

KREMER, Robinson Luis; GOMES, Maria Lúcia de Castro. A eficiência do disfarce em vozes femininas: uma análise da frequência fundamental. *ReVEL*, vol. 12, n. 23, 2014. [www.revel.inf.br].

A EFICIÊNCIA DO DISFARCE EM VOZES FEMININAS: UMA ANÁLISE DA FREQUÊNCIA FUNDAMENTAL

Robinson Luis Kremer¹

Maria Lúcia de Castro Gomes²

malugomes@utfpr.edu.br

roberk2005@gmail.com

RESUMO: Para se avaliar a qualidade vocal, podem-se utilizar dois recursos importantes: a análise perceptivo-auditiva e a análise acústica. Esta última análise é mais objetiva e faz uso de programas computadorizados que demonstram quantitativamente vários aspectos mensuráveis do sinal de voz captado. Um desses aspectos é a frequência fundamental (f_0). Neste trabalho, analisou-se a eficiência de disfarce de voz agregando três aspectos: a análise perceptiva, a qualidade da voz e a curva de f_0 de um trecho da fala (voz normal e com disfarce) de mulheres dispostas em três grupos com 10 integrantes cada, nomeados de Grupo de Garotas (GG), Grupo de Mulheres (GM) e Grupo Controle (GC). Nos resultados preliminares, as falantes que optaram por deixar a voz mais grave não obtiveram grande eficiência no disfarce. Já as falantes que tornaram a voz mais aguda ou que utilizaram outros tipos de ajustes foram mais eficazes na mudança da f_0 , obtendo mais eficiência no disfarce.

Palavras-chave: Análise acústica; Frequência fundamental; Disfarce de voz.

INTRODUÇÃO

Existem diversas definições para o fenômeno da fala, uma delas é dada por Kent e Read ao citar, em seu livro, o pioneiro dos estudos da fala, Raymond H. Stetson, que define a fala como sendo o movimento que se faz audível. No entanto, os próprios autores complementam que a fala seria muito mais do que isso, ganhando sua maior importância como o principal meio pelo qual a linguagem é expressa em todas as culturas da humanidade – exceto entre as comunidades surdas – e tendo o sinal acústico como o produto final (KENT; READ, 2002).

¹ Graduando em Letras Inglês-Português na Universidade Tecnológica Federal do Paraná (UTFPR).

² Professora Doutora na Universidade Tecnológica Federal do Paraná (UTFPR).

Para ocorrer o fenômeno da fala, são utilizados cerca de 80 músculos diferentes, de forma rápida e precisa. Cada falante, ao produzir os movimentos necessários para a produção desse fenômeno, vai apresentar características próprias, que tornará sua fala um elemento individualizante. A análise acústica da fala permite a identificação de traços característicos da produção do falante. Através de elementos gráficos e resultados numéricos e estatísticos, é possível uma análise de eventos que ocorram, mesmo que muito rapidamente, pela dinâmica das informações durante a fala.

A Fonética, nossa área de pesquisa, possui como objeto a fala, mediante estudo da produção e da percepção de seus sons. No âmbito científico, de acordo com Kent e Read (2002, p. 2), “a fala possui três grandes áreas de estudo: fisiológica (ou fonética articulatória), acústica (ou fonética acústica) e perceptual (ou comumente chamada percepção da fala)”.³ Dentro do grupo das ciências fonéticas, tem-se a Fonética Forense, ramo da Linguística Aplicada que oferece conhecimentos e métodos de análise linguística em contextos de investigações de delitos e evidências para processos judiciais que envolvam o uso da linguagem falada. Segundo Braid (2003, pp. 5-6), “a fonética forense não se encerra na identificação de falantes, mas incide em todos os mistérios criminalísticos que envolvam aspectos da fala ou, extrapolando, de sons em geral”. Sendo assim, a análise da fala, principalmente por meio da Fonética, serve como uma ferramenta multidisciplinar, auxiliando diversas áreas, uma delas, a Forense.

Uma das atividades da Fonética Forense é a verificação do locutor, para se determinar se duas falas foram produzidas por um mesmo falante. Desde que gravações passaram a ser utilizadas como meio de prova, no caso do Brasil, um perito pode ser solicitado por um Delegado de Polícia, Procurador, Promotor de Justiça ou Juiz (CARNEIRO; GOMES, no prelo) para analisar as amostras e elaborar um laudo pericial, seja na corte ou em contexto de simples investigação. Com frequência, na busca de determinar a identidade do falante nas gravações, por meio da linguagem utilizada, os peritos podem fazer inferências sobre o perfil do falante, como por exemplo, a classe social ou a origem. Também se recorre aos especialistas para definir o que está sendo dito na amostra, em casos de má qualidade da gravação, ou para confirmar se aquela amostra teve algum tipo de edição. O trabalho com essas

³ “[...] speech has three major arenas of study: the physiologic arena (or *physiologic phonetics*), the acoustic arena (or *acoustic phonetics*), and the perceptual arena (typically called *speech perception*)”. Todas as traduções apresentadas são de nossa responsabilidade.

atividades, assim como uma série de outras que envolvam a fala ou sons em geral gravados em algum tipo de mídia, ficou conhecido como fonética forense (NOLAN, 1999).

