

Classificação de sinais de vozes saudáveis e patológicas por meio da combinação entre medidas da análise dinâmica não linear e codificação preditiva linear

Washington César de Almeida Costa*, Silvana Luciene do Nascimento Cunha Costa, Francisco Marcos de Assis, Benedito Guimarães Aguiar Neto

Resumo Análise acústica tem sido sugerida como uma ferramenta auxiliar não invasiva e de baixo custo ao diagnóstico de patologias laringeas. Diversas são as técnicas empregadas, entre as quais técnicas baseadas no modelo linear de produção da fala e na análise dinâmica não linear de sinais de vozes. O primeiro método é baseado na Teoria fonte-filtro, em que a fonte é a laringe e o filtro é o trato vocal. Nesse modelo, os sons surdos são modelados por uma fonte de ruído aleatório e os sons sonoros por um trem de impulsos na frequência fundamental do locutor. Na abordagem não linear são considerados aspectos da voz humana, não explorados no modelo linear, tais como: variação temporal da forma do trato vocal, as ressonâncias associadas à sua fisiologia, as perdas devido ao atrito nas paredes internas do trato vocal, a radiação do som nos lábios, o acoplamento nasal e o comportamento dinâmico associado à vibração das pregas vocais. Neste trabalho são associadas as duas abordagens e avaliado o desempenho na classificação com as características, de forma individual, e a partir da combinação das mesmas. São empregadas oito medidas oriundas da análise dinâmica não linear (dimensão de correlação, quatro medidas de entropia, expoente de Hurst, maior expoente de Lyapunov e o primeiro mínimo da função de informação mútua), além de coeficientes LPC, obtidos a partir da análise preditiva linear. Os resultados sugerem a viabilidade da técnica empregada para a discriminação entre vozes saudáveis e patológicas em geral, como também entre vozes afetadas por patologias laringeas distintas como edema, nódulos e paralisia nas pregas vocais.

Palavras-chave Patologias laringeas, Análise acústica, Classificação de vozes patológicas, Análise dinâmica não linear, Análise preditiva linear.

Healthy and pathological voice assessment by means of nonlinear dynamic analysis measures and linear predictive coding

Abstract *Acoustic analysis has been suggested as a noninvasive aiding and low cost tool for laryngeal disease diagnosis. Several techniques are employed using either the linear model of speech production, or the nonlinear dynamic analysis of voice signals. The first method is based on source-filter theory, in which the source is the larynx and the filter is the vocal tract. In this model, the unvoiced sounds are modeled by a random noise source and the voiced ones by impulse train at the speaker fundamental frequency. In nonlinear approach, aspects of the human voice are considered, not explored in the linear model, such as temporal variation of the vocal tract shape, resonances associated with its physiology, losses due to friction in the vocal tract inner walls, sound radiation in the lips, nose coupling and dynamic behavior associated with vocal fold vibration. This work combines the two approaches and evaluates the performance in classifying the features individually, and from their combination. Eight measures are employed derived from the nonlinear dynamic analysis (correlation dimension, four entropy measures, Hurst exponent, the largest Lyapunov exponent and the first minimum of mutual information function), besides LPC coefficients obtained from linear predictive analysis. The results suggest the feasibility of the employed technique to discriminate between healthy and pathological voices in general, but also among specific laryngeal diseases as vocal fold edema, nodules and paralysis.*

Keywords *Laryngeal diseases, Acoustic analysis, Pathological voice assessment, Nonlinear dynamic analysis, Linear predictive analysis.*

Introdução

Patologias na laringe têm aumentado muito nos últimos anos devido a hábitos sociais não saudáveis – tabagismo e alcoolismo – e o abuso vocal. As técnicas tradicionais empregadas pela comunidade médica ocorrem pela escuta da voz do paciente, cujo diagnóstico depende da experiência do profissional ou pela inspeção direta das pregas vocais por meio de técnicas laringoscópicas, o que causa certo desconforto ao paciente (Godino-Llorente *et al.*, 2006).

A análise acústica tem sido sugerida, nas últimas décadas, como uma ferramenta auxiliar ao diagnóstico médico de patologias laringeas, acompanhamento de terapias vocais, tratamentos medicamentosos e acompanhamento antes e após cirurgias. É uma técnica não invasiva, já que necessita apenas da gravação da voz do paciente, a qual pode ser armazenada e processada. Dessa forma, é possível extrair e avaliar características e parâmetros do sinal de voz e observar as consequências causadas nos mesmos pela presença de patologias na laringe. Esta técnica não substitui o diagnóstico dos exames laringoscópicos, mas pode ser empregada em pré-diagnóstico, acompanhamentos de tratamentos medicamentosos, pós-cirúrgicos, terapia vocal e indicação da necessidade ou não de exames mais sofisticados para um diagnóstico final.

Por meio da análise acústica, utilizando técnicas de processamento digital de sinais, a voz do paciente pode ser gravada num computador pessoal e enviada para um especialista em outra localidade e este, por meio de *software* específico, avaliar a qualidade vocal. Este sistema poderia, portanto, ser aplicado em redes de telemedicina, nas quais os serviços ofertados podem incluir o compartilhamento de arquivos de prontuários, consultas, exames e segunda opinião. Além disso, é possível a adoção de medidas simples e de baixo custo, como a implantação de sistemas de análise com diagnósticos remotos, que pode contribuir para diminuir a carência de especialistas.

Diversas técnicas têm sido empregadas para quantificar e avaliar as desordens vocais provocadas por patologias laringeas. As técnicas mais tradicionais empregam medidas tais como a frequência fundamental, *jitter*, *shimmer*, APQ, HNR, entre outras (Boyanov *et al.*, 1993; Manfredi *et al.*, 1999). Entretanto, dependendo do grau de severidade da patologia, a obtenção da frequência fundamental se torna uma tarefa bastante complexa e o resultado obtido pode não ser confiável (Godino-Llorente *et al.*, 2006). Outras técnicas se baseiam no modelo linear de produção da fala (modelo fonte-filtro), em que a fonte é a laringe e o filtro é o trato vocal. Nesse modelo, os sons surdos são modelados por uma fonte

de ruído aleatório e os sons sonoros por um trem de impulsos na frequência fundamental do locutor (O’Shaughnessy, 2000; Rabiner e Schafer, 1978).

Uma questão fundamental para que um processo automático de classificação entre vozes saudáveis e patológicas seja eficiente, é a escolha de quais características, extraídas do sinal de voz, são as mais apropriadas para capturar as desordens vocais, provocadas pela presença da patologia.

Apesar do relativo sucesso do uso do modelo linear para a produção da fala em diversas aplicações, estudos mais recentes têm apontado para a evidência do caos na voz humana (Jiang *et al.*, 2006; Kokkinos e Maragos, 2005; Henriquez *et al.*, 2009; Zhang e Jiang, 2008). A análise dinâmica não linear de sinais de voz tem sido considerada por levar em conta aspectos da voz humana, não explorada na abordagem linear, tais como: variação temporal da forma do trato vocal, as ressonâncias associadas à sua fisiologia, as perdas devido ao atrito viscoso nas paredes internas do trato vocal, a suavidade dessas paredes internas, a radiação do som nos lábios, o acoplamento nasal e a flexibilidade (comportamento dinâmico) associada à vibração das pregas vocais (Kumar e Mullik, 1996).

