

Avaliação da Capacidade Discriminante de Distribuições da Proeminência do Pico Cepstral para Identificar Nódulos Vocais

Larissa A. Tokuhara, Bruno C. Bispo, María E. Dajer, Edson A. R. Theodoro e Pedro M. Rodrigues

Resumo— Este trabalho tem como objetivo avaliar as distribuições de CPP e CPPS para a vogal sustentada /a/ e suas estatísticas descritivas como discriminantes entre vozes saudáveis e vozes de pacientes diagnosticados com nódulos vocais. Além de calcular essas medidas na escala decibel como normalmente realizado na literatura, este trabalho também as calculou na escala linear. Após avaliar 9 estatísticas da distribuição de 4 medidas de CPP e 4 medidas de CPPS para 35 voluntários, o 95-ésimo percentil do CPP, calculado a cada 2 ms e na escala decibel, apresentou a maior capacidade discriminante, alcançando um valor- p de $3,58e-05$, uma sensibilidade de 88,67%, especificidade de 85% e precisão de classificação igual a 85,71%.

Palavras-Chave— nódulos vocais, capacidade discriminante, proeminência do pico cepstral, CPP, CPPS.

Abstract— This work aims to investigate CPP and CPPS distributions in sustained vowel /a/ and their descriptive statistics as discriminators between healthy voices and voices from patients diagnosed with vocal fold nodule. In addition to calculating these measures in the decibel scale as normally performed in the literature, this work also calculated them in the linear scale. After evaluating 9 distribution statistics of 4 CPP and 4 CPPS measures for 35 volunteers, the 95th percentile of CPP, calculated every 2 ms and in decibel scale, presented the best discriminant capacity, achieving a p -value of $3.58e-05$, a sensitivity of 88.67%, a specificity of 85% and a classification accuracy of 85.71%.

Keywords— vocal fold nodule, discriminant capacity, cepstral peak prominence, CPP, CPPS.

I. INTRODUÇÃO

Distúrbios vocais afetam pessoas em qualquer estágio da vida e podem trazer diversas deficiências no desenvolvimento pessoal e profissional [1]. Os métodos tradicionais de avaliação vocal são o exame visual laríngeo e a análise perceptiva auditiva. O primeiro é um método invasivo e desconfortável que permite acessar diretamente as pregas vocais por meio de instrumentos de alto custo. O último é um método não-invasivo e de baixo custo onde especialistas analisam sinais vocais gravados ou ao vivo, porém extremamente dependente da experiência do especialista. Ambos os métodos são subjetivos.

Bruno C. Bispo, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil, e-mail: bruno.bispo@ufsc.br; Larissa A. Tokuhara, María E. Dajer e Edson A. R. Theodoro, Departamento de Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Cornélio Procópio-PR, Brasil, e-mail: ltokuhara@alunos.utfpr.edu.br, medajer@utfpr.edu.br, edsontheodoro@utfpr.edu.br; Pedro M. Rodrigues, Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Porto, Portugal, e-mail: prodri-gues@porto.ucp.pt.

Com o intuito de superar tais adversidades, pesquisadores começaram a estudar sinais de voz para obterem parâmetros que pudessem representar índices para diferentes aspectos da voz, incluindo patologias relacionadas com a fala [2]. A análise acústica da voz utiliza o processamento digital de sinais para quantificar, detectar e classificar características e parâmetros da qualidade vocal. Esse campo de estudo é composto por métodos essencialmente não-invasivos, de baixo custo e facilmente executáveis, resultando em vantagens no diagnóstico, tratamento e prevenção de problemas vocais [3].

Inicialmente, parâmetros para a análise acústica da voz encontravam-se no domínio do tempo, principalmente jitter e shimmer [2]. Porém, visto que dependem do conhecimento do início e fim dos ciclos vocais, há uma perda de precisão para sinais altamente perturbados [2], [4]. Além disso, para calcular esses parâmetros, é necessário que a vogal seja analisada como um pico alto e estável, pois qualquer alteração no sinal pode ser considerada como uma perturbação vocal [5].

