

# *Uma Investigação Experimental acerca da Influência da $F_0$ na Percepção do Timbre das Vogais Médias Arredondadas*

AN EXPERIMENTAL INVESTIGATION ABOUT THE INFLUENCE OF  $F_0$  ON THE PERCEPTION OF TIMBRE OF MEDIUM ROUNDED VOWELS

Priscila de Jesus **RIBEIRO\***

Vera **PACHECO\*\***

**Resumo:** O presente trabalho objetiva investigar se a frequência fundamental corrobora para a percepção das vogais médias arredondadas em ouvintes naturais de Vitória da Conquista, na Bahia. Propõe-se investigar se a manipulação da  $F_0$ , tanto para valores ascendentes quanto para valores descendentes, interfere na inteligibilidade das vogais médias arredondadas no dialeto dessa comunidade linguística de sorte a confundir o timbre vocálico a ser percebido. Para isso, foi manipulada a frequência fundamental de forma a obter valores ampliados e reduzidos em 25%, 50% e 75% do valor original. Os resultados apontam que a manipulação da frequência fundamental tanto para valores ascendentes quanto para valores descendentes afetaram a percepção da vogal média não arredondada. Em muitos casos, o ouvinte identificou a vogal média alta como sendo uma vogal média baixa e uma vogal média alta como sendo uma vogal média alta. Houve ainda casos em que o informante identificou a vogal /ô/ como /u/. Assim, tanto a ampliação quanto à redução da  $F_0$  gerou erros na discriminação e identificação da vogal média.

**Palavras-chave:** Frequência fundamental. Percepção. Vogal média não arredondada.

---

\* Mestrado em Linguística na UESB (2012). Contato: priscilla.jribeiro@gmail.com.

\*\* Doutorado em Linguística pela Universidade Estadual de Campinas (2006). Professora titular da Universidade Estadual do Sudoeste da Bahia. Contato: vera.pacheco@gmail.com.

**Abstract:** This paper aims to investigate if the fundamental frequency influences the perception of middle rounded vowels in listeners natural of Vitória da Conquista, Bahia. It is proposed to investigate if manipulation of  $F_0$  for upper values and lower values interfere the intelligibility of middle rounded vowels in this dialect, so that this language community confuses the perception of the vocal timbre. For this, we manipulated the fundamental frequency to obtain values enlarged and reduced by 25%, 50% and 75% of the original value. The results indicated that manipulation of the fundamental frequency for both, upper values as lower values, affected the perception of the middle rounded vowel. In many cases, the listener identified a lower rounded middle vowel as an upper middle rounded vowel and an upper middle vowel as a lower middle vowel. There were also cases in which the informant identified the vowel /ô/ as /u/. Thus, the increase and reducing the  $F_0$  generated errors in identification and discrimination of the middle rounded vowel.

**Key-words:** Fundamental frequency. Perception. Rounded middle vowel.

## Introdução e Justificativa

Uma das questões de principal interesse da fonética experimental consiste em compreender quais são as informações utilizadas pelos ouvintes na identificação das vogais. Alguns estudiosos (HILLENBRAND; GAYVERT, 1993; JENKINS; STRANGE; EDMAN, 1983; STRANGE, 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 propõem que há no estado estacionário um espaço acústico entre as frequências formânticas que determinam a percepção vocálica.

Outros pesquisadores (BERGEM; POLS; BEINUM, 1988; CLAES et al., 1998; 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) cuja proposta é integrar o estado estacionário e a natureza da transição formântica com a informação espectral. Segundo essa autora frequências formânticas juntamente com a duração e a frequência fundamental corroboram para a percepção vocálica.

Estudos que, além da frequência formântica, consideram outros detalhes 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, 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 com a informação acústica da  $F_0$ . Barreda e Nearley (2011) afirmam que essa relação da  $F_0$  na percepção vocálica é indireta, ou seja, nesse ponto de vista, o ouvinte usa a  $F_0$  primordialmente como recurso para a identificação de falantes. 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 a  $F_0$  manipulada tanto para valores superiores quanto inferiores 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 qualidade 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 não arredondadas nos ouvintes naturais de Vitória da Conquista, Bahia. Cumpre salientar que em Vitória da Conquista é recorrente a realização de vogais médias tanto em posição tônica quanto em posição pretônica. Assim, propõe-se investigar se a manipulação da  $F_0$ , tanto para valores ascendentes quanto para valores descendentes, interfere na inteligibilidade das vogais médias arredondadas no dialeto dessa comunidade linguística de sorte a confundir o timbre vocálico a ser percebido.

## 1 Arcabouço Teórico

### 1.1 Percepção de vogais

Teorias acerca 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 a compreensão de como se dá o processo de percepção das vogais (PICKETT, 1996).

### *1.1.1 Modelos 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. O autor propõe que as vogais produzidas por diferentes falantes (homens, mulheres e crianças) podem ser representadas 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 (1985) 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 (1985) 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  $F_3$ - $F_2$ ). Esses resultados sustentam 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 utilizaram estímulos de vogais sintéticas, em que a  $F_0$  foi sistematicamente manipulada, corroboraram com a hipótese de que a identificação de vogais pode estar diretamente influenciada pela  $F_0$ . Dentre esses estudos, pode-se citar Meltzer e Lehist (1972), que sintetizam estímulos vocálicos simulando homens, mulheres e crianças, baseando-se na média dos valores dos formantes dos três grupos. Cada grupo vocálico foi sintetizado nos três valores de  $F_0$  correspondentes 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, era esperado que a percepção fosse mais exata

se ambos,  $F_1$  e  $F_0$ , fossem aqueles correspondentes a um falante de mesma idade e sexo.