Antes de aplicar técnicas não lineares, tais como aquelas inspiradas na teoria do caos, a fenômenos dinâmicos que ocorrem na natureza, é necessário primeiro perguntar se a utilização de tais técnicas avançadas é justificada pelos dados. Enquanto muitos processos na natureza parecem muito improváveis, a priori, serem lineares, a possível natureza não linear pode não ser evidente em aspectos específicos da sua dinâmica. O fato de que um sistema contenha componentes não lineares não implica, necessariamente, que esta não linearidade também se reflita em um sinal específico obtido a partir desse sistema (Schreiber e Schmitz, 2000).

No caso particular da voz humana, diversos trabalhos têm identificado a influência das não linearidades presentes no sistema de produção vocal a partir da análise do próprio sinal de voz (Little *et al.*, 2007; Tokuda *et al.*, 2001). Em um desses trabalhos, Max A. Little (Little, 2006) realizou uma investigação a cerca de que sejam mantidos os pressupostos de um sistema linear variante no tempo para sinais de vozes, apesar de mudanças nos formantes (para diferentes vogais) ou na fonte de energia acústica (ruído acústico em consoantes e vibração das pregas vocais em vogais). Os resultados obtidos indicaram que para a maioria dos sons vocálicos, tanto saudáveis como patológicos, a hipótese de que tais sinais tenham sido gerados a partir de um sistema linear pode ser rejeitada, e que modelos não lineares, estocásticos ou determinísticos, podem ser mais confiáveis. Por outro lado, no caso

de sons consonantais fricativos saudáveis e algumas vozes patológicas (caso da voz soprosa) não se pode descartar o modelo linear.

Recentes pesquisas relacionadas às séries temporais, geradas a partir dos mecanismos de produção da voz humana, têm sido realizadas considerando-se as técnicas da dinâmica não linear e da teoria do caos com objetivos variados, dentre os quais podem ser destacados: classificação de fonemas (Johnson *et al.*, 2005; Kokkinos e Maragos, 2005), reconhecimento automático de locutor (Petry, 2002), discriminação entre vozes saudáveis e patológicas, diagnóstico de patologias laringeas e avaliação de efeitos de tratamentos clínicos (Dajer, 2006; Henríquez *et al.*, 2009; Jiang *et al.*, 2006; Scalassara *et al.*, 2008; Torres *et al.*, 2003; Zhang e Jiang, 2008).

Para que um sistema de auxílio a diagnósticos possa ser implementado com eficiência, torna-se primordial o estudo de técnicas que ofereçam um excelente desempenho, para que o erro de diagnóstico seja mínimo. A investigação acerca de quais são as melhores características que distinguem um sinal de voz saudável de um sinal de voz patológica e, principalmente, que permita distinguir entre patologias ainda é um campo em desenvolvimento.

Na literatura específica, a maioria das pesquisas se baseia em discriminar entre vozes saudáveis e patológicas em geral sem, no entanto, discriminar entre patologias distintas. Neste trabalho, pretende-se investigar o potencial discriminativo de medidas oriundas da análise dinâmica não linear, para classificar sinais de vozes saudáveis e patológicas, bem como distinguir qual patologia está afetando a qualidade vocal. As patologias laringeas consideradas neste trabalho são paralisia nas pregas vocais, edemas de Reinke e nódulos vocais. As medidas empregadas, oriundas da análise dinâmica não linear são: Dimensão de correlação, entropia de correlação, entropia aproximada, entropia de Shannon, entropia de Tsallis, expoente de Hurst, maior expoente de Lyapunov e primeiro mínimo da função de informação mútua. A técnica de análise de discriminante quadrática é aplicada para classificar os sinais em saudáveis ou patológicos, considerando as características de forma individual e combinada.

No intuito de melhorar as taxas de classificação na discriminação entre patologias é realizada uma combinação das medidas da análise não linear com coeficientes LPC (*Linear Prediction Coding*), obtidos a partir da análise linear para o modelo de produção da fala. O objetivo dessa combinação é investigar se características de abordagens diferentes podem extrair

informações que, combinadas, possam refletir mais especificamente as desordens vocais provocadas pela presença de uma determinada patologia na laringe.

Materiais e Métodos

Base de dados

Neste trabalho, os sinais processados são provenientes da base de dados comercialmente disponível, a *Disordered Voice Database, Modelo 4337*, gravada pelo Massachusetts Eye and Ear Infirmary (MEEI) (Kay Elemetrics, 1994). A base de dados contém pronúncias sustentadas da vogal /a/, sendo 53 arquivos de locutores com voz saudável e 657 com vozes patológicas (vozes afetadas por patologias laringeas). Foram selecionados 130 sinais, sendo os 53 sinais de vozes saudáveis e 77 sinais de pacientes com patologia na laringe (31 com edema de Reinke, 28 com paralisia e 18 com nódulos). Os sinais de vozes saudáveis, originalmente amostrados a uma frequência de 50.000 amostras/s, foram subamostrados a 25.000 amostras/s para equiparar o número de amostras aos sinais patológicos. Os sinais de vozes têm duração média de um segundo para os sinais patológicos e de 3 segundos para os sinais de vozes saudáveis. Os sinais disponibilizados na base de dados são previamente processados, por isso, não contém intervalos de silêncio antes ou depois da elocução. Evitando, assim, quaisquer problemas frequentemente encontrados em início e fim de gravação.

Metodologia

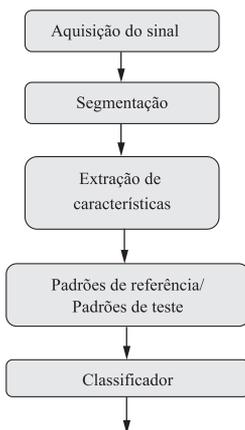
O sistema de classificação é representado de forma resumida pela Figura 1. Após a aquisição do sinal por um microfone, o sinal armazenado em um computador pessoal é processado. São analisados 960 ms de cada sinal de voz, sendo 30 quadros de 32 ms, de forma a preservar a estacionariedade. A seguir é feita a extração de características e logo após a classificação dos sinais. São realizadas duas abordagens distintas para a extração de características: uma abordagem paramétrica, baseada no modelo linear de produção da fala, e outra em medidas obtidas a partir da análise dinâmica não linear. Na abordagem paramétrica, os coeficientes LPC são extraídos de cada sinal, após sua segmentação, configurando um vetor de ordem p (ordem do filtro de predição) para cada segmento. Neste trabalho são utilizados diferentes valores de p (12, 16, 20 e 24), com o intuito de avaliar o efeito do aumento da ordem de predição no desempenho da classificação.

Extração de características

Após a segmentação do sinal é realizada a extração de características, feitas sob as duas abordagens para o modelo de produção de fala (linear e não linear), descritas a seguir.

Análise dinâmica não linear

A análise dinâmica de sinais pode ser feita por meio de um modelo matemático associado ao sistema de produção desses sinais ou por meio da análise de séries temporais. O ponto essencial desta análise é que uma série temporal contém informações sobre variáveis não observáveis do sistema, o que permite a reconstrução do espaço de estados (Savi, 2006). Cada sinal de voz a ser analisado corresponde a uma série temporal.



Resultado: Saudável ou Patológica (paralisia, edema ou nódulo)

Figura 1. Diagrama em blocos do sistema de classificação de vozes patológicas.

Figure 1. Block diagram of pathological voices classification system.

