



Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Departamento de Engenharia Elétrica  
COPELE - Programa de Pós-Graduação em Engenharia Elétrica

Tese de Doutorado

**Análise Dinâmica Não Linear de Sinais de Voz  
para Detecção de Patologias Laríngeas**

Washington César de Almeida Costa

Campina Grande – PB

09 de novembro de 2012

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Departamento de Engenharia Elétrica  
COPELE - Programa de Pós-Graduação em Engenharia Elétrica

## **Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas**

Washington César de Almeida Costa

Tese de Doutorado submetida à COPELE - Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande - como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Processamento da Informação/Comunicações.

Prof. D.Sc. Francisco Marcos de Assis  
Prof. Dr.-Ing. Benedito G. Aguiar Neto  
Orientadores

Campina Grande – PB, Paraíba, Brasil

©Washington César de Almeida Costa

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

C837a Costa, Washington César de Almeida.

Análise dinâmica não linear de sinais de voz para detecção de patologias laríngeas / Washington César de Almeida Costa. – Campina Grande, 2012. 147 f. : il. color.

Tese (Doutorado em Engenharia Elétrica), Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Orientadores: Prof. Dr. Francisco Marcos de Assis, Prof. Dr.-Ing. Benedito Guimarães Aguiar Neto.

Referências

1. Processamento Digital.
2. Sinais de Voz.
3. Análise não Linear.
4. Patologias da Laringe. I. Título.

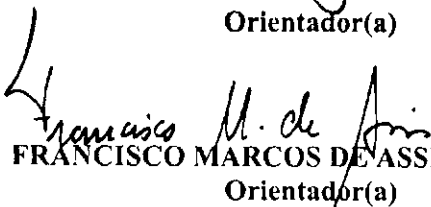
CDU 004.032.3(043)


"Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas"

WASHINGTON CÉSAR DE ALMEIDA COSTA


TESE APROVADA EM 09/11/2012


  
BENEDITO GUIMARÃES AGUIAR NETO, Dr.-Ing., UFCG  
Orientador(a)

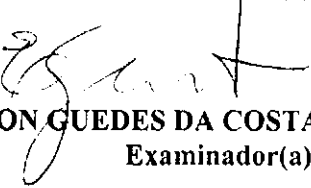
  
FRANCISCO MARCOS DE ASSIS, Dr., UFCG  
Orientador(a)

  
EDUARDO ANTONIO BARROS DA SILVA, Ph.D., COPPE/UFRJ  
Examinador(a)

  
JOSEANA MACEDO FECHINE RÉGIS DE ARAÚJO, D.Sc., UFCG  
Examinador(a)

  
ROSÂNGELA FERNANDES COELHO, Docteur, IME  
Examinador(a)

  
GUILHERME DE ALENCAR BARRETO, Dr, UFC  
Examinador(a)

  
EDSON GUEDES DA COSTA, D.Sc., UFCG  
Examinador(a)

CAMPINA GRANDE - PB

*À minha esposa Silvana, às minhas filhas Isabella e Ana  
Luíza e aos meus filhos Eduardo e Daniel.*

---

---

# Agradecimentos

---

- A Deus, inteligência suprema, causa primária de todas as coisas;
- Aos meus pais que me oportunizaram esta existência no mundo dos homens;
- A minha avó Joaquina (*in memoriam*) pelo amor a mim dedicado durante todo o tempo que durou a nossa convivência;
- Ao meu tio Flávio e a minha tia Terezinha pelo patrocínio de meus estudos, e por acreditarem que a educação é o caminho mais seguro para as conquistas do ser humano;
- À minha esposa Silvana por iluminar os meus caminhos e também pelo apoio, ajuda, compreensão, carinho, paciência e amor;
- Ao professor Benedito Aguiar por acreditar que eu daria "conta do recado" e pela orientação acadêmica;
- Ao professor Francisco Marcos por aceitar participar da minha orientação acadêmica e pela oportunidade que me foi dada em aproveitar seus conhecimentos valiosos e necessários à execução deste trabalho;
- Ao professor Antonio Marcos Moreira do Departamento de Estatística, do *campus* João Pessoa, da Universidade Federal da Paraíba, por sua prestimosa ajuda;
- Aos professores examinadores que fizeram parte das minhas bancas de qualificação e/ou de defesa de tese, pelas valiosas sugestões, que contribuíram para o enriquecimento deste trabalho de pesquisa;
- Aos meus colegas de pesquisa e pós-graduação do Iquanta que me receberam de braços abertos. Em especial, agradeço a Elloá pelo espírito de ajuda sempre presente.
- À Angela Matias, secretária da COPELE, pela presteza e dedicação junto ao Programa de Pós-graduação em Engenharia Elétrica da UFCG;
- Ao Instituto Federal de Educação, Ciência e Tecnologia da Paraíba e à Universidade Federal de Campina Grande pela oportunidade de completar minha formação acadêmica.

*A chave para compreender a aleatoriedade e toda a matemática não é ser capaz de intuir imediatamente a resposta para qualquer problema, e sim possuir as ferramentas para encontrar a solução.*

—LEONARD MLODINOW

---

---

# Resumo

---

Patologias na laringe podem afetar a qualidade vocal, prejudicando a comunicação humana. As técnicas objetivas tradicionais para o diagnóstico dessas patologias fazem uso de exames considerados invasivos, causando certo desconforto ao paciente. Análise acústica, utilizando técnicas de processamento digital de sinais de voz, pode ser utilizada para o desenvolvimento de ferramentas não invasivas de auxílio ao diagnóstico de patologias laríngeas. A precisão do diagnóstico, contudo, depende da escolha das características e parâmetros da fala que melhor representem a desordem vocal provocada por uma determinada patologia. Este trabalho trata da caracterização e da classificação de sinais de vozes saudáveis e vozes afetadas por diferentes patologias laríngeas (edema, paralisia e nódulos nas pregas vocais), por meio da análise dinâmica não linear (e teoria do caos), como também por meio da análise de quantificação de recorrência. No processo de caracterização é investigado, por meio de testes estatísticos, o potencial de cada característica em discriminar os tipos de sinais de voz considerados. Para a classificação é empregada a técnica de análise discriminante com as funções linear ou quadrática, com validação cruzada, sendo considerado um intervalo de confiança de 95% para as médias das taxas de acurácia do classificador. A partir da combinação de características dos conjuntos das medidas de análise não linear (MNL) e das medidas de quantificação de recorrência (MQR), as médias da taxa de acurácia obtidas variaram nos intervalos de confiança: [95,44%; 100%] para a classificação entre vozes saudáveis e patológicas; [94,75%; 100%] entre vozes saudáveis e afetadas por edema, e entre saudáveis e nódulos. Para a classificação entre saudável e paralisia, obteve-se uma acurácia de 100%. Também são avaliados os efeitos do uso de vetores híbridos formados por características MNL, MQR e coeficientes extraídos da análise preditiva linear (LPC). Neste caso, as taxas de acurácia variaram nos intervalos de confiança: [95,02%; 97,62%] na discriminação entre vozes afetadas por paralisia e edema; [98,29%; 99,93%] para paralisia *versus* nódulos e [97,98%; 99,84%] para edema *versus* nódulos. Os resultados encontrados indicam que o método utilizado é promissor, podendo ser empregado no desenvolvimento de uma ferramenta computacional para apoio ao diagnóstico de patologias laríngeas.

**Palavras-chaves:** Processamento Digital, Sinais de Voz, Análise Não Linear, Patologias da Laringe.



---

---

# Abstract

---

Laryngeal pathologies may affect the voice quality, harming human communication. The traditional objective techniques for diagnosing these pathologies make use of exams, considered invasive, causing discomfort to the patient. Acoustic analysis, using digital speech signal processing techniques, can be used for the development of non-invasive tools in order to aid laryngeal diseases diagnosis. The accuracy of diagnosis, however, depends on the choice of parameters and the speech characteristics that better represent the voice disorder caused by a given pathology. This work deals with the characterization and classification of healthy voice signals and voices affected by different laryngeal diseases (edema, paralysis and vocal fold nodules), by means of nonlinear dynamic analysis (and chaos theory) as well as recurrence quantification analysis. In the characterization process, the potential of each feature is investigated to discriminate the types of voice signals considered, by means of statistical tests. For the classification, the technique of discriminant analysis is employed with linear or quadratic functions, with cross-validation. A 95% confidence level was considered for the average of accuracy rates of the classifier performance. From the feature combination of the set of nonlinear analysis measures (MNL) and the quantification recurrence measures (MQR), the average of accuracy rates varied in the following confidence intervals: [95.44%; 100%] for healthy and pathological classification; [94.75%; 100%] between healthy and edema voices, and also between healthy and nodules. The accuracy rate was 100% between healthy and paralysis. We also evaluated the effects of using hybrid vectors formed by MNL, MQR and linear predictive coding (LPC) coefficients. In this case, the accuracy rates ranged in the confidence intervals: [95.02%; 97.62%] in the paralysis versus edema voices discrimination; [98.29%; 99.93%] for paralysis versus nodules and [97.98%; 99.84%] for edema versus nodules. Obtained results indicate that the used method is promising and it can even be used to develop a computational tool to support diagnosis of laryngeal diseases.

**Key-words:** Digital processing, Speech signals, Nonlinear Analysis, Laryngeal pathologies.

---

---

# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivos da Pesquisa . . . . .	4
1.3	Organização do Trabalho . . . . .	4
<b>2</b>	<b>Sistema de Produção da Voz e Patologias da Laringe</b>	<b>6</b>
2.1	Anatomia e Fisiologia do Sistema de Produção da Fala . . . . .	6
2.1.1	A Laringe . . . . .	7
2.1.2	Estrutura Anatômica da Prega Vocal . . . . .	9
2.2	Mecanismo de Produção da Fala . . . . .	10
2.3	Modelo Linear Discreto no Tempo para a Produção da Fala . . . . .	11
2.4	Patologias da Laringe . . . . .	12
2.4.1	Edema de Reinke . . . . .	13
2.4.2	Nódulos nas Pregas Vocais . . . . .	14
2.4.3	Paralisia nas Pregas Vocais . . . . .	15
2.5	Evidências de Não Linearidade e Caos na Voz . . . . .	16
<b>3</b>	<b>Fundamentação Teórica</b>	<b>21</b>
3.1	Reconstrução do Atrator a partir de uma Série Temporal . . . . .	22
3.1.1	Método das Coordenadas Defasadas . . . . .	23
3.1.2	A Escolha do Passo de Reconstrução (tempo de defasagem) . . . . .	23
3.1.3	A Escolha da Dimensão de Imersão . . . . .	24
3.2	Medidas da Análise Dinâmica Não Linear . . . . .	25
3.2.1	Expoentes de Lyapunov . . . . .	25
3.2.2	Dimensão Fractal . . . . .	26
3.2.3	Dimensão de Capacidade e Dimensão de Hausdorff . . . . .	27
3.2.4	Dimensão de Correlação . . . . .	28
3.2.5	Entropia de Shannon . . . . .	29
3.2.6	Entropia de Rényi . . . . .	30
3.2.7	Entropia de Kolmogorov-Sinai . . . . .	30

3.2.8	Entropia de Correlação	31
3.2.9	Entropia Aproximada	32
3.2.10	Entropia de Tsallis	32
3.2.11	Expoente de Hurst	33
3.3	Gráficos de Recorrência e suas Medidas de Quantificação	34
3.3.1	Taxa de Recorrência	36
3.3.2	Determinismo	37
3.3.3	Entropia	37
3.3.4	Comprimento Máximo das Estruturas Diagonais	37
3.3.5	Tendência	38
3.3.6	Laminaridade	39
3.3.7	Tempo de Permanência	39
3.3.8	Comprimento Máximo das Estruturas Verticais	39
3.4	O Uso de Testes Estatísticos	39
<b>4</b>	<b>Caracterização de Sinais de Voz Usando Medidas da Análise Não Linear e Medidas de Recorrência</b>	<b>42</b>
4.1	Metodologia Aplicada	42
4.1.1	A Base de Dados	42
4.1.2	Aspectos Metodológicos	44
4.2	Reconstrução da Dinâmica do Sistema Vocal	45
4.3	Medidas da Análise Dinâmica Não Linear (MNL)	48
4.3.1	Dimensão de Correlação ( $D_2$ )	48
4.3.2	Entropia de Correlação ( $K_2$ )	52
4.3.3	Entropia de Shannon ( $H_1$ )	54
4.3.4	Entropia Aproximada ( $ApEn$ )	57
4.3.5	Entropia de Tsallis ( $H_q$ )	59
4.3.6	Maior Expoente de Lyapunov ( $\lambda_1$ )	61
4.3.7	Primeiro Mínimo da Função de Informação Mútua ( $FMMI$ )	63
4.3.8	Expoente de Hurst ( $H$ )	65
4.4	Resumo dos Resultados do Processo de Caracterização - Medidas Não Lineares (MNL)	68
4.5	Medidas de Quantificação de Recorrência (MQR)	69
4.5.1	Determinismo ( $DET$ )	72
4.5.2	Comprimento Máximo das Linhas Diagonais ( $L_{max}$ )	75
4.5.3	Entropia da Distribuição de Frequências dos Comprimentos das Linhas Diagonais ( $ENTR$ )	77
4.5.4	Tendência ( $TREND$ )	80
4.5.5	Laminaridade ( $LAM$ )	82

4.5.6	Comprimento Máximo das Linhas Verticais ( $V_{max}$ ) . . . . .	84
4.5.7	Tempo de Permanência . . . . .	86
4.6	Resumo dos Resultados do Processo de Caracterização - Medidas de Quantificação de Recorrência . . . . .	88
<b>5</b>	<b>Classificação de Sinais de Voz Usando Medidas da Análise Não Linear e Medidas de Recorrência</b>	<b>90</b>
5.1	Introdução . . . . .	90
5.2	Classificação de Vozes Saudáveis <i>versus</i> Patológicas . . . . .	93
5.2.1	Características Avaliadas Individualmente . . . . .	93
5.2.2	Características Avaliadas de Forma Combinada . . . . .	94
5.3	Classificação de Vozes Saudáveis <i>versus</i> Vozes com Edema nas Pregas Vocais .	96
5.3.1	Características Avaliadas Individualmente . . . . .	96
5.3.2	Características Avaliadas de Forma Combinada . . . . .	97
5.4	Classificação de Vozes Saudáveis <i>versus</i> Vozes com Paralisia nas Pregas Vocais	99
5.4.1	Características Avaliadas Individualmente . . . . .	99
5.4.2	Características Avaliadas de Forma Combinada . . . . .	100
5.5	Classificação de Vozes Saudáveis <i>versus</i> Vozes com Nódulos nas Pregas Vocais	103
5.5.1	Características Avaliadas Individualmente . . . . .	103
5.5.2	Características Avaliadas de Forma Combinada . . . . .	104
5.6	Classificação de Vozes Patológicas - Paralisia <i>versus</i> Edema . . . . .	106
5.6.1	Características Avaliadas Individualmente . . . . .	106
5.6.2	Características Avaliadas de Forma Combinada . . . . .	107
5.7	Classificação de Vozes Patológicas - Paralisia <i>versus</i> Nódulo . . . . .	109
5.7.1	Características Avaliadas Individualmente . . . . .	109
5.7.2	Características Avaliadas de Forma Combinada . . . . .	110
5.8	Classificação de Vozes Patológicas - Edema <i>versus</i> Nódulo . . . . .	113
5.8.1	Características Avaliadas Individualmente . . . . .	113
5.8.2	Características Avaliadas de Forma Combinada . . . . .	114
5.9	Intervalos de Confiança para a Média . . . . .	116
5.9.1	A Desigualdade de Chebyshev . . . . .	116
5.10	Resumo dos Resultados do Processo de Classificação . . . . .	117
5.11	Combinando Análise Linear e Análise Não Linear . . . . .	118
5.11.1	Classificação Saudável x Patológica . . . . .	118
5.11.2	Classificação Saudável x Edema . . . . .	120
5.11.3	Classificação Saudável x Paralisia . . . . .	122
5.11.4	Classificação Saudável x Nódulo . . . . .	123
5.11.5	Classificação Paralisia x Edema . . . . .	125
5.11.6	Classificação Paralisia x Nódulo . . . . .	127

---

5.11.7 Classificação Edema x Nódulo . . . . .	128
5.12 Discussão . . . . .	131
<b>6 Trabalhos Correlatos e Estado da Arte</b>	<b>132</b>
6.1 Caracterização dos Sons da Fala . . . . .	132
6.2 Análise Não Linear de Vozes Saudáveis e Patológicas . . . . .	133
<b>7 Conclusões e Sugestões</b>	<b>138</b>
7.1 Introdução . . . . .	138
7.2 Sumário da Pesquisa . . . . .	138
7.3 Aspectos Relevantes dos Resultados . . . . .	140
7.4 Contribuições da Pesquisa . . . . .	141
7.5 Sugestões para Trabalhos Futuros . . . . .	141
<b>Referências Bibliográficas</b>	<b>143</b>

---

---

# Lista de Figuras

---

2.1	Diagrama esquemático do mecanismo de produção da fala [1]. . . . .	7
2.2	Vista posterior da laringe [2]. . . . .	8
2.3	Pregas vocais saudáveis (a) Abdução e (b) Adução [3]. . . . .	8
2.4	Desenho esquemático da arquitetura histológica da prega vocal do adulto, em seção coronal [4]. . . . .	9
2.5	Modelo simplificado do mecanismo de produção da fala (adaptado de [1]). . . .	10
2.6	Modelo linear discreto no tempo para a produção da fala (adaptado de [1]). . .	11
2.7	Edema de Reinke (a) unilateral e (b) bilateral [3], [5]. . . . .	13
2.8	Edema de Reinke severo [3]. . . . .	14
2.9	Pregas vocais com nódulos: (a) abdução; (b) adução [3]. . . . .	15
2.10	Prega vocal direita paralisada: (a) abdução; (b) adução [3]. . . . .	16
2.11	(a) Interpretação clássica da propagação do som através do sistema vocal. (b) Interpretação da dinâmica não linear de fluidos para a propagação do som ao longo do sistema vocal [1]. . . . .	17
2.12	Formas de ondas para sinais de voz: (a) tipo 1; (b) tipo 2; e (c) tipo 3 [6]. . . . .	19
3.1	Ruído branco (a) Série temporal, (b) Retrato de fase e (c) Gráfico de recorrência; $T = 1.000$ , $m = 3$ , $\tau = 1$ e $\varepsilon = 0, 1$ (norma euclidiana). . . . .	36
3.2	Sinal senoidal (a) Série temporal, (b) Retrato de fase e (c) Gráfico de recorrência; $T = 1.000$ , $m = 2$ , $\tau = 4$ e $\varepsilon = 0, 1$ (norma euclidiana). . . . .	36
4.1	Fluxograma representativo da metodologia aplicada visando a classificação dos sinais de voz. . . . .	43
4.2	Reconstrução do atrator da vogal sustentada /ah/ em um intervalo de 32 ms. Parâmetros: $T = 800$ amostras, $m = 3$ e (a) $\tau = 1$ , (b) $\tau = 5$ e (c) $\tau = 10$ . . . . .	46
4.3	Função de Informação Mútua e a determinação do passo de reconstrução "ideal" para um segmento de 32 ms da vogal sustentada /ah/. . . . .	46
4.4	Histogramas dos valores médios de $\tau$ obtidos para os sinais de vozes (a) saudáveis e (b) patológicas. . . . .	47

4.5	Histogramas dos valores médios de $m$ obtidos para os sinais de vozes (a) saudáveis e (b) patológicas. . . . .	47
4.6	Comportamento dos valores da dimensão de correlação para um segmento de 32 ms para um dos sinais de voz saudável em função da dimensão de imersão e do valor de $\varepsilon$ , em que as curvas correspondem as diferentes dimensões de imersão ( $m = 4, \dots, 10$ ) de baixo para cima, respectivamente. . . . .	48
4.7	Comportamento dos valores da dimensão de correlação para um segmento de 32 ms para um dos sinais de voz com paralisia em função da dimensão de imersão e do valor de $\varepsilon$ , em que as curvas correspondem as diferentes dimensões de imersão ( $m = 4, \dots, 10$ ) de baixo para cima, respectivamente. . . . .	49
4.8	Distribuição dos valores médios da dimensão de correlação para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	50
4.9	Distribuição dos valores médios da dimensão de correlação para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	51
4.10	Comportamento dos valores da entropia de correlação para um segmento de 32 ms para um dos sinais de voz saudável em função da dimensão de imersão e do valor de $\varepsilon$ , em que as curvas correspondem as diferentes dimensões de imersão ( $m = 4, \dots, 10$ ) de baixo para cima, respectivamente. . . . .	52
4.11	Distribuição dos valores médios da entropia de correlação para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	53
4.12	Distribuição dos valores médios da entropia de correlação para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	54
4.13	Distribuição dos valores médios da entropia de Shannon para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	55
4.14	Distribuição dos valores médios da entropia de Shannon para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	56
4.15	Distribuição dos valores médios da entropia aproximada para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	57
4.16	Distribuição dos valores médios da entropia aproximada para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	58
4.17	Distribuição dos valores médios da entropia de Tsallis para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	59
4.18	Distribuição dos valores médios da entropia de Tsallis para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	60
4.19	Comportamento de $S(\tau)$ em função do número de iterações, de $\varepsilon$ e de $m$ . . . . .	61

4.20	Distribuição dos valores médios do maior expoente de Lyapunov para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	62
4.21	Distribuição dos valores médios do maior expoente de Lyapunov para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	62
4.22	Distribuição dos valores médios do primeiro mínimo da função de informação mútua para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	64
4.23	Distribuição dos valores médios do primeiro mínimo da função de informação mútua para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	64
4.24	Estimação do expoente de Hurst para um sinal de voz saudável. . . . .	66
4.25	Distribuição dos valores do expoente de Hurst para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL). . . . .	66
4.26	Distribuição dos valores médios do expoente de Hurst para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL). . . . .	67
4.27	Histogramas dos valores de $\tau$ obtidos para os sinais de vozes (a) saudáveis e (b) patológicas. . . . .	69
4.28	Histogramas dos valores de $m$ obtidos para os sinais de vozes (a) saudáveis e (b) patológicas. . . . .	70
4.29	Sinal de voz saudável (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência; $m = 3$ , $\tau = 4$ , $\varepsilon = 0,1$ (norma euclidiana) . . . . .	71
4.30	Sinal de voz patológica (edema) (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência; $m = 3$ , $\tau = 9$ , $\varepsilon = 0,1$ (norma euclidiana) . . . . .	71
4.31	Sinal de voz patológica (nódulos) (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência; $m = 3$ , $\tau = 8$ , $\varepsilon = 0,1$ (norma euclidiana) . . . . .	71
4.32	Sinal de voz patológica (paralisia) (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência; $m = 3$ , $\tau = 17$ , $\varepsilon = 0,1$ (norma euclidiana) . . . . .	72
4.33	Distribuições dos valores do maior raio para sinais de vozes (a) saudáveis e (b) patológicas para uma taxa de recorrência menor ou igual a 1%. . . . .	73
4.34	Distribuições dos valores do maior raio para sinais de vozes (a) saudáveis e (b) patológicas para uma taxa de recorrência menor ou igual a 5%. . . . .	74
4.35	Distribuições dos valores de $DET$ para sinais de vozes saudáveis (SDL) e patológicas (PTL). . . . .	75
4.36	Distribuições dos valores de $L_{max}$ para sinais de vozes saudáveis (SDL) e patológicas (PTL). . . . .	76
4.37	Distribuições dos valores da medida entropia para sinais de vozes saudáveis (SDL) e patológicas (PTL). . . . .	78



4.38	Distribuições dos valores da medida Trend para sinais de vozes saudáveis (SDL) e patológicas (PTL). . . . .	80
4.39	Distribuições dos valores da medida laminaridade para sinais de vozes saudáveis (SDL) e patológicas (PTL). . . . .	82
4.40	Distribuições dos valores da medida $V_{max}$ para sinais de vozes saudáveis (SDL) e patológicas (PTL). . . . .	84
4.41	Distribuições dos valores da medida Traptime ( $TT$ ) para sinais de vozes saudáveis (SDL) e patológicas (PTL). . . . .	86
5.1	Acurácia Média Máxima - Saudável x Patológica (MNL) . . . . .	94
5.2	Acurácia Média Máxima - Saudável x Patológica (MQR) . . . . .	95
5.3	Acurácia Média Máxima - Saudável x Patológica (MNL e MQR) . . . . .	96
5.4	Acurácia Média Máxima - Saudável x Edema (MNL) . . . . .	98
5.5	Acurácia Média Máxima - Saudável x Edema (MQR) . . . . .	98
5.6	Acurácia Média Máxima - Saudável x Edema (MNL e MQR) . . . . .	99
5.7	Acurácia Média Máxima - Saudável x Paralisia (MNL) . . . . .	101
5.8	Acurácia Média Máxima - Saudável x Paralisia (MQR) . . . . .	102
5.9	Acurácia Média Máxima - Saudável x Paralisia (MNL e MQR) . . . . .	102
5.10	Acurácia Média Máxima - Saudável x Nódulo (MNL) . . . . .	104
5.11	Acurácia Média Máxima - Saudável x Nódulo (MQR) . . . . .	105
5.12	Acurácia Média Máxima - Saudável x Nódulo (MNL e MQR) . . . . .	106
5.13	Acurácia Média Máxima - Paralisia x Edema (MNL) . . . . .	107
5.14	Acurácia Média Máxima - Paralisia x Edema (MQR) . . . . .	108
5.15	Acurácia Média Máxima - Paralisia x Edema (MNL e MQR) . . . . .	109
5.16	Acurácia Média Máxima - Paralisia x Nódulo (MNL) . . . . .	111
5.17	Acurácia Média Máxima - Paralisia x Nódulo (MQR) . . . . .	112
5.18	Acurácia Média Máxima - Paralisia x Nódulo (MNL e MQR) . . . . .	112
5.19	Acurácia Média Máxima - Edema x Nódulo (MNL) . . . . .	114
5.20	Acurácia Média Máxima - Edema x Nódulo (MQR) . . . . .	115
5.21	Acurácia Média Máxima - Edema x Nódulo (MNL e MQR) . . . . .	115
5.22	Acurácia Média - Saudável x Patológica . . . . .	119
5.23	Acurácia Média - Saudável x Patológica . . . . .	120
5.24	Acurácia Média - Saudável x Edema . . . . .	121
5.25	Acurácia Média - Saudável x Edema . . . . .	122
5.26	Acurácia Média - Saudável x Paralisia . . . . .	123
5.27	Acurácia Média - Saudável x Nódulo . . . . .	124
5.28	Acurácia Média - Saudável x Nódulo . . . . .	125
5.29	Acurácia Média - Paralisia x Edema . . . . .	126
5.30	Acurácia Média - Paralisia x Edema . . . . .	127

5.31	Acurácia Média - Paralisia x Nódulo . . . . .	128
5.32	Acurácia Média - Paralisia x Nódulo . . . . .	129
5.33	Acurácia Média - Edema x Nódulo . . . . .	130
5.34	Acurácia Média - Edema x Nódulo . . . . .	130

---

---

## Lista de Tabelas

---

4.1	Média e desvio padrão da dimensão de correlação. . . . .	50
4.2	Média e desvio padrão da dimensão de correlação - separação das patologias. . . . .	51
4.3	Valor-p - dimensão de correlação - separação das patologias. . . . .	51
4.4	Média e desvio padrão para a entropia de correlação. . . . .	53
4.5	Média e desvio padrão da entropia de correlação - separação das patologias. . . . .	54
4.6	Valor-p - entropia de correlação - separação das patologias. . . . .	54
4.7	Média e desvio padrão para a entropia de Shannon. . . . .	55
4.8	Média e desvio padrão da entropia de Shannon - separação das patologias. . . . .	56
4.9	Valor-p - entropia de Shannon - separação das patologias. . . . .	56
4.10	Média e desvio padrão para a entropia aproximada. . . . .	57
4.11	Média e desvio padrão da entropia aproximada - separação das patologias. . . . .	58
4.12	Valor-p - entropia aproximada - separação das patologias. . . . .	58
4.13	Média e desvio padrão para a entropia de Tsallis ( $q = 0,5$ ). . . . .	59
4.14	Média e desvio padrão da entropia de Tsallis - separação das patologias. . . . .	60
4.15	Valor-p - entropia aproximada - separação das patologias. . . . .	61
4.16	Média e desvio padrão para o maior expoente de Lyapunov. . . . .	61
4.17	Média e desvio padrão do maior expoente de Lyapunov - separação das patologias. . . . .	63
4.18	Valor-p - maior expoente de Lyapunov - separação das patologias. . . . .	63
4.19	Média e desvio padrão para o primeiro mínimo da função de informação mútua. . . . .	63
4.20	Média e desvio padrão do primeiro mínimo da função de informação mútua - separação das patologias. . . . .	65
4.21	Valor-p - <i>FMMI</i> - separação das patologias. . . . .	65
4.22	Média e desvio padrão para o expoente de Hurst. . . . .	66
4.23	Média e desvio padrão do expoente de Hurst - separação das patologias. . . . .	67
4.24	Valor-p - Expoente de Hurst - separação das patologias. . . . .	68
4.25	Potencial discriminativo das medidas da análise dinâmica não linear. . . . .	68
4.26	Média e desvio padrão para <i>DET</i> . . . . .	73
4.27	Média e desvio padrão para <i>DET</i> - separação das patologias. . . . .	74

4.28	Potencial estatístico da característica <i>DET</i> na discriminação entre grupos de sinais de voz. . . . .	75
4.29	Média e desvio padrão para $L_{max}$ . . . . .	76
4.30	Média e desvio padrão para $L_{max}$ - separação das patologias. . . . .	77
4.31	Potencial estatístico da característica $L_{max}$ na discriminação entre grupos de sinais de voz. . . . .	77
4.32	Média e desvio padrão para <i>ENTR</i> . . . . .	78
4.33	Média e desvio padrão para <i>ENTR</i> - separação das patologias. . . . .	79
4.34	Potencial estatístico da característica <i>ENTR</i> na discriminação entre grupos de sinais de voz. . . . .	79
4.35	Média e desvio padrão para <i>TREND</i> . . . . .	80
4.36	Média e desvio padrão para <i>TREND</i> - separação das patologias. . . . .	81
4.37	Potencial estatístico da característica <i>ENTR</i> na discriminação entre grupos de sinais de voz. . . . .	81
4.38	Potencial estatístico da característica <i>ENTR</i> na discriminação entre grupos de sinais de voz. . . . .	81
4.39	Média e desvio padrão para <i>LAM</i> . . . . .	82
4.40	Média e desvio padrão para <i>LAM</i> - separação das patologias. . . . .	83
4.41	Potencial estatístico da característica <i>LAM</i> na discriminação entre grupos de sinais de voz. . . . .	83
4.42	Potencial estatístico da característica <i>LAM</i> na discriminação entre grupos de sinais de voz. . . . .	83
4.43	Média e desvio padrão para $V_{max}$ . . . . .	84
4.44	Média e desvio padrão para $V_{max}$ - separação das patologias. . . . .	85
4.45	Potencial estatístico da característica $V_{max}$ na discriminação entre grupos de sinais de voz. . . . .	85
4.46	Média e desvio padrão para <i>TT</i> . . . . .	86
4.47	Média e desvio padrão para <i>TT</i> - separação das patologias. . . . .	87
4.48	Potencial estatístico da característica <i>TT</i> na discriminação entre grupos de sinais de voz. . . . .	87
4.49	Potencial estatístico da característica <i>TT</i> na discriminação entre grupos de sinais de voz. . . . .	87
4.50	Potencial discriminativo das medidas de quantificação de recorrência. . . . .	88
5.1	Matriz de confusão em um teste de detecção da presença/ausência de doença. . . . .	91
5.2	Taxas de classificação (%) Saudável x Patológica - MNL avaliadas individualmente (QDA) . . . . .	93
5.3	Taxas de classificação (%) Saudável x Patológica - MQR avaliadas individualmente (QDA) . . . . .	93

5.4	Taxas de classificação (%) Saudável x Edema - MNL avaliadas individualmente (LDA) . . . . .	97
5.5	Taxas de classificação (%) Saudável x Edema - MQR avaliadas individualmente (LDA) . . . . .	97
5.6	Taxas de classificação (%) Saudável x Paralisia - MNL avaliadas individualmente (QDA) . . . . .	100
5.7	Taxas de classificação (%) Saudável x Paralisia - MQR avaliadas individualmente (QDA) . . . . .	100
5.8	Taxas de classificação (%) Saudável x Nódulo - MNL avaliadas individualmente (LDA) . . . . .	103
5.9	Taxas de classificação (%) Saudável x Nódulo - MQR avaliadas individualmente (QDA) . . . . .	103
5.10	Taxas de classificação (%) Paralisia x Edema - MNL avaliadas individualmente (LDA) . . . . .	106
5.11	Taxas de classificação (%) Paralisia x Edema - MQR avaliadas individualmente (LDA) . . . . .	107
5.12	Taxas de classificação (%) Paralisia x Nódulo - MNL avaliadas individualmente (LDA) . . . . .	110
5.13	Taxas de classificação (%) Paralisia x Nódulo - MQR avaliadas individualmente (LDA) . . . . .	110
5.14	Taxas de classificação (%) Edema x Nódulo - MNL avaliadas individualmente (LDA) . . . . .	113
5.15	Taxas de classificação (%) Edema x Nódulo - MQR avaliadas individualmente (LDA) . . . . .	113
5.16	Intervalos de confiança para a acurácia do sistema de classificação dos sinais de voz analisados (nível de confiança de 95%). . . . .	117
5.17	Taxas de classificação (%) Saudável x Patológica - LPC e $L_{max}$ (LDA) . . . . .	119
5.18	Taxas de classificação (%) Saudável x Patológica - LPC+MNL+MQR (LDA) . . . . .	120
5.19	Taxas de classificação (%) Saudável x Edema - LPC e $D_2$ (QDA) . . . . .	121
5.20	Taxas de classificação (%) Saudável x Edema - LPC+MNL+MQR (QDA) . . . . .	122
5.21	Taxas de classificação (%) Saudável x Paralisia - LPC e $L_{max}$ (LDA) . . . . .	122
5.22	Taxas de classificação (%) Saudável x Nódulo - LPC e $H$ (QDA) . . . . .	123
5.23	Taxas de classificação (%) Saudável x Nódulo - LPC+MNL+MQR (QDA) . . . . .	124
5.24	Taxas de classificação (%) Paralisia x Edema - LPC e $FMMI$ (QDA) . . . . .	125
5.25	Taxas de classificação (%) Paralisia x Edema - LPC+MNL+MQR (QDA) . . . . .	126
5.26	Taxas de classificação (%) Paralisia x Nódulo - LPC e $D_2$ (QDA) . . . . .	127
5.27	Taxas de classificação (%) Paralisia x Nódulo - LPC+MNL+MQR (QDA) . . . . .	128
5.28	Taxas de classificação (%) Edema x Nódulo - LPC e $D_2$ (QDA) . . . . .	129
5.29	Taxas de classificação (%) Edema x Nódulo - LPC+MNL+MQR (QDA) . . . . .	130

---

---

# Lista de Siglas

---

- ABVL - Academia Brasileira de Laringologia e Voz
- APQ - *Amplitude Perturbation Quocient* - Nível de perturbação de amplitude
- CRP - *Cross Recurrence Plot*
- CSLP - Camada superficial da lâmina própria
- EDM - Voz com edema nas pregas vocais
- EEG - Eletroencefalograma
- EER - Taxa de erro igual (*Equal Error Rate*)
- EMG - Eletromiografia
- GMM - *Gaussian Mixture Models* - Modelos de misturas gaussianas
- GRBAS - Escala de avaliação vocal perceptiva no nível glótico (G - grau de rouquidão, R - aspereza, B - sopro, A - astenia, S - tensão)
- HMM - *Hidden Markov Models* - Modelos de Markov Escondidos
- HNR - *Harmonic-to-noise ration* - Relação harmônico ruído
- LDA - *Linear Discriminant Analysis* - Análise de discriminante linear
- LPC - *Linear Predictive Coding* - Codificação por predição linear
- MEEI - *Massachusetts Eye and Ear Infirmary Voice and Speech*
- MFCC - *Mel Frequency Cepstral Coefficients* - Coeficientes cepstrais na escala *mel* de frequência
- MLP - *Multi-Layer Perceptron*
- MNL - Medidas da Dinâmica Não Linear
- MQR - Medidas de Quantificação de Recorrência
- NDL - Voz com nódulos nas pregas vocais
- PPQ - *Pitch Perturbation Quocient* - Nível de perturbação do *pitch*
- PRL - Voz com paralisia nas pregas vocais
- PTL - Voz patológica
- QDA - *Quadratic Discriminant Analysis* - Análise discriminante quadrática
- RP - *Recurrence Plot* - Gráfico de recorrência
- RPDE - Entropia da densidade de probabilidade de recorrência
- RQA - *Recurrence Quantification Analysis*
- SDL - Voz saudável

SNR - *Signal-to-noise ratio* - Relação sinal-ruído

SPSS - *Statistical Package for the Social Sciences*

SVM - *Support Vector Machine* - Máquinas de Vetores de Suporte

TISEAN - *Time Series Analysis*

VRA - *Visual Recurrence Analysis*

---

---

# Lista de Símbolos

---

$\{x_i\}$  - Série temporal

$\vec{\xi}_i$  - Vetores  $m$ -dimensionais da série temporal

$m$  - Dimensão de imersão

$\tau$  - Passo de reconstrução ou tempo de defasagem

$x(t)$  - Sinal no domínio do tempo

$x(t + \tau)$  - Versão de  $x(t)$  defasada de  $\tau$

$I(\tau)$  - Função de informação mútua

$d$  - Dimensão do espaço de fases real dos vetores  $\vec{\xi}_i$

$D_0$  - Dimensão de *Hausdorff* do atrator

$S(\tau)$  - Fator de crescimento (*stretching factor*)

$T$  - Número de pontos da série temporal

$M$  - Número de vizinhos próximos

$d(\vec{\xi}_i, \vec{\xi}_j, \tau)$  - Distância (euclidiana) entre os pontos  $\vec{\xi}_i$  e  $\vec{\xi}_j$

$D$  - Dimensão de Capacidade ou dimensão do atrator

$N(\varepsilon)$  - Número de volumes cheios

$D_2$  - Dimensão de correlação

$q(\varepsilon)$  - Fração de pontos do atrator

$N$  - Número de pontos do atrator ou número de estados

$C(\varepsilon)$  - Integral de correlação

$H_1$  - Entropia de Shannon

$I^{(\alpha)}$  - Entropia de Rényi de ordem  $\alpha$

$p_i$  - Probabilidade do evento  $i$

$K$  - Entropia de Kolmogorov-Sinai

$\lambda_i$  - Expoentes de Lyapunov positivos

$K_2$  - Entropia de Correlação

$ApEn$  - Entropia aproximada

$H_q$  - entropia de Tsallis

$R$  - Distância percorrida por uma partícula aleatória

$\sigma$  - Desvio padrão da amostra

$H$  - Expoente de Hurst



$E$  - Dimensão euclidiana

$R_{i,j}^{m,\varepsilon}$  - Definição para gráfico de recorrência

$\varepsilon$  - Raio da vizinhança (*threshold*) no ponto  $\vec{\xi}_i$ ;

$\Theta(\cdot)$  - Função de Heaviside

$REC$  - Taxa de recorrência ou percentual de recorrência

$P^\varepsilon(l)$  - Distribuição de frequência dos comprimentos das estruturas diagonais

$l$  - Comprimentos das estruturas diagonais no gráfico de recorrência ( $RP$ )

$DET$  - Determinismo em um RP

$l_{min}$  - Número mínimo de estruturas diagonais

$ENTR$  - Entropia de Shannon da distribuição de frequência dos comprimentos das linhas diagonais em um RP

$L$  - Comprimento médio das linhas diagonais em um RP

$L_{max}$  - Comprimento máximo das estruturas diagonais

$DIV$  - Divergência

$TREND$  - Tendência: coeficiente de regressão linear sobre a densidade dos pontos de recorrência das diagonais paralelas em relação a diagonal principal em um RP

$LAM$  - Laminaridade: razão entre os pontos de recorrência que formam as linhas verticais e todo o conjunto de pontos recorrentes

$v_{min}$  - Tamanho mínimo com que se deseja computar uma estrutura vertical

$TT$  - *Traptime* ou *Trapping Time* - Tempo de permanência em um estado

$V_{max}$  - Comprimento máximo das estruturas verticais presentes no gráfico de recorrência

$\mu$  - Média

$P$  - Valor  $p$  em um teste de hipóteses

$\beta$  - Constante de Boltzmann

$k_1$  - Constante da entropia de Tsallis

$N_l$  - Número total de linhas diagonais computadas no RP

$N_v$  - Número total de linhas verticais computadas no RP

# CAPÍTULO 1

---

## Introdução

---

A voz é uma emissão acústica voluntária, considerada principal ferramenta de comunicação humana [4]. Ela serve como a melodia da nossa fala e fornece informações de nossos sentimentos, expressão, intenção e estado de espírito [7]. Além disso, a voz é essencial para o exercício de algumas profissões tais como radialistas, repórteres, apresentadores de televisão, cantores e professores.

A partir da análise da voz é possível extrair informações sobre a saúde geral e bem-estar de um indivíduo. Além de nos revelar quem somos e como nos sentimos, a voz fornece consideráveis percepções sobre a estrutura e função de certas partes do corpo [8].

A voz é uma ferramenta de grande complexidade e possui inúmeros fatores envolvidos na sua produção. As estruturas anatômicas e os processos fisiológicos envolvidos no complexo sistema de produção da fala são responsáveis pelos diversos fonemas emitidos e, conseqüentemente, carregam informações sobre suas formas, fisiologia e dinâmica [4].

Para uma boa produção vocal é necessária uma voz saudável. A qualidade vocal relaciona-se com o timbre, a frequência, a amplitude e a composição dos harmônicos da onda sonora produzida, bem como, com o modo de vibração das pregas vocais, o tamanho e o formato do trato vocal, a tensão e o tônus das paredes faríngeas, podendo ser modificada por fatores anatômicos ou patológicos [9].

### 1.1 Motivação

Há uma grande variedade de doenças relacionadas à laringe que causam modificações na voz. Essas patologias podem ser tanto de origem orgânica como nódulos, cistos ou edemas, quanto de origem neurológica, tal como paralisia nas pregas vocais [10, 11]. Algumas dessas patologias como pólipos, nódulos e edemas de Reinke, por exemplo, ocorrem, principalmente, devido a hábitos sociais não-saudáveis - como tabagismo e alcoolismo - e ao abuso vocal. No caso de paralisia, as principais causas incluem traumas cirúrgicos ou acidentais, doença cardiovascular e doenças neurológicas [7].

Estudos realizados indicam que as patologias nódulos, edema de Reinke e paralisia nas pregas vocais estão entre as patologias de maior incidência em adultos, tanto do gênero masculino quanto do gênero feminino. No Reino Unido, por exemplo, cerca de 2,5 milhões de pessoas têm alguma desordem na comunicação, sendo que em cerca de 800 mil, a desordem é severa, prejudicando a produção da fala [12]. No Brasil, em um estudo realizado em 2003, de 45 mil brasileiros, 25% apresentaram problemas na laringe [13]. Na pesquisa aparecem, com maior prevalência, profissionais que trabalham com a voz, tais como: professores, telefonistas, cantores, atores, leiloeiros, políticos, religiosos, atendentes de *telemarketing* e bancários. Entretanto, a literatura científica é praticamente unânime em apontar os professores como o grupo mais vulnerável [14].

O acompanhamento na melhoria da qualidade vocal, bem como o auxílio ao diagnóstico médico de patologias laríngeas, podem ser levados a efeito por meio de uma análise acústica dos sinais de voz, utilizando programas computacionais de fácil acesso e baixo custo. Os métodos de avaliação acústica são não invasivos quando comparados aos exames tradicionais, possibilitando, ainda, um tratamento à distância - caso em que a voz do paciente pode ser, por exemplo, gravada em um computador pessoal e enviada para um médico em outra localidade, e este avaliar a qualidade da voz, efetuar pré-diagnóstico de patologias laríngeas, bem como a evolução de um tratamento medicamentoso ou pós-cirúrgico.

Dessa forma, o interesse em pesquisas para o desenvolvimento de ferramentas de auxílio ao diagnóstico de patologias laríngeas, bem como de apoio fonoterápico, tem aumentado nos últimos anos.

Para que um sistema computacional de auxílio a diagnósticos possa ser implementado com eficiência, ou seja, com altas taxas de acerto na discriminação de vozes afetadas por algum tipo de patologia, um dos aspectos primordiais é a escolha das características ou parâmetros que representem significativamente as desordens na fala introduzidas por uma determinada patologia.

As técnicas acústicas empregadas para discriminação entre vozes saudáveis e patológicas geralmente são baseadas na análise linear do modelo de produção de fala ou em técnicas de análise dinâmica não linear. No primeiro caso, considerando um modelo linear de produção de fala, é possível representar, de forma geral, o sistema de produção de voz como um sistema fonte-filtro, no qual as pregas vocais são consideradas a fonte sonora e o trato vocal, o filtro. O sinal resultante é o sinal acústico da voz [15].

Várias pesquisas têm sido realizadas utilizando técnicas baseadas no modelo linear de produção da fala, por meio de medidas paramétricas ou não-paramétricas, tais como: Codificação por Predição Linear, Análise Cepstral e suas derivadas (coeficientes cepstrais, delta-cepstrais, cepstrais ponderados e delta-cepstrais ponderados), Mel-cepstral (delta mel e delta-delta mel) com aplicações em sistemas de reconhecimento de fala e de locutor, melhoramento do sinal de voz, classificação de vozes patológicas, entre outras [16–24].

Medições detalhadas sobre a estrutura do fluxo de ar na parte frontal do trato vocal mostram a presença de padrões de fluxo distintos, o que está em discordância com o fluxo de ar de forma laminar assumida na teoria fonte-filtro para a produção da fala. Existem várias não linearidades envolvidas na vibração das pregas vocais e na geração da onda glotal [25]. Estas incluem efeitos devidos às fortes forças de restauração na colisão das pregas vocais, turbulência no fluxo de ar devido a uma constrição do trato vocal durante a produção de sons surdos, e ao acoplamento não linear entre a fonte e o trato vocal durante a pronúncia de alguns sons sonoros [26]. Devido a tais fatores, os métodos clássicos de análise de dados baseados em um modelo linear têm sido enriquecidos com novos métodos que são derivados da teoria dos sistemas dinâmicos não lineares [6].

Do ponto de vista dos sistemas dinâmicos, um sinal de voz pode ser considerado como uma observação do sistema de produção da fala, que pode ser usada para revelar e modelar sua dinâmica usando as técnicas não lineares. Ao longo das últimas duas décadas, pesquisas relacionadas aos mecanismos de produção da voz têm sido realizadas considerando as técnicas da dinâmica não linear e da teoria do caos com objetivos variados, dentre os quais podem ser destacados: classificação de fonemas [27], reconhecimento automático de locutor [28], discriminação entre vozes saudáveis e patológicas, diagnóstico de patologias laríngeas e avaliação de efeitos de tratamentos clínicos [6, 29–32].

Vários outros métodos de análise têm tentado estimar medidas e propriedades não lineares de processos naturais. Os gráficos de recorrência [33] têm surgido, nas duas últimas décadas, como mais uma técnica de análise não linear para o estudo da dinâmica de sistemas dissipativos. Embora pequenas perturbações em tais sistemas causem divergência exponencialmente crescente entre os estados, após algum tempo o sistema retorna a um estado que é arbitrariamente próximo a um estado precedente e passa por uma evolução similar. Os gráficos de recorrência permitem visualizar o comportamento recorrente dos sistemas dinâmicos de baixa ou de alta dimensionalidade.

Os gráficos de recorrência exibem padrões característicos de pequena e grande escala presentes nos comportamentos típicos dos sistemas dinâmicos. Entretanto, as estruturas apresentadas nos gráficos de recorrência nem sempre são de fácil interpretação podendo, numa análise subjetiva, levar a diferentes resultados. Como forma de quantificar o comportamento do sistema sob análise, facilitando a análise dos dados e tornando-a mais confiável, Zbilut e Webber [34] desenvolveram a análise de quantificação de recorrência baseada em um conjunto de medidas objetivas a partir dos gráficos de recorrência.

A análise de recorrência tem sido empregada em diferentes áreas, tais como sistemas biológicos - incluindo dados de eletromiografia (EMG) [35], gravações de Eletroencefalograma (EEG) intracranial [36], análise de mecanismos de deglutição traqueal e de sons pulmonares [37, 38], Ciências da Terra [39], Finanças [40], redes de computadores [41] e avaliação de desempenho de algoritmos de predição de séries temporais [42]. Recentemente, mas ainda pouco explorada, tem sido aplicada em sinais de voz [43–45].

## 1.2 Objetivos da Pesquisa

Esta pesquisa é realizada tendo como principal objetivo a identificação de um conjunto de características, obtidas a partir de medidas relacionadas com a análise dinâmica não linear, que seja capaz de caracterizar e discriminar, de forma eficiente, sinais de vozes saudáveis e vozes afetadas por edema de Reinke, nódulos ou paralisia nas pregas vocais. Para tanto, é analisado o comportamento de sinais de voz por meio de medidas convencionais da análise dinâmica não linear (MNL) tais como: Expoentes de Lyapunov, Dimensão de Correlação, Primeiro mínimo da função de informação mútua, Expoente de Hurst, além de medidas de entropia (Tsallis, Shannon, Aproximada e Entropia de Correlação).

Os sinais também são analisados de forma quantitativa por meio das medidas de quantificação de recorrência (MQR), extraídas dos gráficos de recorrência, a saber: Taxa de recorrência, Determinismo, Entropia, Comprimento máximo das estruturas diagonais, Tendência, laminaridade, Comprimento médio das estruturas verticais e Comprimento máximo das estruturas verticais.

## 1.3 Organização do Trabalho

Os demais capítulos que compõem este trabalho estão estruturados como descrito a seguir. No Capítulo 2, são apresentados os conceitos básicos sobre o sistema de produção da voz. Em seguida, é contextualizada a produção da voz segundo a Teoria do Caos. Na sequência, é apresentado um quadro demonstrativo sobre o grau de incidência de patologias na laringe nos seres humanos, com destaque para edema de Reinke, nódulos e paralisia nas pregas vocais.

No Capítulo 3, são abordados os conceitos básicos referentes à análise não linear de séries temporais. A reconstrução da dinâmica do sistema é explicada a partir da técnica dos atrasos temporais e as invariantes do sistema são então descritas. Também são apresentadas outras medidas de análise no domínio do tempo, inclusive as medidas de quantificação de recorrência.

No Capítulo 4, são apresentados os resultados da caracterização acústica para os grupos de vozes analisados, utilizando as medidas da análise não linear (MNL) e as medidas de quantificação de recorrência (MQR). São apresentados, ainda, o potencial de cada uma dessas medidas, a partir da realização de testes estatísticos, na tarefa de discriminar sinais de vozes saudáveis e vozes afetadas por diferentes patologias na laringe.

No Capítulo 5, são apresentados os resultados obtidos na etapa de classificação dos sinais de voz. Inicialmente, são apresentados os resultados do processo de classificação utilizando as medidas dos conjuntos MNL e MQR, de forma individual e em seguida de forma combinada. Também são apresentados os resultados obtidos por meio do uso de vetores híbridos formados pela combinação de características dos conjuntos MNL e MQR e de coeficientes extraídos da análise de predição linear (LPC).

No Capítulo 6, são apresentados os resultados de vários trabalhos encontrados na literatura que utilizam a análise não linear, seja na caracterização dos sons da fala, seja na discriminação de vozes patológicas.

No Capítulo 7, são apresentadas as conclusões da pesquisa e sugestões para trabalhos futuros.

## CAPÍTULO 2

---

# Sistema de Produção da Voz e Patologias da Laringe

---

Voz, articulação e linguagem são os elementos mais importantes da produção da fala. Quando uma desordem associada a qualquer um desses elementos está presente, a habilidade para se comunicar pode se tornar comprometida. Voz é o elemento da fala que fornece ao locutor o sinal vibratório sobre o qual a fala é transportada [7]. Ela serve como a melodia da nossa fala e fornece informações de nossos sentimentos, expressão, intenção e estado de espírito para nossos pensamentos articulados diariamente. A voz, portanto, pode levar muita informação sobre a saúde geral e bem-estar de um indivíduo [8].

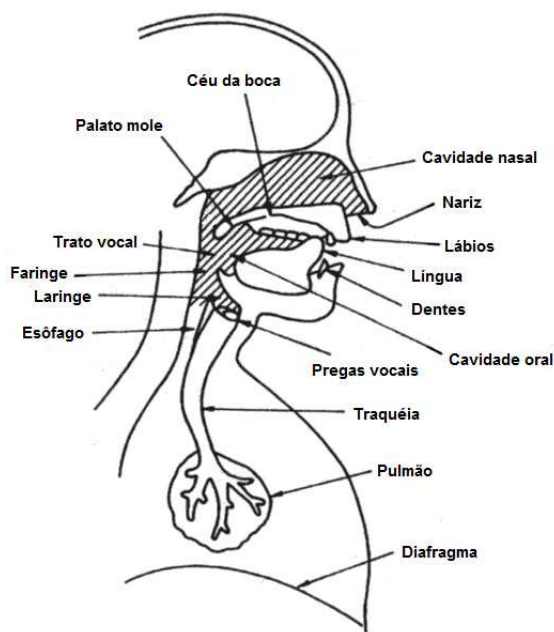
Neste capítulo é realizada uma breve explanação sobre a produção da voz, das causas das desordens vocais e das patologias da função vocal. Mais particularmente, são destacadas duas patologias de origem orgânica: edema de Reinke e nódulos vocais, e uma de origem neurológica: paralisia nas pregas vocais, e os efeitos que as mesmas causam na qualidade vocal.

### 2.1 Anatomia e Fisiologia do Sistema de Produção da Fala

O estudo da anatomia e fisiologia do mecanismo de produção da fala é de fundamental importância para a formação de modelos físicos e matemáticos que sirvam de base para a construção de sistemas de reconhecimento, síntese e codificação de voz. Assim como, permite compreender o processo da fonação normal e patológica e reconhecer o impacto de lesões específicas na produção vocal [4].

A fala é o resultado da ação de um conjunto de estruturas anatômicas que formam um sistema versátil e intrincado para produção de sons, cujas partes mais intimamente associadas à produção são os pulmões, a traqueia, a laringe, a faringe e as cavidades oral e nasal (Figura 2.1).

A fala é produzida a partir da liberação de ar dos pulmões para o trato vocal, formado basicamente por cavidades e órgãos articuladores que começa na abertura entre as pregas vocais e termina nos lábios. A principal função do trato vocal é modular a onda sonora que vem das



**Figura 2.1** Diagrama esquemático do mecanismo de produção da fala [1].

pregas vocais e promover constrictões para a geração de certos tipos de som. O som, que se origina na laringe como um tom fundamental, é modificado por várias câmaras de ressonância acima e abaixo desta, para, finalmente ser convertido em fala por ação da faringe, língua, palato, lábios e estruturas relacionadas.

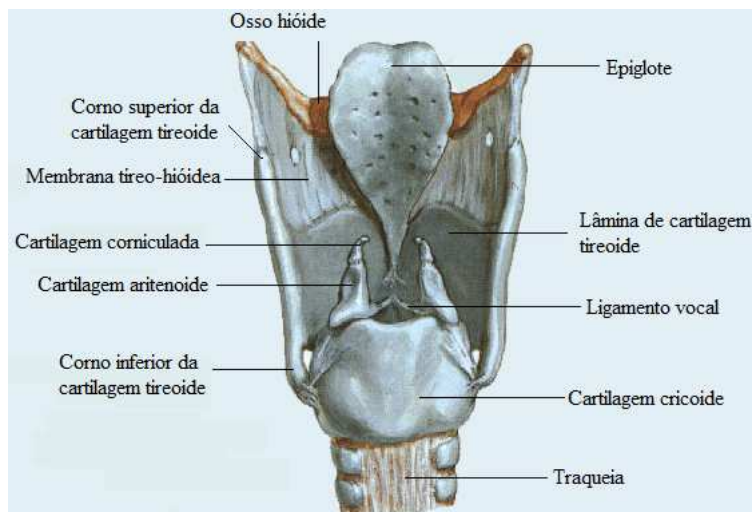
### 2.1.1 A Laringe

A laringe (Figura 2.2) situa-se no pescoço, sendo conectada inferiormente à traquéia e superiormente abre-se na faringe. O esqueleto da laringe é formado por cartilagens, músculos, membranas e mucosa<sup>1</sup>. A laringe como um todo divide-se em três espaços: supraglote, glote e infraglote. A glote é o espaço entre as pregas vocais. O som da voz é produzido na glote, sendo imediatamente acrescido de ressonância na própria supraglote [4].

A maior e mais importante cartilagem da laringe, em respeito à fonação, é a cartilagem tireóidea. Essa cartilagem possui o formato de um escudo, sendo composta de duas lâminas laterais, de forma quadrangular, e dois pares de cornos posteriores. O ângulo de união entre as lâminas direita e esquerda, chamado de proeminência laríngea, varia de acordo com o sexo. Na população masculina, observa-se um ângulo de aproximadamente 90° e na população feminina o ângulo é mais aberto, com cerca de 120°. Esta variação é responsável pela definição do tamanho das pregas vocais e contribui na definição da frequência vocal emitida [4].

<sup>1</sup>tipo de membrana umidificada por secreções glandulares, que recobre cavidades orgânicas em contato direto ou indireto com o meio exterior.

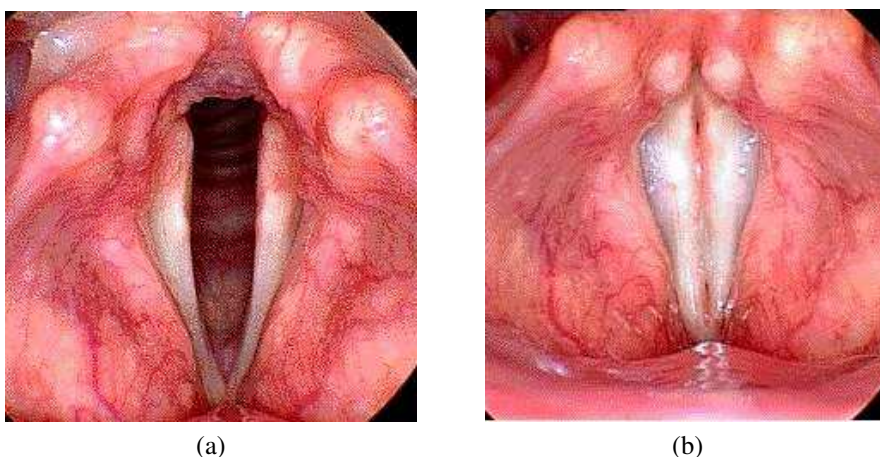




**Figura 2.2** Vista posterior da laringe [2].

As funções básicas da laringe, em ordem de importância são proteção, respiração e fonação. Na função de proteção a laringe atua como esfíncter<sup>2</sup> evitando a entrada de qualquer coisa, exceto o ar, ao pulmão. Na função de respiração, as pregas vocais abduzem ativamente durante o movimento respiratório, contribuindo para regular a troca gasosa com o pulmão e a manutenção do equilíbrio ácido-base. Na função de fonação, as mudanças de tensão e comprimento das pregas vocais, ampliação da abertura glótica e a intensidade do esforço respiratório provocam variações no tom da voz, tom esse que resulta da vibração das pregas vocais [46].

Na Figura 2.3 são ilustrados os processos de (a) abdução (afastamento) e (b) adução (fechamento) das pregas vocais (saudáveis).



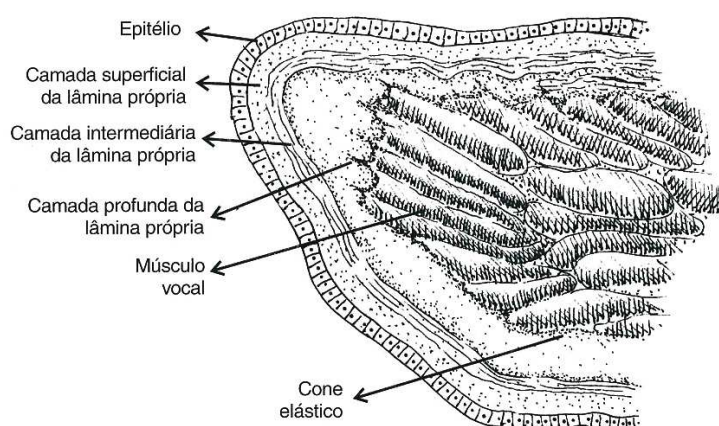
**Figura 2.3** Pregas vocais saudáveis (a) Abdução e (b) Adução [3].

<sup>2</sup>Estrutura muscular que contorna um orifício ou canal natural, permitindo sua abertura ou fechamento, podendo ser constituído de fibras musculares lisas e/ou estriadas.

## 2.1.2 Estrutura Anatômica da Prega Vocal

Segundo o modelo anatômico proposto por Hirano e Sato [47], conhecido como Modelo Corpo-Cobertura (Figura 2.4), as pregas vocais são duas dobras de músculo e mucosa que se estendem horizontalmente na laringe, fixando-se anteriormente na face interna da cartilagem tireóidea e posteriormente à cartilagem aritenóidea (Figura 2.2), cobrindo-a de mucosa. Para a função fonatória, o conceito mais importante é que as pregas vocais são uma estrutura multilaminada, em que cada camada apresenta propriedades mecânicas diferentes. De um modo geral, a prega vocal é composta de mucosa e músculo. A mucosa divide-se em epitélio e lâmina própria [4].

A lâmina própria possui três camadas baseadas nas propriedades mecânicas da prega vocal: superficial, intermediária e profunda, com diferença crescente em rigidez. O epitélio e a camada superficial da lâmina própria constituem a cobertura da prega vocal. As camadas intermediária e profunda formam a transição, o músculo vocal compõe o corpo. O epitélio é como uma fina cápsula vibratória capaz de conter a prega vocal durante a fonação. A camada superficial da lâmina própria é também conhecida como espaço de Reinke. Ela é uma camada frouxa e flexível e com poucos elementos, como se fosse uma gelatina, que vibra intensamente durante a fonação, deslocando-se acentuadamente [4].



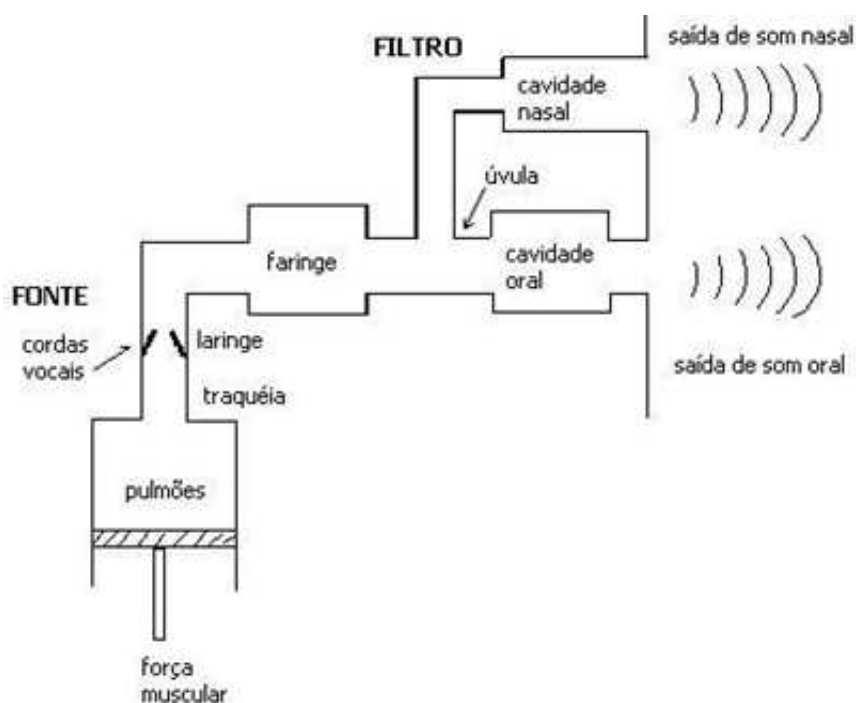
**Figura 2.4** Desenho esquemático da arquitetura histológica da prega vocal do adulto, em seção coronal [4].

