

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
DEPARTAMENTO DE ENGENHARIA ELETRÔNICA
ENGENHARIA ELETRÔNICA

Ana Beatriz Rodrigues Costa de Mattos

**Avaliação da capacidade de estatísticas de distribuições da proeminência do pico
cepstral de vogais sustentadas em distinguir disfonia hipercinética, disfonia hipocinética
e laringite causada por refluxo**

Florianópolis

2020

Ana Beatriz Rodrigues Costa de Mattos

**Avaliação da capacidade de estatísticas de distribuições da proeminência do pico
cepstral de vogais sustentadas em distinguir disfonia hipercinética, disfonia hipocinética
e laringite causada por refluxo**

Trabalho de Conclusão do Curso de Graduação em Engenharia Eletrônica do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito para a obtenção do título de Bacharel em Engenharia Eletrônica.

Orientador: Prof. Dr. Bruno Catarino Bispo

Florianópolis

2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Mattos, Ana Beatriz Rodrigues Costa de
Avaliação da capacidade de estatísticas de distribuições
da proeminência do pico cepstral de vogais sustentadas em
distinguir disfonia hipercinética, disfonia hipocinética e
laringite causada por refluxo. / Ana Beatriz Rodrigues
Costa de Mattos ; orientador, Bruno Catarino Bispo, 2020.
96 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia Elétrica, Florianópolis, 2020.

Inclui referências.

1. Engenharia Elétrica. 2. Processamento Digital de
Sinais. 3. Cepstro. 4. CPP. 5. CPPS. I. Catarino Bispo,
Bruno. II. Universidade Federal de Santa Catarina.
Graduação em Engenharia Elétrica. III. Título.

Ana Beatriz Rodrigues Costa de Mattos

**Avaliação da capacidade de estatísticas de distribuições da proeminência do pico
cepstral de vogais sustentadas em distinguir disfonia hipercinética, disfonia hipocinética
e laringite causada por refluxo**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Engenharia Eletrônica e aprovado em sua forma final pelo Curso Engenharia Eletrônica

Florianópolis, 23 de janeiro de 2020.



Prof. Fernando Rangel de Souza, Dr.
UFSC

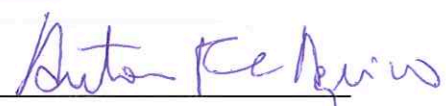
Banca Examinadora:



Prof. Bruno Catarino Bispo, Dr.
UFSC



Prof. Hans Helmut Zürn, PhD.
UFSC



Prof. Antonio Felipe da Cunha de Aquino, Dr.
UFSC

Este trabalho é dedicado aos meus pais e minha irmã.

AGRADECIMENTOS

Gostaria de agradecer primeiramente a UFSC e aos professores por me proporcionarem um ensino de qualidade. Agradecer principalmente ao professor Bruno Bispo por me orientar neste trabalho e por toda a disponibilidade e atenção.

Gostaria de também agradecer aos meus pais e minha irmã Isadora por me apoiarem durante toda a graduação e ao Eduardo por me acompanhar desde o início e sempre me ajudar quando precisei.

"A voz humana é o instrumento mais rico que há."

(Antônio Carlos Jobim)

RESUMO

Este trabalho tem como objetivo avaliar as distribuições da proeminência do pico cepstral (CPP) e da proeminência do pico cepstral suavizado (CPPS) para a vogal sustentada /a/ e suas estatísticas descritivas como discriminantes entre vozes saudáveis e vozes de pacientes patológicos diagnosticados com disfonia hiper e hipocinética e laringite causada por refluxo. As medidas foram calculadas em decibel para 184 vozes, sendo 51 saudáveis e 133 com patologias. As patologias foram separadas em 3 categorias, disfonia hipercinética, disfonia hipocinética e laringite causada por refluxo, contendo 67, 34 e 32 vozes respectivamente e foi realizada a remoção de outliers das 3 categorias e de todas as vozes patológicas e saudáveis. As medidas de CPP e CPPS foram avaliadas por 10 estatísticas distributivas, média, mediana, quinto percentil, 95-ésimo percentil, desvio padrão, curtose, assimetria, intervalo, módulo e variância para a obtenção da maior capacidade discriminante.

Após avaliar as 10 estatísticas da distribuição de 2 medidas de CPP, calculado a cada 2 ms e a cada 10 ms e 2 medidas de CPPS, calculado a cada 2ms e suavização de 7 e 11 amostras para 4 casos, sendo eles saudável versus patológicos, saudável versus disfonia hipercinética, saudável versus disfonia hipocinética e saudável versus laringite causada por refluxo, o quinto percentil do CPP calculado a cada 2 ms para o caso vozes saudáveis versus disfonia hipocinética apresentou a maior capacidade discriminante, alcançando um valor-p de $3,63E-08$, uma precisão de 81,18% e uma área ROC de 0,8326. Utilizando SVM foi possível melhorar este resultado e obter 88,09% de precisão de classificação, com as estatísticas de média, assimetria e quinto percentil.

Palavras-chave: Cepstro; CPP; CPPS; Disfonia Hipercinética; Disfonia Hipocinética; Laringite.

ABSTRACT

This study aims to evaluate the distributions of the cepstral peak prominence (CPP) and the cepstral peak prominence smoothed (CPPS) for the sustained vowel /a/ and their descriptive statistics as discriminating between healthy voices and voices of patients diagnosed with hyper and hypokinetic dysphonia and laryngitis caused by reflux. Measurements were calculated in decibel for 184 voices, 51 healthy and 133 with pathologies. The pathologies were divided into 3 categories, hyperkinetic dysphonia, hypokinetic dysphonia and laryngitis caused by reflux, containing 67, 34 and 32 voices respectively and the removal of outliers from the 3 categories and all pathological and healthy voices was performed. The CPP and CPPS measurements were evaluated by 10 distributive statistics, mean, median, fifth percentile, 95th percentile, standard deviation, kurtosis, asymmetry, interval, mode and variance to obtain the highest discriminant capacity.

After evaluating the 10 distribution statistics of 2 CPP measurements, calculated every 2 ms and every 10 ms and 2 CPPS measurements, calculated every 2ms and smoothing 7 and 11 samples for 4 cases, healthy versus pathological, healthy versus hyperkinetic dysphonia, healthy versus hypokinetic dysphonia and healthy versus reflux laryngitis, the fifth percentile CPP calculated every 2 ms for the case healthy voices versus hypokinetic dysphonia had the highest discriminant capacity, reaching a p-value of 3.63E-08, an accuracy of 81.18% and a ROC area of 0.8326. Using SVM it was possible to improve this result and to obtain 88.09% of classification accuracy, with the statistics of average, asymmetry and fifth percentile.

Keywords: Cepstrum; CPP; CPPS; Hyperkinetic Dysphonia; Hypokinetic Dysphonia; Laryngitis.

LISTA DE FIGURAS

Figura 1: Diafragma na inspiração e expiração.	28
Figura 2: Funcionamento das cordas vocais na fala e na respiração.	29
Figura 3: Esquema do aparelho vocal.	30
Figura 4: Diagrama de blocos da produção de voz.	31
Figura 5: Exemplos de patologias vocais.	32
Figura 6: Cordas vocais na respiração e na fala sem paralisia, com paralisia unilateral e com paralisia bilateral.	35
Figura 7: Cordas vocais normais e cordas vocais inflamadas devido a laringite.	36
Figura 8: Espelho para laringe.	37
Figura 9: Laringoscópio flexível.	38
Figura 10: Laringoscópio rígido.	38
Figura 11: Videostroboscópio.	39
Figura 12: Exame de laringoscopia indireta.	40
Figura 13: Nasolaringoscopia.	41
Figura 14: Exame realizado com Laringoscópio Rígido.	41

Figura 15: Exame de Videoestroboscopia.	43
Figura 16: Cepstro.	45
Figura 17: Cepstro e linha de regressão linear.	47
Figura 18: Cepstro suavizado e linha de regressão linear.	48
Figura 19: Exemplo de resultado de duas distribuições.	51
Figura 20: Exemplo de Curva ROC e área abaixo da curva.	51
Figura 21: Exemplo de outlier.	52
Figura 22: Exemplo de separação entre classes.	54
Figura 23: Boxplot $CPPS_{5prc}$ utilizando $T = 2$ ms, $L = M = 7$ e escala em dB.	68
Figura 24: Valores para cada grupo e limiar de classificação de 12,22 dB para $CPPS_{5prc}$ utilizando $T = 2$ ms, $L = M = 7$ e escala em dB.	68
Figura 25: Boxplot CPP_{median} utilizando $T = 10$ ms e escala dB.	69
Figura 26: Valores para cada grupo e limiar de classificação para CPP_{median} utilizando $T = 10$ ms e escala em dB, limiar 31,57dB.	69
Figura 27: Boxplot $CPPS_{5prc}$ utilizando $T = 2$ ms, $L=M=7$ e escala em dB	72
Figura 28: Valores para cada grupo e limiar de classificação para $CPPS_{5prc}$ utilizando $T = 2$ ms, $L = M = 11$, limiar de 12,16 dB e escala em dB.	73

Figura 29: Boxplot CPP_{std} utilizando $T = 2$ ms e escala em dB, valor-p de $1,00E-03$.	73
Figura 30: Valores para cada grupo e limiar de classificação para CPP_{std} utilizando $T = 2$ ms, limiar de 2,14 dB e escala em dB.	74
Figura 31: Boxplot $CPPS_{5prc}$ utilizando $T = 2$ ms, $L=M=7$ e escala em dB.	77
Figura 32: Valores para cada grupo e limiar de classificação para $CPPS_{5prc}$ utilizando $T = 2$ ms, $L = M = 7$, limiar de 9,40 dB e escala em dB.	77
Figura 33: Boxplot $CPP5_{prc}$ utilizando $T = 2$ ms e escala em dB.	78
Figura 34: Valores para cada grupo e limiar de classificação para CPP_{5prc} utilizando $T = 2$ ms, limiar de 23,16 dB e escala em dB.	78
Figura 35: Boxplot $CPPS_{5prc}$ utilizando $T = 2$ ms, $L=M=7$ e escala em dB.	82
Figura 36: Valores para cada grupo e limiar de classificação para $CPPS_{5prc}$ utilizando $T = 2$ ms, $L = M = 7$, limiar de 9,93 dB e escala em dB.	82
Figura 37: Boxplot $CPPS_{std}$ utilizando $T = 2$ ms, $L=M=11$ e escala em dB.	83
Figura 38: Valores para cada grupo e limiar de classificação para $CPPS_{std}$ utilizando $T = 2$ ms, $L = M = 11$, limiar de 2,11 dB e escala em dB.	83

LISTA DE TABELAS

Tabela 1: Quantidade de vozes saudáveis, patológicas e número total de vozes.	56
Tabela 2: Quantidade de pacientes com disфония hipercinética, hipocinética e laringite.	57
Tabela 3: Remoção de outliers Vozes Saudáveis.	62
Tabela 4: Remoção de outliers Vozes Patológicas.	63
Tabela 5: Remoção de outliers Vozes com Disфония Hipercinética.	64
Tabela 6: Remoção de outliers Disфония Hipocinética.	65
Tabela 7: Remoção de outliers Vozes com Laringite causada por Refluxo.	65
Tabela 8: Valor-p das Estatísticas das Distribuições.	66
Tabela 9: Precisão de Classificação das Estatísticas das Distribuições.	66
Tabela 10: Melhores precisões de classificação utilizando SVM.	70
Tabela 11: Valor-p das Estatísticas das Distribuições para Vozes Saudáveis versus Vozes com Disфония Hipercinética.	71
Tabela 12: Precisão de Classificação das Estatísticas das Distribuições das Vozes Saudáveis versus Vozes com Disфония Hipercinética.	71
Tabela 13: Melhores precisões de classificação utilizando SVM para vozes saudáveis versus disфония hipercinética.	75

Tabela 14: Valor-p das Estatísticas das Distribuições para Vozes Saudáveis versus Vozes com Disfonia Hipocinética.	75
Tabela 15: Precisão de Classificação das Estatísticas das Distribuições das Vozes Saudáveis versus Vozes com Disfonia Hipocinética	76
Tabela 16: Melhores precisões de classificação utilizando SVM para vozes saudáveis versus Disfonia Hipocinética.	79
Tabela 17: Valor-p das Estatísticas das Distribuições para vozes saudáveis versus vozes com laringite causada por refluxo.	80
Tabela 18: Precisão de Classificação das Estatísticas das Distribuições das Vozes Saudáveis versus Vozes com laringite causada por refluxo.	80
Tabela 19: Melhores precisões de classificação utilizando SVM para vozes saudáveis versus vozes com laringite causada por refluxo.	84

LISTA DE ABREVIATURAS E SIGLAS

ANOVA Análise de variância (Analysis of Variance)

AROC Receiver Operating Characteristic Area

Assim. Assimetria

CPP Proeminência do Pico Cepstral (Cepstral Peak Prominence)

CPPS Proeminência do Pico Cepstral Suavizada (Cepstral Peak Prominence Smoothed)

Curt. Curtose

dB Decibel

DTFT Transformada de Fourier de tempo discreto (Discrete-time Fourier transform)

FFT Transformada rápida de Fourier (fast Fourier transform)

IFFT Transformada rápida de Fourier Inversa (Inverse fast Fourier transform)

Interv. Intervalo

ROC Receiver Operating Characteristic

SVM Máquina de Vetores de Suporte (Support Vector Machine)

T Período

Var. Variância

VOICEd VOice ICar fEDerico II Database

2-D Duas Dimensões

5prc Quinto Percentil

95prc 95-percentil

SUMÁRIO

1 INTRODUÇÃO	24
1.1 Objetivos	26
1.1.1 Objetivos gerais	26
1.2 Motivação	26
2 FUNDAMENTAÇÃO TEÓRICA	28
2.1 A Fisiologia e produção da Voz Humana	28
2.2 Patologias vocais	31
2.3 Materiais e métodos de detecção de patologia	36
2.4 Cepstro	43
2.4.1 Proeminência do Pico Cepstral	45
2.4.2 Proeminência de Pico Cepstral Suavizada	47
2.5 Teste estatístico de Kruskal-Wallis	49
2.6 Classificador Binário	49
2.7 Outlier	52
2.8 Máquina de Vetores de Suporte	53
3 MATERIAIS E METODOLOGIA	55
3.1 Materiais	55
3.2 Metodologia	57
3.2.1 Proeminência do Pico Cepstral	57
3.2.2 Proeminência do Pico Cepstral Suavizada	59
4 RESULTADOS	61
4.1 Remoção de Outliers	61
4.2 Vozes saudáveis e vozes patológicas	65
4.3 Vozes saudáveis e vozes com disfonia hipercinética	70
4.4 Vozes saudáveis e vozes com disfonia hipocinética	75
4.5 Vozes saudáveis e vozes com laringite causada por refluxo	79
5 CONCLUSÃO	85
REFERÊNCIAS	88

1 INTRODUÇÃO

É cada vez mais consensual a importância da voz e do cuidado com ela. Além de ser uma das principais ferramentas de trabalho para profissionais como professores, palestrantes, cantores, telefonistas e vários outros, é também por meio da voz que grande parte da comunicação é realizada [30].

Distúrbios vocais afetam pessoas em qualquer estágio da vida e com isso são trazidas diversas deficiências no desenvolvimento pessoal e profissional do indivíduo [52]. Existem diferentes fatores de risco para o desenvolvimento de patologias vocais e para evitá-las e preveni-las são necessárias avaliações médicas e diversos exames para o paciente.

Atualmente para a detecção de patologias vocais são utilizadas principalmente técnicas invasivas, como o exame de laringoscopia, que é a observação direta do trato vocal através de um laringoscópio, ou técnicas subjetivas, através de uma análise perceptiva auditiva, o que pode levar ao diagnóstico incorreto [61].

Buscando superar tais dificuldades de diagnóstico, segundo [61], pesquisadores começaram a estudar sinais de voz para obter-se parâmetros que pudessem representar índices para diferentes aspectos da voz, incluindo patologias relacionadas com a fala. Ainda segundo [61], a análise acústica da voz utiliza o processamento digital de sinais para quantificar, detectar e classificar características e parâmetros da qualidade vocal.

Primeiramente os parâmetros para a análise acústica da voz encontravam-se no domínio do tempo, principalmente como jitter e shimmer [10]. O jitter indica a variabilidade ou perturbação da frequência fundamental e o shimmer refere-se a essa mesma perturbação, mas relacionada à amplitude da onda sonora, ou intensidade da emissão vocal [35],[58]. Ainda segundo [35] e [58], o jitter é alterado principalmente com a falta de controle de vibração das cordas vocais e o shimmer com a redução da resistência glótica e lesões de massa nas cordas vocais, estando correlacionada com a presença de ruído à emissão e com a soproidade.

Porém pelas análises acústicas no domínio do tempo dependerem do conhecimento do início e fim dos ciclos vocais ocorre uma perda de precisão para sinais altamente perturbados [10], [61]. Além disso, é necessário que a vogal seja analisada como um pico alto e estável

para calcular esses parâmetros, porque qualquer alteração no sinal pode ser considerada como perturbação vocal [70].

Buscando superar estas limitações medidas baseadas no espectro e no cepstro foram consideradas tanto para vogal sustentada quanto para fala contínua [34]. O cepstro foi proposto em 1963 como uma melhor alternativa à função de autocorrelação para detectar ecos em sinais sísmicos [62]. E, de acordo com [62], [25], parâmetros cepstrais são os índices mais promissores para detectar a severidade de disfonias na análise de sinais vocais. Na literatura, encontram-se alguns trabalhos em que utilizam a proeminência do pico cepstral (CPP) e a proeminência do pico cepstral suavizado (CPPS), propostos em [26] e [25] respectivamente, com esse objetivo. Como em [25], que o uso do CPP é avaliado para falantes com patologias na laringe. Em [10], estatísticas de distribuições do CPPS são utilizadas para discriminar vozes saudáveis e patológicas. Valores normativos para o CPPS foram estabelecidos como medida de disfonia em [24]. Mais recentemente, em [61], estatísticas do CPP e CPPS foram analisadas para a identificação de nódulos vocais.

O estudo apresentado em [10] obteve como resultado uma área ROC (Receiver Operating Characteristic) igual a 0,95 utilizando apenas uma estatística do CPPS. Tal eficiência em detectar problemas vocais, incluindo 10 diferentes tipos de patologias, com apenas uma estatística do CPPS despertou o interesse em analisar a capacidade desses parâmetros em auxiliar o diagnóstico de disfonias vocais, seguindo o realizado em [10].

Algumas bases de dados contendo sinais de vozes patológicas estão disponíveis publicamente na literatura como, por exemplo, a SVD e MEEI. A base de dados MEEI é amplamente utilizada na literatura e alguns trabalhos envolvendo CPP e CPPS podem ser encontrados como, por exemplo, em [34] e em [60]. No entanto, uma base de dados denominada VOICEd foi publicada recentemente em 2018 contendo gravações de vozes saudáveis e patológicas, divididas nas categorias disfonia hipercinética, disfonia hipocinética e laringite causada por refluxo. Apenas um trabalho, [12], foi encontrado na literatura utilizando essa base de dados. O mencionado trabalho propõe um novo marcador, denominado índice de detecção de disfonia (DDI), para avaliar global e objetivamente a qualidade da voz através da estimativa de parâmetros acústicos apropriados. Em comparação com a utilização de outras bases de dados, a aplicação desse novo marcador na base VOICEd

resultou nos piores resultados. Portanto, até o presente momento, não há conhecimento da aplicação de CPP e CPPS na base de dados VOICEd

Assim, este trabalho tem como objetivo analisar a capacidade de estatísticas de distribuições de CPP e CPPS em distinguir disfonia hipercinética, disfonia hipocinética e laringite causada por refluxo. Para isso, utilizou-se a base de dados pública VOICEd. Este trabalho será apresentado na forma que se segue. Na seção 1, após a introdução, são apresentados os objetivos e a motivação do trabalho; na seção 2, apresenta-se a fundamentação teórica para a realização do trabalho, explicando brevemente a fisiologia e produção da voz e algumas patologias vocais presentes na base de dados, além de alguns métodos invasivos que são utilizados atualmente para a detecção dessas patologias. Ainda na seção 2, é descrita a principal teoria matemática do trabalho, isto é, o cálculo do cepstro, CPP e CPPS. E também é abordado o teste de Kruskal-Wallis, o classificador binário utilizado e máquinas de vetores suporte. Na seção 3 são citados os materiais utilizados e explicada a metodologia. Na seção 4, os resultados são apresentados e discutidos. Por fim, na seção 5, apresenta-se as conclusões do trabalho.

1.1 OBJETIVOS

Na subseção abaixo estão descritos os principais objetivos gerais deste trabalho.

1.1.1 OBJETIVOS GERAIS

O principal objetivo deste trabalho é avaliar a capacidade discriminante de distribuições da proeminência do pico cepstral (CPP) e da proeminência do pico cepstral suavizado (CPPS) na vogal sustentada /a/ e utilizar 10 estatísticas descritivas como discriminantes entre vozes saudáveis e vozes patológicas, vozes saudáveis e com disfonia hipercinética, vozes saudáveis e com disfonia hipocinética e vozes saudáveis e com laringite causada por refluxo.

1.2 MOTIVAÇÃO

A voz, além de ser uma das mais importantes maneiras de expressão, é também um instrumento de trabalho para diversos profissionais, como citado anteriormente. Cada vez mais está sendo conscientizada a importância dos cuidados vocais e com a ideia mais forte de que a voz é uma ferramenta que deve ser preservada.

Como dito anteriormente, os principais meios de identificação de patologias vocais são invasivos e desconfortáveis ou subjetivos. Entendendo a importância da voz, uma das principais motivações deste trabalho é auxiliar no diagnóstico de vozes patológicas, vozes com disfonia hipercinética, vozes com disfonia hipocinética e vozes com laringite causada por refluxo e propor um método de baixo custo e não invasivo, que possa facilitar o diagnóstico e prevenir problemas vocais.

