

DOI:10.20396/cel.v60i3.8651760



## ANÁLISE FONÉTICO-ACÚSTICA DE DISFARCES VOCAIS COM APLICAÇÕES PARA A FONÉTICA FORENSE

RENATA REGINA PASSETTI\*  
PLÍNIO ALMEIDA BARBOSA\*\*

**RESUMO:** Esta pesquisa avaliou quais parâmetros acústicos foram modificados e quais permaneceram inalterados na fala de indivíduos durante a utilização de dois disfarces vocais: um envolvendo a utilização de um lápis posicionado atrás dos dentes incisivos e outro buscando imitar a fala do apresentador de televisão Sílvio Santos. O *corpus* foi composto por amostras de fala de três locutores do sexo masculino que leram a transcrição de um discurso deste apresentador em três situações: (1) utilizando seus padrões de fala habituais; (2) imitando o apresentador Sílvio Santos; e (3) utilizando um lápis na boca. Foram analisadas as frequências dos dois primeiros formantes; frequência fundamental, por meio dos descritores estatísticos mediana e valor de base; ênfase espectral; duração das vogais orais do português brasileiro e das pausas silenciosas; e a duração entre dois picos consecutivos da frequência fundamental. Os resultados mostram que os locutores conseguiram desviar da própria fala por meio de alterações na entonação, no esforço vocal, em alongamentos vocálicos e por alterações no tempo global do discurso. As alterações nas frequências formânticas resultaram na reconfiguração do espaço vocálico. Essas modificações estavam associadas ao desejo de alcançar o registro vocal do locutor-alvo, no caso da imitação, ou às restrições impostas pelo uso do lápis na boca. Não foram atestadas modificações significativas nos valores obtidos para a duração interpicos da frequência fundamental para nenhum dos locutores.

**Palavras-chave:** disfarce de voz; imitação; fonética forense.

**ABSTRACT:** This research evaluated which acoustic parameters were modified and which remained undifferentiated in speakers' speech samples in two vocal disguise situations: placing a pencil behind the incisive teeth, and trying to impersonate Sílvio Santos' speech, a well-famous Brazilian TV host. The corpus consisted in recordings of a discourse transcription from this TV host by three male Brazilian Portuguese speakers in these situations: (1) with speakers' own natural voices and speaking styles, (2) trying to imitate Sílvio Santos' voice and speaking style, and (3) by firmly placing a pencil behind the incisive teeth. The acoustic techniques analyzed were the frequencies of the first two formants, the statistical descriptors median and baseline of fundamental frequency, spectral emphasis, and duration of oral vowels, silent pauses and between two consecutive peaks of fundamental frequency (inter  $F_0$ -peaks duration). The results have shown that speakers succeed in deviating from their own speech by changing fundamental frequency levels, vocal effort, vowel durations, and speech rate. The formant frequencies modifications resulted in a reconfiguration of speakers' acoustic space. All these modifications were related to a desire of speakers to achieve a good performance at impersonating Sílvio Santos' speaking style, as well as to restrictions imposed by the use of the pencil in the mouth. There were not attested any significant modification for inter  $F_0$ -peaks duration.

**Keywords:** vocal disguise; impersonation; forensic phonetics.

\* Universidade Estadual de Campinas (Unicamp), Campinas, SP, Brasil. [re.passetti@gmail.com](mailto:re.passetti@gmail.com)

\*\* Universidade Estadual de Campinas (Unicamp), Campinas SP, Brasil. [pabarbosa.unicampbr@gmail.com](mailto:pabarbosa.unicampbr@gmail.com)

## INTRODUÇÃO

No campo de atuação da fonética forense, as tarefas envolvendo o reconhecimento de locutores podem ser prejudicadas pela utilização de disfarces de fala. Dentre os tipos de disfarces vocais reportados pela literatura na área, Eriksson (2010) afirma que aqueles envolvendo a manipulação da voz por métodos eletrônicos são os que impõem mais dificuldades de análise para as tarefas forenses. No entanto, os de natureza simples, como aqueles envolvendo a manipulação de modos de fonação, são registrados em maior número.

Pesquisas na área dedicaram-se a descrever parâmetros modificados pelo uso de diferentes tipos de disfarces vocais. Utilizando amostras de fala do português brasileiro (doravante PB), Figueiredo e Britto (1996) estudaram os efeitos acústicos na fala de sujeitos que utilizavam como disfarce um lápis posicionado firmemente entre os dentes frontais e abaixo da face inferior da língua. Os autores comentam que este tipo de disfarce é encontrado no Brasil em situações nas quais os criminosos suspeitam que estejam sendo gravados. Os resultados da análise das frequências dos três primeiros formantes de vogais orais do PB mostraram que houve uma modificação na configuração dos formantes provocada pelo abaixamento das vogais altas e elevação das vogais baixas, isto é, uma centralização das vogais em geral. Do ponto de vista perceptual, tais alterações podem obscurecer pistas dialetais importantes para a identificação de locutores.

Künzel (2000) investigou a influência de três tipos de disfarces (aumento e diminuição do *pitch*, e constrição das narinas durante o ato de falar) sobre os valores da frequência fundamental ( $F_0$ ). Os resultados mostraram que a escolha do disfarce vocal estava relacionada à avaliação que os sujeitos faziam de seus próprios comportamentos vocais. Assim, aqueles cujo padrão de  $F_0$  era maior que média do grupo optaram por disfarçar suas vozes aumentando seus níveis de  $F_0$ , enquanto aqueles cujo padrão de  $F_0$  era menor que a média diminuíram ainda mais seus níveis deste parâmetro. Essas modificações causaram alterações na qualidade de voz desses sujeitos, que passou de modal para falsete e crepitante, respectivamente.

