



Samuel Ribeiro de Abreu

**AVALIAÇÃO DA TRANSFORMADA DE
HILBERT-HUANG NA DETECÇÃO DE
DESVIOS VOCAIS**

João Pessoa - PB

Março de 2019



Samuel Ribeiro de Abreu

AVALIAÇÃO DA TRANSFORMADA DE HILBERT-HUANG NA DETECÇÃO DE DESVIOS VOCAIS

Dissertação apresentada à banca examinadora do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal da Paraíba, como requisito necessário à obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Processamento Digital de Sinais

Suzete Élide Nóbrega Correia, Dra.

Orientadora

Silvana Luciene do Nascimento Cunha Costa, Dra.

Coorientadora

João Pessoa - PB, Março de 2019

©Samuel Ribeiro de Abreu- abreu.s.ribeiro@gmail.com

Dados Internacionais de Catalogação na Publicação – CIP
Biblioteca Nilo Peçanha – IFPB, *Campus* João Pessoa

A162a Abreu, Samuel Ribeiro.

Avaliação da transformada de Hilbert-Huang na detecção de desvios vocais / Samuel Ribeiro de Abreu. – 2019.

103 f. : il.

Dissertação (Mestrado em Engenharia Elétrica) – Instituto Federal da Paraíba – IFPB / Coordenação de Pós-Graduação em Engenharia Elétrica, 2019.

Orientadora: Prof.^a Suzete Élda Nóbrega Correia

Coorientadora: Prof.^a Silvana Luciene do Nascimento C. Costa

1. Engenharia elétrica. 2. Processamento digital de sinais - DSP. 3. Análise de sinais. 4. Transformada de Hilbert-Huang. 5. Desvios vocais. 6. Análise em tempo-frequência. I. Título.

CDU 621.391

Ivanise Andrade M. de Almeida
Bibliotecária-Documentalista
CRB-15/0096

Samuel Ribeiro de Abreu

AVALIAÇÃO DA TRANSFORMADA DE HILBERT-HUANG NA DETECÇÃO DE DESVIOS VOCAIS

Dissertação apresentada à banca examinadora do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal da Paraíba, como requisito necessário à obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Dissertação aprovada pela banca examinadora em: 25/03/2019

BANCA EXAMINADORA



Suzete Élide Nóbrega Correia, Dra. – IFPB
Orientadora



Silvana Luciene do Nascimento Cunha Costa, Dra. – IFPB
Coorientadora



Carlos Danilo Miranda Regis, Dr. – IFPB
Examinador Interno



Leonardo Wanderley Lopes, Dr. – UFPB
Examinador Externo



Vinícius Jefferson Dias Vieira, Dr. – UFPB
Examinador Externo

João Pessoa - PB

Março de 2019

Para minha querida e amada esposa, Yana.

AGRADECIMENTOS

A Deus, por Seu cuidado gracioso para comigo desde meu nascimento até hoje, de forma inexplicável Sua Graça me alcançou, por colocar em meu coração sonhos, ideias e inspirações e dar as condições e meios para que pudesse realizá-los. Sou grato a Ele, por ter me capacitado, me dado sabedoria e colocado em minha vida todas as pessoas as quais presto os meus sinceros agradecimentos.

A minha amada esposa, Yana pelo amor, zelo, aliança, compreensão, carinho, tolerância e paciência nesses 11 anos de caminhada juntos. Obrigado por sempre me abrir os olhos e mostrar novas possibilidades, por acreditar e estimular o meu melhor. Ao nosso filho João Pedro, que apesar da sua idade, sempre tentou entender as minhas privações e dedicação para finalização desse estudo. A nossa filha Rebeca, que requer todo nosso amor e tempo disponível.

Aos meu amados pais, Luís Carlos e Lucinete, aos quais devo obediência e honra. Essa conquista jamais será individual. Obrigado por toda educação, amor, apoio, correções e preocupação.

As minhas professoras e orientadoras, Suzete Correia e Silvana Costa, pelos ensinamentos, correções e pelas oportunidades que me deram para trabalhar em seus projetos de pesquisa. Pelo apoio e pela disponibilidade de sempre tentarem me ajudar.

Aos membros da Banca avaliadora desta Dissertação, pela disponibilidade e por suas valiosas contribuições.

E a Capes, pelo suporte financeiro.

RESUMO

O distúrbio vocal é identificado por qualquer dificuldade ou alteração na emissão vocal que dificulta a produção natural de voz, não cumprindo a transmissão da mensagem verbal e/ou emocional. Tais distúrbios podem influenciar negativamente na qualidade de vida de um indivíduo, podendo limitar a comunicação no trabalho, como em outros aspectos sociais. O diagnóstico de uma alteração vocal é um processo que precisa combinar diversas técnicas de avaliação e análise. As técnicas de análise de sinais em tempo-frequência são apropriadas para estudar os sinais biomédicos, como a voz, pois são sinais que se caracterizam por apresentar conteúdo relevante tanto no domínio do tempo quanto da frequência. Um método para análise de sinais não lineares e não estacionários é a transformada de Hilbert-Huang. Este estudo objetivou avaliar a aplicabilidade da transformada de Hilbert-Huang na detecção de desvios vocais. Foram considerados dois estudos de caso para a aplicação da transformada de Hilbert-Huang: 1) Análise acústica de sinais de vozes saudáveis e desviadas (rugosidade, sopro e tensão); e 2) Análise acústica do grau de distúrbios. A base de dados foi cedida pelo Laboratório Integrado de Estudos da Voz, do Departamento de Fonoaudiologia da Universidade Federal da Paraíba, da qual foram selecionados 116 sinais de voz de pessoas do sexo masculino e feminino com idade superior a 18 anos e inferior a 65 anos. Como etapas do processo, foram realizadas extração de características por meio da transformada de Hilbert-Huang; a avaliação do potencial discriminativo por meio de análises estatísticas através de testes de hipóteses; classificação dos sinais usando um classificador *Multilayer Perceptron* (MLP) com o algoritmo de aprendizado supervisionado do gradiente conjugado escalonado. Foi avaliado o desempenho do sistema de classificação e verificou-se que a característica individual com maior potencial discriminativo foi a amplitude instantânea ponderada pela frequência instantânea, extraída na quinta função intrínseca de modo, discriminando vozes normais e desviadas, com acurácia de $92,67\% \pm 4,52\%$. Quando as características foram combinadas, atingiu-se uma acurácia de 100%, na discriminação entre os sinais de vozes normais e com desvio sopro e entre sinais de vozes normais e com desvio rugosidade. Diante dos resultados obtidos, sugere-se o uso da transformada de Hilbert-Huang por parte dos profissionais e pesquisadores na área de voz, uma vez que por meio desta é possível auxiliar no processo de análise, classificação e tomada de decisão clínica no diagnóstico e no tratamento de pacientes com distúrbios vocais.

Palavras-chave: Desvios Vocais, Análise em Tempo-Frequência, Transformada de Hilbert-Huang, Análise Acústica.

ABSTRACT

The voice disorder is identified by any difficulty or alteration in vocal emission that hampers the natural production of voice, not fulfilling the transmission of the verbal and/or emotional message. These disorders may negatively influence the quality of life of an individual, and may limit communication at work, as in other social aspects. The diagnosis of a voice disorder is a process that must combine several techniques of evaluation and analysis. Time-frequency signal analysis techniques are appropriate for studying biomedical signals, such as voice, since they are signals that are characterized by having relevant content in both time and frequency domain. A method for analysis of non-linear and non-stationary signals is the Hilbert-Huang transform. This study aimed to evaluate the applicability of the Hilbert-Huang transform in the detection of vocal deviations. In this research, two case studies were considered for the Hilbert-Huang transform: 1) Acoustic analysis of healthy and vocal deviations signals (roughness, breathiness and strain); and 2) Acoustic analysis of the intensity of degrees voice disorders. The database was provided by the Integrated Laboratory of Voice Studies, Department of Speech Language Pathology of the Federal University of Paraíba, from which they were selected 116 voice signals of males and females aged over 18 and under 65 years. As process steps, the characteristics were extracted through the HilbertHuang transform; the evaluation of the discriminative potential through statistical analysis through hypothesis tests; classification of the signals using an Multilayer Perceptron (MLP) classifier with the supervised learning algorithm of the stepped conjugate gradient. The performance of the classification system was evaluated and it was verified that the individual characteristic with the greatest discriminative potential was the instantaneous power spectral density measure extracted in the fifth intrinsic function so as to discriminate between healthy and disturbed voices with an accuracy of $92.67\% \pm 4.52\%$. When the characteristics were combined, was reached an accuracy of 100%, for discriminating both between healthy and breathy and between healthy and rough voices. In view of the results obtained, we suggest the use of the Hilbert-Huang transform by professionals and researchers in the voice specialist, since it is possible to aid in the process of analysis, classification and clinical decision making in the diagnostic process, treatment of patients with voice disorders.

Keywords: *Vocal Deviations, Time-Frequency Analysis, Hilbert-Huang Transform, Acoustic Analysis.*

LISTA DE FIGURAS

Figura 3.1 – Grau da alteração vocal na escala EAV.	30
Figura 4.1 – Fluxograma do algoritmo do método EMD.	33
Figura 4.2 – Exemplo de um sinal no processo de <i>sifting</i> obtido após k iterações.	34
Figura 4.3 – Exemplo de IMFs e resíduos obtidos após método EMD de um sinal de voz normal.	35
Figura 4.4 – Exemplo do espectro de Hilbert para as 10 primeiras IMFs de um sinal de voz com desvio tensão.	39
Figura 5.1 – Diagrama da metodologia empregada.	43
Figura 5.2 – Quatro sinais de vozes decompostos em 10 IMFs. Da esquerda para a direita: voz normal, voz com desvio rugosidade, voz com desvio soproidade e voz com desvio tensão.	44
Figura 5.3 – Espectro de Hilbet-Huang em relação a IMF ₂ dos sinais de vozes normal, com desvio rugosidade, soproidade e tensão.	45
Figura 5.4 – Espectro de Hilbet-Huang em relação a IMF ₆ dos sinais de vozes normal, com desvio rugosidade, soproidade e tensão.	46
Figura 5.5 – Espectro de Hilbet-Huang em relação a IMF ₈ dos sinais de vozes normal e com desvio tensão.	47
Figura 5.6 – Rede <i>feedforward</i> MLP utilizada.	49
Figura 6.1 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvios: (a) características E_5 ; (b) característica APF_5	53
Figura 6.2 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvio rugosidade: (a) característica E_5 ; (b) característica $DETI_5$, (c) característica $DETI_6$, (d) característica APT_5 , (e) característica APT_6 , (f) característica DFI_4 , (g) característica DFI_{10} , (h) característica APF_1 , (i) característica APF_5 , (j) característica APF_6	55
Figura 6.3 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvio soproidade: (a) característica E_2 ; (b) característica E_3 , (c) característica $DETI_2$, (d) característica $DETI_3$, (e) característica APT_2 , (f) característica APT_3 , (g) característica DFI_3 , (h) característica DFI_{10} , (i) característica APF_3	57
Figura 6.4 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvio tensão: (a) características DFI_1 ; (b) característica DFI_{10}	58

Figura 6.5 – Médias e desvios estatísticos das características dos sinais de vozes com desvio rugosidade e soproisidade: (a) característica E_1 , (b) característica E_2 , (c) característica E_5 , (d) característica E_6 , (e) característica $DETI_1$, (f) característica $DETI_2$, (g) característica $DETI_3$, (h) característica $DETI_5$, (i) característica $DETI_6$, (j) característica APT_2 , (k) característica APT_5 , (l) característica APT_6 , (m) característica DFI_1 , (n) característica DFI_4 , (o) característica DFI_6 , (p) característica DFI_7 , (q) característica APF_2 , (r) característica APF_5	60
Figura 6.6 – Médias e desvios estatísticos das características dos sinais de vozes com desvio rugosidade e tensão: (a) características APF_5 ; (b) característica DFI_3 ; (c) característica DFI_8	61
Figura 6.7 – Médias e desvios estatísticos das características dos sinais de vozes com desvio soproisidade e tensão: (a) característica E_2 , (b) característica E_3 , (c) característica E_7 , (d) característica $DETI_1$, (e) característica $DETI_2$, (f) característica $DETI_3$, (g) característica APT_2 , (h) característica APT_7 , (i) característica DFI_3 , (j) característica DFI_7	63
Figura 6.8 – Médias e desvios das características dos sinais de vozes com intensidades de Grau 1 e Graus 2 e 3: (a) características DFI_1 ; (b) característica APF_5	64
Figura 6.9 – Médias e desvios da característica DFI_5 dos sinais de vozes com grau geral 1 e grau geral 2.	65
Figura 6.10–Médias e desvios das características dos sinais de vozes G1 e G3: (a) característica APT_1 ; (b) característica APT_6 , (c) característica DFI_1 , (d) característica APF_1 , (e) característica APF_5 , (f) característica APF_6	66
Figura 6.11–Médias e desvios da característica APF_1 dos sinais de vozes com grau geral 2 e grau geral 3.	67
Figura 6.12–Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvios.	80
Figura 6.13–Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvio soproisidade.	81
Figura 6.14–Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvio rugosidade.	81

Figura 6.15–Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvio tensão.	82
Figura 6.16–Comparativo entre os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com desvio rugosidade e com desvio soprosidade.	83
Figura 6.17–Comparativo entre os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com desvio rugosidade e com desvio tensão.	83
Figura 6.18–Comparativo entre os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com desvio soprosidade e com desvio tensão.	84
Figura 6.19–Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau geral 1 e grau geral 2.	85
Figura 6.20–Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau geral 1 e grau geral 3.	86
Figura 6.21–Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau geral 2 e grau geral 3.	86
Figura 6.22–Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau 1 e graus 2 e 3.	87

LISTA DE TABELAS

Tabela 5.1 – Quantidade de sinais utilizados para o estudo de caso 1.	42
Tabela 5.2 – Valores médios e desvios padrões de cada classe analisada de acordo com a análise perceptivo-auditiva realizada pela escala EAV.	42
Tabela 5.3 – Quantidade de sinais utilizados para a discriminação com relação às intensidades dos desvios vocais.	43
Tabela 5.4 – Valores médios e desvios padrões de cada classe analisada, com relação as intensidades dos desvios, de acordo com a análise perceptivo-auditiva realizada pela escala EAV.	43
Tabela 5.5 – Matriz de confusão em um teste de diagnóstico da presença ou ausência do desvio.	50
Tabela 6.1 – Potencial discriminativo entre os sinais de vozes normais e com desvios (rugosidade, sopro e tensão).	53
Tabela 6.2 – Potencial discriminativo entre sinais de vozes normais e com desvio rugosidade.	54
Tabela 6.3 – Potencial discriminativo entre os sinais de vozes normais e com desvio sopro.	56
Tabela 6.4 – Potencial discriminativo entre os sinais de vozes normais e com desvio tensão.	58
Tabela 6.5 – Potencial discriminativo entre os sinais de vozes com desvio rugosidade e com desvio sopro.	59
Tabela 6.6 – Potencial discriminativo entre os sinais de vozes com desvio rugosidade e com desvio tensão.	61
Tabela 6.7 – Potencial discriminativo entre os sinais de vozes com desvio sopro e com desvio tensão.	62
Tabela 6.8 – Potencial discriminativo entre os sinais de vozes do grau geral 1 e uma combinação do grau 2 e 3.	64
Tabela 6.9 – Potencial discriminativo entre os sinais de vozes do grau geral 1 e grau geral 2.	65
Tabela 6.10 – Potencial discriminativo entre os sinais de vozes do grau geral 1 e grau geral 3.	66
Tabela 6.11 – Potencial discriminativo entre os sinais de vozes do grau geral 2 e grau geral 3.	67
Tabela 6.12 – Resumo do potencial discriminativo das características.	68
Tabela 6.13 – Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes normais e com desvios (rugosidade, sopro e tensão).	69
Tabela 6.14 – Melhor desempenho das características individuais na classificação entre sinais de vozes saudáveis e sinais de vozes com desvio sopro.	70
Tabela 6.15 – Melhores desempenhos de características combinadas na classificação entre sinais de vozes normais e com desvio sopro.	70
Tabela 6.16 – Melhor desempenho das características individuais na classificação entre sinais de vozes normais e sinais de vozes com desvio rugosidade.	71

Tabela 6.17–Melhores desempenhos de características combinadas na classificação entre sinais de vozes normais e com desvio rugosidade.	71
Tabela 6.18–Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes normais e sinais de vozes com desvio tensão. . .	72
Tabela 6.19–Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes com desvios rugosidade e tensão.	72
Tabela 6.20–Melhor desempenho das características individuais na classificação entre sinais de vozes com desvio rugosidade e soprosidade.	73
Tabela 6.21–Melhores desempenhos de características combinadas na classificação entre sinais de vozes com desvio rugosidade e soprosidade.	74
Tabela 6.22–Melhor desempenho das características individuais na classificação entre sinais de vozes com os desvios soprosidade e tensão.	75
Tabela 6.23–Melhores desempenhos de características combinadas na classificação entre sinais de vozes com desvio soprosidade e tensão.	76
Tabela 6.24–Melhor desempenho das características individuais na classificação entre sinais de vozes G1 e G3.	77
Tabela 6.25–Melhores desempenhos de características combinadas na classificação entre sinais de vozes G1 e G3.	77
Tabela 6.26–Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes com intensidades de Grau 1 e Graus 2 e 3.	78
Tabela 6.27–Resumo dos melhores resultados individuais e combinados.	79
Tabela A.1–Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes normais.	97
Tabela A.2–Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes rugosas.	98
Tabela A.3–Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes soprosas.	99
Tabela A.4–Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes tensas.	100
Tabela A.5–Classificação segundo a Escala Analógica Visual (EAV) para sinais de vozes com predominância de grau 2.	101
Tabela A.6–Classificação segundo a Escala Analógica Visual (EAV) para sinais de vozes com predominância de grau 3.	102
Tabela A.7–Classificação segundo a Escala Analógica Visual (EAV) para sinais de vozes da combinação dos graus 2 e 3.	103

LISTA DE SIGLAS

EAV	Escala Analógico-Visual
EMD	<i>Empirical Mode Decomposition</i>
HHT	<i>Hilbert–Huang Transform</i>
HSA	<i>Hilbert Spectral Analysis</i>
IMF	<i>Intrinsic Mode Functions</i>
LDA	<i>Linear Discriminant Analysis</i>
LIEV	Laboratório Integrado do Estudo da Voz
MEEI	<i>Massachusetts Eye and Ear Infirmary</i>
MLP	<i>Multilayer Perceptron</i>
PCA	<i>Principal Component Analysis</i>
QDA	<i>Quadratic Discriminant Analysis</i>
STFT	<i>Short Time Fourier Transform)</i>
SVM	<i>Support Vector Machine</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Justificativa	16
1.2	Objetivos	17
1.2.1	Objetivo Geral	17
1.2.2	Objetivos Específicos	17
1.3	Organização do Texto	17
2	ESTADO DA ARTE	18
3	DISTÚRBIOS DA VOZ	27
3.1	Avaliação dos Distúrbios Vocais	28
3.1.1	Análise Perceptivo-Auditiva	29
3.1.2	Análise Acústica	30
4	TRANSFORMADA DE HILBERT-HUANG	31
4.1	Método de Decomposição Empírica de Modos	32
4.2	Transformada de Hilbert e Função Analítica	36
5	MATERIAL E MÉTODOS	40
5.1	Base de Dados	40
5.1.1	Análise Perceptivo-Auditiva da Voz	41
5.1.2	Critério para Escolha dos Sinais	41
5.2	Metodologia	43
5.2.1	Extração de Características	45
5.2.2	Processo de Classificação	48
5.2.3	Avaliação e Interpretação	49
6	RESULTADOS E DISCUSSÃO	52
6.1	Análise Estatística	52
6.1.1	Análise Estatística para o Estudo de Caso 1	53
6.1.2	Análise Estatística para o Estudo de Caso 2	63
6.1.3	Resumo da Análise Estatística	67
6.2	Classificação	68
6.2.1	Classificação para o Estudo de Caso 1	69
6.2.2	Classificação para o Estudo de Caso 2	76
6.2.3	Resumo da Classificação	78
6.3	Estudos Comparativos	79
6.3.1	Estudos Comparativos para o Estudo de Caso 1	79
6.3.2	Estudos Comparativos para o Estudo de Caso 2	85
7	CONSIDERAÇÕES	88
7.1	Sugestões para Trabalhos Futuros	89
	REFERÊNCIAS	90

**APÊNDICE A – A CLASSIFICAÇÃO SEGUNDO A ESCALA EAV
PARA OS SINAIS SELECIONADOS DA BASE DE
DADOS LIEV/UFPB 97**

1 INTRODUÇÃO

A voz humana é um dos meios de comunicação e expressão mais importantes para o ser humano. A voz é única para cada pessoa, além de ser um meio rápido, eficiente e econômico para a comunicação e expressão de ideias. Distúrbios na voz podem influenciar negativamente na qualidade de vida de um indivíduo, podendo limitar a comunicação no trabalho, como em outros aspectos sociais. Portanto, as alterações vocais devem ser diagnosticadas e tratadas com celeridade. O diagnóstico de uma alteração vocal é um processo que precisa combinar diversas técnicas de avaliação e análise (LOPES; ALVES; MELO, 2017; BEHLAU, 2008).

A avaliação da voz humana possibilita obter resultados que mostrem a condição de um distúrbio vocal (BEHLAU; PONTES; MORETI, 2018; ULOZA et al., 2011). Esse distúrbio é identificado por qualquer dificuldade ou alteração na emissão vocal que dificulta a produção natural de voz, não cumprindo a transmissão da mensagem verbal e/ou emocional (VERDOLINI; ROSEN; BRANSKI, 2014; BEHLAU, 2001). As dificuldades na produção da voz podem surgir devido a duas condições: uma desordem funcional causada pelo abuso ou uso incorreto da voz, podendo ser corrigido através de terapia vocal ou; por patologias laríngeas, como nódulos nas pregas vocais, úlceras, que podem ser corrigidos através de terapia vocal e/ou cirurgia (MARTINEZ; RUFINER, 2000).

Com o intuito de melhorar o processo de avaliação vocal, nas últimas décadas houve um aumento no número de pesquisas referentes ao estudo da produção da voz humana e suas alterações patológicas. Destas pesquisas, surgiram técnicas e ferramentas de análise como videolaringoscopia, glotografia, análise perceptivo-auditiva e análise acústica. Todas essas ferramentas desenvolvidas ao longo dos anos têm como propósito propiciar informações substanciais para a pesquisa na área da voz, como também aos profissionais clínicos (DAJER, 2010).

A avaliação perceptivo-auditiva é o principal método de avaliação da qualidade da voz no ambiente clínico, mas por ser um método subjetivo, depende de condições tais como: avaliadores ou “juízes” treinados, características acústicas representativas das desordens vocais e a disponibilidade de uma escala universal (DAJER, 2010; KREIMAN; GERRATT, 2010). O conceito de qualidade da voz está diretamente relacionado com as condições fisiológicas, perceptivas e acústicas (AKPINAR et al., 2011; CAMARGO; MADUREIRA, 2010; GUIMARÃES, 2007).

No entanto, a avaliação da qualidade vocal por meio de um único método pode ser frágil e limitada. Dessa forma, é necessário recorrer a análises combinadas que permitam unir conhecimentos adequados e eficazes do sistema de fonação e da qualidade vocal (BEHLAU; GAMA; CIELO, 2010; GUIMARÃES, 2007). A escolha acertada do método ou medidas adequadas para identificar o grau dos desvios ou a presença de patologias laríngeas ainda é objeto de estudo (PINHO, 2017).

Nesse contexto, a análise acústica consiste no uso de técnicas de processamento digital

de sinais de voz que pode proporcionar a construção de um sistema especialista para tipificar desvios vocais, além de ser um método não invasivo de avaliação.

1.1 Justificativa

Nos últimos anos, vem aumentando o interesse dos pesquisadores e profissionais que tratam a voz (fonoaudiólogo e otorrinolaringologista) pelo método de análise acústica para avaliação da qualidade vocal, criação de ferramentas para o auxílio de diagnósticos de desordens vocais e classificadores de patologias laringeas. As técnicas de análise acústica podem ser utilizadas como ferramentas auxiliares e não invasivas. Este tipo de análise pode auxiliar o profissional nas tomadas de decisões para a determinação de alterações da voz, avaliações de cirurgias, tratamentos medicamentoso e de reabilitação (BARCELOS et al., 2018; ROY et al., 2013; BROCKMANN-BAUSER; DRINNAN, 2011; COSTA, 2008).

As técnicas de análise acústica usualmente se baseiam no modelo fonte-filtro proposto por Fant (1970) para a produção de vogais. Outros estudos evidenciam que a produção da voz humana é um processo não linear (TITZE; BAKER; HERZL, 1993). Deste modo, supõe-se que a sua caracterização seja mais precisa com modelos que consideram a não linearidade no processo de produção da voz (DEJONCKERE et al., 2012; DAJER, 2010; SHRIVASTAV; CAMACHO, 2010).

Técnicas de análise de sinais em tempo-frequência também constituem um importante avanço na análise de sinais e são apropriadas para estudar os sinais biológicos, como a voz, pois são sinais que se caracterizam por conteúdos em frequências, são não lineares e não estacionários (PINTO, 2009). Um método para análise de sinais não lineares e não estacionários é a transformada de Hilbert-Huang (HHT, do inglês *Hilbert-Huang Transform*), a qual tem sido aplicada em muitas áreas da ciência e engenharia, incluindo a análise acústica (WANG et al., 2016; MEKYSKA et al., 2015; RAVINDRAN; NAIR, 2015; KALEEM et al., 2013; KALEEM et al., 2011; KALEEM et al., 2010).

A HHT é um método de análise em tempo-frequência composto por duas etapas: a primeira, decompõe o sinal em funções intrínsecas de modo (IMF, do inglês *Intrinsic Mode Functions*) utilizando o método de decomposição empírica de modos (EMD, do inglês *Empirical Mode Decomposition*) que permite obter um conjunto de funções base de forma adaptativa. A segunda, é a análise do espectro de Hilbert (HSA, do inglês *Hilbert Spectral Analysis*), que possibilita a obtenção de frequência instantânea do sinal (HUANG et al., 1998).

Diante da importância da compreensão dos desvios vocais para avaliação da qualidade da voz, para prevenção de agravos relacionados às alterações vocais, bem como para promover melhores informações para tomada de decisão clínica no tratamento ou reabilitação da pessoas que já possuem desvios/patologias vocais, este estudo justifica-se pela relevância de apontar uma técnica promissora para realização da análise acústica baseada na transformada de Hilbert-Huang.

A literatura disponível descreve o uso de diversas transformadas que podem ser utilizadas na análise dos sinais de voz, por exemplo, Fourier, *Wavelet* e Hilbert-Huang. Destas, a transformada de Hilbert-Huang se destaca devido a sua capacidade adaptativa na análise desse tipo de sinal (COELHO et al., 2015; SANTOS, 2015; HUANG; WU; LONG, 2008).

1.2 Objetivos

1.2.1 Objetivo Geral

Avaliar a aplicabilidade da transformada de Hilbert-Huang na detecção de desvios vocais.

1.2.2 Objetivos Específicos

- Caracterizar os sinais de voz e o grau do desvio vocal de pacientes com voz normal e com desvios vocais, por meio das características empregadas extraídas da transformada de Hilbert-Huang;
- Classificar o potencial discriminativo das características na classificação entre vozes normais e desviadas e entre os graus de desvio fonatório dos sinais de voz;
- Aplicar técnicas de classificação de padrões como rede neural artificial, para discriminar entre vozes normais e desviadas e entre os graus de desvio vocal;
- Validar uma característica ou um conjunto de características em relação ao método de decomposição empírica de modos e do espectro de Hilbert, que melhor representem o desvio vocal dos sinais analisados.

1.3 Organização do Texto

No Capítulo 2 é apresentado o estado da arte, em que são apresentados os principais resultados de algumas pesquisas encontradas na literatura, as quais tratam da investigação de métodos de avaliação dos distúrbios da voz.

Os Capítulos 3 e 4 apresentam a fundamentação teórica necessária para compreensão do método aplicado, abordando conceitos básicos como: distúrbios vocais, análise perceptivo-auditiva, análise acústica, método de decomposição empírica de modos e a transformada de Hilbert.

No Capítulo 5 é apresentada a metodologia empregada, tais como informações sobre a base de dados utilizada nesta pesquisa, assim como todo o desenvolvimento metodológico.

No Capítulo 6 são apresentados e discutidos os resultados. E, por fim, no Capítulo 7, são apresentadas as considerações finais, destacando as contribuições desta pesquisa e as sugestões para trabalhos futuros.

2 ESTADO DA ARTE

Dentre as pesquisas encontradas na literatura com o intuito de investigar características que possam avaliar a qualidade vocal de forma objetiva, aplicando a análise acústica, algumas se baseiam no modelo linear de produção da voz e utilizam características tradicionais como energia, frequência fundamental, *jitter*, *shimmer*, quociente de perturbação de amplitude, quociente de perturbação do *pitch*, relação harmônica-ruído, energia de ruído normalizada, índice de turbulência vocal, entre outras (BELHAJ; BOUZID; ELLOUZE, 2016; PANEK et al., 2015). Já outras pesquisas empregam características que não dependem da frequência fundamental, como os coeficientes obtidos da análise preditiva linear, os coeficientes cepstrais e mel-cepstrais (COURAS, 2017; BARAVIEIRA, 2016; MORO-VELÁZQUEZ et al., 2015; COSTA, 2008).

Moro-Velázquez et al. (2015) propuseram um sistema de análise acústica de reconhecimento automático, baseado no modelo de misturas de gaussianas, para classificar o grau geral e a rugosidade (irregularidade na vibração das pregas vocais) dos sinais de vozes em conformidade com a escala GRBAS (*G-overall dysphonia grade*, *R-roughness*, *B-breathiness*, *A-asthenia* e *S-strain*). A base de dados utilizada neste estudo foi a *Massachusetts Eye and Ear Infirmary* (MEEI), sendo utilizados 53 sinais de vozes normais e 657 sinais de vozes patológicas. Para a parametrização foram utilizadas as seguintes características: coeficientes mel-cepstrais e características derivadas dos espectros de modulação. A análise de componentes principais (PCA, do inglês *Principal Component Analysis*) e a análise de discriminantes lineares (LDA, do inglês *Linear Discriminant Analysis*) foram utilizadas para reduzir redundâncias. O classificador utilizado foi baseado no modelo de misturas de gaussianas e foi utilizada a validação cruzada. Os melhores resultados foram em relação as características espectrais, em que para o grau geral a melhor acurácia foi de 81,60% e para a rugosidade a acurácia foi de 84,70%.

No trabalho de Panek et al. (2015) foi realizada uma avaliação de análise acústica na detecção de vozes com distúrbios (disfonia hiperfuncional, disfonia funcional, paralisia e laringite). As características utilizadas neste trabalho foram: frequência fundamental, *jitter*, *shimmer*, energia, curtose, momento de primeira ordem, momento de segunda ordem, momento de terceira ordem, fator de potência, amplitude das formantes, frequência das formantes, valores máximo e mínimo dos sinais e 10 coeficientes mel-cepstrais. As vozes analisadas vieram da base de dados de Saarbrücken, do Instituto de Fonética da Universidade de Sarre. Desta base foram utilizados 671 sinais de vozes saudáveis e 671 sinais de vozes patológicas. Estas características são avaliadas por três métodos: PCA, kernel PCA e uma rede neural autoassociativa com validação cruzada. Foram obtidos resultados quanto a classificação de vozes masculinas e vozes femininas com os distúrbios. Tanto para vozes masculinas, como para as vozes femininas a acurácia obtida foi de 100,00%.

Na pesquisa de Belhaj, Bouzid e Ellouze (2016) foi realizada a classificação entre sinais de vozes femininas saudáveis e patológicas e a discriminação entre os sinais de vozes

femininas patológicas. Os sinais utilizados nesta pesquisa foram selecionados da base de dados MEEI, sendo utilizados 53 sinais de vozes normais e 173 sinais de vozes patológicas. As características utilizadas para classificar os sinais foram: frequência fundamental, quociente de abertura, coeficientes mel-cepstrais e energia. O classificador utilizado foi a máquina de vetores de suporte (SVM, do inglês *Support Vector Machine*). Os resultados na discriminação entre vozes saudáveis e patológicas foi de 99,10% de acurácia e para discriminar as vozes patológicas a acurácia foi de 79,77%.

Em Couras (2017) foram empregadas técnicas de processamento digital de sinais, baseadas no modelo linear de produção da fala, para analisar a qualidade vocal. A base de dados utilizada foi do LIEV (Laboratório Integrado do Estudo da Voz). Desta base foram utilizados 30 sinais de vozes normais e 90 sinais de vozes desviadas, sendo 30 sinais de vozes rugosas, 30 sinais de vozes soprosas e 30 sinais de vozes tensas. Foi avaliado o desempenho de características tradicionalmente empregadas na análise acústica, tais como frequência fundamental, *jitter*, *shimmer*, GNE (*Glottal to Noise Excitation Ratio*) e frequências formânticas. Também foi avaliado o potencial discriminativo dos coeficientes da análise de predição linear, coeficientes cepstrais e mel-cepstrais na classificação de desvios vocais (rugosidade, soprosidade e tensão). Com o auxílio de um classificador, baseado em redes neurais artificiais *Multilayer Perceptron* (MLP) com validação cruzada, foi realizada a classificação dos sinais utilizando as características extraídas individualmente e de forma combinada. Foram obtidas taxas de classificação de 86,76% na discriminação entre sinais de vozes soprosas e vozes saudáveis.

A transformada *wavelet* também tem se mostrado uma ferramenta eficiente na análise acústica, devido às propriedades como: análise tempo-frequência e análise multirresolucional. As características baseadas na energia e na entropia de *Shannon* têm sido usualmente empregadas nas pesquisas de análise acústica que utilizam a transformada *wavelet*.

No estudo de Ali et al. (2016) é apresentada uma abordagem multibanda para detecção de vozes saudáveis e patológicas (nódulos, queratose, pólipos e paralisia). Os sinais utilizados neste estudo foram selecionados da base de dados MEEI. Desta base foram utilizados 53 sinais de vozes saudáveis e 173 sinais de vozes patológicas. Essa abordagem multibanda é baseada na transformada *wavelet* discreta de três níveis e em cada nível a dimensão fractal da potência do espectro foi formada através do algoritmo de Katz (KATZ, 1988) e Higuchi (HIGUCHI, 1988). O estudo mostra que a faixa de frequência entre 1 e 1562 Hz, exibe diferenças significativas no espectro entre sinais de vozes saudáveis e patológicas. O classificador utilizado foi o SVM. A acurácia obtida para discriminar as vozes saudáveis e patológicas foi de 94,71%.

Silva et al. (2017) analisaram a distinção entre os sinais de voz saudáveis e afetados por patologias (nódulos, edema e paralisia) laríngeas, empregando a transformada *wavelet packet* na etapa de extração de características. Os sinais utilizados neste trabalho foram da base de dados MEEI. Desta base foram utilizados 53 sinais de vozes saudáveis, 44 sinais de vozes com edema de Reinke, 18 sinais de vozes com nódulos e 52 sinais de vozes com paralisia nas pregas vocais.

Foram extraídas características de energia e entropia, em seis níveis de resolução, obtidas por meio da função de Daubechies de ordem 4, com a classificação realizada por meio de redes neurais artificiais MLP utilizando validação cruzada. Foram obtidas taxas de acurácia acima de 90,00%, com a medida de entropia, na discriminação entre vozes saudáveis e afetadas por patologias nas pregas vocais.

O estudo de Teager e Teager (1990) mostra as primeiras evidências sobre o fluxo não linear na produção da voz. Titze, Baker e Herzl (1993) descreveram o conceito de não linearidade na vibração das pregas vocais. A produção da voz humana como um sistema caótico foi estabelecido através da análise de voz, de modelagem por computador e das experiências com laringe. A vibração das pregas vocais é um processo não linear e que produz efeitos combinados de eventos aerodinâmicos e biomecânicos com comportamentos vibratórios irregulares, tais como bifurcação e caos. Estes comportamentos têm sido utilizados na compreensão dos distúrbios vocais por meio de métodos dinâmicos não lineares (MACCALLUM et al., 2009).

Na pesquisa de Queiroz (2017), foram utilizadas técnicas de processamento digital de sinais para realizar a classificação de desvios vocais. Foram utilizadas características não lineares, baseadas na teoria do caos, em conjunto com características de quantificação de recorrência para a análise discriminativa destes desvios. Dois estudos de caso foram realizados nessa pesquisa. No estudo de caso 1 foi feita a discriminação de vozes adultas masculinas e femininas normais e desviadas (rugosidade, soprosidade e tensão) e no estudo de caso 2 foi avaliada a discriminação da intensidade dos graus dos desvios vocais de vozes adultas masculinas e femininas. A base de dados utilizada foi do LIEV, do Departamento de Fonoaudiologia da Universidade Federal da Paraíba. Desta base, foram utilizados 30 sinais de vozes normais e 90 sinais de vozes desviadas. As características da análise dinâmica não linear empregada no processo de classificação foram: dimensão de correlação e o primeiro mínimo da função de informação mútua. As características de quantificação empregadas foram: determinismo, entropia de Shannon, o comprimento médio das linhas diagonais, comprimento máximo das linhas verticais e a transitividade. O Passo de Reconstrução também foi utilizado no processo de classificação. Por meio dos testes estatísticos, foi avaliado o potencial de cada característica em discriminar os tipos de sinais de voz considerados. Foi utilizada a rede neural MLP com validação cruzada no processo de classificação. Para a classificação entre sinais de vozes normais e soprosas foi obtida uma acurácia média de 91,17% na distinção. Com relação à discriminação entre a intensidade dos graus dos desvios, obteve-se uma acurácia média de 94,5% entre os Graus 1 e 3.