Com isso, buscou-se seguir parte do trabalho proposto em [10], tendo em vista o resultado obtido com o uso de CPPS, com $T=2\text{ms}$ e $L=M=7$, que foi de uma área ROC de 0,95 utilizando apenas uma estatística descritiva para um banco de dados com 10 patologias diferentes.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção será explicado brevemente o sistema de produção da voz e as principais patologias que serão referenciadas no trabalho. Além disso também serão explicadas as bases matemáticas e estatísticas utilizadas.

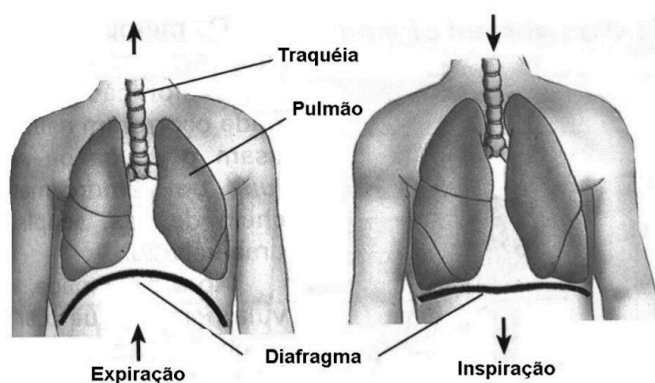
2.1 A FISIOLOGIA E PRODUÇÃO DA VOZ HUMANA

A fala, segundo [16], é usada para disseminar informações de um falante a um ouvinte. Para isso, o indivíduo transmite seu pensamento através de uma série de processos neurológicos e movimentos musculares para produzir uma onda acústica de pressão sonora que é recebida pelo sistema auditivo do ouvinte, processada e convertida novamente em sinais neurológicos [16].

Como dito anteriormente, a voz é uma das principais ferramentas de comunicação humana. A produção vocal nos seres humanos é possível graças ao sistema respiratório pulmonar, que possui diafragma, pulmões, laringe, vias aéreas superiores e vários outros músculos, como pode ser visto nas Figuras 1 e 2.

Segundo [20], o diafragma é o músculo inspiratório mais importante, pois a sua contração aumenta a capacidade das vias aéreas, permitindo a inalação de um volume maior de ar. Durante seu relaxamento, o ar é exalado dos pulmões, que pode voltar ao seu estado volumétrico inicial devido às suas propriedades elásticas, podendo ser visto o processo de expiração e inspiração na Figura 1.

Figura 1: Diafragma na inspiração e expiração.

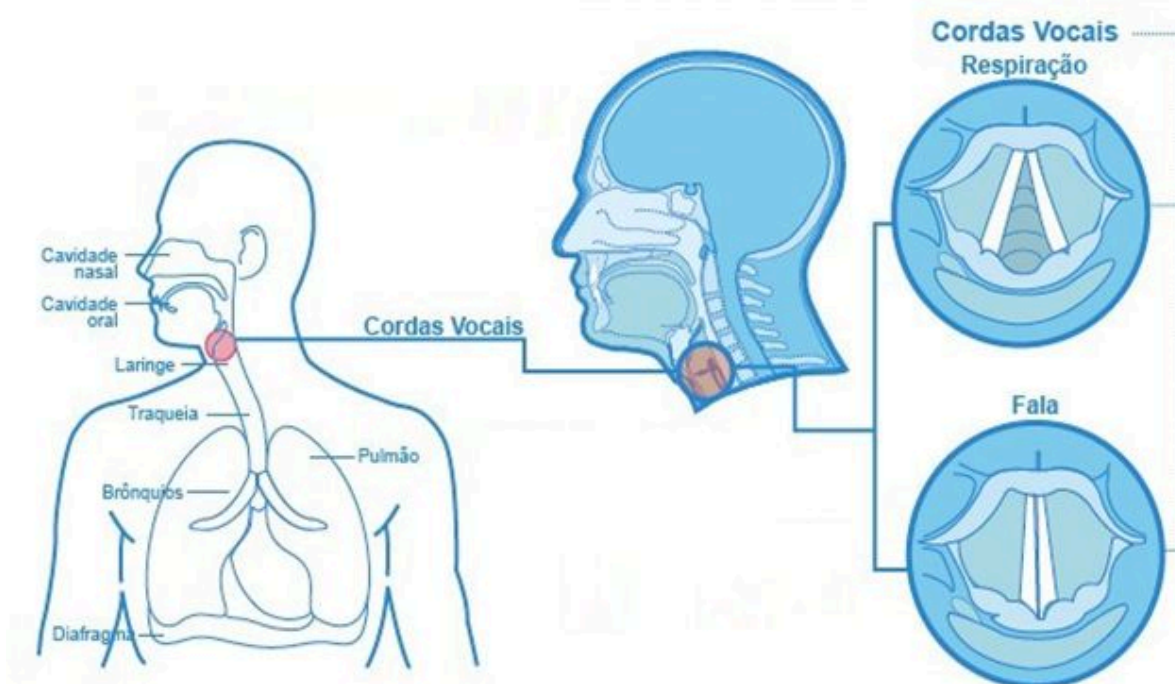


Fonte: [50].

A laringe está localizada acima da traquéia e é formada pela tireóide e cartilagens. As cartilagens fornecem reforço e evitam o colapso das vias aéreas. Os outros componentes da laringe são móveis e formam um mecanismo de fechamento que protege a traquéia durante deglutição [20].

É na laringe que situam-se as cordas vocais, que exercem função chave na produção da voz. As cordas vocais apresentam-se como dois lábios horizontais posicionados na extremidade superior da traquéia, formando saliências na parede interior da laringe, um à direita e outro à esquerda [27]. Quando ocorre a respiração as cordas vocais abrem-se e o ar entra e sai dos pulmões. Para a fala, as cordas vocais aproximam-se e o ar que sai dos pulmões, passa pelas cordas vocais, produzindo a voz. Isso pode ser visto na Figura 2.

Figura 2: Funcionamento das cordas vocais na fala e na respiração.



Fonte:[55].

As cordas vocais são compostas por um músculo denominado tireoaritenóideo e diferentes tecidos as revestem, formando assim camadas.

Existem dois grupos musculares com inserções na laringe: os músculos intrínsecos e os extrínsecos. A musculatura intrínseca está diretamente relacionada ao controle de vibração

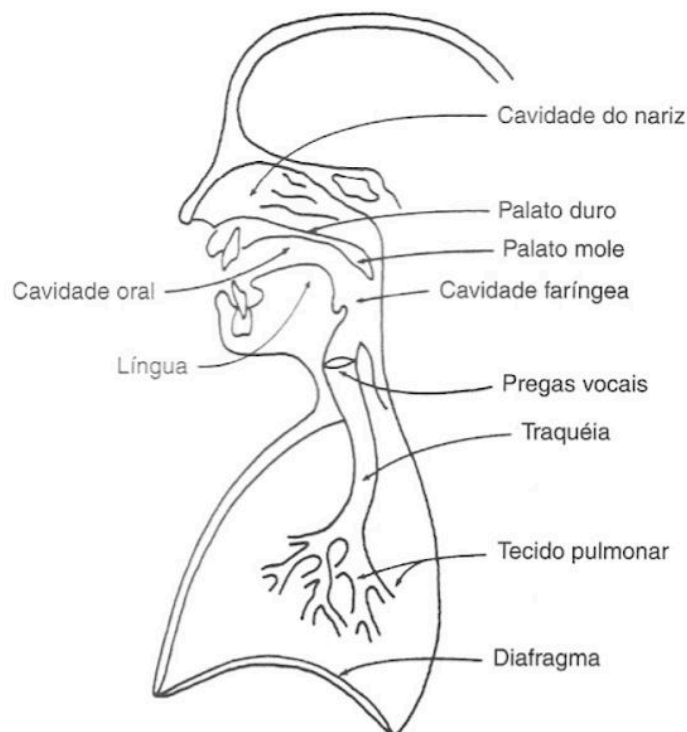
das cordas vocais, enquanto a musculatura extrínseca relaciona-se à elevação e abaixamento da laringe [27].

Segundo [54], a produção da voz começa no córtex cerebral. A ideia da vocalização é planejada e transmitida ao giro pré-central no córtex motor, que transmite outro conjunto de instruções para núcleos motores no tronco cerebral e na medula espinhal. Essas áreas transmitem as mensagens necessárias para coordenar a atividade da musculatura da laringe, torácica e abdominal.

Ainda segundo [54], o refinamento adicional da atividade motora é fornecida pelo sistema nervoso extrapiramidal (córtex cerebral, cerebelo e gânglio basal) e autonômico. Esses impulsos combinam-se para produzir um som transmitido não apenas aos ouvidos dos ouvintes, mas também aos ouvidos da própria pessoa que está falando.

O esquema do aparelho vocal pode ser visto na Figura 3.

Figura 3: Esquema do aparelho vocal.

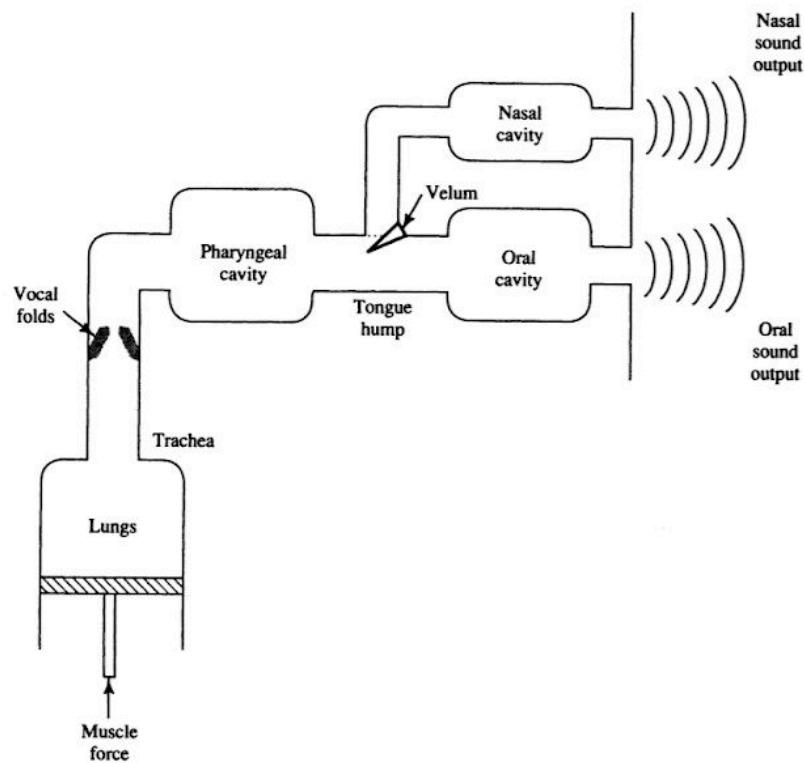


Fonte [27].

Segundo [16], a forma de onda da fala é uma onda acústica que se origina de movimentos voluntários de estruturas anatômicas que compõem o sistema de produção da fala humana, com isso é possível pensar na produção da fala também em termos de operação de filtragem acústica.

As três cavidades principais do sistema de produção da fala, que são o trato vocal e os tratos nasais, compõem o principal filtro acústico [16]. O filtro entra em funcionamento quando é ativado pelos órgãos abaixo dele. Os articuladores, a maioria dos quais são associados ao próprio filtro, são utilizados para alterar as propriedades do sistema [16]. Um modelo acústico simplificado é mostrado na Figura 4.

Figura 4: Diagrama de blocos da produção de voz.



Fonte: [67].

2.2 PATOLOGIAS VOCAIS

Neste tópico será explicado brevemente as três categorias de patologias que estão presentes no banco de dados VOICED e que serão abordadas neste trabalho, que são a disфония

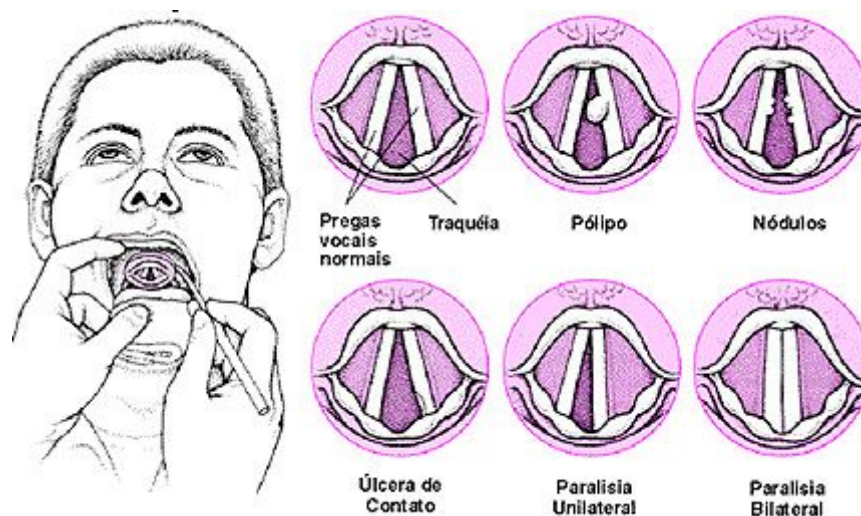
hipercinética, disфонia hipocinética e laringite causada por refluxo e também algumas das principais doenças das categorias.

Segundo [8], em fonoaudiologia, uma voz saudável é caracterizada por ausência de ruído, frequência adequada à idade e sexo do indivíduo, com intensidade apropriada e com variação de frequência e de intensidade de modo que expresse as emoções e sentimentos do falante.

As patologias vocais podem decorrer de interação entre fatores hereditários, comportamentais, estilo de vida e ocupacionais [58]. De modo geral, os distúrbios vocais são caracterizados em dois grupos, funcionais, quando ocorrem devido ao mau uso da voz, ou organofuncionais, quando ocorrem mudanças nas cordas vocais [33].

Alguns exemplos de patologias vocais são: nódulos, pólipos, úlcera de contato e paralisia unilateral e bilateral [51]. Essas patologias podem ser vistas na Figura 5.

Figura 5: Exemplos de patologias vocais.



Fonte [51].

A alteração de algumas das qualidades acústicas da voz, como intensidade, tom, timbre e duração é chamada de disфонia. Como dito anteriormente, ela pode ser causada por um transtorno orgânico ou pelo uso inadequado da voz, podendo ser permanente ou transitório [13].

Nas próximas subseções será feita uma breve descrição das três categorias presentes no banco de dados [11].

2.2.1 DISFONIA HIPERCINÉTICA

A primeira categoria presente no banco de dados é a disfonia hipercinética. A disfonia hipercinética é comum, aparece particularmente em pessoas que utilizam muito a voz e é caracterizada por hipercontração muscular do aparelho pneumo-fônico. Dentro dessa categoria, segundo [11], há várias patologias. Sendo elas prolapse, pólipos nódulos nas cordas vocais, edema de Reinke e Cordite [11]. As quatro primeiras estão presentes em maior quantidade no banco de dados e serão explicadas brevemente abaixo.

A prolapse é a patologia na categoria de disfonia hipercinética com a maior quantidade de vozes dentro do banco de dados. Sendo uma lesão que projeta-se entre as cordas vocais e pode ser diagnosticadas com precisão apenas por biópsia. Essa patologia causa uma lesão primária não relacionada a outras doenças laringeas ou sistêmicas [33].

Após a prolapse, a patologia presente em maior quantidade no banco de dados são os pólipos vocais [11]. Os pólipos são considerados tumores benignos, que muitas vezes desenvolvem-se devido ao esforço excessivo da voz e normalmente para o tratamento necessita-se de cirurgia. O diagnóstico definitivo é importante para a orientação terapêutica correta, sendo muitas vezes utilizado para a determinação a videoestroboscopia [15].

Com 5 indivíduos com nódulos vocais, eles também estão presentes no banco de dados [11]. Os nódulos são pequenas lesões benignas que aumentam devido ao atrito de uma corda vocal [15]. São mais frequentes em pessoas que utilizam muito a voz e que necessitam falar muito alto. O tratamento pode ser feito através de terapia para corrigir os hábitos vocais e a maioria dos indivíduos com nódulos vocais não necessitam de cirurgia para o tratamento [15].

Ainda dentro da categoria de disfonia hipercinética estão os edemas vocais. O edema vocal é a acumulação de líquido ou um gel nas cordas vocais, que em pequeno volume altera a qualidade da voz mas que em alguns casos pode até mesmo dificultar a respiração. Ele está associado na maioria dos casos ao refluxo ácido do estômago, abuso vocal e tabagismo. Para

identificá-lo e corrigi-lo é necessário identificar e corrigir os fatores desencadeantes para assim realizar o tratamento com terapia ou intervenção cirúrgica [15].

2.2.2 DISFONIA HIPOCINÉTICA

Diferentemente da disfonia hipercinética, a disfonia hipocinética é caracterizada por uma redução da adução das cordas vocais durante o ciclo respiratório, em especial durante a fase inspiratória, que produz uma obstrução ao fluxo de ar ao nível da laringe [11]. O fechamento incompleto das cordas vocais leva a uma voz fraca e sem fôlego. No caso da disfonia hipocinética, a voz melhora com o aumento da intensidade vocal e esse fenômeno pode induzir ao abuso vocal e a utilização da voz de maneira incorreta [11].

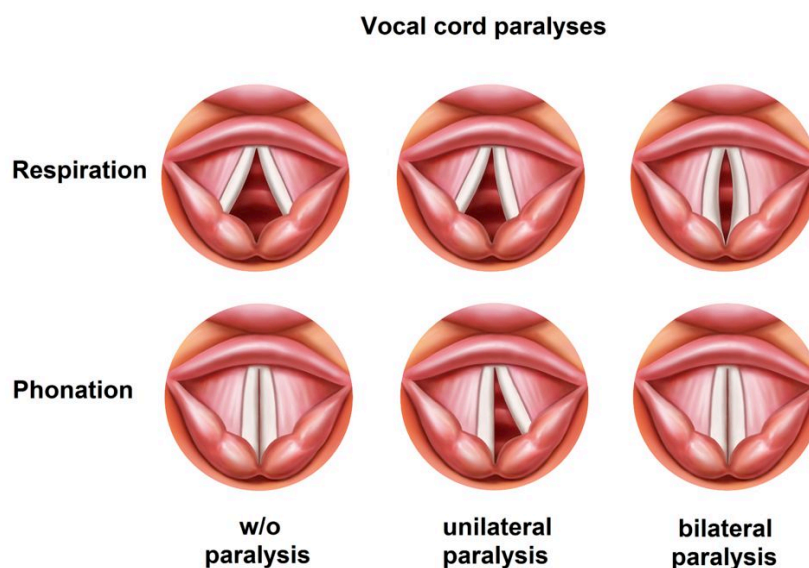
Alguns distúrbios da voz que pertencem à categoria da disfonia hipocinética segundo [11] são a insuficiência glótica, disfonia do sulco cordal, paralisia de corda vocal, déficit de adução, presbifonia, disfonia de conversão, laringite e escape aéreo extraglótico, estando todas presentes no banco de dados, porém as três primeiras presentes em maior quantidade.

A insuficiência glótica é caracterizada pelo fechamento incompleto das cordas vocais ao falar, o que causa vazamento inadequado de ar através da glote, aumentando o risco de aspiração. Com isso, pode ocorrer uma ampla gama de sintomas da laringe, incluindo disfonia, disfagia e infecções recorrentes no peito [42].

A disfonia do sulco cordal é uma lesão em forma de fenda ou depressão longitudinal na corda vocal que dispõe-se paralelamente à sua borda livre. Essa lesão pode estender-se ao longo de todo o comprimento da parte membranosa da corda vocal ou estar limitada a uma parte, variando em extensão e profundidade [14].

Na paralisia das cordas vocais uma ou ambas as cordas vocais deixam de fazer o movimento normal de aproximação, para a fala, ou de afastamento, durante a respiração. Segundo [15], em grande parte das vezes a paralisia é apenas de uma corda vocal, condicionando essencialmente a qualidade da voz. Entretanto, nos casos em que a paralisia é bilateral a respiração também é comprometida e com isso muitas vezes é necessário que o tratamento seja feito urgentemente [15]. Na Figura 6 é possível ver a aparência das cordas vocais na fala e na respiração nos casos de cordas vocais saudáveis e cordas vocais com paralisia unilateral e com paralisia bilateral.

Figura 6: Cordas vocais na respiração e na fala sem paralisia, com paralisia unilateral e com paralisia bilateral.



Fonte:[65].

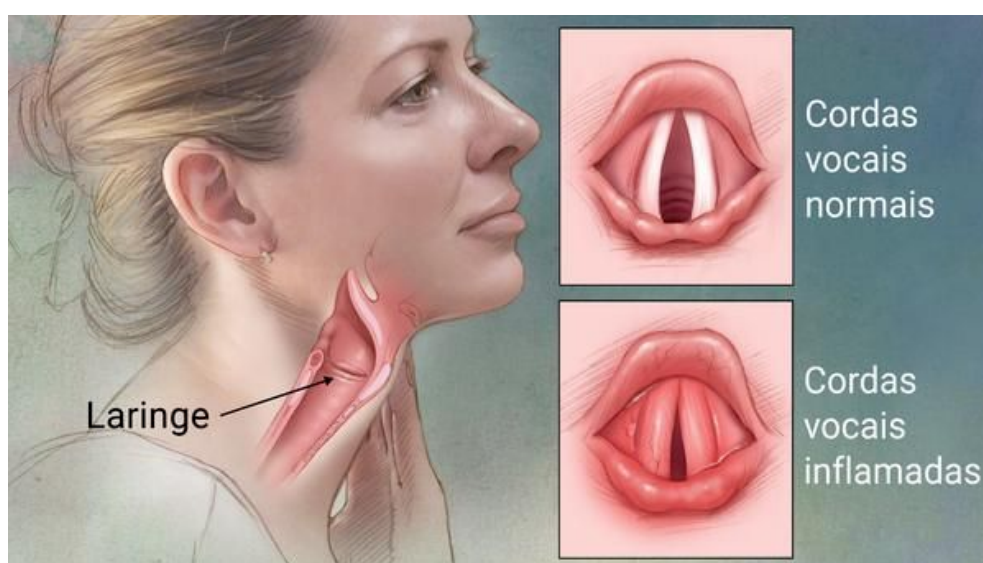
2.2.3 LARINGITE CAUSADA POR REFLUXO

A última categoria abordada é a laringite causada por refluxo. No banco de dados essa é a única patologia presente dentro da categoria [11]. Essa patologia é uma inflamação da laringe causada pelo ácido gástrico que se acumula no estômago e sobe para o esôfago. O sintoma mais comum é uma rouquidão crônica, mas outros sintomas podem também ser pronunciados, como faringite, tontura e, ou tosse noturna, asma, espasmos laríngeos noturnos e halitose [11].

