

ISSN: 1727-1983
ISBN: 85-98739-01-4

IFMBE Proceedings

Volume 5, 2004



*III Latin American Congress on Biomedical Engineering
III Congresso Latino-Americano de Engenharia Biomédica
III Congreso Latinoamericano de Ingeniería Biomédica
XIX Congresso Brasileiro de Engenharia Biomédica*

September 22 - 25, 2004
João Pessoa, Paraíba, Brazil



Proceedings of the International Federation for Medical and Biological Engineering
© 2004 - SBEB, CORAL & IFMBE

Classificador Bayesiano para Otimização de Parâmetros Acústicos no Diagnóstico de Condições da Laringe

Luis Fernando Granato*, José Carlos Pereira**

USP/Engenharia Elétrica, São Carlos, SP, Brazil

e-mail: lgranato@sel.eesc.usp.br

Abstract

In this work, a non invasive technique is used in order to evaluate the vocal tract and vocal folds conditions. This technique allows specialists to obtain a more accurate diagnosis of the larynx. By using both, univariated and multivariated approaches of the Bayesian classifiers onto 38 acoustics parameters of voice, it is possible to evaluate the physiological conditions of the vocal folds and, consequently to determine functional or organic disfunctions of the larynx.

Key words : Baysean classifier, signal and voice.

1. Introdução

A mudança na voz é o principal sintoma que indica uma possível doença na laringe, sendo esta de origem orgânica ou funcional. Tradicionalmente os especialistas aplicam inicialmente o teste de audição da voz do paciente, aliado na maioria das vezes aos exames de videolaringoscopia direta, que visualiza a laringe através de uma microcâmera e videoestroboscopia, possibilitando a visualização da vibração das pregas vocais. Vale lembrar que esses procedimentos são bastante desconfortáveis causando além de dor, um stress grande ao paciente. O resultado desses procedimentos convergem para uma avaliação qualitativa, ou seja, resultados difíceis de serem quantificados, além do que requerem um fato de suma importância para o desfecho do diagnóstico, a experiência do especialista.

Acusticamente, as patologias são diferenciadas em relação à padrões perceptuais, sendo classificadas, de modo genérico, em voz normal, áspera, soprosa e rouca, independente da origem.

Várias técnicas têm sido propostas para auxiliar no diagnóstico do especialista utilizando a análise acústica da voz, tais como a filtragem inversa [2], redes neurais artificiais [6], medidas acústicas da voz [5], teoria bayesiana, dentre outras. Esse caminho propicia ao paciente conforto ao ser avaliado, submetendo-o a uma simples captura do sinal de voz.

A análise acústica da voz através do processamento digital de sinais resulta no formato do traçado da onda sonora, na análise de frequência

fundamental (F_0), nas medidas de ruído, permitindo descrever de forma bastante aceitável a voz humana.

As medidas de F_0 são definidas como sendo o número de vibrações das pregas vocais produzidas por segundo [4]. As medidas de jitter e shimmer são definidas como a perturbação ou variabilidade da F_0 e da amplitude, respectivamente ciclo a ciclo.

As medidas de ruído quantificam o ruído originário da turbulência de ar na glote. Destaca-se SNL (Spectral Noise Level) que quantifica as características de aspereza : HNR (Harmonic to Noise Ratio) que oferece um índice que relaciona o componente harmônico com o componente ruído; NNE (Normalized Noise Energy) um índice acústico que estima o ruído devido ao fechamento insuficiente da glote e BR (Breathiness Ratio) que estima características perceptuais de soprosidade.

O raciocínio probabilístico baseia-se no fato de que os médicos convivem com a incerteza em um grau comparável ao de bem poucos profissionais. Probabilidade, no caso, seria apenas uma maneira de medir essa incerteza. Quando dois médicos dizem, por exemplo, que um paciente provavelmente tem embolia pulmonar, um pode estar pensando em uma probabilidade de 30% e o outro em 90%. No manuseio dessa incerteza o médico (especialista) recorre muitas vezes a exames laboratoriais que a alteram, mas não a eliminam. Na interpretação probabilística desses testes, os médicos (especialistas) experientes usam intuitivamente o chamado teorema de Bayes [1], o qual relaciona sensibilidade, especificidade, prevalência e probabilidade a posteriori. Obviamente, a aplicação de tais conceitos deve idealmente referir-se não apenas a exames laboratoriais, mas também a achados de história ou exames físicos [1].