O uso de disfarces envolvendo técnicas de modificação vocal também merece destaque na área de verificação de locutores, principalmente em situações nas quais um indivíduo pode tentar reproduzir o padrão vocal de outro visando acessar sistemas que utilizam comandos de voz como chaves de segurança, por exemplo. Nessas situações, faz-se necessário acrescentar o estudo da imitação no contexto de análise forense.

Em uma pesquisa sobre este tema, Eriksson e Wretling (1997) investigaram quais os parâmetros mobilizados por um imitador profissional para se aproximar das vozes-alvo. As técnicas de análise envolviam o estudo da organização temporal da fala, da frequência fundamental e dos quatro primeiros formantes de vogais do sueco padrão. Os resultados mostraram que o imitador obteve sucesso ao tentar se aproximar das características vocais dos locutores-alvo ao mobilizar parâmetros como a frequência fundamental e as frequências formânticas. Porém, não foram realizadas modificações significativas na organização temporal da fala

do imitador, sendo este parâmetro considerado robusto em análises desta natureza, isto é, passível de interesse para a área forense.

O estudo de Cole e Shattuck-Hufnagel (2011), por sua vez, investigou quais elementos prosódicos eram imitados na fala espontânea. Os resultados mostraram que os locutores raramente distorciam a presença de um acento de altura (*pitch accent*) ou de uma fronteira de sintagma entoacional, porém mudanças na natureza das pistas fonéticas foram comuns, assim como na duração da pausa ou na ocorrência de períodos glotais irregulares associados a fronteiras e acentos de altura no inglês americano.

De forma a acrescentar às tarefas da área de fonética forense desenvolvidas no cenário pericial brasileiro, este estudo busca avaliar as “estratégias” utilizadas por locutores durante a realização de disfarces que envolvem a manipulação de padrões articulatórios com o intuito de ocultar suas identidades fonético-fonológicas.

## METODOLOGIA

### Sujeitos

Participaram desta pesquisa<sup>1</sup> três locutores do sexo masculino com idade de 21 anos na época da gravação. Todos eram estudantes de graduação da Universidade Estadual de Campinas, exibiam sotaque do sudeste brasileiro (cidades do interior do estado de São Paulo) e não havia queixa por parte deles de problemas cognitivos, auditivos ou fonoarticulatórios.

A escolha por locutores apenas do sexo masculino justifica-se pela tentativa de se aproximar do cenário forense brasileiro. De acordo com o último levantamento estatístico divulgado pelo Departamento Penitenciário Nacional<sup>2</sup>, 93 % da população carcerária brasileira era composta por homens.

### Procedimentos de gravação

As gravações dos registros de fala dos locutores foram realizadas por meio da leitura de um texto de, aproximadamente, 130 palavras correspondente à transcrição de um discurso<sup>3</sup> do apresentador de televisão Sílvio Santos. As leituras ocorreram em três situações e evocaram diferentes tipos de elocução. Primeiramente foi solicitado aos locutores que lessem o texto com seus registros

---

<sup>1</sup> As gravações desta pesquisa foram realizadas de acordo com a norma 196 do CNS que rege pesquisas com seres humanos e foram armazenadas no banco de dados do Grupo de Estudos de Prosódia da Fala.

<sup>2</sup> Dados disponibilizados pelo Departamento Penitenciário Nacional em 2014 no “Formulário Categoria e indicadores preenchidos: todas UF’s. Categoria: Quantidade de Presos/Internados”. Disponível em: <<http://portal.mj.gov.br/data/Pages/MJC4D50EDBPTBRNN.htm>>. Acesso em: 10 jan. 2018.

<sup>3</sup> A transcrição foi obtida a partir de um vídeo disponibilizado no sítio eletrônico do YouTube e está apresentada na seção “ANEXO”.

de fala e voz habituais. Em seguida deveriam ler imitando a voz de Sílvio Santos, cuja identidade só lhes foi revelada neste momento e, por isso, deveriam basear suas imitações em características da fala do apresentador presentes em suas memórias acústicas. Por fim, de forma a replicar parcialmente<sup>4</sup> a metodologia empregada por Figueiredo e Britto (1996), utilizando um lápis posicionado horizontalmente atrás dos dentes incisivos.

Os equipamentos utilizados nas gravações foram um microfone dinâmico Shure, modelo BETA 58A, e uma placa MBox 2, ligada ao programa computacional Pro Tools. O software de análise acústica PRAAT, versão 6.0.36 (BOERSMA & WENNINK, 2017) foi utilizado para a segmentação das vogais orais do PB e das pausas silenciosas. As medições acústicas foram realizadas automaticamente por *scripts* específicos e foram seguidas de correções manuais, quando necessário.

### Parâmetros fonético-acústicos

As técnicas de análise acústica investigadas nessa pesquisa envolveram o cálculo dos seguintes parâmetros:

- Frequência fundamental ( $F_0$ ), calculada em Hertz por meio dos descritores estatísticos mediana e valor de base. O valor de base é o estimador de  $F_0$  correspondente à frequência neutra de vibração das pregas vocais, isto é, o modo de vibração para o qual as pregas vocais retornam após excursões prosódicas (LINDH e ERIKSSON, 2008). Neste estudo, o valor de base da  $F_0$  foi calculado de acordo com a proposta de Lindh e Eriksson (2008), definida pela expressão:

$$F_b = F_{média} - 1,43 \cdot \sigma_{F_0}$$
, em que  $F_b$  representa o valor de base de  $F_0$ , e  $F_{média}$  e  $\sigma_{F_0}$  são, respectivamente, o valor da média aritmética e do desvio-padrão de  $F_0$  da amostra.