Em sua dissertação, Pinho (2017) analisou as alterações causadas por patologias (edema e paralisia) laríngeas e desvios (rugosidade, soprosidade e tensão) vocais baseadas na dinâmica não linear. Foram empregadas duas bases de dados neste estudo, a MEEI para discriminar os sinais com vozes patológicas e normais e a base de dados do LIEV para discriminar os sinais com vozes saudáveis e com desvios. Foi estudada a desordem vocal provocada pela presença de

patologia e/ou desvios vocais, a partir das modificações no espaço de fase dos sinais analisados. Foram utilizados quatro métodos de extração de características: o método da contagem de caixas, o método da diferença, o método da similaridade e o método da contagem de caixas ponderadas. O classificador utilizado foi o MLP com validação cruzada. A classificação foi realizada com as características extraídas individualmente e de forma combinada. Os métodos que obtiveram melhor desempenho foram o da contagem de caixas ponderadas e da similaridade, tanto com características individuais como também combinadas. No segundo caso, obteve-se uma precisão de 99,00% na classificação entre vozes saudáveis e patológicas. Na classificação dos desvios vocais, o melhor resultado foi encontrado para a distinção entre as vozes soprosas e saudáveis, com taxa de precisão de 87% utilizando o método da similaridade.

Ankışhan (2018) em seu estudo realizou a detecção de patologias vocais usando características lineares e não lineares. Este estudo utilizou uma base de dados própria, com 34 sinais de vozes saudáveis e 88 sinais de vozes com patologias. Inicialmente o sinal de voz passou por um pré-processamento. Logo após, o sinal foi modelado utilizando a codificação de predição linear e os coeficientes mel-cepstrais. Depois do processo de modelamento do sinal, foram extraídas as seguintes características: espaço de fase, *jitter*, *shimmer*, características de assimetria, características de curtose, entropia e maior expoente de Lyapunov. O classificador utilizado foi o SVM. As características foram analisadas individualmente e combinadas. A proposta obteve uma acurácia de 100,00% como melhor resultado para discriminar vozes normais e patológicas.

Lopes et al. (2018), em sua pesquisa, analisaram a eficácia das características de quantificação de recorrência na discriminação de pacientes com e sem distúrbios na voz. Foi utilizada uma base de dados própria, com 541 sinais de vozes masculinas e femininas de pessoas adultas com e sem distúrbios. Doze características de quantificação de recorrência foram extraídas usando 1% de taxa de recorrência: determinismo, comprimento médio e máximo das linhas diagonais, entropia de Shannon, laminaridade, tempo de permanência, comprimento máximo das linhas verticais, tempo de recorrência do tipo 1 e do tipo 2, entropia do tempo de recorrência do tipo 1, transitividade e divergência. Características relacionadas com a topologia recorrente do sistema vocal também foram analisadas (dimensão de imersão, passo de reconstrução e raio vizinhança) totalizando um conjunto de 15 características. As características foram classificadas individualmente e combinadas por meio de análise discriminante quadrática (QDA, do inglês Quadratic Discriminant Analysis) com validação cruzada. As combinações das características melhoraram a eficácia da classificação, com melhor equilíbrio entre os valores de sensibilidade e especificidade, atingindo uma acurácia de 83,27%.

A transformada de Hilbert-Huang, mesmo sendo uma técnica para a qual ainda não exista uma formulação analítica, vem sendo empregada com sucesso em muitas áreas do conhecimento, tais como: dinâmica de fluidos no campo da oceanografia (RAO; HSU, 2008; DING et al., 2007), aplicações financeiras (ZHANG; LAI; WANG, 2008; HUANG et al., 2003), vibrações de máquinas (RODRIGUES; BATISTA, 2014; MARQUES, 2013), processamento de sinais biomédicos

(ZAGO et al., 2015; PINTO, 2009) e análise acústica (WANG et al., 2016; RAVINDRAN; NAIR, 2015; MEKYSKA et al., 2015; KALEEM et al., 2013; KALEEM et al., 2011).

Kaleem et al. (2011) classificaram por meio de simulação telefônica sinais de vozes normais e patológicas utilizando a transformada de Hilbert-Huang. A base de dados utilizada foi a MEEI, contendo 51 vozes normais e 161 patológicas de variadas patologias. Para simular a voz em uma transmissão telefônica os sinais foram amostrados a 8 kHz. Foram adicionados aos sinais 3 ruídos gaussianos, cuja relação sinal ruído (SNR, do inglês signal-to-noise ratio) ficou em 30 dB, 40 dB e 50 dB. Em seguida, esses sinais passaram pelo método de decomposição empírica de modos, formando conjuntos de 10 funções intrínsecas de modo para cada sinal. Foram extraídas características temporais e espectrais de cada função de modo intrínseco de cada sinal. A classificação do sistema foi realizada através do LDA com validação cruzada. Os resultados da classificação entre sinais de vozes normais e patológicas de 89,70% de acurácia para SNR de 50 dB, 91,30% de acurácia para SNR de 40 dB e 92,60% de acurácia para SNR de 30 dB.

Na pesquisa de Kaleem et al. (2013) foi realizada a análise e classificação do sinal de fala patológica através da transformada de Hilbert-Huang. Foi utilizada a base de dados MEEI. Os sinais retirados desta base foram frases gravadas de sinais de vozes saudáveis e sinais de vozes patológicas (orgânicas, neurológicas, traumáticas e psicogênicas). As características foram obtidas através do método de decomposição empírica de modos. Foram obtidas 11 funções intrínsecas de modo. Logo após, foi aplicada a transformada de Hilbert em cada função, com a finalidade de obter amplitudes e frequências instantâneas. Com esse conjunto de dados, foram extraídas características temporais e espectrais para discriminar os sinais. Testes estatísticos foram aplicados para selecionar as características a serem classificadas. Foi utilizado o classificador LDA com validação cruzada, obtendo como melhor resultado de classificação entre sinais de vozes normais e patológicas uma acurácia de 95,70% com 10 funções intrínsecas de modo.

Mekyska et al. (2015) propuseram, em sua pesquisa, um estudo das abordagens sobre análise de sinais no campo da fala patológica com foco especial em técnicas de parametrização. Foram utilizadas três base de dados: MEEI com sinais de vozes normais e patológicas (53 sinais de vozes saudáveis e 173 sinais de vozes patológicas); Príncipe de Asturias com sinais de vozes normais e patológicas (239 sinais de vozes saudáveis e 200 sinais de vozes patológicas); *Czech Parkinsonian* com sinais de vozes normais e patológicas (52 sinais de vozes saudáveis e 57 sinais de vozes patológicas). Nesse estudo foi realizado uma revisão de literatura sobre as principais técnicas utilizadas para classificar sinais saudáveis e com distúrbios. Foram aplicadas diversas características para caracterizar o sinal de voz, entre as quais: características que descrevem a fonação, o movimento da língua, a qualidade vocal, que utilizam a segmentação do sinal, baseadas na transformada *wavelet*, no método de decomposição empírica de modos e utilizadas na dinâmica não linear. Os classificadores utilizados foram: SVM e o algoritmo de floresta aleatória. Para a base de dados MEEI a acurácia foi de 100,00%, para a base Príncipe de Asturias

a acurácia foi de 82,10% e para a base *Czech Parkinsonian* a acurácia foi de 67,90%.

Wang et al. (2016) realizaram um estudo sobre avaliação automática da qualidade de voz com patologia usando análise acústica multidimensional baseada na escala GRBAS. A base de dados utilizada foi cedida pelo Departamento de Otorrinolaringologia do Hospital Geral do Exército Popular de Libertação, em Beijing na China. Desta base foram utilizados 100 sinais de vozes normais e 359 sinais de vozes com desvios. Foram utilizadas 65 características para cada sinal, incluindo métodos acústicos tradicionais, coeficientes mel-cepstrais, excitação do ruído glotal, dinâmica não linear e método de decomposição empírica de modos. Foram aplicadas quatro técnicas para reduzir redundâncias: máxima relevância e mínima redundância, algoritmo de Relief, PCA e LDA. O classificador utilizado foi o SVM com validação cruzada. Os resultados da classificação foram correlacionados com classificações da escala GRBAS, com a melhor acurácia de 80,58%.

No Quadro 2.1 são apresentadas de forma sucinta as informações das pesquisas que foram descritas neste Capítulo.

Quadro 2.1 – Pesquisas realizadas correlacionadas à avaliação de distúrbios da voz por meio de análise acústica

Autoria	Características Extraídas	Metodologia	Resultados
Kaleem et al. (2011)	Temporais e espectrais das IMFs de cada sinal.	Transformada de Hilbert-Huang. Base de dados MEEI (51 vozes normais e 161 patológicas). Os sinais foram amostrados a 8 kHz e adicionados aos sinais 3 ruídos gaussianos, cuja relação sinal ruído ficou em 30 dB, 40 dB e 50 dB. Aplicado o método EMD, formando conjuntos de 10 IMFs para cada sinal. Classificador LDA.	Melhores valores de acurácia entre sinais de vozes normais e patológicas: 89,70% para SNR de 50 dB, 91,30% para SNR de 40 dB e 92,60% para SNR de 30 dB. Foi utilizada validação cruzada.
Kaleem et al. (2013)	Temporais e espectrais das IMFs de cada sinal.	Transformada de Hilbert-Huang. Base de dados MEEI. Aplicado o método EMD, formando conjuntos de 10 IMFs para cada sinal. Aplicada a transformada de Hilbert em todas IMFs, para obter amplitudes e frequências instantâneas. Testes estatísticos para selecionar as características. Classificador LDA.	Melhor resultado de acurácia foi de 95,70% entre sinais de vozes normais e patológicas com 10 IMFs. Foi utilizada validação cruzada.
Mekyska et al. (2015)	Características que descrevem a fonação, o movimento da língua, a qualidade vocal, que utilizam a segmentação do sinal, baseadas na transformada <i>wavelet</i> , no método de decomposição empírica de modos e utilizadas na dinâmica não linear.	Foram utilizadas três base de dados: MEEI (53 sinais de vozes saudáveis e 173 sinais de vozes patológicas); Príncipe de Asturias (239 sinais de vozes saudáveis e 200 sinais de vozes patológicas); <i>Czech Parkinsonian</i> (57 sinais de vozes patológicas). Classificadores utilizados: SVM e o algoritmo de floresta aleatória.	Para a base de dados MEEI a acurácia foi de 100,00%, para a base Príncipe de Asturias a acurácia foi de 82,10% e para a base <i>Czech Parkinsonian</i> a acurácia foi de 67,90%.

Continuação do Quadro 2.1.

Autoria	Características Extraídas	Metodologia	Resultados
Moro-Velázquez et al. (2015)	Coeficientes mel-cepstrais e características espectrais.	Base de dados MEEI (53 sinais de vozes normais e 657 sinais de vozes patológicas). PCA e LDA foram utilizadas para reduzir redundâncias. Classificador baseado no modelo de misturas de gaussianas.	Melhores resultados: Acurácia de 81,60% para o grau geral e acurácia de 84,70% para rugosidade. Foi utilizada validação cruzada.
Panek et al. (2015)	Frequência fundamental, <i> jitter</i> , <i> shimmer</i> , energia, curtose, momento de primeira, segunda e terceira ordem, fator de potência, amplitude e frequência das formantes, valores máximo e mínimo dos sinais e 10 coeficientes mel-cepstrais	Base de dados do Instituto de Fonética da Universidade de Sarre (671 sinais de vozes saudáveis e 671 sinais de vozes patológicas). Classificadores utilizados: PCA, kernel PCA e rede neural autoassociativa.	Melhor resultado de acurácia foi de 100,00%. Foi utilizada validação cruzada.
Ali et al. (2016)	Dimensão fractal da potência do espectro.	Base de dados MEEI (53 sinais de vozes saudáveis e 173 sinais de vozes patológicas). Transformada <i>wavelet</i> discreta de três níveis. Potência do espectro obtida através do algoritmo de Katz e Higuchi. Análise espectral na faixa de frequência entre 1 e 1562 Hz. Classificador SVM.	O melhor resultado de acurácia para discriminar as vozes saudáveis e patológicas foi de 94,71%.
Belhaj, Bouzid e Ellouze (2016)	Frequência fundamental, quociente de abertura, coeficientes mel-cepstrais e energia.	Bases de dados MEEI (53 sinais de vozes normais e 173 sinais de vozes patológicas). Classificador SVM.	Os melhores resultados na discriminação entre vozes saudáveis e patológicas foi de 99,10% de acurácia e para discriminar as vozes patológicas a acurácia foi de 79,77%.
Wang et al. (2016)	Métodos acústicos tradicionais, coeficientes mel-cepstrais, excitação do ruído glotal, dinâmica não linear e método de decomposição empírica de modos.	Base de dados do Departamento de Otorrinolaringologia do Hospital Geral do Exército Popular de Libertação (100 sinais de vozes normais e 359 sinais de vozes com desvios). Foram aplicadas quatro técnicas para reduzir redundâncias: máxima relevância e mínima redundância, algoritmo de Relief, PCA e LDA. Classificador SVM.	Os resultados da classificação foram correlacionados com classificações da escala GRBAS, com a melhor acurácia de 80,58%. Foi utilizada validação cruzada.
Couras (2017)	Frequência fundamental, <i> jitter</i> , <i> shimmer</i> , GNE (<i> glottal to noise excitation ratio</i>) e frequências formânticas, coeficientes da análise de predição linear, coeficientes cepstrais e mel-cepstrais.	Base da dados utilizada foi do LIEV (30 sinais de vozes normais e 90 sinais de vozes desviadas, sendo 30 sinais de vozes rugosas, 30 sinais de vozes soprosas e 30 sinais de vozes tensas). Classificador MLP.	Melhores resultados para acurácia: Normal × Desviadas 83,09%, Normal × Soprosa 86,76%, Normal × Rugosa 77,21%, Normal × Tensa 75,00%, Rugosa × Soprosa 76,17%, Rugosa × Tensa 79,34% e Soprosa × Tensa 86,84%. Foi utilizada validação cruzada.

Continuação do Quadro 2.1.

Autoria	Características Extraídas	Metodologia	Resultados
Pinho (2017)	Método da contagem de caixas, método da diferença, método da similaridade e o método da contagem de caixas ponderadas.	Base de dados: MEEI para discriminar os sinais com vozes patológicas e normais e a base de dados do LIEV, para discriminar os sinais com vozes saudáveis e com desvios. Classificador MLP.	Melhores resultados de acurácia para base de dados do LIEV: Normal \times Desviadas 82,50%, Normal \times Soprosa 87,33%, Normal \times Rugosa 81,58%, Normal \times Tensa 78,59%, Rugosa \times Soprosa 85,00%, Rugosa \times Tensa 75,23% e Soprosa \times Tensa 78,00%. Foi utilizada validação cruzada.
Queiroz (2017)	Dimensão de correlação, primeiro mínimo da função de informação mútua, determinismo, entropia de Shannon, o comprimento médio das linhas diagonais, comprimento máximo das linhas verticais, transitividade e passo de reconstrução.	Base da dados LIEV (30 sinais de vozes normais e 90 sinais de vozes desviadas). Testes estatísticos para avaliar o potencial de cada característica. Classificador MLP.	Melhores resultados para acurácia: Normal \times Desviadas 90,33%, Normal \times Soprosa 91,17%, Normal \times Rugosa 89,50%, Normal \times Tensa 76,50%, G1 \times G2 82,33%, G1 \times G3 94,50%, G1 \times G2G3 86,50% e G2 \times G3 81,17%. Foi utilizada validação cruzada.
Silva et al. (2017)	Energia e entropia em seis níveis de resolução.	Transformada <i>wavelet packet</i> . Base de dados MEEI (53 sinais de vozes saudáveis e 114 sinais de vozes patológicas). Função de Daubechies de ordem 4. Classificador MLP.	Em médias as taxas de acurácia foram acima de 90,00%. Foi utilizada validação cruzada.
Ankışhan (2018)	Espaço de fase, <i>jitter</i> , <i>shimmer</i> , características de assimetria, características de curtose, entropia, maior expoente de Lyapunov.	Base de dados própria (34 sinais de vozes saudáveis 88 sinais de vozes com patologias). O sinal foi modelado utilizado a codificação de predição linear e os coeficientes mel-cepstrais. O classificador utilizado foi o SVM.	O melhor resultado da proposta foi acurácia de 100,00% para discriminar vozes normais e patológicas.
Lopes et al. (2018)	Quantificação de recorrência usando 1% de taxa de recorrência: determinismo, comprimento médio e máximo das linhas diagonais, entropia de Shannon, laminaridade, tempo de permanência, comprimento máximo das linhas verticais, tempo de recorrência do tipo 1 e do tipo 2, entropia do tempo de recorrência do tipo 1, transitividade e divergência. Características relacionadas com a topologia recorrente do sistema vocal: dimensão de imersão, passo de reconstrução e raio vizinhança.	Analisar a eficácia das características de quantificação de recorrência na discriminação de pacientes com e sem distúrbios na voz. Base de dados própria, com 541 sinais de vozes masculinas e femininas de pessoas adultas com e sem distúrbios. Classificador QDA.	Melhor valor de acurácia foi de 83,27%. Foi utilizada validação cruzada.

Conclui-se que a revisão do estado da arte dos últimos anos reforça que a avaliação da qualidade vocal tem sido amplamente estudada através da análise acústica e que as técnicas utilizadas se mostram eficazes para a obtenção de relevantes resultados. Nas pesquisas encontradas na literatura, algumas realizam análise estatística com as características de análise acústica, algumas são baseadas no modelo linear da produção de voz, outras baseadas na análise dinâmica não linear. Há, também, técnicas híbridas, contendo combinações de características de métodos diferentes.

Além disso, não foi encontrado nenhum trabalho na literatura que utilize a transformada de Hilbert-Huang para realizar a classificação entre sinais de vozes saudáveis e desviadas e entre as suas intensidades vocais com validação cruzada, como foi empregado nesta pesquisa.

3 DISTÚRBIOS DA VOZ

A voz humana é um instrumento rápido e eficaz de comunicação que, sendo única e singular para cada pessoa, assume um padrão próprio, intrínseco a cada indivíduo (PINHO, 2003). São muitas as definições de voz, estando a sua produção relacionada a aspectos biológicos, genéticos, culturais e psicossociais. Além disso, os traços do estado emocional e a forma de exteriorizar as emoções diferenciam a voz (GOULART; COOPER, 2002). A emissão da voz constitui no som produzido pelos seres humanos, podendo identificar-lhes quanto ao gênero, raça, idade, entre outros (CEBALLOS et al., 2011).

A qualidade vocal traz muitas informações sobre fatores sociais, psicológicos e físicos, viabilizam a compreensão da personalidade e humor de quem fala (BEHLAU, 2001). A qualidade da voz faz parte da ação conjunta da laringe e do trato vocal, manifestando-se da combinação dos ajustes a longo intervalo de tempo, que se fazem presentes, de modo recorrente, na voz de cada pessoa (GAYOTTO, 2005).

A produção da voz de um indivíduo é o fruto da ação de todo organismo, na qual detalhadamente na interação entre os sistemas respiratório, fonador e ressoador (PINHO, 2003). Em comportamentos primários, como o choro, ou até um comportamento bem exercitado, como o canto a voz pode ficar diferente (BEHLAU; GAMA; CIELO, 2010). A classificação vocal considerada “normal” deve ser uma voz produzida com facilidade, confortável e possuir equilíbrio entre o ar e a musculatura das pregas vocais. Quando estas condições não são mantidas e existir dificuldade na emissão vocal em que uma ou mais características acústicas apresentam-se alteradas, em relação ao timbre ou intensidade, a voz é considerada disfônica (GAYOTTO, 2005).

As disfonias podem ser orgânicas, funcionais e mistas. Elas não são caracterizadas como uma doença em si, mas uma manifestação de um mau funcionamento de um dos sistemas ou estruturas que agem na produção da voz (GAYOTTO, 2005). É sintoma presente em vários e diferentes distúrbios vocais, ora mostrando-se como sintoma principal, ora como secundário. Muitas vezes é o sintoma mais importante de uma patologia vocal, as vezes por isso sendo enfrentado como a própria doença, como ocorre nas disfonias funcionais (BEHLAU, 2001).

No contexto de disfonias funcionais o tratamento é realizado pelo fonoaudiólogo, onde este tem mais possibilidades de atuar na recuperação do paciente, dependendo quase exclusivamente do tratamento aplicado para recuperação vocal e representam na verdade, disfonias do comportamento vocal, que decorrem do uso incorreto da voz, devido à falta de conhecimento vocal. A falta de conhecimento vocal pode abranger questões relacionadas à postura corporal e questões ambientais (PINHO, 2003). A intervenção dos especialistas em voz nas disfonias funcionais, é eficaz, eficiente e de caráter preventivo em relação à lesões orgânicas secundárias, tais como: nódulos, pólipos e edemas, que se constituem nas lesões mais comumente observadas nas disfonias (BEHLAU, 2001).

Quando se trata de disfonias orgânicas, estas podem ser consideradas disfonias funcionais diagnosticadas ou tratadas tardiamente, seja por atraso do portador da disfonia na procura da solução para o seu problema ou ainda pelo reconhecimento da potencialidade de se desenvolver uma lesão secundária, alterando todas as características da voz da pessoa portadora da disfonia (PINHO, 2003). É um sintoma associado a outras doenças ou decorrentes de alterações anatômicas importantes do aparelho fonador (BEHLAU, 2001).

A disfonia e todos os sintomas desencadeados por ela acarretam uma série de consequências, tais como, alterações psicoemocionais, depressão e frustração, que interferem direta e indiretamente no desempenho social, profissional e nas atividades de vida diária do indivíduo (LOPES; VILELA, 2016). Diante disto se verifica a importância de realizar a avaliação dos distúrbios vocais.

3.1 Avaliação dos Distúrbios Vocais

A avaliação dos distúrbios da voz é uma atividade multiprofissional e envolve procedimentos da medicina e fonoaudiologia que compreendem uma avaliação detalhada. Assim, a avaliação dos distúrbios vocais tem como objetivo detectar os possíveis motivos do distúrbio vocal, descrevendo as características do perfil vocal do paciente (CASPER; LEONARD, 2006; BEHLAU, 2001).

No ambiente clínico, as alterações vocais mais comuns são: vozes rugosas, quando a impressão auditiva é de irregularidade fonatória; soprosas, na presença de ar não sonorizado ao longo da emissão; e tensas, quando o esforço à fonação é perceptível. A identificação dessas alterações vocais é uma tarefa que se torna muito difícil em determinados casos, pois estas alterações vocais aparecem, com frequência, de forma conjunta no mesmo paciente (MADAZIO, 2009).

Os desvios vocais analisados nesta pesquisa são rugosidade, soprosidade e tensão. Podem ser definidos como (PINHO, 2003; BEHLAU, 2001):

- Rugosidade: irregularidade de vibração das pregas vocais. Engloba o conceito de rouquidão, crepitação, bitonalidade e também aspereza. Assim, a voz é percebida com ruídos presentes em baixas frequências, com características rugosa e ruidosa. Este parâmetro verifica-se em casos de: fenda glótica, presença isolada de uma alteração orgânica ou fenda de qualquer dimensão com alterações da mucosa das pregas vocais (exemplo: nódulos, pólipos ou edemas).
- Soprosidade: presença de ruído de ar na emissão vocal, audível, que corresponde fisiologicamente à fenda glótica (abertura entre as pregas vocais).

- Tensão: associada a esforço vocal por aumento da adução glótica (hiperfunção), geralmente inerente ao aumento da atividade da musculatura extrínseca da laringe, com elevação desta. Exemplo: disfonia espasmódica e síndromes de abuso vocal com conseqüente alteração da mucosa (nódulos ou pólipos).

Os métodos a seguir são mais utilizados (individualmente ou combinados) para o diagnóstico de distúrbios vocais, são eles: autoavaliação, análise visual das pregas vocais, avaliação aerodinâmica, análise perceptivo-auditiva da voz e análise acústica do sinal de voz (COSTA, 2008; CASPER; LEONARD, 2006; BEHLAU, 2001). Dentre estas avaliações esta pesquisa vai destacar a análise perceptivo-auditiva e a análise acústica.

3.1.1 Análise Perceptivo-Auditiva

A avaliação perceptivo-auditiva é uma avaliação tradicional na rotina clínica fonoaudiológica. Permite a inferência de dados anatomofisiológicos importantes e pode prover informações sobre o grau de aceitação da voz avaliada e a limitação social da disfonia.

Apesar de a avaliação perceptivo-auditiva ser a análise mais empregada na clínica, a segurança das avaliações entre os especialistas é incerta, conforme as pesquisas o ouvido dos especialistas parece ser mais preciso para avaliar vozes normais (KREIMAN; GERRATT, 2010). Pesquisas indicam que juízes comumente divergem na análise da qualidade vocal e apontam que os fatores de discordância são o resultado dos métodos utilizados para coletar as avaliações e não problemas nas capacidades perceptivas ou inconsistência dos avaliadores (KREIMAN; GERRATT; ITO, 2007; PATEL; SHRIVASTAV, 2007).

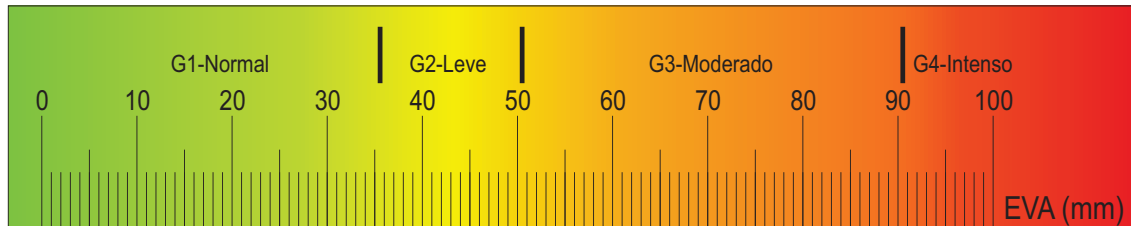
Com a finalidade de diminuir as inconsistências da análise perceptivo-auditiva, foram desenvolvidas diversas escalas de avaliação, com o emprego de diferentes tarefas para a análise perceptivo-auditiva da qualidade vocal (MARTINS; COUTO; GAMA, 2015; KREIMAN; GERRATT, 2010). A mensuração da avaliação perceptivo-auditiva através de escalas numéricas é um processo comumente utilizado. Além dessa aplicação, escalas são utilizadas na área da saúde para quantificar fenômenos subjetivos como dor, ansiedade, náusea, fadiga, dispneia, entre outros (VIEIRA, 2014).

Dentre as escalas existente as mais utilizadas pelos especialistas em voz, conforme a literatura, em relação à avaliação de distúrbios vocais por meio de análise perceptivo-auditiva são: CAPE-V (*Consensus Auditory Perceptual Evaluation of Voice*) (KEMPSTER et al., 2009), GRBAS (*Grade, Roughness, Breathiness, Asthenia, Strain*) (HIRANO, 1981), GRBASI (*Grade, Roughness, Breathiness, Asthenia, Strain, Instability*) (GAMA et al., 2009; BELE, 2005), e EAV (Escala Analógico-Visual) (LOPES et al., 2012; YAMASAKI et al., 2008).

O estudo de Yamasaki et al. (2008) indicou uma correlação entre a escala EAV e a escala numérica da percepção auditiva da voz. No estudo a seleção da escala numérica de 4 pontos foi utilizada por representar a prática na clínica vocal. A determinação dos valores de corte para os

diferentes graus de alteração vocal tornou-se praticável e apresenta, na escala EAV, limites de 0 mm a 35,5 mm para a variabilidade normal da qualidade vocal, denominado grau 1 na escala numérica. O grau 2 representa uma região entre 35,6 mm a 50,5 mm para o desvio leve. O grau 3 compreende uma faixa entre 50,6 mm a 90,5 mm para o desvio moderado e o grau 4 compreende uma faixa entre 90,6 mm a 100 mm para o desvio intenso. A Figura 3.1 ilustra a correlação entre a escala EVA e a escala numérica.

Figura 3.1 – Grau da alteração vocal na escala EAV.



Fonte: Autoria própria, 2018.

3.1.2 Análise Acústica

Os métodos não invasivos para realizar a avaliação da voz, como a análise acústica, embora tenham um papel auxiliar no diagnóstico de distúrbios vocais, possui a vantagem de fornecer dados quantitativos extraídos automaticamente, por meio de um processamento computacional (COURAS, 2017; QUEIROZ, 2017; VIEIRA; ROSA, 2006).

No Brasil, a análise acústica começou a ser introduzida nas clínicas na década de 1990 e, na atualidade, tem sido cada vez mais presente na realidade fonoaudiológica (DAJER, 2010; BEHLAU, 2001). Segundo Vieira e Rosa (2006) este tipo de análise é uma ferramenta valiosa na clínica, devido aos seguintes fatores:

- Pode ser empregada no diagnóstico, apoiando as observações subjetivas nas imagens de videolaringoestroboscopia;
- Em campanhas de triagem, podem potencialmente ser utilizadas para detectar níveis de perturbação fonatória acima dos valores de referência de uma população não disfônica;
- Podem auxiliar o monitoramento terapêutico;
- As imagens e gráficos destas análises uma maior compreensão do problema vocal;
- Registros e medidas acústicas podem servir em perícias onde há questionamento da eficácia de tratamentos vocais.

É possível, ainda, a obtenção de valores normativos, obtendo-se limiares de normalidade ou desvios das medidas acústicas da voz.

4 TRANSFORMADA DE HILBERT-HUANG

Para analisar sinais de voz com e sem desvios vocais, é necessário que se busquem soluções mais apropriadas para esta tarefa, de modo que haja de forma adequada a melhor análise desses sinais. As formas mais usuais de se estudar um sinal de voz é a análise do sinal no domínio do tempo ou da frequência. O uso da transformada de Fourier é comum quando se deseja observar propriedades de frequência de um sinal. Como o sinal de voz é um sinal não estacionário a aplicação da transformada de Fourier implica na perda da informação do tempo e apenas a informação da frequência permanece na análise, pois a transformada de Fourier é mais adequada para sinais estacionários (HUANG et al., 1998).

Gabor, em 1946, fez uma adaptação na Transformada de Fourier com uma técnica chamada janelamento do sinal, a qual é conhecida como transformada de Fourier de Curta Duração (STFT, do inglês *Short Time Fourier Transform*). Essa técnica permite realizar uma análise tempo-frequência do sinal (CASTAÑÓN, 2003). A STFT tem uma precisão limitada pois, após a escolha do tamanho da janela a mesma permanece inalterada para todas as frequências restringindo assim os resultados a serem obtidos (SANTOS, 2015; CROVATO, 2004; PARRAGA, 2002).

A transformada *wavelet* é uma ferramenta matemática que ajudou a superar a limitação da janela fixa da STFT, utilizando a técnica de janelamento variável (CASTAÑÓN, 2003), por proporcionar uma janela de tamanho variável que melhora em precisão as informações sobre tempo e frequência de um sinal. A transformada *wavelet* também serve como ferramenta de análise de fenômenos não estacionários (MOLLA et al., 2013; BURRUS et al., 1998). Porém, ainda é limitada pela relação da incerteza entre as grandezas tempo e frequência, além de apresentar uma grande contaminação no espectro pelos efeitos de bordas (MARTINS, 2015). Outra dificuldade ao se utilizar a transformada *wavelet*, é que esta transformada usa filtros pré-definidos, o que pode acarretar a uma longa busca até encontrar o filtro ideal para realizar a análise sinal (COELHO et al., 2015).

A transformada de Hilbert-Huang (HHT) é um método adaptativo elaborado para análise de dados de processos não lineares e não estacionários. A parte principal do HHT é o método EMD, com o qual qualquer conjunto de dados pode ser decomposto em um número finito de componentes com características locais que são as IMFs. A transformada de Hilbert é usada para calcular a frequência instantânea de cada IMF. A apresentação final dos resultados é uma distribuição amplitude-tempo-frequência (HUANG; WU; LONG, 2008).

O Quadro 4.1 exhibe comparações entre as transformadas de Fourier, *Wavelet* e Hilbert-Huang (MOLLA et al., 2013). Analisando o Quadro 4.1 conclui-se que, dependendo da natureza do sinal, uma determinada transformada pode ser uma ferramenta mais interessante para uma aplicação do que outra. Nesta pesquisa, os sinais em análise são sinais de voz e a transformada de Hilbert-Huang se destaca, pois a sua aplicação pode ser usada para sinais lineares, não lineares,

estacionários e não estacionários. A sua capacidade adaptativa de análise é ideal para lidar com vários tipos de sinais.

Quadro 4.1 – Comparações entre as transformadas de Fourier, *Wavelet* e Hilbert-Huang.

	Fourier	<i>Wavelet</i>	Hilbert-Huang
Base	A priori	A priori	Adaptativa
Frequência	Global, Incerteza	Local, Incerteza	Local, Certeza
Apresentação	Energia-frequência	Energia- tempo-frequência	Energia- tempo-frequência
Sinal não linear	Não	Não	Sim
Não estacionaridade	Não	Sim	Sim
Base teórica	Teoria completa	Teoria completa	Empírica

Fonte: Adaptado de Molla et al., 2013.

4.1 Método de Decomposição Empírica de Modos

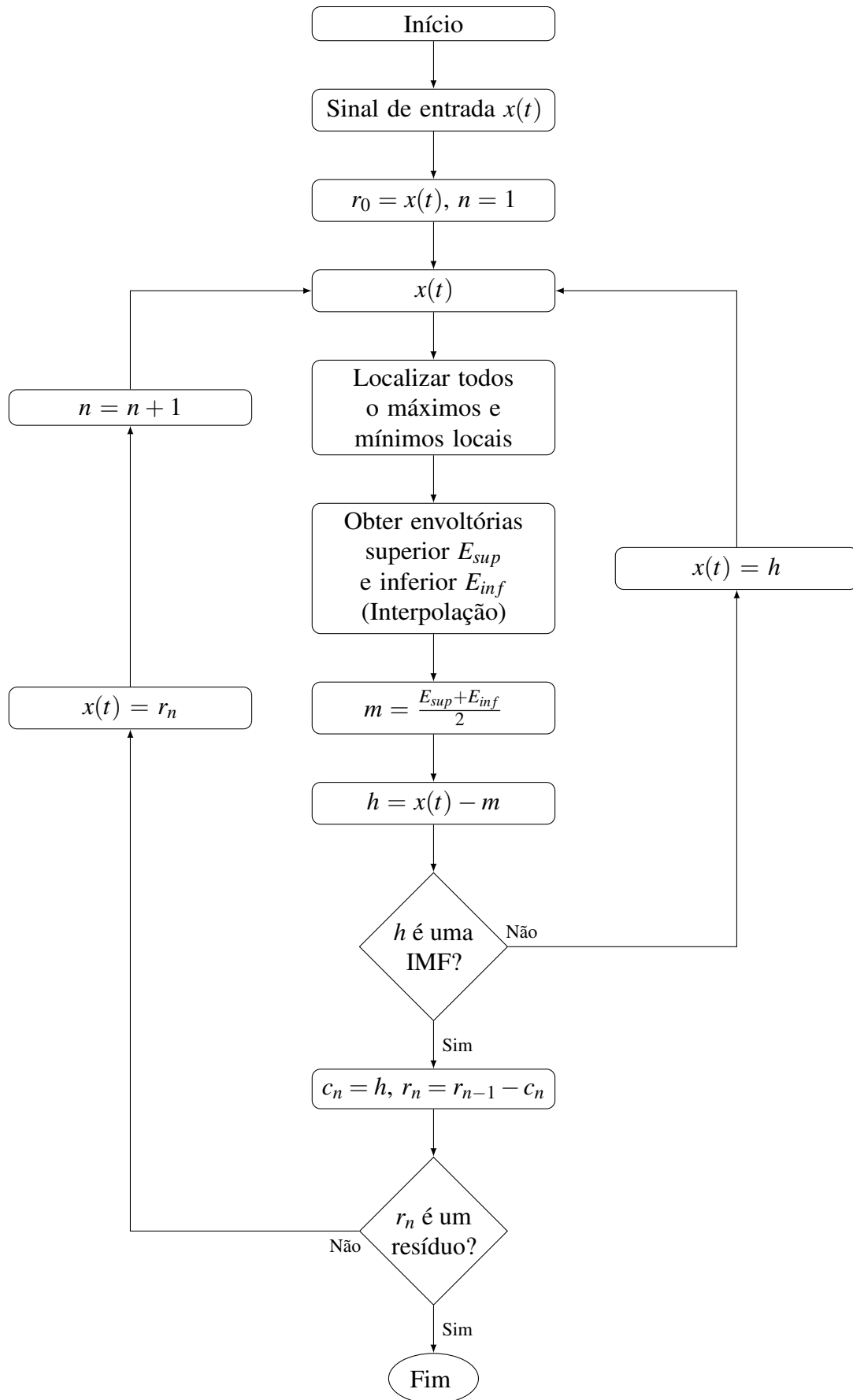
O método EMD está ligado ao conceito de frequência de oscilação local, que é a frequência de oscilação do fenômeno estudado em cada instante. Esse método tem por finalidade reduzir qualquer sinal em uma coleção de IMFs. A essência da EMD é localizar empiricamente todos os modos de oscilações através das características locais do sinal e, por fim, decompor os sinais de acordo com essas características, para as quais a análise espectral de Hilbert pode ser aplicada (HUANG et al., 1998).

A IMF representa um modo oscilatório simples como contrapartida da função harmônica simples, mas é muito mais geral. Em vez de amplitude e frequência constantes em um componente harmônico simples, uma IMF pode ter amplitude e frequência variáveis ao longo do eixo do tempo. Uma IMF tem que satisfazer duas condições: i) o número de máximos ou mínimos locais é igual ao número de vezes que a função cruza o zero; ii) em qualquer instante de tempo, a média local da envoltória formada pela interpolação do extremos (máximos e mínimos) é igual a zero (HUANG et al., 1998).

Um sinal que atende às duas condições destacadas anteriormente recebe o nome de IMF. Um sinal que possui mais de uma frequência de oscilação, possui também mais de uma IMF e é neste tipo de sinal que o método EMD é de grande aplicabilidade, já que tem por finalidade separar o sinal em várias IMFs (FLANDRIN; RILLING; GONCALVES, 2004).