A mucosa da prega vocal vibra em uma velocidade muito acelerada. A frequência da vibração da mucosa ocorre em cerca de 100 Hz no homem, enquanto que na mulher esta frequência é, em média, de 200 Hz [4]. Qualquer modificação (como algum traumatismo ou doença) que torne a mucosa mais rígida resultará em uma menor vibração da prega vocal e, consequentemente, dificuldades na fonação. O aumento de sua espessura ou uma maior flacidez do tecido também gera problemas nos movimentos das pregas vocais (como no edema de Reinke, por exemplo). Em termos patológicos, a lâmina própria (e em particular, a sua camada superficial) é o "lar" de lesões como nódulos e pólipos. Isso ocorre dada a proximidade da superfície das

pregas vocais verdadeiras onde exigências mecânicas extremas da vibração das pregas vocais acabam produzindo esses crescimentos orgânicos [25].

## 2.2 Mecanismo de Produção da Fala

O sistema de produção da fala pode ser pensado em termos de uma operação de filtragem acústica [1]. Na Figura 2.5 é apresentado um modelo simplificado de produção da fala, também denominado sistema fonte-filtro, em que as pregas vocais são consideradas a fonte sonora e o trato vocal, o filtro. Nesse modelo, as saídas produzem ondas acústicas que representam a fala. O ar é conduzido para fora dos pulmões pela traqueia, passando pela laringe, onde estão as pregas vocais.



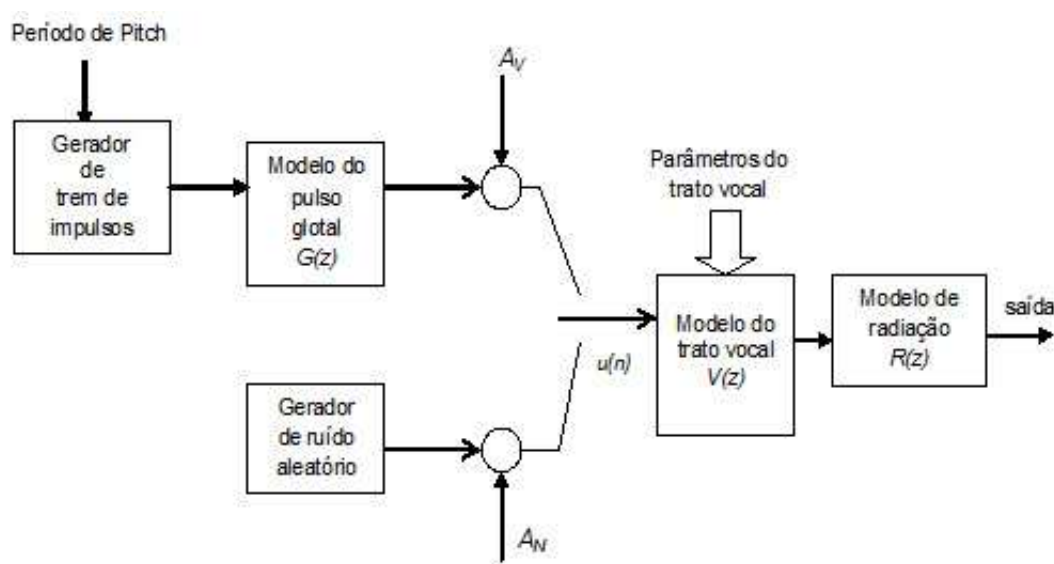
**Figura 2.5** Modelo simplificado do mecanismo de produção da fala (adaptado de [1]).

É no espaço compreendido entre as pregas vocais, ou glote, que o fluxo contínuo de ar dos pulmões é geralmente transformado em vibrações rápidas e audíveis durante a fala. Isso é feito pelo fechamento das pregas vocais, que causa um aumento gradativo da pressão atrás delas, que acaba por fazer com que elas se abram repentinamente, liberando a pressão, para então tornarem a se fechar. O intervalo de tempo entre sucessivas aberturas das pregas vocais é chamado de período fundamental (período de *pitch*),  $T_0$ , enquanto a taxa de vibração é chamada de frequência fundamental da fonação,  $F_0 = 1/T_0$ . Esse processo produz uma sequência de pulsos cuja frequência é controlada pela pressão do ar e pela tensão e comprimento das pregas vocais (frequência fundamental). Os sons assim produzidos são chamados de vozeados, ou

sonoros, que normalmente incluem as vogais; caso contrário, são chamados não-vozeados, ou surdos [15].

## 2.3 Modelo Linear Discreto no Tempo para a Produção da Fala

Na Figura 2.6 é mostrado um modelo linear, discreto no tempo, para a produção da fala. Esse modelo procura representar o processo de produção da fala baseado nas características do próprio sinal de voz. Apesar do fato de não apresentar qualquer hipótese de não linearidade no acoplamento dos subsistemas do modelo, este sistema produz razoável qualidade para propósitos de codificação da fala [1].



**Figura 2.6** Modelo linear discreto no tempo para a produção da fala (adaptado de [1]).

No modelo apresentado na Figura 2.6, o modelo do trato vocal  $V(z)$  e o modelo de radiação  $R(z)$  são excitados por um sinal glotal  $u(n)$ . Durante a produção de sons surdos, a fonte de excitação é modelada por um gerador de ruído aleatório com espectro plano e um controle de ganho  $A_n$ . Durante os períodos de atividade de voz sonora, a excitação usa uma estimativa do período de *pitch* para gerar um trem de impulsos na entrada de um filtro digital  $G(z)$  que simula o efeito de pulsos glotais, que são devidamente selecionados e aplicados ao trato vocal, após um controle de ganho  $A_v$  [19]. A função de transferência do trato vocal  $V(z)$  pode ser modelada como,

$$V(z) = \frac{H_0}{1 - \sum_{k=1}^N b_k z^{-k}}, \quad (2.1)$$

em que  $H_0$  representa um termo de ganho geral e  $b_k$  são chamados coeficientes do filtro.

Muito embora esse modelo seja inadequado para representar certas classes de fonemas - como por exemplo, aqueles que requerem mais de uma fonte de excitação, como é o caso dos sons fricativos sonoros /z/ e /v/ - ele é frequentemente usado para modelar todos os tipos de sons possíveis devido à coerência dos métodos analíticos que seguem esse modelo, e a resultados de desempenho satisfatórios em diversas aplicações como codificação, síntese e reconhecimento [1].

## 2.4 Patologias da Laringe

Alterações na voz podem ou não estar associadas a uma patologia orgânica ou física. A produção de uma voz variável e irregular pode ser a primeira manifestação de um distúrbio neurológico. Por outro lado, um problema orgânico sério, como câncer em fase inicial, pode não produzir qualquer alteração perceptível na qualidade da voz. Como exemplos de doenças por fatores neurológicos, que causam distúrbios na voz, podem ser citadas: doença cerebrovascular, paralisia cerebral, paralisia da prega vocal, refluxo, parkinsonismo e esclerose múltipla. Quanto às doenças orgânicas que causam distúrbios na voz estão: laringite crônica, pólipos de prega vocal, nódulos vocais e edema de Reinke [48].

Patologias na laringe incluem aquelas que causam qualquer alteração na estrutura histológica das pregas vocais. Mudanças nas camadas da mucosa ou no corpo do músculo da prega vocal afetam a massa, o tamanho, a rigidez, a flexibilidade e a tensão do mecanismo de vibração e podem afetar o padrão de fechamento glotal durante a fonação. Qualquer uma destas mudanças nas pregas vocais pode alterar a qualidade vocal, a frequência de vibração das pregas vocais (frequência fundamental ou seu correlato perceptual - *pitch*) e a intensidade.

A busca por métodos que proporcionem uma avaliação de desordens vocais causadas por patologias laríngeas é um assunto de grande interesse econômico e social. A comunidade científica procura, cada vez mais, proporcionar ferramentas de apoio ao diagnóstico médico de patologias laríngeas que sejam eficazes no pré-diagnóstico, bem como no acompanhamento de tratamentos medicamentosos e de terapias vocais. A detecção precoce da patologia pode aumentar significativamente a eficiência dos tratamentos médicos. Entretanto, a escolha das características adequadas que representem as desordens vocais associadas a patologia, ainda é um campo de estudo em aberto. Muito se tem investigado em discriminar vozes afetadas por patologias laríngeas de vozes saudáveis. Mas, muito pouco, ainda, em se tratando da identificação de parâmetros acústicos que melhor discriminem uma patologia específica dentre outras.

Métodos tradicionais de avaliação e diagnóstico de doenças da laringe são realizados, em geral, por meio da escuta da voz do paciente e da inspeção visual das pregas vocais por exames de videolaringoscopia, com ótica telescópica ou fibra ótica acoplados a um sistema de vídeo. Tais métodos, no entanto, requerem um profissional experiente e, em alguns casos, são considerados invasivos, causando certo desconforto ao paciente. Técnicas de processamento digital de sinais têm sido aplicadas nos últimos anos para o desenvolvimento de sistemas de

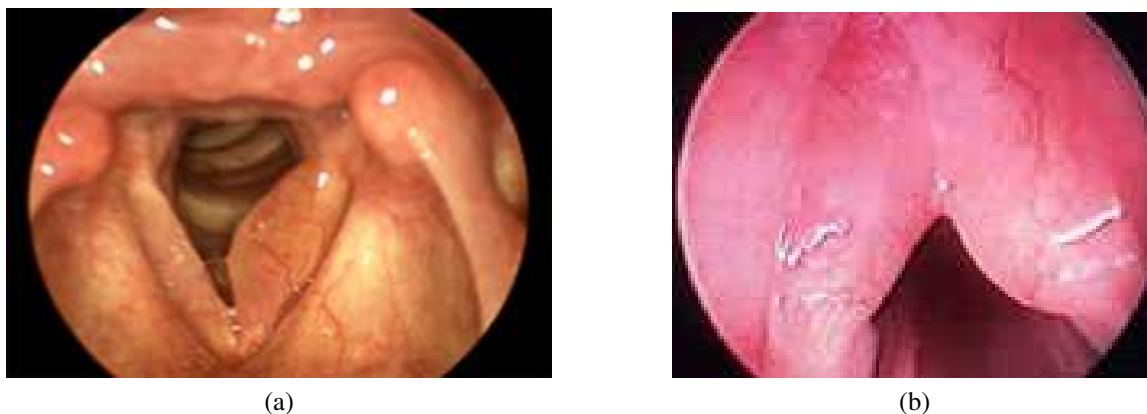
auxílio ao diagnóstico médico. A análise acústica pode ser usada como ferramenta não invasiva de apoio ao diagnóstico médico e acompanhamento de tratamentos e terapias vocais, de forma objetiva. Essa ferramenta não substitui os métodos tradicionais, mas pode auxiliar no diagnóstico, reduzindo a quantidade de exames videolaringoscópicos, podendo, ainda, ser efetuado à distância.

Entre as patologias da laringe, que afetam particularmente as pregas vocais, são consideradas nesta pesquisa duas de origem orgânica: edema de Reinke e nódulos, e uma de origem neurológica: paralisia nas pregas vocais.

### 2.4.1 Edema de Reinke

O edema de Reinke caracteriza-se pela expansão, aumento e inchaço das pregas vocais e pelo acúmulo de líquido ou material gelatinoso (ou ainda semi-sólido) na camada superficial da lâmina própria (espaço de Reinke) das pregas vocais. As pregas vocais ficam enormes com um edema claro e transparente, pobre em vasos e com o aspecto anatômico da região glótica completamente modificado [49], [50]. O edema inicial é translúcido, passando a ser mais consistente e avermelhado com o passar do tempo [4].

O edema de Reinke pode ser unilateral, quando confere à prega vocal uma imagem de fusão (edema fusiforme) ou bilateral, Figura 2.7a e Figura 2.7b, respectivamente. Com o crescimento progressivo do edema, pode haver uma obstrução completa da glote, com conseqüente asfixia, num estágio mais avançado [48].



**Figura 2.7** Edema de Reinke (a) unilateral e (b) bilateral [3], [5].

A incidência de edemas é maior nas mulheres que atingiram entre 50 e 70 anos de idade. Entretanto, encontra-se nos dois sexos e é uma doença de idade adulta. A causa irritativa é o abuso vocal e, especialmente, o tabagismo [48].

Quando o edema de Reinke é severo (Figura 2.8), as grandes bolsas de líquido podem oscilar para dentro e para fora, produzindo uma voz rouca, obstrutiva e roncada durante o sono [48]. No edema de Reinke de longa data, alguns pacientes sofrem de disfonia que, dife-

rente de qualquer tumor benigno de pregas vocais, é de um timbre mais baixo, como produzido por pregas vocais mais flácidas.



**Figura 2.8** Edema de Reinke severo [3].

Na análise acústica, o dado comumente observado é uma frequência fundamental mais baixa que a esperada para sexo e idade. Em um estudo realizado com sinais de voz da vogal sustentada /a/ com edema, foi observado que a frequência fundamental para um grupo de 32 mulheres com edema varia em torno de 160 Hz, enquanto para um grupo de 11 homens, varia em torno de 90 Hz. Para as vozes saudáveis, os valores femininos variam em torno de 240 Hz, enquanto os masculinos variam em torno de 140 Hz [19].

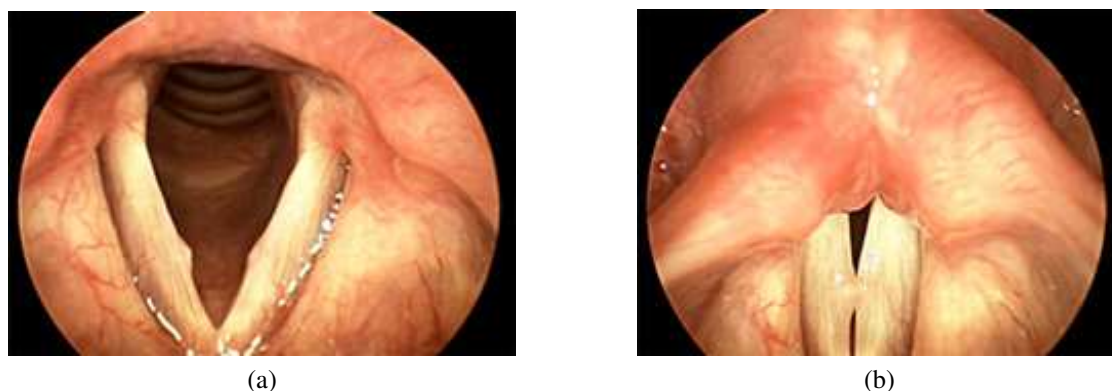
Os valores de *shimmer* (medida da variação na amplitude de um sinal de voz) apresentam-se geralmente bastante elevados, característica dos movimentos mais lentos da vibração das pregas vocais e do ruído espectral, devido à rouquidão. A proporção harmônico-ruído pode apresentar valores dentro da normalidade, porém a energia de ruído glótico está aumentada. Há uma variação no tempo de fonação, podendo ser mais longo, quando o edema é mais severo ou mais curto em casos de edemas discretos e muito flácidos [4].

#### 2.4.2 Nódulos nas Pregas Vocais

Os nódulos vocais são protuberâncias crônicas na extremidade livre e superfície inferior das pregas vocais. Os nódulos são sempre bilaterais e podem variar em tamanho, simetria e coloração. Eles representam uma degeneração inflamatória da camada superficial da lâmina própria com fibrose e edema associadas, que interferem na produção da voz [49], [48].

Na Figura 2.9 é apresentado o caso de uma paciente de 36 anos, não fumante, mas com histórico de abuso vocal. Pode-se observar na Fig. 2.9b o fechamento incompleto da glote devido à presença dos nódulos bilaterais.

Os nódulos vocais são as lesões benignas mais comuns das pregas vocais, tanto em crianças como em adultos. Nas crianças constituem 80% das causas de distúrbios da voz [48]. Nos adultos, são mais comuns em mulheres do que em homens, principalmente em profissionais que usam a voz em demasia. O comportamento vocal inadequado por mau uso e abuso vocal seria o principal motivo que poderia levar à formação de nódulos [4].



**Figura 2.9** Pregas vocais com nódulos: (a) abdução; (b) adução [3].

A massa aumentada dos nódulos nas pregas vocais contribui para uma altura de voz mais grave e maior periodicidade (julgada como rouquidão). Isso leva a um tipo de voz soprosa, monótona, que, muitas vezes, parece carecer de ressonância apropriada. As características acústicas dos nódulos apresentam *jitter* (medida da variação na frequência fundamental de um sinal de voz) e *shimmer* elevados. Outro fato são as evidências de ruído no espectro, de cuja intensidade depende a severidade da rouquidão e o tamanho da lesão [50].

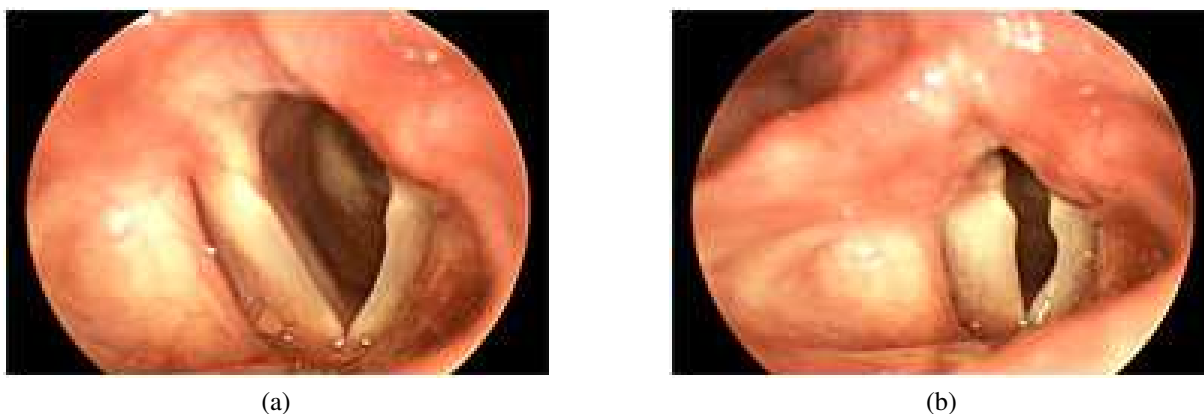
### 2.4.3 Paralisia nas Pregas Vocais

Paralisia das pregas vocais pode ser o resultado acidental de várias cirurgias comuns. Estas incluem o coração e as operações de pulmão. Paralisia da prega vocal não é necessariamente um sinal de que o nervo foi cortado. O nervo também pode parar de funcionar se esticado ou comprimido, e às vezes depois de um simples manuseio. Por esta razão uma das pregas vocais pode ser paralisada após mesmo as cirurgias mais simples. Finalmente, ramos do nervo também pode ser danificados pela colocação do tubo de respiração durante a anestesia geral [7].

A paralisia pode ser unilateral ou bilateral, podendo ser encontrada em lactantes, crianças e adultos. Podem ser ainda completas ou incompletas (paresia), aproximando ou afastando as pregas vocais numa maior ou menor amplitude [49]. Os sintomas perceptuais mais comuns da paralisia unilateral são a sopro e a rouquidão. Ocasionalmente, a diplofonia pode estar presente. A causa principal da paralisia bilateral das pregas vocais em adultos tem sido a tireoidectomia. Entre outras causas, podem ser citadas, ainda, doenças malignas do pescoço ou mediastino, pós-intubação endotraqueal, trauma cervical, doença neurológica ou alguma causa desconhecida.

Na Figura 2.10 está representada a prega vocal paralisada (paralisia unilateral direita). A prega vocal do lado direito da foto foi paralisada após uma cirurgia de tireóide. No lado direito da foto é mostrada a tentativa de fechamento das pregas durante a fonação (Figura 2.10b). No entanto, não ocorre um fechamento completo. Isso contribui para uma voz suave e sussurrada.





**Figura 2.10** Prega vocal direita paralisada: (a) abdução; (b) adução [3].

## 2.5 Evidências de Não Linearidade e Caos na Voz

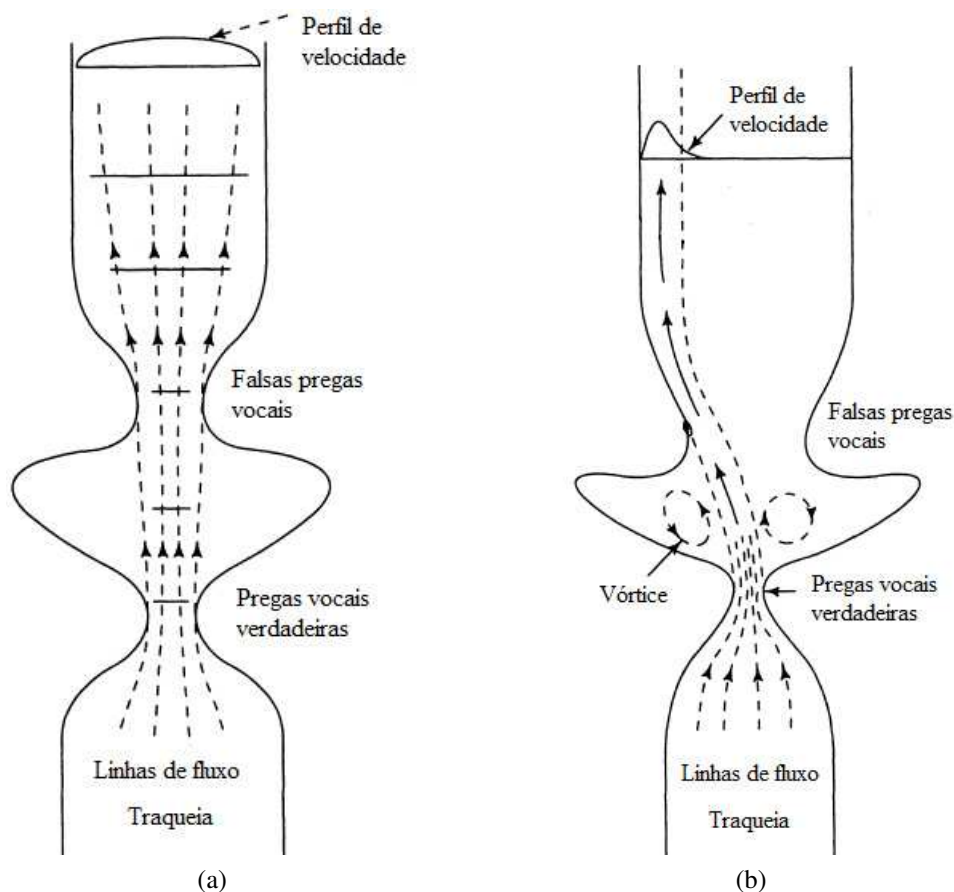
São diversas as teorias que tentam explicar o funcionamento do sistema de produção da fala, entre as quais, a Teoria Fonte-Filtro (Figuras 2.5 e 2.6) é a que tem sido mais investigada nas diversas áreas do processamento digital de sinais de voz. Nesse caso, o modelo é admitido ser a concatenação de subsistemas lineares (fonte, trato vocal e radiação nos lábios) sem perdas de acoplamento entre eles. Por exemplo, um único fonema sonoro, como uma vogal, pode ser representado como o produto de três funções de transferência, como mostrado a seguir:

$$S(\Omega) = U(\Omega)H(\Omega)R(\Omega), \quad (2.2)$$

em que  $U(\Omega)$  representa a excitação da fonte,  $H(\Omega)$  a dinâmica do trato vocal, e  $R(\Omega)$  os efeitos de radiação.

Outra suposição feita na abordagem clássica é que o som da voz se propaga como uma onda plana, em que é assumido que quando as pregas vocais se afastam, uma onda de pressão uniforme, que representa o som, é produzida e se propaga uniformemente através do trato vocal até os lábios (Figura 2.11a). Entretanto, estudos [25, 51–53] sugerem que uma dinâmica não linear de fluidos (Figura 2.11b) parece ser uma abordagem mais realística para caracterizar a propagação do som através do sistema vocal. Neste caso, os vórtices localizados nas falsas pregas vocais fornecem a excitação necessária da fonte durante a fase de fechamento das pregas vocais [1].

Apesar das condições restritivas adotadas pelo modelo fonte-filtro, tem-se que, para muitas aplicações do processamento digital de fala, como a codificação, síntese, e reconhecimento, um bom desempenho pode ser conseguido com um modelo que reflète as características gerais de padrões de tempo e articulatórios, bem como as propriedades de variação da frequência. Por outro lado, métodos mais detalhados de modelagem de produção de fala, com base em características físicas reais, são necessários para áreas como análise do movimento das pregas vocais, efeitos de uma patologia sobre a produção da fala, e outras áreas de aplicação [1].



**Figura 2.11** (a) Interpretação clássica da propagação do som através do sistema vocal. (b) Interpretação da dinâmica não linear de fluidos para a propagação do som ao longo do sistema vocal [1].

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 [54].

No caso particular da voz, diversos trabalhos têm identificado a influência das não linearidades presentes no sistema de produção vocal a partir da análise do próprio sinal de voz [45, 55]. Uma investigação acerca dos pressupostos de um sistema linear invariante no tempo sejam mantidos para todos os sons da fala, apesar das mudanças nos formantes (para diferentes vogais) ou na fonte de energia acústica (ruído acústico em consoantes e vibração das pregas vocais em vogais), foi realizada por Max A. Little [56]. Para tanto, ele utilizou uma técnica baseada na construção de séries sintéticas ou substitutas (*surrogate data*) [57].

As séries substitutas são originadas a partir das séries originais, de maneira que estas séries substitutas conservem algumas propriedades da série original, tais como a distribuição de probabilidade dos dados, espectro de potência, ou as duas características anteriores, mas perdem

características de não linearidade que possam estar presentes na série original [58]. O princípio que rege as séries sintéticas está baseado na geração de um certo número de realizações do sinal original que são especificamente projetadas para atender a uma dada hipótese pré-estabelecida.

No caso do trabalho de Little, a hipótese colocada foi a de que a série temporal original (sinal de voz) tratava-se de um processo estocástico gaussiano, linear e de média zero. Os testes foram aplicados em três diferentes classes de sinais de voz: vogais obtidas de indivíduos saudáveis, vogais sustentadas e consoantes de sujeitos com diferentes desordens vocais. Os sinais avaliados foram obtidos de duas bases de dados: a base de dados TIMIT [59] e base de dados da Kay Elemetrics [60] - a mesma utilizada nesta pesquisa. Desses dois bancos de dados, foram selecionados 28 sinais da base TIMIT e 26 sinais da base da Kay. Para cada um dos sinais utilizados foram construídas 19 séries substitutas. Os testes estatísticos foram realizados considerando o comportamento da função de informação mútua como discriminante estatístico e um nível de significância de 90%. 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.

Titze et al. [61] defendem a presença da teoria do caos na produção da voz. Segundo eles, a laringe, particularmente, as pregas vocais e o fluxo aéreo transglótico, representariam um sistema adequado à análise com base nessa teoria. Um sistema, para ser considerado caótico, deve apresentar as seguintes propriedades: ser um sistema não linear; determinístico; imprevisível e altamente sensível aos fatores de controle, o que significa que uma mínima alteração em suas condições iniciais pode provocar enormes mudanças nos resultados finais.

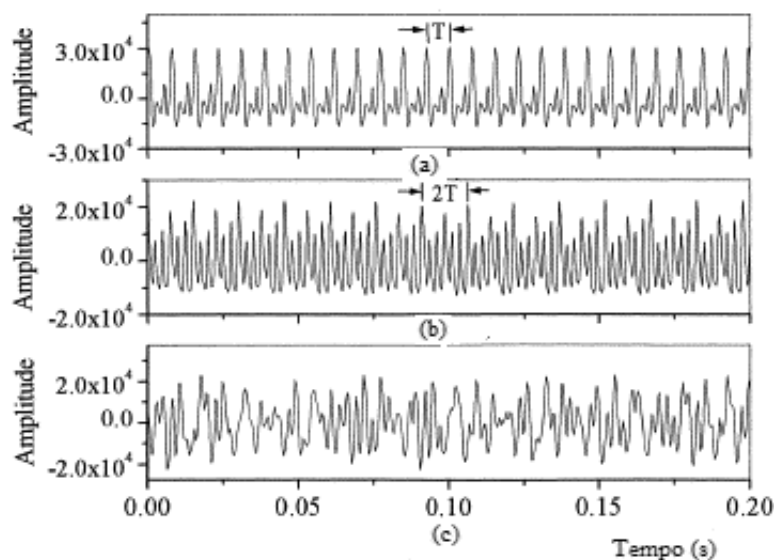
Ao longo das últimas duas décadas, observações em modelos computacionais para as pregas vocais, experimentos com laringes após cirurgia, e análise dinâmica não linear têm apontado para a existência de caos na produção da voz [6]. Como observado por vários pesquisadores, as não linearidades dos mecanismos da fonte da voz (por exemplo, a relação não linear do fluxo de pressão na glote, as curvas não lineares de esforço dos tecidos das pregas vocais e as não linearidades associadas com as colisões das pregas vocais) são fatores que sugerem a presença de caos na voz. Titze et al. [61] propuseram melhorar o entendimento das desordens vocais com conceitos e métodos de análise dinâmica. Desde então, pesquisadores têm aplicado essas novas ferramentas para estudar condições anormais associadas a patologias da laringe, diferenciar vozes saudáveis de patológicas, diagnosticar patologias e avaliar os efeitos dos tratamentos clínicos [6].

O modelo detalhado do processo de produção da fala deve levar em consideração, entre outros aspectos, a 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

associada à vibração das pregas vocais [26]. Alguns dos fatores responsáveis pelo processo de produção da fala são adequadamente modelados por um sistema que combina uma fonte de excitação periódica e outra de ruído branco, aplicadas a um filtro digital variante no tempo [15]. Entretanto, outros fatores são desconsiderados nesse modelo, pela simples dificuldade ou até impossibilidade de descrevê-los em termos de combinações de sinais, filtros digitais, ou equações diferenciais.

A teoria dos sistemas dinâmicos não lineares e a teoria do caos oferecem técnicas para a análise de sinais mesmo quando não se sabe, ou não é conhecido, o modelo detalhado do mecanismo de produção desses sinais. A análise por meio dessas teorias procura avaliar a dinâmica do sinal e, assumindo-se que tais amostras provêm de um sistema dinâmico não linear, medidas qualitativas podem ser obtidas desse sistema. Essas medidas não fornecem informações precisas quanto ao modelo do processo de produção do sinal avaliado, isto é, o modelo analítico é ainda inacessível. Entretanto, pode-se aferir a respeito de suas características qualitativas, como o número de graus de liberdade que esse modelo apresenta, a taxa de perda de informação do modelo relativamente às condições iniciais, ou a estabilidade local de sua trajetória no espaço de fases.

Segundo Titze [8] é possível classificar qualitativamente os sinais de voz em três tipos: tipo 1, tipo 2 e tipo 3 (Fig. 2.12). Os sinais considerados do primeiro tipo são quase-periódicos, os sinais do tipo 2 contêm fortes modulações ou sub-harmônicos, e os sinais do terceiro tipo são irregulares e aperiódicos.



**Figura 2.12** Formas de ondas para sinais de voz: (a) tipo 1; (b) tipo 2; e (c) tipo 3 [6].

Geralmente, a complexidade de um sinal de voz afeta a aplicabilidade de métodos de análise tradicionais. Tem sido sugerido que *jitter* e *shimmer*, por exemplo, duas medidas de perturbação acústica tradicionalmente usadas, são adequadas apenas para sinais do primeiro tipo (quase-periódicos) [6]. Medidas de perturbação tais como *jitter* e *shimmer* pressupõem,

por definição, quase-periodicidade, e assim sua utilidade pode não ser eficiente para sinais do segundo e do terceiro tipos. Estudos examinando a confiabilidade e aplicabilidade de *jitter* e *shimmer* têm encontrado grandes variações e pouca confiabilidade quando se analisa vozes desordenadas ou vozes aperiódicas e sugerem que métodos de análise de dinâmica não linear representam discriminadores mais úteis de vozes irregulares [6].

Métodos baseados na visão espectrográfica e na análise perceptual têm sido recomendadas para os sinais do tipo 2 e tipo 3, respectivamente. Embora a avaliação perceptual da função vocal seja simples, a inerente subjetividade dos julgamentos a torna potencialmente não confiável e também difícil de ser quantificada. Fatores que contribuem para as inconsistências da avaliação perceptual incluem diferenças nos padrões entre os indexadores (estimadores) de um sinal de voz e diferenças em como os ouvintes focalizam suas atenções sobre os vários aspectos da voz. Além disso, tanto a avaliação perceptual como a visão espectrográfica podem falhar em detectar quantitativamente sutis mudanças na função vocal. Devido à limitação dos métodos de análise tradicional, medidas objetivas complementares que analisem sinais vocais quase-periódicos e aperiódicos são desejáveis [6].

Métodos de dinâmica não linear podem analisar quantitativamente atividades irregulares ou caóticas. Pesquisas de atividades caóticas em sistemas fisiológicos sugerem que mudanças nas medidas dinâmicas não lineares podem indicar estados de disfunções patofisiológicas. Os resultados encontrados indicam que a teoria do caos e métodos de dinâmica não linear podem potencialmente ser aplicados para diagnosticar desordens fisiológicas e avaliar os efeitos dos tratamentos clínicos [62].

Os efeitos que as patologias laríngeas causam na qualidade vocal devem ser traduzidos nas características do sinal de voz a serem observadas na análise acústica para fins de diagnóstico e/ou de terapia vocal. As desordens vocais associadas às patologias devem ocasionar mudanças de comportamento nas características ou parâmetros temporais e espectrais do sinal de voz.

A abordagem empregada para a observação da voz e as desordens vocais associadas, deve extrair as características mais adequadas para representar a patologia. Uma característica pode servir para representar bem e discriminar uma voz saudável de uma voz patológica pode ser sensível para uma dada patologia e não ser para outra. Como também pode discriminar entre voz saudável e voz patológica, mas não conseguir separar uma patologia da outra. Não existe, ainda, na literatura, um consenso sobre qual(is) medida(s) é (são) a(s) mais adequada(s) para representar essa ou aquela patologia. Esse é ainda um campo em aberto e bastante promissor. Para a implementação de uma ferramenta de análise acústica eficiente, a escolha das características mais apropriadas para as patologias a serem pesquisadas, bem como a escolha do classificador pode influenciar de forma decisiva no resultado final de classificação obtido.

No próximo capítulo são abordados os fundamentos das técnicas baseadas na teoria do caos, além de outras medidas relacionadas aos sistemas dinâmicos não lineares, visando a sua aplicação na discriminação entre sinais de vozes saudáveis e vozes patológicas.

## CAPÍTULO 3

---

# Fundamentação Teórica

---

A análise dinâmica de sinais pode ser feita por meio de uma modelagem matemática ou por meio da análise de séries temporais. A primeira técnica implica na tentativa de se construir um modelo que possa ser usado para se obter informações úteis associadas à produção do fenômeno em estudo. Por outro lado, a análise de séries temporais considera simplesmente uma série temporal escalar, em geral, associada com uma aquisição experimental para entender o comportamento dinâmico do sistema. O ponto essencial dessa análise é que uma série temporal contém informações sobre variáveis não observáveis do sistema, o que permite a reconstrução do espaço de estados [63].

Segundo Kantz e Schreiber [64], a ligação mais direta entre a teoria do caos e o mundo real é a análise de dados de séries temporais, nos termos da dinâmica não linear. É devido, principalmente, ao desenvolvimento de uma base matemática coerente para a descrição do caos em dinâmicas determinísticas não lineares que se tem dado uma significativa atenção à análise dinâmica não linear e ao modelamento determinístico dos dados de uma série temporal [26, 65].

Uma série temporal é um conjunto discreto numerável de valores de uma variável de estado de um sistema dinâmico [66]. Algumas das dificuldades encontradas no estudo da dinâmica associada a séries temporais experimentais estão ligadas ao não conhecimento das equações que regem o sistema de produção dessas séries. A inevitável presença de ruído em sinais experimentais torna o seu estudo ainda mais difícil, podendo acarretar interpretações incorretas dos resultados. Por exemplo, caos determinístico, processos estocásticos e processos multi-periódicos com períodos incomensuráveis dão origem a séries temporais irregulares e aperiódicas que parecem, à primeira vista, caóticas [67]. Dessa forma, o uso de técnicas adequadas é extremamente importante para a análise de séries temporais.

Os processos regulares (periódicos, multi-periódicos ou quase-periódicos) podem ser identificados a partir de métodos tradicionais de análise de sinais experimentais, como por exemplo: a análise do espectro de potências e da função de autocorrelação. Por outro lado, a análise baseada nessas técnicas não permite, em geral, a distinção entre uma dinâmica caótica determinística e comportamento estocástico. Vários procedimentos têm sido desenvolvidos

com essa finalidade, inclusive nos casos em que não se sabe (ou não é possível) modelar ou descrever a dinâmica em termos de equações diferenciais ou mapas [67].

O comportamento complexo de uma série temporal, que pode inicialmente ser confundido como "aleatório" ou "ruído branco", em uma segunda análise, na verdade, pode ser determinístico. Embora este determinismo não leve a uma predição infinita, pode-se construir modelos determinísticos não lineares dessas séries temporais mais complexas e se fazer predições a curtos intervalos de tempo, as quais podem não ser possíveis com um modelo linear.

Quando se deseja determinar se a dinâmica que deu origem a uma série temporal é determinística (ou não), há de se lançar mão de outros métodos. Se for determinística, pode-se ter condições de desenvolver um modelo capaz de descrever a dinâmica. No caso de comportamento estocástico, ao contrário, trata-se de um processo a muitos graus de liberdade, caso em que só se pode esperar uma descrição estatística do sistema [67].

A resposta para a questão referente à distinção se uma série temporal é realmente aleatória (do ponto de vista da não-predição) ou se é aperiódica, mas caótica (e portanto preditiva) é dada pela estimação de invariantes tais como dimensão, entropia e expoentes de Lyapunov.

Enquanto a definição das invariantes dinâmicas refere-se ao sistema dinâmico em questão, suas estimativas a partir da série temporal tornam-se possíveis por meio de um conjunto de teoremas formulados por Takens [68]. Efetivamente, esses teoremas estabelecem que as invariantes dinâmicas estimadas das séries temporais observadas a partir da reconstrução de sua trajetória no espaço de estados usando coordenadas defasadas serão as mesmas do sistema dinâmico em análise sob certas condições.

### 3.1 Reconstrução do Atrator a partir de uma Série Temporal

De uma maneira geral, um experimento não mede todas as variáveis de estado do sistema e, usualmente, tem-se disponível a evolução no tempo de apenas uma variável observável, representado por uma série temporal,  $x(t)$ , que representa a trajetória de um dado sistema dinâmico. Dessa forma, é interessante analisar o sistema dinâmico a partir dessa série temporal o que é feito a partir de técnicas de reconstrução do espaço de estados [66].

A idéia básica da reconstrução do espaço de estados está calcada no fato de que a história temporal de um sinal contém informações sobre as variáveis de estado não observáveis que podem ser usadas para prever um estado presente. Toda a dinâmica do sistema está contida em uma série temporal associada a uma variável de estado [66].

Para que se possam analisar as propriedades de um possível atrator associado a uma série temporal com comportamento caótico determinístico é necessário em primeiro lugar reconstruir tal atrator num espaço de fases de dimensão adequada [67]. Este espaço reconstruído apresenta

uma suave variação de coordenada em relação ao espaço original, preservando os invariantes geométricos do sistema, tais como, a dimensão do atrator e os expoentes de Lyapunov [66].

Entre os métodos sugeridos para a reconstrução do espaço de estados, o método das coordenadas defasadas [68, 69] é o método citado na literatura como o mais adequado, e por isso, foi escolhido para ser utilizado neste trabalho.

### 3.1.1 Método das Coordenadas Defasadas

A técnica de reconstrução do espaço de estados é baseada no teorema da imersão de Takens [68]. Este teorema permite reconstruir um espaço de estado  $m$ -dimensional similar ao espaço de estado original,  $d$ -dimensional, a partir de uma única variável de estado, a variável medida [66]. Assim, vetores  $\vec{\xi}_i$   $m$ -dimensionais são reconstruídos a partir da série temporal  $\{x_i\}$ , em que  $x_i = x(t_i)$ ,  $i = 1, \dots, T$ , e de suas  $m-1$  versões defasadas no tempo,

$$\vec{\xi}_i = \{x(t_i), x(t_i + \tau), x(t_i + 2\tau), \dots, x(t_i + (m - 1)\tau)\}, \quad (3.1)$$

em que  $m$  é a chamada dimensão de imersão e  $\tau$  é o passo de reconstrução ou tempo de defasagem [67].

O método de Takens é também chamado método dos atrasos temporais e permite a reconstrução do atrator. Embora o atrator reconstruído não seja idêntico ao original, pode-se demonstrar que as propriedades topológicas são preservadas (difeomorfismo). Além disso, a série temporal de uma única variável é suficiente para a reconstrução desde que a dimensão de imersão seja suficientemente grande [68].

### 3.1.2 A Escolha do Passo de Reconstrução (tempo de defasagem)

Para um número infinito de pontos e na ausência de ruído, a escolha do passo de reconstrução  $\tau$  é, na grande maioria dos casos, arbitrária [68]. Entretanto, as séries temporais experimentais são finitas, usualmente contaminadas com ruído externo e obtidas com o uso de filtros. Nessa situação a reconstrução depende, e muito, da escolha correta do passo [67].

Se o passo  $\tau$  for muito pequeno,  $x(t)$  e  $x(t + \tau)$  terão praticamente o mesmo valor. Como consequência, um atrator reconstruído em um espaço bidimensional com um passo de reconstrução menor do que o adequado fica comprimido em torno da diagonal, já que  $\vec{\xi}_1 \approx \vec{\xi}_2$ , ou seja, esse atrator apresentará uma dependência linear entre as variáveis do sistema. Por outro lado, como a trajetória real está restrita a um volume finito do espaço de fases, o passo não pode ser muito grande, sob pena dos vetores reconstruídos serem completamente não-correlacionados cobrindo todo o espaço de fases [67].

O método baseado na função de informação mútua [70] é o método utilizado neste trabalho, por ser recomendado na literatura como o mais eficaz, para a determinação do tempo de defasagem ou passo de reconstrução.



## O Método da Informação Mútua Média

O método da informação mútua média foi proposto por A. M. Fraser e H. L. Swinney [70]. Segundo este método, ao se escolher o passo  $\tau \sim \tau'$ , em que  $\tau'$  é o intervalo de tempo que minimiza a informação mútua contida em vetores vizinhos ao longo de sua evolução, pode-se garantir a reconstrução de vetores com o menor nível de informação redundante (linearmente independentes), mas ainda correlacionados.

A teoria da informação procura identificar o quanto de informação se pode ter de uma medida realizada em um determinado instante de tempo  $t$ , quando se observa outra medida, do mesmo sinal, em um tempo posterior  $t + \tau$ . A informação mútua média fornece as mesmas informações que a função de autocorrelação fornece em sistemas lineares sendo, na realidade, um tipo de generalização para sistemas não lineares [66].

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 dada por

$$I(\tau) = \sum_{i=1}^b \sum_{j=1}^b P_{i,j}(x(t), x(t + \tau)) \log_2 \left[ \frac{P_{i,j}(x(t), x(t + \tau))}{P_i(x(t))P_j(x(t + \tau))} \right], \quad (3.2)$$

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,  $x(t)$  esteja no  $i$ -ésimo intervalo e  $x(t + \tau)$  no  $j$ -ésimo intervalo [64].

Tem-se ainda que  $I(t) \geq 0$ . Se  $x(t)$  e  $x(t + \tau)$  são iguais, então  $I(\tau)$  é máximo. Por outro lado, se  $x(t)$  e  $x(t + \tau)$  são completamente independentes, então o argumento do termo logarítmico é a unidade e  $I(t) = 0$ .

Fraser e Swinney [70] estabelecem que o valor do tempo de defasagem ideal corresponde ao primeiro mínimo local, quando este existir, da função de informação mútua média  $I(\tau)$ . Desta forma, ao traçar  $I(\tau)$  versus  $\tau$ , a melhor defasagem corresponde ao valor de  $\tau$  no primeiro mínimo local. Com isso, garante-se que o espaço de estados reconstruído seja topologicamente equivalente ao espaço de estados do sistema dinâmico original [66].

### 3.1.3 A Escolha da Dimensão de Imersão

A dimensão  $m$  do espaço de fases reconstruído não é necessariamente idêntica à dimensão  $d$  do espaço de fases real dos vetores que representam a dinâmica do sistema físico. Em geral é necessário reconstruir-se o atrator em espaços de fases com dimensão suficientemente elevada ( $m \geq 2D_0 + 1$ , onde  $D_0$  é a dimensão de *Hausdorff* [71] do atrator) para que se tenha segurança com relação aos resultados; caso contrário, o atrator aparecerá dobrado sobre si mesmo como numa projeção. Desta forma, pontos inicialmente distantes tornam-se próximos,

fazendo com que a estatística da medida invariante associada ao atrator reconstruído seja distorcida. Muitas vezes, porém, dependendo das características do atrator, uma reconstrução num espaço de dimensão bem menor que  $2D_0 + 1$  já é suficiente para revelar a sua estrutura métrica. Como não se sabe *a priori* qual a dimensão de *Hausdorff* do atrator associado a uma série de dados experimentais, uma razoável dose de experimentação e de bom senso, torna-se necessária ao interpretar-se os resultados obtidos [67].

Dentre os métodos utilizados para a determinação da dimensão mínima de imersão, o Método das Falsas Vizinhanças ou Método dos Falsos Vizinhos Próximos [72] é o método que oferece os melhores resultados e, por isso, tem sido o mais utilizado na literatura, inclusive neste trabalho.

### Método dos Falsos Vizinhos Próximos

O método dos Falsos Vizinhos Próximos sugerido por Kennel et al. [72], foi desenvolvido baseado na busca de uma dimensão mínima onde não ocorra cruzamento da órbita consigo mesma. Assim, a partir da projeção do sistema em uma determinada dimensão, é verificado se um vizinho é "verdadeiro" ou "falso", comparando com uma projeção em uma dimensão maior. Desta forma, um falso vizinho é um ponto do sinal que só corresponde a um vizinho devido a observação das órbitas em um espaço muito pequeno,  $d < m$ . Quando o espaço está imerso em uma dimensão  $d \geq m$ , todos os pontos vizinhos de todas as órbitas são vizinhos verdadeiros [66].

Sendo assim, é proposto, como método para determinar a dimensão mínima necessária para imergir um atrator, contar o número de falsos vizinhos para cada um dos seus pontos. Quando o número de falsos vizinhos cai a zero, o atrator terá sido suficientemente "desdobrado" e é possível, portanto, identificar a menor dimensão de imersão capaz de representá-lo adequadamente.

## 3.2 Medidas da Análise Dinâmica Não Linear

São as seguintes as medidas da análise dinâmica não linear utilizadas neste trabalho: a dimensão de correlação, a entropia de correlação e o maior expoente de Lyapunov, além da entropia de Shannon, a entropia de Tsallis, a entropia aproximada, o valor do primeiro mínimo da função de informação mútua e o expoente de Hurst.

### 3.2.1 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 [6].

Um sistema com um expoente de Lyapunov positivo apresenta instabilidade local e extrema sensibilidade às condições iniciais. Em outras palavras, qualquer mudança no estado do sistema será exponencialmente amplificada, o que resultará em uma saída totalmente diferente.

Quando o sistema dinâmico possui um modelo matemático estabelecido que permita a sua linearização em torno de uma determinada trajetória, os expoentes de Lyapunov podem ser calculados com precisão a partir do algoritmo de Wolf et al. [73]. Por outro lado, a determinação desses componentes a partir de um sinal experimental é bem mais complicada. Uma das dificuldades no cálculo dos expoentes de Lyapunov associados às séries temporais experimentais reside no desconhecimento das matrizes Jacobianas ou derivadas associadas à dinâmica [66].

Entre os métodos existentes para a determinação dos expoentes de Lyapunov a partir de uma série temporal o mais utilizado é conhecido como o método das trajetórias ou método direto [73]. A idéia básica do método das trajetórias está relacionada com a evolução da distância, no espaço tangente, de duas trajetórias inicialmente muito próximas.

O algoritmo proposto por Kantz [74] para a determinação do maior expoente de Lyapunov está baseado no método das trajetórias e avalia a seguinte quantidade [75]:

$$S(\tau) = \frac{1}{N} \sum_{i=1}^N \ln \left( \sum_{j=1}^M d(\vec{\xi}_i, \vec{\xi}_j; \tau) \right), \quad (3.3)$$

em que  $S(\tau)$  é chamado de fator de crescimento (*stretching factor*),  $N$  é o número de pontos no espaço de fase,  $\vec{\xi}_i$  é um ponto de referência e  $\vec{\xi}_j$  é um ponto vizinho de  $\vec{\xi}_i$ ,  $M$  é o número de vizinhos próximos,  $\tau$  é o índice relativo ao tempo e  $d(\vec{\xi}_i, \vec{\xi}_j; \tau)$  é a distância euclidiana entre os pontos  $\vec{\xi}_i$  e  $\vec{\xi}_j$ .

Se a série temporal sob análise é produzida a partir de um sistema dinâmico não linear apresentando um expoente de Lyapunov positivo, haverá uma inclinação positiva da função  $S(\tau)$  que corresponderá ao máximo expoente de Lyapunov [75].

### 3.2.2 Dimensão Fractal

De uma maneira geral, sistemas dinâmicos que exibem comportamento caótico apresentam trajetórias no espaço de estados que convergem para um atrator estranho. Hausdorff [71] forneceu uma definição rigorosa de dimensão fractal, que corresponde a uma propriedade básica de um atrator. A estranheza de um atrator caótico está associada com sua dimensão fractal, como uma alusão ao termo fracionário, caracterizando uma dimensão não inteira. Pode-se dizer que a dimensão fornece o valor da informação necessária para especificar a posição de um ponto no atrator com certa precisão.

A dimensão é, também, um limite inferior do número de variáveis necessárias para descrever a dinâmica do sistema [66]. Assim, quanto mais complexo for o sistema, maior será o valor de dimensão do mesmo, o que significa que mais variáveis de estado são necessárias para descrever sua dinâmica [76].

Existem muitas maneiras de se quantificar ou definir a dimensão fractal. Farmer et al. [77] definem a dimensão basicamente em dois grupos. Um grupo depende somente de propriedades geométricas, chamado de dimensão fractal ou dimensão métrica, e outro grupo, que depende não só de propriedades geométricas, mas de propriedades probabilísticas, chamado de dimensão da medida natural ou dimensão probabilística. As propriedades complexas dos atratores estranhos tornam a determinação da dimensão do atrator um problema não trivial.

Uma definição geométrica intuitiva para a dimensão pode ser entendida como sendo o expoente que expressa a escala ou graduação de uma *grandeza* (volume, massa, medida de informação) com o seu *tamanho* (distância):  $grandeza \propto tamanho^D$ . Assim, a definição de dimensão é usualmente dada como uma equação da forma [66]:

$$D = \lim_{tamanho \rightarrow 0} \frac{\log(grandeza)}{\log(tamanho)}, \quad (3.4)$$

em que o limite para  $tamanho \rightarrow 0$  é feito para garantir uma invariância sobre uma suave mudança de coordenadas. O mesmo também implica que a dimensão seja uma quantidade local, fazendo com que definições globais para a dimensão requeiram algum tipo de média.

As variações da grandeza e do tamanho caracterizam os diferentes medidores de dimensão. A dimensão de Hausdorff e a de capacidade são exemplos de dimensões métricas, enquanto a de correlação é um exemplo de dimensões probabilísticas.

### 3.2.3 Dimensão de Capacidade e Dimensão de Hausdorff

A dimensão de capacidade ou método das caixas (ou ainda, de contagem das caixas) foi originalmente definida por Kolmogorov e corresponde a uma das formas mais simples de se determinar a dimensão do atrator. Este método consiste em determinar quantos volumes elementares ou caixas, de tamanho  $\varepsilon$ , são necessários para cobrir todo o conjunto de pontos do atrator no espaço de estados [66].

De uma maneira geral, para um atrator de dimensão  $D$  tem-se que o número de volumes elementares necessários para conter todos os pontos é inversamente proporcional a  $\varepsilon^D$ , isto é,

$$N(\varepsilon) \propto \frac{1}{\varepsilon^D}, \quad (3.5)$$

em que  $N(\varepsilon)$  é o número de volumes cheios.

A relação dada pela Equação 3.5 é válida para a maioria dos conjuntos fractais, exceto se  $D$  for menor do que 1 [78]. Por analogia com os casos clássicos, pode-se interpretar  $D$  como uma dimensão, geralmente chamada de dimensão de capacidade ou dimensão de caixa do atrator. Uma definição equivalente é:

$$D = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}, \quad (3.6)$$

a dimensão  $D$  é então determinada pela inclinação da curva formada por  $N(\varepsilon)$  versus  $\varepsilon$  em um gráfico log-log.

A dimensão de capacidade pode ser vista como uma simplificação da dimensão de Hausdorff,  $D_0$ , que exige que  $N(\varepsilon)$  seja o número mínimo de caixas de lado  $\varepsilon$  necessário para cobrir todo o atrator, ou seja, há de se escolher o modo particular pelo qual se dividirá o espaço de fases em caixas.

Na prática a otimização requerida pela dimensão de Hausdorff é inexequível na maioria dos casos, apresentando certa dificuldade de implementação [67]. Farmer et. al. [77] mostraram que a dimensão de capacidade e a dimensão de Hausdorff são geralmente iguais, mas rigorosamente  $D \geq D_0$ .

### 3.2.4 Dimensão de Correlação

A dimensão de correlação,  $D_2$ , corresponde a um tipo de dimensão probabilística e representa uma das formas mais populares de se calcular a dimensão do atrator. Muitos atratores estranhos não são homogêneos, de maneira que algumas de suas regiões são mais visitadas do que outras. As dimensões do tipo probabilística levam em consideração as não-homogeneidades do atrator e se baseiam na frequência relativa  $f_i$  com que cada caixa  $i$  é visitada. A frequência  $f_i$  é definida como

$$f_i = \lim_{N \rightarrow \infty} \frac{N_i}{N}, \quad (3.7)$$

sendo  $N$  o número total de pontos que pertencem ao atrator e  $N_i$  o número de pontos na  $i$ -ésima caixa.

A dimensão de correlação mede a probabilidade de se encontrar um par aleatório de pontos dentro de um determinado volume elementar e é definida da seguinte forma,

$$D_2 = \lim_{\varepsilon \rightarrow 0} \frac{\log \sum_{i=1}^{N(\varepsilon)} (f_i)^2}{\log(\varepsilon)}, \quad (3.8)$$

em que o termo  $(f_i)^2$  representa a frequência relativa com que dois pontos de um atrator estejam dentro da  $i$ -ésima caixa de tamanho  $\varepsilon$  [79].

A partir da estimativa da dimensão de correlação, o caos determinístico pode ser distinguido do ruído branco aleatório: o cálculo de  $D_2$  do ruído branco não converge com o aumento da dimensão de imersão  $m$ , por outro lado, a estimativa de  $D_2$  para um sistema caótico converge para um valor finito [6].

Grassberger e Procaccia [80] 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  $\varepsilon$

é aproximada pela probabilidade de que a distância entre dois pontos seja menor que  $\varepsilon$ . Esse algoritmo fornece um limite inferior e um valor aproximado para a dimensão fractal [67].

Seja  $q_i(\varepsilon)$  a fração de pontos do atrator que está dentro de uma hiper-esfera de raio  $\varepsilon$ , centrada no ponto  $\vec{\xi}_i$  ( $i$ -ésimo vetor de imersão  $m$ -dimensional). Essa fração é expressa pela relação:

$$q_i(\varepsilon) = \frac{1}{N} \sum_{j=1}^N \Theta(\varepsilon - \|\vec{\xi}_i - \vec{\xi}_j\|), \quad (3.9)$$

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  $C(\varepsilon)$  como:

$$C(\varepsilon) = \frac{1}{N} \lim_{N \rightarrow \infty} \sum_{i=1}^N q_i(\varepsilon), \quad (3.10)$$

ou ainda [67],

$$C(\varepsilon) = \frac{1}{N^2} \lim_{N \rightarrow \infty} \sum_{i=1}^N \sum_{j=1}^N \Theta(\varepsilon - \|\vec{\xi}_i - \vec{\xi}_j\|). \quad (3.11)$$

Para regiões de escala suficientemente pequenas e quando a dimensão de imersão  $m$  é maior do que a dimensão de capacidade do atrator, tem-se que

$$C(\varepsilon) \propto \varepsilon^{D_2}. \quad (3.12)$$

Uma vez que não se conhece a dimensão de capacidade do atrator *a priori*, o valor da dimensão de correlação é estimado numa região de convergência dos valores de  $D_2$  quando se varia  $m$  (dimensão de imersão). A dimensão de correlação é então calculada, para cada valor de  $m$ , como

$$D_2 \simeq \lim_{\varepsilon \rightarrow 0} \frac{\log C(\varepsilon)}{\log(\varepsilon)}, \quad (3.13)$$

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

### 3.2.5 Entropia de Shannon

Numa primeira aproximação, a entropia pode ser associada à desordem e à ausência de regulação do comportamento de determinado sistema. Essa desordem é proporcional à quantidade de estados diferentes do sistema. Assim, a entropia é crescente à medida que a quantidade de trajetórias possíveis cresce com a complexidade do sistema.

Dado um determinado sinal, a entropia de Shannon é definida como [81]

$$H_1 = - \sum_{i=1}^M p_i \log_2(p_i), \quad (3.14)$$

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  $H_1$  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 [82].

### 3.2.6 Entropia de Rényi

A entropia de Rényi, uma generalização da entropia de Shannon, é uma família de funcionais para quantificação da diversidade, incerteza ou aleatoriedade de um sistema.

A entropia de Rényi de ordem  $\alpha$ , para  $\alpha > 0$  é definida como [83]

$$I^{(\alpha)} = \frac{1}{1 - \alpha} \log \left( \sum_{i=1}^M p_i^\alpha \right), \quad (3.15)$$

em que  $p_i$  é a probabilidade do evento  $i$ .

### 3.2.7 Entropia de Kolmogorov-Sinai

Uma entropia de Kolmogorov-Sinai não nula, positiva e finita é condição suficiente para a ocorrência de caos determinístico [6, 67]. Como consequência, a estimativa dessa entropia é importante para se determinar se o sistema dinâmico estudado apresenta dinâmica caótica.

A entropia de Kolmogorov-Sinai ( $K$ ) é definida como a taxa média de criação de informação no sistema (ou a taxa média de perda de informação do observador), isto é,

$$K = - \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{m \rightarrow \infty} \frac{1}{m\tau} \sum_{i_1, \dots, i_m} p(i_1, \dots, i_m) \ln p(i_1, \dots, i_m), \quad (3.16)$$

em que  $p(i_1, \dots, i_m)$  é a probabilidade conjunta de se observar o sistema na caixa  $i_1$  em  $t = 0$ ,  $i_2$  em  $t = \tau \dots$  e  $i_m$  em  $t = m\tau$  quando se divide o espaço de fases em caixas de hiper-volume  $\varepsilon^m$  [67].

Num processo regular, pontos inicialmente próximos continuam assim ao longo da trajetória engendrada pela dinâmica associada. Nenhuma informação adicional é produzida no sistema e a quantidade de informação do observador não é alterada. Consequentemente deve-se esperar uma entropia associada nula, isto é,  $K_{regular} = 0$ .

Num processo aleatório (estocástico), ao contrário, pontos inicialmente próximos são distribuídos com igual probabilidade para qualquer nova posição consistente com os vínculos

do sistema. Assim, existe geração de informação assintoticamente infinita no sistema, ou seja,  $K_{estocastico} \rightarrow \infty$ .

Para processos caóticos, devido à dependência as condições iniciais, pontos inicialmente adjacentes separam-se exponencialmente com a evolução temporal. Existe geração de informação no sistema e a entropia é finita e positiva, isto é,  $0 < K < \sum \lambda_i$ , em que os  $\lambda_i$  são os expoentes de Lyapunov positivos.

### 3.2.8 Entropia de Correlação

As primeiras tentativas de cálculo da entropia de Kolmogorov-Sinai ( $K$ ) baseavam-se na sua definição, que, além da difícil implementação, apresentavam resultados muitas vezes insatisfatórios [67]. A abordagem mais adequada para se estimar a entropia de Kolmogorov-Sinai foi proposta por Grassberger e Procaccia [80] e baseia-se no cálculo da integral de correlação  $C(\varepsilon)$  (Equação 3.11).

A partir da Equação 3.16 é possível generalizar o conceito de entropia, definindo as entropias generalizadas de Rényi de ordem  $q$  como [67]

$$K_q = - \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{m \rightarrow \infty} \frac{1}{m\tau} \frac{1}{q-1} \ln \sum_{i_1, \dots, i_m} [p(i_1, \dots, i_m)]^q. \quad (3.17)$$

Pode-se demonstrar que  $K_1 \equiv K$  e que vale a desigualdade  $K_{q'} \leq K_q$  se  $q < q'$ . Em particular  $K_2 \leq K$  e  $K_2$ , conhecida como entropia de segunda ordem ou entropia de correlação, fornece um limite inferior para a entropia de Kolmogorov-Sinai.

O procedimento de Grassberger-Procaccia permite estimar  $K_2$  a partir das integrais de correlação  $C(\varepsilon)$  (Equação 3.11) calculadas para dimensões de imersão  $m$  sucessivas, ou seja,  $C_m(\varepsilon)$  e  $C_{m+1}(\varepsilon)$ , da seguinte forma:

$$K_2(\varepsilon) = \frac{1}{\tau} \ln \left[ \frac{C_m(\varepsilon)}{C_{m+1}(\varepsilon)} \right], \quad (3.18)$$

em que  $\tau$  é o passo utilizado na reconstrução [67].

Para se estimar  $K_2$ , calcula-se o valor médio da Equação 3.18 para diversos valores de  $\varepsilon$  (pequenos) na região linear dos gráficos de  $\ln C_m(\varepsilon) \times \ln(\varepsilon)$ . À medida que  $m$  cresce esse valor médio converge para  $K_2$ . Uma outra possibilidade é traçar o gráfico da Equação 3.18 em função de  $m$  para diferentes valores de  $\varepsilon$ ; observa-se o comportamento assintótico para  $m$  grande e  $\varepsilon$  pequeno.



### 3.2.9 Entropia Aproximada

A entropia aproximada,  $\text{ApEn}$ , é uma medida da informação condicional média gerada por pontos divergentes numa trajetória no espaço de fase [84]. Considerando valores fixos para  $m$  e  $\varepsilon$  a entropia aproximada é definida como [85]:

$$\text{ApEn}(m, \varepsilon) = [\Phi^m(\varepsilon) - \Phi^{m+1}(\varepsilon)], e \quad (3.19)$$

$$\Phi^m(\varepsilon) = \frac{1}{N} \sum_{i=1}^N \log q_i(\varepsilon), \quad (3.20)$$

em que  $q_i(\varepsilon)$  é definida pela Equação 3.9 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 faculta a aplicação da  $\text{ApEn}$  em uma variedade de contextos, incluindo processos determinísticos, caóticos e estocásticos [85]. Embora a medida da entropia aproximada não seja uma medida de caos, ela quantifica a regularidade do processo de imersão em séries temporais [82].

### 3.2.10 Entropia de Tsallis

Em meados de 1870, ao associar uma variável macroscópica (entropia) a conceitos microscópicos, Boltzmann deu início a formulação da Mecânica Estatística. Segundo Boltzmann, a medida da desordem  $B$  na distribuição dos estados de um sistema sob os microestados permitidos pode ser estimada como [86]:

$$B = \beta \ln W, \quad (3.21)$$

em que  $\beta$  é conhecida como a constante de Boltzmann e  $W$  é o número total de estados microscópicos relativos ao estado macroscópico do sistema. Nesta definição de Boltzmann, a entropia é vista como o grau de desordem de um sistema. Na medida em que o número de microestados possíveis do sistema aumenta, a sua entropia também aumenta indicando que existem mais configurações possíveis para aquele sistema e que as moléculas deste sistema estão mais dispersas (mais desorganizadas) [58].

