ANÁLISE DINÂMICA NÃO-LINEAR DE SINAIS DE VOZES AFETADOS POR PATOLOGIAS NA LARINGE

Silvana CUNHA COSTA (1); Washington COSTA (2); Benedito G. AGUIAR NETO (3)

(1) IFPB/Campus J.Pessoa, e-mail: silvana@ifpb.edu.br
 (2) IFPB/Campus J. Pessoa\ UFCG/Campus C. Grande, e-mail:washington@ifpb.edu.br
 (3) UFCG/Campus C. Grande, e-mail: bganeto@dee.ufcg.edu.br

RESUMO

Patologias na laringe podem causar desordens na fala, afetando a sua qualidade e prejudicando a comunicação oral do indivíduo. Elas podem ser de origem orgânica, como pólipos, cistos, edemas e nódulos ou de origem neurológica tais como paralisia unilateral ou bilateral nas dobras vocais, mal de Parkinson ou de Alzheimer, entre outras. Este trabalho trata do estudo de técnicas de análise dinâmica não-linear para a caracterização de sinais de vozes afetados por patologias da laringe de origem orgânica e/ou neurológica e vozes saudáveis. São focalizadas as patologias edemas de Reinke e paralisia nas dobras vocais. As medidas empregadas na análise são: o primeiro mínimo da função de informação mútua, a entropia de Shannon e a entropia de correlação. O comportamento dessas medidas será analisado para os sinais patológicos comparados com as mesmas medidas obtidas para vozes saudáveis. Esse estudo deve servir de base para a futura construção de um sistema que possa identificar desordens presentes na fala devido à presença de patologias na laringe. Isso possibilitará o desenvolvimento de ferramentas de baixo custo, baseadas em técnicas de processamento digital de sinais para auxiliar médicos especializados em doenças da laringe a efetuarem pré-diagnósticos e acompanhamento de tratamentos, diminuindo a necessidade do uso de exames mais invasivos como os de videolaringoscopia.

Palavras-chave: análise dinâmica não-linear de sinais de voz, patologias na laringe, entropia de Shannon, informação mútua, entropia de correlação

1 INTRODUÇÃO

A voz é uma emissão acústica voluntária, considerada principal ferramenta de comunicação humana, de alta complexidade e inúmeros fatores envolvidos na sua produção. As estruturas anatômicas e os processos fisiológicos envolvidos no complexo sistema de produção da fala humana são responsáveis pelos diversos fonemas emitidos e conseqüentemente carregam informações sobre suas formas, fisiologia e dinâmica.

Várias pesquisas têm sido feitas utilizando técnicas baseadas no modelo linear de produção da fala, por meio de medidas paramétricas ou não-paramétricas, tais como: Codificação por Predição Linear, Análise Cepstral e suas derivadas (coeficientes cepstrais, delta-cepstrais, cepstrais ponderados e delta-cepstrais ponderados), Mel-cepstral (delta mel e delta-delta mel) com aplicações em sistemas de reconhecimento de fala e de locutor, melhoramento do sinal de voz, classificação de vozes patológicas, entre outras (COSTA, 2008; AGUIAR NETO et al, 2007; GODINO-LLORENTE et al,2006; BAHOURA and PELLETIER, 2004).

A maioria dos esquemas de produção da fala está baseada na hipótese do modelo linear para a produção da voz. Tal hipótese, no entanto, carrega intrinsecamente suas limitações, o que leva à busca de estudos com o objetivo de se incorporar modelos não-lineares para representar a produção da fala. O estudo da complexidade dos sinais de voz tem levado em consideração os aspectos não-lineares da voz e o comportamento dinâmico das dobras vocais na produção da fala (KUMAR and MULLICK, 1996; JIANG et al, 2006; ZHANG et al, 2005; TORRES et al, 2003).

Os sistemas dinâmicos não-lineares têm como ponto central o processamento dos sinais caóticos. De forma simplificada, um sistema dinâmico é um sistema com uma memória de estados, ou seja, o conhecimento sobre o comportamento passado do sistema é armazenado em variáveis de estado e influencia o comportamento presente e o futuro do sistema. Os sistemas dinâmicos são representados em um espaço de estados (por meio das equações de transição de estados correspondentes) ou por equações de diferenças/diferenciais.

