

UNIVERSIDADE ESTADUAL DO SUDOESTE DA BAHIA (UESB)
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA (PPGLIN)

PRISCILA DE JESUS RIBEIRO

**O PAPEL DA FREQUÊNCIA FUNDAMENTAL NA PERCEPÇÃO DO TIMBRE
VOCÁLICO: UMA AVALIAÇÃO EXPERIMENTAL DAS VOGAIS MÉDIAS**

VITÓRIA DA CONQUISTA - BA

2013

PRISCILA DE JESUS RIBEIRO

**O PAPEL DA FREQUÊNCIA FUNDAMENTAL NA PERCEPÇÃO DO TIMBRE
VOCÁLICO: UMA AVALIAÇÃO EXPERIMENTAL DAS VOGAIS MÉDIAS**

Dissertação apresentada ao Programa de Pós-Graduação em Linguística (PPGLIN), da Universidade Estadual do Sudoeste da Bahia (UESB), como requisito parcial e obrigatório para obtenção do título de Mestre em Linguística.

Área de Concentração: Linguística

Linha de Pesquisa: Descrição e Análise de Línguas Naturais

Orientadora: Profa. Dra. Vera Pacheco

VITÓRIA DA CONQUISTA

2013

Ribeiro, Priscila de Jesus.

R371p O papel da frequência fundamental na percepção do timbre vocálico: uma avaliação experimental das vogais média/Priscila de Jesus Ribeiro, 2013.
79f.: il; algumas col.

Orientador (a): Vera Pacheco.

Dissertação (mestrado) – Universidade Estadual do Sudoeste da Bahia, Programa de Pós-graduação em Linguística, Vitória da Conquista, 2013.

Referências: f. 76-79.

1. Vogais médias – Percepção da fala. 2. Fonologia. I. Pacheco, Vera. II. Universidade Estadual do Sudoeste da Bahia, Programa de Pós-graduação em Linguística. III. T.

CDD: 414

Catálogo na fonte: Elinei Carvalho Santana - CRB 5/1026

UESB – Campus Vitória da Conquista – BA

Título em inglês: The role of fundamental frequency in the perception of vowel tone: an experimental evaluation of the mean vowels

Palavras-chave em inglês: Middle vowels. Perception of the speech. Fundamental Frequency.

Área de concentração: Linguística

Titulação: Mestre em Linguística

Banca examinadora: Profa. Dra. Vera Pacheco (Presidente-Orientadora); Profa. Dra. Consuelo Paiva Godinho (UESB) Prof. Dr. Luiz Carlos Cagliari (UNESP)

Data da defesa: 18 de fevereiro de 2013

Programa de Pós-Graduação: Programa de Pós-Graduação em Linguística.

PRISCILA DE JESUS RIBEIRO

**O PAPEL DA FREQUÊNCIA FUNDAMENTAL NA PERCEPÇÃO DO TIMBRE
VOCÁLICO: UMA AVALIAÇÃO EXPERIMENTAL DAS VOGAIS MÉDIAS**

Dissertação apresentada ao Programa de Pós-Graduação em Linguística (PPGLIN), da Universidade Estadual do Sudoeste da Bahia (UESB), como requisito parcial e obrigatório para obtenção do título de Mestre em Linguística.

Data da apresentação: 18 de fevereiro de 2013.

BANCA EXAMINADORA



Profa. Dra. Vera Pacheco (UESB)
(Orientadora)

Profa. Dra. Consuelo de Paiva Godinho (UESB)



Prof. Dr. Luiz Carlos Cagliari (UNESP)

*“O céu anuncia a glória de Deus
e nos mostra aquilo que suas mãos fizeram ...
Não há discurso, nem palavras
e não se ouve nenhum som, no entanto, a voz do céu
se espalha pelo mundo inteiro
e as suas palavras alcançam a toda terra.” (Salmos 19:1,3)*

AGRADECIMENTOS

À Vera Pacheco, com quem tive, mais uma vez, o prazer de aprender, por suas críticas, sugestões e pela forma respeitosa e paciente com que me tratou durante toda construção deste trabalho, por contribuir significativamente para meu crescimento intelectual e para meu interesse pela Fonética e Fonologia.

A Capes, por apoiar financeiramente a concretização deste projeto.

A Deus, por abençoar imensamente a minha vida no decorrer de todo este trabalho e fazer-me vencer todos os obstáculos e dificuldades que surgiram nesse percurso. Pela graça de poder contemplar a excelência de sua grandeza revelada em tudo àquilo que Ele criou.

A Sisínio, meu pai, *in memoriam*, por todos os seus ensinamentos e pela forma que me amou.

A Minha mãe, uma grande amiga, por estar presente em todos os momentos da minha vida.

A Raquel e Lays, por serem amigas e companheiras, por suas palavras de incentivo e por partilharem comigo sonhos, alegrias e tristezas.

A Celton e Celmário, meus queridos primos, por estarem durante todo esse percurso, mesmo com tantos afazeres, dispostos a ajudar e a dispor do seu tempo.

Ao Johnny e ao Rodrigo, por sempre estarem dispostos a ajudar.

À Tia Marlene, por sua amizade e colaboração. Tia Lila, uma das pessoas mais divertidas da família, com ela risadas e diversão são garantidas, o que tornou os momentos de ansiedade muito mais leves. A tia Maria, por sua companhia sempre animada e pelos deliciosos lanches da tarde, quando os longos testes de percepção foram aplicados.

À Jussara Matos, grande amiga, que mesmo vivenciando tantas mudanças em sua vida e ainda estar em fase de adaptação a tantas novas situações, por dispor de seu tempo nas leituras e correções desta dissertação.

A meus colegas do mestrado, por todo conhecimento compartilhado, especialmente àlva, uma boa amiga que encontrei nesse percurso, com quem pude partilhar minhas ansiedades e que pacientemente soube ouvir e incentivar.

Ao Jefferson Lopes, pela amizade prazerosa e por sua disponibilidade de sempre.

Aos informantes, agradeço-lhes imensamente, por disporem voluntariamente do seu tempo e por contribuírem para a concretização desse trabalho.

Aos meus avós, Áurea, Hotalina e Arlindo, pelo exemplo de vida e coragem.

A Quindim, Mel, Blue, Tom, turminha da agitação e diversão. Quito, por suas “dancinhas” muito engraçadas.

Aos meus alunos, por me estimularem na busca de novos conhecimentos.

RESUMO

Estudos experimentais acerca de como os ouvintes identificam os sons vocálicos apontam que, além da frequência formântica, outros aspectos do espectro vocálico podem ser igualmente importantes para a compreensão de como se dá a discriminação vocálica. Considerando que a F_0 pode ser uma importante informação utilizada pelo ouvinte para a percepção de vários aspectos da fala, o presente trabalho propõe investigar se a frequência fundamental (F_0) interfere na percepção das vogais médias em ouvintes naturais de Vitória da Conquista – BA. Para isso, foi elaborado um *corpus* com palavras compostas por vogais médias. Em seguida, a F_0 das vogais em questão foi manipulada de forma a obter 25%, 50% e 75% do valor original, tanto para valores ascendentes quanto para valores descendentes. Os dados apontam que a manipulação desse parâmetro tanto para valores ascendentes quanto para valores descendentes afetaram a percepção das vogais em questão. Os resultados assinalam, portanto, que além da configuração formântica, a F_0 constitui outro parâmetro espectral envolvido no processo de percepção das vogais médias.

PALAVRAS CHAVE

Vogais médias. Percepção de fala. Frequência Fundamental.

ABSTRACT

Several experimental studies about how listeners identify vowel sounds point out that, in addition of the formant frequency, other aspects of the vowel spectrum are important to understand how the vowel discrimination happens. Considering that F_0 can be an important information used for perception of the some aspects of the speech, this article proposes to investigate whether the F_0 interferes on the perception of the middle vowels innative speakers from Vitória da Conquista - BA. To implement this, a corpus with words composed by high-middle and rounded and non-rounded vowels was set up. Each word was inserted in a sentence to obtain a significant context for the listener. Then, the vowel frequency was manipulated to obtain 25%, 50% e 75% from the original value, for both ascending and descending values. The data point out that manipulation of the vowel frequency affects the vowel perception. The results show that, beyond the formant configuration, other spectral parameters are involved in the perception process of vowel perception.

KEYWORDS

Middle vowels. Perception of the speech. Fundamental Frequency.

LISTA DE IFIGURAS

Figura 1 – Posição do trato vocal na produção da vogal /a/	17
Figura 2 – Posição do trato vocal na produção da vogal /i/.....	18
Figura 3 – Posição da Língua em /i/ e em /u/ traço ponteados	18
Figura 4 – Trato vocal comparado a um tubo de 17,5 cm.....	19
Figura 5 – Forma assumida pelo trato vocal na produção da vogal /a/	23
Figura 6 - Forma assumida pelo trato vocal na produção da vogal /i/.....	24
Figura 7 - orma gerada pelo sinal acústico de fala para a vogal /a/ e /ã/	35
Figura 8 – Espectro gerado por Stevens e Blumstein (1979) para as oclusivas bilabial, alveolar e velar	36
Figura 9 - Praat – manipulação da F_0	48
Figura 10 - Manipulação – ampliação da F_0	48
Figura 11 - Manipulação - redução da Manipulação - redução da F_0	49
Figura 12 - Transcriber – Aplicação dos testes de percepção	50

LISTA DE GRÁFICOS

Gráfico 1 – Índice de erro de discriminação da vogal média baixa não arredondada para ampliação da F_0	54
Gráfico 2 – Índice de erro de discriminação da vogal média alta para valores ampliados da F_0	56
Gráfico 3 – Índice de erro de discriminação da vogal média baixa arredondada para ampliação da F_0	57
Gráfico 4 – Índice de erro de discriminação da vogal média alta arredondada para ampliação da F_0	58
Gráfico 5 – Índice de erro de discriminação da vogal média baixa não arredondada para valores reduzidos da F_0	60
Gráfico 6 – Índice de erro de discriminação da vogal média alta não arredondada para valores reduzidos da F_0	61
Gráfico 7 – Índice de erro de discriminação da vogal média baixa arredondada para valores reduzidos da F_0	62
Gráfico 8 – Índice de erro de discriminação da vogal média arredondada para valores reduzidos da F_0	63
Gráfico 9 – Índice de erro de identificação da vogal média baixa não arredondada para valores ampliados da F_0	65
Gráfico 10 – Índice de erro de identificação da vogal média alta não arredondada para valores ampliados da F_0	66
Gráfico 11 – Índice de erro de identificação da vogal média baixa arredondada para valores ampliados da F_0	67
Gráfico 12 – Índice de erro de identificação da vogal média alta arredondada para valores ampliados da F_0	68
Gráfico 13 – Índice de erro de identificação da vogal média baixa não arredondada para valores reduzidos da F_0	69
Gráfico 14 – Índice de erro de identificação da vogal média alta não arredondada para valores reduzidos da F_0	70
Gráfico 15 – Índice de erro de identificação referente à vogal média baixa arredondada para valores reduzidos da F_0	71

Gráfico 16 – Índice de erro de identificação referenta à vogal média alta arredondada para valores reduzidos da F_0	72
---	----

SUMÁRIO

1 INTRODUÇÃO	14
2 ARCABOUÇO TEÓRICO	16
2.1 DESCRIÇÃO ACÚSTICA E ARTICULATÓRIA DOS SEGMENTOS VOCÁLICOS	16
2.1.1 Descrição articulatória dos segmentos vocálicos	16
<i>2.1.1.1 Teoria linear de produção de fala fonte-filtro.....</i>	<i>18</i>
2.1.2 Descrição acústica dos segmentos vocálicos	22
<i>2.1.2.1 Padrões formânticos.....</i>	<i>22</i>
<i>2.1.2.2 Disposição espectral.....</i>	<i>25</i>
<i>2.1.2.3 Duração.....</i>	<i>26</i>
<i>2.1.2.4 Frequência fundamental (F_0).....</i>	<i>26</i>
<i>2.1.2.5 Intensidade.....</i>	<i>29</i>
2.2 O SISTEMA VOCÁLICO DO PORTUGUÊS BRASILEIRO (PB).....	30
2.3 TEORIAS DE PERCEPÇÃO DE FALA	31
2.3.1 Teoria motora	31
2.3.2 Teoria quântica.....	34
2.4 PERCEPÇÃO DE VOGAIS.....	37
2.4.1 Teoria de especificação intrínseca.....	37
2.4.2 Teoria de especificação extrínseca	39
2.5 EXEMPLOS DE MODELOS DE PERCEPÇÃO DE VOGAIS.....	39
2.5.1 Simple Target Model.....	39
2.5.2 Elaborated Simple Target Model	41
2.5.3 Dynamicspecification	42
3 METODOLOGIA.....	45
3.1 INFORMANTES	45
3.2 SELEÇÃO DE PALAVRAS E PREPARAÇÃO DAS FRASES.....	45
3.3 GRAVAÇÃO DAS FRASES E PREPARAÇÃO DO ESTÍMULO: A MANIPULAÇÃO DA F_0	47
3.4 ELABORAÇÃO DOS TESTES DE PERCEPÇÃO	49
3.5. AS TAREFAS DOS TESTES DE PERCEPÇÃO	50
3.5.1 Teste de discriminação	51
3.5.2 Teste de identificação	51
3.6 PROCEDIMENTOS PARA APLICAÇÃO DOS TESTES DE PERCEPÇÃO	52
4 RESULTADOS E DISCUSSÃO	54

4.1 TESTE 1: DISCRIMINAÇÃO	54
4.1.1 Análise dos testes de discriminação com valores de F_0 ampliados a 25%, 50% e 75%.	54
4.1.2 Análise dos testes de discriminação com valores de F_0 reduzidos a 25%, 50% e 75%.	59
4.2 TESTE 2: IDENTIFICAÇÃO	64
4.3.2 Análise dos testes de identificação para valores reduzidos da F_0.....	69
4.4 DISCUSSÕES	73
5 CONSIDERAÇÕES FINAIS.....	75
REFERÊNCIAS	76

1 INTRODUÇÃO

Uma das questões de principal interesse da fonética experimental consiste em compreender quais são as informações utilizadas pelos ouvintes para a identificação das vogais. Alguns estudiosos (HILLENBRAND; GAYVERT, 1993; JENKINS et al., 1983; TUERNER; PATTERSON, 2003) propõem que a percepção dos segmentos vocálicos está diretamente relacionada às frequências dos dois formantes mais baixos (F_1 e F_2). Esses autores afirmam que há, no estado estacionário, um espaço acústico entre as frequências formânticas que determina a percepção vocálica.

Outros pesquisadores (BERGEM et al., 1988; CLAES, et al., 1998; JOOS, 1972; STRANGE, 1995) adotam a hipótese de que, além das frequências dos dois formantes mais baixos, a discriminação de vogais está também relacionada a outras informações espectrais. Dentre esses estudiosos, pode-se citar Strange (1995) que propõe integrar o estado estacionário e a natureza da transição formântica com outras informações espectrais. A autora propõe que, juntamente com as frequências formânticas, a duração e a frequência fundamental corroboram para a percepção vocálica.

Estudos que, além da frequência formântica, consideram outros aspectos do espectro vocálico apontam que as demais informações espectrais podem ser igualmente importantes para a compreensão de como se dá a discriminação vocálica. Dentre esses estudiosos, podem-se citar pesquisas que, utilizando estímulos de vogais sintéticas em que a frequência fundamental (F_0) foi sistematicamente manipulada, evidenciam que a percepção vocálica pode estar relacionada à informação acústica da F_0 . Barreda e Nearley (2011) afirmam que a relação da F_0 na percepção vocálica é indireta. De acordo com os autores, o ouvinte usa a F_0 como recurso para identificar as características de idade e sexo do falante, mais do que para a identificação da qualidade vocálica. Por outro lado, Ryalls e Liberman (1982) relatam que a F_0 pode influenciar de forma direta a percepção vocálica. Segundo os autores, pesquisas utilizando vogais sintetizadas com valores formânticos correspondentes aos padrões utilizados por homens, mulheres e crianças, mas com F_0 manipulada tanto para valores superiores quanto inferiores ao padrão apontam que tal variação influenciou o modo como os ouvintes perceberam as vogais. Essas pesquisas, segundo os autores, corroboram com a hipótese de que a F_0 pode influenciar diretamente a percepção vocálica.

Considerando que a F_0 pode ser uma importante informação utilizada pelo ouvinte para a percepção da vocálica e considerando que as vogais médias possuem padrões formânticos com valores aproximados, o presente trabalho propõe investigar se a F_0 interfere

e o quanto interfere na percepção das vogais médias em ouvintes naturais de Vitória da Conquista – BA. Propõe-se investigar se a manipulação da F_0 , tanto para valores ascendentes quanto para valores descendentes, interfere e o quanto interfere na perceptibilidade das vogais médias no dialeto dessa comunidade linguística.

A hipótese a ser testada é que a alteração da F_0 compromete a inteligibilidade das vogais médias de sorte a confundir o timbre vocálico a ser percebido.

2 ARCABOUÇO TEÓRICO

2.1 DESCRIÇÃO ACÚSTICA E ARTICULATÓRIA DOS SEGMENTOS VOCÁLICOS

As ondas sonoras são produzidas ao nível da glote e na cavidade supraglóticas incluindo a cavidade oral e a cavidade velofaríngea que funcionam como ressoadores (KENT; READ, 1992). Esses sons, que saem do aparelho fonador, podem ser produzidos com algum tipo de obstrução na glote ou nas cavidades supraglóticas de maneira que pode ocorrer obstrução total ou parcial da passagem da corrente de ar; por outro lado, a corrente de ar pode não ser interrompida na linha central (KENT; READ, 1992).

Os sons pronunciados com passagem de ar livre chamam-se vogais, neste caso, as cavidades supraglóticas só fazem modificar o timbre do som laríngeo. Já os sons produzidos com algum tipo de obstrução, seja ela total ou parcial, chamam-se consoantes, devido à formação, na cavidade supraglotal, de diferentes espécies de ruídos (KENT; READ, 1992).

De acordo com Kent e Read (1992), os modernos métodos de eletroacústica permitem aos foneticistas analisar qualquer som linguístico e apresentar o resultado da análise sob forma de um espectro. Como cada som tem um espectro próprio, é possível, com o auxílio de um filtro, tornar visíveis os espectros vocálicos.

Os sons vocálicos estão diretamente associados a um estado estacionário de articulação, bem como a um estado estacionário acústico. Assim, podem ser caracterizados como sendo um fenômeno acústico e articulatório. A seguir segue-se uma breve discussão a cerca das características articulatória dos segmentos vocálicos.

2.1.1 Descrição articulatória dos segmentos vocálicos

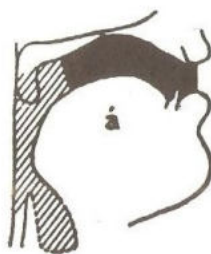
Os segmentos vocálicos podem ser descritos levando-se em consideração os seguintes aspectos articulatórios: papel da cavidade oral e nasal, zona de articulação e arredondamento dos lábios (MALMBERG, 1954).

Durante a articulação de uma vogal, o véu palatino pode estar levantado ou abaixado. Se o véu palatino estiver levantado, a corrente de ar passará somente pela cavidade oral, produzindo assim, as *vogais orais*, do contrário, se o véu palatino estiver abaixado, o fluxo de ar penetrará a cavidade nasal, sendo expelido pelas narinas e pela cavidade oral concomitantemente, produzindo assim uma *vogal nasal*. (MALMBERG, 1954).

A *zona de articulação* está diretamente relacionada à região da cavidade oral onde as vogais são articuladas. Segundo Ladefoged (1996), os parâmetros básicos para a descrição dos sons vocálicos das diferentes línguas do mundo são as três dimensões tradicionalmente chamadas de: alta e baixa, anterior e posterior, arredondada e não arredondada. Estes termos são usados para descrever as características dos sons vocálicos em concordância com a altura da língua e forma do trato vocal. Segundo Ladefoged (1996), a grande maioria das línguas do mundo tem uma relação previsível entre as dimensões de anterioridade e arredondamento, assim, vogais anteriores são usualmente não arredondadas – para o quê, o francês é um contraexemplo - e vogais posteriores geralmente são arredondadas – para o quê, o inglês é um contraexemplo. Deste modo, são definidas como:

Vogal central: articulada com levantamento do dorso da língua em direção ao palato médio. Nessa ocorrência, a cavidade oral está dividida em duas caixas de ressonância de tamanhos semelhantes, ficando os dois formantes principais ocupando o meio do espectro - no caso da vogal /ə/ a cavidade oral 1300HZ e cavidade faríngea cerca de 720 HZ, trata-se, portanto, de uma vogal acusticamente compacta¹;

Figura 1 – Posição do trato vocal na produção da vogal /a/

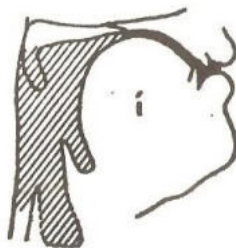


(cf. MALMBERG, p. 62, 1954)

Vogais anteriores: articuladas com a língua elevada em direção ao palato duro, próxima aos dentes incisivos superiores, como por exemplo, na produção da sequência fônica [ʌ♦♦ʌ⊕]. Na produção das vogais anteriores, a língua se direciona ao palato duro gerando, assim, uma diminuição do volume da boca e um consequente aumento na parte posterior da obstrução, desse modo o valor do padrão formântico mais elevado sobe e o formante mais baixo diminui, no caso do [ʌ] a F₂ sobe até 2500 Hz e F₁ diminui até 280 Hz.

¹ Classificação fonológica gerativa proposta por Jakobson, Fant e Halle (1952).

Figura 2 – Posição do trato vocal na produção da vogal /i/



(cf. MALMBERG, p. 62, 1954)

Vogais posteriores: articuladas quando a língua se dirige ao véu palatino, dividindo a cavidade oral em duas caixas de ressonância de tamanhos diferentes - a anterior bucal em maior extensão e a outra cavidade posterior faríngea em menor extensão, como exemplificado na figura. Essa posição abaixa o tom próprio da cavidade oral e reforça o tom laríngeo.

Figura 3 – Posição da Língua em /i/ e em /u/ traço pontilhado



(cf. MALMBERG, p. 62, 1954)

Durante a articulação de um segmento vocálico, os lábios podem estar estendidos [ɸ]; [ɱ]; [ɔ̃]; [ɰ] ou arredondados [ɸ̹]; [ɱ̹]; [ɔ̹̃]; [ɰ̹]. Já o timbre está relacionado com a abertura da boca: têm-se vogais abertas [ɔ̃]; [ɰ̹]; [ɸ̹] e fechadas [ɱ̹]; [ɱ̹]; [ɰ̹]; [ɰ̹].

2.1.1.1 Teoria linear de produção de fala fonte-filtro

A Teoria Acústica de Produção de Fala, desenvolvida por Fant (1960), busca explicar como os sons da fala são gerados e modificados. Segundo o autor²,

²A onda sonora da fala é uma resposta do sistema de filtro do trato vocal de uma ou mais fontes sonoras. Esta regra simples, expressa na terminologia da engenharia acústica e elétrica, implica que a onda sonora da fala pode ser exclusivamente especificada em termos de características de fonte e filtro.

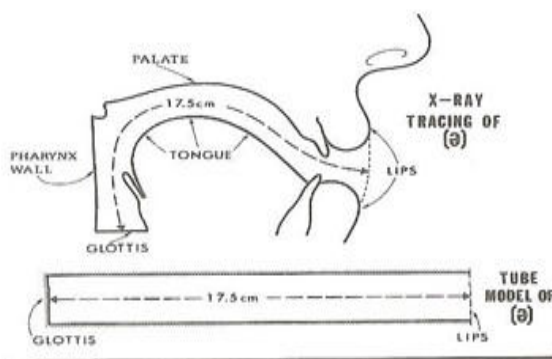
The speech wave is the response of the vocal tract filter system to one or more sound sources. This simple rule, expressed in the terminology of acoustical and electrical engineering, implies that the speech wave may be uniquely specified in terms of source and filter characteristics (FANT, 1960, p. 15).

