

Avaliação de Desempenho na Classificação de Patologias Laríngeas por Análise LPC de Sinais de Voz e Redes Neurais MLP

Vinícius J. D. Vieira, Silvana C. Costa,
Washington C. de A. Costa e Suzete E. N. Correia

Unidade Acadêmica de Indústria
Instituto Federal de Educação,
Ciência e Tecnologia, João Pessoa, Paraíba
Emails: viniciusjdv@gmail.com, silvana@ifpb.edu.br,
washington@ifpb.edu.br, suzete@ifpb.edu.br

Joseana M. Fechine Régis de Araújo

Unidade Acadêmica de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, Paraíba
Email: joseana@computacao.ufcg.edu.br

Resumo— Este artigo trata da aplicação da Codificação por Predição Linear (LPC – *Linear Predictive Coding*) e das redes neurais *Multilayer Perceptron* (MLP) na classificação entre sinais de vozes saudáveis e vozes afetadas por patologias laríngeas (edema, nódulos e paralisia nas pregas vocais). O desempenho do sistema de classificação entre vozes saudáveis e patológicas é avaliado variando-se a ordem dos coeficientes LPC de 8 a 32. Foram obtidas taxas de classificação superiores a 98%, na discriminação entre vozes saudáveis e patológicas. Entre patologias, os valores de acurácia obtidos estão acima de 90%.

Palavras-Chave— Processamento de Sinais de Voz, Codificação por Predição Linear, Redes Neurais MLP, Patologias Laríngeas.

Abstract— This article deals with the application of Linear Predictive Coding (LPC) and MLP neural networks in classifying healthy and voices affected by laryngeal pathologies (vocal fold edema, nodules and paralysis). The system performance is evaluated by varying the order of the LPC coefficients from 8 to 32. Classification rates above 98% were obtained in the healthy and pathological voice discrimination. Among pathologies, the accuracy values were higher than 90%.

Keywords— Processing of Speech Signals, Linear Predictive Coding, Neural Network MLP, Laryngeal Pathologies.

I. INTRODUÇÃO

Sistemas de detecção automática de patologias na laringe, por meio de análise acústica da voz, podem ser ferramentas úteis em tratamentos preventivos, bem como no auxílio a tratamentos pré e pós cirúrgicos [1]. O acesso a uma ferramenta desse tipo traz, entre outras facilidades, a possibilidade de acompanhamento da qualidade vocal de profissionais como professores, músicos, locutores, entre outros. Além disso, diagnósticos podem ser compartilhados por meio de redes de Telemedicina, a fim de reduzir custos de traslado de pacientes e de exames desconfortáveis, como a videolaringoscopia.

Técnicas de processamento digital de sinais de voz têm sido empregadas com o objetivo de investigar parâmetros quantitativos que possam diferenciar uma voz alterada de uma voz normal [1-10]. Patologias na laringe alteram o padrão de vibração das pregas vocais, fazendo com que o sinal resultante da produção vocal possua características diferentes

de um sinal de voz saudável. Quanto mais sensíveis forem as características extraídas do sinal para representar a desordem vocal provocada pela patologia, melhores serão os resultados da classificação.

Algumas medidas tradicionais de análise acústica, baseadas na frequência fundamental, tais como *jitter* e *shimmer*, apresentam certas limitações quando o sinal de voz em análise é afetado por uma patologia considerada severa. Uma vogal sustentada pode perder seu aspecto quase periódico, apresentando características de ruído aleatório (Figura 1), prejudicando a detecção correta da frequência fundamental e das medidas derivadas da mesma [1], [3], [7], [8].

Dessa forma, outras medidas que independem da frequência fundamental têm sido investigadas em diversos trabalhos. Algumas pesquisas anteriores se baseiam no modelo linear de produção da voz, empregando análise LPC (*Linear Predictive Coding*) e análise cepstral [4-6], além de combinações destas medidas, empregando outros tipos de classificadores, apresentando resultados relevantes [2], [11], [12].

