

Classificação de Patologias da Fala a partir do PPM

Hildegard Paulino Barbosa, Joseana Macêdo Fechine, José Eustáquio Rangel
Centro de Engenharia Elétrica e Computação
Universidade Federal de Campina Grande (UFCG)
Campina Grande, Brasil
hildegardpaulino@gmail.com, joseana@dsc.ufcg.edu.br, rangeldequeiroz@gmail.com

Abstract—Speech organs are very susceptible to several types of pathologies, which may harm voice production. Several techniques have been traditionally used to detect these pathologies. However, they present drawbacks concerning the accuracy and the comfort of patients during application. Moreover, results obtained by computing techniques have not yet matured to a reliable tool for application in clinics. In this research, a classification approach based on a method not previously employed in classification of vocal tract diseases is proposed. It is based on Prediction by Partial Matching (PPM), which uses acoustical and temporal features to feed models. It were obtained very promising results in the presence or absence of pathologies (at least 92%). With regard to pathology discrimination, preliminary results confirmed that PPM is a high potential technique for voice pathology classification, although its clinical application for the diagnosis of voice pathologies still needs deeper investigation.

Keywords—speech pathologies; prediction by partial matching; acoustical and temporal features;

I. INTRODUÇÃO

A voz é o meio de comunicação mais importante e mais natural do ser humano, a partir da qual são expressas vontades, pensamentos, ordens e informações. Entretanto, para que a comunicação seja efetiva, é necessário o entendimento correto da voz enunciada por parte do interlocutor do processo. Se isto não ocorrer, haverá maior propensão a equívocos, o que desestimulará a comunicação causando, até mesmo, o constrangimento do locutor. Tal problema, denominado *disfonia*, é causado muitas vezes por patologias da fala, às quais o sistema fonador humano é muito suscetível. Estima-se que entre 3 e 10% da população geral tenha o sistema fonador comprometido por alguma patologia [1], além do que é comum que o mesmo indivíduo possa ser acometido por até 8 patologias [2], as quais podem ser causadas por alterações psicoemocionais, doenças neurodegenerativas, mau uso da voz ou hábitos sociais não saudáveis, tais como o tabagismo e a ingestão de álcool [1] [3] [4]. Algumas destas razões explicam a ocorrência mais freqüente de patologias da fala em fumantes e em categorias de profissionais que utilizam a voz como seu principal instrumento de trabalho, e.g., professores, cantores, radialistas, jornalistas [4]. Atualmente, há conhecimento de mais de 120 patologias [5], mas as mais conhecidas são *Nódulo*, *Edema*, *Paralisia* e *Pólipo*.

Na detecção de patologias da fala, são usados, tradicionalmente, dois tipos de mecanismos. O primeiro, consiste da escuta da elocução vocal do paciente por um

profissional (normalmente, um fonoaudiólogo ou um otorrinolaringologista), visando a decidir sobre a presença ou ausência de uma patologia. Até poucos anos, este era o método mais empregado [6]. Contudo, não é difícil perceber seu caráter altamente subjetivo e propenso à indução de erros, principalmente nos casos em que a patologia se encontra em estágios iniciais, devido à alta dependência da experiência, da acurácia, do nível de fadiga e da sensibilidade do sistema auditivo do profissional^a [7] [8].

O segundo mecanismo consiste de procedimentos clínicos nos quais a voz do paciente é avaliada por meio de recursos visuais. Dentre os exames mais comuns desta natureza estão a *videolaringoscopia* e a *videostroboscopia* [3]. Embora precisos, estes exames são bastante invasivos e desconfortáveis para o paciente, causando, em alguns casos, a ação de reflexo durante a aplicação, em função de sua sensibilidade laríngea, o que pode acarretar falsos diagnósticos [4]. Além disto, comprometem financeiramente ambas as partes, já que os equipamentos requeridos para executá-los são caros e sofisticados, obrigando o repasse dos custos ao paciente e restringindo seu acesso a grande parte da população.