Dessa maneira, para superar tais limitações, medidas baseadas no espectro e cepstro foram consideradas tanto para vogal sustentada quanto para fala contínua [6]. De acordo com [2], [7], parâmetros cepstrais são os índices mais promissores para detectar a severidade de disфонia na análise de sinais vocais. Em [2], estatísticas descritivas da proeminência do pico cepstral suavizado (CPPS, do inglês *cepstral peak prominence smoothed*), proposta em [8] como uma alternativa para a proeminência do pico cepstral (CPP, do inglês *cepstral peak prominence*) apresentada em [9], foram aplicadas para discriminar vozes saudáveis e patológicas.

Apesar do sucesso em detectar problemas vocais de 41 pacientes, cada um diagnosticado com um tipo de disфонia e incluindo 10 patologias diferentes, apenas 2 pacientes com nódulos vocais foram considerados. Nódulos vocais são protuberâncias bilaterais, não cancerígenas, que se formam nas pregas vocais devido a traumas em sua mucosa decorrentes do abuso ou uso incorreto da voz [10]. Desse modo, testes envolvendo mais pacientes são necessários para tal evidenciar a aplicabilidade e eficiência para diagnosticar nódulos vocais.

Este trabalho tem como objetivo avaliar as distribuições de CPP e CPPS na vogal sustentada /a/ e suas respectivas estatísticas descritivas como discriminantes entre vozes saudáveis e vozes de pacientes com nódulos vocais. Este artigo está organizado da seguinte maneira: a Seção II explica as medidas de CPP e CPPS; a Seção III descreve o banco de dados de voz utilizado; a Seção IV apresenta e discute os resultados obtidos; e, por fim, a Seção V conclui o trabalho.

II. MEDIDAS DE CPP E CPPS

A. Análise cepstral

A análise cepstral é uma técnica de análise de sinais baseada numa transformação homomórfica que resulta no cepstro. O cepstro foi proposto em 1963 como uma melhor alternativa à função de autocorrelação para detectar ecos em sinais sísmicos [11].

A representação cepstral permite que a convolução de dois sinais no domínio do tempo, portanto não-linear no domínio da frequência, seja representada como uma combinação linear no domínio do cepstro [11], [12], [13]. Devido a isto, a análise cepstral é bastante adequada para deconvolução e tem sido amplamente utilizada no processamento da fala para a detecção de pitch [12].

Existem diferentes tipos de cepstros dependendo da sua definição matemática [12]. O cepstro de potência de um sinal $x(n)$ é definido como [12]

$$c_p(n) = \left| \mathcal{F}^{-1} \left\{ \log \left[|\mathcal{F} \{x(n)\}|^2 \right] \right\} \right|^2, \quad (1)$$

onde $\mathcal{F} \{ \cdot \}$ and $\mathcal{F}^{-1} \{ \cdot \}$ denotam a Transformada de Fourier de tempo discreto (DTFT) e sua inversa, respectivamente. Na prática, uma Transformada Rápida de Fourier (FFT) de N -pontos é usada no lugar da DTFT. Embora qualquer base possa ser usada nesse cálculo, este trabalho assumiu o logaritmo natural como em [12]. No cepstro, o índice de tempo n é usualmente chamado de quefrência [12].

As referências [2], [8], [9] afirmam que o cepstro é o espectro de potência logarítmico de um espectro de potência logarítmico, mas as definições matemáticas não foram fornecidas. Combinando a definição textual e os resultados gráficos que mostram o cepstro de sinais de voz com valores apenas positivos e iguais a dezenas de decibels, pode-se inferir que em [2], [8], [9] o cepstro é definido como

$$c_h(n) = \left| \mathcal{F} \left\{ \log \left[|\mathcal{F} \{x(n)\}|^2 \right] \right\} \right|^2. \quad (2)$$

Entretanto, em casos práticos onde uma FFT de N -pontos é utilizada no cálculo, pode-se demonstrar que

$$c_h(n) = c_p(n) \times N^2. \quad (3)$$

Portanto, as medidas CPP e CPPS utilizando $c_p(n)$ e $c_h(n)$ serão proporcionais e resultarão na mesma precisão de diagnóstico. Assim, neste trabalho, optou-se por utilizar $c_p(n)$.

