

Auxílio ao Diagnóstico de Doenças Laríngeas utilizando Análise Cepstral e DD-HMM

R. D. R. Fagundes¹, I. C. Zwetsch¹, D. Scolari¹

¹IDEIA-PUCRS, Prédio 30, sala 301-03, Av. Ipiranga 6681, Porto Alegre, RS, 90619-900, Brasil

Resumo: O presente trabalho apresenta técnicas de DSP (Digital Signal Processing) para análise de vozes com doenças laríngeas, para obter-se um método computacional eficiente de identificação dos distúrbios da laringe através das alterações na análise cepstral e do reconhecimento via DD-HMM (discrete density Hidden Markov Model). Esta metodologia de análise e diagnóstico laríngeo obteve resultados com precisão superior a 80% na avaliação de casos reais.

Palavras-Chave: Doenças laríngeas, análise cepstral, DD-HMM.

Abstract: This work presents DSP's (Digital Signal Processing) techniques to larynx diseases voices analysis, to obtain an efficient computational method for recognition of this diseases through cepstrum analysis and DD-HMM (discrete density Hidden Markov Model). This method of larynx diseases diagnostics had superior to 80% precision results with real cases.

Keywords: Larynx diseases, cepstrum analysis, DD-HMM.

I. Introdução

Este trabalho apresenta um modelo de análise da voz, com uso da técnica cepstral, como método de caracterização e diagnóstico de distúrbios laríngeos. O diagnóstico destas alterações é atualmente realizado principalmente pelo exame de videolaringoscopia. Certas doenças, mesmo para médicos especialistas experientes, podem trazer dificuldade diagnóstica, pois às vezes são muito parecidas no aspecto, apesar de apresentarem origens e alterações fisiopatológicas diferentes.

Estas dificuldades também são encontradas na realização de técnicas computacionais de processamento de sinais que, em determinados casos, não são eficientes o suficiente para a diferenciação das alterações.

No presente trabalho, aplicou-se a análise Cepstral com DD-HMM em casos normais e nas seguintes doenças: nódulo vocal, cisto vocal, pólipos vocais, edema de Reinke e sulco vocal.

Estas representam a grande maioria dos atendimentos de pacientes com alteração da voz, que não sejam as alterações transitórias por infecções das vias respiratórias, onde temos em alguns dias a melhora do quadro geral.

Na maioria dos casos, as doenças citadas produzem uma rouquidão com características típicas de cada uma, principalmente quando analisadas por ouvintes mais experientes, tais como médicos otorrinolaringologistas ou profissionais da área da fonoaudiologia.

Várias técnicas de análise de sinais da voz são estudadas para a identificação de alterações da laringe [1][2][3][4][5][6][7][8].

A proposta deste trabalho é utilizar a análise cepstral com DD-HMM como método de análise das alterações acústicas da voz nas doenças da laringe.

A análise cepstral do sinal de voz para o estudo de distúrbios da laringe é muito útil, permitindo trabalhar com o sinal da glote (excitação) separadamente das repercussões ressonantes do trato vocal, facilitando o entendimento das alterações que as doenças causam nas pregas vocais.

Esta técnica de DSP no estudo do sinal acústico da emissão da voz permitirá detectar modificações nas ondas que se relacionem com as doenças e, conseqüentemente, a criação de modelos para uma classificação, permitindo a obtenção de ferramenta de diagnóstico não-invasiva.

É utilizado o sistema DD-HMM de reconhecimento de padrões para distinguir as diferenças cepstrais de cada doença, e com isto, têm-se um sistema totalmente automático, não-invasivo de diagnóstico de alterações laríngeas.

II. Fundamentação Teórica

A. Análise Cepstral

A análise cepstral do sinal de voz permite trabalhar com o sinal da glote (excitação) e do trato vocal (ressonância) separadamente, pelas suas propriedades homomórficas, facilitando o estudo das alterações que as doenças causam nas pregas vocais. Temos a separação das características do filtro do trato vocal da seqüência de excitação.

