

# ANÁLISE CEPSTRAL USANDO MÚLTIPLOS CLASSIFICADORES PARA DETECÇÃO DE VOZES PATOLÓGICAS

**Raissa TAVARES (1); Nathália BRUNET (1); Suzete CORREIA (1); Silvana COSTA (1);  
Benedito AGUIAR (2); Joseana FECHINE (2)**

(1) Instituto Federal de Educação Ciência e Tecnologia da Paraíba, Av. 1º de Maio 720 Jaguaribe CEP:58.015-430 João Pessoa - PB, e-mail: [raissa@ieee.org](mailto:raissa@ieee.org), [nathaliabrunet@gmail.com](mailto:nathaliabrunet@gmail.com), [suzete@ifpb.edu.br](mailto:suzete@ifpb.edu.br), [silvana@ifpb.edu.br](mailto:silvana@ifpb.edu.br)

(2) Universidade Federal de Campina Grande, Av. Aprígio Veloso 882 Bodocongó CEP 58.109-970 Campina Grande -PB, e-mail: [bganeto@dee.ufcg.edu.br](mailto:bganeto@dee.ufcg.edu.br), [joseana@dsc.ufcg.edu.br](mailto:joseana@dsc.ufcg.edu.br)

## RESUMO

Este artigo trata da avaliação de desempenho de classificadores baseados em análise cepstral aplicados na detecção de vozes patológicas. O objetivo é analisar o desempenho individual e da combinação de classificadores baseados em análise cepstral, por meio dos coeficientes cepstral, cepstral ponderado, delta cepstral, e delta cepstral ponderados. Duas estratégias de combinações diferentes são empregadas, produzindo um classificador múltiplo que é mais eficiente do que qualquer um dos parâmetros cepstrais aplicados individualmente. As taxas de eficiência obtidas variam de 87% usando o classificador individual delta cepstral ponderado para 98% considerando a combinação de classificadores.

**Palavras-chave:** análise de sinais, vozes patológicas, análise cepstral, classificadores múltiplos

## 1 INTRODUÇÃO

Patologias na laringe como nódulos, pólipos, cistos e edemas podem provocar modificações significativas na produção natural da voz. Sabe-se que a maioria dessas doenças causa alterações acústicas no sinal de voz, transformando-o em uma ferramenta útil para diagnosticá-las. São utilizados testes como a escuta de voz do paciente e inspeção direta da laringe para diagnósticos das patologias. A primeira técnica é subjetiva e depende de experiência do profissional, podendo levar a diferentes resultados. A segunda técnica é mais exata e objetiva. No entanto causa desconforto ao paciente, além de exigir equipamentos mais sofisticados. Técnicas de processamento digital de sinais surgem como uma ferramenta que propicia um exame menos desconfortável ao paciente e de fácil utilização, que detecta automaticamente as desordens vocais provocadas por patologias da laringe. Pode também ser utilizado para a avaliação dos tratamentos cirúrgicos e farmacológicos e processos de reabilitação (GODINO-LLORENTE et al. 2006).

Diversos são os métodos utilizados e diferentes níveis de sucesso são obtidos na separação entre vozes saudáveis e patológicas. A seleção das características acústicas que diferenciem uma voz patológica de uma voz normal, ou que seja capaz de fazer a distinção entre diversas patologias é uma etapa importante para um bom desempenho de um sistema de discriminação de vozes patológicas (SHAMA et al, 2007; MURPHY and AKANDE, 2007; DIBAZAR et al, 2006; UMAPATHY et al, 2005; BAHOURA and PELLETIER, 2004).

Na literatura há um grande número de parâmetros acústicos que podem ser extraídos e analisados, mas não é absolutamente clara a sua utilidade para resolver o problema. A seleção dos parâmetros mais adequados é ainda uma questão em aberto. O objetivo deste trabalho é aplicar uma solução baseada na análise cepstral para discriminar vozes patológicas afetadas por edema nas dobras vocais. Os parâmetros cepstral (CEP), cepstral ponderado (CEPP), delta cepstral (DCEP), e delta cepstral ponderado (DCEPP) são usados como recursos para detectar as irregularidades das vozes patológicas, em comparação com a voz normal. A técnica de Quantização Vetorial foi utilizada em cada parâmetro associada à medida de distorção Euclidiana, para o processo de classificação do sinal de fala. O quantizador vetorial foi treinado com vozes afetadas pela patologia considerada e os resultados serão utilizados para construir um método eficaz para a detecção de vozes patológicas.

