Índice de Não Estacionariedade aplicado à Classificação de Desordens Vocais

Vinicus J. D. Vieira, Silvana C. Costa e Suzete E. N. Correia

***Resumo*— Este artigo apresenta um estudo sobre a classificação de vozes saudáveis e vozes afetadas por patologias laríngeas usando, como características acústicas, medidas do Índice de Não Estacionariedade (INS). O INS é uma abordagem tempo-frequência que, além de testar a estacionariedade de sinais, quantifica-a por meio de diferentes escalas de observação. Nos experimentos são consideradas dez escalas. Como classificador, é utilizada a Análise Discriminante Linear (*Linear Discriminant Analysis* – LDA). As patologias laríngeas consideradas são: paralisia, edema e nódulos nas pregas vocais. O desempenho do classificador atinge uma acurácia de 91,82% com a combinação de nove escalas de observação. Os resultados indicam que o INS é uma medida promissora na caracterização de distúrbios da voz.**

***Palavras-Chave*— Processamento de sinais de voz, Classificação de patologias laríngeas, Índice de não estacionariedade.**

***Abstract*— This article presents a study on the classification of healthy voices and voices affected by laryngeal pathologies using measures of the Index of Non-Stationarity (INS) as acoustic characteristics. The INS is a time-frequency approach that makes a stationarity test of signals, and also quantifies it through different observation scales. In the experiments, ten scales are considered. As a classifier, it is used the Linear Discriminant Analysis (LDA). The laryngeal pathologies considered are: paralysis, edema and nodules in the vocal folds. The classifier performance reaches an accuracy of 91.82% with the combination of nine observation scales. The results indicate that the INS is a promising measure in the characterization of voice disorders.**

***Keywords*— Speech signal processing, Laryngeal pathologies classification, Index of non stationarity.**

I. Introdução

O sinal de voz é um processo resultante de um complexo sistema de produção que envolve fatores neurológicos e fisiológicos [1]. Tal forma de onda carrega informações que não estão restritas apenas ao conteúdo linguístico. É possível, também, verificar a identidade do falante e sua saúde [2].

Diversas aplicações envolvendo processamento digital de sinais de voz têm sido estudadas ao longo das últimas décadas, por ser a fala um dos principais meios de comunicação do ser humano [3–7]. Sistemas de reconhecimento de fala, por exemplo, são utilizados nos dias atuais em *smartphones* para facilitar a interface homem-máquina e otimizar processos como pesquisa e digitação de longos textos [8], [9]. Outras propostas comuns na literatura estão imersas em contextos como reconhecimento de locutor [4], [10], reconhecimento de emoções [11], [12] e detecção de distúrbios vocais [7], [13]. Esta última aplicação é importante na prática clínica de profissionais de saúde, como fonoaudiólogos e demais especialistas da área, pois permite a triagem, o diagnóstico e o acompanhamento de pacientes de forma presencial ou remota (por meio de mecanismos de Telessaúde) [2], [14].

A análise de distúrbios vocais pode estar relacionada a aspectos patológicos ou hiperfuncionais. Em relação às patologias, estas podem ter origem neurológica ou fisiológica, fatores estes que se subdividem em diferentes tipos de condições patológicas [15]. No contexto dos aspectos hiperfuncionais, as disfonias podem ser analisadas em relação ao grau do desvio fonatório [2], ou, ainda, em relação a fatores como tensão, rugosidade e soprosidade na emissão sonora [16]. A separação entre um sinal de voz considerado saudável e um sinal de voz considerado patológico (ou detecção de algum tipo de disfonia) depende da robustez da medida acústica empregada para capturar informações da forma de onda.

Na literatura são encontrados trabalhos que propõem medidas acústicas baseadas em um modelo linear de produção da fala [15], [17] e trabalhos que propõem medidas baseadas em um modelo não linear [7], [18], [19]. O modelo linear considera o sistema de produção vocal como sendo um sistema fonte-filtro [20]. Por outro lado, o modelo não linear considera a produção da fala como um sistema dinâmico não linear, sujeito a comportamento caótico [2].

