

Análise de Sinais de Voz Baseada em Medidas de Quantificação de Recorrência

Washington C. de A. Costa, Silvana C. Costa, Vinícius J. D. Vieira, Benedito G. Aguiar Neto e F. M. de Assis

Resumo—Este artigo apresenta uma análise de sinais de voz baseada em medidas de quantificação obtidas de seus gráficos de recorrência. São avaliados três grupos de sinais: vozes saudáveis, vozes com paralisia e vozes com edema nas dobras vocais. O objetivo deste trabalho é caracterizar os grupos de sinais de acordo com as medidas de quantificação de recorrência, a fim de potencializar o desenvolvimento de um sistema que discrimine os diferentes tipos de sinais. São abordadas quatro medidas: determinismo, entropia, taxa de recorrência e laminaridade.

Palavras-Chave—Processamento digital de sinais de voz, gráficos de recorrência, medidas de quantificação.

Abstract—This article presents a speech signal analysis based on quantification measures obtained from their recurrence plots. Three signal groups are evaluated: healthy voices and voices affected by vocal fold paralysis and edema. This work aims to characterize these signals by their quantification recurrence measurements to enable the development of a system capable of discriminating among the different signal types. Four measures are explored: determinism, entropy, recurrence rate and laminarity.

Keywords—Digital speech processing, recurrence plots, quantification measurements.

I. INTRODUÇÃO

Nas últimas décadas os métodos clássicos de análise de dados baseados em um modelo linear têm sido enriquecidos com novos métodos que são derivados da teoria do caos. Vários métodos de análise tem tentado estimar medidas e propriedades não-lineares de processos naturais. Entretanto, a maioria dos métodos da análise não-linear exigem séries um tanto longas ou estacionaridade [1].

Recentemente, outro método baseado na análise não-linear de dados tem sido usado: os gráficos de recorrência proposto por Eckmann et al [2]. Recorrência é uma propriedade fundamental dos sistemas dinâmicos dissipativos. A principal vantagem dos gráficos de recorrência é a possibilidade do uso dessa técnica para análise de séries curtas e não-estacionárias [1], [3].

A maioria dos esquemas de produção da fala está baseada na hipótese na qual o sinal de voz é modelado como sendo a saída de um filtro linear variante no tempo excitado por uma dada fonte. Tal hipótese carrega intrinsecamente suas limitações, o que leva a busca de estudos com o objetivo de se

incorporar modelos não-lineares para representar a produção da fala [4]. O estudo da complexidade dos sinais de voz tem levado em consideração os aspectos não-lineares da voz e o comportamento dinâmico das dobras vocais na produção da fala. Dessa forma, o estudo de modelos não-lineares para representar a produção da voz tem se destacado em recentes pesquisas [5].

Sabe-se que um sinal de voz é caracterizado por sua não-estacionaridade a longo intervalo de tempo (maior que 32 ms) e algumas técnicas de análise dinâmica não-linear a partir das séries temporais exigem séries de comprimentos longos (tipicamente 500 amostras ou mais). A técnica de análise baseada nos gráficos de recorrência é utilizada, neste trabalho, com o objetivo de caracterizar sinais de vozes saudáveis e vozes afetadas por patologias na laringe, analisadas a longo intervalo de tempo (400 ms).

A análise acústica de sinais de vozes afetadas por patologias laringeas possibilita uma ferramenta auxiliar ao diagnóstico médico e no acompanhamento de tratamentos de forma objetiva e não-invasiva, quando comparados a exames de videolaringoscopia, frequentemente empregados. Além disso, com o crescente avanço nas redes de telemedicina, esta técnica torna possível o acompanhamento à distância da evolução de tratamentos farmacológicos por meio da análise vocal, bem como do auxílio a diagnósticos [6].

Neste trabalho, o comportamento do sinal de voz é avaliado a partir de quatro medidas de quantificação obtidas dos gráficos de recorrência: determinismo, entropia, taxa de recorrência e laminaridade. Uma análise estatística, baseada em testes de hipótese, é realizada para investigar sobre uma possível diferenciação entre os sinais de vozes saudáveis e os sinais com patologias (paralisia ou edema) nas dobras vocais.

