



O Uso de Características Temporais de Sinais de Voz na Detecção de Patologias Laríngeas

Maria do Carmo de Luna Malheiros Frazão¹, Silvana Luciene do Nascimento Cunha Costa², Suzete Élide Nóbrega Correia², Washington César de Almeida Costa³

¹Graduanda do Curso Superior em Engenharia Elétrica - IFPB. Bolsista do CNPq. e-mail: marafrazao@gmail.com

²Professoras da Coordenação de Engenharia Elétrica - IFPB. e-mail: silvana@ifpb.edu.br; suzete.correia@gmail.com

³Professor da Coordenação de Ciências da Natureza, Matemática e suas Tecnologias - IFPB. e-mail: washington@ifpb.edu.br

Resumo: A avaliação acústica de vozes patológicas está diretamente relacionada à extração de características. Quanto mais representativa for à característica e proporcione uma diferenciação entre sinais de classes diferentes, melhor o desempenho do sistema. Este estudo apresenta a classificação dos sinais de voz como saudáveis ou patológicos usando as medidas temporais de frequência fundamental, *jitter* e *shimmer*, aplicando como classificador, análise discriminante com as funções linear (LDA) e quadrática (QDA). A avaliação de desempenho do classificador é realizada com as medidas empregadas de forma individual e de forma combinada. Espera-se que o método empregado possibilite o desenvolvimento de ferramentas baseadas em técnicas de processamento digital de sinais, para diagnóstico de patologias laríngeas de forma não invasiva. Os resultados obtidos demonstram a eficácia dos métodos empregados na discriminação entre vozes saudáveis e vozes afetadas por edema de Reinke e paralisia nas pregas vocais.

Palavras-chave: análise discriminante, classificação de vozes patológicas, frequência fundamental, *jitter*, *shimmer*

1. INTRODUÇÃO

Várias técnicas tem sido empregadas na análise acústica de sinais de vozes afetados por patologias na laringe. Algumas baseiam-se nas medidas da frequência fundamental (frequência de vibração das pregas vocais), como *jitter* (perturbações na frequência) e *shimmer* (perturbações na amplitude). Outras técnicas se utilizam do modelo linear de produção da fala, analisando-as em termos dos coeficientes de predição linear (coeficientes LPC) e da análise cepstral (GODINO-LLORENTE et al, 2006; DIBAZAR et al, 2006, VIEIRA et al, 2012).

As patologias na laringe podem ser de origem orgânica ou neurológica. No primeiro tipo encontram-se patologias como nódulos, pólipos, cistos, cânceres, edemas de Reinke, entre outras. Nas patologias do tipo neurológicas ou neuro-degenerativas situam-se as patologias causadas pelo mal de Parkinson, paralisias nas pregas vocais entre outras (GODINO-LLORENTE et al, 2006). No entanto, para este trabalho são focalizadas as patologias edemas de Reinke e paralisia nas pregas vocais.

A análise acústica usando técnicas de processamento digital de sinais para auxiliar diagnósticos de médicos especializados em doenças da laringe é um procedimento de medição simples, de baixo custo e não invasivo para os pacientes. Podendo ser empregada como ferramenta auxiliar em pré-diagnósticos e acompanhamento de tratamentos, diminuindo a necessidade do uso de exames mais invasivos como os de videolaringoscopia.

As alterações nas vozes desordenadas podem ser observadas tanto nas modificações da frequência fundamental quanto da envoltória espectral do sinal (GODINO-LLORENTE et al, 2006). O pré-diagnóstico fornece uma detecção precoce de patologias laríngeas e aumenta significativamente a eficácia do tratamento.

A chave para a modelagem acústica de vozes desordenadas é o entendimento das mudanças, referentes às medidas acústicas, produzidas pelos efeitos da fonte de excitação e do trato vocal (COSTA et al, 2008a). Para que se possa obter um bom modelamento acústico da patologia em estudo, as características ou parâmetros devem ser escolhidos de tal forma que representem o sinal, de forma a diferenciar um sinal saudável de um patológico, para uma discriminação eficaz da patologia.

Neste trabalho são empregadas as características temporais de frequência fundamental, bem como de duas medidas de perturbação obtidas a partir desta: *jitter* e *shimmer*. A classificação dos sinais em saudável ou patológico é feito por meio de um classificador baseado em análise discriminante linear ou quadrática. É realizada a análise de desempenho da classificação para verificar a viabilidade de implementação da técnica empregada em discriminar vozes patológicas, afetadas por paralisia ou edema nas dobras vocais de vozes saudáveis.