Uma vez que o sinal em análise pode ser separado em IMFs, o método EMD é simples e objetivo, não sendo necessário o uso de ferramentas complexas. Rilling et al. (2003) descrevem um algoritmo para se obter as IMFs de um sinal não linear e não estacionário. O processo sistemático para extrair as IMFs é chamado de *sifting* (ESPER, 2016; FLANDRIN; RILLING; GONCALVES, 2004; RILLING et al., 2003; HUANG et al., 1998). A Figura 4.1 apresenta o fluxograma do algoritmo do método EMD.

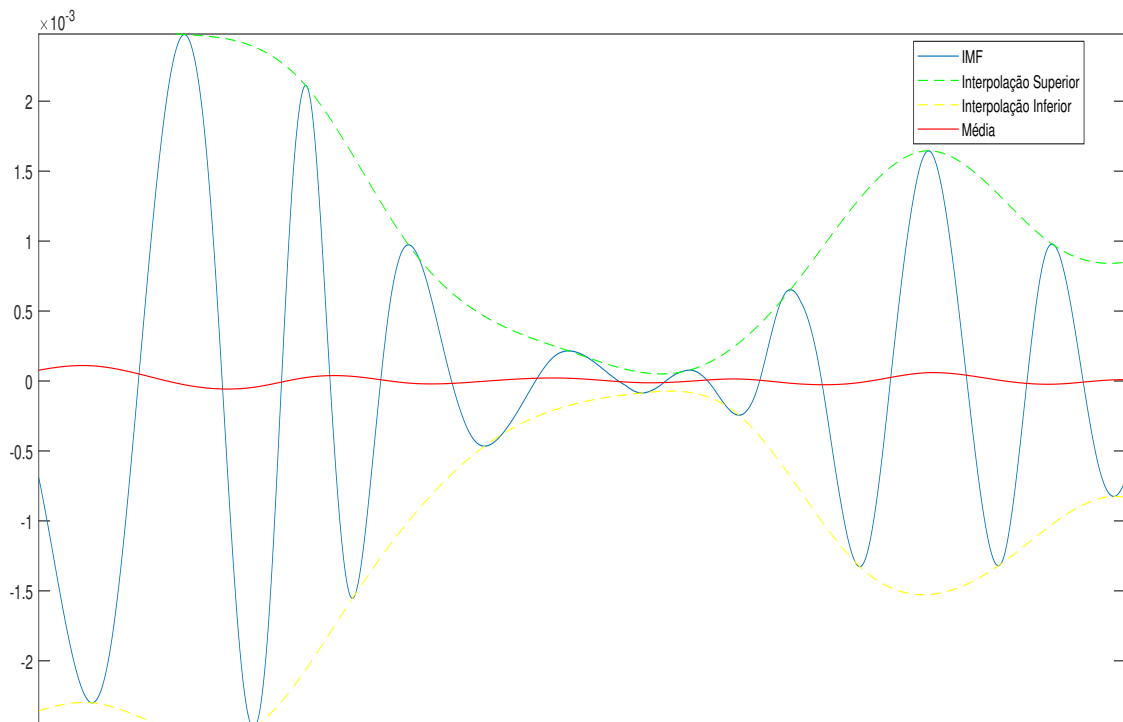
Figura 4.1 – Fluxograma do algoritmo do método EMD.



Os itens de 1 a 6 detalham o fluxograma apresentado na Figura 4.1.

1. Para um determinado sinal $x(t)$, todos os mínimos e máximos de $x(t)$ são identificados;
2. A envoltória superior é calculada usando uma interpolação *spline* cúbica para conectar todos os máximos locais. Da mesma forma, a envoltória inferior é calculada a partir dos mínimos locais. As envoltórias superior e inferior devem cobrir entre elas todos os dados em $x(t)$;
3. A média m_1 das envoltórias superior e inferior é calculada, é feita uma atualização através de uma operação de subtração denominada de h_1 : $h_1 = x(t) - m_1$ e h_1 será o primeiro componente;
4. Se h_1 não for uma IMF, h_1 será tratado como o sinal original e as três etapas anteriores são executadas. Então $h_1 - m_{11} = h_{11}$ este processo será repetido até k vezes, até que h_{1k} seja reduzido a uma IMF, ou seja, $h_{1k} = h_{1(k-1)} - m_{1k}$. Esse processo para extrair as IMFs é denominado de *sifting*. Então h_{1k} que agora é uma IMF será denominado como c_1 . A primeira IMF contém as maiores frequências de oscilação encontradas dos dados originais de $x(t)$. A Figura 4.2 apresenta um exemplo de um sinal de voz normal após algumas interações de *sifting*;

Figura 4.2 – Exemplo de um sinal no processo de *sifting* obtido após k iterações.

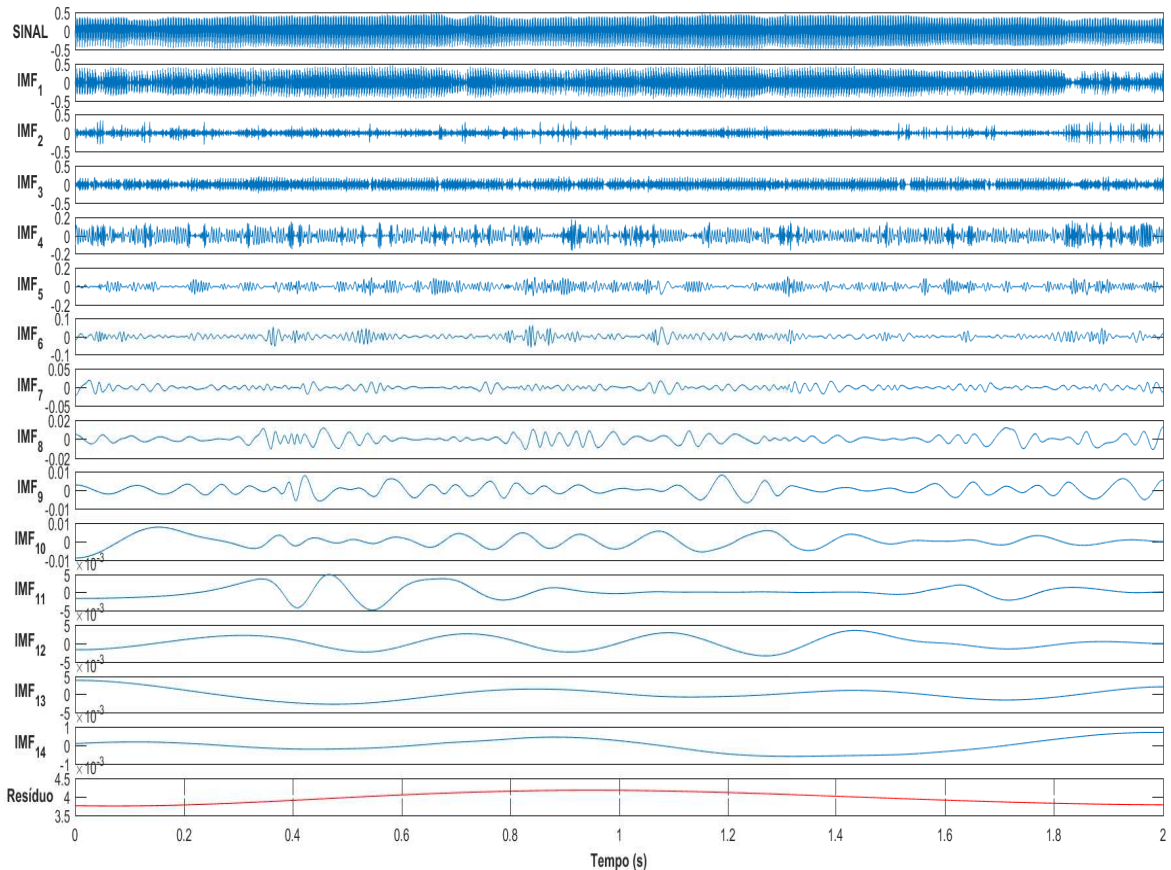


Fonte: Dados da pesquisa, 2018.

5. A primeira IMF c_1 é subtraída de $x(t)$ para obter o resíduo r_1 : $r_1 = x(t) - c_1$;

6. O resíduo r_1 é agora considerado o ponto de partida. As etapas mencionadas anteriormente são repetidas n vezes para encontrar todas as IMFs e para que o resíduo r_n se torne uma constante, uma função monótona ou ainda, uma função com um único máximo e mínimo do qual nenhum outro IMF pode ser extraído: $r_n = r_{n-1} - c_n$. A Figura 4.3 ilustra os gráficos obtidos para cada IMF de um sinal de voz normal que foi submetido ao algoritmo do método EMD.

Figura 4.3 – Exemplo de IMFs e resíduos obtidos após método EMD de um sinal de voz normal.



Fonte: Dados da pesquisa, 2018.

O algoritmo do método EMD assegura que qualquer sinal $x(t)$ pode ser decomposto em um número finito de n interações formando suas IMFs. Esse sinal pode ser reconstruído como a soma de n IMFs e um resíduo r_n , conforme apresentado na Equação 4.1.

$$x(t) = \sum_{i=1}^n c_i + r_n. \quad (4.1)$$

Devido o processo de *sifting* se tratar de um processo numérico, o critério de convergência foi desenvolvido para verificar o instante de parada desse processo. Este critério tem que atender às duas condições, descritas no início desta seção, para que uma IMF possa existir.

Para atender a segunda condição, o critério de convergência tem por finalidade suavizar altas disparidades de amplitudes, onde o valor do desvio padrão (SD, do inglês *Standard*

Deviation) deve ser menor que um valor previamente estabelecido. Foi sugerido por Huang et al. (1998) valores entre 0,2 e 0,3. O cálculo de SD é realizado pela Equação 4.2.

$$SD = \sum_{t=0}^T \left[\frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_{1(k-1)}^2(t)} \right], \quad (4.2)$$

em que k é a quantidade de interações que o sinal vai fazer até alcançar o critério de parada do processo de *sifting* e começa com o valor 1.

Devido ao fato de que o método EMD apresenta algumas limitações em relação ao critério de parada, à interpolação dos extremos e obtenção da média nula, surgiram pesquisas que trazem propostas para rever estas limitações.

Rilling et al. (2003) abordaram um critério de parada do processo de *sifting* que permite garantir globalmente pequenas variações na média m de modo que, localmente, permita oscilações significativas. Os autores também propuseram alteração no algoritmo original através de uma decomposição *online*, com o intuito de otimizar a interpolação dos extremos. O algoritmo é processado para um conjunto de pontos (5 pontos no caso da interpolação *spline* cúbica, pois é o mínimo de pontos possível) e depois o processo é repetido para o restante do sinal através de uma janela deslizante. Pustelnik, Borgnat e Flandrin (2014) propuseram um método de otimização convexa e não suave para realizar a interpolação dos extremos do processo de *sifting*.

Com a intenção de alcançar a média nula de forma mais rápida, Rilling e Flandrin (2006) sugeriram uma decomposição EMD local. Nesta decomposição, é criada uma função peso e determinadas condições para que se alcance a média nula mais rápida que o algoritmo original.

4.2 Transformada de Hilbert e Função Analítica

Concluídas as etapas anteriores é obtida um conjunto de IMFs, as quais se pode aplicar a transformada de Hilbert para obter o espectro amplitude-tempo-frequência, que é a segunda etapa da HHT.

Tomando as IMFs como $c_i(t)$ como indicado na Equação 4.1, a transformada de Hilbert possui as seguintes propriedades: i) um sinal $c_i(t)$ e sua transformada de Hilbert $\tilde{c}_i(t)$ têm o mesmo espectro de amplitudes; ii) se $\tilde{c}_i(t)$ é a transformada de Hilbert de $s(t)$, então, $-c_i(t)$ é a transformada de Hilbert de $\tilde{c}_i(t)$; iii) a transformada de Hilbert é uma transformação linear; iv) um sinal $c_i(t)$ e sua transformada de Hilbert $\tilde{c}_i(t)$ são ortogonais (CRESPO, 2018).

Partindo das IMFs obtidas pelo método EMD, Huang et al. (1998) demonstra que é possível calcular a transformada de Hilbert para cada IMF resultante, para posteriormente se obter o sinal original como a soma da parte real das transformadas. Que é expresso matematicamente através da Equação 4.3 (YU; CHENG; YANG, 2005).

$$\tilde{c}_i(t) = H[c_i(t)] = c_i(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(\tau)}{t - \tau} d\tau \quad (4.3)$$

Como a integral é imprópria, é necessário calcular o valor principal de Cauchy. Para sinais discretos esta preocupação não é necessária, já que a operação de integração passa a ser de somatório (OPPENHEIM; SCHAFER, 2012).

A Transformada de Hilbert pode ser definida a partir de um sinal analítico. Um sinal analítico é um sinal que não apresenta componentes negativos de frequência e que, por consequência, é complexo no tempo. É possível demonstrar a relação entre a Transformada de Hilbert e o sinal analítico aplicando o teorema da convolução à Equação 4.3 (MELO, 2015):

$$\begin{aligned}\mathcal{F}\{\tilde{c}_i(t)\} &= \mathcal{F}\{c_i(t)\} * \mathcal{F}\left\{\frac{1}{\pi t}\right\} \\ \tilde{C}_i(f) &= -jC_i(f)\text{sgn}(f),\end{aligned}\quad (4.4)$$

em que sgn é a função sinal:

$$\text{sgn}(c_i) = \begin{cases} -1, & c_i < 0 \\ 0, & c_i = 0. \\ 1, & c_i > 0 \end{cases} \quad (4.5)$$

Substituindo a definição de sgn da Equação 4.5 na Equação 4.4:

$$\tilde{C}_i(f) = \begin{cases} jC_i(f), & f < 0 \\ 0, & f = 0. \\ -jC_i(f), & f > 0 \end{cases} \quad (4.6)$$

Logo:

$$\tilde{C}_i(f) = -jC_i(f_+) + jC_i(f_-). \quad (4.7)$$

Multiplicando ambos os lados da Equação 4.7 por j :

$$j\tilde{C}_i(f) = C_i(f_+) - C_i(f_-). \quad (4.8)$$

Somando $C_i(f)$ em ambos os lados da Equação 4.8:

$$C_i(f) + j\tilde{C}_i(f) = C_i(f) + C_i(f_+) - C_i(f_-). \quad (4.9)$$

Substituindo $C_i(f)$ por $C_i(f_-) + C_i(f_0) + C_i(f_+)$ no lado direito da Equação 4.9:

$$C_i(f) + j\tilde{C}_i(f) = C_i(f_0) + 2C_i(f_+). \quad (4.10)$$

Nota-se que o lado direito da Equação 4.10 não possui componentes negativas de frequência o que está em conformidade com a definição de função analítica. Sendo que a função analítica pode ser representada em frequência, por:

$$Z(f) = \begin{cases} 2C_i(f), & f > 0 \\ C_i(f), & f = 0. \\ 0, & f < 0 \end{cases} \quad (4.11)$$

Já no domínio do tempo a função analítica pode ser representada por:

$$z(t) = \mathcal{F}^{-1} [Z(f)]. \quad (4.12)$$

Aplicando a definição da Equação 4.11 no lado direito da Equação 4.10:

$$C_i(f) + j\tilde{C}_i(f) = Z(f), \quad (4.13)$$

calculando a Transformada Inversa de Fourier na Equação 4.13:

$$c_i(t) + j\tilde{c}_i(t) = z(t). \quad (4.14)$$

Substituindo a 4.3 na Equação 4.14:

$$z(t) = c_i(t) + jH[c_i(t)]. \quad (4.15)$$

Em que $H[c_i(t)]$ é a transformada de Hilbert de $c_i(t)$. Reescrevendo a Equação 4.15 para forma polar (YU; CHENG; YANG, 2005):

$$z_i(t) = a_i(t)e^{j\theta_i(t)} \quad (4.16)$$

A amplitude instantânea é representada por a_i e a fase instantânea é representada por θ_i . A amplitude instantânea e fase instantânea são calculadas conforme as Equações 4.17 e 4.18, respectivamente (YU; CHENG; YANG, 2005).

$$a_i(t) = \sqrt{c_i^2(t) + H^2[c_i(t)]} \quad (4.17)$$

$$\theta_i(t) = \text{tg}^{-1} \frac{H[c_i(t)]}{c_i(t)} \quad (4.18)$$

A frequência instantânea, f_i , pode ser obtida conforme Equação 4.19 (YU; CHENG; YANG, 2005).

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt} \quad (4.19)$$

Uma vez que a HHT tenha sido aplicada a todas as IMFs, e calculadas as amplitudes e frequências instantâneas, o sinal original $x(t)$ pode ser expresso conforme a Equação 4.20 (YU; CHENG; YANG, 2005).

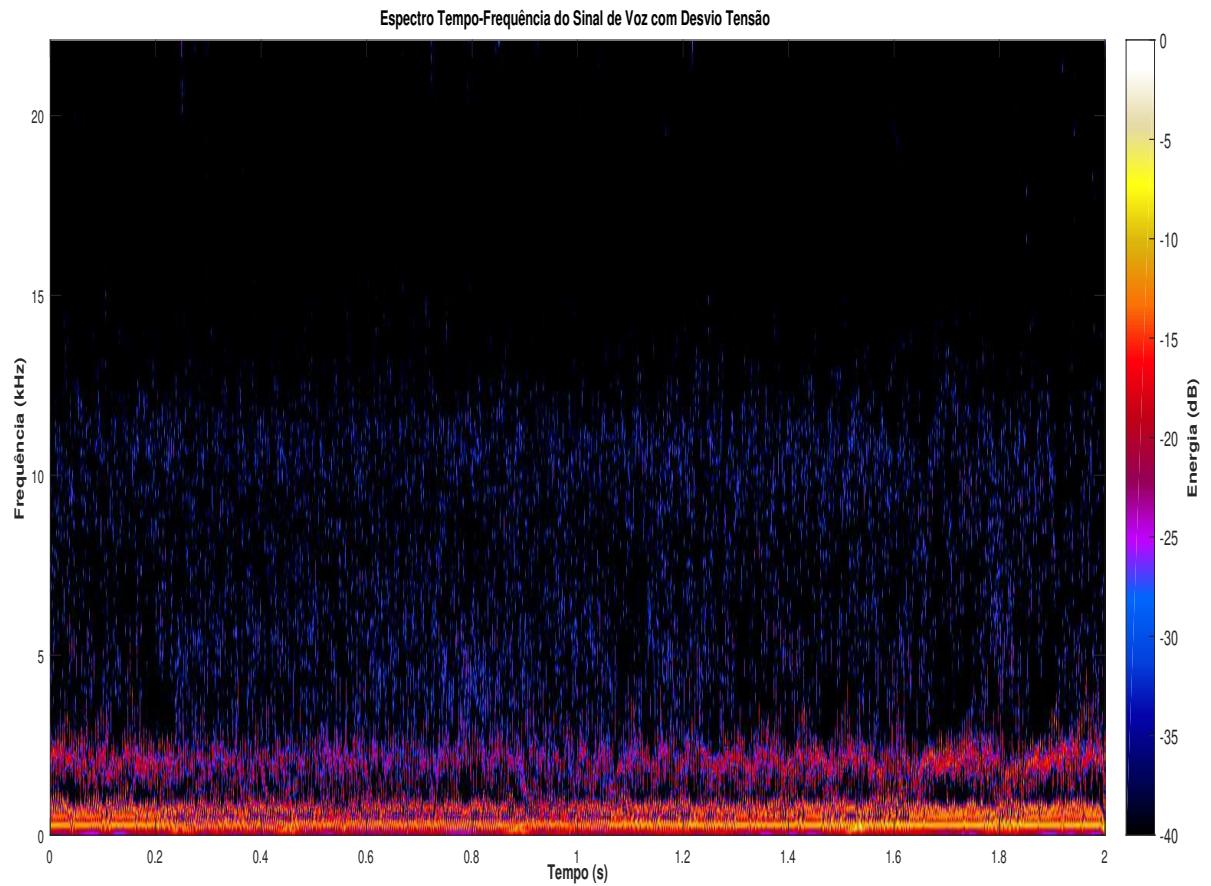
$$x(t) = \sum_{i=1}^n a_i(t)e^{j\theta_i(t)} = \sum_{i=1}^n a_i(t)e^{\frac{j}{2\pi} \int f_i(t) dt} \quad (4.20)$$

A parte real da Equação 4.20 permite uma distribuição tempo-frequência-amplitude aliada ao espectro de Hilbert, $H(f, t)$, que é expresso através da Equação 4.21 (YU; CHENG; YANG, 2005).

$$H(f, t) = \text{Re} \sum_{i=1}^n a_i(t)e^{\frac{j}{2\pi} \int f_i(t) dt} \quad (4.21)$$

Em que $H(f,t)$ representa o espectro de Hilbert. A Equação 4.21, quando comparada com a expansão clássica do sinal numa série de Fourier é mais favorável, visto que as amplitudes e as frequências variam com o tempo, permitindo uma melhor caracterização de processos que têm uma natureza não estacionária. A Figura 4.4, ilustra o espectro de Hilbert para as 10 primeiras IMFs de um sinal de voz com desvio tensão.

Figura 4.4 – Exemplo do espectro de Hilbert para as 10 primeiras IMFs de um sinal de voz com desvio tensão.



Fonte: Dados da pesquisa, 2018.

5 MATERIAL E MÉTODOS

Nesta pesquisa, foram considerados dois estudos de caso para a aplicação da transformada de Hilbert-Huang: 1) Análise acústica de sinais de vozes saudáveis e desviadas (rugosidade, soprosidade e tensão); e 2) Análise acústica da intensidade dos desvios vocais. Para análise acústica, foram selecionados trechos de 2 segundos dos sinais, cada sinal foi composto por 88.200 amostras. A base de dados, extração de características dos sinais, bem como o método de classificação empregado estão descritos a seguir.

5.1 Base de Dados

A base de dados dos sinais utilizados neste estudo é proveniente do Laboratório Integrado de Estudos da Voz (LIEV), do Departamento de Fonoaudiologia da Universidade Federal da Paraíba (UFPB), o qual faz parte de um projeto intitulado “Integração das medidas acústicas, perceptivas, fisiológicas e de autoavaliação em pacientes com e sem alteração laríngea”, avaliado e aprovado pelo Comitê de Ética em Pesquisa do Centro de Ciências da Saúde/UFPB, com o parecer número 52492/12 (LOPES et al., 2016).

Foram considerados como critérios de elegibilidade para participação do estudo ter idade superior a 18 anos e inferior a 65 anos, do sexo masculino e feminino, que apresentassem queixa vocal e tivessem realizado exame visual laríngeo, com laudo otorrinolaringológico. Foram excluídos indivíduos com alterações cognitivas ou neurológicas que impossibilitassem o preenchimento do questionário utilizado para avaliação. Todos os participantes assinaram o Termo de Consentimento Livre e Esclarecido, autorizando a pesquisa.

As vozes foram coletadas em cabine de gravação com tratamento acústico e ruído inferior a 50 dB NPS (Nível de Pressão Sonora), com taxa de amostragem de 44.100 Hz, com 16 bits por amostra e distância de 10 cm entre o microfone e os lábios do paciente.

Os sinais de vozes dos pacientes foram previamente classificados pela avaliação perceptivo-auditiva, a partir da vogal sustentada / ϵ / (“é”), em frequência e intensidade consideradas como confortáveis e de uso habitual. A vogal / ϵ / foi selecionada para este estudo, pois é uma vogal considerada com a posição mais média no português brasileiro, o que permite uma posição mais neutra e intermediária do trato vocal. Além disso, é a vogal mais utilizada para avaliação da qualidade vocal e exame laríngeo no Brasil (REBELO et al., 2009).

Posteriormente, as vozes foram editadas no *software SoundForge* versão 10.0, sendo eliminados os dois segundos iniciais e finais da emissão da vogal, devido a maior irregularidade nesses trechos, preservando-se o tempo mínimo de três segundos para cada emissão. Para a avaliação perceptivo-auditiva, os sinais foram normalizados, a fim de obter uma padronização na saída de áudio entre -6 dB e 6 dB.

5.1.1 Análise Perceptivo-Auditiva da Voz

A análise acústica realizada nesta pesquisa foi feita a partir da análise perceptivo-auditiva, feita pelos colaboradores do LIEV, responsáveis pela base de dados, da seguinte forma: A sessão de avaliação perceptivo-auditiva ocorreu em ambiente silencioso e foi realizada por um fonoaudiólogo especialista em voz e com mais de 10 anos de experiência em avaliação vocal perceptivo-auditiva.

Para a análise perceptivo-auditiva da voz, foi utilizada uma escala analógica visual, com uma métrica de 0 a 100 mm, avaliando-se a intensidade do desvio vocal (grau geral). A marcação mais próxima do 0 representa menor desvio vocal, e, quanto mais próxima do 100, maiores são as alterações.

Antes da avaliação perceptivo-auditiva, foram utilizados 16 estímulos-âncora da vogal sustentada /ε/ para treinamento do juiz, contendo 4 amostras de indivíduos vocalmente saudáveis, 4 amostras de indivíduos com desvio vocal leve, 4 amostras de indivíduos com desvio vocal de leve a moderado e 4 amostras de indivíduos com desvio vocal intenso. Em cada grau do desvio vocal, havia 2 arquivos de vozes masculinas e 2 arquivos de vozes femininas. O juiz foi orientado a escutar os estímulos-âncora imediatamente antes da análise das vozes desta pesquisa. Cada emissão da vogal sustentada foi apresentada através de caixa de som, em intensidade confortável segundo o avaliador. Em seguida, fez-se a identificação da intensidade do desvio vocal e, caso a emissão fosse considerada desviada, o juiz deveria identificar a qualidade vocal predominante (rugosa, sopro ou tensa).

Ao final da avaliação perceptivo-auditiva, 10% das amostras foram repetidas aleatoriamente, para a análise da confiabilidade da avaliação do juiz por meio do Coeficiente Kappa de Cohen. Os resultados da análise perceptivo-auditiva apresentou coeficiente Kappa de 0,79.

Nesta pesquisa a análise acústica é relacionada com a análise perceptivo-auditiva, cedida pelo Laboratório do Estudo Integrado da Voz da Universidade Federal da Paraíba, onde o estudo foi realizado a partir da escala EAV. Os registros da avaliação do especialista, para fins de análise do resultado da escala EAV, foram classificados de acordo com os pontos de corte definidos na escala numérica proposta no estudo de Yamasaki et al. (2008).

5.1.2 Critério para Escolha dos Sinais

Estudo de Caso 1

Neste Estudo de Caso foram avaliados, do banco de dados, 116 sinais, sendo 30 sinais de vozes normais, 29 sinais de vozes com desvio rugosidade, 28 sinais de vozes com desvio sopro e 29 sinais de vozes com desvio tensão.

Para escolher os sinais foram utilizados critérios relacionados ao grau do parâmetro do desvio, ou seja, caso fosse escolhido mais sinais com grau 2, por exemplo, os resultados poderiam

ser tendenciosos. Logo, fazendo-se uma proporcionalidade com relação aos graus dos sinais que foram avaliados, supõe-se que haja maior confiabilidade nos resultados do classificador.

A Tabela 5.1 apresenta a quantidade de sinais utilizados referentes, ao sinal normal e à intensidade do de cada tipo de desvio (soprosidade, rugosidade e tensão), relacionando-se a quantidade de sinais ao grau correspondente de cada desvio.

Tabela 5.1 – Quantidade de sinais utilizados para o estudo de caso 1.

	Normal	Rugosidade	Soprosidade	Tensão
Grau 1	30	0	0	0
Grau 2	0	10	16	12
Grau 3	0	16	11	15
Grau 4	0	3	1	2
Total	30	29	28	29

Na Tabela 5.2 são apresentados os valores médios e seus desvios padrões dos desvios vocais de acordo com a avaliação perceptivo-auditiva realizada pela escala EAV, para as vozes selecionadas, de acordo com as Tabelas A.1–A.4 apresentadas no Apêndice A.

Tabela 5.2 – Valores médios e desvios padrões de cada classe analisada de acordo com a análise perceptivo-auditiva realizada pela escala EAV.

Classe	Média	Desvio Padrão
Normal	28,13	6,07
Rugosidade	57,58	15,90
Soprosidade	51,46	12,73
Tensão	54,84	13,91

Estudo de Caso 2

Para o Estudo de Caso 2 foram avaliados, da base de dados, 30 sinais de vozes com grau geral 1 (são os mesmos sinais de vozes normais apresentados na Tabela 5.1), 30 sinais de vozes com grau geral 2 de cada tipo de desvio (rugosas, soprosas e tensas) e 30 sinais de vozes com grau geral 3 de cada tipo de desvio (rugosas, soprosas e tensas). Também foram utilizados 30 sinais de vozes combinadas de cada tipo de desvio (rugosas, soprosas e tensas) entre os graus de desvios 2 e 3 (G2G3). Para esta etapa não foram utilizados sinais de vozes com grau geral 4 devido sua quantidade não ser suficiente para realizar a classificação, pois conforme apresentado na Tabela 5.1 o total de sinais com grau geral 4 são apenas 6 sinais.

Para as discriminações entre os graus dos desvios foram escolhidos os sinais de vozes com desvio rugosidade, soprosidade e tensão com a mesma proporcionalidade. A escolha dos sinais foi feita de forma que não apresentasse uma tendência nos resultados. Já que escolha de um tipo de desvio predominante pode levar ao aumento de acurácia.

Na Tabela 5.3, é apresentada a quantidade de sinais utilizados referentes a cada grau de intensidade, relacionando-os aos desvios correspondentes.

Tabela 5.3 – Quantidade de sinais utilizados para a discriminação com relação às intensidades dos desvios vocais.

	G1	G2	G3	G2G3
Normal	30	0	0	0
Rugosidade	0	10	12	10
Soprosidade	0	10	8	10
Tensão	0	10	10	10
Total	30	30	30	30

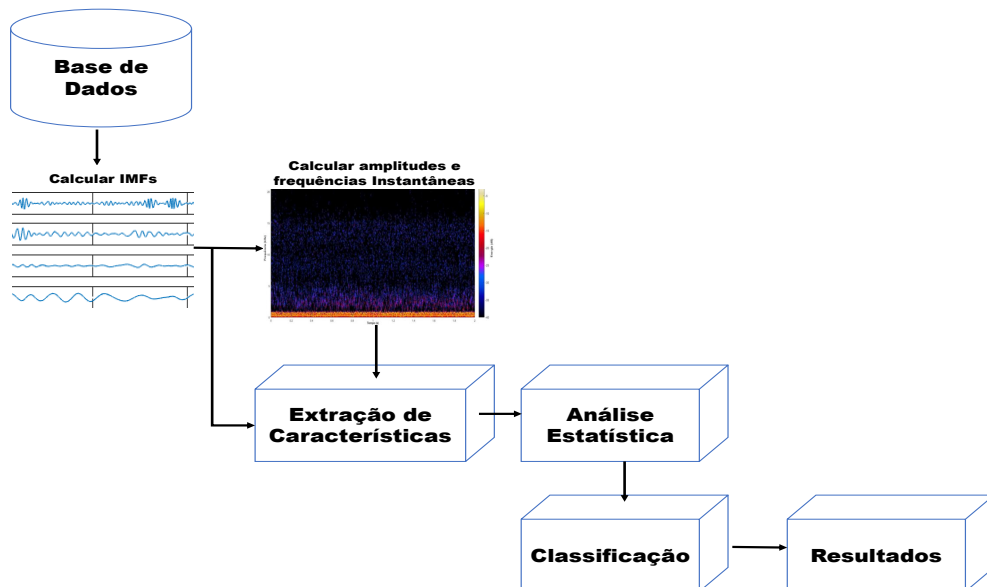
Na Tabela 5.4 são apresentados os valores médios e seus desvios padrões de intensidade do grau do desvio de acordo com a avaliação perceptivo-auditiva realizada pela escala EAV, para as vozes selecionadas, de acordo com as Tabelas A.5–A.7 apresentadas no Apêndice A.

Tabela 5.4 – Valores médios e desvios padrões de cada classe analisada, com relação as intensidades dos desvios, de acordo com a análise perceptivo-auditiva realizada pela escala EAV.

Classe	Média	Desvio Padrão
G1	28,13	6,07
G2	43,58	4,11
G3	61,11	8,09
G2G3	52,50	10,96

5.2 Metodologia

A metodologia empregada é apresentada conforme o diagrama da Figura 5.1.

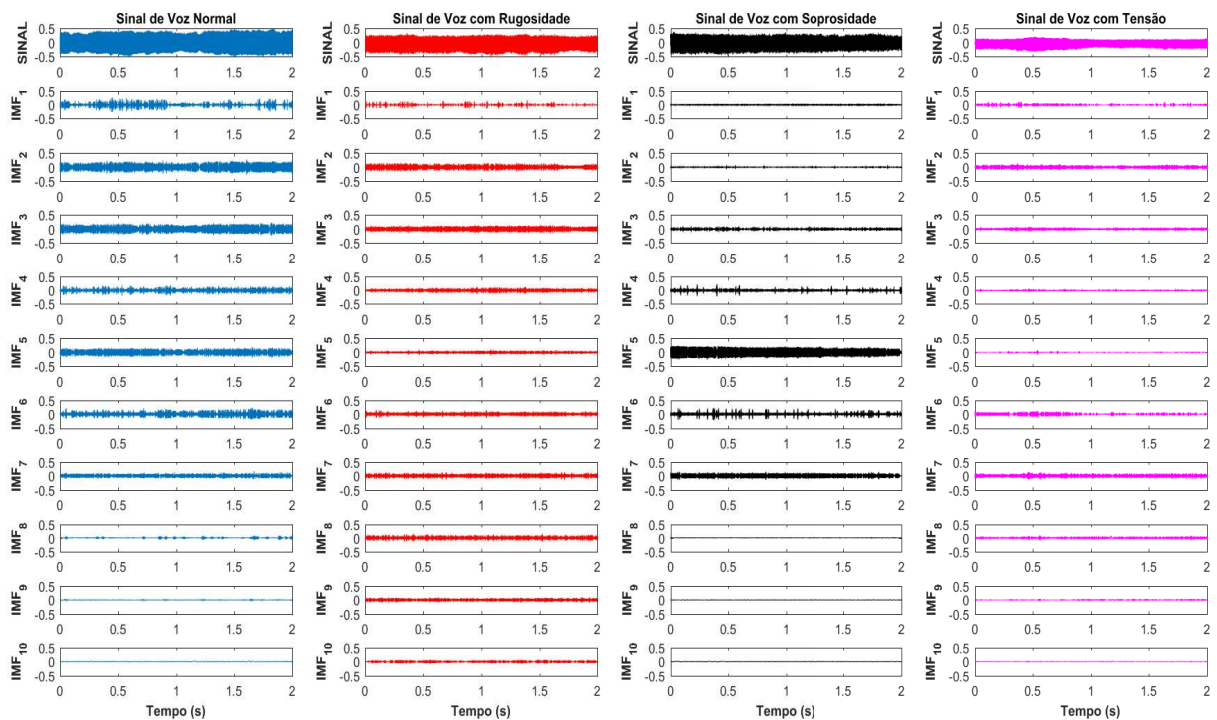
Figura 5.1 – Diagrama da metodologia empregada.

Fonte: Elaboração própria, 2018.

Os estudos de simulação foram implementados usando o software Matlab®R2016a. Para a decomposição dos sinais em IMF's foram usadas funções da biblioteca disponibilizadas por Flandrin et al. (2007). Para calcular as amplitudes e frequências instantâneas foram usadas funções disponibilizadas por Auger et al. (1995).

Os sinais de vozes selecionados da base de dados do LIEV/UFPB passaram pelo método EMD, a fim de obter um conjunto de IMFs de cada sinal de voz. Obtidas as IMFs, foram calculadas amplitudes e frequências instantâneas do sinais de vozes em análise, conforme descrito nas Equações 4.17 a 4.21. Na extração de características foram obtidas as características representativas de cada sinal, oriundas da transformada de Hilbert-Huang. Como cada sinal gerou um número de IMFs diferentes, decidiu-se então trabalhar com as dez primeiras IMFs de cada sinal, tomando como base o estudo de Kaleem et al. (2013). A Figura 5.2 mostra o resultado final da EMD decompostas em 10 IMFs. Para avaliar o potencial discriminativo das características referente a cada IMF, foi realizada uma análise estatística, através de testes de hipótese. As características representativas das IMFs de cada sinal de voz foram submetidas a um classificador, que faz a discriminação entre voz alterada ou normal, classificando através do grau ou tipo de desvio vocal. O classificador utilizado foi o *Multilayer Perceptron* com o algoritmo de aprendizado supervisionado do gradiente conjugado escalonado.

Figura 5.2 – Quatro sinais de vozes decompostos em 10 IMFs. Da esquerda para a direita: voz normal, voz com desvio rugosidade, voz com desvio soproidade e voz com desvio tensão.



Fonte: Dados da pesquisa, 2018.

Observando a Figura 5.2 nota-se que na primeira, segunda e terceira IMFs referente ao sinal de voz normal possuem energia superior em relação aos outros sinais de voz, enquanto que na quinta IMF este papel passa a ser do sinal de voz com desvio soproidade. Já na oitava IMF a maior energia está no sinal de voz com soproidade. Pode ser notado, ainda, que nas mais baixas frequências, no caso na décima IMF, a forma de onda já não contém tanta informação agregada ao sinal quanto as IMFs anteriores. Este é um indício de que a decomposição baseada em EMD fornece as informações tempo-frequência que permitem uma diferenciação entre os sinais com desvios vocais dos sinais saudáveis.

