

QUALIDADE VOCAL DE INDIVÍDUOS SUBMETIDOS A LARINGECTOMIA TOTAL: ASPECTOS ACÚSTICOS DE CURTO E DE LONGO TERMO EM MODALIDADES DE FONAÇÃO ESOFÁGICA E TRAQUEOESOFÁGICA

Voice quality from total laryngectomee subjects: short and long-term acoustic parameters in esophageal and tracheoesophageal phonation

Mariana Silva Jorge ⁽¹⁾, Fabiana Nogueira Gregio ⁽²⁾, Zuleica Camargo ⁽³⁾

RESUMO

Objetivo: investigar os parâmetros acústicos de curto e de longo termo da qualidade vocal em modalidades de fonação esofágica e traqueoesofágica. **Métodos:** gravação de amostras de fala (vogais sustentadas, entrecortadas, sentenças, texto padronizado e emissão espontânea) de dois indivíduos laringectomizados totais (72 e 71 anos), um com fonação esofágica e um com fonação traqueoesofágica, e um indivíduo referência (76 anos), todos do sexo masculino. As amostras de fala foram submetidas a modalidades de análise acústica de curto termo (frequência fundamental, frequência dos três primeiros formantes e tempo máximo de fonação) e de longo termo (espectros de longo termo, frequência e intensidade dos picos espectrais). **Resultados:** a análise acústica de curto termo evidenciou frequência fundamental mais grave e valores de formantes mais agudos para os laringectomizados totais. Verificou-se que a prótese traqueoesofágica propiciou resultados mais próximos ao indivíduo referência em termos de intensidade, frequência fundamental e tempo máximo de fonação, os quais são compatíveis com aspectos da fisiologia referentes aos fatores aerodinâmicos da fonação. Na análise de longo termo, os falantes laringectomizados totais revelaram menor amplitude do sinal de fala e escassa definição de picos espectrais em relação ao indivíduo referência. **Conclusões:** da análise acústica de curto e de longo termo foi possível comparar as modalidades de fonação sem laringe entre si e prever localização da atividade esfíntérica, além de supor aspectos de conformação das cavidades ressoadoras.

DESCRITORES: Qualidade da voz; Voz esofágica; Voz alaríngea; Acústica da fala; Laringectomia

■ INTRODUÇÃO

O ser humano se integra ao meio em que vive graças à condição de comunicação oral, possibilitando que se expresse de forma mais concreta graças às

inflexões, às modulações e às variações da qualidade vocal. Quando ocorre a privação da voz laríngea em decorrência do tratamento para o câncer, o sujeito passa por uma série de mudanças na vida familiar, social e profissional.

Segundo o Ministério da Saúde, o câncer figura como a segunda causa de morte no Brasil, perdendo apenas para as doenças cardiovasculares. Sendo assim, configura-se como um problema de saúde pública que tende a se tornar uma das principais causas de morte no país. O aumento da incidência de câncer na população brasileira é atribuído a diversos fatores, como a maior expectativa de vida da população, e também à maior exposição a fatores cancerígenos, tais como agentes infecciosos, álcool, má nutrição, tabaco e poluição ¹.

Nesse universo, o câncer de laringe é um dos

⁽¹⁾ Fonoaudióloga, Especializanda em Voz.

⁽²⁾ Fonoaudióloga, Mestranda em Lingüística Aplicada e Estudos da Linguagem – PUC-SP, Pesquisadora do Laboratório Integrado de Análise Acústica e Cognição- LIAAC- PUC-SP

⁽³⁾ Fonoaudióloga, Doutora em Lingüística Aplicada e Estudos da Linguagem – PUC-SP, Professora assistente-doutora do Departamento de Lingüística da Faculdade de Comunicação e Filosofia da PUC-SP, Pesquisadora do Laboratório Integrado de Análise Acústica e Cognição- LIAAC- PUC-SP.

mais freqüentes a atingir a região de cabeça e pescoço e compreende 1 a 2% de todos os tumores malignos que acometem a população mundial. É mais encontrado em homens na faixa etária compreendida entre 50 e 60 anos. O tabagismo e o consumo excessivo de álcool são os principais fatores de risco. Além desses, foram descritas a exposição a metais, componentes da indústria têxtil e a hereditariedade ².

Além do local e extensão do tumor, as seqüelas decorrentes dos tratamentos do câncer dependem de alguns fatores, tais como: extensão da ressecção, grau de mobilidade das estruturas remanescentes, tipo de reconstrução, motivação do paciente, envolvimento da família e, também, da experiência dos profissionais da equipe.

Quando o câncer não é diagnosticado precocemente e uma cirurgia funcionalmente conservadora não é possível, a laringectomia total torna-se a opção cirúrgica ³. Tal procedimento causa importante mudança na região cervical ao desconectar via aérea de via digestiva, de forma que a respiração é processada pelo traqueostoma definitivo.

De maneira geral, o fonoaudiólogo atuante na área de cirurgia de cabeça e pescoço realiza um trabalho de reabilitação focado no estímulo ao desenvolvimento de mecanismos compensatórios para a superação das alterações impostas pela doença, os quais podem se manifestar em funções vegetativas (respiração e deglutição) e na comunicação oral ⁴.

A voz esofágica, referida na literatura anteriormente à primeira laringectomia total realizada, promove qualidade vocal aperiódica, uma vez que o segmento faringo-esofágico e as paredes do esôfago e da hipofaringe não possuem mucosa organizada em camadas como as pregas vocais. Além disso, a voz é caracterizada por instabilidade, ruído, freqüência de vibração aproximada de 80 Hz e variações de freqüência e intensidade restritas ⁵. A freqüência fundamental da voz esofágica é a metade da média encontrada no homem adulto ⁶.

O reservatório de ar para a voz esofágica é de apenas 80 ml, enquanto que para falantes laríngeos saudáveis é de 2200 a 46000 ml. Esse fato implica o máximo de três segundos de duração de emissão para bons falantes esofágicos ⁵.