2. FUNDAMENTAÇÃO TEÓRICA

Patologias na laringe provocam mudanças significativas na voz causadas por uma modificação na morfologia de excitação, produzindo um padrão de vibração irregular, pois a distribuição de massa na dobra vocal e sua rigidez são maiores. A escolha de características que representem bem a desordem no sinal provocada pela patologia é de extrema importância. Portanto, a busca por características que melhor representem o sinal em análise é ainda um campo bastante promissor (FRAZÃO et al, 2011). Na Figura 1 estão representados trechos de 50 ms de sinais com as patologias edema de Reinke (Fig. 1a), paralisia nas pregas vocais (Fig. 1b) e para um sinal saudável (Fig. 1c) da base de dados utilizada no trabalho.

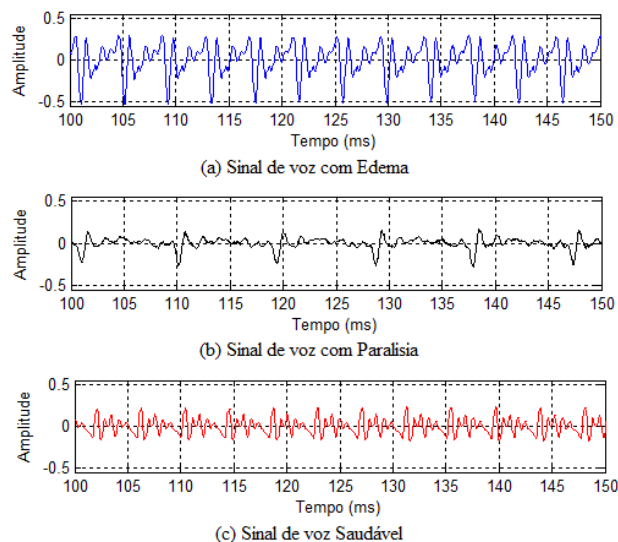


Figura 1 – Formas de onda para trechos de tipos de sinais: (a) Edema – JXF11AN.NSP, (b) Paralisia – JFN11AN.NSP e (c) Saudável – JEG1NAL.NSP.

Os tópicos a seguir apresentam uma descrição de alguns conceitos relacionados a algumas características temporais do sinal de voz, para avaliar a importância da utilização de técnicas de análise linear na descrição de desordens vocais, diagnósticos de patologias na laringe e avaliação de tratamento clínico.

2.1 Frequência Fundamental

A presença de patologia pode afetar a frequência fundamental (*pitch*), fazendo com que homens e mulheres produzam uma frequência excessivamente elevada ou baixa, respectivamente, ou seja, fora dos valores médios usuais para os gêneros masculino e feminino (COSTA, 2008).

A frequência fundamental, F_0 , corresponde à frequência do sinal de excitação proveniente da glote, ou seja, é o número de vibrações das pregas vocais por segundo (COSTA, 2008). Mais especificamente, corresponde ao componente periódico mais baixo (grave) do som produzido na glote (GOLDFIELD, 2000).



2.2 Variações na Frequência Fundamental

Há muitos sinais acústicos que podem ser associados a qualquer patologia, entre os quais estão o *jitter* e o *shimmer*. O *jitter* (perturbação de frequência) é um dos índices que reflete anomalias das dobras vocais e pode ser de fácil medição (PARRAGA, 2002).

Os falantes normais apresentam uma pequena perturbação, que pode representar uma variação de massa, tensão, atividade muscular ou atividade neural das dobras vocais. Do mesmo modo que para a frequência fundamental, é possível que a amplitude do som da prega vocal varie de um ciclo para o seguinte. Essa característica é denominada perturbação de amplitude ou *shimmer* (PARRAGA, 2002).

Os parâmetros de perturbação acústica para estimação das propriedades de variação de vozes patológicas foram medidos durante a fonação de vogais sustentadas. Esses parâmetros definem o grau de instabilidade de amplitude e *pitch* e indica o nível de componentes aperiódicos, determinado pela presença de ruído turbulento, modulações de frequência e amplitude de sinais de voz, devido às alterações nas propriedades dinâmicas das pregas vocais (ALMEIDA, 2009).