E, de fato, os resultados encontrados vão nessa direção: 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 salientam 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, foi verificado nesses estudos que 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, segundo os autores, indicam que a  $F_0$  afeta a identificação de vogais.

Ryalls e Liberman (1982) relatam resultados similares em um estudo em que vogais produzidas por homens, mulheres e crianças foram sintetizadas com três diferentes valores de  $F_0$ : i) típicos do sexo (135 Hz, para homem e 185 Hz, para mulher); ii) mais baixos do que o típico (100 Hz); e, iii) maiores do que o padrão (250 Hz). Segundo os autores, ambos os grupos de vogais, com  $F_0$  típico e com  $F_0$  mais baixo, as vogais foram mais bem identificadas do que aquelas 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 foram menos bem percebidas, independente de as frequências formânticas serem as de vogais produzidas por homens, mulheres e crianças. Para esses autores, a variação da  $F_0$  na síntese de vogais no estado estacionário influencia o modo como os falantes percebem o padrão formântico.

### *1.1.2 Modelos de normalização extrínseca*

Os modelos de normalização extrínseca defendem a hipótese 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, as informações sobre idade e sexo do falante são avaliadas em padrões de fala corrente e utilizadas como base para interpretar as vogais acusticamente ambíguas.

De acordo com esse modelo, os ouvintes interpretam os padrões formânticos das vogais em relação ao contexto em que elas ocorrem. Destarte, vogais individuais são percebidas relacionadas aos padrões de frequência dos formantes do inventário completo de um dado 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 frequência 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.

Para demonstrar que o “*frame references*”, estabelecido pelas características acústicas de fala, afeta a percepção de vogais, Ladefoged e Broadbent (1957) investigam a identificação de vogais quando os níveis de  $F_1$  e  $F_2$  foram manipulados em sílabas sintéticas do tipo /b-vogal-t/ contendo vogais /I, ε, æ/ ou /Λ/. Os autores encontraram mudanças significativas na identificação dos testes silábicos quando a frequência dos formantes na sentença foi alterada. Por exemplo, quando a  $F_1$  foi alterada de forma a possuir um valor mais baixo do que o padrão, a sílaba do sinal original /bIt/ foi percebida como /bæt/. Alternativamente, quando o  $F_1$  foi alterado para um valor mais alto, a sílaba /bæT/ foi percebida como /bet/.

## 1.2 Exemplos de modelos de percepção da fala

Dentre os principais modelos de percepção de fala pode-se citar o *Simple Target Model*, *Elaborated Simple Target Model* e *Specific Dynamic Model*. Abaixo segue uma breve descrição desses modelos.

### 1.2.1 *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 a vogal está inserida. Assim, na *articulação*, a forma canônica é representada pela configuração estática assumida pelo trato vocal em fala corrente; em contextos de coarticulação e 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 ou três primeiros formantes (F1, F2 e F3). Nessa abordagem, a informação espectral dos dois primeiros formantes constitui a informação acústica suficiente para a *percepção* de vogais.

Esse modelo, no entanto, segundo Kent e Read (1992), 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 normalização de fala for aplicado<sup>1</sup>. 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 mesmo falante produz uma mesma vogal em diferentes contextos com diferentes valores formânticos. Uma possível solução apontada por Kent e Read (1992) é 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.

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. Abaixo segue uma breve descrição desse modelo.

---

<sup>1</sup> Entende-se por normalização de fala o processo que elimina ou corrige as diferenças entre os padrões formânticos produzidos pelos diferentes falantes.

### 1.2.2 *Elaborated Simple Target Model*

O *Elaborated Simple Target Model* visa principalmente dar conta da normalização de falantes. Seu principal objetivo é tentar 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 e as diferenças perceptuais entre as vogais sejam maximizadas. O trabalho de Syrdal e Golpal (1985) é um bom exemplo desses esforços. Nesse modelo, a frequência fundamental e a formântica são transformadas em banda crítica psicofisicamente motivada ou em escala *Bark*. Nessa proposta, as vogais são caracterizadas por 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*.

Alguns pesquisadores (GERSTMAN, 1968; SKINNER, 1977, MILLER, 1989), utilizando o *Elaborated Simple Target Model*, propõem diferentes algoritmos para a transformação dos dados acústicos em dados perceptuais. Dentre esses, pode-se citar o de Miller (1989), que é um modelo de percepção-auditiva, no qual a discriminação do sinal de fala é caracterizada como um espaço auditivo tridimensional definido em termos de níveis formânticos, em uma referência sensorial baseada na frequência média do falante.

Segundo Strange (1989), os modelos *Simple Target Model* e *Elaborated Simple Target Model* diferem em alguns aspectos, mas possuem algumas características em comum. De acordo com a autora, os dados acústicos para uma representação auditiva da vogal são tomados, em ambas as propostas, a partir de uma única secção transversal espectral da sílaba acústica. Segundo a autora, todas essas tentativas pressupõem 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.

Além do *Elaborated Simple Target Model*, dentre os modelos criados que visam a 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 *Elaborated Simple Target Model* poderiam dar conta adequadamente da percepção de vogais. Ela propõe outra abordagem definida como *Dynamic specification* 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.

A seguir apresenta-se uma breve descrição acerca dos pressupostos desse modelo.