5.2.1 Extração de Características

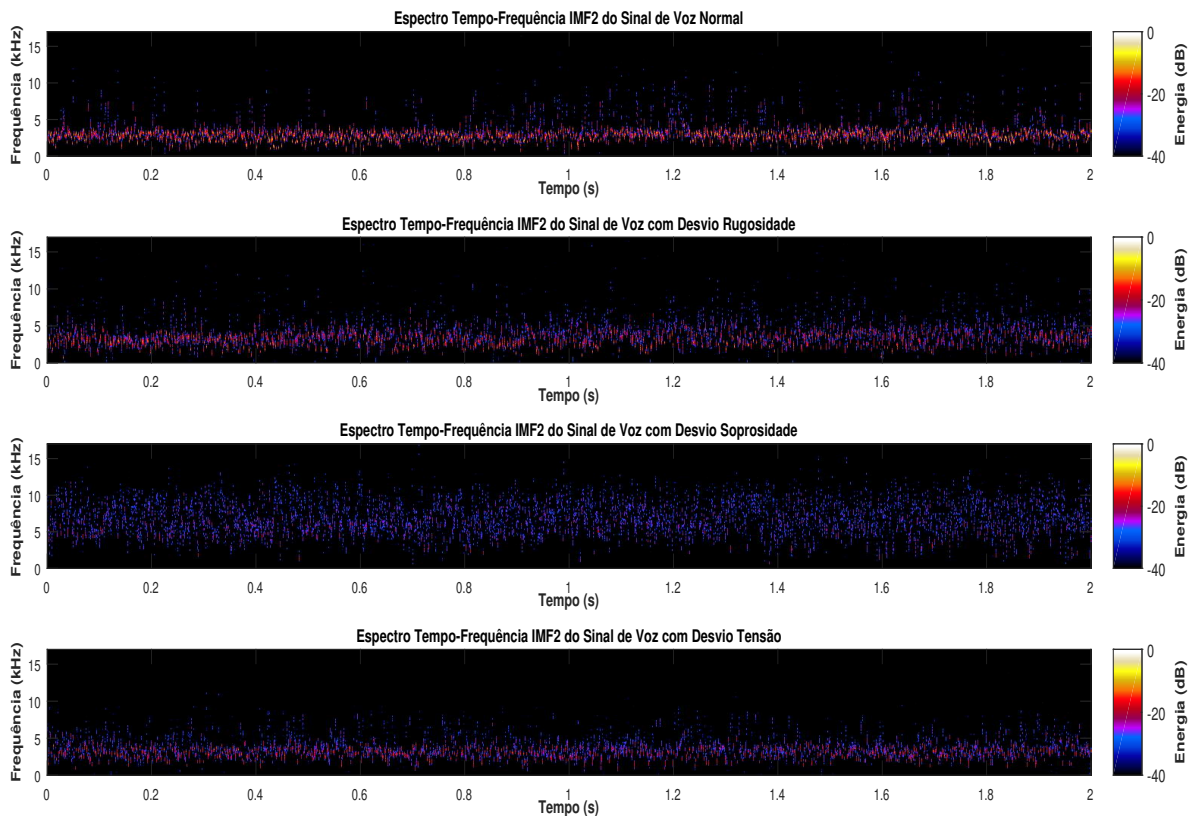
Foram extraídas características temporais e espectrais de cada IMF. A primeira característica extraída foi a energia do sinal, que representa sua intensidade sonora. A Equação 5.1 define como é calculada a energia (E_i) do sinal para cada IMF (RAVINDRAN; NAIR, 2015; KALEEM et al., 2013).

$$E_i = \sum_{n=1}^N c_i^2[n], i = 1, 2, 3, \dots, 10 \quad (5.1)$$

Para as Equações de 5.1 a 5.5, N representa o tamanho das amostras da IMF e o índice i representa a IMF associada. A energia foi escolhida como característica, pois a amplitude do sinal em vozes saudáveis tem menos variações do que em sinais com desvios (KALEEM et al., 2010).

Devido a voz com desvio soproiedade ter ruídos em frequências mais altas (LOPES; ALVES; MELO, 2017), espera-se que esses ruídos se espalhem pelo espectro do HSA. A Figura 5.3 apresenta o espectro HSA da IMF₂ entre os sinais de vozes normal, com desvio rugosidade, soproiedade e tensão. Pode ser observado que para essa IMF, o sinal com desvio soproiedade tem mais ruídos em frequências mais altas do que os sinais de vozes normal, com o desvio rugosidade e tensão.

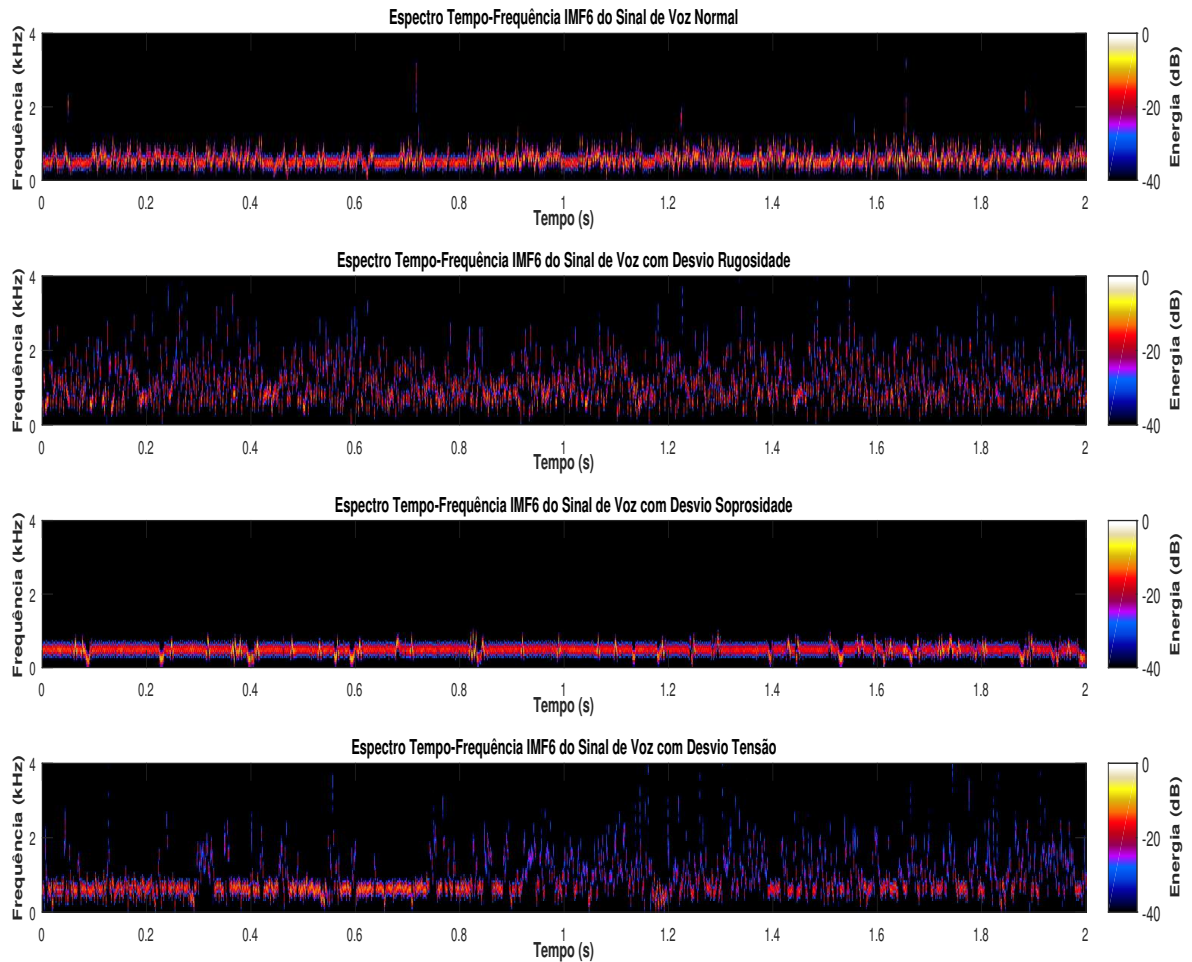
Figura 5.3 – Espectro de Hilbet-Huang em relação a IMF₂ dos sinais de vozes normal, com desvio rugosidade, soproiedade e tensão.



Fonte: Dados da pesquisa, 2018.

Já as vozes com desvio rugosidade possuem mais ruídos em frequências médias (LOPES; ALVES; MELO, 2017). A Figura 5.4 mostra o espectro HSA da IMF₆ entre os sinais de vozes normal, com desvio rugosidade, soproidade e tensão. Pode ser observado na Figura 5.4 que, para essa IMF, o sinal com desvio rugosidade tem mais ruídos em frequências média quando comparado aos sinais de vozes normal, com desvio soproidade e tensão.

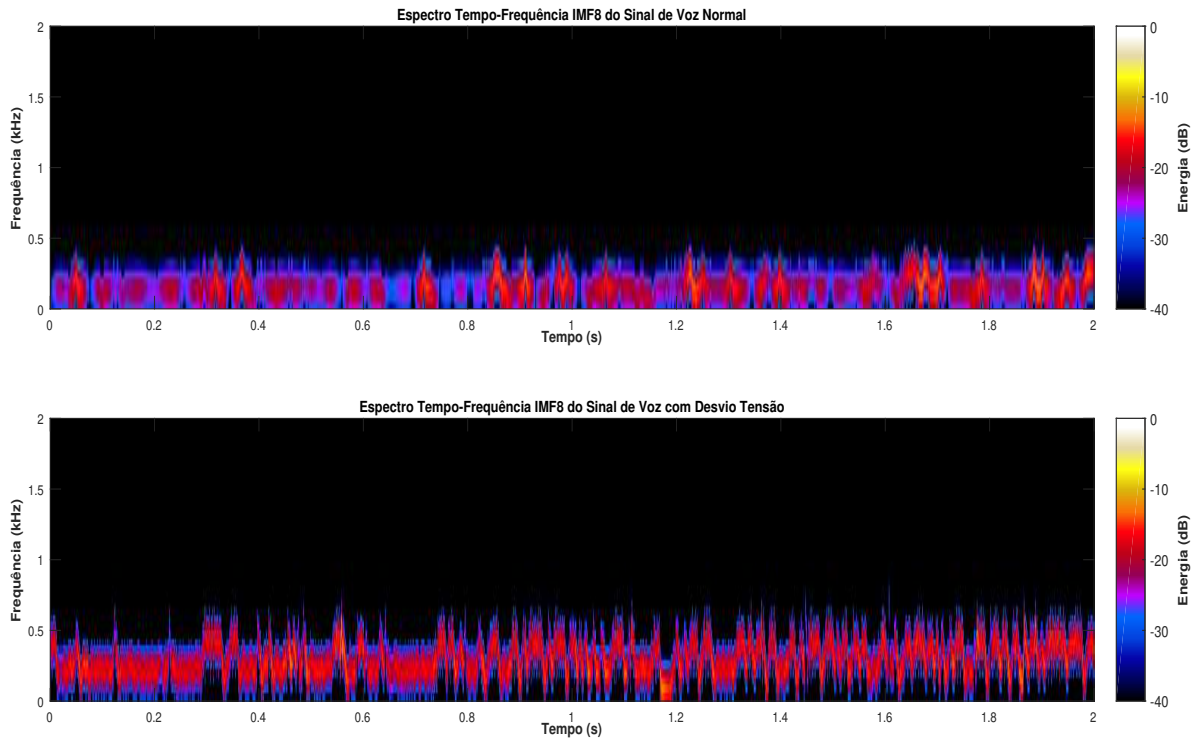
Figura 5.4 – Espectro de Hilbet-Huang em relação a IMF₆ dos sinais de vozes normal, com desvio rugosidade, soproidade e tensão.



Fonte: Dados da pesquisa, 2018.

Lopes, Alves e Melo (2017), em seu estudo observaram frequências dos traçados espectrográficos de sinais de vozes com desvio tensão entre 1 kHz e 3 kHz. Eles identificaram que esses sinais podem ser confundidos com sinais de vozes normal no traçado espectrográfico, pois a presença de harmônicos nessa faixa do espectro relaciona-se tanto a uma boa qualidade vocal, quanto à presença de aumento de tensão vocal no trato supraglótico. Examinando essas informações, foram observadas as IMFs com frequências abaixo de 1 kHz e foi percebido que o sinal de voz com desvio tensão em alguns trechos tem uma dispersão de frequência instantânea maior que no sinal de voz normal. Esse desvio pode ser visualizado na Figura 5.5.

Figura 5.5 – Espectro de Hilbet-Huang em relação a IMF₈ dos sinais de vozes normal e com desvio tensão.



Fonte: Dados da pesquisa, 2018.

Com base nas informações do comportamento do sinais de vozes na escala energia-tempo-frequência, foram extraídas quatro características do espectro HSA de cada IMF, duas em relação a contribuição do tempo e duas em relação a contribuição da frequência. Com relação às contribuições referentes ao tempo, tem-se a dispersão da densidade de energia temporal instantânea (*DETI*), que é demonstrada pela Equação 5.2 (RAVINDRAN; NAIR, 2015; KALEEM et al., 2013). Esta característica destaca o quanto os dados da energia instantânea são semelhantes, pois calcula a distância de uma energia instantânea para seu valor esperado.

$$DETI_i = \frac{1}{N} \sum_{n=1}^N |a_i[n] - \bar{a}_i|^2, i = 1, 2, 3, \dots, 10 \quad (5.2)$$

Nas Equações 5.2, 5.3 e 5.5, *a* representa amplitude instantânea. A outra característica analisada é a amplitude instantânea ponderada pelo tempo (*APT*), apresentada na Equação 5.3 (RAVINDRAN; NAIR, 2015; KALEEM et al., 2013). Como os valores de energia instantânea são comuns ao tempo e a frequência instantânea esta característica tem a finalidade de encontrar um valor marginal referente a energia instantânea com o tempo.

$$APT_i = \frac{1}{N} \sum_{n=1}^N (n \times a_i[n]), i = 1, 2, 3, \dots, 10 \quad (5.3)$$

Com relação às contribuições referentes à frequência, tem-se a dispersão da frequência instantânea (*DFI*), descrita pela Equação 5.4 (RAVINDRAN; NAIR, 2015; KALEEM et al.,

2013). Esta característica destaca o quanto os dados da frequência instantânea são semelhantes, pois calcula a distância de uma frequência instantânea para seu valor esperado.

$$DFI_i = \frac{1}{N} \sum_{n=1}^N |f_i[n] - \bar{f}_i|^2, i = 1, 2, 3, \dots, 10 \quad (5.4)$$

Nas Equações 5.4 e 5.5, f representa frequência instantânea. A outra característica analisada é a amplitude instantânea ponderada pela frequência instantânea (APF). A Equação 5.5 demonstra esta relação (RAVINDRAN; NAIR, 2015; KALEEM et al., 2013). Como os valores de energia instantânea são comuns ao tempo e a frequência instantânea esta característica tem o objetivo de encontrar um valor marginal referente a energia instantânea com a frequência instantânea.

$$APF_i = \frac{1}{N} \sum_{n=1}^N (a_i[n] \times f_i[n]), i = 1, 2, 3, \dots, 10 \quad (5.5)$$

5.2.2 Processo de Classificação

Apoiada nos estudos sobre o cérebro humano, a rede neural artificial é um sistema inteligente com capacidade de aprendizado e reconhecimento de padrões nas informações apresentadas em suas entradas, além de desenvolver esta capacidade para a análise de novos dados. A rede neural artificial é um processador que interliga um conjunto de células computacionais (HAYKIN, 2007; MACKAY; KAY, 2003).

A capacidade de aprender e armazenar as informações são atributos das redes neurais que se comparam ao cérebro humano. Seus neurônios interconectados operam de forma complexa, paralela e não linear. A vantagem do processamento paralelo é possuir certa tolerância a falha, pois se alguns neurônios falharem não prejudica substancialmente o desempenho do restante do sistema. O atributo de não linearidade, por sua vez, permite que o sistema trabalhe com sinais de entrada não lineares (HAYKIN, 2007).

Dentre os diversos modelos de rede neural artificial, a rede *Multilayer Perceptron* tem sido muito utilizada para o reconhecimento de padrões em sinais de voz (BARAVIEIRA, 2016; MORO-VELÁZQUEZ et al., 2015; STRÁNÍK; ČMEJLA; VOKŘÁL, 2014). Este modelo possui diversas camadas, dentre elas a camada de entrada, uma ou mais camadas ocultas e a camada de saída. A camada oculta possui neurônios ocultos que intervêm de maneira útil entre a camada neural de entrada e a de saída, tornando a rede capaz de realizar cálculos mais complexos. O treinamento de uma rede *Multilayer Perceptron* é do tipo supervisionado por meio de algoritmo de retropropagação do erro (DUDA; HART; STORK, 2001).

De acordo Panchal F. e Panchal M. (2014) não há uma equação exata que dará o número correto de neurônios na camada oculta para se ter uma quantidade menor de erro, o que se deve fazer é por tentativa e erro, tentar utilizar um número pequeno de neurônios, e ir aumentando gradualmente e acompanhar os resultados. A equação que é sugerida para se aproximar do

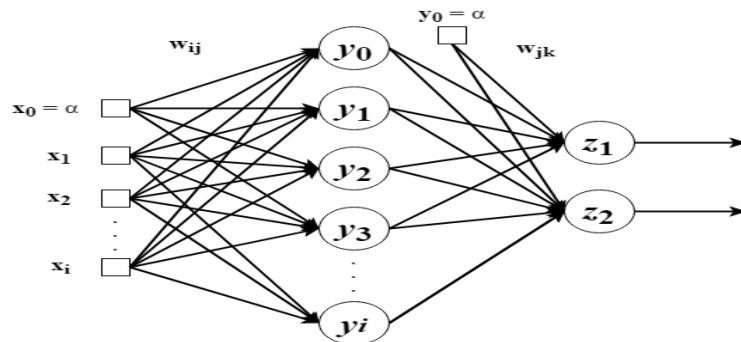
número de neurônios escondidos é dada pela Equação 5.6.

$$N_{oc} = \frac{2}{3}(N_{en} + N_{sa}), \quad (5.6)$$

em que N_{oc} representa o número de neurônios em camada oculta, N_{en} representa o número de neurônios na entrada e N_{sa} representa o número de neurônios na saída.

O classificador utilizado nesta pesquisa foi o *Multilayer Perceptron* com o algoritmo de aprendizado supervisionado do gradiente conjugado escalonado. Este algoritmo, foi desenvolvido por Moller em 1993, para evitar a busca unidimensional a cada iteração, utilizando uma abordagem de Levenberg-Marquardt, cujo objetivo é fazer um escalonamento do passo de ajuste. A busca unidimensional envolve um grande número de avaliações da função ou de sua derivada, tornando o processo com alto custo computacional (SILVA; BARBOSA, 2002). A arquitetura utilizada neste estudo foi a rede *feedforward* de múltiplas camadas por possuir uma camada oculta ou intermediária. Os sinais de vozes utilizados foram separados, aleatoriamente, em conjuntos distintos, dos quais 2/3 foram utilizados para treinamento e 1/3 para validação e teste. Para maior confiabilidade, a rede neural foi executada 10 vezes, a fim de realizar a validação cruzada. Para a camada oculta foi utilizado o critério de neurônios conforme a Equação 5.6 e foram utilizados dois neurônios na camada de saída. Na camada de entrada, a quantidade de neurônios irá variar de acordo com o número de características empregadas. A Figura 5.6 representa a topologia da rede neural empregada.

Figura 5.6 – Rede *feedforward* MLP utilizada.



Fonte: Autoria própria, 2018.

5.2.3 Avaliação e Interpretação

Um das formas de avaliar e validar um conjunto de dados que foram classificados é por meio de uma matriz de confusão. Esta matriz é uma ferramenta importante que determina se o valor classificado correspondeu ou não ao valor real, de forma que todos os casos em cada categoria são contabilizados e os totais são exibidos na matriz (CARDOSO; LIMA, 2018).

A validação representa a capacidade que um método de teste possui para identificar corretamente a classe de um indivíduo dentro de uma população. Para avaliar a precisão dos resultados obtidos através do classificador foram empregadas, neste trabalho, três medidas normalmente

usadas no meio científico, são elas: acurácia, sensibilidade e especificidade (STOJANOVIĆ et al., 2014).

A Tabela 5.5 mostra que essas medidas estão relacionadas com os resultados de uma classificação e o diagnóstico verdadeiro. O teste é considerado positivo (desvio) ou negativo (saudável), e o desvio presente ou ausente. O teste está correto quando ele é positivo na presença do desvio (Verdadeiro Positivo-*VP*), ou negativo na ausência do desvio (Verdadeiros Negativo-*VN*). Também, o teste está errado quando ele é positivo na ausência do desvio (Falso Positivo-*FP*), ou negativo quando o desvio está presente (Falso Negativo-*FN*) (CARDOSO; LIMA, 2018).

Tabela 5.5 – Matriz de confusão em um teste de diagnóstico da presença ou ausência do desvio.

Resultado	Desvio	
	Presente	Ausente
Positivo	Verdadeiro Positivo	Falso Positivo
Negativo	Falso Negativo	Verdadeiro Negativo

Fonte: Adaptado de Costa, 2012.

Acurácia (*ACU*) mede a capacidade do teste de identificar corretamente quando há e quando não há presença do desvio. É definida como a relação entre o número de casos corretamente classificados e todos os casos expostos ao classificador (VIEIRA, 2014; COSTA, 2012):

$$ACU = \frac{VP + VN}{VP + VN + FP + FN} \quad (5.7)$$

A medida de sensibilidade (*SENS*) é a capacidade do teste em identificar corretamente o desvio entre aqueles que o possuem. É definida pela relação entre o número de casos corretamente classificados com a presença do distúrbio e a quantidade total de casos com o distúrbio (VIEIRA, 2014; COSTA, 2012):

$$SENS = \frac{VP}{VP + FN} \quad (5.8)$$

A medida de especificidade (*ESP*) é a capacidade do teste em excluir corretamente aqueles que não possuem o desvio. É definida pela relação entre o número de casos corretamente classificados como saudável e a quantidade total de casos saudáveis (VIEIRA, 2014; COSTA, 2012):

$$ESP = \frac{VN}{VN + FP} \quad (5.9)$$

A representação das medidas de sensibilidade e especificidade é mais clara quando se trata da discriminação entre um sinal de voz com normal e um sinal de voz com desvio. Quando há discriminação entre classes com desvios, é necessário que seja definido, no classificador, qual grupo de sinais terá sua correta classificação medida pela sensibilidade e especificidade (VIEIRA, 2014). No Quadro 5.1, estão apresentados todos os casos de discriminação considerados neste estudo, relacionando as medidas de sensibilidade e especificidade com cada classe envolvida.

Quadro 5.1 – Relação das medidas medidas de sensibilidade e especificidade com as classes envolvidas nos casos de discriminação.

Casos de Discriminação	Sensibilidade	Especificidade
Normal × Desvios	Taxa de correta classificação dos sinais de vozes com desvios	Taxa de correta classificação dos sinais de vozes normais
Normal × Soproso	Taxa de correta classificação dos sinais de vozes soprosas	Taxa de correta classificação dos sinais de vozes normais
Normal × Rugoso	Taxa de correta classificação dos sinais de vozes rugosas	Taxa de correta classificação dos sinais de vozes normais
Normal × Tenso	Taxa de correta classificação dos sinais de vozes tensas	Taxa de correta classificação dos sinais de vozes normais
Rugoso × Tenso	Taxa de correta classificação dos sinais de vozes tensas	Taxa de correta classificação dos sinais de vozes rugosas
Rugoso × Soproso	Taxa de correta classificação dos sinais de vozes soprosas	Taxa de correta classificação dos sinais de vozes rugosas
Soproso × Tenso	Taxa de correta classificação dos sinais de vozes tensas	Taxa de correta classificação dos sinais de vozes soprosas
G1 × G2	Taxa de correta classificação dos sinais de vozes com grau geral 2	Taxa de correta classificação dos sinais de vozes com grau geral 1
G1 × G3	Taxa de correta classificação dos sinais de vozes com grau geral 3	Taxa de correta classificação dos sinais de vozes com grau geral 1
G2 × G3	Taxa de correta classificação dos sinais de vozes com grau geral 3	Taxa de correta classificação dos sinais de vozes com grau geral 2
G1 × G2G3	Taxa de correta classificação dos sinais de vozes com graus 2 e 3	Taxa de correta classificação dos sinais de vozes com grau geral 1

6 RESULTADOS E DISCUSSÃO

Neste Capítulo, são apresentados os resultados obtidos nesta pesquisa para os dois estudos de caso por meio das medidas de acurácia, sensibilidade e especificidade, que compreendem conforme o diagrama da metodologia empregada referente a Figura 5.1, às etapas da análise estatística, classificação e resultados. As características utilizadas para as discriminações referentes aos dois estudos de caso, são: energia, dispersão da densidade de energia temporal instantânea, amplitude instantânea ponderada pelo tempo, dispersão da frequência instantânea e amplitude instantânea ponderada pela frequência instantânea. São apresentados os resultados dos testes estatísticos com o objetivo de evidenciar o potencial discriminativo de cada uma dessas características.

6.1 Análise Estatística

Após a extração das características, foram realizados os testes de hipóteses com a utilização do Matlab®R2016a, com o objetivo de validar as conclusões a respeito do potencial discriminativo de cada característica. O nível de significância considerado em todos os testes estatísticos realizados foi de 5%, isso significa que para valores de $p \leq 0,05$ a hipótese nula foi rejeitada.

Os testes estatísticos utilizados para inferência das características requerem o conhecimento sobre a distribuição dos valores da medida considerada, buscando saber se esta distribuição é normal ou não (MOTA; FILHO, 2009). Para o teste de normalidade, foi feito o teste de Shapiro-Wilk. Esse teste é considerado quando o número de amostras dentro de um mesmo grupo for inferior a 50. (SHAPIRO; WILK, 1965 apud VIEIRA, 2014).

Quando os testes de normalidade sugerem que a distribuição dos resultados é normal, são aplicados testes paramétricos. Que para esse estudo foi aplicado *t-student* (OWEN, 1965 apud VIEIRA, 2014). Caso contrário, os testes de hipótese a serem aplicados são testes não-paramétricos. Que para esse estudo foi aplicado o teste de Mann-Whitney (MANN; WHITNEY, 1947 apud VIEIRA, 2014). Para que sejam aplicados testes paramétricos, é necessário que ambas as distribuições dos valores obtidos da medida em questão sejam normais. Das Tabelas 6.1 a 6.11 foram destacadas em negrito as medidas com hipótese nula rejeitada, apenas as características que tiveram a hipótese nula rejeitada foram usadas no classificador.

Para o estudo de caso 1 foram realizados testes de hipóteses entre sinais de vozes normais e com desvios (rugosidade, sopro e tensão), entre os sinais de vozes normais e com desvio rugosidade, entre os sinais de vozes normais e com desvio sopro, entre os sinais normais e com desvio tensão, entre os sinais de vozes com desvio rugosidade e com desvio sopro, entre os sinais de vozes com desvio rugosidade e com desvio tensão e entre os sinais de vozes com desvio sopro e com desvio tensão.

Para o estudo de caso 2 os testes de hipóteses foram realizados para identificar o potencial discriminativo entre o grau geral 1 e uma combinação do grau geral 2 e 3, entre o grau geral 1 e o grau geral 2, entre grau geral 1 e o grau geral 3, entre o grau geral 2 e o grau geral 3.

6.1.1 Análise Estatística para o Estudo de Caso 1

A Tabela 6.1 apresenta o *valor-p*, representado o potencial discriminativo das características de cada IMF entre os sinais de vozes normais e com desvios (rugosidade, sopro e tensão). Do total de 50 características somente 2 medidas tiveram a hipótese nula rejeitada, essas características foram a energia e a amplitude instantânea ponderada pela frequência instantânea referente a IMF₅.

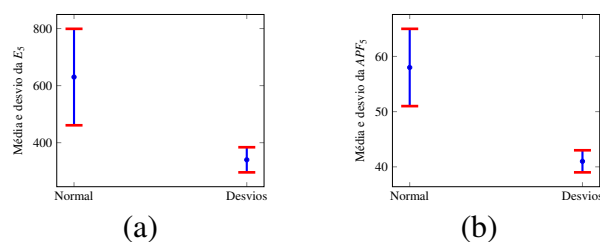
Tabela 6.1 – Potencial discriminativo entre os sinais de vozes normais e com desvios (rugosidade, sopro e tensão).

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,225	0,248	0,458	0,256	0,043
$DETI_i$	0,267	0,218	0,507	0,147	0,143
APT_i	0,238	0,221	0,420	0,402	0,068
DFI_i	0,123	0,694	0,279	0,080	0,747
APF_i	0,402	0,216	0,727	0,549	0,045
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,147	0,821	0,511	0,606	0,979
$DETI_i$	0,086	0,958	0,251	0,776	0,882
APT_i	0,191	0,781	0,536	0,597	0,861
DFI_i	0,209	0,474	0,490	0,327	0,624
APF_i	0,193	0,638	0,656	0,670	0,718

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

Conforme Figura 6.1 observa-se que para E_5 e APF_5 as médias são maiores em sinais de vozes normais do que em vozes com desvios. Com isso, é possível entender que para a IMF₅ no espectro HSA o subconjunto energia-frequência a energia instantânea tem em média valores maiores para sinais de vozes normais quando comparado aos sinais de vozes com desvios (rugosidade, sopro e tensão).

Figura 6.1 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvios: (a) características E_5 ; (b) característica APF_5 .



Fonte: Dados da pesquisa, 2018.

A Tabela 6.2 apresenta o *valor-p*, representando o potencial discriminativo das características de cada IMF entre os sinais de vozes normais e com desvio rugosidade. Do total de 50 características, 10 características tiveram a hipótese nula rejeitada. Observa-se que dentre as características com hipótese nula rejeitada, 7 características estão entre as IMF₅ e IMF₆.

Tabela 6.2 – Potencial discriminativo entre sinais de vozes normais e com desvio rugosidade.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,544	0,096	0,105	0,709	0,010
$DETI_i$	0,488	0,088	0,375	0,236	0,020
APT_i	0,589	0,278	0,132	0,651	0,002
DFI_i	0,995	0,183	0,075	0,011	0,448
APF_i	0,008	0,933	0,170	0,133	0,001
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,053	0,181	0,732	0,632	0,937
$DETI_i$	0,014	0,166	0,875	0,813	0,447
APT_i	0,027	0,457	0,863	0,664	0,989
DFI_i	0,665	0,426	0,212	0,146	0,003
APF_i	0,022	0,093	0,260	0,400	0,626

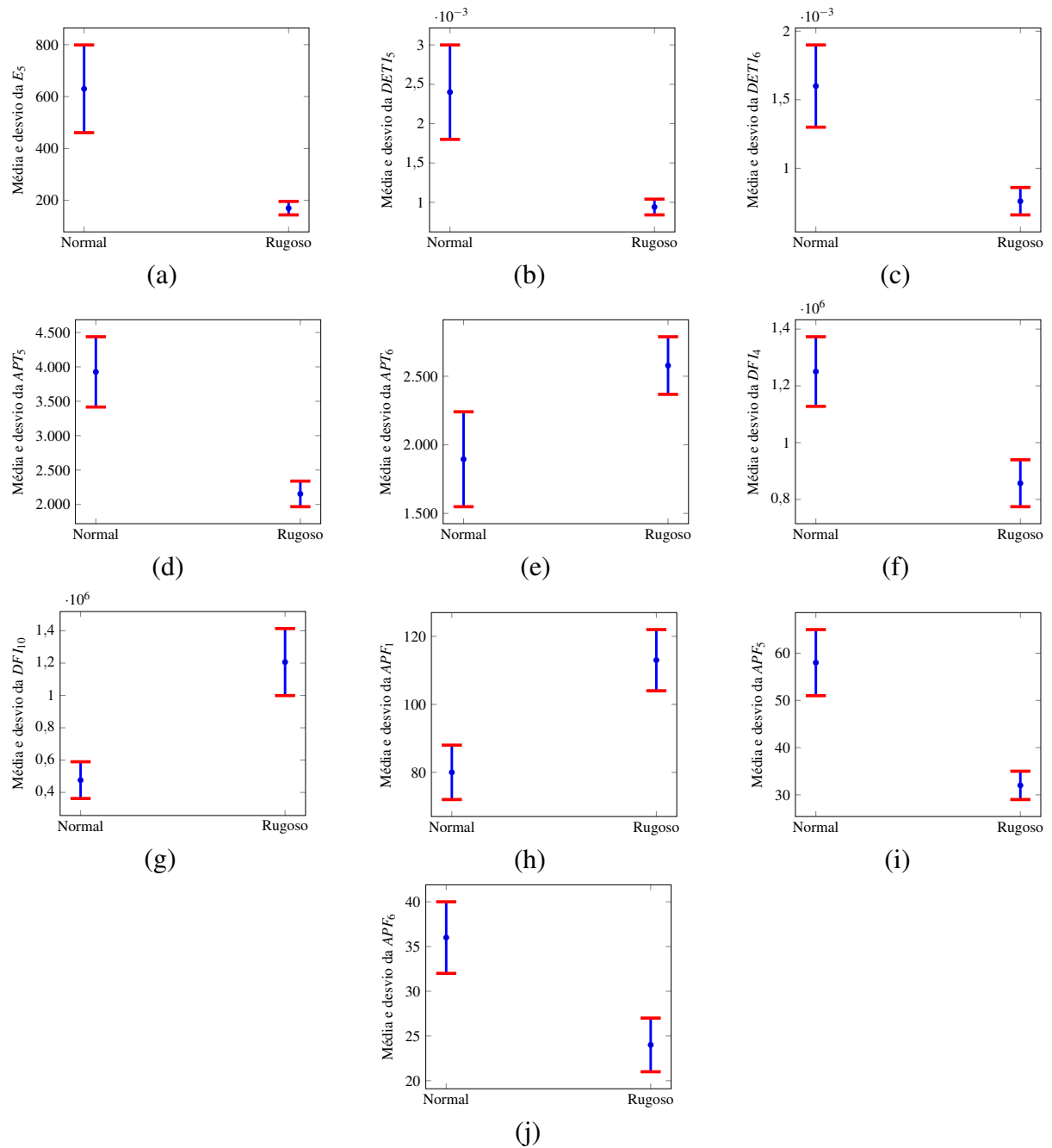
E: Energia; *DETI*: Dispersão da Densidade de Energia Temporal Instantânea; *APT*: Amplitude Instantânea ponderada pelo Tempo; *DFI*: Dispersão da Frequência Instantânea; *APF*: Amplitude Instantânea ponderada pela Frequência Instantânea.

De acordo com a Figura 6.2 para APF_5 e APF_6 as médias indicam que no subconjunto energia-frequência do HSA a energia instantânea têm maiores valores para sinais de vozes normais que em sinais de vozes com rugosidade. Já em APF_1 , a energia instantânea no subconjunto energia-frequência do HSA têm valores maiores para sinais de vozes com desvio rugosidade do que sinais de vozes normais.

A média da intensidade sonora da IMF₅ para sinais de vozes normais é maior do que em sinais de vozes com rugosidade. Para DFI_4 as variâncias em sinais de vozes normais são maiores que em sinais de vozes com desvio rugosidade. Isso indica que a frequência instantânea dos sinais de vozes normais tem uma dispersão maior do que em sinais de vozes com desvio rugosidade para a IMF₄. Já em DFI_{10} as variâncias em sinais de vozes normais são menores que em sinais de vozes com desvio rugosidade. Isso indica que a frequência instantânea dos sinais de vozes normais tem uma dispersão menor do que em sinais de vozes com desvio rugosidade para a IMF₁₀.

Para $DETI_5$ e $DETI_6$ suas variâncias em sinais de vozes normais são maiores que em sinais de vozes com desvio rugosidade. Isso aponta que a energia instantânea dos sinais de vozes normais tem uma dispersão maior do que em sinais de vozes com desvio rugosidade para a IMF₅ e IMF₆. As médias de APT_5 e APT_6 apontam que no subconjunto energia-tempo do HSA a energia instantânea têm maiores valores para sinais de vozes normais que em sinais de vozes com desvio rugosidade.

Figura 6.2 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvio rugosidade: (a) característica E_5 ; (b) característica $DETI_5$, (c) característica $DETI_6$, (d) característica APT_5 , (e) característica APT_6 , (f) característica DFI_4 , (g) característica DFI_{10} , (h) característica APF_1 , (i) característica APF_5 , (j) característica APF_6 .



Fonte: Dados da pesquisa, 2018.

A Tabela 6.3 apresenta o *valor-p*, representando o potencial discriminativo das características de cada IMF entre os sinais de vozes normais e com desvio soprosoidade. Do total de 50 características, 9 características tiveram a hipótese nula rejeitada. Observa-se que dentre as características com hipótese nula rejeitada, 8 características estão entre as IMF₂ e IMF₃.

Tabela 6.3 – Potencial discriminativo entre os sinais de vozes normais e com desvio soprosoidade.

