



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ESTUDOS DA LINGUAGEM

Renata Regina Passetti

**Desviando da própria fala: implicações para a
verificação de locutor em situações de disfarces
vocais**

CAMPINAS

2012

Renata Regina Passetti

**Desviando da própria fala: implicações para a
verificação de locutor em situações de disfarces
vocais**

Monografia apresentada ao Instituto de Estudos da
Linguagem da Universidade Estadual de Campinas
como requisito parcial para a obtenção do título de
“Bacharel em Linguística”.

Orientador: Prof. Dr. Plínio Almeida Barbosa

CAMPINAS

2012

Dedico esta monografia aos meus pais, Amaury e Elizabete, que sempre apoiaram minhas escolhas e estiveram ao meu lado em todos os momentos.

AGRADECIMENTOS

A FAPESP, por financiar essa pesquisa;

ao professor, PLÍNIO A. BARBOSA, pelos ensinamentos que muito acrescentaram à minha formação como linguista, orientação, oportunidades e por ter me aceitado em sua equipe;

a minha família, meus pais AMAURY e ELIZABETE e minha irmã ANA BEATRIZ, pelo carinho, pelo empenho em minha educação, pelos ensinamentos e por estarem ao meu lado em todos os momentos e decisões;

aos meus avós, CONSTANTINA e APARECIDO, ADILSE e DOMINGOS, pelas orações, pelos ensinamentos, pelas refeições maravilhosas e apoio financeiro;

ao GUILHERME, pelo carinho, por sempre estar presente e me dar forças nos momentos difíceis;

a minhas amigas, AMANDA, ANA LUIZA, ANNELIESE, GEORGIA e JOYCE, pelas risadas, companheirismo e amizade;

a RAYSSA pela gentileza em revisar meus trabalhos e pelos conselhos;

aos meus colegas ALINE MACHADO, ANA CAROLINA CONSTANTINI, CRISTIANE SILVA e WELLINGTON DA SILVA, pelo gesto solícito de me ajudarem com a minha pesquisa em um momento difícil;

aos PROFESSORES do curso de linguística pelo conhecimento compartilhado em suas aulas;

aos colegas do curso de linguística 2009, pela convivência e troca de experiência;

aos meus colegas do GRUPO DE ESTUDOS DA PROSÓDIA DA FALA, pelo convívio e troca de experiência;

a todos que gentilmente participaram das gravações utilizadas nesta pesquisa.

“De modo suave, você pode sacudir o mundo.”

Mahatma Gandhi

RESUMO

Esta monografia avalia as pistas acústicas de locutores utilizadas em disfarces vocais que envolvem a utilização de um objeto na boca, entre os dentes frontais, e a imitação de uma amostra de fala familiar aos sujeitos, evocando memória de longo prazo. O corpus se constitui de 9 gravações de 3 sujeitos do sexo masculino, com idade de 21 anos. Cada locutor leu um texto com cerca de 130 palavras que consistia na transcrição de um discurso do apresentador de televisão, Sílvio Santos, evocando três diferentes tipos de elocução: (1) com o estilo de fala habitual, com voz e estilo de fala naturais; (2) imitando o apresentador de televisão, Sílvio Santos, cuja identidade foi apresentada aos locutores somente nesse momento; e (3) utilizando o disfarce vocal “lápiz na boca” pela utilização de um lápis posicionado firmemente entre os dentes frontais. O principal objetivo foi determinar quais características acústicas e prosódicas eram modificadas e quais permaneciam na fala de indivíduos diante da utilização destes disfarces. Foram estudadas técnicas de análise acústica que envolviam análise de formantes, frequência fundamental (f_0), ênfase espectral e duração em vogais orais do português brasileiro (PB) e também das pausas silenciosas e duração entre picos de f_0 presentes no discurso. A validação das análises acústicas foi conduzida por análises comparativas intra-locutor dos disfarces vocais em relação à voz habitual, seguido da aplicação de um teste T de variáveis independentes, com um nível de significância de 5%, com a finalidade de testar a hipótese nula de que as amostras de fala comparadas (fala disfarçada e fala habitual) não apresentavam diferenças significativas para cada um dos parâmetros acústicos testados. Os resultados mostram que os locutores conseguiram desviar da própria fala, por meio de alterações na entonação, no esforço vocal, expresso pela ênfase espectral, empregado durante a realização dos disfarces e também por meio do alongamento vocálico e de alterações no tempo global do discurso. Outra modificação importante diz respeito às frequências de formantes e a reconfiguração do espaço vocálico. Não foram atestadas modificações nos valores obtidos para a duração inter-picos de f_0 para nenhum dos locutores testados.

Palavras-chave: fonética forense, verificação de locutor, disfarce vocal, modificações acústicas.

ABSTRACT

This monograph aims at analyzing the speaker's acoustic strategies in two vocal disguise situations: (1) with an object between the front teeth, and (2) the imitation of a familiar sample from a well-known person (impersonation), evoking long-term memory. The corpus used consisted of 9 recordings from three men, all of them at the mean age of 21-year-old. Each speaker read a 130-word text, which consisted in a transcription of a Brazilian TV host speech, Sílvia Santos. The speakers were recorded three different times: (1) with their own natural voice and speaking style, (2) trying to imitate Sílvia Santos and (3) with a vocal disguise "object in the mouth" by the use of a pencil firmly placed between the frontal teeth. The focus was to determinate which were the acoustic and prosodic cues which were modified and which cues remained undifferentiated in the speaker's speech when the speakers deviated from their own voice. The acoustic techniques included the evaluation of the frequency of the first four formants, fundamental frequency (f_0), duration and spectral emphasis in Brazilian Portuguese oral vowels. The distributions of silent pauses throughout the utterances and the inter f_0 peak duration were also analyzed. A comparative intra-speaker analysis between the natural voice speech sample and the disguised voices speech samples (impersonation and "object in the mouth") was followed by a T-test of independent variables with a significance level of 5% to evaluate the null hypothesis that there were no significant differences between the natural voice and the disguised ones. The results indicate that there was a reconfiguration in the speakers' vowel space via modification of vowels' formants. Changes in pitch were also attested, as well as in vocal effort (via spectral emphasis), in vowel duration and in speech rate, parameters involved in helping the speakers to achieve a good performance with the vocal disguises.

Keywords: Forensic Phonetics, Speaker Verification, vocal disguise, acoustic modification.

SUMÁRIO

| | |
|--|----|
| 1. INTRODUÇÃO | 1 |
| 2. VERIFICAÇÃO DE LOCUTOR E DISFARCES VOCAIS | 1 |
| 2.1 <i>Verificação de locutor e análises acústicas em estudos envolvendo imitação da fala</i> | 2 |
| 2.2 <i>Verificação de locutor e análises acústicas em estudos envolvendo outros disfarces vocais</i> | 4 |
| 3. ESTUDO DE TÉCNICAS ACÚSTICAS PARA VERIFICAÇÃO DE LOCUTOR EM SITUAÇÕES DE DISFARCES VOCAIS | 6 |
| 3.1 OBJETIVOS | 6 |
| 3.2 METODOLOGIA | 6 |
| 3.2.1 <i>Corpus</i> | 7 |
| 3.2.2 <i>Condução do experimento</i> | 8 |
| 3.3 PARÂMETROS ACÚSTICOS | 11 |
| 3.4 ANÁLISES ESTATÍSTICAS | 13 |
| 4. RESULTADOS | 13 |
| 4.1 <i>Sujeito GB</i> | 13 |
| 4.2 <i>Sujeito RH</i> | 16 |
| 4.3 <i>Sujeito KM</i> | 18 |
| 5. ANÁLISE COMPARATIVA INTRA-LOCUTOR PARA OS PARÂMETROS TESTADOS | 20 |
| 5.1 <i>Frequência fundamental média (f0 média)</i> | 20 |
| 5.2 <i>Duração das vogais orais do PB</i> | 20 |
| 5.3 <i>Duração das pausas silenciosas</i> | 21 |
| 5.4 <i>Ênfase espectral</i> | 21 |
| 5.5 <i>Formantes</i> | 22 |
| 5.6 <i>Duração inter-picos de f0</i> | 27 |
| 6. CONSIDERAÇÕES FINAIS | 27 |
| 7. REFERÊNCIAS | 28 |
| 8. APÊNDICE | 29 |

1. INTRODUÇÃO

A ocorrência de casos que envolvem amostras de fala de sujeitos merece destaque no âmbito da ciência forense. Estudos relacionados à verificação de locutor encontram em um campo linguístico, o da fonética, a possibilidade de avaliar os padrões de parâmetros acústicos (a análise dos formantes e frequência fundamental, entre outros), como ferramentas para o reconhecimento de locutor. Esses estudos auxiliam, também, no desenvolvimento de tecnologias que podem, semi ou automaticamente, fornecer pistas para o reconhecimento de um indivíduo a partir de sua fala.

Hollien (2002) explica que, na área de reconhecimento de fala, o termo “Verificação de Locutor” (*Speaker Verification*) busca a identificação de um locutor em ambientes acústicos favoráveis, com o uso de equipamentos de alta precisão e amostras de fala que estão sob a supervisão do operador. Nesses casos, há colaboração do falante, que permite ser reconhecido. É utilizada, geralmente, em sistemas que utilizam chaves de segurança como forma de reconhecimento de indivíduo. Difere-se da “Identificação de Locutor” (*Speaker Identification*) pelas condições em que as amostras de fala são colhidas, pois geralmente nesta, a análise de fala ocorre em ambientes com ruídos e distorções e não há cooperação por parte do locutor, que pode ser desconhecido. É comumente relacionada à área forense, pois, em situações em que não se há pistas, pode haver um criminoso totalmente desconhecido em meio a uma comunidade inteira de suspeitos.

Na prática, a Verificação de Locutor associa-se, também, à área forense, pois o trabalho de reconhecimento de um sujeito envolve, no fim, um conjunto finito de suspeitos. E ainda existe, em alguns casos, a possibilidade do sujeito gravado como referência ser cooperativo.