A intensidade da voz esofágica é situada em aproximadamente 5 a 8 dB abaixo da voz laríngea, enquanto o potencial máximo de variação da intensidade da voz esofágica é de 20 dB. A voz apresenta *pitch* grave e *loudness* reduzida ⁴. O falante esofágico bem sucedido apresenta fala de baixa intensidade, o que é notado principalmente quando se está em ambiente ruidoso ⁷.

Os indivíduos submetidos a altas doses de radioterapia na região do pescoço podem apresentar dificuldades quanto à utilização da laringe eletrônica, uma vez que o enrijecimento dos tecidos atrapalha a

transferência da vibração da prótese para o pescoço ⁸.

Os indivíduos que utilizam a prótese traqueoesofágica tendem a produzir a voz com freqüência fundamental (f₀) mais próxima de falantes normais ⁹⁻¹¹. Não há significativa variação de f₀ entre a fala traqueoesofágica e a fala laríngea ¹² e, embora ocorra variação de f₀, é menor que a encontrada em falantes esofágicos ¹⁰. A freqüência fundamental média observada foi de 100 Hz em indivíduo com voz traqueoesofágica ¹³ e máxima de 127 Hz ¹⁴.

A análise acústica de longo termo envolve a observação dos espectros de longo termo, enquanto representação da distribuição de energia ao longo de uma faixa de freqüência, num período de 30 a 60 segundos ¹⁵. Os aspectos definidos em tal análise são recorrentes no tempo, ou seja, configuram-se como traços estáveis do falante e independem do conteúdo da mensagem falada, mais especificamente da composição no plano segmental. Dessa forma, utilizando o espectro de longo termo (ELT), é possível identificar aspectos da qualidade vocal do falante, a qual envolve não só as características físicas e estruturais do trato vocal, como controle neurológico, estado emocional e saúde do falante ¹⁶.

Não foram encontrados estudos que utilizam o procedimento da análise acústica de longo termo para investigar qualidade vocal de indivíduos submetidos a laringectomia total.

A análise de longo termo para vozes de idosos do gênero masculino e feminino revelou rebaixamento dos valores dos três primeiros picos espectrais, especialmente quando comparados às emissões dos adultos jovens. Para população masculina, acima de 60 anos, destacaram-se valores de freqüência do primeiro pico espectral em 442 Hz e do segundo em 1640 Hz. Os resultados estão de acordo com os dados anatômicos, sugerindo que o envelhecimento resulta no alongamento da porção supraglótica do trato vocal, mais acentuado em falantes do sexo feminino ¹⁷. O mesmo tipo de análise foi recentemente referido para um grupo de indivíduos portadores de disfonia ¹⁸.

Do ponto de vista acústico, as freqüências dos formantes são mais agudas em falantes esofágicos, devido ao encurtamento do trato vocal decorrente da laringectomia total ¹⁹⁻²⁰.

Em estudo acústico da freqüência fundamental e dos três primeiros formantes das vogais do português brasileiro da fonação esofágica, observou-se instabilidade no registro dos formantes ao longo do tempo. O primeiro formante foi o mais alterado pela laringectomia total, elevando-se em média 26% em relação ao indivíduo de voz normal. Os valores absolutos das freqüências dos três primeiros formantes das vogais orais emitidas apresentaram-se sistematicamente mais agudos que os indivíduos normais, o que reflete redução do comprimento do trato vocal. Para a vogal [a], a média do primeiro

formante foi de 954 Hz, do segundo de 1575 Hz e do terceiro de 2819 Hz. Os valores da frequência fundamental foram significativamente mais graves para a voz esofágica em comparação à voz laríngea, situando-se em torno de 72,51 Hz¹⁹.

Outros parâmetros acústicos do português brasileiro em fonação esofágica foram explorados do ponto de vista das medidas de duração de frases, vogais, consoantes e tempo de início de sonorização (VOT), com correlatos fisiológicos descritos por meios endoscópicos e videofluoroscópicos²¹. Estudo explorando a relação entre qualidade vocal esofágica e videofluoroscopia do esôfago, também em falantes do português brasileiro, revelou relações entre dimensão do segmento faringo-esofágico e qualidade vocal, bem como presença de barra cricofaríngea em maior frequência nos falantes considerados bons e moderados, destacando-se a relação entre configuração do segmento faringo-esofágico e qualidade vocal esofágica²².

A intensidade da fala traqueoesofágica é aparentemente menor do que os níveis produzidos pelos falantes laríngeos. A variação de intensidade pode ser maior do que para os falantes normais¹⁰. No entanto, quando comparado com falantes esofágicos, a fala traqueoesofágica apresenta valores de intensidade mais próximos dos encontrados para falantes laríngeos⁹. A intensidade habitual média encontrada foi de 66 dB; intensidade habitual máxima de 76 dB e variação de intensidade média de 21 dB¹³.

Estudos concluíram que os falantes traqueoesofágicos atingem resultados superiores com relação à qualidade vocal^{14,23}. Pela análise perceptivo-auditiva e acústica da produção vocal de indivíduos laringectomizados totais, pode-se verificar que falantes traqueoesofágicos apresentam qualidade vocal rouca, forte e grave. O falante esofágico, por sua vez, apresenta *loudness* reduzido²⁴.

Quanto ao tempo máximo de fonação (TMF), a fala traqueoesofágica apresenta maiores valores em relação à fala esofágica e está mais próxima do falante normal^{9,14,25}. Tal fato possibilita emissões mais longas, com adequada modulação vocal e entoação²⁶. Para o falante esofágico o valor máximo encontrado foi 3 segundos (s)²⁶.

Considerando as alterações sofridas no aparelho fonador em decorrência do câncer laríngeo e as possibilidades de comunicação oral dos indivíduos submetidos a laringectomia total, o presente estudo tem como objetivo investigar os parâmetros acústicos de curto e de longo termo da qualidade vocal de falantes laringectomizados totais, que se comunicam por meio da fonação esofágica e traqueoesofágica.

