

# USO DA DIMENSÃO DE CORRELAÇÃO NA DISCRIMINAÇÃO ENTRE VOZES SAUDÁVEIS E PATOLÓGICAS

**Washington César COSTA(1); Benedito G. AGUIAR NETO (2); Silvana C. COSTA (3)**

(1) UFCG/Campus C. Grande, IFPB/Campus J. Pessoa, e-mail: washington@ifpb.edu.br

(2) UFCG/Campus C. Grande, e-mail: bganeto@dee.ufcg.edu.br

(3) IFPB/Campus J. Pessoa, e-mail: silvana@ifpb.edu.br

## RESUMO

Este trabalho trata da aplicação de técnicas baseadas na teoria dos sistemas dinâmicos não-lineares para o processamento digital de sinais de voz visando à discriminação entre vozes saudáveis e vozes patológicas. Essas técnicas podem ser empregadas como ferramenta auxiliar no pré-diagnóstico de patologias laringeas, reduzindo a quantidade de exames baseados em videolaringoscopia, considerados desconfortáveis aos pacientes por ser esta uma técnica invasiva. A dimensão de correlação, uma das invariantes da dinâmica não-linear, é empregada para descrever a complexidade da voz humana a partir da análise de um banco de dados contendo vozes saudáveis e vozes afetadas por diferentes patologias da laringe. A análise realizada demonstra uma característica de dinâmica de baixa dimensão tanto para as vozes saudáveis quanto para as vozes patológicas, bem como significativas diferenças estatísticas entre estes dois grupos. A variação da dimensão de correlação em relação ao gênero dos locutores também é avaliada. Os resultados encontrados sugerem o uso, em potencial, de técnicas de análise não linear, como a dimensão de correlação, no processamento de sinais de vozes patológicas para aplicações clínicas.

**Palavras-chave:** sistemas não-lineares, processamento digital de sinais de voz; patologias da laringe.

## 1. INTRODUÇÃO

A produção da voz é um processo fortemente não-linear envolvendo efeitos biomecânicos e aerodinâmicos (JIANG e ZHANG, 2002). A teoria do caos determinístico, um ramo da teoria dos sistemas dinâmicos não-lineares, oferece uma alternativa para a análise do sinal de voz. Caos é um termo que descreve um comportamento pseudo-aleatório gerado por um sistema que é ao mesmo tempo determinístico e não-linear (JIANG et al., 2006).

A saída de um sistema caótico é não-predizível e extremamente sensível a pequenas diferenças nas condições iniciais. Contudo, métodos baseados na dinâmica não-linear são capazes de descrever de forma quantitativa o comportamento caótico.

Os métodos de dinâmica não-linear tem sido usados no processamento de séries temporais associadas a sinais de voz em aplicações que vão da caracterização deste tipo de sinal (PETRY, 2002; JOHNSON et al., 2005) a aplicações na análise da voz apresentado algum tipo de patologia na laringe (JIANG e ZHANG, 2002, BUTTE et al., 2009, ARIAS-LONDOÑO et al., 2009; HENRIQUEZ et al., 2009).

Observações clínicas demonstram que a maioria das doenças na laringe causa mudanças na qualidade da voz, o que resulta em irregularidades e instabilidades vocais. Métodos de dinâmica não-linear são capazes de analisar comportamentos irregulares e podem ser valiosos em várias áreas do estudo da voz, incluindo avaliação de tratamentos clínicos, classificação de vozes de acordo com a taxonomia da desordem proposto por Titze (TITZE, 1995), diferenciação de vozes saudáveis e patológicas, e possibilidade de diagnóstico de patologias da laringe (JIANG et al., 2006).

Durante a produção de alguns tipos de sons da fala, considerados sonoros, como por exemplo, as vogais, as dobras vocais vibram impulsionadas pelo ar que vem dos pulmões, durante a expiração (O'SHAUGNESSY, 2000). A presença de uma patologia, como um nódulo, pólipos, cisto ou mesmo, paralisia nas dobras vocais pode afetar o grau de fechamento da glote, modificando a natureza do som produzido. A voz pode se apresentar rouca, soprosa e com dificuldades de manter-se uma elocução sustentada. Exames clínicos tradicionais fazem uso da escuta da voz do paciente e da inspeção visual da laringe, durante os quais o paciente produz o som de uma vogal de forma sustentada, e um patologista de voz observa a vibração das dobras vocais durante a produção do som e avalia o grau de modificação morfológica da anatomia laríngea. Estas duas avaliações, uma subjetiva e outra objetiva dão um diagnóstico da patologia (GODINO-LLORENTE et al, 2006; COSTA, 2008). Em uma avaliação sistemática o patologista de voz utiliza os seus conhecimentos sobre a anatomia e a fisiologia dos mecanismos laríngeos e de fatores etiológicos comuns para o seu diagnóstico a cerca de uma dada patologia (STEMPLE et al., 2010).