Segundo [2], a laringite causada por refluxo ou o refluxo gastroesofagolaríngeo é um tema bastante controverso e seu diagnóstico é baseado muitas vezes em critérios subjetivos e que mesmo através da videolaringoscopia é de difícil identificação. Também é dito em [2] que na suspeita de laringite causada por refluxo, normalmente o paciente deve ser encaminhado pelo otorrinolaringologista para o gastroenterologista para realização de endoscopia digestiva alta para avaliação do esôfago e de possíveis complicações e somente assim comprovando o refluxo.

Para o tratamento dessa patologia normalmente são utilizados medicamentos que visam reduzir a secreção ácida do estômago e outros remédios que visam diminuir o tempo de contato do esôfago com o agente refluído, chamados de procinéticos. As medidas não medicamentosas, como reduzir a quantidade de ingestão de alimentos nas refeições, respeitar um intervalo de duas horas entre uma alimentação e o horário de deitar e evitar certos alimentos, também são indicadas.[2] Na figura 7 é possível ver um exemplo de laringite causada por refluxo.

Figura 7: Cordas vocais normais e cordas vocais inflamadas devido a laringite.



Fonte:[2].

2.3 MATERIAIS E MÉTODOS DE DETECÇÃO DE PATOLOGIA

Nesta subseção serão explicados alguns dos métodos e materiais utilizados para a detecção de patologias vocais. Eles são em geral métodos invasivos e incômodos para o paciente.

Atualmente, para a detecção das patologias o médico fará perguntas detalhadas sobre os problemas vocais e examinará o indivíduo. Para isso, um anestésico pode ser aplicado para anestesiar os tecidos antes que o médico use uma das seguintes ferramentas para examinar as cordas vocais.

Os principais materiais utilizados para este exame são:

- Espelho: semelhante a um espelho dental que é inserido pela boca para examinar o paciente. É possível ver um exemplo de espelho na Figura 8.

Figura 8: Espelho para laringe.



Fonte:[21].

- Laringoscópio flexível: tubo flexível contendo luz e câmera e que é inserido pelo nariz. Ilustrado na Figura 9.

Figura 9: Laringoscópio flexível.



Fonte : [43].

- Laringoscópio rígido: tubo rígido de visualização que é inserido pela boca. Ilustrado na Figura 10.

Figura 10: Laringoscópio rígido.



Fonte : [66].

- Videostroboscópio:, que é uma câmera combinada com uma luz piscante para fornecer uma visão em câmera lenta das cordas vocais do paciente à medida que elas se movem. Ilustrado na Figura 11.

Figura 11: Videostroboscópio.



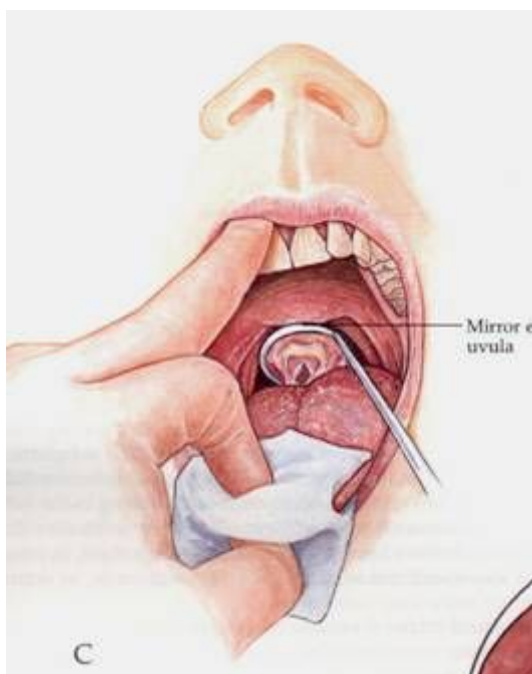
Fonte : [59].

Os métodos utilizados para a detecção de patologias vocais utilizam os materiais explicados acima. Explicando brevemente os métodos, a forma mais simples de analisar a laringe é através de um pequeno espelho que o médico insere até o final da boca do paciente e inclina em direção à laringe. Esse exame é chamado de laringoscopia indireta [17].

A laringoscopia indireta é feita de maneira rápida e oferece uma boa visualização tridimensional da laringe na cor verdadeira. No entanto, algumas pessoas têm um forte ânsia de vômito e não conseguem realizar o exame do espelho, por ele ficar muito próximo a

garganta. Outro ponto negativo do exame é que não pode ser realizado durante a fala normal do paciente, sendo que o espelho deve permanecer no fundo da garganta [17]. Um exemplo de como é realizado o exame de laringoscopia indireta pode ser visto na Figura 12.

Figura 12: Exame de laringoscopia indireta.

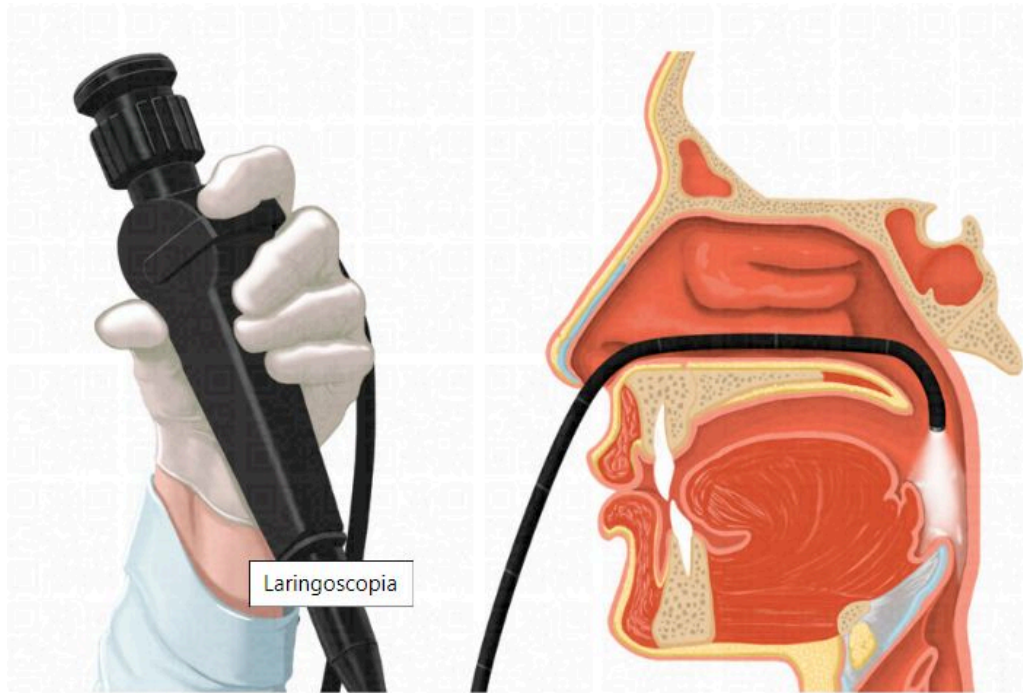


Fonte: [3].

Outro método bastante utilizado é a laringoscopia. A laringoscopia é um exame da porção mais alta das vias aéreas, que é realizado por meio de um aparelho endoscópico chamado laringoscópio.

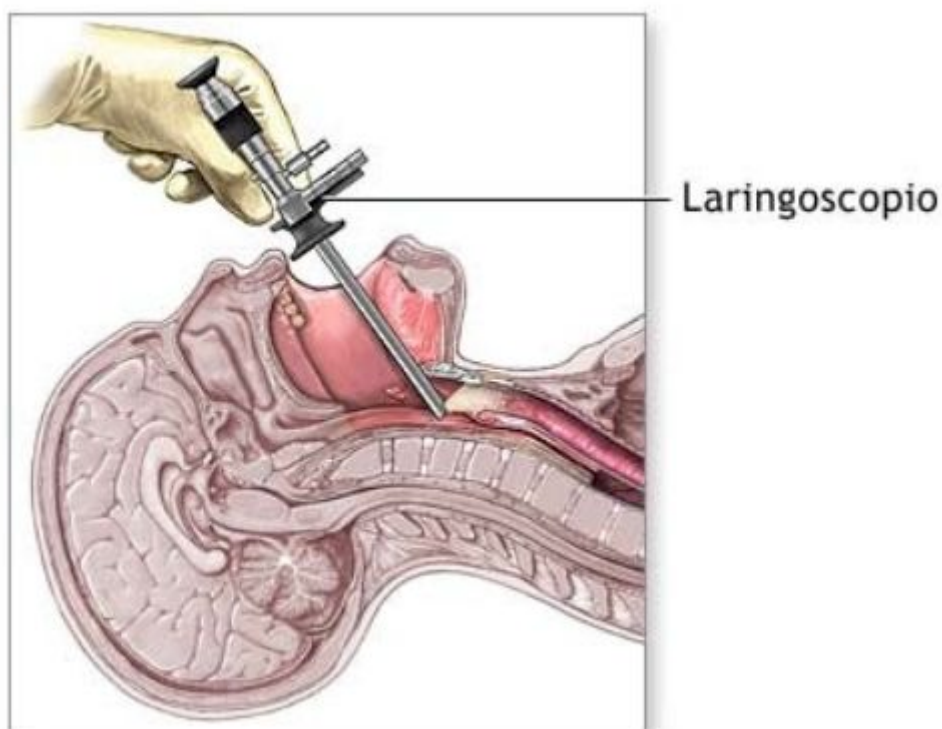
Como dito anteriormente, há dois tipos de aparelhos para laringoscopia direta: um aparelho rígido, que normalmente é introduzido pela boca, e um aparelho que consiste de um fino tubo flexível de fibras óticas, que é introduzido através do nariz (nasolaringoscopia), ambos portando em sua extremidade uma minicâmera que permite ao médico visualizar, por via direta ou através de um monitor de vídeo, o interior das vias aéreas superiores e gravar as imagens correspondentes, se desejar [1]. Segundo [1], no caso de visualização em vídeo, o exame recebe o nome de videolaringoscopia ou videonasolaringoscopia. Dois exemplos de exame de videolaringoscopia podem ser vistos nas Figuras 13 e 14, sendo na Figura 13 um exemplo de nasolaringoscopia.

Figura 13: Nasolaringoscopia.



Fonte [1].

Figura 14: Exame realizado com Laringoscópio Rígido.



Fonte:[49].

Segundo [46], o principal objetivo da videolaringoscopia é a observação da cavidade oral, a orofaringe, a hipofaringe e a laringe. O exame é feito para buscar lesões ou quaisquer outros sinais que indicam a existência de qualquer doença presente. É também usado para ver infecção ou qualquer problema na base da língua, as cordas vocais, e áreas próximas.

Ainda segundo [46], as imagens são exibidas em um monitor no momento do exame e gravada para referência futura. Quando exibido as imagens são ampliadas permitindo exame detalhado da laringe.

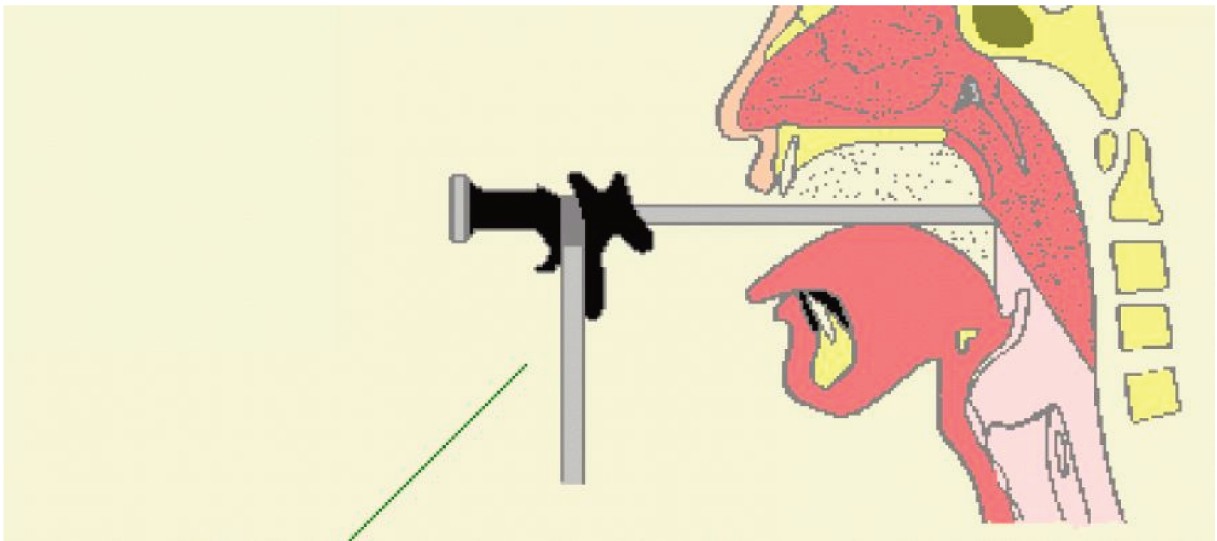
Um problema da laringoscopia direta é que como na indireta, uma parcela dos pacientes pode apresentar ânsia de vômito, sendo assim necessária a sedação deste para a realização do exame.

Porém, durante a fala, as cordas vocais vibram 100 vezes por segundo ou mais [17]. e isso é rápido demais para ser visto pelos olhos humanos e não conseguindo ser observado nos métodos anteriores, portanto outro exame muito utilizado é o exame de videolaringostroboscopia, que é utilizado para obter uma visão detalhada de toda a estrutura

morfológica das cordas vocais vocais. Como o videoestroboscópio apresenta uma luz estroboscópica, é facilitada assim a busca de lesões nas partes internas das cordas vocais [46].

Segundo [46], para diagnosticar os pacientes o exame de videolaringoestroboscopia permite obter detalhes mais precisos das cordas vocais, como a visão de fibroses e sulcos. Pode-se ver na Figura 15 um exemplo de como funciona a videoestroboscopia.

Figura 15: Exame de Videoestroboscopia. .



Fonte:[47].

2.4 CEPSTRO

Segundo [16], a fala pode ser modelada com o resultado da convolução entre a resposta ao impulso do modelo do sistema vocal e um sinal representando a excitação da glote. Normalmente tem-se acesso apenas a voz propriamente dita, mas algumas vezes é desejado o conhecimento de um dos componentes para que possa ser examinado, codificado, modelado ou usado em um algoritmo de reconhecimento [16]. Como as partes individuais do sinal da fala não são combinadas linearmente, as técnicas lineares habituais não fornecem ajuda aparente [16].

O cepstro é o resultado da aplicação da transformada inversa de Fourier ao logaritmo da transformada de Fourier de um sinal [16]. A representação cepstral permite que a convolução de dois sinais no domínio do tempo, portanto não-linear no domínio da frequência, seja representada como uma combinação linear no domínio do cepstro [61]. Assim, se o objetivo é avaliar separadamente as componentes da fala, o cepstro pode ser suficiente para fornecer as informações necessárias [61]. Devido a isto, a análise cepstral é bastante adequada para deconvolução e tem sido amplamente utilizada para avaliar separadamente as componentes da fala como na detecção de pitch [16].

Existem alguns tipos de cepstro, sendo eles o cepstro complexo, o cepstro real e o cepstro de potência. O cepstro de potência em particular tem aplicações na análise da fala humana e pode ser definido como [16]

$$c_p(n) = \left| F^{-1} \{ \log [|F \{ x(n) \}|^2] \} \right|^2 \quad (1)$$

onde $F \{ \cdot \}$ e F^{-1} denotam a Transformada de Fourier de tempo discreto (DTFT) e sua inversa, respectivamente. Na literatura, também encontra-se a seguinte definição para o cepstro de potência [61]

$$c_h(n) = |F \{ \log [|F \{ x(n) \}|^2] \}|^2 \quad (2)$$

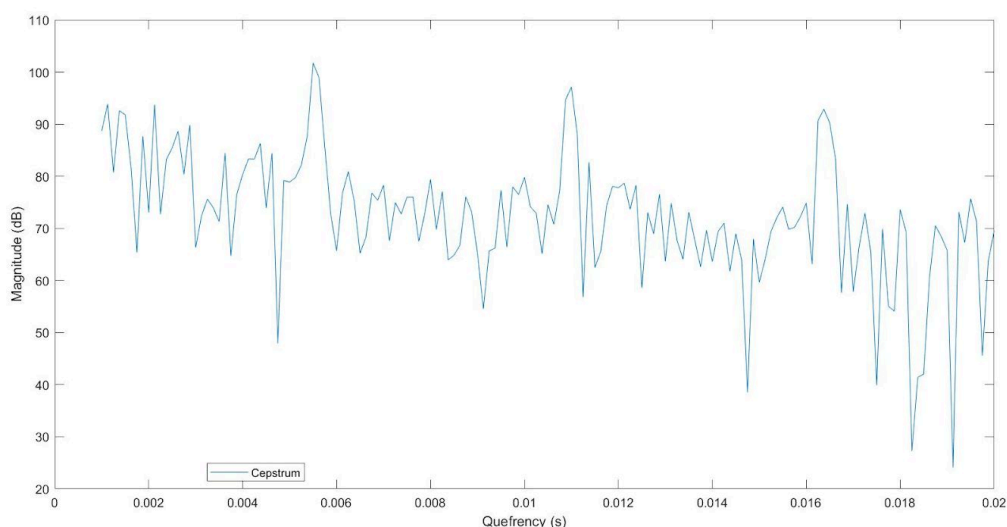
Entretanto, em casos práticos onde uma transformada rápida de Fourier (FFT) de N-pontos é utilizada no cálculo, é sabido que [61]

$$c_h(n) = c_p(n) * N^2 \quad (3)$$

Segundo [61], as medidas calculadas utilizando $c_p(n)$ e $c_h(n)$ serão proporcionais e resultarão na mesma precisão de diagnóstico. Sendo assim, neste trabalho, optou-se pela definição do cepstro de potência apresentada na equação (1).

Segundo [25], o cepstro é um espectro de potência logaritmica. É afirmado também que para sinais periódicos, o primeiro espectro de potência mostrará energia em frequências harmonicamente relacionadas, e o segundo espectro (sendo isso o espectro do espectro) mostrará um componente forte correspondente à regularidade dos picos harmônicos. O tempo (quefrência) do pico cepstral corresponde ao período fundamental do sinal. Um sinal cujo espectro mostra uma estrutura harmônica bem definida mostrará um pico cepstral muito proeminente [25]. Um exemplo de cepstro pode ser visto na Figura 16.

Figura 16: Cepstro.



Fonte: a autora.

O cepstro apresentará um pico proeminente devido à regularidade dos picos harmônicos presentes no espectro. O instante de tempo (queffrência) do pico cepstral corresponderá ao período fundamental do sinal. Um sinal com uma estrutura harmônica mais bem definida apresentará um pico cepstral mais proeminente do que um sinal menos periódico [25]. Essa proeminência do pico cepstral é explorada nas medidas de CPP e CPPS, que serão explicadas na próxima subseção.

2.4.1 PROEMINÊNCIA DO PICO CEPSTRAL

A proeminência do pico cepstral (Cepstral Peak Prominence, CPP) é uma medida acústica que pode ajudar a determinar a qualidade da voz e que tem sido qualificada como uma das medidas acústicas promissoras para a detecção da disfonia [38][69][23].

O CPP é obtido através da medida da amplitude do primeiro pico cepstral, que é localizado entre os valores mínimo e máximo do período fundamental esperado [25] e normalizado em relação a amplitude de todo o cepstro [26]. Com isso é representado quanto o primeiro pico cepstral emerge do cepstro.

Sendo o CPP um sinal quasiperiódico, possui-se então uma estrutura harmônica bem definida e um pico cepstral mais proeminente do que um sinal menos periódico. Porém, não é somente a periodicidade que afeta a amplitude do pico cepstral, mas o tamanho da janela de

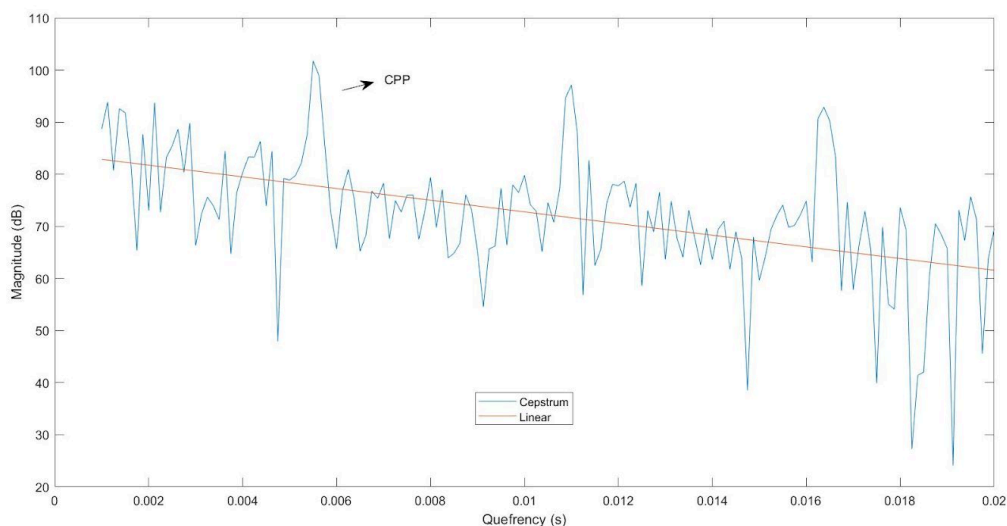
análise cepstral e a energia também afetam [26]. Portanto, é necessário fazer a medida da proeminência do pico cepstral.

A quefrência no pico cepstral geralmente corresponde ao período fundamental, e a amplitude do pico cepstral reflete o nível de organização harmônica e a amplitude geral do sinal. Para normalizar a amplitude geral, uma linha de regressão linear é calculada relacionando a quefrência à magnitude cepstral [25].