■ MÉTODOS

A coleta de amostras envolveu uma sessão de gravação da fala de dois indivíduos laringectomizados

totais, do sexo masculino, que utilizam fonação esofágica e traqueoesofágica como meio de comunicação oral, referidos respectivamente como C1 e C2. Como referência para as análises de ordem acústica foram adotadas as emissões de um indivíduo do mesmo gênero e faixa etária, sem alteração de aparelho fonador, referido como CR. Todas as amostras constam do banco de dados do Laboratório Integrado de Análise Acústica (LIAAC) – PUC/SP e foram coletadas seguindo os mesmos procedimentos e a mesma seqüência de estímulos.

C1, 72 anos, foi submetido a laringectomia total, com ressecção dos três primeiros anéis traqueais, esvaziamento cervical seletivo bilateral (níveis II, III e IV), tireoidectomia subtotal, esvaziamento recorrential bilateral e fechamento primário da faringe em dois planos submucosos. Foi submetido a cerca de 40 sessões de radioterapia. Desenvolveu a modalidade de comunicação oral por meio de fonação esofágica, como resultado de reabilitação fonoaudiológica.

C2, 71 anos, foi submetido a laringectomia total com esvaziamento cervical esquerdo (níveis II a V) e direito (níveis II a V), com remoção do músculo esternocleidomastoideo e veia jugular interna; miotomia do músculo cricofaríngeo; fechamento primário da faringe em dois planos submucosos mais um plano muscular até a base da língua e confecção primária de fístula traqueoesofágica. Foi submetido a 50 sessões de radioterapia. Desenvolveu a modalidade de comunicação oral por meio de fonação traqueoesofágica, com suporte de reabilitação fonoaudiológica.

Finalmente, CR, 76 anos, com ausência de alteração de aparelho fonador, foi adotado como parâmetro de referência por não ter sido submetido a laringectomia total e por não apresentar queixas vocais.

A gravação das amostras de fala foi realizada no Laboratório de Rádio da Faculdade de Comunicação e Filosofia (COMFIL) da Pontifícia Universidade Católica de São Paulo (PUC/SP), em ambiente tratado acusticamente, dispondo de recursos para gravação e digitalização do sinal de fala.

Os resultados apresentados são analisados segundo os referenciais da teoria acústica da produção da fala²⁷ e do modelo fonético de descrição de qualidade vocal²⁸.

Para a coleta dos dados, os falantes permaneceram em posição sentada a 10cm de distância do microfone da marca Sennheiser, modelo e835, que foi conectado à entrada de linha da mesa de som Mackie Microseries 1202 VLZ (12 canais). Os estímulos captados pelo microfone foram digitalizados com o auxílio da placa de som Delta 44 e processados pelo *software* Sound Forge 4.0.

Como estímulo para gravação das amostras de fala, destacou-se a emissão da vogal [a] em padrão

sustentado (duração máxima) e entrecortado; sentença-veículo com estrutura “Diga _____ baixinho”, que foi preenchida com vocábulos contendo sons plosivos surdos ([p], [t], [k]) e sonoros ([b], [d], [g]) do Português Brasileiro em posição inicial na estrutura [___ ata]; leitura de um texto, composta exclusivamente de segmentos sonoros; e um trecho de fala espontânea. A composição de texto exclusivo por segmentos sonoros segue tendência de estudos envolvendo a análise acústica de longo termo, em que o centro dos estudos repousa na emissão de sons sonoros dada a influência da sonoridade neste aspecto. O texto elaborado no Laboratório Integrado de Análise Acústica e Cognição (LIAAC)²⁹ da PUC-SP é o seguinte:

“Li uma alegoria de um jovem nobre, abordando a viagem de uma ave na Índia. A líder das aves é uma bela e animada arara. É a rainha e ninguém duvida da realeza dela. Reina, venerada na região do Vale Dourado. A arara exibiu os dons maravilhosos a ela doados e originados de um dom divino. Aos brados diz: “Vamos dominar o ódio e ganhar o reino do Deus amado”. Divulga às demais aves as verdades e misérias da vida e julga a maioria dos inimigos dos valores morais de maneira ardorosa. Diz às aves da árdua jornada aos vales do amor, da humildade e da nobreza e ganha aliados valorosos. Ela ainda verbaliza e já formas aladas voavam mirando o ideal de vida a elas divulgado. Baseadas nos novos valores, ajudaram uma as outras de verdade e ganharam o galardão real, banhado em ouro”.

O falante C1 realizou duas emissões da seqüência de sentenças-veículo e apenas uma leitura, uma vez que apresentou sinais de fadiga durante a gravação.

As amostras de fala foram digitalizadas na frequência de amostragem de 44100 Hz e 16 bits, formato wav e modalidade mono. Posteriormente, foram analisadas com o auxílio do *software Multispeech*. Tais amostras foram submetidas a procedimentos de análise acústica de curto e de longo termo.

Os trechos de leitura e fala espontânea com 60 segundos de duração foram eleitos para análise de longo termo. Os trechos de silêncio e pausas, incluindo ruídos do traqueostoma e de deglutição de ar (no caso do falante esofágico) foram removidos. Nessa etapa de análise, as amostras tiveram sua frequência de amostragem rebaixada para 22050 Hz.

Os espectros e medidas de longo termo do presente estudo foram extraídos de acordo com os seguintes parâmetros de análise: 256 pontos, sem suavização, aplicação de pré-ênfase de 0.098, janelamento *hanning*, intensidade mínima de 0dB e máxima de 80 dB e faixa de frequência de 0 a 11025 Hz.

O processamento dos dados por análise de longo termo gerou o espectro de longo termo, obtido pela análise de distribuição de energia em 128 pontos de frequência para cada uma das leituras dos fa-

lantes. A análise envolveu a inspeção dos traçados de longo termo, a definição de picos espectrais em frequência e suas respectivas variações de intensidade.