Inúmeras pesquisas sobre a detecção de patologias da fala por computador têm sido desenvolvidas, com o intuito de reduzir significativamente a necessidade e a frequência de exames visuais. Dentre as técnicas utilizadas, incluem-se Modelos de Markov Escondidos (*Hidden Markov Models - HMM*) [3], Quantização Vetorial, Redes Neurais [4], Gráficos de Recorrência [1] [9], Modelos de Misturas Gaussianas (*Gaussian Mixture Models - GMM*) [4] [10], fase TEO [11], Máquinas de Suporte Vetorial [12], Dinâmica Não Linear com medidas de complexidade [13], Expoente de Hurst [14] dentre outras. Porém, um tipo de abordagem sobre a qual não foi encontrado registro na revisão de literatura foi o uso de métodos de compressão de dados. Embora eles tenham sido projetados inicialmente para comprimir dados, percebeu-se que o rico modelo estatístico gerado por alguns destes algoritmos pode ser empregado também em atividades de classificação. Por esta razão, se afigura importante o estudo da eficácia deste tipo de método na discriminação de patologias.

Um dos métodos mais eficazes de compressão de dados é a Predição por Casamento Parcial (*Prediction by Partial Matching - PPM*) [18]. Seu princípio de funcionamento será descrito na Seção II.B. Contudo, deve-se considerar que bons

^a Diferentes diagnósticos podem ser dados por diferentes profissionais ou, até mesmo, pelo mesmo profissional, em ocasiões diferentes.

resultados têm sido obtidos a partir do seu uso em atividades de compressão e classificação de arquivos binários, textos, sinais de eletrocardiograma e imagens, dentre outros tipos de sinais [15][16][17][18].

II. FUNDAMENTAÇÃO TEÓRICA

Nesta seção, serão inicialmente apresentadas medidas extraídas dos sinais de voz, as quais serão utilizadas como entradas para os modelos utilizados na etapa de classificação. Em seguida, será apresentado brevemente o funcionamento do método proposto.

A. Parâmetros Utilizados

É importante mencionar de antemão que os sinais de voz apresentam invariabilidade estatística em janelas de até 32 ms porque, mesmo sendo a voz invariante no tempo, o trato vocal apresenta natureza dinâmica^b, o que afeta os parâmetros que representam a voz e, conseqüentemente, sua produção. Sendo assim, para que os valores extraídos representem a realidade acerca do sinal de voz manipulado, todas as medidas consideradas foram extraídas em janelas dentro daquele intervalo, de modo que um arquivo era representado por um conjunto de valores de determinada medida.

Energia - a soma dos quadrados dos valores de amplitude. Para obter seu valor em decibéis, basta multiplicar o logaritmo do valor encontrado por 10.

$$E_{seg} = N_A \cdot E\{[s(n)]^2\} = \sum_{n=0}^{N_A-1} [s(n)]^2 \quad (1)$$

$$E_{seg}(dB) = 10 \cdot \log[E_{seg}]$$

Em (1), N_A é o tamanho da janela e $s(n)$ é a n -ésima amostra do sinal de voz (amplitude).

Taxa de Cruzamento por Zero - número de vezes em que a forma de onda do sinal cruza o eixo das abscissas (tempo), i.e, quantas vezes um valor positivo de amplitude é sucedido por um negativo e vice-versa.

$$\begin{aligned} TCZ &= N_A \cdot E\{\text{sgn}[s(n)] - \text{sgn}[s(n-1)]\} \\ &= \sum_{n=1}^{N_A-1} |\text{sgn}[s(n)] - \text{sgn}[s(n-1)]| \end{aligned} \quad (2)$$

em que

$$\text{sgn}[s(n)] = \begin{cases} 1, & \text{se } s(n) \geq 0 \\ -1, & \text{se } s(n) < 0 \end{cases} \quad (3)$$

Os elementos dessa equação apresentam o mesmo significado dos elementos de (1).

Número Total de Picos (NTP) - quantidade de picos, positivos e negativos, existente na forma de onda do sinal.