A técnica de reconstrução do espaço de estados é baseada no teorema da imersão de Takens (1981). Ele demonstrou que, com o uso da técnica dos tempos de retardo, ou método das coordenadas defasadas, é possível reconstruir certas propriedades topológicas do espaço de estados (atrator) a partir da série temporal, $\{x_i\}$ em que vetores ξ_i m -dimensionais são reconstruídos, de forma que:

$$\xi_i = \{x(t_i), x(t_i + \tau), \dots, x(t_i + (m-1)\tau)\} \quad (1)$$

onde, $i = 1, 2, \dots, T - (m-1)\tau$, T é o número total de amostras, m é a chamada dimensão de imersão e τ é o passo de reconstrução ou tempo de defasagem (Fiedler-Ferrara e Prado, 1994).

A determinação eficiente dos parâmetros de reconstrução (tempo de defasagem e dimensão de imersão) é de extrema importância para a identificação dos invariantes geométricos do sistema, tais como a dimensão do atrator e os expoentes de Lyapunov. A partir da reconstrução do espaço de fases dos sinais analisados é realizada a extração das medidas de dinâmica não linear.

Na Figura 2 está ilustrada a reconstrução em duas dimensões do atrator de um dos sinais de voz analisados utilizando: (a) um passo de reconstrução bem menor do que o adequado ($\tau = 1$); (b) utilizando um passo de reconstrução “ótimo” determinado com o uso da função de informação mútua ($\tau = 5$); e (c) usando um passo bem maior do que o adequado ($\tau = 10$).

O método utilizado para obtenção do passo de reconstrução, utilizado neste trabalho, é devido a Fraser e Swinney (1986) e está baseado na função de informação mútua. Neste método, o valor do tempo de defasagem ideal corresponde ao primeiro mínimo local, quando este existir, da função de informação mútua média $I(\tau)$.

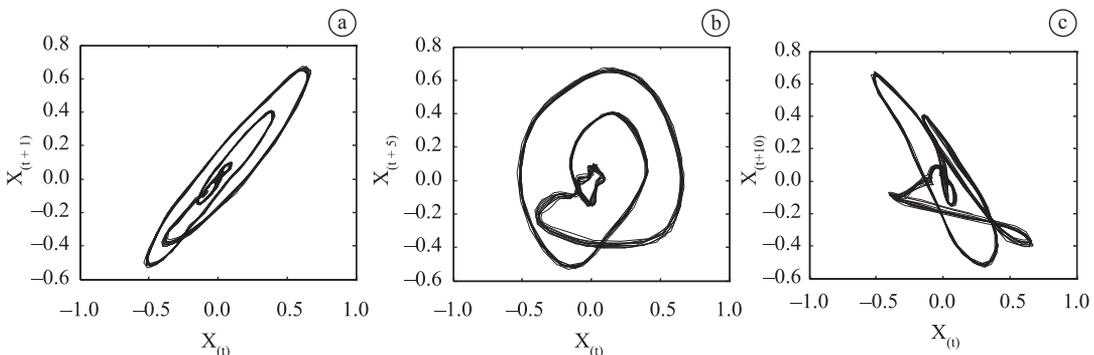


Figura 2. Reconstrução do atrator da vogal sustentada /a/ em um intervalo de 32 ms. Parâmetros: $m = 2$ e a) $\tau = 1$ amostra, b) $\tau = 5$ amostras e c) $\tau = 10$ amostras.

Figure 2. Atractor reconstruction of the sustained vowel /a/ for a 32 ms frame. Parameters: $m = 2$ and a) $\tau = 1$ sample, b) $\tau = 5$ samples and c) $\tau = 10$ samples.

Função de informação mútua

A informação mútua média fornece as mesmas informações que a função de correlação fornece em sistemas lineares sendo, na realidade, um tipo de generalização para sistemas não lineares (Savi, 2006). Na Figura 3 está ilustrado o comportamento da $I(\tau)$ em função de τ , para o mesmo segmento do sinal cujo atrator está mostrado na Figura 2. A linha tracejada indica o ponto onde ocorre o primeiro mínimo de $I(\tau)$ (FMMI – *First Minimum of Mutual Information*) e assim a determinação do passo de reconstrução “ótimo” para este segmento ($\tau = 5$).

A partir de um histograma de b intervalos de classe (*bins*), criado para estimar a distribuição de probabilidade dos dados de um sinal $x(t)$, a informação mútua média entre $x(t)$ e sua versão defasada $x(t + \tau)$ é estimada por meio da expressão:

$$I(\tau) = \sum_{i=1}^b \sum_{j=1}^b P_{i,j}(x(t), x(t+\tau)) \log_2 \left[\frac{P_{i,j}(x(t), x(t+\tau))}{P_i(x(t))P_j(x(t+\tau))} \right] \quad (2)$$

em que $P_i(x(t))$ é a probabilidade de que o sinal $x(t)$ assuma um valor dentro do i -ésimo intervalo do histograma, $P_j(x(t+\tau))$ é a probabilidade de que $x(t + \tau)$ esteja no j -ésimo intervalo e $P_{i,j}(x(t), x(t+\tau))$ é a probabilidade de que, simultaneamente, o sinal de $x(t)$ esteja no i -ésimo intervalo e $x(t + \tau)$ no j -ésimo intervalo (Kantz e Schreiber, 2004).

Dimensão de correlação

A dimensão de correlação, uma das formas mais comuns de se calcular a dimensão de um atrator, é uma medida geométrica de uma trajetória no espaço de fase que descreve o quanto dois pontos dessa trajetória são correlacionados. Ela mede a probabilidade de se encontrar um par aleatório de pontos dentro de um determinado volume elementar (Savi, 2006).

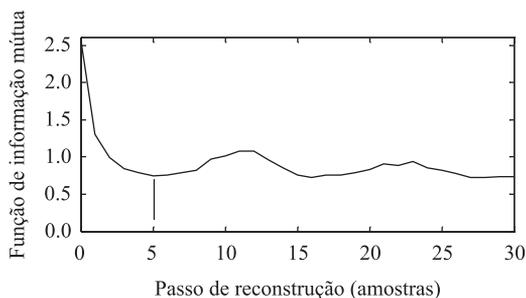


Figura 3. Função de Informação Mútua e determinação do passo de reconstrução ótimo para um segmento de 32 ms da vogal sustentada /a/.
Figure 3. Mutual Information Function and the determination of optimal reconstruction step for a 32 ms frame of /a/ sustained vowel.

Grassberger e Procaccia (1983) desenvolveram um algoritmo para o cálculo da dimensão de correlação, em que a probabilidade de se ter dois pontos do atrator numa caixa de lado ϵ é aproximada pela probabilidade de que a distância entre dois pontos seja menor que ϵ . Esse algoritmo fornece um limite inferior e um valor aproximado para a dimensão fractal (Fiedler-Ferrara e Prado, 1994).

Seja $q(\epsilon)$ a fração de pontos do atrator que está dentro de uma hipersfera de raio ϵ , centrada no ponto ξ_i (i -ésimo vetor de imersão m -dimensional). Essa fração é expressa pela relação:

$$q(\epsilon) = \frac{1}{N} \sum_{j=1}^N \theta(\epsilon - \|\xi_i - \xi_j\|) \quad (3)$$

sendo N o número de pontos no atrator. A função degrau $\theta(x)$ é tal que $\theta = 1$ se $x \geq 0$, e $\theta = 0$ se $x < 0$. Define-se a integral de correlação $C_m(\epsilon)$ como:

$$C_m(\epsilon) = \frac{1}{N-1} \lim_{N \rightarrow \infty} \sum_{i=1(i \neq j)}^N q(\epsilon) \quad (4)$$

A dimensão de correlação é então calculada, para cada valor de m , como:

$$D_2 \square \lim_{\epsilon \rightarrow 0} \frac{\log C_m(\epsilon)}{\log(\epsilon)} \quad (5)$$

ou seja, D_2 é a inclinação de uma reta definida pelo gráfico $\log C_m(\epsilon) \times \log(\epsilon)$, para um valor infinitesimal de ϵ .