Como observado por vários pesquisadores, as não-linearidades dos mecanismos da fonte da voz humana (por exemplo, a relação não-linear do fluxo de pressão na glote, as curvas não-lineares de esforço dos tecidos das dobras vocais e as não-linearidades associadas com as colisões das dobras vocais) são fatores que sugerem a presença de caos na voz (ZHANG et al, 2004; TORRES et al, 2003).

Titze et al (TITZE, 1993) sugeriram como melhorar o entendimento das desordens vocais com conceitos e métodos de análise dinâmica. Desde então, pesquisadores têm aplicado essas novas ferramentas para estudar condições anormais associadas com patologias da laringe, para diferenciar vozes normais de patológicas e diagnosticar patologias e avaliar os efeitos dos tratamentos clínicos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Doenças da Laringe

Desordens vocais são resultantes das alterações morfológicas nas estruturas do trato vocal ou de seu mau funcionamento. A presença da patologia pode ser percebida por meio de sintomas relatados por pacientes aos seus médicos como queixa de sensações associadas à fonação ou dores na região da garganta. Alguns sintomas podem ser verificados, outros não. Outros sintomas podem referir-se às características perceptuais da voz, como a rouquidão, garganta arranhando ou tremor na voz (PARRAGA, 2002).

Quanto aos sinais acústicos, é importante lembrar que a voz é produzida por movimentos das dobras vocais interrompendo o fluxo de ar regressivo que vem dos pulmões. Os movimentos das dobras são controlados pelas características biomecânicas das próprias dobras vocais, pela magnitude da pressão de ar abaixo das dobras e por seu controle neural. A patologia pode afetar tais movimentos interferindo em quaisquer dessas variáveis (PARRAGA, 2002).

Alterações na voz podem não estar associadas a uma patologia orgânica ou física, sendo denominadas de disfonia funcional. Algumas das vozes mais ásperas podem não ter relação com uma patologia orgânica. Por outro lado, um problema orgânico sério, como câncer em fase inicial pode não produzir qualquer alteração na voz. A natureza da mudança da voz, por vezes, sugere a presença da patologia. A duração da anormalidade da voz pode ser de apenas alguns dias, com a laringite que sucede a uma infecção do trato respiratório superior, ou pode ser de muitos anos, em que a rouquidão é causada pelo edema de Reinke (BENJAMIN, 2000).

A produção de voz variável e irregular, quando as palavras são difíceis de compreender, pode ser a primeira manifestação de um distúrbio neurológico. Alguns sintomas laríngeos podem ser a característica inicial de uma doença sistêmica generalizada ou de condições neurológicas. Como exemplos de doenças por fatores neurológicos, que causam distúrbios na voz, podem ser citadas: doença cerebrovascular, paralisia cerebral, paresia ou paralisia da prega vocal, refluxo gastroesofágico decorrente de incoordenação neuromuscular, parkinsonismo, tremor essencial, disfonia espamódica e esclerose múltipla.

Quanto às doenças orgânicas que causam distúrbios na voz estão: laringite crônica, pólipo de prega vocal, nódulo vocal, edemas de Reinke, granuloma vocal idiopático, paquidermia de contato, hemorragia de prega vocal, trauma laríngeo por intubação prolongada, trauma laríngeo externo, papilomas respiratórios múltiplos, doença laríngea cística, hiperqueratose, displasia, carcinoma ou outra malignidade laríngea, anormalidades congênitas, artrite reumática, entre outras (BENJAMIN, 2000).

Na Figura 1 está ilustrado o aspecto das dobras vocais saudáveis.

As pregas vocais têm papel preponderante na fonação. São estruturas multi-laminadas e cada camada apresenta propriedades mecânicas diferentes (ZITTA, 2005). De modo geral, as pregas vocais são duas dobras de músculos e mucosas que se estendem horizontalmente na laringe.

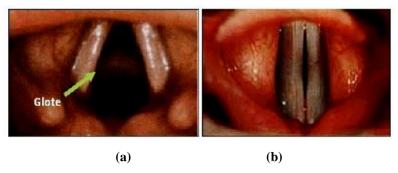


Figura 1 - Pregas vocais normais em: (a) abdução e (b) adução - visão endoscópica.

A presença da patologia pode afetar o padrão vibratório das dobras vocais e o grau de fechamento da mesma, modificando a qualidade da voz.

Entre as patologias da laringe, que afetam particularmente as dobras vocais, serão aqui destacadas: edemas de Reinke e paralisia.

