

EFEITOS DA PROSÓDIA E DA DISFLUÊNCIA SOBRE A QUALIDADE DE VOZ EM DIFERENTES ESTILOS DE ELOCUÇÃO NO PORTUGUÊS BRASILEIRO

Pablo ARANTES¹

Aveliny Mantovan LIMA²

DOI: <http://dx.doi.org/10.21165/gel.v20i2.3516>

Resumo: Neste trabalho apresentamos dois experimentos que articulam as temáticas da qualidade de voz, disfluências, estilos de elocução e prosódia. O primeiro experimento investiga aspectos segmentais e prosódicos que afetam a ocorrência de laringalização, de disfluências e de fronteiras de grupo acentual. O segundo estuda o uso de voz não modal em diferentes contextos prosódicos, bem como em três estilos de elocução: fala espontânea, leitura de frases e de palavras isoladas. O achado inédito no português brasileiro e também o mais robusto encontrado nos dois experimentos é a predominância de casos de qualidade de voz não modal em posições prosodicamente fortes. Embora ambos tenham adotado critérios diferentes para a definição dessas posições, a semelhança nos resultados indica que o acento no nível da frase e da palavra é um fator relevante que condiciona a distribuição das instâncias de vozeamento não modal. Os resultados do primeiro experimento indicam diferença no comportamento do vozeamento laringalizado em relação aos segmentos e ao sexo do falante, uma vez que houve predominância de ocorrência nas vogais e maior taxa de vozeamento não modal em falantes do sexo feminino. No segundo experimento, observamos que os estilos de elocução exercem efeitos significativos sobre a distribuição do vozeamento não modal. Ainda, vimos que a ocorrência de vozeamento não modal em eventos de disfluência se comporta de maneira diferente em relação ao estilo de elocução. De modo geral, concluímos que os diferentes aspectos linguísticos, a saber, fluência de fala, fonoestilo e prosódia podem afetar a qualidade de voz.

Palavras-chave: Qualidade de voz. Fonoestilo. Fonética acústica. Disfluência. Prosódia.

1 Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil; pabloarantes@ufscar.br; <http://orcid.org/0000-0001-9707-8493>

2 Universidade de Brasília (UnB), Brasília, Distrito Federal, Brasil; avelinylima@gmail.com; <https://orcid.org/0000-0002-4305-4864>

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

EFFECTS OF PROSODY AND DISFLUENCY ON VOICE QUALITY IN DIFFERENT SPEAKING STYLES

Abstract: The paper presents two studies that investigate possible interactions between voice quality, disfluency, speaking styles and prosody. The first study investigates the effects of segmental and prosodic factors on the occurrence of laryngealization, disfluency and boundaries of accentual group in different prosodic conditions. The second study investigates non-modal phonation occurrences in different prosodic contexts in three speaking styles: semi-spontaneous interview, sentence reading and word reading. The most robust finding in both studies is that nonmodal phonation occurrences are highly associated with strong prosodic positions, such as stress group boundaries and lexically stressed syllables. Results of the first experiment suggest that creaky phonation occur more in vowels than consonants and that female speakers are more prone to present instances of creaky voice. The results of the second experiment show that in read speech occurrences of nonmodal phonation are more common in final (prepausal) position and in the semi-spontaneous style there are more occurrences in initial and medial positions. Overall, results from both studies show that changes in voice quality is sensitive to prosodic structure, fluency, and speaking styles.

Keyword: Voice quality. Phonostylistics. Acoustic phonetics. Disfluency. Prosody.

Introdução

Apresentamos neste artigo resultados de dois estudos experimentais que articulam áreas que foram até o momento pouco exploradas em conjunto e de forma sistemática na literatura a respeito do português brasileiro: qualidade de voz, prosódia, disfluências e estilos de fala. No restante desta introdução, apresentaremos brevemente conceitos e achados de estudos relevantes que relacionam uma ou mais dessas dimensões linguísticas como forma de destacar a relevância da tentativa de apresentar todas elas em um mesmo texto e destacar a originalidade dos resultados relatados aqui.

A descrição articulatória das atividades laríngicas é fundamental no estudo da produção sonora das línguas. Durante as últimas décadas, registra-se um crescimento no interesse científico pela descrição articulatória e acústica de enunciados que sejam realizados em condições naturais de fala, bem como de diferentes fonoestilos, e, por isso, destaca-se a importância de estudos que busquem entender a possível relação entre os tipos de fonação e de qualidade de voz, tendo em vista que, nessas condições de elocução, fatores paralinguísticos, como emoção e atitude, e extralinguísticos, como

o cansaço, podem afetar a fonação e a qualidade de voz de maneira importante. Nesse sentido, o falante, ao manipular o sistema laríngeo para executar as funções prosódicas de proeminência, demarcação ou dialógica, por exemplo, pode ter a sua qualidade vocal afetada.

O termo “qualidade de voz”, assim como definido por Laver (1980), é bastante geral e não está necessariamente relacionado apenas a características articulatórias ligadas à laringe, mas também a outros ajustes, incluindo os referentes a estruturas supralaríngeas e traços suprasegmentais. Por outro lado, na literatura fonética, expressões mais específicas como “tipo de fonação” ou “estado da glote” são encontradas e podem em alguns casos ser tomadas como sinônimos para qualidade de voz em um entendimento mais estreito do termo, que circunscreve o fenômeno às suas manifestações laríngeas (Ladefoged, 2001). Para os propósitos do presente trabalho, recorreremos a termos mais específicos, como “tipo de fonação” ou “estado da glote” para referenciar a descrição articulatória dos dados sem, contudo, nos comprometermos com uma dessas visões em particular a respeito da definição do termo “qualidade de voz”. Na seção de métodos, explicaremos mais detalhadamente os procedimentos que usamos nos dois estudos relatados no presente trabalho para identificar diferentes tipos de qualidade de voz.

A configuração laríngea considerada neutra é característica do tipo de fonação modal. Para Zeroual, Esling e Crevier-Buchman (2008), a consequência acústica da vibração laríngea regular é a produção de formas de onda periódicas. Essa periodicidade, segundo Laver (1980, p. 16), implica a ausência de *jitter* e *shimmer*.

Em relação aos tipos de fonação não-modais, a descrição articulatória que abrange uma primeira proposta de contínuo de estados da glote, feita por Ladefoged (1972), varia de “não vozeado” (aritenóides mais afastadas), passando pela “soproso”, “modal” (vibração regular), “*creaky voice*”, “*creak*”, e, por último, a oclusiva glotal (aritenóides mais fechadas). Ladefoged (1972) esclarece que entre os estados “soproso” e “modal” há um gradiente de “voz relaxada” (murmúrio ou sussurro), e entre os estados “modal” e “*creaky voice*”, há a “voz tensa”, embora não se possam definir pontos pré-determinados entre esses diferentes estados. A fim de ampliar a proposta de Ladefoged (1972), Laver (1980) propõe categorias para os tipos de fonação, que seriam: “modal” e “falsete”, “*creak*” e “sussurro”, “áspero” e “soproso”. Tais categorias podem ocorrer sozinhas ou se combinar com outras, embora algumas não possam ser combinadas entre si.

Essa classificação dos modos de fonação tem sido estudada do ponto de vista da fonética geral, mas pode ser utilizada também para análise de voz como uma ferramenta que auxilia na descrição de alterações patológicas da voz, como as disfonias. No caso do

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

presente estudo, pretende-se compreender como os modos de fonação se comportam em diferentes estilos de elocução, como a fala espontânea (entrevistas a um *podcast* e narrativa pessoal) e a leitura de frases e de palavras isoladas. Nesse âmbito, dois estudos no português brasileiro (doravante PB) estudaram especialmente o tipo de fonação que é chamado de laringalização (Lima, 2011; Oliveira, 2017). No estudo de Oliveira (2017), o autor buscou compreender como a ocorrência de laringalização pode contribuir para a tarefa de comparação de locutores. Os resultados mais importantes de seu estudo foram que: (a) a taxa de laringalização vocálica e total foram as medidas que permitiram maior diferenciação de pares; (b) a taxa de laringalização foi maior no grupo do sexo feminino; (c) a ocorrência de laringalização alinhada à fronteira de unidades prosódicas pareceu revelar uma possível relação entre estrutura prosódica e a sua motivação ou restrição. O estudo de Lima (2011) investigou justamente a possível relação entre as fronteiras prosódicas e discursivas e o fenômeno de laringalização. Como resultado, observou-se relação estatisticamente significativa entre as variáveis fronteiras prosódicas, fronteiras discursivas e laringalização, e, também, entre a produção de disfluências. Falaremos sobre a relação entre qualidade de voz e disfluência logo mais adiante. Vimos nesse último estudo de Lima (2011) uma investigação que relaciona o interesse por qualidade de voz e prosódia. Tal interesse está ligado à própria definição de prosódia que, segundo Barbosa (2019), “é o componente da nossa fala que organiza nossos enunciados, moldando nossa maneira de falar através do concurso de modificações articulatórias que se manifestam acusticamente em unidades prosódicas”.

Trabalhos que estudaram outras línguas também se dedicaram à descrição dos efeitos da interface entre prosódia e modos de fonação, especialmente o modo laringalização. Em trabalho sobre o inglês, Dilley e colegas (1996) mostram que a laringalização afeta vogais que ocupam, principalmente, a posição inicial de uma frase entoacional, independente da tonicidade da vogal. Ainda em relação ao inglês, Redi e Shattuck-Hufnagel (2001) observaram que a laringalização pode servir como um marcador de fronteira prosódica, sinalizando o final do turno quando ocorre em fim de enunciado ou contrastar o trecho proeminente, seja antecedendo ou sucedendo um trecho proeminente. Andreeva e colegas (2013), em trabalho sobre o norueguês e o alemão, comparou essas duas línguas e observou que é possível encontrar estados da glote mais fechados, que resultam na produção de oclusiva glotal e laringalização, em trechos proeminentes dos enunciados. Em trabalho que abrangeu sueco, inglês e finlandês, Gordon e Ladefoged (2001) mostram que o modo de fonação “laringalização” ocorre com mais frequência em sílabas acentuadas, além de servir como marcador de fronteira prosódica, sendo mais comum nas palavras iniciadas por vogal, em algumas categorias gramaticais e fronteira de frase entoacional. Kim e Valdovinos (2014), em estudo sobre a língua mariteco cora, falada no

México, mostram que vogais produzidas com modo de fonação “laringalização” podem ocorrer tanto em trechos prosodicamente fortes quanto fracos, embora a queda na frequência fundamental, que pode ser considerada um dos correlatos desse modo de fonação, seja mais acentuada nos trechos prosodicamente fortes. A respeito do vietnamita, Jannedy (2007) relata a ocorrência de laringalização em vogais preferencialmente em posição de ênfase. Em conjunto, os resultados reportados na literatura a respeito de um número diverso de línguas, o português brasileiro entre elas, sugerem que há uma relação privilegiada entre posições prosódicas fortes, sejam elas sílabas tônicas, trechos postos em ênfase ou fronteiras prosódicas, e a ocorrência de modos de fonação não modais. Essa relação poderia ser interpretada como evidência para o uso de fonação não modal como correlato auxiliar para a realização das funções prosódicas de demarcação e ênfase.

Em relação à disfluência, é preciso esclarecer que este termo se refere a uma linha teórica que está orientada para o entendimento da gagueira do desenvolvimento, conforme explica Merlo (2006). Nessa linha, as disfluências são vistas como índices de dificuldades motoras durante a fala e os estudos conduzidos sob essa perspectiva dividem seus participantes em pessoas que gaguejam e pessoas que não gaguejam. Embora o estudo 1 apresentado no presente artigo tenha utilizado material de fala produzido por pessoas que não gaguejam, por ter sido desenvolvido no âmbito do curso de Fonoaudiologia da Universidade de Brasília, a classificação das disfluências se deu seguindo esta linha. A tipologia adotada será apresentada na seção de Materiais e Métodos do Estudo 1.

Ainda sobre esse tema, Merlo (2006, p. 209) explica que outra linha teórica

[...] denomina seu objeto de “hesitação” e apresenta uma forte orientação psicolinguística, focando seus estudos na produção falada de pessoas sem distúrbios de comunicação. Nesta linha, as hesitações são vistas como índices de dificuldades com o processamento da língua falada.

A hesitação se relaciona com a prosódia por meio daquilo que Koch e Souza e Silva (1996) e Marcuschi (1999) chamam de “marcas formais”. Tais marcas são tipificadas de maneira diferente a depender da linha teórica adotada pelo analista. Exemplos dessas marcas, segundo os autores citados, seriam pausas (preenchidas, silenciosas ou hesitativas), repetições hesitativas, prolongamentos hesitativos e falsos inícios.