Tais dados foram estatisticamente tratados por meio de análise fatorial de componentes principais e análise de construção de *clusters* com auxílio dos programas SPSS 10.0 for Windows e SPAD 3.

Para a análise acústica de curto termo, foi eleita a inspeção do espectrograma de banda estreita e estimativas de frequência fundamental, bem como do espectrograma de banda larga e análise dos três primeiros formantes. Nesta etapa de análise as amostras também tiveram sua frequência de amostragem diminuída para 22050 Hz.

As emissões da vogal [a] em padrão sustentado, entrecortado e inserida na sílaba tônica do vocábulo “arara” da leitura e das sentenças-veículo foram selecionadas para extração dos valores de formantes e de frequência fundamental.

No âmbito da extração de frequência fundamental, os espectrogramas de banda estreita e os espectros por *FFT (Fast Fourier Transform)* foram gerados de acordo com os seguintes parâmetros: 1024 pontos, pré-ênfase de 0,098, janelamento *hanning*, intensidade mínima de 0 dB e máxima de 80 dB e faixa de frequência de 0 a 2000 Hz.

Para análise do padrão de formantes, os espectrogramas de banda larga e os espectros por *LPC (Linear Predictive Coding)* foram gerados de acordo com os seguintes parâmetros: 75 pontos, aplicação de pré-ênfase de 0,098, janelamento *hanning*, intensidade mínima de 0 dB e máxima de 80 dB e faixa de frequência de 0 a 4000 Hz.

Os valores obtidos pela análise de curto e longo termo foram comparados entre os indivíduos do grupo estudado (C1 e C2) e o indivíduo referência (CR).

Vale ressaltar que o procedimento em questão não foi invasivo e que os indivíduos participantes da pesquisa assinaram e receberam uma cópia de um termo de consentimento, de acordo com as normas vigentes em nosso meio para pesquisas envolvendo seres humanos.

O projeto desta pesquisa foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Fonoaudiologia da PUC/SP sob o número 221/02.

■ RESULTADOS

Os resultados são apresentados de acordo com as informações obtidas a partir da análise acústica de curto e de longo termo correspondente ao processamento das amostras de fala do falante esofágico (C1), traqueoesofágico (C2) e o sujeito referência (CR).

A inspeção geral dos traçados registrados revelou aperiodicidade da emissão sonora, adição de ruído, incluindo ruído decorrente de irregularidade vibratória,

além daqueles adventícios, provenientes de ruído de traqueostoma (respiração) e de deglutição de ar no caso do falante esofágico, bem como de oclusão de traqueostoma, no caso do falante com prótese traqueoesofágica.

Os dados referentes aos valores dos dois primeiros picos espectrais dos traçados de longo termo de C1, C2 e CR encontram-se na Tabela 1.

Tabela 1 - Valores de frequência (Hz) e intensidade (dB) dos dois primeiros picos espectrais dos traçados de longo termo para os estímulos de leitura de texto (LT) e de fala espontânea (FE) dos indivíduos do grupo estudado (C1 e C2) e do indivíduo referência (CR).

Sujeito	1º Pico Espectral				2º Pico Espectral			
	Frequência. (Hz)		Intensidade (dB)		Frequência (Hz)		Intensidade(dB)	
	LT	FE	LT	FE	LT	FE	LT	FE
C1	126	147	37	33	945	903	25	20
C2	186	252	47	43	1651	3402	25	27
CR	86	86	49	53	1357	1314	14	18

Na Tabela 2 são expostas as variações de intensidade (em dB) do primeiro e segundo picos espectrais das emissões de C1, C2 e CR, a qual revela a característica de declínio espectral.

Tabela 2 - Valores da variação de intensidade (em dB) entre os dois primeiros picos espectrais dos traçados de longo termo referentes aos estímulos de leitura de texto (LT) e de fala espontânea (FE).

Sujeito	Variação de Intensidade - dB	
	LT	FE
C1	12	12
C2	22	16
CR	35	36

Os dados de análise acústica de curto termo referentes aos valores de frequência fundamental encontram-se na tabela 3.

Tabela 3- Valores de frequência fundamental (em Hz) das emissões da vogal [a] em padrão sustentado [a:], entrecortado [aaa] e na sílaba tônica do vocábulo "arara" dos indivíduos laringectomizados totais (C1 e C2) e do indivíduo referência (CR).

Amostra	Frequência Fundamental (Hz)		
	C1	C2	CR
[a:]	55	121	*
[aaa]	55	99	110
[a ra ra]	60	71	93
Média	56,6	97	101,5

* amostra não disponível para extração dos valores

Nas Tabelas 4, 5 e 6 são apresentadas as respectivas médias de frequência (em Hz) dos três primeiros formantes da emissão dos laringectomizados totais (C1 e C2) e do indivíduo referência (CR).

Tabela 4 – Medidas das médias de frequência (em Hz) do primeiro formante (F1) da vogal [a] em padrão sustentado [a:], entrecortado [aaa] e em posição tônica nas emissões [__ata], completadas por vocábulos surdos e sonoros, inseridas na sentença "Diga _____baixinho".

Amostra	Frequência de F1		CR
	C1	C2	
[a:]	1055	659	*
[aaa]	1043	637	725
Pata	1043	913	648
Bata	967	945	615
Cata	1132	1044	610
Gata	1049	925	654
Tata	1077	934	703
Data	1187	941	643
Média	1069	875	657

*amostra não disponível para extração dos valores.

Tabela 5 – Medidas das médias de frequência (em Hz) do segundo formante (F2) da vogal [a] em padrão sustentado [a:], entrecortado [aaa] e em posição tônica nas emissões [__ata], completadas por vocábulos surdos e sonoros, inseridas na sentença "Diga _____baixinho".

