v. 15, n. 2, p. 81-94, 2026
ISSN 2237-9223

REVISTA BRASILEIRA DE CRIMINALÍSTICA

DOI: http://dx.doi.org/10.15260/rbc.v15i2.1013

Descrição Fonética do Disfarce por Modulação da Altura Tonal (*Pitch*): Um Estudo Piloto

G. A. da Silva[a*]

[a] *Instituto de Criminalística, Superintendência de Polícia Técnico-Científica do Estado de São Paulo, São Paulo (SP), Brasil*

**Endereço de e-mail para correspondência: gerson.gas@policiacientifica.sp.gov.br Tel.: +55-11-980363902*

Recebido em 21/07/2025; Revisado em 09/05/2026; Aceito em 09/05/2026

Resumo

Nas tipologias tradicionais, estratégias de disfarce que envolvem o aumento ou a diminuição da frequência fundamental (F_0) costumam ser classificadas como modificações da fonte glótica. No entanto, essa categorização ignora que o disfarce vocal constitui, em essência, uma forma de alteração da qualidade de voz. Conforme o modelo fonético proposto por Laver (1980), a qualidade vocal não se restringe à vibração das pregas vocais, mas resulta de um conjunto integrado de ajustes fonatórios, articulatórios e de tensão muscular — os chamados *settings* — que moldam sistematicamente o perfil vocal de um indivíduo, em curto e longo prazo. O objetivo deste trabalho é investigar os ajustes articulatórios envolvidos nesse tipo de disfarce, de modo a descrevê-lo foneticamente dentro do arcabouço teórico da qualidade de voz proposto por Laver. Para quantificar esses ajustes, foram utilizadas duas métricas principais: o comprimento estimado do trato vocal (CTV_e) e a área do espaço vocálico.

Palavras-Chave: Disfarce; Qualidade de Voz; *Settings* articulatórios; Fonética Forense; Física da Fala.

Abstract

In traditional typologies, disguise strategies involving increases or decreases in fundamental frequency (F_0) are typically classified as modifications of the glottal source. However, this categorization overlooks the fact that voice disguise is, in essence, a form of voice quality alteration. According to the phonetic model proposed by John Laver (1980), voice quality is not limited to the vibration of the vocal folds, but results from an integrated set of phonatory, articulatory, and muscular tension adjustments — referred to as *settings* — that systematically shape an individual's vocal profile over both short- and long-term timescales. The present study investigates the articulatory adjustments underlying this type of disguise, with the aim of providing a phonetic description within Laver's theoretical framework of voice quality. To quantify these adjustments, two main metrics were used: estimated vocal tract length (VTLe) and vowel space area.

Keywords: Voice disguise; Voice quality; Articulatory settings; Forensic phonetics; Physics of Speech.

1. INTRODUÇÃO

A ocorrência de disfarces vocais configura um fenômeno relevante no âmbito pericial, embora permaneçam escassos e heterogêneos os levantamentos disponíveis na literatura. Hermann J. Künzel, por exemplo, com base em dados acumulados ao longo de mais de duas décadas no setor do Bundeskriminalamt (BKA), responsável por perícias de áudio e fonética forense, reporta que entre 15% e 25% dos casos analisados anualmente apresentam, no mínimo, uma forma de disfarce vocal. Tais estratégias incluem, entre outras, o uso de falsete, *creaky voice*, voz sussurrada, emulação de sotaque estrangeiro e pinçamento nasal [1]. Por outro lado, a JP French Associates — considerada o principal laboratório de fonética forense e acústica do mundo, sediado no Reino Unido — apresenta uma estimativa mais modesta, segundo a qual aproximadamente um em cada quarenta casos envolve algum tipo de disfarce vocal [2].

Dados observacionais acumulados pelo autor ao longo de mais de duas décadas de atuação junto ao setor responsável pelas perícias em áudio e ciência forense da fala do Núcleo de Identificação Criminal (NIC) do Instituto de Criminalística de São Paulo sugerem que tal proporção não se distribui de maneira homogênea entre as diferentes naturezas delitivas, assumindo maior relevância quando estas são consideradas individualmente. Em crimes contra a honra, por exemplo, observam-se proporções superiores a 20%, enquanto, em casos que envolvem extorsão mediante sequestro, o uso de disfarce vocal tende a se tornar predominante na casuística observada.

Como já observado em [1], o tipo de disfarce vocal empregado pode guardar relação com a modalidade do crime. Observações empíricas derivadas da casuística do

Núcleo indicam que casos de importunação sexual, por exemplo, estão frequentemente associados ao uso de *creaky voice* ou de voz sussurrada — estratégias que mascaram parcialmente a identidade vocal e, ao mesmo tempo, conferem à fala um tom sugestivo ou insinuante, em consonância com o teor íntimo da interação. Já nos crimes de estelionato e extorsão, não é raro observar a simulação de sotaques estrangeiros ou de regionalismos.

Tais estratégias visam à alteração da emissão habitual do indivíduo, modificando a qualidade vocal, aqui entendida como a impressão global produzida pela voz em suas múltiplas dimensões — psicológica, biológica e socioeducacional, entre outras.

À luz dessa concepção, o presente artigo tem por objetivo apresentar um estudo piloto sobre disfarces vocais baseados na manipulação da frequência fundamental (F_0), reinterpretando-os como um fenômeno de reorganização da qualidade de voz. Para tanto, utiliza-se o arcabouço proposto por John Laver [3] para a descrição da qualidade vocal.

Para situar o artigo no amplo espectro da ciência forense da fala, a seção seguinte apresenta um panorama de estudos sobre disfarce vocal que acompanha a evolução das metodologias adotadas ao longo do tempo, desde abordagens perceptivas até métodos fonético-acústicos e sistemas automáticos supervisionados

2. REVISÃO CIENTÍFICA

Como recurso voltado à alteração da qualidade de voz, o disfarce vocal possui uma dimensão intrinsecamente perceptiva. Por essa razão, muitos estudos sobre disfarce vocal baseiam-se em experimentos perceptivos, conduzidos com ouvintes treinados ou não, a fim de avaliar a eficácia das modificações vocais na detecção ou ocultação da identidade. Para esses pesquisadores, a investigação da forma como ouvintes não treinados percebe os disfarces vocais justifica-se pelo fato de que, em situações reais, são essas pessoas — vítimas ou testemunhas — que normalmente presenciam o crime [4]. Seguindo essa abordagem, em um estudo frequentemente citado [5] investigou a efetividade do uso deliberado da voz crepitante[1] como técnica de disfarce vocal em um contexto forense. Em um experimento do tipo ABX, os autores testaram a capacidade de ouvintes treinados em fonética de identificar um falante cuja voz havia sido propositalmente modificada. Os resultados mostraram que pares de amostras com voz modal foram corretamente identificados em 90% dos casos, enquanto a taxa de acerto caiu para 65% quando a amostra "X" apresentava *creaky voice* — indicando que esse tipo de fonação interfere de maneira significativa na tarefa discriminativa, embora não a inviabilize completamente.

Outros estudos perceptivos subsequentes passaram a destacar os desafios na identificação de disfarces vocais, bem como as diferenças entre os tipos de disfarce quanto à sua dificuldade de reconhecimento. Nessa categoria, encontra-se o estudo de Hollien, Majewski e Doherty [6]. O estudo investigou a identificação perceptiva de vozes sob três condições de fala: normal, estresse e disfarce, com o objetivo de estimar a capacidade dos ouvintes de reconhecer os falantes e avaliar o impacto da familiaridade sobre o desempenho na tarefa. O experimento contou com três grupos de ouvintes: (a) familiarizados com os falantes; (b) não familiarizados com os falantes; e (c) não familiarizados nem com os falantes nem com a língua. Os resultados indicaram que a condição de disfarce levou a uma redução significativa nas taxas de identificação em todos os grupos, em comparação com a condição não disfarçada (Grupo A: 98% → 79%; Grupo B: 40% → 21%; Grupo C: 27% → 18%).

Em linha com esses achados, um estudo conduzido pelo Bundeskriminalamt [7], evidenciou que mesmo vozes familiares tornam-se substancialmente mais difíceis de reconhecer por ouvintes não treinados quando o falsete é empregado como estratégia de disfarce. Nesse caso, as taxas de identificação, elevadas na condição de fala normal (97%), caíram drasticamente para apenas 4% sob fonação em falsete.

A introdução do espectrograma permitiu os primeiros estudos sobre disfarce vocal, conduzidos principalmente nas décadas de 1970 e início de 1980. Esses estudos devem ser compreendidos como uma resposta à chamada técnica do *voiceprint* [8-10], cujos proponentes afirmavam que seu desempenho não seria significativamente afetado por disfarces. Todos esses trabalhos, críticos à leitura de espectrogramas na abordagem proposta por Kersta [10], faziam referência a uma investigação clássica [11], que demonstrou que o disfarce vocal provoca, de fato, alterações significativas em parâmetros espectrográficos — como as frequências centrais e as larguras de banda dos formantes, além da frequência fundamental, entre outros.