Para sinais quasi-periódicos, como sons de vogais, o cepstro apresenta um pico proeminente na quefrência correspondente ao período fundamental do sinal [9], [12], que é explorado nas medidas de CPP e CPPS.

B. Proeminência do Pico Cepstral

A proeminência do pico cepstral (CPP, do inglês *cepstral peak prominence*) é uma medida da amplitude do primeiro pico cepstral, que é localizado entre os valores mínimo e máximo do período fundamental esperado [8], normalizada em relação a amplitude de todo o cepstro [9]. E, portanto, representa quanto o primeiro pico cepstral emerge do cepstro.

A idéia por trás da medida de CPP é que um sinal quasi-periódico possui uma estrutura harmônica mais bem definida

e, conseqüentemente, um pico cepstral mais proeminente do que um sinal menos periódico. No entanto, a amplitude do pico cepstral é afetada não somente pela periodicidade mas também pela energia e tamanho da janela de análise cepstral [9]. Por isso, torna-se necessária uma medida da proeminência do pico cepstral ao invés da sua amplitude absoluta [9].

Nesse sentido, em [9], foi proposto uma normalização do pico cepstral através de uma regressão linear calculada relacionando quefrência e magnitude cepstral. E o CPP é a diferença em amplitude entre o primeiro pico cepstral e o valor na regressão linear que está diretamente abaixo do pico [9].

Em [2], [8], [9], a diferença supracitada foi calculada na escala decibel (dB) pois o $c_h(n)$, assim como o $c_p(n)$, é real positivo. Neste trabalho, para efeito de estudos, o CPP foi calculado não só na escala dB mas também na escala linear.

Neste trabalho, como em [2], [8], [9], o primeiro pico cepstral foi procurado de 3,3 a 16,7 ms devido à frequência fundamental em adultos geralmente ser entre 60 e 300 Hz [2]. E, como também em [2], [8], [9], quefrências abaixo de 1 ms foram desconsideradas na regressão linear porque elas são mais afetadas pelo envelope espectral, que varia lentamente, do que pela periodicidade do espectro [14].

Além disso, como em [8], [9], o CPP foi calculado a cada $T = 10$ ms. E, para efeito de comparação, ele também foi calculado a cada $T = 2$ ms como sugerido em [2], [8] para o CPPS. Em ambos os casos, as medidas de CPP foram calculadas utilizando uma janela de análise retangular de 46,4 ms (1024 amostras) e uma FFT de $N = 1024$ pontos.

C. Proeminência do Pico Cepstral Suavizado

A Proeminência do pico cepstral suavizado (CPPS, do inglês *cepstral peak prominence smoothed*) foi proposta como uma modificação do CPP, consistindo de 2 processos de suavização do cepstro antes do cálculo do CPP [8].

Primeiramente, médias são realizadas ao longo do tempo, isto é, o cepstro da k -ésima janela é substituído pela média do cepstro da $(k - \frac{L-1}{2})$ à $(k + \frac{L-1}{2})$ -ésima janela. Posteriormente, médias são realizadas ao longo da quefrência, isto é, a l -ésima amostra do cepstro da k -ésima janela é substituída pela média da $(l - \frac{M-1}{2})$ à $(l + \frac{M-1}{2})$ -ésima amostra da k -ésima janela.

Neste trabalho, como em [2], [8], as medidas de CPPS foram calculadas a cada $T = 2$ ms. E, para efeito de comparação, o procedimento de suavização foi realizado utilizando $L = M = 7$ como em [2] e $L = M = 11$ como em [8].

