

Análise perceptiva e acústica em fonética forense: uma pesquisa em disfarce de voz

Perceptive and acoustic analysis in forensic phonetics: research in voice disguise

Maria Lúcia de Castro Gomes*

Denise de Oliveira Carneiro**

Andrea Alves Guimarães Dresch***

RESUMO: O objetivo maior deste trabalho de pesquisa foi introduzir um grupo de alunos de graduação em Letras nas atividades de análise perceptiva e acústica em contexto forense. Para isso um texto de 75 palavras foi criado para simular uma situação de sequestro. Cinquenta participantes, divididos em grupos de dez de acordo com gênero e faixa etária, foram gravados lendo o texto em voz normal e, em seguida, disfarçando a sua voz. Para a análise perceptiva, foi utilizado o protocolo VPAS (LAVER, 1980; CAMARGO E MADUREIRA, 2008), realizada por seis pessoas (juízes). Para a análise acústica, usou-se o software PRAAT para medição de duração, F0, F1 e F2. Vozes femininas e masculinas foram comparadas e alguns resultados confirmam algumas tendências universais, como: a duração intrínseca das vogais, na qual as vogais mais baixas são mais longas que as vogais mais altas; informantes do sexo feminino apresentam valores de frequência de F0, F1 e F2 mais altas que os do sexo masculino; as mulheres apresentam espaço vocálico maior que os homens. Os ajustes mais frequentes para o disfarce de voz foram: a mandíbula aberta e a protrusão labial, como ajustes supralaríngeos, o falsete e a voz áspera, como ajustes de fonação.

PALAVRAS-CHAVE: Análise perceptiva. Análise acústica. Fonética forense. Disfarce de voz. Análise de vogais.

ABSTRACT: The main objective of this research was to introduce a group of undergraduate students in activities of perceptual and acoustic analysis in a forensic context. A 75-word text was created to simulate a kidnapping situation. Fifty participants, in groups of ten according to gender and age, were recorded reading the text in a normal voice and then disguising their voice. For the perceptive analysis, we used the VPAS protocol (LAVER 1980; CAMARGO AND MADUREIRA, 2008), and six researchers (judges) analyzed the voices. For the acoustic analysis, we used PRAAT for measuring duration, F0, F1 and F2. Male and female voices were compared and some results confirm some universal trends: vowel intrinsic duration, in which the lower vowels are longer than the higher vowels; female informants have F0, F1 and F2 higher frequency values than males; women have larger vowel spaces than men. The most common settings for voice disguise were: open jaw and lip protrusion, as supralaryngeal settings, falsetto and harsh voice, as phonation settings.

KEYWORDS: Perceptive analysis. Acoustic analysis. Forensic phonetics. Voice disguise. Vowel analysis.

* Doutora em Letras, Linguista, Professora no Departamento de línguas Estrangeiras Modernas – DALEM da UTFPR Câmpus Curitiba.

** Mestranda em Engenharia Biomédica, Fonoaudióloga, Perita Criminal no Instituto de Criminalística do Paraná – IC-PR.

*** Mestre em Engenharia Biomédica, Engenheira Eletricista, Perita Criminal no Instituto de Criminalística do Paraná – IC-PR.

1. Introdução

À medida em que as atividades e relações humanas adquirem maior complexidade, os estudos interdisciplinares tornam-se mais necessários. A linguística vem estabelecendo parcerias com diversas áreas de conhecimento para melhor compreensão dos variados processos que envolvem a linguagem. Para possibilitar melhor entendimento sobre a aquisição da linguagem, a linguística e a psicologia se associam e assim nasce a psicolinguística. Aliando-se à sociologia, a linguística consegue explicar fatos de variação e mudança na língua, e dessa forma surge a sociolinguística. Os pesquisadores do discurso vão buscar na psicanálise e no marxismo noções importantes sobre a constituição do sujeito e da produção de sentidos, e mais uma área importante na linguística é criada – a análise do discurso.

Com os avanços tecnológicos e o desenvolvimento da tecnologia da fala para, por exemplo, reconhecimento de voz e comparação de locutor, novas parcerias têm sido formadas. Os estudiosos da linguagem veem a necessidade, mais do que nunca, de ultrapassar suas áreas de pesquisa e trabalhar com estudiosos de outras áreas, como as engenharias, as ciências da computação, a fonoaudiologia, a psicologia.

Diversos ramos da linguística aplicada, a partir de tais parcerias, têm se dedicado ao estudo das particularidades da fala com objetivos diversos. Dentre essas áreas está a ciência forense, que tem por objetivo atribuir autoria a falas por meio de comparação. Segundo Braid (2003), “a verificação de locutor é a atividade pericial dentro da Fonética forense capaz de determinar se duas falas foram produzidas por um mesmo falante” (p.6). O exame consiste “na comparação de uma gravação de voz suspeita com um padrão de voz coletado do provável autor da voz suspeita” (ESPINDULA, 2005, p. 338).

Sob a ótica forense, a fala é analisada como um sinal biológico eletronicamente processado, armazenado em mídia. Isso significa que os estudos não podem ser restritos à produção da fala, mas devem se expandir ao processamento do sinal. Por isso, diferentes métodos são empregados para a análise, sem ainda haver um consenso sobre qual seria o mais eficaz, conforme pesquisa de Gold e French (2011), que agruparam os métodos em quatro possibilidades: análise perceptiva somente, análise acústica somente, combinação entre análise perceptiva e acústica, e análise automática para reconhecimento de locutor. Tem sido crescente o uso de sistemas automáticos e muita discussão sobre a utilização de tais sistemas tem acontecido, principalmente em eventos como a Conferência Anual da Associação para Fonética e Acústica Forense (IAFPA, 2015).

Este trabalho, porém, concentra-se nos métodos de análise perceptiva e acústica, escolhidos por um grupo de pesquisa organizado com o intuito de introduzir alunos de um curso de graduação em Letras nos mistérios da fonética forense. O objetivo da pesquisa foi comparar vozes de 50 falantes, distribuídos em grupos de acordo com gênero e faixa etária, em simulação de um telefonema para pedido de resgate durante uma situação de sequestro, em suas vozes normais e usando estratégia de disfarce.

O objetivo deste texto é, primeiramente, compor uma revisão dos conceitos básicos e fundamentais sobre qualidade de voz e sobre alguns dos principais parâmetros acústicos para análise de vogais: duração, primeiro e segundo formantes e frequência fundamental. Em segundo lugar, descreve-se a metodologia da pesquisa, os principais resultados da análise acústica, e a relação entre a análise perceptiva dos disfarces com a análise acústica das vogais.

2. Pressupostos Teóricos

Este trabalho teve três focos de análise: a qualidade vocal, os parâmetros acústicos das vogais do português brasileiro e o disfarce de voz.

2.1 Qualidade Vocal

Muito pode se dizer de uma pessoa a partir de sua voz. A voz pode demonstrar características físicas e psicológicas, estados emocionais e de humor. Dada a diversidade de perspectivas de análise, definir voz ou qualidade de voz tem sido objeto de discussão. A definição de qualidade de voz de Abercrombie (1967) é citada por Laver (1980, p.1) e por Kreiman e Sidtis (2013, p.8) e indica “as características que estão mais ou menos presentes o tempo todo em que a pessoa está falando; trata-se da qualidade quase permanente que perpassa todo som que sai de sua boca”¹.

Kreiman e Sidtis (2013) apontam para a dificuldade de atribuir uma definição de voz que sirva a todos os propósitos, e que possa servir a uma variedade de disciplinas. Segundo as autoras, a definição de voz pode ter um sentido estreito ou amplo. Um som produzido pela vibração das pregas vocais é uma definição estreita do termo voz, pois exclui os efeitos acústicos do trato vocal. Definir voz dessa forma aproxima-se ao que os linguistas definem como traço

¹ “those characteristics which are present more or less all the time that a person is talking; it is a quasi-permanent quality running through all the sound that issues from his mouth” (ABERCROMBIE, 1967, *apud* LAVER, 1980 e KREIMAN; SIDTIS, 2013). Todas as traduções neste artigo foram realizadas pela primeira autora.

de vozeamento, que faz com que /z/ seja diferente /s/, por exemplo. Nesse caso, “voz seria sinônimo do termo ‘fonte laríngea’, que enfatiza o fato de que as vibrações das pregas vocais são a fonte de energia acústica para grande parte da fala”². Num sentido amplo, voz seria sinônimo de ‘fala’, pois incluiria “os resultados acústicos da ação coordenada entre o sistema respiratório, língua, mandíbula, lábios e palato mole”³.

A *qualidade de voz*, para Kreiman e Sidtis (2013), também pode ser definida de forma estreita ou ampla, o que seriam na verdade dois lados de uma mesma moeda – de um lado, características da percepção e, de outro, características da produção. No sentido estreito, a qualidade de voz é vista como um aspecto único do processo de fonação, como a percepção da corrente de ar no sinal vocal. Num sentido não tão estreito, é tida como o resultado percebido do processo de fonação. E num sentido mais amplo, é a resposta do ouvinte ao som da fala, ou seja, conforme Denes e Pinson (1993, *apud* KREIMAN; SIDTIS, 2013), a qualidade de voz é o resultado final de uma sequência complexa de eventos acústicos, psicológicos e cognitivos, a conhecida ‘cadeia da fala’. O foco que se dá a essa cadeia de eventos depende do interesse do pesquisador ou profissional envolvido, se um cirurgião, um engenheiro, um psicólogo, cada um vai adotar uma definição diferente e ter uma preocupação específica.

Kreiman e Sidtis (2013) também discutem questões sobre a distinção entre produção e percepção da voz, que têm sido mantidas separadas nos trabalhos de pesquisa. No entanto, defendem uma perspectiva adotada em trabalhos mais recentes no dialogismo dentro da linguística, na qual percepção e produção são inseparáveis. A voz é um evento concreto produzido por uma pessoa num contexto comunicativo que, necessariamente, inclui o ouvinte, mas também manifesta a consciência abstrata e não observável do falante. As autoras concluem, afirmando que “a voz expressa quem nós somos, tanto de forma isolada como em relação a outros indivíduos”⁴.

De acordo com Behlau (2004), que também afirma que a terminologia é bastante imprecisa e confusa, a qualidade vocal é um termo empregado para denominar a impressão total criada por uma voz, é “nossa avaliação perceptiva principal”. Behlau e Pontes (*apud* BEHLAU,

² “Voice in this sense is also synonymous with the term ‘laryngeal source’, which emphasizes the fact that vocal fold vibrations are the acoustic energy source for much of speech” (KREIMAN; SIDTIS, 2013, p.5).

