóteses, as quais podem ser verificadas ou negadas, caso se julgue necessário e, além disso, trata-se de um sistema aberto, pois é auto-adaptável e se regula de acordo com o meio em que está inserido.

2.1.2 Sistemas, intervalos e a proposta de Brooks (1997)

Antes de apresentarmos o modelo em si, destacamos que entendemos um sistema como "um conjunto de elementos quaisquer ligados entre si por cadeias de relações de modo a constituir um todo organizado" (MACIEL, 1974, p. 13). Ashby (1957) levava em consideração a noção de *mudança* para compor um sistema 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)

Entendemos, pois, que a noção de mudança dentro de um sistema responde à natureza evolutiva dele próprio. Deste modo, supomos que além de ser um todo organizado em uma série de relações internas ele possuirá estados ($S_1, S_2, S_3, ..., S_n$), de acordo com sua evolução temporal. No caso desta pesquisa entendemos que haverá um estado neutro (S_1) e outros tantos estados quanto forem os valores de F_0 que estejam para além dos *limiares de diferenciação tonal* (T´HART; COLLIER; COHEN, 1990) e possuam algum significado (seja pragmático, seja sintático) para o ouvinte.

Nossa proposta, neste sentido, é um sistema dentro de um outro sistema, sendo o sistema principal, o sistema de produção da fala. O sistema de produção dos sons da fala pode ser compreendido por modelos como os de Stevens (2000) e Fant (1966), bem como por abordagens mais modernas como as de Browman e Goldstein (1992). Em cada um dos modelos temos componentes básicas [2] a partir das quais o sistema pode evoluir no tempo, gerando uma série de elementos, como uma vogal ou uma sílaba ou todo um discurso. Este sistema seria descontínuo pois a sua percepção seria ajustada de acordo com os segmentos produzidos. Em nosso entender o sistema da entoação vem associado a este produzindo as curvas melódicas da fala. As componentes da entoação, em nosso modelo, seriam: (i) o Tom médio; (ii) a finalização; (iii) os limiares de diferenciação tonal e por fim (iv) a variação da frequência fundamental (FERREIRA-NETTO, 2008). Além disso, a entoação seria um sistema contínuo, uma vez que sua percepção não é ajustada em pulsos, mas sim ao longo de sua produção.

Os elementos acima elencados podem ser definidos da seguinte maneira. O (i) tom médio (TM) é um valor de referência contínuo e variável, que se define no instante t_n pela média acumulada de $t_0 \rightarrow t_n$. Em suma, o tom médio é uma média acumulada no tempo. De um ponto-de-vista linguístico e cognitivo o tom médio é definido como um valor tonal de referência, o qual o falante estabelece no início de sua fonação e tende a

[2] A pressão, o fluxo e os harmônicos no caso dos primeiros e o gesto articulatório no caso da Fonologia articulatória, de maneira simplista

mantê-lo ao longo de um ato de fala. Ao longo de uma fonação haverá declinações e acentuações pontuais da frequência fundamental, por esta razão Ferreira-Netto (2008) nomeia de *sustentação* o esforço feito pelo falante para retornar e manter-se próximo a esse valor de referência do tom médio, a despeito destas declinações e acentuações. O autor nomeia de ritmo tonal, a consequência deste esforço, uma vez que as forças que o compõem operam em sentidos opostos.A declinação decorrente da suspensão deste esforço representa a (ii) *finalização* e, pragmaticamente, tem a função de marcar o fecho de um ato de fala.

Os limiares de diferenciação tonal, objeto de estudo desta dissertação, possuem a função linguística de marcar elementos linguísticos dentro de um ato de fala, além disso, são em número de dois: um superior e outro inferior e tem como referência o tom médio. Consoni (2011) propõe a existência de duas proeminências para o Português Brasileiro, uma para além do limiar superior e outra entre o limiar inferior e a finalização. Nossa suposição é a de que as proeminências para além do limiar superior, ou seja, aquelas que são positivas, tem um caráter de focalização do item, enquanto aquelas entre o limiar inferior e a finalização aparentam ser correções de uma palavra ou termo usado pelo falante.

As acentuações e declinações em si referem-se a (iv) variação da frequência fundamental, a qual é o valor original de F_0 sem nenhum tratamento matemático. Deste modo, como propõe Ferreira-Netto (2008), podemos dizer que as componentes *tom médio, limiares de diferenciação tonal* são componentes estruturadoras, uma vez que organizam as variações de F_0. As componentes *finalização e proeminências superior e inferior* são componentes semântico-funcionais, pois comunicam informações aos ouvintes. No caso da primeira, o fecho do ato de fala e, a segunda, uma marcação pragmática/sintática. A figura 1, retirada de Consoni (2011), ilustra os componentes de nosso modelo, bem como nos traz uma representação visual de como poderiam ser as curvas com marcações tonais.

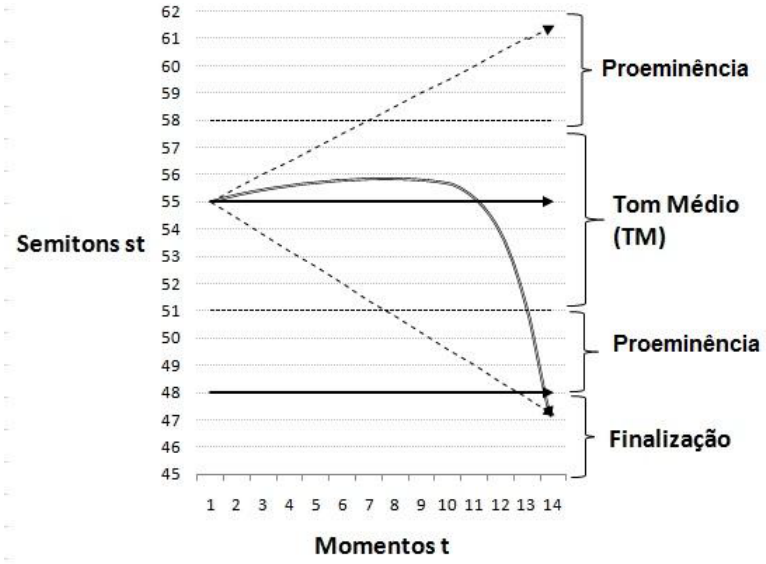

Figura 1: Componentes para o sistema de decomposição e análise da entoação. Retirado de Consoni (2011)

Na figura 1,as linhas escuras e contínuas representam o tom médio e a finalização, respectivamente. As linhas tracejadas mais finas representam os limiares de diferenciação tonal, enquanto as linhas tracejadas com setas nas pontas representam as possíveis direções de uma curva entoacional. Os valores no eixo lateral (y) estão expressos em MIDI e não em Hertz, como é usual. Consoni (2011) assume que os limiares de diferenciação tonal e o tom médio formam uma faixa de banda tonal de tom médio, sendo este intervalo todo considerado o tom médio. Enquanto o intervalor entre o limiar de diferenciação tonal inferior e a finalização seria a proeminência inferior.

Por esta razão e dado o escopo de nossa pesquisa nos apropriaremos também da noção matemática de *intervalos*. Em termos leigos, um intervalo pode ser definido como o conjunto de todos os números entre dois extremos diferentes entre si, incluindo ou não os próprios extremos (GUIDORIZZI, 2008; TáBOAS, 2008). Por exemplo, o intervalo de dias em um mês é definido como $[1, 30]$, ou ainda como $[1, 31]$ ou $[1, 28]$ onde a extremidade a esquerda representa seu início, ao passo que a extremidade oposta, representa seu término. É importante destacar que a noção de *intervalo* é ligeiramente diferente de *conjunto*, uma vez que o segundo é uma coleção de elementos, sequenciais

ou não, enquanto que no primeiro se pretende certa continuidade[3].

Estendendo-se a definição matemática podemos, por implicação, derivar uma série de aplicações para os intervalos, tais como, a organização sistemática e previsível de algum evento, o qual em última instância pode ser visto como um conjunto de elementos dentro de um dado intervalo. Um exemplo natural capaz de ilustrar bem tal conceito é a chance de se engravidar, que se dá em um dado intervalo de dias em um mês, bem como em um intervalo de meses, dentro de anos. Em termos matemáticos, o intervalo é o espaço limitado de possível ocorrência deste evento, dada uma função previamente estabelecida ou a se estabelecer.

A partir do princípio dos intervalos pode se definir, por exemplo, as chances de que ocorra uma fecundação e, por consequência, uma gravidez. Em uma primeira análise ela terá de ocorrer em um intervalo de 7 dias, aproximadamente, em um mês, ou seja sua probabilidade será representada por 1/7 (em uma fecundação para sete tentativas). Em um ano temos que serão intervalos de 7 dias dentro do intervalo de 12 meses, assim a probabilidade será de 1/7 * 1/12 = 1/84 (uma fecundação para oitenta e quatro tentativas). Por questões biológicas temos de lembrar que serão intervalos de 12 meses, em um conjunto de aproximadamente de 25 anos, desta maneira teremos, 1/84 * 1/25 = 1/2100 (uma fecundação para duas mil e cem tentativas). A probabilidade da fecundação (o processo de união dos gametas) é de 1/2 (ou se fecunda ou não se fecunda), estendendo nossa conta temos 1/2100 * 1/2 = 1/4200, algo em torno de 0,00023. Nestes termos, a probabilidade de que um bebê seja gerado é particularmente baixa e, no entanto, vemos crianças nascendo a todo momento. A evolução contornou de algum modo este problema, fazendo com que um dos pares tenham uma maior quantidade de gametas, por exemplo.

A partir do conceito de intervalo mais duas aplicações podem ser propostas: (i) a primeira refere-se à possibilidade de se prever a ocorrência de fenômenos, com base em um intervalo pré-definido, como acima feito com a gravidez; (ii) a segunda

[3]Tende-se a assumir que os intervalos são subconjuntos de conjuntos maiores, como o conjunto dos números naturais ou reais (GUIDORIZZI, 2008)

refere-se à possibilidade de se analisar a recorrência de um fenômeno dentro deste intervalo, caracterizando-se determinados grupos de ocorrência, como "gravidez na adolescência" ou "gravidez de risco", etc. Nossa intenção, neste aspecto não é discutir as propriedades ou a natureza dos intervalos em matemática ou estatística, mas antes, trazer à tona o conceito, uma vez que ele terá implicações ao longo de todo o nosso trabalho. Deste modo, nosso objetivo com estas observações é a de colocar em foco a análise relacional, baseando-a em termos de intervalo numéricos, o mecanismo central de nossa hipótese

Somem-se assim os dois conceitos: a ideia de controle e a noção de intervalos e poderemos pensar em um *sistema*. A engenharia e a administração são exemplos de como podemos vislumbrar essa ligação. Ao pensarmos em sistemas, temos necessariamente de pensar em *componentes*. Um jogo de futebol, como o que ilustramos, é um exemplo de sistema. A administração de uma universidade e suas divisões em pró-reitorias e comissões é também outro exemplo. Um motor de carro é o exemplo extremo de um sistema, que podemos aqui chamar, de concatenado. Em um sistema deste tipo, uma série de componentes concorrem para o funcionamento ótimo. Uma falha, ou um erro de projeto e temos um funcionamento enviesado ou um não-funcionamento.

Em engenharia aplicada a Inteligência Artificial existem dois modos de se encarar um sistema, como pontua Brooks (1997). A primeira opção seria por meio de uma (i) *decomposição por função*, ao passo que a segunda seria uma (ii) *decomposição por atividade* (BROOKS, 1997, p. 402) . A ideia por trás dessa proposta é a de se analisar por engenharia reversa um sistema complexo, podendo ao final remontá-lo. No caso de Brooks a aplicação refere-se à construção de um sistema inteligente capaz de se mover de maneira autônoma em um ambiente "real", isto é, um ambiente com obstáculos e dificuldades, como os enfrentados por um ser humano.

Assim, a análise por função teria como proposta a decomposição elementar verificando-se passo-a-passo o papel de cada módulo dentro do sistema global. Nesta perspectiva um sistema qualquer seria definido por um conjunto de módulos separa-

dos, unidos por um centro de processamento. Desta maneira, qualquer *input* em um dos módulos seria, necessariamente, processado e enviado na forma de um sinal (químico, elétrico, mecânico, linguístico, simbólico, etc.) para o centro de processamento, onde receberia um tratamento e um *feedback* em forma de um *output* interrompendo o sistema ou ativando outros módulos a fim de se continuar o processo global. 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. As hierarquias de Chomsky (CHOMSKY, 1959) tem espaço neste modelo, bem como os primeiros desdobramentos da teoria gerativa

Para a outra visão de decomposição, a decomposição por análise, vale o mesmo princípio das funções, porém, a diferença está no fato de que em vez de módulos, teríamos *camadas* de processamento. Tais camadas seriam relativamente autônomas, isto é, não haveria entre elas um módulo mediador ou centro de processamento, pelo contrário, as camadas seriam responsáveis por si só por gerar *atividades*, as quais se tornariam *habilidades* dado o uso contínuo e proposital de algum periférico. Importante destacar que o processamento de cada camada seria autônomo do todo, de forma que o *input* perceptivo não se distinguiria do *output* de ação ou vice-versa. Assim, o centro de processamento se perderia, não havendo a necessidade de um mecanismo de transmissão da informação entre camadas. Neste caso, o que ocorreria seria um processamento interno, o qual poderia, em certos termos, ter algum tipo de sistema de comunicação. A comunicação entre camadas seria feita por meio de um sistema de desligamento e acionamento parcial ou total de determinadas camadas. A diferença entre as abordagens se dá justamente neste ponto. Ao invés de se propor ou imaginar sistema de combinações com um nível de representação, podemos elaborar um sistema de elementos relativamente autônomos com uma base plural, ou melhor dito, com uma base multi-modal [4].

[4] A noção de "multi-modal" está associada ao conceito de *modo* e, não, de *modularidade*. Para Spinoza (SPINOZA, 2004) a noção de *modo* está relacionado as propriedades de um determinado elemento e esta parece ser a noção mais bem aceita até os dias de hoje. Em contrapartida o conceito de módulo está

Neste trabalho, apropriar-nos-emos da ideia de que a entoação é, em síntese, parte de um sistema complexo, no qual uma série de camadas/elementos concorrem para a produção final dos sons, embora ela venha ao longo dos tempos sendo compreendida como uma coleção de categorias e inter-relações entre elementos passíveis de descrição. As propostas teórico-metodológicas de Pierrehumbert (1980) vão justamente nesta direção: a catalogação de componentes entoacionais a partir da distinção entre tons altos e baixos (High-Low), como acima discutido. O método de análise ToBI (SILVERMAN et al., 1992) tem justamente este princípio de funcionamento analítico. Embora, seja um método válido e muito poderoso de descrição nossa interpretação é a de que ele apenas consegue relativizar tons em uma sequência temporal, sem que haja, necessariamente, uma análise complexa das variações de F_0. Isto é, ao se categorizar tons em categorias bem distintas, perde-se o entremeio, as variações mediais se tornam apenas complementares e sem necessidade de serem analisadas.

Nossa proposta é a de se analisar a entoação como a componente de um sistema complexo e não-estático, no caso, a fala. Nesta abordagem compreenderemos a entoação como uma *camada*[5] que concorre em paralelo com algumas outras camadas. Destacamos que esta camada tem, em nosso entender, certa autonomia com relação a outras componentes da fala, sendo que algumas *atividades* intrínsecas a ela são acionadas em conjunto com o início da fonação. O controle destas atividades é feito com base na percepção do que é ouvido e tem alguns correspondentes que podem ser capturados dentro da análise dos sinais e da percepção da fala.

2.1.3 Um esboço de modelo para a entoação

Seguindo a linha de raciocínio apresentada no parágrafo anterior, que trata das análises por função e por decomposição,podemos definir a entoação como um fenômeno linguístico dotado de um mecanismo complexo, o qual envolve tanto a percepção dos sons, quanto o processamento da produção. A despeito da fonética dos segmentos, a

ligada a ideia de objetos construídos com a finalidade de serem parte de um sistema maior.

[5]Os termos aqui utilizados são encontrados em Brooks (1997).

entoação não possui referências/categorias internalizadas capazes de fornecer uma resposta instantânea para um determinado estímulo sonoro. Pelo contrário, a entoação pode ser entendida como uma construção não-linear, embora previsível dentro do decurso de um ato de fala. Além disso, a entoação pode ser compreendida como um efeito secundário, isto é, um subproduto resultante da produção da fala, embora seja possível, a nosso ver, controlá-las fisiologicamente, como resposta linguística, gerando frequências moduladas (XU; WANG, 1997, 2001; XU, 2005). Estas modulações de frequência, em última análise seriam a "tessitura entoacional", isto é, a região de variação da frequência fundamental (F_0).

A entoação, portanto, 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...z_{n-1},z_n, com $n-1>0$, para $z_n = F_0(t)$),tomadas nos instantes t (t_1,t_2...t_{n-1},t_n, onde $n-1>0$).

A equação (2.1) é a forma matemática do que foi acima descrito e retorna o 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:

$$\theta = z(t) = [z_1(t_1), z_2(t_2), ..., z_n(t_n)] \tag{2.1}$$

Equação (2.1). Entoação como série temporal.

Em termos matemáticos, a equação (2.1) demonstra que o conjunto das observações z estaria contido no eixo y de um plano cartesiano, ao passo que os instantes t estariam contidos no eixo x. Entender as curvas de F_0 como série temporal permite-nos um melhor refinamento analítico, uma vez que podemos eliminar de nossa análise a acelaração angular ω e passar a operar com um coeficiente angular α, de modo a se tornar possível a comparação entre alturas em função do tempo. Neste caso não se analisariam ondas, mas sim funções lineares, onde os instantes seriam intrinsecamente

dependentes, isto é, o instante z_2 estaria ligado e dependeria do instante z_1.

A proposta que se segue deste modelo é a de que haveria intervalos laterais para as alturas, definidos com base em propriedades da percepção humana. t´Hart, Collier e Cohen (1990) nomeiam tais intervalos laterais de *limiares de diferenciação tonal*[6]. t´Hart (1981), com base em modelos experimentais, estipulou que estes valores seriam de 3 semitons para o holandês, isto é, para que se note qualquer diferença tonal no curso de uma fonação no holandês é necessário que o F_0 varie ao menos 3 semitons para cima ou para baixo. Consoni et al. (2009), Consoni (2011), replicando os experimentos de t'Hart, determinaram os valores de $+3$st e -4st para o Português Brasileiro, estabelecendo, assim, uma faixa de banda de 7st,como demonstrado na figura 2. Para se estipular esse limites laterais Ferreira-Netto (2006, 2008) propõe uma componente cognitivo-funcional chamada *tom médio* (τ), melhor definido na subseção 2.1.3.1. Na Figura 2 pode se ver os limiares de diferenciação tonal inferior e superior (as linhas sólidas) gerados a partir de um tom médio (a linha tracejada) obtido na análise da oração "A Maria gosta de quindim".

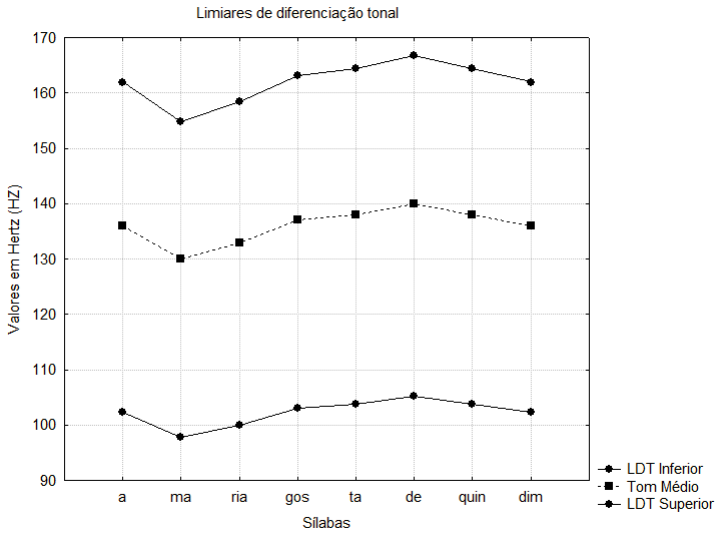

Figura 2: Análise usando os LDTs e o tom médio como parâmetros de análise.

Os valores de frequência, apresentados na figura 2, em cada sílaba foram estabelecidos pela soma das frequências válidas (maiores que 50Hz e menores que 350Hz)

[6]Aqui, por vezes, referiremo-nos a estes limiares pela sigla LDT

LDT inferior	Tom Médio	LDT superior
102,3	136	162,0
97,8	130	154,8
100,0	133	158,4
103,0	137	163,2
103,8	138	164,4
105,3	140	166,7
103,8	138	164,4
102,3	136	162,0

Tabela 1: Tabela dos valores de Hz médio para cada sílaba da frase "A Maria gosta de quindim", demonstrada na Figura 2

encontradas a cada 25ms dividido pela duração total da sílaba dividida por 25ms, obtendo-se assim uma média, a equação (2.2) representa como foi feito o cálculo

$$F_m = \frac{\sum_{i=1}^{n} F_0}{\Delta_t / 0,025} \tag{2.2}$$

Equação (2.2) Equação para o cálculo da frequência média em uma sílaba.

Onde, temos:

F_m = Frequência média.

$\Delta_t / 0,025$ = Duração total da sílaba (Δ_t), divida por *frames* de 0,025.

$\sum i = 1^n F_0$ = Soma de todos os valores de F_0 maiores que zero

2.1.3.1 O Tom Médio

De acordo com o modelo de Ferreira-Netto (2008), durante o processo inicial de fonação, o falante estabeleceria um valor de referência de modo que, durante o curso de fala, ele teria a tendência a se esforçar para manter os valores de frequência em torno desta referência. Desta maneira, o tom médio seria uma *baseline* de referência para se estipular o intervalo lateral de a para b $[a \rightarrow b]$[7] contido no eixo y da série temporal θ, como representado na Figura 2, pelas linhas sólidas. A equação (2.4) define

[7]Em termos matemáticos a e b representam os limiares de diferenciação tonal.

a forma matemática do tom médio, ao passo que o sistema (2.5) define como seriam estabelecidos os valores de a e b. Importante ressaltar que a equação (2.4) é uma média aritimética 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:

$$z(t) = [z_1(t_1), z_2(t_2), ..., z_n(t_n)] \tag{2.3}$$

Equação (2.3). Análise da entoação como série temporal.

Temos o tom médio (τ) definido pelas médias da série temporal θ:

$$\tau = \sum_{i=1}^{n} z_n(t_n) \tag{2.4}$$

Equação (2.4). Equação que define o tom médio.

A partir da qual podemos pressupor o seguinte sistema, onde a e b são os extremos do intervalo de frequência, os LDTs:

$$p/[a,b] = \begin{cases} a = \tau * 1,19 \\ b = \tau - (\tau * 0,22) \end{cases} \tag{2.5}$$

Equação (2.5). Sistema que estabelece os limiares de diferenciação tonal.(T´HART; COLLIER; COHEN, 1990)

Os valores de a e b (os LDT's) 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.

O modelo de Ferreira-Netto (2006, 2008) supõe uma terceira componente, além do intervalo lateral e do tom médio, a saber: a finalização. De acordo com Peres, Consoni e Ferreira-Netto (2011) 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 necessário a inclusão de um terceiro elemento no sistema (2.5), como demonstrado abaixo.

$$p/[a \to b, c] = \begin{cases} a = \tau * 1,19 & \text{estabelece o LDT superior.} \\ b = \tau - (\tau * 0,22) & \text{estabelece o LDT inferior.} \\ c = \tau - (\tau * 0,35) & \text{estabelece o valor de finalização.} \end{cases} \tag{2.6}$$

Equação (2.6). Sistema que demonstra o processamento de percepção da entoação.

O valor c de 0,35 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, com vistas a futura computação e representação gráfica, mais dois elementos, definindo o sistema $\epsilon(t)$[8], para todo $F_0 > 0$:

[8]Usaremos a letra grega ϵ para indicar "Entoação"

$$= \epsilon(t) \begin{cases} a = \tau * 1,19 & \text{estabelece o LDT superior..} \\ F_0 = z(t) & \text{variação da frequência fundamental} \\ \tau(t) & \text{tom médio} \\ b = \tau - (\tau * 0,22) & \text{estabelece LDT inferior.} \\ c = \tau - (\tau * 0,35) & \text{estabelece o valor de finalização.} \end{cases} \tag{2.7}$$

Equação (2.7). Sistema que demonstra o processamento global da entoação

A figura 3 [9] apresenta uma análise da oração "O Domenico entrar aqui...como assim?" utilizando-se o algoritmo acima descrito. Note-se que a linha central marcada com "x" refere-se ao tom médio, nosso principal objeto de estudo. É notório que quase não ocorrem variações sistemáticas do valor do tom médio, indicando que ele possui uma relativa estabilidade. Observa-se também que a curva de F_0, toda em preto, somente irá ultrapassar as margens -3 e $+4$ ao fim da oração, quando há um aumento paulatino da frequência fundamental com a intenção de se 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.

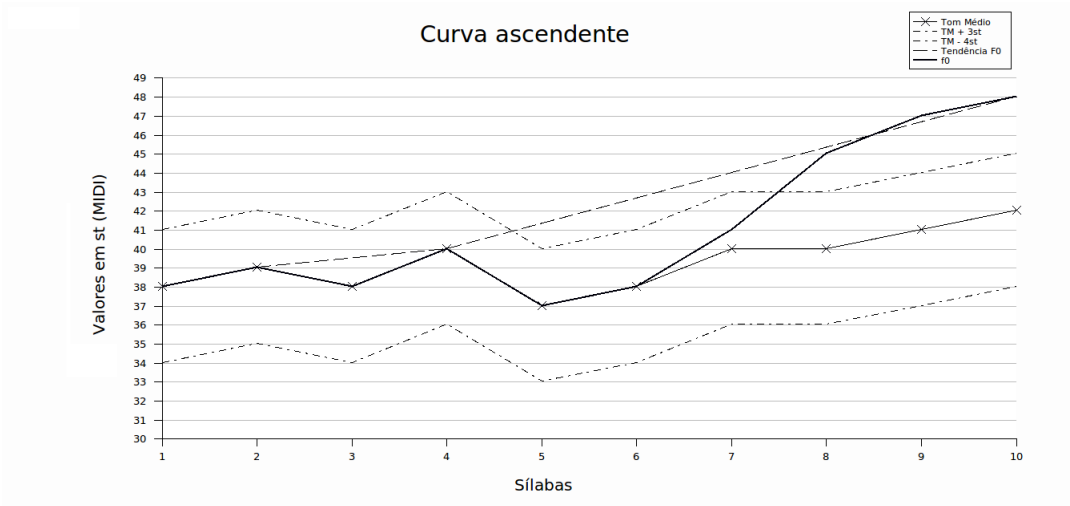

Figura 3: Exemplo de análise se usando o tom médio para a oração: "O Domenico entrar aqui...como assim?".

[9] A mesma figura pode ser vista ampliada no Anexo IX

2.2 Visão do Capítulo

Nossa proposta até aqui ligeiramente desenhada é a de que não se pode compreender a fala como um *locus* unicamente de relações simbólicas e de processamentos locais. Em nosso entender a fala é um sistema complexo, o qual demanda um modelo de decomposição que leve em conta também estes aspectos. O modelo de decomposição por nós escolhido é o que se encontra em Brooks (1997) e está ligeiramente resenhado neste capítulo. Em nossa interpretação teórica o modelo proposto por Ferreira-Netto (2008) atende aos pressupostos metodológicos da decomposição por atividade sendo que fica claro que a entoação possui por si só um processamento de uma série *atividades*, as quais se combinam com outras presentes na linguagem, como o uso da frequência fundamental para expressar conteúdos pragmáticos.

Ressaltamos que os cálculos aqui demonstrados são apenas a base teórica para a implementação da rotina ExProsodia, sendo que o objetivo principal deste trabalho é o de rever tais valores, principalmente aqueles que dizem respeitos as componentes a e b do (2.7) os quais foram descritos em Consoni et al. (2009), Consoni (2011). Contudo, destaca-se que no trabalho dos autores o foco de análise foram palavras, ao passo que no presente trabalho tomaremos a sentença como unidade de análise. A hipótese é a de que haveria diferenças, no que se refere a percepção de palavras, com relação a percepção de enunciados mais longos, como uma sentença, por conta da duração de cada um dos enunciados. A explicação para esta possível diferença é o efeito *tau* (SHIGENO, 1986) que trata da percepção diferencial como uma função do tempo. Em tese, dois estímulos podem ser percebidos como semelhantes ou não em função do espaço temporal entre eles colocado. Segundo experimentos de Shigeno (1986), a distância temporal entre dois estímulos afeta, de fato, a percepção tonal: quanto menor o tempo entre dois tons, menor a frequência do segundo tom é percebida pelos sujeitos.

3 Entoação

É fato que as línguas do mundo têm uma melodia e ritmo próprios. Desde os gregos, a observação das características acentuais e do tom melódico foram tidas como parte da prosódia (do Gr.*Προσωδια* junto / canto). Contudo, na perspectiva grega estes elementos eram tratados como parte da poesia, não sendo considerados elementos linguísticos propriamente (COLVIN, 2007). Desta maneira, na tradição ocidental, foi apenas no Renascimento, com os estudos de versificação que se verá uma análise linguística mais criteriosa, de modo que, um estudo sistemático do uso destes elementos, considerando-os parte integrante de um sistema linguístico, somente passou a ser feito com o avanço da Linguística no século XX.

É preciso destacar que a definição do termo entoação é, ainda hoje, relativamente, intrincada, pois como destaca Vaissière (2004) a dificuldade em se definir o termo está, também, em sua abrangência:

> "There is currently no universally accepted definition of intonation. The term may be strictly restricted to the perceived F_0 pattern, or include the perception of other prosodic parameters fulfilling the same functions: pauses, relative loudness, voice quality, duration, and segmental phenomena related to varying strengthening of the speech organs" (VAISSIÈRE, 2004, p. 238).

Segundo a autora, portanto, há duas possíveis abordagens para a entoação. Uma primeira, a qual podemos chamar de *abordagem restrita*, irá analisar apenas a percepção das variações de F_0 sem se ocupar de outras componentes como o acento ou a duração, ou seja, elementos prosódicos. A segunda seria uma *abordagem ampla*, de modo que esta irá se comprometer com os demais elementos prosódicos, tais como o acento, ou as pausas e ainda elementos entendidos como paralinguísticos, como a qualidade de

voz ou a amplitude do sinal de fala.