Para melhorar o desempenho dos classificadores cepstrais, uma abordagem baseada em múltiplos classificadores é avaliada. Para isso, duas regras de combinação são consideradas: a combinação da média

aritmética e a combinação do produto, que são modificações das estratégias utilizadas em Oliveira Jr. et al. (2004).

## 2 FUNDAMENTAÇÃO TEÓRICA

Considerando que o sinal de voz (Eq. 01) é o resultado da convolução da excitação ( $e(n)$ ) com a resposta do trato vocal ( $\theta(n)$ ), através da análise cepstral, é possível separar estas duas componentes. Todas as características cepstrais relacionadas descritas são obtidas através da Codificação por Predição Linear (análise LPC – *Linear Predictive Coding*) (RABINER et al. 1978; O'SHAUGHNESSY, 2000).

$$s(n) = e(n) * \theta(n) \quad [\text{Eq. 01}]$$

O método de predição linear estima cada amostra de voz baseado numa combinação linear de  $p$  amostras anteriores; um  $p$  maior permite um modelo mais preciso. Esta análise fornece um conjunto de parâmetros da fala que representa o trato vocal (RABINER et al. 1978). Espera-se que, qualquer mudança na estrutura anatômica do trato vocal, devido à patologia, afete os coeficientes LPC como também os cepstrais e seus derivados. Um preditor linear com coeficientes de predição,  $\alpha(k)$ , é definido como um sistema cuja saída é:

$$\tilde{s}(n) = \sum_{k=1}^p \alpha(k)s(n-k), \quad [\text{Eq. 02}]$$

onde  $p$  é a ordem do preditor. O método de autocorrelação e o da covariância são dois métodos padrões para o cálculo dos coeficientes do preditor. Os coeficientes de predição são calculados usando o algoritmo de Levinson-Durbin (RABINER et al. 1978).

Os cepstrais derivados podem melhorar a representação das propriedades espectrais da fala. Vozes patológicas apresentam significativas diferenças espectrais quando comparadas com vozes normais. Os coeficientes cepstrais podem ser calculados recursivamente a partir dos coeficientes de predição linear,  $\alpha(k)$ , por meio de (TOHKURA, 1987):

$$\begin{cases} c(1) = -\alpha(1) \\ c_i(n) = -\alpha(n) - \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \alpha(j) c(n-j) \end{cases} \quad 1 < n \leq p \quad [\text{Eq. 03}]$$

Os coeficientes cepstrais obtidos pela Eq. 03 fornecem uma boa medida das diferenças na envoltória espectral dos segmentos de voz em análise. Estes coeficientes são utilizados para observar as informações das transições do sinal de voz patológico versus sinal de voz normal (MAMMONE et al. 1996).

A primeira derivada dos coeficientes cepstrais (Coeficientes Delta Cepstrais) é dada por (MAMMONE et al. 1996)

$$\frac{\Delta c(n, t)}{\Delta t} = \Delta c_i(n) \approx \phi \sum_{k=-K}^K k c(n, t+k), \quad [\text{Eq. 04}]$$

onde  $c(n, t)$  é o  $n$ -ésimo coeficiente da predição linear no tempo  $t$ ,  $\phi$  é uma constante de normalização e  $2K+1$  é o número de quadros sobre os quais o cálculo é realizado.

Os coeficientes delta cepstrais são obtidos como uma versão simplificada da (Eq. 04), a partir de:

$$\Delta c_i(n) = \left[ \sum_{q=-K}^K k c_{i-q}(n) \right] G, \quad 1 \leq n \leq p, \quad [\text{Eq. 05}]$$

onde  $G$  é um termo de ganho (por exemplo: 0,375),  $p$  é o número dos coeficientes delta cepstrais,  $K = 2$ ,  $n$  representa o índice de coeficiente e  $i$  o quadro de análise (FECHINE, 2000).