Este estudo consiste na verificação de locutor em contextos que envolvam disfarce da fala com o uso de instrumento e imitação da fala de um sujeito com acesso à memória de longo prazo, por envolver a imitação de uma figura pública cuja fala se encontra frequentemente na mídia.

2. VERIFICAÇÃO DE LOCUTOR E DISFARCES VOCAIS

Neste capítulo, serão apresentados estudos de verificação de locutor que envolvem a utilização de disfarces vocais. Discussões acerca desse tema, os principais parâmetros acústicos e prosódicos utilizados, resultados e conclusões presentes na literatura serão

expostas com o objetivo de compreender e, posteriormente, aprimorar os estudos na área de Fonética Forense.

2.1 Verificação de locutor e análises acústicas em estudos envolvendo imitação da fala

A verificação de um locutor pode ser afetada por aspectos relacionados à língua, ao dialeto, ao sotaque e a disfarces de fala. Eriksson (2010) afirma que disfarces de voz podem causar sérios problemas à verificação de um locutor, especialmente se forem utilizados métodos eletrônicos de manipulação de voz. Entretanto, a maioria dos casos envolvendo disfarces vocais utiliza métodos de natureza simples.

Alguns tipos de disfarces de voz provocam alterações na frequência fundamental e nos formantes. Eriksson (2011), no seu artigo *Aural/acoustic vs. automatic methods in forensic phonetic case work*, afirma que tipos comuns de disfarce de voz são aqueles que consistem no aumento ou diminuição da frequência fundamental. Já um estudo de Endres, Bambach & Flösser (1971), em que se procurava saber se a configuração de formantes poderia ser alterada por meio de um disfarce de voz, e ainda, se havia a possibilidade de imitar a configuração dos formantes de outra pessoa, concluiu que existe a possibilidade de ocorrerem mudanças consideráveis na configuração dos formantes de vogais em disfarces de voz. No entanto, o grau dessas mudanças varia entre indivíduos. Quanto às imitações, as análises mostraram que os imitadores conseguiram variar a frequência fundamental e configuração dos formantes de suas falas, mas não foram capazes de adaptar esses parâmetros de maneira similar aos utilizados pelos indivíduos imitados.

Eriksson e Wretling (1997) investigaram um caso de imitação, no qual um imitador profissional tinha como objetivo imitar três figuras públicas conhecidas na Suécia. O objetivo da pesquisa era avaliar em que medida as imitações combinavam com determinados parâmetros acústicos das gravações originais e o quanto o artista modificava sua voz habitual para se aproximar das vozes-alvo. Os autores analisaram a organização temporal da fala, a frequência fundamental e os quatro primeiros formantes para todos os alofones vocálicos do sueco presentes na gravação.

Os resultados mostraram que o imitador alcançou valores próximos da taxa de elocução global das vozes dos locutores-alvo, porém, ao nível segmental, a organização temporal da fala mostrou pouca ou nenhuma mudança na direção dos valores estabelecidos para as vozes-alvo. A frequência fundamental média e as frequências dos formantes obtiveram valores muito próximos aos dos locutores-alvo. Para duas das três imitações, houve

mudanças no espaço vocálico do imitador, situando-se entre a configuração vocálica do locutor-alvo e de sua própria voz. Em relação às vogais analisadas individualmente, em algumas situações, as frequências dos formantes das vogais imitadas alcançavam valores próximos às da voz do artista.

Zetterholm *et al.* (1997, 2001, 2004, 2006) desenvolveram várias pesquisas no campo da imitação com o objetivo de verificar quais aspectos da voz habitual de imitadores e características do comportamento de fala eram manipuladas durante tal ato. Em geral, os resultados de seus estudos mostram que, mesmo com diferentes sujeitos imitando diferentes locutores-alvo, as imitações apresentaram modificações na qualidade vocal, entonação e prosódia, a fim de aproximar-se auditivamente das vozes-alvo. Os imitadores modificavam suas vozes de acordo com o estilo de fala característico de cada locutor-alvo, atentando-se ao ritmo da fala, às pausas, acentuação das palavras, dialeto, etc. No que diz respeito às análises acústicas conduzidas, modificações em relação à duração das pausas, média da frequência fundamental e ajustes na posição dos formantes aos do locutor-alvo, mostraram flexibilidade da voz habitual dos imitadores para um bom desempenho durante a imitação.

A prosódia é outro importante parâmetro a ser considerado em estudos de verificação de locutor. Cole e Shattuck-Hufnagel (2011) argumentam que diferenças prosódicas interlocutor são especialmente salientes na fala espontânea, e contribuem para distinguir o estilo de fala entre indivíduos. Um de seus estudos pretendia investigar quais elementos prosódicos eram imitados por ouvintes na fala espontânea. Para tanto, um experimento de imitação testou a hipótese de que, na reprodução de um enunciado espontâneo, os locutores imitariam com mais precisão a prosódia do estímulo em sua estrutura fonológica do que suas pistas fonéticas. Os resultados mostraram que os locutores raramente distorciam a presença de um acento de altura (*pitch accent*) ou de uma fronteira de sintagma entoacional¹ (*intonational phrase boundary*), porém mudanças na natureza das pistas fonéticas eram comuns, assim como na duração da pausa ou na ocorrência de períodos glotais irregulares associados com fronteiras e acentos de altura no inglês americano. Esses resultados indicam que a estrutura fonológica é codificada separadamente das pistas fonéticas que a sinalizam, e que essas pistas variam entre locutores.

Em estudos de verificação de locutor que envolvem a análise acústica de imitações, Eriksson (2011) discute que, para ciência forense, este tipo de disfarce interessa no grau de

¹ Alguns autores brasileiros, como Leda Bisol e Carmem Metzenauer, utilizam o termo “frase entoacional” para se referir à unidade menor (*phrase*) que a sentença (*sentence*).

precisão utilizado pelo imitador para reproduzir a fala de um indivíduo em específico. O critério de avaliação da imitação é de dois tipos: em que medida o imitador consegue enganar os ouvintes que estão familiarizados com a voz do locutor-alvo e o quão próximo o imitador consegue reproduzir determinadas características acústicas da voz-alvo. Estas, quando consideradas na análise de imitação, geralmente são frequência fundamental, configuração dos formantes, taxa de elocução, e características prosódicas, como já discutidas anteriormente.

2.2 Verificação de locutor e análises acústicas em estudos envolvendo outros disfarces vocais

Figueiredo e Britto (1996) estudaram os efeitos acústicos na fala de sujeitos que utilizavam como disfarce vocal um lápis posicionado firmemente entre os dentes frontais, paralelo à língua. Os autores comentam que este tipo de disfarce é encontrado no Brasil em crimes relacionados a sequestros, que envolvem ligações telefônicas por parte do criminoso, em que este suspeita que esteja sendo gravado.

Sabe-se que a utilização de um instrumento na boca, como um lápis, tende a dificultar estratégias de articulação por parte do falante. O lápis limita a movimentação dos lábios e da língua, além de bloquear a mandíbula. Alterações no comprimento entre as cavidades anterior e posterior também são causadas pela restrição de movimento da laringe.

O estudo considerou amostras de fala de três sujeitos masculinos, pronunciadas no estilo habitual e com restrição do movimento da mandíbula pelo uso de um lápis. O parâmetro considerado foi a análise dos três primeiros formantes das vogais orais do português brasileiro. Os resultados mostraram que um dos efeitos da restrição da mandíbula é a alteração do timbre vocálico, há uma redução do espaço vocálico pelo abaixamento das vogais altas e elevação das vogais baixas, o que provocou uma modificação na disposição relativa da configuração dos formantes. Os autores comentam que a redução de vogais altas posteriores é característica de dialetos do sul do país, e um disfarce que possua esse efeito poderia conduzir uma investigação erroneamente. Apesar de não terem sido consideradas nesse estudo, as consoantes também são afetadas pela utilização de um lápis na boca; [s] transforma-se em [ʃ], como em sotaques do sudeste brasileiro, especialmente na fala de indivíduos do Rio de Janeiro. Esta informação foi confirmada por um criminoso que utilizava esse tipo de disfarce para soar como um carioca, segundo os autores.

Em um recente estudo de Perrot & Chollet (2012), os autores propuseram a análise de dois tipos diferentes de disfarces: o primeiro deles envolvia a modificação da voz por meios não eletrônicos, mas intencionais; enquanto que o segundo envolvia modificações da fala através de meios eletrônicos e intencionais. Foram estudados quatro tipos de disfarces: voz extremamente aguda (*high pitched voice*), voz extremamente grave (*low pitched voice*), mão sobre a boca (*hand over the mouth*) e narinas fortemente apertadas pelos dedos (*pinched nostrils*). A análise tinha por base modelos acústicos que foram usados para medir mudanças específicas na fala e também, modelos automáticos utilizados para detectar disfarces vocais. Os parâmetros estudados na primeira parte da análise foram extraídos pelo programa Matlab e por scripts do PRAAT (Boersma, P., Weenink, D., 2009) e podiam ser divididos em duas categorias: parâmetros relacionados à qualidade vocal e parâmetros relacionados ao domínio espectral. São eles: taxa de elocução, HNR (*Harmonics to Noise Ratio*), *jitter*, *shimmer*, frases desvozeadas (*unvoiced frames*), frequência fundamental e formantes (F1 e F2). A análise descritiva dos parâmetros acústicos permitiu aos autores avaliarem o impacto de disfarces vocais em descritores acústicos ou instrumental-auditivos e mostrou-se importante devido à possibilidade de compreendermos, com base na comparação inter-disfarces, as modificações envolvidas durante a execução destes.

A segunda parte da análise limitou-se aos métodos automáticos de detecção de disfarces vocais. Os autores consideraram que a melhor técnica encontrada para detecção de disfarces foi o *Support Vector Machine* (SVM)², com um nível de desempenho de 0,79 para uma área sob a curva (AUC).