Além disso, como feito para o CPP, o CPPS foi calculado em escala linear e em dB utilizando uma janela de análise retangular de 46,4 ms (1024 amostras) e uma FFT de $N = 1024$ pontos, o primeiro pico cepstral foi procurado de 3,3 a 16,7 ms, e quefrências abaixo de 1 ms não foram consideradas na regressão linear.

III. BASE DE DADOS DE VOZ

A base de dados de voz foi fornecido pelo Grupo de Pesquisa em Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEM/CNPq).

TABELA I: Resumo do gênero e idade do banco de dados.

	Gênero		Idade	
	Masculino	Feminino	Média	Desvio Padrão
Controles	12	8	31,31	9,26
Pacientes	2	13	35,87	7,54
Total	14	21	33,52	8,65

A base de dados consiste em 35 gravações, 20 pertencentes a indivíduos saudáveis (grupo de controle) e 15 a pacientes diagnosticados com nódulos vocais (grupo de pacientes), da vogal sustentada /a/. A vogal /a/ foi escolhida devido à sua ampla utilização na análise acústica da voz [2]. As gravações foram realizadas no Ambulatório de Voz do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) utilizando uma frequência de amostragem de $f_s = 22050$ Hz.

As informações sobre o gênero e idade dos voluntários são apresentadas na Tabela I. A duração dos sinais de voz saudáveis tem média de 2,85 s e desvio padrão de 0,15 s, enquanto a duração dos sinais patológicos tem média de 2,57 s e desvio padrão de 0,38 s. Como recomendado em [15], a duração de cada fonema foi sempre superior a 2 s.

O nível de potência ativa de cada sinal foi normalizado para -26 dBov de acordo com o algoritmo da Recomendação ITU-T P.56 [16].

IV. RESULTADOS

Séries temporais dos valores de CPP e CPPS foram obtidas para cada gravação, utilizando $c_p(n)$ e os parâmetros descritos nas seções II-B e II-C, e tratadas como distribuições. Similarmente ao feito em [2], as seguintes estatísticas descritivas foram calculadas para as distribuições de CPP e CPPS: média (CPP_{mean} e $CPPS_{\text{mean}}$), mediana (CPP_{median} e $CPPS_{\text{median}}$), módulo (CPP_{mode} e $CPPS_{\text{mode}}$), quinto percentil ($CPP_{5\text{prc}}$ e $CPPS_{5\text{prc}}$) e o 95-ésimo percentil ($CPP_{95\text{prc}}$ e $CPPS_{95\text{prc}}$) como medidas de localização da distribuição; desvio padrão (CPP_{std} e $CPPS_{\text{std}}$) e o intervalo entre o valor máximo e o mínimo (CPP_{range} e $CPPS_{\text{range}}$) como medidas de variância; assimetria (CPP_{skew} e $CPPS_{\text{skew}}$) e curtose (CPP_{kurt} e $CPPS_{\text{kurt}}$) para a caracterização do formato de distribuição.

O teste de Kruskal-Wallis foi utilizado para determinar se as estatísticas dos grupos de controle e patológico são oriundos da mesma distribuição. Os valores- p resultantes são apresentados na Tabela II, onde o menor valor para cada

configuração de medidas de CPP e CPPS está destacado. Vale a pena ressaltar que valores- p menores que 0,05 indicam que a hipótese nula é rejeitada, isto é, existe uma diferença significativa entre as médias das distribuições e portanto as estatísticas dos grupos de controle e patológico originam de distribuições diferentes [17].

A partir da Tabela II, pode-se observar que, para os mesmos parâmetros (tempo e escala), o menor valor- p foi obtido pelas medidas de CPP, o que indica que ele pode ser mais discriminante que o CPPS. E a melhor estatística descritiva foi diferente para o CPP e o CPPS, sendo $CPP_{95\text{prc}}$ e CPP_{std} para CPP enquanto $CPPS_{\text{mean}}$, $CPPS_{\text{mode}}$ e $CPPS_{5\text{prc}}$ para CPPS. Além disso, pode-se notar que, em comparação com a escala linear, os resultados obtidos utilizando a escala dB são muito inferiores para CPPS e levemente superiores para CPP.