Essa teoria visa compreender a relação entre os aspectos acústico-articulatórios envolvidos na produção dos sons da fala, relacionando uma propriedade acústica específica a um determinado correlato articulatório. O principal objetivo do autor com essa teoria é contribuir para uma melhor compreensão de como se dá as relações entre os aspectos articulatórios de produção de fala e os aspectos físicos da descrição da onda sonora.

Essa teoria, também conhecida como teoria fonte-filtro, baseia-se em um modelo linear no qual a fonte de energia é independente do filtro, ou seja, assume que qualquer som produzido pelo aparelho fonador é resultado de uma *fonte* de ruído (prega vocal ou ainda a constrição do trato vocal em algum ponto de sua extensão³) que permite a entrada acústica no sistema, e de um *filtro* (trato vocal) que modifica essa fonte e realiza os mais variados sons existentes nas línguas naturais.

Nesse modelo, os sons vocálicos são representados por meio de um aparato que consiste de um vibrador (membrana elástica com um corte estreito ao meio) e um tubo reto. Entende-se por vibrador uma fonte de energia acústica que se propaga através de um tubo, o qual funciona como um ressoador. Assim, nesse modelo, o trato vocal é representado por um tubo fechado em uma extremidade e aberto na outra, como mostra a figura 1.

Figura 4—Trato vocal comparado a um tubo de 17,5 cm



(cf. PICKETT, p. 36, 1998)

Uma excitação na membrana provocará a vibração das partículas de ar do tubo e produzirá energia sonora. O tubo funcionará como um ressoador ou filtro gerando uma onda

³ Ruídos fricativos

sonora, cujas características dependerão da forma e tamanho desse. Segundo Kent Read (1992), para que a comparação do tubo com o trato vocal seja relevante, é necessário considerar dois importantes fatores: 1) a média de comprimento do trato vocal de um homem adulto que é de aproximadamente 17.5 cm, compreendido entre a glote e os lábios e 2) o trato vocal tem aproximadamente a mesma frequência de ressonância de um tubo, ou seja, o mesmo comprimento e área da secção transversal.

De acordo com esse modelo, cada tubo possui um número infinito de ressonâncias, dado pela seguinte fórmula matemática:

$$Fn = (2n-1)c / 4l$$

Onde:

- n é um número inteiro

- c corresponde à velocidade do som (aproximadamente 35.000 cm/seg), e

- l , ao tamanho do tubo

Essa fórmula nos diz que um tubo ressoa com máxima amplitude um som, cujo comprimento de onda é quatro vezes o comprimento do tubo. Assim, um tubo de 17.5 cm (média para o trato vocal masculino) ressoará as seguintes frequências: $F1 = (2.1-1)c / 4.17,5 = 500$ Hz; $F2 = (2.2-1)c / 4.17,5 = 1500$ Hz, etc. Considerando um tubo de 35.0 cm, dobro do valor anterior, as frequências ressoadas assumirão valores muito mais baixos, a saber, 250, 750, 1250 e 1750 Hz. Por outro lado, um tubo de 8.75 cm, metade do valor referido acima, por exemplo, terá as seguintes frequências ressoadas: 1000, 3000, 5000 e 7000 Hz. (KENT; READ, 1992). O valor relativo da frequência é determinado pelo tamanho do trato vocal. Sendo assim, uma criança possui frequências mais altas do que um adulto, pois possui um trato vocal bem menor, já os homens, por apresentarem trato vocal mais longo, possuem, no geral, frequências mais baixas do que uma mulher e uma criança. As frequências de ressonância variam conforme as características do falante, sendo os principais fatores determinantes a idade e o sexo.

Esses diferentes grupos de frequências, determinados pelas caixas de ressonância que reforçam o som laríngeo no harmônico necessário para que se obtenha o timbre desejado, são chamados de *formantes*. Segundo Kent e Read (1992), os formantes são um modo natural de vibração (ressonância) do trato vocal, ou seja, são faixas de frequências em que ocorre elevação da amplitude dos componentes espectrais em decorrência da ressonância do trato vocal. Nesse sentido, podemos dizer que a concentração de energia acústica em determinadas

faixas de frequência correspondem aos formantes. Cada pico de frequência reforçada pela ressonância do espectro representa um formante que são conhecidos como F_1 , F_2 , F_3 , F_n . A frequência do primeiro formante (F_1) está relacionada à posição da língua no plano vertical e é influenciada pelo grau de abertura da boca, a frequência do segundo formante (F_2) está relacionada à posição da língua no plano horizontal (grau de anterioridade) (KENT; READ 1992). De acordo com os autores, teoricamente há um número infinito de formante, no entanto, somente os três ou quatro primeiros (F_1 , F_2 , F_3) são suficientes para caracterizar acusticamente os vários sons da fala humana.

Segundo Ladefoged (1996), o termo “formantes” é usado para definir os picos de onda produzidos pela vibração das pregas vocais. Segundo ele, esses picos correspondem às frequências formânticas das vibrações do ar no trato vocal. Nessa perspectiva, formantes referem-se à configuração do aparelho fonador na produção dos sons da fala, ou seja, diz respeito às propriedades correspondentes à posição dos articuladores.

É possível perceber na figura 2 que o trato vocal assume diferentes tipos de constrição para a produção dos sons vocálicos. Assim, na produção da vogal [ɪ], a cavidade oral apresenta uma constrição pré-palatal e uma expansão próxima à faringe; na vogal [ʊ] é possível observar que há uma expansão tanto na região próxima a abertura labial quanto na região próxima à faringe e à laringe; na vogal [ʌ] há uma constrição na parte posterior da cavidade oral, uma interrupção na parte central e um arredondamento dos lábios que pode ser visualizado por meio da constrição nas partes posterior e anterior do tubo. As frequências de ressonância dos sons produzidos pelo trato vocal são determinadas pelos diferentes tipos de constrição do trato que podem ser modificados tanto no sentido horizontal quanto no sentido vertical.

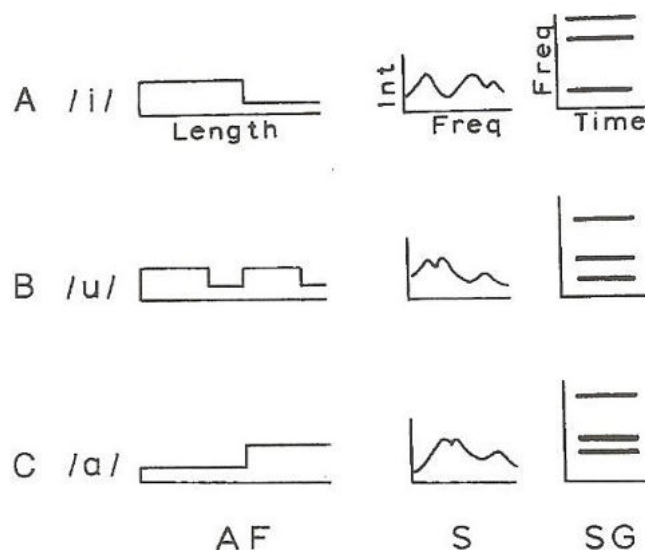


Figura 5: Tubo correspondente à forma assumida pelo trato vocal na produção de /i/, /u/ e /a/ e as respectivas informações espectrais: duração e intensidade (cf. KENT; READ, 1992, p.21).

2.1.2 Descrição acústica dos segmentos vocálicos

Considerando os diferentes modelos de percepção vocálica, os parâmetros candidatos para a descrição acústica das vogais são: padrões formânticos, espectro, duração e frequência fundamental (KENT; READ, 1992).

2.1.2.1 Padrões formânticos

Pesquisas em análise de fala apontam que a configuração formântica é um importante recurso utilizado pelo ouvinte para a identificação das vogais. De acordo com Kent e Read (1992), essas pesquisas mostram que quando as vogais são sintetizadas usando as frequências formânticas típicas da fala natural os resultados são satisfatórios. Segundo os autores, uma vez que os valores dos padrões formânticos das vogais de fala natural são usados para a síntese de fala, os formantes são de especial importância para a caracterização acústica do sistema vocálico de uma dada língua natural.

Embora seja notável a importância da frequência formântica para a caracterização acústica das vogais, alguns experimentos, também baseados na síntese vocálica, lançam dúvidas a cerca do papel dos formantes para a discriminação vocálica. Pode-se citar um experimento realizado por Carlson, Fant e Granstom (1998), no qual foi analisada a

identificação de várias combinações das frequências de F_1 e F_2 . Os autores realizaram um estudo em que o F_1 foi mantido dentro dos padrões da fala natural e o F_2 foi experimentalmente variado, incluindo valores acima do nível esperado para fala natural. Os resultados apontam que embora os valores de F_2 das vogais estivessem acima dos valores da fala natural, a percepção dessas não foi prejudicada.

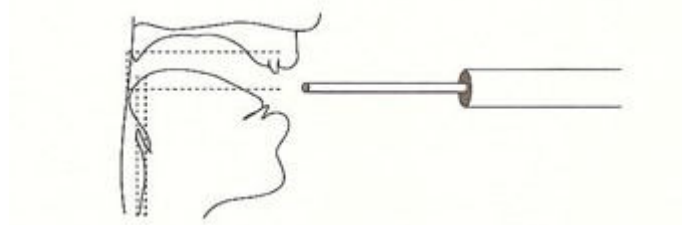
Segundo Kent e Read (1992), os formantes são a melhor informação acústica para a descrição das vogais. De acordo com os autores, este fato deve-se a diversos fatores: 1) os formantes são utilizados como base para a produção de fala sintética; 2) são econômicos, uma vez que, em muitos casos, para a produção de determinado som é necessário apenas considerar os três primeiros mais baixos; 3) são contínuos em relação às consoantes vizinhas; 4) são de fácil visualização no espectro acústico de fala. De modo geral, portanto, é possível classificar acusticamente as vogais considerando apenas a frequência formântica, uma vez que as frequências dos dois formantes mais baixos são suficientes para caracterizar os sons vocálicos.

Considerando que os formantes principais (F_1 e F_2) podem ocupar diferentes faixas no espectro, a saber, o meio ou as extremidades, as vogais podem ser classificadas em dois tipos acústicos: compacto e difuso⁴.

a) Tipo compacto

Na produção do tipo compacto, podemos considerar o trato vocal como dois tubos⁵, um estreito na faringe e um mais largo na cavidade oral, como ilustrado na figura 3.

Figura 5—Forma assumida pelo trato vocal na produção da vogal /a/



(cf. LADEFOGED, p.123, 1996).

⁴ Classificação fonológica gerativa proposta por Jakobson, Fant e Halle (1952).

⁵ Abaixo segue uma breve descrição da teoria fonte-filtro.

Como temos dois tubos de tamanhos semelhantes⁶, os dois primeiros formantes (F_1 e F_2) ocupam posição neutra na faixa em que se encontram os três primeiros formantes do espectro, um lugar intermediário, como é o caso do vogal [ɐ], ilustrada na figura 1.

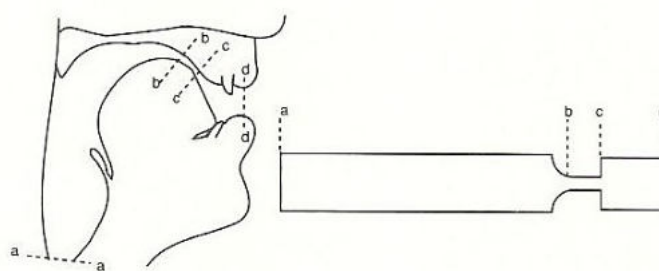
Considerando que o trato vocal possui em média 17.5 cm, e que nesse modelo a produção da vogal [ɐ] é metade do comprimento do trato vocal, suas frequências de ressonâncias serão o dobro da frequência de ressonância do comprimento total do tubo. Assim, teremos frequências de 1000, 3000 Hz, etc. No tipo compacto, portanto, os formantes F_1 e F_2 ocupam posição central no espectro (LADEFOGED, 1996).

b) Tipo difuso

No tipo difuso, o trato vocal está configurado de forma a constituir duas caixas de ressonância de diferentes tamanhos, gerando assim uma informação espectrográfica na qual os dois formantes mais baixos encontram-se nitidamente separados um do outro, cada um ocupando uma das extremidades do espectro. É o caso das vogais [ɤ] e [ɔ].

Na produção da vogal [ɤ], a parte anterior da língua fica levantada em direção ao palato duro, formando um canal estreito, cuja configuração lembra o “gargalo de uma garrafa”. Essa constrição da língua produz um pequeno volume de ar atrás da boca e um amplo volume de ar na faringe, dividindo o trato vocal em dois tubos com diferentes extensões, conforme ilustrado na figura 4.

Figura 6 - Forma assumida pelo trato vocal na produção da vogal /i/



(cf. LADEFOGED, p.127, 1996).

⁶Podemos observar que um tubo é bem mais largo do que o outro, mas matematicamente esse fato é insignificante, uma vez a frequência de ressonância depende somente do comprimento do tubo.

No caso da vogal [◆], a língua está posicionada para trás, em direção ao véu palatino, dividindo o trato vocal em duas caixas de ressonância: a cavidade oral anterior, mais extensa e a cavidade faríngea posterior, mais estreita.

Estes tipos são basicamente os mesmos para todas as línguas do mundo, no entanto, cada uma dessas utiliza apenas um número restrito de todas as possibilidades vocálicas do trato vocal. Algumas línguas possuem apenas as oposições citadas acima (⋈, ◆, ☉), outras, por sua vez, possuem um sistema vocálico mais extenso, com exemplo o Português Brasileiro que possui uma série de vogais agudas [☞], [ℳ] [⋈] e o Francês que possui além da série aguda portuguesa, outra série de vogais agudas, [⊕] e [⊗]. Há línguas que, assim como o português, possuem apenas uma série de vogais agudas, como exemplo, o italiano e o espanhol e outras que possuem grau intermediário entre o grave e o agudo, chamadas vogais mistas, como exemplo, o inglês e o grego. (MALMBERG, 1954).

O sistema vocálico do Português Brasileiro é caracterizado por apresentar uma dupla oposição vocálica:

a	i	u	
Compacto	+-	-	
Difuso	-+	+	
Flat	--	+	(labializado ou “grave”)

O Português Brasileiro e o Francês, como lembra Malmberg (1954, p. 29), dentre um número pequenos de línguas, conta ainda com uma série de vogais nasais, cuja peculiaridade é o fato de possuírem um formante especial e por uma modificação “dos outros formantes em relação à vogal oral correspondente”.

2.1.2.2 *Disposição espectral*

O sistema vocálico de uma dada língua natural pode também ser descrito considerando o espectro. Segundo Kent e Read (1992), para alguns autores a informação espectral propicia uma caracterização mais completa das vogais do que apenas os padrões formânticos. Assim, para alguns pesquisadores, o espectro é de fundamental importância para a descrição acústica das vogais, pois este além de nos fornecer os padrões formânticos de uma dada vogal, fornece-nos também informações adicionais de importância crucial para a distinção vocálica.

Estudos comparativos a cerca da informação formântica e espectral apontam, segundo Kent e Read (1992), que o espectro fornece um grupo de parâmetros acústicos de fundamental

importância para a descrição acústica das vogais, a saber, *duração*, *frequência fundamental* e *intensidade*. Abaixo segue uma pequena descrição a cerca desses correlatos.

2.1.2.3 Duração

Os segmentos vocálicos podem ser identificados a partir dos valores relativos à duração. Segundo Kent e Read (1992), embora, muitas vezes, a duração seja negligenciada no gráfico tradicional de F_1 e F_2 , esta sempre deve ser avaliada como sendo um ponto importante no sinal acústico de fala, uma vez que, muitas línguas exploram a duração como sendo uma característica distintiva da vogal, como exemplo, o inglês, alemão, sueco, russo, etc.

As vogais possuem diferenças inerentes quanto a sua duração. Dentre os fatores que influenciam a duração pode-se citar: o traço tenso ou relaxado, altura, acento, velocidade de fala, ponto de articulação, qualidade do segmento anterior ou posterior, fatores sintáticos ou semânticos. Alguns desses fatores são inerentes à duração outros são determinados pelas propriedades suprasegmentais ou pelo contexto fonético (KENT; READ, 1992).

Estudos apontam que, embora a duração não seja suficiente para a discriminação de vogais de uma dada língua natural, essa constitui uma importante informação para o ouvinte na discriminação de vogais que possuem espectros semelhantes, tais como // versus // ou para vogais que pertencem à categoria tensa *versus* relaxada. Hillenbrand, Clark e Houde (2000) concluíram, por meio de um experimento usando a síntese de fala, que os efeitos da duração na percepção de vogais são de fundamental importância para um determinado grupo de vogais, sendo essas // - // - // e o par // - //.

2.1.2.4 Frequência fundamental (F_0)

Outra questão de grande importância para a caracterização acústica das vogais é a F_0 . Mateus (2004) define F_0 como sendo a frequência com que os impulsos glotais se sucedem, ou seja, o número de ciclos que as pregas vocais realizam por segundo. Segundo a autora, a frequência fundamental de um ponto de vista articulatório relaciona-se com as pregas vocais, uma vez que, quanto mais delgadas, maior o número de vibrações e, portanto, maior a altura do som.

A frequência fundamental do sinal de fala corresponde, portanto, à medida física da taxa de vibração das pregas vocais, assim, quanto maior o número de vibrações mais aguda é a voz e mais alta a frequência e quanto mais lenta a vibração mais grave é a voz. Segundo

Titze (1992), as pregas vocais de um homem são geralmente cerca de 60% mais espessas do que as cordas vocais femininas, o que gera uma frequência fundamental reduzida em relação aos valores da F_0 de uma voz feminina.

Na perspectiva da teoria acústica de produção de fala, o correlato acústico correspondente às características referentes à idade e sexo dos falantes é atribuído a Frequência Fundamental (F_0) e as Frequências formânticas. Pode-se citar pesquisas realizadas por Coleman (1976), GelferandMikos (2005) e Whiteside (1998) que realizaram um estímulo organizado em quatro diferentes tipos: (1) a F_0 de uma voz masculina foi atrelada às frequências formânticas de uma voz masculina, (2) a F_0 de uma voz feminina foi vinculada às frequências formânticas de uma voz feminina, (3) a F_0 de uma voz masculina foi vinculada às frequências formânticas de uma voz feminina, e por fim, (4) a F_0 de uma voz masculina foi vinculada aos valores das frequências formânticas de uma voz feminina.

Os dados apontaram, segundo os autores, que quando a F_0 e as frequências formânticas foram equivalentes a um determinado gênero de fala natural, a taxa de identificação para o sexo masculino foi superior ao feminino. Quando a F_0 e as frequências formânticas não foram equivalentes a um determinado sexo, os resultados dos testes de percepção apontaram discrepâncias. Gelfer e Mikos (2005), assim como, Whiteside (1998) concluíram que os ouvintes utilizaram mais a Frequência Fundamental do que as frequências formânticas, para a identificação de falantes. Pesquisas realizadas por Siu-FungPoon e Manwa L. Ng (2011) apontaram que a F_0 foi o principal recurso utilizado pelo ouvinte para a percepção de gênero.

Uma das questões de principal interesse da fonética experimental é compreender se essas informações referentes às características de idade e sexo dos falantes interagem ou não com as informações acústicas referentes à qualidade vocálica. Alguns pesquisadores defendem a hipótese de que a F_0 é um recurso utilizado pelo ouvinte apenas para a identificação de falantes (idade e sexo), outros defendem que a F_0 é um recurso utilizado pelo ouvinte também para a percepção da qualidade vocálica.

Não há, portanto, consenso entre os pesquisadores a cerca de como se dá a interação entre Frequência Fundamental e qualidade acústica vocálica. Alguns estudiosos como Syrdal e Gopal(1986) e Miller (1989) propõem que a F_0 está diretamente relacionada com a qualidade acústica da vogal. Outros autores, a exemplo deJohnson(1990) defendem o pressuposto de que não existe uma relação direta entre a F_0 e a qualidade vocálica, estes argumentam a favor de uma independência da F_0 e da qualidade acústica dos segmentos vocálicos. Há ainda autores (SMITH et al., 2005; Jenkinset al. 1983) que defendem a

hipótese de que a F_0 se relaciona com a qualidade vocálica somente pelo fato de essa contribuir para a determinação das características de um possível falante.

Nessa perspectiva, há diferentes teorias referentes à relação que a F_0 estabelece com a qualidade acústica dos sons vocálicos. Essas teorias dividem-se em: direta, indireta e independente⁷.

Teoria indireta

A teoria indireta defende a hipótese de que a F_0 é um importante recurso utilizado pelo ouvinte para a identificação das características de idade e sexo do falante, mais do que para a especificação da qualidade vocálica. De acordo com essa teoria, a F_0 se relaciona com a qualidade vocálica somente pelo fato de essa contribuir para a determinação das características de um falante presumível.

Dentre os defensores desse modelo, pode-se citar Johnson (1990) que defende a hipótese de que o ouvinte cria uma representação mental dos segmentos percebidos e, com base nesse modelo preestabelecido, interpreta os sons da fala e Nearey (1989), o qual postula que a normalização de vogais envolve o uso de vogais de qualidade acústica já conhecida pelo ouvinte por meio da qual outras vogais são discriminadas.

Teoria independente

Essa teoria defende o pressuposto de que não existe uma relação entre a F_0 e a qualidade acústica vocálica. Dentre os estudiosos que defendem essa hipótese, podem-se citar Irino e Peterson (2002) que argumentam a favor de uma independência da F_0 e da qualidade acústica vocálica. Esses pesquisadores, utilizando experimentos em que o envelope espectral e a F_0 foram alterados separadamente, apontam que mudanças na F_0 não afetaram a percepção da qualidade acústica vocálica.

Postula-se nessa teoria que o sistema auditivo segrega informações de idade e sexo da informação referente à forma do trato vocal, possibilitando, assim, a discriminação dos sons da fala, bem como a percepção de características do falante. De acordo com os autores, em um estágio precoce do processamento auditivo, os ouvintes transformam o *input* acústico em uma imagem acústica de forma e tamanho invariáveis, na qual os padrões espectrais de um

⁷Traduções da autora para: Indirect F_0 theories, F_0 -free theories, Indirect F_0 theories. In: BARREDA, S.; NEAREY, T. M., 2011.

som ficam representados. Nesse ponto de vista, mudanças na F_0 não desempenham nenhum papel na determinação da qualidade da vogal.

Teoria Direta

Essa teoria baseia-se no fato de que o ouvinte é capaz de distinguir precisamente os sons vocálicos produzidos por diferentes falantes e ainda distinguir quem os produziu (homem, mulher ou criança) com base na covariância existente entre a F_0 e as frequências formânticas dos diferentes falantes, ou seja, apesar de, no geral, crianças possuírem valores de F_0 e frequências formânticas mais elevados do que adultos, os ouvintes são capazes de distinguir os sons vocálicos utilizando a covariância existente entre a F_0 e as frequências formânticas.

Pesquisas em síntese de fala (LEHISTE; MELTZER, 1972; NEAREY; ASSMANN, 2007) apontam que a taxa de acerto dos ouvintes na identificação das vogais foi superior quando a relação entre a F_0 e as frequências formânticas foram mantidas dentro dos padrões de fala natural. Com base nessas ocorrências, a hipótese levantada é a de que a F_0 , assim como a frequência formântica, está diretamente relacionada com a qualidade acústica da vogal (SYRDAL; GOPAL, 1986). Segundo os autores, a F_0 é usada pelo ouvinte como um recurso para eliminar as diferenças entre os falantes e interpretar a relação da F_0 com os padrões formânticos.

2.1.2.5 Intensidade

Apesar de ser objeto de estudo deste trabalho, à semelhança da duração, não se pode deixar de mencionar outro parâmetro importante na descrição vocálica, qual seja a intensidade. Segundo Behlau e Pontes (1995), a intensidade vocal está diretamente relacionada com a pressão subglótica da coluna aérea que, por sua vez, depende de fatores como amplitude de vibração e tensão das pregas vocais.

Matheus (2004, p.6) afirma que “a intensidade do som decorre da amplitude da onda sonora (o valor da distância entre a pressão 0 e a pressão máxima da onda)”, ou seja, quanto maior a amplitude de vibração das partículas sonoras, maior a quantidade de energia transportada e, portanto, maior a percepção auditiva. A autora aponta ainda que “a proeminência do som a que chamamos ‘acento’ decorre dessa intensidade” (MATHEUS, 2004, p.6).