Com o objetivo de minimizar a sensibilidade dos coeficientes cepstrais de baixa ordem em relação à envoltória espectral e à sensibilidade dos coeficientes cepstrais de alta ordem em relação ao ruído, é empregada a ponderação cepstral, também conhecida como *liftering* ou suavização.

Os coeficientes cepstrais ponderados (CEPP),  $cw_i(n)$ , são obtidos por (MAMMONE et al. 1996; FECHINE, 2000):

$$cw_i(n) = c_i(n) \cdot w(n). \quad [\text{Eq. 06}]$$

O tipo de janela utilizada neste trabalho foi a passa-faixa (BPL – *Bandpass Liftering*), dada por (MAMMONE et al 1996).

$$w(n) = \begin{cases} 1 + \frac{L}{2} \sin\left(\frac{n\pi}{L}\right), & n = 1, 2, \dots, L \\ 0, & \text{caso contrário.} \end{cases}, \quad [\text{Eq. 07}]$$

onde  $L$  é o tamanho da janela. A técnica baseada em BPL pondera uma sequência cepstral pela janela da (Eq. 07) tal que as componentes de ordem mais alta e de ordem mais baixa são de-enfatizadas.

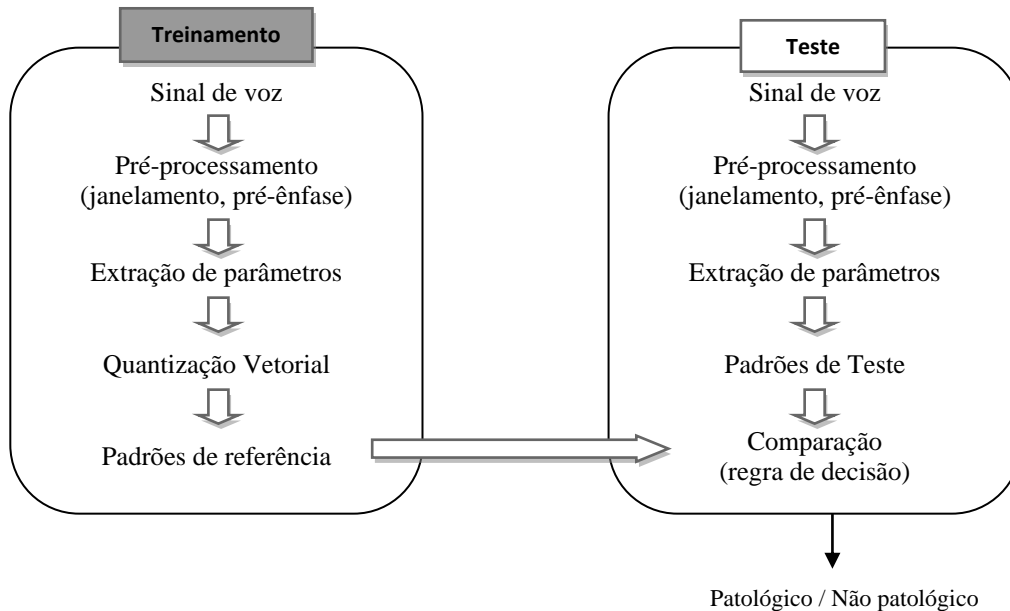
Os coeficientes delta-cepstrais ponderados (DCEPP) são obtidos a partir de (Eq. 05) e (Eq. 06), associando as características dos coeficientes cepstrais ponderados com os delta-cepstrais, resultando em:

$$\Delta cw_i(n) = \Delta c_i(n) \cdot w(n). \quad [\text{Eq. 08}]$$

### 3 BASE DE DADOS E METODOLOGIA

O banco de dados usado foi gravado pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* construído para aplicações clínicas e de pesquisa (KAY ELEMENTRICS, 1994). Os casos selecionados, a partir da gravação da vogal sustentada /a/ foram: 44 pacientes apresentando edema nas dobras vocais - 33 mulheres (17-85 anos) e 11 homens (23-63 anos), a maioria deles (32) com edema bilateral e 53 pacientes com voz normal, composto por 21 homens (26-59 anos), e 32 mulheres (22-52 anos). Durante a emissão da vogal sustentada, as dobras vocais vibram, permitindo analisar a influência da patologia no padrão de vibração das mesmas.