O ser humano possui uma grande habilidade em reconhecer pessoas pela voz mesmo em situações adversas, e um grande desafio para os cientistas da fala é o de estabelecer um modelo que reproduza essa habilidade. Tal desafio tem motivado inúmeros estudos na área de identificação de falantes, e com o desenvolvimento de sistemas de processamento digital de sinal, tem-se percebido um grande impulso da área que, no entanto, continua bastante carente de pesquisas.

Com o objetivo de desenvolver estudos em fonética forense, um grupo multidisciplinar foi composto por professores e alunos do curso de Letras, juntamente com um grupo de peritos criminais, na Universidade Tecnológica Federal do Paraná (UTFPR). Esse grupo iniciou, a partir da criação de um banco de dados, uma série de experimentos com propósito formativo, ou seja, de aprendizado das atividades relacionadas ao trabalho da perícia.

Este artigo traz resultados de um desses experimentos, em que se pesquisou a diferença produzida em fala normal e disfarçada de vozes femininas, através de análise perceptiva e acústica. Para a análise perceptiva, em duas sessões diferentes, os avaliadores, primeiramente, testaram a eficiência do disfarce por reconhecimento ou não das vozes e, em um segundo momento, analisaram as estratégias de disfarces escolhidas por cada uma das participantes. Para a análise acústica, realizada também em dois momentos, primeiramente, foi medida a frequência fundamental (f_0) de um trecho de fala e verificada a diferença entre os valores em voz normal e em voz disfarçada. Em um segundo momento, foi analisada a curva de f_0 também nas duas modalidades de voz.

1 FREQUÊNCIA FUNDAMENTAL

A frequência fundamental da voz – conhecida também pela sigla f_0 – é o menor componente periódico resultante da vibração das pregas vocais. É a primeira frequência produzida na glote. Na voz, a f_0 indica tanto as variações de altura (sons agudos ou graves) como as de intensidade (sons fortes ou fracos). Para Braid (2003, p. 51):

Em análise de fala, as taxas de repetição dos padrões periódicos numa forma de onda são denominadas frequência fundamental, entretanto, perceptualmente, esta impressão auditiva é chamada de *pitch*. O *pitch* corresponde à sensação de som grave ou agudo, e suas escalas.

As pessoas com vozes dentro de uma faixa mais aguda são consideradas como tendo *pitch* alto, por conseguinte, as de vozes mais graves, *pitch* baixo. No entanto, independentemente da característica pessoal de voz grave ou aguda, as pessoas alteram o *pitch* durante sua fala, pois esse elemento é responsável pela entonação. Segundo Roach (2009), não há uma definição completamente satisfatória para o termo entonação, que muitas vezes se confunde com o termo “prosódia”, que também não é facilmente definido, conforme explicam Kent e Read (2002, p. 229).

Uma grande discordância está entre os termos “prosódia” e “entonação”. Alguns escritores se referem a eles como sinônimos, enquanto outros apontam uma importante distinção entre eles. [...] Entonação é parecida com prosódia no sentido de que seus parâmetros são frequência vocal, intensidade e duração, mas entonação se refere a uma gama mais restrita de fenômenos, geralmente os padrões de altura de *pitch* e de tonicidade de uma língua. A prosódia inclui esses efeitos, mas também abrange o tempo (pausa e duração), esforço vocal, sonoridade, e outros fenômenos.⁴

Ao proferir um turno de fala, o falante acaba por produzir variação do *pitch*, uma vez que se utiliza de tons diversos para se expressar. A análise acústica da f_0 permite o traçado de gráficos que se denominam *curvas de contorno de pitch*, em que se encontram os valores instantâneos da frequência fundamental em função do tempo (BRAID, 2003).

A frequência fundamental é, certamente, um parâmetro fonético promissor na identificação de falantes. Segundo Künzle (2000), é o parâmetro mais conhecido nas pesquisas de reconhecimento de voz/identificação de falantes. Em sua pesquisa, esse autor reconhece que as formas mais frequentes de disfarce afetam diretamente a f_0 .

2 A FREQUÊNCIA FUNDAMENTAL E A IDENTIFICAÇÃO DE FALANTES

Num relato de Tosi (1979), em 1937, Frances McGehee – primeira pesquisadora de testemunhas auditivas – realizou o primeiro experimento

⁴ One major disagreement is with the pair of terms, prosody and intonation. Some writers regard them as synonyms, while others mark an important distinction between them. [...] Intonation is similar to prosody in that its parameters are vocal frequency, intensity, and duration, but intonation refers to a narrower range of phenomena, generally the patterns of pitch rises and falls and the patterns of stress in a given language. Prosody includes these effects but also embraces *tempo* (pause and lengthening), vocal effort, loudness, and other phenomena.

significativo para a área de análise acústica, valendo-se da memória de longo prazo. Como informantes, a pesquisadora utilizou um total de 31 homens e 18 mulheres, todos estudantes de pós-graduação, que leram um parágrafo de 56 palavras. Depois disso, 740 estudantes de graduação participaram como ouvintes de duas sessões, que variaram de um dia a cinco meses entre uma sessão e outra. Nesse experimento, McGehee investigou, entre outros aspectos, os efeitos do disfarce na voz pela mudança do *pitch* e observou que alterações da frequência fundamental reduziam o reconhecimento consideravelmente. Segundo Künzel (2000), a f_0 é um parâmetro bastante popular no reconhecimento de falantes da área forense, pois se trata de um parâmetro de fácil extração e medição se comparado com outros como, por exemplo, os formantes⁵.