Sendo a regressão linear um modelo que tem por objetivo resumir o relacionamento entre duas ou mais variáveis por meio de uma linha, e assim o resultado da função dessa linha pode ser utilizado para estimar valores, quando conhece-se as variáveis que a afetam [19], [57].

O CPP, então, é a diferença em amplitude entre o primeiro pico cepstral e o valor na regressão linear que está diretamente abaixo do pico [26]. O seu cálculo é realizado a cada período T , utilizando uma janela de análise retangular de x ms e uma FFT de N pontos, como feito em [26]. Um exemplo de CPP pode ser visto na Figura 17.

Figura 17: Cepstro e linha de regressão linear.



Fonte: a autora.

2.4.2 PROEMINÊNCIA DE PICO CEPSTRAL SUAVIZADA

Segundo [25], foi descoberto que através de uma simples modificação no algoritmo de CPP produziria uma melhoria notável na previsão. A etapa de processamento adicional envolve suavizar os cepstros individuais antes de extrair o pico cepstral e calcular o pico de destaque.

Em [38] é destacado a importância e relevância do CPPS e afirmado que o CPPS satisfaz os critérios meta-analíticos nas vogais sustentadas e na fala contínua. O CPPS também apresentou um bom resultado para o julgamento perceptivo do grau geral de disфонia e diferentes tipos de qualidade vocal, mostrado em [22], [40]. Além disso, valores significativamente diferentes de CPPS entre o grupo disfônico e o controle foram encontrados na vogal / a / [9].

Apesar de diversas vezes apresentar um bom resultado, há ainda pouca informação e uma falta de investigação sobre a precisão diagnóstica do CPPS na literatura existente, como também afirmado em [10].

O CPPS é uma modificação no CPP que considera duas etapas de suavização antes de calcular o seu valor [11]. Para estimá-lo será seguido o procedimento descrito por

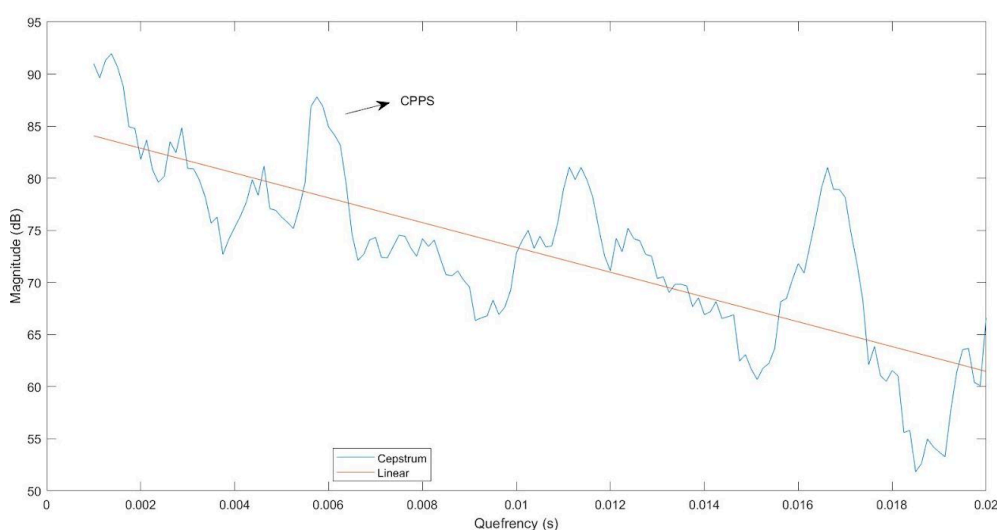
Hillenbrand e Houde [25]. Utilizando o método do CPPS, os cepstros são calculados a cada 2 ms ao invés de a cada 10 ms.

Primeiramente, os cepstros são calculados através da média ao longo do tempo. Cada cepstro não suavizado é substituído pela média de um número de quadros cepstrais à esquerda do quadro atual e algum número de quadros cepstrais à direita do quadro atual, sendo assim, cada k -ésima janela será substituída pela média da $\frac{(k-(L-1))}{2}$ à $\frac{(k+(L-1))}{2}$ -ésima janela. Por exemplo, para uma janela de suavização de sete *frames*, a saída suavizada para uma determinada *frame* consistiria na média da *frame* atual com os três *frames* anteriores e os três *frames* subsequentes [11].

Após esse procedimento, as médias são realizadas ao longo da quefrência. Assim a l -ésima amostra do cepstro da k -ésima janela é substituída pela média de $\frac{(l-(M-1))}{2}$ à $\frac{(l+(M-1))}{2}$ -ésima amostra da k -ésima janela. Por exemplo, para uma janela de média de 7 compartimentos, cada magnitude cepstral é substituída pela média do compartimento atual pelos três compartimentos adjacentes de menor quefrência e os três compartimentos adjacentes de maior quefrência. [11]

Na Figura 18 é possível ver o cepstro suavizado, com $M=L=7$ e a linha de regressão linear, sendo apresentado o CPPS.

Figura 18: Cepstro suavizado e linha de regressão linear.



Fonte: a autora.

2.5 TESTE ESTATÍSTICO DE KRUSKAL-WALLIS

O teste ANOVA é utilizado para testar a igualdade em mais de duas médias populacionais [41]. Quando as médias de k populações são comparadas e sabe-se que as populações não apresentam variações, o teste não paramétrico de Kruskal-Wallis é utilizado.

O teste de Kruskal-Wallis é o equivalente não paramétrico de uma ANOVA de mão única e é utilizado para testar se as amostras são originárias da mesma distribuição [41]. Sendo uma extensão do teste de duas amostras de Wilcoxon-Mann-Whitney para mais de duas amostras independentes. Segundo [45], o teste de Kruskal-Wallis não faz suposições sobre a normalidade. No entanto, parte do pressuposto de que as observações em cada grupo provêm de populações com a mesma forma de distribuição e que as amostras são aleatórias e independentes. A estatística de teste para análise de variância unidirecional é calculada como a razão entre a soma dos quadrados do tratamento e a soma residual dos quadrados. O teste de Kruskal-Wallis usa o mesmo método, mas, como em muitos testes não paramétricos, as fileiras dos dados são utilizados no lugar dos dados brutos [45].

Com isso, valores-p são gerados e determinam se estatísticas de grupos diferentes são oriundos da mesma distribuição [61]. Os valores-p precisam ser menores que 0,05 para que uma hipótese nula seja rejeitada e para significar que os grupos originam de distribuições diferentes [45].

2.6 CLASSIFICADOR BINÁRIO

Classificadores são separadores de grupos que mediante determinadas características organizam os dados agrupando elementos que apresentam traços semelhantes, o que permite reconhecimento de padrões e a identificação de elementos que não se encaixam dentro do padrão. No modelo binário, assume-se que cada documento é representado por um vetor de atributos binários de modo que cada atributo indica a ocorrência ou não de um evento no documento [5].

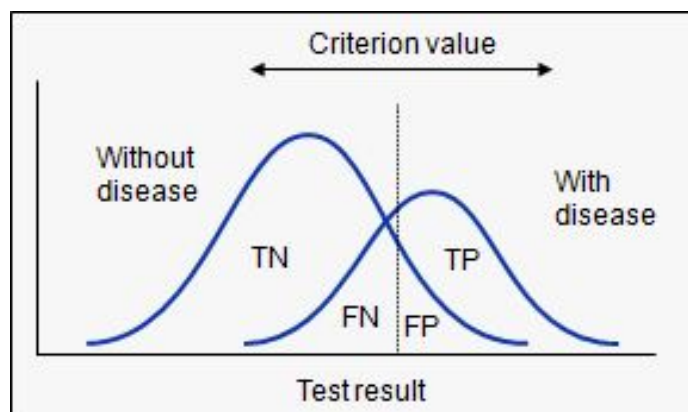
Segundo [5], a saída real de vários algoritmos de classificação binária é uma pontuação de previsão. A pontuação indica a certeza do sistema de que determinada

observação pertence à classe positiva. Para decidir se a observação deve ser classificada como positiva ou negativa um limite deverá ser definido e a pontuação deve ser comparada a ele. Com isso, todas as observações com pontuações maiores que o limite serão previstas como classe positiva, e as pontuações menores que o limite serão previstas como classe negativa [5].

No classificador binário, as previsões podem ser classificadas em quatro grupos com base na resposta conhecida e na resposta prevista, os grupos são: previsões de positivos corretas (verdadeiros positivos, em inglês True Positive, TP), previsões de negativos corretas (verdadeiros negativos, em inglês True Negative, TN), previsões de positivos incorretas (falsos positivos, em inglês False Positive, FP) e previsões de negativos incorretas (falsos negativos, em inglês False Negative, FN) [5].

A análise ROC (Receiver Operating Characteristic) é um método gráfico para avaliação, organização e seleção de sistemas de diagnóstico e/ou predição [48]. Segundo [71] ela é muito utilizada na medicina, para analisar a qualidade de determinados testes clínicos. A ROC é uma curva 2-D, que pode ser desenhada para visualizar o desempenho do classificador binário, sendo que o eixo x representa a taxa de falso positivo e o eixo y representa a taxa de verdadeiro positivo [48]. Quando os resultados de um teste entre duas populações é obtido, uma com indivíduos saudáveis e outra com indivíduos com patologias, raramente observará uma separação perfeita entre os dois grupos [39]. Portanto, a distribuição dos resultados do teste se sobrepõe, como mostra a Figura 19.

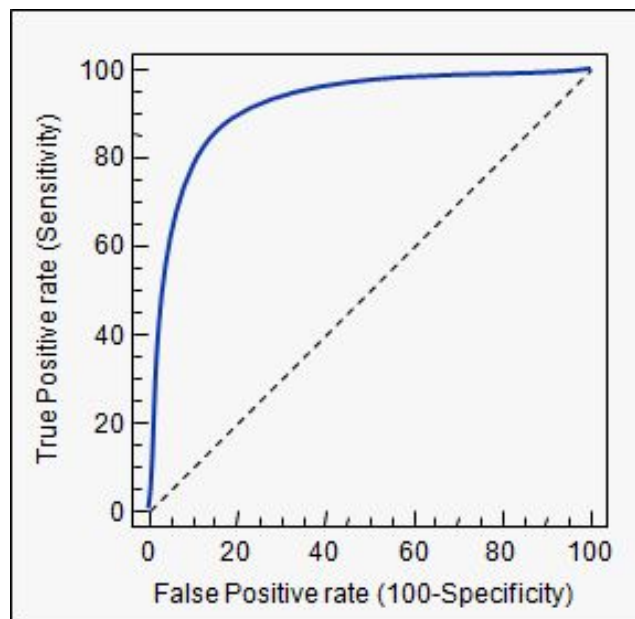
Figura 19: Exemplo de resultado de duas distribuições.



Fonte: [39].

A área abaixo da curva ROC também é comumente utilizada como um índice de desempenho geral da precisão do diagnóstico de um marcador e indica a probabilidade de que duas medições sejam ordenadas corretamente [45]. O seu valor sempre estará entre 0 e 1, pois a adivinhação aleatória produz a linha diagonal entre (0, 0) e (1, 1), que tem uma área de 0,5, e assim nenhum classificador realista deve ter uma área ROC menor que 0,5 [18]. Um exemplo de curva ROC pode ser visto na Figura 20.

Figura 20: Exemplo de Curva ROC e área abaixo da curva.



Fonte: [39].

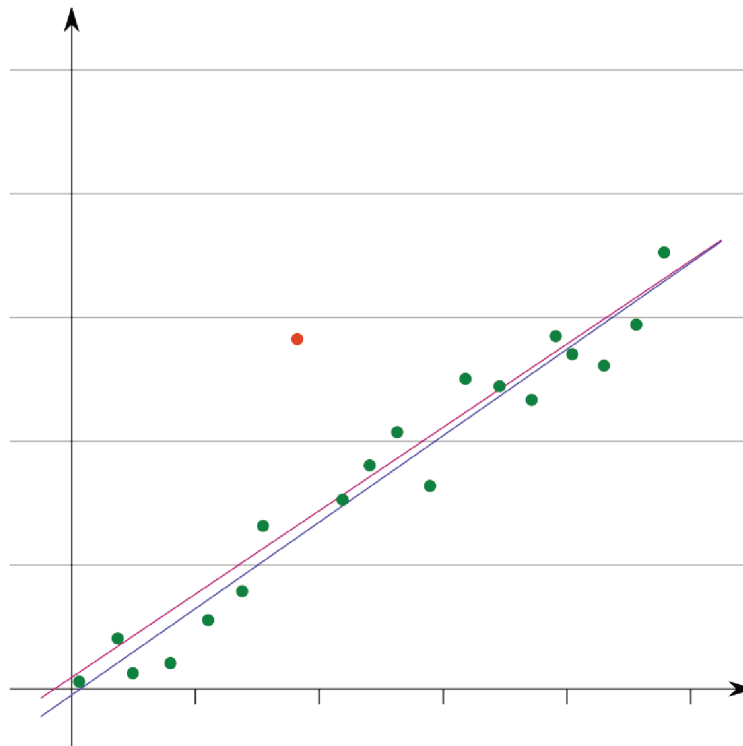
Com isso, o classificador que será utilizado neste trabalho, será o classificador OB-ROC, que é um classificador binário que utiliza a distância quadrática entre médias de distribuições para encontrar uma boa classificação para discriminar classes [45]. Ele utiliza a ROC (Receiver Operating Characteristic) para analisar o resultado da classificação [45].

2.7 OUTLIER

Segundo [28], um outlier é um valor atípico que se encontra a uma distância anormal de outros valores em uma amostra aleatória de uma população. Os outliers podem prejudicar a interpretação dos resultados dos testes estatísticos aplicados às amostras [67].

Um exemplo de outlier é apresentado na Figura 21, sendo ele o ponto vermelho que apresenta um valor distante dos outros.

Figura 21: Exemplo de outlier.



Fonte: [68].

Entre os métodos de detecção de outliers, os três mais comumente utilizados são: método da média e desvio padrão; método de desvio absoluto mediano; e método de desvio mediano e Interquartil [44]. Neste trabalho, optou-se por utilizar o método da média e desvio padrão e para este método de detecção, a média e o desvio padrão dos resíduos são calculados e comparados [44]. Portanto, se um valor estiver a um certo número de desvios-padrão da média, esse ponto de dados é identificado como um erro externo [44].

O número especificado de desvios padrão é chamado de limite. O valor padrão é 3 porque, numa distribuição normal, 99,73% dos valores encontram-se dentro de uma banda em torno da média de seis desvios padrões [63]. E, no caso de distribuições não-normais, pelo

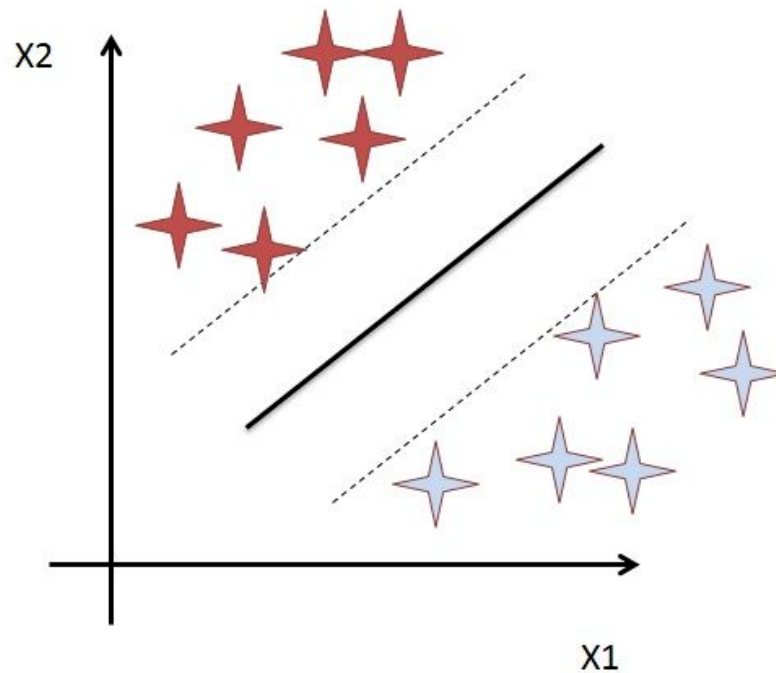
menos 88% dos valores são encontrados dentro de uma banda em torno da média no mesmo caso.

2.8 MÁQUINA DE VETORES DE SUPORTE

Uma máquina de vetores de suporte (SVM, do inglês: support vector machine) é um método de aprendizagem de máquina que tenta tomar dados de entrada e classificá-los em uma entre duas categorias. Para que uma máquina de vetores de suporte seja eficaz, primeiramente é necessário utilizar um conjunto de dados de entrada e de saída de treinamento para construir o modelo de máquina de vetores de suporte que pode ser utilizado para classificação de novos dados [30].

Esse modelo é desenvolvido tomando as entradas de treinamento, mapeando-as no espaço multidimensional e utilizando regressão para encontrar um hiperplano que melhor separe duas classes de entradas, sendo um hiperplano uma superfície em espaço de n dimensões que o separa em duas metades de um espaço [30]. Uma vez que a máquina de vetores de suporte tenha sido treinada, ela é capaz de avaliar novas entradas em relação ao hiperplano divisor e classificá-las em uma entre duas categorias [30]. A Figura 22 apresenta um exemplo de separação de classes.

Figura 22: Exemplo de separação entre classes.



Fonte: [64].

O número de entradas para qualquer máquina de vetores de suporte especificada varia de um a infinito, teoricamente. Entretanto, em termos práticos, a capacidade computacional limita a quantidade de entradas que pode ser utilizada [30].

Assim, dado um conjunto de pares, a máquina de vetores de suporte, para otimização, a SVM consiste em mapear os vetores de instâncias a um espaço de dimensão mais elevada e encontrar, nesse espaço, um hiperplano de separação com margem máxima. Para isso, associa-se um parâmetro, que atribui um fator de penalização ao erro, e uma função kernel, que pode assumir, tipicamente, quatro formas básicas: linear, polinomial, base radial (RBF) e sigmoide [30]. Neste trabalho, optou-se por utilizar a função kernel do tipo RBF.

3 MATERIAIS E METODOLOGIA

Nesta seção serão explicados os principais materiais e metodologias que foram utilizados e aplicados no decorrer deste trabalho.

3.1 MATERIAIS

Primeiramente, para desenvolvimento do código deste trabalho foi utilizado um computador com Matlab na versão 2019b. O Matlab é um software interativo de alta performance voltado para o cálculo numérico e análise numérica, cálculo com matrizes, processamento de sinais e construção de gráficos.

Como material de voz, foi usado o banco de dados VOICEd. O VOICEd é um banco de dados livre, disponível em [11], que contém 208 sinais de voz, sendo eles 150 com patologia e 58 saudáveis. O banco de dados foi disponibilizado em 2018 e, por isso, ainda não foi muito explorado.

Os sinais consistem em gravações de uma vocalização da vogal /a/ com cinco segundos de duração sem qualquer interrupção. De acordo com [11] todas as amostras foram registradas em ambiente silencioso, com menos de 30 dB de ruído de fundo. As gravações de voz foram feitas usando um sistema m-health apropriado, o Vox4Health, capaz de adquirir em tempo real o sinal de voz usando o microfone de um dispositivo móvel. Este sistema foi instalado em um Samsung Galaxy S4, versão Android 5.0.1 e foram realizadas a uma distância de cerca de 20 cm do paciente em um ângulo de cerca de 45 graus [11].

Todas as gravações foram amostradas a 8 kHz e com uma resolução de 32 bits. Ainda de acordo com [11], cada gravação foi preenchida com um filtro apropriado para remover qualquer ruído acidentalmente adicionado durante a aquisição. No entanto, ao ouvir os sinais gravados, percebeu-se que em 22 das 208 gravações ainda havia ruídos, em alguns casos era possível escutar a fala de outras pessoas enquanto ocorria a gravação.

O banco de dados contém gravações de vozes afetadas por diferentes tipos de patologia, as quais estão divididas em três categorias: disfonia hipercinética, disfonia hipocinética e laringite causada por refluxo. Dentro das categorias de disfonia hipercinética e disfonia hipocinética estão presentes várias outras patologias [11], elas citadas na seção 2.

Segundo [11], as vozes saudáveis e patológicas foram clinicamente verificadas antes de serem incluídas no banco de dados. Além disso, o banco de dados inclui informações como sexo, idade, patologia, hábitos de vida (como tabagismo, consumo de álcool e café) e status ocupacional [11]. Entretanto, duas vozes não foram especificadas corretamente, uma delas estava com a especificação e o número da voz anterior e a outra estava com a especificação da voz posterior. O que fez a autora questionar a credibilidade das especificações. Além disso, grande parte das especificações das vozes foi feita somente por categoria, não constando a patologia exata do indivíduo e impossibilitando a separação das vozes por doenças.

Os voluntários envolvidos no estudo foram adultos com idade entre 18 e 70 anos [11]. A idade média é de cerca de 40 anos, tanto para mulheres quanto para homens, porque segundo [11], pessoas com idade entre 40 e 60 anos representam a categoria de indivíduos que mais sofrem com distúrbios da voz.

Cada item do banco de dados é composto por um arquivo .wav , dois arquivos .txt e arquivos .hea e .dat. O arquivo .wav contém o sinal de áudio, os arquivos .txt incluem as amostras de cada sinal e sua especificação e o arquivo .hea contém o número de identificação da amostra, informações sobre o formato de armazenamento e frequência de amostragem, além da idade, sexo e diagnóstico do sujeito. E por fim, o arquivo .dat inclui as amostras digitalizadas do sinal [11].

Após a escolha do banco de dados, todas as 208 vozes foram escutadas, o nível de potência ativa foi normalizado de acordo com o algoritmo de Recomendação ITU T P.56 [29] em -26dBov e 22 vozes foram descartadas por conter ruído atrás e 2 por não ter as especificações de patologia, como dito anteriormente.

Com isso, o total de pacientes com vozes patológicas e vozes saudáveis analisados no trabalho foi de 184, sendo 51 saudáveis e 133 patológicos, como pode ser visto na Tabela 1.