Um experimento descrito em [12] avaliou os efeitos de diferentes estratégias de disfarce vocal na discriminação de falantes por ouvintes. Os participantes produziram sentenças em sua voz habitual e em cinco modos distintos de disfarce, incluindo a voz presbifônica, rouquidão, hipernasalidade, diminuição do tempo de fala e disfarce livre. Os ouvintes, com distintos níveis de treinamento, foram então solicitados a julgar se pares de sentenças haviam sido proferidos pela mesma pessoa ou por indivíduos diferentes. Quando ambos os estímulos envolviam vozes não disfarçadas, a taxa de discriminação

[1] Também conhecida como registro de pulso, registro basal, *creaky voice* ou *glottal fry*

correta foi de 92%. No entanto, a introdução de uma amostra disfarçada reduziu significativamente o desempenho dos ouvintes, com taxas de acerto variando entre 59% e 81%, a depender do tipo de disfarce. Embora todas as estratégias tenham impactado negativamente a acurácia perceptiva, os autores observaram diferentes graus de eficácia entre elas. As estratégias baseadas em hipernasalidade e na redução da velocidade de fala mostraram-se menos efetivas. Segundo os autores, a primeira depende fortemente das características anatômicas do falante, o que limita sua capacidade de alterar a identidade vocal; já a segunda exerce pouca influência sobre os parâmetros espectrais do trato vocal, como as frequências de ressonância, que mantêm alto valor discriminativo na tarefa de Comparação de Locutores.

No início da década de 1980, contudo, a confiabilidade da identificação de locutor por meio de análise de *voiceprint* foi severamente questionada pela comunidade científica. Como consequência, esse método não resistiu ao teste do tempo e acabou sendo substituído por outras abordagens, como as baseadas na fonética acústica.

Sob essa abordagem, a influência de certos disfarces na configuração formântica vocálica foi abordada em [13]. Essas abordagens são particularmente interessantes porque estudos recentes de base acústica demonstram o elevado poder discriminatório das frequências de formantes, ainda que sensível a variações de estilo de fala e às condições de comparação entre amostras [14-16].

O estudo em [13] envolveu a imitação da voz de um comentarista político; o disfarce por meio do uso de regionalismos; um disfarce envolvendo mudança de gênero — em que os sujeitos emulavam uma voz feminina —; e o disfarce por meio da aplicação de um lenço à frente da boca. Os pesquisadores observaram que, de maneira geral, o primeiro e o segundo formantes caem drasticamente, independentemente do tipo de disfarce. Ainda segundo os autores, a variação dos formantes de ordem superior depende do tipo de disfarce. Assim, enquanto F3 e F4 apresentam diferenças significativas quando se imita a voz de um comentarista político ou uma voz feminina, os demais disfarces analisados no estudo não provocaram variação significativa. Uma observação interessante acerca do artigo é que disfarces por meio do falsete não foram analisados, uma vez que o programa utilizado confundia harmônicos com formantes.

Em um estudo longitudinal, com análise sincrônica [1], foram analisados os efeitos de três estratégias de disfarce— elevação da frequência fundamental, diminuição da frequência fundamental e produção de voz hiponasal por meio da compressão do nariz — com foco nas alterações da F0. O autor observou diferenças entre os sexos quanto à preferência pelas formas de disfarce: homens tendem a baixar a F0, enquanto mulheres preferem produzir uma voz nasalmente abafada, pressionando o nariz. Além disso, os homens elevam sua F0 de forma mais acentuada do que as mulheres; por outro lado, as mulheres produzem reduções de F0 mais expressivas do que os homens. De modo geral, o autor destaca que a produção de voz hiponasal não altera significativamente a F0; que a frequência fundamental pode ser recuperada com relativa facilidade quando o disfarce envolve o abaixamento da laringe; mas que, nos casos de elevação da F0, a recuperação da F0 normal do indivíduo torna-se consideravelmente mais difícil.

Os estudos revisados até aqui convergem ao indicar que os efeitos acústicos dos disfarces vocais são heterogêneos e dependem fortemente do tipo de manipulação envolvida — seja articulatória, fonatória ou de ressonância. Em conjunto, essas evidências indicam que os disfarces vocais não apenas afetam a percepção — inclusive de ouvintes treinados —, mas também produzem efeitos acústicos mensuráveis, cuja magnitude varia conforme a natureza da manipulação. Em consequência, tanto abordagens auditivas quanto análises espectrográficas e demais métodos acústicos enfrentam desafios específicos, os quais variam conforme a estratégia adotada pelo locutor.

Nas últimas décadas, a literatura tem direcionado atenção crescente à análise dos efeitos do disfarce vocal sobre sistemas de reconhecimento de locutores supervisionados por peritos[2].Em um desses estudos, os pesquisadores analisaram dez estratégias distintas em 20 participantes do sexo masculino, utilizando tal sistema [17]. As estratégias incluíram manipulações da F0 (elevação e abaixamento), variação da velocidade de fala (aumento e redução), sussurro, nasalidade, uso de máscara, *bite block*, objeto na boca (pastilha elástica) e simulação de sotaque estrangeiro. Com exceção do sotaque simulado, todas as estratégias afetaram significativamente o desempenho do sistema, reduzindo a acurácia de reconhecimento. O uso de máscara, elevação de F0 e sussurro mostraram-se os disfarces mais impactantes. Notavelmente, o estudo demonstrou que tanto a variabilidade intra quanto inter-falante persiste sob disfarce, indicando que não há uniformidade no efeito das estratégias — sua eficácia depende da interação entre o tipo de manipulação e as características individuais do locutor.

Em linha semelhante, um estudo baseado em uma amostra extensa de 100 falantes alemães demonstrou que disfarces clássicos — como o aumento ou abaixamento da frequência fundamental e o bloqueio da cavidade nasal — afetam a performance do sistema apenas de forma marginal, quando a população de referência também apresenta esse mesmo tipo de disfarce [18]. Entretanto, quando o sistema é calibrado com amostras naturais, as

[2] Frequentemente referidos, embora de forma imprecisa, como sistemas automáticos.

taxas de degradação aumentam sensivelmente, chegando a até 70% de falhas significativas no caso do disfarce nasal.

Esses achados indicam que, em abordagens automáticas supervisionadas, o disfarce deve ser tratado como uma fonte adicional de variabilidade — no caso, uma variabilidade dependente do locutor [19]. Assim, o sistema deve ser treinado e avaliado com gravações que incluam amostras produzidas sob estratégias de disfarce [20]. À semelhança do que ocorre em contextos sem disfarce, abordagens híbridas — que combinem análise fonética e sistemas automáticos — configuram o padrão de excelência na análise fonético-forense.

Grande esforço tem sido empreendido no sentido de definir o fenômeno do disfarce vocal. Dada a ampla variedade de estratégias envolvidas, parte desses esforços tem se dedicado à sua sistematização em tipologias. Na seção seguinte, apresentam-se as definições mais utilizadas do disfarce vocal e a tipologia mais amplamente adotada na literatura.

2.1. Definição e Tipologia

Lato sensu, conforme a classificação proposta em [1,21], as alterações vocais observadas em contextos periciais podem ser deliberadas ou não, resultando tanto de manipulações fisiológicas do trato vocal quanto do uso de técnicas de processamento digital de sinal ou de algoritmos de inteligência artificial.Alterações não deliberadas, não mediadas por tecnologia, costumam estar associadas a condições fisiológicas ou emocionais, como infecções respiratórias, obstruções nasais, fadiga vocal, estados ansiosos ou uso de substâncias como álcool e cigarro [1]. Já as alterações não deliberadas mediadas por tecnologia correspondem a distorções introduzidas de forma não intencional durante o processo de transmissão ou gravação do sinal de voz. Essas distorções podem ser provocadas por compressão de dados, limitações de largura de banda — como ocorre em chamadas telefônicas —, ruído de fundo ou instabilidades em conexões digitais.

As alterações deliberadas podem ocorrer com ou sem o uso de tecnologia. No primeiro caso, incluem-se estratégias que envolvem o uso intencional de modulação vocal artificial — por meio de *voice changers* — ou clonagem por inteligência artificial, cada vez mais acessíveis por meio de aplicativos ou plataformas de síntese neural. No segundo, encontram-se os disfarces produzidos por manipulações fisiológicas conscientes, como o uso de *creaky voice*, falsete, sussurro, emulação de sotaques estrangeiros ou de dialetos regionais, entre outras possibilidades [1,21].

São diversas as definições de disfarce vocal na literatura. Para Francis Nolan [22], por exemplo, trata-se da "exploração da flexibilidade do trato vocal com vistas à produção de um efeito comunicativo específico — que, no contexto forense, se orienta à ocultação ou à distorção de características relevantes à identificação do falante". Em contraste, abordagens mais abrangentes [21], definem o disfarce como "qualquer alteração, distorção ou desvio em relação ao padrão vocal habitual, independentemente de sua causa". Outras abordagens propõem uma delimitação mais restrita, ao caracterizá-lo como uma "modificação voluntária de características da voz, da fala e da linguagem, realizada pelo falante com o objetivo de ocultar sua identidade" [1].Sob essa definição, o disfarce vocal pode ser compreendido como a modificação voluntária das características acústicas da voz, realizada pelo próprio falante por meio de ajustes fisiológicos ou articulatórios, com o objetivo de dificultar ou inviabilizar sua identificação.Nessa delimitação, excluem-se tanto as modificações involuntárias — como a rouquidão decorrente de laringite ou a hiponasalidade associada a quadros gripais — quanto aquelas resultantes de fatores emocionais ou farmacológicos, bem como as alterações mediadas exclusivamente por recursos tecnológicos, ainda que intencionais. Da mesma forma, não se incluem os desvios decorrentes do próprio procedimento de coleta, os quais, em regra, não são intencionais, mas refletem fatores como o contexto comunicativo, os efeitos do paradoxo do observador [23] e a consciência, por parte do falante, de que sua produção linguística está sendo analisada.