O modelo mostrado na figura 1, é o mais freqüentemente usado, onde se assume que o sinal de voz, $s(t)$ é composto por um sinal de excitação $e(t)$ aplicado ao filtro do trato vocal, com uma resposta impulsional $v(t)$.

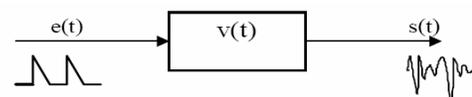


Fig.1: Modelo simplificado da produção da voz no domínio do tempo.

Onde $s(t)$ é a convolução de $e(t)$ com $v(t)$, que é definida por:

$$s(t) = e(t) \otimes v(t) \quad (1)$$

No domínio da freqüência a equação é definida como:

$$S(\omega) = E(\omega) \times V(\omega) \quad (2)$$

Em que $S(\omega)$, $E(\omega)$ e $V(\omega)$ são as transformadas de Fourier das funções contínuas no tempo $s(t)$, $e(t)$ e $v(t)$ ou as transformadas discretas de Fourier das seqüências de amostras temporais $s(n)$, $e(n)$ e $v(n)$. Assim, para executarmos uma análise cepstral do sinal de voz, o sinal $s(t)$ será processado via

Transformada de Fourier (usualmente uma FFT) resultando o espectro de frequência $S(\omega)$

Contudo, o sinal de excitação $E(\omega)$ e o filtro do trato vocal $V(\omega)$ não podem ser diretamente identificados no sinal de voz $S(\omega)$ resultante, visto que o sinal de voz é a resposta em frequência do trato vocal (modelado algebricamente como um filtro) pela função excitação $E(\omega)$. Como se deseja determinar as alterações laríngeas a partir dos efeitos analisados no sinal de voz, será necessário dissociar os efeitos da excitação e os feitos do trato vocal diretamente do sinal $S(\omega)$. Neste sentido, a análise cepstral, descrita a seguir, propiciará esta dissociação:

Lembrando a propriedade matemática dos logaritmos:

$$\log(a \times b) = \log(a) + \log(b) \quad (3)$$

Assim, tomando o logaritmo de $S(\omega)$ e aplicando (2):

$$\log(S(\omega)) = \log(E(\omega) \times V(\omega)) \quad (4)$$

e, a seguir (3):

$$\log(E(\omega).V(\omega)) = \log(E(\omega)) + \log(V(\omega)) \quad (5)$$

Na expressão (5) o sinal de voz $S(\omega)$ está sendo em sua forma logarítmica. No entanto, as componentes da excitação e do trato ainda são indistinguíveis. Então, lembrando que:

$$\mathfrak{S}^{-1}(f + g) = \mathfrak{S}^{-1}(f) + \mathfrak{S}(g)^{-1} \quad (6)$$

A propriedade aditiva do espectro logarítmico continua a se verificar quando lhe for efetuada a transformada inversa de Fourier, pela aplicação de (5) em (6), sendo o resultado dessa operação chamada de função cepstral ou cepstro.

O processo resumido de estimação do cepstro pode ser visto no diagrama de blocos apresentado na figura 2:

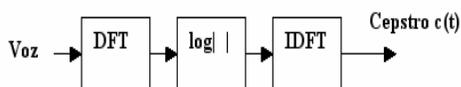


Fig.2: Diagrama de blocos da estimação do cepstro.

A concentração da componente periódica no espectro logarítmico de um sinal de voz num intervalo de frequências equivalente ao inverso do período fundamental T, aparece no Cepstro como um pico.

O eixo horizontal, da função cepstral tem dimensões temporais e o nome de quefrências. Com isto, na voz se obtém uma clara distinção entre a componente de excitação e a

contribuição do trato vocal, que aparece como um aglomerado de componentes aos baixos valores de quefrência afastado da componente do período fundamental que aparece em valores mais altos de quefrência.

Na figura 3 vê-se o Cepstro de um segmento de voz onde o pico correspondente ao período fundamental está próximo da quefrência de 10 ms, separado das componentes do trato vocal às de baixas quefrências. Nesta figura são apresenta das apenas as componentes do Cepstro superiores a sensivelmente 1 ms, pois as componentes de mais baixas quefrências têm valores comparativamente muito superiores aos restantes e a sua apresentação não deixaria claro o pico correspondente à frequência fundamental [9] [10] [11] [12] [13] [14].