³ “the acoustic results of the coordinated action of the respiratory system, tongue, jaw, lips, and soft palate” (KREIMAN; SIDTIS, 2013, p. 6).

⁴ “voice expresses who we are, both in isolation and with respect to other individuals” (KREIMAN; SIDTIS, 2013, p.10)

2004) estipulam 23 tipos de vozes, das quais podem ser extraídas, perceptivamente, as características de fenômenos biológicos, psicológicos e socioeducacionais.

Rejeitando o sentido estreito de qualidade de voz, apenas como uma atividade laríngea, Laver (1980) adota um sentido amplo, como a ‘coloração característica auditiva’ da voz de uma pessoa. Para o autor, não somente as características laríngeas, mas também as supralaríngeas são importantes para a qualidade da voz. Inspirando-se em Abercrombie (1967, *apud* Laver, 1980) e o conceito de ajuste muscular (*muscular adjustment*), e em Honikman (1964, *apud* Laver, 1980), que traz o termo ajuste articulatorio (*articulatory setting*), Laver propõe uma taxonomia descritiva baseada em componentes fonéticos da qualidade de voz. O sistema descritivo proposto apresenta-se com uma fundamentação de base auditiva, mas que se correlaciona com outros níveis de análise, todos passíveis de verificação instrumental, quais sejam, os níveis articulatorio, fisiológico e acústico.

A perspectiva fonética da proposta de Laver tem como resultado o roteiro VPAS - *Vocal Profile Analysis Scheme* que, segundo Camargo e Madureira (2008), oferece a possibilidade de estudos a partir de uma noção básica: a plasticidade do aparelho vocal. As autoras adaptaram o protocolo de Laver para o contexto brasileiro, o BP-VPAS (*Brazilian Portuguese Vocal Profile Analysis Scheme*).

Esse protocolo oferece uma perspectiva a partir de unidades analíticas que, segundo Camargo e Madureira (2008), são tendências musculares de longo termo, os chamados ajustes. Tais ajustes podem ser de ordem fonatória, definidos pela vibração das pregas vocais (como o falsete e a voz áspera), ou de ordem supraglotal ou articulatoria (como a protrusão labial ou mandíbula aberta) ou, ainda, de tensão, seja laríngea ou do trato vocal. Segundo as autoras, os ajustes são definidos a partir de variações de um ponto neutro de referência, em termos de forças de adução ou tensão longitudinal.

O uso do protocolo BP-VPAS para avaliação de voz permite também a classificação por níveis, escalados de 1 a 6, de moderado a extremo. Como no âmbito forense, a avaliação não é ensejada por queixas vocais ou alterações trazidas pelos locutores, nesta pesquisa, o modelo foi utilizado apenas como roteiro para identificação dos ajustes utilizados nos disfarces de voz, sem considerar-se os níveis escalares, conforme será descrito na seção destinada à metodologia.

Defende-se que a avaliação da qualidade vocal deva ser predominantemente perceptiva. Kreiman e Gerratt (1993) apontam que a qualidade de voz é, por natureza, perceptual, por isso a área clínica valoriza muito mais as medidas perceptivas do que as medidas instrumentais. No

entanto, essa avaliação é cunhada em subjetividade, gerando discordância entre os ouvintes e dificuldade em assumir um consenso em torno do uso desta ou daquela terminologia. Também se defende que a avaliação forense da qualidade vocal, assim como a avaliação clínica, deva ser predominantemente perceptivo-auditiva e, em caso de dúvidas, proceda-se à busca de fenômenos que possam ancorar ou refutar os achados por meio da avaliação acústica. A soberania da avaliação perceptivo-auditiva é citada por pelo menos dois autores, pesquisadores de parâmetros acústicos, indicando a necessidade de corroborar os achados de modo mais objetivo. (DAJER, 2010; MASTER, 2005).

Partimos, então, para a descrição de alguns parâmetros acústicos importantes em análises de comparação de locutor, que serão correlacionados com a análise perceptiva da qualidade de voz.

2.2 Teoria da produção da fala

A teoria acústica da produção da fala descrita por Fant (1960) é conhecida como teoria fonte-filtro, segundo a qual o sinal da fala é resultante da combinação de uma fonte de energia sonora e de filtros resultantes da ressonância da energia da fonte nas cavidades supraglotais. Segundo tal modelo linear (KENT; READ, 2002), a fonte de energia acústica (dos sons vozeados) é proveniente dos pulsos glóticos, gerados a partir da vibração das pregas vocais situadas na laringe, em movimentos sucessivos de abertura e fechamento. A frequência de tal vibração é medida em Hertz (Hz), constituindo o valor da frequência fundamental (F0) e tem relação direta com a massa, tamanho e geometria das pregas vocais (GILLIER, 2011). As pregas vocais em vibração produzem um espectro, que de acordo com Kent e Read (2002), poderia ser idealizado por frequências discretas, múltiplas da frequência vibratória fundamental, conforme demonstrando na Figura 1.

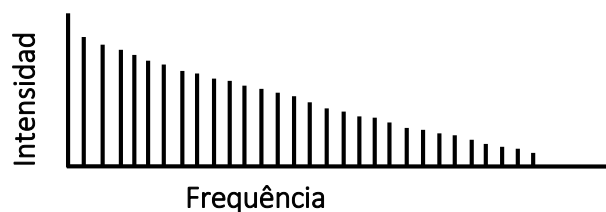


Figura 1: espectro idealizado, no qual a energia está distribuída em frequências discretas em múltiplos inteiros da frequência fundamental (KENT; READ, 2002, p. 23).

Conforme já mencionado acima, as cavidades do trato vocal supralaríngeo atuam como ressoadores e atenuadores (filtros) da energia proveniente da fonte. As frequências realçadas são os chamados formantes. Os “formantes consistem em realces de energia acústica concentrada que representa a ressonância em um tubo acústico (modelo proposto pela teoria fonte-filtro). Sendo relacionados à anatomofisiologia do trato vocal, suas medidas constituem dados relevantes para a individualização de um sinal” (GILLIER, 2011). A partir da fonte produtora da voz e da filtragem pelos articuladores, produzem-se os segmentos que formam as palavras: as vogais e as consoantes. Neste trabalho, o foco se dá nos segmentos vocálicos.

2.3 Vogais e parâmetros acústicos

Segundo Camara Jr. (2002), a divisão mínima na fonologia, a chamada segunda articulação⁵ da língua, “é a dos sons vocais elementares, que podem ser vogais ou consoantes” (p. 33). O autor afirma que há dois critérios para diferenciar vogais de consoantes, um de caráter mais fonético, que considera a vogal como um som produzido pela ressonância do trato vocal, com passagem livre do ar pela boca. Na mesma linha, Marchal e Reis (2012) explicam que “a vogal é produzida com um canal aerífero aberto, sem constricção maior e na ausência de geração de ruído de fricção” (p. 142). Já o segundo critério refere-se ao fato de a vogal constituir-se como núcleo de uma sílaba. Embora as consoantes nasais e líquidas também possam figurar como centro de sílaba, no português apenas vogais podem estar no centro ou ápice, ficando as consoantes como elementos marginais (CAMARA JR., 2002; MARCHAL; REIS, 2012).

As vogais, tidas como os sons mais simples de se analisar e descrever (KENT; READ, 2002), têm como parâmetros principais para descrição acústica a duração, a frequência fundamental, o padrão formântico e o espectro (este não considerado nesta pesquisa).

2.3.1 Duração

Um dos importantes parâmetros acústicos para análise de vogais é a duração, que pode figurar como elemento distintivo nas línguas. No entanto, mesmo na inexistência desse tipo de distinção, as vogais variam substancialmente em sua duração. Lindblom (1967) aborda a variação sistemática da duração da vogal como um problema da fonética clássica. O problema

⁵ André Martinet, linguista do Círculo Linguístico de Praga propôs a teoria da dupla articulação da linguagem, segundo a qual a morfologia é a primeira articulação e a fonologia, a segunda articulação (PIETROFORTE, 2005).

consiste na composição de fatores que exercem influência na duração, quais sejam, a característica da vogal em si e o ambiente consonantal adjacente. No primeiro caso, fatores aerodinâmicos são responsáveis pela diferença – vogais mais abertas tendem a ser mais longas que vogais mais fechadas (duração intrínseca). Quanto ao ambiente fonológico, diz-se que uma vogal é geralmente mais longa antes de consoante vozeada do que precedendo uma consoante não vozeada.

Kent e Read (2002) ilustram esses dois fatores com as palavras *bet*, *bed* e *bad*, da língua inglesa, conforme Figura 2. A vogal em *bed* é mais longa que a vogal em *bet*, pelo vozeamento de /d/, e a vogal de *bad* é mais longa que a de *bed* porque /æ/ é mais baixa, ou seja, mais aberta que /ɛ/.

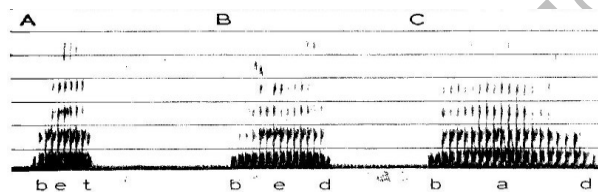


Figura 2: ilustração em espectrogramas de variação na duração da vogal. A – [bet]; B – [bed]; C – [bæd] (KENT; READ, 2002, p. 127).

Estudos de Keating (1985), com o uso de eletromiografia, atestam que línguas como o chinês, o tcheco e o árabe não apresentam essa característica, ou seja, as vogais não se alongam diante de consoantes vozeadas. A autora defende, por conta disso, que essa é uma característica específica de determinadas línguas.

Quanto à duração intrínseca das vogais, Escudero *et al* (2009) analisaram as sete vogais orais do português brasileiro e europeu e concluíram com confiança que as vogais mais baixas são mais longas que as vogais mais altas na língua portuguesa. Afirmam também que o ouvinte de língua portuguesa usa a duração como pista para a identidade da vogal em maior grau do que ouvintes de outras línguas.

Se a duração intrínseca da vogal tem relação com abertura que, segundo Marchal e Reis (2012), para as vogais “descreve normalmente a distância que separa o ponto mais elevado da cúpula da língua até o palato” (p. 137) e que com esse parâmetro é possível distinguir vogais mais altas de vogais mais baixas, como /i/ de /e/, por exemplo, conclui-se que a duração tem estreita relação com o primeiro formante.