- Frequências dos dois primeiros formantes (doravante F1 e F2), calculados em Hertz no ponto médio das vogais por meio da técnica Linear Predictive Coding (LPC);
- Ênfase espectral, medida que corresponde à diferença de energia em diferentes faixas de frequência, é considerada um correlato do esforço vocal. Neste estudo, a ênfase espectral foi calculada em decibéis, de acordo com a proposta de Traunmüller e Eriksson (2000) que considera a diferença de energia entre a energia total da vogal e sua energia até 1,5 vezes a  $F_0$  máxima da vogal.
- Duração: das vogais orais do PB, das pausas silenciosas e entre dois picos consecutivos de  $F_0$ , calculadas em milissegundos. A duração entre picos de  $F_0$  permite analisar a sucessão de acentos entoacionais que o locutor produz e a forma como ele organiza o enunciado do ponto de vista entoacional. O cálculo deste parâmetro foi precedido pela suavização da curva de  $F_0$  com o valor limiar de 1,5 Hz.

---

<sup>4</sup> Faz-se necessário justificar que a metodologia empregada neste estudo difere em parte daquela descrita por Figueiredo e Britto (1996) por não solicitar aos locutores que posicionassem o lápis sob a face inferior da língua, como o fizeram os autores citados (cf.: FIGUEIREDO e BRITTO, 1996: 169).

## **Técnicas de análise acústica**

As análises estatísticas objetivaram comprovar a significância da diferença entre as médias dos parâmetros supracitados considerando comparações para o conjunto de locutores entre suas amostras de fala habituais e cada uma das suas amostras de disfarces vocais (imitação e uso do lápis na boca). Para isso, aplicou-se o teste de Kruskal-Wallis, alternativa não paramétrica para a Análise de Variância (ANOVA) de um fator, para o qual assumiu-se o nível de significância de 5 %. A hipótese nula testada afirmava que não havia diferenças significativas entre as médias do conjunto de parâmetros entre os disfarces vocais e a voz habitual desses locutores.

A fim de analisar as modificações acústicas provocadas pelo uso dos disfarces no espaço vocálico dos locutores, aplicou-se o método de Lobanov (1971) para normalização dos valores das frequências formânticas das vogais orais do PB. De acordo com Barbosa e Madureira (2015: 302), este procedimento permite a comparação de dados entre locutores distintos, uma vez que características formânticas relacionadas à fisiologia do trato desses locutores são normalizadas de maneira extrínseca às vogais e intrínseca aos locutores e formantes.

As análises estatísticas e a técnica de normalização de Lobanov foram conduzidas no ambiente estatístico R (R DEVELOPMENT CORE TEAM, 2017).

## **RESULTADOS**

### **Teste estatístico de Kruskal-Wallis**

Os resultados do teste estatístico de Kruskal-Wallis estão apresentados na Tabela 1. Para os parâmetros que apresentaram resultados significativos, conduzimos o teste *post hoc* não-paramétrico de Wilcoxon para múltiplas comparações utilizando a correção de Bonferroni para o valor de *p*.

**Tabela 1:** Resultados do teste estatístico Kruskal-Wallis para comparação entre os valores de variância dos parâmetros acústicos para o fator DISFARCE.

PARÂMETROS		KRUSKAL-WALLIS
Frequências de F1		$\chi^2(2) = 6,3999$ ; $p = 0,04$
Frequências de F2		$\chi^2(2) = 3,0871$ ; $p = 0,21$
Frequência Fundamental	Mediana	$\chi^2(2) = 133,01$ ; $p < 2,2 \cdot 10^{-16}$
	Valor de base	$\chi^2(2) = 87,122$ ; $p < 2,2 \cdot 10^{-16}$
Ênfase espectral		$\chi^2(2) = 83,79$ ; $p < 2,2 \cdot 10^{-16}$
Duração	Vogais orais do PB	$\chi^2(2) = 9,6371$ ; $p = 8,0 \cdot 10^{-3}$
	Pausas silenciosas	$\chi^2(2) = 14,527$ ; $p = 7,0 \cdot 10^{-4}$
	Inter picos de $F_0$	$\chi^2(2) = 4,1754$ ; $p = 0,12$

A análise mostrou que os locutores modificam determinados parâmetros em função dos níveis do fator DISFARCE. Os parâmetros que apresentaram resultados significativos foram: frequências de F1; mediana e valor de base de  $F_0$ ; ênfase espectral; e duração das vogais orais do PB e das pausas silenciosas. A condução do teste *post hoc* de Wilcoxon permitiu analisar a quais comparações entre os níveis do fator DISFARCE estavam relacionados os resultados significativos do teste de Kruskal-Wallis para o conjunto de parâmetros supracitados. Os resultados do teste de Wilcoxon estão apresentados na Tabela 2.

**Tabela 2:** Resultados do teste *post hoc* de Wilcoxon com correção de Bonferroni para comparação entre os níveis do fator DISFARCE entre os parâmetros significativos (ns = “não significativo”).