A entropia de Tsallis é uma generalização da entropia de Boltzmann e é definida como [87]:

$$H_q = k_1 \frac{1 - \sum_{i=1}^W p_i^q}{q - 1}, \quad (3.22)$$

em que  $q \in \mathbb{R}$  caracteriza a generalização,  $k_1$  é uma constante positiva que define a unidade em que a entropia é medida e  $\{p_i\}$  são as probabilidades associadas com as  $W$  ( $W \in \mathbb{N}$ )

configurações possíveis do sistema. A entropia de Boltzmann é encontrada quando, no limite,  $q \rightarrow 1$  [88].

A principal característica da entropia de Tsallis é a sua não aditividade [88]. 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 [58].

### 3.2.11 Expoente de Hurst

O expoente ou parâmetro de Hurst ( $H$ ), também denominado de grau de dependência temporal ou de escala, de um processo estocástico é definido pela taxa de decaimento da função coeficiente de auto-correlação  $\rho(k)$  quando  $k \rightarrow \infty$ . Esta medida foi identificada e definida por Harold Edwin Hurst [89] segundo observações de fenômenos da natureza realizadas durante seu trabalho de dimensionamento de reservatórios de água no rio Nilo. Esta medida foi posteriormente designada na literatura como parâmetro de Hurst.

Considerando um sinal de voz representado pelo processo estocástico  $X(t)$  com variância finita e função autocorrelação definida por

$$\rho(k) = \frac{Cov[X(t), X(t+k)]}{Var[X(t)]}, \quad k = 0, 1, 2, \dots \quad (3.23)$$

em que  $\rho(k)$  toma valores na faixa  $[-1, 1]$  e  $\lim_{k \rightarrow \infty} \rho(k) = 0$ . O comportamento assintótico da função  $\rho(k)$  é dado por

$$\rho(k) \sim H(2H - 1)k^{2(H-1)} \quad (3.24)$$

que é uma função de variação lenta no infinito e  $H$  ( $0 < H < 1$ ) é o expoente da função autocorrelação [90]. Os processos estocásticos podem ser classificados segundo o valor de  $H$  como,

- Processos anti-persistentes -  $0 < H < \frac{1}{2}$ . Nestes processos, a função autocorrelação é negativa, em particular,  $\sum_{k=-\infty}^{\infty} \rho(k) = 0$ .
- Processos com dependência ou memória de curto alcance (*Short-Range-Dependence* - SRD<sup>1</sup>) -  $H = \frac{1}{2}$ . A função autocorrelação  $\rho(k)$  apresenta decaimento exponencial para zero tal que  $\sum_{k=-\infty}^{\infty} \rho(k) = c$ , onde  $c > 0$  é uma constante finita.
- Processos com dependência ou memória de longo alcance. (*Long-Range-Dependence* - LRD) -  $\frac{1}{2} < H < 1$ . A função  $\rho(k)$  apresenta decaimento lento indicando dependência mesmo entre amostras distantes. Neste caso, temos  $\sum_{k=-\infty}^{\infty} \rho(k) = \infty$ .

<sup>1</sup>Processos Markovianos e Movimento Browniano puro são exemplos de processos SRD.

## Estatística R/S

A estatística R/S [89] [91] foi proposta por Harold Edwin Hurst para estimação do parâmetro de Hurst.

Considere uma sequência de amostras aleatórias  $\{X_i\}$  com somas parciais  $Y(n) = \sum_{i=1}^n X_i$  e variância amostral  $S^2(n) = (1/n) \sum_{i=1}^n X_i^2 - (1/n)^2 Y(n)^2$ . A estatística R/S é dada por

$$R(n)/S(n) = \frac{1}{S(n)} \left[ \max_{0 \leq t \leq n} (Y(t) - (t/n)Y(n)) - \min_{0 \leq t \leq n} (Y(t) - (t/n)Y(n)) \right], \quad (3.25)$$

Verifica-se que  $E[R(n)/S(n)] \sim C_1 n^H$  para  $n \rightarrow \infty$ , onde  $C_1$  é uma constante positiva.

Considerando-se uma sequência de tamanho  $N$ , devemos subdividi-la em  $K$  blocos de tamanho  $N/K$ . Para valores de  $n < N/K$  são obtidas  $K$  estimativas de  $R(n)/S(n)$  calculadas em cada bloco. Caso  $n$  seja maior que o tamanho do bloco ( $n > N/K$ ) obtém-se tantos valores de  $R(n)/S(n)$  quanto possível, partindo-se dos pontos iniciais de cada bloco. O valor do parâmetro  $H$  pode ser estimado através de regressão linear sobre os pontos obtidos, em um gráfico log-log de  $R(n)/S(n)$  versus  $n$ .

## 3.3 Gráficos de Recorrência e suas Medidas de Quantificação

Nas últimas décadas, os métodos clássicos de análise de dados baseados em modelo linear têm sido enriquecidos com novos métodos que são derivados da teoria do caos. Vários métodos de análise têm 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 estacionariedade [92]. Behrman e Baken [93] mostraram que não estacionariedade, ruído e um comprimento finito do sinal podem afetar a dimensão de correlação de dados de eletroglotografia (EGG).

Recentemente, outro método baseado na análise não linear de dados tem sido usado: os Gráficos de Recorrência (ou *Recurrence Plots - RP*) criados como método de análise de sistemas dinâmicos por Eckmann et al. [33]. A ideia de recorrência surgiu no século XIX, a partir do Teorema da Recorrência, do físico e filósofo francês J. H. Poincaré (1890).

Recorrência é uma propriedade fundamental dos sistemas dinâmicos dissipativos e a principal vantagem dos gráficos de recorrência é a possibilidade do uso dessa técnica para análise de séries curtas e não-estacionárias [92, 94].

O primeiro passo para a construção do gráfico de recorrência de uma série temporal de comprimento  $T$  é a imersão dessa série em um espaço  $m$ -dimensional, em geral, por meio do método dos atrasos temporais [68]. Nesse método são criadas  $m - 1$  cópias defasadas de  $\tau, 2\tau$

, ...,  $(m - 1)\tau$  amostras, a partir de cada sinal de voz disponível. Daí são construídos  $N$  vetores (Eq. 3.1), de dimensão  $m$ , representativos da dinâmica do sistema em análise no espaço de fase.

De acordo com a definição originalmente dada em [33], o gráfico de recorrência é uma matriz  $N \times N$  (matriz gráfica), em que  $N$  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 nas 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)  $\varepsilon$ , fixada no centro do estado corrente.

Uma definição alternativa comumente utilizada para o Gráfico de Recorrência é dada por [92]:

$$\mathcal{R}_{i,j}^{m,\varepsilon} = \Theta(\varepsilon - \|\vec{\xi}_i - \vec{\xi}_j\|), \quad \vec{\xi}_i \in R^m, \quad i, j = 1 \dots N, \quad (3.26)$$

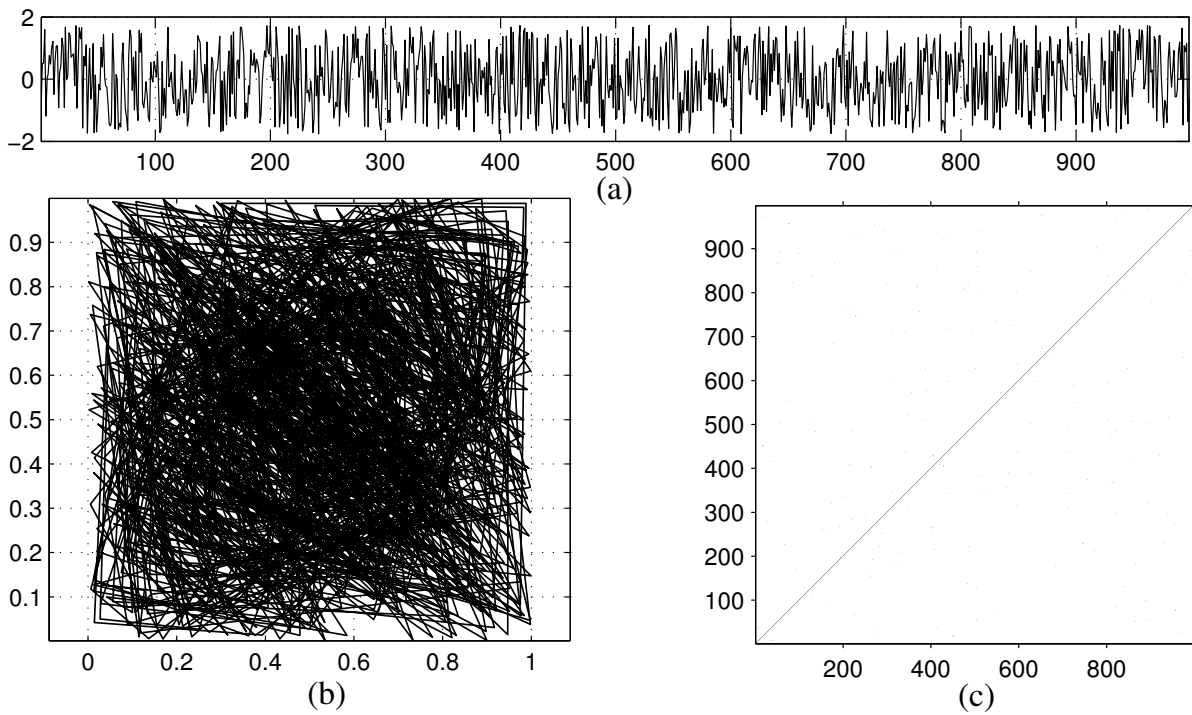
em que:

- $N$  é o número de estados  $\vec{\xi}_i$  considerados;
- $\varepsilon$  é o raio da vizinhança (*threshold*) no ponto  $\vec{\xi}_i$ ;
- $\|\cdot\|$  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).

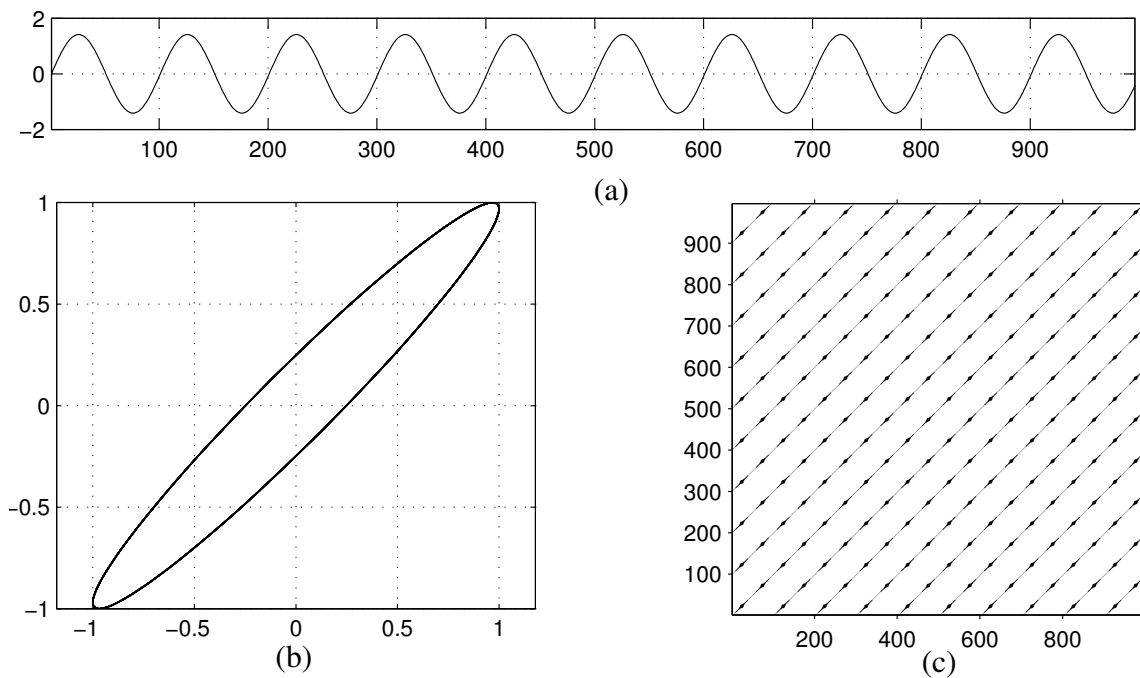
Pontos isolados, linhas diagonais, como também linhas verticais e horizontais (e a combinação destas para formar estruturas variadas) são estruturas típicas presentes nos gráficos de recorrência [94].

A análise visual dos gráficos de recorrência fornece resultados de forma qualitativa. Por exemplo, processos com comportamento estocástico tendem a não apresentar estruturas diagonais (Figura 3.1). Por outro lado, processos determinísticos causam diagonais mais longas e menos pontos de recorrência isolados (Figura 3.2).

Objetivando quantificar a dinâmica de um sistema não linear, Webber Jr. e Zbilut desenvolveram medidas de complexidade, as chamadas Medidas de Quantificação de Recorrência [34, 95]. As medidas inicialmente propostas - a taxa de recorrência, o determinismo, o comprimento máximo das linhas diagonais, a entropia de Shannon da distribuição de frequências dos comprimentos das linhas diagonais e a tendência - estão relacionadas a densidade dos pontos de recorrência e as estruturas diagonais presentes nos gráficos de recorrência. Posteriormente, Marwan [92] propôs outras três medidas de recorrência, desta vez baseadas nas estruturas verticais (ou horizontais), são elas: a laminaridade, o tempo de permanência e o comprimento máximo das estruturas verticais.



**Figura 3.1** Ruído branco (a) Série temporal, (b) Retrato de fase e (c) Gráfico de recorrência;  $T = 1.000$ ,  $m = 3$ ,  $\tau = 1$  e  $\varepsilon = 0, 1$  (norma euclidiana).



**Figura 3.2** Sinal senoidal (a) Série temporal, (b) Retrato de fase e (c) Gráfico de recorrência;  $T = 1.000$ ,  $m = 2$ ,  $\tau = 4$  e  $\varepsilon = 0, 1$  (norma euclidiana).

### 3.3.1 Taxa de Recorrência

A taxa de recorrência ou percentual de recorrência (*REC*) quantifica a porcentagem de pontos de recorrência presentes no gráfico de recorrência dentro de um raio especificado,  $\varepsilon$ , e é definida como

$$REC = \frac{1}{N^2} \sum_{i,j=1}^N \mathcal{R}_{i,j;i \neq j}^{m,\varepsilon}. \quad (3.27)$$

A *REC* varia na faixa de 0% (nenhuma recorrência) a 100% (todos os pontos são recorrentes). Para uma implementação adequada das medidas de recorrência é necessário que a matriz de recorrência seja esparsa ( $REC < 5\%$ ), caso contrário, vetores distantes podem ser, erroneamente, considerados vizinhos [96].

### 3.3.2 Determinismo

A razão entre o número de pontos de recorrência que formam as estruturas diagonais e todos os pontos de recorrência é introduzida como uma medida chamada determinismo (*DET*), que está relacionada com a previsibilidade do sistema e é definida como:

$$DET = \frac{\sum_{l=l_{min}}^N l \times P^\varepsilon(l)}{\sum_{i,j}^N \mathcal{R}_{i,j}^{m,\varepsilon}}, \quad (3.28)$$

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.

Quando o percentual de determinismo é igual a 0% nenhuma linha diagonal é formada no gráfico de recorrência (sistemas estocásticos), linhas curtas no *RP* indicam um sistema caótico ( $DET > 0$ ) e a presença de linhas longas caracteriza um sistema periódico ( $DET = 100\%$ ).

### 3.3.3 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.

$$ENTR = - \sum_{l=l_{min}}^N p(l) \ln p(l), \quad \text{em que} \quad (3.29)$$

$$p(l) = \frac{P^\varepsilon(l)}{\sum_{l=l_{min}}^N P^\varepsilon(l)} \quad (3.30)$$

### 3.3.4 Comprimento Máximo das Estruturas Diagonais

As estruturas diagonais mostram a distância em que um segmento da trajetória está em relação a outro segmento da mesma trajetória em outro instante de tempo. Assim, estas linhas informam sobre a divergência dos segmentos da trajetória.

O comprimento médio das linhas diagonais,  $L$ , é definido como o tempo médio que dois segmentos de uma trajetória estão próximos um do outro e pode ser interpretado como o tempo médio de predição. O valor de  $L$  é calculado da seguinte forma:

$$L = \frac{\sum_{l=l_{min}}^N l \times P^\varepsilon(l)}{\sum_{l=l_{min}}^N P^\varepsilon(l)}. \quad (3.31)$$

A análise de quantificação de recorrência utiliza o comprimento máximo das estruturas diagonais (ou sua inversa, a divergência), ao invés do comprimento médio. Assim,

$$L_{max} = \max(\{l_i; i = 1 \dots N_l\}), \quad (3.32)$$

em que  $N_l$  é o número total de linhas diagonais.

O inverso do máximo comprimento das diagonais  $L_{max}$  é definido como sendo a *divergência (DIV)*, ou seja,

$$DIV = \frac{1}{L_{max}}. \quad (3.33)$$

A *divergência* é uma medida do quão rapidamente trajetórias paralelas divergem uma da outra. Assim, quanto menor o valor de  $L_{max}$  mais divergente são as trajetórias do sistema analisado.

### 3.3.5 Tendência

A medida de quantificação chamada de tendência (*TREND*) é um coeficiente de regressão linear sobre a densidade dos pontos de recorrência das diagonais paralelas a diagonal principal (linha de identidade). Essa medida fornece informações a respeito da não-estacionaridade do processo e é definida por:

$$TREND = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(REC_i - \langle REC_i \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2} \quad (3.34)$$

em que  $N - \tilde{N}$  deve ser cerca de dez vezes a ordem de magnitude do tempo de autocorrelação [92].

A medida *TREND* é calculada como a inclinação da reta que melhor se ajusta aos valores do percentual de recorrência calculados a partir de valores inteiros de distância tomadas em relação a diagonal central. Neste caso, se a densidade dos pontos de recorrência permanecer aproximadamente constante ao longo do *RP*, *TREND* tenderá a zero. Se a densidade diminui

quando se afasta da diagonal central, a *TREND* será negativa. Caso contrário, se a densidade aumenta, a *TREND* será positiva.

Valores de *TREND* próximos de zero podem ser usados para determinar se o sistema é estacionário no tempo, valores diferentes de zero podem ser usados para determinar se o sistema está no regime transitório [96].

### 3.3.6 Laminaridade

É definida como a 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, assim:

$$LAM = \frac{\sum_{v=v_{min}}^N v \times P^\varepsilon(v)}{\sum_{i,j}^N \mathcal{R}_{i,j}^{m,\varepsilon}}, \quad (3.35)$$

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

### 3.3.7 Tempo de Permanência

O comprimento médio das estruturas verticais é chamado de tempo de permanência em um estado (*trapping time*), e é definido como

$$TT = \frac{\sum_{v=v_{min}}^N v \times P^\varepsilon(v)}{\sum_{v=v_{min}}^N P^\varepsilon(v)}. \quad (3.36)$$

A medida *TT* contém informação acerca da quantidade e do comprimento das estruturas verticais no *RP*. Ela mede o tempo médio que o sistema permanece em um estado específico.

### 3.3.8 Comprimento Máximo das Estruturas Verticais

O comprimento máximo das estruturas verticais presentes no gráfico de recorrência é calculado de forma análoga à medida  $L_{max}$ , Eq.(3.33), ou seja,

$$V_{max} = \max(\{V_l; l = 1 \dots N_v\}), \quad (3.37)$$

em que  $N_v$  é o número total de linhas verticais.

## 3.4 O Uso de Testes Estatísticos

O uso de testes estatísticos em uma pesquisa tem por objetivo fornecer um caráter científico às conclusões auferidas a partir dos resultados obtidos nos experimentos realizados. Para chegar a uma decisão objetiva sobre se uma determinada hipótese é confirmada, ou não, por um



conjunto de dados, é preciso ter um procedimento objetivo para aceitar ou rejeitar uma hipótese estabelecida [97].

Dois pontos importantes que devem ser observados para se ter confiabilidade nos resultados obtidos de um teste estatístico são a escolha adequada do tipo de teste estatístico e o nível de significância ( $\alpha$ ) a ser adotado.

De uma forma geral, os testes de hipótese são classificados em dois tipos: paramétricos e não paramétricos. Os testes paramétricos utilizam o pressuposto que os dados disponíveis formam uma amostra aleatória de uma (ou de mais de uma) população cuja distribuição é a normal. Por outro lado, os testes não paramétricos não necessitam de suposições sobre o tipo de distribuição da população a qual pertencem os dados da amostra utilizada.

Em geral, os procedimentos não-paramétricos não utilizam todas as informações proporcionadas pela amostra. Como resultado, um teste não-paramétrico será menos eficiente que o procedimento paramétrico correspondente quando a população em análise é a normal. Esta perda de eficiência é refletida por uma exigência de uma amostra de tamanho maior para o processo não paramétrico do que seria exigido pelo procedimento paramétrico correspondente, de modo a se obter o mesmo poder de teste. Por outro lado, esta perda de eficiência geralmente não é grande, e muitas vezes a diferença no tamanho da amostra é muito pequena. Quando as distribuições sob estudo não estão próximas da normal, os métodos não paramétricos têm muito a oferecer. Eles costumam oferecer melhoria considerável em relação aos métodos paramétricos [98].

Geralmente, se os dois métodos paramétricos e não paramétricos são aplicáveis a um determinado problema, deve ser usado o procedimento paramétrico por ser mais eficiente. No entanto, os pressupostos para uso dos métodos paramétricos podem ser difíceis ou impossíveis de justificar [98].

O nível de significância ( $\alpha$ ) deve ser estabelecido antes do experimento ser realizado e corresponde ao risco que se corre de rejeitar uma hipótese verdadeira. É o limite que se toma como base para afirmar que um certo desvio é decorrente do acaso ou não. São comumente utilizados como estatisticamente significativos os níveis  $\alpha = 0,05$  e  $\alpha = 0,01$ , ou seja, 5% e 1%, respectivamente. A partir de um nível de significância convencionalizado ( $\alpha$ ) os desvios são devidos à lei do acaso e o resultado é considerado não significativo.

Existem dois tipos de erros que podem ser cometidos ao chegar a uma decisão sobre uma hipótese. O primeiro, chamado erro do tipo I, envolve a rejeição da hipótese quando ela é, de fato, verdadeira. O segundo, chamado erro do tipo II, envolve não rejeitar a hipótese quando, de fato, ela é falsa.

Em um teste de hipótese, o nível de significância ( $\alpha$ ) é a sua probabilidade máxima permissível para cometer um erro tipo I. A probabilidade de ocorrência de um erro tipo II é denotado por  $\beta$ .

Se a hipótese nula for verdadeira, um valor  $P$  (ou valor de probabilidade) de um teste de hipótese é a probabilidade de se obter uma estatística amostral com valores tão extremos ou

mais extremos do que aquela determinada a partir dos dados da amostra. Para usar um valor  $P$  para chegar a uma conclusão em um teste de hipótese, é preciso comparar o valor  $P$  com o nível de significância ( $\alpha$ ). Se  $P \leq \alpha$ , então a hipótese nula deve ser rejeitada. Se  $P > \alpha$ , a hipótese nula não pode ser rejeitada [99].

Com o propósito de analisar o comportamento das desordens vocais que as patologias edema, nódulos e paralisia nas dobras vocais provocam no sinal de voz, no capítulo a seguir, é mostrada uma caracterização estatística dessas patologias por meio das medidas da análise dinâmica não linear (MNL) e das medidas de quantificação de recorrência (MQR).

## CAPÍTULO 4

---

# Caracterização de Sinais de Voz Usando Medidas da Análise Não Linear e Medidas de Recorrência

---

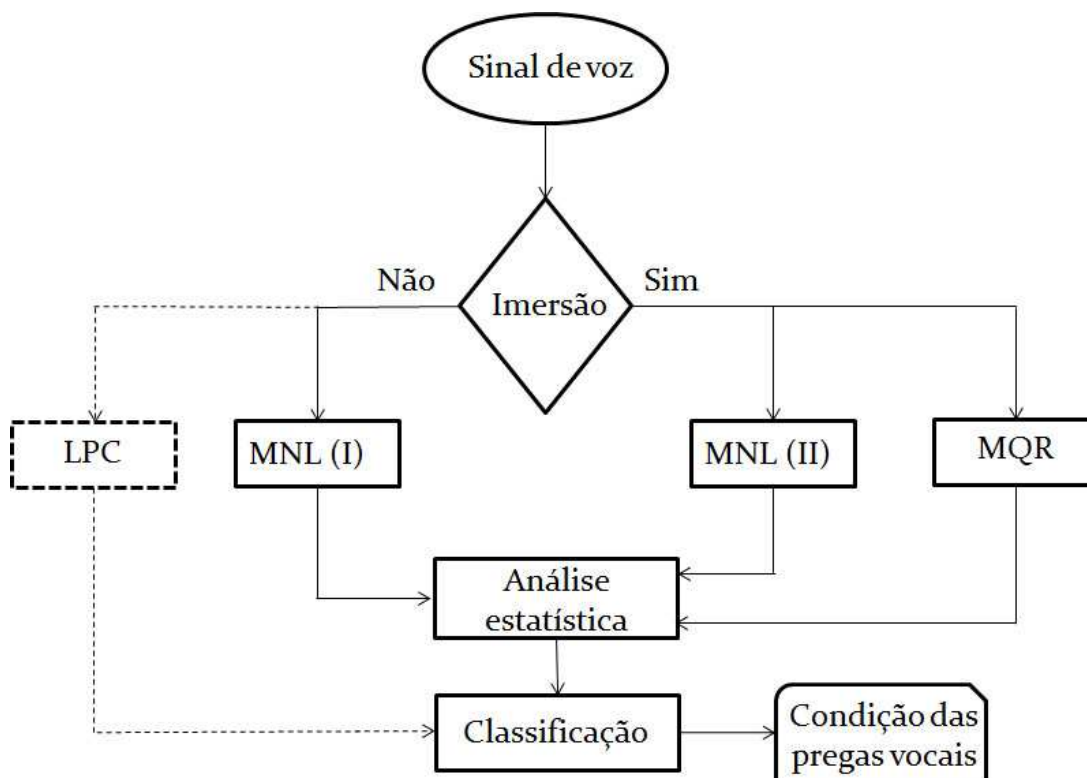
Neste capítulo é apresentada a metodologia aplicada nos experimentos computacionais realizados neste trabalho. Também são apresentados os valores experimentais obtidos para a caracterização dos sinais de vozes saudáveis e patológicas por meio das seguintes medidas associadas à teoria dos sistemas não lineares: Dimensão de correlação, entropia de correlação, maior expoente de Lyapunov, entropia de Shannon, entropia de Tsallis, entropia aproximada, o valor do primeiro mínimo da função de informação mútua e o expoente Hurst. Além disso, são apresentados os resultados obtidos considerando as seguintes medidas de quantificação de recorrência: Taxa de recorrência, determinismo, comprimento máximo das estruturas diagonais, entropia da distribuição de frequências dos comprimentos das linhas diagonais, tendência, laminaridade, comprimento máximo das estruturas verticais e tempo de permanência em um estado. São apresentados, ainda, os resultados dos testes estatísticos que procuram evidenciar o potencial discriminativo de cada uma dessas características, no sentido de caracterizar sinais de vozes saudáveis e patológicas.

### 4.1 Metodologia Aplicada

A metodologia aplicada nas simulações computacionais realizadas nesta pesquisa pode ser sintetizada na Figura 4.1.

#### 4.1.1 A Base de Dados

Os sinais de voz utilizados para as avaliações realizadas neste trabalho fazem parte de um banco de dados comercial, o *Disordered Voice Database*, Modelo 4337, desenvolvido por



**Figura 4.1** Fluxograma representativo da metodologia aplicada visando a classificação dos sinais de voz.

*Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab*. [60]. Este banco de dados vem sendo empregado em pesquisas de âmbito internacional envolvendo análise acústica de vozes desordenadas e na discriminação entre vozes saudáveis e patológicas [32, 100–102].

A base de dados inclui amostras de pacientes com uma ampla variedade de patologias orgânicas, neurológicas, psicogênicas e traumáticas, entre outras disfonias. Ela contém mais de 1.400 sinais de voz, gravados a partir de 700 locutores, sendo pronúncias da vogal sustentada /ah/ - com 53 arquivos de locutores com voz saudável e 657 sinais de vozes patológicas -, além de gravações de 12 segundos de voz contínua da sentença "*Rainbow Passage*" de todos os locutores.

Todos os arquivos de sinais e informações clínicas para as vozes patológicas foram coletadas e estruturadas no laboratório MEEI. Praticamente todos os pacientes tiveram extensa análise da função vocal incluindo estroboscopia, análise aerodinâmica e análise acústica. Todas as amostras foram coletadas em um ambiente controlado (baixo ruído ambiente, a distância constante do microfone, com 16 bits por amostra, e frequências de amostragem de 25 kHz (sinais de vozes patológicas) e 50 kHz (sinais de vozes saudáveis).

### 4.1.2 Aspectos Metodológicos

Do banco de dados utilizado, foram inicialmente selecionados 171 sinais de vozes da vogal sustentada /ah/, sendo 53 sinais de vozes saudáveis e 118 de vozes afetadas por algum tipo de patologia nas pregas vocais (45 com edema, 55 com paralisia e 18 com nódulos).

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 contínua das pregas vocais) na pronúncia deste tipo de som. Assim, vários tipos de anomalias desses mecanismos podem ser detectados [32].

Os sinais de vozes saudáveis, originalmente amostrados a uma taxa de 50.000 amostras/s, foram sub-amostrados a 25.000 amostras/s para equiparar o número de amostras aos sinais patológicos.

Algumas das características, dentre as que fazem parte do conjunto das medidas de dinâmica não linear (MNL), foram extraídas diretamente da série temporal que representa o sinal de voz. Esse conjunto de características formado pela entropia de Shannon, o primeiro mínimo da função de informação mútua, a entropia de Tsallis e o expoente de Hurst, está representado na Figura 4.1 como MNL (I).

As demais características que fazem parte do conjunto das medidas de dinâmica não linear, representado na Figura 4.1 como MNL (II), ou seja, a dimensão de correlação, a entropia de correlação, a entropia aproximada e o maior expoente de Lyapunov, bem como o conjunto de todas as medidas de quantificação de recorrência, representado na Figura 4.1 como MQR, foram extraídas após a aplicação do processo de imersão (*embedding*) de cada um dos sinais de voz em um espaço de fases adequado, por meio da reconstrução dos estados do sistema que os produziu, utilizando o método dos atrasos temporais (secção sec:atrator).

Como ferramentas para extração das medidas são empregados os seguintes *softwares*: TISEAN - *Time Series Analysis* v. 3.01 [103], o *Visual Recurrence Analysis* (VRA) [104], o RQA (*Recurrence Quantification Analysis*) [105], o CRP (*Cross Recurrence Plot*) [106], além de elaboração de rotinas (*scripts*) e uso de funções pré-definidas no ambiente Matlab@v.7.9.

Todas as características do conjunto das medidas não lineares foram obtidas a partir da segmentação do sinal analisado, procurando-se ter o máximo número de amostras em cada segmento, porém, levando-se também em consideração o intervalo de tempo em que o sinal de voz é admitido com sendo estacionário (16-32 ms). Assim, cada sinal é segmentado em 30 quadros de 800 amostras (sem sobreposição) usando uma janela retangular.

As medidas de quantificação de recorrência foram obtidas a partir de um único quadro de 10.000 amostras (entre 100 ms e 500 ms) de cada um dos sinais selecionados da base de dados.

A extração das medidas da análise dinâmica não linear foi realizada por meio de programas desenvolvidos no ambiente Matlab@v.7.9, como também, com o auxílio do pacote de rotinas TISEAN (*Time Series Analysis*) v. 3.01 [103] que implementam em *software* os algorit-

mos básicos referentes ao cálculo dessas medidas. Este pacote de rotinas tem sido utilizado em diversos trabalhos presentes na literatura especializada em dinâmica não linear [32, 107–109]. As medidas de quantificação de recorrência foram obtidas a partir do pacote de rotinas RQA (*Recurrence Quantification Analysis*) [105].

Após a extração de todas as características consideradas neste trabalho, foram efetuados testes de hipóteses estatísticos, por meio do *software Statistical Package for the Social Sciences* - SPSS Statistics®, versão 19.0, com o intuito de validar as conclusões a respeito do potencial discriminativo de cada medida no sentido de discriminar os sinais de voz analisados. Em todos os testes estatísticos realizados foi considerado um nível de significância de 5%.

Após a etapa da análise estatística, é realizada a classificação dos sinais como saudáveis ou patológicos - e entre sinais de vozes apresentando diferentes patologias - empregando a técnica da análise de discriminante. Neste trabalho, empregam-se a análise de discriminante linear (LDA) e a análise discriminante quadrática (QDA). Inicialmente, é feita a classificação com cada uma das medidas consideradas individualmente e depois, com o objetivo de melhorar os resultados individuais, as medidas são combinadas entre elas duas a duas, três a três, quatro a quatro, e assim por diante até ser tomado o conjunto formado por todas as características. Também é realizada a classificação dos sinais de voz por meio da formação de vetores híbridos, combinando as características dos conjuntos MNL e MQR e vetores de coeficientes obtidos por meio da análise LPC - Codificação por Predição Linear.

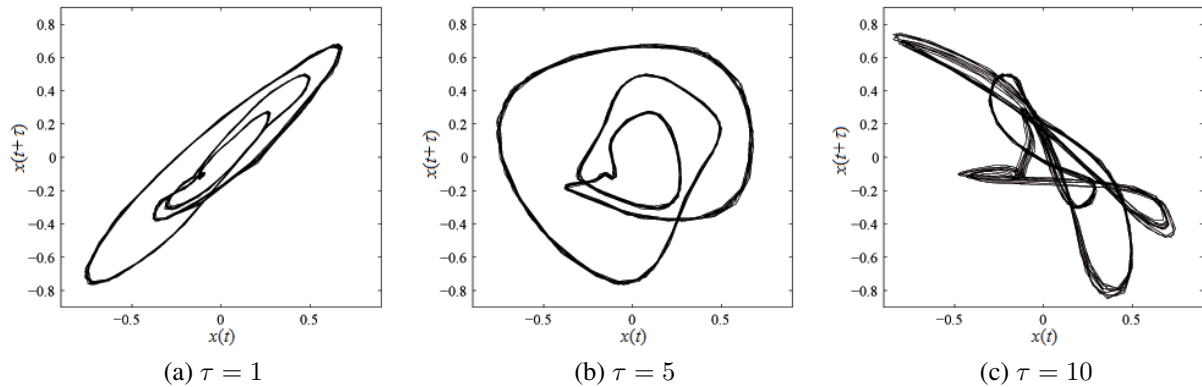
## 4.2 Reconstrução da Dinâmica do Sistema Vocal

A técnica de reconstrução do espaço de estados, baseada no teorema de imersão [68], permite reconstruir um espaço de estados  $m$ -dimensional similar ao espaço de estados original,  $d$ -dimensional, a partir de uma única variável de estado, no caso, o próprio sinal de voz.

A determinação eficiente dos parâmetros de reconstrução (tempo de defasagem,  $\tau$ , e dimensão de imersão,  $m$ ) é de extrema importância para a identificação dos invariantes geométricos. A partir da reconstrução do espaço de fases dos sinais analisados é realizada a extração das medidas de dinâmica não linear. O método utilizado para obtenção do tempo de defasagem utilizado neste trabalho está baseado na função de informação mútua [70].

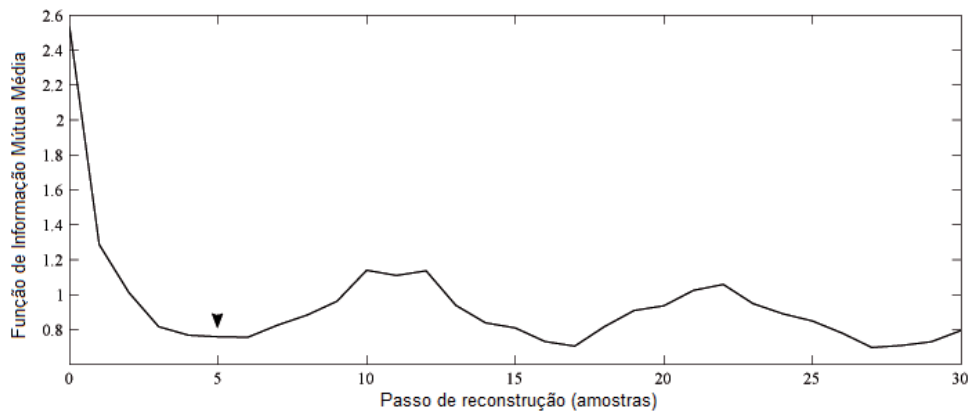
Na Figura 4.2 está ilustrada uma projeção bidimensional de um atrator tridimensional ( $m = 3$ ) reconstruído a partir de 800 amostras de um dos sinais de voz saudável analisados. No caso apresentado na Figura 4.2a foi utilizando um passo de reconstrução,  $\tau = 1$ , bem menor que o adequado. Na Figura 4.2b, o atrator foi reconstruído utilizando um passo de reconstrução "ideal",  $\tau = 5$ , obtido com o uso da função de informação mútua. Na Figura 4.2c o atrator foi reconstruído usando um passo bem maior que o adequado ( $\tau = 10$ ).

Na Figura 4.3 está ilustrado o comportamento da função de informação mútua,  $I(\tau)$  em função de  $\tau$ , para o mesmo segmento do sinal cujo atrator está mostrado na Figura 4.2. A seta



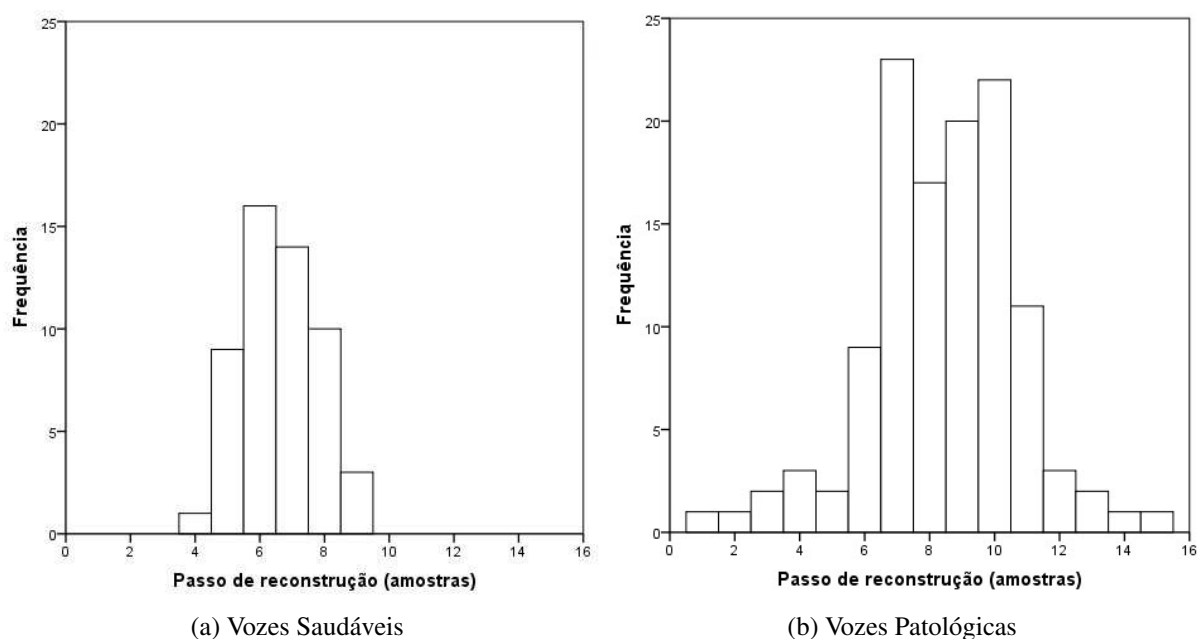
**Figura 4.2** Reconstrução do atrator da vogal sustentada /ah/ em um intervalo de 32 ms. Parâmetros:  $T = 800$  amostras,  $m = 3$  e (a)  $\tau = 1$ , (b)  $\tau = 5$  e (c)  $\tau = 10$ .

indica o ponto onde ocorre o primeiro mínimo da função e assim a determinação do passo de reconstrução "ideal" para este segmento ( $\tau = 5$ ).



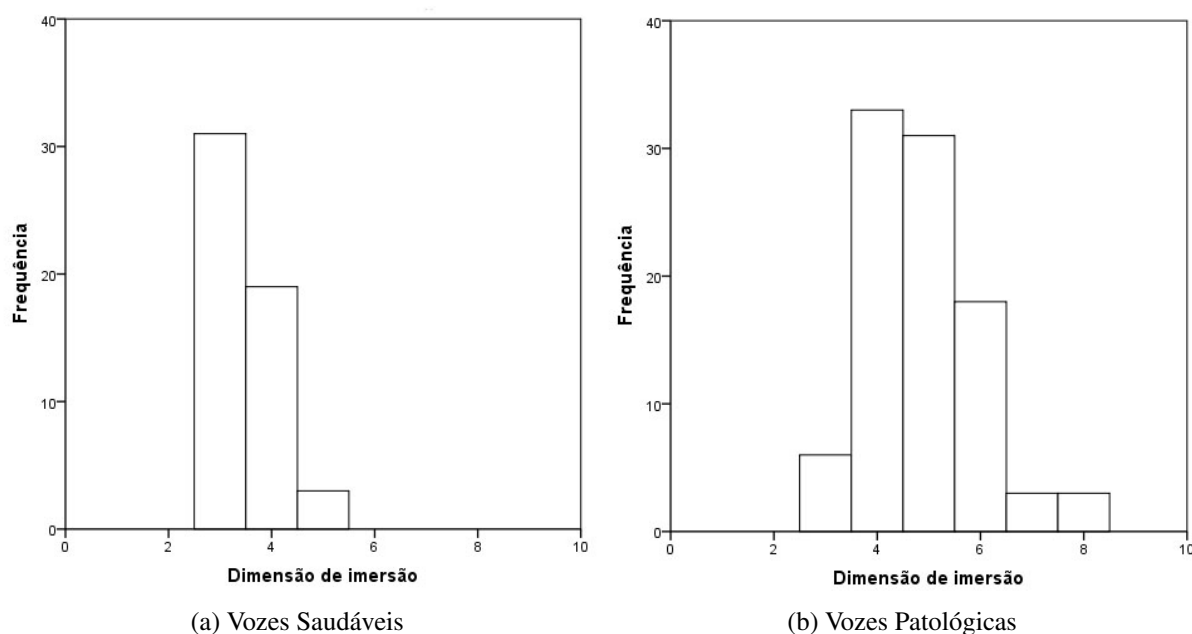
**Figura 4.3** Função de Informação Mútua e a determinação do passo de reconstrução "ideal" para um segmento de 32 ms da vogal sustentada /ah/.

Na Figura 4.4 estão apresentados os histogramas dos valores médios do passo de reconstrução obtidos para cada um dos 53 sinais de vozes saudáveis (Fig. 4.4a) e 118 sinais de vozes patológicas (Fig. 4.4b) calculados em segmentos de 32 ms. Foi observado que, para os sinais de vozes saudáveis analisados há uma predominância dos valores de  $\tau \leq 8$  (94,3%). Por outro lado, para os sinais de vozes patológicas observa-se que  $\tau > 8$  para 50,8% dos casos.



**Figura 4.4** Histogramas dos valores médios de  $\tau$  obtidos para os sinais de vozes (a) saudáveis e (b) patológicas.

O método dos falsos vizinhos próximos [72] foi utilizado neste trabalho para determinação da dimensão mínima de imersão,  $m$ . Na Figura 4.5 estão apresentados os histogramas dos valores médios da dimensão de imersão mínima,  $m$ , para os sinais de vozes saudáveis e patológicas analisadas, respectivamente.



**Figura 4.5** Histogramas dos valores médios de  $m$  obtidos para os sinais de vozes (a) saudáveis e (b) patológicas.



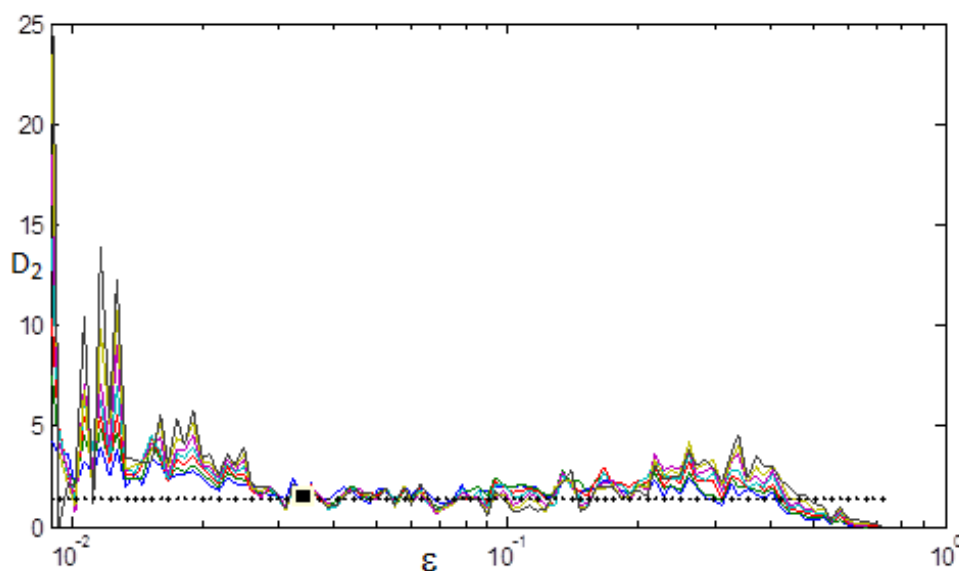
É possível observar que os sinais de vozes patológicas (Fig. 4.5b) necessitam, em geral, de uma maior dimensão de imersão quando comparados aos sinais de vozes saudáveis (Fig. 4.5a). Para as vozes saudáveis analisadas o máximo valor encontrado para a dimensão de imersão foi igual a 5 e  $m \leq 4$  para 94,3% dos sinais. Para as vozes patológicas foi observado valores de  $m > 4$  para 59,1% dos casos.

### 4.3 Medidas da Análise Dinâmica Não Linear (MNL)

Neste trabalho são calculadas as seguintes medidas da análise dinâmica não linear: a dimensão de correlação, a entropia de correlação e o maior expoente de Lyapunov, além da entropia de Shannon, a entropia de Tsallis, a entropia aproximada, o valor do primeiro mínimo da função de informação mútua e o expoente de Hurst visando a caracterização de sinais de vozes saudáveis e patológicas.

#### 4.3.1 Dimensão de Correlação ( $D_2$ )

Na Figura 4.6 é apresentado o comportamento da dimensão de correlação,  $D_2$  (definida na seção 3.2.4 na página 28), em função de  $m$  e de  $\varepsilon$ , para um segmento de 32 ms de um sinal de voz saudável.

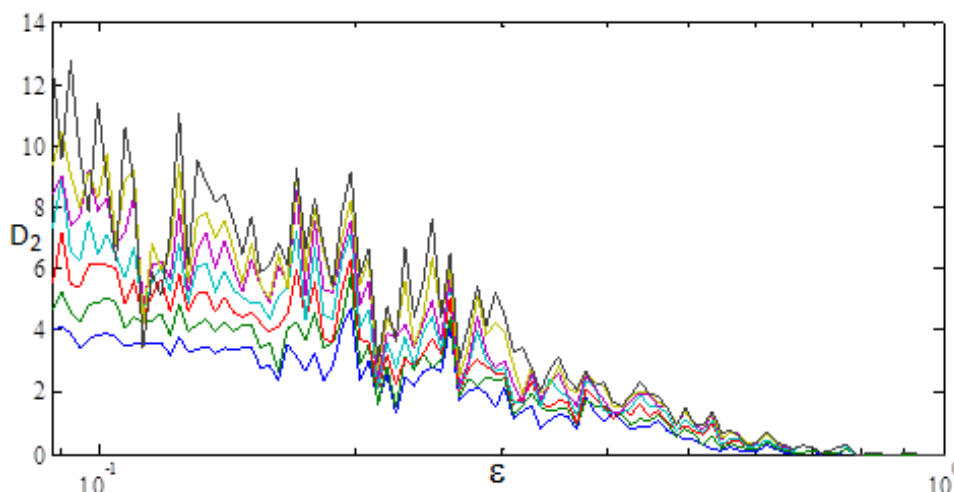


**Figura 4.6** Comportamento dos valores da dimensão de correlação para um segmento de 32 ms para um dos sinais de voz saudável em função da dimensão de imersão e do valor de  $\varepsilon$ , em que as curvas correspondem as diferentes dimensões de imersão ( $m = 4, \dots, 10$ ) de baixo para cima, respectivamente.

A partir da inspeção visual do comportamento dos valores da dimensão de correlação em função de  $m$  (variando de 4 a 10) e de  $\varepsilon$  é levada a efeito a determinação do valor de  $\varepsilon$  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  $\varepsilon$ . A dimensão de correlação do sinal é tomada como sendo a média calculada sob todos os 30 quadros considerados. No caso do segmento ilustrado na Figura 4.6, o quadrado sobreposto ao gráfico indica os valores encontrados ( $\varepsilon = 0,034$  e  $D_2 = 1,47$ ).

Durante os experimentos foi observado que a região de convergência não se apresenta de maneira clara para todos os sinais, nem mesmo para todos os segmentos do mesmo sinal. Em relação aos sinais de vozes patológicas, para cerca de 47% dos sinais afetados por paralisia, 26% com edema e 17% com nódulos, não foi possível a constatação visual de uma região de convergência para estimação da dimensão de correlação. Na Figura 4.7 está ilustrado um exemplo desse fato. Nesses casos, tais sinais foram descartados para o cálculo da medida da dimensão de correlação. Portanto, dos 118 sinais de vozes patológicas inicialmente considerados, apenas 74 sinais de pacientes com algum tipo de patologia na laringe (31 com edema de Reinke, 28 com paralisia e 15 com nódulos) foram efetivamente avaliados para essa medida.

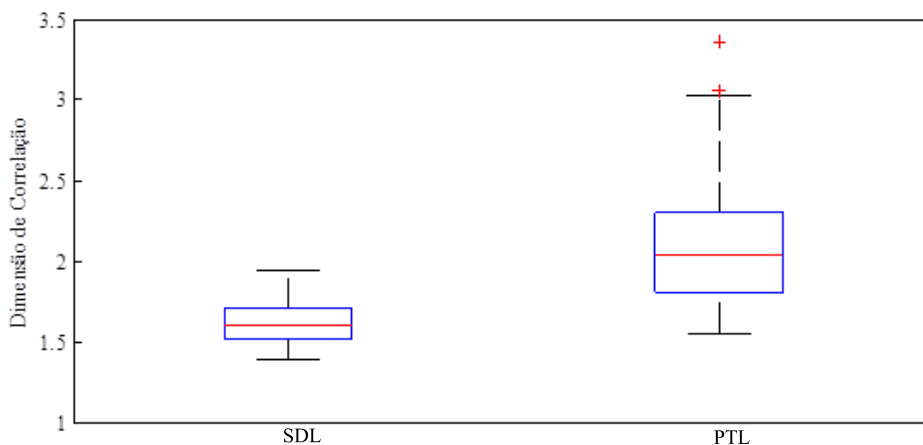


**Figura 4.7** Comportamento dos valores da dimensão de correlação para um segmento de 32 ms para um dos sinais de voz com paralisia em função da dimensão de imersão e do valor de  $\varepsilon$ , em que as curvas correspondem as diferentes dimensões de imersão ( $m = 4, \dots, 10$ ) de baixo para cima, respectivamente.

Na Figura 4.8 está ilustrada a distribuição dos valores médios da dimensão de correlação para os 53 sinais de vozes saudáveis (SDL) e 74 sinais de vozes patológicas (PTL) utilizados nos experimentos. Os sinais de vozes patológicas são: 28 com paralisia, 31 com edema e 15 com nódulos.

No gráfico de caixas (*boxplot*) apresentado na Figura 4.8 a linha interior em cada caixa (retângulo) indica o valor da mediana, as bases do retângulo mostram o 25-ésimo e o 75-ésimo percentil, os traços horizontais exteriores às caixas (*whiskers*) delimitam as extremidades do intervalo dos dados quando não há pontos fora de uma determinada faixa de valores e os pequenos círculos acima destas são valores atípicos (*outliers*).

Da Figura 4.8 pode ser observada uma maior variabilidade nos valores da dimensão de correlação dos sinais de vozes patológicas em relação aos sinais de vozes saudáveis. Pode ser



**Figura 4.8** Distribuição dos valores médios da dimensão de correlação para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

observada ainda uma distribuição aproximadamente simétrica dos valores de  $D_2$  para as vozes saudáveis em contraste a uma distribuição claramente assimétrica dos valores de  $D_2$  para as vozes patológicas.

Na Tabela 4.1 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da medida de dimensão de correlação para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.1** Média e desvio padrão da dimensão de correlação.

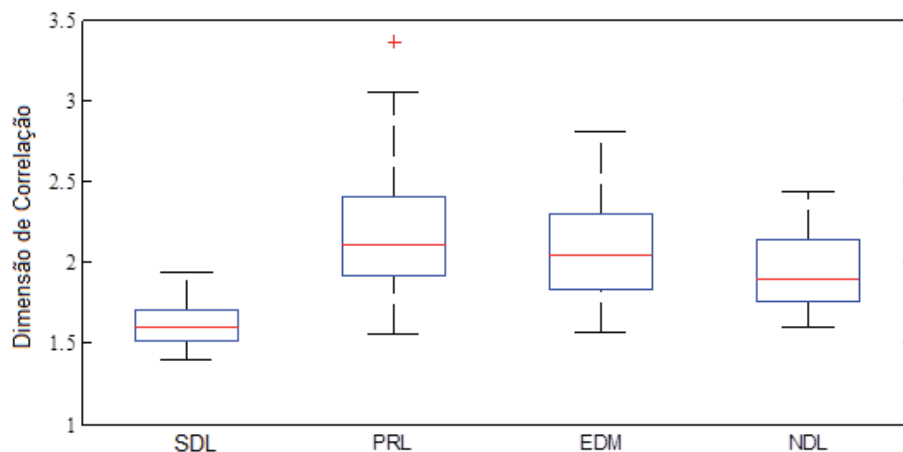
Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
$D_2$	1,62	0,14	2,09	0,35	< 0,001

Como podem ser observados a partir da Tabela 4.1, os resultados sugerem como característica dos sinais de voz analisados uma dinâmica de baixa dimensão ( $D_2 < 3$ ), tanto no caso da voz saudável como para a maioria dos sinais de vozes patológicas. Além disso, a análise dos resultados apresentados na Tabela 4.1 sugere diferenças estatísticas entre os dois grupos.

Após a realização do teste não-paramétrico U de Mann-Whitney [97], os resultados revelaram que o grupo de vozes patológicas apresentam valores de dimensão de correlação significativamente diferentes dos valores de  $D_2$  para o grupo de vozes saudáveis. Isto é um indicador de uma estrutura geométrica mais complexa em uma voz patológica.

Na Figura 4.9 está ilustrada a distribuição dos valores médios da dimensão de correlação para os mesmos sinais considerados na análise anterior, porém é feita uma estratificação dos sinais de vozes patológicas pelo tipo de patologia apresentada, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.2 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da medida de dimensão de correlação para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.



**Figura 4.9** Distribuição dos valores médios da dimensão de correlação para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

**Tabela 4.2** Média e desvio padrão da dimensão de correlação - separação das patologias.

$D_2$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	1,62	2,20	2,07	1,95
$\sigma$	0,14	0,43	0,30	0,26

A fim de testar as possíveis diferenças estatísticas entre os quatro grupos de vozes foi realizado o teste não-paramétrico de Kruskal-Wallis [97], seguido do teste de Mann-Witney para uma comparação dois a dois entre os grupos. A aplicação do teste ANOVA não-paramétrico revelou a existência de diferenças estatísticas significativas nos valores de  $D_2$  entre, pelo menos, dois dos grupos analisados ( $P < 0,001$ ). Os resultados obtidos a partir do teste de Mann-Witney estão apresentados na Tabela 4.3.

**Tabela 4.3** Valor-p - dimensão de correlação - separação das patologias.

$D_2$	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P < 0,001$	$P < 0,001$	$P < 0,001$
Paralisia	-	$P = 0,347$	$P = 0,036$
Edema	-	-	$P = 0,178$

A aplicação do teste de Mann-Witney revelou diferenças estatísticas significativas para os valores da dimensão de correlação entre o grupo de vozes saudáveis e cada um dos três grupos de vozes patológicas. Os resultados do teste apontaram ainda uma diferença significativa entre os valores de  $D_2$  do grupo dos sinais afetados por paralisia e aqueles do grupo de vozes afetadas por nódulos. Entretanto, os valores da dimensão de correlação entre o grupo de vozes afetadas por edema e os outros dois grupos de vozes patológicas (paralisia e nódulos) não apresentaram diferenças estatísticas para um nível de significância de 0,05.

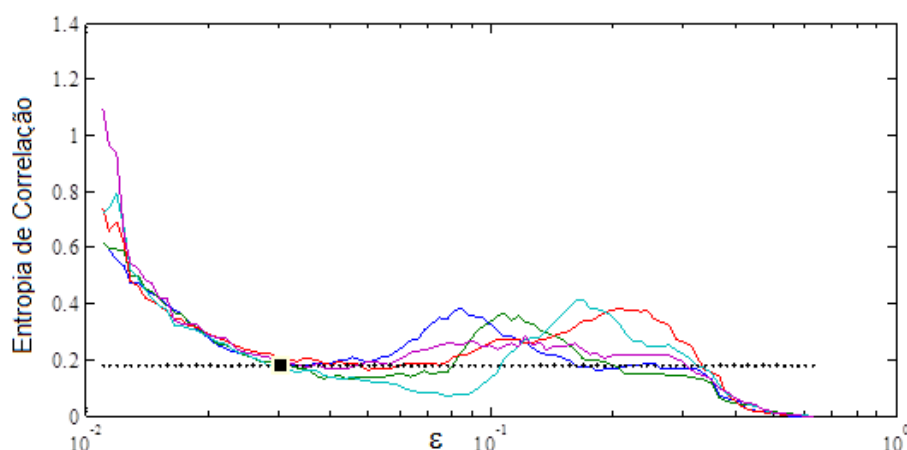
### 4.3.2 Entropia de Correlação ( $K_2$ )

Na Figura 4.10 é apresentado o comportamento da entropia de correlação,  $K_2$  (definida na seção 3.2.8 na página 31), em função de  $m$  e de  $\varepsilon$ , para um segmento de 32 ms de um sinal de voz patológico (com nódulos).

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  $\varepsilon$ .

A partir da inspeção visual do comportamento dos valores da entropia de correlação em função de  $m$ , e de  $\varepsilon$ , é levada a efeito a determinação do valor de  $\varepsilon$  que oferece o menor coeficiente de variação entre as curvas dentro de um intervalo de convergência (novamente, nem sempre presente em todos os sinais analisados). O valor de  $K_2$  para o segmento é estimado como sendo a média dos valores obtidos de cada curva para esse valor de  $\varepsilon$ .

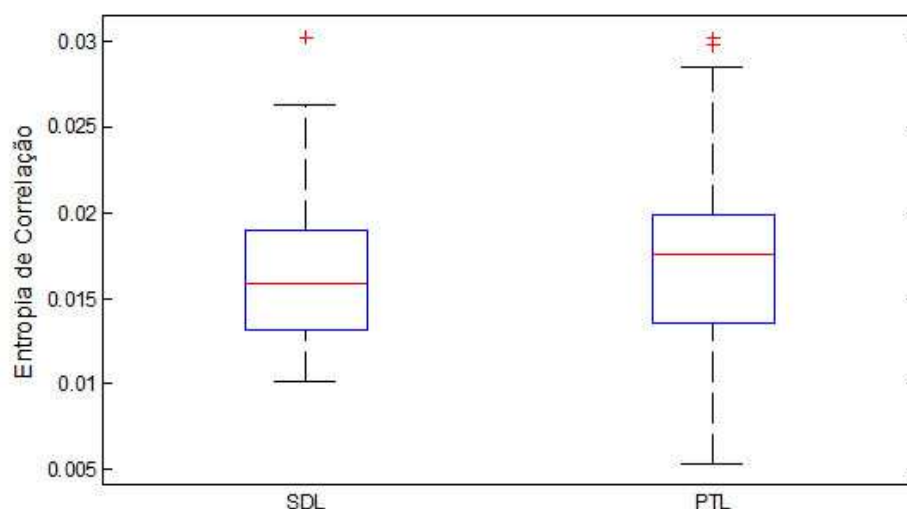
A entropia de correlação do sinal é tomada como sendo a média calculada sob todos os 30 segmentos do sinal sob análise. No caso do segmento ilustrado na Figura 4.10, o quadrado sobreposto ao gráfico indica os valores encontrados ( $\varepsilon = 0,031$  e  $K_2 = 0,184$ ).



**Figura 4.10** Comportamento dos valores da entropia de correlação para um segmento de 32 ms para um dos sinais de voz saudável em função da dimensão de imersão e do valor de  $\varepsilon$ , em que as curvas correspondem as diferentes dimensões de imersão ( $m = 4, \dots, 10$ ) de baixo para cima, respectivamente.

Na Figura 4.11 está ilustrada a distribuição dos valores médios da entropia de correlação para os 53 sinais de vozes saudáveis (SDL) e 74 sinais de vozes patológicas (PTL) utilizados nos experimentos.

Da Figura 4.11 pode-se perceber uma maior variabilidade nos valores da entropia de correlação dos sinais de vozes patológicas em relação aos sinais de vozes saudáveis. Entretanto, pode-se verificar uma leve assimetria na distribuição dos valores de  $K_2$  para os sinais saudáveis em contraste com uma distribuição aproximadamente simétrica dos valores desta medida para os sinais patológicos. O sinal (+) indica a presença de um valor atípico na distribuição dos valores de  $K_2$  para as vozes saudáveis.



**Figura 4.11** Distribuição dos valores médios da entropia de correlação para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

Na Tabela 4.4 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da medida de entropia de correlação para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.4** Média e desvio padrão para a entropia de correlação.

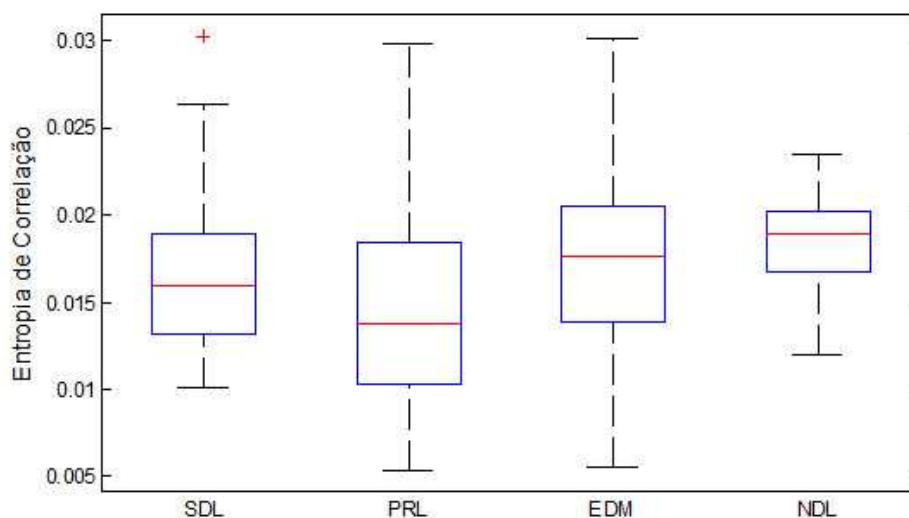
Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
$K_2$	0,016	0,004	0,017	0,005	0,214

Como podem ser verificados com base na Tabela 4.4, os resultados sugerem como característica dos sinais de voz analisados a quase coincidência dos valores das médias e dos desvios padrão entre os dois grupos de sinais analisados. Após a realização do teste de Mann-Whitney para amostras independentes, os resultados revelaram que a diferença entre os valores de  $K_2$  entre os grupos saudável e patológico é não-significativa. Assim, pode-se dizer que, segundo os dados disponíveis, a entropia de correlação tem valores iguais para ambos os grupos.

