

# Parâmetros espectrais de vozes saudáveis e patológicas

Comparação de resultados entre duas bases de dados

## *Spectral features of healthy and pathological voices*

*Results comparison between two databases*

Hugo Cordeiro

Depart. de Engenharia Eletrónica, Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa  
Instituto Politécnico de Lisboa  
Lisboa, Portugal  
hcordeiro@deetc.isel.ipl.pt

Carlos Meneses

Depart. de Engenharia Eletrónica, Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa  
Instituto Politécnico de Lisboa  
Lisboa, Portugal  
cmeneses@deetc.isel.ipl.pt

**Resumo** — Este artigo apresenta um estudo comparativo entre três parâmetros espectrais na discriminação entre vozes saudáveis e patológicas. Os parâmetros avaliados envolvem a análise do primeiro pico espectral, uma medida da relação sinal-ruído e o declive entre duas bandas de baixa frequência do sinal de fala. O declive entre as bandas de baixa frequência é proposto como otimização do primeiro pico espectral, de modo a colmatar os casos de erro de classificação devido à degradação da qualidade vocal com o avanço da doença. Os três parâmetros são avaliados em duas bases de dados. O declive entre as bandas de baixa frequência obtém os melhores resultados, com 100% de acurácia na base de dados da USP e 83,5% de acurácia na base de dados da MEEI.

**Palavras Chave** - Vozes patológicas, Diagnóstico, Parâmetros espectrais.

**Abstract** — This paper presents a comparative study between three spectral parameters in the discrimination between healthy and pathological voices. The evaluated parameters involve the analysis of the first spectral peak, the Relative Power of the Periodic Component, which corresponds to a measure of the signal-to-noise ratio and the Low Band Spectral Tilt. The Low Band Spectral Tilt is proposed as optimization of the first spectral peak, to resolve the cases of classification error due to the degradation of vocal quality with the disease progression. The three parameters are evaluated in two databases. The Low Band Spectral Tilt achieves the best results, with 100% accuracy in the USP database and 83.5% accuracy in the MEEI database.

**Keywords** - Pathological Voices, Diagnostic, Spectral Features.

### I. INTRODUÇÃO

O diagnóstico de patologias da voz envolve normalmente métodos invasivos que requerem esforços conjuntos de equipas multidisciplinares. A utilização de um método automático baseado em técnicas de processamento de fala, sendo não invasivo e rápido, pode ser um método de rastreio ou de diagnóstico preliminar ao realizado por especialistas [1].

Existem duas abordagens clássicas na caracterização de vozes patológicas: a análise acústica e a análise espectral. Na primeira são avaliadas características como o *jitter* e *shimmer* da frequência fundamental [2], [3]. Na segunda são avaliados métodos de relação sinal-ruído do sinal e análise da energia espectral [4], [5].

O uso de classificadores de alto nível com recurso a redes neuronais [6], SVM [7] e outros [8]–[10] são recorrentes no reconhecimento de vozes patológicas e amplamente usados em reconhecimento de padrões. Contudo, apesar de normalmente os resultados apresentados estarem acima de 90%, nem sempre existe o foco na interpretação dos parâmetros que podem ser mais eficientes na discriminação de vozes patológicas.

Em [11] verificou-se que em sistemas baseados em modelos de misturas gaussianas e parâmetros espectrais que é possível reconhecer vozes patológicas com fala contínua de sinais com uma largura de banda de 2 kHz. Este trabalho motivou uma análise mais detalhada do espectro dos sinais nas baixas frequências sendo os resultados preliminares apresentados em [12].

O objetivo deste artigo é apresentar o parâmetro *Low Banda Spectral Tilt (LBST)*, baseado na análise do declive entre duas bandas de baixa frequência, para caracterização de vozes patológicas na base de dados da *Massachusetts Eye and Ear Infirmary* (MEEI) [13] e compará-los com os resultados já obtidos para a base de dados da Universidade de São Paulo (USP) [14], em que foi garantida a total discriminação entre sujeitos saudáveis e não saudáveis. Os resultados são comparados com outros dois parâmetros [3], o *first spectral peak (P1)* e o *Relative Power of the Periodic Component (RPPC)*. Estes parâmetros são retirados do sinal de fala e são interpretados fisicamente quer face à condição patológica quer face ao estado de desenvolvimento da patologia.

Este artigo está organizado da forma seguinte: na secção II descrevem-se os métodos de caracterização de vozes patológicas, na secção III é apresentada a descrição das bases

de dados, na secção IV são apresentados os resultados obtidos e sua discussão, sendo a secção V dedicada às conclusões.