II. GRÁFICOS DE RECORRÊNCIA E MEDIDAS DE QUANTIFICAÇÃO

A ideia de recorrência surgiu no século XIX, a partir do Teorema da Recorrência, do físico e filósofo francês J. H. Poincaré (1890). Os Gráficos de Recorrência (ou *Recurrence Plots - RP*) foram criados como método de análise dos sistemas dinâmicos por Eckmann et al. [2].

O primeiro passo para a construção do gráfico de recorrência de uma série temporal de comprimento L é a imersão dessa série em um espaço m -dimensional, em geral, por meio do método dos atrasos temporais [7]. Nesse método são criadas $m - 1$ cópias defasadas de $\tau, 2\tau, \dots, (m - 1)\tau$ amostras, a partir de cada sinal de voz disponível. Daí são construídos N vetores, de dimensão m , representativos da dinâmica do sistema de produção da voz no espaço de fase.

Washington C. de A. Costa, Silvana C. Costa e Vinícius J. Dias Vieira, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, Brasil, F. M. de Assis, Universidade Federal de Campina Grande, Campina Grande, Brasil, Benedito G. Aguiar Neto, Universidade Mackenzie, São Paulo, Brasil. E-mails: washington@ifpb.edu.br, silvana@ifpb.edu.br, viniciusjdv@gmail.com, benedito.aguiar@mackenzie.br, fmarcos@dee.ufcg.edu.br. Este trabalho foi parcialmente financiado pelo CNPq.

De acordo com a definição originalmente dada em [2] o gráfico de recorrência é uma matriz $N \times N$ (matriz gráfica), onde N é o número de vetores (estados do sistema) de dimensão m , preenchida por pontos brancos e pretos. O ponto preto, chamado de ponto recorrente, é colocado na matriz de recorrência com coordenadas (i, j) somente se a distância $d(i, j)$ nos instantes $n = i$ e $n = j$ (entre o estado corrente do sistema e o estado a ser comparado) for menor que certa distância (raio) ε , fixada no centro do estado corrente.

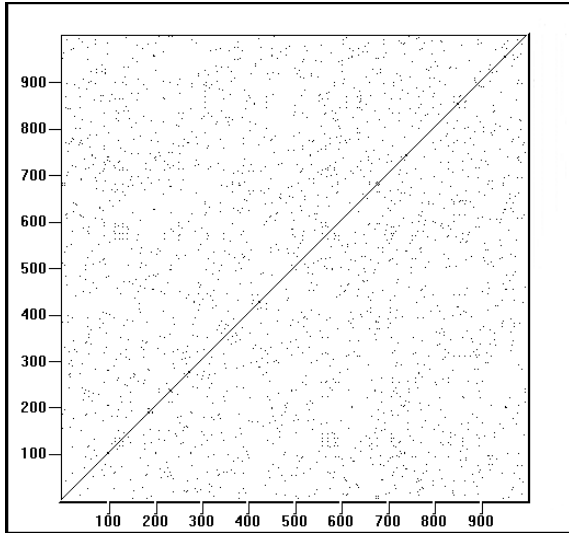


Fig. 1. Gráfico de recorrência para o ruído branco, $L = 1.000$, $\tau = 1$ e $m = 3$.

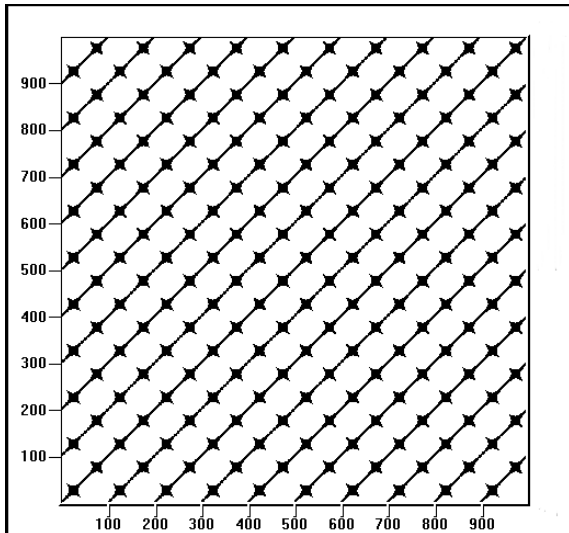


Fig. 2. Gráfico de recorrência para um sinal senoidal, $L = 1.000$, $\tau = 4$ e $m = 2$.