Na Figura 4.12 está ilustrada a distribuição dos valores médios da entropia de correlação para os sinais de vozes saudáveis (SDL) e os sinais de vozes patológicas discriminadas pelo tipo de patologia, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.5 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da medida de entropia de correlação para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

A fim de testar as possíveis diferenças entre as médias dos valores de  $K_2$  dos quatro grupos de vozes foi realizado o teste estatístico não-paramétrico de Kruskal-Wallis. Os resultados revelaram a existência de uma diferença estatísticas significativa entre, pelo menos, dois dos quatro grupos analisados ( $P = 0,018$ ). Os resultados obtidos a partir da aplicação do teste de Mann-Witney para comparação dois a dois entre os grupos estão apresentados na Tabela 4.6.



**Figura 4.12** Distribuição dos valores médios da entropia de correlação para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

**Tabela 4.5** Média e desvio padrão da entropia de correlação - separação das patologias.

$K_2$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	0,016	0,015	0,018	0,019
$\sigma$	0,004	0,006	0,005	0,003

**Tabela 4.6** Valor-p - entropia de correlação - separação das patologias.

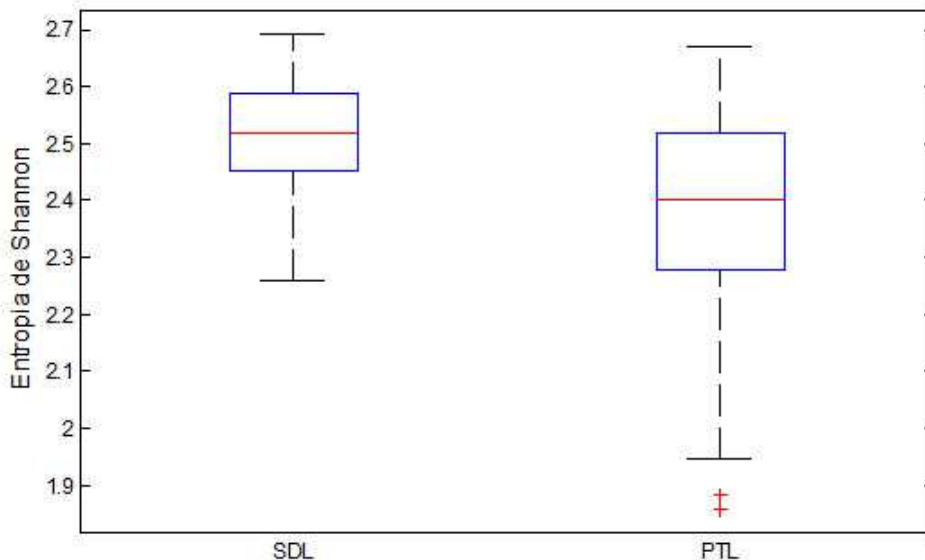
$K_2$	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P = 0,348$	$P = 0,174$	$P = 0,005$
Paralisia	-	$P = 0,093$	$P = 0,019$
Edema	-	-	$P = 0,117$

Como conclusão, a entropia de correlação não oferece, a princípio, um poder discriminativo entre vozes saudáveis e vozes afetadas por paralisia ou edema. Entretanto, há diferenças estatísticas significativas para os valores de  $K_2$  entre os grupos de vozes saudáveis e vozes com nódulos, e também entre este último e o de vozes com paralisia.

### 4.3.3 Entropia de Shannon ( $H_1$ )

A entropia de Shannon,  $H_1$  (definida na seção 3.2.5 na página 29), foi calculada considerando intervalos de 32 ms (800 amostras) do sinal analisado. A entropia de Shannon atribuída ao sinal é então tomada como a média desta medida calculada sob todos os 30 segmentos considerados.

Na Figura 4.13 está ilustrada a distribuição dos valores médios da entropia de Shannon para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos.



**Figura 4.13** Distribuição dos valores médios da entropia de Shannon para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

Na Tabela 4.7 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da entropia de Sahnnon para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.7** Média e desvio padrão para a entropia de Shannon.

Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
$H_1$	2,51	0,09	2,38	0,17	< 0,001

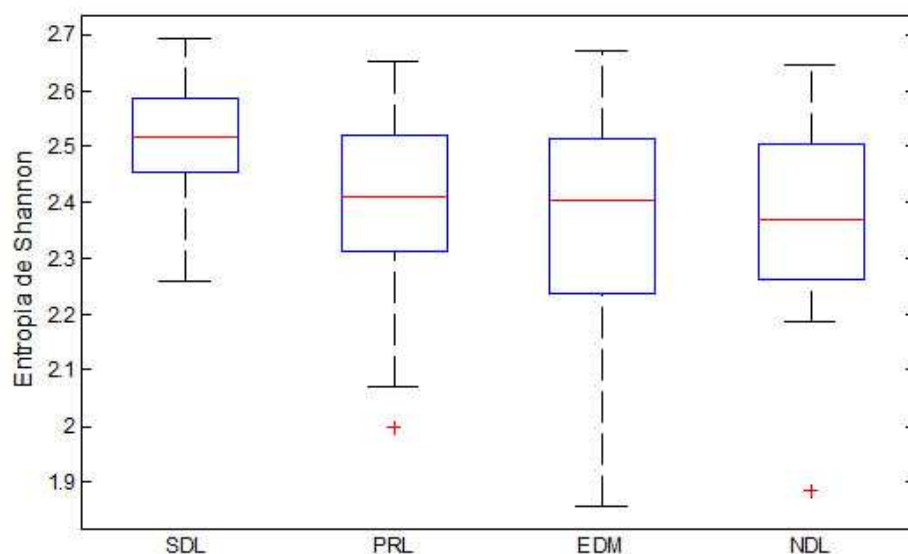
Após a realização do teste  $t$  [110] para duas amostras independentes, os resultados indicam diferenças estatísticas significativas entre as médias dos dois grupos analisados. Assim, a entropia de Shannon pode ser utilizada, em potencial, como uma medida de discriminação entre vozes saudáveis e patológicas.

Na Figura 4.14 está ilustrada a distribuição dos valores médios da entropia de Shannon para os sinais de vozes saudáveis (SDL) e os sinais de vozes patológicas discriminadas pelo tipo de patologia, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.8 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da entropia de Shannon para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

A fim de testar as possíveis diferenças estatísticas entre os quatro grupos de vozes foi realizado o teste paramétrico de análise de variância (ANOVA) [110] que indicou a existência de diferenças estatísticas significativas ( $P < 0,001$ ) envolvendo, pelo menos, dois dos quatro





**Figura 4.14** Distribuição dos valores médios da entropia de Shannon para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

**Tabela 4.8** Média e desvio padrão da entropia de Shannon - separação das patologias.

$H_1$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	2,51	2,41	2,36	2,36
$\sigma$	0,09	0,15	0,20	0,18

grupos analisados. Os resultados obtidos a partir da aplicação do teste de Dunnett (T3) para uma comparação dois a dois entre os grupos estão apresentados na Tabela 4.9.

**Tabela 4.9** Valor-p - entropia de Shannon - separação das patologias.

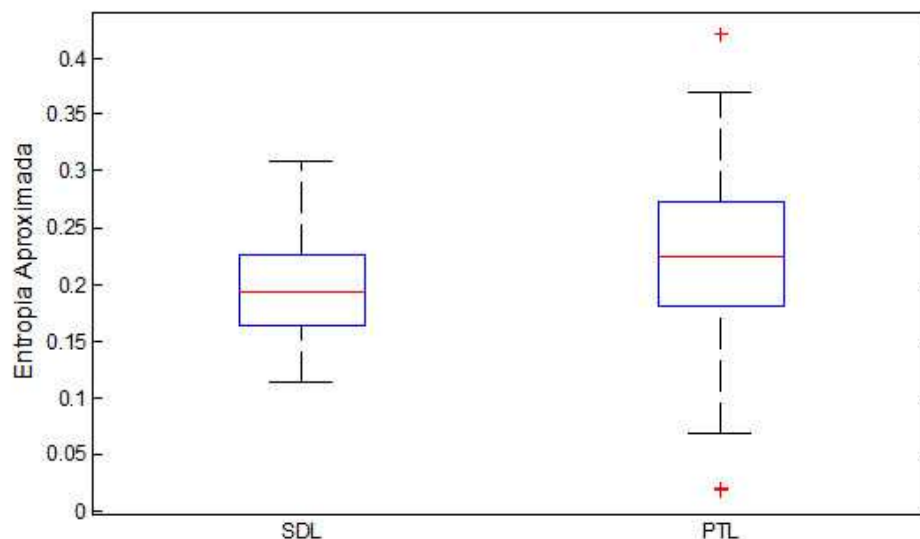
$H_1$	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P = 0,001$	$P < 0,001$	$P = 0,022$
Paralisia	-	$P = 0,644$	$P = 0,902$
Edema	-	-	$P = 1,000$

A análise dos resultados apresentados na Tabela 4.9 indica diferenças estatísticas significativas dos valores da entropia de Shannon entre o grupo de vozes saudáveis e cada um dos três grupos de vozes patológicas. Entretanto, os testes revelaram não haver evidências estatísticas para o uso da entropia de Shannon no sentido de diferenciar entre os grupos de vozes patológicas.

Como conclusão, a entropia de Shannon pode ser usada, em potencial, como um discriminativo entre vozes saudáveis e as patologias da laringe consideradas neste trabalho, mas não para diferenciar entre os diferentes tipos de patologia.

### 4.3.4 Entropia Aproximada ( $ApEn$ )

Na Figura 4.15 está ilustrada a distribuição dos valores médios da entropia aproximada,  $ApEn$  (definida na seção 3.2.9 na página 32), para os 53 sinais de vozes saudáveis (SDL) e 93 sinais de vozes patológicas (PTL) utilizados nos experimentos.



**Figura 4.15** Distribuição dos valores médios da entropia aproximada para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

Na Tabela 4.10 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da entropia aproximada para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.10** Média e desvio padrão para a entropia aproximada.

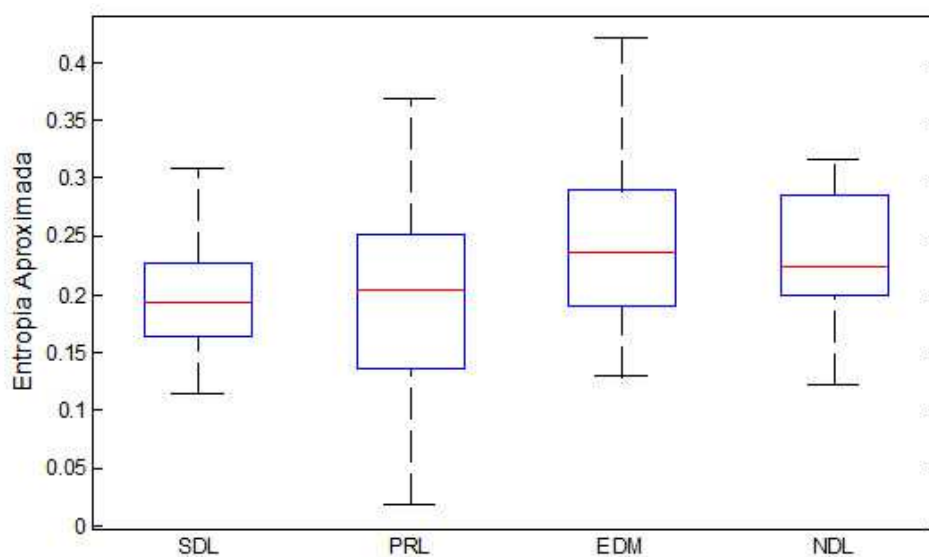
Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
$ApEn$	0,198	0,050	0,215	0,087	0,148

A aplicação do teste  $t$  revelou não haver diferenças estatísticas significativas entre as médias dos dois grupos analisados. Assim, a entropia aproximada não apresenta, em potencial, poder discriminativo entre vozes saudáveis e patológicas.

Na Figura 4.16 está ilustrada a distribuição dos valores médios da entropia aproximada para os sinais de vozes saudáveis (SDL) e os sinais de vozes patológicas discriminadas pelo tipo de patologia, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.11 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) para a entropia aproximada dos sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

A aplicação do teste ANOVA de um caminho revelou a existência de diferença estatísticas significativa entre as médias de, pelo menos, dois dos grupos analisados ( $P = 0,010$ ).



**Figura 4.16** Distribuição dos valores médios da entropia aproximada para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

**Tabela 4.11** Média e desvio padrão da entropia aproximada - separação das patologias.

$ApEn$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	0,198	0,185	0,235	0,232
$\sigma$	0,050	0,091	0,087	0,056

Diante deste fato, foi aplicado o teste *post hoc* de Dunnett ( $T_3$ ) para a comparação de múltiplas médias. Os resultados estão apresentados na Tabela 4.12.

**Tabela 4.12** Valor-p - entropia aproximada - separação das patologias.

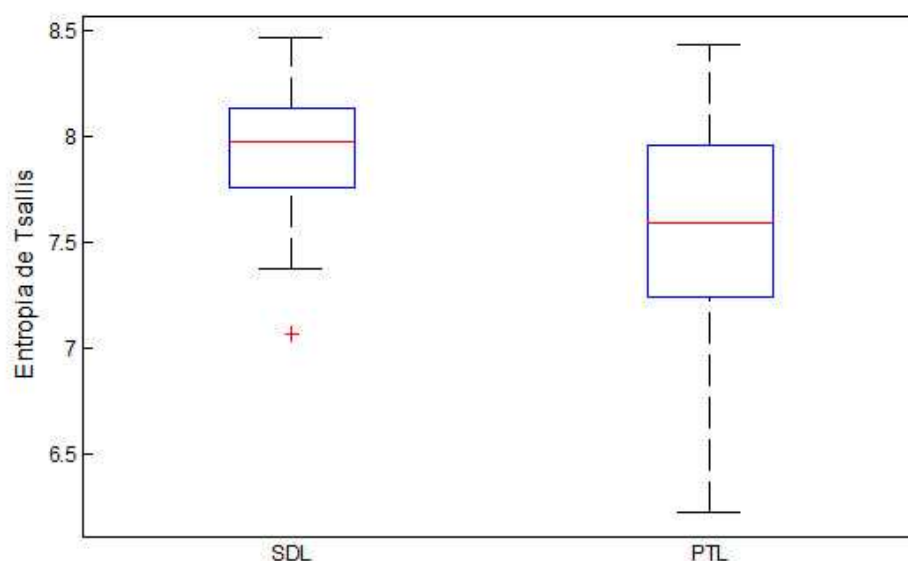
$ApEn$	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P = 0,961$	$P = 0,100$	$P = 0,239$
Paralisia	-	$P = 0,086$	$P = 0,160$
Edema	-	-	$P = 1,000$

Após o teste a aplicação do teste *post hoc* de Dunnett ( $T_3$ ), os resultados revelaram não existir diferenças estatísticas significativas entre quaisquer dos grupos considerados. A aparente contradição entre o resultado fornecido pelo teste anterior (ANOVA) e os resultados apresentados pelo teste de Dunnett ( $T_3$ ), pode ser explicado considerando que, caso as variâncias entre os grupos fossem estatísticas iguais, outros testes *post hoc* seriam adequados, como por exemplo, o teste de Tukey que revelaria diferença estatísticas significativa entre os grupos de vozes com paralisia e com edema ( $P = 0,017$ ).

### 4.3.5 Entropia de Tsallis ( $H_q$ )

Após experimentos realizados utilizando a entropia de Tsallis,  $H_q$  (definida na seção 3.2.10 na página 32), com diferentes valores do parâmetro  $q$ , Falcão et al. [111] relatam que, os resultados mais satisfatórios na discriminação entre vozes saudáveis e patológicas foram obtidos com o valor  $q = 0,5$ . Por isso, neste trabalho, também foi adotado esse mesmo valor de  $q$  para o cálculo de  $H_q$ .

Na Figura 4.17 está ilustrada a distribuição dos valores médios da entropia de Tsallis,  $H_q$ , para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos.



**Figura 4.17** Distribuição dos valores médios da entropia de Tsallis para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

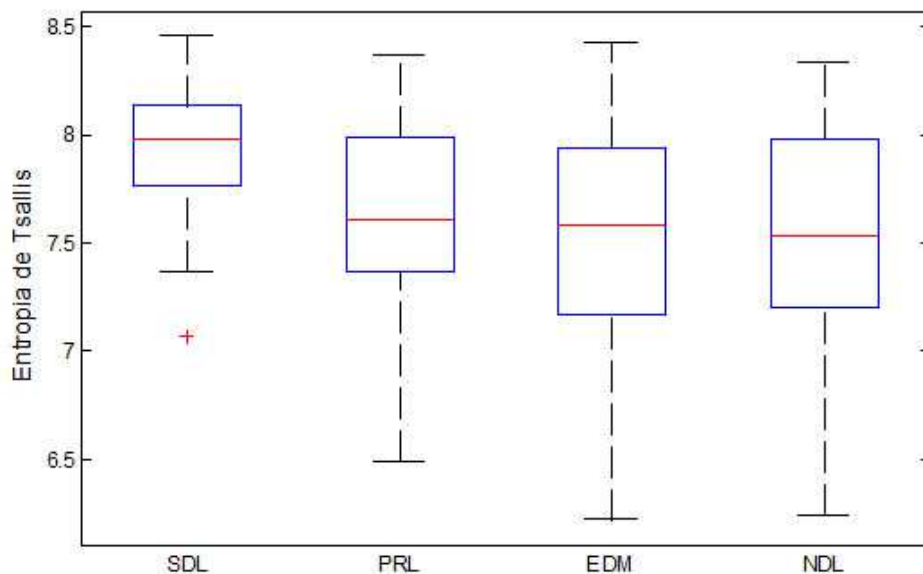
Na Tabela 4.13 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) da entropia de Tsallis para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.13** Média e desvio padrão para a entropia de Tsallis ( $q = 0,5$ ).

Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
$H_q$	7,93	0,29	7,57	0,50	< 0,001

Após a realização do teste não-paramétrico U de Mann-Whitney, os resultados indicam que o grupo de vozes saudáveis apresenta valores da entropia de Tsallis significativamente diferentes dos valores desta medida para o grupo de vozes patológicas. Assim, a entropia de Tsallis pode ser utilizada, em potencial, como uma medida de discriminação entre vozes saudáveis e patológicas.

Na Figura 4.18 está ilustrada a distribuição dos valores médios da entropia de Tsallis para os sinais de vozes saudáveis (SDL) e os sinais de vozes patológicas discriminadas pelo tipo de patologia, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).



**Figura 4.18** Distribuição dos valores médios da entropia de Tsallis para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.14 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) para a entropia de Tsallis dos sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.14** Média e desvio padrão da entropia de Tsallis - separação das patologias.

$H_q$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	7,93	7,63	7,51	7,54
$\sigma$	0,29	0,45	0,57	0,50

A fim de testar as possíveis diferenças dos valores da entropia de Tsallis entre os quatro grupos de vozes foi realizado o teste ANOVA não paramétrico de Kruskal-Wallis, que indicou a existência de diferenças estatísticas significativas ( $P < 0,001$ ) envolvendo, pelo menos, dois dos grupos analisados. Na Tabela 4.15 estão apresentados os resultados da aplicação do teste estatístico de Mann-Whitney para a comparação dois a dois entre os grupos.

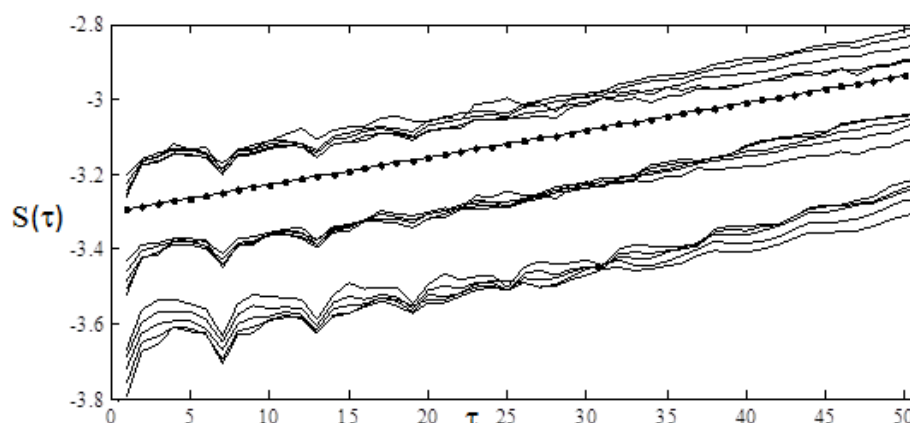
Após a aplicação do teste de Mann-Whitney, os resultados revelaram diferenças significativas dos valores da entropia de Tsallis entre o grupo de vozes saudáveis e os três grupos de vozes patológicas. Entretanto, não houve diferenças estatísticas significativas entre os grupos de vozes patológicas.

**Tabela 4.15** Valor-p - entropia aproximada - separação das patologias.

$H_q$	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P < 0,001$	$P < 0,001$	$P = 0,002$
Paralisia	-	$P = 0,377$	$P = 0,450$
Edema	-	-	$P = 0,994$

### 4.3.6 Maior Expoente de Lyapunov ( $\lambda_1$ )

Na Figura 4.19 está ilustrado o comportamento de  $S(\tau)$  (Eq.3.3) em função do número de iterações  $\tau$ , para três diferentes valores de  $\varepsilon$  e  $m$  variando de 4 a 8. A linha reta representa a inclinação média das curvas mostradas e sua inclinação é a estimacão para o maior expoente de Lyapunov,  $\lambda_1$  (definido na seção 3.2.1 na página 25).

**Figura 4.19** Comportamento de  $S(\tau)$  em função do número de iterações, de  $\varepsilon$  e de  $m$ .

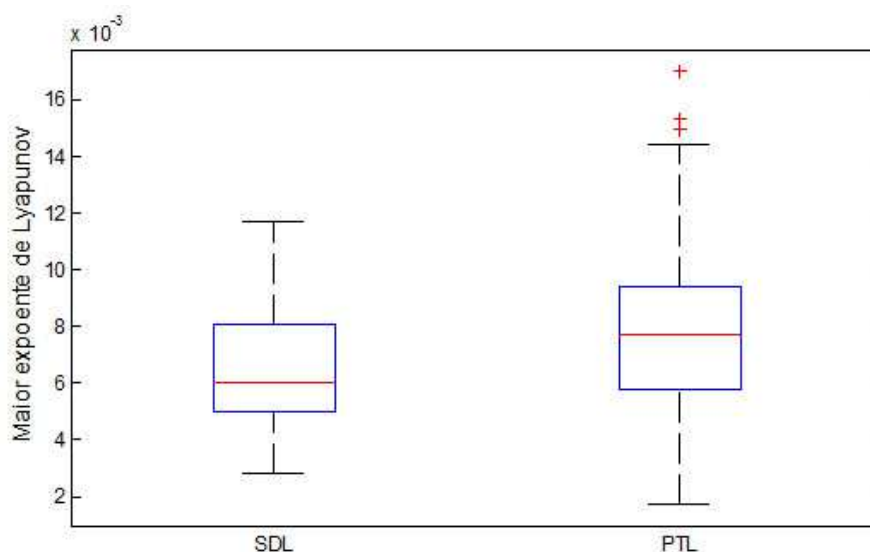
Na Figura 4.20 está ilustrada a distribuição dos valores médios do maior expoente de Lyapunov,  $\lambda_1$ , para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos.

Na Tabela 4.16 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) do maior expoente de Lyapunov para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.16** Média e desvio padrão para o maior expoente de Lyapunov.

Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
$\lambda_1$	0,0067	0,0024	0,0080	0,0030	0,006

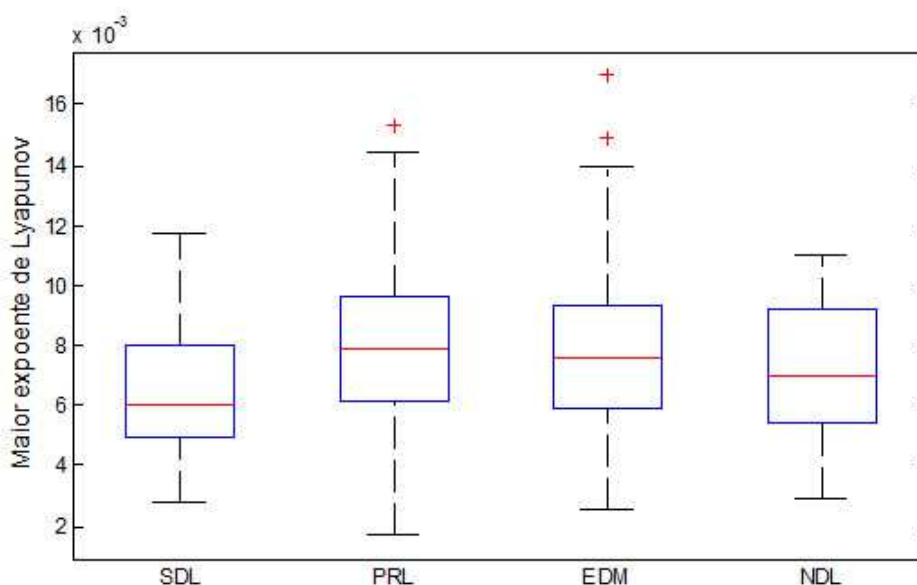
Após a realização do teste U não-paramétrico de Mann-Whitney, foi verificado que há diferenças estatísticas significativas dos valores do maior expoente de Lyapunov entre os grupos de vozes saudáveis e patológicas. A partir dos resultados fornecidos pelo teste estatístico



**Figura 4.20** Distribuição dos valores médios do maior expoente de Lyapunov para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

realizado pode-se dizer que, segundo os dados disponíveis, a medida do maior expoente de Lyapunov tem valores diferentes para os dois grupos. Sendo, a princípio, uma medida de discriminação entre vozes saudáveis e patológicas.

Na Figura 4.21 está ilustrada a distribuição dos valores médios do maior expoente de Lyapunov para os sinais de vozes saudáveis (SDL) e os sinais de vozes patológicas discriminadas pelo tipo de patologia, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).



**Figura 4.21** Distribuição dos valores médios do maior expoente de Lyapunov para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.17 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) para o maior expoente de Lyapunov dos sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.17** Média e desvio padrão do maior expoente de Lyapunov - separação das patologias.

$\lambda_1$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	0,0067	0,0074	0,0075	0,0072
$\sigma$	0,0024	0,0037	0,0027	0,0024

Após aplicação do teste não-paramétrico de Kruskal-Wallis foi observado que há diferenças estatísticas significativas entre, pelo menos, dois dos dos quatro grupos analisados ( $P = 0,035$ ). Os resultados obtidos após a aplicação do teste de Mann-Whitney para comparação estatística dois a dois entre os grupos estão apresentados na Tabela 4.18.

**Tabela 4.18** Valor-p - maior expoente de Lyapunov - separação das patologias.

$\lambda_1$	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P = 0,010$	$P = 0,016$	$P = 0,362$
Paralisia	-	$P = 0,806$	$P = 0,318$
Edema	-	-	$P = 0,349$

De acordo com os resultados encontrados e para os dados disponíveis é possível concluir que a medida do maior expoente de Lyapunov tem, a princípio, o poder de discriminar entre vozes saudáveis e patológicas afetadas por paralisia ou edema nas pregas vocais. Entretanto, o teste aplicado revelou que não há diferenças estatísticas significativas entre os grupos de vozes saudáveis e com nódulos, nem entre os grupos de sinais apresentando diferentes patologias.

### 4.3.7 Primeiro Mínimo da Função de Informação Mútua (FMMI)

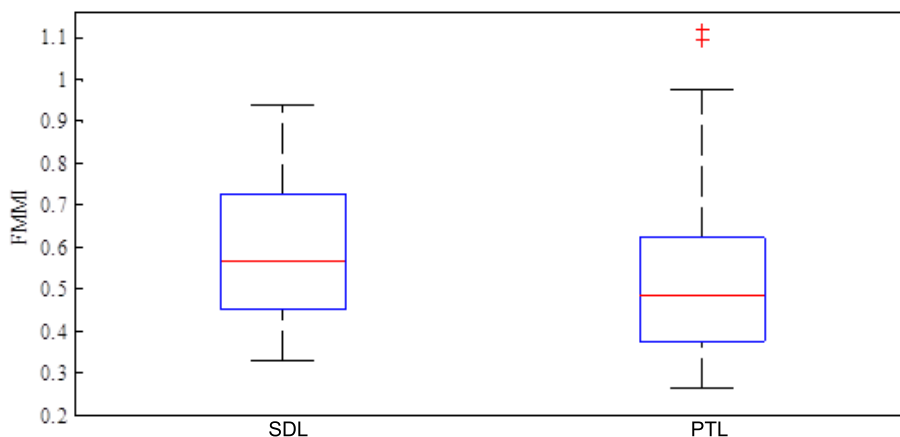
Na Figura 4.22 está ilustrada a distribuição dos valores médios do primeiro mínimo da função de informação mútua, FMMI (definido na seção 3.1.2 na página 24), para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos.

Na Tabela 4.19 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) do primeiro mínimo da função de informação mútua para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.19** Média e desvio padrão para o primeiro mínimo da função de informação mútua.

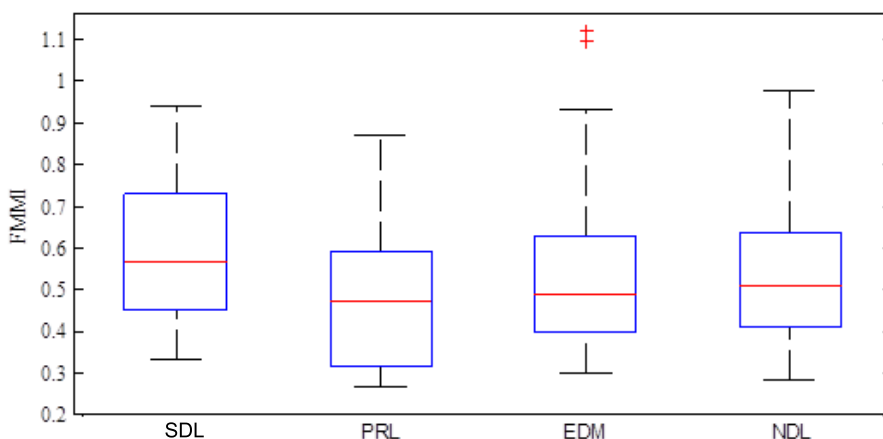
Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
FMMI	0,581	0,153	0,441	0,202	$P < 0,001$





**Figura 4.22** Distribuição dos valores médios do primeiro mínimo da função de informação mútua para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

Após a realização do teste  $t$  para amostras independentes, foi verificado que há evidências estatísticas para afirmar que a diferença entre as médias dos grupos saudável e patológico é significativa. Assim, o valor do primeiro mínimo da função de informação mútua pode ser utilizado, em potencial, como uma medida de discriminação entre vozes saudáveis e patológicas.



**Figura 4.23** Distribuição dos valores médios do primeiro mínimo da função de informação mútua para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Figura 4.23 está ilustrada a distribuição dos valores médios do primeiro mínimo da função de informação mútua para os sinais de vozes saudáveis (SDL) e os sinais de vozes patológicas discriminadas pelo tipo de patologia, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.20 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) para o primeiro mínimo da função de informação mútua dos sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

A fim de testar as possíveis diferenças entre as médias dos quatro grupos de vozes foi realizado o teste estatístico ANOVA não paramétrico de Kruskal-Wallis, que indicou a existência de diferenças estatísticas significativas ( $P < 0,001$ ) envolvendo, pelo menos, dois dos

**Tabela 4.20** Média e desvio padrão do primeiro mínimo da função de informação mútua - separação das patologias.

<i>FMMI</i>	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	0,580	0,368	0,492	0,536
$\sigma$	0,153	0,182	0,205	0,180

grupos analisados. Os resultados obtidos após a aplicação do teste de Mann-Whitney para uma comparação dois a dois entre os grupos estão apresentados na Tabela 4.21.

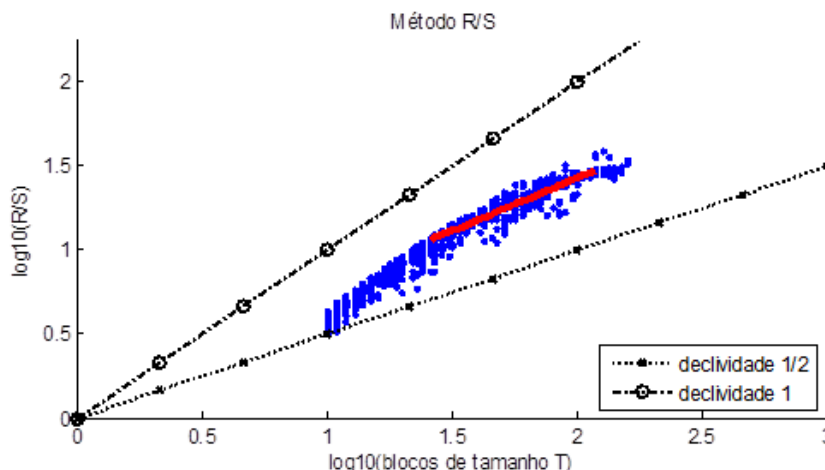
**Tabela 4.21** Valor-p - *FMMI* - separação das patologias.

<i>FMMI</i>	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P < 0,001$	$P = 0,002$	$P = 0,239$
Paralisia	-	$P = 0,001$	$P = 0,001$
Edema	-	-	$P = 0,257$

Os resultados apresentados na Tabela 4.21 revelaram diferenças significativas dos valores do primeiro mínimo da função de informação mútua entre o grupo de vozes saudáveis e os grupos de vozes afetadas por paralisia e por edema. Há diferenças estatísticas significativas entre o grupo de vozes afetadas com paralisia e os grupos de vozes com edema e com nódulos. Assim, é possível concluir que a medida do primeiro mínimo da função de informação mútua pode ser utilizada para diferenciar vozes saudáveis de vozes afetadas por paralisia e por edema. Entretanto, essa medida não se mostrou capaz de evidenciar diferenças estatísticas significativas entre vozes saudáveis e vozes patológicas afetadas por nódulos.

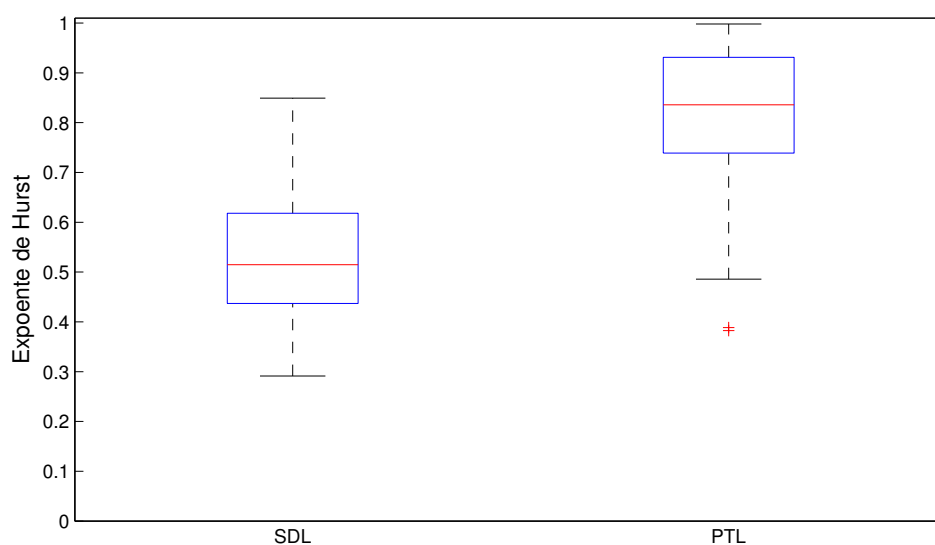
### 4.3.8 Expoente de Hurst (*H*)

Na Figura 4.24 está ilustrada a estimação do expoente de Hurst, *H* (definido na seção 3.2.11 na página 33), dado pela inclinação do segmento entre as retas de declividade 1/2 e 1, para um segmento de 32 ms (800 amostras) de um sinal de voz saudável. Para esse segmento o valor estimado foi de  $H = 0,6254$ .



**Figura 4.24** Estimação do expoente de Hurst para um sinal de voz saudável.

Na Figura 4.25 está ilustrada a distribuição dos valores médios do expoente de Hurst para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos.



**Figura 4.25** Distribuição dos valores do expoente de Hurst para sinais de vozes saudáveis (SDL) e vozes patológicas (PTL).

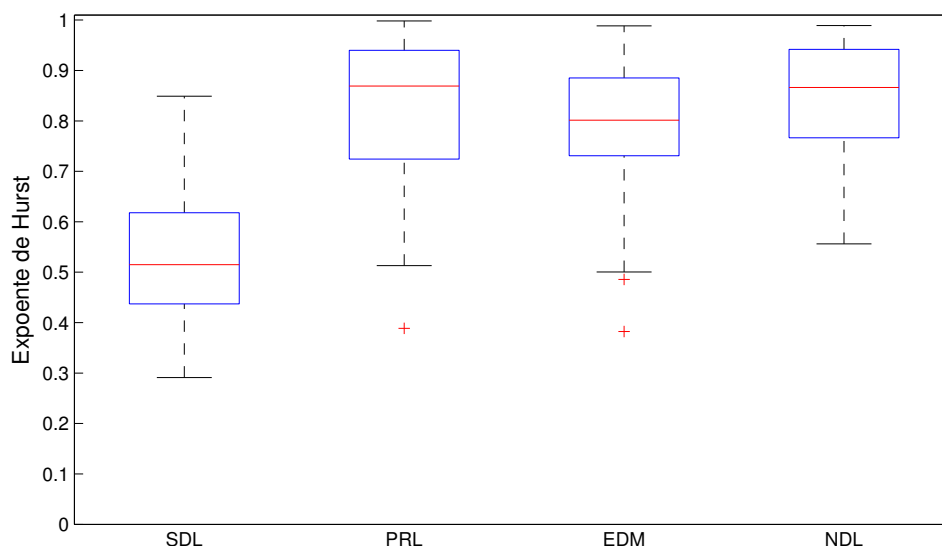
Na Tabela 4.22 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) do expoente de Hurst para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.22** Média e desvio padrão para o expoente de Hurst.

Medida	Voz Saudável		Voz Patológica		Valor-p
	$\mu$	$\sigma$	$\mu$	$\sigma$	
$H$	0,527	0,124	0,830	0,162	$P < 0,001$

A partir da realização do teste  $t$  para amostras independentes, foi verificado que há evidências estatísticas para afirmar que a diferença entre as médias dos grupos saudável e patológico é significativa.

Na Figura 4.26 está ilustrada a distribuição dos valores médios do expoente de Hurst para os sinais de vozes saudáveis (SDL) e os sinais de vozes patológicas discriminadas pelo tipo de patologia, ou seja, paralisia (PRL), edema (EDM) e nódulos (NDL).



**Figura 4.26** Distribuição dos valores médios do expoente de Hurst para sinais de vozes saudáveis (SDL) e patológicas com paralisia (PRL), edema (EDM) e nódulos (NDL).

Na Tabela 4.23 estão apresentados os valores da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) para o expoente de Hurst dos sinais de vozes saudáveis e dos sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.23** Média e desvio padrão do expoente de Hurst - separação das patologias.

$H$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$\mu$	0,527	0,844	0,802	0,857
$\sigma$	0,124	0,173	0,158	0,130

A aplicação do teste ANOVA de um caminho revelou a existência de diferenças estatísticas significativas entre as médias de, pelo menos, dois dos grupos analisados ( $P < 0,001$ ). Diante deste fato, foi aplicado o teste *post hoc* de Tukey para a comparação de múltiplas médias [110]. Os resultados estão apresentados na Tabela 4.24.

Após o teste *post hoc* aplicado, os resultados revelaram existirem diferenças estatísticas significativas entre a média do grupo de vozes saudáveis e as médias dos três grupos de vozes patológicas ( $P < 0,001$ ). Por outro, os resultados mostraram não existirem diferenças estatísticas significativas entre os três diferentes grupos de vozes patológicas.

**Tabela 4.24** Valor-p - Expoente de Hurst - separação das patologias.

$H$	Tipo de Voz		
	Paralisia	Edema	Nódulo
Saudável	$P < 0,001$	$P < 0,001$	$P < 0,001$
Paralisia	-	$P = 0,517$	$P = 0,988$
Edema	-	-	$P = 0,560$

#### 4.4 Resumo dos Resultados do Processo de Caracterização - Medidas Não Lineares (MNL)

Na Tabela 4.25 está apresentado o resumo dos resultados obtidos a partir dos testes estatísticos realizados, visando identificar o potencial discriminativo de cada uma das medidas de análise não linear (MNL) avaliadas neste trabalho, na caracterização de sinais de vozes saudáveis e patológicas. Os campos marcados com x representam os casos em que foram verificadas diferenças estatísticas significativas entre os grupos analisados, considerando-se um nível de significância igual a 5%.

**Tabela 4.25** Potencial discriminativo das medidas da análise dinâmica não linear.

Medidas	Medidas da análise dinâmica não linear							
	$D_2$	$K_2$	$H_1$	$ApEn$	$H_q$	$\lambda_1$	$FMMI$	$H$
<i>SDL x PTL</i>	x		x		x	x	x	x
<i>SDL x PRL</i>	x		x		x	x	x	x
<i>SDL x EDM</i>	x		x		x	x	x	x
<i>SDL x NDL</i>	x	x	x		x			x
<i>PRL x EDM</i>							x	
<i>PRL x NDL</i>	x	x					x	
<i>EDM x NDL</i>								

A Dimensão de Correlação ( $D_2$ ) e o Primeiro Mínimo da Função de Informação Mútua ( $FMMI$ ) são as medidas que apresentaram potencial discriminativo para a maior parte dos casos analisados. A maior parte das medidas consideradas sugere diferenças estatisticamente significativas entre sinais de vozes saudáveis e vozes patológicas, quando se agrupam as patologias numa mesma classe, ou em classes distintas. Entretanto, a grande dificuldade está na identificação de medidas que apresentem potencial discriminativo entre vozes afetadas por diferentes patologias. Na discriminação entre edema e nódulos, por exemplo, nenhuma das medidas da análise não linear investigadas apresentou potencial discriminativo entre essas patologias, com os testes estatísticos aplicados.

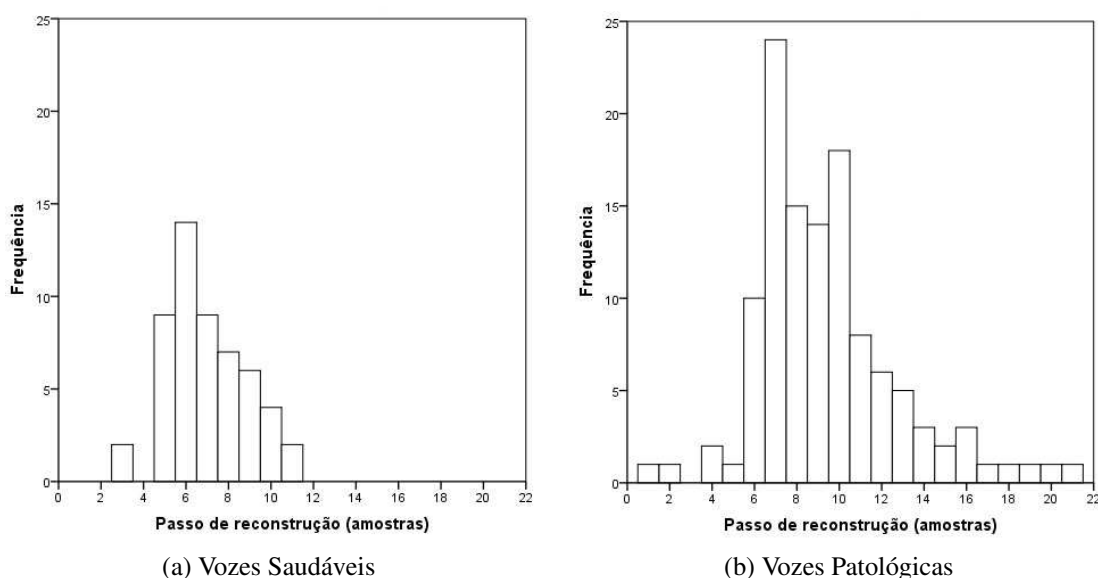
## 4.5 Medidas de Quantificação de Recorrência (MQR)

O objetivo principal desta seção é investigar a aplicação das medidas de quantificação de recorrência (MQR) na caracterização de sinais de vozes saudáveis e vozes patológicas. As medidas avaliadas são as seguintes: determinismo, comprimento máximo das linhas diagonais, entropia da distribuição de frequências dos comprimentos das linhas diagonais, tendência, laminaridade, comprimento máximo das linhas verticais e tempo de permanência em um estado.

Do banco de dados disponível [60] foram utilizados 171 sinais de vozes da vogal sustentada /ah/, sendo 53 sinais de vozes saudáveis e 118 de vozes afetadas por algum tipo de patologia nas pregas vocais (45 com edema, 55 com paralisia e 18 com nódulos). Para efeito de análise dos sinais foi considerado um bloco único de 10.000 amostras (400 ms), extraídas na faixa entre 100 ms e 500 ms, visando excluir os efeitos iniciais presentes no processo de gravação.

O primeiro passo realizado foi a determinação dos parâmetros de imersão ( $\tau$  e  $m$ ) visando a obtenção dos vetores responsáveis pela reconstrução do espaço de estados do sistema de produção da voz. Neste caso, foi utilizado o *software Visual Recurrence Analysis (VRA)* [104] para a obtenção dos valores ideais destes parâmetros. O VRA utiliza o método que estima o valor ideal de  $\tau$  como sendo o primeiro mínimo da função de informação mútua [70] e o método baseado na determinação dos falsos vizinhos próximos [112] para a determinação da dimensão de imersão mínima.

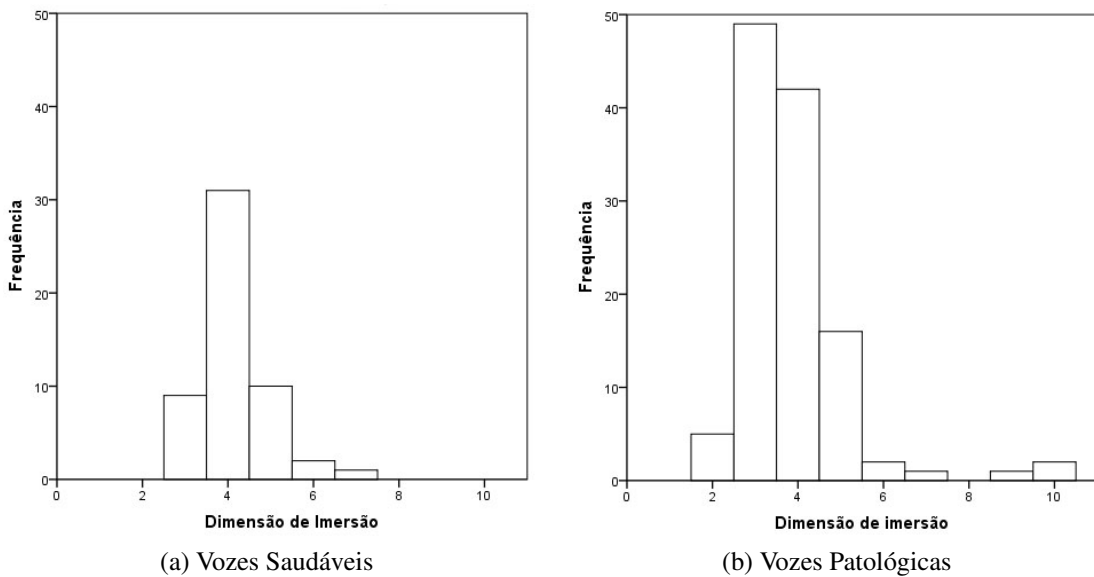
Na Figura 4.27 estão apresentados os histogramas dos valores do passo de reconstrução calculados para um segmento de 400 ms de cada um dos (a) 53 sinais de vozes saudáveis e (b) 118 sinais de vozes patológicas.



**Figura 4.27** Histogramas dos valores de  $\tau$  obtidos para os sinais de vozes (a) saudáveis e (b) patológicas.

Como pode ser observado, para os sinais de vozes saudáveis há uma predominância do valor de  $\tau = 6$  (26,4%), além disso, tem-se que  $\tau \leq 9$  para 88,70% dos casos. Por outro lado, para os sinais de vozes patológicas há uma predominância do valor de  $\tau = 7$  (20,3%) e  $\tau \geq 9$  para 54,2% dos casos.

Na Figura 4.28 estão apresentados os histogramas dos valores da dimensão de imersão mínima para os sinais de vozes (a) saudáveis e (b) patológicas analisadas. Para as vozes saudáveis o valor predominante de  $m = 4$  ocorreu para 58,5% dos sinais. Para as vozes patológicas o valor  $m = 3$  ocorreu com maior frequência, para 41,5% dos casos.



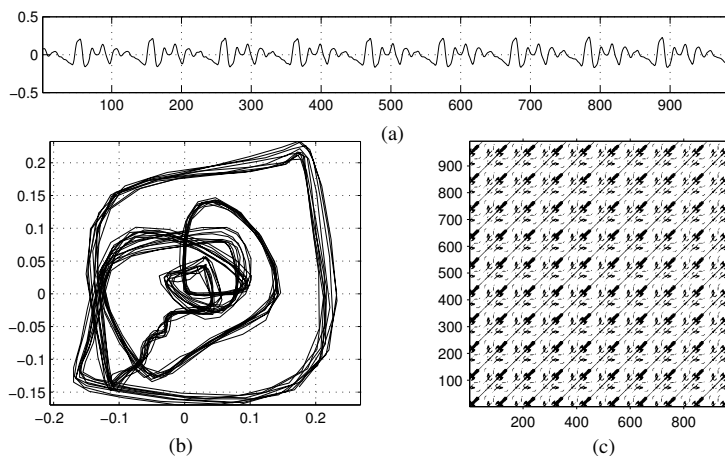
**Figura 4.28** Histogramas dos valores de  $m$  obtidos para os sinais de vozes (a) saudáveis e (b) patológicas.

Após a determinação dos parâmetros de imersão do sinal, um total de  $N = 10.000 - (m - 1)\tau$  vetores  $m$ -dimensionais são construídos para cada sinal.

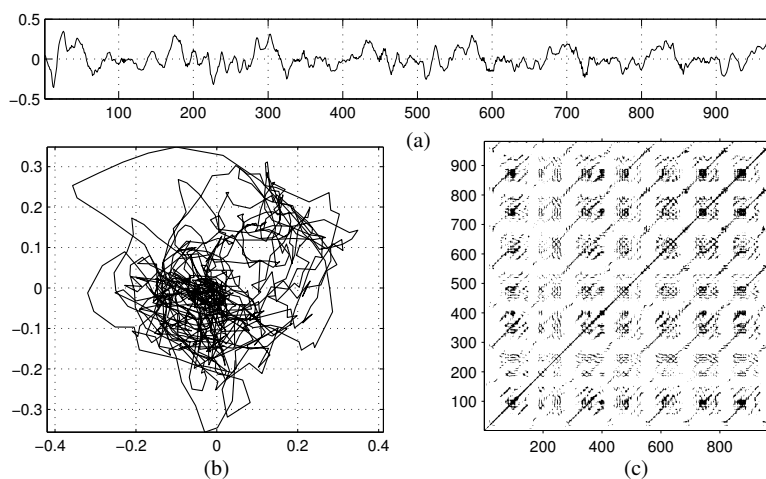
Nas Figuras 4.29 a 4.32 estão apresentados exemplos dos gráficos de recorrência representando a vogal sustentada /ah/ para um segmento de 40 ms de um sinal de voz (a) saudável, (b) afetada por edema, (c) com nódulos e (d) com paralisia nas dobras vocais, respectivamente. Pode ser observado diferenças nas estruturas presentes em cada sinal analisado.

Realizado o processo de imersão, o pacote de rotinas *RQA - software13.1* [105] é utilizado para obtenção dos parâmetros e medidas de quantificação relativas à análise de recorrência. Após a determinação dos vetores representativos da dinâmica do sinal de voz, uma matriz de distâncias é então formada calculando-se a norma euclidiana entre eles. A etapa seguinte é a escolha do limiar de distância (ou raio),  $\varepsilon$ , a fim de se construir a matriz de recorrência. No *software RQA* o valor do raio é sempre relativo à máxima distância entre os vetores (*Maxdist*).

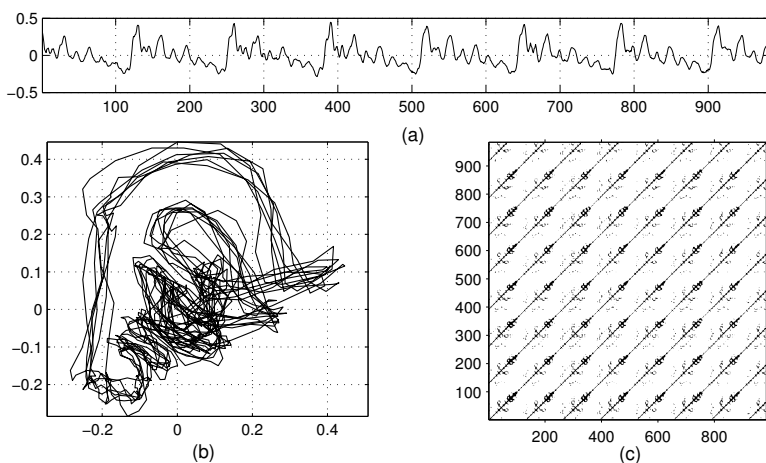
Quanto maior o raio, mais ampla é a vizinhança para se determinar a recorrência dos estados de um sistema. Assim, caso o valor do limiar de distância seja fixado muito alto, algumas medidas como a taxa de recorrência e o determinismo tenderão a saturar em 100%.



**Figura 4.29** Sinal de voz saudável (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência;  $m = 3, \tau = 4, \varepsilon = 0,1$  (norma euclidiana)

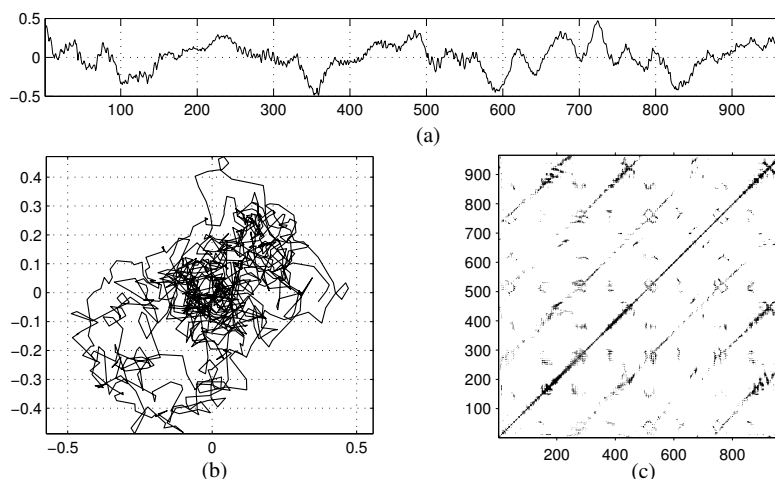


**Figura 4.30** Sinal de voz patológica (edema) (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência;  $m = 3, \tau = 9, \varepsilon = 0,1$  (norma euclidiana)



**Figura 4.31** Sinal de voz patológica (nódulos) (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência;  $m = 3, \tau = 8, \varepsilon = 0,1$  (norma euclidiana)





**Figura 4.32** Sinal de voz patológica (paralisia) (a) série temporal, (b) projeção bidimensional do atrator (c) gráfico de recorrência;  $m = 3$ ,  $\tau = 17$ ,  $\varepsilon = 0,1$  (norma euclidiana)

Portanto, o valor escolhido para o raio deve ser baixo, mas não tão baixo a ponto de levar a taxa de recorrência a se aproximar de zero [105].

Entre os três diferentes procedimentos propostos por Webber e Zbilut [113] para a determinação adequada do raio, o método escolhido para ser utilizado neste trabalho está baseado na obtenção de um valor para este parâmetro que assegure a manutenção de uma baixa taxa de recorrência (tipicamente menor ou igual a 5%). A taxa de recorrência, *REC* (definida na seção 3.3.1 na página 36), quantifica a porcentagem de pontos de recorrência dentro do raio especificado.

Nos experimentos computacionais realizados foi investigado o comportamento do limiar de distância de modo que a taxa de recorrência máxima estivesse abaixo de (ou igual a) cinco diferentes patamares, ou seja, 1%, 2%, 3%, 4% e 5%.

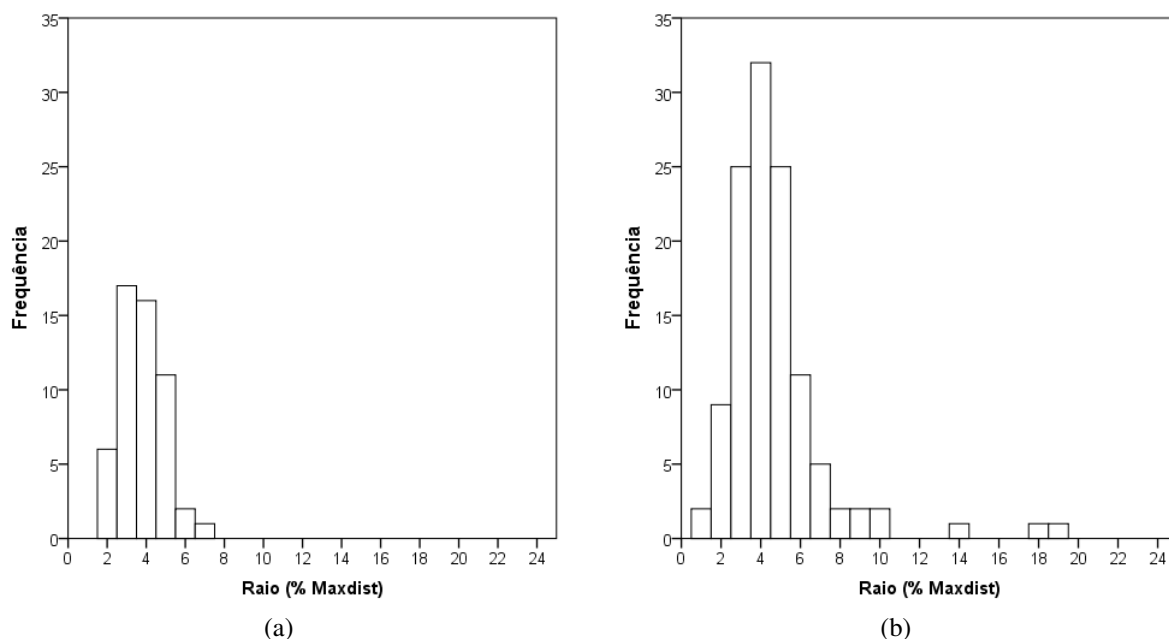
Na Figura 4.33 estão apresentadas as distribuições dos valores do maior raio que mantém uma taxa de recorrência menor ou igual a 1%.

Como pode ser observado, neste caso, há uma maior concentração dos valores do raio na faixa entre 3% a 5% da máxima distância entre os vetores (*Maxdist*), tanto para vozes saudáveis ( $\mu = 4$ ;  $\sigma = 1,13$ ) como para vozes patológicas ( $\mu = 5$ ;  $\sigma = 2,63$ ).

Na medida em que é permitido uma taxa de recorrência maior os valores do raio aparecem mais largamente distribuídos, numa faixa de valores superior ao caso anterior, para ambos os grupos de sinais. No caso da taxa de recorrência máxima ser de 5% (Figura 4.34), tem-se que ( $\mu = 11$ ;  $\sigma = 2,64$ ) para as vozes saudáveis e ( $\mu = 10$ ;  $\sigma = 3,41$ ) no caso das vozes patológicas.

#### 4.5.1 Determinismo (*DET*)

Na Figura 4.35 está ilustrada a distribuição dos valores da medida determinismo, *DET* (definida na seção 3.3.2 na página 37), para os 53 sinais de vozes saudáveis (SDL) e 118 si-



**Figura 4.33** Distribuições dos valores do maior raio para sinais de vozes (a) saudáveis e (b) patológicas para uma taxa de recorrência menor ou igual a 1%.

nais de vozes patológicas (PTL) utilizados nos experimentos, considerando cinco diferentes limitantes aplicados à taxa de recorrência: 1%, 2%, 3%, 4% e 5%.

Na Tabela 4.26 estão apresentados os valores da média e do desvio padrão do determinismo para os sinais de vozes saudáveis e de vozes patológicas analisados.

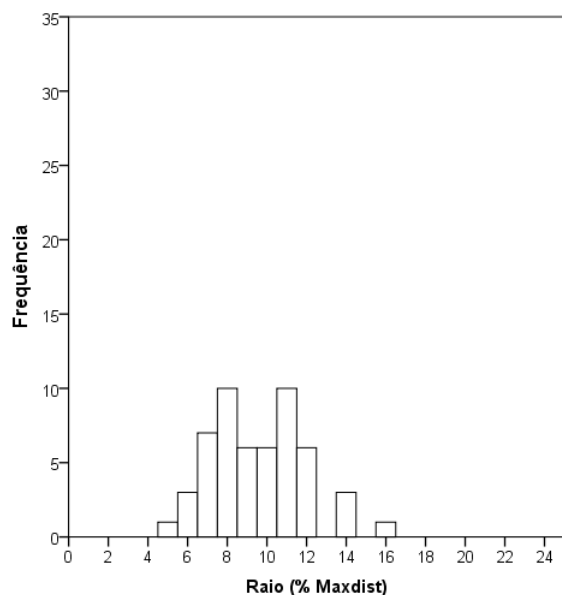
**Tabela 4.26** Média e desvio padrão para *DET*.

Medida <i>DET</i>	Tipo de Voz		
	Saudável	Patológica	Valor P
$REC \leq 1\%$	$88,21 \pm 5,68$	$60,25 \pm 25,18$	$P < 0,001$
$REC \leq 2\%$	$92,28 \pm 4,24$	$71,82 \pm 23,03$	$P < 0,001$
$REC \leq 3\%$	$93,79 \pm 3,72$	$77,23 \pm 21,14$	$P < 0,001$
$REC \leq 4\%$	$94,69 \pm 3,38$	$80,48 \pm 19,33$	$P < 0,001$
$REC \leq 5\%$	$95,16 \pm 3,15$	$82,87 \pm 18,11$	$P < 0,001$

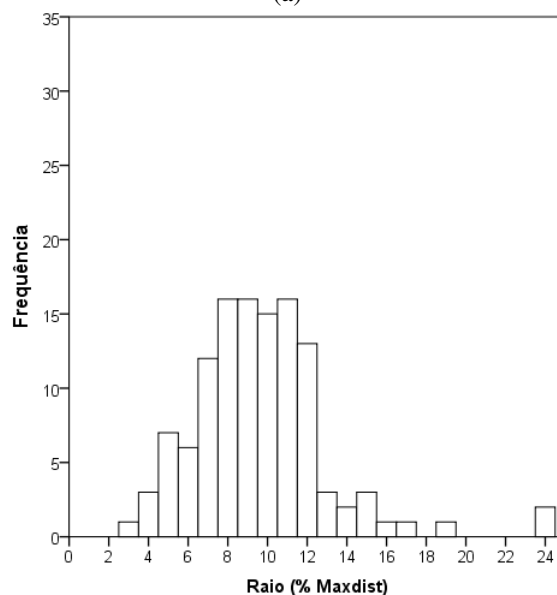
Após a realização do teste não-paramétrico U de Mann-Whitney, os resultados revelaram que os grupos de vozes saudáveis e patológicas apresentam diferenças estatísticas significativas para todos os limiares de taxa de recorrência considerados ( $P < 0,001$ ).