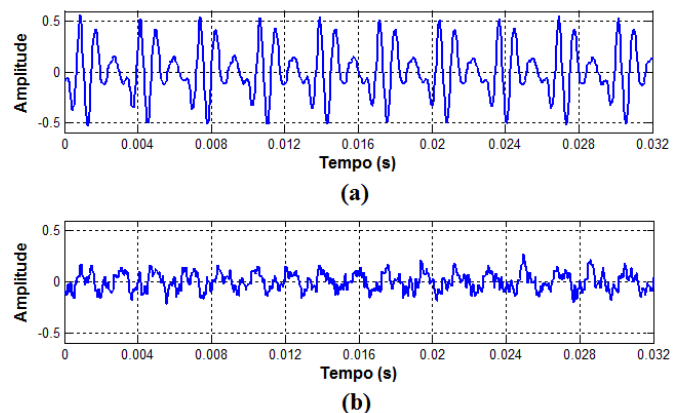


Fig. 1. Forma de onda de 32 ms de uma vogal sustentada /ah/ para: (a) sinal de voz saudável; (b) sinal de voz afetada por edema de Reinke.

Godino *et al.* [1] sugerem que a detecção de desordens vocais pode ser realizada usando vetores formados pela

combinação de coeficientes cepstrais na escala *mel* (*MFCC – Mel Frequency Cepstral Coefficients*), mais a sua primeira derivada (delta mel-cepstrais) e a energia de cada segmento do sinal analisado. Nos experimentos são utilizados 53 sinais de vozes saudáveis e 173 sinais de vozes afetadas por diferentes patologias [13]. A classificação é feita com o uso de modelos de misturas gaussianas (*GMM – Gaussian Mixture Models*) que fornecem uma decisão sobre a presença ou ausência da patologia para cada segmento do sinal. O melhor resultado de classificação foi obtido com o uso de 24 MFCC e um modelo de 6 misturas gaussianas, com uma taxa máxima de acurácia de $94,07\% \pm 3,28\%$.

Em seu trabalho de Tese, Costa [3] apresenta um estudo sobre a discriminação da patologia edema na pregas vocais, empregando-se análise por predição linear (LPC), análise cepstral e mel-cepstral, com 12 coeficientes, para cada segmento de 20 ms dos sinais analisados. A classificação é realizada em duas etapas: uma pré-classificação, em que é utilizada a medida de distância do mínimo erro médio quadrático e, na fase final, utiliza-se uma modelagem paramétrica por meio de Modelos de Markov Escondidos (*HMM – Hidden Markov Models*). Os resultados indicam os coeficientes provenientes da análise LPC como sendo os que obtiveram melhores valores de acurácia no processo de classificação, tendo sido obtido, na fase final de classificação, a taxa máxima de 100% na discriminação entre sinais de vozes saudáveis e vozes afetadas por edema e de 96% na discriminação entre sinais de vozes com edema e vozes com outras patologias numa mesma classe (nódulos, cistos e paralisia). No entanto, no referido trabalho, não foram realizadas: validação cruzada, variação da ordem dos coeficientes e discriminação entre patologias.

Fonseca e Pereira [9] utilizaram uma abordagem baseada na análise *wavelet* para avaliar diferenças no *jitter* de 71 sinais de voz, sendo 30 saudáveis, 25 de pacientes com nódulos e 16 de pacientes com edema de Reinke. Os resultados obtidos a partir do classificador SVM (*Support Vector Machine*) para a discriminação entre os diferentes grupos de sinais apresentaram as seguintes taxas máximas de acurácia: saudável *versus* patológica, 88,20%; saudável *versus* edema, 85,30%; saudável *versus* nódulos, 90,10% e nódulos *versus* edema 82,40%.

Wang *et al.* [10] utilizaram um subconjunto da base de dados da Kay Elemetrics [13] composto por 53 sinais de vozes saudáveis e 173 sinais de vozes afetadas por diversas patologias nas pregas vocais. Nesse trabalho é utilizada análise mel-cepstral para extração de vetores de dimensão 16 como entrada para um classificador baseado na combinação de um modelo com 16 misturas Gaussianas e SVM (GMM-SVM). Os resultados indicam que um sinal da vogal sustentada /ah/ pode ser classificado como sendo de uma voz saudável ou patológica com uma acurácia de $96,1\% \pm 2,51\%$. Também não houve discriminação entre patologias.