Em nosso entender, ambas as abordagens são válidas, de modo que a diferença entre as duas está mais relacionada a interpretação dos dados do que a definição de entoação em si. Neste sentido, nos ocuparemos de apresentar os modelos teóricos mais recorrentes afim de se estabelecer a nossa definição de entoação e como esta se relaciona com prosódia. Nossa consideração é a de que, para um estudo explicitamente fonético, a definição estrita é mais bem aceita, ao passo que a abordagem ampla seria melhor aceita em um estudo de cunho explicitamente fonológico. Dado que estamos nas margens de ambas as ciências preferimos a capacidade de manipulação e experimentação da primeira, associada ao poder analítico da segunda. Neste sentido, dividiremos o presente capítulo em duas grandes seções, a primeira tratará de maneira resumida das principais teorias da Fonologia Entoacional, ao passo que a segunda seção atentará nos aspectos físico-fisiológicos da produção da entoação. Uma terceira seção terá por objetivo tratar da percepção e da percepção da entoação.

3.1 A Fonologia da entoação

3.1.1 Os primeiros passos

Foi no bojo da teoria estruturalista europeia que se postulou que a menor unidade linguística seria o segmento (HJELMSLEV, 2006; MARTINET, 1975), ou ainda os elementos constituintes deste; os traços do nível merismático (BENVENISTE, 1976). Nesta abordagem, os fatos prosódicos se sobreporiam aos segmentos, sendo, portanto, parte da oração ou da sílaba (TROUBETZKOY, 1976), e não da vogal, como no modelo de Chomsky e Halle (1968). Para Troubetzkoy os fatos prosódicos deveriam ser chamados de *prosodemas*, devendo ser interpretados como modulações, 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, autremente dit la syllabe dans les langues qui comptent les

> syllabes et la more dans les langues qui comptent les mores. Les particularités peuvent être diviseés 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 liason ce ne sont pas les prospodèmes 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, adiante, 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 exemplifica esta premissa com a entoação de *continuidade* que significa ela-mesmo 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) (Grifos do autor).

Com relação a entoação, para o autor, o estado dos estudos linguísticos em sua época não eram ainda suficientemente avançados para se ter uma *fonologia da frase*, de maneira geral o autor 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. Desta maneira, o autor coloca um importante critério para a compreensão da entoação em termos de análise, o qual seria a questão de ser o tom intrínseco ou não a um sistema, isto é, 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. Caso o tom não tivesse 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 diferenciar as frases" (TROUBETZKOY, 1976).

Seguindo estes princípios o autor pontua que haveriam dois perfis de entoação para as línguas europeias não-tonais (i) *ascendente* e (ii) *descendente*, de modo que a primeiro perfil estaria ligado a continuidade de uma oração, ao passo que o segundo estaria ligado a finalização. O modelo de Troubetzkoy ainda incorpora um terceiro perfil o (iii) *enumerativo*, que estaria ligado a entoação dada quando se listam elementos de um conjunto, por exemplo.

Em nosso entender, há duas importantes questões em torno do tema, a primeira seria: onde a entoação se encaixaria no sistema de uma língua? Isto é, como se poderiam estabelecer pares mínimos também para entoação? Ou seriam apenas diferenciações categoriais, na superfície do fenômeno fonético em si? Para Troubetzkoy (1976), aparentemente, vale a segunda alternativa para a entoação, ao passo que a primeira valeria para o estudo da tonalidade. Esta resposta nos guia para nossa segunda pergunta: dentro desta visão, qual o papel da entoação na fonologia de uma língua? Aparentemente, o próprio modelo de Troubetzkoy responde esta pergunta, uma vez que a entoação seria parte da fonologia da oração. Deste modo, a entoação não seria composta de elementos duplamente articulados, nem mesmo distintos, mas, sim, uma componente composta de modo linear e distinguível apenas entre si.

A ideia de que a entoação seria uma componente não duplamente articulada poderia excluí-la do programa estruturalista de pesquisa. Em verdade, até o ponto que pudemos averiguar, grande parte dos desdobramentos posteriores destas corrente não levaram em conta as questões referentes a entoação. Generalizando, é possível dizer que a corrente estruturalista e posteriormente a funcionalista tiveram como principal preocupação a fonologia do segmento, deixando em segundo plano uma possível fonologia da entoação. Em nosso entender, a entoação recebeu um tratamento paralinguístico, isto é, foi tratada antes como parte da pragmática, do que propriamente

parte da fonética ou fonologia (FONàGY, 1991).

3.2 Fonologia e Entoação ou Fonologia da entoação?

O grande impulso em direção a uma "Fonologia Entoacional" se deu com a tese de Janet Pierrehumbert (PIERREHUMBERT, 1980). O impacto principal da tese de Pierrehumbert está na forma como se aborda metodologicamente a questão da entoação, tendo em vista que sua tese só tenha se tornado possível dentro da proposta da Fonologia Métrico-autossegmental (LADD, 1996). A Fonologia Métrico-Autossegmental seria a reunião das propostas da Fonologia Métrica, desenvolvida por Liberman (1975) e seus posteriores desdobramentos em Liberman e Prince (1977), com as propostas da Fonologia Autossegmental de Goldsmith (1976), além do estudo do *pitch accent* do Sueco feito por Bruce (apud CHO, 2010)[1].

Para Ladd (1996) a Fonologia da Entoação deveria seguir, ao menos, alguns princípios da Fonologia Segmental. O principal seria a capacidade de se recuperar elementos distintivos categorialmente (fonemas, traços, articulações) do *continuum* variável que constitui a língua, a nosso ver o mesmo princípio que regem os pares mínimos de Troubetzkoy (1976). Desta maneira, o estudo da entoação deveria antes categorizar elementos mínimos da entoação, com vistas a formar uma teoria com o escopo o suficiente para se analisar outros casos. Destaca-se que tal como Troubetzkoy (1976) o autor faz tais considerações, sempre fazendo distinção entre entoação e tom, sendo a segunda sempre aquelas variações de tom próprias do item lexical. Neste sentido, uma teoria fonológica da entoação deveria se pautar sobretudo, por entender, quais seriam os princípios universais que regulariam a produção da entoação. De modo geral pode se dizer que tal teoria teria de averiguar quais seriam as variáveis mínimas para um algoritmo da produção da entoação.

Lévi-Strauss (2010) propõe que haja, na história das ideias, uma "sonologia" anterior a uma "fonologia". De fato, veja-se por exemplo, a organização temática dos *Prin-*

[1] BRUCE, G. *Swedish Word Accents in Sentence perspective.* Lund: LiberLaromedel/Gleerup, 1977

cipes de Phonologie (TROUBETZKOY, 1976), ou ainda de *Sound Patter of English* (CHOMSKY; HALLE, 1968). Teorias linguísticas partem de uma Fonologia, e esta Fonologia parte de uma descrição fonética, ainda que contestável, ou seja de uma "sonologia". Contudo, será apenas em fins da década de 70 que se terá uma organização mais sistemática de uma "tonologia": uma proposta teórico/metodológica que faça ligar o nível do som ao nível da sintaxe, sem a necessidade do léxico.

A entoação está justamente neste ponto. Sua existência sem a cadeia segmental é vista como a música, apenas a organização temporal de elementos harmônicos, na definição de Helmhotz (1895), em que apenas a frequência fundamental pode ser compreendida como a base da harmonia. Ao passo que seu estudo ligando-a a cadeia segmental pode ser interpretado como um estudo da tonalidade, a despeito do seu aspecto global, o aspecto entoacional.

A questão da tonalidade foi formulada por Leben (1973), enquanto este estudava as línguas tonais em seu modelo suprassegmental. A diferença com relação a Leben (1973) está no fato de que o autor procurava determinar se o tom era intrínseco à sílaba ou ao segmento, tomando como base que a tonalidade é intrínseca ao léxico. O corpo teórico que se formaria em torno destes estudos e chamado por Ladd (1996) de *Autossegmental /Metrical Phonology* (LEBEN, 1973; GOLDSMITH, 1976; LIBERMAN; PRINCE, 1977; PIERREHUMBERT, 1980) assumiria que os processos entoacionais estariam 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, 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 e por força da lógica, um nível abaixo dos processos sintáticos. (PIERREHUMBERT, 1980) propõe que

estes tons pertencentes ao domínio da entoação seriam linguisticamente estruturados em um eixo de contradição: ou seriam altos (H - high), ou baixos (L -low). A proposta da autora não é nova e, sim, uma releitura dos trabalhos de Pike (1945), Bolinger (1964), Maeda (1976) e Liberman (1975).

Destaca-se que Pike (1945) ateve-se aos níveis tonais, estabelecendo quatro deles: *E (Extra High),H (high), M (mid) e L (low)*, ao passo que Bolinger (1964) foi atento às questões de como se configuram os padrões tonais em um eixo sintagmático, base da teoria de Pierrehumbert e abaixo apresentado na figura 4. Liberman (1975) e Maeda (1976) usaram-se da análise instrumental para construir o corpo de análise em suas teses, outro importante aspecto na metodologia usada na tese de Pierrehumbert (1980). O segundo passo da autora, após definir os níveis tonais, foi o de estabelecer as restrições sintagmáticas, as quais os tons estariam sujeitos. A Figura 4 demonstra como seria a distribuição possível dos tons em uma cadeia sintagmática.

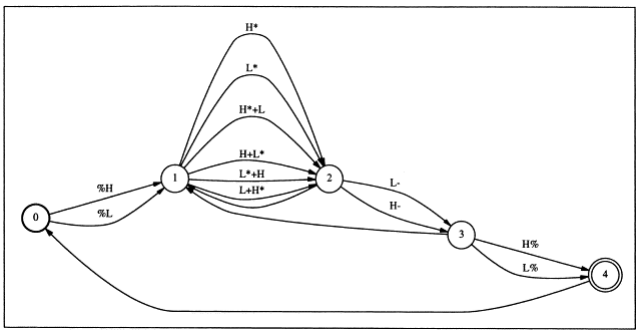

Figura 4: Esquema de organização sintagmática no modelo de Pierrehumbert (1980)

Os tons que vão de 0 a 1 são chamados de tons de fronteira (*Boundary tones*), 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*). O modelo de (PIERREHUMBERT, 1980) é bastante competente no que se refere a descrição de padrões tonais e foi a base do modelo ToBI (SILVERMAN et al., 1992), o qual tem por objetivo, justamente, contribuir para a descrição dos padrões entoacionais nas mais variadas línguas (JUN, 2006).

O modelo de Pierrehumbert (1980) não é único na área, embora seja o recorrente

em certas análises. Diversos outros modelos foram propostos, a seguir apresentaremos os principais estudos na área da entoação, o objetivo é apenas traçar um panorama dos estudiosos e trabalhos e não fazer uma análise crítica de cada um deles.

3.3 Os estudos de entoação.

Os estudos envolvendo entoação podem ser divididos em duas categorias, como destaca Ladd (1996): uma primeira de cunho impressionístico ou proto-fonológica e outra de cunho instrumentalista. Sendo a primeira marcada por trabalhos como os Pike (1945, 1964) e de Trager e Smith (1957). A segunda abordagem é marcada por nomes como de Liberman (1975) e Maeda (1976), tendo Lehiste (1970) e Ladefoged (2004) feito um breve resumo dos principais nomes desta abordagem.A principal característica metodológica da linha impressionística é o de sistematizar a entoação com um determinado número de elementos pertencentes à categoria distintas ou níveis, contudo não indo além das descrições de oitiva das curvas entoacionais (LADD, 1996). Ao passo que a segunda abordagem analisaria os movimentos de pitch obtidos em medições experimentais, com aparelhos próprios para este fim.

Maeda (1976) citando Lieberman (1965) pontua que a principal falha da abordagem impressionística é o de não verificar por meio de experimentos se tais análises estão corretas. Em nossa interpretação, uma melhor integração destes fenômenos acontecerá na tese de Pierrehumbert (1980), uma vez que seu modelo se baseia na análise dos padrões de variação de f0 observados em análises experimentais. 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). Citando Pierrehumbert (1980), Bolinger (1986, 1989) e Fonàgy (1991), a autora demonstra como as abordagens da entoação podem se diferenciar diametralmente.

> "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).

Hirst e Cristo (1998) ao proporem um modelo universal de análise da entoação, destacam a ambiguidade existente decorrente do termo entoação, 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)

As abordagens mais recentes do tema da entoação como as de Dilley (2005) e Cho (2010) tem se baseado na ideia de que as relações entre as variações de F_0 e outros aspectos linguísticos não são óbvios, de modo que dado o a natureza do objeto e o método de análise escolhidos deve se atentar para questões além entoação. Xu (2005) discute o mesmo tema, tendo em vista, que processos como entoação e tonalidade podem ser compreendidos sobre um mesmo aspecto o de melodia da fala, se tomados como etapas de um sistema de transmissão de informação.

t´Hart, Collier e Cohen (1990) propõem que a entoação seja compreendida basicamente como a variação de F_0 no decurso de uma elocução, de modo que certos movimentos tonais (pitch movements) são relevantes para o ouvinte. Estes movimentos tonais no plano fisiológico/cognitivo corresponderiam a comandos para movimentos específicos das cordas vocais, os quais podem ser recuperados no sinal da fala. Foi a partir deste princípio que grande parte dos trabalhos do IPO (*Instiuut voor Perceptie Onderzoek*, em holândes) (T´HART; COHEN, 1973; T´HART, 1981; T´HART; COLLIER; COHEN, 1990) em Eindhoven se apoiou. Desta maneira, seguindo este princípio propomos, dado o escopo desta pesquisa, um estudo sistemático dos processos fisiológicos e físicos da produção de F_0, bem como quanto à percepção deste.

3.4 A Fonética da entoação

A questão da produção da melodia da fala, no caso da frequência fundamental (F_0), parece ser de maneira geral bem aceita e explicada nos detalhes na literatura fonética. Assume-se que a frequência fundamental é o produto físico das vibrações das cordas vocais e o *pitch* sua contraparte perceptiva. Em Stevens (2000) temos uma descrição baseada no modelo físico de "massa mola", em que sistemas oscilatórios são analisadas a partir de um ponto de equilíbrio (quando a força da mola sobre a massa é nula $F = 0$) em direção a um estado S_n (quando a força da mola sobre a massa é diferente de zero $F \neq 0$) e de volta ao estado de equilíbrio, seguiremos este modelo e descrição em nosso trabalho.

3.4.1 A Laringe

Anatomicamente, a laringe é uma 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 cordas vocais; as cordas vocais/glote e abaixo pela tráquea.

A figura 5 é uma ilustração de Gray (1918) e demonstra a posição das cordas vocais e do ventrículo da laringe, em relação à epiglote e a tráquea, em um corte coronal.

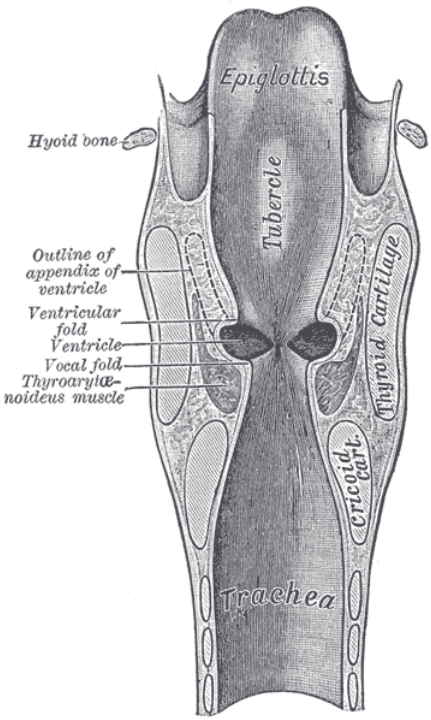

Figura 5: Desenho anatômico da laringe, vista em um corte coronal. Retirado de Gray (1918)

A figura 6 demonstra a posição das cordas vocais e do ventrículo da laringe, em relação à epiglote e a tráquea, em um corte sagital.

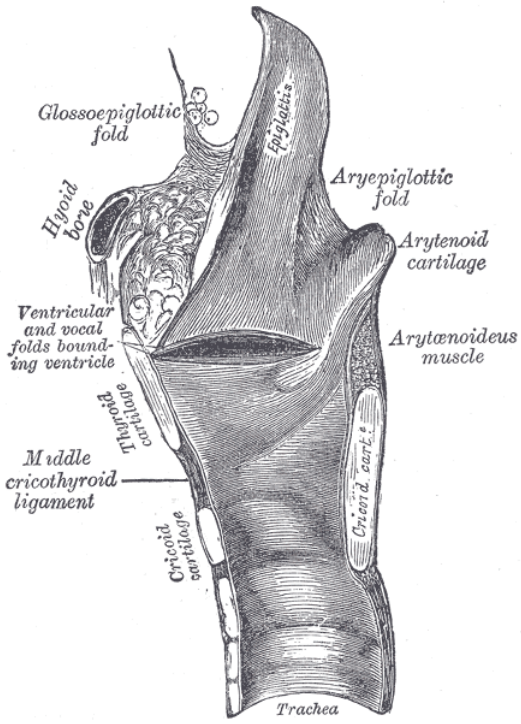

Figura 6: Desenho anatômico da laringe, vista em um corte sagital. Retirado de Gray (1918)

Stevens (2000) sugere que a laringe é apenas um sistema suporte para as cordas vocais, ao passo que Gray (1918) propõe que a laringe seja também chamada de "órgão da fala". A diferença se encontra no fato de que o segundo tende a atribuir funções à laringe, enquanto o primeiro a vê, apenas como um mecanismo suporte, sem funções específicas. 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 para proteção deste. Em geral, entende-se que essa proteção visa não deixar que corpos estranhos atinjam o sistema respiratório. Outros procedimentos como o de deglutição e regulação da pressão torácica levaram a laringe a possuir uma extensão maior do que a necessária para a produção da fala unicamente (BAER, 1975). Negus (apud BAER, 1975) sugere que a única diferença da laringe humana para a laringe dos demais animais não é de ordem estrutural, mas, sim, na habilidade adquirida pelo homem de controlá-la. Na seção, a seguir faremos melhores considerações sobre a glote e as cordas vocais, responsáveis pela produção da frequência fundamental.

3.4.2 A glote

A glote corresponde as cordas vocais (*plicoe vocales*) e o espaço entre elas, a *rima glottidis*. (GRAY, 1918) descreve anatomicamente as cordas vocais da seguinte maneira:

> As cordas vocais (*plicoe vocales*) tem a função de produzir os sons e é envolta por duas fortes faixas chamadas de *ligamentos vocais* (*ligamenta vocales*). Cada um destes ligamentos consiste de uma faixa de tecido elástico e amarelo, ligado pela frente com o ângulo da cartilagem tireoide e atrás pelo processo vocal[2] da aritenoide. Seu limite inferior é contínuo com a parte lateral delgada do *conus elasticus*, seu limite superior forma a divisão inferior do ventrículo da laringe. Lateralmente está o *vocalis muscle* (*músculo vocal/músculo tireoaritenoideo*), o qual está em paralelo aos ligamentos vocais. As cordas vocais são cobertas por uma membrana mucosa, a qual é extremamente delgada e aderente a sua superfície. [Tradução, notas e grifos nossos](GRAY, 1918)

As figuras 7 e 8 são ilustrações de Gray (1918) para os músculos, cartilagens e posições relativas das partes que compõem a glote. A figura 7 apresenta a estrutura muscular e cartilaginosa da glote. Pode se notar nesta figura os ligamentos vocais (com a anotação *chorda vocalis*), os músculos tireoaritenoideos (em vermelho à esquerda), a ligação abaixo com o ângulo da cartilagem tireoide e a ligação acima com os processos vocais (anotadas na figura como cartilagem aritenoide), responsáveis pelos movimentos da glote. Morfologicamente, as cordas vocais são compostas de três partes: o músculo vocal, a *lamina propria*[3] e o epitélio (HIRANO; KURITA; NAKASHIMA, 1981 apud COLTON; CASPER; LEONARD, 2006).

[2] O processo vocal é uma pequena parte frontal da cartilagem aritenoide, onde as cordas vocais se ligam é ele o responsável pelo movimento da glote.

[3] A *lamina propria* é um tipo de tecido conjuntivo frouxo, o qual recobre o epitélio e junto dele forma a mucosa epitelial (COLTON; CASPER; LEONARD, 2006)

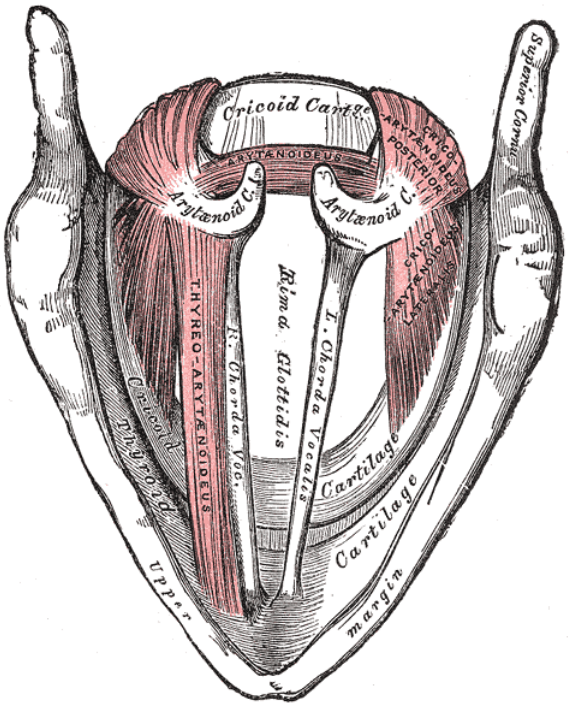

Figura 7: Desenho anatômico da glote, vista de cima em um corte transversal. Retirado de Gray (1918)

A figura 8 apresenta a *rima glottidis* aberta, embora não aparente na imagem pode-se notar a ligação com a cartilagem tireoide na parte superior da imagem, abaixo da raiz da epiglote, no ponto em que as cordas vocais se encontram. É possível observar a cartilagem aritenoide na parte inferior, onde se encontram os processos vocais, neste ponto a anotação que se lê é cartilagem cuneiforme.

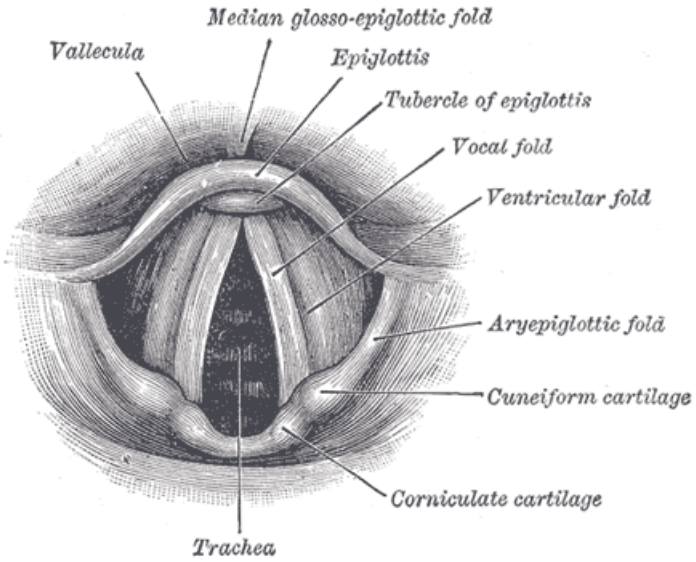

Figura 8: Desenho anatômico da *rima glottidis* vista de cima, a partir da entrada da laringe. Retirado de Gray (1918)

.

Gray (1918) propõe que os músculos da 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 cordas vocais . De acordo com autor (ver figura 7):

> Os *Cricoaritenoides posteriores* separam as cordas vocais, e consequentemente, abrem a glote, rotacionando as cartilagens aritenoides para fora em torno de um eixo vertical, passando através das juntas cricoaritenoides; então o processo vocal e as cordas vocais a ele ligadas se distanciam. Os *Cricoaritenoides laterais* fecham a glote, rotacionando as cartilagens aritenoides para dentro, de modo que elas aproximem os processos vocais. O *Aritenoide* aproxima a cartilagem aritenoide e então fecha a abertura da glote, especialmente em sua parte posterior.[Tradução nossa](GRAY, 1918)

Os músculos que tomam parte no grau de tensão das cordas vocais são respectivamente o *cricotireoideo* e o *tireoaritenoideo*, ver figura abaixo:

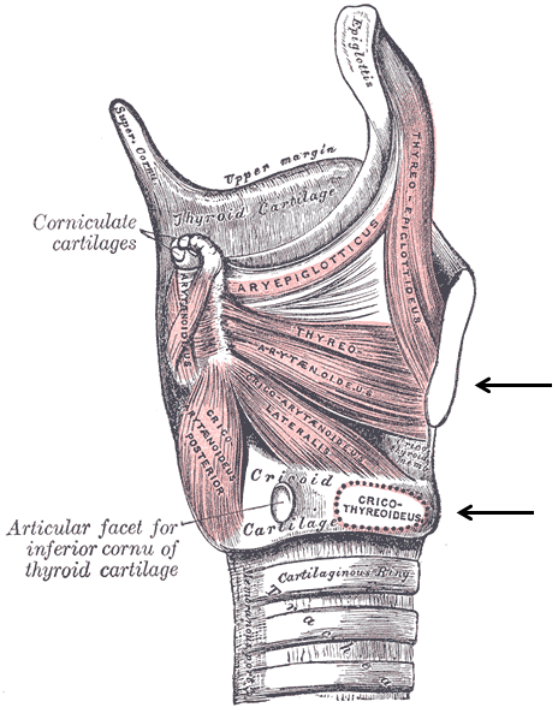

Figura 9: Vista lateral dos músculos da laringe. Retirado de Gray (1918). As setas indicam a posição dos músculos *tireoaritenoideo* e *cricotireoideo*, respectivamente

.

O primeiro alonga ou tensiona as cordas vocais, por meio da elevação da ângulo da membrana da tireoide, gerando um aumento da distância os processos vocais. O segundo, por sua vez, está ligado às cordas vocais e controla sua elasticidade e direção, rotacionando a cartilagem aritenoide e deixando a *rima glottidis* mais estreita ou mais larga.

Estes movimentos glotais são diretamente responsáveis pela variação da frequência fundamental. De acordo com Stevens (2000), devido a morfologia das cordas vocais tem-se que as diversas camadas vibram em modos diferentes. Este fato decorre da rigidez própria de cada um dos músculos e tecidos epiteliais que compõem a glote. (KAKITA; HIRANO; OHMARU, 1981 apud 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 cordas vocais. O "Módulo de Young" é representado pela fórmula (STEVENS, 2000):

$$\gamma = \frac{\Delta F / A}{\Delta x / x} \tag{3.1}$$

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/cm^2 (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 modulo de Young (cerca de 10^4 dynes/cm^2) para que se alonguem cerca de 10% do seu tamanho original. Por outro lado, a mucosa necessitaria de cerca de 10^6 dynes/cm^2 para sofrer o mesmo alongamento.

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

3.4.3 A produção da frequência fundamental

O processo de produção de uma frequência fundamental (ou natural) na glote pode ser simplificada, embora os mecanismos que lhe são subjacentes possuam uma complexidade ainda não compreendida em um modelo geral. Apenas dois fatores são necessários para que o processo de fonação se inicie. O primeiro refere-se à diferença entre a pressão sub-glotal (P_s) e a pressão acima da glote (P_a); o segundo

fator necessário é a existência de um fluxo de ar entre os pulmões e o meio externo. Assume-se que a pressão sub-glotal 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. Simplificadamente, o processo produção de uma frequência fundamental se inicia quando a pressão sub-glotal atinge um determinado limiar de pressão de fonação (TITZE, 1988, 1989; LUCERO et al., 2009) levando a uma abertura lateral das cordas vocais (painel B da figura 10). A citada movimentação lateral gera um movimento de rotação das cartilagem aritenoide (Painéis C à F, na figura 10). Em C pode-se observar uma movimentação em sentido interno, ao passo que em D, E e F pode-se observar o movimento em sentido externo. O painel C mostra um deslocamento longitudinal em direções opostas dos pontos de conexão nas cartilagens aritenoides levando a um enrijecimento das cordas vocais. Devido aos valores de pressão sub-glotal estarem acima do limiar de pressão de fonação tem-se novamente um movimento longitudinal de separação das cartilagens aritenoides (Painéis D, E e F) e por consequência um esticamento das cordas vocais.

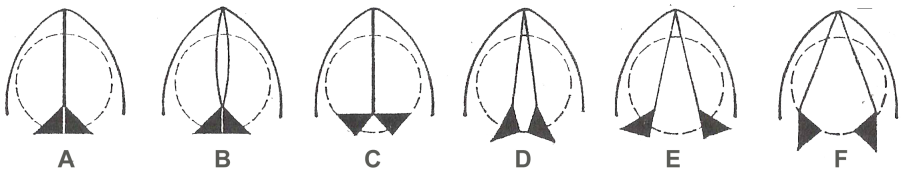

Figura 10: Vista superior das etapas de abertura da glote, os triângulos negros representam as cartilagens aritenoides. Retirado da Internet

Conforme o fluxo de ar entre as cordas vocais se mantenha contínuo e a pressão sub-glotal mantenha-se em um valor que supere o valor necessário para as forças laterais das cordas vocais se anularam e voltarem ao estado repouso, teremos uma frequência natural. 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 corda 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 cordas

vocais, levando as cartilagens aritenoides a movimentar-se em sentido longitudinal de aproximação, como apresentado nos painéis F, E e D da figura 10.