Na Figura 4 é apresentado o comportamento da dimensão de correlação, em função de m e de ϵ , para um segmento de 32 ms de um sinal de voz saudável. A fim de se estimar o valor de D_2 em cada segmento do sinal são consideradas as inclinações da integral de correlação para diferentes dimensões de imersão, com m variando de 4 a 10.

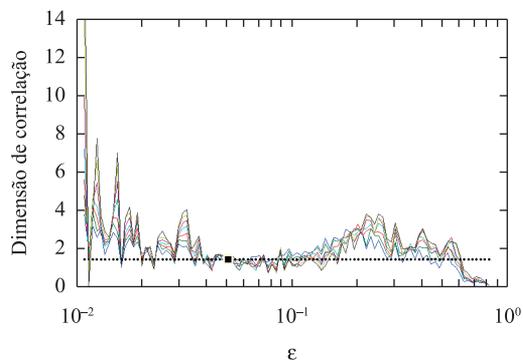


Figura 4. Comportamento da dimensão de correlação para um segmento de 32 ms de um sinal de voz saudável em função da dimensão de imersão e do valor de ϵ , em que as curvas correspondem as diferentes dimensões de imersão ($m = 4, \dots, 10$) de baixo para cima, respectivamente.

Figure 4. Correlation dimension behavior for a 32 ms segment of a healthy speech signal in function of the embedded dimension and ϵ value, wherein the curves correspond to the different embedded dimensions ($m = 4, \dots, 10$) upwards, respectively.

A partir da inspeção visual do comportamento dos valores da dimensão de correlação em função de m e de ϵ é levado a efeito a determinação do valor de ϵ que oferece o menor coeficiente de variação entre as curvas dentro de um intervalo de convergência. O valor de D_2 para o segmento é estimado como sendo a média dos valores obtidos de cada curva para este valor de ϵ . Para o segmento ilustrado na Figura 4, o quadrado sobreposto ao gráfico indica os valores encontrados ($\epsilon = 0,050$ e $D_2 = 1,43$).

Entropia de Shannon

Dado um determinado sinal, a entropia de Shannon, H_1 , é definida como (Shannon, 1948):

$$H_1 = -\sum_{i=1}^M p_i \log_2(p_i) \quad (6)$$

em que p_i é a probabilidade de que o sinal pertença a um intervalo i e M é o número de partições em seu espaço de fase.

A entropia de Shannon, H_1 , é a medida de informação necessária para localizar um sistema em um determinado estado, significando que H_1 é a medida da incerteza sobre o sistema físico (Torres *et al.*, 2003). As medidas de entropia avaliam o grau de desordem de um sinal, podendo ser empregadas para avaliar as desordens nos sinais de vozes afetados por patologias na laringe (Little *et al.*, 2007; Scalassara *et al.*, 2008).

Entropia de correlação

O procedimento de Grassberger-Procaccia, permite estimar a entropia de Correlação, K_2 , a partir das integrais de correlação $C_m(\epsilon)$ (Equação 4), da seguinte forma:

$$K_2(\epsilon) = \frac{1}{\tau} \ln \frac{C_m(\epsilon)}{C_{m+1}(\epsilon)} \quad (7)$$

em que τ é o passo utilizado na reconstrução e m é a dimensão de imersão.

Na Figura 5 é apresentado o comportamento da entropia de correlação, em função de m e de ϵ , para um segmento de 32 ms de um sinal de voz patológico. A fim de se estimar o valor de K_2 em cada segmento do sinal são consideradas as curvas com os valores da entropia de correlação para diferentes dimensões de imersão, com m variando de 4 a 10 e diferentes valores de ϵ .

A partir da inspeção visual do comportamento dos valores da entropia de correlação em função de m e de ϵ é levada a efeito a determinação do valor de ϵ que oferece o menor coeficiente de variação entre as curvas dentro de um intervalo de convergência. O valor de K_2 para o segmento é estimado como sendo

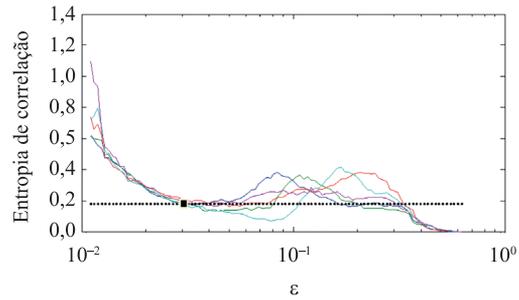


Figura 5. Comportamento da entropia de correlação para um segmento de 32 ms de um sinal de voz patológico, em função da dimensão de imersão e do valor de ϵ em que as curvas correspondem a diferentes dimensões de imersão ($m = 4, \dots, 10$), de baixo para cima, respectivamente.

Figure 5. Behavior of entropy correlation for a 32 ms frame of a pathological speech signal related to embedded dimension and ϵ value, wherein the curves correspond to the different embedded dimensions ($m = 4, \dots, 10$), upwards, respectively.

a média dos valores obtidos de cada curva para este valor de ϵ . A entropia de correlação do sinal é tomada como sendo a média calculada sob todos os segmentos do sinal analisado. No caso do segmento ilustrado na Figura 5, o quadrado sobreposto ao gráfico indica os valores encontrados ($\epsilon = 0,031$ e $K_2 = 0,184$).

Entropia aproximada

A entropia aproximada, $ApEn$, é uma medida da informação condicional média gerada por pontos divergentes numa trajetória no espaço de fase (Arias-Londono *et al.*, 2011). Considerando valores fixos para m e ϵ , e considerando os vetores dados na Equação 1, a entropia aproximada é definida como (Pincus, 1991):

$$ApEn(m, \epsilon) = \lim_{N \rightarrow \infty} \left[\phi^m(\epsilon) - \phi^{m+1}(\epsilon) \right] \quad (8)$$

em que,

$$\phi^m(\epsilon) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log q(\epsilon) \quad (9)$$

e $q(\epsilon)$ é definida pela Equação 3 e N é o número de vetores do atrator reconstruído em uma dimensão de imersão mínima m .

A capacidade de discriminar mudanças de complexidade a partir de uma quantidade relativamente pequena de dados facultada a aplicação da $ApEn$ em uma variedade de contextos, incluindo processos determinísticos, caóticos e estocásticos (Pincus, 1991). Embora a medida da entropia aproximada não seja uma medida de caos, ela quantifica a regularidade do processo de imersão em séries temporais (Torres *et al.*, 2003).

Entropia de Tsallis

A entropia de Tsallis é uma generalização da entropia de Boltzmann-Gibbs (Chakrabarti e De, 2000). É definida como (Tsallis, 1988):

$$H_q = k \frac{1 - \sum_{i=1}^W p_i^q}{q-1} \quad (10)$$

em que $q \in \mathbb{R}$ caracteriza a generalização e $\{p_i\}$ são as probabilidades associadas com as W ($W \in \mathbb{N}$) configurações possíveis do sistema. A entropia de Boltzmann-Gibbs é encontrada quando, no limite, $q \rightarrow 1$.