Quanto à variação, ao menos em línguas não tonais⁶, a frequência fundamental não é afetada pelo conteúdo lexical da fala, todavia, são conhecidos muitos fatores não linguísticos que afetam a f_0 , como o estado de saúde, o estado emocional, o gênero do discurso, os ruídos do ambiente e se a pessoa está ou não falando ao telefone. Kinoshita *et al.*, (2009) referem pesquisas de French (1994) e Braun (1995), em que se notou que uma única pessoa pode demonstrar uma grande variação na f_0 dependendo da ocasião e até mesmo durante uma única sessão de gravação.

Para Kinoshita *et al.* (2009), uma vez que a força inerente dos parâmetros de identificação de falantes na área forense (*Forensic Speaker Recognition – FSR*) se concentra primariamente nas taxas de variação intrafalantes e interfalantes, a f_0 não é considerada muito eficaz como um parâmetro dessa identificação, embora alguns pesquisadores (NOLAN, 1983; SAMBUR, 1975; JIANG, 1996 *apud* KINOSHITA *et al.*, 2009) tenham sugerido que a f_0 seja um forte parâmetro. Kinoshita *et al.* (2009) citando seu trabalho de 2005, afirma que, devido a sua baixa taxa de variação, a duração da f_0 de longo termo mostra-se pouco forte como evidência na identificação. De todo modo, a análise forense da f_0 tem se concentrado até o momento em sua média de duração e desvio padrão.

⁵ Segundo Kent e Read (2002, p. 302), formante é “uma ressonância do trato vocal. Um formante é especificado por sua frequência central (geralmente chamada frequência de formante) e largura de banda. Os formantes são indicados por números inteiros que aumentam de acordo com a localização da frequência. F1 é o formante de frequência mais baixa, F2 é o próximo, e assim por diante”. No original: *Formant is a resonance of the vocal tract. A formant is specified by its center frequency (commonly called formant frequency) and bandwidth. Formants are denoted by integers that increase with the relative frequency location of the formants. F1 is the lowest-frequency formant, F2 is the next highest, and so on.*

⁶ Língua tonal é aquela em que o tom participa como elemento distintivo das suas unidades lexicais.

Antes de dar prosseguimento, é necessário definir algumas nomenclaturas utilizadas em contexto forense para análise de voz, como a diferença entre identificação e verificação de falantes e também entre identificação técnica e não técnica.

2.1 IDENTIFICAÇÃO VS. VERIFICAÇÃO DE FALANTES

Tradicionalmente autores como Hollien (2002) e Nolan (1999) têm feito uma distinção entre dois termos para o reconhecimento de falantes, *verificação* e *identificação*. Começando pela verificação de falantes, a definição é bem simples. De acordo com Hollien (2002, p. 5), na verificação, “não é necessário determinar o que está sendo dito, mas quem está falando”.⁷ O uso desse tipo de reconhecimento pode ser ilimitado. Um exemplo desse uso seria o acesso a uma área restrita por comando de voz, e em situações assim, o falante quer e necessita ser reconhecido a partir de um banco de dados (HOLLIEN, 2002).

Na identificação de falantes, é preciso identificar um falante desconhecido por análise de voz em contextos em que, geralmente, residem alguns problemas: ruídos do ambiente ou baixa qualidade do sinal – em uma conversa por telefone, por exemplo – e também distorções na voz, como em caso em que a fala é proferida em situação de muito estresse. Essa distorção na voz pode ainda ser resultado de medo ou de excitação (quando cometendo algum delito) ou na tentativa de disfarçar a voz. Nesse caso, o falante não quer ser identificado (HOLLIEN, 2002).

No Brasil, no entanto, o termo “verificação de locutor” tem sido mais comumente utilizado no contexto forense (BRAID, 2003). Em textos mais recentes, aparecem outros termos como comparação de falante (no Brasil, comparação de locutor), comparação de voz, reconhecimento de voz (GOMES; CARNEIRO, no prelo). Seja qual for o termo utilizado, a identificação de falante pode ser feita por uma pessoa especializada ou por alguém que não tenha nenhum conhecimento técnico.

