Universidade de São Paulo
Departamento de Engenharia Elétrica

Fernando Araujo de Andrade Sobrinho

Detecção de patologias em pregas vocais
utilizando a seção Poincaré do espaço de fase
tridimensional de um sinal de voz

São Carlos
2016

FERNANDO ARAUJO DE ANDRADE SOBRINHO

DETECÇÃO DE PATOLOGIAS EM PREGAS VOCAIS UTILIZANDO A SEÇÃO POINCARÉ DO ESPAÇO DE FASE TRIDIMENSIONAL DE UM SINAL DE VOZ

Doutor em Ciências, Programa de Engenharia
Elétrica.

Área de concentração: Sistemas Dinâmicos.

Orientador: Prof. Dr. Luis Fernando Costa
Alberto

SÃO CARLOS

2016

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Araujo de Andrade Sobrinho, Fernando

A553m Detecção de patologias em pregas vocais utilizando a seção de Poincaré do espaço de fase tridimensional de um sinal de voz / Fernando Araujo de Andrade Sobrinho; orientador Luis Fernando Costa Alberto; coorientadora Maria Eugênia Dager. São Carlos, 2016.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em sistemas Dinâmicos — Escola de Engenharia de São Carlos da Universidade de São Paulo, 2016.

1. Processamento digital de sinais de voz. 2. Seção de Poincaré. 3. Análise não linear. 4. Espaço de fase. 5. Análise de voz. I. Título.

FOLHA DE JULGAMENTO

Candidato: Engenheiro **FERNANDO ARAÚJO DE ANDRADE SOBRINHO**.

Título da tese: "Detecção de patologias em pregas vocais utilizando a seção *Poincaré* do espaço de fase tridimensional de um sinal de voz".

Data da defesa: 02/09/2016.

Comissão Julgadora:

Resultado:

Prof. Associado **Luís Fernando Costa Alberto (Orientador)**
(Escola de Engenharia de São Carlos/EESC)

Aprovado

Prof. Titular **José Carlos Pereira**
(Escola de Engenharia de São Carlos/EESC)

Aprovado

Prof. Dr. **Marcio Eisencraft**
(Escola Politécnica/EP-USP)

Aprovado

Prof. Dr. **Alan Petrônio Pinheiro**
(Universidade Federal de Uberlândia/UFU)

APROVADO

Prof. Dr. **Bruno Catarino Bispo**
(Universidade Tecnológica Federal do Paraná/UTFPR)

Aprovado

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica:
Prof. Associado **Luis Fernando Costa Alberto**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Luis Fernando Costa Alberto**

Dedicatória

À minha esposa, com amor, admiração e gratidão por sua compreensão, carinho, presença e incansável apoio ao longo do período da elaboração deste trabalho.

Agradecimentos

Agradeço à Deus pelo seu grande amor por mim.

Aos meus pais Sérgio e Silvana por todo esforço que fizeram para que eu chegasse até aqui e por mostrar desde cedo a importância dos estudos.

Ao Prof. Dr. Luis Fernando Costa Alberto por ter me concedido a oportunidade de realizar este trabalho, pela orientação e amizade, contribuindo para meu crescimento científico e intelectual.

À Doutora Maria Eugênia Dajer, pela amizade, companheirismo e as publicações geradas em conjunto que contribuíram para o enriquecimento desse trabalho.

Ao Prof. Dr. José Carlos Pereira que deu início a todo esse trabalho.

Aos amigos da República dos Largados: Farelo, Bicudo, Spin, Velhinho, Luizinho, Sinotti, Mori, Maru e agregados pelo apoio e incentivo durante todo o período da pós-graduação.

À Roseli Ribeiro e João Paulo Brognoni Casati por todo apoio na solução de problemas com meu notebook.

Ao amigo Cláudio Cyrino Duarte de Souza pela amizade e todo apoio durante a graduação e pós-graduação.

Resumo

Sobrinho, Fernando A. de A. **Detecção de patologias em pregas vocais utilizando a seção de Poincaré do espaço de fase tridimensional de um sinal de voz**. 2016. 105 f. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2016.

Diversos estudos foram realizados para detecção de patologias na laringe. Essas patologias causam alteração na frequência, amplitude e formato de onda do sinal de voz e podem ser estudadas através dos parâmetros convencionais de análise como *jitter* e *shimmer*, ou sob o enfoque da dinâmica não linear. Essas técnicas são não invasivas e servem de apoio ao especialista da área de fonoaudiologia para o diagnóstico de patologias nas pregas vocais. As técnicas de análise acústica baseiam-se no formato de onda vocal no domínio do tempo e domínio da frequência, enquanto que a técnica de análise não linear utilizada nesse trabalho baseia-se no atrator reconstruído do sinal de voz. O objetivo dessa tese é diferenciar vozes normais e patológicas e entre patologias usando a técnica de análise não linear conhecida como Seção de Poincaré. Foram analisados 48 sinais de vozes humanas, divididos em 3 grupos (16 normais, 16 com nódulo e 16 com edema de Reinke). Em seguida foram selecionados 3 trechos de 500 ms nos intervalos 0.5s-1.0s, 2.0s-2.5s e 4.0s-4.5s chamado de primeiro critério e um trecho 500ms no trecho de maior variação de *pitch*, chamado de segundo critério. Em seguida, o atrator foi reconstruído em 3 dimensões, determinado o atrator médio, e de cada ponto do atrator médio foi extraída a seção de Poincaré. De cada seção de Poincaré foi calculada a dispersão dos pontos do atrator no plano através da média e desvio padrão das dispersão dos pontos da seção de Poincaré em relação ao ponto médio da seção. A validação da ferramenta desenvolvida para essa tese foi realizada utilizando um sinal senoidal inserindo *jitter* gradativamente, onde verificou-se uma variação proporcional da média da dispersão. Os resultados obtidos mostraram que não foi possível diferenciar patologias mas foi possível classificar vozes normais das patológicas. O melhor intervalo para classificar as vozes normais das patológicas utilizando o primeiro critério foi entre 0.5s - 1.0s pois nesse intervalo todas as vozes normais foram classificadas corretamente. No entanto, 6 vozes patológicas foram classificadas como normais com 2 vozes patológicas na fronteira que separa as vozes normais das patológicas. O segundo critério classificou todas as vozes normais corretamente e apenas uma voz patológica foi classificada como patológica. Concluiu-se que a ferramenta proposta utilizando o segundo critério mostrou-se superior em relação ao primeiro critério para diferenciar vozes normais das patológicas.

Palavras-chave: Análise acústica de voz, Dinâmica não linear, Reconstrução do espaço de fase, Seção de Poincaré.

Abstract

Sobrinho, Fernando A. de A. **Detection of pathologies in vocal fold by means of Poincaré section of the tridimensional phase space of a voice signal.** 2016. 105 f. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2016.

Several studies have been performed to detect pathologies of the larynx. These pathologies cause changes in the frequency, amplitude, and waveform of the voice signal. They can be studied by means of conventional analysis parameters such as jitter and shimmer, or from nonlinear dynamics concepts. These techniques are noninvasive and can help the speech therapist to better diagnose the pathologies in the vocal folds. The acoustic analysis techniques are based on the voice waveform in the time and frequency domains, while the non-linear analysis techniques are based on the attractor reconstructed from the speech signal. The aim of this thesis is to differentiate normal and pathological voices using a nonlinear analysis technique named Poincaré section. We analyzed 48 human voice signals divided into 3 groups (16 normal, 16 nodule and 16 Reinke's edema). Then, we analyzed 3 stretches of 500ms in the intervals 0.5s-1.0s, 2.0s-2.5s e 4.0-4.5s, denominated first criteria, and a stretch of 500ms in a higher variation in *pitch*, denominated second criteria. The attractor was then reconstructed in three dimensions, the average attractor was determined, and at each point of the average attractor, a Poincaré section was extracted. From each Poincaré section, the dispersion of the points of the attractor was calculated in the plane by means of the statistical average and standard deviation related to the medium point of the section. The validation of the tool developed for this thesis was achieved by inserting jitter gradually in a sinusoidal wave, where there was a proportional variation of average's dispersion was observed. The results obtained for this set of voices showed that the average and standard deviation of dispersion of the points in the Poincaré section differentiate the groups of voices, but not the pathological groups. The Statistical tests of Anova and Tukey were used to analyze the 3 groups and all group pairings, two by two, with a statistical significance of 5%. The best interval to classify normal voices from pathological voices by means of the first criteria was between 0.5s - 1.0s, given the fact that in this interval, all normal voices were correctly classified. However, 6 pathological voices were classified as normal voices, with 2 voices border lining the frontier between normal voices from pathological voices. The second criteria classified all normal voices correctly, with only one pathological voice incorrectly classified. In conclusion, the second criteria tool proposed by this thesis was proven superior to differentiate normal voices from pathological ones.

Key-words: Acoustic analysis of voice, Nonlinear dynamic, Phase space reconstruction, Poincaré section.

Lista de Figuras

1	Sistema fonatório. Modificado de (DELLER; PROAKIS; HANSEN, 2000). . . .	25
2	Esboço da laringe, vista de cima, na posição de fonação e respiração. Adaptada de http://www.phys.unsw.edu.au/jw/voice.html	26
3	Sequência de um ciclo completo da fonação (DELLER; PROAKIS; HANSEN, 2000).	27
4	Forma de onda do pulso glotal. Adaptado de Dajer (2010).	28
5	Espectro de frequências do sinal de voz. Modificado de Scalassara (2009). .	29
6	Modelo fonte-filtro proposto por Fant (1970)	30
7	Modelo sistema/fonte do sinal de voz. Adaptado de Rabiner e Schafer (2007). .	30
8	Nódulos bilaterais indicado pelas setas (MARTINS; TRINDADE, 2003) (Esquerda). Prega vocal saudável (extraído de voicedoctor.net)(Direita). . . .	34
9	Espectro de frequência de um sinal de voz normal e voz com nódulo. A voz normal, apresentada à esquerda, é mais rica em componentes de frequência quando comparada à voz com nódulo, apresentada à direita.	35
10	Edema de Reinke na prega vocal. Fonte: http://www.fonologo.es	36
11	Os diferentes pontos de cruzamento no eixo x de cada ciclo do sinal de voz sobrepostos ilustram o jitter. A figura da direita é uma ampliação da figura à esquerda para melhor visualização (Fonte própria).	39
12	Série temporal da equação de <i>Van Der Pol</i>	44
13	Gráfico da informação mútua da série temporal da equação de <i>Van Der Pol</i>	45
14	Atrator original (superior) e atrator reconstruído (inferior). O tempo de atraso foi escolhido usando como critério o mínimo da função de informação mútua. Observa-se que o atrator reconstruído é topologicamente semelhante ao original (FRASER; SWINNEY, 1986b).	46

15	Gráfico da bifurcação da equação recursiva. A partir do parâmetro $r \geq 3$ o sistema apresenta uma série de pontos até se tornar caótico.	47
16	Figura ilustrativa para o cálculo do expoente de Lyapunov	49
17	Seção de Poincaré	51
18	Etapas realizadas desde a análise do sinal de voz no tempo até a análise da dispersão dos pontos na seção de Poincaré.	55
19	Sinal de voz completo	56
20	Trecho do sinal normalizado	57
21	Informação Mútua	57
22	Espaço de fase	58
23	(a) Espaço de fase rotacionado em 2 dimensões do sinal do sinal de voz da Figura 20. (b) Ampliação do espaço de fase. As linhas pontilhadas estabelecem os limites da região do corte e a orientação da linha tangente em relação ao fluxo da trajetória.	59
24	Espaço de fase ampliado. A seta mostra a ocorrência de mais de um ponto no trecho em estudo.	61
25	Ilustração dos segmentos de trajetória	63
26	Determinação dos pontos pertencentes ao plano de Poincaré através do produto escalar	64
27	A figura da esquerda mostra o trecho selecionado com os pontos resultantes do produto escalar (em vermelho) pertencentes ao plano de Poincaré e a reta tangente à trajetória. A figura da direita mostra o hiperplano de Poincaré cortando o trecho do espaço de fase.	64
28	(a) Seção de Poincaré extraída do conjunto X' mostrado na Figura 26. (b) Os pontos em vermelho representam os pontos do sinal (no tempo) que cortam o plano de Poincaré no espaço de fase.	66
29	Seções de Poincaré em vermelho obtidas em todo o espaço de fase de um sinal de voz normal.	67
30	Espaço de fase médio de um sinal de voz normal (destacado em vermelho).	68
31	Espaço de fase e seções de Poincaré extraídas de 2 sinais de vozes com nódulo.	68

32	Espaço de fase e seções de Poincaré extraídas de um sinal de voz normal (c) e um sinal de voz com edema (d).	69
33	Pontos da seção de Poincaré no tempo	69
34	Espaço de fase de um sinal de voz. As setas mostram os pontos inacessíveis utilizando o método utilizado em Sobrinho (2011).	70
35	O espectro de frequências de um sinal de voz com nódulo mostra dois picos próximos.	71
36	Frequência fundamental extraída através da seção de Poincaré em 51 ciclos do espaço de fase.	72
37	Dispersão dos pontos da seção de Poincaré para um sinal senoidal.	74
38	Sinal de voz completo normalizado. As barras vermelhas limitam um trecho com menor variação em amplitude utilizando critério visual.	76
39	Variação do <i>pitch</i> no sinal de voz. A reta destacada em verde possui a maior inclinação num intervalo de 500ms.	77
40	Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio. Intervalo: 0.5s - 1.0s.	80
41	Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio. Intervalo: 2.0s - 2.5s.	80
42	Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio. Intervalo: 4.0s - 4.5s.	80
43	Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - trechos com maior variação de <i>pitch</i>	81
44	Teste Anova - Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - 1° critério. Trecho (1.0s; 1.5s).	82
45	Teste Anova - Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - 1° critério. Trecho (2.0s;2.5s).	83
46	Teste Anova - Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - 1° critério. Trecho (4.0s;4,5s).	84
47	Classificação utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 1° critério - intervalo: 0.5s - 1.0s.	85

48	Classificação utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 1º critério -intervalo: 2.0s - 2.5s.	86
49	Classificação utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 1º critério -intervalo: 4.0s - 4.5s.	87
50	Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio em trechos com maior variação de <i>pitch</i> - 2º critério.	88
51	Classificação dos grupos de vozes utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 2º critério.	89
52	Comparação entre a metodologia utilizada proposta nessa tese (esquerda) e utilizada em Sobrinho (2011) (direita).	91

Lista de Tabelas

1	Frequência fundamental extraída usando Transformada de Fourier e seção de Poincaré	73
2	Média e desvio padrão de F_0	73
3	Média da dispersão para uma senóide.	74
4	Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo extraídos do intervalo de 0.5s a 1.0s.	78
5	Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo extraídos do intervalo de 2.0s a 2.5s.	78
6	Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo extraídos do intervalo de 4.0s a 4.5s.	79
7	Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo - trechos com maior variação de pitch.	79
8	Valores de p obtidos com o teste Anova no intervalo 0.5s - 1.0s entre as vozes normais, com edema e com nódulo.	82
9	Valores de p obtidos com o teste de <i>Tukey</i> no intervalo 0.5s - 1.0s.	82
10	Valores de p obtidos com o teste Anova no intervalo 2.0s - 2.5s entre as vozes normais, com edema e com nódulo.	83
11	Valores de p obtidos com o teste de <i>Tukey</i> no intervalo 2.0s - 2.5s.	83
12	Valores de p obtidos com o teste Anova no intervalo 4.0s - 4.5s entre as vozes normais, com edema e com nódulo.	84
13	Valores de p obtidos com o teste de <i>Tukey</i> no intervalo 4.0s - 4.5s.	84
14	Valores de p obtidos com o teste Anova utilizando o 2° critério entre as vozes normais, com edema e com nódulo.	88
15	Valores de p obtidos com o teste de <i>Tukey</i> utilizando o 2° critério para as vozes tomadas duas a duas.	88

Sumário

1	INTRODUÇÃO	21
1.1	Contribuição	23
2	FORMAÇÃO DA VOZ	25
3	VOZ NORMAL E PATOLÓGICA	33
3.1	Nódulo	34
3.2	Edema de Reinke	35
4	TÉCNICAS DE DETECÇÃO DE PATOLOGIAS	37
4.1	Método clássico	37
4.1.1	Avaliação perceptiva-auditiva	37
4.1.2	Medidas acústicas	38
4.2	Análise não linear	40
4.2.1	Reconstrução do espaço de fase	41
4.2.1.1	Tempo de atraso - τ	42
4.2.2	Entropia	45
4.2.3	Bifurcação	46
4.2.4	Dimensão de correlação	47
4.2.5	Expoente de Lyapunov - λ	48
4.2.6	Seção de Poincaré	50
4.3	Outras técnicas	52
5	METODOLOGIA PROPOSTA PARA EXTRAÇÃO DA SEÇÃO DE	

POINCARÉ DE UM SINAL DE VOZ	55
5.1 Procedimento para extrair a seção de Poincaré	56
5.1.0.1 Cálculo da frequência fundamental (F_0) através da seção de Poincaré	70
5.2 Validação da ferramenta utilizando senóides	73
6 RESULTADOS E DISCUSSÕES	75
6.1 Testes estatísticos	77
6.1.1 Análise com 1° critério	82
6.1.1.1 Intervalo: 0.5s - 1.0s	82
6.1.1.2 Intervalo: 2.0s - 2.5s	83
6.1.1.3 Intervalo: 4.0s - 4.5s	84
6.1.1.4 Análise do 1° critério utilizando <i>Support Vector Machine</i> (<i>SVM</i>)	85
6.1.2 Análise com 2° critério	87
6.1.2.1 Análise do 2° critério utilizando <i>Support Vector Machine</i> (<i>SVM</i>)	88
6.1.3 Comparação entre metodologias e Contribuição	89
7 CONCLUSÃO E TRABALHOS FUTUROS	93
REFERÊNCIAS BIBLIOGRÁFICAS	95

1 Introdução

Os primeiros estudos sobre produção de voz datam de meados do século XVIII quando os pesquisadores tentavam desvendar os órgãos responsáveis pela formação da voz. Em 1700, Dodart (n.d.) provou que a produção das vogais tinham origem na laringe. Ele acreditava que a única função da laringe e das pregas vocais era variar a área da glote e que "redemoinhos" na glote geravam o som. Em 1741, Ferrein (n.d.) realizou experimentos com animais e provou que a vibração das pregas vocais eram essenciais para produção de voz. Helmholtz (1865), mostrou que fluxos na forma de rajadas de ar escapando através da glote eram a fonte primária da produção do som. Em 1954, Van Den Berg desenvolveu a teoria aerodinâmica mioelástica da produção de voz e, em 1993, foi introduzida a teoria do Caos para explicar o fenômeno da produção da voz (TITZE; BAKEN; HERZEL, 1993a).

Enquanto os pesquisadores tentavam explicar o processo de formação da voz, equipamentos eram desenvolvidos para visualizar o sinal de voz graficamente. Por volta de 1920 foi inventado o oscilograma, onde se verificava a leitura da amplitude dos sinais de voz em função do tempo, e a partir da década de 1940, com a invenção do espectrógrafo, passou a ser possível visualizar o espectro de frequências dos sinais de voz em função da amplitude, bem como a frequência fundamental e frequências formantes da voz. Em 1961, Lieberman propôs o primeiro parâmetro acústico e a partir da década de 1970, vários trabalhos foram desenvolvidos utilizando a análise acústica para detecção de patologias (ARROYAVE; BONILLA; TREJOS, 2012).

