

COMPARAÇÃO DE MODELOS MECÂNICOS DE LARINGE NA PRODUÇÃO DE VOGAIS

Edson Cataldo

Universidade Federal Fluminense, Departamento de Matemática Aplicada, Centro, Niterói
Programa de Pós-graduação em Engenharia de Telecomunicações,
PGMEC – Programa de pós-graduação em Engenharia Mecânica
Rua Mário Santos Braga, s/ no - 24020-140 – Centro – Niterói – RJ – Brasil
ecataldo@mec.puc-rio.br

Rubens Sampaio

Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Eng. Mecânica
Rua Marquês de São Vicente, 225 - 22453-900 – Gávea – Rio de Janeiro – Brasil ,
rsampaio@mec.puc-rio.br

Lucas Nicolato

Universidade Federal Fluminense, Departamento de Engenharia de Telecomunicações
Rua Passo da Pátria, 156 - 24120-240 - São Domingos – Niterói – RJ- Brasil
lucasnicolato@yahoo.com.br

Resumo. *O mecanismo mais importante relacionado à produção de voz é o das cordas vocais. O ar proveniente dos pulmões é forçado através do pequeno espaço existente entre as cordas vocais, causando seu movimento em uma freqüência determinada pela tensão dos músculos associados. Este movimento causa a modificação do fluxo de ar, resultando em pulsos de ar que serão amplificados e modificados pelas cavidades oral e nasal até ser irradiado pela boca. Este complexo processo é modelado por um sistema de equações integro-diferenciais. Neste artigo, apresentam-se três modelos mecânicos das cordas vocais e um deles, o mais complexo (modelo de Titze), é descrito completamente. Neste caso, discute-se todo o processo usado para a produção de vogais. Os resultados obtidos com os três modelos são comparados. Embora os resultados obtidos sejam preliminares, são satisfatórios na produção de vogais e dão razões para acreditar que podemos continuar nesse caminho para a síntese articulatória da voz.*

Palavras-chave: Modelagem, simulação, produção de vogais, processamento de sinais

1. INTRODUÇÃO

Uma das principais motivações para o entendimento do mecanismo da produção de voz está no fato que a voz humana é um dos principais meios de comunicação.

A produção da voz se inicia com uma contração-expansão dos pulmões. Cria-se, assim, uma diferença entre a pressão do ar nos pulmões e a pressão do ar na frente da boca, causando um escoamento de ar. O escoamento passa pela laringe e, antes homogêneo, vai se transformando em uma série de pulsos (conhecidos como trem de pulsos ou sinal glotal) de ar que chegam na boca e na cavidade nasal. Os pulsos de ar são modulados pela língua, pelos dentes e lábios; isto é, pela geometria destes órgãos, de forma a produzir o que conhecemos por voz. O sinal glotal, porém, possui propriedades importantes de difícil

reprodução que estão intimamente ligadas às características anatômicas e fisiológicas da laringe.

Atualmente, a teoria mais aceita para a descrição do sinal glotal (isto é, o aparecimento do trem de pulsos) é a teoria chamada de aerodinâmica-mioelástica, proposta por van den Berg (1958) e Titze (1980). Esta teoria postulou que o movimento de abrir e fechar as cordas vocais é regido pelas propriedades mecânicas dos tecidos musculares que constituem, principalmente, as cordas vocais e pelas forças aerodinâmicas que se distribuem ao longo da laringe durante a fonação. A ação neural consiste apenas em aproximar as cordas vocais de tal forma que a superfície destas vibrem.

Para facilitar o estudo do sistema de produção da voz, normalmente ele é reduzido a quatro grupos distintos, em relação à onda sonora que é produzida ou modificada pelos órgãos. O primeiro grupo, chamado de grupo da *respiração*, corresponde à produção de um fluxo de ar; que inicia-se nos pulmões e termina no final da traquéia. Na faringe, encontram-se os órgãos do segundo grupo, responsáveis pela produção do sinal glotal, chamado de grupo da *vocalização*. O sinal glotal é um sinal de baixa intensidade, que necessita ser amplificado e que determinadas componentes harmônicas sofram "ênfase", de maneira que os fonemas sejam caracterizados. Este é o grupo chamado de grupo de *ressonância*. Esse fenômeno ocorre na passagem do ar pelo chamado trato vocal (porção que vai da laringe até a boca). Finalmente, as ondas de pressão são irradiadas quando chegam à boca. Esse grupo é chamado de grupo da *irradiação*.