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 cordas vocais, de modo que ocorrerão perdas em forma de fricção no contato entre as partes. Tais perdas de energia levarão há uma oscilação do tecido e por consequência a vibração das cordas vocais. Embora soe simplista, o mecanismo de oscilação possui uma complexa gama de relações. Titze (1988) propõe um diferencial de tempo de reação entre as parte inferior e superior das cordas vocais. 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 cordas vocais, enquanto a parte superior levaria mais tempo para se submeter à esta força. A figura 11 demonstra esquematicamente o mecanismo de oscilação, proposto por Titze (1988)

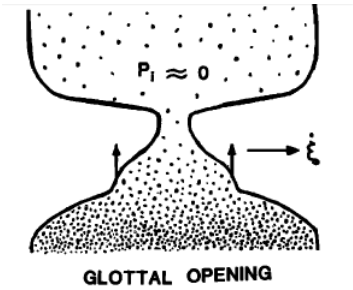

Figura 11: Retirado de Titze (1988)

No caso da abertura da glote, em (a), tem-se uma força lateral como a acima explicado, com a auxílio da figura 10, com uma velocidade $\dot{\xi}$ com direção lateral. Observa-se que a parte inferior sofre alterações, enquanto a parte superior mantem-se inalterada. No instante em que a parte superior mudar sua forma, a parte inferior terá voltado ao seu estado natural, gerando uma alternância entre as partes, levando a geração de uma frequência natural. Será a velocidade entre essas alternâncias que determinará o valor desta frequência. Com a troca de energia entre o fluxo de ar e o ambiente e diminuição da pressão sub-glotal tem se uma mudança de sentindo

na velocidade $\dot{\xi}$, a qual passa a se mover em sentido contrário. Novamente, a parte inferior sofre uma alteração antes da superior, ver figura 12.

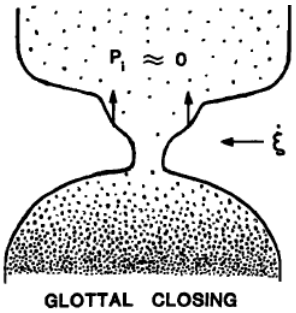

Figura 12: Retirado de Titze (1988)

A figura 13 demonstra a movimentação das massas da glote. Os painéis de 1 a 3 ilustram a movimentação da parte inferior antes da parte superior das cordas 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. Os painéis de 4 a 5 demonstram o processo intermediário entre a abertura e o fechamento. Se nos fosse possível fazer uma *motion picture* com os painéis de 1 a 10 notar-se-ia a oscilação ondulatória das cordas vocais, como acima dito.

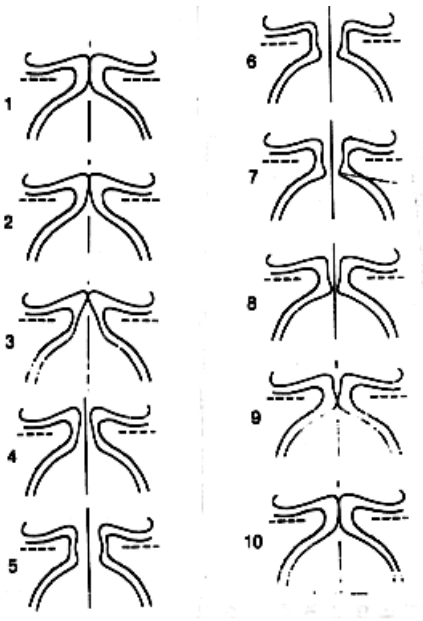

Figura 13: Vista lateral das etapas de abertura da glote, baseado em Baer (1975)

Em seu trabalho posterior, Titze (1989) propõe mais duas variáveis para a variação da frequência fundamental. A primeira delas é o grau de expansão da corda vocal, em relação ao seu estado neutro; a segunda o comprimento da corda vocal em estado de repouso. Em um segundo momento o autor relaciona estas duas variáveis a pressão sub-glotal e o comprimento da corda vocal em um modelo matemático unificado. De acordo com a modelagem do autor, a frequência fundamental é menos sensível ao comprimento quando as cordas vocais são curtas, do que quando elas são mais longas. No caso das cordas vocais curtas, a pressão afeta de maneira mais direta o controle dos valores de frequência fundamental. No caso das cordas vocais curtas, temos uma maior variação de F_0, conforme eleva-se a pressão sub-glotal, ao passo que em cordas vocais mais longas não se vê o mesmo efeito.(TITZE, 1989a) demonstra que cordas vocais masculinas, em média, são cerca de 1,7 vezes maiores que as femininas[4]. De modo que vozes masculinas têm a frequência fundamental em torno de 120 Hz e vozes feminina em torno de 210 Hz, ou seja, cerca de 1,75 vezes maiores que as masculinas; confirmando, em partes, a hipótese do autor.

3.5 A Percepção da entoação

3.5.1 Propriedades da percepção

Bertrand Russell propõe, a certa altura de sua *Análise da Matéria* (RUSSELL, 1978), uma *Teoria Casual da percepção*. A proposta inicial do autor é a de que por um lado, o objeto percebido está em nós, ao passo que por outro lado o mesmo objeto está exteriorizado. Assim, nas palavras do autor, ao se olhar o Sol, "vemos o Sol", enquanto a ciência demonstra que há um circuito de transformações, que vão do "Sol", ao "ver o Sol". A viagem da luz do Sol a Terra, aos nosso olhos, dos olhos a mudança de condição para que seja processado pelo nervo óptico e por fim ao cérebro, onde é finalmente "visto". O mesmo princípio guiou a segunda geração de físicos quânticos,

[4]O comprimento das cordas vocais masculinas é de 17 mm em média, enquanto as cordas vocais femininas tem cerca de 10mm

principalmente Heisenberg, com seu princípio da incerteza e Schrödinger, com o famoso paradoxo do gato meio morto e a questão do observador.A noção de *percepção* e *observação*, ou ainda de *percepto vs. observador* é um conceito não claro. O objeto percebido está em uma relação de necessidade com observador: o objeto percebido só existe em função do observador e vice-versa. Contudo, como colocado por Russell (1978) esta relação tende a ser falseada pela própria lógica da observação, a qual passa por tantos níveis, dependendo da "distância" entre objeto e observador.

A saída para este problema já havia sido dada pelo trabalho de Max Wertheimer. Em seu artigo *Laws of Organization in Perceptual Forms* (WERTHEIMER, 1938), o autor propõe que a percepção das formas segue pressupostos de ordenação e que os matizes que compõem uma cena não são do domínio da forma, e sim, da percepção, de modo que o percepto é uma construção do observador. Em uma análise menos criteriosa vale dizer que é o observador quem reconstrói a cena, dado um conjunto de regras e domínios a ele intrínsecos. Observar uma criança brincando em um parque não se trata de analisar as matizes de azul, verde e vermelho que compõem a cena, mas, sim, é ver "uma criança", "uma ação de brincar" e "um parque".

Vernon (1974) assume que a percepção está associada à uma ideia de forma, isto é, o elemento percebido tem de possuir uma forma característica ou própria, a qual se distingua de outros objetos, os quais estão em uma relação de proximidade ou contiguidade em um "ambiente natural" (WERTHEIMER, 1938). A questão da forma, precisa ser associada a outros processos cognitivos, caso da memória ou da atenção, por exemplo. No caso da memória, a noção de "permanência do objeto" é a principal unidade de estudo, pois ela assegura uma referência cognitiva, também, chamada de imagem mental. A permanência do objeto contribui para a organização do universo prático, uma vez que assegura a visualização dos objetos percebidos mesmo quando fora do alcance de visão. Na caso da percepção, a permanência contribui para a percepção da mudança de estado dos objetos, ou de posição ou deslocamento no espaço (PIAGET; INHELDER, 1989). O estudo de Baird et al. (2002) procura demonstrar,

por meio de espectroscopia, como o processo de permanência do objeto é aprendido por bebês. Não nos fica claro como o processo global ocorre em um nível cognitivo, uma vez que o caráter do estudo é de ordem neurocientífica. No entanto, o estudo aponta que a maturação do lobo frontal parece ter um papel fundamental, sendo esta a região responsável pelo movimento e linguagem. De algum modo, o percepto está condicionado a existência de um objeto que lhe seja semelhante ou totalmente oposto. Nesse caso, a memória exerce uma função auxiliar no processo de significação. A seguir faremos uma pequena análise usando-se da percepção visual tentando concatenar estas ideias.

3.5.2 Jack LaMotta e a percepção

Como dito acima, a memória tem uma importante função no processo de atribuição de significado ao objeto percebido. No filme *Raging Bull* Martin Scorsese relata a vida do pugilista Jack LaMotta (Robert DeNiro), um lutador diversas vezes campeão e que teve seu auge em meados da década de 40. O filme produzido em finais da década de 70 procurava reproduzir o ambiente de Nova York, no período em que viveu o lutador. O primeiro recurso óbvio é o uso do figurino e cenário típicos. O segundo passo na caracterização do ambiente foi o uso da fotografia.

Scorsese optou pelo uso de uma fotografia preto e branco, com um certo nível de saturação nos tons escuros para certas externas e escuras, e uma saturação menor em ambientes de maior claridade, revelando com maior intensidade o branco. No primeiro caso tem se uma maior evidência dos tons escuros, em que a oposição preto e branco se torna particularmente chamativo. No segundo caso há uma maior valorização dos tons de cinza. As figuras 14 e 15 apresentam duas cenas distintas do filme, uma de pouca claridade e uma segunda já no ringue com um maior nível de luminosidade.

Figura 14: Cena externa do filme "Touro indomável"

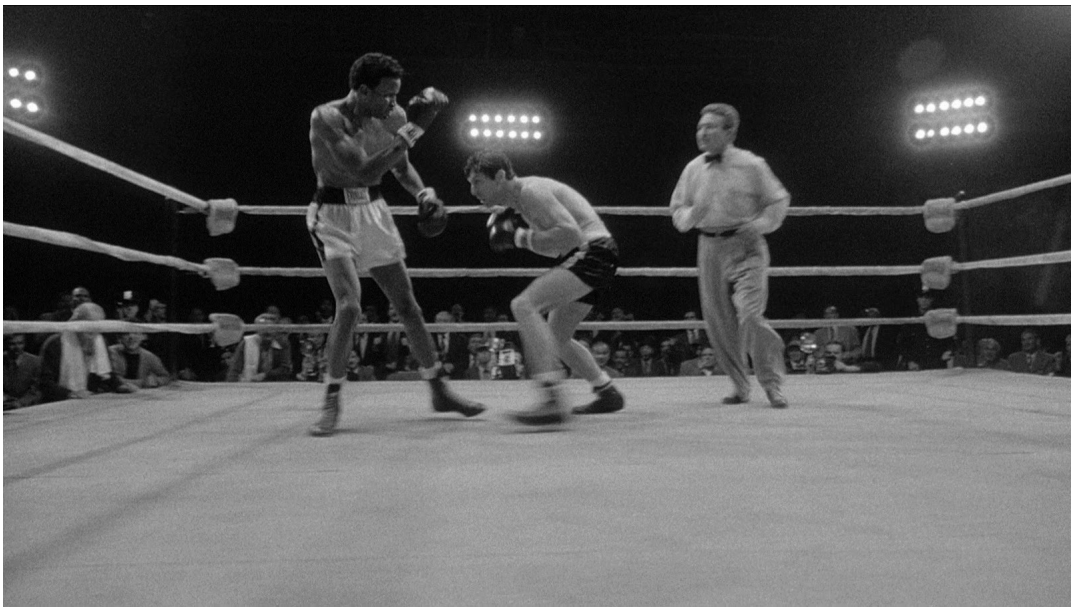

Figura 15: Cena *indoor* do filme "Touro indomável"

O tipo de fotografia escolhida, como apresentada, remete aos filmes preto e branco produzidos até meados da década de 40 (HAYES, 2005) . Dessa maneira, o tipo de fotografia nos transporta para um tempo discursivo anterior ao do filme, mas que é ao mesmo tempo o da narrativa. E, podemos dizer, que é a percepção e a memória quem produzem esse efeito de sentido. A memória, por meio da permanência do objeto, remete a uma tipologia de filmes, que podemos chamar de "filmes antigos" e nos faz perceber o objeto, no caso o filme, como parte deste conjunto, embora seu lançamento

tenha ocorrido na década de 80, emprestando uma nova discursividade a releitura da biografia de LaMotta.

3.5.3 Percepção e língua

Até o momento tratamos do tema em dois aspectos, o primeiro refere-se ao seu dinamismo com relação ao objeto percebido. O segundo aspecto trata da ideia de que é a percepção está associada a forma e que a forma está ligada ao conceito de "permanência do objeto" e da sua relação com a memória. Nesta seção tentaremos estabelecer uma relação entre a percepção e a língua, ou melhor, qual fator que nos leva a entender uma língua como língua e não como um conjunto de ruídos aleatórios.

Há cerca de oito anos, o *Massachusetts Institute of Technology* promoveu uma série de conferências, denominadas "From sound to sense". O objetivo era avaliar os avanços ocorridos nos estudos que lidam com som e sentido, bem como, planejar os próximos passos na pesquisa. De um ponto-de-vista da história da ideias institucionais chama-nos a atenção o fato de que tais conferências não foram organizadas ou proferidas por pessoas ligadas ao Departamento de Linguística, mas, sim, pelo Departamento de Engenharia, mais especificamente pelo *Research Laboratory of Electronics*. O que nos chama mais a atenção é o fato de que não se tratava de engenheiros em busca de um volapuque computacional, mas, sim, linguistas ou cientistas ligados a Linguística, buscando a solução para o problema da variabilidade dos sons e a percepção destes como sons linguísticos.

Gunnar Fant na introdução de seu livro *Speech acoustics and phonetics* (FANT, 2004b) conta que durante sua estada, em meados da década de 50, nos Laboratórios do MIT e de psicoacústica em Harvard procurava entender como os processos de "abstração" do fenômeno físico eram processados pela mente. Nos termos do autor, queria entender como aquele contínuo sonoro se tornava categórico e dotado de significação, após serem percebidos pelos falantes. A resposta veio com o auxílio de um renomado linguista: Roman Jakoboson. R. Jakobson havia trazido em sua concepção de estudos

linguísticos a noção de "par mínimo" de Troubetzkoy (TROUBETZKOY, 1976) e dizia que um modelo fonológico universal deveria trazer a mesma concepção em sua base metodológica. A associação destas duas necessidades fez surgir *Preliminaries to Speech Analysis* (JAKOBSON; FANT; HALLE, 1952).O mesmo princípio seria aplicado anos mais tarde no *Sound Patterns of English* (CHOMSKY; HALLE, 1968), embora não haja uma citação direta da obra de Jakobson, Fant e Halle (1952). Na realidade, o trabalho de Chomsky e Halle exclui da análise a fala e por esta razão a percepção. A passagem do contínuo ao categórico se perde, uma vez que é a relação do categórico com o categórico se torna o objeto da análise, o que o compõe não é do escopo da Linguística e sim de uma ciência da fala, a qual demoraria cerca de 30 anos para surgir, no próprio MIT.

A ciência da percepção de fala teve de se manter junto dos foneticistas e psicólogos não conseguindo tomar forma como parte da Linguística ao longo das décadas de 70, 80 e 90. Este fato levou uma sistematização curiosa do estudo da percepção da fala nos primeiros anos. Um primeiro passo nesse domínio foi compreender como o "speech code" era interpretado em meio a um conjunto de ruídos não-linguísticos. Podemos chamar essa fase de "cocktail party effect" (CHERY, 1953), nome dado pelos cientistas do período e se refere a capacidade humana de obter informações linguísticas, a partir de um ambiente em ruído, como em uma festa. A face contrária era a busca por compreender como ocorria a integração entre sinais não linguístico e informações linguísticas, como a leitura facial ou o arqueamento dos olhos. As etapas consequentes tiveram um caráter cerebral/cognitivo se preocupando com processos já internalizados (SUMBY; POLLACK, 1954).

A partir de meados década de 70 e finais da década de 60, os efeitos da teoria de Noam Chomsky começavam a ser percebidos e em já em 1972 Kenneth Stevens (STEVENS, 1972) propõe a teoria quântica de percepção e produção da fala. Em um período anterior Alvin Liberman e seus colegas apresentam a Teoria Motora da percepção de fala (LIBERMAN et al., 1957), contudo continuam seus estudos até uma

revisão completa do modelo (LIBERMAN A.M. et al., 1967; LIBERMAN; MATTINGLY, 1985) e embora, as teorias sejam distintas em seus detalhes e objetivos, são parecidas em suas essências. Se detectar o "speech code" era um problema para seus antecessores, detectar a natureza do que era percebido e servia de base para o *speech code* era o trabalho dessa nova geração. Em verdade, o paradigma que se assentou com estas teorias é ainda a lógica dos estudos atuais: quais características acústicas são determinantes para a percepção/distinção de um segmento? É a pergunta que todo foneticista em começo de carreira deve se fazer.

De volta a conferência do M.I.T. em 2004. O principal convidado naquele evento foi Gunnar Fant, o responsável, por assim dizer, junto de R. Jakobson e Morris Halle, pela forma como a percepção deveria ser vista dentro da análise fonética. A opinião final do foneticista era a de que a variabilidade é uma condição de existência da língua, embora a mesma variabilidade só exista em termos relacionais, de modo que ela deve ser testada "ceteris paribus[5]" e completa: "Absolute invariance is a property of the perceptual-cognitive process induced by linguistic competence rather than a property of the physical form" (FANT, 2004a)

Deste modo, os estudos feitos até ali estavam corretos, de acordo com Fant (2004a). A menor *forma* perceptiva da língua seria o segmento e a "essência" do segmento era o que os faria diferentes entre si, ainda que esta essência seja tão complexa quanto a própria linguagem. O *speech code* é, então, a forma como a língua se manifesta e deixa ser percebida como tal. Embora, não exista uma definição precisa deste objeto, ele pode ser tanto o conjunto de elementos que compõem a fala (segmentos, prosódia, entoação), quanto um conjunto de objetos aleatórios que compõem um domínio mais amplo, como aqueles que formam a essência do sons da fala (sonoridade, oclusão, bilabialidade, etc.). De modo que a percepção da fala seria um processo de identificação e organização destes elementos em uma série de níveis cada vez mais elevados na língua. A percepção e interpretação de cada um desses níveis levaria a compreensão global da informação contida naquele ato de fala. Em nosso entender, a entoação

[5]Mantidas inalteradas todas as outras coisas

é parte deste *speech code*, sendo um destes níveis de percepção. Na seção a seguir faremos algumas considerações sobre o assunto.

3.5.4 A proposta de t´Hart, Collier e Cohen (1990)

t´Hart, Collier e Cohen (1990) propõem um estudo perceptivo da entoação para o holandês. Os autores argumentam que os estudos linguísticos se organizam, ora em busca de unidade mínimas da entoação, ora na análise das variações da frequência fundamental. No primeiro caso, pode se cometer o erro de não se apreender as informações necessárias transmitidas pela melodia da fala em um sistema de distinções linguísticas. No segundo caso, determinar que qualquer variação de frequência fundamental em curvas melódicas é relevante pode estar equivocado, posto que pode-se ter variações não intencionais por parte do falante. A alternativa proposta pelos autores é de se criar um modelo baseado no ouvinte, um "modelo do ouvinte". Desta forma, o estudo da entoação se voltaria para a análise das variações tonais relevantes para o ouvinte, isto é, quais seriam as estruturas de percepção e interpretação da informação entoacional relevante, assim as minímas variações seriam desconsideradas, tornando o modelo mais robusto. Para determinar o que seria ou não relevante seriam estabelecidos uma série de limiares de percepção, sendo que as variações de F_0 seriam relevantes, apenas, quando os valores de frequência fundamental atingissem estes valores de limiar.

t´Hart, Collier e Cohen (1990) propõem quatro limiares de percepção, a saber: (i) O limiar de diferenciação tonal; (ii) o limiar de diferenciação de distância tonal; (iii) o limiar absoluto de mudança tonal e o (iv) limiar de diferenciação de mudança tonal. O primeiro deles trata da acurácia da percepção de F_0, isto é, a partir de qual valor relativo (em porcentagens) o sujeito é capaz de diferenciar dois tons sucessivos. O segundo limiar refere-se a habilidade de se identificar a diferença (distância) entre dois tons sucessivos ou simultâneos. O terceiro limiar faz referência a velocidade de mudança de F_0, ou seja, quão rápido deve variar a frequência fundamental para

que o ouvinte note a diferenças. Por último, o limiar de diferenciação de mudança tonal relaciona a velocidade de mudança a mudança tonal em si, qual seria o tempo necessário entre dois tons para que eles sejam percebidos como diferentes.

3.6 Discussão

Embora o assunto seja vasto e os modelos não tenham dado conta de maneira completa do objeto, assumimos que os processos aqui elencados são os princípios básicos para a compreensão do funcionamento da glote e, por consequência da produção da frequência fundamental. Partimos do princípio que a laringe não é apenas um sistema suporte para a produção da fala, mas, sim, o componente principal, embora demande-se mais estudos sobre a sua função na produção da fala. De outro ponto-de-vista, a glote precisa ser contemplada dentro deste aspecto maior, não como componente principal, mas como parte do sistema, junto da laringe.

A componente *sustentação* de nosso sistema, contemplado no capítulo 2, pode ser compreendida nessa chave. A sustentação pode ser visto fisiologicamente como uma reposta ao comprimento das cordas. Dentro do modelo de Titze (1989) o controle tonal é feito, principalmente, por essa variável em conjunto da pressão sub-glotal, tendo essa um papel secundário. Por outro lado, o experimento do autor utilizando a pressão sub-glotal como variável, possuía uma regulagem fina dos valores desta. Em nosso entender, a mesma regulagem fina não ocorre no discurso normal, enquanto uma regulagem da força aplicada à corda vocal parece ser plausível.

A extensão das cordas vocais, segundo Gray (1918), é regulada pelos músculos cricotireoideos e tireoaritenoideo, ambos inervados pelo Nervo Laringal Superior. O músculo cricotireoideo é inervado pelo Ramo Externo do nervo laringal superior (RENLS) (FREITAS et al., 2007; CAMPOS, 2008), um nervo motor capaz de controlar a distensão e extensão do músculo. Estudos conduzidos com pacientes que sofreram tireoidectomia mostram que, em cerca de 25% dos casos, estes pacientes sofreram disfonias permanentes ao ter o ramo externo do nervo laringal superior afetado (KARK

et al., 1984). O mesmo estudo demonstrou que quando o ramo externo foi localizado e preservado, apenas 5% dos pacientes sofreram disfonias. Segundo Souza (2008), os casos de disfonia decorrentes de lesão no ramo externo do nervo superior laringal são, em geral, o de se conseguir emitir registros mais agudos. Estes estudos demonstram o papel funcional dos músculos e nervos citados, bem como a existência de uma articulação primária na glote.

O controle muscular citado tem uma resposta na curva final de F_0. A contração dos músculos cricotireoideo e tireoaritenoideo leva a um aumento da frequência fundamental, enquanto a distensão gera um abaixamento da frequência fundamental. Nossa hipótese é de que tal mecanismo fisiológico corresponda, em um nível linguístico, a uma possível articulação glotal. Xu e Wang (2001) propõe que a articulação dos tons possua duas contrapartes: uma dinâmica e outra estática. A dinâmica seria responsável pelos movimentos lineares, como *fall* e *rising*, ao passo que à estática caberia o controle de registros específicos como alto, baixo ou neutro (XU, 2005). Estas articulações são movimentos em direção aos *pitch targets* (alvos tonais) que seriam as menores unidades melódicas, associadas à funções linguísticas. Os alvos tonais seriam, em última análise, a forma fonológica a ser implementada foneticamente. O ritmo tonal, proposto por Ferreira-Netto (2008), seria um mecanismo de ajustamento fisiológico da citada implementação.

Dado que o Português Brasileiro não possui tons fonológicos, como no caso do chinês, ou tons lexicais, como no inglês (XU; WANG, 2001; PROM-ON; XU; THIPAKON, 2009), assumimos que os alvos tonais do Português Brasileiro se ligam a proeminência tonal (CONSONI, 2011). Em nosso entender, a proeminência tonal é um alvo tonal, o qual está para além dos limiares de diferenciação tonal (T´HART; COLLIER; COHEN, 1990). Dessa maneira, em nosso modelo apresentado no capítulo 2 teríamos um alvo tonal estático, o tom médio (FERREIRA-NETTO, 2008) e dois alvos tonal dinâmicos, as proeminências tonais positiva e negativa. Além disso, um mecanismo de ajustamento dos valores de F_0, o ritmo tonal e um mecanismo de ajustamento perceptivo, os limiares

de diferenciação tonal. Os limiares determinam, neste quadro, o que seria ou não uma proeminência tonal.

4 Metodologia

-Estatística é como gravidade...
Você não pode negociar com a
gravidade.

James "Sonny" Crockett - Miami Vice

No capítulo a seguir faremos uma série de considerações em torno da metodologia do experimento de percepção usado em nosso trabalho. Iremos, para este fim, debater a questão teórica dos experimentos em Linguística, principalmente no que refere-se ao delineamento, à título de ilustrar nossas opções metodológicas. Além dessas questões apresentaremos o nosso método de análise estatística e de organização dos dados. A primeira parte do capítulo se dedica ao debate teórico acima proposto, ao passo que as duas partes subsequentes tratam, respectivamente, da montagem, execução e método de análise estatística usada por nós.

4.1 A questão experimental nos estudos linguísticos

Do ponto de vista estritamente epistemológico, e baseado em seus textos fundadores, a Linguística não é uma ciência experimental [1]. A máxima de Saussure (1977) "O ponto de vista cria o objeto", obviamente coloca em xeque nossa afirmação. Entretanto, uma análise mais criteriosa das ideias do genebrino leva-nos a imaginar que seu gesto inicial previa uma ciência puramente descritiva e explicativa, no sentido de que os

[1] Tomamos como textos fundadores, as obras *Curso de Linguística Geral* (SAUSSURE, 1977) e *Syntactic Structures*(CHOMSKY, 2002), por considerarmo-las as principais obras das duas correntes linguísticas mais estudadas em nosso país.

dados estariam postos *a priori* e o analista, os linguistas, teriam por objetivo revelar as estruturas e a organização por detrás destes dados, no caso, revelar a estrutura de uma língua à partir da própria língua. Nesta perspectiva, o sistema de dicotomias saussureanas (SAUSSURE, 1977) é uma forma de se atomizar o objeto, reduzindo-o a um maior número de categorias analisáveis. A metalinguagem tem papel principal, neste caso, uma vez que é ela o mecanismo de organização destas análises e precisa ser robusta o suficiente sem se esgotar em si mesma: o temido ponto em que somente o sistema linguístico é capaz de explicar a si mesmo, a despeito das suas falhas e de suas características intrínsecas.

As ideias de Chomsky vão além, pois ao propor uma sistema gerativo baseado em um princípio universal (o caso da Gramática Universal), o autor precisou reinventar a metalinguagem, como método de análise. Neste ponto, pode se notar certa aproximação com os conceitos da álgebra linear e na Teoria dos Autômatos (HOPCROFT; MOTWANI; ULLMAN, 2005), seja por meio da Teoria dos Grafos (largamente usada nas árvores sintáticas) ou por meio das matrizes de traços, base para as análises fonológicas do modelo gerativista (CHOMSKY; HALLE, 1968; CHOMSKY, 1965). O analista, segundo a Teoria Gerativa, tem a função de enquadrar fenômenos linguísticos a uma estrutura mental abstrata, tal ligação é demonstrada ou apresentada como resultado por meio dessa nova metalinguagem criada com base na própria análise, entretanto menos dependente das vicissitudes da linguagem em si (SEARLE, 1972). Contudo, foi necessário estabelecer tal metalinguagem como um sistema de oposição e contradição em que os fenômenos são todos categóricos, não havendo espaço para a gradiência ou para a continuidade, como as que ocorrem na fala.

Nos dois casos a função do linguista é explicar uma estrutura subjacente a linguagem, seja por meio da própria linguagem, seja por meio de uma estrutura organizada para este fim. Do lado do estruturalismo, o linguista tem de se tornar capaz de sistematizar todos os fenômenos possíveis, enquanto do lado gerativista sua obrigação se torna a de explicar os mesmos fenômenos dentro de um sistema universal. O fim (e não a

finalidade) de ambas as abordagens é o colecionismo de dados. A Linguística acaba por se tornar um grande exemplário de fenômenos da linguagem e por não explicar o porque de suas ocorrências. A variação linguística ou a variabilidade de elementos linguísticos passa a se tornar uma questão problemática, uma vez que ela vai contra os princípios axiomáticos tomados *a priori*. Willian Labov é o primeiro a ir em outra direção, uma vez que o fator social, além do fator linguístico, tem papel fundamental no que se refere a análise da variação (LABOV, 2010). A ideia central de Labov é simples, porém ousada, pois pressupõe que as análises linguísticas qualitativas sejam feitas com base em um método quantitativo, analisando-se a distribuição dos fenômenos e não as mudanças categóricas, unicamente. Ousada, pois métodos quantitativos tiveram o seu grande apogeu computacional apenas na década de 90, antes deste período a execução de qualquer análise estatística era complexa demais, demandando tempo e não permitindo erros.

A mudança, dentro deste quadro, está na metodologia de análise, de um lado processos categóricos podem ser descritos de maneira sistemática sem que se haja uma preocupação com o uso. Na abordagem sociolinguística, a análise se pauta pelo uso, condicionando o estudo de um determinado fenômeno linguístico a sua ocorrência sistemática dentro de um cenário social. O avanço causado pela Sociolinguística está justamente em propor conjuntos de variáveis (ou fatores) que sejam a base para o estudo de uma dada variabilidade intrínseca a língua, isto é, toda variabilidade está associada a uma variação sistemática de uma dada variável. Tal suposição é fundamental para uma "Linguística Experimental", posto que podemos assumir que não apenas fatos sociais tomam parte na variação linguística, mas elementos de ordem biológica e psicológica também têm suas devidas funções. A ideia de que fatores biológicos ou psicológicos têm papel fundamental na produção de fala e podem ser quantificados em função dos elementos próprios da língua não é nova e tem espaço desde os primeiros trabalhos de Gunnar Fant, como aponta Stevens (2000) na introdução de sua obra *Acoustic Phonetics*. O objetivo de uma *Fonologia de Laboratório* é justamente preencher essa lacuna construindo uma teoria que ligue os aspectos fonéticos (biológicos) aos

aspectos psicológicos/cognitivos (fonológicos) em um programa de pesquisa adequado para a análise da produção e percepção da fala [2].

Nesse sentido, o objetivo de um experimento em Linguística necessita de certos cuidados, principalmente no que concerne ao seu delineamento. Este cuidado se deve para que se tenha claro o que realmente foi analisado ao fim do teste. Um exemplo de delineamento truncado refere-se aos julgamentos de agramaticalidade, conduzidos por sintaticistas gerativistas, o objetivo destes testes é avaliar a competência linguística dos falantes de uma dada língua, contudo, ao se apresentar sentenças agramaticais e requisitar um julgamento, não se estará mais avaliando a competência linguística, mas, sim, a performance individual de cada um. Questões de amostragem, análise quantitativa e interpretação se tornam esvaziadas, dentro do objetivo inicial destas pesquisas. Deste modo, usar variações tênues para detectar variabilidades linguísticas precisa ter um delineamento claro e não enviesado. Na seção a seguir apresentaremos o delineamento de nosso experimento de percepção, tendo em vistas as questões aqui esboçadas.