Em geral, as medidas da análise linear são extraídas em trechos considerados estacionários do sinal de voz, compreendidos entre 16 ms e 40 ms [15]. Em contrapartida, características não lineares, a exemplo das medidas de quantificação de recorrência, não possuem o pré-requisito da estacionariedade, o que permite a análise em trechos de duração superior a 40 ms [18]. Assim, a escolha do tamanho do trecho do sinal, do qual são extraídas as medidas acústicas, tem influência do tipo de análise empregada nele. Quanto mais adequada for a análise, mais robusto tende a ser o sistema homem-máquina de reconhecimento de desordens vocais.

Apesar de existirem diversas pesquisas que investiguem a presença de padrões acústicos nas disfonias, ainda não há um consenso sobre qual medida acústica pode ser utilizada universalmente para discriminar sinais considerados saudáveis de sinais oriundos de distúrbios patológicos ou hiperfuncionais. Além disso, há trabalhos que levam em consideração a combinação de características para atingir um melhor desempenho de classificação [7], [13], [18].

A escolha de medidas acústicas confiáveis que representem cada tipo de disfonia pode se tornar uma tarefa difícil, uma vez que depende de vários fatores como o tipo e o grau da lesão, a severidade dos efeitos causados pela desordem vocal e a quantidade de ruído presente no sinal de voz analisado [2], [17], [18]. Embora diversas medidas aproximem a voz a

Vinicius J. D. Vieira, Silvana C. Costa e Suzete E. N. Correia, Programa de Pós-Graduação em Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (PPGEE/IFPB). E-mails: {viniciusjdv@gmail.com, silvana@ifpb.edu.br, suzete@ifpb.edu.br}. Este trabalho foi financiado pela Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ-PB).

processos estacionários por meio de métodos de segmentação e janelamento, a natureza deste tipo de sinal é não estacionária.

A consideração de que sinais de voz são estacionários em curtos intervalos de tempo [21] é originalmente concebida para sinais diagnosticados como normais (ou saudáveis). Os estudos que avaliam desordens vocais fazem a extração de medidas a curto intervalo de tempo por padrão para todas as classes, sem se preocupar se os sinais disfônicos realmente podem ser considerados estacionários no mesmo intervalo de tempo que os sinais saudáveis. Estudos [2], [18] mostraram que a condição patológica do sistema de produção vocal pode introduzir ruído na emissão sonora. Isto é um indício de que o padrão de estacionariedade (ou não estacionariedade) de sinais patológicos pode ser diferente do convencional para sinais saudáveis.

Pesquisas recentes têm empregado uma medida chamada Índice de Não Estacionariedade (INS – *Index of Non-Stationarity*) [22], que realiza o teste de não estacionariedade e ainda fornece o seu grau, a fim de observar como sinais acústicos de diferentes classes se comportam sob este ponto de vista [6], [12]. Em um trabalho observando a influência acústica de sinais ruidosos na inteligibilidade da fala [6] foram observadas diferenças do INS entre sinais de ruído de balbucio, motosserra, fábrica e britadeira. Outro trabalho [12] apresentou diferenças de INS entre sinais de voz com variações emocionais e, ainda, empregou esta medida no processo de classificação, aprimorando o desempenho do classificador.

Este estudo tem duas contribuições relevantes: 1) a aplicação do INS em distúrbios da voz para fins de classificação de patologias na laringe; e 2) a verificação da robustez do INS como medida acústica em sinais sem variações de fala (apenas a emissão da vogal sustentada).

O restante deste trabalho está organizado da seguinte forma: Na Seção II são apresentadas brevemente as definições relacionadas ao INS. Na Seção III é apresentada a metodologia empregada neste estudo. Na Seção IV são apresentados os resultados obtidos nos experimentos realizados e, na Seção V, são apresentadas as considerações finais.

