INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
Coordenação do Curso de Engenharia Elétrica

VINICIUS JEFFERSON DIAS VIEIRA

Análise do Efeito de Variações Acústicas Não Estacionárias em Sinais de Voz Afetados por Patologias Laríngeas

João Pessoa
2024

VINICIUS JEFFERSON DIAS VIEIRA

Análise do Efeito de Variações Acústicas Não Estacionárias em Sinais de Voz Afetados por Patologias Laríngeas

Trabalho de Conclusão de Curso submetido à Coordenação do Curso de Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, como parte dos requisitos para a obtenção do grau de Engenheiro Eletricista.

Orientadora: Silvana Luciene do Nascimento Cunha Costa

João Pessoa
2024

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Nilo Peçanha do IFPB, *campus* João Pessoa

V657a Vieira, Vinicius Jefferson Dias.

Análise do efeito de variações acústicas não estacionárias em sinais de voz afetados por patologias laríngeas / Vinicius Jefferson Dias Vieira. – 2024.
50 f. : il.
TCC (Graduação – Engenharia Elétrica) –Instituto Federal de Educação da Paraíba / Coordenação do Curso Superior em Engenharia Elétrica, 2024.
Orientação : Profa. Silvana Luciene do Nascimento Cunha Costa.
1. Processamento digital de sinais. 2. Análise acústica de voz. 3. Patologias na laringe. 4. Índice de não-estacionariedade. I. Título.

CDU 621.391(043)

Bibliotecária responsável: Lucrecia Camilo de Lima – CRB 15/132

VINICIUS JEFFERSON DIAS VIEIRA

Análise do Efeito de Variações Acústicas Não Estacionárias em Sinais de Voz Afetados por Patologias Laríngeas

Trabalho de Conclusão de Curso submetido à Coordenação do Curso de Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, como parte dos requisitos para a obtenção do grau de Engenheiro Eletricista.

BANCA EXAMINADORA

gov.br Documento assinado digitalmente
Silvana Luciene do Nascimento Cunha Costa
Data: 27/09/2024 11:19:04-0300
Verifique em https://validar.iti.gov.br

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

gov.br Documento assinado digitalmente
Suzete Elida Nobrega Correia
Data: 27/09/2024 16:36:00-0300
Verifique em https://validar.iti.gov.br

Suzete Élida Nóbrega Correia, Dra. – IFPB
Membro da Banca

gov.br Documento assinado digitalmente
Carlos Danilo Miranda Regis
Data: 27/09/2024 10:50:20-0300
Verifique em https://validar.iti.gov.br

Carlos Danilo Miranda Regis, Dr. – IFPB
Membro da Banca

João Pessoa, 24 de setembro de 2024.

Agradecimentos

A gratidão é um dos sentimentos mais bonitos que são guardados no coração e que nos dá a certeza de que não estamos sozinhos na jornada da vida. Sendo assim, primeiramente (e acima de tudo) gostaria de agradecer a Deus por tudo. É difícil elencar todas as suas bênçãos na minha vida, por isso sou grato por cada segundo de desafios e vitórias pelos quais passei até aqui.

Agradeço a meus pais, Francisco e Verônica, e à minha irmã, Fernanda, por todo amor e apoio desde quando comecei na minha primeira graduação, em 2009. Eles entenderam minhas ausências, me abraçaram nas necessidades e celebraram minhas conquistas.

Gratidão à minha noiva, Maria Luiza, meu amor, uma pessoa admirável que tem me inspirado a buscar meus sonhos e a ser cada dia mais uma pessoa melhor.

Minha eterna gratidão à Professora Silvana, que não é apenas a orientadora deste trabalho, mas também uma referência, uma inspiração, uma amiga que Deus colocou na minha vida. Mais do que ter me ajudado a ser o profissional que sou hoje, sou grato por todas as conversas fora do contexto acadêmico. Gratidão a seu esposo, Professor Washington, por ter sido a primeira pessoa que acreditou em meu potencial.

Agradeço à Professora Suzete, não apenas por fazer parte desta banca, mas por toda parceria ao longo da minha jornada no IFPB (que neste ano completa 15 anos). Seus conselhos também me ajudaram a construir minha persona profissional.