A Tabela 1, adaptada de [1], [21] e [24], sumariza os tipos de disfarces que não se valem de meios eletrônicos normalmente encontrados em fonética forense.

Tabela 1 – Tipos frequentes de disfarce na análise forense da fala não mediados por tecnologia

Categoria	Estratégia de disfarce
Características da fonte	Elevação da frequência fundamental (com ou sem mudança de registro)
	Abaixamento da frequência fundamental
	Creaky voice / vocal fry (voz crepitante)
	Rouquidão artificial
	Uso das pregas vocais falsas – voz diplofônica
	Fala sussurrada
Características do filtro	Objeto estranho introduzido no trato vocal (ex.: caneta)
	Uso de ressonador adicional (ex.: lata de cerveja próxima à boca)
	Hiponasalidade / hipernasalidade (principalmente hiponasalidade, por pinçamento nasal)
	Obstrução com tecido sobre a boca (ex.: lenço)
Linguagem	Alteração deliberada de traços fonético-fonológicos e prosódicos associados à identidade dialetal
	Uso de outro dialeto da mesma língua
	Simulação de sotaque estrangeiro
Modo de falar	Redução da variação melódica (monotonia artificial)
	Aumento da variação melódica
	Alteração no ritmo de fala (geralmente lentificação)
	Modificação dos padrões acentuais (ex.: "voz robótica")

Observa-se que a tipologia apresentada na Tabela 1 reflete uma concepção segmentada da produção da voz, ao distinguir entre estratégias que afetam a "fonte" e aquelas que interferem no "filtro". Nesse enquadramento, disfarces vocais baseados na manipulação da frequência fundamental (F_0), como o seu aumento ou redução, são classificados como modificações das características da fonte glótica. Tal categorização, no entanto, revela-se conceitualmente limitada, uma vez que o disfarce vocal configura, em essência, uma alteração da qualidade de voz em seu sentido mais amplo. Conforme proposto em [3], a qualidade vocal não deve ser compreendida apenas como produto da vibração das pregas vocais, mas como resultado de ajustes coordenados — ou *settings* — que envolvem todo o trato vocal, desde a laringe até as estruturas supraglóticas.

Na seção seguinte, descrevem-se alguns ajustes articulatórios descritos no arcabouço teórico da qualidade de voz de Laver (1980) se relacionam com a modulação da frequência fundamental (F_0) empregada como estratégia de disfarce.

3. OS *SETTINGS* LONGITUDINAIS DE QUALIDADE DE VOZ

Embora o termo qualidade de voz às vezes seja utilizado para descrever propriedades supralaríngeas da fala, como a nasalidade, ele é frequentemente empregado em um sentido mais restrito, referindo-se exclusivamente ao tipo de fonação, como nos casos de voz crepitante ou soprosa [25–26]. No entanto, conforme definido em [27], qualidade de voz compreende "aquelas características que estão presentes, em maior ou menor grau, durante todo o tempo em que uma pessoa está falando (...) uma qualidade quase permanente que permeia todos os sons produzidos". Essa concepção mais abrangente é considerada mais adequada por muitos autores, por se alinhar melhor ao uso cotidiano do termo "voz"., em comparação com definições mais restritivas, que a reduzem ao tipo de fonação [28-29].

Alguns componentes da qualidade de voz decorrem de características orgânicas individualizadas dos órgãos da fala. A massa das pregas vocais, o comprimento do trato vocal, o comprimento traqueal, o tamanho da mandíbula e da língua, bem como o volume da cavidade nasal, enquadram-se nessa categoria e podem fornecer informações sobre idade, sexo, constituição física e saúde. Outros aspectos da qualidade de voz derivam da maneira como o falante utiliza habitualmente os órgãos vocais durante a fala. Esses aspectos podem incluir padrões sociofonéticos adquiridos pela influência de uma determinada comunidade de fala, efeitos emocionais e psicológicos ou outros padrões de fala puramente idiossincráticos. Nesse sentido, a qualidade de voz apresenta caráter indexical [30], funcionando como um índice de propriedades biológicas, psicológicas e sociais do falante, ao refletir tanto características de base orgânica quanto ajustes resultantes de hábitos articulatórios e fonatórios, além de influências de natureza psicológica e sociolinguística.

Laver propôs um modelo fonético de qualidade de voz baseado no conceito de "ajustes" (*settings*) dos órgãos da fala [3]. Nesse quadro, o ajuste constitui a unidade analítica central, definida como uma postura de curto ou longo termo que se mantém durante a fala e influencia a qualidade dos segmentos fônicos. Tais ajustes correspondem a padrões habituais de tensão muscular no sistema de produção da fala, os quais determinam uma qualidade de voz específica.

Esses ajustes podem ainda ser classificados em supralaríngeos e laríngeos, que constituem a maior parte das configurações observáveis, além dos ajustes globais de tensão, geralmente limitados às categorias tenso e relaxado. Todos são definidos em relação a um estado de referência arbitrariamente estabelecido, denominado ajuste neutro. Esse estado é caracterizado por um trato vocal cujo viés articulatório tende a uma área de secção transversal aproximadamente uniforme ao longo de sua extensão, sem alongamento ou encurtamento significativo, seja na região dos lábios ou da laringe. Nessa condição, a nasalidade ocorre apenas quando linguisticamente requerida, e as pregas vocais vibram de forma regular e eficiente.

As configurações de qualidade de voz que constituem o perfil vocal de um indivíduo podem ser sistematicamente avaliadas por meio do Voice Profile Analysis (VPA), conforme proposto em [29]. Esse sistema foi posteriormente adaptado para aplicações em fonética forense, a exemplo do protocolo desenvolvido por linguistas da Universidade de York, o qual possibilita a análise de um conjunto abrangente de parâmetros supralaríngeos, fonatórios, velofaríngeos e de tensão muscular, totalizando 32 dimensões descritivas [31].

Em seu construto teórico, Laver estabeleceu duas categorias de ajustes do trato vocal: ajustes latitudinais e ajustes longitudinais. Segundo o autor, os ajustes latitudinais do trato vocal supralaríngeo envolvem tendências quase permanentes de manter um determinado efeito constritivo (ou expansivo) sobre a área de secção transversal correspondente ao trato vocal em posição neutra. Esses ajustes envolvem modificações nos articuladores, como os lábios (ajuste labial), a língua (ajuste lingual), a região do istmo orofaríngeo (ajuste faucal), a faringe (ajuste faríngeo) e a mandíbula (ajuste mandibular). Por sua vez, os ajustes longitudinais dizem respeito ao alongamento ou encurtamento do trato vocal ao longo de seu eixo anteroposterior.

De acordo com Laver [3], os ajustes longitudinais do trato vocal podem ser realizados de, no mínimo, quatro formas, sendo que as duas primeiras envolvem,

respectivamente, a elevação e o abaixamento da laringe[3]. Tais movimentos modificam significativamente a geometria do trato vocal, alterando sua extensão e, por consequência, a posição dos formantes, o que afeta diretamente a qualidade perceptiva da voz.

A elevação da laringe a partir de sua posição neutra pode ocorrer por duas vias: pela estabilização do osso hioide, seguida da tração da laringe pelo músculo tireo-hióideo, ou pelo deslocamento conjunto do complexo hioide-laringe por ação dos músculos supra-hióideos, que promovem sua ascensão em bloco [32,33].

Por outro lado, a laringe pode ser rebaixada pela ação do grupo de músculos infra-hióideos, que conectam o osso hioide ao esterno e às escápulas, permitindo o deslocamento inferior do sistema laríngeo.

Tanto a elevação quanto o rebaixamento da laringe, ao modificarem o comprimento da laringofaringe, produzem alterações acústicas sistemáticas que afetam simultaneamente a frequência fundamental (F0) [3] e as frequências formânticas [34], com impacto direto na acurácia de tarefas de identificação de locutor. Longe de constituírem variações superficiais, tais modificações refletem uma reorganização efetiva da configuração do trato vocal, evidenciando a natureza estrutural do fenômeno.

O abaixamento da laringe reduz sistematicamente as frequências formânticas, com destaque para aquelas associadas à cavidade posterior. Segundo [34], para cada centímetro de abaixamento da laringe, a frequência do primeiro formante (F1) pode ser reduzida entre 5% e 6% para a maioria das vogais. O segundo formante (F2) pode apresentar reduções de até 8% em vogais anteriores, como [i], em função de sua associação com ressonâncias da cavidade posterior, enquanto o quarto formante (F4) apresenta uma queda média de aproximadamente 5%. Em contraste, o terceiro formante (F3) mostra-se relativamente insensível a esse tipo de manipulação para a maior parte das vogais analisadas, à exceção de /u/.