Na Figura 1 é ilustrado o processo de discriminação de vozes patológicas empregando classificadores individuais. O processo é realizado em duas fases principais: treinamento e teste. Na fase de treinamento os sinais são pré-processados, etapa em que os sinais de voz são segmentados em 20 ms com janelamento de Hamming (sobreposição de 50%). Um filtro de pré-ênfase (0.95) também é utilizado. A etapa de extração de parâmetros consiste na obtenção dos coeficientes cepstrais de ordem  $p$  (neste caso,  $p=12$ ), para cada segmento.

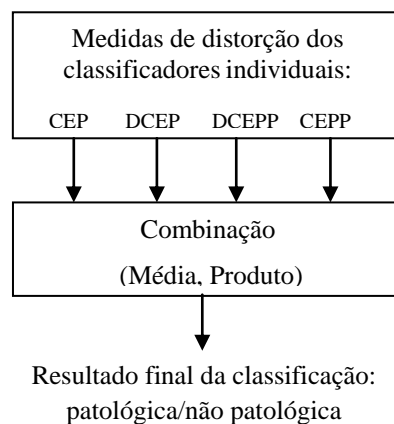


**Figura 1 – Processo de Discriminação entre vozes normais e vozes patológicas.**

Para a redução da dimensionalidade dos dados, uma técnica de Quantização Vetorial (QV) é aplicada aos parâmetros cepstrais, associada a medida de distância Euclidiana (MAKHOUL et al.1985). A quantização é feita individualmente para cada característica usando apenas vozes com edema nas dobras vocais na fase de treinamento. Nesta fase, são usados 50% de casos de edema vocal. Para a fase de teste são usados os outros 50% de vozes com edema nas dobras vocais, e todos os 53 de vozes normais. Após a extração de características, um dicionário é gerado a partir da medição de distância euclidiana e da regra do vizinho mais próximo, empregando-se o algoritmo LBG (LINDE et al. 1980).

Uma pré-classificação é feita, para cada parâmetro e depois os parâmetros são combinados empregando as regras da média e do produto, visando melhorar o desempenho dos classificadores individuais (OLIVEIRA Jr.et al,2004).

Na Figura 2 está representado o método de classificadores múltiplos aplicado neste trabalho.



**Figura 2 – Múltiplos classificadores baseado em análise cepstral – regras da média e do produto.**

#### 4 RESULTADOS E DISCUSSÃO

As medidas utilizadas para avaliar o desempenho dos métodos são as seguintes (GODINO-LLORENTE et al. 2006): Correta aceitação (CA) - é detectada corretamente a presença da patologia; Correta rejeição (CR) - é detectada corretamente a ausência da patologia; Falsa aceitação (FA) - é detectada erroneamente a presença da patologia; Falsa rejeição (FR) - a presença da patologia é rejeita quando ela está presente; e Eficiência (E) - representa a correta classificação de uma determinada classe quando está presente, dada por:

$$E = \frac{CA + CR}{CA + CR + FA + FR} \times 100 \quad [\text{Eq. 09}]$$

A Tabela 1 mostra os resultados obtidos para cada parâmetro individualmente. Pode ser visto que os parâmetros Delta Cepstrais (DCEP) apresentam a melhor eficiência e as menores taxas de falsa aceitação. No entanto, este método apresenta uma maior taxa de falsa rejeição em relação ao método cepstral (CEP).