Gratidão ao Professor Carlos Danilo, não apenas por fazer parte desta banca, mas também por me abrir as portas do laboratório do GPDS (grupo de processamento digital de sinais), no qual pude compartilhar pesquisas e desenvolver contribuições para os colegas.

A todos os colegas do GPDS pelo acolhimento. Aos amigos que direta ou indiretamente me ajudaram nesta segunda graduação: Igor Forcelli, Gabriel Diniz, Ildja Queiroz e Taciana Souza. Aos colegas de trabalho no SiDi por todo apoio: Juliana Inácio, Rafael Pertum e Renato Candido.

Gostaria também de mencionar alguns professores do IFPB pelas palavras de incentivo: Alfredo, Álvaro, Franklin, Michel e Joabson. Obrigado a todas as pessoas que, mesmo não sendo mencionadas aqui, torceram pelo meu sucesso.

Por fim, minha gratidão à pessoa que mais me incentivou: eu mesmo. Deixo a modéstia de lado por um instante para poder expressar minha alegria e orgulho de mim mesmo, por nunca ter desistido e, ainda, por sempre olhar no espelho em meio às dificuldades acreditando que dias melhores podem vir.

"De que adianta um homem ganhar o mundo inteiro,
se perde e destrói a si mesmo?"
(Jesus Cristo, segundo Lucas, capítulo 9, versículo 25)

RESUMO

A análise acústica é uma abordagem associada ao processamento digital de sinais (PDS) que desempenha um papel auxiliar no contexto da avaliação de distúrbios da voz. Por meio de técnicas de PDS, é possível extrair informações que podem ser características relevantes do sistema de produção vocal. Ao digitalizar e processar o sinal de voz, profissionais de saúde podem detectar alterações sutis na qualidade vocal que podem estar associadas a patologias, permitindo diagnósticos mais precisos e intervenções terapêuticas adequadas. Neste trabalho é realizado um estudo sobre a análise acústica baseada na não estacionariedade dos sinais de voz. Assim, é investigado como variações acústicas não estacionárias podem afetar a classificação de sinais de laringes saudáveis e patológicas. O Índice de Não-Estacionariedade (INS – *Index of Non-Stationarity*) é empregado como medida do grau de não estacionariedade da voz. De maneira mais detalhada, dois estudos de caso são conduzidos: 1) uso do INS como medida acústica; e 2) uso do INS para estabelecer o tamanho de segmento ideal para extração de outras medidas acústicas. Dois classificadores são utilizados neste trabalho: LDA (*Linear Discriminant Analysis*) e QDA (*Quadratic Discriminant Analysis*). Como resultados do estudo de caso 1, é verificado que a combinação de escalas de observação do INS proporciona um desempenho de classificação acima de 90% de acurácia. No contexto do estudo de caso 2, é observado que a utilização de segmentação adaptativa baseada em INS proporciona um desempenho de classificação acima de 92% de acurácia. Então, o desenvolvimento deste trabalho traz como principal contribuição a observação de que o INS é uma característica relevante da voz, tendo potencial para ser aplicado no contexto da classificação de distúrbios vocais.

Palavras-chave: Processamento Digital de Sinais. Análise Acústica da Voz. Patologias na Laringre. Índice de Não-Estacionariedade.

ABSTRACT

Acoustic analysis is an approach associated with digital signal processing (DSP) which plays an auxiliary role in the voice disorders assessment. Through DSP techniques, it is possible to extract information that may be relevant characteristics of the speech production system. By digitizing and processing the voice signal, speech pathologists can detect subtle changes in vocal quality that may be associated with pathologies, allowing for more accurate diagnoses and appropriate therapeutic interventions. This work presents a study on acoustic analysis based on the non-stationarity of voice signals. Thus, it is investigated how non-stationary acoustic variations can affect the classification of signals from healthy and pathological larynges. The Index of Non-Stationarity (INS) is employed as a measure of the degree of voice non-stationarity. In more detail, two case studies are conducted: 1) use of the INS as an acoustic feature; and 2) use of the INS to establish the ideal segment size for other acoustic features extraction. Two classifiers are used in this work: LDA (Linear Discriminant Analysis) and QDA (Quadratic Discriminant Analysis). As a result of case study 1, it is verified that the combination of INS-based observation scales provides a classification performance above 90% accuracy. In the context of case study 2, it is observed that the use of INS-based adaptive segmentation provides a classification performance above 92% accuracy. Then, the development of this work brings as its main contribution the observation that INS is a relevant characteristic of the voice, with potential to be applied in the context of voice disorders classification.