2.2 IDENTIFICAÇÃO TÉCNICA E NÃO TÉCNICA

A identificação não técnica é basicamente realizada por pessoas não treinadas. “Na área forense esse tipo de identificação aparece sob o nome de ‘evidência de

⁷ “[...] it is not necessary to determine what is being said but rather who is talking”.

testemunha auditiva” (NOLAN, 1999, p. 755).⁸ Nesse contexto, o criminoso pode ser identificado durante o tempo em que o delito está sendo cometido, no caso de se tratar de uma pessoa conhecida da testemunha. Em outra situação, a testemunha pode vir a ser chamada para prestar depoimento e identificar, dentre diversas vozes mostradas, qual pertence ao sujeito que cometeu o crime.

Na identificação técnica, a gravação é obtida durante o ato criminoso, seja uma ameaça de bomba ou um pedido de resgate em um sequestro, por exemplo. O suspeito é chamado a depor e, durante uma entrevista em gravação, o perito tenta controlar o contexto linguístico para que essa gravação seja comparada com a gravação anterior (a do crime). Ainda dentro desse tipo de identificação, usa-se a percepção auditiva e a análise acústica (NOLAN, 1999).

A percepção auditiva é realizada por profissionais especializados, foneticistas ou fonoaudiólogos, por exemplo, que irão analisar a fala, a fim de verificar a qualidade das vogais e consoantes, diferenças de pronúncia, características do falante. No ambiente de análise acústica, se obtém muito mais do que apenas a análise da fala. Para Nolan (1999, p. 762), “também se revelam informações que nosso sistema auditivo obscurece”.⁹

Uma vez esclarecidas essas nomenclaturas, este artigo descreve o início de um trabalho de análise da eficiência do disfarce com foco na frequência fundamental ao comparar trechos de fala em duas modalidades: voz normal e disfarçada. A pesquisa aqui relatada teve como objetivo verificar a eficiência de disfarce de voz agregando três aspectos: (1) análise perceptiva; e análise acústica em que se observa (2) a qualidade vocal; e (3) a curva de f_0 . O objetivo era saber se apenas alterando a frequência fundamental se conseguiria eficiência (deixando a voz mais aguda ou grave), ou se seria utilizado algum outro recurso para modificar a voz (como por exemplo, voz nasal ou áspera) para se obter um disfarce eficiente. Conforme Eriksson, 2010, nas pesquisas em fonética forense e na identificação de falantes, o disfarce tem merecido atenção de alguns pesquisadores, embora aparentemente não seja de ocorrência muito comum. Segundo esse autor, 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, ou seja, apenas a

⁸ “Naive speaker recognition in the forensic arena often goes under the heading ‘Earwitness evidence’”.
⁹ “[...] also reveals information which our auditory system obscures”.

habilidade do próprio falante para alterar sua voz. Esse foi o recurso utilizado pelos participantes desta pesquisa.

3 MÉTODO E MATERIAIS

O estudo que vem sendo realizado por nosso grupo de pesquisa¹⁰ está analisando amostras de fala de 50 participantes. As gravações foram realizadas no laboratório de fonética da universidade, e os participantes leram um texto de 75 palavras, simulando uma ligação telefônica. A leitura foi feita, primeiramente, com a voz normal e, em seguida, com voz disfarçada. O disfarce utilizado foi de escolha livre de cada participante.

Os participantes foram divididos em 5 grupos de 10 pessoas, sendo nomeados: GH (grupo de homens entre 30-55 anos), GR (grupo de rapazes entre 19-25 anos), GM (grupo de mulheres entre 30-55 anos), GG (grupo de garotas entre 19-25 anos) e GC (grupo controle, mulheres entre 25-55 anos). O Grupo Controle é formado pelas pesquisadoras do grupo e é chamado de controle porque se inferiu que seja um grupo que teria maior eficiência na realização dos disfarces, pelos estudos que realizam.

O texto lido simulava um pedido de resgate em um sequestro e, para a análise apresentada neste artigo, apenas um trecho foi selecionado. A coleta de dados foi conduzida em ambiente com tratamento acústico, e os dados foram amostrados a uma taxa de 44100 Hz. Usou-se um computador Pentium Dual Core 5.300 2.60 GHz, 1.99 Gb RAM Processador XP 2002 Service Pack 3, placa de som externa M-AudioFastTrack Pro 4x4, Microfone AKG C 3000 B. O software Audacity foi utilizado para fazer as gravações.

O texto foi elaborado para uma pesquisa maior de análise de vogais em sílaba tônica, por isso as palavras objeto do estudo são paroxítonas contendo as sete vogais do português entre as plosivas¹¹ /p/ /t/ /k/ (Figura 1). Vários experimentos estão sendo realizados, como por exemplo, análise de formantes, de intensidade, de duração da vogal, de qualidade vocal, de frequência fundamental, sempre comparando voz normal e voz disfarçada.

¹⁰ O grupo é formado por professores, acadêmicos e peritos, das áreas da linguística, engenharia e fonoaudiologia.

¹¹ A decisão pela escolha de palavras com vogais entre plosivas se deu pela maior facilidade de medição de duração das vogais, posto que essas consoantes facilitam a visualização dos pontos inicial e final de cada vogal, tanto nos oscilogramas como nos espectrogramas.