Diferença no Número de Picos (DNP) - difere da anterior no sinal usado na operação: enquanto no NTP somam-se as quantidades de picos, na DNP subtraem-se dos picos positivos os picos negativos

Frequência Fundamental - O som é decorrente de vibrações mecânicas que se propagam por meio da interação com um meio físico. Se há um padrão nessas vibrações, diz-se que o som tem uma forma de onda *periódica*. Se não há nenhum padrão, ele é classificado como ruído. A repetição de uma forma de onda periódica dá-se o nome de *ciclo*. O número de ciclos por segundo que ocorre na transmissão de um som representa a *Frequência Fundamental* desse som [19].

No contexto de sinais de voz, F_0 representa a frequência na qual as dobras vocais vibram (abrem-se e fecham-se a cada ciclo) e influi diretamente na tonalidade da voz. Homens apresentam valores de Frequência Fundamental mais baixos (entre 100 e 137 Hz) e mulheres, mais altos (entre 177 e 244 Hz), enquanto crianças apresentam valores ainda mais elevados (entre 206 e 281 Hz). Porém, esses valores ocorrem apenas em vozes saudáveis, pelo fato de as patologias da laringe afetarem o padrão vibratório das dobras vocais, sempre reduzindo sua velocidade de vibração e, conseqüentemente, o valor da Frequência Fundamental da voz. Em [20] foi feita uma pesquisa com um grupo de homens e mulheres que apresentavam Edema e percebeu-se que as mulheres tinham uma frequência fundamental média de 108 Hz, enquanto o grupo de homens com a mesma patologia apresentou Frequência Fundamental média de 91 Hz.

Jitter - consiste da perturbação dos valores de período fundamental (T_0) do sinal. Sendo assim, para obtê-lo, devem ser extraídos primeiramente os valores de T_0 (inverso de F_0) de pequenas janelas do sinal e em seguida verificar, grosso modo, quanto difere cada valor de seus vizinhos. O *Jitter* consiste no somatório dessas diferenças, sendo útil para verificar a estabilidade do sistema fonador, mais especificamente da vibração das dobras vocais, reduzida na presença de patologias [21]. Há 4 tipos de *Jitter* comumente usados. As equações (4) a (7) são utilizadas na obtenção de cada um.

$$jitta = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \quad (4)$$

$$jitt = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (5)$$

$$rap = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - (\frac{1}{3} \sum_{n=i-1}^{i+1} T_n)|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (6)$$

^b Sua configuração varia com o tempo durante a produção da fala

$$ppq5 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} |T_i - (\frac{1}{5} \sum_{n=i-2}^{i+2} T_n)|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (7)$$

Nessas equações, T_i é o valor do período fundamental da i -ésima janela e N é o tamanho da janela.

Shimmer - similar ao *Jitter*, mas empregada na análise da perturbação existente nos valores de amplitude dos picos do sinal. A forma de obter os valores de amplitude para o cálculo é similar: a partir de segmentos do sinal (janelas), é obtida a distância entre os dois picos desse segmento (o mais alto e o mais baixo). Esta medida é útil para verificar a estabilidade da intensidade vocal, afetada pela pressão subglótica e, por sua vez, influenciada pela amplitude de vibração e pela tensão das dobras vocais [22].

Semelhante ao *Jitter*, existem vários tipos de *Shimmer*. As equações (8) a (10) são utilizadas na obtenção de cada um. Vale ressaltar que no cálculo do APQ, foram encontradas várias quantidades de amplitudes adjacentes consideradas, além da utilizada (10): 5, 11 e 55.

$$Shim = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (8)$$

$$ShdB = \frac{1}{N-1} \sum_{i=1}^{N-1} |20 \cdot \log(A_{i+1}/A_i)| \quad (9)$$

$$apq3 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - (\frac{1}{3} \sum_{n=i-2}^{i+2} A_n)|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (10)$$

Nessas equações, A_i é o valor do período fundamental da i -ésima janela e N é o tamanho da janela.