Uma definição alternativa comumente utilizada para o Gráfico de Recorrência é dada por [8]:

$$\mathcal{R}_{i,j}^{m,\varepsilon} = \Theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|), \quad \vec{x}_i \in R^m, \quad i, j = 1 \dots N. \quad (1)$$

Em que:

- N é o número de estados \vec{x}_i considerados;

- ε é o raio da vizinhança (*threshold*) no ponto \vec{x}_i ;
- $\|\cdot\|$ é a norma da vizinhança, comumente a norma euclidiana;
- $\Theta(\cdot)$ é a função de Heaviside;
- m é a dimensão de imersão (graus de liberdade).

Pontos isolados, linhas diagonais, como também linhas verticais e horizontais (e a combinação destas para formar estruturas variadas) são estruturas típicas presentes nos gráficos de recorrência [3].

A análise visual dos gráficos de recorrência fornece resultados de forma qualitativa. Como, por exemplo, processos com comportamento estocástico tendem a não apresentar estruturas diagonais (Figura 1). Por outro lado, processos determinísticos causam diagonais mais longas e menos pontos de recorrência isolados (Figura 2).

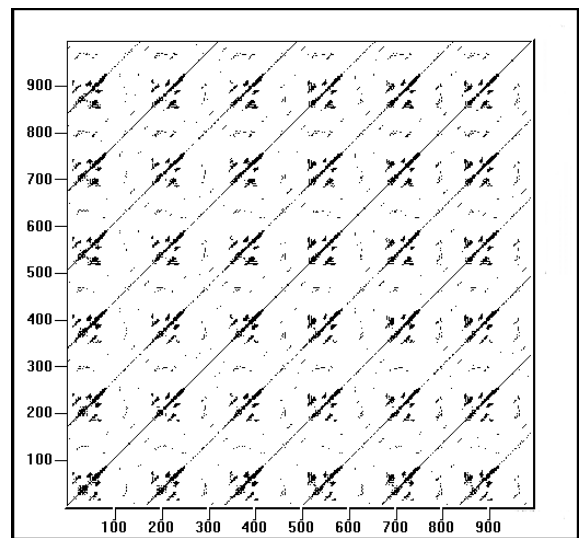


Fig. 3. Gráfico de recorrência da vogal sustentada /ah/ para um sinal de voz saudável, $L = 1.000$, $\tau = 4$ e $m = 3$.

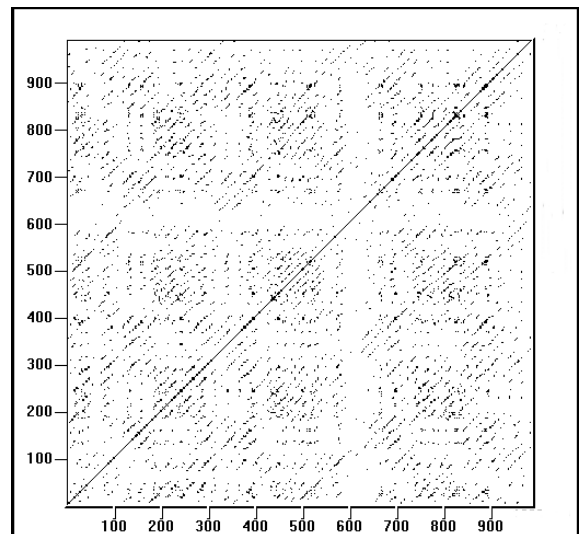


Fig. 4. Gráfico de recorrência da vogal sustentada /ah/ para um sinal de voz patológica (paralisia), $L = 1.000$, $\tau = 6$ e $m = 3$.