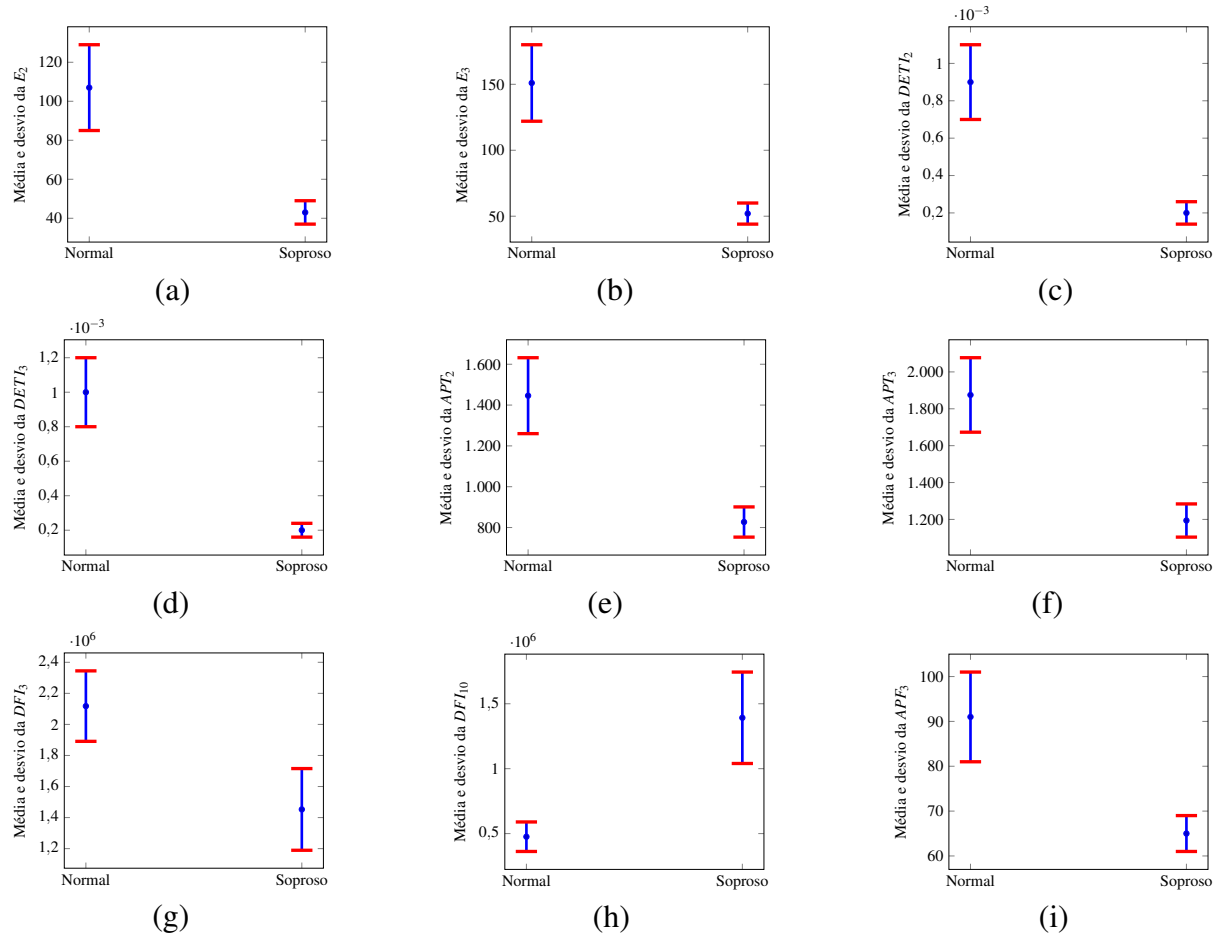
Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,056	0,002	0,003	0,440	0,351
$DETI_i$	0,068	0,001	0,002	0,850	0,307
APT_i	0,446	0,004	0,004	0,374	0,512
DFI_i	0,088	0,201	0,015	0,593	0,073
APF_i	0,060	0,097	0,028	0,056	0,210
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,752	0,283	0,430	0,783	0,167
$DETI_i$	0,720	0,926	0,134	0,575	0,217
APT_i	0,779	0,201	0,700	0,567	0,299
DFI_i	0,297	0,408	0,114	0,506	0,014
APF_i	0,241	0,236	0,053	0,076	0,140

E: Energia; *DETI*: Dispersão da Densidade de Energia Temporal Instantânea; *APT*: Amplitude Instantânea ponderada pelo Tempo; *DFI*: Dispersão da Frequência Instantânea; *APF*: Amplitude Instantânea ponderada pela Frequência Instantânea.

A Figura 6.3 demonstra as médias e os desvios entre os sinais de vozes normais e com desvio soprosoidade das características individuais. Em conformidade com essa figura é possível identificar que todas as médias são maiores para os sinais de vozes normais do que para os sinais de vozes com desvio soprosoidade, exceto para DFI_{10} que os sinais de vozes com desvio soprosoidade possuem média maior que os sinais de vozes normais.

Conforme Figura 6.3 as médias e desvios estatísticos apontam que para energia, energia instantânea com suas ponderações, variância energia-tempo do HSA, variância energia-frequência do HSA em média são maiores em sinais de vozes normais quando comparados a sinais de vozes com desvio soprosoidade para a IMF₂ e IMF₃. Isso pode ser explicado pois os sinais de vozes com desvio soprosoidade têm componentes de energia em frequências mais altas e quando são decompostas pela EMD as componentes de energia ficam mais evidentes através das IMFs. Como a EMD evidencia a característica local a IMF de um sinal de voz normal não necessariamente tem a mesma faixa de frequência de um sinal de voz com soprosoidade, isso explica o motivo das médias do sinal de voz normal serem maiores do que as médias dos sinais de vozes com desvio soprosoidade.

Figura 6.3 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvio soprosidade: (a) característica E_2 ; (b) característica E_3 , (c) característica $DETI_2$, (d) característica $DETI_3$, (e) característica APT_2 , (f) característica APT_3 , (g) característica DFI_3 , (h) característica DFI_{10} , (i) característica APF_3 .



Fonte: Dados da pesquisa, 2018.

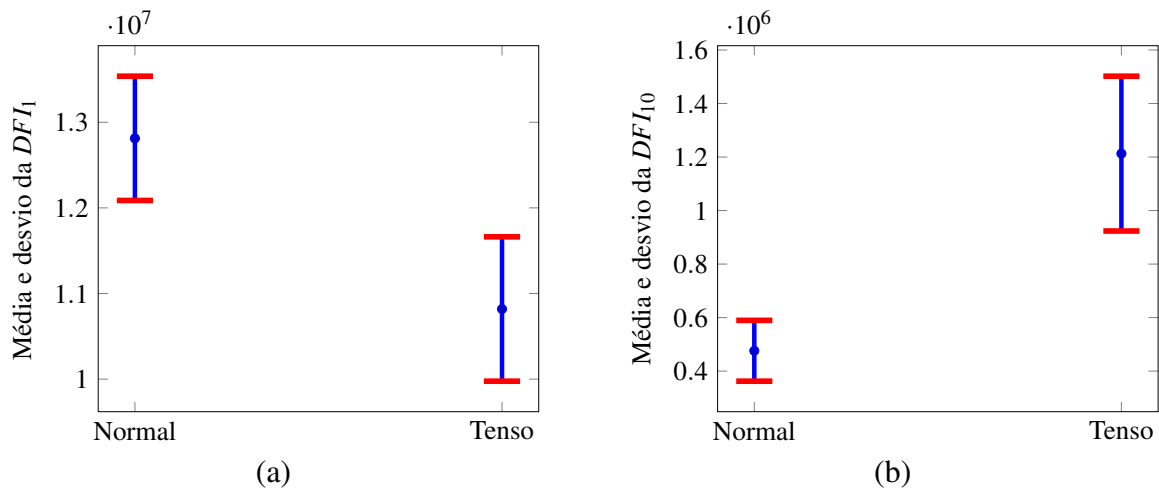
A Tabela 6.4 apresenta o *valor-p*, representando o potencial discriminativo das características de cada IMF entre os sinais de vozes normais e com desvio tensão. Do total de 50 características somente 2 características tiveram a hipótese nula rejeitada. Essas características foram a dispersão da frequência instantânea referente a IMF_1 e a dispersão da frequência instantânea referente a IMF_{10} .

De acordo com a Figura 6.4 observa-se que para as frequências mais altas como no caso da DFI_1 a variância é maior para os sinais de vozes normais, já para as frequências mais baixa como o caso da DFI_{10} a variância é maior em sinais de vozes com o desvio tensão. Isso colabora com a hipótese que foi explicada a partir da Figura 5.5, que para frequências abaixo de 1 kHz existem maiores dispersões de frequência instantânea para os sinais de voz com desvio tensão quando comparados quando comparado aos sinais de vozes normais.

Tabela 6.4 – Potencial discriminativo entre os sinais de vozes normais e com desvio tensão.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,335	0,767	0,228	0,569	0,246
$DETI_i$	0,343	0,732	0,358	0,490	0,306
APT_i	0,335	0,755	0,205	0,665	0,239
DFI_i	0,031	0,549	0,790	0,654	0,375
APF_i	0,200	0,933	0,265	0,779	0,099
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,426	0,184	0,298	0,897	1,000
$DETI_i$	0,258	0,375	0,417	0,861	0,873
APT_i	0,313	0,200	0,306	0,921	0,945
DFI_i	0,685	0,135	0,710	0,579	0,023
APF_i	0,216	0,471	0,549	0,699	0,767

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

Figura 6.4 – Médias e desvios estatísticos das características dos sinais de vozes normais e com desvio tensão: (a) características DFI_1 ; (b) característica DFI_{10} .

Fonte: Dados da pesquisa, 2018.

A Tabela 6.5 apresenta através do *valor-p* os resultados do potencial discriminativo de cada IMF entre os sinais de vozes com desvio rugosidade e com desvio soprosideade. Do total de 50 características, 18 características tiveram a hipótese nula rejeitada. Observa-se que dentre as características com hipótese nula rejeitada 15 características estão entre as IMF₁, IMF₂, IMF₅ e IMF₆.

Tabela 6.5 – Potencial discriminativo entre os sinais de vozes com desvio rugosidade e com desvio soproidade.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,014	0,005	0,065	0,260	0,001
$DETI_i$	0,009	0,002	0,048	0,058	0,020
APT_i	0,051	0,006	0,070	0,538	0,002
DFI_i	0,046	0,879	0,841	0,014	0,528
APF_i	0,719	0,031	0,260	0,904	0,007
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,013	0,127	0,581	0,274	0,767
$DETI_i$	0,016	0,615	0,098	0,420	0,707
APT_i	0,029	0,080	0,603	0,260	0,672
DFI_i	0,035	0,016	0,755	0,177	0,411
APF_i	0,153	0,477	0,334	0,210	0,267

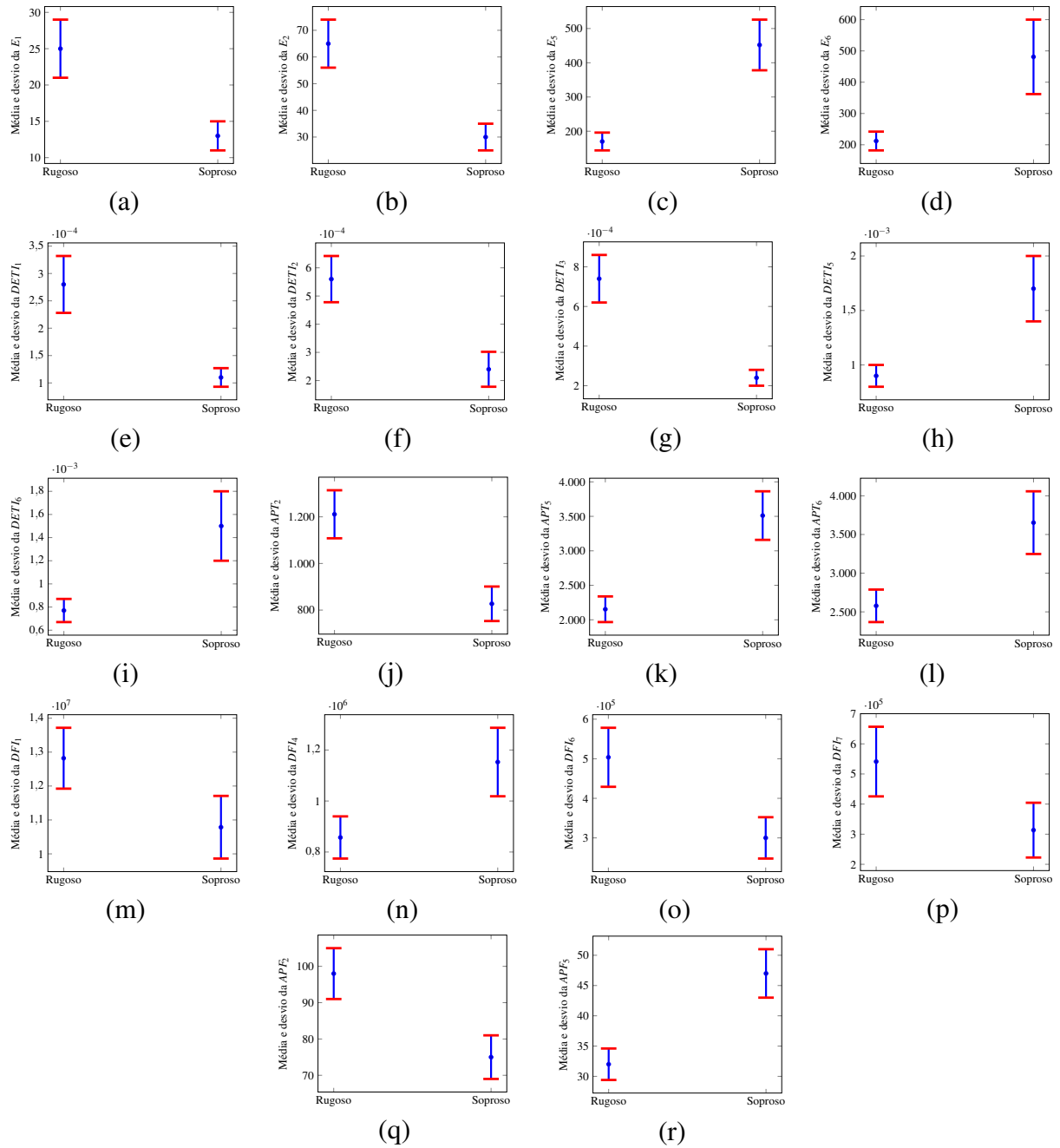
E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

A Figura 6.5 ilustra a média e desvios das características que foram selecionadas. Com isso, é possível identificar que para as frequências relacionadas a IMF₁, IMF₂ e IMF₃ os sinais de vozes com rugosidade possuem valores maiores em todas as características que foram selecionadas. Para as frequências relacionadas a IMF₄, IMF₅ e IMF₆ os sinais de vozes com soproidade possuem valores maiores em todas as características selecionadas, com exceção das características DFI_6 e DFI_7 que tiveram os seus valores maiores para os sinais de vozes com desvio rugosidade.

Inicialmente, se esperava que os sinais de vozes com desvio soproidade tivessem em média valores maiores para as altas frequências com a característica dispersão da frequência instantânea, mas o resultado foi contrário, os maiores valores ficaram com os sinais de vozes com desvio rugosidade. A explicação para isso são as propriedades do método EMD, a decomposição depende exclusivamente do sinal e as IMFs são baseadas nas propriedades locais do sinal. Então o que acontece na prática é que para uma mesma IMF os sinais de vozes com desvio rugosidade estão em uma faixa de frequência e os sinais de vozes com desvio soproidade estão em outra faixa de frequência.

Com isso é possível entender que para as altas frequências o sinal de voz com desvio soproidade possui ruídos acima de 4 kHz e sua variância é menor do que o sinal de voz com desvio rugosidade. Já o sinal de voz com desvio rugosidade possui componentes de energia abaixo de 4 kHz e sua variância é maior para a mesma IMF quando comparado ao sinal de voz com desvio rugosidade.

Figura 6.5 – Médias e desvios estatísticos das características dos sinais de vozes com desvio rugosidade e sopro: (a) característica E_1 , (b) característica E_2 , (c) característica E_5 , (d) característica E_6 , (e) característica $DETI_1$, (f) característica $DETI_2$, (g) característica $DETI_3$, (h) característica $DETI_5$, (i) característica $DETI_6$, (j) característica APT_2 , (k) característica APT_5 , (l) característica APT_6 , (m) característica DFI_1 , (n) característica DFI_4 , (o) característica DFI_6 , (p) característica DFI_7 , (q) característica APF_2 , (r) característica APF_5 .



Fonte: Dados da pesquisa, 2018.

A Tabela 6.6 apresenta o *valor-p*, representando o potencial discriminativo das características de cada IMF entre os sinais de vozes com desvio rugosidade e com desvio tensão. Do total de 50 características somente 3 características tiveram a hipótese nula rejeitada. Essas características foram dispersão da frequência instantânea referente a IMF₃, dispersão da frequência instantânea referente a IMF₈ e a amplitude instantânea ponderada pela frequência instantânea referente a IMF₅.

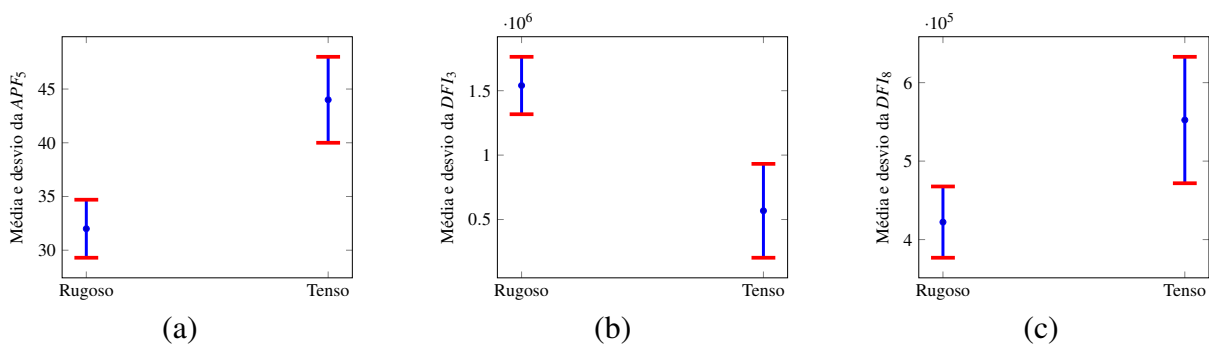
Tabela 6.6 – Potencial discriminativo entre os sinais de vozes com desvio rugosidade e com desvio tensão.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,586	0,925	1,000	0,102	0,090
$DETI_i$	0,876	0,975	0,755	0,099	0,207
APT_i	0,350	0,755	0,864	0,157	0,157
DFI_i	0,076	0,427	0,015	0,068	0,876
APF_i	0,096	0,375	0,803	0,116	0,027
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,283	0,350	0,297	0,864	0,720
$DETI_i$	0,073	0,720	0,249	0,925	0,464
APT_i	0,304	0,304	0,327	0,864	1,000
DFI_i	0,327	0,544	0,048	0,297	0,708
APF_i	0,081	0,663	0,900	0,851	0,839

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

A Figura 6.6 ilustra a média e desvios das características que foram selecionadas com isso é possível identificar que para as frequências relacionadas a IMF₃ os sinais de vozes com rugosidade possuem maiores valores de variância. Para as frequências relacionadas a IMF₈ os sinais de vozes com desvio tensão possuem maiores variâncias.

Figura 6.6 – Médias e desvios estatísticos das características dos sinais de vozes com desvio rugosidade e tensão: (a) características APF_5 ; (b) característica DFI_3 ; (c) característica DFI_8 .



Fonte: Dados da pesquisa, 2018.

A Tabela 6.7 apresenta o *valor-p*, representando o potencial discriminativo das características de cada IMF entre os sinais de vozes com desvio soproidade e com desvio tensão. Do total de 50 características, 10 características tiveram a hipótese nula rejeitada. Observa-se que dentre as características com hipótese nula rejeitada 9 características se concentram entre as IMF₂, IMF₃ e IMF₇.

Tabela 6.7 – Potencial discriminativo entre os sinais de vozes com desvio soproidade e com desvio tensão.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,112	0,003	0,023	0,649	0,253
$DETI_i$	0,013	0,001	0,010	0,968	0,581
APT_i	0,438	0,016	0,056	0,393	0,295
DFI_i	0,879	0,538	0,024	0,866	0,695
APF_i	0,411	0,204	0,342	0,067	0,570
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,310	0,030	0,660	0,594	0,615
$DETI_i$	0,497	0,402	0,477	0,517	0,438
APT_i	0,326	0,017	0,637	0,448	0,603
DFI_i	0,528	0,010	0,098	0,779	0,517
APF_i	0,917	0,779	0,507	0,326	0,267

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

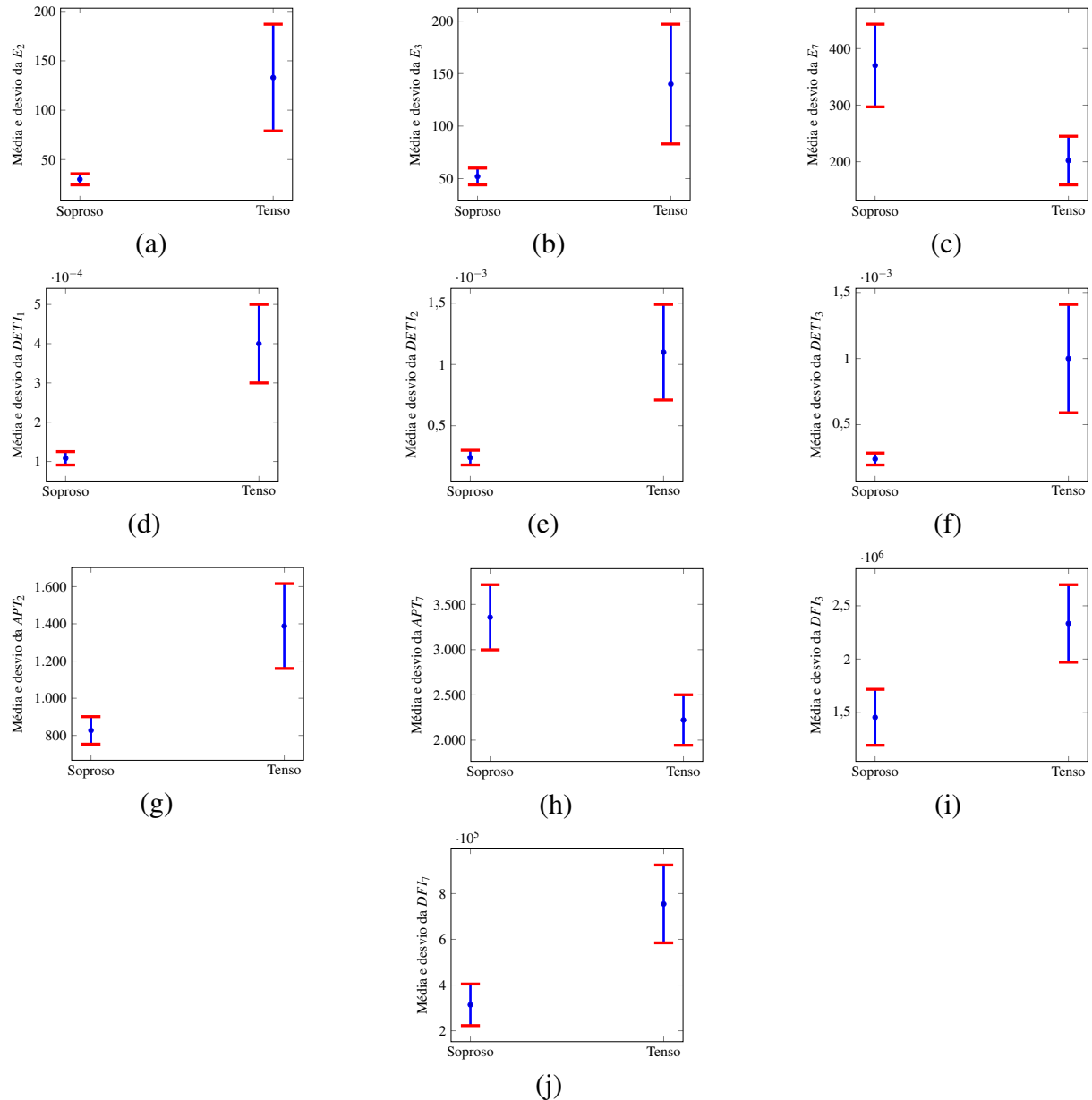
A Figura 6.7 ilustra a média e desvios estatísticos das características que foram selecionadas. Com isso, é possível identificar que para as frequências relacionadas a IMF₁, IMF₂ e IMF₃ os sinais de vozes com desvio tensão possuem valores maiores em todas as características que foram selecionadas.

Para as frequências relacionadas as IMF₃ e IMF₇ os sinais de vozes com desvio tensão possuem valores maiores para as características DFI_3 e DFI_7 . Observando a Figura 5.4, é possível ter um melhor entendimento em identificar que a dispersão da frequência instantânea é maior no sinal de voz com o desvio tensão quando comparado ao sinal com o desvio soproidade e essa dispersão se espalha para as demais IMFs.

Com isso é possível entender o espalhamento das componentes de energia pelas frequências instantâneas através do espectro HSA, pois se espera que essas componentes de energia se espalhem através pelas frequências das IMFs.

Também é possível observar que para as altas frequências todas as características que foram selecionadas para classificar os sinais possuem valores maiores para os sinais de voz com desvio tensão quando comparados aos sinais de voz com desvio soproidade.

Figura 6.7 – Médias e desvios estatísticos das características dos sinais de vozes com desvio soprosidade e tensão: (a) característica E_2 , (b) característica E_3 , (c) característica E_7 , (d) característica $DETI_1$, (e) característica $DETI_2$, (f) característica $DETI_3$, (g) característica APT_2 , (h) característica APT_7 , (i) característica DFI_3 , (j) característica DFI_7 .



Fonte: Dados da pesquisa, 2018.

6.1.2 Análise Estatística para o Estudo de Caso 2

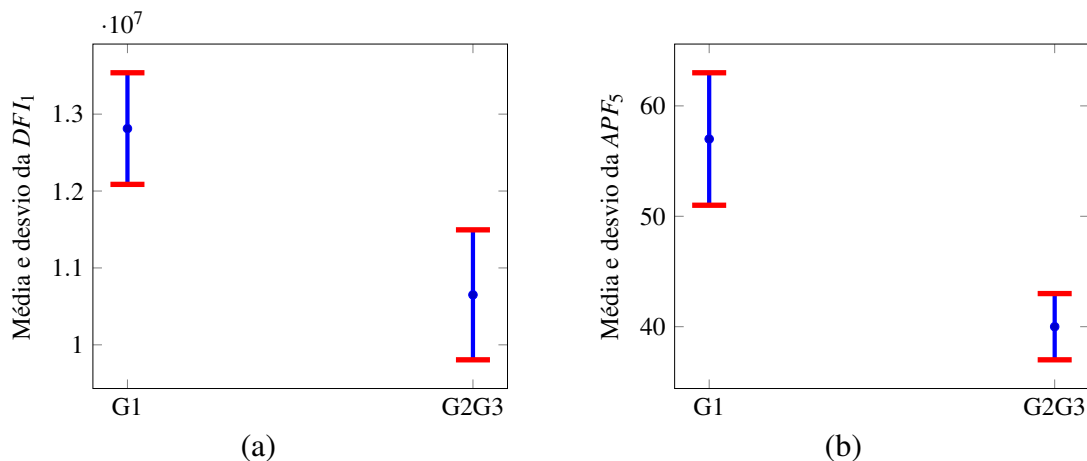
A Tabela 6.8 apresenta através do *valor-p* os resultados do potencial discriminativo de cada IMF entre os sinais de vozes do grau geral 1 e uma combinação do graus 2 e 3. Do total de 50 características somente 2 características tiveram a hipótese nula rejeitada. Essas características foram dispersão da frequência instantânea referente a IMF₁ e a amplitude instantânea ponderada pela frequência instantânea referente a IMF₅.

Tabela 6.8 – Potencial discriminativo entre os sinais de vozes do grau geral 1 e uma combinação do graus 2 e 3.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,395	0,166	0,180	0,728	0,145
$DETI_i$	0,946	0,102	0,195	0,652	0,162
APT_i	0,133	0,277	0,171	0,853	0,133
DFI_i	0,015	0,072	0,099	0,228	0,133
APF_i	0,057	0,630	0,162	0,579	0,029
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,162	0,239	0,206	0,290	0,304
$DETI_i$	0,077	0,363	0,206	0,180	0,277
APT_i	0,137	0,318	0,228	0,379	0,355
DFI_i	0,549	0,297	0,270	0,662	0,195
APF_i	0,096	0,340	0,176	0,283	0,510

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

De acordo com a Figura 6.8, observa-se que para as altas frequências como no caso da característica DFI_1 a dispersão da frequência instantânea é maior para os sinais de vozes com grau geral 1. Já para as baixas frequências como o caso da característica APF_5 o valor do subconjunto energia-tempo do HSA é maior em sinais de vozes combinadas em graus 2 e 3.

Figura 6.8 – Médias e desvios das características dos sinais de vozes com intensidades de Grau 1 e Graus 2 e 3: (a) características DFI_1 ; (b) característica APF_5 .

Fonte: Dados da pesquisa, 2018.

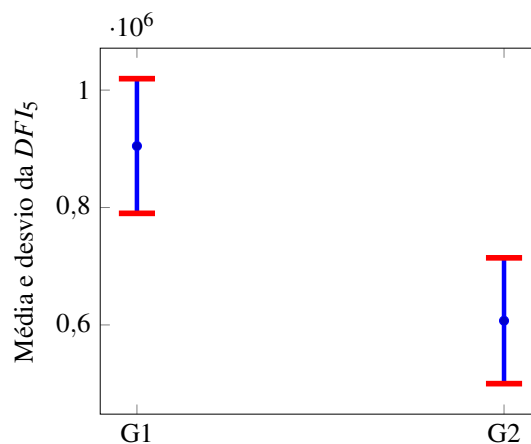
A Tabela 6.9 apresenta através do *valor-p* os resultados do potencial discriminativo de cada IMF entre os sinais de vozes do grau geral 1 e grau geral 2. Do total de 50 características apenas 1 característica teve a hipótese nula rejeitada. Essa característica foi dispersão da frequência instantânea referente a IMF₅.

Tabela 6.9 – Potencial discriminativo entre os sinais de vozes do grau geral 1 e grau geral 2.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,579	0,171	0,065	0,549	0,239
$DETI_i$	0,728	0,099	0,074	0,589	0,206
APT_i	0,491	0,270	0,090	0,549	0,258
DFI_i	0,559	0,411	0,501	0,437	0,037
APF_i	0,355	0,412	0,055	0,761	0,074
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,888	0,217	0,464	0,728	0,706
$DETI_i$	0,111	0,270	0,549	0,589	0,673
APT_i	0,876	0,277	0,491	0,899	0,830
DFI_i	0,994	0,446	0,129	0,264	0,125
APF_i	0,482	0,228	0,195	0,264	0,520

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

Nota-se que como os graus de intensidade são próximos um do outro, a quantidade de características diminuiu e o desvio estatístico aumentou quando comparado a classificação dos sinais de vozes com desvios. A Figura 6.9 demonstra as médias e desvios da característica DFI_5 . Com isso, é possível observar que, para esta característica, a dispersão da frequência instantânea referente a IMF₅ dos sinais de vozes com grau geral 1 são maiores do que em sinais com grau geral 2.

Figura 6.9 – Médias e desvios da característica DFI_5 dos sinais de vozes com grau geral 1 e grau geral 2.

Fonte: Dados da pesquisa, 2018.

A Tabela 6.10 apresenta através do *valor-p* os resultados do potencial discriminativo de cada IMF entre os sinais de vozes do grau geral 1 e grau geral 3. Do total de 50 características, 6 características tiveram a hipótese nula rejeitada. Observa-se que dentre as medidas com hipótese nula rejeitada 5 características estão entre as IMF₁ e IMF₆.

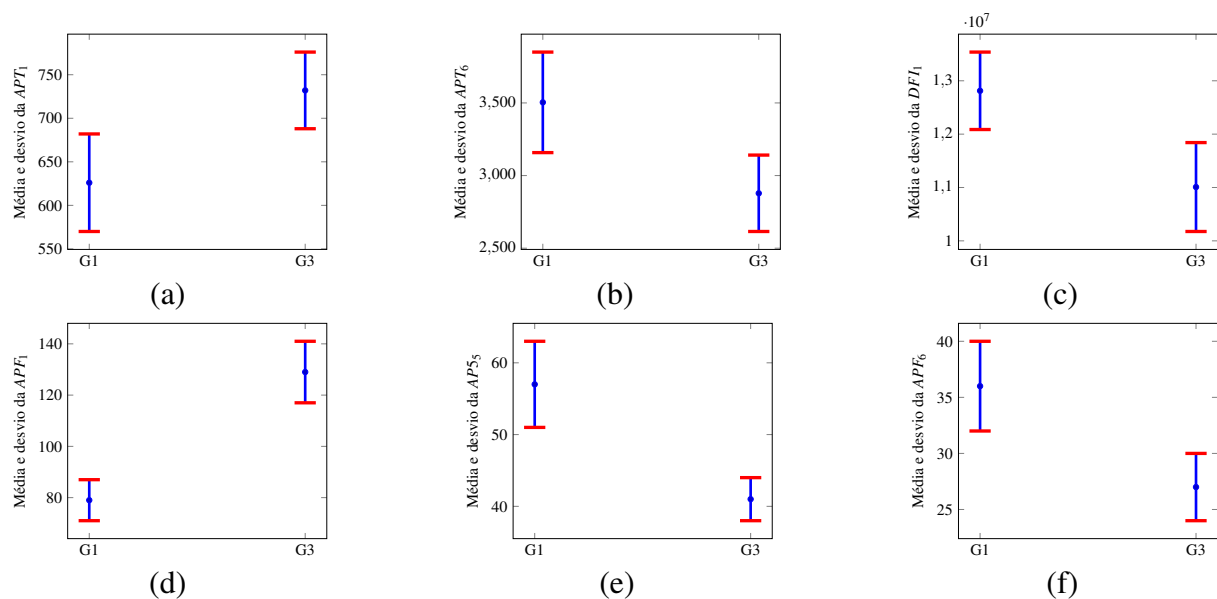
Tabela 6.10 – Potencial discriminativo entre os sinais de vozes do grau geral 1 e grau geral 3.

Característica	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
E_i	0,087	0,395	0,149	0,395	0,102
$DETI_i$	0,363	0,245	0,157	0,750	0,166
APT_i	0,011	0,529	0,171	0,340	0,074
DFI_i	0,018	0,153	0,096	0,105	0,549
APF_i	0,001	0,761	0,332	0,108	0,029
Característica	IMF ₆	IMF ₇	IMF ₈	IMF ₉	IMF ₁₀
E_i	0,065	0,888	0,946	0,641	0,784
$DETI_i$	0,251	0,994	0,630	0,473	0,630
APT_i	0,045	0,784	0,982	0,784	0,876
DFI_i	0,549	0,841	0,420	0,982	0,145
APF_i	0,037	0,761	0,684	0,501	0,982

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

A Figura 6.10 ilustra a média e desvios estatísticos das características que foram selecionadas, com isso é possível identificar que, para as frequências relacionadas a IMF₅ e IMF₆, os sinais de vozes com grau geral 1 possuem valores maiores em todas as características que foram selecionadas. Para as frequências relacionadas a IMF₁, os sinais de vozes com grau geral 2 possuem valores maiores em todas as características selecionadas, com exceção da características DFI_1 que tiveram os seus valores maiores para os sinais de vozes com grau geral 1.

Figura 6.10 – Médias e desvios das características dos sinais de vozes G1 e G3: (a) característica APT_1 ; (b) característica APT_6 , (c) característica DFI_1 , (d) característica APF_1 , (e) característica APF_5 , (f) característica APF_6 .



Fonte: Dados da pesquisa, 2018.

A Tabela 6.11 apresenta através do *valor-p* os resultados do potencial discriminativo de cada IMF entre os sinais de vozes do grau geral 2 e grau geral 3. Do total de 50 características apenas 1 característica teve a hipótese nula rejeitada. Essa característica foi amplitude instantânea ponderada pela frequência instantânea referente a IMF_1 .

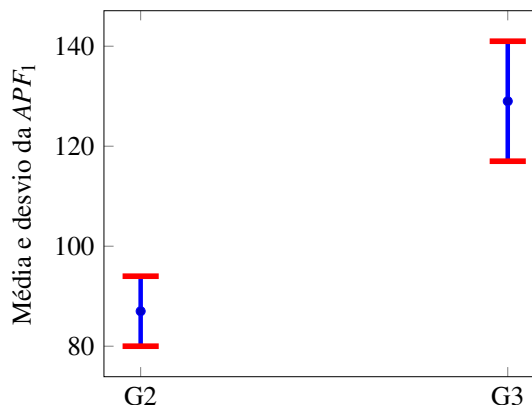
Tabela 6.11 – Potencial discriminativo entre os sinais de vozes do grau geral 2 e grau geral 3.

Característica	IMF_1	IMF_2	IMF_3	IMF_4	IMF_5
E_i	0,340	0,684	0,589	0,264	0,599
$DETI_i$	0,652	0,761	0,652	0,311	0,853
APT_i	0,200	0,529	0,549	0,211	0,455
DFI_i	0,093	0,437	0,180	0,347	0,090
APF_i	0,003	0,102	0,217	0,122	0,428
Característica	IMF_6	IMF_7	IMF_8	IMF_9	IMF_{10}
E_i	0,082	0,099	0,446	0,935	0,784
$DETI_i$	0,784	0,222	0,795	0,970	0,706
APT_i	0,082	0,115	0,446	0,911	0,652
DFI_i	0,510	0,501	0,501	0,239	0,761
APF_i	0,195	0,311	0,304	0,539	0,539

E : Energia; $DETI$: Dispersão da Densidade de Energia Temporal Instantânea; APT : Amplitude Instantânea ponderada pelo Tempo; DFI : Dispersão da Frequência Instantânea; APF : Amplitude Instantânea ponderada pela Frequência Instantânea.

A Figura 6.11 demonstra as médias e desvios da característica APF_1 . Com isso, é possível observar que para esta característica que os valores para os sinais de vozes com grau geral 3 são maiores quem em sinais com grau geral 2.

Figura 6.11 – Médias e desvios da característica APF_1 dos sinais de vozes com grau geral 2 e grau geral 3.



Fonte: Dados da pesquisa, 2018.

6.1.3 Resumo da Análise Estatística

Na Tabela 6.12 está apresentado o resumo das características obtidas para cada caso de classificação a partir dos testes estatísticos realizados, visando identificar o potencial discrimina-

tivo de cada uma das medidas temporais e espectrais avaliadas neste estudo, na caracterização de sinais de vozes com e sem desvios. Foram verificadas diferenças estatísticas significativas entre os grupos analisados, considerando-se um nível de significância igual a 5%.