Tabela 1: Quantidade de vozes saudáveis, patológicas e número total de vozes

Vozes	Quantidade de indivíduos
Vozes saudáveis	51
Vozes patológicas	133
Total	184

Fonte: autora.

Como já dito anteriormente, dentro dos grupos de disfonia hipercinética e hipocinética estão presentes diversas outras patologias, mas por a maioria das vezes não estarem especificadas no banco de dados não foi possível identificá-las. Com isso as vozes foram divididas somente em categorias e o número de vozes de cada grupo pode ser visto na Tabela 2.

Tabela 2: Quantidade de pacientes com disfonia hipercinética, hipocinética e laringite causada por refluxo.

Categorias	Quantidade de indivíduos
Disfonia hipercinética	67
Disfonia hipocinética	34
Laringite causada por refluxo	32
Total	133

Fonte: a autora.

3.2 METODOLOGIA

Após ter sido feita a pesquisa para a escolha do banco de dados, optou-se pelo VOICEd, por ser um banco de dados publicado recentemente, por não existir trabalhos avaliando as vozes desse banco de dados utilizando CPP e CPPS e por ser aberto.

As vozes com ruído e sem especificação foram descartadas e todas as outras normalizadas em -26dBov. Com isso foi inicializado o desenvolvimento do código de CPP e CPPS.

3.2.1 PROEMINÊNCIA DO PICO CEPSTRAL

O CPP, como dito anteriormente, é a medida da amplitude do primeiro pico cepstral, normalizada em relação a amplitude de todo cepstro [25].

Neste trabalho o CPP foi calculado a cada período de 10ms, $T=10\text{ms}$, como proposto em [25] e também a cada período de 2ms, como utilizado em [26] e em [10] para o cálculo do CPPS, sendo calculado assim como forma de comparação. Para o cálculo foi utilizada uma janela de Hanning de 1024 amostras, equivalente a 46,4 ms, e uma FFT de 1024 pontos. O primeiro pico cepstral foi procurado entre 3,3ms e 16,7 ms, como em [61], uma vez que a frequência fundamental em adultos costumam ser de 60Hz a 300Hz. As quefrências abaixo de 1 ms foram desconsideradas porque elas são as mais afetadas pelo envelope cepstral do que a periodicidade [61].

Para cada voz séries temporais de CPP foram obtidas e tratadas como distribuições. Seguindo [10] e [61], as seguintes estatísticas descritivas foram calculadas para as distribuições de CPP: média (CPP_{mean}), mediana ($\text{CPP}_{\text{median}}$), módulo (CPP_{mode}), quinto percentil ($\text{CPP}_{5\text{prc}}$) e o 95-ésimo percentil ($\text{CPP}_{95\text{prc}}$), desvio padrão (CPP_{std}), o intervalo entre o valor máximo e o mínimo ($\text{CPP}_{\text{range}}$), assimetria (CPP_{skew}) e curtose (CPP_{kurt}) para a caracterização do formato de distribuição. Além dessas estatísticas também foi calculada a variância (CPP_{var}), que não havia sido calculado em [10] e em [61].

Após as estatísticas terem sido calculadas foi utilizado o teste de Kruskal-Wallis para determinar se as estatísticas de diferentes grupos de pacientes são oriundos da mesma distribuição. Para isso foram analisadas as especificações das vozes e elas foram separadas em grupos. Primeiramente separando-as entre vozes saudáveis e vozes patológicas. Dentro das vozes patológicas foi feita a divisão de disfonia hipercinética, disfonia hipocinética e laringite causada por refluxo.

Com a separação feita, foi utilizado o teste de Kruskal-Wallis para vozes saudáveis *versus* patológicas, vozes saudáveis *versus* disfonia hipercinética, vozes saudáveis *versus* disfonia hipocinética e vozes saudáveis *versus* laringite causada por refluxo. Como dito em [45] e na seção 2, os valores-p precisam ser menores que 0,05 para que uma hipótese nula seja rejeitada e para significar que os grupos originam de distribuições diferentes. Com isso, foram determinados os valores-p de cada estatística para CPP, com $T=2\text{ms}$ e $T=10\text{ms}$ e para cada um dos casos descritos anteriormente.

Para averiguar a eficiência das estatísticas descritivas foi utilizado um classificador OB-ROC, que como explicado na seção 2 é um classificador binário que utiliza a distância

quadrática entre médias de distribuições para encontrar a melhor classificação para discriminar classes [41].

Após realizada a classificação foi possível obter os valores de precisão e área ROC de cada estatística para CPP, com $T=2\text{ms}$ e $T=10\text{ms}$ e também a área ROC de cada uma delas. Porém, buscando ainda um melhor resultado, foi feita a remoção dos outliers de cada caso, das vozes saudáveis, patológicas, com disfonia hipercinética, com disfonia hipocinética e com laringite causada por refluxo. Com isso algumas vozes com o valor atípico para a categoria foram descartadas.

Para a remoção de outliers, primeiramente foi calculado a médias e o desvios padrão de todos os CPP de cada estatística. Após isso foram descartados os valores que estivessem acima do valor da média mais 3 vezes o desvio padrão e os valores que estivessem abaixo do valor da média menos 3 vezes o valor do desvio padrão para cada caso.

Com os outliers removidos, os valores de precisão de classificação e valor-p foram recalculados com o classificador OB-ROC. Além disso, foi utilizado um SVM com kernel função de base radial (RBF) para obter a melhor precisão com a combinação das estatísticas.

3.2.2 PROEMINÊNCIA DO PICO CEPSTRAL SUAVIZADA

O CPPS foi proposto por [26] como uma modificação do CPP, consistindo de 2 processos de suavização do cepstro.

Com isso, após feito o código do CPP foram realizadas alterações no código para o cálculo do CPPS. Foram adicionadas a suavização no tempo e na quefrência, como explicado anteriormente na seção 2. As medidas foram calculadas a cada período de 2ms, como em [10] e em [61] e na escala de decibel (dB). O cepstro é real e positivo, pois foi calculado o cepstro de potência. Os valores de L e M utilizados foram 7, como em [10] e 11, como em [26].

Para cada voz, como para CPP, séries temporais de CPPS foram obtidas e tratadas como distribuições. Ainda como em [10], as mesmas estatísticas descritivas calculadas para CPP foram calculadas para as distribuições de CPPS, que foram: média ($\text{CPPS}_{\text{mean}}$), mediana ($\text{CPPS}_{\text{median}}$), módulo ($\text{CPPS}_{\text{mode}}$), quinto percentil ($\text{CPPS}_{5\text{pre}}$) e o 95-ésimo percentil ($\text{CPPS}_{95\text{pre}}$), desvio padrão (CPPS_{std}), o intervalo entre o valor máximo e o mínimo ($\text{CPPS}_{\text{range}}$), assimetria ($\text{CPPS}_{\text{skew}}$) e curtose ($\text{CPPS}_{\text{kurt}}$) para a caracterização do formato de distribuição.

Assim como para CPP, também foi calculada a variação ($CPPS_{var}$), que não havia sido calculado em [10] e em [61].

Como na subseção anterior, com as estatísticas calculadas foi então utilizado o teste de Kruskal-Wallis para determinar se as estatísticas dos grupos de pacientes saudáveis e patológicos, saudáveis e com disfonia hipercinética, saudáveis e com disfonia hipocinética e saudáveis e com laringite causada por refluxo são oriundos da mesma distribuição. Os valores-p foram encontrados para cada estatística para CPPS, com $T=2ms$, $L=M=7$ e $L=M=11$.

Com o objetivo de averiguar a eficiência das estatísticas descritivas explicadas anteriormente, foi utilizado um classificador OB-ROC e após realizada a classificação foi possível obter então os valores de precisão e área ROC de cada estatística para CPPS, com $T=2ms$, $L=M=7$ e $L=M=11$ e também a área ROC de cada uma delas.

Além disso, como para o cálculo de CPP, para CPPS foi utilizada uma janela de Hanning de 1024 amostras e uma FFT de 1024 pontos. O primeiro pico cepstral também foi procurado entre 3,3 ms e 16,7 ms e as quefrências abaixo de 1ms foram desconsideradas.

Buscando melhores resultados, foi realizada a remoção dos outliers de cada caso, das vozes saudáveis, patológicas, com disfonia hipercinética, com disfonia hipocinética e com laringite causada por refluxo. Descartando assim, vozes com valor atípico para a categoria e estatística.

Assim como para CPP, para a remoção de outliers, foram calculadas a média e o desvio padrão de todos os CPPS de cada estatística. Após isso foram descartados os valores que estivessem acima do valor da média mais 3 vezes o desvio padrão e os valores que estivessem abaixo do valor da média menos 3 vezes o valor do desvio padrão de cada estatística. Sendo possível recalculando os valores de precisão de classificação e valor-p, como feito para CPP.

Por fim, novamente foi utilizado o SVM com kernel função de base radial (RBF) para gerar a combinação de estatísticas que resulte na maior precisão.

4 RESULTADOS

Nesta seção serão mostrados e explicados os resultados obtidos neste trabalho após a metodologia ter sido aplicada. Ela será dividida em cinco subseções, primeiramente mostrando a quantidade de outliers retirados das vozes saudáveis, patológicas, com disfonia hipercinética, disfonia hipocinética e com laringite causada por refluxo.

Após isso, serão mostrados os resultados da discriminação de CPP e CPPS entre vozes saudáveis e vozes patológicas, vozes saudáveis e vozes com disfonia hipercinética, vozes saudáveis e vozes com disfonia hipocinética e vozes saudáveis e vozes com laringite causada por refluxo. Serão apresentados para cada caso o valor-p do teste estatístico de Kruskal-Wallis, o valor de precisão de classificação obtido pelo classificador OB-ROC e as melhores precisões obtidas pelo SVM.

4.1 REMOÇÃO DE OUTLIERS

Na tentativa de obter melhores resultados, antes do cálculo do valor-p do teste estatístico de Kruskal-Wallis e do cálculo de precisão de classificação das estatísticas foi realizada a remoção dos outliers. Como dito na seção 2, os outliers podem prejudicar a interpretação dos resultados dos testes estatísticos, por isso a sua remoção pode ajudar a obter melhores resultados. A tabela 3 apresenta o número de outliers retirados das vozes saudáveis nos casos de CPP, T=10ms e T=2ms, e nos casos de CPPS, T=2ms e L=M=7 e L=M=11.

Tabela 3: Remoção de outliers Vozes Saudáveis.

Parâmetros CPP e CPPS			Número de Outliers retirados de cada estatística para vozes saudáveis										
Tempo	Suav	Escala	Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda	
CPP	T=10ms	-	dB	1	1	1	0	2	0	2	0	0	0
	T=2ms	-	dB	0	1	0	1	2	0	0	0	0	0
CPPS	T=2ms	L=M=7	dB	0	0	1	0	0	1	0	3	3	0
	T=2ms	L=M=11	dB	0	0	0	0	2	1	0	1	1	0

Fonte: a autora.

Pode-se ver que as maiores quantidades de outliers apresentou-se nas estatísticas: curtose, desvio padrão, e variância. No total, para este caso, foram retirados 24 outliers.

Após a retirada dos outliers das vozes saudáveis, foram removidos os outliers das vozes patológicas, vozes com disfonia hiper e hipocinética e laringite causada por refluxo. Das vozes patológicas o total de outliers retirados foi de 13 e as quantidades para cada estatística podem ser vistas na Tabela 4, sendo 12 deles das estatísticas de desvio padrão, variância e assimetria.

Tabela 4: Remoção de outliers Vozes Patológicas.

Parâmetros CPP e CPPS				Número de Outliers retirados de cada estatística									
	Tempo	Suav.	Escala	Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var .	Moda
CPP	T=10ms	-	dB	0	0	0	0	0	0	2	2	2	0
	T=2ms	-	dB	0	0	0	0	1	0	1	2	2	0
CPPS	T=2ms	L=M=7	dB	0	0	0	0	0	0	1	0	0	0
	T=2ms	L=M=1 1	dB	0	0	0	0	0	0	0	0	0	0

Fonte: a autora.

Para as vozes com disfonia hipercinética, o total de outliers removidos pode ser visto na Tabela 5. Foram retirados 18 outliers e novamente, a maior quantidade de outliers retirados foram nas estatísticas de assimetria, desvio padrão e variância, representando 16 dos 18 outliers.

Fonte: a autora.

A Tabela 7 apresenta o número de outliers removido para cada estatística das vozes com laringite causada por refluxo. Neste caso, foram removidos 21 outliers e novamente estavam mais presentes no CPP.

Tabela 7: Remoção de outliers Vozes com Laringite causada por Refluxo.

Parâmetros CPP e CPPS				Número de Outliers retirados de cada estatística									
Tempo	Suav.	Escala	Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda	
CPP	T= 10ms	-	dB	0	1	0	0	1	0	0	5	5	0
	T=2ms	-	dB	2	2	2	0	0	0	0	1	1	0
CPPS	T= 2ms	L=M=7	dB	0	0	0	0	1	0	0	0	0	0
	T=2ms	L=M=11	dB	0	0	0	0	0	0	0	0	0	0

Fonte: a autora.

4.2 VOZES SAUDÁVEIS E VOZES PATOLÓGICAS

Após os outliers terem sido removidos, foi realizado o teste de Kruskal-Wallis, para a obtenção do valor-p e obtidas as precisões de classificação através do classificados OB-ROC e pelo SVM.

Para a discriminação entre vozes saudáveis e vozes patológicas os resultados obtidos para o valor-p do Teste Estatístico de Kruskal-Wallis e as precisões do classificador binário na discriminação entre vozes saudáveis e patológicas são apresentados na Tabela 8 e 9, respectivamente.

Tabela 8: Valor-p das Estatísticas das Distribuições.

Parâmetros CPP e CPPS				Valor-p das Estatísticas das Distribuições									
Tempo	Suav.	Escala		Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda
CPP	T= 10ms	-	dB	2,42E-04	3,32E-04	4,35E-04	3,90E-03	1,10E-03	3,90E-01	8,99E-01	7,95E-02	7,95E-02	1,58E-02
	T=2ms	-	dB	5,18E-04	3,50E-03	1,55E-04	4,39E-02	3,84E-01	5,34E-01	3,20E-02	1,93E-02	1,93E-02	8,58E-02
CPPS	T= 2ms	L=M=7	dB	4,87E-02	4,53E-02	4,02E-02	1,21E-01	3,28E-01	1,39E-02	8,57E-01	5,87E-01	5,87E-01	2,87E-02
	T=2ms	L=M=11	dB	3,45E-02	2,91E-02	1,17E-02	5,36E-02	7,25E-01	1,42E-02	5,86E-01	6,72E-01	6,72E-01	5,99E-02

Fonte: a autora

Tabela 9: Precisão de Classificação das Estatísticas das Distribuições.

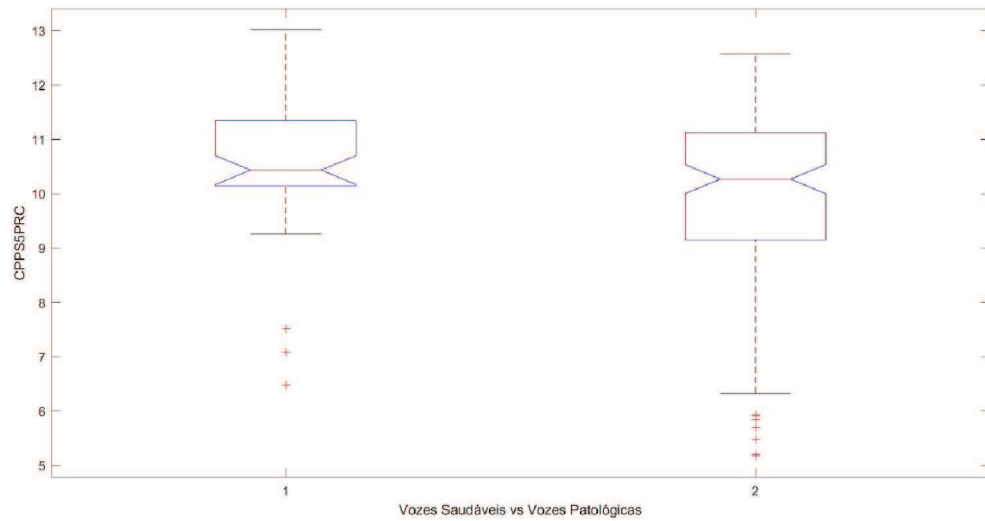
Parâmetros CPP e CPPS				Precisão de Classificação das Estatísticas das Distribuições (%)									
Tempo	Suavização	Escala		Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda
CPP	T= 10ms	-	dB	75,96	76,50	72,28	72,28	74,03	72,28	72,77	72,22	72,22	72,83
	T=2ms	-	dB	72,97	73,22	74,46	72,38	72,93	72,28	73,77	73,07	73,07	72,83
CPPS	T= 2ms	L=M=7	dB	72,28	72,28	74,32	72,28	72,28	72,67	72,67	73,48	73,48	73,37
	T=2ms	L=M=11	dB	73,37	72,83	73,91	73,37	72,28	73,22	75,00	74,32	74,32	74,46

Fonte: a autora

O $CPPS_{5prc}$ calculado com $T=2$ e $L=M=7$, parâmetro utilizado em [10], apresentou um valor-p igual a $4,02E-02$ e uma precisão de classificação igual a $74,32\%$. A área ROC obtida foi $0,5983$, muito inferior ao $0,95$ alcançado em [10]. O boxplot do $CPPS_{5prc}$ para as vozes saudáveis e patológicas é mostrado na Figura 23, onde nota-se que não há uma separação clara entre os dois grupos como sugerido pelo valor-p inferior a $0,05$. Os valores do $CPPS_{5prc}$ para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 24, onde observa-se a tendência do classificador em considerar quase todas as vozes (abaixo do limiar) como patológicas. Vale ressaltar que uma precisão de $72,28\%$ seria obtida ao classificar todas as vozes como patológicas. Considerando o CPPS, a melhor estatística descritiva é o $CPPS_{skew}$ calculado com $L=M=11$, o qual obteve um valor-p igual a $5,86E-01$, uma precisão de classificação igual a $75,00\%$ e uma AROC igual a $0,5171$.

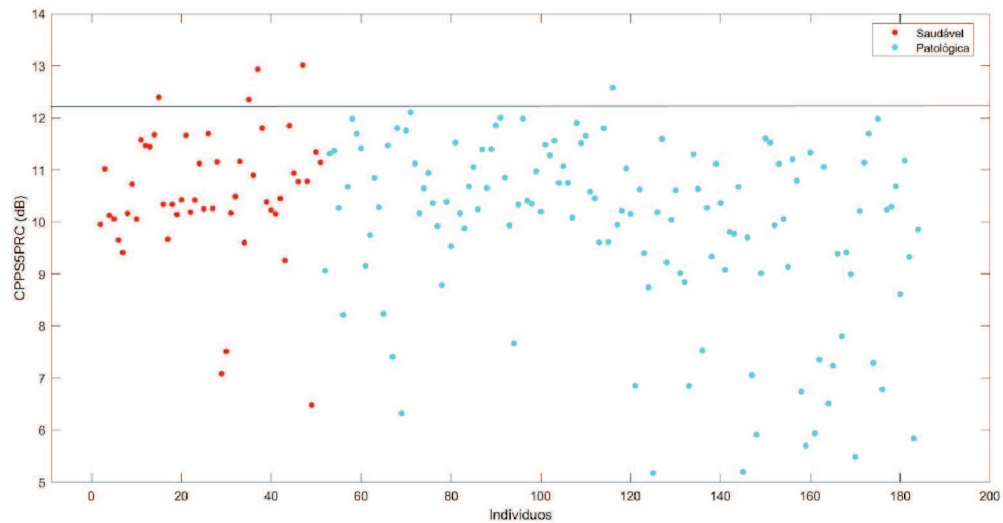
Neste trabalho, como pode ser observado nas Tabelas 8 e 9, a melhor estatística descritiva foi o CPP_{median} calculado com $T=10ms$, o qual obteve um valor-p igual a $3,32E-04$, uma precisão de classificação igual a $76,50\%$ e uma área ROC igual a $0,6723$. O boxplot do CPP_{median} para as vozes saudáveis e patológicas é mostrado na Figura 25, onde nota-se uma separação entre os grupos mais nítida que a obtida com $CPPS_{5prc}$, porém também não muito evidente. Os valores do CPP_{median} para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 26, onde observa-se mais uma vez a tendência em considerar a grande maioria das vozes como patológicas. É possível observar na Tabela 8 que para os valores-p de CPP, $T=10ms$ e $T=2ms$, os menores resultados foram os do teste estatístico da mediana e do quinto percentil, respectivamente.

Figura 23: Boxplot CPPS_{5prc} utilizando T = 2 ms, L = M = 7 e escala em dB.



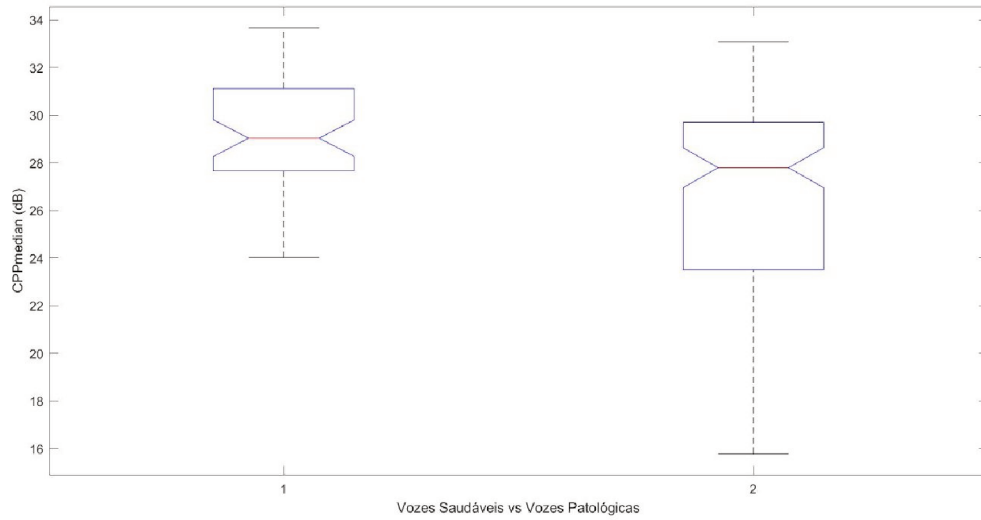
Fonte: a autora.