Exames à base de videolaringoscopia são considerados desconfortáveis ao paciente e de custo relativamente alto, devido à necessidade de equipamentos especiais. Técnicas baseadas em análise acústica utilizam apenas um computador pessoal e um microfone associados a um programa especialista de baixo custo computacional, que pode auxiliar o médico em acompanhamentos de tratamentos a base de medicamentos, pós-cirúrgicos e durante terapias fonoaudiológicas (GODINO-LLORENTE et al, 2006).

A presente pesquisa trata de um estudo experimental no qual medidas de dinâmica não-linear, como a dimensão de correlação (*Correlation Dimension* - CD), são usadas para caracterizar a complexidade presente em sinais de voz falados de forma sustentada por locutores com vozes saudáveis e locutores apresentando algum tipo de patologia nas dobras vocais. A dimensão de correlação oferece uma estimativa do número de variáveis independentes necessárias para modelar o trato vocal (KUMAR e MULLICK, 1990).

Os resultados demonstram uma característica dinâmica de baixa dimensão, tanto para sinais saudáveis como para sinais patológicos. A análise da dimensão de correlação indica diferenças estatísticas significativas entre os grupos de sinais estudados (saudáveis e patológicos) e mesmo quando se considera a distinção de gênero. Independentemente do gênero dos locutores cujas vozes têm sido analisadas os sinais patológicos apresentam uma maior complexidade, refletida em valores mais altos da dimensão de correlação, em relação ao grupo das vozes saudáveis.

## 2. FUNDAMENTAÇÃO TEÓRICA

Uma série temporal representando o sinal de voz é obtida pela indexação no tempo dos valores de amplitude de suas amostras, matematicamente:  $x(t_1), x(t_2), \dots, x(t_N)$ , onde  $x(t_i) \in R$ ,  $i = 1, 2, \dots, N$  e  $N$  é o número total de amostras do sinal.

A dinâmica do sistema não-linear pode ser reconstruída pela técnica dos tempos de retardo ou Método das Coordenadas Defasadas (TAKENS, 1981) em que vetores  $\xi_i$   $m$ -dimensionais são reconstruídos a partir da série temporal,

$$\xi_i = \{x(t_i), x(t_i + \tau), \dots, x(t_i + (m-1)\tau)\} \quad [\text{Eq. 01}]$$

onde  $i = 1, 2, \dots, N - (m-1)\tau$ ,  $m$  é a dimensão de imersão e  $\tau$  é o passo de reconstrução ou tempo de defasagem. Teoricamente, quando  $m > 2D + 1$  ( $D$  é a dimensão fractal), o espaço de fase reconstruído é topologicamente equivalente (difeomorfismo) ao sistema original (FIEDLER-FERRARA e PRADO, 1994).

A determinação eficiente dos parâmetros de imersão (tempo de defasagem e dimensão de imersão) é de extrema importância para a identificação dos invariantes geométricos. A evolução temporal dos vetores (composição das amostras do sinal original e suas versões defasadas) no espaço de fase reconstruído é chamada de atrator. A escolha adequada do tempo de defasagem permite a reconstrução otimizada do atrator no espaço de fase (JIANG et al., 2006).

## 2.1 A Dimensão de Correlação

A dimensão de correlação ( $D_2$ ) é uma medida geométrica de uma trajetória no espaço de fase que descreve o quanto dois pontos dessa trajetória são correlacionados. Esta medida representa a complexidade dos sistemas. Assim, um sistema mais complexo apresenta uma maior dimensão, o que significa que um maior número de variáveis de estado é necessário para descrever sua dinâmica. O valor de  $D_2$  de uma fonte de ruído aleatório, por exemplo, não converge com aumento da dimensão de imersão  $m$ ; por outro lado, um sistema determinístico apresenta um valor finito para  $D_2$  (JIANG et al., 2006).