Um dos principais diferenciadores do som é a forma de excitação. Identificamos dois tipos básicos de excitação: (1) vocálico e (2) não-vocálico. Neste trabalho, discutiremos apenas a produção de sons vocálicos, no nosso caso, a produção de vogais.

2. MODELAGEM MATEMÁTICA

Na produção de vogais, o fluxo de ar proveniente dos pulmões é interrompido pela vibração quase-periódica das cordas vocais, conforme ilustrado na figura a seguir.

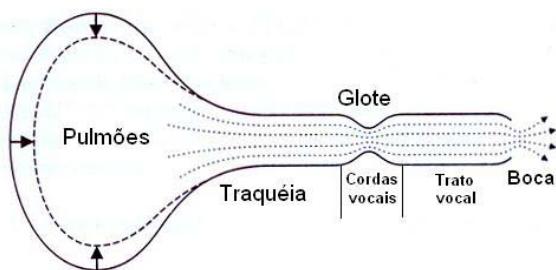


Figura1 – Representação esquemática do sistema de produção da voz (adaptado de Titze (1994)).

A partir da representação esquemática do sistema de produção de voz, podem-se observar os quatro grupos destacados anteriormente: respiração (pulmões e traquéia), vocalização (cordas vocais), ressonância (trato vocal) e radiação (boca).

Nas últimas décadas, a dinâmica das cordas vocais tem sido extensivamente estudada e alguns modelos mecânicos foram desenvolvidos. Esses modelos diferem pela representação que fazem das cordas vocais, vistas como sistemas mecânicos que modulam a passagem do ar.

Dentre esses modelos podem-se destacar o modelo de uma única massa (Flanagan and Landgraf, 1968), modelos de duas massas (Ishizaka and Flanagan, 1972, 1977; Flanagan e Ishizaka, 1978; Koizumi et al, 1987), modelo de três massas (Titze, 1994), modelo com múltiplas massas (Titze, 1973, 1974), e modelos contínuos (Titze e Strong, 1975; Titze e Talkin, 1979; Titze, 1984). Há, ainda, modelos tridimensionais, bem mais sofisticados, como o apresentado por Rosa (2002).

Como a produção da voz é caracterizada pela mudança da forma do trato vocal, pode-se esperar que um modelo mais realista de trato vocal consista de um tubo que varia como um função do tempo e da posição ao longo do eixo de propagação do som. A formulação de um modelo desse tipo do trato vocal pode ser extremamente complexo. Um método de simplificação deste modelo é o de representar o trato vocal como uma série de tubos acústicos concatenados.

Para o caso da produção de vogais, considera-se que a área da seção do tubo varia apenas com a posição e não com o tempo.

3. APRESENTAÇÃO DOS MODELOS

3.1. Introdução

Nesta seção, serão apresentados os três modelos a serem usados para a produção de vogais. Para os dois primeiros modelos será feita uma breve discussão, sem considerar muitos detalhes. Porém, para o terceiro modelo, será feita uma descrição mais detalhada, pois é o modelo considerado mais completo dentre os estudados aqui. Servirá como base para a discussão dos resultados desejados.

3.2. Modelo de Flanagan-Landgraf (1968)

O primeiro modelo a ser discutido será o modelo de Flanagan e Landgraf (1968), cujo circuito acústico usado para representar o sistema vocal é o da Fig.2:

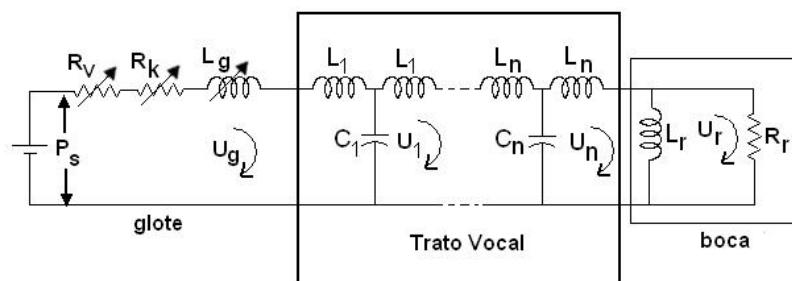


Figura 2 – Representação do circuito acústico para a produção de vogais.

