

Função de transferência das vogais orais do Português brasileiro: análise acústica comparativa

*Maria Inês Rebelo Gonçalves*¹, *Paulo Augusto de Lima Pontes*², *Vanessa Pedrosa Vieira*³, *Antônio Augusto de Lima Pontes*⁴, *Daniella Curcio*⁵, *Noemi Grigoletto De Biase*⁶

Transfer function of Brazilian Portuguese oral vowels: a comparative acoustic analysis

Palavras-chave: acústica da fala, espectrografia do som, fonética, voz.
Keywords: speech acoustics, sound spectrography, phonetics, voice.

Resumo / Summary

O trato vocal transfere suas características ao som produzido na glote, de acordo com sua configuração tridimensional. **Objetivo:** Determinar qual das sete vogais orais do Português brasileiro sofre a menor interferência acústica das modificações do trato vocal. **Material e método:** Estudo transversal prospectivo. Os indivíduos foram 23 homens e 23 mulheres, na faixa etária entre 20 e 45 anos (médias de 28,95 e 29,79 respectivamente), sem queixas vocais e com qualidade vocal normal na avaliação perceptivo-auditiva. 322 emissões vocais sustentadas foram digitalizadas e analisadas acusticamente utilizando-se a combinação de três programas computadorizados. Os resultados foram comparados à distribuição das frequências de ressonância de um tubo reto fechado em uma das extremidades. **Resultados:** A análise estatística mostrou que a vogal /ε/ apresentou diferença significativa em comparação com as demais vogais, com valores harmônicos médios mais altos e menores desvios-padrão para ambos os sexos. **Conclusão:** Para o Português brasileiro, a vogal /ε/ sofre menor influência das modificações do trato vocal e apresenta atenuação significativamente menor para ambos os sexos. A inclusão dessa vogal nos protocolos padronizados de avaliação vocal pode contribuir para a qualidade da informação obtida por meio das análises espectrográficas e acústica quantitativa.

The vocal tract transfers its characteristics onto the sounds produced at the glottis, depending on its tridimensional configuration. **Aim:** this study aims to determine which of the seven oral vowels in Brazilian Portuguese is acoustically less impacted by changes to the vocal tract. **Materials and method:** this is a cross-sectional prospective study. Twenty-three males and 23 females with ages ranging between 20 and 45 years (mean values of 28.95 and 29.79 years respectively) were enrolled in the study; none had voice complaints and their voices were normal under perceptive-auditory evaluation. Three-hundred and twenty-two sustained vocal emissions were digitized and acoustically analyzed by three computer programs combined. Results were compared against the distribution of resonance frequencies in a straight tube with one end sealed. **Results:** statistical analysis showed that vowel /ε/ was significantly different when compared to the other vowels, with higher mean harmonic values and lower standard deviation for both genders. **Conclusion:** in Brazilian Portuguese, vowel /ε/ is less impacted by changes to the vocal tract and is significantly less attenuated in both genders. The inclusion of this vowel in voice assessment standard protocols may contribute to improve the quality of the information obtained as a result of quantitative spectrographic and acoustic tests.

¹ Pós-doutora, Professora Adjunto do Departamento de Fonoaudiologia da Universidade Federal de São Paulo, São Paulo, SP, Brasil. Chefe do Serviço Integrado de Fonoaudiologia do Hospital São Paulo, São Paulo, SP, Brasil. Pesquisadora Associada do Instituto da Laringe - INLAR, São Paulo, SP, Brasil.

² Professor Titular do Departamento de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço da Universidade Federal de São Paulo, São Paulo, SP, Brasil. Diretor do Instituto da Laringe - INLAR, São Paulo, SP, Brasil.

³ Fonoaudióloga; Mestre em Ciências pela Universidade Federal de São Paulo; Especialista em Voz.

⁴ Médico Otorrinolaringologista do Instituto da Laringe - INLAR, São Paulo, SP, Brasil.