A dimensão de correlação pode ser escrita como (GRASSBERGER e PROCACCIA, 1983)

$$D_2 = \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log C(N, \varepsilon)}{\log \varepsilon}, \quad [\text{Eq. 02}]$$

onde na integral de correlação

$$C(N, \varepsilon) = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \theta(\varepsilon - \|\xi_i - \xi_j\|), \quad [\text{Eq. 03}]$$

$\varepsilon$  é o raio em torno de  $\xi_i$  e a função  $\theta(x)$  satisfaz à condição:  $\theta(x) = 1$  para  $x > 0$  e  $\theta(x) = 0$  para  $x \leq 0$ .  $\xi_i$  e  $\xi_j$  representam vetores reconstruídos a partir de defasagens na série temporal associada ao sinal de voz.

A Eq. 02 converge muito lentamente quando  $\varepsilon \rightarrow 0$ . Alternativamente, pode ser usado o método de cálculo da declividade local

$$D_2 = \frac{d \log C(N, \varepsilon)}{d \log \varepsilon} \simeq \lim_{\Delta \log \varepsilon \rightarrow 0} \frac{\Delta \log C(N, \varepsilon)}{\Delta \log \varepsilon}. \quad [\text{Eq. 04}]$$

Em algumas regiões de escala, a curva de inclinação local é relativamente plana. Quando o tamanho da série  $N$  é suficientemente grande, o valor de  $D_2$  irá convergir com o aumento da dimensão de imersão  $m$ .

## 3. METODOLOGIA, RESULTADOS E DISCUSSÃO

Os sinais de voz utilizados para as avaliações realizadas neste trabalho fazem parte de um banco de dados comercial, o *Massachusetts Eye and Ear Infirmary (MEEI) Voice Disorders Database* distribuído pela Kay Elemetrics (KAY ELEMETRICS, 1994). Este banco de dados vem sendo largamente empregado em pesquisas de âmbito internacional envolvendo análise acústica de vozes desordenadas e na discriminação entre vozes saudáveis e patológicas. Ele contém pronúncias sustentadas da vogal /a/, sendo 53 arquivos de locutores com voz saudável e 657 sinais patológicos. Os sinais de voz saudáveis são amostrados a uma frequência de 50 kHz, enquanto que os sinais patológicos são amostrados a 25 kHz.

Vogais sustentadas são comumente utilizadas em experimentos semelhantes em virtude do sistema de produção da voz usar a maior parte de seus mecanismos (por exemplo, fluxo de ar glotal constante e vibração

continua das dobras vocais) na pronúncia deste tipo de som. Assim, vários tipos de anomalias destes mecanismos podem ser detectados. Além disso, pronúncias sustentadas são independentes da linguagem (HENRÍQUEZ et al., 2009).

Na presente pesquisa foi utilizado aproximadamente um segundo de gravação de cada um dos 53 sinais provenientes de locutores saudáveis, como também de 113 sinais de pacientes com algum tipo de patologia na laringe (paralisia, nódulo ou edema). Dentre os dois grupos de sinais usados, 58 são de locutores do gênero masculino e 98 do gênero feminino.

