

ACÚSTICA

Princípios da Produção e Análise de Voz

XV Escola de Inverno

Departamento de Física/ICEx/UFMG

2004

Princípios da Produção e Análise de Voz

Maurílio Nunes Vieira

XV Escola de Inverno, 26-30 de Julho 2004

Departamento de Física/ICEx/UFMG

CP 30.161-970 - Belo Horizonte/MG

1. INTRODUÇÃO

A informação transmitida pela voz humana pode ser separada em componentes que são determinadas intencionalmente ou não pelo locutor. Dentre estas componentes, destacam-se o conteúdo linguístico da elocução, isto é, o que está sendo falado (ou cantado), a identidade de quem fala (ou pistas sobre o sexo, porte físico, idade, origem sócio-cultural, etc.), e a condição de saúde ou estado emocional do locutor. No estudo da fala, a importância destas componentes dependerá da aplicação. Aplicações de interesse atual incluem [1-8]:

- *Síntese de voz* (conversão de texto para fala),
- *Reconhecimento de voz* (conversão de fala para texto),
- *Compreensão da fala* (determinação do significado da elocução),
- *Reconhecimento/identificação de locutor*,
- *Codificação de voz a baixas taxas* (compressão),
- *Análise de disfonias* (patologias da laringe).

O detalhamento de qualquer uma destas aplicações está além dos objetivos desta apostila. A ciência da fala é abrangente, multidisciplinar, envolvendo áreas como anatomia da cabeça e pescoço [9, 10], acústica [1, 11-13], psicofísica [1, 14], linguística [e.g., 15], e processamento de sinais [5, 16, 17]. Este texto tem por objetivo ser um guia para o estudo da análise e síntese de sinais de voz. Os tópicos serão apresentados de forma simplificada, mas informação adicional poderá ser obtida na bibliografia indicada.

2. TRATO VOCAL E NOÇÕES DE FONÉTICA

Os principais órgãos e cavidades do corpo humano associados com a acústica da produção da fala estão indicados esquematicamente na Fig. 1. Na produção da fala, o ar expelido pelos pulmões é conduzido pela traquéia, passa pelas pregas vocais, atinge a cavidades oral e/ou nasal, e o som (produzido de alguma forma dentro do trato vocal) é irradiado na boca e/ou narinas. O grau de acoplamento do trato nasal é determinado pela elevação do véu palatino. Os movimentos da língua, maxilar, e lábios, dão forma à cavidade oral, dependendo do som desejado. Articulação não é possível no trato nasal, mas as cavidades nasais podem ser alteradas, por exemplo, pelo excesso de secreção durante uma infecção.

As cavidades do trato vocal (*filtros*) são excitadas pela combinação de pelo menos três tipos de estímulos (*fontes*):

1. uma sequência de *pulsos de ar* produzidos pela vibração quase-periódica das pregas vocais durante a geração de sons “sonoros” (e.g.,

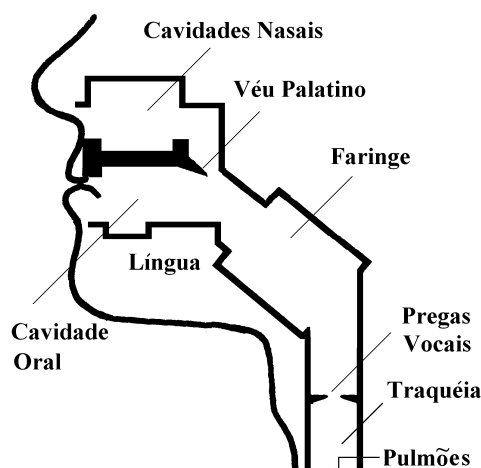


Fig. 1. Representação Esquemática do Trato Vocal. Adaptado de [10], p. 160.

vogais e consoantes como /l/ e /m/);

2. o fluxo *turbulento* de ar (*ruído*) causado por constrictões em alguma parte do trato vocal, como nos *fricativos “surdos”* (i.e., não-sonoros) /f/ e /s/;
3. um único pulso produzido pela liberação súbita do ar comprimido em uma cavidade formada dentro do trato vocal, como nos *oclusivos* (surdos) /p/ e /t/.

Fricativos sonoros (e.g., /v/, /z/) são produzidos pela combinação de ruído e vibração das pregas vocais. A diferença entre fricativos surdos e fricativos sonoros pode ser percebida tocando-se o pescoço e alternando-se /sss/ e /zzz/, ou /fff/ e /vvv/, por exemplo. A diferença entre oclusivos surdos e sonoros (e.g., /p/ versus /b/, ou /t/ versus /d/) pode ser detectada da mesma forma, ressaltando-se que em /b/ e /d/ as pregas vocais vibram durante a fase de acumulação de ar.

Noções básicas sobre fonação são apresentadas a seguir. A física da geração de turbulências, que tem sido relegada a um segundo plano no estudo da fala [18], não é simples e não será discutida neste texto. Os fenômenos associados com a vibração das pregas vocais (i.e., fonação) têm sido mais estudados, principalmente por estarem diretamente relacionados com a produção de vogais. Na maioria das aplicações, a qualidade da síntese e a precisão da análise de vogais são, em geral, mais importantes que a análise ou síntese dos outros fonemas.