Nas Figuras 3 e 4, estão apresentados os gráficos de recorrência da vogal sustentada /ah/ para um sinal de voz saudável e para um sinal de voz afetado por paralisia nas dobras vocais, respectivamente. Como pode ser observado nestas representações, o gráfico de recorrência do sinal de voz saudável apresenta a formação de várias linhas diagonais evidenciando, desta forma, a sua natureza quase periódica. Por outro lado, a desordem vocal produzida devido à presença da paralisia nas dobras vocais apresenta uma quantidade maior de pontos isolados no gráfico de recorrência.

Objetivando quantificar as riquezas presentes na dinâmica de um sistema, Webber Jr. e Zbilut desenvolveram medidas de complexidade, as chamadas Medidas de Quantificação de Recorrência [9]. Neste trabalho são analisadas quatro dessas medidas, descritas a seguir.

Determinismo: Mede a quantidade dos pontos de recorrência presentes na formação das linhas diagonais em relação a todo o conjunto dos pontos de recorrência. Está relacionada com o grau de previsibilidade do sistema.

$$DET = \frac{\sum_{l=l_{min}}^N lP^\varepsilon(l)}{\sum_{i,j}^N \mathcal{R}_{i,j}^{m,\varepsilon}}, \quad (2)$$

em que $P^\varepsilon(l)$ representa a distribuição de frequência dos comprimentos l das estruturas diagonais no RP e l_{min} é o número mínimo de estruturas diagonais que se deseja contabilizar dentro do gráfico de recorrência.

Entropia: Refere-se à entropia de Shannon da distribuição de frequência dos comprimentos das linhas diagonais. Ela reflete a complexidade da estrutura determinística no sistema.

$$ENTR = - \sum_{l=l_{min}}^N p(l) \ln p(l), \quad (3)$$

em que $p(l) = \frac{P^\varepsilon(l)}{\sum_{l=l_{min}}^N P^\varepsilon(l)}$

Taxa de Recorrência: Quantifica a porcentagem de pontos de recorrência presentes dentro de um raio especificado.

$$REC = \frac{1}{N^2} \sum_{i,j=1}^N \mathcal{R}_{i,j;i \neq j}^{m,\varepsilon}. \quad (4)$$

Laminaridade: é uma razão entre os pontos de recorrência que formam as linhas verticais e todo o conjunto de pontos recorrentes. Representa a ocorrência de estados laminares sem, contudo, descrever o comprimento destas fases laminares.

$$LAM = \frac{\sum_{v=v_{min}}^N vP^\varepsilon(v)}{\sum_{i,j}^N \mathcal{R}_{i,j}^{m,\varepsilon}}, \quad (5)$$

em que v_{min} é o tamanho mínimo com que se deseja computar uma estrutura vertical.

III. METODOLOGIA

São avaliadas 145 locuções da vogal sustentada /ah/, sendo 53 sinais de vozes saudáveis, 49 sinais de vozes com paralisia e 43 sinais com edema nas dobras vocais. Os sinais são oriundos da base de dados *Disordered Voice Database*, Model 4337, da *Kay Elemetrics*, desenvolvida pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* [10].

Os sinais de vozes saudáveis, originalmente amostrados a uma frequência de 50 kHz, são sub-amostrados a 25 kHz para equiparar ao número de amostras dos sinais patológicos.

Para a obtenção dos parâmetros de imersão (passo de reconstrução τ e dimensão de imersão m) e dos gráficos de recorrência é utilizado o *software VRA (Visual Recurrence Analysis)* [11].

As medidas de quantificação são obtidas a partir do *software RQA (Recurrence Quantification Analysis)* [12] utilizando um único quadro de 10.000 amostras (entre 100 ms e 500 ms) de cada um dos sinais selecionados da base de dados. É considerando, ainda, um raio de vizinhança ε de 2% da máxima distância entre os estados de cada sinal analisado [13].

IV. RESULTADOS OBTIDOS

Nas Figuras 5, 6, 7 e 8 estão representados os histogramas para as medidas de recorrência consideradas: Determinismo, Entropia, Taxa de Recorrência e Laminaridade.