⁵ Fonoaudióloga. Doutora em Ciências pela Universidade Federal de São Paulo, SP, Brasil. Professora Instrutora do Departamento de Morfologia da Faculdade de Ciências Médicas da Santa Casa de São Paulo, SP, Brasil.

⁶ Pós-doutora, Professora Associada do Departamento de Fundamentos da Faculdade de Fonoaudiologia da Pontifícia Universidade Católica, São Paulo, SP, Brasil. Orientadora de Pós-Graduação do Departamento de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço da Universidade Federal de São Paulo, São Paulo, SP, Brasil. Instituto da Laringe.

Endereço para correspondência: Noemi Grigoletto De Biase - Rua Madre Rita Amada de Jesus 106 04721-050 São Paulo SP Brasil.

Tel.: (+5511) 5683.2903 - Fax: (+5511) 5683.2903 - E-mail: ngdebiase@gmail.com

Este artigo foi submetido no SGP (Sistema de Gestão de Publicações) da BJORL em 18 de março de 2008. cod. 5770

Artigo aceito em 4 de setembro de 2006.

INTRODUÇÃO

A geometria do trato vocal adulto assemelha-se a um tubo reto fechado em uma extremidade com comprimento médio de 17 cm, cujos três primeiros picos de ressonância estão ao redor de 500, 1500 e 2500Hz¹⁻³.

O trato vocal transfere suas características ao som produzido na glote, de acordo com sua configuração tridimensional, a qual é resultado do posicionamento de suas estruturas. Essa transferência, também denominada fator de transferência, modificará a intensidade dos harmônicos, consequência do fenômeno da ressonância. As frequências dos harmônicos que coincidem com as frequências de ressonância do trato vocal sofrerão modificações menores e serão denominadas formantes, enquanto as demais terão suas intensidades reduzidas (ou não amplificadas)^{2,4}. Tais formantes variam de acordo com a geometria tridimensional do trato vocal, sendo os três primeiros fundamentais para a identidade acústica das vogais.

Protocolos padronizados de avaliação acústica vocal preferencialmente utilizam as vogais /a/, /i/ e /u/ por serem acusticamente muito diferentes entre si, o que favoreceria a observação de possíveis distúrbios nas diferentes regiões do espectro, de acordo com as correspondentes configurações do trato vocal.

A observação da configuração do trato vocal durante a emissão das vogais por meio de videofluoroscopia mostra secção transversal do tubo mais uniforme durante a fonação da vogal /ε/. Também a observação na prática clínica da espectrografia das vogais orais do Português brasileiro mostra, em geral, harmônicos com maior intensidade na emissão da vogal /ε/, o que sugere haver menor modificação na forma do trato vocal em comparação às demais vogais orais. Assim, a vogal /ε/ é a que mais se assemelha, em configuração, a um tubo reto, o que nos leva a supor que também apresentaria maior intensidade e, conseqüentemente, melhor definição de seus harmônicos. Ainda considerando, visualmente, a configuração do trato vocal, a vogal /u/, por apresentar maior modificação, apresentaria mais obstáculos à passagem do som, com conseqüente redução da intensidade e pior definição dos harmônicos.

Desta forma, é importante a identificação objetiva da vogal que permite a relação mais acurada entre os achados visuais das imagens laringoscópicas e os dados acústicos obtidos no laboratório de voz.

Portanto, o objetivo deste estudo foi comparar a identidade acústica das sete vogais orais do Português brasileiro e determinar qual delas sofre a menor influência das modificações do trato vocal quando comparada a um tubo reto fechado em uma das extremidades.

MATERIAL E MÉTODO

Foram gravadas as sete vogais orais do Português brasileiro (/a/, /ε/, /e/, /i/, /É/, /o/ e /u/) de 23 homens e 23 mulheres, entre 20 e 45 anos de idade (valores médios: 28,95 e 29,79 anos, respectivamente). Assim, cada um dos indivíduos gravou a emissão de cada vogal, totalizando 322 gravações.