Em busca das relações a que se refere o presente trabalho, apresentamos, por fim, alguns resultados de trabalhos na interface entre disfluência e qualidade de voz encontrados na literatura internacional. Tilsen (2006), em trabalho sobre fala espontânea em inglês, observou que a laringalização pode aparecer antecedendo a disfluência de

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

repetição, principalmente quando se repete a palavra três vezes. Kohler (2000) observou que, em alemão, a laringalização reforça as pausas e os finais prolongados, tanto de palavras quanto de frases, sendo que na última a marcação fica mais evidente. As possíveis inter-relações entre modos de fonação não modal, estrutura prosódica e disfluências foram muito pouco estudadas até o momento, o que torna os resultados que apresentamos no presente artigo bastante relevantes e inovadores.

Em relação aos estilos de elocução, a tradição de pesquisa que relaciona os diferentes estilos a variáveis de natureza fonético-fonológica centra seus esforços em caracterizar as possíveis especificidades de estilos que se distribuem num *continuum* em cujos polos normalmente são colocadas, de um lado, a fala espontânea e, do outro, a chamada fala de laboratório. Sobre essas definições e as controvérsias em torno desses termos, ver, entre outros, Beckman (1997), Lucente (2012, p. 21-23) e Xu (2010). Uma variada gama de estilos pode ser identificada entre os pólos, definidos por relações específicas entre conteúdo e a forma, como, por exemplo, o estilo jornalístico, a fala atuada, a narração esportiva, entre outros. No caso de variáveis fonético-fonológicas, Llisterri (1992) apresenta uma revisão abrangente da literatura disponível até o momento de sua publicação, na qual sumariza resultados a respeito da influência de estilos de elocução sobre fatores tais como qualidade vocálica, taxa de elocução, frequência fundamental, entre outras dimensões fonéticas. Entre os 35 estudos incluídos na revisão, apenas um apresenta evidência de correlações sistemáticas entre mudanças no modo de fonação por efeito do estilo de elocução (Greisbach, 1991): no caso, relata-se a ocorrência de voz laringalizada em encontros vocálicos e ditongos na fala muito rápida em contraste com a fala em taxa neutra).

Na literatura mais recente, alguns trabalhos mostram a influência do estilo de elocução na implementação de correlatos acústicos do acento lexical em línguas como o PB, inglês britânico, sueco, estoniano e checo. Essa série de estudos conta com *corpora* de fala coletados de maneira semelhante no âmbito de um mesmo projeto de pesquisa. O procedimento de coleta permitiu a elicitação de três estilos de elocução, leitura de palavras isoladas, leitura de frases e entrevista semiespontânea³.

Nos trabalhos, são investigados parâmetros tradicionalmente considerados relevantes na pesquisa sobre correlatos acústicos do acento lexical, como a duração, f_0 e uma medida de intensidade relativa, chamada ênfase espectral. A medida da ênfase

³ Veja <https://wordstress.ling.su.se/> para mais informações sobre os objetivos do projeto, os procedimentos de coleta de dados e informações demográficas dos participantes. O uso do material para o presente estudo foi autorizado pelo coordenador do projeto.

espectral é definida como a contribuição relativa da banda mais alta de energia para o nível de intensidade total de um segmento⁴. A justificativa para a adoção dessa medida é a observação segundo a qual contribuições relativas maiores da banda alta para a energia global do sinal resultam de maior esforço expiratório e articulatorio, que podem implicar maior tensão da musculatura laríngea. A elevação do esforço articulatorio na atividade de fonação pode ter impacto na percepção da qualidade de voz. No estudo sobre o PB, os resultados referentes ao parâmetro duração indicam que o estilo leitura de palavras é caracterizado por gerar maior diferença no alongamento da sílaba tônica em relação às átonas em comparação aos outros dois estilos. Em relação à f_0 , registra-se mais variação, medida pelo desvio-padrão de f_0 nas vogais, no estilo leitura de palavras do que nos outros dois estilos. Na medida de ênfase espectral, o efeito do estilo se manifesta na forma de valores maiores para esse parâmetro na entrevista semiespontânea em comparação com os outros dois estilos, observação que é explicada como resultado de maior esforço associado a trechos mais longos nas narrativas nos quais são produzidos enunciados em níveis elevados de f_0 por trechos relativamente longos. Apesar do avanço representado por essas contribuições recentes, ainda é marcante a escassez de estudos que relacionam a qualidade de voz com os estilos de elocução. Muito embora a medida de ênfase espectral reflita a atividade laríngea, ela mede modulações relativamente pequenas na tensão da musculatura laríngea e faz isso de forma indireta e inespecífica em termos das estruturas envolvidas na fonação. Portanto, ainda se faz necessário explorar a questão das interações entre qualidade de voz e estilos de fonação de maneira mais sistemática.

Em função dos diagnósticos feitos acima a respeito da carência de estudos que procurem compreender simultaneamente uma ou mais das dimensões mencionadas, especialmente tomando como base o português brasileiro, reunimos neste artigo o relato de dois experimentos que abordam as inter-relações entre variáveis prosódicas, estilos de elocução, qualidade de voz (em especial a laringalização, mas não apenas) e disfluências. Os dois estudos serão apresentados a seguir de maneira separada e, ao final, os resultados serão discutidos de maneira conjunta.

O Estudo 1 tem como foco de estudo a qualidade de voz laringalizada (também conhecida como *creak* ou *creaky voice*), que, na perspectiva gradiente dos estados da glote, encontra-se entre os estados neutro (modal) e mais fechado (oclusiva glotal). O objetivo do estudo é investigar os contextos segmentais e prosódicos que podem favorecer a ocorrência de laringalização, a estrutura rítmica (a posição no grupo acentual) e a presença de disfluência comum de fala.

⁴ “Trechos de fonação ininterrupta ladeados por silêncio” (tradução dos autores).

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

O Estudo 2 tem como objetivo determinar se o uso de qualidade de voz não modal (laringalização e outros tipos) é afetado por um conjunto de variáveis de natureza diferente: a) prosódicas: a tonicidade da sílaba onde a voz não modal acontece e a posição da palavra em relação à posição inicial e final de enunciado e b) estilo de elocução, em especial a comparação entre fala espontânea e fala lida.

Estudo 1

Materiais e métodos

Apresentamos nesta seção uma caracterização das amostras de fala que compõem o *corpus* estudado. Descrevemos em seguida os parâmetros fonéticos extraídos das amostras de fala e os critérios usados para a extração, além da organização da segmentação e etiquetagem dos dados. Finalmente, descrevemos os critérios estatísticos para a análise dos resultados obtidos.

Material de fala

A amostra foi composta por 9 falantes no total, sendo 4 do sexo feminino e 5 do sexo masculino, em idade adulta, com diversas formações profissionais, todos com histórico de atuação profissional na área da moda. Todos os falantes da amostra são nativos do PB. Nas amostras selecionadas para o estudo, nenhum apresentou qualquer tipo de disfonia e/ou alterações vocais.

As amostras de fala analisadas no estudo são trechos de episódios de *podcasts* veiculados por diferentes emissoras de rádio, disponibilizados publicamente nos *websites* das emissoras em formato de áudio MP3 ou em canais do YouTube, que também foram extraídos em formato MP3, e convertidos posteriormente para o formato sem compressão WAV para as etapas de análise acústica. Após essa coleta inicial, os arquivos de áudio foram analisados em relação à qualidade acústica, seguindo os seguintes critérios: (a) excelente relação sinal/ruído; (b) ausência da fala de outros que não seja o sujeito alvo; (c) ausência de música de fundo durante a fala do sujeito. Tais critérios foram seguidos para garantir a possibilidade de extração das informações acústicas desejadas. Foram excluídos da amostra inicialmente selecionada trechos em que um dos critérios não pode ser atendido. A duração total das amostras é de 33 minutos e 15 segundos de áudio, entre amostras de falantes do sexo feminino e masculino.

Análises fonéticas

Laringalização

Episódios de laringalização foram identificados através do exame visual sistemático da forma de onda, do espectrograma de banda larga do sinal acústico e dos traçados de f_0 e intensidade gerados pelo programa Praat que podem ser superpostos ao espectrograma de cada amostra. Para a identificação, foram usados os correlatos acústicos associados a essa qualidade de voz ou modo de fonação reunidos por Lima (2011, p. 16), a saber: 1) queda ou rebaixamento local no traçado de f_0 , (2) irregularidade dos pulsos glotais, que pode manifestar-se tanto no espaçamento temporal entre os estriamentos verticais no espectrograma e entre espículas no oscilograma tanto quanto na intensidade dos pulsos e 3) queda local no contorno do traçado de intensidade ou na amplitude da forma de onda. A inclinação espectral, mencionada por Lima (2011), não foi usada, uma vez que o efeito da laringalização sobre esse correlato não é facilmente detectável pela simples inspeção visual no espectrograma. Não foi feita nenhuma tentativa de automatizar o reconhecimento das pistas mencionadas anteriormente. Fez-se a inspeção visual completa de todas as amostras de áudio do *corpus* em busca da ocorrência dos padrões visuais no espectrograma e oscilograma para a identificação dos trechos com laringalização. A oitava dos trechos identificados visualmente como ocorrências de laringalização foi usada como critério confirmatório adicional à análise das pistas visuais.

Disfluências

A identificação das disfluências seguiu a tipologia de disfluências típicas ou não gagas apresentada por Schiefer e Arcuri (2000), amplamente usada na avaliação da fluência da fala no contexto da prática fonoaudiológica. A tipologia identifica as seguintes classes de disfluências típicas: hesitação, prolongamento, repetição, revisão, interjeição, bloqueio e palavra não terminada. Nessa tipologia, o prolongamento é considerado uma disfluência atípica, ou seja, comum na produção de falantes com gagueira. Embora enquadrem os prolongamentos entre as disfluências atípicas, os autores reconhecem que elas não são exclusivas dos quadros de gagueira e podem ocorrer, embora com menor frequência, também na fala de indivíduos sem alteração de fluência. Como notamos durante a audição das amostras a ocorrência de um número não negligenciável de casos de prolongamento, decidimos incluir essa disfluência entre as possibilidades de disfluência em nosso estudo.

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

Fronteira prosódica

O evento prosódico que escolhemos analisar no Estudo 1 são as fronteiras de grupo acentual. Barbosa (2019, p. 43-45) define grupo acentual como um agrupamento delimitado por dois acentos frasais consecutivos. Ainda segundo a definição, acento frasal é o acento que recai sobre a sílaba lexicalmente acentuada de uma palavra fonológica que é a mais proeminente no enunciado de que faz parte. Por essa definição, um grupo acentual contém pelo menos uma palavra fonológica; pode conter mais de uma, desde que apenas um dos acentos lexicais seja proeminente no nível do enunciado. Barbosa (2006) define um conjunto de procedimentos baseados em critérios de produção para instrumentalizar essa definição e permitir a identificação de grupos acentuais em amostras de fala. Esse procedimento identifica acentos frasais como picos no contorno de duração de um enunciado segmentado em unidades do tamanho da sílaba ou unidades VV e um grupo acentual como o conjunto de unidades VV delimitado entre dois picos consecutivos no contorno. Para o propósito de identificação de acentos frasais, o contorno de duração bruta (a duração em unidades de tempo real definida pela segmentação do sinal acústico) passa primeiro por uma sequência de duas transformações. Primeiramente, a normalização pela técnica do z-score estendido, que minimiza os efeitos da duração intrínseca dos segmentos e enfatiza os alongamentos e encurtamentos motivados pela atribuição de proeminências prosódicas. Em seguida, o contorno normalizado passa por uma etapa de suavização por meio da aplicação de uma média móvel de cinco pontos com o objetivo de minimizar oscilações no contorno de duração que não tenham sido eliminadas na etapa de normalização. Uma vez que o contorno bruto tenha sido normalizado e suavizado, os picos (ou pontos máximos locais) desse contorno podem ser identificados. Esses pontos são considerados acentos frasais e cada um desses pontos pode ser considerado uma fronteira de grupo acentual segundo a definição de Barbosa. A descrição técnica detalhada das operações de normalização e suavização é apresentada em Barbosa (2006), juntamente com o código-fonte de um *script* do programa de análise acústica Praat que implementa computacionalmente os dois procedimentos. Na etapa de análise, serão computados os eventos de laringalização que coincidem ou não com as fronteiras de grupo acentual.