### *1.2.3 Dynamic specification*

Esse modelo enfatiza principalmente a percepção da coarticulação 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 coarticulação deve-se a resultados inesperados de pesquisas realizadas com vogais inseridas em contexto silábico de consoantes oclusivas produzidas por diferentes falantes. Estes estudos apontaram que vogais coarticuladas 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 coarticuladas 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 essas estavam isoladas<sup>2</sup>. Os resultados apontaram que a identificação de vogais inseridas em um contexto CVC foi mais precisa do que quando apresentadas isoladamente ao ouvinte.

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 evidenciam que vogais isoladas foram menos precisamente percebidas do que vogais em contexto de coarticulação. Strange (1989) cita pesquisas realizadas por Macchi (1980), Diehl, Kluender e Walsh (1981), Assmann, Nearey e Hogan (1982), Rakerd e Verbrugge (1982), segundo as quais vogais coarticuladas foram melhores percebidas do que vogais isoladas. Com base nesses dados, Strange (1989) propõe que há, no sinal acústico, outras informações, além da

---

<sup>2</sup> Contexto em que as frequências formânticas de  $F_1$  e  $F_2$  são mais distintas.

configuração formântica, que os ouvintes utilizam para a identificação de vogais, assim, a autora volta-se a investigar a natureza da informação dinâmica disponível na sílaba coarticulada.

Strange (1989) e seus seguidores desenvolveram uma técnica em que a informação da duração e a informação dinâmica espectral (trajetória formântica) do núcleo silábico 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. De acordo com o autor, esta técnica permitiu explorar três tipos de informações perceptuais: primeiro, as informações perceptuais provenientes do núcleo silábico, ou seja, a informação da porção central de sílabas CVC em que os formantes se encontram em estado estacionário ou onde as variações nos níveis dessas frequências são menores; segundo, a informação proveniente da duração intrínseca 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.

De acordo com os resultados obtidos, 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. 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. Para Strange (1989), isso ocorre porque as porções do centro silábico não são suficientes nem necessárias para a percepção exata da coarticulação.

## 2 Metodologia

### 2.1 Informantes

Nessa pesquisa, foram investigadas três informantes do sexo feminino, com idade compreendida entre 20 e 30 anos, todas naturais de Vitória da Conquista-BA.

### 2.2 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, em relação de oposição, como discriminado no quadro 1<sup>3</sup>.

| Vogal média alta arredondada | Vogal média baixa arredondada |
|------------------------------|-------------------------------|
| <i>Côrte</i>                 | <i>Córtē</i>                  |
| <i>Chôro</i>                 | <i>Chórō</i>                  |
| <i>Acôrdō</i>                | <i>Acórdō</i>                 |
| <i>Fôrma</i>                 | <i>Fórma</i>                  |
| <i>Sôu</i>                   | <i>Sól</i>                    |

**Quadro 1:** Palavras inseridas nas sentenças dos testes de percepção

Em seguida, cada palavra foi inserida em uma determinada 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 pesquisa e cuja aceitabilidade ou não era

---

<sup>3</sup> As vogais foram acentuadas a fim de fornecer uma melhor visualização da oposição do timbre vocálico das palavras selecionadas.

objeto de julgamento pelos julgadores. Frases incoerentes distratoras foram acrescentadas às frases-alvo. O julgador, ao ler o teste, deveria atribuir graus de aceitabilidade as frases, sendo: 0 para frases 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 respostas desse teste foram tabuladas e as sentenças que obtiveram índice de rejeição maior ou igual a 30% foram descartadas.

As palavras cujas sentenças foram classificadas pelos julgadores como inadequadas foram inseridas na frase veículo “Digo \_\_\_\_\_ baixinho”.

As frases que se encontram no quadro 2 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 Pesquisa e Estudo em Fonética e Fonologia da Uesb (LAPEFF). Antes da gravação, as sentenças foram apresentadas aos sujeitos seguida da orientação de eles deveriam lê-las em tempo natural de fala.

|    |   |
|----|---|
| 1. | Maria disse: eu não quero ficar com essa fôrma de bolo        |
|    | Maria disse: eu não quero ficar com essa fórmula de bolo.     |
|    |   |
| 2. | João disse aos garotos: Côrte, tudo o que as meninas desejam. |
|    | João disse aos garotos: Córte, tudo o que as meninas desejam. |
|    |   |
| 3. | Acórdo feliz, com você em qualquer tempo e lugar.             |
|    | Acôrdo feliz, com você em qualquer tempo e lugar.             |
|    |   |
| 4. | Digo chôro baixinho.  |
|    | Digo chóro baixinho.  |
|    |   |
| 5. | Sól a coisa mais fascinante do universo.                      |
|    | Sôu a coisa mais fascinante do universo.                      |

**Quadro 2** – Frases utilizadas nos testes de percepção

## 2.3 Gravação das Frases e Preparação do Estímulo: a manipulação da $F_0$

A frequência fundamental 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. Foram elaborados, portanto, diferentes sinais sonoros com a  $F_0$  manipulada tanto para valor inferior quanto superior ao original 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 ser obtida a média do referido valor para ser manipulado a 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 ainda necessária de forma a atingir de fato os valores desejados.

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.

A elaboração dos testes de percepção contou com a utilização do programa *Transcriber* desenvolvido por Boudahmane et al. (2000), o qual 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 nas diferentes taxas aqui avaliadas ( $F_0$  acrescida e diminuída em 25%, 50% e 50%). O programa permitiu que as informantes ouvissem os sinais e simultaneamente respondessem aos testes escritos.

## 2.4 As Tarefas dos Testes de Percepção

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

### 2.4.1 Teste de discriminação

Para a aplicação desse teste, as informantes ouviram um dos sinais, original ou manipulado, e escolheram 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 dez, a saber, cinco vogais médias altas arredondadas, cinco vogais médias baixas arredondadas. As informantes, ao ouvir determinado sinal, deveriam marcar no teste escrito aquele que correspondia à sentença percebida.