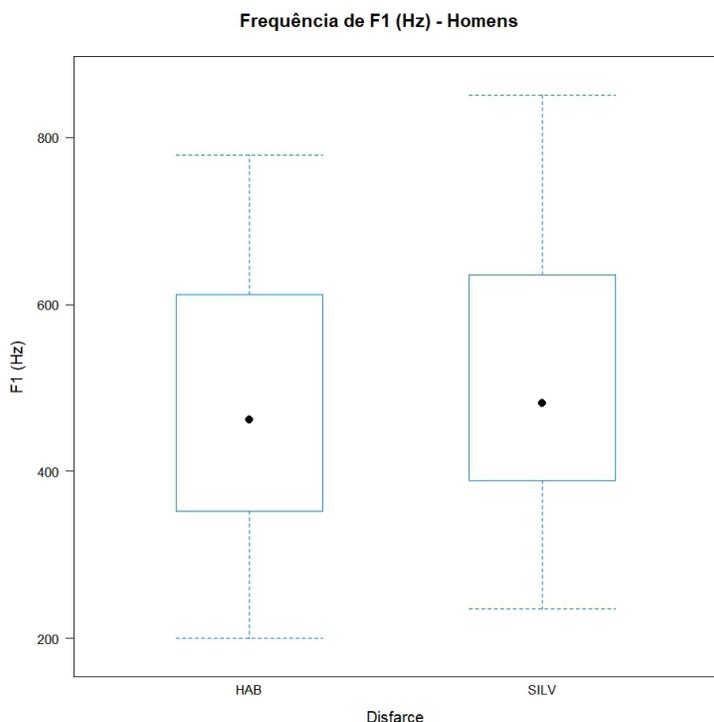
Parâmetros		Comparações	
		Habitual × Lápis na boca	Habitual × Imitação
Frequências de F1		ns	$p = 0,07$
Frequências de F2		ns	ns
Frequência Fundamental	Mediana	$p = 2,0 \cdot 10^{-4}$	$p < 2,2 \cdot 10^{-16}$
	Valor de base	$p = 4,0 \cdot 10^{-4}$	$p < 2,2 \cdot 10^{-16}$
Ênfase espectral		ns	$p < 2,2 \cdot 10^{-16}$
Duração	Vogais orais do PB	ns	$p = 7,6 \cdot 10^{-3}$
	Pausas silenciosas	$p = 6,6 \cdot 10^{-3}$	$p = 1,2 \cdot 10^{-3}$
	Inter picos de $F_0$	ns	ns

Os resultados do teste *post hoc* de Wilcoxon mostram que, para a comparação entre a fala habitual e o disfarce que utilizava um lápis na boca, os parâmetros frequência fundamental, pelos descritores estatísticos “mediana” e “valor de base”, e a duração das pausas silenciosas apresentam resultados significativos. Já para a comparação entre a fala habitual e o disfarce de imitação do apresentador Sílvio Santos, os parâmetros que apresentaram modificações significativas foram: frequências de F1 (marginalmente), descritores “mediana” e “valor de base” de  $F_0$ , ênfase espectral e duração das vogais orais do PB e das pausas silenciosas. A seguir, comentaremos as modificações atestadas para cada um desses parâmetros.

## Frequências formânticas

Em relação aos valores das frequências do primeiro e segundo formantes, os testes estatísticos atestaram diferença marginalmente significativa apenas para os valores de F1 na comparação entre o estilo de elocução habitual dos locutores e a imitação do apresentador Sílvio Santos. A Figura 1 apresenta os diagramas de caixas<sup>5</sup> (*boxplots*), nos quais é possível visualizar a variação dos valores de F1 em função dos níveis do fator “disfarce” que se mostraram significativos.

**Figura 1:** Diagramas de caixas (*boxplots*) contendo a variação das medidas do primeiro formante (Hz) para os níveis “habitual” (HAB) e “imitação” (SILV) do fator DISFARCE.



A análise dos diagramas de caixa revela que os valores das frequências de F1 dos locutores aumentaram<sup>6</sup> durante a imitação do Sílvio Santos. Este aumento

<sup>5</sup> O ponto central em negrito no interior de cada diagrama corresponde à mediana do parâmetro correspondente. As extremidades superior e inferior da caixa retangular indicam, respectivamente, o primeiro e terceiro quartil. As linhas pontilhadas exteriores aos diagramas indicam as variabilidades externas aos quartis.

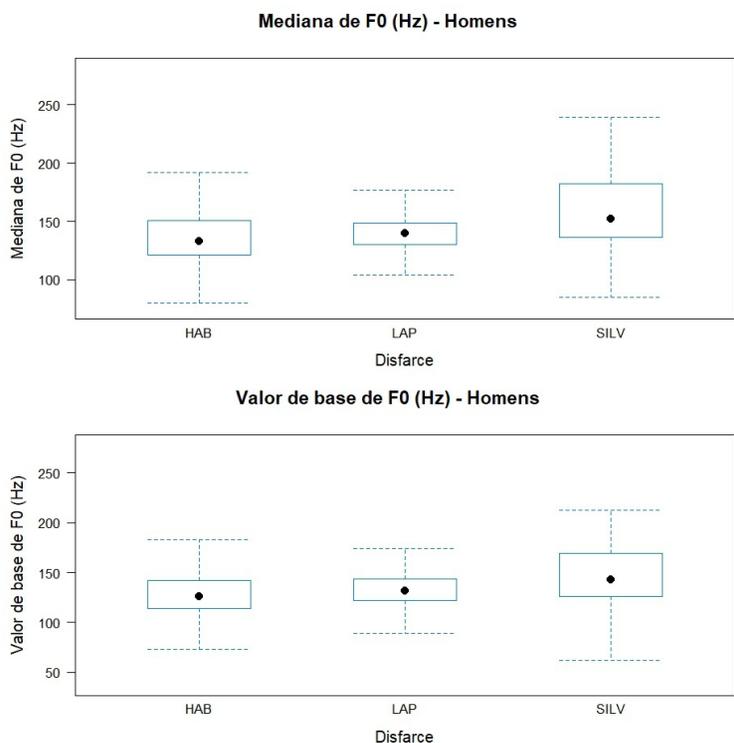
<sup>6</sup> Faz-se necessário esclarecer que os resultados estatísticos apresentados para a análise das frequências de F1 para comparação “habitual” x “imitação” dizem respeito às vogais investigadas em sua totalidade. O maior número de dados da vogal /a/ possivelmente influenciou nos resultados obtidos para este parâmetro (elevação dos valores médios). As modificações específicas nas frequências de F1 das demais vogais podem ser observadas na Figura 6.

foi de 21 Hz, uma vez que o valor médio de F1 desses sujeitos para o estilo de elocução habitual que era de 487 Hz passou a ser de 508 Hz durante a imitação do apresentador.