A principal característica da entropia de Tsallis é a sua não aditividade (Tavares, 2003). Isto é, a informação contida em um sistema composto não é a simples soma da informação dos subsistemas que o compõem. As entropias de Shannon e de Boltzmann apresentam a aditividade como uma de suas características. A ideia principal da utilização da entropia de Tsallis é verificar a contribuição do parâmetro entrópico (q) na avaliação de características não extensivas (interações não lineares) presentes na dinâmica das séries temporais (Silva, 2010).

Expoentes de Lyapunov

Os expoentes de Lyapunov estão relacionados à taxa média de divergência ou de convergência exponencial de órbitas vizinhas no espaço de fase. Geralmente, um sistema contendo pelo menos um expoente de Lyapunov positivo é definido como caótico, por outro lado, um sistema sem nenhum expoente positivo é regular (Jiang *et al.*, 2006).

Neste trabalho é utilizada uma rotina do pacote TISEAN (Hegger *et al.*, 1999) a fim de se obter o crescimento exponencial médio das distâncias entre duas trajetórias, que é baseada no algoritmo de Kantz (1994). O programa basicamente gera os resultados da expressão (Kantz e Schreiber, 2004):

$$S(\Delta n) = \frac{1}{N} \sum_{n_0=1}^N \ln \left(\frac{1}{\nu(S_{n_0})} \sum_{S_n \in \nu(S_{n_0})} |S_{n_0+\Delta n} - S_{n+\Delta n}| \right) \quad (11)$$

onde, $S(\Delta n)$ é chamado de fator de crescimento (*stretching factor*), S_{n_0} são os vetores de imersão e $\nu(S_{n_0})$ é a vizinhança de S_{n_0} com diâmetro ϵ . Na Figura 6 está ilustrado o comportamento de $S(\Delta n)$ em função do número de iterações Δn , para três diferentes valores de ϵ e m variando de 4 a 8. A linha reta representa a inclinação média das curvas mostradas e sua inclinação é a estimativa para o maior expoente de Lyapunov, λ_1 .

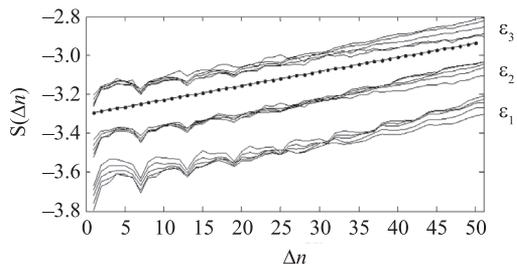


Figura 6. Comportamento de $S(\Delta n)$ em função do número de iterações, de ϵ e de m ($m = 4 \dots 8$).

Figure 6. Behavior of $S(\Delta n)$ related to iteration number, ϵ and m ($m = 4 \dots 8$).

Expoente de Hurst

O parâmetro ou expoente de Hurst (Hurst, 1951) é uma medida de autossimilaridade que vem sendo empregada em diversas áreas, tais como: gerenciamento de tráfego autossimilar de redes de computadores (Gomes *et al.*, 2001), reconhecimento de padrões sonoros para avaliação de sons pulmonares (Custódio, 1999), sistemas de reconhecimento de locutor (Sant'Ana *et al.*, 2006), detecção de diferentes áreas do cérebro (Rodríguez-Sánchez *et al.*, 2008) e análise de sinais EEG (eletroencefalograma) (Natarajan *et al.*, 2004), entre outros. Em processamento de sinais de voz, este parâmetro tem sido empregado por representar o comportamento estocástico da voz (Fantinato, 2008). Além disso, apresenta um baixo custo computacional, podendo ser obtido em tempo real, durante atividade da fala (Sant'Ana *et al.*, 2006).

Quanto maior o valor do parâmetro de Hurst, maior é o impacto das autocorrelações no desempenho do sistema. Além disso, pequenas variações no parâmetro de Hurst podem significar mudanças significativas no comportamento dos sinais, caracterizando o grau de desordem do mesmo (Gomes *et al.*, 2001). O estimador R/S, empregado neste trabalho, é um dos mais conhecidos e simples métodos para estimar o grau de dependência de uma sequência amostral. A estatística R/S é a soma parcial dos desvios, em relação a sua média, de uma sequência reescalada pelo seu desvio padrão (Alvarez-Ramirez *et al.*, 2008).

Se o processo estocástico associado à sequência é escalonada sobre certo domínio n , a estatística R/S segue a equação:

$$R/S = an^H \quad (12)$$

em que R/S é a distância reescalada (dividida pelo desvio padrão), a é uma constante e H é o expoente de Hurst.

Análise preditiva linear

A ideia básica na predição linear é a de que o valor de uma amostra pode ser aproximado (predita), por

combinação linear dos valores das amostras anteriores, levando em consideração a correlação entre estas. Os coeficientes de predição linear ou coeficientes LPC (*Linear Predictive Coding*) são estimados por minimização do erro quadrático entre a amostra atual e a sua predição (O'Shaughnessy, 2000; Rabiner e Juang, 1993).

O modelo linear de produção da fala incorpora os efeitos dos pulsos glotais, trato vocal e da radiação dos lábios como um filtro linear (Rabiner e Schafer, 1978). A fonte é uma sequência de impulsos quase periódicos utilizados para gerar sons sonoros ou a adição de uma sequência de ruído aleatório para sons surdos.

Um fator de ganho, G , é ajustado para controlar a intensidade da excitação. Combinando os efeitos dos pulsos glotais do trato vocal e da radiação, o modelo pode ser representado por uma função de transferência de apenas polos, $H(z)$, como:

$$H(z) = \frac{G}{1 - \sum_{k=1}^p \alpha(k)z^{-k}} \quad (13)$$

Os termos $\alpha(k)$ representam os coeficientes de predição linear (coeficientes LPC) e p a ordem do filtro de predição.

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. A análise LPC fornece um conjunto de parâmetros da fala que representa o trato vocal. Espera-se que, qualquer mudança na estrutura anatômica do trato vocal, devido à patologia, afete os coeficientes LPC.

Classificação

Após a extração das características, tanto pela análise linear quanto pela não linear, é realizado o processo de classificação. Neste processo, é aplicada a análise discriminante quadrática (QDA), com validação cruzada, em que são utilizados 10 subconjuntos (*folds*). São avaliados os desempenhos de classificadores implementados com cada uma das características, de forma individual, e de classificadores implementados a partir da combinação dessas características.

Outros classificadores são implementados com coeficientes LPC de ordem p , variando de 12 a 24 ($p = 12, 16, 20$ e 24). Após a avaliação individual de cada método (linear e não linear), são construídos vetores híbridos, em que a cada vetor LPC de ordem p , são adicionadas as medidas não lineares. Uma nova avaliação de desempenho é realizada a partir dos vetores híbridos.

Para avaliação de desempenho do processo de classificação, três medidas são utilizadas: acurácia, sensibilidade e especificidade. A Acurácia mede a

taxa de classificação correta global do teste, ou seja, a capacidade do teste de identificar corretamente quando há e quando não há presença da doença. A Acurácia (Equação 14) é definida como a relação entre todos os sinais corretamente classificados e o número total de sinais avaliados

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (14)$$

em que VP representa os verdadeiros positivos, ou seja, o número de sinais classificados corretamente como patológicos; VN representa os verdadeiros negativos, ou seja, em quantos sinais foram detectados a ausência correta da patologia; FP , os falsos positivos, ou a quantidade de sinais saudáveis classificados erroneamente como patológicos e FN os falsos negativos, ou seja, a quantidade de sinais patológicos classificados erroneamente como saudáveis.