Um grupo com experiência em voz, composto por dois fonoaudiólogos e dois otorrinolaringologistas, selecionou os indivíduos para o estudo; os critérios de inclusão foram ausência de queixas vocais e qualidade vocal normal na avaliação perceptivo-auditiva.

Foi solicitado aos indivíduos que emitissem cada vogal de modo sustentado e em sua frequência e intensidade habituais e a gravação da voz foi realizada em cabina acústica, mantendo-se uma distância de 15 cm entre os lábios e o microfone (unidirecional SHURE SM58). A digitalização foi realizada por meio do programa SoundScope, da Macintosh. Uma vez registradas, as 322 ondas sonoras foram equalizadas na amplitude de 4 volts. As amostras foram padronizadas em segmentos de 5 segundos na porção intermediária da onda sonora.

As ondas foram então convertidas para o formato wave (.wav) utilizando-se o programa Wave Converter 1.5. Os dados quantitativos foram obtidos pela combinação de três programas computadorizados: Vocal Assessment, Screen Size Capture e Carnoy 2.0, como descrito a seguir:

Passo 1. Vocal Assessment - utilizado para selecionar o segmento intermediário de 5 segundos para análise e para separar os harmônicos do ruído, restando apenas picos de harmônicos.

Passo 2. Carnoy 2.0 - utilizado para obter a quantificação dos picos dos harmônicos dos arquivos de imagem a partir do gráfico gerado sem ruído.

Todos os valores da intensidade de cada harmônico foram registrados para cada amostra. Os valores médios e desvios-padrão foram calculados para permitir a comparação intra e inter indivíduos das intensidades dos harmônicos para cada vogal.

No estudo da frequência dos harmônicos primeiramente foi realizado o cálculo computadorizado do valor da f0 para cada vogal de cada indivíduo, extraída pelo programa SoundScope; a seguir, multiplicamos esse valor pelo número do harmônico de maior intensidade na região correspondente aos três primeiros formantes de cada vogal, para determinar a frequência de cada formante. Por exemplo, se a f0 fosse 100 Hz e o quinto harmônico fosse o de maior intensidade na região do primeiro formante, então o valor deste seria 500 Hz. Posteriormente, obteve-se a média dos valores da frequência

dos três harmônicos de maior intensidade (Peterson, 1959), tanto para o grupo masculino quanto para o feminino. Estas médias, representando as regiões dos três primeiros formantes, foram comparadas à distribuição de ressonância de um tubo reto e fechado em uma das extremidades, no qual a frequência dos demais formantes são múltiplos inteiros e ímpares do primeiro. No caso específico de um tubo com comprimento ao redor de 17 cm estas frequências correspondem a aproximadamente 500, 1500 e 2500 Hz. Desta forma, consideraremos a vogal cujo trato vocal assume a forma mais próxima possível da tubular e que menos modificações causaria no som glótico, isto é, aquela cujas médias estão mais próximas destes valores. Esta vogal será destacada mediante escrita em negrito na tabela referente aos valores das frequências dos harmônicos.

Considerando-se a natureza das variáveis estudadas, foi realizado o Teste de Friedman (teste não-paramétrico) para comparar se estes postos médios eram estatisticamente diferentes ou não, adotando-se o nível de significância $p < 0,001$. Como houve significância, então pelo menos um dos sete desvios era diferente do restante. Assim, aplicamos também o Teste t de Student para Dados Pareados (teste paramétrico), para determinação dos pares com diferença significativa, e o Teste dos Postos Sinalizados de Wilcoxon (teste não-paramétrico), para a comparação da distribuição classificatória (Rosner, 1986). Foi adotado o nível de significância de 0,050 (5%) e foram assinalados, com um asterisco, os valores estatisticamente significantes.