II. Índice de Não Estacionariedade

O INS é uma medida tempo-frequência que analisa de forma objetiva a não estacionariedade de um sinal [22]. Na proposta original do INS, um sinal é definido estacionário em relação a uma escala de observação se o seu espectro local de tempo curto em diferentes instantes de tempo for estatisticamente similar ao seu espectro global. Na Figura 1 são apresentados espectrogramas globais (800 ms) e locais (40 ms) obtidos de sinais de voz da vogal sustentada /a/ de duas classes: laringe saudável e laringe patológica (paralisia, cuja fonação com menos intensidade resulta em um sinal de menor amplitude). A não estacionariedade é detectada à medida em que o espectrograma local diverge estatisticamente do espectrograma global.

O teste de estacionariedade é realizado pela comparação de componentes espectrais do sinal com referenciais estacionários, chamados *surrogates*, obtidos do próprio sinal [22]. Para tanto, os espectrogramas do sinal e dos *surrogates* são obtidos por meio da Transformada de Fourier de Tempo Curto (STFT - *Short Time Fourier Transform*). A distância Kullback-Leibler (KL) [23] é aplicada para medir a divergência entre o espectro de tempo curto do sinal analisado e seu espectro global, bem como a diferença entre cada *surrogate* e seu respectivo espectro global.

Para o cálculo do INS, seja $D_n^{(x)}$ a divergência do espectrograma do sinal analisado em diferentes escalas de tempo $t_n(n = 1, ..., N)$. De maneira similar, seja $D_n^{(s_j)}$ a distância KL medida entre os espectrogramas do $j-$ésimo *surrogate* ($n = 1, ..., N$; $j = 1, ..., J$). Neste trabalho, são consideradas

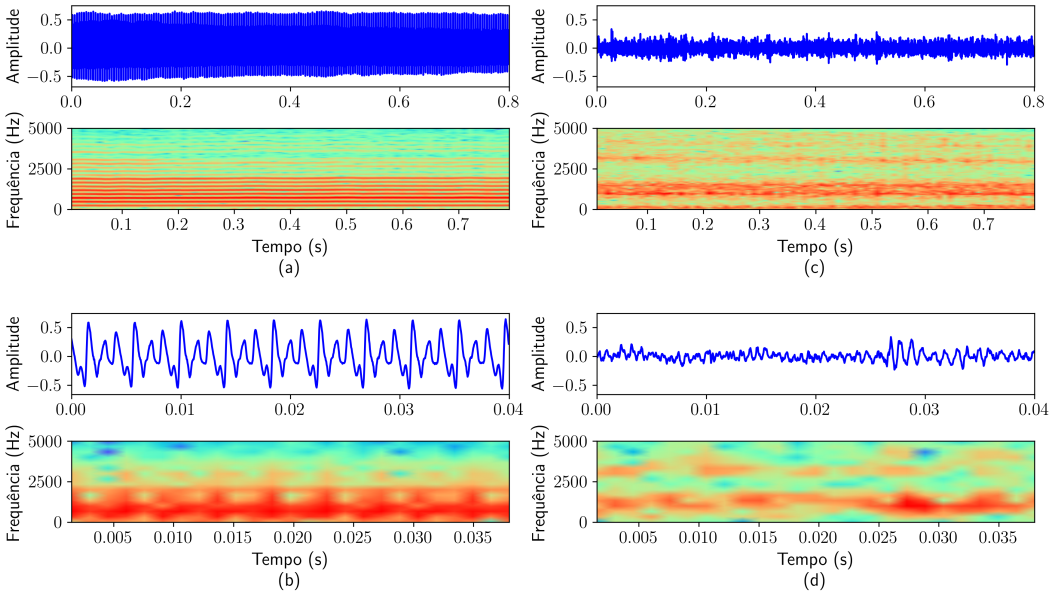

Fig. 1. Sinal de voz da vogal sustentada /a/ e seu espectrograma considerando trechos de 800 ms e 40 ms, respectivamente, para as classes saudável ((a) e (b)) e patológico ((c) e (d)).