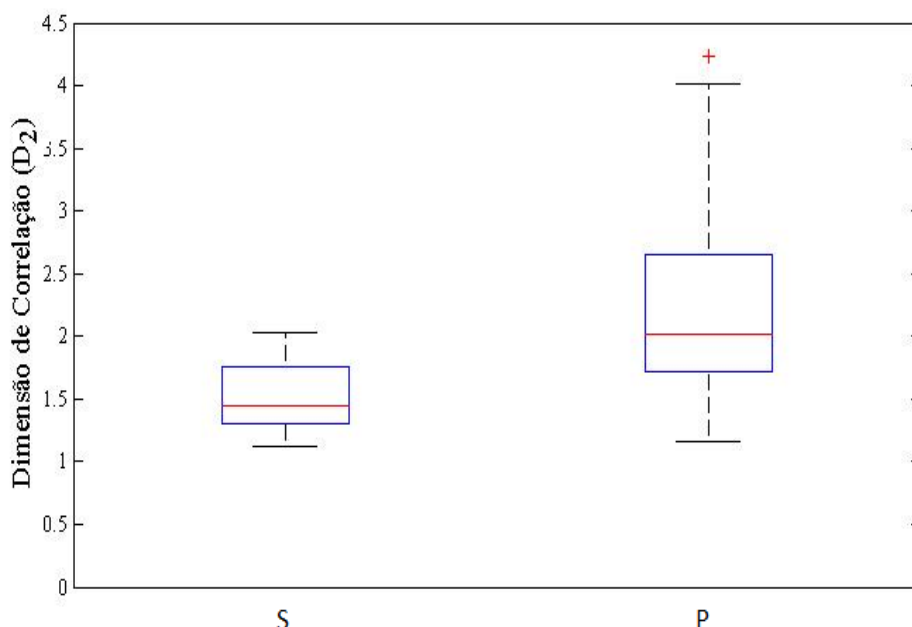
O processo usado neste experimento é dividido em três partes: pré-processamento do sinal, extração das medidas de análise não-linear e caracterização dos sinais considerados. Na etapa de pré-processamento, todos os sinais de voz saudável são re-amostrados com uma frequência de 25 kHz.

Todas as medidas de análise dinâmica do sinal de voz utilizadas neste trabalho foram determinadas com o auxílio do pacote de rotinas TISEAN (HEGGER et al., 1999).

O primeiro passo visando à reconstrução do atrator no espaço de fase é a determinação do tempo de defasagem ( $\tau$ ) ótimo para cada sinal considerado. Levando-se em consideração o intervalo de tempo em que o sinal de voz é admitido com sendo estacionário (16-32 ms), cada sinal é segmentado em 30 blocos de 800 amostras (sem sobreposição) usando uma janela retangular. O tempo de defasagem ótimo é encontrado, para cada bloco de 32 ms, como sendo o valor do primeiro mínimo da função de informação mútua (FRASER e SWINNEY, 1986). Finalmente, o atraso ótimo é tomado como sendo o valor médio destes valores.

Determinado o tempo de defasagem ótimo para cada sinal, a dimensão de correlação  $D_2$  é calculada para cada bloco de 32 ms. O valor da dimensão de correlação do sinal analisado é tomado como sendo o valor médio determinado em cada bloco.

Na Figura 1 está ilustrada a distribuição dos valores da dimensão de correlação,  $D_2$ , de todos os sinais de voz utilizados nos experimentos. As alturas das bases dos retângulos indicam o primeiro e o terceiro quartis, respectivamente. A linha dentro dos retângulos indica a mediana da distribuição e o sinal (+) marca um valor atípico (*outliers*).