As 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 350 sinais. As vogais em questão foram divididas em diferentes grupos: *original* – 25 sinais referentes à vogal média alta arredondada, 25 sinais referentes à vogal média baixa; *manipulado* – sinais referentes às vogais médias altas e baixas manipuladas em 25%, 50% e 75% acima do valor original, totalizando 150 sinais, e 25%, 50% e 75% abaixo do valor original, totalizando 150 sinais.

#### 2.4.2 Teste de identificação

No teste de identificação, foi montada uma tarefa em que os informantes ouviram dois sinais de fala, um original e outro manipulado ao mesmo tempo. 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 os dois sinais acústicos referiam-se a uma mesma sentença; 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, 350 sinais agrupados do mesmo modo.

## 2.5 Procedimentos para Aplicação dos Testes de Percepção

As informantes tiveram acesso aos estímulos por meio de um fone conectado a um computador, procedimento que permitiu às informantes terem acesso aos sinais de fala sem interferência dos possíveis sons externos. Os sujeitos tiveram acesso, por meio do software *Transcriber*, 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 às informantes como seria a aplicação dos testes, como deveriam marcar as respostas. Além disso, elas foram informadas sobre a possibilidade de pausa caso se sentissem cansadas ou por qualquer outro motivo. Antes da aplicação dos testes, as informantes foram submetidas a um teste inicial para que pudesse compreender efetivamente todos os procedimentos necessários.

As tarefas foram aplicadas em diferentes etapas. Primeiramente foi realizado o teste de discriminação e, em seguida, o de identificação. Os testes foram organizados de forma que as informantes tivessem acesso aleatório aos três sinais: a) com a  $F_0$  original; b)  $F_0$  com manipulação ascendente; e c)  $F_0$  com manipulação e descendente. No teste de identificação, as informantes tinham acesso a um grupo de vinte e oito sentenças, organizadas em pares mínimos, como exemplo, “Digo *sêde* baixinho / Digo *sêde* baixinho”. A cada teste as informantes faziam uma pausa de cinco minutos e depois retornavam ao teste. A cada três testes, as informantes faziam um intervalo de vinte minutos e, depois, voltavam 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 por um 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, foi aguardado o período de três dias para ser aplicado o teste de identificação. Assim como o teste de identificação, as informantes tiveram acesso ao sinal de fala dentro de um grupo de vinte e oito sentenças, sendo que, a cada teste, as informantes faziam uma pausa de cinco minutos e, depois, voltavam a responder aos testes. A cada três testes, as informantes fizeram uma pausa de vinte minutos, depois voltaram a responder às tarefas dentro do mesmo período de tempo anterior. Devido também a uma questão de tempo das julgadoras, 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.

### 3 Resultados e Discussões

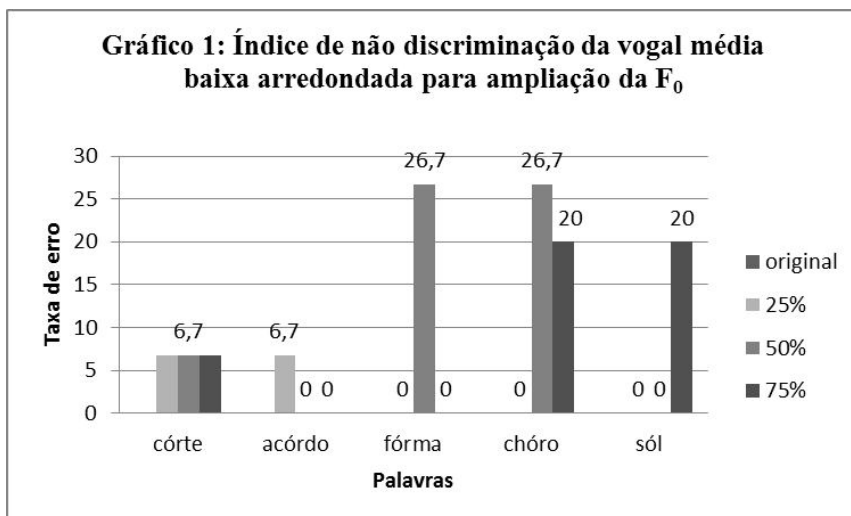
A seguir, observam-se os resultados obtidos nesta pesquisa e algumas discussões. Os resultados foram agrupados da seguinte forma: 1) teste de discriminação e 2) teste identificação.

#### 3.1 Teste 1: Discriminação

Abaixo seguem os gráficos correspondentes aos resultados dos testes de discriminação.

##### 3.1.1 Análise intervocálica dos testes de discriminação com valores de $F_0$ ampliados a 25%, 50% e 75%

Resultados dos testes de discriminação referentes às vogais médias arredondadas para valores ampliados da  $F_0$ .



Os dados apontam que o índice de alteração da inteligibilidade da palavra *córte* foi o mesmo para os três valores manipulados, ou seja, em relação ao original, a ampliação da  $F_0$  suscitou mesmo índice de erro, 6,7%.



Assim, a vogal em questão foi alterada nos três referidos casos de modo semelhante.

A ampliação da  $F_0$  na palavra *acórdo* acarretou interferência apenas para a taxa percentual de 25%, ou seja, para 75% e 50%, os ouvintes conseguiram recuperar a vogal média baixa arredondada em 100% dos casos. A alteração da  $F_0$ , portanto, gerou uma pequena interferência na discriminação da vogal média em questão quando essa foi ampliada a 25% do valor original.