RESULTADOS

Os resultados são apresentados nas Tabelas 1 a 3. A Tabela 1 apresenta, para cada vogal, os valores das médias das intensidades dos harmônicos com seus respectivos desvios-padrão, obtidas para cada sexo. A Tabela

Tabela 1. Médias dos valores das intensidades dos harmônicos (em db) e respectivos desvios-padrão, para cada vogal, para ambos os sexos.

Vogal	MULHERES		HOMENS	
	X	DP	X	DP
/a/	42,92	9,48	36,91	11,02
/ε/	45,04	6,88	39,37	9,27
/e/	43,88	8,32	38,85	10,3
/i/	41,49	10,7	36,73	10,97
/ɔ/	39,94	11,56	36,33	11,63
/o/	36,5	13,25	33,84	12,91
/u/	35,29	13,56	32,78	13,02

2 apresenta os dados estatísticos para a comparação entre as intensidades dos harmônicos, par a par. A Tabela 3 apresenta os valores das frequências dos três primeiros formantes, para cada vogal e sexo.

DISCUSSÃO

Foram estudadas apenas as vogais orais do Português brasileiro, por duas razões: a primeira refere-se ao fato de que, para a análise vocal e durante a avaliação laringológica, geralmente não são emitidas vogais nasais, a não ser em casos para estudo do pórtico velofaríngeo. Na avaliação vocal clínica fonoaudiológica são analisadas mais comumente as vogais orais /a/, /i/ e /u/; na avaliação laringológica, geralmente utilizam-se as vogais /a/ ou /ε/ (para o registro modal) e /i/ (este para avaliação da emissão aguda ou em falsete), embora não haja consenso entre os profissionais. A segunda razão diz respeito à análise espectrográfica das vogais nasais, que apresenta leitura mais difícil pela presença de formantes adicionais com intensidade reduzida; além disso, as nasais geram maior número de erros de identificação em relação às orais⁵.

Foi realizada a equalização da entrada de energia das amostras ao redor de 4 volts para padronizar os níveis de saída da intensidade, uma vez que os indivíduos foram solicitados a emitir as vogais em frequência e intensidade habituais e confortáveis.

A vogal /e/ apresentou valores médios da intensidade dos harmônicos maiores e com menor desvio-padrão (Tabela 1). Com estudo pareado das médias (Tabela 2), observamos que a vogal /ε/ diferiu de modo significativo das demais vogais, com exceção da vogal /e/, para ambos os sexos. O estudo pareado dos desvios-padrão também, por sua vez, mostrou diferença estatisticamente significativa entre a vogal /ε/ e todas as outras vogais. Assim, podemos afirmar que, além de apresentar a maior média e o menor desvio-padrão das intensidades dos harmônicos, a vogal /ε/ apresentou diferença significativa em relação às demais vogais (com exceção em relação à média da vogal /e/), para ambos os sexos. Essa semelhança entre as vogais /ε/ e /e/ talvez se deva ao fato de que ambas apresentam configurações do trato vocal semelhantes. Ainda assim, a vogal /ε/ possui um trato vocal com a língua em posição mais baixa do que na emissão da vogal /e/, ou seja, com menor constrição e, conseqüentemente, menos obstáculos à passagem da energia sonora. Desta forma, embora a diferença entre os valores das duas vogais não tenha sido estatisticamente significativa, estes foram maiores, em ambos os sexos, na emissão da vogal /ε/ e o valor do p (0,052) mostra tendência à significância.

Maiores médias e menores valores de desvio-padrão estão relacionados à menor atenuação do trato

Tabela 2. Valores de p no estudo pareado das médias e respectivos desvios-padrão das intensidades dos harmônicos das vogais, para ambos os sexos.