3. FONAÇÃO

3.1 O ciclo vibratório

De forma simplificada, as pregas vocais fazem parte de um sistema — a laringe — que atua como uma válvula para (1) abrir e parcialmente fechar a entrada da traquéia nas fases inspiratória e expiratória da respiração, respectivamente, e (2) bloquear a entrada da traquéia rapidamente e a qualquer momento (especialmente durante a deglutição), impedindo que líquidos, alimentos, ou secreções atinjam os pulmões. A biomecânica da laringe é discutida em [19].

Fonação ocorre, normalmente, durante a fase expiratória da respiração. Para que as vibrações se

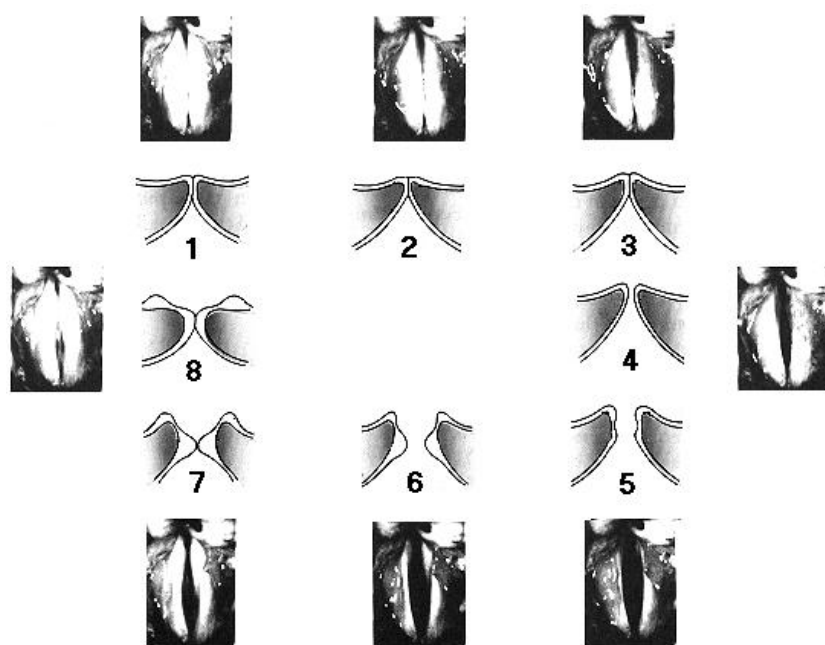


Fig. 2. Ciclo vibratório. Fotografias no plano horizontal e desenhos num plano frontal (i.e., um plano vertical entre os lados direito e esquerdo do corpo). O espaçamento temporal entre os quadros não é linear: a fase de fechamento (quadros 5-7) é mais rápida. Baseado em [7], p. 222.

iniciem, as pregas vocais devem estar tensas, fechadas, e sob uma ligeira força muscular de compressão. A pressão subglótica exercida pelos pulmões separa as pregas vocais gradualmente a partir da parte inferior e em direção à borda superior (Fig. 2). Após a separação da borda superior, as bordas inferiores aproximam-se novamente e o ciclo se repete na *frequência fundamental* (F_0). Valores médios de F_0 em adultos são 120 Hz (homens) e 220 Hz (mulheres).

Durante as vibrações, a reaproximação das bordas inferiores das pregas vocais ocorre devido a dois fatores principais: (1) a elasticidade dos tecidos que compõem as pregas vocais e

(2) uma força (“acoplamento aerodinâmico”) causada pelo movimento horizontal das bordas superiores [20]. Há também uma componente de sucção (“Efeito Bernoulli”) devido à queda de pressão no espaço entre as pregas vocais após a separação, mas este fenômeno não é essencial para a manutenção das vibrações [21]. *Deve ser enfatizado que o modelamento do movimento vibratório é uma das áreas mais ativas no estudo da fala.*

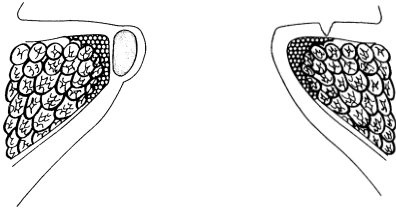


Fig. 3. Cisto (esquerda) e Sulcus Vocalis (direita). Adaptado de [23].

pode causar rouquidão. A Fig. 3 mostra duas causas comuns de rouquidão: um *cisto* (produzido, por exemplo, pela oblituração de uma glândula produtora de muco) e um *sulcus vocalis* (onde a mucosa que reveste a camada externa das pregas vocais se adere à camada interna). A análise de voz pode ser utilizada para auxiliar a detecção de doenças na laringe e para avaliar os resultados de uma cirurgia e/ou tratamento fonoterapêutico [7, 8, 32].

3.2 Modelo a duas massas

O mérito deste modelo da vibração das pregas vocais está na sua aparente simplicidade, capaz de resgatar as condições essenciais para a existência de vibrações. O modelo se baseia (1) nos efeitos aerodinâmicos dos movimentos horizontais das bordas superiores das pregas vocais sobre as partes inferiores e (2) na assimetria (atraso) entre o deslocamento das bordas superiores e inferiores das pregas vocais. Veja detalhes em [20, 22].