Keywords: Digital Signal Processing. Acoustic Analysis of Voice. Laryngeal Pathologies. Index of Non-Stationarity.

LISTA DE FIGURAS

Figura 1 – Pregas vocais saudáveis. 18
Figura 2 – Paralisia nas pregas vocais: (a) unilateral direita e (b) bilateral. 19
Figura 3 – Pregas vocais com edema de Reinke. 19
Figura 4 – Pregas vocais com nódulos. 20
Figura 5 – Gráficos de recorrência obtidos de: (a) um sinal de voz de uma laringe saudável; (b) um sinal de voz de uma laringe afetada por nódulos. . . . 24
Figura 6 – Sinal de voz da vogal sustentada /a/ e seu espectrograma considerando trechos de 800 ms e 40 ms, respectivamente, para as classes saudável ((a) e (b)) e patológico ((c) e (d)). 26
Figura 7 – Fluxograma resumido da extração do INS. 28
Figura 8 – INS calculado em diferentes escalas de sinais de voz: (a) saudável; (b) paralisia; (c) edema; (d) nódulo. A linha vermelha indica o valor do INS do sinal original. A linha verde representa o limiar obtido da distribuição Gamma dos *surrogates* para cada escala de tempo. 29
Figura 9 – Metodologia conduzida no estudo de caso 1, em que as escalas do INS são empregadas como características acústicas. 31
Figura 10 – Metodologia conduzida no estudo de caso 2, em que as medidas de análise linear (LPC, MFCC e GFCC) e não linear (MQRs) empregadas como características acústicas. 32
Figura 11 – Boxplots obtidos do INS para sinais saudáveis (SDL) e patológicos (PTL) considerando as dez escalas de observação. 34
Figura 12 – Melhores valores de acurácia média (%) obtidos com a classificação individual (Ind.) e combinada das MQRs utilizando LDA considerando os diferentes tipos de segmentação. 39
Figura 13 – Melhores valores de acurácia média (%) obtidos com a classificação individual (Ind.) e combinada das MQRs utilizando QDA considerando os diferentes tipos de segmentação. 41

LISTA DE TABELAS

Tabela 1 – Escalas de extração do INS. 31
Tabela 2 – Desempenho da classificação individual com LDA utilizando as medidas de INS de cada escala. 35
Tabela 3 – Desempenho da classificação individual com QDA utilizando as medidas de INS de cada escala. 35
Tabela 4 – Desempenho da classificação com LDA por meio das melhores combinações das escalas INS. 36
Tabela 5 – Desempenho da classificação com QDA por meio das melhores combinações das escalas INS. 37
Tabela 6 – Resultado da classificação com LDA considerando as medidas de análise linear em diferentes tipos de segmentação. 38
Tabela 7 – Resultado da classificação com QDA considerando as medidas de análise linear em diferentes tipos de segmentação. 39
Tabela 8 – Melhores resultados de classificação com LDA considerando as MQRs em diferentes tipos de segmentação. 40
Tabela 9 – Melhores resultados de classificação com QDA considerando as MQRs em diferentes tipos de segmentação. 41

SUMÁRIO

1 INTRODUÇÃO 12
1.1 Justificativa 13
1.2 Objetivos 16
1.2.1 Objetivo Geral 16
1.2.2 Objetivos Específicos 16
1.3 Organização deste Trabalho 16

2 FUNDAMENTAÇÃO TEÓRICA 17
2.1 Patologias na Laringe 18
2.1.1 Paralisia 18
2.1.2 Edema 19
2.1.3 Nódulos 19
2.2 Análise Acústica de Vozes Saudáveis e Patológicas 20
2.2.1 Análise Linear 20
2.2.1.1 LPC 21
2.2.1.2 MFCC 21
2.2.1.3 GFCC 22
2.2.2 Análise Não Linear 23
2.2.2.1 MQRs 23
2.2.3 Análise baseada em não estacionariedade 25
2.2.3.1 INS 26
2.2.3.2 Segmentação Adaptativa 28