Com o intuito de averiguar a eficiência das estatísticas descritivas explicadas anteriormente como discriminantes entre vozes saudáveis e patológicas, um classificador OB-ROC foi utilizado. O OB-ROC é um classificador binário que usa a distância quadrática entre médias de distribuições para encontrar uma boa classificação para discriminar classes [18]. Ele também utiliza a curva ROC (do inglês *Receiver Operating Characteristic*) para analisar o resultado da classificação [18]. As precisões do classificador são apresentadas na Tabela III, onde o maior valor para cada configuração do CPP e do CPPS está destacado. Em caso de duas ou mais estatísticas resultarem na mesma precisão de classificação, destacou-se aquela que apresentou a maior área da curva ROC (AROC).

Em [2], a melhor estatística descritiva para sinais gravados com microfone no ar foi o $CPPS_{5\text{prc}}$. Neste trabalho, considerando o CPPS calculado com $T = 2$ ms, $L = M = 7$ e escala dB como em [2], a melhor estatística descritiva também foi o $CPPS_{5\text{prc}}$, como pode ser observado na Tabela II. O *boxplot* do $CPPS_{5\text{prc}}$ para cada grupo é apresentado na Figura 1a. Pode-se observar que não há uma separação clara entre os grupos, o que é confirmado por um p -valor de $4,92e-02$ e uma AROC de 0,6967 (muito menor que o 0,95 obtido em [2]). Os valores do $CPPS_{5\text{prc}}$ para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 1b. O limiar ficou entre 11,07 e 11,09 dB, alcançando uma sensibilidade de 93,33%, uma especificidade de 50% e uma precisão de 68,57%.

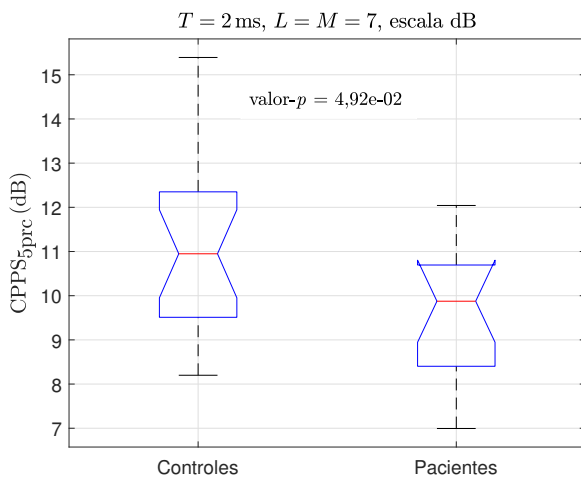
Neste trabalho, a melhor estatística descritiva foi o $CPP_{95\text{prc}}$ utilizando $T = 2$ ms e escala dB, como pode ser observado nas Tabelas II e III. O *boxplot* do $CPP_{95\text{prc}}$ para cada grupo é apresentado na Figura 2a. Pode-se observar que existe uma separação entre os grupos mais nítida, comparativamente ao

 TABELA II: Valores- p para cada estatística das distribuições.