2.3.2 Frequência fundamental

A frequência fundamental ou F0 corresponde à taxa de vibração das pregas vocais e é, como visto acima, a fonte de energia para a produção da voz. A F0 é o correlato acústico da vibração das pregas vocais, enquanto o *pitch* é o seu correlato perceptivo (KREIMAN; SIDTIS, 2012). É muito comum, no entanto, o uso do termo *pitch* para se referir à frequência fundamental (NOOTBOOM, 1997), ou o tratamento dos dois termos como sinônimos, mas na verdade, a frequência fundamental é a pista mais importante para a percepção do *pitch* (KREIMAN; SIDTIS, 2012).

Assim como a frequência de formantes, a frequência fundamental também é determinada por questões anatômicas, constituindo em fator importante na caracterização de locutores. Como a F0 de cada indivíduo depende do tamanho e massa das pregas vocais, os fatores sexo e idade também são cruciais. Nootboom (1997) reporta que, para os homens, os valores ficam entre 80 e 200 Hz, para as mulheres, entre 180 e 400 Hz, e para crianças os valores são consideravelmente mais altos. Em pesquisa com o português brasileiro, Madazio (2009) reporta que “para os homens, a faixa da frequência fundamental varia de 80 a 150Hz; e, para as mulheres, de 150 a 250Hz. Para falantes do português de São Paulo, a frequência fundamental média encontrada foi de 113Hz para homens e 205Hz para mulheres” (p. 37). Essa autora ainda argumenta que a F0 é um parâmetro robusto por ser resistente aos diferentes sistemas de análise acústica e menos sensível aos meios de gravação.

Rose (2002) afirma que a frequência fundamental é uma medida extremamente importante para a fonética em geral e para a fonética forense, em particular. Kreiman e Sidtis (2012) também argumentam sobre a importância da F0, afirmando que “a F0 apresenta correlações com as características físicas e o estado interno de um indivíduo, e ainda pode ser muito bem controlada e bastante saliente aos ouvintes”⁶. Essa combinação de fatores, segundo as autoras, é ideal para fornecer informações sobre um locutor, além de ser um parâmetro robusto mesmo com ruído no ambiente. Rose (2002) afirma ainda que a F0 pode ser extraída mesmo de gravações de má qualidade. Além de Rose, outros autores defendem a F0 como um parâmetro robusto para o trabalho de reconhecimento de voz e comparação de locutor (NOLAN, 1983; SAMBUR, 1975; JIANG, 1996, *apud* KINHOSHITA, 2009). Mas também,

⁶ “F0 displays correlations with an individual’s physical characteristics and internal state, but is also finely controllable and highly salient to listeners” (KREIMAN; SIDTIS, 2012, p. 139)

há argumentos em contrário. Kinoshita (2009), por exemplo, não considera a F0 um parâmetro eficaz, por apresentar razão de variância bastante fraca.

Em fonética forense, além da F0, os formantes das vogais também se destacam como parâmetros de grande importância.

2.3.3 Padrão formântico das vogais

Stevens (1997) afirma que “as frequências dos formantes, em particular os dois primeiros, F1 e F2, são dependentes do formato do espaço entre a glote e os lábios, e esse formato, por sua vez, é determinado pela posição do corpo da língua e dos lábios”⁷. O valor de F1 varia com a elevação e abaixamento da língua, ou seja, com o movimento vertical, enquanto F2 varia com o movimento horizontal, para frente e para trás. Kent e Read (2002) advertem para o uso dessa regra porque há exceções, no entanto, descrevem vários experimentos que confirmam a precisão dessa escala multidimensional (RAKERD; VERBRUGGE, 1995; PETERSON-BARNEY, 1952; HILLENBR *et al*, 1995, *apud* KENT e READ, 2002).

Essa relação das medidas acústicas com movimentos articulatorios tem sido simbolizada, a partir de trabalhos de Bell (1879) e Jones (1862), em um espaço vocálico, representado por um trapézio que imita o espaço do trato bucal. A delimitação do espaço é feita pelas chamadas vogais cardeais, que indicam os limites das pronúncias vocálicas, conforme Figura 3. Segundo Marchal e Reis (2012), “o sistema das vogais cardeais representa um instrumento para descrição das vogais com base articulatória e auditiva” (p. 145).

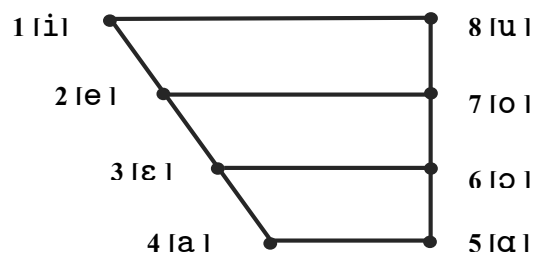
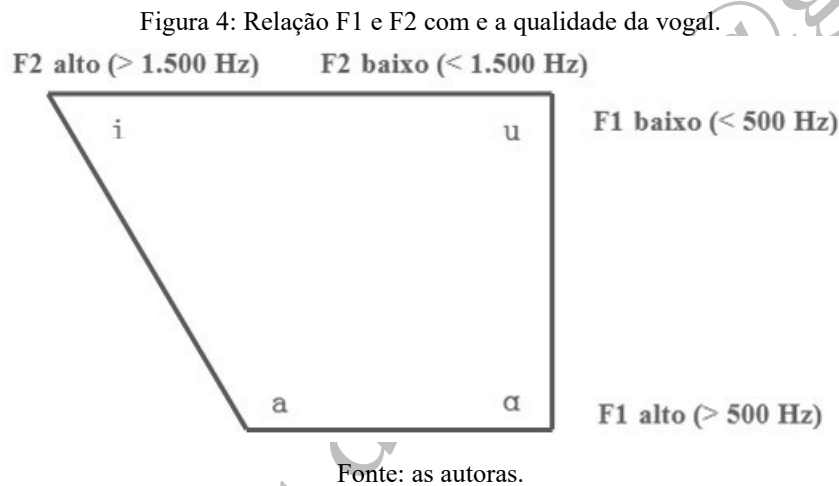


Figura 3: Vogais cardeais primárias (MARCHAL; REIS, 2012).

⁷ “The frequency of the formants, particularly the first two formants F1 and F2, are dependent on the shape of the airway between the glottis and the lips, and this shape in turn is determined by the position of the tongue body and the lips” (STEVENS, 1997, p. 472)

A posição das vogais no trapézio indica a posição da língua durante pronúncia da vogal e, acusticamente, as constrictões definem as frequências dos formantes (medidas em Hz). Conforme já mencionado, o primeiro formante está relacionado com o levantamento e abaixamento da língua, concomitante à abertura e fechamento de mandíbula. A elevação do corpo da língua diminui o valor de F1, enquanto que o abaixamento da língua aumenta o valor de F1. O segundo formante, por sua vez, tem relação com a anterioridade ou posterioridade da língua. Se o corpo da língua estiver anteriorizado, o valor de F2 será alto; se a língua estiver recuada, o valor será baixo. A Figura 4 ilustra essa relação.



Os valores dos formantes dependem da geometria do trato vocal e o cálculo se baseia em modelos de tubo de ressonância (BRAID, 2003). O comprimento do tubo determina a frequência da ressonância, ou seja, dos formantes. Quanto mais longo o tubo de ressonância, menor os valores de frequência de formantes. Isso significa que os valores das frequências de formantes variam de acordo com as características dos falantes, e os dois fatores principais são sexo e idade (KENT; READ, 2002). Como exemplo, podemos citar valores médios encontrados por Escudero *et al* (2008) para a vogal /ε/ do português brasileiro: nas vozes femininas, F1: 646 Hz, F2: 2.271 Hz, F3: 2.897 Hz; nas vozes masculina, F1: 518 Hz, F2: 1.831 Hz, F3: 2.572 Hz. Vale mencionar, ainda, que existe um número infinito de formantes. Duckworth *et al* (2011) reportam sobre diversos autores que demonstraram que os formantes mais altos são mais relacionados a características dos falantes, o que em tese seria muito interessante para análise em contexto forense. No entanto, o sinal analisado em situações forenses é comumente originado de interceptações telefônicas e a largura da banda desse canal deve ser considerada.

Embora os formantes mais altos sejam ricos em informações que levem a particularização, essas frequências não são transmitidas pelo canal telefônico. Kunzel (2001) cita que a faixa a ser considerada encontra-se entre 350-3400 Hz. Assim, serão considerados, nesta pesquisa, apenas o primeiro e o segundo formantes.

2.4 Disfarce de voz

O uso de disfarce de voz em fonética forense tem chamado a atenção de alguns pesquisadores (MASTHOFF, 1996; KÜNZEL, 2000; GILLIER, 2001) embora, segundo Eriksson (2010), não seja muito comum sua utilização na prática de delitos. De acordo com esse autor, um disfarce pode causar sérios problemas para a identificação de locutor, especialmente se for realizado com o uso de equipamentos eletrônicos.

É importante considerar que o estudo do disfarce não é exaurido nele próprio: pode trazer informações importantes sobre o comportamento dos parâmetros acústicos quando submetidos a condições fisiologicamente distintas. Por exemplo, se um sujeito realiza como disfarce a voz soprosa, a comparação de sua voz normal com sua voz disfarçada exibirá quais medidas são alteradas e de que forma, possibilitando correções em casos específicos. Essa observação carece, ainda, de maiores estudos.

Para reconhecer/detectar um disfarce, é necessário ter conhecimento da voz natural do falante, ou não será possível afirmar se uma característica especial do dado de fala é disfarçada ou natural (KÜNZEL, 2000). Esse autor analisou a preferência de homens e mulheres no uso de disfarce. Os resultados de sua pesquisa demonstram que pessoas, ao disfarçar a voz, variam bastante a frequência fundamental. Os sujeitos com F0 mais alta que a média tendem a elevá-la ainda mais. O contrário também é verdadeiro, ou seja, pessoas com F0 mais baixa que a média, no disfarce, são percebidas como apresentando *pitch* mais baixo. Sua pesquisa demonstra também que as mulheres são mais relutantes que os homens em alterar drasticamente a F0. A preferência de tipos de disfarce já tinha sido abordada na pesquisa realizada por Masthoff (1996), a qual revelou a preferência por alterações na fonação. Revelou também que as alterações realizadas no disfarce afetavam no máximo dois parâmetros fonéticos, deixando muitos outros aspectos do comportamento vocal disponíveis para análise forense.