Variações deste modelo foram utilizadas para o estudo de paralisias [23] e tumores [24] na laringe. Modelos mais elaborados são discutidos em [26-28]. Numa visão futurística, modelos computacionais que incluam mais massas e que considerem as forças exercidas pela musculatura da laringe poderão ser úteis até mesmo para o planejamento de cirurgias.

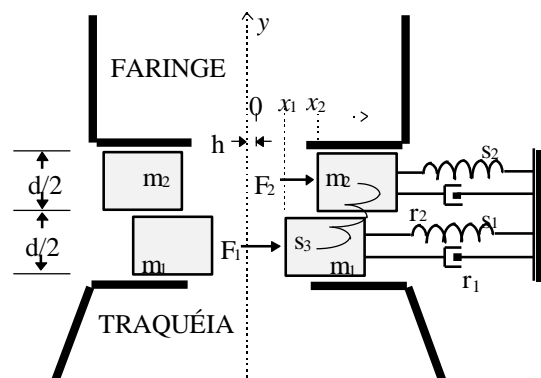


Fig. 4. Modelo a duas massas das pregas vocais. m_1 e m_2 são as massas, s_1 e s_2 são molas, r_1 e r_2 são amortecedores. O comprimento horizontal “ l ” das massas não está indicado na figura. Baseado em [20], p. 219.

3.3 Características da excitação glotal

O resultado da simulação de pulsos glotais e de um sinal acústico (vogal /a/) é apresentado na Fig. 5. Identifique, inicialmente, a frequência fundamental, F_0 . Observe, em seguida, que as variações de

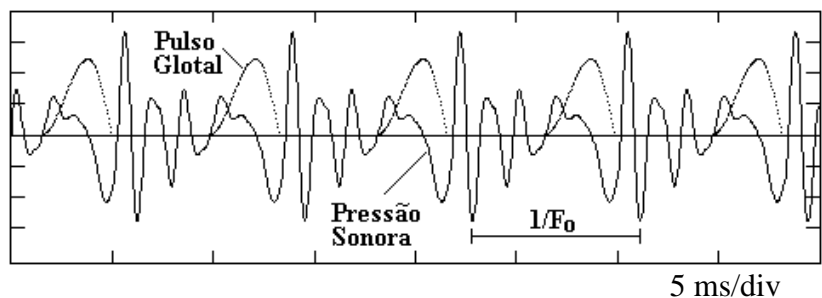


Fig. 5. Pulsos glotais e vogal /a/. Simulações do autor.

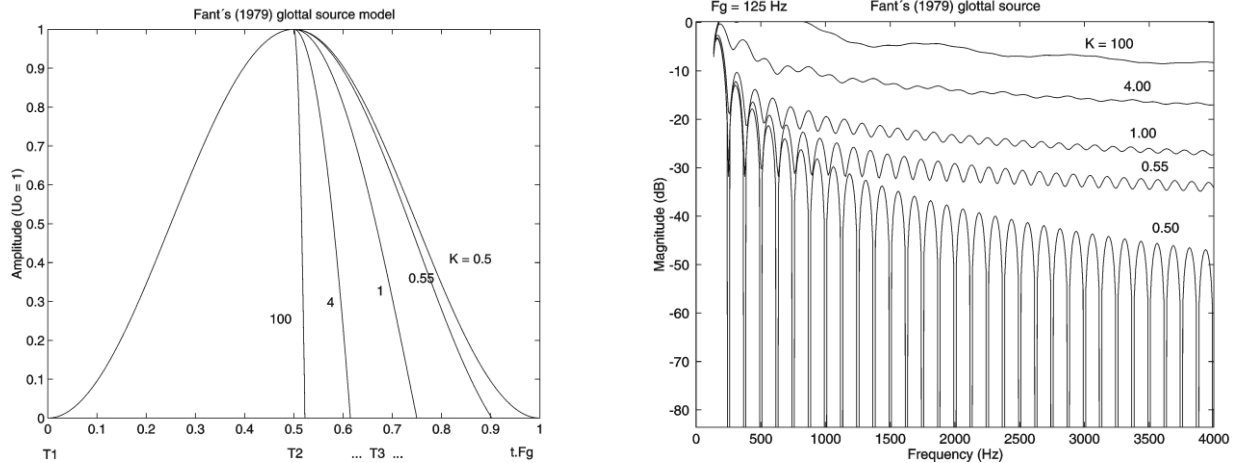


Fig. 6. Características temporais e espectrais do pulso glotal. Simulações baseadas em [30].

pressão sonora são induzidas, principalmente, pelo fechamento das pregas vocais. Note, ainda, que os pulsos glotais são assimétricos, isto é, inclinados para a direita. Esta assimetria relaciona-se com a maior velocidade da fase de fechamento do ciclo glotal, como fora mencionado na Fig. 2. A diferença entre as velocidades de abertura e fechamento é um fator fundamental para a transferência de energia do fluxo de ar para o movimento das pregas vocais, possibilitando a *manutenção* das oscilações; veja detalhes em [21, 29].