Dessa forma, este trabalho tem como objetivo utilizar o classificador Bayesiano para a classificação de patologias da laringe até então nunca aplicado com esse intuito, usando-se duas vertentes de parâmetros acústicos : univariados e multivariados, ou seja, analisando-se a porcentagem de acerto de classificação para vozes normais através do classificador Bayesiano para cada um dos 38 parâmetros (vide tabela 1) individualmente e analisando as combinações dos 38 parâmetros, ou seja, 2 a 2, 3 a 3, 4 a 4 até 38 a 38. Com isso, é possível avaliar qual (univariado) ou quais (multivariado) parâmetros são interessantes para o diagnóstico de patologias da laringe.

Tabela 1. Lista dos 38 parâmetros acústicos em análise.

Exerc	Parâmetro	Normal
1	Fo	61
2	Jitter_PPF	0
3	Jitter_DPF	8
4	Jitter_JR	5
5	Jitter_JF	5
6	Shimmer_PPF	2
7	Shimmer_DPF	2
8	Shimmer_JR	0
9	Shimmer_JF	11
10	Br	11
11	Hfpr_>6000Hz	0
12	Hfpr_50-400Hz	20
13	Hfpr_400-2000Hz	51
14	Hfpr_2000-5000Hz	8
15	Srl_Media_100-2600Hz	14
16	Srl_Mediana_100-2600Hz	20
17	Srl_Media_2600-5100Hz	11
18	Srl_Mediana_2600-5100Hz	0
19	Srl_Media_5100-8000Hz	20
20	Srl_Mediana_5100-9000Hz	64
21	Srl_Media_100-5100Hz	55
22	Srl_Mediana_100-5100Hz	55
23	Srl_Media_100-8000Hz	50
24	Srl_Mediana_100-8000Hz	47
25	Ra_Media	2
26	Ra_Mediana	0
27	Hr_Media	5
28	Hr_Mediana	5
29	Ns_Medio	8
30	Ns_Mediana	8
31	Hnr_Media	8
32	Hnr_Mediana	0
33	Nne_50-400Hz	23
34	Nne_400-2000Hz	32
35	Nne_2000-5000Hz	32
36	Nne_5000-8000Hz	20
37	Pvi	26
38	Avi	35

2. Classificador Bayesiano

Admitindo que x é um vetor que representa parâmetros acústicos, da seguinte forma :

$$x = [F_o, \text{Jitter}, \text{Shimmer}, \text{Br}, \text{Hfpr}, \text{Snl}, \text{Ra}, \text{Hr}, \text{Ns}, \text{Hnr}, \text{Nne}, \text{Avi}, \text{Pvi}] \quad (1)$$

Considerando o problema de classificação de um dado parâmetro acústico x em uma das situações, w_1 e w_2 , respectivamente, normal e nódulo, partindo-se do pressuposto que cada parâmetro acústico x pode ser alocado em uma, e somente uma das situações w_i . O objetivo do classificador é encontrar solução ótima minimizando a probabilidade de erro.

Uma das possibilidades de se encontrar essa solução ótima é a regra de decisão Bayesiana, que parte da idéia de que as funções de probabilidade (fdp) dentro de cada situação w_i , denotada por $p(x/w_i)$, são conhecidas e que há disponibilidade de informações adicionais sobre as situações w_i , as quais são chamadas de probabilidade a priori. Daí, a partir da informação das fdp de cada situação w_i e das probabilidades *a priori* é possível aplicar o teorema de Bayes para obter a probabilidade *a posteriori* de um parâmetro acústico x classificar a situação w_i :

$$P(w_i/x) = \frac{P(w_i) p(x/w_i)}{\sum_{j=1}^n P(w_j) p(x/w_j)} \quad (2)$$

onde :

- $P(w_i)$ = probabilidade *a priori* de um parâmetro acústico classificar uma das situações w_i ;
- n = número de situações;
- $p(x/w_i)$ = função de densidade de probabilidade de x nas situações w_i e
- $P(w_i/x)$ = probabilidade *a posteriori* para as situações w_i dado parâmetro acústico x .