10 escalas de observação ($N = 10$) e 50 *surrogates* ($J = 50$). A variância calculada a partir dos valores de divergência é dada por:

$$\begin{cases} \Theta_0(j) = \text{var}\left(D_n^{(s_j)}\right)_{n=1,\ldots,N.}, & j = 1, ..., J. \\ \Theta_1 = \text{var}\left(D_n^{(x)}\right)_{n=1,\ldots,N.} \end{cases} \quad (1)$$

Finalmente, o INS é dado por:

$$\text{INS} := \sqrt{\frac{\Theta_1}{\langle \Theta_0(j) \rangle}}, \quad (2)$$

em que $\langle \cdot \rangle$ é o valor médio de $\Theta_0(j)$. Na proposta do INS [22], os autores consideram que a distribuição dos valores da distância KL são aproximados por uma distribuição Gamma. Por isso, para cada escala de tempo T_h, um limiar γ, com 95% de precisão, pode ser definido para o teste de estacionariedade. Desta forma, o sinal é considerado não estacionário se o valor de INS estiver acima deste limiar. Ou seja,

$$\text{INS} \begin{cases} \leq \gamma & \text{, sinal estacionário;} \\ > \gamma & \text{, sinal não estacionário.} \end{cases} \quad (3)$$

Exemplos de INS são apresentados na Figura 2. Os valores de INS são obtidos em diferentes escalas de observação, considerando um sinal de voz saudável e três sinais patológicos: paralisia, edema e nódulo. A escala de tempo T_h/T indica a relação entre o tamanho adotado para a observação local (T_h) e o tamanho total do sinal ($T = 800$ ms). Os valores de INS são denotados por linhas vermelhas, enquanto as linhas verdes indicam o limiar de estacionariedade. O comportamento do sinal saudável apresenta um crescimento nos valores de INS após a quarta escala de observação (80 ms, T_h/T=0.1), em que o teste aponta não estacionariedade. O sinal de paralisia mostra-se não estacionário em todas as escalas de observação, o que pode ter sido causado pela severidade da patologia, que é de origem neurológica. As duas patologias de origem orgânica (edema e nódulo) apresentam comportamento semelhante em relação à variação do INS ao longo das escalas. Contudo, apesar de serem sinais de vogal sustentada, com pouca variação acústica, diferenças são observadas entre sinais saudáveis e patológicos.

III. Metodologia

Neste trabalho, é utilizada a *Disordered Voice Database*, *Model 4337*, desenvolvida pelo *Kay Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* [24]. Desta base, são analisados 167 sinais de voz (vogal sustentada */a/*), sendo 53 sinais de vozes saudáveis e 114 sinais de vozes afetadas por patologias laríngeas (52 sinais de vozes afetadas por paralisia nas pregas vocais, 44 sinais de vozes afetadas por edema de Reinke, e 18 sinais de vozes afetadas por nódulos vocais.

De cada sinal, foi utilizado um trecho de 800 ms para a extração do INS. Esta escolha tem como razão a padronização das escalas do INS, visto que nem todos os sinais da base de dados têm o mesmo tamanho. Na Tabela I são apresentados os valores de fator de escala aplicados ao trecho de 800 ms dos sinais, a fim de se obter o INS de diferentes tamanhos de segmento. Dessa forma, o valor de INS obtido de cada escala (tamanho de segmento) é considerado no processo de classificação como sendo uma medida acústica.

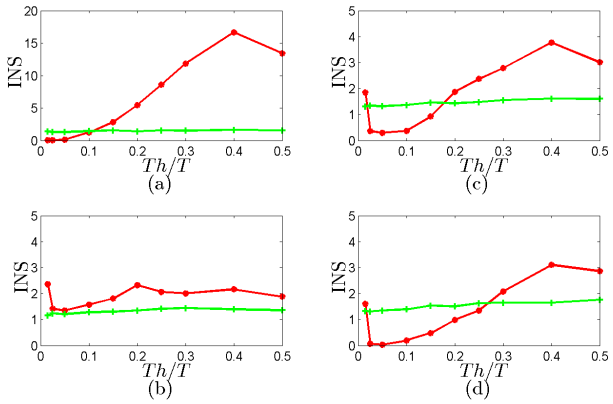

Fig. 2. INS calculado em diferentes escalas de sinais de voz: (a) saudável; (b) paralisia; (c) edema; (d) nódulo.