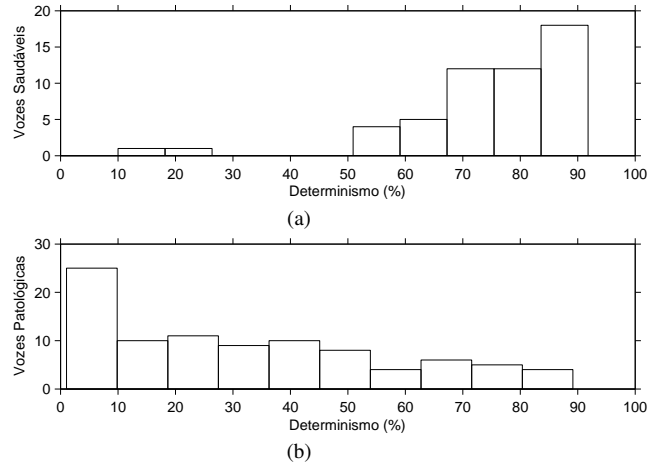


Fig. 5. Histogramas dos valores da medida do determinismo para (a) sinais de vozes saudáveis e (b) sinais de vozes patológicas.

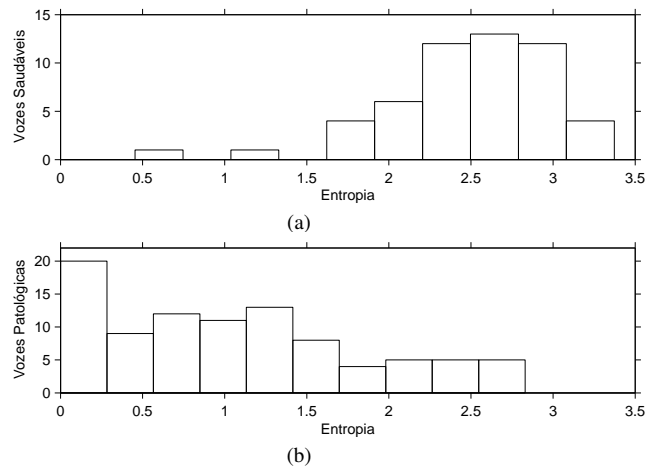


Fig. 6. Histogramas dos valores da medida da entropia de Shannon para (a) sinais de vozes saudáveis e (b) sinais de vozes patológicas.

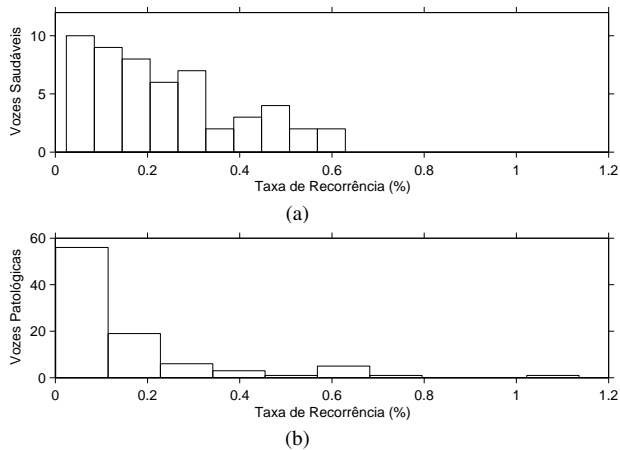


Fig. 7. Histogramas dos valores da medida da taxa de recorrência para (a) sinais de vozes saudáveis e (b) sinais de vozes patológicas.

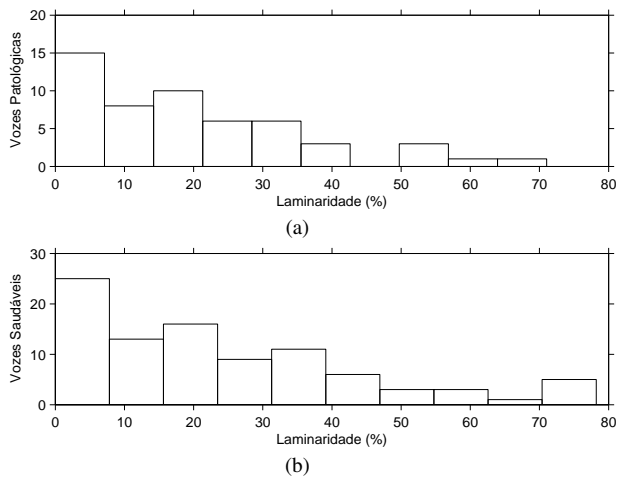


Fig. 8. Histogramas dos valores da medida de laminaridade para (a) sinais de vozes saudáveis e (b) sinais de vozes patológicas.