**Tabela 1 – Avaliação de desempenho das características individuais**

Classificadores	CR (%)	CA (%)	FA (%)	FR (%)	E (%)
CEP	89	91	11	9	90
CEPP	94	86	6	14	90
<b>DCEP</b>	<b>98</b>	<b>86</b>	<b>2</b>	<b>14</b>	<b>92</b>
DCEPP	91	82	9	18	87

Para avaliar a combinação das características, faz-se a suposição de que a um sinal de voz devem ser atribuídos uma das  $K$  classes possíveis e assumir que os  $L$  classificadores estão disponíveis. A medida de distância utilizada pelo  $i$ -ésimo classificador é denominado  $d_i$ . Duas regras de combinação foram empregadas:

- Combinação da média aritmética:

$$D = \frac{1}{K} \sum_{i=1}^L d_i, \quad [\text{Eq. 10}]$$

- Combinação do produto:

$$D = \prod_{i=1}^L d_i, \quad [\text{Eq. 11}]$$

onde  $D$  é a distorção obtida após a combinação.

Para garantir uma padronização das saídas dos classificadores, os valores de distorção obtidos em cada classificador individual foram normalizados (valores entre 0 e 1). É aplicado um limiar de distorção, escolhido de tal forma que proporcione a melhor separação entre as classes.

Os resultados da combinação da média e do produto são apresentados nas Tabelas 2 e 3, respectivamente.

**Tabela 2 – Avaliação de Desempenho – Regra da média**

Classificadores	CR (%)	CA (%)	FA (%)	FR (%)	E (%)
CEP e DCEP	92	95	8	5	94
CEP e DCEPP	94	95	6	5	95
CEP e CEPP	96	95	4	5	96
DCEP e DCEPP	98	82	2	18	90
DCEP e CEPP	94	93	6	7	94
DCEPP e CEPP	92	95	8	5	94
CEP, DCEP e DCEPP	94	91	6	9	93
CEP, DCEP e CEPP	96	95	4	5	96
<b>CEP, DCEPP e CEPP</b>	<b>96</b>	<b>98</b>	<b>4</b>	<b>2</b>	<b>97</b>
DCEP, DCEPP e CEPP	94	93	6	7	94
CEP, DCEP, DCEPP e CEPP	94	98	6	2	96

Os resultados mostram que, para quase todas as combinações, as taxas de falsa rejeição diminuem significativamente. No caso individual DCEP (a melhor eficiência), a taxa de falsa rejeição foi de 14%. Quando foi combinado CEP, DCEPP e CEPP na regra da média (Tabela 2) a taxa de falsa rejeição caiu para 2%. Para esta combinação obteve-se 97% na regra da média, enquanto obteve-se 92% no melhor resultado individual (DCEP).

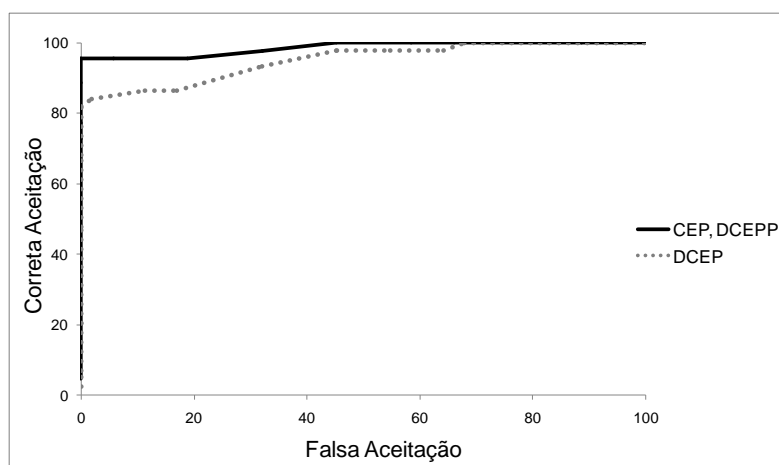
**Tabela 3 – Avaliação de Desempenho – Regra do Produto**

Classificadores	CR (%)	CA (%)	FA (%)	FR (%)	E (%)
CEP, DCEP	100	93	0	7	97
<b>CEP, DCEPP</b>	<b>100</b>	<b>95</b>	<b>0</b>	<b>5</b>	<b>98</b>
CEP, CEPP	92	98	8	2	95
DCEP, DCEPP	96	86	4	14	91
DCEP, CEPP	94	95	6	5	95
DCEPP, CEPP	92	98	8	2	95
CEP, DCEP e DCEPP	94	95	6	5	95
CEP, DCEP e CEPP	94	95	6	5	95
CEP, DCEPP e CEPP	94	98	6	2	96
DCEP, DCEPP e CEPP	94	93	6	7	94
CEP, DCEP, DCEPP e CEPP	74	98	26	2	86