Figura 24: Valores para cada grupo e limiar de classificação de 12,22 dB para CPPS_{5prc} utilizando T = 2 ms, L = M = 7 e escala em dB.



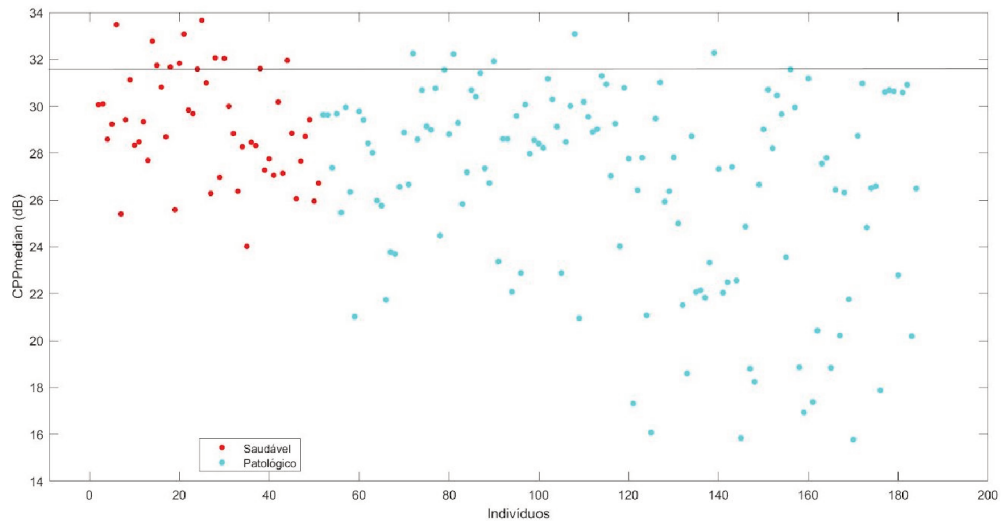
Fonte: a autora.

Figura 25: Boxplot CPP_{median} utilizando $T = 10ms$ e escala dB.



Fonte: a autora.

Figura 26: Valores para cada grupo e limiar de classificação para CPP_{median} utilizando $T = 10 ms$ e escala em dB, limiar 31,57dB.



Fonte: a autora.

Após o cálculo da precisão de classificação de CPP, $T=10ms$ e $T=2ms$, e CPPS, $T=2ms$ e $L=M=7$ e $L=M=11$ através do classificador OB-ROC, foi calculada a melhor precisão com a combinação das estatística utilizando o SVM.

Combinando diferentes estatísticas da mesma distribuição para discriminar vozes saudáveis e patológicas, os resultados de classificação obtidos com o SVM são apresentados na Tabela 10. É pertinente ressaltar que esses são os resultados médios obtidos utilizando uma validação cruzada k-fold com k=10. Como pode-se observar, o melhor resultado foi obtido combinando CPP_{mean} , CPP_{kurt} , CPP_{mode} e CPP_{std} calculados com T=2ms, resultando numa precisão de classificação igual a 80,80%. Esse resultado representa um acréscimo de 4,3% na melhor classificação obtida com apenas um parâmetro. Todavia, esse resultado não representa um grande avanço em frente aos 72,28% obtidos considerando todas as vozes patológicas. Isso indica que as estatísticas de distribuições do CPP e CPPS não se mostraram muito efetivas para discriminar vozes saudáveis e patológicas, conclusão que vai de encontro ao apresentado em [10]. Não se deve descartar a hipótese da causa para esse baixo desempenho ser a base de dados VOICEd que, como descrito na seção 3, aparenta ter sido gravada e rotulada sem os cuidados esperados.

Tabela 10: Melhores precisões de classificação utilizando SVM.

Parâmetros CPP e CPPS			Estatísticas que resultam na melhor precisão		
	Tempo	Suavização	Escala		Melhor resultado (%)
CPP	T= 10ms	-	dB	Média, Mediana, Assimetria, 95prc	79,08
	T=2ms	-	dB	Média, Curtose, Moda, Desvio Padrão	80,80
CPPS	T= 2ms	L=M=7	dB	Média, Mediana, Curtose, Intervalo, Desvio Padrão, 5prc	75,52
	T=2ms	L=M=11	dB	Média, Curtose, Assimetria, 5prc	75,54

Fonte: a autora

4.3 VOZES SAUDÁVEIS E VOZES COM DISFONIA HIPERCINÉTICA

Após obtidos os resultados para o caso de vozes saudáveis versus vozes patológicas, foi realizado o teste de Kruskal-Wallis e feita a classificação para o caso vozes saudáveis e com disfonia hipercinética. Os resultados são apresentados na Tabela 11 e 12, respectivamente.

Tabela 11: Valor-p das Estatísticas das Distribuições para Vozes Saudáveis versus Vozes com Disfonia Hipercinética.

Parâmetros CPP e CPPS				Valor-p das Estatísticas das Distribuições									
Tempo	Suav.	Escala		Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda
CPP	T=10ms	-	dB	9,90E-03	3,60E-03	5,90E-03	3,72E-02	1,80E-03	2,80E-01	3,29E-01	1,00E-03	1,00E-03	4,36E-02
	T=2ms	-	dB	8,75E-02	8,73E-03	2,00E-02	1,92E-01	2,98E-02	8,00E-02	1,46E-01	1,43E-02	1,43E-02	1,52E-01
CPPS	T=2ms	L=M=7	dB	7,62E-01	7,36E-01	1,74E-01	3,60E-01	4,38E-01	2,70E-03	9,43E-01	2,00E-01	2,00E-01	7,90E-02
	T=2ms	L=M=11	dB	3,69E-01	2,31E-01	5,76E-02	3,11E-01	4,03E-01	1,20E-03	7,71E-01	7,47E-01	7,47E-01	1,22E-01

Fonte: a autora.

Tabela 12: Precisão de Classificação das Estatísticas das Distribuições das Vozes Saudáveis versus Vozes com Disfonia Hipercinética.

Parâmetros CPP e CPPS				Precisão de Classificação das Estatísticas das Distribuições (%)									
Tempo	Suavização	Escala		Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda
CPP	T=10ms	-	dB	61,54	64,10	63,24	60,17	65,52	57,63	59,13	71,55	71,55	61,86
	T=2ms	-	dB	59,32	62,39	65,25	58,12	61,86	63,56	63,56	67,52	67,52	61,02
CPPS	T=2ms	L=M=7	dB	59,32	58,47	60,68	57,63	61,86	65,81	61,54	61,4	61,4	61,86
	T=2ms	L=M=11	dB	58,97	58,97	60,17	60,17	62,93	67,52	61,54	60,34	60,34	61,86

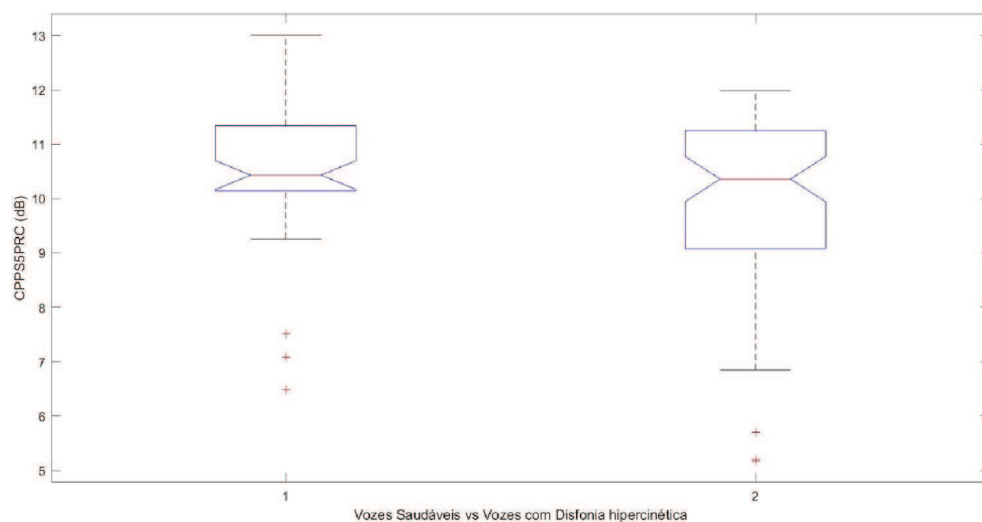
Fonte: a autora.

Para este caso, o $CPPS_{5prc}$ calculado com T=2 e L=M=7, parâmetro utilizado em [10], apresentou um valor-p igual a 1,74E-01, maior do que 0,05, e uma precisão de classificação

igual a 60,68%. A área ROC obtida foi 0,5737, valor muito inferior ao 0,95 alcançado em [10]. O boxplot do $CPPS_{5prc}$ para as vozes saudáveis e patológicas é mostrado na Figura 27, onde nota-se que novamente não há uma separação definida entre os dois grupos. Os valores do $CPPS_{5prc}$ para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 28, onde é possível observar a tendência do classificador em considerar quase todas as vozes (abaixo do limiar) como patológicas. Ressaltando que uma precisão de 56,77 % seria obtida ao classificar todas as vozes como patológicas. Considerando o CPPS, a melhor estatística descritiva é o $CPPS_{range}$ calculado com $L=M=11$, o qual obteve um valor-p igual a $1,20E-03$, uma precisão de classificação igual a 67,62% e uma AROC igual a 0,6761.

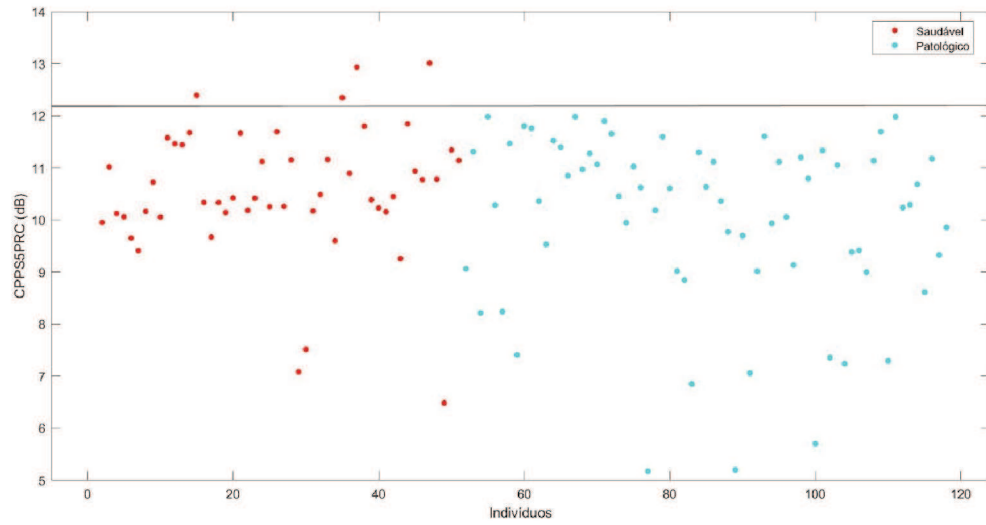
Como pode ser observado nas tabelas 11 e 12, neste trabalho a melhor estatística descritiva foi o CPP_{std} , com $T=10ms$, que obteve um resultado similar a variância, o que era esperado, pois o desvio padrão é a raiz quadrada da variância. Foi obtida uma precisão de 71,55%, uma área ROC de 0,6773, um valor-p de $1,00E-03$ e um limiar de 2,14 dB. O boxplot do CPP_{std} para as vozes saudáveis e com disfonia hiperkinética é mostrado na Figura 29, onde nota-se uma separação entre os grupos mais nítida que a obtida com $CPPS_{5prc}$. Os valores do CPP_{std} para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 30.

Figura 27: Boxplot $CPPS_{5prc}$ utilizando $T = 2 ms$, $L=M=7$ e escala em dB.



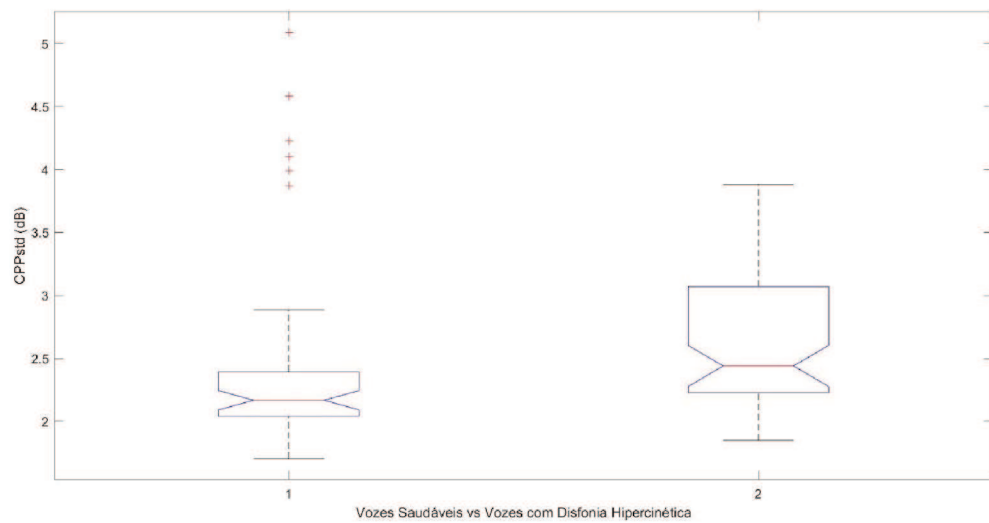
Fonte: a autora.

Figura 28: Valores para cada grupo e limiar de classificação para $CPPS_{sprc}$ utilizando $T = 2$ ms, $L = M = 11$, limiar de 12,16 dB e escala em dB.



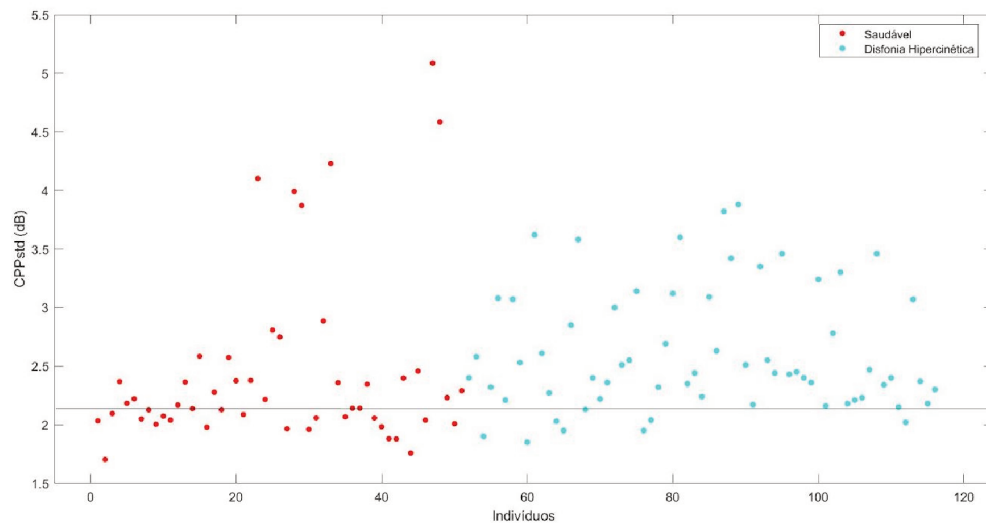
Fonte: a autora.

Figura 29: Boxplot CPP_{std} utilizando $T = 2$ ms e escala em dB, valor-p de $1,00E-03$.



Fonte: a autora.

Figura 30: Valores para cada grupo e limiar de classificação para CPP_{std} utilizando $T = 2$ ms, limiar de 2,14 dB e escala em dB.



Fonte: a autora.

Buscando melhores resultados de precisão, foi utilizado novamente o SVM para fazer a combinação das estatísticas da mesma distribuição e os resultados de classificação obtidos são apresentados na Tabela 13. Como na subseção anterior, é pertinente ressaltar que esses são os resultados médios obtidos utilizando uma validação cruzada k-fold com $k=10$.

Como pode-se observar, o melhor resultado foi obtido combinando CPP_{mode} , CPP_{Range} , CPP_{std} e CPP_{5prc} calculados com $T=2$ ms, resultando numa precisão de classificação igual a 75,85%. Esse resultado representa um acréscimo de 4,3 % na melhor classificação obtida com apenas um parâmetro. Portanto, representando um grande avanço em frente aos 56,77% obtidos considerando todas as vozes patológicas. Isso indica que as estatísticas de distribuições do CPP e CPPS mostram-se mais efetivas para discriminar vozes saudáveis e com disfonia hiperkinética, do que vozes saudáveis e o grupo de todas as patologias.

Tabela 13: Melhores precisões de classificação utilizando SVM para saudável versus Disfonia Hipercinética.

Parâmetros CPP e CPPS				Estatísticas que resultam na melhor precisão	
	Tempo	Suavização	Escala		Melhor resultado (%)
CPP	T= 10ms	-	dB	Moda, Intervalo, Desvio Padrão, 95prc	75,84
	T=2ms	-	dB	Média, Curtose , Moda, Desvio Padrão	73,65
CPPS	T= 2ms	L=M=7	dB	Média, Moda , Assimetria, 5prc	72,71
	T=2ms	L=M=11	dB	Mediana, Moda, Intervalo, 5prc	67,52

Fonte: a autora

4.4 VOZES SAUDÁVEIS E VOZES COM DISFONIA HIPOCINÉTICA

O procedimento da seção 3 foi repetido para o caso de vozes saudáveis e vozes com disfonia hipocinética. Com isso, os resultados de valor-p do teste de Kruskal-Wallis e de precisão de classificação após a remoção dos outliers podem ser vistos abaixo nas Tabelas 14 e 15, respectivamente.

Tabela 14: Valor-p das Estatísticas das Distribuições para Vozes Saudáveis versus Vozes com Disfonia Hipocinética.

Parâmetros CPP e CPPS				Valor-p das Estatísticas das Distribuições									
	Tempo	Suav.	Escala	Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var,	Moda
CPP	T= 10ms	-	dB	7,36E-07	8,12E-08	2,81E-06	2,69E-06	1,49E-05	8,40E-01	7,79E-02	1,67E-01	1,67E-01	1,71E-04
	T=2ms	-	dB	6,02E-07	2,38E-06	3,63E-08	4,01E-06	1,40E-05	6,82E-01	5,60E-01	7,79E-02	7,79E-02	2,80E-03
CPPS	T= 2ms	L=M=7	dB	3,30E-03	2,40E-02	9,00E-04	2,21E-05	3,00E-03	1,67E-01	3,70E-01	2,51E-01	2,51E-01	5,90E-03
	T=2ms	L=M=11	dB	2,04E-04	7,90E-03	1,48E-04	1,90E-04	2,78E-03	2,36E-01	3,24E-01	1,21E-01	1,21E-01	2,70E-03

Fonte: a autora.

Tabela 15: Precisão de Classificação das Estatísticas das Distribuições das Vozes Saudáveis versus Vozes com Disfonia Hipocinética.

Parâmetros CPP e CPPS			Precisão de Classificação das Estatísticas das Distribuições (%)										
	Tempo	Suav.	Escala	Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda
CPP	T= 10ms	-	dB	79,76	79,76	78,57	76,47	62,65	63,53	67,47	68,75	68,75	74,12
	T=2ms	-	dB	80	80,95	81,18	80,95	62,39	66,67	60	69,05	69,05	69,41
CPPS	T= 2ms	L=M=7	dB	70,93	70,93	73,81	75,29	63,39	64,28	64,28	64,19	64,19	70,93
	T=2ms	L=M=11	dB	70,59	71,76	75,29	70,59	63,85	63,09	63,53	64,28	64,28	71,76

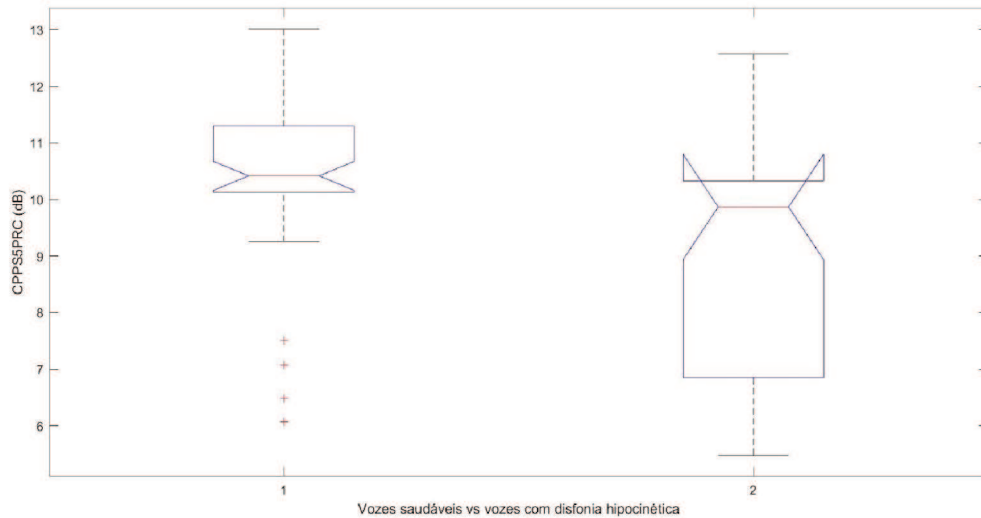
Fonte: a autora.

Observando novamente o melhor caso obtido por [10], o $CPPS_{5prc}$, $T=2ms$, $L=M=7$, novamente não obteve-se uma melhor precisão que o CPP. Para $CPPS_{5prc}$ apresentou-se um valor-p de $9,00E-04$, uma precisão de classificação de 73,81%, com o limiar de 9,4034 dB e área ROC de 0,7141, resultado ainda inferior do que o 0,95 obtido em [10]. O boxplot do $CPPS_{5prc}$ para as vozes saudáveis e com disfonia hipocinética é mostrado na Figura 31, onde nota-se que há uma separação entre os dois grupos como sugerido pelo valor-p inferior a 0.05. Os valores do $CPPS_{5prc}$ para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 32. Ressaltando que uma precisão de 60% seria obtida ao classificar todas as vozes como saudáveis. Considerando o CPPS, a melhor estatística descritiva é o $CPPS_{5prc}$ calculado com $L=M=11$, o qual obteve um valor-p igual a $1,48E-04$, uma precisão de classificação igual a 75,29% e uma AROC igual a 0,7439.