As técnicas típicas de diagnóstico de patologias de voz são consideradas invasivas pois utilizam instrumentos introduzidos dentro da boca ou nariz até obter uma visualização das pregas vocais. Essas técnicas causam desconforto no paciente que necessitam de anestésico local para evitar a reação de náusea, além de utilizar equipamentos sofisticados e de alto custo, mas permitem a visualização direta da laringe e das pregas vocais. Algumas dessas técnicas básicas são a laringoscopia indireta por espelho circular, laringoscopia direta tradicional e laringoscopia indireta com fibra ótica (BEHLAU, 2001).

Utilizando as técnicas de processamento digital de sinais, especificamente, sinais de voz, as análises passaram a ser não invasivas, onde o sinal de voz é coletado por um profissional treinado e armazenado para processamento posterior. Através de *softwares* específicos vários parâmetros passaram a ser determinados a partir do sinal no tempo, como variação do período ciclo a ciclo (*jitter*), variação em amplitude ciclo a ciclo (*shimmer*),

frequência fundamental (F_0), dentre outros; e os parâmetros no domínio da frequência como o nivelamento espectral de ruído, nivelamento espectral do filtro, relação harmônico ruído e energia normalizada de ruído (SCALASSARA, 2009). Com base nesses parâmetros, vários estudos foram feitos para detecção de patologias na voz, como pode ser visto em Lieberman (1963), Yumoto, Gould e Baer (1982), Kasuya et al. (1986), Ludlow et al. (1987), Yunik e Boyanov (1990), Kasuya, Endo e Saliu (1993) e Vieira, McInnes e Jack (1997). No entanto, Brockmann-Bauser e Drinnan (2011) questionaram o uso dos parâmetros *jitter* e *shimmer* como indicadores absolutos de patologia vocal ou rouquidão.

As técnicas para extração de parâmetros no domínio do tempo e frequência são baseadas no modelo linear fonte - filtro, proposto por Fant (1970). No entanto, Titze, Baken e Herzel (1993a) mostraram que a vibração das pregas vocais é um fenômeno não linear e em 1995 classificaram os sinais de voz em três tipos de acordo com a sua periodicidade: (i) sinais quase periódicos, (ii) sinais contendo modulações e subharmônicas, e (iii) sinais irregulares e aperiódicos. Sob esse enfoque, novos parâmetros foram obtidos para diagnóstico e detecção de patologias, como pode ser visto em Pitsikalis, Kokkinos e Maragos (2003) - que usaram medidas de dimensão fractal e expoente de Lyapunov, Scalassara et al. (2009) - que extraíram medidas de entropia, Matassini et al. (2000) - que utilizaram medidas de entropia, dimensão de correlação e expoente de Lyapunov, Alonso et al. (2005b) - que utilizaram dimensão de correlação e expoente de Lyapunov, Sobrinho (2011) - que extraiu parâmetros a partir da seção de Poincaré e Dajer et al. (2011) - que estabeleceram parâmetros visuais do espaço de fase, dentre outros trabalhos.

Os mais recentes estudos para diagnósticos e classificação de patologias envolvem redes neurais, inteligência artificial, transformada *wavelet* e *support vector machine*, além de patentes para diagnóstico desenvolvidas na última década (ARROYAVE; BONILLA; TREJOS, 2012).

Através dessa técnica de análise não linear é possível extrair medidas quantitativas que possam diferenciar vozes normais das vozes patológicas. Devido à formação de trajetórias bem definidas no espaço de fase de vozes normais e patológicas, essa técnica torna-se útil e eficaz para quantificar os padrões geométricos obtidos.

Na literatura há poucos trabalhos relacionados à extração da seção de Poincaré. A técnica proposta em Kubin (1997) para extrair a seção de Poincaré usa uma transformação de coordenadas locais para detectar os pontos de interseção da trajetória com a seção. Especificamente, essa técnica utiliza novas coordenadas que são obtidas através de decomposição em valores singulares de uma matriz cujas colunas são os pontos de interseção da

trajetória com a seção. Essa técnica não oferece nenhuma medida quantitativa sobre a dispersão dos pontos na seção mas fornece uma alta resolução, análise multidimensional do sinal de voz e o desvio de sua periodicidade. No trabalho desenvolvido em Hegger, Kantz e Schreiber (1999), a seção de Poincaré foi construída tomando os pontos onde a derivada é zero no sinal temporal, o que corresponde a tomar os pontos de máximo e mínimo do sinal temporal.

Ressalta-se que as técnicas não invasivas de análise de voz (acústica e não linear) são auxiliares à técnica perceptivo-auditiva e não as descartam. Ainda, as medidas extraídas através das técnicas de análise não linear são complementares às técnicas de análise acústica e não se pode afirmar, até o momento, que uma se sobreponha à outra (ARROYAVE; BONILLA; TREJOS, 2012).

Essa tese está dividida da seguinte forma: o Capítulo 2 aborda o processo de formação da voz. O Capítulo 3 aborda voz normal e patológica, especificamente, pregas vocais com edema e nódulo. O Capítulo 4 descreve sobre as técnicas de detecção de patologias através da abordagem clássica e as técnicas de análise não linear, com destaque para a seção de Poincaré, técnica de detecção de patologia estudada nessa tese. O Capítulo 5 descreve em detalhes a técnica de extração da seção de Poincaré, o cálculo da frequência fundamental (F_0) e a validação da técnica proposta. O Capítulo 6 traz os resultados e discussões sobre a técnica proposta para detecção de patologias. O capítulo 7 traz, respectivamente, a conclusão e as propostas de trabalhos futuros.

1.1 Contribuição

A contribuição deste trabalho é o desenvolvimento de uma ferramenta baseada em dinâmica não linear capaz de detectar patologias laríngeas como nódulos e edemas. A técnica utilizada é a determinação da seção de Poincaré do atrator tridimensional reconstruído do sinal de voz em todo o espaço de fase. Cada seção de Poincaré é determinada em cada ponto da trajetória média do atrator e, em seguida, é calculada a média da dispersão dos pontos em cada seção. Dessas medidas extrai-se a média final da dispersão dos pontos dos 3 tipos de vozes estudadas que são comparadas entre si utilizando o teste Anova para diferenciação entre os 3 grupos e o teste de Tukey, para diferenciação entre os grupos de vozes tomadas duas a duas. A técnica de extração da seção de Poincaré proposta nessa tese foi apresentada no Congresso de Dinâmica e Controle - Dincon (2015), na cidade de Natal - RN. O caráter inovador dessa tese em comparação ao que foi realizado em Sobrinho (2011) é extrair a seção de Poincaré em todo espaço de fase, inclusive em trechos do

espaço de fase que eram inacessíveis utilizando o método anterior.

Uma vantagem do algoritmo apresentado nesse trabalho em relação a Kubin (1997) é a garantia de que somente um ponto estará associado a cada cruzamento da trajetória na seção de Poincaré, enquanto que em Kubin (1997) são tomados os pontos na vizinhança do plano de Poincaré. Em relação ao que foi proposto em Hegger, Kantz e Schreiber (1999), o algoritmo proposto é capaz de (i) extrair múltiplas seções de Poincaré ao longo de todo o espaço de fase e (ii) extrair o espaço de fase médio. Hegger, Kantz e Schreiber (1999) também afirmam que a correta posição da seção de Poincaré é de alta relevância para a utilidade dos resultados obtidos. O algoritmo proposto elimina esse problema pois extrai múltiplas seções de Poincaré através de cortes perpendiculares ao longo da trajetória.

Outra importante contribuição da técnica proposta é que através dela é possível determinar da frequência fundamental F_0 de um sinal de voz. A frequência fundamental foi calculada pela média do tempo de retorno da trajetória na seção de Poincaré (SOBRINHO; DAJER; ALBERTO, 2016). Outra contribuição é que o método proposto nessa tese possibilita visualizar o *jitter* de um sinal de voz.

2 Formação da voz

O processo de formação da voz começa quando se deseja transmitir uma mensagem falada para um ouvinte. Para que isso ocorra, o cérebro articula o que se deseja transmitir e, em conjunto com os músculos responsáveis pela fala, produz uma onda acústica a ser interpretada pelo ouvinte. Do ponto de vista físico, a voz é produzida pela vibração das pregas vocais localizadas na laringe e modificada pelas cavidades localizadas abaixo e acima dela, chamadas de cavidades de ressonância. A Figura 1 mostra um corte médio sagital do sistema fonatório indicando os elementos que estão envolvidos no processo de formação da voz: pulmões, laringe, cavidade faríngea, cavidade oral e nasal. A cavidade faríngea e oral formam o trato vocal, limitado pela laringe e os lábios; e a cavidade nasal, limitada pelo velum e narinas. Uma análise mais refinada dos elementos que envolvem a formação da voz incluem as pregas vocais, a língua, dentes, queixo e traquéia.

Figura 1: Sistema fonatório. Modificado de (DELLER; PROAKIS; HANSEN, 2000).

A produção dos sons é feita pela vibração das pregas vocais localizadas na laringe sendo radiada e filtrada através desse trato quando o ar é expelido pelos pulmões. O fluxo de ar resultante é perturbado pelo estreitamento ou abertura da região do trato vocal, envolvendo a cavidade bucal. A cavidade bucal atua como um filtro ampliando ou atenuando as frequências produzidas durante a fala. Além da função fonação, a laringe também tem as funções de respiração e deglutição, não abordadas nesse trabalho, mas discutidas em Behlau (2001). A Figura 2 mostra a posição das pregas vocais para a produção do som e respiração. No processo de deglutição, a traquéia fecha-se totalmente num processo involuntário utilizando a epiglote para evitar a entrada de alimentos na traquéia.

Figura 2: Esboço da laringe, vista de cima, na posição de fonação e respiração. Adaptada de <http://www.phys.unsw.edu.au/jw/voice.html>.

A mecânica da produção da voz começa com os músculos das pregas vocais inicialmente em repouso. Uma pressão vinda diretamente dos pulmões age abaixo da glote forçando os músculos das pregas vocais a se afastarem um do outro num movimento acelerado, conforme ilustrado na Figura 3 (a) e (b). A prega vocal começa a se abrir (Figura 3 (c)). Conforme o fluxo de ar passa por entre as pregas vocais (Figura 3 (d)), a pressão local é reduzida conforme o efeito de *Bernoulli*, e força as pregas vocais a voltarem à sua posição inicial (Figura 3 (e) e (f)). Uma vez que as pregas estão novamente próximas uma da outra, o fluxo de ar diminui enquanto que a pressão subglotal estática aumenta. A partir daí o ciclo é novamente repetido obtendo-se a onda de mucosa.

Usando o conceito de conservação de energia, a energia cinética é representada como a velocidade do ar (através da glote) elevada ao quadrado; e a energia potencial é representada pela pressão do ar. Conforme as pregas vocais se afastam, a velocidade do

Figura 3: Sequência de um ciclo completo da fonação (DELLER; PROAKIS; HANSEN, 2000).

ar aumenta causando uma queda de pressão local na glote. Portanto, quando as pregas vocais estão fechadas, a pressão do ar e a energia potencial são altas.

A glote continua a se abrir até que a tensão elástica natural se iguale à força de separação devido à pressão do ar. Nesse ponto, a abertura glotal e o fluxo de ar através da glote atingiram o máximo. A energia cinética que foi recebida pelas pregas vocais durante a abertura é armazenada na forma de energia potencial elástica, que por sua vez, forçam as pregas vocais a se fecharem, conforme ilustrado na Figura 3 (e). O efeito de sucção causado pela força de *Bernoulli* ocorre quando as pregas estão suficientemente próximas. O fechamento abrupto ocorre devido à ação conjunta da força de *Bernoulli* e a força e restauração elástica (DELLER; PROAKIS; HANSEN, 2000).

Esse movimento das pregas vocais determina a frequência fundamental do sinal de voz que é da ordem de 125 Hz para o homem e 250 Hz para as mulheres (BEHLAU, 2001). O mecanismo de vibração das pregas vocais permite um pulso glotal quase periódico suficiente para excitar o sistema acústico acima das pregas vocais. A Figura 4 mostra a forma de onda de um pulso glótico.

O espectro de potência do sinal glotal decresce com o aumento da frequência. Para produção da fala esse sinal é filtrado no trato vocal atenuando ou amplificando as componentes de frequência. Em seguida, num último processo de filtragem, o sinal é radiado pelos lábios, produzindo a fala do modo como a ouvimos, conforme ilustração da Figura 5.

Figura 4: Forma de onda do pulso glotal. Adaptado de Dajer (2010).

Diversas teorias foram desenvolvidas para explicar a produção de voz; algumas descritas brevemente a seguir. A teoria da corda vibrante explicava que a corrente de ar, ao passar entre as pregas vocais, provocava sua vibração e a produção do som; a teoria da palheta dizia que a laringe era comparada a um instrumento de palheta, com as pregas vocais aproximando-se temporariamente e separando-se em seguida, permitindo a passagem de ar num fluxo intermitente que produzia o som; a teoria neurocronática, que apresentou a laringe com um sistema de ativação periódica onde a fonação era apresentada como um fenômeno neuromuscular ativo, e os movimentos periódicos das pregas vocais originados por estímulos do nervo laríngeo na mesma frequência do som emitido, controlados por um centro cerebral que regularia a velocidade de vibração de cada uma das pregas vocais, ciclo a ciclo. Essa teoria foi descartada, pois pregas vocais com paralisia produzem som e até mesmo cadáveres são capazes de produzir som quando um fluxo de ar atravessa sua laringe. Uma importante contribuição para o processo de formação da voz foi a teoria mioelástica-aerodinâmica de Van Den Berg (1954). Essa teoria estabeleceu uma combinação entre as forças elásticas dos músculos da laringe e as forças físicas aerodinâmicas da respiração para produção da voz. A mais recente teoria para formação da voz utiliza conceitos de análise não linear e teoria do caos. Essa teoria foi desenvolvida por Titze, Baken e Herzel (1993a), que defendem que a produção da voz é um fenômeno não linear, determinístico, imprevisível, de baixa dimensão e sensível às condições de controle da voz (BEHLAU, 2001).

Sob o ponto de vista da engenharia, a produção da voz pode ser imaginada como um processo de filtragem acústica (DELLER; PROAKIS; HANSEN, 2000). Sob esse enfoque, o sistema vocal pode ser representado por um modelo simplificado, indicando as 3 principais cavidades responsáveis pela formação e filtragem da voz (faríngea, oral e nasal), mas foi Flanagan (1972) que inicialmente interpretou o sistema vocálico como um sistema fonte-filtro sendo o conjunto pulmão + traquéia a fonte de energia; e o filtro, o trato vocal, ilustrado na Figura 6.

Figura 5: Espectro de frequências do sinal de voz. Modificado de Scalassara (2009).

Diversas técnicas são utilizadas para simulação de um sinal vocálico. Rabiner e Schafer (2007) propuseram um modelo linear variante no tempo indicado na Figura 7, chamado "modelo sistema/fonte".

Nesse modelo, o sistema linear discreto e variante no tempo à direita da Figura 7 simula a envoltória de frequências do trato vocal. O gerador de excitação, à esquerda da Figura 7, simula os diferentes modos de produção de som no trato vocal. Nesse caso, as amostras do sinal de voz são resultantes de um sistema linear variante no tempo. Para intervalos da ordem de 10ms, pode-se assumir o sistema como linear e invariante no tempo, cuja função de transferência na transformada Z , que relaciona a saída $S[z]$ e a entrada $E[z]$, pode ser representada pela expressão (2.1):

$$H(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{1 - \sum_{k=1}^N a_k z^{-k}} = \frac{b_0 \prod_{k=1}^M (1 - d_k z^{-1})}{\prod_{k=1}^N (1 - c_k z^{-1})} \quad (2.1)$$

Figura 6: Modelo fonte-filtro proposto por Fant (1970)

Figura 7: Modelo sistema/fonte do sinal de voz. Adaptado de Rabiner e Schafer (2007).

onde os coeficientes a_k e b_k , chamados de parâmetros do trato vocal, mudam a uma taxa de 50 à 100 vezes por segundo. Os coeficientes c_k são os polos da função e alguns desses polos estão próximos ao círculo unitário criando as frequências de ressonância (formantes). Os zeros d_k são usados para modelar os sons nasais e fricativos, mas muitas das aplicações do modelo sistema/fonte usam um modelo de função de transferência utilizando apenas os polos da função. Isso permite simplificar as análises necessárias para estimar os parâmetros do trato vocal tudo-polo para o sinal de voz (DELLER; PROAKIS; HANSEN, 2000).

As técnicas de análise de voz utilizam sinais de vogais sustentadas, mas alguns estudos também sugerem o uso de sinais não vozeados (sons fricativos) para análise (NARAYANAN; ALWAN, 1995). Por esse motivo, será feita a seguir uma breve descrição dos diversos tipos de sons produzidos pela fala humana.

Os sons da fala são gerados de diversas maneiras. Sons "sonoros", ou vocálicos, são produzidos quando o trato vocal é excitado por pulsos glóticos quase periódicos através das pregas vocais. Sons "surdos" são produzidos através de uma constrição em alguma parte do trato vocal, forçando a passagem do ar através dessa constrição, criando um fluxo turbulento e ruidoso de ar. Exemplos desse tipo de som são os fonemas /x/ e /s/. Uma outra forma de produção de som ocorre quando o trato vocal é parcialmente fechado, causando um fluxo turbulento de ar devido à constrição e ao mesmo tempo permitindo um fluxo quase periódico devido à vibração das pregas vocais. Exemplo desse tipo de som são os fonemas /v/ e /z/. Os sons plosivos resultam do fechamento completo e momentâneo, da boca por exemplo, criando uma pressão por trás dos lábios e abertura abrupta. Exemplos desse tipo de sons são os fonemas /t/, /k/ e /p/. Todas essas formas de produção de som criam um sinal de excitação de banda larga, onde o trato vocal atua como uma linha de transmissão acústica, realçando algumas frequências e atenuando outras, para o som finalmente ser irradiado através dos lábios.

3 Voz normal e patológica

No contexto da fonoaudiologia, uma voz normal se caracteriza quando é agradável e ausente de ruído, com frequência adequada à idade e sexo do falante, com intensidade apropriada e com variações de frequência e intensidade que expressem as emoções e sentimentos do indivíduo (BEHLAU, 2001). Nesse caso, a classificação da voz entre normal ou patológica é feita através de técnicas de percepção de um profissional treinado, como será visto adiante. No contexto da bioengenharia, parâmetros são extraídos e comparados com uma base de dados que contém informações de vozes normais e patológicas previamente analisadas por exame estroboscópico ou pela análise perceptivo-auditiva. Como exemplo, uma voz normal produzirá um *jitter* menor do que uma voz patológica. Montagnoli e Pereira (2009) consideram que porcentagens abaixo de 0,35% de *jitter* e menores do que 2,25% de *shimmer* são valores para vozes sem patologia. Outros estudos mostram que uma voz patológica tem uma dimensão de correlação maior do que uma voz normal (ALONSO et al., 2005b).