A busca por características que retratem bem as mudanças na voz que a patologia provoca é objeto de pesquisa constante, bem como o tipo de classificador que consiga separar de forma eficiente, os dados dos sinais analisados e discriminá-los. Algumas características conseguem modelar bem uma patologia, mas não outra, dado que os aspectos fisiológicos e articulatórios variam na produção dos sons, com a presença

da patologia.

Neste trabalho, os coeficientes LPC são usados como características representantes dos sinais de voz analisados. São investigadas diferentes quantidades de coeficientes do filtro de predição linear (8 a 32), com o intuito de determinar a ordem ideal do preditor, que aumente a eficácia do classificador. A avaliação da ordem dos coeficientes LPC é relevante em aplicações em que se deseja utilizar o menor número possível de coeficientes, como por exemplo, em sistemas embarcados para situações em tempo real, que necessitem de baixo custo computacional [14]. As redes neurais *Multilayer Perceptron* são usadas para a classificação dos sinais de voz.

A maioria dos trabalhos existentes na literatura proporciona a discriminação apenas entre vozes saudáveis e patológicas em geral. Por outro lado, a discriminação entre patologias é um fator importante na análise clínica. Portanto, é investigada também, neste trabalho, a classificação dos sinais de voz entre diferentes patologias. As patologias laringeas consideradas são paralisia nas pregas vocais, edemas de Reinke e nódulos vocais.

O trabalho está organizado como segue. Na Seção II é apresentada a Análise LPC. Os materiais e métodos empregados no trabalho são apresentados na Seção III. Na Seção IV são apresentados os resultados obtidos e, na Seção V, as conclusões.

II. ANÁLISE LPC

A análise de sinais de voz pelo método da predição linear baseia-se na Teoria do sistema fonte-filtro (Figura 2), em que a fonte é o resultado acústico da vibração das pregas vocais, enquanto a função de transferência do trato vocal (filtro) é o resultado acústico de uma determinada forma e comprimento do trato vocal [15].

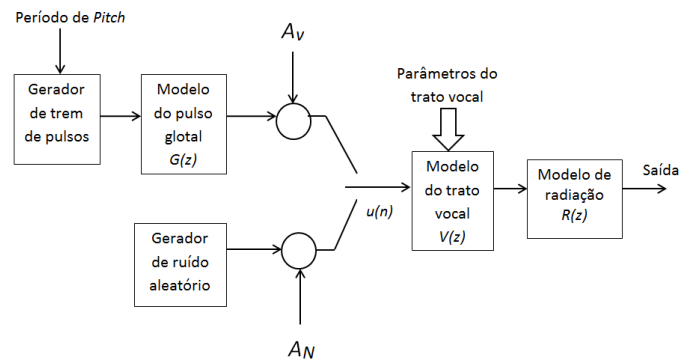


Fig. 2. Modelo geral discreto no tempo para produção da fala [16].

Na Figura 2, $u(n)$ é o sinal de excitação que chega ao trato vocal, A_V e A_N agem controlando a intensidade da excitação da voz e do ruído, respectivamente. As funções do pulso glotal ($G(z)$), da radiação do som nos lábios ($V(z)$) e das componentes do trato vocal ($R(z)$), combinadas, constituem uma única função de transferência ($H(z)$) que representa o processo de produção da fala, conforme (1) [16].

$$H(z) = G(z) \cdot V(z) \cdot R(z). \quad (1)$$

As características acústicas das vogais são geradas pela ação das propriedades de ressonância do trato vocal, que interagem modificando o conteúdo acústico do som quase periódico produzido pela laringe, o sinal da fonte. As frequências naturais de ressonância do trato vocal (formantes) dependem da forma, do tamanho e do material de que é constituído. Em vozes irregulares, com características ruidosas, a determinação das frequências de formantes do som se torna mais complexa e menos confiável. Neste caso, é recomendada a análise por predição linear [15].