No caso da produção de vogais, os pulmões aparecem como uma fonte de pressão constante e de baixa impedância. Além disso, a queda de pressão através das (grandes) áreas dos brônquios e da traquéia é relativamente pequena. Assim, aproxima-se a pressão subglotal pela bateria esquematizada acima com valor de tensão P_s . Usando resultados experimentais obtidos por van den Berg (1957), a impedância glotal (R_v) dependente do fluxo, a resistência cinética (R_k) dependente do fluxo e a inertância (L_g) devido à massa do pulso de ar glotal são dadas em termos da viscosidade cinemática do ar, da espessura das cordas vocais, dos comprimentos das cordas vocais, das áreas do orifício glotal, da densidade do ar e do fluxo do ar através do orifício glotal. Esses valores podem ser encontrados em Flanagan e Landgraf (1968).

Neste modelo, as cordas vocais são consideradas como um sistema massa-mola-amortecedor (massa M , constante da mola K e amortecimento B). O sistema é forçado por uma força $F(t)$, dada pelo produto da pressão do ar na glote pela área da superfície intraglotal. A força age na face da corda vocal, como descrito na figura a seguir. A força é distribuída e sua resultante, que não aparece na Fig. 3, pode ser pensada como aplicada na massa M .

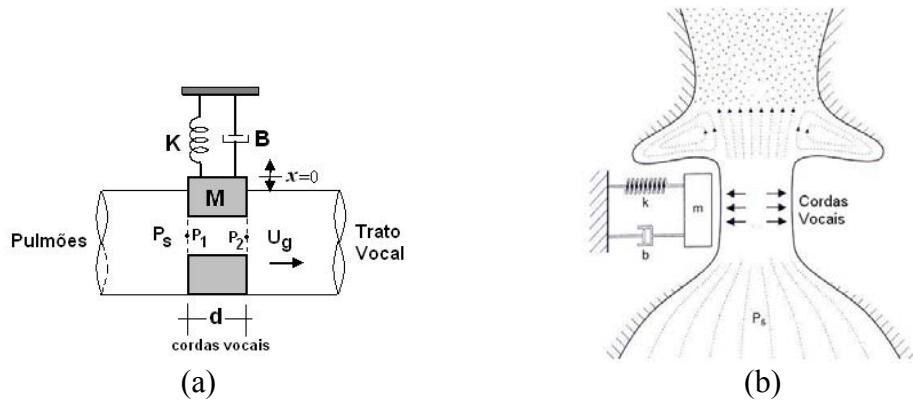


Figura 3. (a) Modelo mecânico de segunda ordem para as cordas vocais. (b) Sistema vocal para o Modelo de Flanagan e Landgraf (1968) (adaptado de Titze (1994)).

3.3. Modelo de Ishizaka e Flanagan (1972)

Repetidas observações do movimento das cordas vocais, em câmera lenta, têm demonstrado que as cordas vocais não se movem como corpos rígidos. A mucosa vibra em um movimento que pode ser pensado quase que independente do corpo que ela recobre. Devido a essas observações, dois modelos podem ser criados, considerando-se o movimento de duas ou três massas.

Nesta seção, será apresentado o modelo de duas massas e na próxima seção, o modelo com três massas.

O sistema considerado para este modelo é mostrado na figura a seguir:

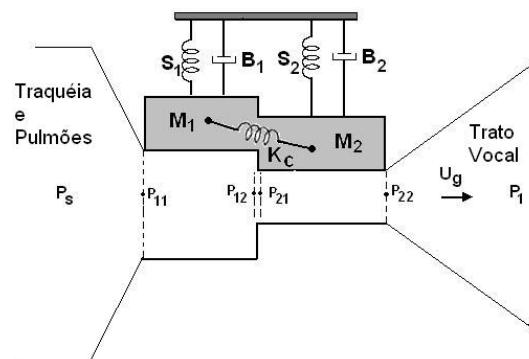


Figura 4. Modelo mecânico de segunda ordem para as cordas vocais - modelo de Flanagan e Ishizaka (1972).

O circuito acústico utilizado para este modelo será o mesmo do modelo de Titze, discutido a seguir. Desta forma, sua apresentação será feita na discussão do referido modelo.