Na Tabela 4.27 estão apresentados os valores da média e do desvio padrão da medida de determinismo para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

A fim de testar as possíveis diferenças estatísticas entre os quatro grupos de sinais de voz, foi realizado o teste de análise de variância ANOVA não paramétrico de Kruskal-Wallis. Segundo os resultados obtidos, existem diferenças estatísticas significativas entre pelo menos dois dos grupos analisados, para todos os patamares da taxa de recorrência considerados ( $P < 0,001$ ).



(a)



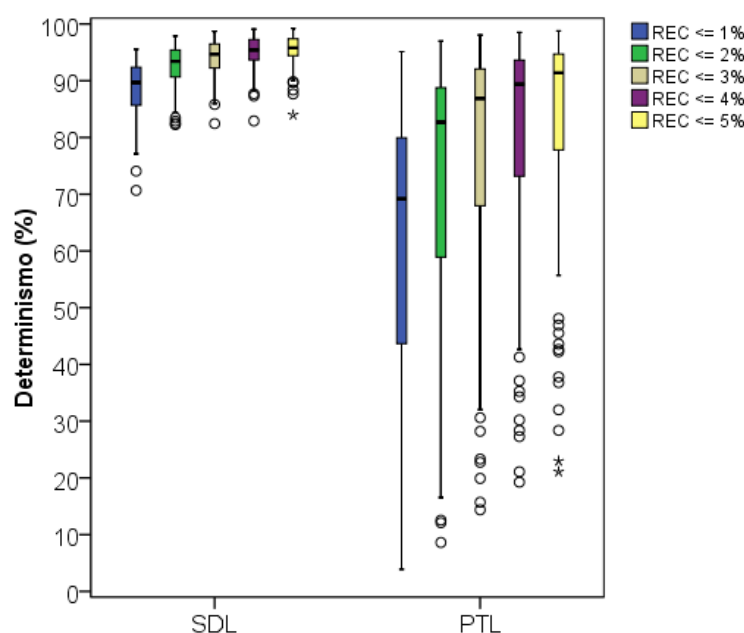
(b)

**Figura 4.34** Distribuições dos valores do maior raio para sinais de vozes (a) saudáveis e (b) patológicas para uma taxa de recorrência menor ou igual a 5%.

**Tabela 4.27** Média e desvio padrão para *DET* - separação das patologias.

Medida <i>DET</i>	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$REC \leq 1\%$	88,21 ± 5,68	52,37 ± 26,83	67,20 ± 22,45	66,95 ± 19,89
$REC \leq 2\%$	92,28 ± 4,24	64,05 ± 25,00	77,63 ± 20,53	81,05 ± 14,09
$REC \leq 3\%$	93,79 ± 3,72	70,18 ± 23,57	82,04 ± 18,37	86,77 ± 10,87
$REC \leq 4\%$	94,69 ± 3,38	74,27 ± 21,90	84,67 ± 16,61	89,02 ± 9,30
$REC \leq 5\%$	95,16 ± 3,15	76,84 ± 21,08	86,91 ± 14,68	91,18 ± 7,32

Na Tabela 4.28 estão apresentados os resultados da análise estatística realizada visando a comparação dos valores da característica *DET* obtidos para sinais de vozes saudáveis (SDL), sinais de vozes afetados por paralisia (PRL), edema (EDM) e nódulos (NDL) nas pregas vocais.



**Figura 4.35** Distribuições dos valores de  $DET$  para sinais de vozes saudáveis (SDL) e patológicas (PTL).

**Tabela 4.28** Potencial estatístico da característica  $DET$  na discriminação entre grupos de sinais de voz.

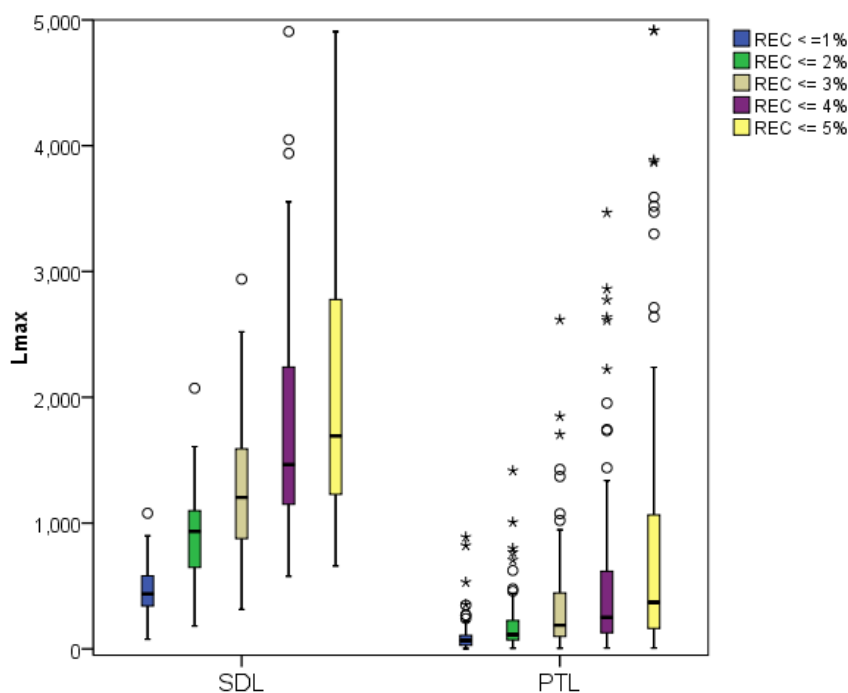
Medida $DET$	Pares de Grupos de Sinais					
	SDL x PRL	SDL x EDM	SDL x NDL	PRL x EDM	PRL x NDL	EDM x NDL
$REC \leq 1\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,007^*$	$P = 0,044^*$	$P = 0,897$
$REC \leq 2\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,001^*$	$P = 0,005^*$	$P = 0,744$
$REC \leq 3\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,002^*$	$P = 0,001^*$	$P = 0,442$
$REC \leq 4\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,002^*$	$P = 0,001^*$	$P = 0,390$
$REC \leq 5\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,002^*$	$P = 0,001^*$	$P = 0,297$

\* $P < 0,05$

A utilização do teste de Mann-Whitney para uma comparação dois a dois entre os grupos, revelou diferenças estatísticas significativas dos valores do determinismo entre o grupo de vozes saudáveis e cada um dos grupos de vozes afetadas por patologias em todos os patamares da taxa de recorrência. O mesmo fato ocorreu na comparação entre os grupos de vozes afetadas pelas patologias paralisia e edema e entre paralisia e nódulos. Entretanto, não foram observadas diferenças significativas entre os grupos de vozes com edema e nódulos para qualquer patamar da taxa de recorrência.

#### 4.5.2 Comprimento Máximo das Linhas Diagonais ( $L_{max}$ )

Na Figura 4.36 está ilustrada a distribuição dos valores do comprimento máximo das linhas diagonais,  $L_{max}$  (definido na seção 3.3.4 na página 37), para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos, considerando cinco diferentes limitantes aplicados à taxa de recorrência: 1%, 2%, 3%, 4% e 5%.



**Figura 4.36** Distribuições dos valores de  $L_{max}$  para sinais de vozes saudáveis (SDL) e patológicas (PTL).

Na Tabela 4.29 estão apresentados os valores da média e do desvio padrão para o comprimento máximo das linhas diagonais dos sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.29** Média e desvio padrão para  $L_{max}$ .

Medida $L_{max}$	Tipo de Voz		Valor P
	Saudável	Patológica	
$REC \leq 1\%$	$466,38 \pm 211,80$	$97,52 \pm 128,63$	$P < 0,001$
$REC \leq 2\%$	$938,91 \pm 369,98$	$185,96 \pm 207,34$	$P < 0,001$
$REC \leq 3\%$	$1320,75 \pm 573,03$	$329,15 \pm 397,90$	$P < 0,001$
$REC \leq 4\%$	$1779,02 \pm 965,48$	$523,81 \pm 659,56$	$P < 0,001$
$REC \leq 5\%$	$2121,04 \pm 1161,18$	$830,64 \pm 1049,18$	$P < 0,001$

Após a realização do teste não-paramétrico U de Mann-Whitney, os resultados revelaram que os grupos de vozes saudáveis e patológicas apresentam diferenças estatísticas significativas para todos os limiares de taxa de recorrência considerados ( $P < 0,001$ ).

Na Tabela 4.30 estão apresentados os valores da média e do desvio padrão do comprimento máximo das linhas diagonais para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos. A fim de testar as possíveis diferenças estatísticas entre os quatro grupos de sinais de voz, foi realizado o teste de análise de variância ANOVA não paramétrico de Kruskal-Wallis. Segundo os resultados obtidos, existem diferenças estatísticas significativas entre pelo menos dois dos grupos analisados para todos os patamares da taxa de recorrência considerados ( $P < 0,001$ ).

**Tabela 4.30** Média e desvio padrão para  $L_{max}$  - separação das patologias.

Medida $L_{max}$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$REC \leq 1\%$	466,38 ± 211,80	71,60 ± 71,02	140,58 ± 183,90	69,06 ± 48,88
$REC \leq 2\%$	938,91 ± 369,98	127,20 ± 132,73	265,40 ± 274,61	166,89 ± 128,27
$REC \leq 3\%$	1320,75 ± 573,03	202,96 ± 218,89	457,38 ± 505,01	394,17 ± 421,67
$REC \leq 4\%$	1779,02 ± 965,48	333,84 ± 468,69	729,58 ± 749,20	589,83 ± 780,99
$REC \leq 5\%$	2121,04 ± 1161,18	518,04 ± 728,26	1168,78 ± 1270,35	940,44 ± 1040,57

Na Tabela 4.31 estão apresentados os resultados da análise estatística realizada visando a comparação dos valores da característica  $L_{max}$  obtidos para sinais de vozes saudáveis (SDL), sinais de vozes afetados por paralisia (PRL), edema (EDM) e nódulos (NDL) nas pregas vocais.

**Tabela 4.31** Potencial estatístico da característica  $L_{max}$  na discriminação entre grupos de sinais de voz.

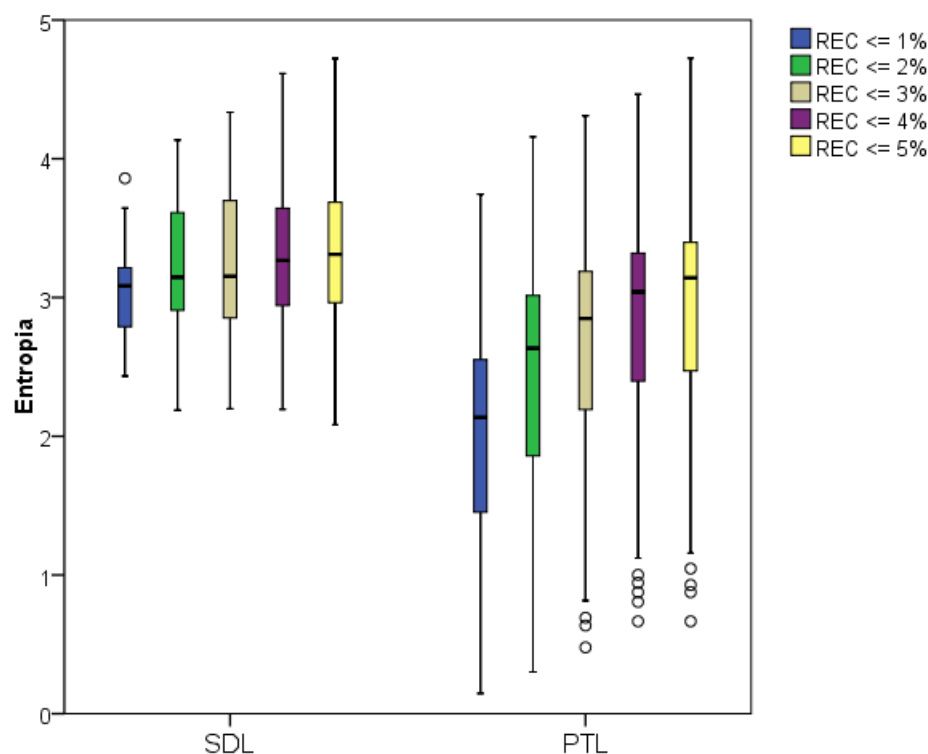
Medida $L_{max}$	Pares de Grupos de Sinais					
	SDL x PRL	SDL x EDM	SDL x NDL	PRL x EDM	PRL x NDL	EDM x NDL
$REC \leq 1\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,010^*$	$P = 0,556$	$P = 0,132$
$REC \leq 2\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,001^*$	$P = 0,097$	$P = 0,201$
$REC \leq 3\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,001^*$	$P = 0,018^*$	$P = 0,802$
$REC \leq 4\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,001^*$	$P = 0,031^*$	$P = 0,676$
$REC \leq 5\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,001^*$	$P = 0,018^*$	$P = 0,548$

\* $P < 0,05$

A utilização do teste de Mann-Whitney para uma comparação dois a dois entre os grupos, revelou diferenças significativas dos valores de  $L_{max}$  entre o grupo de vozes saudáveis e cada um dos três grupos de vozes patológicas para todos os patamares da taxa de recorrência. O mesmo fato ocorreu na discriminação entre os grupos com paralisia e com edema. Entretanto, não foram observadas diferenças significativas entre os grupos de vozes com paralisia e com nódulos para  $REC \leq 1\%$  e  $REC \leq 2\%$ . A comparação entre os grupos de vozes com edema e com nódulos não apresentou diferenças estatísticas significativas para qualquer limiar da taxa de recorrência.

### 4.5.3 Entropia da Distribuição de Frequências dos Comprimentos das Linhas Diagonais ( $ENTR$ )

Na Figura 4.37 está ilustrada a distribuição dos valores da entropia de Shannon da distribuição de frequências dos comprimentos das linhas diagonais,  $ENTR$  (definida na seção 3.3.3 na página 37), para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos, considerando cinco diferentes limitantes aplicados à taxa de recorrência: 1%, 2%, 3%, 4% e 5%.



**Figura 4.37** Distribuições dos valores da medida entropia para sinais de vozes saudáveis (SDL) e patológicas (PTL).

Na Tabela 4.32 estão apresentados os valores da média e do desvio padrão da entropia da distribuição de frequências dos comprimentos das linhas diagonais para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.32** Média e desvio padrão para *ENTR*.

Medida <i>ENTR</i>	Tipo de Voz		
	Saudável	Patológica	Valor P
$REC \leq 1\%$	$3,04 \pm 0,32$	$1,98 \pm 0,81$	$P < 0,001$
$REC \leq 2\%$	$3,23 \pm 0,43$	$2,40 \pm 0,83$	$P < 0,001$
$REC \leq 3\%$	$3,27 \pm 0,52$	$2,63 \pm 0,82$	$P < 0,001$
$REC \leq 4\%$	$3,33 \pm 0,58$	$2,78 \pm 0,81$	$P < 0,001$
$REC \leq 5\%$	$3,37 \pm 0,58$	$2,90 \pm 0,80$	$P = 0,002$

Como apresentado na Tabela 4.32, os resultados obtidos a partir da aplicação do teste não-paramétrico U de Mann-Whitney revelaram que os grupos de vozes saudáveis e patológicas apresentam diferenças estatísticas significativas para todos os limiares de taxa de recorrência considerados.

Na Tabela 4.33 estão apresentados os valores da média e do desvio padrão da medida de entropia da distribuição de frequências dos comprimentos das linhas diagonais, para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.33** Média e desvio padrão para *ENTR* - separação das patologias.

Medida <i>ENTR</i>	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$REC \leq 1\%$	$3,04 \pm 0,32$	$1,73 \pm 0,84$	$2,22 \pm 0,76$	$2,15 \pm 0,65$
$REC \leq 2\%$	$3,23 \pm 0,43$	$2,09 \pm 0,85$	$2,65 \pm 0,77$	$2,74 \pm 0,56$
$REC \leq 3\%$	$3,27 \pm 0,52$	$2,32 \pm 0,86$	$2,84 \pm 0,74$	$3,04 \pm 0,54$
$REC \leq 4\%$	$3,33 \pm 0,58$	$2,49 \pm 0,85$	$2,99 \pm 0,74$	$3,17 \pm 0,52$
$REC \leq 5\%$	$3,37 \pm 0,58$	$2,62 \pm 0,85$	$3,10 \pm 0,72$	$3,29 \pm 0,51$

A fim de testar as possíveis diferenças estatísticas entre os quatro grupos de sinais de voz, foi realizado o teste de análise de variância ANOVA não paramétrico de Kruskal-Wallis. Segundo os resultados obtidos, existem diferenças estatísticas significativas entre pelo menos dois dos grupos analisados, para todos os patamares da taxa de recorrência considerados ( $P < 0,001$ ).

Na Tabela 4.34 estão apresentados os resultados da análise estatística realizada visando a comparação dos valores da característica *ENTR* obtidos para sinais de vozes saudáveis (SDL), sinais de vozes afetados por paralisia (PRL), edema (EDM) e nódulos (NDL) nas pregas vocais.

**Tabela 4.34** Potencial estatístico da característica *ENTR* na discriminação entre grupos de sinais de voz.

Medida <i>ENTR</i>	Pares de Grupos de Sinais					
	SDL x PRL	SDL x EDM	SDL x NDL	PRL x EDM	PRL x NDL	EDM x NDL
$REC \leq 1\%$	$P < 0,001^*$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,007^*$	$P = 0,063$	$P = 0,740$
$REC \leq 2\%$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,003^*$	$P = 0,001^*$	$P = 0,003^*$	$P = 0,873$
$REC \leq 3\%$	$P < 0,001^*$	$P = 0,011^*$	$P = 0,124$	$P = 0,001^*$	$P = 0,001^*$	$P = 0,342$
$REC \leq 4\%$	$P < 0,001^*$	$P = 0,101$	$P = 0,318$	$P = 0,002^*$	$P = 0,003^*$	$P = 0,579$
$REC \leq 5\%$	$P < 0,001^*$	$P = 0,170$	$P = 0,596$	$P = 0,003^*$	$P = 0,002^*$	$P = 0,548$

\* $P < 0,05$

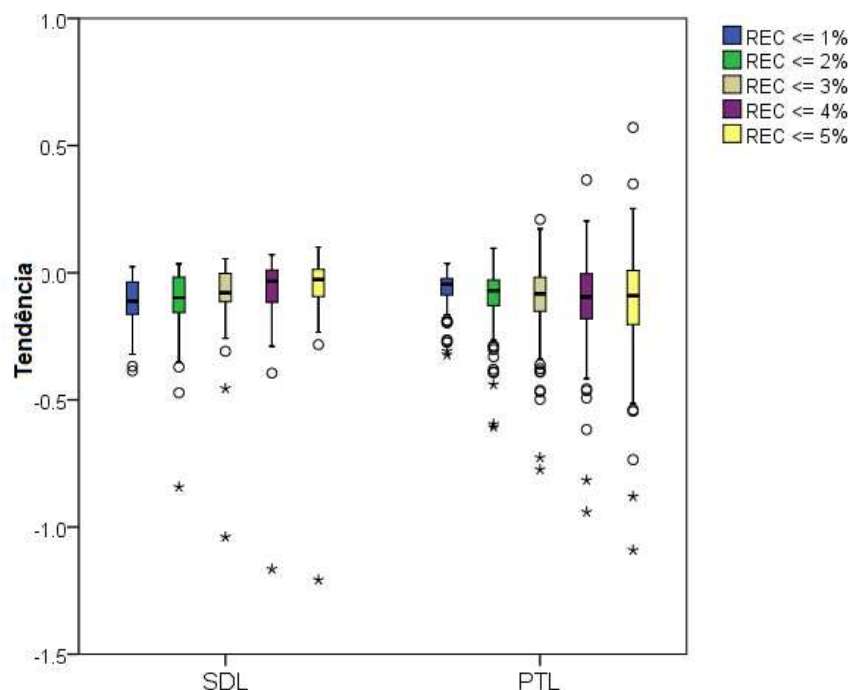
A utilização do teste de Mann-Whitney para uma comparação dois a dois entre os grupos, revelou diferenças significativas dos valores da *ENTR* entre o grupo de vozes saudáveis e o grupo de vozes com paralisia para todos os patamares da taxa de recorrência. Entretanto, não houve diferenças significativas entre os grupos de vozes saudáveis e com edema para  $REC > 3\%$ , nem entre os grupos de vozes saudáveis e com nódulos para os casos em que  $REC > 2\%$ .

Considerando a comparação entre os grupos de vozes afetadas por patologias, os resultados apontaram diferenças significativas entre os grupos de vozes com paralisia e com edema para todos os patamares da taxa de recorrência. Na discriminação entre os grupos de vozes com paralisia e com nódulos não foram identificadas diferenças significativas para  $REC \leq 1\%$ . A comparação entre os grupos de vozes com edema e com nódulos não apresentou diferenças estatísticas significativas para qualquer limiar da taxa de recorrência.



#### 4.5.4 Tendência (*TREND*)

Na Figura 4.38 está ilustrada a distribuição dos valores da tendência, *TREND* (definida na seção 3.3.5 na página 38), para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos, considerando cinco diferentes limitantes aplicados à taxa de recorrência: 1%, 2%, 3%, 4% e 5%.



**Figura 4.38** Distribuições dos valores da medida Trend para sinais de vozes saudáveis (SDL) e patológicas (PTL).

Na Tabela 4.35 estão apresentados os valores da média e do desvio padrão da medida tendência para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.35** Média e desvio padrão para *TREND*.

Medida <i>TREND</i>	Tipo de Voz		Valor P
	Saudável	Patológica	
$REC \leq 1\%$	$-0,113 \pm 0,095$	$-0,064 \pm 0,069$	$P = 0,001$
$REC \leq 2\%$	$-0,114 \pm 0,146$	$-0,097 \pm 0,121$	$P = 0,557$
$REC \leq 3\%$	$-0,094 \pm 0,163$	$-0,107 \pm 0,153$	$P = 0,272$
$REC \leq 4\%$	$-0,076 \pm 0,178$	$-0,112 \pm 0,180$	$P = 0,039$
$REC \leq 5\%$	$-0,066 \pm 0,183$	$-0,115 \pm 0,212$	$P = 0,019$

Como apresentado na Tabela 4.35 os resultados obtidos a partir da aplicação do teste não-paramétrico U de Mann-Whitney revelaram que os grupos de vozes saudáveis e patológicas apresentam diferenças estatísticas significativas apenas quando  $REC \leq 1\%$ ,  $4\%$  e  $5\%$ .

Na Tabela 4.36 estão apresentados os valores da média e do desvio padrão da medida tendência para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.36** Média e desvio padrão para *TREND* - separação das patologias.

Medida <i>TREND</i>	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$REC \leq 1\%$	$-0,113 \pm 0,095$	$-0,068 \pm 0,083$	$-0,059 \pm 0,060$	$-0,066 \pm 0,044$
$REC \leq 2\%$	$-0,114 \pm 0,146$	$-0,107 \pm 0,147$	$-0,081 \pm 0,092$	$-0,105 \pm 0,098$
$REC \leq 3\%$	$-0,094 \pm 0,163$	$-0,125 \pm 0,186$	$-0,084 \pm 0,119$	$-0,111 \pm 0,113$
$REC \leq 4\%$	$-0,076 \pm 0,178$	$-0,136 \pm 0,219$	$-0,086 \pm 0,143$	$-0,108 \pm 0,120$
$REC \leq 5\%$	$-0,066 \pm 0,183$	$-0,145 \pm 0,255$	$-0,083 \pm 0,177$	$-0,104 \pm 0,130$

Na Tabela 4.37 estão apresentados os resultados obtidos, a partir da aplicação do teste de análise de variância ANOVA não paramétrico de Kruskal-Wallis, a fim de testar as possíveis diferenças estatísticas entre os quatro grupos de vozes. Segundo os resultados obtidos, existem diferenças estatísticas significativas entre pelo menos dois dos grupos analisados, apenas quando  $REC \leq 1\%$ ; nos demais casos não existem indícios de diferenças significativas.

**Tabela 4.37** Potencial estatístico da característica *ENTR* na discriminação entre grupos de sinais de voz.

Medida <i>TREND</i>	Limiares da Taxa de Recorrência				
	$REC \leq 1\%$	$REC \leq 2\%$	$REC \leq 3\%$	$REC \leq 4\%$	$REC \leq 5\%$
Valor-P	$P = 0,009^*$	$P = 0,834$	$P = 0,584$	$P = 0,165$	$P = 0,102$

\* $P < 0,05$

Na Tabela 4.38 estão apresentados os resultados da análise estatística realizada visando a comparação dos valores da característica *TREND* obtidos para sinais de vozes saudáveis (SDL), sinais de vozes afetados por paralisia (PRL), edema (EDM) e nódulos (NDL) nas pregas vocais, com  $REC \leq 1\%$ .

**Tabela 4.38** Potencial estatístico da característica *ENTR* na discriminação entre grupos de sinais de voz.

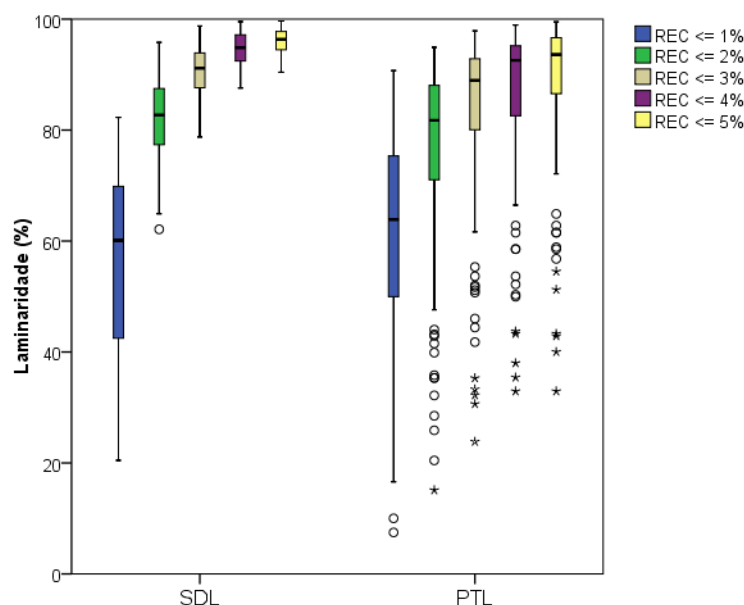
Medida <i>TREND</i>	Pares de Grupos de Sinais					
	SDL x PRL	SDL x EDM	SDL x NDL	PRL x EDM	PRL x NDL	EDM x NDL
$REC \leq 1\%$	$P < 0,006^*$	$P < 0,003^*$	$P < 0,081$	$P = 0,925$	$P = 0,485$	$P = 0,377$

\* $P < 0,05$

Assim, para  $REC \leq 1\%$ , a utilização do teste de Mann-Whitney para uma comparação dois a dois entre os grupos, revelou diferenças significativas dos valores de *TREND* entre o grupo de vozes saudáveis e dois dos três grupos de vozes patológicas: saudável x paralisia e saudável x edema. Na comparação saudável x nódulos, bem como na comparação entre os grupos de vozes afetadas por patologias, não foram reveladas diferenças estatisticamente significativas.

### 4.5.5 Laminaridade (*LAM*)

Na Figura 4.39 está ilustrada a distribuição dos valores da laminaridade, *LAM* (definida na seção 3.3.6 na página 39), para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos, considerando cinco diferentes limitantes aplicados à taxa de recorrência: 1%, 2%, 3%, 4% e 5%.



**Figura 4.39** Distribuições dos valores da medida laminaridade para sinais de vozes saudáveis (SDL) e patológicas (PTL).

Na Tabela 4.39 estão apresentados os valores da média e do desvio padrão da laminaridade para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.39** Média e desvio padrão para *LAM*.

Medida <i>LAM</i>	Tipo de Voz		
	Saudável	Patológica	Valor P
$REC \leq 1\%$	$55,97 \pm 16,42$	$60,32 \pm 19,46$	$P = 0,078$
$REC \leq 2\%$	$81,98 \pm 8,25$	$75,91 \pm 17,83$	$P = 0,213$
$REC \leq 3\%$	$90,50 \pm 4,91$	$82,95 \pm 16,60$	$P = 0,008$
$REC \leq 4\%$	$94,32 \pm 3,23$	$85,95 \pm 14,93$	$P < 0,001$
$REC \leq 5\%$	$95,89 \pm 2,41$	$88,13 \pm 13,99$	$P < 0,001$

Como apresentado na Tabela 4.39 os resultados obtidos a partir da aplicação do teste não-paramétrico U de Mann-Whitney revelaram que os grupos de vozes saudáveis e patológicas apresentam diferenças estatísticas significativas quando  $REC \leq 3\%$ ,  $4\%$  e  $5\%$ . Por outro lado, quando  $REC \leq 1\%$  e  $REC \leq 2\%$  as diferenças apresentadas não foram estatísticas significativas.

Na Tabela 4.40 estão apresentados os valores da média e do desvio padrão da medida laminaridade para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.40** Média e desvio padrão para *LAM* - separação das patologias.

Medida <i>LAM</i>	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$REC \leq 1\%$	55,97 ± 16,42	58,42 ± 21,85	62,12 ± 18,82	61,63 ± 12,27
$REC \leq 2\%$	81,98 ± 8,25	71,73 ± 20,06	78,79 ± 16,74	81,48 ± 8,62
$REC \leq 3\%$	90,50 ± 4,91	78,26 ± 18,50	84,88 ± 15,71	89,20 ± 7,03
$REC \leq 4\%$	94,32 ± 3,23	82,12 ± 16,70	88,15 ± 14,05	92,11 ± 6,09
$REC \leq 5\%$	95,89 ± 2,41	84,24 ± 16,08	90,38 ± 12,41	94,37 ± 4,89

Na Tabela 4.41 estão apresentados os resultados obtidos, a partir da aplicação do teste de análise de variância ANOVA não paramétrico de Kruskal-Wallis, a fim de testar as possíveis diferenças estatísticas entre os quatro grupos de vozes. Segundo os resultados obtidos, existem diferenças estatísticas significativas entre pelo menos dois dos grupos analisados, para os patamares da taxa de recorrência  $REC \leq 3\%$ ,  $4\%$  e  $5\%$ ; nos demais casos não existem indícios de diferenças significativas.

**Tabela 4.41** Potencial estatístico da característica *LAM* na discriminação entre grupos de sinais de voz.

Medida <i>LAM</i>	Limiares da Taxa de Recorrência				
	$REC \leq 1\%$	$REC \leq 2\%$	$REC \leq 3\%$	$REC \leq 4\%$	$REC \leq 5\%$
Valor-P	$P = 0,295$	$P = 0,062$	$P = 0,001^*$	$P < 0,001^*$	$P < 0,001^*$

\* $P < 0,05$

Na Tabela 4.42 estão apresentados os resultados da aplicação do teste U de Mann-Whitney para uma comparação dois a dois entre os grupos de sinais nas situações em que  $REC \leq 3\%$ ,  $4\%$  e  $5\%$ . Na comparação entre o grupo de vozes saudáveis e o grupo de vozes com paralisia foram observadas diferenças significativas dos valores de *LAM* para os três patamares. No caso da comparação saudável x edema, houve diferença significativa apenas para  $REC \leq 5\%$  e entre os grupos de sinais de vozes saudáveis e com nódulos não foram reveladas diferenças significativas para nenhum patamar da taxa de recorrência.

Considerando a comparação entre os grupos de vozes afetadas por patologias, os resultados apontaram diferenças significativas entre os grupos paralisia x edema e paralisia x nódulos para os três patamares da taxa de recorrência. A comparação entre os grupos edema x nódulos não apresentou diferenças estatísticas significativas para qualquer limiar da taxa de recorrência.

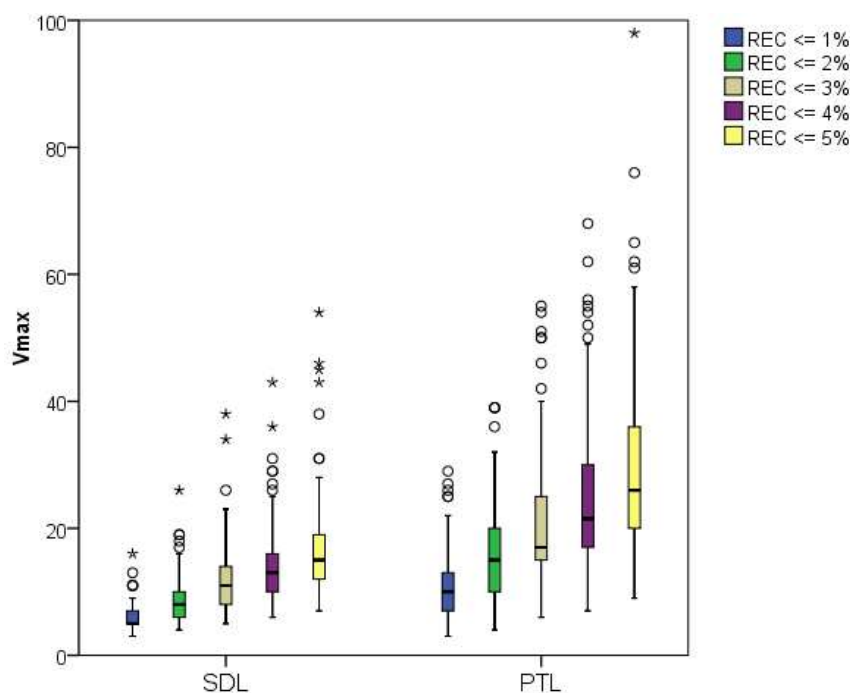
**Tabela 4.42** Potencial estatístico da característica *LAM* na discriminação entre grupos de sinais de voz.

Medida <i>LAM</i>	Pares de Grupos de Sinais					
	SDL x PRL	SDL x EDM	SDL x NDL	PRL x EDM	PRL x NDL	EDM x NDL
$REC \leq 3\%$	$P < 0,001^*$	$P = 0,327$	$P = 0,792$	$P = 0,013^*$	$P = 0,015^*$	$P = 0,654$
$REC \leq 4\%$	$P < 0,001^*$	$P = 0,060$	$P = 0,383$	$P = 0,005^*$	$P = 0,005^*$	$P = 0,732$
$REC \leq 5\%$	$P < 0,001^*$	$P = 0,023^*$	$P = 0,615$	$P = 0,004^*$	$P = 0,001^*$	$P = 0,365$

\* $P < 0,05$

### 4.5.6 Comprimento Máximo das Linhas Verticais ( $V_{max}$ )

Na Figura 4.40 está ilustrada a distribuição dos valores do comprimento máximo das linhas verticais,  $V_{max}$  (definido na seção 3.3.8 na página 39), para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos, considerando cinco diferentes limitantes aplicados à taxa de recorrência: 1%, 2%, 3%, 4% e 5%.



**Figura 4.40** Distribuições dos valores da medida  $V_{max}$  para sinais de vozes saudáveis (SDL) e patológicas (PTL).

Na Tabela 4.43 estão apresentados os valores da média e do desvio padrão do comprimento máximo da linhas verticais para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.43** Média e desvio padrão para  $V_{max}$ .

Medida $V_{max}$	Tipo de Voz		
	Saudável	Patológica	Valor P
$REC \leq 1\%$	$6,09 \pm 2,58$	$10,98 \pm 5,14$	$P < 0,001$
$REC \leq 2\%$	$9,30 \pm 4,33$	$16,36 \pm 7,68$	$P < 0,001$
$REC \leq 3\%$	$12,57 \pm 6,64$	$20,76 \pm 10,18$	$P < 0,001$
$REC \leq 4\%$	$15,47 \pm 7,76$	$24,87 \pm 11,57$	$P < 0,001$
$REC \leq 5\%$	$18,45 \pm 10,39$	$29,45 \pm 14,22$	$P < 0,001$

Após a realização do teste não-paramétrico U de Mann-Whitney, os resultados revelaram que os grupos de vozes saudáveis e patológicas apresentam diferenças estatísticas significativas para todos os limiares de taxa de recorrência considerados ( $P < 0,001$ ).

Na Tabela 4.44 estão apresentados os valores da média e do desvio padrão do comprimento máximo das linhas verticais para os sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.44** Média e desvio padrão para  $V_{max}$  - separação das patologias.

Medida $V_{max}$	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$REC \leq 1\%$	6,09 ± 2,58	11,45 ± 5,13	11,16 ± 5,47	9,11 ± 3,97
$REC \leq 2\%$	9,30 ± 4,33	17,07 ± 7,67	16,76 ± 8,08	13,17 ± 6,07
$REC \leq 3\%$	12,57 ± 6,64	21,96 ± 11,40	20,64 ± 9,31	17,39 ± 7,68
$REC \leq 4\%$	15,47 ± 7,76	26,33 ± 12,11	24,89 ± 11,94	20,39 ± 7,66
$REC \leq 5\%$	18,45 ± 10,39	31,04 ± 16,11	29,56 ± 13,35	24,33 ± 8,35

A fim de testar as possíveis diferenças estatísticas entre os quatro grupos de vozes, foi realizado o teste de análise de variância ANOVA não paramétrico de Kruskal-Wallis. Segundo os resultados obtidos, existem diferenças estatísticas significativas entre pelo menos dois dos grupos analisados, para todos os patamares da taxa de recorrência considerados  $P < 0,001$ .

Na Tabela 4.45 estão apresentados os resultados da análise estatística realizada visando a comparação dos valores da característica  $V_{max}$  obtidos para sinais de vozes saudáveis (SDL), sinais de vozes afetados por paralisia (PRL), edema (EDM) e nódulos (NDL) nas pregas vocais.

**Tabela 4.45** Potencial estatístico da característica  $V_{max}$  na discriminação entre grupos de sinais de voz.

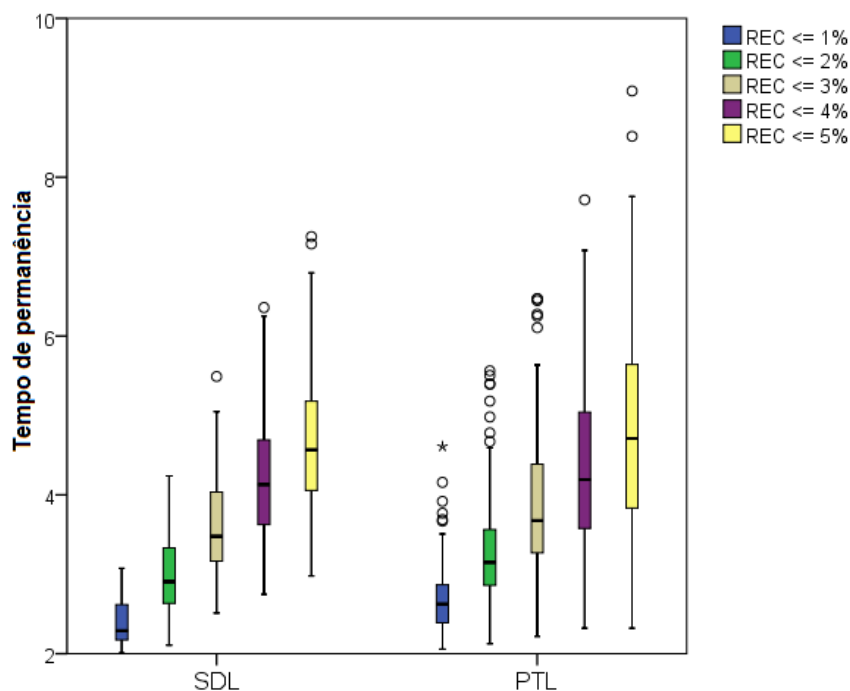
Medida $V_{max}$	Pares de Grupos de Sinais					
	SDL x PRL	SDL x EDM	SDL x NDL	PRL x EDM	PRL x NDL	EDM x NDL
$REC \leq 1\%$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,001^*$	$P = 0,487$	$P = 0,048^*$	$P = 0,179$
$REC \leq 2\%$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,002^*$	$P = 0,605$	$P = 0,030^*$	$P = 0,121$
$REC \leq 3\%$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,002^*$	$P = 0,652$	$P = 0,061$	$P = 0,173$
$REC \leq 4\%$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,004^*$	$P = 0,367$	$P = 0,055$	$P = 0,259$
$REC \leq 5\%$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,002^*$	$P = 0,795$	$P = 0,187$	$P = 0,226$

\* $P < 0,05$

A utilização do teste U de Mann-Whitney para uma comparação dois a dois entre os grupos, revelou diferenças significativas dos valores de  $V_{max}$  entre o grupo de vozes saudáveis e cada um dos três grupos de vozes patológicas, para todos os patamares da taxa de recorrência. Considerando a comparação entre os grupos de vozes afetadas por patologias, os resultados apontaram diferenças significativas entre os grupos paralisia x nódulos apenas nos casos em que  $REC \leq 1\%$  e  $REC \leq 2\%$ . A comparação entre os grupos paralisia x edema e edema x nódulos não apresentou diferenças estatísticas significativas sob qualquer condição analisada.

### 4.5.7 Tempo de Permanência

Na Figura 4.41 está ilustrada a distribuição dos valores do tempo de permanência em um estado,  $TT$  (definido na seção 3.3.7 na página 39), para os 53 sinais de vozes saudáveis (SDL) e 118 sinais de vozes patológicas (PTL) utilizados nos experimentos, considerando cinco diferentes limitantes aplicados à taxa de recorrência: 1%, 2%, 3%, 4% e 5%.



**Figura 4.41** Distribuições dos valores da medida Traptime ( $TT$ ) para sinais de vozes saudáveis (SDL) e patológicas (PTL).

Na Tabela 4.46 estão apresentados os valores da média e do desvio padrão do tempo de permanência em um estado para os sinais de vozes saudáveis e de vozes patológicas analisados.

**Tabela 4.46** Média e desvio padrão para  $TT$ .

Medida $TT$	Tipo de Voz		Valor P
	Saudável	Patológica	
$REC \leq 1\%$	$2,38 \pm 0,28$	$2,69 \pm 0,46$	$P < 0,001$
$REC \leq 2\%$	$2,98 \pm 0,47$	$3,30 \pm 0,74$	$P = 0,006$
$REC \leq 3\%$	$3,60 \pm 0,63$	$3,87 \pm 0,95$	$P = 0,078$
$REC \leq 4\%$	$4,20 \pm 0,79$	$4,33 \pm 1,11$	$P = 0,362$
$REC \leq 5\%$	$4,66 \pm 0,95$	$4,86 \pm 1,36$	$P = 0,274$

Como apresentado na Tabela 4.46 os resultados obtidos a partir da aplicação do teste não-paramétrico U de Mann-Whitney revelaram que os grupos de vozes saudáveis e patológicas apresentam diferenças estatísticas significativas, apenas quando  $REC \leq 1\%$  e  $REC \leq 2\%$ .

Na Tabela 4.47 estão apresentados os valores da média e do desvio padrão para a medida do tempo de permanência em um estado dos sinais de vozes saudáveis e para os sinais de vozes patológicas estratificados pelo tipo de patologia, ou seja, paralisia, edema e nódulos.

**Tabela 4.47** Média e desvio padrão para *TT* - separação das patologias.

Medida <i>TT</i>	Tipo de Voz			
	Saudável	Paralisia	Edema	Nódulo
$REC \leq 1\%$	2,38 ± 0,28	2,69 ± 0,47	2,74 ± 0,49	2,56 ± 0,27
$REC \leq 2\%$	2,98 ± 0,47	3,23 ± 0,79	3,42 ± 0,78	3,24 ± 0,45
$REC \leq 3\%$	3,60 ± 0,63	3,70 ± 1,01	4,00 ± 0,96	4,02 ± 0,66
$REC \leq 4\%$	4,20 ± 0,79	4,12 ± 1,17	4,57 ± 1,18	4,53 ± 0,70
$REC \leq 5\%$	4,66 ± 0,95	4,50 ± 1,38	5,17 ± 1,44	5,15 ± 0,74

Na Tabela 4.48 estão apresentados os resultados obtidos, a partir da aplicação do teste de análise de variância ANOVA não paramétrico de Kruskal-Wallis, a fim de testar as possíveis diferenças estatísticas entre os quatro grupos de vozes. Segundo os resultados obtidos, existem diferenças estatísticas significativas entre pelo menos dois dos grupos analisados, para os patamares da taxa de recorrência  $REC \leq 1\%$  e  $REC \leq 2\%$ ; nos demais casos não existem indícios de diferenças significativas.

**Tabela 4.48** Potencial estatístico da característica *TT* na discriminação entre grupos de sinais de voz.

Medida <i>TT</i>	Limiares da Taxa de Recorrência				
	$REC \leq 1\%$	$REC \leq 2\%$	$REC \leq 3\%$	$REC \leq 4\%$	$REC \leq 5\%$
Valor-P	$P < 0,001^*$	$P = 0,015^*$	$P = 0,078$	$P = 0,103$	$P = 0,103$

\* $P < 0,05$

Na Tabela 4.49 estão apresentados os resultados da aplicação do teste U de Mann-Whitney para uma comparação dois a dois entre os grupos de sinais nas situações em que  $REC \leq 1\%$  e  $REC \leq 2\%$ .

**Tabela 4.49** Potencial estatístico da característica *TT* na discriminação entre grupos de sinais de voz.

Medida <i>TT</i>	Pares de Grupos de Sinais					
	SDL x PRL	SDL x EDM	SDL x NDL	PRL x EDM	PRL x NDL	EDM x NDL
$REC \leq 1\%$	$P < 0,001^*$	$P < 0,001^*$	$P = 0,018^*$	$P = 0,693$	$P = 0,506$	$P = 0,330$
$REC \leq 2\%$	$P = 0,131$	$P = 0,003^*$	$P = 0,036^*$	$P = 0,085$	$P = 0,321$	$P = 0,533$

\* $P < 0,05$

A utilização do teste U de Mann-Whitney para uma comparação dois a dois entre os grupos, revelou diferenças significativas dos valores de *TT* entre os grupos saudável x paralisia apenas para  $REC \leq 1\%$ ; saudável x edema e saudável x nódulos para  $REC \leq 1\%$  e também quando  $REC \leq 2\%$ . Na comparação entre os grupos de vozes afetadas por patologias, os testes estatísticos não apontaram diferenças significativas entre quaisquer dois grupos.



## 4.6 Resumo dos Resultados do Processo de Caracterização - Medidas de Quantificação de Recorrência

Na Tabela 4.50 está apresentado o resumo dos resultados obtidos a partir de testes estatísticos realizados, visando identificar o potencial discriminativo de cada uma das medidas de quantificação de recorrência avaliadas neste trabalho, na caracterização de sinais de vozes saudáveis e patológicas sob diferentes taxas de recorrência.

**Tabela 4.50** Potencial discriminativo das medidas de quantificação de recorrência.

Medidas	Taxa de Recorrência ( <i>REC</i> )				
	$\leq 1\%$	$\leq 2\%$	$\leq 3\%$	$\leq 4\%$	$\leq 5\%$
<i>SDL x PTL</i>	Todas, exceto <i>LAM</i>	Todas, exceto <i>LAM</i> e <i>TREND</i>	Todas, exceto <i>TREND</i> e <i>TT</i>	Todas, exceto <i>TT</i>	Todas, exceto <i>TT</i>
<i>SDL x PRL</i>	Todas, exceto <i>LAM</i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>V<sub>max</sub></i>	Todas, exceto <i>TREND</i> e <i>TT</i>	Todas, exceto <i>TREND</i> e <i>TT</i>	Todas, exceto <i>TREND</i> e <i>TT</i>
<i>SDL x EDM</i>	Todas, exceto <i>LAM</i>	Todas, exceto <i>LAM</i> e <i>TREND</i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>V<sub>max</sub></i>	<i>DET</i> , <i>L<sub>max</sub></i> e <i>V<sub>max</sub></i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>LAM</i> e <i>V<sub>max</sub></i>
<i>SDL x NDL</i>	Todas, exceto <i>LAM</i> e <i>TREND</i>	Todas, exceto <i>LAM</i> e <i>TREND</i>	<i>DET</i> , <i>L<sub>max</sub></i> e <i>V<sub>max</sub></i>	<i>DET</i> , <i>L<sub>max</sub></i> , e <i>V<sub>max</sub></i>	<i>DET</i> , <i>L<sub>max</sub></i> e <i>V<sub>max</sub></i>
<i>PRL x EDM</i>	<i>DET</i> , <i>L<sub>max</sub></i> e <i>ENTR</i>	<i>DET</i> , <i>L<sub>max</sub></i> e <i>ENTR</i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>LAM</i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>LAM</i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>LAM</i>
<i>PRL x NDL</i>	<i>DET</i> e <i>V<sub>max</sub></i>	<i>DET</i> , <i>ENTR</i> e <i>V<sub>max</sub></i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>LAM</i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>LAM</i>	<i>DET</i> , <i>L<sub>max</sub></i> , <i>ENTR</i> e <i>LAM</i>
<i>EDM x NDL</i>	–	–	–	–	–

Os experimentos computacionais realizados sugerem que, entre as medidas de quantificação de recorrência avaliadas, a laminaridade e o tempo de permanência em um estado oferecem, em potencial, uma capacidade de discriminação entre os grupos de sinais de vozes (saudáveis e patológicas) que depende sensivelmente da máxima taxa de recorrência permitida (ou equivalentemente do maior raio). Os resultados encontrados mostram que para uma escolha do valor do raio que mantém uma taxa de recorrência próxima (menor ou igual) a 2%, a medida de laminaridade não apresenta, em potencial, a capacidade de discriminar entre os dois grupos de sinais de voz analisados.

Por outro lado, para a escolha do maior raio de forma que a taxa de recorrência esteja próxima (ou acima) de 3%, a medida do tempo de permanência em um estado não oferece diferenças estatísticas significativas entre saudável e patológica. As demais medidas analisadas mostraram uma menor dependência, em relação à variação da taxa de recorrência, no sentido de manter o potencial discriminativo entre esses dois grupos de sinais de voz.

Observa-se que nenhuma das medidas apresenta potencial para discriminar entre vozes com edema e vozes afetadas por nódulos. A semelhança nos aspectos perceptuais e histológicos

que dificulta o diagnóstico médico para diferenciar essas patologias [49, 114], é evidenciada nos resultados apresentados pelos testes estatísticos. Na discriminação entre as patologias de origem orgânica (edema e nódulos) e neurológica (paralisia), as medidas oferecem potencial para discriminar entre paralisia e nódulos, bem como entre paralisia e edema.

De uma forma geral, os resultados obtidos por meio da análise estatística baseada em testes de hipóteses, sugerem que a utilização da análise dinâmica não linear se mostra como uma técnica promissora na tarefa de discriminação de sinais de vozes saudáveis e vozes patológicas.

A análise de desempenho da classificação dos sinais de vozes entre saudáveis e patológicos é apresentada no capítulo a seguir.

## CAPÍTULO 5

---

# Classificação de Sinais de Voz Usando Medidas da Análise Não Linear e Medidas de Recorrência

---

### 5.1 Introdução

Neste capítulo são apresentados os resultados obtidos do processo de classificação de sinais de vozes saudáveis e patológicas realizado a fim de investigar o potencial discriminativo de um conjunto de medidas da análise não linear (MNL), das medidas de quantificação de recorrência (MQR), bem como da combinação de ambos os conjuntos (MNL e MQR). O objetivo é detectar a presença de desordens vocais causadas pelas patologias edema, paralisia e nódulos nas pregas vocais. Para esse fim são utilizadas as seguintes medidas da análise não linear: O máximo expoente de Lyapunov ( $\lambda_1$ ), o primeiro mínimo da função de informação mútua média (*FMMI*), a entropia de Shannon ( $H_1$ ), a entropia de Tsallis ( $H_q$ ), o expoente de Hurst ( $H$ ), a entropia aproximada (*ApEn*), a dimensão de correlação ( $D_2$ ) e a entropia de correlação ( $K_2$ ). As medidas de quantificação de recorrência utilizadas são: A taxa de recorrência (*REC*), o determinismo (*DET*), o comprimento máximo das estruturas diagonais ( $L_{max}$ ), a entropia da distribuição das linhas diagonais (*ENTR*), a tendência (*TREND*), a laminaridade (*LAM*), o comprimento máximo das estruturas verticais ( $V_{max}$ ) e o comprimento médio das linhas verticais ou tempo de permanência (*TT*).

Em um teste de laboratório para detecção de uma doença, com os possíveis resultados sendo positivo ou negativo, podem-se obter erros de precisão ou mesmo de interpretação. Quando o resultado for positivo, o indivíduo pode ter a doença (verdadeiro positivo - VP) ou pode não tê-la (falso positivo - FP). Da mesma forma, quando o resultado for negativo, o indivíduo pode não ter a doença (verdadeiro negativo - VN) como pode tê-la (falso negativo - FN). Na Tabela 5.1 é dada a matriz de confusão que fornece as possíveis situações encontradas em um teste deste tipo.

Para saber quão precisa é a classificação, três medidas são comumente 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 sensibilidade mede a capacidade do teste em identificar corretamente a doença entre aqueles que a possuem, ou seja, o quão sensível é o teste. A especificidade mede a capacidade do teste em excluir corretamente aqueles que não possuem a doença, ou seja, o quão específico é o teste.

**Tabela 5.1** Matriz de confusão em um teste de detecção da presença/ausência de doença.

Resultado	Doença	
	Presente	Ausente
Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

A Acurácia é definida como a relação do número de casos corretamente classificados e o número total de sinais analisados:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}. \quad (5.1)$$

A Sensibilidade é a fração dos que obtiveram resposta positiva no teste entre aqueles que possuem a doença:

$$Sensibilidade = \frac{VP}{VP + FN}. \quad (5.2)$$

A Especificidade é a fração dos que obtiveram resposta negativa no teste entre aqueles que não possuem a doença:

$$Especificidade = \frac{VN}{VN + FP}. \quad (5.3)$$

Um bom teste deve possuir além de um alto valor para a acurácia, um alto valor para a sensibilidade e também para a especificidade, pois ele identificará corretamente aqueles que têm a doença e aqueles que não a têm.

Os experimentos computacionais realizados neste trabalho utilizaram um total de 171 sinais de voz da vogal sustentada /ah/, sendo: 118 sinais de vozes patológicas (45 de pacientes apresentando edema, 55 pacientes com paralisia e 18 com nódulos nas pregas vocais) e 53 pacientes com vozes saudáveis. A fim de se obter um bom classificador, é recomendável usar o máximo possível de dados para o treinamento do modelo. Por outro lado, para se obter um boa estimativa de erro é recomendável usar o máximo possível dos dados na etapa de teste [115]. Desse modo, visando utilizar a maior quantidade de dados possível em ambas as fases do processo de classificação, é criada uma partição aleatória para definir os conjuntos de treinamento e teste para a validação do modelo estatístico usando a validação cruzada com  $k$  subconjuntos (Cross-Validation  $k$ -fold) estratificada.

O método de validação cruzada com  $k$  subconjuntos consiste em dividir o conjunto total de dados em  $k$  subconjuntos mutuamente exclusivos aproximadamente do mesmo tamanho e, a partir disto, um desses subconjuntos é utilizado para teste e os  $k-1$  restantes são utilizados para estimação dos parâmetros. Este processo é realizado  $k$  vezes alternando o subconjunto de teste. A cada uma das  $k$  iterações são obtidas as medidas de desempenho (Acurácia, Sensibilidade e Especificidade) sobre os erros e acertos da predição e ao final de todas as iterações, calculados a média e o desvio padrão, obtendo assim uma estimação da capacidade do modelo de representar o processo gerador dos dados. O termo "estratificada" significa que todos os grupos envolvidos no processo de classificação estarão representados, de forma proporcional, nos conjuntos de teste e de treino.

Nesta pesquisa é utilizado um valor de  $k$  igual a 10 no processo de validação cruzada. Segundo Witten *et al.* [115], extensivos testes sobre vários conjuntos de dados usando diferentes técnicas de aprendizado automático, têm mostrado que  $k = 10$  é o número de subconjuntos ("*folds*") que tem apresentado as melhores estimativas de erro, além de existir algumas evidências teóricas, embora não conclusivas, que sugerem esse valor. Por isso, a validação cruzada com 10 subconjuntos tem se tornado padrão em termos práticos [116].

A classificação dos sinais é realizada por meio do método da análise de discriminante considerando dois tipos de funções: linear (LDA - *Linear Discriminant Analysis*) e quadrática (QDA - *Quadratic Discriminant Analysis*). Os métodos de discriminação são muito afetados pela natureza das matrizes de covariâncias dos grupos. No método de discriminação linear, ou método de Fisher é assumido a igualdade das matrizes de covariância para todos os grupos. Quando a hipótese de homogeneidade da covariância dentro dos grupos é violada, a análise de discriminação quadrática é recomendada. Entretanto, ambos os métodos supõem a hipótese da distribuição normal multivariada para os dados dentro dos grupos, caracterizada pelo vetor média e a matriz covariância para cada grupo. Apenas os resultados que ofereceram as melhores taxas de classificação, em termos de acurácia média, estão apresentados nas seções seguintes.

Inicialmente, é analisado o desempenho de classificação considerando apenas dois grupos de sinais: um de vozes patológicas - formado conjuntamente por vozes com edema, paralisia ou nódulos - e o outro de vozes saudáveis. Posteriormente, é analisado o desempenho de classificação entre cada dois dos quatro diferentes grupos (saudável  $\times$  paralisia, saudável  $\times$  edema, saudável  $\times$  nódulos, paralisia  $\times$  edema, paralisia  $\times$  nódulos e edema  $\times$  nódulos). A avaliação de desempenho na classificação dos sinais entre os grupos é realizada considerando-se o poder individual de cada uma das medidas consideradas (MNL e MQR), como também de todas as combinações possíveis entre essas medidas. Após identificar, entre as características analisadas, aquelas que fornecem as melhores taxas de classificação entre os diferentes grupos de sinais de vozes considerados, são avaliados os efeitos do uso de vetores híbridos, formados pela combinação das características MNL e MQR e por coeficientes extraídos da análise de predição linear (LPC - *Linear Predictive Coding*).

## 5.2 Classificação de Vozes Saudáveis *versus* Patológicas

Nesta seção são apresentados os resultados obtidos da classificação entre os sinais de vozes saudáveis e patológicas. Os sinais de vozes patológicas compreendem, neste caso, os sinais de todas as patologias (edema, paralisia e nódulos) agrupadas numa mesma classe.

### 5.2.1 Características Avaliadas Individualmente

Os resultados do processo de classificação utilizando cada característica de forma individual estão apresentados nas Tabelas 5.2 e 5.3.

#### I - Medidas da Análise Não Linear (MNL)

Na Tabela 5.2 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de análise não linear utilizadas de forma individual na classificação entre sinais de vozes saudáveis e vozes patológicas (Saudável x Patológica), por meio da função discriminante quadrática (QDA).

**Tabela 5.2** Taxas de classificação (%) Saudável x Patológica - MNL avaliadas individualmente (QDA)

Medidas	$\lambda_1$	FMMI	$H_1$	$H_q$	$H$	ApEn	$D_2$	$K_2$
Acurácia	54,51±11,50	62,55±8,03	60,85±11,62	62,61±12,88	87,09*±8,26	56,60±9,39	79,36±15,06	52,88±7,12
Sensibilidade	44,77±15,42	59,17±10,93	51,82±18,53	53,48±18,42	84,85±8,65	37,68±16,76	71,96±20,38	30,71±16,10
Especificidade	75,33±19,70	70,00±13,05	81,33±12,39	83,33±9,94	92,67±12,35	83,00±13,74	89,00±12,38	83,00±9,99

\*acurácia média máxima

Como pode ser observado, a medida do expoente de Hurst ( $H$ ) apresentou a maior taxa de acurácia média do classificador. Além disso, essa mesma característica ofereceu o melhor desempenho nos critérios de sensibilidade média e especificidade média.

#### II - Medidas de Quantificação de Recorrência (MQR)

Na Tabela 5.3 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de quantificação de recorrência utilizadas de forma individual na classificação entre vozes saudáveis e vozes patológicas, por meio da função discriminante quadrática (QDA).

**Tabela 5.3** Taxas de classificação (%) Saudável x Patológica - MQR avaliadas individualmente (QDA)

Medidas	DET	$L_{max}$	ENTR	TREND	LAM	$V_{max}$	TT
Acurácia	76,63±11,35	91,27*±6,21	80,75±5,96	71,37±12,13	51,50±14,47	66,70±8,14	53,20±10,41
Sensibilidade	69,39±16,57	95,83±5,89	75,53±8,91	83,71±15,15	46,74±20,80	55,83±10,19	42,35±10,95
Especificidade	92,00±10,33	81,00±12,38	92,67±9,53	43,67±13,56	62,33±18,06	91,33±12,09	77,33±22,43

\*acurácia média máxima

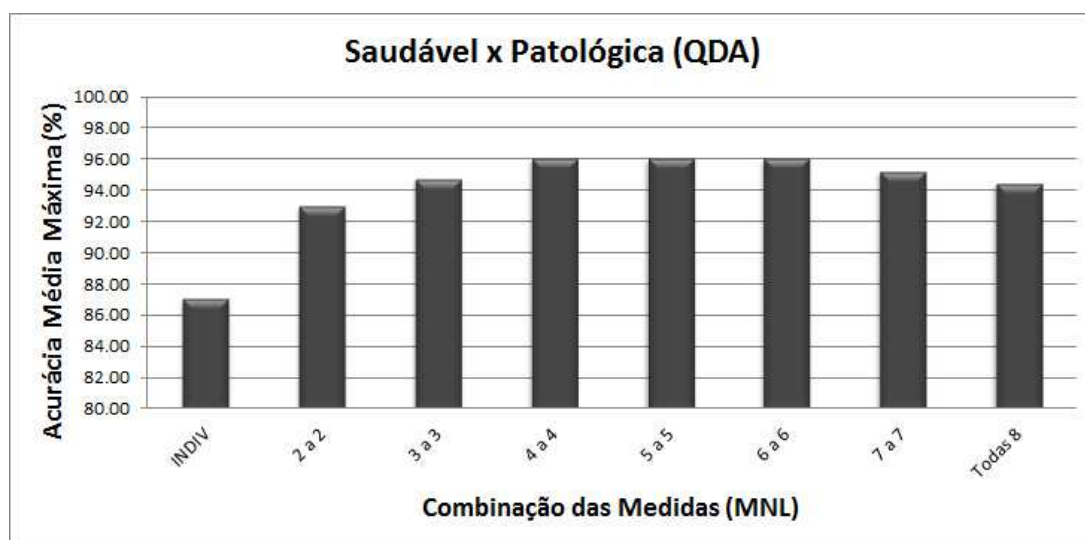
Os resultados apresentados na Tabela 5.3 mostram que o comprimento máximo das linhas diagonais ( $L_{max}$ ) oferece a maior taxa de acurácia média, como também o melhor desempenho na sensibilidade do classificador. Por outro lado, a medida de entropia ( $ENTR$ ) apresenta melhor resultado no critério de especificidade.

## 5.2.2 Características Avaliadas de Forma Combinada

A combinação das características visa aumentar a quantidade de informação fornecida ao classificador, no intuito de melhorar as taxas de classificação. Para isso, foram avaliadas as combinações entre as medidas da análise não linear (MNL), entre as medidas de recorrência (MQR) e também entre todas as características desses dois conjuntos (MNL e MQR).

### I - Medidas da Análise Não Linear (MNL)

Na Figura 5.1 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas da análise não linear investigadas, bem como de todas as combinações possíveis entre elas, na classificação entre vozes saudáveis e vozes patológicas.