Pares das Vogais	MULHERES		HOMENS	
	p X	DP	p X	DP
/a/ x /ε/	0,005*	0,001*	0,007*	0,001*
/a/ x /e/	0,139	0,010*	0,013*	0,062
/a/ x /i/	0,148	0,031*	0,832	0,913
/a/ x /ɔ/	0,003*	0,001*	0,519	0,076
/a/ x /o/	0,001*	0,001*	0,001*	0,001*
/a/ x /u/	0,001*	0,001*	0,001*	0,001*
/ε/ x /e/	0,052	0,001*	0,349	0,007*
/ε/ x /i/	0,001*	0,001*	0,001*	0,001*
/ε/ x /ɔ/	0,001*	0,001*	0,001*	0,001*
/ε/ x /o/	0,001*	0,001*	0,001*	0,001*
/ε/ x /u/	0,001*	0,001*	0,001*	0,001*
/e/ x /i/	0,015*	0,001*	0,011*	0,085
/e/ x /ɔ/	0,001*	0,001*	0,006*	0,001*
/e/ x /o/	0,001*	0,001*	0,001*	0,001*
/e/ x /u/	0,001*	0,001*	0,001*	0,001*
/i/ x /ɔ/	0,170	0,064	0,614	0,093
/i/ x /o/	0,001*	0,001*	0,003*	0,001*
/i/ x /u/	0,001*	0,001*	0,001*	0,001*
/ɔ/ x /o/	0,001*	0,001*	0,004*	0,001*
/ɔ/ x /u/	0,001*	0,001*	0,001*	0,001*
/o/ x /u/	0,168	0,266	0,194	0,547

Teste t de Student para Dados Pareados. Significância estatística marcada com asterisco, para $p < 0,050$.

Tabela 3. Médias dos valores das frequências dos harmônicos correspondentes aos três primeiros formantes (F1, F2, F3), em Hz, para cada vogal, para ambos os sexos.

VOGAIS	MULHERES			HOMENS		
	F1	F2	F3	F1	F2	F3
/a/	1002,90	1549,95	2959,70	753,87	1278,70	2483,44
/ε/	672,45	2242,93	3018,60	588,44	1745,11	2566,00
/e/	437,03	2429,76	3087,09	406,63	1955,60	2540,33
/i/	361,90	2583,89	3378,14	297,80	2150,85	2925,14
/ɔ/	715,34	1073,27	2981,69	580,15	947,25	2525,52
/o/	444,89	914,26	2899,80	411,62	832,84	2376,13
/u/	461,82	763,41	2902,55	345,27	799,51	2351,50

vocal^{4,6,7}. Tal fato nos permite considerar que, ao passar pelo trato vocal, a posição dos ressoadores na emissão da vogal /ε/ permite a passagem da maior quantidade de energia, com menor interferência sobre o som produzido pela vibração das pregas vocais, indicando um fator de transferência cujas frequências coincidem com as frequências

de ressonância desse trato, sem deformações. Mais especificamente, indicam que o posicionamento das estruturas permite um fluxo quase totalmente livre, através do trato vocal, o que observamos na vogal /ε/. Devemos lembrar, no entanto, que a amplitude de uma onda é, essencialmente, a medida de seu tamanho, e independe

da frequência da onda. Sendo assim, a energia sonora está presente em cada harmônico, mas a amplitude de cada harmônico sofrerá a influência da função do filtro e da amplitude do harmônico específico da fonte glótica⁴.

Como já afirmava Hermann⁸, as frequências dos formantes são, na realidade, aquelas nas quais o filtro supralaríngeo permite a passagem da maior quantidade de energia. Isto ocorre devido à mínima interferência do trato vocal no som produzido pela vibração das pregas vocais, uma vez que as modificações do trato tendem a reduzir a informação acústica do som gerado na laringe. Assim, a vogal /ε/, no Português brasileiro, é produzida com menor atenuação dos harmônicos, representando melhor o espectro da fonte glótica. Nossos achados concordam com outros autores⁹ que afirmam que, durante a articulação da vogal /ε/ o trato vocal age acusticamente com um tubo fechado numa extremidade, baseado nos achados da reconstrução tridimensional do trato vocal por meio de ressonância magnética.