Relação Harmônico Ruído (*Harmonic-to-Noise Ratio - HNR*) - permite avaliar a quantidade de ruído presente em um sinal, em comparação a sua componente periódica. Se mostra adequada ao contexto do processamento digital de sinais de voz, pelo fato de a componente periódica ser decorrente da vibração das dobras vocais e a aperiódica, do ruído glótico [7]. As patologias interferem diretamente no padrão vibratório das dobras vocais, deixando a voz com um aspecto ruidoso, de modo que uma baixa HNR representa um forte indício da presença de uma patologia. Em [23] há a afirmação de que 83% das vozes patológicas apresentam baixa HNR, além de que quanto mais ruído um sinal de voz apresenta, mais avançado é o estágio da patologia.

Análise por Predição Linear - considerada uma das técnicas mais poderosas de análise da fala, parte da ideia de que é possível estimar uma amostra de um sinal de voz por combinação linear de amostras passadas. Isso significa que o sinal como um todo pode ser reconstruído, a partir da estimação das amostras. Essa estimativa deve ser feita de modo a minimizar a soma das diferenças quadradas entre as amostras reais e amostras previstas, o que permite encontrar um conjunto único de p coeficientes preditores. Quanto maior o valor de p , mais precisas serão as estimativas e, conseqüentemente, mais íntegro o sinal reconstruído. É comum seu uso na transmissão de fala, com o intuito de possibilitar a transmissão em uma baixa taxa de bits.

A Análise por Predição Linear é originária da concepção da produção da fala como um sistema linear variante no tempo, excitado por um trem de pulsos quase periódicos e ruído aleatório. A predição linear provê métodos robustos e precisos para estimar os parâmetros que caracterizam este sistema, dentre eles a Codificação por Predição Linear (*Linear Predictive Coding - LPC*), os quais foram usados neste trabalho.

Entropia - medida da incerteza de uma variável aleatória. É fundamental para a Teoria da Informação, sendo empregada como o limiar de compressão máximo possível de ser obtido na compressão de uma mensagem proveniente de uma fonte de informação. O cálculo da entropia é feito por (11).

$$H = -\sum_{i=0}^N p_i \log_2(p_i) \quad (11)$$

Em (11), p_i é a probabilidade de surgimento do i -ésimo símbolo da fonte de informação analisada e N é o tamanho do alfabeto dessa fonte.

B. Método de Compressão PPM

O PPM é um método adotado para comprimir fluxos de dados. Do ponto de vista do poder de compressão, é o compressor de dados mais eficaz, i.e., o estado da arte na área [18] [24]. Seu uso em escala comercial ainda é muito limitado, restringindo-se ao âmbito da pesquisa acadêmica, pelo fato de ser armazenado, durante sua execução, um modelo muito preciso da fonte de dados sendo comprimida, o que acarreta alto consumo de memória e velocidade de execução relativamente baixa, principalmente se comparado com os compressores mais utilizados.

Métodos estatísticos destinados à compressão de dados, tal como o PPM, podem ter sua operação dividida em duas etapas: *modelagem e codificação* [24]. A *modelagem* consiste no armazenamento das probabilidades de símbolos originários de um fluxo de dados. Tendo em vista que o modelo PPM é contextual, também podem ser armazenadas as probabilidades de sequências dos símbolos. A intenção é estimar, o mais precisamente possível, as probabilidades da fonte de dados estacionária abstrata que gerou o fluxo de dados. A escolha do tamanho máximo de contexto armazenado é livre, mas vale salientar que o consumo de memória cresce exponencialmente com esse valor. Além disso, contextos muito altos guardam

muitas informações que aparecem com pouca frequência e, segundo [24], a curva de aprendizado do modelo PPM cessa seu crescimento a partir de um dado tamanho de contexto, o que ocasiona redução na taxa de compressão. Sendo assim, merece atenção o tamanho máximo de contexto empregado em qualquer utilização do PPM.