O melhor resultado é obtido usando uma combinação do produto entre os classificadores CEP e DCEPP (Tabela 3). Pode-se observar uma melhora de pelo menos 6% na taxa de eficiência, em comparação com o classificador DCEP individual. Nesse caso, a probabilidade de detectar a presença da patologia edema quando, na realidade, ela não está presente (FA), é nula.

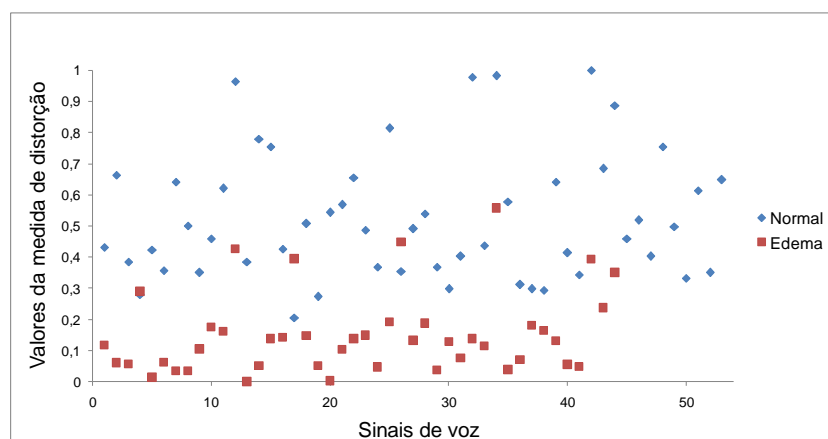
Na Figura 3 estão ilustradas as curvas ROC (*Receive Operator Characteristic*) para o melhor desempenho do classificador individual DCEP e para o melhor desempenho da combinação dos classificadores CEP, DCEPP obtida pela regra do produto. As curvas ROC são comumente empregadas para avaliação de desempenho. A área sob a curva mede a discriminação, ou seja, a capacidade do teste em classificar corretamente os sinais com ou sem a patologia. Quanto maior a área sob a curva, melhor o desempenho do método empregado.

Comparando-se a curva do desempenho individual com a da combinação, observa-se que na segunda, a curva alcança a extremidade mais superior e esquerda do gráfico, aproximando-se do resultado ideal. Isso confirma que a combinação propicia melhores taxas de Correta Aceitação (%) e Falsa Aceitação (%).



**Figura 3. Curva ROC para a combinação CEP/DCEPP e para DCEP individualmente.**

Nas Figuras 4 e 5 são mostradas as distribuições dos sinais de vozes normais e vozes afetadas por edemas nas dobras vocais, de acordo com a medida de distorção, aplicada após a quantização vetorial dos parâmetros cepstrais na avaliação individual e na combinação, respectivamente. Observa-se que após a combinação é clara a melhoria na separação entre as classes normal e patológica (Fig. 5).



**Figura 4 - Distribuição dos sinais de voz - avaliação individual (DCEP)**

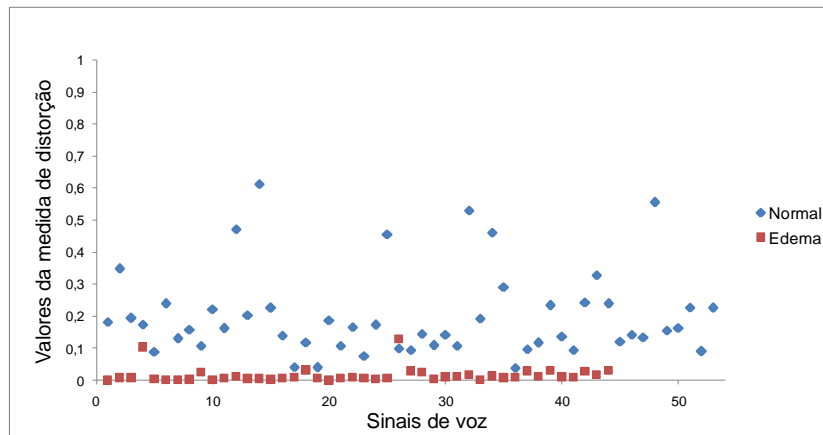


Figura 5 - Distribuição dos sinais de voz - avaliação da combinação (CEP e DCEPP)