4.2 A preparação experimental

Esta seção tem por objetivo explicar e apresentar a etapa preparatória, feita antes da aplicação do teste de percepção planejada, explicaremos como foram coletados os estímulos, bem como foram feitas as manipulações a serem utilizadas. Como dito anteriormente o teste de percepção aqui reportado tem por objetivo avaliar os limiares de diferenciação tonal para o Português Brasileiro tem como unidade de análise a sentença. Para esse fim, sentenças terão seus valores F_0 manipulados e serão comparadas com a original não-manipulada, com o intuito de que os falantes possam responder se notam ou não diferenças entre elas. Neste ponto, temos que a variável dependente de nosso teste será a sentença não-manipulada, bem como suas variáveis independentes serão as sentenças cujos valores de F_0 sofreram manipulação.

[2] Albano (2009) discute a mesma questão em seu capítulo introdutório, principalmente sobre a divisão esboçada por Troubetzkoy entre uma ciência "física" (Fonética) e outra "social" (Fonologia)

As manipulações seguiram o espaço de uma oitava, ou seja, serão variadas em 12 semitons (st), sendo seis para cima e seis para baixo. Se tomado um mesmo ponto de F_0 em uma mesma sentença com manipulação de $+6$st e -6st, o valor deve, em Hz, ser o dobro, se comparados. Para o desenvolvimento do trabalho criamos o seguinte roteiro:

1. A primeira etapa consiste na gravação de uma série de frases-piloto, a fim de se escolher uma dentre elas para a elaboração do teste de percepção;
2. a segunda etapa baseia-se na análise linguística, focando o aspecto entoacional da gravação escolhida, com o intuito de se verificar suas propriedades entoacionais;
3. a terceira etapa refere-se a montagem do teste em questão;
4. a quarta e última etapa refere-se a aplicação do teste.

4.2.1 Gravação dos estímulos

A primeira etapa para o desenvolvimento do teste de percepção, seguindo as propostas metodológicas de Consoni et al. (2009), as quais foram baseadas em t´Hart (1981), refere-se a coleta de um *corpus* mínimo para a posterior manipulação das curvas de F_0, as quais serviram de estímulo para o teste de percepção. Como frases-piloto para a manipulação escolhemos as seguintes:

(i) O João mora na casa verde
(ii) O gato está naquele muro
(iii) O mercado, que fica na rua debaixo, fechou
(iv) O professor de Matemática foi viajar
(v) A Maria gosta de quindim.

Todas as orações acima tem como característica a possível produção com um elemento focalizado por meio da entoação. Na primeira frase podemos ter o item

lexical "verde", sendo focalizado, a título de se demarcar que a casa em que se mora não é azul ou vermelha, mas, sim, verde. O mesmo vale para as orações (ii), (iv) e (v) e para os itens "naquele", "matemática" e "quindim". Quanto a oração de número (iii) é notável que a elocução da encaixada, será marcada com duas pausas com relação ao todo da oração, o que necessariamente trará uma diferença de tom, uma possível queda seguida de um aumento de *pitch*, a nosso ver, isso se dá devido a inserção de novas informações na oração, trazidas pela encaixada.

Deste modo, as frases-piloto foram gravadas em série, três vezes cada, sendo usadas apenas as últimas gravações de cada qual. As gravações foram feitas em um ambiente de ruído controlado na Faculdade de Filosofia, Letras e Ciências Humanas da USP, com um microfone de captura unidirecional (ver o equipamento) e utilizando o software Audacity (Versão 1.1.13 Beta). Para nossa gravação escolhemos um sujeito do sexo masculino, no momento de nossa gravação com 22 anos, estudante universitário, brasileiro e falante de português nativo, além de dominar o inglês, não fazer uso profissional da voz e ter declarado ser não-fumante ou ter alguma alteração de voz, tendo preenchido nosso termo de consentimento, bem como o questionário de identificação (ver anexos I e II).

A seguir apresentamos uma análise melhor detalhada das frases-piloto escolhidas, além de uma apresentação de como foi montado nosso teste de percepção. Escolhemos para o teste de percepção as orações (i) e (ii) por questões relacionadas à forma como ela foi entoada pelo nosso sujeito. Principalmente no que se refere a pouca variação de F_0 ao longo da elocução e a não tentativa de se focalizar alguns elementos lexicais das sentenças escolhidas. As durações totais de ambas foram (i) = 5,36s e (ii) = 5,17s

4.2.2 Análise das gravações

A seguir, faremos uma análise quantitativa dos estímulos escolhidos para a manipulação.

4.2.2.1 Análise da sentença *O João mora na casa verde*

A figura 16 apresenta a curvas de *pitch* para a oração o "João mora na casa verde". Note-se que a frase possui uma relativa estabilidade dos valores de F_0, possuindo apenas um pico no item "verde", mas com um decaimento, a seguir, bastante característico, demarcando a finalização. A causa provável para esse pico na sílaba tônica pode ser uma tentativa do falante em focalizar o item, como dito anteriormente. Por outro lado, Ohala (1983) demonstra que é natural haver um certo acréscimo nos valores de F_0 antes que ocorra a finalização, na tentativa de melhor marcar a queda de F_0 ressaltando a finalização.

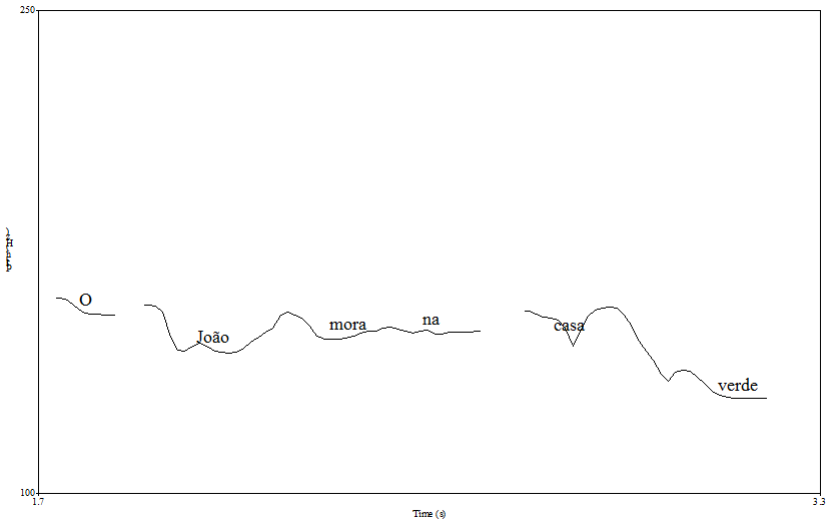

Figura 16: Curva de pitch para a oração "O João mora na casa verde"

A figura 17 apresenta o mesmo gráfico, com auxílio do espectrograma e da forma de onda da sentença.

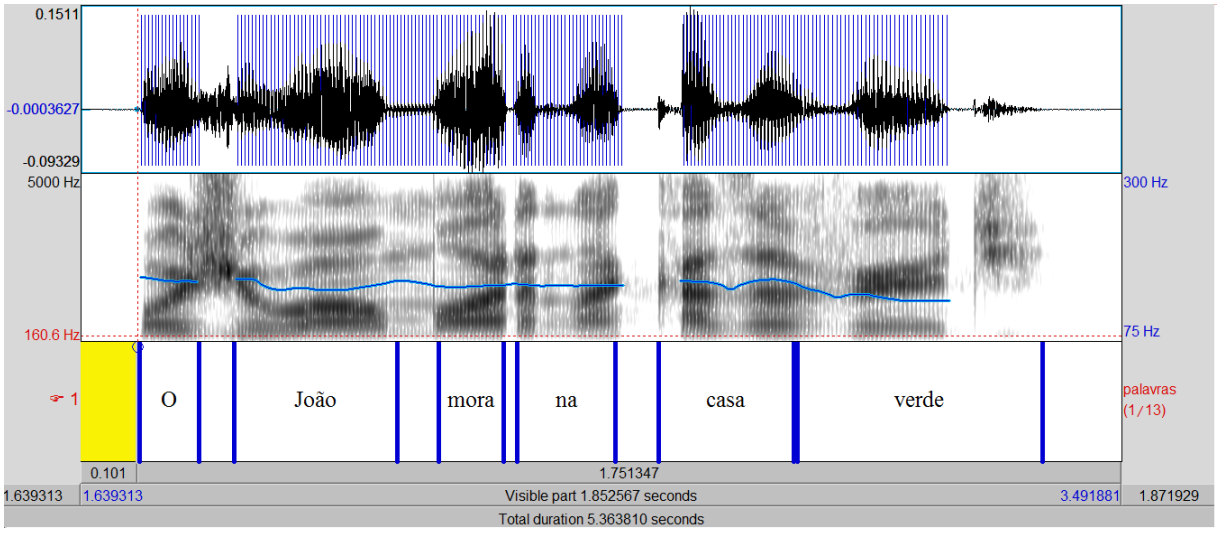

Figura 17: Curva de pitch e espectrograma para a oração "O João mora na casa verde"

A tabela 2 nos dá uma dimensão mais clara dos valores de *pitch* obtidos com a análise da sentença pelo comando *voice report*[3]do *software* Praat.

Pitch mediano	149,82 Hz
Pitch médio:	147,68 Hz
Desvio padrão	8,19 Hz
Pitch mínimo	129,27 Hz
Pitch máximo	160,64 Hz

Tabela 2: Tabela dos valores de pitch para a frase "O João mora na casa verde"

Os dados da tabela 2 também nos deixam claro que a média geral do falante em sua elocução foi relativamente alto (em torno de 150 Hz), em relação ao que se espera para um falante adulto do sexo masculino, cujos valores esperados oscilam em torno dos 80Hz e 120Hz, tendo como média 100 Hz (BEHLAU; PONTES, 1995). Os valores de dispersão dados pelo desvio padrão (8,2Hz) e amplitude[4] (31,4Hz) sugerem uma pequena variação interna de F_0. A figura 18 apresenta a análise em termos de semitons.

[3]Para se utilizar este comando o *pitch* foi analisado por um método de autocorrelação, sendo que os valores de F_0 foram tomados na aba "Pulses" do Praat

[4]Amplitude = *Pitch* máximo - *pitch* mínimo

O valor plotado no gráfico em st é uma média dos valores de F_0 maiores que 0 e convertidos para semitons a este cálculo damos o nome de *média silábica*.

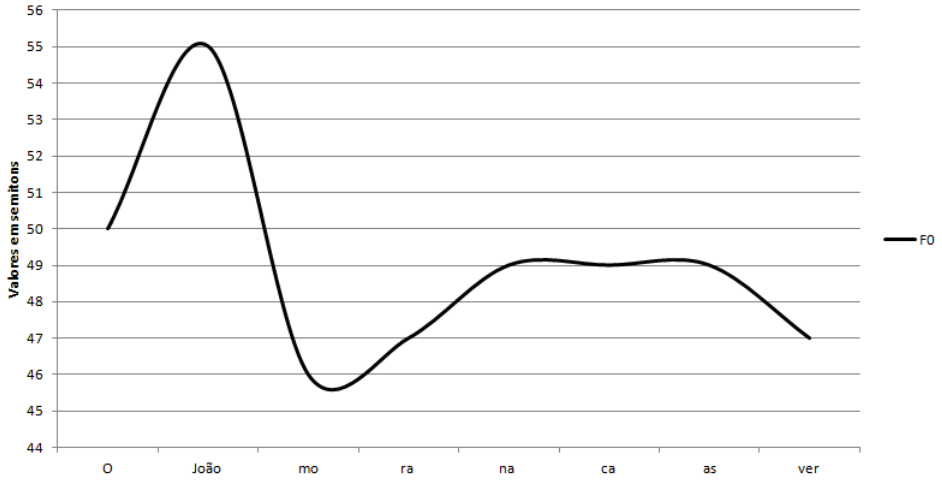

Figura 18: Curva de média silábica, em st, para a oração "O João mora na casa verde", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010)

Pode-se notar que, em termos de média silábica, o item que possui maior valor de F_0 é "João", ao passo que a menor é "mora". Estes resultados aparecem por vicissitudes do programa, o qual detectou de maneira automática os valores de F_0. O vocábulo "João", dada suas características acústicas tem tendência a uma continuidade e maior extensão, o que implica em um aumento da média, ao passo que "mora" não sofre o mesmo efeito. A figura 19 apresenta a média acumuladas, o tom médio dos valores de F_0 tomando as sílabas como referência podemos notar que a variação da média é de apenas 3st.

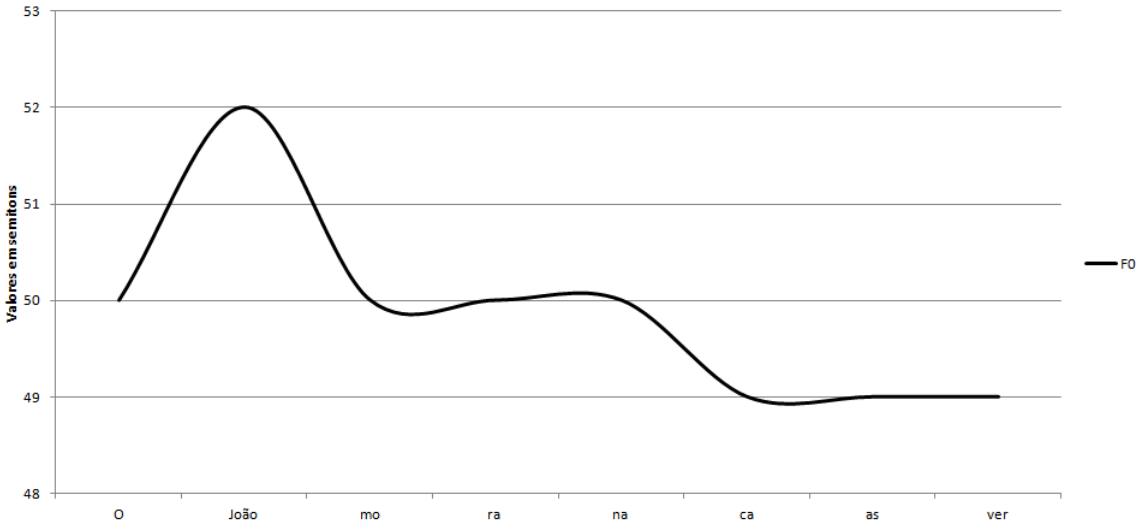

Figura 19: Curva de tom médio, em st, para a oração "O João mora na casa verde", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010)

4.2.2.2 Análise da sentença *O gato está naquele muro*

A oração (ii) foi escolhida por suas características entoacionais, a seguir apresentadas. A primeira delas refere-se a variabilidade dos valores de F_0, como se pode ver na Figura 20 é notável que há nela picos e vales, contudo sem que haja um decaimento grande a ponto de configurar como uma finalização, bem como um crescimento exagerado, de modo que ocorra uma focalização em um determinado item lexical.

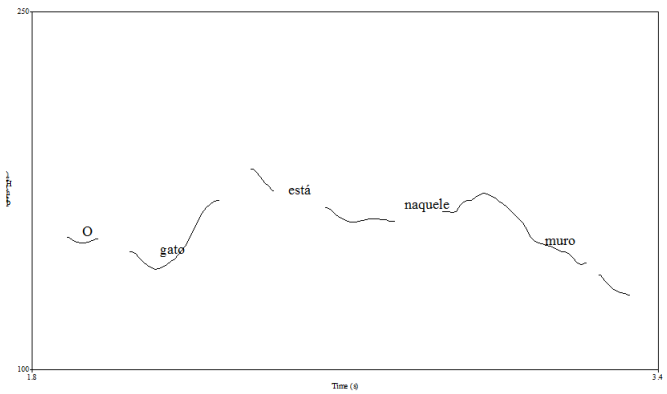

Figura 20: Curva de pitch para a oração "O gato está naquele muro'

Soma-se a essas duas características o fato de se notar um decaimento paulatino dos valores de F_0, a partir do pico mais saliente de *pitch*. Podemos dizer que na

primeira parte, anterior ao item lexical "está", não temos um controle relativo das variações de F_0, sendo que a superfície do sinal nos revela isso, com uma série de descidas e subidas. A segunda parte, iniciada ao fim do item "está", também tem como característica o fato de decair com relação ao tom mais alto. Este decaimento a nosso entender é uma resposta da glote e do mecanismo sub-glotal, ou seja, o falante ao estabelecer um limite máximo passa apenas a controlar o fluxo de ar restante em direção a finalização, representada pelo decaimento paulatino dos valores de F_0.

A figura 21 apresenta o mesmo gráfico, com auxílio do espectrograma e da forma de onda da sentença, como feito anteriormente.

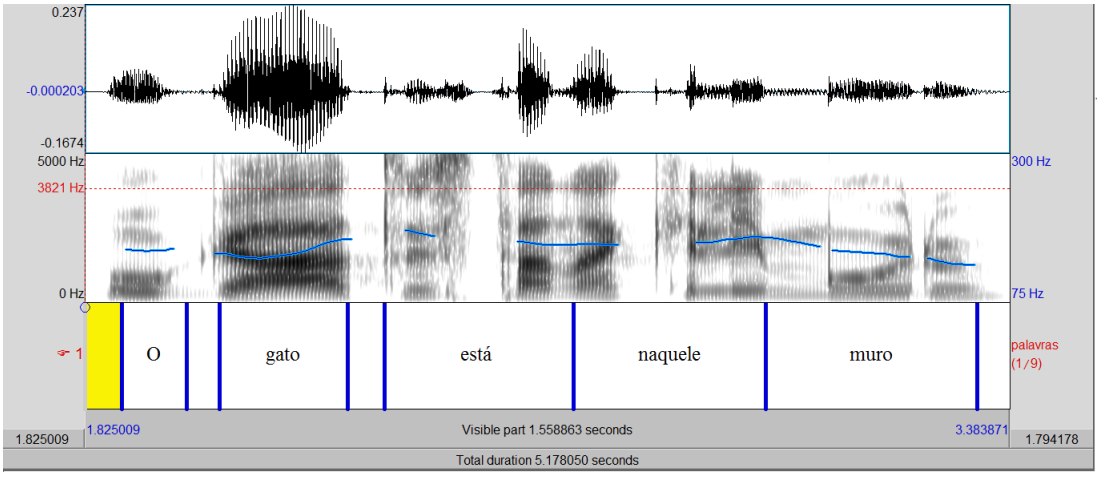

Figura 21: Curva de pitch e espectrograma para a oração "O gato está naquele muro"

A tabela 3 nos dá uma dimensão mais clara dos valores de *pitch* obtidos com a análise da oração pelo comando *voice report* do *software* Praat.

Pitch Mediano	160,35 Hz
Pitch Médio	157,67 Hz
Desvio-Padrão	12,25 Hz
Pitch Mínimo	130,83 Hz
Pitch Máximo	185,94 Hz

Tabela 3: Tabela dos valores de F_0 para a frase "O gato está naquele muro"

Os valores apresentados na tabela 3 nos demonstram se tratar de um registro ligeiramente superior com relação a um registro masculino, bem como apontado na

análise anterior. As medidas de dispersão, desvio-padrão (12,25Hz) e amplitude (55,11 Hz) demonstram que há uma certa variação interna dos valores de F_0, sem que haja grandes sobressaltos na variação. As figuras 22 e 23, apresentam, respectivamente, as análise de média silábica e de tom médio da sentença.

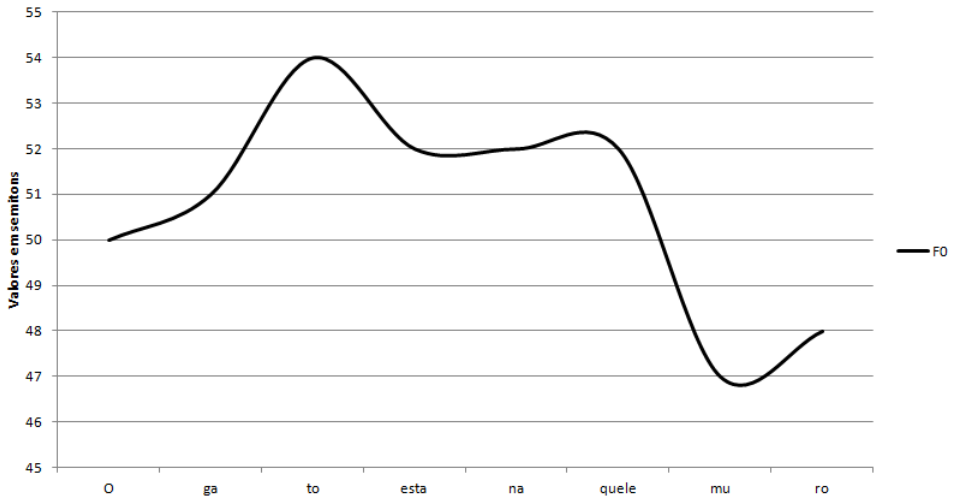

Figura 22: Curva de média silábica, em st, para a oração "O gato está naquele muro", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010)

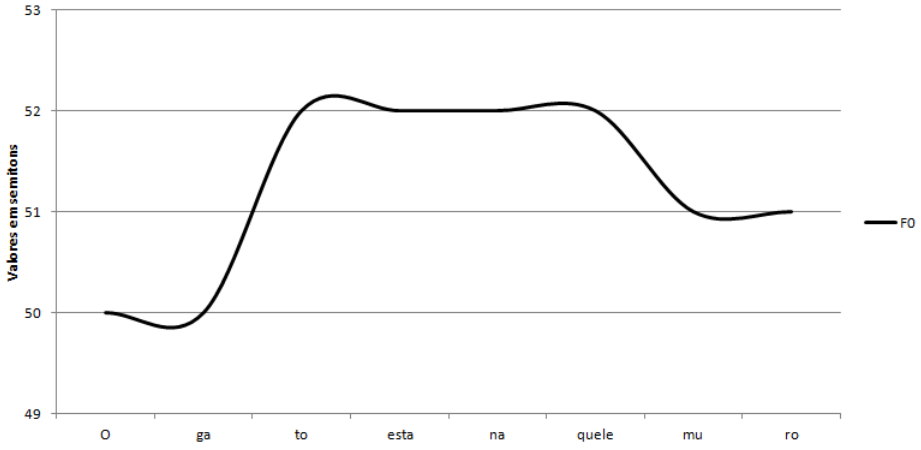

Figura 23: Curva de tom médio, em st, para a oração "O gato está naquele muro", obtidos com a rotina ExProsodia (FERREIRA-NETTO, 2010)

Nota-se na figura 23 que a frase (ii) tem uma menor amplitude de frequências na média acumulada, cerca de 2 semitons, ao passo que a Figura 22 demonstra um amplitude de 7 semitons entre as médias silábicas, além disso as figuras nos revelam a a curva entoacional, a qual segue uma forma de platô, com um decaimento de

finalização ao fim, como esperado.

Estes valores registrados em ambas as gravações nos ajudam no momento da manipulação, de modo que as variáveis manipuladas para cima e para baixo não perderão suas características próprias, isto é, a manipulação não gerará grandes distúrbios nas manipulações a serem obtidas. É importante ressaltar que um registro muito grave ou muito agudo ao serem manipulados soariam falsos ou então já dotados de alguma significação em item com alguma variação relativa de F_0. A seguir, apresentaremos como foram executadas as manipulações, contudo, antes faremos uma breve apreciação em torno da nossa escolha em usar os semitons como forma de medida.

4.2.2.3 Hz *versus* semitons: uma escolha metodológica

Outro importante passo na elaboração de nosso teste de percepção refere-se à forma como serão medidos os valores de frequência fundamental. A forma mais comum e padrão é a mensuração por Hertz, contudo, como pontua t´Hart, Collier e Cohen (1990) a medição de valores em Hertz é uma medida não-relativizada, isto é, trata-se de uma medida linear, onde para cada instante há um único valor de frequência. Por medida linear entenda-se uma forma de se mensurar a distância absoluta entre dois pontos, neste caso está em jogo a diferença entre duas frequência absolutas, quando o verdadeiro interesse está na relação entre estes dois valores. Ao longo dos anos escalas de mensuração vêm sendo propostas, grande parte destas escalas tem como principal característica o fato de serem baseadas em logaritmos, ou seja, são escalas não lineares, mas, sim, escalas de produto, de modo que a relação entre o eixo temporal (x) e o eixo das frequências (y) é não linear, com intervalos que são por vezes múltiplos entre si. Um exemplo destas escalas seria a escala de semitons, largamente usada na música.

Tais escalas, chamadas de "escalas psicoacústicas" são obtidas, em geral, a partir de experimentos de percepção dos sons. Dentre elas, destaca-se a escala conhecida por *mels* e elaborada por Stevens S.S; Volkman; J. & Newman (1937). Tal escala visa medir a "magnitude" do *pitch*, isto é, os intervalos em que um tom pode ser interpretado

como um mesmo tom. Segundo os autores, acima de 500 Hz se tornam maiores e maiores os intervalos de frequência, dentro dos quais os valores de frequência não são distinguíveis. Outras escalas (como barks, ERB-rate) foram propostas na mesma linha de lógica matemática, possuindo certa linearidade para valores até 500Hz e uma distribuição logarítmica para frequências maiores que 600Hz (HERMES; GESTEL, 1991).

Para estudos de percepção de entoação linguística, escalas logarítmicas são mais adequadas do que as log-lineares ou lineares, uma vez que os intervalos entoacionais podem assumir diversos valores de frequência, os quais precisam ser relativizados de alguma maneira. Neste ponto a escala de semitons pode ser aplicada por ter como característica principal o fato de medir duas alturas distintas relativizando-as entre si. t´Hart, Collier e Cohen (1990) propõem a equação a abaixo para se calcular a distância, em semitons, entre duas frequências dadas em Hertz, tendo justamente como pressuposto a mensuração relativizada entre duas frequências

$$D = 12 \cdot \log_2 \frac{f_1}{f_2} = \frac{12}{\log_{10} 2} \cdot \log_{10} \frac{f_1}{f_2} \tag{4.1}$$

Tomemos como exemplo um intervalo de oitava, tendo frequências de 100Hz e 200Hz, aplicada a equação devemos obter com reposta 12 st:

$D = \frac{12}{\log_{10} 2} \cdot \log_{10} \frac{100}{200}$

Onde:

$D = 39,86 \cdot 0,301$

Logo,

$D = 11,9987 \cong 12\text{st}$

Caso aplicássemos a mesma fórmula para valores como 200Hz e 400Hz, ou 400Hz e 800Hz, ainda obteríamos $D = 12$, demonstrando a capacidade da escala de semitons em se relativizar valores com grande variação.

Em Martins e Ferreira-Netto (2010) tem-se

> Em uma escala temperada o intervalo de uma oitava contem 12 intervalos i, os 12 semitons, assim tal valor pode ser obtido pela multiplicação sucessiva do valor i, sendo $i_1 = i_2$, até que este resulte igual a 2:
> $i + i + i + ... \rightarrow 12 \cdot i = i^{12} \rightarrow i^{12} = 2$
> Desta maneira:
> $i^12 = 2 \rightarrow \sqrt[12]{2} = 1,0594631$
> É notável que os valores desta escala estejam em uma progressão geométrica de razão q = 1,0594631. Pierce (1983) estabelece que a primeira frequência desta série seja um lá de 27, 50 Hz, com base na nota central lá de 440Hz, de modo que podemos calcular o valor da primeira nota dó desta escala, a qual estaria por volta de 16,35 Hz, valor em torno dos 20 Hz , estabelecidos como sendo a primeira frequência distinguível pelos seres humanos(RUSSO, 1993).(MARTINS; FERREIRA-NETTO, 2010, p. 290)

A formulação apresentada pelos autores retorna um gráfico não-linear, como o apresentado a seguir na Figura24.

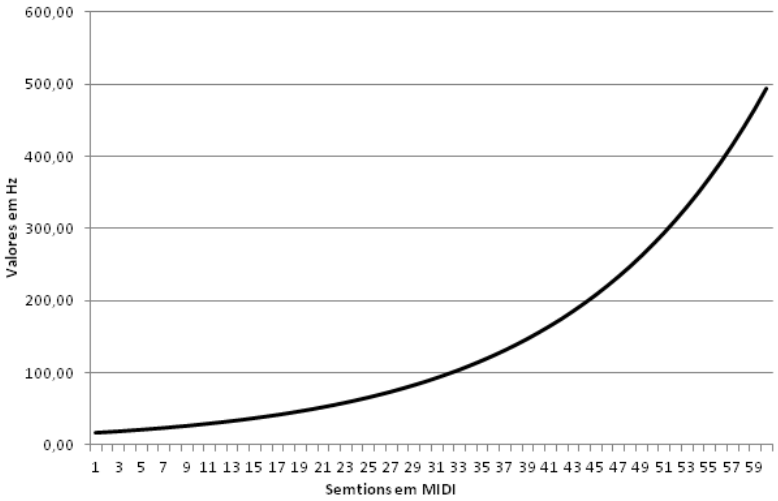

Figura 24: Gráfico de relação entre Hz e semitons plotados em MIDI.

O gráfico demonstra que quanto maiores os valores de frequência maiores serão os intervalos de semitons, o que pode enviesar os resultados, isto é, quanto maiores os valores de Hz, maior terá de ser a faixa de banda para comportar estes valores, contudo, destaca-se que em estudos entoacionais opera-se com escalas de no máximo 500Hz. O que pretendemos demonstrar é nossa preferência metodológica por operar com "bandas de frequência", em detrimento de valores absolutos.

> A explicação para esta postura está nas *j.n.d.* (just notable difference) fenômeno, o qual demonstra que, aproximadamente, 3% de cada

> frequência não será percebida como variação para frequências abaixo de 500 Hz. Essa mesma proporção diminui até, mais ou menos, 2.000 Hz, que parece ser o nível ótimo da audição humana, quando as j.n.d. atingem os valores mínimos de 0,5% (ROEDERER, 2002).(MARTINS; FERREIRA-NETTO, 2010, p. 291)

Em estudos por nós conduzidos obteve-se que a relação entre a mensuração em Hertz e semitons não tem diferenças significativas, isto é, não há perdas sistemáticas na qualidade da análise ao se converter valores de frequência fundamental medidas originalmente em Hertz para a escala de semitons (MARTINS; FERREIRA-NETTO, 2011), de modo que a escolha do método de análise dos dados dependerá unicamente do analista e seus objetivos.