## II. PARÂMETROS ESPECTRAIS

### A. Comparação entre sujeitos saudáveis e não saudáveis

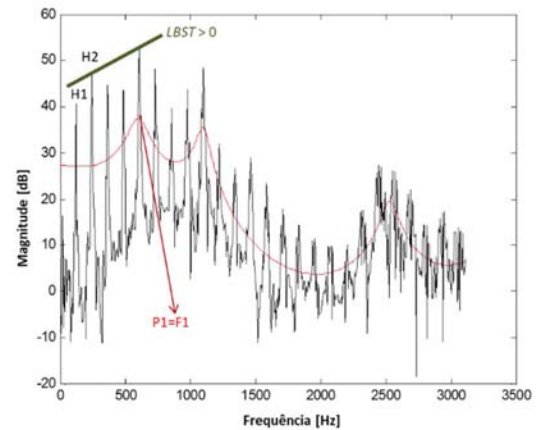
Para as vozes saudáveis as harmônicas com maior energia são as correspondentes ao primeiro formante ( $F1$ ), que coincidem com o primeiro pico da envolvente espectral ( $P1$ ), como pode ser visto na Fig. 1a). Para estimar esta envolvente espectral foi utilizado um filtro LPC de ordem 30. Pode também ocorrer um pico na envolvente espectral a modelar as duas primeiras harmônicas, mas com uma energia menor que a do primeiro formante e com uma largura de banda que tende a aumentar. Devido à vibração das pregas vocais, modeladas por dois polos na origem, surge um declive espectral negativo que diminui a energia das altas frequências. Contudo, o declive espectral abaixo do primeiro formante é positivo.

Para a voz produzida por sujeitos não saudáveis, no entanto, as duas primeiras harmônicas têm uma energia maior em relação às harmônicas de ordem superior, nomeadamente em relação ao primeiro formante, o que favorece a sua modelação, como pode ser visto na Fig. 1b). Estas duas harmônicas, cujas amplitudes são designadas na literatura por  $H1$  e  $H2$ , são alvo de vários estudos [15], [16]. Esses estudos revelam que a diferença entre estas duas amplitudes está relacionada com as vozes soprosas, presentes em muitas das patologias das pregas vocais, pois a segunda harmônica tem menos energia que a primeira harmônica, ao contrário das vozes não soprosas. O primeiro pico da envolvente espectral também pode modelar apenas a primeira harmônica da frequência fundamental. Nestes casos, a energia da primeira harmônica prevalece sobre as restantes podendo as restantes harmônicas serem ocultadas por ruído. A largura de banda é normalmente menor do que quando é modelada a primeira e a segunda harmônicas. Ocorre também, embora num número inferior de casos, a modelação apenas da segunda harmônica.

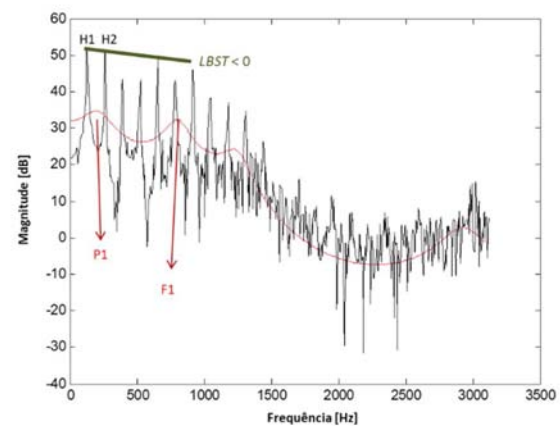
Devido a este pico de muito baixa frequência o declive espectral abaixo da frequência do típico formante da vogal /a/ torna-se negativo. Dependendo da evolução da doença, as pregas vocais aumentam a aperiodicidade e o sinal tende a tornar-se mais ruidoso, principalmente nas frequências mais elevadas. O declive espectral tende a diminuir e a energia nas altas frequências a aumentar, o que prejudica a modelação das duas primeiras harmônicas, como pode ser visto na Fig. 1c).