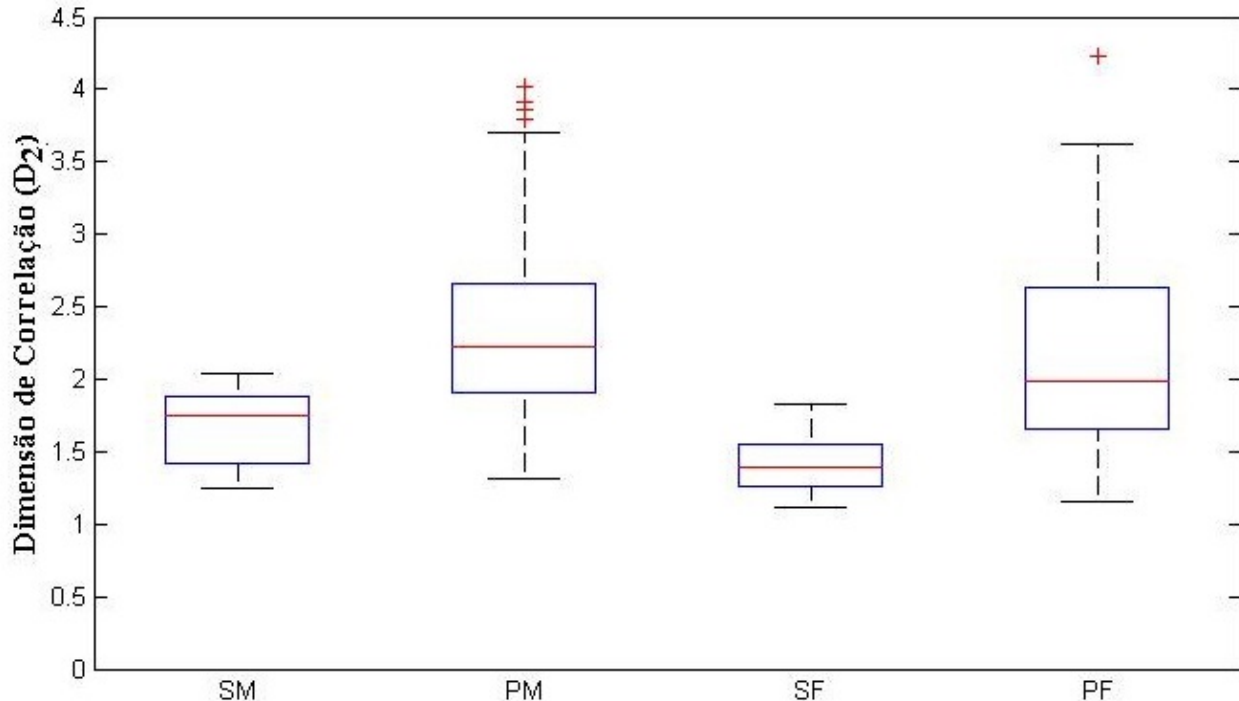


**Figura 1 – Distribuição dos valores da dimensão de correlação para sinais de vozes saudáveis (S) e patológicas (P).**

Como pode ser observado na Fig. 1, os resultados demonstram como característica dos sinais de voz uma dinâmica de baixa dimensão, tanto no caso da voz saudável como na voz patológica ( $D_2 < 3$ ) para a maioria dos sinais analisados. Além disso, a análise da dimensão de correlação indica diferenças estatísticas significativas entre os dois grupos. Os valores da média e do desvio padrão ( $\mu = 1,52$ ;  $\sigma = 0,26$ ) de  $D_2$  dos

sinais de vozes saudáveis são mais baixos do que aqueles encontrados para os sinais de vozes patológicas ( $\mu = 2,23$ ;  $\sigma = 0,69$ ).

Na Figura 2 está ilustrada a distribuição dos valores da dimensão de correlação para os mesmos sinais considerados na análise anterior, porém desta vez é realizada uma investigação do comportamento desta invariante geométrica quando se faz uma discriminação de gênero.



**Figura 2 – Distribuição dos valores da dimensão de correlação para sinais de vozes saudáveis masculinas (SM), patológicas masculinas (PM), saudáveis femininas (SF) e patológicas femininas (PF).**

A partir da análise ilustrada na Figura 2 pode-se observar que os sinais dos locutores de gênero feminino apresentam, em média, valores de  $D_2$  inferiores em relação aos calculados para os sinais de vozes masculinas em ambos os grupos (saudável e patológico). Contudo, a significativa diferença entre os valores das medianas dos dois grupos (masculino e feminino) de sinais saudáveis é sensivelmente diminuída quando se compara os sinais patológicos de ambos os grupos.

Independentemente do gênero do locutor, observa-se que o grupo de vozes patológicas apresenta maior complexidade e dimensões maiores do que o grupo de vozes normais. Assim, a análise da dimensão de correlação efetivamente descreve e classifica os dois tipos de sinais e pode-se concluir que a influência de patologias na qualidade da voz pode ser descrita por meio de métodos de dinâmica não-linear.

#### 4. CONSIDERAÇÕES FINAIS

Neste trabalho são empregadas técnicas de análise não-linear em sinais de voz saudáveis e patológicos no qual é constatada a baixa dimensionalidade de ambos os tipos de sinais. A observação de que a maioria dos sinais de voz analisados é de baixa dimensionalidade ( $D_2 < 3$ ), ou seja, tem atratores de baixa dimensão, permite o uso de um pequeno número de variáveis independentes para reconstrução de um modelo funcional do trato vocal.

A dimensão de correlação apresenta diferenças estatísticas significativas entre os dois grupos considerados, como também na distinção de gênero dos locutores. Desta forma, tem-se a confirmação de que a análise dinâmica não-linear tem potencial suficiente para funcionar como uma medida auxiliar de processamento de sinais de voz para o estabelecimento de um modelo acústico que permita avaliar e detectar patologias da laringe.

