

**INSTITUTO FEDERAL DE EDUCAÇÃO,
CIÊNCIA E TECNOLOGIA DA PARAÍBA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA -PPGEE**

MARIA DE FÁTIMA KALLYNNA BEZERRA COURAS

**CLASSIFICAÇÃO DE DESVIOS VOCAIS UTILIZANDO
CARACTERÍSTICAS BASEADAS NO MODELO LINEAR
DE PRODUÇÃO DA FALA**

João Pessoa, PB
Fevereiro de 2017

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

C658c Couras, Maria de Fátima Kallynna Bezerra.
Classificação de desvios vocais utilizando características baseadas no modelo linear de produção da fala / Maria de Fátima Kallynna Bezerra Couras. – 2017.
115 f. : il.
Dissertação (Mestrado em Engenharia Elétrica) – Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB / Coordenação de Pós-Graduação em Engenharia Elétrica, 2017.
Orientadora: Prof^ª. Silvana Luciene N. Cunha Costa.
1. Redes neurais. 2. Desvios vocais. 3. Sinais de voz.
I. Título.

CDU 621.391

MARIA DE FÁTIMA KALLYNNA BEZERRA COURAS

**CLASSIFICAÇÃO DE DESVIOS VOCAIS UTILIZANDO
CARACTERÍSTICAS BASEADAS NO MODELO LINEAR
DE PRODUÇÃO DA FALA**

Dissertação de Mestrado submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba como requisito necessário à obtenção do grau de Mestre em Engenharia Elétrica.

Orientadora: Prof. Dra. Silvana Luciene do Nascimento Cunha Costa
Coorientadora: Prof. Dra. Suzete Élide Nóbrega Correia

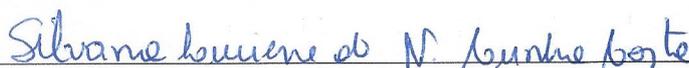
João Pessoa, PB
Fevereiro de 2017

MARIA DE FÁTIMA KALLYNNA BEZERRA COURAS

**CLASSIFICAÇÃO DE DESVIOS VOCAIS UTILIZANDO
CARACTERÍSTICAS BASEADAS NO MODELO LINEAR
DE PRODUÇÃO DA FALA**

Dissertação aprovada pela Banca Examinadora em: 24/02/2017.

BANCA EXAMINADORA



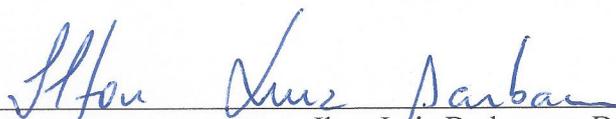
Silvana Luciene do Nascimento Cunha Costa, Dra. – IFPB
Orientadora



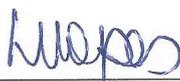
Suzete Elida Nobrega Correia, Dra. – IFPB
Coorientadora



Daniella Dias Cavalcante da Silva, Dra. – IFPB
Membro da Banca



Ilton Luiz Barbacena, Dr. – IFPB
Membro da Banca



Leonardo Wanderley Lopes, Dr. – UFPB
Membro da Banca

João Pessoa – PB, Fevereiro de 2017

© Maria de Fátima Kallynna Bezerra Couras – kallynna.mary@gmail.com

“O sucesso não tem a ver com o lugar de onde você veio, e sim com a confiança que você tem e o esforço que você está disposto a investir.”

Michelle Obama

AGRADECIMENTOS

- ❖ A Deus, que permitiu que tudo isso acontecesse, ao longo de minha vida, e não somente nestes anos como mestranda, mas que em todos os momentos é o maior mestre que alguém pode conhecer;
- ❖ Ao Instituto Federal de Educação, Ciência e Tecnologia da Paraíba - IFPB, seu corpo docente, direção e administração que oportunizaram a janela que hoje vislumbro. Além do apoio financeiro ofertado pela COPEX, onde fui bolsista durante o mestrado;
- ❖ À minha Orientadora, professora Silvana Costa, por dedicar grande parte do seu tempo para compartilhar comigo seus preciosos conhecimentos, tanto na área acadêmica, como na vida pessoal;
- ❖ À minha Coorientadora Suzete Correia, por ter compartilhado também grande parte do seu tempo na transmissão de conhecimentos e valores essenciais no processo de formação profissional;
- ❖ Aos professores examinadores Leonardo Lopes, Daniella Dias e Ilton Barbacena que fizeram parte das minhas bancas de qualificação e de defesa de dissertação, pelas valiosas sugestões, que contribuíram para o enriquecimento deste trabalho de pesquisa;
- ❖ À minha família, em especial aos meus pais Francisco Bezerra (*in memoriam*) e Pastoura Gomes aos meus irmãos Carlos Alberto e Vicente Bezerra, ao meu esposo Pablo Henrique, as minhas cunhadas Edina Alves e Célia Lima, aos meus sobrinhos e sobrinhas Ianca Lavine, Karla Vitória, Carlos Daniel e Igo Alexandre pelo apoio incondicional;
- ❖ Aos meus amigos, em especial Layne Leal, Ellen Castro, Alisson Cavalcante, Pedro Lucas, Gabriel Gonçalves e Monalisa Serafim, que foram meus irmãos na amizade e que vão continuar presentes em minha vida.
- ❖ Aos amigos do programa, em especial Evandson Dantas, Nairon Vasconcelos, Moisés Hamsés, Giulliana Lacerda, Gilberto Sobreira, Andrécia Pereira, Laís Felix e Gilliane Oliveira, que foram meus companheiros de trabalho e irmãos na amizade que fizeram parte da minha formação e que vão continuar presentes em minha vida.

RESUMO

A avaliação perceptivo-auditiva tem papel fundamental na avaliação da qualidade vocal. No entanto, por ser uma avaliação subjetiva, está sujeita a imprecisões e variações, sendo necessária a utilização de técnicas que tragam maior confiabilidade aos resultados. A análise acústica surge como uma ferramenta que proporciona a avaliação da qualidade vocal de forma objetiva. Neste trabalho, são empregadas técnicas de processamento digital de sinais, baseadas no modelo linear de produção da fala, para analisar a qualidade vocal. É avaliado o desempenho de medidas tradicionalmente empregadas na análise acústica, tais como frequência fundamental, medidas de perturbação (*jitter* e *shimmer*), GNE (*Glottal to Noise Excitation Ratio*) e frequências formânticas. Também é avaliado o potencial discriminativo dos coeficientes da análise de predição linear (*Linear Predictive Coding*- LPC), coeficientes cepstrais e mel-cepstrais na classificação de desvios vocais (rugosidade, sopro e tensão). Com o auxílio de um classificador, baseado em redes neurais artificiais MLP (*Multilayer Perceptron*), é realizada a classificação dos sinais utilizando as medidas extraídas individualmente e de forma combinada. Foram obtidas taxas de classificação de 86% na discriminação entre vozes soprosas e vozes saudáveis.

Palavras-chave: Processamento de Sinais de Voz, Modelo Linear de Produção da Fala, Desvios vocais, Redes Neurais.

ABSTRACT

The perceptual-auditory evaluation plays a fundamental role in the evaluation of vocal quality. However, because it is a subjective technique, it is subject to inaccuracies and variations, being necessary to make technical evaluations that bring greater reliability to the results. Acoustic analysis appears as a tool that provides an objective evaluation of the vocal quality. In this work, digital signal processing techniques, based on the linear model of speech production, are used to analyze vocal quality. The performance of measures traditionally employed in acoustic analysis, such as fundamental frequency, jitter and shimmer measurements, GNE (Glottal to Noise Excitation Ratio) and formant frequencies are analyzed. The discriminatory potential of the Linear Predictive Coding (LPC) coefficients, the coefficients of the cepstral and mel-cepstral in the classification of vocal deviations (roughness, breathiness and strain) is also evaluated. With the aid of a classifier based on neural networks (Multilayer Perceptron) the classification of the signals is performed using the measurements individually extracted and in combination. Classification rates of 86% were obtained in the discrimination between breathy and healthy voices.

Keywords: Speech Signal Processing, Linear Model of Speech Production, voice deviations, Neural Networks.

LISTA DE FIGURAS

Figura 1 – Sistema de produção da voz humana.	27
Figura 2 - Modelo simplificado do mecanismo de produção da fala.	29
Figura 3 - Modelo linear discreto no tempo para a produção da fala.	30
Figura 4 – AMDF para vogal sustentada /É/ (“é”).	31
Figura 5 – Fluxograma apresentando a metodologia de extração do GNE.	34
Figura 6 - Comparação dos espectros LPC e FFT para um sinal de voz.	37
Figura 7 – Frequências formânticas para a vogal sustentada /É/ (“é”). F_1 representa 1º formante, F_2 o 2º formante e F_3 o 3º formante.	38
Figura 8 - Banco de filtros digitais na escala mel.	41
Figura 9 - Diagrama em Blocos da Metodologia empregada.	48
Figura 10 – Frequência Fundamental por segmento para sinal de voz: (a) rugoso (b) tenso (c) soproso e (d) saudável.	49
Figura 11 - Representação do <i>Codebook</i>	52
Figura 12 – Espectro da Frequência formânticas para cada segmento do sinal de voz: (a) rugoso, (b) soproso, (c) tenso, e (d) saudável.	54
Figura 13 - Rede <i>feedforward</i> MLP utilizada.	55
Figura 14 – Comportamento da frequência fundamental para vozes saudáveis e vozes desviadas.	58
Figura 15 - <i>Boxplots</i> da frequência fundamental para sinais de voz rugosa (RUG), soproso (SOP), tenso (TEN) e saudável (SDL).	59
Figura 16 - <i>Boxplots</i> da frequência fundamental para os sinais de vozes desviadas (DES) e saudáveis (SDL).	59
Figura 17 - Comportamento das medidas: (a) <i>jitter</i> (b) <i>shimmer</i> e (c) GNE para vozes saudáveis e vozes desviadas.	60
Figura 18 - <i>Boxplots</i> do (a) <i>jitter</i> , (b) <i>shimmer</i> e (c) GNE para sinais de voz rugosa (RUG), soproso (SOP), tenso (TEN) e saudável (SDL).	61
Figura 19 - <i>Boxplots</i> do (a) <i>jitter</i> , (b) <i>shimmer</i> e (c) GNE para sinais de voz desviadas (DES) e saudável (SDL).	62
Figura 20 - Comportamento das medidas: (a) F_{1med} , (b) F_{2med} e (c) F_{3med} para vozes saudáveis e vozes desviadas.	62
Figura 21 - <i>Boxplots</i> (a) F_{1med} , (b) F_{2med} e (c) F_{3med} para sinais de voz rugosa (RUG), soproso (SOP), tenso (TEN) e saudável (SDL).	63

Figura 22 - <i>Boxplots</i> (a) F_{1med} , (b) F_{2med} e (c) F_{3med} para sinais de voz desviada (DES) e saudável (SDL).....	64
Figura 23 Comparação dos resultados obtidos com as medidas individuais, combinadas, resultados obtidos por Lopes <i>et al.</i> (2016) e por Costa (2008).....	69
Figura 24 - Comportamento da distorção para vozes saudáveis e vozes desviadas para os diversos números de níveis com os coeficientes LPC: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.....	71
Figura 25 - Comportamento da distorção para vozes rugosas (RUG), tensas (TEN), soprosa (SOP) e saudáveis (SDL) para os diversos número de níveis com os coeficientes LPC: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.	72
Figura 26 - Comportamento da distorção para vozes desviadas (DES) e saudáveis (SDL) para os diversos números de níveis com os coeficientes LPC: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.....	73
Figura 27 - Comportamento da distorção para vozes saudáveis e vozes desviadas para os diversos números de níveis com os coeficientes cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.....	74
Figura 28 - Comportamento da distorção para vozes rugosas (RUG), tensas (TEN), soprosa (SOP) e saudáveis (SDL) para os diversos números de níveis com os coeficientes cepstrais: (a) 16; (b) 32; (c) 64; (d) 128 níveis; (e) 256; (f) 512 e (g) 1024.....	75
Figura 29 Comportamento da distorção para vozes desviadas (DES) e saudáveis (SDL) para os diversos números de níveis com os coeficientes cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.....	76
Figura 30 - Comportamento da distorção para vozes saudáveis e vozes desviadas para os diversos números de níveis com os coeficientes mel-cepstrais. (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.	77
Figura 31 – Comportamento da distorção para vozes rugosas (RUG), tensas (TEN), soprosa (SOP) e saudáveis (SDL) para diversos números de níveis com os coeficientes mel-cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.....	78
Figura 32 – Comportamento da distorção para vozes desviadas (DES) e saudáveis (SDL) para diversos números de níveis, com os coeficientes mel-cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.	79
Figura 33 - Comportamento da curva ROC com as distorções dos coeficientes LPC para cada número de níveis de quantização. (a) vozes Rugosas Vs. Saudáveis; (b) vozes Tensas Vs.	

Saudáveis; (c) vozes Soprosas Vs. Saudáveis; (d) vozes Desviadas Vs. Saudáveis; (e) vozes Rugosas Vs. Tensas; (f) vozes Rugosas Vs. Soprosas e (g) vozes Tensas Vs. Soprosas..... 82

Figura 34 - Comportamento da curva ROC com as distorções dos coeficientes cepstrais para cada número de níveis de quantização. (a) vozes Rugosas Vs. Saudáveis; (b) vozes Tensas Vs. Saudáveis; (c) vozes Soprosa Vs. Saudáveis; (d) vozes Desviadas Vs. Saudáveis; (e) vozes Rugosas Vs. Tensas; (f) vozes Rugosas Vs. Soprosas e (g) vozes Tensas Vs. Soprosas..... 83

Figura 35 - Comportamento da curva ROC com as distorções dos coeficientes mel-cepstrais para cada nível de quantização. (a) vozes Rugosas Vs. Saudáveis; (b) vozes Tensas Vs. Saudáveis; (c) vozes Soprosa Vs. Saudáveis; (d) vozes Desviadas Vs. Saudáveis; (e) vozes Rugosas Vs. Tensas; (f) vozes Rugosas Vs. Soprosas e (g) vozes Tensas Vs. Soprosas..... 84

Figura 36 - Comparação entre os resultados obtidos com as medidas tradicionais e com as medidas dos coeficientes LPC, cepstrais e mel-cepstrais..... 91

LISTA DE TABELAS

Quadro I - Estado da Arte.....	43
Tabela A 1 – Classificação segundo a Escala Analógica-Visual (EAV) para os sinais de vozes rugosas.....	101
Tabela A 2 – Classificação segundo a Escala Analógica-Visual (EAV) para os sinais de vozes tensas.	102
Tabela A 3 - Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes soprosas.	103
Tabela B 1 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Rugosas Vs. Saudáveis.....	104
Tabela B 2 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Tensas Vs. Saudáveis.	104
Tabela B 3 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Soprosas Vs. Saudáveis.	104
Tabela B 4 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Desviadas Vs. Saudáveis.	104
Tabela B 5 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Rugosas Vs. Tensas.	105
Tabela B 6 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Rugosas Vs. Soprosas.....	105
Tabela B 7 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Tensas Vs. Soprosas.	105
Tabela B 8 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Rugosas Vs. Saudáveis.....	105
Tabela B 9 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Tensas Vs. Saudáveis.	106
Tabela B 10 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Soprosas Vs. Saudáveis.	106
Tabela B 11 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Desviadas Vs. Saudáveis.....	106
Tabela B 12 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Rugosas Vs. Tensas.....	106

Tabela B 13 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Rugosas Vs. Soprosas.....	106
Tabela B 14 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Tensas Vs. Soprosas.	107
Tabela B 15 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Rugosas Vs. Saudáveis.	107
Tabela B 16 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Tensas Vs. Saudáveis.	107
Tabela B 17 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Soprosas Vs. Saudáveis.	107
Tabela B 18 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Desviadas Vs. Saudáveis.	108
Tabela B 19 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Rugosas Vs. Tensas.	108
Tabela B 20 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Rugosas Vs. Soprosas.....	108
Tabela B 21 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Tensas Vs. Soprosas.	108
Tabela C 1 - Classificação com os <i>codebooks</i> LPC (Classificador MLP) – Vozes Rugosas Vs. Saudáveis.	109
Tabela C 2 - Classificação com os <i>codebooks</i> LPC (Classificador MLP) – Vozes Tensas Vs. Saudáveis.	109
Tabela C 3 - Classificação com os <i>codebooks</i> LPC (Classificador MLP) – Vozes Soprosas Vs. Saudáveis.	109
Tabela C 4 - Classificação com os <i>codebooks</i> LPC (Classificador MLP) – Vozes Desviadas Vs. Saudáveis.	109
Tabela C 5 - Classificação com os <i>codebooks</i> LPC (Classificador MLP) – Vozes Rugosas Vs. Tensas.	109
Tabela C 6 - Classificação com os <i>codebooks</i> LPC (Classificador MLP) – Vozes Rugosas Vs. Soprosas.....	110
Tabela C 7 - Classificação com os <i>codebooks</i> LPC (Classificador MLP) – Vozes Tensas Vs. Soprosas.....	110
Tabela C 8 - Classificação com os <i>codebooks</i> cepstrais (Classificador MLP) – Vozes Rugosas Vs. Saudáveis.	110

Tabela C 9 - Classificação com os <i>codebooks</i> cepstrais (Classificador MLP) – Vozes Tensas Vs. Saudáveis.	111
Tabela C 10 - Classificação com os <i>codebooks</i> cepstrais (Classificador MLP) – Vozes Soprosas Vs. Saudáveis.	111
Tabela C 11 - Classificação com os <i>codebooks</i> cepstrais (Classificador MLP) – Vozes Desviadas Vs. Saudáveis.	111
Tabela C 12 - Classificação com os <i>codebooks</i> cepstrais (Classificador MLP) – Vozes Rugosas Vs. Tensas.	111
Tabela C 13 - Classificação com os <i>codebooks</i> cepstrais (Classificador MLP) – Vozes Rugosas Vs. Soprosas.	111
Tabela C 14 - Classificação com os <i>codebooks</i> cepstrais (Classificador MLP) – Vozes Tensas Vs. Soprosas.	112
Tabela C 15 - Classificação com os mel-cepstrais (Classificador MLP) – Vozes Rugosas Vs. Saudáveis.	112
Tabela C 16 - Classificação com os <i>codebooks</i> mel-cepstrais (Classificador MLP) – Vozes Tensas Vs. Saudáveis.	112
Tabela C 17 - Classificação com os <i>codebooks</i> mel-cepstrais (Classificador MLP) – Vozes Soprosas Vs. Saudáveis.	112
Tabela C 18 - Classificação com os <i>codebooks</i> mel-cepstrais (Classificador MLP) – Vozes Desviadas Vs. Saudáveis.	113
Tabela C 19 - Classificação com os <i>codebooks</i> mel-cepstrais (Classificador MLP) – Vozes Rugosas Vs. Tensas.	113
Tabela C 20 - Classificação com os <i>codebooks</i> mel-cepstrais (Classificador MLP) – Vozes Rugosas Vs. Soprosas.	113
Tabela C 21 - Classificação com os <i>codebooks</i> mel-cepstrais (Classificador MLP) – Vozes Tensas Vs. Soprosas.	113
Tabela D 1 - Classificação com a média dos coeficientes LPC (LPC_{med}) realizada pelo Classificador MLP.	114
Tabela D 2 - Classificação com a média dos coeficientes LPC (LPC_{med}) realizada pelo Classificador MLP.	114
Tabela D 3 - Classificação com a média dos coeficientes cepstrais ($CEPS_{med}$) realizada pelo Classificador MLP.	114
Tabela D 4 - Classificação com a média dos coeficientes cepstrais ($CEPS_{med}$) realizada pelo Classificador MLP.	114

Tabela D 5 - Classificação com a média dos coeficientes mel-cepstrais (MEL_{med}) realizada pelo Classificador MLP.....	115
Tabela D 6 - Classificação com a média dos coeficientes mel-Cepstrais (MEL_{med}) realizada pelo Classificador MLP.....	115

LISTA DE ABREVIATURAS E SIGLAS

ACUR – Acurácia

AMDF – *Average Magnitude Difference Function*

APQ – *Amplitude Perturbation Quotient*

CA – Correta aceitação

CHNR – *Cepstrum based Harmonics-to-Noise Ratio*

CR – Correta rejeição

EAV – Escala Analógica Visual

ESP - Especificidade

FA - Falsa aceitação

FFT - *Fast Fourier Transform*

FR – Falsa rejeição

G1 – Grau normal

G2 – Grau leve

G3 – Grau moderado

G4 – Grau intenso

GNE – *Glottal to Noise Excitation Ratio*

GRBASI – G: grau geral do distúrbio, R: *roughness*, B: *breathness*, A: *asteny*, S: *strain* e I: *instability*

HNR – *Harmonics-to-Noise Ratio*

LIEV – Laboratório Integrado de Estudos da Voz

LBG – *Linde-Buzo-Gray*

LPC – *Linear Predictive Coding*

MFCC – *Mel-frequency Cepstral Coefficients*

MLP – *Multilayer Perceptron*

NFFT – Número de pontos da Transformada de Fourier

NNE - *Normalized Noise Energy*

FPQ – *Frequency Perturbation Quotient*

QV – Quantização Vetorial

RNA – Rede Neural Artificial

ROC - *Receive Operator Characteristic Curve*

RUG - Rugosa

SCG – *Scaled Conjugate Gradient*

SEN – Sensibilidade

SDL – Saudável

SOP – Soprosa

TEN – Tensa

UFPB – Universidade Federal da Paraíba

LISTA DE SÍMBOLOS

$a(k)$ – Coeficientes LPC

A_n, A_v – Ganho

A_i – Amplitude do impulso glotal

c_1, c_2, \dots, c_n – Conjunto de mapeamento

$c(n)$ – Coeficientes cepstrais

$c_{mel}(n)$ – Coeficientes mel-cepstrais

$CEPS$ – Coeficientes cepstrais

$CEPS_{med}$ – Média dos coeficientes cepstrais

Codebook CEPS – *Codebook* com os coeficientes cepstrais

Codebook LPC – *Codebook* com os coeficientes LPC

Codebook MEL – *Codebook* com os coeficientes mel-cepstrais

$d(v_N, c_N)$ – Distorção do erro médio quadrático

dB – Decibel

E – Energia do erro

$e(n)$ – Erro de estimação

f - Frequência

F_0 – Frequência fundamental

F_{0med} – Frequência fundamental média

F_1 – 1º Formante

F_{1med} - 1º Formante médio

F_2 – 2º Formante

F_{2med} - 2º Formante médio

F_3 – 3º Formante

F_{3med} - 3º Formante médio

f_a – Frequência de amostragem

F_{linear} – Frequência percebida

F_{mel} – Frequência mel

G - Ganho

$G(z)$ – Filtro digital

$H(z)$ – Função de transferência

k – Amostras atrasadas

LPC – Coeficientes LPC

LPC_{med} – Média dos coeficientes LPC

MEL – Coeficientes mel-cepstrais

MEL_{med} – Média dos coeficientes mel-cepstrais

mm – Milímetro

ms – Milissegundos

N – Amostras

N_f – Número de filtros

p – Ordem do preditor

R – Matriz de autocorrelação dos coeficientes LPC

$s(n)$ – Sinal

$\tilde{s}(n)$ – Sinal estimado

$S_{FFT}(k)$ – Sinal de saída dos filtros digitais

T_i – Período

$u(n)$ – Sinal de excitação

v_1, v_2, \dots, v_n – Dados de entrada

W_{ij}, W_{jk} – Pesos da rede

$W(j)$ – Janelas de ponderação

$X(j)$ – Espectro de magnitude

$x(n)$ – Sinal

SUMÁRIO

1. INTRODUÇÃO	22
1.1 Formulação do Problema	22
1.2 Justificativa	23
1.3 Objetivos	25
1.3.1 Geral.....	25
1.3.2 Específicos	25
1.4 Organização do Trabalho	26
2. CARACTERÍSTICAS DO MODELO LINEAR DE PRODUÇÃO VOCAL.....	27
2.1 Sistema de Produção da Voz Humana	27
2.2 Teoria Fonte-Filtro	28
2.3 Frequência Fundamental	30
2.4 Medidas de Perturbação	32
2.5 <i>Glottal to Noise Excitation Ratio</i> (GNE)	33
2.6 Análise de Voz por Predição Linear	35
2.6.1 Coeficientes LPC.....	35
2.6.2 Frequências formânticas.....	36
2.7 Análise Cepstral	39
2.8 Análise Mel-Cepstral.....	40
2.9 Avaliação Perceptivo-auditiva	42
2.10 Trabalhos Relacionados	43
3. MATERIAIS E MÉTODOS.....	46
3.1 Base de Dados – LIEV/UFPB	46
3.2 Metodologia	47
3.2.1 Frequência fundamental	48
3.2.2 Extração do <i>jitter</i>	49
3.2.3 Extração do <i>shimmer</i>	50
3.2.4 Extração do <i>glottal to noise excitation ratio</i> (GNE)	50

3.2.5 Extração dos coeficientes LPC, cepstrais e mel-cepstrais	51
3.2.6 Extração das frequências formânticas	53
3.2.7 Processo de Classificação	54
4. RESULTADOS.....	58
4.1 Caracterização Acústica dos Desvios Vocais com as Medidas Tradicionais.....	58
4.2 Resultados da Classificação dos Desvios Vocais com as Medidas Tradicionais	64
4.2.1 Discussão dos resultados com as medidas tradicionais	68
4.3 Resultados obtidos – Análise LPC, Cepstral e Mel-Cepstral.....	70
4.3.1 Análise LPC.....	70
4.3.2 Análise cepstral – coeficientes cepstrais	73
4.3.3 Análise mel-cepstral – coeficientes mel-cepstrais	77
4.4 Resultados da Classificação dos Desvios Vocais com as Análises LPC, Cepstral e Mel-Cepstral	80
4.4.1 Resultados obtidos com as curvas ROC	80
4.4.2 Resultados obtidos com classificador MLP.....	85
4.4.3 Discussão dos resultados com as análises LPC, cepstral e mel-cepstral	89
5. CONSIDERAÇÕES FINAIS	92
5.2 Contribuições da Pesquisa.....	93
5.3 Sugestão para Trabalhos Futuros	93
REFERÊNCIAS BIBLIOGRÁFICAS	95
APÊNDICES.....	101
A Classificação segundo a Escala Analógica-Visual (EAV) pra os sinais selecionados da base de dados LIEV/UFPB	101
A 1. Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes rugosas	101
A 2. Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes tensas.	102
A 3. Classificação segundo a Escala Analógica-Visual (EAV) para sinais de vozes soprosas	103
B Tabelas com a classificação realizada a partir da curva ROC pra as distorções com os métodos empregados	103
B 1. Tabelas com a classificação realizada a partir da curva ROC para as distorções com os coeficientes LPC	104

B 2. Tabelas com a classificação realizada a partir da curva ROC para as distorções com os coeficientes cepstrais.....	105
B 3. Tabelas com a classificação realizada a partir da curva ROC para as distorções com os coeficientes mel-cepstrais	107
C Tabelas apresentando a classificação realizada com o classificador MLP com os <i>codebooks</i> como características de entrada.....	108
C 1. Classificação realizada com o classificador MLP com os codebooks LPC como características de entrada	109
C 2. Classificação realizada com o classificador MLP com os codebooks Cepstrais como características de entrada.....	110
C 3. Classificação realizada com o classificador MLP com os codebooks mel-cepstrais como características de entrada.....	112
D Tabelas apresentando a classificação realizada com o classificador MLP com a média dos coeficientes como características de entrada	113
D 1. Classificação realizada com o classificador MLP com a média dos coeficientes LPC como características de entrada.....	114
D 2. Classificação realizada com o classificador MLP com a média dos coeficientes cepstrais como características de entrada.....	114
D 3. Classificação realizada com o classificador MLP com a média dos coeficientes mel-cepstrais como características de entrada.....	115

1. INTRODUÇÃO

A voz é um veículo de comunicação e traz informações sobre as características físicas, psicológicas e sociais, sendo considerada a principal ferramenta de comunicação humana. Ela contém informações a respeito do falante, tais como identidade, estado emocional e humor (BRANDI, 2002; VIEIRA, 2014).

A qualidade vocal é o termo empregado para designar o conjunto de características que identificam uma voz (BEHLAU, 2008). A avaliação da qualidade vocal possui duas formas de análise que se complementam: a avaliação perceptivo-auditiva e a análise acústica. A avaliação perceptivo-auditiva é realizada por um profissional treinado para ouvir e identificar características presentes no sinal de voz que indiquem se há alterações ou não na qualidade vocal (COLTON; CASPER; LEONARD, 2006).

A análise acústica da voz emprega técnicas de processamento digital de sinais para extrair características a partir da forma de onda do sinal de voz. Digitalizada, essa forma de onda torna-se uma série temporal que traz consigo informações relevantes do sistema de produção vocal (COSTA, W. *et al.*, 2013). Estas técnicas proporcionam a obtenção dos parâmetros da série temporal (sinal de voz) que podem ou não identificar a presença de um desvio da voz. Além de ser um método não invasivo de avaliação (COSTA, S., 2008).

1.1 Formulação do Problema

Os desvios vocais podem ter implicações na vida social e laboral de um indivíduo, já que a voz humana é uma ferramenta importante de comunicação (ALMEIDA, 2010). A presença de patologias laríngeas pode causar mudanças estruturais ou funcionais resultando em alterações na produção vocal, afetando a qualidade vocal. Por isso, há uma grande correspondência entre a fisiologia laríngea e as medidas acústicas (FANT, 2004; NAYAK, *et al.*, 2005; COSTA, W., 2012).

A avaliação da qualidade vocal é mais frequentemente realizada por meio da avaliação perceptivo-auditiva, de caráter subjetivo e impressionístico, dependente da experiência do profissional, que faz a escuta da voz e emite seu parecer técnico.

Diversas pesquisas tem sido realizadas nos últimos anos que investigam medidas que possam avaliar a qualidade vocal de forma objetiva, aplicando a análise acústica, em que diversas medidas são investigadas. De forma mais tradicional, medidas como frequência fundamental (F_0), o *jitter*, o *shimmer*, as medidas de ruído glótico (*Glottal to Noise Excitation*

ratio (GNE), *Harmonics-to-Noise Ratio* (HNR), *Normalized Noise Energy* (NNE)) são empregadas na análise clínica com relativo sucesso (CARRASCO *et al.*, 2010; BANDEIRA & NETO, 2016; BARAVIEIRA, 2016), detectando e quantificando o grau de desvios vocais (rugosidade, sopro e tensão). No entanto, medidas que dependam da frequência fundamental, podem não ser muito eficientes quando os distúrbios vocais apresentam grau elevado ou são decorrentes de patologias laríngeas que inserem alto nível de ruído nos sinais analisados (COSTA, 2008; GODINO-LLORENTE *et al.*, 2006).

Diante do exposto, questiona-se quais medidas obtidas da análise linear do sistema de produção vocal podem identificar e quantificar desvios vocais e serem úteis na prática clínica. A análise por predição linear e a análise cepstral, já empregadas para detectar desordens vocais provocadas por patologias na laringe (COSTA, 2008) podem quantificar os desvios vocais rugosidade, tensão e sopro, de forma eficiente?

1.2. Justificativa

Neste trabalho propõe-se o uso da análise acústica para detectar a presença de desvios vocais e discriminar os tipos de desvios, dado que diagnóstico precoce dos desvios vocais pode influenciar no tratamento e na correção do desvio. A análise acústica pode auxiliar, de maneira efetiva, no acompanhamento terapêutico e fonoaudiológico utilizando os parâmetros acústicos para avaliar de forma objetiva a qualidade vocal e sua evolução durante o tratamento (BEHLAU, 2008; LOPES, *et al.*, 2016).

A análise acústica tem caráter objetivo e é realizada por meio da extração de parâmetros ou características representativas do sinal de fala e que possam capturar, da melhor forma possível, as alterações no sinal, causadas pela presença do desvio vocal. Por ser de caráter objetivo, esta análise pode fornecer uma ferramenta adicional para avaliação da qualidade vocal.

De acordo com Behlau (2008) a qualidade vocal relaciona-se à impressão total criada por uma voz, e, embora a qualidade vocal varie de acordo com o contexto de fala e as condições físicas do indivíduo, há sempre um padrão básico de emissão que o identifica. Esse padrão é o índice mais completo dos atributos da emissão de um indivíduo, sendo capaz de fornecer desde informações sobre suas características físicas até dados de sua formação educacional.

A análise perceptivo-auditiva é a avaliação clássica da qualidade vocal. Esta avaliação pode ser de caráter exclusivamente impressionístico ou envolver diversas escalas e índices

para uma determinação menos subjetiva e mais confiável do desvio encontrado (BEHLAU, 2008). Já a análise acústica da voz, diferentemente da avaliação perceptivo-auditiva, cujo caráter é subjetivo, proporciona uma avaliação objetiva, realizada por meio de programas computadorizados, fornecendo dados quantitativos que podem indicar a presença do desvio vocal, a sua intensidade e o tipo do desvio observado.

As medidas utilizadas para análise devem estar relacionadas ao padrão vibratório das pregas vocais, à forma do trato vocal, e às suas mudanças no tempo. A interpretação desses parâmetros varia com a idade, o sexo, o tipo de fonação e o treino vocal (COSTA, W., 2012). Geralmente estas medidas dependem da frequência fundamental e em sinais com desvios vocais, provenientes ou não de patologias na laringe, podem apresentar dificuldades na obtenção da frequência fundamental e das medidas de perturbação (COSTA, S. *et al.*, 2014). Outras características independentes da frequência fundamental (F_0) tem sido sugeridas, na literatura, para atuar na discriminação entre sinais de voz, com o fim de identificar e classificar desordens vocais. Além dessas medidas não dependerem da frequência fundamental, são de simples processamento, com baixo custo computacional e também são bastante exploradas na caracterização de sinais de voz (COSTA, S., 2008; COSTA, W., 2012; VIEIRA, 2014).

Diversos trabalhos na literatura têm investigado quais as medidas que melhor caracterizam os desvios vocais e, em sua maior parte, utilizam tanto as medidas tradicionais da análise linear como *jitter*, *shimmer*, GNE, entre outras (COSTA, S., 2008; TELES & ROSINHA, 2008; LOPES *et al.*, 2016), como se baseiam em algumas características da análise dinâmica não linear (DAJER, 2010; COSTA, W., 2012; VIEIRA, 2014).

Algumas medidas que não dependem da frequência fundamental são os coeficientes obtidos da análise preditiva linear (LPC – *Linear Predictive Coding*), as frequências formânticas, os coeficientes cepstrais e mel-cepstrais. Os coeficientes LPC, cepstrais e seus derivados, os coeficientes mel-cepstrais e as frequências formânticas têm sido empregados, como parâmetros de classificação, em sistemas de reconhecimento de fala, sistemas de identificação e verificação de locutor, como também na discriminação entre sinais de vozes afetados por patologias laríngeas (FECHINE, 2000; COSTA, S., 2008; TAVARES, *et al.*, 2010; BARAVIEIRA, 2016).

Uma determinada característica, ou um vetor de características, pode representar bem a presença de um desvio vocal, mas não ser relevante para outro. Então a procura pelas características ou parâmetros mais representativos de desordens vocais é uma busca

incessante, sendo necessário investigar outros parâmetros que possam representar melhor a presença dos desvios vocais (COSTA, S., *et al.*, 2014).

Para que a análise da qualidade vocal seja considerada mais confiável, é necessário que o método empregado seja preciso, que as características do sinal sejam representativas das desordens que o sinal de voz apresenta, devido ao distúrbio vocal provocado ou não por lesão laríngea. Diversos métodos e características/parâmetros têm sido investigados na literatura. No entanto, observa-se que determinada característica ou parâmetro pode representar bem um determinado tipo de desvio vocal, mas não conseguir detectar a presença de outro tipo ou quantificar mais precisamente o grau do desvio presente na fala (COSTA, S., *et al.*, 2014). Dessa forma, a busca da identificação de quais características detectam e quantificam melhor os distúrbios vocais é de grande relevância para a prática clínica.

A realização de um estudo a respeito da classificação de desvios vocais utilizando parâmetros do modelo linear, baseadas na análise LPC, cepstral e mel-cepstral ainda é um campo em aberto. Sua análise é simples computacionalmente e esses parâmetros têm sido bastante empregadas em sistemas de reconhecimento de voz e de locutor, o que as torna potencialmente relevantes para serem investigadas na caracterização de desvios vocais.

1.3 Objetivos

1.3.1 Geral

Avaliar o desempenho vocal por meio das medidas obtidas a partir da análise linear do modelo de produção vocal: análise LPC, análise cepstral e mel-cepstral para a detecção de desvios vocais.