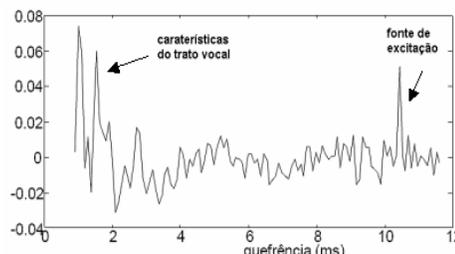


Fig.3: Cepstro de um segmento de fala.

B. Liftering

A função de transferência do trato vocal e a função de excitação da voz aparecem em partes separadas da escala de quefrências, e podem ser separadas em duas funções, pelo processo de “lifteragem”, facilitando o estudo individualizado das alterações na excitação e da parte ressonantal, como apresentado na figura 3.

C. Quantização Vetorial

Na quantização vetorial por amplitude, a idéia básica é indexar um vetor em um código numérico, como mostrado na figura (5) [1].

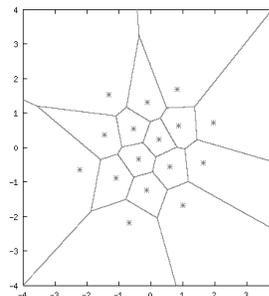


Fig. 5. Exemplo de Quantização vetorial de 4 bits e 2 dimensões.

Vetores de N-dimensão de um determinado espaço vetorial podem ser agrupados em células. Cada célula tem um vetor representante de N-dimensão chamado de *code word*, e cada *code word* é representado por um índice numérico de 1 a M, onde M, é o número máximo de células. Para treinar os

code words, será aplicado o algoritmo de LBG (Linde-Buzo-Gray) que é baseado em distâncias euclidianas. O algoritmo provê todas as code words treinados em uma tabela chamada codebook [1,15].

A fig. (4) mostra o sistema de reconhecimento de um DD-HMM utilizando a quantização vetorial.

$$d(X, Y) = \|X - Y\|^2 \quad (7)$$

A expressão (7) mostra a fórmula da distância euclidiana, onde X e Y são vetores N-dimensionais.

D. DD-HMM

Um dos problemas do DD-HMM é achar a seqüência de estados $S = S_1, S_2, \dots, S_N$, onde N é o número de estados individuais, que melhor modela a seqüência de observações de entrada $O = O_1, O_2, \dots, O_M$, onde M é o número de símbolos de observação distintas por estado. É denotado o estado no tempo t como q_t e os símbolos individuais como $V = \{V_1, V_2, \dots, V_M\}$. A seqüência de observação de entrada é composta por code words, quantizados pelo algoritmo de LBG. Um DD-HMM é caracterizado por $\lambda = (A, B, \pi)$, onde $A = \{a_{ij}\}$ é a matriz de probabilidades de transição de estados e $a_{ij} = P(q_{t+1} = j | q_t = i)$, que é a probabilidade de transição do estado i no tempo t para o estado j no tempo $t+1$, para $1 \leq i, j \leq N$; $B = \{b_j(k)\}$ é a matriz de probabilidade de observação e $b_j(k) = P(O_t = V_k | q_t = j)$, que é a probabilidade de observar o símbolo V_k estando no estado j no tempo t , para $1 \leq k \leq M$ e para $j = 1, 2, \dots, N$. A matriz B define a função de probabilidade de distribuição para cada estado (fig. 6); $\pi = \{\pi_i\}$ é a matriz de probabilidades iniciais e $\pi_i = P(q_1 = i)$, para $1 \leq i \leq N$ [1].

Para treinar as matrizes A e B , é usado o algoritmo de Baum-Welch e para achar a melhor seqüência de estados para uma dada seqüência de observações é utilizado o algoritmo de Viterbi (fig. 7) [1,15].