3 METODOLOGIA 30
3.1 Base de Dados 30
3.2 Estudos de Caso 30
3.2.1 Estudo de Caso 1: INS como característica acústica 31
3.2.2 Estudo de Caso 2: INS como base para extração de características acústicas 31
3.3 Etapa de Classificação 32

4 RESULTADOS 34
4.1 Resultados do Estudo de Caso 1 34
4.1.1 Classificação com as escalas do INS individualmente 35
4.1.2 Classificação com as escalas do INS combinadas 36
4.1.3 Discussão dos resultados do estudo de caso 1 37

4.2 Resultados do Estudo de Caso 2 37
4.2.1 Classificação com as medidas de análise linear de produção da fala . . . 38
4.2.2 Classificação com as medidas de análise não linear de produção da fala . 39
4.2.3 Discussão dos resultados do estudo de caso 2 42

5 CONSIDERAÇÕES FINAIS 43
5.1 Publicações desta Pesquisa 44

REFERÊNCIAS 46

1 Introdução

O sinal de voz é um processo resultante de um complexo sistema de produção que envolve fatores neurológicos e fisiológicos (BEHLAU, 2001). Tal forma de onda carrega informações que não estão restritas apenas ao conteúdo linguístico. É possível, também, verificar a identidade do falante e sua saúde (VIEIRA, 2014). Uma vez que a fala é considerada como sendo um dos principais meios de comunicação dos seres humanos, além do seu uso intenso em diversas profissões, o seu estudo possui relevância na área de processamento de sinais.

Diversas aplicações envolvendo processamento digital de sinais de voz têm sido estudadas ao longo das últimas décadas (BENZEGHIBA et al., 2007; TACHBELIE; ABATE; BESACIER, 2014; KINNUNEN; LI, 2010; BAI; ZHANG, 2021; TAVARES; COELHO, 2015; VIEIRA et al., 2018). Sistemas de reconhecimento de fala, por exemplo, são utilizados nos dias atuais em smartphones para facilitar a interface homem-máquina e otimizar processos como pesquisa e digitação de longos textos (CHERN et al., 2017; ISMAIL; ABDLERAZEK; EL-HENAWY, 2020). Outras propostas comuns na literatura estão imersas em contextos como reconhecimento de locutor (KINNUNEN; LI, 2010; VENTURINI; ZAO; COELHO, 2014), reconhecimento de emoções (WANG et al., 2015; VIEIRA; COELHO; ASSIS, 2020) e reconhecimento de distúrbios vocais (VIEIRA et al., 2018; AKBARI; ARJMANDI, 2014). Essa última aplicação é importante na prática clínica de profissionais de saúde, como fonoaudiólogos e demais especialistas da área, pois permite a triagem, o diagnóstico e o acompanhamento de pacientes de forma presencial ou remota (por meio de mecanismos de Telessaúde) (VIEIRA, 2014; RANGARATHNAM et al., 2015).

A análise de distúrbios vocais pode estar relacionada a aspectos patológicos ou hiperfuncionais. Em relação às patologias, essas podem ter origem neurológica ou fisiológica, fatores estes que se subdividem em diferentes tipos de condições patológicas (COSTA, 2008). No contexto dos aspectos hiperfuncionais, as disfonias podem ser analisadas em relação ao grau do desvio fonatório (VIEIRA, 2014), ou, ainda, em relação a fatores como tensão, rugosidade e soprosidade na emissão sonora (QUEIROZ, 2018). A separação entre um sinal de voz considerado saudável e um sinal de voz considerado disfônico (ou mesmo a caracterização das diferentes disfonias) depende da robustez da medida acústica empregada para capturar informações da forma de onda.