Jitt (*Jitter* percentual) (%) é uma avaliação relativa da variabilidade período-a-período do *pitch* (em curtíssima duração), para a amostra de voz analisada. Os intervalos de quebra da voz são excluídos. *Jitt* é calculado a partir dos valores de *pitch* extraídos período-a-período do sinal de voz, como:

$$Jitt = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |To^{(i)} - To^{(i+1)}|}{\frac{1}{N} \sum_{i=1}^N To^{(i)}} \quad (1)$$

Em que $To^{(i)}$ na Equação 1 é o período de *pitch* extraído, com $i=1,2,\dots, N$, e N = número de períodos de *pitch* extraídos (BRANDT, 2011).

Shimm (*Shimmer* percentual) (%) é a avaliação relativa da variabilidade período-a-período de curtíssima duração da amplitude pico-a-pico dentro da amostra de voz analisada. Os intervalos de quebra da voz são excluídos. *ShdB* é calculado a partir dos dados de amplitude pico-a-pico do sinal de voz, como:

$$Shimm = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A^{(i)} - A^{(i+1)}|}{\frac{1}{N} \sum_{i=1}^N A^{(i)}} \quad (2)$$

Em que $A^{(i)}$ na Equação 2, é o valor pico-a-pico extraído da amplitude, com $i=1,2,\dots, N$, e N = número de impulsos extraídos (BRANDT, 2011).

3. MATERIAL E MÉTODOS

Os sinais de voz analisados são oriundos de uma base de dados comercial, o *Massachusetts Eye and Ear Infirmary (MEEI) Disordered Voice Database* distribuído pela Kay Elemetrics (KAY ELEMETRICS, 1994a). Foram gravadas as vozes de 710 indivíduos (657 vozes patológicas e 53 vozes normais), contando com mais de 1400 amostras de voz. As amostras contêm de 1 a 3 segundos da vogal sustentada /a/. A escolha da vogal sustentada deve-se ao fato de que durante a emissão da vogal sustentada, as dobras vocais vibram, permitindo a observação de seu comportamento na presença da patologia (VIEIRA, 2011).

Essa base de dados, desenvolvida com o objetivo de auxiliar a análise perceptual de vozes desordenadas para aplicações clínicas e de pesquisa, tem sido amplamente empregada em trabalhos científicos internacionais, relacionados ao foco da pesquisa (VAZARI et al, 2008; COSTA et al, 2008; ZHANG et al, 2005; UMAPATHY et al, 2005; GODINO-LLORENTE, 2006). Para este trabalho são



selecionados os seguintes sinais: 53 arquivos de vozes saudáveis, além de 43 sinais com edema e 47 afetados por paralisia nas pregas vocais (VIEIRA et al, 2012).

Os valores das medidas temporais de frequência fundamental, *jitter* e *shimmer* foram adquiridos pelo *software* MULTI-SPEECH da Kay Elemetrics (KAY ELEMETRICS, 1994b). As medidas foram separadas pelo gênero e submetidas ao *software* Matlab, v.7.11. O classificador utilizado neste trabalho utiliza a função *classify* do Matlab, empregando a análise discriminante linear e quadrática.

A análise discriminante linear (LDA) é uma técnica comumente usada para a classificação de dados e redução de dimensionalidade (BALAKRISHNAMA and GANAPATHIRAJU). Essa técnica clássica em reconhecimento de padrões permite uma previsão direta do grupo que a variável pertence. Neste trabalho isso corresponde a classificar sinais de vozes como sendo patológicos ou saudáveis. Assim, a LDA classifica um objeto da amostra em uma de duas categorias baseadas em determinadas propriedades do objeto. Quando a regra de classificação assume que as variâncias das populações são iguais, as funções discriminantes são ditas lineares e quando não são funções discriminantes quadráticas (QDA). O objetivo da análise discriminante, com as funções linear e quadrática, inclui identificar a contribuição relativa das variáveis empregadas na classificação, para a separação dos grupos. Assim, é avaliado o desempenho dos classificadores com as características individuais e, posteriormente, é realizada uma combinação das medidas, com o propósito de melhorar o desempenho na classificação.