Frequência fundamental

Conforme dissemos anteriormente, a literatura sugere que há um rebaixamento repentino da frequência fundamental durante o trecho da fonação laringalizada e que esse rebaixamento localizado pode ser considerado um correlato da laringalização. Para testar

em que medida essa associação se verifica em dados do português brasileiro, decidimos quantificar a amplitude de variação da f_0 nos trechos identificados como ocorrências de laringalização por meio da seguinte relação:

$$\log_2 \left(\frac{f_{lar.}}{f_{mod.}} \right),$$

em que $f_{lar.}$ é a mediana de f_0 no trecho laringalizado e $f_{mod.}$ é a mediana de f_0 da unidade VV em que o evento de laringalização acontece, excluídos os valores de f_0 do trecho laringalizado. O valor resultante será negativo quando houver uma queda no contorno de f_0 no trecho laringalizado e será positivo se houver uma subida no contorno de f_0 no trecho laringalizado. Além da amplitude de variação, calculou-se também o desvio-padrão dos valores de f_0 em semitons, ao longo da unidade VV em que houve ocorrência de laringalização. Valores altos de desvio-padrão indicam uma variabilidade alta no contorno, constituindo evidência auxiliar para a ocorrência de movimentos abruptos no contorno.

A extração do contorno de f_0 de cada amostra de áudio foi feita com o auxílio do *script* `better_f0` (Arantes, 2019) para o programa Praat. Esse *script* implementa um algoritmo que otimiza alguns parâmetros da função nativa *Get pitch* do programa Praat de forma a minimizar erros de extração. O contorno resultante foi analisado em seguida pelo *script* `f0_outliers` (Arantes, 2021a), que identifica como pontos suspeitos de erro de extração trechos em que duas amostras de f_0 consecutivas apresentam variação positiva ou negativa acima de um determinado limiar. Os casos suspeitos apontados pelo *script* foram posteriormente analisados por um dos autores para verificar se se tratava de um erro de extração verdadeiro ou um falso alarme.

Segmentação e etiquetagem dos dados

As amostras de fala foram segmentadas e anotadas por meio do *software* livre de análise acústica Praat (Boersma, 2001). A anotação das análises relevantes foi distribuída em cinco camadas em um objeto TextGrid, como indicado na figura 1. Os trechos laringalizados são marcados na primeira camada, identificados pela etiqueta “l”; os trechos com voz modal imediatamente circundantes estão identificados pela etiqueta “m”. Na segunda camada está a segmentação em unidades VV. Os segmentos consonânticos e vocálicos presentes em cada unidade VV são identificados por etiquetas do sistema de transcrição *Ortofon* (Albano; Moreira, 1996), que associa cada um dos fones do PB a um símbolo do alfabeto ASCII. A informação presente nessa camada é usada para a

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

realização da operação de normalização da duração. A terceira camada identifica, dentro de cada unidade VV, os segmentos consonantais, indicados pela etiqueta “c”, e os vocálicos, indicados pela etiqueta “v”. Na quarta camada estão registradas as ocorrências de disfluências, identificadas por sete rótulos textuais diferentes que correspondem aos casos previstos na tipologia mencionada na seção “Disfluências”. A segmentação e etiquetagem dessas quatro camadas foi feita manualmente por uma das autoras.

A quinta camada usa os mesmos marcos temporais da segunda camada, isto é, a que contém a marcação das unidades VV. Para cada unidade VV, a quinta camada indica se nessa unidade recai uma fronteira de grupo acentual ou não. Caso recaia, a unidade recebe a etiqueta “1”; caso não, recebe a etiqueta “0”. Essa camada foi obtida de forma automática por meio do script *duration_suite* (Arantes, 2021b), que implementa computacionalmente o procedimento para identificação de grupos acentuais a partir do contorno de duração que é explicado em Barbosa (2006, p. 169-170). A partir das informações dessa camada, um *script ad hoc* extraiu a informação acerca da presença ou não de fronteira de grupo acentual na vizinhança das ocorrências da laringalização, considerando uma janela de três unidades VV, centrada naquela em que a laringalização ocorreu. O resultado dessa verificação gerou padrões com três posições, que podem ser ocupadas pelos símbolos “1”, “0” e “_”; o símbolo “_” representa o início ou final absolutos do contorno de duração extraído em um arquivo de áudio analisado. Os quatro padrões principais são listados e explicados a seguir:

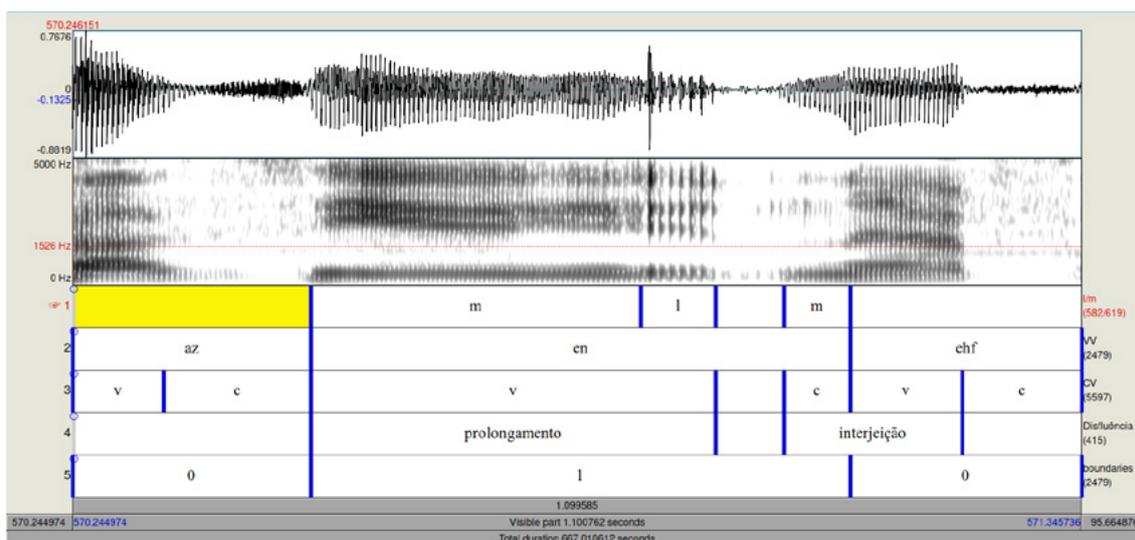
1. 000: não há fronteira de grupo acentual na janela;
2. 100: fronteira de grupo acentual antes da laringalização;
3. 010: fronteira coincide com a ocorrência de laringalização;
4. 001: fronteira de grupo acentual na unidade seguinte à laringalização;

Além dos quatro padrões listados, há, ainda, aqueles em que há a presença do símbolo “_”, que indica a presença do início ou fim absolutos do contorno de duração analisado. Exemplos de padrão com a presença desse símbolo são, entre outros, “10_”, fronteira de grupo acentual acontece antes da laringalização, seguida do final absoluto do contorno de duração e “_10”, fronteira de grupo acentual coincide com a laringalização na primeira unidade VV do contorno de duração. Como a ocorrência desses padrões é relativamente limitada, uma vez que podem ocorrer apenas no início ou fim dos contornos, optamos por agrupar as ocorrências dos padrões que contêm “_” aos demais, respeitando a similaridade: ocorrências do padrão “10_” foram contadas como ocorrências do padrão “100”; ocorrências de “01_” foram contadas como ocorrências de “010” e assim por

diante. Finalmente, em função da natureza relacional dos contrastes prosódicos, não são considerados padrões nos quais há a presença de duas fronteiras em sequência, como “110” ou “011”.

Figura 1. Ilustração da organização das anotações em camadas de objetos TextGrid do Praat.

Na camada 1, estão marcados os trechos modais e laringalizados; na camada 2, as unidades VV; na camada 3, são individualizados os segmentos vocálicos e consonantais; na camada 4, as disfluências; na camada 5, anota-se, para cada unidade VV, se ela coincide ou não com fronteira de grupo acentual



Fonte: Elaboração própria

Análise estatística

Nos resultados de estatística descritiva, os valores de média e mediana das variáveis dependentes são acompanhados, respectivamente, pelo desvio-padrão ou desvio mediano absoluto, abreviados respectivamente como DP e DMA. Nos casos em que o teste-*t* ou o teste de Análise de Variância (ANOVA) foram usados, testou-se previamente o pressuposto de homogeneidade de variância por meio do teste de Fligner-Killeen. Quando esse pressuposto não foi atendido, testes não paramétricos foram usados como alternativa: Wilcoxon, no caso do teste-*t*, e Kruskal-Wallis, no caso da ANOVA. A magnitude de efeito para o teste-*t* foi medida por meio do *d* de Cohen e para ANOVA e Kruskal-Wallis reporta-se o *eta* quadrado parcial (η^2).

Toda a estatística descritiva, bem como os testes inferenciais, foi feita com o auxílio do ambiente de computação estatística R (R CORE TEAM, 2020). Em todos os testes estatísticos inferenciais adotou-se o limiar de 5% para declarar a hipótese nula refutada.

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

Resultados

Caracterização geral das ocorrências de laringalização

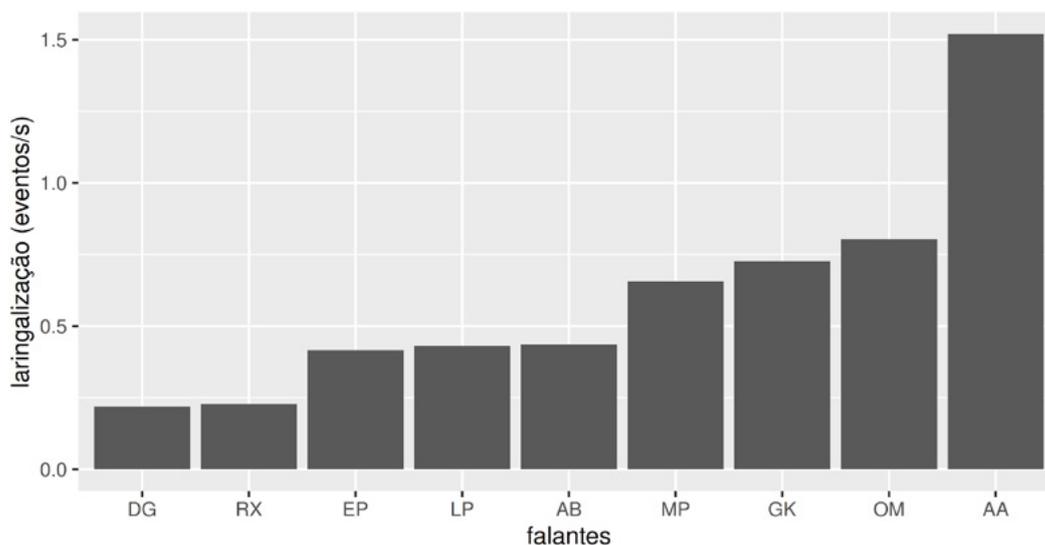
O conjunto de dados analisados consiste em 18.560 segmentos, 8.381 dos quais são vogais. Foram registradas 926 ocorrências de laringalizações nos dados. Destas, 94% das ocorrências se dão em fones vocálicos e 4% em fones consonantais. Esse resultado corrobora os achados de Oliveira (2017), que relata um percentual máximo de 4% de ocorrências de laringalização nos segmentos consonantais entre os participantes de seu estudo. A explicação sugerida pelo autor para esse resultado é que os segmentos vocálicos estão mais sujeitos à ocorrência de laringalização por terem maior índice de sonoridade. Uma vez que a laringalização é um fenômeno que afeta a vibração das pregas vocais, somente os segmentos em que há vozeamento estão sujeitos a ele. Na mesma direção, Sousa e Lima (2015) também observaram que as laringalizações ocorrem em grande parte nas vogais e nas consoantes soantes, mais especificamente, [l], [m] e [n]. Um estudo a respeito da língua Totonac-Tepehua (Levy Podolsky, 1987) reporta ocorrências de laringalização em soantes, porém sem especificar em quais delas houve essas ocorrências. Kohler (2000), em um estudo a respeito do alemão, encontrou ocorrências de laringalização principalmente nas soantes nasais. Os segmentos mais afetados em ambos os estudos são os vocálicos, embora não sejam apresentados resultados quantitativos em relação a essa prevalência.