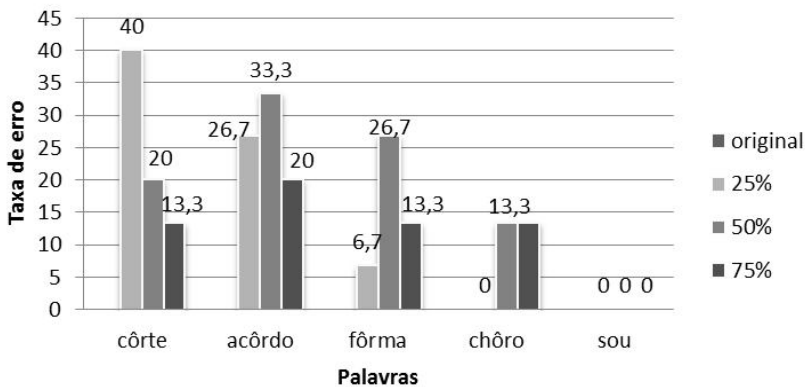
A discriminação da palavra *fórma* sofreu interferência quando os valores da  $F_0$  foram ampliados para a taxa de 50% do valor original, ou seja, nesse contexto, os índices de 25% e 75% não interferiram na discriminação da vogal média arredondada.

Na palavra *chóro*, a ampliação da  $F_0$  gerou prejuízos para a inteligibilidade da palavra. Assim, a taxa referente a 50% gerou índice de erro de 26,7%. Quando a manipulação foi ampliada para 75% do valor do sinal original, a taxa de erro caiu em 6,7%. A ampliação da  $F_0$  em 25% não gerou prejuízo para a discriminação da palavra em questão. Assim, nesse contexto, a ampliação da  $F_0$  acarretou erros na discriminação da vogal média baixa arredondada.

Na palavra *sól*, as taxas de ampliação da  $F_0$  em 25% e 50% não implicaram em prejuízos para a inteligibilidade da palavra, ou seja, os ouvintes conseguiram recuperar o sinal em 100% dos casos; no entanto, quando o índice de manipulação foi ampliado para 75%, os ouvintes não conseguiram recuperar o sinal original em 20% das ocorrências. Assim, nesse contexto, a ampliação da  $F_0$  gerou prejuízos para a discriminação das vogais médias baixas arredondadas.

No gráfico 2, estão dispostos os resultados obtidos no teste de discriminação da vogal média alta arredondada com  $F_0$  ampliada

**Gráfico 2: Índice de não discriminação da vogal média alta arredondada para ampliação da F<sub>0</sub>**



Quando a F<sub>0</sub> foi ampliada para os valores de 25, 50 e 75% do valor do sinal original, a inteligibilidade da palavra *côrte* foi afetada em proporção decrescente. Assim, para o índice de 25%, o ouvinte não conseguiu recuperar o sinal original em 40% dos casos, quando a taxa de manipulação foi ampliada para 50%, esse valor caiu para 20% e para 13,3% quando a taxa de ampliação ascendeu para 75% do sinal original. Nesse contexto, portanto, a ampliação da F<sub>0</sub> gerou prejuízos para a discriminação da vogal média arredondada.

Na palavra *acôrdô*, a ampliação da F<sub>0</sub> para os três índices de manipulação ocasionou prejuízos para a inteligibilidade do sinal. Na taxa de 25%, o ouvinte não recuperou o sinal em 26,7% dos casos, essa taxa de não discriminação aumentou com a ampliação da F<sub>0</sub> para 50% e decresceu com a taxa de ampliação para 75%. Assim, o índice de ampliação da F<sub>0</sub> afetou a percepção da vogal média alta arredondada nesse contexto.

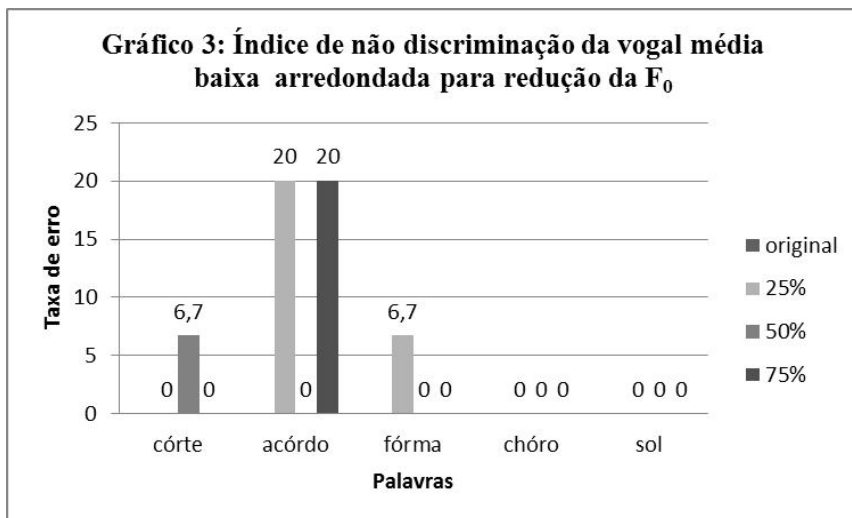
A ampliação da F<sub>0</sub> gerou erros na discriminação da palavra *fôrma* nos três valores investigados. Nesse sentido, o maior índice de erro foi para a ampliação da F<sub>0</sub> em 50%, o menor índice foi para 25% e índice de erro intermediário foi para 75%. Desse modo, para taxa de 25%, o índice percentual de erro foi de 6,7, quando a taxa de manipulação da F<sub>0</sub> foi ampliada para 50%, o índice percentual de erro de discriminação aumentou para 26,7 e caiu quando a taxa de manipulação foi ampliada para 75% do

valor original. Logo, a ampliação da  $F_0$  afetou a discriminação da vogal média alta arredondada nesse contexto de fala.