Analisando as tabelas 14 e 15 é possível observar que o melhor resultado apresentado neste trabalho foi de CPP_{5prc} , $T=2ms$, o qual obteve um valor-p de $3,63E-08$, uma precisão de classificação de 81,18%, uma área ROC de 0,8541 e um limiar de 23,16 dB. O boxplot do CPP_{5prc} para as vozes saudáveis e com disfonia hipocinética é mostrado na Figura 33, onde nota-se uma separação entre os grupos mais nítida e evidente do que a obtida com $CPPS_{5prc}$. Os valores do CPP_{5prc} para cada indivíduo e o melhor limiar de classificação são mostrados na

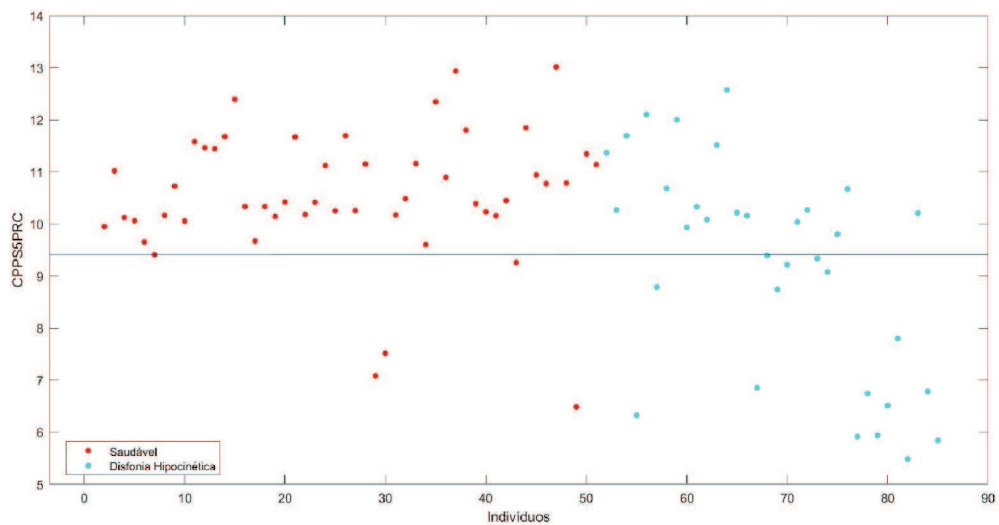
Figura 34, onde devido ao bom resultado é possível que acima do limiar há uma maior quantidade de vozes saudáveis e abaixo uma maior quantidade de vozes com disfonia hipocinética.

Figura 31: Boxplot CPPS_{5prc} utilizando T = 2 ms, L=M=7 e escala em dB.



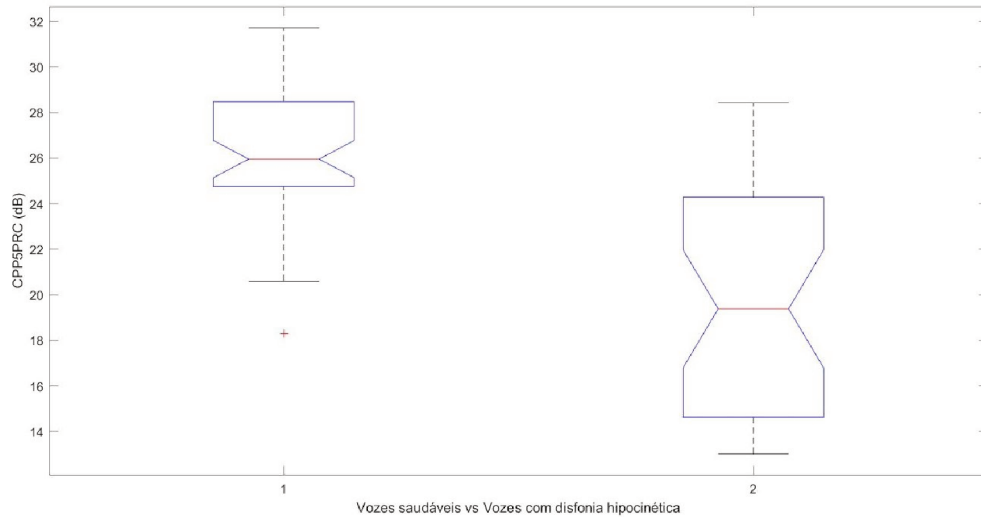
Fonte: a autora.

Figura 32: Valores para cada grupo e limiar de classificação para CPPS_{5prc} utilizando T = 2 ms, L = M = 7, limiar de 9,40 dB e escala em dB.



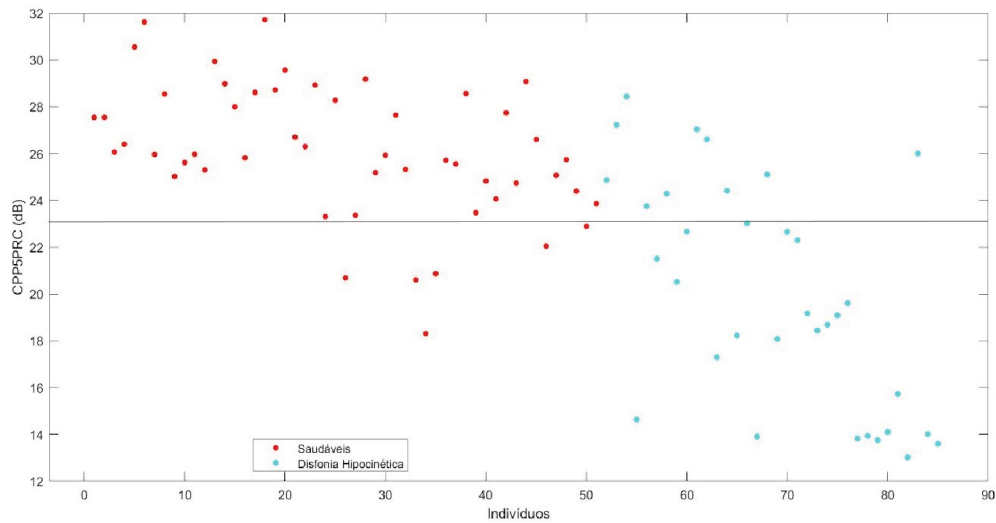
Fonte: a autora.

Figura 33: Boxplot CPP_{5prc} utilizando $T = 2$ ms e escala em dB.



Fonte: a autora.

Figura 34: Valores para cada grupo e limiar de classificação para CPP_{5prc} utilizando $T = 2$ ms, limiar de 23,16 dB e escala em dB.



Fonte: a autora.

Pode-se concluir, que neste caso, como nos anteriores, a capacidade discriminante de CPPS não foi maior do que a do CPP, sendo o melhor resultado de CPP 5,89% maior do que o melhor caso de CPPS.

Buscando novamente uma maior precisão de classificação, foram combinadas diferentes estatísticas da mesma distribuição para discriminar vozes saudáveis e com disfonia hipocinética, os resultados de classificação obtidos com o SVM são apresentados na Tabela 15. Esses são os resultados médios obtidos utilizando uma validação cruzada k-fold com $k=10$. Como pode-se observar, o melhor resultado foi obtido combinando CPP_{median} , CPP_{skew} e CPP_{5prc} calculados com $T=10$, resultando numa precisão de classificação igual a 88,09%. Esse resultado representa um acréscimo de 6,91 % na melhor classificação obtida com apenas um parâmetro. Representando um grande avanço em frente aos 60% obtidos considerando todas as vozes saudáveis. Mostrando assim que as estatísticas de distribuições do CPP e CPPS são efetivas para discriminar vozes saudáveis e vozes com disfonia hipocinética.

Tabela 16: Melhores precisões de classificação utilizando SVM para Vozes Saudáveis versus Vozes com Disfonia Hipocinética.

Parâmetros CPP e CPPS			Estatísticas que resultaram na melhor precisão		
	Tempo	Suavização	Escala		Melhor resultado (%)
CPP	T= 10ms	-	dB	Mediana, Assimetria, 5prc	88,09
	T=2ms	-	dB	Mediana, Moda, Curtose, Desvio Padrão, 5prc	87,87
CPPS	T= 2ms	L=M=7	dB	Mediana, Moda, Assimetria, 5prc	87,90
	T=2ms	L=M=11	dB	Moda, Variância, 5prc, 95prc	75,29

Fonte: a autora

4.5 VOZES SAUDÁVEIS E VOZES COM LARINGITE CAUSADA POR REFLUXO

Por último, foi realizada a análise das vozes saudáveis *versus* vozes com laringite causada por refluxo. Foi executado novamente o teste de Kruskal-Wallis para a obtenção do valor-p das estatísticas de distribuições e utilizado o classificador OB-ROC para determinação da precisão de classificação, os resultados são apresentados nas Tabelas 17 e 18.

Tabela 17: Valor-p das Estatísticas das Distribuições para Vozes Saudáveis versus Vozes com laringite causada por refluxo.

Parâmetros CPP e CPPS				Valor-p das Estatísticas das Distribuições									
Tempo	Suav.	Escala		Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda
CPP	T=10ms	-	dB	7,90E-02	5,60E-01	7,05E-01	5,19E-02	6,97E-01	2,46E-01	1,78E-01	6,55E-01	6,55E-01	9,26E-01
	T=2ms	-	dB	1,90E-01	1,83E-01	9,14E-01	2,00E-01	2,00E-01	2,70E-01	1,67E-01	1,44E-01	1,44E-01	7,00E-01
CPPS	T=2ms	L=M=7	dB	1,10E-02	9,40E-03	6,76E-01	1,91E-01	5,89E-01	7,38E-01	2,74E-01	2,11E-02	2,11E-02	8,72E-01
	T=2ms	L=M=11	dB	9,55E-01	5,19E-01	6,88E-01	9,11E-01	4,36E-01	6,53E-01	2,23E-01	6,60E-03	6,60E-03	9,48E-01

Fonte: a autora.

Tabela 18: Precisão de Classificação das Estatísticas das Distribuições das Vozes Saudáveis versus Vozes com laringite causada por refluxo.

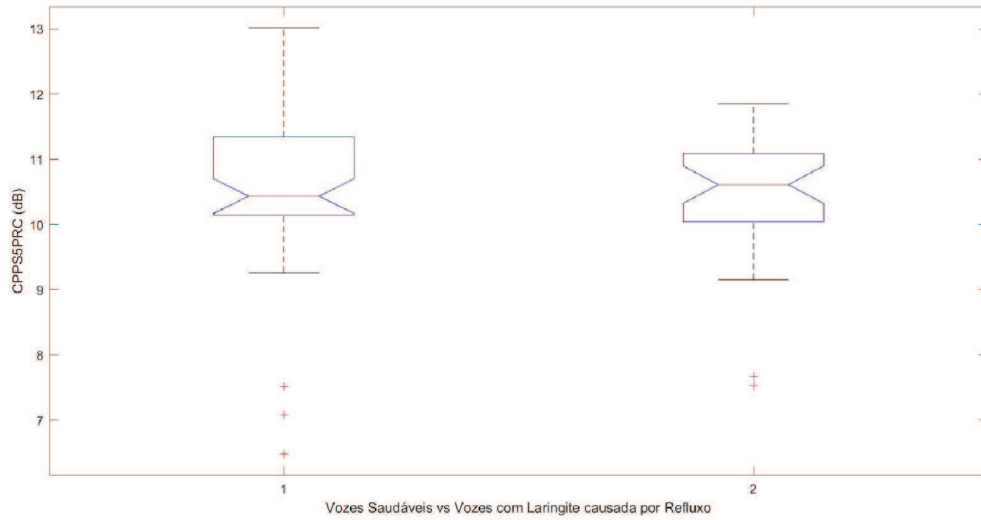
Parâmetros CPP e CPPS				Precisão de Classificação das Estatísticas das Distribuições (%)									
Tempo	Suavização	Escala		Média	Mediana	5prc	95prc	Curt.	Range	Assim.	Desvio Padrão	Var.	Moda
CPP	T=10ms	-	dB	67,07	61,73	60,97	61,45	62,5	66,27	62,96	65,38	65,38	62,65
	T=2ms	-	dB	64,19	62,5	62,96	62,19	64,63	66,27	65,06	65,47	65,47	62,65
CPPS	T=2ms	L=M=7	dB	66,67	66,67	60,98	62,65	65,85	65,85	66,67	65,00	65,00	61,45
	T=2ms	L=M=11	dB	62,65	63,86	62,65	62,65	61,47	65,85	61,45	69,51	69,51	61,45

Fonte: a autora.

Para a discriminação de vozes saudáveis e vozes com laringite causada por refluxo, novamente, diferentemente do que ocorre em [10], o $CPPS_{5prc}$, $T=2ms$ e $L=M=7$, não obteve o melhor resultado, tanto para o valor-p quanto para a precisão de classificação. Apresentou um valor-p igual a 0,6758, significativamente maior do que 0,05, e uma precisão de classificação igual a 60,98%. A área ROC obtida foi 0,5272, muito inferior ao 0,95 alcançado em [10]. O boxplot do $CPPS_{5prc}$ para as vozes saudáveis e com laringite causada por refluxo é mostrado na Figura 35, onde é visível que não há uma separação clara entre os dois grupos, o que já era esperado devido ao alto valor-p. Os valores do $CPPS_{5prc}$ para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 36, onde observa-se a tendência do classificador em considerar quase todas as vozes (acima do limiar) como saudáveis. Vale ressaltar que uma precisão de 61,45 % seria obtida ao classificar todas as vozes como saudáveis.

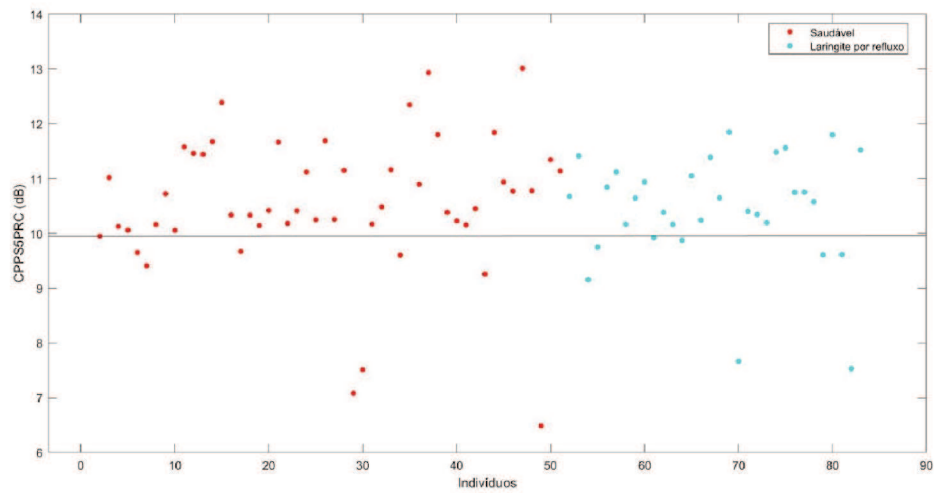
Neste caso, a maior precisão foi de 69,51%, de $CPPS_{std}$, para $T=2ms$ e $L=M=11$, sendo o único caso neste trabalho que a maior precisão foi obtida por CPPS. O valor-p obtido também foi menor do que 0,05, sendo 0,00660, com área ROC de 0,6787 e com um limiar de 2,11 dB. O boxplot do $CPPS_{std}$ para as vozes saudáveis e patológicas é mostrado na Figura 37, onde nota-se uma separação entre os grupos um pouco mais nítida que a obtida com $CPPS_{5prc}$, porém também não muito evidente. Os valores do $CPPS_{5prc}$ para cada indivíduo e o melhor limiar de classificação são mostrados na Figura 38, onde observa-se mais uma vez a tendência em considerar a grande maioria das vozes como saudáveis.

Figura 35: Boxplot CPPS_{sprc} utilizando T = 2 ms, L=M=7 e escala em dB.



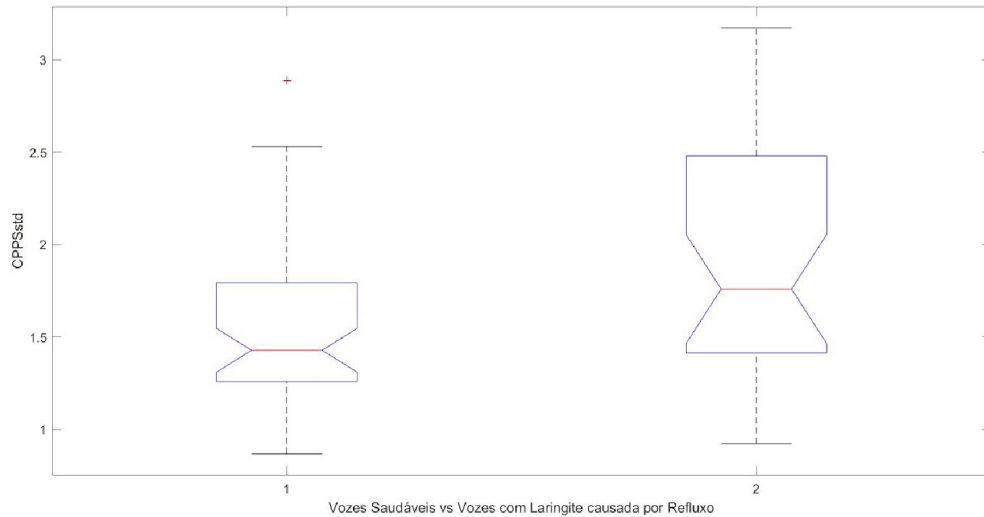
Fonte: a autora.

Figura 36: Valores para cada grupo e limiar de classificação para CPPS_{sprc} utilizando T = 2 ms, L = M = 7, limiar de 9,93 dB e escala em dB.



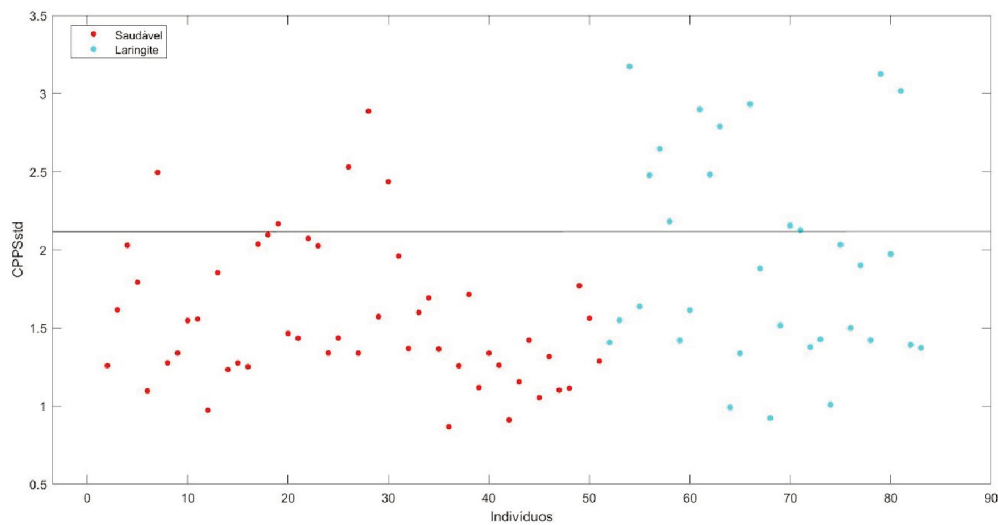
Fonte: a autora.

Figura 37: Boxplot $CPPS_{std}$ utilizando $T = 2$ ms, $L=M=11$ e escala em dB.



Fonte: a autora.

Figura 38: Valores para cada grupo e limiar de classificação para $CPPS_{std}$ utilizando $T = 2$ ms, $L = M = 11$, limiar de 2,11 dB e escala em dB.



Fonte: a autora.

Em busca de um melhor resultado, novamente foi utilizado o SVM e os resultados obtidos são apresentados na Tabela 19. Como pode-se observar, o melhor resultado foi obtido através da combinação de CPP_{mean} , CPP_{median} e CPP_{mode} calculados com $T=2ms$, resultando numa precisão de classificação igual a 73,57%. Esse resultado representa um acréscimo de

4,06 % na melhor classificação obtida com apenas um parâmetro. Todavia, esse resultado não representa um avanço em frente aos 61,45% obtidos considerando todas as vozes saudáveis. Isso indica que as estatísticas de distribuições do CPP e CPPS não se mostraram muito efetivas para discriminar vozes saudáveis e vozes com laringite causada por refluxo, conclusão que vai de encontro ao apresentado em [10]. Novamente, não se deve descartar a hipótese da causa para esse baixo desempenho ser a base de dados VOICEd que, como descrito na seção 3, aparenta ter sido gravada e rotulada sem os cuidados esperados.

Tabela 19: Melhores precisões de classificação utilizando SVM para vozes saudáveis versus vozes com laringite causada por refluxo.

Parâmetros CPP e CPPS			Estatísticas que resultaram na melhor precisão		
	Tempo	Suavização	Escala		Melhor resultado (%)
CPP	T= 10ms	-	dB	Média, Mediana, Variância	72,66
	T=2ms	-	dB	Média, Mediana, Moda	73,57
CPPS	T= 2ms	L=M=7	dB	Média, Mediana, Curtose, Assimetria, 95prc	70,23
	T=2ms	L=M=11	dB	Média, Mediana, Intervalo, Assim., Desvio Padrão, 5prc	69,51

Fonte: a autora

5 CONCLUSÃO

Foi visto que os métodos atuais utilizados para identificar patologias vocais são invasivos e desconfortáveis para o paciente ou subjetivos. Com isso, o trabalho apresentado buscou investigar as distribuições de CPP e CPPS e suas estatísticas descritivas como discriminantes entre vozes saudáveis e patológicas, saudáveis e com disfonia hipercinética, saudáveis e com disfonia hipocinética e saudáveis e com laringite causada por refluxo.