As patologias em pregas vocais surgem devido à acidentes, doenças, mal uso da voz ou cirurgias afetando as pregas vocais e têm um profundo impacto na vida dos pacientes (LITTLE et al., 2006), sendo a rouquidão o primeiro sintoma da manifestação de uma patologia. De modo geral, elas são classificadas como patologias funcionais, quando ocorrem devido ao mau uso da voz, ou organofuncionais, quando ocorrem mudanças na anatomia das pregas vocais. As patologias em pregas vocais causam aumento de componentes ruidosas no sinal, alteração na frequência fundamental, amplitude, qualidade vocal, além do aparecimento de componentes de frequência sub-harmônicas (DELIYSKI, 1993);(DAVIS, 1979). Em casos mais graves de patologias, como câncer, toda a laringe pode ser extraída, incluindo as pregas vocais, comprometendo totalmente o mecanismo de produção vocal (HAGMÜLLER, 2009).

De acordo com Behlau (2001), as 3 principais lesões organofuncionais são o pólipos, o nódulo e o edema de Reinke, os 2 últimos analisados nesse trabalho por pertencerem ao banco de dados disponível para essa tese. Essas disfonias são caracterizadas por lesões benignas nas pregas vocais devido ao mau uso da voz e que usualmente são diagnosticadas tardiamente. A seguir uma breve descrição das patologias estudadas nesse trabalho é apresentada.

3.1 Nódulo

Os nódulos são lesões de massa e bilaterais de cor esbranquiçada que se desenvolvem na região anterior das pregas vocais, na metade de maior vibração glótica, que são causadas por abuso vocal e é considerada a lesão laríngea mais comum. Consistem principalmente de tecido edematoso e/ou fibras colágenas. Essas lesões são comuns em mulheres jovens adultas, na faixa etária de 25 a 35 anos, e também em crianças de ambos os sexos.

Esse tipo de lesão nas pregas vocais pode causar um movimento de comportamento caótico entre as pregas (JIANG; ZHANG; MCGILLIGAN, 2006). A rouquidão e sopro são os principais sinais da avaliação perceptivo-auditiva que indicam a presença de nódulos (BEHLAU, 2001). A Figura 8 (esquerda) mostra um nódulo nas pregas vocais e a Figura 8 (direita), mostra uma prega vocal saudável para comparação.

Figura 8: Nódulos bilaterais indicado pelas setas (MARTINS; TRINDADE, 2003) (Esquerda). Prega vocal saudável (extraído de voicedoctor.net) (Direita).

De acordo com Yanagihara (1967), a análise espectrográfica de um sinal de voz com rouquidão mostrou componentes ruidosas nos formantes principais de várias vogais, componentes de ruído de altas frequências e perda de componentes harmônicas de alta frequência. A Figura 9, desenvolvida no *Matlab - versão student*, mostra um sinal de voz normal típico, à esquerda, mais rico em componentes de frequência do que uma voz com nódulo, à direita.

O grau de rouquidão ou sopro depende do tamanho do nódulo e da rigidez de seus tecidos. A rouquidão correlaciona-se ao grau de irregularidade de vibração das pregas vocais, enquanto que a sopro refere-se ao fechamento glótico incompleto. Alguns estudos relacionados à detecção de nódulo podem ser vistos em Campisi et al. (2000), Jiang et al. (2009) e Birkent et al. (2004).

Figura 9: Espectro de frequência de um sinal de voz normal e voz com nódulo. A voz normal, apresentada à esquerda, é mais rica em componentes de frequência quando comparada à voz com nódulo, apresentada à direita.

3.2 Edema de Reinke

De acordo com Behlau (2001), o edema de Reinke é uma lesão difusa na camada superficial da prega vocal, bilateral e assimétrica, de coloração rosada, caracterizada por acúmulo de fluido, de modo irregular, em alguma região da porção membranosa ou nela toda. Ocorre em indivíduos adultos de ambos os sexos, com idade entre 45 e 65 anos, que apresentam uma frequente associação de uso intensivo da voz, abusos vocais e tabagismo.

O edema ocorre por reação natural do tecido ao trauma fonatório associado ao consumo de cigarro, conforme verificado em Matsuo, Kamimura e Hirano (1983). De acordo com Neves, Neto e Pontes (2004), o aspecto extenso do edema é uma característica importante para distingui-lo dos pólipos. O abuso vocal também é responsável pelo aparecimento desta lesão, mas não é uma etiologia tão importante como o é para o nódulo vocal, que é considerado típico do trauma vocal. A Figura 10 mostra uma prega vocal com edema de Reinke.

O edema de Reinke mostra uma variedade de modelos clínicos e diferenças em análise de voz dependendo de seu tipo. A severidade do edema é verificada através da eletroglografia, que também mostra a evolução pré e pós cirurgia (LIM et al., 2006). Zaleska-Krecicka, Krecicki e Cyganek (1993) afirmam que por meio de micro-cirurgia pode-se melhorar a qualidade vocal em quase todos os casos de edema de Reinke. A melhora da qualidade vocal está relacionada às características perceptivas típicas que são voz grave para a idade e sexo do paciente e rouquidão.

Figura 10: Edema de Reinke na prega vocal. Fonte: <http://www.fonologo.es>

4 Técnicas de detecção de patologias

Anteriormente, a análise e o diagnóstico de voz eram feitos através da técnica perceptiva-auditiva que dependia de treinamento e experiência do profissional da área de fonoaudiologia. Com o avanço da computação digital e das técnicas de processamento digital de voz, foi possível aumentar a precisão dos diagnósticos em laringologia (ARAÚJO et al., 2002). Deve-se ressaltar que todas as técnicas aplicadas (acústica e não linear) servem como auxílio ao profissional da área de voz, e não o substitui. Através do processamento do sinal de voz, são extraídos parâmetros do sinal completo ou segmentos dele que passam por um classificador; este por sua vez produz uma resposta com base nos modelos existentes de patologias ou voz normal. A seguir será feita uma revisão das técnicas existentes (linear e não linear) e dos parâmetros comumente utilizados para análise e diagnóstico de voz.

4.1 Método clássico

4.1.1 Avaliação perceptiva-auditiva

A análise perceptiva-auditiva se baseia no julgamento que um avaliador treinado faz a respeito das características de uma voz. Embora questionada por ser uma avaliação subjetiva, a avaliação perceptiva-auditiva é uma ferramenta fundamental na avaliação vocal e faz parte da prática dos profissionais da área de fonoaudiologia. Uma desvantagem da análise perceptiva-auditiva é que especialistas podem divergir quanto às suas opiniões e vários avaliadores são necessários para se obter uma avaliação confiável (AS-BROOKS et al., 2006).

As características avaliadas por percepção são a rouquidão, aspereza e soproidade, às quais são atribuídas um valor de 0 a 3, de acordo com a escala GRBAS (acrônimo de *Grade, Roughness, Breathness, Asthenicity* e *Strained*) ou a escala RASAT, acrônimo de Rouquidão, Aspereza, Soproidade, Astenia e Tensão (PINHO; PONTES, 2002)), que é uma adaptação da escala GRBAS. Na escala GRBAS, o valor 0 é atribuído a uma voz normal, 1 para uma voz com disfonia leve, 2 para voz com disfonia moderada e 3 para extremo. A avaliação perceptiva também pode fornecer informações de não linearidades num sinal de voz (HERZEL et al., 1994a). A rugosidade intermitente, por exemplo, pode indicar frequências subharmônicas. Se essas frequências subharmônicas estiverem abaixo

de 70Hz, o sinal vocálico é percebido como crepitante.

4.1.2 Medidas acústicas

O sinal de voz é variante no tempo uma vez que as características do sistema que o produz mudam ao longo do tempo. Por isso, as medidas acústicas são tomadas em pequenas porções do sinal, para que as medidas sejam avaliadas sob o conceito de um sistema invariante no tempo (DELLER; PROAKIS; HANSEN, 2000).

As medidas acústicas extraídas de um sinal de voz podem trazer informações auxiliares às características acústicas obtidas pela avaliação perceptiva (AS-BROOKS et al., 2006). Essas medidas são extraídas a partir de técnicas lineares como transformada discreta de Fourier ou do próprio sinal no tempo (TALKIN, 1995). Algumas dessas medidas são o *jitter* (perturbação ou variabilidade da frequência fundamental ciclo a ciclo), *shimmer* (perturbação ou variabilidade da amplitude ciclo a ciclo), frequência fundamental (número de vibrações por segundo produzidas pelas pregas vocais); e as medidas de ruído que quantificam o ruído originado na turbulência do ar ao nível da glote como nível de ruído espectral, relação harmônico ruído e energia de ruído normalizada.

Rosa, Pereira e Grellet (2000) estudaram 21 tipos de patologias através de medidas acústicas. O estudo foi feito através da filtragem inversa do sinal de voz utilizando filtro de Kalman e filtro de Wiener onde foi estimado o resíduo do sinal e 7 parâmetros acústicos: resíduo médio quadrático, coeficiente de excesso, nivelamento espectral do resíduo, nivelamento espectral do filtro AR, *jitter*, pico (média das maiores diferenças do tamanho do sinal residual) e pico de autocorrelação. Os resultados através do teste *Mann-Whitney* indicaram que o melhor parâmetro para discriminação de patologias foi o *jitter*, com 54,79% de acertos para discriminação de patologias tomadas duas a duas.

A Figura 11(direita) mostra o *jitter* quando se sobrepõe 5 ciclos completos de um sinal de voz. A variação da frequência (ou período) do sinal está indicada nos diferentes pontos onde cada ciclo de cada sinal sobreposto "corta" o eixo x. A Figura 11(direita) é uma ampliação para melhor visualização do *jitter*. Essa figura foi um dos resultados obtidos através do método proposto nessa tese.

De acordo com Zhang et al. (2004), o *jitter* pode ser calculado através da expressão (4.1):

$$j = \frac{[1/(K-1)] \sum_{i=1}^{K-1} |T_0^{(i)} - T_0^{(i+1)}|}{(1/K) \sum_{i=1}^K T_0^{(i)}} \quad (4.1)$$

e $i = 1, 2, \dots, K$ sendo K o número de períodos de *pitch* e $T_0^{(i)}$ os períodos de *pitch* extraídos do sinal de voz.

Figura 11: Os diferentes pontos de cruzamento no eixo x de cada ciclo do sinal de voz sobrepostos ilustram o jitter. A figura da direita é uma ampliação da figura à esquerda para melhor visualização (Fonte própria).

Araújo et al. (2002) normatizaram os parâmetros acústicos para obtenção de parâmetros da voz normal. Foram obtidos os valores médios de frequência fundamental, *jitter*, *shimmer*, *spectral noise level*, *harmonic-to-noise ratio*, *harmonic-to-noise ratio cepstrum*, *normalized noise energy* e *breathiness ratio*. Nesse trabalho verificou-se a necessidade de se normalizar as medidas acústicas para compreensão da voz normal. Essas medidas também podem ser utilizadas por uma rede neural que classifica as vozes como normal ou patológica como pode ser visto em Jesus, José e Miguel (2001) e (BOYANOV; HADJITODOROV, 1997)). A desvantagem dessas técnicas é o alto custo computacional para o treinamento adequado da rede neural (LITTLE et al., 2006).

Yumoto, Gould e Baer (1982) utilizaram a relação harmônico ruído para avaliação quantitativa do grau de rouquidão. Anteriormente, essa medida era subjetiva e se baseava na extensão de ruído que se sobrepunha ao espectrograma de uma vogal sustentada. Nesse trabalho, 50 períodos de *pitch* de uma vogal sustentada foram analisados; em seguida foi calculada a energia H da forma de onda e a energia média N, tomada das diferenças de energia dos períodos individuais e da forma de onda média. Foram extraídas

medidas de ruído relativo a cada componente harmônica de 42 vozes normais e 41 vozes com algum grau de rouquidão através do espectrograma. Os resultados mostraram uma concordância de 84,9% entre as duas medidas mostrando ser um parâmetro eficaz para medir rouquidão. Na detecção automática de patologia apresentaram uma taxa de acerto de 83,3%. Num outro estudo, Krom (1993) mediu a relação harmônico ruído utilizando a técnica denominada *cepstrum* e mostrou ser uma medida eficaz para análise da qualidade vocal. Por outro lado, Morente et al. (2001) relataram que esse não é um parâmetro sensível para diferenciar uma voz disfônica de uma voz normal.

Kasuya et al. (1986) propuseram uma medida de energia de ruído normalizada - (NNE - *Normalized Noise Energy*) como uma medida acústica para avaliar ruídos em vozes patológicas devido ao fechamento incompleto da glote. NNE é a relação entre a energia do ruído e a energia total do sinal, medida em decibéis (dB), e foi eficaz na detecção câncer na glote, paralisia no nervo recorrente e nódulo nas cordas vocais. Através dessa medida detectou-se câncer glótico com 88,8% de acerto sobre a base de dados estudada.

Michaelis, Gramss e Strube (1997) posteriormente mediram a *Glottal to Noise Excitation Ratio* para descrever patologia vocal. Os parâmetros dessa medida são baseados no coeficiente de correlação do envelope de Hilbert e indicam se o sinal é originado da vibração da prega vocal, ou gerado do ruído turbulento do trato vocal, indicando sopro.

Dibazar, Narayanan e Berger (2002) coletaram o sinal de voz da vogal sustentada 'a', extraíram os coeficientes cepstrais de frequência da escala mel (MFCC) e através de um classificador baseado nos modelos ocultos de Markov (HMM) detectaram patologias com uma taxa de 98,59% de acerto. Escala "MEL" é uma escala criada para levar em consideração a percepção não linear do som pelo ouvido humano.

4.2 Análise não linear

Classicamente, estuda-se a acústica da voz humana através de um modelo linear de produção de voz proposto por Fant (1960) e Davis (1979). Nesse modelo, o sinal de voz é o resultado de uma excitação periódica filtrada pela glote, trato vocal e os lábios. No entanto, padrões de vibrações e imprevisibilidade existentes em vozes patológicas (SCALLASSARA, 2009) são melhores compreendidos se analisados por intermédio de um modelo de laringe não linear de baixa dimensão, pois mesmo os mais simples modelos de pregas vocais incorporam não linearidades (ISHIZKA; FLANAGAN, 1972) e exibem dinâmica caótica de baixa dimensão (HERZEL; WENDLER, 1991). Little et al. (2006) também afir-

mam que a complexidade de um sinal vocal é um fator limitante da técnica de análise linear.

Deliyski (1993) desenvolveu um modelo acústico que descreve os efeitos de não linearidade que ocorrem na forma de onda acústica em vozes patológicas como irregularidades na frequência e amplitude. A partir desse modelo foram extraídos 33 parâmetros acústicos típicos de avaliação (frequência média fundamental, relação harmônico ruído, componentes sub-harmônicas, dentre outras) e comparados com 68 voluntários com vozes normais e patológicas para se definir os limiares de vozes normais e patológicas. A motivação para se construir modelos para produção de voz é a possibilidade de controle dos parâmetros que envolvem o sinal de voz, permitindo a produção e estudo de sinais com irregularidades características de vozes patológicas (JIANG; ZHANG; MCGILLIGAN, 2006).

Diversas medidas da dinâmica não linear, utilizadas para discriminação de vozes normais e patológicas, são descritas a seguir.

4.2.1 Reconstrução do espaço de fase

Os métodos de análise não linear em séries temporais são baseados na teoria dos sistemas dinâmicos, onde a evolução da série no tempo é definida no espaço de fase. Os sistemas dinâmicos que modelam a acústica da voz humana podem ser constituídos por um conjunto de equações ordinárias de primeira ordem.

O espaço de fase pode ser considerado como um espaço de dimensão finita $\mathbb{R}^m$ onde o estado é especificado por um vetor $x \in \mathbb{R}^m$. A dinâmica do sistema é descrita por um sistema de m equações diferenciais de primeira ordem contínuo (ou discreto) de acordo com a equação (4.2):

$$\frac{dx(t)}{dt} = f(x(t)), \quad t \in R \quad (4.2)$$

A solução, $x(t)$, da equação (4.2) é chamada de trajetória do sistema dinâmico, com $x(0)$ sendo a condição inicial. Considera-se que o sistema de equações diferenciais (4.2) admite solução única para cada condição inicial.

No problema de análise de voz, o modelo (4.2) e a função f não são bem conhecidos e recorre-se a uma técnica de recuperação do espaço de fase a partir das medidas temporais do sinal de voz. Para a construção do espaço de fase é necessário conhecer as variáveis dinâmicas que governam o sistema, mas nem sempre essas variáveis são conhecidas ou

estão disponíveis, como é o caso da dinâmica vocal. Num sinal de voz tem-se disponível apenas a série temporal; para esse caso, deve-se converter a série temporal em vetores de estado para reconstruir o espaço de fase utilizando a técnica de tempo de atraso (PACKARD et al., 1980).

O espaço de fase reconstruído é formado por vetores em $\mathbb{R}^m$, cujos elementos são amostras atrasadas da série temporal original conforme indicado na equação (4.3):

$$X = \{x[t], x[t - \tau], x[t - 2\tau], \dots, x[t - (m - 1)\tau]\} \quad (4.3)$$

sendo m a dimensão de imersão e τ o tempo de atraso, definido pelo mínimo do gráfico da informação mútua.

As amostras na série de tempo devem possuir um tamanho ideal para que o espaço de fase reconstruído seja equivalente ao atrator original. A garantia para que o atrator reconstruído seja semelhante ao original é estabelecida quando $m > 2D + 1$, onde D é a dimensão de *Hausdorff* do atrator (TAKENS, 1981).

4.2.1.1 *Tempo de atraso - τ*

Um parâmetro importante para reconstrução do espaço de fase é o tempo de atraso τ . Esse tempo de atraso está relacionado à série original da seguinte forma: dada uma série temporal (ou discreta) $x_1 = x(n)$, o tempo de atraso irá definir um vetor $x_2 = x(n - \tau)$. Inicialmente, imaginava-se que o tempo de atraso τ podia ser escolhido arbitrariamente. Contudo, experimentos mostraram que a qualidade do atrator reconstruído dependia da escolha correta de τ . Ainda nos estudos iniciais sobre a reconstrução do espaço de fase não havia na literatura nenhum critério para a escolha de τ .

De acordo com Fraser e Swinney (1986b), se a escolha de τ para a reconstrução do espaço de fase for muito pequena, então as trajetórias serão definidas aproximadamente por uma linha dada por $x_0(t) = x_1(t)$ e os dados estariam correlacionados. Se τ for muito grande, perde-se a correlação e os pontos no espaço de fase ficarão muito dispersos. A escolha de τ deve garantir que as amostras atrasadas do sinal no tempo sejam linearmente independentes.

Uma outra escolha de τ para garantir a independência linear é o valor na qual a função de autocorrelação passa pelo zero. A desvantagem desse método é que ele é baseado na estatística linear e não explora as características da dinâmica não linear da voz (HERZEL,

1993), (JIANG; ZHANG, 2002), (TITZE; BAKEN; HERZEL, 1993a).

A determinação da dimensão de imersão usando decomposição em valores singulares também pode ser utilizada em sistemas reais, onde há presença de ruídos. Nessa técnica, o espaço de estado é dividido em 2 dois subespaços, um contendo o sinal e o outro contendo o ruído. Eliminando o ruído, reduz-se a dimensão e remove-se os efeitos indesejados, obtendo-se o atrator reconstruído (BROOMHEAD; KING, 1986).