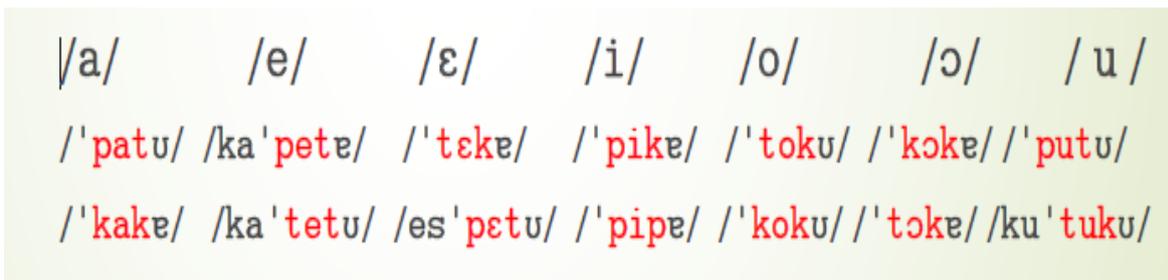


Figura 1: Vogais do português brasileiro analisadas durante a pesquisa.

O experimento objeto deste artigo, particularmente, tem como objetivo a análise da qualidade da voz comparando a voz normal com a disfarçada, com foco na frequência fundamental. Os grupos-alvo desta análise foram GM, GG e GC, em que se obteve a análise do disfarce de vozes femininas. O GG é formado por estudantes universitárias; o GM, por estudantes e professoras universitárias; e o GC, por mulheres participantes do grupo de estudo, estudantes, professoras universitárias e peritas criminais.

É importante deixar claro que aqui se demonstram apenas os primeiros resultados do estudo em que se busca comparar uma análise perceptual com os gráficos da curva de f_0 de um trecho escolhido aleatoriamente: “*estamos com o pato do teu marido na toca e vamos matar ele, picar e tacar dentro de uma garrafa de coca*”. A extração das medidas de f_0 e dos gráficos das curvas da voz normal e disfarçada foi realizada no software PRAAT¹² e foi considerada apenas a primeira passagem de cada modalidade (normal e disfarçada), pois, de acordo com Nolan (1995, p. 132), “não há dois enunciados idênticos, mesmo que sejam ditos pela mesma pessoa”.¹³

4 PROCEDIMENTOS E RESULTADOS

A análise perceptual foi realizada por três integrantes do grupo de estudo, que deveriam avaliar a eficiência do disfarce. Uma vez que as vozes gravadas eram de pessoas conhecidas dos três avaliadores (os pesquisadores mais um membro do grupo), as inferências se deram pelo reconhecimento ou não da pessoa com o disfarce. Primeiro, a avaliação foi feita individualmente, e depois os avaliadores chegaram a um acordo quanto ao nível atribuído: eficiente e não eficiente.

¹² Paul Boersma e David Weenink (autores) – Universidade de Amsterdã. Disponível em: <http://www.fon.hum.uva.nl/praat/> acesso em novembro de 2012.

¹³ “No utterances are identical, even if they are by the same speaker”.

Após análise perceptiva, chegou-se à conclusão de que grande parte dos disfarces se apresentou eficiente - das 30 vozes analisadas (Tabelas 1, 2 e 3), quatorze não foram reconhecidas no disfarce, sendo seis delas de informantes do GC.

Para a análise da qualidade vocal, foi utilizado o protocolo *Voice Profile Analysis Scheme Profile* (VPAS) criado por Laver (1980) e adaptado para o português brasileiro – BP-VPAS (CAMARGO; MADUREIRA, 2008). Esse protocolo permite realizar uma análise perceptual das configurações de qualidade de voz, tais como elementos do trato vocal (por exemplo, mandíbula aberta, protrusão labial e língua recuada); e elementos fonatórios (por exemplo, falseto, voz áspera e sussurro). É importante deixar claro que apenas os itens do protocolo foram utilizados para essa análise, isto é, não se considerou a graduação, que vai de 1 a 6, sendo que de 1 a 3 é considerado moderado e de 4 a 6 é considerado extremo (CAMARGO; MADUREIRA, 2008).

Na análise da qualidade vocal, várias foram as estratégias de disfarce detectadas pelo grupo avaliador, desde ajustes do trato vocal (como lábios protraídos ou mandíbula aberta), até elementos fonatórios (como falseto ou voz áspera). Essa etapa foi realizada por cinco integrantes do grupo de estudo, sob a orientação de uma fonoaudióloga. A exemplo da análise da eficiência, essa avaliação também foi feita, em primeiro lugar, individualmente. Depois, os avaliadores compararam os ajustes encontrados nas vozes e chegaram a um acordo para uma única avaliação final.

Para as medidas de f_0 , considerou-se o valor médio, ou seja, o valor central exibido na janela do espectrograma com valores de *pitch* no PRAAT, conforme Figura 2.