Na Tabela I estão apresentados os valores médios e respectivos valores do desvio padrão obtidos para as seguintes medidas: Determinismo (*DET*), Entropia (*ENTR*), Taxa de Recorrência (*REC*) e Laminaridade (*LAM*).

TABELA I

VALOR MÉDIO E DESVIO PADRÃO DAS MEDIDAS DE QUANTIFICAÇÃO DE RECORRÊNCIA PARA OS SINAIS DE VOZES SAUDÁVEIS E PATOLÓGICAS.

| Medida | Voz Saudável | | Voz Patológica | |
|----------------|--------------|----------|----------------|----------|
| | μ | σ | μ | σ |
| <i>DET</i> (%) | 75,17 | 15,60 | 31,72 | 25,59 |
| <i>ENTR</i> | 2,47 | 0,53 | 1,06 | 0,79 |
| <i>REC</i> (%) | 0,24 | 0,16 | 0,15 | 0,19 |
| <i>LAM</i> | 19,65 | 16,76 | 24,04 | 19,95 |

Nos dados apresentados na Tabela I, observa-se que os valores do desvio padrão são relativamente altos evidenciando a complexidade inerente ao sistema de produção da fala notadamente para os sinais de vozes afetadas por patologias.

A fim de verificar a existência de possíveis diferenças estatísticas entre os dois grupos de vozes considerados, fo-

ram realizados testes de hipóteses admitindo-se um nível de significância igual a 0,05 [14]. Após a realização do teste *t* para duas amostras independentes, os resultados encontrados revelaram que os valores de entropia apresentam diferenças estatisticamente significativas ($P < 0,001$) entre os grupos de vozes saudáveis e o de vozes patológicas (edema ou paralisia) considerados.

A aplicação do teste não-paramétrico de Mann-Whitney, para o caso de duas amostras independentes, apresentou evidências estatisticamente significativas ($P < 0,001$) nos valores das medidas de Determinismo e Taxa de Recorrência entre os dois grupos de vozes avaliados. Por outro lado, o mesmo teste revelou a ausência de diferenças estatisticamente significativas ($P = 0,175$) nos valores da medida de Laminaridade para os dois grupos.

Foi avaliado ainda o comportamento das medidas para os sinais de vozes afetadas por patologias específicas (paralisia ou edema). Na Tabela II estão apresentados os valores médios e os valores do desvio padrão para as medidas de quantificação dos dois diferentes tipos de vozes patológicas analisados.

TABELA II

VALORES MÉDIOS E DESVIO PADRÃO DAS MEDIDAS DE QUANTIFICAÇÃO PARA VOZES COM PARALISIA E VOZES COM EDEMA.

| Medida | Paralisia | | Edema | |
|----------------|-----------|----------|-------|----------|
| | μ | σ | μ | σ |
| <i>DET</i> (%) | 23,67 | 23,39 | 40,90 | 25,13 |
| <i>ENTR</i> | 0,83 | 0,74 | 1,32 | 0,77 |
| <i>REC</i> (%) | 0,14 | 0,16 | 0,17 | 0,22 |
| <i>LAM</i> | 20,65 | 17,66 | 27,90 | 21,84 |

A fim de verificar a existência de possíveis diferenças estatísticas significativas entre os três grupos de vozes: saudáveis, com paralisia ou com edema foi realizado, para cada uma das medidas, o teste estatístico de análise de variância ANOVA não paramétrico de Kruskal-Wallis. Os resultados indicaram que, para as medidas de entropia, taxa de recorrência e determinismo, há diferenças estatisticamente significativas ($P < 0,001$) entre pelo menos dois dos grupos considerados. Entretanto, não foram encontradas evidências de diferenças estatísticas significativas entre os três grupos para a medida de Laminaridade.