2.1.1 Edema de Reinke

O edema de Reinke caracteriza-se pela expansão, aumento e inchaço das dobras vocais e pelo acúmulo de líquido ou material gelatinoso (ou ainda semi-sólido) na camada superficial da lâmina própria (espaço de Reinke) das dobras vocais (Figura 2) (KUHL, 1982). As dobras vocais ficam enormes com um edema clarohialino, pobre em vasos e que modifica completamente o aspecto anatômico da região glótica (KUHL, 1982; PARRAGA, 2002).

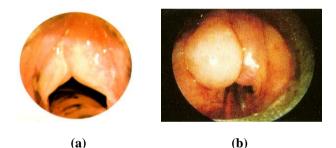


Figura 2 – (a) Edema de Reinke; (b) Edema de Reinke severo.

A incidência é maior nas mulheres que atingiram a menopausa. Entretanto, encontra-se nos dois sexos e é uma doença de idade adulta. A causa irritativa é o abuso vocal e, especialmente, o fumo, principalmente em mulheres entre 50 e 70 anos de idade (BENJAMIN, 2002).

2.1.2 Paralisia

Segundo Benjamin (2002), a paralisia da dobra vocal pode ser causada por patologia central ou periférica e pode envolver as conexões centrais, o nervo vago, o nervo laríngeo recorrente e/ou o nervo laríngeo superior.

Entre as causas de paralisia unilateral das dobras vocais em lactantes e crianças estão: lesão por trauma de parto, anomalias congênitas do coração e grandes vasos, cirurgias para corrigir essas anomalias, cirurgias intratorácicas de cistos e tumores, pós-intubação endotraqueal, entre outras ou, ainda, alguma causa não-definida. Acomete mais freqüentemente o lado esquerdo do que o direito.

As principais causas de paralisia unilateral de prega vocal em adultos são: viral ou idiopático, pós-cirúrgico (tireoidectomia, cirurgia cardíaca, entre outras), doença maligna do pescoço ou mediastino, pós-intubação endotraqueal, trauma cervical, entre outras. Os sintomas perceptuais mais comuns da paralisia unilateral são a soprosidade e a rouquidão (BENJAMIN, 2002).



Figura 3 – Paralisia nas dobras vocais (dobra paralisada durante a fonação).

Fonte: (DE BIASE et al., 2003).

2.2 Análise Não-Linear de Sinais de Voz

As medidas empregadas, neste trabalho, para análise dos sinais de voz são: O primeiro mínimo da função de informação mútua (FMMI), entropia de Shannon (H) e entropia de correlação ou entropia de Rényi de segunda ordem (h_2), descritas a seguir.

2.2.1 Função de Informação Mútua

A teoria da informação procura identificar o quanto de informação se pode ter uma medida realizada em um determinado instante de tempo t, quando se observa outra medida, do mesmo sinal, em um tempo posterior $t+\tau$. A informação mútua média fornece as mesmas informações que a função de correlação fornece em sistemas lineares sendo, na realidade, um tipo de generalização para sistemas não-lineares (SAVI, 2006).

Para um sinal x(t) e sua versão defasada a informação mútua média, $I(\tau)$, é calculada por meio da Equação 1:

$$I(\tau) = \sum_{t=\tau} \Gamma_{t+\tau} \left(x(t), x(t+\tau) \right) \log_2 \left[\frac{\Gamma_{t+\tau} \left(x(t), x(t+\tau) \right)}{\Gamma_{t+\tau} \left(x(t) \right) \Gamma_{t+\tau} \left(x(t+\tau) \right)} \right],$$
 [Eq. 01]

Em que $\Gamma_{t+\tau}(x(t))$ é a probabilidade da medida x(t), $\Gamma_{t+\tau}(x(t+\tau))$ é a probabilidade da medida $x(t+\tau)$ e $\Gamma_{t+\tau}(x(t),x(t+\tau))$ é a probabilidade da medida combinada de x(t) e $x(t+\tau)$ e $x(t+\tau)$

As informações mútuas médias não são funções das variáveis x(t), mas de um funcional da probabilidade combinada de x(t) e $x(t+\tau)$. Se x(t) e $x(t+\tau)$ são iguais, então $I(\tau)$ é máximo. Por outro lado, se x(t) e $x(t+\tau)$ são completamente independentes então o argumento do termo logarítmico é a unidade e $I(\tau)$ =0.