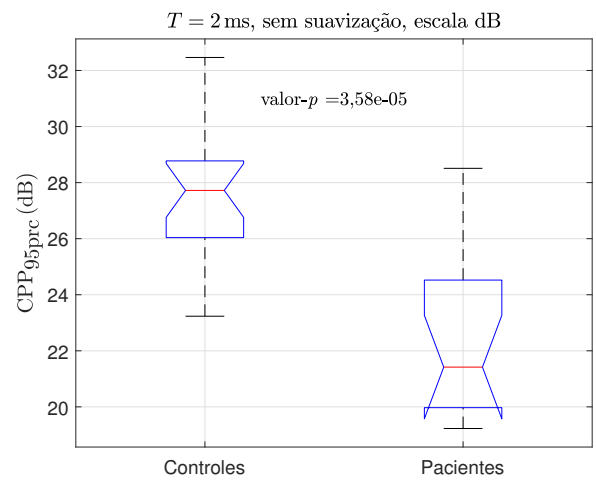
	Parâmetros CPP and CPPS			Valor- p das Estatísticas das Distribuições								
	Tempo	Suavização	Escala	média	mediana	módulo	5prc	95prc	desvio	intervalo	assimetria	curtose
CPP	$T = 10$	—	dB	4,77e-05	6,33e-05	5,50e-05	5,50e-05	4,13e-05	7,14e-01	2,86e-01	2,34e-02	3,35e-03
	$T = 2$	—	dB	4,77e-05	5,50e-05	1,65e-04	4,77e-05	3,58e-05	7,14e-01	5,49e-01	3,57e-02	1,96e-02
	$T = 10$	—	linear	1,10e-04	1,45e-04	7,66e-03	2,70e-03	4,77e-05	4,13e-05	1,45e-04	4,92e-02	3,03e-02
	$T = 2$	—	linear	1,45e-04	1,65e-04	9,32e-03	2,16e-03	4,77e-05	7,29e-05	1,22e-03	1,42e-01	5,74e-02
CPPS	$T = 2$	$L = M = 7$	dB	8,30e-02	8,30e-02	1,82e-01	4,92e-02	1,42e-01	1,34e-01	9,20e-01	5,27e-01	7,39e-01
	$T = 2$	$L = M = 11$	dB	7,73e-02	6,19e-02	5,32e-02	2,78e-02	9,56e-02	1,25e-01	8,94e-01	7,64e-01	3,34e-01
	$T = 2$	$L = M = 7$	linear	1,10e-04	1,10e-04	3,62e-04	3,62e-04	1,26e-04	7,61e-04	5,66e-03	8,16e-01	4,24e-01
	$T = 2$	$L = M = 11$	linear	9,62e-05	9,62e-05	1,10e-04	2,16e-04	1,26e-04	1,54e-03	8,46e-03	9,20e-01	3,86e-01

TABELA III: Precisão de classificação para cada estatística das distribuições.

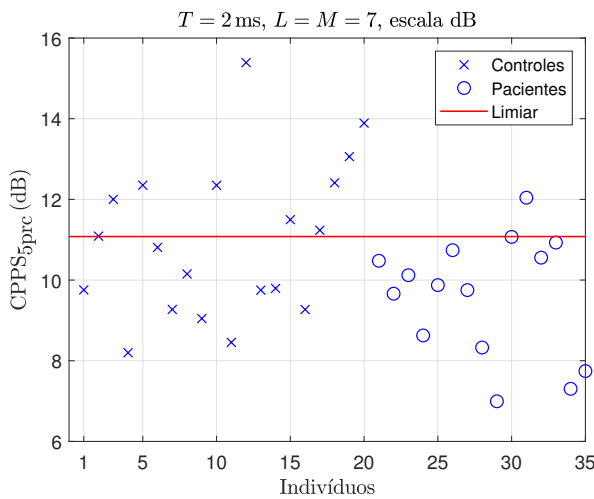
	Parâmetros CPP and CPPS			Precisão de Classificação das Estatísticas das Distribuições (%)								
	Tempo	Suavização	Escala	média	mediana	módulo	5prc	95prc	desvio	intervalo	assimetria	curtose
CPP	$T = 10$	—	dB	85,71	85,71	82,86	82,86	85,71	60,00	62,86	74,29	77,14
	$T = 2$	—	dB	85,71	85,71	85,71	85,71	85,71	60,00	68,57	77,14	80,00
	$T = 10$	—	linear	82,86	80,00	77,14	77,14	82,86	82,86	82,86	68,57	71,43
	$T = 2$	—	linear	80,00	80,00	74,29	77,14	82,86	82,86	80,00	71,43	71,43
CPPS	$T = 2$	$L = M = 7$	dB	65,71	65,71	65,71	68,57	68,57	71,43	60,00	65,71	65,71
	$T = 2$	$L = M = 11$	dB	71,43	71,43	65,71	71,43	68,57	71,43	62,86	62,86	68,57
	$T = 2$	$L = M = 7$	linear	82,86	82,86	80,00	80,00	82,86	82,86	74,29	60,00	60,00
	$T = 2$	$L = M = 11$	linear	82,86	82,86	85,71	82,86	82,86	80,00	74,29	60,00	65,71