## Frequência Fundamental

Os descritores estatísticos de  $F_0$  apresentaram resultados significativos para a comparação entre o estilo de elocução habitual e os dois disfarces vocais. A Figura 2 apresenta os diagramas de caixas para a mediana e o valor de base de  $F_0$ .

**Figura 2:** Diagramas de caixas (*boxplots*) contendo a variação das medidas da mediana de  $F_0$  (Hz) (superior) e valor de base de  $F_0$  (Hz) (inferior) para os níveis “habitual” (HAB), “lápiz na boca” (LAP) e “imitação” (SILV) do fator DISFARCE.



Pela análise das figuras, é possível notar que, para a mediana de  $F_0$ , o uso de ambos os disfarces aumentou os valores deste parâmetro em relação aos valores originais, relacionados ao estilo de elocução habitual desses locutores. Esses aumentos foram, em média, de 5 Hz para o disfarce do lápis na boca ( $F_{0MED} = 141$  Hz) e 22 Hz para a imitação ( $F_{0MED} = 158$  Hz). A análise dos diagramas mostra, ainda, que a diminuição da variabilidade dos valores deste parâmetro durante o uso do disfarce “lápiz na boca” em comparação com o estilo de elocução habitual

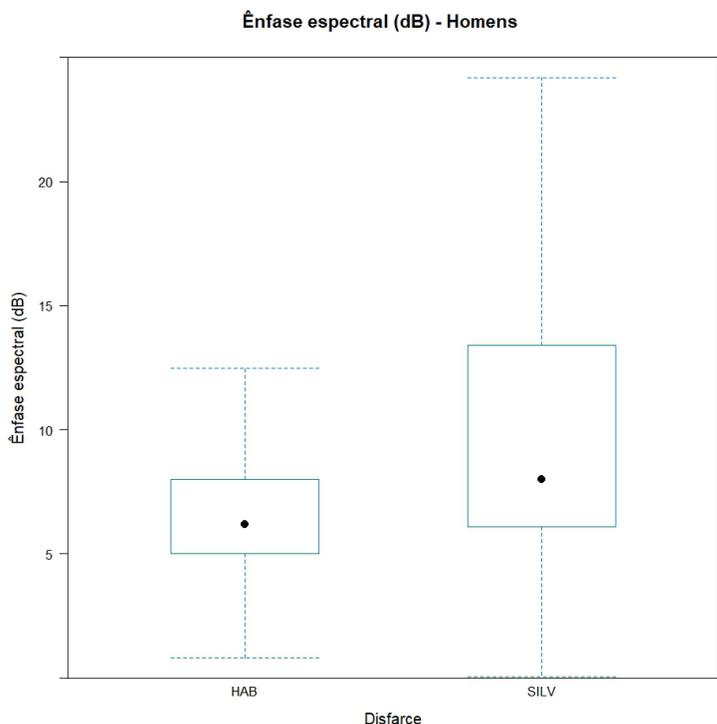
dos locutores. Por outro lado, a variabilidade nos valores deste parâmetro durante a imitação aumentou.

Os resultados referentes ao valor de base da frequência fundamental também indicaram um aumento nos valores deste parâmetro para ambos os disfarces em comparação com o estilo de elocução habitual. Este aumento foi de 4 Hz para o disfarce do lápis na boca ( $F_b = 133$  Hz) e 18 Hz para a imitação ( $F_b = 146$  Hz). Quanto à variabilidade das medidas, mais uma vez a imitação apresentou maior variabilidade se comparada às demais situações.

### Ênfase espectral

Para o parâmetro “ênfase espectral” foram atestadas modificações significativas apenas para a imitação do apresentador Sílvio Santos, conforme apresentado na Figura 3.

**Figura 3:** Diagramas de caixas (*boxplots*) contendo a variação das medidas da ênfase espectral (dB) para os níveis “habitual” (HAB) e “imitação” (SILV) do fator DISFARCE.



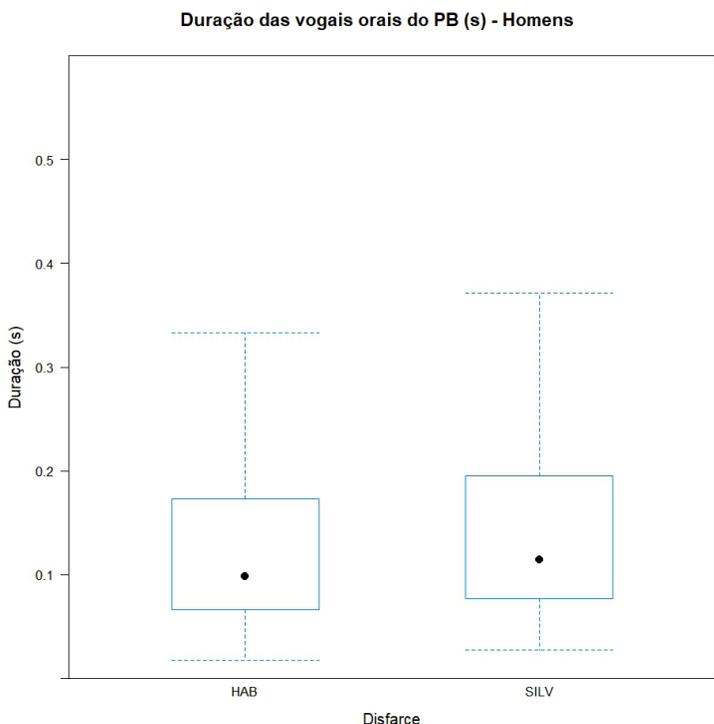
A análise dos diagramas de caixa mostra que o aumento e maior variabilidade nos valores deste parâmetro durante a imitação do apresentador se comparado ao

estilo de elocução habitual desses sujeitos. Este aumento foi de, aproximadamente 3 dB ( $\hat{\text{Ênfase}}_{\text{imitação}} = 9,6 \text{ dB}$ ).