Fraser e Swinney (1986) estabelecem que o valor do tempo de defasagem ideal para reconstrução do diagrama de estados de um sistema dinâmico corresponde ao primeiro mínimo local, quando este existir, da função de informação mútua média $I(\tau)$. Desta forma, ao traçar $I(\tau)$ versus τ , a melhor defasagem corresponde ao valor de τ no primeiro mínimo local (FMMI – First Minimum Mutual Information). Com isso, garante-se que o espaço de estado reconstruído seja topologicamente equivalente ao espaço de estado do sistema dinâmico original.

2.2.2 Entropia de Shannon

A noção de entropia está ligada ao grau de desordem existente na informação. Dessa forma, pode ser uma ferramenta interessante para modelar o comportamento ruidoso e desordenado das vozes afetadas por algumas patologias, quando comparado às vozes normais.

Dado um determinado sinal, a entropia de Shannon é definida como (SHANNON, 1948):

$$H = -\sum_{i=1}^{M} p_i \log_2(p_i),$$
 [Eq. 02]

em que p_i é a probabilidade de que o sinal pertença a um intervalo i e M é o número de partições em seu espaco de fase.

A entropia H é a medida de informação necessária para localizar um sistema em um determinado estado, e reflete a medida da incerteza sobre o sistema físico (TORRES et al., 2003).

2.2.3 Entropia de Rényi

A entropia de Rényi, uma generalização da entropia de Shannon, é uma família de funcionais para quantificação da diversidade, incerteza ou aleatoriedade de um sistema.

A entropia de Rényi de ordem α , para $\alpha > 0$ é definida como (RÉNYI, 1961):

$$I^{(\alpha)} = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{M} p_i^{\alpha} \right),$$
 [Eq. 03]

em que p_i é a probabilidade do evento i.

No limite, quando $\alpha \to 1$, a entropia de Rényi converge para a entropia de Shannon. A entropia de Rényi de segunda ordem ($\alpha = 2$) é denominada entropia de correlação.

3 DESCRIÇÃO DA PROPOSTA

O objetivo principal desse trabalho é analisar sinais de vozes de pessoas saudáveis (sem patologia na laringe) e de pessoas com as patologias edemas de Reinke ou paralisia usando medidas das técnicas de análise não-linear: Tempo de atraso ótimo, primeiro mínimo da função de informação mútua e entropias de Shannon e de correlação. Pretende-se investigar o comportamneto dessas medidas nesses sinais e avaliar a possibilidade da utilização das mesmas para diferenciar vozes normais de vozes patológicas.

4 METODOLOGIA, RESULTADOS, ANÁLISE E INTERPRETAÇÃO DOS DADOS

A base de dados usada neste trabalho foi desenvolvida pelo Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab (KAY ELEMETRICS, 1994). Nesta pesquisa são empregados 53 sinais de vozes normais (32 masculinas e 21 femininas), 53 de vozes afetadas por paralisia (24 masculinas e 29 femininas) e 44 com edema de Reinke (12 masculinas e 32 femininas). A base contém vozes contínuas, com frase de 12 segundos, como também gravações da vogal sustentada /a/. Estas últimas são empregadas neste trabalho, tendo em vista que na emissão da vogal sustentada há vibração das dobras vocais, possibilitando assim, uma análise do padrão de vibração e do quanto o sinal de voz é afetado pelas patologias em estudo.

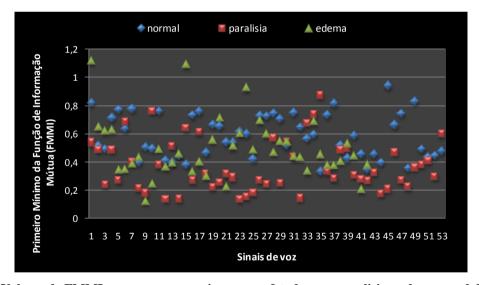
As medidas são obtidas por meio de rotinas provenientes do pacote de rotinas TISEAN package (HEGGER et al, 1999), em linguagem C, combinadas com programação em Matlab.

Na Tabela 1estão apresentados os valores médios obtidos para as variáveis empregadas. O primeiro mínimo da função de Informação Mútua (FMMI) foi menor para os sinais afetados por paralisia, mas não apresenta uma grande discrepância entre os sinais do gênero masculino afetados pelas outras patologias. Para as vozes normais ou saudáveis, esse valor é consideravelmente maior, indicando que os sinais são mais correlacionados ou "mais comportados" estatisticamente que os sinais patológicos. Observa-se, ainda, a diferença de valores para sinais de locutores do gênero masculino em relação ao feminino.