Técnicas de conversão vocal também foram propostas e aplicadas em dois cenários forenses: imitação de um político em uma gravação da Internet e aplicação de reversibilidade ao disfarce de voz. Diferentes testes foram conduzidos com o intuito de avaliar a relevância dos resultados obtidos. Os autores consideraram que a melhor conversão foi obtida pelo conversor de voz GMM-ALISP.

² De acordo com os autores, a técnica SVM tornou-se uma importante ferramenta para classificação discriminante. SVMs são eficazes classificadores discriminantes, capazes de maximizar a margem de erro. Devido ao uso da função kernel, os SVMs podem fazer decisões em contextos que não podem ser separados linearmente. A escolha de uma função kernel adequada permite que os SVMs simulem outros classificadores que são baseados em discriminação linear. O algoritmo cria um hiperplano marginal máximo para discriminar duas classes entre si através de separação linear em um espaço de maior dimensão. O SVM não-linear mapeia o espaço n-dimensional de entrada original em um espaço de atributos de maior dimensão, no qual um classificador linear é construído. (Perrot & Chollet, 2012: 490)

Os autores justificaram a necessidade de desenvolvimento em estudos na área de disfarce vocal devido aos riscos que os falsos positivos por análises de disfarces vocais incorretas representam à prática de verificação de locutor em situações reais.

O exame das estratégias perceptuais e acústicas de locutores empregadas em disfarces de voz proporciona importantes contribuições aos estudos na área forense. Esta pesquisa busca avaliar tais estratégias utilizadas em disfarces envolvendo modificações externas dos padrões articulatórios com o intuito de ocultar características fonético-acústicas do locutor e a imitação de uma amostra de fala familiar evocando memória de longo prazo. Pretende-se determinar o grau de afastamento de si mesmo, ou seja, o desvio e manipulação da identidade fonético-fonológica do locutor para realização dos disfarces.

3. ESTUDO DE TÉCNICAS ACÚSTICAS PARA VERIFICAÇÃO DE LOCUTOR EM SITUAÇÕES DE DISFARCES VOCAIS

No presente capítulo será apresentada a pesquisa, conduzida como Iniciação Científica. Foram estudadas técnicas de análise acústica que envolviam a utilização de procedimentos semiautomáticos com o intuito de avaliar as modificações de características acústicas e prosódicas em sujeitos diante da utilização de disfarces de voz selecionados em comparação à voz habitual.

3.1 OBJETIVOS

O principal objetivo dessa pesquisa era avaliar as modificações prosódicas e acústicas na fala de indivíduos através da utilização de dois tipos diferentes de disfarces vocais – imitação e uso de um objeto na boca e, dessa forma, determinar quais características acústicas e prosódicas eram modificadas e quais permaneciam na fala de indivíduos diante da utilização destes disfarces de voz.

3.2 METODOLOGIA

Sabe-se que os disfarces de voz provocam alterações na fala de um indivíduo e conseqüentemente no reconhecimento de locutor, como atestado por Eriksson (2010, 2011), Künzel (2000), Markham (1999) e Sjöström *et al* (2006). Neste trabalho, foram estudadas técnicas de análise acústica para as vogais orais do português brasileiro (PB), como os

formantes, frequência fundamental (f_0), ênfase espectral (cf. Eriksson, 2001)³ e duração, também das pausas silenciosas e duração entre picos de f_0 presentes no discurso. Aspectos relacionados à constituição do corpus e descrição dos procedimentos conduzidos às gravações serão explicitados neste item.

3.2.1 *Corpus*

O corpus utilizado neste estudo foi constituído de 9 gravações de 3 sujeitos do sexo masculino, todos possuíam 21 anos de idade, eram estudantes de graduação da Universidade Estadual de Campinas, exibiam um sotaque do sudeste brasileiro (interior do estado de São Paulo) e não apresentavam problemas auditivos ou fonoarticulatórios. Foi pedido aos locutores que lessem um texto de aproximadamente 130 palavras, que consistia na transcrição da fala do apresentador de televisão Sílvio Santos⁴, em três momentos diferentes, evocando diferentes tipos de elocução: na primeira vez de forma natural, com sua voz habitual; na segunda vez imitando o apresentador de televisão, cuja identidade foi revelada aos participantes apenas neste momento, e por fim, na terceira vez, mordendo um instrumento (lápiz) com os dentes frontais.

As gravações foram feitas utilizando um microfone Shure, modelo BETA 58A, e uma placa MBox 2, ligada ao programa computacional Pro Tools. Os equipamentos utilizados são de responsabilidade do Grupo de Estudos de Prosódia da Fala. O programa PRAAT⁵ (versão 5.3.16) foi utilizado para a segmentação das vogais orais do PB e também para a análise dos formantes das vogais e das pausas silenciosas. O script para o PRAAT “*RatesExtractor*” (Barbosa, 2011) foi utilizado para o cálculo da frequência fundamental média, duração entre picos de f_0 e duração das vogais orais do PB, enquanto o script “*SpectralEmphasis*” (Arantes, 2011) foi utilizado para o cálculo da ênfase espectral. Ambas as análises automáticas foram seguidas por correção manual. As gravações farão parte do banco de dados do Grupo de Pesquisas de Estudos sobre a Fala da PUC-SP, sob responsabilidade das professoras doutoras Sandra Madureira e Zuleica Antonia de Camargo.

³ Eriksson *et al* (2001) define “ênfase espectral” como a diferença de energia, em decibéis, entre a energia total da vogal e sua energia até 1,5 vezes a frequência fundamental máxima do locutor no trecho lido.

⁴ Veja em APÊNDICE.

⁵ BOERSMA, P. & WEENINK (2012). Praat: doing phonetics by computer (Versão 5.3.16) [Programa computacional]. Disponível em: <<http://www.fon.hum.uva.nl/praat/>>

3.2.2 *Condução do experimento*

As gravações foram editadas no programa PRAAT e as vogais orais do português brasileiro foram segmentadas e etiquetadas manualmente. Posteriormente à segmentação das vogais orais deu-se início ao cálculo dos parâmetros acústicos. Cada gravação continha, aproximadamente, 150 vogais orais e a duração de, aproximadamente, 1 minuto. Calculou-se manualmente para todas as vogais selecionadas presentes em cada gravação, os quatro primeiro formantes, extraídos por meio do algoritmo de detecção dos formantes via LPC do PRAAT (com número de formantes igual a 5 e frequência máxima de 5000Hz) e a duração das pausas silenciosas. Foram utilizados os scripts já referidos para o cálculo da frequência fundamental, duração inter-picos de f_0 (com grau de suavização da curva de f_0 igual a 5Hz), ênfase espectral e duração das vogais orais do PB. Os valores obtidos para esses parâmetros foram compilados no programa Excel, onde foram conduzidos os testes T de variáveis independentes ($\alpha = 0,05$).

A seguir, as figuras 1, 2 e 3 mostram exemplos da segmentação das vogais orais para a frase “*Não, eu vô jogar aqui dentro*” do sujeito RH, retirada do texto transcrito, pronunciadas nos estilos de elocução: voz habitual, imitação do Sílvio Santos e uso do objeto (lápiz) na boca; respectivamente. Para a segmentação das vogais orais seguiram-se os critérios habituais de aparecimento de F2 para o início e desaparecimento do mesmo formante para o fim da vogal. A transcrição foi feita manualmente em caracteres ASCII.

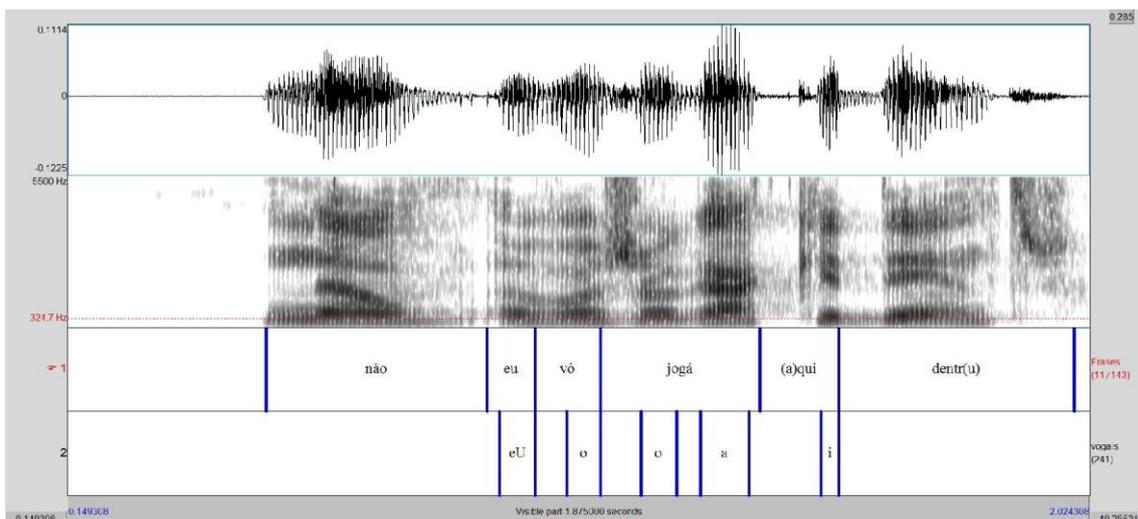


Figura 1: Ilustração da segmentação das vogais orais na frase “*Não, eu vô jogar aqui dentro*”, para o sujeito RH, no estilo de elocução “voz habitual”. Forma de onda (superior), espectrograma (centro) e segmentação das frases e vogais orais do PB (inferior).