O sistema considerado para a dinâmica das cordas vocais é um sistema de dois graus de liberdade. As molas S_1 e S_2 são não-lineares e a mola K_c é linear.

3.4. Modelo de Titze (1994)

Este modelo considera a inclusão de uma terceira massa em relação ao modelo anterior, conforme descrito na Fig. 5:

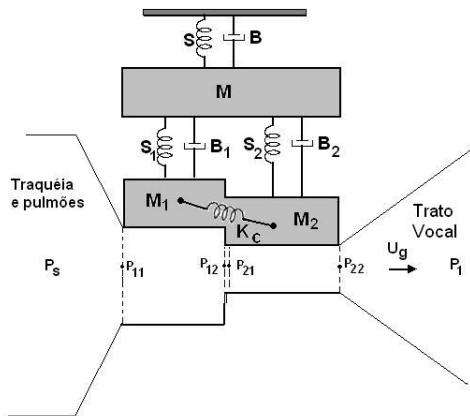


Figura 5. Modelo das cordas vocais para o modelo de Titze (1994).

O sistema é composto por três massas (M , M_1 e M_2). A essas massas são permitidos apenas movimentos perpendiculares ao tubo vocal x , x_1 e x_2 . As duas massas M_1 e M_2 são acopladas por uma mola linear, de rigidez k_c . As molas S , S_1 e S_2 são representações equivalentes da tensão nas cordas vocais e têm características não-lineares. A relação não-linear entre a deflexão a partir da posição de equilíbrio e a força requerida para produzir tal deflexão é dada por $f = kx(1 + \eta x^2)$, no caso da mola k , onde f é a força requerida para produzir x , k é a rigidez linear e η é o coeficiente que descreve a não-linearidade da mola S .

Durante o fechamento da glote, considera-se uma força de contato agindo quando as massas colidem. Essa força causará deformação nas cordas vocais. A força de restauração durante esse processo de colisão pode ser representada por uma mola equivalente s_{hj} , $j = 1, 2$. Considera-se uma característica não-linear para essa mola equivalente s_{hj} ; isto é,

$$f_{hj} = h_j \left(x_i + \frac{A_{g0i}}{2l_g} \right) \left\{ 1 + \eta_{hj} \left(x_j + \frac{A_{g0i}}{2l_g} \right)^2 \right\} \quad \text{para } x_i + \frac{A_{g0i}}{2l_g} \leq 0, \quad j = 1, 2 \quad (1)$$

sendo f_{hj} a força requerida para produzir a deformação na massa, M_i , durante a colisão, h_j é a rigidez linear e η_{hj} é um coeficiente positivo representando a não-linearidade das cordas vocais em contato. A força resultante agindo em M_i durante o fechamento é, então, a soma de f_{si} com f_{hi} .

Toma-se como base as relações entre as diferenças de pressões na glote, como descrito em Flanagan e Ishizaka (1972), a impedância acústica do orifício glotal constitui o circuito equivalente mostrado na Fig. 6.

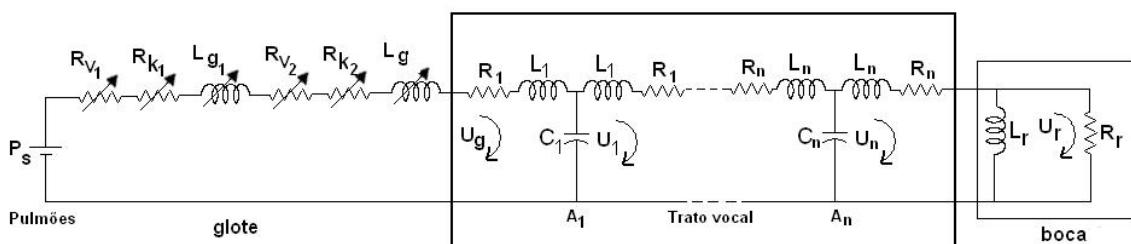


Fig. 6 – Circuito equivalente para o modelo de Flanagan e Ishizaka (1972) e de Titze (1994).

Considerou-se a viscosidade dinâmica do ar (μ) igual a $1,84 \times 10^{-5}$, as espessuras d_1 e d_2 das cordas vocais iguais a 0,0032 m, o comprimento ℓ_g das cordas vocais igual a $1,8 \times 10^{-2}$ m, as áreas neutras do orifício glotal A_{g01}, A_{g02} (cordas vocais em repouso) iguais a $0,05 \text{ cm}^2$ e a densidade do ar (ρ) igual a $1,3 \text{ g/cm}^3$.