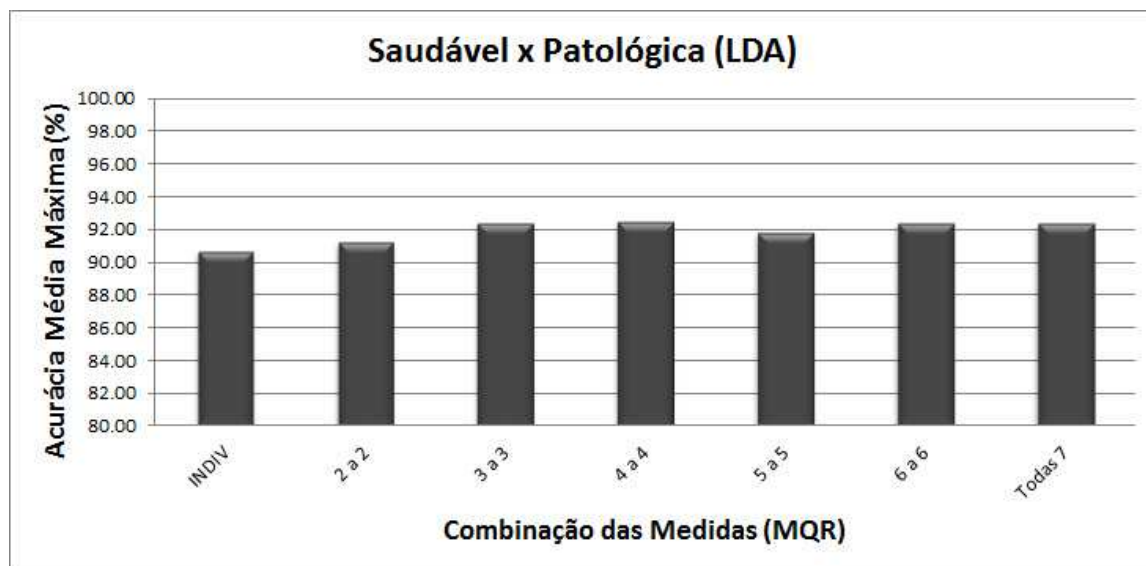


**Figura 5.1** Acurácia Média Máxima - Saudável x Patológica (MNL)

Como pode ser observado no gráfico apresentado na Figura 5.1, o maior valor atingido pela acurácia média máxima pode ser obtido pela combinação de quatro medidas entre as oito empregadas (usando QDA). Uma acurácia média máxima de valor igual a  $96,03 \pm 5,75\%$  foi encontrada na combinação das seguintes medidas: entropia de Shannon ( $H_1$ ), entropia de Tsallis ( $H_q$ ), expoente de Hurst ( $H$ ) e dimensão de correlação ( $D_2$ ). Uma sensibilidade média de  $94,46 \pm 7,17\%$  e uma especificidade média de  $98,00 \pm 6,32\%$  foram encontradas para este mesmo conjunto de características.

## II - Medidas de Quantificação de Recorrência (MQR)

Na Figura 5.2 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas de quantificação de recorrência investigadas, bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes saudáveis e patológicas.



**Figura 5.2** Acurácia Média Máxima - Saudável x Patológica (MQR)

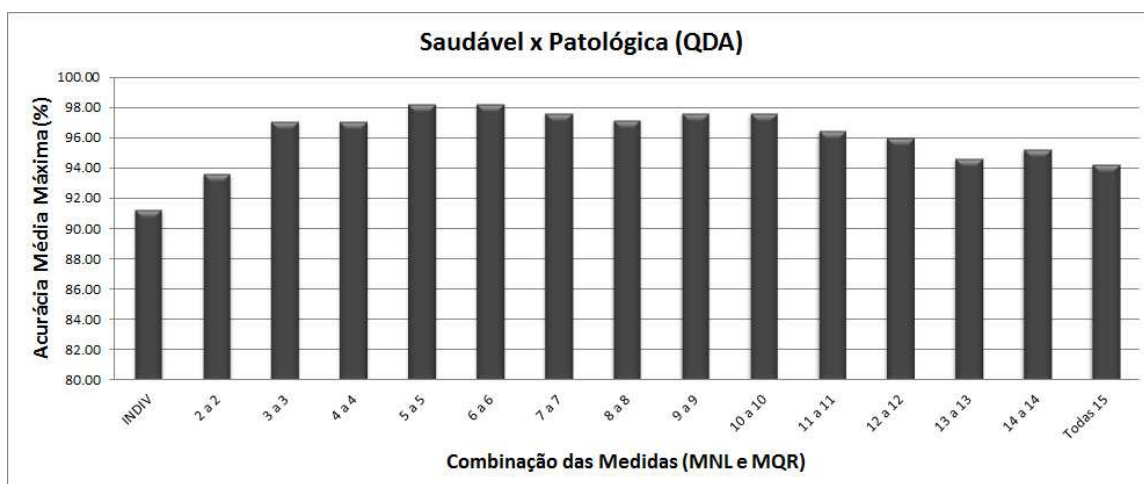
No caso da combinação das medidas de recorrência, o maior valor encontrado para a acurácia média máxima foi de  $92,48 \pm 6,49\%$ , para o conjunto das quatro medidas:  $L_{max}$ ,  $ENTR$ ,  $LAM$  e  $V_{max}$  (usando LDA). Para este mesmo conjunto de características foi encontrado um valor médio de  $93,26 \pm 6,60\%$  para a sensibilidade e um valor médio de  $90,67 \pm 12,65\%$  para a especificidade. Entretanto, como pode ser observado no gráfico apresentado na Figura 5.2, desempenho semelhante em termos de acurácia média pode ser obtido por diferentes combinações de apenas três medidas, entre as oito consideradas, ou seja,  $DET$ ,  $L_{max}$  e  $V_{max}$  ( $92,35 \pm 6,82\%$ ) ou  $L_{max}$ ,  $ENTR$  e  $TT$  ( $92,35 \pm 7,87\%$ ).

## III - Medidas da Análise Não Linear e Medidas de Quantificação de Recorrência

Na Figura 5.3 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual de todas as medidas dos conjuntos MNL e MQR, bem como de todas as combinações possíveis entre elas, na classificação entre vozes saudáveis e vozes patológicas.

O maior valor encontrado para a acurácia média máxima foi de  $98,27 \pm 2,79\%$ , para o conjunto das seis medidas: entropia de Shannon ( $H_1$ ), entropia de Tsallis ( $H_q$ ), expoente de Hurst ( $H$ ),  $DET$ ,  $LAM$  e  $TT$  (usando QDA). Para este mesmo conjunto de características foi encontrado um valor médio de  $97,50 \pm 4,03\%$  para a sensibilidade e a taxa máxima de 100% para a especificidade média.





**Figura 5.3** Acurácia Média Máxima - Saudável x Patológica (MNL e MQR)

Como pode ser observado no gráfico apresentado na Figura 5.3, desempenho semelhante ao apresentado pela combinação de seis características, em termos de acurácia média, pode ser obtido por meio da combinação de apenas cinco medidas: *FMMI*, entropia de Shannon ( $H_1$ ), entropia de Tsallis ( $H_1$ ), expoente de Hurst ( $H$ ) e *ENTR*. Neste caso, o valor da acurácia média máxima foi de  $98,24 \pm 2,84\%$ . A sensibilidade e a especificidade apresentaram valores médios iguais ao caso da combinação com seis medidas.

## 5.3 Classificação de Vozes Saudáveis *versus* Vozes com Edema nas Pregas Vocais

Nesta seção são apresentados os resultados da classificação entre sinais de vozes saudáveis e vozes afetadas por edema nas pregas vocais.

### 5.3.1 Características Avaliadas Individualmente

Os resultados do processo de classificação utilizando cada característica de forma individual estão apresentados nas Tabelas 5.4 e 5.5.

#### I - Medidas da Análise Não Linear (MNL)

Na Tabela 5.4 estão apresentados os resultados das medidas de desempenho referentes à utilização das medidas de análise não linear na classificação entre sinais de vozes saudáveis e vozes com edema por meio da função discriminante linear (LDA).

Os melhores resultados encontrados mostram desempenhos praticamente idênticos, em termos de acurácia média, das características expoente de Hurst ( $H$ ) e dimensão de correlação  $D_2$ . Por outro lado, enquanto o expoente de Hurst apresenta os melhores resultados para a

medida de sensibilidade, a dimensão de correlação é a característica que fornece resultados superiores no critério especificidade.

**Tabela 5.4** Taxas de classificação (%) Saudável x Edema - MNL avaliadas individualmente (LDA)

Medidas	$\lambda_1$	FMMI	$H_1$	$H_q$	H	ApEn	$D_2$	$K_2$
Acurácia	62,11±17,93	57,89±11,48	69,56±14,01	69,56±12,32	85,78±8,33	67,92±11,14	85,97*±11,67	56,81±15,76
Sensibilidade	53,50±34,08	65,50±21,27	58,50±24,27	63,00±26,27	84,50±10,92	61,67±33,38	73,33±26,29	51,67±22,84
Especificidade	68,00±20,14	53,00±27,42	79,67±16,21	75,67±15,56	86,67±15,40	71,00±14,99	92,67±12,35	59,67±21,46

\*acurácia média máxima

## II - Medidas de Quantificação de Recorrência (MQR)

Na Tabela 5.5 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de quantificação de recorrência na classificação entre vozes saudáveis e vozes com edema por meio da função discriminante linear (LDA).

**Tabela 5.5** Taxas de classificação (%) Saudável x Edema - MQR avaliadas individualmente (LDA)

Medidas	DET	$L_{max}$	ENTR	TREND	LAM	$V_{max}$	TT
Acurácia	78,78±12,87	82,78*±10,69	80,89±15,17	65,22±14,50	57,22±13,25	73,44±13,93	65,11±20,00
Sensibilidade	62,00±25,84	92,00±19,32	71,00±31,16	73,00±23,36	67,50±21,63	57,50±25,41	55,00±27,39
Especificidade	92,33±13,70	76,33±18,62	89,00±20,48	58,00±20,68	48,33±20,08	86,67±13,24	73,33±21,49

\*acurácia média máxima

Os resultados apresentados na Tabelas 5.5 mostram que a maior taxa de acurácia média foi obtida a partir da característica  $L_{max}$ . A utilização desta mesma característica também ofereceu o melhor desempenho na sensibilidade do classificador. Por outro lado, a característica DET apresentou melhor desempenho no critério especificidade.

### 5.3.2 Características Avaliadas de Forma Combinada

Nesta seção são apresentados os resultados dos experimentos em que foram avaliadas as combinações entre as medidas da análise não linear, entre as medidas de recorrência e também entre todas as características desses dois conjuntos, com o objetivo de classificar os sinais de vozes saudáveis e com edema nas pregas vocais.

#### I - Medidas da Análise Não Linear (MNL)

Na Figura 5.4 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas da análise não linear (MNL), bem como de todas as combinações possíveis entre elas, na classificação de vozes saudáveis e vozes afetadas por edema.

O maior valor da acurácia média máxima, igual a 98,75±3,95%, foi obtido usando a análise discriminante linear a partir da combinação de cinco características: entropia de Shannon

( $H_1$ ), entropia de Tsallis ( $H_q$ ), expoente de Hurst ( $H$ ), dimensão de correlação ( $D_2$ ) e entropia de correlação ( $K_2$ ). Para este mesmo conjunto de características foi verificado um valor de sensibilidade média igual a  $96,67 \pm 10,54\%$  e uma taxa de especificidade média de 100%.

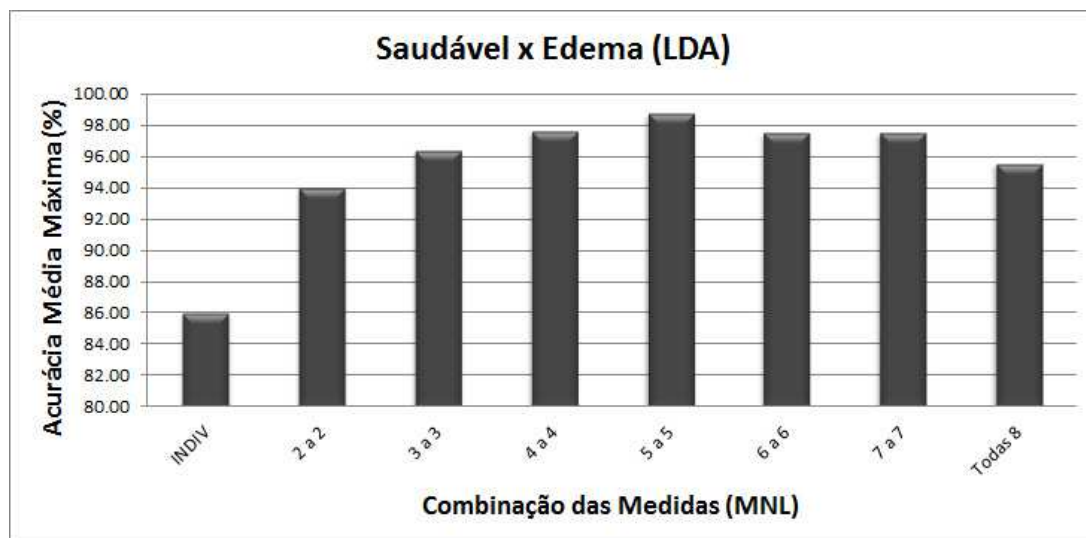


Figura 5.4 Acurácia Média Máxima - Saudável x Edema (MNL)

## II - Medidas de Quantificação de Recorrência (MQR)

Na Figura 5.5 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas de quantificação de recorrência investigadas, bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes saudáveis e vozes afetadas por edema.

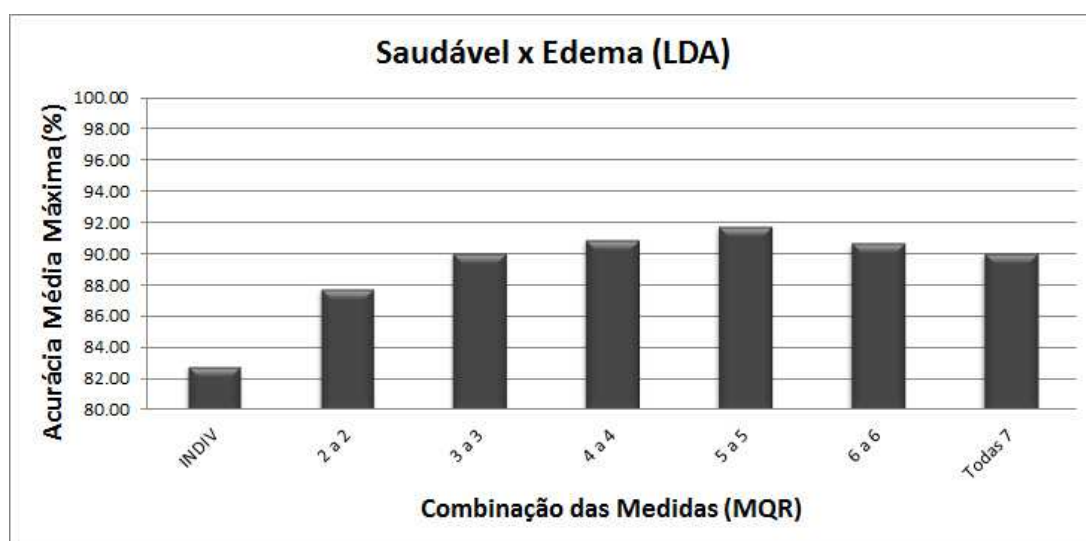


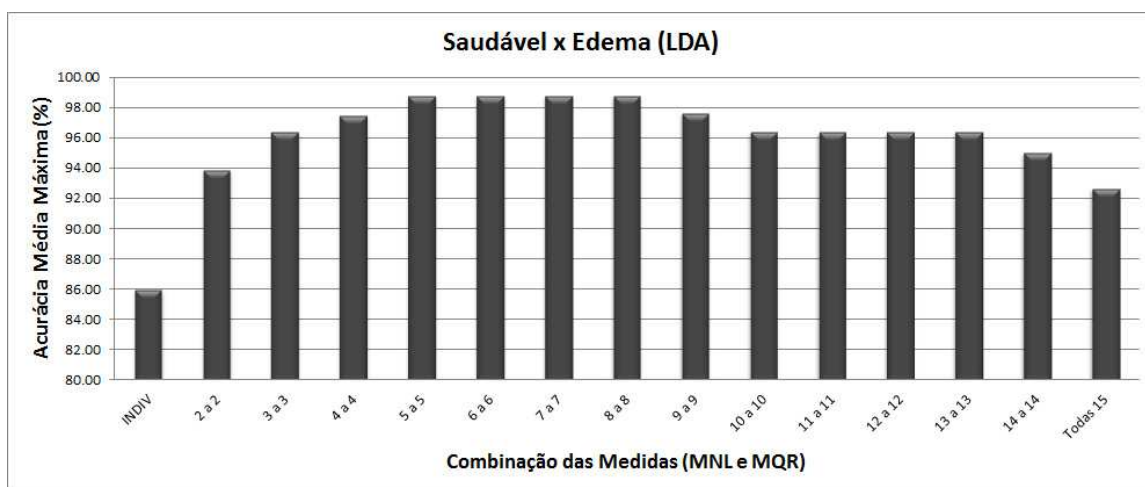
Figura 5.5 Acurácia Média Máxima - Saudável x Edema (MQR)

O maior valor da acurácia média máxima, igual a  $91,78 \pm 7,96\%$ , foi obtido usando a análise discriminante linear com a combinação de cinco características:  $DET$ ,  $L_{max}$ ,  $ENTR$ ,

$V_{max}$  e  $TT$ . Para este mesmo conjunto de características foi verificado um valor de sensibilidade média igual a  $89,00 \pm 11,74\%$  e uma especificidade média de  $94,00 \pm 9,66\%$ .

### III - Medidas da Análise Não Linear e Medidas de Quantificação de Recorrência

Na Figura 5.6 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual de todas as medidas dos conjuntos das MNL e MQR, bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes saudáveis e vozes com edema.



**Figura 5.6** Acurácia Média Máxima - Saudável x Edema (MNL e MQR)

No caso da combinação de todas as características, a fim de classificar as vozes em saudáveis ou afetadas por edema, o maior valor encontrado para a acurácia média máxima foi de  $98,75 \pm 3,95\%$ , para o conjunto das cinco medidas: entropia de Shannon ( $H_1$ ), entropia de Tsallis ( $H_q$ ), expoente de Hurst ( $H$ ), dimensão de correlação ( $D_2$ ) e entropia de correlação ( $K_2$ ), usando LDA. Para este mesmo conjunto de características foi encontrado um valor médio de  $96,67 \pm 10,54\%$  para a sensibilidade e para a especificidade média a taxa máxima de 100%.

## 5.4 Classificação de Vozes Saudáveis *versus* Vozes com Paralisia nas Pregas Vocais

Nesta seção são apresentados os resultados da classificação entre sinais de vozes saudáveis e vozes afetadas por paralisia nas pregas vocais.

### 5.4.1 Características Avaliadas Individualmente

Os resultados do processo de classificação utilizando cada característica de forma individual estão apresentados nas Tabelas 5.6 e 5.7.

## I - Medidas da Análise Não Linear (MNL)

Na Tabela 5.6 estão apresentados os resultados das medidas de desempenho referentes à utilização das medidas de análise não linear na classificação de vozes saudáveis e vozes com paralisia por meio da função discriminante quadrática (QDA).

Os melhores resultados encontrados mostram desempenhos praticamente idênticos, em termos de acurácia média, das características expoente de Hurst ( $H$ ) e dimensão de correlação ( $D_2$ ). Em relação as medidas de sensibilidade e de especificidade, o expoente de Hurst é a característica que forneceu resultados superiores.

**Tabela 5.6** Taxas de classificação (%) Saudável x Paralisia - MNL avaliadas individualmente (QDA)

Medidas	$\lambda_1$	$FMMI$	$H_1$	$H_q$	$H$	$ApEn$	$D_2$	$K_2$
Acurácia	58,55±13,97	68,64±14,12	65,82±8,48	68,55±10,71	88,09*±10,52	69,29±18,88	87,50±11,79	61,96±16,75
Sensibilidade	42,67±17,34	67,00±11,49	53,00±11,60	58,33±14,76	82,00±15,81	46,67±31,23	76,67±26,29	45,00±30,48
Especificidade	76,33±18,62	69,67±22,74	79,00±14,06	79,00±16,43	94,00±9,66	81,00±18,19	92,67±12,35	69,67±22,74

\*acurácia média máxima

## II - Medidas de Quantificação de Recorrência (MQR)

Na Tabela 5.7 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de quantificação de recorrência na classificação entre vozes saudáveis e vozes afetadas por paralisia por meio da função discriminante quadrática (QDA).

**Tabela 5.7** Taxas de classificação (%) Saudável x Paralisia - MQR avaliadas individualmente (QDA)

Medidas	$DET$	$L_{max}$	$ENTR$	$TREND$	$LAM$	$V_{max}$	$TT$
Acurácia	85,18±4,77	92,64*±7,20	88,09±6,10	65,55±13,54	60,45±15,00	78,91±10,32	61,36±19,19
Sensibilidade	75,67±15,56	94,67±8,64	82,33±13,70	79,33±11,74	49,67±23,44	67,00±14,86	44,00±25,08
Especificidade	94,00±13,50	90,33±13,92	94,33±9,17	51,33±22,29	71,67±16,57	91,00±12,38	79,67±22,58

\*acurácia média máxima

Os resultados encontrados mostram que a maior taxa de acurácia média foi obtida a partir da característica  $L_{max}$ . A utilização desta mesma característica oferece, também, o melhor desempenho para o critério de sensibilidade. Por outro lado, as características  $DET$  e  $ENTR$  apresentaram as maiores taxas no critério especificidade.

### 5.4.2 Características Avaliadas de Forma Combinada

Nesta seção são apresentados os resultados dos experimentos em que foram avaliadas as combinações entre as medidas da análise não linear (MNL), entre as medidas de recorrência (MQR) e também entre ambas (MNL e MQR), com o objetivo de classificar os sinais de vozes saudáveis e com paralisia nas pregas vocais.

## I - Medidas da Análise Não Linear (MNL)

Na Figura 5.7 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas da análise não linear investigadas, bem como de todas as combinações possíveis entre elas, na classificação de vozes saudáveis e com paralisia.

Como pode ser observado, o melhor desempenho do classificador, em termos de acurácia média máxima, pode ser obtido com a combinação de quatro características, usando (QDA). Dois diferentes conjuntos apresentaram o valor de acurácia média máxima igual a  $97,50 \pm 5,27\%$ . Esses conjuntos são formados pelas características: *FMMI*, expoente de Hurst (*H*), dimensão de correlação (*D*<sub>2</sub>) e, além dessas, em um deles aparece a entropia de Shannon (*H*<sub>1</sub>) e no outro a entropia de Tsallis (*H*<sub>q</sub>). Para os dois conjuntos foi verificado um valor de sensibilidade média igual a  $93,33 \pm 14,05\%$  e uma taxa de especificidade média de 100%.

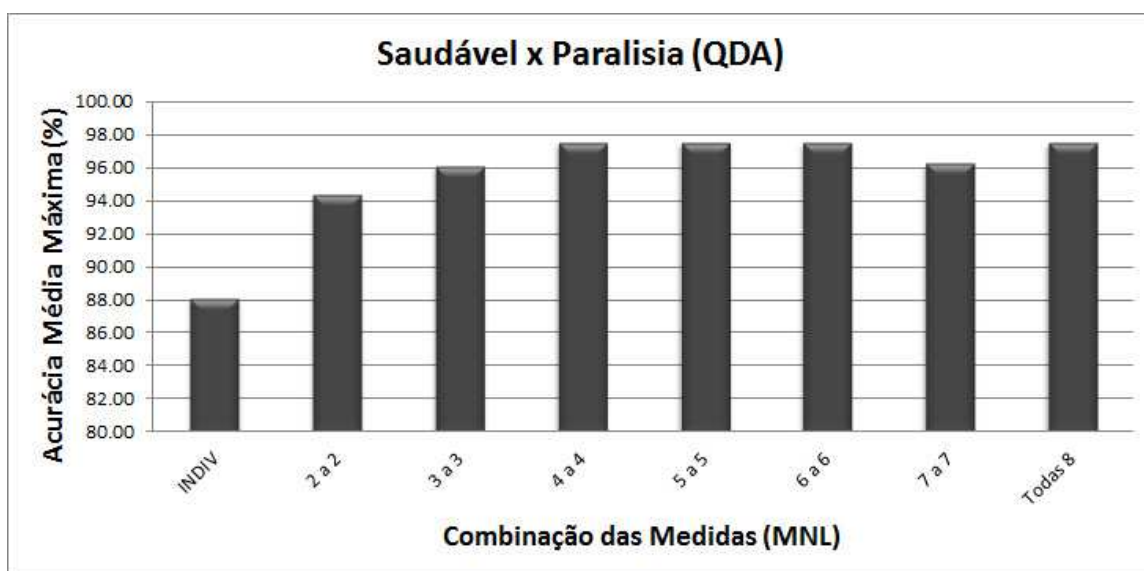


Figura 5.7 Acurácia Média Máxima - Saudável x Paralisia (MNL)

## II - Medidas de Quantificação de Recorrência (MQR)

Na Figura 5.8 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas de quantificação de recorrência investigadas, bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes saudáveis e com paralisia.

Como pode ser observado a partir do gráfico apresentado na Figura 5.8 o maior valor para a acurácia média máxima ocorre a partir da combinação de cinco medidas de recorrência. Um valor máximo de  $95,45 \pm 4,79\%$  foi obtido do conjunto das medidas: *DET*, *L*<sub>max</sub>, *ENTR*, *V*<sub>max</sub> e *TT*. Para este mesmo conjunto de características a sensibilidade média atingiu o valor de  $92,67 \pm 9,53\%$  e a especificidade média o valor de  $98,33 \pm 5,27\%$ .

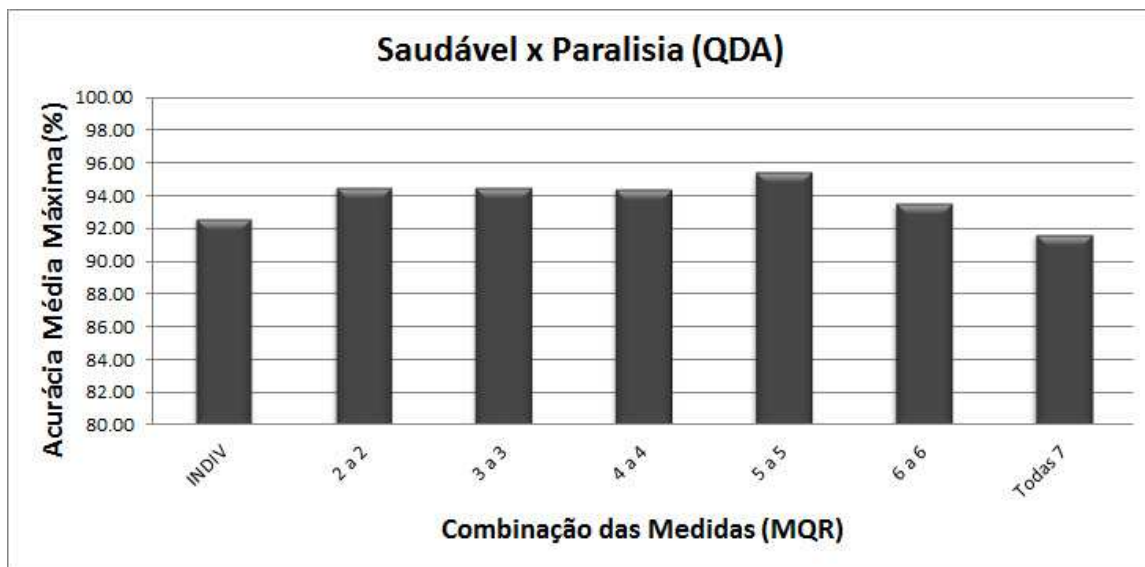


Figura 5.8 Acurácia Média Máxima - Saudável x Paralisia (MQR)

### III - Medidas da Análise Não Linear e Medidas de Quantificação de Recorrência

Na Figura 5.9 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual de todas as características dos conjuntos MNL e MQR, bem como de todas as combinações possíveis entre elas, na classificação entre vozes saudáveis e vozes afetadas por paralisia.

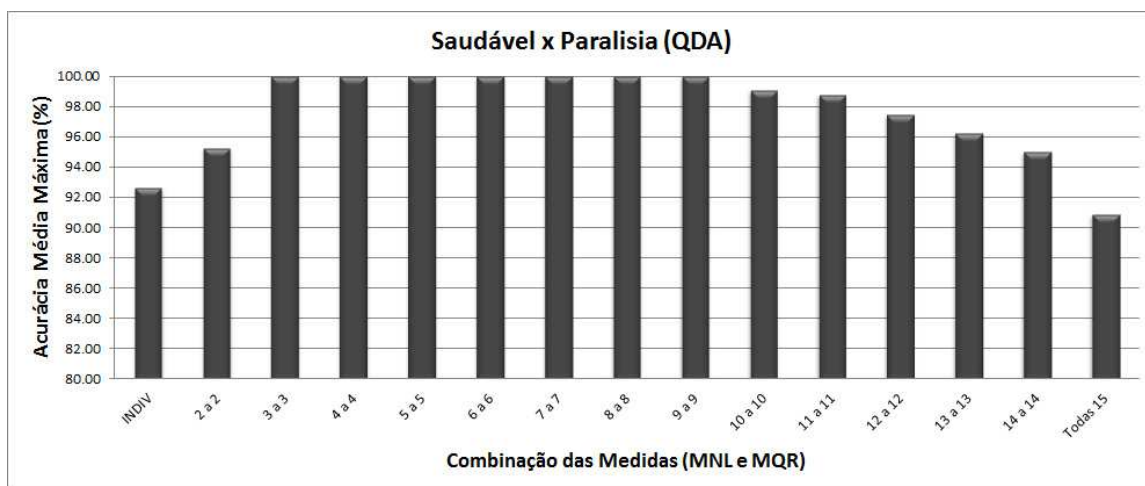


Figura 5.9 Acurácia Média Máxima - Saudável x Paralisia (MNL e MQR)

Como pode ser observado a partir do gráfico mostrado na Figura 5.9, a taxa de acurácia média máxima atingiu o maior valor possível a partir da combinação de apenas três características. Dois diferentes conjuntos com três características apresentaram a taxa de sucesso máxima de 100% para as medidas de desempenho utilizadas. Ambos são formados pelas características: *FMMI* e expoente de Hurst; além dessas, em um deles aparece o *DET* e no outro a *ENTR*.

## 5.5 Classificação de Vozes Saudáveis versus Vozes com Nódulos nas Pregas Vocais

Nesta seção são apresentados os resultados da classificação entre sinais de vozes saudáveis e vozes afetadas por nódulos nas pregas vocais.

### 5.5.1 Características Avaliadas Individualmente

Os resultados do processo de classificação utilizando cada característica de forma individual estão apresentados nas Tabelas 5.8 e 5.9.

#### I - Medidas da Análise Não Linear (MNL)

Na Tabela 5.8 estão apresentados os resultados das medidas de desempenho referentes à utilização das medidas de análise não linear na classificação entre vozes saudáveis e vozes afetadas por nódulos vocais, por meio da função discriminante linear (LDA).

**Tabela 5.8** Taxas de classificação (%) Saudável x Nódulo - MNL avaliadas individualmente (LDA)

Medidas	$\lambda_1$	$FMMI$	$H_1$	$H_q$	$H$	$ApEn$	$D_2$	$K_2$
Acurácia	60,71±13,98	51,96±14,15	74,46±17,70	73,04±18,53	94,29*±7,38	65,95±19,57	84,29±17,10	66,19±16,41
Sensibilidade	50,00±47,14	60,00±31,62	60,00±31,62	60,00±31,62	95,00±15,81	75,00±26,35	75,00±42,49	75,00±35,36
Especificidade	64,33±21,08	49,33±14,04	79,33±20,95	77,33±21,82	94,33±9,17	66,33±25,65	86,33±13,37	64,00±16,76

\*acurácia média máxima

Apresentando desempenhos de classificação idênticos para LDA e QDA, o expoente de Hurst é a característica que ofereceu os melhores resultados para as três medidas de desempenho.

#### II - Medidas de Quantificação de Recorrência (MQR)

Na Tabela 5.9 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de quantificação de recorrência na discriminação entre vozes saudáveis e vozes afetadas por nódulos, por meio da função discriminante quadrática (QDA).

**Tabela 5.9** Taxas de classificação (%) Saudável x Nódulo - MQR avaliadas individualmente (QDA)

Medidas	$DET$	$L_{max}$	$ENTR$	$TREND$	$LAM$	$V_{max}$	$TT$
Acurácia	84,64±10,11	93,04*±7,36	87,14±14,21	56,43±17,96	53,75±16,72	75,89±11,94	63,21±21,69
Sensibilidade	65,00±24,15	95,00±15,81	80,00±25,82	95,00±15,81	80,00±25,82	50,00±40,82	65,00±33,75
Especificidade	92,67±9,53	92,67±9,53	90,33±13,92	43,33±19,88	44,67±17,79	84,33±18,33	63,00±26,03

\*acurácia média máxima

O maior valor para a acurácia média foi obtido com a utilização da característica  $L_{max}$ . A maior taxa da medida de sensibilidade foi encontrada com o uso de duas diferentes caracte-



rísticas  $L_{max}$  e  $TREND$ . Por outro lado, o maior valor para a especificidade média foi obtido também com o uso de duas características  $DET$  e  $L_{max}$ .

## 5.5.2 Características Avaliadas de Forma Combinada

Nesta seção são apresentados os resultados dos experimentos em que foram avaliadas as combinações entre as medidas da análise não linear (MNL), entre as medidas de recorrência (MQR) e também entre todas as características dos dois conjuntos (MNL e MQR), com o objetivo de discriminar entre os sinais de vozes saudáveis e os sinais com nódulos nas pregas vocais.

### I - Medidas da Análise Não Linear (MNL)

Na Figura 5.10 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas da análise não linear (MNL), bem como de todas as combinações possíveis entre elas, na discriminação entre vozes saudáveis e vozes afetadas por nódulos.

Como pode ser observado, o maior valor da acurácia média máxima pode ser obtido com a combinação de apenas três características, usando LDA. Tais características são: as entropias de Shannon ( $H_1$ ) e de Tsallis ( $H_q$ ) e o expoente de Hurst ( $H$ ). Neste caso, a acurácia média máxima atingiu o valor de  $98,57 \pm 4,52\%$ , a sensibilidade média o valor de  $95,00 \pm 15,81\%$  e a especificidade a taxa máxima de 100%.

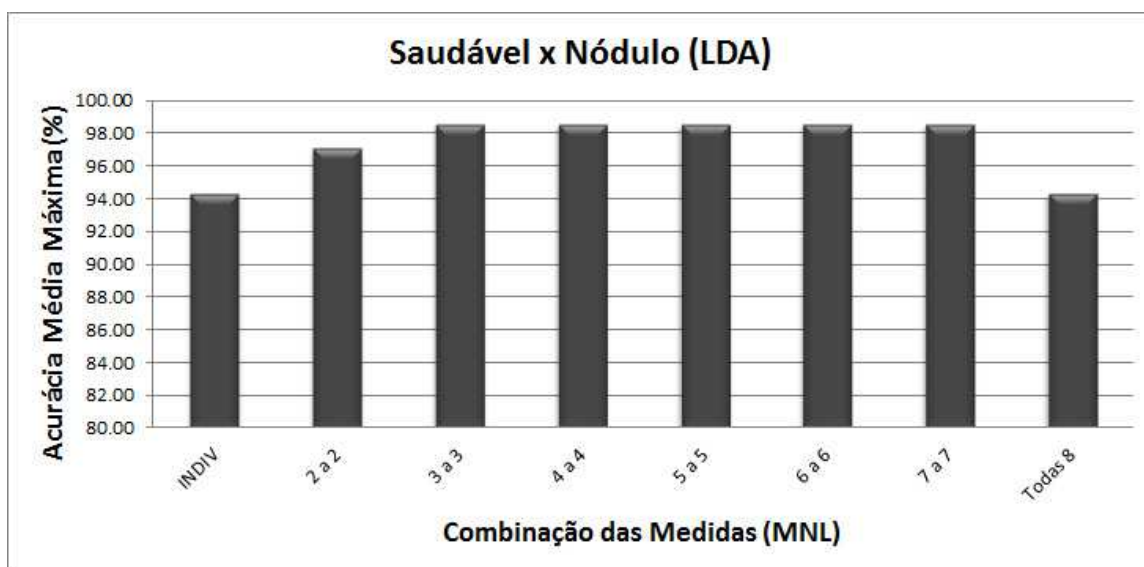
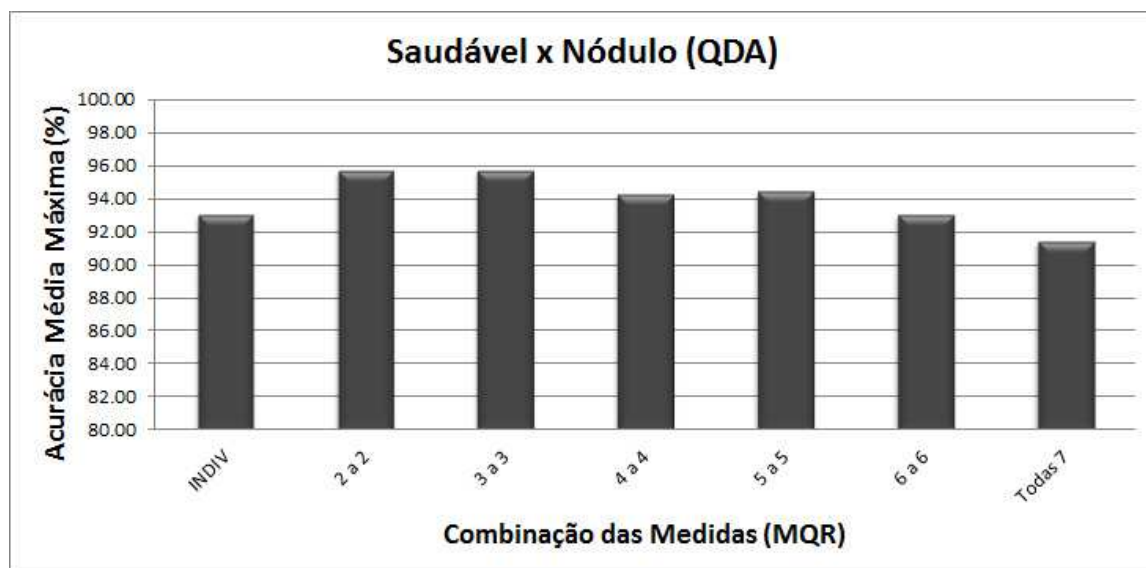


Figura 5.10 Acurácia Média Máxima - Saudável x Nódulo (MNL)

## II - Medidas de Quantificação de Recorrência (MQR)

Na Figura 5.11 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas de quantificação de recorrência (MQR), bem como de todas as combinações possíveis entre elas, na classificação entre vozes saudáveis e vozes com nódulos.



**Figura 5.11** Acurácia Média Máxima - Saudável x Nódulo (MQR)

Como pode ser observado a partir do gráfico apresentado na Figura 5.11 o maior valor para a acurácia média máxima ocorre a partir da combinação de apenas duas medidas de recorrência. Um valor máximo de  $95,71 \pm 6,90\%$  foi obtido na combinação de quadro diferentes conjuntos de características, envolvendo em todos eles a medida  $L_{max}$ . As outras características foram  $DET$ ,  $ENTR$ ,  $LAM$  e  $TT$ . Em todos esses casos a sensibilidade média atingiu o valor de  $95,00 \pm 15,81\%$  e a especificidade média o valor de  $96,00 \pm 8,43\%$ .

## III - Medidas da Análise Não Linear e Medidas de Quantificação de Recorrência

Na Figura 5.12 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das características dos grupos MNL e MQR, bem como de todas as combinações possíveis entre elas, na classificação entre vozes saudáveis e vozes afetadas por nódulos nas pregas vocais.

Dois diferentes conjuntos formados por cinco características apresentaram o maior valor encontrado para a acurácia média máxima, usando QDA. Em ambos os conjuntos estão presentes as características: expoente de Hurst ( $H$ ),  $DET$  e  $LAM$ ; Além dessas, em um deles aparecem  $FMMI$  e  $ENTR$  e no outro as entropias de Shannon ( $H_1$ ) e de Tsallis ( $H_q$ ). Para tais conjuntos, a acurácia máxima foi de  $98,75 \pm 3,95\%$ , a sensibilidade média de  $95,00 \pm 15,81\%$  e especificidade média de  $100\%$ .

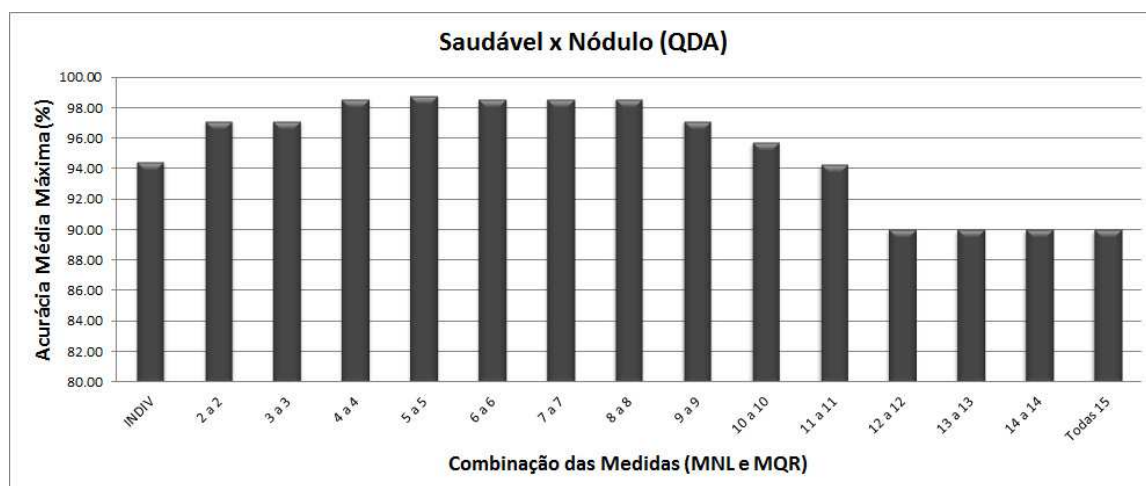


Figura 5.12 Acurácia Média Máxima - Saudável x Nódulo (MNL e MQR)

## 5.6 Classificação de Vozes Patológicas - Paralisia versus Edema

Nesta seção são apresentados os resultados da classificação entre sinais de vozes afetadas por paralisia e vozes com edema nas pregas vocais.

### 5.6.1 Características Avaliadas Individualmente

Os resultados do processo de classificação utilizando cada característica de forma individual estão apresentados nas Tabelas 5.10 e 5.11.

#### I - Medidas da Análise Não Linear (MNL)

Na Tabela 5.10 estão apresentados os resultados das medidas de desempenho referentes à utilização das medidas de análise não linear (MNL) na discriminação entre sinais de vozes afetadas por paralisia e vozes afetadas por edema, por meio da função discriminante linear (LDA).

Tabela 5.10 Taxas de classificação (%) Paralisia x Edema - MNL avaliadas individualmente (LDA)

Medidas	$\lambda_1$	$FMMI$	$H_1$	$H_q$	$H$	$ApEn$	$D_2$	$K_2$
Acurácia	41,00±11,97	64,00*±18,38	52,00±19,32	52,00±19,32	59,00±14,49	57,33±19,17	54,67±24,10	52,67±21,93
Sensibilidade	30,00±19,86	61,50±28,29	45,00±26,67	47,00±28,60	58,00±19,75	61,67±24,91	58,33±30,68	52,50±24,23
Especificidade	49,00±10,78	67,00±16,81	56,67±19,12	54,67±19,76	61,00±22,88	51,67±30,88	53,33±31,23	51,67±35,53

\*acurácia média máxima

Como pode ser observado na Tabela 5.10, os maiores valores encontrados para a acurácia média e para a especificidade média foram obtidos com o uso do primeiro mínimo da função de informação mútua ( $FMMI$ ). No caso da medida de sensibilidade média as características ( $FMMI$ ) e a entropia aproximada ( $ApEn$ ) apresentaram desempenhos semelhantes.

## II - Medidas de Quantificação de Recorrência (MQR)

Na Tabela 5.11 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de quantificação de recorrência (MQR) na classificação entre vozes afetadas por paralisia e vozes afetadas por edema, por meio da função discriminante linear (LDA).

**Tabela 5.11** Taxas de classificação (%) Paralisia x Edema - MQR avaliadas individualmente (LDA)

Medidas	DET	$L_{max}$	ENTR	TREND	LAM	$V_{max}$	TT
Acurácia	63,00*±18,29	63,00*±17,67	57,00±18,29	46,00±20,66	55,00±15,81	49,00±9,94	52,00±12,29
Sensibilidade	77,14±30,24	44,71±21,54	61,21±20,90	55,62±21,00	64,31±29,54	59,95±30,42	42,71±28,87
Especificidade	48,32±29,03	72,82±29,70	47,96±29,01	41,95±27,96	45,05±23,87	41,54±15,85	57,55±26,40

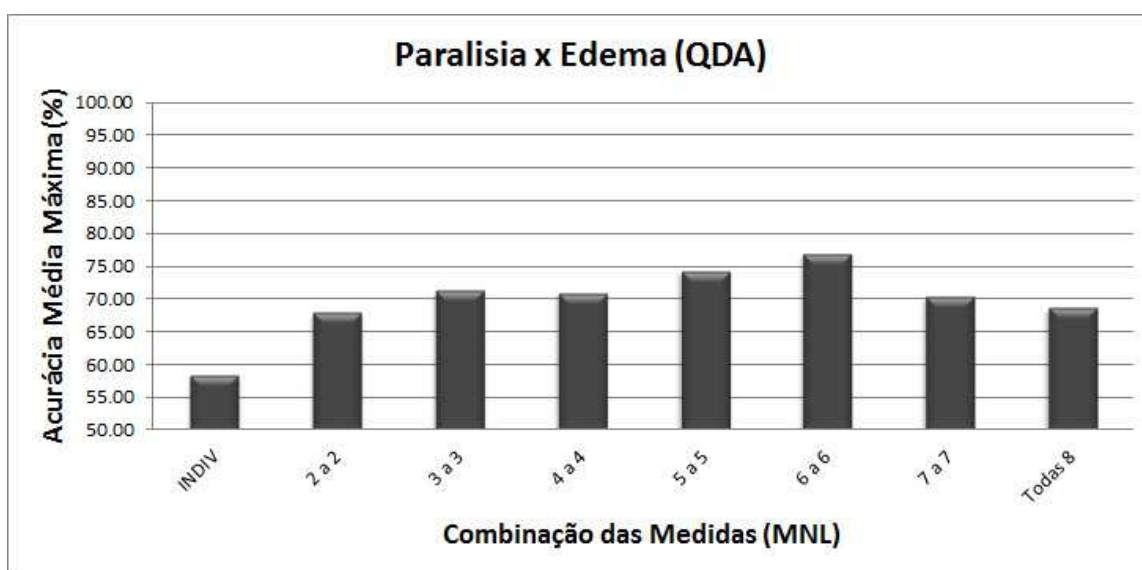
\*acurácia média máxima

As características  $DET$  e  $L_{max}$  apresentaram desempenhos praticamente idênticos, em termos dos valores de acurácia média máxima. O melhor resultado relativo à sensibilidade média foi obtido com o uso da medida  $DET$  e a especificidade média alcançou taxa mais alta por meio do uso de  $L_{max}$ .

### 5.6.2 Características Avaliadas de Forma Combinada

Nesta seção são apresentados os resultados dos experimentos em que foram avaliadas as combinações entre as medidas da análise não linear (MNL), entre as medidas de recorrência (MQR) e entre as características desses dois conjuntos (MNL e MQR), com o objetivo de classificar os sinais de vozes afetadas por paralisia e vozes com edema nas pregas vocais.

#### I - Medidas da Análise Não Linear (MNL)



**Figura 5.13** Acurácia Média Máxima - Paralisia x Edema (MNL)

Na Figura 5.13 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas da análise não linear (MNL), bem como de todas as combinações possíveis entre elas, na discriminação entre os sinais de vozes com paralisia e vozes com edema, por meio da função discriminante quadrática (QDA).

O maior valor para a acurácia média máxima foi de  $77,00 \pm 20,39\%$ , obtido com a combinação de seis características,  $\lambda_1$ ,  $FMMI$ ,  $H_q$ ,  $ApEn$ ,  $D_2$  e  $K_2$ . A medida de sensibilidade média alcançou  $74,17 \pm 26,19\%$  e a especificidade média o valor máximo de  $81,67 \pm 25,40\%$  para esse mesmo conjunto de características.

## II - Medidas de Quantificação de Recorrência (MQR)

Na Figura 5.14 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas de quantificação de recorrência (MQR), bem como de todas as combinações possíveis entre elas, na classificação entre vozes afetadas por paralisia e vozes com edema, por meio da função discriminante linear (LDA).

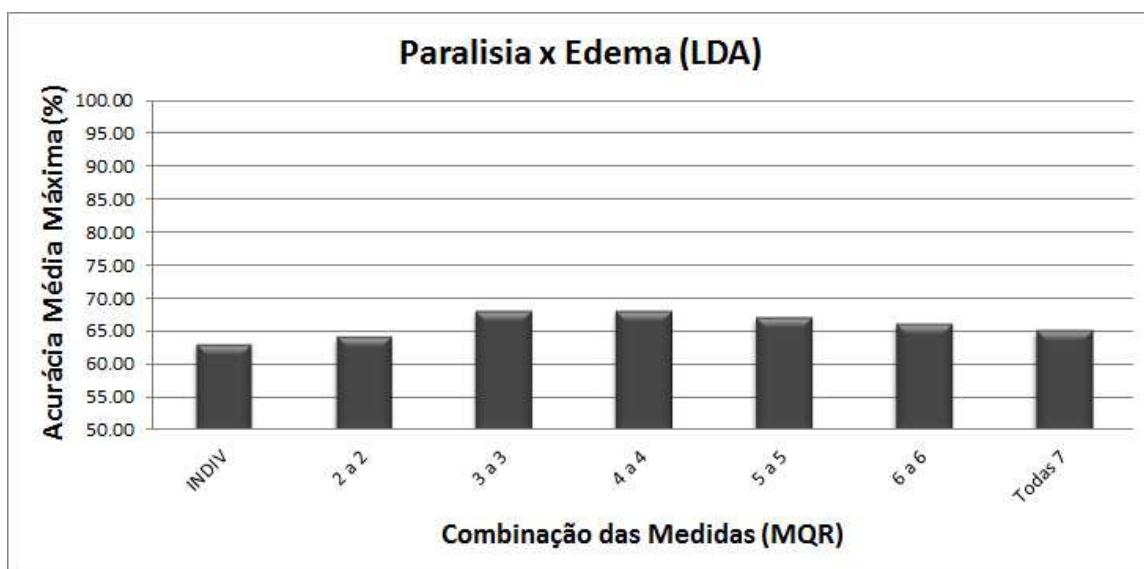
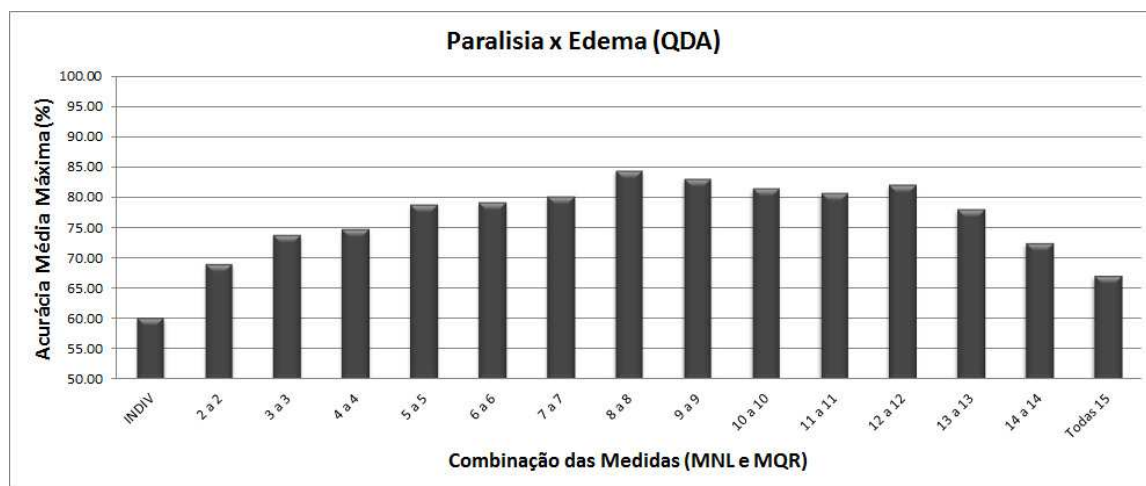


Figura 5.14 Acurácia Média Máxima - Paralisia x Edema (MQR)

O maior valor da acurácia média máxima foi obtido por três diferentes conjuntos com a combinação de três características. Na combinação das características  $DET$ ,  $L_{max}$  e  $TREND$  a acurácia máxima foi de  $68,00 \pm 11,35\%$ . Para esta mesma combinação foi atingido o melhor valor para a sensibilidade,  $75,00 \pm 20,00\%$ . Em outra combinação:  $DET$ ,  $LAM$  e  $TT$ , com acurácia máxima de  $68,00 \pm 16,19\%$ , foi obtida o melhor valor para a especificidade média de  $72,33 \pm 20,43\%$  e na terceira combinação:  $DET$ ,  $TREND$  e  $LAM$ , acurácia média máxima foi de  $68,00 \pm 14,76\%$ .

### III - Medidas da Análise Não Linear e Medidas de Quantificação de Recorrência

Na Figura 5.15 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das características referentes às medidas MNL e MQR, bem como de todas as combinações possíveis entre elas, na classificação de sinais de vozes afetadas por paralisia e de vozes com edema nas pregas vocais, por meio da função discriminante quadrática (QDA).



**Figura 5.15** Acurácia Média Máxima - Paralisia x Edema (MNL e MQR)

Como pode ser observado a partir do gráfico mostrado na Figura 5.15, o maior valor para a acurácia média máxima foi obtido na combinação de oito características:  $\lambda_1$ ,  $ApEn$ ,  $D_2$ ,  $K_2$ ,  $DET$ ,  $ENTR$ ,  $TREND$  e  $TT$ . Para este conjunto foi obtida uma acurácia média máxima de  $84,33 \pm 13,52\%$ , uma sensibilidade média igual a  $86,67 \pm 17,21\%$  e especificidade média de  $81,67 \pm 25,40\%$ .

## 5.7 Classificação de Vozes Patológicas - Paralisia versus Nódulo

Nesta seção são apresentados os resultados da classificação entre sinais de vozes afetadas por paralisia e vozes com nódulos nas pregas vocais.

### 5.7.1 Características Avaliadas Individualmente

Os resultados do processo de classificação utilizando cada característica de forma individual estão apresentados nas Tabelas 5.12-5.13.

## I - Medidas da Análise Não Linear (MNL)

Na Tabela 5.12 estão apresentados os resultados das medidas de desempenho referentes à utilização das medidas de análise não linear (MNL) na classificação entre sinais de vozes afetadas por paralisia e vozes com nódulos, por meio da função discriminante linear (LDA).

**Tabela 5.12** Taxas de classificação (%) Paralisia x Nódulo - MNL avaliadas individualmente (LDA)

Medidas	$\lambda_1$	$FMMI$	$H_1$	$H_q$	$H$	$ApEn$	$D_2$	$K_2$
Acurácia	52,32±23,34	66,43*±23,87	55,89±17,74	49,64±15,65	38,39±20,29	58,00±17,51	64,00±25,03	63,50±26,67
Sensibilidade	60,00±45,95	65,00±33,75	60,00±39,44	45,00±43,78	30,00±42,16	70,00±42,16	75,00±42,49	75,00±42,49
Especificidade	50,67±21,99	68,33±26,49	56,33±20,03	51,33±20,32	39,33±22,38	48,33±21,44	58,33±25,15	55,00±35,18

\*acurácia média máxima

O primeiro mínimo da função de informação mútua ( $FMMI$ ) é a característica que apresentou os maiores valores para as medidas de acurácia média e especificidade média. O maior valor para a sensibilidade média foi obtido por meio da dimensão de correlação ( $D_2$ ) e, também, com o uso da entropia de correlação ( $K_2$ ).

## II - Medidas de Quantificação de Recorrência (MQR)

Na Tabela 5.13 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de quantificação de recorrência (MQR) na classificação entre sinais de vozes afetadas por paralisia e vozes com nódulos, por meio da função discriminante linear (LDA).

**Tabela 5.13** Taxas de classificação (%) Paralisia x Nódulo - MQR avaliadas individualmente (LDA)

Medidas	$DET$	$L_{max}$	$ENTR$	$TREND$	$LAM$	$V_{max}$	$TT$
Acurácia	55,00±21,95	39,82±13,10	51,07±24,47	36,31±11,40	50,36±22,03	55,18*±10,86	49,46±18,79
Sensibilidade	65,00±41,16	23,33±25,09	56,67±43,89	46,67±37,52	65,00±41,16	73,33±37,02	53,33±44,31
Especificidade	52,67±24,88	43,33±19,12	51,00±25,09	34,00±17,55	47,33±25,03	48,67±9,19	48,00±19,58

\*acurácia média máxima

O comprimento máximo das estruturas verticais ( $V_{max}$ ) e o determinismo ( $DET$ ) apresentaram desempenho semelhante, no caso dos maiores valores para a acurácia média. O melhor desempenho para a sensibilidade média foi registrado com uso de  $V_{max}$  e o maior valor para especificidade média foi encontrado com a característica  $DET$ .

### 5.7.2 Características Avaliadas de Forma Combinada

Nesta seção são apresentados os resultados dos experimentos em que foram avaliadas as combinações entre as medidas da análise não linear (MNL), entre as medidas de recorrência (MQR) e também entre todas as características de ambos os conjuntos (MNL e MQR), com o objetivo de classificar os sinais de vozes afetadas por paralisia e com nódulos nas pregas vocais, por meio da função discriminante quadrática (QDA).

## I - Medidas da Análise Não Linear (MNL)

Na Figura 5.16 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas da análise não linear (MNL), bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes afetadas por paralisia e vozes com nódulos.

Como pode ser observado no gráfico apresentado na Figura 5.16, o maior valor para a acurácia média máxima foi obtido com a combinação de cinco características:  $\lambda_1$ ,  $H$ ,  $ApEn$ ,  $D_2$  e  $K_2$ . Para esse conjunto de características o valor da acurácia máxima foi de  $86,00 \pm 15,42\%$ ; a medida de sensibilidade média alcançou  $65,00 \pm 41,16\%$  e a especificidade média o valor máximo de  $96,67 \pm 10,54\%$ .

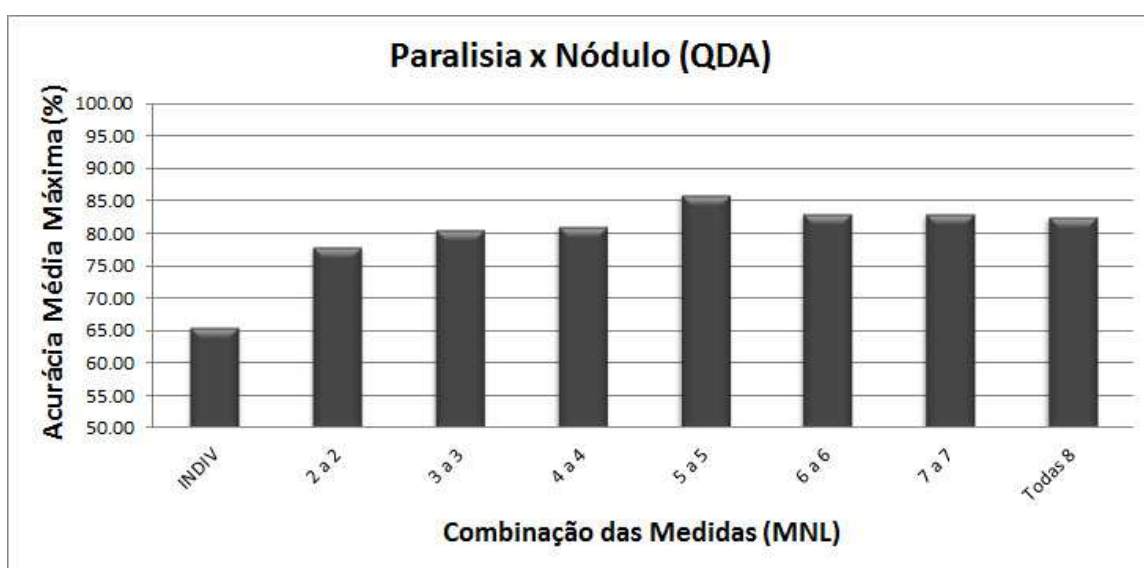


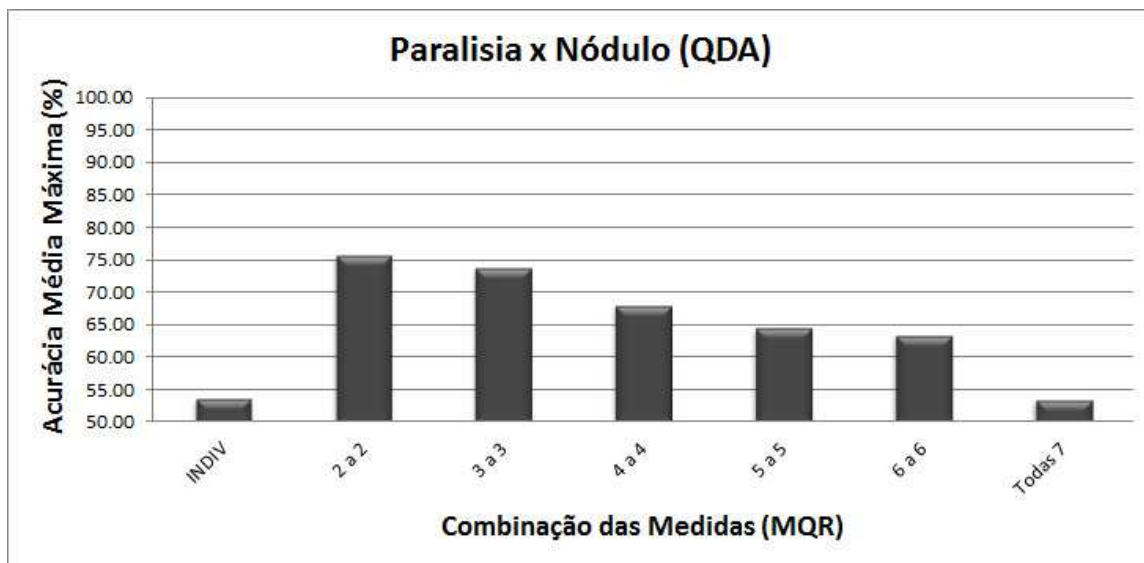
Figura 5.16 Acurácia Média Máxima - Paralisia x Nódulo (MNL)

## II - Medidas de Quantificação de Recorrência (MQR)

Na Figura 5.17 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas de quantificação de recorrência investigadas, bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes afetadas por paralisia e vozes com nódulos.

Como pode ser observado no gráfico apresentado na Figura 5.17, o maior valor para a acurácia média máxima foi obtido com a combinação de apenas duas características, usando QDA. Tal conjunto é formado pelas medidas:  $DET$  e  $LAM$  e apresentou valor máximo de acurácia média igual a  $75,71 \pm 16,46\%$ ; sensibilidade média igual a  $65 \pm 47,43\%$  e especificidade média de  $81,24 \pm 19,99\%$ .