## Duração

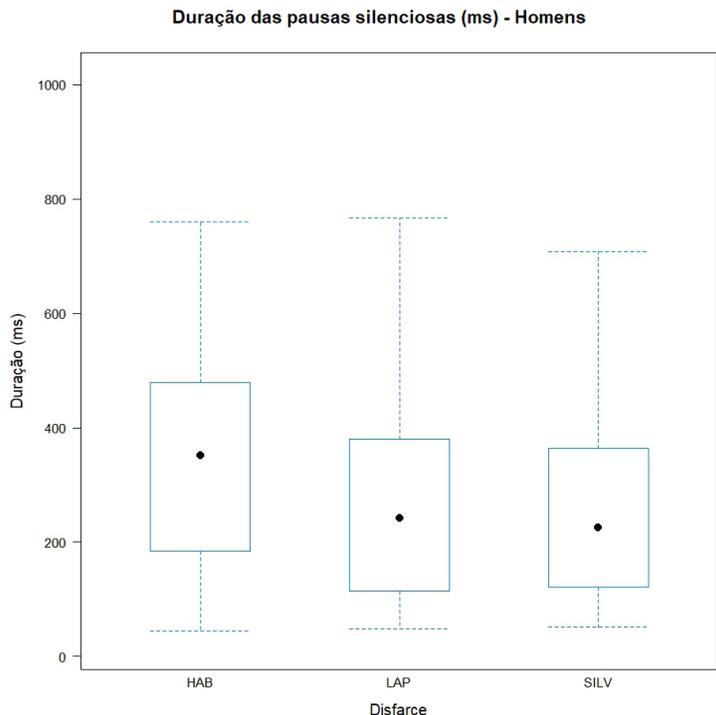
Os testes estatísticos apresentaram resultados significativos para a análise da duração das vogais orais do PB e das pausas silenciosas. As Figuras 4 e 5 apresentam os diagramas de caixa para esses parâmetros.

**Figura 4:** Diagramas de caixas (*boxplots*) contendo a variação das medidas da duração das vogais orais do PB (s) para os níveis “habitual” (HAB) e “imitação” (SILV) do fator DISFARCE.



A análise da Figura 4 revela que as vogais orais do PB foram produzidas com maior duração durante a imitação do apresentador Sílvio Santos ( $\text{dur}_{\text{imitação}} = 150 \text{ ms}$ ) se comparado ao estilo de fala habitual ( $\text{dur}_{\text{habitual}} = 128 \text{ ms}$ ). Apesar de significativa, a diferença média de 22 ms entre os dois estilos de elocução não é audível do ponto de vista perceptivo. A observação do diagrama de caixa deste parâmetro revela, também, que as durações das vogais orais durante a imitação apresentaram maior variabilidade se comparadas ao estilo de elocução habitual.

**Figura 5:** Diagramas de caixas (*boxplots*) contendo a variação das medidas da duração das pausas silenciosas (ms) para os níveis “habitual” (HAB), “lápiz na boca” (LAP) e “imitação” (SILV) do fator DISFARCE.

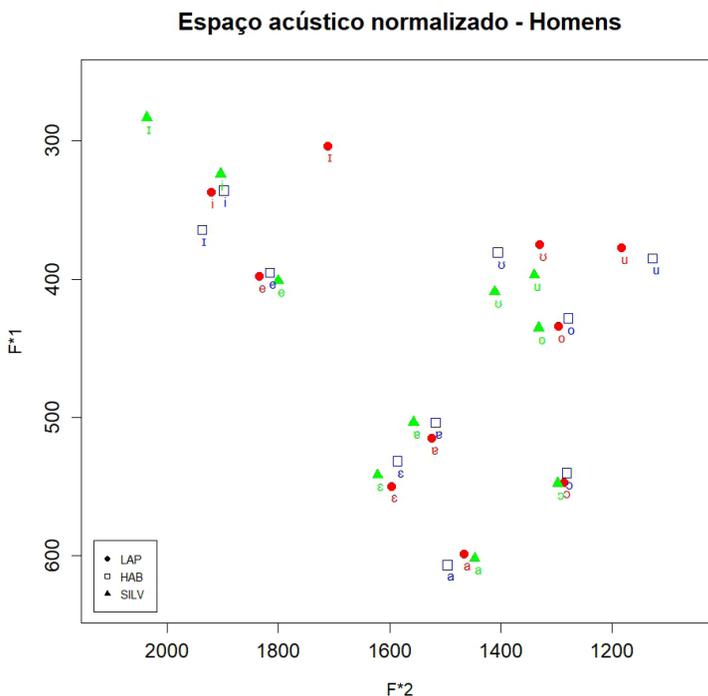


Em relação à duração das pausas silenciosas, a análise estatística mostrou significância para ambos os disfarces vocais. Se comparados ao estilo de elocução habitual, nota-se a diminuição na duração das pausas silenciosas. Essa diminuição foi, em média, de 79 ms para o disfarce “lápiz na boca” ( $\text{durPS}_{\text{lápiz}} = 268 \text{ ms}$ ) e de 92 ms para a imitação ( $\text{durPS}_{\text{imitação}} = 255 \text{ ms}$ ).

### Espaço acústico dos locutores

O método de normalização de Lobanov (LOBANOV, 1971) para os valores das frequências de F1 e F2 permitiu avaliar o efeito dos disfarces vocais no espaço acústico dos locutores. A Figura 6 apresenta a disposição espacial das vogais orais do PB nas três condições de análise: estilo de elocução habitual, uso do lápis na boca e imitação do apresentador Sílvio Santos.