A dispersão da frequência instantânea (*DFI*) e a amplitude instantânea ponderada pela frequência instantânea (*APF*) são as características que apresentaram potencial discriminativo para a maior parte dos casos analisados. A maior parte das características consideradas sugere diferenças estatisticamente significativas entre sinais de vozes com desvios rugosidade e sopro-sidade, sinais de vozes com desvios sopro-sidade e tensão, sinais de vozes normais e sinais de vozes com desvio sopro-sidade, sinais de vozes normais e sinais de vozes com desvio rugosidades. Entretanto quando se compara sinais de vozes normais e sinais de vozes com desvio tensão, sinais de vozes com desvios rugosidade e tensão, sinais de vozes normais e sinais de vozes com desvios (rugosidade, sopro-sidade e tensão), sinais de vozes com grau geral 1 e sinais de vozes com grau geral 2, sinais de vozes com grau geral 2 e sinais de vozes com grau geral 3, sinais de vozes com grau geral 1 e sinais de vozes com graus 2 e 3, há dificuldade na identificação de características que apresentem potencial discriminativo, pois para essas classificações apresentaram poucas características com os testes estatísticos aplicados.

Tabela 6.12 – Resumo do potencial discriminativo das características.

Classificação	Características
NORM × DESV	$E_5 APF_5$
NORM × SOP	$E_2 E_3 DETI_2 DETI_3 APT_2 APT_3 DFI_3 DFI_{10} APF_3$
NORM × RUG	$E_5 DETI_5 DETI_6 APT_5 APT_6 DFI_{10} APF_1 APF_5 APF_6$
NORM × TEN	$DFI_1 DFI_{10}$
RUG × SOP	$E_1 E_2 E_5 E_6 DETI_1 DETI_2 DETI_3 DETI_5 DETI_6 APT_2 APT_5 APT_6 DFI_1 DFI_4 DFI_6 DFI_7 APF_2 APF_5$
RUG × TEN	$APF_5 DFI_3 DFI_8$
SOP × TEN	$E_2 E_3 E_7 DETI_1 DETI_2 DETI_3 APT_2 APT_7 DFI_3 DFI_7$
G1 × G2	DFI_5
G1 × G3	$APT_1 APT_6 DFI_1 APF_1 APF_5 APF_6$
G2 × G3	APF_1
G1 × G2G3	$DFI_1 APF_5$

NORM × DESV: classificação entre os sinais de vozes normais e com desvios; NORM × SOP: classificação entre os sinais de vozes normais e com desvio sopro-sidade; NORM × RUG: classificação entre os sinais de vozes normais e com desvio rugosidade; NORM × TEN: classificação entre os sinais de vozes normais e com desvio tensão; RUG × SOP: classificação entre os sinais de vozes com desvio rugosidade e sopro-sidade; RUG × TEN: classificação entre os sinais de vozes com desvio rugosidade e tensão; SOP × TEN: classificação entre os sinais de vozes com desvio sopro-sidade e tensão; G1 × G2: classificação entre os sinais de vozes com grau geral 1 e grau geral 2; G1 × G3: classificação entre os sinais de vozes com grau geral 1 e grau geral 3; G2 × G3: classificação entre os sinais de vozes com grau geral 2 e grau geral 3; G1 × G2G3: classificação entre os sinais de vozes com grau 1 e graus 2 e 3.

6.2 Classificação

Nesta Seção estão apresentados os resultados referentes à classificação dos dois estudos de caso. A metodologia empregada no processo de classificação está descrita no Capítulo 5.

6.2.1 Classificação para o Estudo de Caso 1

Classificação entre sinais de vozes normais e com desvios (rugosidade, soprosidade e tensão)

Para realizar a classificação dos sinais de vozes normais e sinais com desvios (rugosidade, soprosidade e tensão) as características foram analisadas individualmente e combinadas. Conforme Tabela 6.1 apenas 2 características foram selecionadas. A Tabela 6.13 apresenta os resultados da classificação das características individuais e combinadas.

Tabela 6.13 – Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes normais e com desvios (rugosidade, soprosidade e tensão).

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
E_5	Individual	84,52±1,19	100,00±0,00	81,67±1,67
APF_5	Individual	92,67±4,52	90,00±10,00	96,00±4,00
E_5+APF_5	Todas	91,67±4,81	100,00±0,00	90,00±5,77

E_5 : Energia referente a IMF₅; APF_5 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₅; ACU: Acurácia; SENS: Sensibilidade; ESP: Especificidade.

A melhor acurácia da classificação individual foi de 92,67% obtida pela característica APF_5 , indicando que APF_5 é a principal característica para a classificação considera. Já APF_5 combinada com E_5 obteve acurácia de 91,67%, as características combinadas obtiveram resultados de acurácia de valor menor que a característica APF_5 classificada individualmente.

Classificação entre sinais de vozes normais e sinais de vozes com desvio soprosidade

Para realizar a classificação dos sinais de vozes normais e com desvio soprosidade as características foram analisadas individualmente e combinadas 2 a 2, 3 a 3, 4 a 4, 5 a 5, 6 a 6, 7 a 7, 8 a 8 e todas as características.

Na Tabela 6.14 estão apresentados os resultados com características individuais da classificação entre sinais de vozes saudáveis e com desvio soprosidade. A E_2 foi a característica que obteve acurácia de 90,83%, indicando a melhor característica individual. Considerando os desvios estatísticos das características todas obtiveram acurácia acima de 90,00%, com exceção da característica DFI_3 que obteve 74,07% de acurácia.

Na Tabela 6.15 estão apresentados os melhores resultados de características combinadas entre os sinais de vozes saudáveis e sinais de vozes com desvio soprosidade. O destaque desses resultados são as combinações formadas por características 3 a 3, 4 a 4, 5 a 5, 6 a 6 e 7 a 7 que alcançaram 100,00% de acurácia. Dentre as características se destaca a energia referente a IMF₂, pois está em todas as combinações que atingiram 100,00% de acurácia. As características combinadas obtiveram um aumento de acurácia de aproximadamente 10,00% quando comparadas às características individuais.

Tabela 6.14 – Melhor desempenho das características individuais na classificação entre sinais de vozes saudáveis e sinais de vozes com desvio soproidade.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
E_2	Individual	90,83±4,72	100,00±0,00	86,67±6,94
E_3	Individual	87,50±5,16	95,00±5,00	86,67±6,94
$DETI_2$	Individual	87,04±5,22	100,00±0,00	81,48±7,58
$DETI_3$	Individual	86,46±9,70	87,50±12,50	85,42±9,68
APT_2	Individual	88,10±5,71	92,86±7,14	90,48±6,15
APT_3	Individual	90,48±6,15	85,71±9,22	100,00±0,00
DFI_3	Individual	74,07±10,80	77,78±12,11	77,78±12,11
DFI_{10}	Individual	90,28±6,24	94,44±5,56	91,67±8,33
APF_3	Individual	84,72±6,94	91,67±8,33	86,11±9,04

E_2 : Energia referente a IMF₂; E_3 : Energia referente a IMF₃; $DETI_2$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₂; $DETI_3$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₃; APT_2 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₂; APT_3 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₃; DFI_3 : Dispersão da Frequência Instantânea referente a IMF₃; DFI_{10} : Dispersão da Frequência Instantânea referente a IMF₁₀; APF_3 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₃; ACU: Acurácia; SENS: Sensibilidade; ESP: Especificidade.

Tabela 6.15 – Melhores desempenhos de características combinadas na classificação entre sinais de vozes normais e com desvio soproidade.

Características	Combinações	ACU (%)	SENS (%)	ESP (%)
$E_2 + DETI_2$	2 a 2	95,24±4,76	92,86±7,14	100,00±0,00
$E_2 + DETI_2 + APT_2$	3 a 3	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + DETI_2 + DETI_3 + APF_3$	4 a 4	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_3 + DETI_3 + APT_3 + DFI_3$	5 a 5	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_3 + DETI_2 + DETI_3 + APT_2 + APT_3$	6 a 6	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_3 + DETI_2 + DETI_3 + APT_2 + APT_3 + DFI_{10}$	7 a 7	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_3 + DETI_2 + DETI_3 + APT_2 + APT_3 + DFI_{10} + APF_3$	8 a 8	94,44±3,67	100,00±0,00	92,59±4,90
$E_2 + E_3 + DETI_2 + DETI_3 + APT_2 + APT_3 + DFI_3 + DFI_{10} + APF_3$	Todas	91,67±5,46	88,10±7,90	100,00±0,00

E_2 : Energia referente a IMF₂; E_3 : Energia referente a IMF₃; $DETI_2$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₂; $DETI_3$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₃; APT_2 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₂; APT_3 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₃; DFI_3 : Dispersão da Frequência Instantânea referente a IMF₃; DFI_{10} : Dispersão da Frequência Instantânea referente a IMF₁₀; APF_3 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₃; ACU: Acurácia; SENS: Sensibilidade; ESP: Especificidade.

Classificação entre sinais de vozes normais e sinais de vozes com desvio rugosidade

Para realizar a classificação dos sinais de vozes normais e com desvio rugosidade as características foram analisadas individualmente e combinadas 2 a 2, 3 a 3, 4 a 4, 5 a 5, 6 a 6, 7 a 7, 8 a 8, 9 a 9 e todas as características. Na Tabela 6.16 estão apresentados os resultados com características individuais da classificação entre sinais de vozes saudáveis e sinais de vozes com desvio rugosidade. A APF_5 foi a característica que obteve acurácia de 93,05%, indicando a melhor característica individual. Considerando os desvios estatísticos das características todas podem obter acurácia acima de 85,00%, com exceção de APT_6 que obteve 76,04% de acurácia.

Tabela 6.16 – Melhor desempenho das características individuais na classificação entre sinais de vozes normais e sinais de vozes com desvio rugosidade.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
E_5	Individual	86,90±6,26	92,85±7,14	88,09±7,89
$DETI_5$	Individual	82,40±5,63	94,44±5,56	79,63±8,22
$DETI_6$	Individual	81,67±7,63	100,00±0,00	73,33±11,30
APT_5	Individual	88,63±6,57	90,91±6,09	87,87±9,29
APT_6	Individual	76,04±5,98	83,33±8,33	79,16±8,18
DFI_4	Individual	80,20±5,87	93,75±6,25	77,08±8,87
DFI_{10}	Individual	80,55±6,51	79,63±8,22	88,88±7,35
APF_1	Individual	86,90±6,26	88,09±7,89	92,85±7,14
APF_5	Individual	93,05±3,67	100,00±0,00	90,27±5,21
APF_6	Individual	88,88±7,02	100,00±0,00	83,33±10,54

E_5 : Energia referente a IMF₅; $DETI_5$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₅; $DETI_6$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₆; APT_5 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₅; APT_6 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₆; DFI_4 : Dispersão da Frequência Instantânea referente a IMF₄; DFI_{10} : Dispersão da Frequência Instantânea referente a IMF₁₀; APF_1 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₁; APF_5 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₅; APF_6 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₆; *ACU*: Acurácia; *SENS*: Sensibilidade; *ESP*: Especificidade.

Na Tabela 6.17 estão apresentados os melhores resultados de características combinadas entre os sinais de vozes saudáveis e sinais de vozes com desvio rugosidade. O destaque desses resultados são as combinações formadas por características 3 a 3, 4 a 4, 5 a 5, 7 a 7 e 8 a 8 que alcançaram 100,00% de acurácia. Dentre as características se destaca a energia referente a IMF₅ e a amplitude instantânea ponderada pela frequência instantânea referente a IMF₅, pois estão em todas as combinações que atingiram 100,00% de acurácia. As características combinadas obtiveram um aumento de acurácia de aproximadamente 7,00% quando comparadas com as características individuais.

Tabela 6.17 – Melhores desempenhos de características combinadas na classificação entre sinais de vozes normais e com desvio rugosidade.

Características	Combinações	ACU (%)	SENS (%)	ESP (%)
$DETI_5 + DFI_{10}$	2 a 2	96,43±3,57	100,00±0,00	95,24±4,76
$E_5 + DETI_5 + APF_5$	3 a 3	100,00±0,00	100,00±0,00	100,00±0,00
$E_5 + APT_6 + DFI_4 + APF_5$	4 a 4	100,00±0,00	100,00±0,00	100,00±0,00
$E_5 + DETI_5 + APT_5 + APT_6 + APF_5$	5 a 5	100,00±0,00	100,00±0,00	100,00±0,00
$E_5 + DETI_5 + DETI_6 + APT_6 + APF_1 + APF_5$	6 a 6	96,43±3,57	95,24±4,76	100,00±0,00
$E_5 + DETI_6 + APT_5 + APT_6 + DFI_{10} + APF_1 + APF_5$	7 a 7	100,00±0,00	100,00±0,00	100,00±0,00
$E_5 + DETI_5 + DETI_6 + APT_6 + DFI_4 + DFI_{10} + APF_5 + APF_6$	8 a 8	100,00±0,00	100,00±0,00	100,00±0,00
$E_5 + DETI_6 + APT_5 + APT_6 + DFI_4 + DFI_{10} + APF_1 + APF_5 + APF_6$	9 a 9	94,17±3,94	100,00±0,00	91,67±5,69
$E_5 + DETI_5 + DETI_6 + APT_5 + APT_6 + DFI_4 + DFI_{10} + APF_1 + APF_5 + APF_6$	Todas	88,64±6,58	87,88±6,39	90,91±9,09

E_5 : Energia referente a IMF₅; $DETI_5$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₅; $DETI_6$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₆; APT_5 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₅; APT_6 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₆; DFI_4 : Dispersão da Frequência Instantânea referente a IMF₄; DFI_{10} : Dispersão da Frequência Instantânea referente a IMF₁₀; APF_1 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₁; APF_5 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₅; APF_6 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₆; *ACU*: Acurácia; *SENS*: Sensibilidade; *ESP*: Especificidade.

Classificação entre sinais de vozes normais e sinais de vozes com o desvio tensão

Para realizar a classificação dos sinais de vozes normais e sinais de vozes com o desvio tensão, as características foram analisadas individualmente e combinadas. Conforme Tabela 6.4 apenas 2 características foram selecionadas. A Tabela 6.18 apresenta os resultados das características individuais e combinadas. A melhor acurácia desta classificação foi de 88,10% obtida pela característica DFI_{10} , indicando que DFI_{10} é a principal característica para a classificação considerada. Já DFI_{10} combinada com DFI_1 obteve acurácia de 84,09%, as características combinadas obtiveram resultado de acurácia de valor menor que a característica DFI_{10} individualmente.

Tabela 6.18 – Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes normais e sinais de vozes com desvio tensão.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
DFI_1	Individual	77,08±7,51	72,92±8,30	87,50±8,18
DFI_{10}	Individual	88,10±7,90	92,86±7,14	85,71±9,22
$DFI_1 + DFI_{10}$	Todas	84,09±4,69	84,85±6,53	92,42±5,20

DFI_1 : Dispersão da Frequência Instantânea referente a IMF₁; DFI_{10} : Dispersão da Frequência Instantânea referente a IMF₁₀; ACU: Acurácia; SENS: Sensibilidade; ESP: Especificidade.

Classificação entre sinais de vozes com os desvios rugosidade e tensão

Para realizar a classificação dos sinais de vozes com os desvios rugosidade e tensão as características foram analisadas individualmente e combinadas. Conforme Tabela 6.6, apenas 3 características foram selecionadas.

A Tabela 6.19 apresenta os resultados das características individuais e combinadas entre sinais de vozes com desvio rugosidade e tensão. A melhor acurácia desta classificação foi de 93,75% obtida pela característica DFI_3 , indicando que DFI_3 é a principal característica para a classificação considerada. Esta característica indica que a variância referente a IMF₃ da frequência instantânea foi uma característica determinante para diferenciar estes desvios. O melhor resultado de acurácia com características combinadas foi de 91,66%, que é um valor menor do que a acurácia de DFI_3 classificada individualmente.

Tabela 6.19 – Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes com desvios rugosidade e tensão.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
APF_5	Individual	75,00±9,21	77,78±12,11	74,07±11,82
DFI_3	Individual	93,75±6,25	91,67±8,33	100,00±0,00
DFI_8	Individual	90,48±6,15	92,86±7,14	92,86±7,14
$APF_5 + DFI_3$	2 a 2	87,03±6,67	88,88±7,34	88,88±7,34
$APF_5 + DFI_8$	2 a 2	80,20±6,48	81,25±7,34	87,50±8,18
$DFI_3 + DFI_8$	2 a 2	90,47±6,14	92,86±7,14	92,86±7,14
$APF_5 + DFI_3 + DFI_8$	Todas	91,66±5,27	88,88±7,02	100,00±0,00

APF_5 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₅; DFI_3 : Dispersão da Frequência Instantânea referente a IMF₃; DFI_8 : Dispersão da Frequência Instantânea referente a IMF₈; ACU: Acurácia; SENS: Sensibilidade; ESP: Especificidade.

Classificação entre sinais de vozes com desvio rugosidade e soproidade

A Tabela 6.20 apresenta os resultados das características que foram selecionadas individualmente entre sinais de vozes com desvio rugosidade e soproidade. A melhor acurácia desta classificação foi de 93,33% obtida pela característica $DETI_3$, indicando que $DETI_3$ é a principal característica para a classificação considera. Considerando os desvios estatísticos das características, todas podem obter acurácia acima de 86,00%, com exceção de APT_6 que obteve 75,00% de acurácia. Também observa-se que mais duas características estão acima de 90,00% que são E_5 e DFI_6 . A característica $DETI_3$ representa no espectro HSA a variância da energia instantânea no subconjunto do tempo referente a IMF_3 , a característica E_5 representa a intensidade sonora referente a IMF_5 e a característica DFI_6 representa no espectro HSA a variância da frequência instantânea. Nota-se conforme as IMFs selecionadas que foram determinantes para esta classificação as características de altas e médias frequências.

Tabela 6.20 – Melhor desempenho das características individuais na classificação entre sinais de vozes com desvio rugosidade e soproidade.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
E_1	Individual	86,67±8,16	90,00±10,00	90,00±10,00
E_2	Individual	82,50±6,97	80,00±10,18	90,00±6,67
E_5	Individual	91,67±5,46	87,50±8,18	100,00±0,00
E_6	Individual	80,00±8,16	90,00±10,00	80,00±12,25
$DETI_1$	Individual	81,25±5,62	89,58±7,00	83,33±8,33
$DETI_2$	Individual	86,11±5,56	94,44±5,56	85,19±7,58
$DETI_3$	Individual	93,33±6,67	100,00±0,00	90,00±10,00
$DETI_5$	Individual	84,52±7,51	92,86±7,14	80,95±9,22
$DETI_6$	Individual	88,89±7,03	100,00±0,00	83,33±10,54
APT_2	Individual	83,33±5,56	90,00±6,67	85,00±7,64
APT_5	Individual	86,90±6,26	80,95±9,22	100,00±0,00
APT_6	Individual	75,00±8,33	75,00±9,45	81,25±13,15
DFI_1	Individual	84,52±7,81	80,95±9,22	92,86±7,14
DFI_4	Individual	88,33±7,26	100,00±0,00	83,33±10,54
DFI_6	Individual	91,67±5,46	95,24±4,76	92,86±7,14
DFI_7	Individual	83,33±5,38	92,59±4,90	83,33±8,33
APF_2	Individual	84,72±6,94	86,11±9,04	91,67±8,33
APF_5	Individual	85,19±5,86	88,89±7,35	88,89±7,35

E_1 : Energia referente a IMF_1 ; E_2 : Energia referente a IMF_2 ; E_5 : Energia referente a IMF_5 ; E_6 : Energia referente a IMF_6 ; $DETI_1$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF_1 ; $DETI_2$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF_2 ; $DETI_3$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF_3 ; $DETI_5$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF_5 ; $DETI_6$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF_6 ; APT_2 : Amplitude Instantânea ponderada pelo Tempo referente a IMF_2 ; APT_5 : Amplitude Instantânea ponderada pelo Tempo referente a IMF_5 ; APT_6 : Amplitude Instantânea ponderada pelo Tempo referente a IMF_6 ; DFI_1 : Dispersão da Frequência Instantânea referente a IMF_1 ; DFI_4 : Dispersão da Frequência Instantânea referente a IMF_4 ; DFI_6 : Dispersão da Frequência Instantânea referente a IMF_6 ; DFI_7 : Dispersão da Frequência Instantânea referente a IMF_7 ; APF_2 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_2 ; APF_5 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_5 ; ACU : Acurácia; $SENS$: Sensibilidade; ESP : Especificidade.

Na Tabela 6.21 estão apresentados os melhores resultados de características combinadas entre os sinais de vozes saudáveis e sinais de vozes com desvio rugosidade. O destaque desses resultados são as combinações formadas por características 3 a 3 até 13 a 13 que alcançaram

100,00% de acurácia. Dentre as características avaliadas, se destacam a energia referente a IMF₂ e a dispersão da densidade de energia temporal instantânea referente a IMF₃, pois estão na maioria das combinações que atingiram 100,00% de acurácia. As características combinadas obtiveram um aumento de acurácia de aproximadamente 7,00% quando comparadas as características individuais.

Tabela 6.21 – Melhores desempenhos de características combinadas na classificação entre sinais de vozes com desvio rugosidade e soprosidade.

Características	Combinações	ACU (%)	SENS (%)	ESP (%)
$E_5 + DETI_3$	2 a 2	96,67±3,33	100,00±0,00	95,00±5,00
$E_2 + E_6 + APT_5$	3 a 3	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_6 + DETI_3 + APT_5$	4 a 4	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_5 + E_6 + DETI_3 + APT_5$	5 a 5	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_5 + DETI_3 + DETI_6 + DFI_4 + DFI_6$	6 a 6	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_5 + DETI_3 + DETI_6 + APT_5 + DFI_4 + DFI_6$	7 a 7	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_2 + E_5 + DETI_3 + DETI_6 + APT_5 + DFI_4 + DFI_6$	8 a 8	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_2 + E_5 + DETI_3 + DETI_6 + APT_5 + DFI_4 + DFI_6 + APF_5$	9 a 9	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_2 + E_5 + DETI_2 + DETI_3 + DETI_6 + APT_5 + DFI_4 + DFI_6 + APF_5$	10 a 10	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_2 + E_5 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_5 + DFI_4 + DFI_6 + APF_5$	11 a 11	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_2 + E_5 + E_6 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_5 + DFI_4 + DFI_6 + APF_5$	12 a 12	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_2 + E_5 + E_6 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_5 + DFI_4 + DFI_6 + APF_2 + APF_5$	13 a 13	100,00±0,00	100,00±0,00	100,00±0,00
$E_1 + E_2 + E_5 + E_6 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_5 + DFI_1 + DFI_4 + DFI_6 + APF_2 + APF_5$	14 a 14	98,08±1,92	100,00±0,00	97,44±2,56
$E_1 + E_2 + E_5 + E_6 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_2 + APT_5 + DFI_1 + DFI_4 + DFI_6 + APF_2 + APF_5$	15 a 15	96,15±2,60	94,87±3,47	100,00±0,00
$E_1 + E_2 + E_5 + E_6 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_2 + APT_5 + DFI_1 + DFI_4 + DFI_6 + DFI_7 + APF_2 + APF_5$	16 a 16	95,83±4,17	100,00±0,00	93,75±6,25
$E_1 + E_2 + E_5 + E_6 + DETI_1 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_2 + APT_5 + DFI_1 + DFI_4 + DFI_6 + DFI_7 + APF_2 + APF_5$	17 a 17	95,24±4,76	100,00±0,00	92,86±7,14
$E_1 + E_2 + E_5 + E_6 + DETI_1 + DETI_2 + DETI_3 + DETI_5 + DETI_6 + APT_2 + APT_5 + APT_6 + DFI_1 + DFI_4 + DFI_6 + DFI_7 + APF_2 + APF_5$	Todas	91,67±4,35	90,91±6,10	96,97±3,03

E_1 : Energia referente a IMF₁; E_2 : Energia referente a IMF₂; E_5 : Energia referente a IMF₅; E_6 : Energia referente a IMF₆; $DETI_1$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₁; $DETI_2$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₂; $DETI_3$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₃; $DETI_5$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₅; $DETI_6$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₆; APT_2 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₂; APT_5 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₅; APT_6 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₆; DFI_1 : Dispersão da Frequência Instantânea referente a IMF₁; DFI_4 : Dispersão da Frequência Instantânea referente a IMF₄; DFI_6 : Dispersão da Frequência Instantânea referente a IMF₆; DFI_7 : Dispersão da Frequência Instantânea referente a IMF₇; APF_2 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₂; APF_5 : Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF₅; ACU : Acurácia; $SENS$: Sensibilidade; ESP : Especificidade.

Classificação entre sinais de vozes com os desvios soproidade e tensão

Para realizar a classificação entre os sinais de vozes com os desvios soproidade e tensão, as características foram analisadas individualmente e combinadas 2 a 2, 3 a 3, 4 a 4, 5 a 5, 6 a 6, 7 a 7, 8 a 8, 9 a 9 e todas as características.

Na Tabela 6.22 estão apresentados os resultados com características individuais da classificação entre sinais de vozes com os desvios soproidade e tensão. A dispersão da frequência instantânea referente a IMF₇ (*DFI*₇) foi a característica que obteve acurácia de 95,00%, indicando a melhor característica individual. Essa característica indica que a variância referente a IMF₇ da frequência instantânea foi determinante para diferenciar esses desvios. Considerando os desvios estatísticos das características todas podem obter acurácia acima de 90,00%, com exceção das características *E*₇ e *APT*₇.

Tabela 6.22 – Melhor desempenho das características individuais na classificação entre sinais de vozes com os desvios soproidade e tensão.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
<i>E</i> ₂	Individual	94,44±5,56	91,67±8,33	100,00±0,00
<i>E</i> ₃	Individual	88,54±5,66	93,75±6,25	89,58±7,00
<i>E</i> ₇	Individual	75,76±8,29	80,30±10,74	74,24±10,14
<i>DETI</i> ₁	Individual	81,67±10,00	78,33±10,56	90,00±10,00
<i>DETI</i> ₂	Individual	83,33±7,45	75,00±11,18	100,00±0,00
<i>DETI</i> ₃	Individual	93,75±6,25	91,67±8,33	100,00±0,00
<i>APT</i> ₂	Individual	87,50±7,16	86,67±6,94	90,00±10,00
<i>APT</i> ₇	Individual	78,79±9,29	81,82±10,16	81,82±10,16
<i>DFI</i> ₃	Individual	91,67±8,33	87,50±12,50	100,00±0,00
<i>DFI</i> ₇	Individual	95,00±5,00	100,00±0,00	93,33±6,67

*E*₂: Energia referente a IMF₂; *E*₃: Energia referente a IMF₃; *E*₇: Energia referente a IMF₇; *DETI*₁: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₁; *DETI*₂: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₂; *DETI*₃: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₃; *APT*₂: Amplitude Instantânea ponderada pelo Tempo referente a IMF₂; *APT*₇: Amplitude Instantânea ponderada pelo Tempo referente a IMF₇; *DFI*₃: Dispersão da Frequência Instantânea referente a IMF₃; *DFI*₇: Dispersão da Frequência Instantânea referente a IMF₇; *ACU*: Acurácia; *SENS*: Sensibilidade; *ESP*: Especificidade.

Na Tabela 6.23 estão apresentados os melhores resultados de características combinadas entre os sinais de vozes com os desvios soproidade e tensão. O destaque desses resultados são as combinações formadas por características combinadas 2 a 2, 3 a 3, 4 a 4, 5 a 5, 6 a 6 e 7 a 7 que alcançaram 100,00% de acurácia. Dentre as características avaliadas, se destacam a energia referente a IMF₂ e IMF₃ e a dispersão da frequência instantânea referente a IMF₃, pois estão em quase todas as combinações que atingiram 100,00% de acurácia. As características combinadas obtiveram um aumento de acurácia de aproximadamente 5,00%, quando comparadas às características individuais.

Tabela 6.23 – Melhores desempenhos de características combinadas na classificação entre sinais de vozes com desvio soproidade e tensão.

Características	Combinações	ACU (%)	SENS (%)	ESP (%)
$E_2 + DETI_3$	2 a 2	100,00±0,00	100,00±0,00	100,00±0,00
$E_3 + DFI_3 + DFI_7$	3 a 3	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_3 + DETI_3 + DFI_3$	4 a 4	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + DETI_1 + DETI_2 + DETI_3 + DFI_3$	5 a 5	100,00±0,00	100,00±0,00	100,00±0,00
$E_3 + APT_2 + APT_7 + DETI_1 + DETI_2 + DFI_3$	6 a 6	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_3 + E_7 + DETI_1 + DETI_3 + DFI_3 + DFI_7$	7 a 7	100,00±0,00	100,00±0,00	100,00±0,00
$E_2 + E_3 + APT_2 + APT_7 + DETI_2 + DETI_3 + DFI_3 + DFI_7$	8 a 8	96,43±3,57	95,24±4,76	100,00±0,00
$E_2 + E_3 + APT_2 + APT_7 + DETI_1 + DETI_2 + DETI_3 + DFI_3 + DFI_7$	9 a 9	90,74±4,69	90,74±6,28	96,30±3,70
$E_2 + E_3 + E_7 + APT_2 + APT_7 + DETI_1 + DETI_2 + DETI_3 + DFI_3 + DFI_7$	Todas	87,87±5,07	86,36±7,04	95,45±4,55

E_2 : Energia referente a IMF₂; E_3 : Energia referente a IMF₃; E_7 : Energia referente a IMF₇; $DETI_1$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₁; $DETI_2$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₂; $DETI_3$: Dispersão da Densidade de Energia Temporal Instantânea referente a IMF₃; APT_2 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₂; APT_7 : Amplitude Instantânea ponderada pelo Tempo referente a IMF₇; DFI_3 : Dispersão da Frequência Instantânea referente a IMF₃; DFI_7 : Dispersão da Frequência Instantânea referente a IMF₇; ACU : Acurácia; $SENS$: Sensibilidade; ESP : Especificidade.

6.2.2 Classificação para o Estudo de Caso 2

Os resultados, referentes à classificação realizada no Estudo de Caso 2, que representa a classificação entre os graus dos desvios vocais são apresentados nesta Seção. A classificação, assim como no Estudo de Caso 1, foi realizada com as características que foram rejeitadas nos testes de hipótese. A metodologia aplicada neste estudo de caso está apresentada no Capítulo 5.

Classificação entre as intensidades de Grau Geral 1 e Grau Geral 2 dos desvios vocais - (G1×G2)

Para realizar a classificação dos sinais de vozes normais (Grau 1 - G1) e de grau leve (Grau 2 - G2), conforme Tabela 6.9 apenas a característica DFI_5 foi selecionada para o processo de classificação. O resultado de acurácia dessa classificação foi de 83,33 ± 10,54, a sensibilidade foi de 80,00 ± 12,25 e a especificidade foi de 90,00 ± 10,00. Para esta classificação a dispersão da frequência instantânea referente a IMF₅ representa a variância da frequência instantânea, que mostra ser uma medida representativa para realizar esta classificação.

Classificação entre as intensidades de Grau Geral 1 e Grau Geral 3 dos desvios vocais - (G1×G3)

Para realizar a classificação dos sinais de vozes normais (Grau 1 - G1) e de grau moderado (Grau 3 - G3) as características foram analisadas individualmente e combinadas 2 a 2, 3 a 3, 4 a 4, 5 a 5 e todas as características.

A Tabela 6.24 apresenta os resultados com características individuais da classificação

entre sinais de vozes normais (Grau 1 - G1) e de grau moderado (Grau 3 - G3). A APF_5 foi a característica que obteve acurácia de 92,71%, indicando a melhor característica individual.

Tabela 6.24 – Melhor desempenho das características individuais na classificação entre sinais de vozes G1 e G3.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
APT_1	Individual	82,50±6,14	90,00±6,67	81,67±7,64
APT_6	Individual	77,50±11,11	76,67±10,89	80,00±13,13
DFI_1	Individual	88,54±5,66	93,75±6,25	89,58±7,00
APF_1	Individual	84,03±6,02	91,67±5,62	80,56±7,06
APF_5	Individual	92,71±4,84	93,75±6,25	95,83±4,17
APF_6	Individual	77,27±5,96	86,36±7,04	75,76±7,23

APT_1 : Amplitude Instantânea ponderada pelo Tempo referente a IMF_1 ; APT_6 : Amplitude Instantânea ponderada pelo Tempo referente a IMF_6 ; DFI_1 : Dispersão da Frequência Instantânea referente a IMF_1 ; DFI_1 : Dispersão da Frequência Instantânea referente a IMF_7 ; APF_1 Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_1 ; APF_5 Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_5 ; APF_6 Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_6 ; ACU : Acurácia; $SENS$: Sensibilidade; ESP : Especificidade.

Na Tabela 6.25 estão apresentados os melhores resultados de características combinadas entre os sinais de vozes com grau geral 1 e de vozes com grau geral 3. Todos os resultados combinados desta classificação obtiveram acurácia acima de 90%, com destaque para combinação formadas por características combinadas 4 a 4 que obteve 98,44% de acurácia. As características combinadas obtiveram um aumento de acurácia de aproximadamente 5,00% quando comparadas as características individuais.

Tabela 6.25 – Melhores desempenhos de características combinadas na classificação entre sinais de vozes G1 e G3.

Características	Combinações	ACU (%)	SENS (%)	ESP (%)
$APF_1 + APF_5$	2 a 2	97,50±2,50	100,00±0,00	96,67±3,33
$APT_1 + APF_1 + APF_5$	3 a 3	98,33±1,67	97,78±2,22	100,00±0,00
$APT_1 + APT_6 + APF_1 + APF_5$	4 a 4	98,44±1,56	100,00±0,00	97,92±2,08
$APT_1 + APT_6 + APF_1 + APF_5 + APF_6$	5 a 5	95,00±5,00	95,56±4,44	93,33±6,67
$APT_1 + APT_6 + DFI_1 + APF_1 + APF_5 + APF_6$	Todas	93,94±4,06	95,45±4,55	95,45±4,55

APT_1 : Amplitude Instantânea ponderada pelo Tempo referente a IMF_1 ; APT_6 : Amplitude Instantânea ponderada pelo Tempo referente a IMF_6 ; DFI_1 : Dispersão da Frequência Instantânea referente a IMF_1 ; DFI_1 : Dispersão da Frequência Instantânea referente a IMF_7 ; APF_1 Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_1 ; APF_5 Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_5 ; APF_6 Amplitude Instantânea ponderada pela Frequência Instantânea referente a IMF_6 ; ACU : Acurácia; $SENS$: Sensibilidade; ESP : Especificidade.

Classificação entre as intensidades de Grau Geral 2 e Grau Geral 3 dos desvios vocais - (G2×G3)

Para realizar a classificação dos sinais de vozes de grau leve (Grau 2 - G2) e de grau moderado (Grau 3 - G3), conforme Tabela 6.11 apenas a característica APF_1 foi selecionada para o processo de classificação. O resultado de acurácia desta classificação foi de 87,12% ± 4,55%, a sensibilidade foi de 95,45% ± 4,55% e a especificidade foi de 86,36% ± 5,87%. Para esta classificação, a amplitude instantânea ponderada pela frequência instantânea referente a IMF_1 representa um valor marginal energia-frequência do HSA, que mostra ser uma medida representativa para realizar esta classificação.

Classificação entre as intensidades de Grau 1 e Graus 2 e 3 dos desvios vocais - (G1×G2G3)

Para realizar a classificação dos sinais de vozes grau 1 e de vozes com graus 2 e 3 foram analisadas individualmente e combinadas. Conforme Tabela 6.8 apenas 2 características foram selecionadas. A Tabela 6.26 apresenta os resultados das características individuais e combinadas. A melhor acurácia desta classificação foi de 87,96% obtida pela característica APF_5 , indicando que APF_5 é a principal característica para a classificação considerada. Já DFI_1 combinada com APF_5 obteve acurácia de 87,50%. As características combinadas obtiveram resultado de acurácia de valor praticamente igual a característica APF_5 individualmente.

Tabela 6.26 – Melhor desempenho das características individuais e combinadas na classificação entre sinais de vozes com intensidades de Grau 1 e Graus 2 e 3.

Característica	Combinação	ACU (%)	SENS (%)	ESP (%)
DFI_1	Individual	83,33±6,03	88,10±7,90	88,10±7,90
APF_5	Individual	87,96±4,83	100,00±0,00	83,33±6,80
$DFI_1 + APF_5$	Todas	87,50±5,59	83,33±7,45	100,00±0,00

DFI_1 : Dispersão da Frequência Instantânea referente a IMF_1 ; DFI_{10} : Dispersão da Frequência Instantânea referente a IMF_{10} ; ACU : Acurácia; $SENS$: Sensibilidade; ESP : Especificidade.

6.2.3 Resumo da Classificação

Na Tabela 6.27 são apresentados resumidamente os melhores resultados obtidos nesta pesquisa. São mostrados os resultados nos casos em que a classificação foi realizada considerando-se o desempenho das características de forma individual, bem como a partir da combinação de características. Neste caso, os resultados apresentados indicam a combinação feita com o menor número de características que atingiu o maior valor de acurácia.

Pode ser observado, a partir dos resultados apresentados na Tabela 6.27, que as características dispersão da frequência instantânea (DFI) e a amplitude instantânea ponderada pela frequência instantânea (APF) obtiveram os melhores resultados individuais de acurácia na maioria das classificações. Já os resultados das características que foram combinadas, em quatro casos a acurácia resultou em 100,00% e em outros casos o valor de acurácia foi menor que o valor classificado individualmente. Nos casos em que a acurácia atingiu 100,00% a quantidade máxima de características combinadas foram três e a mínima foram duas características. Ressalta-se que as características combinadas que obtiveram valor de acurácia menor que quando classificadas individualmente, foram os casos que houveram poucas características para realizar a classificação. Também pode ser observado a dificuldade em se discriminar entre os sinais de vozes com grau geral 1 e com grau geral 2, pois esta classificação foi a que obteve o menor valor de acurácia 83,33% e o maior desvio de acurácia que ficou em 10,54%.