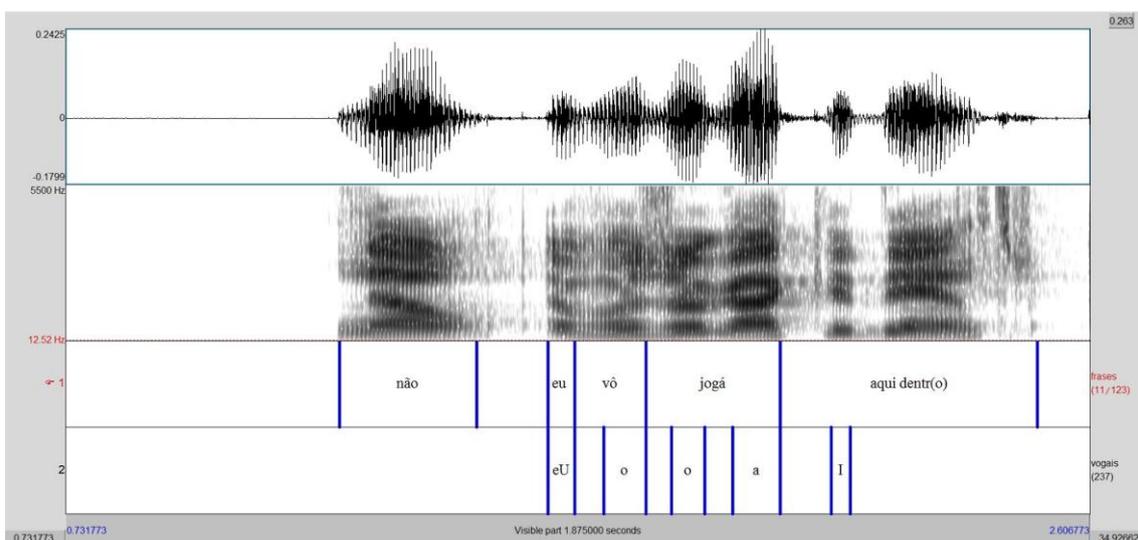


Figura 2: Ilustração da segmentação das vogais orais na frase “*Não, eu vô jogar aqui dentro*”, para o sujeito RH, no estilo de elocução “imitação de Sílvio Santos”. Forma de onda (superior), espectrograma (centro) e segmentação das frases e vogais orais do PB (inferior).

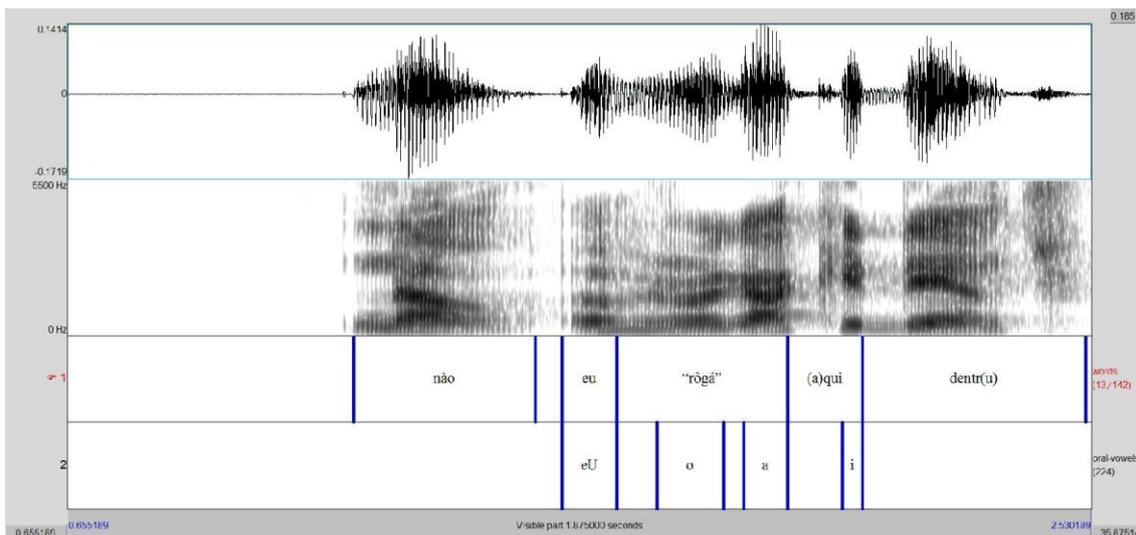


Figura 3: Ilustração da segmentação das vogais orais na frase “*Não, eu vô jogar aqui dentro*”, para o sujeito RH, no estilo de elocução “lápiz na boca”. Forma de onda (superior), espectrograma (centro) e segmentação das frases e vogais orais do PB (inferior).

Nesta pesquisa, a opção apenas de sujeitos do sexo masculino é justificada: **(a)** com base em estudos anteriores que preveem que os formantes da fala produzidos por vozes masculinas são mais fáceis de medir por algoritmo de detecção de formantes via técnica de LPC, em comparação aos produzidos por vozes femininas. A razão é que a frequência fundamental média de vozes masculinas é, aproximadamente, uma oitava menor que a f_0 média de vozes femininas e, tendo em vista que os formantes podem ser considerados como “feixes” de harmônicos, pode-se dizer que quanto menor é a f_0 , maior é a quantidade de seus harmônicos contida em um “feixe” de determinado tamanho (comprimento de banda). Dessa forma, utilizando um filtro de banda de 300 Hz, o primeiro formante da vogal [i] (frequência central média de 270 Hz e comprimento de banda de 300 Hz, aproximadamente) produzido por um homem irá conter, essencialmente, 3 harmônicos (200, 300, 400 Hz). Uma vez que estas frequências estejam aproximadamente espaçadas, haverá a incorporação em uma mesma banda, o que torna a tarefa de determinar um formante mais fácil. (Künzel, 2001: 97) e **(b)** com o intuito de se aproximar de situações forenses reais, pois crimes cometidos por homens são mais frequentes no cenário forense real (correspondente a, aproximadamente, 94% da população carcerária do Brasil em 2011⁶).

⁶ Dados disponibilizados pelo Departamento Penitenciário Nacional no “Formulário Categoria e indicadores preenchidos: todas UF’s. Categoria: Quantidade de Presos/Internados”. Disponível em: <goo.gl/d3Hpc>. Acesso em: 5 nov. 2012.

A idade dos sujeitos participantes, 21 anos de idade, também foi escolhida com base em dados disponibilizados pelo Departamento Penitenciário Nacional que apontam que, em 2011, aproximadamente 29% da população carcerária do sexo masculino encontra-se na faixa etária de 18 a 24 anos de idade (média de idade de 21 anos) ⁷.

Posteriormente à compilação de todos os dados dos 3 sujeitos no Excel, as amostras de fala foram organizadas da seguinte maneira: (a) “voz habitual” vs. “imitação do Sílvio Santos” e (b) “voz habitual” vs. “uso do lápis na boca”. Para os pares de amostras de fala combinados, foram conduzidas análises estatísticas comparativas intra-locutores, para todos os parâmetros acústicos, por meio da aplicação de um teste T de variáveis independentes e nível de significância de 5%, que tinha como objetivo testar a hipótese nula de que não havia diferenças significativas entre os disfarces vocais e a voz habitual desses locutores.

3.3 PARÂMETROS ACÚSTICOS

Lima *et al* (2007) afirma que as vogais são importantes correlatos da posição dos articuladores e da configuração das cavidades ressoantes, pois são consideradas como resultados de ressonância do trato vocal supraglótico. Dessa forma, a análise de determinados parâmetros na produção das vogais orais mostra-se essencial aos estudos de verificação de locutor, já que por meio de características do trato vocal de um falante pode-se compreender o que é próprio de um indivíduo e o que é comum a outros sujeitos.

Nesta pesquisa foram estudados, para as vogais orais do português brasileiro, os seguintes parâmetros acústicos: frequência fundamental (f_0), a frequência dos quatro primeiros formantes (F1, F2, F3 e F4), ênfase espectral, duração e também duração inter-picos de f_0 e duração das pausas silenciosas.

A frequência fundamental (f_0) é o correlato acústico do número de ciclos completos da vibração das pregas vocais em uma unidade de tempo por segundo. A f_0 está relacionada à entonação da fala e mostra-se sensível a fatores como emoção, idade do falante, doenças relacionadas ao trato vocal e disfarces vocais. Künzel (2000), com o intuito de estudar a influência de diferentes tipos de disfarces no comportamento vocal através de técnicas auditivas e acústicas para reconhecimento de locutor, investigou os efeitos de disfarces vocais selecionados na frequência fundamental. Os resultados mostraram que a maioria dos sujeitos eram, de fato, consistentes em modificar seus valores de f_0 de acordo com o modo de disfarce

⁷ Dados disponibilizados pelo Departamento Penitenciário Nacional no “Formulário Categoria e indicadores preenchidos: todas UF’s. Indicador: Quantidade de Presos por Faixa Etária”. Disponível em: <goo.gl/d3Hpc>. Acesso em: 5 nov. 2012.

selecionado e que havia uma relação constante entre o valor de f_0 para o comportamento de fala habitual de um sujeito e o tipo de disfarce escolhido. No que diz respeito aos resultados apresentados por esse estudo, o autor afirma que a frequência fundamental certamente é um importante parâmetro acústico a ser considerado em estudos envolvendo disfarces vocais, já que muitos destes disfarces afetam diretamente esse parâmetro.

Os formantes caracterizam-se como uma concentração de energia acústica local que reflete o modo de vibração ressoante em um tubo acústico. No caso das vogais, a combinação de frequências predominantes é responsável por produzir as qualidades distintivas de uma vogal. De acordo com Lima *et al (id.)*, as frequências dos dois primeiros formantes conferem a identidade fonética das vogais. A frequência de F1 está relacionada à disposição da faringe atrás do ponto de máxima constrição lingual e sofre influências da altura da língua e do movimento da mandíbula, enquanto a frequência de F2 relaciona-se à cavidade oral e à constrição lingual. Já a frequência de F3 está relacionada à cavidade faríngea. O F4 está relacionado ao comprimento do tubo laríngeo e permanece mais estável que os três primeiros formantes.