A frequência fundamental média no Português brasileiro está em torno de 105 Hz para os homens, 213 Hz para as mulheres. Logo a separação pelo gênero é importante para a eficiência dos classificadores. Foram utilizados 54 sinais de vozes masculinas, distribuídos em: 21 saudáveis, 11 com edemas de Reinke e 22 com paralisia nas pregas vocais. Para os femininos foram selecionados 89 sinais de vozes, sendo: 32 sinais saudáveis, 32 com edemas de Reinke e 25 com paralisia nas pregas vocais.

O processo de classificação dos sinais é realizado em duas fases: treinamento e teste/classificação (Figura 2). Na etapa de treinamento, o sinal de voz é pré-processado, o que consiste em segmentação do sinal para manter a estacionariedade (intervalos de 16 a 32 ms). Neste trabalho são utilizados segmentos de 20 ms. As características dos sinais são extraídas e alguns dos sinais são utilizados para treinamento do sistema (11 saudáveis e 18 patológicos masculinos; 22 saudáveis e 37 patológicos femininos), constituindo os padrões de referência. Os outros sinais (10 saudáveis e 14 patológicos masculinos; 10 saudáveis e 20 patológicos femininos) são utilizados na fase de teste. Na etapa de teste ou classificação, os sinais são também pré-processados e as características são extraídas e comparadas com os padrões de referência. São empregados classificadores de análise discriminante: linear (LDA) e quadrática (QDA).

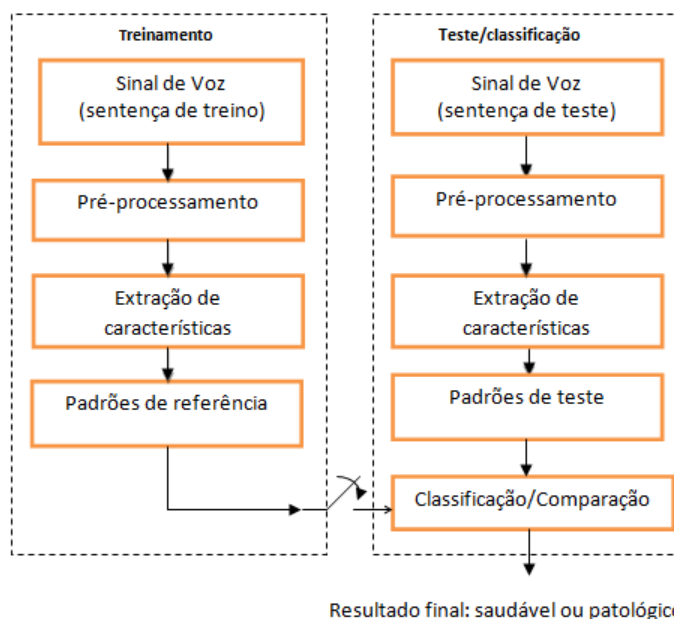


Figura 2 – Diagrama em Blocos do sistema de classificação de sinais.

Numa primeira classificação, os sinais são classificados como saudável ou patológico, sem especificar qual a patologia presente no sinal de voz. Dois classificadores distintos são empregados, um para vozes femininas e outro para vozes masculinas. O resultado final da classificação é obtido a partir da média entre os resultados individuais. Na sequência, é feita a classificação especificando a patologia: paralisia ou edema nas pregas vocais.

4. RESULTADOS E DISCUSSÃO

Para a avaliação de desempenho são empregadas as seguintes medidas: Correta rejeição (*CR*), que representa a detecção correta da ausência da patologia; Correta aceitação (*CA*), indica que a patologia está presente, e ela realmente está; Precisão, que consiste na taxa de classificação correta, ou seja, é o número de sinais classificados corretamente dividido pelo número total de sinais.

Nas Tabelas 1, 2 e 3 são apresentados os resultados obtidos para os classificadores individuais de análise discriminante linear (LDA) e quadrática (QDA). Os sinais são denominados como sinais patológicos (PTL), sinais afetados por edema de Reinke (EDM), paralisia (PRL) e sinais de vozes saudáveis (SDL). Observa-se que a medida temporal *Shimmer*, usando o classificador QDA apresentou os melhores resultados para as medidas de desempenho, chegando a obter uma precisão de 100% na classificação entre vozes saudáveis e vozes afetadas por paralisia.

Tabela 1 – Classificação individual SDLxPTL das medidas utilizadas.