Quanto mais rápida a fase de fechamento do ciclo glotal, maior a intensidade do sinal acústico. Isto é ilustrado na Fig. 6 (acima), onde o espectro do pulso glotal é plotado para vários graus de assimetria da forma de onda; em vozes normais, “K” varia entre 0,55 e 1,00 aproximadamente. Observe, na Fig. 6, como a intensidade das componentes de alta frequência aumenta com o grau de assimetria do pulso. Isto contribui para aumentar a percepção de volume sonoro (“loudness”).

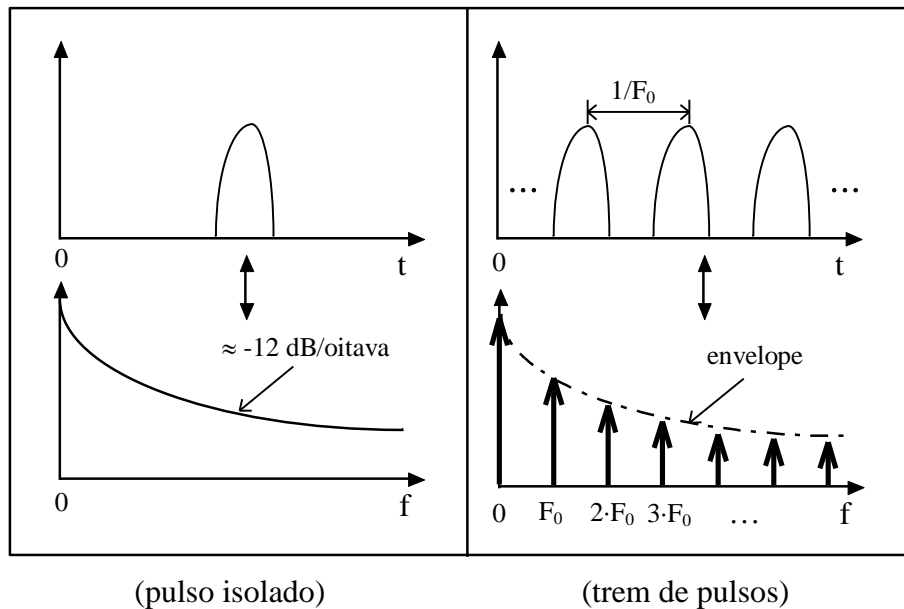


Fig. 7. Representações da excitação glottal. O espectro do trem de pulsos (direita) é dado por uma frequência fundamental (F_0) e uma série de harmônicos cujas amplitudes (envelope) são determinadas pelo espectro do pulso isolado (esquerda). A inclinação de -12 dB/oitava — mostrada fora de escala — é um valor tipicamente observado em vozes normais entre 0 e 2 kHz aproximadamente (isto pode ser medido na Fig. 6).

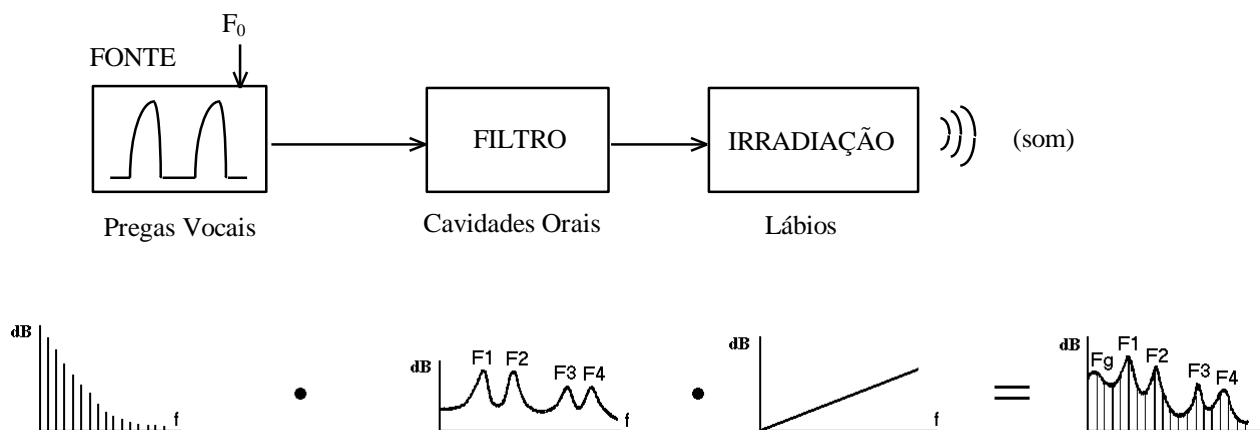


Fig. 8. Modelo fonte-filtro. Representação no domínio da frequência. F_1 -- F_4 são ressonâncias (“formantes”) do trato oral, $H(f)$; F_g relaciona-se com a forma do pulso glottal (veja detalhes em [30]). A transformação do fluxo de ar em pressão sonora é modelada por uma função de transferência tipo passa-alta (“impedância de radiação); veja detalhes em [1].

Lembre-se, das propriedades da Transformada de Fourier, que o *espectro de uma sequência* de pulsos que se repetem numa taxa F_0 é dado por *amostras do espectro do pulso*, o espaçamento entre as amostras sendo igual a F_0 . (Fig. 7, página anterior). Consequentemente, quanto mais grave for o tom da voz (“pitch”), menor será o espaçamento entre as raias espectrais e menor a intensidade relativa das componentes de alta frequência. Isto afeta a voz masculina cantada, em particular. Como o ouvido humano é mais sensível a frequências em torno de 3 kHz [14], cantores clássicos (homens) usam certas técnicas vocais [11, 33] para aumentar a intensidade dos harmônicos incidindo nesta faixa de frequências privilegiada pelo sistema auditivo.