TABELA I
Escalas de extração do INS.

Escala	Fator de Escala	Tamanho do Segmento
1	0,015	12 ms
2	0,025	20 ms
3	0,05	40 ms
4	0,1	80 ms
5	0,15	120 ms
6	0,2	160 ms
7	0,25	200 ms
8	0,3	240 ms
9	0,4	320 ms
10	0,5	400 ms

A classificação é realizada por meio da função *classify* do Matlab® 2021 (*Mathworks*), utilizando análise discriminante linear (*Linear Discriminant Analysis* – LDA). Para dar mais confiabilidade aos resultados, foi empregado o método *k-fold* de validação cruzada, com $k = 10$, que tem sido comumente utilizado em classificação de distúrbios vocais [2], [17], [18]. Ainda, a fim de balancear a quantidade de sinais entre as classes, no processo de classificação foram selecionados aleatoriamente, entre todas as patologias, 57 sinais (metade do total), objetivando retirar qualquer influência do tamanho amostral nos resultados.

Para analisar o desempenho do classificador empregado neste trabalho, três medidas são utilizadas: acurácia, sensibilidade e especificidade. Essas medidas estão relacionadas à capacidade de um classificador em diagnosticar uma doença em um paciente doente (Verdadeiro Positivo – VP) ou saudável (Falso Positivo – FP), ou, ainda, diagnosticar um estado saudável em um paciente saudável (Verdadeiro Negativo – VN) ou doente (Falso Negativo – FN) [18].

A acurácia (Ac) representa a taxa global de acerto:

$$Ac = \frac{VP + VN}{VP + VN + FP + FN}. \quad (4)$$

A sensibilidade (Sen) é a relação entre o número de casos corretamente classificados como presença do distúrbio e a quantidade total de casos com o distúrbio:

$$Sen = \frac{VP}{VP + FN}. \quad (5)$$

A especificidade (Esp) mede a relação entre o número de casos corretamente classificados como saudáveis e a quantidade total de casos de estado saudável:

$$Esp = \frac{VN}{VN + FP}. \quad (6)$$

IV. Resultados

Na Figura 3 é apresentada a dispersão das medidas de INS em cada escala de observação, para as classes saudável (SDL) e patológico (PTL). Para as três primeiras escalas de observação, nota-se que os valores de INS para sinais de laringes patológicas são maiores e mais dispersos, com destaque para uma maior diferença observada entre as classes na escala 1. Na escala 4, a dispersão é mais equilibrada entre as classes. Da escala 5 até a escala 10 percebe-se um crescimento e um espalhamento nos valores de INS para a classe saudável. Isto indica que, à medida em que se aumenta a escala de observação (segmento do sinal), sinais de vogal sustentada oriundos de laringes saudáveis possuem um crescimento de INS. Por outro lado, os valores de INS para a classe patológica foram menores que aqueles obtidos para a classe saudável nas maiores escalas de observação. Isto pode ter sido influenciado pela presença de componentes ruidosas (por exemplo, ruído estacionário), que são intrinsecamente adicionadas no processo de produção vocal de laringes patológicas.

A. Classificação Individual das Medidas de INS

Os resultados da classificação considerando, individualmente, cada escala de observação do INS como medida acústica são apresentados na Tabela II. As medidas de INS nas escalas 1, 2 e 10 proporcionam ao classificador LDA uma acurácia média acima de 70%. Como destaque, nota-se que o maior valor de acurácia foi obtido na primeira escala (80%) com um desvio-padrão de 2,64%. Além disso, com a escala 1, o classificador atingiu máximo desempenho (100%) na identificação de patologias (sensibilidade). Por outro lado, o maior valor de especificidade foi obtido com a escala 10, com média 91% e desvio-padrão de 3,02%. Contudo, percebe-se que algumas escalas de observação (1, 2 e 3) têm valores mais elevados de sensibilidade, enquanto outras (escala 4 à 10) contribuem com um incremento nos valores de especificidade.

TABELA II
DESEMPENHO DA CLASSIFICAÇÃO INDIVIDUAL COM AS MEDIDAS DE INS DE CADA ESCALA.