O método LPC estima cada amostra de voz baseado numa combinação linear de p amostras anteriores. Um valor de p maior representa um modelo mais preciso. O princípio básico da análise LPC é determinar um conjunto de coeficientes do preditor, $\alpha(k)$, diretamente do sinal de voz, de forma a obter uma boa estimativa das propriedades espectrais do sinal [16]. Desta forma, (1) pode ser escrita da forma:

$$H(z) = \frac{G}{1 - \sum_{k=1}^p \alpha(k)z^{-k}}, \quad (2)$$

em que G é um fator de ganho, que é ajustado para controlar a intensidade da excitação e p a ordem do preditor.

Considerando que o trato vocal seja saudável, estima-se que as desordens no sinal de voz sejam causadas por alterações na fonte (laringe). Desta forma, os coeficientes LPC podem ser empregados para modelar as variações do sinal provenientes de problemas associados à fonte.

III. MATERIAIS E MÉTODOS

Nesta seção é apresentada a base de dados utilizada na pesquisa, bem como a metodologia empregada nas etapas de extração de parâmetros e classificação dos sinais de voz analisados.

A. Base de Dados

Neste trabalho, foram analisados 167 sinais de voz, oriundos da *Disordered Voice Database, Model 4337*, desenvolvida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* e comercializada pela kay Pentax [13]. São utilizados 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 nódulo vocal). Os sinais foram amostrados com uma taxa de amostragem de 25 ou 50 kHz e 16 bits/amostra.

B. Metodologia

Para análise, cada sinal de voz amostrado em 50 kHz foi reamostrado com uma taxa de 25 kHz e particionado em 30 segmentos consecutivos de 32 ms de duração (quadros de 800 amostras). Com a segmentação, obteve-se 1.590 segmentos para os sinais de vozes saudáveis, 1.560 para sinais afetados por paralisia nas pregas vocais, 1.320 para sinais com edema e 540 para os sinais de vozes afetadas por nódulos. De cada segmento, foram extraídos os coeficientes LPC com as seguintes ordens: 8, 12, 16, 20, 24, 28 e 32.

No modelo LPC, a ordem do filtro de predição é um compromisso entre acurácia espectral, custo computacional e largura de faixa de transmissão. Recomenda-se, em geral, que o número de coeficientes seja o valor da frequência de amostragem em kHz mais dois a quatro pólos. Neste caso, o valor mínimo ideal recomendado seria de 27 coeficientes [14].

Para a classificação dos sinais, foram utilizadas as redes neurais MLP [17], implementadas a partir do *software WEKA* [18]. As classes observadas foram: sinais de vozes saudáveis (SDL), sinais de vozes patológicas (PTL) envolvendo todas as patologias numa mesma classe, sinais de paralisia (PRL), sinais de edema (EDM) e sinais de nódulo (NDL). A classificação é realizada caso a caso: SDLxPTL, SDLxPRL, SDLxEDM, SDLxNDL, PRLxEDM, PRLxNDL e EDMxNDL.

Foi empregado, durante a fase de treinamento das redes neurais, o algoritmo *Backpropagation* [19] com uma camada oculta. Alguns parâmetros da rede MLP são os mesmos para todos os casos de classificação. Foi utilizada uma taxa de aprendizagem igual a 0,3 e uma taxa de momento igual a 0,2. A fase de teste do classificador é realizada com validação cruzada, na qual o conjunto de dados é dividido em n subconjuntos. Desses, $n - 1$ são utilizados na fase de treino do classificador, e o subconjunto restante é utilizado na fase de teste. O procedimento se repete n vezes, de maneira que todos os subconjuntos (*folds*) sejam empregados na fase de teste. Neste trabalho, utiliza-se n igual a 10.

A topologia de cada rede MLP implementada neste trabalho dependeu da quantidade de coeficientes LPC na camada de entrada. Para definição da quantidade de neurônios na camada oculta foi utilizada a regra descrita em (3).

$$N_h = \frac{N_f + N_c}{2}, \quad (3)$$

em que N_h é a quantidade de neurônios na camada oculta, N_f é a quantidade de características na entrada da rede MLP, e N_c é a quantidade de classes envolvidas no reconhecimento.