A ênfase espectral trata-se de uma medida de identidade relativa de bandas e permite analisar a produção de vogais proeminentes por meio da presença de mais energia em faixas de frequências mais altas, como atestado anteriormente por Traunmüller & Eriksson (2000) e Arantes (2011).

O estudo da duração das vogais orais indica a extensão de tempo empregada na articulação dessas vogais, o que permite que características referentes ao ritmo da fala de indivíduos sejam analisadas.

As curvas da frequência fundamental possuem picos que, quando prosodicamente audíveis, assinalam proeminência ao ouvinte (acentos entoacionais ou *pitch accents*). A sucessão destes picos prosodicamente audíveis ao longo do tempo marca a quantidade de acentos entoacionais que o falante produz e mede como ele organiza a informação no enunciado, no que diz respeito à entonação. Dessa forma, ao realizarmos a análise da duração inter-picos de f_0 por meio da suavização da curva de f_0 , obtemos relatos importantes das características entoacionais de um indivíduo.

O estudo da duração das pausas silenciosas entre as amostras dos sujeitos aqui estudados permite analisar como estes sujeitos organizam seu discurso, a fim de aproximarem-se da fala do locutor-alvo, no caso do disfarce de imitação, e distanciarem-se de suas vozes habituais. Dessa forma, a produção de mais ou menos pausas silenciosas em um

discurso apresenta-se como uma importante estratégia durante a produção de um disfarce vocal.

3.4 ANÁLISES ESTATÍSTICAS

Como já mencionado anteriormente, as vogais orais do PB de cada uma das gravações de cada sujeito foram organizadas em planilhas do programa Excel, juntamente com os valores obtidos para o cálculo de cada um dos parâmetros acústicos testados. A tabela 1 exemplifica a compilação dos dados no Excel para os 12 primeiros segmentos para a amostra de fala “imitação de Sílvio Santos” para o sujeito RH.

Tabela 1: amostra dos cálculos realizados para o sujeito RH, para a gravação utilizando o disfarce vocal “imitação do Sílvio Santos”.

| Segmento | f_0 média (Hz) | F1 (Hz) | F2 (Hz) | F3 (Hz) | F4 (Hz) | Ênfase espectral (dB) | Duração (ms) | Pausas silenciosas (ms) | | Duração inter picos de f_0 (ms) | |
|----------|------------------------|------------|------------|------------|------------|-----------------------------|-----------------|-------------------------------|-----|---|------|
| eU | 124 | 553 | 1438 | 2607 | 3116 | 12,01 | 49 | 1 | 130 | 1 | 0,54 |
| o | 137 | 543 | 1485 | 2739 | 3431 | 6,32 | 76 | 2 | 191 | 2 | 0,24 |
| o | 146 | 513 | 1388 | 2666 | 3065 | 11,21 | 61 | 3 | 191 | 3 | 0,79 |
| a | 149 | 622 | 1362 | 2094 | 3127 | 13,7 | 87 | 4 | 51 | 4 | 0,58 |
| l | 156 | 283 | 1751 | 2443 | 3176 | 10,13 | 35 | 5 | 123 | 5 | 0,54 |
| ae | 138 | 432 | 1880 | 2646 | 3245 | 15,51 | 405 | 6 | 392 | 6 | 0,66 |
| a | 121 | 828 | 1372 | 2228 | 3309 | 15,65 | 63 | 7 | 76 | 7 | 0,87 |
| oh | 134 | 733 | 1975 | 3012 | 3074 | 13,06 | 53 | 8 | 325 | 8 | 0,09 |
| A | 146 | 522 | 1337 | 2119 | 3189 | 14,02 | 35 | 9 | 462 | 9 | 0,7 |
| eU | 147 | 577 | 1435 | 1966 | 3250 | 9,28 | 60 | 10 | 122 | 10 | 0,19 |
| o | 125 | 545 | 993 | 2186 | 3386 | 11,78 | 177 | 11 | 386 | 11 | 0,35 |
| o | 138 | 645 | 1483 | 2535 | 3402 | 9,67 | 67 | 12 | 232 | 12 | 0,59 |

4. RESULTADOS

Nesta seção, serão apresentados os resultados obtidos para as análises comparativas intra-locutor das vogais orais do português brasileiro, através da aplicação de um teste T de variáveis independentes ($\alpha = 0,05$) para os parâmetros acústicos selecionados. Cada locutor teve suas amostras de fala comparadas da seguinte maneira: valores da amostra de fala “voz habitual” vs. “imitação do Sílvio Santos” e “voz habitual” vs. “uso do objeto lápis na boca”

4.1 Sujeito GB

A tabela 2 apresenta os resultados obtidos nos testes T para todos os parâmetros acústicos testados para as vogais orais do PB, entre os dois pares de amostras de fala comparados. A análise estatística comparativa entre o par de amostras de fala “voz habitual” e “imitação do Sílvio Santos”, para esse locutor, apresentou mudanças significativas ($p <$

0,05) para os parâmetros “frequência fundamental média”; frequências do primeiro (F1), terceiro (F3) e quarto (F4) formantes, ênfase espectral e duração das vogais orais do PB. Já, para o par de amostras “voz habitual” e “uso do lápis na boca”, a hipótese nula foi rejeitada para os parâmetros “frequência fundamental média”; terceiro (F3) e quarto (F4) formantes e duração das pausas silenciosas.

Tabela 2: Resultados do teste estatístico comparativo entre as amostras de fala “voz habitual” e “imitação do Sílvio Santos” / “uso do lápis na boca” do sujeito GB, para todos os parâmetros acústicos analisados para as vogais orais do PB.

Resultados do teste T ($\alpha = 0,05$): valor da probabilidade p

| <i>Parâmetros acústicos (vogais orais do PB)</i> | <i>Voz habitual vs. Imitação do Sílvio Santos</i> | <i>Voz habitual vs. Uso do lápis na boca</i> |
|--|---|--|
| <i>Frequência fundamental média</i> | 4,0 E-16 | 4,65 E-11 |
| <i>F1</i> | 2,0 E-15 | 0,19 |
| <i>F2</i> | 0,18 | 0,33 |
| <i>F3</i> | 0,006 | 0,005 |
| <i>F4</i> | 0,04 | 2,13E-12 |
| <i>Ênfase Espectral</i> | 0,0004 | 0,38 |
| <i>Duração</i> | 0,0002 | 0,06 |
| <i>Duração inter-picos de f_0</i> | 0,15 | 0,06 |
| <i>Duração das pausas Silenciosas</i> | 0,07 | 0,0006 |

4.2 Sujeito RH

Os resultados obtidos no teste estatístico para o sujeito RH estão apresentados na tabela 3. Os parâmetros acústicos que indicaram uma diferença significativa entre a amostra de fala “voz habitual” e o disfarce vocal “imitação do Sílvio Santos” foram: frequência fundamental média; frequências do primeiro (F1) e quarto (F4) formantes e ênfase espectral. Na comparação entre do disfarce vocal “uso do lápis na boca” com a amostra de fala “voz habitual”, foram relatadas mudanças significativas para os parâmetros “frequência fundamental média”; terceiro (F3) e quarto (F4) formantes e ênfase espectral.

Tabela 3: Resultados do teste estatístico comparativo entre as amostras de fala “voz habitual” e “imitação do Sílvio Santos” / “uso do lápis na boca” do sujeito RH, para todos os parâmetros acústicos analisados para as vogais orais do PB.

Resultados do teste T ($\alpha = 0,05$): valor da probabilidade p

| <i>Parâmetros acústicos (vogais orais do PB)</i> | <i>Voz habitual vs. Imitação do Sílvio Santos</i> | <i>Voz habitual vs. Uso do lápis na boca</i> |
|--|---|--|
| <i>Frequência fundamental média</i> | 1,05E-08 | 3,84E-07 |
| <i>F1</i> | 0,01 | 0,26 |
| <i>F2</i> | 0,49 | 0,24 |
| <i>F3</i> | 0,32 | 2,13E-11 |
| <i>F4</i> | 7,56E-18 | 2,20E-20 |
| <i>Ênfase Espectral</i> | 7,66E-20 | 5,5E-09 |
| <i>Duração</i> | 0,28 | 0,08 |
| <i>Duração inter-picos de f_0</i> | 0,52 | 0,97 |
| <i>Duração das pausas Silenciosas</i> | 0,29 | 0,08 |

4.3 Sujeito KM

Nos testes conduzidos para o sujeito KM na comparação entre as amostras de fala “voz habitual” e “imitação do Sílvio Santos”, os parâmetros que apresentaram uma diferença significativa entre esses dois pares de amostras foram: “frequência fundamental média”, primeiro (F1) e quarto (F4) formantes, ênfase espectral, duração das vogais orais e duração das pausas silenciosas. Na comparação entre as amostras de fala “voz habitual” e “uso do lápis na boca”, os parâmetros “frequência fundamental média”, os quatro primeiros formantes (F1, F2, F3, F4) e duração das vogais orais rejeitaram a hipótese nula, apresentando diferenças significativas entre os dois pares de amostras. Os resultados dos testes T estão detalhados na tabela 4.