Medidas	LDA			QDA		
	<i>CR</i> (%)	<i>CA</i> (%)	Precisão (%)	<i>CR</i> (%)	<i>CA</i> (%)	Precisão (%)
<i>F0</i>	60,0	85,0	73,33	100	58,57	74,17
<i>Jitter</i>	90,0	46,43	62,92	90,0	63,57	73,75
<i>Shimmer</i>	100	69,64	81,25	100	90,36	91,67

Tabela 2 – Classificação individual SDLxEDM das medidas utilizadas.

Medidas	LDA			QDA		
	CR (%)	CA (%)	Precisão (%)	CR (%)	CA (%)	Precisão (%)
<i>F0</i>	75,0	72,5	71,07	75,0	67,5	68,57
<i>Jitter</i>	95,0	42,5	76,79	90,0	42,5	71,43
<i>Shimmer</i>	100	57,5	86,79	100	70,0	90,36

Tabela 3 – Classificação individual SDLxPRL das medidas utilizadas.

Medidas	LDA			QDA		
	CR (%)	CA (%)	Precisão (%)	CR (%)	CA (%)	Precisão (%)
<i>F0</i>	90,0	35,0	62,5	95,0	45,0	70,0
<i>Jitter</i>	95,0	65,0	80,0	90,0	75,0	82,5
<i>Shimmer</i>	100	75,0	87,5	100	100	100

Para melhorar o desempenho na classificação entre as classes saudável e patológica e saudável e edema, as medidas foram combinadas duas a duas e três a três. Nas Tabelas 4 e 5 estão representados os resultados obtidos na classificação mediante a combinação duas a duas das medidas empregadas.

Tabela 4 – Classificação SDLxPTL – Medidas combinadas 2 a 2.

Medidas	LDA			QDA		
	CR (%)	CA (%)	Precisão (%)	CR (%)	CA (%)	Precisão (%)
<i>F0 e Jitter</i>	100	69,64	81,25	90,0	69,64	77,5
<i>F0 e Shimmer</i>	100	84,29	90,42	100	93,93	96,25
<i>Jitter e Shimmer</i>	100	73,57	84,17	90,0	81,79	85,0

Os melhores resultados obtidos através da combinação de duas medidas simultâneas, para a classificação entre sinais saudáveis e patológicos (Tabela 4), representam uma CR de 100%, CA de 93,93% e uma Precisão de 96,25%, conseguidos pela combinação das medidas entre *F0* e *Shimmer*, através do classificador QDA. A combinação aumentou em cerca de 4,58% a Precisão, comparada ao melhor resultado individual (*Shimmer*, 91,67%).



Tabela 5 – Classificação SDLxEDM – Medidas combinadas 2 a 2.

Medidas	LDA			QDA		
	CR (%)	CA (%)	Precisão (%)	CR (%)	CA (%)	Precisão (%)
<i>F0 e Jitter</i>	95,0	77,5	88,93	85,0	65,0	78,21
<i>F0 e Shimmer</i>	100	70,0	90,36	95,0	87,5	92,86
<i>Jitter e Shimmer</i>	100	70,0	90,36	90,0	70,0	84,29

Para a classificação entre os sinais saudáveis e com edema de Reinke (Tabela 5) as melhores medidas de desempenho correspondem a uma $CR = 95\%$, $CA = 87,5\%$ e uma $Precisão = 92,86\%$, conseguidos com o classificador QDA. Houve um aumento de aproximadamente 2,5%, em relação ao melhor resultado individual (*Shimmer*, 90,36%). Combinando as três medidas, no entanto, não foram apresentadas melhorias relevantes, em comparação aos demais casos.

5. CONCLUSÕES

Os resultados obtidos apontam a medida *Shimmer* como a que, individualmente, consegue melhor capturar as desordens vocais provocadas pelas patologias edema e paralisia. As medidas temporais frequência fundamental e *Shimmer*, quando combinadas, conseguem com maior eficiência discriminar vozes saudáveis de vozes patológicas, melhorando os resultados obtidos com as medidas individuais do conjunto selecionado.

Os resultados são bastante promissores e apontam para o uso dessas medidas em sistema de classificação de sinais como ferramenta auxiliar a pré-diagnósticos de patologias laríngeas. Assim, definição das características, empregadas no diagnóstico, que representem bem a desordem no sinal provocada por uma determinada patologia, é de extrema importância para o sucesso do diagnóstico.