**Figura 6:** Espaço acústico normalizado F1 x F2, em Hertz, das vogais orais dos locutores para os níveis “habitual” (HAB), “lápiz na boca” (LAP) e “imitação” (SILV). Os rótulos F\*1 e F\*2 correspondem aos valores normalizados das frequências do primeiro e segundo formantes, respectivamente.



De modo geral, a análise do espaço acústico revela que as vogais altas foram mais afetadas pelo uso dos disfarces vocais. Para o disfarce vocal “lápiz na boca”, a vogal alta anterior em posição átona (/i/) apresentou valores de F1 e F2 menores se comparados ao estilo de elocução habitual. Essas alterações fizeram com que esta vogal ocupasse uma posição mais alta e mais centralizada do que sua posição original. A vogal alta posterior em posição átona (/u/), por sua vez, manteve os valores de F1 próximos aos valores originais (estilo de elocução habitual), mas teve os valores de F2 diminuídos, o que provocou sua posteriorização no espaço acústico. A vogal alta posterior tônica (/u/) sofreu ligeira diminuição nos valores de F1 e aumento nos valores de F2, o que resultou na sua elevação e anteriorização em relação a sua posição original. As demais vogais permaneceram em posições próximas às originais.

Para a imitação do apresentador Sílvio Santos, a vogal alta anterior em posição átona (/i/) sofreu uma diminuição nos valores de F1 e aumento nos valores de F2, situando-se numa posição mais elevada e anterior se comparada a sua posição para o estilo de elocução habitual. Já a vogal alta posterior em posição átona (/u/) sofreu apenas um aumento nos seus valores de F1 provocando seu abaixamento

em relação ao estilo de elocução habitual. A vogal alta posterior tônica (/u/), por sua vez, teve seus valores de F1 e F2 aumentados em relação a sua posição original e passou a ocupar uma posição mais baixa e mais centralizada no espaço acústico. Não foram atestadas modificações significativas para as demais vogais.

## DISCUSSÃO

As análises conduzidas mostraram que a utilização de ambos os disfarces vocais modificou parâmetros acústicos da fala dos locutores investigados. Comparado ao estilo de elocução habitual, o disfarce “lápis na boca” apresentou diferenças significativas em descritores estatísticos da frequência fundamental e na duração das pausas silenciosas.

As modificações relacionadas à frequência fundamental mostraram que o aumento nos valores deste parâmetro foi acompanhado da diminuição na variabilidade das medidas. Este aumento, ainda que pequeno, pode estar relacionado a uma estratégia dos locutores para compensar a restrição articulatória imposta pela presença do lápis entre os dentes frontais, impossibilitando a livre movimentação dos articuladores. Assim, os locutores podem ter optado por fazer compensações a nível fonatório visando completar a tarefa (leitura do texto) proposta, uma vez que a articulação dos sons no trato oral estava prejudicada pela presença do lápis.

A menor variabilidade nos valores deste parâmetro e a diminuição da duração das pausas silenciosas, que está associada à elevação da taxa de elocução, também estão relacionadas à restrição articulatória devido ao uso do lápis, pois a presença do objeto também impossibilitou a livre movimentação dos articuladores durante a produção dos sons da fala, justificando a variabilidade reduzida. A elevação da taxa de elocução, manifestada pela diminuição do número e duração das pausas silenciosas, pode estar relacionada a um possível estranhamento que o uso do lápis na boca, uso incomum para este objeto, possa ter causado nos locutores fazendo com que se comportassem dessa maneira.

A imitação do apresentador televisivo Sílvio Santo, por sua vez, foi o disfarce que apresentou o maior número de diferenças significativas em comparação ao estilo de elocução habitual dos locutores. Para este disfarce, foram atestadas modificações significativas para os descritores da frequência fundamental, ênfase espectral, duração das vogais orais do PB e das pausas silenciosas, e marginalmente para as frequências de F1.

Os resultados indicam que ao imitar o Sílvio Santos a voz dos locutores foi produzida com valores globais da frequência fundamental mais elevados e seus registros de fala foram produzidos com maior esforço vocal, modificação relacionada a variações na ênfase espectral. Além disso, as análises duracionais mostraram que a imitação foi produzida com maior número de alongamentos vocálicos e menor número de pausas silenciosas. A produção de sentenças marcadas por alongamentos vocálicos é uma característica conhecida da fala do apresentador Sílvio Santos. A redução na produção de pausas silenciosas está

associada à taxa de elocução mais elevada. Possivelmente, a imagem acústica que os locutores possuem da fala deste apresentador está associada à uma taxa de elocução mais elevada que suas taxas habituais. As modificações desse conjunto de parâmetros acústicos revelam as estratégias que os locutores empregaram para se aproximarem da voz do locutor-alvo.

Quanto à disposição das vogais no espaço acústico normalizado, o uso de ambos os disfarces afetou principalmente as vogais altas em posição átona (/ɪ/ e /ʊ/) e a tônica /u/. Para o disfarce “lápiz na boca”, a vogal /ɪ/ ocupou uma posição mais alta e central em comparação com sua produção original, a vogal /ʊ/ posteriorizou-se e a vogal /u/ ocupou uma posição mais alta e central. Já para a imitação, a vogal /ɪ/ ocupou uma posição mais alta e anterior, a vogal /ʊ/ abaixou-se e a vogal /u/ ocupou uma posição mais baixa e central.