Tabela 4: Resultados do teste estatístico comparativo entre as amostras de fala “voz habitual” e “imitação do Sílvio Santos” / “uso do lápis na boca” do sujeito KM, para todos os parâmetros acústicos analisados para as vogais orais do PB.

| Resultados do teste T ($\alpha = 0,05$): valor da probabilidade p | | |
|--|---|--|
| <i>Parâmetros acústicos (vogais orais do PB)</i> | <i>Voz habitual vs. Imitação do Sílvio Santos</i> | <i>Voz habitual vs. Uso do lápis na boca</i> |
| <i>Frequência fundamental média</i> | 2,00E-26 | 1,89E-12 |
| <i>F1</i> | 9,90E-11 | 0,0001 |
| <i>F2</i> | 0,07 | 0,01 |
| <i>F3</i> | 0,23 | 4,76E-10 |
| <i>F4</i> | 0,03 | 3,13E-05 |
| <i>Ênfase Espectral</i> | 0,0003 | 0,67 |
| <i>Duração</i> | 1,03E-12 | 1,83E-05 |
| <i>Duração inter-picos de f_0</i> | 0,16 | 0,80 |
| <i>Duração das pausas Silenciosas</i> | 0,0003 | 0,35 |

5. ANÁLISE COMPARATIVA INTRA-LOCUTOR PARA OS PARÂMETROS TESTADOS

5.1 *Frequência fundamental média (f0 média)*

Os testes estatísticos conduzidos para os parâmetros acústicos para as vogais orais do PB, mostraram que, para todos os sujeitos analisados, houve uma modificação significativa na produção da frequência fundamental média nas vogais orais para ambos os disfarces vocais comparados à voz habitual desses sujeitos. Visto que a entonação caracteriza-se como um dos principais correlatos perceptivos de f_0 (Nooteboom, 1997), as mudanças nas taxas médias de f_0 podem indicar uma alteração na produção dos acentos tonais das vogais analisadas na realização dos disfarces vocais em relação à voz habitual e caracterizam-se como uma mudança de registro. Para o falante GB, a sua voz ficou globalmente mais grave para o disfarce vocal “lápiz na boca” (m=143)⁸ e mais aguda para a imitação do Sílvio Santos (m=179) em comparação à amostra de voz habitual (m=152). Já para o falante RH, em ambos os disfarces vocais, a sua voz ficou globalmente mais aguda (imitação do Sílvio Santos, m=144 e lápis na boca, m=146) em comparação à voz habitual (m=129). Para o falante KM, a voz habitual ficou globalmente mais aguda (m=123) na realização de ambos os disfarces vocais (imitação do Sílvio Santos, m=154 e lápis na boca, m=134). Os falantes modificaram a frequência fundamental de suas vozes como estratégia para se distanciarem da voz habitual e se aproximarem à voz do apresentador Sílvio Santos, no caso do disfarce de imitação, ou porque o uso do lápis na boca causou perturbações à produção da fala e afetou a produção laríngea, no caso do disfarce “uso do lápis na boca”.

5.2 *Duração das vogais orais do PB*

Os valores obtidos para as medidas da duração apresentaram diferenças significativas apenas na comparação entre as amostras de fala “voz habitual” e “imitação do Sílvio Santos” para os sujeitos GB e KM. Esse resultado mostra que esses locutores realizam um alongamento vocálico para a amostra de fala “imitação

⁸ Valores aproximados.

do Sílvio Santos” em relação às suas vozes habituais. O alongamento vocálico pode ser compreendido como uma estratégia desses sujeitos em aproximar seu estilo de fala ao do apresentador de TV e, dessa forma, ocultar características fonético-acústicas do seu estilo de fala habitual. A análise desse parâmetro acústico não apresentou modificações significativas para as amostras de fala comparadas para o sujeito RH.

5.3 Duração das pausas silenciosas

No que diz respeito à análise da duração das pausas silenciosas, o teste T apresentou diferenças significativas entre os resultados obtidos para o par de amostras de fala “voz habitual” e “uso do lápis na boca” para o sujeito GB e para o par de amostras “voz habitual” e “imitação do Sílvio Santos” para o sujeito KM. Para o primeiro sujeito, uma possível explicação para a realização de um discurso menos pausado, no caso do disfarce vocal, pode ser atribuída à restrição dos movimentos dos lábios, língua e mandíbula, causados pelo posicionamento do lápis entre os dentes frontais. O comprometimento de movimentos articulatórios pode ter afetado a fluência do discurso desse sujeito, o que refletiu na realização natural das pausas. No caso do sujeito KM, houve um alongamento das pausas silenciosas na produção do discurso de imitação do Sílvio Santos. Mais uma vez, essa modificação em relação à sua voz habitual pode ser entendida como uma estratégia em aproximar-se da voz do locutor-alvo (Sílvio Santos) e distanciar-se do seu estilo de fala natural.

5.4 Ênfase espectral

O parâmetro acústico ênfase espectral apresentou diferenças significativas entre as amostras de fala do disfarce de imitação do Sílvio Santos em comparação com a voz habitual dos três sujeitos testados. Para a comparação entre as amostras de fala “uso do lápis na boca” e “voz habitual”, este parâmetro rejeitou a hipótese nula apenas para o sujeito RH. O estudo deste parâmetro, um correlato do esforço vocal, mostra que os locutores GB e RH utilizaram um maior esforço vocal para a produção das vogais orais na imitação do apresentador de TV. Esperava-se que não houvesse alterações na ênfase espectral para as amostras de fala do disfarce “uso do lápis na boca”, visto que a

utilização do lápis não requer esforço laríngeo (esforço vocal) particular, ao contrário da imitação, na qual o objetivo é tentar se aproximar (imitar) a voz do locutor-alvo.

5.5 Formantes

A análise dos formantes entre os pares de amostras de fala “voz habitual x imitação do Sílvio Santos” e “voz habitual x lápis na boca” mostrou uma reconfiguração no espaço vocálico dos locutores pela disposição relativa dos dois primeiros formantes durante a realização dos disfarces vocais. As figuras⁹ de 4 a 9 apresentam as modificações encontradas no espaço vocálico de cada um dos locutores entre a amostra de fala “voz habitual” e os disfarces vocais.

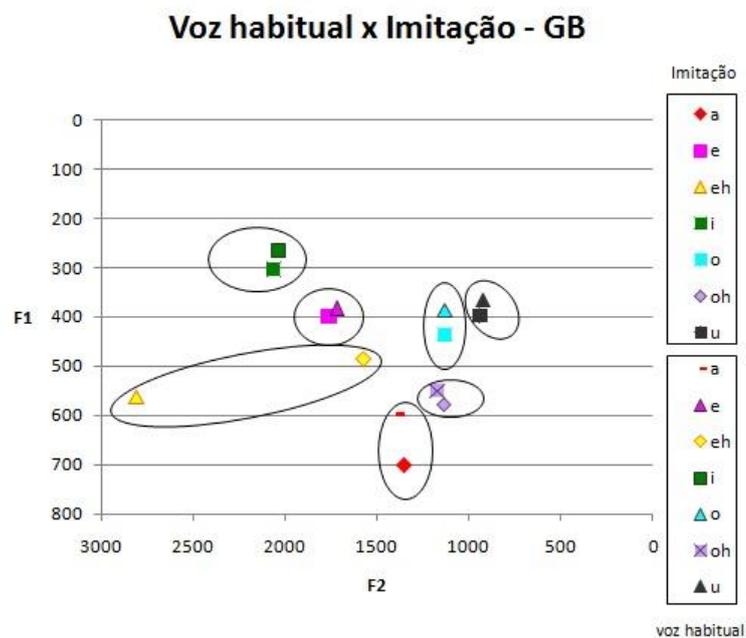


Figura 4: Gráfico de F1 x F2 correspondente à comparação entre as amostras de fala “voz habitual” x “imitação de Sílvio Santos” para o sujeito GB.

⁹ As vogais presentes nas figuras de 4 a 9 estão representadas em caracteres ASCII.

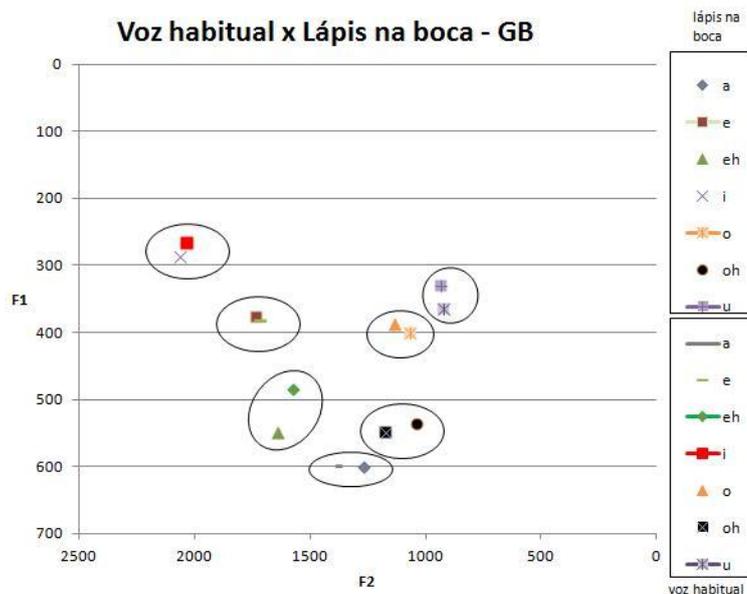


Figura 5: Gráfico de F1 x F2 correspondente à comparação entre as amostras de fala “voz habitual” x “lápiz na boca” para o sujeito GB.

Para o sujeito GB, na comparação entre os espaços vocálicos das amostras de fala “voz habitual” e “imitação de Sílvio Santos”, nota-se que não houve um padrão de mudanças entre as vogais orais. O aumento de F1 implicou em um abaixamento de vogais pertencentes a diferentes posições: [i] (anterior), [a] (central) e [o], [ɔ] e [u] (posteriores), para a amostra de fala “imitação do Sílvio Santos” em relação à voz habitual. O aumento concomitante de F1 e F2 resultou em um abaixamento associado a uma anteriorização da vogal média-baixa anterior [ɛ]. A comparação entre as amostras de fala “voz habitual” e “lápiz na boca” apresentaram uma posteriorização das vogais [o] e [ɔ] e da vogal central [a] pela diminuição de F2. As modificações em F1 provocaram um abaixamento na vogal anterior [ɛ] e uma ligeira elevação na vogal posterior [u].