Na literatura são encontrados trabalhos que propõem medidas acústicas baseadas em um modelo linear de produção da fala (COSTA, 2008; LOPES et al., 2017) e trabalhos que propõem medidas baseadas em um modelo não linear (VIEIRA et al., 2018; COSTA, 2012; JIANG; ZHANG; MCGILLIGAN, 2006). O modelo linear considera o sistema de

produção vocal como sendo um sistema fonte-filtro (FANT, 1981). Entre as medidas utilizadas baseando-se neste modelo estão a Frequência Fundamental (F0), *Jitter*, *Shimmer*, Coeficientes LPC (*Linear Predictive Coding*) e Coeficientes MFCC (*Mel-Frequency Cepstrum Coefficients*) (COSTA, 2008; LOPES et al., 2017; VIEIRA et al., 2013). Por outro lado, o modelo não linear considera a produção da fala como um sistema dinâmico não linear, sujeito a comportamento caótico (VIEIRA, 2014). Algumas das medidas não lineares comumente empregadas são a dimensão de correlação, o expoente de Lyapunov e as medidas de quantificação de recorrência (COSTA, 2012).

Em geral, as medidas da análise linear são extraídas em trechos considerados estacionários do sinal de voz, compreendidos entre 16 ms e 40 ms (COSTA, 2008). Em contrapartida, características não lineares, a exemplo das medidas de quantificação de recorrência, não possuem o pré-requisito da estacionariedade, o que permite a análise em trechos de duração superior a 40 ms (COSTA, 2012). Assim, a escolha do tamanho do trecho do sinal, do qual são extraídas as medidas acústicas, tem influência do tipo de análise empregada nele. Quanto mais adequada for a análise, mais robusto tende a ser o sistema homem-máquina de reconhecimento de disfonias.

1.1 Justificativa

Apesar de existirem diversas pesquisas que investiguem a presença de padrões acústicos nas disfonias, ainda não há um consenso sobre qual medida acústica pode ser utilizada universalmente para discriminar sinais considerados saudáveis de sinais oriundos de distúrbios patológicos ou hiperfuncionais. Além disso, há trabalhos que levam em consideração a combinação de características para atingir um melhor desempenho de classificação (CHERN et al., 2017; AKBARI; ARJMANDI, 2014; COSTA, 2012).

A escolha de medidas acústicas confiáveis que representem cada tipo de disfonia pode se tornar uma tarefa difícil, uma vez que depende de vários fatores como o tipo e o grau da lesão, a severidade dos efeitos causados pela desordem vocal e a quantidade de ruído presente no sinal de voz analisado (VIEIRA, 2014; LOPES et al., 2017; COSTA, 2012). Embora diversas medidas aproximem a voz a processos estacionários por meio de métodos de segmentação e janelamento, a natureza deste tipo de sinal é não estacionária.

A consideração de que sinais de voz são estacionários em curtos intervalos de tempo (RABINER; SCHAFER, 2007) é originalmente concebida para sinais diagnosticados como normais (ou saudáveis). Os estudos que avaliam desordens vocais fazem a extração de medidas a curto intervalo de tempo por padrão para todas as classes, sem se preocupar se os sinais disfônicos realmente podem ser considerados estacionários no mesmo intervalo de tempo que os sinais saudáveis. Estudos mostraram que a condição patológica do sistema de produção vocal pode introduzir ruído na emissão sonora (VIEIRA, 2014; COSTA, 2012). Caso fatores como este influenciem em um alto grau de

não estacionariedade nos trechos do sinal, inicialmente considerados estacionários, o uso de medidas acústicas clássicas não seria possível do ponto de vista teórico, uma vez que quebra o requisito da estacionariedade.

Em um contexto multidisciplinar, a metodologia de extração supracitada é comum na área da Fonoaudiologia, em que muitas pesquisas têm procurado caracterizar acusticamente os distúrbios da voz usando diferentes técnicas (LOPES et al., 2017; GOY et al., 2013; KNIGHT; AUSTIN, 2020; MUNIER et al., 2020). A análise acústica é o auxílio computacional do Fonoaudiólogo para triagem, diagnóstico e acompanhamento. Com as novas demandas tecnológicas, o uso do Teleatendimento (STROHL et al., 2020) pode se tornar comum em um futuro próximo, ocasionando a adição de ruído[1] nos sinais e, por consequência, tornar menos fidedigna a extração de medidas acústicas tradicionais. A proposta de uma ferramenta computacional que forneça um pré-processamento adequado, observando se os trechos curtos podem realmente ser considerados estacionários e realizando uma análise acústica mais apropriada, pode ser um grande avanço contra a falta de consenso existente a respeito da técnica de análise acústica mais adequada para auxílio clínico.