1.3.2 Específicos

- Entender os aspectos teóricos referentes às características do modelo linear de produção da fala;
- Analisar o comportamento das características tradicionais (frequência fundamental, *jitter*, *shimmer*, GNE e frequências formânticas) e dos coeficientes LPC, cepstrais e mel-cepstrais nos sinais a serem analisados;
- Classificar os desvios vocais (rugosidade, sopro e tensão) utilizando as características do modelo linear de produção vocal;

- Investigar a influência da variação do número de níveis empregados na quantização vetorial no desempenho da classificação.
- Comparar o desempenho das medidas tradicionais e as medidas obtidas da análise LPC, cepstral e mel cepstral de forma individual e combinada;
- Avaliar o desempenho do sistema de classificação implementado.

1.4 Organização do Trabalho

No presente capítulo é apresentada a contextualização da pesquisa, sua importância na avaliação da qualidade vocal e as respectivas justificativas, bem como os objetivos do trabalho e as linhas gerais da metodologia empregada. Além de destacar o uso da análise acústica como ferramenta auxiliar em procedimentos de diagnóstico de desvios vocais.

No Capítulo 2 são destacados os métodos propostos descrevendo o comportamento dos parâmetros utilizados para avaliar os desvios vocais, tais como frequência fundamental, *jitter*, *shimmer*, GNE e frequências formânticas. No Capítulo 2, também são descritos os métodos utilizados para obtenção dos parâmetros para a análise por predição linear, a análise cepstral e mel-cepstral, ou seja, a obtenção dos coeficientes LPC, cepstrais e mel-cepstrais. Além de uma descrição de trabalhos relevantes encontrados na literatura que possam contribuir para o desenvolvimento desta pesquisa, abordando os principais aspectos e resultados obtidos na classificação dos sinais que apresentam desordens vocais.

No Capítulo 3 é apresentada a descrição base de dados e da metodologia utilizada no trabalho. No Capítulo 4 são mostrados e analisados os melhores resultados obtidos no processo de classificação e no Capítulo 5 são apresentadas as considerações finais e sugestões para trabalhos futuros.

A identificação dos sinais utilizados da base de dados, conforme lá denominados e algumas características dos locutores, como avaliação perceptivo-auditiva pela Escala Analógica-Visual (EAV) e presença ou ausência de alteração laríngea são apresentados em quadro no Apêndice A. Nos Apêndices B, C e D são apresentados os resultados obtidos para cada método empregado.

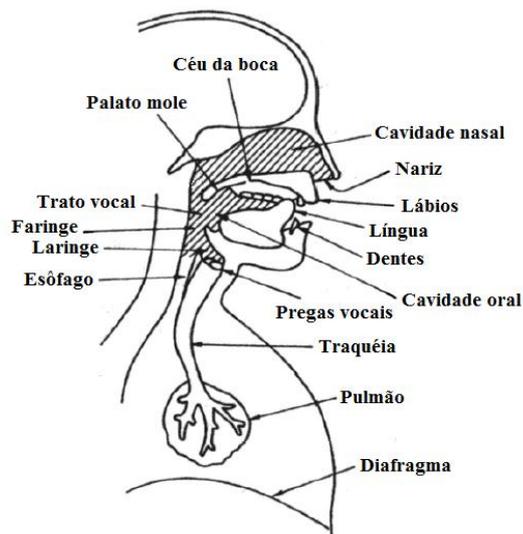
2. CARACTERÍSTICAS DO MODELO LINEAR DE PRODUÇÃO VOCAL

O estudo da anatomia e fisiologia do mecanismo de produção da fala é de fundamental importância para a formatação de modelos físicos e matemáticos que sirvam de base para a construção de sistemas de reconhecimento, síntese e codificação de voz. Assim como, permite compreender o processo da fonação normal e patológica e reconhecer o impacto de lesões específicas na produção vocal (BEHLAU, 2008). A seguir será descrito o sistema de produção vocal e a teoria fonte filtro, bem como alguns parâmetros que podem ser extraídos da voz.

2.1 Sistema de Produção da Voz Humana

Voz, articulação e linguagem são os elementos mais importantes da produção da fala. Quando uma desordem associada a qualquer um desses elementos está presente, a habilidade para se comunicar pode se tornar comprometida (COSTA, W., 2012). A emissão da voz é um fenômeno que comporta grandes variações (LE HUCHE & ALLALI, 2005). Além das consideráveis diferenças de uma pessoa para outra, a voz apresenta em um mesmo indivíduo múltiplos aspectos. Observa-se que a produção da fala pode ser descrita como a inter-relação das forças físicas aerodinâmicas da respiração e da força elástica dos tecidos musculares da laringe que é o principal órgão da voz, já que nela estão as pregas vocais (SANTOS, 2004). O sistema de produção da fala pode ser visto na Figura 1.

Figura 1 – Sistema de produção da voz humana.



Fonte: Adaptado de DELLER & PROAKIS, 2000, apud COSTA, S., 2008.

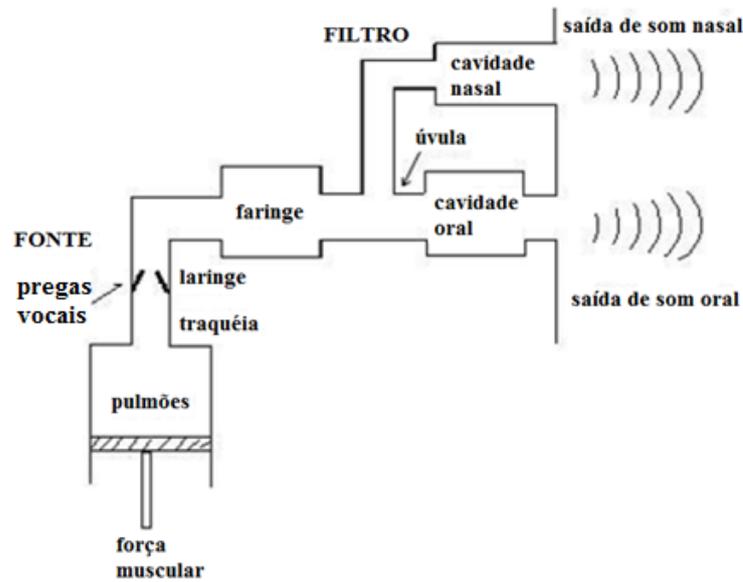
A fala é produzida a partir da liberação de ar dos pulmões para o trato vocal. O trato vocal é formado basicamente por cavidades e órgãos articuladores que começa na abertura entre as pregas vocais e termina nos lábios (RABINER & SCHAFER, 1978). A principal função do trato vocal é modular a onda sonora que vem das pregas vocais e promover constrições para a geração de certos tipos de sons (COSTA, S., 2008). As ressonâncias associadas ao trato vocal têm um efeito transformador sobre a excitação glótica que consiste na amplificação ou atenuação seletiva de componentes espectrais. O som, que se origina na laringe é modificado por várias câmaras de ressonância que se encontram acima e abaixo desta, para, finalmente ser convertido em fala por ação da faringe, língua, palato, lábios e estruturas relacionadas (COSTA, W., 2012).

Esse sistema de produção da fala funciona como um sistema fonte-filtro que também é chamado de modelo linear de produção da fala. Ele é muito útil para entender a relação entre aspectos articulatórios e acústicos além de proporcionar uma fundamentação teórica para vários procedimentos de análise acústica da fala (COSTA, S., 2008; COSTA, W., 2012). A seguir será descrito esse sistema e em que esta teoria se baseia.

2.2 Teoria Fonte-Filtro

O modelo fonte-filtro consiste numa simplificação do sistema de produção de voz, sob as perspectivas físicas e de processamento de sinal. De acordo com este modelo, a excitação das pregas vocais é representada por um trem de impulsos e as propriedades do trato vocal, no que se refere à amplificação de determinadas frequências nas suas cavidades, oral e nasal, são representadas por um filtro variante no tempo e linear na maior parte dos casos. As pregas vocais são consideradas a fonte sonora e o trato vocal, o filtro. Nesse modelo, as saídas produzem ondas acústicas que representam a fala (SOUSA, 2011; COSTA, W., 2012). Na Figura 2 é apresentado um modelo simplificado de produção da fala, também denominado sistema fonte-filtro.

Figura 2 - Modelo simplificado do mecanismo de produção da fala.

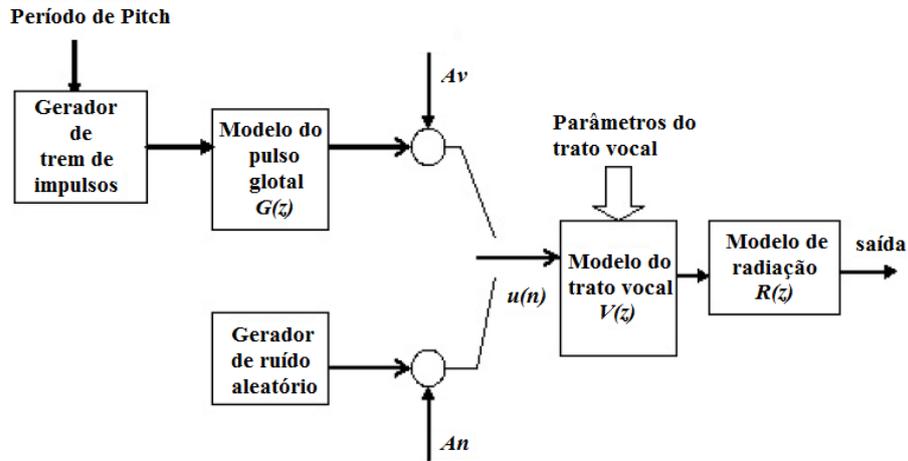


Fonte: Adaptado de DELLER & PROAKIS, 2000, apud COSTA, S., 2008.

O modelo linear de produção da fala é muito útil para entender a relação entre aspectos articulatórios e acústicos além de proporcionar uma fundamentação teórica para vários procedimentos de análise acústica da fala. O princípio dessa teoria é que as características do sistema de produção da fala (o trato vocal) podem ser inferidas através da análise da saída acústica desse sistema. Esse modelo foi construído baseado nas características do modelo vocal humano, em que a fonte de excitação e o aparelho vocal são considerados como dois sistemas separados (ALCAIM & OLIVEIRA, 2011; COSTA, S., 2008).

O modelo completo do sistema de produção da fala é mostrado na Figura 3. Para a produção dos sinais sonoros é gerado um trem de impulsos unitários cuja periodicidade é determinada pelo período de *pitch* (inverso da frequência fundamental) do sinal da fala. Esse trem de impulsos é aplicado a um filtro digital $G(z)$ que simula o efeito dos pulsos glotais, que são devidamente selecionados e aplicados ao trato vocal, após um controle de ganho A_v . Para a produção dos sinais não-sonoros é utilizado um gerador de ruído aleatório com espectro plano e um controle de ganho A_n . As variáveis A_v e A_n controlam a intensidade da excitação do sinal de voz e do ruído respectivamente e $u(n)$ representa o sinal de excitação (RABINER & SCHAFER, 1978, apud COSTA, S., 2008).

Figura 3 - Modelo linear discreto no tempo para a produção da fala.



Fonte: Adaptado de DELLER & PROAKIS, 2000, apud COSTA, S., 2008.

O modelo linear de produção da fala é apresentado na Figura 3. Este modelo insere os efeitos dos pulsos glotais, trato vocal e da radiação dos lábios como um filtro linear. A fonte é uma sequência de impulsos quase periódicos ou adiciona uma sequência de ruído aleatório. No primeiro caso gera sons sonoros e no segundo caso gera sons surdos. Um fator de ganho, G , é ajustado para controlar a intensidade da excitação. Combinando os efeitos dos pulsos glotais do trato vocal e da radiação o modelo pode ser representado por uma função de transferência de apenas polos, $H(z)$, como apresentado na Equação (1) (RABINER & SCHAFER, 1978, apud COSTA, S., 2008)

$$H(z) = \frac{G}{1 - \sum_{k=1}^p \alpha(k)z^{-k}} = \frac{G}{A(z)}, \quad (1)$$

em que G representa o ganho que é ajustável, os termos $a(k)$ representam os coeficientes de predição linear e p a ordem do filtro de predição.

2.3 Frequência Fundamental

A frequência fundamental é definida como a menor frequência de uma onda periódica. Quando aplicada no estudo da voz, a frequência fundamental é resultado da vibração das pregas vocais, indicando as variações que ocorrem entre sons agudos e graves e também a variação de sua intensidade (FRAZÃO; COSTA, S.; CORREIA, 2013). Os fatores de determinação da frequência fundamental e sua variação em extensão vocal são o comprimento natural da prega vocal, o alongamento, a massa em vibração e a tensão envolvida. A frequência fundamental é enormemente afetada pelo sexo e pela idade, com uma distribuição

média de 80 a 250 Hz nos adultos jovens, sendo que nos homens a faixa está entre 80 a 150 Hz, nas mulheres de 150 a 250 Hz e nas crianças está entre 250 a 400 Hz (BEHLAU, 2008).

Estes valores não são estacionários uma vez que, além de variarem com o sexo e a idade, pensa-se que dependem também de fatores como o estado de espírito da pessoa, o período do dia em que se enquadram, os hábitos de vida, o uso profissional da voz e o desvio vocal (TEIXEIRA; FERREIRA; CARNEIRO, 2011).

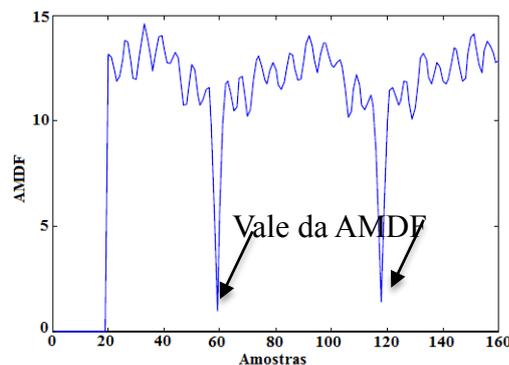
A frequência fundamental pode ser extraída de várias formas como, por exemplo, pelo método da Função da Média de Diferenças de Amplitudes (AMDF - *Average Magnitude Difference Function*) (ZHANG; XU; WANG, 2002), método da função de autocorrelação (SONDHI, 1968), algoritmos que utilizam análise cepstral e a partir do resíduo da análise LPC, sendo os dois primeiros os mais utilizados (SANTOS, 2015).

O método utilizado nesta pesquisa para a extração da frequência fundamental foi o método da Função da Média de Diferenças de Amplitudes (AMDF) (ZHANG; XU; WANG, 2002). Este método é considerado simples e eficiente (COSTA S., 2008). Considera-se o sinal $s(n)$ periódico, de período P . Seja a Equação (2):

$$AMDF(k) = \frac{1}{N} \sum_{n=1}^{N-k} |s(n+k) - s(n)|, \quad (2)$$

em que $s(n)$ é o sinal de voz e N é o número de amostras no segmento do sinal. A AMDF é mínima para os valores de k , múltiplos do período. A frequência fundamental é estimada calculando o mínimo global da AMDF considerando os limites admissíveis do período como apresentado na Equação (2), pois se verifica que a função AMDF é mínima no período correspondente à frequência fundamental e que não há mínimos comparáveis nos segmentos sem voz (COSTA S., 2008; SOUSA, 2011). Na Figura 4 a frequência fundamental pode ser determinada pelos vales da função AMDF.

Figura 4 – AMDF para vogal sustentada /ε/ (“é”).



AMDF da vogal “e”.

Fonte: COSTA, S., 2015.

2.4 Medidas de Perturbação

Jitter e *shimmer* são medidas de perturbação obtidas a partir de segmentos sonoros do sinal de voz (GODINO; GOMEZ; BLANCO, 2006). O *Jitter* é definido como as variações na periodicidade da frequência fundamental (KASUYA; ENDO; SALIU, 1993). Ele pode ser expresso através de medidas absolutas ou medidas relativas à frequência fundamental.

As medidas absolutas de *jitter* são os valores de perturbação do período que ignoram a frequência fundamental. É expresso em segundos, milissegundos ou microssegundos. Já as medidas relativas consideram a frequência fundamental média. Essas medidas são expressas em porcentagem e o valor limite de normalidade é 0,5% (BEHLAU, 2008). O cálculo é baseado na diferença entre os picos mais altos de cada impulso glotal, onde serão determinados os índices dos respectivos picos máximos para calcular a diferença entre cada pico (TEIXEIRA; FERREIRA; CARNEIRO, 2011). Deve ser medido por análise de vogais sustentadas, em que a vogal utilizada deve ser indicada, em uma única frequência, sem que haja variação musical ou de intensidade.

Ele é calculado usando a definição do quociente de perturbação da frequência (*FPQ* - *Frequency Perturbation Quotient*) (TEIXEIRA; FERREIRA; CARNEIRO, 2011), conforme a Equação (3).

$$FPQ = \frac{\left(\frac{1}{N-1}\right) \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (3)$$

onde T_i representa o comprimento de cada período de F_0 e N é o número de períodos de F_0 .

Esses valores sofrem interferência do sexo e da idade do falante e indica a pequena variabilidade entre os sucessivos ciclos glóticos. Altos níveis de *jitter* normalmente estão associados às vozes desviadas. A instabilidade da frequência fundamental pode ser atribuída a mudanças no tamanho, formato ou firmeza das pregas vocais (KASUYA; ENDO; SALIU, 1993). No caso dos desvios vocais, este valor altera-se principalmente com a falta de controle da vibração das pregas vocais, que ocorre nos desvios de origem neurológica e correlaciona-se com a aspereza (BEHLAU, 2008; COSTA S., 2004). É aceitável que qualquer voz possua alguma instabilidade devido a fatores de ordem neurológica, emocional, entre outras. Por isso, é previsível a presença de um pequeno grau de perturbação e irregularidade no sinal vocal. A maioria dos investigadores considera como valor típico “normal” a variação entre 0,5 e os 1,0% para as fonações sustentadas em jovens adultos (TEIXEIRA; FERREIRA; CARNEIRO, 2011).

O *shimmer* é uma medida de perturbação que reflete a variação nas amplitudes consecutivas de uma onda sonora a curto intervalo de tempo (WERTZNER; SCHREIBER; AMARO, 2005). Esta medida representa a variação na intensidade dos ciclos adjacentes de vibração das pregas vocais e altera-se com a redução da resistência glótica e lesões de massa nas pregas vocais, estando correlacionado com a presença de ruído à emissão e com a soproiedade (BEHLAU, 2008).

O *shimmer* também é conhecido como *Amplitude Perturbation Quotient* (APQ). O cálculo do *shimmer* é baseado na diferença entre os picos mais altos de cada impulso glotal, em que é realizado o sincronismo e calculada a diferença entre as amplitudes sucessivas do sinal nesses instantes e é calculada pela Equação (4) (TEIXEIRA; FERREIRA; CARNEIRO, 2011).

$$APQ = \frac{\left(\frac{1}{N-1}\right) \sum_{i=1}^{N-1} |A_i - A_{i-1}|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (4)$$

onde A_i representa a amplitude do impulso glotal i , e N o número de impulsos glotais.

É uma medida de estabilidade fonatória em que observa se o sistema fonatório é absolutamente estável. Altos níveis de *shimmer* normalmente estão associados às vozes patológicas. Isto pode ser atribuído devido às mudanças no tamanho, formato ou firmeza das pregas vocais (KASUYA; ENDO; SALIU, 1993).

Quanto mais estável o *shimmer*, seu valor será mais próximo de zero. Oferece uma percepção indireta do ruído na produção vocal e seus valores crescem quanto maior a quantidade de ruído numa emissão (rouquidão), como ocorre na soproiedade (COSTA S., 2004). Esse parâmetro deve ser medido por análise de vogais sustentadas numa única frequência, sem nenhuma variação, pois há variação nos valores do *shimmer* para as diferentes vogais, para sexo e para as idades (COSTA S., 2004).

2.5 *Glottal to Noise Excitation Ratio* (GNE)

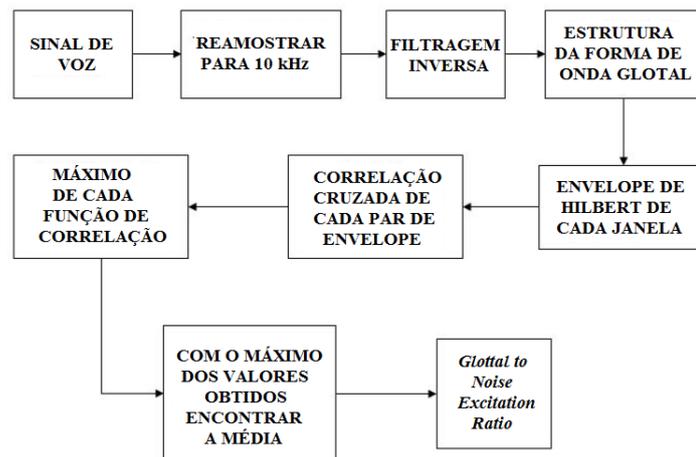
Glottal to Noise Excitation ratio (GNE) corresponde à proporção do sinal glótico da excitação pelo ruído. É uma medida acústica utilizada para calcular o ruído em uma série de pulsos produzidos pela oscilação das pregas vocais (CARRILLO & ORTIZ, 2007; CARRASCO; OLIVEIRA; BEHLAU, 2010). O GNE representa uma abordagem interessante para quantificar a quantidade de excitação devido às oscilações da prega vocal versus a excitação dada pelo ruído. Assim, ele está intimamente relacionado com a respiração e é

considerada uma medida confiável para o nível de ruído relativo, mesmo na presença de fortes perturbações de amplitude e frequência (LLORENTE *et al.*, 2010).

Este parâmetro é baseado na hipótese que pulsos resultantes da colisão das pregas vocais geram uma excitação síncrona de diferentes faixas de frequência. Por outro lado, o ruído produzido pelas pregas vocais comprimidas gera excitações não correlacionadas (MICHAELIS & STRUBE, 1995).

O cálculo do ruído é realizado em uma série de pulsos, o qual indica se o sinal vocal está sendo originado pela vibração das pregas vocais ou pela corrente de ar turbulenta gerada no trato vocal (PIFAIA; MADAZIO; BEHLAU, 2013). Na Figura 5 é apresentado o fluxograma contendo a metodologia utilizada para extração do GNE, utilizada pelo *software VoxMetria 3.0*, empregado nesta pesquisa.

Figura 5 – Fluxograma apresentando a metodologia de extração do GNE.



Fonte: Adaptado de VISAVE, *et al.*, 2015.

O fluxograma da Figura 5 pode ser descrito da seguinte forma (MICHAELIS & STRUBE, 1995): o sinal de voz é reamostrado para 10 kHz, a partir do qual é obtida a forma de onda glotal do LPC residual usando um preditor de ordem 13. É utilizada uma janela de *Hanning* de 30 ms com sobreposição de 10 ms para o processamento do sinal. Posteriormente é realizado o enquadramento da forma de onda glotal encontrada do sinal. Na etapa seguinte deve-se calcular os envelopes de Hilbert de diversas bandas de frequência para cada janela, com a banda fixa e diferentes frequências centrais. É realizada a correlação cruzada para cada par de envelopes, onde a distância entre as frequências de centro é igual ou superior a metade da largura de banda. Deve ser escolhido para cada par da banda de frequência o valor máximo de cada função de correlação e, por conseguinte, deve-se escolher os valores máximos obtidos anteriormente. Por fim, é calculada a média das medidas obtidas encontrando o GNE.

Os valores do GNE são considerados normais quando maiores ou iguais a 0,5 dB (CARRASCO; OLIVEIRA; BEHLAU, 2010). Está intimamente relacionado com a soprosidade. É aplicável mesmo para oscilações glotais altamente irregulares (MICHAELIS; GRAMSS; STRUBE, 1997). Se o valor de GNE for próximo a um, significa excitação normal da glote, enquanto que os valores próximos a zero resultam de uma excitação do tipo ruído. O GNE não necessita do cálculo da frequência fundamental e gera informações mais independentes de *jitter*, *shimmer* e correlação que outros parâmetros utilizados para descrever ruído como o *Normalized Noise Energy* (NNE) e o *Cepstrum based Harmonics to Noise Ratio* (CHNR) (MICHAELIS & STRUBE, 1995).

2.6 Análise de Voz por Predição Linear

A análise de voz por predição linear ou simplesmente análise LPC fornece um conjunto de parâmetros da fala que representam o trato vocal. A seguir serão descritos os coeficientes LPC e as frequências formânticas.

2.6.1 Coeficientes LPC

O método da análise de voz por predição linear estima cada amostra de voz baseando-se em uma combinação linear de p amostras anteriores. Um valor de p maior representa um modelo mais preciso. Na Equação (5) é definida a saída do preditor linear com coeficientes de predição, $a(k)$ (RABINER & SCHAFER, 1978, apud COSTA, S., 2008).

$$\tilde{s}(n) = \sum_{k=1}^p a(k)s(n-k), \quad (5)$$

sendo p a ordem do preditor e k representa o atraso das amostras.

Existem várias formulações diferentes para a predição linear, sendo que algumas delas são equivalentes entre si. O método de autocorrelação e o da covariância são dois métodos padrões para o cálculo dos coeficientes do preditor. Ambos os métodos são baseados na minimização do valor médio quadrático do erro de estimação $e(n)$, ou sinal residual, como dado pela Equação (6) (RABINER & SCHAFER, 1978; O'SHAUGHNESSY, 2000, apud COSTA, S., 2008).

$$e(n) = s(n) - \sum_{k=1}^p a(k)s(n-k), \quad (6)$$

sendo p a ordem do preditor, n é a amostra atual e k é a amostra atrasada.

Um dos métodos utilizados para calcular os coeficientes LPC é o método da autocorrelação. Este método baseia-se na minimização da energia do erro em função dos coeficientes $a(k)$ do filtro da Equação (5) (SOUSA, 2011). A determinação dos coeficientes do filtro é obtida pela minimização da energia do erro de predição, E , entre as sequências de amostras do sinal $s(n)$ e as sequências de amostras estimadas ($\tilde{s}(n)$) como apresentado nas Equações (7) e (8) (RABINER & SCHAFER, 1978, apud COSTA, S., 2008).

$$E = \sum_n e(n)^2 \quad (7)$$

sendo $e(n)$ o erro do preditor estimado apresentado na Equação (6). Para realizar a minimização da energia do erro em função dos coeficientes $a(k)$ do filtro é necessário realizar a derivada de E em relação a $a(k)$, como pode ser visto na Equação (8).

$$\frac{\partial E}{\partial a(k)} = 0 \quad k = 1, \dots, p \quad (8)$$

sendo p a ordem do preditor. Obtendo um sistema de p equações com p incógnitas, que pode ser representado por uma matriz expressa na Equação (9).

$$\begin{bmatrix} R_s[1] \\ R_s[2] \\ \dots \\ R_s[p] \end{bmatrix} = \begin{bmatrix} R_s[0] & R_s[1] & \dots & R_s[p-1] \\ R_s[1] & R_s[0] & \dots & R_s[p-2] \\ \dots & \dots & \dots & \dots \\ R_s[p-1] & R_s[p-2] & \dots & R_s[0] \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \dots \\ a(k) \end{bmatrix} \quad (9)$$

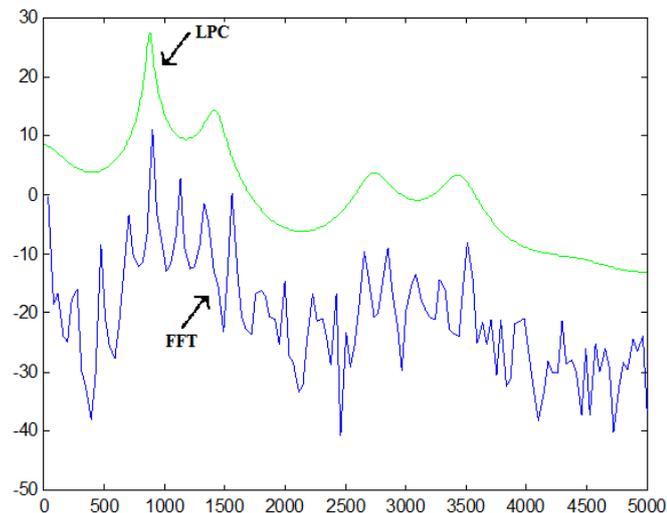
Sendo R a matriz de autocorrelação do sinal degradado e $a(k)$ são os coeficientes de predição linear.

2.6.2 Frequências formânticas

As frequências formânticas ou formantes representam ressonâncias naturais do trato vocal, na posição articulatória da vogal falada (BEHLAU, 2008). Correspondem os picos de energia em uma região do espectro sonoro e são particularmente importantes na determinação da fala. São definidos por seus picos de frequência, amplitude e largura de banda, mas a amplitude e a largura de banda têm pouco interesse na clínica vocal e raramente são demonstrados (COSTA S., 2004; BEHLAU, 2008).

Para extração das frequências formânticas são utilizados os sinais correspondentes às vogais, devido ao seu caráter sonoro. No caso das consoantes, elas possuem zonas de incremento de energia denominadas de zona de alta intensidade, mas isso não representa um formante (COSTA S., 2004). As frequências formânticas são retiradas a partir da Análise por Predição Linear (LPC), sendo um filtro inverso dos coeficientes $a(k)$. Geralmente são expressos através do seu valor médio, em Hertz (Hz) ou ciclos por segundo. As frequências formânticas são extraídas tanto da análise FFT como da análise LPC (BEHLAU, 2008). Na Figura 6 pode ser observada a comparação do espectro LPC e do espectro da FFT para um sinal de voz. Os coeficientes LPC representam uma média suavizada do espectro de Fourier do sinal (COSTA, S.,2008). Quanto maior a ordem do preditor, mais próximo do espectro da FFT será o espectro LPC.

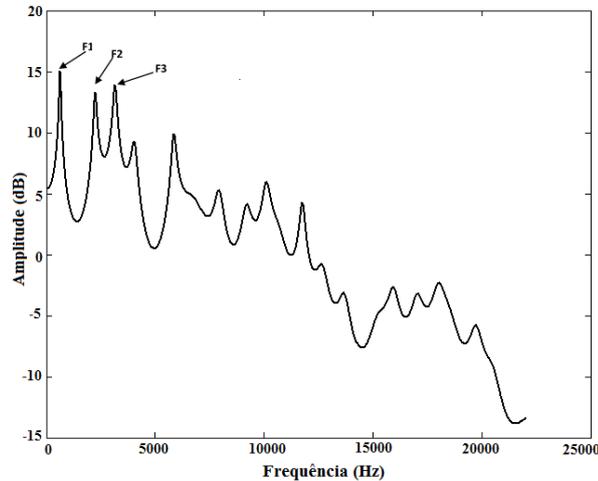
Figura 6 - Comparação dos espectros LPC e FFT para um sinal de voz.



Fonte: Adaptado de TEIXEIRA, 2005.

As frequências formânticas são definidas no domínio espectral. Um dos métodos utilizados para realizar este cálculo, baseia-se na determinação dos polos do filtro do trato vocal estimado na Equação (1). O LPC define um envelope da magnitude do espectro do sinal que é obtida a partir dos coeficientes do filtro, determinando a resposta em frequência deste. Também podem ser estimadas pela localização dos picos no espectro logarítmico (espectro LPC) (Figura 5) (DINIZ & THOMÉ, 1998; SOUSA, 2011). Na Figura 7 podem ser observadas as três primeiras frequências formânticas para a vogal sustentada /É/ (“é”).

Figura 7 – Frequências formânticas para a vogal sustentada /ɛ/ (“ê”). F_1 representa 1º formante, F_2 o 2º formante e F_3 o 3º formante.



Fonte: Autoria própria.

A descrição das vogais quase nunca ultrapassa a identificação das três primeiras frequências formânticas, sendo que são elas que determinam a qualidade vocal em termos acústicos e sua identidade em termos auditivos. A primeira frequência formântica (F_1) está interligada com o abaixamento da mandíbula, a segunda frequência formântica (F_2) está relacionada com o deslocamento da língua e a terceira frequência formântica (F_3) está relacionada com as duas cavidades estabelecidas pela posição da língua, ou seja, à cavidade atrás da constrição da língua e aquela à frente. A frequência de F_3 pode ser influenciada também por ajuste de lábios, laringe e faringe (MAGRI *et al.*, 2007; BEHLAU, 2008). A seguir, são apresentados na Tabela 1 os valores médios do primeiro formante (F_1), do segundo formante (F_2) e do terceiro formante (F_3) para homens e mulheres, falantes do português brasileiro da cidade de São Paulo (BEHLAU, 2008).

Tabela 1 – Tabela com os valores médios para as frequências formânticas das vogais orais em falantes da cidade de São Paulo.

Grupos	Frequências formânticas (Hz)	Vogais						
		“i”	“ê”	“e”	“a”	“o”	“ô”	“u”
Homens	F_1	398	563	699	807	715	558	400
	F_2	2456	2339	2045	1440	1201	1122	1182
	F_3	3320	2995	2848	2524	2481	2520	2452
Mulheres	F_1	425	628	769	956	803	595	462
	F_2	2984	2712	2480	1634	1317	1250	1290
	F_3	3668	3349	3153	2721	2602	2668	2528
Crianças	F_1	465	698	902	1086	913	682	505
	F_2	3176	2825	2606	1721	1371	1295	1350
	F_3	3980	3637	3243	2873	2793	2823	2667

F_1 : primeiro formante; F_2 : segundo formante; F_3 : terceiro formante. Fonte: BEHLAU, 2008.

Cada vogal apresenta suas frequências formânticas características, devido à configuração do trato vocal durante a locução do som e cada indivíduo apresenta suas frequências formânticas particulares, para uma determinada vogal, devido às dimensões das estruturas do trato vocal e do padrão articulatorio que são únicos para cada indivíduo. Os valores das frequências formânticas (F_1 , F_2 e F_3) também possuem correspondência com os ajustes supralaríngeos da qualidade vocal de indivíduos que apresentam desvios vocais. Os ajustes supraglóticos (laringe baixa, mandíbula fechada, lábios arredondados, corpo de língua abaixado, corpo de língua retraído e labiodentalização) da qualidade vocal apresentam correspondência principalmente aos valores das frequências formânticas F_1 e F_3 com diferenças estatisticamente significantes apresentando a diminuição destas frequências em relação aos valores de referência para o Português-Brasileiro (MAGRI *et al.*, 2007).

2.7 Análise Cepstral

A análise cepstral do sinal de voz para o estudo das alterações laríngeas pode ser muito útil, pois permite trabalhar com o sinal da excitação da glote separadamente das repercussões ressonantes do trato vocal. Os coeficientes cepstrais são usados para descrever a envoltória espectral do sinal de voz a curto intervalo de tempo. A aplicação dessa técnica, no estudo do sinal acústico de vozes alteradas, pode detectar modificações no sinal de voz que se relacionem com as alterações laríngeas e, conseqüentemente, identificar modelos para uma classificação dos sinais (FECHINE, 2000; ZWETSCH, *et al.*, 2006, apud COSTA, S., 2008).

Os coeficientes cepstrais podem ser calculados tanto da FFT, como a partir dos coeficientes LPC (FECHINE, 2000; SILVA, 2011). A partir dos coeficientes de predição linear, $a(k)$, são calculados recursivamente por meio da Equação (10). O uso dessa recursão permite um cálculo eficiente dos coeficientes e evita fatoração polinomial (MAMMONE *et al.*, 1996, apud COSTA, S., 2008).

$$\begin{cases} c(1) = -a(1) \\ c_1(n) = -a(n) - \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) a(j)c(n-j), \end{cases} \quad 1 < n < p' \quad (10)$$

em que n corresponde ao n -ésimo coeficiente da predição linear.

Os coeficientes cepstrais representam as condições da fonte (laringe) e do filtro (o trato vocal), separadamente. O sinal de voz (Equação (11)) é resultado da convolução da

excitação, $ex(n)$, com a resposta do trato vocal, $\theta(n)$. Então seria útil separar ou deconvoluir as duas componentes. (ZWETSCH *et al.*, 2006, apud COSTA, S., 2008).

$$s(n) = ex(n) * v(n) \quad (11)$$

A deconvolução cepstral converte um produto de dois espectros (como na Equação (11)) na soma de dois sinais, separando-os por um processo de filtragem linear, a “lifteragem” ou *liftering* facilitando o estudo individual das modificações ocorridas na fonte e no filtro. Das propriedades matemáticas envolvidas nesta operação destacam-se as FFT e as funções logarítmicas, que resultam em uma função chamada cepstral ou cepstro, responsável pela dissociação do sinal de voz. A transformação desejada é logarítmica, dada pela Equação (12):

$$\log(Ex(\omega).V(\omega)) = \log(Ex(\omega)) + \log(V(\omega)) \quad (12)$$

onde $Ex(\omega)$ e $V(\omega)$ são as transformadas de Fourier da forma de onda da excitação e da resposta do trato vocal, respectivamente.