Hollien (2002) analisou a efetividade do disfarce no contexto de identificação de locutor. Segundo o autor, por exemplo, o sussurro pode eliminar ou reduzir informações sobre

a frequência fundamental e, conseqüentemente, a percepção do *pitch*, dificultando a identificação.

Gillier (2011) também pesquisou sobre disfarce de voz em fonética forense, analisando não somente a F0, mas também as frequências de formantes. Alguns dos pontos importantes no relato da autora são: os dois parâmetros são eficientes para discriminação de indivíduos; nem todos os disfarces são eficazes, pois não alteram as frequências dos parâmetros; o efeito do disfarce não é homogêneo entre as várias vogais, sendo possível recuperar marcas específicas de cada falante através dos triângulos vocálicos⁸.

Esses trabalhos, em suma, revelam que o estudo de disfarce é relevante em pesquisa de comparação de locutor. No entanto, o objetivo maior da decisão do nosso grupo de pesquisa no uso de disfarce foi obter parâmetros de comparação de voz para treinamento em análise perceptiva e acústica.

3. Metodologia

Conforme já exposto, o objetivo inicial deste projeto de pesquisa foi a realização de um trabalho multidisciplinar para inserir os alunos de um curso de Letras em uma pesquisa em fonética forense. Os objetivos específicos deste recorte do trabalho foram (i) analisar a qualidade vocal e a estratégia de disfarce de acordo com o protocolo VPAS (LAVÉRE, 1980) – VPAS PB (CAMARGO; MADUREIRA, 2008); (ii) examinar quatro correlatos acústicos das sete vogais do português: duração; frequência fundamental; primeiro formante – F1, segundo formante – F2 (ESCUDEIRO *et al*, 2009) em voz normal e disfarçada; (iii) comparar vozes femininas com vozes masculinas; (iv) relacionar os espaços vocálicos verificados nos trapézios com as estratégias de disfarce percebidas pelos juízes.

3.1 Informantes, texto e gravação

Para o projeto de pesquisa como um todo, os informantes foram reunidos em cinco grupos de 10 pessoas cada, de acordo com gênero e faixa etária, com as seguintes características: mulheres pesquisadoras com idade entre 25 e 55 anos (média de 37), chamado Grupo Controle (GC); homens de 30 a 55 anos (média de 44,4), chamado Grupo de Homens (GH); mulheres de 30 a 55 anos (média de 42,3), chamado Grupo de Mulheres (GM); rapazes

⁸ Usamos aqui o termo “triângulo” porque assim foi utilizado por Gillier (2011). Ao longo do texto, porém, utilizamos o termo “trapézio vocálico”.

de 19 a 25 anos (média de 22,4), chamado Grupo de Rapazes (GR); garotas de 19 a 25 anos (média de 22,2), chamado Grupo de Garotas (GG). Neste recorte, estamos comparando as vozes femininas com as vozes masculinas, sem considerar o grupo controle como tal, nem as diferenças de idades. Todos os informantes assinaram termo de consentimento para uso dos dados exclusivamente para pesquisa.

A coleta de dados foi realizada em laboratório com tratamento acústico com o seguinte ferramental: Computador Pentium Dual Core 5.300 2.60 GHz, 1.99 Gb RAM Processador XP 2002 Service Pack 3, placa de som externa M-Audio Fast Track Pro 4x4, Microfone AKG C 3000 B. Para as gravações o programa utilizado foi o Audacity⁹. Os locutores forneceram suas vozes no interior de uma cabine acústica, na qual encontrava-se o microfone, recebendo a orientação de não variar sua posição corpórea, evitando tanto o distanciamento do microfone, como também uma inclinação cervical que pudesse modificar o movimento laríngeo.

O texto elaborado simulava um pedido de resgate durante um telefonema e era lido pelo informante por duas vezes em sua voz normal e duas vezes disfarçando a voz. A instrução era para que, no disfarce, buscasse não ser reconhecido. O texto continha para análise 14 palavras paroxítonas, contemplando as sete vogais orais do português /i, ε, a, o, u/, presentes em suas sílabas tônicas e localizadas entre consoantes plosivas não vozeadas /p, t, k/. As palavras inseridas no texto para análise eram:

/a/	/e/	/ε/	/i/	/o/	/o/	/u/
/'patu/	/ka'petε/	/'tεkε/	/'pike/	/'toku/	/'køkε/	/'putu/
/'kake/	/ka'tetu/	/es'petu/	/'pipe/	/'koku/	/'tøkε/	/ku'tuku/

Em cada coleta, o registro de áudio foi capturado com as seguintes configurações: formato *wave* monocal, frequência de amostragem de 44,1 KHz e 16 bits. Os arquivos foram nominados de acordo com o grupo e o número sequencial do informante, para facilitar os procedimentos seguintes.

⁹Software livre, versão 2.0.3, disponível em www.audacity.sourceforge.net/.

3.2 A análise perceptiva das vozes disfarçadas

A análise dos disfarces foi feita por um grupo de seis pessoas¹⁰ (juízes), com orientação prévia sobre a aplicação do BP-VPAS (ZULEICA; MADUREIRA, 2008). Faz-se necessário reforçar que o protocolo não foi aplicado em toda extensão, nem foram utilizados os níveis de graduação. O objetivo era apoiar na detecção de qual teria sido o ajuste escolhido pelo participante para disfarçar sua voz, para posterior comparação com alterações de F1 e F2, ou seja, analisar a relação entre a percepção dos juízes e a variação no trato vocal visualizada por meio da análise acústica. Os juízes ouviam as vozes quantas vezes considerassem suficiente para preencher de seus formulários individualmente. Depois, as respostas eram comparadas e, voltando a ouvir novamente quantas vezes fossem necessárias, uma resposta final era negociada para que houvesse um consenso de grupo. Também é importante relatar que a diferença que aqui se faz entre vozes normais e vozes disfarçadas se refere à voz natural do informante em relação à sua voz com a estratégia de disfarce. Não foram considerados os possíveis ajustes que o locutor pudesse realizar naturalmente em sua emissão.

3.2.1 Resultados das análises

Como se observa na Tabela 1, a escolha dos disfarces se deu tanto em ajustes supralaríngeos, tal como a protrusão labial e a mandíbula aberta, como em ajustes fonatórios, incluindo o falsete e a voz crepitante. Nos dados de variação de *pitch*, o segundo valor refere-se à composição desse ajuste com outro, isto é, o informante protruiu os lábios e baixou o *pitch*. Algumas das estratégias foram usadas em combinação, por exemplo, mandíbula aberta + voz nasal. A voz nasal, na verdade, esteve na maioria das vezes combinada com outro ajuste.

Tabela 1: Estratégias de disfarce utilizadas pelos informantes.

Estratégia de disfarce	Vozes femininas (30)	Vozes masculinas (20)
Mandíbula aberta	2	7
Lábios protruídos	6	1
Língua avançada	2	1
Língua recuada	2	0
Ajustes supralaríngeos		
Língua abaixada	1	0
Língua elevada	0	1
Voz nasal	4	5
Voz denasal	1	0

¹⁰ Cinco alunos e a professora de fonética de um curso de Letras.

Ajustes fonatórios	Hiperfunção laríngea	4	1
	Falsete	4	0
	Voz crepitante	1	3
	Voz áspera	2	4
	Pitch elevado	5 + 2	1 + 7
	Pitch abaixado	2 + 6	1

Verifica-se, pelos resultados, que a protrusão labial e a mandíbula aberta foram os ajustes mais frequentes no trato vocal, enquanto que a voz áspera foi um ajuste fonatório bastante utilizado. Diferente de resultados anteriores (KÜNZEL, 2000), muitas mulheres baixaram o *pitch*, enquanto vários homens elevaram o *pitch*. Foi interessante notar que houve uma relação dessa variação de *pitch* com duas estratégias específicas de disfarce, a protrusão labial por informantes do sexo feminino e a voz áspera por informantes do sexo masculino. As seis informantes que usaram a estratégia de protrusão labial foram também percebidas com o *pitch* abaixado. Segundo Laver (1980), a protrusão labial aumenta o eixo longitudinal do trato vocal e, conforme o modelo do tubo de ressonância, o comprimento maior do trato diminui as frequências (KENT; READ, 2002). Levantamos a hipótese de que, na tentativa de parecerem mais agressivas, pelo contexto de um pedido de resgate em sequestro, essas informantes escolheram baixar o *pitch* da voz, conseqüentemente, baixando os valores da frequência fundamental.

Também, como será apresentado mais adiante, as frequências de formantes tiveram valores mais baixos quando no ajuste de protrusão labial, fato também exposto por Laver (1980). Alguns informantes masculinos que elevaram o *pitch* combinaram essa estratégia com o ajuste de voz áspera (quatro informantes). As duas informantes do sexo feminino que usaram a voz áspera como estratégia de disfarce, por outro lado, baixaram o *pitch*. Aqui também foi levantada a hipótese de que a escolha da voz áspera teve relação com o contexto e deve ter sido utilizada para causar a impressão de agressividade. Laver (1980) afirma que a voz áspera é usada no inglês como sinal paralinguístico de raiva, e Moisk (2012) constata que a emoção exagerada na voz áspera pode gerar atributos agressivos e transgressivos.

Quatro dos ajustes usados pelos informantes foram escolhidos para uma relação com a análise acústica, dois ajustes de fonação - a voz áspera e o falsete, e dois do trato vocal - a protrusão labial e a abertura de mandíbula. Essa análise será apresentada mais adiante.

3.3 Os procedimentos de medição para análise acústica

A partir dos arquivos de áudio, e com o recurso de etiquetagem (*textgrids*) do PRAAT¹¹, realizou-se a segmentação manual de cada vogal alvo da análise, bem como de sua respectiva palavra, conforme pode ser observado na Figura 5. A delimitação dos trechos das vogais foi realizada conforme Escudero *et al* (2009), sendo os pontos inicial e final definidos pelo primeiro e último períodos que continham amplitude considerável e com formatos compatíveis com os dos períodos mais centrais.