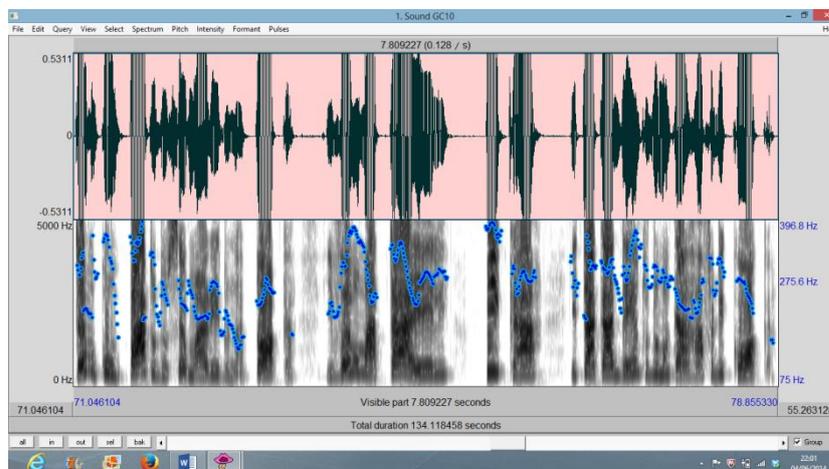


Figura 2: Trecho de voz disfarçada da informante GC10 – valor médio do *pitch* de 275 Hz.

As Tabelas 1, 2 e 3 mostram os resultados das duas análises perceptivas – de eficiência de disfarce e da qualidade vocal, assim como os valores médios da frequência fundamental, na voz normal e na voz disfarçada, e a diferença entre esses valores.

Grupo Mulheres (GM)					
	Eficiência do disfarce – percepção	Tipo de ajuste percebido	F₀ normal (Hz)	F₀ disfarce (Hz)	Diferenças de F₀ Disfarce – normal (Hz)
GM1	Não eficiente	Aumento de <i>pitch</i>	194	228	+34
GM2	Eficiente	Voz nasal	223	234	+11
GM3	Eficiente	Hiperfunção laríngea Mandíbula aberta	220	222	+2
GM4	Eficiente	Lábios protraídos <i>Pitch</i> , baixo	217	201	-16
GM5	Eficiente	Elevação e hiperfunção laríngea	220	183	-37
GM6	Não eficiente	<i>Pitch</i> + baixo	234	198	-36
GM7	Não eficiente	Apertou o nariz	265	278	+13
GM8	Não eficiente	<i>Pitch</i> + baixo	240	223	-17
GM9	Não eficiente	Aumento de <i>pitch</i> Mudou sotaque	242	299	+57
GM10	Não eficiente	Língua recuada Sotaque estrang.	213	234	+21

Tabela 1: Valores da Frequência Fundamental em Hz do Grupo Mulheres.

Grupo de Garotas (GG)					
	Eficiência do disfarce – percepção	Tipo de ajuste percebido	F₀ normal (Hz)	F₀ disfarce (Hz)	Diferenças de F₀ Disfarce – normal (Hz)
GG1	Não eficiente	Leve aumento do <i>pitch</i>	280	295	+15
GG2	Eficiente	Lábios protraídos <i>Pitch</i> + baixo Usou pano na boca	251	218	-33
GG3	Não eficiente	Aumento de <i>pitch</i> Denasal	282	302	+20
GG4	Não eficiente	<i>Pitch</i> + baixo	254	231	-23
GG5	Não eficiente	Lábios protraídos	232	236	+4
GG6	Não eficiente	Lábios protraídos <i>Pitch</i> +baixo	271	254	-17
GG7	Eficiente	Avanço de língua Aumento de <i>pitch</i>	269	334	+65
GG8	Eficiente	Tensão laríngea Mudou sotaque	249	334	+76
GG9	Eficiente	Voz áspera	214	187	-27
GG10	Não eficiente	Lábios protraídos <i>Pitch</i> + baixo	220	202	-18

Tabela 2: Valores da Frequência Fundamental em Hz do Grupo Garotas.

Grupo Controle (GC)					
	Eficiência do disfarce – percepção	Tipo de ajuste percebido	F ₀ normal (Hz)	F ₀ disfarce (Hz)	Diferenças de F ₀ Disfarce – normal (Hz)
GC1	Eficiente	Língua recuada Hiperf do trato vocal Sotaque estrang.	230	288	+58
GC2	Eficiente	Voz áspera	250	217	-33
GC3	Eficiente	Falsete	222	330	+108
GC4	Não eficiente	Corpo de língua rec. e abaixado	260	307	+47
GC5	Não eficiente	Aumento de <i>pitch</i> Voz infantil	213	304	+91
GC6	Não eficiente	Mandíbula aberta	226	238	+12
GC7	Não eficiente	Lábio prot. <i>Pitch</i> +baixo	242	215	-27
GC8	Eficiente	Voz nasal Hiperf.e elevação da laringe	238	364	+126
GC9	Eficiente	Aumento de <i>pitch</i> Voz nasal	262	371	+109
GC10	Eficiente	Língua recuada Aumento de <i>pitch</i>	231	275	+44

Tabela 3: Valores da Frequência Fundamental em Hz do Grupo Controle.