Nesse sentido, nas estratégias de disfarce baseadas na modulação da frequência fundamental (F0), as variações observadas nos formantes podem ser interpretadas como manifestação direta dos ajustes longitudinais do trato vocal descritos no arcabouço teórico de Laver [3].

Como se verá na próxima seção, essas variações formânticas podem ser utilizadas para estimar, com alta acurácia, o comprimento do trato vocal, com base nas propriedades de impedância acústica observadas nos lábios.

4. DA ESTIMATIVA DO COMPRIMENTO DO TRATO VOCAL

4.1. Da Equação do Quarto de Onda

Para descrever o comportamento acústico de um tubo com variação espacial de área — como ocorre no trato vocal humano — é possível utilizar uma formulação baseada nas leis de conservação da massa e da quantidade de movimento, uma vez que a velocidade de volume U(x,t) e a pressão p(x,t) para uma onda unidimensional em um tubo acústico são relacionadas entre si por equações derivadas das leis de Newton e considerações sobre compressibilidade [35]. Considerando uma propagação unidimensional ao longo do eixo x, com perturbações acústicas de pequena amplitude, obtêm-se as seguintes equações diferenciais lineares para a pressão acústica p(x,t) e a velocidade de volume U(x,t):

$$\frac{\partial p}{\partial x} = -\frac{\rho}{A}\frac{\partial U}{\partial t} \quad (1)$$

$$\frac{\partial U}{\partial x} = -\frac{A}{\gamma P_0}\frac{\partial p}{\partial t} \quad (2)$$

Nas equações acima, A é a área transversal do tubo, P_0 é a pressão ambiental, ρ é a densidade do ar e γ é a razão entre o calor específico do ar à pressão constante e o calor específico do ar à volume constante, o que, no ar, resulta em 1.4. Podemos assumir que a dependência no tempo se dá de forma exponencial, de forma que podemos escrever

$$\mathrm{p(x,t)} = p(x)e^{j2\pi ft} \quad (3)$$

$$\mathrm{U(x,t)} = U(x)e^{j2\pi ft} \quad (4)$$

em que f é a frequência. Desta forma, as equações (1) e (2) podem ser reescritas como:

$$\frac{\mathrm{dp}}{\mathrm{dx}} = -\frac{\mathrm{j}2\pi f\rho}{A}U \quad (5)$$

$$\frac{\mathrm{dU}}{\mathrm{dx}} = -\frac{\mathrm{j}2\pi fA}{\gamma P_0}p \quad (6)$$

Isolando U nas equações (5) e (6), obtemos a seguinte equação, conhecida como *Webster horn equation* [36]

$$\frac{d^2p}{dx^2} + \frac{1}{A}\frac{dA}{dx}\frac{dy}{dx} + k^2\mathrm{p} = 0 \quad (7),$$

onde $k = 2\pi f/c$ e

$$c = \sqrt{\frac{\gamma P_0}{\rho}}$$

[3] O terceiro ajuste diz respeito à protrusão labial, enquanto o quarto envolve a elevação e a retração do lábio inferior, caracterizando a voz labiodentalizada.

Considerando-se as condições típicas do ar ambiente, adota-se c = 3.5 x 10^4 cm/s (350 m/s). As frequências naturais de um tubo acústico com A(x) de função de área são os valores da frequência f para as quais a equação (7) possui soluções, sujeitas às condições de contorno.

No caso, temos impedância nula na boca e impedância infinita na glote, ou seja, U (-l) =0 e p (0) = 0 e também $\frac{dp}{dx}$ = 0 em x= - l (de (5)).

4.1.1 Frequência em um tubo uniforme

Em um tubo uniforme, como o da Figura 1 $\frac{dA}{dx}$ = 0, de forma que a equação se resume a

$$\frac{d^2p}{dx^2} + k^2 p = 0 \quad (8),$$

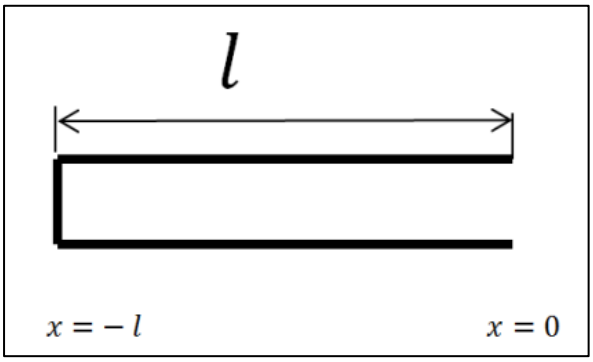

Figura 1 - Tubo acústico aberto em uma extremidade e fechado em outra.

A solução desta equação é p(x) = $P_m$sinkx, onde P_m representa o maior volume da pressão. Da equação (5), obtém-se a seguinte relação entre pressão e velocidade de volume:

$$U(x) = jP_m \frac{A}{\rho c} \cos kx \quad (9)$$

Sendo U(-l) = 0, então cos kl = 0 , de tal forma que k l = (2n-1)$\frac{\pi}{2}$, de tal forma que a frequência no tubo é

$$F_n = \frac{2n-1}{4} \frac{c}{l} \quad (10)$$

O diagrama articulatório apresentado na Figura 1 representa a configuração do trato vocal durante a produção da vogal [ə], sendo a extremidade fechada correspondente à glote e a extremidade aberta aos lábios. A expressão (10) demonstra que as frequências dos formantes variam em função do comprimento do trato vocal e, por conseguinte, refletem características anatômicas individuais [37].

É importante destacar, entretanto, que a proporcionalidade expressa pela Equação (10) é válida apenas sob a suposição de que o trato vocal seja uniforme em sua estrutura. Na realidade, há diferenças anatômicas significativas entre os grupos populacionais. Mulheres, por exemplo, tendem a apresentar uma abertura bucal proporcionalmente maior e uma cavidade faríngea relativamente menor em comparação aos homens. Crianças, por sua vez, não apenas têm trato vocal mais curto, mas também proporções distintas ao longo do tubo vocal. Essas variações impedem que se realizem normalizações acústicas precisas com base apenas em fatores de escala, como o simples alongamento ou encurtamento do trato vocal. Fant destacou que tais diferenças exigem modelos mais complexos que considerem a distribuição não uniforme das áreas transversais ao longo do trato [38]. A Figura 2, adaptada de [39], apresenta os valores médios do terceiro formante (F3) em função do comprimento do trato vocal para três indivíduos adultos. O terceiro formante foi escolhido porque, segundo o autor, varia menos em função das vogais do que o primeiro e o segundo formantes, oferecendo, assim, uma indicação mais precisa do comprimento do trato vocal do que aquela baseada em F1 ou F2.

A curva em azul escuro representa os dados empíricos obtidos a partir da produção de vogais, sendo o comprimento do trato vocal medido por meio de radiografias [40], enquanto a linha tracejada ilustra a variação esperada caso todas as dimensões do trato vocal fossem escaladas proporcionalmente ao seu comprimento total. Verifica-se que, embora exista uma relação inversa entre o comprimento do trato vocal e a frequência do terceiro formante, os valores empíricos não se alinham perfeitamente à reta teórica correspondente ao modelo de trato uniforme.

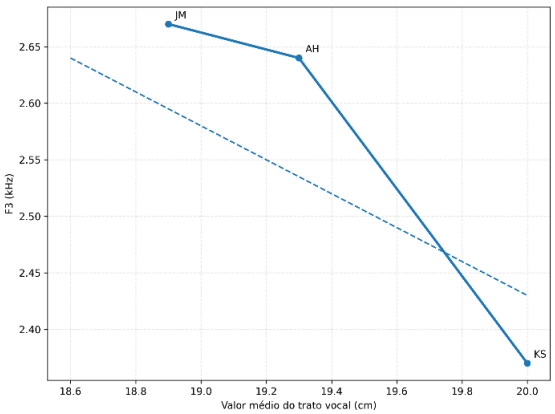

Figura 2 - Valores médios do terceiro formante (F3) em função do comprimento do trato vocal (adaptado de Fant [39])

4.2. Da Extensão da Equação do Quarto de Onda

A Equação (10) apresenta limitações ao assumir um trato vocal uniforme, não contemplando a complexidade articulatória associada a vogais não centrais. Propõe-se, portanto, uma formulação expandida que integra informações provenientes das frequências formânticas. A estimativa do comprimento do trato vocal é, assim, obtida a partir dessas frequências, com base em uma medida global derivada da impedância acústica observada nos lábios, conforme o modelo proposto em [41]. Os autores de [41] formularam o problema a partir da estrutura de polos e zeros da impedância de entrada do trato vocal, representando-o como uma cadeia de tubos acoplados em cascata. Os detalhes matemáticos apresentados a seguir têm por finalidade fundamentar a estimativa física do CTV_e utilizada nas análises experimentais. Assume-se que, embora existam infinitas configurações geométricas capazes de produzir os mesmos valores de formantes, é possível restringir a solução a uma distribuição de área A(x) que se aproxime, tanto quanto possível, de um tubo uniforme. Essa aproximação é obtida por meio da minimização da seguinte função de erro logarítmica:

$$\varepsilon[\mathrm{L}] = \frac{1}{L}\int_0^L |\ln A(x)|^2 \qquad (11)$$

A função de erro dada pela Equação (11)variações abruptas na área do trato vocal e seleciona, entre as soluções compatíveis com os formantes observados, aquela mais próxima de um tubo uniforme, isto é, mais fisiologicamente plausível. Nesse contexto, define-se uma frequência de referência f_0, que sintetiza globalmente a estrutura espectral do trato vocal. A expressão (12) estima f_0 como uma média ponderada dos zeros da impedância acústica, ajustada aos múltiplos ímpares ideais de um tubo uniforme:

$$f_0 = f_1 \frac{\Sigma\left(\frac{f_k}{kf_1}\right)^2}{\Sigma\left(\frac{f_k}{kf_1}\right)} \qquad (12)$$