Uma sílaba acentuada, segundo Silva (2007, p.77), é caracterizada por ser pronunciada com maior pulso torácico, ou seja, com um “jato de ar” mais forte. Segundo a autora,

A vogal acentuada é auditivamente percebida como tendo duração mais longa e também como sendo pronunciada de maneira mais alta (no sentido de falar alto). Este aumento de volume permite-nos distinguir as vogais acentuadas das vogais não acentuadas e, portanto, percebê-las auditivamente de maneira mais distinta.

Alguns autores (MATEUS, 1990; MORAES BARBOSA, 1965; DELGADO-MARTINS, 2002;) afirmam que o acento está relacionado à duração, frequência fundamental, intensidade e energia. Segundo Moraes Barbosa (1965, p.12), as vogais acentuadas têm uma estrutura acústica mais precisa e destacam-se claramente dos sons que as rodeiam pela sua intensidade e, sobretudo, pela sua duração relativa, apresentando, portanto, uma diferença considerável em relação às vogais não acentuadas.

2.2 O SISTEMA VOCÁLICO DO PORTUGUÊS BRASILEIRO (PB)

Estudos sobre o sistema fonológico do Português Brasileiro apontam que os segmentos vocálicos estão condicionados pelo acento. Em posição tônica, o quadro vocálico do PB é constituído por sete vogais [a, e, i, o, u, ɐ, ʊ], nessa posição, a função distintiva dos segmentos vocálicos é plenamente estabelecida, como mostra os exemplos abaixo:

s/a/la

s/e/la

s/i/da

s/o/da

s/u/da

s/o/pa

s/u/ja

O quadro vocálico do Português Brasileiro são apresentadas por Câmara Jr. (1970) como sendo um sistema triangular, cuja base é constituída pela vogal /a/. Nesse sistema, a elevação gradual da língua, na parte anterior ou na parte posterior, origina a classificação articulatória de vogal aberta, vogais médias de 1º grau (abertas), vogais médias de 2º grau (fechadas) e vogais altas, como ilustrado no quadro 1.

altas	/u/			/i/	
médias	/ô/		/ê/	(2º grau)	
médias	/ò/		/è/	(1º grau)	
baixa	/a/				
	/posteriores/	/central/	/anteriores/		
Diante de consoante nasal na sílaba seguinte:					
altas	/u/		/i/		
médias	/o/		/e/		
baixa	/a/				
	[â]				

Quadro 1: Sistema vocálico do Português do Brasil (cf. CÂMARA Jr., p.43, 1970).

Em posição átona, o quadro vocálico do PB sofre redução, sendo interpretada por Câmara Jr. (1970) como neutralização, conceito da fonologia de Praga usado para referir-se à perda de um traço distintivo - evento fonológico no qual uma oposição distintiva, que opera em certos contextos da língua, deixa de operar em outras posições, neutralizando a distinção.

2.3 TEORIAS DE PERCEPÇÃO DE FALA

Há diversos pesquisadores que visam explicar como o ouvinte é capaz de perceber os sons vocálicos. Tais pesquisas baseiam-se tanto em modelos auditivos, os quais assumem que as diferentes formas do trato vocal são alternativas para alcançar diferentes alvos auditivos - destacando-se a teoria quântica. Há ainda os modelos que assumem a existência de gestos articulatórios como alternativa aos traços distintivos - destaca-se a teoria motora. A seguir segue-se uma breve descrição dessas duas principais teorias de percepção de fala.

2.3.1 Teoria motora

A teoria motora defende a proposição de que os ouvintes interpretam o sinal acústico em termos de padrões articulatórios, ou seja, para esse modelo, padrões articulatórios produzem padrões acústicos. A hipótese defendida é que os ouvintes, ao perceber determinada categoria fonética, reconstroem as intenções gestuais do falante, como por exemplo, arredondamento dos lábios, posição da língua, sonoridade, etc. Para essa teoria, o que os ouvintes percebem são as intenções articulatórias que controlam os movimentos do trato vocal e que por sua vez produzem uma versão perfeita das intenções do falante. A proposição é que

os ouvintes percebem gestos e não os movimentos individuais de articuladores individuais (FOWLER; GALANTUCCI, 2005).

Segundo Ignatius, Mattingly e Liberman (1985), apesar de os segmentos fônicos serem produzidos em sobreposição de período de tempo, eles são discretos, ou seja, a forma assumida pelo trato vocal difere de segmento para segmento. Liberman e Mattingly (1985) definem gestos como sendo dinâmicos tanto na forma como são produzidos quanto na forma como são percebidos. Os autores citam como exemplo a sílaba /bi/ que é produzida com uma constrição nos lábios, uma sobreposição temporal do corpo da língua e uma constrição no palato. De acordo os autores, ambas as constrições são discretas no sentido de serem distintas uma da outra, assim, de forma geral, as descrições dos dois gestos são livres do contexto, ou seja, os lábios sempre formam uma constrição para /b/, e uma constrição palatal sempre será feita para /i/, independente do contexto articulatorio. Apesar de, em um nível mais refinado de descrição, devido à coarticulação, os movimentos que atingem os gestos serem sensíveis ao contexto, às relações sinérgicas entre os articuladores contribuem para um gesto, assim, a mandíbula pode contribuir mais para o fechamento dos lábios no contexto de uma vogal de coarticulação próximos como /i/ do que no contexto de uma vogal aberta /a/.

Ignatius, Mattingly e Liberman (1985), afirmam que a grande contribuição da teoria motora está em mostrar que as línguas podem ter a propriedade de relacionar produção e percepção. Assim, considerando que para essa teoria, as formas fonológicas são ações em que os falantes se envolvem quando falam, estas devem ser preservadas através de uma troca comunicativa, portanto, elas são os átomos dos planos de fala dos falantes e das atividades do seu trato vocal. Além disso, por serem causas imediatas da estrutura do sinal acústico de fala por serem estruturas distintivas, os gestos estruturam distintamente o sinal, ou seja, o sinal pode fornecer informações sobre os gestos. Nesse sentido, se o ouvinte usar essas informações gestuais e acompanhar os gestos, então formas de linguagem fonológicas são preservadas ao longo de um intercâmbio comunicativo (FOWLER; GALANTUCCI, 2005).

Ignatius, Mattingly e Liberman (1995) propõem que a relação entre produção e percepção se deve a um módulo de fala, ou seja, a um componente especializado de circuitos neurais que evoluíram como uma especificação para a produção e percepção da fala. A existência de um módulo não foi inferida apenas com base nas restrições teóricas impostas pela exigência de paridade entre os mecanismos de produção e percepção. Sua existência, segundo os autores, foi sugerida por observações empíricas e pelo grande número de resultados de pesquisa que acompanhou o desenvolvimento de tecnologia de fala. O primeiro passo que levou a proposta do módulo de fala foi a constatação de que a fala não é um

alfabeto acústico. A essência da descoberta de Liberman (1957) era que um alfabeto acústico comparável a um alfabeto ortográfico não seria viável (cf. também HARRIS, 1953), mesmo se se criasse alfabetos cuidadosamente projetados e treinamentos extensivos. Esta conclusão indica, segundo Liberman (1957), a possibilidade de um mecanismo dedicado para lidar com a fala, ou seja, se a fala não pode ser substituída por um alfabeto acústico, então sua percepção exige diferentes tipos de mecanismos que lidam com percepção, diferente da máquina que manipula seqüências acústicas tais como o alfabeto Morse.

A segunda constatação de que produção e percepção estão diretamente relacionadas ocorreu quando uma compreensão mais completa do sinal de fala tornou possível devido ao advento do espectrograma. Uma das primeiras descobertas foi que, na fala real, informações acústicas sobre fonemas não são temporalmente discretas, ou seja, uma pequena porção de um dado de sinal acústico pode conter informações sobre vários fonemas e, inversamente, um fonema pode influenciar o sinal acústico, tornando-o mais longo. Estas descobertas aprofundaramas questões referentes à relação entre a instanciação física da fala e suas unidades linguísticas, permitindo concluir que o sinal da fala é contínuo, que os códigos de informações fonéticas são altamente paralelos, mas que a percepção fonética é discreta e sequencial.

Segundo Liberman (1957), o que parece ser um quebra-cabeça complicado para alguns pesquisadores pode ser uma solução ideal para a natureza da fala. De acordo com o autor, se por um lado a instanciação física da fala é um sinal seqüencial feito de unidades discretas, então, a fala poderia ser considerada como sendo uma portadora de comunicação altamente ineficiente. Se por outro lado, as unidades linguísticas, especificadas em paralelo no sinal fossem combinadas com a percepção, então os ouvintes não perceberiam as unidades particulares que são os átomos do produtivo sistema fonológico. Parece ser fundamental, portanto, segundo o autor, considerar a existência de uma transformação contínua-paralela, ou seja, um sistema perceptivo de fala capaz de transformar uma estrutura acústica em estrutura fonologia. Segundo Ignatius, Mattingly e Liberman (1985), essa peculiaridade em transformar, rapidamente e com precisão, uma estrutura acústica em uma estrutura fonologia discreta e seqüencial deve-se a um módulo especializado.

Ignatius, Mattingly e Liberman (1985) propõem identificar o módulo de percepção dentro da arquitetura do sistema auditivo. Para isto, distinguem dois diferentes módulos, o módulo aberto e o módulo fechado. Módulos abertos, também chamados de sistemas horizontais, são abertos no sentido de que podem adaptar-se e ajustar-se a novas situações ambientais. Os percepto são homomórficos, assim, no caso do sistema auditivo, por exemplo,

a altura é a percepção homomórfica da frequência e a sonoridade é percepção homomórfica da intensidade. Módulos fechados (também chamados de sistemas verticais) são fechados no sentido de que, sendo altamente especializados para um determinado tipo de estímulo, não podem adaptar-se a novas situações ambientais.

Pode-se concluir assim que a proposição da teoria motora é que os ouvintes percebem os gestos, e que a percepção é alcançada por meio de um módulo do sistema nervoso especializado à produção e à percepção da fala, e que a percepção recruta o sistema motor de fala.

2.3.2 Teoria quântica

As duas maiores questões que levaram Stevens (apud KENT e READ, 1992) a desenvolver a teoria quântica foram: a) por que alguns sons são muito comuns para algumas línguas e para outras não? b) por que o grande número de sons que as pessoas produzem faz parte de número restrito em suas mentes? Segundo Pickett (1996), a primeira questão diz respeito a um dos fenômenos mais fascinantes da linguística e o outro diz respeito à percepção categórica.

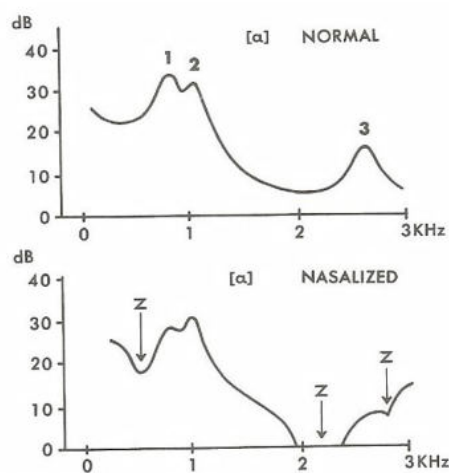
A teoria quântica da fala argumenta que algumas vezes o sinal de fala é especialmente rico em informações linguísticas relevantes, as quais são invariantes, quer devido à relação acústico-articulatória de fala ou devido à resposta quântica do sistema auditivo a um tipo particular de propriedade acústica. A hipótese desse modelo é que para toda unidade linguística há uma propriedade acústica invariante no sinal para a qual o sistema auditivo é especializado. Assim, esse modelo assume que propriedades invariantes do sinal de fala podem estar diretamente relacionadas às unidades linguísticas. Nesse sentido, difere da teoria motora em dois aspectos: primeiro, assume que as invariâncias são parte do sinal ou de um estágio precoce em resposta ao sistema auditivo para um dado sinal acústico de fala; segundo, assume que as unidades linguísticas são menores do que os fones ou fonemas, ou seja, assumem como menor unidade os traços distintivos da fonologia. Assim, a proposição de Stevens (1989) é que as propriedades invariantes do sinal de fala são correlatos acústicos e auditivos dos traços fonológicos distintivos.

Utilizando o sistema de traços articulatórios proposto por Chomsky e Halle (1968), Stevens (1989) argumenta que cada traço articulatório está diretamente relacionado a um correlato acústico correspondente, uma vez que a configuração articulatória tem uma consequência acústica predizível. Stevens (1989) utiliza os princípios acústicos e

aerodinâmicos para mostrar que a relação entre a configuração articulatória e o *output* acústico explicaria a estrutura linguística e a percepção fonética. O autor enfatiza que dentro da teoria acústica é possível observar que, para um dado parâmetro articulatório, haverá ocorrências em que uma pequena mudança articulatória provocará uma mudança significativa no *output* acústico. No entanto, em outros casos uma mudança articulatória igualmente pequena não provocará uma mudança auditivamente significativa no *output* acústico.

A hipótese de Stevens (1989) é que há uma relação acústico-auditiva, cuja base são os traços fonológicos distintivos. O autor não afirma que haja uma propriedade auditiva subjacente a cada traço distintivo, mas mantém um forte interesse pelas pesquisas neurológicas, argumentando que estas poderiam conduzir a descoberta de um sistema de respostas. Considerando tais pesquisas, Hawkins e Stevens (1985) sugerem que um sistema potencial deve estar na base da distinção entre vogais nasais e não nasais. Assim, como mostra a figura 6, quando a vogal é nasalizada, o espectro vizinho do primeiro formante é mais amplo e menos proeminente do que quando a mesma vogal não é nasalizada. Os autores sugerem que estes diferentes padrões de resposta estão por trás da distinção entre vogais nasais e não nasais. Assim, para toda unidade linguística há uma propriedade acústica invariante, e são essas propriedades invariantes que o ouvinte lança mão para discriminar os sons vocálicos.

Figura 7 - orma gerada pelo sinal acústico de fala para a vogal /a/ e /ã/

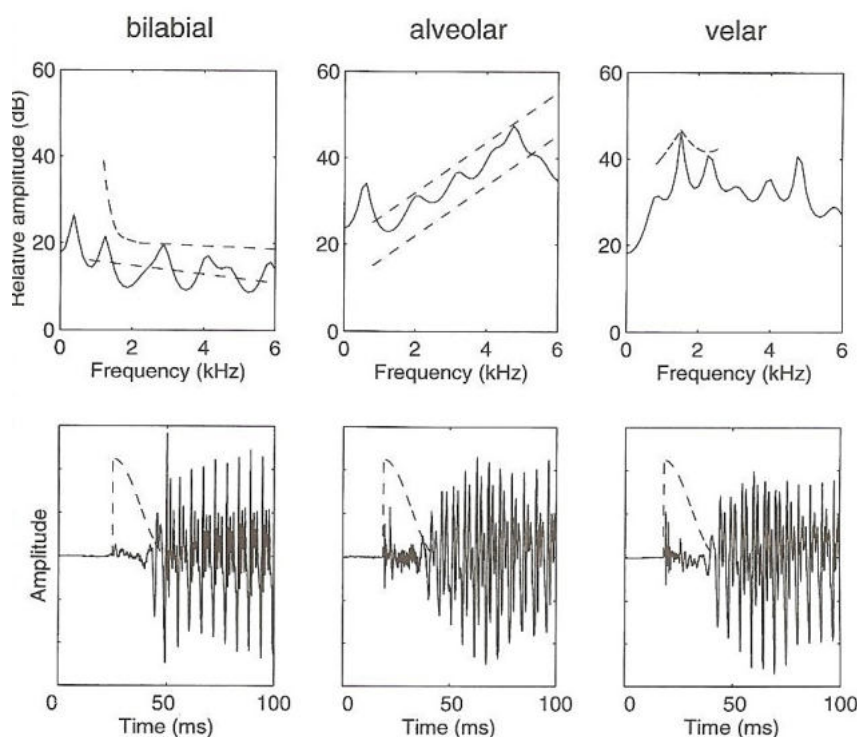


(cf. PICKETT, P.71,1998).

Stevens (1989) modelou a interpretação tradicional dos traços distintivos como descrição articulatório-estática ou estado-acústico. É possível observar essa invariação acústica do sinal de fala a partir do espectro de som. Na figura 7, por exemplo, há diferentes

espectros, uma vez que há diferentes movimentos do trato vocal. Assim, há uma bilabial, uma alveolar e outra velar.

Figura 8–Espectro gerado por Stevens e Blumstein (1979) para as oclusivas bilabial, alveolar e velar



(apud. PICKETT, 1996, p.219).

No topo esquerdo da figura 7, encontra-se um espectro de aproximadamente 26 ms considerando a explosão e *onset* de /b/, o espectro é relativamente baixo e inclinado (o que mostra que sua amplitude é maior em frequências baixas mais do que em altas), todos os picos desse espectro repousam em duas porções de linhas paralelas que aparecem abaixo do espectro. Stevens e Blumsteins (1970) propõem que qualquer explosão espectral, cujos picos repousam dentro dessas linhas paralelas seria identificada como uma bilabial. A alveolar consiste, assim como a bilabial, em linhas paralelas, no entanto, sua inclinação é para cima, ou seja, com frequência crescente, ao invés de para baixo como a bilabial. Além disso, os picos do espectro da bilabial, diferentemente da alveolar, ultrapassam toda a zona de queda delimitada pelas duas linhastracejadas. A velar, por sua vez, é a única que possui uma curvatura na frequência média com o pico dentro dessa região. Segundo Blumstein e Stevens (1970, apud PICKETT, 1996), essas mudanças espectrais refletem a forma do trato vocal no

momento de produção dos referidos sons. (ele diz que essa descrição não corresponde aos fatos ilustrados nos gráficos.)

Assim, para a teoria quântica, o *output* acústico fornece ao ouvinte as informações necessárias para discriminação dos sons da fala, uma vez que cada *output* terá um correlato acústico correspondente.

2.4 PERCEPÇÃO DE VOGAIS

Teorias a cerca de como os ouvintes são capazes de identificar vogais produzidas por diferentes falantes (homens, mulheres e crianças) podem ser divididas em: *especificação intrínseca* e *especificação extrínseca*. Ambos os modelos, utilizando tanto a fala natural quanto a sintética, têm obtido evidências empíricas para se compreender o processo de percepção das vogais (PICKETT, 1996).

2.4.1 Teoria de especificação intrínseca

Os modelos de especificação intrínseca baseiam-se na hipótese de que existem informações suficientes nos padrões acústicos das vogais que permitem ao ouvinte identificar as diferentes vogais inequivocamente. Esse modelo considera que o correlato acústico de uma dada vogal possui importantes aspectos espectrais representados em um modelo apropriado.

Pesquisas dentro do modelo de normalização intrínseca têm explorado a contribuição dos formantes mais altos (F_3) e da frequência fundamental (F_0) para especificar a identificação das vogais. Miller (1989) defende a hipótese de que as vogais são percebidas com base nos percentuais de formantes, assim, o autor propõe que as vogais produzidas por diferentes falantes, homens, mulheres e crianças podem ser representados em um espaço auditivo perceptual tridimensional definido por três intervalos (mensuração das unidades algorítmicas de frequências F_1-F_2 , F_2-F_1 , F_1-SR , onde a SR é a referencia sensorial, estimada na F_0 do falante). Syrdal e Gopal (1986,) propõem uma representação similar de vogais em um espaço bidimensional em que as coordenadas são F_1-F_0 e F_3-F_2 . Tanto Miller (1989) quanto Syrdal e Gopal (1986) afirmam que a sobreposição de vogais entre as categorias vocálicas foram reduzidas ou eliminadas quando tal transformação foi produzida (F_1-F_2 , F_2-F_1 , F_1-SR / F_1-F_0 e F_3-F_2). Tal resultado sustenta indiretamente a hipótese de que os ouvintes utilizam ambas as informações F_3 e F_0 para a discriminação de vogais produzidas por homens, mulheres e crianças.

Estudos de percepção que utilizam estímulos de vogais sintéticas em que a F_0 foi sistematicamente manipulada evidenciam que a identificação de vogais é diretamente influenciada pela F_0 . Dentre esses estudiosos, pode-se citar Meltzer e Lehist (1972) que sintetizam estímulos vocálicos simulando homens, mulheres e crianças baseando-se nos valores dos formantes dos três grupos. Cada grupo vocálico foi sintetizado nos três valores de F_0 , correspondendo às médias de valores masculino, feminino e infantil com base nos níveis de F_1 e F_0 . Se os ouvintes normalizam vogais produzidas por homens, mulheres e crianças com base nos níveis de F_1 e F_0 , então, esperar-se-ia que a percepção fosse mais exata se ambos F_1 e F_0 fossem aqueles correspondentes a um falante de mesma idade e sexo.

Segundo os autores, os resultados apontam que, como esperado, a síntese de vogais com formantes correspondentes a uma voz feminina foi mais exata quando a F_0 foi apropriada para uma mulher, no entanto, os autores apontam que vogais contendo padrões formânticos correspondentes a uma voz masculina foram identificadas igualmente bem quando sintetizadas com F_0 de voz masculina bem como com F_0 de voz feminina. Além disso, vogais com formantes apropriados para uma criança foram melhores identificadas quando sintetizadas com F_0 de voz feminina adulta do que quando sintetizadas com F_0 apropriada para uma criança. Embora a melhor percepção nem sempre tenha sido obtida para vogais sintéticas que contivessem F_1 - F_0 apropriados para falantes de uma idade e sexo particulares, estes resultados indicam que a F_0 afeta a identificação de vogais.

Ryalls e Liberman (1982) relatam resultados similares em um estudo em que as vogais produzidas por homens, mulheres e crianças foram sintetizadas com três diferentes valores de F_0 : típico do sexo (135 Hz, para homem e 185 Hz, para mulher), mais baixo do que o típico (100 Hz) e maior do que o típico (250 Hz). Segundo os autores, ambos os grupos de vogais, típico e mais baixo, foram melhores identificados do que vogais sintetizadas com valores superiores. Esse resultado, portanto, vai contra a hipótese de que o ouvinte normaliza as vogais com base nos níveis de F_0 e F_1 , uma vez que, vogais produzidas com F_0 maior do que o padrão tendem a ser bem menos percebidas, independente de as frequências de formantes serem as de vogais produzidas por homens, mulheres e crianças. Esses resultados apontam, portanto, que a variação da F_0 na síntese de vogais no estado estacionário influencia o modo como os falantes percebem o padrão formântico.

2.4.2 Teoria de especificação extrínseca

A hipótese defendida por essa teoria é a de que o ouvinte identifica as vogais com base nos “*frame references*” (arcabouço de referências), estabelecidos a partir de predições de padrões de fala, ou seja, de acordo com esse ponto de vista, as informações sobre idade e sexo do falante são avaliadas em padrões de fala corrente e utilizadas como base para interpretar os segmentos vocálicos.

De acordo com esse modelo, os ouvintes interpretam os padrões formânticos das vogais em relação ao contexto em que elas ocorrem. De acordo esse pressuposto, vogais isoladas são percebidas e relacionadas aos padrões de frequência dos formantes do inventário completo de qualquer falante (JOOS, 1972), a hipótese é que enquanto os formantes de uma mesma vogal produzida por diferentes falantes variam na frequência absoluta, a relação entre os padrões formânticos de diferentes vogais mantém-se relativamente constantes. Assim, por exemplo, qualquer falante seja homem mulher ou criança, a vogal /i/ tem frequências de F_2 maior e de F_1 mais baixa do que a vogal /I/. Além disso, para qualquer falante a vogal /i/ e a vogal /a/ possuem, respectivamente, as frequências de F_2 e de F_1 mais elevadas.

2.5 EXEMPLOS DE MODELOS DE PERCEPÇÃO DE VOGAIS

Dentre os principais modelos de percepção de vogais pode-se citar o *Simple Target Model*, *ElaboratedSimpleTargetModel* e *SpecificDynamicModel*. De um modo geral esses modelos ora privilegiam a hipótese da especificação intrínseca e ora a extrínseca. Abaixo segue uma breve descrição desses modelos.