REFERÊNCIAS

ALMEIDA, N.C. **Estudo da Análise Dinâmica Não-Linear na Discriminação de Vozes Patológicas**. 2009. 58f. Monografia (Tecnólogo em Sistemas de Telecomunicações) – Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, 2009.

BALAKRISHNAMA, S. and GANAPATHIRAJU, A. **Linear Discriminant Analysis - A Brief Tutorial**. Institute for Signal and Information Processing. Department of Electrical and Computer Engineering, Mississippi State University.

BRANDT, R. R. **Modelagem de Vozes Patológicas baseadas na Estimação Espectral do Ruído Glotal**. 2011. 88f. Dissertação (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande, Campina Grande, 2011.

COSTA S. C., CORREIA, S. E. N., FALCÃO, H. H., ALMEIDA, N. C., AGUIAR NETO, B. G. and FECHINE, J.M. **Pathological Voice Discrimination based on Entropy Measurements**. Proceedins of the 23rd Annual ACM Symposium on Applies Computing, pp. 1410-1414, 2008a.

COSTA, S. C. **Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas**. Tese de Doutorado. Universidade Federal de Campina Grande, Campina Grande, 2008.



COSTA, S. C., CORREIA, S. E. N., COSTA, W. C. A. **Análise Dinâmica Não Linear Aplicada ao Diagnóstico de Patologias Orgânicas e Neurológicas da Laringe.** In: VI Congresso de Pesquisa e Inovação da Rede Norte e Nordeste de Educação Tecnológica – VI CONNEPI, 2011. Anais do VI CONNEPI, 2011.

DIBAZAR, A. A., BERGER, T.W., and NARAYANAN, S. S. **Pathological Voice Assessment.** Proceedings of the 28th IEEE EMBS Annual International Conference, New York, USA, pp. 1669-1673, August, 2006.

FRAZÃO, M. C. L. M., COSTA, S. C., CORREIA S. E. N., COSTA, W. C. A. **Análise Acústica de Sinais de Vozes por Meio de Medidas Temporais.** In: VI Congresso de Pesquisa e Inovação da Rede Norte e Nordeste de Educação Tecnológica – VI CONNEPI, 2011. Anais do VI CONNEPI, 2011.

GODINO-LLORENTE, J. I., GÓMEZ-VILDA, P., BLANCO VELASCO, M. **Dimensionality Reduction of a Pathological 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.

GOLDFELD, M. **A criança surda – linguagem e cognição numa perspectiva sóciointeracionista.** São Paulo: Plexus, 1997.

KAY ELEMETRICS CORP. **Disordered Voice Database**, model 4337, 03 Ed, 1994a.

KAY ELEMETRICS, Kay Elemetrics Corp. **Multi-Speech**, model 3700, 1994b.

PARRAGA, A. **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas.** Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, 2002.

UMAPATHY, K., KRISHNAN, S., PARSAR, V. and JAMIESON, D. G. **Discrimination of Pathological Voices Using a Time-Frequency Approach.** IEEE Trans. on Biomedical Engineering, Vol. 52., No. 3, Março, 2005.

VAZIRI, G., ALMASGANJ, F. and JENABI, M. S. **On the Fractal Self-Similarity of Laryngeal Pathologies Detection: The estimation of Hurst parameter,** Proceedings of the 5th International Conference on Information Technology and Application in Biomedicine, p. 383-386, 2008.

VIEIRA, R. T. **Processamento Digital de Sinais de Voz para Detecção de Patologias na Laringe.** 2011. 66f. Monografia (Bacharel em Engenharia Elétrica) – Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, 2011.

VIEIRA, R. T., MONTEIRO, N. A. B., COSTA, S. C., CORREIA, S. E. N., AGUIAR NETO, B. G., FECHINE, J. M. **Combining Entropy Measurements and Cepstral Analysis for Pathological Voice Assessment.** Journal of Medical and Biological Engineering, 2012. DOI: <<http://jmbe.bme.ncku.edu.tw/AIP/AIP-JMBE%20928.pdf>>.

ZHANG, Y., JIANG, J. J., BIAZZO L. and JORGENSEN, M. **Perturbation and Nonlinear Dynamic Analyses of Voices from Patients with Unilateral Laryngeal Paralysis,** Journal of Voice, Vol. 19 (4), pp. 519-528, Dezembro, 2005.