O denominador do teorema de Bayes é constante para as situações w_i , dessa forma este pode ser desconsiderado na regra. Com isso, podemos utilizar a regra de classificação baseada no numerador do teorema. Assim, um dado parâmetro acústico x classificará à situação w_i se :

$$P(w_i/x) > P(w_j/x), \text{ para } i \neq j \quad (3)$$

onde a função de densidade de probabilidade $p(x/w_i)$ para as situações w_i ($i=1,2$) são gaussianas e tem a seguinte forma [3]:

$$x \text{ classifica } w_i \text{ se } F_i(x) > F_j(x) \forall i \neq j \quad (4)$$

onde :

- para a etapa univariada :

$$F_i(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad (5)$$

onde : μ : média e σ : desvio-padrão

- para a etapa multivariada :

$$F_i = \frac{1}{(2\pi)^{\frac{p}{2}} \left| \sum \right|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \sum^{-1} (x-\mu)} \quad (6)$$

onde : $-\infty < x_i < \infty, i=1,2,\dots,p$ (combinações) e \sum : matriz de covariância.

3. Materiais e métodos

Foram utilizadas 60 vozes de pacientes normais (sem nenhuma patologia diagnosticada), as quais foram obtidas através da gravação da vogal /a/ através do software desenvolvido por Rosa [6] na EESC – USP.

A gravação foi feita diretamente num PC Pentium 100 MHz com placa de aquisição SoundBlaster SB16, em ambiente isolado. O sinal foi captado por microfone tipo cardióide unidirecional com ganho dinâmico posicionado à 5 cm da boca do paciente, o qual emitiu de forma sustentada a vogal /a/ por aproximadamente 7s. Foi orientado para uma fonação em níveis de altura e intensidade confortáveis ao emissor, evitando tensão excessiva das pregas vocais.

O sinal, foi armazenado no PC, foi pré-processado com a remoção de suas porções instáveis iniciais e finais padronizando os sinais em 5s, as amplitudes padronizadas entre -1 e +1 e utilizando de um algoritmo para a retirada da tendência linear. O intuito da padronização se deu para obtenção de uma análise uniforme sem que características da gravação influenciassem os parâmetros acústicos.

As medidas estudadas foram de frequência, perturbação e ruído. A F_0 foi obtida através do processamento de sinais por *waveform matching*. As medidas de perturbação foram para *jitter* : DPF (*Directional Perturbation Factor*), JR(*Jitter Ratio*), JF (*Jitter Factor*), PVI (*Perturbation Variation Index*) e para *shimmer* : DPF (*Directional Perturbation Factor*), JR(*Jitter Ratio*), JF (*Jitter Factor*), AVI (*Amplitude Variation Index*). As medidas de ruído foram : SNL (*Spectral Noise Level*), HNR(*Harmonic to Noise Ratio*), NNE (*Normalized Noise Energies*) e BR (*Breathiness Ratio*).

4. Resultados experimentais

Foram organizados em uma tabela 67 pacientes com voz normal e seus respectivos valores numéricos dos 38 parâmetros acústicos extraídos (vide tabela 1), onde 50% desse conjunto de vozes foram destinados a fase de treinamento e os outros 50% para a fase de teste. A partir daí aplica-se o classificador Bayesiano, lembrando que separamos em etapas univariada e multivariada. Na etapa univariada, para cada um dos 38 parâmetros acústicos são calculadas média e variância (fase de treinamento) que entram como variáveis na fdp univariada (5) na qual retornará um valor numérico (fase de teste). Dessa forma, tem-se 38 valores obtidos pela fdp univariada (5) e de acordo com o teste (4) obtem-se o maior valor e portanto, o correspondente parâmetro ótimo para classificar voz normal. Os resultados obtidos na etapa univariada, ou seja, o parâmetro ótimo para classificar voz normal e aqueles com pelo menos 50% de acerto são mostrados na figura 1 e tabela 2, respectivamente.

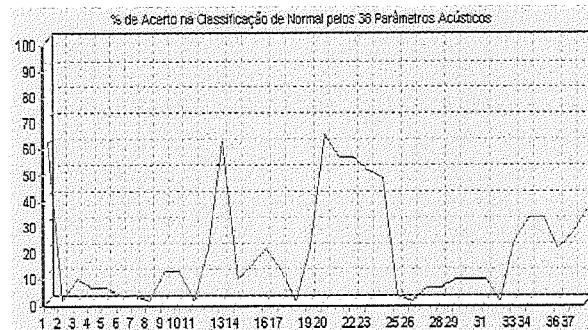


Figura 1. Porcentagem dos 38 parâmetros acústicos na classificação de voz normal.

Tabela 2. Parâmetros acústicos com pelo menos 50% de classificação correta de voz normal.