2.5.1 Simple Target Model

O objetivo central desse modelo é integrar as informações articulatórias, acústicas e perceptuais dos segmentos vocálicos. A hipótese é que existe uma forma canônica vocálica livre do contexto em que está inserida. Assim, na *articulação*, a forma canônica é representada pela configuração estática assumida pelo trato vocal em fala corrente, em contextos de coarticulação, essa posição articulatória estática é o alvo desse modelo. Em um contexto acústico, as vogais são consideradas como espaço acústico estático multidimensional, cujas coordenadas são os dois primeiros formantes (F_1 e F_2), ou os três

formantes (F1, F2 e F3). Nesse modelo, a informação espectral dos dois primeiros formantes constitui a informação acústica suficiente para a *percepção* de vogais.

Este modelo, no entanto, apresenta limitações, uma vez que, as frequências formânticas diferem a depender das características do falante, ou seja, os padrões formânticos produzidos por uma mulher, homem ou criança diferem entre si, já que as frequências de ressonância são em parte determinadas pelo tamanho do trato vocal. Considerando tais limitações, Kent e Read (1992) propõem que esse modelo somente pode ser usado se algum tipo de normalização⁸ de fala for aplicado. De acordo com os autores, a normalização não é um problema trivial e representa uma possível solução especialmente devido ao rápido desenvolvimento das máquinas de reconhecimento de fala.

Outra limitação do modelo é a variação dinâmica e temporal dos padrões formânticos, ou seja, as frequências formânticas de uma vogal isolada não são as mesmas quando esta está inserida em uma sílaba CVC. Além disso, pesquisas apontam que um falante produz uma mesma vogal em diferentes contextos com diferentes valores formânticos. Uma possível solução apontada por Kent e Read (1992), é a proposta de que aquelas frequências formânticas que não alcançam o valor padrão são compensadas pelo ouvinte por meio de uma frequência perceptual, cujos valores são superiores ao padrão, compensando assim as discrepâncias acústicas.

Além da configuração formântica, as vogais diferem entre si em muitos outros parâmetros. Lehiste e Peterson (1961) apontam dois desses parâmetros: duração e trajetória formântica. Assim, as vogais possuem fonologicamente diferenças inerentes quanto à duração - vogais tensas possuem duração maior do que vogais frouxas. Quanto à trajetória formântica, os sons vocálicos tendem a diferir a sua trajetória a depender do contexto em que estão inseridos, por exemplo, vogais tensas tendem a ter estado estacionário mais longo quando inseridas em um contexto Vogal-Consoante (V-C); e vogais frouxas, quando inseridas nesse mesmo contexto V-C, possuem estado estacionário mais curto.

Considerando as limitações do *Simple Target Model*, alguns autores têm proposto outros modelos, a fim de dar conta das limitações desse e, assim, explicitar de que forma o ouvinte discrimina os diferentes sons vocálicos. Dentre esses, pode-se citar o *Elaborated Simple Target Model*, cuja solução encontrada para a questão da trajetória formântica foi transformar os valores acústicos dos formantes vocálicos em espaços perceptuais e psicofísicos. Parte-se do pressuposto de que o sistema auditivo realiza uma

⁸Entende-se por normalização de fala o processo que elimina ou corrige as diferenças entre os padrões formânticos produzidos pelos diferentes falantes.

reorganização dos dados acústicos. A maioria desses pesquisadores (MILLER, 1989; SYRDAL, GOLPAL, 1986; LEHISTE, MELTZER, 1973) leva em conta principalmente a normalização de fala.

2.5.2 Elaborated Simple Target Model

O *Elaborated Simple Target Model* visa principalmente dar conta da normalização de falantes. Seu principal objetivo é caracterizar as vogais produzidas por falantes de diferentes idades e sexo de tal forma que as variações entre a vogal produzida e a percebida sejam minimizadas.

O recente trabalho de Syrdal e Golpal (1986) é um bom exemplo desses esforços. Nessa proposta, as vogais são caracterizadas como tendo um espaço auditivo multidimensional definido pelas coordenadas F_1-F_0 , F_1-F_2 , F_2-F_3 . Os autores classificam as vogais dentro de duas categorias ao longo de cada dimensão, conforme ultrapassem ou não uma diferença crítica das três escalas *Bark*. Segundo Strange (1989), este modelo mostra que nos dados de Peterson e Barney (apud, STRANGE, 1989), os quais utilizam um corpus com vários falantes no qual as vogais ficam sobrepostas, houve uma redução da sobreposição dessas diferentes vogais produzidas por diferentes falantes. A dimensão corresponde ainda aos tradicionais traços fonéticos que descrevem as mudanças no trato vocal em termos de língua, lábios e posição da mandíbula. (STRANGE, 1989)

Outros pesquisadores, utilizando o *Simple Target Model*, idealizado por Gerstman (1968) e Skinner (1977) propõem diferentes algoritmos para a transformação dos dados acústicos em dados perceptuais. Dentre esses, pode-se citar Miller (1989) que propõe um modelo de percepção-auditiva em que a discriminação do sinal de fala é caracterizada como sendo um espaço auditivo tridimensional definido em termos de níveis formânticos e uma referencia sensorial baseada na frequência média do falante.

Ainda outros autores levantam algumas questões referentes à normalização das vogais, tais como: se um falante não é capaz de perceber as diferentes propriedades acústicas de seu sistema vocálico, então como os ouvintes sabem que vogal um determinado falante está produzindo? Alguns propõem responder tal pergunta (CLAES et al.; JOSS, 1948;), afirmando que os ouvintes criam protótipos idiossincráticos para cada falante, ou seja, afirmam que os ouvintes desenvolvem padrões acústicos com base na percepção de um pequeno número de elocuições de um determinado falante. De acordo com essa hipótese, o contexto acústico geral fornece a informação acústica essencial para que o ouvinte possa

construir um padrão vocálico referencial para um dado falante. Outra hipótese é a possibilidade de o ouvinte estimar o tamanho do trato vocal a partir da frequência média da F_3 .

Assim, de acordo com essa teoria, as vogais são interpretadas dentro de um espaço vocálico. A hipótese é que o ouvinte constrói um espaço vocálico referencial, ou seja, um “dispositivo” vocálico. Nessa proposição, os ouvintes adquirem modelos baseados em suas experiências de longo prazo, ouvindo a fala de várias pessoas e estabelecendo médias acústicas determinadas por homens, mulheres e crianças. Assim, ao ouvir uma vogal desconhecida, o ouvinte a marca a partir de um modelo vocálico apropriado com base no *pitch* e timbre, estabelecendo, assim, a média para homens, mulheres e crianças. O argumento dessa teoria é que os ouvintes podem identificar acuradamente tanto vogais isoladas, descontextualizadas, como aquelas produzidas por falantes de diferentes idades e sexo.

Além do *ElaboratedSimple Target Model*, dentre os modelos criados que visam dar conta das limitações do *Simple Target Model*, pode-se citar o *Modelo de Especificação Dinâmica* criado por Strange (1989). A autora acredita que nem o *Simple Target Model* nem *ElaboratedSimple Target Model* poderiam dar conta adequadamente da percepção de vogais. Assim, ela propõe outra abordagem definida como *DinamicSpecification* que busca integrar o estado estacionário, a informação temporal ou dinâmica desse estado e a natureza da transição formântica antes e depois do estado estacionário. Dentro dessa perspectiva, F_1 e F_2 sozinhos não são suficientes para representar adequadamente os sons vocálicos, sendo necessário incluir a informação espectral.

Segundo Strange (1989), os modelos *Simple Target Model* e *ElaboratedSimple Target Model* diferem em alguns aspectos, mas possuem algumas características em comum. Os dados acústicos para uma representação auditiva da vogal são tomados, em ambas as teorias, a partir de uma única secção transversal espectral da sílaba acústica. De acordo com Strange (1989) todas essas tentativas pressupõe que uma configuração espectral estática capta a essência do input acústico para a identificação da vogal, pelo menos no que se refere a vogais monotongas.

2.5.3 Dinamicspecification

Esse modelo enfatiza principalmente a percepção da co-articulação das vogais. Propõe investigar como o ouvinte é capaz de discriminar os segmentos vocálicos diante da variabilidade acústica associada aos efeitos contextuais, além da variabilidade de falantes.

Segundo Strange (1989), o interesse por pesquisas referentes à co-articulação deve-se a resultados inesperados de pesquisas realizadas com vogais inseridas em contexto silábico de consoantes oclusivas produzidas por diferentes falantes. Estes estudos apontaram, segundo a autora, que vogais co-articuladas foram percebidas com surpreendente precisão apesar da considerável ambiguidade acústica das frequências formânticas, ou seja, as frequências formânticas F_1 e F_2 de diferentes vogais co-articuladas foram mais similar do que quando estas estavam isoladas e, apesar de tal similaridade, o ouvinte foi capaz de discriminar tais vogais mais precisamente do que quando estas estavam isoladas - contexto em que a frequência formântica de F_1 e F_2 é mais distinta. Os resultados apontaram, segundo a autora, que a identificação de vogais inseridas em um contexto CVC foram melhor discriminadas do que quando estas foram apresentadas ao ouvinte isoladas.

A partir desses resultados um grande número de pesquisas foi realizado a fim de comprovar a importância da coarticulação para a percepção vocálica. Segundo Strange (1989), essas pesquisas apontam que vogais isoladas foram menos precisamente percebidas do que vogais em contexto de coarticulação. Strange (1989) cita pesquisas realizadas por Macchi (1980), Diehl et al. (2003), Assmann et al. (1984), Rakerdet et al. (1982), os quais apontam que vogais co-articuladas foram melhor percebidas do que vogais isoladas. Com base nesses dados, Strange (1989) propõe que há no sinal acústico, outras informações além da configuração formântica que os ouvintes utilizam para a identificação de vogais, assim, o autor volta-se a investigar a natureza da informação dinâmica disponível na configuração silábica.

Strange (1989) e seus seguidores desenvolveram uma técnica em que a informação do núcleo silábico, da duração e a informação dinâmica espectral (trajetória formântica) foram independentemente manipuladas. As sílabas CVC foram armazenadas em um computador em forma de onda e diferentes estímulos foram gerados apagando porções da sílaba e mudando a relação temporal. Segundo Strange (1989), esta técnica permitiu explorar três tipos de informações perceptuais: primeiro, as informações perceptuais provenientes do núcleo silábico, ou seja, informação da porção central de sílabas CVC em que os formantes encontram-se em estado estacionário ou onde as variações nos níveis dessas frequências são menores; segundo, a informação proveniente da duração das vogais em que a combinação com outros fatores, tais como, consoantes vozeadas, acentuadas e nível de fala determinam a duração silábica; e por fim a informação da transição formântica dentro e fora do núcleo silábico. Segundo Strange (1989), estas transições formânticas podem refletir diferenças

sistemáticas das características de oclusão e abertura dos gestos vocálicos, os quais determinam os parâmetros que moldam a trajetória formântica dentro e fora da sílaba.

Strange (1989) elaborou três diferentes condições de manipulação: (1) *silent-center*, em que a toda a porção central foi eliminada ficando apenas a porção inicial e final; (2) *neutral-duration*, em que a duração intrínseca foi removida igualando os intervalos de silêncio entre a transição das porções inicial e final; *centers* em que o núcleo central foi apresentado sem a porção inicial e final.

Os resultados apontam que a discriminação de vogais no estímulo *silent-center* foi notavelmente correta, em alguns casos os ouvintes as identificaram tão bem quanto identificaram a sílaba não modificada. Assim, quando a informação espectral dinâmica e a informação da duração relativa foram mantidas a percepção vocálica manteve-se altamente exata apesar do fato da porção do núcleo estar ausente. Quando a informação espectral da duração foi removida os ouvintes tiveram problemas para identificar a vogal do centro silábico. A percepção foi ainda menos precisa quando a informação nuclear foi mantida e a transição inicial e final foi removida. Segundo Strange (1989), estes resultados apontam que as porções do centro silábico não são suficientes nem necessárias para a percepção exata da co-articulação.

3 METODOLOGIA

3.1 INFORMANTES

Objetiva-se investigar um grupo de informantes composto por seis pessoas, três do sexo feminino e três do sexo masculino. Até o momento, foram investigadas três informantes do sexo feminino, com idade compreendida entre 20 e 30 anos, todas naturais de Vitória da Conquista-BA.

3.2 SELEÇÃO DE PALAVRAS E PREPARAÇÃO DAS FRASES

Foi elaborado um *corpus* constituído por palavras que continham em sua composição vogais médias abertas e fechadas arredondadas e não arredondadas em relação de oposição, como discriminado no quadro 2⁹.

Quadro 2 - Palavras inseridas nas sentenças dos testes de percepção

Vogal média alta arredondada	Vogal média baixa arredondada	Vogal média alta não arredondada	Vogal média baixa não arredondada
Côrte	Cóрте	Zêlo	Zélo
Chôro	Chóрто	Pêso	Péso
Acôrdo	Acórрто	Sêde	Séde
Fôrma	Fórрто	Vê-lo	Vélo
Sou	Sol	Pê	Pé

Em seguida, cada palavra foi inserida em uma determina sentença, a fim de criar um ambiente significativo e coerente para o ouvinte. Com vistas a investigar se tais sentenças eram de fato coerentes e significativas para a comunidade linguística investigada, elaboramos um teste de aceitabilidade, o qual foi aplicado com informantes que correspondiam ao perfil daqueles que seriam submetidos ao teste de percepção.

O teste de aceitabilidade foi constituído por trinta sentenças, dentre as quais tínhamos aquelas alvo da nossa pesquisa e cuja aceitabilidade ou não era objeto de fato de julgamento pelos nossos julgadores, bem tínhamos frases incoerentes que eram distratoras¹⁰. O julgador, ao ler o teste, deveria atribuir graus de aceitabilidade as frases, sendo: 0 para frases

⁹As palavras foram acentuadas, a fim de favorecer melhor visualização nos pares mínimos investigados.

¹⁰As “distratoras” foram incluídas visando analisar a eficiência do teste, se de fato o informante estava atento as informações contidas nos testes, constituíram 26% do teste.

incoerentes, 1 para frases pouco coerentes e 2 frases totalmente coerentes. O teste foi aplicado a um grupo de vinte pessoas, sendo que cada um respondeu ao teste individualmente.

As palavras cujas sentenças foram classificadas pelos julgadores como inadequadas foram inseridas na frase veículo “Digo _____ baixinho”. A partir desse procedimento metodológico, as frases usadas para manipulação da F_0 foram as que são apresentadas no quadro 3:

Quadro 3–Frases utilizadas nos testes de percepção

1.	Por você zêlo sempre independente do que acontecer
	Por você zêlo sempre independente do que acontecer.
2.	Maria disse: eu não quero ficar com essa fôrma de bolo
	Maria disse: eu não quero ficar com essa fóрма de bolo.
3.	Estou sempre atenta: pêsô, sempre o recomendado pelo médico.
	Estou sempre atenta: péso, sempre o recomendado pelo médico.
4.	João disse aos garotos: Côrte, tudo o que as meninas desejam.
	João disse aos garotos: Córte, tudo o que as meninas desejam.
5.	Acórdo feliz, com você em qualquer tempo e lugar.
	Acôrdo feliz, com você em qualquer tempo e lugar.
6.	Ela sorrindo perguntou isso é um pê de pôrco?
	Ela sorrindo perguntou isso é um pé de pôrco?
7.	Digo chôro baixinho.
	Digo chóro baixinho.
8.	Digo vêlo baixinho.
	Digo vélo baixinho.
9.	Digo séde baixinho.
	Digo sêde baixinho.
10.	sol a coisa mais fascinante do universo.
	sou a coisa mais fascinante do universo.

As frases que se encontram no quadro 3 foram lidas por dois informantes do sexo masculino natural de Vitória da Conquista - BA e gravadas por meio do software *Audacity*, em uma cabine acusticamente tratada no Laboratório de Fonética e Fonologia da Uesb (LAPEFF). Antes da gravação, apresentamos as sentenças aos sujeitos e explicamos que esses

deveriam lê-las em tempo natural de fala. Em seguida, analisamos as gravações, e dentre essas escolhemos aquela que apresentava melhor sinal para o objetivo proposto¹¹.

3.3 GRAVAÇÃO DAS FRASES E PREPARAÇÃO DO ESTÍMULO: A MANIPULAÇÃO DA F_0

A frequência fundamental das vogais em questão foi alterada na porcentagem de 25%, 50% e 75% do valor original, tanto para valores ascendentes quanto para valores descendentes. Elaboramos, portanto, diferentes sinais sonoros com a F_0 manipulada tanto para valor inferior ao original, quanto superior nas taxas discriminadas acima.

A fim de obter os valores da frequência fundamental pretendidos, foi necessário primeiramente manipular o sinal da sentença completa e em seguida a vogal em questão, ou seja, inicialmente consideramos o menor valor de F_0 na sentença, para posteriormente obtermos a média do referido valor vocálico e manipularmos em 25%, 50% e 75%, tanto para padrões ascendentes quanto descendentes.

Com a sentença já manipulada, o valor da frequência fundamental da vogal foi alterado na porcentagem necessária de forma a atingir os valores desejados. Tal procedimento foi necessário, uma vez que, quando toda a sentença foi manipulada considerando somente os padrões das vogais em questão, a manipulação do sinal ocorria somente em alguns pontos, já que dentro de uma sentença os valores da frequência fundamental de cada segmento variam de um para outro, assim, havia segmentos, cuja frequência, ao ser manipulada com o valor considerando para vogal, ficava abaixo do padrão auditivo e não era, portanto, manipulada. Assim, foi necessário, considerar primeiramente o menor valor da Frequência Fundamental dos segmentos da sentença, manipulá-lo nos valores pretendidos e, em seguida, obter a média percentual ainda necessária para que a vogal seguramente atingisse os valores esperados de 25%, 50% e 75%, tanto superior quanto inferior ao sinal original.

Para a manipulação da F_0 , utilizamos o *Praat*, programa desenvolvido por Paul Boersma e David Weenink, do *Institute of Phonetic Sciences*, Universidade de Amsterdã. O programa pode ser obtido gratuitamente pelo site <http://www.praat.org>. Primeiramente transformamos o sinal original em um tipo de arquivo cujo formato permite a manipulação dos dados (obtem-se tal efeito clicando no botão *To manipulation*), em seguida, clicando no botão *View & Edit*, obtemos o sinal a ser manipulado, conforme ilustrado na figura 8. Posteriormente, no botão *Pitch*, comando *shift pitch frequencies*, ampliamos e reduzimos os

¹¹Os informantes possuíam vozes com diferentes timbres, um bem mais grave do que o outro, optamos pela voz menos grave para facilitar a manipulação da F_0 .

valores conforme a porcentagem estipulada, como ilustrado na figura 9 (ampliação da F_0) e figura 10 (redução da F_0). Salvamos os arquivos em formato wave e inserimos os arquivos de som no *Transcriber*.

Figura 9 - Praat – manipulação da F_0 ¹²

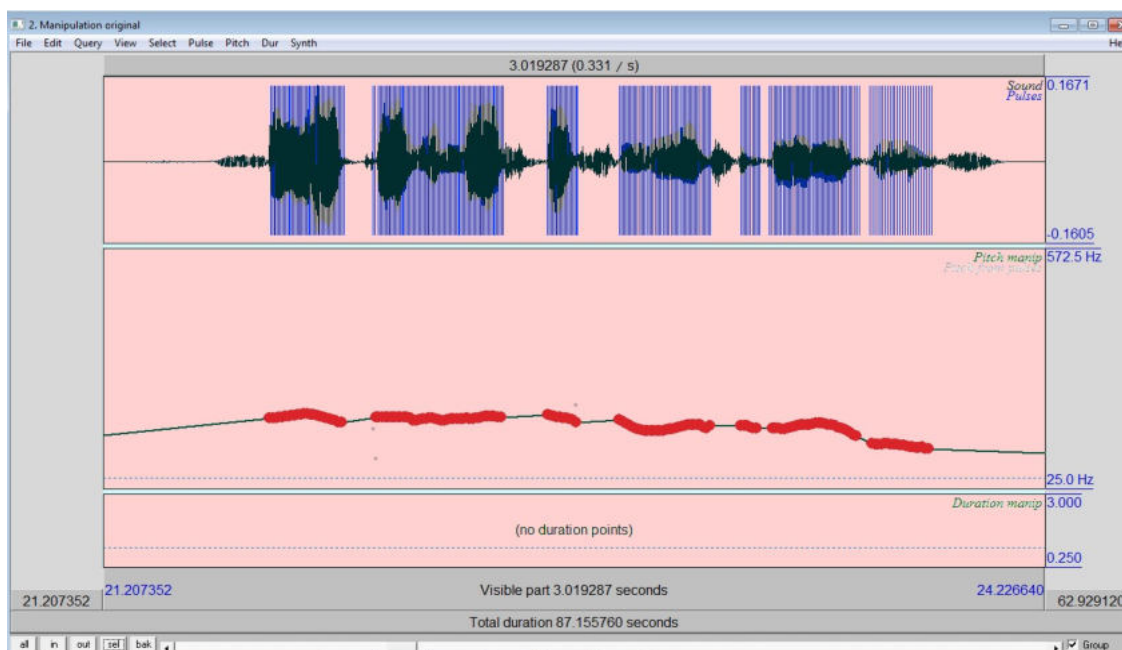
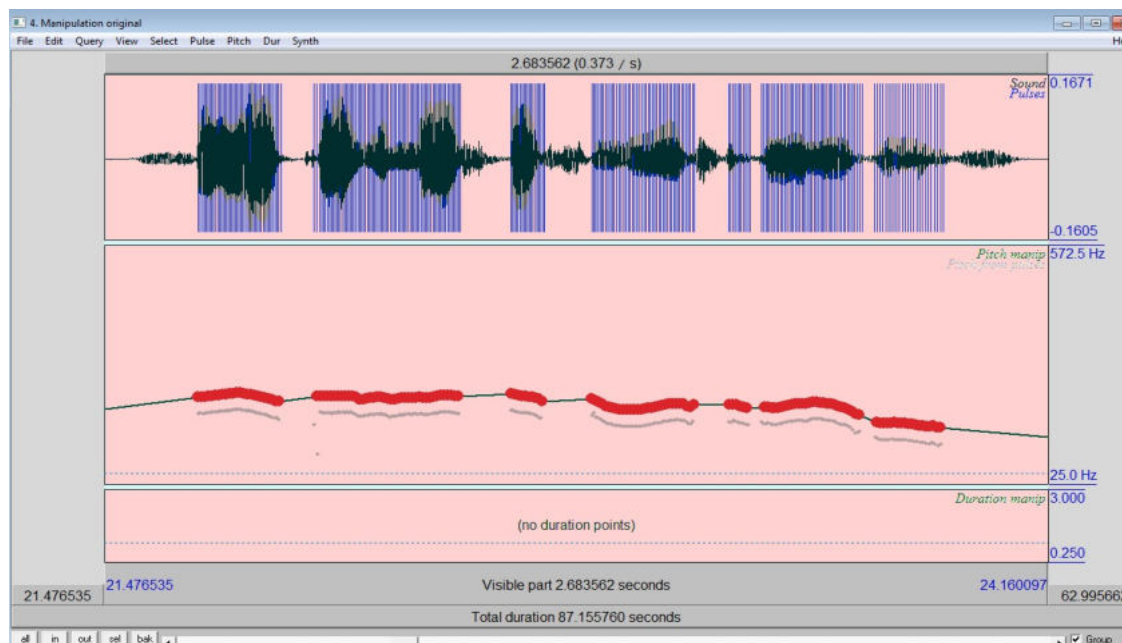


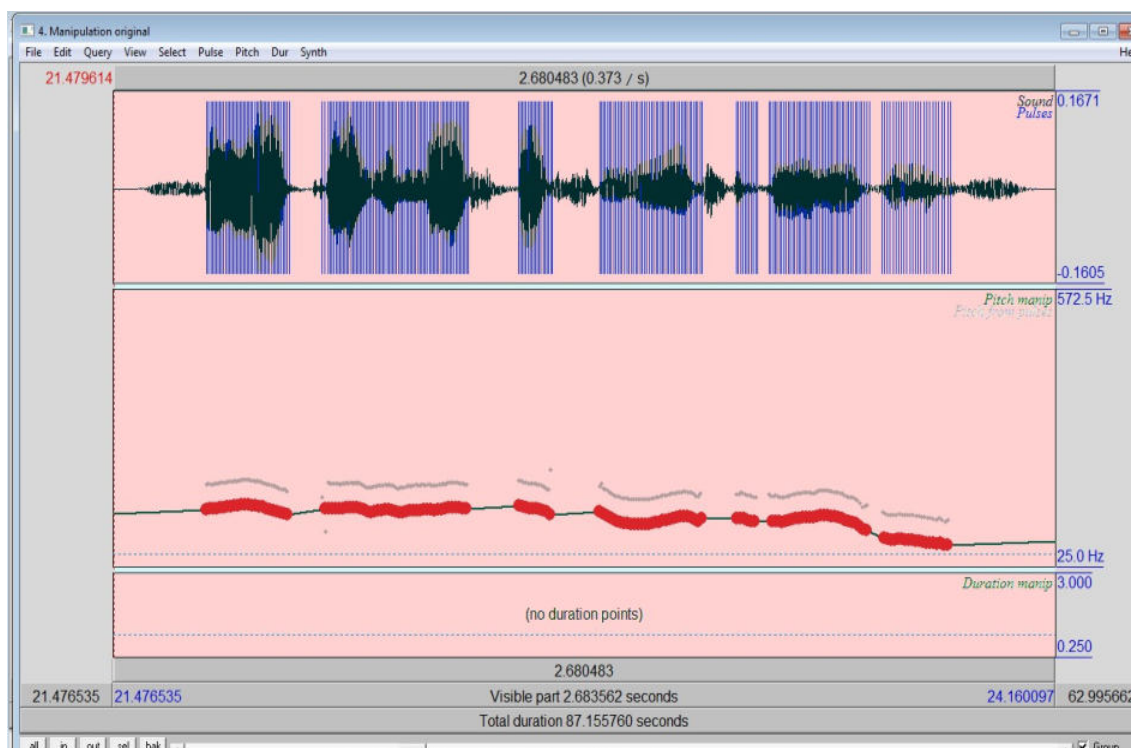
Figura 10 - Manipulação – ampliação da F_0 ¹³



¹² Figura referente à sentença: “sol, a coisa mais fascinante do universo”.