Amostra	Frequência de F2		CR
	C1	C2	
[a:]	1648	1429	*
[aaa]	1747	1549	1352
Pata	1874	2059	1407
Bata	1868	1912	1297
Cata	1901	1967	1379
Gata	1880	1954	1418
Tata	2033	1895	1401
Data	1973	1788	1390
Média	1865,5	1819,25	1377,71

*amostra não disponível para extração dos valores.

Em relação aos valores de tempo máximo de fonação (TMF) da vogal [a] realizada pelos falantes laringectomizados totais, foram obtidos os seguintes resultados: 1,5 s para C1 e 11,13 s para C2. Devido à indisponibilidade da amostra foi adotado como referência o valor de 21 s de TMF descrito na literatura para falantes do sexo masculino do Português Brasileiro³⁰ para a fonação laríngea, representativos dos valores de CR.

Tabela 6 – Medidas das médias de freqüência (em Hz) do terceiro formante (F3) da vogal [a] em padrão sustentado [a:], entrecortado [aaa] e em posição tônica nas emissões [__ata], completadas por vocábulos surdos e sonoros, inseridas na sentença “Diga ____ baixinho”.

Amostra	Freqüência de F3 – Hz		CR
	C1	C2	
[a:]	2945	3176	*
[aaa]	2868	2934	2637
Pata	2786	3315	2709
Bata	2857	3154	2747
Cata	3038	3538	2467
Gata	2879	3601	2566
Tata	3044	2901	2692
Data	2885	3378	2632
Média	2912,75	3249,62	2635,71

*amostra não disponível para extração dos valores.

Nas Figuras 1, 2 e 3 são expostos os traçados de longo termo dos dois indivíduos do grupo estudado (C1 e C2) e do indivíduo referência (CR).

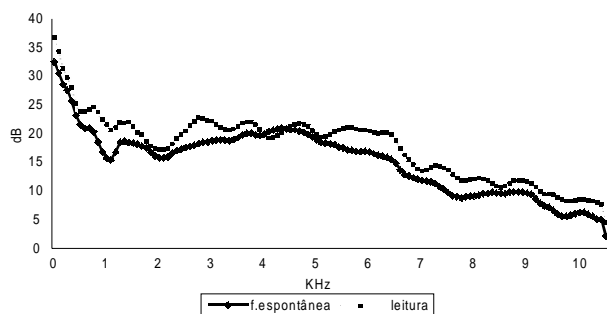


Figura 1- Espectros de longo termo de trecho de fala espontânea e de leitura, realizadas pelo falante esofágico (C1).

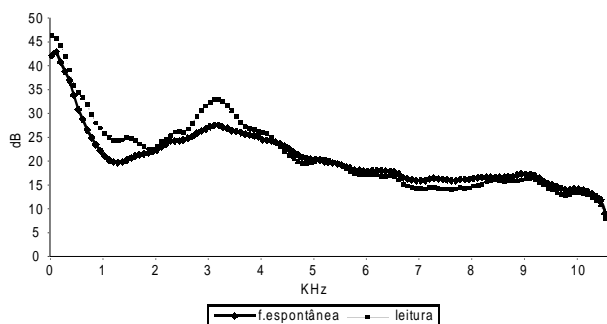


Figura 2- Espectros de longo termo de trecho de fala espontânea e de leitura, realizadas pelo falante traqueoesofágico (C2).

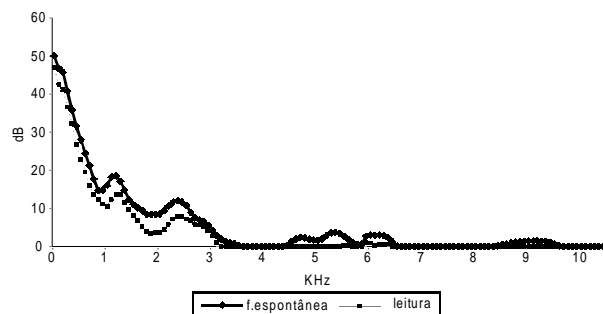


Figura 3- Espectros de longo termo de trecho de fala espontânea e de leitura, realizadas pelo indivíduo referência (CR).

Na Figura 4 encontra-se a média dos traçados de longo termo referentes às emissões dos indivíduos laringectomizados totais (C1 e C2) e do indivíduo referência (CR).

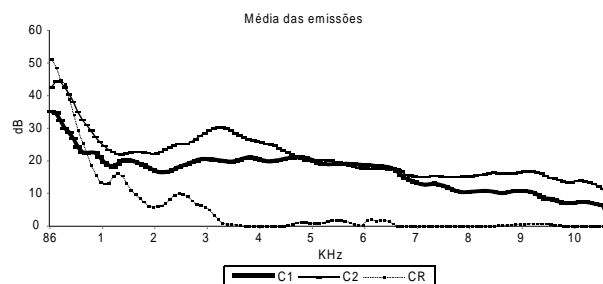


Figura 4- Média dos espectros de longo termo das emissões de cada um dos indivíduos do grupo estudado (C1 e C2) e do indivíduo referência (CR).

A análise fatorial dos dados obtidos revelou a formação de agrupamentos distintos para as informações referentes às várias emissões de um mesmo falante. Para cada falante, as duas emissões analisadas formaram um grupo distinto.

■ DISCUSSÃO

Os parâmetros de análise acústica de curto termo envolveram a inspeção de espectrogramas de banda estreita e espectro *FFT*, bem como espectrogramas de banda larga e espectro *LPC*. Nos espectrogramas de banda estreita foram detectados pulsos decorrentes da ação do segmento faringo-esofágico e da vibração de paredes do esôfago e hipofaringe, característico da fonte de voz alaríngea, à semelhança de estudo anterior¹¹.

As medidas de freqüência fundamental, extraídas pelos espectros *FFT*, (Tabela 3), revelaram que o falante esofágico apresenta f_0 mais grave do grupo estudado (média de 56,6 Hz), compatível com aumento de massa vibrante (tecidos faríngeos e esofágicos). Vários autores referiram resultados semelhantes^{4,6,19}.