Todos os métodos estudados apresentaram bons resultados, mas o melhor tempo de atraso foi determinado pelo método da informação mútua (FRASER; SWINNEY, 1986a). Nesse estudo, foi reconstruído o espaço de fase da reação química chamada *Belousov-Zhabotinskii* utilizando dois critérios para escolha de τ : o primeiro mínimo da informação mútua do sistema e o primeiro zero da função de autocorrelação. Foi verificado que o critério do primeiro mínimo da informação mútua resultava em um atrator topologicamente semelhante ao original. A seguir, uma breve descrição do conceito de informação mútua é apresentada.

Dada duas variáveis $x(t)$ e $x(t - \tau)$, a informação mútua é a informação que se tem sobre o valor $x(t - \tau)$ se o valor $x(t)$ for conhecido. O valor $x(t - \tau)$ é a máxima informação obtida em relação a $x(t)$. O mesmo resultado é válido para variáveis discretas, $x(n)$ e $x(n - \tau)$. Uma expressão para determinação da informação mútua é atribuída à Shannon (1968): dado um conjunto de dados, cria-se um histograma de resolução ϵ para a distribuição de probabilidade de dados. Indicando por p_i a probabilidade do sinal assumir um valor dentro do i -ésimo intervalo do histograma, $p_{ij}(\tau)$ a probabilidade de $s(\tau)$ pertencer ao intervalo i e $s(t + \tau)$ pertencer ao intervalo j , a informação mútua para o tempo de atraso τ é dada pela equação (4.4) (HEGGER; KANTZ; SCHREIBER, 1999):

$$I_\epsilon(\tau) = \sum_{ij} p_{ij}(\tau) \ln(p_{ij}(\tau)) - 2 \sum_i p_i \ln p_i \quad (4.4)$$

Um exemplo de sistema dinâmico não linear onde as variáveis de estado são conhecidas é a equação de *Van Der Pol*, indicada pela equação (4.5).

$$\ddot{x} + 2\epsilon(x^2 - 1)\dot{x} + x = 0 \quad (4.5)$$

A série temporal desse sistema dinâmico para uma determinada condição inicial está representada na Figura 12.

Para determinar as variáveis de estado basta fazer

Figura 12: Série temporal da equação de *Van Der Pol*.

$$\dot{x} = y \quad (4.6)$$

$$\ddot{x} = -2\epsilon(x^2 - 1)y - x \quad (4.7)$$

O espaço de fase pode ser reconstruído a partir da série temporal. Nesse caso, determina-se a gráfico da informação mútua, indicado na Figura 13, onde observa-se que o primeiro mínimo ocorre na amostra de número 21. O espaço de fase reconstruído em 2 dimensões toma o eixo x com o sinal original $x(t)$ e o eixo y toma as amostras atrasadas $x(t - \tau)$. O espaço de fase reconstruído está representado na parte superior da Figura 14.

A parte inferior da Figura 14 mostra espaço de fase reconstruído. Essa figura mostra um atrator em forma de ciclo limite no espaço de fase para o sistema associado. Verifica-se que a forma do atrator reconstruído no espaço de fase é muito similar à do atrator original, representado na parte superior da Figura 14.

Através do espaço de fase reconstruído Lindgren, Johnson e Povinelli (2003) desenvolveram um método para reconhecimento de voz. Neves, Femandes e Restivo (2006) melhoraram processos de filtragem utilizando o espaço de fase de uma série temporal. O espaço de fase foi utilizado para uma melhor compreensão do sinal e do ruído, quando comparado às análises no domínio do tempo ou da frequência. A filtragem baseou-se no fato de que o ruído se comportava como uma nuvem de pontos ao redor do sinal principal

Figura 13: Gráfico da informação mútua da série temporal da equação de *Van Der Pol*

representado no espaço de fase. Em sinais de voz, Dajer et al. (2011) utilizou o espaço de fase para determinar padrões visuais da dinâmica vocal para vozes normais e patológicas e Sobrinho (2011) extraiu a seção de Poincaré do espaço de fase tridimensional para medir a dispersão da periodicidade de sinais de voz normal e patológica.

4.2.2 Entropia

A entropia é um conceito estatístico usado em termodinâmica. Ela descreve o grau de desordem ou complexidade de um sistema dinâmico e mede a taxa média de perda de informação a medida que o sistema evolui no tempo. A entropia também pode ser definida como uma medida da incerteza de uma variável aleatória X . Se essa variável assumir um valor $x \in X$ num conjunto finito, a entropia $H(X)$ pode ser definida pela equação (4.8):

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (4.8)$$

sendo $p(x)$ a probabilidade de ocorrência da amostra x e X é o número total de amostras.

Em análise de sinais de voz, Costa et al. (2007) verificaram que a entropia relativa é mais eficaz do que a entropia de *Shannon* para discriminar vozes normais das patológicas. Também Henríquez et al. (2009) encontraram valores maiores de entropia de *Rényi* de

Figura 14: Atrator original (superior) e atrator reconstruído (inferior). O tempo de atraso foi escolhido usando como critério o mínimo da função de informação mútua. Observa-se que o atrator reconstruído é topologicamente semelhante ao original (FRASER; SWINNEY, 1986b).

1° e 2° ordem e entropia de *Shannon*¹ em vozes normais em relação às vozes patológicas, concluindo que estruturas geométricas mais complexas estão relacionadas às vozes patológicas. No estudo de Scalassara et al. (2009) verificaram-se altos valores de entropia relativa em vozes patológicas devido à sua complexidade e imprevisibilidade. Outro trabalho relacionado à entropia pode ser visto em Scalassara et al. (2008).

4.2.3 Bifurcação

Em sistemas dinâmicos, há o interesse em estudar o comportamento qualitativo dos atratores através de seus pontos de equilíbrio e órbitas periódicas. Também há o interesse em verificar o que ocorre quando há pequenas alterações nos parâmetros do sistema. Quando essas alterações implicam na mudança qualitativa de comportamento dinâmico e em particular dos pontos de equilíbrio ou das órbitas periódicas, elas são chamadas de bifurcações. Em sistemas dinâmicos existem quatro tipos principais de bifurcação: sela-nó, transcritical, *pitchfork* e *Hopf* (KHALIL, 2002). A Figura 15 abaixo ilustra um diagrama de bifurcação a partir da equação de recorrência definida por $x(n) = r \cdot x(n-1) \cdot (1 - x(n-1))$. Nessa figura observa-se uma duplicação do período para $r = 3.1$ e para

¹A entropia de Shannon é uma particularidade da entropia de Rnyi. Para mais informaes consultar Kantz e Schreiber (2004)

$r = 3.45$ tem-se 4 períodos. O aumento do parâmetro m conduz o sistema para o caos.

Figura 15: Gráfico da bifurcação da equação recursiva. A partir do parâmetro $r \geq 3$ o sistema apresenta uma série de pontos até se tornar caótico.

Em análise de voz, bifurcações são relatadas nos trabalhos descritos resumidamente a seguir. Mende, Herzel e Wermke (1990) analisaram vozes de recém-nascidos usando espectrograma e métodos da dinâmica não linear. Eles verificaram uma rica variedade de bifurcações e irregularidades no sinal de voz. Analisaram também a seção de Poincaré do atrator e sugeriram que essas características de ruído poderiam ser tratadas como caos determinístico de baixa dimensão. Em vozes patológicas, bifurcações surgem na forma de duplicação do período. Awrejcewicz (1990) estudou as bifurcações em sinais de voz utilizando equações diferenciais de 5° ordem que simulam o comportamento das pregas vocais onde foi possível verificar triplicação de períodos e bifurcações de *Hopf*.

Steinecke e Herzel (1995) utilizaram o modelo não linear da laringe composto por duas-massas (ISHIZAKA, 1972) e trabalharam com os parâmetros de pressão subglotal e rigidez das pregas vocais, verificando que o desbalanceamento entre as pregas esquerda e direita induzia harmônicos em regime subharmônico, oscilações toroidais e caos. Outros trabalhos sobre bifurcação em análise de vozes podem ser encontrados em Herzel (1993) e Berry et al. (1996).

4.2.4 Dimensão de correlação

A dimensão de correlação é um conceito introduzido por Grassberger e Procaccia (1983) e caracteriza a complexidade da dinâmica de um atrator. Quanto mais complexo o

sistema, maior será a dimensão de correlação. Essa dimensão está relacionada ao número de variáveis independentes necessárias para reconstrução do atrator. A dimensão de correlação também dá uma estimativa do número de variáveis independentes para modelar o trato vocal (KUMAR; MULLICK, 1990). Um importante resultado é dado por Kantz e Schreiber (2004), onde a máxima dimensão que se pode encontrar em sinal de comprimento N é dada por

$$D \simeq 2 \log_{10} (N - 2) \quad (4.9)$$

Assim, para um sinal de 22Khz, a dimensão máxima num trecho de 1 segundo será 8.

Vários trabalhos foram realizados utilizando o conceito de dimensão de correlação para análise de voz. Os estudos de Alonso et al. (2005a), Henríquez et al. (2009), Zhang e Jiang (2003), Jiang e Zhang (2002) e Costa, Neto e Costa (2010) foram capazes de distinguir vozes normais de patológicas através da dimensão de correlação.

4.2.5 Expoente de Lyapunov - λ

Algumas áreas do conhecimento humano como engenharia, medicina e biologia, lidam com dados extraídos de um sistema dinâmico na forma de uma série temporal. Esses dados podem vir acrescidos de ruído o que dificulta a análise. Se os dados forem caóticos, a dificuldade de análise aumenta pois o espectro de um sistema caótico e ruidoso são praticamente indistinguíveis. Um parâmetro para se medir o quanto um sinal é caótico é o expoente de Lyapunov (KINSNER, 2006).

O princípio básico para o cálculo de λ em uma série temporal é determinar a divergência de duas órbitas próximas no espaço de fase, conforme ilustrado na Figura 16. Para análise de sinais de voz, pode-se determinar o máximo expoente de Lyapunov conforme descrito a seguir. Sejam $\mathbf{s}_{n_1}$ e $\mathbf{s}_{n_2}$ dois pontos no espaço de fase a uma distância $\|\mathbf{s}_{n_1} - \mathbf{s}_{n_2}\| = \delta_0 \ll 1$. Seja $\delta_{\Delta n}$ a distância depois de um tempo Δn entre as duas trajetórias percorrendo esse espaço de fase dada por $\delta_{\Delta n} = \|\mathbf{s}_{n_1+\Delta n} - \mathbf{s}_{n_2+\Delta n}\|$. Então λ será dado pela expressão (4.10):

$$\delta_{\Delta n} \simeq \delta_0 e^{\lambda \Delta n}, \delta_{\Delta n} \ll 1, \Delta n \gg 1 \quad (4.10)$$

sendo o expoente λ dado pela equação (4.11):

$$\lambda \simeq \frac{1}{\Delta n} \ln\left(\frac{\delta_{\Delta n}}{\delta_0}\right) \quad (4.11)$$

Figura 16: Figura ilustrativa para o cálculo do expoente de Lyapunov

Se λ for positivo, significa que há uma divergência exponencial entre trajetórias próximas, o que é característico de comportamentos caóticos. A equação (4.11) é válida somente para intervalos Δn sobre o qual $\delta_{\Delta n}$ permanece pequeno (KANTZ; SCHREIBER, 2004).

O cálculo do expoente de Lyapunov foge do escopo desse trabalho mas há vários métodos propostos literatura que podem ser encontrados em Wolf et al. (1985), Eckmann et al. (1986) e Sano e Sawada (1985).

Sob o enfoque da teoria do caos, o expoente de Lyapunov é utilizado para análise de vozes normais e patológicas mas, de acordo com Matassini et al. (2000), é um parâmetro que deve ser utilizado com cautela. Kumar e Mullick (1996) encontraram um máximo expoente de Lyapunov positivo, característica de caos, em vozes normais (vogais) e consoantes. Por outro lado, Narayanan e Alwan (1995) calcularam um λ máximo igual a zero, indicando um ciclo limite, para vozes normais (vogais) e um λ positivo para sons fricativos. Steinecke e Herzel (1995) encontraram um máximo λ igual a zero para pronúncia de vogais de vozes saudáveis e um expoente de Lyapunov positivo para vogais disfônicas. Contudo, os autores advertiram quanto à sensibilidade do cálculo do λ em uma quantidade pequena de dados e trechos não estacionários. E ainda, a sensibilidade ao ruído

é prejudicial ao cálculo correto do λ (KANTZ; SCHREIBER, 1995). No entanto, Giovanni, Ouaknine e Triglia (1999) estudaram 12 vozes normais e 26 vozes com paralisia unilateral da laringe e encontraram o expoente de Lyapunov de $0,38 \pm 0,182$ para vozes normais e $0,57 \pm 0,337$ para vozes patológicas, mostrando ser uma característica útil para detectar vozes patológicas.

4.2.6 Seção de Poincaré

Uma técnica clássica para análise de sistemas dinâmicos é devido a Henry Poincaré. Essa técnica transforma um sistema dinâmico contínuo no tempo para um sistema discreto utilizando o mapa de Poincaré ou seção de Poincaré.

Dado um sistema dinâmico autônomo, n -dimensional em seu espaço de fase, a seção de Poincaré reduz a dimensão desse espaço de fase para um sistema de dimensão $(n - 1)$ e é determinada realizando-se um corte perpendicular às trajetórias do espaço de fase e tomando somente os pontos que interceptam essa seção.

A Figura 17 ilustra a seção de Poincaré realizada numa trajetória $\mathbf{x}$. Na figura x_i é um ponto da trajetória e Σ é o plano de dimensão $(n-1)$ transversal à trajetória e que contém x_i . A trajetória iniciando em x_i cortará o plano Σ em T segundos no ponto x_{i+1} . De acordo com Mann e McLaughlin (1997) e Kubin (1997), o intervalo entre cada corte no espaço de fase de um sinal de voz corresponde ao período de *pitch*.

Devido à continuidade do fluxo da trajetória, uma trajetória periódica ou quase periódica iniciando no plano Σ , numa vizinhança suficiente pequena nas vizinhanças de x_i , irá interceptar o plano Σ nas vizinhanças de x_i . Daí pode-se dizer que Σ define um mapeamento P de alguma vizinhança V contida em Σ de x_i , dentro de outra vizinhança V contida em Σ de x . P é chamado de mapa de Poincaré ou primeiro mapa de retorno (PARKER; CHUA, 1987). Pela seção de Poincaré é possível verificar que as trajetórias de alguns sistemas dinâmicos no espaço de fase nunca repetem o mesmo percurso, mas estão confinadas dentro de um feixe bem limitado (KUBIN, 1997). Em sistemas caóticos, a seção de Poincaré apresenta dimensão fractal e autosimilaridade.

Em análise de voz, alguns trabalhos foram realizados utilizando a seção de Poincaré. Herzel e Wendler (1991) observaram que a presença de estruturas regulares na seção de Poincaré de sinais de voz era uma indicação da presença de uma dinâmica de baixa dimensão. Em sinais de vozes fortemente soprosas, a seção de Poincaré apresentou um aspecto nebuloso, indicando um comportamento caótico. Na dissertação de mestrado de

Figura 17: Seção de Poincaré

Sobrinho (2011) foi possível diferenciar grupos de vozes normais de patológicas (edema e nódulo) utilizando a seção de Poincaré. Nesse trabalho, foram feitos 10 cortes no espaço de fase tridimensional e tomados pontos pertencentes ao plano e à trajetória. Foi calculada a dispersão média dos pontos de vozes normais, com edema e com nódulo e constatado que, na média, a dispersão dos pontos no plano de Poincaré de vozes patológicas é maior do que em vozes normais; mas através desse método não foi possível diferenciar grupos patológicos.

Em Dajer, Sobrinho e Pereira (2010), a seção de Poincaré foi utilizada como uma medida quantitativa dos padrões visuais da dinâmica vocal. Em Dajer et al. (2011), a seção de Poincaré foi utilizada para verificar a evolução das terapias vocais onde verificou-se que, após as terapias, a dispersão dos pontos da seção de Poincaré foram menores do que as medidas pré-terapia. Outros trabalhos nessa linha utilizando a seção de Poincaré podem ser encontrados em Dajer, Sobrinho e Pereira (2009) e Dajer, Sobrinho e Pereira (2009b).

Little et al. (2006) também afirmam que a complexidade de um sinal vocal é um fator limitante da técnica de análise linear. Nesse estudo, foi utilizada a seção de Poincaré para determinar a diferença dos intervalos que cada trajetória interceptava a seção. A partir dessas diferenças, estabeleceu-se uma medida de entropia e concluiu-se que sinais de vozes

patológicas possuíam entropia maior do que os sinais normais.

A tese de doutorado de Hagmüller (2009) apresentou técnicas de avaliação do aprimoramento de vozes com neoplasia da laringe. Ele verificou que a redução do *pitch* estava relacionada à melhora da voz. A marcação do *pitch* foi feita utilizando a seção de Poincaré.

O trabalho de Jafari, Almasganj e Bidhendi (2010) estudou o reconhecimento de voz combinando a *Mel Frequency Cepstral Coefficients* (MFCCs) e a seção de Poincaré. Nesse estudo houve uma eficiência de reconhecimento 5,7% maior do que a técnica MFCC aplicada isoladamente.

Em estudos não relacionados à análise de voz, Propes (1996) utilizou a seção de *Poincaré* para analisar o comportamento de longo termo de um sistema periódico forçado (circuito Diodo-R-L). Roux, Simoyi e Swinney (1983) estudaram o atrator reconstruído de uma reação química homogênea não equilibrada. No atrator resultante, extraíram a seção de Poincaré e constataram o alongamento e dobramento do atrator, característica típica dos atratores estranhos, que se caracterizam por ter dimensão não inteira (dimensão fractal), e dinâmica com comportamento caótico.

4.3 Outras técnicas

Além das técnicas tradicionais de análise da dinâmica de sistemas não lineares descritas nos itens anteriores, algumas técnicas recentes desenvolvidas para detecção de patologias nas pregas vocais merecem atenção. Dajer et al. (2011) analisou vozes normais e patológicas aplicando análise de padrões visuais da dinâmica vocal (PVDV) em conjunto com análise acústica e perceptiva - auditiva. Nesse trabalho o espaço de fase foi reconstruído e através dele foram analisados os parâmetros de *loops*, regularidade e convergência de traçados. Utilizando esses 3 parâmetros foi possível diferenciar as vozes normais das vozes com nódulo e edema de Reinke. Scalassara (2009) utilizou medidas de previsibilidade em sinais de voz onde partiu-se da hipótese que uma voz com alteração era menos previsível do que uma voz normal. Essa técnica foi eficiente para discriminar vozes normais e vozes com edema de Reinke e nódulo. Em Accardo, Fabbro e Mumolo (1992) e Baken (1990) foi possível classificar e avaliar as irregularidades fisiológicas e acústicas das pregas vocais utilizando a dimensão fractal, um conceito da teoria do caos. Fonseca et al. (2007) utilizaram transformada discreta *wavelet* de *Daubechies* (DWT-db), coeficiente de predição linear (LPC), e *Least Squares Support Vector Machines* (LS-SVM) para analisar distúrbios na voz. Godino-Llorente et al. (2005) e Guido et al. (2006) utilizaram *Suport*

Vector Machine para classificação de patologias. Godino-Llorente et al. (2005) usaram *support vector machine* para detecção de alteração de voz.