### B. First Spectral Peak - $P1$ .

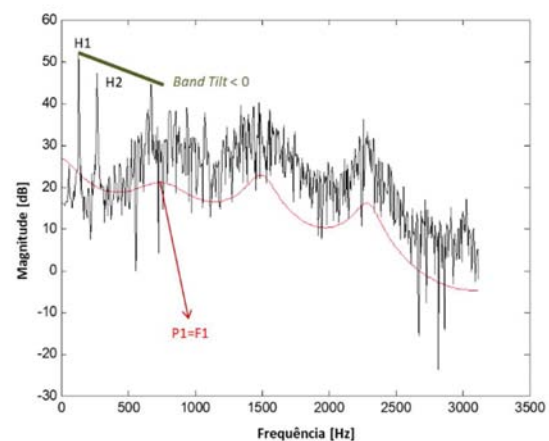
Como descrito anteriormente e pode ser observado na Fig. 1b), nos sujeitos não saudáveis em que a patologia não está avançada, surge na envolvente espectral um pico que modela as duas primeiras harmônicas [4], antes do primeiro formante, que para a vogal /a/ se situa normalmente entre os 550 Hz e os 800 Hz. Para sons vozeados, como a vogal /a/, as vibrações das pregas vocais impõem um declive espectral que diminui a energia das frequências mais elevadas, favorecendo a modelação do primeiro pico espectral nos sujeitos não saudáveis. Nos sujeitos saudáveis, este pico espectral ou não existe, Fig. 1a), ou se existir exibe uma maior largura de banda e menor energia do que nos sujeitos não saudáveis.



a) Sujeito Saudável,  $P1=F1 > 550$  Hz,  $RPPC=0,979$ ,  $LBST > 0$



b) Sujeito Não Saudável,  $P1 < 550$  Hz,  $RPPC=0,977$ ,  $LBST < 0$



c) Sujeito Não Saudável,  $P1=F1 > 550$  Hz,  $RPPC=0,54$ ,  $LBST < 0$

Figure 1. Espectros e envolventes espectrais

Este pico, denominado primeiro pico espectral ( $P1$ ), pode ser utilizado na detecção de vozes patológicas. Contudo, como

se verifica na Fig. 1c), com a evolução da doença o ruído aumenta e o declive espectral diminui inibindo a detecção do primeiro pico espectral na zona das duas primeiras harmónicas, e consequentemente a detecção de vozes patológicas falha só com este parâmetro.

### C. Relative Power of the Periodic Component - RPPC

O *RPPC*, *Relative Power of the Periodic Component* [4] corresponde à autocorrelação normalizada dada pela Eq. 1, mas calculado para um atraso correspondente ao período fundamental das pregas vocais,  $T_0$ , como na Eq. 2.

$$r'_x(\tau) = \frac{r(\tau)}{r(0)} \quad (1)$$

$$RPPC = r'_x(T_0) = \frac{r(T_0)}{r(0)} \quad (2)$$

A autocorrelação normalizada possui um valor máximo de 1 para um atraso de 0 amostras. Se o sinal fosse totalmente periódico existiriam também máximos de 1 em múltiplos de  $T_0$ . Para uma voz saudável, quase periódica, o *RPPC* toma mesmo assim valores perto de um. Para vozes não saudáveis, devido ao ruído e à aperiodicidade do sinal (*jitter* e *shimmer*) o seu valor baixa. Contudo, se a doença estiver num estado inicial, o valor do *RPPC* pode ser elevado e a detecção de vozes patológicas falha só com este parâmetro. As legendas das figuras anteriores contêm os valores deste parâmetro para vozes saudáveis e não saudáveis.

### D. Low Band Spectral Tilt - LBST

Quando analisados os espectros dos sinais de fala dos sujeitos saudáveis correspondentes à vogal /a/ sustentada, Fig. 1a), e os espectros de sujeitos não saudáveis, Fig. 1b) e 1c), pode-se verificar em todos a presença do típico primeiro formate entre os 550 Hz e os 800 Hz, representado por um pico na envolvente espectral. Na banda das duas primeiras harmónicas pode ou não a envolvente espectral modular um pico nestas harmónicas. Contudo, se tomarmos em consideração diretamente o espectro do sinal e não a envolvente espectral, a energia na banda das duas primeiras harmónicas (*FBME* – *First Band Maximum Energy*) é superior à energia das harmónicas do típico formante do /a/ (*SBME* – *Second Band Maximum Energy*) em vozes patológicas, quer para a situação da Fig. 1b) em que a patologia está num estado inicial, quer da Fig. 1c) em que a patologia está num estado mais avançado.