A filtragem exercida pelo trato vocal supra-laríngeo sobre o espectro da excitação glotal será discutida em seguida. Por simplicidade, as influências das cavidades subglóticas [34] e nasais [1] serão desconsideradas. Nasalidade é um tópico relativamente complexo e cercado de controvérsias, ao passo que os efeitos acústicos das cavidades subglóticas (traquéia e, talvez, os pulmões) não são claramente compreendidos e seriam demasiadamente específicos para os propósitos deste texto.

4. MODELO FONTE-FILTRO PARA A PRODUÇÃO DE VOGAIS

Numa abordagem simplificada da acústica da produção da voz, o trato vocal pode ser decomposto em três blocos independentes: fonte(s), filtro(s) e efeitos da irradiação acústica (Fig. 8). Nesta introdução, a “fonte” será limitada à vibração periódica das pregas vocais e o “filtro” relacionará apenas às cavidades orais, isto é, admitir-se-á que o trato nasal esteja desacoplado pela elevação do véu palatino (veja Fig. 1).

No domínio da frequência, o sinal acústico pode ser visto como o *produto* dos espectros de cada um dos blocos mencionados acima. Isto está representado na Fig. 8. O espectro da vibração glotal já foi discutido na seção anterior. O efeito dos lábios (modelados como uma pequena abertura num plano refletor, i.e., a cabeça) é equivalente a um filtro passa-altas, como indicado na Fig. 8; veja detalhes em [1]. As ressonâncias do trato vocal serão analisadas em seguida, admitindo-se que área da abertura entre as pregas vocais (glote) seja desprezível em relação à área das cavidades superiores.

4.1 Modelo de tubos

Numa primeira aproximação, as ressonâncias do trato vocal na vogal neutra (o “schwa”, /ə/) podem ser estimadas a partir de um tubo cilíndrico fechado na glote e aberto nos lábios (Fig. 9).

Admitindo-se que na faixa de interesse (0-4 kHz, aproximadamente) a propagação acústica se dê apenas na direção glote-lábios (i.e, que não haja propagação transversal), os modos de ressonância do tubo corresponderão aos comprimentos de onda em que haja um nó de pressão na glote e um anti-nó nos lábios, isto é:

$$(2n + 1) \cdot \frac{\lambda}{4} = L, \quad (1)$$

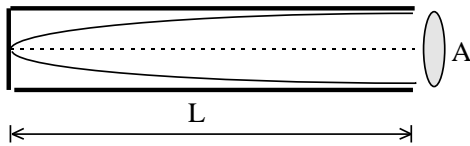


Fig. 9. Ressonâncias do trato vocal. O padrão de ondas estacionárias indicado corresponde ao modo de ressonância de menor frequência.

onde n é um número inteiro ($n \geq 0$), λ é o comprimento de onda, e L é o comprimento do tubo. O caso $n = 0$ está representado na Fig. 9. Expressando-se o comprimento de onda em termos da frequência, F , e da velocidade do som no ar, v (i.e., $\lambda = v/F$), segue que

$$F = (2n + 1) \cdot \frac{v}{4 \cdot L}. \quad (2)$$

Particularizando a Eq. 2 para $L = 17,5$ cm (homem adulto) e $v = 350$ m/s (ar úmido, 37° C), vê-se que as ressonâncias do trato vocal “neutro” ([\emptyset]) ocorrem em

múltiplos ímpares de 500 Hz, isto é,

$$F = (2n + 1) \cdot 500 \text{ [Hz]} \quad (3)$$

As frequências derivadas do tubo sem perdas ($F_1 = 500$ Hz, $F_2 = 1500$ Hz, $F_3 = 2500$ Hz, etc.) correspondem aos valores medidos em vozes reais para a vogal neutra. Contudo, o modelo não inclui perdas e não permite determinar as amplitudes e larguras de faixa das ressonâncias. Perdas de energia, que são causadas, por exemplo, por absorção nas paredes internas do trato vocal [1], alteram a forma dos picos de ressonância: quanto maior as perdas, menos aguçados serão os picos. A amplitude relativa dos picos também é alterada, porque as perdas aumentam com a frequência. Isto está representado na

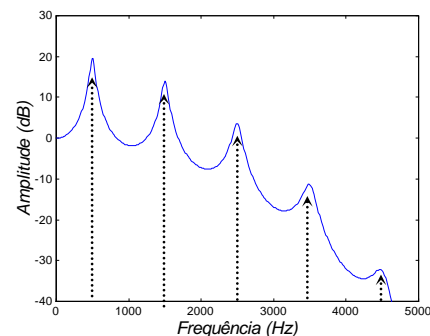


Fig. 10. Efeito das perdas. A resposta em frequência do tubo sem perdas (raias) é alterada por perdas de energia, resultando em ressonâncias mais largas e alterações na amplitude relativa dos picos (linha contínua).