A ampliação da  $F_0$  não afetou, contudo, a inteligibilidade da palavra *sól*, ou seja, o ouvinte conseguiu recuperar o sinal apesar dos índices de ampliação da  $F_0$ . Nesse contexto, a vogal média alta não sofreu prejuízos de discriminação com a ampliação da  $F_0$ .

### 3.1.2 Análise intervocálica dos testes de discriminação com valores de $F_0$ reduzidos a 25%, 50% e 75%

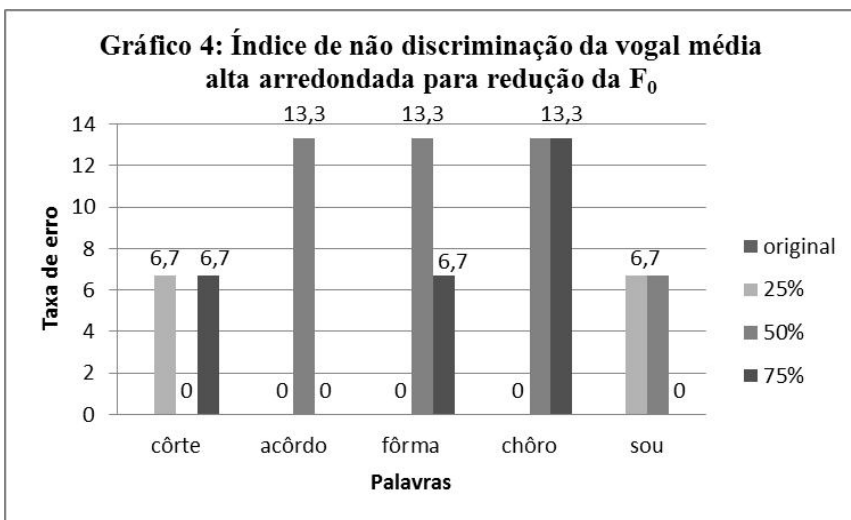
Abaixo segue uma breve discussão referente aos resultados obtidos sobre a interferência da redução da  $F_0$  na inteligibilidade da vogal média arredondada.



É possível observar, a partir do gráfico 3, que a discriminação das palavras *sól* e *chóro* não foi afetada com a redução da  $F_0$ . No caso das palavras *córte*, *acórdo* e *fórma*, a redução da  $F_0$  gerou interferências na inteligibilidade do timbre vocálico. Assim, na palavra *córte*, a redução da  $F_0$  em 50% gerou interferência de 6,7%. No caso da palavra *acórdo*, as taxas percentuais 25 e 50 interferiram a inteligibilidade da vogal em 20% dos casos, e a palavra *fórma*

obteve índice de erro de discriminação de 6,7%. A redução da  $F_0$  em alguns contextos, portanto, interferiu na discriminação das vogais médias baixas arredondas.

Com vistas a avaliar os efeitos da redução  $F_0$  na inteligibilidade da vogal alta não arredondada, os resultados aqui obtidos são dispostos no gráfico 4.

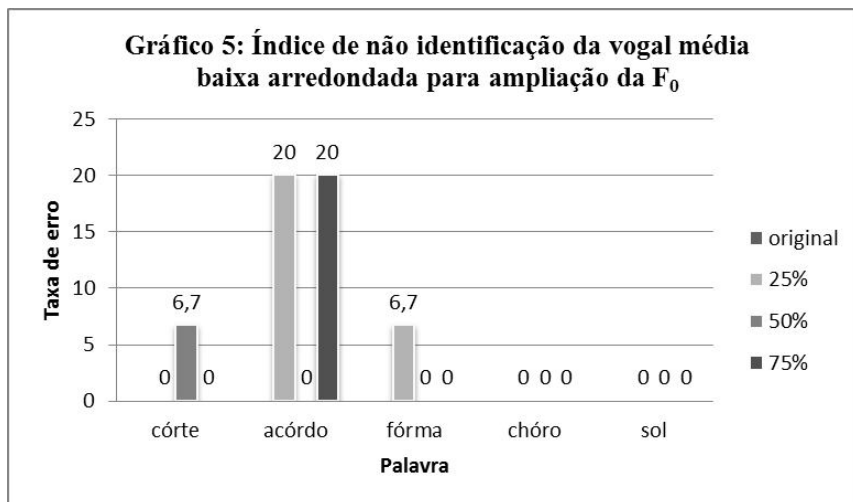


Observa-se que as palavras *acôrdo*, *fôrma*, *chôro* não apresentaram taxa de erro de discriminação para a redução de 25%; já as palavras *sou* e *côrte* obtiverem taxa de erro de discriminação de 6,7% para a taxa de redução de 25%. Para as taxas de 50%, as palavras *acôrdo*, *fôrma*, *chôro* apresentaram mesma taxa de erro de discriminação, 13,3%. Para os índices de 50% de redução da  $F_0$ , a palavra *corte* obteve 0% de erro e a palavra *sou* 6,7% de erro. A redução da  $F_0$  acarretou, portanto, alteração na discriminação da vogal média alta arredonda nos referidos contextos.