4.2.3 Manipulação da frequência fundamental

O objetivo da manipulação[5] da frequência fundamental é criar diferentes pares de estímulos, a fim de que o participante de nosso teste venha a dizer se detecta ou não diferenças entre elas. As manipulações foram executadas usando o *Praat* e modificadas em termos de semitons. O método usado pelo programa é simples, cada valor de F_0 detectado pelo software sofre uma mudança, de acordo com o que é estabelecido pelo usuário. Essa mudança pode ser linear, isto é, somando-se Hertz, ou pode ser geométrica quando os valores de frequência são multiplicados por alguma constante. No caso de nosso teste usaremos o método geométrico (utilizando os semitons), de maneira que os valores serão multiplicados pela constante acima exposta (0, 059) elevada ao número de semitons escolhidos. Como dito anteriormente, optamos por utilizar o espectro de uma oitava, 12 semitons, de modo que os valores de frequência das sentenças manipuladas com -6 st terão metade da frequência daquelas manipuladas com +6 st.

As manipulações foram executadas usando o comando "To Manipulation" do Praat (BOERSMA; WEENINK, 2009). Os parâmetros de análise estão arrolados na Tabela 4, logo

[5]Os processos manipulação e síntese de fala, bem como seu uso em ciência, é objeto de estudos há diversos anos, principalmente nos Estados Unidos e Japão. Tal fato levou a criação de um departamento dedicado ao tema no Museu Smithsonian em Washington (EUA), além de ser tema de diversos artigos e pesquisa ao longo dos tempos, como apontado em Mattingly (1974).

abaixo:

Time step	0.01 s
Pitch máximo	600 Hz
Pitch mínimo	50 Hz

Tabela 4: Tabela dos valores com os parâmetros utilizados para a manipulação.

A partir destes valores obtemos a janela exposta na Figura 25. As linhas cinza indicam (a mais claras, abaixo) os valores originais de frequência fundamental, ao passo que as linhas verdes (as mais escuras, acima) indicam os novos valores manipulados com mais 6 semitons. Todos os novos estímulos foram manipulados, a partir da gravação original, a fim de que não se enviesassem o teste. O intervalo de semitons foi de 1 a +6 e de -1 a -6, obtendo-se assim um total de 12 manipulações. Após essa etapa de manipulação dos valores de F_0 utilizamos o comando "Publish Resynthesis" na aba "File" da janela do "Manipulation Editor", a seguir, os dados eram salvos no formato .wav, por meio do comando "Save as .wav file" na janela do Praat *Objects*.

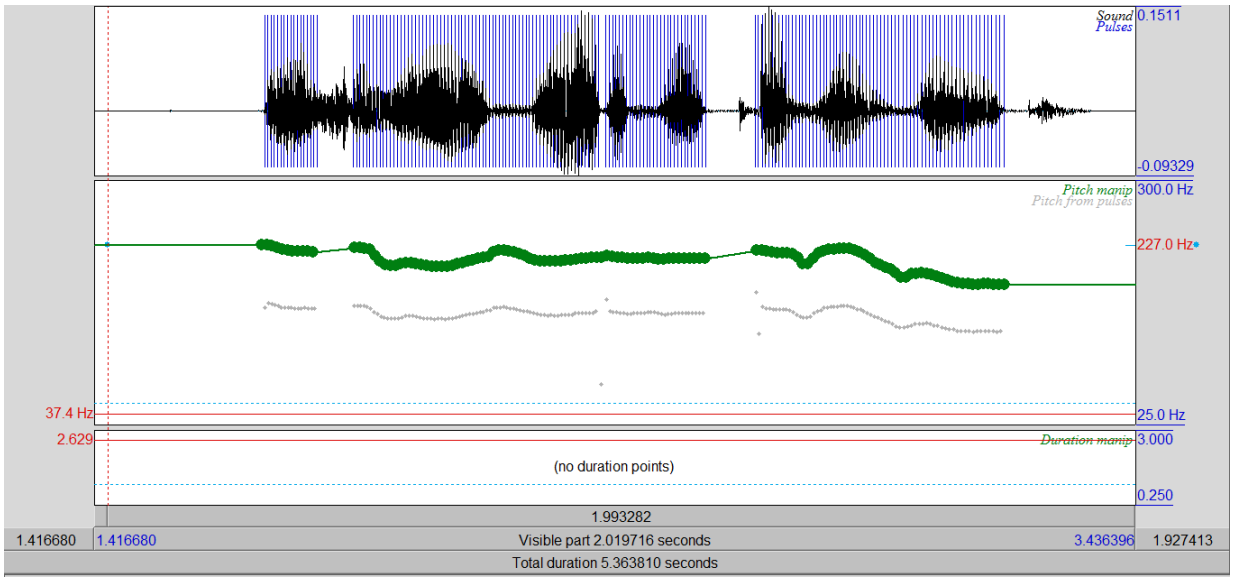

Figura 25: Exemplo de manipulação com mais 6st para a frase "O João mora na casa verde".

A tabela abaixo mostra o *Voice Report* das manipulações de +6 e -6 semitons. Nosso intuito é demonstrar a eficiência do software, tendo em vista que todos os valores de

frequência expostos na coluna da direita (+6st) devem ser, aproximadamente[6], o dobro dos da esquerda (-6st), posto que há uma oitava de diferença entre elas.

	-6 semitons	6 semitons
Pitch médio	104,49 Hz	209, 65 Hz
Pitch máximo	91,47 Hz	182, 97 Hz
Pitch mínimo	112, 60 Hz	225,89 Hz

Tabela 5: Tabela contrastiva dos valores de pitch, após manipulação com o Praat

Os números apresentados na Tabela 5 demonstram que a manipulação foi feita com relativo sucesso, uma vez que os valores apresentados à direita são o dobro quase exato dos valores apresentados à esquerda. Entretanto, a análise numérica dos valores de F_0 manipulados não suplanta uma análise qualitativa. Nesse sentido, foi feito um pequeno teste com um informante falante nativo de Língua Portuguesa, com 21 anos na época, com algum treinamento musical e de computação. A tarefa dirigida a ele foi a de julgar se as manipulações tinham ou não alguma diferença, não no quesito tonalidade, mas, sim no que se refere à qualidade sonora de cada uma delas. Foi dado a ele total liberdade para a utilização de recursos computacionais, bem como aqueles que ele julgasse necessários. Ao fim do teste, o qual tomou por volta de uma hora, o falante dizia estar satisfeito com as manipulações e que não detectava diferenças entre elas. Foram feitos outros testes não sistemáticos com cerca de 8 pessoas, nos quais era apenas requisitado que se julgasse a qualidade de algumas das manipulações. Em todos os casos a resposta foi positiva, fato que, associado aos valores numéricos, demonstra a fidelidade da manipulação.

4.3 O delineamento experimental

Esta seção tem por objetivo apresentar o processo de montagem e delineamento do experimento conduzido.

[6] A imprecisão se deve ao método de detecção dos valores de *pitch* do próprio software, tais desvios são da casa de décimos de Hertz

4.3.1 A montagem do experimento

A montagem do experimento foi desenvolvida em duas etapas, uma primeira em um teste-piloto, o qual determinaria quais seriam as condições de análise e uma segunda em que o teste em si seria aplicado. O teste piloto contava de treze estímulos, ou seja, a combinação de todas as 12 manipulações com a gravação original (chamada aqui de *sentença 0*) mais um par formado de duas sentenças 0. O teste foi aplicado utilizando o *script ExperimentMFC* do Praat. O *script* utilizado pode ser encontrado no Anexo III. O teste consistia de duas partes, uma primeira com um "treino" com 3 pares de estímulo, a fim de que o participante se acostumasse com a interface do software e o uso do *mouse* e uma segunda, em que eram apresentados os estímulos a serem analisados, ambos eram feitos no Praat. O texto de instrução era o seguinte:

Este é um experimento de percepção. A seguir serão apresentados pares de uma mesma oração. Sua tarefa será responder se há ou não diferenças de tom entre elas. Responda SIM, caso note alguma semelhança, responda NÃO, caso não note diferenças.
Clique para continuar.

A seguir era tocado um estímulo e em seguida outro, com um intervalo 1,2 segundos entre ambos. O tempo total entre o início do estímulo 1 e do estímulo 2 era de cerca de 4,5 segundos, ao passo que o tempo medido de reação [7] foi por volta dos 10s. A Figura 26 abaixo apresentada a interface utilizada para o treino, bem como para o teste-piloto de percepção.

[7] O tempo de reação é calculado somando-se o tempo total do par de estímulos e o tempo que o informante demora a escolher uma das opções.

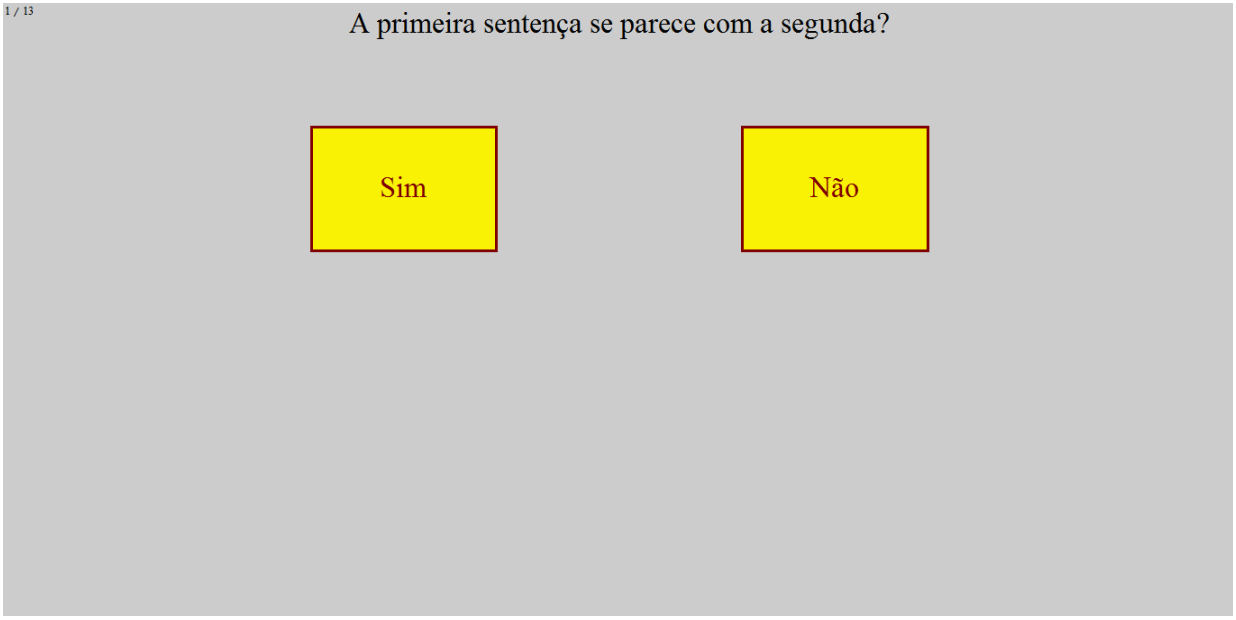

Figura 26: Interface usada para o experimento de percepção

Na fase piloto o teste foi aplicado em cerca de 5 pessoas e a principal dúvida referia-se ao que devia ser julgado: o domínio semântico, sintático ou fonético. Alguns dos participantes buscavam encontrar erros de gramática, enquanto outros buscavam diferenças segmentais mínimas, ou ainda significados diferentes. Em geral, os participantes afirmavam ter se confundido com a instrução ou com a interface do teste, de modo que pediam para refazer o teste, de maneira que nesta segunda vez se atentariam unicamente para as diferenças de entoação. Com base nestes dados desenvolvemos três propostas de testes, as quais estão esquadrinhadas abaixo.

- **Proposta 1**

A proposta 1 tinha por objetivo usar os 13 pares de uma mesma sentença em três condições distintas: sem instrução, com instrução parcial e com instrução completa. Na primeira condição o participante não teria nenhuma informação sobre quais seriam as diferenças, apenas seria alertado de que há diferenças entre os pares. Na segunda condição, haveria uma orientação no sentido de se dizer que há uma diferença de tons e nada mais.Na terceira condição seria explicado que há diferença entre tons, que todo o *pitch* da sentença foi alterado para cima e para baixo, de modo que se deva

notar se há ou não diferenças significativas entre grave e agudo. O teste deveria ser conduzido em três etapas distintas (se planejado dentre participantes) com intervalos de tempo regulares, a fim de não se enviesar a análise. Caso planejado para ser entre participantes seria preciso, no mínimo, 3*n da amostra anterior, além de se supor um enviesamento da análise por conta dos erros de participante (as diferenças individuais).

- **Proposta 2**

A proposta 2 tinha por objetivo criar 3 condições e dividindo a variável dependente em duas, uma primeira com cadeia segmental e outra sem cadeia segmental, obtendo-se, assim, um total de 26 pares. As condições seriam as mesmas que as da proposta anterior. A diferença, nesse caso, seria o uso de um estímulo com a cadeia segmental e outro sem a cadeia segmental. O processo para a coleta desse material, sem cadeia segmental, seria por meio da função *to pitch* do Praat. A questão em desse teste refere-se à mudança das VD's. A alteração delas, muda também o propósito do teste, pois, ao invés de se medir os LDT's do Português Brasileiro, estaríamos avaliando o papel da cadeia segmental na percepção da entoação.

- **Proposta 3**

A proposta 3 tem como pressuposto usar dois estímulos, mas com métodos de procedimento diferentes. No primeiro caso, o participante seria forçado a responder "sim" ou "não", ao passo que no segundo haveria a opção "neutra". A ideia por detrás desse delineamento é diminuir os erros de participante, bem como procurar um número balanceado de respostas. Em geral, os testes do tipo "resposta forçada 2" levam a um valor nebuloso nos pontos de inflexão (pressupostos em +3 e -4 semitons (CONSONI et al., 2009)), ao passo que um teste de *Multiple Forced Choice* com 3 opções leva a apresentação de um ponto neutro em que a faixa de nebulosidade se torna evidente. Nesses pontos nebulosos o participante não sabe, ao certo, se há ou não diferenças.

Ressalta-se que entendemos que a percepção não é categórica nesses casos e, sim, está em uma gradiência (SCHOUTEN, 1987). Em termos matemáticos, O resultado do

teste será assintoticamente positivo e negativo, isto é, as margens serão modalizadas (ou bimodalizada, no modelo geral), contudo, haverá, por pressuposição, um ponto a partir do qual a mudança da percepção começará a ter efeito. Este ponto (o ponto de inflexão) não pode ser definido como fixo e determinado, mas, sim, por um valor de tendência contínuo no seu limite. Ou seja, haverá uma faixa de transição não detectável pela matemática linear das análises estatísticas, a menos que seja feito um estudo unicamente com este propósito. Uma opção "neutra" tenta compensar essa falha.

Propõe-se, portanto dois testes paralelos: em um se apresentará os pares de estímulo com as respostas forçadas 2 (sim ou não) e em outro haverá 3 opções (sim, não e talvez). Os estímulos utilizados serão diferentes para não habituar o participante, eles serão informados completamente sobre a tarefa e terão o tempo que precisarem para cumpri-la (a ideia é medir o tempo que os valores nos pontos de inflexão demoram a ser respondidos). Os resultados do primeiro teste apontaram os valores modais (a partir de qual valor o ouvinte começa a notar diferenças) enquanto o segundo teste, em princípio, responderá sobre qual é o ponto de transição.

Após algumas outras análises dos resultados dos testes-pilotos , optamos pelo uso da proposta 3. Em nossa opinião, tal proposta satisfaz os métodos utilizados nos testes de t´Hart (1981) e Consoni et al. (2009), por usar o método de resposta forçada 2, além de atender a demanda de se analisar os pontos de nebulosidade, como acima explicado.

4.3.2 O experimento

Tomando como base a proposta 3 foram montados dois testes distintos, com 13 pares cada qual. Os fundamentos dos testes: cuidados com o delineamento, as questões a serem feitas, tamanho de amostra e aplicação, foram feitos com base em Rea e Parker (2000), Cozby (2003) e Dancey e Reidy (2006). O teste com duas respostas foi inspirado nos testes chamados de *two-alternative forced choice tasks* e comumente abreviado por 2AFC *tasks*. A lógica destes testes refere-se a capacidade dos indivíduos fazerem

escolhas com base em julgamentos daquilo que consideram ser adequados dado uma determinada tarefa, sendo que destas respostas pode-se analisar a lógica das tomadas de decisão (HILL, 1898; BOGACZ et al., 2006), uma visão completa sobre as variações do referido teste, bem como a matemática utilizada pode ser vista em Zacksenhouse, Bogacz e Holmes (2010).

O teste de percepção com 3 alternativas foi feito inspirado na escala de Ramsay, a qual avalia o nível de sedação em pacientes internados em UTI's com seis níveis de avaliação. Cada nível corresponde a um grau de sedação e eles se opõem internamente (1 se opõe a 6, 2 se opõe a 3, 4 se opõe a 5)e são obtidas a partir de uma avaliação feita com o próprio paciente. Para a desenvolvimento da escala de 3 níveis baseamo-nos na escala Likert (LIKERT, 1932) de cinco níveis. Na escala de Likert os níveis são valorados em um sistema de oposição e contradição, tendo sempre um ponto neutro, abaixo demonstramos um exemplo de escala de Likert.

(1) Não concordo totalmente.

(2) Não concordo parcialmente.

(3) Indiferente.

(4) Concordo parcialmente.

(5) Concordo totalmente.

Os valores de 1 e 2 / 4 e 5 estão em um sistema de oposição, ao passo que os valores de 1 e 4 / 2 e 5 estão em um sistema de contradição, enquanto 3 é neutro em relação a ambos. Optamos em nosso teste por usar uma escala com apenas três níveis, com apenas um eixo de contradição e um nível neutro. Uma descrição mais adequada sobre escalas e os desenvolvimentos e métodos para se criar uma podem ser encontrados em Pereira (2004).

As sentenças eram iguais nos testes, mas diferentes entre testes. O teste com apenas duas respostas seguiu os parâmetros apresentados na descrição do teste-piloto,

tendo uma fase de treino seguida da aplicação em si. O *script* utilizado para a sua montagem e aplicação pode ser visto no Anexo V, enquanto o Anexo IV demonstra o *script* utilizado para a fase de treino. A interface para as respostas é igual aquela apresentada na Figura 26, de modo que foi feita apenas uma modificação no texto de apresentação, apresentado abaixo. A sentença ouvida pelos participantes foi "O João mora na casa verde".

Olá,

Este é um teste de percepção. A seguir serão apresentados pares de uma mesma sentença. Elas tiveram seus tons manipulados, algumas ficaram mais graves e outras mais agudas, ou seja, algumas parecem ser faladas por uma voz mais fina e outra por uma voz mais grossa, embora elas sejam da mesma pessoa. Sua tarefa será dizer se você observa ou não alguma diferença entre os pares.

Vamos fazer um teste antes de começar.

Agora vamos ao teste

A figura 27 demonstra como o texto de apresentação é visto pelos participantes do teste, antes de respondê-lo.

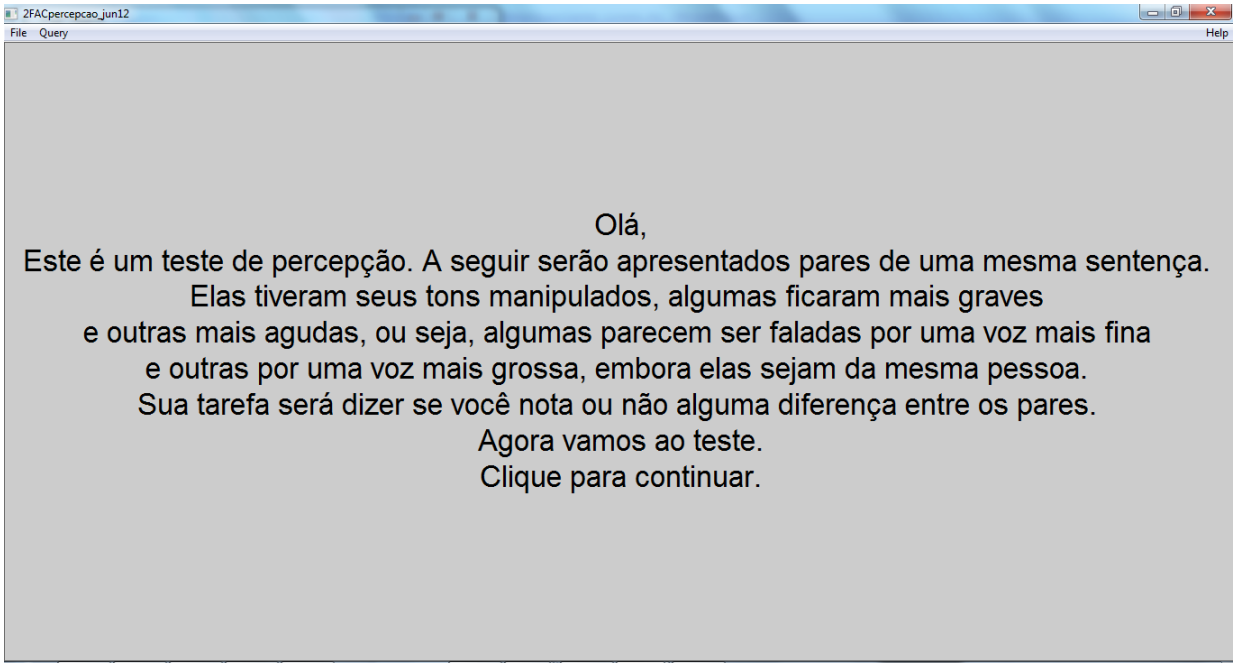

Figura 27: Texto inicial, como visto pelos participantes do teste.

O teste de 3 alternativas seguiu o mesmo procedimento, sofrendo uma alteração no texto de apresentação, como apresentado abaixo e na interface de respostas como demonstrado na Figura 28. A sentença ouvida pelos participantes foi "O gato está naquele muro".

Olá

Este é um teste de percepção. A seguir serão apresentados pares de uma mesma sentença. Elas tiveram seus tons manipulados, algumas ficaram mais graves e outras mais agudas, ou seja, algumas parecem ser faladas por uma voz mais fina e outras por uma voz mais grossa, embora elas sejam da mesma pessoa. Sua tarefa será dizer se você nota ou não alguma diferença entre os pares, se você não tiver certeza se há ou não diferenças, você poderá responder talvez.

Agora vamos ao teste.

Clique para continuar

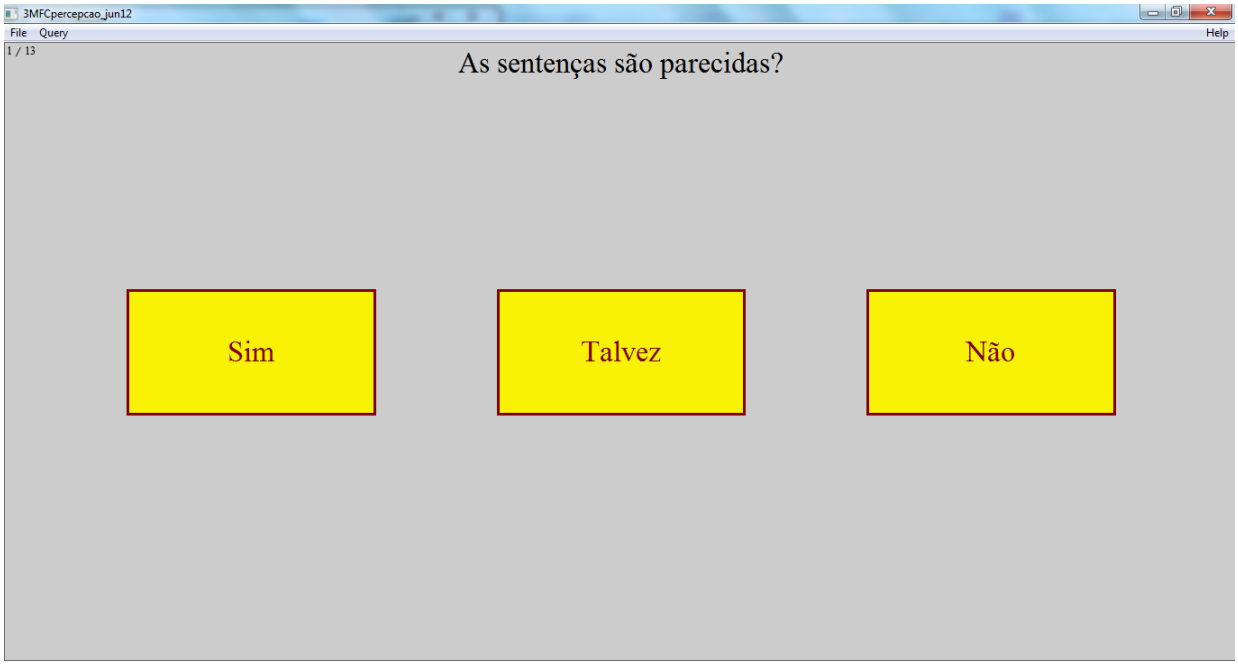

Figura 28: Interface do teste com três alternativas.

Não houve intervalo entre os estímulos, uma vez que havia um silêncio anterior ao início da fala na gravação original não manipulada. O tempo total entre o início do primeiro estímulo e o fim do segundo é da ordem de 1 segundo para os dois testes. Foi escolhida a ordem randômica de apresentação dos pares, isto é, o Praat escolhia quais seriam os pares apresentados de maneira arbitrária, com base em um algoritmo de aleatoriedade intrínseco a ele. O tempo médio total para aplicação do teste, isto é somado o tempo para as respostas, mais o de instrução e o de adaptação ao equipamento pelo participante foi da ordem de 9 minutos. A Tabela 6 resume os valores obtidos na fase de verificação do teste.

	Duração (tempo gasto)
Duração de cada estímulo	5,17s e 5,36s
Duração de cada par de estímulo	10,34s e 10,72s
Duração média total do teste mais treino	9 min
Duração média do tempo de respostas	5 min
Duração média para instrução e treino	4 min

Tabela 6: Tabela com os valores de tempo para cada etapa do teste

Os valores acima resumidos servirão de base para a aplicação do teste e para controle de *outliers*. Por exemplo, tempos de repostas maiores que 15 ou 20 segundos,

demonstram que o participante em algum ponto confundiu os pares a serem analisados, bem como a quantidade de participantes que poderiam ser testados simultaneamente. A seguir, trataremos do tamanho da amostra, ou seja, o número de participantes de nosso teste.

4.3.3 O tamanho da amostra

Questões de amostragem são central em estudos que envolvam propriedades comportamentais, ou que demandem ser replicadas e encontradas sistematicamente em uma dada população. Um exemplo seria a habilidade de dirigir quando alcoolizado, um único sujeito não seria capaz de revelar o quão afetada é a habilidade de dirigir, dado certas quantias de álcool. Por outro lado, toda a população de motoristas de uma localidade não seria, também, capaz de revelar os efeitos do álcool em combinação com o volante. A questão é, um único indivíduo deixa o estudo vulnerável às características individuais exclusivas a ela. Toda a população nos mostrará unicamente um valor central e não uma estimativa próxima do real, uma vez que será avaliada toda sorte de características individuais: abstêmios e alcoólatras estarão no mesmo conjunto de dados a ser analisado.

Obviamente, a situação ideal seria aquela em que se pode entrevistar ou avaliar toda uma população estudada. Contudo, o alargamento ou estreitamento da população ocorrido a cada instante teria de ser levado em conta. O processo de coleta de dados seria como aquele ilustrado no Paradoxo de Zenon [8]. Por esta razão, entre outras, desenvolveu-se ao longo dos anos diversos métodos, a fim de se determinar o número mínimo de sujeitos (geralmente ilustrado pela letra n) a serem utilizados, sejam em entrevista de tendências ou testes populacionais. Os cálculos que determinam o tamanho da população[9], em geral, são complexos ou demandam uma quantidade

[8]Conta a história de uma corrida entre uma tartaruga e Aquiles. Dado que Aquiles era o guerreiro mais veloz de toda a Grécia, ele permitiu que a tartaruga saísse na sua frente. No entanto, ele nunca conseguiria alcançar a tartaruga, uma vez que ao chegar ao ponto A, onde estava a tartaruga, ela teria se descolado para o ponto B, ao chegar ao ponto B, ela estaria no ponto C, assim até o infinito.

[9]Recomendamos a consulta aos documentos da Food and Agriculture Organization (FAO) da ONU, os quais, em geral, lidam com populações animais ou a estimativa de estoques de comida. Disponíveis

extensa de informações para serem feitos (KRAEMER; THIEMANN, 1987). Pode-se dizer que seria necessário um trabalho unicamente voltado para este tipo de cálculo. A fim de sanar este problema e calcular o número de participantes necessários usaremos o *software* G*Power (ERDFELDER; FAUL; BUCHNER, 1996) desenvolvido na Düsseldorf Universität. Os métodos matemáticos utilizados pelo programa podem ser consultados em sua documentação.

Para este fim, foram determinados de antemão o tipo de análise estatística, bem como o modo de delineamento. Assim, foi escolhido como teste de hipótese a ANOVA dentre fatores, isto é, no teste de percepção todos os falantes participariam de todas as condições experimentais. Consideramos, para este fim, os dois testes como sendo dois grupos independentes, embora tivessem o mesmo número de repetições. Consideramos também os testes como sendo dentre sujeitos, de modo que todos os sujeitos participariam dos dois testes. Para um $\alpha = 0,05$ e um $\beta = 0,95$, tendo 1 grupo de análise e 13 repetições (neste caso a avaliação é feita com relação à sentença 0, a qual repete 13 vezes), obteve-se um $n = 17$. O resultado apresentado demonstra a necessidade de apenas dezessete sujeitos para que o teste de hipótese tenha um tamanho de efeito significativo. Neste caso específico, o tamanho de efeito escolhido foi de $f = 0,25$, o que é considerado médio. A Figura 29 apresenta os valores de significância a serem assumidos de acordo com o tamanho da amostra coletada.