Exec	Parâmetro	Normal
20	Srl_Mediana_5100-8000Hz	64
1	F_0	61
13	Hfpr_400-2000Hz	61
21	Srl_Media_100-5100Hz	55
22	Srl_Mediana_100-5100Hz	55
23	Srl_Media_100-8000Hz	50

Na etapa multivariada, a idéia é a mesma da univariada porém, o objetivo é o de realizar combinações, ou seja, 2 a 2, 3 a 3, 4 a 4, ..., 38 a 38 entre os 38 parâmetros até conseguir 100% na classificação. Por exemplo, com-binando-se 2 a 2, cada parâmetro será combinado com outros 37 parâmetros (onde $F_0 - Jitter_PFF = Jitter_PFF - F_0$, portanto, considera-se somente uma vez) e para cada uma delas são obtidas média e matriz de covariância que entram como variáveis na fdp multivariada (6) na qual retornará um valor numérico. Iniciamos com as combinações 2 a 2, um total de 703 combinações realizadas entre os 38 parâmetros acústicos, obtendo-se 96% de classificação correta ; depois passamos para as combinações 3 a 3, um total de 8.436 combinações realizadas entre os 38 parâmetros acústicos, obtendo-se um total de 100% de classificação correta. Dessa forma, alcançamos o objetivo já nas combinações 3 a 3, classificando com 100% vozes normais. Os resultados obtidos nas combinações 2 a 2 e 3 a 3, ou seja a combinação ótima e aquelas próximas a ótima para classificar voz normal são mostradas nas tabelas 3 e 4.

Tabela 3. Combinação 2 a 2 ótima e as mais próximas da ótima para classificar voz normal.

Combinação	Parâmetro	Normal
659	Ns_Media-Ns_Mediana-	95
62	Jitter_DPF+Hfpr_30-400Hz-	93
59	Jitter_DPF-Ns_Media-	93
676	Hnr_Media-Hnr_Mediana-	93
79	Jitter_DPF-Shimmer_JF-	91
81	Jitter_DPF+Hfpr_>8000Hz-	91
613	Ra_Media-Fz_Mediana-	91
454	Srl_Media_2600-5100Hz-Ns_Media-	87
50	Jitter_DPF-Srl_Mediana_5100-3000Hz-	84
53	Jitter_DPF-Srl_Media_100-8000Hz-	84
212	Shimmer_DPF-Hfpr_50-400Hz-	84
523	Srl_Media_5100-8000Hz-Ns_Media-	84
541	Srl_Mediana_5100-8000Hz-Ns_Media-	84
542	Srl_Mediana_5100-8000Hz-Ns_Mediana-	84
584	Srl_Media_100-8000Hz-Srl_Mediana_100-E	84
589	Srl_Media_100-8000Hz-Ns_Media-	84
603	Srl_Mediana_100-8000Hz-Ns_Media-	84
89	Jitter_DPF-Srl_Media_E100-8000Hz-	81
54	Jitter_DPF-Srl_Mediana_100-8000Hz-	81
55	Jitter_DPF-Fz_Media-	81
100	Jitter_DPF-Ns_Mediana-	81
209	Shimmer_DPF-Shimmer_JF-	81
229	Shimmer_DPF-Ns_Media-	81
475	Srl_Media_2600-5100Hz-Srl_Mediana_5100	81
537	Srl_Mediana_5100-8000Hz-Fz_Media-	81
551	Srl_Media_100-E100-Hz-Srl_Mediana_100-E	81
580	Srl_Media_100-8000Hz-Ns_Mediana-	81

Os resultados obtidos nas combinações 3 a 3, ou seja a combinação ótima e aquelas próximas a ótima para classificar voz normal são mostradas na tabela 4.

5. Conclusões

O uso de parâmetros acústicos e classificador Bayesiano para avaliar características de voz normal dentro do contexto de patologias da laringe, mostra-se eficiente e uma aplicação praticamente isenta de custo instrumentais para o apoio ao diagnóstico médico para patologias da laringe.

Ficou claro, de acordo com os resultados obtidos, que a combinação dos parâmetros acústicos é de fundamental importância para a obtenção de classificação com maior probabilidade de se conseguir 100% de acerto.

Os resultados mostram que o parâmetro ótimo e os com pelo menos 50% de classificação correta, estão presente entre as melhores combinações, ou seja, o uso do classificador Bayesiano mostrou-se eficaz nesse tipo de análise retornando os parâmetros que contribuem sensivelmente para a classificação correta.