Fig. 10. A largura de faixa dos formantes é particularmente crítica para síntese de voz e para a estimação das formas de onda do pulso glotal a partir do sinal acústico (*filtragem inversa* [37]).

Na produção de outras vogais, o posicionamento da língua, lábios e maxilar, altera a forma do trato vocal, criando cavidades múltiplas e invalidando a aplicação do modelo baseado em um único tubo. *Análise Preditiva Linear* (ou LPC, Linear Predictive Coding) pode ser utilizada para um modelamento que é equivalente a uma série de tubos cilíndricos concatenados. Detalhes sobre LPC — provavelmente a técnica mais popular para a análise de sinais de voz — podem ser obtidos em

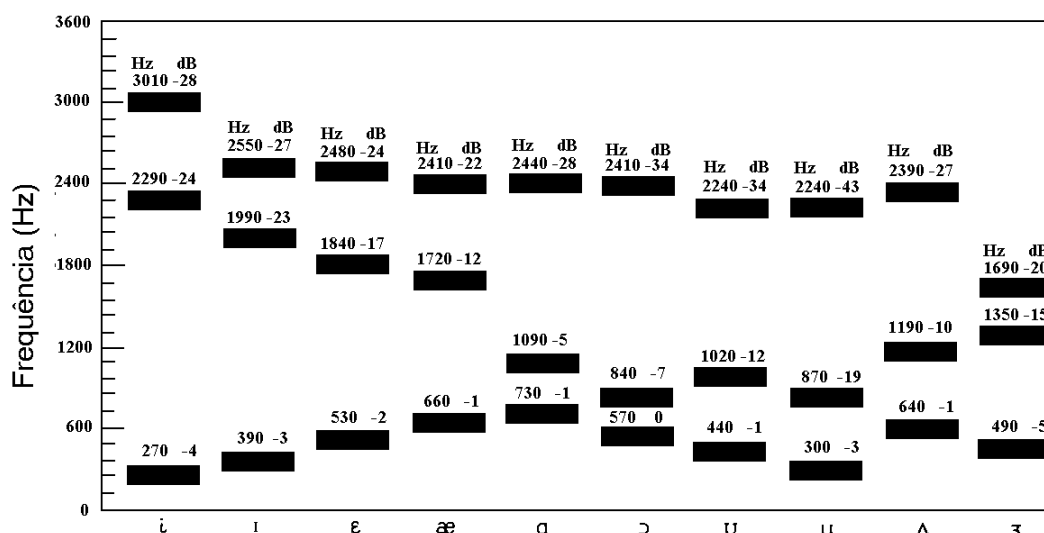


Fig. 11. Frequências e amplitudes dos formantes de vogais. Valores médios para locutores masculinos e femininos, de acordo com [1].

[3, 5]. A próxima seção apresenta um resumo das características dos formantes do trato vocal durante a produção de outras vogais.

4.2 Formantes e tipos de vogais

Os três primeiros formantes (F_1 , F_2 , F_3) estão associados com o tipo de vogal, enquanto os formantes de ordem superior estão mais relacionados com a identidade vocal do locutor. Valores médios de F_1 , F_2 , e F_3 para vogais da língua inglesa estão indicados na Fig. 11. Observe como a frequência do segundo formante aumenta significativamente em vogais próximas ao /i/, diminuindo, por outro lado, em vogais próximas ao /u/. Esta variação em F_2 é correlacionada positivamente com o avanço/recuo da ponta da língua. Correlatos acústicos de outros movimentos articulatorios são descritos em [36].

Parâmetros relacionados com F_1 , F_2 , e F_3 formam a base de sistemas de reconhecimento de voz. Deve ser enfatizado, contudo, (1) que existem *diferenças significativas* entre indivíduos, e (2) que os formantes são influenciados pelas consoantes que antecedem e sucedem as vogais.

4.3 Espectrogramas

O espectrograma é uma representação do sinal de voz no plano $F \times t$

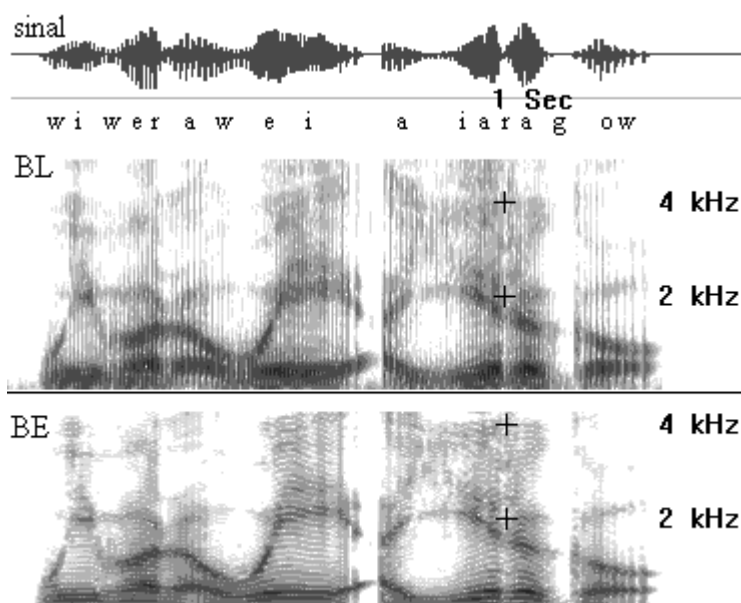


Fig. 12. Espectrograma. A sentença "we were way a year ago" (homem adulto, inglês da Escócia) é analisada em um espectrogram de banda larga (BL) e banda estreita (BE). A intensidade aumenta com a tonalidade de cinza.