Visando realizar a comparação, dois a dois, entre os grupos de vozes saudáveis, com paralisia e com edema para as medidas de entropia, taxa de recorrência e determinismo foi aplicado o teste não-paramétrico de Mann-Whitney.

Os resultados obtidos revelam diferenças estatisticamente significativas entre o grupo de vozes saudáveis e os outros dois grupos de vozes patológicas para todas as três medidas ($P < 0,001$). Foram observadas ainda diferenças significativas entre os dois grupos de vozes patológicas (paralisia e edema) por meio das medidas de determinismo ($P = 0,001$) e entropia ($P = 0,002$). Por outro lado, não foram reveladas diferenças entre os dois grupos de vozes patológicas para a medida de taxa de recorrência ($P = 0,402$).

Nas Figuras 9, 10, 11 e 12 estão representadas as distribuições dos valores das medidas avaliadas para os sinais de vozes saudáveis, com paralisia, ou edema.

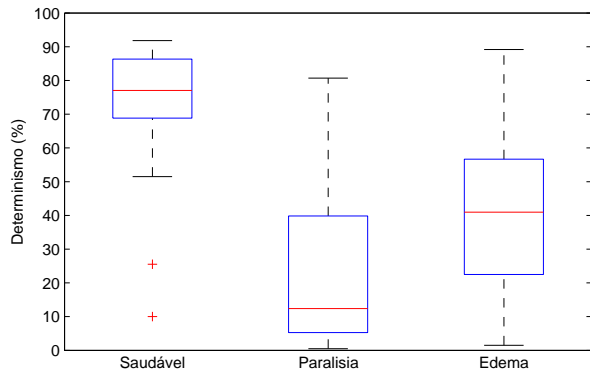


Fig. 9. Distribuição dos valores do determinismo para os sinais de vozes saudáveis, com paralisia ou edema.

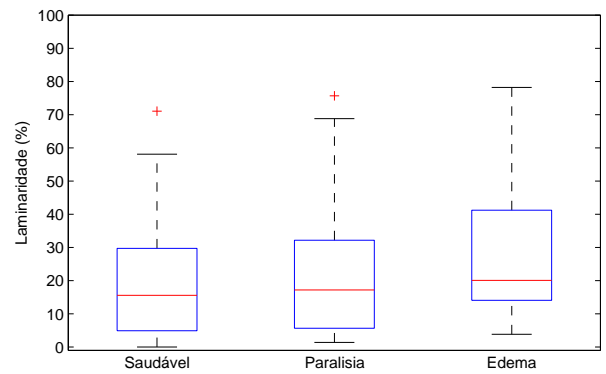


Fig. 12. Distribuição dos valores de laminaridade para os sinais de vozes saudáveis, com paralisia ou edema.

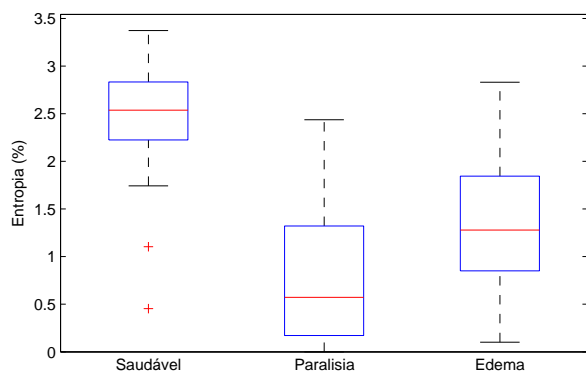


Fig. 10. Distribuição dos valores de entropia para os sinais de vozes saudáveis, com paralisia ou edema.