IV. RESULTADOS

Inicialmente, no processo de classificação, os sinais foram separados em duas classes distintas: saudáveis (SDL) ou patológicos (PTL). A primeira classificação foi realizada entre estas duas classes, em que a classe patológica se refere à junção de todos os sinais com diferentes patologias laríngeas numa mesma classe. Neste trabalho, as patologias analisadas são: paralisia (PRL), edema (EDM) e nódulos (NDL) vocais. Numa segunda etapa de classificação, foi realizada a discriminação entre os sinais saudáveis e cada uma das patologias separadamente (SDLxEDM, SDLxNDL e SDLxPRL). Além disso, foi realizada a discriminação entre as patologias (EDMxNDL, EDMxPRL e PRLxNDL). Os resultados obtidos na classificação são apresentados a seguir.

A. Classificação entre vozes saudáveis e patológicas

Na Figura 3 estão apresentados os valores de acurácia obtidos da classificação dos sinais de vozes saudáveis e dos sinais de vozes patológicas (SDLxPTL), para todas as patologias avaliadas agrupadas em uma única classe.

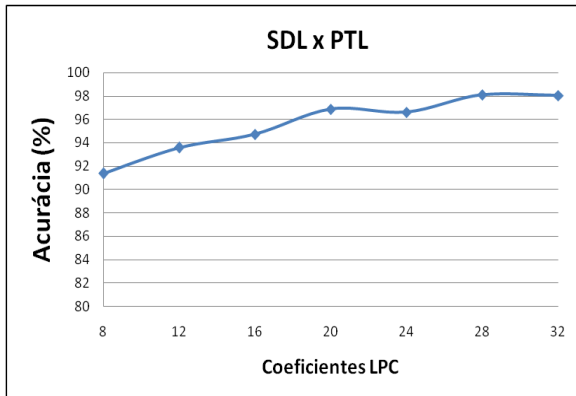


Fig. 3. Resultado da classificação na rede MLP das classes Saudável (SDL) e Patológica (PTL).

O melhor resultado foi obtido com 28 coeficientes LPC, com uma acurácia média de $98,19\% \pm 0,70\%$ (sendo $97,72\% \pm 1,61\%$ de correta classificação dos sinais saudáveis e $98,35\% \pm 0,55\%$ de correta classificação dos sinais patológicos). Portanto, o método baseado na codificação preditiva linear consegue perceber bem a desordem provocada nos sinais de voz pela presença da patologia, comparativamente aos sinais considerados saudáveis.

O grupo de sinais de vozes saudáveis foi comparado com os sinais agrupados em patologias distintas (paralisia, edema ou nódulos). Na Figura 4 estão apresentados os valores de acurácia obtidos da classificação com a rede MLP para os três casos investigados neste trabalho: sinais de vozes saudáveis *versus* sinais de vozes afetadas por paralisia nas pregas vocais (SDLxPRL), sinais de vozes saudáveis *versus* sinais de vozes afetadas por edema de Reinke (SDLxEDM) e sinais de vozes saudáveis *versus* sinais de vozes afetadas por nódulo vocal (SDLxNDL).

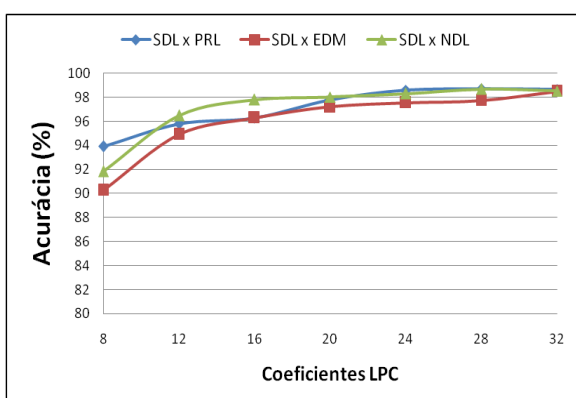


Fig. 4. Resultado da classificação na rede MLP entre as classes Saudável e Paralisia (SDLxPRL), Saudável e Edema (SDLxEDM), e Saudável e Nódulo (SDLxNDL).

No caso da classificação entre sinais saudáveis e sinais de paralisia, a melhor taxa de correta classificação foi obtida com 28 coeficientes LPC, com uma acurácia média de $98,73\% \pm 0,64\%$ (sendo $98,55\% \pm 0,89\%$ de correta classificação dos sinais saudáveis e $98,92\% \pm 0,76\%$ de cor-

reta classificação dos sinais de paralisia nas pregas vocais). No entanto, com 24 coeficientes, já se consegue resultado próximo ao que foi alcançado com 28 coeficientes ($98,54\% \pm 0,64\%$ de acurácia média).