A deconvolução realizada resulta em uma função chamada cepstro, responsável, justamente, pela separação da fonte e do filtro. Na prática, o cepstrum complexo não é necessário, sendo suficiente o cepstrum real, definido como a transformada inversa do logaritmo do espectro de magnitude (Equação (13)). (ZWETSCH *et al.*, 2006, apud COSTA, S., 2008).

$$c(n) = \frac{1}{2\pi} \int_0^{2\pi} \log|X(e^{j\omega})| e^{j\omega} d\omega \quad (13)$$

Para sinais reais $x(n)$, $c(n)$ é a parte par do cepstro $\hat{x}(n)$.

2.8 Análise Mel-Cepstral

Os coeficientes mel-cepstrais (*Mel-frequency Cepstral Coefficients* – MFCC) surgiram devido aos estudos que mostraram que a percepção humana das frequências de tons puros ou de sinais de voz não segue uma escala linear. Para cada tom com frequência f , medida em Hz, define-se um tom subjetivo medido em uma escala chamada escala mel. O mel é uma unidade de medida da frequência percebida de um tom (DELLER; PROAKIS; HANSEN, 1987, apud

COSTA, 2008). A frequência acústica em Hz pode ser transformada para uma escala de frequências percebidas Mel da seguinte maneira (Equação (14)) (O'SHAUGHNESSY, 2000, apud COSTA, S., 2008):

$$F_{mel} = 2595 \cdot \log_{10} \left(1 + \frac{F_{linear}(Hz)}{700} \right) \quad (14)$$

em que F_{linear} é a frequência em Hz e F_{mel} é a frequência percebida (em mel).

Para realizar o cálculo dos coeficientes mel-cepstrais primeiramente é calculado o espectro de magnitude do sinal, $x(n)$ a partir do módulo da transformada de Fourier ($|FFT(x(n))|^2$) para cada janela. Posteriormente é aplicado um banco de filtros triangulares em escala mel com o formato triangular e não são separados linearmente como pode ser visto na Figura 8.

Figura 8 - Banco de filtros digitais na escala mel.



Fonte: Adaptado de COSTA, S., 2008.

A quantidade de filtros é baseada na frequência de amostragem que é dada por $(Fa) \cdot (3 \ln(Fa))$. Em seguida é feito o cálculo do logaritmo da energia de saída de cada filtro para a obtenção do cepstro e por fim é realizada a obtenção dos coeficientes mel-cepstrais $c_{mel}(n)$, através da Equação (15) (O'SHAUGHNESSY, 2000, apud COSTA, S., 2008).

$$c_{mel}(n) = \sum_{k=1}^{N_f} \log(S_{FFT}(k)) \cdot \cos \left[n \left(k - \frac{1}{2} \right) \right] \cdot \frac{\pi}{N_f}, \quad n = 0, 1, \dots, N_f \quad (15)$$

N_f : número de filtros digitais utilizados; $c_{mel}(n)$: n -coeficientes mel-cepstrais; $S_{FFT}(k)$: sinal de saída do bando de filtros digitais que é obtido através da Equação (16).

$$S_{FFT}(k) = \sum_{j=1}^{N_{FFT}} W(j) \cdot X(j), \quad k = 1, \dots, N_f, \quad (16)$$

$W(j)$ correspondem às janelas de ponderação triangulares associadas a escala mel e $X(j)$ corresponde o espectro dada *FFT* para n pontos (O'SHAUGHNESSY, 2000, apud COSTA, S., 2008).

2.9 Avaliação Perceptivo-auditiva

A Avaliação Perceptivo-auditiva utiliza diversas escalas e índices, entre elas estão a escala japonesa GRBASI (G – grau global do desvio, R – *roughness*, B – *breathness*, A – *asteny*, S – *strain*, I - *instability*) e a Escala Analógica Visual (EAV).

A escala GRBASI é uma escala japonesa desenvolvida pelo *Commite for Phonatory Function Test* da *Japan Society of Logopedics and Phoniatics*. Seu método de avaliação é simples, onde é observado o grau global de desvio (G) pela identificação da contribuição de cinco fatores independentes que são: rugosidade (R), soproiedade (B), astenia (A), tensão (S) e instabilidade (I). A rugosidade representa irregularidade na vibração das pregas vocais, a soproiedade é o escape de ar audível na voz, astenia significa perda de potência ou fraqueza vocal, a tensão é a impressão de estado hiperfuncional apresentando frequência aguda e a instabilidade corresponde à flutuação na frequência fundamental e/ou na qualidade vocal (BEHLAU, 2008; BARAVIEIRA, 2016). A escala possui quatro pontos que são usados para a identificação do grau de desvio de cada um dos fatores, onde “0” significa normal ou ausente, “1” discreto, “2” moderado, e “3” severo. A avaliação é feita através da escuta do paciente e os avaliadores julgam cada parâmetro de acordo com os pontos citados anteriormente e os resultados são anotados com os níveis de avaliação subscritos (BEHLAU, 2008).

A Escala Analógica Visual (EAV) é amplamente utilizada na área da saúde por diferentes profissionais, principalmente como instrumento de auto percepção do paciente quanto à dor ou ao incômodo. Também é usada para mensurar a gravidade de uma doença ou alteração, como no caso dos fonoaudiólogos na avaliação perceptivo-auditiva. Esta escala é composta por uma reta, geralmente com comprimento de 100 *mm*, em que os ouvintes devem marcar o ponto que corresponde à extensão da variação de uma dada característica, como o grau de desvio da voz. São considerados 4 graus de desvio vocal: o grau normal (G1) que varia entre 0 e 35,5 *mm* na escala; o grau leve a moderado (G2) que varia entre 35,6 a 50,5 *mm*; o grau moderado (G3) que varia entre 50,6 e 90,5 *mm*; e o grau intenso (G4) que varia entre 90,5 e 100 *mm*. (MARTINS; COUTO; GAMA, 2015).

2.10 Trabalhos Relacionados

Nesta Seção são apresentadas, de forma sucinta, algumas pesquisas encontradas na literatura que buscam desenvolver métodos de classificação utilizando as características tradicionais, a análises LPC, cepstral e mel-cepstral para dar maior robustez à avaliação da qualidade vocal. No Quadro I, são apresentadas informações de pesquisas que utilizam a análise acústica como ferramenta no auxílio do diagnóstico dos desvios vocais através dos métodos propostos.

A maioria dos trabalhos relacionados aborda a classificação de desvios vocais utilizando as frequências formânticas e também características tradicionais como *jitter*, *shimmer*, medidas de ruído, entre outras medidas que dependem da frequência fundamental.

Quadro I - Estado da Arte.

Autoria	Características Extraídas	Metodologia	Resultados e Conclusão
Ma & Yu (2006)	<ol style="list-style-type: none"> 1. Tempo máximo de fonação; 2. Pressão intraoral; fluxo de ar e pressão na fonação; 3. Perfil de extensão vocal; 4. Frequência fundamental; 5. <i>Jitter</i>; 6. <i>Shimmer</i>; 7. Proporção harmônico ruído. 	<p>Base de dados: 112 sinais com desvio vocal e 41 sinais de vozes saudáveis.</p> <p>Avaliaram a precisão das características extraídas comparando ao grau geral de desvio vocal, graduado por meio da avaliação perceptivo-auditiva.</p>	<p>Melhores parâmetros observados: tempo máximo de fonação da vogal /a/, o pico de pressão intraoral na sílaba /pi/, o perfil de extensão vocal e o <i>jitter</i>. O uso destes quatro parâmetros combinados discriminou corretamente 67,3% dos níveis de desvio vocal.</p>
Costa S. (2008)	<ol style="list-style-type: none"> 1. Frequência fundamental; 2. Estrutura dos formantes 3. Energia a curtos intervalos; 4. Coeficientes da Análise por Predição Linear; 5. Coeficientes da Análise Cepstral e seus derivados; 6. Coeficientes da Análise Mel-cepstral. 	<p>Base de dados: 120 sinais de vozes distribuídos entre sinais patológicos e saudáveis.</p> <p>Foram avaliadas as características tradicionais e adicionalmente foi realizada uma análise por meio de abordagem Paramétrica utilizando as características extraídas; As características foram processadas por modelagem paramétrica baseada no Modelo de Markov Escondido (HMM).</p>	<p>Os coeficientes LPC, cepstrais e seus derivados e os coeficientes mel cepstrais foram eficientes na discriminação da patologia, apresentando uma taxa de classificação acima de 90%.</p> <p>Estes métodos são eficientes em modelar os efeitos provocados pela patologia e permitir uma discriminação eficiente da patologia quando comparada a vozes normais.</p>

Autoria	Características Extraídas	Metodologia	Resultados e Conclusão
Teles & Rosinha (2008)	1. Frequência fundamental; 2. <i>Jitter</i> ; 3. <i>Shimmer</i> ; 4. <i>Amplitude Perturbation Quocient</i> (APQ); 5. <i>Period Perturbation Quotient</i> (PPQ); Taxa de Harmônico Ruído (HNR); Três primeiras frequências formânticas.	Criação de um banco de dados utilizando as características extraídas das sete vogais do português brasileiro. 107 mulheres participaram do estudo, que foram divididas em 3 grupos pela faixa etária. Avaliação perceptivo-auditiva por meio da escala GRBASI	Não houve diferença estatisticamente significativa entre os grupos nas medidas avaliadas, mas os valores das medidas de perturbação <i>jitter</i> , APQ, PPQ, HNR foram maiores no Grupo 3 (faixa etária maior). Concluíram que a criação de um banco de dados normativos pode favorecer as pesquisas com vozes desviadas e vozes saudáveis.
Tavares <i>et al.</i> (2010)	1. Coeficientes LPC Parâmetros cepstrais e seus derivados	Base de dados 149 sinais de vozes distribuídas em patológicas (Edema de Reinke e Paralisia) e saudáveis; Utilizou classificadores de maneira individual e combinada, em que na combinação utilizou como regra de classificação a média aritmética e a soma ponderada.	A análise acústica pode ser empregada como ferramenta adicional aos exames tradicionais para pré-diagnóstico de patologias laringeas e avaliação de tratamentos pré e pós-cirúrgicos. As taxas de eficiência obtidas variam de 85% usando somente os coeficientes LPC e de 99%, considerando a combinação de classificadores.
Bandeira & Neto (2013)	Frequência fundamental	Base de dados: 30 sinais de vozes de professoras distribuídos em sinais desviados e sem alteração vocal; Programa PRAAT para análise do parâmetro.	Na análise dos sinais apresentaram desvio vocal, a média da frequência fundamental foi mais alta do que nos sinais sem alteração vocal, concluindo que as vozes desviadas apresentaram ajustes de frequência diferentes das não desviadas.
Costa S. <i>et al.</i> (2014)	1. Análise por predição linear, análise cepstral e mel-cepstral e seus respectivos coeficientes; 2. Dinâmica não linear (análise de quantificação de recorrência) e; Transformada <i>Wavelet</i> (escalogramas).	Artigo de revisão das características citadas; Avaliaram a utilização da análise linear, não-linear e Transformada <i>Wavelet</i> na classificação dos sinais patológicos e saudáveis.	Os métodos são promissores, sendo necessário que seja realizada a escolha do melhor método para cada aplicação.

Autoria	Características Extraídas	Metodologia	Resultados e Conclusão
Baravieira (2016)	1. Característica Transformada da <i>Wavelet Packet</i> ; 2. <i>Jitter</i> ; 3. <i>Shimmer</i> ; 4. Amplitude derivada; da Amplitude do <i>pitch</i> .	Base de dados: 123 sinais de vozes desde neutras até sinais com presença de grau intenso de rugosidade e/ou sopro; Avaliação perceptivo-auditiva foi realizada através da EAV e pela escala numérica de 4 pontos; Utilizou um classificador de RNA (Rede Neural Artificial) validado por meio de parametrização.	Na caracterização de sinais de voz com sopro, as melhores características foram <i>jitter</i> , amplitude de <i>pitch</i> e frequência fundamental com taxa de acertos de 74%. Na caracterização de sinais de voz com rugosidade as melhores características foram <i>jitter</i> , <i>shimmer</i> , <i>Wavelet Packet</i> com 1 nível de decomposição, amplitude do <i>pitch</i> e frequência fundamental, obtendo 73%.
Lopes <i>et al.</i> (2016)	1. Frequência fundamental; 2. <i>Jitter</i> ; 3. <i>Shimmer</i> ; 4. GNE; 5. Três primeiros formantes.	Base de dados: 302 pacientes do sexo feminino, com queixa Vocal; Utilizaram a vogal /É/ (“é”) sustentada; Avaliação perceptivo-auditiva pela Escala Analógica Visual (EAV); A classificação foi realizada através da análise discriminante quadrática utilizando a média e o desvio padrão das características extraídas.	A medida acústica GNE, mostrou-se a única capaz de discriminar a intensidade do desvio vocal e a qualidade vocal predominante, com acurácia superior a 71%. Houve um ganho no desempenho da classificação com a combinação das medidas acústicas tradicionais e formânticas atingindo acurácia de 84%.

Os trabalhos aqui citados, que utilizaram a análise por predição linear e a análise cepstral classificam sinais de vozes patológicas, sendo que ainda não foi encontrado na literatura pesquisas que utilizem as medidas tradicionais, aliadas à análise por predição linear, a análise cepstral e análise mel-cepstral para realizar a classificação de sinais que apresentam desvio vocal.

3. MATERIAIS E MÉTODOS

Nesta pesquisa é realizado um estudo sobre a avaliação da qualidade vocal utilizando parâmetros extraídos através da análise acústica. A base de dados e os métodos utilizados para extração das medidas dos sinais, bem como os métodos de classificação empregados estão descritos a seguir.

3.1 Base de Dados – LIEV/UFPB

Esta base foi desenvolvida e disponibilizada pelo Laboratório Integrado de Estudos da Voz (LIEV) da Universidade Federal da Paraíba, situada no Campus João Pessoa, Paraíba. A base faz parte de um projeto avaliado e aprovado pelo Comitê de Ética em Pesquisa do Centro de Ciências da Saúde/UFPB (parecer número 52492/12) (LOPES *et. al*, 2016). Os pacientes possuem faixa etária entre 18 e 65 anos, e apresentam queixa vocal, além de já terem realizado avaliação laringológica, com laudo otorrinolaringológico. A base de dados em questão possui mais de 900 amostras de sinais de voz. 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.

Foram extraídos dos pacientes sinais de vozes para vogal sustentada /ɛ/ (“é”). A coleta dos dados foi realizada em um ambiente tratado acusticamente (cabine de gravação), com ruído ambiental inferior a 50dB NPS, a taxa de amostragem foi de 44.100 Hz. Para a gravação, o microfone situou-se a uma distância média de 10 cm dos lábios. Posteriormente ao momento da coleta, 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. Esses sinais, inicialmente, foram classificados com análise perceptivo-auditiva por meio da escala analógico visual (EAV) de acordo com o grau geral de intensidade do desvio vocal onde, o grau 1 corresponde a voz saudável, grau 2 corresponde a voz com desvio leve e grau 3 corresponde a voz com desvio moderado. Não há, nessa base de dados, casos de sinais classificados como grau geral 4 (com desvio intenso).

Para este estudo foram selecionados, da base de dados 120 amostras de sinais apenas do sexo feminino por haver influência da variável sexo na frequência fundamental e nas frequências formânticas, e que segundo alguns estudos a incidência de distúrbios vocais

ocorre principalmente em mulheres (MOTA *et al.*, 2012; LOPES, *et al.*, 2016). 30 sinais são de vozes saudáveis e 90 sinais são de vozes desviadas, sendo 30 sinais são de vozes rugosas, 30 sinais de vozes soprosas e 30 sinais de vozes tensas. Na Tabela 2 são apresentados os valores médios e o desvio-padrão da avaliação perceptivo-auditiva realizada por meio da EAV para as vozes selecionadas, de acordo com as tabelas apresentadas no Apêndice A.

Tabela 2 - Valores médios da avaliação perceptivo auditiva com a Escala Analógica Visual (EAV) para cada classe.

Classes	Média e Desvio padrão – EAV
Rugosa	53,03 ± 15,10
Soprosa	56,88 ± 15,87
Tensa	52,50 ± 12,01

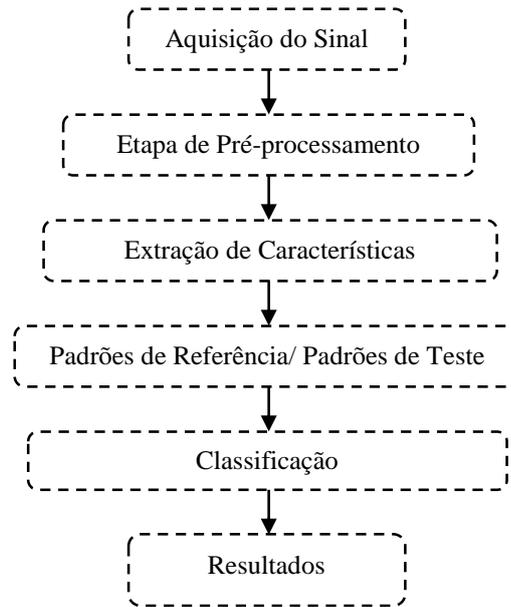
3.2 Metodologia

Inicialmente, é a aquisição do sinal que, nesta pesquisa foram obtidas da base dados apresentada na seção anterior. A seguir é realizada a etapa de pré-processamento, na qual o sinal é segmentado em quadros de 20ms, assegurando a sua estacionaridade, no intervalo considerado. Usualmente, são utilizados de 16 a 32ms (RABNER & SCHAFER, 1978). É realizado o janelamento (*Hamming*) para limitar a extensão do sinal de voz em análise. É realizada a sobreposição de 50% no janelamento e feita a pré-ênfase. A próxima etapa é a extração de características. Nesta pesquisa, foram consideradas a frequência fundamental, as frequências formânticas, o *jitter*, o *shimmer*, o GNE, os coeficientes LPC, os coeficientes cepstrais e os coeficientes mel-cepstrais. A frequência fundamental, as frequências formânticas, os coeficientes LPC, cepstrais e mel-cepstrais foram extraídos através de rotinas implementadas no *MATLAB*® R2015a e o *jitter*, *shimmer* e GNE foram extraídos através do *software VoxMetria* 3.0. Os coeficientes LPC, cepstrais e mel-cepstrais foram submetidos a quantização vetorial, com *N* níveis (variando entre 16 e 1024) e dimensão *k* = 46.

Esses valores foram armazenados como padrões de referência onde são utilizados pelo classificador MLP (*Multilayer Perceptron*) para realizar a classificação dos sinais em saudável ou desviado, ou ainda, discriminando o desvio vocal (rugosidade, soprosidade e tensão). Posteriormente, é realizada uma avaliação de desempenho do método considerado.

Na Figura 9 é apresentado o diagrama em blocos da metodologia empregada neste trabalho.

Figura 9 - Diagrama em Blocos da Metodologia empregada.



Fonte: Autoria própria.

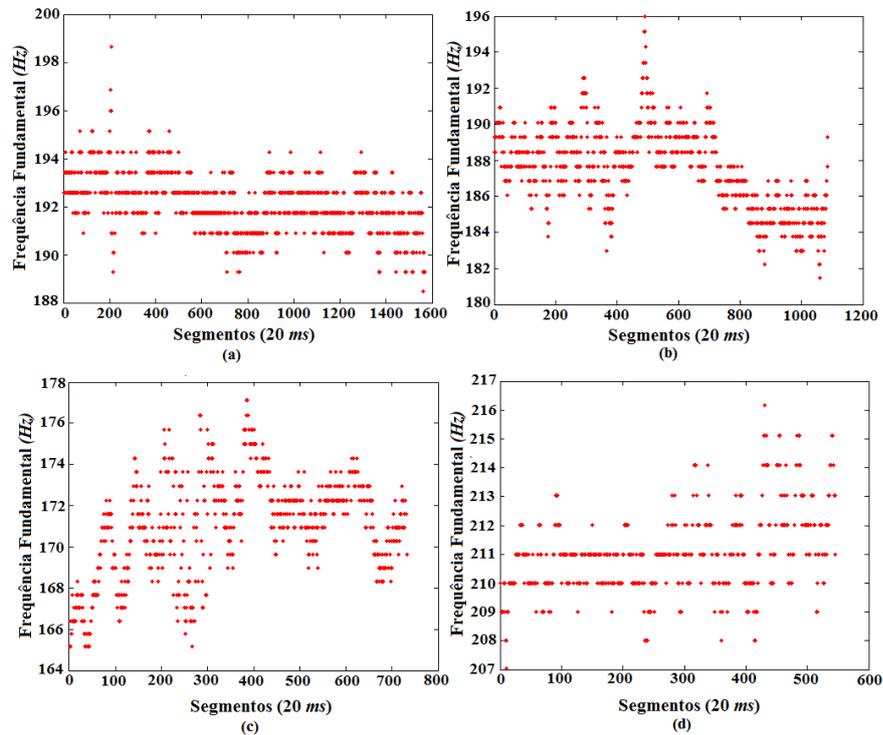
A seguir serão descritos os métodos de extração de característica utilizados neste trabalho.

3.2.1 Frequência fundamental

A frequência fundamental foi extraída a partir do método da AMDF, implementado em rotinas do *MATLAB*® *R2015a*, versão 8.5. A AMDF foi calculada para cada segmento. Posteriormente, foi realizada a média das frequências segmentais para encontrar a F_0 média do sinal de voz. Na Figura 10 pode ser observado o comportamento encontrado para a F_0 por segmento de um sinal de voz saudável e sinais de voes com presença de desvio vocal (rugosidade, tensão ou sopro).

A média e o desvio padrão dos valores da F_0 encontrada para cada classe da base de dados utilizada estão apresentados na Tabela 3.

Figura 10 – Frequência Fundamental por segmento para sinal de voz: (a) rugoso (b) tenso (c) soproso e (d) saudável.



Fonte: Autoria própria.

Tabela 3 - Valores médios e desvio da frequência fundamental (F_0) encontrado para cada classe.

Classes	Média dos valores de F_0 (Hz)
Saudável	$194,57 \pm 22,70$
Rugosa	$192,72 \pm 29,88$
Soprosa	$202,80 \pm 18,82$
Tensa	$187,31 \pm 22,96$

Observa-se, pelos gráficos da Figura 10 e da Tabela 3 que os sinais tensos apresentam frequência fundamental mais baixa e os sinais soprosos, ao contrários, F_0 mais altas. O maior desvio, no entanto, está presente nas vozes soprosas. Pela Tabela 3 observa-se que a frequência fundamental média apresentou valores bem próximos. Sendo assim, a utilização de F_{0med} sozinho não é capaz de classificar os desvios vocais.

3.2.2 Extração do jitter

A extração da medida de perturbação *jitter* foi realizada através do programa *Voxmetria 3.0*. No *VoxMetria* o *jitter percentual* é calculado usando a definição do quociente de perturbação da Frequência (FPQ – *Frequency Perturbation Quotient*) de uma sequência de períodos (KASUYA, ENDO, SALIU, 1993), conforme a Equação (3). O valor limite de

normalidade estabelecido pelo programa *VoxMetria* para o *jitter* é de até 0,6%. A média e o desvio padrão dos valores do *jitter* encontrado para cada classe da base de dados utilizada estão apresentados na Tabela 4.

Tabela 4 - Valores médios do *jitter* encontrado para cada classe.

Classes	Média dos valores do <i>jitter</i>
Saudável	0,14 ± 0,55
Rugosa	0,51 ± 3,93
Soprosa	0,27 ± 3,03
Tensa	0,22 ± 3,41

Analisando a Tabela 4, pode-se observar que para as quatro classes de sinais analisados, embora o valor do *jitter* Médio esteja dentro do padrão considerado normal, pelo desvio padrão, observa-se a alteração presente nos sinais analisados.

3.2.3 Extração do *shimmer*

A extração da medida de perturbação *shimmer* foi realizada através do programa *VoxMetria 3.0*. Ele calcula o *shimmer* usando a definição do Quociente de Perturbação da Amplitude (*Amplitude Perturbation Quotient* - APQ) de uma sequência de impulsos glotais (KASUYA *et al.*, 1993) conforme a Equação (4). O valor limite de normalidade estabelecido pelo programa *VoxMetria* para o *shimmer* é de até 6,5%. A média e desvio padrão dos valores do *shimmer* encontrado, para cada classe da base de dados utilizada, estão apresentados na Tabela 5. Observa-se a normalidade, considerando-se o desvio padrão, apenas para os sinais identificados como saudáveis. O desvio padrão para os sinais desviados é maior que o limite da normalidade de 6,5%, mostrando o distúrbio nesses sinais.

Tabela 5 - Valores médios do *shimmer* encontrados para cada classe.

Classes	Média dos valores do <i>shimmer</i>
Saudável	2,97 ± 2,38
Rugosa	6,34 ± 8,19
Soprosa	6,00 ± 5,68
Tensa	6,45 ± 7,83

3.2.4 Extração do *glottal to noise excitation ratio* (GNE)

A extração da medida do GNE também foi realizada através do programa *VoxMetria 3.0*. Os valores para o GNE são considerados normais quando maiores ou iguais a 0,5 dB. A

média e desvio padrão dos valores do GNE encontrado para cada classe da base de dados utilizada estão apresentados na Tabela 6.

Tabela 6 - Valores médios do GNE encontrado para cada classe.

Classes	Média dos valores do GNE
Saudável	$0,93 \pm 0,10$
Rugosa	$0,82 \pm 0,19$
Soprosa	$0,60 \pm 0,20$
Tensa	$0,85 \pm 0,15$

Analisando a Tabela 6, pode-se observar que para as quatro classes o valor do GNE estão dentro da normalidade.

3.2.5 Extração dos coeficientes LPC, cepstrais e mel-cepstrais

Os coeficientes LPC, cepstrais e mel-cepstrais foram extraídos através de programa implementado no *software MATLAB® R2015a, versão 8.5* e armazenado em uma matriz de coeficientes (cada linha da matriz corresponde a um segmento do sinal). Estas matrizes encontradas foram concatenadas para realizar a quantização vetorial (QV).

A quantização vetorial é uma técnica clássica de compressão de dados (FECHINE, 2000). Trata-se de uma técnica que normalmente acarreta perdas, uma vez que um dicionário de tamanho limitado pode não permitir a recuperação da informação original na íntegra (HAIKIN, 2001).

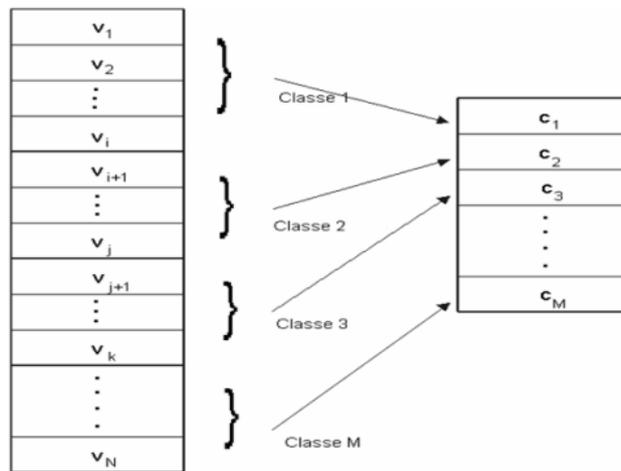
A quantização vetorial é realizada para a criação de dicionários que contém um conjunto de dados $\{c_1, c_2, \dots, c_M\}$, de tamanho finito M , onde $M < N$, composto por vetores de dimensão k . Este dicionário possui dados extraídos do conjunto de entrada $\{v_1, v_2, \dots, v_N\}$ constituídos por N vetores também de dimensão k . Este dicionário é chamado de *codebook* e os seus vetores c_i são denominados *codevectors*. Com o projeto do *codebook* deseja-se minimizar a perda de informação do conjunto de dados codificados (SIMÕES *et al.*, 2009).

Para gerar o *codebook*, utiliza-se um conjunto de dados de treinamento representativo dos dados a serem posteriormente codificados. Os dados são agrupados em um subconjunto de vetores próximos entre si (chamados classes ou *clusters*), de acordo com algum critério de proximidade. No caso desta pesquisa, será utilizada a distância euclidiana como critério. O número de classes utilizadas depende dos dados de treinamento, que pode ser determinado ou não, gerando assim um *codebook* que possua tamanho previamente conhecido. Quanto maior o número de classes, menor o erro de quantização médio encontrado. Posteriormente à criação do *codebook*, é feita uma varredura em busca do *codevector* mais próximo, associando ao vetor de entrada o índice relativo a este *codevector*. Dessa forma, o conjunto de vetores de

entrada é transformado em um conjunto de índices do *codebook* (MAKHOUL *et al.*, 1985; SIMÕES *et al.*, 2009).

Com o *codebook* criado para cada classe são calculadas as medidas de distorção, para realizar a classificação dos sinais. A medida de distorção é uma função de atribuição de um valor não negativo para o par entrada/saída de um sistema (FECHINE, 2000). Muitas medidas de distorção são propostas, tais como: Medida de Distorção do Erro Médio Quadrático (MAKHOUL *et al.*, 1985), Erro Médio Quadrático Ponderado, Medida de Distorção de Itakura-Saito (LINDE *et al.*, 1980), dentre outras. Na Figura 11 é apresentada a estrutura da criação de um *codebook*.

Figura 11 - Representação do Codebook.



Fonte: Adaptado de SIMÕES, 2009.

Para o projeto do dicionário do quantizador vetorial, neste trabalho, é utilizado o algoritmo LBG (*Linde-Buzo-Gray*) (LINDE *et al.*, 1980). O dicionário inicial é constituído a partir de um conjunto de amostras iniciais da sequência de treino, tomadas de forma aleatória a partir da matriz dos coeficientes LPC, cepstrais e mel-cepstrais concatenados. São criados diversos *codebooks*, onde cada um possuirá uma determinada quantidade de níveis (16, 32, 64, 128, 256, 512 e 1024 níveis). É realizado este teste visando investigar qual número de níveis fornece um resultado que melhor se adapte às características do sistema de classificação e que possua um custo computacional menos elevado. A medida de distorção utilizada no quantizador vetorial é a medida de Distância do Erro Médio Quadrático mínimo. É uma medida simples e comum, por sua simplicidade e tratamento matemático (FECHINE, 2000). Esta distorção é calculada pela Equação (17)

$$d(v_N, c_M) = \frac{1}{K} \sum_{i=1}^K |v_i - c_i|^2 \quad (17)$$

sendo $d(v_N, c_M)$ a distorção do erro médio quadrático, K a dimensão dos espaços Euclidianos de entrada e de reprodução, v_i corresponde aos vetores de entrada e c_i corresponde aos *codevectors*.

Nesta pesquisa, a quantização vetorial é realizada utilizando uma matriz de vetores concatenados contendo os coeficientes LPC, cepstrais e mel-cepstrais para cada classe. A ordem do preditor ($p = 46$) neste trabalho para todos os coeficientes corresponde à dimensão do quantizador. Esta técnica é utilizada na pesquisa para reduzir a dimensão dos dados para a etapa de classificação dos sinais em saudável ou desviada. Através da quantização vetorial é gerado o *codebook* (dicionário) para cada classe, para determinar a similaridade entre as elocuições dos sinais a serem analisados (COSTA, S., 2008).

3.2.6 Extração das frequências formânticas

As frequências formânticas foram extraídas através de programa implementado no *software MATLAB® R2015a, versão 8.5*, onde foi realizada a Transformada de Fourier a curto intervalo de tempo dos coeficientes LPC (para cada segmento). Posteriormente, foi realizada a média das transformadas segmentais e detectados quais os três primeiros picos, representando o primeiro, segundo e terceiro formante respectivamente. As frequências formânticas são dadas pelos valores médios segmentais encontrados em Hz.

Na Tabela 7 é apresentada a média e o desvio padrão dos valores encontrados para as três primeiras frequências formânticas para cada classe de sinais utilizados, onde F_1 representa o primeiro formante, F_2 o segundo formante e F_3 o terceiro formante. Os sinais que apresentam sopro, apresentam o maior desvio padrão no terceiro formante, que os outros desvios analisados.

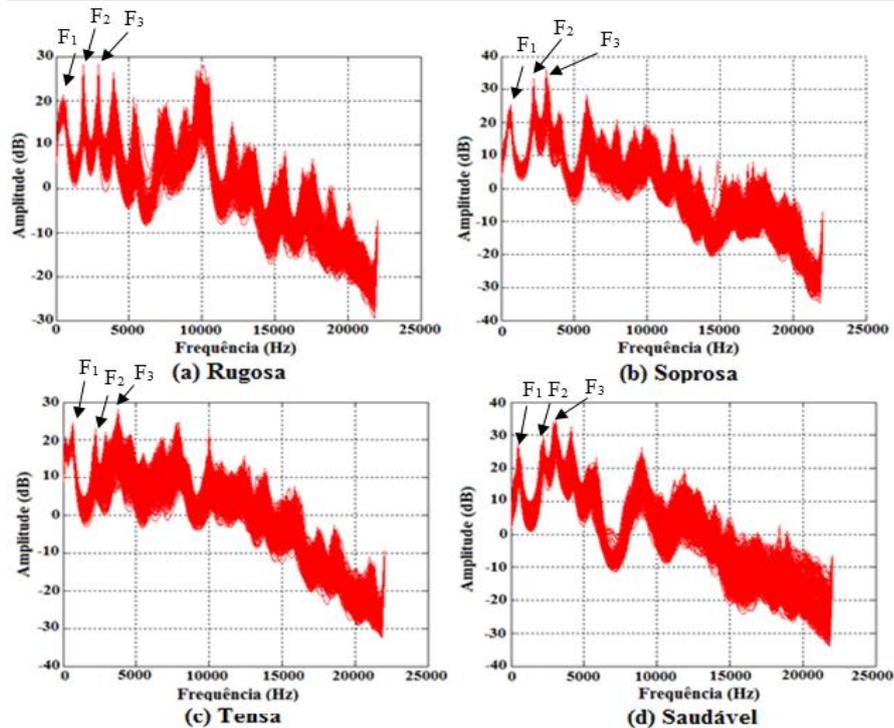
Tabela 7 - Valores médios encontrados para as três primeiras frequências formânticas (F_1 , F_2 , F_3) de cada classe.

Classes	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)
Saudável	522,76 ± 100,34	2185,13 ± 241,82	2970,34 ± 386,62
Rugosa	514,01 ± 110,41	2052,80 ± 278,48	2890,59 ± 236,16
Sopro	516,80 ± 157,70	2028,33 ± 405,47	2992,03 ± 588,54
Tensa	575,64 ± 149,16	2114,38 ± 327,04	3006,39 ± 296,33

F_1 : primeiro formante; F_2 : segundo formante; F_3 : terceiro formante.

Na Figura 12 são apresentadas as frequências formânticas dados pelos picos do espectro LPC para cada segmento de um sinal rugoso, sopro, tenso e voz saudável. Observa-se na Figura 12 que cada segmento corresponde a um vetor.

Figura 12 – Espectro da Frequência formânticas para cada segmento do sinal de voz: (a) rugoso, (b) soproso, (c) tenso, e (d) saudável.



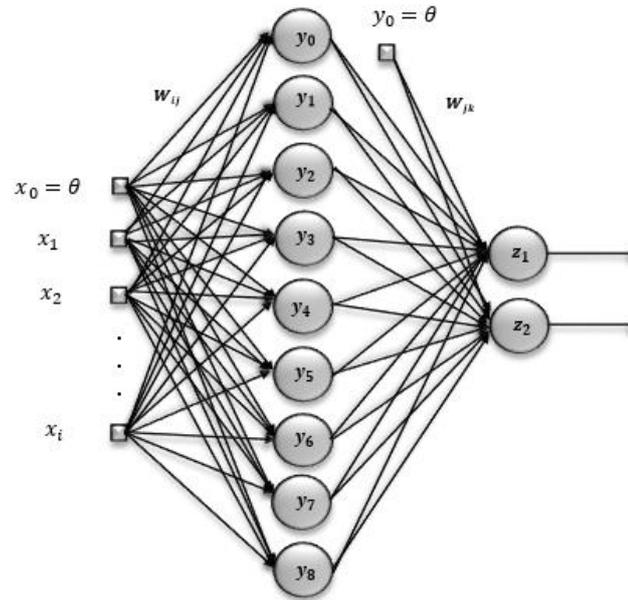
Fonte: Autoria própria.

3.2.7 Processo de classificação

Nesta pesquisa foi utilizado um classificador MLP (*Multilayer Perceptron*) com o algoritmo de aprendizado supervisionado do Gradiente Conjugado Escalonado (SCG) (SIMSEK *et al.*, 2016). Este algoritmo é uma variação do algoritmo do Gradiente Conjugado e é baseado no trabalho de Moller (MOLLER, 1993), e tem por objetivo diminuir o tempo de convergência e realizar a otimização da rede, através de informações da derivada de segunda ordem da função de erro (PINHATI, 2014). Este classificador é uma rede neural *feedforward* formada por unidades de processamento interligadas conhecidas como neurônios, e tem a tendência natural para armazenar conhecimento experimental e torná-lo disponível para uso. Cada camada de neurônios se conecta à próxima camada, sem caminho de volta (HAYKIN, 2001).