(frequência versus tempo). Uma terceira dimensão, a amplitude das componentes espectrais, pode ser criada através do nível de cinza (ou cor) das curvas. Um espectrograma é, basicamente, a indicação temporal da energia na saída de um banco de filtros. Ajustando-se de alguma forma a largura de faixa (seletividade) dos filtros é possível ressaltar-se (1) os formantes ou (2) os harmônicos da excitação. Isto está ilustrado na Fig. 12. No espectrograma de banda larga (BL), há uma melhor definição dos formantes, ao passo que no espectrograma de banda estreita (BE) há uma melhor definição dos harmônicos da frequência fundamental.

5. NECESSIDADE DE PROCESSAMENTO DIGITAL DE SINAIS

O espectrograma, introduzido após o final da Segunda Guerra Mundial, é amplamente utilizado, especialmente em linguística aplicada para a análise semi-automática do sinal de voz. Espectrogramas, podem ser obtidos usando-se a Transformada Discreta de Fourier [17], como nos exemplos da Fig. 12. Contudo, em várias aplicações, é desejável ou necessário que parâmetros acústicos associados com a F_0 e com os formantes sejam estimados *automaticamente*. Isto requer o uso de técnicas mais elaboradas de processamento digital (numérico) de sinais. Tópicos fundamentais em processamento digital de sinais de voz são:

- Transformada Z [16,17],
- Transformada Discreta de Fourier/FFT [16,17],
- Implementação de filtros FIR/IIR [17],
- Estimação de formantes e F_0 usando LPC [3, 5] e análise cepstral [5].

Deve ser ressaltado que o processamento digital de sinais oferece inúmeras técnicas extremamente úteis na investigação experimental. A vogal /a/ mostrada na Fig. 5 (p. 3), por exemplo, foi simulada através da filtragem (convolução) dos pulsos com um filtro digital representando cinco formantes; veja detalhes em [5, 30, 32].

Nos últimos 10 anos, o processamento digital de sinais tem penetrado cada vez mais em diversas áreas e vários livros-texto foram publicados. Abordagens básicas são dadas em [16, 38]; a referência [17] é um texto intermediário, enquanto [35] é um livro mais avançado. As referências [3, 5] são específicas para sinais acústicos de voz.

6. CONCLUSÃO

A compreensão dos mecanismos de vibração das pregas vocais pode levar a avanços significativos na análise de patologias da laringe e na síntese de voz com características emotivas (tristeza, alegria, etc.). Nesta apostila, os princípios fundamentais da anatomia e da acústica da fala foram discutidos de forma concisa, com o intuito de ser um guia para estudos posteriores. O principal objetivo da discussão foi apresentar o modelo Fonte-Filtro na produção de vogais. Um certo enfoque foi dado à biomecânica da vibração das pregas vocais, que é uma das áreas mais ativas na pesquisa básica da ciência da fala. Outra área de interesse atual, com implicações na fonoaudiologia, é o estudo de relações entre medidas acústicas e impressões subjetivas da qualidade de voz (e.g., rouquidão, sopro, aspereza). Por fim, deve ser ressaltado que o estudo experimental da produção da fala não se limita apenas a sinais acústicos: a análise de outros sinais (e.g., sinais eletroglotográficos [7, 32]) e imagens endoscópicas [39] também se faz necessária.

7. REFERÊNCIAS¹

- [1] Flanagan, J. (1972). *Speech Analysis, Synthesis and Perception (Second Edition)*. Springer-Verlag, New York.
- [2] Holmes, J. N. (1988). *Speech Synthesis and Recognition*. Chapman & Hall, London.

¹ A referência [1] é um texto clássico e altamente recomendado como uma leitura inicial. Os livros [2] e [4] são os textos mais básicos desta lista.