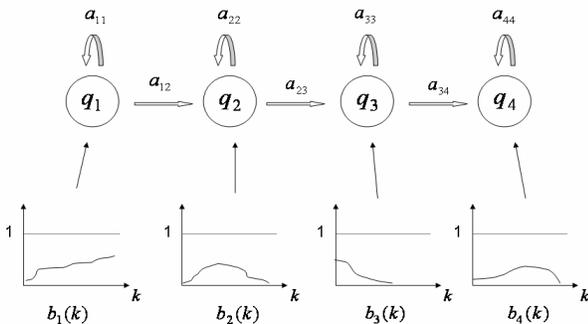


Fig. 6. Exemplo de estrutura DD-HMM de 4 estados.

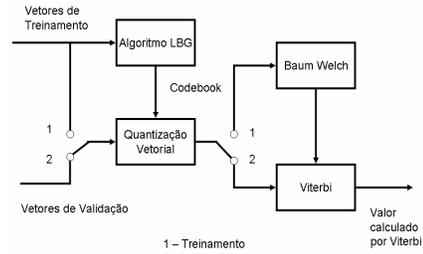


Fig. 7. Estrutura de treino e validação do DD-HMM.

E. Fisiopatologia das Doenças das Pregas Vocais

A qualidade da voz depende do modo de fechamento e abertura da glote e da vibração das pregas vocais. Certas doenças laríngeas impedem que pregas vocais tenham uma vibração glotal harmônica, resultando em uma área vocal onde o trato vocal é excitado em duas frequências fundamentais diferentes[10,16].

III. Materiais e Métodos

Foram utilizados sinais acústicos gravados de pacientes com alterações vocais e de normais, atendidos em consultório, que realizaram videolaringscopia. A distribuição das doenças usadas como modelos das alterações acústicas foi 10 cistos femininos, 5 edemas de Reinke femininos, 7 nódulos femininos, 5 sulcos (3 masculinos e 2 femininos) e 7 pólipos masculinos. Também foram obtidos 8 sinais femininos e 5 masculinos de vozes sem lesões nas pregas vocais e sem alteração acústica perceptível, que compõem o grupo normal. Todos os sinais de voz foram adquiridos em três vogais, “A”, “E” e “I”.

O sinal acústico de foi digitalizado em modo mono, com frequência de amostragem de 22 Khz e 16 bits de resolução[1,9, 15,17].

Os dados obtidos foram submetidos à análise de processamento digital de sinais em programa rotina de análise Cepstral no Matlab[19], que consta dos seguintes passos (Figura 4):

- 1) obtenção de fragmento da vocalização e janelamento
- 2) análise cepstral
- 3) análise por DD-HMM com as três vogais de cada paciente.

A pré-ênfase que permite a filtragem de sons labiais não foi usada, porque altera também o sinal de excitação (da glote), fenômeno descrito na literatura e constatado também neste trabalho[17].

O janelamento foi ajustado para a análise do frame da vocalização em estudo e com o uso da análise Cepstral[17].



Fig.8: Esquema em blocos da análise do sinal acústico.

IV. Resultados

Os achados Cepstrais da análise das vocalizações de cada alteração diagnosticada foram analisados automaticamente utilizando a técnica de DD-HMM. Com a

ajuda das características fisiopatológicas das alterações laríngeas já conhecidas foram descritos os achados que podem ser particularmente atribuídas às diferentes doenças.

Na comparação entre as cinco alterações laríngeas em estudo é possível observar que existem diferenças no perfil cepstrográfico, quando comparadas com o grupo normal e também entre si. Tais diferenças são mais evidentes no edema de Reinke e no sulco vocal.

Para a análise automática do DD-HMM, 70% dos dados foram utilizados para treinamento e o restante foi utilizado para validação do sistema. O *codebook* foi treinado com os dados de treino com 32 vetores código (*code vectors*), cada vetor com 10 elementos, apresentado a distribuição na figura 10:

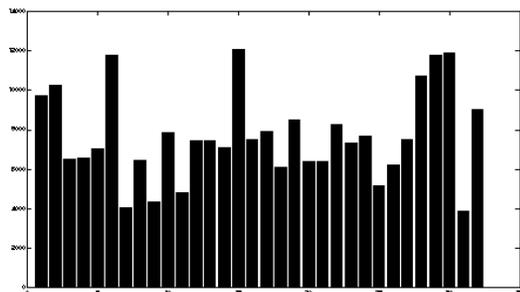


Fig.10. Histograma do codebook após treino.