Dessa forma, sob a hipótese de trato quase uniforme e múltiplos ímpares, a minimização de (11) conduz a um estimador global f_0 dado por (13):

$$f_0 = \frac{\Sigma\left(\frac{F_i}{(2\mathrm{i}-1)}\right)^2}{\Sigma\left(\frac{F_i}{(2\mathrm{i}-1)}\right)} \qquad (13)$$

do qual se obtém o comprimento do trato vocal estimado:

$$\mathrm{CTVe} = \frac{c}{4f_0} \qquad (14)$$

A equação (14) pode ser interpretada como uma extensão da equação do quarto de onda (Equação 10), ao incorporar informações provenientes de múltiplos formantes, em vez de depender exclusivamente da frequência fundamental associada ao modo de ressonância dominante. Desse modo, os valores de CTVe podem ser estimados diretamente a partir dos valores médios dos formantes, tornando o modelo aplicável a diferentes configurações vocálicas, e não apenas à vogal neutra schwa [ə] [42].

A equação (14) foi avaliada com base nos dados de formantes das vogais reportados em [43], amplamente aceitos como referência experimental para estimativas do comprimento do trato vocal. Os valores de CTVe obtidos por meio da formulação proposta apresentaram forte concordância com as medidas empíricas, com erros percentuais inferiores a 5%. Especificamente, para a vogal /a/, o comprimento do trato vocal reportado por Fant foi de 17,0 cm, enquanto o modelo forneceu uma estimativa de 17,3 cm (erro de +1,8%). Para a vogal /i/, o valor de referência foi de 16,5 cm, com estimativa de 16,4 cm (erro de −0,6%). Por fim, para a vogal /u/, obteve-se uma estimativa de 19,0 cm, em comparação com os 19,5 cm reportados por Fant (erro de −2,6%) [41]. Esses resultados, por sua vez, indicam que a formulação proposta é capaz de capturar, com boa precisão, as variações do comprimento efetivo do trato vocal.

O uso da Equação (14) foi demonstrado em [42] e posteriormente aprofundado em [43–45]. No experimento descrito na próxima seção, essa formulação será empregada para avaliar a variação do CTVe quando os falantes utilizam estratégias de disfarce vocal baseadas na modulação da frequência fundamental, cujo correlato perceptivo é a altura tonal.

5. DO EXPERIMENTO

O estudo piloto teve como objetivo avaliar a estimativa do CTV como parâmetro para a análise de estratégias de disfarce vocal baseadas em ajustes longitudinais da laringe — notadamente, variações na altura laríngea. Participaram do experimento seis indivíduos do sexo masculino, falantes do português brasileiro padrão e com formação superior, com idades de 25, 27, 32, 42, 45 e 55 anos. Nenhum deles possuía formação em Linguística. Os participantes identificados como L1, L2 e L3 tinham conhecimentos em técnicas de canto. Nenhum participante era fumante, nem apresentava queixas vocais ou histórico de patologias laríngeas. As gravações foram realizadas em ambiente controlado, sob condições acústicas estáveis, visando minimizar fontes externas de variabilidade.

Conforme proposto em [46], a espontaneidade da fala está diretamente relacionada ao grau de intervenção do experimentador na eliciação. Nesse sentido, as amostras obtidas neste estudo, por terem sido produzidas sob

condições experimentais controladas, com um grau extremo de intervenção, não podem ser classificadas como amostras de fala espontânea. Por outro lado, cabe destacar que, diferentemente da maioria dos estudos em fonética forense, nos quais se busca a fala espontânea por ela se aproximar do vernáculo — definido em [23] como o estilo de fala mais natural e isento de monitoramento consciente —, nas investigações envolvendo disfarce, a espontaneidade deixa de ser o objetivo principal. Isso porque o disfarce caracteriza-se justamente como um afastamento deliberado do vernáculo: trata-se de uma produção intencionalmente modificada, com o propósito de ocultar a identidade do falante. Nesses casos, o foco do analista desloca-se da busca por naturalidade para a identificação e descrição dos desvios articulatórios e prosódicos empregados pelo locutor como estratégia de disfarce. Ademais, é comum que o material questionado esteja associado à leitura de roteiros, textos decorados ou mesmo à fala dirigida por terceiros. Em tais situações, observa-se que a ameaça, o insulto ou a exigência extorsiva pode ser vocalizada por um membro distinto daquele que detém o interesse direto na ação delitiva, recorrendo-se, muitas vezes, a uma leitura instrumentalizada, num estilo distante do vernáculo.

Diante desse cenário, e visando a maior controle experimental, optou-se pela produção das sete vogais orais do português brasileiro em estrutura fonética e prosódica padronizada. As vogais foram inseridas no núcleo tônico de palavras paroxítonas trissílabas com estrutura CVCVCV, as quais foram integradas à frase veículo “Digo _____ baixinho”. Essa construção buscou assegurar uniformidade prosódica e controle do contexto fonético.

Para cada vogal, selecionaram-se duas palavras, totalizando quatorze itens lexicais. Sempre que possível, foram escolhidos vocábulos comumente encontrados em contextos delitivos, incluindo gírias e expressões de baixo calão. Assim, por exemplo, para a vogal [ɔ], utilizou-se a frase “Digo pistola baixinho”; para [a], “Digo cadáver baixinho”. Na segmentação, a duração da vogal foi determinada entre o primeiro e o último pulso regular da vogal e os formantes foram calculados na porção estável da vogal, seguindo sempre as recomendações apresentadas em [47].

Antes da gravação, foi realizado um treino breve para garantir a compreensão das instruções. Cada sujeito leu as frases sob três contextos experimentais distintos: fala neutra, aumento de F0 e abaixamento de F0. No cálculo dos formantes, considerou-se a média dos valores obtidos em duas ocorrências de cada vogal. As gravações foram realizadas com equipamento de captação de alta qualidade, assegurando condições acústicas adequadas para a análise dos parâmetros fonético-acústicos relevantes. As coletas foram realizadas com o consentimento dos participantes, exclusivamente para fins de pesquisa.

5.1 Resultados e Discussão

Na Figura 3, observa-se o valor médio do Comprimento do Trato Vocal estimado (CTV_e) em função da modulação da F0 Os resultados mostram boa concordância com os valores reportados por [38], indicando um CTV_e médio de aproximadamente 17,5 cm, valor compatível com o esperado para locutores adultos do sexo masculino. A figura também evidencia que o CTV_e aumenta significativamente na condição de abaixamento de F0, alcançando valores próximos a 18,0 cm, e diminui na condição de aumento de F0, aproximando-se de 15,3 cm, tomando como referência a condição de F0 em fala neutra. Essas variações estão em consonância com os modelos anatômicos e acústicos da produção da fala, segundo os quais há uma relação entre os ajustes longitudinais da laringe e F0. Vale ressaltar que o valor do CTV_e constitui uma estimativa indireta, sendo influenciado por fatores como constituição física, sexo, idade e grupo étnico, além das características da população analisada [38], [48 – 49].

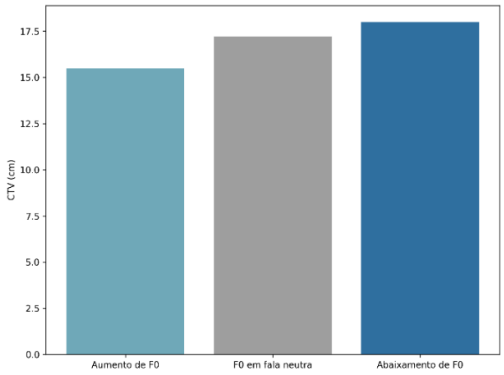

Figura 3 – Variação CTVe em função da modulação de F0

A análise restrita aos valores extremos de CTV_e (< 14 cm ou > 21 cm), apresentados na Figura 4, revelou uma correlação negativa quase perfeita com a F0 ($r = -0,97$). Esse resultado indica que, nesses limites anatômicos, há um vínculo extremamente forte entre o comprimento estimado do trato vocal e a frequência fundamental.