Escala	Ac (%)	Sens (%)	Esp (%)
1	80,00 ± 2,64	100,00 ± 0,00	61,00 ± 5,19
2	70,91 ± 3,26	98,33 ± 1,67	45,33 ± 7,44
3	63,64 ± 3,59	94,67 ± 3,69	34,67 ± 8,92
4	59,09 ± 3,65	32,00 ± 7,03	84,33 ± 3,07
5	64,55 ± 3,44	39,67 ± 6,82	87,67 ± 3,69
6	65,45 ± 3,26	43,33 ± 7,70	85,67 ± 3,48
7	65,45 ± 3,26	43,33 ± 7,70	86,00 ± 3,44
8	69,09 ± 3,37	47,67 ± 7,60	89,33 ± 2,93
9	68,18 ± 3,39	45,00 ± 7,03	89,33 ± 2,93
10	71,82 ± 3,44	51,00 ± 7,03	91,00 ± 3,02

B. Classificação das Medidas de INS Combinadas

Na Tabela III são apresentados os resultados da classificação utilizando a combinação (Comb.) das medidas de INS de cada escala. Nota-se que, além da elevação nos valores de acurácia em relação à classificação individual, a combinação das medidas de INS proporciona um maior equilíbrio entre os valores de sensibilidade e especificidade. O maior valor de acurácia (91,82%) é obtido a partir da combinação de nove escalas, com um desvio-padrão de 2,86%. Este resultado indica que a combinação de nove escalas de INS incrementa a acurácia em aproximadamente 11 pontos percentuais (p.p.) em relação ao uso de uma única escala e, ainda proporciona um aumento de aproximadamente 9 p.p. na acurácia em relação à combinação de todas as dez escalas.

C. Discussão

A classificação de distúrbios da voz por meio de uma medida acústica que efetivamente caracterize um estado patológico ainda é um desafio. A avaliação de patologias laríngeas por meio do sinal de voz é, em geral, realizada por meio da vogal sustentada [15], [18]. Para este tipo de sinal, a detecção de variações acústicas não estacionárias pode representar padrões relacionados a disfonias. Os resultados encontrados nesta pesquisa foram obtidos considerando patologias de diferentes

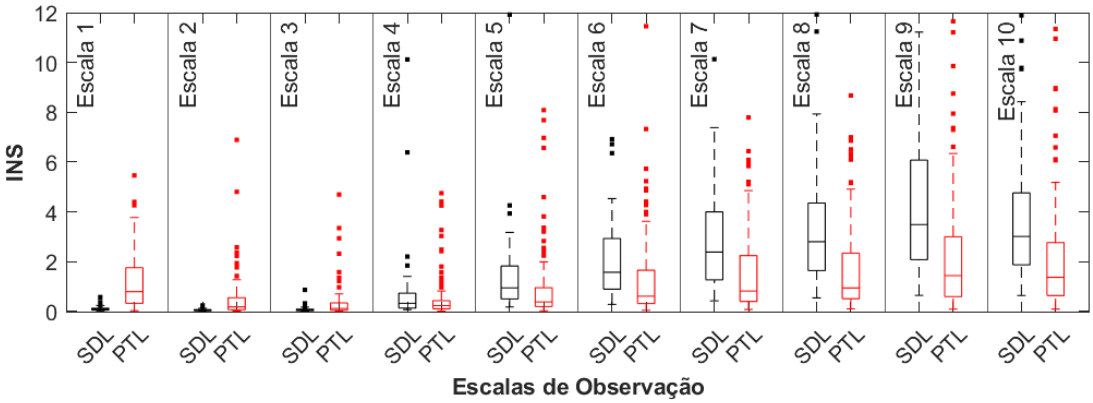

Fig. 3. Boxplots obtidos do INS para sinais saudáveis (SDL) e patológicos (PTL) considerando as dez escalas de observação.

TABELA III
DESEMPENHO DA CLASSIFICAÇÃO POR MEIO DA COMBINAÇÃO DAS MEDIDAS DE INS DE CADA ESCALA.