Para saber o quanto o fenômeno da laringalização é frequente na produção de fala dos participantes do *corpus*, calculamos a taxa de ocorrência de eventos de laringalização, definida como o número de ocorrências de laringalização por segundo de fala – o cálculo não levou em conta a duração de cada evento, apenas sua ocorrência. A figura 2 mostra os valores da taxa calculada separadamente para cada falante. A taxa mediana dos nove falantes reunidos é 0,43 laringalizações/s (DMA 0,32) com uma gama que vai de 0,23 a 1,52. Nenhum dos falantes deixou de produzir laringalizações. Pode-se notar na figura 2 quatro grupos com valores semelhantes de taxa: DG-RX, AB-EP-LP, OM-MP-GK e AA. Oliveira (2017) opta por um cálculo de taxa de laringalização definido como a porcentagem de tempo de fala tomado por fonação laringalizada. Fazendo o cálculo da taxa segundo essa definição, a mediana dos falantes no *corpus* é 3,52% (DMA 2,08), com uma gama que vai de 1,51% a 12,27%. Em Oliveira (2017), a taxa mediana é de 3% (DMA 2,97) com uma gama que vai de 3% a 14%.

É possível que essa taxa seja uma característica idiossincrática, embora os dados não permitam confirmar essa possibilidade. É possível, ainda, que essa taxa seja uma influência

do estilo de elocução, hipótese que os dados não permitem verificar, uma vez que todos os dados do Estudo 1 podem ser classificados como pertencentes a um mesmo estilo: fala espontânea em contexto de entrevista em veículo de imprensa.

Figura 2. Taxa de laringalização, medida em eventos por segundo, em função dos falantes



Fonte: Elaboração própria

Em termos da duração dos eventos de laringalização, a média e a mediana gerais são respectivamente 68 (48) e 57 (31) milissegundos. Separando por tipo de segmento afetado, as laringalizações que afetam as vogais têm média e mediana de 69 (49) e 57 (31) ms e as que afetam as consoantes têm média de 60 (29) e 57 (24) ms. A amplitude das médias calculadas separadamente por falante abrange valores que vão de 54 a 82 ms. O resultado de uma ANOVA univariada que usa os falantes como fator aponta que o efeito do fator é significativo [$F(8, 917) = 2,86$ $p = 0,004$; $\eta^2 = 0,024$], indicando que a duração média dos trechos laringalizados não é a mesma para todos os falantes.

Para entender o significado desse achado, podemos comparar a duração dos eventos laringalizados aos valores de duração de vogais e consoantes no *corpus*, cujas médias e medianas são 106 (95) ms e 79 (47) ms para as vogais e 78 (43) e 72 (34) ms para as consoantes. Os valores sugerem que a laringalização toma uma proporção alta da duração dos segmentos que ela afeta. A porcentagem de tempo ocupado pela laringalização nos segmentos afetados têm média e mediana de 54% (35) e 47% (44) no caso das vogais e 75% (26) e 83% (24) no caso das consoantes. As vogais têm uma proporção menor do seu tempo tomado pelas ocorrências de laringalização em comparação com as consoantes [$\chi^2(1) = 18,5$ $p < 0,001$; $\eta^2 = 0,017$], embora a variabilidade no caso das vogais seja

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

estatisticamente maior [$\chi^2(1) = 5,88 p = 0,015$]. Casos em que toda a duração do segmento é tomada pela laringalização correspondem a 29% das vogais e 43% das consoantes afetadas. Analisando a proporção dos segmentos tomada pela laringalização pelo ponto de vista da variação individual, o teste estatístico mostra que os falantes apresentam médias que são significativamente diferentes entre si [$\chi^2(8) = 146 p < 0,001; \eta^2 = 0,15$], assim como são também as variâncias [$\chi^2(8) = 26,8 p < 0,001$]. A magnitude do efeito é considerada grande, o que aponta para diferenças importantes entre os falantes.

Embora não seja o caso do PB, que não tem a qualidade de voz como traço de contraste fonêmico, um estudo que analisou línguas nas quais a qualidade de voz é um traço fonêmico e que comparou a proporção entre o trecho laringalizado e o trecho modal nos segmentos em que houve laringalização, verificou-se que o trecho modal tinha duração maior do que o trecho laringalizado (Blankenship, 2002). Esse resultado pode indicar que falantes de línguas em que o tipo de fonação não modal é contrastivo desenvolvem um controle mais preciso dessa característica da fala do que os falantes de línguas em que a fonação não modal está mais associada a outros níveis da organização linguística, como parece ser o caso do PB, conforme indicam as análises que estão reportadas a seguir neste estudo.

Laringalização e f_0

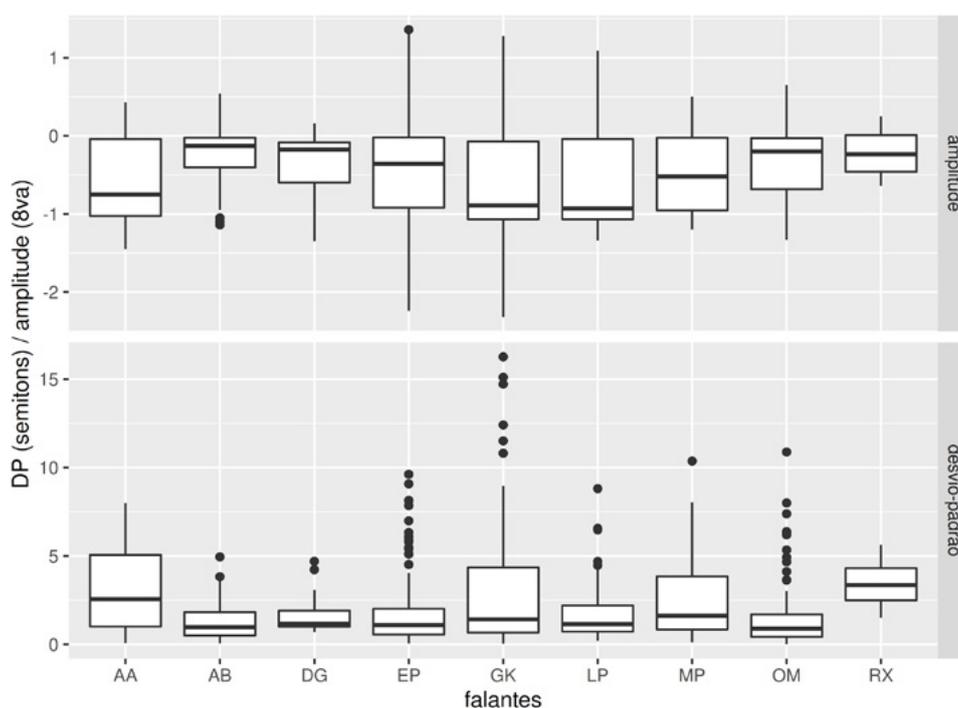
Foi possível determinar a amplitude de variação da f_0 no trecho laringalizado em comparação com o trecho não laringalizado envolvente em 71% das 926 ocorrências de laringalização identificadas no *corpus*. Casos em que isso não foi possível têm duas causas. Primeiramente, a ausência de trecho de f_0 não laringalizado envolvente no âmbito da unidade VV. Essa situação pode acontecer, por exemplo, nos casos em que a laringalização tomou 100% da duração da vogal e as consoantes que a envolvem são desvozeadas. Casos em que a laringalização tomou toda a duração do segmento afetado correspondem a 45,5% dos casos em que não foi possível determinar a amplitude. Uma segunda causa possível para a não determinação da amplitude são casos em que o algoritmo de extração da f_0 não conseguiu determinar o valor no trecho laringalizado em função da irregularidade dos pulsos glotais nesse modo de vozeamento.

A amplitude média nos casos em que foi possível determinar essa medida é de -0,47 oitava (DP 0,57), isto é uma queda no valor da f_0 de aproximadamente 50% no trecho laringalizado em relação ao trecho não laringalizado envolvente. Um teste-*t* bicaudal de amostra única comparando a média da amplitude com o valor de referência para essa variável igual a 0 aponta que a média amostral é significativamente diferente de 0, com

magnitude de efeito considerada alta [$t(657) = -20,8$ $p < 0,001$; $d = -0,81$]. A distribuição dos valores de amplitude em função dos falantes é mostrada na figura 3 por meio de gráficos de caixa (*box plots*). Como o exame visual da figura indica, o valor típico da amplitude de todos os falantes é negativo, embora haja variabilidade entre os falantes. A média calculada separadamente por falante varia entre -0,65 8va e -0,22 8va.

O teste de homogeneidade de variância indica que as os falantes têm valores de variância significativamente diferentes [$\chi^2(8) = 41,7$ $p < 0,001$]. O teste de comparação de médias indica que o efeito dos falantes sobre a média da amplitude é significativo, embora a magnitude do efeito seja considerada baixa [$\chi^2(8) = 38,2$ $p < 0,001$; $\eta^2 = 0,033$].

Figura 3. Distribuição da amplitude de f_0 (em oitavas) da porção laringalizada relativamente à porção não laringalizada e desvio-padrão do contorno das unidades VV afetadas por laringalização (em semitons)



Fonte: Elaboração própria

No caso do desvio-padrão do contorno de f_0 das unidades em que ocorre a laringalização, foi possível calcular a medida em 77,5% dos casos. A explicação para os casos em que isso não foi possível é similar à apontada anteriormente para o caso da amplitude. Dos casos em que não foi possível determinar o desvio-padrão, 83,1% deles correspondem a ocorrências em que a laringalização tomou toda a duração do segmento afetado. O desvio-padrão médio foi de 2,1 st (DP 2,3), resultado que sugere que há

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

uma variação no contorno das unidades VV afetadas pela laringalização, que pode ser atribuída à sua presença. Um teste-*t* bicaudal de amostra única comparando a média do desvio-padrão com o valor de referência para essa variável igual a 0 aponta que a média amostral é significativamente diferente da referência, com magnitude de efeito considerada alta [$t(717) = 24,8$ $p < 0,001$; $d = 0,93$]. O resultado mostra que a variabilidade do desvio-padrão não pode ser considerada nula e que a magnitude dessa diferença é considerável. Considerando o limiar necessário para a percepção de movimentos no contorno de f_0 (Martins; Ferreira Netto, 2010), uma queda com magnitude média de meia oitava (ou 6 semitons) tem o potencial para ser saliente do ponto de vista perceptual. A variabilidade associada a essa média é bastante elevada. A figura 3 mostra a distribuição dos valores em função dos falantes. A observação da figura confirma que a variabilidade alta do desvio-padrão é em parte atribuível a diferenças entre os falantes. Falantes como AB e DG apresentam valores baixos e pouco variados; falantes como EP e OM apresentam valores baixos, mas com grau maior de variabilidade em sua amostra. O teste de homogeneidade de variância confirma que as amostras dos diferentes falantes têm variâncias significativamente diferentes [$\chi^2(8) = 63$ $p < 0,001$]. O teste de comparação de médias indica que o efeito dos falantes sobre a média do desvio-padrão é significativo, embora a magnitude do efeito seja considerada baixa [$\chi^2(8) = 39,5$ $p < 0,001$; $\eta^2 = 0,034$].

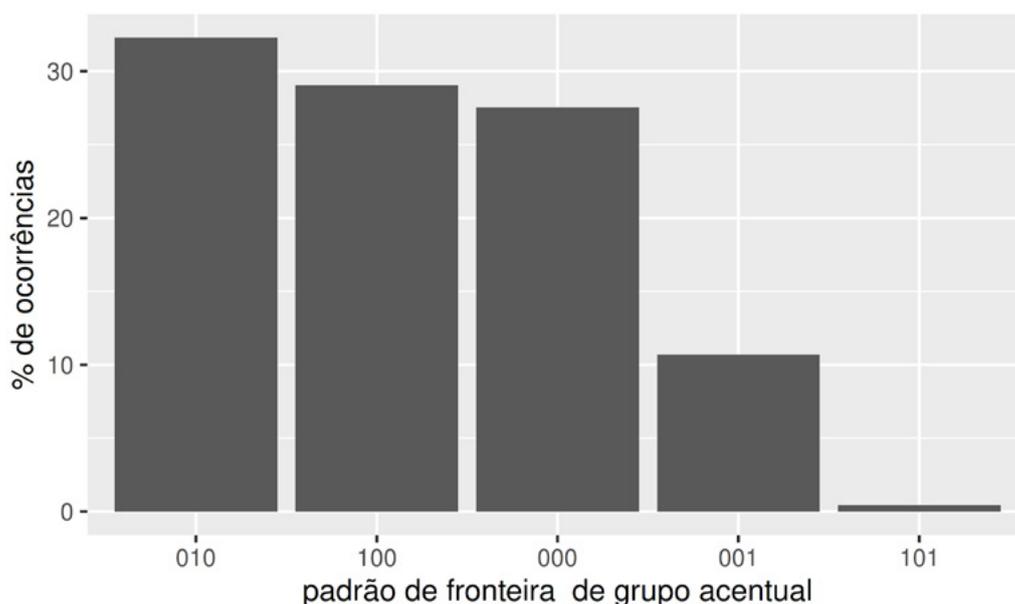
Tomados em conjunto, os resultados das duas variáveis corroboram a sugestão, comum na literatura, segundo a qual a fonação laringalizada tende a causar um rebaixamento no contorno de f_0 coextensivo à sua duração (Zeroual; Esling; Crevier-Buchman, 2008, p. 311). Os dados analisados aqui mostram que o abaixamento abrupto é uma pista robusta da presença de fonação laringalizada, embora não seja um critério que, sozinho, possa ser usado para a identificação de sua presença.