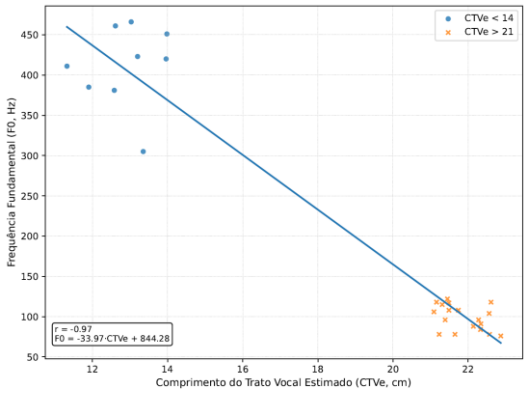

Figura 4 – Correlação entre o CTVe (para valores inferiores a 14cm e superiores a 21cm) e F0.

O gráfico da Figura 5 evidencia a variação sistemática do Comprimento do Trato Vocal estimado (CTV_e) em função da modulação da F0, analisada separadamente para cada um dos seis locutores (L1 a L6). Para todos os indivíduos, observa-se um padrão consistente: o CTV_e é menor na condição de aumento de F0, intermediário na condição de F0 em fala neutra e maior na condição de abaixamento de F0. Essa regularidade interindividual reforça a robustez do efeito anatômico associado à elevação e ao rebaixamento laríngeo, os quais promovem, respectivamente, o encurtamento e o alongamento do trato vocal, modificando de forma sistemática as frequências de ressonância envolvidas na produção da fala.

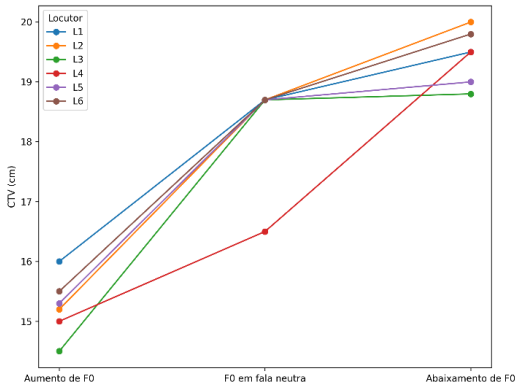

Figura 5 – CTVe Médio por Modulação de F0 para cada Locutor (L1 a L6)

Ainda que os valores absolutos variem entre os locutores — o que é esperado devido a diferenças fisiológicas individuais (como altura, estrutura craniofacial e idade) — , a direção da variação permanece constante. Esse resultado é compatível com os modelos clássicos da produção vocal descritos em [38]. Observa-se um aumento progressivo do CTV_e da condição de aumento de F0 para a condição de abaixamento de F0, tendo como referência a F0 em fala neutra, refletindo diretamente o alongamento do trato vocal associado ao rebaixamento laríngeo.

A análise de correlação entre o CTV_e e os formantes F1 a F4 é apresentada na Figura 6. Observa-se uma correlação negativa consistente entre o CTV_e e as frequências dos formantes, em correspondência com o padrão já evidenciado no estudo experimental apresentado na Figura 2. Os coeficientes de correlação de Pearson calculados foram: F1 (r = -0.585), F2 (r = -0.696), F3 (r = -0.523) e F4 (r = -0.727). Esses resultados indicam que alterações no comprimento do trato vocal afetam todos os formantes de forma sistemática e mensurável.

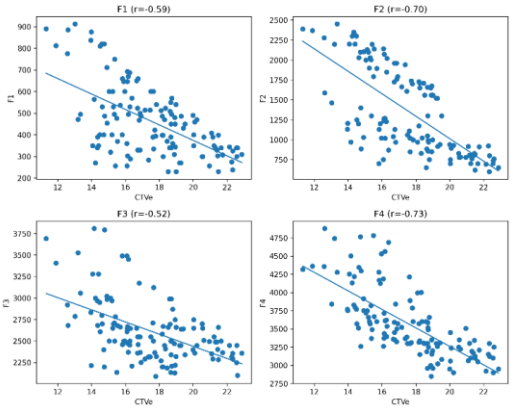

Figura 6- Correlação entre o CTV_e e os formantes

Na Figura 7, apresenta-se a variação média do comprimento do trato vocal estimado (CTV_e) por vogal, tomando-se como referência a posição neutra ($\Delta CTV_e = 0$). Os resultados indicam que, para a maioria das vogais, há um aumento consistente do CTV_e na condição de diminuição de F0 e uma redução na condição de aumento de F0, o que evidencia a influência direta da altura laríngea sobre a configuração do trato vocal. Note-se, contudo, que a magnitude dessas variações não é uniforme entre as vogais, sugerindo a interação entre ajustes laríngeos e configurações articulatórias supralaríngeas. Esse comportamento sugere que, ao alterar a posição da laringe, os falantes ajustam outros articuladores — como a língua, os lábios e a abertura da boca — para compensar essas mudanças e manter a produção da vogal.

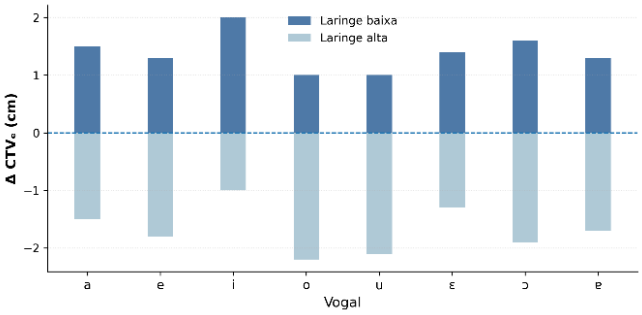

Figura 7 – Variação Média do CTVe por Vogal (Referência = Posição Neutra)

O gráfico da Figura 8 mostra a relação entre a dispersão de formantes (D*f*) e o comprimento do trato vocal estimado (CTV_e). Cada ponto representa uma amostra em que foram mensurados os quatro primeiros formantes (F1 a F4), e o valor de D*f* definida em [50] por:

$$Df = \frac{1}{N-1}\sum_{i=1}^{N-1}\left(F_{(i+1)} - F_i\right) \qquad (21)$$

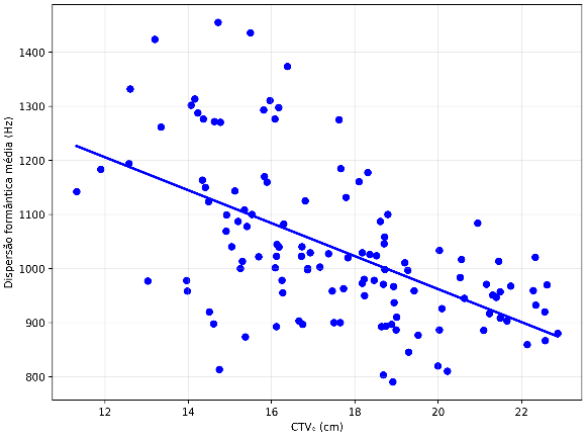

Figura 8 – Dispersão dos formantes X CTVe

Conforme discutido por [50], a relação entre a dispersão dos formantes e o comprimento do trato vocal está associada à percepção do tamanho corporal do falante. Nesse contexto, o sistema auditivo utiliza pistas espectrais para inferir propriedades físicas, como o comprimento do trato vocal, geralmente correlacionado ao porte físico. Essa associação pode ser explorada em estratégias de disfarce vocal, nas quais o locutor ajusta sua configuração vocal de modo a sugerir maior estatura ou dominância.

Na Figura 9 apresentamos, para o falante L1, a relação entre o comprimento estimado do trato vocal e as frequências formânticas. Os resultados indicam que os elevados coeficientes de variação de F1 e F2, por um lado, e os baixos valores de CV para F3 e F4, por outro, são compatíveis com achados da literatura [34], [39].

Ao discutir estratégias de abaixamento laríngeo, Lindblom & Sundberg [34], mostram que o terceiro formante (F3) é amplamente insensível a esse tipo de manipulação para a maioria das vogais. De modo semelhante, [40] assinala que, para alguns falantes, há um quarto ou quinto formante relativamente estável que se mantém ao longo de diferentes configurações vocálicas. De acordo com [39], uma ressonância desse tipo pode ocorrer na região laríngea do trato vocal, entre as pregas vocais e o ponto em que o tubo laríngeo se abre para a cavidade faríngea mais ampla (orofaringe). No presente estudo, para L1, F3 (CV = 8,5%) apresenta menor variabilidade relativa do que F4 (CV = 9,3%), e de acordo com [34], esse efeito de redução da distância espectral entre F3 e F4, contribui para a percepção de uma "voz mais encorpada".

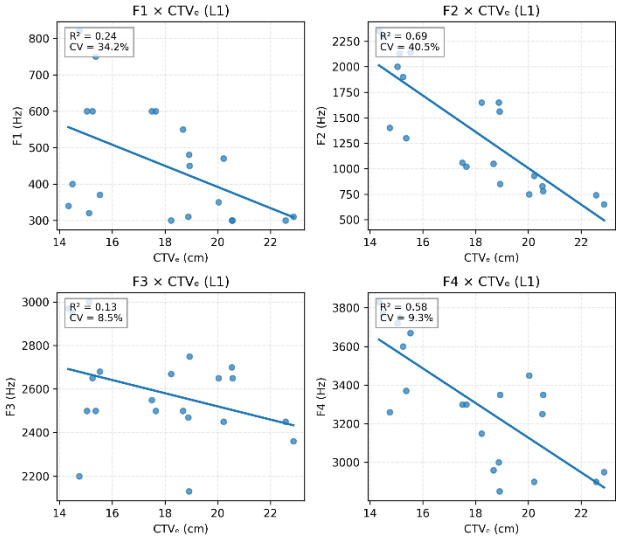

Figura 9 – Gráficos de dispersão com ajuste de regressão linear dos formantes F1–F4 em função do comprimento estimado do trato vocal (CTV_e) para o falante L1, considerando todas as vogais e condições experimentais.