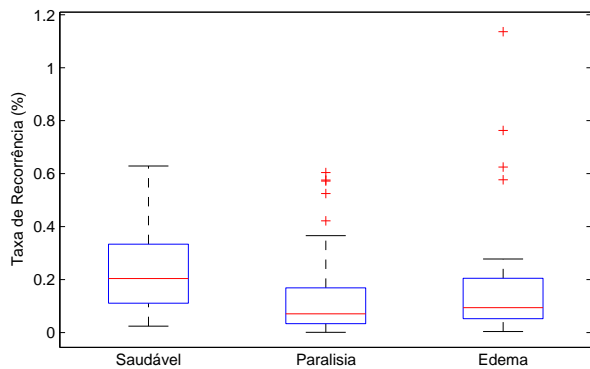


Fig. 11. Distribuição dos valores da taxa de recorrência para os sinais de vozes saudáveis, com paralisia ou edema.

V. CONCLUSÕES

Os resultados apresentados mostram que há diferenças estatisticamente significativas entre os sinais de vozes saudáveis e os sinais de vozes patológicas (paralisia e edema) para as medidas Determinismo, Entropia e Taxa de Recorrência. Porém, não foram constatadas diferenças entre os grupos de sinais avaliados em relação à medida Laminaridade.

As medidas de Entropia e Determinismo apresentam ainda evidências de diferenças estatísticas relevantes entre os sinais de vozes com a patologia edema daqueles afetados por para-

lisia. A aplicação das medidas de quantificação baseadas nos Gráficos de Recorrência como técnica de análise do sistema de produção da fala sugere resultados promissores, no que diz respeito à potencialidade de discriminação entre diferentes grupos de sinais de voz.

Por meio de uma análise visual dos gráficos de recorrência pode-se observar diferenças significativas entre os grupos de sinais. No entanto, a análise visual é subjetiva, podendo levar a diferentes diagnósticos, dependendo do profissional. Por outro lado, as medidas quantitativas fornecem uma forma mais objetiva de análise, podendo levar a resultados mais confiáveis.

REFERÊNCIAS

- [1] N. Marwan. "Encounters With Neighbours - Current Developments of Concepts Based on Recurrence Plots and Their Applications". Ph.D. thesis, University of Potsdam, 2003.
- [2] J. P. Eckmann, S. O. Kamphorst and D. Ruelle. "Recurrence plots of dynamical systems". *Europhys. Letters*, vol. 4(9), pp. 973-977, 1987.
- [3] N. Marwan and J. Kurths. "Line structure in recurrence plots". *Physics Letters A*, vol. 336, pp. 349-357, 2004.
- [4] A. Kumar and S. K. Mullick. "Nonlinear dynamical analysis of speech". *Journal Acoustic Society of America*, vol. 100(1), pp. 615-629, July 1996.
- [5] J. Jiang, Y. Zhang and C. McGilligan. "Chaos in voice, from modeling to measurement". *Journal of Voice*, vol. 20 (1), pp. 2-17, January 2006.
- [6] 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.
- [7] F. Takens. "Detecting strange attractors in turbulence Dynamical Systems and Turbulence". In *Lecture Notes in Mathematics*, edited by D. Rand and L.-S. Young, pp. 366-381. Berlin: Springer, 1981.
- [8] E. G. Souza. "Caracterização de sistemas dinâmicos através de gráficos de recorrência". Master's thesis, Universidade Federal do Paraná, 2008.
- [9] C. L. Webber-Jr. and J. P. Zbilut. "Dynamical assessment of physiological systems and states using recurrence plot strategies". *J. Appl. Physiol.*, vol. 76, pp. 965-973, 1994.
- [10] K. Elemetrics. "Kay Elemetrics Corp. Disordered Voice Database". Model 4337, 03 Ed., 1994.
- [11] E. Kononov. "Visual Recurrence Analysis 4.9", March 2006. Software disponível em <http://nonlinear.110mb.com/vra/>.
- [12] C. L. Webber-Jr. "Recurrence Quantification Analysis, v. 13.1", June 2009. software package disponível em <http://homepages.luc.edu/~cwebber/>.
- [13] C. L. Webber-Jr. and J. P. Zbilut. "Recurrence quantification analysis of nonlinear dynamical systems". In *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, edited by M. A. Riley and G. C. V. Orden, 2005.
- [14] V. T. Motta and P. F. de Oliveira Filho. *SPSS - Análise de Dados Biomédicos*. Medbook, 2009.