Como continuidade da pesquisa pretende-se fazer uso de outras medidas obtidas por meio de técnicas não-lineares e a aplicação de métodos de classificação usando medidas de distorção, a fim de se obter um maior grau de eficiência dos métodos utilizados na discriminação entre vozes saudáveis e patológicas.

## 5. REFERÊNCIAS

- ARIAS-LONDOÑO, J.D.; GODINO-LLORENTE, J.I.; CASTELLANOS-DOMÍNGUEZ, G.; SÁENZ-LECHÓN, N.; OSMA-RUIZ, V. Complexity Analysis of Pathological Voices by means of Hidden Markov Entropy measurements. **1st Annual International Conference of the IEEE EMBS**, Minneapolis, Minnesota, USA, Sept. 2009, p. 2248-2251.
- BUTTE, C. J.; ZHANG, Y.; SONG, H.; JIANG, J.J. Perturbation and Nonlinear Dynamic Analysis of Different Singing Styles. **Journal of Voice**, v. 23, n. 6, p. 647-652.
- COSTA, S. L.N.C. , **Análise Acústica, Baseada no Modelo Linear de Produção da Fala , para Discriminação de Vozes Patológicas**, Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Campina Grande, nov. 2008.
- FRASER, A. M.; SWINNEY, H. L. Independent Coordinates for Strange Attractors from Mutual Information, **Physical Review A**, v.33, p. 1134-1140, 1986.
- FIEDLER-FERRARA, N.; PRADO C. P. C., **Caos: Uma Introdução**. São Paulo: Ed. Edgard Blucher LTDA, 1994.
- GODINO-LLORENTE, J. I.; 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**, v. 53, n. 10, Oct. 2006.
- GRASSBERGER, P. and PROCACCIA, I. Measuring the Strangeness of Strange Attractors. **Physica D**, v. 9, p.189-208, 1983.
- HEGGERR, R.;KANTZ, h.; SCHREIBER, T. Practical implementation of nonlinear time series methods: The TISEAN package, **Chaos**, v. 9, p. 413–435, Jun. 1999.
- HENRÍQUEZ, P.; ALONSO, J. B.; FERRER, M. A.; TRAVIESO, C. M.; GODINO-LLORENTE; J., I. and DÍAZ-de-MARIA, F., Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics, **IEEE Trans. on Audio, Speech, and Language Processing**, vol. 17, N0. 6, p. 1186-1195, August, 2009.
- JIANG, J. J.; ZHANG, Y.; MCGILLIGAN, C. Chaos in Voice, From Modeling to Measurement. **Journal of Voice**, Vol. 20, No. 1, p. 2-17, 2006.
- JIANG, J.J.; ZHANG, Y. Nonlinear dynamic analysis of speech from pathological subjects. **Electronics Letters**, v. 38, n. 6, p. 294-295, Mar. 2002.
- JOHNSON, M. T.; POVINELLI, R. J.; LINDGREN A. C.; YE, J.; LIU, X. and INDREBO, K. M., Time-Domain Isolated Phoneme Classification Using Reconstructed Phase Spaces, **IEEE Trans. Speech and Audio Processing**, v. 13, n.4, p. 458-466, July, 2005.
- KAY ELEMETRICS, *Kay Elemetrics Corp. Disordered Voice Database*, Model 4337, 03 Ed,1994.
- KUMAR, A.; MULLICK, S.K. Attractor dimension, entropy and modeling of speech time series. **Electronics Letters**, v. 26, p. 1790-1792, oct. 1990.
- O'SHAUGHNESSY, D., **Speech Communications: Human and Machine**, 2nd Edition, NY, IEEE Press, 2000.

PETRY, A., **Reconhecimento Automático de Locutor Utilizando Medidas de Invariantes Dinâmicas Não-Lineares**, Tese (Doutorado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Ago. 2002.

STEMPLE, J.C.; GLAZE, L. e KLABEN, B. Clinical Voice Pathology, Theory and Management, 4th Edition, San Diego, Plural Publishing Inc., 2010.

TAKENS, F. Detecting strange attractors in turbulence. **Dynamical Systems and Turbulence**. New York : Springer-Verlag, v. 898, p. 366, 1980.

TITZE, I. R. **Workshop on Acoustic Voice Analysis: Summary Statement**, p. 1-36, Denver, 1995.