Nas Figuras 10 e 11 observamos o polígono vocálico associado às diferentes posições laríngeas para o locutor L1. Nota-se que o espaço vocálico correspondente à condição de elevação de F0 (representado em azul) está associado a ajustes articulatórios que ampliam o espaço acústico no plano F1 × F2. Especificamente, observa-se um aumento de F1 — frequentemente relacionado à maior abertura da boca — e de F2 — indicativo de avanços da língua no eixo anteroposterior. Esses deslocamentos sugerem que, na tentativa de elevar a frequência fundamental como estratégia de disfarce, o falante simultaneamente modifica a configuração do trato vocal, promovendo um alargamento do espaço vocálico. Por outro lado, na mesma Figura, o espaço vocálico correspondente à condição de abaixamento de F0 (representado em vermelho) está associado a ajustes articulatórios que comprimem o espaço acústico no plano F1 × F2. Especificamente, observa-se uma redução nos valores de F1 — comumente associada a menor abertura da boca — e de F2 — que pode refletir retração da língua no eixo anteroposterior. Esses deslocamentos sugerem que, na tentativa de reduzir a frequência fundamental como estratégia de disfarce, o falante também altera a configuração do trato vocal, resultando em uma contração do espaço vocálico. Esse padrão é recorrente entre os falantes, como evidenciado inicialmente pelas representações do espaço vocálico do sujeito L2 (Figura 11).

Vale notar que, mesmo diante dessa tendência de expansão, os limites anatômicos do trato vocal impõem restrições à dispersão dos formantes [3,51]. O polígono vocálico não cresce indefinidamente: vogais como /i/ e /u/ mantêm seus valores relativamente estáveis, sugerindo que, embora os ajustes laríngeos permitam modulações expressivas no espaço acústico, essas variações ocorrem

dentro de uma faixa controlada pelas propriedades físicas do trato vocal de cada falante. Dessa forma, as Figuras mostram que, em disfarces vocais baseados na elevação da frequência fundamental (F0), observa-se uma tendência de expansão do espaço vocálico em relação à posição neutra — especialmente nas vogais médias e anteriores. Esse padrão é consistente entre os falantes, com aumentos tanto de F1 quanto de F2. De modo análogo, nas estratégias que envolvem o abaixamento de F0, os locutores tendem a contrair o espaço vocálico nessas mesmas regiões. Essas alterações de configuração laríngea impactam diretamente na qualidade fonética percebida. A vogal /a/ (baixa), por exemplo, quando produzida na posição neutra da laringe, a depender do locutor, pode a apresentar uma qualidade próxima à de uma vogal média-baixa articulada na posição de laringe alta. Inversamente, a mesma vogal, quando produzida com laringe rebaixada, também pode adquirir características fonéticas similares às da posição neutra.

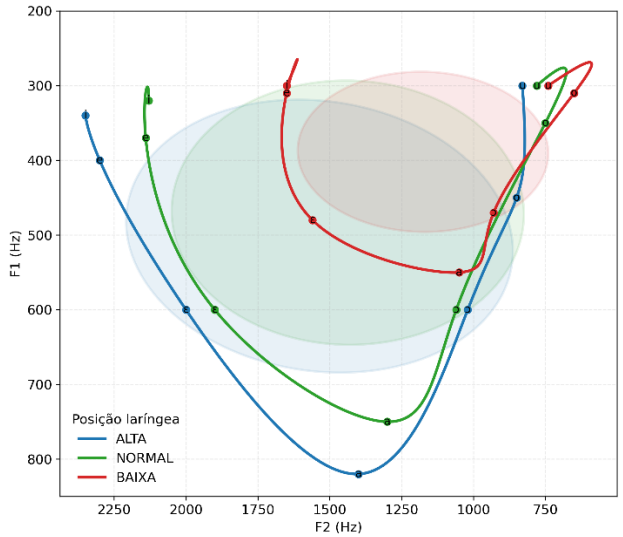

Figura 10 – Polígono vocálico L1 nas três posições da laringe.

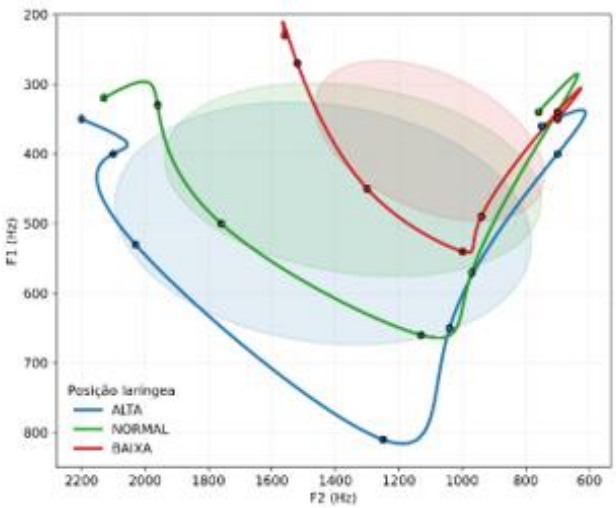

Figura 11 – Polígono vocálico L2 nas três posições da laringe.

As Figuras 10 e 11 ilustram o modelo proposto por Francis Nolan [51], segundo o qual a voz resulta da interação entre as restrições impostas pelas propriedades físicas do trato vocal e as escolhas realizadas pelo falante para atingir determinados objetivos comunicativos por meio dos recursos oferecidos pelos diferentes componentes de seu sistema linguístico. Nelas, as curvas suavizadas do espaço vocálico representam trajetórias associadas às escolhas articulatórias do locutor, enquanto as elipses de dispersão delimitam o campo de variabilidade permitido pelo sistema fonético.

A Tabela 2 apresenta a relação entre as áreas dos polígonos vocálicos associadas às estratégias de disfarce por modulação da F0, tomando como referência a posição neutra da laringe e considerando, além dos locutores L1 e L2, a média dos locutores analisados.

Tabela 2 – Relação entre as Áreas dos Espaços Vocálicos

Sujeito	Posição Laríngea	Área (Hz²)	Razão em relação à NEUTRA
L1	ALTA	474,650.00	1.19
L1	BAIXA	168,150.00	0.42
L1	NEUTRA	398,950.00	1.00
L2	ALTA	390,700.00	1.55
L2	BAIXA	118,800.00	0.47
L2	NEUTRA	252,400.00	1.00
MÉDIA	ALTA	431,890.39	1.43
MÉDIA	BAIXA	194,319.57	0.65
MÉDIA	NEUTRA	301,267.61	1.00

6. Considerações Finais

O presente artigo apresentou um estudo piloto dedicado à análise dos efeitos de estratégias de disfarce vocal por meio da modulação da frequência fundamental (F0), à luz do arcabouço de qualidade de voz proposto por John Laver [3]. A partir da reinterpretação da tipologia tradicional, apoiada em métricas como o comprimento estimado do trato vocal (CTV_e), os resultados indicam que, embora tais manipulações sejam convencionalmente descritas como modificações da fonte glótica [1], o disfarce vocal deve ser compreendido de forma mais abrangente como uma alteração da qualidade de voz, resultante de ajustes coordenados entre fonte e filtro.Nesse sentido, verificou-se que estratégias de elevação da F0 estão associadas não apenas ao aumento da tensão laríngea, mas também a um encurtamento efetivo do trato vocal, decorrente da elevação da laringe, frequentemente acompanhado por configurações articulatórias que promovem expansão do espaço vocálico no plano F1 × F2.

Por outro lado, a redução da F0 mostrou-se relacionada ao alongamento do trato vocal, associado ao rebaixamento laríngeo, com consequente compressão do espaço vocálico. Tais padrões são consistentes com a noção de *settings* latitudinais descrita por Laver [3][4], refletindo tendências sistemáticas de configuração do trato vocal.

Do ponto de vista fonético-forense, os resultados reforçam que a análise de disfarce vocal não deve se restringir a parâmetros da fonte. Em particular, evidenciou-se que formantes e medidas globais do trato vocal, como o CTV_e, devem ser considerados de forma integrada à F0. Assim, variações na frequência fundamental não constituem, isoladamente, critério suficiente para exclusão de locutor, especialmente quando os parâmetros do filtro permanecem dentro de padrões esperados para o indivíduo.

Adicionalmente, a observação de desacoplamento entre fonte e filtro — como em manipulações artificiais simples (e.g., *pitch shifting*) — requer análise cautelosa, uma vez que tais estratégias não reproduzem integralmente os ajustes fisiológicos envolvidos no disfarce natural, podendo indicar estratégias de manipulação não fisiológica.

Os achados corroboram a literatura científica da área ao indicar que os formantes superiores, particularmente F3 e F4, apresentam maior estabilidade relativa frente a estratégias de disfarce baseadas na modulação de F0. Tais resultados sugerem que esses parâmetros se configuram como indicadores mais robustos da geometria global do trato vocal, sendo, portanto, menos suscetíveis a variações articulatórias específicas.