No método proposto em [12] o espectro do sinal foi analisado em duas bandas. A primeira banda contém as duas primeiras harmónicas, aonde é detetado o pico nas vozes patológicas. A segunda banda é caracterizada pelo intervalo entre a terceira e a décima harmónica, aonde se situa o típico formante da vogal /a/. Este intervalo é suficiente para conseguir estimar este formante em sujeitos com a frequência fundamental mais baixa. Por exemplo, garante que num orador com frequência fundamental  $F_0=80$  Hz e primeiro formante  $F_1=800$  Hz, o pico na segunda banda é estimado na 10ª harmónica.

O espectro do sinal e as energias máximas das bandas do espectro são calculados por um algoritmo semelhante ao aplicado em [17], mas é ignorada a componente de ruído e

apenas é a analisada energia das harmónicas. De um modo sucinto o algoritmo consiste na estimativa da frequência fundamental através do método da autocorreção, sendo depois o espectro do sinal calculado com uma dimensão de 60 períodos fundamentais. Esta análise permite um espectro de longo termo com maior resolução. Para cada orador foram selecionados 500 ms de fala numa zona estável do sinal sendo a sua amplitude normalizada. O declive das duas bandas é dado pelo quociente entre as diferenças de amplitude máxima das duas bandas e a diferença das respetivas frequências, Eq. (3),

$$LBST = \frac{SBME - FBME}{SBME_{freq} - FBME_{freq}} \quad (3)$$

Como se pode verificar na Fig. 1, o *LBST* é positivo para o caso do orador saudável e negativo nos dois casos dos sujeitos não saudáveis, independentemente da evolução da doença.

## III. BASES DE DADOS

Na realização deste trabalho foi usado um subconjunto da base de dados da MEEI. No total foram utilizados 53 sujeitos saudáveis e 153 sujeitos com várias patologias (15 com nódulos; 37 com Edema; 15 com pólipos, 21 com Keratosis\Leukoplakia; 65 com paralisia das pregas vocais). Não foi realizada qualquer separação de género. O sinal de fala utilizado neste trabalho foi a vogal /a/ sustentada, sendo os dados semelhantes aos usados em vários trabalhos anteriores, nomeadamente em [3]–[5].

A base de dados da Universidade de São Paulo, USP, é composta por 15 sujeitos saudáveis e 31 sujeitos com patologias da voz (16 com Edema de Reinke e 15 com nódulos). O sinal de fala corresponde à vogal /a/ sustentada.

## IV. RESULTADOS E DISCUSSÃO

Os resultados apresentados nesta secção têm como objetivo verificar a qualidade da discriminação entre sujeitos saudáveis e não saudáveis usado apenas um parâmetro, *P1*, *RPPC* ou *LBST*, como tal não foi adotada a típica classificação de conjunto treino\teste. Neste caso todos os sujeitos foram comparados com um limiar calculado de modo a se obter uma taxa de erro igual para sujeitos saudáveis e não saudáveis (*EER* – *Equal Error Rate*). Os resultados para os parâmetros *P1* e *LBST* para a base de dados da USP foram já publicados em [12], e os resultados para o parâmetro *RPPC* estão publicados em [4].

### A. Análise dos Resultados do LBST na base de dados da MEEI

Os resultados apresentados de seguida para a análise da energia das bandas do sinal e *LBST* consistem na aplicação do algoritmo descrito na secção II, ponto C, para a vogal sustentada /a/, em 500 ms do sinal, com um andamento de 5 ms, resultando em 100 tramas para cada orador.

Na Fig. 2 pode-se verificar a dispersão da energia máxima da primeira banda, *FBME*, por trama e a dispersão da energia máxima da segunda banda, *SBME*. Os dados apresentados não são separáveis, no entanto existem informações relevantes a reter, nomeadamente que tendencialmente os sujeitos saudáveis têm menor energia na primeira banda do sinal comparativamente aos sujeitos não saudáveis e a dispersão dos

dados nos sujeitos não saudáveis é significativamente maior que nos sujeitos saudáveis em ambas as bandas. O primeiro caso é espectável pela análise espectral realizada na secção II. O segundo caso prende-se com o facto do ruído existente nas vozes patológicas mais deterioradas provocarem uma menor energia na primeira banda devido a dificuldades de vozeamento.