Os deslocamentos das massas, x_1 e x_2 , são obtidos pelas equações a seguir, que descrevem a dinâmica das cordas vocais para este modelo.

$$\begin{cases} M\ddot{x} + B\dot{x} + S(x) + S_1(x - x_1) + B_1(\dot{x} - \dot{x}_1) + S_2(x - x_2) + B_2(\dot{x} - \dot{x}_2) = 0 \\ M_1\ddot{x}_1 + S_1(x_1 - x) + B_1(\dot{x}_1 - \dot{x}) + k_c(x_1 - x_2) = F_1 \\ M_2\ddot{x}_2 + S_2(x_2 - x) + B_2(\dot{x}_2 - \dot{x}) + k_c(x_2 - x_1) = F_2 \end{cases} \quad (2)$$

As molas S, S_1 e S_2 seguem as características não-lineares descritas anteriormente. As forças F_1 e F_2 são calculadas em termos das pressões do ar na glote. O trato vocal é representado por uma linha de transmissão de n cilindros, de paredes rígidas, áreas das seções transversais A_1, A_2, \dots, A_n , e os comprimentos dos cilindros $\ell_1, \ell_2, \dots, \ell_n$. Os valores das indutâncias, das capacitâncias e das resistências podem ser obtidos em Ishizaka e Flanagan (1972).

Usando leis das malhas, chegamos às seguintes equações para o circuito da figura:

Malha 0: correspondente à seção das cordas vocais

$$\begin{aligned} (R_{k1} + R_{k2})U_g|U_g + (R_{v1} + R_{v2})U_g + (L_{g1} + L_{g2})\frac{dU_g}{dt} + L_1\frac{dU_g}{dt} + R_1U_g \\ + \frac{1}{C_1} \int_0^t (U_g - U_1)dt - P_s = 0 \end{aligned} \quad (3)$$

Malha 1: correspondente à primeira seção do trato vocal

$$(L_1 + L_2)\frac{dU_1}{dt} + (R_1 + R_2)U_1 + \frac{1}{C_2} \int_0^t (U_1 - U_2)dt + \frac{1}{C_1} \int_0^t (U_1 - U_g)dt = 0 \quad (4)$$

Malha 2: correspondente à segunda seção do trato vocal

$$(L_2 + L_3)\frac{dU_2}{dt} + (R_2 + R_3)U_2 + \frac{1}{C_3} \int_0^t (U_2 - U_3)dt + \frac{1}{C_2} \int_0^t (U_2 - U_1)dt = 0 \quad (5)$$

e, assim por diante, até a penúltima malha.

Malha n: correspondente à última seção do trato vocal – n -ésima seção

$$(L_n + L_R)\frac{dU_n}{dt} + R_nU_n - L_R\frac{dU_R}{dt} + \frac{1}{C_n} \int_0^t (U_n - U_{n-1})dt = 0 \quad (6)$$

Malha da irradiação na boca:

$$L_R \frac{d(U_R - U_n)}{dt} + R_R U_R = 0 \quad (7)$$

Para resolver o sistema obtido, usaram-se equações de diferença, considerando:

$$\frac{df}{dt} \approx \frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}} \quad \text{e} \quad \int_a^b f(t) dt \approx T \sum_{j=0}^{i-1} f(t_j), \text{ sendo } T = \frac{b-a}{i} \quad (8)$$

conforme proposto por Flanagan e Landgraf (1972).

Comparando os modelos, verificou-se que o terceiro requeria mais tempo de CPU, mas, em compensação, apresentou os melhores resultados, principalmente no que diz respeito ao som produzido. Como resultado, pode-se obter a função que descreve a área da seção glotal Ag (observa-se que esse valor depende do trato vocal), a função que descreve o fluxo glotal Ug , a pressão acústica na boca (representada pelo produto da corrente U_n pela impedância de radiação) e a função resposta em freqüência do trato vocal. Um outro resultado, igualmente importante, é o som obtido a partir da modelagem. Para obtenção dos resultados descritos acima, consideram-se tratos vocais de comprimento 17cm. A comparação é feita considerando a produção da vogal 'a'. Os resultados obtidos foram comparados com os encontrados nos trabalhos de Flanagan e Landgraf (1968) e de Ishizaka e Flanagan (1972). Além disso, o Modelo de Titze foi implementado.