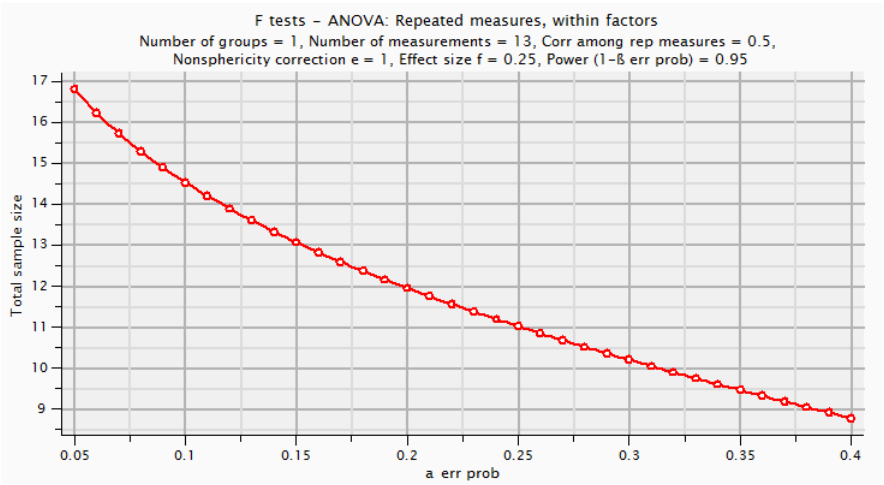

Figura 29: Relação entre o valor de α e o tamanho n da amostra.

em http://www.fao.org/corp/statistics/en/.

A Figura 29 demonstra que para um $\alpha = 0,10$, o que pode ser usado para estudos linguísticos, dados os tamanhos das populações, tem-se que seriam necessários apenas 15 sujeitos, enquanto que para um $\alpha = 0.2$ seriam necessários apenas 12 sujeitos e assim por diante. Vale dizer, portanto, que ao se admitir um menor número de sujeitos, admite-se também uma maior chance de se cometer erros. Ressaltamos que diversos outros tipos de testes e análise, bem como tamanho de amostras foram testados, sendo que o mais adequado é o acima reportado, por duas razões básicas. A primeira refere-se à confiabilidade e aceitação no meio acadêmico da ANOVA e segundo por se tratar de um teste básico sem a necessidade de uma modelagem complexa. A seguir iremos tratar da aplicação dos testes.

4.4 Aplicação dos testes

Os testes foram conduzidos em salas de aula da Faculdade de Filosofia, Letras e Ciências Humanas da USP em horários de baixo ou nenhum movimento, de modo que os ruídos e interferências do ambiente foram minimizados. Foram utilizados um notebook de 14" para a realização da tarefa, um mouse óptico com fio e um fone de ouvido *Sennheiser* HD-280 em cada teste os participantes eram testados isoladamente, isto é, apenas um sujeito respondia ao teste por vez. A aplicação seguia o seguinte roteiro:

1. Instrução oral, em que o informante recebia as orientações do que deveria ser feito pelo analista;

2. Treino para o teste de duas respostas. Em um primeiro momento se recebia a orientação escrita, como a apresentada na Figura 27 e a seguir a janela de respostas como aquela apresentada na Figura 26. Nesta etapa tinha-se apenas 3 pares de estímulos;

3. Teste de duas respostas. O informante respondia ao teste completo com os 13 estímulos;

4. Nova orientação oral, com o intuito de alertar o falante de que agora seriam 3 alternativas;

5. Treino para o teste de três respostas, mudança na orientação escrita, como citado anteriormente e a seguir a janela de respostas como aquela da Figura 28, novamente haviam apenas 3 estímulos.

6. Teste de três respostas. O informante respondia ao teste completo com os 13 estímulos.

A fase de treino tinha por objetivo habituar o participante ao equipamento como um todo. Ao fim do treino era perguntado se havia restado alguma dúvida sobre o teste ou se algum dos periféricos não estava operando de maneira satisfatória. A seguir o teste era aplicado. O teste foi aplicado, até o momento, a 17 pessoas, atendendo o número mínimo de nosso delineamento. No capítulo a seguir trataremos dos resultados e das conclusões parciais a que chegamos.

5 Análise

Neste capítulo trataremos da análise dos dados obtidos no experimento apresentado no capítulo 4. Primeiro, demonstraremos como foi feito o tratamentos dos dados e a seguir como procedemos com as análises estatísticas. As análises aqui apresentadas foram feitas com base em Dancey e Reidy (2006), Magalhães e Lima (2008), Pereira (2004)

5.1 Tratamento dos dados

As respostas dos experimentos apresentados no Capítulo 4, e obtidas pelo software *Praat*, vinham em forma de vetores não-sequenciados e salvos no formato .txt. Em arquivos não-sequenciados (ou sem separação) os dados se organizam de maneira não-lógica e a tentativa de extração pode retornar valores não absolutos (como letras ou sinais de pontuação). Por esta razão, foi preciso desenvolver um modo de recuperar estes dados em forma de planilhas do MSExcel. O próprio *Praat* possui em seu sistema um mecanismo de conversão, o qual foi utilizado, embora com algumas ressalvas, uma vez que ele manteve os dados aleatorizados. Os dados foram organizados em quatro vetores: (i) identificação do participante, (ii) estímulo, (iii) resposta [sim, não ou talvez] e (iv) tempo de resposta. As respostas foram etiquetadas como: 1 - sim, 2 - não e 3 - talvez, ao passo que o tempo de respostas faz referência à soma do tempo do estímulo mais o tempo que os participantes demoraram a responder. Ressaltamos que o teste foi aplicado em 16 pessoas ($n = 16$), sendo que em cada teste havia 13 respostas possíveis. Logo temos um total de repostas igual a 208 ($n' = 208$) (13 respostas x 16 sujeitos).

Definimos:

- Fatores: os 13 pares de sentenças comparados, no caso os 13 semitons em análise de +6 a -6, mais o 0 e 0.
- Fator fixo: o par 0 e 0.
- Variáveis: os demais 12 semitons analisados.

5.2 Primeiras análises

O primeiro passo da análise foi calcular os coeficientes de correlação entre os dados obtidos no teste de duas respostas e no teste de três resposta, ou seja, foram comparadas as respostas obtidas em um e outro experimento, tendo o sujeito com variável independente e os estímulos como variáveis dependentes. Como podemos ver na tabela 7 a média geral dos coeficiente de correlação ficou em torno de $\bar{\rho} = 0,46$, com um coeficiente de variação de Pearson de 0,59, um desvio padrão de 0,27 e uma amplitude de 1,14 (maior $\rho = 0,82$ - menor $\rho = -0,32$). Estes dados indicam que não há homogeneidade clara entre os resultados obtidos nos dois experimentos, ou seja, é preciso investigar os dois conjuntos de dados como conjuntos não correlacionados, embora eles tenham alguma relação, haja vista, o valor de maior $\rho = 0,82$, indicando uma correlação forte.

1	0,78
2	0,27
3	0,61
4	0,35
5	0,82
6	0,35
7	0,76
8	0,29
9	0,33
10	0,43
11	0,41
12	0,42
13	0,68
14	0,53
15	0,64
16	-0,32
Média	0,46
Desvio-Padrão	0,27
Coef. De Variação	0,59
Amplitude	1,14

Tabela 7: Tabela com os valores de correlação entre as respostas para o teste com duas e três respostas

Após esta etapa de análise primária os dados dos 16 participantes foram transpostos para uma matriz, a qual tem as linhas como sujeito e as colunas como as treze variáveis estudadas. Obtivemos, então, quatro matrizes distintas, uma com dados do teste de duas respostas, outra com os resultados dos teste de três respostas e outras duas com os valores do tempo de resposta. O segundo passo, após a montagem das matrizes, foi verificar a distribuição dos dados. Para este fim foram montados os gráficos de

distribuição apresentados nas Figuras 30 e 31. Os gráficos indicam, respectivamente, a quantidade de repostas “sim” e “não” obtidas para cada variável. Lembramos que a pergunta de nosso teste era “As sentenças são parecidas?”. Deste modo, a resposta “sim” indica que as sentenças são parecidas ou iguais, ao passo que a resposta “não” representa o contrário, que há diferenças entre os pares. Note-se que obtivemos um número maior de resposta “não” (n=140) e um menor número de respostas “sim” (n=68).

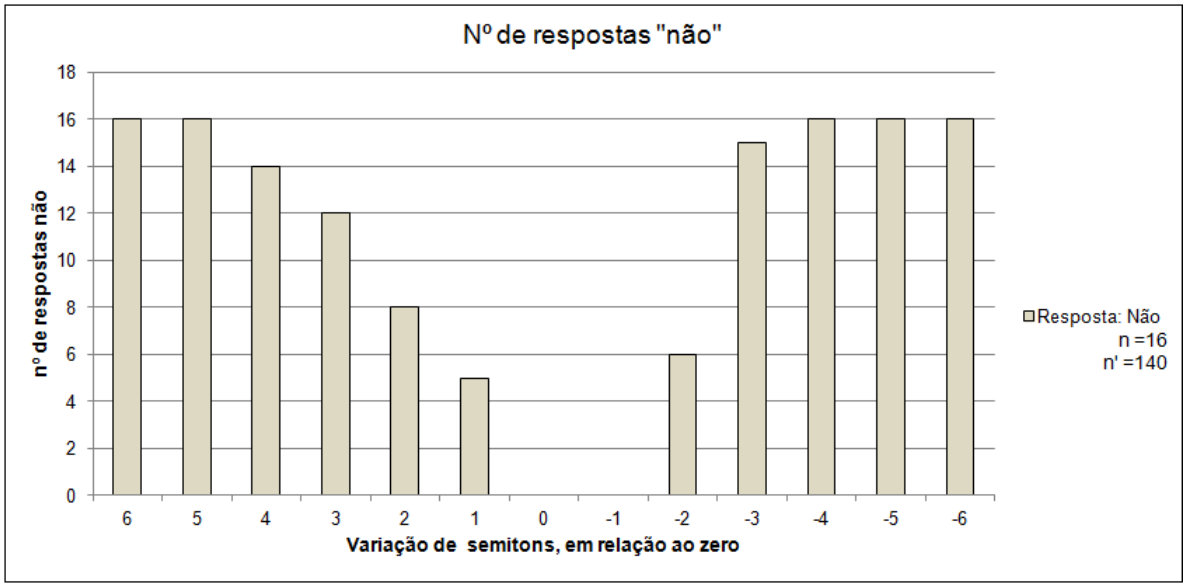

Figura 30: Número de repostas “não” obtidas no teste 2AFC.

Como se pode ver na Figura 30 a distribuição dos dados para a as resposta “não” é bimodal e simétrica, ou ainda multimodal simétrica. Em suma, a distribuição possui ao menos duas modas distintas, porém simétricas. Neste tipo de distribuição temos um centro com um vale bem denotado, ao passo que vemos picos em suas duas extremidades. Estes picos apontam para uma distribuição crescente em direção a um *plateau* de estabilidade, a partir do qual os testes apontam padrão um padrão homogêneo de respostas.

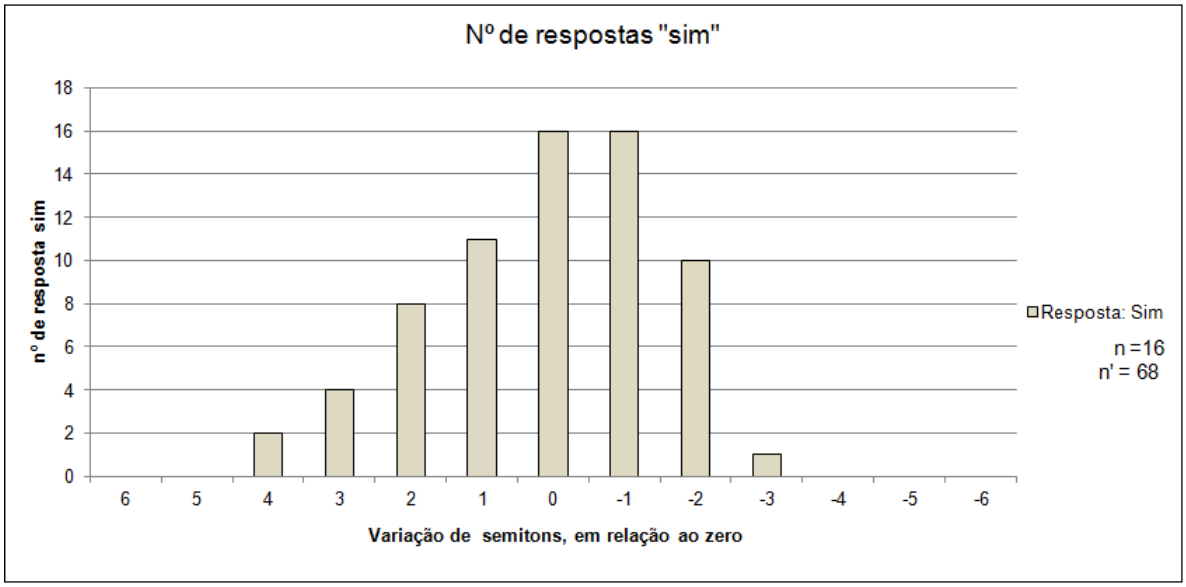

Figura 31: Número de repostas "sim" obtidas no teste 2AFC para cada variável

A Figura 31 apresenta uma distribuição normalizada, embora não simétrica das respostas para cada variável. A rigor, deve-se dizer que a distribuição neste caso é positivamente assimétrica. Em ambas as distribuições temos um valor central relativamente estável, o qual vai decrescendo em direção ao infinito, sem que se toque o eixo das abcissas. Contudo este tipo de curva é apenas uma idealização e verificável apenas em formulações teóricas. Esta segunda curva, apresentada na Figura 31, permite-nos aplicar alguns testes de hipóteses, com a ressalva de que a normalidade não seja completamente assumida na aplicação dos citados testes. Cremos que os valores de curtose e assimetria não tenham relevância em nossa análise, uma vez que todos eles ficaram em torno de 0,8 demonstrando que as curvas são simétricas e mesocúrticas[1].

5.3 Análise de dados: testes de hipóteses

Executamos uma ANOVA entre todas as variáveis de análise para $n = 16$, como previsto pelo nosso dimensionamento tratado na última seção do Capítulo 4. A

[1]Distribuições mesocúrticas são aquelas em que se tem um padrão de achatamento como o da curva normal, as outras formas são: *platicúrtica*(em forma de platô) ou *leptocúrtica* (em forma de um funil invertido), sendo a mesocúrtica um entremeio das outras duas.

hipótese nula assumida é a de que não há diferenças entre as respostas dadas para cada um dos tons variados. Obtivemos $p < 0,001$, ($F = 31,71$, para um F crítico = 1,84), para $\alpha = 5\%$ e n = 16, gl=12 tomando como base uma ANOVA sem repetição e de múltiplos fatores dentre sujeitos. Tal *p-value* permite-nos rejeitar a hipótese nula e supor H_1, uma vez que ele é estatisticamente significativo, temos também que o valor F ($F = 33,07$) é maior que seu valor crítico (F crítico = 1,81), o que corrobora a hipótese alternativa, dizendo que há diferenças entre as variáveis selecionadas.Os cálculos foram executados com o software MsExcel.

A Tabela 8 apresenta os valores de variância para as treze variáveis levadas em consideração em nosso trabalho

	1	**2**	**3**	**4**	**5**	**6**
	0,23	0,27	0,2	0,12	0	0
0	**-1**	**-2**	**-3**	**-4**	**-5**	**-6**
0	0	0,25	0,0625	0	0	0

Tabela 8: Valores de variância para as 13 variáveis em análise

A partir da tabela 8 podemos sugerir a sistematização dos dados em dois grupos distintos de variáveis: um grupo de variáveis *variável* e outro de variáveis *invariável*. No primeiro, temos que a variância é maior que zero $(Var(X) > 0)$ enquanto no segundo temos que a variância é igual a zero $(Var(X) = 0)$. As variáveis, portanto, podem ser organizadas da maneira apresentada na Tabela 9:

Condição	Variáveis
Invariáveis	6 ,5, -1, -4, -5 e -6
Variáveis	4, 3, 2, 1, -2 e -3
Fixo	0

Tabela 9: Agrupamento de variáveis com base na variância

Esta organização visa unicamente sistematizar os dados para uma análise explo-

ratória melhor elaborada. Desta maneira, nossas variáveis foram organizados em três categorias: (i) *variáveis*, (ii) *invariáveis* e (iii) *fixo*. O fator fixo, não citado acima, representa o grupo de controle. Assim, tal fator é um grupo de referência, a partir do qual podemos sistematizar análises contrastivas. Além da organização dos dados, pensar na variância como critério de organização permite-nos um melhor refinamento analítico, pois alguns testes de hipótese, como a própria ANOVA ou o *teste-t* tendem a desconsiderar aquelas variáveis cuja variância é igual à zero por uma questão matemática. Uma vez que os computadores não identificam os valores binários das respostas (sim ou não) e, sim, apenas o valor absoluto da variância preferimos analisar aqueles cuja variância tenha sido maior que zero $(Var(x) > 0)$.

Prosseguindo em nossa análise, coube comparar, por meio de *testes-t*, o grupo variável com o fator fixo zero (o par amostral 0 e 0). Os valores de significância encontrados na citada comparação encontram-se resumidos na tabela 10, logo abaixo:

	t	gl	Sig. (Bi-caudal)
mais3	15,652	15	,000
mais4	21,958	15	,000
mais2	11,619	15	,000
mais1	10,967	15	,000
menos2	11,000	15	,000
menos3	31,000	15	,000

Tabela 10: Valores de significância do teste-t comparando-se o grupo variável ao fator fixo.

Podemos notar que em nenhum dos casos obteve-se $p > 0,05$, em verdade, todos os valores de significância são da ordem de $p < 0,001$, para $n = 16$. Foi também aplicado um teste-t entre o fator fixo e os grupo variável utilizando a técnica de re-amostragem *bootstraping* [2] para *n=10000*. Os resultados de *p-value* obtidos foram os mesmos que

[2] A técnica de *bootstraping* consiste em reorganizar aleatoriamente, com base em um método de amostragem aleatória simples, um certo conjunto de dados. Em tese, o conjunto re-amostrado mantem

aqueles em uma amostra de $n = 16$. A partir destes testes-t podemos rejeitar a hipótese nula de não-diferença entre o grupo variável e o fator fixo. Podemos também assumir que o grupo invariável é semelhante ao do fator fixo, posto que as variâncias são iguais a zero, levando o *p-value* a ser maior que 1 ou menor que 0, fato que é impossível, sendo incalculável para este grupo.

Executamos uma nova ANOVA levando em conta apenas as seis variáveis do grupo variável. Obtivemos um $p - value < 0,001$,($F = 9,69$, para um *F crítico* = 2,33), tendo $\alpha = 5\%$ e $n = 16$, $gl = 12$. Com base nestes números podemos, novamente, excluir a hipótese nula de não- diferenças entre grupos e assumir que haja uma sub-organização entre os elementos que compõem o grupo variável. A fim de tratar da questão da subsistematização, fizemos uma análise dos intervalos de confiança, apresentados na Figura 32, logo abaixo.

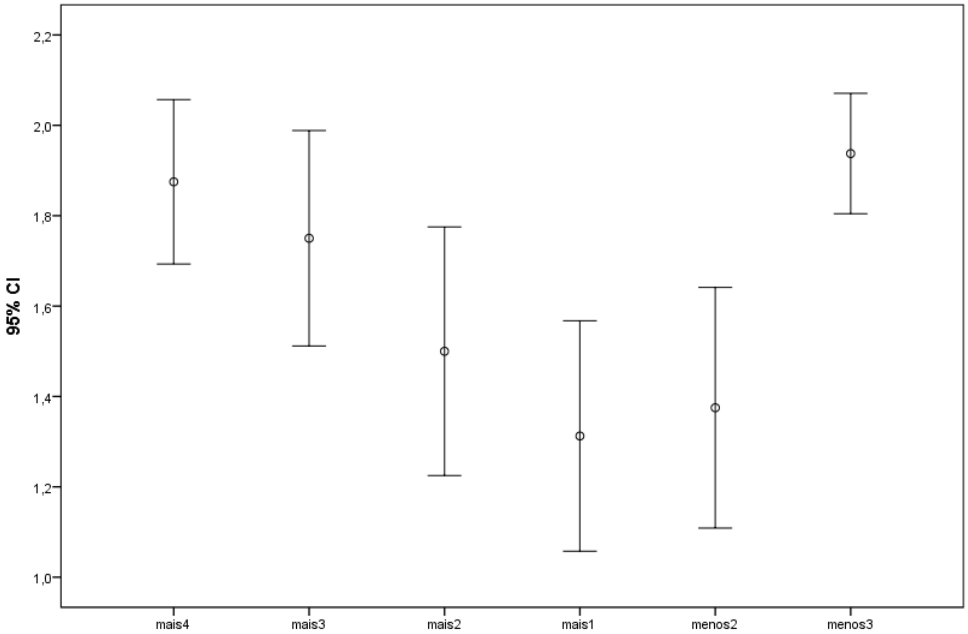

Figura 32: Intervalos de confiança para o grupo variável de variáveis.

Intervalos de confiança nos fornecem um determinado intervalo de valores dentro do qual podemos, com certa segurança, encontrar a média amostral. No caso deste

as mesmas características de distribuição do conjunto de dados iniciais, uma vez que eles são amostras aleatórias de uma dada população.

trabalho, a média central seria de 1,5, uma vez que as respostas analisadas estão opostas entre 1 e 2. Por exemplo, a média de um grupo de "sim" (1) absoluto com outro, o qual teve "não" (2)absoluto é de 1,5. Neste sentido, intervalos com centro (o pequeno círculo na barra de erros) em, aproximadamente, 1,5 podem ser entendidos como elementos de confusão, ao passo que aqueles que possuem seu centro deslocado de 1,5 tem uma maior tendência ser quasi-categóricos. Quando o centro da barra de erros for maior que 1,5 e próximo a 2 é possível assumir que eles tendem a ser compreendidos como diferentes, por outro lado, quanto mais próximos de 1 mais as variáveis tenderão a ser compreendidas como semelhantes. O tamanho da barra de erros tem importante papel, quanto mais largo o intervalo, maior é a probabilidade de se cometer e vice-versa.

Nesse sentido, o gráfico da Figura 32 permite-nos dizer que as variáveis mais4 e menos3 são *quasi-categóricas por diferença*, uma vez que possuem a média próxima a 2 e têm comprimentos intervalares pequenos, em relação ao demais, i.e., a probabilidade de se confundi-los com o zero é menor. As variáveis mais2 e mais3 podem ser analisados como *variáveis de confusão*, pois possuem seus centros próximos a 1,5 e dimensões intervalares mais extensas que as variáveis mais4 e menos3, ou seja, a probabilidade de associá-las com o zero é maior. As variáveis mais1 e menos2 também podem ser considerados *quasi-categóricas por semelhança*, dado a proximidade do centro com 1 e pelo tamanho dos intervalos. Neste caso, temos que a probabilidade de se associá-las com o zero é maior do que se comparado com as variáveis de confusão. O grupo de variáveis invariáveis pode ser considerado *categórico por semelhança* e *categórico por diferença*, posto que a probabilidade de se associá-las ou diferenciá-las do zero é igual 0%. O diagrama de Venn apresentado na Figura 33 resume o que apresentamos até o momento.

Figura 33: Diagrama de Venn ilustrando as categorias de cada uma das variáveis

As variáveis ficam organizadas da seguinte maneira:

- Variáveis categóricas por diferenças: mais6, mais5, menos4, menos5, menos6;
- Variáveis categóricas por semelhança: zero, menos1;
- Variáveis de confusão: mais 2 e mais 3;
- Variáveis quasi-categóricas por semelhança: mais1 e menos2;
- Variáveis quasi-categóricas por diferença: mais4 e menos3;

5.4 Análise fatorial

Para melhor definir os elementos pertencentes a cada uma das categorias expostas na Figura 33 executamos uma análise de fatores. A análise de fatores é um método de estatística exploratória que nos permite agrupar fatores com base em seus valores

de correlação. Suponha-se que temos três vetores de respostas para três variáveis distintas, um deles contêm a avaliação de 0 a 10 de uma determinada marca de batom, um segundo contêm a avaliação de 0 a 10 para uma determinada marca de sapatos masculinos e a terceira contêm a avaliação de um determinado modelo de carro de luxo. O teste com estas perguntas foi aplicado a um grupo de pessoas não estratificado, isto é, pode estar na amostra qualquer sujeito capaz de responder as perguntas propostas, não sendo fatores sociais a base para a construção da amostra.

O teste de fatores usa o método de correlação de Pearson entre as variáveis para determinar se tais variáveis são pareadas ou não, quando pareadas tem-se um fator. Por exemplo, uma forte correlação entre as respostas dadas para a marca de sapato e as respostas dadas para o carro indicam que elas formam um fator, como "produtos de luxo". Uma baixa correlação não permite o agrupamento de variáveis, enquanto uma correlação negativa permite um agrupamento contrário. No caso dado uma correlação negativa entre "batom" e "sapato" pode indicar que trata-se de um "produto masculino" vs. um "produto feminino", ou uma "peça de vestuário" vs. um "cosmético". Vale ressaltar que o método de análise de fatores retorna apenas o agrupamento de variáveis como *fator1, fator2 ... fator n*, sendo que cabe ao analista atribuir a eles nomes que coincidam com a natureza das variáveis.

Aplicamos aos nossos dados uma análise de fatores, contudo aqueles dados que havíamos agrupados como "invariáveis" não podem ser analisados por este tipo de teste, uma vez que suas variâncias são iguais a zero. Assim, apenas os dados do grupo "variável" foi analisado. Usamos para essa análise a extração de dados pelo método de análise de componentes principais e o método de rotação de matrizes denominado *Varimax*.

A Tabela 11 apresenta os valores de correlação entre as variáveis de grupo variável. Note-se que os valores são espelhados, a partir da diagonal.

Correlation Matrix							
		mais4	mais3	mais2	mais1	menos2	menos3
Correlation	mais4	1,000	,655	,378	,255	,293	,683
	mais3	,655	1,000	,577	,389	,447	,447
	mais2	,378	,577	1,000	,405	,516	,258
	mais1	,255	,389	,405	1,000	,313	,174
	menos2	,293	,447	,516	,313	1,000	,200
	menos3	,683	,447	,258	,174	,200	1,000

Tabela 11: Tabela de correlação entre as variáveis do grupo com variância maior que zero

Os valores de correlação apresentados na Tabela 11 apenas indicam o quão ligadas são as respostas dadas em cada uma das variáveis. Note-se, por exemplo, que na primeira linha das correlações a variável "mais4" tem uma tendência a ser mais correlacionada a suas variáveis próximas (como o mais3) e simétricas (menos3) e menos correlacionada com as variáveis centrais. No caso de uma variável central, como o mais1 vemos a tendência a se correlacionar com suas variáveis mais próximas e também centrais, ao passo que possui uma menor correlação com aquelas que estão nas extremidades (mais4 e menos3).

O próximo passo foi agrupar os valores de correlação em fatores, para isso executamos a rotação da matriz apresentada na Tabela 11. Para o processo de rotação de matrizes pelo processo *varimax* seguiremos Kaiser (1958). A vantagem de uma análise usando uma matriz rotacionada é a de se ter uma maior simplicidade na leitura da união de variáveis. No caso do método *varimax*, por exemplo, tende-se a maximizar as máximas semelhanças e a minimizar as mínimas semelhanças, obtendo assim uma separação mais clara das correlações entre variáveis. A Tabela 12 apresenta os grupos de fatores para aquelas variáveis que possuem variância maior que zero, uma vez que, como discutido acima, o teste não analisa variâncias iguais a zero.

Rotated Component Matrix		
	Componentes	
	1	2
mais4		,887
mais3	,596	,608
mais2	,797	
mais1	,700	
menos2	,765	
menos3		,897

Tabela 12: Matriz rotacionada e cargas fatorais

Os números da tabela 12 acima são chamados de *cargas fatorais* e estão relacionados aos valores de correlação, sendo um valor obtido por meio do cálculo da correlação do valor de correlação (cf. Tabela 11) com o fator. As cargas fatoriais ao quadrado explicam o quanto da variância pode ser explicada unicamente por aquele fator, deste modo quanto maior a carga de uma variável, maior será a variância por ela explicada (HAIR et al., 2009). Utilizamos para nossa análise 0,3 como valor mínimo de carga para agrupamento de variáveis, o valor deste número é uma opção semi-aleatória, uma vez que acreditamos que cargas em torno de 0,7 podem ser consideradas fortes para uma análise linguística. Deste modo estabelecer um piso de 0,3 nos dá alguma dimensão de análise, embora faltem estudos que corroborem nossa escolha numérica. Os espaços em branco representam que as cargas fatoriais naqueles pontos eram menor que 0,3.

Podemos notar que a análise fatorial nos retornou duas componentes distintas, chamaremos a estas componentes de Fator 1 e Fator 2. O Fator 1 tem como característica o fato de agrupar os valores mais centrais ou próximos de zero ao seu redor, tais como mais1, menos2 e mais2 e pode ser entendido como "o fator magnético de aproximação". Por outro lado, o Fator 2 tem uma maior dispersão e reúne em seu entorno as variáveis localizadas mais próximas das extremidades, mais4 e menos3. Como pode se ver na

Tabela 12 a variável mais3 possui dois valores de carga fatorial, tendo apenas 0,012 de diferença entre uma e outra carga. Podemos considerá-la parte dos dois fatores, isto é, ela seria uma "variável de confusão", utilizando a ideia apresentada na figura 33. Podemos nomear o Fator 1 de "Variáveis quasi-categóricas", enquanto o Fator 2 pode ser considerado "os limiares de diferenciação tonal". No capítulo a seguir retomaremos esta questão em nossas conclusões.

5.5 Síntese dos testes aplicados

Neste capítulo tratamos da análise dos dados para um de nossos experimentos, identificado por 2Alternative Forced Choice. Estruturamos nossa análise da seguinte maneira:

- Em um primeiro instante executamos uma ANOVA entre todos valores e participantes, a fim de determinarmos se havia ou não diferenças significativas entre as 13 variáveis escolhidas para a análise. Obtivemos um resultado positivo, com $p < 0,001$;

- A segunda parte da análise consistiu em organizar os dados em grupos de análise. Tais grupos foram organizados com base nos valores de variância: para $Var(x) = 0$ temos o grupo *invariável*, para $Var(X) > 0$ temos o grupo *variável* e o grupo de referência, chamado de *fator fixo* ou zero.