## 3.2 Teste 2: Identificação

### 3.2.1 Análise intervocálica dos testes de identificação para valores ampliados da $F_0$

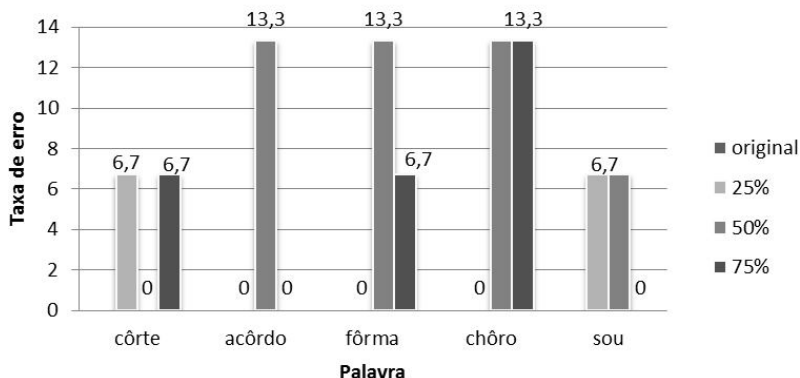
Breve discussão e explanação dos resultados referentes à influência da ampliação da  $F_0$  na identificação da vogal média arredondada são apresentadas nesta sessão.



A ampliação da  $F_0$  em 25% gerou erro de identificação para as palavras *acôrdo* e *fórmula*, com taxa de erro de 20% e 6,7% respectivamente, conforme gráfico 5. Na palavra *côrte* a ampliação da  $F_0$  causou prejuízo para a identificação da vogal em questão no índice de 50%, as palavras *choro* e *sol* não obtiveram erro de identificação com a ampliação da  $F_0$  para os três valores investigados.

Conforme apresentado no gráfico 6, as vogais médias altas também sofre perda de sua identificação quando a  $F_0$  é ampliada.

**Gráfico 6: Índice de não identificação da vogal média alta arredondada para ampliação da  $F_0$**

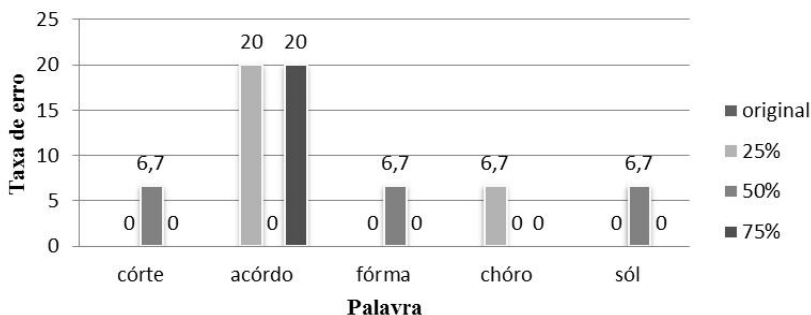


Na palavra *côrte*, a ampliação da  $F_0$  em 25% e 50% gerou erro na identificação da vogal. Na palavra *acôrdo*, a ampliação da  $F_0$  em 50% afetou a identificação da vogal média em questão. A ampliação da  $F_0$  em 50% gerou erro de identificação em 13,3% para a palavra *fôrma*, com a ampliação da  $F_0$  em 75%, o índice de erro caiu para 6,7%. A palavra *chôro*, nos índices percentuais de 50 e 75, obteve erro de identificação. A palavra *sou* obteve taxa de erro de 6,7% para os índices de redução da  $F_0$  em 25% e 50%, quando a taxa foi ampliada para 75%, o ouvinte conseguiu recuperar a vogal média alta em 100% dos caso. Os resultados apontam, portanto, que a vogal média alta não arredondada, nesse contexto, sofreu interferência com a redução da  $F_0$ .

### 3.2.2 Análise intervocálica dos testes de identificação para valores reduzidos da $F_0$

Os gráficos seguintes correspondem aos resultados dos testes de discriminação para redução da  $F_0$  referentes às vogais médias arredondadas.

**Gráfico 7: Índice de não identificação da vogal média baixa arredondada para redução da  $F_0$**



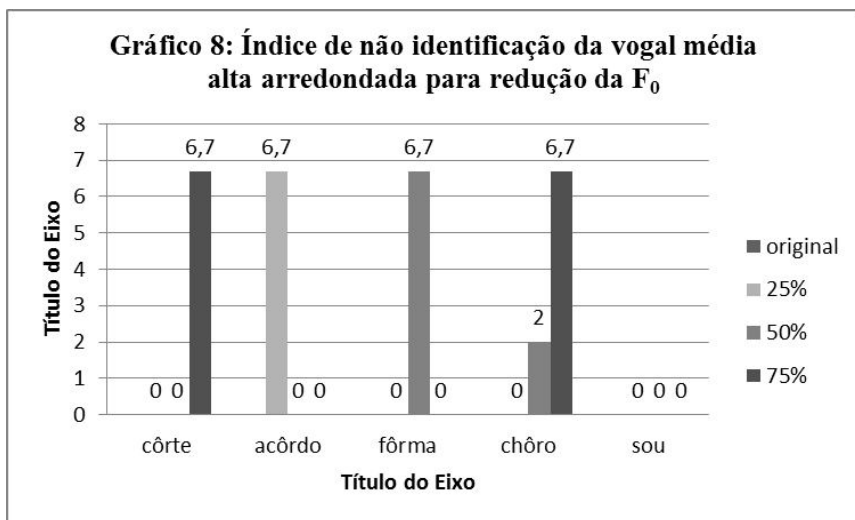
Os dados apontam que quando a  $F_0$  foi reduzida em 50% do valor original, o índice de alteração na identificação da palavra *córte* foi de 6,7%. Já os índices de 25% e 75% de redução, não geraram erros de identificação na palavra *córte*. Assim, a vogal em questão sofreu uma pequena alteração quando a  $F_0$  é reduzida em 50% do valor original.