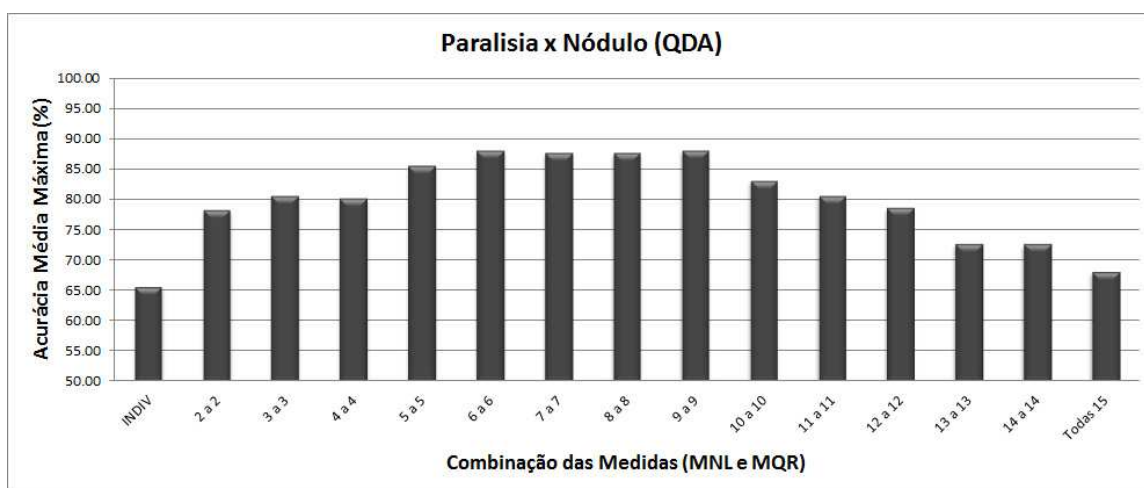




**Figura 5.17** Acurácia Média Máxima - Paralisia x Nódulo (MQR)

### III - Medidas da Análise Não Linear e Medidas de Quantificação de Recorrência

Na Figura 5.18 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual de todas as características dos conjuntos MNL e MQR, bem como de todas as combinações possíveis entre elas, na classificação de sinais de vozes afetadas por paralisia e de vozes com nódulos nas pregas vocais.



**Figura 5.18** Acurácia Média Máxima - Paralisia x Nódulo (MNL e MQR)

O conjunto com o menor número de características que atingiu o maior valor para acurácia média máxima é composto por seis medidas:  $\lambda_1$ ,  $H$ ,  $ApEn$ ,  $K_2$ ,  $ENTR$  e  $TT$ , usando QDA. Para esse conjunto de características, a acurácia foi de  $88,00 \pm 17,35\%$ , a sensibilidade atingiu  $75,00 \pm 42,49\%$  e a especificidade foi de  $96,67 \pm 10,54\%$ .

## 5.8 Classificação de Vozes Patológicas - Edema versus Nódulo

Nesta seção são apresentados os resultados da classificação entre sinais de vozes afetadas por edema e vozes com nódulos nas pregas vocais.

### 5.8.1 Características Avaliadas Individualmente

Os resultados do processo de classificação utilizando cada característica de forma individual estão apresentados nas Tabelas 5.14-5.15.

#### I - Medidas da Análise Não Linear (MNL)

Na Tabela 5.14 estão apresentados os resultados das medidas de desempenho referentes à utilização das medidas de análise não linear (MNL) na classificação entre sinais de vozes afetadas por edema e vozes com nódulos por meio da função discriminante linear (LDA).

**Tabela 5.14** Taxas de classificação (%) Edema x Nódulo - MNL avaliadas individualmente (LDA)

Medidas	$\lambda_1$	<i>FMMI</i>	$H_1$	$H_q$	<i>H</i>	<i>ApEn</i>	$D_2$	$K_2$
Acurácia	52,86±21,32	60,00±16,56	39,76±13,89	39,76±13,89	57,62±18,81	52,00±19,47	65,00*±19,00	59,00±27,37
Sensibilidade	50,00±40,82	45,00±49,72	30,00±42,16	20,00±34,96	55,00±43,78	60,00±45,95	70,00±42,16	65,00±47,43
Especificidade	53,00±22,39	63,50±22,49	42,00±20,98	47,00±22,39	58,50±23,46	47,50±18,45	60,00±21,08	55,83±23,59

\*acurácia média máxima

A dimensão de correlação ( $D_2$ ) é a característica que apresentou os maiores valores para as medidas de acurácia média e sensibilidade média. O maior valor para a especificidade média foi obtido por meio da característica *FMMI*.

#### II - Medidas de Quantificação de Recorrência (MQR)

Na Tabela 5.15 estão apresentados os resultados das medidas de desempenho referentes à aplicação das medidas de quantificação de recorrência (MQR) na classificação entre sinais de vozes afetadas por edema e vozes com nódulos, por meio da função discriminante linear (LDA).

**Tabela 5.15** Taxas de classificação (%) Edema x Nódulo - MQR avaliadas individualmente (LDA)

Medidas	<i>DET</i>	$L_{max}$	<i>ENTR</i>	<i>TREND</i>	<i>LAM</i>	$V_{max}$	<i>TT</i>
Acurácia	39,29±21,77	55,24*±16,64	45,24±18,10	51,67±18,75	44,29±20,27	51,43±27,38	53,10±23,62
Sensibilidade	23,33±34,43	91,67±18,00	36,67±42,16	43,33±43,89	26,67±37,02	66,67±40,82	66,67±40,82
Especificidade	48,50±36,21	41,50±24,27	49,00±28,07	56,00±28,94	52,00±30,20	46,50±29,73	48,50±20,96

\*acurácia média máxima

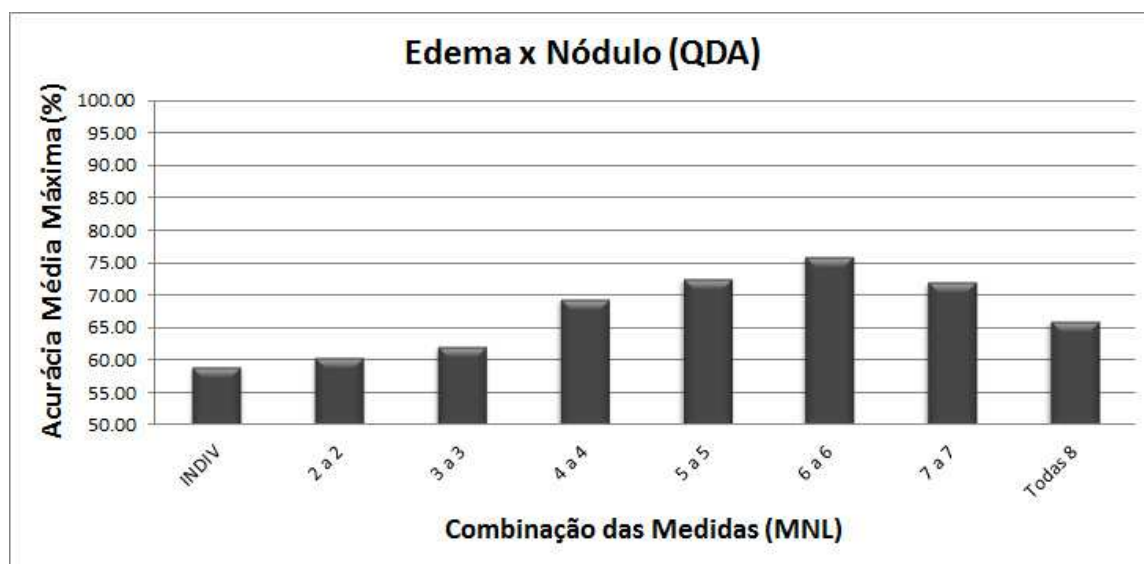
O comprimento máximo das linhas diagonais ( $L_{max}$ ) é a característica que apresentou os maiores valores para as medidas de acurácia média e sensibilidade média. Por outro lado, o uso da característica *TREND* apresentou o maior valor para a especificidade média.

## 5.8.2 Características Avaliadas de Forma Combinada

Nesta seção são apresentados os resultados dos experimentos em que foram avaliadas as combinações entre as medidas da análise não linear (MNL), entre as medidas de quantificação recorrência (MQR) e também entre todas as características desses dois conjuntos (MNL e MQR), com o objetivo de classificar os sinais de vozes afetadas por edema e com nódulos nas pregas vocais.

### I - Medidas da Análise Não Linear (MNL)

Na Figura 5.19 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas da análise não linear (MNL), bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes afetadas por edema e vozes com nódulos, por meio da análise discriminante quadrática (QDA).

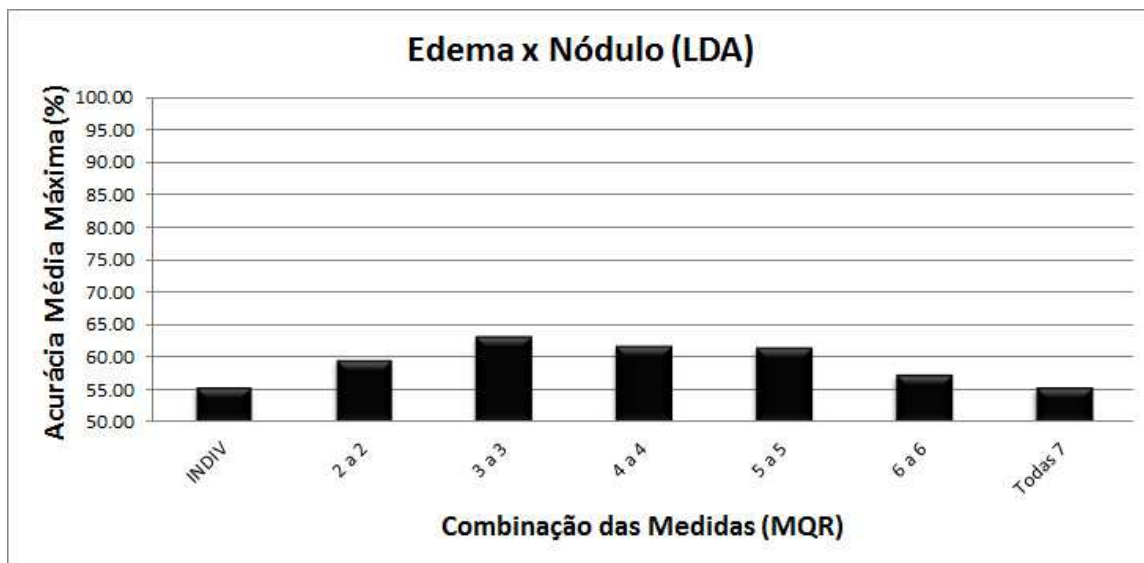


**Figura 5.19** Acurácia Média Máxima - Edema x Nódulo (MNL)

Como pode ser observado no gráfico apresentado na Figura 5.19, o maior valor para a acurácia média máxima foi obtido com a combinação de seis características:  $\lambda_1$ ,  $FMMI$ ,  $H_1$ ,  $H_q$ ,  $H$  e  $D_2$ . Para esse conjunto de características o valor da acurácia média máxima foi de  $76,00 \pm 13,29\%$ , a medida de sensibilidade média alcançou  $55,00 \pm 36,89\%$  e a especificidade média o valor máximo de  $86,67 \pm 17,21\%$ .

### II - Medidas de Quantificação de Recorrência (MQR)

Na Figura 5.20 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual das medidas de quantificação de recorrência (MQR), bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes afetadas por edema e vozes com nódulos, por meio da análise discriminante linear (LDA).

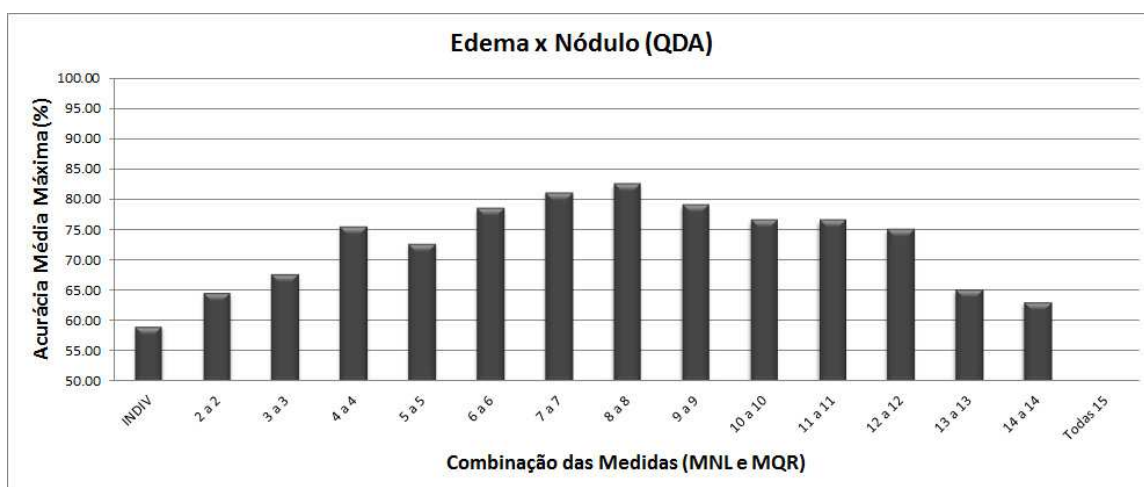


**Figura 5.20** Acurácia Média Máxima - Edema x Nódulo (MQR)

Como pode ser observado no gráfico apresentado na Figura 5.20, o maior valor para a acurácia média máxima, igual a  $63,10 \pm 13,89\%$ , foi obtido com a combinação de três características:  $L_{max}$ ,  $TREND$  e  $V_{max}$ . Para esse mesmo conjunto de características, o melhor valor para a sensibilidade média foi de  $81,67 \pm 24,15\%$  e para a especificidade média de  $58,00 \pm 19,75\%$ .

### III - Medidas da Análise Não Linear e Medidas de Quantificação de Recorrência

Na Figura 5.21 estão representados os valores da acurácia média máxima determinados a partir da avaliação individual de todas as características dos conjuntos MNL e MQR, bem como de todas as combinações possíveis entre elas, na classificação entre sinais de vozes afetadas por edema e de vozes com nódulos nas pregas vocais, por meio da análise discriminante quadrática (QDA).



**Figura 5.21** Acurácia Média Máxima - Edema x Nódulo (MNL e MQR)

Como pode ser observado a partir do gráfico mostrado na Figura 5.21, o maior valor para a acurácia média máxima foi obtido na combinação de oito características:  $\lambda_1$ ,  $H_1$ ,  $H_q$ ,  $H$ ,  $ApEn$ ,  $D_2$ ,  $L_{max}$  e  $ENTR$ , para o qual obteve-se acurácia de  $82,50 \pm 9,50\%$ ; sensibilidade de  $55,00 \pm 43,78\%$  e especificidade de  $93,33 \pm 14,50\%$ .

## 5.9 Intervalos de Confiança para a Média

Muito embora a média da amostra seja útil como um estimador não tendencioso da média da população, não existe maneira de expressar o grau de acurácia de um estimador por ponto. De fato, matematicamente falando, a probabilidade de que a média da amostra seja exatamente correta como um estimador é nula. Um intervalo de confiança para a média é um intervalo estimado, construído com respeito à média da amostra, pelo qual pode ser especificada a probabilidade de o intervalo incluir o valor da média da população. O grau de confiança, associado com um intervalo de confiança, indica a percentagem de tais intervalos que incluiriam o parâmetro que se está estimando [117].

O uso da distribuição normal na estimação da média populacional é garantido para qualquer grande amostra (tipicamente maior ou igual a 30), sendo-o para uma pequena amostra (menor que 30), somente se a população for normalmente distribuída e o desvio padrão da população conhecido. No caso em que a amostra é pequena e a população normalmente distribuída, mas o desvio padrão da população é desconhecido, os valores são distribuídos de acordo com a distribuição  $t$  de Student [117].

### 5.9.1 A Desigualdade de Chebyshev

Quando a amostra é pequena (menor que 30) e se supõe que a população não é normalmente distribuída, nem a distribuição normal de probabilidade, nem a distribuição  $t$  de Student podem ser usadas na construção de um intervalo de confiança [117]. Contudo, um teorema geral desenvolvido pelo matemático russo Chebyshev estabelece que, a probabilidade de que a média da amostra esteja dentro de  $k$  unidades de erro padrão da média da população é dada por

$$P(|\bar{x} - \mu| \leq k s_{\bar{x}}) \geq 1 - \frac{1}{k^2}, \quad (5.4)$$

em que  $\bar{x}$  é a média amostral,  $\mu$  a média populacional e  $s_{\bar{x}}$  é o erro padrão da amostra.

A expressão 5.4 é geralmente referida como a desigualdade de Chebyshev. A desigualdade de Chebyshev é, de fato, raramente usada na construção de intervalos de confiança para a média, mas é o único método apropriado, dada uma população decididamente não normal e uma amostra pequena [117].

Ao usar a desigualdade de Chebyshev na estimação por intervalo, iguala-se o fator  $(1 - 1/k^2)$  ao grau de confiança desejado, resolve-se para  $k$ , construindo-se, então, o intervalo de confiança utilizando a média da amostra como termo central. Os intervalos de confiança

mais frequentemente utilizados são os de 90%, 95% e 99%. No caso desta pesquisa foi adotado um nível de confiança de 95%, dessa forma, o valor calculado para o parâmetro  $k$  foi de 4,47.

## 5.10 Resumo dos Resultados do Processo de Classificação

Na Tabela 5.16 são apresentados os intervalos de confiança para a acurácia, construídos a partir dos resultados obtidos nos testes de validação cruzada com 10 iterações e utilizando a desigualdade de Chebychev. São mostrados os resultados nos casos em que a classificação foi realizada considerando-se o desempenho das características de forma individual (*INDIV*), bem como a partir da combinação de características (*COMBO*). Neste caso, os resultados apresentados indicam a combinação feita com o menor número de características que atingiu o maior valor de acurácia média máxima.

**Tabela 5.16** Intervalos de confiança para a acurácia do sistema de classificação dos sinais de voz analisados (nível de confiança de 95%).

Classificação	MNL		MQR		MNL e MQR
	INDIV	COMBO	INDIV	COMBO	COMBO
<i>SDL x PTL</i>	[75,41%;98,77%] <i>H</i> (QDA)	[87,90%; 100%] <i>H</i> <sub>1</sub> , <i>H</i> <sub>q</sub> , <i>H</i> e <i>D</i> <sub>2</sub> (QDA)	[82,49%; 100%] <i>L</i> <sub>max</sub> (QDA)	[83,31%; 100%] <i>L</i> <sub>max</sub> , <i>ENTR</i> , <i>LAM</i> e <i>V</i> <sub>max</sub> (LDA)	[94,33%; 100%] <i>H</i> <sub>1</sub> , <i>H</i> <sub>q</sub> , <i>H</i> , <i>DET</i> , <i>LAM</i> e <i>TT</i> (QDA)
<i>SDL x EDM</i>	[69,47%; 100%] <i>D</i> <sub>2</sub> (LDA)	[93,17%; 100%] <i>H</i> <sub>1</sub> , <i>H</i> <sub>q</sub> , <i>H</i> , <i>D</i> <sub>2</sub> e <i>K</i> <sub>2</sub> (LDA)	[67,67%; 97,89%] <i>L</i> <sub>max</sub> (LDA)	[80,53%; 100%] <i>DET</i> , <i>L</i> <sub>max</sub> , <i>ENTR</i> , <i>V</i> <sub>max</sub> e <i>TT</i> (LDA)	[93,17%; 100%] <i>H</i> <sub>1</sub> , <i>H</i> <sub>q</sub> , <i>H</i> , <i>D</i> <sub>2</sub> e <i>K</i> <sub>2</sub> (LDA)
<i>SDL x PRL</i>	[73,22%; 100%] <i>H</i> (QDA)	[90,05%; 100%] <i>FMMI</i> , <i>H</i> , <i>D</i> <sub>2</sub> e <i>H</i> <sub>1</sub> ou <i>H</i> <sub>q</sub> (QDA)	[82,46%; 100%] <i>L</i> <sub>max</sub> (QDA)	[88,68%; 100%] <i>DET</i> , <i>L</i> <sub>max</sub> , <i>ENTR</i> <i>V</i> <sub>max</sub> e <i>TT</i> (QDA)	100% <i>FMMI</i> , <i>H</i> e <i>DET</i> ou <i>ENTR</i> (QDA)
<i>SDL x NDL</i>	[83,86%; 100%] <i>H</i> (LDA e QDA)	[92,18%; 100%] <i>H</i> <sub>1</sub> , <i>H</i> <sub>q</sub> e <i>H</i> (LDA)	[82,64%; 100%] <i>L</i> <sub>max</sub> (QDA)	[89,08%; 100%] <i>L</i> <sub>max</sub> e <i>DET</i> ou <i>ENTR</i> ou <i>LAM</i> ou <i>TT</i> (QDA)	[93,17%; 100%] <i>H</i> , <i>DET</i> , <i>LAM</i> com <i>FMMI</i> e <i>ENTR</i> ou com <i>H</i> <sub>1</sub> e <i>H</i> <sub>q</sub> (QDA)
<i>PRL x EDM</i>	[38,02%; 89,98%] <i>FMMI</i> (LDA)	[48,18%; 100%] $\lambda_1$ , <i>FMMI</i> , <i>H</i> <sub>q</sub> , <i>ApEn</i> , <i>D</i> <sub>2</sub> e <i>K</i> <sub>2</sub> (QDA)	[38,02%; 87,98%] <sup>(1)</sup> [37,15%; 88,85%] <sup>(2)</sup> <sup>(1)</sup> <i>L</i> <sub>max</sub> <sup>(2)</sup> <i>DET</i> (LDA)	[51,96%; 84,04%] <sup>(3)</sup> [45,11%; 90,89%] <sup>(4)</sup> [47,14; 88,86%] <sup>(5)</sup> <sup>(3)</sup> <i>DET</i> , <i>L</i> <sub>max</sub> , <i>TREND</i> <sup>(4)</sup> <i>DET</i> , <i>LAM</i> e <i>TT</i> <sup>(5)</sup> <i>DET</i> , <i>TREND</i> e <i>LAM</i> (LDA)	[65,22%; 100%] $\lambda_1$ , <i>ApEn</i> , <i>D</i> <sub>2</sub> , <i>K</i> <sub>2</sub> , <i>DET</i> , <i>ENTR</i> <i>TREND</i> e <i>TT</i> (QDA)
<i>PRL x NDL</i>	[32,69%; 100%] <i>FMMI</i> (LDA)	[64,20%; 100%] $\lambda_1$ , <i>H</i> , <i>ApEn</i> , <i>D</i> <sub>2</sub> e <i>K</i> <sub>2</sub> (QDA)	[39,83%; 70,53%] <i>V</i> <sub>max</sub> (LDA)	[52,44%; 98,98%] <i>DET</i> e <i>LAM</i> (QDA)	[63,48%; 100%] $\lambda_1$ , <i>H</i> , <i>ApEn</i> , <i>K</i> <sub>2</sub> , <i>ENTR</i> e <i>TT</i> (QDA)
<i>EDM x NDL</i>	[38,14%; 91,86%] <i>D</i> <sub>2</sub> (LDA)	[57,21%; 94,79%] $\lambda_1$ , <i>FMMI</i> , <i>H</i> <sub>1</sub> , <i>H</i> <sub>q</sub> , <i>H</i> e <i>D</i> <sub>2</sub> (QDA)	[31,72%; 78,76%] <i>L</i> <sub>max</sub> (LDA)	[43,47%; 82,73%] <i>L</i> <sub>max</sub> , <i>TREND</i> e <i>V</i> <sub>max</sub> (LDA)	[69,07%; 95,93%] $\lambda_1$ , <i>H</i> <sub>1</sub> , <i>H</i> <sub>q</sub> , <i>ApEn</i> , <i>D</i> <sub>2</sub> , <i>H</i> <i>L</i> <sub>max</sub> e <i>ENTR</i> (QDA)

Pode ser observado, a partir dos resultados apresentados na Tabela 5.16, a grande dificuldade em se discriminar entre as patologias avaliadas usando as medidas consideradas (MNL e MQR). Diante disso, é investigado na próxima seção os efeitos no processo de classificação entre os diferentes grupos de sinais de vozes, do uso de vetores híbridos obtidos da combinação das características dos conjuntos MNL e MQR com coeficientes extraídos na análise de predição linear.

## 5.11 Combinando Análise Linear e Análise Não Linear

Visando avaliar os efeitos da utilização simultânea de características extraídas da análise linear e da análise não linear nas taxas de classificação entre sinais de vozes saudáveis e vozes patológicas, e entre sinais de vozes apresentando diferentes patologias, foram considerados vetores híbridos, formados pela combinação de características dos conjuntos MNL e MQR e vetores de coeficientes obtidos por meio da análise LPC - Codificação por Predição Linear.

A predição linear é usada intensivamente em sistemas de codificação de voz [118]. Ela usa uma combinação linear de amostras passadas do sinal para construir uma previsão para o valor que melhor se aproxima das novas entradas. Assim, para um sinal  $s(n)$  tem-se que

$$s_p(n) = \sum_{k=1}^{k=p} a_k s(n-k), \quad (5.5)$$

fornece o valor previsto do sinal no instante  $n$ . Em que  $p$  é a ordem do preditor e os coeficientes  $a_k$  são chamados de coeficientes de predição linear (LPC) e devem ser escolhidos de modo a aproximar  $s_p(n)$  de  $s(n)$ .

Neste trabalho de pesquisa os coeficientes LPC foram calculados por meio do método da autocorrelação [119]. Inicialmente é investigado o potencial discriminativo dos coeficientes LPC obtidos a partir de preditores de diferentes ordens (8, 12, 16, 20, 24, 28 e 32) na classificação entre as diferentes classes de sinais consideradas. Em seguida, são avaliados os efeitos do processo de classificação considerando-se a utilização de vetores híbridos, formados pela combinação dos coeficientes LPC e as características que apresentaram os melhores valores de acurácia, de forma individual e combinada, conforme apresentado na Tabela 5.16.

### 5.11.1 Classificação Saudável x Patológica

De acordo com os resultados apresentados na Tabela 5.16, a medida de quantificação de recorrência  $L_{max}$  é a característica que apresentou, entre todas as avaliadas (MNL e MQR) de maneira individual, o maior valor de acurácia média na classificação entre sinais de vozes saudáveis e sinais de vozes patológicas.

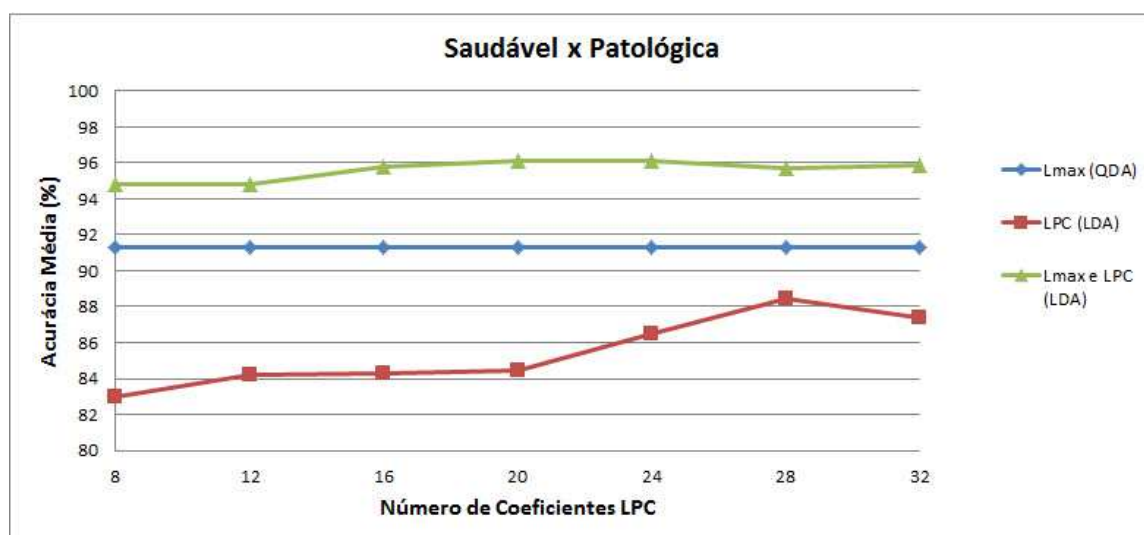
Na Tabela 5.17 estão apresentados os resultados das medidas de desempenho referentes à combinação de  $L_{max}$  com os vetores de coeficientes LPC na classificação entre sinais de vozes saudáveis e sinais de vozes patológicas, por meio da função discriminante linear (LDA).

**Tabela 5.17** Taxas de classificação (%) Saudável x Patológica - LPC e  $L_{max}$  (LDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	94,79±0,59	94,85±0,99	95,77±0,69	96,08±0,87	96,12*±0,75	95,69±1,18	95,90±0,67
Sensibilidade	92,83±2,02	93,46±2,34	96,60±1,04	96,16±1,55	96,23±2,03	95,85±1,86	96,73±1,56
Especificidade	95,67±0,97	95,47±0,80	95,39±1,12	96,04±1,07	96,07±1,08	95,62±1,19	95,53±1,19

\*acurácia média máxima

Na Figura 5.22 estão representados os valores da acurácia média determinados a partir da avaliação individual da característica  $L_{max}$ , da análise LPC e da combinação dessas características na classificação entre sinais de vozes saudáveis e de vozes patológicas.



**Figura 5.22** Acurácia Média - Saudável x Patológica

Como pode ser observado, o uso de vetores híbridos formados pela combinação da medida  $L_{max}$  e coeficientes LPC resultou em um aumento significativo nos valores da acurácia do sistema de classificação. Por exemplo, o valor de 94,79±0,59% (Tabela 5.17) obtido para a acurácia média com o uso do vetor híbrido formado por  $L_{max}$  e oito coeficientes LPC está em torno de seis pontos percentuais acima do melhor resultado obtido usando apenas análise LPC com 28 coeficientes (88,41±1,47%) e três pontos percentuais acima do maior valor de acurácia média obtida com o uso apenas de  $L_{max}$  (91,27±6,21%).

Ainda de acordo com os resultados mostrados na Tabela 5.16, a combinação de características  $H_1$ ,  $H_q$ ,  $H$ ,  $DET$ ,  $LAM$  e  $TT$  foi a que apresentou o maior valor para a acurácia média na classificação entre sinais de vozes saudáveis e vozes patológicas.

Na Tabela 5.18 estão apresentados os resultados das medidas de desempenho referentes à combinação das características  $H_1$ ,  $H_q$ ,  $H$ ,  $DET$ ,  $LAM$  e  $TT$  com os vetores de coeficientes LPC, por meio da função discriminante linear (LDA).

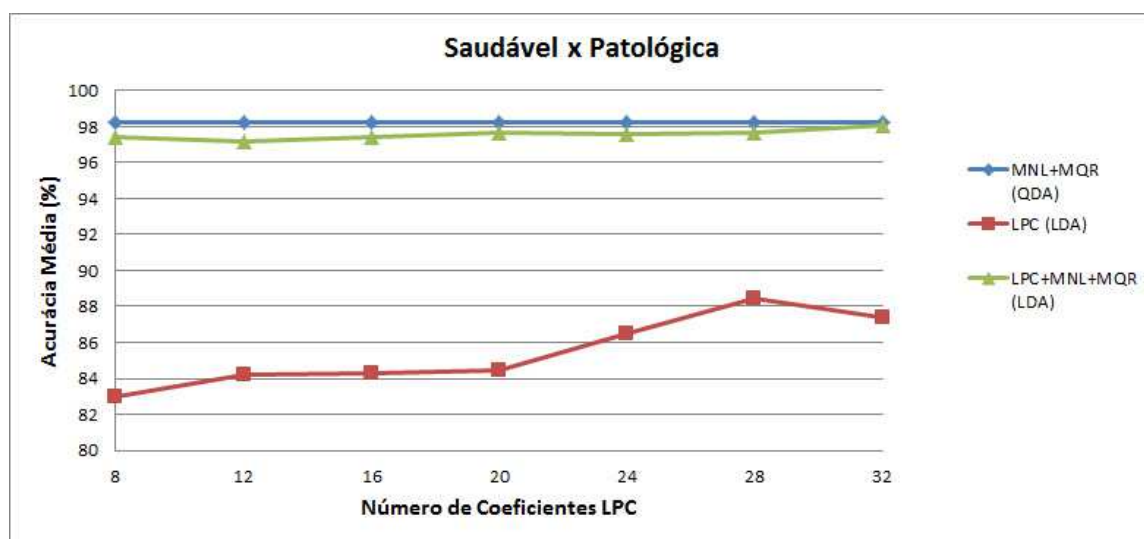


**Tabela 5.18** Taxas de classificação (%) Saudável x Patológica - LPC+MNL+MQR (LDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	97,40±0,66	97,19±0,71	97,39±0,64	97,66±0,52	97,56±0,69	97,64±1,08	98,11*±0,43
Sensibilidade	98,11±1,19	98,05±1,20	98,81±0,86	98,74±0,59	98,43±1,20	99,18±0,79	99,94±0,20
Especificidade	97,09±0,92	96,80±0,94	96,75±0,98	97,17±0,74	97,17±0,69	96,95±1,52	97,28±0,58

\*acurácia média máxima

Na Figura 5.23 estão apresentados os valores médios da acurácia obtidos na combinação das características das MNL e MQR, bem como os obtidos pela análise LPC e também pela combinação dos coeficientes LPC e esse conjunto de características.

**Figura 5.23** Acurácia Média - Saudável x Patológica

Como pode ser observado, o valor máximo da acurácia média apresentado na Tabela 5.18 foi de  $98,11 \pm 0,43\%$ , obtido com o acréscimo de 32 coeficientes LPC à combinação das características das MNL e MQR. Assim, pode-se concluir que a formação dos vetores híbridos não contribuiu para uma melhora significativa, em termos da acurácia média, em relação ao obtido apenas com a combinação das características MNL e MQR, ou seja,  $98,27 \pm 2,79\%$ .

### 5.11.2 Classificação Saudável x Edema

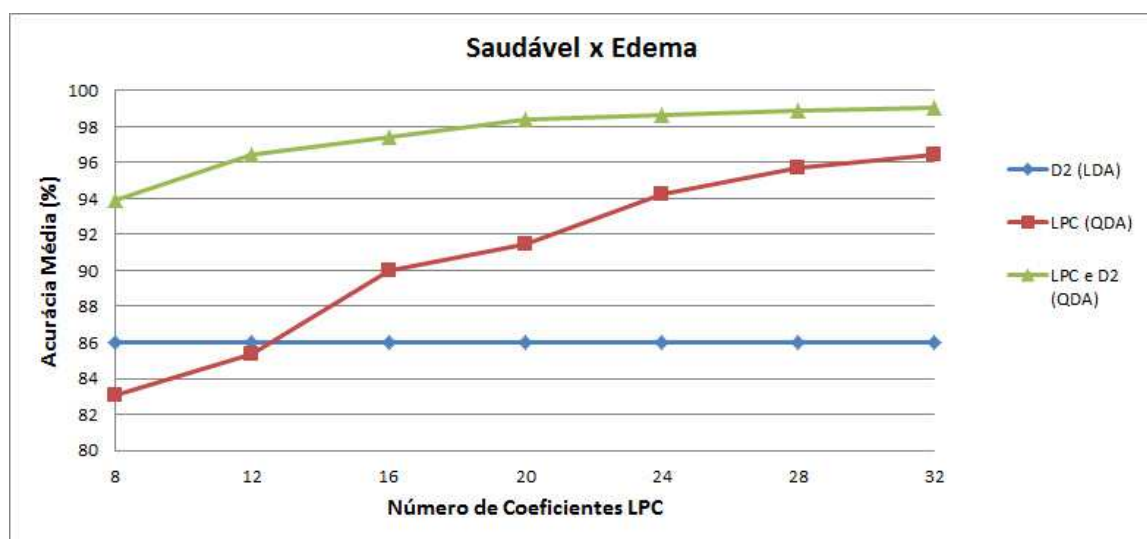
De acordo com os resultados apresentados na Tabela 5.16, a dimensão de correlação ( $D_2$ ) é a característica que apresentou o maior valor de acurácia média na classificação entre sinais de vozes saudáveis e sinais de vozes afetadas por edema nas pregas vocais. Na Tabela 5.19 estão apresentados os resultados das medidas de desempenho referentes à combinação de  $D_2$  com os vetores de coeficientes LPC na classificação entre sinais de vozes saudáveis e sinais de vozes afetadas por edema, por meio da função discriminante quadrática (QDA).

**Tabela 5.19** Taxas de classificação (%) Saudável x Edema - LPC e  $D_2$  (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	93,89±2,06	96,43±1,15	97,38±1,22	98,37±0,82	98,61±0,94	98,85±0,80	99,05*±0,65
Sensibilidade	95,28±1,97	97,36±0,93	98,18±1,71	98,93±1,03	98,99±1,12	99,37±0,66	99,56±0,67
Especificidade	91,51±3,42	94,84±2,77	96,02±1,68	97,42±1,26	97,96±1,79	97,96±1,56	98,17±1,61

\*acurácia média máxima

Na Figura 5.24 estão representados os valores da acurácia média determinados a partir da avaliação individual da característica  $D_2$ , da análise LPC e da combinação dessas características na classificação entre sinais de vozes saudáveis e de vozes afetadas por edema.

**Figura 5.24** Acurácia Média - Saudável x Edema

É possível observar um ganho em torno de oito pontos percentuais na taxa de acurácia média no caso da combinação de  $D_2$  com oito coeficientes LPC, em relação à classificação utilizando apenas a característica  $D_2$ . Além disso, uma acurácia média acima de 98% foi obtida com o uso de vetores híbridos contendo a partir de 20 coeficientes da análise LPC. O valor máximo da acurácia média foi de 99,05±0,65% com o uso de 32 coeficientes LPC combinados com  $D_2$ .

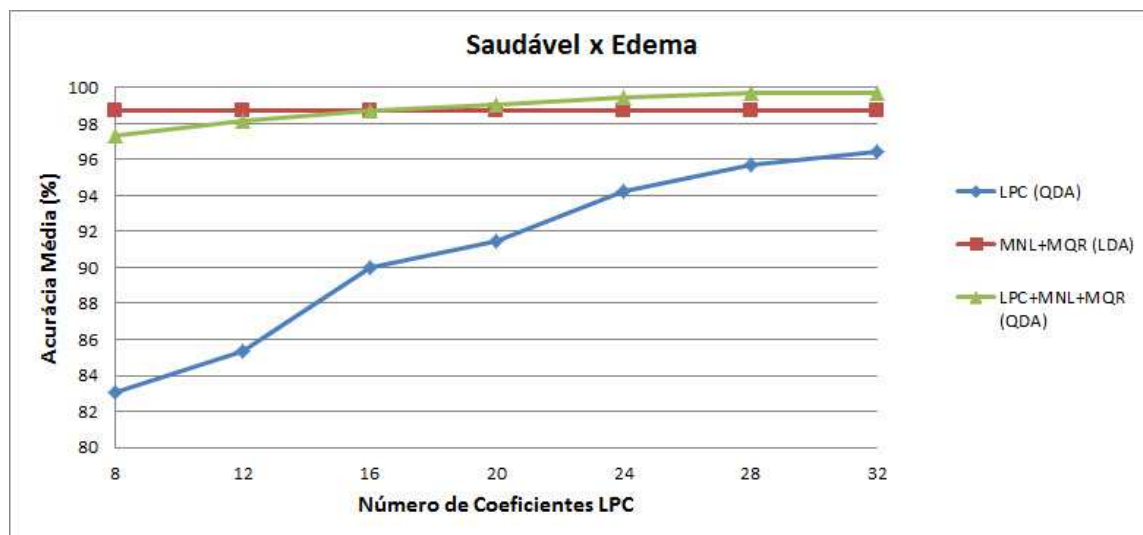
Ainda de acordo com os resultados mostrados na Tabela 5.16, a combinação das características  $H_1$ ,  $H_q$ ,  $H$ ,  $D_2$  e  $K_2$  foi a que apresentou o maior valor para a acurácia média na classificação entre sinais de vozes saudáveis e vozes afetadas por edema. Na Tabela 5.20 estão apresentados os resultados das medidas de desempenho referentes à combinação desse conjunto de características com os vetores de coeficientes LPC, por meio da função discriminante quadrática (QDA).

Na Figura 5.25 estão apresentados os valores médios da acurácia obtidos na combinação das características dos conjuntos MNL e MQR, bem como os obtidos pela análise LPC e ainda pela combinação dos coeficientes LPC e esse conjunto de medidas.

**Tabela 5.20** Taxas de classificação (%) Saudável x Edema - LPC+MNL+MQR (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	97,30±0,95	98,17±0,50	98,69±0,82	99,09±0,65	99,44±0,60	99,68±0,36	99,72*±0,27
Sensibilidade	98,11±1,29	99,31±0,63	99,87±0,27	100	99,81±0,30	99,94±0,20	99,94±0,20
Especificidade	95,91±1,81	96,24±1,04	96,67±2,24	97,53±1,76	98,82±1,56	99,25±1,02	99,35±0,75

\*acurácia média máxima

**Figura 5.25** Acurácia Média - Saudável x Edema

Como pode ser observado, há um ganho com uso de vetores híbridos a partir de 20 coeficientes LPC. A acurácia média atinge um valor máximo de  $99,72 \pm 0,27\%$  com o uso de 32 coeficientes LPC e o conjunto das características  $H_1$ ,  $H_q$ ,  $H$ ,  $D_2$  e  $K_2$ .

### 5.11.3 Classificação Saudável x Paralisia

De acordo com os resultados apresentados na Tabela 5.16, a medida de quantificação de recorrência  $L_{max}$  é a característica que apresentou o maior valor de acurácia média na classificação entre sinais de vozes saudáveis e sinais de vozes com paralisia.

Na Tabela 5.21 estão apresentados os resultados das medidas de desempenho referentes à combinação da característica  $L_{max}$  com os vetores de coeficientes LPC na classificação entre sinais de vozes saudáveis e sinais de vozes com paralisia, por meio da função discriminante linear (LDA).

**Tabela 5.21** Taxas de classificação (%) Saudável x Paralisia - LPC e  $L_{max}$  (LDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	97,09±0,67	97,28±0,73	97,31±1,01	97,40±0,88	97,74*±1,10	97,28±0,52	97,53±0,80
Sensibilidade	96,79±1,31	97,42±1,24	97,67±1,48	97,67±1,07	98,55±1,07	97,23±1,16	96,67±0,94
Especificidade	97,39±1,11	97,14±1,22	96,96±1,68	97,14±1,59	96,96±1,80	97,33±0,91	97,39±1,31

\*acurácia média máxima

Na Figura 5.26 estão representados os valores da acurácia média determinados a partir da avaliação individual da característica  $L_{max}$ , da análise LPC e da combinação dessas características na classificação entre sinais de vozes saudáveis e de vozes com paralisia. Como pode ser observado, o uso de vetores híbridos oferece uma melhora significativa na acurácia média do processo de classificação, em torno de cinco pontos percentuais para o caso do uso de oito coeficientes LPC, em relação à classificação realizada apenas com a característica  $L_{max}$ .

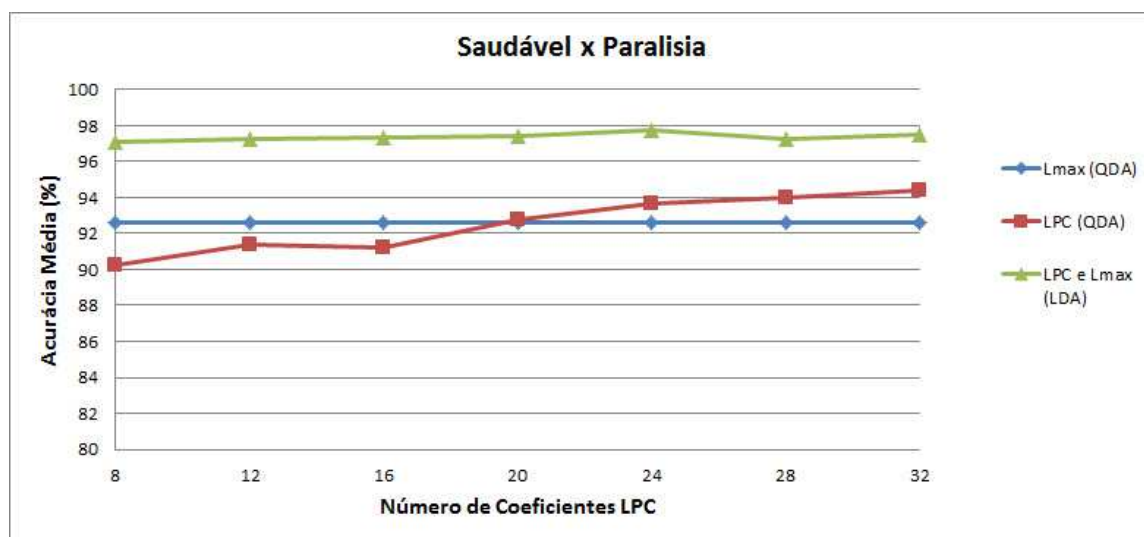


Figura 5.26 Acurácia Média - Saudável x Paralisia

Ainda de acordo com os resultados mostrados na Tabela 5.16, as combinações das características  $FMMI$  e  $H$  com  $DET$  ou com  $ENTR$  apresentaram a taxa máxima de 100% para a acurácia média na classificação entre sinais de vozes saudáveis e vozes afetadas por paralisia.

#### 5.11.4 Classificação Saudável x Nódulo

De acordo com os resultados apresentados na Tabela 5.16, o expoente de Hurst ( $H$ ) é a característica que apresentou o maior valor de acurácia média na classificação entre sinais de vozes saudáveis e sinais de vozes com nódulos.

Na Tabela 5.22 estão apresentados os resultados das medidas de desempenho referentes à combinação do expoente de Hurst com os vetores de coeficientes LPC na classificação entre sinais de vozes saudáveis e sinais de vozes com nódulos, por meio da função discriminante quadrática (QDA).

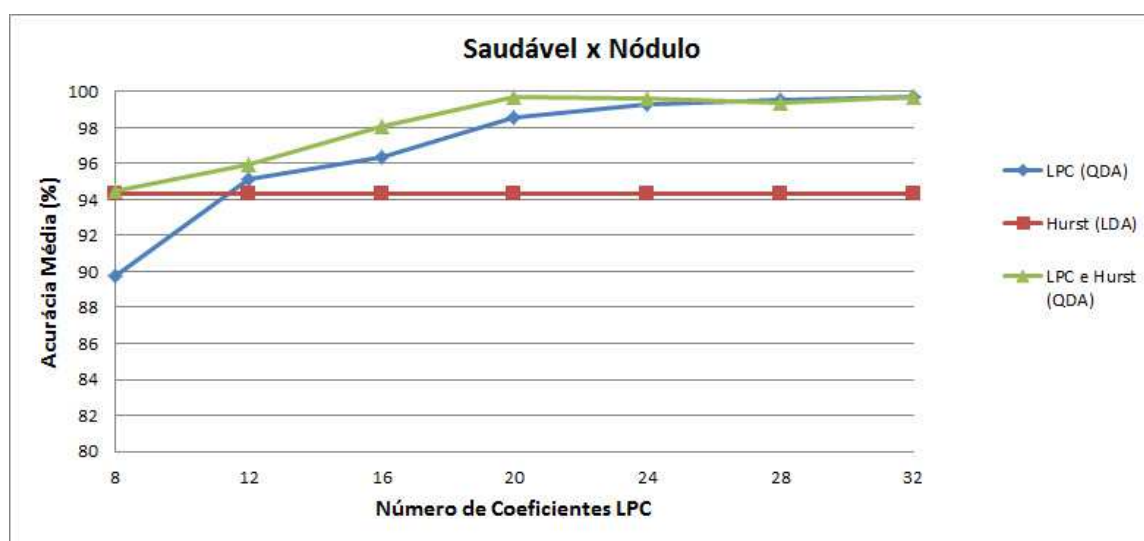
Tabela 5.22 Taxas de classificação (%) Saudável x Nódulo - LPC e  $H$  (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	94,51±1,57	95,92±0,91	98,08±1,37	99,72*±0,40	99,62±0,62	99,34±0,50	99,67±0,39
Sensibilidade	93,33±2,16	95,53±1,61	97,80±1,71	99,87±0,27	99,75±0,61	99,43±0,55	99,75±0,44
Especificidade	97,96±1,84	97,04±2,50	98,89±1,29	99,26±1,56	99,26±0,96	99,07±1,31	99,44±0,89

\*acurácia média máxima

Pode ser observado que com o uso de vetores híbridos formados pela combinação do expoente de Hurst ( $H$ ) com 20 coeficientes LPC a acurácia média atingiu um valor de  $99,72 \pm 0,40\%$ . Este resultado é cerca de cinco pontos percentuais acima do resultado obtido apenas com o uso de  $H$  ( $94,29 \pm 7,38\%$ ) e está acima, inclusive, do melhor resultado encontrado utilizando combinações dos conjuntos MNL e MQR ( $98,75\% \pm 3,95\%$ ).

Na Figura 5.27 estão representados os valores da acurácia média determinados a partir da avaliação individual da característica  $H$ , da análise LPC e da combinação dessas características na classificação entre sinais de vozes saudáveis e de vozes com nódulos.



**Figura 5.27** Acurácia Média - Saudável x Nódulo

Ainda de acordo com os resultados mostrados na Tabela 5.16, as combinações das características  $H$ ,  $DET$  e  $LAM$  com  $FMMI$  e  $ENTR$  ou com  $H_1$  e  $H_q$  foram as que apresentaram os melhores valores para a acurácia média na classificação entre sinais de vozes saudáveis e vozes afetadas por nódulos.

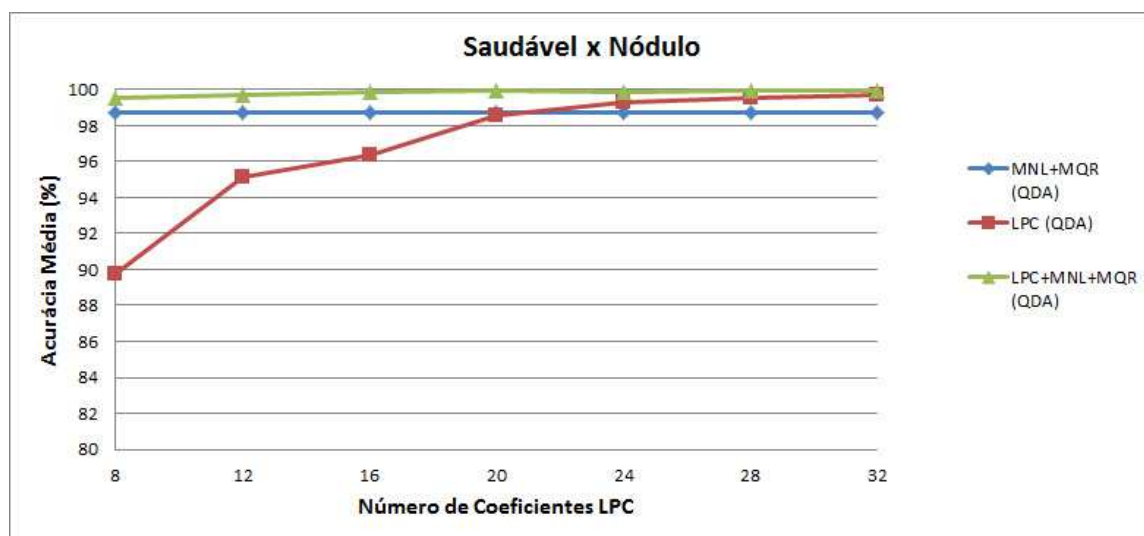
Na Tabela 5.23 estão apresentados os resultados das medidas de desempenho referentes à combinação do conjunto de características  $H$ ,  $DET$ ,  $LAM$ ,  $FMMI$  e  $ENTR$  com os vetores de coeficientes LPC.

**Tabela 5.23** Taxas de classificação (%) Saudável x Nódulo - LPC+MNL+MQR (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	$99,53 \pm 0,38$	$99,72 \pm 0,45$	$99,86 \pm 0,23$	$99,95^* \pm 0,15$	$99,86 \pm 0,23$	$99,95^* \pm 0,15$	$99,91 \pm 0,20$
Sensibilidade	$99,43 \pm 0,55$	100	100	100	100	100	100
Especificidade	$99,81 \pm 0,59$	$98,89 \pm 1,79$	$99,44 \pm 0,89$	$99,81 \pm 0,59$	$99,44 \pm 0,89$	$99,81 \pm 0,59$	$99,63 \pm 0,78$

\*acurácia média máxima

Na Figura 5.28 estão apresentados os valores médios da acurácia obtidos na combinação das características  $H$ ,  $DET$ ,  $LAM$ ,  $FMMI$  e  $ENTR$ , bem como os obtidos pela análise LPC e ainda pela combinação dos coeficientes LPC e esse conjunto de características. Como pode ser observado a combinação das cinco características com o acréscimo de 12 coeficientes LPC



**Figura 5.28** Acurácia Média - Saudável x Nódulo

apresentou resultado idêntico ao encontrado com o uso combinado de  $H$  com 20 coeficientes LPC (Tabela 5.22). A acurácia média atingiu o valor de  $99,95 \pm 0,15\%$  com o uso combinado das características  $H$ ,  $DET$ ,  $LAM$ ,  $FMMI$  e  $ENTR$  e 20 coeficientes LPC.

### 5.11.5 Classificação Paralisia x Edema

De acordo com os resultados apresentados na Tabela 5.16, o primeiro mínimo da função de informação mútua média  $FMMI$  é a característica que apresentou o maior valor de acurácia média na classificação entre sinais de vozes com paralisia e sinais de vozes com edema.

Na Tabela 5.24 estão apresentados os resultados das medidas de desempenho referentes à combinação da característica  $FMMI$  com os vetores de coeficientes LPC na classificação entre sinais de vozes com paralisia e sinais de vozes com edema, por meio da função discriminante quadrática (QDA).

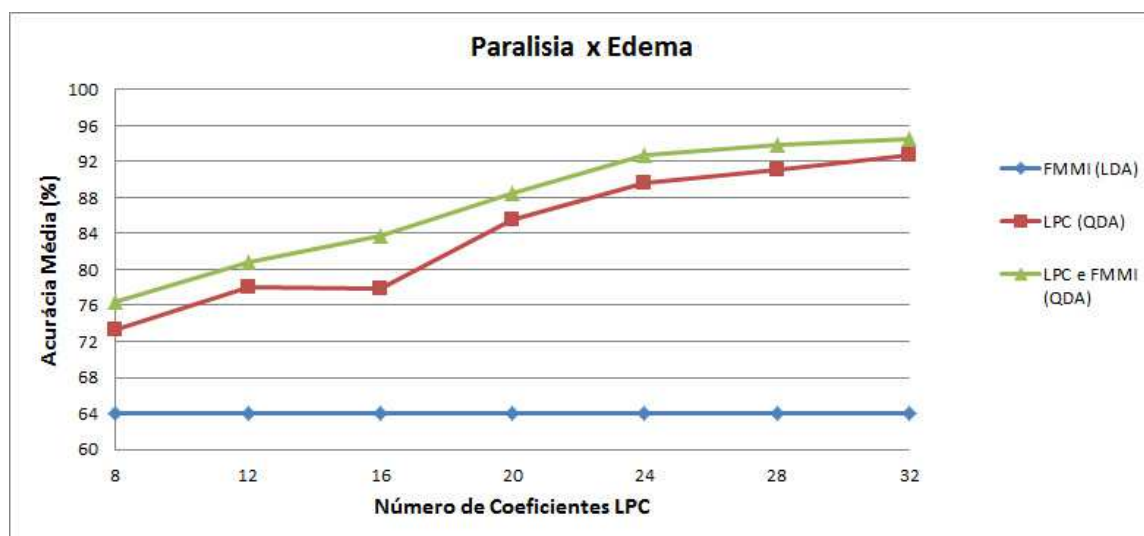
**Tabela 5.24** Taxas de classificação (%) Paralisia x Edema - LPC e  $FMMI$  (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	76,43 $\pm$ 2,56	80,84 $\pm$ 2,32	83,74 $\pm$ 2,80	88,41 $\pm$ 1,67	92,75 $\pm$ 1,54	93,86 $\pm$ 1,78	94,52* $\pm$ 1,30
Sensibilidade	69,19 $\pm$ 4,83	75,99 $\pm$ 4,03	78,42 $\pm$ 3,65	86,02 $\pm$ 2,05	94,29 $\pm$ 1,55	92,34 $\pm$ 2,00	93,37 $\pm$ 2,08
Especificidade	85,26 $\pm$ 3,17	86,74 $\pm$ 2,22	90,22 $\pm$ 3,74	91,33 $\pm$ 2,59	90,89 $\pm$ 3,53	95,70 $\pm$ 1,77	95,93 $\pm$ 2,16

\*acurácia média máxima

Na Figura 5.29 estão representados os valores da acurácia média determinados a partir da avaliação individual da característica  $H$ , da análise LPC e da combinação dessas características na classificação entre sinais de vozes com paralisia e de vozes com edema.

Como pode ser observado, o acréscimo dos coeficientes LPC oferece um alto ganho em relação ao resultado da classificação quando se usa apenas a característica  $H$ . Entretanto, o valor



**Figura 5.29** Acurácia Média - Paralisia x Edema

máximo observado para acurácia máxima fica abaixo de 95%, ou seja, em torno do mesmo valor obtido usando apenas a análise LPC com 32 coeficientes.

De acordo com os resultados mostrados na Tabela 5.16, a combinação de características  $\lambda_1$ ,  $ApEn$ ,  $D_2$ ,  $K_2$ ,  $DET$ ,  $ENTR$ ,  $Trend$  e  $TT$  foi a que apresentou o maior valor para a acurácia média na classificação entre sinais de vozes afetadas por paralisia e vozes afetadas por edema.

Na Tabela 5.25 estão apresentados os resultados das medidas de desempenho referentes à combinação conjunto de características  $\lambda_1$ ,  $ApEn$ ,  $D_2$ ,  $K_2$ ,  $DET$ ,  $ENTR$ ,  $Trend$  e  $TT$  com os vetores de coeficientes LPC, por meio da função discriminante quadrática (QDA).

**Tabela 5.25** Taxas de classificação (%) Paralisia x Edema - LPC+MNL+MQR (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	96,32±1,29	97,95±1,33	99,42±0,73	99,88*±0,25	99,82±0,39	99,77±0,56	99,47±0,43
Sensibilidade	95,13±2,48	97,31±2,52	99,62±0,62	99,74±0,54	99,74±0,81	99,62±0,87	99,36±0,91
Especificidade	97,31±2,04	98,49±1,26	99,25±1,14	100	99,89±0,34	99,89±0,34	99,57±0,56

\*acurácia média máxima

Na Figura 5.30 estão apresentados os valores médios da acurácia obtidos na combinação das características das MNL e MQR, bem como os obtidos pela análise LPC e ainda pela combinação dos coeficientes LPC e esse conjunto de características.

Como pode ser observado, o uso de vetores híbridos formados pela combinação das características  $\lambda_1$ ,  $ApEn$ ,  $D_2$ ,  $K_2$ ,  $DET$ ,  $ENTR$ ,  $Trend$  e  $TT$  com oito coeficientes LPC apresenta ganho em torno de 12 pontos percentuais na taxa de acurácia média, em relação ao resultado obtido usando apenas as medidas dos conjuntos MNL e MQR. A acurácia média máxima atinge 99,88±0,25% quando a combinação é feita com 20 coeficientes LPC.

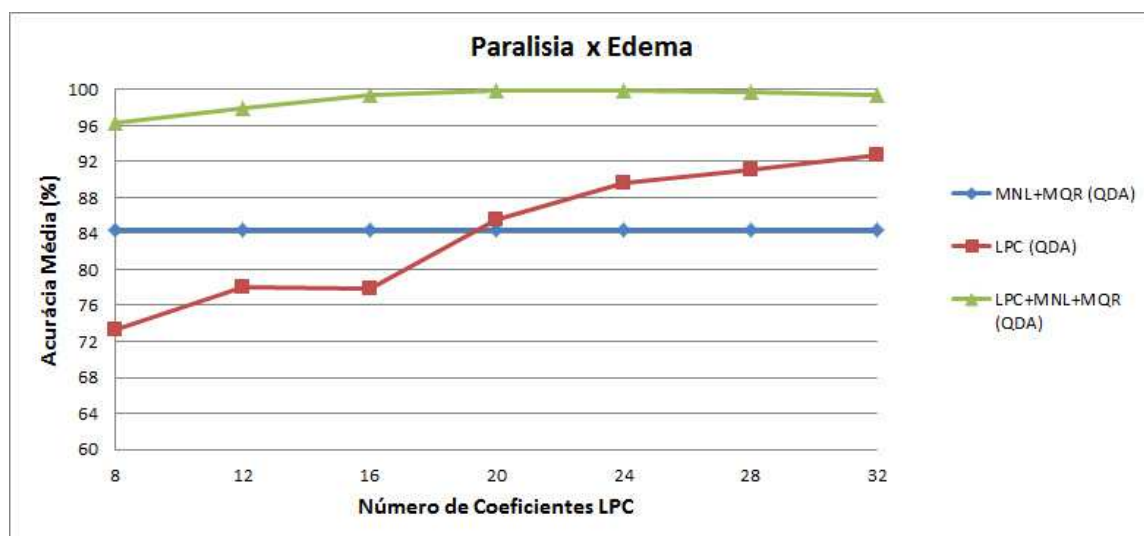


Figura 5.30 Acurácia Média - Paralisia x Edema

### 5.11.6 Classificação Paralisia x Nódulo

De acordo com os resultados apresentados na Tabela 5.16, o primeiro mínimo da função de informação mútua média *FMMI* é a característica que apresentou o maior valor de acurácia média na classificação entre sinais de vozes com paralisia e sinais de vozes com nódulos.

Na Tabela 5.26 estão apresentados os resultados das medidas de desempenho referentes à combinação da característica *FMMI* com os vetores de coeficientes LPC na classificação entre esses dois grupos de sinais por meio da função discriminante quadrática (QDA).

Tabela 5.26 Taxas de classificação (%) Paralisia x Nódulo - LPC e  $D_2$  (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	85,26±3,04	91,72±2,35	94,65±1,53	96,98±1,38	98,63±0,86	98,90*±0,54	98,08±0,71
Sensibilidade	81,88±3,90	90,15±2,95	93,86±2,12	97,02±1,42	98,97±0,81	99,21±0,41	98,36±1,00
Especificidade	95,56±3,05	96,48±2,95	97,04±1,79	96,85±2,48	97,59±2,15	97,96±1,62	97,22±2,66

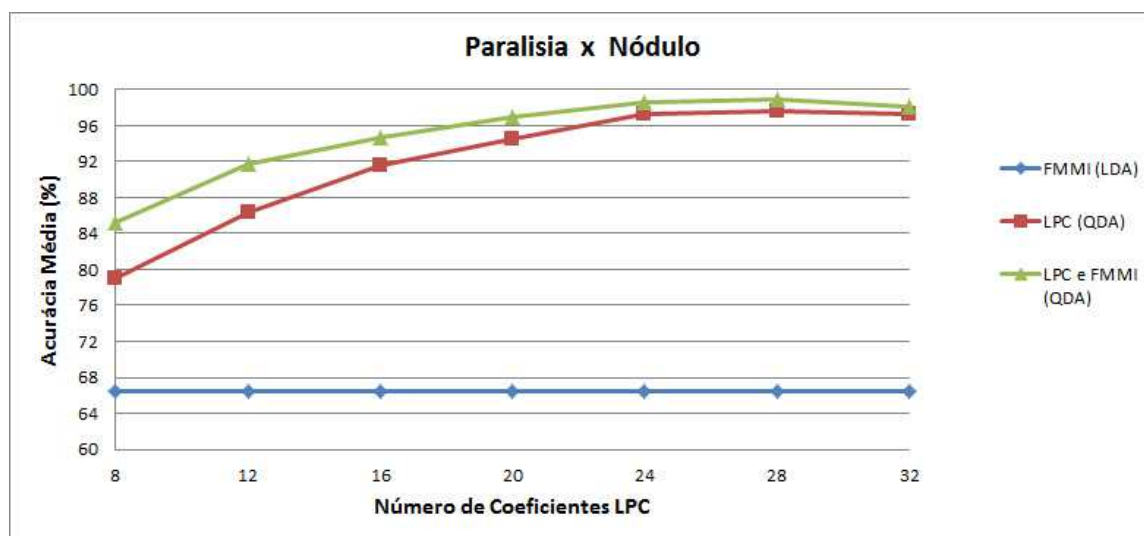
\*acurácia média máxima

Na Figura 5.31 estão representados os valores da acurácia média determinados a partir da avaliação individual da característica *FMMI*, da análise LPC e da combinação dessas características na classificação entre sinais de vozes com paralisia e de vozes com nódulos.

Como pode ser notado, o uso de vetores híbridos oferece alto ganho no processo de classificação quando é comparado aos resultados obtidos com o uso individual da característica *FMMI*. Por outro lado, pode-se observar a partir da Figura 5.31 que o ganho da combinação diminui consideravelmente, em relação ao oferecido pela análise LPC, com o aumento do número de coeficientes.