A transformada *wavelet* é um processo de filtragem digital no domínio do tempo, através da convolução, seguida de *downsampling* por 2. Ela é útil quando se deseja estudar o sinal em suas componentes de frequência separadamente. A vantagem de se utilizar a transformada *wavelet* é que o sinal transformado contém informações sobre as frequências que compõe o sinal original, e também onde tais frequências se localizam dentro do sinal. Segundo Hirano et al. (1990), a rouquidão provoca ruídos de baixa frequência e através da transformada *wavelet*, pode-se isolar as baixas frequências e extrair informações sobre o sinal. Os mesmos autores afirmam que a aspereza provoca ruídos de alta frequência e novamente pode-se explorar as técnicas da transformada *wavelet* para obter novas informações.

Essa técnica é utilizada em diversas áreas do conhecimento como medicina, física e engenharia. Em análise de voz, alguns estudos foram feitos usando essa técnica para análise e classificação de vozes normais e patológicas. Parraga (2002) utilizou a transformada *wavelet packet* e o algoritmo *Best Basis* para classificação automática de vozes em patológicas ou normais. A técnica utilizada permitiu separar vozes normais de patológicas com um erro de classificação de 23,07% para falsos positivos e de 14,58% para falsos negativos. Carvalho, Cavalcante e Cortez (2011) utilizaram a transformada *wavelet* e redes neurais artificiais para caracterizar sinais de vozes normais, vozes com nódulo, edema de Reinke e disfonia neurológica. No recente trabalho de Saeedi e Almasganj (2013), extraíram-se parâmetros de uma transformada *wavelet* e algoritmo genético. Através da *wavelet* gerada, construiu-se um banco de filtros e os sinais de vozes foram decompostos para se efetuar o cálculo de 8 características de energia. Em seguida foi utilizado *Support Vector Machine-SVM* onde, segundo os autores, foi possível classificar paralisia, nódulos, polipo, edema, disfonia espasmódica e queratose vocal.

Também no estudo de Almeida, Soares e Bresolin (2012), foram extraídos os coeficientes utilizando a transformada *wavelet packet*, família *Daubechies-5*, de sinais de vozes normais, com edema e nódulo e aplicados no SVM. Obtiveram um taxa de 98,75% de acerto na classificação entre vozes patológicas e normais. Em outro trabalho recente, Hariharan, Polat e Yaacob (2013) utilizaram *wavelet* e decomposição em valor singular para detecção de patologias em pregas vocais. Nesse estudo foram utilizados 4 classificadores (*k-nearest neighbour*, *least-square support vector machine*, rede neural probabilística e rede neural de regressão geral) e duas base de dados de sinais de voz. De acordo com os auto-

res, as patologias foram classificadas com 100% de acerto. Outros trabalhos envolvendo análise de voz utilizando *wavelet* para detecção de patologias podem ser encontrados em Murugesapandian, Yaacob e Hariharan (2008) e Arjmandi e Pooyan (2012).

No trabalho de Bassi (2014) foi utilizada uma técnica não invasiva para detecção de patologias utilizando classificadores baseados em distância euclidiana e redes neurais. Foram extraídos a média e desvio padrão das energias das faixas de frequência da escala *Bark* e esses dados foram utilizados como entrada para os classificadores propostos. Esse trabalho mostrou-se eficaz na detecção de falsos negativos, ou seja, vozes patológicas reconhecidas como normais.

5 Metodologia proposta para extração da seção de *Poincaré* de um sinal de voz

Neste capítulo será apresentado o método proposto para extração da seção de Poincaré de um sinal de voz. O diagrama da Figura 18 mostra as principais etapas do método proposto.

Figura 18: Etapas realizadas desde a análise do sinal de voz no tempo até a análise da dispersão dos pontos na seção de Poincaré.

Para reconstrução do espaço de fase foi utilizado o programa de análise de dinâmica não linear desenvolvido a partir do pacote *TISEAN* (HEGGER; KANTZ; SCHREIBER, 1999). Em seguida, foi extraída a seção de Poincaré de todo o espaço de fase através de um algoritmo desenvolvido pelo autor desse trabalho no software MATLAB 7.0 - versão *student*.

5.1 Procedimento para extrair a seção de Poincaré

A ferramenta proposta nessa tese inicialmente considera um sinal de voz completo, indicado na Figura 19, normalizado entre os valores -1 e 1, conforme a equação (5.1):

$$x'(t_k) = \frac{x(t_k)}{\max_k x(t_k)} \quad (5.1)$$

Sendo:

$x'(t_k)$: Sinal amostrado normalizado

$x(t_k)$: Sinal amostrado original

$\max_k x(t_k)$: Amplitude máxima do sinal amostrado original

As amostras são extraídas em intervalos regulares de tempo dado por $t_k = t_0 + k\Delta t$, onde Δt é o intervalo de tempo entre duas amostras consecutivas.

Desse procedimento, resulta a Figura 19.

Figura 19: Sinal de voz completo

A partir do sinal da Figura 19, seleciona-se um trecho de 500 ms (indicado pelas barras tracejadas). O trecho de 500ms é suficiente para que se tenha vários ciclos no espaço de fase. Trechos maiores do que 500ms aumentam o tempo computacional para análise além de impossibilitar a visualização do espaço de fase. Em seguida, o sinal é interpolado linearmente em 10 vezes para melhor exatidão das medidas. Dessa forma, obtém-se o sinal de voz normalizado representado na Figura 20.

Figura 20: Trecho do sinal normalizado

Em seguida, obtém-se o gráfico da função de informação mútua no trecho de 500ms utilizando o pacote *TISEAN* (HEGGER; KANTZ; SCHREIBER, 1999), e dessa função seleciona-se o primeiro mínimo, conforme indicado pelo cruzamento das linhas vermelhas na Figura 21.

Figura 21: Informação Mútua

De acordo com o gráfico da informação mútua (Figura 21), uma estimativa do primeiro mínimo ocorre na oitava amostra do sinal. No exemplo da Figura 20, o intervalo entre 2 amostras consecutivas corresponde ao período de amostragem $T = 1/22050$ s. Então o tempo de atraso será $\tau = 9(1/22050) = 408,16\mu s$.

A partir desse sinal, gera-se o espaço de fase reconstruído em 2 dimensões utilizando a técnica de tempo de atraso, conforme indicado na Figura 22. Nas figuras dos espaços de fase gerados considera-se $n = t_k$. O espaço de fase foi reconstruído utilizando o pacote Tisean (*Time series analysis*). O tempo de atraso τ , para reconstrução do espaço de fase, é determinado tomando-se o primeiro mínimo da informação mútua (FRASER; SWINNEY, 1986b).

Na espaço de fase representado na Figura 22, selecionam-se manualmente 2 pontos $P_1(x_1, y_1)$ e $P_2(x_2, y_2)$ de modo a formar uma reta tangente às trajetórias do espaço de fase. Esse 2 pontos devem ser selecionados num trecho do espaço de fase que não tenha curvas ou irregularidades. A partir desses 2 pontos, rotaciona-se todo o espaço de fase de um ângulo θ , de acordo com a expressão (5.2). Essa rotação é feita para se trabalhar com o eixo x como referência na horizontal.

$$\theta = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (5.2)$$

Figura 22: Espaço de fase

Após a rotação, obtém-se a Figura 23a e uma ampliação desta figura nas vizinhanças dos pontos P_1 e P_2 resulta na Figura 23b. Nessa figura, cinco pontos indicados por $P_3(x_3; y_3)$, $P_4(x_4; y_4)$, $P_5(x_5; y_5)$, $P_6(x_6; y_6)$ e $P_7(x_7; y_7)$ são manualmente selecionados entre os pontos P_1 e P_2 . A distância entre os pontos P_3 e P_4 e P_5 e P_6 deve conter pelo menos um ponto. O ponto P_7 é necessário para o algoritmo delimitar os conjuntos de pontos X' e X'' , descritos a seguir. A distância adotada na Figura 23b serve apenas para ilustrar o método proposto nesse trabalho.

Figura 23: (a) Espaço de fase rotacionado em 2 dimensões do sinal de voz da Figura 20. (b) Ampliação do espaço de fase. As linhas pontilhadas estabelecem os limites da região do corte e a orientação da linha tangente em relação ao fluxo da trajetória.

Em seguida, o algoritmo extrai automaticamente 2 grupos de pontos X' e X'' . O conjunto X' é composto de todos os pontos $(x(t_k), x(t_k - \tau), x(t_k - 2\tau))$ em $\mathbb{R}^3$ que pertencem ao intervalo limitado pelos pontos P_3 e P_4 como indicado na Figura 23b. A terceira coordenada se faz necessária para se obter a seção de Poincaré em 2 dimensões. Apesar da Figura 23b representar apenas 2 dimensões, a rotina que extrai a seção de Poincaré armazena as coordenadas da terceira dimensão indexada às 2 coordenadas $(x(t_k), x(t_k - \tau))$ para uso futuro. Matematicamente, esse conjunto é definido como:

$$X' = \{ (x(t_k), x(t_k - \tau), x(t_k - 2\tau)) \in \mathbb{R}^3, k \in \mathbb{I} \\ x_3 \leq x(t_k) \leq x_4, y_7 \leq x(t_k - \tau) \leq \min(y_3, y_4) \}$$

Onde $\mathbb{I} = \{1, 2, \dots, N\}$ e N é o número de amostras no sinal de voz normalizado.

Da mesma forma, o conjunto X'' é composto de todos os pontos $(x(t_k), x(t_k - \tau), x(t_k - 2\tau))$ em $\mathbb{R}^3$ que pertencem ao intervalo limitado pelos pontos P_5 e P_6 como indicado na Figura 23b. Matematicamente, esse conjunto é definido como:

$$X'' = \{ (x(t_k), x(t_k - \tau), x(t_k - 2\tau)) \in \mathbb{R}^3, k \in \mathbb{I} \\ x_5 \leq x(t_k) \leq x_6, y_7 \leq x(t_k - \tau) \leq \min(y_5, y_6) \}$$

Uma ampliação da Figura 23b, entre os pontos P_3 e P_4 , resulta na Figura 24. Essa

figura mostra que várias trajetórias contêm mais de um ponto no intervalo.

Figura 24: Espaço de fase ampliado. A seta mostra a ocorrência de mais de um ponto no trecho em estudo.

Agora, para obter a seção de *Poincarè*, o ponto médio O do conjunto X' entre os pontos P_3 e P_4 e o ponto médio A do conjunto X'' entre os pontos P_5 e P_6 são definidos de acordo com a expressão (5.3) and (5.4).

$$O = \{\overline{x'(t_k)}; \overline{x'(t_k - \tau)}; \overline{x'(t_k - 2\tau)}\} \quad (5.3)$$

$$A = \{\overline{x''(t_k)}; \overline{x''(t_k - \tau)}; \overline{x''(t_k - 2\tau)}\} \quad (5.4)$$

onde

$$\overline{x'(t_k)} = \frac{\sum_{k=1}^n x(t_k)}{n}, \quad (5.5)$$

$$\overline{x'(t_k - \tau)} = \frac{\sum_{k=1}^n x(t_k - \tau)}{n} \quad (5.6)$$

$$\overline{x'(t_k - 2\tau)} = \frac{\sum_{k=1}^n x(t_k - 2\tau)}{n} \quad (5.7)$$

$$\overline{x''(t_k)} = \frac{\sum_{k=1}^m x(t_k)}{m}, \quad (5.8)$$

$$\overline{x''(t_k - \tau)} = \frac{\sum_{k=1}^m x(t_k - \tau)}{m} \quad (5.9)$$

$$\overline{x''(t_k - 2\tau)} = \frac{\sum_{k=1}^m x(t_k - 2\tau)}{m} \quad (5.10)$$

E n é o número de pontos limitados pelos pontos P_3 e P_4 , e m é o número de pontos limitados pelos pontos P_5 e P_6 , ambos pertencentes à Figura 23.

O vetor $\overrightarrow{OA}$ define a reta tangente à trajetória média nas vizinhanças de P_1 e P_2 e o plano de Poincaré será o plano que contém o ponto O e é perpendicular ao vetor $\overrightarrow{OA}$.

Em seguida, determinam-se os pontos da terceira coordenada (Eixo z), ou seja, $x[n - 2\tau]$ e os índices referentes ao início de cada segmento de trajetória para que elas sejam isoladas umas das outras. Em verdade, o conjunto X' é composto de todos os segmentos da trajetória $x(t_k)$ que foram limitados entre os pontos P_3 e P_4 , conforme ilustra a Figura 24. Para facilitar a identificação de cada segmento, define-se cada um desses segmentos pelo indexador i escrito como $x_i(t_{j_n})$, onde j representa a posição ao longo do segmento da trajetória i . O indexador n serve para diferenciar cada segmento uma vez que eles não possuem, necessariamente, o mesmo número de pontos. Para se obter o mesmo número de posições j em cada segmento, normaliza-se a quantidade de posições j_n em relação ao seu valor máximo através de interpolação. A Figura 25 ilustra os segmentos da trajetória i e as posições j_n , antes da normalização dos segmentos. As coordenadas da terceira dimensão são armazenadas e indexadas às outras 2 coordenadas.

Depois da interpolação, definem-se os pontos $B(i, j)$, de acordo com a equação (5.11).

$$B(i, j) = \{x'_i(t_j), x'_i(t_j - \tau), x'_i(t_j - 2\tau)\} \quad (5.11)$$

Figura 25: Ilustração dos segmentos de trajetória

Os pontos $B(i, j)$ são aqueles definidos por X' .

Agora, o algoritmo calcula o produto escalar entre os vetores $\overrightarrow{OA}$ e $\overrightarrow{OB(i, j)}$. A condição para um ponto $B(i, j)$ pertencer ao plano de Poincaré é $\overrightarrow{OA} \cdot \overrightarrow{OB(i, j)} = 0$, ou seja, o produto escalar entre os vetores $\overrightarrow{OA}$ e $\overrightarrow{OB(i, j)}$ em $\mathbb{R}^3$ é igual a zero (Figura 26). Devido à amostragem, o produto escalar pode não ser exatamente zero. Nesse caso constrói-se uma matriz M_{ixj} , onde a i -ésima linha da matriz é o resultado do produto escalar dos elementos da trajetória x_i e a j -ésima coluna representa o elemento de cada trajetória limitada pelos pontos P_3 e P_4 . O ponto com o menor produto escalar para cada segmento de trajetória, em módulo, será aquele mais próximo do plano de Poincaré. A Figura 27a ilustra os pontos $B(i, j)$ destacados em vermelho pertencentes à seção de Poincaré.

Uma ilustração desse procedimento pode ser feita utilizando um trecho de 1000 ms. Realizando os procedimentos anteriores obteve-se a Figura 27b.

Figura 26: Determinação dos pontos pertencentes ao plano de Poincaré através do produto escalar

Figura 27: A figura da esquerda mostra o trecho selecionado com os pontos resultantes do produto escalar (em vermelho) pertencentes ao plano de Poincaré e a reta tangente à trajetória. A figura da direita mostra o hiperplano de Poincaré cortando o trecho do espaço de fase.

Na Figura 27 pode-se visualizar a reta tangente à trajetória, a região do corte selecionada (formada pelos pontos pretos) e os pontos da seção de Poincaré, em vermelho, que obedecem à equação do produto escalar.

Depois de selecionar os pontos pertencentes à seção de Poincaré para todos os segmentos, esses pontos são rotacionados 2 vezes para representá-los no plano. A primeira rotação é feita em relação ao eixo y de acordo com a expressão (5.12)

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (5.12)$$

onde (x, y, z) são os pontos pertencentes à seção de Poincaré e (x', y', z') são os pontos (x, y, z) depois da primeira rotação e θ é definido conforme equação (5.13)

$$\theta = \arctan\left(\frac{A_z - O_z}{A_x - O_x}\right) \quad (5.13)$$

onde A_z, O_z, A_x e O_x são as coordenadas dos pontos O e A , respectivamente.

O vetor $\overrightarrow{OA}$ também deve ser rotacionado de acordo com a expressão (5.14) e (5.15) para ser usado como referência para segunda rotação.

$$\begin{bmatrix} O_{x'} \\ O_{y'} \\ O_{z'} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} O_x \\ O_y \\ O_z \end{bmatrix} \quad (5.14)$$

$$\begin{bmatrix} A_{x'} \\ A_{y'} \\ A_{z'} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} \quad (5.15)$$

onde O' e A' são os pontos que definem o vetor $\overrightarrow{O'A'}$ depois da primeira rotação.

A segunda rotação é feita em relação ao eixo z , de acordo com a expressão (5.16)

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (5.16)$$

onde (x'', y'', z'') são os pontos pertencentes à seção de Poincaré depois da rotação de (x', y', z') e ϕ é definido de acordo com a (5.17).

$$\phi = \arctan\left(\frac{A'_y - O'_y}{A'_x - O'_x}\right) \quad (5.17)$$

onde A'_y, O'_y, A'_x e O'_x são as coordenadas dos pontos O' e A' .

Depois das duas rotações, os pontos são projetados no plano, obtendo-se a seção de Poincaré, ilustrada na Figura 28a. Os respectivos pontos da seção em relação ao tempo estão destacados em vermelho na Figura 28b.

Figura 28: (a) Seção de Poincaré extraída do conjunto X' mostrado na Figura 26. (b) Os pontos em vermelho representam os pontos do sinal (no tempo) que cortam o plano de Poincaré no espaço de fase.

Depois que todos os pontos da seção são encontrados, o plano de Poincaré percorre todo o espaço de fase tomando como referência o espaço de fase médio. Para um espaço de fase médio de tamanho L , é possível extrair n seções de Poincaré. O n -ésimo ponto do espaço de fase médio que contém a seção de Poincaré é representado pelo plano Σ_n .

Conforme o plano de Poincaré percorre o espaço de fase medio, uma subrotina cria uma matriz $S_{n \times k}$, onde n representa o n -ésimo plano Σ e k representa o k -ésimo ponto pertencente ao plano Σ . Aqui, o indexador k armazena as coordenadas (x, y) dos pontos da seção. A Matriz S é representada a seguir.

$$S = \begin{bmatrix} P(1,1) & P(1,2) & P(1,3) & . & . & . & P(1,k) \\ P(2,1) & P(2,2) & P(2,3) & . & . & . & P(2,k) \\ P(3,1) & P(3,2) & P(3,3) & . & . & . & P(3,k) \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ P(n,1) & P(n,2) & . & . & . & . & P(n,k) \end{bmatrix}$$

Na matriz S , a i -ésima linha contém a sequência de pontos da trajetória que pertencem à sua correspondente seção de Poincaré e a j -ésima coluna é formada pelos pontos que pertencem à j -ésima seção de Poincaré. A Figura 29 mostra 90 seções de Poincaré com 58 pontos em cada seção, extraídas da matriz S .

Figura 29: Seções de Poincaré em vermelho obtidas em todo o espaço de fase de um sinal de voz normal.

Calculando-se a média de todas as seções de Poincaré, uma a uma, ao longo do espaço de fase, pode-se determinar o espaço de fase médio, conforme ilustrado na Figura 30. Essa figura dá uma visão geral do comportamento do espaço de fase no trecho do sinal de voz selecionado através de uma única curva no espaço de fase.

A Figura 31 e a Figura 32 são alguns exemplos de espaço de fase e seções de Poincaré extraídas da base de dados utilizadas nessa tese. As Figuras 31(a) e (b) representam as seções de Poincaré extraídas de duas vozes com nódulo. As Figuras 32(a) e (b) representam uma voz normal e uma voz com edema, respectivamente.

Figura 30: Espaço de fase médio de um sinal de voz normal (destacado em vermelho).