A ampliação da  $F_0$  na palavra *acórdo* suscitou interferência para as taxas percentuais de 25% e 75%, ou seja, somente para o índice de redução 50% foi que os ouvintes conseguiram recuperar a vogal média baixa arredondada em 100% dos casos. A  $F_0$  gerou interferência para a identificação da vogal média baixa arredondada quando essa foi reduzida em 25% e 75% do valor original.

A discriminação da palavra *fórma* sofreu interferência de 6,7% para o índice de redução de 50%. A redução da  $F_0$  em 25% e 75% não interferiu na identificação da vogal média arredondada. Assim, a redução da  $F_0$  gerou interferência na inteligibilidade do sinal apenas quando essa foi reduzida em 25% do valor original.

Na palavra *chóro*, a redução da  $F_0$  em 25% gerou índice percentual de erro de 6,7. Quando a manipulação foi reduzida para 25% e 75% do valor do sinal original, o ouvinte conseguiu recuperar em 100% a vogal média em questão. Assim, nesse contexto, a redução da  $F_0$  acarretou índices de erro na discriminação da vogal média baixa arredondada.

Na palavra *sól*, as taxas de redução da  $F_0$  em 25% não implicaram em prejuízos para a inteligibilidade da palavra. Quando o índice de manipulação foi reduzido em 50% em relação ao valor original, o ouvinte não conseguiu recuperar o sinal original em 6,7% das ocorrências. Assim, nesse contexto, a redução da  $F_0$  acarretou alteração na discriminação das vogais médias baixas arredondadas.



Com base no gráfico 8, tem-se que na palavra *côrte*, a ampliação da  $F_0$  em 25% e 50% não gerou erro na identificação da vogal, já o índice de 75% suscitou taxa de erro de 6,7%. Na palavra *acôrdo*, a ampliação da  $F_0$  em 25% afetou a identificação da vogal média em questão em 6,7% dos casos investigados, as taxas de manipulação 50% e 75%, nesse contexto, não suscitaram prejuízos para a inteligibilidade do sinal. A ampliação da  $F_0$  em 50% gerou erro de identificação de 6,7% para a palavra *fôrma*, já os índices de 25% e 75%, nesse contexto, não interferiram na inteligibilidade do sinal. A palavra *chôro* nos índices percentuais de 50 e 75 obteve erro de identificação de 2% e 6,7% respectivamente. Na palavra *sou*, os índices de manipulação investigados não afetaram a identificação da vogal média analisada.



## Discussões 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 alta como sendo uma vogal média alta. Houve ainda casos em que o informante identificou a vogal /ô/ como /u/. Assim, tanto a ampliação quanto à redução da  $F_0$  geraram erros na discriminação e identificação da vogal média arredondada.

Os dados sinalizam, portanto, que a frequência fundamental parece ser um importante recurso utilizado pelo ouvinte para a discriminação da vogal média arredondada, 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. Vogais em que a configuração formântica dos dois formantes mais baixos ( $F1$  e  $F2$ ) possuem valores aproximados, como é caso das vogais aqui investigadas, médias altas arredondadas e médias baixas arredondadas, a  $F_0$  parece contribuir juntamente com a frequência formântica para a discriminação do timbre vocálico.

Os resultados assinalam que, além da configuração formântica, outros parâmetros espectrais contribuem significativamente para a compreensão de como se dá a percepção das vogais médias arredondadas, ou seja, o ouvinte utiliza 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*, Canadá, 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*, Canadá, v. 127, n. 3, p. 219, 2011.

BERGEM, D. R. van; POLS, L. C. W.; BEINUM, K. van Perceptual normalization of the vowels of a man and a child in various contexts. *Speech Communication*, v. 7, p. 1-20, 1988.

BOERSMA, P.; WEENINK, D. *Praat*: doing phonetics by computer. Version 5.1.01, Disponível em: <<http://www.praat.org>>. Acesso em: 6 abr. 2012.

BOUDAHMANE, K. et al. *Transcriber*: a free tool for segmenting, labeling and transcribing speech. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 1., Filadélfia, 2000. Filadélfia: Oasis, 2000. Disponível em: <<http://trans.sourceforge.net>>. Acesso em: 17 maio 2012.

CLAES, T. et al. 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, 1988.

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. In: 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.

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

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.

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

JOOS, M. Acoustic phonetics Language Monograph. In: SUEN, C. Y.; MICHAEL, P. B. Discrimination of vowels sound of very short duration. *Attention, Perception & Psychophysics*, 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.; BROADBENT, D. E. Information conveyed by vowels. *Journal of the Acoustical Society of America*, Cambridge, 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.

MELTZER, D.; LEHISTE, I. Vowel and speaker identification in natural synthetic speech. *Journal of the Acoustical Society of America*, Columbus, v. 51, n. 1A, p. 131-131, 1972.

MILLER, J. D. Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, Missouri, v. 85, p. 2114-2134, 1989.

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.

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

STRANGE, W. Dynamic specification of coarticulated vowels spoken in sentences context. *Journal of the Acoustical Society of America*, Florida, v. 85, p. 2135-2153, 1983.

STRANGE, W. Evolving theories of vowel perception. *Journal of the Acoustical Society of America*, Florida, v. 85, n. 5, p. 2081-2087, 1989.

STRANGE, W. Cross-language studies of speech perception. In: STRANGE, W. *Speech perception and linguistic experience: issues in cross-language research*, Baltimore: York Press, 1995.

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, 1985.

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.