A Sensibilidade mede a capacidade do teste em identificar corretamente a doença entre aqueles que a possuem, ou seja, o quão sensível é o teste. É a relação entre os sinais patológicos corretamente classificados e o número total de sinais patológicos, dada por:

$$Sensibilidade = \frac{VP}{VP + FN} \quad (15)$$

A Especificidade mede a capacidade do teste em excluir corretamente aqueles que não possuem a doença, ou seja, o quão específico é o teste. É a relação entre os sinais saudáveis corretamente classificados e o número total de sinais saudáveis:

$$Especificidade = \frac{VN}{VN + FP} \quad (16)$$

Os resultados obtidos no processo de classificação são apresentados a seguir.

Resultados

Na Tabela 1 são apresentados os valores de acurácia obtidos com o uso de cada uma das medidas da análise dinâmica não linear (MNL), para a classificação entre vozes saudáveis (SDL) e patológicas (PTL), usando a análise discriminante quadrática (QDA – *Quadratic Discriminant Analysis*). A classe dos sinais patológicos inclui os sinais afetados por edema, nódulos e paralisia nas pregas vocais. O expoente de Hurst, H , foi o parâmetro mais eficiente em capturar as desordens vocais presentes nos sinais de vozes patológicas, comparadas aos sinais de vozes saudáveis, tendo a maior acurácia no processo de classificação.

Na Tabela 2 são apresentados, inicialmente, os resultados obtidos na classificação entre vozes saudáveis e patológicas (SDL \times PTL), mediante a

combinação das medidas. Observa-se que, com a combinação de apenas quatro medidas, entre as oito, foi obtida a maior acurácia ($96,03 \pm 4,11$), com as medidas entropia de Shannon (H_1), entropia de Tsallis (H_q), expoente de Hurst (H) e a dimensão de correlação (D_2). O acréscimo das três medidas ao expoente de Hurst proporcionou um aumento de cerca de 9% no valor da acurácia, alcançada com o uso individual desta característica.

As patologias são separadas em classes diferentes (paralisia, edema ou nódulo) e comparadas aos sinais saudáveis, constituindo as seguintes comparações: saudável \times edema (SDL \times EDM), saudável \times paralisia (SDL \times PRL) e saudável \times nódulo (SDL \times NDL). Como apresentado na Tabela 2, nesses casos, as taxas de acurácia média são superiores a 97%, com todas as taxas de especificidade chegando a 100%, o que significa que nenhum sinal saudável testado foi considerado patológico. O expoente de Hurst e as entropias de Shannon e de Tsallis aparecem em todas as combinações, associando-se a outras medidas no caso da discriminação entre saudável e edema e entre saudável e paralisia.

Na distinção entre patologias, observa-se pela Tabela 2, mesmo com a combinação das medidas,

a acurácia média atingiu no máximo 86%, enquanto que os valores de sensibilidade média não chegaram a 75%. Dessa forma, visando aumentar o desempenho na classificação, foi investigada a influência do uso combinado das medidas da análise não linear com os coeficientes LPC na discriminação entre as patologias consideradas. Nas Figuras 7 a 9 são mostrados os resultados obtidos na discriminação entre paralisia e edema, paralisia e nódulo e entre edema e nódulo, respectivamente.

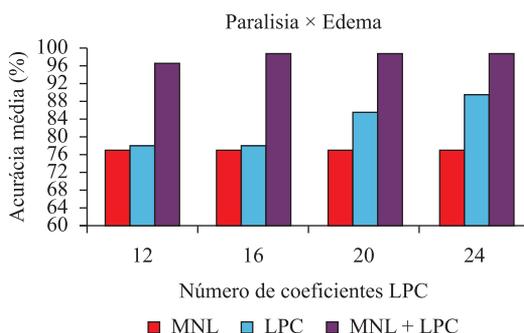


Figura 7. Classificação entre vozes afetadas por paralisia e sinais afetados por edema nas pregas vocais.

Figure 7. Classification between voices affected by vocal fold paralysis and vocal fold edema.

Tabela 1. Avaliação de desempenho das medidas individuais – valores de Acurácia (%) obtidos na classificação entre vozes saudáveis e vozes patológicas.

Table 1. Performance evaluation of individual measures – values of the obtained accuracy in classification between healthy and pathological voices.

Medidas	Saudável \times Patológica	Saudável \times Edema	Saudável \times Paralisia	Saudável \times Nódulo
λ_1	54,51 \pm 8,23	61,22 \pm 11,05	58,55 \pm 9,99	52,32 \pm 13,08
FMMI	62,55 \pm 5,74	60,22 \pm 8,48	68,64 \pm 10,10	58,21 \pm 14,46
H_1	60,85 \pm 8,31	69,44 \pm 8,82	65,82 \pm 6,07	74,64 \pm 8,03
H_q	62,61 \pm 9,21	67,44 \pm 7,18	68,55 \pm 7,66	76,07 \pm 6,82
H	87,09 \pm 5,91*	85,67 \pm 5,13*	88,09 \pm 7,53*	94,29 \pm 5,28*
ApEn	56,60 \pm 6,72	68,06 \pm 6,15	69,29 \pm 13,51	66,67 \pm 10,37
D_2	79,36 \pm 10,77	84,58 \pm 8,87	87,50 \pm 8,09	82,62 \pm 12,54
K_2	52,88 \pm 5,09	60,97 \pm 10,32	61,96 \pm 11,98	57,38 \pm 8,81

*Acurácia máxima.

*Maximum accuracy.

Tabela 2. Avaliação de desempenho das medidas combinadas – valores de acurácia obtidos na classificação entre vozes saudáveis e patologias e entre as patologias.

Table 2. Performance evaluation of combined measures - values of the obtained accuracy in classification between healthy and pathological voices and among pathologies.

Classificação	Características	Acurácia (%)	Sensibilidade (%)	Especificidade (%)
SDL \times PTL	H_1, H_q, H e D_2	96,03 \pm 4,11	94,46 \pm 5,13	98,00 \pm 4,52
SDL \times EDM	H_1, H_q, H, D_2 e K_2	98,75 \pm 2,83	96,67 \pm 7,54	100
SDL \times PRL	FMMI, H, D_2 e H_1 ou H_q	97,50 \pm 3,77	93,33 \pm 10,05	100
SDL \times NDL	H_1, H_q e H	98,57 \pm 3,23	95,00 \pm 11,31	100
PRL \times EDM	$\lambda_1, FMMI, H_q, ApEn, D_2$ e K_2	77,00 \pm 14,59	74,17 \pm 18,74	81,67 \pm 18,17
PRL \times NDL	$\lambda_1, H, ApEn, D_2$ e K_2	86,00 \pm 11,03	65,00 \pm 29,66	96,67 \pm 14,94
EDM \times NDL	$\lambda_1, FMMI, H_1, H_q, H$ e D_2	76,00 \pm 9,51	55,00 \pm 26,39	86,67 \pm 12,31

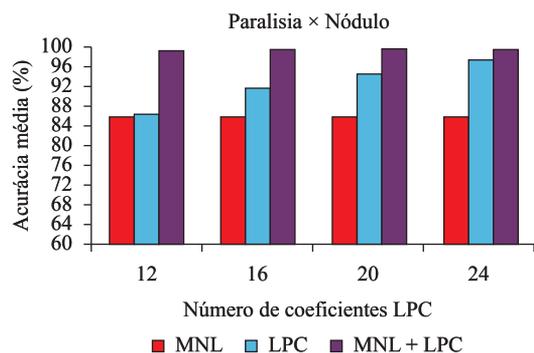


Figura 8. Classificação entre vozes afetadas por paralisia e sinais afetados por nódulos nas pregas vocais.