Laringalização e fronteira de grupo acentual

Dada a totalidade de ocorrências de laringalização identificados no *corpus*, 72% acontecem em unidades VV que coincidem com a fronteira de grupo acentual ou em unidade VV imediatamente adjacente. Um teste binomial exato bicaudal indica que essa proporção de laringalizações nas imediações de fronteira de grupo acentual é significativamente maior do que a proporção esperada de 50%, caso a ocorrência ou não de fronteira de grupo acentual em unidades VV afetadas por eventos de laringalização se desse ao acaso ($p < 0,001$; intervalo de confiança de $\pm 3\%$). Esse resultado indica uma ligação forte entre a laringalização e pontos fortes prosodicamente na cadeia da fala.

Como explicamos na seção “Fronteira prosódica”, a análise que reportamos aqui identifica a presença de picos de duração, que marcam a fronteira de grupo acentual, na janela de três unidades VV centrada na unidade onde ocorre o evento de laringalização. A proporção de 72% diz respeito a todas as situações em que a fronteira aconteceu, não importando sua posição dentro da janela. A figura 4 mostra a distribuição de ocorrências entre todas as configurações em que a fronteira ocorre dentro da janela. Em torno de 60% dos casos em que há a presença de fronteira de grupo acentual na janela, a laringalização coincide com a unidade VV frasalmente acentuada, padrão “010”, ou ela acontece imediatamente após a fronteira, padrão “100”, normalmente uma sílaba pós-tônica, uma vez que os acentos frasais normalmente coincidem com sílabas tônicas. São relativamente menos comuns, em torno de 10%, casos em que a laringalização precede o acento frasal.

Figura 4. Porcentagem de ocorrência dos padrões de janela de três unidades VV centradas naquela em que a laringalização acontece. O número “1” indica que uma determinada posição na janela é ocupada por fronteira de grupo acentual



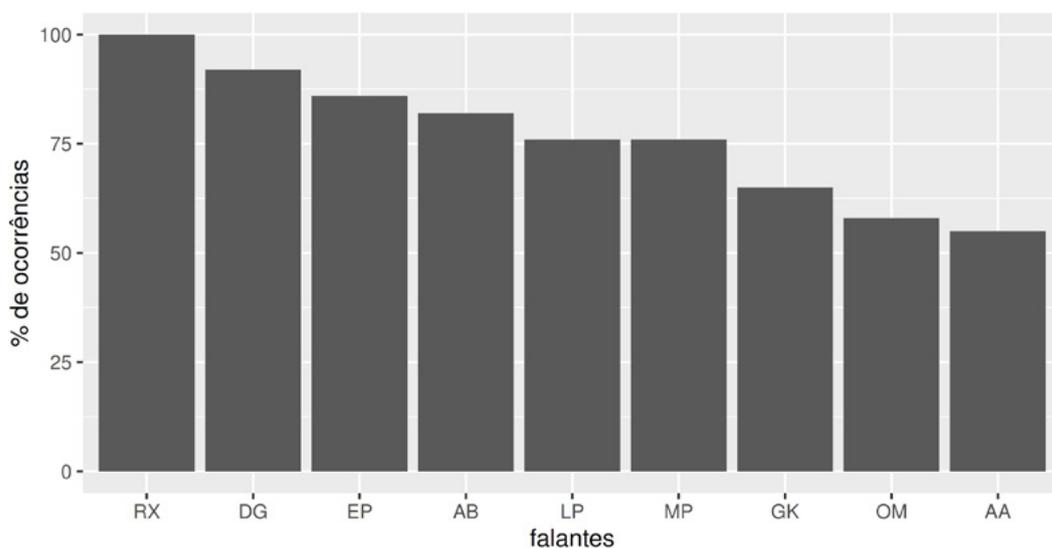
Fonte: Elaboração própria

Calculando a porcentagem de casos em que a janela de três unidades VV centrada na laringalização contém uma fronteira de grupo acentual separadamente por falante, constata-se que essa porcentagem tem uma amplitude que vai 55% a 100%. A figura 5 mostra a distribuição dos valores para os nove falantes. Um teste χ^2 que testa a hipótese nula segundo a qual a proporção de coincidências na ocorrência de laringalização e fronteira de grupo acentual é homogênea entre os nove falantes refuta essa hipótese, indicando que há diferença significativa entre os falantes [$\chi^2 (8) = 67 p < 0,001$]. Apesar

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

da variabilidade observada, mantém-se presente a tendência à associação entre laringalização e proximidade em relação à fronteira de grupo acentual. A porcentagem é maior ou igual a 65% para sete dos nove falantes. Oliveira (2017) usa uma metodologia semelhante para a determinação da posição das fronteiras de grupo acentual e reporta que 83% das laringalizações em seu estudo coincidem com fronteiras de grupo acentual, resultado que corrobora o que descrevemos aqui.

Figura 5. Porcentagem de laringalizações que acontecem nas imediações de fronteira de grupo acentual por falante



Fonte: Elaboração própria

Laringalização e disfluências

Considerando todos os falantes do *corpus*, 26% das laringalizações aconteceram concomitantemente a algum tipo de disfluência da fala. A tabela 1 mostra a distribuição dessa porcentagem em função dos sete tipos de disfluência descritos na seção “Disfluências”. A hesitação foi o tipo predominante, correspondendo sozinha a quase metade das ocorrências. Um teste χ^2 de homogeneidade rejeita a hipótese nula de uma distribuição uniforme das ocorrências entre os sete tipos de disfluência [$\chi^2(6) = 280, p < 0,001$].

A tabela 2 mostra a distribuição de laringalizações acompanhadas de algum tipo de disfluência em função dos falantes. A inspeção dos valores mostra que há uma grande variabilidade entre eles: o falante RX, por exemplo, produziu um número muito

baixo de disfluências (quatro em uma amostra de 30 segundos de fala), nenhuma delas acompanhada de laringalização; o falante EP apresenta um percentual de quase 51% de laringalizações acompanhadas de alguma disfluência. Um teste de homogeneidade rejeita a hipótese nula de uma distribuição similar das proporções entre os falantes [$\chi^2(8) = 381, p < 0,001$], indicando que os falantes diferem em termos dos tipos de disfluência que predominam em sua fala.

Tabela 1. Distribuição dos tipos de disfluência que coincidem com ocorrências de laringalização

Tipo de disfluência	Porcentagem
Hesitação	49,1
Prolongamento	17,2
Repetição	16,3
Revisão	10,5
Interjeição	4,2
Bloqueio	2,1
Palavra não terminada	0,4

Fonte: Elaboração própria

A figura 6 mostra a porcentagem de cada tipo de disfluência entre as ocorrências que foram acompanhadas de laringalização em função dos falantes (o falante RX foi omitido porque não apresentou nenhum caso desse tipo). Podemos observar que há uma diversidade de padrões entre os diferentes falantes, embora a hesitação esteja entre os tipos mais comuns para uma boa parte deles. Em função da ocorrência de muitas células vazias na tabela de contingência, não foi possível fazer o teste estatístico para determinar se a distribuição de ocorrências entre os tipos de disfluência é uniforme entre os falantes.

Tabela 2. Distribuição de ocorrências de laringalização que são acompanhadas de algum tipo de disfluência em função dos falantes

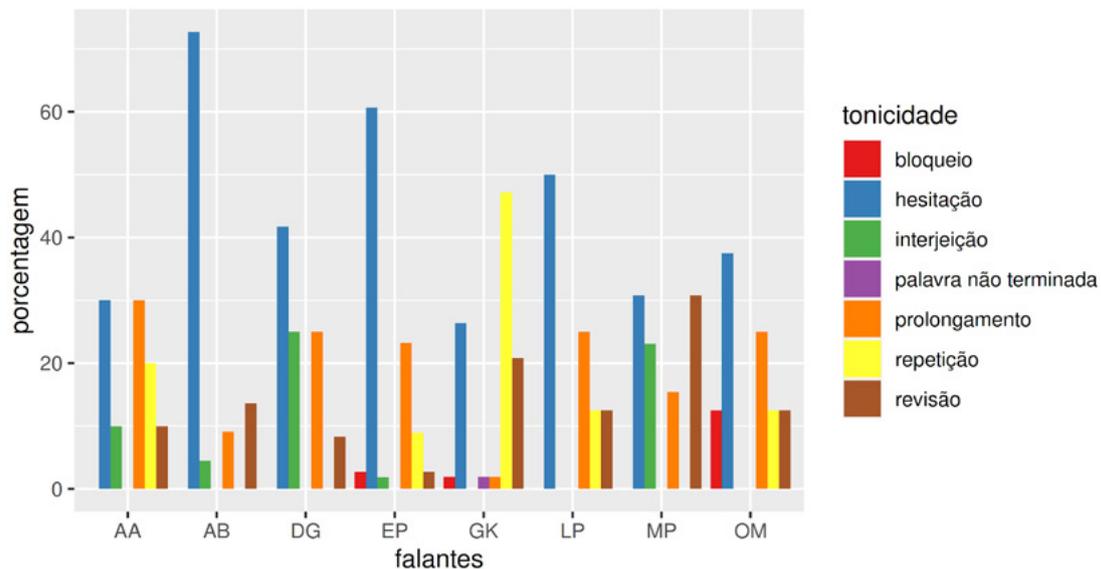
Falante	AA	AB	DG	EP	GK	LP	MP	OM
%	14,1	20	46,2	50,9	22	15,1	22,4	6,3

Fonte: Elaboração própria

Considerando agora o total de ocorrências de disfluência, observamos que 62% delas aconteceram acompanhadas de laringalização. Tomando os dois dados em conjunto, podemos dizer que aproximadamente 3 em cada 5 disfluências são acompanhadas de laringalização, mas apenas em torno de uma a cada quatro das laringalizações são acompanhadas de disfluência.

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

Figura 6. Distribuição das ocorrências de laringalização acompanhadas de disfluência em função do tipo e dos falantes



Fonte: Elaboração própria

A hesitação foi a disfluência que mais apareceu, tanto no conjunto dos falantes quanto quando se consideram os falantes individualmente. Com relação às outras disfluências, houve grande variabilidade entre os sujeitos, o que também foi encontrado em outros estudos (Maclay; Osgood, 1959; Ribeiro, 2006).

Estudo 2

Materiais e métodos

Material de fala

O material de fala analisado no estudo é parte do acervo de gravações de falantes do português brasileiro do projeto “A typology for word stress and speech rhythm based on acoustic and perceptual considerations”, coordenado por Anders Eriksson da Universidade de Estocolmo, Suécia, cujo objetivo é estudar a produção e percepção do acento lexical de uma perspectiva interlinguística.

O *corpus* contempla três estilos de elocução: fala espontânea, leitura de frases e leitura de palavras isoladas. A fala espontânea foi elicitada por meio de entrevistas

informais, conduzidas por um representante do projeto, nativo da mesma língua do participante. Transcrições ortográficas das gravações da entrevista foram usadas para produzir uma versão textual da entrevista. A partir da versão transcrita da entrevista, os pesquisadores do projeto selecionaram frases em trechos fluentes e sem erros de fala. Em momento posterior, os participantes foram chamados novamente para gravar a leitura das frases selecionadas na etapa anterior. De cada frase selecionada, os pesquisadores escolheram uma palavra e a leitura da lista dessas palavras também foi gravada na segunda sessão. O procedimento permitiu a produção do mesmo material em três estilos de elocução diferentes. Dez participantes nativos do português brasileiros foram gravados no contexto do projeto, cinco do sexo feminino, cinco do sexo masculino. A idade média dos participantes é de 23 anos (gama de 18 a 32 para o sexo feminino e 20 a 30 no caso dos falantes masculinos). Todos falam a variedade linguística típica do interior do estado de São Paulo e eram estudantes universitários em nível de graduação ou pós-graduação no momento da gravação.