4. COMPARAÇÃO ENTRE OS GRÁFICOS DA ÁREA DA SEÇÃO GLOTAL (Ag), DO FLUXO GLOTAL (Ug) e DA PRESSÃO ACÚSTICA NA BOCA

A comparação entre os gráficos da área glotal (Ag), do fluxo glotal (Ug) e da pressão acústica na boca, para os três modelos considerados será feita. Será mostrado, primeiro, na Fig. 7, o sinal obtido de um sinal de voz (real) na emissão de uma vogal 'a' sustentada.

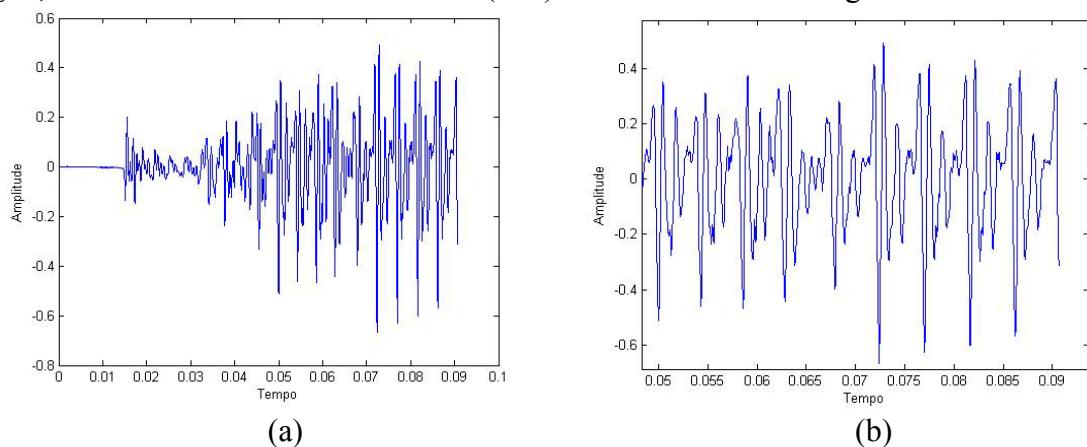


Figura 7. Sinal de voz correspondente à vogal 'a' (sustentada): (a) sinal durante todo o tempo da emissão da voz, (b) apenas um trecho do sinal-destaque para a peridodicidade do sinal.

Serão mostrados, na Fig. 8, os gráficos relacionados à produção da vogal 'a', usando os três modelos discutidos. São usadas dezessete seções do trato vocal, com valores sugeridos por Koizumi et al (1987).

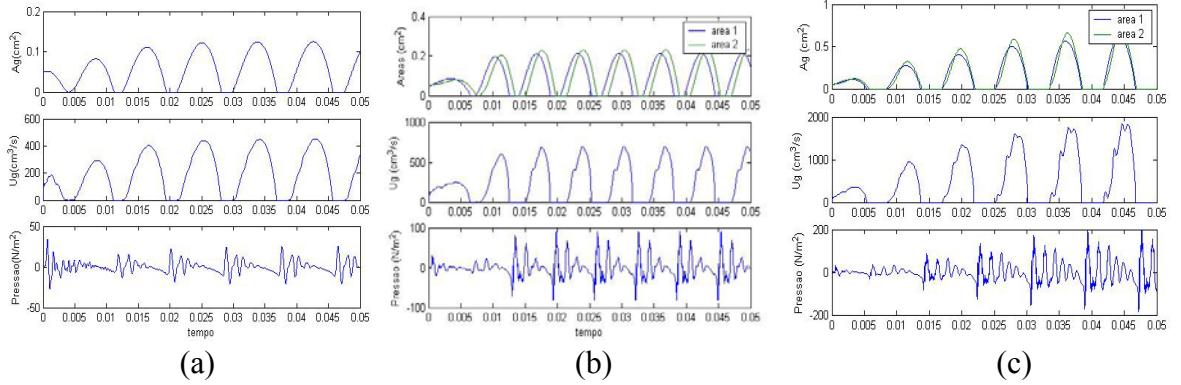


Figura 8. Gráficos da(s) área(s) da seção glotal, do fluxo glotal e da pressão irradiada para os três modelos considerados, no caso da emissão da vogal ‘a’ (vogal ‘a’ sustentada): (a) Modelo de 1968; (b) Modelo de 1972; (c) Modelo de Titze.