O falante traqueoesofágico apresentou 97 Hz de valor médio de f_0 , valor aproximado do descrito na literatura^{13,26}. Os resultados de C2, que estão próximos aos do indivíduo referência, novamente enfatizaram a importância do fluxo aéreo para a fonação alaríngea⁹⁻¹¹, diferenciando-se a condição de fluxo contínuo de ar pela fonte vibratória no caso de C2, diferente da situação de C1, que não apresenta uso de fluxo pulmonar.

No caso dos espectrogramas de banda larga e dos espectros *LPC*, destacou-se a abordagem dos três primeiros formantes das emissões apresentadas nas tabelas 4, 5 e 6. Com relação ao primeiro formante (F1) observa-se maior estabilidade dos valores de C1 e CR em relação a C2, cuja variação de F1 foi maior nas diferentes emissões. C1 possui a média da frequência de F1 mais alta, seguida de C2 e CR. Tal aspecto pode ser atribuído à variação na posição de língua, em que se destaca possível ajuste de língua baixa e mandíbula aberta para C1 em contraposição a ajuste de língua alta e mandíbula fechada para C2, além de aspectos da remoção cirúrgica.

Para o segundo formante (F2), C1 revelou maior média do grupo, porém com mínima diferença em relação a C2. Provavelmente C1 mantém a língua em posição mais posterior que C2, fato confirmado pela medida do terceiro formante (F3) em faixa mais aguda para C2, compatível com a localização de constricção mais anterior em cavidade oral³¹.

Comparando-se os valores de formantes das emissões dos laringectomizados totais ao indivíduo referência, os falantes laringectomizados totais apresentaram aumento de todas as frequências dos formantes, denotando diminuição da extensão do trato vocal¹⁹. Tal apreciação pode revelar interessantes aspectos também relativos à extensão do trato vocal, em que a localização da nova fonte vibratória refere-se à extremidade distal. Este fator também pode ter influenciado as medidas de frequência de F1, uma vez que se refere à dimensão da cavidade posterior. Na comparação dos valores entre ambos os falantes laringectomizados totais pode-se observar aumento de F1 para C1, podendo-se levantar a hipótese de localização de fonte vibratória em plano superior a C2.

Em termos de duração das emissões, o tempo máximo da emissão da vogal [a] alcançado pelo falante esofágico foi 1,5 s, enquanto o valor máximo encontrado de TMF esofágica na literatura foi 3 segundos (s)²⁶. O falante traqueoesofágico apresentou 11,13 s de TMF, valor próximo ao referido na literatura de 12 s¹³. Considerando que a média do TMF para o indivíduo do gênero masculino é de 21s³⁰, nota-se que o valor encontrado para C2 está mais próximo do falante laríngeo do que o esofágico^{9,14,25}, novamente graças à possibilidade de retomada do uso de fluxo aéreo expiratório para fonação.

Quanto aos resultados da análise acústica, os traçados de longo termo (figuras 1, 2 e 3) dos falantes esofágico (C1), traqueoesofágico (C2) e do indivíduo referência (CR), foram detectadas reduções diferenciadas de intensidade para as provas em modalidades de leitura e de fala espontânea. Além disso, C1 apresenta menor amplitude do sinal e declínio espectral, o que pode ser justificado pelo fato de as paredes e mucosa do esôfago e da hipofaringe não apresentarem condições de realização de movimentos vibratórios refinados como as pregas vocais, além da fonação esofágica não utilizar fluxo aéreo pulmonar.

Os traçados de longo termo referentes às emissões a C1 revelaram maior proeminência do primeiro pico espectral (126 Hz para leitura e 147 Hz para fala espontânea). C2, por sua vez, apresentou o primeiro pico espectral em faixa de frequência mais aguda e com maior amplitude em relação a C1. Ao comparar o falante esofágico (C1) àquele que utiliza a fonação traqueoesofágica (C2), o último apresenta amplitude máxima do sinal em faixa próxima ao indivíduo referência (CR), o qual, por sua vez, revela a maior resolução espectral do grupo, com clara definição dos três picos, ao contrário dos falantes laringectomizados totais. Neste ponto, cabe ressaltar que a inclinação espectral presente em indivíduos com pregas vocais deve-se ao refinamento de atividade glótica. Na medida em que tais estruturas são removidas, as paredes do esôfago e, inclusive, de hipofaringe, passam a assumir o papel vibratório e, conseqüentemente, a definição espectral e a conformação dos picos se alteram de forma considerável.

Na figura 4, referente à média dos espectros de longo termo das emissões de C1, C2 e CR, pode-se notar mais claramente a diferença dos espectros de longo termo de tais falantes. O falante esofágico apresenta a intensidade mais baixa do grupo, com escassa definição de picos espectrais. C2 apresenta valor de intensidade mais elevado que C1 e se aproxima do falante laríngeo, que teve a emissão de maior intensidade. A diferença entre C2 e CR reside no fato de que o último tem melhor definição dos picos espectrais e maior intensidade do sinal de fala. Tal observação revela a efetividade das compensações realizadas por C2 na tentativa de superar a impossibilidade da fonação alaríngea e acentua a importância do uso de fluxo aéreo pulmonar expiratório na atividade fonatória.

A análise fatorial das emissões de C1, C2 e CR reforça os aspectos acima indicados ao distinguir três categorias distintas, o que mostra que tais indivíduos realizam diferentes ajustes durante a fonação, porém guardando a característica de individualidade dos ajustes referentes à qualidade vocal, o que volta a reforçar os aspectos de refinamento da emissão,

inclusive na ausência de porção do aparelho fonador, como é o caso da laringectomia total.

Os valores em intensidade dos dois primeiros picos espectrais (tabela 1), que correspondem ao nível de pressão sonora do sinal de fala gerado pelos falantes, revelaram que o falante esofágico possui o menor nível de intensidade geral do sinal no grupo, o que é compatível com referências a aspectos perceptivo-auditivos de *loudness* fraco para a fonação esofágica^{4,7,13,24}.