## 5 CONCLUSÕES

Neste trabalho foram avaliados o desempenho individual e da combinação de classificadores baseados nos parâmetros cepstral, cepstral ponderado, delta cepstral e delta cepstral ponderado para a detecção de vozes patológicas. Os resultados mostram que a combinação destes classificadores pode produzir uma melhoria significativa no desempenho comparando com os parâmetros individuais. A melhor taxa de eficiência no caso individual foi de 92% e após as combinações, cerca de 98%. Isto significa que os parâmetros utilizados são complementares e podem ser aplicados para detectar distúrbios vocais causados pela presença de patologias nas dobras vocais.

## REFERÊNCIAS

- BAHOURA, M. and PELLETIER, C. **Respiratory Sounds Classification using Analysis and Gaussian Mixture Models**. Proceedings of the 26th Annual Conference of the IEEE EMBS, September, 2004.
- DIBAZAR, A., BERGER, T.W., and NARAYANAN S. S., **Pathological Voice Assessment**. Proc. of the 28th IEEE EMBS Annual International Conference, NY, USA, Aug., 2006.
- FECHINE, J. M. **Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística**. Tese de doutorado, Engenharia Elétrica. Universidade Federal de campina Grande, Paraíba, Brasil, 2000.
- GODINO-LLORENTE, J. I., GOMES-VILDA P. and BLANCO-VELASCO, M. **Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters**. IEEE Trans. on Biom. Engineering, Vol. 53, No. 10, pp. 1943-1953, October, 2006.
- KAY ELEMETRICS, Kay Elemetrics Corp. **Disordered Voice Database**. Model 4337, 03 Ed, 1994.
- LINDE, Y., BUZO, A., and GRAY, R. M., **An Algorithm for Vector Quantizer Design**. IEEE Transaction on Communications, Vol. COM-28, No. 1, pp 84-95, January, 1980.
- MAKHOUL, J., ROUCOS, S., and GISH, H. **Vector Quantization in Speech Coding**. Proceedings of the IEEE, Vol. 73, No. 11, November, pp. 1551-1588, 1985.
- MAMMONE, J. R., ZHANG, X., and RAMACHANDRAN, R. P. **Robust Speaker Recognition - A Feature-Based Approach**. IEEE Signal Processing Magazine, Vol. 13, No. 5, pages 58-71, September, 1996.

MURPHY, P. J. and AKANDE, O. O. **Noise Estimation in Voice Signals Using Short-term Cepstral.** Journal of the Acoust. Society of America, pp. 1679-1690, Vol. 121, No. 3, March, 2007.

O'SHAUGHNESSY, D. **Speech Communications: Human and Machine**, 2nd Edition, NY, IEEE Press, 2000.

OLIVEIRA JÚNIOR J. J. de, KAPP, M. N., FREITAS, C. O. A., CARVALHO, J. M., SABOURIN, R. **Handwritten Recognition with Multiple Classifiers for Restricted Lexicon.** Proceedings of the 17th Brazilian Symposium on Computer Graphics and Image Processing, v. 1. p. 82-89, 2004.

RABINER, L. R. and SCHAFER, R. W. **Digital Processing of Speech Signals.** New Jersey: Prentice-Hall, 1978.

SHAMA, K., KRISHNA, A., and CHOLAYYA, N. U. **Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology.** EURASIP Journal on Advances in Signal Processing, Vol. 2007, 2007.

TOHKURA, Y. **A Weighted Cepstral Distance Measure for Speech Recognition.** IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 35, No. 10, pp.1414-1422, October, 1987.

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