Figura 31: Espaço de fase e seções de Poincaré extraídas de 2 sinais de vozes com nódulo.

Figura 32: Espaço de fase e seções de Poincaré extraídas de um sinal de voz normal (c) e um sinal de voz com edema (d).

Tomando-se a posição dos pontos do espaço de fase referido ao sinal temporal, conforme indicado na Figura 33, verifica-se de fato que a seção de Poincaré também é útil para marcação de *pitch* (KUBIN, 1997).

Figura 33: Pontos da seção de Poincaré no tempo

Uma vantagem desse novo algoritmo é a possibilidade de analisar trechos inacessíveis

do espaço de fase, tanto de vozes normais quanto patológicas, que não foram analisados em Sobrinho (2011). As setas ilustradas na Figura 34 mostra um exemplo de um sinal de voz com trechos onde não foi possível extrair a seção de Poincaré pelo método utilizado em Sobrinho (2011).

Figura 34: Espaço de fase de um sinal de voz. As setas mostram os pontos inacessíveis utilizando o método utilizado em Sobrinho (2011).

Outra vantagem do método proposto nessa tese é a possibilidade de se referir cada trajetória no tempo. Sobrepondo-as uma a uma pode-se visualizar as diferenças de amplitude (*shimmer*) e período (*jitter*), conforme já ilustrado na Figura 11. Essa técnica para extração da seção de *Poincaré* foi apresentada em Sobrinho, Dajer e Alberto (2015).

5.1.0.1 Cálculo da frequência fundamental (F_0) através da seção de *Poincaré*

Um dos parâmetros mais importante para análise de voz é o cálculo da frequência fundamental (F_0). A ferramenta clássica para determinar F_0 de um sinal de voz é a transformada de Fourier (RABINER; SCHAFER, 1979), onde o sinal de voz é representado em seu espectro de frequência. A frequência fundamental é dada pelo primeiro pico do espectro.

Em geral, a transformada de Fourier é aplicada em sinais periódicos que possuem várias componentes harmônicas. Nas últimas décadas, vários pesquisadores descobriram que sinais de voz apresentam não-linearidade (ALONSO et al., 2005b), (KUMAR; MULLICK, 1996), (NARAYANAN; ALWAN, 1995) e a aplicação da transformada de Fourier pode não extrair a frequência fundamental exata. Caso o sinal de voz possua forte não-linearidade, o espectro de frequência pode apresentar os primeiros picos próximos um do outro, conforme ilustrado na Figura 35.

Figura 35: O espectro de frequências de um sinal de voz com nóculo mostra dois picos próximos.

De acordo com Kubin (1997), o fluxo da trajetória no espaço de fase realiza um ciclo completo depois de um período de pitch, T_0 . O período de *pitch* pode ser marcado por meio da seção transversal através do espaço de fase ou seção de Poincaré. Para verificar a validade do algoritmo proposto para extração da frequência fundamental, foram utilizadas 13 vozes de crianças pertencentes ao banco de dados do Grupo de pesquisa e Engenharia Médica e a fundação CNPQ. Cada sinal de voz representa 5 segundos da vogal sustentada 'a' da língua portuguesa. As vozes foram amostradas a uma taxa de 22050 Hz. A frequência fundamental foi extraída de um trecho de 200ms numa região de maior estacionariedade em amplitude através da seção de Poincaré (apresentada nessa tese) e comparada com a frequência fundamental do banco de dados. Foi utilizado o teste

estatístico T com um nível de significância p de 5%. O teste T é usado para verificar se dois grupos de dados pertencem a uma mesma amostra. Isso significa que se o valor de p for menor do que 5%, as amostras pertencem a grupos diferentes. Caso contrário, os grupos pertencem a uma mesma amostra.

Considerando que o trecho de 200ms possui vários ciclos no espaço de fase, o tempo médio de retorno pode ser calculado através da expressão (5.18):

$$T_0 = \frac{\sum_{i=1}^n T_i}{n} \quad (5.18)$$

onde n é o número de ciclos. A frequência fundamental será dada por $F_0 = T_0^{-1}$.

O gráfico da Figura 36 mostra 51 ciclos do espaço de fase que corresponde a 200ms do sinal de voz da criança número 2. A frequência fundamental extraída pela seção de Poincaré foi de 260Hz.

Figura 36: Frequência fundamental extraída através da seção de Poincaré em 51 ciclos do espaço de fase.

Os resultados obtidos estão dispostos na Tabela 1.

A média e o desvio padrão de ambas as frequências estão dispostos na Tabela 2.

O teste estatístico T apresentou um valor p igual a 0.94%. Isso significa que as

Tabela 1: Frequência fundamental extraída usando Transformada de Fourier e seção de Poincaré

Frequência fundamental, F_0		
Crianças	Transformada de Fourier [Hz]	Seção de Poincaré. [Hz]
1	172.3	191.49
2	258.4	260.14
3	301.5	324.93
4	258.4	256.4
5	215.3	201.75
6	301.5	297.00
7	258.4	244.73
8	215.3	202.65
9	215.3	198.33
10	258.4	236.59
11	215.3	222.49
12	215.3	228.02
13	258.4	263.48

Tabela 2: Média e desvio padrão de F_0
Transformada de Fourier Seção de Poincaré

μ	241.83	240.62
σ	37.47	39.91

frequências extraídas pela Transformada de Fourier e seção de Poincaré são estatisticamente iguais ((SOBRINHO; DAJER; ALBERTO, 2016)), o que comprova que a ferramenta proposta nessa tese é útil para determinar a frequência fundamental de um sinal de voz.

5.2 Validação da ferramenta utilizando senóides

Para se extrair a seção de Poincaré do espaço de fase é necessário determinar o plano Σ de dimensão (n-1), que pode ser feito de várias maneiras (PARKER; CHUA; PARKER, 1989). Nesse trabalho, para verificar a eficácia do algoritmo, foi feita a validação da ferramenta utilizando o espaço de fase reconstruído de um sinal senoidal inserindo *jitter* gradativamente e considerando *shimmer* nulo. A dispersão dos pontos da seção foi calculada para cada plano gerado e a média da dispersão foi extraída plano a plano. Como será visto posteriormente, o cálculo da frequência fundamental através da seção de Poincaré foi igual ao medido manualmente, o que também comprova a eficácia do algoritmo.

Procedendo dessa forma foram gerados os valores indicados na Tabela 3.

A partir da Tabela 3 foi gerada a Figura 37, onde se verifica um aumento proporcional

Tabela 3: Média da dispersão para uma senóide.

Resultados	
Jitter [%]	Media
1,25	0,0001
2,5	0,0003
3,75	0,0004
5,00	0,0005
6,25	0,0006
7,50	0,0008
8,75	0,0009
10,00	0,0011
11,25	0,0012
12,50	0,0014

da dispersão a cada incremento do *jitter*. Como vozes patológicas possuem maiores valores de *jitter* em relação as vozes normais (ZHANG et al., 2004), a validação da ferramenta garante que a dispersão dos pontos na seção de Poincaré de um sinal de voz patológica será maior do que em vozes em vozes normais.

Figura 37: Dispersão dos pontos da seção de Poincaré para um sinal senoidal.

6 Resultados e discussões

Neste capítulo, os sinais de vozes analisados pertencem ao Grupo de Pesquisa em Engenharia Médica (CNPq). Foram selecionadas 16 vozes normais, 16 vozes com edema e 16 vozes com nódulo de Reinke. Esses sinais foram retirados do banco de dados do laboratório de processamento de sinais do departamento de engenharia elétrica da escola de engenharia de São Carlos (EESC) da Universidade de São Paulo, e utilizados em diversos trabalhos (ROSA; PEREIRA; GRELET, 2000), (DAJER; PEREIRA; MACIEL, 2005), (FONSECA et al., 2007), (DAJER; SOBRINHO; PEREIRA, 2010), (DAJER et al., 2011) e (SOBRINHO, 2011).

Os sinais pertencem a pessoas com idade entre 18 e 60 anos que foram diagnosticadas por médicos dos setores de otorrinolaringologia e cirurgia de cabeça e pescoço do hospital de clínicas da faculdade de medicina de Ribeirão Preto. Descreve também que foram utilizados videolaringoscópio e estroboscópio para esses diagnósticos e análise final feita por uma fonoaudióloga.

Foram utilizados 2 critérios para extração da seção de Poincaré. O primeiro critério utiliza trechos de 500ms extraídos do início do sinal (0.5s;1.0s), no trecho intermediário do sinal (2.0s;2.5s) e no final do sinal (4.0s;4.5s). A Figura 38 ilustra 5.0s de um sinal de voz da vogal sustentada /a/, e o intervalo de onde vai se extrair a seção de Poincaré, delimitado pelas barras vermelhas.

Figura 38: Sinal de voz completo normalizado. As barras vermelhas limitam um trecho com menor variação em amplitude utilizando critério visual.

O segundo critério utiliza trechos de 500ms com maior variação no *pitch*. Inicialmente, uma subrotina calcula a variação do *pitch* em todo o trecho do sinal de voz, conforme ilustrado na Figura 39. Em seguida, uma janela de 500ms percorre o gráfico do *pitch* gerado calculando o coeficiente angular da reta formada pelo ponto final e inicial de cada janela. O intervalo de 500ms que possuir o maior coeficiente angular será considerado aquele com maior variação do *pitch*. Na Figura 39, o intervalo de 500ms com maior variabilidade do *pitch* calculado está aproximadamente no trecho entre 2.8s e 3.3s, limitado pelas barras vermelhas. A reta destacada em verde mostra a inclinação da reta nesse trecho.

Figura 39: Variação do *pitch* no sinal de voz. A reta destacada em verde possui a maior inclinação num intervalo de 500ms.

6.1 Testes estatísticos

Procedendo de acordo com o 1º critério, foram extraídas a média e desvio padrão da dispersão dos pontos de todas as seções de Poincaré em todo espaço de fase nos 3 tipos de sinais (normal, edema e nódulo), obtendo-se as Tabelas 4, 5 e 6. Testes preliminares mostraram que a média e o desvio padrão dos 3 tipos de vozes provém de uma distribuição normal.

Em seguida foram extraídas as mesmas medidas usando o segundo critério, obtendo-se a Tabela 7.

Das Tabelas 4, 5 e 6 foram gerados os gráficos comparativos das Figuras 40, 41 e 42.

Tabela 4: Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo extraídos do intervalo de 0.5s a 1.0s.

N° da voz	Voz normal		Voz com Edema		Voz com nódulo	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0.0360	0.0206	0.0880	0.0633	0.0942	0.0487
2	0.0073	0.0036	0.0765	0.0425	0.1290	0.0618
3	0.0376	0.0230	0.1125	0.0603	0.0635	0.0302
4	0.0518	0.0260	0.1121	0.0546	0.0635	0.0302
5	0.0314	0.0158	0.1160	0.0540	0.1265	0.0676
6	0.0302	0.0177	0.1195	0.0656	0.1322	0.0739
7	0.0432	0.0238	0.0752	0.0428	0.1590	0.0855
8	0.0725	0.0965	0.2706	0.2500	0.2242	0.1239
9	0.0629	0.0358	0.0634	0.0354	0.1581	0.0865
10	0.0323	0.0169	0.2042	0.1004	0.0937	0.0430
11	0.0690	0.0365	0.2221	0.1367	0.1001	0.0521
12	0.0484	0.0662	0.0400	0.0203	0.1220	0.0704
13	0.0581	0.0377	0.0455	0.0248	0.0962	0.0521
14	0.0353	0.0199	0.1267	0.0581	0.2891	0.1739
15	0.0802	0.0642	0.1223	0.0616	0.0527	0.0339
16	0.0350	0.0199	0.0320	0.0155	0.1096	0.0532
Média	0.0457	0.0328	0.1142	0.0679	0.1259	0.0679

Tabela 5: Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo extraídos do intervalo de 2.0s a 2.5s.

N° da voz	Voz normal		Voz com Edema		Voz com nódulo	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0.0530	0.0273	0.0640	0.0386	0.0886	0.0511
2	0.0482	0.0245	0.0716	0.0382	0.1174	0.0587
3	0.0532	0.0282	0.0979	0.0446	0.0445	0.0222
4	0.0465	0.0247	0.0646	0.0322	0.0905	0.0409
5	0.0349	0.0198	0.1383	0.0638	0.1038	0.1038
6	0.0213	0.0133	0.1123	0.0615	0.1037	0.0499
7	0.0344	0.0176	0.0633	0.0288	0.0954	0.0442
8	0.0479	0.0327	0.0696	0.0360	0.1036	0.0888
9	0.0619	0.0301	0.0520	0.0251	0.0882	0.0520
10	0.0236	0.0131	0.1560	0.0892	0.0594	0.0278
11	0.0559	0.0314	0.1543	0.1684	0.1016	0.0527
12	0.0420	0.0221	0.0728	0.0380	0.0655	0.0339
13	0.0598	0.0276	0.0869	0.0518	0.1367	0.0661
14	0.0416	0.0216	0.0601	0.0278	0.1460	0.0614
15	0.0596	0.0335	0.1273	0.0607	0.0558	0.0294
16	0.0459	0.0268	0.0518	0.0218	0.1155	0.0507
Média	0.0456	0.0246	0.0902	0.0517	0.0948	0.0521

Tabela 6: Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo extraídos do intervalo de 4.0s a 4.5s.

N° da voz	Voz normal		Voz com Edema		Voz com nódulo	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0.0581	0.0354	0.0694	0.0365	0.0671	0.0349
2	0.0498	0.0262	0.0572	0.0291	0.0919	0.0449
3	0.0589	0.0322	0.0553	0.0297	0.056	0.0271
4	0.0333	0.019	0.0834	0.0406	0.0792	0.0368
5	0.042	0.0273	0.1173	0.0556	0.1577	0.0734
6	0.017	0.0091	0.1107	0.0589	0.1004	0.0672
7	0.0472	0.0253	0.0703	0.036	0.0628	0.0331
8	0.0315	0.0238	0.0811	0.0373	0.0624	0.0297
9	0.0452	0.0237	0.0356	0.0205	0.0928	0.0446
10	0.0258	0.0124	0.1124	0.0681	0.0592	0.0298
11	0.0545	0.0304	0.1018	0.052	0.0916	0.0427
12	0.0462	0.0233	0.0633	0.0339	0.0655	0.0345
13	0.0474	0.0364	0.0921	0.0427	0.0937	0.0521
14	0.0268	0.0151	0.094	0.0515	0.1087	0.0997
15	0.1035	0.0772	0.1449	0.07	0.056	0.0275
16	0.0266	0.0144	0.0415	0.0221	0.1263	0.0609
Média	0.0446	0.02695	0.0831	0.0428	0.0857	0.0461

Tabela 7: Valor médio das dispersões da seção de Poincaré nos sinais de voz normal, com edema e nódulo - trechos com maior variação de pitch.

N° da voz	Voz normal		Voz com Edema		Voz com nódulo	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
1	0.0549	0.0287	0.1168	0.0603	0.0886	0.0511
2	0.0474	0.0240	0.0873	0.0421	0.1180	0.0583
3	0.0455	0.0253	0.0780	0.0393	0.0480	0.0235
4	0.0453	0.0229	0.0736	0.0363	0.0988	0.0514
5	0.0334	0.0193	0.1430	0.0738	0.1268	0.0680
6	0.0202	0.0122	0.1228	0.0644	0.1679	0.0628
7	0.0325	0.0172	0.0692	0.0300	0.1075	0.0579
8	0.0376	0.0220	0.2005	0.0761	0.2855	0.1063
9	0.0591	0.0283	0.0915	0.0394	0.1905	0.0913
10	0.0218	0.0114	0.1554	0.0855	0.0863	0.0398
11	0.0468	0.0238	0.1543	0.1684	0.1184	0.0558
12	0.0372	0.0206	0.0742	0.0380	0.0797	0.0408
13	0.0322	0.0159	0.1137	0.0508	0.1052	0.0523
14	0.0371	0.0200	0.0945	0.0458	0.1544	0.0640
15	0.0519	0.0252	0.1273	0.0607	0.0630	0.0363
16	0.0317	0.0168	0.0883	0.0335	0.1188	0.0557
Média	0.0397	0.02085	0.1119	0.0590	0.1223	0.0572

Figura 40: Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio. Intervalo: 0.5s - 1.0s.

Figura 41: Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio. Intervalo: 2.0s - 2.5s.

Figura 42: Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio. Intervalo: 4.0s - 4.5s.

Da Tabela 7 foi gerado o gráfico comparativo da Figura 43.

Figura 43: Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - trechos com maior variação de *pitch*.

Observando os gráficos das Figuras 40, 41, 42 e 43 nota-se uma diferenciação entre vozes normais e patológicas utilizando os 2 critérios. Para uma análise mais precisa, foi realizado o teste estatístico de variância ANOVA unilateral com 5% de significância com o objetivo de comparar os resultados de média e desvio padrão entre vozes saudáveis, com edema e com nódulo. Em seguida, foi feito o teste *Tukey* para analisar os grupos normal x edema, normal x nódulo e edema x nódulo. O teste Anova é útil quando as amostras pertencem a mais de 2 grupos. O teste de *Tukey* é útil para análise de grupos tomados dois a dois. O grau de confiança utilizado significa que se o valor de p para o teste estatístico for menor do que 5 %, as amostras pertencem a grupos diferentes.

6.1.1 Análise com 1º critério

6.1.1.1 Intervalo: 0.5s - 1.0s

Os resultados obtidos pelo teste Anova utilizando o primeiro critério no intervalo entre 0.5s - 1.0s estão dispostos na Figura 44. A Tabela 8 mostra os valores de p para esse teste.

Figura 44: Teste Anova - Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - 1º critério. Trecho (1.0s; 1.5s).

Tabela 8: Valores de p obtidos com o teste Anova no intervalo 0.5s - 1.0s entre as vozes normais, com edema e com nódulo.

	p
Média	0.0%
Desvio padrão	3.0%

Os resultados obtidos utilizando o teste de *Tukey* para análise do valor médio e desvio padrão de grupos tomados dois a dois estão indicados na Tabela 9, onde mostra os valores de p para esse teste.

Tabela 9: Valores de p obtidos com o teste de *Tukey* no intervalo 0.5s - 1.0s.

	Normal x Edema	Normal x Nódulo	Edema x Nódulo
Média	0.0%	0.0%	60.9%
Desvio padrão	3.0%	0.4%	99.7%

Para o intervalo entre 0.5s- 1.0s, o teste Anova mostra que os grupos de sinais de vozes são estatisticamente diferentes. O teste de Tukey mostra que as vozes normais são estatisticamente diferentes das vozes com patologias mas não diferenciou entre patologias.

6.1.1.2 Intervalo: 2.0s - 2.5s

Os resultados obtidos pelo teste Anova utilizando o primeiro critério no intervalo entre 2.0s - 2.5s estão dispostos na Figura 45. A Tabela 10 mostra os valores de p para esse teste.

Figura 45: Teste Anova - Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - 1º critério. Trecho (2.0s;2.5s).

Tabela 10: Valores de p obtidos com o teste Anova no intervalo 2.0s - 2.5s entre as vozes normais, com edema e com nódulo.

	p
Média	0.0%
Desvio padrão	3.3%

Os resultados obtidos utilizando o teste de *Tukey* para análise do valor médio e desvio padrão de grupos tomados dois a dois estão dispostos na Tabela 11, onde mostra os valores de p para esse teste.

Tabela 11: Valores de p obtidos com o teste de *Tukey* no intervalo 2.0s - 2.5s.