Considerando a classificação entre sinais saudáveis e sinais de edema, o maior valor de acurácia média é $98,25\% \pm 0,94\%$ com 32 coeficientes LPC (sendo $98,58\% \pm 0,92\%$ de correta classificação dos sinais saudáveis e $97,81\% \pm 1,39\%$ de correta classificação dos sinais de edema). Um número mais alto de coeficientes para obter-se uma acurácia de mais de 98%, neste caso, pode ser atribuído ao fato dos sinais afetados por edema, contidos na base de dados empregada, não serem separados por grau de severidade da patologia. Dessa forma, a presença da patologia menos grave pode caracterizar, dependendo dos parâmetros da voz analisada, um sinal saudável. Uma maior precisão poderia ser obtida, caso os sinais fossem separados por grau de severidade ou tipo de patologia, como por exemplo edema unilateral direito, edema unilateral esquerdo ou edema bilateral.

No caso dos sinais, cujos pacientes apresentam paralisia nas pregas vocais, também não houve uma divisão na base de dados entre os sinais de vozes afetados por paralisia discriminando a prega vocal paralisada (esquerda, direita, ou ambas). Tal separação não foi feita, neste trabalho, para não causar uma diminuição significativa no número de sinais disponíveis para análise, em cada caso particular.

A maior taxa de correta classificação entre sinais saudáveis e sinais de nódulos foi obtida com 28 coeficientes LPC, com $98,69\% \pm 0,85\%$ de acurácia média (sendo $99,31\% \pm 0,64\%$ de correta classificação dos sinais saudáveis e $96,84\% \pm 2,62\%$ de correta classificação dos sinais de nódulos).

A partir dos resultados obtidos, acurácia acima de 98%, constata-se a eficácia da análise paramétrica com os coeficientes LPC na detecção das patologias edema, nódulos e paralisia quando comparados aos sinais de vozes saudáveis.

B. Classificação entre as patologias

Os sinais de vozes patológicos foram comparados entre si, empregando a mesma variação nos coeficientes LPC, descrita anteriormente, com o intuito de discriminar entre as patologias edema, nódulos e paralisia nas pregas vocais.

Na Figura 5 estão apresentados os valores de acurácia obtidos da classificação com a rede MLP entre patologias: paralisia *versus* edema (PRLxEDM), paralisia *versus* nódulos (PRLxNDL), edema *versus* nódulos (EDMxNDL).

A classificação de sinais de vozes afetadas por paralisia e sinais de vozes afetadas por edema obtém uma acurácia média máxima de $96,56\% \pm 0,92\%$ (sendo $96,85\% \pm 1,68\%$ de correta classificação dos sinais de paralisia e $96,20\% \pm 1,48\%$ de correta classificação dos sinais de edema), com 32 coeficientes LPC.

Em se tratando da classificação entre sinais de vozes afetadas por paralisia e sinais de vozes afetadas por nódulos, a rede apresentou acurácia média máxima, com 28 coeficientes LPC, de $97,88\% \pm 0,55\%$ (sendo $98,56\% \pm 0,76\%$ para correta classificação de paralisia e $95,16\% \pm 2,79\%$ para correta classificação de nódulos).

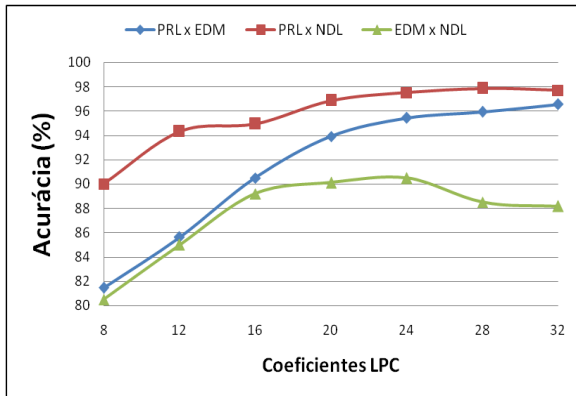


Fig. 5. Resultado da classificação com a rede MLP entre as classes de sinais patológicos: Paralisia e Edema (PRLxEDM), Paralisia e Nódulo (PRLxNDL), e Edema e Nódulo (EDMxNDL).