Na Figura 13 encontra-se a topologia da rede, onde a quantidade de neurônios na camada de entrada depende da quantidade de características selecionadas, na camada oculta são empregados 09 neurônios e na camada de saída 02 neurônios. Os pesos da rede W_{ij} e W_{jk} são ajustados automaticamente.

Figura 13 - Rede *feedforward* MLP utilizada.



Fonte: Autoria própria.

São considerados sete casos de classificação: (i) Vozes Desviadas Vs. Saudável, em que na classe de Vozes Desviadas estão incluídos os sinais rugosos, soprosos e tenso; (ii) Rugoso Vs. Saudável; (iii) Tenso Vs. Saudável; (iv) Soproso Vs. Saudável; (v) Rugoso Vs. Tenso; (vi) Rugoso Vs. Soproso e (vii) Tenso Vs. Soproso.

Foram consideradas três formas de classificação: 1) medidas tradicionais; 2) classificação com os *codebooks* dos coeficientes LPC, cepstrais e mel-cepstrais e 3) classificação com as médias segmentais dos coeficientes LPC, cepstrais e mel-cepstrais.

Os sinais de vozes utilizados foram separados, aleatoriamente, em dois conjuntos distintos, dos quais 2/3 foram utilizados para treinamento e 1/3 para teste. Para maior confiabilidade, a rede neural foi executada 10 vezes, a fim de realizar a validação cruzada.

Para os testes foram realizadas diversas combinações das características de entrada, a fim de encontrar a combinação que pudesse maximizar os resultados. Na Tabela 8 são apresentadas as características de entrada para cada caso de classificação.

Tabela 8 - Características de Entrada para os casos de classificação.

Caso de Classificação	Características de Entrada
Medidas Tradicionais	(i) Frequência Fundamental Média (F_{0med}),
	(ii) 1º Formante Médio (F_{1med}),
	(iii) 2º Formante Médio (F_{2med}),
	(iv) 3º Formante Médio (F_{3med})
	(v) <i>Jitter</i>
	(vi) <i>Shimmer</i> , e
	(vii) <i>GNE</i> .
Codebooks dos coeficientes LPC, cepstrais e mel-cepstrais	(i) Coeficientes LPC;
	(ii) Coeficientes cepstrais; e
	(iii) Coeficientes mel-cepstrais
Médias segmentais dos coeficientes LPC, cepstrais e mel-cepstrais	(i) Média dos coeficientes LPC (LPC_{med}),
	(ii) Média dos coeficientes cepstrais ($CEPS_{med}$) e
	(iii) Média dos coeficientes mel-cepstral (MEL_{med})

Os coeficientes LPC, os coeficientes cepstrais e os coeficientes mel-cepstrais foram empregados no classificador separadamente. Foram observadas as medidas de desempenho acurácia, sensibilidade e especificidade. Os resultados são descritos como os valores médios e seus respectivos desvios padrões.

A acurácia mede a capacidade do classificador de identificar corretamente a presença ou a ausência do desvio, como pode ser visto na Equação (18).

$$ACURÁCIA = \frac{CA+CR}{N} \times 100 \quad (18)$$

onde CA representa a correta aceitação, ou seja, a presença do desvio é detectada quando está realmente presente e CR representa a correta rejeição, ou seja, a classificação correta da ausência do desvio e N corresponde a quantidade total de sinais utilizados na classificação.

A sensibilidade representa a proporção de pessoas com o desvio de interesse que têm o resultado do teste positivo, indicando o quão bom é o teste para identificar os indivíduos com o desvio vocal. É dada pela Equação (19).

$$SENSIBILIDADE = \frac{CA}{CA+FR} \times 100, \quad (19)$$

em que CA representa a correta aceitação, ou seja, a presença do desvio é detectada quando ela está realmente presente e FR representa a falsa rejeição ou seja, a presença do desvio é rejeitada quando de fato ela está presente.

A especificidade representa a proporção de pessoas sem a doença, cujo teste dá negativo, indicando o quão bom é o método empregado na identificação dos sinais de vozes não desviadas. A especificidade é dada pela Equação (20).

$$ESPECIFICIDADE = \frac{CR}{CR+FA} \times 100, \quad (20)$$

onde CR representa a correta rejeição, ou seja, é detectada corretamente a ausência do desvio e FA representa a falsa aceitação ou seja, quando é detectada a presença do desvio quando ela não está presente.

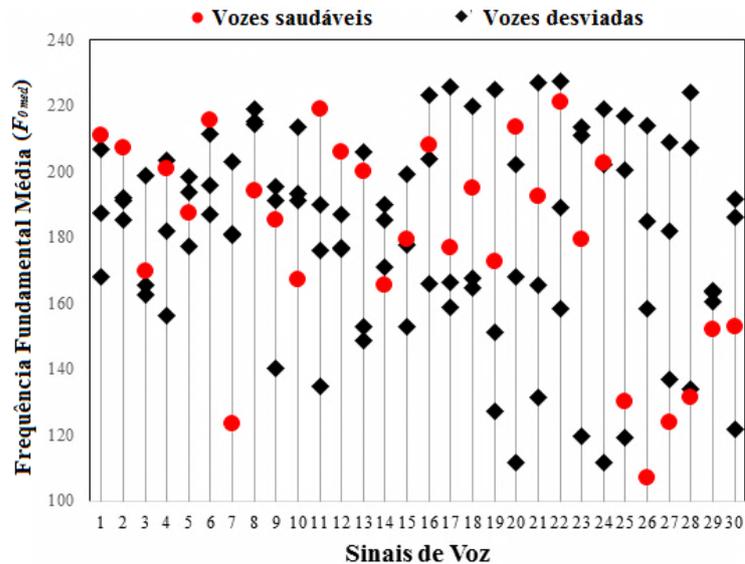
4. RESULTADOS

A seguir são descritos os resultados obtidos os métodos empregando: medidas tradicionais, Análises LPC, cepstral e mel-cepstral.

4.1 Caracterização Acústica dos Desvios Vocais com as Medidas Tradicionais

Nesta Seção serão apresentados os gráficos com os valores médios encontrados para os métodos tradicionais, comparando os sinais desviados (sinais rugosos, soprosos e tensos) com os sinais saudáveis. Na Figura 14 está apresentado o comportamento da frequência fundamental para os sinais desviados e saudáveis.

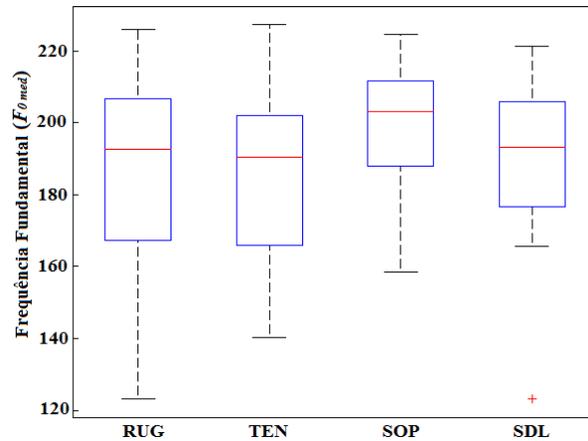
Figura 14 – Comportamento da frequência fundamental para vozes saudáveis e vozes desviadas.



Na Figura 14 observa-se que, através da frequência fundamental, não é possível distinguir os sinais de vozes saudáveis dos sinais desviados, pois os valores encontrados variam de 100 Hz a 230 Hz para ambas as classes.

Na Figura 15 estão ilustrados os *boxplots* com os resultados para a frequência fundamental, ilustrando os valores da média e do desvio padrão encontrado para cada classe: rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).

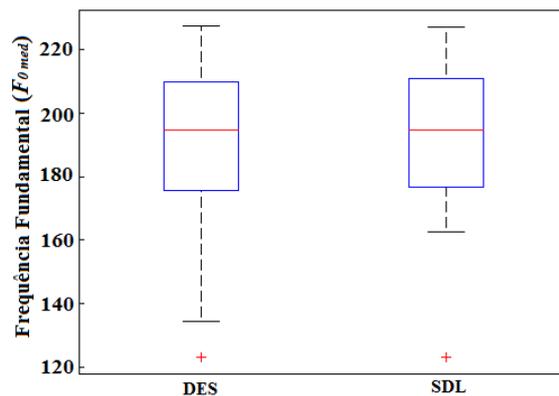
Figura 15 - Boxplots da frequência fundamental para sinais de voz rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).



Analisando os *boxplots* da Figura 15, observa-se que as medianas encontradas para a frequência fundamental diferenciam apenas os sinais soprosos, sendo que as classes de sinais rugosos, tensos e saudáveis apresentam mediana semelhante. A dispersão apresentada para os sinais rugosos é superior a dispersão apresentada para os outros sinais, onde a classe que apresentou menor dispersão nos valores da frequência fundamental foram os sinais saudáveis.

Na Figura 16 estão ilustrados os *boxplots* com os resultados para a frequência fundamental, ilustrando os valores da média e do desvio padrão encontrado para as vozes desviadas (DES) e saudáveis (SDL).

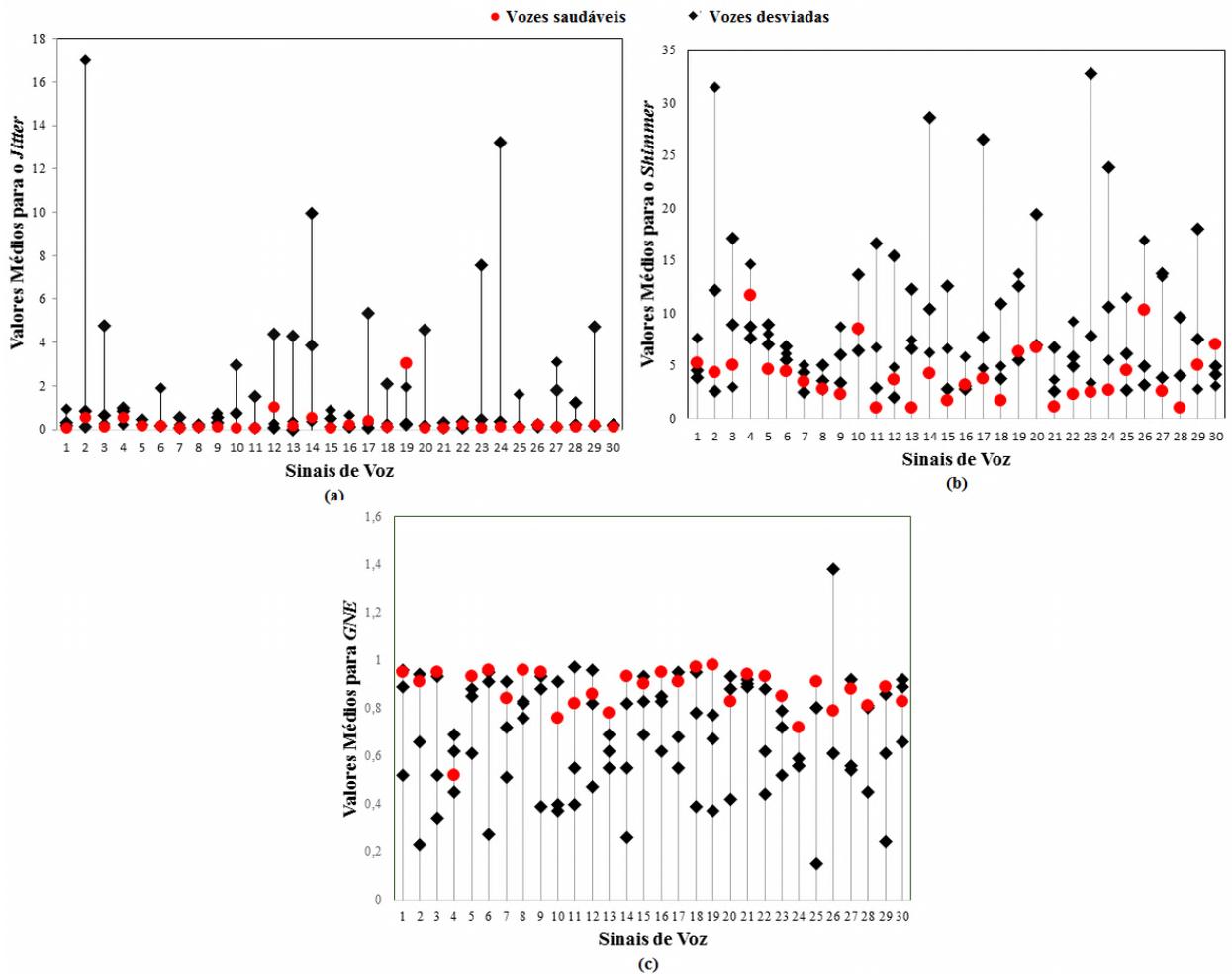
Figura 16 - Boxplots da frequência fundamental para os sinais de vozes desviadas (DES) e saudáveis (SDL).



Analisando os *boxplots* da Figura 16, observa-se que a frequência fundamental para os sinais desviados possuem dispersão em torno da mediana maior que os sinais saudáveis, em que a mediana é semelhante para as duas classes.

Na Figura 17 pode ser observado o comportamento do *jitter*, do *shimmer* e do GNE para os sinais desviados e saudáveis.

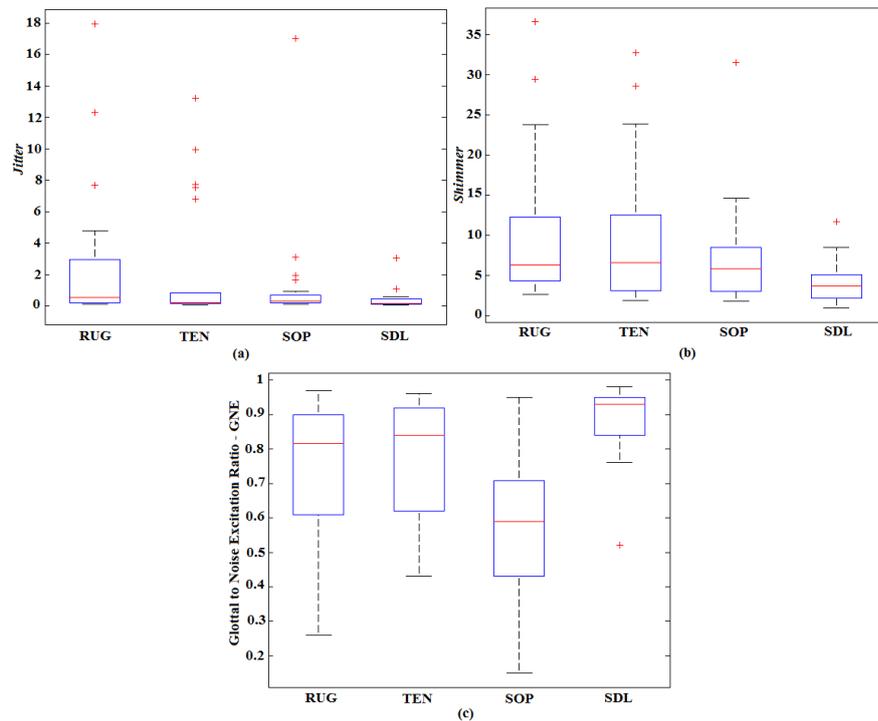
Figura 17 - Comportamento das medidas: (a) *jitter* (b) *shimmer* e (c) GNE para vozes saudáveis e vozes desviadas.



Na Figura 17 (a) observa-se que nos valores do *jitter*, para os sinais desviados, há uma grande variação quando comparados aos sinais saudáveis. Na Figura 17 (b) observa-se que a maioria dos valores obtidos para o *shimmer* são maiores para os sinais desviados comparados aos sinais saudáveis, pois representa a variação na intensidade dos ciclos adjacentes de vibração das pregas vocais, estando correlacionado com a presença de ruído à emissão e com a soprosidade (TEIXEIRA, FERREIRA, CARNEIRO, 2011). Na Figura 17 (c) observa-se que o GNE na maioria dos casos apresenta valores menores que um (onde quanto mais próximo de 1, mais o sinal se comporta como saudável) para vozes desviadas, sendo resultante de uma excitação do tipo ruído.

Na Figura 18 estão ilustrados os *boxplots* com os resultados para o *jitter*, *shimmer* e GNE, ilustrando os valores da média e do desvio padrão encontrado para cada classe: rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).

Figura 18 - Boxplots do (a) jitter, (b) shimmer e (c) GNE para sinais de voz rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).

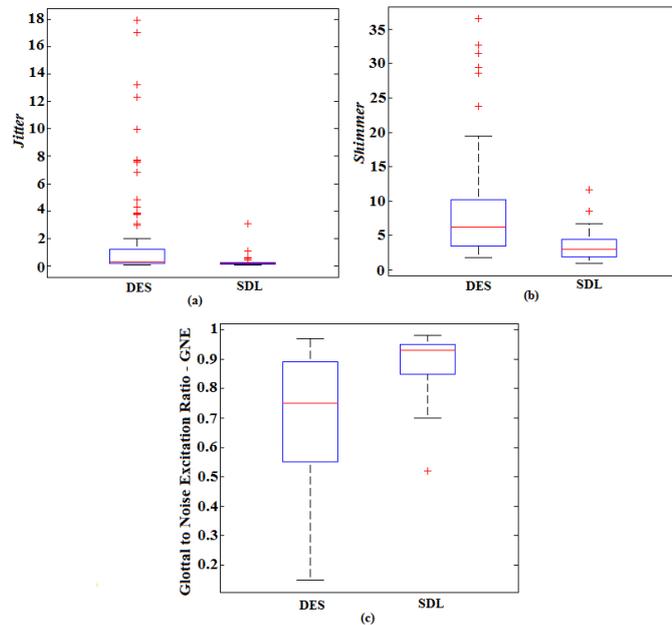


Observa-se, pelos *boxplots* da Figura 18 (a), que o *jitter* apresentou maior dispersão para a classe rugosa e menor dispersão para as demais classes. Quanto à mediana, observou-se que não é possível distinguir as classes, pois apresentou valores semelhantes. Os *boxplots* da Figura 18 (b) demonstram que há uma maior dispersão dos valores do *shimmer* para as classes de sinais tensos e rugosos. Observa-se, também, que a alta dispersão diminui o desempenho da classificação. Na Figura 18 (c) observa-se que o GNE apresentou medianas que possibilitem a distinção entre as classes. Houve maior dispersão para os sinais soprosos e menor dispersão para os sinais saudáveis.

Na Figura 19 estão ilustrados os *boxplots* com os resultados para o *jitter*, *shimmer* e GNE, ilustrando os valores da média e do desvio padrão encontrado para as classes desviadas (DES) e saudáveis (SDL).

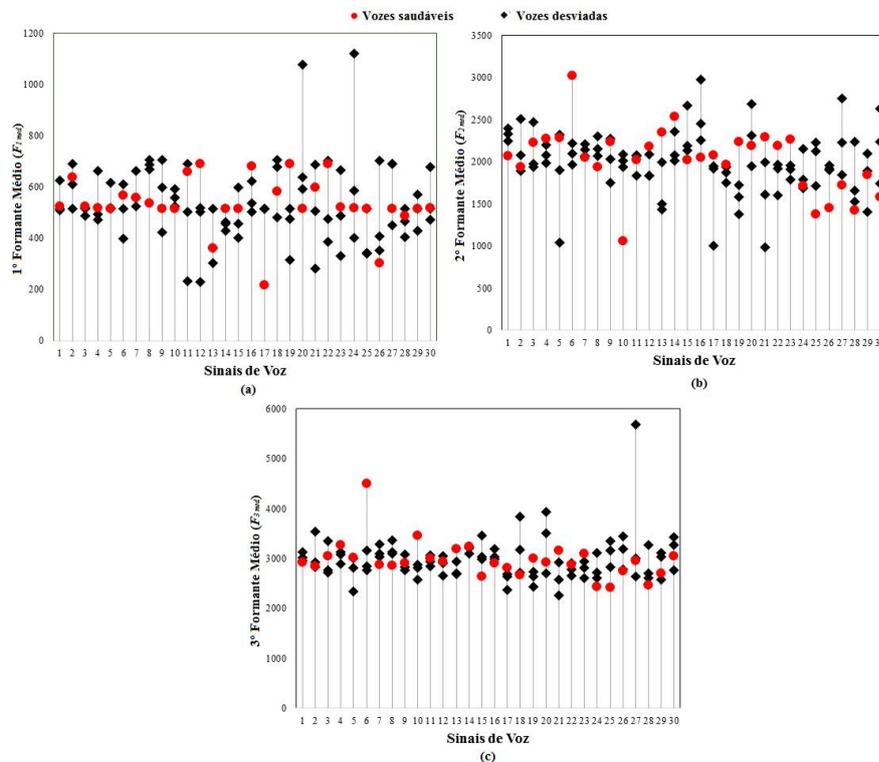
Os *boxplots* da Figura 19 (a) apresentam maior dispersão para os sinais desviados e menor dispersão para sinais saudáveis. Quanto à mediana, observou-se que não é possível distinguir as classes, pois apresentou valores semelhantes. Os *boxplots* da Figura 19 (b) demonstram que há uma maior dispersão dos valores do *shimmer* para as classes de sinais desviados. Na Figura 19 (c) observa-se que o GNE apresentou medianas que possibilitem a distinção entre as classes e houve maior dispersão para os sinais desviados e menor dispersão para os sinais saudáveis.

Figura 19 - Boxplots do (a) jitter, (b) shimmer e (c) GNE para sinais de voz desviadas (DES) e saudável (SDL).



Na Figura 20 pode ser observado o comportamento do F_{1med} , F_{2med} e F_{3med} para os sinais desviados e saudáveis.

Figura 20 - Comportamento das medidas: (a) F_{1med} , (b) F_{2med} e (c) F_{3med} para vozes saudáveis e vozes desviadas.

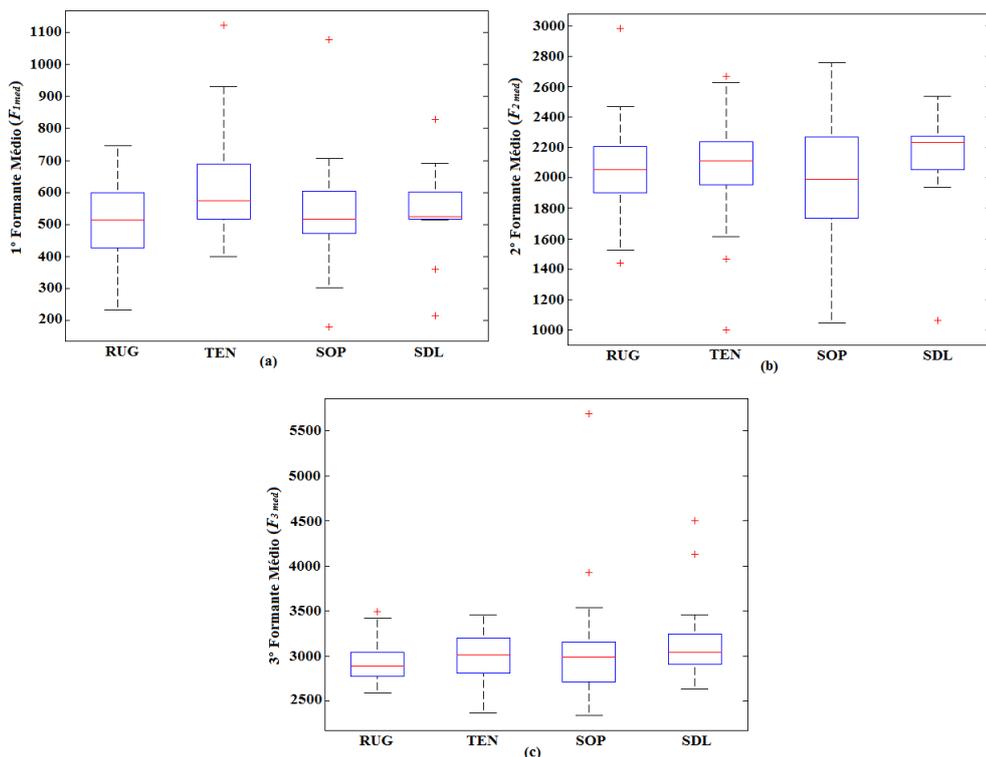


Observa-se, na Figura 20, que os valores para (a) F_{1med} , para (b) F_{2med} e para (c) F_{3med} sozinho não são bons para a classificação entre vozes desviadas e vozes saudáveis, no caso dos sinais selecionados, pois houve uma variação considerável para os valores encontrados.

Na Figura 21 estão ilustrados os *boxplots* com as medidas do F_{1med} , F_{2med} e F_{3med} , ilustrando os valores da média e do desvio padrão encontrados para cada classe: rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).

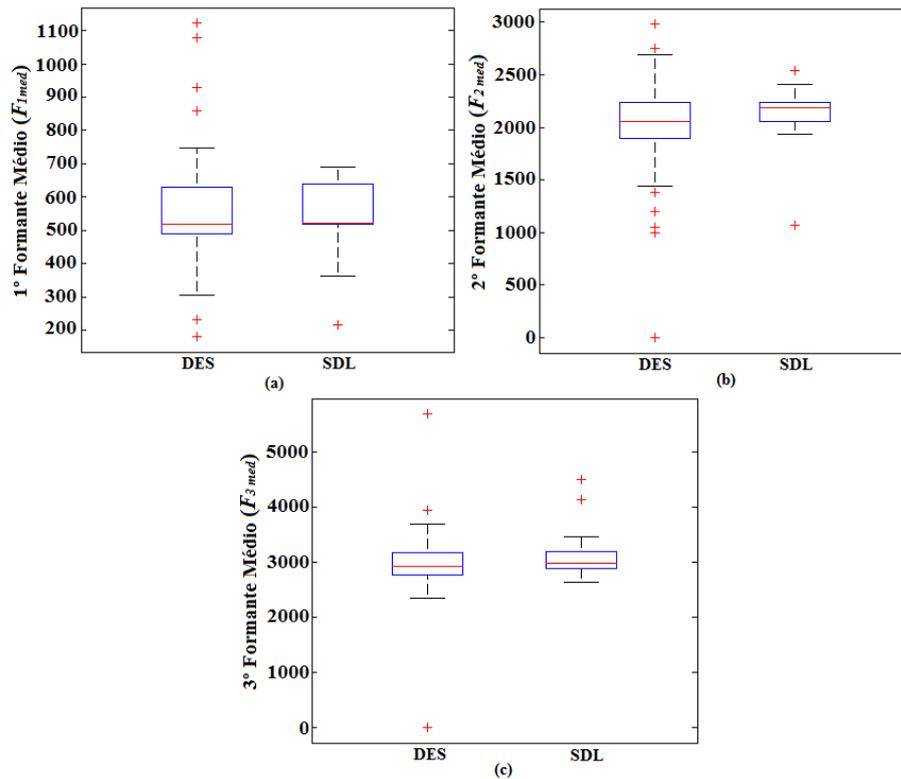
Na Figura 21 (a) os *boxplots* para o F_{1med} não demonstram boa classificação entre os sinais rugosos, soprosos e saudáveis. Os valores obtidos para os sinais saudáveis apresentaram menor dispersão quando comparados com os demais sinais. Pode ser observado na Figura 21 (b) que os sinais apresentaram os valores das medianas semelhantes utilizando o F_{2med} para os sinais rugosos, tensos e saudáveis, diferentemente dos sinais soprosos. Os sinais soprosos possuem maior dispersão dos valores e os sinais saudáveis possuem menor dispersão. Os *boxplots* da Figura 21 (c) para o F_{3med} apresentam medianas semelhantes, impossibilitando a separação das classes. A menor dispersão encontrada foi para os sinais rugosos e a maior para os sinais soprosos.

Figura 21 - Boxplots (a) F_{1med} , (b) F_{2med} e (c) F_{3med} para sinais de voz rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).



Na Figura 22 estão ilustrados os *boxplots* com as medidas do F_{1med} , F_{2med} e F_{3med} , ilustrando os valores da média e do desvio padrão encontrados para as vozes desviadas (DES) e saudáveis (SDL).

Figura 22 - *Boxplots* (a) F_{1med} , (b) F_{2med} e (c) F_{3med} para sinais de voz desviada (DES) e saudável (SDL).



Na Figura 22 (a) os *boxplots* para o F_{1med} não demonstram boa classificação entre os desviados e saudáveis. Os valores obtidos para os sinais saudáveis apresentaram menor dispersão quando comparados com os sinais desviados. Pode ser observado na Figura 22 (b) que os sinais apresentaram dispersão maior utilizando o F_{2med} para os sinais desviados, diferentemente dos sinais saudáveis. Os *boxplots* da Figura 22 (c) para o F_{3med} apresentam medianas e dispersões semelhantes, dificultando a separação das classes.

4.2 Resultados da Classificação dos Desvios Vocais com as Medidas Tradicionais

Na Tabela 9 estão apresentados os melhores resultados com acurácia (ACUR), sensibilidade (SEN) e especificidade (ESP), demonstrando de forma individual e combinada a classificação entre Desviada Vs. Saudável, Rugosa Vs. Saudável, Tensa Vs. Saudável e Soprosa Vs. Saudável. A classificação foi realizada seguindo os critérios de Hosmer-Lemeshow (2000) em que taxas de classificação abaixo de 70% não são significativas (entre 70 e 80% são aceitáveis, entre 80 e 90%, excelentes e acima de 90% são consideradas excepcionais) (HOSMER & LEMESHOW, 2000). No entanto, para algumas classificações os melhores resultados obtiveram taxas de acurácia inferiores a 70%, sendo aqui demonstrados para uma posterior comparação entre os métodos aplicados na pesquisa.

Tabela 9 - Melhores resultados para a classificação com as características individuais.

Classificação	ACUR (%)	SEN (%)	ESP (%)	Características
Desviada Vs. Saudável	80,15 ± 0,98	81,88 ± 1,33	80,25 ± 1,98	<i>Shimmer</i>
Rugosa Vs. Saudável	75,37 ± 0,98	77,51 ± 1,68	76,31 ± 1,56	GNE
Tensa Vs. Saudável	73,53 ± 2,71	75,38 ± 3,13	73,04 ± 7,40	<i>Shimmer</i>
Soprosa Vs. Saudável	72,06 ± 3,25	72,73 ± 3,48	70,00 ± 0,96	GNE
Desviada Vs. Saudável	67,65 ± 2,76	74,54 ± 3,76	64,09 ± 4,73	<i>Shimmer</i>
Rugosa Vs. Saudável	80,88 ± 0,85	86,21 ± 1,83	77,81 ± 0,65	GNE

ACUR: acurácia; SEN: sensibilidade; ESP: especificidade; GNE: *Glottal to Noise Excitation Ratio*.

Observa-se pela Tabela 9 que as melhores classificações ocorreram na distinção entre vozes soprosas e saudáveis e na distinção entre vozes desviadas e saudáveis. Para o primeiro caso, a medida utilizada foi o GNE e para o segundo caso a medida utilizada foi o *shimmer*, onde ambas as classificações apresentaram acurácia acima de 80%. Entre as quatro classes, o caso de classificação com menor taxa de classificação foi para a distinção entre vozes tensas e vozes saudáveis, proporcionando uma acurácia média de 67% para a melhor medida, apresentando o *shimmer* como melhor medida.

Na Tabela 10 observam-se as matrizes de confusão para a classificação Rugosa Vs. Soprosa, apresentando uma acurácia de 68,66% utilizando como medida o GNE.

Tabela 10 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz soprosos – medidas individuais

CLASSE	RUGOSA	SOPROSA
RUGOSA	18	12
SOPROSA	11	19

Na Tabela 11 observam-se as matrizes de confusão para a classificação Rugosa Vs. Tensa, apresentando uma acurácia de 65,45% utilizando o F_{1med} como medida.

Tabela 11 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz tensos – medidas individuais.

CLASSE	RUGOSA	TENSA
RUGOSA	23	7
TENSA	11	19

Na Tabela 12 observam-se as matrizes de confusão para a classificação Tensa Vs. Soprosa, apresentando uma acurácia de 72,22% utilizando o GNE.

Tabela 12 - Matriz de confusão para a discriminação entre sinais de voz tensos e sinais de voz soprosos – medidas individuais.

CLASSE	TENSA	SOPROSA
TENSA	22	8
SOPROSA	7	23

Nas classificações realizadas entre os desvios vocais observou-se que a melhor medida individual foi o GNE realizando a classificação entre sinais tensos e soprosos. A melhor classificação com medidas individuais entre os desvios deu-se para a distinção entre sinais tensos e soprosos.

Na Tabela 13, é apresentada a classificação com as medidas combinadas. As taxas de acurácia média encontradas são superiores a 70%, onde a característica GNE, aparece em todas as combinações, associada a outras medidas para a discriminação entre Desviada Vs. Saudável, Rugosa Vs. Saudável, Tensa Vs. Saudável e Soprosa Vs. Saudável.

Tabela 13 - Melhores resultados para a classificação com as características combinadas.

Classificação	ACUR (%)	SEN (%)	ESP (%)	Características
Desviada Vs. Saudável	83,09 ± 0,98	86,19 ± 1,18	72,52 ± 2,28	F_{1med} , F_{2med} , Jitter, Shimmer, GNE
Rugosa Vs. Saudável	77,21 ± 1,05	84,31 ± 1,14	72,30 ± 1,65	F_{2med} , GNE, Shimmer
Tensa Vs. Saudável	75,00 ± 2,31	80,71 ± 2,95	70,77 ± 2,27	F_{1med} , F_{2med} , F_{3med} , F_{0med} , GNE
Soprosa Vs. Saudável	86,76 ± 0,72	87,88 ± 1,48	85,71 ± 0,75	F_{1med} , F_{2med} , GNE

ACUR: acurácia; SEN: sensibilidade; ESP: especificidade; GNE: *Glottal to Noise Excitation Ratio*; F_{med} : frequência fundamental média; F_{1med} : primeiro formante médio; F_{2med} : segundo formante médio; F_{3med} : terceiro formante médio.

Na distinção entre vozes desviadas e vozes saudáveis observa-se, pela Tabela 13, que a melhor combinação utilizou três características (F_{1med} , F_{2med} , GNE) para distinção entre vozes soprosas e vozes saudáveis, com acurácia média de 86%. A classificação entre vozes desviadas e saudáveis também obteve uma boa classificação utilizando cinco medidas (F_{1med} , F_{2med} , jitter, shimmer e GNE), apresentando acurácia superior a 80%. A combinação menos significativa de acordo com os critérios de Hosmer & Lemeshow (2000), se deu para a distinção entre vozes tensas e vozes saudáveis apresentando cinco características (F_{1med} , F_{2med} , F_{3med} , F_{0med} , GNE), obtendo acurácia de 75%, considerada aceitável. Na classificação entre desviada e saudável e entre soprosa e saudável, a acurácia é considerada excelente sob os critérios de Hosmer & Lemeshow (2000).

Na Tabela 14 observam-se as matrizes de confusão para a classificação Rugosa Vs. Soprosa, em que as características combinadas que melhor classificaram os desvios vocais foram F_{1med} , F_{2med} , *jitter*, *shimmer* e GNE, apresentando uma acurácia de 74,58%.

Tabela 14 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz soprosos – medidas combinadas.

CLASSE	RUGOSA	SOPROSA
RUGOSA	23	7
SOPROSA	7	23

Na Tabela 15 observam-se as matrizes de confusão para a classificação Rugosa Vs. Tensa, apresentando uma acurácia de 67,27% em que as características combinadas que melhor classificaram foram F_{1med} e F_{3med} .

Tabela 15 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz tensos – medidas combinadas.

CLASSE	RUGOSA	TENSA
RUGOSA	21	9
TENSA	10	20

Na Tabela 16 observam-se as matrizes de confusão para a classificação Tensa Vs. Soprosa, apresentando uma acurácia de 81,48%, considerada excelente, em que as características combinadas que melhor classificaram foram F_{0med} , F_{1med} , F_{2med} , F_{3med} , *jitter* e GNE.

Tabela 16 - Matriz de confusão para a discriminação entre sinais de voz tensos e sinais de voz soprosos – medidas individuais.

CLASSE	TENSA	SOPROSA
TENSA	25	5
SOPROSA	6	24

Na Tabela 17 é apresentado um resumo da avaliação do desempenho para a classificação entre os desvios utilizando as características tradicionais.

Tabela 17- Resumo da avaliação de desempenho na discriminações entre os desvios vocais.