No presente estudo, pode-se verificar que a intensidade da fala traqueoesofágica é superior à voz esofágica e mais próxima da fala laríngea (CR). Tal resultado é semelhante ao encontrado por outros autores^{9,13-14,23}. Os valores da variação de intensidade entre os dois primeiros picos espectrais (tabela 2), são indicativos do declínio espectral. Nota-se maior declínio e nível de intensidade geral do sinal para C2 em relação a C1, provavelmente atribuído ao fato de apresentar uso de fluxo aéreo pulmonar.

Para a população masculina sem alteração laríngea, acima de 60 anos, destacam-se valores de frequência do primeiro pico espectral em 442 Hz e do segundo em 1640 Hz¹⁷. Contrapondo tais dados aos encontrados no presente estudo verifica-se que os valores de C1 estão mais próximos de CR. O deslocamento dos picos espectrais de C1 para faixas de frequência mais graves em relação a C2 indica possíveis aspectos de alongamento do aparelho fonador¹⁷. Tratando-se de falantes sem laringe, um aspecto a interferir em tal fenômeno seria a localização de fonte vibratória em porção mais baixa do trato vocal remanescente para C1.

Tais observações de longo termo da emissão revelaram importantes aspectos acústicos a colaborarem para a compreensão da qualidade vocal gerada e necessitam ser explorados em maior proporção na população em questão, como forma de colaborar para a reabilitação de tais indivíduos.

A análise acústica em modalidades de curto e longo termo para falantes laringectomizados totais possibilitou a reflexão a respeito dos mecanismos subjacentes às diferentes modalidades de comunicação (esofágica e traqueoesofágica), os quais podem influenciar o processo terapêutico dos

indivíduos em questão. Tais aspectos foram indicados pela definição dos traçados e picos espectrais na análise acústica de longo termo e de medidas acústicas de curto termo de frequência fundamental e formantes.

Para além da mera classificação dos falantes em termos de eficiência ou medidas pautadas na normalidade, tal concepção de análise permitiu a abordagem do detalhamento da produção sonora em situação de mutilação do aparelho fonador, como é o caso da laringectomia total. Como possibilidades futuras, sugere-se incorporação de componentes de descrição fisiológica e detalhamento de parâmetros acústicos de longo termo em maior número de falantes de ambas as modalidades de comunicação oral. Tal refinamento de análise poderá revelar detalhes importantes ao planejamento terapêutico e à compreensão dos refinados mecanismos envolvidos na produção de fala pelos indivíduos laringectomizados totais.

■ CONCLUSÕES

Do presente estudo pode-se concluir que houve correspondência entre os resultados da análise acústica de curto e longo termo. Foi possível comparar modalidade que não envolve fluxo de ar pulmonar (fonação esofágica) e que o envolve (prótese traqueoesofágica), além de compará-las à situação de fonação laríngea. Os resultados possibilitaram, inclusive, prever a localização da atividade esfíntérica, a ser posteriormente complementada por avaliação fisiológica, e supor aspectos de conformação das cavidades ressoadoras e, conseqüentemente, de posicionamento dos articuladores.

Na análise de longo termo, verificou-se que os falantes laringectomizados totais não apresentaram boa definição dos picos espectrais e apresentaram menor amplitude do sinal de fala, caracterizando emissões com menor nível de intensidade em relação ao indivíduo referência.

Na análise de curto termo, os falantes laringectomizados totais apresentaram emissões de frequência mais graves que o indivíduo referência, bem como valores de frequência de formantes mais agudos.

ABSTRACT

Purpose: to investigate the short and long-term acoustic parameters of voice quality in esophageal and tracheoesophageal phonation. **Methods:** the methodology used included the recording of speech samples (vowels, sentences, a standardized text and spontaneous emission) of two men (72 and 71 years old) who were total laryngectomee. One of them produced the esophageal phonation; the other produced the tracheoesophageal phonation. A speech sample of a third man (aged 76) was also recorded as a reference. The speech samples were all submitted to both short term analysis (fundamental frequency, the first three formants and maximum phonation time), and long-term analysis (long-term spectrum, intensity and frequency of spectrum peaks). **Results:** the short term analysis showed lower fundamental frequency and higher formant values for those individuals who were total laryngectomee. It was noticed that the tracheoesophageal prosthesis provided results closer to the individual used as a reference in terms of intensity, fundamental frequency and maximum phonation time, which are compatible with physiological aspects referring to the change caused by the coupling of the tracheoesophageal prosthesis and the aerodynamic events. The long-term analysis showed that the total laryngectomee speakers had less amplitude of the speech signal and a deficient definition of spectrum peaks, compared to the reference subject. **Conclusions:** considering the short and long term acoustic analysis, it was possible to compare phonation forms, in cases that there was no larynx, and to predict the place of the sphincteral activity. It was also possible to make assumptions about some aspects related to the conformation of the resonant cavities.

KEYWORDS: Voice quality; Speech, esophageal; Speech, alaryngeal; Speech acoustics; Laryngectomy