Comb.	Ac (%)	Sens (%)	Esp (%)	Escalas
2 a 2	88,18 ± 3,85	87,00 ± 5,17	89,67 ± 2,83	3 e 5
3 a 3	80,91 ± 4,59	96,67 ± 2,22	66,33 ± 8,26	1, 9 e 10
4 a 4	82,73 ± 4,38	98,33 ± 1,67	67,67 ± 9,34	1, 2, 7 e 10
5 a 5	85,45 ± 2,78	98,00 ± 2,00	73,33 ± 5,92	1, 3, 8, 9 e 10
6 a 6	86,36 ± 2,79	98,33 ± 1,67	75,00 ± 5,40	1, 5, 6, 8, 9 e 10
7 a 7	89,09 ± 2,64	98,00 ± 2,00	81,33 ± 5,24	1, 2, 4, 6, 7, 8 e 9
8 a 8	90,00 ± 2,12	98,00 ± 2,00	82,67 ± 3,54	1, 2, 3, 4, 5, 7, 8 e 9
9 a 9	91,82 ± 2,86	98,00 ± 2,00	85,33 ± 5,73	1, 2, 3, 5, 6, 7, 8, 9 e 10
Todas 10	82,73 ± 3,16	96,00 ± 2,67	70,00 ± 6,65	Todas

naturezas (orgânica e neurológica) e, mesmo com esta heterogeneidade, foi constatado que sinais saudáveis se comportam de maneira diferente dos sinais patológicos no que diz respeito à estacionariedade.

Individualmente, quando cada escala de observação é analisada, nota-se que o aumento do INS a partir de 80 ms nos sinais saudáveis acaba provocando uma diminuição na taxa de acerto do classificador, cuja acurácia atinge mais de 70% novamente em uma janela de observação de 400 ms. Isto é um indício de que há escalas de tempo em que são enfatizadas as variações acústicas provocadas por distúrbios da voz. Nos experimentos de combinação das medidas obtidas das escalas de INS, foi observado que há um aumento nas taxas de acerto.

Quando comparado ao estado da arte, o INS tem desempenho similar às medidas de quantificação de recorrência (análise dinâmica não linear), atingindo mais de 91% de acurácia [7]. Por outro lado, trabalhos utilizando medidas lineares (tais como frequência fundamental, *jitter* e *shimmer*) [17] e medidas não lineares baseadas em espaço de fase [25] chegaram a uma acurácia de aproximadamente 70% na discriminação entre vozes saudáveis e patológicas.

V. CONCLUSÃO

Este trabalho apresentou a análise da classificação de vozes saudáveis e vozes patológicas utilizando, como medida acústica, o índice de não estacionariedade. Para tanto, O INS foi extraído em dez escalas de observação dos sinais de vogal sustentada. Mesmo se tratando de um único tipo de emissão sonora (diferente de tarefas de fala encadeada), foram observadas diferenças de INS entre os sinais de laringes saudáveis e os sinais de laringes patológicas. A classificação, realizada com LDA, mostrou que há escalas de observação em que a medida de INS proporciona acurácia de mais de 70%. Além disso, com a combinação de medidas, foi verificado que o desempenho do classificador atinge mais de 91% de acerto. Estes resultados indicam que o INS pode ser uma efetiva medida na caracterização e classificação de patologias laríngeas por meio da voz.

REFERÊNCIAS

[1] M. Behlau, *Voz: O Livro do Especialista*. Revinter, 2001.

[2] V. J. D. Vieira, "Avaliação de distúrbios da voz por meio de análise de quantificação de recorrência," Master's thesis, Programa de Pós-Graduação em Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, 2014.

[3] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, *et al.*, "Automatic speech recognition and speech variability: A review," *Speech communication*, vol. 49, no. 10-11, pp. 763–786, 2007.

[4] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech communication*, vol. 52, no. 1, pp. 12–40, 2010.

[5] Z. Bai and X.-L. Zhang, "Speaker recognition based on deep learning: An overview," *Neural Networks*, vol. 140, pp. 65–99, 2021.