	Normal x Nódulo	Normal x Edema	Edema x Nódulo
Média	0.0%	0.0%	69.2%
Desvio padrão	0.6%	0.0%	72.8%

Para o intervalo entre 2.0s-2.5s, o teste Anova mostra que os grupos de sinais de vozes são estatisticamente diferentes. O teste de *Tukey* mostra que as vozes normais são estatisticamente diferentes das vozes com patologias mas não diferenciou entre patologias.

6.1.1.3 Intervalo: 4.0s - 4.5s

Os resultados obtidos pelo teste Anova utilizando o primeiro critério no intervalo entre 4.0s - 4.5s estão dispostos na Figura 46. A Tabela 12 mostra os valores de p para esse teste.

Figura 46: Teste Anova - Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio - 1º critério. Trecho (4.0s;4,5s).

Tabela 12: Valores de p obtidos com o teste Anova no intervalo 4.0s - 4.5s entre as vozes normais, com edema e com nódulo.

	p
Média	0.0%
Desvio padrão	0.6%

Os resultados obtidos utilizando o teste de *Tukey* para análise do valor médio e desvio padrão de grupos tomados dois a dois estão dispostos na Tabela 13, onde mostra os valores de p para esse teste.

Tabela 13: Valores de p obtidos com o teste de *Tukey* no intervalo 4.0s - 4.5s.

	Normal x Nódulo	Normal x Edema	Edema x Nódulo
Média	0.0%	0.0%	80.5%
Desvio padrão	0.7%	0.5%	59.38%

Para o intervalo entre 4.0s-4.5s, o teste Anova mostra que os grupos de sinais de vozes são estatisticamente diferentes. O teste de Tukey mostra que as vozes normais são estatisticamente diferentes das vozes com patologias mas não diferenciou entre patologias.

6.1.1.4 Análise do 1º critério utilizando Support Vector Machine (SVM)

Apesar do teste Anova e o teste de *Tukey* diferenciarem vozes normais das patológicas nos três intervalos estudados, há o interesse em classificar cada voz. Para isso foi utilizada uma rede neural baseada em *Support Vector Machine* (SVM) onde o eixo x corresponde à média da dispersão dos pontos na seção de Poincaré e o eixo y corresponde ao desvio padrão da dispersão dos pontos na seção. Os resultados estão dispostos nos gráficos das Figuras 47, 48, 49. Os gráficos foram gerados através de uma subrotina do *Matlab - versão student*. Nesses gráficos há uma linha (fronteira) que divide os grupos de vozes onde os pontos à esquerda da linha são classificados como vozes normais e os pontos à direita são classificados como vozes patológicas. Não foi utilizado SVM para diferenciar vozes com edema das vozes com nódulo porque o teste de *Tukey* mostrou que esses dois grupos de vozes são estatisticamente iguais.

Figura 47: Classificação utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 1º critério - intervalo: 0.5s - 1.0s.

Na Figura 47(a) verifica-se que todas as vozes normais foram classificadas corretamente com uma voz normal na fronteira e quatro vozes com edema foram classificadas

como normais com dois edemas na fronteira. Na Figura 47(b), todas as vozes normais foram classificadas corretamente e duas vozes com nódulo foram classificadas como normais. Na Figura 47(c), comparando-se vozes normais e patológicas, verifica-se que todas as vozes normais foram classificadas corretamente, seis vozes patológicas foram classificadas como normais com 2 vozes patológicas na fronteira.

Figura 48: Classificação utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 1º critério -intervalo: 2.0s - 2.5s.

Na Figura 48(a) verifica-se que duas vozes normais foram classificadas como edema e 3 vozes com edema foram classificadas como normais. Na Figura 48(b), todas as vozes normais foram classificadas corretamente e 4 vozes com edema foram classificadas como normais. Na Figura 48(c), comparando-se vozes normais e patológicas, verifica-se que todas as vozes normais foram classificadas corretamente, dez vozes patológicas foram classificadas como normais com uma voz patológica na fronteira.

Figura 49: Classificação utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 1º critério -intervalo: 4.0s - 4.5s.

Na Figura 49(a) verifica-se que uma voz normal foi classificada como edema, 4 vozes com edema foram classificadas como normais e uma voz com edema na fronteira. Na Figura 49(b), uma voz normal foi classificada como nódulo, 4 vozes com nódulo foram classificadas como normais com 2 vozes na fronteira. Na Figura 49(c), comparando-se vozes normais e patológicas, verifica-se que uma normal foi classificada como patológica, sete vozes patológicas foram classificadas como normais com duas vozes patológicas na fronteira.

6.1.2 Análise com 2º critério

Os resultados obtidos pelo teste Anova utilizando o segundo critério estão dispostos na Figura 50. A Tabela 14 mostra os valores de p para esse teste. O nível de significância adotado foi de $p=5\%$, ou seja, valores de p menores do que 5% indicam que as amostras são estatisticamente diferentes.

Figura 50: Gráfico comparativo da dispersão dos pontos da seção de Poincaré em relação ao ponto médio em trechos com maior variação de *pitch* - 2° critério.

Tabela 14: Valores de p obtidos com o teste Anova utilizando o 2° critério entre as vozes normais, com edema e com nódulo.

	p
Média	0.0%
Desvio padrão	0.0%

O Teste Anova resultou um valor de $p=0\%$ indicando que as amostras são estatisticamente diferentes. Os resultados obtidos utilizando o teste de *Tukey* para análise do valor médio e desvio padrão de grupos tomados dois a dois estão dispostos na Tabela 15, onde mostra os valores de p para esse teste.

Tabela 15: Valores de p obtidos com o teste de *Tukey* utilizando o 2° critério para as vozes tomadas duas a duas.

	Normal x Edema	Normal x Nódulo	Edema x Nódulo
Média	0.0%	0.0%	54.4%
Desvio padrão	0.0%	0.0%	85.39%

O teste de *Tukey* mostra que foi possível diferenciar vozes normais e patológicas mas não foi possível diferenciar entre patologias.

6.1.2.1 Análise do 2° critério utilizando *Support Vector Machine (SVM)*

Utilizando novamente as medidas de média da dispersão para o eixo x e desvio padrão da dispersão dos pontos da seção de Poincaré para o eixo y vozes normais e patológicas

através de SVM, obteve-se a Figura 51. Não foi utilizado SVM para classificar vozes com edema das vozes com nódulo porque o teste de *Tukey* mostrou que esses dois grupos de vozes são estatisticamente iguais.

Figura 51: Classificação dos grupos de vozes utilizando SVM: Normal x Edema, Normal x Nódulo e Normal x Patológica - 2º critério.

Na Figura 51(a) verifica-se que todas as vozes normais e com edema foram classificadas corretamente. Na Figura 51(b), todas as vozes normais foram classificadas corretamente, 1 voz com nódulo foi classificada como normal com uma voz com edema na fronteira. Na Figura 51(c), comparando-se vozes normais e patológicas, verifica-se que todas as vozes normais foram classificadas corretamente, uma voz patológica foi classificada como normal e uma voz patológica na fronteira.

6.1.3 Comparação entre metodologias e Contribuição

Em Sobrinho (2011), a técnica da seção de Poincaré também foi utilizada para detecção de patologias na laringe. Naquele trabalho foi selecionado um trecho de 500ms de cada sinal de voz, na região considerada de maior estacionariedade em amplitude utili-

zando critério visual. Foi feita a extração da seção de Poincaré em 10 porções do espaço de fase do sinal de voz. Duas medidas foram levantadas: o desvio padrão em relação ao ponto médio da dispersão (DPD) e o desvio padrão em relação aos eixos x e y (DPE). A maior dificuldade encontrada nesse trabalho foi a extração manual de apenas 10 seções de Poincaré em trechos bem comportados do espaço de fase. A técnica proposta nessa tese extrai automaticamente várias seções de Poincaré do espaço de fase. Isso é possível tomando o espaço de fase médio como referência. Nesse caso, cada ponto do espaço de fase médio pertence à sua respectiva seção de Poincaré.

Observando os resultados obtidos com o primeiro critério nos três intervalos estudados para classificação entre vozes normais e patológicas, conclui-se que o melhor intervalo foi entre 0.5s - 1.0s pois a dispersão dos pontos da seção de Poincaré classificou todas as vozes normais corretamente. No entanto, 6 vozes patológicas foram classificadas como normais e duas patológicas ficaram na fronteira da linha que divide vozes normais e patológicas.

Observando os resultados obtidos com o segundo critério na análise entre vozes normais e patológicas, todas as vozes normais foram classificadas corretamente, uma voz patológica foi classificada como normal e apenas uma voz patológica ficou na fronteira. Portanto, os resultados obtidos com o segundo critério demonstraram-se superiores ao primeiro critério pois este último classificou 6 vozes patológicas como normais e o segundo critério classificou apenas uma voz patológica como normal.

A Figura 52 mostra uma comparação do que foi apresentado em Sobrinho (2011) e o proposto nessa tese utilizando o segundo critério. Nessa tese considera-se que as vozes patológicas possuem maiores médias e desvios padrões em relação às vozes normais. Os resultados para o segundo método da metodologia proposta nessa tese, representado na (Figura 52 - esquerda), verifica-se que apenas uma voz patológica foi classificada incorretamente, indicada abaixo da linha que divide vozes normais e patológicas; e todas as vozes normais foram classificadas corretamente. O trabalho proposto em Sobrinho (2011), ilustrado na Figura (52- direita), 8 vozes patológicas foram classificadas como normais e 2 vozes normais foram classificadas como patológicas, de acordo com a linha divisória entre vozes normais e patológicas. Nesse caso, considera-se que as vozes patológicas possuem valores de DPE e DPD maiores do que as vozes normais. Em ambos os casos, a linha divisória foi determinada utilizando *Support Vector Machine*, uma função pré-determinada no *Matlab - versão student*.

Um resultado importante sobre o uso da seção de Poincaré pode ser visto em Dajer et al. (2011)). Nesse trabalho foi possível quantificar uma melhoria da qualidade vocal

Figura 52: Comparação entre a metodologia utilizada proposta nessa tese (esquerda) e utilizada em Sobrinho (2011) (direita).

pós-terapia onde a melhoria da qualidade vocal está relacionada a uma menor dispersão dos pontos na seção de Poincaré.

7 Conclusão e Trabalhos Futuros

Nesse trabalho, uma ferramenta para extração da seção de Poincaré de todo o espaço de fase tridimensional de um sinal de voz foi desenvolvida. Essa ferramenta mostrou-se útil para extrair medidas quantitativas da seção de Poincaré e classificar vozes normais e patológicas. No entanto, ainda não é possível concluir que uma voz patológica, tomando-a isoladamente, possui alguma patologia específica (edema ou nódulo) através da média da dispersão dos pontos da seção de Poincaré pois há vozes normais com valores de dispersão maior do que as 2 patologias estudadas. Ressalta-se que a ferramenta mostrou-se superior ao que foi apresentado em Sobrinho (2011), uma vez que essa última classificou 8 vozes patológicas como normais e duas vozes normais como patológicas. A ferramenta proposta nessa tese utilizando o segundo critério classificou todas as vozes normais corretamente e apenas uma voz patológica foi classificada como normal.

O método proposto mostra-se útil para extrair a frequência fundamental do sinal de voz e supõe-se que seja útil para a extração do *jitter*, um importante parâmetro para análise de sinais de voz e clínica médica. O método proposto também mostrou-se útil para visualizar o *jitter* e o espaço de fase médio, informação útil para uma análise visual do comportamento do espaço de fase através de uma única curva.

A vantagem do método proposto nessa tese é possibilidade de extração de seções de Poincaré de todo o espaço de fase, inclusive em trechos que eram inacessíveis na metodologia proposta em Sobrinho (2011), onde se extraía apenas 10 seções de Poincaré do espaço de fase em regiões bem comportadas.

O cálculo da média e desvio padrão foi extraído de todas as seções de Poincaré ao longo de todo espaço de fase. É possível que haja perda de informações considerando uma média geral de todas as seções. Uma análise da média e desvio padrão de cada seção pode trazer outras informações que tornem possível a diferenciação entre patologias.

As próprias medidas de média e desvio padrão também podem não ser as medidas mais adequadas para análise dos dados. Pode-se extrair outras medidas estatísticas não abordadas nesse trabalho como curtose, média ponderada, média geométrica, etc.

A principal contribuição dessa tese foi diferenciar vozes patológicas das vozes normais através da média e desvio padrão. Verificou-se que vozes patológicas possuem uma média da dispersão maior do que as vozes normais. Sugere-se também que através de terapias de

voz é possível reduzir a média da dispersão dos pontos da seção de Poincaré. No entanto, mais estudos devem ser feitos para explorar melhor a seção de Poincaré do atrator do sinal vocálico para diferenciar patologias específicas.

Nos trabalhos futuros utilizando a seção de *Poincarè* pode-se utilizar matriz de recorrência e outras medidas estatísticas, calcular a média e desvio padrão de todas as seções separadamente, utilizar outras medidas estatísticas (média ponderada, media geométrica, curtose, etc.), calcular o *jitter* do sinal de voz, combinar os parâmetros extraídos da seção de Poincaré com outros parâmetros da análise acústica e não linear como entrada de uma rede neural para diferenciar patologias.

REFERÊNCIAS BIBLIOGRÁFICAS

- ACCARDO, A.; FABBRO, F.; MUMOLO, E. Analysis of normal and pathological voices via short-time fractal dimension. In: **IEEE. Engineering in Medicine and Biology Society, 1992 14th Annual International Conference of the IEEE**. [S.l.], 1992. v. 3, p. 1270–1271.
- ALMEIDA, N.; SOARES, H.; BRESOLIN, A. de A. Performance of wavelets in the diagnosis of laryngeal pathologies. 2012.
- ALONSO, J. B. et al. **Using nonlinear features for voice disorder detection**: in proc. 3rd int. conf. nonlinear speech process. 2005. 94-106 p.
- ALONSO, J. B. et al. Using nonlinear features for voice disorder detection. In: **ISCA Tutorial and Research Workshop (ITRW) on Non-Linear Speech Processing**. [S.l.: s.n.], 2005.
- ARAÚJO, S. A. et al. Normatização de medidas acústicas da voz normal. **Rev Bras Otorrinolaringol**, SciELO Brasil, v. 68, n. 4, p. 540–4, 2002.
- ARJMANDI, M. K.; POOYAN, M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. **Biomedical Signal Processing and Control**, Elsevier, v. 7, n. 1, p. 3–19, 2012.
- ARROYAVE, J. R. O.; BONILLA, J. F. V.; TREJOS, E. D. Acoustic analysis and non linear dynamics applied to voice pathology detection: A review. **Recent Patents on Signal Processing**, Bentham Science Publishers, v. 2, n. 2, p. 96–107, 2012.
- AS-BROOKS, C. J. van et al. Acoustic signal typing for evaluation of voice quality in tracheoesophageal speech. v. 20, n. 3, 2006.
- AWREJCEWICZ, J. Bifurcation portrait of the human vocal cord oscillations. **Journal of Sound and Vibration**, Academic Press, v. 136, n. 1, p. 151–156, 1990.
- BAKEN, R. Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice. **Journal of Voice**, Elsevier, v. 4, n. 3, p. 185–197, 1990.
- BASSI, R. D. S. **Identificação inteligente de patologias no trato vocal**. Tese (Doutorado) — Universidade de São Paulo, 2014.
- BEHLAU, M. **A voz. O livro do especialista**. [S.l.: s.n.], 2001.
- BERG, J. Van den. Myoelastic-aerodynamic theory of voice production. **Journal of Speech, Language and Hearing Research**, ASHA, v. 1, n. 3, p. 227, 1958.
- BERG, V. D. Myoelastic - aerodynamic theory of voice production. p. 227–244, 1958.

- BERRY, D. A. et al. Bifurcations in excised larynx experiments. **Journal of Voice**, Elsevier, v. 10, n. 2, p. 129–138, 1996.
- BIRKENT, H. et al. Results of voice therapy in functional voice disorders]. **Kulak burun boğaz ihtisas dergisi: KBB= Journal of ear, nose, and throat**, v. 12, n. 5-6, p. 120, 2004.
- BOYANOV, B. Analysis of pathological voice. **Report under contract ERB-CIPA-CT-92-0170 with the European Community**, 1993.
- BOYANOV, B.; HADJITODOROV, S. Acoustic analysis of pathological voices. a voice analysis system for the screening of laryngeal diseases. **Engineering in Medicine and Biology Magazine, IEEE**, IEEE, v. 16, n. 4, p. 74–82, 1997.
- BROCKMANN-BAUSER, M.; DRINNAN, M. J. Routine acoustic voice analysis: time to think again? **Current opinion in otolaryngology & head and neck surgery**, LWW, v. 19, n. 3, p. 165–170, 2011.
- BROOMHEAD, D.; KING, G. P. On the qualitative analysis of experimental dynamical systems. **Nonlinear Phenomena and Chaos**, Adam Hilger, v. 113, p. 114, 1986.
- CAMPISI, P. et al. Multidimensional voice program analysis in children with vocal cord nodules. **The Journal of otolaryngology**, v. 29, n. 5, p. 302–308, 2000.
- CARVALHO, R. T. S.; CAVALCANTE, C. C.; CORTEZ, P. C. Wavelet transform and artificial neural networks applied to voice disorders identification. In: **IEEE. Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on**. [S.l.], 2011. p. 371–376.
- COSTA, S. et al. Uso da entropia na discriminação de vozes patológicas. In: **II Congresso de Inovação da Rede Norte Nordeste de Educação Tecnológica, João Pessoa, Paraíba**. [S.l.: s.n.], 2007.
- COSTA, W. C. A.; NETO, B. G. A.; COSTA, S. C. Uso da dimensão de correlação na discriminação entre vozes saudáveis e patológicas. In: **V CONNEPI-2010**. [S.l.: s.n.], 2010.
- DAJER, M. et al. Vocal dynamic visual pattern for voice characterization. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2011. v. 332, n. 1, p. 012026.
- DAJER, M.; SOBRIHNO, F. A. A.; PEREIRA, J. C. Técnicas não lineares e convencionais para análise de vozes de crianças. In: **XVII Congreso Argentino de Bioingenieria SABI**. [S.l.: s.n.].
- DAJER, M. E. Análise de sinais de voz por padrões visuais da dinâmica vocal. **São Paulo: Universidade de São Paulo–Escola de Engenharia de São Carlos**.
- DAJER, M. E.; PEREIRA, J. C.; MACIEL, C. D. Nonlinear dynamical analysis of normal voices. In: **IEEE. Multimedia, Seventh IEEE International Symposium on**. [S.l.], 2005. p. 6–pp.