Tabela 1 – Valores médios para o primeiro mínimo da função de informação mútua (FMMI), entropia de Shannon e entropia de correlação.

	FMMI		Entropia de Shannon		Entropia de Correlação	
Sinais de Voz	fem	masc	fem	masc	fem	masc
Normal	0,67	0,44	2,54	2,45	0,67	0,68
Paralisia	0,40	0,33	2,45	2,35	0,75	0,71
Edema	0,54	0,36	2,39	2,28	0,65	0,67

Na Figura 4 estão representados os valores de FMMI para o caso de sinais de vozes afetados por paralisia, edema e vozes normais. Observa-se que os valores para vozes normais estão concentrados acima de 0,4, enquanto que para paralisia concentram-se, em sua maioria, abaixo desse valor, havendo uma maior confusão dos valores de edema com os valores obtidos para vozes normais. A partir da Tabela 1 pode-se observar ainda que, os valores do primeiro mínimo da função de informação mútua, para feminino e masculino apresentam diferenças expressivas, sendo normais ou patológicas. Esse resultado deve ser levado em conta quando esses dados forem aplicados a um processo de classificação. Resultados melhores podem ser obtidos aplicando-se classificadores distintos para vozes masculinas e vozes femininas.



 $Figura\ 4-Valores\ de\ FMMI\ para\ vozes\ normais\ e\ vozes\ afetadas\ por\ paralisia\ e\ edema\ nas\ dobras\ vocais.$

Para a Entropia de Shannon (Figura 5), observa-se uma grande similaridade entre os valores obtidos para o gênero masculino e o gênero feminino, seja entre vozes normais ou mesmo entre as vozes patológicas. Há uma maior separação, no entanto, para vozes de pessoas com edemas em relação às vozes normais.

Na Figura 6, pode-se observar o comportamento da Entropia de correlação (h_2) para as vozes com edema, paralisia e vozes normais. Os valores médios obtidos (Tabela 1) não apresentam muitas diferenças para estes sinais. No entanto, observa-se uma maior regularidade dos valores para vozes normais e vozes afetadas por paralisia, ou seja, uma maior concentração entre 0,6 e 0,8. Já para os sinais com edema, os valores de h_2 são bastante dispersos e variam de acordo com o grau da patologia.

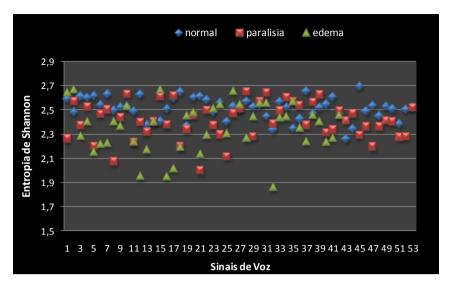


Figura 5 – Entropia de Shannon para sinais de vozes normais, com paralisia e edema.



Figura 6 – Entropia de Correlação para sinais de vozes normais e vozes com edema e paralisia.

5 DISCUSSÃO E CONSIDERAÇÕES FINAIS

Os resultados obtidos aplicando-se as medidas obtidas por meio da análise não-linear, consideradas neste trabalho, demonstram que os sinais de vozes afetados por patologias nas dobras vocais apresentam comportamento diferenciado dos sinais de vozes ditas saudáveis.

No caso da Informação mútua, as vozes afetadas por paralisia apresentam aspecto mais diferenciado em relação às vozes normais do que as vozes com edema.

Já para o caso da entropia de Shannon, as vozes com edema se diferenciam mais das vozes normais que as vozes com paralisia.

Para o caso da Entropia de Correlação ou de Renyi, torna-se necessário avaliar melhor o poder de discriminação desta medida, principalmente entre voz afetada por paralisia e voz normal.

Os resultados obtidos ainda estão em estudo e pretende-se investigar o uso de medidas estatísticas específicas para avaliação de desempenho e métodos de classificação de padrões para observar o grau de confiabilidade dessas medidas para discriminar vozes de diferentes patologias e fazer um pré-diagnóstico das mesmas.

Outras medidas serão investigadas, bem como a combinação destas medidas como um vetor de parâmetros não-lineares, buscando verificar a viabilidade desta técnica em discriminar vozes patológicas de vozes normais.