A duração média líquida (em segundos, desvio-padrão entre parênteses) das amostras de áudio é de 565 (166) para as entrevistas, 191 (40) para a leitura de frases e 38 (9) para a leitura de palavras. A duração média líquida (em segundos, desvio-padrão entre parênteses) das amostras de áudio é de 565 (166) para as entrevistas, 191 (40) para a leitura de frases e 38 (9) para a leitura de palavras. O número mediano (desvio mediano absoluto entre parênteses) de unidades interpausais em cada entrevista é de 274 (54); o número mediano de frases no estilo leitura de frases é 67 (27); o número mediano de palavras é de 46 (1) no estilo leitura de palavras.

Análise fonética

Vozeamento não modal

Usamos como critérios para identificação de instâncias de vozeamento não modal a oitiva cuidadosa das amostras de fala e o exame visual tanto da forma de onda quanto do espectrograma nos casos suspeitos identificados pela oitiva. As principais pistas visuais usadas para guiar a identificação foram o padrão de estriamento irregular ou muito espaçado no caso da laringalização, associado ou não à diminuição concomitante da amplitude na forma de onda. A duração das instâncias de laringalização foram marcadas temporalmente em arquivos *TextGrid* do programa Praat. Diferentemente do Estudo 1, fizemos o registro de todos os tipos de vozeamento não modal que ocorreram nas amostras do *corpus*, embora não tenhamos discriminado os diferentes tipos com etiquetas

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

diferentes e o tipo de vozeamento não modal não tenha sido usado como variável independente. Predominou largamente o tipo laringalização, embora tenham ocorrido em menor número também a aspereza vocal (*harsh voice*, em inglês) e falsete.

Estrutura prosódica

Dada cada ocorrência de vozeamento não modal, registramos a tonicidade da sílaba afetada. Em palavras com duas ou mais sílabas foram registradas a tonicidade que poderia ser pretônica, tônica e postônica; monossílabos foram classificados em tônicos ou átonos. A grande maioria dos casos afetaram apenas a vogal e por isso não contamos separadamente ocorrências em consoantes e vogais, como no Estudo 1. Houve casos em que a ocorrência de vozeamento não modal ocorreu em trechos de disfluência. Esses casos foram codificados juntamente com a variável tonicidade, mas com uma etiqueta separada; entraram nessa categoria hesitações, reparos, marcadores conversacionais (ocorrências como *né, ah, hum, ahn*), palavras de conteúdo ou gramaticais que não são usadas com seu significado usual na fala (como o “é” prolongado com valor de hesitação).

Registramos também as ocorrências de vozeamento não modal em relação à sua posição quanto às fronteiras prosódicas. A unidade prosódica cujas fronteiras foram observadas foi definida usando como critério delimitador a presença de pausas silenciosas, tomando por base Kendall (2013, p. 138), que define enunciado fonético (*phonetic utterance* ou *interpausal stretches*), como “stretches of uninterrupted phonation surrounded by silence”. Escolhemos o critério da pausa para definir a presença de fronteira por uma razão prática: pausas silenciosas são facilmente identificáveis de maneira visual ou de forma semiautomática e sem recurso à análise morfológica ou sintática. Além disso, as fronteiras dessa unidade podem ser consideradas fortes e nisso o enunciado fonético se aproxima, embora não corresponda, exatamente, às fronteiras do enunciado fonológico, o constituinte prosódico mais alto na hierarquia definida por Nespor e Vogel (2007). Aspectos que aproximam o enunciado fonético ao enunciado fonológico são a presença de alongamento final (*final lengthening*) e abaixamento de f_0 em posição final. A variável posição tem três possibilidades de marcação: início, meio e final de enunciado fonético. No caso dos estilos de elocução entrevista e leitura de frases, a atribuição das etiquetas tomou por base a palavra prosódica (palavra morfológica mais seus clíticos), de modo que ocorrências que afetam qualquer sílaba da primeira palavra prosódica que vem após pausa inicial ganha a etiqueta “início” e as que ocorrem na última palavra fonológica antes de pausa recebem a etiqueta “final”; no caso do estilo leitura de palavras, o critério para a definição da etiqueta foi a própria palavra, de modo que a etiqueta “início” foi aplicada se a ocorrência de vozeamento não modal afetou a primeira sílaba da palavra, a etiqueta

“final” se a ocorrência afetou a última sílaba e a etiqueta “meio” apenas se a palavra tinha três ou mais sílabas e uma sílaba não inicial ou não final foi afetada.

Análise estatística

As variáveis dependentes no estudo são a taxa de vozeamento não modal, duração dos eventos de vozeamento não modal e percentual de duração dos eventos de vozeamento não modal relativamente ao total do tempo de fala.

As variáveis independentes no estudo são:

- **Estilo** de elocução: entrevista, leitura de frases e leitura de lista de palavras;
- **Tonicidade** da sílaba: pré-tônica, tônica, pós-tônica, monossílabo (tônicos e átonos) e trecho de disfluência;
- **Posição** do evento de vozeamento não modal no enunciado fonético: inicial, medial ou final;
- **Sexo** do falante: feminino e masculino.

O principal tipo de análise foi o da distribuição das ocorrências de vozeamento não modal em função das diferentes variáveis independentes e de cruzamentos entre elas. Testes χ^2 de homogeneidade foram usados para determinar se os subníveis das variáveis independentes geram efeitos significativos sobre a distribuição das frequências de ocorrência do vozeamento não modal.

Analizamos também a taxa de ocorrência de vozeamento não modal por meio da divisão do número de eventos desse tipo em uma determinada amostra de fala pela duração (em segundos) da amostra. Essa medida será usada para estimar quão comum é esse tipo de ocorrência na fala e será calculada em função dos estilos de elocução e dos falantes. Assim, teremos 30 valores (10 falantes \times 3 estilos) no total, que podem ser agrupados em função das variáveis independentes. Os dados de taxa de vozeamento não modal serão analisados estatisticamente por meio do teste de análise de variância (ANOVA). O pressuposto de homogeneidade de variância é verificado por meio do teste Fligner-Killeen (abreviado FK na apresentação dos resultados). O teste não paramétrico Kruskal-Wallis (abreviado KW) foi usado quando os pressupostos foram violados.

Para todas as análises estabelecemos um nível alfa de 5% para rejeição da hipótese nula de todos os testes estatísticos. O ambiente de computação estatística R (R CORE TEAM, 2020) foi usado para todas as fases da análise estatística.

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

Resultados

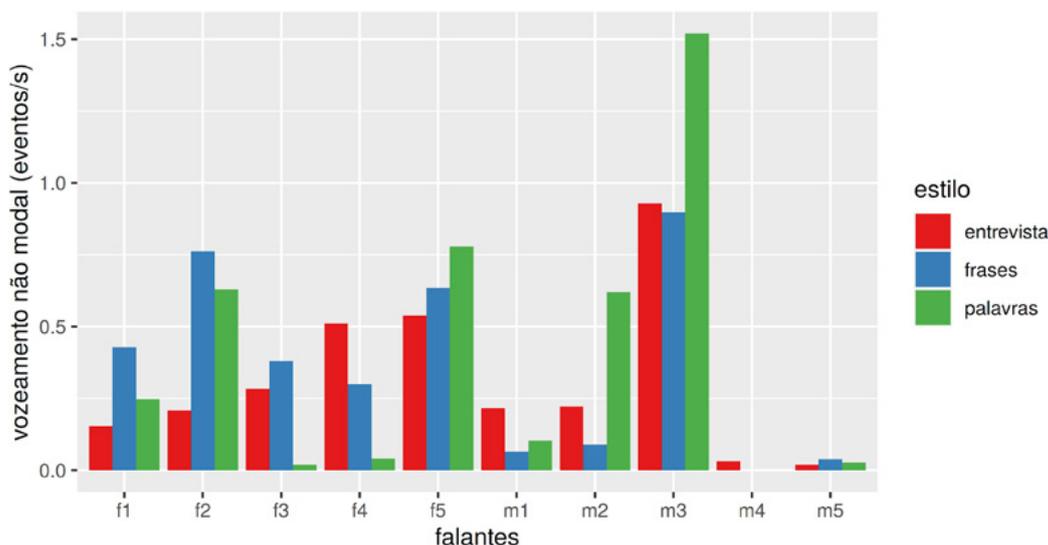
Descrição geral dos eventos de fonação não modal

Foram registrados 2481 casos de vozeamento não modal, considerando todos os falantes e todos os estilos conjuntamente. A taxa de ocorrência é de 0,31 eventos por segundo. A duração média dos eventos é de 134 ms (DP 108). A duração total da fonação não modal tem uma média geral de 4,5% do tempo total de fala no *corpus*; o percentual calculado por falantes varia entre 0,21% e 16,8%. Todos os falantes produziram ocorrências de vozeamento não modal nos três estilos, com exceção do falante *m4*, que só mostrou ocorrência no estilo entrevista.

Calculando a taxa de ocorrência por falante, os valores variam entre 0,01 (*speaker m4*) e 1,12 (*m3*). A amostra é homocedástica [$\chi^2(9) = 5,3$ *ns*] e os resíduos do modelo linear são distribuídos de acordo com uma distribuição normal. Uma ANOVA univariada mostrou um efeito significativo dos falantes sobre a taxa média de fonação não modal [$F(9, 20) = 8,07$ $p < 0,001$]. A figura 7 mostra a taxa de vozeamento não modal em função tanto do falante quanto do estilo de elocução. A inspeção visual da figura sugere que há variação intra quanto interfalante. Os falantes variam em termos do quanto eles usam vozeamento não modal de maneira geral: *m1*, *m4* e *m5* exibem taxas menores, enquanto *f2*, *f5* e *m3* mostram médias maiores. A influência do estilo de elocução sobre a taxa também não é uniforme entre os falantes: para alguns, a entrevista tem a maior média (*f4*, *m1*, *m2*), enquanto para outros os estilos leitura de frases (*f1*, *f3*) ou de palavras (*f5*, *m2*, *m3*) têm a maior média. Entre todos os falantes, *m3* é quem apresenta as maiores taxas. Esse resultado pode ser explicado pelo fato de esse falante ter um caso leve de *sulcus vocalis*, conforme a avaliação de duas fonoaudiólogas que analisaram a produção do falante de forma independente a pedido dos autores. A fala de *m3* apresenta mais ocorrências da qualidade que pode ser descrita como aspereza (*harsh voice*) do que os demais falantes, muito embora esse tipo de qualidade de fala não seja considerado patológico, nem seja permanente na produção desse falante.

A figura 7 sugere a existência de um efeito do sexo dos falantes sobre o vozeamento não modal. Falantes masculinos tendem a apresentar taxas menores do que os femininos, com exceção do falante *m3*, que mostra valores de taxas mais elevados pelas razões apontadas no parágrafo anterior. A taxa média para falantes femininos é 0,39 (DP 0,24) e 0,32 (DP 0,46) para os masculinos; o valor da média masculina cai para 0,12 (DP 0,17) quando o falante *m3* é excluído da amostra. Testes mostram que o efeito do sexo dos falantes é significativo quando *m3* é excluído da amostra masculina [$\chi^2(1) = 9,16$ $p = 0,002$], mas não quando *m3* é mantido [$\chi^2(1) = 2,68$ *ns*].

Figura 7. Taxa de ocorrência de eventos de fonação não modal por segundo em função dos falantes e estilos de elocução



Fonte: Elaboração própria

De maneira similar à taxa de ocorrência de vozeamento não modal, a duração dos eventos também é afetada pelos falantes [FK: $\chi^2(9) = 170, p < 0,001$; KW: $\chi^2(9) = 334, p < 0,001$]. A duração média calculada por falante varia entre 94 ms e 248 ms. Resultados de comparações pareadas indicam que os falantes *m2*, *m3*, *m4* e *m5* formam um grupo com médias de duração cujos valores não diferem significativamente entre si. Não há agrupamentos homogêneos entre as falantes do sexo feminino.

Eventos de vozeamento não modal são mais longos no estilo entrevista (média 145 ms, DP 123) do que no estilo leitura de frases (média 108 ms, DP 55) ou leitura de palavras (média 106 ms, DP 53) [FK: $\chi^2(2) = 57,4, p < 0,001$; KW: $\chi^2(2) = 43,9, p < 0,001$].