¹³ Figura referente à ampliação da F_0 em 25% do valor original para a sentença: “sol, a coisa mais fascinante do universo”.

Figura 11 - Manipulação - redução da Manipulação - redução da F_0 ¹⁴

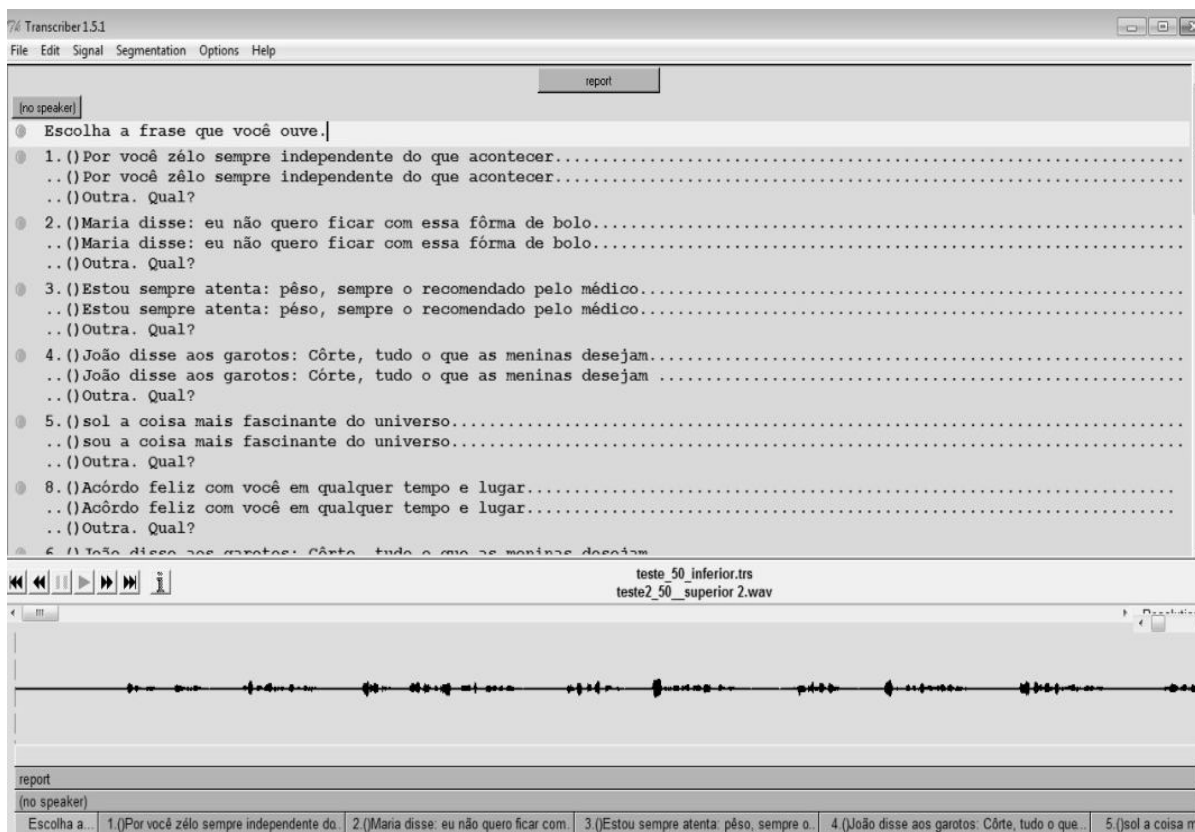


3.4 ELABORAÇÃO DOS TESTES DE PERCEPÇÃO

Para a elaboração dos testes de percepção, utilizamos o programa *Transcriber* que possui duas diferentes áreas de tarefas: um editor de textos, no qual foi possível inserir os testes escritos e outro editor de áudio, no qual foi possível inserir os sinais de fala, tanto o original quanto os manipulados (F_0 acrescida e diminuída em 25%, 50% e 75%), nas diferentes taxas aqui avaliadas. O programa permitiu que os informantes ouvissem os sinais e simultaneamente respondessem aos testes escritos a partir da janela do *Transcriber* como demonstrado na figura 11.

¹⁴ Figura referente à redução da F_0 em 25% do valor original para a sentença: “sol, a coisa mais fascinante do universo”.

Figura 12 - Transcriber – Aplicação dos testes de percepção



O *Transcriber* é um software gratuito distribuído sob licença da *General Public License* GPL, desenvolvido por KarimBoudahmane - DGA/CEP, Mathieu Manta - DGA/CEP, Fabien Antoine - DGA/CEP, Sylvain Galliano - DGA/CEP, Claude Barras - CNRS/LIMSI e coordenado por Edouard Geoffrois - DGA/CEP/GIP, Mark Liberman & Zhibiao Wu - LDC. Esse programa permite a transcrição e anotações do sinal de voz, oferece ferramentas que possibilitam ao usuário segmentar, etiquetar e transcrever o sinal de voz, bem como fazer alterações de tópicos e condições acústicas. O programa foi desenvolvido com a linguagem de script Tcl / Tk e extensões C. Pode ser adquirido pelo site <http://trans.sourceforge.net/en/presentation.php>.

3.5. AS TAREFAS DOS TESTES DE PERCEPÇÃO

No presente trabalho foram aplicados dois testes de percepção de percepção, um de discriminação e outro de identificação.

3.5.1 Teste de discriminação

No teste de discriminação, foi montada uma tarefa em que os informantes ouviram dois sinais de fala, um original e outro manipulado simultaneamente, em seguida deveriam responder ao teste informando se ambos os sinais, manipulado e original, correspondiam a uma mesma sentença ou não. Havia duas alternativas () *sim* e () *não*, o informante deveria responder ao teste com resposta positiva se considerassem que se tratavam de uma mesma sentenças e resposta negativa se considerassem que se tratavam de sentenças diferentes. Essa tarefa continha o mesmo número de vogais do teste de discriminação, ou seja, 450 vogais agrupadas do mesmo modo.

3.5.2 Teste de identificação

Para a aplicação desse teste, foi necessário que o sujeito ouvisse um dos sinais, original ou manipulado, e escolhesse dentre duas alternativas qual correspondia aquela que foi ouvida. No teste escrito, havia duas opções contendo os pares de palavras com as vogais em questão, ou seja, uma sentença contendo a vogal média alta e outro contendo a vogal média baixa, tais como, Maria disse: eu não quero ficar com essa *fôrma* de bolo / Maria disse: eu não quero ficar com essa *fôrma* de bolo, dentre outras, totalizando vinte e oito vogais, a saber, sete vogais médias altas arredondadas, sete vogais altas não arredondadas, sete vogais médias baixas arredondadas, sete vogais médias baixas não arredondadas. Assim, o informante ao ouvir determinado sinal, deveria marcar no teste escrito aquele que correspondia à sentença percebida.

Os informantes ouviram, portanto, dois sinais de fala, um original e outro manipulado nos valores de 25%, 50% e 75%, tanto para valores descendentes quanto ascendentes. O teste continha cinco repetições do sinal original e cinco repetições do sinal manipulado, totalizando 450 vogais. As vogais em questão foram divididas em diferentes grupos: *original* - 14 sinais referentes às vogais médias baixas, 14 sinais referentes às vogais médias altas; *manipulado* – sinais referentes às vogais médias altas e baixas manipuladas em 25%, 50% e 75% acima do valor original e abaixo desse.

3.6 PROCEDIMENTOS PARA APLICAÇÃO DOS TESTES DE PERCEPÇÃO

Os informantes tiveram acesso aos estímulos por meio de um fone conectado a um computador, o qual permitiu que o informante tivesse acesso aos sinais de fala sem interferência dos possíveis sons externos. Os sujeitos tiveram, por meio do software *Transcriber*, acesso simultaneamente ao sinal de fala e ao teste escrito, assim, responderam ao teste no monitor, marcando com um *x* a resposta escolhida. Antes da aplicação do teste, explicamos detalhadamente ao informante como seria a aplicação dos testes, como deveria marcar as respostas, além disso, este foi informado a cerca da possibilidade de pausa caso se sentisse cansado ou por qualquer outro motivo. Antes da aplicação dos testes, o informante foi submetido a um teste inicial para que pudesse compreender efetivamente todos os procedimentos necessários.

Considerando que os testes eram relativamente longos e que o tempo dispensado para a realização desses dependia da disponibilidade de tempo dos informantes investigados, as tarefas foram aplicadas em diferentes etapas, com vistas a favorecer a aplicação dos testes em tempo hábil e que não ficasse cansativo para os informantes. Organizamos os testes de forma que os informantes tiveram acesso aos três sinais de forma aleatória: sinal com a F_0 original, F_0 com manipulação ascendente e descendente. Primeiramente realizamos o teste de discriminação e em seguida o de identificação.

As tarefas foram aplicadas em diferentes etapas. Primeiramente realizamos o teste de discriminação e em seguida o de identificação. Organizamos os testes de forma que os informantes tiveram acesso aos três sinais de forma aleatória: sinal com a F_0 original, F_0 com manipulação ascendente e descendente. No teste de identificação, o informante tinha acesso a um grupo de vinte e oito sentenças, organizadas em pares mínimos, como exemplo, “Digo séde baixinho / Digo séde baixinho”, a cada teste o informante fazia uma pausa de cinco minutos e depois retornava. A cada três testes, o informante fazia um intervalo de vinte minutos, depois voltava a responder as tarefas, considerando o mesmo tempo de aplicação anterior. Devido a uma questão de tempo dos informantes para a realização das tarefas, em um dia foi possível aplicar seis testes, totalizando 168 vogais. Ao completar tal quantidade, foi necessário aguardar período de vinte e quatro horas para a realização dos testes seguintes. O teste de identificação foi aplicado em três dias.

Depois de aplicado o teste de discriminação, aguardamos o período de três dias e aplicamos o teste de identificação. Assim como o teste de identificação, o informante teve acesso ao sinal de fala dentro de um grupo de vinte e oito sentenças, sendo que a cada teste o

informante fazia uma pausa de cinco minutos e depois voltava a responder os testes. A cada três testes, o informante fez uma pausa de vinte minutos, depois voltou a responder as tarefas dentro do mesmo período de tempo anterior. Devido também a uma questão de tempo dos informantes, para a realização das tarefas, foi possível aplicar seis testes em um dia, totalizando 168 vogais. Ao completar tal quantidade, foi necessário aguardar período de vinte e quatro horas para a realização dos testes seguintes. O teste de identificação foi aplicado, assim como o de discriminação, em três dias. Para ambos os testes, os informantes tiveram plena liberdade de pedir pausa antes do período de tempo proposto.

4 RESULTADOSE DISCUSSÃO

A seguir seguem os resultados obtidos os quais foram agrupados da seguinte forma: 1) teste de discriminação; 2) teste de identificação.

4.1 TESTE 1: DISCRIMINAÇÃO

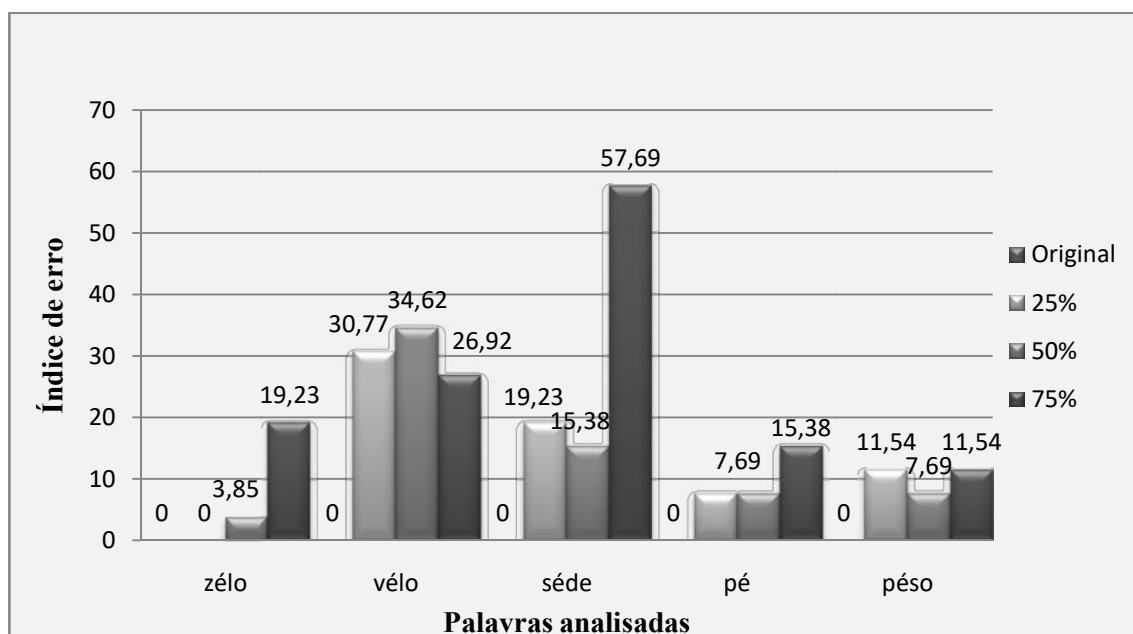
O teste de discriminação foi aplicado com vistas a avaliar se o ouvinte é capaz de discriminar mudanças na fundamental em relação à sua produção original, de sorte a distinguir entre dois *inputs* sonoros qual se referia a F_0 original e qual a manipulada. Com esse teste foi possível avaliar se o ouvinte é ou não sensível à mudanças na F_0 .

Abaixo seguem os gráficos correspondentes aos resultados obtidos nos testes de discriminação aplicados nesta pesquisa.

4.1.1 Análise dos testes de discriminação com valores de F_0 ampliados a 25%, 50% e 75%.

A seguir, no gráfico 1, estão dispostos os resultados dos testes de discriminação referentes às vogais médias para valores ampliados da F_0 .

Gráfico 1 – Índice de erro de discriminação da vogal média baixa não arredondada para ampliação da F_0



A taxa de perda de recuperação da qualidade vocálica correspondente ao sinal manipulado de F_0 varia razoavelmente de palavra para palavra. Os dados apontam que na palavra *zélo*, a ampliação da F_0 em 25% não gerou taxa de erro de discriminação. Taxas maiores de manipulação, contudo, alteraram essa tendência e comprometeram a inteligibilidade do timbre vocálico. Quando, por exemplo, a manipulação foi de 50%, o ouvinte não recuperou o sinal em 3,85% dos casos, e a ampliação em 75% gerou taxa de erro de discriminação de 19,23%.

Na palavra *vélo*, os índices de 25%, 75% e 50% de ampliação da F_0 acarretaram prejuízos maiores na identificação da vogal. Assim, para valores de 25%, a ampliação da F_0 acarretou 30,77% de erro. A manipulação da F_0 em 50% gerou taxa de erro de 34,62%. A ampliação para 75% acarretou erro de inteligibilidade da vogal em questão em 26,92%.

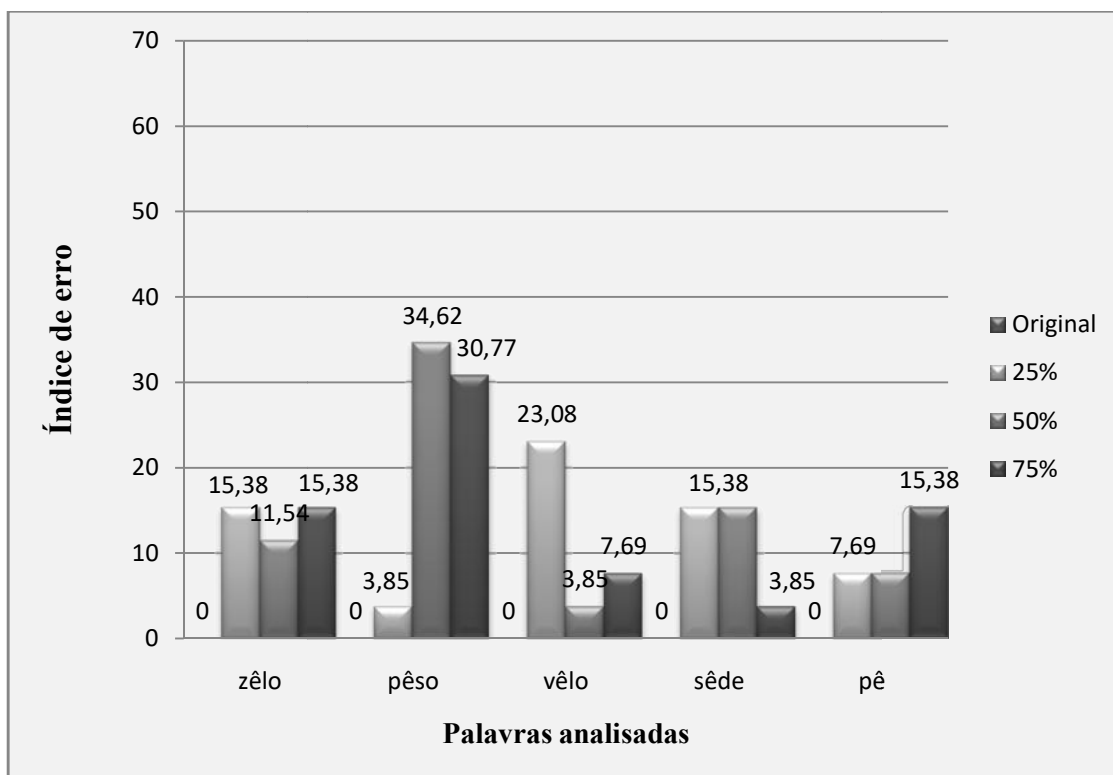
A ampliação da F_0 para a palavra *pé* gerou índice de interferência de 7,69% para os índices de manipulação 25% e 50%, já a ampliação da F_0 em 75% gerou taxa de erro de discriminação de 15,38%. A ampliação da F_0 , portanto, nesse contexto, afetou a inteligibilidade da vogal /E/ para os três índices de ampliação investigados.

Na palavra *péso*, os índices percentuais de 25 e 75 geraram 11,54% de erro de percepção do sinal. Para os valores de 50%, a ampliação da F_0 suscitou uma taxa de erro de 7,9%, em relação ao sinal original. Com a ascensão da manipulação para os valores de 50%, a taxa de erro ficou reduzida a 7,69%, e voltou a ascender para 11,54% na taxa de manipulação de 75%.

A identificação da vogal /E/ na palavra *séde* foi afetada em 19,23% para a taxa de manipulação de 25%. Para a taxa de 50%, o índice de não discriminação caiu para 15,38% e aumentou para 57,69% com a ampliação da F_0 em 75%. Nesse caso, curiosamente a taxa de 25% afetou a percepção da vogal /E/ em maior escala do que a taxa de manipulação de 50%.

Os dados apontam, portanto, que a ampliação da F_0 afetou a percepção das vogais médias baixas não arredondadas, uma vez que em todos os casos analisados houve taxa de erro de discriminação da vogal em questão em menor ou maior escala. Curiosamente, as taxas de erro não obedeceram ao padrão previsível, ou seja, esperava-se que quanto maior as taxas de ampliação da F_0 maiores os índices de erro de discriminação, mas, ao contrário do que se esperava, em muitos casos a ampliação em 25% gerou taxa de erro superior às taxas de 75% e 50%.

Resultado muito semelhante foi também encontrado para a média alta não arredondada, como podemos observar no gráfico 2.

Gráfico 2 – Índice de erro de discriminação da vogal média alta para valores ampliados da F_0 

É possível observar a partir do gráfico a cima que a discriminação da palavra *zêlo* foi afetada para os três índices de manipulações analisadas. As taxas de ampliação de 25% e 75% geraram mesmo índice de erro de discriminação, 15,38%. A ampliação em 50% gerou índice de erro de 11,54%. A ampliação da F_0 , portanto, nesses contextos interferiu a inteligibilidade da vogal média alta não arredondada.

A inteligibilidade da palavra *pêso* também sofreu perda com a F_0 manipula nos três valores. A taxa de 25% gerou índice de erro de 3,85%, índice que aumenta para 34,62%, quando a ampliação da F_0 é 50%, e que cai para 30,77% quando é 75%.

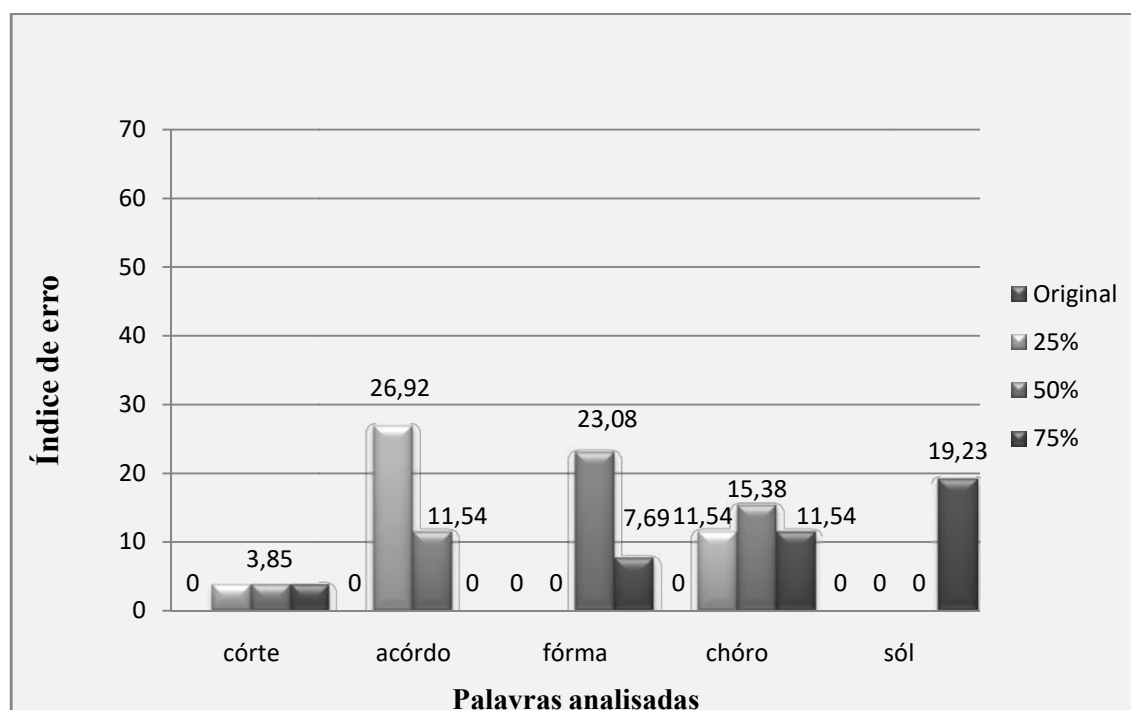
Na palavra *quêlo*, a ampliação da F_0 em 25% ocasionou alteração na discriminação da vogal em relação ao sinal original, em 23,08%. Quando essa foi manipulada em 50%, o índice de erro de discriminação da vogal em questão caiu para 3,85%, e quando manipulada em 75%, a taxa de erro de discriminação foi levemente elevada para 7,69. Os dados apontam que, curiosamente, os ouvintes recuperaram com melhor acuidade a vogal média em questão para valores ampliados a 75% e 50% do que para a taxa de 25%.