Este trabalho mostra então, o potencial de informações obtidas por essas combinações para se avaliar o comportamento e a importância de cada parâmetro acústico diante do objeto em estudo, ou seja,

Tabela 4. Combinações 3 a 3 ótimas e as mais próximas das ótimas para classificar voz normal.

Combinação	Frequência	Nº de
1432 _lto_0PF-SrImel_3-3n_Medaa_100-2000Hz-	100	
1434 _lto_0PF-SrImel_3-3n_Medaa_100-3000Hz-	100	
7631 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-S1_vedara_1C0800Hz-	100	
7632 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Ra_Medaa-	100	
7633 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Ra_Medana-	100	
7634 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Hr_Medaa-	100	
7635 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Hr_Medana-	100	
7636 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Ns_Medaa-	100	
7637 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Ns_Medana-	100	
7638 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Hm_Medaa-	100	
7639 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Hm_Medana-	100	
7640 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Hmz_BD-C0H-	100	
7641 Sr_Meda_10051CH-3n_Medaa_100-3000Hz-Hmz_400-2000Hz-	100	
7642 Sr_Meda_10051CH-3n_Medaa_100-3000Hz_Hmz_400-2000Hz_200-500Hz-	100	
7643 Sr_Meda_10051CH-3n_Medaa_100-3000Hz_Hmz_500-800Hz-	100	
7644 Sr_Meda_10051CH-3n_Medaa_100-3000Hz_Pi+	100	
7651 Sr_Meda_10051CH-3n_Medaa_100-3000Hz_Av+	100	
7662 Sr_Meda_10051CH-3n_Medana_1C-8000Hz-Hre_2C00500Hz-	100	
7663 Sr_Meda_10051CH-3n_Medana_1C-8000Hz-nv+	100	
7782 Sr_Medana_100-5100Hz-Sr_Medana_1C0800Hz-Av-	100	
8013 Fe_Meda-Ra_Medana-Ns_Medana-	100	
8017 Fe_Meda-Ra_Medana-H1_Meda-	100	
8017 Fe_Meda-Ra_Medana-H1_Medana-	100	
802 Fe_Meda-Ra_Medana-Ns_80-400Hz-	100	
8081 Fe_Meda-Ra_Medana-Ns_400-3000Hz-	100	
8081 Fe_Meda-Ra_Medana-Ns_200C-8000Hz-	100	
8082 Fe_Meda-Ra_Medana-Ns_500C-8000Hz-	100	
8083 Fe_Meda-Ra_Medana-Pi+	100	
8084 Fe_Meda-Ra_Medana-Av+	100	
1404 _lto_0PF-SrImel_CPF-Sr_Medana_100-2000Hz-	36	
1462 _lto_0PF-SrImel_3-Sr_Meda_10C-2500Hz-	36	
1463 _lto_0PF-SrImel_3-Sr_Medana_100-2500Hz-	36	
1892 _lto_JR_lto_JF-SrImel_FPF-	36	
3947 Shriner_JF-SrImel_3-Sr_Meda_1C-350CH-	36	
3948 Shriner_JF-SrImel_3-Sr_Medana_100-2000Hz-	36	

indicando o melhor e os melhores parâmetros acústicos para classificar voz normal e estendendo-se as patologias da laringe.

5. Referências Bibliográficas

- [1] ANDRADE, P.J.N., "Sistemas especialistas de apoio ao diagnóstico em medicina. Relações com o Teorema de Bayes e com a Lógica do raciocínio diagnóstico", Arq. Bras. Cardiologia, vol.73, nº6, 1999.
- [2] CHILDERS, D.G., Bal, K.S., "Detection of laryngeal function using speech and electroglottographic data", IEEE Trans. On Medical Eng, 39, 19-25, 1992.
- [3] JOHNSON, R. A., *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey, 1998.
- [4] MURRAY, K.D., Zubich, H.H., *Evaluation of vocal function.*, In: Fried M.D.. The Larynx, Mosby, 2^a-ed., cap.11, 115-24,1996.
- [5] RAZERA, D.E., Determinadores de *Pitch*, Dissertação de Mestrado, EESC-USP, São Carlos, SP, Brasil, 2004.
- [6] ROSA, M.O., Análise acústica da voz para pré-diagnóstico de patologias da laringe, Dissertação de Mestrado, EESC-USP, São Carlos, SP, Brasil, 1998.