Eventos de vozeamento não modal são mais longos em trechos disfluentes (246 ms); são mais breves (80 ms) em sílabas pré-tônicas; eventos em monossílabos (121 ms), sílabas acentuadas (122 ms) e sílabas pós-tônicas (117 ms) apresentam valores intermediários [FK: $\chi^2(6) = 256, p < 0,001$; KW: $\chi^2(2) = 309, p < 0,001$].

Efeitos da posição no enunciado fonético

Iniciamos a descrição dos resultados pelo efeito da variável posição (as demais variáveis independentes agrupadas) sobre a distribuição das ocorrências de vozeamento não modal: posição inicial 16%, medial 48% e final 36%. O teste χ^2 mostra um efeito

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

significativo [$\chi^2(2) = 375,98 p < 0,001$], indicando que a distribuição entre as três posições não é homogênea. Os dados mostram uma preferência por ocorrências de vozeamento não modal em posição medial ou final. A prevalência de casos na posição medial não surpreende, em especial nos estilos entrevista e leitura de frases, uma vez que só há uma palavra possível nas posições inicial ou final, mas em geral há muitas palavras em posição medial; portanto, em posição medial há mais oportunidades para a ocorrência de eventos de vozeamento não modal. Além disso, os casos em posição medial podem coincidir com fronteiras prosódicas menos fortes, que não coincidem com pausa silenciosa. Considerando essa vantagem natural da posição medial, o fato de haver 36% de ocorrência em posição final indica que essa posição tem uma força considerável de atração para ocorrência de vozeamento não modal.

Considerando todos os enunciados fonéticos do *corpus* (3911) e a quantidade total de eventos de vozeamento não modal (2481), a porcentagem deles em que a posição inicial é alvo de vozeamento não modal é de 10%, medial 30% e final 23%. Desta perspectiva, a vantagem da posição medial se mantém, embora menos pronunciada e a posição final continua em segundo lugar; 1 em cada 4 das palavras que ocupam a posição final são alvo de eventos de vozeamento não modal.

Uma vez que o estilo leitura de palavras difere dos outros dois no sentido de que naquele cada enunciado fonético coincide com uma palavra e nos outros dois os enunciados são compostos normalmente por mais do que uma palavra, é importante discutir o efeito do estilo sobre a distribuição do vozeamento não modal relativamente às posições em que eles podem ocorrer dentro de cada enunciado fonético. A tabela 3 mostra os valores dessa distribuição em porcentagens.

Tabela 3. Porcentagem dos eventos de vozeamento não modal em função da posição relativamente às fronteiras e dos estilos de elocução

	Inicial	Medial	Final
Entrevista	18	50	32
Frases	12	42	46
Palavras	14	46	40

Fonte: Elaboração própria

O exame da tabela 3 indica que o maior contraste entre os estilos é aquele observado entre entrevista e leitura de frases: há uma migração das ocorrências em posição inicial e medial na entrevista para a posição final na leitura de frases; a maior migração se dá na posição medial, que perde 14 pontos percentuais, em comparação com a posição inicial, que perde 6% percentuais. O estilo leitura de palavras fica em situação intermediária

entre os outros dois. A diferença causada pelo estilo de elocução sobre a distribuição do vozeamento não modal entre as diferentes posições é estatisticamente significativa [$\chi^2(4) = 46,3 p < 0,001$].

A tabela 4 mostra o efeito do sexo dos falantes sobre a distribuição dos eventos de vozeamento não modal entre as posições em que eles podem ocorrer dentro de cada enunciado fonético. A variável sexo não tem efeito significativo sobre a distribuição ao longo das posições no enunciado fonético [$\chi^2(2) = 2,46 ns$]. O teste foi repetido com a exclusão do falante *m3* da amostra geral e o resultado continuou sendo não significativo [$\chi^2(2) = 4,01 ns$].

Tabela 4. Porcentagem dos eventos de vozeamento não modal em função da posição relativamente às fronteiras e do sexo dos falantes

	Inicial	Medial	Final
Femininas	37	17	46
Masculinos	35	16	50

Fonte: Elaboração própria

Efeito da tonicidade das sílabas

Discutimos agora o efeito da variável tonicidade sobre as ocorrências de vozeamento não modal. Casos em que o vozeamento não modal afetou mais do que uma sílaba sempre envolveram a tônica, isto é, o vozeamento não modal abrangeu a pretônica e a tônica ou tônica e postônica. Casos desse tipo foram observados nos dados de sete dos dez falantes, mas foram muito pouco numerosos, representaram entre 1 e 2% do total. Por isso, agrupamos esses casos juntamente com as ocorrências em sílaba tônica nos dados reportados aqui. A distribuição do total dos eventos de vozeamento não modal pelos graus de tonicidade foi a seguinte: monossílabos átonos 14%, pretônicas 7%, tônicas 58%, postônicas 11% e disfluências 10%. O teste estatístico confirma o exame das porcentagens e indica que a distribuição dos casos entre os graus de tonicidade não é homogênea [$\chi^2(4) = 2310 p < 0,001$]. Sílaba tônica atrai em torno de 60% das ocorrências de vozeamento não modal; os demais graus de tonicidade têm em torno de 10%, bem como os casos de trechos disfluentes.

O resultado do cruzamento das variáveis tonicidade e estilos de elocução gera os dados mostrados na tabela 5. Na comparação entre entrevista e leitura de frases, se destaca o aumento de 16% nas ocorrências em posição tônica e diminuição de 12% em ocorrências em trechos disfluentes. O resultado faz sentido, uma vez que a leitura de frases oferece a mensagem já elaborada, permitindo um grau maior de planejamento antecipado da

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

fala, que resulta na diminuição da interrupção da fluência. No estilo leitura de palavras, as palavras são pelo menos dissílabos, o que justifica a falta de ocorrências na categoria monossílabos átonos. Não houve ocorrências de disfluências nesse estilo. Nesse estilo nota-se também um aumento no caso de sílabas postônicas afetadas por vozeamento não modal. Esse aumento pode ser justificado pelo fato de que, no caso da leitura de palavras, a sílaba postônica coincide com a posição final do enunciado fonético e vimos antes que a posição final tende a atrair eventos de vozeamento não modal. O teste estatístico indica um efeito significativo do estilo sobre a distribuição das ocorrências entre os graus de tonicidade [$\chi^2(8) = 157 p < 0,001$].

Tabela 5. Porcentagem dos eventos de vozeamento não modal em função da posição relativamente às fronteiras e do sexo dos falantes

	monossílabos átonos	pretônica	tônica	postônica	disfluência
Entrevista	15	7	54	10	14
Frases	11	8	70	9	2
Palavras	0	15	62	23	0

Fonte: Elaboração própria

A tabela 6 mostra a distribuição das ocorrências de vozeamento não modal em função da tonicidade da sílaba afetada e do sexo dos falantes. A principal observação é que os falantes do sexo masculino apresentam um aumento expressivo, de 14%, de ocorrências em trechos disfluentes. O aumento vem quase todo das ocorrências em sílabas tônicas. O efeito do sexo do falante sobre a distribuição dos eventos entre os graus de tonicidade é estatisticamente significativo [$\chi^2(8) = 137 p < 0,001$].

Tabela 6. Porcentagem dos eventos de vozeamento não modal em função da posição relativamente às fronteiras e do sexo dos falantes

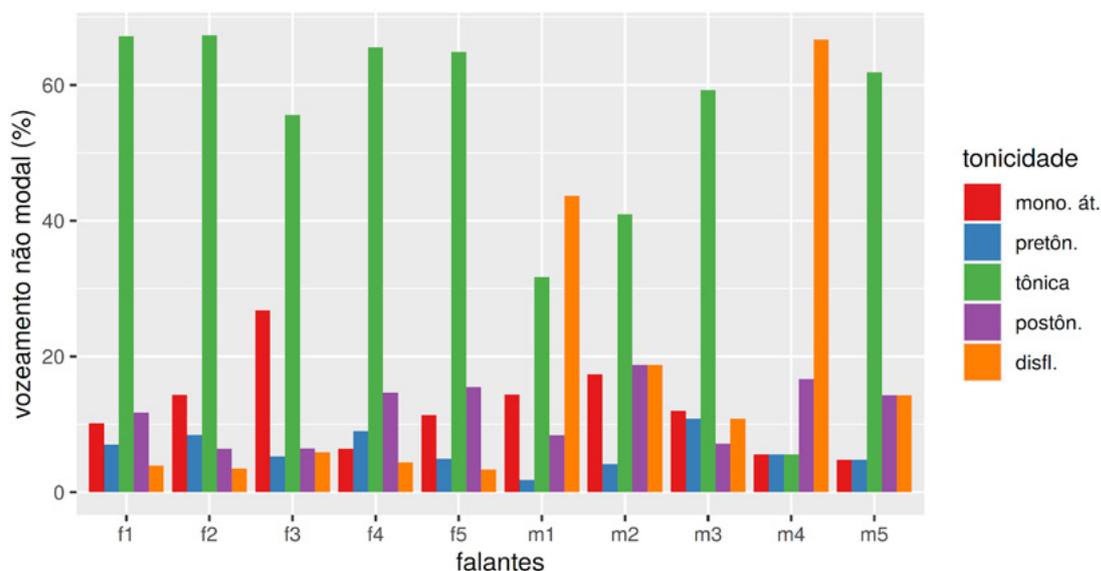
	monossílabos átonos	pretônica	tônica	postônica	disfluência
Femininas	14	7	63	12	4
Masculinas	13	8	51	9	18

Fonte: Elaboração própria

A figura 8 ilustra a distribuição das ocorrências de vozeamento não modal em função da tonicidade da sílaba afetada e dos falantes, organizados por sexo e permite ver a variação individual. Falantes do sexo feminino apresentam maior homogeneidade. Predominam as ocorrências em sílaba tônica (em torno de 60% das ocorrências em todas as falantes). Falantes do sexo masculino apresentam menos homogeneidade: *m3* e *m5* mais parecidos com o padrão do sexo feminino; os falantes *m2*, *m1* e *m4* apresentam quantidades crescentes (em torno de 20%, 40% e 60%) de ocorrências de vozeamento

em trechos disfluente (barras laranjas), que parecem ser compensadas por reduções nas ocorrências em sílabas tônicas. O teste estatístico indica que os falantes exercem um efeito significativo sobre a distribuição dos eventos de vozeamento não modal entre os graus de tonicidade [$\chi^2(36) = 487 p < 0,001$]. Esse resultado se mantém ainda que sejam analisados separadamente apenas falantes do mesmo sexo: feminino [$\chi^2(16) = 93,9 p < 0,001$] ou masculino [$\chi^2(16) = 176 p < 0,001$].

Figura 8. Porcentagem de eventos de vozeamento não modal em função da tonicidade da sílaba afetada e do falante



Fonte: Elaboração própria

Considerações finais

Retomamos agora o que consideramos os resultados mais relevantes aportados pelos dois estudos relatados no trabalho. Destacamos aqueles que confirmam achados anteriores, tanto no que diz respeito ao português brasileiro (Lima, 2011; Oliveira, 2017) quanto a outras línguas (Dilley; Shattuck-Hufnagel; Ostendorf, 1996; Gordon; Ladefoged, 2001) e também os resultados que entendemos representarem as maiores novidades no campo.

O primeiro resultado que destacamos é o que aponta para a relação entre a frequência fundamental e a laringalização. Os achados do presente trabalho confirmam resultados anteriores que sugerem que rebaixamentos repentinos, transitórios e com magnitudes de cerca de meia oitava acompanham uma proporção bastante alta de ocorrências de laringalização. Foram observados também nos dados casos relativamente

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

numerosos nos quais a irregularidade dos pulsos glotais impediu a adequada detecção automática de f_0 . Esses resultados reafirmam uma série de apontamentos encontrados na literatura a respeito da relação entre f_0 e laringalização, fortalecendo o argumento em favor do uso de rebaixamentos de f_0 com certas características como uma pista confiável para a detecção de laringalização.