REFERÊNCIAS

- AGUIAR NETO, B. G., FECHINE, J. M., COSTA, S. CUNHA, MUPPA, M., Feature Estimation for Vocal Fold Edema Detection Using Short-Term Cepstral Analysis. Proceedings of the 7th International Conference on Bioinformatics and Bioengineering, 14-17 Oct., p. 1158-1162, 2007.
- BAHOURA, M., PELLETIER, C. Respiratory Sound Classification using Cepstral Analysis and Gaussian Mixture. Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, September, 2004.
- BENJAMIN, B. Cirurgia Endolaríngea. Editora Revinter: Rio de Janeiro, 2000.
- COSTA, S. L. DO N. C. Análise Acústica Baseada no Modelo Linear de Produção da Fala para Discriminação de Vozes Patológicas. 2008. 160p. Tese (Doutorado em Engenharia Elétrica) -, Universidade Federal de campina Grande, Paraíba, Brasil, 2008.
- DE BIASE, N., PONTES, P., Nóbrega, J. A. M., DE BIASE, S. **Paralisia do músculo cricoaritenóideo lateral: relato de caso como contribuição ao estudo da anatomia funcional da laringe**. Revista Brasileira de Otorrinolaringologia, vol.69, no.1, São Paulo, Jan./Feb., 2003.
- FRASER, A. M. and SWINNEY, H. L. **Independent Coordinates for Strange Attractors from Mutual Information.** Physical Review A., v.33, p. 1134-1140, 1986.
- FURUI, S. Cepstral Analysis Technique for Automatic Speaker Verification. IEEE Trans. on ASSP, Vol. 29, No. 2, pp 254-272, 1981.
- GODINO-LLORENTE, J. I., GÓMEZ-VILDA, P.,BLANCO VELASCO, M. Dimensionality Reduction of a Patholgical Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters. IEEE Transactions on Biomedical Engineering, Vol. 53, No. 10, October, 2006.
- HEGGER, R.; KANTZ, H. and SCHREIBER, T. **Practical implementation of nonlinear time series methods: The TISEAN package**, CHAOS 9, 413, 1999.
- JIANG, J. J., ZHANG, Y. and McGILLIGAN, C. Chaos in Voice, From Modeling to Measurement. Journal of Voice, Vol. 20, No. 1, pp.2-17, 2006.
- KAY ELEMETRICS. Kay Elemetrics Corp. Disordered Voice Database. Model 4337, 1994.
- KUHL, I. **Manual Prático de Laringologia**. Livro Texto/II. Editora da Universidade UFRGS. Porto Alegre, 1982.
- KUMAR, A. and MULLICK, S. K. **Nonlinear dynamical analysis of speech**. J. Acoust. Soc. Am, 100 (1), July, 1996.
- PARRAGA, A. Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas. Universidade Federal do Rio Grande do Sul. Dissertação de Mestrado, 2002.
- RÉNYI, A. **On Measures of Entropy and Information**. Proc. Fourth Berkeley Symp. Math. Stat. and Probability, vol. 1. Berkeley, CA: University of California Press, pp. 547-561, 1961.
- SAVI, M. A. Dinâmica não-linear e Caos. Rio de Janeiro: E-papers, 2006.
- SHANNON, C., A **Mathematical Theory of Communication**. The Bell System Technical Journal, vol. 27, no. 3, pp. 623-656, 1948.
- TITZE, IR, BAKEN R., HERZEL H. **Evidence of chaos in vocal fold vibration**. Vocal Fold Physiology: New Frontiers in Basic Science, San Diego, CA: Singular Publishing Group, 143-188, 1993.
- TORRES, M.E., GAMERO, L.G. and RUFINER, H.L. **Study of Complexity in Normal and Pathological Speech Signals.** Proceedings of the 25 Annual International Conferences of the IEEE EMBS, Cancun, México, September 17-21, 2003.
- ZHANG, Y., JIANG, J,J. and FEROZE, F.A. Wavelet-based denoising for improving nonlinear dynamic analisys of pathological voices. Eletronics Letters 4th, Vol. 42, No. 16, August, 2005.
- ZITTA, S. M. Análise Perceptivo-Auditiva e Acústica em Mulheres com Nódulos Vocais. Dissertação. Mestrado em Engenharia Biomédica. CEFET-PR, Curitiba, Paraná, 2005.