A segunda etapa de um processo de compressão estatístico é a codificação, que normalmente consiste em atribuir códigos a cada símbolo, visando a gerar um fluxo de dados menor que o original. Para tanto, os símbolos e sequências mais frequentes (identificados na etapa de modelagem) devem receber os menores códigos, com o mínimo de bits possível, enquanto os menos frequentes recebem os maiores. Entretanto, para que seja obtido um código comprimido de tamanho ideal, é necessário atribuir quantidades fracionárias de bits a cada símbolo do fluxo original. O codificador Aritmético alcança este objetivo ao transformar a mensagem original em um número com quantidade ilimitada de casas decimais, dentro de um intervalo limitado durante sua operação. No início do processo, o intervalo é $[0, 1)$. À medida que mais probabilidades são recebidas, este intervalo diminui, mas o número de casas decimais de cada limite aumenta. No fim, tem-se um intervalo pequeno, mas com limites contendo várias casas decimais. O fluxo comprimido pode ser qualquer número dentro deste intervalo. Dessa forma, são atribuídas quantidades fracionárias de bits para cada símbolo do fluxo original.

III. MATERIAIS E MÉTODOS

Nesta pesquisa, foi utilizada a base de dados da Kay Elemetrics, gravada no Voice Speech Lab da Massachusetts Eye and Ear Infirmary (MEEI) [2]. Esta base é composta por 1.381 arquivos, em formato NSP, com 16 bits/amostra, dos quais 666 contêm o pronunciamento da vogal *a* sustentada (conjunto utilizado nesta pesquisa). Desses, 53 correspondem a sinais de vozes saudáveis (cada um dos quais com 3s de duração) e o restante a sinais de vozes patológicas (com duração de 1s). É importante salientar que não há unanimidade entre as frequências de amostragem (taxas de amostragem) dos arquivos. Todos os arquivos contendo sinais de vozes saudáveis têm frequência de amostragem de 50 kHz, ou seja, 50 mil amostras por segundo. Os arquivos contendo sinais de vozes patológicas, por sua vez, têm frequência de amostragem de 25 kHz, mas alguns também têm 50 kHz.

Dos 666 arquivos que contêm a elocução da vogal *a* sustentada, foram utilizados 189, sendo 53 de vozes Normais, 34 de vozes com Edema, 46 de vozes com Paralisia e 56 de vozes com Outras patologias. Em uma classificação, 60% dos arquivos são utilizados para treinamento e os 40% restantes, para testes. A seleção dos arquivos para treinamento é feita aleatoriamente, com o intuito de evitar vieses nos resultados e de ser obtida uma boa estimativa de erro. Todas as classificações incluem etapas como segmentação (com superposição de 50%) e janelamento, que consistem, respectivamente, de extrair apenas uma parte (janela) do sinal e diminuir o efeito das extremidades de cada janela, mantendo o do centro. Os modelos PPM são alimentados com medidas extraídas de cada janela.

A primeira etapa da investigação consistiu em identificar o melhor tipo de entrada para cada classificação executada, dentre parâmetros temporais, acústicos, entropia, coeficientes LPC e combinações entre eles. A identificação do melhor tipo de entrada foi levado a efeito utilizando Projeto Experimental de fator único, apresentado por [25].

A segunda etapa consistiu em analisar o impacto de atividades de pré-processamento como pré-ênfase, distinção de gênero (classificações utilizando sinais de vozes de um único gênero) e subamostragem (utilização de apenas metade das amostras dos sinais que contêm 50 kHz de frequência de amostragem), sobre os resultados, isto é, se implicam ou não em ganho significativo dos percentuais. Paralelamente, foi investigada também a viabilidade do aumento do tamanho do contexto do classificador, já que ele implica em aumento da utilização de recursos computacionais. Em outras palavras, se o aumento desse tamanho implica em aumento significativo dos percentuais de acerto.

A investigação se deu mediante a comparação par a par entre dois casos, sendo um deles advindo da etapa anterior. Por exemplo, caso na etapa anterior tenha se verificado que os melhores resultados da classificação de sinais de vozes Normais quando confrontados com sinais de vozes com Edema foram obtidos utilizando apenas coeficientes LPC, nesta etapa, estes resultados foram comparados par a par com os resultados obtidos ao aplicar filtro de pré-ênfase, distinção de gênero e subamostragem e ao utilizar tamanhos de contexto entre 1 e 4.