Com relação às frequências dos três primeiros formantes (Tabela 3), observamos que a vogal /ε/ apresentou valores mais aproximados aos da função de transferência de uma vogal neutra, ou seja, 500, 1500 e 2500 Hz do que as demais vogais, para o sexo masculino. O mesmo aconteceu em relação ao sexo feminino, mas com uma maior margem de diferença, uma vez que, por apresentarem f0 mais aguda e, conseqüentemente harmônicos mais espaçados e menor número deles no espectro, tornou-se mais difícil sua mensuração. Mas, de qualquer modo, cabe aqui ressaltar que, mesmo com valores mais distantes, a proporção entre os três formantes manteve-se mais próxima também para a vogal /ε/ nesse sexo.

No entanto, outros autores, estudando outras línguas, concluíram que a vogal /æ/ é a mais aproximada à função de transferência do trato vocal, ou seja, com frequências dos formantes F1, F2 e F3 próximas a 500, 1500 e 2500Hz¹⁰. Outros estudos descrevem a vogal schwa do Inglês (/ε/) como a que mais se aproxima da função de transferência, denominada vogal neutra¹¹⁻¹⁵. Assim, podemos sugerir a inclusão da vogal /ε/ nos protocolos padronizados de avaliação vocal no Brasil.

CONCLUSÃO

De acordo com este estudo concluímos que, observando as características acústicas de cada vogal do Português brasileiro, podemos afirmar que a distribuição dos três primeiros formantes da vogal /ε/ é a mais próxima às frequências de ressonância de um tubo reto fechado em uma das extremidades e, conseqüentemente, à vogal neutra, sofrendo menor influência das modificações do trato vocal para ambos os sexos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Kent RD. Vocal tract acoustics. *J Voice*. 1993(7)2:97-117.
2. Sundberg J. The science of the singing voice. DeKalb: Northern Illinois University Press; 1987. 216p.
3. Fant G. Speech sounds and features. Cambridge: The MIT Press; 1973. 227p.
4. Lieberman P. Speech physiology and acoustic phonetics: an introduction. New York: Macmillan Publishing; 1977. 206p.
5. Behlau M, Pontes P, Tosi O, Ganaça M. Análise espectrográfica de formantes das vogais do português brasileiro falado em São Paulo. *Acta AWHO*. 1988;7:74-85.
6. Stevens KN, House AS. An acoustical theory of vowel production and some of its implications. *J Speech Hear Res*. 1961;4(4):303-20.
7. Hiraoka N, Kitazoe Y, Ueta H, Tanaka S, Tanabe M. Harmonic-intensity analysis of normal and hoarse voices. *J Acoust Soc Am*. 1984b;76(6):1648-51.
8. Hermann (1894) Sulter AM, Miller DG, Wolf RF, Schutte HK, Wit HP, Mooyaart EL. On the relation between the dimensions and resonance characteristics of the vocal tract: a study with MRI. *Magn Reson Imaging*. 1992 10(3):365-73.
9. Jakobson R, Fant CGM, Halle M. Preliminaries to speech analysis: the distinctive features and their correlates. Cambridge, The MIT press, 1972, 64p.
10. Fant G. Acoustic theory of speech production. The Hague: Mouton; 1960.
11. Borden GJ, Harris KS. Speech science primer: physiology, acoustics, and perception of speech. Baltimore: The Williams & Wilkins Company; 1980. 297p.
12. Johnson K. Acoustic and auditory Phonetics. Oxford. Blackwell, 1997. 169p.
13. Lieberman P, Blumstein SE. Speech physiology, speech perception, and acoustic phonetics. Cambridge: Cambridge University Press; 1988. 249p.
14. Speaks CE. Introduction to sound. Acoustics for the hearing and speech sciences. San Diego: Singular Publishing Group; 1992. p.163-90.