REFERÊNCIAS

1. Câncer no Brasil. Dados dos registros de base populacional.[citado 2002 jun 28]. Disponível em: http://www.inca.gov.br/regpop/2003/index.asp?link=conteudo_view.asp&ID=13
2. Kowalski LP, Miguel REV, Ulbrich FS. Câncer de laringe. In: Carrara-de-Angeli E, et al. A atuação da fonoaudiologia no câncer de cabeça e pescoço. São Paulo: Lovise; 2000. p. 97-104.
3. Bertelli AP. Câncer da laringe. São Paulo: Departamento de Cirurgia de Cabeça e Pescoço, Fundação Antonio Prudente; 1980.
4. Camargo ZA. Reabilitação fonoaudiológica em câncer de laringe. In: Pinho SMR. Fundamentos em fonoaudiologia: tratando os distúrbios da voz. Rio de Janeiro: Guanabara Koogan; 1998. p. 99-114.
5. Fúria CLB, Mourão LF, Angelis EC. Reabilitação fonoaudiológica nas laringectomias totais. In: Carrara-de-Angeli E, et al. A atuação da fonoaudiologia no câncer de cabeça e pescoço. São Paulo: Lovise; 2000. p. 227-38.
6. Doyle PC. Alaryngeal voice and speech options. In: Doyle PC. Foundations of voice and speech rehabilitation following laryngeal cancer. San Diego: Singular; 1994. p.123-9.
7. Greene MCL. Distúrbios da voz. São Paulo: Manole; 1989.
8. Gonçalves MI, Behlau M. Laringectomia total- perspectivas de reabilitação vocal. In: Lopes Filho OC. Tratado de fonoaudiologia. São Paulo: Roca; c1997. p. 1063- 78.
9. Robbins J, Fisher HB, Blom EC, Singer MI. A comparative acoustic study of normal, esophageal, and traqueoesophageal speech production. J Speech Hear Disord. 1984; 49(2): 202-10.
10. Casper JK, Colton RH. Clinical manual for laryngectomy and head/neck cancer rehabilitation. San Diego: Singular; c1993.
11. Van As CJ. Tracheoesophageal speech – a multidimensional assessment of voice quality. Niuwegein: Budde Grafische; 2001.
12. Robbins J. Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech. J Speech Hear Res. 1984; 27(4): 577-85.
13. Ceccon FP, Carrara-de Angelis E. Próteses traqueoesofágicas na reabilitação vocal após laringectomias totais. In: Carrara-de-Angeli E, et al. A atuação da fonoaudiologia no câncer de cabeça e pescoço. São Paulo: Lovise; 2000. p.119-25.
14. Rics LA, Oliveira E, Leite A, Fernandez CRF, Souza BB, Naranjo DA, et al. Avaliação da qualidade vocal em pacientes com prótese traqueoesofágica. Recife; 2001. In: Painel da

- Federação Sulamericana em Cirurgia de Cabeça e Pescoço, 3. Recife, 2001.
15. Pinho SMR, Camargo Z. Introdução à análise da voz e da fala. In: Pinho SMR. Tópicos em voz. Rio de Janeiro: Guanabara Koogan; c2001. p. 19-44.
 16. Figueiredo RM. A eficácia de medidas extraídas do espectro de longo termo para a identificação de falantes. *Cad Estud Linguist.* 1993; 25: 129-60
 17. Linville SE, Rens J. Vocal tract resonance analysis of aging voice using long-term average spectra. *J Voice.* 2001; 15 (3): 323-30.
 18. Camargo ZA, Vilarim GS, Cukier S. Parâmetros perceptivo-auditivos e acústicos de longo termo da qualidade vocal de indivíduos disfônicos. *Rev. CEFAC.* 2004; 6(2):189-96.
 19. Gonçalves MIR. Análise computadorizada de frequência fundamental e espectrográfica de formantes em fonação esofágica [dissertação mestrado]. São Paulo: Universidade Federal de São Paulo. Escola Paulista de Medicina; 1989.
 20. Gonçalves I, Behlau M, Pontes P, Tosi O. Análise computadorizada da frequência fundamental e espectrográfica de formantes em fonação esofágica. *Acta AWHO.* 1994; 13(1): 21-8.
 21. Vieira CH. Fala esofágica: um estudo de caso embasado nos achados anátomo-fisiológicos e na investigação acústica das medidas de duração. [tese]. São Paulo: Pontifícia Universidade Católica de São Paulo; 2002.
 22. Fouquet ML. Relação entre a avaliação videofluoroscópica do esôfago e a qualidade vocal esofágica. [dissertação mestrado]. São Paulo: Universidade de São Paulo. Faculdade de Medicina; 2003.
 23. Williams SE, Watson JB. Differences in speaking proficiencies in laryngectomee groups. *Arch Otolaryngol.* 1985; 111(4): 216-9.
 24. Hammarberg B, Nord L. The alaryngeal voice source as analysed by videofluoroscopy, fiberendoscopy, and perceptual – acoustic assessment. Elenius K, Branderud P, editors. In: XIII International Congress of Phonetic Sciences. Proceedings. Stockholm; 1995. v.4, p. 590-3.
 25. Sedory SE, Hamlet SL, Connor NP. Comparisons of perceptual and acoustic characteristics of traqueoesophageal and excellent esophageal speech. *J Speech Hear Disord.* 1989; 54(2):209-14.
 26. Vieira MBM, Gama ACC, Bernardes FF. Fístulas traqueoesofágicas: técnica de “amatsu”. In: Carrara-de-Angeli E, et al. A atuação da fonoaudiologia no câncer de cabeça e pescoço. São Paulo: Lovise; 2000. p. 113-8.
 27. Fant G. Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations. The Hague: Mouton; 1970.
 28. Laver J. The phonetic description of voice quality. Cambridge [Eng]; New York: Cambridge University Press; 1980.
 29. Camargo Z, Madureira S, Tsuji DH. Analysis of dysphonic voices based on the interpretation of acoustic, physiological and perceptual data In: 6th International Seminar on Speech Production. Proceedings. Sydney; 2003.
 30. Ferreira LP, Pontes PAL. Avaliação fonoaudiológica da voz: o valor discriminatório das provas respiratórias. In: Ferreira LP. Um pouco de nós sobre voz. 4. ed. São Paulo: Pró- fono; 1995. p. 1-28.
 31. Ladefoged P. Elements of acoustic phonetics. Chicago: University of Chicago Press; 1996.

RECEBIDO EM: 25/04/04

ACEITO EM: 02/09/04

Endereço para correspondência:
Rua Guairá 50 ap 162 – Jd Barbosa
Guarulhos - SP
CEP: 07111-320
e-mail: mary_sj@hotmail.com