O sistema de DD-HMM formado por 9 estados, analisou as vogais “A”, “E” e “I” e treinou as matrizes de transição de estados e a matriz de probabilidade de observação pelo algoritmo de Baum-Welch. Em seguida o sistema foi validado pelo algoritmo de Viterbi, o qual compara a probabilidade da sequência analisada ser representada por uma doença específica.

O resultado obtido de validação foi 81,9% de acerto no máximo, pois varia com o número de iterações do algoritmo Baum-Welch, de acordo com a figura 11, sendo que o número máximo de iterações foi 10, pois num número maior de iterações o sistema começa a perder a característica de generalização, ficando mais especialista, ou seja, reconhecendo apenas os dados de treino. Esse resultado valida significativamente o método proposto, contribuindo de forma efetiva para a elaboração de um protocolo de diagnóstico laringeo, baseado na técnica de análise cepstral.

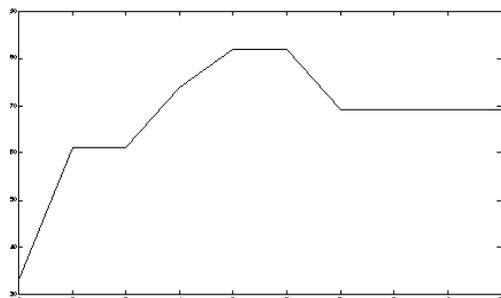


Fig.11. Número de iterações no eixo X por percentagem de acerto no eixo Y.

V. Discussão

Tabela 1. Doenças das Pregas Vocais, suas características principais e aspectos diagnósticos pelos métodos convencionais e o proposto na pesquisa.

Definição ^(2,3,5)	Características acústicas ^(2,3,5)	Imagem da prega vocal	Sistema cepstral de análise da voz (proposto)
Normal Pregas vocais sem lesões	Som perceptivelmente adequado		
Cisto Lesão: cística com fluido em geral unilateral	aumento da massa vibratória, rigidez e assimetria vibratória, gera pitch agravado, diplofonia, aspereza e rouquidão moderada		
Edema Lesão: edema da mucosa das pregas	aumento da massa, onda aperiódica, rigidez diminuída, gera pitch agravado e rouquidão intensa		
Sulco Lesão: falha em forma de sulco, uni ou bilateral	diminuição da massa vibratória, onda “trava” no sulco e a rigidez é intensa, gera pitch agudizado, sopro, pode ter bitonalidade e rouquidão		
Nódulo Lesão: quase sempre bilateral e simétrica	Interfere na vibração dependendo de sua firmeza, aumento de massa gera rouquidão e pitch variáveis, sopro, pode ter diplofonia		
Pólipo Lesão: em geral unilateral	aumento da massa vibratória, aumento da rigidez em geral, vibração assimétrica e aperiódica, gera pitch agravado e rouquidão variada		

Algumas doenças geram alterações significativas nos achados do Cepstro, outras nem tanto.

Os achados cepstrográficos do sulco vocal e do edema de Reinke, são os que mais se diferenciam de todas as outras doenças. O sulco vocal deixa a prega vocal rígida e este é um dos fatores que produzem alterações significativas na voz. O

edema de Reinke deixa a sensação de voz muito grave e rouca e por isto produz uma grande alteração no pitch.

Destaca-se também que algumas doenças apresentam alterações características e constantes que servem como método diagnóstico. A tabela 1, apresenta de maneira condensada, os dados das alterações e os achados do método proposto, comparando as cinco doenças e um caso normal com determinados fragmentos da vocalização.