Tabela 6.27 – Resumo dos melhores resultados individuais e combinados.

Classificação	Acurácia Individual	Característica Individual	Acurácia Combinada	Características Combinadas
NORM × DESV	92,67 ± 4,52	APF_5	91,67 ± 4,81	$E_5 + APF_5$
NORM × SOP	90,83 ± 4,72	E_2	100,00 ± 0,00	$E_2 + DETI_2 + APT_2$
NORM × RUG	93,05 ± 3,67	APF_5	100,00 ± 0,00	$E_5 + DETI_5 + APF_5$
NORM × TEN	88,10 ± 7,90	DFI_{10}	84,09 ± 4,69	$DFI_1 + DFI_{10}$
RUG × SOP	93,33 ± 6,67	$DETI_3$	100,00 ± 0,00	$E_2 + E_6 + APT_5$
RUG × TEN	93,75 ± 6,25	DFI_3	91,66 ± 5,27	$APF_5 + DFI_3 + DFI_8$
SOP × TEN	95,00 ± 5,00	DFI_7	100,00 ± 0,00	$E_2 + DETI_3$
G1 × G2	83,33 ± 10,54	DFI_5	–	–
G1 × G3	92,71 ± 4,84	APF_5	98,44 ± 1,56	$APT_1 + APT_6 + APF_1 + APF_5$
G2 × G3	87,12 ± 4,55	APF_1	–	–
G1 × G2G3	87,96 ± 4,83	APF_5	87,50 ± 5,59	$DFI_1 + APF_5$

NORM × DESV: classificação entre os sinais de vozes normais e com desvios; NORM × SOP: classificação entre os sinais de vozes normais e com desvio soproisidade; NORM × RUG: classificação entre os sinais de vozes normais e com desvio rugosidade; NORM × TEN: classificação entre os sinais de vozes normais e com desvio tensão; RUG × SOP: classificação entre os sinais de vozes com desvio rugosidade e soproisidade; RUG × TEN: classificação entre os sinais de vozes com desvio rugosidade e tensão; SOP × TEN: classificação entre os sinais de vozes com desvio soproisidade e tensão; G1 × G2: classificação entre os sinais de vozes com grau geral 1 e grau geral 2; G1 × G3: classificação entre os sinais de vozes com grau geral 1 e grau geral 3; G2 × G3: classificação entre os sinais de vozes com grau geral 2 e grau geral 3; G1 × G2G3: classificação entre os sinais de vozes com grau 1 e graus 2 e 3.

6.3 Estudos Comparativos

Nesta Seção, os resultados obtidos neste estudo serão discutidos com resultados publicados na literatura. Para o estudo de caso 1 os resultados desta pesquisa serão discutidos com os resultados de Queiroz (2017), Couras (2017) e Pinho (2017). Já para o estudo de caso 2 os resultados desta pesquisa serão discutidos com os resultados de Queiroz (2017).

A pesquisa de Queiroz (2017) faz uso da análise dinâmica não linear e análise de quantificação de recorrência para discriminar os desvios vocais, a base de dados utilizada procede do LIEV/UFPB e foram utilizadas vozes adultas masculinas e femininas. Couras (2017) empregou técnicas de processamento digital de sinais baseadas no modelo linear de produção da fala para discriminar os desvios vocais, a base de dados utilizada foi proveniente do LIEV/UFPB, porém foram utilizadas vozes adultas femininas. Na pesquisa de Pinho (2017) foi feita uma avaliação das desordens vocais por meio de características do espaço de fase reconstruído, no qual foram utilizadas duas bases de dados em sua pesquisa, mas para critério de discussão só serão considerados os resultados referentes a base de dados do LIEV/UFPB e desta base foram utilizadas vozes adultas masculinas e femininas.

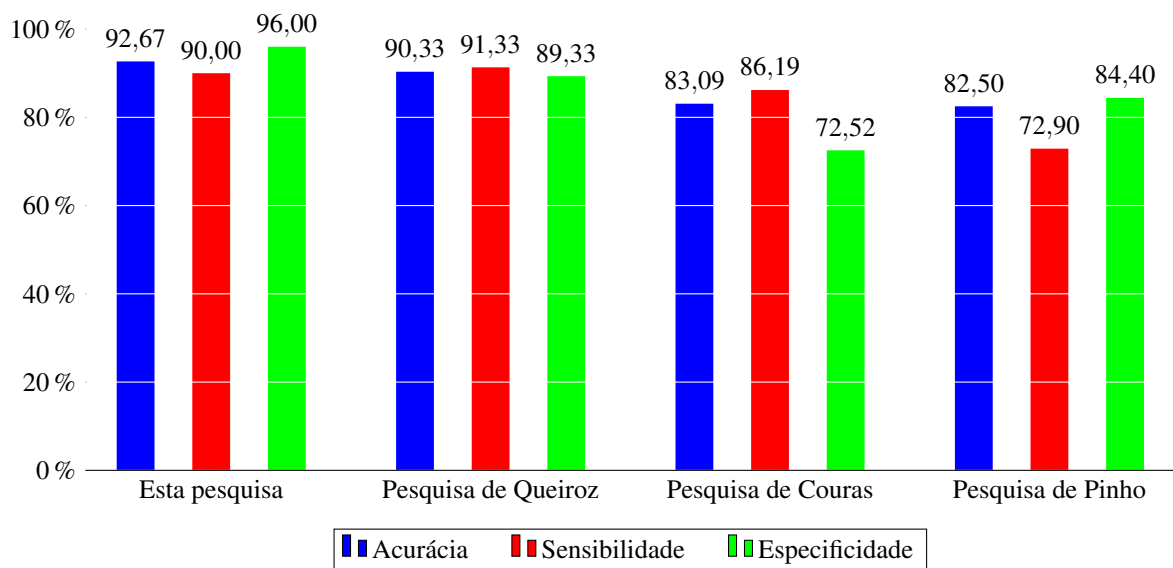
6.3.1 Estudos Comparativos para o Estudo de Caso 1

A Figura 6.12 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade obtidos das pesquisas de Queiroz (2017), Couras (2017), Pinho (2017) e desta pesquisa. Para atingir a acurácia de 90,33% Queiroz (2017) combinou seis características, para obter 83,09% de acurácia Couras (2017) combinou cinco características, para obter 82,50%

de acurácia Pinho (2017) combinou três características e, nesta pesquisa, utilizou-se apenas uma característica para alcançar a acurácia de 92,67%.

Apesar dos valores de acurácia deste estudo e de Queiroz (2017) estarem próximos, o que se destaca neste estudo, além de ter atingido uma maior acurácia quando comparado aos outros estudos, para obter praticamente a mesma acurácia que Couras (2017) obteve, nesta pesquisa utilizou apenas uma característica.

Figura 6.12 – Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvios.

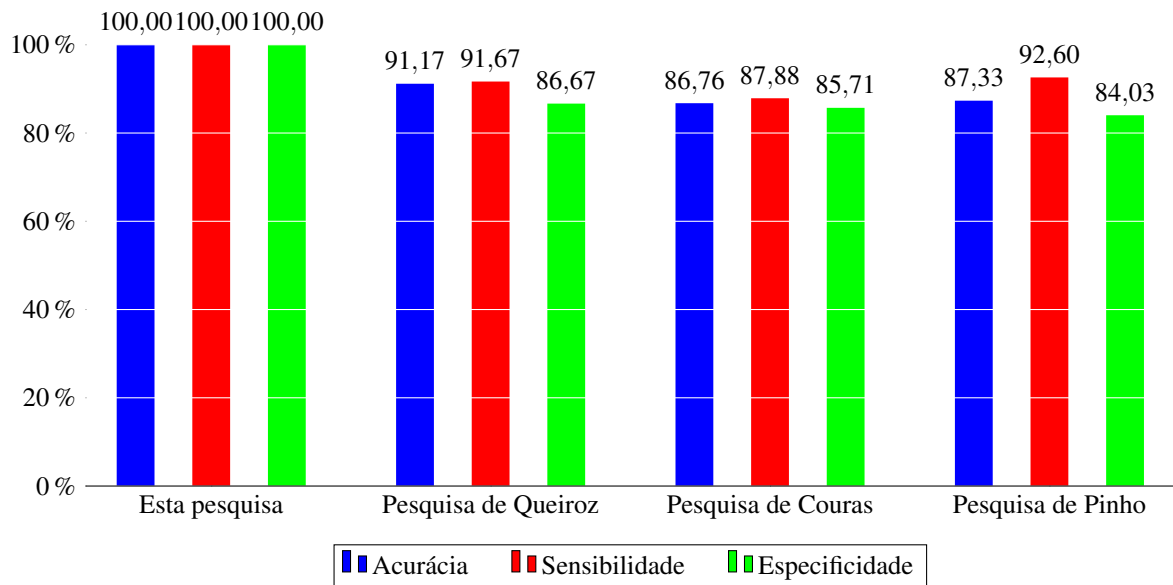


Fonte: Dados da pesquisa, 2018.

A Figura 6.13 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os resultados dos sinais de vozes normais e com desvio soproside obtidos das pesquisas de Queiroz (2017), Couras (2017), Pinho (2017) e desta pesquisa. Para atingir a acurácia de 91,17% Queiroz (2017) combinou duas características, para obter 86,76% de acurácia Couras (2017) combinou três características, para obter 87,33% de acurácia Pinho (2017) combinou três características e, nesta pesquisa, utilizou uma combinação de três características para alcançar a acurácia de 100,00%.

Para esta comparação os resultados desta pesquisa mostram-se os melhores resultados como também utilizando combinações com menores quantidade de características. Chegando a possuir uma acurácia de aproximadamente 9,00% maior quando comparado ao estudo de Queiroz (2017). Indicando que tanto o método proposto como as características que foram selecionadas e combinadas para esta classificação mostraram-se eficientes para realizar a discriminação entre sinais de vozes normais e com desvio soproside. Além da acurácia esta pesquisa se destaca pelos valores de sensibilidade e especificidade, chegando a possuir uma sensibilidade de aproximadamente 7,00% maior que o estudo de Pinho (2017) e uma especificidade aproximadamente 13,00% maior que o estudo de Queiroz (2017).

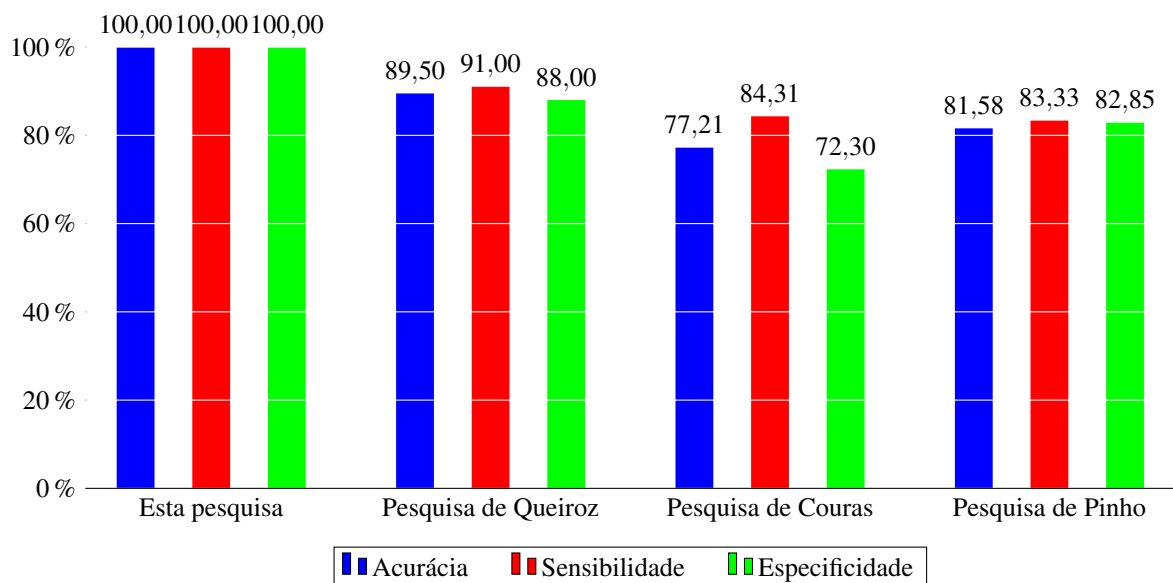
Figura 6.13 – Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvio soproside.



Fonte: Dados da pesquisa, 2018.

A Figura 6.14 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os resultados dos sinais de vozes normais e sinais de vozes com desvio rugosidade obtidos das pesquisas de Queiroz (2017), Couras (2017), Pinho (2017) e desta pesquisa.

Figura 6.14 – Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvio rugosidade.



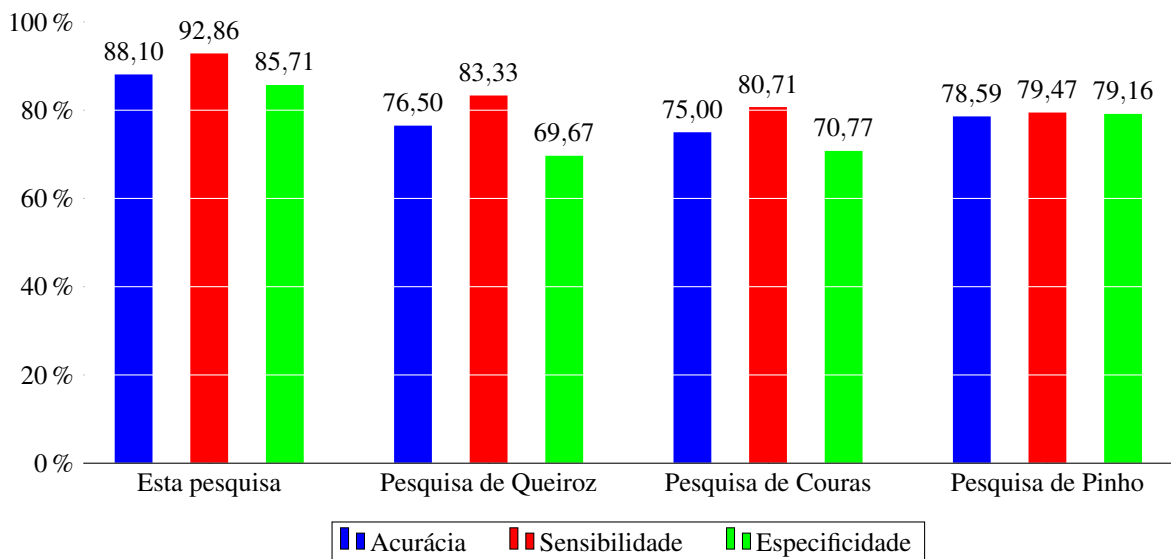
Fonte: Dados da pesquisa, 2018.

Para atingir a acurácia de 89,50% Queiroz (2017) combinou quatro características, para obter 77,21% de acurácia Couras (2017) combinou três características, para obter 81,58% de

acurácia Pinho (2017) combinou três características e, nesta pesquisa, utilizou uma combinação de três características para obter uma acurácia de 100,00% . Os resultados desta pesquisa mostram-se os melhores resultados como também utilizando combinações com menores quantidade de características. Chegando a possuir uma acurácia de aproximadamente 10,00% maior quando comparado ao estudo de Queiroz (2017).

A Figura 6.15 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os resultados dos sinais de vozes normais e com desvio tensão obtidos das pesquisas de Queiroz (2017), Couras (2017), Pinho (2017) e desta pesquisa. Para atingir a acurácia de 76,50% Queiroz (2017) combinou seis características, para obter 75,00% de acurácia Couras (2017) combinou cinco características, para obter 78,59% de acurácia Pinho (2017) combinou duas características e, nesta pesquisa, utilizou apenas uma característica para obter a acurácia de 88,10%. Esta pesquisa apresenta o melhor resultado de acurácia para esta classificação como também utiliza a menor quantidade de características para atingir estes resultados. A acurácia desta pesquisa chegar ser aproximadamente 9,00% maior quando comparado ao resultado de acurácia de Pinho (2017).

Figura 6.15 – Comparativo entre os resultados obtidos por Queiroz (2017), com os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes normais e com desvio tensão.

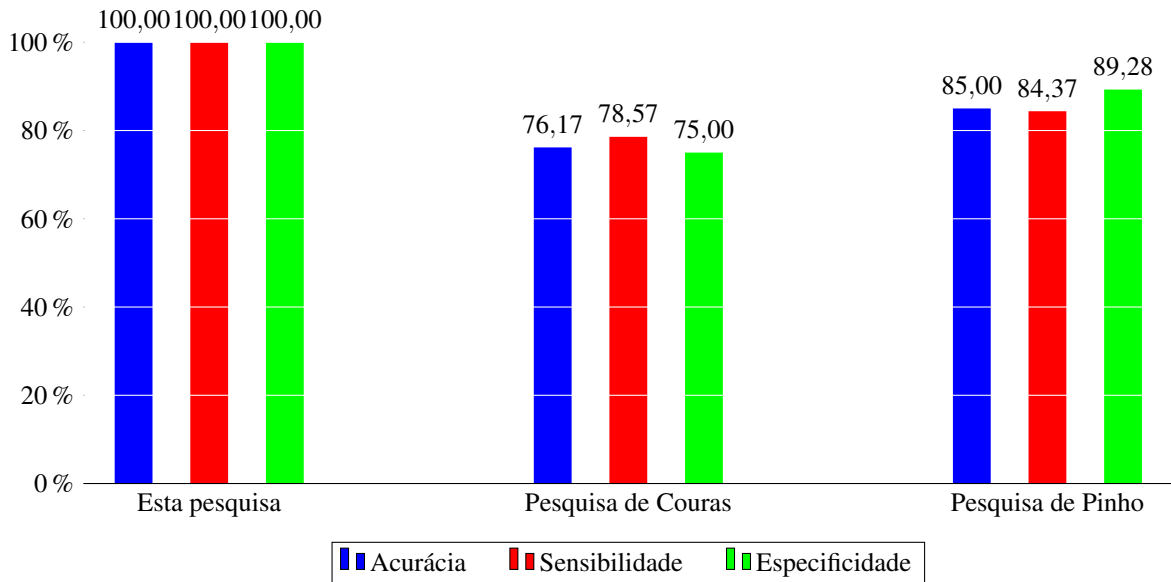


Fonte: Dados da pesquisa, 2018.

A Figura 6.16 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os resultados dos sinais de vozes com desvio rugosidade e com desvio sopro obtidos das pesquisas de Couras (2017), Pinho (2017) e desta pesquisa. Para atingir a acurácia de 76,17% Couras (2017) utilizou um *codebook* com 256 níveis, para obter 85,00% de acurácia Pinho (2017) combinou quatro características e, nesta pesquisa, foram combinadas três características para obter a acurácia de 100,00%. Esta pesquisa apresenta o melhor resultado de acurácia para esta classificação como também utiliza a menor quantidade

de características para atingir estes resultados. A acurácia desta pesquisa é aproximadamente 15,00% maior quando comparado ao resultado de acurácia de Pinho (2017).

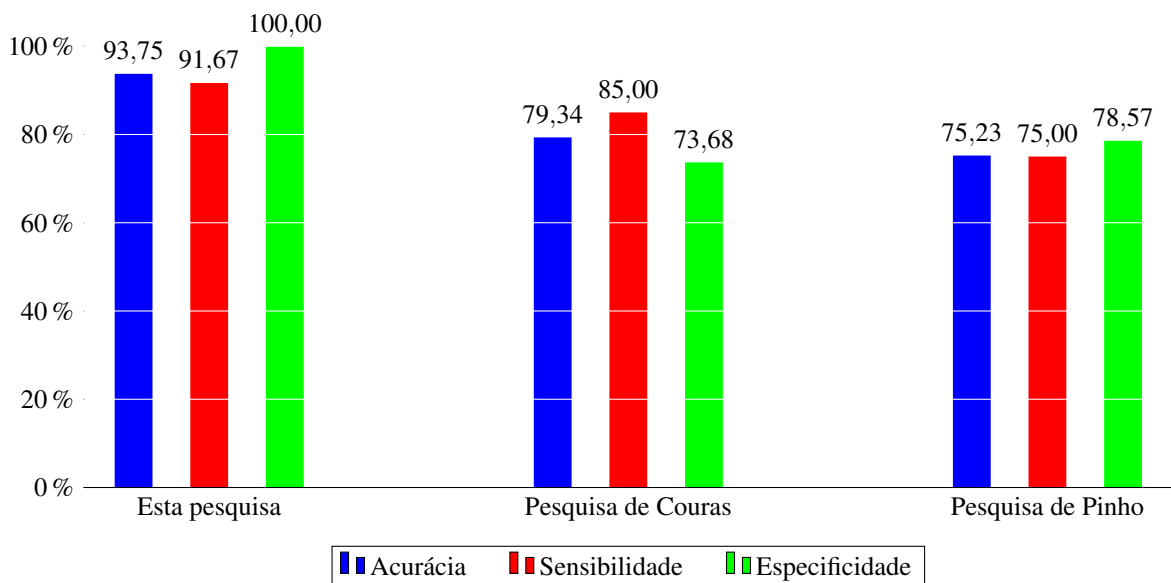
Figura 6.16 – Comparativo entre os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com desvio rugosidade e com desvio soproidade.



Fonte: Dados da pesquisa, 2018.

A Figura 6.17 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os resultados dos sinais de vozes com desvio rugosidade e com desvio tensão obtidos das pesquisas de Couras (2017), Pinho (2017) e desta pesquisa.

Figura 6.17 – Comparativo entre os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com desvio rugosidade e com desvio tensão.

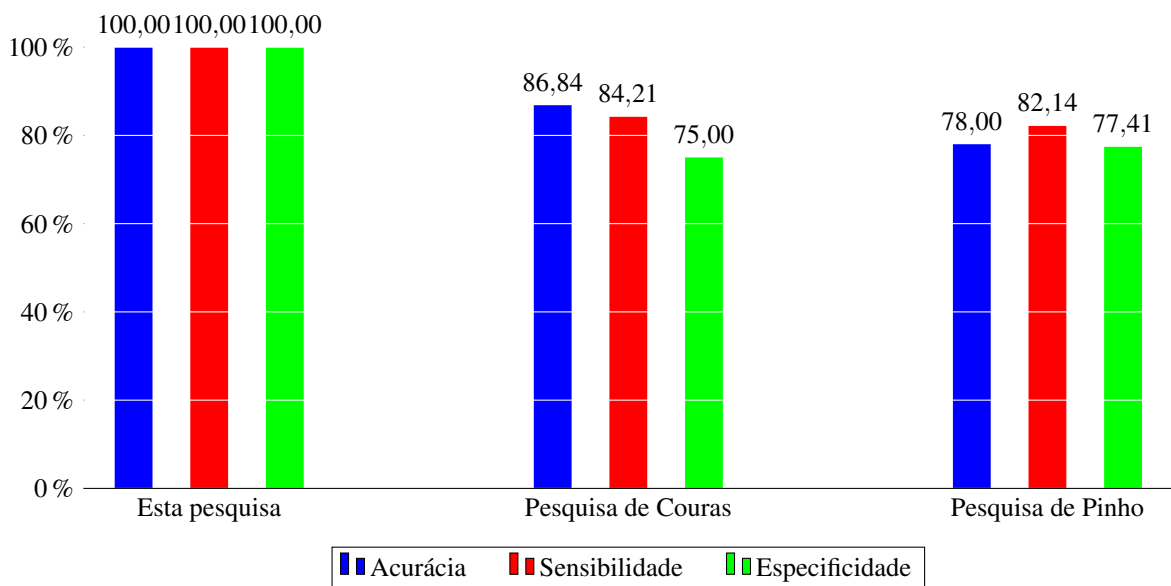


Fonte: Dados da pesquisa, 2018.

Para atingir a acurácia de 79,34% Couras (2017) utilizou coeficientes mel-cepstrais com 128 níveis, para obter 75,23% de acurácia Pinho (2017) combinou três características e, nesta pesquisa, foi utilizada apenas uma característica para alcançar a acurácia de 93,75%. Esta pesquisa apresenta o melhor resultado de acurácia para esta classificação como também utiliza a menor quantidade de característica para atingir estes resultados. A acurácia desta pesquisa é aproximadamente 14% maior quando comparado ao resultado de acurácia de Couras (2017).

A Figura 6.18 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os resultados dos sinais de vozes com desvio soproside e com desvio tensão obtidos das pesquisas de Couras (2017), Pinho (2017) e desta pesquisa. Para atingir a acurácia de 86,84% Couras (2017) utilizou coeficientes mel-cepstrais com 1024 níveis, para obter 78,00% de acurácia Pinho (2017) combinou duas características e, nesta pesquisa, foram combinadas duas características para obter a acurácia de 100,00%. Esta pesquisa apresenta o melhor resultado de acurácia para esta classificação, como também utiliza a menor quantidade de características para atingir estes resultados. A acurácia desta pesquisa é aproximadamente 13,00% maior, quando comparado ao resultado de acurácia de Couras (2017).

Figura 6.18 – Comparativo entre os resultados obtidos por Couras (2017), com os resultados obtidos por Pinho (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com desvio soproside e com desvio tensão.



Fonte: Dados da pesquisa, 2018.

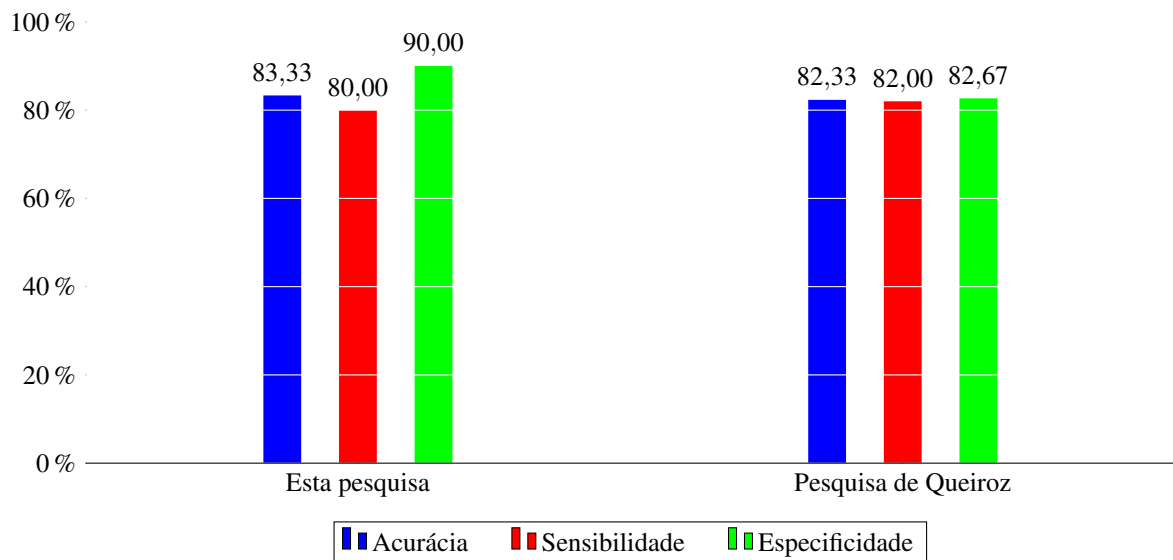
De um modo geral, as características combinadas demonstraram um potencial discriminativo melhor do que as características individuais. Os resultados desta pesquisa, para o Estudo de Caso 1, obtiveram em todas as situações analisadas valores de acurácia maiores quando comparados aos estudos de Queiroz (2017), Couras (2017) e Pinho (2017), chegando a ter uma diferença de 15,00% no valor de acurácia para a classificação entre sinais de vozes com os desvios rugosidade e soproside. Para analisar os resultados qualitativamente, foram adotados os padrões de Hosmer, Lemeshow e Sturdivant (2013) que, para o Estudo de Caso 1,

são considerados excelentes para todas as classificações, com exceção da classificação entre os sinais de vozes normais e os sinais de vozes com desvio tensão que é considerado bom. Com os resultados apresentados, é possível afirmar que, para a base de dados do LIEV/UFPB, o modelo adotado nesta pesquisa para classificação de sinais de vozes com e sem desvios vocais, através da transformada de Hilbert-Huang, foi superior aos demais métodos empregados nas outras pesquisas comparadas.

6.3.2 Estudos Comparativos para o Estudo de Caso 2

A Figura 6.19 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os sinais de vozes normais (Grau 1 - G1) e de grau leve (Grau 2 - G2), obtidos da pesquisa de Queiroz (2017) e desta pesquisa. Para atingir a acurácia de 82,33% Queiroz (2017) combinou quatro características e, nesta pesquisa, para alcançar a acurácia de 83,33% foi utilizada apenas uma característica. Apesar dos valores de acurácia deste estudo e de Queiroz (2017) estarem com valores aproximados, o que se destaca neste estudo, quando comparado ao estudo de Queiroz (2017), é que a quantidade de característica utilizada foi somente uma característica, enquanto Queiroz (2017) utilizou uma combinação com quatro características.

Figura 6.19 – Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau geral 1 e grau geral 2.

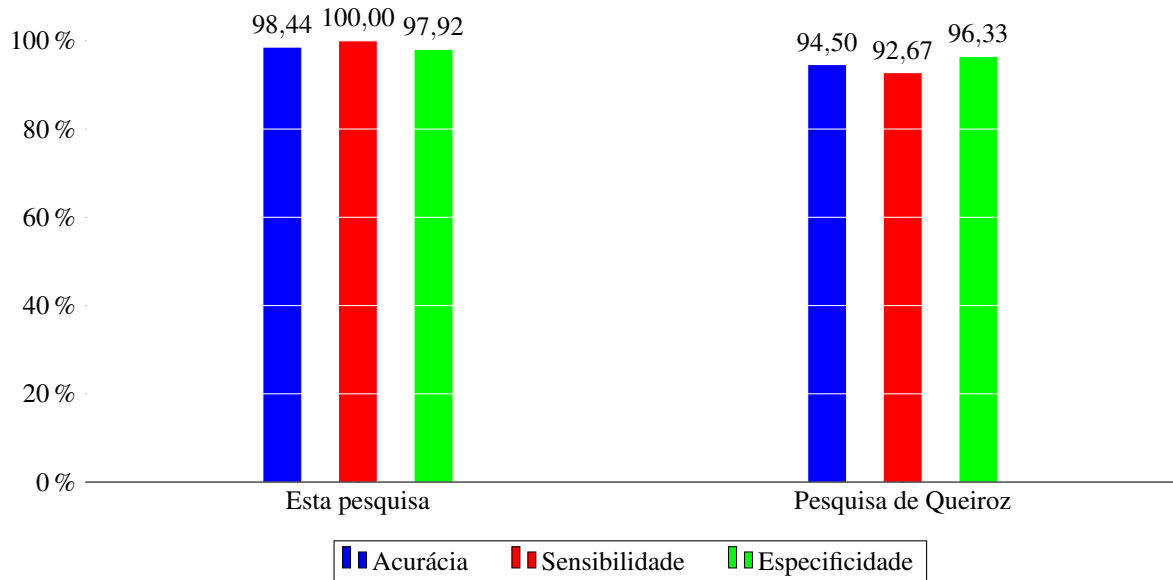


Fonte: Dados da pesquisa, 2018.

A Figura 6.20 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os sinais de vozes normais (Grau 1 - G1) e de grau moderado (Grau 3 - G3), obtidos da pesquisa de Queiroz (2017) e desta pesquisa. Para atingir a acurácia de 94,50% Queiroz (2017) combinou cinco características e, nesta pesquisa, para alcançar a acurácia de 98,44% foi utilizada uma combinação com quatro características. Neste estudo o

valor da acurácia é maior aproximadamente em 4,00% em relação a acurácia obtida por Queiroz (2017) para esta classificação.

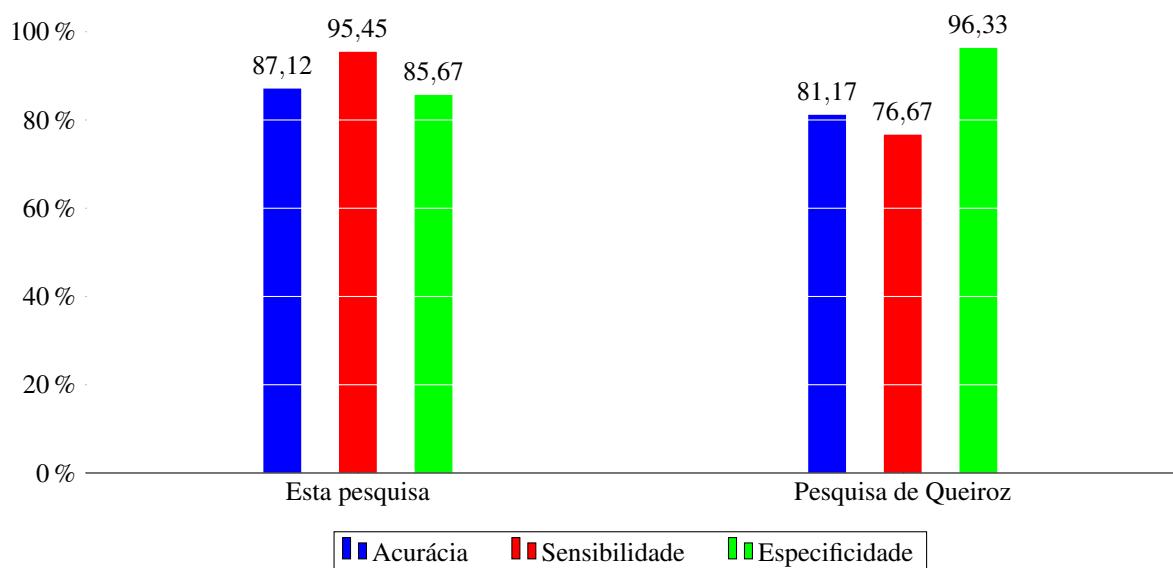
Figura 6.20 – Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau geral 1 e grau geral 3.



Fonte: Dados da pesquisa, 2018.

A Figura 6.21 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os sinais de vozes de grau leve (Grau 2 - G2) e de grau moderado (Grau 3 - G3), obtidos da pesquisa de Queiroz (2017) e desta pesquisa.

Figura 6.21 – Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau geral 2 e grau geral 3.



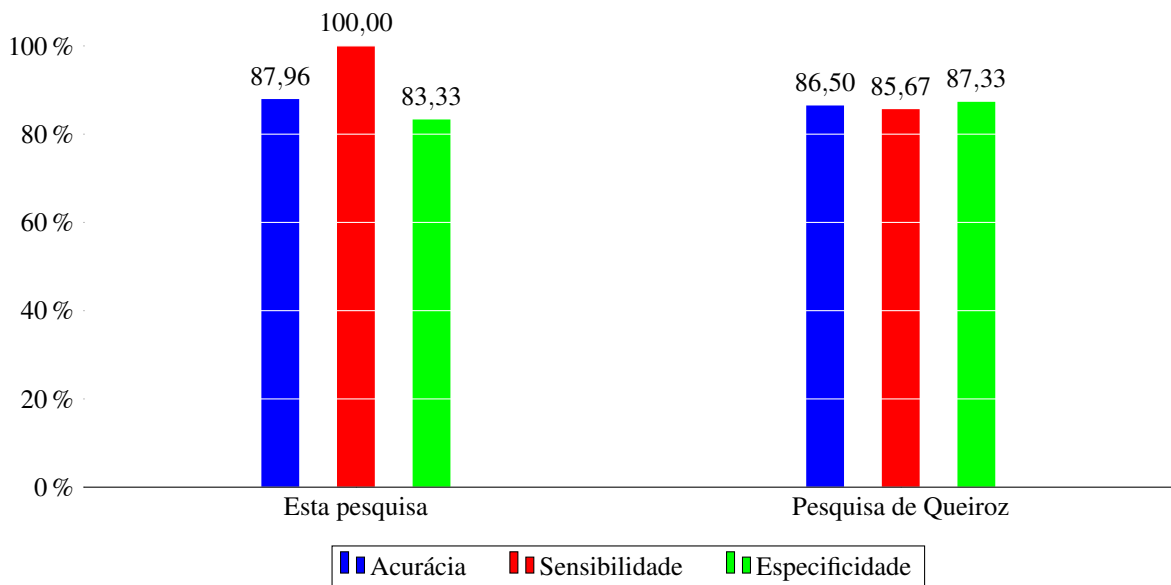
Fonte: Dados da pesquisa, 2018.

Para atingir a acurácia de 81,17% Queiroz (2017) combinou quatro características e, nesta pesquisa, para alcançar a acurácia de 87,12% foi utilizada apenas uma característica. Neste

estudo, o valor da acurácia é maior aproximadamente em 6,00% em relação a acurácia obtida por Queiroz (2017) para esta classificação.

A Figura 6.22 demonstra um comparativo entre os principais valores de acurácia, sensibilidade e especificidade entre os sinais de vozes de grau 1 e de graus 2 e 3, obtidos da pesquisa de Queiroz (2017) e desta pesquisa. Para atingir a acurácia de 86,50% Queiroz (2017) combinou sete características e, nesta pesquisa, para alcançar a acurácia de 87,96% foi utilizada apenas uma característica. Os valores de acurácia deste estudo e do estudo de Queiroz (2017) são bem próximos. Neste estudo, o destaque para esta classificação é a quantidade de característica, pois foi utilizada uma característica para obtenção desses resultados enquanto no estudo de Queiroz (2017) foram utilizadas sete características.

Figura 6.22 – Comparativo entre os resultados obtidos por Queiroz (2017) e com os resultados obtidos nesta pesquisa, na classificação entre sinais de vozes com grau 1 e graus 2 e 3.



Fonte: Dados da pesquisa, 2018.

De um modo geral o estudo de caso 2, as características individuais demonstraram um potencial discriminativo melhor do que as características combinadas. Os resultados desta pesquisa, para o Estudo de Caso 2, obtiveram em todas as situações analisadas valores de acurácia maiores quando comparados a pesquisa de Queiroz (2017), chegando a ter uma diferença de aproximadamente 6,00% no valor de acurácia para a classificação entre sinais de vozes com grau geral 2 e com grau geral 3. Para analisar os resultados qualitativamente, foram adotados os padrões de Hosmer, Lemeshow e Sturdivant (2013) que, para o Estudo de Caso 2, são considerados bons para todas as classificações, com exceção da classificação entre os sinais de vozes grau geral 1 e grau geral 3 que é considerado excelente.