Um segundo conjunto de resultados que destacamos diz respeito à predominância das vogais como os segmentos afetados preferencialmente pelo vozeamento laringalizado, que confirma achados anteriores. Do mesmo modo, a estimativa da taxa de laringalização, em especial calculada como a porcentagem de tempo de fala tomado por vozeamento não modal, confirma resultados anteriores reportados por Oliveira (2017). As médias gerais e mesmo a amplitude de variação das médias dos falantes são próximas. Como se trata de três *corpora* independentes, um de Oliveira e dois no presente estudo, os resultados mostram-se robustos. A presença de variabilidade interfalante e de um efeito do sexo do falante merecem exploração em pesquisas futuras – em especial no segundo estudo, as falantes do sexo feminino apresentam taxa de vozeamento não modal maior do que os do sexo masculino. Destacamos que, na revisão da literatura que deu base para a elaboração do trabalho, não encontramos referências de estimativa para a taxa de ocorrência de vozeamento não modal para outras línguas.

A literatura prévia, em especial Dilley *et al.* (1996) e Gordon e Ladefoged (2001), aponta para uma relação entre posições prosodicamente fortes e a ocorrência de vozeamento não modal em diferentes línguas. Uma das contribuições do presente trabalho que consideramos mais original foi a adoção, em especial no Estudo 1, da fronteira de grupo acentual, fundamentada em aspectos de produção e percepção, como critério para a definição de “posição prosódica forte”. No Estudo 2, as posições fortes foram consideradas as sílabas tônicas, quer fossem portadoras de acentos frasais ou não, e posições de final de enunciado, seguidas sempre de pausa silenciosa. Nos dois estudos, a predominância de casos nessas posições é marcante e indica que a prosódia é um fator relevante que condiciona a distribuição das instâncias de vozeamento não modal nos enunciados. O fato de os resultados dos dois estudos serem convergentes, apesar da diferença de critérios adotados para a definição de posição prosódica forte, pode ser considerada evidência para a força e robustez dos achados. Tomamos o conjunto dos resultados como uma indicação de que o vozeamento não modal, em especial a laringalização, tem como um uso linguístico no português brasileiro funcionar como correlato adicional do acento, tanto no nível da palavra quanto no nível do enunciado. Não encontramos na literatura a respeito do português brasileiro outro estudo que investigue essa relação da mesma maneira que fizemos aqui, de modo que não é possível comparar os nossos resultados com o de outros trabalhos.

Em relação às disfluências, resultados que podem ser considerados inéditos pelo menos no que diz respeito ao PB mostram que em torno de um quarto das laringalizações são acompanhadas de episódios de disfluências, a hesitação sendo o tipo mais comum nesses casos. Tomando como referência o total de hesitações, em torno de 60% das ocorrências são acompanhadas de laringalizações. Esse resultado sugere que as hesitações são um lugar preferencial para a ocorrência de laringalização, embora o contrário não seja verdadeiro. A taxa de produção de eventos de laringalização acompanhada de disfluência é um comportamento variável entre os falantes e é outro aspecto que pode ser investigado com mais profundidade em trabalhos posteriores.

O Estudo 2 mostrou que o estilo de elocução exerce efeitos significativos sobre a distribuição do vozeamento não modal. Um desses efeitos manifesta-se na duração dos eventos de vozeamento não modal, que são mais longos no estilo entrevista do que nos estilos leitura de frases e palavras. Do ponto de vista da posição do evento de vozeamento não modal relativamente às fronteiras do enunciado fonético, um dos efeitos mais importantes são as diferenças entre o estilo entrevista e o estilo leitura de frases – no segundo, há uma migração de ocorrências das posições inicial e medial para a posição final. Em termos da tonicidade da sílaba onde ocorrem os eventos de vozeamento não modal, os efeitos do estilo manifestam-se na forma de maior ocorrência relativa nas tônicas na leitura de frases e maior ocorrência relativa nas pós-tônicas na leitura de palavras. No estilo entrevista, o que mais chama a atenção é a maior ocorrência relativa de eventos de vozeamento não modal em episódios de disfluência, que ocorrem em baixo número na leitura de frases e são ausentes na leitura de palavras.

Agradecimentos

O presente trabalho baseia-se parcialmente nos resultados do projeto de pesquisa “Efeitos do vozeamento não modal na descrição estatística de amostras de frequência fundamental”, financiado pela FAPESP (processo 2016/16544-8) na modalidade Iniciação Científica, concedida a Isabela de Jesus Silveira e orientado pelo primeiro autor. O primeiro autor agradece a Anders Eriksson, da Universidade de Estocolmo, por ceder as gravações do português brasileiro analisadas no presente trabalho, que compõem o acervo do projeto “A typology for word stress and speech rhythm based on acoustic and perceptual considerations”. O presente trabalho também baseia-se parcialmente nos resultados dos projetos de pesquisa “Identificação da laringalização em *corpus* de fala espontânea” e “Relação entre laringalização e disfluência comum em *corpus* de fala espontânea do português brasileiro”, financiados pelos programas de iniciação científica PIBIC-CNPq e PROIC-FAPDF, respectivamente, que concederam bolsa a Nathalia Elyne Vasconcelos de Sousa, e orientados pela segunda autora.

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

Referências

ALBANO, E. C.; MOREIRA, A. A. Archisegment-based letter-to-phone conversion for concatenative speech synthesis in Portuguese. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96. **Anais...** In: PROCEEDING OF FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. ICSLP '96. Philadelphia, PA, USA: IEEE, 1996. Disponível em: <http://ieeexplore.ieee.org/document/607956/>. Acesso em: 18 jan. 2023.

ANDREEVA, B. *et al.* Prosodic effects of implicit causality in German and Norwegian? (E. L. Asu, P. Lippus, Eds.) Nordic prosody: proceedings of the XIth conference, Tartu 2012. **Anais...** In: NORDIC PROSODY: PROCEEDINGS OF THE XITH CONFERENCE, TARTU 2012. Tartu, Estonia: 2013.

ARANTES, P. **better_f0**: A Praat script for better f0 extraction. [s.l.] Zenodo, 2019.

ARANTES, P. **parantes/f0-outliers**: Version 1.0. Zenodo, 8 jul. 2021a. Disponível em: <https://zenodo.org/record/5083453>. Acesso em: 7 fev. 2023.

ARANTES, P. **parantes/duration_suite**: Initial release. Zenodo, 22 set. 2021b. Disponível em: <https://zenodo.org/record/5522291>. Acesso em: 22 set. 2021.

BARBOSA, P. A. **Incursões em torno do ritmo da fala**. Campinas: Pontes, 2006.

BARBOSA, P. A. **Prosodia**. 1. ed. São Paulo: Parábola, 2019.

BECKMAN, M. E. A Typology of Spontaneous Speech. In: SAGISAKA, Y.; CAMPBELL, N.; HIGUCHI, N. (ed.). **Computing Prosody**. New York, NY: Springer US, 1997.

BLANKENSHIP, B. The timing of nonmodal phonation in vowels. **Journal of Phonetics**, v. 30, n. 2, p. 163-191, abr. 2002.

BOERSMA, P. Praat, a system for doing phonetics by computer. **Glott International**, v. 5, n. 9/10, p. 341-345, 2001.

DILLEY, L.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M. Glottalization of word-initial vowels as a function of prosodic structure. **Journal of Phonetics**, v. 24, n. 4, p. 423-444, out. 1996.

GORDON, M.; LADEFOGED, P. Phonation types: a cross-linguistic overview. **Journal of Phonetics**, v. 29, n. 4, p. 383-406, out. 2001.

GREISBACH, R. Some aspects of maximally fast reading style. ESCA Workshop Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication. **Anais... In: ESCA WORKSHOP PHONETICS AND PHONOLOGY OF SPEAKING STYLES: REDUCTION AND ELABORATION IN SPEECH COMMUNICATION.** Barcelona: 1991.

HELDNER, M. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. **Journal of Phonetics**, v. 31, n. 1, p. 39-62, 2003.

JANNEDY, S. Prosodic Focus in Vietnamese. **Interdisciplinary Studies on Information Structure**, v. 8, p. 209-230, 2007.

KENDALL, T. **Speech Rate, Pause, and Sociolinguistic Variation: Studies in Corpus Sociophonetics.** London: Palgrave Macmillan, 2013.

KIM, Y.; VALDOVINOS, M. The interaction of laryngealized vowels, stress, and falling pitch in Mariteco Cora. Proceedings of the Workshop on the Sound Systems of Mexico and Central America. **Anais...In: WORKSHOP ON THE SOUND SYSTEMS OF MEXICO AND CENTRAL AMERICA.** New Haven, CT: 2014.

KOCH, I.; SOUZA E SILVA, M. C. P. de. **Gramática do Português Falado**, vol. IV, Campinas: Editora da Unicamp-FAPESP, 1996.

KOHLER, K. Linguistic and paralinguistic functions of non-modal voice in connected speech. Proceeding of 5th Seminar on Speech Production: Models and Data. **Anais...** 2000.

LADEFOGED, P. **Preliminaries to Linguistic Phonetics.** Chicago: University of Chicago Press, 1972.

LADEFOGED, P. **A Course in Phonetics.** 4th. ed. Boston, MA: Heinle & Heinle, 2001.

LAVER, J. **The phonetic description of voice quality.** New York: Cambridge University Press, 1980.

- | Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro

LEVY PODOLSKY, P. **Fonología del totonaco de Papantla, Veracruz**. 1. ed. México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Filológicas, Seminario de Lenguas Indígenas, 1987.

LIMA, A. M. **Oclusiva glotal e laringalização em sujeitos com fissura palatina**: um estudo segundo abordagem dinamicista. 2011. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2011.

LLISTERRI, J. Speaking styles in speech research. ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language. **Anais...** Dublin, Ireland: 1992.

LUCENTE, L. **DaTo**: um sistema de notação entoacional para o português brasileiro baseado em princípios dinâmicos. Ênfase no foco e na fala espontânea. 2012. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2012.

MACLAY, H.; OSGOOD, C. E. Hesitation Phenomena in Spontaneous English Speech. **WORD**, v. 15, n. 1, p. 19-44, jan. 1959.

MARCUSCHI, L. A. **Cognição, explicitude e autonomia do texto falado e escrito**. Recife: UFPE, 1999.

MARTINS, M. V.; FERREIRA NETTO, W. Prosódia e escalas de frequência: um estudo em torno da escala de semitons. **Revista Virtual de Estudos da Linguagem**, v. 8, p. 286-296, 2010.

MERLO, S. **Hesitações na fala semi-espontânea**: análise por séries temporais. 2006. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2006.

NESPOR, M. A.; VOGEL, I. **Prosodic phonology**: with a new foreword. 2. ed. Berlin: Mouton de Gruyter, 2007.

OLIVEIRA, J. C. C. **Laringalização no português brasileiro**: uma análise em torno do fenômeno laríngeo e implicações para a comparação de locutor. 2017. Dissertação (Mestrado em Linguística) – Universidade Federal de Alagoas, Maceió, 2017.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2020.

REDI, L.; SHATTUCK-HUFNAGEL, S. Variation in the realization of glottalization in normal speakers. **Journal of Phonetics**, v. 29, n. 4, p. 407-429, 2001.

RIBEIRO, S. E. B. A. **Contributos para a caracterização do discurso oral: um estudo de disfluências**. 2006. Dissertação (Mestrado em Linguística) – Universidade de Aveiro, Aveiro, 2006.

SCHIEFER, A. M.; ARCURI, C. F. Avaliação da fluência da fala. *In*: MARCHESAN, I. Q.; JUSTINO, H.; TOMÉ, M. C. (ed.). **Tratado de especialidades em fonoaudiologia**. [s.l.] Editora Roca, 2000.

SOUSA, N. E. V.; LIMA, A. M. Identificação da laringalização em *corpus* de fala espontânea. 21º Congresso de Iniciação Científica da UnB e 12º do DF. **Anais... In**: 21º CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UnB E 12º DO DF. Brasília: 2015.

TILSEN, S. Syllable Timing in Repetition Disfluency. **UC Berkeley Phonology Lab Annual Reports**, v. 2, 2006.

XU, Y. In defense of lab speech. **Journal of Phonetics**, v. 38, n. 3, p. 329-336, jul. 2010.

ZEROUAL, C.; ESLING, J. H.; CREVIER-BUCHMAN, L. The contribution of supraglottic laryngeal adjustments to voice: Phonetic evidence from Arabic. **Logopedics Phoniatics Vocology**, v. 33, n. 1, p. 3-11, jan. 2008.

COMO CITAR ESTE ARTIGO: ARANTES, Pablo; LIMA, Aveliny Mantovan. Efeitos da prosódia e da disfluência sobre a qualidade de voz em diferentes estilos de elocução no português brasileiro. **Revista do GEL**, v. 20, n. 2, p. 62-98, 2023. Disponível em: <https://revistadogel.gel.org.br/>

Submetido em: 16/03/2023 | Aceito em: 12/09/2023.