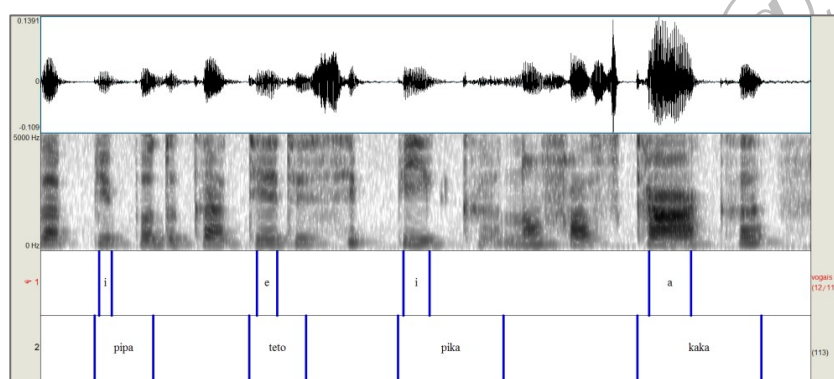


Figura 5: Tela do software do PRAAT, com visualização da etiquetagem do trecho da amostra GG9: “*pipa do Cateto e se pica, não faz caca*”. Na primeira camada está a forma da onda (oscilograma), na segunda o espectrograma, na terceira etiquetagem dos trechos correspondentes às vogais alvo etiquetadas e, na quarta e última, as palavras etiquetadas.

As transcrições das vogais (‘a’, ‘e’, ‘\ef’, ‘i’, ‘o’, ‘\ct’ ‘u’), serviram de referência para extração das medidas de duração, F0, F1 e F2 das vogais. A extração dos valores foi realizada com aplicação de script, desenvolvido pela terceira autora com os recursos disponibilizados no próprio Praat, baseando-se em tutoriais para aplicações semelhantes^{12,13}. A seguir são detalhadas as rotinas utilizadas para as medições, sendo importante ressaltar que após a extração os valores foram confirmados, sendo as medidas realizadas manualmente quando necessário.

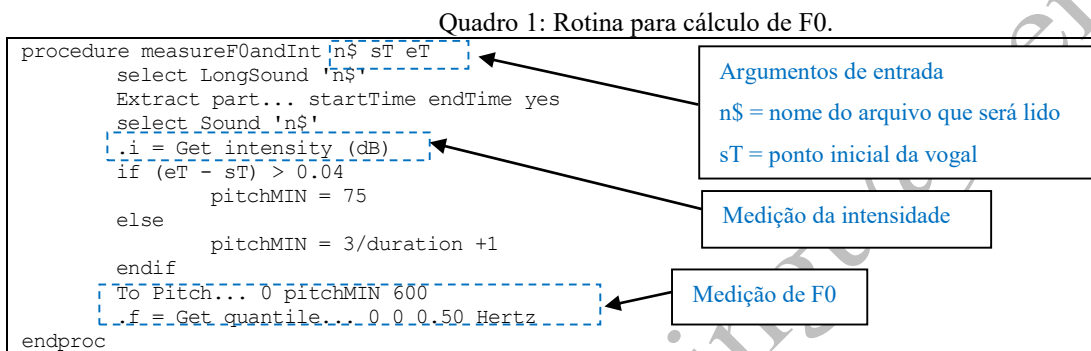
- a) Duração: determinada através de cálculo simples, pela subtração do ponto final e inicial de cada vogal.

¹¹ Software livre criado por Boersma e Weenick, 2008, disponível em <http://www.fon.hum.uva.nl/praat/>.

¹² Tutorial para Análise automática de formantes e plotagem de vogais no Praat, disponibilizado por Andréia Rauber em http://www.nupffale.ufsc.br/rauber/Praat_tutorial.pdf.

¹³ Scripts para extração de valores de F0 e de formantes, disponibilizados por Mietta Lennes (2003) em <http://www.helsinki.fi/~lennes/praat-scripts/>.

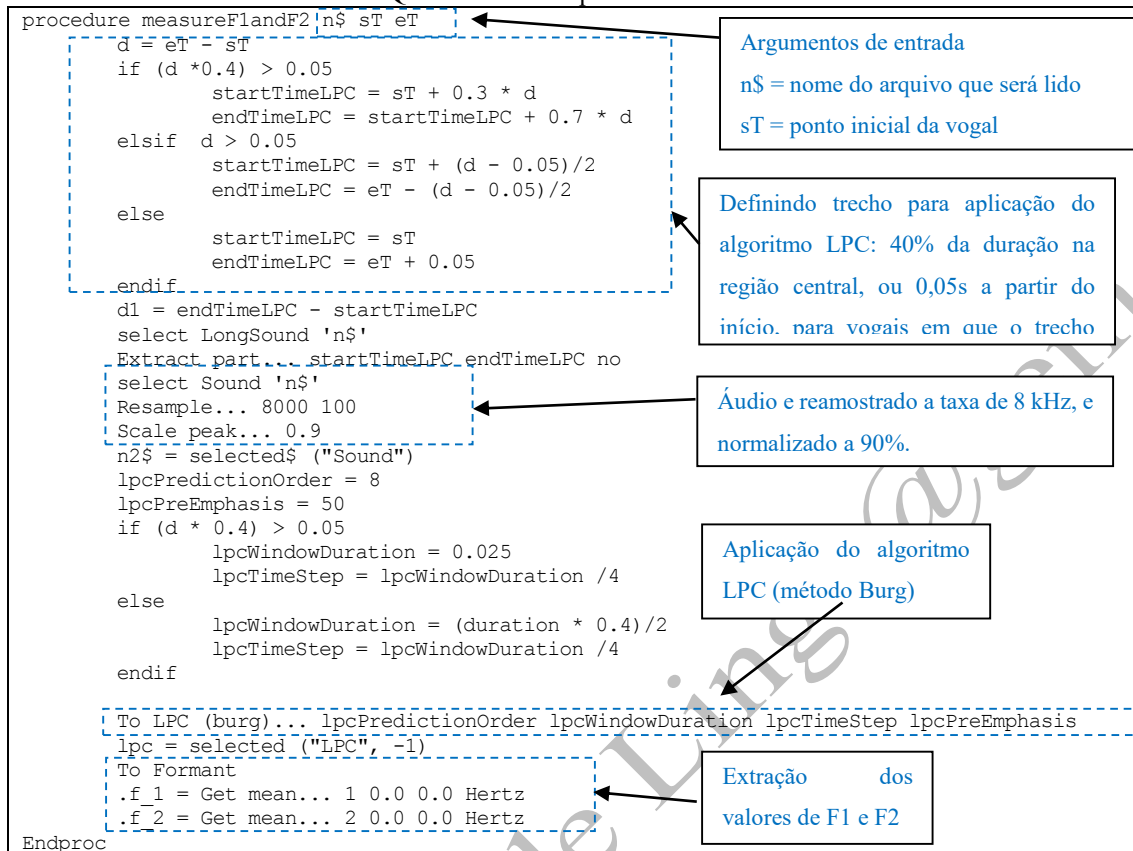
- b) F0: procedeu-se à extração do *pitch*¹⁴ do intervalo em análise (duração da vogal), sendo consideradas duas casas decimais no resultado. Foram mantidos os valores *default* do Praat para *ceiling* (valor máximo) de 600 Hz e o *floor* (valor mínimo) de 75 Hz. Para durações menores que 40 ms o algoritmo utilizado exige que o valor de *floor* seja recalculado, para atender a condição de resultar em no mínimo $(3/dur + 1)$. A seguir rotina utilizada para efetuar tal medição:



- c) F1 e F2: selecionou-se a porção central da vogal, correspondente a 40% da duração, com reamostragem do sinal para 8 kHz, precisão de 100 amostras e escala ajustada para 90% do valor de pico (normalização). Utilizou-se o algoritmo LPC, método *Burg*, sendo estabelecida ordem de predição igual a 8 (número de coeficientes utilizados no algoritmo, que deve corresponder a pelo menos o dobro do número de formantes a serem detectados), frequência pré-ênfase de 50 Hz, largura de janela de 25 ms para trechos maiores que 50 ms (em caso contrário, foi considerada a metade da duração do trecho) e um *timestep* de 25% da duração da janela. Daí, então, mediu-se a média do primeiro e do segundo formante. A seguir rotina utilizada para tal medição:

¹⁴ Vale lembrar que o *pitch* é o correlato perceptivo da frequência fundamental.

Quadro 2: Rotina para cálculo de F1 e F2.



4. Hipóteses e Resultados

4.1 Duração

Levando em conta os estudos de Escudero *et al.* (2009) sobre as vogais do português, levantaram-se as seguintes hipóteses em relação à duração: a) as mulheres produzem vogais mais longas que os homens; b) as vogais baixas são mais longas que as vogais altas (duração intrínseca); c) vogais posteriores tendem a ser mais longas que suas correspondentes anteriores. Também foi levantada a hipótese de que d) as vogais seriam mais longas no disfarce do que na voz normal, considerando que em situação de instrução a fala pode ser mais articulada.

A partir de testes não paramétricos Wilcoxon, foram encontrados os seguintes resultados:

- a) Hipótese não confirmada. Embora as mulheres tenham produzido sistematicamente vogais mais longas que os homens, tanto na voz normal como disfarçada, os resultados não foram estatisticamente significativos;

- b) Hipótese parcialmente confirmada. As diferenças foram significativas para as seguintes vogais: [e] > [i], [a] > [ɛ] e [ɔ] > [o], com $p < 0,05$ tanto na voz normal, como na voz disfarçada. Embora [ɛ] tenha sido mais longa que [e], as diferenças não foram significativas, com $p > 0,05$, nas duas modalidades. A vogal [o] foi mais longa que [u] na voz normal e mais curta no disfarce, mas também sem diferença significativa.
- c) Hipótese confirmada, salvo pela diferença entre [e] e [o] no disfarce, que teve a vogal anterior mais longa que a posterior, mas com diferença não significativa.
- d) Hipótese parcialmente confirmada. Os resultados apontam para uma tendência real de maior duração das vogais no disfarce, mas as diferenças não são significativas em todas as vogais nos diferentes grupos. A diferença de duração entre as vogais na fala normal e disfarçada parece ser mais relevante nos grupos masculinos, que tiveram diferença significativa ($p < 0,05$) nas vogais [e], [ɛ] e [o] no GH, e nas vogais [i], [e], [ɛ], [ɔ] e [o] no GR. Nos grupos femininos, GC teve resultados significativos em [i] e [o], GG em [o] e GM, em [i] e [e]. [a] e [u] não tiveram diferenças significativas de duração entre voz normal e disfarçada em nenhum dos grupos.