CLASSE	ACURÁCIA	Medidas
RUGOSA Vs. SOPROSA	74,58%	F_{1med} , F_{2med} , <i>jitter</i> , <i>shimmer</i> e GNE
RUGOSA Vs. TENSA	67,27%	F_{1med} e F_{3med}
TENSA Vs. SOPROSA	81,48%	F_{0med} , F_{1med} , F_{2med} , F_{3med} , <i>jitter</i> e GNE

ACUR: acurácia; SEN: sensibilidade; ESP: especificidade; GNE: *Glottal to Noise Excitation Ratio*; F_{0med} : frequência fundamental média; F_{1med} : primeiro formante médio; F_{2med} : segundo formante médio; F_{3med} : terceiro formante médio.

Pode-se concluir, a partir dos resultados obtidos com as medidas tradicionais, que apresentaram uma boa classificação utilizando as medidas lineares extraídas atingindo acurácia de 86%, exceto para a distinção entre vozes rugosas e tensas. Vale ressaltar que entre as medidas tradicionais as que melhor classificaram foram as medidas de perturbação da frequência fundamental (*jitter*) e de ruído (GNE), demonstrando a presença clara dessas componentes nos desvios vocais.

Neste trabalho também são avaliadas outras medidas que não dependam da frequência fundamental como coeficientes LPC, coeficientes cepstrais e mel-cepstrais para observar como se comportam na classificação dos desvios vocais.

4.2.1 Discussão dos resultados com as medidas tradicionais

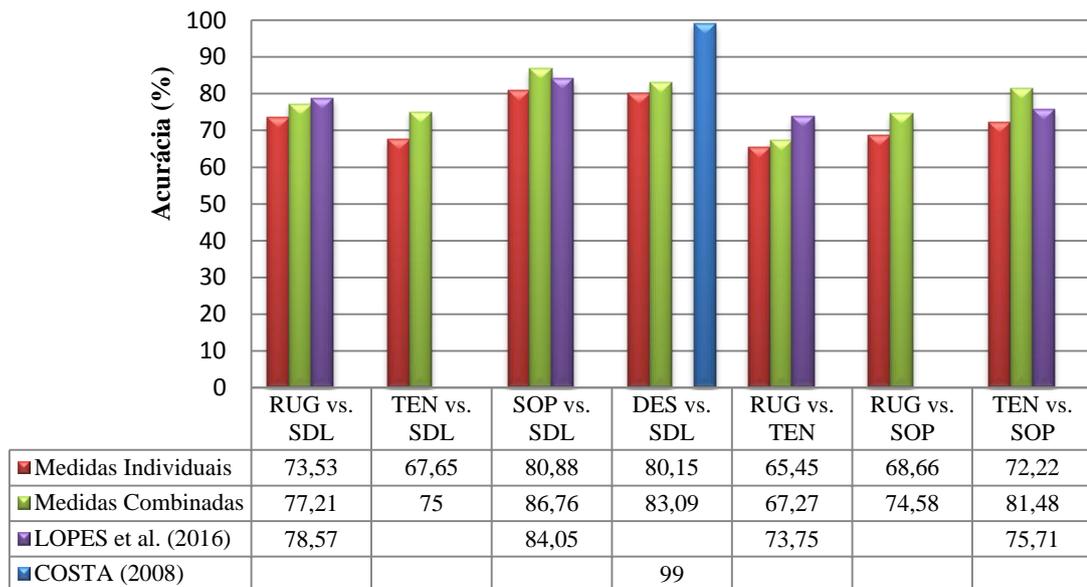
Entre as medidas tradicionais empregadas algumas obtiveram maior potencial discriminativo na classificação entre sinais desviados e saudáveis (como o *shimmer* e o GNE atingindo acurácia de 80% na classificação entre vozes soprosas e saudáveis e vozes desviadas e saudáveis), bem como entre os desvios (GNE apresentou acurácia de 72% na classificação entre sinais tensos e soprosos), onde houve maior destaque quando as medidas foram combinadas, pois a classificação atingiu acurácia de 86% na classificação entre sinais de vozes soprosas e saudáveis e de 83% na classificação entre sinais de vozes desviadas (rugosidade, tensão e sopro) e sinais saudáveis. De maneira individual as características que mais se destacaram foram o *shimmer* e o GNE e na combinação das medidas as frequências formânticas, a frequência fundamental e o *jitter* apareceram, influenciando no aumento do potencial discriminativo das medidas.

O GNE foi a medida com maior destaque. Acredita-se que isso ocorreu por que o GNE demonstra o ruído glotal presente nos sinais, em que os sinais estudados possuem como característica a sopro, influenciando na quantificação do desvio, como demonstrado na pesquisa de Lopes *et. al.* (2016). Quanto ao *shimmer*, sua presença foi mais evidente na classificação entre os sinais rugosos e saudáveis e entre os sinais desviados e saudáveis, como também foi observado por Baravieira (2016). De acordo com Costa (2008) a presença do *shimmer* demonstra uma percepção indireta do ruído na produção vocal, presente em alguns desvios.

Na Figura 23 é apresentado um gráfico comparativo dos resultados obtidos com as medidas tradicionais de maneira individual e combinada, bem como com os resultados

obtidos por Lopes *et al.* (2016) e por Costa (2008), onde para o trabalho de Costa será comparado apenas a classificação obtida para a distinção entre sinais de vozes afetados por Edema nas pregas vocais e vozes saudáveis com a classificação obtida para a discriminação entre vozes desviadas e saudáveis.

Figura 23 Comparação dos resultados obtidos com as medidas individuais, combinadas, resultados obtidos por Lopes *et al.* (2016) e por Costa (2008).



Na Figura 23 observa-se que as medidas tradicionais de maneira combinada demonstraram um potencial discriminativo melhor que as medidas individuais na maioria dos casos. Mas, na distinção entre vozes rugosas e tensas, a combinação das medidas não melhorou a discriminação, onde a acurácia foi inferior a 70%, não classificando bem os sinais. Quando comparado os resultados obtidos nesta pesquisa com os resultados de Lopes *et al.* (2016) observa-se que, em alguns casos, foi semelhante o resultado da classificação realizada (exceto para a distinção entre vozes rugosas e tensas, onde a classificação de Lopes *et al.* (2016), se mostrou melhor).

Vale ressaltar que que foram utilizados classificadores diferentes (Lopes utilizou a análise discriminante quadrática e, nesta pesquisa, foi empregada a rede MLP). Os resultados indicam que o classificador MLP empregado possui um potencial discriminativo semelhantes ao classificador Discriminante Quadrático, o que demonstra a grande relevância das características empregadas e não, exclusivamente, do classificador. Destaca-se, ainda, que

nesta pesquisa foram realizadas mais classificações que o trabalho de Lopes *et al.* Na Figura 23 onde está vazio, não foi realizada a classificação pelo(s) autor(es).

Costa, S. (2008) avaliou o desempenho da classificação na detecção de desordens vocais em sinais afetados por edema nas pregas vocais, utilizando os coeficientes LPC, cepstrais e mel cepstrais, com quantização vetorial. A autora, em sua tese, obteve acurácia de 90%, 97% e 99%, na detecção da patologia, utilizando os coeficientes cepstrais, mel-cepstrais e LPC, respectivamente. No entanto, vale ressaltar, que na pesquisa citada, não foi utilizada a validação cruzada, ficando os resultados restritos ao contexto dos sinais analisados. Mudando o *codebook*, ou os sinais de teste, os resultados poderiam variar.

4.3 Resultados obtidos – Análise LPC, Cepstral e Mel-Cepstral

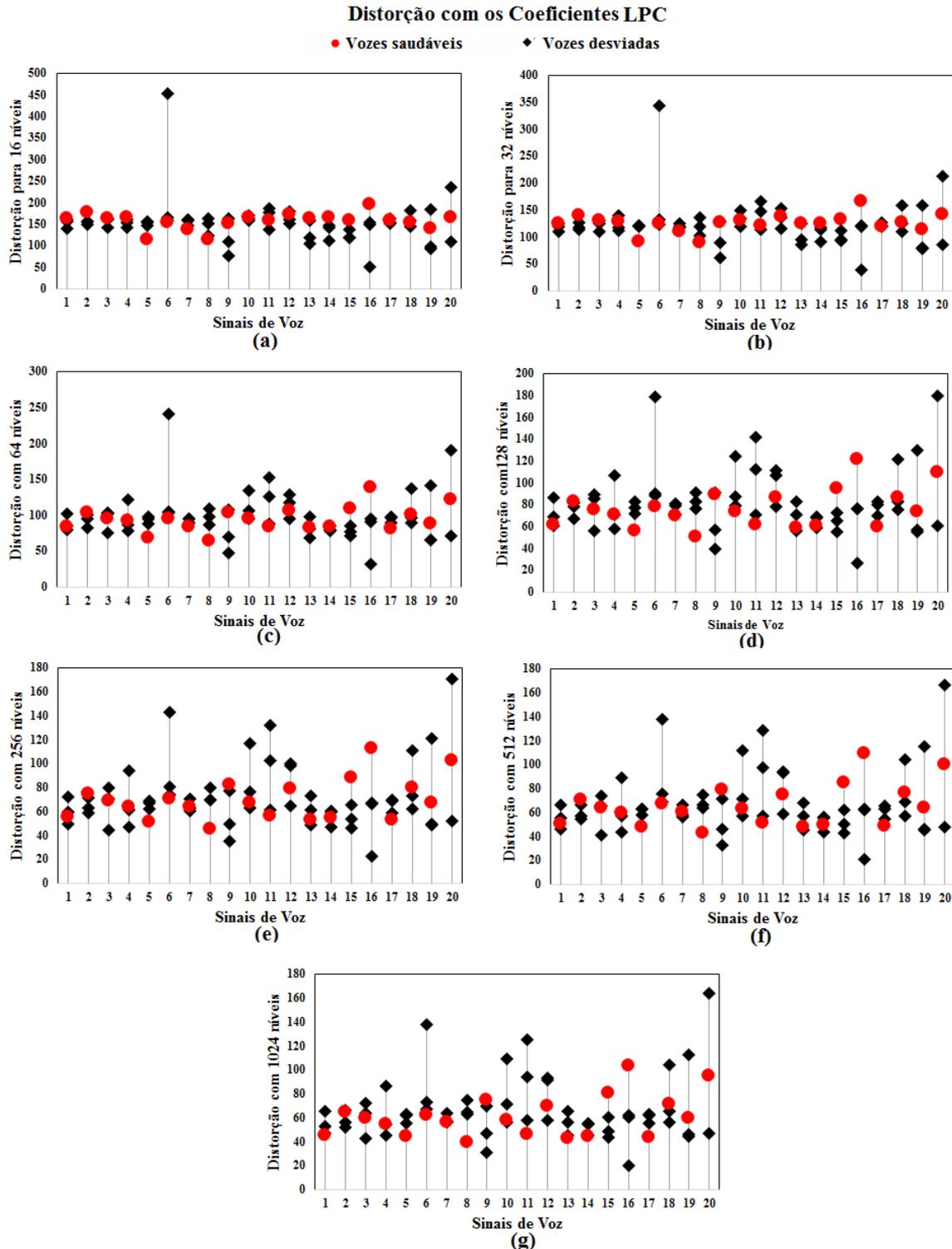
A seguir, são apresentados os resultados obtidos através Análises LPC (coeficientes LPC), cepstral (coeficientes cepstrais) e mel-cepstral (coeficientes mel-cepstrais).

4.3.1 Análise LPC

Na Figura 24 é apresentada a distribuição dos sinais de vozes saudáveis e desviadas (rugosas, soprosas e tensas) de acordo com a distorção euclidiana encontrada para 16, 32, 64, 128, 256, 512 e 1024 níveis, empregada na quantização vetorial com os coeficientes LPC.

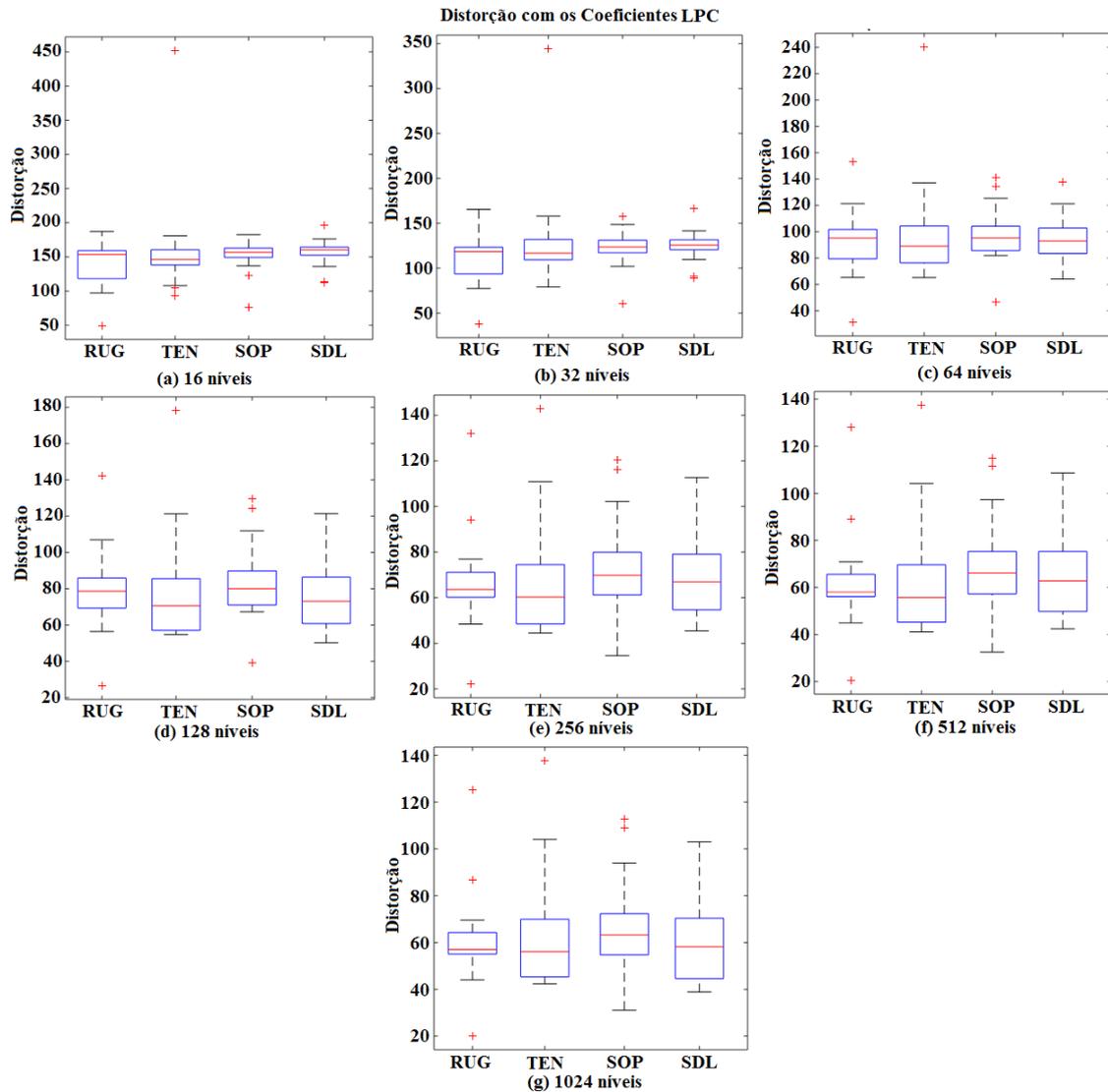
Observa-se pela Figura 24 que não há uma clara separação entre as vozes saudáveis e desviadas, mostrando que os coeficientes LPC não conseguem distinguir entre as classes de forma clara. Observa-se, ainda, que quanto maior o número de níveis, maior a variabilidade dos valores das distorções encontradas com os coeficientes LPC para os sinais saudáveis.

Figura 24 - Comportamento da distorção para vozes saudáveis e vozes desviadas para os diversos números de níveis com os coeficientes LPC: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.



Na Figura 25 estão ilustrados os *boxplots* com os resultados para as distorções com os coeficientes LPC para os diversos níveis de quantização as classes rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).

Figura 25 - Comportamento da distorção para vozes rugosas (RUG), tensas (TEN), soprosa (SOP) e saudáveis (SDL) para os diversos número de níveis com os coeficientes LPC: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.



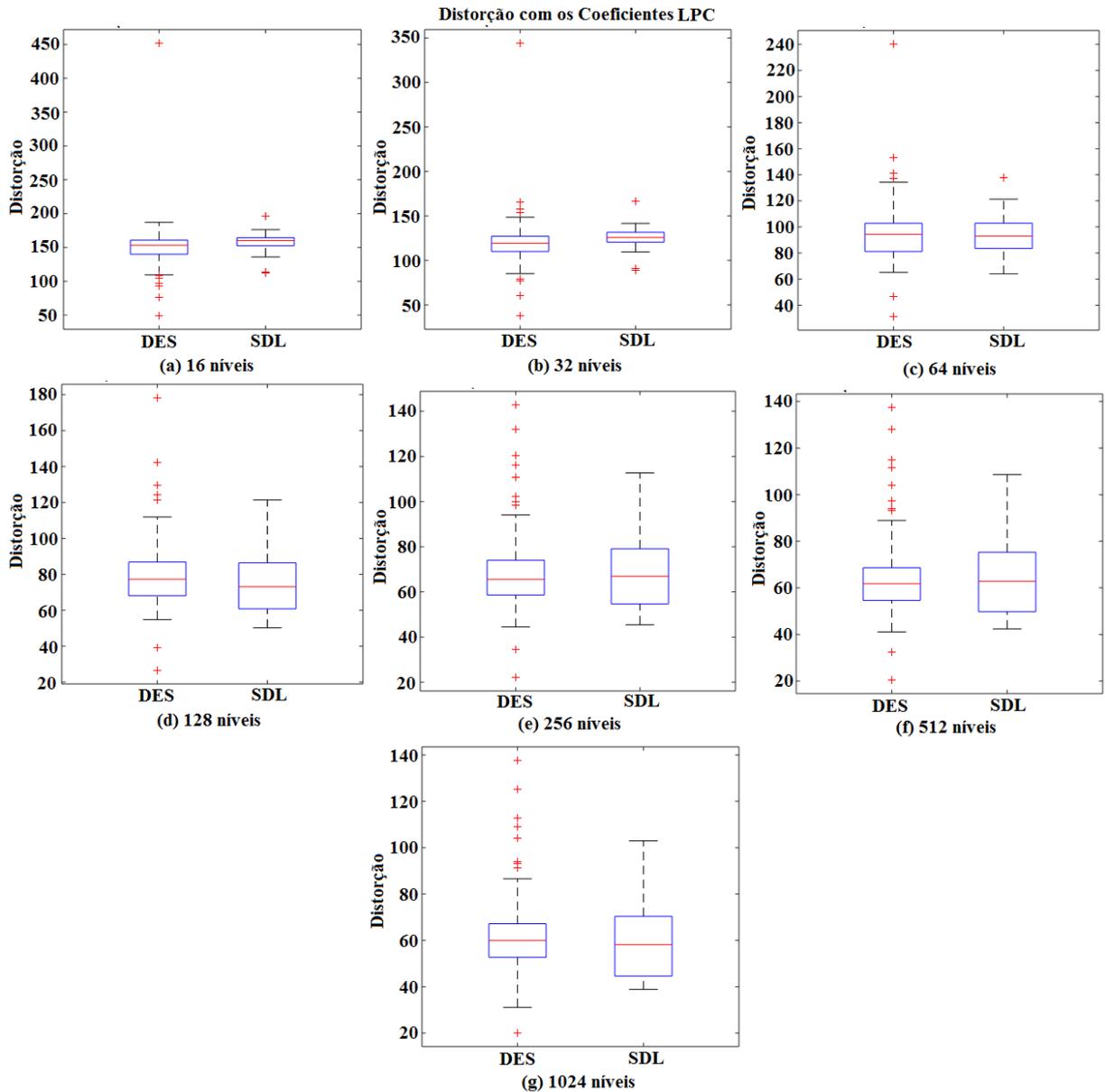
Observa-se na Figura 25 que as dispersões para as quatro classes possuem intervalos semelhantes e as medidas não se diferenciam. Quanto à avaliação das distorções pelos níveis, observa-se que quanto maior o número de níveis de quantização, maiores são as dispersões para as distorções, não havendo muita variação para as distorções com os sinais rugosos.

Na Figura 26 estão ilustrados os *boxplots* com os resultados para as distorções com os coeficientes LPC para os diversos níveis de quantização para os sinais desviados (DES) e saudáveis (SDL).

Observa-se, na Figura 26, que as dispersões para os sinais saudáveis e desviados para 16 e 32 níveis de quantização possuem intervalos semelhantes e as medidas não se diferenciam. A partir 128 níveis os sinais de vozes saudáveis apresentam dispersão maior que

os sinais de vozes desviadas para os coeficientes LPC. Não é possível ver separação entre as medianas em todos os níveis.

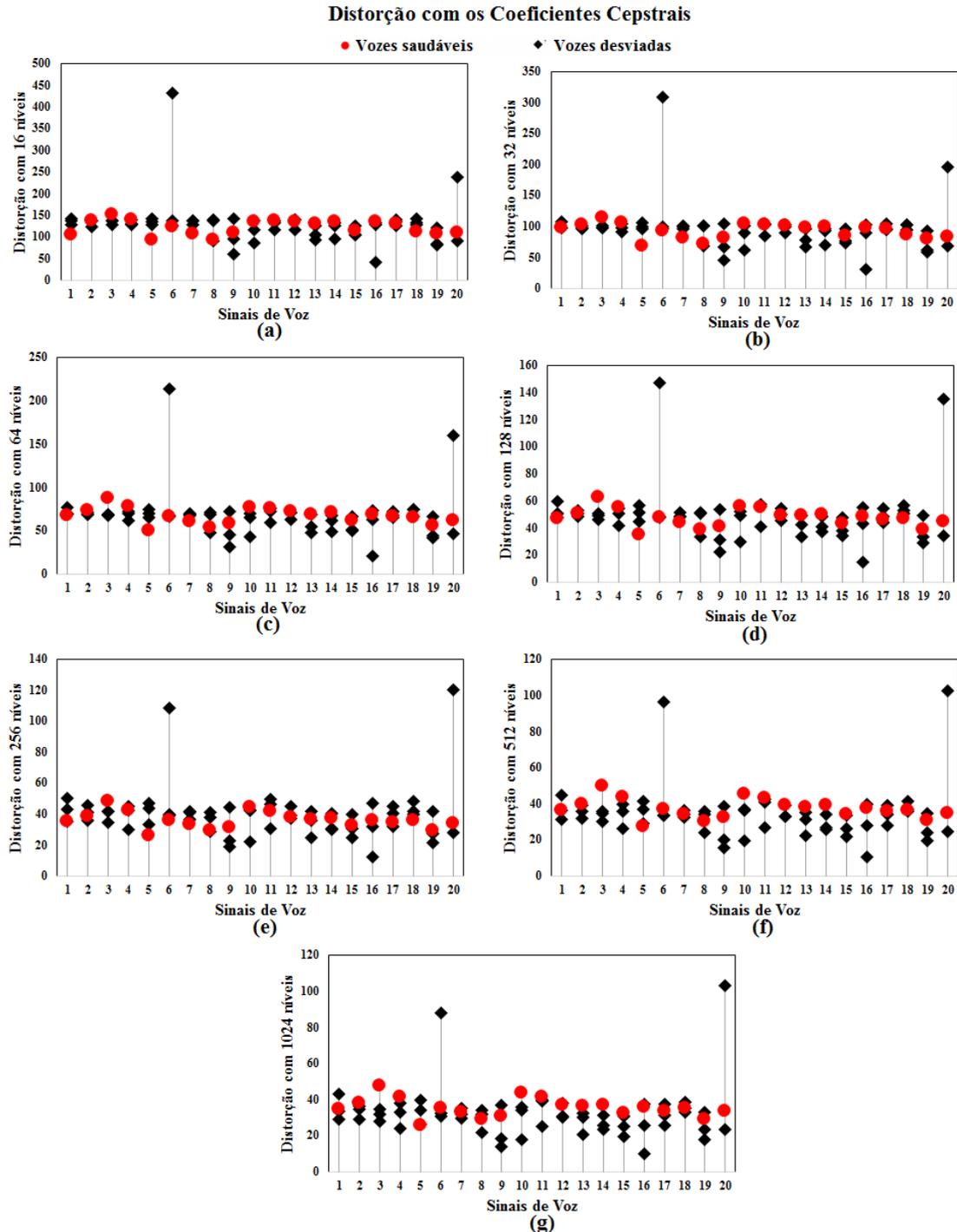
Figura 26 - Comportamento da distorção para vozes desviadas (DES) e saudáveis (SDL) para os diversos números de níveis com os coeficientes LPC: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.



4.3.2 Análise cepstral – coeficientes cepstrais

Na Figura 27 é apresentada a distribuição dos sinais de vozes saudáveis e desviadas (rugosas, soprosas e tensas) de acordo com a distorção euclidiana encontrada para cada número de níveis, empregada na quantização vetorial com os coeficientes cepstrais.

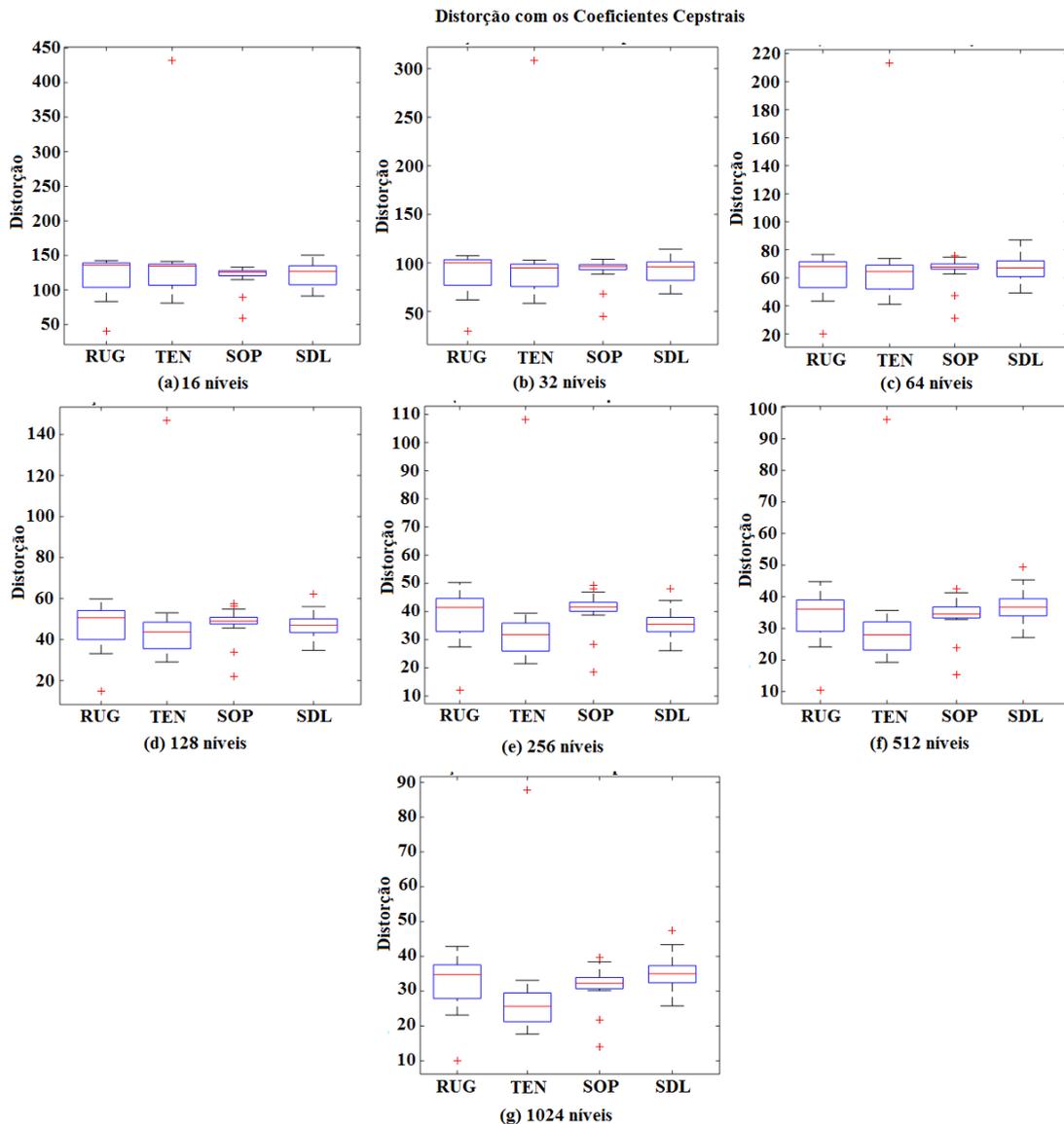
Figura 27 - Comportamento da distorção para vozes saudáveis e vozes desviadas para os diversos números de níveis com os coeficientes cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.



Observa-se pela Figura 27 que os valores encontrados para as distorções dos sinais saudáveis estão dentro dos valores para as distorções dos sinais desviados, não havendo uma clara separação entre as vozes saudáveis e desviadas, mostrando que a distorção encontrada para os coeficientes cepstrais com os níveis utilizados não ajudam na classificação dos sinais.

Na Figura 28 estão ilustrados os *boxplots* com os resultados para as distorções com os coeficientes cepstrais para os diversos níveis de quantização, para as classes: rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).

Figura 28 - Comportamento da distorção para vozes rugosas (RUG), tensas (TEN), soprosa (SOP) e saudáveis (SDL) para os diversos números de níveis com os coeficientes cepstrais: (a) 16; (b) 32; (c) 64; (d) 128 níveis; (e) 256; (f) 512 e (g) 1024.

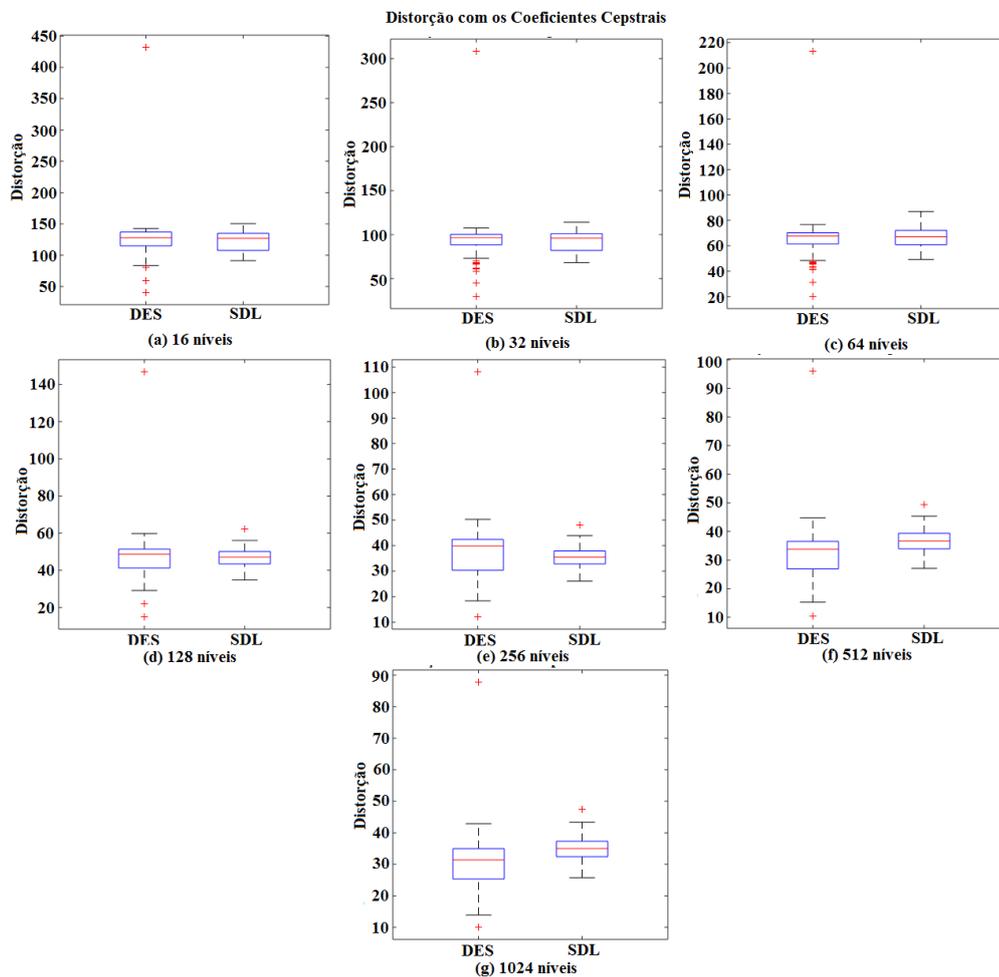


Na Figura 28 as dispersões obtidas com os coeficientes cepstrais para os diferentes níveis de quantização para os sinais rugosos, tensos e saudáveis possuem intervalos semelhantes em 16, 32, 64 e 128 níveis, onde apenas para os sinais soprosos a dispersão é pequena e seus valores estão dentro das demais classes. Para os sinais saudáveis observa-se que quanto maior o número de níveis, menor é a distorção encontrada e a classe que possui

maior intervalo de dispersão são os sinais rugosos. Não é possível realizar distinção entre as classes visto que os valores encontrados possuem intervalos semelhantes, dificultando na classificação. Quanto à avaliação das distorções pelos níveis, observa-se que quanto maior o número de níveis de quantização maiores são as dispersões para as distorções em algumas classes.

Na Figura 29 estão ilustrados os *boxplots* com os resultados para as distorções com os coeficientes cepstrais, para os diversos números níveis de quantização, para os sinais desviados (DES) e saudáveis (SDL), empregados nesta pesquisa.

Figura 29 Comportamento da distorção para vozes desviadas (DES) e saudáveis (SDL) para os diversos números de níveis com os coeficientes cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.

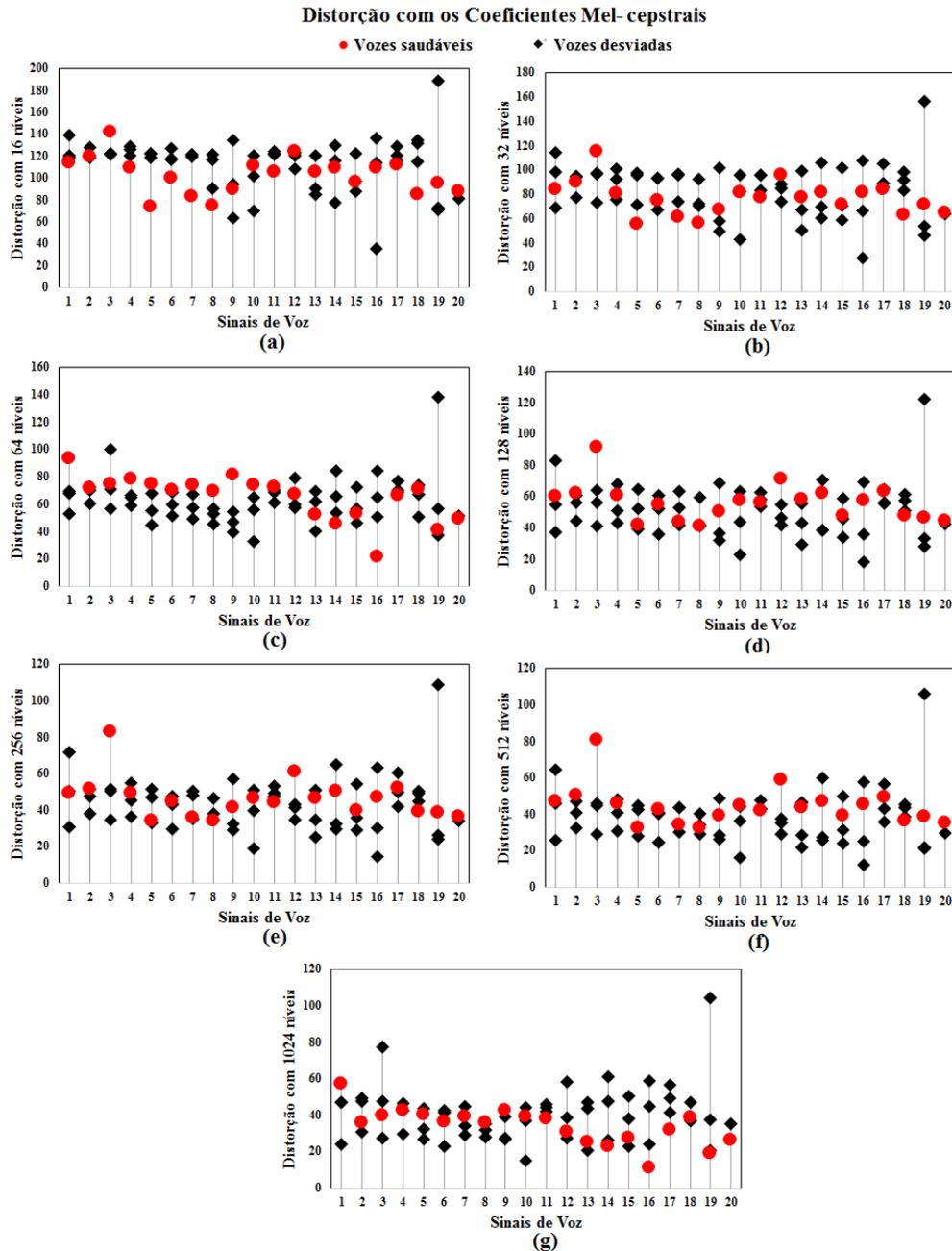


Na Figura 29 as dispersões obtidas com os coeficientes cepstrais para os diferentes níveis de quantização para os sinais desviados e saudáveis possuem intervalos semelhantes para 16, 32, 64 e 128 níveis. Para os sinais saudáveis observa-se que quanto maior o número de níveis, menor é a distorção encontrada. Não é possível realizar distinção entre as classes visto que os valores encontrados possuem intervalos semelhantes, dificultando na classificação.