(a)

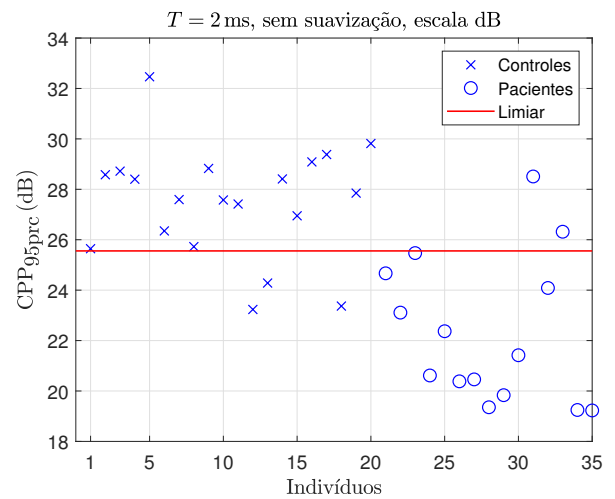


(a)



(b)

Fig. 1: CPPS_{5prc} utilizando $T = 2$ ms, $L = M = 7$ e escala dB: (a) *boxplot* para cada grupo; (b) valores para cada grupo e limiar de classificação.



(b)

Fig. 2: CPP_{95prc} utilizando $T = 2$ ms e escala dB: (a) *boxplot* para cada grupo; (b) valores para cada grupo e limiar de classificação.

caso anteriormente discutido, o que é confirmado por um p -valor de $3,58e-05$ e uma AROC de 0,9133 (próxima da 0,95 obtida em [2] para CPPS_{5prc} utilizando $T = 2$ ms, $L = M = 7$ e escala dB). Os valores de CPP_{95prc} para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 2b. O limiar ficou entre 25,47 and 25,64 dB, resultando numa sensibilidade de 88,67%, especificidade de 85% e precisão de classificação igual a 85,71%.

O melhor desempenho geral do CPP, em comparação com o CPPS, indica que a suavização realizada no CPPS não melhora a classificação de nódulos vocais. Tal suavização tende a reduzir a variação do pico cepstral normalizado ao longo do tempo. Esse fato pode contribuir para a baixa capacidade discriminante do CPPS.

Deve-se enfatizar que o trabalho apresentado em [2] investigou estatísticas capazes de discriminar vozes disfônicas (10 tipos de doenças) e saudáveis. Além de não focar em nódulos vocais, apenas 2 indivíduos com essa patologia foram considerados. Embora o CPPS_{5prc}, calculado utilizando $T = 2$ ms, $L = M = 7$ e escala dB, tenha apresentado bom resultado na detecção de vozes patológicas, essa pequena quantidade de dados não permite a generalização da sua eficácia para a detecção de nódulos vocais. Neste trabalho, onde 15 pacientes com nódulos vocais foram incluídos na avaliação, essa estatística descritiva não se mostrou eficiente e foi superada pelo CPP_{95prc} calculado utilizando $T = 2$ ms e escala dB.

V. CONCLUSÃO

Este trabalho investigou distribuições de CPP e CPPS na vogal sustentada /a/ e suas estatísticas descritivas como discriminantes entre vozes saudáveis e vozes de pacientes diagnosticados com nódulos vocais. Além de calcular essas medidas na escala decibel como normalmente realizado na literatura, este trabalho também as calculou na escala linear.