Conforme afirmam Escudero *et al* (2009), as vogais do português não apresentam a duração como traço fonológico, mas isso não exclui a possibilidade de haver diferenças fonéticas na duração das vogais, entre dialetos e entre gêneros. A pesquisa de Escudero *et al* (2009) comparou português brasileiro e português europeu, e vozes femininas e masculinas. Embora, na pesquisa que aqui se descreve, nem todas as hipóteses tenham se confirmado, tanto na diferença entre gêneros como no efeito da altura, duração intrínseca, a tendência se confirma. Quanto à diferença de duração na relação anterioridade-posterioridade da vogal, na pesquisa de Escudero *et al* (2009) para as vogais do português brasileiro, apenas a diferença entre [ɛ] e [o] não foi significativa. Nesta pesquisa, o que não foi significativa foi a diferença entre [e] e [o].

4.2 Frequência fundamental

Para a frequência fundamental, reportamos aqui as seguintes hipóteses em relação à frequência fundamental: a) as mulheres falantes de português têm valores de F0 mais altos que

dos homens¹⁵. A mudança do *pitch* é considerada uma das estratégias preferidas em disfarce de voz, conforme relatado inicialmente nos estudos de Eriksson (2010). Mais uma hipótese que aqui se levanta, então, é que conseqüentemente: b) haverá diferenças significativas nos valores de F0 entre voz normal e disfarçada, nos diversos grupos.

Extraídas as medidas de F0, testes estatísticos não paramétricos foram aplicados e os resultados foram:

- a) Hipótese confirmada. O teste de Mann Whitney revela que para todas as vogais, tanto na voz normal quanto disfarçada, as mulheres tiveram F0 mais altas que os homens, $p < 0,05$.
- b) Os resultados apontam para significativa diferença entre F0 das sete vogais na voz normal e disfarçada, quando se comparam os cinco grupos, conforme se observa nos resultados do teste Kruskal Wallis apresentados na Tabela 2.

Tabela 2 - Teste Kruskal Wallis – Comparação de F0 entre os cinco grupos.

		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/
Voz Normal	Chi square	31,862	28,840	27,443	28,587	22,418	26,873	32,184
	Asymp Sig.	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Voz Disfarçada	Chi square	23,824	14,640	16,473	22,704	10,087	20,676	20,411
	Asymp Sig.	0,000	0,007	0,002	0,000	0,039	0,000	0,000

O efeito do gênero é realmente significativo no tocante à F0, já que todas as vogais tiveram valores mais altos nas vozes femininas do que nas vozes masculinas. Além disso também comprova-se, pelos resultados das diferenças entre a voz normal e disfarçada que, além das questões anatômicas que determinam os valores de F0, esse é um parâmetro acústico que pode ser muito bem controlado pelo falante, conforme afirmam Kreiman e Sidtis (2012). As autoras também afirmam que esse é um parâmetro saliente ao ouvinte, fato comprovado pelos resultados da análise perceptiva.

Outros trabalhos de análise de F0 foram realizados a partir da mesma análise perceptiva (KREMER; GOMES, 2014; KREMER, 2015), nos quais foram comparadas as vozes femininas

¹⁵ Embora essa diferença universal entre vozes de homens e mulheres seja um dado confirmado em várias pesquisas (vide ESCUDERO *ET AL* 2009), levantou-se a hipótese pelo caráter formativo da pesquisa.

e masculinas a partir da curva de F0 e da estratégia de disfarce. Resultados dessas pesquisas mostraram que a maioria dos participantes elevaram o *pitch*, uns mais significativamente que outros, sendo que a maior elevação de *pitch* aconteceu com um participante masculino e o maior abaixamento do *pitch* foi percebido em duas participantes do sexo feminino. A escolha de elevação do *pitch* foi mais frequente entre os homens, e o abaixamento mais frequente entre as mulheres. Esses resultados são consistentes com os de Masthoff (1996), mas contrários aos resultados de Künzel (2000).

4.3 F1 e F2

Para os valores de F1 e F2, a análise foi feita na diferença entre as vozes femininas e masculinas, e duas hipóteses foram levantadas: a) mulheres tendem a ter valores de F1 e F2 mais altos que de homens (ESCUDERO et al, 2009); b) mulheres tendem a ter espaço vocálico maior que dos homens (GILLIER, 2011).

Para os valores de F1, o teste não paramétrico de Mann Whitney revelou que as diferenças são significativas para todas as vogais na voz normal e para quase todas na voz disfarçada ($p < 0,05$), consistente com o trabalho de Escudero *et al* (2009), que afirmam que essa observação já foi reportada de forma abundante na literatura. Nesta pesquisa, as vogais altas [i] e [u] tiveram diferenças não significativas no disfarce ($p > 0,05$).

Os valores de F2, por outro lado, apresentaram diferenças significativas apenas nas vogais anteriores nas vozes normais, e na vogal [ɔ] nas vozes normais e disfarçadas. Diferente do trabalho de Escudero *et al* (2009), que afirmam que as mulheres têm valores de F2 mais altos que dos homens, neste trabalho apenas a vogal média baixa apresentou tal resultado. As demais vogais posteriores tiveram diferenças estatisticamente não significativas nas vozes normais. Nas vozes disfarçadas, apenas a vogal [ɔ] apresentou valores com diferenças significativas entre vozes femininas e masculinas. Isso demonstra que os informantes, ao escolherem ajustes para disfarçar sua voz, realizaram movimentação no trato vocal, especificamente no corpo da língua, que resultaram em aproximação dos valores de F2 nas duas modalidades de voz.

O uso de padrão formântico das vogais em contexto forense, como um parâmetro distintivo robusto, tem sido defendido por vários autores (por exemplo, NOLAN *et al*, 2009, 2011; NOLAN; GRIGORAS, 2005). Duckworth *et al* (2011) alertam, no entanto, para a

importância na consistência no processo de medição de formantes. Recomendam que se estabeleça um protocolo para a análise forense para facilitar a validação das evidências.

Quanto à hipótese de que as mulheres apresentam espaço vocálico mais amplo que os homens, utilizou-se um gráfico com a média dos valores de F1 e F2. Nesse caso, para equilíbrio dos valores, apenas valores de 40 informantes foram utilizados. Não foram considerados os valores do chamado grupo controle, com vozes femininas.

Gillier (2011) defende que o triângulo (ou trapézio) vocálico pode ser uma importante ferramenta para análise, uma vez que revelam tendências articulatórias e características anatômicas do trato vocal dos indivíduos. É interessante verificar algumas questões que se apresentam nos gráficos das Figura 6 e 7, que representam os dados de voz normal e disfarçada, respectivamente, dos quatro grupos (GM, GG, GH e GR).

No gráfico da Figura 6, pode-se compreender o que representam as diferenças significativas dos valores de F1 entre a produção das vozes femininas (GM e GG) com a das masculinas (GH e GR), sempre mais altos para as mulheres, trazendo para baixo as linhas em tons de rosa. Para F2, conforme reportado acima, as vogais anteriores e a vogal posterior média baixa tiveram diferenças significativas entre as vozes das mulheres e dos homens, mas as vogais posteriores média alta e alta não apresentaram diferenças significativa. Nota-se que as regiões de [u] e [o] estão muito próximas nos quatro grupos.

No gráfico da Figura 7, observa-se que os valores de F1 nas vozes disfarçadas apresentaram diferenças significativas para quase todas as vogais, salvo as altas. Isso se constata pela proximidade, no gráfico, dos quatro grupos nas regiões de [i] e [u]. Como F2 se refere à anterioridade-posterioridade e, no disfarce, as diferenças não foram significativas, as regiões de produção ficaram muito próximas em quase todas as vogais.

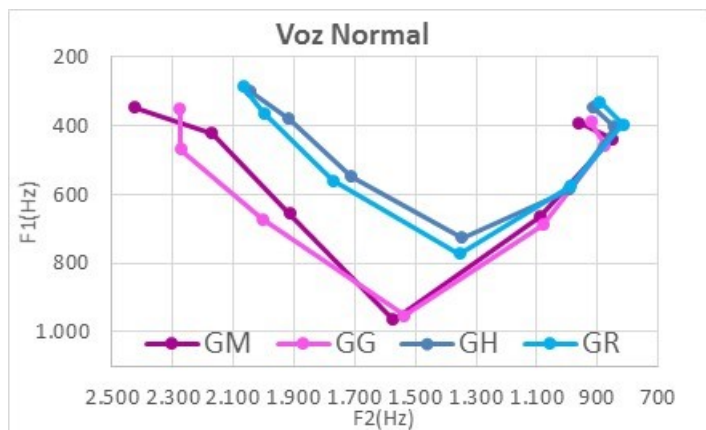


Figura 6: Espaço vocálico da média dos valores de F1 e F2 – voz normal.



Figura 7: Espaço vocálico da média dos valores de F1 e F2 – voz disfarçada.

5. Relação entre análise perceptiva e análise acústica

Para a melhor compreensão da relação entre a articulação empreendida pelo locutor, a percepção do ouvinte e o efeito acústico da voz, alguns gráficos de F1-F2 individuais agrupados por alguns ajustes na produção do disfarce serão analisados.

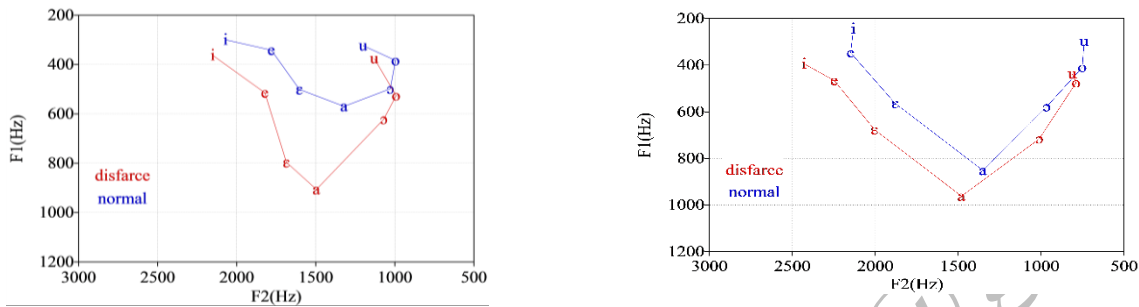
5.1 Estratégia de disfarce e efeitos acústicos

Após a análise auditiva dos disfarces, quatro ajustes foram selecionados para comparação com a análise acústica nas vozes normal e disfarçada. Por vezes, esses ajustes foram combinados com outros, por exemplo, o arredondamento dos lábios combinando com o abaixamento do *pitch*. Alguns dos ajustes mais comuns na análise dos 50 participantes, ao disfarçar a voz foram: a) mandíbula aberta – nove participantes (duas vozes femininas; sete vozes masculinas); b) protrusão labial – sete participantes (seis vozes femininas; uma voz masculina); c) voz áspera – seis participantes (duas vozes femininas; quatro vozes masculinas); d) falsete – quatro participantes (todas vozes femininas). Esses resultados reforçam a afirmação de Künzel (2000) de que há diferenças entre os gêneros na preferência por modos de disfarce, assim como de suas estratégias articulatórias para implementá-los. Nos participantes desta pesquisa, observa-se que a protrusão labial e o falsete são mais usados pelas participantes do sexo feminino, ao passo que a mandíbula aberta e a voz áspera são mais usadas por participantes do sexo masculino.