Por outro lado, apesar de algumas características serem extraídas de trechos de voz considerados não estacionários, elas obtêm diretamente a informação desses sinais sem necessariamente avaliar o grau de não estacionariedade deles. Em um trabalho comparando o desempenho a curto intervalo de tempo com o desempenho a longo intervalo de tempo das medidas de quantificação de recorrência (VIEIRA et al., 2014), observou-se que o uso de segmentos não estacionários (longo intervalo de tempo) proporciona uma maior discriminação entre sinais saudáveis e sinais patológicos. Isto é um indício de que os distúrbios vocais podem ser mais perceptíveis quando a não estacionariedade é levada em consideração.

Outro cenário em que a análise da não estacionariedade dos sinais de voz pode ser importante é no pré-processamento para a aplicação de extração de características, com posterior classificação utilizando aprendizado profundo (*Deep Learning*). Neste contexto costuma-se utilizar redes neurais profundas (DNN – *Deep Neural Networks*), as quais, em geral, precisam de muitos dados para fornecer resultados confiáveis (DIAS, 2020). O uso de um teste de não estacionariedade na etapa de segmentação do sinal pode ajudar a diminuir o tamanho do trecho sem perder a não estacionariedade. Isto tem grande utilidade quando: 1) o estudo não possui uma base de dados com muitos sinais; 2) os sinais têm duração pequena; 3) a técnica de extração de características não tem pré-requisito de estacionariedade.

Pesquisas recentes têm empregado uma medida chamada Índice de Não Estacionariedade (INS – *Index of Non-Stationarity*) (BORGNAT et al., 2010), que realiza o teste de

[1] ruído típico de canais de comunicação.

não estacionariedade e ainda fornece o seu grau, a fim de observar como sinais acústicos de diferentes classes se comportam sob este ponto de vista (TAVARES; COELHO, 2015; VIEIRA; COELHO; ASSIS, 2020). Em um trabalho observando a influência acústica de sinais ruidosos na inteligibilidade da fala (TAVARES; COELHO, 2015) foram observadas diferenças do INS entre sinais de ruído de balbucio, motosserra, fábrica e britadeira. Outro trabalho (VIEIRA; COELHO; ASSIS, 2020) apresentou diferenças de INS entre sinais de voz com variações emocionais e, ainda, empregou esta medida no processo de classificação, aprimorando o desempenho do classificador.

Na literatura, existem algumas outras propostas, além do INS, para análise de não estacionariedade de sinais em diferentes aplicações (CAPPONI et al., 2017; MARTIN; MAILHES, 2009). Porém, não foi encontrada nenhuma aplicação envolvendo distúrbios vocais no que diz respeito a testar a estacionariedade de diferentes trechos de sinais disfônicos ou, ainda, na caracterização do grau de não estacionariedade das disfonias como uma medida acústica.

Uma vez que alguns trabalhos apontaram que algumas classes de sinais acústicos podem apresentar variações não estacionárias, surgem as perguntas norteadoras deste estudo:

1. Sinais de voz afetados por disfonias apresentam diferentes graus de não estacionariedade?

2. Trechos de sinais de voz considerados estacionários, quando afetados por patologias (de origem orgânica ou neurológica) permanecem estacionários?

A resposta destas perguntas pode influenciar na escolha das técnicas empregadas e das medidas acústicas extraídas dos sinais, levando a resultados diferentes de desempenho de classificação. Neste contexto, uma das hipóteses motivadoras deste trabalho é que o grau de não estacionariedade pode ser uma característica relevante na análise e classificação dos diferentes distúrbios vocais. Outra hipótese é de que pode haver um tamanho de segmento, determinado de forma adaptativa, em que as propriedades acústicas não estacionárias das disfonias são enfatizadas.

Como contribuições deste trabalho estão a proposta de novas rotinas de extração de características no ponto de vista da não estacionaridade e, ainda, a proposta do INS como medida acústica para discriminação de disfonias. A escolha de características com alto grau de confiabilidade impactará no desempenho dos sistemas de classificação e avaliação da qualidade vocal, melhoramento na terapia vocal, tão importante para a comunicação, especialmente em profissionais que utilizam a voz em suas profissões, a exemplo de professores, cantores, apresentadores de rádio e televisão, operadores de telemarketing, dubladores, entre outros.