- DAJER, M. E.; SOBRINHO, F. A. A.; PEREIRA, J. C. Técnicas no lineales y técnicas tradicionales para análisis de la calidad de voz en niños. In: **XI Congreso Argentino de Otorrinolaringología y Fonoaudiología Pediátrica, Argentina, Salta**. [S.l.: s.n.], 2009b.
- DAJER, M. E.; SOBRINHO, F. A. A.; PEREIRA, J. C. Qualitative and quantitative analysis of vocal dynamic visual patterns. In: PANHELLENIC ASSOCIATION OF LOGOPEDISTS, 2010. **28th International Congress of the International Association of Logopedists and Phoniatrics (IALP)**, Athens. [S.l.], 2010. I, p. 789–794.
- DAJER, M. E. et al. Análise não linear identifica evolução na terapia de voz. In: **19 congresso brasileiro de fonoaudiologia e 8 internacional de fonoaudiologia, 2011, São Paulo**. [S.l.: s.n.], 2011.
- DAVIS, S. **Acoustic Characteristics of Normal and Pathological Voices. Speech and Language Research and Theory**. [S.l.: s.n.], 1979.
- DELIYSKI, D. D. Acoustic model and evaluation of pathological voice production. In: **Eurospeech**. [S.l.: s.n.], 1993. v. 93, p. 1969–1972.
- DELLER, J. R.; PROAKIS, J. G.; HANSEN, J. H. **Discrete-time processing of speech signals**. [S.l.]: Ieee New York, NY, USA:, 2000.
- DIBAZAR, A. A.; NARAYANAN, S.; BERGER, T. W. Feature analysis for automatic detection of pathological speech. In: IEEE. **Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint**. [S.l.], 2002. v. 1, p. 182–183.
- ECKMANN, J.-P. et al. Liapunov exponents from time series. **Physical Review A**, v. 34, n. 6, p. 4971–4979, 1986.
- FANT, C. **Acoustic theory of Speech Production**. [S.l.: s.n.], 1960.
- FANT, G. **Acoustic theory of speech production**. [S.l.]: Walter de Gruyter, 1970.
- FLANAGAN, J. L. **Speech analysis, synthesis and perception**. 2a ed. ed. [S.l.: s.n.], 1972.
- FONSECA, E. S. et al. Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. **Computers in Biology and Medicine**, Elsevier, v. 37, n. 4, p. 571–578, 2007.
- FRASER, A.; SWINNEY, H. Using mutual information to find independent coordinates for strange attractors. **Phys. Rev. A**, v. 33, p. 1134–1140, 1986.
- FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. **Physical review A**, APS, v. 33, n. 2, p. 1134, 1986.
- GIOVANNI, A.; OUAKNINE, M.; TRIGLIA, J.-M. Determination of largest lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis. **Journal of Voice**, Elsevier, v. 13, n. 3, p. 341–354, 1999.

- GODINO-LLORENTE, J. I. et al. Support vector machines applied to the detection of voice disorders. In: **Nonlinear Analyses and Algorithms for Speech Processing**. [S.l.]: Springer, 2005. p. 219–230.
- GRASSBERGER, P.; PROCACCIA, I. Measuring the strangeness of strange attractors. **Physica D: Nonlinear Phenomena**, Elsevier, v. 9, n. 1, p. 189–208, 1983.
- GUIDO, R. C. et al. Support vector machines and wavelets for voice disorder sorting. In: **IEEE. System Theory, 2006. SSST'06. Proceeding of the Thirty-Eighth Southeastern Symposium on**. [S.l.], 2006. p. 434–438.
- HAGMÜLLER, M. **Speech Enhancement for Disordered and Substitution Voices**. Tese (Doutorado) — Citeseer, 2009.
- HARIHARAN, M.; POLAT, K.; YAACOB, S. A new feature constituting approach to detection of vocal fold pathology. **International Journal of Systems Science**, Taylor & Francis, n. ahead-of-print, p. 1–13, 2013.
- HEGGER, R.; KANTZ, H.; SCHREIBER, T. Practical implementation of nonlinear time series methods: The tisean package. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, American Institute of Physics, v. 9, n. 2, p. 413–435, 1999.
- HELMHOLTZ, H. von. **Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik, von H. Helmholtz,... 2te Ausgabe**. [S.l.]: F. Vieweg und Sohn, 1865.
- HENRÍQUEZ, P. et al. Characterization of healthy and pathological voice through measures based on nonlinear dynamics. **Audio, Speech, and Language Processing, IEEE Transactions on**, IEEE, v. 17, n. 6, p. 1186–1195, 2009.
- HERZEL, H. Bifurcations and chaos in voice signals. **Applied Mechanics Reviews**, v. 46, p. 399, 1993.
- HERZEL, H. et al. Analysis of vocal disorders with methods from nonlinear dynamics. **Journal of Speech, Language and Hearing Research**, ASHA, v. 37, n. 5, p. 1008, 1994.
- HERZEL, H. et al. Analysis of vocal disorders with methods from nonlinear dynamics. **Journal of Speech, Language and Hearing Research**, ASHA, v. 37, n. 5, p. 1008, 1994.
- HERZEL, H. et al. Detecting bifurcations in voice signals. In: **Nonlinear analysis of physiological data**. [S.l.]: Springer, 1998. p. 325–344.
- HERZEL, H.; WENDLER, J. Evidence of chaos in phonatory samples. In: **Second European Conference on Speech Communication and Technology**. [S.l.: s.n.], 1991.
- HIRANO, M. et al. Sulcus vocalis: functional aspects. **The Annals of otology, rhinology & laryngology**, Annals Publishing Compagny, v. 99, n. 9, p. 679–683, 1990.

- ISHIZAKA, K. Synthesis of voiced sounds from a two-mass model of the vocal. **Cords. Bell Syst. Tech. J.**, v. 51, n. 6, p. 1233–1268, 1972.
- ISHIZKA, K.; FLANAGAN, J. L. Synthesis of voice sounds from a two-mass model of the vocal cords. v. 50, n. 6, p. 1233–1268, 1972.
- JAFARI, A.; ALMASGANJ, F.; BIDHENDI, M. N. Statistical modeling of speech poincaré sections in combination of frequency analysis to improve speech recognition performance. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, American Institute of Physics, v. 20, n. 3, p. 033106–033106, 2010.
- JESUS, B. A.; JOSÉ, A. I. de L.; MIGUEL, A. F. Automatic detection of pathologies in the voice by hos based parameters. **EURASIP Journal on Advances in Signal Processing**, 2001.
- JIANG, J.; ZHANG, Y. Nonlinear dynamic analysis of speech from pathological subjects. **Electronics Letters**, IET, v. 38, n. 6, p. 294–295, 2002.
- JIANG, J. J. et al. Objective acoustic analysis of pathological voices from patients with vocal nodules and polyps. **Folia Phoniatica et Logopaedica**, Karger Publishers, v. 61, n. 6, p. 342–349, 2009.
- JIANG, J. J.; ZHANG, Y.; MCGILLIGAN, C. Chaos in voice, from modeling to measurement. **Journal of Voice**, Elsevier, v. 20, n. 1, p. 2–17, 2006.
- KANTZ, H.; SCHREIBER, T. Dimension estimates and physiological data. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, v. 5, n. 1, p. 143–154, 1995.
- KANTZ, H.; SCHREIBER, T. **Nonlinear time series analysis**. [S.l.]: Cambridge university press, 2004.
- KASUYA, H.; ENDO, Y.; SALIU, S. Novel acoustic measurements of jitter and shimmer characteristics from pathological voice. In: **Third European Conference on Speech Communication and Technology**. [S.l.: s.n.], 1993.
- KASUYA, H. et al. Normalized noise energy as an acoustic measure to evaluate pathologic voice. **The Journal of the Acoustical Society of America**, v. 80, p. 1329, 1986.
- KHALIL, H. K. **Nonlinear systems**. [S.l.]: Prentice hall Upper Saddle River, 2002.
- KINSNER, W. Characterizing chaos through lyapunov metrics. **Systems, Man, and Cybernetics, Part C: Applications and Reviews**, IEEE Transactions on, IEEE, v. 36, n. 2, p. 141–151, 2006.
- KROM, G. d. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. **Journal of Speech, Language and Hearing Research**, ASHA, v. 36, n. 2, p. 254, 1993.
- KUBIN, G. Poincaré section techniques for speech. In: IEEE. **Speech Coding For Telecommunications Proceeding, 1997, 1997 IEEE Workshop on**. [S.l.], 1997. p. 7–8.

- KUMAR, A.; MULLICK, S. Attractor dimension, entropy and modelling of speech time series. **Electronics Letters**, IET, v. 26, n. 21, p. 1790–1792, 1990.
- KUMAR, A.; MULLICK, S. K. Nonlinear dynamical analysis of speech. v. 100, n. 1, p. 615–629, 1996.
- LIEBERMAN, P. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. **The Journal of the Acoustical Society of America**, v. 35, p. 344, 1963.
- LIM, J.-Y. et al. Voice analysis of patients with diverse types of reinke's edema and clinical use of electroglottographic measurements. **Acta oto-laryngologica**, Informa UK Ltd UK, v. 126, n. 1, p. 62–69, 2006.
- LIN, S. et al. Phase space reconstruction of nonlinear time series based on kernel method. In: IEEE. **Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on**. [S.l.], 2006. v. 1, p. 4364–4368.
- LINDGREN, A. C.; JOHNSON, M. T.; POVINELLI, R. J. Speech recognition using reconstructed phase space features. In: IEEE. **Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on**. [S.l.], 2003. v. 1, p. I–60.
- LITTLE, M. et al. Nonlinear, biophysically-informed speech pathology detection. In: IEEE. **Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on**. [S.l.], 2006. v. 2, p. II–II.
- LUDLOW, C. et al. The validity of using phonatory jitter and shimmer to detect laryngeal pathology. **Laryngeal function in phonation and respiration**, p. 492–508, 1987.
- MA, Y.; GU, M.-l. Pitch detection method based on poincare sections for speech signals. **Audio Engineering**, v. 8, p. 011, 2006.
- MANN, I.; MCLAUGHLIN, S. Poincare maps and pitch detection in speech. In: IET. **Signals Systems and Chaos (Ref. No. 1997/393), IEE Colloquium on**. [S.l.], 1997. p. 5–1.
- MARTINS, R. H.; TRINDADE, S. H. A criança disfônica: diagnóstico, tratamento e evolução clínica. **Rev Bras Otorrinolaringol**, SciELO Brasil, v. 69, n. 6, p. 801–6, 2003.
- MATASSINI, L. et al. Analysis of vocal disorders in a feature space. **Medical engineering & physics**, Elsevier, v. 22, n. 6, p. 413–418, 2000.
- MATSUO, K.; KAMIMURA, M.; HIRANO, M. Polypoid vocal folds. a 10-year review of 191 patients. **Auris Nasus Larynx**, Elsevier, v. 10, p. S37–S45, 1983.
- MENDE, W.; HERZEL, H.; WERMKE, K. Bifurcations and chaos in newborn infant cries. **Physics Letters A**, Elsevier, v. 145, n. 8, p. 418–424, 1990.

- MICHAELIS, D.; GRAMSS, T.; STRUBE, H. W. Glottal-to-noise excitation ratio a new measure for describing pathological voices. **Acta Acustica united with Acustica**, S. Hirzel Verlag, v. 83, n. 4, p. 700–706, 1997.
- MONTAGNOLI, A.; PEREIRA, J. C. Análise de voz. versão 4.10. 2009.
- MORENTE, J. C. et al. Estudio objetivo de la voz en población normal y en la disfonía por nódulos y pólipos vocales. **Acta Otorrinolaringológica Española**, Elsevier, v. 52, n. 6, p. 476–482, 2001.
- MURUGESAPANDIAN, P.; YAACOB, S.; HARIHARAN, M. Feature extraction based on mel-scaled wavelet packet transform for the diagnosis of voice disorders. In: SPRINGER. **4th Kuala Lumpur International Conference on Biomedical Engineering 2008**. [S.l.], 2008. p. 790–793.
- NARAYANAN, S. S.; ALWAN, A. A. A nonlinear dynamical systems analysis of fricative consonants. **The Journal of the Acoustical Society of America**, v. 97, p. 2511, 1995.
- NEVES, B. M.; NETO, J. G.; PONTES, P. Diferenciação histopatológica e imunoistoquímica das alterações epiteliais no nódulo vocal em relação aos pólipos e ao edema de laringe. **Rev Bras Otorrinolaringol**, SciELO Brasil, v. 70, n. 4, p. 439–48, 2004.
- NEVES, J. A. C.; FEMANDES, J.; RESTIVO, F. Phase space signal filtering. In: IEEE. **Industrial Technology, 2006. ICIT 2006. IEEE International Conference on**. [S.l.], 2006. p. 1805–1809.
- OTT, E. **Chaos in dynamical systems**. [S.l.]: Cambridge university press, 2002.
- PACKARD, N. et al. **Geometry from a time series**. 1980.
- PARKER, T. S.; CHUA, L. O. Chaos: A tutorial for engineers. **Proceedings of the IEEE**, IEEE, v. 75, n. 8, p. 982–1008, 1987.
- PARKER, T. S.; CHUA, L. O.; PARKER, T. S. **Practical numerical algorithms for chaotic systems**. [S.l.]: Springer New York, 1989.
- PARRAGA, A. Aplicação da transformada wavelet packet na análise e classificação de sinais de vozes patológicas. 2002.
- PETRÔNIO, A. extração e avaliação da dinâmica não linear da voz utilizando vídeo de alta velocidade da laringe. **São Paulo: Universidade de São Paulo–Escola de Engenharia de São Carlos**, 2012.
- PINHO, S.; PONTES, P. Escala de avaliação perceptiva da fonte glótica: Rasat. In: **Vox Brasilis - ano 8**. [S.l.: s.n.], 2002.
- PITSIKALIS, V.; KOKKINOS, I.; MARAGOS, P. Nonlinear analysis of speech signals: generalized dimensions and lyapunov exponents. In: CITESEER. **INTERSPEECH**. [S.l.], 2003.

- PROPEs, N. C. Poincare sections and the rl-diode circuit. In: IEEE. **System Theory, 1996., Proceedings of the Twenty-Eighth Southeastern Symposium on.** [S.l.], 1996. p. 195–197.
- RABINER, L. R.; SCHAFER, R. W. **Digital processing of speech signals.** [S.l.]: IET, 1979.
- RABINER, L. R.; SCHAFER, R. W. Introduction to digital speech processing. **Foundations and trends in signal processing**, Now Publishers Inc., v. 1, n. 1, p. 1–194, 2007.
- ROSA, M. de O.; PEREIRA, J. C.; GRELLET, M. Adaptive estimation of residue signal for voice pathology diagnosis. **Biomedical Engineering, IEEE Transactions on**, IEEE, v. 47, n. 1, p. 96–104, 2000.
- ROUX, J.-C.; SIMOYI, R. H.; SWINNEY, H. L. Observation of a strange attractor. **Physica D: Nonlinear Phenomena**, Elsevier, v. 8, n. 1, p. 257–266, 1983.
- SAEEDI, N. E.; ALMASGANJ, F. Wavelet adaptation for automatic voice disorders sorting. **Computers in biology and medicine**, Elsevier, 2013.
- SAEEDI, N. E.; ALMASGANJ, F.; TORABINEJAD, F. Support vector wavelet adaptation for pathological voice assessment. **Computers in Biology and Medicine**, Elsevier, v. 41, n. 9, p. 822–828, 2011.
- SÁENZ-LECHÓN, N. et al. Methodological issues in the development of automatic systems for voice pathology detection. **Biomedical Signal Processing and Control**, Elsevier, v. 1, n. 2, p. 120–128, 2006.
- SALHI, L.; MOURAD, T.; CHERIF, A. Voice disorders identification using multilayer neural network. **pathology**, v. 2, n. 7, p. 8, 2008.
- SANO, M.; SAWADA, Y. Measurement of the lyapunov spectrum from a chaotic time series. **Physical review letters**, APS, v. 55, n. 10, p. 1082, 1985.
- SCALASSARA, P. Utilização de medidas de previsibilidade em sinais de voz para discriminação de patologias da laringe. **São Paulo: Universidade de São Paulo–Escola de Engenharia de São Carlos**, 2009.
- SCALASSARA, P. R. et al. Voice signals characterization through entropy measures. In: **BIOSIGNALS (2)**. [S.l.: s.n.], 2008. p. 163–170.
- SCALASSARA, P. R. et al. Relative entropy measures applied to healthy and pathological voice characterization. **Applied Mathematics and Computation**, Elsevier, v. 207, n. 1, p. 95–108, 2009.
- SCALASSARA, P. R. et al. Analysis of voice pathology evolution using entropy rate. In: IEEE. **Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on.** [S.l.], 2008. p. 580–585.
- SOBRINHO, F. A. de A. Medida da dispersão da periodicidade de um sinal de voz normal e voz patológica através da seção de poincaré. **São Paulo: Universidade de São Paulo–Escola de Engenharia de São Carlos**, 2011.

- SOBRINHO, F. A. de A.; DAJER, M. E.; ALBERTO, L. F. C. Técnica de extração da seção de poincaré para análise de sinais de voz: Dincon conferencia brasileira de dinâmica, controle e aplicações. 2015.
- SOBRINHO, F. A. de A.; DAJER, M. E.; ALBERTO, L. F. C. Estimation of pitch period using poincare section: 6th international conference on nonlinear science and complexity. 2016.
- STEINECKE, I.; HERZEL, H. Bifurcations in an asymmetric vocal-fold model. **The Journal of the Acoustical Society of America**, v. 97, p. 1874, 1995.
- STEPP, C. E. et al. Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction: Journal of speech, language and hearing research. v. 54, p. 1260–1266, 2011.
- TAKENS, F. Detecting strange attractors in turbulence. In: **Dynamical systems and turbulence, Warwick 1980**. [S.l.]: Springer, 1981. p. 366–381.
- TALKIN, D. A robust algorithm for pitch tracking (rapt). **Speech coding and synthesis**, New York: Elsevier, v. 495, p. 518, 1995.
- TITZE, I. R.; BAKEN, R.; HERZEL, H. Evidence of chaos in vocal fold vibration. **Vocal fold physiology: New frontier in basic science**, Singular, p. 143–188, 1993.
- TITZE, I. R.; BAKEN, R.; HERZEL, H. **Evidence of chaos in vocal fold vibration. Vocal Fold Physiology**. [S.l.: s.n.], 1993.
- VIEIRA, M. N.; MCINNES, F. R.; JACK, M. A. Comparative assessment of electroglottographic and acoustic measures of jitter in pathological voices. **Journal of Speech, Language and Hearing Research**, ASHA, v. 40, n. 1, p. 170, 1997.
- WOLF, A. et al. Determining lyapunov exponents from a time series. **Physica D: Nonlinear Phenomena**, Elsevier, v. 16, n. 3, p. 285–317, 1985.
- YANAGIHARA, N. Significance of harmonic changes and noise components in hoarseness. **Journal of Speech, Language and Hearing Research**, ASHA, v. 10, n. 3, p. 531, 1967.
- YUMOTO, E.; GOULD, W. J.; BAER, T. Harmonics-to-noise ratio as an index of the degree of hoarseness. **The Journal of the Acoustical Society of America**, v. 71, p. 1544, 1982.
- YUNIK, M.; BOYANOV, B. **Method for evaluation of the noise-to-harmonic-component ratios in pathological and normal voices**. [S.l.]: S HIRZEL VERLAG POSTFACH 10 10 61, D-70 009 STUTTGART, GERMANY, 1990. 89–91 p.
- ZALESSKA-KRECICKA, M.; KRECICKI, T.; CYGANIEK, P. A clinical study of reinke's edema]. **Otolaryngologia polska. The Polish otolaryngology**, v. 47, n. 2, p. 153, 1993.

ZHANG, Y.; JIANG, J. J. **Nonlinear dynamic analysis in signal typing of pathological human voices:** Electron. lett. Jun. 2003. 1021-1023 p.

ZHANG, Y. et al. Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps. **The Journal of the Acoustical Society of America**, v. 115, p. 2270, 2004.