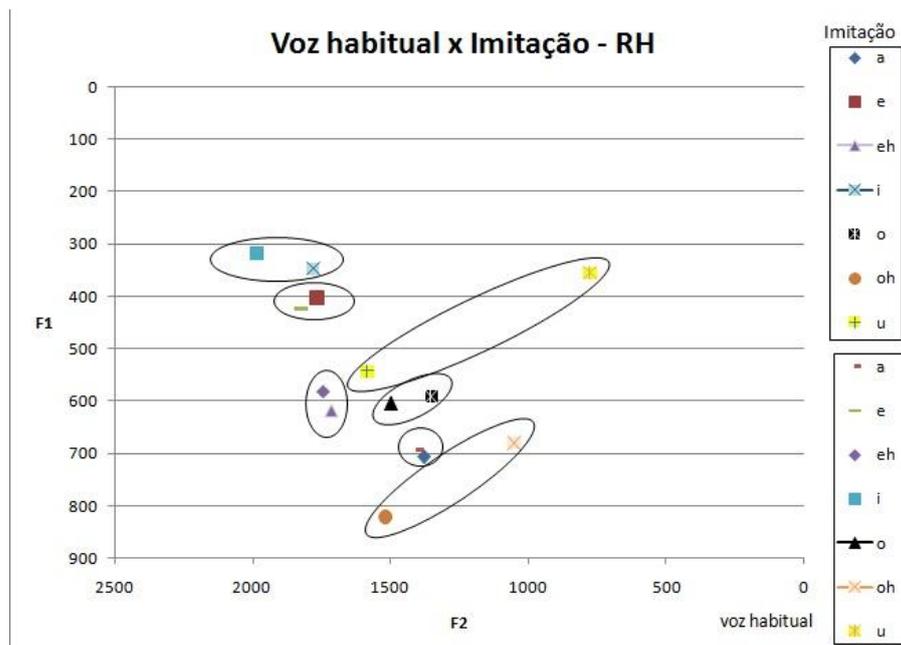


Figura 6: Gráfico de F1 x F2 correspondente à comparação entre as amostras de fala “voz habitual” x “imitação de Sílvia Santos” para o sujeito RH.

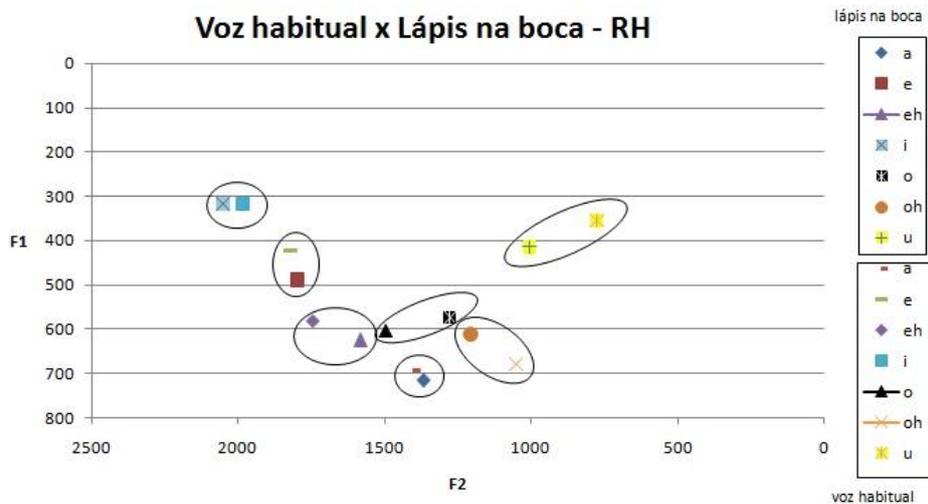


Figura 7: Gráfico de F1 x F2 correspondente à comparação entre as amostras de fala “voz habitual” x “lápiz na boca” para o sujeito RH.

Uma análise geral dos valores obtidos para F1 e F2 para o sujeito RH, apresentaram, para ambos os disfarces vocais em comparação à voz habitual, um abaixamento de vogais

altas ([i]¹⁰ e [u]) e médias ([e]¹¹, [ɛ] e [ɔ]). Além disso, as vogais médias [ɛ](anterior) e [ɔ] (posterior) sofreram uma centralização na amostra de fala “lápiz na boca” pelo aumento nos valores de F1 e diminuição nos valores de F2. A vogal média-alta posterior, [o], sofreu uma posteriorização e uma elevação pela diminuição nos valores de F2 e F1, respectivamente, para ambos os disfarces vocais.

A vogal alta posterior [u] sofreu uma anteriorização pelo aumento nos valores de F1 e F2, em ambos os disfarces vocais. Enquanto, a vogal alta anterior [i] sofreu uma posteriorização devido à diminuição nos valores de F2, para o disfarce vocal “imitação de Sílvio Santos”. A vogal central [a] não atestou mudanças significativas para ambos os disfarces vocais.

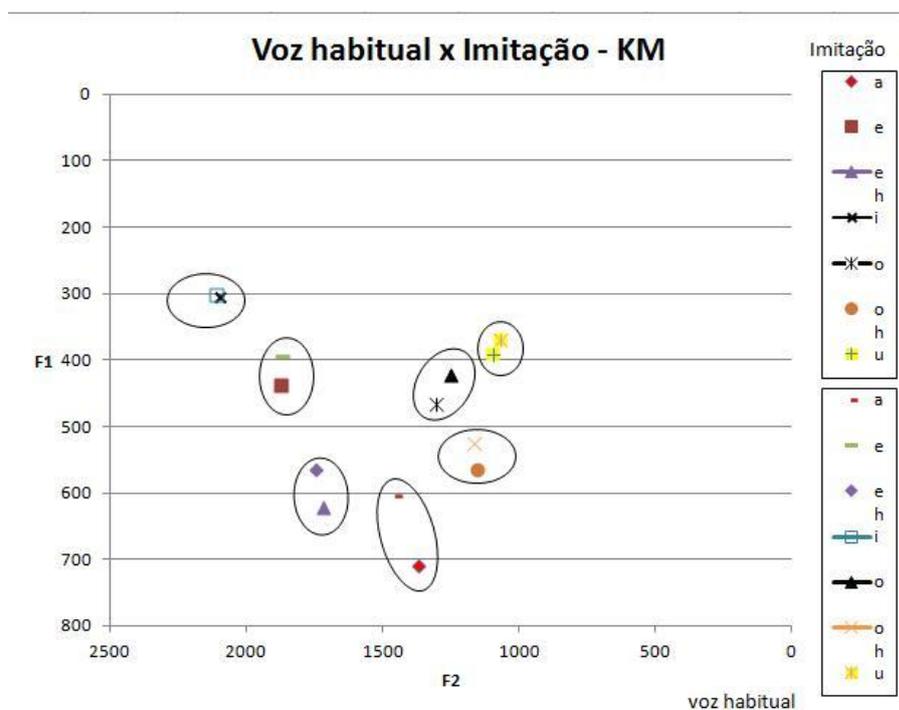


Figura 8: Gráfico de F1 x F2 correspondente à comparação entre as amostras de fala “voz habitual” x “imitação de Sílvio Santos” para o sujeito KM.

¹⁰ A vogal alta anterior [i], para esse locutor, sofreu um abaixamento durante a realização do disfarce vocal “imitação de Sílvio Santos” e uma posteriorização pela diminuição de F2, para o disfarce vocal “lápiz na boca”.

¹¹ A vogal média-alta [e], para esse locutor, sofreu um abaixamento durante a realização do disfarce vocal “lápiz na boca” e uma ligeira elevação pela diminuição de F1 e F2, para o disfarce vocal “imitação de Sílvio Santos”.

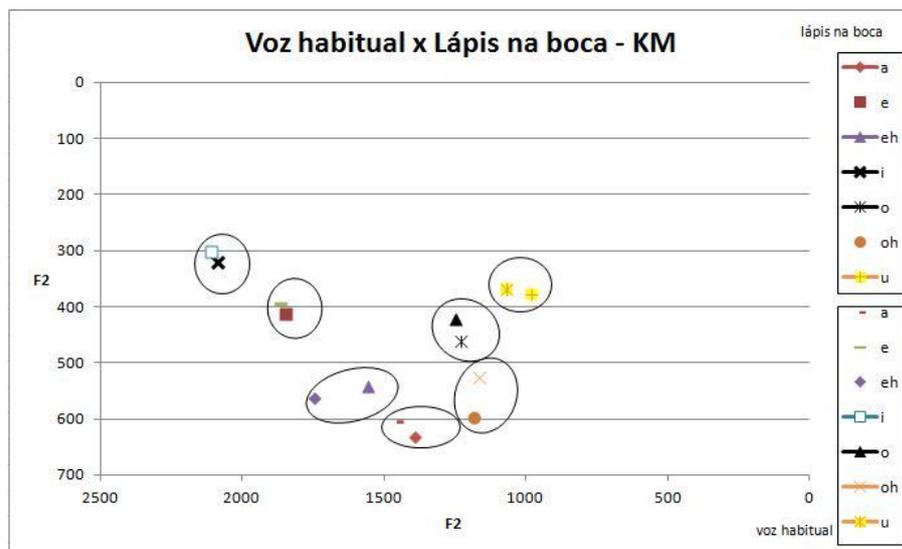


Figura 9: Gráfico de F1 x F2 correspondente à comparação entre as amostras de fala “voz habitual” x “lápiz na boca” para o sujeito KM.

Para esse sujeito houve uma reconfiguração na área do espaço vocálico, durante a realização de ambos os disfarces vocais, pelo aumento nos valores de F1 para a maioria das vogais, resultando em um abaixamento dessas vogais. Para a vogal alta posterior [u] foi atestada uma ligeira posteriorização durante a realização do disfarce vocal “lápiz na boca”. Para a vogal média-baixa [ɛ], houve uma posteriorização e elevação pela diminuição de F1 e F2 durante a realização do disfarce vocal “lápiz na boca”.