Figure 8. Classification between voices affected by vocal fold paralysis and vocal nodules.

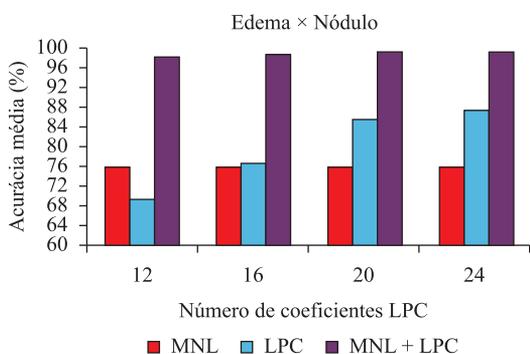


Figura 9. Classificação entre vozes afetadas por edema e sinais afetados por nódulos nas pregas vocais.

Figure 9. Classification between voices affected by vocal fold Edema and vocal nodules.

Na discriminação entre paralisia e edema, a taxa de acurácia sai de $96,55\% \pm 1,51\%$, com 12 coeficientes LPC para $99,24\% \pm 0,52\%$, com 24 coeficientes. Além do acréscimo na acurácia média em mais de 18%, em relação ao uso dos melhores resultados individuais (LPC), com a combinação (MNL e LPC), o erro padrão caiu consideravelmente, aumentando a confiabilidade do teste.

Na discriminação entre paralisia e nódulo, a acurácia de 86% só com as medidas não lineares, cresce para valores maiores que 99%. Com 12 coeficientes LPC, a acurácia é de $99,19\% \pm 0,55\%$, com sensibilidade 100% e especificidade média de aproximadamente 98%.

A acurácia média na classificação entre edema e nódulo, antes de 76% com as medidas não lineares e cerca de 70% com LPC, chega a mais de 98% com o vetor híbrido, com apenas 12 coeficientes. E, passa dos 99% a partir de 16 coeficientes, quando associados às medidas não lineares.

Discussão

Pelos resultados apresentados, constata-se que o desempenho da classificação entre vozes saudáveis e patológicas aumentou significativamente com a combinação das medidas da análise dinâmica não linear. Nos melhores resultados, o expoente de Hurst, que se destacou como melhor medida individual, aparece em todas as combinações, juntamente com as entropias de Shannon e de Tsallis, para a classificação entre os sinais de vozes saudáveis e sinais afetados por paralisia, edema ou nódulo. No entanto, essas medidas não foram eficazes em discriminar entre patologias, apresentando taxas inferiores às apresentadas na discriminação entre vozes saudáveis e patológicas.

A análise LPC, comparada às outras medidas empregadas, mostra-se mais eficaz em discriminar entre patologias. Observa-se que o valor da acurácia aumenta com o aumento da ordem de preditor. Além disso, o desempenho na classificação entre patologias melhora significativamente ao combinar as medidas não lineares com os coeficientes LPC.

É importante considerar que o aumento na ordem dos coeficientes LPC está associado a um maior volume de dados a serem processados, aumentando o custo computacional. No entanto, caso seja mais relevante a precisão do diagnóstico do que a realização da análise em tempo real, é possível optar por empregar uma maior quantidade de coeficientes LPC. Observa-se pelos resultados, que o valor da acurácia não teve um aumento tão considerável com o aumento da ordem de predição além de 16 coeficientes.

Conclusão

A partir dos resultados obtidos, constata-se que as medidas não lineares conseguem capturar bem as desordens vocais provocadas por patologias nas pregas vocais, como nódulos, edemas e paralisia. A consideração das não linearidades inerentes ao processo de produção da fala e sua consequente análise utilizando medidas não lineares é eficiente em distinguir um sinal saudável de um sinal patológico, comprovadamente com as patologias consideradas no trabalho (edema, paralisia e nódulos nas pregas vocais). Entretanto, observa-se que a abordagem pelo modelo linear de produção da fala, em que os coeficientes LPC modelam o trato vocal, consegue distinguir melhor as especificidades existentes nas patologias analisadas, notadamente quando combinadas à abordagem não linear. As mudanças nas estruturas anatômicas do trato vocal e da fonte (laringe) foram bem retratadas com a combinação das duas abordagens, sendo a principal contribuição deste trabalho. Com a abordagem híbrida, o desempenho na distinção entre

as patologias obteve um aumento considerável. Isto denota que o método empregado é bastante promissor na tarefa de discriminar entre vozes saudáveis e vozes afetadas por patologias laringeas, bem como entre patologias. Espera-se que os métodos apresentados nesta pesquisa contribuam ainda mais para a detecção precoce das doenças da laringe, auxiliando no seu diagnóstico ou para monitorar seu progresso, buscando assegurar os cuidados clínicos adequados e mais conforto aos pacientes.