- Embora a organização em grupos baseados na variação seja uma medida positiva de análise, ela ainda não é capaz de traduzir com exatidão o fenômeno estudado. Por esta razão, reorganizamos as variáveis como apresentado na figura 33, utilizando-se das barras de erros, de ANOVA's e testes-t entre os elementos do grupo variável para determinar o grau de diferenças entre os seus elementos;

- A parte final de nossa análise consistiu em executar uma análise de fatores, baseada no método de extração de componentes principais para que pudésse-

mos determinar de maneira mais clara como as variáveis semi-categóricas e de confusão estão organizadas.

5.6 Análise dos dados do experimento 3Multiple Forced Choice

O teste com três alternativa visa avaliar os pontos de nebulosos, aqueles em que os falantes não sabe ao certo se há ou não diferenças significativas. Como demonstrado na tabela 7 os testes, aparentemente tiveram respostas parecidas, dado o valor de correlação. As figuras 34 e 35. As siglas n', nt, ns e nn significam, respectivamente, número de participantes no teste, número de respostas talvez, número de respostas sim e número de respostas não.

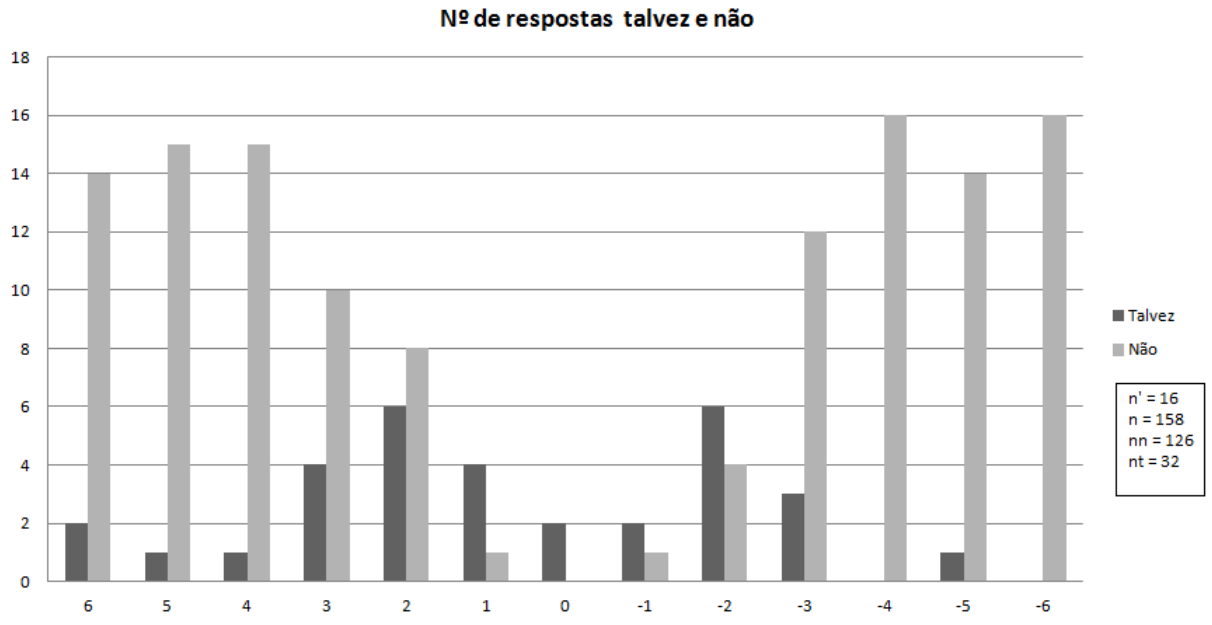

Figura 34: Número de repostas "não" e "talvez" obtidas no teste 3MFC.

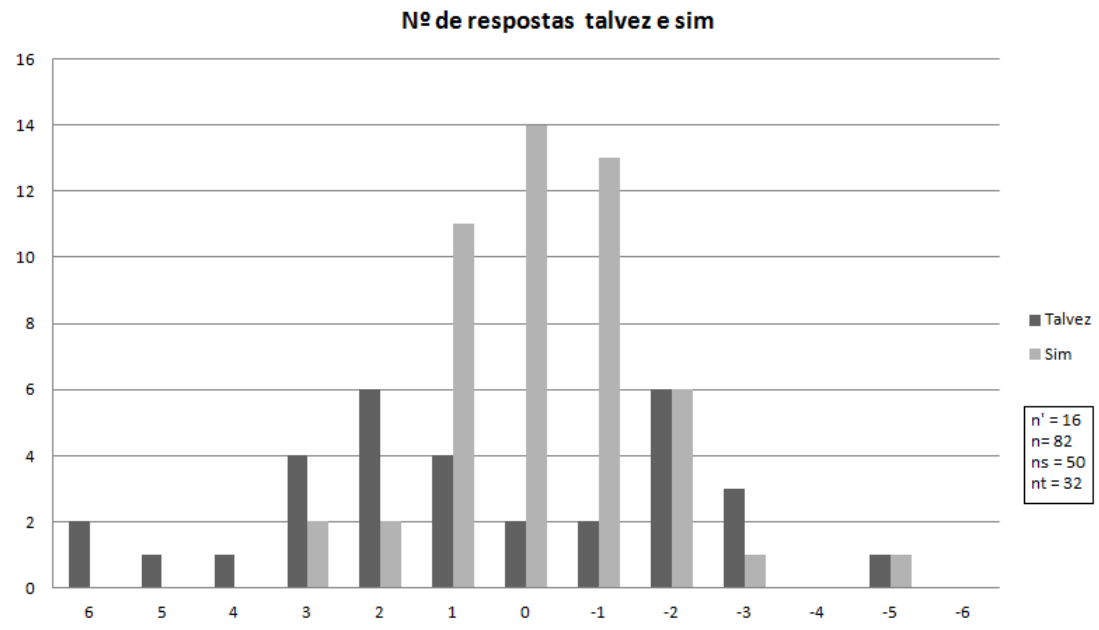

Figura 35: Número de repostas "sim" e "talvez" obtidas no teste 3MFC.

Os gráficos demonstram que há uma maior distribuição dos valores de "não" (nn=126) e um número de respostas "talvez" relativamente baixo (nt=32), ao passo que há um total de 50 respostas "sim" (ns=50), totalizando 208 respostas, como esperado. No primeiro teste, com duas respostas, os números ns e nn foram de 140 e 68, demonstram uma diminuição de 10% de respostas no primeiro caso e um total de 17,5 %. As variáveis que sofrem alterações no quesito da variância, foram as variáveis mais6, mais5, zero, menos1 e menos5. No entanto, como pode se notar pelas figuras 34 e 35 os valores de "talvez" encontram-se na faixa que vai de mais3 a menos3, especificamente nos valores de mais2 e menos2. Nas variáveis mais6, mais5, mais4, zero, menos1 e menos5 o número de repostas "talvez" oscilou entre um e dois. Aplicamos uma ANOVA entre estas variáveis utilizando o número de repostas sim e não comparadas ao de talvez, obteve-se um $p = 0,33$ ($F = 1,28$, para um F crítico = 4,46), para $\alpha = 5\%$ e $n = 96$.

Nossa interpretação é a de que a alternativa "talvez", no caso das variáveis acima elencadas (mais6, mais5, mais4, zero, menos1 e menos5), foi utilizada pelos parti-

cipantes não como uma alternativa de respostas, mas, sim, como uma alternativa compensatória. Analisando-se os tempos de reposta para cada alternativa, nota-se que ao responder "talvez" o participante demorou cerca de 50% a mais de tempo do que nas alternativas "sim" e "não", cerca de 15 segundos para talvez e 10 segundos para "sim" e "não". Ao nosso ver, isso demonstra uma desatenção ao par apresentado ou uma leve fuga da tarefa, sendo que a maneira de se compensar é utilizando-se da respostas talvez. Entendemos que no caso das variáveis mais2 e menos2, por exemplo, se tem justamente o oposto, em que os participante demonstraram ter algum tipo de dúvida com relação a semelhança ou diferenças dos pares.

Devido ao fato de que grande parte das demais respostas "talvez" encontrem-se no intervalo que cremos ser o dos limiares de diferenciação tonal não iremos prosseguir com a aplicação de testes de hipóteses. No capítulo de conclusão faremos melhores considerações sobre estas variáveis, por ora, podemos dizer que as variáveis mais2 e menos2 aparentam ser os indícios de variação, isto é, o valor a partir do qual os ouvintes tem a percepção de que há uma diferença, embora ela esteja ainda muito difusa.

6 Considerações Finais

6.1 Um breve resumo

Para uma boa conclusão sobre nosso trabalho precisamos recorrer a uma metáfora. Imaginemos uma grande sala com quatro portas, atrás de cada uma delas há um prêmio em dinheiro: um é considerado bom, dois são considerados regulares e um visto como ruim. Na mesma sala há dez pessoas, cada uma delas deverá escolher uma porta. Por matemática simples, podemos dizer que a chance individual de cada uma destas pessoas ganhar um prêmio bom ou ruim é de 25% e um prêmio médio é de 50%. O prêmio será multiplicado pela quantidade de pessoas que escolherem cada uma das portas. Temos então que a probabilidade de ganhar cada um dos prêmios precisa ser dividida por 10, ou seja, serão 2,5% para os prêmios bom e ruim e 5% para o prêmio regular. A melhor estratégia para o grupo seria organizar três grupos, dois com três integrantes e outro com quatro, escolher três portas e dividir o prêmio total por dez. Como os prêmios somente serão revelados após todos terem escolhido uma porta, não há como se montar estratégia individual. O única estratégia individual que pode ser usada é escolher a mesma porta que a maioria. Ainda assim, tal estratégia beneficiaria o grupo, antes de beneficiar o indivíduo [1].

Temos então dois aspectos importantes: um deles é a escolha acima debatida, o outro é a execução desta escolha. Embora a percepção individual tenha parte, ela apenas terá algum valor na interação, o que não foi estudado neste trabalho, ou seja, notar diferenças tonais entre pares tonais é uma atividade comportamental. Em uma

[1] Este tipo de raciocínio é base de teoria probabilísticas de análise da dados e é comumente chamada de teoria dos jogos

atividade linguística a percepção de semelhanças e diferenças teria de receber algum tipo de modalização linguística, como a diferença entre uma palavra neutra e uma com foco tonal. Em nosso entender, a interação em nosso trabalho se dá em um nível não-pragmático, diferentemente do que ocorre em uma análise de foco frasal, ou de emoções, por exemplo. A interação em nosso experimento se dá entre elementos fonéticos: o par zero, sendo comparado a doze semitons. A modalização surgirá, como por nós proposto, quando o tom de referência escolhido pelo falante (estratégia individual) ultrapassar o limiar de diferenciação tonal (o padrão de grupo).

Em nosso trabalho estamos lidando com o mesmo tipo de questão. As escolhas individuais, embora tenham impacto na avaliação do grupo não são determinantes para a organização geral da percepção dos tons. Por exemplo, se um único participante de nossa pesquisa houvesse escolhido "sim", ao invés de "não" em uma variável por nós nomeada de *categórica*, teríamos de repensar todo o nosso instrumental de análise e organização de variáveis. Como não foi este o caso pudemos seguir nossa análise como proposta no capítulo 5. Nossa metáfora pretende dizer que a escolha individual, baseada no repertório de algum tipo, como habilidade musical ou conhecimento de entoação do Português, não afeta diretamente o grupo.

6.2 Os limiares de diferenciação tonal

As análises estatísticas apresentadas no capítulo 5 nos levam a alguns pontos de debate. O primeiro é com relação as variáveis categóricas: mais6, mais5, zero, menos4, menos5 e menos6. Nomeamo-nas com tal nome por entender que a percepção de tais variáveis é sempre categórica, ou seja, não há semelhanças ou diferenças evidentes entre elas e o fator fixo zero. Nos casos em que o par zero foi comparado a uma destas variáveis não houve ambiguidade nas respostas, isto é, os falantes não oscilaram em suas respostas, tendo todos os participantes dado a mesma resposta. Variáveis que tiveram um número de "não" igual a n como respostas (cf. Figura30) nomeamos de *categóricas por diferença*. Aquelas que receberam um número de "sim" igual a n,

(cf. Figura 31) nomeamos de *categóricas por semelhança*. Contudo, para a apresentação final de resultados chamaremos todos eles de variáveis categóricas e incluiremos neste conjunto as variáveis: mais6, mais5, zero, menos4, menos5 e menos6. Excluímos a variável mais1 deste conjunto por crermos que ela faz par com outras variáveis abaixo descrita

As variáveis cujas respostas oscilaram entre "sim" e "não" receberam o nome de *variáveis quasi-categóricas*, seja por semelhança, seja por diferença. A análise de fatores junto da análise de intervalos de confiança nos ajudam a melhor agrupar as tais variáveis. Em verdade, a análise de fatores nos permite agrupar as variáveis quasi-categóricas, ao passo que a análise de intervalos de confiança nos ajuda a determinar quais são quasi-categóricas por semelhança e por diferença. Poderíamos usar um critério quantitativo, como:

(1) Se *sim* > *não* $\rightarrow$ semelhantes;

(2) Se *não* > *sim* $\rightarrow$ diferentes.

Contudo, este critério não seria adequado, uma vez que não poderia ser replicado à exaustão em outros estudos. Por esta razão usamos como critério de separação os gráficos de barras de erros. Há também uma questão semântica, nomeá-las por critérios de semelhança ou diferença nos levaria a criar limiares de limiares. A noção de limiares de limiares é próxima da definição de limite em matemática, por exemplo:

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

leia-se: o limite de x tendendo a zero é igual a infinito.

Não existem números reais para a divisão de qualquer número real por zero. Desse modo, no denominador da função serão inseridos números próximos de zero, mas

nunca o número zero, isto é, 0,001, 0,000001, 0,0000000001 e assim por adiante. Quanto mais próximo de zero, mais próximo do infinito. A ideia de limiares de limiares segue o mesmo princípio quanto mais próximos estivermos de um limiar, mais próximos estaremos de um outro limiar ainda não pesquisado, o qual tem uma natureza distinta do primeiro. Obviamente, não chegaremos nunca ao infinito, pois a percepção humana possui algum grau de finitude.

Wertheimer discute esta questão em seu artigo *Laws of Organization in Perceptual Forms* (WERTHEIMER, 1938). Para o autor a percepção não poderia ser entendida como a coleção de pequenas percepções de inúmeros objetos, mas, sim, um conjunto uno, regulados por certas regras. Os fisiologistas do século XIX, como van Helmholtz (HELMHOTZ, 1895), acreditavam no contrário, que embora houvesse uma organização externa, haveria uma lógica intrínseca ao objeto percebido, como a qualidade de uma nota musical, a qual possui uma harmonia interna e uma harmonia na sua relação com demais notas. A proposta adequada para nosso trabalho parece ser a de Patricia Kuhl (KUHL, 1991; IVERSON; KUHL, 1995; KUHL et al., 2006). Nesta abordagem, haveria formas prototípicas, as quais por uma questão de gradiência iriam se tornando "não-prototípicas", mas mantendo alguma conexão perceptiva com aquela considerada prototípica. A diferença entre estas formas estaria associada a mudanças acústicas, por exemplo. O princípio é parecido com os conceitos de proximidade e similaridade de Wertheimer (1938). Objetos próximos tendem a ser compreendidos como um único, ao passo que objetos distantes são percebidos como diferentes; objetos semelhantes são percebidos como equivalentes, enquanto objetos não-semelhantes são vistos como elementos distintos.

O conceito de *Perceptual Magnet Effect* (PME) (KUHL, 1991; IVERSON; KUHL, 1995) reúne todas as pontas acima descritas. Os elementos acústicos são levados em consideração, uma vez que são a base para a gradiência, atendendo as necessidades das teorias fisiológicas. Ao mesmo tempo, as características de percepção da forma são trazidas à tona. O *Perceptual Magnet Effect* pode ser compreendido como um elemento

prototípico para o qual a percepção converge elementos gradientes, desde que estes estejam em fase com aquele prototípico. Assim definimos por variáveis *quasi-categóricas* : mais2, mais1, menos1 e menos2. Entendemos que estas variáveis estão, de fato, sob a influência do efeito magnético de percepção (PME), isto é, a percepção delas tende a ser categorizada de acordo com os fatores de proximidade e similaridade, dada a proximidade e a semelhança com o tom neutro (zero), tendem a ser compreendidas como esta. Embora, elas tenham de 6% à 13% de variação acústica, este fator não afetou de maneira absoluta como se dá a percepção dos tons. Neste sentido, não vemos a necessidade de distingui-las entre variáveis quasi-categóricas por semelhança e diferença, por esta razão as nomearemos unicamente de *variáveis centrais* e incluiremos nesta categoria a variável menos1.

A variável mais3 possui um caráter ambíguo, embora quantitativamente suas respostas sejam de "não" (4 para 12). Estatisticamente ela não pode ser considerada uma variável central, bem como, não pode ser considerado o limiar de diferenciação tonal. Os valores de carga fatorial apresentados na Tabela 12 nos deixam fazer tal afirmação, uma vez que o valor de carga fatorial para a variável mais4 é cerca de 0,279 maior do que a carga fatorial de mais3. Os intervalos de confiança também demonstram que a probabilidade de se confundir a variável mais3 com as variáveis centrais ou com a variável mais4 é aproximadamente igual. A Figura 36 demonstra este fato.

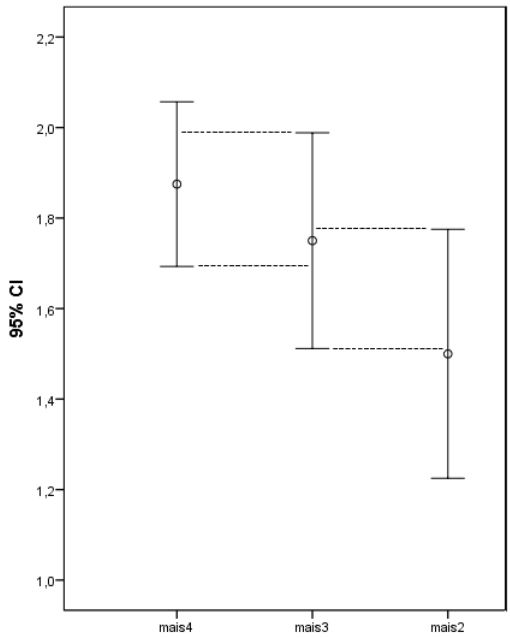

Figura 36: Intervalos de confiança para as variáveis mais4, mais3 e mais2. As linhas tracejadas indicam a intersecção entre as barras de erros.

Como se pode observar na Figura 36 as intersecções, representadas pelas linhas pontilhadas, entre a variável mais3 e as demais variáveis são bastantes parecidas, ou seja, não é possível determinar com clareza qual a probabilidade de se considerar mais3 uma variável central ou o limiar de diferenciação tonal. Novamente, o modelo de PME pode ser aplicado a interpretação desta variável, se pensarmos que ela é atraída para as variáveis centrais ou para o limiar de diferenciação tonal, de acordo com o modo de fala ou o *pitch range* do falante, por exemplo, ou ainda pela capacidade de diferenciação do ouvinte.

É importante ressaltar que o trabalho de Consoni (2011) utilizou-se de uma frase portadora. A frequência fundamental desta frase não foi alterada, de modo que os únicos elementos alterados foram, ora a palavra toda, ora apenas a sílaba tônica. Em um outro trabalho (CONSONI; FERREIRA-NETTO, 2008) a autora utilizou somente palavras isoladas procurando determinar a sensibilidade dos falantes de Português para variação tonal ascendente. Partindo do princípio de que nosso trabalho não possui uma frase portadora, sendo, portanto, sentenças isoladas, aplicaremos o mesmo

método de análise utilizada pela autora. O método utilizado consiste em cruzar as respostas as variáveis reconhecidas como diferentes com aquelas não reconhecidas como tal. O ponto em que as curvas de dados se cruzam é entendido como o limiar de diferenciação tonal. As figuras 37 38 apresentam os gráficos obtidos:

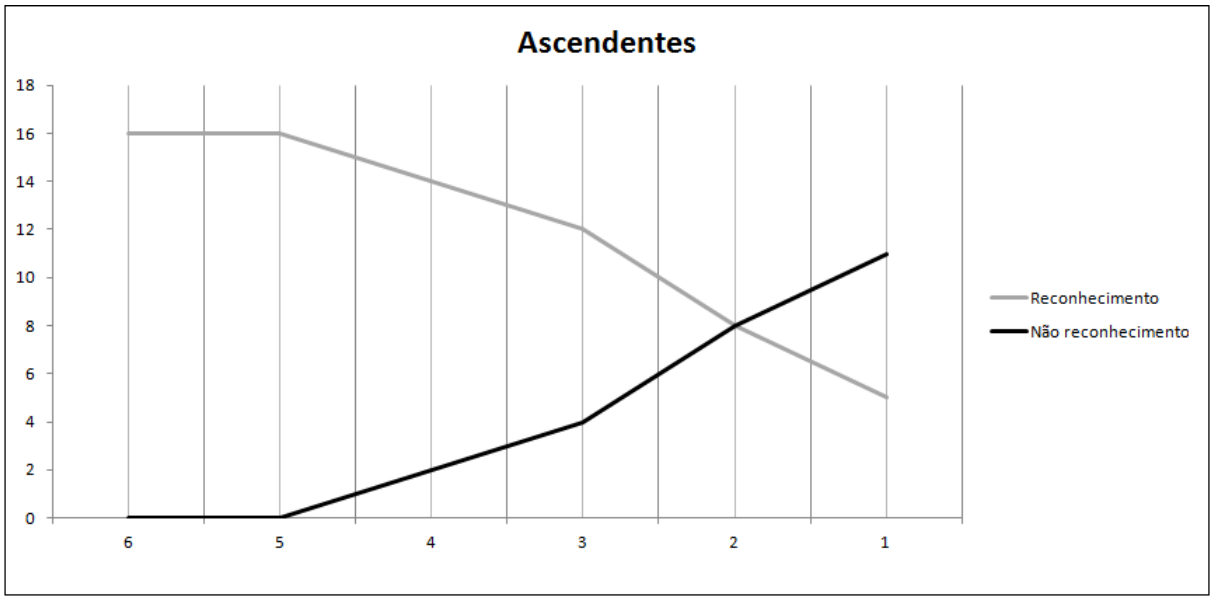

Figura 37: Gráfico que demonstra o cruzamento das curvas de reconhecimento e de não-reconhecimento das diferenças entre os pares das manipulações ascendentes.

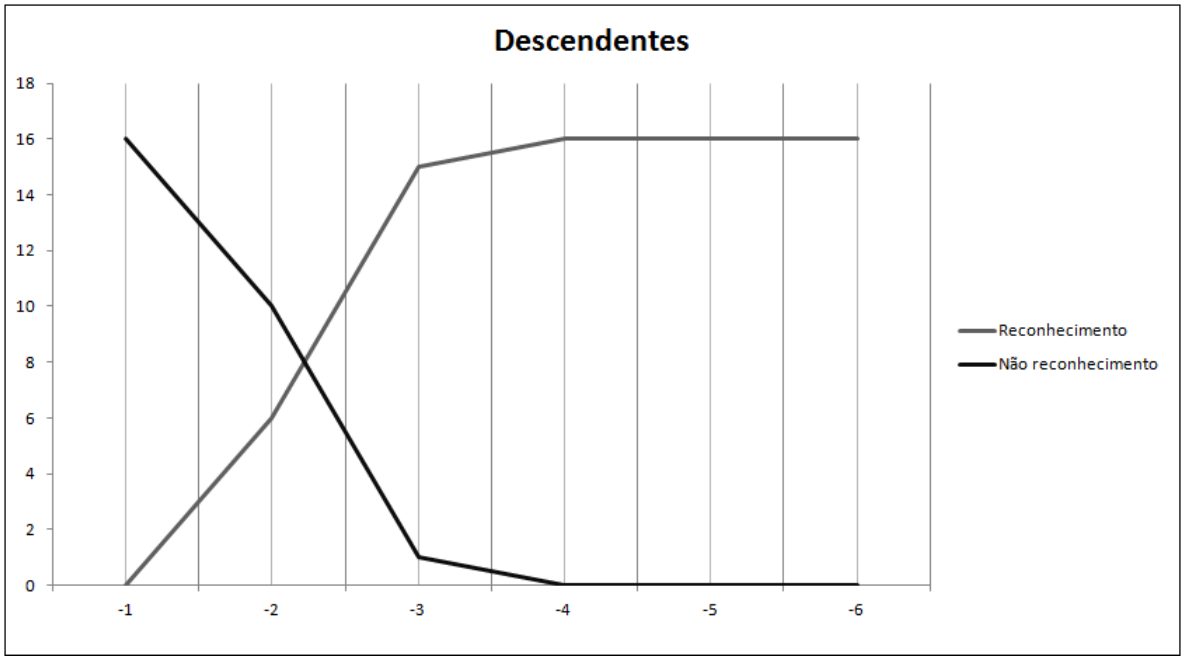

Figura 38: Gráfico que demonstra o cruzamento das curvas de reconhecimento e de não-reconhecimento das diferenças entre os pares das manipulações descendentes.

Como pode se observar nas figuras acima, os pontos de intersecção das curvas de reconhecimento e não reconhecimento se deram por volta das variáveis mais e menos3. Como verificado no teste 3MFC as variáveis mais2 e menos2 representam os pontos de uma possível modalização perceptual, contudo não são os pontos de mudança definitivos, mas, sim, os pontos de confusão. Estes pontos estão mais associados a uma escolha individual ou ao contexto em que foram percebidos, do que a uma propriedade da percepção em si. O mesmo princípio vale para a variável mais4, embora esta pareça estar em uma relação de confusão mais complexa com o limiar de diferenciação tonal. De fato, a partir da figura 36, dos valores de carga fatorial obtidos na tabela 12 do capítulo 5, bem como com os dados do teste 3MFC, podemos deduzir que a variável mais4 apenas acentua a diferença, ou seja, a variável mais4 representa a percepção já modalizada e não um limiar, o ponto em que a percepção passa a ser modalizada. Nesse sentido, em nosso entender os limiares de diferenciação tonal, objeto de pesquisa deste trabalho, encontram-se nas variáveis mais3 e menos3. De acordo com os resultados obtidos, para que se note alguma diferença tonal entre duas sentenças do Português Brasileiro é necessário que se varie três semitons ascendentes ou três semitons descendentes. No caso dos tons descendentes os achados aqui reportados são encontram-se diferentes daqueles apresentados por Consoni (2011) em um semitom, uma vez que os achados da autora apontavam para uma variação de 3 semitons ascendentes e 4 semitons descendentes.

Uma de nossas hipóteses explicativas até o momento é a de que o efeito temporal toma parte na diferença encontrada entre nossa pesquisa e a de Consoni et al. (2009). Em nosso entender, uma maior janela temporal para a percepção tonal permite uma refinação na acurácia, o que é levado para a tarefa de responder se há ou não diferenças entre os pares tonais. Embora os estudo que tomem isso com objeto levam em conta apenas a distância temporal entre estímulos, acreditamos que podemos melhor definir sobre este tema. Por um viés fisiológico, encontramos respostas na comparação entre a teoria tonotópica e a teoria temporal. Para os primeiros, a percepção de frequências é determinada por espaços definidos na membrana basilar, enquanto a segunda afirma

que a percepção da frequência é baseada no tempo em que os neurônios respondem a um som na cóclea, a variação destas janelas temporais altera o processo de percepção (MOORE, 2005).

Outra hipótese para a diferença entre os achados nesta pesquisa e os de Consoni (2011) refere-se ao método. No trabalho de 2009 os autores utilizaram uma voz feminina para a comparação, enquanto nós usamos uma voz masculina. A manipulação por semitons tem a característica de ser uma progressão geométrica: uma manipulação de 1 semitom para uma frequência de 100 Hz resultará em uma frequência de 106 Hz, para 200 Hz serão 212 Hz. De acordo com Behlau e Pontes (1995) a frequência fundamental de uma voz masculina é 100 Hz e uma voz feminina 200 Hz. Deste modo, a manipulação da voz masculina comF_0 na faixa dos 140 Hz, como a utilizada em nossa pesquisa, em 6 semitons ascendentes a eleva para a faixa dos 200 Hz, o *default* feminino. Por outro lado, a diminuição em 6 semitons descendentes de uma voz feminina em 200 Hz, a deixará com cerca de 140 também. O que temos, então, é uma variação das frequências fundamentais, devido a sexo da voz usada como estímulo, o qual causa uma diferença entre as respostas nos testes. Voltamos, novamente, a questão dos limiares dos limiares e do paradoxo da divisão finita de um número infinito, ou seja, quanto mais modularmos o envelope de F_0, mais específicos teremos de ser *ad infinitum*.