VI. Conclusão

Com os resultados da proposta de análise cepstral com DD-HMM do sinal da voz iniciados neste trabalho, permite-se prever que este método será uma ferramenta diagnóstica muito útil e promissora, pois é um método não-invasivo, de custo mais baixo e fácil execução. Para uma maior qualificação deste método, deve-se realizar trabalhos com a inclusão de mais amostras, de outras doenças mais raras e de realização de estudo científico estatisticamente adequado para a validação, como um estudo prospectivo duplo cego.

Também é necessário desenvolver modelos de doenças com o diagnóstico mais preciso possível para o treinamento do sistema, através de diagnóstico de certeza das doenças da laringe. Outro aspecto relevante é determinar as possibilidades de que o sexo da pessoa testada possa influenciar o resultado, já que as doenças tendem a alterar o pitch da voz, como observado na pesquisa.

VII. Referências Bibliográficas

- [1]. Rabiner RL, Schafer R. *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
- [2]. Rabiner R.L., Gold B. *Theory and Application of Digital Processing*, Prentice Hall, 1975.
- [3]. Gomez P, Godino JI, Rodriguez F, et al. *Evidence of Vocal Cord Pathology from the mucosal wave Cepstral contents*, IEEE/Universidade Politécnica de Madrid, Madrid, 2004;45-52.
- [4]. Wilpon JG, Rabiner LR, Lee CH et al. "Automatic recognition of keywords in unconstrained speech using Markov Models", *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol.38, N° 11, Nov.1990;1870-1878.
- [5]. Hadjittodorov S, Mitev, P. "A Computer system for acoustic of pathological voices and laryngeal diseases screening, Technical note", *Medical Engineering, & Physics*, Sofia, n. 24, 2002;419-429.
- [6]. Manfredi C, D'aniello M, Brusciaglioni P, et al. "A comparative analysis of fundamental frequency estimation methods with application to pathological voices", *Medical Engineering & Physics*, Firenze, n. 22, 2000, 135-147.
- [7]. Wszolek W, Tadeusiewicz R, Izworski A, et al. "Automated understanding of selected voice tract pathologies based on the speech signal analysis", *Proceedings of the 23rd Annual EMBS International Conference, EMBS, Istanbul, October*, 2001;25-28.
- [8]. Mitev P, Hadjittodorov S. "Fundamental frequency estimation of voice of patients with laryngeal disorders", *Information Sciences*, Sofia, n. 156, 2003;3-19.
- [9]. Rosa MO, Pereira JC, Greller M and Carvalho A. "Signal processing and statistical procedures to identify laryngeal pathologies", *IEEE Transactions on Biomedical Engineering* IEEE/EESC-USP, 1999; 423-426.
- [10]. Dedivitis RA, Barros, APB. *Métodos de Avaliação e Diagnóstico de Laringe e Voz*, 2ª ed Lovise, São Paulo, 2002.
- [11]. Minoru H. Diane MB. *Exame videostroboscópico da laringe*; Porto Alegre, Artes Médicas, 1997.
- [12]. Erich CM, Lupercio LB, Osiris CB, et al. Incidência de lesões laríngeas não neoplásicas em pacientes com queixas vocais; *Revista Brasileira de Otorrinolaringologia*, vol.67, n.6, nov/dez 2001;788-94.
- [13]. Khul I. *Manual prático de Laringologia*; Editora da universidade, Porto Alegre 1982.
- [14]. Martinez CE, Rufiner HL. "Acoustic analysis of speech for detection of laryngeal pathologies", *Proceedings of the 22nd Annual EMBS International Conference, EMBS, Chicago, July 2000*, 23-28.
- [15]. Fagundes RDR. *Reconhecimento de Voz, Linguagem Contínua*, usando Modelos de Markov, Dissertação de Mestrado, Universidade de São Paulo, São Paulo, 1993.
- [16]. Hansen JH, Gavdida-Ceballos L, Kaiser RJF. "A Nonlinear operator-Based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment". *IEEE Transactions on Biomedical Engineering*, Vol. 45, N°. 3, March 1998;937-940.
- [17]. Furui S. *Digital Speech Processing, Synthesis and Recognition*, Marcel Dekker, Inc., 2001.