Os resultados das frequências formânticas referentes ao disfarce “lápiz na boca” diferem, em parte, daqueles reportados por Figueiredo e Britto (1996: 170 – 171). Em relação às frequências de F1, os autores atestam modificações em maior grau para as vogais altas – resultado também verificado em nosso estudo. Porém, a direção dessas modificações é diferente da aqui encontrada. Os autores reportam que, de maneira geral, há um aumento nos valores de F1 para as vogais altas e que esse aumento é ainda maior para as vogais posteriores arredondadas. Em nosso estudo, no entanto, houve pouca variação nos valores de F1 dessas vogais e as modificações estavam, no geral, associadas à diminuição nos valores desse formante.

Em relação aos resultados obtidos para as frequências de F2, Figueiredo e Britto (1996: 170 – 171) reportam aumento nas frequências desse formante para as vogais posteriores arredondadas (em maior grau quanto mais arredondada e posterior for a vogal), enquanto para as demais vogais, os valores de F2 tendem a diminuir durante o uso do disfarce. Em nosso estudo, os resultados de F2 atestados para as vogais posteriores seguem os reportados pelos autores supracitados, sendo a vogal /u/ aquela cujos valores de F2 foram mais afetados pelo uso do disfarce. A vogal /a/, por sua vez, também teve seus valores de F2 ligeiramente diminuídos durante o uso do disfarce. No entanto, em relação às vogais anteriores, apesar de pouca variação em relação ao estilo de elocução habitual, as mudanças estavam associadas, em sua maioria, ao aumento dos valores de F2 na situação do disfarce. Acreditamos que as diferenças reportadas entre os dois estudos se devam, principalmente, a diferenças metodológicas associadas ao posicionamento do lápis, uma vez que, diferentemente de nosso estudo, Figueiredo e Britto (1996) pediram aos seus locutores que posicionassem o lápis sob a superfície inferior da língua. A distinção na disposição exata do lápis entre os dois estudos pode ter afetado a movimentação dos articuladores (principalmente da língua) de maneira diferente.

Além de se investigar os efeitos do uso de disfarces vocais em parâmetros do sinal da fala, visando contribuir com as tarefas desenvolvidas na área de fonética forense, é importante considerar também os parâmetros que não apresentaram modificações significativas pelo uso dos disfarces vocais, isto é, aqueles que permaneceram inalterados independentemente da técnica de disfarce adotada.

Dentre os parâmetros inalterados, faz-se necessário ressaltar aqui a duração interpícos de  $F_0$ , parâmetro relacionado à sucessão de acentos entoacionais e que revela como os falantes organizam a informação do enunciado do ponto de vista entoacional. Esta medida rítmica mostrou-se resistente às modificações acústicas impostas por ambos os disfarces e pode ser considerada como um importante parâmetro a ser utilizado na identificação de falantes em análise forense.

## CONSIDERAÇÕES FINAIS

Este trabalho teve como principal objetivo analisar os efeitos fonético-acústicos de dois tipos de disfarces de voz (o uso de um lápis posicionado atrás dos dentes incisivos e a imitação de um apresentador televisivo) no estilo de elocução habitual de locutores do sexo masculino.

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 seus estilos de elocução habituais, por meio de estratégias acústicas e independentemente da produção linguística, esta como consequência de restrições causadas pelo uso de instrumentos, no caso da utilização do lápis na boca.

Outro resultado interessante diz respeito ao parâmetro “duração interpícos de  $F_0$ ” que não apresentou modificações significativas para nenhum dos disfarces vocais analisados. A preservação da hipótese nula se mostra importante para as tarefas forenses, pois permite reconhecer o que permanece, ou seja, o que não modifica na fala de sujeitos mesmo em situações de disfarces vocais.

A análise de modificações acústicas no sinal da fala motivadas pelo uso de disfarces vocais e evidenciadas nesta pesquisa possui aplicações para as tarefas práticas desenvolvidas no campo da fonética forense e busca contribuir para o trabalho de peritos e outros profissionais da área.

## AGRADECIMENTOS

Os autores agradecem o financiamento concedido pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo nº. 2011/23054-3. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não necessariamente refletem a visão da FAPESP.

---

## REFERÊNCIAS

BARBOSA, P.; MADUREIRA, S. *Manual de Fonética Acústica Experimental: aplicações a dados do português*. São Paulo: Editora Cortez, 2015.

- BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer* (Versão 6.0.36). Disponível em: <http://www.praat.org>. Acesso em 2017.
- BRASIL. Ministério da Justiça, 2014. Departamento Penitenciário Nacional. *Levantamento Nacional de Informações Penitenciárias*. INFOPEN – Dezembro 2014.
- 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.
- ERIKSSON, A. 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
- ERIKSSON, A. ; WRETLING, P. How flexible is the human voice? – A case study of mimicry. In: *Proceedings 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.
- KÜNZEL, H. J. 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.
- LINDH, J.; ERIKSSON, A. Robustness of Long Time Measures of Fundamental Frequency. In: *Proceedings of the Interspeech 2007*. Anwerp, Belgium, 27-31 August, p. 2025-2028, 2007.
- LOBANOV, B. M. Classification of Russian vowels spoken by different listeners. In: *Journal of Acoustical Society of America*, v. 49, p. 606-608, 1971.
- R DEVELOPMENT CORE TEAM, version 3.4.3. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <<http://www.Rproject.org>>. Acessado em 2017.
- 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.

## ANEXO

### TRANSCRIÇÃO DO VÍDEO DO APRESENTADOR SÍLVIO SANTOS<sup>7</sup>

“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 ((Sílvia Santos joga o dinheiro para a plateia)).  
Ai... Ai... Agora... Agora... Agora... Vamos lá... Vamos lá... Agora! Você aí... Eu vô jogar!  
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 ((Sílvia 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?”

---

<sup>7</sup> Vídeo disponível em <<http://youtu.be/-CpTVCaSVJE>>. Acesso em: 16 nov. 2012.