A maior taxa de correta classificação na rede MLP, entre edema e nódulos, é obtida com 24 coeficientes. A acurácia média máxima foi de $90,48\% \pm 2,36\%$ (sendo $94,22\% \pm 3,76\%$ para correta classificação de edema e $81,27\% \pm 7,38\%$ para correta classificação de nódulos vocais).

Na tarefa de discriminar entre as patologias paralisia e edema e entre paralisia e nódulos, o classificador MLP apresentou uma acurácia média acima de 95%. Este resultado expressa a eficiência do método empregado em capturar as diferenças significativas na desordem vocal provocada nos sinais analisados, mediante as mudanças do padrão vibratório das pregas vocais, específicas para cada patologia em estudo (edema e nódulos como lesões de massa e paralisia como lesão neuromuscular [15]).

C. Resumo da classificação

Ao se observar os gráficos apresentados nas seções precedentes, a diferença percentual na acurácia para as classificações entre SDLxNDL, SDLxPRL e PRLxNDL, quando se varia o número de coeficientes de 20 para 28, não ultrapassa 1%. Nos casos de discriminação entre PRLxEDM, o ganho obtido é de cerca de 2% e, entre SDLxPTL e SDLxEDM, o aumento na taxa de classificação é cerca de 1,2%. Em se tratando de EDMxNDL, de 20 para 24 coeficientes, o aumento na acurácia é de 0,4%.

O número de coeficientes necessários para alcançar a acurácia máxima em cada caso de classificação, seja entre vozes saudáveis e vozes patológicas, foi de, no mínimo, 24 coeficientes. Dado que o aumento da acurácia obtida na variação de 20 a 32 coeficientes não ultrapassa 2%, recomenda-se utilizar um número mínimo de 20 coeficientes, caso seja necessário diminuir o custo computacional.

Na Tabela I estão apresentados os melhores resultados obtidos para os casos de classificação investigados neste trabalho. O sistema de classificação conseguiu discriminar entre as vozes saudáveis e patológicas com acurácia acima de 98% para 28 coeficientes por segmento.

O menor valor de acurácia ($90,48\% \pm 2,36\%$) foi obtido na discriminação entre edema e nódulos, chegando a taxas superiores a 96% na maioria dos casos, mesmo entre patologias.

TABELA I
VALORES MÁXIMOS DE ACURÁCIA (%) PARA CADA CASO DE CLASSIFICAÇÃO.

Casos de Classificação	Acurácia Máxima (%)	Quantidade de coeficientes LPC
SDL x PTL	$98,19 \pm 0,70\%$	28
SDL x PRL	$98,73 \pm 0,64\%$	28
SDL x EDM	$98,25 \pm 0,94\%$	32
SDL x NDL	$98,69 \pm 0,85\%$	28
PRL x EDM	$96,56 \pm 0,92\%$	32
PRL x NDL	$97,88 \pm 0,55\%$	28
EDM x NDL	$90,48 \pm 2,36\%$	24

Com um desvio padrão de 2,36%, na classificação EDMxNDL, o número de coeficientes pode ser reduzido até 16. Dado que as duas patologias são lesões na porção membranosa das pregas vocais, apresentando aspectos similares, não só na fisiologia, como também nos aspectos perceptuais, o resultado obtido é considerado satisfatório.

V. CONCLUSÕES

Neste trabalho foi avaliado o uso dos coeficientes LPC como parâmetros representativos dos sinais de voz. Para classificação, foram utilizadas redes neurais MLP. Os resultados obtidos fornecem indicativos para o emprego do método em ferramentas de apoio ao diagnóstico de patologias laringeas. O método empregado detecta a presença da patologia com precisão acima de 98%. Comparativamente aos trabalhos referenciados [1], [2], [10], [20], os resultados obtidos neste trabalho, tanto na discriminação entre vozes saudáveis e patológicas, quanto entre patologias, foram superiores, embora sejam empregadas outras características e outros tipos de classificadores. Os valores de acurácia, superiores a 90%, demonstram a aplicabilidade do método não só na discriminação entre vozes saudáveis e patológicas, como também entre as patologias edema, nódulos e paralisia nas pregas vocais.