- [3] Markel, J. D., & Gray, A. H. (1976). *Linear Prediction of Speech*. Springer-Verlag, New York.
- [4] Owens, F. J. (1993). *Signal Processing of Speech*. MacMillan, Hampshire.
- [5] Rabiner, L. R., & Schafer, R. W. (1978). *Digital Processing of Speech Signals*. Prentice-Hall, New Jersey.
- [6] Rabiner, L. R., & Juang, B. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall, New Jersey.
- [7] Baken, R. J. (1987). *Clinical Measurement of Speech and Voice*. Taylor & Francis, London.
- [8] Hirano, M. (1981). *Clinical Examination of Voice*. Springer-Verlag, New York.
- [9] Kaplan, H. M. (1971). *Anatomy and Physiology of Speech (Second Edition)*. McGraw-Hill, New York.
- [10] Perkins, W. H., & Kent, R. D. (1986). *Textbook of Functional Anatomy of Speech, Language, and Hearing*. Taylor & Francis, London.
- [11] Titze, I. R. (1994). *Principles of Voice Production*. Prentice-Hall, New Jersey.
- [12] Fant, G. (1970). *Acoustic Theory of Speech Production*. Mouton, The Hague.
- [13] Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (1982). *Fundamentals of Acoustics (Third Edition)*. John Wiley & Sons, New York.
- [14] Roederer, J. G. (1995). *The Physics and Psychophysics of Music: An Introduction (Third Edition)*. Springer-Verlag, New York.
- [15] Gleason, H. A. (1961). *An Introduction to Descriptive Linguistics (Revised Edition)*. Holt, Rinehart and Winston, New York.
- [16] Oppenheim, A. V., Willsky, A. S., & Nawab, S. H. (1997). *Signals & Systems (Second Edition)*. Prentice-Hall, New Jersey.
- [17] Oppenheim, A. V., & Schafer, R. W. (1975). *Digital Signal Processing*. Prentice Hall, New Jersey.
- [18] Teager, H. M., & Teager, S. M. (1990). Evidence for nonlinear sound production mechanisms in the vocal tract. In: *Speech Production and Speech Modelling*, edited by H. W. Hardcastle and A. Marchal (Kluwer), pp. 241-261.
- [19] Fink, B. R., & Demarest, R. J. (1978). *Laryngeal Biomechanics*. Harvard University Press, Cambridge.
- [20] Broad, D. (1979). The new theories of vocal fold vibration. In: *Speech and Language: Advances in Basic Research and Practice (vol. 2)*, edited by N. Lass (Academic Press, New York), pp. 203-256.
- [21] Titze, I. R., (1988). The physics of small-amplitude oscillation of the vocal folds. *Journal of the Acoustical Society of America* 83, pp. 1536-1552.
- [22] Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal* 51, pp. 1233-1268.
- [23] Hirano, M., & Bless, D. M. (1993). *Videostroboscopic Examination of the Larynx*. Singular Publishing Group, San Diego.
- [24] Isshiki, N., Tanabe, M., Ishizaka, K., & Broad, D. (1977). Clinical significance of asymmetrical vocal cord tension. *Annals of Otolaryngology* 86, pp. 58-66.
- [25] Koizumi, T., Taniguchi, S. & Itakura, F. (1993). An analysis-by-synthesis approach to the estimation of vocal cord polyp features. *Laryngoscope* 103, pp. 1035-1042.
- [26] Titze, I. R., (1973). The human vocal cords: a mathematical model (Part I). *Phonetica* 28, pp. 129-170.
- [27] Titze, I. R., & Strong, D. T. (1975). Normal modes in vocal cord tissues. *Journal of the Acoustical Society of America* 57, pp. 736-744.
- [28] Titze, I. R., (1989). A four-parameter model of the glottis and vocal fold contact area. *Speech Communication* 8, pp. 191-201.

- [29] Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica* 34, pp. 264-279.
- [30] Fant, G. (1979). Glottal source and excitation analysis. *Speech Transmission Laboratory Quarterly Progress and Status Report 1* (Royal Institute of Technology, Stockholm) pp. 85-107.
- [31] Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *Speech Transmission Laboratory Quarterly Progress and Status Report 4* (Royal Institute of Technology, Stockholm) pp. 1-13.
- [32] Vieira, M. N. (1997). *Automated Measures of Dysphonias and the Phonatory Effects of Asymmetries in the Posterior Larynx*. Ph.D. Thesis, University of Edinburgh.
- [33] Sundberg, J. (1987). *The Science of the Singing Voice*. Northern Illinois University Press, Dekalb.
- [34] Hanson, H. M., & Stevens, K. N. (1995). Sub-glottal resonances in female speakers and their effect on vowel spectra. *Proceedings of ICPHS95: International Conference on Phonetic Sciences (Stockholm)*, vol. 3, pp. 182-185.
- [35] Orfanidis, S. J. (1988). *Optimum Signal Processing: An Introduction (Second Editoin)*. McGraw-Hill, New York.
- [36] Lindblom, B. E. F., & Sundberg, J. E. F. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movements. *Journal of the Acoustical Society of America* 50, pp. 1166-1179.
- [37] Javkin, H. R., Antoñanzas-Barroso, N., & Maddieson, I. (1987). Digital inverse filtering for linguistic research. *Journal of Speech and Hearing Research* 30, pp. 122-129.
- [38] Damper, R. I. (1995). *Introduction to Discrete-Time Signals and Systems*. Chapman & Hall, London.
- [39] Wittenberg, T., Mergell, P., Tigges, M., & Eysholdt, U., (1997). Quantitative characterization of functional voice disorders using motion analysis of highspeed video and modelling. *Proceedings of ICASSP'97: International Conference on Acoustic, Speech, and Signal Processing (IEEE; Munich)*, pp. 1663-1666.

ANEXO A: Periódicos de Interesse

- *Journal of the Acoustical Society of America*,
- *Journal of Speech, Language and Hearing Research*, cujo nome, até 1996, era *Journal of Speech and Hearing Research*,
- *Journal of Voice* (USA),
- *IEEE Signal Processing Letters*,
- *IEEE (Institute of Electrical and Electronic Engineers) Transaction on Acoustics, Speech, and Signal Processing*, que, por volta de 1993, foi dividido em:
 - *IEEE Transactions on Speech and Audio Processing*, e
 - *IEEE Transactions on Signal Processing*.