Referências

- Alvarez-Ramirez J, Echeverria JC, Rodriguez E. Performance of a high-dimensional R/S method for Hurst exponent estimation. *Physica A: Statistical Mechanics and its Applications*. 2008; 387(26):6452-62. <http://dx.doi.org/10.1016/j.physa.2008.08.014>
- Arias-Londono J, Godino-Llorente J, Sáenz-Lechón N, Osma-Ruiz V, Castellanos-Dominguez G. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *IEEE Transactions on Biomedical Engineering*. 2011; 58(2):370-9. <http://dx.doi.org/10.1109/TBME.2010.2089052>
- Boyanov B, Ivanov T, Hadjitodorov S, Chollet G. Robust hybrid pitch detector. *Electronic Letters*. 1993; 29(22):1924-6. <http://dx.doi.org/10.1049/el:19931281>
- Chakrabarti CG, De K. Boltzmann-Gibbs entropy: axiomatic characterization and application. *International Journal of Mathematics and Mathematical Sciences*. 2000; 23(4):243-51. <http://dx.doi.org/10.1155/S0161171200000375>
- Custódio RF. Análise Não-linear no reconhecimento de padrões sonoros: Estudo de caso para sons pulmonares [tese]. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul; 1999. 119 p.
- Dajer ME. Padrões visuais de sinais de voz através de técnica de análise de não-linear [dissertação]. São Paulo: Escola de Engenharia de São Carlos; 2006. 154 p.
- Fantinato PC. Segmentação de voz baseada na análise fractal e na transformada wavelet [dissertação]. São Paulo: Instituto de Física de São Carlos da Universidade de São Carlos; 2008. 123 p.
- Fiedler-Ferrara N, Prado CPC. *Caos: Uma introdução*. São Paulo: Ed. Edgard Blucher Ltda, 1994.
- Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. *Physical Review A*. 1986; 33:1134-40. <http://dx.doi.org/10.1103/PhysRevA.33.1134>
- Godino-Llorente JI, Gómez-Vilda P, Blanco Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*. 2006; 53(10):1943-53. <http://dx.doi.org/10.1109/TBME.2006.871883>
- Gomes D, Fonseca N, Agoulmine N, Souza JN. Neurocomputação na estimativa do parâmetro de hurst. In: SBRC'2001: Anais do 19 Simpósio Brasileiro de Redes de Computadores; 2001 Maio; Florianópolis, Santa Catarina. Florianópolis; 2001. p. 306-21.
- Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. *Physica D*. 1983; 9:189-208. [http://dx.doi.org/10.1016/0167-2789\(83\)90298-1](http://dx.doi.org/10.1016/0167-2789(83)90298-1)
- Hegger R, Kantz H, Schreiber T. Practical implementation of nonlinear time series methods: the tisean package. *Chaos*. 1999; 9(2):413-35. <http://dx.doi.org/10.1063/1.166424>
- Henríquez P, Alonso JB, Ferrer MA, Travieso CM, Godino-Llorente JI, Díaz-De-Maria F. Characterization of healthy and pathological voice through measures based on nonlinear dynamics. *IEEE Transactions on Audio, Speech, and Language Processing*. 2009; 17(6):1186-95. <http://dx.doi.org/10.1109/TASL.2009.2016734>
- Hurst E. Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*. 1951; 116:770-9.
- Jiang JJ, Zhang Y, McGilligan C. Chaos in voice, from modeling to measurement. *Journal of Voice*. 2006; 20(1):2-17. <http://dx.doi.org/10.1016/j.jvoice.2005.01.001>
- Johnson MT, Povinelli RJ, Lindgren AC, Ye J, Liu X, Indrebo KM. Time-domain isolated phoneme classification using reconstructed phase spaces. *IEEE Transactions Speech and Audio Processing*. 2005; 13(4):458-66. <http://dx.doi.org/10.1109/TSA.2005.848885>
- Kantz H. A Robust Method to Estimate the maximal lyapunov exponent of a time series. *Physics Letters A*. 1994; 185:77-87. [http://dx.doi.org/10.1016/0375-9601\(94\)90991-1](http://dx.doi.org/10.1016/0375-9601(94)90991-1)
- Kantz H, Schreiber T. *Nonlinear time series analysis*. 2nd ed. Cambridge: University Press; 2004.
- Kay *Elementrics*. *Disordered Voice Database*. Model 4337. 3th ed. United States; 1994.
- Kokkinos I, Maragos P. Nonlinear speech analysis using models for chaotic systems. *IEEE Transactions on Speech and Audio Processing*. 2005; 13(6):1098-109. <http://dx.doi.org/10.1109/TSA.2005.852982>
- Kumar A, Mullick SK. Nonlinear dynamical analysis of speech. *Journal Acoustic Society America*. 1996; 100(1):615-29. <http://dx.doi.org/10.1121/1.415886>
- Little M. *Biomechanically informed nonlinear speech signal processing [tese]*. Oxford: University of Oxford, England; 2006. 164 p.
- Little M, McSharry P, Roberts S, Costello D, Moroz I. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine*. 2007; 6(23):1-19. <http://dx.doi.org/10.1186/1475-925X-6-23>
- Manfredi C, Pierazzi L, Brusaglioni P. Pitch estimation for noise retrieval in time and frequency domain. *Medical & Biological Engineering & Computing*. 1999; 37(2):532.
- Natarajan K, Acharya UR, Alias F, Tiboleng T, Puthusserypadu SK. Nonlinear analysis of EEG signals at different mental states. *BioMedical Engineering OnLine*. 2004; 3:7. <http://dx.doi.org/10.1186/1475-925X-3-7>

- O'Shaughnessy D. Speech communications: human and machine. 2nd ed. New York: IEEE Press; 2000.
- Petry A. Reconhecimento automático de locutor utilizando medidas de invariantes dinâmicas não-lineares [tese]. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul; 2002. 155 p.
- Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences. 1991; 88:2297-301. <http://dx.doi.org/10.1073/pnas.88.6.2297>
- Rabiner LR, Schafer RW. Digital processing of speech signals. Upper Saddle River: Prentice Hall; 1978.
- Rabiner LR, Juang BH. Fundamentals of speech recognition. Englewood Cliffs. New Jersey: Prentice Hall, 1993.
- Rodríguez-Sánchez A, Delgado-Trejos E, Orozco-Gutiérrez A, Castellanos-Domínguez G. Nonlinear dynamics techniques for the detection of the brain areas using MER signals. International Conference on BioMedical Engineering and Informatics. 2008; 2:198-202. <http://dx.doi.org/10.1109/BMEI.2008.330>
- Sant'Ana R, Coelho R, Alcaim A. Text-independent speaker recognition based on the hurst parameter and the multidimensional fractional brownian motion model. IEEE Transactions on Audio, Speech and Language Processing. 2006; 14(3):931-40. <http://dx.doi.org/10.1109/TSA.2005.858054>
- Savi MA. Dinâmica não-linear e Caos. Rio de Janeiro: E-papers; 2006.
- Scalassara PR, Dajer ME, Maciel CD, Pereira JC. Voice signals characterization through entropy measures. In: BIOSIGNALS 2008: Proceedings of the International Conference on Bio-inspired Systems and Signal Processing; 2008 Jan 28-31; Madeira, Portugal. Madeira; 2008. v. 2, p. 163-170.
- Schreiber T, Schmitz A. Surrogate time series. Physica D. 2000; 142:346-82. [http://dx.doi.org/10.1016/S0167-2789\(00\)00043-9](http://dx.doi.org/10.1016/S0167-2789(00)00043-9)
- Shannon C. A Mathematical theory of communication. The Bell System Technical Journal. 1948; 27(3):623-56.
- Silva LEV. Ferramentas computacionais na análise da variabilidade da frequência cardíaca através do paradigma não extensivo no estudo de cardiopatias [dissertação]. São Paulo: Universidade de São Paulo; 2010. 97 p.
- Takens F. Detecting Strange Attractors in Turbulence. In: Rand DA, Young L-S. Dynamical systems and turbulence. New York: Springer-Verlag; 1981. v. 898, p. 366-81.
- Tavares AHMP. Aspectos matemáticos da entropia [dissertação]. Portugal: Universidade de Aveiro; 2003. 98 p.
- Tokuda I, Miyano T, Aihara K. Surrogate analysis for detecting nonlinear dynamics in normal vowels. The Journal of the Acoustical Society of America. 2001; 110(6):3207-17.
- Torres ME, Gamero LG, Rufiner HL. Study of complexity in normal and pathological speech signals. In: EMBC 2003: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003 17-21 Sept., Cancun, México. IEEE; 2003. p. 2339-2342. <http://dx.doi.org/10.1109/IEMBS.2003.1280385>
- Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics. 1988; 52(1/2):479-87. <http://dx.doi.org/10.1007/BF01016429>
- Zhang Y, Jiang JJ. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. Journal of Voice. January 2008; 22(1):1-9. <http://dx.doi.org/10.1016/j.jvoice.2006.08.003>

Autores

Washington César de Almeida Costa*, Francisco Marcos de Assis

Departamento de Engenharia Elétrica – DEE, Universidade Federal de Campina Grande – UFCG, Campina Grande, PB, Brasil.

Washington César de Almeida Costa*, Silvana Luciene do Nascimento Cunha Costa

Coordenação do Curso Superior de Bacharelado em Engenharia Elétrica – CCSBEE, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB, Av. Primeiro de Maio, 720, Jaguaribe, CEP 58015-430, João Pessoa, PB, Brasil.

Benedito Guimarães Aguiar Neto

Universidade Presbiteriana Mackenzie, CEP 01302-907, São Paulo, SP, Brasil.