Na comparação entre os ajustes percebidos auditivamente e as análises dos gráficos das vogais formados pelos valores de F1 e F2, foi possível compreender a relação entre os movimentos realizados no trato vocal com a estratégia do disfarce utilizada. Os gráficos abaixo

mostram alguns exemplos. A cor azul indica as médias para as vogais produzidas na voz normal, e a cor vermelha, as médias das vogais produzidas na voz disfarçada.

5.1.1 Mandíbula aberta



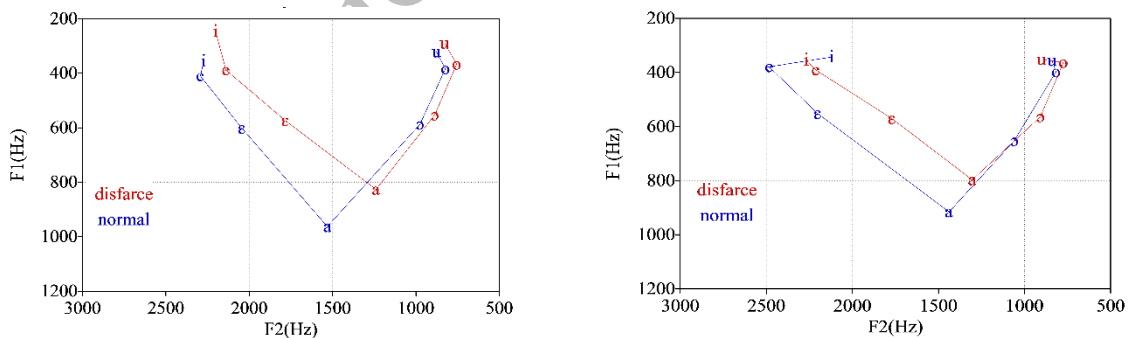
(a) Informante GH2.

(b) Informante GR9.

Figura 8: Gráficos de F1/F2 com as médias para voz normal e voz com disfarce de dois dos informantes que usaram o ajuste de mandíbula aberta.

Os gráficos demonstram como a percepção do ajuste de mandíbula aberta, escolhida pelos juízes, coincide com aumento do valor de F1, comprovando o abaixamento da língua pela abertura da mandíbula. Laver (1980) afirma que o efeito mais importante na posição da mandíbula é verificado nas mudanças de F1, com aumento de frequência na medida em que a mandíbula se torna mais aberta.

Protrusão labial



(a) Informante GM4.

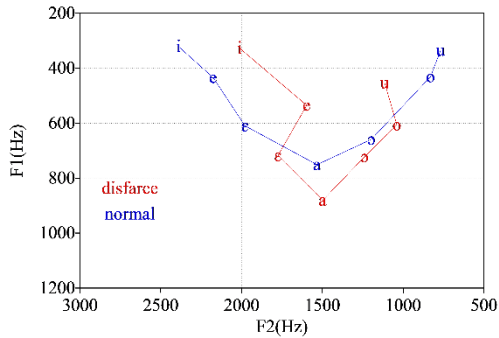
(b) Informante GG10.

Figura 9: Gráficos de F1/F2 com as médias para voz normal e voz com disfarce de dois dos informantes que usaram o ajuste de protrusão labial.

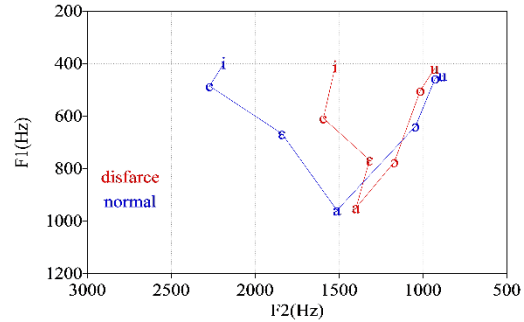
No ajuste de arredondamento dos lábios ou protrusão labial, percebe-se o recuo da língua pela diminuição de F2, principalmente nas vogais anteriores. A protrusão labial aumenta o eixo longitudinal do trato vocal, adicionando mais uma seção de comprimento variado ao

trato, o que resulta no efeito acústico de frequência mais baixa em todos os formantes (LAVÉ, 1980).

Falsete



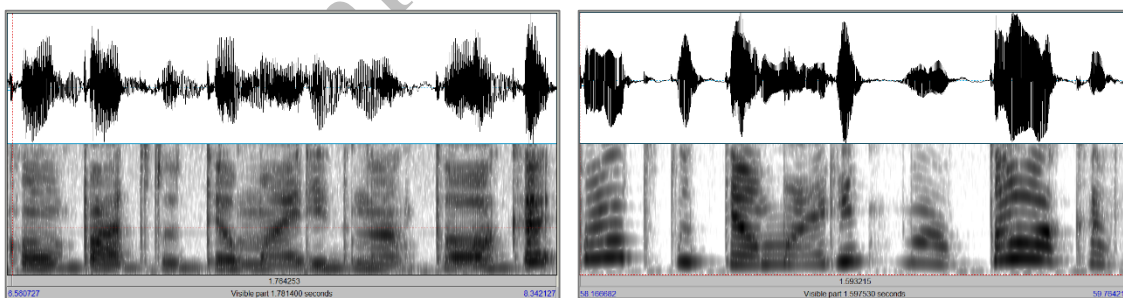
(a) Informante GC5



(b) Informante GG1.

Figura 10: Gráficos de F1/F2 com as médias para voz normal e voz com disfarce de duas das informantes que usaram o ajuste de falsete.

Pelos gráficos, observa-se que o efeito do falsete parece operar com um estreitamento do espaço vocálico. Segundo Laver (1980), o falsete é caracterizado acusticamente por vários fatores: F0 mais alta do que em voz modal; interação da F0 alta e o modo de vibração das pregas vocais; o espectro apresenta queda mais íngreme, caindo aproximadamente -20 dB por oitava. Enquanto a voz modal tem uma porção fechada da forma da onda laríngea como um componente abrupto, no falsete a porção aberta é mais inclinada.

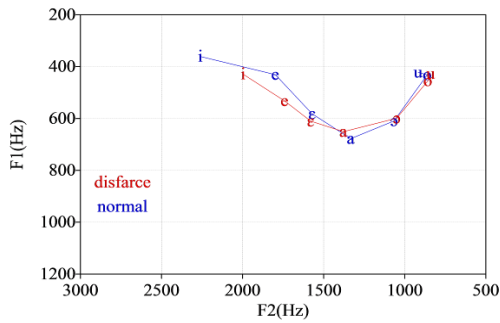


(a) Voz normal.

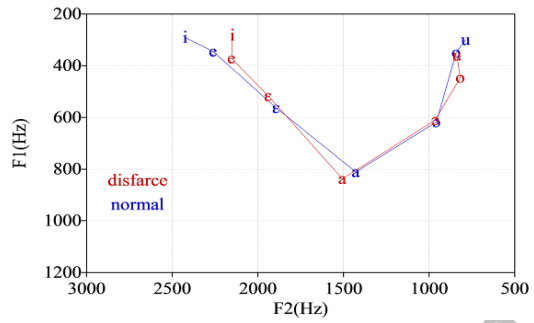
(b) Voz disfarçada com falsete.

Figura 11: Oscilogramas e espectrogramas de trechos da informante GC3.

Verifica-se pelas formas da onda nas vozes normal e disfarçada da informante GC3, na Figura 11, as inclinações mais abruptas na forma da onda. O espectrograma, na Figura 11 (b), do falsete reflete a tensão das pregas vocais.

Voz áspera

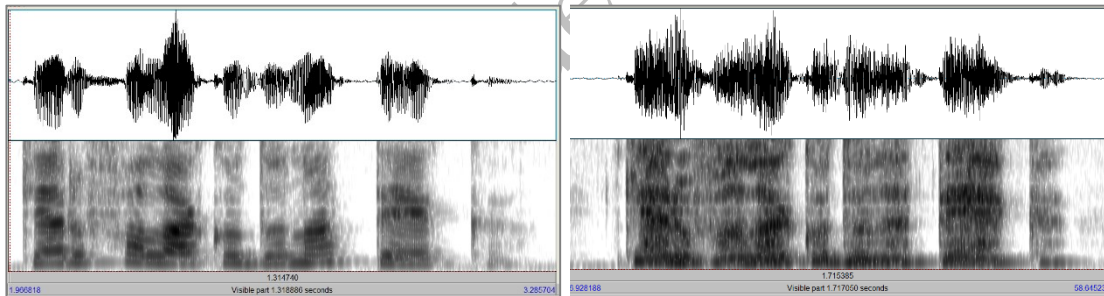
(a) Informante GH1.



(b) Informante GR6.

Figura 12: Gráficos de F1/F2 com as médias para voz normal e voz com disfarce de dois dos informantes que usaram o ajuste de voz áspera.

Na voz áspera praticamente não há alteração na configuração das vogais. Esse parece ser um ajuste que, realizado no nível laríngeo, não requer movimentos importantes na região supralaríngea. As características acústicas da voz áspera estão relacionadas principalmente com a irregularidade da forma da onda glotal e ruído espectral (LAVER, 1980), o que pode ser observado na Figura 13.



(a) Voz normal.

(b) Disfarce com voz áspera.

Figura 13: Oscilogramas e espectrogramas de trechos da informante GC2.

O ruído característico da voz áspera pode ser observado tanto na forma da onda como no espectrograma da Figura 13(b). Segundo Moisk (2012) a correlação fisiológica da voz áspera é o aumento de tensão na região laríngea e faríngea, e atividade muscular intensa nas pregas vocais.