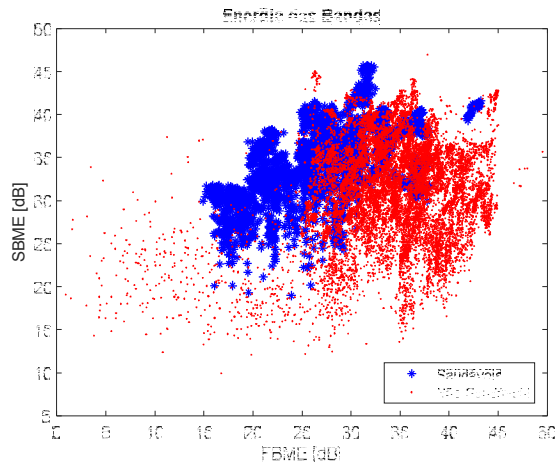


Figure 2. Energias das bandas, SBME e FBME

Na Fig. 3 são apresentados os valores do *LBST* por trama, para todos os sujeitos. Pode-se verificar que neste caso os sujeitos saudáveis têm tendencialmente valores positivos e que os sujeitos não saudáveis têm valores tendencialmente negativos e com uma maior dispersão que os sujeitos saudáveis.

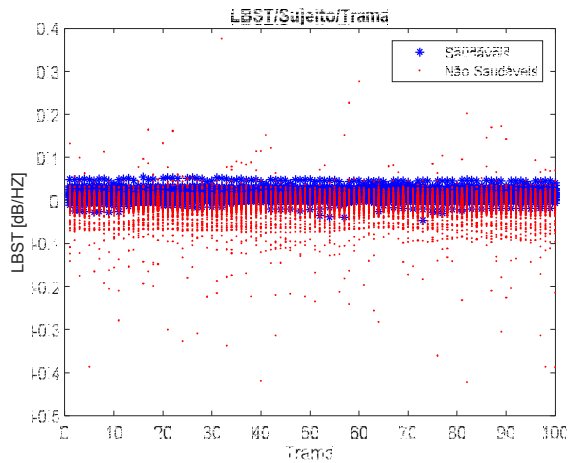


Figure 3. *LBST* para 100 tramas do sinal.

A Fig. 4 apresenta o histograma dos valores médios por orador do *LBST*. Para os sujeitos saudáveis, 51 dos 53 sujeitos apresentam um declive do *LBST* positivo, que corresponde a 96% de taxa de acerto nesta classe. Para os sujeitos não saudáveis, 96 dos 153 sujeitos apresentam um declive negativo, que corresponde a 63% de taxa de acerto nesta classe. A taxa total de acerto (*ACC* – *Accuracy*) é de 71,4%, com 147 sujeitos

diagnosticados corretamente em 206. Contudo, é notório que existe um número considerável de sujeitos não saudáveis que apresentam um valor de *LBST* positivo. De facto, para se obter uma taxa de erro igual para sujeitos saudáveis e não saudáveis o limiar de decisão do *LBST* terá de ser 0,0092 [dB/Hz]. Neste caso a taxa total de acerto é de 83,5% com 44 sujeitos saudáveis e 128 sujeitos não saudáveis diagnosticados corretamente. De notar que a estimativa da taxa de acerto tendo por base o *EER* minimiza o facto de existir um desbalançamento entre classes, neste caso um maior número de sujeitos não saudáveis comparativamente ao número de sujeitos saudáveis.

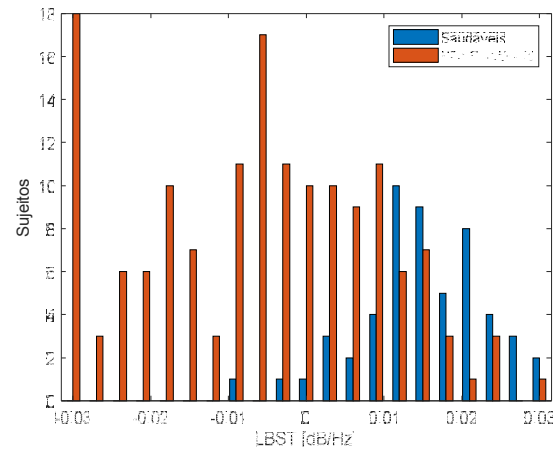


Figure 4. Sujeitos classificados por intervalo de *LBST*

#### B. Comparação dos resultados entre bases de dados

Os três parâmetros em estudo foram utilizados para medir a taxa total de acerto para as duas bases de dados. O limiar de decisão de cada um dos parâmetros foi calculado de modo a obter igual de taxa de erro (*EER*) entre sujeitos saudáveis e não saudáveis. Os resultados são apresentados na Table 2.

TABLE I. RESULTADOS COMPARATIVOS ENTRE PARÂMETROS E BASE DE DADOS.