Foram avaliadas 10 distribuições estatísticas tanto para CPP com $T=2\text{ms}$ e $T=10\text{ms}$ quanto para CPPS com $T=2\text{ms}$, $L=M=7$ e $L=M=11$ utilizando 184 vozes da base de dados VOICEd, que apesar de conter um total de 208 vozes, 24 foram descartadas por conter ruído ou por falta de especificação. Das 184 gravações, 51 são de indivíduos saudáveis e 133 indivíduos com patologias. Na tentativa de obter melhores resultados, antes do teste de Kruskal-Wallis e de realizar a classificação os outliers foram retirados.

No caso da discriminação entre vozes saudáveis e patológicas a diferença entre as maiores precisões de classificação CPP e CPPS foi de apenas 1,50%. Demonstrando que para essa base de dados o CPP apresenta uma maior capacidade discriminante que o CPPS em grande parte dos casos.

Dentre a avaliação das vozes de indivíduos saudáveis e vozes de indivíduos com cada categoria de patologia, o melhor resultado foi na discriminação da disfonia hipocinética. Com a maior precisão de 81,18% pelo classificador OB-ROC para CPP_{SPRC} , com um valor-p igual a $3,63\text{E}-08$ e uma área ROC de 0,8326. Contendo neste caso a maior diferença entre as maiores precisões do classificador OB-ROC de CPP e CPPS que foi de 5,89%.

Dentre os quatro casos analisado, a discriminação entre vozes saudáveis e vozes com disfonia hipercinética e entre vozes saudáveis e vozes com laringite causada por refluxo apresentaram resultados similares. Sendo a maior precisão obtida para o primeiro caso 70,34%, para CPP_{std} , $T=2\text{ms}$, e o menor valor-p de $1,10\text{E}-03$, e para o segundo caso a maior precisão de 69,51% para CPPS_{std} , $T=2\text{ms}$ e $L=M=11$, apresentando um valor-p de $6,60\text{E}-03$. Sendo este o único caso no trabalho que o CPPS obteve o maior valor de precisão e não o CPP.

Em comparação com os resultados obtidos por [10], a maior área ROC encontrada neste trabalho foi de 0,8326 sendo que em [10] foi obtido um valor de 0,95. Para os casos de

discriminação entre vozes saudáveis e patológicas, saudáveis e com disfonia hipercinética e saudáveis e laringite causada por refluxo as áreas ROC foram de 0,6723, 0.6773 e 0.6787 respectivamente. Portanto, a menor área ROC foi a da discriminação entre vozes saudáveis e vozes patológicas, sendo um valor muito menor do que o obtido em [10].

Buscando melhorar os resultados, foi utilizado o SVM para obter a melhor combinação de estatísticas da mesma distribuição para a discriminação dos 4 casos. Com o uso do SVM todos os melhores resultados de precisão aumentaram. A maior precisão foi obtida novamente para a discriminação de vozes com disfonia hipocinética e vozes saudáveis, apresentando um valor de 88,09% de precisão e um aumento de 6,91% comparado com o resultado obtido pelo classificador OB-ROC.

Para os casos de discriminação entre vozes saudáveis e vozes patológicas, vozes saudáveis e vozes com disfonia hipercinética e vozes saudáveis e vozes com laringite causada por refluxo após o uso do SVM as maiores precisões de classificação foram de 80,80%, 75,84% e 73,57%.

Com isso, pode-se concluir que o uso do SVM para as estatísticas utilizadas neste trabalho e para esta base de dados auxilia na obtenção de maiores resultados de precisão de classificação.

Porém, utilizando somente o classificador OB-ROC, como dito anteriormente, os valores de área ROC foram menores do que os encontrados por [10], acredita-se que a grande diferença de valores ocorra devido a grande quantidade de doenças presentes no banco de dados, que são diferentes das doenças do banco de dados de [10] e devido a presença de pequenos ruídos nas gravações das vozes podem ter auxiliado neste resultado. Acredita-se também que o modo com que a base de dados foi gravada também tenha influência nos resultados finais e influência no melhor desempenho obtido por [10]. Enquanto o banco de dados utilizado neste trabalho foi gravado através do microfone de um Samsung Galaxy S4, em [10] as gravações foram feitas através de dois microfones, Mipro MU55HN e um ECM AE38, que foi fixado na jugular de cada locutor por meio de uma banda cirúrgica. Sendo o ECM AE38 um microfone que detecta as vibrações da pele induzidas pela atividade das pregas vocais e assim foi conectado a um gravador [10].

Entretanto, mesmo com a base de dados VOICEd, em geral após o uso do SVM foi possível obter resultados de precisão maiores do que 70% para CPP e para CPPS para todos

os casos. Como dito anteriormente, diferente de [10], o uso do CPPS não causou uma grande melhora dos resultados neste trabalho, exceto no caso da discriminação entre vozes saudáveis e vozes com laringite causada por refluxo.

Para a obtenção de melhores resultados, uma proposta para trabalhos futuros seria utilizar outras formas de análise, não analisando a primeira harmônica por exemplo. Outra proposta seria realizar a comparação entre todos os grupos para fazer a discriminação e não somente vozes saudáveis e vozes com patologias, mas também a discriminação de patologia *versus* patologia. Por fim, propõe-se a utilização de uma nova base de dados, pois foram encontrados diversos problemas na base de dados utilizada, como gravações com ruído, gravações sem especificação das patologias só da categoria e gravações sem especificação alguma, o que coloca em dúvida a credibilidade da mesma.

REFERÊNCIAS

- [1] ABCMED. “Laringoscopia: o que é? Como é realizada? Quem deve e quem não deve fazer?”. Abc.med. Disponível em: <<https://www.abc.med.br/p/exames-e-procedimentos/359364/laringoscopia-o-que-e-como-e-realizada-quem-deve-e-quem-nao-deve-fazer.htm>>. Acesso em: 20 dez. 2019.
- [2] Ablv, “O QUE É LARINGITE DE REFLUXO?”. Ablv, 06 ago. 2013. Disponível em: .<http://www.ablv.com.br/secao_detalhes.asp?s=12&id=267>. Acesso em: 01 jan. 2020.
- [3] Allmedtech. “LARYNGOSCOPE AND ACCESSORIES”. Allmedtech. Disponível em: <<https://allmedtech.com/labac.html>>. Acesso em: 31 dez. 2019.
- [4] ALVES, C., ““Laringite: o que é, Remédios, Tratamento, Sintomas e Causas.” 17 mar. 2018. Disponível em: <<https://opas.org.br/laringite-o-que-e-remedios-tratamento-sintomas-e-causas/>>. Acesso em: 20 dez. 2019.
- [5] Amazon Machine Learning. Guia do desenvolvedor. Classificação binária. Amazon. Disponível em: <https://docs.aws.amazon.com/pt_br/machine-learning/latest/dg/binary-classification.html>. Acesso em: 20 dezembro. 2019.
- [6] AWAN, S. N., “Analysis of dysphonia in speech and voice: an application guide”, Montalve NJ:KayPENTAX, 2011.
- [7] BARNES, D., HOLINGER, L., PELLETTIERE, E. “Prolapse of the Laryngeal Ventricle. Otolaryngology and head and neck surgery”. 88. 165-71. 10.1017/S0022215100067311. 1967.
- [8] BEHLAU, M. “A voz. O livro do especialista.”, 2001.

- [9] BRINCA, L. F., BATISTA, P. F., TAVARES, A. I., GONCALVES, C., MORENO, M. L., “Use of cepstral analyses for differentiating normal from dysphonic voices: A comparative study of connected speech versus sustained vowel in European portuguese female speakers,” *J. Voice*, vol. 28, no. 3, pp. 282–286, 2014.
- [10] CASTELLANA, A., CARULLO, A., CORBELLINI, S., ASTOFOLI, A., “Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel,” *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 646–654, March 2018.
- [11] CESARI, U., PIETRO, G. de, MARCIANO, E., NIRI, C., SANNINO, G., VERDE, L., “A new database of healthy and pathological voices”, *Computers & Electrical Engineering*, vol. 68, pp. 310-321, 5, 2018.
- [12] CESARI, U., PIETRO, G. de, MARCIANO, E., NIRI, C., SANINO, G., VERDE, L., “Dysphonia Detection Index (DDI): A New Multi-Parametric Marker to Evaluate Voice Quality. *IEEE Access*, v. 7, p. 55689-55697, 2019.
- [13] Clínica San Vicente. “Alterações da Voz: DISFONIA”, Clínica San Vicente, 2016. Disponível em: <https://www.clinicasanvicente.es/pt-pt/servicios/alteracoes-da-voz-disfonia/>. Acesso em: 20 dez. 2019.
- [14] Colunista Portal-Educação. “Alterações Vocais - Sulco Vocal”. Disponível em: <https://www.portaleducacao.com.br/conteudo/artigos/fonoaudiologia/alteracoes-vocais-sulco-vocal/27872>>. Acesso em: 21 dez. 2019.
- [15] CUF. “Nódulos Vocais”. Saude CUF, 2019. Disponível em: <https://www.saudecuf.pt/mais-saude/doencas-a-z/nodulos-vocais>>. Acesso em: 20 dez. 2019.

- [16] DELLER, J. R., HANSEN, J. H. L., PROAKIS, J. G., “Discrete-Time Processing of Speech Signals.” Piscataway, New Jersey: IEEE Press, 2000.
- [17] Eastern Virginia Medical School. “Examining the larynx”, Evms. Disponível em: <https://www.evms.edu/patient_care/specialties/ent_surgeons/services/laryngology/examinin_g_the_larynx/>. Acesso em 31 dez. 2019.
- [18] FAWCETT, T. “An introduction to ROC analysis”, Pattern Recognition Letters, Volume 27, pp. 861-874, 2006.
- [19] GARCIA, N. L. “Regressão Linear Simples. IME Unicamp”. IME UNICAMP. Disponível em: <<https://www.ime.unicamp.br/~nancy/Cursos/me104/regressao.pdf>>. Acesso em: 05 jan. 2020.
- [20] GIOVANNI, A., DUFLO, S., “Physiology of Voice Production”. Em: Remacle M., Eckel H. (eds) Surgery of Larynx and Trachea. Springer, Berlin, Heidelberg. 2009.
- [21] HÄRTE, “Espelho para laringe tamanho nº7”, Harte Instrumentos, 2019. Disponível em: <https://www.harteinstrumentos.com.br/produto/espelho-para-laringe-tamanho-no7/https://pt.wikipedia.org/wiki/Paralisia_da_laringe#/media/Ficheiro:Vocalcordparalyses.png>. Acesso em: 20 dez. 2019.
- [22] HEMAN-ACKAH, Y. D., et al. “Cepstral peak prominence: A more reliable measure of dysphonia,” Ann. Otol. Rhinol. Laryngol., vol. 112, no. 4, pp. 324–333, 2003.
- [23] HEMAN-ACKAH, Y. D., MICHAEL, D. D., CODING, G. S., “The relationship between cepstral peak prominence and selected parameters of dysphonia”, J. Voice 16, 20-27, 2002.

- [24] HEMAN-ACKAH, Y. D., SATALOFF, T., LAUREYNS, G., LURIE, D., MICHAEL, D. D., et al. “Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia”.
- [25] HILLENBRAND, J., HOUDE, R. A., “Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech,” *J. Speech Hearing Res.*, vol. 39, no. 2, pp. 311–321, 1996.
- [26] HILLENBRAND, J., CLEVELAND, R. A., ERICKSON, R. L., “Acoustic correlates of breathy vocal quality,” *Journal of Speech and Hearing Research*, vol. 37, pp. 769–778, August 1994.
- [27] HUCHE, F. Le, ALLALI, A., “La voz: anatomía y fisiología de los órganos de la voz y el habla”. Barcelona, 1993.
- [28] Itl NIST. “What are outliers in the data?”. Itl NIST. Disponível em: <<https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>>. Acesso em 10 jan. 2020.
- [29] ITU-T P. 56, “Objective measurement of active speech level,” International Telecommunications Union, Geneva, Switzerland 2011.
- [30] Iped. “Por que cuidar da voz?”. Iped. Disponível em: <<https://www.iped.com.br/materias/fonoaudiologia/cuidar-voz.html>>. Acesso em: 05 jan. 2020.
- [31] KRUSKAL, W. H., WALLIS, W. A., “Use of ranks in one-criterion variance analysis”. *Journal of the American Statistical Association*, v. 47, n. 260, p. 583–621, 1952.
- [32] KUMAR, B. R., BHAT, J. S., PRASAD, N., “Cepstral Analysis of Voice in Persons With Vocal Nodules”.

- [33] LITTLE, M. et al. “Nonlinear, biophysically-informed speech pathology detection. IEEE. Acoustics, Speech and Signal Processing”, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on [S.I], 2006. v. 2, p. II-II.
- [34] LOWELL, S. Y., COLTON, R. H., KELLEY, T. R., HAHN, Y. C., “Spectral and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker,” *Journal of Voice*, vol. 25, no. 5, p. 223–232, September 2011.
- [35] MAG, S., PIMENTEL, R. M., PILZ, W., “A Saúde vocal do professor: levantamento junto a escolas particulares de Porto Alegre” . *Pro-Fono*. 1996;8(2):25-30.
- [36] MARTINS, R. H. G., AMARAL, H. A., TAVARES, E. L. M., MARTINS, M. G., GONÇALVES, T. M., DIAS, N. H., “Voice Disorders: Etiology and Diagnosis,” *Journal of Voice*, vol. 30, no. 6, pp. 761.e1–761.e9, 2016.
- [37] MARYN, Y., BODT, M. de, ROY, N., “The acoustic voice quality index: Toward improved treatment outcomes assessment in voice disorders,” *J. Commun. Disorders*, vol. 43, no. 3, pp. 161–174, 2010.
- [38] MARYN, Y., ROY, N., DE BODT, M., VAN CAUWENBERGE, P., CORTHALS, P., “Acoustic measurement of overall voice quality: a meta-analysis”, *J. Acoust. Soc. Am.* 126, 2619-2634, 2009.
- [39] Medcalc. “ROC curve analysis”. Medcalc. Disponível em: <https://www.medcalc.org/manual/roc-curves.php>. Acesso em: 15 jan. 2020.
- [40] MOERS, C., MÖBIUS, B., ROSANOWSKI, F., NÖTH, EYSHOLDT, U., HADERLEIN, T., “Vowel- and text-based cepstral analysis of chronic hoarseness,” *J. Voice*, vol. 26, no. 4, pp. 416–424, 2012.

- [41] NAKAS, C. T., YIANNOUTSOS, C. T., “Ordered multiple-class ROC analysis with continuous measurements,” *Statistics in Medicine*, vol. 23, no. 22, p. 3437–3449, November 2004.
- [42] ONWORDI, L. N., YAGHCHI, C. AI, “Airway Glottic Insufficiency”. StatPearls. Treasure Island (FL): StatPearls Publishing, 04 nov. 2019. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK538207/>>. Acesso em: 05 jan. 2020.
- [43] Optim LCC. “FIBROSCÓPIO FLEXÍVEL / INDUSTRIAL / PARA LED / PORTÁTIL”. Direct Industry. Disponível em: <<https://www.directindustry.com/pt/prod/optim-llc/product-111823-1919040.html>>. Acesso em: 30 dez. 2019.
- [44] Oracle. “Outlier Detection Methods”. Oracle. Disponível em: <https://docs.oracle.com/cd/E17236_01/epm.1112/cb_statistical/frameset.htm?ch07s02s10s01.html>. Acesso em: 12 jan. 2020.
- [45] OSTERTAGOVÁ, E., OSTERTAG, O., KOVÁČ, J., “Methodology and application of the Kruskal-Wallis test,” *Applied Mechanics and Materials*, vol.611, p. 115–120, October 2014.
- [46] Otoface. “VIDEOLARINGOSCOPIA E VIDEOLARINGOESTROBOSCOPIA”. Otoface. Disponível em: <<https://www.otoface.com/exames/videolaringoscopia-e-videolaringoestroboscopia.html>>. Acesso em: 30 dez. 2019.
- [47] Otorrinovillas. “Videolaringostroboskopi”. Otorrinovillas, 2016. Disponível em: <<https://otorrinovillas.com.br/exames-complementares/videolaringoestroboscopia/>>. Acesso em: 31 dez. 2019.

[48] PRATI, R. C., BATISTA, G., MONARD, C., “Curvas ROC para avaliação de classificadores”. Revista IEEE América Latina, v. 6, n. 2, p. 215-222, 2008.

[49] Pulmocenter, “Laringoscopia”, Pulmocenter, 18 mai. 2016. Disponível em: <<http://www.pulmocenter.com/portfolio/laringoscopia/>>. Acesso em: 20 dez. 2019.

[50] RAMOS, Bruno Thadeu Reis, “AS SEIS CANÇÕES TROVADORESAS DE FRUCTUOSO VIANNA: Aspectos intertextuais e perspectivas interpretativas para voz de contratenor na canção de câmara brasileira”, Escola de Música UFMG, 2013.

[51] Redação Tasabendo. “Patologias vocais e seus tratamentos”, Tasabendo, 03 nov. 2014. Disponível em: <<https://tasabendo.com.br/saude/patologias-vocais-e-seus-tratamentos/>>. Acesso em: 20 dez. 2019.

[52] ROY, N., MERRIL, R. M., GRAY, S. D., SMITH, E. M., “Voice disorders in the general population: prevalence, risk factors, and occupational impact,” The Laryngoscope, vol. 115, no. 11, p. 1988–1995, November 2005.

[53] SANTOS, C. J., “Avaliação do uso de classificadores para verificação de atendimento a critérios de seleção em programas sociais”, 2017.

[54] SATALOFF, R., HEMAN-ACKAH, Y., HAWKSHAW. “Clinical Anatomy and Physiology of the Voice. Otolaryngologic clinics of North America.” 40. 909-29, v. 10.1016/j.otc.2007.05.002. 2007.

[55] SARMENTO, P., “O que é a voz?”. Jun. 2015, Disponível em: <<https://repositoriodavoz.wordpress.com/2015/06/30/o-que-e-a-voz/>>. Acesso em: 01 dez. 2019.

[56] SCHLEGEL, P., KUNDUK, M., STINGL, M., SEMMLER, M., DÖLLINGER, M., et al. “Influence of spatial camera resolution in high-speed videoendoscopy on laryngeal

parameters”, PLOS ONE 14(4): e0215168. Abr. 2019. Disponível em <https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0215168.g001>.

Acesso em: 02 dez. 2019.

[57] SORRENTINO, J., “Guia de estatística — Regressão linear. Ecommerce Brasil, 19 abr. 2019. Disponível em:

<https://www.ecommercebrasil.com.br/artigos/guia-de-estatistica%E2%80%8A-%E2%80%8Aregressao-linear/>. Acesso em 02 jan. 2020.

[58] SOUZA, Carla Lima de et al . Fatores associados a patologias de pregas vocais em professores. Rev. Saúde Pública, São Paulo , v. 45, n. 5, p. 914-921, Oct. 2011 . Disponível em:<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102011000500013&lng=en&nrm=iso>. Epub Aug 05, 2011.

<http://dx.doi.org/10.1590/S0034-89102011005000055>. Acesso em: 07 Jan. 2020.

[59] Strattner, “O que é a Laringoscopia?”, Strattner, 05 jan. 2018. Disponível em: <https://www.strattner.com.br/blog/categorias/cirurgia-minimamente-invasiva/o-que-e-laringoscopia.asp>. Acesso em: 30 dez. 2019.

[60] SZKLANNY, K., WRZECIONO, P.. "Relation of RBH Auditory-Perceptual Scale to Acoustic and Electroglottographic Voice Analysis in Children With Vocal Nodules," in IEEE Access, vol. 7, pp. 41647-41658, 2019.

[61] TOKUHARA, L., BISPO, B., DAJER, M., THEODORO, E., et al. “Avaliação de Capacidade Discriminante de Distribuições da Proeminência do Pico Cepstral para Identificar Nódulos Vocais. XXXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBRT2019. 29 set. 2019 , PETRÓPOLIS, RJ.

[62] TRIBOLET, J. M., “Seismic Applications of Homomorphic Signal Processing”, Englewood Cliffs, New Jersey: Prentice Hall, 1979.

- [63] Wheeler, D. J.; Chambers, D. S. (1992). Understanding Statistical Process Control. SPC Press.
- [64] Wikipédia. “Máquina de vetores de suporte”. Wikipédia, 7 jan. 2020. Disponível em: <https://pt.wikipedia.org/wiki/M%C3%A1quina_de_vetores_de_suporte>. Acesso em: 2 jan. 2019
- [65] Wikipédia. “Vocal Cord Paralysis”. Wikipédia, 20 abr. 2016. Disponível em: <<https://pt.wikipedia.org/wiki/Ficheiro:Vocalcordparalyses.png>>. Acesso em: 20 dez. 2019.
- [66] Wikipédia. “Percentil”. Wikipédia, 13 mai. 2019. Disponível em: <<https://pt.wikipedia.org/wiki/Percentil>>. Acesso em: 01 jan. 2020.
- [67] Wikipédia. “Outlier”. Wikipédia, 18 dez. 2019. Disponível em: <<https://en.wikipedia.org/wiki/Outlier>>. Acesso em 11 jan. 2020.
- [68] Wikipédia. “Outlier”. Wikipedia, 19 de mai. 2018. Disponível em: <https://pt.wikipedia.org/wiki/Outlier#/media/Ficheiro:Outlier_statistics.svg>. Acesso em: 10 jan. 2020.
- [69] WOLFE, V. I., MARTIN, D. P., PALMER, C. I., “Perception of dysphonic voice quality by naive listeners”, J. Speech Lang. Hears. Res. 43, 697-705, 2000.
- [70] ZHANG, Y., JIANG, J. J., “Acoustic analyses of sustained and running voices from patients with laryngeal pathologies,” Journal of Voice, vol. 22, no. 1, p. 1–9, January 2008.
- [71] ZHOU, X.-H., MCCLISH, D. K., OBUCHOWSKI, N. A., “Statistical Methods in Diagnostic Medicine”. John Wiley & Sons Inc, 2002.