As análises foram feitas utilizando Intervalos de Confiança (procedimento referido em [25] como *t-test*) ou teste de Mann-Whitney, a depender da distribuição amostral dos resultados comparados. A utilização destes testes é justificada pelo fato de que o intuito desta investigação era verificar se a aplicação de atividades de pré-processamento ou aumento do tamanho do modelo forneciam resultados significativamente maiores que os resultados obtidos nas classificações sem essas etapas, de modo a justificar sua inclusão.

Por fim, identificada a melhor configuração do classificador para cada caso, foram obtidos percentuais que caracterizassem a eficácia do PPM no contexto de diagnóstico de patologias da fala, por meio do procedimento conhecido como Validação Cruzada com 4 parcelas. Nesta etapa, diferentemente das anteriores, não é possível inserir aleatoriedade, pela própria natureza do processo.

IV. RESULTADOS

Nesta seção serão apresentados e discutidos os resultados obtidos em todas as etapas da investigação.

A. Identificação do melhor tipo de entrada

Nesta etapa da investigação, objetivou-se identificar o melhor tipo de entrada para cada caso de classificação. Conforme mostrado na Tabela 1, descobriu-se que não há um tipo de entrada que retorne os melhores percentuais em todos os casos. O melhor tipo de entrada varia com a classificação executada.

TABELA 1. OS MELHORES TIPOS DE ENTRADA E OS RESULTADOS OBTIDOS

Classificação	Mediana em % da primeira classe e tipo de entrada	Mediana em % da segunda classe e tipo de entrada
Normal x Tudo	100,0 - Temporais; Entropia; Temporais e Entropia	96,3 - Temporais e Entropia
Normal x Edema	100,0 - Temporais; Temporais e Entropia	100,0 - Temporais e Entropia
Normal x Paralisia	100,0 - Temporais; Temporais e Entropia	94,7 - Temporais
Normal x Outras	100,0 - Temporais; Temporais e Entropia	95,6 - Temporais
Edema x Paralisia	64,2 - LPC	73,7 - Temporais e Acústicos
Edema x Outras	57,1 - LPC	65,2 - Acústicos
Paralisia x Outras	68,4 - Temporais	65,2 - Temporais, Acústicos e Entropia

É possível perceber que há casos em que mais de um tipo de entrada fornece resultados semelhantes. Para tanto, é recomendável a utilização daquele que utilize menos memória. Por exemplo, na classificação de arquivos Normais confrontados com arquivos de vozes com Edema, recomenda-se a utilização apenas de parâmetros temporais, por utilizarem menos memória.

É válido destacar também a forte participação dos parâmetros temporais na maioria dos casos, seja isoladamente ou em combinação com outros parâmetros, o que indica o potencial deste tipo de dado na classificação de patologias da fala.

B. Influência de pré-processamentos nos resultados

Nenhuma atividade de pré-processamento executada implicou o aumento significativo dos resultados. Ao invés disto, notou-se com frequência a redução significativa nos resultados devido ao pré-processamento dos sinais de entrada. Cenário semelhante foi observado na variação do tamanho do contexto do classificador. Sendo assim, é possível concluir que, em todos os casos, a melhor configuração consistiu da não inclusão de nenhuma etapa de pré-processamento e da adoção do contexto 0 em todo o processo.

C. Percentuais via validação cruzada

Tendo sido encontradas as melhores e mais viáveis configurações para cada caso de classificação, foi utilizado o procedimento de Validação Cruzada, a fim de obter os percentuais de acerto para cada um deles, com o intuito de caracterizar o PPM quanto à eficácia, na classificação de patologias da fala. Os resultados obtidos encontram-se na Tabela 2

TABELA 2. PERCENTUAIS OBTIDOS COM VALIDAÇÃO CRUZADA UTILIZANDO AS MELHORES CONFIGURAÇÕES

Classificação	Resultados obtidos (mediana %)
Normal x Tudo - Normal	100,0
Normal x Tudo - Tudo	94,1
Normal x Edema - Normal	100,0
Normal x Edema - Edema	95,0
Normal x Paralisia - Normal	100,0
Normal x Paralisia - Paralisia	92,3
Normal x Outras - Normal	100,0
Normal x Outras - Outras	96,4
Edema x Paralisia - Edema	50,0
Edema x Paralisia - Paralisia	84,6
Edema x Outras - Edema	40,0
Edema x Outras - Outras	75,0
Paralisia x Outras - Paralisia	57,6
Paralisia x Outras - Outras	64,2

D. Eficiência da abordagem metodológica

A eficiência desta abordagem metodológica pode ser analisada quanto à velocidade de execução e à utilização de memória.