Uma terceira hipótese refere-se ao fato de os estímulos utilizados em Consoni (2011) teriam uma referência fixa para todos os estímulos, o tom da frase portadora. Neste sentido, os participantes da pesquisa da autora tinha uma percepção abrupta da variação tonal, ao passo que nossos participantes poderiam realizar o julgamento durante o tempo de realização da frase. Roederer (2002) sugere que variações súbitas são detectadas com uma j.n.d. mais baixa do que aquelas percebidas em um tempo normal de processamento. Isto é, dado um pulso com um tom invariável, a primeira variação súbita será detectada, mesmo que esta variação seja pequena em relação ao tom original. Contudo, esta percepção de variação súbita tende a diminuir conforme

os tons se tornam mais graves. Roederer (2002) sugere que a resolução para tons em torno de 500 Hz é de 3%, ao passo que para tons em torno de 2000Hz esta resolução é de apenas 0,5%. Dessa forma, a diminuição do valor do limiar de diferenciação tonal inferior pode ser explicado pelo fato de que no caso da percepção abrupta os participantes tem uma pior definição na percepção do tom grave, o que leva a diferença entre os dados aqui encontrados e os encontrados em Consoni (2011)

6.3 Palavras finais

Podemos ilustrar nossos resultados no esquema apresentado na Figura 39:

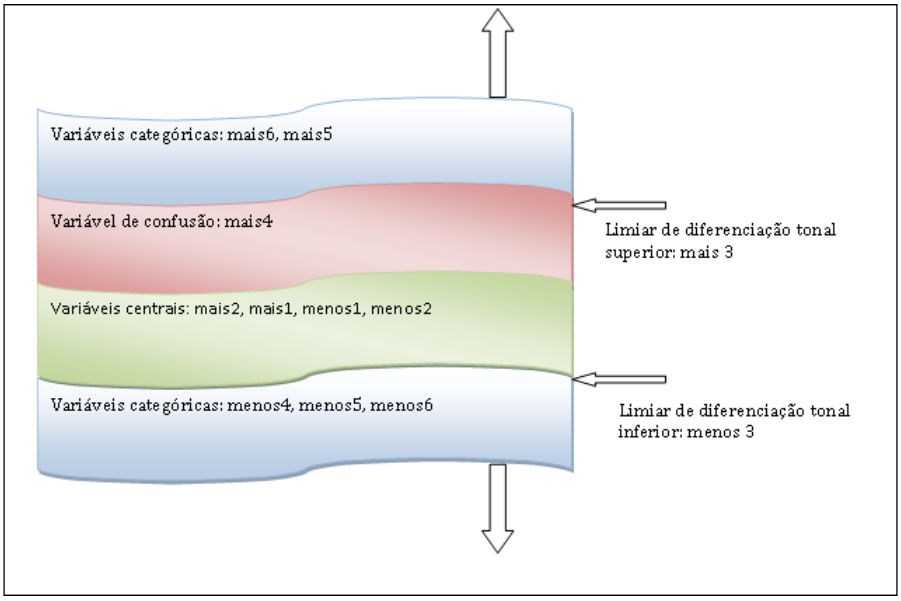

Figura 39: Esquema representando os limiares de diferenciação tonal para o Português Brasileiro, como relação a um tom natural

A Figura 39 apresenta como seria a percepção entoacional do Português Brasileiro. Os limiares de diferenciação tonal estão representados pelas setas horizontais a direita, enquanto as setas verticais representam a variação categórica. As faixas verticais superior e inferior, de cor cinza, representam as variáveis categóricas, aquelas quais os participantes tiveram a tendência a dizer 100% das vezes que são diferentes. As faixas centrais ilustram, respectivamente, as variáveis centrais e a variável de confusão, as primeiras são aquelas, as quais tendem a sempre serem percebidas como iguais a

sentença zero, não manipulada, ao passo que a segunda possui uma ambiguidade em sua percepção, podendo ser percebida, ora como categórica, ora como central.

De acordo com o que ilustramos anteriormente podemos afirmar que a inversão dos resultados entre esta pesquisa e a de Consoni (2011) é causada pela ambiguidade da frequência fundamental entre os sexos masculino e feminino. Tal fato demonstra que os limiares de diferenciação tonal não são valores estáticos, mas, sim, valores de transição, os quais estão em uma gradiência, a qual pode ser quantificada. Nossa proposta, para continuar esta pesquisa é cruzar os valores de ambos os trabalhos em busca de um valor relativamente mais central. Além disso, os próximos passos desta pesquisa são: (i) analisar os valores temporais mínimos para a modalização de frequência, de acordo com os valores do limiar de diferenciação tonal. (ii)Além disso, pretendemos estabelecer um modelo matemático, o qual preveja essas variações de frequência tomando a variação temporal δ_t como função relativa da variação dos limiares de diferenciação tonal. (iii) um terceiro passo refere-se a análise conjuntas de dados com durações variadas.

Referências

ALBANO, E. C. *O gesto e suas bordas: esboço de Fonologia-Articulatória do Português*. Campinas: Mercado das Letras, 2009.

AQUINO, T. de. *Summa Theologica (Part I)*. New York: Cosimo, 2007.

ASHBY, W. R. *An Introduction to Cybernetics*. London: Chapmann & Hall, 1957.

BAER, T. *Investigation of phonation using excised larynx*. Tese (Doutorado) — M.I.T., Cambridge, 1975.

BAIRD, A. A. et al. Frontal lobe activation during object permanence: Data from near-infrared spectroscopy. *NeuroImage*, v. 16, p. 1120–1126, 2002.

BEHLAU, M.; PONTES, P. *Avaliação e tratamento das disfonias*. São Paulo: Lovise, 1995.

BENVENISTE Émile. *Problemas de Linguística Geral I*. São Paulo: Nacional/Edusp, 1976.

BERTALANFFY, L. von. *Teoria Geral dos Sistemas*. Petrópolis: Ed. Vozes, 1975.

BOERSMA, P.; WEENINK, D. [Programa de computador], *Praat: doing phonetics by computer (Version 5.1.41)*. 2009. From: www.praat.org. Acessado em Agosto de 2011.

BOGACZ, R. et al. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, v. 113, n. 4, p. 700–765, 2006.

BOLINGER, D. L. Intonation: Level vs. configuration. *Word*, v. 7, n. 3, p. 191–210, 1961.

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.

BROOKS, R. A. Inteligence without representation. In: HAUGELAND, J. (Ed.). *Mind Design II*. Cambridge: MIT Press, 1997. p. 395–420.

BROWMAN, C. P.; GOLDSTEIN, L. Articulatory gestures as phonological units. *Phonology*, v. 6, p. 201–251, 1989.

BROWMAN, C. P.; GOLDSTEIN, L. Articulatory phonology: an overview. *Phonetica*, v. 49, p. 155–180, 1992.

BRUCE, G. *Swedish Word Accents in Sentence perspective*. Lund: LiberLaromedel/Gleerup, 1977.

CAMPOS, D. de. *Avaliação morfométrica do nervo laríngeo recorrente em cadáveres humanos*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

CHERY, E. C. Some experiments on the recognition of speech, with one and two ears. *Journal of Acoustical Society of America*, v. 25, p. 975–979, 1953.

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. *Aspects of the theory of syntax*. Cambridge: MIT Press, 1965.

CHOMSKY, N. *Knowledge of Language: Its Nature, Origins and Use*. New York: Praeger, 1986. (Converge).

CHOMSKY, N. *New Horizons in the Study of Language and Mind*. [S.l.]: Cambridge University Press, 2000.

CHOMSKY, N. *Syntactic Structures*. Netherlands: Mouton: The Haugue Mouton, 2002.

CHOMSKY, N. The mysteries of nature: How deeply hidden? *Journal of Philosophy*, v. 106, n. 4, p. 167–200, April 2009.

CHOMSKY, N.; HALLE, M. *The sound pattern of the English*. New York: Harper & Row, 1968.

COLTON, R. H.; CASPER, J. K.; LEONARD, R. *Understanding Voice Problems: A Physiological Perspective For Diagnosis And Treatment*. Philadelphia: Lippincott Williams & Wilkins, 2006.

COLVIN, S. *A Historical Greek Reader: Mycenaean to the Koiné*. Oxford: Oxford University Press, 2007.

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 Lingusitic Meeting*. Póznan: [s.n.], 2009.

COZBY, P. *Métodos de Pesquisa em Ciências do Comportamento*. São Paulo: Editora Átlas, 2003.

DANCEY, C. P.; REIDY, J. *Estatística Sem Matemática para Psicologia*. Porto Alegre: Artmed - Bookman, 2006.

DILLEY, L. *The phonetics and phonology of tonal systems*. Tese (Doutorado) — M.I.T., Cambridge, 2005.

ECO, U. *A busca da Língua Perfeita na Cultura Europeia*. Bauru: EDUSC, 2001.

ERDFELDER, E.; FAUL, F.; BUCHNER, A. Gpower: A general power analysis program. *Behavior Research Methods, Instruments, & Computers,*, v. 28, p. 1–11, 1996.

EWERTY, A. Dante's theory of language. *The Modern Language Review*, v. 35, n. 3, p. 355–366, 1940.

FANT, G. *Acoustic Theory of speech production*. Netherlands: Mouton: The Haugues, 1966.

FANT, G. Phonetics and phonology in the last 50 years. In: M.I.T. *From sound to sense.* Cambridge, 2004. p. B20–B41.

FANT, G. *Speech acoustics and phonetics*. [S.l.]: Kluwer Academic Publisher, 2004. (Text, speech, and language technology, v. 24).

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. Exprosodia. processo nº08992-2. *Revista da Propriedade Industrial*, v. 2038, p. 167, 2010.

FONàGY, I. *La vive voix: Essais de psycho-phonetique*. Paris: Payot, 1991.

FREITAS, C. A. F. de et al. Anatomia cirúrgica do ramo externo do nervo laríngeo superior. *Revista Brasileira Cirurgia de Cabeça e Pescoço*, v. 36, n. 2, p. 58–61, 2007.

GöDEL, K. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme. *Monatshefte für Mathematik und Physik*, v. 38, p. 173–198, 1930.

GOLDSMITH, J. *Autosegmental Phonology*. Tese (Doutorado) — M.I.T., Cambridge, 1976.

GRAY, H. *Anatomy of the Human Body*. Philadelphia: Lea & Febiger, 1918; Bartleby.com, 2000, 1918.

GUIDORIZZI, H. L. *Um Curso de Cálculo - Vol.1*. 5. ed. Rio de Janeiro: LTC, 2008.

HAIR, J. F. et al. *Análise Multivariada de Dados*. Porto Alegre: Bookman, 2009. 688 p.

HAYES, B. *A metrical theory of stress rule*. Tese (Doutorado) — M.I.T., Cambridge, 1981.

HAYES, K. J. Introduction: the heritage and legacy of raging bull. In: HAYES, K. J. (Ed.). *Martin Scorcese's Raging Bull*. [S.l.]: Cambridge University Press, 2005.

HELMHOTZ, H. L. F. *On the sensations of tone as a physiological basis for the theory of music*. London: Longmans, Green and Co., 1895.

HERMES, D. J.; GESTEL, J. C. van. The frequency scale of speech intonation. *Journal of Acoustical Society of America*, v. 90, n. (1), p. 97–102, 1991.

HILL, C. M. On choice. *The American Journal of Psychology, Vol. 9, No. 4 (Jul., 1898), pp.Published*, v. 9, n. 4, p. 587–590, Julho 1898.

HIRANO, M.; KURITA, S.; NAKASHIMA, T. The strucuture of vocal folds. In: STEVENS, K. N.; HIRANO, M. (Ed.). *Vocal fold physiology*. Tokyo: University of Tokyo Press, 1981.

HIRST, D.; CRISTO, A. di. A survey of intonation systems. In: HIRST, D.; CRISTO, A. di (Ed.). *Intonation Systems*. Cambridge: Cambridge University Press, 1998.

HJELMSLEV, L. *Prolegômenos a uma teoria da linguagem*. São Paulo: Editora Perspectiva, 2006.

HOPCROFT, J.; MOTWANI, R.; ULLMAN, J. *Introdução à Teoria de Autômatos, Linguagens e Computação*. Rio de Janeiro: Editora Campus-Elsevier, 2005.

IVERSON, P.; KUHL, P. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of Acoustical Society of America*, v. 97, p. 553–562, 1995.

JAKOBSON, R.; FANT, G.; HALLE, M. *Preliminaries to Speech Analysis*. Cambridge: MIT Press, 1952.

JUN, S.-A. *Prosodic typology : the phonology of intonation and phrasing*. Oxford: Oxford University Press, 2006.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, v. 23, n. 3, p. 187–200, 1958.

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.

KARK, A. E. et al. Voice changes after thyroidectomy: role of the external laryngeal nerve. *British Medical Journal*, v. 289, n. 6465, p. 1412–1415, Nov. 1984.

KENNEDY, J. Kurt gödel. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*. Stanford: Stanford University, 2011.

KLEENE, S. C. *Mathematical Logic*. New York: Dover Publications, 2002.

KRAEMER, H.; THIEMANN, S. *How many subjects?: Statistical power analysis in research*. Newbury Park: Sage Publications, Inc, 1987.

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.

KUHL, P. et al. *Língua, cultura, mente, cérebro: progresso nas fronteiras entre disciplinas*. São Paulo: Editora Paulistana, 2006.

LABOV, W. *Padrões Sociolinguísticos*. São Paulo: Parábola Editorial, 2010.

LADD, D. R. *Intonational Phonology*. Cambridge: Cambridge University Press., 1996.

LADEFOGED, P. Phonetics and phonology in the last 50 years. In: MIT. *Proceedings of From Sound to sense: 50 years of Discoveries in Speech Communication at MIT*. Cambridge, 2004.

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.

LEBEN, W. *Suprasegmental Theory*. Tese (Doutorado) — M.I.T., Cambridge, 1973.

LEHISTE, I. *Suprasegmentals*. Cambridge: M.I.T. Press, 1970.

LIBERMAN, A. et al. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, v. 54, p. 358–368, 1957.

LIBERMAN, A.; MATTINGLY, I. The motor theory of speech perception revised. *Cognition*, v. 21, p. 1–36, 1985.

LIBERMAN A.M., C. et al. Perception of the speech code. *Psychological Review*, v. 74, p. 431–461, 1967.

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.

LIEBERMAN, P. On the acoustic basis of perception of intonation by linguists. *Word*, v. 21, p. 40–54, 1965.

LIKERT, R. A technique for the measurement of attitudes", archives of psychology 140: pp. 1-55. *Archives of Psychology*, v. 140, p. 1–55, 1932.

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, Fev. 2009.

LéVI-STRAUSS, C. *Olha escutar ler*. São Paulo: Companhia das Letras, 2010.

MACIEL, F.; ROTHE-NEVES, R. Investigações experimentais da entonação no português brasileiro: revisão de literatura. *IX Congresso da Sociedade Brasileira de Fonética/III Congresso Internacional de Fonética e Fonologia, Belo Horizonte*, v. 1, p. 1–48, 2006.

MACIEL, J. *Elementos de Teoria Geral dos Sistemas*. Petrópolis: Editora Vozes, 1974.

MAEDA, S. *A Characterization of American English Intonation*. Tese (Doutorado) — M.I.T., Cambridge, 1976.

MAGALHãES, M. N.; LIMA, A. C. P. de. *Noções de Probabilidade e Estatística*. São Paulo: Edusp, 2008. 383 p.

MARRONE, S. P. Medieval philosophy in context. In: MCGRADE, A. S. (Ed.). *The Cambridge Companion to Medieval Philosophy*. Cambridge: Cambridge University Press, 2003. p. 10–50.

MARTINET, A. *Elementos de linguística geral*. São Paulo: Martins Fontes, 1975.

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.; FERREIRA-NETTO, W. Speech intonation and perception: A study of frequency scales for brazilian portuguese. *Journal of Acoustical Society of America*, v. 129, n. 4, p. 2657, 2011.

MATTINGLY, I. G. Speech synthesis for phonetic and phonological models. In: SEBEOK, T. A. (Ed.). *Current trends in Linguistics*. Mouton: The Haugue Mouton, 1974. v. 12, p. 2451–2487.

MOORE, B. J. *An Introduction to the Psychology of Hearing*. Londres: Academic Press, 2005.

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.

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.

NEWTON, I. *Principia - Princípios Matemáticos de Filosofia Natural*. [S.l.]: Editora da Folha, 2010.

SCHOUTEN, M. (Ed.). *NATO Advanced Research Workshop on "The Psychophysics of Speech Perception"*. North Atlantic Treaty Organization, Utrecht: Kluwer Academic Publisher, 1987.

OHALA, J. J. Cross-language use of pitch: an ethological view. *Phonetica*, v. 40, p. 1–18, 1983.

PEREIRA, J. C. R. *Análise de Dados Qualitativos*. 3ª. ed. São Paulo: Edusp, 2004.

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

PIAGET, J.; INHELDER, B. *A psicologia da criança*. 10. ed. Rio de Janeiro: Bertrand Brasil, 1989.

PIERCE, J. R. *The science of musical sounds*. New York: Scientific American Library, 1983.

PIERREHUMBERT, J. *The Phonology of Phonetic of English Intonation*. Tese (Doutorado) — M.I.T., Cambridge, 1980.

PIERREHUMBERT, J. Tonal elements and their alignment. In: HORNE, M. (Ed.). *Prosody, Theory and Experiment: studies presented to Gösta Bruce*. Dodrecht: Kluwer Academic Publisher, 2000. p. 11–36.

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.

PROM-ON, 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.

REA, L. M.; PARKER, R. A. *Metodologia de pesquisa: do planejamento à execução*. São Paulo: Editora Pioneira, 2000.

ROEDERER, J. G. *Introdução a Física e Psicofísica da Música*. São Paulo: Edusp, 2002.

RUSSELL, B. A análise da matéria (cap. xvii a xx). In: *Bertrand Russell: Ensaios Escolhidos*. São Paulo: Ed. Abril, 1978, (Os Pensadores). p. 161–202.

RUSSO, I. C. P. *Acústica e psicoacústica aplicadas à fonoaudiologia*. São Paulo: Lovise, 1993.

SAUSSURE, F. de. *Curso de Linguística Geral*. São Paulo: Editora Cultrix, 1977.

SEARLE, J. R. Chomsky's revolution in linguistics. *The New York Review of Books*, v. 18, n. 12, June 1972. Disponível em: <http://www.nybooks.com/articles/archives/1972/jun/29/a-special-supplement-chomskys-revolution-in-lingui/?pagination=false>.

SHANNON, C. E. *An algebra for theoretical genetcis*. Tese (Doutorado) — M.I.T., Cambridge, 1940.

SHANNON, C. E.; WEAVER, W. *The Mathematical Theory of Communication*. Urbana: Univ of Illinois Press, 1949.

SHIGENO, S. The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception & Psychophysics*, v. 40, p. 9–19, 1986.

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.

SKYTTNER, L. *General Systems Theory*. Singapore: World Scientific Publishing Co., 2001.

SOUZA, L. S. *Lesão do nervo laríngeo superior pós tireoidectomia sob anestesia local e hipnosedação: frequência, alterações vocais e endoscópicas subsequentes da laringe*. Dissertação (Mestrado) — Unicamp, Campinas, 2008.

SPINOZA, B. *Ethics Demonstrated in Geometrical Order*. [S.l.]: www.earlymoderntexts.com, 2004.

STEELE, J. *An essay towards establishing the melody and measure of speech to be expressed and perpetuated by peculiar symbols*. Londres: J. Almon, 1775.

STEVENS, K. The quantal nature of speech: Evidence from articulatory-acoustic data. In: DAVID, E.; DENES, P. (Ed.). *Human Communication: A Unified View*. New York: McGraw-Hill, 1972. p. 51–66.

STEVENS, K. N. *Acoustic Phonetics*. Cambridge: MIT Press, 2000.

STEVENS S.S; VOLKMAN; J. & NEWMAN, E. A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, v. 8, n. 3, p. 185 – 190, 1937.

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.

SUMBY, W.; POLLACK, I. Visual contribution to speech intelligibility in noise. *Journal of Acoustical Society of America*, v. 26, p. 212, 1954.

TáBOAS, P. Z. *Cálculo em uma variável real*. São Paulo: Edusp, 2008.

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.; COHEN, A. Intonation by rule: a perceptual quest. *Journal of Phonetics*, n. 1, p. 309–327, 1973.

T´HART, J.; COLLIER, R.; COHEN, A. *A perceptual study of intonation*. Cambridge: Cambridge University Press., 1990.

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, Apr. 1989a.

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.

TURING, A. Computing machinery and intelligence. In: HAUGELAND, J. (Ed.). *Mind Design II*. Cambridge: MIT Press, 1997. p. 29–56.

VAISSIÈRE, J. Perception of intonation. In: PISONI, D. B.; REMEZ, R. E. (Ed.). *Handbook of Speech Perception*. Oxford: Blackwell, 2004. p. 236–263.

VERNON, M. *Percepção e Experiência*. São Paulo: Editora Perspectiva, 1974.

VIVAI, C. B. S. *Uma leitura do De vulgari eloquentia de Dante Alighieri*. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2009.

WERTHEIMER, M. Laws of organization in perceptual forms. In: ELLIS, W. D. (Ed.). *A source book of Gestalt psychology*. London: Routledge & Kegan Paul, 1938.

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>.

XU, Y. Speech melody as articulatorily implemented communicative functions. *Speech Communication*, v. 46, p. 220–251, 2005.

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 Comunnication 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.

ZACKSENHOUSE, M.; BOGACZ, R.; HOLMES, P. Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology*, v. 54, p. 230–246, 2010.

Anexos

Anexo I

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

Termo de consentimento para a participação voluntária em pesquisa.

Eu,.....RG..... concordo em participar da pesquisa referente a dissertação de mestrado em Filologia e Língua Portuguesa do mestrando Marcus Vinícius Moreira Martins, com o tema: "Aspectos da percepção e do controle entoacional do Português Brasileiro". A participação consta da gravação de voz de 3 séries de 5 frases com duração média de 4 segundos a serem utilizada na construção de um experimento de percepção. De maneira que, minha identidade (do participante) será preservada.
Considero que a pesquisa tem caráter científico.

.....
Participante

.....
Mestrando

.....
Orientadora

Data: ..../..../.....

Anexo II

Questionário de identificação

Nome: .....

Idade: ..... Sexo: ( )Masculino ( )Feminino

Nacionalidade: .....

Profissão: .....

Formação/Escolaridade: ( ) Nível Superior ( )Pós-Graduação

( ) Completo ( ) Incompleto

Curso:.....

Monolíngue ( ) Bilíngue ( )

Queixa ou alteração de voz: ( ) sim não ( )

Faz uso profissional da voz: ( ) sim não ( )

É fumante: ( ) sim não ( )

Anexo III

```
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
stimulusFileNameHead = "\Manipulações\"
stimulusFileNameTail = ""
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1 seconds
stimulusMedialSilenceDuration = 1.2
numberOfDifferentStimuli = 13
"-1st.wav,0.wav" ""
"+3st.wav,0.wav" ""
"0.wav,-4st.wav" ""
"-2st.wav,0.wav" ""
"+6st.wav,0.wav" ""
"0.wav,+5st.wav" ""
"0.wav,-3st.wav" ""
"0.wav,+1st.wav" ""
"-5st.wav,0.wav" ""
"+2st.wav,0.wav" ""
"0.wav,-6st.wav" ""
"+4st.wav,0.wav" ""
"0.wav,0.wav" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 22
randomize = <PermuteBalancedNoDoublets>
startText = "Este é um experimento de percepção.
A seguir serão apresentados pares de uma mesma oração.
Sua tarefa será responder se há ou não diferenças de tom entre elas.
Responda SIM, caso note alguma diferença, responda NÃO, caso não note diferenças.
Clique  para continuar."
runText = "A primeira sentença se parece com a segunda?"
pauseText = "Clique para continuar."
endText = "Fim do experimento.
Muito Obrigado."
maximumNumberOfReplays = 0
;replayButton = 0.4 0.8 0.02 0.08 "Nog een keer luisteren" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
0.25 0.4 0.6 0.8 "Sim" "" "A"
0.6 0.75 0.6 0.8 "Não" "" "B"
numberOfGoodnessCategories = 0
```

Anexo IV

Script para a para a preparação de teste de percepção de resposta forçada, baseado no *ExperimentMFC* do software *Praat*. Mais informações podem ser encontradas em http://www.fon.hum.uva.nl/ praat/manual/ExperimentMFC.html . O código abaixo refere-se ao treino para o teste em si.

```
{}
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
stimulusFileNameHead = "\Manipulações\"
stimulusFileNameTail = ""
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1 seconds
stimulusMedialSilenceDuration = 0.0
numberOfDifferentStimuli = 3
"casa0.wav,casa-6.wav" ""
"casa+4.wav,casa0.wav" ""
"casa0.wav,casa0.wav" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 22
randomize = <PermuteBalancedNoDoublets>
startText = "Olá,
Este é um teste de percepção. A seguir serão apresentados pares de uma mesma senten
Eles tiveram seus tons manipulados, algumas ficaram mais graves
e outras mais agudas, ou seja, algumas parecem ser faladas por uma voz mais fina
e outras por uma voz mais grossa, embora elas sejam da mesma pessoa.
Sua tarefa será dizer se você nota ou não alguma diferença entre os pares.
Vamos fazer um treino antes de começar.
Clique para continuar"
runText = "As sentenças são parecidas?"
pauseText = "Clique para continuar."
endText = "Fim do Treino.Agora vamos ao experimento."
maximumNumberOfReplays = 0
;replayButton = 0.4 0.8 0.02 0.08 "Nog een keer luisteren" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
0.25 0.4 0.6 0.8 "Sim" "" "A"
0.6 0.75 0.6 0.8 "Não" "" "B"
numberOfGoodnessCategories = 0
```

Anexo V

Código do *script* utilizado na preparação do teste de percepção com 2 respostas.

```
{}

"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
stimulusFileNameHead = "\manipulações\"
stimulusFileNameTail = ""
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1 seconds
stimulusMedialSilenceDuration = 0.0
numberOfDifferentStimuli = 13
"casa-1.wav,casa0.wav" ""
"casa+3.wav,casa0.wav" ""
"casa0.wav,casa-4.wav" ""
"casa-2.wav,casa0.wav" ""
"casa+6.wav,casa0.wav" ""
"casa0.wav,casa+5.wav" ""
"casa0.wav,casa-3.wav" ""
"casa0.wav,casa+1.wav" ""
"casa-5.wav,casa0.wav" ""
"casa+2.wav,casa0.wav" ""
"casa0.wav,casa-6.wav" ""
"casa+4.wav,casa0.wav" ""
"casa0.wav,casa0.wav" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 22
randomize = <PermuteBalancedNoDoublets>
startText = "Olá,
Este é um teste de percepção. A seguir serão apresentados pares de uma mesma senten
Elas tiveram seus tons manipulados, algumas ficaram mais graves
e outras mais agudas, ou seja, algumas parecem ser faladas por uma voz mais fina
e outras por uma voz mais grossa, embora elas sejam da mesma pessoa.
Sua tarefa será dizer se você nota ou não alguma diferença entre os pares.
Agora vamos ao teste.
Clique para continuar."
runText = "As sentenças são parecidas?"
pauseText = "Clique para continuar."
endText = "Fim do experimento.
Muito Obrigado."
maximumNumberOfReplays = 0
;replayButton = 0.4 0.8 0.02 0.08 "Nog een keer luisteren" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
```

```
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
0.25 0.4 0.6 0.8 "Sim" "" "1"
0.6 0.75 0.6 0.8 "Não" "" "2"
numberOfGoodnessCategories = 0
```

Anexo VI

Código do *script* utilizado na preparação do treino para o teste de percepção com 3 respostas.

```
{}

"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
stimulusFileNameHead = "\Manipulações\"
stimulusFileNameTail = ""
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1 seconds
stimulusMedialSilenceDuration = 0.0
numberOfDifferentStimuli = 3
"gato0.wav,gato-6.wav" ""
"gato+4.wav,gato0.wav" ""
"gato0.wav,gato0.wav" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 22
randomize = <PermuteBalancedNoDoublets>
startText = "Olá,
Este é um teste de percepção. A seguir serão apresentados pares de uma mesma senten
Elas tiveram seus tons manipulados, algumas ficaram mais graves
e outras mais agudas, ou seja, algumas parecem ser faladas por uma voz mais fina
e outras por uma voz mais grossa, embora elas sejam da mesma pessoa.
Sua tarefa será dizer se você nota ou não alguma diferença entre os pares.
Vamos fazer um treino antes de começar.
Clique para continuar"
runText = "As sentenças são parecidas?"
pauseText = "Clique para continuar."
endText = "Fim do Treino.Agora vamos ao experimento."
maximumNumberOfReplays = 0
;replayButton = 0.4 0.8 0.02 0.08 "Nog een keer luisteren" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 3
0.1 0.3 0.4 0.6 "Sim" "" "A"
0.4 0.6 0.4 0.6 "Talvez" "" "C"
0.7 0.9 0.4 0.6 "Não" "" "B"
numberOfGoodnessCategories = 0
```

Anexo VII

Código do *script* utilizado na preparação do teste de percepção com 3 respostas.

```
{}
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
stimulusFileNameHead = "\manipulações\"
stimulusFileNameTail = ""
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1 seconds
stimulusMedialSilenceDuration = 0.0
numberOfDifferentStimuli = 13
"gato-1.wav,gato0.wav" ""
"gato+3.wav,gato0.wav" ""
"gato0.wav,gato-4.wav" ""
"gato-2.wav,gato0.wav" ""
"gato+6.wav,gato0.wav" ""
"gato0.wav,gato+5.wav" ""
"gato0.wav,gato-3.wav" ""
"gato0.wav,gato+1.wav" ""
"gato-5.wav,gato0.wav" ""
"gato+2.wav,gato0.wav" ""
"gato0.wav,gato-6.wav" ""
"gato+4.wav,gato0.wav" ""
"gato0.wav,gato0.wav" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 22
randomize = <PermuteBalancedNoDoublets>
startText = "Olá,
Este é um teste de percepção. A seguir serão apresentados pares de uma mesma senten
Elas tiveram seus tons manipulados, algumas ficaram mais graves
e outras mais agudas, ou seja, algumas parecem ser faladas por uma voz mais fina
e outras por uma voz mais grossa, embora elas sejam da mesma pessoa.
Sua tarefa será dizer se você nota ou não alguma diferença entre os pares.
Agora vamos ao teste.
Clique para continuar"
runText = "As sentenças são parecidas?"
pauseText = "Clique para continuar."
endText = "Fim do experimento.
Muito Obrigado."
maximumNumberOfReplays = 0
;replayButton = 0.4 0.8 0.02 0.08 "Nog een keer luisteren" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
```

```
numberOfDifferentResponses = 3
0.1 0.3 0.4 0.6 "Sim" "" "1"
0.4 0.6 0.4 0.6 "Talvez" "" "3"
0.7 0.9 0.4 0.6 "Não" "" "2"
numberOfGoodnessCategories = 0
```

Anexo VII

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

Termo de consentimento para a participação voluntária em pesquisa.

Eu,.....
RG..... concordo em participar da pesquisa referente a dissertação de mestrado em Filologia e Língua Portuguesa do mestrando Marcus Vinícius Moreira Martins. A participação se resume a responder dois testes de percepção da entoação do Português Brasileiro, de maneira que, minha identidade (do participante) será preservada.
Considero que a pesquisa tem caráter científico e a participação é voluntária.

.....
Participante

.....
Mestrando - Marcus Vinícius Moreira Martins

.....
Orientadora - Profª. Drª. Beatriz Raposo de Medeiros

Data: ..../..../.....

Anexo VIII

Questionário de identificação

Idade: ..... Sexo: ( )Masculino ( )Feminino

Nacionalidade: .....

Profissão: .....

Formação/Escolaridade: ( ) Nível Superior ( )Pós-Graduação

( ) Completo ( ) Incompleto

Curso:.....

Monolíngue ( ) Bilíngue ( )

Faz uso profissional da voz: ( ) sim não ( )

Possui algum hobby? Se sim e ligado a música, qual?
.....

.....
.....
.....

Toca algum instrumento? Se sim, qual?
.....

.....
.....
.....