5. COMPARAÇÃO ENTRE AS FUNÇÕES RESPOSTA EM FREQÜÊNCIA

Será feita uma comparação entre as funções resposta em freqüência obtidas através da razão entre o espectro de amplitude da corrente U_m do circuito acústico e o espectro de amplitude da corrente de entrada U_g . Com isso, obtém-se os formantes (freqüências naturais do trato vocal). Quer-se mostrar a diferença encontrada quando aumenta-se o número de seções do trato vocal. Usamos para comparação apenas o modelo de 1968. Os gráficos das funções resposta em freqüência são mostrados na Fig. 9. Valores típicos para os quatro primeiros formantes para a vogal ‘a’, de acordo com Rabiner e Schafer (1978) são: 650 Hz, 1075 Hz, 2463 Hz e 3558 Hz.

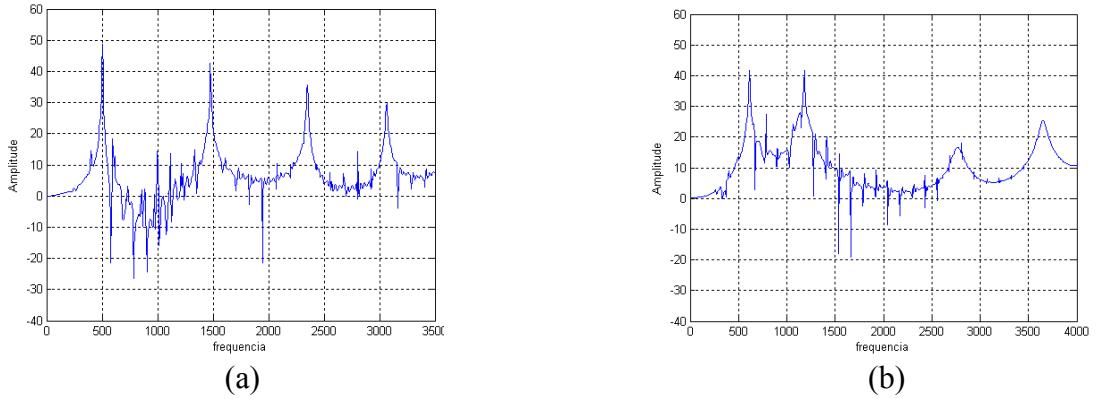


Figura 9. Espectros obtidos com as simulações do modelo de 1968, considerando: (a) aproximação de quatro seções para o trato vocal e (b) aproximação de dezessete seções para o trato vocal.

6. CONCLUSÕES

Ao compararem-se os três modelos discutidos para a produção de vogais, observa-se que o terceiro modelo fornece os melhores resultados. Os gráficos do fluxo glotal, da área glotal e da pressão acústica dos três modelos e verificar que, no terceiro modelo, com o acréscimo da terceira massa, há uma demora para a estabilização dos pulsos, porém os sinais, em cada período, são mais próximos dos obtidos com um sinal real. Quanto ao som produzido, conseguimos identificar uma sensível melhora, em relação ao obtido com o segundo modelo e uma melhora substancial com o som produzido pelo primeiro modelo. Embora não tenhamos mostrado os gráficos da função resposta em freqüência obtidos com os três modelos, observamos que os gráficos obtidos ao usarmos dezessete seções do trato vocal são bem próximas no caso real.

7. REFERÊNCIAS

Flanagan, J. e Landgraf, L., 1968, "Self-oscillating source for vocal-tract synthesizers", IEEE Trans. On Audio and Eletroacoustics, Vol. 16, pp. 57-64.

Ishizaka, K. and Flanagan, J., 1972, "Synthesis of voiced sounds from two-mass model of the vocal cords", Bell Syst. Tech. Journal, Vol. 51, pp. 1233-1268.

Ishizaka, K. and Flanagan, J., 1977, "Acoustic properties of longitudinal displacement in vocal cord vibration", Bell Syst. Tech. Journal, Vol. 56, pp. 889-918.