Em conjunto, os resultados reforçam a necessidade de adoção, em casos de disfarce envolvendo modulações de F0, de abordagens integradas que articulem parâmetros locais de fonte e filtro, bem como a configuração global do trato vocal. Por fim, é importante notar que se trata de um estudo piloto, o que implica que investigações futuras, com amostras mais amplas e maior diversidade de perfis vocais, são necessárias para avaliar a generalização dos padrões aqui observados e refinar os critérios propostos para sua aplicação em contextos fonético-forenses.

[4] *"Latitudinal settings of the subpharyngeal vocal tract involve quasi-permanent tendencies to maintain a particular constrictive (or expansive) effect on the cross- sectional area (...) (Laver, 1980, pp.34").*

REFERÊNCIAS BIBLIOGRÁFICAS

[1] H.J. Künzel. Effects of voice disguise on speaking fundamental frequency. The International Journal of Speech, Language and the Law 7(2): 149-179 (2000).

[2] J. Clark; P. Foulkes. Identification of voices in electronically disguised speech. The International Journal of Speech, Language and the Law 14(2): 195-221 (2008).

[3] J. Laver. The Phonetic Description of Voice Quality. Cambridge University Press, Cambridge (1980).

[4] F. McGehee. The reliability of the identification of the human voice. The Journal of General Psychology 17(2): 249-271 (1937).

[5] A. Hirson; M. Duckworth. Glottal fry and voice disguise: a case study in forensic phonetics. Journal of Biomedical Engineering 15(3): 193-200 (1993).

[6] H. Hollien; W. Majewski; P.A. Hollien. Perceptual identification of voices under normal, stress, and disguised speaking conditions. The Journal of the Acoustical Society of America 56(S1): S53-S53 (1974).

[7] I. Wagner; O. Köster. Perceptual recognition of familiar voices using falsetto as a type of voice disguise. Proceedings of the 14th International Congress of Phonetic Sciences: 1381-1384 (1999).

[8] W. Koenig. Visual spectrography. Journal of the Acoustical Society of America 18: 19-49 (1946).

[9] R.K. Potter; G.G. Kopp; H.C. Green. Visible Speech. D. Van Nostrand Company, New York (1947).

[10] L.G. Kersta. Voiceprint identification. Nature 196: 1253-1257 (1962).

[11] W. Endres; W. Bambach; G. Flösser. Voice spectrograms as a function of speaker and phonetic context. Phonetica 23(2): 48-65 (1971).

[12] A.R. Reich; K.L. Moll; J.F. Curtis. Effects of selected vocal disguises upon spectrographic speaker identification. The Journal of the Acoustical Society of America 60(4): 919-925 (1976).

[13] P. Rose; R. Simmons. The acoustic characteristics of speaker disguise. Proceedings of the International Conference on Forensic Phonetics (1996).

[14] J.C. Cavalcanti; A. Eriksson; P.A. Barbosa; S. Madureira. Revisiting the speaker discriminatory power of vowel formant frequencies under a likelihood ratio-based paradigm: The case of mismatched speaking styles. PLOS ONE 19(12): e0311363 (2024).

[15] J.C. Cavalcanti; A. Eriksson; P.A. Barbosa. On the speaker discriminatory power asymmetry regarding acoustic-phonetic parameters and the impact of speaking style. Frontiers in Psychology 14: 1101187 (2023).

[16] J.C. Cavalcanti; A. Eriksson; P.A. Barbosa. Acoustic analysis of vowel formant frequencies in genetically-related and non-genetically related speakers with implications for forensic speaker comparison. PLOS ONE 16(2): e0246645 (2021).

[17] Z. Zhang; Z.-H. Tan. An evaluation of disguised speech on speaker recognition performance. Proceedings of the 9th International Conference on Signal Processing. IEEE (2008).
[18] H.J. Künzel; J. González-Rodríguez; J. Ortega-García. Automatic speaker recognition under voice disguise. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2004).
[19] J.H.L. Hansen; T. Hasan. Speaker recognition by machines and humans: A tutorial review. IEEE Signal Processing Magazine 32(6): 74-99 (2015).
[20] G.S. Morrison; E. Enzinger; C. Zhang. Forensic speech science. In: I. Freckelton; H. Selby (Eds.). Expert Evidence, Ch. 99. Thomson Reuters, Sydney, Australia (2018).
[21] R. Rodman; M. Powell. Computer recognition of speakers who disguise their voice. Proceedings of the International Conference on Signal Processing Applications and Technology (ICSPAT2000) (2000).
[22] F. Nolan. The Phonetic Bases of Speaker Recognition. Cambridge University Press, Cambridge (1983).
[23] W. Labov. Some principles of linguistic methodology. Language in Society 1(1): 97-120 (1972).
[24] H. Masthoff. A report on a voice disguise experiment. The International Journal of Speech, Language and the Law 3(1): 160-167 (1996).
[25] P. Foulkes. Current trends in British sociophonetics. University of Pennsylvania Working Papers in Linguistics 8(3): 75-86 (2002).
[26] R.J. Podesva. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. Journal of Sociolinguistics 11(4): 478-504 (2007).
[27] D. Abercrombie. Elements of General Phonetics. Edinburgh University Press, Edinburgh (1967).
[28] J. Kreiman; D. Sidtis. Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Wiley-Blackwell, Malden (2011).
[29] J.M. Beck. Perceptual analysis of voice quality: the place of vocal profile analysis. In: A Figure of Speech: 285-322. Routledge (2014).
[30] J. Laver. Voice Quality and Indexical Information. British Journal of Disorders of Communication 3(1): 43-54 (1968).
[31] E. San Segundo et al. Voice quality analysis in forensic voice comparison: developing the vocal profile analysis scheme. International Association of Forensic Phonetics and Acoustics (IAFPA), York: 24-27 (2016).
[32] H.M. Kaplan. Anatomy and Physiology of Speech. McGraw-Hill, New York (1960).
[33] C.G. Van Riper; J.V. Irwin. Voice and Articulation. Prentice-Hall, Englewood Cliffs, NJ (1958).
[34] B.E.F. Lindblom; J.E.F. Sundberg. Acoustical consequences of lip, tongue, jaw, and larynx movement. The Journal of the Acoustical Society of America 50(4B): 1166-1179 (1971).
[35] K.N. Stevens. Acoustic Phonetics. MIT Press (2000).
[36] P.M. Morse. Vibration and Sound. 2nd ed. McGraw-Hill Book Company, New York (1948).
[37] R.D. Kent; C. Read. Acoustic Analysis of Speech. 2nd ed. Singular Publishing Group, San Diego (2000).
[38] G. Fant. Acoustic Theory of Speech Production: With Calculations Based on X-ray Studies of Russian Articulations. Walter de Gruyter (1971).
[39] K.N. Stevens. Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. In: A. Rigault; R. Charbonneau (Eds.). Actes du Septième Congrès International des Sciences Phonétiques: 206-232. De Gruyter Mouton, Berlin/Boston (1972).
[40] K.N. Stevens; A.S. House. Perturbation of vowel articulations by consonantal context: An acoustical study. Journal of Speech and Hearing Research 6(2): 111-128 (1963).
[41] A. Paige; V. Zue. Calculation of vocal tract length. IEEE Transactions on Audio and Electroacoustics 18(3): 268-270 (1970). DOI: 10.1109/TAU.1970.1162113.
[42] F. Clermont. Advances in acoustic modeling of vocal tract dynamics. Actes du Colloque de Phonétique Expérimentale. Presses Universitaires du Midi, Toulouse (2007).
[43] F. Clermont; P. Mokhtari. Analysis and interpretation of vocal tract length estimates in speech. Speech Communication 25(1-3): 185-199 (1998).
[44] F. Clermont. Acoustic modeling of the vocal tract: theoretical foundations and applications. Proceedings of the International Conference on Speech Science (2002).
[45] P. Mokhtari; F. Clermont. New perspectives on linear-prediction modelling of the vocal tract: Uniqueness, formant-dependence and shape parameterisation. Proceedings of the Eighth Australian International Conference on Speech Science and Technology: 478-483 (2000).
[46] P.A. Barbosa. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. Revista de Estudos da Linguagem 20(1): 11-27 (2012).
[47] P.A. Barbosa; S. Madureira. Manual de Fonética Acústica Experimental: Aplicações a Dados do Português. Cortez, São Paulo (2023).
[48] J. Hao. Cross-racial Studies of Human Vocal Tract Dimensions and Formant Structures. Ph.D. Thesis, Ohio University (2002).
[49] S.A. Xue; G.J.P. Hao; R. Mayo. Volumetric measurements of vocal tracts for male speakers from different races. Clinical Linguistics & Phonetics 20(9): 691-702 (2006).
[50] W.T. Fitch. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. The Journal of the Acoustical Society of America 102(2): 1213-1222 (1997).
[51] F. Nolan. Speaker recognition and forensic phonetics. In: W.J. Hardcastle; J. Laver (Eds.). The Handbook of Phonetic Sciences. Blackwell, Oxford (1997).