A discriminação da palavra *quêlo* não sofreu interferência quando manipulada para os valores de 25% e 50%, observa-se uma pequena interferência quando o sinal foi manipulado em 75%.

Na palavra *sêde*, a ampliação da fundamental em 25% e 50% acarretou prejuízo de 7,69% e de 13,3% em percentual de 75%. Nesse contexto, portanto, a ampliação da F_0 interferiu a discriminação da vogal média baixa não arredondada.

Passemos agora, a partir dos gráficos 3 e 4, a avaliar os resultados obtidos para a ampliação da F_0 para as vogais médias arredondadas.

Gráfico 3—Índice de erro de discriminação da vogal média baixa arredondada para ampliação da F_0



Os dados apontam que o índice de alteração da inteligibilidade da palavra *córte* foi o mesmo para os três valores manipulados, 3,85.

A ampliação da F_0 na taxa de manipulação de 25% na palavra *acórdo* suscitou taxa de erro de 26,92%. Para 50%, os ouvintes não conseguiram recuperar a vogal média baixa arredondada em 11,54% dos casos. A F_0 , portanto, gerou interferência para a discriminação da vogal média em questão quando essa foi ampliada em 25% e 50% do valor original. A ampliação em 75%, curiosamente não comprometeu a inteligibilidade da vogal. Nesse contexto de fala, a taxa inferior de manipulação 25%, que previsivelmente geraria menor taxa de erro de discriminação em relação às taxas de 50% e 75%, não ocorreu, ao contrário, taxas superiores de ampliação superiores geraram taxas de erro inferiores.

A discriminação da palavra *fórmula* sofreu interferência de 23,08% quando os valores de F_0 foram ampliados para a taxa de 50%. Na ampliação de 75%, a taxa de erro de

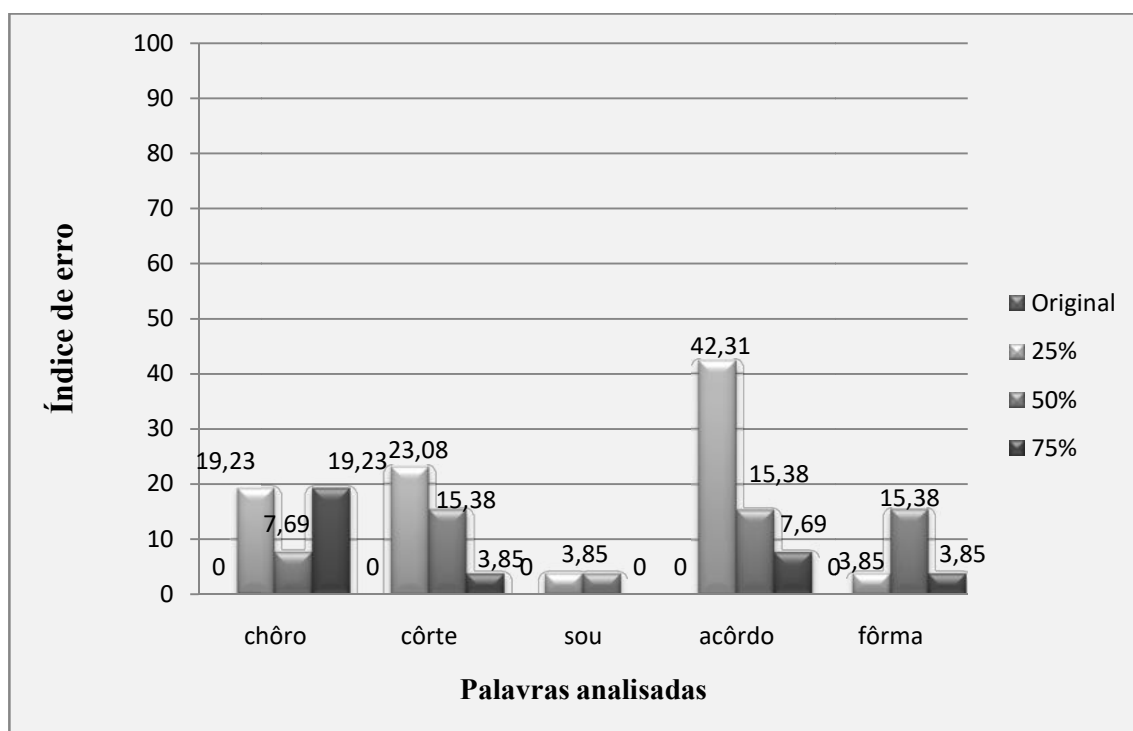
discriminação caiu para 7,69%; já o índice de 25% não afetou a discriminação da vogal em questão, diferentemente dos casos anteriores.

Na palavra *chôro*, a ampliação da F_0 gerou prejuízos para a inteligibilidade da palavra nos três casos investigados. A 25%, foi computado erro de discriminação de 11,54% e a 50% 15,38%. Quando a manipulação foi de para 75% a mais do valor do sinal original, a taxa de erro caiu para 11,54%, a mesma taxa de erro foi encontrada para a ampliação em 25%. Nessa palavra a ampliação da F_0 acarretou erros na discriminação da vogal média baixa arredondada.

A ampliação da F_0 em 25% e 50% na palavra *sól*, não trouxe prejuízos para a inteligibilidade dessa palavra, ou seja, o ouvinte conseguiu recuperar o sinal em 100% dos casos; no entanto, quando o índice de ampliação foi de 75%, o ouvinte não conseguiu recuperar o sinal original em 19,23% das ocorrências.

Seguem os resultados obtidos referentes à vogal média alta arredondada.

Gráfico 4 – Índice de erro de discriminação da vogal média alta arredondada para ampliação da F_0



A ampliação da F_0 em 25% na palavra *chôro* interferiu a inteligibilidade da palavra em 19,23% dos casos, já a ampliação em 50%, curiosamente, gerou taxa de erro de discriminação de 7,69%, inferior ao esperado comparando-se a 25%. Quando a manipulação foi ampliada

para 75% do valor do sinal original, a taxa de erro sobe para 19,23%, mesmo valor obtido para 25%.

A inteligibilidade da palavra *côrte* foi afetada em proporção decrescente: para taxaíndice de 25% de ampliação, o ouvinte não conseguiu recuperar o sinal original em 23,08% dos casos, para 50%, o índice de não discriminação caiu para 15,38% e para 3,85% quando a taxa de ampliação foi de 75% do sinal original.

Na palavra *acôrdo*, a ampliação da F_0 para os três índices de manipulação ocasionou prejuízos para a inteligibilidade do sinal de 42,31% para a fundamental ampliada. Essa taxa de não discriminação diminuiu para 15,38% com a ampliação da F_0 em 50% e para 7,69%, quando a ampliação foi de 75%.

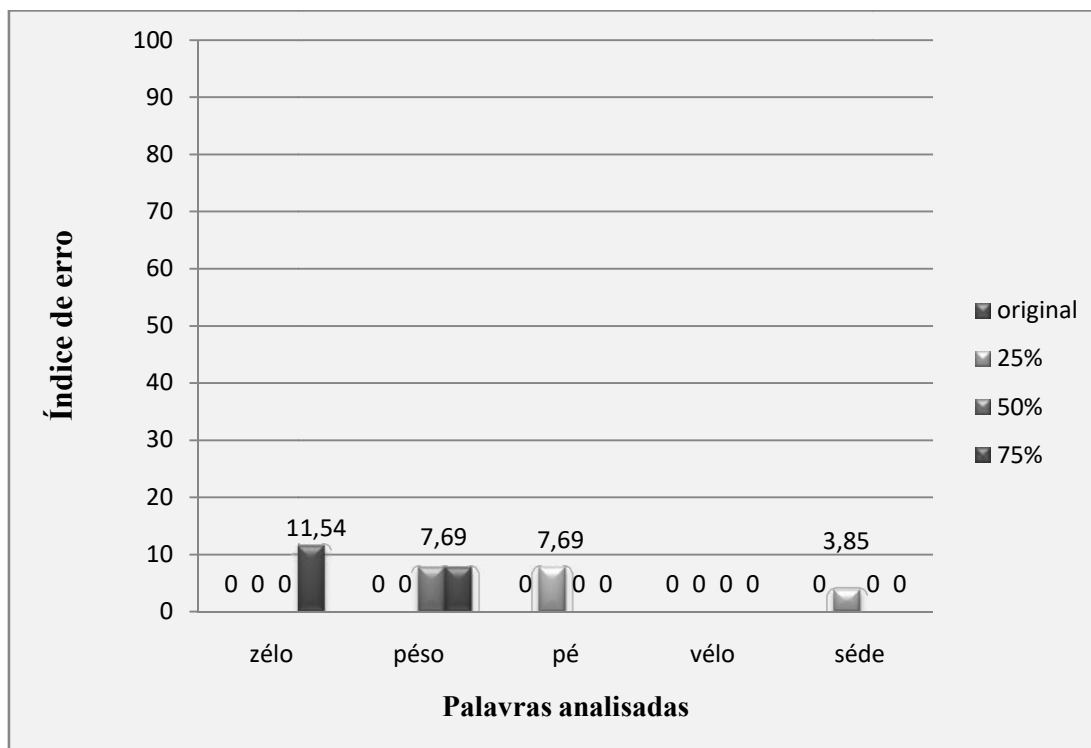
A ampliação da F_0 gerou prejuízos para a discriminação da palavra *fôrma* nos três valores investigados. Diferentemente do previsível, o maior índice de erro de discriminação foi para a ampliação da F_0 em 50%, o menor índice de erro foi para 25% e índice de erro intermediário para 75%. Desse modo, para a taxa de 25%, o índice percentual de erro foi de 3,85%; de 15,38% para 50% e de 3,85% para 75%

A 25% e a 50% as ampliações da frequência fundamental acarretaram prejuízos de perceptibilidade da palavra *sôu*. A ampliação da F_0 em 75% não gerou índices de erro, ou seja, o ouvinte conseguiu recuperar o sinal apesar do índice de ampliação da F_0 .

4.1.2 Análise dos testes de discriminação com valores de F_0 reduzidos a 25%, 50% e 75%.

Abaixo segue uma breve discussão referente aos resultados obtidos sobre a interferência da redução da F_0 na inteligibilidade da vogal média.

Gráfico 5 –Índice de erro de discriminação da vogal média baixa não arredondada para valores reduzidos da F_0



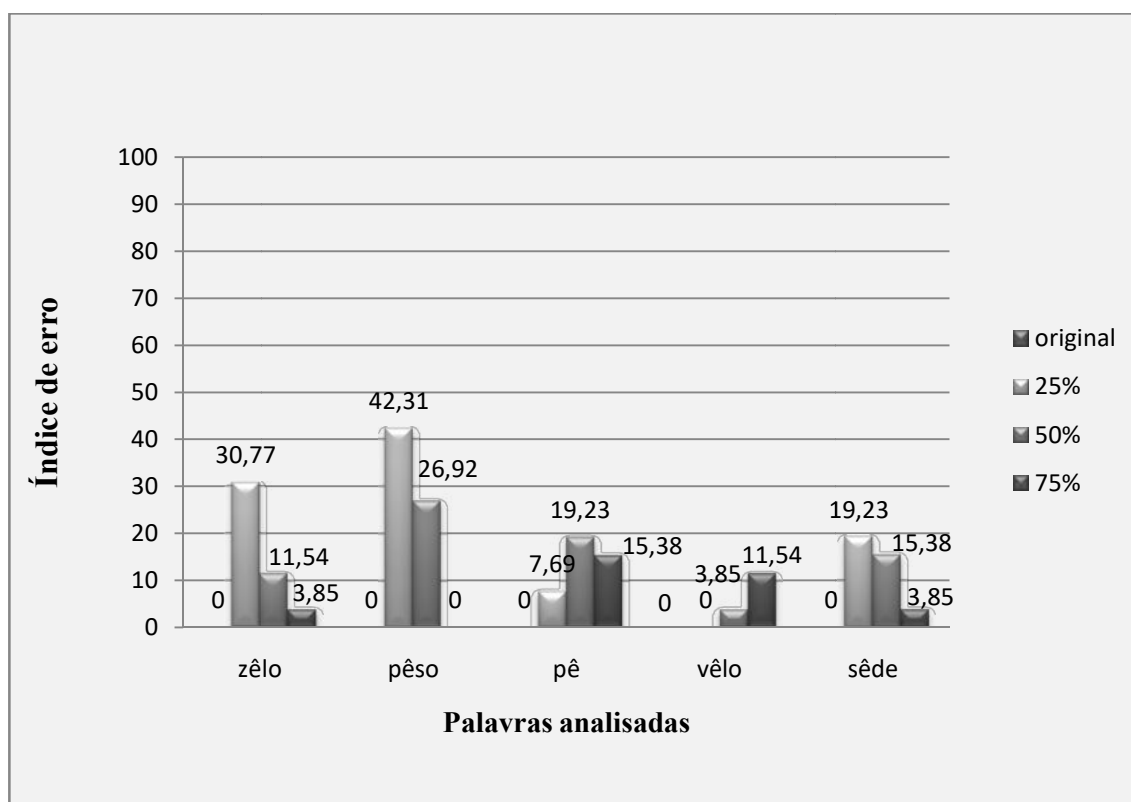
A partir da análise do gráfico 5, é possível observar que, na palavra *vélo* a redução da F_0 para os três valores investigados não acarretou taxa de erro de discriminação, ou seja, nesse contexto de fala, a redução da F_0 não interferiu na inteligibilidade da vogal em questão.

No caso das palavras *zélo*, as reduções da F_0 em 25% e em 50% não comprometeram a discriminação da vogal em questão, já a redução em 75% do valor original suscitou erros de discriminação de 11,54%.

Para alguns contextos de fala e para as diferentes taxas de manipulação investigadas, a redução da F_0 trouxe, em diferentes escalas e contextos, prejuízos para a discriminação da vogal /E/.

Passemos a analisar o gráfico 6 com os resultados referentes à vogal média baixa.

Gráfico 6 – Índice de erro de discriminação da vogal média alta não arredondada para valores reduzidos da F_0



A discriminação da palavra *zêlo* foi afetada para as três taxas de redução aqui investigadas. A redução da F_0 em 25% gerou erro de discriminação de 30,77%. Esse valor caiu para 11,54% quando a redução foi 50% e para 3,85%, quando de 75%.. Desse modo, os resultados apontam que, nesse contexto, a redução da F_0 interferiu a discriminação da vogal média alta arredondada. Como ocorre para os valores ampliados, os índices de erro de discriminação com a fundamental reduzida não obedeceram ao previsível, ao contrário, com a ampliação da F_0 , para os valores de 50 % e 75% os índices de erro de discriminação caíram consideravelmente, ou seja, a inteligibilidade da vogal em questão foi mais acurada para as taxas de ampliação superiores a 25%.

Na palavra *pêso*, assim como na palavra *zêlo*, a redução da frequência fundamental a 25% gerou índice de erro de identificação superior aos valores obtidos para 50% e 75%. Em 25% de redução, o índice de erro de inteligibilidade da palavra em questão foi de 40%; 50% de redução gerou índice de erro de discriminação de 26,92%; e a 75%, a taxa caiu para 0%, ou seja, o falante conseguiu recuperar em 100% o sinal original, ao contrário do que ocorreu com a manipulação da taxa de manipulação inferior, 25%.

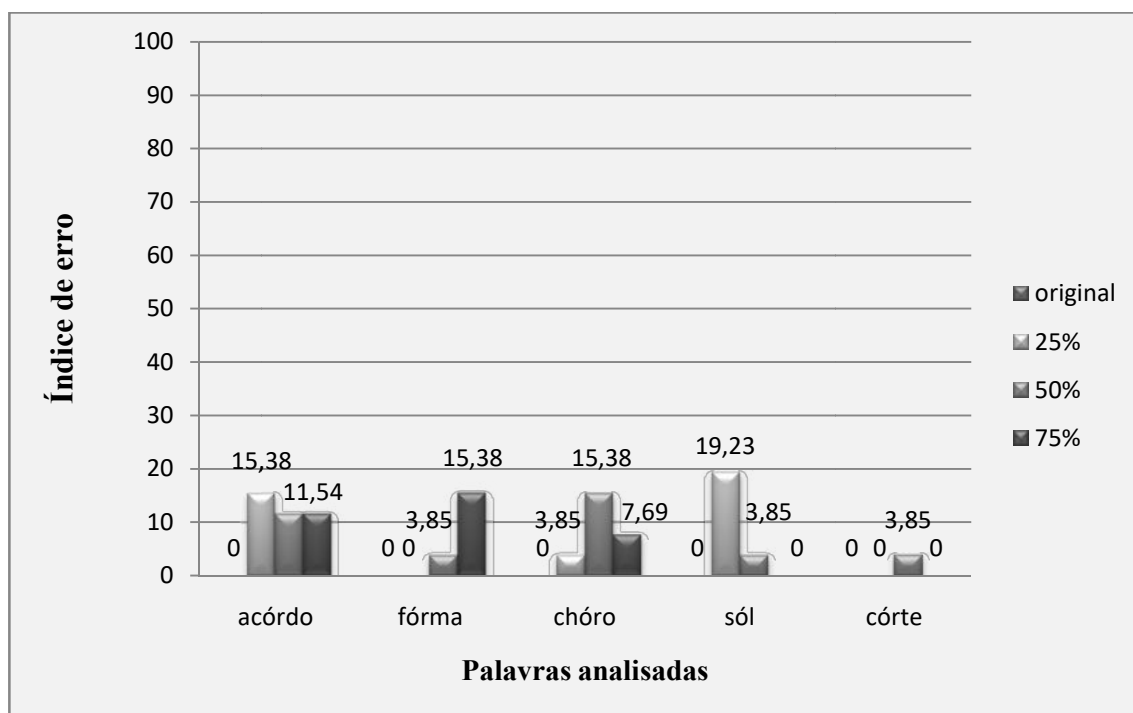
Na palavra *pê*, a redução da F_0 em 25% gerou índice de erro de discriminação 7,69%, o índice sobe para 19,23% com a redução em 50%, e voltou a cair para 15,28% com a redução em 75% do valor original.

Na palavra *vêlo*, a taxa de redução em 25% não afetou a inteligibilidade do sinal, ou seja, os ouvintes recuperaram em 100% dos casos investigados o sinal original, no entanto, a redução em 50% e 75% afetou a inteligibilidade do sinal em 3,85 e 11,54% respectivamente.

A redução da F_0 em *sêde* gerou, assim como para outros casos investigados, padrão gráfico não previsível, ou seja, a redução da F_0 não suscitou erro de discriminação crescente conforme a redução foi ampliada em 25%, 50% e 75%, mas, ao contrário, quanto maior a redução entre os valores analisados, menor a interferência para a percepção do sinal de fala.

Segue abaixo a descrição dos resultados obtidos para a redução da F_0 em palavras que contém em sua composição vogais médias arredondadas.

Gráfico 7 – Índice de erro de discriminação da vogal média baixa arredondada para valores reduzidos da F_0



É possível observar a partir do gráfico 7 que a redução da F_0 gerou, para os diferentes contextos de fala e taxas de manipulação, interferências para a inteligibilidade do timbre vocálico. Assim, na palavra *acórdo*, a redução da F_0 em 25% gerou interferência na discriminação do timbre vocálico em 15,38%; já os índices de 50% e 75% de redução levaram

a índices de erros similares de 11,54%, menor índice do que a taxa de redução de 25%, resultado também não previsível.

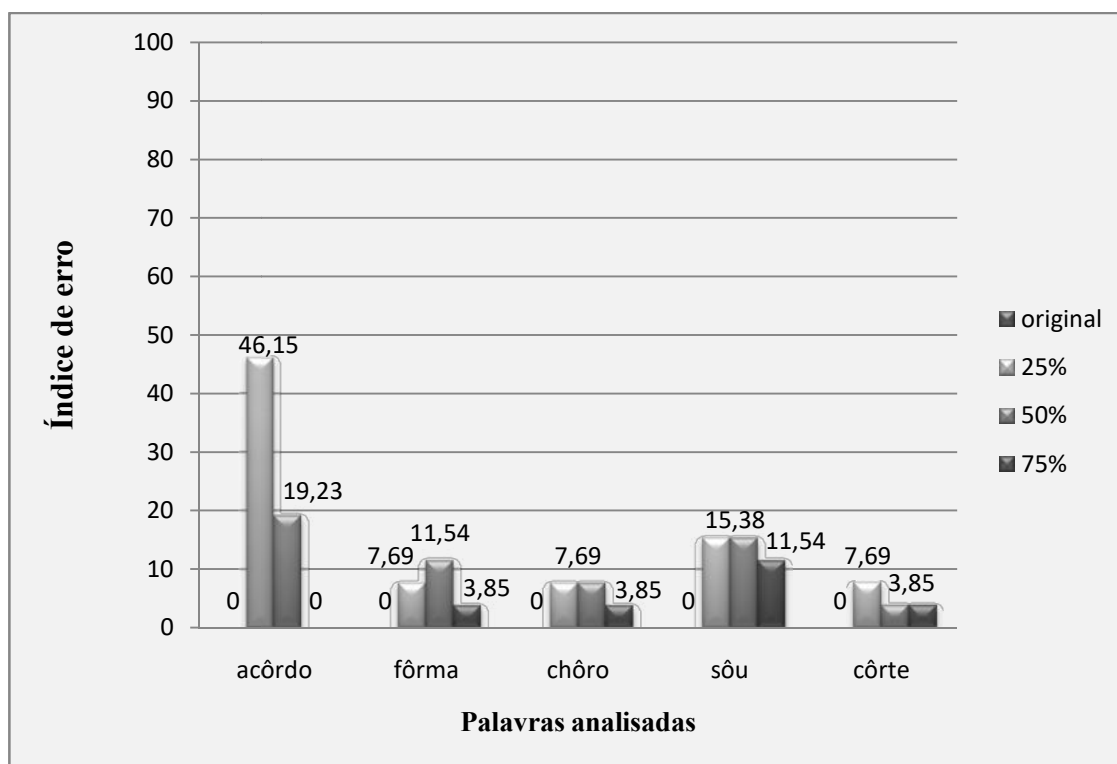
No caso da palavra *chôro*, a taxa percentual 25 interferiu na inteligibilidade da vogal em 3,85% dos casos, a manipulação em 50% suscitou índice de erro de discriminação de 15,38%, e a manipulação em 75% gerou índice de 7,69%. A redução da F_0 , para as três taxas de manipulação analisadas, interferiu na inteligibilidade do timbre vocálico da vogal em questão.

Na palavra *sól*, a taxa percentual 25 gerou índice de erro de 19,23%, na taxa de 50%, o índice de erro de inteligibilidade caiu para 3,85%, e para 0% com a ascensão da redução para a taxa de 75%.

Na palavra *côrte*, a redução da F_0 em 25% e 50% não interferiu na percepção do sinal; já a taxa de manipulação de 75% gerou, nesse contexto, uma interferência na inteligibilidade do sinal de 3,85%.