	MEEI		USP	
	Limiar	ACC	Limiar	ACC
<i>P1</i>	560[Hz]	77,1%	300[Hz]	93,5%
<i>RPPC</i>	0,98435	83%	0,9853	54,4%
<i>LBST</i>	0,0092[db/Hz]	83,5%	0,002[db/Hz]	100%

Para a base de dados da MEEI, o *LBST* e o *RPPC* têm resultados idênticos, com 83,5% e 83% de taxa total de acerto respetivamente. Pelo contrário, utilizando *P1* a taxa total de acerto desce para 77,1%. Este resultado é esperado uma vez que esta base de dados apresenta sujeitos com patologias em estado avançado, aonde o parâmetro *RPPC* é mais fiável do que o *P1*.

Para a base de dados da USP, que apresenta sujeitos com patologias em estado inicial, o *LBST* apresenta uma taxa total de acerto de 100% e o *P1* de 93,5%. Pelo contrário, utilizando

o *RPPC* o valor deste parâmetro não desce consideravelmente para sujeitos com patologias em estado inicial e a taxa total de acerto é de apenas 54,4%.

Ao nível dos limiares de decisão verifica-se que na base de dados da MEEI, devido à maior degradação dos sinais, os limiares de *P1* e de *LBST* são maiores que na base de dados da USP. Este facto deve-se à maior dispersão dos valores destes parâmetros nos sujeitos não saudáveis devido à maior presença de ruído e consequente degradação dos sinais. Como tal o aumento do limiar tende a privilegiar a deteção de sujeitos com vozes patologias em detrimento dos sujeitos saudáveis.

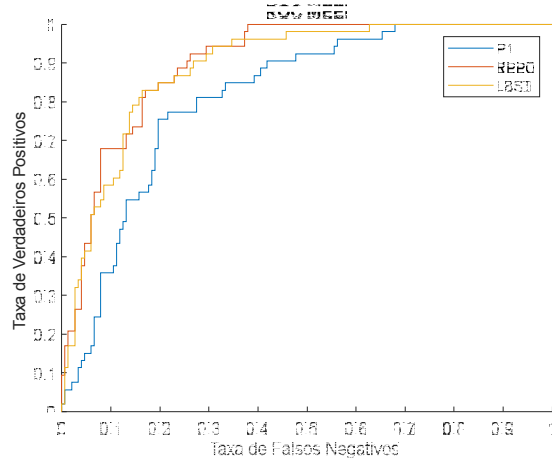


Figure 5. Curvas ROC na base de dados da MEEI,

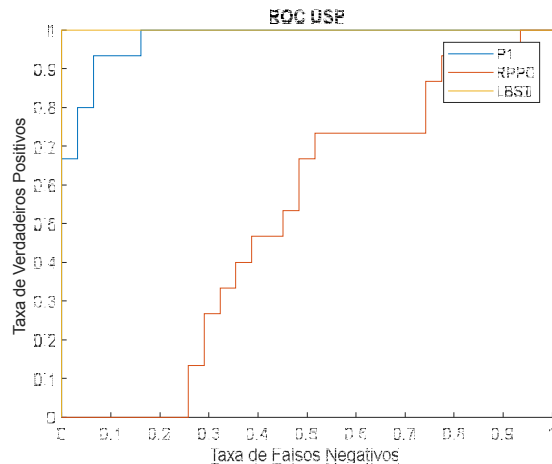


Figure 6. Curvas ROC da base de dados da USP

Nas Fig. 5 e Fig. 6 são apresentadas as ROC dos 3 parâmetros avaliados para a base de dados da MEEI e da USP, respetivamente. Estes resultados confirmam na sua generalidade os resultados mostrados na Table I. Nomeadamente é patente a melhoria dos resultados utilizando *LBST* relativamente a *P1* na base de dados da MEEI. Na ROC da base de dados da USP pode constatar-se a ineficiência do parâmetro *RPPC* e o melhor desempenho do parâmetro *LBST*. A Table II apresenta os valores da Área Abaixo da Curva ROC

(AUC – Area Under the ROC Curve) para as ROC apresentadas na Fig. 5 e Fig. 6, de onde são tiradas genericamente as mesmas conclusões.

TABLE II. AUC - ÁREA ABAIXO DAS CURVAS ROC.