4.3.3 Análise mel-cepstral – coeficientes mel-cepstrais

Na Figura 30 encontra-se a distribuição dos sinais de vozes saudáveis e desviadas (rugosas, soprosas e tensas) de acordo com a distorção euclidiana encontrada para cada número de níveis, empregada na quantização vetorial com os coeficientes mel-cepstrais.

Figura 30 - Comportamento da distorção para vozes saudáveis e vozes desviadas para os diversos números de níveis com os coeficientes mel-cepstrais. (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.

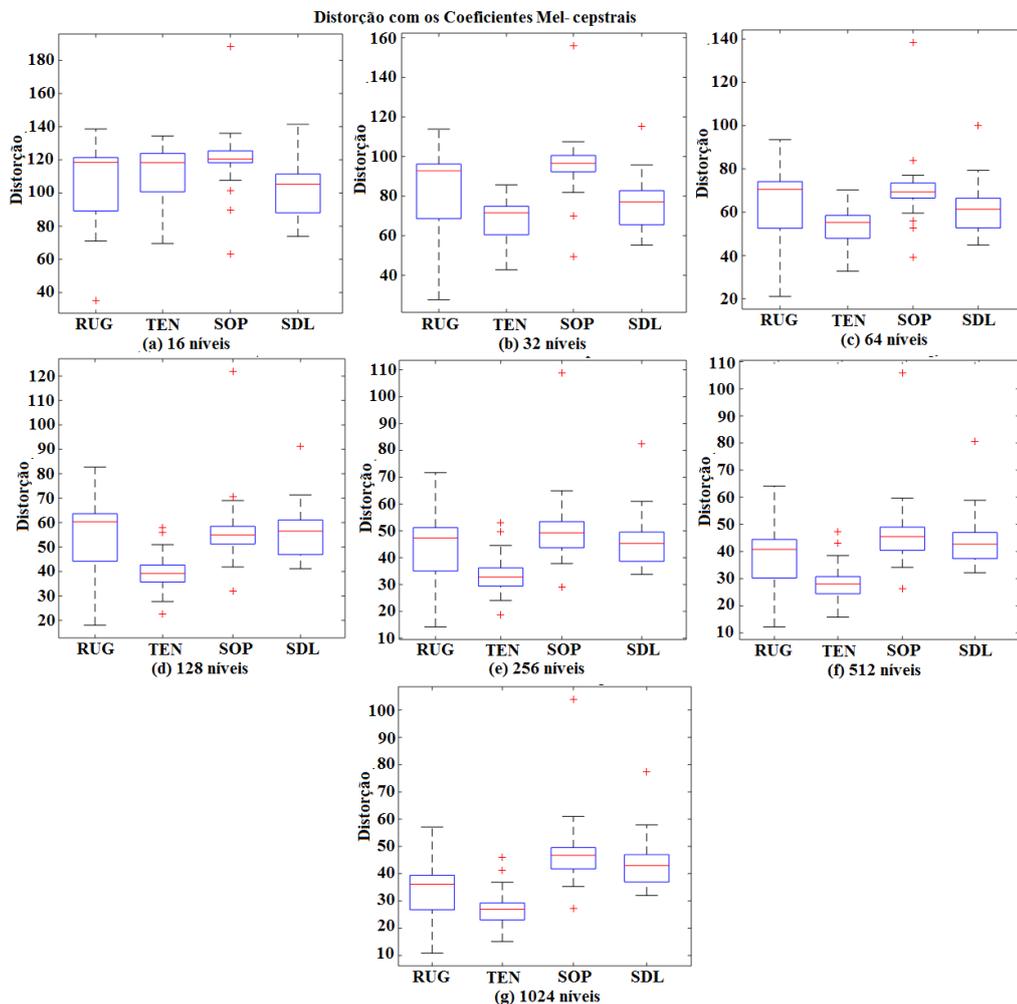


Observa-se pela Figura 30 que os valores encontrados para as distorções dos sinais saudáveis estão dentro dos valores para as distorções dos sinais desviados, não havendo uma clara separação entre as vozes saudáveis e desviadas. Em alguns níveis é possível observar

que as distorções para os sinais saudáveis são menores que para os sinais desviados (por exemplo, 16 e 1024 níveis).

Na Figura 31 estão ilustrados os *boxplots* com os resultados para as distorções com os coeficientes mel-cepstrais para os diversos níveis de quantização, ilustrando os valores da média e do desvio padrão encontrado para cada classe: rugosa (RUG), soprosa (SOP), tensa (TEN) e saudável (SDL).

Figura 31 – Comportamento da distorção para vozes rugosas (RUG), tensas (TEN), soprosa (SOP) e saudáveis (SDL) para diversos números de níveis com os coeficientes mel-cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.

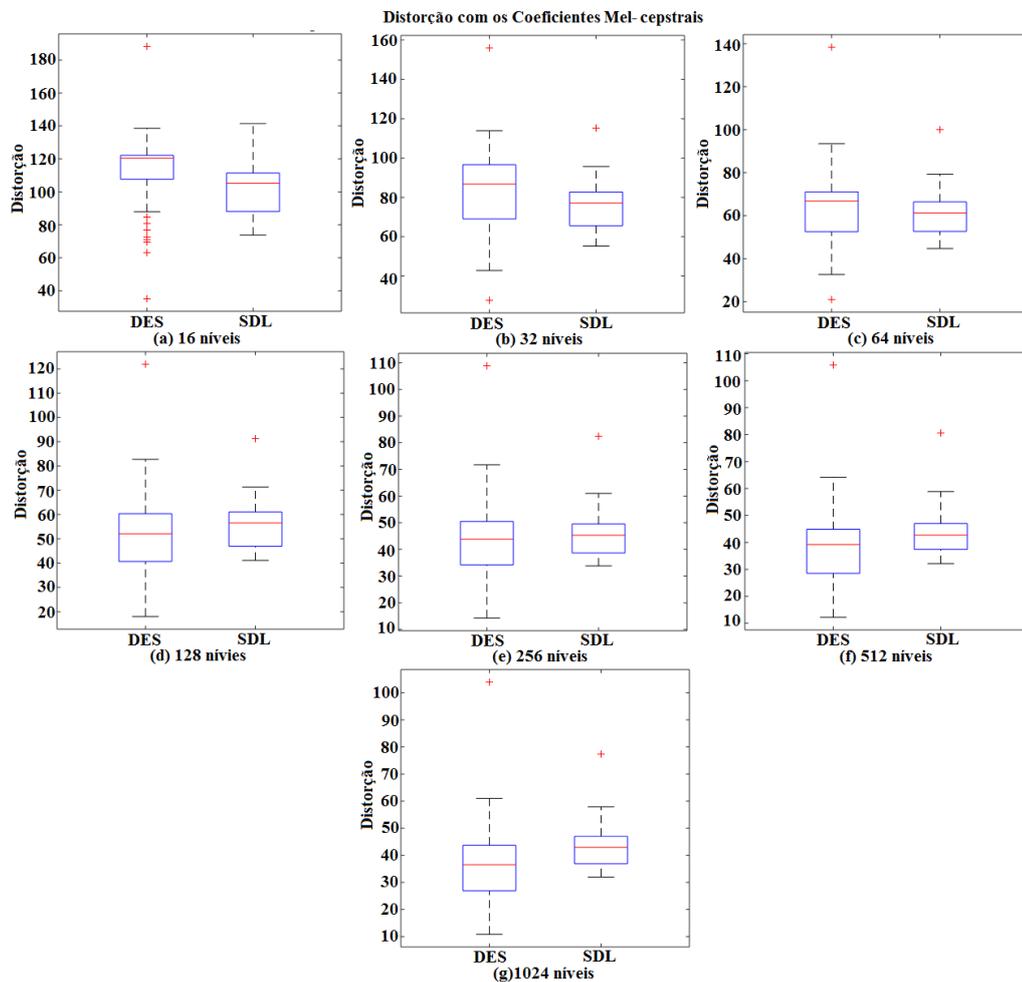


Na Figura 31 é possível observar os *boxplots* obtidos com as distorções euclidianas encontrada com os coeficientes mel-cepstrais para os diferentes números de níveis de quantização. Observa-se que os sinais rugosos comparados aos demais sinais apresentaram maior dispersão em todos os níveis de quantização. Houve grande variação para as distorções quanto ao número de níveis com os sinais tensos, soprosos e saudáveis com 32 e 64 níveis de quantização é possível distinguir os sinais soprosos dos sinais tensos. Os sinais soprosos, tensos e saudáveis possuem menor dispersão, comparado com os sinais rugosos. Quanto aos

níveis de quantização, observa-se que quanto maior o número de níveis menor é a dispersão para as classes rugosa, tensa e saudável ocorrendo o inverso para a classe soprosa.

Na Figura 32 estão ilustrados os *boxplots* com os resultados para as distorções com os coeficientes mel-cepstrais para os diversos números de níveis de quantização, ilustrando os valores da média e do desvio padrão encontrado para os sinais de vozes desviadas (DES) e saudáveis (SDL).

Figura 32 – Comportamento da distorção para vozes desviadas (DES) e saudáveis (SDL) para diversos números de níveis, com os coeficientes mel-cepstrais: (a) 16; (b) 32; (c) 64; (d) 128; (e) 256; (f) 512 e (g) 1024.



Na Figura 32 é possível observar os *boxplots* obtidos com as distorções euclidianas encontrada com os coeficientes mel-cepstrais para os diferentes níveis de quantização. Observa-se que os sinais desviados comparados aos sinais saudáveis apresentaram menor dispersão para 16 níveis, e para os demais níveis sua dispersão é maior. Para 16, 32 e 64 níveis, é possível distinguir os sinais desviados e saudáveis pelas medianas. Quanto aos níveis de quantização, observa-se que quanto maior o número de níveis maior é a dispersão para os sinais desviados.

4.4 Resultados da Classificação dos Desvios Vocais com as Análises LPC, Cepstral e Mel-Cepstral

Nesta secção são apresentados os resultados obtidos com as distorções apresentando as tabelas de classificação obtidas com as curvas ROC (*Receive Operator Characteristic Curve*) e as curvas ROC obtidos para os coeficientes LPC, cepstrais e mel-cepstrais, bem como também serão demonstrados os resultados obtidos através do classificador utilizando as médias dos coeficientes LPC, cepstrais e mel-cepstrais na classificação dos sinais saudáveis e desviados (rugosos, tensos e soprosos). A classificação foi realizada seguindo os critérios de Hosmer-Lemeshow (2000), mas para algumas classificações os melhores resultados obtiveram taxas de acurácia inferiores a 70%, sendo aqui demonstrados para uma posterior comparação entre os métodos aplicados na pesquisa.

4.4.1 Resultados obtidos com as curvas ROC

Para realizar a avaliação da classificação e demonstrar os resultados obtidos, foram utilizados gráficos das curvas ROC. A Curva ROC (*Receive Operator Characteristic Curve*) é um método muito utilizado para avaliação de desempenho em testes de diagnósticos médicos.

Para o cálculo da curva ROC emprega-se uma regra de decisão (pertence ou não pertence a classe) baseada na busca por um ponto de corte, de forma que valores menores ou iguais a este ponto de corte são classificados com pertencentes a classe e, analogamente, valores com resposta ao teste maiores que o ponto de corte são classificados como não pertencentes a classe (ou vice-versa). Dessa forma, para diferentes pontos de corte, dentro dos possíveis valores que o teste produz, é possível estimar características como sensibilidade e especificidade explicadas anteriormente (MARTINEZ; LOUZADA-NETO; PEREIRA, 2003). Com este método de classificação as curvas de diferentes testes diagnósticos podem ser comparadas, quanto melhor o resultado da comparação, mais perto está a curva do canto superior esquerdo dos gráficos (COSTA, 2008).

Na Tabela 18 são apresentados os melhores resultados obtidos com a curva ROC para os coeficientes cepstrais e mel-cepstrais na distinção de sete casos de classificação (vozes Rugosas Vs. Saudáveis, vozes Tensas Vs. Saudáveis, vozes Soprosas Vs. Saudáveis, vozes Desviadas Vs. Saudáveis, vozes Rugosas Vs. Tensas, vozes Rugosas Vs. Soprosas e vozes Tensas Vs. Soprosas) apresentando qual foi o melhor número de níveis de quantização para realizar a classificação baseado nas tabelas apresentadas no Apêndice A. As taxas de classificação utilizando os coeficientes LPC não obtiveram bons resultados (taxas inferiores

(menores que 70%) às encontradas com os coeficientes cepstrais e mel-cepstrais na classificação), então por este motivo, não foram expostas.

Tabela 18– Melhores casos de classificação com as medidas de distorção dos coeficientes - Curva ROC.

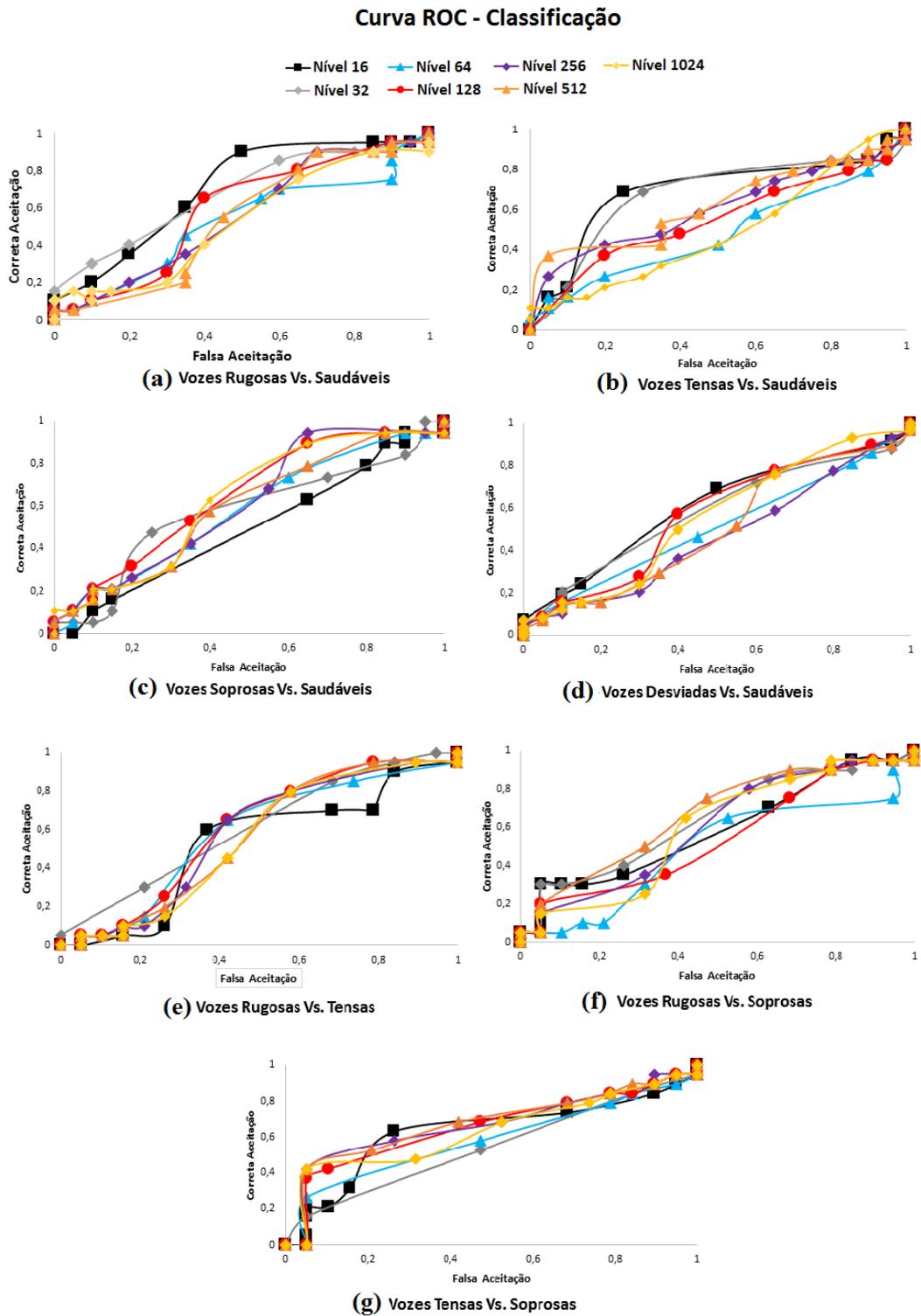
CLASSE	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)	Medida	Nível
Desviada Vs. Saudável	0,15	0,70	0,85	77,84	70,69	85,00	MEL	16
Rugosa Vs. Saudável	0,15	0,65	0,85	75,00	65,00	85,00	MEL	16
	0,25	0,70	0,75	72,50	70,00	75,00	MEL	32
	0,2	0,65	0,80	72,50	65,00	80,00	MEL	256
Tensa Vs. Saudável	0,2	0,84	0,80	82,10	84,21	80,00	MEL	256
	0,15	0,84	0,85	84,60	84,21	85,00	MEL	512
	0,20	0,89	0,80	84,73	89,47	80,00	CEPS	1024
Soprosa Vs. Saudável	0,20	0,78	0,80	79,47	78,95	80,00	MEL	16
	0,15	0,78	0,85	81,97	78,95	85,00	MEL	32
	0,20	0,84	0,80	82,10	84,21	80,00	CEPS	256
Rugosa Vs. Tensa	0,15	0,70	0,84	77,10	70,00	84,21	MEL	32
	0,26	0,85	0,73	79,34	85,00	73,68	MEL	128
	0,10	0,65	0,89	77,23	65,00	89,47	CEPS	512
	0,21	0,75	0,78	76,97	75,00	78,94	CEPS	1024
Rugosa Vs. Soprosa	0,15	0,65	0,84	74,60	65,00	84,21	CEPS	32
	0,15	0,60	0,84	72,10	60,00	84,21	MEL	1024
Tensa Vs. Soprosa	0,10	0,73	0,89	81,57	73,68	89,47	CEPS	256 e 32
	0,15	0,84	0,84	84,21	84,21	84,21	MEL	512
	0,10	0,84	0,89	86,84	84,21	89,47	MEL	1024

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia; SEN: sensibilidade; ESP: especificidade; MEL: coeficientes mel cepstrais; CEPS: coeficientes cepstrais.

Observa-se pela Tabela 18 que os melhores resultados de classificação foram obtidos com os coeficientes cepstrais (*CEPS*) e mel-cepstrais (*MEL*). O maior destaque na classificação, utilizando as medidas de distorções, foram a distinção entre vozes tensas e saudáveis, soprosas e saudáveis, rugosas e soprosas, todas utilizando os coeficientes mel-cepstrais como melhor medida. O número de níveis igual a 16 obteve menor destaque na classificação. Isso pode ter corrido por que, quando é realizada a quantização vetorial, ocorre perda de informação, já que quanto menor o número de níveis de quantização, menor é a quantidade de informação sobre o sinal. Quanto maior o número de níveis, maiores são as taxas de classificação, como pode ser observado na Tabela 18. No entanto, o custo computacional aumenta proporcionalmente ao número de níveis.

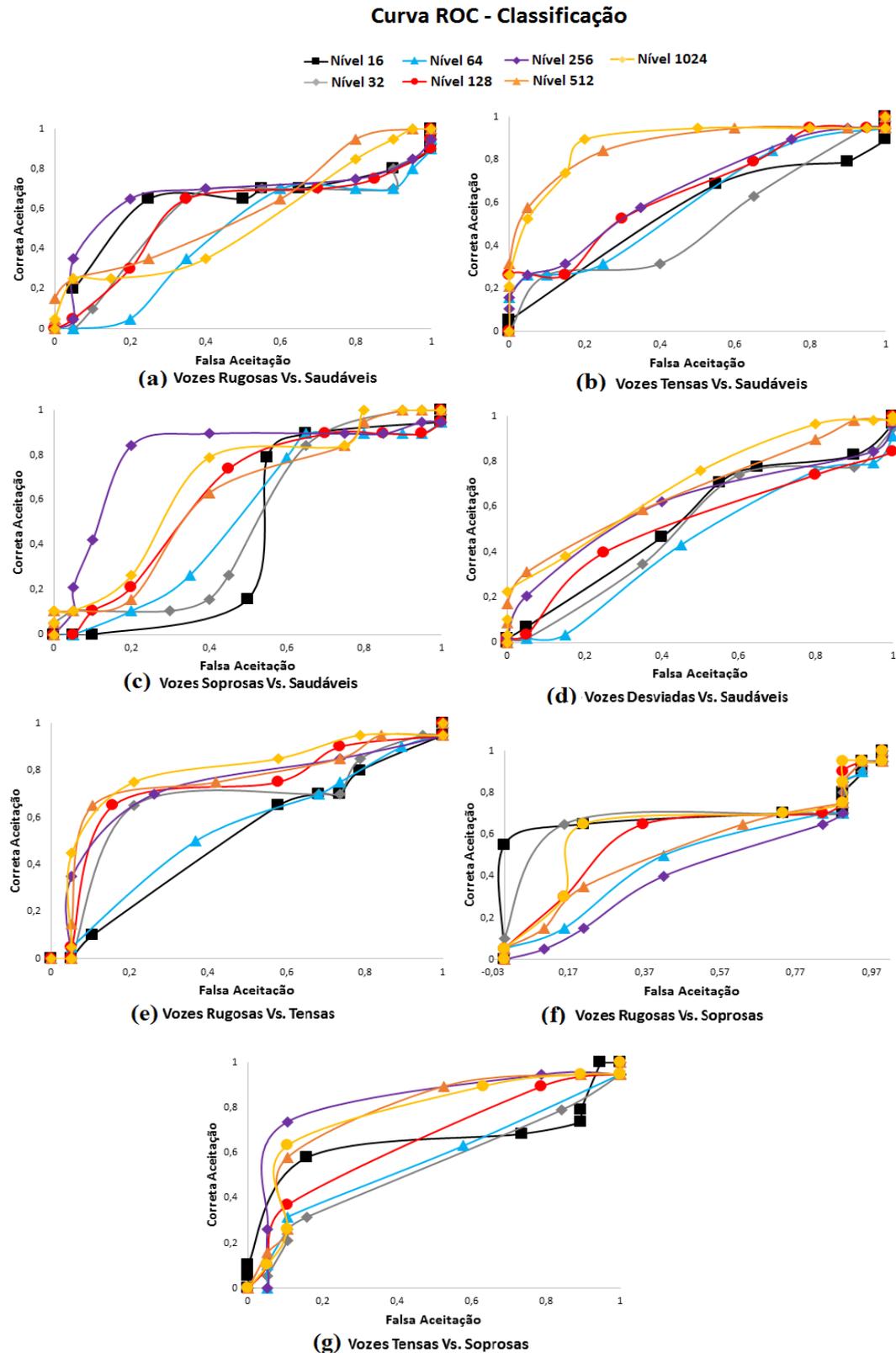
Na Figura 33 são apresentados os gráficos contendo as curvas ROC para distorções euclidianas com os coeficientes LPC, demonstrando como se comportou a curva para cada nível de quantização na classificação dos sinais.

Figura 33 - Comportamento da curva ROC com as distorções dos coeficientes LPC para cada número de níveis de quantização. (a) vozes Rugosas Vs. Saudáveis; (b) vozes Tensas Vs. Saudáveis; (c) vozes Soprosas Vs. Saudáveis; (d) vozes Desviadas Vs. Saudáveis; (e) vozes Rugosas Vs. Tensas; (f) vozes Rugosas Vs. Soprosas e (g) vozes Tensas Vs. Soprosas.



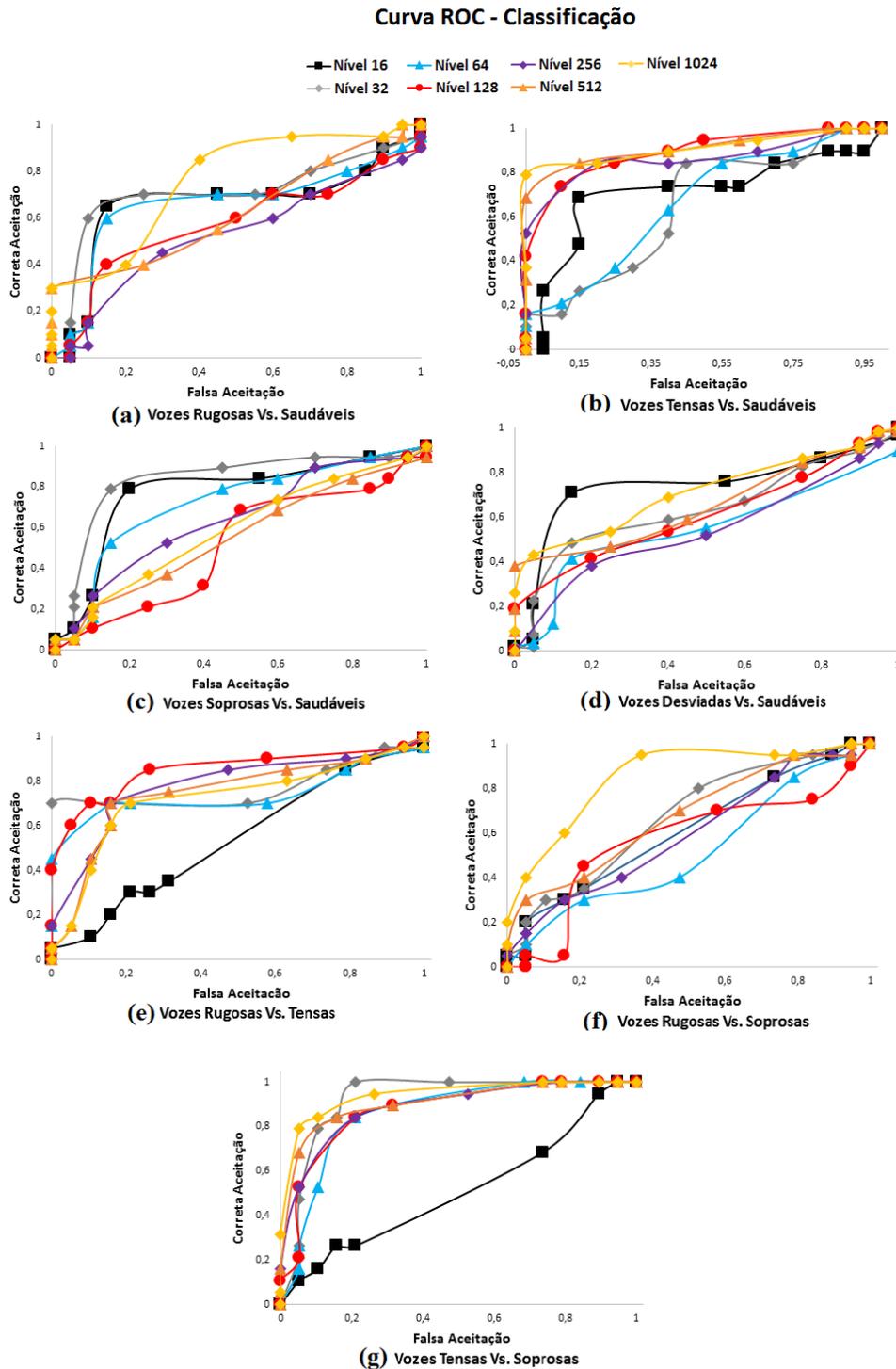
Na Figura 34 são apresentados os gráficos contendo as curvas ROC para as distorções euclidianas com os coeficientes cepstrais, demonstrando como se comportou a curva para cada número de níveis de quantização na classificação dos sinais.

Figura 34 - Comportamento da curva ROC com as distorções dos coeficientes cepstrais para cada número de níveis de quantização. (a) vozes Rugosas Vs. Saudáveis; (b) vozes Tensas Vs. Saudáveis; (c) vozes Soprosas Vs. Saudáveis; (d) vozes Desviadas Vs. Saudáveis; (e) vozes Rugosas Vs. Tensas; (f) vozes Rugosas Vs. Soprosas e (g) vozes Tensas Vs. Soprosas.



Na Figura 35 são apresentados os gráficos contendo as curvas ROC's para as distorções euclidianas com os coeficientes mel-cepstrais, demonstrando como se comportou a curva para cada nível de quantização na classificação dos sinais.

Figura 35 - Comportamento da curva ROC com as distorções dos coeficientes mel-cepstrais para cada nível de quantização. (a) vozes Rugosas Vs. Saudáveis; (b) vozes Tensas Vs. Saudáveis; (c) vozes Soprosa Vs. Saudáveis; (d) vozes Desviadas Vs. Saudáveis; (e) vozes Rugosas Vs. Tensas; (f) vozes Rugosas Vs. Soprosas e (g) vozes Tensas Vs. Soprosas.



Dessa forma, é possível observar o desempenho entre os métodos LPC, cepstrais e mel-cepstrais na classificação e qual o número de níveis que apresenta melhor resultado. Os resultados das curvas demonstrados nas Figuras 33, 34 e 35 confirmam a superioridade dos coeficientes cepstrais e mel-cepstrais em relação aos coeficientes LPC na classificação dos sinais, onde a classificação utilizando os coeficientes mel-cepstrais obteve o maior destaque entre os três métodos. Quanto ao número de níveis, houve uma grande variação na classificação, não sendo possível ainda definir um valor único que sirva para os sete casos de classificação.

4.4.2 Resultados obtidos com classificador MLP

Nesta seção são apresentados os resultados obtidos através do classificador na classificação dos sinais desviados (rugosos, soprosos e tensos) e sinais saudáveis. A classificação foi realizada de duas maneiras. A primeira classificação utilizou os *codebooks* dos coeficientes LPC, cepstrais e mel-cepstrais, com os diversos níveis para realizar a classificação e assim observar quais as características e o número de níveis que melhor diferenciasse as classes, aqui será chamado de Caso 1. E a segunda classificação utilizou como características de entrada para cada sinal de voz um vetor com as médias dos coeficientes LPC (LPC_{med}), cepstrais ($CEPS_{med}$) e mel-cepstrais (MEL_{med}) obtidos para cada sinal e será chamado de Caso 2.

4.4.2.1 Caso 1 de classificação com o classificador MLP

Na Tabela 19 são apresentados os melhores resultados obtidos (baseados nas tabelas do Apêndice C) com o classificador MLP utilizando como características de entrada os *codebooks* com os coeficientes LPC (*codebook LPC*), cepstrais (*codebook CEPS*) e mel-cepstrais (*codebook MEL*) com os diferentes números de níveis de quantização na distinção de quatro casos de classificação (vozes Rugosas Vs. Saudáveis, vozes Tensas Vs. Saudáveis, vozes Soprosas Vs. Saudáveis e vozes Desviadas Vs. Saudáveis) apresentando o melhor número de níveis de quantização encontrado no processo de classificação, seguindo os critérios de Hosmer-Lemeshow (2000). No entanto, para algumas classificações os melhores resultados obtiveram taxas de acurácia inferiores a 70%, sendo aqui demonstrados para uma posterior comparação entre os métodos aplicados na pesquisa.

Tabela 19 - Melhores casos de classificação com os *codebooks* para os coeficientes LPC, cepstrais e mel-cepstrais – Classificador MLP, com *N* níveis.

CLASSE	ACUR (%)	SEN (%)	ESP (%)	Medida	N
Desviada Vs. Saudável	73.14 ± 0.61	67.71 ± 0.68	84.00 ± 0.25	<i>codebook LPC</i>	256
	75.00 ± 0.01	75.00 ± 0.02	75.07 ± 0.31	<i>codebook CEPS</i>	256
	75.00 ± 0.00	75.00 ± 0.00	75.00 ± 0.04	<i>codebook MEL</i>	32
Rugosa Vs. Saudável	57.42 ± 0.91	56.07 ± 0.70	55.23 ± 1.22	<i>codebook LPC</i>	128
Tensa Vs. Saudável	71.29 ± 0.18	81.70 ± 0.67	66.03 ± 0.34	<i>codebook LPC</i>	256
Soprosa Vs. Saudável	73.14 ± 0.61	67.71 ± 0.68	84.00 ± 0.25	<i>codebook CEPS</i>	256

ACUR: acurácia; SEN: sensibilidade; ESP: especificidade; LPC: coeficientes predição linear; MEL: coeficientes mel-cepstrais; CEPS: coeficientes cepstrais.

Pela Tabela 19 pode-se observar que a melhor classificação foi encontrada para a discriminação entre sinais de vozes tensas e saudáveis, soprosas e saudáveis e entre sinais de vozes desviadas e saudáveis atingindo taxas de acurácia de 70% utilizando coeficientes LPC, cepstrais e mel-cepstrais. A classificação entre sinais rugosos e saudáveis não foi efetiva, já que a taxa de acurácia foi inferior a 60% utilizando como característica o *codebook* com os coeficientes LPC. Neste caso de classificação as características *codebook CEPS* e *codebook MEL* não estão presentes, pois as taxas de classificação foram inferiores a 55%. Quanto ao número de níveis de quantização, o que obteve maior destaque foi $N=256$, não sendo efetivo apenas na classificação de sinais de vozes rugosas e saudáveis.

Na Tabela 20 observa-se a matriz de confusão para a classificação Rugosa Vs. Soprosa, em que a característica que melhor classificou foi o *codebook CEPS* com 256 níveis, apresentando uma acurácia de 76,17%.

Tabela 20 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz soprosos com o *codebook*.

CLASSE	RUGOSA	SOPROSA
RUGOSA	22	8
SOPROSA	6	24

Na Tabela 21 observa-se a matriz de confusão para a classificação Rugosa Vs. Tensa, apresentando uma acurácia de 71,19% em que a característica utilizada para a classificação foi *codebook LPC* com 256 níveis.

Tabela 21 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz tensos com o *codebook*.

CLASSE	RUGOSA	TENSA
RUGOSA	25	5
TENSA	10	20

Na Tabela 22 observa-se a matriz de confusão para a classificação Tensa Vs. Soprosa, apresentando uma acurácia de 75%. Apresentando esta taxa de acurácia tanto na classificação com o *codebook LPC* como com o *codebook CEPS*, ambos com 256 níveis de quantização.

Tabela 22 - Matriz de confusão para a discriminação entre sinais de voz soprosos e sinais de voz tensos com o *codebook*.

CLASSE	TENSA	SOPROSA
TENSA	25	5
SOPROSA	9	21

Na Tabela 23 encontra-se a um resumo da avaliação do desempenho na discriminação entre os desvios vocais utilizando os *codebooks* dos coeficientes LPC, cepstrais e mel-cepstrais.

Tabela 23 - Resumo da avaliação de desempenho na discriminações entre os desvios vocais – *codebooks LPC, CEPS e MEL*.

CLASSE	ACURÁCIA	Nº de Níveis	Medidas
RUGOSA Vs. SOPROSA	76,17%	256	<i>Codebook CEPS</i>
RUGOSA Vs. TENSA	71,19%	256	<i>codebook LPC</i>
TENSA Vs. SOPROSA	75%	256	<i>codebook CEPS, codebook LPC</i>

LPC: coeficientes predição linear; MEL: coeficientes mel cepstrais; CEPS: coeficientes cepstrais.

Observa-se pelas matrizes de confusão apresentadas (Tabelas 20, 21 e 22) e pela Tabela 23, que os coeficientes cepstrais e LPC foram mais eficientes na classificação entre os desvios, onde os coeficientes mel-cepstrais não foram tão efetivos. O melhor número de níveis foi 256, estando presente nos três casos de classificação.

4.4.2.2 Caso 2 de classificação com o classificador MLP

Na Tabela 24 são apresentados os melhores resultados (baseados nas tabelas do Apêndice D) obtidos com o classificador MLP utilizando como características de entrada um vetor com as médias dos coeficientes LPC (LPC_{med}), cepstrais ($CEPS_{med}$) e mel-cepstrais (MEL_{med}) na distinção de quatro casos de classificação (vozes Rugosas Vs. Saudáveis, vozes Tensas Vs. Saudáveis, vozes Soprosas Vs. Saudáveis e vozes Desviadas Vs. Saudáveis). A classificação foi realizada seguindo os critérios de Hosmer-Lemeshow (2000).