Ishizaka, K. and Flanagan, J., 1978, "Computer model to characterize the air volume displaced by the vibrating vocal cords", Journal of the Accoustical Society American, Vol. 63, pp. 1559-1565.

Koizumi, T.; Taniguchi, S.; Hiromitsu, S., 1987, "Two-mass models of the vocal cords for natural sounding voice synthesis", Journal of the Accoustical Society American, Vol. 82(4), pp. 1179-1192.

Rabiner, L.R. e Schafer, R.W., 1978, "Digital Processing of Speech Signals", Prentice Hall, Englewood Cliffs, New Jersey.

Rosa, M. O., 2002, "Laringe Digital", Tese de doutorado, Universidade Federal de São Carlos.

Titze, I. R., 1994, "Principles of Voice Production", Prentice Hall, Englewood Cliffs, New Jersey.

Titze, I. R., 1973, "The Human Vocal Cords: A Mathematical Model-Part I", Phonetica Vol. 28, pp. 129-170.

Titze, I. R., 1974, "The Human Vocal Cords: A Mathematical Model-Part II", Phonetica Vol. 29, pp. 1-21.

Titze, I. R., 1975, "Normal modes in vocal cord tissues", Journal of the Accoustical Society American, Vol. 57(3), pp. 736-744.

Titze, I. R., 1979, "A Theoretical study of the effects of various laryngeal configurations on the acoustics of phonation", Journal of the Accoustical Society American, Vol. 66(1), pp. 60-74.

Titze, I. R., 1980, "Comments on the myoelastic-aerodynamic theory of phonation", The Journal of the Acoustical Society of America, Vol. 23, pp. 495-510.

Titze, I. R., 1984, "Parametrization of the glottal area, glottal flow, and vocal fold contact area", Journal of the Accoustical Society American, Vol. 75, pp. 570-580.

Titze, I. R., "Principles of voice production". Prentice-hall, NJ: Englewood Cliffs, NJ, 1994.

Van den Berg, J. W., Zantema, J. T. e Doornenbal, P. Jr., 1957, "On the air resistance and the Bernoulli effect of the human larynx, Journal of Accoustical Society American, Vol. 29, pp. 626-631.

Van den Berg, J., 1958, "Myoelastic-aerodynamic theory of voice production", Journal of Speech and Hearing Research, Vol. 1, pp. 227-244.

8. DIREITOS AUTORAIS

Os autores são os únicos responsáveis pelo conteúdo do material impresso incluído no seu trabalho.

COMPARISON OF THREE MODELS OF LARYNX IN THE PRODUCTION OF VOWELS

Edson Cataldo

Universidade Federal Fluminense, Departamento de Matemática Aplicada, Centro, Niterói
Programa de Pós-graduação em Engenharia de Telecomunicações,
PGMEC – Programa de pós-graduação em Engenharia Mecânica
Rua Mário Santos Braga, s/ no - 24020-140 – Centro – Niterói – RJ – Brasil
ecataldo@mec.puc-rio.br

Rubens Sampaio

Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Eng. Mecânica
Rua Marquês de São Vicente, 225 - 22453-900 – Gávea – Rio de Janeiro – Brasil ,
rsampaio@mec.puc-rio.br

Lucas Nicolato

Universidade Federal Fluminense, Departamento de Engenharia de Telecomunicações
Rua Passo da Pátria, 156 - 24120-240 - São Domingos – Niterói – RJ- Brasil
lucasnicolato@yahoo.com.br

***Abstract** The most important mechanism involved in the production of the voice is played by the vocal cords. Air coming from the lungs are forced through the narrow space between the two sides of the vocal cords that are set in motion in a frequency that is governed by the tension in the attached muscles. The vocal cords change the type of flow that comes from the lungs and as the flow passes through the oral and nasal cavities it is amplified and changed until it is radiated from the mouth. This complex process can be modeled by a system of integral-differential equations. In this paper are presented three mechanical models of the vocal cords and the most complex one, the Titze model, is described completely. In this case, it is discussed all of the process used for the production of vowels. The results of the three models are compared. Although the models are rather crude, the results obtained in the production of vowels are satisfactory and give reason to believe that this way can be followed to the articulatory synthesis of voice.*

Keywords: Modeling, Simulation, Voice production, Signal processing