	AUC	
	MEEI	USP
<i>P1</i>	0,8154	0,9763
<i>RPPC</i>	0,9013	0,51
<i>LBST</i>	0,8907	1

### C. Discussão

Pela análise dos resultados das Table I e II verifica-se que na base de dados na MEEI as taxas de acerto do *RPPC* e do *LBST* estão bastante próximas. O *RPPC* falha principalmente para sujeitos em estado inicial da patologia, devido à pouca aperiodicidade e ruído. O *LBST*, embora bastante mais robusto que *P1*, falha principalmente para sujeitos em estado mais avançado da patologia, como se pode verificar na Fig. 7.

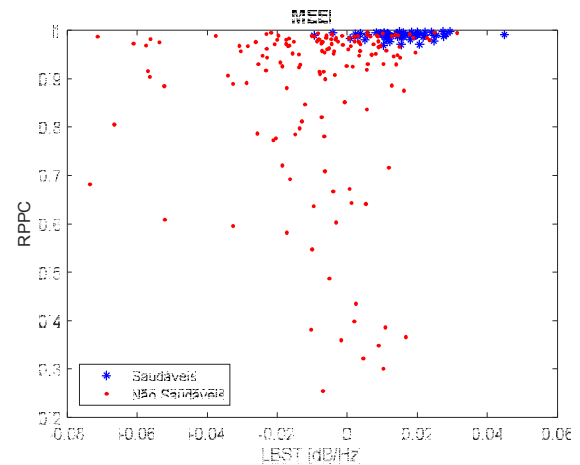


Figure 7. Valores médios de *RPPC/LBST* na base de dados da MEEI.

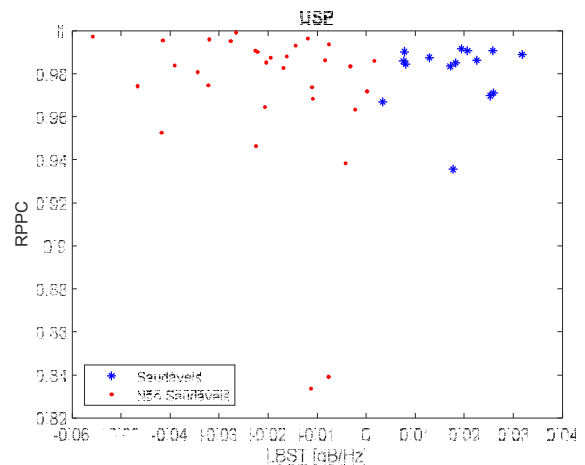


Figure 8. Valores médios de *RPPC/LBST* na base de dados da USP.

Esta situação deve-se ao facto de em alguns casos a quase ausência de vozeamento comprometer a detecção das harmónicas para o cálculo de *LBST* assim como os valores das respectivas frequências. Desde modo, otimizações no algoritmo para detecção das bandas do *LBST* estão a ser estudadas de modo a diminuir a dependência em relação à estimativa da frequência fundamental.

De notar que na base de dados da USP o *RPPC* tem uma taxa total de acerto de cerca de 54,5%, assim como uma *AUC* de cerca de 0,5. Estes dois valores indicam que nesta base dados o *RPPC*, que é um parâmetro que avalia indiretamente a relação sinal-ruído, não consegue realizar a discriminação entre sujeitos saudáveis e não saudáveis. Verifica-se nesta base de dados que não existe assim uma degradação relevante dos sinais de fala dos sujeitos não saudáveis, como se pode verificar na Fig. 8. Nestas situações, onde a degradação vocal não é severa, são favorecidos parâmetros como o *P1* e principalmente o *LBST*, sendo possível com este último distinguir todos os oradores saudáveis dos não saudáveis.

## V. CONCLUSÕES

Este artigo apresenta o parâmetro *LBST*, baseado na análise do declive entre duas bandas de baixa frequência, para caracterização de vozes patológicas na base de dados da MEEI e compara os resultados com os resultados anteriormente obtidos com a base de dados da USP. Os resultados são ainda comparados com outros dois parâmetros de análise espectral, o primeiro pico espectral, *P1*, e o *Relative Power of the Periodic Component*, *RPPC*. O parâmetro *LBST* supera os resultados dos parâmetros *P1* e *RPPC* nas duas bases de dados.

As análises dos resultados com os parâmetros apresentados apesar de não envolverem classificadores de alto nível, conseguem individualmente taxas de acerto total acima dos 77% na base de dados da MEEI. Tendo em conta que se trata de avaliações de parâmetros individuais são resultados bastante promissores. Um classificador de alto nível poderá integrar a informação destes 3 parâmetros de modo a melhorar a taxa total de acerto.