Após avaliar 9 distribuições estatísticas de 4 medidas de CPP e 4 medidas de CPPS utilizando uma base de dados contendo 35 gravações, 20 pertencentes a indivíduos saudáveis e 15 a pacientes com nódulos vocais, foi demonstrado que o CPP apresenta uma maior capacidade discriminante que o CPPS. Isso indica que, apesar de trabalhos anteriores afirmarem uma eficiência do CPPS em distinguir vozes saudáveis e patológicas, o procedimento de suavização realizado no CPPS não melhora a classificação de nódulos vocais.

O 95-ésimo percentil do CPP, calculado a cada 2 ms e na escala dB, apresentou a maior capacidade discriminante, alcançando um valor- p de $3,58e-05$, uma sensibilidade de 88,67%, especificidade de 85% e precisão de classificação igual a 85,71%.

AGRADECIMENTOS

Os autores gostariam de agradecer o Grupo de Pesquisa em Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEM/CNPq) por fornecer a base de dados de voz.

REFERÊNCIAS

- [1] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, "Voice disorders in the general population: prevalence, risk factors, and occupational impact," *The Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, November 2005.
- [2] A. Castellana, A. Carullo, S. Corbellini, and A. Astolfi, "Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 646–654, March 2018.
- [3] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological voice: sustained vowels versus continuous speech," *Journal of Speech Language and Hearing Research*, vol. 44, no. 2, pp. 327–339, May 2001.
- [4] S. Bielamowicz, J. Kreiman, B. R. Gerratt, M. S. Dauer, and G. S. Berke, "Comparison of voice analysis systems for perturbation measurement," *Journal of Speech Language and Hearing Research*, vol. 39, no. 1, pp. 126–124, February 1996.
- [5] Y. Zhang and J. J. Jiang, "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," *Journal of Voice*, vol. 22, no. 1, pp. 1–9, January 2008.
- [6] S. Y. Lowell, R. H. Colton, R. T. Kelley, and Y. C. Hahn, "Spectral- and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker," *Journal of Voice*, vol. 25, no. 5, pp. 223–232, September 2011.
- [7] S. Y. Lowell, R. H. Colton, R. T. Kelley, and S. A. Mizia, "Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech," *Journal of Voice*, vol. 27, no. 4, pp. 393–400, July 2013.
- [8] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech," *Journal of Speech and Hearing Research*, vol. 39, pp. 311–321, April 1996.
- [9] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech and Hearing Research*, vol. 37, pp. 769–778, August 1994.
- [10] R. B. Nunes, M. Behlau, M. B. Nunes, and J. G. Paulino, "Clinical diagnosis and histological analysis of vocal nodules and polyps," *Brazilian Journal of Otorhinolaryngology*, vol. 79, no. 4, pp. 434–440, June 2013.
- [11] J. M. Tribolet, *Seismic Applications of Homomorphic Signal Processing*. Englewood Cliffs, New Jersey: Prentice Hall, 1979.
- [12] J. R. Deller Jr., J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. Piscataway, New Jersey: IEEE Press, 2000.
- [13] A. V. Oppenheim and R. W. Schaffer, "From frequency to quefrequency: A history of the cepstrum," *IEEE Signal Processing Magazine*, vol. 21, pp. 95–106, September 2004.
- [14] G. de Krom, "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals," *Journal of Speech and Hearing Research*, vol. 36, no. 2, pp. 254–266, April 1993.
- [15] R. F. Coleman, "Sources of variation in phonetograms," *Journal of Voice*, vol. 7, no. 1, pp. 1–14, March 1993.
- [16] ITU-T P.56, "Objective measurement of active speech level," International Telecommunications Union, Geneva, Switzerland 2011.
- [17] E. Ostertagová, O. Ostertag, and J. Kováč, "Methodology and application of the Kruskal-Wallis test," *Applied Mechanics and Materials*, vol. 611, pp. 115–120, October 2014.
- [18] C. T. Nakas and C. T. Yiannoutsos, "Ordered multiple-class ROC analysis with continuous measurements," *Statistics in Medicine*, vol. 23, no. 22, pp. 3437–3449, November 2004.