Ainda de acordo com os resultados mostrados na Tabela 5.16, a combinação de características  $\lambda_1$ ,  $H$ ,  $ApEn$ ,  $K_2$ ,  $ENTR$  e  $TT$  foi a que apresentou o maior valor para a acurácia média





**Figura 5.31** Acurácia Média - Paralisia x Nódulo

na classificação entre sinais de vozes afetadas por paralisia e vozes afetadas por nódulos nas pregas vocais.

Na Tabela 5.27 estão apresentados os resultados das medidas de desempenho referentes à combinação das características  $\lambda_1$ ,  $H$ ,  $ApEn$ ,  $K_2$ ,  $ENTR$  e  $TT$  com os vetores de coeficientes LPC, por meio da função discriminante quadrática (QDA).

**Tabela 5.27** Taxas de classificação (%) Paralisia x Nódulo - LPC+MNL+MQR (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	99,11±0,81	99,27±1,05	99,51±0,57	99,76±0,39	99,76±0,39	99,84±0,51	99,92*±0,26
Sensibilidade	99,87±0,41	99,74±0,54	100	100	100	100	100
Especificidade	97,78±1,81	98,44±2,11	98,67±1,55	99,33±1,07	99,33±1,07	99,56±1,41	99,78±0,70

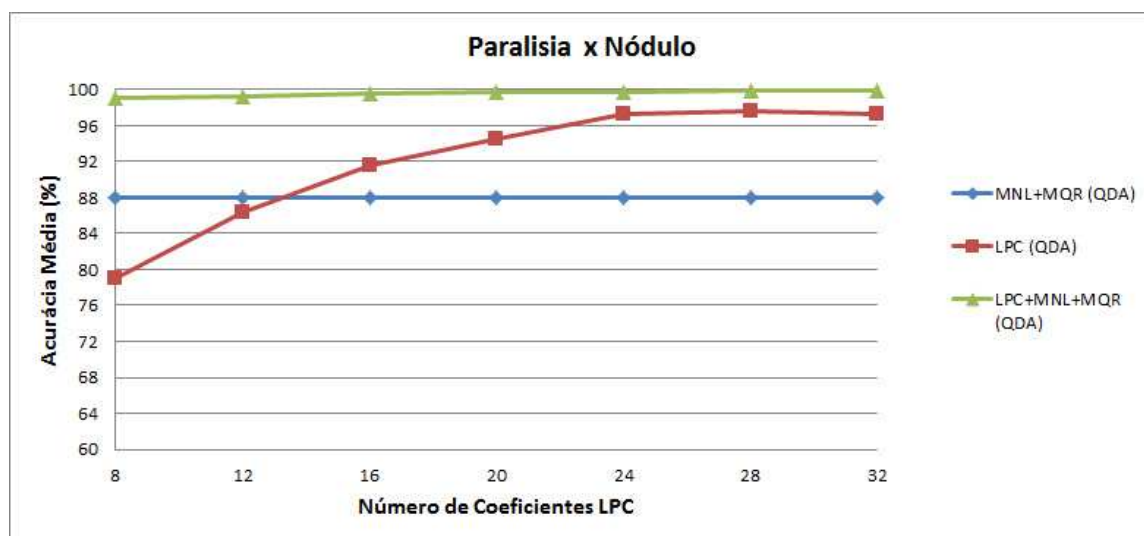
\*acurácia média máxima

Na Figura 5.32 estão apresentados os valores médios da acurácia obtidos na combinação das características das MNL e MQR, bem como os obtidos pela análise LPC e ainda pela combinação dos coeficientes LPC e esse conjunto de características.

Como pode ser observado, a acurácia média já atinge valores acima de 99% com o uso de vetores híbridos formados pelo conjunto  $\lambda_1$ ,  $H$ ,  $ApEn$ ,  $K_2$ ,  $ENTR$  e  $TT$  e oito coeficientes LPC. O maior valor obtido para a acurácia média foi de 99,92±0,26% quando a combinação das características dos conjuntos MNL e MQR é feita com 32 coeficientes LPC.

### 5.11.7 Classificação Edema x Nódulo

De acordo com os resultados apresentados na Tabela 5.16, a dimensão de correlação ( $D_2$ ) é a característica que apresentou o maior valor de acurácia média na classificação entre sinais de vozes com paralisia e sinais de vozes com nódulos.



**Figura 5.32** Acurácia Média - Paralisia x Nódulo

Na Tabela 5.28 estão apresentados os resultados das medidas de desempenho referentes à combinação de  $D_2$  com os vetores de coeficientes LPC na classificação entre sinais de vozes com paralisia e sinais de vozes com nódulos, por meio da função discriminante quadrática (QDA).

**Tabela 5.28** Taxas de classificação (%) Edema x Nódulo - LPC e  $D_2$  (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	79,78±4,33	88,70±2,74	92,32±2,00	97,46±1,33	97,32±1,08	98,70*±1,22	85,65±2,60
Sensibilidade	75,38±6,01	87,74±4,09	92,15±2,49	98,17±1,44	97,74±0,79	99,35±0,91	90,54±3,61
Especificidade	88,89±4,80	90,67±5,42	92,67±3,64	96,00±2,04	96,44±4,08	97,33±2,73	75,56±7,91

\*acurácia média máxima

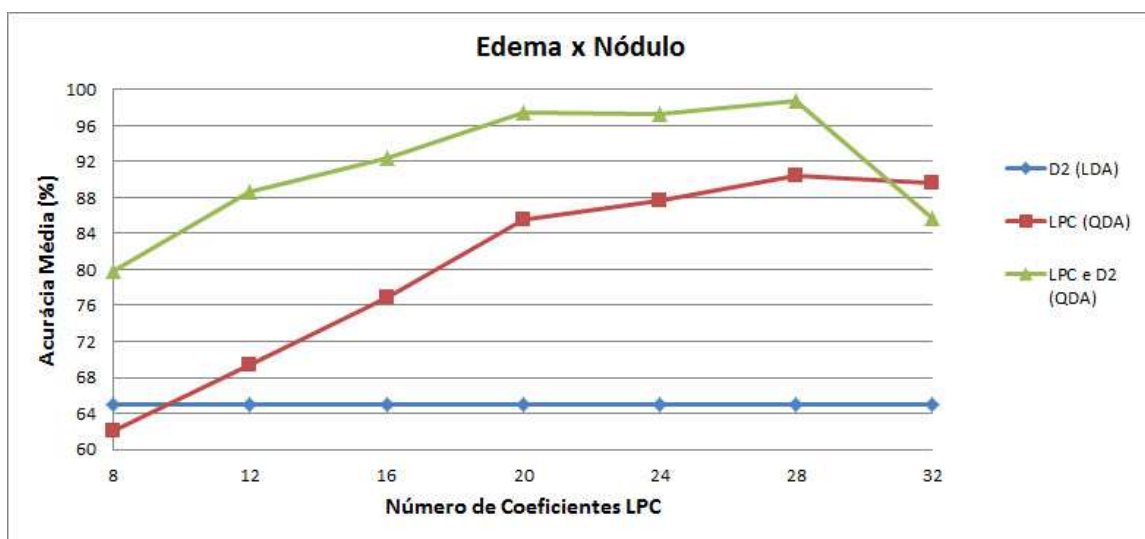
Na Figura 5.33 estão representados os valores da acurácia média determinados a partir da avaliação individual da característica  $H$ , da análise LPC e da combinação dessas características na classificação entre sinais de vozes com paralisia e de vozes com nódulos.

Como pode ser observado, a combinação de  $H$  com a os resultados da análise LPC apresentou alto ganho na taxa de acurácia do classificador. O maior valor para a acurácia média foi de 98,70±1,22 para a combinação com 28 coeficientes LPC.

De acordo com os resultados mostrados na Tabela 5.16, a combinação de características  $\lambda_1$ ,  $H_1$ ,  $H_q$ ,  $ApEn$ ,  $D_2$ ,  $L_{max}$  e  $ENTR$  foi a que apresentou o maior valor para a acurácia média na classificação entre sinais de vozes afetadas por edema e vozes afetadas por nódulos nas pregas vocais.

Na Tabela 5.29 estão apresentados os resultados das medidas de desempenho referentes à combinação do conjunto de características  $\lambda_1$ ,  $H_1$ ,  $H_q$ ,  $ApEn$ ,  $D_2$ ,  $L_{max}$  e  $ENTR$  com os vetores de coeficientes LPC, por meio da função discriminante quadrática (QDA).

Na Figura 5.34 estão apresentados os valores médios da acurácia obtidos na combinação das características das MNL e MQR, bem como os obtidos pela análise LPC e ainda pela



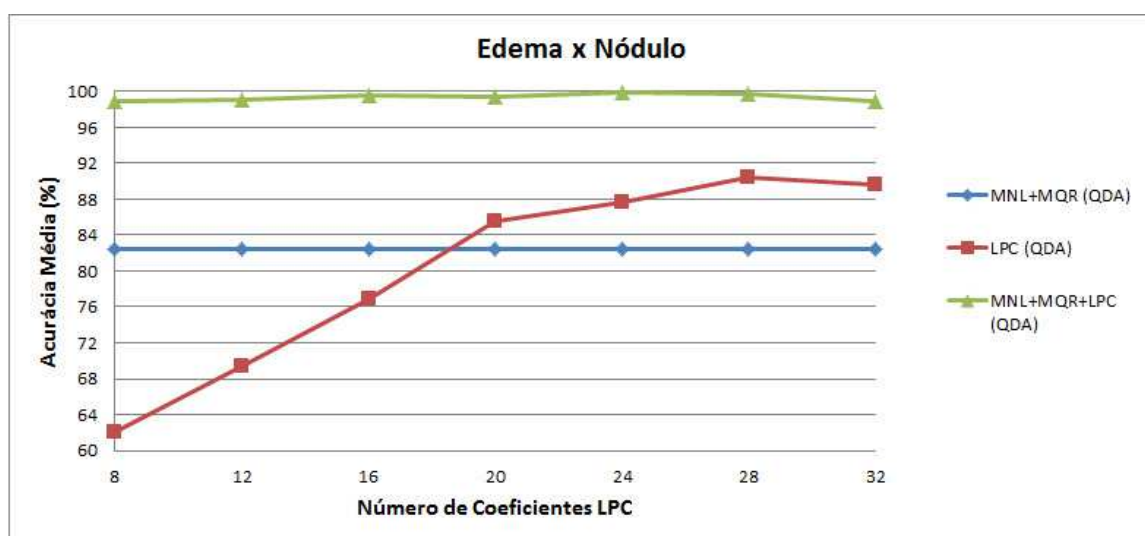
**Figura 5.33** Acurácia Média - Edema x Nódulo

**Tabela 5.29** Taxas de classificação (%) Edema x Nódulo - LPC+MNL+MQR (QDA)

Medidas	Número de coeficientes LPC						
	8	12	16	20	24	28	32
Acurácia	98,91±0,92	99,13±0,75	99,57±0,51	99,42±0,75	99,86*±0,31	99,78±0,35	98,99±0,85
Sensibilidade	99,03±0,94	99,89±0,34	100	100	100	100	98,82±1,18
Especificidade	98,67±1,87	97,56±2,45	98,67±1,55	98,22±2,30	99,56±0,94	99,33±1,07	99,33±1,07

\*acurácia média máxima

combinação dos coeficientes LPC e esse conjunto de características. Como pode ser observado, no caso da combinação das medidas MNL e MQR com 8 coeficientes LPC, a acurácia média do classificador atingiu  $98,91 \pm 0,92\%$ . Este resultado está cerca de 16 pontos percentuais acima do valor máximo obtido sem a inclusão dos coeficientes LPC. O valor máximo para a acurácia média foi de  $99,86 \pm 0,31\%$  com o uso de 24 coeficientes LPC.



**Figura 5.34** Acurácia Média - Edema x Nódulo

## 5.12 Discussão

A análise dos resultados encontrados na classificação realizada com as características de forma individual mostra que as medidas de quantificação de recorrência (MQR) apresentaram resultados relativamente superiores aos obtidos com a utilização das medidas da análise não linear (MNL) para o caso da classificação entre sinais de vozes saudáveis e vozes com paralisia nas pregas vocais. Em contrapartida, a situação é oposta nos casos de discriminação entre os sinais de vozes saudáveis e afetadas por edema, ou na classificação entre os sinais com as patologias paralisia e nódulos, ou ainda, no caso edema *versus* nódulos. Os desempenhos das características de ambos os grupos (MNL e MQR) podem ser considerados equivalentes nos casos da discriminação entre os grupos de vozes saudáveis e afetados por nódulos ou no caso da classificação paralisia *versus* edema.

Ainda em relação ao desempenho de cada característica individual na tarefa de discriminar dois a dois os diversos grupos de sinais de vozes avaliados, é possível notar a predominância de três medidas: expoente de Hurst, dimensão de correlação e do primeiro mínimo da função de informação mútua, para o caso das medidas da análise não linear (MNL). Por outro lado, em relação as medidas de quantificação de recorrência (MQR), a característica referente à medida do comprimento máximo das linhas diagonais é responsável pelo melhor desempenho na maioria dos testes realizados.

Como já esperado, a combinação adequada de características apresentou resultados superiores na taxa de sucesso do sistema de classificação, em relação ao uso de qualquer característica avaliada de forma individual. Entretanto, o aumento do ganho revelado nas medidas de desempenho não é diretamente proporcional ao número de características envolvidas na combinação. Por exemplo, em nenhum dos casos avaliados a combinação de todas as características resultou na máxima taxa de classificação encontrada. Daí, a importância de se buscar a combinação adequada entre as características a fim de se obter o melhor desempenho do classificador.

A partir dos resultados apurados, é possível observar que a combinação, entre elas, das medidas de análise não linear (MNL) avaliadas apresenta resultados superiores, em termos de acurácia média, aos obtidos com a combinação das medidas de quantificação de recorrência (MQR). Por outro lado, a combinação das características de ambos os conjuntos (MNL e MQR) oferece ganho significativo no desempenho do sistema de classificação para a maioria dos casos analisados.

Em relação à utilização de vetores híbridos, formados pela combinação de medidas dos conjuntos MNL e MQR e por coeficientes extraídos da análise LPC, os resultados mostram que houve ganho na maioria dos casos analisados. Principalmente, nos casos de discriminação entre sinais de vozes afetados por diferentes patologias.

## CAPÍTULO 6

---

# Trabalhos Correlatos e Estado da Arte

---

Neste capítulo, são apresentados os resultados de vários trabalhos encontrados na literatura que utilizam a análise não linear, seja na caracterização dos sons da fala, seja na discriminação de vozes patológicas.

### 6.1 Caracterização dos Sons da Fala

Martínez et al. [120] fazem uso do maior expoente de Lyapunov como estimativa da quantidade de caos presente no sistema de produção da fala para uma variedade de sons sonoros da Língua Espanhola. Os resultados obtidos mostram a presença de um expoente de Lyapunov máximo positivo em todos os sinais considerados, confirmando, segundo os autores, a natureza caótica desses sinais de voz.

Em seu trabalho de Tese, Petry [28] utiliza medidas de invariantes dinâmicas não lineares para desenvolver um sistema de reconhecimento automático de locutor. As invariantes dinâmicas consideradas são: dimensão fractal, dimensão de correlação e maior expoente de Lyapunov. O desempenho do sistema proposto, incluindo as medidas das invariantes não lineares, é comparado com o desempenho apresentado por um sistema de referência (sistema-base) composto por vetores formados com 16 coeficientes mel-cepstrais. O sistema proposto apresenta uma redução de 17,65% na taxa de erro igual (EER) em relação ao sistema de referência.

Pitsikalis et al. [121] investigaram o uso da análise dinâmica não linear em sinais de voz para a classificação de fonemas isolados, por meio da extração das medidas invariantes: dimensões fractais generalizadas e expoentes de Lyapunov, obtendo uma taxa de sucesso de 78%. Por outro lado, como medida de comparação, foi realizado outro processo de classificação com um vetor por fonema, composto por 12 coeficientes cepstrais, obtendo-se 67% de taxa de sucesso na classificação dos fonemas.

Lindgren et al. [122] e Johnson et al. [123] realizaram experimentos de reconhecimento de fonemas isolados, distribuídos em 39 classes, usando características extraídas do espaço de fase reconstruído (expoentes de Lyapunov e dimensão de correlação). Os resultados obtidos

ficaram entorno de 75% da acurácia obtida com o uso de um sistema de referência baseado em coeficientes mel-cepstrais. Fazendo-se uso de uma combinação de ambos os métodos, a acurácia do processo de classificação aumentou cerca de 5%.

## 6.2 Análise Não Linear de Vozes Saudáveis e Patológicas

Alguns dos trabalhos referenciados a seguir são baseados em resultados de testes estatísticos, ou seja, apenas sugerem o potencial discriminativo das medidas de dinâmica não linear sem, contudo, levar a efeito um processo efetivo de classificação entre os grupos analisados.

Titze et al. [61] encontraram que a dimensão de correlação de um sujeito disfônico com polipose é mais alta que a de um sujeito saudável.

Kakita e Okamoto [124] encontraram que o valor médio da dimensão de correlação de dezessete sujeitos disfônicos foi superior ao de oito sujeitos saudáveis, o que sugere que vozes perceptualmente roucas tem dinâmica mais complexa que vozes saudáveis.

Hertrich et al. [125] encontraram que os sinais eletrográficos da glote de pessoas com doença de Parkinson e doenças cerebrais têm significativamente maiores dimensões fractais que os sinais eletrográficos da glote de sujeitos saudáveis.

Giovanni et al. [126] encontraram diferenças estatisticamente significativas ( $P = 0,031$ ) no valor do maior expoente de Lyapunov entre um grupo formado por 12 sinais de vozes saudáveis e um grupo de 26 pacientes com paralisia unilateral das pregas vocais.

Jiang e Zhang [76] empregaram métodos de dinâmica não linear para descrever a complexidade da voz entres locutores com vozes saudáveis e pessoas apresentando pólipos vocais como patologia da voz. Os resultados demonstram como característica uma dinâmica de baixa dimensão, tanto no caso da voz saudável como na voz patológica ( $D_2 < 3$ ). Além disso, a análise da dimensão de correlação indica diferenças estatísticas significativas entre os dois grupos. A média e o desvio padrão de  $D_2$  das amostras dos sinais de vozes saudáveis são mais baixas que das amostras dos sinais de vozes patológicas.

Em outro trabalho, Zhang e Jiang [127] utilizaram um total de 122 sinais de voz, referentes à vogal sustentada /ah/, derivados de vozes patológicas de pacientes com doença de Parkinson, nódulos e pólipos vocais, paralisia da laringe e com câncer de origem epitelial. Os sinais foram separados em três grupos segundo a classificação sugerida por Titze [8]. Os resultados baseados em testes estatísticos sugerem que a dimensão de correlação potencialmente discrimina os três tipos de sinais.

Torres et al. [82] utilizaram diferentes medidas de complexidade a fim de caracterizar os sinais de vozes patológicas. Foi realizado um estudo comparativo da complexidade destes sinais com a de sinais de vozes saudáveis. As medidas utilizadas foram: a medida de complexidade de Ziv-Lempel, a entropia de Shannon, a entropia de Tsallis, a entropia relativa de Kullback e suas derivadas (a entropia generalizada e o divergente). Os experimentos foram realizados considerando a estratificação das vozes em relação ao gênero. Os resultados de testes estatísticos

mostram que as medidas utilizadas não oferecem o mesmo potencial discriminativo para ambos os sexos. A entropia de Shannon e a medida de complexidade de Ziv-Lempel não apresentaram diferenças estatisticamente significativas em nenhum dos casos. A entropia de Tsallis e as entropias derivadas da entropia relativa apresentaram diferenças estatisticamente para ambos os grupos ( $P < 0,001$ ). A entropia relativa apresentou diferenças significativas apenas no caso do grupo de vozes masculinas.

Jiang et al [6] realizaram um estudo comparativo entre a análise dinâmica não linear (dimensão de correlação) e análise de perturbação (*jitter* e *shimmer*) em termos dos efeitos do comprimento do sinal analisado, taxa de amostragem e nível de ruído. Nos experimentos foram usadas vogais sustentadas de vozes saudáveis e de vozes afetadas por patologias na laringe. Segundo os autores, as grandes variações nos valores das medidas de perturbação para os casos dos sinais aperiódicos e caóticos tornam inapropriado o uso dessas medidas para análise de tais sinais. Em contrapartida, a análise dinâmica não linear pode quantificar sinais caóticos. Além disso, a dimensão de correlação oferece uma medida de análise mais confiável para sinais quase periódicos de menor comprimento, taxa de amostragem mais baixa e maiores níveis de ruído.

Em outro estudo, Aghazadeh et al. [128] investigaram o uso de duas medidas não lineares para quantificar a complexidade da voz e na discriminação entre vozes saudáveis e patológicas: a entropia aproximada e um parâmetro de escala proporcional ao expoente de Hurst. Segundo os autores, as medidas avaliadas mostram ser ferramentas efetivas de classificação.

El-Imam e Elwakil [129] utilizam a análise de quantificação de recorrência (% determinismo) e a entropia espaço-temporal para analisar sinais de voz e caracterizar sons vocálicos sonoros e nasais (precedendo /m/ e /n/). Segundo os resultados obtidos, as medidas avaliadas potencialmente discriminam os dois tipos de sinais.

Little et al. [45] utilizam uma medida baseada na ideia de recorrência (RPDE - entropia da densidade de probabilidade de recorrência) combinada com um fator de escala fractal (auto-similaridade) para classificar vozes saudáveis *versus* vozes patológicas. Os resultados obtidos, utilizando análise discriminante quadrática, apontam para uma acurácia de  $91,8\% \pm 2\%$ . Como parâmetro de comparação, os autores realizaram diferentes combinações (duas a duas) envolvendo algumas medidas de perturbação tradicionais, obtendo nestes casos taxas de classificação inferiores às encontradas com o método proposto (*jitter* e *shimmer*,  $81,4\% \pm 3,9\%$ ; *shimmer* e HNR (Relação harmônico-ruído),  $80,7\% \pm 4\%$  e *jitter* e HNR,  $76,4\% \pm 4,8\%$ ).

Zhang e Jiang [130] investigam as características de vogais faladas de forma contínua e de forma sustentada de locutores com vozes saudáveis e pacientes com patologias da laringe. Foram usados métodos de perturbação (incluindo *jitter* e *shimmer*), relação sinal-ruído (SNR), e métodos de dinâmica não linear (dimensão de correlação e entropia de segunda ordem). Os resultados obtidos mostraram que em todos os casos estudados, ou seja, fala contínua, sustentada, saudável e/ou patológica, os sinais de voz têm características dinâmicas de baixa dimensão. Além disso, para as vogais sustentadas, a análise de *jitter*, *shimmer*, dimensão de correlação, e entropia de segunda ordem revelaram diferenças significativas entre vozes saudáveis e patoló-

gicas. Por outro lado, para voz contínua, *jitter* e *shimmer* não discriminaram, estatisticamente, vozes saudáveis de patológicas, porém uma diferença significativa foi encontrada para a SNR (Relação sinal-ruído), dimensão de correlação, e entropia de segunda-ordem.

Em outro trabalho, Falcão et al. [111] apresentam um estudo sobre a aplicação da entropia de Shannon, entropia relativa e a entropia de Tsallis para a classificação automática de voz saudável ou patológica, com edema, cistos, nódulos nas cordas vocais e paralisia dos nervos laríngeos. Segundo os resultados encontrados a entropia de Tsallis mostrou-se mais robusta na separação de vozes saudáveis e patológicas sendo, no geral, o modelo mais eficiente, pois apresenta uma maior probabilidade de correta detecção da patologia. Porém, a entropia relativa se revelou mais precisa na correta rejeição da presença de patologias, ou seja, apresenta um melhor diagnóstico de vozes saudáveis, com o menor índice de falsos positivos. Os resultados para a entropia de Shannon também foram significativos, principalmente no correto diagnóstico da presença de patologias. Segundo a avaliação dos autores, os resultados obtidos mostram que o uso da entropia para diagnosticar a presença de patologias nas dobras vocais é bastante interessante e promissor. As medidas de entropia utilizadas não se mostraram eficazes na separação dos diferentes tipos de patologia abordados.

Vaziri et al. [131] fazem uso do expoente de Hurst para efeito de classificação entre vozes saudáveis e vozes patológicas. Os resultados encontrados, utilizando análise discriminante linear (LDA), apresentam diferentes taxas de acurácia dependendo do método utilizado para estimação do parâmetro de Hurst, ou seja, 95% com o uso do método da variância agregada e 95,24% para o método R/S.

Henríquez et al. [32] investigam a possibilidade de se quantificar a qualidade da voz (saudável ou patológica) por meio do uso de seis medidas não lineares, são elas: as entropias de Rényi de primeira e segunda ordem, a entropia de correlação (ou entropia de segunda ordem), a dimensão de correlação, a entropia de Shannon e o valor do primeiro mínimo da função de informação mútua. A fim de se verificar a utilidade dessas medidas são usados dois bancos de dados: um banco de dados não comercial, composto de quatro diferentes tipos de qualidade da voz (voz saudável e três níveis de vozes patológicas - alta, média e baixa severidade) de acordo com o parâmetro do grau de rouquidão (*Grade of hoarseness*), o (G) da escala GRBAS; e um banco de dados comercial, "MEEI Voice Disorders", composto de dois níveis de qualidade vocal (vozes saudáveis e vozes patológicas). Para um classificador baseado em redes neurais MLP (*Multi-Layer Perceptron*) e usando todas as características combinadas, a taxa global de sucesso de um sistema de classificação foi de 82,47% (banco de dados não-comercial) e a do outro 99,69% (banco de dados MEEI). Segundo os autores, a diferença entre as taxas de classificação obtidas a partir dos dois bancos de dados deve-se ao fato da não existência, no banco de dados MEEI, de vozes patológicas de baixa severidade.

Em outro trabalho, Vaziri et al. [30] faz uso da medidas de dinâmica não linear: dimensão de correlação, maior expoente de Lyapunov, entropia aproximada, dimensão fractal, medida de complexidade de Ziv-Lempel, bem como duas medidas clássicas de perturbação: o



quociente de perturbação de amplitude (APQ - *Amplitude Perturbation Quotient*) e o quociente de perturbação de *pitch* (PPQ - *Pitch Perturbation Quotient*), na tarefa de diagnóstico da qualidade da voz. Nos experimentos foram analisados 51 sinais de vozes saudáveis e 112 sinais de vozes afetadas por patologias nas pregas vocais. Utilizando um classificador baseado em SVM (*Support Vector Machine*) os resultados encontrados revelam as seguintes taxas de classificação: dimensão de correlação, 94,44% ; maior expoente de Lyapunov, 88,89%; Ziv-Lempel, 88,80%; dimensão fractal, 87,85%; entropia aproximada, 79,78%, PPQ, 73,33% e APQ, 71,12%. Segundo os autores, isto significa que a dimensão de correlação e o maior expoente de Lyapunov podem ser usados como um método confiável no diagnóstico clínico de patologias na laringe e em outras aplicações na área.

Em um trabalho mais recente, Chai et al. [29] observaram que a dimensão de correlação de sinais de voz de pessoas fumantes apresenta valores superiores aos valores dessa medida com vozes de pessoas não-fumantes ( $P < 0,001$ ).

Arias-Londono et al. [44] utilizam a combinação de dois classificadores para realizar a discriminação entre vozes saudáveis e vozes patológicas. Em um dos classificadores é usado um conjunto de onze características derivadas da análise não linear. Duas dessas características são baseadas nas medidas da teoria do caos (o maior expoente de Lyapunov e a dimensão de correlação), duas são baseadas na análise de recorrência (RPDE) e de auto-similaridade, e as demais são diferentes estimativas da entropia. O outro classificador utiliza as informações extraídas de métodos clássicos de parametrização (parâmetros de ruído e coeficientes mel-cepstrais). A combinação dos dois classificadores e a utilização do classificador SVM gerou uma taxa de classificação global (acurácia) de  $98,23\% \pm 0,01\%$ .

A seguir são apresentados alguns trabalhos, dentre vários existentes, que estão no mesmo contexto desta pesquisa, porém que utilizam técnicas baseadas em medidas de perturbação ou no modelo linear de produção da fala.

Godino *et al.* [21] sugerem que a detecção de desordens vocais pode ser realizada usando vetores formados pela combinação de coeficientes cepstrais na escala *mel* (*MFCC - mel frequency cepstral coefficients*), mais a sua primeira derivada (delta mel-cepstrais) e a energia de cada segmento do sinal analisado. Nos experimentos são utilizados 53 sinais de vozes saudáveis e 173 sinais de voz com uma larga faixa de anomalias nas pregas vocais. A classificação é feita com o uso de modelos de misturas gaussianas (*GMM - Gaussian mixture models*) que fornecem uma decisão sobre a presença ou ausência da patologia para cada segmento do sinal. O melhor resultado de classificação foi obtido com o uso de 24 MFCC e um modelo de 6 misturas gaussianas, com uma taxa máxima de acurácia de  $94,07\% \pm 3,28\%$ .

Em seu trabalho de Tese, Costa [19] apresenta um estudo sobre a discriminação da patologia edema na pregas vocais. A patologia é caracterizada utilizando-se análise por predição linear (LPC), análise cepstral e mel-cepstral. Para a estimação dos coeficientes cepstrais, é utilizada uma abordagem paramétrica derivada da análise por predição linear e para os mel-cepstrais uma abordagem não paramétrica, baseada na transformada rápida de Fourier. Para cada tipo de

parâmetro, foram extraídos 12 coeficientes. A classificação é realizada em duas etapas: uma pré-classificação, em que é utilizada a medida de distância do mínimo erro médio quadrático para o cálculo dos valores de distorção, a partir daí a voz é classificada como patológica ou não-patológica, de acordo com um limiar pré-estabelecido, que proporcione a melhor separação entre as classes; na fase final utiliza-se uma modelagem paramétrica por meio de Modelos de Markov Escondidos (*HMM-Hidden Markov Models*) [132]. Nessa pesquisa, foram utilizados 53 arquivos de vozes normais, disponíveis na base de dados da Kay Elemetrics [60], 44 sinais de vozes afetados por edema nas pregas vocais e 23 sinais com outras patologias da laringe como nódulos, cistos, e paralisia nas pregas vocais. Os resultados indicam os coeficientes provenientes da análise LPC como sendo os que obtiveram melhores valores de acurácia no processo de classificação, tendo sido obtido, na fase final de classificação, a taxa máxima de 100% na discriminação entre sinais de vozes saudáveis e vozes afetadas por edema e de 96% na discriminação entre sinais de vozes com edema e vozes com outras patologias (nódulos, cistos e paralisia).

Fonseca e Pereira [133] utilizaram uma abordagem baseada na análise *wavelet* para avaliar diferenças no *jitter* de 71 sinais de voz, sendo 30 saudáveis, 25 de pacientes com nódulos e 16 de pacientes com edema de Reinke. Os resultados obtidos a partir do classificador SVM para a discriminação entre os diferentes grupos de sinais apresentaram as seguintes taxas máximas de acurácia: saudável *versus* patológica, 88,20%; saudável *versus* edema, 85,30%; saudável *versus* nódulos, 90,10% e nódulos *versus* edema 82,40%.

Wang *et al.* [17] utilizaram um subconjunto da base de dados da Kay Elemetrics [60] composto por 53 sinais de vozes saudáveis e 173 sinais de vozes afetadas por diversas patologias nas pregas vocais. Nesse trabalho é utilizada análise mel-cepstral para extração de vetores de dimensão 16 como entrada para um classificador baseado na combinação de um modelo com 16 misturas Gaussianas e SVM (GMM-SVM). Os resultados indicam que um sinal da vogal sustentada /ah/ pode ser classificado como sendo de uma voz saudável ou patológica com uma acurácia de  $96,1\% \pm 2,51\%$ .

## CAPÍTULO 7

---

# Conclusões e Sugestões

---

### 7.1 Introdução

Muito embora métodos de análise de sinais de voz baseados no modelo linear de produção da fala apresentem, ainda, um campo aberto para a pesquisa em diversas aplicações que envolvem o processamento digital de sinais como, por exemplo, codificação, reconhecimento de fala e de locutor e discriminação entre sinais de vozes saudáveis e patológicas, outras abordagens têm sido investigadas para esses mesmos fins. A análise baseada na teoria dos sistemas dinâmicos não lineares e na teoria do caos tem oferecido caminhos alternativos para a extração de informações do sistema de produção da fala a partir da saída desse sistema representado por meio de uma série temporal (sinal de voz).

Tratando-se particularmente da discriminação de vozes patológicas por meio do uso de técnicas de processamento digital de sinais, objeto desta pesquisa, o estudo é fundamentado na busca de características que possam representar, de forma eficiente, sinais de vozes obtidos de locutores saudáveis, isto é, sem apresentarem patologias na laringe - mais especificamente nas pregas vocais - e sinais de vozes oriundos de indivíduos com algum tipo de patologia laríngea.

Neste capítulo são apresentadas de forma resumida as principais etapas do procedimento experimental, além dos aspectos mais relevantes dos resultados obtidos e, por fim, sugestões para trabalhos futuros.

### 7.2 Sumário da Pesquisa

Neste trabalho foram realizados experimentos computacionais (simulações) reunindo um conjunto de oito medidas decorrentes da análise não linear e da teoria do caos que, via de regra, são abordadas de forma fragmentada na literatura. O potencial discriminativo de cada uma dessas medidas é avaliado, primeiro, por meio de testes estatísticos, e posteriormente, como características (de forma individual e combinada) nos processos de classificação. Além desse

conjunto de medidas, outro conjunto é apresentado e avaliado no contexto da discriminação de vozes patológicas, sendo formado por oito medidas de quantificação de recorrência.

São apresentados a seguir os procedimentos realizados durante a fase de obtenção dos resultados desta pesquisa.

1. Determinação dos parâmetros de imersão (passo de reconstrução e dimensão de imersão mínima) para cada um dos 171 sinais de voz analisados, sendo 53 sinais de vozes saudáveis e 118 sinais de vozes afetadas por patologias laríngeas (paralisia, edema e nódulos nas pregas vocais);
2. Caracterização acústica dos sinais de vozes saudáveis e patológicas por meio da extração das invariantes relativas à Teoria do Caos e outras medidas da análise não linear:
  - Dimensão de Correlação ( $D_2$ );
  - Maior Expoente de Lyapunov ( $\lambda_1$ );
  - Entropia de Correlação ( $K_2$ );
  - Entropia de Shannon ( $H_1$ );
  - Entropia de Tsallis ( $H_q$ );
  - Entropia Aproximada ( $ApEn$ );
  - Expoente de Hurst ( $H$ ).
3. Investigação do potencial discriminativo das medidas de dinâmica não linear (MNL) por meio da análise estatística com a realização de testes de hipóteses;
4. Caracterização acústica dos sinais de vozes saudáveis e patológicas por meio da extração das medidas de quantificação de recorrência (MQR):
  - Taxa de recorrência ( $REC$ );
  - Determinismo ( $DET$ );
  - Comprimento máximo das estruturas diagonais ( $L_{max}$ );
  - Entropia da distribuição de frequências dos comprimentos das linhas diagonais ( $ENTR$ );
  - Tendência ( $TREND$ );
  - Laminaridade ( $LAM$ );
  - Comprimento médio das estruturas verticais ( $TT$ );
  - Comprimento máximo das estruturas verticais ( $V_{max}$ ).
5. Investigação do potencial discriminativo das medidas de quantificação de recorrência (MQR) por meio da análise estatística com a realização de testes de hipóteses;

6. Classificação entre sinais de vozes saudáveis e patológicas (paralisia, edema e nódulos como um só grupo) utilizando as medidas de análise não linear (MNL) e as medidas de quantificação de recorrência (MQR), de forma individual e combinada, com os classificadores de discriminante linear (LDA) e de discriminante quadrática (QDA);
7. Classificação entre vozes saudáveis e vozes afetadas por cada uma das patologias consideradas individualmente (edema, paralisia e nódulos) usando as medidas convencionais de análise não linear, de forma individual e combinada, com os classificadores LDA e QDA;
8. Discriminação de vozes afetadas por diferentes patologias (edema *versus* paralisia, edema *versus* nódulos e paralisia *versus* edema) usando as medidas de análise não linear (MNL) e as medidas de quantificação de recorrência (MQR), de forma individual e combinada, com os classificadores LDA e QDA;
9. Utilização de vetores híbridos formados pelas medidas de análise não linear (MNL), medidas de quantificação de recorrência (MQR) e coeficientes de predição linear (análise LPC) para a classificação entre vozes saudáveis e patológicas (paralisia, edema e nódulos), considerando as patologias como um só grupo e também em três diferentes grupos de patologias específicas;
10. Utilização de vetores híbridos formados pelas medidas de análise não linear (MNL), medidas de quantificação de recorrência (MQR) e coeficientes de predição linear (análise LPC) para a discriminação entre sinais de vozes afetadas por diferentes patologias (edema *versus* paralisia, edema *versus* nódulos e paralisia *versus* edema).

A fim de avaliar o desempenho do sistema proposto, foi aplicada a técnica de validação cruzada (*Cross-Validation k-fold*) nas etapas de treino/classificação, com a divisão do conjunto de características em 10 subconjuntos mutuamente exclusivos.

### 7.3 Aspectos Relevantes dos Resultados

Os resultados obtidos por meio dos processos de classificação sugerem que, de um modo geral, as medidas investigadas podem ser utilizadas de forma eficaz na construção de sistemas de discriminação automática de patologias laríngeas. Foi observado, contudo, que cada patologia apresenta particularidades que requerem diferentes conjuntos de características para representar as desordens vocais a ela associadas, permitindo a discriminação entre patologias, ou entre vozes patológicas e vozes saudáveis. Neste trabalho, foi possível identificar, entre as medidas avaliadas, as combinações de características que oferecem melhor desempenho nos processos de classificação.

Outro ponto de relevância foi a verificação de que a combinação de medidas da análise linear (especificamente os coeficiente LPC) e de análise não linear, por meio do uso de vetores híbridos, oferece melhoras significativas no desempenho do classificador, notadamente no caso da discriminação entre grupos de vozes afetadas por diferentes patologias.

A partir dos experimentos realizados com o uso de duas diferentes funções discriminante (linear e quadrática), foi possível observar a influência do classificador no desempenho final do processo de classificação. A análise discriminante quadrática (QDA) apresentou, na maioria dos casos analisados, taxas de classificação superiores às obtidas com o uso da análise discriminante linear (LDA), sobretudo, quando é feita a combinação de características de diferentes conjuntos de medidas (medidas das análises linear e não linear).

## 7.4 Contribuições da Pesquisa

As medidas de análise dinâmica não linear empregadas neste trabalho vem sendo utilizadas há algum tempo, embora de forma esparsa, no intuito de servir como base para sistemas que buscam discriminar sinais de vozes saudáveis e de vozes patológicas. Nesse sentido, um fato relevante desta pesquisa é a caracterização e a identificação do potencial individual das MNL e de conjuntos formados pela combinação dessas medidas, que oferecem o melhor desempenho na classificação entre sinais de vozes saudáveis e patológicas, bem como entre sinais de vozes saudáveis e algumas patologias específicas (paralisia, edema e nódulos). Adicionalmente, é introduzido, para esta aplicação, um novo conjunto de medidas - as medidas de quantificação de recorrência (MQR). Estas medidas apresentam desempenho equivalente a diversas outras medidas utilizadas para o mesmo fim, mas com o mérito de serem obtidas a partir de algoritmos relativamente simples. Importante também destacar o uso de vetores híbridos, formados a partir da combinação adequada das características dos conjuntos MNL e MQR com vetores extraídos da análise LPC, com resultados significativos na capacidade de discriminação entre as patologias consideradas neste trabalho.

## 7.5 Sugestões para Trabalhos Futuros

Embora os resultados obtidos nesta pesquisa possam ser, de uma forma geral, considerados relevantes, são apresentadas, a seguir, algumas sugestões que poderão oferecer melhorias no processo que se refere à classificação de vozes patológicas, bem como, na utilização das medidas estudadas em outras aplicações na área do processamento digital de sinais de voz.

1. Utilização das medidas de recorrência calculadas a partir da segmentação dos sinais de voz;

2. Verificação da eficiência no processo de classificação a partir do uso da combinação das medidas de quantificação de recorrência e medidas de perturbação (como *jitter* e *shimmer*) e da análise linear (*e.g.*, coeficientes Mel Cepstrais e suas derivadas);
3. Utilização de classificadores mais sofisticados, como por exemplo, SVM, GMM e redes neurais;
4. Investigação do uso de técnicas de redução de dimensionalidade (*e.g.*, PCA, Relação-F e o discriminante de Fisher) que possam minimizar o número de características que compõem os vetores usados no processo de classificação sem, contudo, apresentar perda significativa de desempenho;
5. Aplicação das medidas de recorrência na discriminação de outras patologias;
6. Avaliação do desempenho das medidas no processo de classificação após a separação por gênero dos locutores e grau de severidade das patologias;
7. Construção de uma base de dados composta, por exemplo, por vozes de professores para avaliação e um pré-diagnóstico relativo à presença de possíveis patologias nas pregas vocais;

Além disso, pode ser investigada a aplicação das medidas de recorrência em sistemas de reconhecimento de fala e de locutor.

---

---

## Referências Bibliográficas

---

- [1] J. Deller, J. Proakis, and J. Hansen, *Discrete-time Processing of Speech Signals*. Wiley-IEEE Press, 2000.
- [2] A. F. Dalley, “Dynamic human anatomy, versão 2.0, (cd-rom, institutional edition),” Lippincott Williams and Wilkins.
- [3] M. Lucian Sulica. (2011) Reinke’s edema. Voice medicine. [Online]. Available: [http://www.voicemedicine.com/reinkes\\_edema.htm](http://www.voicemedicine.com/reinkes_edema.htm)
- [4] M. Behlau, *Voz, O Livro do Especialista*, E. Revinter, Ed., 2001, vol. I.
- [5] M. James P. Thomas, March 2010. [Online]. Available: [http://www.voicedoctor.net/media/photo/mucosal/reinke/reinke\\_m.jpg](http://www.voicedoctor.net/media/photo/mucosal/reinke/reinke_m.jpg)
- [6] J. Jiang, Y. Zhang, and C. McGilligan, “Chaos in voice, from modeling to measurement,” *Journal of Voice*, vol. 20, no. 1, pp. 2–17, 2006.
- [7] J. C. Stemple, L. Glaze, and B. Klaben, *Clinical Voice Pathology - Theory and Management*, 4th ed. Plural Publishing, 2010.
- [8] I. Titze, *Workshop on acoustic voice analysis: Summary statement*. National Center for Voice and Speech, 1995.
- [9] G. Fant, *Speech acoustics and phonetics*. Springer, 2004, vol. 24.
- [10] S. B. Davis, “Acoustic characteristics of normal and pathological voices,” *Speech and Language: Advances in Basic Research and Practice*, vol. 1, pp. 271–335, 1979.
- [11] F. Quek, M. Harper, Y. Haciahmetoglou, L. Chen, and L. O. Raming, “Speech pauses and gestural holds in parkinson’s disease,” in *Proc. of International Conference on Spoken Language Processing*, 2002, pp. 2485–2488.
- [12] L. Cummings, *Clinical linguistics*. Edinburgh Univ Pr, 2008.
- [13] C. A. Araújo, “Cuidados com a voz,” *Revista Associação Médica Brasileira*, p. 48:24., 2004.



- 
- [14] A. Barbosa-Branco and M. Romariz, “Doenças das cordas vocais e sua relação com o trabalho,” *Comum Ciênc Saúde*, vol. 17, no. 1, pp. 37–45, 2006.
- [15] L. Rabiner and R. Schafer, *Digital processing of speech signals*. Prentice-hall Englewood Cliffs, NJ, 1978, vol. 100.
- [16] R. K. Balasubramaniam, J. S. Bhat, S. F. III, and R. R. III, “Cepstral analysis of voice in unilateral adductor vocal fold palsy,” *Journal of Voice*, vol. 25, no. 3, pp. 326–329, 2011.
- [17] X. Wang, J. Zhang, and Y. Yan, “Discrimination between pathological and normal voices using gmm-svm approach,” *Journal of Voice*, vol. 25, pp. 38–43, 2011.
- [18] B. R. Kumar, J. Bhat, and N. Prasad, “Cepstral analysis of voice in persons with vocal nodules,” *Journal of Voice*, vol. 24(6), pp. 651–653, 2010.
- [19] S. L. do Nascimento Cunha Costa, *Análise Acústica Baseada no Modelo Linear de Produção da Fala para Discriminação de Vozes Patológicas*. Tese de doutorado, UFCG - Universidade Federal de Campina Grande, 2008.
- [20] J. Fechine, *Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística*. Tese de doutorado, Universidade Federal da Paraíba, Campina Grande, Brasil, 2000.
- [21] 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.
- [22] E. Fonseca, R. Guido, A. Silvestre, and J. Pereira, “Discrete wavelet transform and support vector machine applied to pathological voice signals identification,” *Seventh IEEE International Symposium on Multimedia*, pp. 785–789, 2005.
- [23] M. Bahoura and C. Pelletier, “Respiratory sounds classification using cepstral analysis and gaussian mixture models,” in *Engineering in Medicine and Biology Society, 2004. IEMBS’04. 26th Annual International Conference of the IEEE*, vol. 1. IEEE, 2004, pp. 9–12.
- [24] C. Adnene, B. Lamia, and M. Mounir, “Analysis of pathological voices by speech processing,” in *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, vol. 1. IEEE, 2003, pp. 365–367.
- [25] M. de Oliveira Rosa, *Laringe Digital*. Tese de doutorado, Escola de Engenharia de São Carlos da Universidade de São Paulo, 2002.

- 
- [26] A. Kumar and S. K. Mullick, "Nonlinear dynamical analysis of speech," *Journal of Acoustical Society of America*, vol. 100, No. 1, pp. 615–629, 1996.
- [27] I. Kokkinos and P. Maragos, "Nonlinear speech analysis using models for chaotic systems," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 6, pp. 1098–1109, 2005.
- [28] A. Petry, *Reconhecimento Automático de Locutor Utilizando Medidas de Invariantes Dinâmicas Não-Lineares*. Tese de doutorado, Universidade Federal do Rio Grande do Sul, 2002.
- [29] L. Chai, A. Sprecher, Y. Zhang, Y. Liang, H. Chen, and J. Jiang, "Perturbation and nonlinear dynamic analysis of adult male smokers," *Journal of Voice*, vol. 25(3), pp. 342–347, 2011.
- [30] G. Vaziri, F. Almasganj, and R. Behroozmand, "Pathological assessment of patients' speech signals using nonlinear dynamical analysis," *Computers in Biology and Medicine*, vol. 40, no. 1, pp. 54–63, 2010.
- [31] S. Awan, N. Roy, and J. Jiang, "Nonlinear dynamic analysis of disordered voice: The relationship between the correlation dimension ( $d_2$ ) and pre-/post-treatment change in perceived dysphonia severity," *Journal of Voice*, vol. 24(3), no. 3, pp. 285–293, 2010.
- [32] P. Henríquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. D. de Maria, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1186–1195, August 2009.
- [33] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhys. Letters*, vol. 4(9), pp. 973–977, 1987.
- [34] 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.
- [35] A. Schmied and M. Descarreaux, "Reliability of emg determinism to detect changes in motor unit synchrony and coherence during submaximal contraction," *Journal of neuroscience methods*, vol. 196(2), pp. 238–246, 2011.
- [36] U. R. Acharya, S. V. Sree, S. Chattopadhyay, W. Yu, and P. Ang, "Application of recurrence quantification analysis for the automated identification of epileptic eeg signals," *International Journal of Neural Systems*, vol. 21, no. 3, pp. 199–211, 2011.
- [37] M. Aboofazeli and Z. Moussavi, "Comparison of recurrence plot features of swallowing and breath sounds," *Chaos, Solitons & Fractals*, vol. 37, no. 2, pp. 454–464, 2008.

- 
- [38] J. Zbilut, N. Thomasson, and C. Webber, "Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals," *Medical engineering & physics*, vol. 24, no. 1, pp. 53–60, 2002.
- [39] V. Cermak, L. Bodri, and J. Safanda, "Tidal modulation of temperature oscillations monitored in borehole yaxcopoil-1 (yucatán, mexico)," *Earth and Planetary Science Letters*, vol. 282, no. 1-4, pp. 131–139, 2009.
- [40] J. A. Bastos and J. Caiado, "Recurrence quantification analysis of global stock markets," *Physica A: Statistical Mechanics and its Applications*, vol. 390(7), pp. 1315–1325, 2011.
- [41] F. Palmieri and U. Fiore, "A nonlinear, recurrence-based approach to traffic classification," *Computer Networks*, vol. 53, no. 6, pp. 761–773, 2009.
- [42] J. Menezes and G. Barreto, "Long-term time series prediction with the narx network: An empirical evaluation," *Neurocomputing*, vol. 71, no. 16, pp. 3335–3343, 2008.
- [43] M. Little, D. Costello, and M. Harries, "Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures," *Journal of Voice*, vol. 25, no. 1, pp. 21–31, 2011.
- [44] J. Arias-Londono, J. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Dominguez, "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients." *IEEE Transactions on bio-medical engineering*, vol. 58, no. 2, pp. 370–379, 2011.
- [45] M. Little, P. McSharry, S. Roberts, D. Costello, and I. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering OnLine*, vol. 6, no. 1, p. 23, 2007.
- [46] S. Zitta, "Análise perceptivo-auditiva e acústica em mulheres com nódulos vocais," Dissertação de Mestrado, Centro Federal de Educação Tecnológica do Paraná, 2010.
- [47] M. Hirano and K. Sato, *Histological Color Atlas of the Human Larynx*. Singular Pub. Group, 1993.
- [48] B. Benjamin, *Cirurgia Endolaríngea*, E. Revinter, Ed., Rio de Janeiro, 2000.
- [49] I. Kuhl, *Manual Prático de Laringologia*, E. da Universidade UFRGS, Ed., Porto Alegre, 1982.
- [50] A. Parraga, "Aplicação da transformada wavelet packet na análise e classificação de sinais de vozes patológicas," Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, 2002.

- 
- [51] H. M. Teager, "Some observations on oral air flow during phonation," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 5, pp. 599–601, 1980.
- [52] H. M. Teager and S. M. Teager, "A phenomenological model for vowel production in the vocal tract," in *Speech Science: Recent Advances*, R. G. Daniloff, Ed. San Diego, CA: College-Hill Press, 1985, ch. 3, pp. 73–109.
- [53] H. M. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract," in *Speech Production and Speech Modelling*, W. J. Hardcastle and A. Marchal, Eds. Boston, MA: Kluwer Academic, 1990, pp. 241–261.
- [54] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D: Nonlinear Phenomena*, vol. 142, no. 3, pp. 346–382, 2000.
- [55] I. Tokuda, T. Miyano, and K. Aihara, "Surrogate analysis for detecting nonlinear dynamics in normal vowels," *The Journal of the Acoustical Society of America*, vol. 110, p. 3207, 2001.
- [56] M. Little, "Biomechanically informed nonlinear speech signal processing," Ph.D. dissertation, University of Oxford, 2006.
- [57] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doynne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1, pp. 77–94, 1992.
- [58] L. E. V. da Silva, *Ferramentas computacionais na análise da variabilidade da frequência cardíaca através do paradigma não extensivo no estudo de cardiopatias*. Dissertação de mestrado, USP - Universidade de São Paulo, 2010.
- [59] W. Fisher, G. Doddington, and K. Goudie-Marshall, "The darpa speech recognition research database: specifications and status," in *Proc. DARPA Workshop on Speech Recognition*, 1986, pp. 93–99.
- [60] K. Elemetrics, *Kay Elemetrics Corp. Disordered Voice Database*, Model 4337, 03 Ed., 1994.
- [61] I. Titze, R. Baken, and H. Herzel, "Evidence of chaos in vocal fold vibration," pp. 143–188, 1993.
- [62] Y. Zhang, C. McGilligan, L. Zhou, M. Vig, and J. Jiang, "Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps," *The Journal of the Acoustical Society of America*, vol. 115, p. 2270, 2004.
- [63] M. Savi, "Chaos and order in biomedical rhythms," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 27, no. 2, pp. 157–169, 2005.

- 
- [64] H. Kantz and T. Schreiber, *Nonlinear time series analysis*. Cambridge Univ Pr, 2004, vol. 7.
- [65] G. Kubin, “Nonlinear processing of speech,” *Speech coding and synthesis*, pp. 557–610, 1995.
- [66] M. Savi, *Dinâmica não-linear e caos*. Editora E-papers, 2006.
- [67] N. Fiedler-Ferrara and C. P. C. Prado, *Caos - uma Introdução*. Editora Edgard Blücher, 1994.
- [68] F. Takens, “Detecting strange attractors in turbulence,” *Dynamical systems and turbulence, Warwick 1980*, pp. 366–381, 1981.
- [69] N. Packard, J. Crutchfield, J. Farmer, and R. Shaw, “Geometry from a time series,” *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.
- [70] A. Fraser and H. Swinney, “Independent coordinates for strange attractors from mutual information,” *Physical Review A*, vol. 33, no. 2, p. 1134, 1986.
- [71] F. Hausdorff, “Der wertvorrat einer bilinearform,” *Mathematische Zeitschrift*, vol. 3, pp. 314–316, 1919.
- [72] M. Kennel, R. Brown, and H. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Physical review A*, vol. 45, no. 6, p. 3403, 1992.
- [73] A. Wolf, J. Swift, H. Swinney, and J. Vastano, “Determining lyapunov exponents from a time series,” *Physica D: Nonlinear Phenomena*, vol. 16, no. 3, pp. 285–317, 1985.
- [74] H. Kantz, “A robust method to estimate the maximal lyapunov exponent of a time series,” *Physics Letters A*, vol. 185, no. 1, pp. 77–87, 1994.
- [75] W. Steeb, W. Steeb, Y. Hardy, and R. Stoop, *The nonlinear workbook: chaos, fractals, cellular automata, neural networks, genetic algorithms, gene expression programming, support vector machine, wavelets, hidden Markov models, fuzzy logic with CC++, Java and Symbolic C++ programs*. World Scientific Pub Co Inc, 2005.
- [76] J. Jiang and Y. Zhang, “Nonlinear dynamic analysis of speech from pathological subjects,” *Electronics Letters*, vol. 38, no. 6, pp. 294–295, 2002.
- [77] J. Farmer, E. Ott, and J. Yorke, “The dimension of chaotic attractors,” *Physica D: Nonlinear Phenomena*, vol. 7, no. 1-3, pp. 153–180, 1983.
- [78] S. Strogatz, *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Westview Pr, 1994.

- 
- [79] L. H. A. Monteiro, *Sistemas Dinâmicos*. Livraria da Física, 2006.
- [80] P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189–208, 1983.
- [81] C. Shannon, “A mathematical theory of communication,” *ACM Sigmoble Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.
- [82] M. Torres, L. Gamero, H. Rufiner, C. Martínez, D. Milone, and G. Schlotthauer, “Study of complexity in normal and pathological speech signals,” in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 3. IEEE, 2003, pp. 2339–2342.
- [83] A. Rényi, “On measures of entropy and information,” in *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1961, pp. 547–561.
- [84] J. Arias-Londoño, J. Godino-Llorente, G. Castellanos-Dominguez, N. Sáenz-Lechón, and V. Osma-Ruiz, “Complexity analysis of pathological voices by means of hidden markov entropy measurements,” in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 2248–2251.
- [85] S. Pincus, “Approximate entropy as a measure of system complexity,” *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, p. 2297, 1991.
- [86] C. Chakrabarti and K. De, “Boltzmann-gibbs entropy: axiomatic characterization and application,” *International Journal of Mathematics and Mathematical Sciences*, vol. 23, no. 4, pp. 243–251, 2000.
- [87] C. Tsallis, “Possible generalization of boltzmann-gibbs statistics,” *Journal of statistical physics*, vol. 52, no. 1, pp. 479–487, 1988.
- [88] A. Tavares, *Aspectos matemáticos da entropia*. Dissertação de mestrado, Universidade de Aveiro, Portugal, 2003.
- [89] H. Hurst, “Long-term storage capacity of reservoirs,” *Trans. Amer. Soc. Civil Eng.*, vol. 116, pp. 770–808, 1951.
- [90] J. Beran, *Statistics for Long-Memory Processes*. Chapman & Hall, 1994.
- [91] M. Taqqu, V. Teverovsky, and W. Willinger, “Estimators for long-range dependence: An empirical study,” *Fractals*, vol. 3, no. 4, pp. 785–788, 1995.
- [92] N. Marwan, “Encounters with neighbours - current developments of concepts based on recurrence plots and their applications,” Ph.D. dissertation, University of Potsdam, 2003.

- 
- [93] A. Behrman and R. Baken, "Correlation dimension of electroglottographic data from healthy and pathologic subjects," *The Journal of the Acoustical Society of America*, vol. 102, p. 2371, 1997.
- [94] N. Marwan and J. Kurths, "Line structure in recurrence plots," *Physics Letters A*, vol. 336, pp. 349–357, 2004.
- [95] J. P. Zbilut and C. L. Webber-Jr., "Embeddings and delays as derived from quantification of recurrence plots," *Physics Letters A*, vol. 171 (3-4), pp. 199–203, 1992.
- [96] C. Webber Jr and J. Zbilut, "Recurrence quantifications: Feature extractions from recurrence plots," *Int. J. Bifurcation and Chaos*, vol. 17, pp. 3467–3475, 2007.
- [97] S. Siegel and N. J. C. Jr., *Estatística Não-Paramétrica para Ciências do Comportamento*, 2nd ed. Artmed, 2006.
- [98] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, 3rd ed. Wiley, 2003.
- [99] R. Larson and B. Farber, *Estatística Aplicada*, 4th ed. Pearson Prentice Hall, 2010.
- [100] N. Sáenz-Lechón, J. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120–128, 2006.
- [101] S. Hadjitodorov and P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening," *Medical engineering & physics*, vol. 24, no. 6, pp. 419–429, 2002.
- [102] V. Parsa and D. Jamieson, "Identification of pathological voices using glottal noise measures," *Journal of speech, language, and hearing research*, vol. 43, no. 2, pp. 469–485, 2000.
- [103] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The tisean package," *Chaos*, vol. 9, no. 2, pp. 413–435, 1999.
- [104] E. Kononov, "Visual recurrence analysis 4.9," 2006, software disponível em <http://nonlinear.110mb.com/vra/>.
- [105] C. L. Webber-Jr. (2009, June) Recurrence quantification analysis, v. 13.1. Software package disponível em <http://homepages.luc.edu/~cwebber/>. [Online]. Available: <http://homepages.luc.edu/~cwebber/>
- [106] N. Marwan. (2010, Jan) Cross recurrence plot toolbox, v. 5.15, r(28.5). Software disponível em <http://www.agnld.uni-potsdam.de/~marwan/toolbox/>. [Online]. Available: <http://www.agnld.uni-potsdam.de/~marwan/toolbox/>

- 
- [107] P. Scalassara, *Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias da Laringe*. Tese de doutorado, Escola de Engenharia de São Carlos da Universidade de São Paulo, 2009.
- [108] E. G. Souza, *Caracterização de Sistemas Dinâmicos Através de Gráficos de Recorrência*. Dissertação de mestrado, Universidade Federal do Paraná, 2008.
- [109] G. H. Erban, *Análise de Séries de Tempo Financeiras - Uma Aplicação da Teoria do Caos em Finanças Empíricas*. Dissertação de mestrado, Escola de Administração de Empresas de São Paulo-Fundação Getúlio Vargas, São Paulo, 2004.
- [110] V. T. Motta and P. F. de Oliveira Filho, “Spss: análise de dados biomédicos,” 2009.
- [111] H. Falcão, S. Correia, S. Costa, J. Fachine, and B. Aguiar Neto, “O uso da entropia na discriminação de vozes patológicas,” in *Anais do 21<sup>o</sup> Congresso Brasileiro de Engenharia Biomédica*. Salvador: Sociedade Brasileira de Engenharia Biomédica, 2008, pp. 1599–1602.
- [112] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Phy*, vol. 45 (6), pp. 3403–3411, 1992.
- [113] 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*, M. A. Riley and G. C. V. Orden, Eds., 2005. [Online]. Available: <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>
- [114] E. Melo, L. Brito, O. Brasil, M. Behlau, and D. Melo, “Incidência de lesões laríngeas não neoplásicas em pacientes com queixas vocais,” *Revista Brasileira de Otorrinolaringologia*, vol. 67, no. 6, pp. 788–794, 2001.
- [115] I. Witten, E. Frank, and M. Hall, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2011.
- [116] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *International joint Conference on artificial intelligence*, vol. 14, 1995, pp. 1137–1145.
- [117] L. J. Kazmier, *Estatística Aplicada à Economia e Administração*. McGraw-Hill do Brasil, São Paulo, 1982.
- [118] O. Hersent, D. Guide, and J.-P. Petit, *Telefonia IP: Comunicação multimídia baseada em pacotes*. Pearson Education do Brasil, São Paulo, 2002.



- 
- [119] D. O'shaughnessy, *Speech communications: human and machine*. Universities press, 2000.
- [120] F. Martinez, A. Guillamon, J. Alcaraz, and M. Alcaraz, "Detection of chaotic behaviour in speech signals using the largest lyapunov exponent," in *Digital Signal Processing, 14th International Conference on DSP*, vol. 1. IEEE, 2002, pp. 317–320.
- [121] V. Pitsikalis, I. Kokkinos, and P. Maragos, "Nonlinear analysis of speech signals: Generalized dimensions and lyapunov exponents," in *Proc. of Eurospeech*. Citeseer, 2003.
- [122] A. Lindgren, M. Johnson, and R. Povinelli, "Speech recognition using reconstructed phase space features," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), 2003*, vol. 1. IEEE, 2003, pp. 6–63.
- [123] M. Johnson, R. Povinelli, A. Lindgren, J. Ye, X. Liu, and K. Indrebo, "Time-domain isolated phoneme classification using reconstructed phase spaces," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 4, pp. 458–466, 2005.
- [124] Y. Kakita and H. Okamoto, "Visualizing the characteristics of vocal fluctuation from the viewpoint of chaos: an attempt toward qualitative quantification," *Vocal Fold Physiology–Voice Quality Control. San Diego, CA: Singular Publishing Group*, pp. 235–248, 1995.
- [125] I. Hertrich, W. Lutzenberger, S. Spieker, and H. Ackermann, "Fractal dimension of sustained vowel productions in neurological dysphonias: An acoustic and electroglottographic analysis," *The Journal of the Acoustical Society of America*, vol. 102, p. 652, 1997.
- [126] A. Giovanni, M. Ouaknine, and J. Triglia, "Determination of largest lyapunov exponents of vocal signal: Application to unilateral laryngeal paralysis," *Journal of Voice*, vol. 13, no. 3, pp. 341–354, 1999.
- [127] Y. Zhang and J. Jiang, "Nonlinear dynamic analysis in signal typing of pathological human voices," *Electronics Letters*, vol. 39, no. 13, pp. 1021–1023, 2003.
- [128] B. S. Aghazadeh, H. Khadivi, and M. Nikkhah-Bahrami, "Nonlinear analysis and classification of vocal disorders," in *Proceedings of 29th Annual International Conference of the IEEE EMBS*, August 2007, pp. 6199–6202.
- [129] Y. A. El-Imam and A. S. Elwakil, "Applying recurrence quantification and spectral analysis to represent nasalization in speech signals," in *Proceedings of the 7th WSEAS International Conference on Multimedia Systems & Signal Processing*, April 2007, pp. 129–132.
- [130] Y. Zhang and J. J. Jiang, "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," *Journal of Voice*, vol. 22, no. 1, pp. 1–9, January 2008.

- [131] G. Vaziri, F. Almasganj, and M. Jenabi, "On the fractal self-similarity of laryngeal pathologies detection: The estimation of hurst parameter," in *Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on.* IEEE, 2008, pp. 383–386.
- [132] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [133] E. S. Fonseca and J. C. Pereira, "Normal versus pathological voice signals," *IEEE EMB Magazine*, pp. 44–48, 2009.