A avaliação do *LBST* na base de dados da MEEI, com uma maior degradação vocal comparativamente à da USP, permitiu também perceber algumas fragilidades deste parâmetro e verificar que existe ainda margem para a sua otimização.

Valores baixos do *RPPC* podem então ser utilizados na detecção de patologias num estado mais avançado, enquanto que a frequência e largura de banda do primeiro pico da envolvente espectral podem ser utilizados em estágios menos avançados da doença. O *LBST* apresenta-se como um método mais robusto, conseguindo caracterizar vozes patológicas independentemente do estado da doença e apresenta bons resultados nas duas bases de dados.

Este tipo de análise favorece uma maior compreensão das características e diferenças das vozes saudáveis e patológicas. Os autores julgam que este tipo de interpretação é

extremamente importante para comunidade científica, nomeadamente na mais ligada à investigação médica.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Z. Ali, M. S. Hossain, G. Muhammad, and A. K. Sangaiah, "An intelligent healthcare system for detection and classification to discriminate vocal fold disorders," *Futur. Gener. Comput. Syst.*, vol. 85, pp. 19–28, 2018.
- [2] J. P. Teixeira and A. Gonçalves, "Algorithm for Jitter and Shimmer Measurement in Pathologic Voices," *Procedia Comput. Sci.*, vol. 100, no. March, pp. 271–279, 2016.
- [3] L. C. Oliveira, D. G. Silva, and M. Andrea, "Jitter estimation algorithms for detection of pathological voices," *EURASIP J. Adv. Signal Process.*, vol. 2009, 2009.
- [4] H. T. Cordeiro, C. M. Ribeiro, H. T. Cordeiro, and C. M. Ribeiro, "Spectral envelope first peak and periodic component in pathological voices: A spectral analysis," *Procedia Comput. Sci.*, vol. 138, pp. 64–71, 2018.
- [5] K. Shama, A. Krishna, and N. U. Cholayya, "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology," *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007.
- [6] V. Majidnezhad, "A novel hybrid of genetic algorithm and ANN for developing a high efficient method for vocal fold pathology diagnosis," *EURASIP J. Audio, Speech, Music Process.*, vol. 2015, no. 1, p. 3, 2015.
- [7] M. K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, and A. Moqarehzadeh, "Identification of voice disorders using long-time features and support vector machine with different feature reduction methods," *J. Voice*, vol. 25, no. 6, 2011.
- [8] H. Cordeiro, J. Fonseca, I. Guimarães, and C. Meneses, "Hierarchical Classification and System Combination for Automatically Identifying Physiological and Neuromuscular Laryngeal Pathologies," *J. Voice*, vol. 31, no. 3, p. 384.e9-384.e14, 2017.
- [9] Z. Ali *et al.*, "Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions," *IEEE Access*, vol. 6, pp. 6961–6974, 2017.
- [10] Z. Ali, I. Elamvazuthi, M. Alsulaiman, and G. Muhammad, "Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals," *J. Med. Syst.*, vol. 40, no. 1, pp. 1–10, 2016.
- [11] H. Cordeiro and C. Meneses, "Low band continuous speech system for voice pathologies identification," *2018 Signal Process. Algorithms, Archit. Arrange. Appl.*, pp. 315–320, 2018.
- [12] H. Cordeiro and C. Meneses, "Low band spectral tilt analysis for pathological voice discrimination," *6th IEEE Port. Meet. Bioeng.*, 2019.
- [13] M. E. and E. Infirmary, *Voice disorders database, (Version 1.03 cd-rom)*. Kay Elemetrics Corp., Lincoln Park, NJ, 1994.
- [14] P. R. Scalassara, M. E. Dajer, C. D. Maciel, R. C. Guido, and J. C. Pereira, "Relative entropy measures applied to healthy and pathological voice characterization," *Appl. Math. Comput.*, vol. 207, no. 1, pp. 95–108, 2009.
- [15] R. Wayland and A. Jongman, "Acoustic correlates of breathy and clear vowels: The case of Khmer," *J. Phon.*, vol. 31, no. 2, pp. 181–201, 2003.
- [16] B. R. Gerratt, J. Kreiman, and M. Garellek, "Comparing Measures of Voice Quality From Sustained Phonation and Continuous Speech," *Am. J. Speech-Language Pathol.*, vol. 59(5), no. October, pp. 994–1001, 2016.
- [17] A. Stráňík and R. Čmejla, "an Analysis of Iterative Algorithm for Estimation of Harmonics-To-Noise Ratio in Speech," pp. 0–6.