Quanto à velocidade de execução, foram analisados os tempos de execução de 10 classificações com as classes mais bem representadas da base de dados: Normal e Tudo, com 53 e 56 arquivos cada uma. Essas classes foram escolhidas por serem as mais bem representadas, de modo que os tempos das outras classificações serão certamente menores que os relatados na Tabela 3. Os intervalos apresentados correspondem aos intervalos de confiança a 95% de significância estatística dos tempos de execução do treinamento e de todo o processo, o que inclui treinamento e testes com vários sinais de voz.

Com relação ao uso de memória, também foi registrada a utilização máxima de memória em 10 execuções desta mesma classificação. O intervalo de confiança a 95% de significância destes registros foi extraído, tendo sido obtido o intervalo [15,02; 15,77] MB. A ferramenta de monitoramento utilizada foi a VisualVM, fornecida pela distribuidora da plataforma Java.

TABELA 3. TEMPOS DE EXECUÇÃO DA CLASSIFICAÇÃO NORMAL X TUDO

Etapa	Intervalo de confiança (ms)
Treinamento	[405,3; 436,6]
Todo o processo	[804,0; 844,0]

E. Discussão

É possível constatar que os melhores resultados estão associados a classificações entre sinais de vozes Normais e sinais de vozes que apresentam alguma patologia, executadas com o intuito de detectar a presença de patologias. Nelas, obtiveram-se resultados entre 92,3 e 100%. Porém, estes resultados apenas confirmam o potencial do PPM no contexto da classificação de padrões vocais, haja vista que resultados semelhantes foram alcançados em pesquisas anteriormente conduzidas, tais como [1] [3] [4].

Um resultado importante, almejado por pesquisadores da área, consiste da discriminação precisa de patologias distintas. Na pesquisa ora descrita, foram obtidos resultados deste tipo que podem ser considerados bons e medianos, a depender da classificação. Os piores resultados são oriundos das classificações em que o intuito é diagnosticar sinais de voz com Edema. Isto se deve à pequena representatividade desta patologia na base de dados utilizada, haja vista que dos 43 registros de arquivos de vozes com Edema, apenas 3 são de vozes que apresentam apenas a patologia Edema. Todos os demais apresentam outras patologias, além desta. Alguns deles chegaram a apresentar outras 5 (cinco) patologias. Quanto mais patologias um registro de voz apresenta, menos tal registro representará a patologia principal que contém [5], o que pode confundir os classificadores utilizados, comprometendo os resultados.

Contudo, bons resultados de discriminação de patologias também foram obtidos, especialmente aqueles de classificações com o intuito de discriminar outras patologias em confronto com sinais de vozes com Edema, que se encontram entre 75 e 84,6%. Mesmo assim, eles podem ser considerados insuficientes para um ambiente clínico real.

V. CONSIDERAÇÕES FINAIS

O presente trabalho visou a apresentar uma nova abordagem de utilização para um método já conhecido no contexto da compressão de dados. É possível constatar, pelos resultados apresentados na Seção IV, que a abordagem metodológica proposta se mostra promissora, haja vista que boa parte dos resultados de classificação obtidos é superior a 90%. Porém, os casos de discriminação de patologias ainda carecem de investigação adicional, já que os resultados podem ser considerados inadequados para a diagnose em ambientes clínico reais.

Como sugestão de trabalhos futuros, recomenda-se a identificação de entradas e/ou combinações que permitam melhor distinção entre patologias, isto é, que sejam afetadas de forma diferente para as diferentes patologias às quais o sistema de produção da fala humano é suscetível.

AGRADECIMENTOS

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro durante o período da pesquisa, e à Universidade Federal de Campina Grande (UFCG), especialmente ao prof. Benedito, pela disponibilização da base de dados.