[6] R. Tavares and R. Coelho, "Speech enhancement with nonstationary acoustic noise detection in time domain," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 6–10, 2015.

[7] V. J. Vieira, S. C. Costa, S. L. Correia, L. W. Lopes, W. C. d. A. Costa, and F. M. de Assis, "Exploiting nonlinearity of the speech production system for voice disorder assessment by recurrence quantification analysis," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 8, p. 085709, 2018.

[8] A. Chern, Y.-H. Lai, Y.-P. Chang, Y. Tsao, R. Y. Chang, and H.-W. Chang, "A smartphone-based multi-functional hearing assistive system to facilitate speech recognition in the classroom," *IEEE Access*, vol. 5, pp. 10339–10351, 2017.

[9] A. Ismail, S. Abdlerazek, and I. M. El-Henawy, "Development of smart healthcare system based on speech recognition using support vector machine and dynamic time warping," *Sustainability*, vol. 12, no. 6, p. 2403, 2020.

[10] A. Venturini, L. Zao, and R. Coelho, "On speech features fusion, α-integration gaussian modeling and multi-style training for noise robust speaker classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1951–1964, 2014.

[11] K. Wang, N. An, B. N. Li, Y. Zhang, and L. Li, "Speech emotion recognition using fourier parameters," *IEEE Transactions on affective computing*, vol. 6, no. 1, pp. 69–75, 2015.

[12] V. Vieira, R. Coelho, and F. M. de Assis, "Hilbert–huang–hurst-based non-linear acoustic feature vector for emotion classification with stochastic models and learning systems," *IET Signal Processing*, vol. 14, no. 8, pp. 522–532, 2020.

[13] A. Akbari and M. K. Arjmandi, "An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features," *Biomedical Signal Processing and Control*, vol. 10, pp. 209–223, 2014.

[14] B. Rangarathnam, G. H. McCullough, H. Pickett, R. I. Zraick, O. Tulunay-Ugur, and K. C. McCullough, "Telepractice versus in-person delivery of voice therapy for primary muscle tension dysphonia," *American Journal of Speech-Language Pathology*, vol. 24, no. 3, pp. 386–399, 2015.

[15] S. L. do Nascimento Cunha Costa, *Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas*. PhD thesis, Programa de Pós-Graduação em Engenharia Elértica, Universidade Federal de Campina Grande, 2008.

[16] G. K. L. P. Queiroz, "Análise dinâmica não linear e análise de quantificação de recorrência aplicadas na classificação de desvios vocais," Master's thesis, Programa de Pós-Graduação em Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, 2018.

[17] L. W. Lopes, L. B. Simões, J. D. da Silva, D. da Silva Evangelista, A. C. d. N. e Ugulino, P. O. C. Silva, and V. J. D. Vieira, "Accuracy of acoustic analysis measurements in the evaluation of patients with different laryngeal diagnoses," *Journal of Voice*, vol. 31, no. 3, pp. 382–e15, 2017.

[18] W. C. de Almeida Costa, *Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas*. PhD thesis, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, 2012.

[19] J. J. Jiang, Y. Zhang, and C. McGilligan, "Chaos in voice, from modeling to measurement," *Journal of Voice*, vol. 20, no. 1, pp. 2–17, 2006.

[20] G. Fant, *Speech acoustics and phonetics: Selected writings*, vol. 24. Springer Science & Business Media, 2004.

[21] L. R. Rabiner and R. W. Schafer, *Introduction to digital speech processing*, vol. 1. Now Publishers Inc, 2007.

[22] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3459–3470, 2010.

[23] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal processing*, vol. 18, no. 4, pp. 349–369, 1989.

[24] K. Elemetrics, "Kay elemetrics corp. disordered voice database." Model 4337, 03 Ed., 1994.

[25] M. O. Santos, J. M. d. Assis, V. J. D. Vieira, and F. M. d. Assis, "Ksg estimation of reconstruction delay to detect vocal disorders in nonlinear dynamical analysis," *Research on Biomedical Engineering*, vol. 34, pp. 217–225, 2018.