7 CONSIDERAÇÕES FINAIS

Neste estudo, foi possível avaliar a aplicabilidade da transformada de Hilbert-Huang em dois estudos de caso: 1) Análise acústica de sinais de voz saudáveis e desviadas (rugosidade, sopro e tensão); e 2) Análise acústica do grau de intensidade de distúrbios. Para os dois estudos foi aplicado o método de decomposição empírica, no qual, os sinais de voz foram decompostos em dez faixas de frequência. Diante dos resultados obtidos a transformada de Hilbert-Huang se destaca como uma ferramenta aplicável para análise acústica.

Dessa forma, aponta-se a possibilidade de uso da transformada de Hilbert-Huang por parte dos profissionais e pesquisadores na área de voz, uma vez que por meio desta é possível auxiliar no processo de análise, classificação e tomada de decisão clínica no processo diagnóstico e tratamento de pacientes com distúrbios vocais.

No que se refere a validação de características que melhor representaram os desvios vocais dos sinais analisados, observou-se que as características dispersão da frequência instantânea (*DFI*) e a amplitude instantânea ponderada pela frequência instantânea (*APF*) obtiveram os melhores resultados individuais de acurácia na maioria das classificações. Tais achados indicam que as características espectrais tiveram melhor representação das oscilações locais dos sinais de voz estudados.

Como caráter inovador, esta pesquisa destaca-se pelo fato de que não foi encontrado na literatura da área um estudo que relacione a transformada de Hilbert-Huang à avaliação de desvios vocais, assim como também a intensidade destes desvios, bem como à análise do espectro de Hilbert em relação aos desvios vocais (rugosidade, sopro e tensão) e a intensidade destes desvios.

O estudo apontou como principais resultados em relação ao desempenho das características individuais, a amplitude instantânea ponderada pela frequência instantânea, extraída na quinta função intrínseca de modo, mostrou o melhor potencial de classificação para discriminar vozes normais e com desvios, com acurácia de $92,67\% \pm 4,52\%$. Quando as características foram combinadas, atingindo uma acurácia de 100%, para discriminar tanto entre os sinais de vozes normais e com desvio sopro quanto entre sinais de vozes normais e com desvio rugosidade.

Diante dos resultados obtidos para os sinais de voz utilizados da base de dados do LIEV nos dois estudos de caso pode-se considerar a transformada de Hilbert-Huang em conjunto com o classificador *Multilayer Perceptron*, que utiliza o algoritmo de aprendizado supervisionado do gradiente conjugado escalonado, tem possibilidades de contribuir para detectar desvios vocais e auxiliar com informações relevantes para que os profissionais de saúde possam desenvolver novos estudos e estratégias ligadas ao diagnóstico e tratamentos dos pacientes.

De forma geral, as principais contribuições deste estudo foram:

- Obtenção de características por meio do uso da transformada de Hilbert-Huang para a discriminação entre sinais de vozes normais e com desvios vocais (rugosidade, sopro e tensão), como também em relação às intensidades dos graus dos desvios vocais;
- Abordagem com decomposição tempo-frequência baseada em EMD para obtenção de espectros de Hilbert com características das oscilações locais de cada sinal de voz estudado;
- Aplicação da rede neural MLP, com as características temporais e espectrais obtidas com a aplicação da transformada de Hilbert-Huang.
- Proporcionar técnicas que podem se tornar ferramentas de apoio ao diagnóstico e tratamento de desvios vocais para profissionais da área da voz;
- Contribuição para o estudo, aprofundamento e teorização de métodos de análise de desvios vocais para pesquisadores e estudiosos da área.

7.1 Sugestões para Trabalhos Futuros

- Avaliar o uso dos espectros de Hilbert das IMFs para treinamento e aperfeiçoamento de especialistas;
- Aplicar processamento digital de imagens para extrair novas características do espectro de Hilbert para realizar a classificação dos sinais;
- Avaliar a influência da presença de lesão laríngea nos desvios vocais e em seu processo de classificação;
- Utilizar outros classificadores na discriminação dos desvios tais como, máquinas de vetor de suporte e redes neurais profundas.

REFERÊNCIAS

- AKPINAR, M. E. et al. Effects of soft palate implants on acoustic characteristics of voice and articulation. **Journal of Voice**, Elsevier, v. 25, n. 3, p. 381–386, 2011.
- ALI, Z. et al. Detection of voice pathology using fractal dimension in a multiresolution analysis of normal and disordered speech signals. **Journal of medical systems**, Springer, v. 40, n. 1, p. 20, 2016.
- ANKIŞHAN, H. A new approach for detection of pathological voice disorders with reduced parameters. **Electrica**, v. 18, n. 1, p. 60–71, 2018.
- AUGER, F. et al. **Time-frequency toolbox**. 1995. Disponível em: <<http://tftb.nongnu.org>>. Acesso em: 03 de março de 2018.
- BARAVIEIRA, P. B. **Aplicação de uma Rede Neural Artificial para a Avaliação da Rugosidade e Soprosidade Vocal**. Tese (Doutorado em Bioengenharia) — Universidade de São Paulo, São Carlos-SP, 2016.
- BARCELOS, C. B. et al. Multidimensional effects of voice therapy in patients affected by unilateral vocal fold paralysis due to cancer. **Brazilian journal of otorhinolaryngology**, SciELO Brasil, v. 84, n. 5, p. 620–629, 2018.
- BEHLAU, M. **Voz: o livro do especialista**. Rio de Janeiro: Revinter, 2001.
- BEHLAU, M. Oficina de técnicas de avaliação de voz. In: XV CONGRESSO BRASILEIRO DE FONOAUDIOLOGIA. **Sociedade Brasileira de Fonoaudiologia**. [S.l.], 2008. v. 1, n. 1.
- BEHLAU, M.; GAMA, A. C.; CIELO, C. A. Técnicas vocais. In: FERNANDES, F. D. M.; MENDES, B. C. A.; NAVAS A. L. G. P. **Tratado de Fonoaudiologia**. 2ª edição. ed. São Paulo: Editora Roca, 2010.
- BEHLAU, M.; PONTES, P.; MORETI, F. **Higiene vocal: cuidando da voz**. [S.l.]: Thieme Revinter Publicações LTDA, 2018.
- BELE, I. V. Reliability in perceptual analysis of voice quality. **Journal of Voice**, Elsevier, v. 19, n. 4, p. 555–573, 2005.
- BELHAJ, A.; BOUZID, A.; ELLOUZE, N. Normal and organic pathology classification of female voices using svm classifier. **International Journal of Applied Engineering Research**, v. 11, n. 15, p. 8702–8707, 2016.
- 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.
- BURRUS, C. S. et al. **Introduction to wavelets and wavelet transforms: a primer**. 1. ed. New Jersey: Prentice hall, 1998.
- CAMARGO, Z.; MADUREIRA, S. Análise acústica: aplicações na fonoaudiologia. **Tratado de Fonoaudiologia**, Roca, São Paulo, v. 2, p. 695–699, 2010.

CARDOSO, L. F.; LIMA, N. C. Utilização de redes neurais artificiais para detecção de arritmias cardíacas. In: ANAIS DO ENCONTRO DE COMPUTAÇÃO DO OESTE POTIGUAR ECOP/UFERSA. **III Encontro de Computação do Oeste Potiguar**. [S.l.], 2018. v. 2, n. 1.

CASPER, J. K.; LEONARD, R. **Understanding voice problems: A physiological perspective for diagnosis and treatment**. [S.l.]: Lippincott Williams & Wilkins, 2006.

CASTAÑÓN, C. A. B. **Recuperação de imagens por conteúdo através de análise multiresolução por wavelets**. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, São Carlos-SP, 2003.

CEBALLOS, A. G. d. C. d. et al. Avaliação perceptivo-auditiva e fatores associados à alteração vocal em professores. **Revista Brasileira de Epidemiologia**, SciELO Public Health, v. 14, p. 285–295, 2011.

COELHO, R. F. et al. **Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing**. [S.l.]: CRC Press, 2015.

COSTA, S. L. do N. C. **Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas**. Tese (Doutorado em Engenharia Elétrica) — Universidade Federal de Campina Grande, Campina Grande-PB, 2008.

COSTA, W. C. A. **Análise dinâmica não linear de sinais de voz para detecção de patologias laríngeas**. Tese (Doutorado em Engenharia Elétrica) — Universidade Federal de Campina Grande, Campina Grande-PB, 2012.

COURAS, M. F. K. B. **Classificação de Desvios Vocais Utilizando Características Baseadas no Modelo Linear de Produção da Fala**. Dissertação (Mestrado em Engenharia Elétrica) — Instituto Federal de educação, ciência e tecnologia da Paraíba, João Pessoa–PB, 2017.

CRESPO, J. H. A. **Utilização do método de decomposição empírico no processamento de dados de mobilidade urbana**. Dissertação (Mestrado em Matemática Aplicada) — Fundação Getúlio Vargas, Rio de Janeiro-RJ, 2018.

CROVATO, C. D. P. **Classificação de Sinais de Voz Utilizando a Transformada Wavelet Packet e Redes Neurais Artificiais**. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, 2004.

DAJER, M. E. **Análise de sinais de voz por padrões visuais de dinâmica vocal**. Tese (Doutorado) — Universidade de São Paulo, 2010.

DEJONCKERE, P. H. et al. To what degree of voice perturbation are jitter measurements valid? a novel approach with synthesized vowels and visuo-perceptual pattern recognition. **Biomedical Signal Processing and Control**, Elsevier, v. 7, n. 1, p. 37–42, 2012.

DING, H. et al. Hilbert–huang transform based signal analysis for the characterization of gas–liquid two-phase flow. **Flow measurement and instrumentation**, Elsevier, v. 18, n. 1, p. 37–46, 2007.

DUDA, R.; HART, P.; STORK, D. Multilayer neural networks. **Pattern classification**, Wiley, p. 287, 2001.

- ESPER, N. B. **Utilização de métodos de decomposição empíricos no pré-processamento de dados de ressonância magnética funcional**. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio Grande do Sul, 2016.
- FANT, G. **Acoustic Theory of Speech Production**. [S.l.]: Walter de Gruyter, 1970.
- FLANDRIN, P.; RILLING, G.; GONCALVES, P. Empirical mode decomposition as a filter bank. **IEEE signal processing letters**, IEEE, v. 11, n. 2, p. 112–114, 2004.
- FLANDRIN, P. et al. **Empirical Mode Decomposition**. 2007. Disponível em: <<http://perso.ens-lyon.fr/patrick.flandrin/emd.html>>. Acesso em: 10 de fevereiro de 2018.
- GAMA, A. C. C. et al. Correlation between acoustic-perceptual data and voice-related quality of life in elderly women. **Pró-Fono Revista de Atualização Científica**, SciELO Brasil, v. 21, n. 2, p. 125–130, 2009.
- GAYOTTO, L. H. Dinâmicas de movimento da voz. **Distúrbios da Comunicação**, v. 17, n. 3, 2005.
- GOULART, D.; COOPER, M. **Por todo canto: método de técnica; vocal música popular**. [S.l.]: G4 Ed., 2002.
- GUIMARÃES, I. A ciência e a arte da voz humana. **Alcoitão, Escola Superior de Saúde de Alcoitão**, 2007.
- HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2007.
- HIGUCHI, T. Approach to an irregular time series on the basis of the fractal theory. **Physica D: Nonlinear Phenomena**, Elsevier, v. 31, n. 2, p. 277–283, 1988.
- HIRANO, M. Clinical examination of voice. **Disorders of human communication**, Springer-Verlag, v. 5, p. 1–99, 1981.
- HOSMER, D. W. J.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. [S.l.]: John Wiley & Sons, 2013. v. 398.
- HUANG, N. E. et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. **Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 454, n. 1971, p. 903–995, 1998.
- HUANG, N. E. et al. A confidence limit for the empirical mode decomposition and hilbert spectral analysis. **Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 459, n. 2037, p. 2317–2345, 2003.
- HUANG, N. E.; WU, Z.; LONG, S. R. Hilbert–huang transform. **Scholarpedia**, v. 3, n. 7, p. 2544, 2008.
- KALEEM, M. et al. Telephone-quality pathological speech classification using empirical mode decomposition. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE. **Engineering in Medicine and Biology Society**. [S.l.], 2011. p. 7095–7098.
- KALEEM, M. et al. Pathological speech signal analysis and classification using empirical mode decomposition. **Medical & biological engineering & computing**, Springer, v. 51, n. 7, p. 811–821, 2013.

KALEEM, M. F. et al. Application of empirical mode decomposition and teager energy operator to eeg signals for mental task classification. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY. **Engineering in Medicine and Biology Society**. [S.l.], 2010. v. 1, n. 1, p. 4590–4593.

KATZ, M. J. Fractals and the analysis of waveforms. **Computers in biology and medicine**, Elsevier, v. 18, n. 3, p. 145–156, 1988.

KEMPSTER, G. B. et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. **American Journal of Speech-Language Pathology**, ASHA, 2009.

KREIMAN, J.; GERRATT, B. R. Perceptual assessment of voice quality: past, present and future. **Perspectives on Voice and Voice Disorders**, ASHA, v. 20, n. 2, p. 62–67, 2010.

KREIMAN, J.; GERRATT, B. R.; ITO, M. When and why listeners disagree in voice quality assessment tasks. **The Journal of the Acoustical Society of America**, ASA, v. 122, n. 4, p. 2354–2364, 2007.

LOPES, L. W. et al. Severity of voice disorders in children: correlations between perceptual and acoustic data. **Journal of Voice**, Elsevier, v. 26, n. 6, p. 819–e7, 2012.

LOPES, L. W. et al. Relationship between vocal symptoms, severity of voice disorders, and laryngeal diagnosis in patients with voice disorders. In: SCIELO BRASIL. **CoDAS**. [S.l.], 2016. v. 28, n. 4, p. 439–445.

LOPES, L. W. et al. Effectiveness of recurrence quantification measures in discriminating subjects with and without voice disorders. **Journal of Voice**, Elsevier, 2018.

LOPES, L. W.; VILELA, E. G. Autoavaliação e prontidão para mudança em pacientes disfônicos. In: SCIELO BRASIL. **CoDAS**. [S.l.], 2016. v. 28, n. 3, p. 295–301.

LOPES, W. L.; ALVES, G. A. S.; MELO, M. L. Evidência de conteúdo de um protocolo de análise espectrográfica. **Revista CEFAC**, Instituto Cefac, v. 19, n. 4, 2017.

MACCALLUM, J. K. et al. Acoustic analysis of aperiodic voice: perturbation and nonlinear dynamic properties in esophageal phonation. **Journal of Voice**, Elsevier, v. 23, n. 3, p. 283–290, 2009.

MACKAY, D. J.; KAY, D. J. M. **Information theory, inference and learning algorithms**. [S.l.]: Cambridge university press, 2003.

MADAZIO, G. **Diagrama de desvio fonatório na clínica vocal**. Tese (Doutorado) — Escola Paulista de Medicina, 2009.

MARQUES, J. R. **Aplicação da transformada de Hilbert-Huang na análise das vibrações dos motores de indução de máquinas ferramentas**. Tese (Doutorado em Engenharia Elétrica) — Universidade de São Paulo, 2013.

MARTINEZ, C. E.; RUFINER, H. L. Acoustic analysis of speech for detection of laryngeal pathologies. In: PROCEEDINGS OF THE 22ND ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE. **Engineering in Medicine and Biology Society**. [S.l.], 2000. v. 3, p. 2369–2372.

MARTINS, L. G. N. **Usando a Decomposição em Modos Empíricos para determinação de fluxos turbulentos entre oceano/atmosfera**. Tese (Doutorado em Física) — Universidade Federal de Santa Maria, 2015.

MARTINS, P. C.; COUTO, T. E.; GAMA, A. C. C. Avaliação perceptivo-auditiva do grau de desvio vocal: correlação entre escala visual analógica e escala numérica. In: SCIELO BRASIL. **CoDAS**. [S.l.], 2015. v. 27, n. 3, p. 279–284.

MEKYSKA, J. et al. Robust and complex approach of pathological speech signal analysis. **Neurocomputing**, Elsevier, v. 167, p. 94–111, 2015.

MELO, T. M. **Aplicação da Transformada de Hilbert e Separação em Bandas Clínicas para Redução do Eletroencefalograma de Longa Duração**. Dissertação (Mestrado em Engenharia Biomédica) — Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, 2015.

MOLLA, M. K. I. et al. Empirical mode decomposition for advanced speech signal processing. **Journal of Signal Processing**, Research Institute of Signal Processing, Japan, v. 17, n. 6, p. 215–229, 2013.

MORO-VELÁZQUEZ, L. et al. Modulation spectra morphological parameters: a new method to assess voice pathologies according to the grbas scale. **BioMed Research International**, Hindawi Publishing Corporation, 2015.

MOTA, V. T.; FILHO, P. F. O. **SPSS: Análise de Dados Biomédicos**. [S.l.]: Medbook, 2009.

OPPENHEIM, A. V.; SCHAFER, R. W. **Processamento em Tempo Discreto de Sinais**. [S.l.]: PEARSON, 2012.

PANCHAL, F. S.; PANCHAL, M. Review on methods of selecting number of hidden nodes in artificial neural network. **International Journal of Computer Science and Mobile Computing**, v. 3, n. 11, p. 455–464, 2014.

PANEK, D. et al. Acoustic analysis assessment in speech pathology detection. **International Journal of Applied Mathematics and Computer Science**, De Gruyter Open, v. 25, n. 3, p. 631–643, 2015.

PARRAGA, A. **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. Dissertação (Mestrado em Engenharia Elétrica e de Computação) — Universidade Federal do Rio Grande do Sul, Porto Alegre - RS, 2002.

PATEL, S.; SHRIVASTAV, R. Perception of dysphonic vocal quality: some thoughts and research update. **Perspectives on voice and voice disorders**, ASHA, v. 17, n. 2, p. 3–7, 2007.

PINHO, P. H. U. d. **Avaliação de Desordens Vocais por Meio de Características do Espaço de Fase Reconstruído**. Dissertação (Mestrado em Engenharia Elétrica), João Pessoa-PB, 2017.

PINHO, S. M. R. **Fundamentos em fonoaudiologia: tratando os distúrbios da voz**. [S.l.]: Guanabara Koogan, 2003.

PINTO, R. M. d. F. **Novas abordagens ao estudo de sinais biomédicos: análise em tempo-frequência e transformada de Hilbert-Huang**. Tese (Doutorado em Engenharia Biomédica e Biofísica) — Universidade de Lisboa, Lisboa-Portugal, 2009.

- PUSTELNIK, N.; BORGNAT, P.; FLANDRIN, P. Empirical mode decomposition revisited by multicomponent non-smooth convex optimization. **Signal Processing**, Elsevier, v. 102, p. 313–331, 2014.
- QUEIROZ, G. K. L. P. **Análise Dinâmica não Linear e Análise de Quantificação de Recorrência Aplicadas na Classificação de Desvios Vocais**. Dissertação (Mestrado em Engenharia Elétrica) — Instituto Federal de educação, ciência e tecnologia da Paraíba, João Pessoa–PB, 2017.
- RAO, A. R.; HSU, E.-C. **Hilbert-Huang transform analysis of hydrological and environmental time series**. [S.l.]: Springer Science & Business Media, 2008. v. 60.
- RAVINDRAN, P.; NAIR, V. V. Analysis of vocal tract disorders using mel-frequency cepstral coefficients and empirical mode decomposition based features. In: INTERNATIONAL CONFERENCE ON CONTROL COMMUNICATION. **Control Communication & Computing India**. [S.l.], 2015. p. 505–510.
- REBELO, G. M. I. et al. Função de transferência das vogais orais do português brasileiro: análise acústica comparativa. **Brazilian Journal of Otorhinolaryngology**, Associação Brasileira de Otorrinolaringologia e Cirurgia Cérvico-Facial, v. 75, n. 5, 2009.
- RILLING, G.; FLANDRIN, P. On the influence of sampling on the empirical mode decomposition. In: INTERNATIONAL CONFERENCE ON ACOUSTICS SPEECH AND SIGNAL PROCESSING PROCEEDINGS. **Acoustics Speech and Signal Processing**. [S.l.], 2006. v. 3, p. III–III.
- RILLING, G. et al. On empirical mode decomposition and its algorithms. In: PROCEEDINGS OF IEEE-EURASIP WORKSHOP ON NONLINEAR SIGNAL AND IMAGE PROCESSING. **Nonlinear Signal and Image Processing**. [S.l.], 2003. v. 3, p. 8–11.
- RODRIGUES, G. W.; BATISTA, F. B. Suavização de sinais de vibração utilizando a emd. In: XIV CONEMI. **Congresso Nacional de Engenharia Mecânica e Industrial**. Salvador-BA, 2014.
- ROY, N. et al. Evidence-based clinical voice assessment: a systematic review. **American Journal of Speech-Language Pathology**, ASHA, v. 22, n. 2, p. 212–226, 2013.
- SANTOS, M. O. **Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet**. Dissertação (Mestrado em Engenharia Elétrica) — Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa-PB, 2015.
- SHRIVASTAV, R.; CAMACHO, A. A computational model to predict changes in breathiness resulting from variations in aspiration noise level. **Journal of Voice**, Elsevier, v. 24, n. 4, p. 395–405, 2010.
- SILVA, A. F. T. et al. Classificação de sinais de voz através da aplicação da transformada wavelet packet e redes neurais artificiais. **Revista Principia**, v. 1, n. 37, p. 34–41, 2017.
- SILVA, C. S. d.; BARBOSA, P. S. F. **Algoritmo do gradiente conjugado escalonado em RNA para a previsão do consumo horário de água em sistemas de abastecimento**. 2002. Disponível em: <<http://repositorio.furg.br>>. Acesso em: 03 de março de 2018.
- STOJANOVIĆ, M. et al. Understanding sensitivity, specificity and predictive values. **Vojnosanitetski pregled**, v. 71, n. 11, p. 1062–1065, 2014.

STRÁNÍK, A.; ČMEJLA, R.; VOKŘÁL, J. Acoustic parameters for classification of breathiness in continuous speech according to the grbas scale. **Journal of Voice**, Elsevier, v. 28, n. 5, p. 653–e9, 2014.

TEAGER, H. M.; TEAGER, S. M. Evidence for nonlinear sound production mechanisms in the vocal tract. **Speech Production and Speech Modeling**, Kluwer Academic, p. 241–261, 1990.

TITZE, I.; BAKER, R.; HERZL, H. Evidence of chaos in vocal fold vibration. **Vocal Fold Physiology, Frontiers in Basic Science**, Singular Pub., p. 143–188, 1993.

ULOZA, V. et al. Categorizing normal and pathological voices: automated and perceptual categorization. **Journal of Voice**, Elsevier, v. 25, n. 6, p. 700–708, 2011.

VERDOLINI, K.; ROSEN, C. A.; BRANSKI, R. C. **Classification manual for voice disorders-I**. [S.l.]: Psychology Press, 2014.

VIEIRA, M. N.; ROSA, L. L. C. Avaliação acústica na prática fonoaudiológica. In: PINHO S. M. R.; TSUJI, D. H.; BOHADANA, S. C. **Fundamentos em laringologia e voz**. [S.l.]: Revinter, 2006. v. 1, n. 1, p. 33–52.

VIEIRA, V. J. D. **Avaliação de Distúrbios da Voz por meio de Análise de Quantificação de Recorrência**. Dissertação (Mestrado em Engenharia Elétrica) — Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa-PB, 2014.

WANG, Z. et al. Automatic assessment of pathological voice quality using multidimensional acoustic analysis based on the grbas scale. **Journal of Signal Processing Systems**, Springer, v. 82, n. 2, p. 241–251, 2016.

YAMASAKI, R. et al. Correspondência entre escala analógico-visual ea escala numérica na avaliação perceptivo-auditiva de vozes. In: XVI CONGRESSO BRASILEIRO DE FONOAUDIOLOGIA. **Sociedade Brasileira de Fonoaudiologia**. [S.l.], 2008. p. 24–27.

YU, D.; CHENG, J.; YANG, Y. Application of emd method and hilbert spectrum to the fault diagnosis of roller bearings. **Mechanical systems and signal processing**, Elsevier, v. 19, n. 2, p. 259–270, 2005.

ZAGO, G. T. et al. Utilização do método da decomposição de modo empírico para a remoção de ruído em sinais de eletrocardiograma. **Revista Ifes Ciência-ISSN 2359-4799**, v. 1, n. 1, 2015.

ZHANG, X.; LAI, K. K.; WANG, S.-Y. A new approach for crude oil price analysis based on empirical mode decomposition. **Energy economics**, Elsevier, v. 30, n. 3, p. 905–918, 2008.

APÊNDICE A – A CLASSIFICAÇÃO SEGUNDO A ESCALA EAV PARA OS SINAIS SELECIONADOS DA BASE DE DADOS LIEV/UFPB

Nas Tabelas A.1 a A.7, são apresentados os valores obtidos pela avaliação perceptivo-auditiva, realizada pelo LIEV/UFPB, referente aos sinais escolhidos para esta pesquisa nos dois casos de estudo.

Tabela A.1 – Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes normais.

VOZES NORMAIS	EAV (0mm–100mm)			EAV (Grau)		
	RUG	SOP	TEN	RUG	SOP	TEN
VGE6	10,00	2,00	29,00	1	1	1
VGE20	20,00	26,50	17,00	1	1	1
VGE27	27,00	35,50	6,00	1	1	1
VGE32	4,00	22,00	28,00	1	1	1
VGE55	27,50	36,00	11,50	1	2	1
VGE61	17,00	3,00	25,00	1	1	1
VGE71	21,00	2,50	17,00	1	1	1
VGE91	8,00	2,00	19,00	1	1	1
VGE93	36,00	5,00	21,00	2	1	1
VGE104	4,50	2,00	12,00	1	1	1
VGE113	14,50	2,50	26,00	1	1	1
VGE123	21,50	2,50	27,00	1	1	1
VGE170	34,00	17,00	23,00	1	1	1
VGE259	34,00	4,50	20,00	1	1	1
VGE266	16,50	2,00	23,00	1	1	1
VGE272	22,50	4,00	16,00	1	1	1
VGE273	25,00	13,00	10,00	1	1	1
VGE279	35,00	5,00	32,00	1	1	1
VGE282	33,00	4,00	27,00	1	1	1
VGE287	32,00	3,00	29,00	1	1	1
VGE303	34,00	4,00	10,00	1	1	1
VGE320	20,00	6,00	27,00	1	1	1
VGE321	19,50	6,00	28,00	1	1	1
VGE356	21,50	11,50	24,00	1	1	1
VGE374	34,00	3,00	31,00	1	1	1
VGE400	31,00	7,00	24,00	1	1	1
VGE458	7,00	5,00	20,00	1	1	1
VGE468	2,00	3,00	21,00	1	1	1
VGE484	32,50	27,00	23,00	1	1	1
VGE490	33,00	3,00	27,50	1	1	1

Tabela A.2 – Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes rugosas.

VOZES RUGOSAS	EAV (0mm–100mm)			EAV (Grau)		
	RUG	SOP	TEN	RUG	SOP	TEN
VGE1	79,00	48,00	63,00	3	2	3
VGE2	66,00	58,00	42,00	3	3	2
VGE3	47,00	38,00	12,00	2	2	1
VGE5	42,00	39,00	33,00	2	2	1
VGE9	51,50	41,00	37,00	3	2	2
VGE11	94,00	68,00	54,00	4	3	3
VGE13	54,00	48,00	44,00	3	2	2
VGE16	41,00	36,50	10,50	2	2	1
VGE17	40,00	30,00	37,00	2	1	2
VGE18	51,00	45,00	39,00	3	2	2
VGE23	38,50	26,00	35,00	2	1	1
VGE25	52,00	47,00	33,00	3	2	1
VGE28	66,00	50,00	61,00	3	2	3
VGE29	92,50	66,00	78,00	4	3	3
VGE34	53,50	49,00	43,00	3	2	2
VGE39	42,00	37,00	20,00	2	2	1
VGE41	50,00	45,00	38,00	2	2	2
VGE43	52,50	32,50	46,00	3	1	2
VGE53	59,50	53,00	38,50	3	3	2
VGE56	58,00	52,00	33,00	3	3	1
VGE75	62,00	52,50	49,00	3	3	2
VGE76	55,50	49,50	30,50	3	2	1
VGE80	61,50	52,00	46,00	3	3	2
VGE96	58,50	40,00	37,00	3	2	2
VGE 97	78,00	69,00	69,50	3	3	3
VGE 98	93,00	67,50	60,00	4	3	3
VGE 106	39,50	06,00	35,00	2	1	1
VGE 107	45,00	38,00	33,50	2	2	1
VGE 108	48,50	07,00	19,50	2	1	1

Tabela A.3 – Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes soprosas.

VOZES SOPROSAS	EAV (0mm–100mm)			EAV (Grau)		
	RUG	SOP	TEN	RUG	SOP	TEN
VGE10	31,00	46,00	14,00	1	2	1
VGE12	50,00	64,00	57,00	2	3	3
VGE15	42,50	50,00	10,00	2	2	1
VGE19	36,00	41,50	04,00	2	2	1
VGE22	40,50	45,50	37,00	2	2	2
VGE36	47,00	57,00	31,00	2	3	1
VGE45	62,00	72,00	57,50	3	3	3
VGE49	52,00	60,00	22,00	3	3	1
VGE58	18,00	39,50	33,00	1	2	1
VGE65	39,00	46,00	20,50	2	2	1
VGE68	44,00	49,50	10,00	2	2	1
VGE78	36,00	47,00	07,50	2	2	1
VGE84	18,00	38,00	04,00	1	2	1
VGE94	28,00	40,00	06,00	1	2	1
VGE109	35,50	41,00	19,00	1	2	1
VGE114	38,50	50,50	11,50	2	2	1
VGE118	43,00	51,00	36,00	2	3	2
VGE119	41,00	48,50	35,00	2	2	1
VGE121	68,00	76,00	43,00	3	3	2
VGE125	31,00	44,00	03,00	1	2	1
VGE159	32,50	43,00	15,00	1	2	1
VGE166	36,50	52,00	05,00	2	3	1
VGE172	45,00	55,00	19,00	2	3	1
VGE178	38,00	49,00	08,00	2	2	1
VGE227	47,50	56,00	34,00	2	3	1
VGE304	37,00	52,00	05,00	2	3	1
VGE319	70,00	96,00	83,50	3	4	3
VGE327	55,50	65,50	13,00	3	3	1

Tabela A.4 – Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes tensas.

VOZES TENSAS	EAV (0mm–100mm)			EAV (Grau)		
	RUG	SOP	TEN	RUG	SOP	TEN
VGE4	36,00	11,00	41,00	2	1	2
VGE14	57,00	10,00	64,00	3	1	3
VGE21	57,00	07,00	64,00	3	1	3
VGE30	55,50	61,50	68,00	3	3	3
VGE31	40,00	28,50	46,50	2	1	2
VGE37	40,00	05,00	46,50	2	1	2
VGE38	38,00	16,00	43,00	2	1	2
VGE42	61,00	03,50	93,00	3	1	4
VGE60	45,50	03,00	52,00	2	1	3
VGE70	46,00	18,50	54,00	2	1	3
VGE81	43,00	05,00	55,50	2	1	3
VGE148	45,00	27,00	50,00	2	1	2
VGE156	50,00	03,00	61,00	2	1	3
VGE180	66,00	05,00	80,00	3	1	3
VGE200	56,00	03,00	64,00	3	1	3
VGE248	44,00	06,00	55,00	2	1	3
VGE284	44,50	08,00	52,00	2	1	3
VGE299	39,50	08,00	48,50	2	1	2
VGE312	39,00	03,00	47,00	2	1	2
VGE530	43,50	05,00	54,50	2	1	3
VGE561	41,00	09,00	54,00	2	1	3
VGE583	36,00	06,50	42,50	2	1	2
VGE607	44,00	37,00	51,50	2	2	3
VGE654	37,00	04,00	45,00	2	1	2
VGE711	37,00	04,00	44,00	2	1	2
VGE719	37,00	08,00	50,00	2	1	2
VGE722	39,50	04,50	54,00	2	1	3
VGE736	85,00	69,50	96,00	3	3	4
VGE763	48,00	06,00	48,50	2	1	2

Tabela A.5 – Classificação segundo a Escala Analógica Visual (EAV) para sinais de vozes com predominância de grau 2.

PREDOMINANTE GRAU 2 (G2)	EAV (0mm-100mm)			TIPO DE DESVIO
	RUG	SOP	TEN	
VGE3	47,00	38,00	12,00	RUG
VGE5	42,00	39,00	33,00	RUG
VGE16	41,00	36,50	10,50	RUG
VGE17	40,00	30,00	37,00	RUG
VGE23	38,50	26,00	35,00	RUG
VGE39	42,00	37,00	20,00	RUG
VGE41	50,00	45,00	38,00	RUG
VGE50	36,00	03,50	32,50	RUG
VGE106	39,50	06,00	35,00	RUG
VGE107	45,00	38,00	33,50	RUG
VGE10	31,00	46,00	14,00	SOP
VGE15	42,50	50,00	10,00	SOP
VGE19	36,00	41,50	04,00	SOP
VGE22	40,50	45,00	37,00	SOP
VGE54	41,00	46,00	38,00	SOP
VGE78	36,00	47,00	07,50	SOP
VGE58	18,00	39,50	33,00	SOP
VGE65	39,00	46,00	20,50	SOP
VGE68	44,00	49,50	10,00	SOP
VGE94	28,00	40,00	06,00	SOP
VGE4	36,00	11,00	41,00	TEN
VGE31	40,00	28,50	46,50	TEN
VGE37	40,00	05,00	46,50	TEN
VGE38	38,00	16,00	43,00	TEN
VGE148	45,00	27,00	50,00	TEN
VGE299	39,50	08,00	48,50	TEN
VGE583	36,00	06,50	42,50	TEN
VGE711	37,00	04,00	44,00	TEN
VGE719	37,00	08,00	50,00	TEN
VGE763	48,00	06,00	48,50	TEN

Tabela A.6 – Classificação segundo a Escala Analógica Visual (EAV) para sinais de vozes com predominância de grau 3.

PREDOMINANTE GRAU 3 (G3)	EAV (0mm-100mm)			TIPO DE DESVIO
	RUG	SOP	TEN	
VGE1	79,00	48,00	63,00	RUG
VGE2	66,00	58,00	42,00	RUG
VGE9	51,50	41,00	37,00	RUG
VGE13	54,00	48,00	44,00	RUG
VGE18	51,00	45,00	39,00	RUG
VGE25	52,00	47,00	33,00	RUG
VGE28	66,00	50,00	61,00	RUG
VGE53	59,50	53,00	38,50	RUG
VGE75	62,00	52,50	49,00	RUG
VGE76	55,50	49,50	30,50	RUG
VGE80	61,50	52,00	46,00	RUG
VGE96	58,50	40,00	37,00	RUG
VGE12	50,00	64,00	57,00	SOP
VGE36	47,00	57,00	31,00	SOP
VGE45	62,00	72,00	57,50	SOP
VGE322	48,00	59,00	12,00	SOP
VGE327	55,50	65,50	13,00	SOP
VGE121	68,00	76,00	43,00	SOP
VGE172	45,00	55,00	19,00	SOP
VGE118	43,00	51,00	36,00	SOP
VGE14	57,00	10,00	64,00	TEN
VGE21	57,00	07,00	64,00	TEN
VGE30	55,50	61,50	68,00	TEN
VGE60	45,50	03,00	52,00	TEN
VGE70	46,00	18,50	54,00	TEN
VGE81	43,00	05,00	55,50	TEN
VGE156	50,00	03,00	61,00	TEN
VGE180	66,00	05,00	80,00	TEN
VGE200	56,00	03,00	64,00	TEN
VGE248	44,00	06,00	55,00	TEN

Tabela A.7 – Classificação segundo a Escala Analógica Visual (EAV) para sinais de vozes da combinação dos graus 2 e 3.

COMBINAÇÃO G2G3	EAV (0mm-100mm)			TIPO DE DESVIO	GRAU PREDOMINANTE
	RUG	SOP	TEN		
VGE3	47,00	38,00	12,00	RUG	2
VGE5	42,00	39,00	33,00	RUG	2
VGE16	41,00	36,50	10,50	RUG	2
VGE17	40,00	30,00	37,00	RUG	2
VGE23	38,50	26,00	35,00	RUG	2
VGE10	31,00	46,00	14,00	SOP	2
VGE15	42,50	50,00	10,00	SOP	2
VGE19	36,00	41,50	04,00	SOP	2
VGE22	40,50	45,00	37,00	SOP	2
VGE54	41,00	46,00	38,00	SOP	2
VGE4	36,00	11,00	41,00	TEN	2
VGE31	40,00	28,50	46,50	TEN	2
VGE37	40,00	05,00	46,50	TEN	2
VGE38	38,00	16,00	43,00	TEN	2
VGE100	36,00	09,00	40,00	TEN	2
VGE1	79,00	48,00	63,00	RUG	3
VGE2	66,00	58,00	42,00	RUG	3
VGE9	51,50	41,00	37,00	RUG	3
VGE13	54,00	48,00	44,00	RUG	3
VGE18	51,00	45,00	39,00	RUG	3
VGE12	50,00	64,00	57,00	SOP	3
VGE36	47,00	57,00	31,00	SOP	3
VGE45	62,00	72,00	57,50	SOP	3
VGE322	48,00	59,00	12,00	SOP	3
VGE327	55,50	65,50	13,00	SOP	3
VGE14	57,00	10,00	64,00	TEN	3
VGE21	57,00	07,00	64,00	TEN	3
VGE30	55,50	61,50	68,00	TEN	3
VGE60	45,50	03,00	52,00	TEN	3
VGE70	46,00	18,50	54,00	TEN	3