É possível observar, a partir das análises a cima, que a redução da F_0 gerou padrões diferenciados para os diferentes contextos de fala em proporções também distintas.

Gráfico 8 – Índice de erro de discriminação da vogal média arredondada para valores reduzidos da F_0



Observa-se, com base nos gráfico 8, que as palavras *acôrdo* apresentou, para a redução da F_0 em 25%, taxa de erro de discriminação de 46%, para a taxa de manipulação em 50%, a taxa de erro cai para 19,23, e com a redução em 75%, a taxa de discriminação foi de 100%, ou seja, os ouvintes conseguiram recuperar 100% do sinal original, apesar da redução da F_0 em 75%.

Na palavra *fôrma*, a redução de 25% gerou erro de 7,69; na redução de 50%, o erro aumentou para 11,54%, e a redução da F_0 para 75%, o erro de discriminação decresceu cerca de 7%, ou seja, a interferência da F_0 na percepção do sinal foi em menor escala com a ascensão da redução.

Na palavra *chôro*, as taxas de 50% e 25% de redução da F_0 , acarretaram índice de erro de discriminação, 7,69%. Para 75% de redução da F_0 , a inteligibilidade do sinal foi afetada em 3,85%.

Para a palavra *sôu*, as taxa de manipulação de 50% e 75% geraram mesmo índice de erro de discriminação 3,85%. Para a taxa de redução de 25%, a inteligibilidade do sinal sofreu interferência de 7,69%. A redução da F_0 acarretou, portanto, alteração na discriminação da vogal média alta arredonda para os referidos contextos.

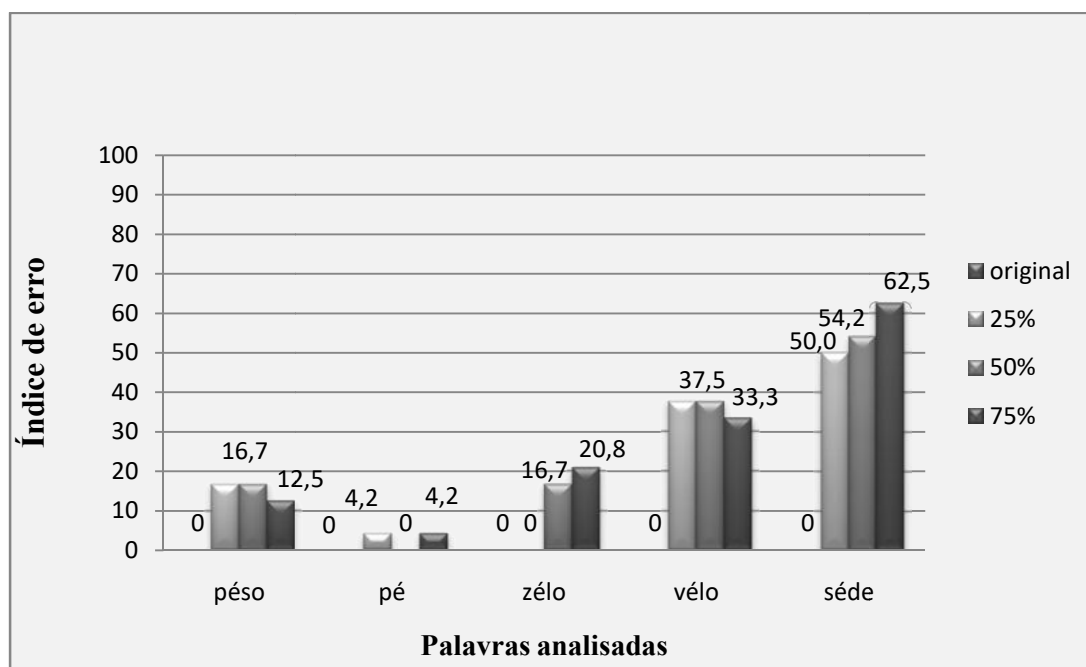
4.2 TESTE 2: IDENTIFICAÇÃO

Uma vez atestado que o ouvinte é sensível às mudanças na F_0 , foi objetivo ainda desse trabalho avaliar se essa sensação auditiva de alteração da fundamental tinha implicações para a percepção do timbre vocálico. Para dar conta desse objetivo, foi aplicado o teste de identificação, que buscou avaliar se o sujeito da pesquisa era capaz de determinar o timbre da vogal presente no estímulo sonoro que lhe era apresentado. Com esse procedimento metodológico, foi possível verificar se a manipulação da F_0 acarreta prejuízos ou não na percepção das vogais médias abertas. Foi possível avaliar se a inteligibilidade dessas vogais pode estar ou não associada aos valores da fundamental ou se sua inteligibilidade está garantida única e exclusivamente com o seu padrão formântico.

Breve discussão e explanação dos resultados referentes à influência da redução e ampliação da F_0 na identificação da vogal média são apresentadas nesta sessão.

4.2.1 Análise dos testes de identificação para valores ampliados da F_0

Gráfico 9 – Índice de erro de identificação da vogal média baixa não arredondada para valores ampliados da F_0



Observa-se, a partir do gráfico acima, que a ampliação da F_0 na palavra *péso* para os valores de 25% e 50% afetou a identificação da vogal média em questão em índices similares, 16,7%. Com a ampliação da F_0 para 75%, o índice de erro de identificação caiu para 12,5%, ou seja, apesar de maior taxa de ampliação, os ouvintes conseguiram ampliar a acuidade perceptiva.

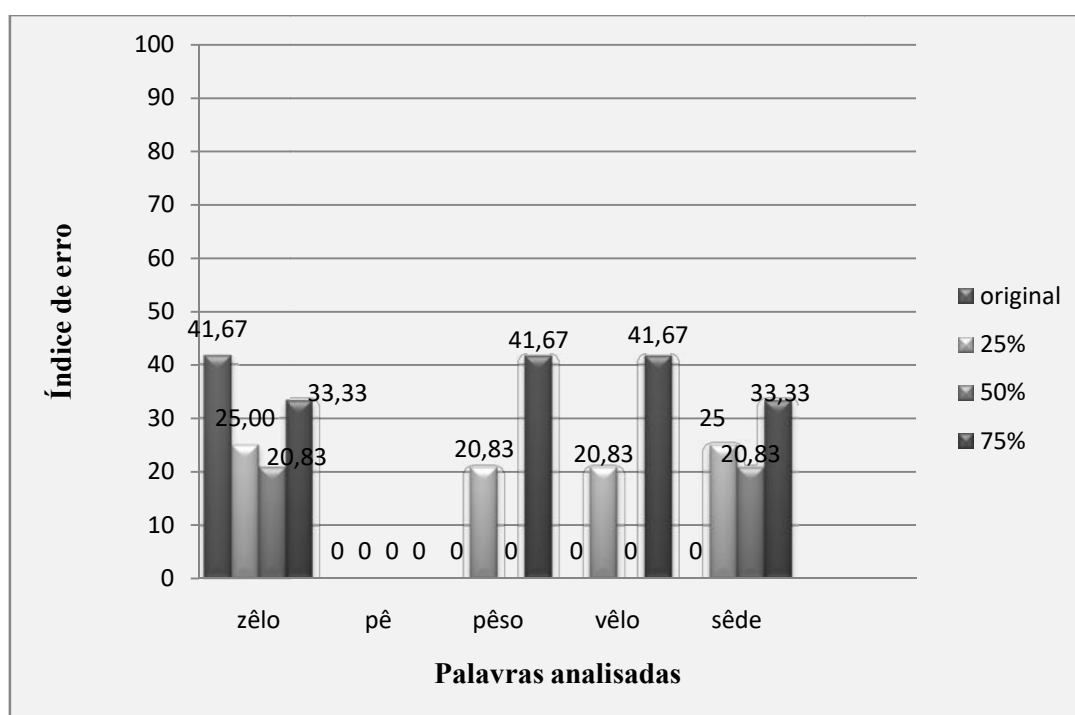
Na palavra *pé*, a ampliação da F_0 em 25% e 75% suscitou interferências na identificação da palavra em questão de 4,2%. No entanto, quando a F_0 foi ampliada em 50% do valor original, os ouvintes conseguiram identificar essa vogal em 100% dos casos, ou seja, a ampliação do referido valor não acarretou prejuízos para a identificação da vogal em questão.

Na palavra *zélo*, a ampliação da F_0 gerou para os diferentes valores percentuais 25, 50 e 100 investigados, diferentes taxas de erro de identificação. A ampliação em 25%, não gerou prejuízo na inteligibilidade do sinal, o ouvinte recuperou em 100% dos casos o sinal investigado. Já a ampliação em 50% gerou erro de identificação de 16,7% e a ampliação em 75%, erro de identificação de 20,8%. A ampliação da F_0 interferiu, portanto, na percepção do sinal de fala analisado.

A palavra *vêlot* também sofreu interferência para os três valores investigados. As manipulações de 25% e 50% geraram erro de 37,5%, já a ampliação em 75% ocasionou erro de 33,3%.

O mesmo ocorreu para a palavra *sêde*, ou seja, a manipulação da F_0 interferiu na inteligibilidade do timbre vocálico. A ampliação em 25% suscitou taxa de erro de 50%, com a ampliação da F_0 para o índice percentual de 50%, o índice de erro de identificação ascendeu para 54,2%. Com a ampliação da F_0 em 75% em relação ao valor original, a percepção do timbre vocálico foi ainda mais afetada, 62,5%, ou seja, com a ampliação da F_0 em 75%, o ouvinte somente conseguiu recuperar o sinal original em 37,5%. A ampliação da F_0 nesse contexto, portanto, acarretou interferências para a identificação da vogal média baixa não arredondada.

Gráfico 10 – Índice de erro de identificação da vogal média alta não arredondada para valores ampliados da F_0



Os resultados apontam que na palavra *pê*, a ampliação da F_0 em 25%, 50% e 75% do valor original não gerou erros de identificação. No entanto, a ampliação da F_0 para os valores de ampliação investigados nas palavras *pêso*, *vêlo*, *sêde* e *zêlo* ocasionou erros de identificação.

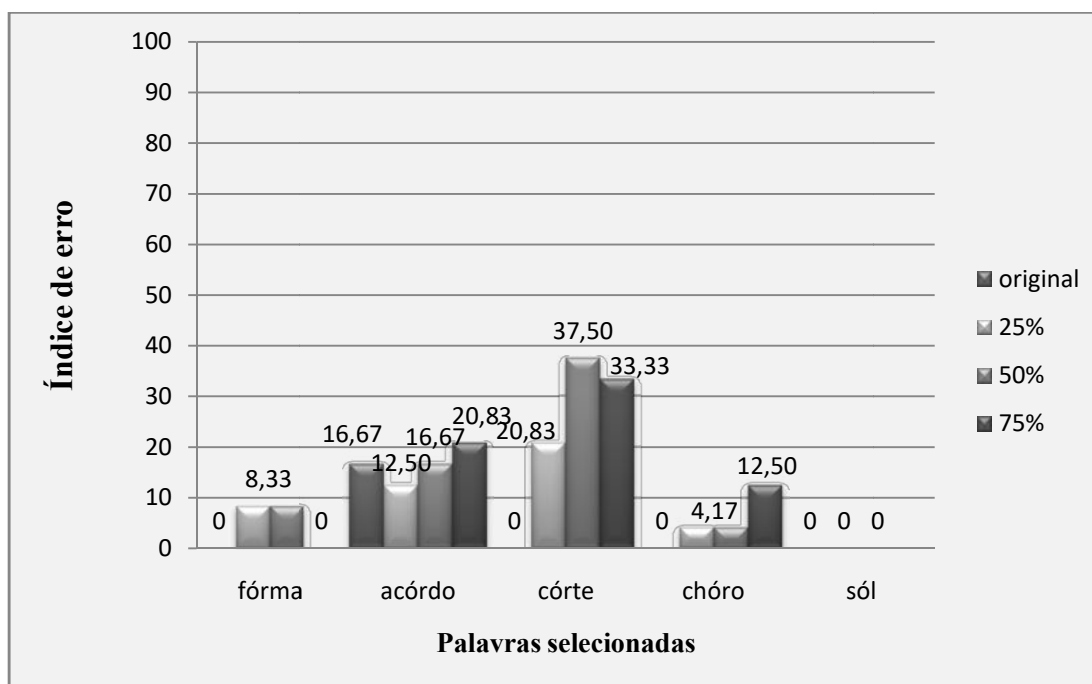
A ampliação da F_0 em 25% para a palavra *zêlo* gerou índice de erro de identificação de 25%. Esse índice caiu para 20% quando a manipulação foi ampliada para 50%, e voltou a

aumentar para 33,3% com a ampliação de 75%. No sinal original, a palavra *zêlo* sofreu de identificação de 41,67%.

Na palavra *pêso*, a ampliação da F_0 em 25% gerou erro de 20,83%. Quando a palavra foi manipulada em 50% do valor original da fundamental, o ouvinte conseguiu recuperar a palavra em questão em 100% dos casos investigados. Para a ampliação em 75%, o erro de identificação foi de 41,67%.

A manipulação da F_0 na palavra *êde* interferiu na inteligibilidade do timbre vocálico nos três valores de ampliação da F_0 investigados. A ampliação da F_0 em 25% ocasionou erro de 25%. Quando a manipulação foi ampliada para 50%, o índice de erro ascendeu para 20,8% e decresceu para 33,3% quando a manipulação da F_0 foi ampliada para 75%. Nesses contextos, com a ampliação da F_0 , os ouvintes não conseguiram recuperar a vogal média alta não arredondada em mais de 40% dos casos investigados.

Gráfico 11 – Índice de erro de identificação da vogal média baixa arredondada para valores ampliados da F_0

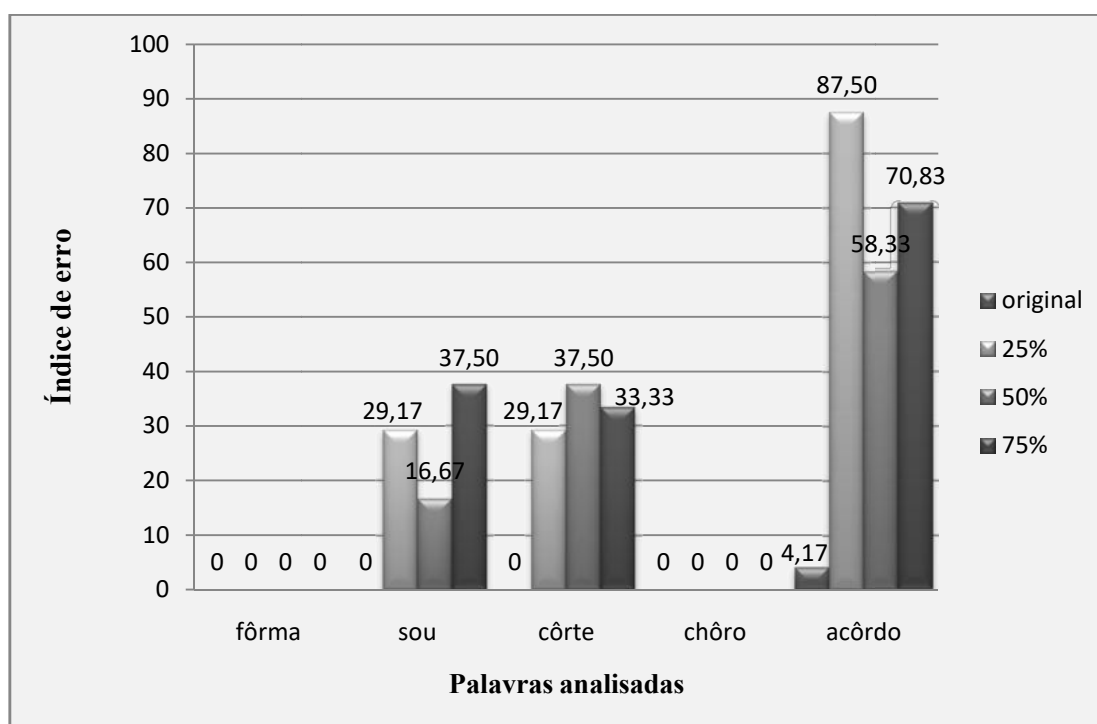


A ampliação da F_0 na palavra *fórma* gerou erro de identificação de 8,33% tanto para a taxa de manipulação de 25% quanto para a taxa de 50%, já a taxa de ampliação de 75% não ocasionou erro de identificação nesse contexto; no sinal original, o ouvinte conseguiu recuperar a palavra *fórma* em 100% dos casos investigados. Na palavra *acórdo*, o sinal original ocasionou taxa de erro de 16,6%; a ampliação da F_0 em 25% gerou erro de 12,5%, o índice de

erro sofreu uma elevação de 4,1%, com a ampliação da fundamental para 50%; a 75% de ampliação, o índice de erro de identificação foi de 8,33%. Na palavra *côrte*, a ampliação da F_0 gerou prejuízo de identificação: a 25% os ouvintes não recuperaram o sinal em 20,8% dos casos investigados, a 50% o erro de identificação ascendeu para 37,5%; a 75% ocasionou erro de 33,3%.

A palavra *chôro* obteve índice de erro para os valores de 25% e 50% similares, qual seja, 4,1%. Com a ampliação da F_0 para 75%, o erro de identificação foi ampliado para 12,5%. Na palavra *sól*, as taxas de ampliação da F_0 para os valores investigados não geraram prejuízos de identificação.

Gráfico 12 – Índice de erro de identificação da vogal média alta arredondada para valores ampliados da F_0



Na palavra *fôrma*, a ampliação da F_0 nas referidas taxas investigadas não gerou erro de identificação da vogal. Já na palavra *sôu*, a ampliação da F_0 em 25% afetou a identificação da vogal média em questão em 29,1% dos casos investigados; com a ampliação da F_0 em 50%, o erro de identificação foi de 16,6% e para a taxa de manipulação da F_0 em 75%, o erro de identificação foi de 37,5%.

Na palavra *côrte*, a ampliação da F_0 em 25% gerou índice de erro de identificação de 29,1%, esse índice de erro ascendeu para 37,5% com a ampliação da F_0 em 50% e caiu para

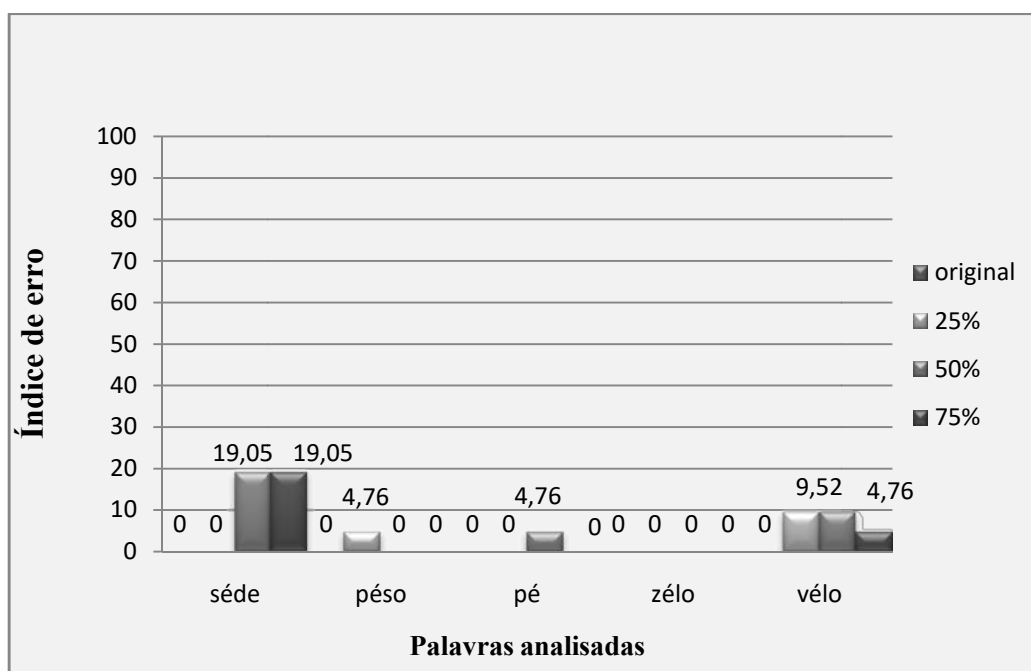
33,3% com a ampliação da F_0 para 75%. A palavra *chóro*, nas taxas percentuais de ampliação da F_0 investigadas, não sofreu erros de identificação.

A palavra *acôrd* obteve erro de 4,7% para o sinal original. Na ampliação 25%, o erro de discriminação foi de 87,5%, e caiu para 58,3% com a ampliação da F_0 para 50%. A 75% de ampliação o erro de identificação foi de 70,8%. Os resultados apontam que a vogal média alta não arredondada, nesse contexto, sofreu prejuízos de identificação com a ampliação da F_0 para os valores de ampliação investigados.

4.3.2 Análise dos testes de identificação para valores reduzidos da F_0

Os gráficos abaixo correspondem aos resultados parciais dos testes de identificação para redução da F_0 referentes às vogais médias.

Gráfico 13 – Índice de erro de identificação da vogal média baixa não arredondada para valores reduzidos da F_0



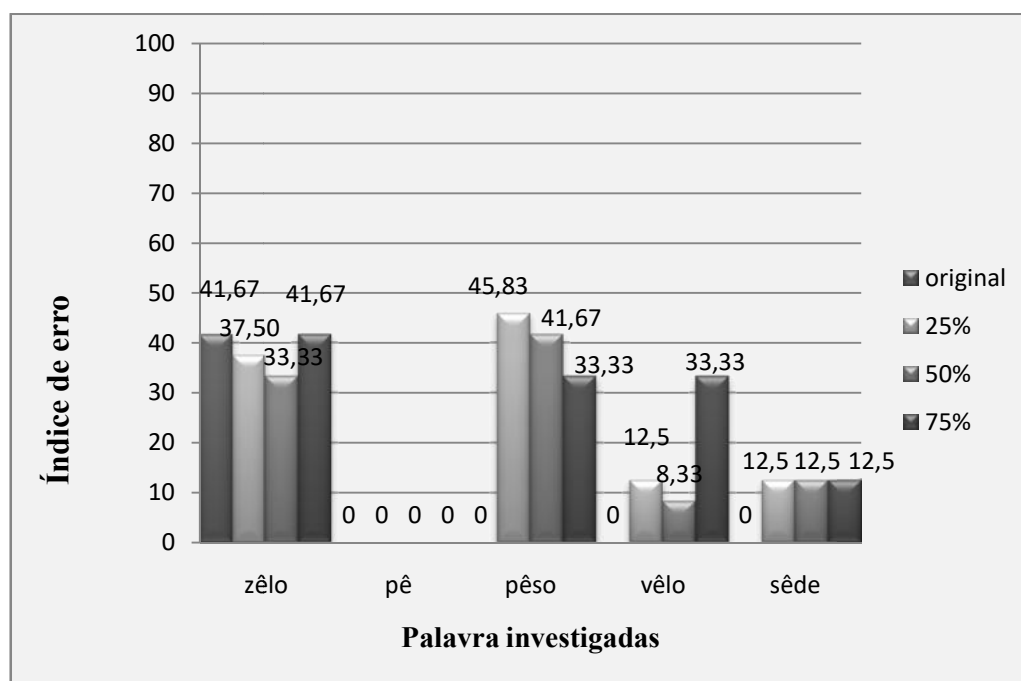
Na palavra *séde*, a taxa de ampliação da F_0 a 25% não gerou prejuízos para a identificação da vogal, os ouvintes recuperaram o sinal original em 100% dos casos investigados; já as taxas de redução de 50% e 75% de redução da F_0 geraram prejuízos para a identificação da vogal com erro de identificação para ambos os casos de 19%.

As reduções de 50% e 75% na palavra *péso* não geraram prejuízos para a identificação da vogal, já a redução de 25% ocasionou taxa de erro de 4,76%.

Na palavra *pé*, a taxa de redução da F_0 ocasionou erro de identificação apenas para redução de 50%, com índice de erro de 4,76%. A redução da F_0 na palavra *zêlo* não afetou a identificação das vogais médias baixas nos três valores de redução da F_0 investigados.