No GM (Tabela 1), quatro das dez participantes foram consideradas eficientes em seus disfarces, duas com elevação de *pitch* e outras duas com diminuição do *pitch*. No GG (Tabela 2), também quatro foram eficientes, da mesma forma, duas delas aumentando o *pitch* e duas, baixando-o. No GC (Tabela 3), seis foram eficientes, cinco delas com aumento de *pitch*.

Como se pode observar a partir das tabelas, a maioria dos disfarces considerados mais eficientes foi percebida pela variação de *pitch*, seja por voz mais grave ou mais aguda, em conjunto com algum outro ajuste, como protrusão labial ou hiperfunção laríngea, por exemplo. As estratégias comuns nos disfarces considerados eficientes foram voz nasal, hiperfunção laríngea e falsete, essas três com aumento de *pitch*, e lábios protraídos e voz áspera, essas duas com diminuição do *pitch*.

Outro fato que se verificou ao analisar os dados presentes nas tabelas é o de que o grupo controle optou por elevar o *pitch*, ou seja, teve maior uso da voz aguda, tendo apenas uma participante com o *pitch* diminuído (voz mais grave). Dentro desse grupo, houve maior número de disfarces percebidos como eficientes. A hipótese inicial do grupo era a de que as mulheres que optassem pelo aumento do *pitch* obteriam também mais eficiência no disfarce, pois, segundo Künzel (2000), a mulher

tem uma tendência maior a elevar o *pitch*. A outra hipótese, de que o grupo controle teria mais eficiência no disfarce, também foi confirmada.

A outra análise realizada foi a dos gráficos da curva de f_0 . Na análise dos gráficos, pôde-se constatar e observar visualmente a diferença entre as duas modalidades de voz, considerando a frequência fundamental, em que a linha preta representa a voz normal e a linha vermelha, a voz disfarçada (exemplos nas Figuras 3 e 4).

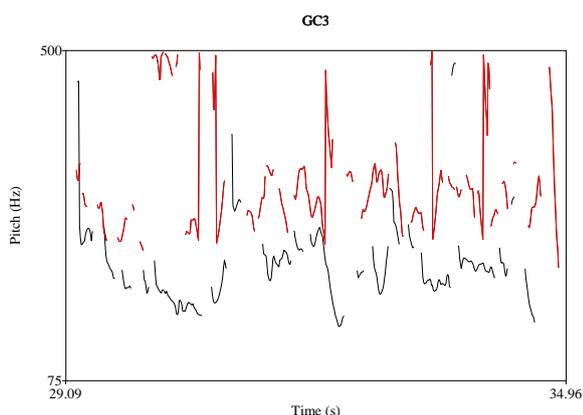


Figura 3: Disfarce eficiente.

Voz tornada aguda e com grande variação de f_0 .

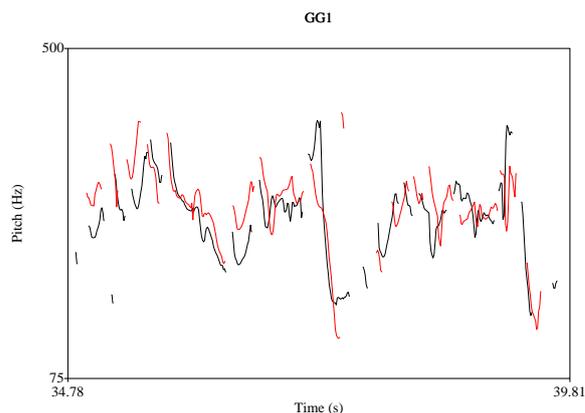


Figura 4: Disfarce pouco eficiente.

Voz ligeiramente agravada com ligeiro aumento de intensidade.

Nota-se que, na Figura 3, o sinal da fala normal está distante do sinal da fala disfarçada, não possuindo quase nenhum ponto em comum e com diferença considerável na curvatura dentro do gráfico. Já na Figura 4, tanto o sinal da fala disfarçada quanto da fala normal possuem quase os mesmos pontos de início e fim, e mesmas curvaturas. Essa análise visual foi comparada com as análises auditivas e numéricas da f_0 . Como se pode ver, as diferenças nas curvas de GC3 na Figura 3 combinam com a avaliação de disfarce eficiente na Tabela 3, enquanto as de GG1 na Figura 4, com a de disfarce não eficiente na Tabela 2.