Tabela 24 - Melhores casos de classificação com a média dos coeficientes LPC (LPC_{med}), cepstrais ($CEPS_{med}$) e mel-cepstrais (MEL) – Classificador MLP.

Classificação	ACUR (%)	SEN (%)	ESP (%)	Medida
Desviada Vs. Saudável	77.73 ± 1.71	80.40 ± 1.19	79.23 ± 1.04	LPC_{med}
	77.54 ± 1.93	84.54 ± 1.90	81.32 ± 1.54	$CEPS_{med}$
	73.93 ± 2.79	75.11 ± 0.35	74.40 ± 1.25	MEL_{med}
Rugosa Vs. Saudável	81.67 ± 3.22	77.95 ± 3.45	82.97 ± 3.09	LPC_{med}
	77.12 ± 3.88	82.74 ± 4.51	73.13 ± 7.88	$CEPS_{med}$
Tensa Vs. Saudável	71.19 ± 3.93	71.19 ± 3.93	73.78 ± 4.00	LPC_{med}
Soprosa Vs. Saudável	83.05 ± 2.30	82.29 ± 2.51	82.05 ± 2.44	$CEPS_{med}$
	75.83 ± 4.38	75.00 ± 5.49	81.93 ± 3.95	LPC_{med}
	75.86 ± 3.70	77.75 ± 4.03	81.44 ± 3.68	MEL_{med}

ACUR: acurácia; SEN: sensibilidade; ESP: especificidade; LPC_{med} : médias dos coeficientes LPC; $CEPS_{med}$: média dos coeficientes cepstrais e MEL_{med} : médias dos coeficientes mel-cepstrais.

Pela Tabela 24 pode-se observar que a melhor classificação foi encontrada para a discriminação entre sinais de vozes rugosas e saudáveis e entre sinais de vozes soprosas e saudáveis atingindo taxas de acurácia superiores a 80% utilizando a média dos coeficientes LPC e cepstrais respectivamente. Todas as classificações obtiveram ótimas taxas de acurácia, se comparada às taxas obtidas com as curvas ROC e os *codebooks*, tendo destaque a classificação entre sinais de vozes rugosas e saudáveis que apresentou taxa de acurácia inferior a 60% na classificação com o *codebook*. Neste caso de classificação as características que estiveram mais presentes foram LPC_{med} e $CEPS_{med}$.

Na Tabela 25 observa-se a matriz de confusão para a classificação Rugosa Vs. Soprosa, apresentando uma acurácia de 71,19%. Apresentando esta taxa de acurácia tanto na classificação com o LPC_{med} , como com o $CEPS_{med}$.

Tabela 25 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz soprosos com a média dos coeficientes.

CLASSE	RUGOSA	SOPROSA
RUGOSA	21	9
SOPROSA	7	23

Na Tabela 26 observa-se a matriz de confusão para a classificação Rugosa Vs. Tensa, apresentando uma acurácia de 74,17% em que a característica utilizada para a classificação foi $CEPS_{med}$.

Tabela 26 - Matriz de confusão para a discriminação entre sinais de voz rugosos e sinais de voz tensos com a média dos coeficientes.

CLASSE	RUGOSA	TENSA
RUGOSA	22	8
TENSA	6	24

Na Tabela 27 observa-se a matriz de confusão para a classificação Tensa Vs. Soprosa, apresentando uma acurácia de aproximadamente 78%. Apresentando esta taxa de acurácia na classificação com $CEPS_{med}$.

Tabela 27 - Matriz de confusão para a discriminação entre sinais de voz soprosos e sinais de voz tensos com a média dos coeficientes.

CLASSE	TENSA	SOPROSA
TENSA	24	6
SOPROSA	7	23

Na Tabela 28 encontra-se a um resumo da avaliação do desempenho na discriminação entre os desvios vocais utilizando a média dos coeficientes LPC, cepstrais e mel-cepstrais.

Tabela 28 - Resumo da avaliação de desempenho na discriminações entre os desvios vocais – codebooks das médias dos coeficientes LPC, CEPS e MEL.

CLASSE	ACURÁCIA	Medidas
RUGOSA Vs. SOPROSA	71,19%	LPC_{med} e $CEPS_{med}$
RUGOSA Vs. TENSA	74,17%	$CEPS_{med}$
TENSA Vs. SOPROSA	78,00%	$CEPS_{med}$

LPC_{med} : média dos coeficientes LPC; $CEPS_{med}$: média dos coeficientes cepstrais; MEL_{med} : média dos coeficientes mel-cepstrais.

A partir das matrizes de confusão demonstradas nas Tabelas 18, 19 e 20 e da Tabela 28 observa-se que a classificação realizada entre os desvios com a média dos coeficientes LPC e cepstrais apresentou taxas de acurácia superiores a 70%, sendo semelhante às taxas obtidas para a classificação realizada com os *codebooks* (caso 1 de classificação), apesar de não serem as mesmas características. Como no caso 1, os coeficientes mel-cepstrais não foram aceitáveis, de acordo com a classificação Hosmer & Lemeshow (2000) na discriminação realizada pelo classificador.

4.4.3 Discussão dos resultados com as análises LPC, cepstral e mel-cepstral

A classificação realizada com os desvios vocais utilizando os coeficientes extraídos das análises LPC, cepstral e mel-cepstral foram realizadas de três maneiras. A primeira maneira utilizou as distorções euclidianas dos coeficientes quantizados em diferentes níveis para classificar os desvios a partir da curva ROC observando qual a melhor característica e qual o melhor nível. A segunda forma utilizou os *cobedooks* dos coeficientes como características de entrada em um classificador MLP para classificar os desvios avaliando qual melhor medida e qual o melhor nível. E a terceira forma de classificação utilizou a média

segmental dos coeficientes como características de entrada em um classificador MLP, para classificar os desvios vocais, avaliando qual seria a melhor característica.

Na primeira forma de classificação (através da curva ROC) as taxas de acurácia foram superiores a 72%, onde as características que obtiveram maior destaque foram as distorções euclidianas com os coeficientes cepstrais e mel-cepstrais, onde não houve ocorrência dos coeficientes LPC, com classificações acima de 70%. Os números de níveis que obtiveram maior destaque foram 32, 256 e 1024. Observa-se que, quanto maior o número de níveis de quantização, há mais informações a respeito dos sinais, apresentando, no entanto, maior custo computacional. A menor taxa de classificação se deu entre os sinais de vozes rugosas e soprosos.

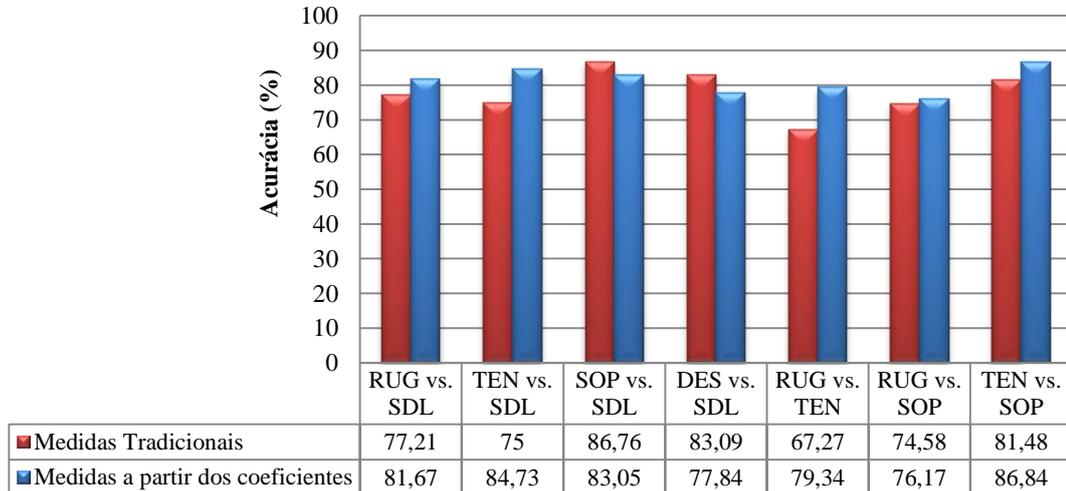
Na segunda forma de classificação (*codebooks* dos coeficientes LPC, cepstrais e mel-cepstrais como características de entrada do classificador MLP) as taxas de acurácia foram superiores a 70% não sendo efetiva apenas na classificação entre vozes rugosas e saudáveis com taxa de acurácia inferior a 60%. As características que obtiveram maior destaque foram os *codebooks* com os coeficientes LPC e cepstrais dando destaque para o nível 256. Nesta forma de classificação os *codebooks* com os coeficientes mel-cepstrais não foram tão eficientes, comparada a classificação utilizando as distorções ou a média dos coeficientes, estando apenas presente a distinção entre vozes desviadas e vozes saudáveis. A classificação realizada com os *codebooks* não atingiu taxas superiores a 75%. Acredita-se que, a perda de informação causada pela quantização pode causar alguns prejuízos na classificação. Além disso, as medidas não conseguiram quantificar precisamente as desordens existentes nos sinais.

Na terceira forma de classificação (média dos coeficientes LPC, cepstrais e mel-cepstrais como características de entrada do classificador MLP) as taxas de acurácia foram semelhantes às taxas obtidas no segundo caso de classificação, exceto para a distinção entre vozes rugosas e saudáveis e entre vozes desviadas e saudáveis (apresentaram acurácia superior a 80% na classificação). A característica que obteve maior destaque foi a média dos coeficientes cepstrais, que apareceu em quase todos os casos (das sete classificações, os coeficientes cepstrais estavam em seis classificações). Observa-se que os valores médios dos coeficientes LPC, cepstrais e mel-cepstrais podem representar os sinais apresentando taxas de classificação entre 70 e 86%.

Na Figura 36 é possível observar o gráfico comparativo dos melhores resultados, obtidos com a classificação dos sinais desviados, utilizando as medidas tradicionais de forma combinada e utilizando as medidas obtidas a partir dos coeficientes LPC, cepstrais e mel-

cepstrais. O detalhamento das medidas que apresentaram melhor caso de classificação está apresentado na Tabela 24.

Figura 36 - Comparação entre os resultados obtidos com as medidas tradicionais e com as medidas dos coeficientes LPC, cepstrais e mel-cepstrais.



Observa-se que a classificação utilizando os coeficientes LPC, cepstrais e mel-cepstrais foi eficiente, onde a maneira como é extraído os coeficientes influenciou nos resultados. Apresentou taxas de acurácia semelhantes as obtidas com as características tradicionais, onde as características que apresentaram melhor resultado estão combinadas, possuindo um custo computacional mais elevado. Observa-se, também, que houve ganho na classificação da maioria dos casos atingindo acurácia 86%, exceto na classificação entre vozes desviadas e saudáveis, onde a acurácia com as medidas tradicionais foi superior à obtida com as medidas dos coeficientes.

A partir desta análise pode-se dizer que as análises LPC, cepstrais e mel-cepstrais são eficientes na classificação dos desvios vocais, onde é necessário observar como são extraídas as medidas a partir destes métodos, pois a extração das características influencia no resultado de classificação.

Outro fator que pode ter influenciado na acurácia, diferenciando os resultados entre o de Lopes *et al.* (2016) e os resultados desta pesquisa, é o fato de que os sinais utilizados para avaliação não terem sido os mesmos, mesmo sendo oriundos da mesma base de dados. Além disso, o grau dos desvios dos sinais selecionados podem mudar a representatividade nos *codebooks* de treino, alterando os resultados na fase de teste.

5. CONSIDERAÇÕES FINAIS

Esta pesquisa investiga a aplicabilidade das medidas tradicionais (frequência fundamental, *jitter*, *shimmer*, GNE, frequências formânticas) e das medidas extraídas a partir das análises LPC, cepstral e mel-cepstral na análise acústica de desvios vocais (rugosidade, soprosidade e tensão). Foram feitos dois estudos, onde o primeiro estudo (Caso 1) empregou métodos de extração das características tradicionais que dependam (como *jitter*, *shimmer*, frequência fundamental) e não dependam da frequência fundamental (frequências formânticas e GNE) para investigar o potencial destes parâmetros na classificação dos desvios vocais, visto que em alguns casos não é possível realizar a obtenção da frequência fundamental. O segundo estudo (Caso 2) avaliou parâmetros extraídos a partir da Análises LPC (coeficientes LPC), cepstral (coeficientes cepstrais) e mel-cepstral (coeficientes mel-cepstrais) com o intuito de investigar o potencial discriminativo destes coeficientes na classificação de desvios vocais, sendo que são parâmetros que não dependem da frequência fundamental e na literatura não foram encontrados trabalhos que utilizem estas análises como parâmetros na avaliação de desvios.

No primeiro estudo conclui-se que as melhores medidas para classificação dos desvios foram o *shimmer* e o GNE. O GNE pode ser utilizada para realizar a triagem dos pacientes, pois foi eficiente na discriminação entre os sinais desviados e saudáveis detectando a presença do desvio vocal (rugosidade, soprosidade e tensão). O *shimmer* seria uma medida de classificação identificando qual o desvio vocal presente, já que ele foi eficiente na discriminação de vozes soprosas e saudáveis. A combinação das medidas frequências formânticas e a frequência fundamental aumentou o desempenho da classificação.

No segundo estudo de caso (análise LPC, cepstral e mel-cepstral), comparando as três formas de classificação (curva ROC, *codebooks* com QV e coeficientes médios com MLP), é possível observar que a maneira como é realizada a classificação utilizando as técnicas propostas pode influenciar nos resultados.

Comparando os métodos empregados (medidas tradicionais com as análises LPC, cepstral e mel-cepstral) observa-se que os resultados obtidos com cada método são semelhantes, sendo necessário destacar que, para alguns casos de classificação, um método se comporta melhor que o outro não podendo definir uma métrica para a classificação de todos os sinais. É preciso também observar a influência do desvio vocal na classificação com as medidas utilizadas, visto que alguns sinais de vozes possuem a presença de mais de um

desvio, sendo necessário também avaliar se o grau do desvio presente e as lesões que podem ou não estar presentes na laringe influenciam na classificação dos sinais.

Observa-se, ainda, que pelas Tabelas A1 a A3, Apêndice A, dos 30 sinais selecionados como rugosos, 5 deles (apenas 16,67%) não apresentam lesão laríngea. Dos tensos, 26,67% não tem lesão na laringe associada e dos soprosos, também 16,67% não sofrem de lesão laríngea. Dos sinais selecionados, portanto, a maioria dos desvios podem ter relação direta com a lesão laríngea, o que pode ser melhor investigado. Nesta pesquisa, procurou-se avaliar apenas o distúrbio vocal presente e não a causa, em si.

5.2 Contribuições da Pesquisa

As contribuições desta pesquisa de Dissertação são as seguintes:

1. Aplicação dos métodos tradicionais dependentes e não dependentes da frequência fundamental (*jitter*, *shimmer*, frequência fundamental, frequências formânticas, GNE, Análise LPC, Análise cepstral e mel-cepstral) em sinais de vozes para discriminar sinais que apresentam desvios vocais de sinais saudáveis, utilizando técnicas de processamento digital de sinais de voz;
2. Comparação das medidas tradicionais com as medidas obtidas da Análise LPC, Análise cepstral e mel-cepstral na classificação dos desvios vocais, observando quais demonstram maior potencial na discriminação dos desvios vocais;
3. Investigação do uso da quantização vetorial na classificação dos desvios (a quantidade de níveis e de coeficientes utilizados na extração de características), observando quais implicações ela poderia trazer na discriminação dos sinais e sua influência sobre os coeficientes LPC, cepstrais e mel-cepstrais;
4. Investigação do uso de redes neurais MLP com as medidas tradicionais, que dependam e não dependam da frequência fundamental, na classificação dos desvios vocais.
5. Investigação do uso de redes neurais MLP com a Análise LPC, cepstral e mel-cepstral na classificação dos desvios vocais.

5.3 Sugestão para Trabalhos Futuros

As sugestões para trabalhos futuros estão elencadas a seguir.

1. Realização de uma investigação à cerca da combinação das medidas tradicionais com as medidas das análises LPC, cepstral e mel-cepstral para observar se o potencial discriminativo dessas medidas, quando combinadas, podem influenciar na classificação dos desvios vocais;
2. Realização de um estudo a respeito de outras características extraídas a partir da Análise Cepstral (coeficientes delta-cepstrais, cepstrais ponderados, delta-cepstrais ponderados, entre outros) para observar se eles podem quantificar melhor os desvios comparados às características estudadas nesta pesquisa.
3. Analisar melhor a obtenção dos coeficientes mel-cepstrais como o número de coeficientes, número de filtros utilizados, combinação dos coeficientes delta-cepstrais, cepstrais ponderados e delta-cepstrais ponderados na extração de características;
4. Realização de um estudo a respeito da influência do grau do desvio vocal e a presença ou ausência de lesão laríngea na classificação dos sinais utilizando as medidas estudadas na pesquisa;
5. Investigação do uso de outros classificadores na discriminação dos desvios, como máquinas de vetor de suporte e outras redes neurais além das redes MLP, HMM (Modelo de Markov Escondido), bem como a investigação do potencial das medidas extraídas dos sinais de vozes com a combinação desses classificadores.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALCAIM, A.; OLIVEIRA, C. A. dos S. **Fundamentos do Processamento de Sinais de Voz e Imagem**. Ed. Interciência, 1ªEd. Rio de Janeiro, RJ, 2011.
- ALMEIDA, N. C. **Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor de Suporte**. 119 f. Dissertação de Mestrado, Universidade Federal do Rio Grande do Norte, RN, 2010. Disponível em: <https://repositorio.ufrn.br/jspui/bitstream/123456789/15149/1/NathaleeCA DISSERT.pdf>.
- ARJMANDI, M.; POOYAN, M. An optimum algorithm in pathological voice quality assessment using wavelet packet-based features, linear discriminant analysis and support vector machine. **Biomedical Signal Processing and Control**, v. 7, n. 1, p. 3– 19, jan 2012.
- AUGUSTI, A. C. V.; OLIVEIRA, I. B. de. Disfonia Pós Laringectomia Parcial: Análise Perceptivo-Auditiva da Voz e Correlatos Com A Análise Acústica. In: **Anais do XVI Encontro de Iniciação Científica e I Encontro de Iniciação em Desenvolvimento Tecnológico e Inovação da PUC-Campinas**, SP, 2011.
- BANDEIRA, M. P. A.; NETO, O. P. **Análise da Frequência Fundamental Comparativa das Vozes Disfônicas e Normais do Professor**. Disponível em: http://unicastelo.br/epginic2016/edicoes_antiores/files/2014/EPG/Engenharias/291%20-%20EPG225.pdf
- BARAVIEIRA, P. B. **Aplicação de uma rede neural artificial para avaliação da rugosidade e soprovidade vocal**. 101 f. Tese de doutorado, Universidade de São Paulo, SP, 2016.
- BEBER, B. C.; CIELO, C. A. Características vocais acústicas de Homens com Voz e Laringe Normal. **Revista CEFAC**, 2011. Mar-Abr, ed. 13, vol 2, pp. 340-351.
- BEHLAU, M., Voz, **O Livro do Especialista**, vol. 1. Reimpressão, Rio de Janeiro: Revinter, 2008.
- BRANDI, E. A qualidade vocal. In: Brandi E. **Educação da voz falada – a terapêutica da conduta vocal**. São Paulo: Atheneu; 2002. p.157-92.
- CARRASCO, E. R.; OLIVEIRA, G.; BEHLAU, M. Análise Perceptivo-Auditiva e Acústica da Voz de Indivíduos Gagos. **Revista CEFAC**, São Paulo, 2010. Disponível em <http://www.scielo.br/pdf/rcefac/2010ahead/179-09.pdf>.
- CARRILLO, L.; ORTIZ, K. Z. Análise vocal (auditiva e acústica) nas disartrias. **Pró-Fono Revista de Atualização Científica**, 2007 out-dez, vol. 19, nº 4. Disponível em: <http://www.scielo.br/pdf/pfono/v19n4/a10v19n4.pdf>
- COLTON, R. H.; CASPER, J. K.; LEONARD, R. **Understanding voice problems: A physiological perspective for diagnosis and treatment**. Wolters Kluwer Health, 2006.
- COSTA, S. L. do N. C. **Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas**. 161 f. Tese de Doutorado, Universidade

Federal de Campina Grande, PB, 2008. [Online]. Disponível em http://www.difusaocientifica.com.br/dt/Tese_Silvana_Cunha.pdf.

COSTA, S. L. do N. C. **Pré-processamento de Sinais de Voz: Análise de sinais de voz a curto intervalo de tempo.** Apresentação *Power Point*. Mar. 2015.

COSTA, S. C.; COSTA, W. C. de A.; CORREIA, S. E. N.; ARAÚJO, J. M. F. R. de; VIEIRA, V. J. D. Análise de Sinais de Voz para Caracterização de Patologias na Laringe. **Revista de Tecnologia da Informação e Comunicação**. Vol. 4, nº. 2, Out. 2014 [Online]. Disponível em: www.rtic.com.br/en/artigos/v04n02/v04n02a09.pdf.

COSTA, W. C. d. A. **Análise dinâmica não linear de sinais de voz para detecção de patologias laríngeas.** 176 f. Tese de Doutorado, Universidade Federal de Campina Grande, PB, 2012.

COSTA, W. C. de A.; COSTA, S. L. do N. C.; ASSIS, F. M.de; AGUIAR NETO, B. G. Classificação de Sinais de Vozes Saudáveis e Patológicas por meio da Combinação entre Medidas da Análise Dinâmica Não Linear e Codificação Preditiva Linear. **Revista Brasileira de Engenharia Biomédica**, [Online]. Vol. 29, nº 1, pp. 3-14, Mar, 2013 [Online]. Disponível em: <http://dx.doi.org/10.4322/rbeb.2013.010>.

COSTA, W. C. de A.; ASSIS, F. M.; NETO, B. G. A.; COSTA, S. C.; VIEIRA, V. J. D. Pathological Voice Classification Based on Recurrence Quantification Measures. In: **XXX Simpósio Brasileiro de Telecomunicações- SBRT** 12. Set. 2012. [Online]. Disponível em: http://sbrt.org.br/sbirt2012/publicacoes/98742_1.pdf.

DAJER, M. E. **Análise de sinais de voz com padrões visuais de dinâmica vocal.** 119 f. Tese de doutorado, Escola de Engenharia de São Carlos, Universidade de São Paulo, SP, 2010. [Online]. Disponível em: www.teses.usp.br/teses/disponiveis/82/82131/tde-16052006-095323/pt-br.php.

DELLER, J.; PROAKIS, J.; HANSEN, J. *Discrete-time Processing of Speech Signals*. Wiley-IEEE Transactions on Information Theory Press, 2000.

DINIZ, S. dos S.; THOMÉ, A. C. G. Uso de Técnicas Neurais para o reconhecimento de comandos a voz. **Revista Militar de Ciência e Tecnologia**, v. 15, nº 01, jan-mar, 1998. Disponível em: http://rmct.ime.eb.br/arquivos/RMCT_1_tri_1998/uso_tec_neurais_recon_cmd_voz.pdf.

ESMAILI, I.; DABANLOO, N. J.; MAGHOOLI, K., Bidirectional Neural Network for Pathological Voice Detection. In: **20th Iranian Conference on Biomedical Engineering (ICBME)**, 2013.

FANT, G. *Speech Acoustics and Phonetics [: Selected Writings]*, vol. 24. Springer, 2004.

FECHINE, J. M. **Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística.** 237 f. Tese de Doutorado, Universidade Federal da Paraíba, 2000.

FRAZÃO, M. do C. de L.; COSTA, S. L. do N. C.; CORREIA, S. E. N. Análise Acústica de Sinais de Vozes Patológicas por meio de Características Temporais. **Revista Princípios**, jun. 2013, João Pessoa, PB, n.º 22, p. 28-32. Disponível em: periodicos.ifpb.edu.br/index.php/principia/article/download/123/98.

GODINO, L. J. I. P. GOMEZ, V.; BLANCO, V. M. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. Biomedical Engineering. In: **IEEE Transactions**, vol. 53, no. 10, pp. 1943–1953, 2006.

GUSMÃO, C. de S.; CAMPOS, P. H.; MAIA, M. E. O. O Formante do Cantor e os Ajustes Laríngeos Utilizados para realizá-lo: Uma Revisão Descritiva. **Revista Acadêmica de Música**, n.º 21, 120 p., jan.- jul., 2010. Disponível em: <http://www.scielo.br/pdf/pm/n21/a05n21.pdf>.

HAYKIN, S. **Redes Neurais, Princípios e Práticas**. Porto Alegre, RS. BOOKMAN, 2001.

HOSMER, D. W. JR.; LEMESHOW, S. **Applied logistic regression**. 2.ed. New York: John Wiley & Sons; 2000.

KASUYA, H.; ENDO, Y.; SALIU, S. Novel acoustic measurements of *jitter* and *shimmer* characteristics from pathological voice. In: **Third European Conference on Speech Communication and Technology**, 1993.

KAY ELEMETRICS. **Disordered Voce Database**. Model 4337. 3th ed. United States, 1994.

LE HUCHE, F.; ALLALI, A. **A Voz – Patologia de Origem Funcional**. 2ª Edição, Porto Alegre: Artmed Editora, 2005.

LINDE, Y.; BUZO, A.; GRAY, R.M. An Algorithm for Vector Quantizer Design. **IEEE Transactions on Communications**, Vol. COM - 28, No. 1, pp. 84-95, Jan, 1980.

LOPES, L. W.; SILVA, J. D. da; SIMÕES, L. B.; EVANGELISTA, D. da S.; SILVA, P. O. C.; ALMEIDA, A. A.; LIMA-SILVA, M. F. B. de. Relationship Between Acoustic Measurements and Self-evaluation in Patients With Voice Disorders. **Journal of Voice**, Articles in Press, 2016. [Online]. Disponível em: <http://dx.doi.org/10.1016/j.jvoice.2016.02.021>.

LOPES, L. W.; SILVA, H. F. da; EVANGELISTA, D. da S.; SILVA, J. D. da; SIMÕES, L. B.; SILVA, P. O. C.; LIMA-SILVA, M. F. B. de; ALMEIDA, A. A. F. de. Relationship between vocal symptoms, severity of voice disorders, and laryngeal diagnosis in patients with voice disorders. **CoDAS** [Online], vol.28, n.º.4, pp.439-445, Jun, 2016. Disponível em: http://www.scielo.br/scielo.php?pid=S2317-17822016000400439&script=sci_abstract&tlng=pt.

LOPES, L. W.; EVANGELISTA, D. da S.; FRANÇA, F. P.; SIMÕES, L. B.; SILVA, J. D. da; VIEIRA, V. J. D. **Acurácia das medidas acústicas tradicionais e formânticas na**

discriminação de vozes saudáveis e desviadas. [Online]. Out. 2016. In: Anais do XXIV Congresso Brasileiro de Fonoaudiologia.

LLORENTE, J. I. G.; RUIZ, V. O.; LECHÓN, N. S.; VILDA, P. G.; VELASCO, M. B.; ROLDÁN, F. C. The Effectiveness of the Glottal to Noise Excitation Ratio for the Screening of Voice Disorders. **Journal of Voice**, v. 24, nº 1, 2010.

MA, E. P. M.; YIU, E. M. L. Multiparametric Evaluaton of Dysphonic Severity. **Journal of Voice**, v. 20, nº 3, p.380-390, 2006. Disponível em: <http://dx.doi.org/10.1016/j.jvoice.2005.04.007>.

MAGRI, A.; CUKIER-BLAJ, S.; KARMAN, D. de F.; CAMARGO, Z. A. de. Correlatos Perceptivos e Acústicos dos Ajustes Supraglóticos na Disfonia. **Revista CEFAC**, São Paulo, vol.9, nº.4, pp.512-518, out-dez, 2007. Disponível em: <http://www.scielo.br/pdf/rcefac/v9n4/10.pdf>.

MAKHOUL, J.; ROUCOS, S.; GISH, H. Vector Quantization in Speech Coding. In: **Proceedings of the IEEE**, Vol. 73, nº.11, pp.1551-1588, Nov, 1985.

MARTINEZ, E. Z.; LOUZADA-NETO, F.; PEREIRA, B. B. A Curva ROC para Testes Diagnósticos. **Caderno de Saúde Coletiva**, Rio de Janeiro, vol. 11, nº 1, pp. 7-31, 2007.

MARTINS, R.; RIBEIRO, C.; MELLO, B.; BRANCO, A.; TAVARES, E. Dysphonia in Children. **Journal of Voice**, v. 26, nº 5, 2012.

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. **CoDAS**, vol. 27 nº. 3, pp.279-84, 2015. Disponível em: http://www.scielo.br/pdf/codas/v27n3/pt_2317-1782-codas-27-03-00279.pdf.

MICHAELIS, D. GRAMSS, T. STRUBE. H.W. Glottal-to-Noise Excitation Ratio – a New Measure for Describing Pathological Voices. **ACUSTICA, Acta Acústica**, Vol. 83, 1997. Disponível em: <http://wwwuser.gwdg.de/~istuerm/doc/acustica97.pdf>

MICHAELIS, D.; STRUBE, H. W. Empirical study to test the independence of different acoustic voice parameters on a large voice database. In: **Eurospeech**. 1995.

MOLLER, M. F. **A scaled conjugate gradient algorithm for fast supervised learning.** *Neural Networks*, vol. 6, nº 4, p. 525–533, 1993.

MORETI F.; ZAMBOM, F.; OLIVEIRA, G.; BEHLAU, M. Adaptação transcultural da versão brasileira da escala de sintomas de voz: VoiSS. **Jornal Sociedade Brasileira de Fonoaudiologia**, 2011, ed. 23, vol. 4, pp.398-400.

MOTA, L. A. A.; SANTOS, C. M. B.; VASCONCELOS, J. M. de; MOTA, B. C.; MOTA, H. de S. C. Aplicação da técnica de emissão em tempo máximo de fonação em paciente com disfonia espasmódica adutora: relato de caso. **Revista Sociedade Brasileira de Fonoaudiologia**, 2012, vol. 17, nº.3, pp. 351-356.

NAYAK, J.; BHAT, P. S.; ACHARYA, R.; AITHAL, U. V. **Classification and analysis of speech abnormalities**. ITBM-RBM, 2005. Vol. 26, pp. 319-27.

OLIVEIRA, R.; TEIXEIRA, L.; GAMA, A.; MEDEIROS, A. **Análise perceptivo-auditiva, acústica e autopercepção vocal em crianças**. *Jornal da Sociedade Brasileira de Fonoaudiologia*, vol 23, nº 2, pp. 158-63, 2011.

O'SHAUGHNESSY, D. **Speech Communications: Human and Machine**. Ed. 2ª, Edition: NY, IEEE Press, 2000.

PARRAGA, A. **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. 163 f. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, RS, 2002.

PIFAIA, L. R.; MADAZIO, G.; BEHLAU, M. Diagrama de Desvio Fonatório e Análise Perceptivo Auditiva Pré e Pós-Terapia Vocal. **CoDAS**, [Online]. Vol. 25, nº 2, 2013. Disponível em <http://www.scielo.br/pdf/codas/v25n2/a09v25n2.pdf>

PINHATI, F.; Ferlin, C. Musik: um Ambiente Inteligente de Auxílio à Composição. **Revista Faeterj** – Rio. Mai, 2014. [Online]. Disponível em: http://www.faeterj-rio.edu.br/layout/downloads/artigos/art_musik_rev.pdf.

RABINER, L. R. and SCHAFER, R. W. **Digital Processing of Speech Signals**. Prentice Hall, Upper Saddle River, New Jersey, 1978.

SANTOS, M. L. R. dos. **Segmentação da Fala utilizando as Transformadas Wavelets Discretas – DWT**. 105 f. Trabalho de Conclusão de Curso, Universidade Federal do Rio Grande do Sul, RS, 2004. [Online]. Disponível em http://www.inf.ufrgs.br/~danielnm/docs/TCC_Marcio_Santos.pdf.

SANTOS, M. O. **Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet**. 79 f. Dissertação de mestrado, Instituto Federal da Paraíba, PB, 2015. Disponível em: https://editor.ifpb.edu.br/campi/joao-pessoa/cursos/pos-graduacao-stricto-sensu/mestrado-em-engenharia-eletrica/arquivos/Analise_Acustica_de_Desvios_Vocais_Infantis_utilizando_a_Transformada_Wavelet.pdf.

SILVA, D. D. C. da. **Reconhecimento de Fala Contínua para o Português Brasileiro em Sistemas Embarcados**. 194 f. Tese de doutorado, Universidade Federal de Campina Grande, PB, 2011.

SIMSEK, B.; TANSEL, Y.; HÜSEYİN, E. Hybridizing a fuzzy multi-response Taguchi optimization algorithm with artificial neural networks to solve standard ready-mixed concrete optimization problems. In: **Int. J. of Computational Intelligence Systems**, Vol. 9, No. 3; pp.525-543, 2016.

SIMÕES, F. O.; NETO, M. U.; MACHADO, J. B.; NAGLE, E. J.; RUNSTEIN, F. O.; GOMES, L. de C. T. Compressão de fala utilizando quantização vetorial e redes neurais não supervisionadas. **Cadernos CPqD Tecnologia**, Campinas, v. 5, n. 1, p. 33-48, jan./jun. 2009.

SONDHI, M. M. New methods of pitch extraction, *Audio and Electroacoustics*, In: **IEEE Transactions on**, vol. 16, no. 2, pp. 262–266, 1968.

SOUSA, R. J. T. de. **Metodologias de Avaliação Perceptiva e Acústica do Sinal de Voz em Aplicações de Ensino do Canto e Diagnóstico/Reabilitação da Fala**. Dissertação de mestrado, Universidade do Porto, POR, 2011.

TAVARES, R.; BRUNET, N. A.; CORREIA, S.; COSTA, S. C.; AGUIAR NETO, B. G.; FECHINE, J. M. Combinação de Classificadores Baseados em Análise LPC e Cepstral para a Detecção de Paralisia nas Dobras Vocais. In: **anais do XXII Congresso Brasileiro de Engenharia Biomédica**, 2010.

TEIXEIRA, A. **Conceitos de Sinais e Sistemas Mestrado em Ciências da Fala e da Audição**, 2004. [Online]. Notas de Aula. Disponível em: sweet.ua.pt/ajst/css/aula12css.pp. Acessado em 11/02/17.

TEIXEIRA, J. P.; FERREIRA, D. B.; CARNEIRO, S. M. **Análise Acústica Vocal – Determinação do jitter e do shimmer para Diagnóstico de Patologias da Fala**, 2011. Escola Superior de Tecnologia e Gestão Instituto Politécnico de Bragança - Bragança, POR. Disponível em https://bibliotecadigital.ipb.pt/bitstream/10198/7282/1/artigo_publicado.pdf. Acessado em 05/11/15.

TELES, V. de C.; ROSINHA, A. C. U. **Análise Acústica dos Formantes e das Medidas de Perturbação do Sinal Sonoro em Mulheres Sem Queixas Vocais, Não Fumantes e Não Etilista**. *Arq. Int. Otorrinolaringol. / Intl. Arch. Otorhinolaryngol*, São Paulo, vol. 12, nº 4, p. 523-530, 2008. Disponível em <http://www.arquivosdeorl.org.br/conteudo/pdfForl/567.pdf>.

VIEIRA, V. J. D. **Avaliação de Distúrbios da Voz por meio de Análise de Quantificação de Recorrência**. 218 f. Dissertação de mestrado, Instituto Federal da Paraíba, PB, 2014.

VISAVE, A.; KACHARE, P.; JEYAKUMAR, A.; CHEERAN, A.; NIRMAL, J. Glottal Pathology Discrimination using ANN and SVM. In: **International Conference on Advances in Computing, Communications and Informatics (ICACCI)**, 2015.

WERTZNER, H. F.; SCHREIBER, S.; AMARO, L. Análise da frequência fundamental, *jitter*, *shimmer* e intensidade vocal em crianças com transtorno fonológico. **Revista Brasileira de Otorrinolaringologia** [Online]. Vol.71 no.5 São Paulo Sept./Oct. 2005. Disponível em: <http://dx.doi.org/10.1590/S0034-72992005000500007>.

ZHANG W.; XU, G.; WANG, Y. Pitch estimation based on circular AMDF. In: **Proceedings of International Conference on Acoustics, Speech, and Signal Processing**. Vol. 1, 2002.