AGRADECIMENTOS

À CAPES, pelo financiamento parcial da pesquisa, e à Universidade Federal de Campina Grande pela base de dados.

REFERÊNCIAS

- [1] J. I. Godino-Llorente, P. Gomes-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *IEEE Transactions on Biomedical Engineering*, vol. 53(10), pp. 1943–1953, 2006.
- [2] W. C. de A Costa, F. M. Assis, B. G. A. Neto, S. C. Costa, and V. J. D. Vieira, "Pathological voice assessment by recurrence quantification analysis," in *Biosignals and Birobotics Conference (BRC), 2012 ISSNIP*. IEEE, 2012, pp. 1–6.
- [3] S. L. d. N. C. Costa, "Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas," *Universidade Federal de Campina Grande. Tese de Doutorado*, 161 p., 2008.
- [4] 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, p. 327, 2001.
- [5] M. Marinaki, C. Kotropoulos, I. Pitas, and N. Maglaveras, "Automatic detection of vocal fold paralysis and edema," in *Proc. 2004 Int. Conf. Spoken Language Processing*, 2004.

- [6] J. V. M. H. M. G. J. M. F. S. C. Costa, "Detecção automática de patologias da laringe usando codificação por predição linear e redes neurais mlp," in *Anais do IX Congresso Brasileiro de Redes Neurais, Ouro Preto-MG*, 2009.
- [7] B. Boyanov, T. Ivanov, S. Hadjitodorov, and G. Chollet, "Robust hybrid pitch detector," *Electronics letters*, vol. 29, no. 22, pp. 1924–1926, 1993.
- [8] C. Manfredi, L. Pierazzi, and P. Brusaglioni, "Pitch estimation for noise retrieval in time and frequency domain," *Med. Biol. Eng. Comput.*, vol. 37, no. 2, pp. 532–533, 1999.
- [9] E. S. Fonseca and J. C. Pereira, "Normal versus pathological voice signals," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 28, no. 5, pp. 44–48, 2009.
- [10] X. Wang, J. Zhang, and Y. Yan, "Discrimination between pathological and normal voices using GMM-SVM approach," *Journal of Voice*, vol. 25, no. 1, pp. 38–43, 2011.
- [11] R. Tavares, N. Brunet, S. Correia, S. C. Costa, B. G. A. Neto, and J. Macêdo, "Pathological voice detection by cepstral analysis using multiple classifiers," *analysis*, vol. 8, p. 14.
- [12] S. C. Costa, S. Correia, H. Falcão, N. Almeida, B. G. A. Neto, and J. M. Fechine, "Pathological voice discrimination based on entropy measurements," in *Proceedings of the 2008 ACM symposium on Applied computing*. ACM, 2008, pp. 1424–1425.
- [13] K. Elemetrics, "Kay elemetrics corp. disordered voice database," Model 4337, 03 Ed., 1994.
- [14] D. O'Shaughnessy, "Speech communications: Human and machine, the institute of electrical and electronics engineers-ieee," *Inc., New York*, 2000.
- [15] M. Behlau, *A Voz do Especialista*. Rio de Janeiro: Revinter, 2001, vol. 1.
- [16] L. R. Rabiner and R. W. Schafer, *Digital Signal Processing of Speech Signals*. Prentice Hall, Inc., 1978.
- [17] S. Russell and P. Norvig, "Artificial intelligence: A modern approach. 2^o ed." *Prentice Hall, Inc.*, 2003.
- [18] M. H. E. F. G. H. B. P. P. Reutemann and I. H. Witten. (2009) The weka data mining software: An update; sigkdd explorations, volume 11, issue 1. Machine Learning Group at the University of Waikato. Disponível em <http://www.cs.waikato.ac.nz/ml/weka/index.html>. Acesso em março de 2013.
- [19] S. S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall, Inc., 1999.
- [20] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.