REFERÊNCIAS

- [1] W. C. de A. Costa, F. M. Assis, B. G. Aguiar Neto, S. C. Costa e V. J. D. Vieira, "Pathological voice assessment by recurrence quantification analysis", BRC 2012, in press.
- [2] K. Elemetrics, "Kay elemetrics corp. disordered voice database", Model 4337, 03 Ed., 1994.
- [3] S. L. N. C. Costa, "Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas", Tese (Doutorado em Engenharia Elétrica) - UFCG, Campina Grande, 2008.
- [4] J. V. M. L. Marinus, "Estudo de técnicas para classificação de vozes afetadas por patologias", Dissertação (Mestrado em Ciência da Computação) - UFCG, Campina Grande, 2010.
- [5] R. R. Brandt, "Classificação de vozes patológicas utilizando análise paramétrica e não-paramétrica", Tese (Doutorado em Engenharia Elétrica) - UFCG, Campina Grande, 2012.
- [6] Y. Hu e P. C. Loizou, "Evaluation of objective quality measures for speech enhancement", IEEE Trans. on Audio, Speech and Lang. Proc., vol. 16, pp. 229-238, 2008.
- [7] J. Lopes et al., "A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa", IJTS, in press.
- [8] J. Oates, "Auditory-perceptual evaluation of disordered voice quality", Fol. Phon. et Logop., vol. 61, pp. 49-56, 2009.
- [9] J. V. Vieira, S. C. Costa e W. C. Costa, "Análise de quantificação de recorrência e análise discriminante aplicadas à classificação de sinais de vozes saudáveis e sinais de vozes patológicas", VII CONNEPI, in press.
- [10] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz e G. Castellanos-Domínguez, "Automat detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients", IEEE Trans. on Biom. Engin., vol. 58, pp. 370-379, 2011.
- [11] H. A. Patil e P. N. Baljekar, "Classification of normal and pathological voices using TEO phase and mel cepstral features", ICSPCS 2012, in press.
- [12] N. Raju, T. L. Priya, S. Mathini e P. Preethi, "Normal versus pathology voice-an analysis", ICCCA 2012, in press.
- [13] J. R. Orozco et al., "Voice pathology detection in continuous speech using nonlinear dynamics", ISSPA 2012, in press.
- [14] J. dos S. Lima, T. T. C. Palitó, S. C. Costa e S. E. N. Correia, "Classificação de sinais vozes patológicas por meio do parâmetro de Hurst e LDA", XXX SBTr, in press.
- [15] B. C. Coutinho, J. L. de M. Macêdo, A. Rique Júnior e L. V. Batista, "Atribuição de autoria usando PPM", XXV CSBC, in press.
- [16] Barufaldi et al. "Text classification by literary period using PPM-C data compression", STIL 2009, in press.
- [17] T. C. de S. Honório, L. V. Batista e R. C. M. Duarte, "Texture classificatoim using prediction by partial matching models", WVC 2009, in press.
- [18] T. F. L. Medeiros et al. "Heart arrhythmia classification using the PPM algorithm", BRC 2011, in press.
- [19] C. Roads, Computer Music Tutorial. Massachusetts: MIT Press, 1995.
- [20] S. Bennett, S. Bishop e S. M. Lumpkin, "Phonatory characteristics associated with bilateral diffuse polypoid degeneration", The Laryngoscope, vol. 97, pp. 446-450, 1987.
- [21] J. P. Teixeira, D. Ferreira e S. Carneiro, "Análise acústica vocal - determinação do jitter e shimmer para diagnóstico de patologias da fala", CLME 2011, in press.
- [22] M. Farrús e J. Hernando, "Using jitter and shimmer in speaker verification", IET Sig. Proc., vol. 3, pp. 247-257, 2008.
- [23] V. Parsa e D. Jamieson, "Identification of pathological voices using glottal measures", J. of Speech and Hear. Res., vol. 43, pp. 469-485, 2000.
- [24] D. Salomon, Data Compression: The Complete Reference, 3rd ed. New York: Springer, 2004.
- [25] R. Jain, The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling. New York: Wiley-Interscience, 1991.