DISCUSSÃO E CONSIDERAÇÕES FINAIS

A proposta deste artigo foi a de analisar a eficiência do disfarce em vozes femininas tendo como parâmetro a frequência fundamental. Como já dito anteriormente, a análise perceptual foi feita dentro de um consenso entre os avaliadores, integrantes do grupo de estudo. Dentre as 30 participantes analisadas, 14 foram consideradas eficientes no disfarce, tanto na análise perceptual quanto na

visual com a ajuda dos gráficos. Até o momento, ao se comparar esses grupos de mulheres, o experimento mostrou que, quando se realiza uma maior mudança da f_0 , combinada com algum outro tipo de ajuste, obtém-se uma maior eficiência no disfarce. Segundo Künzel (2000, p. 173), “as formas de disfarce que afetam características da frequência fundamental ocorrem com mais frequência do que outros tipos de disfarces em situações forenses. Os falantes, obviamente, as consideram fáceis de produzir, eficientes o bastante para transmitir a mensagem linguística e, ao mesmo tempo, garantir o anonimato”.¹⁴ Isso revela a importância do estudo da frequência fundamental para o desenvolvimento da pesquisa na área forense.

Os recursos utilizados pelos falantes não foram analisados de forma detalhada neste estudo, ainda que se tenha analisado a qualidade vocal por meio do roteiro BP-VPAS. A análise se concentrou apenas na identificação dos ajustes utilizados nos disfarces, sem avaliar os graus de escala de cada um deles (CAMARGO; MADUREIRA, 2008). Mas já foi possível combinar duas técnicas de análise, a auditiva com a acústica, conforme recomenda Nolan (1999). Segundo esse autor, os foneticistas e dialetologistas podem detectar diferenças sutis de pronúncia numa análise auditiva, e com a análise acústica podem quantificar os aspectos da fala reconhecidos pela audição.

Futuros experimentos ainda levarão em conta uma comparação entre grupos de homens e mulheres, para comparação de estratégias de disfarce selecionadas por ambos os sexos, pois Künzel (2000) afirma que há diferenças nas preferências por modos de disfarce, assim como por estratégias articulatórias, entre homens e mulheres. Além de comparar as diferenças entre as mulheres e os homens, pretende-se também fazer uma comparação entre os grupos mais jovens e mais velhos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. BRAID, Antonio Cesar Morant. *Fonética Forense: tratado de perícias criminalísticas*. Campinas: Millenium, 2003.

¹⁴ “[...] disguise modes affecting fundamental frequency characteristics occur more often than other disguise in forensic situations. Speakers obviously find them easy to produce, efficient enough to transmit the linguistic message and, at the same time, to ensure their anonymity”.

2. CAMARGO, Zuleica, MADUREIRA, Sandra. *Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS)*. Campinas, 2008.
3. CARNEIRO, Denise de Oliveira, GOMES, Maria Lúcia de Castro. *A fonética forense no Brasil: cenários e atores (no prelo)*.
4. ERIKSSON, Anders. The disguised voice: imitating accents or speech styles and impersonating individuals. In: Eriksson, Anders. *Language and Identity*. Edinburg: Edinburg University Press, 2010. p. 86-96.
5. HOLLIEN, Harry. *Forensic voice identification*. London: Academic Press, 2002.
6. KENT, Ray D., READ, Charles. *The acoustic analysis of speech*. Connecticut, USA: Thomson Learning, 2002.
7. KÜNZEL, Hermann J. *Effects on voice disguise on speaking fundamental frequency*. Birmingham: University of Birmingham Press, 2000.
8. KINOSHITA, Y., ISHIHARA, S., ROSE, P. *Exploring the discriminatory potential of F_0 distribution parameters in traditional forensic speaker recognition*. Equinox Publishing Online, 2009.
9. LAVER, John. *The phonetic description of voice quality*. Cambridge: Cambridge University Press, 1980.
10. NOLAN, Francis. Can the definition of each speaker be expected to come from the laboratory in the next decades? *Phonetic Sciences: Proceedings of the XIIIth International Congress of Phonetic Science, ICPhS95, Stockholm, Sweden, v. 2, pp. 130-137, 13-19 August, 1995*.
11. NOLAN, Francis. Speaker Recognition and Forensic Phonetics. In: HARDCASTLE, William J.; LAVER, John. (Eds.). *The Handbook of Phonetic Sciences*. The United States: Blackwell Publishing, 1999.
12. ROACH, Peter. *English Phonetics and Phonology: A practical course*. Cambridge: CUP, 2009.
13. TOSI, Oscar. *Voice identification: theory and legal applications*. Baltimore: University Park Press, 1979.

ABSTRACT: In order to evaluate vocal quality, it is possible to use two important resources: perceptual analysis and acoustic analysis. The latter is objective and uses computerized programs that quantitatively show different measurable aspects of the captured voice signal. One of these aspects is the fundamental frequency (f_0). In this work, the efficiency of voice disguise was analyzed using three aspects: perceptual analysis, vocal quality and the curve of f_0 from part of a speech (in normal and

disguised voice) produced by women in three different groups with 10 participants each, labeled Girls Group, Women Group and Control Group. In the preliminary results, speakers who opted to use a low pitch in their voices did not acquire great efficiency in disguise. On the other hand, speakers who used a high pitch or used other types of adjustment acquired more efficiency in the change of f_0 , achieving higher efficiency in disguise.

Keywords: Acoustic analysis; Fundamental frequency; Disguised voice.

Recebido no dia 05 de junho de 2014.

Aceito para publicação no dia 18 de agosto de 2014.