ZWETSCH, I. C.; FAGUNDES, R. D. R.; RUSSOMANO, T. E, SCOLARI, D. Processamento digital de sinais no diagnóstico diferencial de doenças laringeas benignas. **Scientia Medica**, Porto Alegre: PUCRS, v. 16, n. 3, jul./set. 2006. Disponível em <http://revistaseletronicas.pucrs.br/ojs/index.php/scientiamedica/article/viewFile/1912/1417>. Data de acesso: 06/02/2017.

APÊNDICES

A Classificação segundo a Escala Analógica-Visual (EAV) pra os sinais selecionados da base de dados LIEV/UFPB

A seguir são apresentadas as tabelas com a classificação segundo a Escala Analógica Visual. Nas tabelas são apresentados os valores obtidos pela avaliação perceptivo-auditiva para os sinais rugosos, soprosos e tensos dos sinais selecionados da base de dados LIEV/UFPB, mostrando os pontos da escala EAV e o grau para cada tipo de desvio. E se o sinal possui ou não lesão na laringe.

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

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

VOZES RUGOSAS	EAV (0-100 mm)			EAV (Grau)			LESÃO NA LARINGE
	RUG	SOP	TEN	RUG	SOP	TEN	
VGE 03	47	38	12	2	2	1	SIM
VGE 13	54	48	44	3	2	2	SIM
VGE 16	41	36,5	10,5	2	2	1	SIM
VGE 39	42	37	20	2	2	1	NÃO
VGE 29	92	66	78	4	3	3	SIM
VGE 41	50	45	38	2	2	2	SIM
VGE 56	58	52	33	3	3	1	SIM
VGE 75	62	52,5	49	3	3	2	SIM
VGE 80	61,5	52	46	3	3	2	SIM
VGE 97	78	69	69,5	3	3	3	SIM
VGE 106	39,5	6	35	2	1	1	SIM
VGE 107	45	38	33,5	2	2	1	SIM
VGE 46	50,5	46	37	2	2	2	SIM
VGE 112	47,5	40,5	15,5	2	2	1	SIM
VGE 48	50,5	38	46,5	2	2	2	SIM
VGE 79	39	9	32	2	1	1	SIM
VGE 85	41	36	27	2	2	1	SIM
VGE 90	47	41	35,5	2	2	1	SIM
VGE 93	36	5	21	2	1	1	SIM
VGE 95	36	28,5	23	2	1	1	NÃO
VGE 116	63,5	38,5	59,5	3	2	3	SIM
VGE 120	51,5	45,5	38	3	2	2	SIM
VGE 127	48	17	41,5	2	1	2	SIM
VGE 128	81	36	36,5	3	2	2	SIM
VGE 137	38	22,5	20	2	1	1	NÃO
VGE 145	38	33	28	2	1	1	NÃO
VGE 161	90	46,5	70	3	2	3	SIM
VGE 165	45	11	37	2	1	2	NÃO
VGE 189	64	56	51	3	3	3	SIM
VGE 199	54	43	24	3	2	1	SIM

EAV: Escala Analógica-visual; RUG: rugosa; TEN: tensa; SOP: soprosa.

A partir da Tabela A 1 pode-se observar que dos 30 sinais rugosos selecionados 11 possuem grau 3 (rugosidade grau moderado), 18 sinais possuem grau 2 (rugosidade grau leve) e apenas 1 sinal possui grau 4 (rugosidade grau intenso), onde a maioria possui grau leve para sopro ou tensão. Observa-se também que 75% dos sinais (25 sinais) possuem lesão na laringe (edemas, nódulos, paralisia nas pregas vocais, entre outros).

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

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

VOZES TENSAS	EAV (0-100 mm)			EAV (Grau)			LESÃO NA LARINGE
	RUG	SOP	TEN	RUG	SOP	TEN	
VGE 04	36	11	41	2	1	2	SIM
VGE 30	55,5	61,5	68	3	3	3	SIM
VGE 77	32	8,5	38	1	1	2	SIM
VGE 37	40	5	46,5	2	1	2	SIM
VGE 38	38	16	43	2	1	2	SIM
VGE 60	45,5	3	52	2	1	3	SIM
VGE 81	43	5	55,5	2	1	3	SIM
VGE 100	36	9	40	2	1	2	SIM
VGE 180	66	5	80	3	1	3	NÃO
VGE 248	44	6	55	2	1	3	SIM
VGE 299	39,5	8	48,5	2	1	2	SIM
VGE 312	39	3	47	2	1	2	NÃO
VGE 371	41	2	48	2	1	2	SIM
VGE 454	34	6	59	1	1	3	NÃO
VGE 494	16	52	88	1	3	3	SIM
VGE 117	31	1,5	39	1	1	2	SIM
VGE 198	37	4	63	2	1	3	NÃO
VGE 62	28	3	37,5	1	1	2	SIM
VGE 284	44,5	8	52	2	1	3	SIM
VGE 21	57	7	64	3	1	3	NÃO
VGE 231	33	7	65	1	1	3	SIM
VGE 241	28	4	45	1	1	2	NÃO
VGE 252	34	3	41	1	1	2	SIM
VGE 267	35	6	44,5	1	1	2	SIM
VGE 561	41	9	54	2	1	3	NÃO
VGE 347	49,5	35	60	2	1	3	SIM
VGE 148	45	27	50	2	1	2	SIM
VGE 31	40	28,5	46,5	2	1	2	SIM
VGE 156	50	3	61	2	1	3	SIM
VGE 192	34	3	43	1	1	2	NÃO

EAV: Escala Analógica-visual; RUG: rugosa; TEN: tensa; SOP: sopro.

A partir da Tabela A 2 pode-se observar que dos 30 sinais tensos selecionados 14 possuem grau 3 (tensão grau moderado) e 16 sinais possuem grau 2 (tensão grau leve), onde a maioria possui grau normal para sopro e grau leve para rugosidade. Observa-se também que 66% dos sinais (22 sinais) possuem lesão na laringe (edemas, nódulos, paralisia nas pregas vocais, entre outros).

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

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

VOZES SORPOSAS	EAV (0-100 mm)			EAV (Grau)			LESÃO NA LARINGE
	RUG	SOP	TEN	RUG	SOP	TEN	
VGE 304	37	52	5	2	3	1	SIM
VGE 319	70	96	83,5	3	4	3	SIM
VGE 327	55,5	65,5	13	3	3	1	SIM
VGE 335	42	50	25	2	2	1	SIM
VGE 350	36	44	5	2	2	1	SIM
VGE 358	55,5	75	6	3	3	1	SIM
VGE 361	64	83,5	44	3	3	2	SIM
VGE 364	39	47	7,5	2	2	1	SIM
VGE 366	50	56,5	12	2	3	1	SIM
VGE 12	50	64	57	2	3	3	SIM
VGE 15	42,5	50	10	2	2	1	SIM
VGE 36	47	57	31	2	3	1	SIM
VGE 78	36	47	7,5	2	2	1	SIM
VGE 114	38,5	50,5	11,5	2	2	1	SIM
VGE 121	68	76	43	3	3	2	SIM
VGE 125	31	44	3	1	2	1	SIM
VGE 183	40	59	36,5	2	3	2	SIM
VGE 376	65	77	8	3	3	1	SIM
VGE 45	62	72	57,5	3	3	3	SIM
VGE 417	65	93,5	83	3	4	3	SIM
VGE 243	44	54,5	4,5	2	3	1	SIM
VGE 10	31	46	14	1	2	1	SIM
VGE 58	18	39,5	33	1	2	1	SIM
VGE 84	18	38	4	1	2	1	NÃO
VGE 94	28	40	6	1	2	1	SIM
VGE 232	32,5	41,5	7,5	1	2	1	SIM
VGE 239	35	41	11	1	2	1	NÃO
VGE 245	42,5	54	5	2	3	1	NÃO
VGE 257	38	51	6	2	3	1	NÃO
VGE 19	36	41,5	4	2	2	1	NÃO

EAV: Escala Analógica-visual; RUG: rugosa; TEN: tensa; SOP: soprosa.

A partir da Tabela A 3 pode-se observar que dos 30 sinais tensos selecionados 14 possuem grau 3 (soprosidade grau moderado), 14 sinais possuem grau 2 (soprosidade grau leve) e 2 sinais possuem grau 4 (soprosidade grau intenso), onde a maioria possui grau normal para tensão e graus leve e moderado para rugosidade. Observa-se também que 75% dos sinais (25 sinais) possuem lesão na laringe (edemas, nódulos, paralisia nas pregas vocais, entre outros).

B Tabelas com a classificação realizada a partir da curva ROC pra as distorções com os métodos empregados

A seguir são apresentadas as tabelas com a classificação obtida pela curva ROC com as distorções encontradas para os coeficientes LPC, cepstrais e mel-cepstrais. São

demonstradas as taxas de correta aceitação (CA), falsa aceitação (FA), correta rejeição (CR), acurácia (ACUR), sensibilidade (SEM) e especificidade (ESP) pra cada método e também para cada nível.

B 1. Tabelas com a classificação realizada a partir da curva ROC para as distorções com os coeficientes LPC

Tabela B 1 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Rugosas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,35	0,60	0,65	62,50	60,00	65,00
32	0,60	0,85	0,40	62,50	85,00	40,00
64	0,55	0,65	0,45	55,00	65,00	45,00
128	0,40	0,65	0,60	62,50	65,00	60,00
256	0,60	0,70	0,40	55,00	70,00	40,00
512	0,45	0,55	0,55	55,00	55,00	55,00
1024	0,40	0,40	0,60	50,00	40,00	60,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 2 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Tensas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,25	0,68	0,75	71,71	68,42	75,00
32	0,30	0,68	0,70	69,21	68,42	70,00
64	0,60	0,57	0,40	48,94	57,89	40,00
128	0,40	0,47	0,60	53,68	47,37	60,00
256	0,45	0,57	0,55	56,44	57,89	55,00
512	0,45	0,57	0,55	56,44	57,89	55,00
1024	0,50	0,42	0,50	46,05	42,11	50,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 3 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Soprosas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,65	0,63	0,35	49,07	63,16	35,00
32	0,25	0,47	0,75	61,18	47,37	75,00
64	0,35	0,42	0,65	53,55	42,11	65,00
128	0,35	0,52	0,65	58,81	52,63	65,00
256	0,55	0,68	0,45	56,71	68,42	45,00
512	0,40	0,57	0,60	58,94	57,89	60,00
1024	0,40	0,63	0,60	61,57	63,16	60,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 4 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Desviadas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,50	0,68	0,50	59,48	68,97	50,00
32	0,60	0,72	0,40	56,20	72,41	40,00
64	0,45	0,46	0,55	50,77	46,55	55,00
128	0,40	0,56	0,60	58,44	56,90	60,00
256	0,65	0,58	0,35	46,81	58,62	35,00
512	0,55	0,51	0,45	48,36	51,72	45,00
1024	0,40	0,50	0,60	55,00	50,00	60,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 5 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Rugosas Vs. Tensas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,36	0,60	0,63	61,57	60,00	63,15
32	0,21	0,30	0,78	54,47	30,00	78,94
64	0,42	0,65	0,57	61,44	65,00	57,89
128	0,42	0,65	0,57	61,44	65,00	57,89
256	0,42	0,65	0,57	61,44	65,00	57,89
512	0,42	0,45	0,57	51,44	45,00	57,89
1024	0,42	0,45	0,57	51,44	45,00	57,89

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 6 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Rugosas Vs. Soprosas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,63	0,70	0,36	53,42	70,00	36,84
32	0,26	0,40	0,73	56,84	40,00	73,68
64	0,52	0,65	0,47	56,18	65,00	47,36
128	0,68	0,75	0,31	53,28	75,00	31,57
256	0,31	0,35	0,68	51,71	35,00	68,42
512	0,31	0,50	0,68	59,21	50,00	68,42
1024	0,42	0,65	0,57	61,44	65,00	57,89

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 7 - Classificação Curva ROC para as distorções com os coeficientes LPC - Vozes Tensas Vs. Soprosas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,26	0,63	0,73	68,42	63,16	73,68
32	0,47	0,52	0,52	52,63	52,63	52,63
64	0,47	0,57	0,52	55,26	57,89	52,63
128	0,47	0,68	0,52	60,52	68,42	52,63
256	0,26	0,57	0,73	65,78	57,89	73,68
512	0,42	0,68	0,57	63,15	68,42	57,89
1024	0,52	0,68	0,47	57,89	68,42	47,36

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

B 2. Tabelas com a classificação realizada a partir da curva ROC para as distorções com os coeficientes cepstrais

Tabela B 8 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Rugosas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,25	0,65	0,75	70,0	65,00	75,00
32	0,35	0,65	0,65	65,0	65,00	65,00
64	0,60	0,70	0,4	55,0	70,00	40,00
128	0,35	0,65	0,65	65,0	65,00	65,00
256	0,20	0,65	0,80	72,5	65,00	80,00
512	0,60	0,65	0,40	52,5	65,00	40,00
1024	0,40	0,35	0,60	47,5	35,00	60,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 9 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Tensas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,55	0,68	0,45	56,71	68,42	45,00
32	0,65	0,63	0,35	49,07	63,16	35,00
64	0,25	0,31	0,75	53,28	31,58	75,00
128	0,30	0,52	0,70	61,31	52,63	70,00
256	0,35	0,57	0,65	61,44	57,89	65,00
512	0,25	0,84	0,75	79,60	84,21	75,00
1024	0,20	0,89	0,80	84,73	89,47	80,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 10 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Soprosas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,55	0,78	0,45	61,97	78,95	45,00
32	0,30	0,10	0,70	40,65	26,32	55,00
64	0,60	0,78	0,40	59,47	78,95	40,00
128	0,45	0,73	0,55	64,34	73,68	55,00
256	0,20	0,84	0,80	82,10	84,21	80,00
512	0,40	0,63	0,60	61,57	63,16	60,00
1024	0,40	0,78	0,60	69,47	78,95	60,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 11 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Desviadas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,40	0,46	0,60	53,27	46,55	60,00
32	0,35	0,34	0,65	49,74	34,48	65,00
64	0,45	0,43	0,55	49,05	43,10	55,00
128	0,25	0,39	0,75	57,32	39,66	75,00
256	0,40	0,62	0,60	61,03	62,07	60,00
512	0,35	0,58	0,65	61,81	58,62	65,00
1024	0,50	0,75	0,50	62,93	75,86	50,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 12 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Rugosas Vs. Tensas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,57	0,65	0,42	53,55	65,00	42,10
32	0,21	0,65	0,78	71,97	65,00	78,94
64	0,36	0,5	0,63	56,57	50,00	63,15
128	0,15	0,65	0,84	74,60	65,00	84,21
256	0,26	0,70	0,73	71,84	70,00	73,68
512	0,10	0,65	0,89	77,23	65,00	89,47
1024	0,21	0,75	0,78	76,97	75,00	78,94

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 13 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Rugosas Vs. Soprosas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,21	0,65	0,78	71,97	65,00	78,94
32	0,15	0,65	0,84	74,60	65,00	84,21
64	0,42	0,50	0,57	53,94	50,00	57,89
128	0,36	0,65	0,63	64,07	65,00	63,15
256	0,42	0,40	0,57	48,94	40,00	57,89
512	0,63	0,65	0,36	50,92	65,00	36,84
1024	0,21	0,65	0,78	71,97	65,00	78,94

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 14 - Classificação Curva ROC para as distorções com os coeficientes cepstrais - Vozes Tensas Vs. Soprosas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,15	0,57	0,84	71,05	57,89	84,21
32	0,15	0,31	0,84	57,89	31,58	84,21
64	0,57	0,63	0,42	52,63	63,16	42,10
128	0,10	0,36	0,89	63,15	36,84	89,47
256	0,10	0,73	0,89	81,57	73,68	89,47
512	0,10	0,57	0,89	73,68	57,89	89,47
1024	0,10	0,63	0,89	76,31	63,16	89,47

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

B 3. Tabelas com a classificação realizada a partir da curva ROC para as distorções com os coeficientes mel-cepstrais

Tabela B 15 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Rugosas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,15	0,65	0,85	75,00	65,00	85,00
32	0,25	0,70	0,75	72,5	70,00	75,00
64	0,15	0,60	0,85	72,5	60,00	85,00
128	0,50	0,60	0,50	55,0	60,00	50,00
256	0,60	0,60	0,40	50,0	60,00	40,00
512	0,45	0,55	0,55	55,0	55,00	55,00
1024	0,40	0,85	0,60	72,5	85,00	60,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 16 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Tensas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,40	0,73	0,60	66,84	73,68	60,00
32	0,40	0,52	0,60	56,31	52,63	60,00
64	0,40	0,63	0,60	61,57	63,16	60,00
128	0,25	0,84	0,75	79,60	84,21	75,00
256	0,20	0,84	0,80	82,10	84,21	80,00
512	0,15	0,84	0,85	84,60	84,21	85,00
1024	0,20	0,84	0,80	82,10	84,21	80,00

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 17 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Soprosas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,2	0,78	0,8	79,47	78,95	80,0
32	0,15	0,78	0,85	81,97	78,95	85,0
64	0,45	0,78	0,55	66,97	78,95	55,0
128	0,5	0,68	0,5	59,21	68,42	50,0
256	0,3	0,52	0,7	61,31	52,63	70,0
512	0,6	0,68	0,4	54,21	68,42	40,0
1024	0,6	0,73	0,4	56,84	73,68	40,0

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 18 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Desviadas Vs. Saudáveis.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,15	0,70	0,85	77,84	70,69	85,0
32	0,4	0,58	0,6	59,31	58,62	60,0
64	0,5	0,55	0,5	52,58	55,17	50,0
128	0,4	0,53	0,6	56,72	53,45	60,0
256	0,5	0,51	0,5	50,86	51,72	50,0
512	0,45	0,58	0,55	56,81	58,62	55,0
1024	0,4	0,68	0,6	64,48	68,97	60,0

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 19 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Rugosas Vs. Tensas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,31	0,35	0,68	51,71	35,00	68,42
32	0,15	0,7	0,84	77,10	70,00	84,21
64	0,21	0,7	0,78	74,47	70,00	78,94
128	0,10	0,7	0,89	79,34	85,00	73,68
256	0,15	0,7	0,84	77,10	70,00	84,21
512	0,31	0,75	0,68	71,71	75,00	68,42
1024	0,21	0,7	0,78	74,47	70,00	78,94

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 20 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Rugosas Vs. Soprosas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,21	0,35	0,78	56,97	35,00	78,94
32	0,52	0,8	0,47	63,68	80,00	47,36
64	0,47	0,4	0,52	46,31	40,00	52,63
128	0,21	0,45	0,78	61,97	45,00	78,94
256	0,31	0,4	0,68	54,21	40,00	68,42
512	0,47	0,7	0,52	61,31	70,00	52,63
1024	0,15	0,6	0,84	72,10	60,00	84,21

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela B 21 - Classificação Curva ROC para as distorções com os coeficientes mel-cepstrais - Vozes Tensas Vs. Soprosas.

Nível	FA	CA	CR	ACUR (%)	SEN (%)	ESP (%)
16	0,73	0,68	0,26	47,36	68,42	26,315
32	0,15	0,84	0,84	84,21	84,21	84,21
64	0,21	0,84	0,78	81,57	84,21	78,94
128	0,21	0,84	0,78	81,57	84,21	78,94
256	0,21	0,84	0,78	81,57	84,21	78,94
512	0,15	0,84	0,84	84,21	84,21	84,21
1024	0,10	0,84	0,89	86,84	84,21	89,47

FA: falsa aceitação; CA: correta aceitação; CR: correta rejeição; ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

C Tabelas apresentando a classificação realizada com o classificador MLP com os *codebooks* como características de entrada

A seguir são apresentadas as tabelas com a classificação obtida pelo classificador MLP com os *codebooks* dos coeficientes LPC, cepstrais e mel-cepstrais. São demonstradas as taxas

de acurácia (ACUR), sensibilidade (SEN) e especificidade (ESP) pra cada método e também para cada nível de quantização empregado.

C 1. Classificação realizada com o classificador MLP com os *codebooks* LPC como características de entrada

Tabela C 1 - Classificação com os *codebooks* LPC (Classificador MLP) – Vozes Rugosas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	50.00 ± 0.00	50.00 ± 1.02	51.40 ± 1.02
32	53.91 ± 0.85	52.91 ± 1.05	51.33 ± 1.35
64	53.91 ± 0.96	52.88 ± 1.20	56.01 ± 1.91
128	57.42 ± 0.91	56.07 ± 0.70	55.23 ± 1.22
256	50.00 ± 0.89	50.00 ± 0.43	50.00 ± 1.02
512	50.00 ± 0.00	51.20 ± 0.45	52.11 ± 0.88
1024	50.00 ± 0.00	50.00 ± 0.00	50.00 ± 0.00

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 2 - Classificação com os *codebooks* LPC (Classificador MLP) – Vozes Tensas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	54.69 ± 1.14	55.49 ± 3.08	54.58 ± 0.85
32	56.25 ± 1.37	59.09 ± 1.55	54.76 ± 4.09
64	55.47 ± 0.28	58.06 ± 1.35	54.32 ± 0.55
128	54.10 ± 0.76	57.28 ± 1.03	52.70 ± 1.15
256	71.29 ± 0.18	81.70 ± 0.67	66.03 ± 0.34
512	53.81 ± 1.79	52.61 ± 2.13	56.22 ± 2.99
1024	61.67 ± 1.45	58.33 ± 2.01	68.65 ± 1.95

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 3 - Classificação com os *codebooks* LPC (Classificador MLP) – Vozes Soprosas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	50.00 ± 1.06	50.00 ± 2.95	51.20 ± 1.35
32	50.00 ± 0.72	50.65 ± 1.54	50.00 ± 0.89
64	52.34 ± 0.47	51.59 ± 4.90	52.09 ± 2.45
128	53.32 ± 0.83	54.15 ± 0.78	52.78 ± 1.36
256	53.13 ± 0.50	52.31 ± 1.33	54.78 ± 0.99
512	51.81 ± 0.56	51.20 ± 0.79	54.22 ± 1.95
1024	53.52 ± 0.59	52.23 ± 1.18	57.74 ± 1.10

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 4 - Classificação com os *codebooks* LPC (Classificador MLP) – Vozes Desviadas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	75.00 ± 0.94	75.00 ± 0.18	75.82 ± 1.09
32	75.00 ± 0.07	75.00 ± 0.02	76.13 ± 0.55
64	75.00 ± 0.42	75.00 ± 0.08	75.92 ± 1.08
128	75.00 ± 0.66	75.00 ± 0.19	75.00 ± 0.23
256	75.00 ± 0.15	75.00 ± 0.04	75.00 ± 0.34
512	75.00 ± 0.00	75.00 ± 0.00	75.00 ± 0.03
1024	75.00 ± 0.04	75.00 ± 0.01	75.27 ± 0.27

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 5 - Classificação com os *codebooks* LPC (Classificador MLP) – Vozes Rugosas Vs. Tensas.

Nível	ACUR (%)
16	56.25 ± 1.14
32	54.69 ± 0.97
64	58.59 ± 1.42
128	60.35 ± 0.88
256	71.19 ± 0.16
512	47.27 ± 2.20
1024	61.65 ± 2.33

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 6 - Classificação com os *codebooks* LPC (Classificador MLP) – Vozes Rugosas Vs. Soprosas.

Nível	ACUR (%)
16	53.13 ± 1.40
32	51.56 ± 0.43
64	50.39 ± 0.96
128	56.05 ± 0.44
256	53.52 ± 0.28
512	52.29 ± 0.51
1024	53.49 ± 0.19

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 7 - Classificação com os *codebooks* LPC (Classificador MLP) – Vozes Tensas Vs. Soprosas.

Nível	ACUR (%)
16	53.13 ± 1.02
32	52.34 ± 1.30
64	54.30 ± 1.03
128	55.27 ± 0.33
256	75.00 ± 0.26
512	51.22 ± 1.18
1024	52.32 ± 1.77

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

C 2. Classificação realizada com o classificador MLP com os *codebooks* Cepstrais como características de entrada

Tabela C 8 - Classificação com os *codebooks* cepstrais (Classificador MLP) – Vozes Rugosas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	50.00 ± 0.98	50.00 ± 1.41	51.05 ± 0.83
32	50.00 ± 0.33	51.23 ± 0.53	52.04 ± 0.83
64	50.00 ± 0.73	50.00 ± 0.95	50.00 ± 1.01
128	50.78 ± 0.15	51.55 ± 0.45	50.63 ± 0.21
256	52.25 ± 0.20	51.56 ± 1.10	53.48 ± 0.54
512	51.51 ± 0.41	56.81 ± 2.74	50.82 ± 0.60
1024	51.20 ± 0.27	51.58 ± 1.86	52.07 ± 2.09

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 9 - Classificação com os *codebooks* cepstrais (Classificador MLP) – Vozes Tensas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	50.00 ± 0.47	50.48 ± 0.97	50.00 ± 0.52
32	51.56 ± 0.34	50.91 ± 1.68	50.86 ± 0.83
64	50.39 ± 0.54	50.23 ± 0.85	50.58 ± 2.81
128	50.00 ± 0.13	50.33 ± 0.56	50.69 ± 0.76
256	50.88 ± 0.26	51.22 ± 0.48	50.68 ± 0.29
512	52.34 ± 0.43	51.31 ± 0.24	60.77 ± 1.95
1024	52.20 ± 0.42	51.50 ± 0.62	57.36 ± 1.49

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 10 - Classificação com os *codebooks* cepstrais (Classificador MLP) – Vozes Soprosas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	51.56 ± 1.23	51.92 ± 4.96	51.76 ± 2.06
32	51.56 ± 0.59	51.04 ± 0.89	51.62 ± 4.94
64	50.00 ± 0.52	50.00 ± 4.95	50.00 ± 2.56
128	51.17 ± 0.62	51.46 ± 0.78	51.32 ± 6.01
256	73.14 ± 0.61	67.71 ± 0.68	84.00 ± 0.25
512	51.42 ± 0.74	50.77 ± 0.86	54.06 ± 4.24
1024	51.27 ± 0.21	50.77 ± 0.10	53.79 ± 1.12

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 11 - Classificação com os *codebooks* cepstrais (Classificador MLP) – Vozes Desviadas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	75.00 ± 0.23	75.00 ± 0.06	75.00 ± 0.56
32	75.00 ± 1.69	75.00 ± 0.14	75.60 ± 0.96
64	75.00 ± 0.03	75.00 ± 0.01	75.00 ± 0.32
128	75.00 ± 0.04	75.00 ± 0.01	75.34 ± 0.40
256	75.00 ± 0.01	75.00 ± 0.02	75.07 ± 0.31
512	75.00 ± 0.10	75.00 ± 0.03	75.45 ± 0.70
1024	75.68 ± 0.00	75.65 ± 0.00	79.17 ± 1.30

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 12 - Classificação com os *codebooks* cepstrais (Classificador MLP) – Vozes Rugosas Vs. Tensas.

Nível	ACUR (%)
16	50.00 ± 0.00
32	51.56 ± 0.34
64	50.78 ± 0.24
128	50.59 ± 0.40
256	53.61 ± 0.57
512	53.91 ± 0.16
1024	49.93 ± 0.56

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 13 - Classificação com os *codebooks* cepstrais (Classificador MLP) – Vozes Rugosas Vs. Soprosas.

Nível	ACUR (%)
16	56.25 ± 1.41
32	51.56 ± 0.91
64	51.17 ± 0.39
128	52.15 ± 0.54
256	76.17 ± 0.46
512	52.00 ± 0.32
1024	49.98 ± 0.08

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 14 - Classificação com os *codebooks* cepstrais (Classificador MLP) – Vozes Tensas Vs. Soprosas.

Nível	ACUR (%)
16*	53.13 ± 1.02
32	52.34 ± 1.30
64	54.30 ± 1.03
128	55.27 ± 0.33
256	75.00 ± 0.26
512	51.22 ± 1.18
1024	52.32 ± 1.77

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

C 3. Classificação realizada com o classificador MLP com os *codebooks* mel-cepstrais como características de entrada

Tabela C 15 - Classificação com os *codebooks* mel-cepstrais (Classificador MLP) – Vozes Rugosas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	50.00 ± 0.62	50.00 ± 0.02	50.00 ± 0.75
32	51.56 ± 0.31	50.32 ± 0.01	51.07 ± 0.98
64	53.13 ± 0.63	52.04 ± 0.70	55.10 ± 0.99
128	51.76 ± 0.66	51.42 ± 0.08	52.46 ± 1.01
256	52.73 ± 0.20	52.38 ± 0.14	53.22 ± 0.56
512	50.98 ± 0.41	50.99 ± 0.51	51.20 ± 0.76
1024	53.83 ± 0.42	52.87 ± 0.28	55.68 ± 0.74

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 16 - Classificação com os *codebooks* mel-cepstrais (Classificador MLP) – Vozes Tensas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	50.00 ± 1.04	50.00 ± 1.43	50.00 ± 1.22
32	49.22 ± 0.89	49.98 ± 0.76	49.42 ± 5.03
64	53.91 ± 0.62	54.58 ± 0.82	53.22 ± 0.05
128	53.91 ± 0.54	56.02 ± 1.08	52.76 ± 1.34
256	55.37 ± 0.35	58.98 ± 0.80	53.91 ± 0.23
512	52.05 ± 0.38	52.49 ± 0.53	52.15 ± 0.50
1024	51.81 ± 0.52	51.32 ± 0.59	53.03 ± 0.61

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 17 - Classificação com os *codebooks* mel-cepstrais (Classificador MLP) – Vozes Soprosas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	50.00 ± 0.72	50.00 ± 0.64	50.00 ± 0.03
32	52.34 ± 0.46	53.59 ± 0.90	52.09 ± 0.70
64	50.00 ± 0.20	50.00 ± 0.33	50.00 ± 0.05
128	49.80 ± 0.33	49.86 ± 0.69	49.97 ± 0.98
256	50.00 ± 0.18	50.00 ± 0.28	50.00 ± 0.03
512	51.71 ± 0.25	51.11 ± 0.27	52.03 ± 0.64
1024	50.53 ± 0.98	50.00 ± 0.92	50.89 ± 0.45

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 18 - Classificação com os *codebooks* mel-cepstrais (Classificador MLP) – Vozes Desviadas Vs. Saudáveis.

Nível	ACUR (%)	SEN (%)	ESP (%)
16	75.00 ± 6.43	75.23 ± 3.49	75.09 ± 2.87
32	75.00 ± 0.00	75.00 ± 0.00	75.00 ± 0.04
64	75.00 ± 0.00	75.00 ± 0.04	75.40 ± 0.22
128	75.00 ± 0.13	75.00 ± 0.03	75.00 ± 0.27
256	75.00 ± 0.05	75.00 ± 0.06	75.00 ± 0.32
512	75.00 ± 0.00	75.00 ± 0.00	75.25 ± 0.32
1024	74.99 ± 0.01	75.00 ± 0.00	74.78 ± 0.06

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 19 - Classificação com os *codebooks* mel-cepstrais (Classificador MLP) – Vozes Rugosas Vs. Tensas.

Nível	ACUR (%)
16	51.56 ± 1.68
32	56.25 ± 1.02
64	57.03 ± 1.00
128	55.27 ± 1.30
256	58.59 ± 0.38
512	53.61 ± 0.43
1024	51.76 ± 0.31

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 20 - Classificação com os *codebooks* mel-cepstrais (Classificador MLP) – Vozes Rugosas Vs. Soprosas.

Nível	ACUR (%)
16	50.00 ± 0.41
32	53.13 ± 0.58
64	51.17 ± 0.43
128	51.76 ± 0.64
256	53.13 ± 0.46
512	49.95 ± 0.40
1024	51.05 ± 0.23

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

Tabela C 21 - Classificação com os *codebooks* mel-cepstrais (Classificador MLP) – Vozes Tensas Vs. Soprosas.

Nível	ACUR (%)
16	51.56 ± 0.90
32	51.56 ± 1.37
64	54.30 ± 0.85
128	51.37 ± 0.91
256	54.49 ± 0.39
512	51.17 ± 0.78
1024	51.81 ± 0.15

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade.

D Tabelas apresentando a classificação realizada com o classificador MLP com a média dos coeficientes como características de entrada

A seguir são apresentadas as tabelas com a classificação obtida pelo classificador MLP com a média dos coeficientes LPC, cepstrais e mel-cepstrais. São demonstradas as taxas de

acurácia (ACUR), sensibilidade (SEN) e especificidade (ESP) e seus respectivos desvios na classificação entre vozes desviadas e saudáveis (Desviadas Vs. Saudáveis, Rugosas Vs. Saudáveis, Tensas Vs. Saudáveis, Soprosas Vs. Saudáveis) e a taxa de acurácia (ACUR) e seu desvio na classificação entre os desvios vocais (Rugosas Vs. Tensas, Rugosas Vs. Soprosas e Tensas Vs. Soprosas) pra cada método empregado.

D 1. Classificação realizada com o classificador MLP com a média dos coeficientes LPC como características de entrada

Tabela D 1 - Classificação com a média dos coeficientes LPC (LPC_{med}) realizada pelo Classificador MLP.

Classificação	ACUR (%)	SEN (%)	ESP (%)
DES Vs. SDL	77.73 ± 1.71	80.40 ± 1.19	79.23 ± 1.04
RUG Vs. SDL	81.67 ± 3.22	77.95 ± 3.45	82.97 ± 3.09
TEN Vs. SDL	71.19 ± 3.93	71.19 ± 3.93	73.78 ± 4.00
SOP Vs. SDL	75.83 ± 4.38	75.00 ± 5.49	81.93 ± 3.95

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade; DES: desviada; RUG: rugosa, TEN: tensa, SOP: soprosa; SDL: saudável; LPC_{med} : média dos coeficientes LPC.

Tabela D 2 - Classificação com a média dos coeficientes LPC (LPC_{med}) realizada pelo Classificador MLP.

Classificação	ACUR (%)
RUG Vs. TEN	71.19 ± 3.81
RUG Vs. SOP	57.50 ± 2.90
TEN Vs. SOP	61.02 ± 3.64

ACUR: acurácia; RUG: rugosa, TEN: tensa, SOP: soprosa; LPC_{med} : média dos coeficientes LPC.

D 2. Classificação realizada com o classificador MLP com a média dos coeficientes cepstrais como características de entrada

Tabela D 3 - Classificação com a média dos coeficientes cepstrais ($CEPS_{med}$) realizada pelo Classificador MLP.

Classificação	ACUR (%)	SEN (%)	ESP (%)
DES Vs. SDL	77.54 ± 1.93	84.54 ± 1.90	81.32 ± 1.54
RUG Vs. SDL	77.12 ± 3.88	82.74 ± 4.51	73.13 ± 7.88
TEN Vs. SDL	58.62 ± 3.17	64.64 ± 3.65	67.24 ± 2.05
SOP Vs. SDL	83.05 ± 2.30	82.29 ± 2.51	82.05 ± 2.44

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade; DES: desviada; RUG: rugosa, TEN: tensa, SOP: soprosa; SDL: saudável; $CEPS_{med}$: média dos coeficientes cepstrais.

Tabela D 4 - Classificação com a média dos coeficientes cepstrais ($CEPS_{med}$) realizada pelo Classificador MLP.

Classificação	ACUR (%)
RUG Vs. TEN	71.19 ± 4.24
RUG Vs. SOP	74.17 ± 3.94
TEN Vs. SOP	77.97 ± 2.99

ACUR: acurácia; RUG: rugosa, TEN: tensa, SOP: soprosa; $CEPS_{med}$: média dos coeficientes cepstrais.

D 3. Classificação realizada com o classificador MLP com a média dos coeficientes mel-cepstrais como características de entrada

Tabela D 5 - Classificação com a média dos coeficientes mel-cepstrais (MEL_{med}) realizada pelo Classificador MLP.

Classificação	ACUR (%)	SEN (%)	ESP (%)
DES Vs. SDL	73.93 ± 2.79	75.11 ± 0.35	74.40 ± 1.25
RUG Vs. SDL	51.69 ± 1.40	53.27 ± 1.33	54.05 ± 2.04
TEN Vs. SDL	57.76 ± 2.86	56.96 ± 3.32	58.77 ± 3.28
SOP Vs. SDL	75.86 ± 3.70	77.75 ± 4.03	81.44 ± 3.68

ACUR: acurácia, SEN: sensibilidade; ESP: especificidade; DES: desviada; RUG: rugosa, TEN: tensa, SOP: soprosa; SDL: saudável.
 MEL_{med} : média dos coeficientes mel-cepstrais.

Tabela D 6 - Classificação com a média dos coeficientes mel-Cepstrais (MEL_{med}) realizada pelo Classificador MLP.

Classificação	ACUR (%)
RUG Vs. TEN	60.17 ± 3.64
RUG Vs. SOP	59.32 ± 3.13
TEN Vs. SOP	54.31 ± 2.24

ACUR: acurácia; RUG: rugosa, TEN: tensa, SOP: soprosa; MEL_{med} : média dos coeficientes mel-cepstrais.