Na palavra *pé*, a redução da F_0 em 25% e 50% ocasionou mesmo índice de erro de identificação, 9,5%. Para a taxa de 75%, o erro para a referida palavra caiu para 4,7%.

Gráfico 14 – Índice de erro de identificação da vogal média alta não arredondada para valores reduzidos da F_0



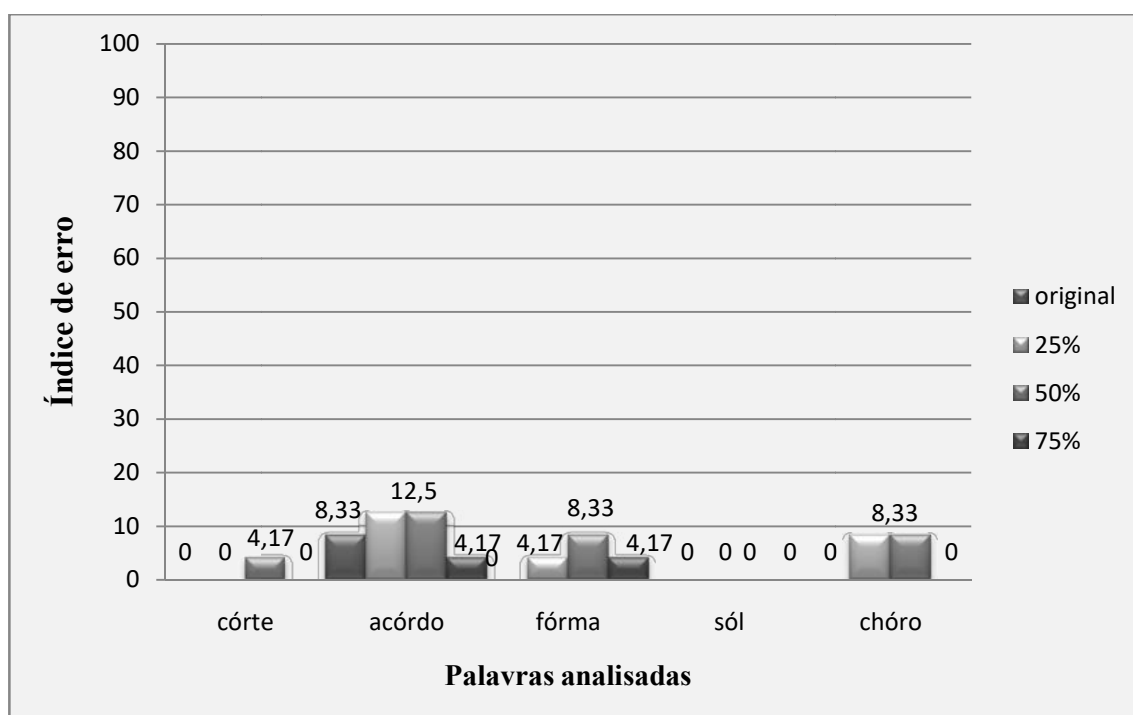
É possível observar a partir do gráfico 14 que a identificação da palavra *zêlo* foi afetada para as três taxas de manipulações investigadas, bem como para o sinal original. Para os valores manipulados, o maior índice de erro foi para a redução em 75%, a menor para 50%, e interferência intermediária para a taxa de 25%. A redução da F_0 interferiu na inteligibilidade da vogal média alta não arredondada para esses contextos em mais de 40% dos casos investigados.

A inteligibilidade da palavra *pê* não sofreu interferência das reduções da F_0 investigadas. Já na palavra *pêso*, a redução da F_0 em 25%, 50%, 75% ocasionou alteração na identificação da vogal em relação ao sinal original em relação inversa: quanto maior a taxa de redução, menor o índice de erro de identificação e vice-versa. A redução da fundamental em 25% ocasionou erro de 45,8%, erro de 41,6% na redução a 50% e 33,3%, na redução de 75%.

A discriminação da palavra *vêlo* sofreu interferência de 12,5% quando a fundamental foi reduzida em 50%. O índice de erro caiu para 8,3% com a redução em 50% e aumentou para 33,3% para a redução de 75%. A palavra *sêde* obteve mesmo índice de erro de identificação para os três casos de manipulação investigados. A percepção da vogal média baixa não arredondada, em alguns contextos, sofreu interferência com a redução da F_0 .

No gráfico 15, são apresentados os resultados obtidos para a palavra *córte*, *acórdo*, *fórma*, *chóro* e *sól*.

Gráfico 15 – Índice de erro de identificação referente à vogal média baixa arredondada para valores reduzidos da F_0



A identificação da palavra *córte* sofreu interferência em 4,1% na redução de 50%. As reduções de 25% e 75%, bem como o original não ocasionaram erro de identificação.

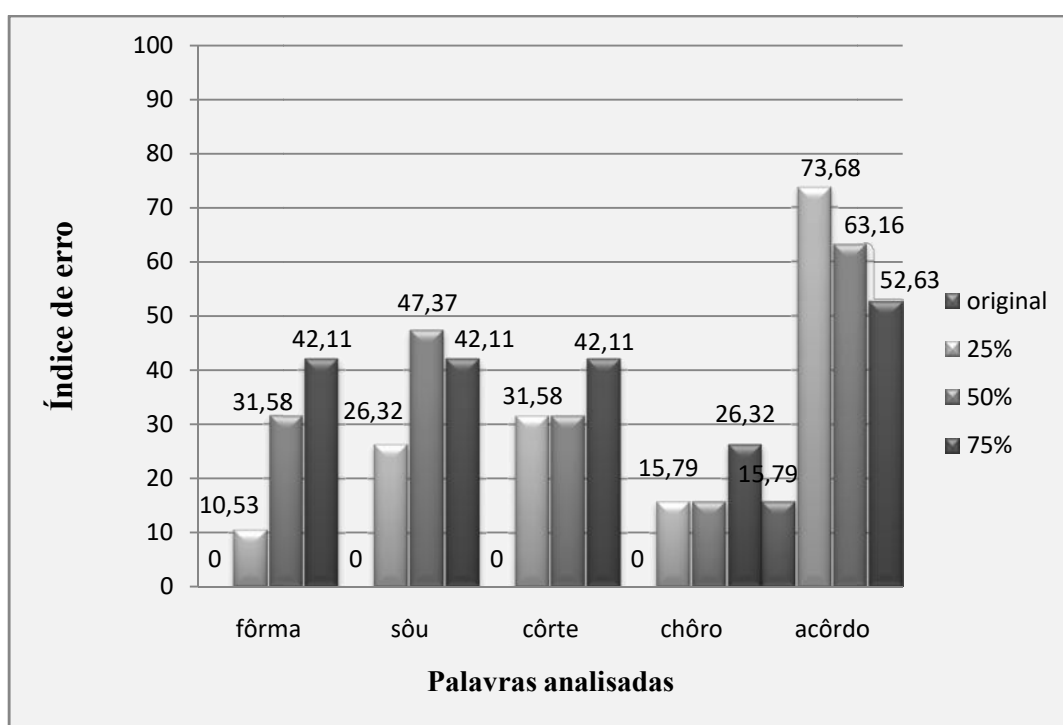
Na palavra *acórdo*, a redução da F_0 nas três taxas investigadas, bem como a ausência da redução suscitaram interferência na perceptibilidade. Assim, no sinal original a taxa de erro foi de 8,3%. Para a taxa de manipulação de 25% e 50%, os ouvintes não conseguiram recuperar a vogal média baixa arredondada em 12,5% dos casos analisados, essa porcentagem de erro caiu para 4,1% com a redução da F_0 para 75%.

Os dados apontam que na palavra *fórma*, o sinal original obteve de erro de identificação igual a 4,17%. Quando a F_0 foi reduzida em 25% do valor original, a interferência

na identificação foi de 8,3%. O índice de redução de 50% gerou erro de 4,1%, 0% quando a redução foi para 75%.

Na palavra *sól*, as taxas de redução da F_0 em 25%, 50% e 75% não implicaram em prejuízos para a inteligibilidade da palavra. Na palavra *chóro*, a redução da F_0 em 25% e 50% gerou erro de 8,33%; quando a manipulação foi reduzida para 75% do valor do sinal original, o índice de erro caiu para 0%.

Gráfico 16 – Índice de erro de identificação referente à vogal média alta arredondada para valores reduzidos da F_0



Os dados apontam que na palavra *fôrma*, no sinal original, o erro foi de 0%, já as a redução de 25% gerou erro de discriminação de 10,5%. A redução da F_0 em 50% gerou erro de discriminação de 31,5% e erro de identificação de 42,1% com a redução da F_0 a 75% do valor original.

Na palavra *sôu*, o sinal original não ocasionou erro de identificação, enquanto a redução da F_0 em 25% afetou a identificação da vogal média em 26,3% dos casos investigados. A redução da F_0 em 50% gerou erro de identificação de 47,3%. Com a redução da F_0 em 75%, o erro caiu para 42,1%. Na palavra *côrte*, as taxas de redução de 25% e 50% geraram erro de identificação similar, sendo o valor 31,5%.

A palavra *chôro* não obteve erro para o sinal original, já para as taxas de redução da F_0 nos índices de 25% e 50% gerou erro de identificação de 15,7%. Para as taxas de 75%, a identificação da vogal foi afetada em 26,3%.

A palavra *acôrdo* foi afetada no sinal original em 15,7%, as taxas de redução da F_0 a 25%, 50% e 75% geraram erro de identificação em escala decrescente, assim, para 25% de redução, o erro foi de 73,68%; para 50%, o erro de identificação foi de 63,1%, e caiu para 52,6% com a redução da F_0 a 75%. Os dados apontam, portanto, que a redução da F_0 para os valores de redução investigados gerou porcentagem de erro de identificação entre 40% e 70%.

4.4 DISCUSSÕES

Diante dos resultados obtidos no teste de discriminação e no teste de identificação, é possível afirmar que a alteração da F_0 compromete a inteligibilidade das vogais médias, levando o ouvinte a confundir um timbre pelo outro.

Esses resultados podem ser explicados à luz da Teoria Quântica da percepção da fala, idealizada por Stevens, como descrito no item 2.4.2. Se se considerar, com base nessa teoria, que a percepção de um segmento ocorre a partir das respostas quânticas do sistema auditivo, alterações no sinal acústico em taxas diversas nessas regiões podem alterar a percepção dos segmentos vocálicos, o que explicaria os diferentes índices de erros nas diferentes taxas de manipulação.

Alterações da F_0 podem alterar as respostas quânticas do sistema auditivo. Assim, o que estaria em jogo não seria a dimensão da manipulação da fundamental (25%, 50% ou 75%), mas o quanto esta afeta as respostas quânticas do sistema auditivo. Isso pode explicar, por exemplo, porque em muitos casos, a manipulação da F_0 a 25% provoca mais prejuízos do que a manipulação a 75%. A hipótese aqui aventada é a de que a manipulação a 25%, naquele sinal acústico, altera uma região quântica mais importante para a identificação do timbre vocálico do que a 75%, que consegue ter ainda nos 25% restantes uma região quântica eficaz para a percepção desses vocóides.

Se alterações da F_0 podem alterar as respostas quânticas do sistema auditivo na percepção das vogais médias, tem-se, portanto, que esse parâmetro acústico presente no espectro sonoro não é descartado pelo ouvinte, no ato da percepção, como sugerem Irino e Peterson (2002), na Teoria independente, segundo a qual não existe uma relação entre a F_0 e a qualidade acústica vocálica.

Os resultados obtidos evidenciam que, ao contrário do que propõem aqueles autores, haja uma relação entre a fundamental e o padrão formântico da vogal, como sugerem LehisteeMeltzer(1972); SyrdaleGopal(1986) e Assmann(2007).

Nesse sentido, a frequência fundamental seria uma informação no sinal acústico importante para a identificação da vogal. A alteração dos valores da F_0 compromete em maior ou menor escala, a depender da resposta quântica do sistema auditivo, a referência sensorial, calculada a partir da relação algorítmica da fundamental com os formantes (MILLER, 1989; SYRDAL; GAPOL, 1986).

A relação algorítmica entre F_0 e F_1 , F_2 e F_3 pode, pois, explicar a ausência de uma relação diretamente proporcional entre taxa de manipulação e índice de erro. A perceptibilidade do timbre vocálico não deve ser explicada pelo quanto de F_0 foi manipulado, mas a partir da relação da F_0 que ainda restou no sinal acústico e o padrão formântico original, sem perder de vista as respostas quânticas do sistema auditivo.

5 CONSIDERAÇÕES FINAIS

Os resultados apontam que a manipulação da Frequência Fundamental tanto para valores ascendentes quanto para valores descendentes afetam a percepção da vogal média. Em muitos casos o ouvinte identificou a vogal média alta como sendo uma vogal média baixa e uma vogal média baixa como sendo uma vogal média alta. Houve ainda casos em que o informante identificou a vogal /□/ como /u/. Assim, tanto a ampliação quanto à redução da F_0 geraram erros na discriminação e na identificação da vogal média.

Os dados sinalizam, portanto, que a F_0 é um importante recurso utilizado pelo ouvinte para a discriminação da vogal média, uma vez que, em alguns contextos, a F_0 afetou diretamente a percepção das vogais em questão. Os resultados apontam que a frequência fundamental pode ser um recurso utilizado pelo ouvinte não apenas para a discriminação de falantes, mas um importante recurso utilizado pelo ouvinte para diferenciar o timbre vocálico.

Os resultados assinalam que além da configuração formântica, outros parâmetros espectrais podem contribuir significativamente para a compreensão de como se dá a percepção das vogais médias, ou seja, o ouvinte talvez utilize outras informações do sinal acústico como recurso para a discriminação do timbre vocálico.

REFERÊNCIAS

- ASSMANN, P. F.; NEAREY, T. M.; HOGAN, J. T. Vowel identification: Orthographic, perceptual and acoustic aspects. **Journal of the Acoustical Society of America**, Canada, v. 71, p. 975-989, 1982.
- BARREDA, S.; NEAREY, T. M. The relationship between fundamental frequency and vowel quality. **Journal of the Acoustical Society of America**, Canada, v. 127, n. 3, p. 219, 2011.
- BEHLAU M.; PONTES P. **Avaliação e Tratamento das Disfonias**. São Paulo: Lovise, 1995.
- BERGEM, D. R van; POLS, L. C. W; F. J; BEINUM, K van, F.J. Perceptual normalization of the vowels of a man and a child in various contexts. **Speech Communication**, Amsterdam, v. 7, p. 1-20, 1988.
- BOERSMA, P; WEENINK, D. **Praat**: Doing phonetics by computer. Version 5.1.01, <http://www.praat.org>, 2009.
- BOUDAHMANE, K; MANTA, M.; ANTONOINE, F.; GALLIANO, S.; BARRAS, C; EDOUARD, G.; LIBERMAN, M.; WU, Z. **Transcriber**: a Free Tool for Segmenting, Labeling and Transcribing Speech. First International Conference on Language Resources and Evaluation, Filadélfia: Oasis, 2000, <http://trans.sourceforge.net>.
- CÂMARA JR, J.M. **Estrutura da Língua Portuguesa**. Rio de Janeiro: Vozes, 1992.
- CARLSON, R.; FANT, G; GRANSTROM, B. Two-formant models, pitch, and vowel perception. In: SYRDAL, A. K; GOLPAL, H. S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. **Journal of the Acoustical Society of America**, Naperville, v. 79, p.1086-1100, 1998.
- CHOMSKY, N.; HALLE, M. **The Sound Pattern of English**. New York: Harper e Row, 1968.
- CLAES, T.; DOLOGLOU, I; BOSCH, L. T; COMPERNOLLE, V. D. A novel feature transformation for vocal tract length normalization in automatic speech recognition. In: Auditory Sensitivity to Formant Ratios: Toward an Account of Vowel Normalization. **NHI Public Access Author Manuscript**, Maryland, v. 25, n. 6, p. 808–839, 1998.
- CLEMENTS, J. N.; HUME, E. V. The internal organization of speech sounds. In: GOLDSMITH, J. (ed.). **The handbook of phonological theory**. London: Brasil Blackwell, 1995.
- COLEMAN, R.O. A comparison of the contribution of two voice quality characteristics to the perception of maleness and femaleness in the voice. **Journal of Speech and Hearing Research**, Oregon, v.19, p. 168-180, 1976.
- DELGADO-MARTINS, Maria Raquel. **Fonética do Português: Trinta anos de Investigação**. Lisboa: Caminho, 2002.

DIEHL, R. L; KLUENDER, K. R; WALSH, M. A. Some auditory bases of speech perception and production. In: Effects of contrast between onsets of speech and other complex spectra. COADY, J. A; KLUENDER, K. R; RHODE, W. S. **Journal of the Acoustical Society of America**, Madison, v.114, n. 4, p.1-11, 2003.

FANT, G. **Acoustic theory of speech production**: with calculations based on x-ray studies of Russian articulators. The Hague: Mouton, 1960.

GELFER, M.P., MIKOS, V.A. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. **Journal Voice**, Milwaukee, v. 19, n. 4, p. 544-554, 2005.

GERSTMAN, L. J. Classification of self-normalized vowels, IEEE Trans. **Audio and Electroacoustic**. AU-16, p. 78-80, 1968.

HAWKINS, S; STEVENS. K. N. Acoustic end perceptual correlates of non-nasal/ nasal distinction of vowels. **Journal of the Acoustical Society of America**, Michigan, v. 77, n. 4, p. 1560-1575, 1985.

HARRIS, M. C. A Study of the Building Blocks in Speech. **Journal of the Acoustical Society of America**, Columbia, v.25, n. 5, p. 962-969, 1953.

HILLENBRAND, J. M., CLARK, M. J.; HOUDE, R. A. Some effects of duration on vowel recognition. **Journal of the Acoustical Society of America**, Kalamazoo, v. 108, n. 6, p. 1-10, 2000.

HILLENBRAND, J; GAYVERT, R. T. Identification of steady-state vowels synthesized from the Peterson and Barney measurements. **Journal of the Acoustical Society of America**, Kalamazoo, v.94, n. 2, p.668-674, 1993.

IGNATIUS, G.; MATTINGLY, I. G; A. M. LIBERMAN. **The speech code and the physiology of language**. New York City: Storrs, 1985.

IRINO, T.; PATTERSON, R. D. Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform. **Speech Commun.** Cambridge, v. 36, p. 181–203, 2002.

JENKINS, J. J.; STRANGE, W.; EDMAN, T. R. Identification of vowels in "vowelless" syllables. **Perception and Psychophysics**, Minneapolis, v. 34, p. 441-450, 1983.

JOHNSON, K. The role of perceived speaker identity in f0 normalization of vowels. **Journal of the Acoustical Society of America**, Bloomington, v. 88, p. 642–654, 1990.

JOOS. M. Acoustic phonetics Language Monograph. In: SUEN, C. Y; MICHAEL P. B. Discrimination of vowels sound of very short duration. **Attention, Perception, & Psychophysics**, Canada, v. 11, n. 6, p. 24, 1972.

KENT, R. D.; READ, C. **The acoustic analysis of speech**. San Diego: Singular Publishing Group, 1992.

- LADEFOGED, P. **Elements of acoustic phonetics**. Chicago: University of Chicago Press, 1996.
- LADEFOGED, P.; BROADBENT, D. E. Information conveyed by vowels. **Journal of the Acoustical Society of America**, Cambridge, v. 29, n. 1, p. 1-7, 1957.
- LEHISTE, I; PETERSON, G. E. Transitions, glides, and diphthongs. **Journal of the Acoustical Society of America**, Michigan, v. 33, n. 3, p. 268-277, 1961.
- LEHISTE, I; MELTZER, D. Vowel and speaker identification in natural synthetic speech. **Journal of the Acoustical Society of America**, Columbus, v. 51, n. 1a, p. 131-131, 1972.
- LIBERMAN, A. M. Some results of research on speech perception. **Journal of the Acoustical Society of America**, Connecticut, v. 29, n. 1, p. 1-7, 1957.
- MACCHI, M. J. Identification of vowels spoken in isolation and versus vowels spoken in consonantal context. **Journal of the Acoustical Society of America**, New Jersey, v. 68, n. 6, p. 1636-1642, 1980.
- MALMBERG, B. **A fonética**. Lisboa: Livros do Brasil, 1954.
- MATEUS, Maria Helena Mira. **Estudando a melodia da fala: Traços prosódicos e constituintes prosódicos**. Setúbal: APLS, 2004.
- MATEUS, Maria Helena Mira; ANDRADE, Amália; VIANA, Maria do Céu; VILLALVA, Alina. **Fonética, Fonologia e Morfologia do Português**. Lisboa: Universidade Aberta, 1990.
- MILLER J. D. Auditory-perceptual interpretation of the vowel. **Journal of the Acoustical Society of America**, Missouri, v. 85, p. 2114-2134, 1989.
- MORAIS-BARBOSA, Jorge. **Études de phonologie portugaise**. Lisboa: Junta de investigações Nacional de Ultramar, 1965.
- NASCENTES, A. V. **O linguajar carioca**. Rio de Janeiro: Organização Simões, 1953.
- NEAREY, T. M. Static, dynamic, and relational properties in vowel perception. **Journal of the Acoustical Society of America**, Bloomington, v. 85, p. 2088-2113, 1989.
- NEAREY, T. M.; ASSMANN P. F. Probabilistic 'sliding template' models for indirect vowel normalization. In: **Experimental Approaches to Phonology**, ed. M. J. SOLE', P. S. BEDDOR, e M. OHALA, Oxford, p. 246-269, 2007.
- OLIVEIRA, J.N.; RIBEIRO, P.J.; PACHECO, V. As vogais médias abertas em posição pretônica no dialeto baiano. In: SEMINÁRIO DO GEL, 53, 2005, Campinas. **Caderno de Resumos [...]**, Campinas, 2005.
- PICKETT, J. M. **The acoustics of speech communication: Fundamentals, speech perception, theory and technology**. Boston: Allyn and Bacon, 1996.

RAKERD, B.; VERBRUGGE, R. R. Evidence that the dynamic information for vowels is talker independent in form. *Journal Memory Language*, v. 26, p. 558-563, 1982.

RYALLS, J. H. E.; LIEBERMAN, P. Fundamental frequency and vowel perception. **Journal of the Acoustical Society of America**, Providence, v. 72, p.1-4, 1982.

SILVA, M. B. **Breve notícia sobre as vogais pretônicas na variedade culta de Salvador**. Salvador, EstudosLinguísticos e Literários, 1993.

SIU-FUNG P.; MANWA L. N. Contribution of voice fundamental frequency and formants to the identification of speaker's gender. **International Congress of Phonetic Sciences**, Hong Kong, v. 17, p. 17-21, 2011.

SKINNER, T. E. Speaker invariant characterizations of vowels, liquids, and glides using relative variant formant frequencies. **ournal of the Acoustical Society of America**, Saint Paul, v. 162, 1977.

SMITH, D. R. R.; PATTERSON, R. D.; TURNER, R.; KAWAHARA, H.; IRINO, T. The processing and perception of size information in speech sounds. **Journal of the Acoustical Society of America**, Cambridge, v 117, n. 1, p. 305–318, 2005.

STRANGE, W. Cross-language studies of speech perception. In: Winifred Strange. **Speech Perception and Linguistic Experience: Issues in Cross-Language Research**. Baltimore: York Press, 1995.

STRANGE, W. Dynamic specification of coarticulated vowels spoken in sentences context. **Journal of the Acoustical Society of America**, Florida, v. 85, p. 2135-2153, 1989.

SYRDAL, A. K; GOLPAL, H. S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. **Journal of the Acoustical Society of America**, Naperville, v. 79, p.1-15, 1986.

TITZE, I. R. Phonation threshold pressure: A missing link in glottal aerodynamics. *The Journal of the Acoustical Society of America*, Iowa City, v. 91, p. 2926-2935, 1992.

TUERNER, R. E; PATTERSON, R. D. An analysis of the size information in classical formant data: Peterson and Barney (1952) revisited. **The Acoustical Society of Japan**, Cambridge, v. 33, n. 9, p. 585-589, 2003.

WHITESIDE, S.P. The identification of a speaker's sex from synthesized vowels. **Perceptual and Motor Skill**, Sheffield, v. 87, p. 595-600, 1998.