Podemos, assim, resumir as estratégias dos participantes na realização dos seus disfarces: a) a maioria buscou alterar o *pitch* na voz disfarçada, utilizando diferentes estratégias, ou de elementos do trato vocal ou de elementos fonatórios (ERIKSSON, 2010); b) elementos do trato vocal são mais efetivos para a variação de F1 e F2; c) dos elementos fonatórios, o

falsete foi o mais efetivo para a variação de F1 e F2, conforme pesquisa anterior (GILLIER, 2011); d) as vogais foram mais longas no disfarce, possivelmente por ser uma situação de instrução; e) a variação de F0 foi mais significativa nos grupos masculinos; f) a variação de F1 foi mais significativa nos grupos masculinos (vários usaram abertura de mandíbula), e de F2 nos grupos femininos (várias usaram o arredondamento dos lábios e avanço de língua).

6. Considerações finais

Este trabalho ressalta a importância da abordagem multidisciplinar da análise de fala em fonética forense. O cunho instrucional proposto foi levado a termo, tendo sido os membros do grupo instruídos sobre os procedimentos para realização de gravações e análises acústicas relacionadas à pesquisa. Foram capazes de atuar como juízes para a realização da análise perceptiva da qualidade vocal normal e disfarçada, identificando ajustes. Estiveram aptos também a analisar os parâmetros extraídos, estabelecendo relações entre a avaliação perceptiva e os achados acústicos, relacionando-os.

As hipóteses levantadas em relação à duração não foram em sua totalidade confirmadas, mas os resultados revelaram tendências que confirmam resultados de trabalhos anteriores, como o de Escudero *et al* (2009) para as vogais do português brasileiro, principalmente na diferença entre os gêneros e no efeito de altura.

Em relação à frequência fundamental, os resultados das diferenças entre a voz normal e disfarçada revelam a facilidade de controle desse parâmetro acústico pelo falante. As diferentes preferências por elevação ou abaixamento do *pitch* por informantes do sexo masculino e feminino também foram relevantes nesta pesquisa, pois foram condizentes com o contexto estabelecido para coleta de dados. O abaixamento do pitch, preferido por informantes do sexo feminino, pode ter relação com uma vontade de parecerem agressivas.

A análise do padrão formântico, que neste trabalho focou apenas o primeiro e segundo formantes, foi importante para que os participantes em instrução pudessem refletir sobre a relação entre a análise perceptiva que fizeram dos disfarces realizados e os movimentos articulatórios necessários para implementá-los. A partir da análise dos gráficos, esses participantes foram capazes de visualizar a abertura da mandíbula causando um aumento do valor de F1, o recuo da língua causando a diminuição do valor de F2, e o estreitamento do espaço vocálico causado pelo falsete.

Já defendemos em outros textos (GOMES *et al*, 2012; GOMES; CARNEIRO, 2014) a carência das pesquisas em fonética forense e a importância da abordagem interdisciplinar. Reforçamos aqui essa defesa, conclamando mais pesquisadores, profissionais e estudantes a realizar pesquisas para o desenvolvimento da área.

Referências

- BEHLAU, M. **Voz: O livro do especialista**. Rio de Janeiro: Revinter, 2004. Vol. 1.
- BELL, A. G. Vowel theories. In: **American Journal of Otology**, Vol. 1, p. 163-180, 1879.
- BRAID, A. C. M. **Fonética Forense**. Campinas: Millennium, 2003, 2. ed.
- CAMARA JR., J. M. **Estrutura da língua portuguesa**. Petrópolis: Editora Vozes, 2002. 35. ed.
- CAMARGO, Z.; MADUREIRA, S. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In: **Proceedings Of The Fourth Conference On Speech Prosody**. Campinas, 2008.
- DAJER, M. E. **Análise de Sinais de Voz por padrões visuais da dinâmica vocal**. Tese de Doutorado. Programa de Engenharia Elétrica da Escola de Engenharia da USP, 2010.
- DUCKWORTH, M.; MCDUGAL, K.; DE JONG, G.; SHOCKEY, L. Improving the consistency of formant measurement. In: **The International Journal of Speech, Language and the Law**, Vol. 18.1, 2011, p. 35-51. **crossref** <http://dx.doi.org/10.1558/ijsl.v18i1.35>
- ERIKSSON, A. The Disguised Voice: Imitating Accents or Speech Styles and Impersonating Individuals. In: **Language and Identities**, 8. Edinburg: Kapitell, 2010. p.86 – 96.
- ESCUADERO, P.; BOERSMA, P.; RAUBER, A. S.; BION, R. A. H. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. **Journal of Acoustic Society of America**, 126 (3), 2009, 1379–1393. **crossref** <http://dx.doi.org/10.1121/1.3180321>
- ESPINDULA, A. **Perícia Criminal e Cível: Uma visão geral para peritos e usuários da perícia**. Campinas: Millennium Editora Ltda, 2005, 2. ed.
- FANT, G. **The Acoustic Theory of Speech Production**. The Hague: Mouton, 1960.
- GILLIER, R. **O disfarce de voz em fonética forense**. Master Thesis, Departamento de Linguística Geral e Românica, Faculdade de Letras, Universidade de Lisboa, 2011.
- GOMES, M. L. C.; CARNEIRO, D. O. **A fonética forense no Brasil: cenários e atores**. *Language and Law/Linguagem e Direito*. Vol. 1(1), 2014. p. 751-754.

GOMES, M. L. C.; RICHERT, L. C.; MALAKOSKI, J. Identificação de locutor na área forense: a importância da pesquisa interdisciplinar. **Anais do X Encontro do CELSUL – Círculo de Estudos Linguísticos do Sul UNIOESTE - Universidade Estadual do Oeste do Paraná Cascavel-PR**, 2012.

GOLD, E.; FRENCH, P. An International Investigation of Forensic Speaker Comparison. In: **ICPhS XVII**. Hong Kong, 2011. p. 751-754.

HOLLIEN, H. **Forensic Voice Identification**. London, UK: Academic Press, 2002.

IAFPA. **Proceedings of IAFPA 2015**. 24th Annual Conference of the International Association for Forensic Phonetics and Acoustics. Leiden: Leiden University, 2015.

JONES, D. **An outline of English phonetics**. Cambridge: Heffer, 1962.

KEATING, P. Universal phonetics and the organization of grammars. In: FROMKIN, V. **Phonetic Linguistics: essays in honor of Peter Ladefoged**. New York: Academic Press, 1995. p. 115-132.

KENT, R. D.; READ, C. **Acoustic Analysis of Speech**. Albany: Singular, 2002. 2. ed.

KINOSHITA, Y., ISHIHARA, S., ROSE, P. **Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition**. Equinox Publishing Online, p. 92, 2009.

KREIMAN, J.; GERRATT, B. R. Perceptual Evaluation of Voice Quality: Review, Tutorial, and a Framework for Future Research. **Journal of Speech and Hearing Research**. Vol. 36, p. 21-40, Fev. 1993. **crossref** <http://dx.doi.org/10.1044/jshr.3601.21>

KREIMAN, J.; SIDTIS, D. **Foundations of Voice Studies**. Malden: Wiley-Blackwell, 2013.

KREMER, R. L.; GOMES, M. L. C. A eficiência do disfarce em vozes femininas: uma análise da frequência fundamental. **ReVEL**, vol. 12, n. 23, 2014.

KREMER, R. L. **Fundamental frequency: an analysis of age and gender**. Trabalho de Conclusão de Curso apresentado para conclusão de grau de licenciado no curso de Licenciatura em Letras Português/Inglês da UTFPR Curitiba, 2015.

KÜNZEL, H. J. Effects on voice disguise on speaking fundamental frequency. In: **Forensic Linguistics**, 7 (2). Birmingham: University of Birmingham Press, 2000. p. 149-179.

_____. Beware of the ‘telephone effect’: the influence of telephone transmission on the measurement of formant frequencies. **Forensic Linguistics** 8(1) 2001. p. 1350-1771.

LAVER, J. **The Phonetic Description of Voice Quality**. Cambridge: Cambridge University Press, 1980.

LINDBLOM, B. Vowel duration and a model of lip-mandible coordination. **Speech Transmission Laboratory – Quarterly Progress and Status Report**. Stockholm, 1, 29.

MADAZIO, G. **Diagrama de desvio fonatório na clínica vocal**. Tese apresentada à Universidade Federal de São Paulo - Escola Paulista de Medicina, para a obtenção do Título de Doutor em Ciências, 2009.

MARCHAL, A.; REIS, C. **Produção da fala**. Belo Horizonte: Editora UFMG, 2012.

MASTER, S. **Análise acústica e perceptivo-auditiva da voz de atores e não atores masculinos**: Long term average spectrum e o “formante do ator”. Tese de Doutorado no Programa de Pós-Graduação em Distúrbios da Comunicação Humana da Escola Paulista de Medicina da USP, 2005.

MASTHOFF, H. A report on voice disguise experiment. In: **Forensic Linguistics**, Vol.3 (1), Equinox Publishing Online, UK, 1996, p. 160-167.

MOISIK, S. R. Harsh Voice Quality and Its Association with Blackness in Popular American Media. In: **Phonetica**, 69, 2012. p. 193-215. **crossref** <http://dx.doi.org/10.1159/000351059>

NOLAN, F.; GRIGORAS, C. A case for formant analysis in forensic speaker identification. In: **International Journal of Speech Language and the Law**, 12(2), 2005. p. 143-173. **crossref** <http://dx.doi.org/10.1558/sll.2005.12.2.143>

NOLAN, F.; MCDUGALL, K.; DE JONG, G.; HUDSON, T. The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. In: **International Journal of Speech Language and the Law**, 16(1), 2009. p. 31-57.

NOLAN, F.; MCDUGALL, K.; HUDSON, T. Some acoustic correlates of perceived (dis)similarity between same-accent voices. In LEE, W. S.; ZEE, E. **Proceedings of the 17th International Congress of Phonetic Sciences**, Hong Kong, 2011, p. 1506-1509.

NOOTEBOOM, S. The Prosody of Speech: Melody and Rhythm. In: HARDCLATLE, W. J.; LAVER, J. **The Handbook of Phonetic Sciences**. Oxford: Blackwell, 1997. p. 640-673.

PIETROFORTE, A. V. A língua como objeto da linguística. In: MUSSALIN, F.; BENTES, A. C. **Introdução à Linguística 3: fundamentos epistemológicos**. São Paulo: Cortez, 2005. p. 75-93.

ROSE, P. **Forensic Speaker Identification**. Canberra: Taylor & Francis, 1997.

STEVENS, K. N. Articulatory-Acoustic-Auditory Relationships. In: HARDCLATLE, W. J.; LAVER, J. **The Handbook of Phonetic Sciences**. Oxford: Blackwell, 1997. p. 462-506.

Artigo recebido em: 20.09.2015

Artigo aprovado em: 20.03.2016