Não foram atestadas diferenças significativas entre os disfarces vocais comparados à voz habitual para a vogal anterior [i]. Esse é um bom resultado em termos forenses, pois, mesmo diante da utilização de disfarces vocais capazes de provocar alterações na disposição vocálica dos locutores, a vogal anterior alta desse falante não teve sua produção afetada pelos disfarces vocais, sua produção foi semelhante à produção na voz habitual.

A análise dos dois primeiros formantes e do espaço vocálico dos sujeitos mostrou que, no geral, a utilização desses disfarces vocais resultou em alterações nos movimentos da língua e da mandíbula, visto que houve modificações nos valores dos dois primeiros formantes, principalmente de F1, encontrados para essas vogais orais.

A análise do terceiro e quarto formantes (F3 e F4) apresentou diferenças significativas para praticamente todos os locutores testados entre as amostras de fala comparadas. Esses resultados sugerem que, durante a realização dos disfarces vocais selecionados, ocorrem modificações no trato vocal destes falantes por meio da constrição faríngea (relacionada ao terceiro formante) e de modificações no tubo laríngeo (relacionado ao quarto formante).

5.6 Duração inter-picos de f_0

O teste T não apresentou mudanças significativas para o parâmetro “duração inter-picos de f_0 ” para nenhum dos pares de amostras comparados para os três sujeitos. A análise da frequência do segundo formante (F_2) também não apresentou diferenças significativas, para nenhum dos sujeitos, para as amostras de fala “voz habitual” e “imitação do Sílvia Santos”. A preservação da hipótese nula se mostra importante para essa pesquisa, pois permite reconhecer o que permanece na fala de sujeitos em situações de disfarces vocais.

6. CONSIDERAÇÕES FINAIS

A análise das alterações acústicas e prosódicas na fala de sujeitos do sexo masculino diante da utilização de disfarces vocais selecionados mostrou que o disfarce vocal que apresentou o maior número de diferenças significativas em comparação à voz habitual de todos os locutores, foi o disfarce “imitação do Sílvia Santos”. Para este disfarce os parâmetros acústicos que mais apresentaram modificações em relação à voz habitual foram: frequência fundamental média, primeiro e quarto formantes e ênfase espectral. O disfarce “lápiz na boca” apresentou modificações em relação à voz habitual para os parâmetros: frequência fundamental média, terceiro e quarto formantes.

Os resultados sugerem que os locutores conseguiram desviar da própria fala e alcançar a voz do locutor-alvo (disfarce de imitação), principalmente, no que diz respeito a alterações na entonação, conseguindo que suas vozes mudassem de registro, ficando globalmente mais grave ou aguda e também por produzirem um discurso com maior ou menor esforço vocal. O alongamento vocálico e as alterações no tempo global do discurso mostraram-se como estratégias desses locutores em se distanciarem de suas vozes habituais. Outra modificação importante diz respeito às frequências de formantes e a reconfiguração do espaço vocálico. Estudos anteriores (cf. Figueiredo e Brito, 1996) já haviam comprovado que a utilização do lápis boca entre os dentes frontais e paralelo à língua, limitava os movimentos articulatorios e, dessa forma, a produção de sílabas proeminentes. Por meio da análise dos quatro primeiros formantes, nesta pesquisa foram comprovadas alterações nos movimentos mandibular e lingual e também no comprimento do trato vocal e constrição faríngea.

As modificações encontradas neste estudo estão relacionadas ao empenho desses sujeitos em alcançar a voz do locutor-alvo, no caso do disfarce de imitação, e também à tarefa de se distanciarem de suas vozes habituais, por meio de estratégias acústicas e prosódicas e

como consequência de restrições causadas por instrumentos independentes da produção linguística (como é o caso da utilização do lápis na boca).

Reconhecer o que é próprio do sujeito e o que desvia da sua voz habitual, interessa aos estudos de Verificação de Locutor, pois permite identificar o modifica e o que permanece na fala disfarçada de sujeitos em relação à sua voz habitual.

Esta pesquisa evidenciou que a utilização de disfarces vocais mobiliza um grande número de parâmetros acústicos e prosódicos a serem considerados em uma análise forense, o que a tornou uma tarefa exigente e minuciosa, porém com importantes resultados aos estudos em Fonética Forense.

7. REFERÊNCIAS

ARANTES, P. “Correlatos prosódicos da estrutura informacional em português brasileiro”. In: III Colóquio Brasileiro de Prosódia da Fala, 2011, Belo Horizonte. Resumos do III Colóquio Brasileiro de Prosódia da Fala, 2011.

COLE, J.; SHATTUCK-HUFNAGEL, S. “The phonology and phonetics of perceived prosody: What do listeners imitate?”. In: Proceedings of the Interspeech 2011. Florence, Italy, 21-31 August, p. 969-972, 2011.

ENDRES, W.; BAMBACH, W.; FLÖSSER, G. “Voice Spectrograms as a Function of Age, Voice Disguise, and Voice Imitation“. In the Journal of the Acoustical Society of America, v.48, n.6 (part 2), 1971.

ERIKSSON, A. “Aural/acoustic vs. automatic methods in forensic phonetic case work”. In: Neustein, A.; Patil, H. A. (Orgs.). “Forensic Speaker Recognition: Law Enforcement and Counter-terrorism”. Nova York: Springer-Verlag New York Inc, p. 41-69, 2011.

_____. “The disguised voice: imitating accents or speech styles and impersonating individuals.” In: LLAMAS, C; WATT, D. (Orgs.) “Language and Identities”. Edinburgh: Edinburgh University Press, 2010. Cap.08, p. 86-96. 15

_____; THUNBERG, G.; TRAUNMÜLLER, H. “Syllable prominence: a matter of vocal effort, phonetic distinctness and top-down processing”. In the Proceedings of the Interspeech, p.399-402, 2001.

_____; WRETLING, P. “How flexible is the human voice? – A case study of mimicry”. In: Proceedings of of the European Conference on Speech Technology, v.2, 1997.

FIGUEIREDO, R.; BRITTO, H. “A report on the acoustic effects of one type of disguise”. In: Forensics Linguistics 3 (1), 1996. P. 168-175.

HOLLIEN, H. “Forensic Voice Identification”. London: Academic Press, 2002.

KÜNZEL, H. J. “Beware of the ‘telephone effect’: the influence of the telephone transmission on the measurement of formant frequencies”. In: Forensic Linguistics, v.8, n.1, p.80-99, 2001.

_____. “Effects of voice disguise on speaking fundamental frequency”. In the International Journal of Speech Language and the Law, v.7, n.2, p.150-179, 2000.

LIMA, M.F.B; CAMARGO, Z.A.; FERREIRA, L.P.; MADUREIRA, S. “Qualidade vocal e formantes das vogais de falantes adultos da cidade de João Pessoa”. Rev. CEFAC. 2007; 9(1):99-109.

MARKHAM, D. “Listeners and disguised voices: the imitation and perception of dialectal accent”. Forensic Linguistics 6 (2), 1999. P. 289-299.

NOOTEBOOM, S. G. The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), The Handbook of Phonetic Sciences. Oxford: Blackwell, 1997, P. 640-673.

PERROT, P.; CHOLLET, G. “Helping the Forensic Research Institute of the French Gendarmerie to Identify a Suspect in the Presence of Voice Disguise or Voice Forgery”. In: NEUSTEIN, A; PATIL, H.A. (Eds.). “Forensic Speaker Recognition: Law Enforcement and Counter-terrorism”. New York, NY, Springer Verlag, p. 469-503, 2012.

SJÖSTRÖM, M.; ERIKSSON, E.J.; ZETTERHOLM, E.; SULLIVAN, K.P.H. “A switch of dialect as disguise”. Proceedings of the 19th Swedish Phonetics Conference (FONETIK 2006). Swedish: University of Stockholm, p. 113- 116, 2006.

TRAUNMÜLLER, H., ERIKSSON, A. “Acoustic effects of variation in vocal effort by men, women, and children”. Journal of the Acoustical Society of America, 107(6): 3438-3451, 2000.

ZETTERHOLM, E. “Impersonation: a phonetic case study of the imitation of a voice”. 1997.

_____. “Impersonation – reproduction of speech”. In: Working Papers, 49, p.176-179. 2001.

_____; BLOMBERG, M.; ELENIOUS, D. “A comparison between human perception and a speaker verification system score of a voice imitation”. Tenth Australian International Conference on Speech Science & Technology, 2004, Australia.

_____. “Same speaker – different voices: a study of one impersonator and some of his different imitations”. In: Proceedings of the 11th Australian International Conference on Speech Science & Technology. University of Auckland, New Zealand, 6-8 December, p.70-75, 2006.

8. APÊNDICE

TRANSCRIÇÃO DO VÍDEO DO APRESENTADOR SÍLVIO SANTOS¹²

“Não! Eu vô jogar aqui dentro. Aê! Agora eu vô, vô jogar, aê!

Dá licença aí, dá licença aí, dá licença, eu vô jogar. Olha só, quero ver quem vai pegar a nota de cem? Aê, aê — Sílvio Santos joga o dinheiro para a platéia.

Ai, ai, agora, agora, agora... Vamos lá, vamos lá, agora! Você aí, eu vô jogar!

¹² Disponível em: < <http://youtu.be/-CpTVCaSVJE>>. Acesso em: 16 nov. 2012

Vô jogar! Lá, lá, lá, lá, lá... Ei, você aí, me dá um dinheiro aí! Vô jogar! Aê! Aê!

Ih, caramba! Ih, caramba! Ri, ri, ri — Sílvio Santos cai no chão.

Ei, ei, ai, ai... Segura o velho! Ai, ai, ri, ri, ri.

Me derrubaram, não é possível! Nunca mais eu faço isso! Ai... ai, ai... ai, ai, ai... ri, ri...

Quem quer o aviãozinho? Aê! Quem quer?"