1.2 Objetivos

1.2.1 Objetivo Geral

Analisar o grau de não estacionariedade de sinais de voz e seu efeito na classificação de desordens vocais.

1.2.2 Objetivos Específicos

- Extrair o índice de não estacionariedade de sinais de voz provenientes de laringes saudáveis e patológicas;
- Avaliar o índice de não estacionariedade dos sinais de voz como medida acústica para classificação de desordens vocais;
- Verificar a influência do índice de não estacionaridade na extração de características baseadas nos modelos linear e não linear de produção da fala para classificação de desordens vocais.

1.3 Organização deste Trabalho

No Capítulo 2 é apresentada a fundamentação teórica deste trabalho, com uma breve descrição das patologias laríngeas consideradas neste estudo e as abordagens de análise acústica empregadas: modelos linear e não de produção da fala, bem como o índice de não estacionariedade. as características acústicas. A metodologia e os cenários experimentais, com dois estudos de caso, são apresentados no Capítulo 3. Os resultados de classificação, considerando dois diferentes classificadores, são apresentados no Capítulo 4, em que são apresentadas discussões sobre cada estudo de caso. Finalmente, no Capítulo 5, são apresentadas as considerações finais deste trabalho, em que também são colocadas sugestões para trabalhos futuros.

2 Fundamentação Teórica

A avaliação de distúrbios da voz (triagem, o diagnóstico ou o acompanhamento de tratamento) no contexto de patologias laríngeas pode contemplar pelo menos uma das metodologias descritas como segue (COLTON; CASPER; LEONARD, 2006):

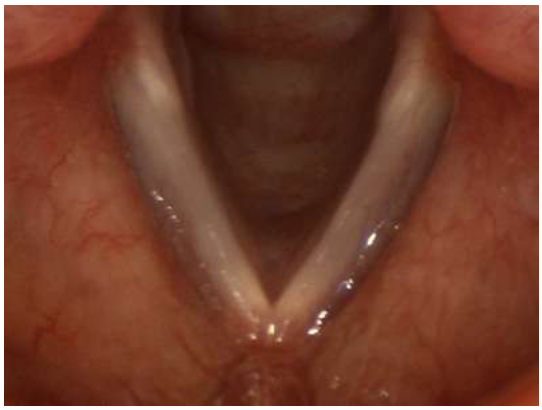
- Autoavaliação
  - como o próprio nome sugere, está relacionada à percepção do paciente em relação ao seu problema vocal. A partir de alguns sintomas, o paciente tenta identificar as possiveis causas do distúrbio da voz.
- Análise otorrinolaringológica
  - geralmente é constituída por um ou mais procedimentos invasivos que podem causar desconforto ao paciente, tais como a laringoscopia e a estroboscopia.
- Avaliação aerodinâmica
  - técnica clínica utilizada para medir a vazão e a pressão do ar com base nas mudanças do fluxo ao passar pela laringe e pelo trato vocal. Esse método permite identificar uma adução inadequada das pregas vocais.
- Análise perceptivo-auditiva da voz
  - conduzida por um especialista capacitado (em geral um(a) profissional da Fonoaudiologia) para ouvir e detectar características no sinal de voz que possam revelar possíveis alterações na qualidade do sistema de produção vocal.
- Análise acústica da voz
  - utiliza técnicas de processamento digital de sinais para extrair características a partir da forma de onda do sinal vocal. Uma vez digitalizada, essa forma de onda se transforma em uma série temporal que contém informações importantes sobre o sistema de produção vocal.

Este trabalho está situado no contexto da análise acústica da voz. Nas Seções seguintes, são apresentadas as patologias consideradas no estudo e as técnicas de processamento digital de sinais empregadas. Entre essas abordagens, é apresentada a análise baseada em não estacionariedade dos sinais de voz, que compreende o escopo principal desta pesquisa.

2.1 Patologias na Laringe

Ao se investigar desordens vocais associadas a patologias na laringe, é importante ressaltar que a região mais afetada nesse órgão é conhecida como pregas vocais. As pregas vocais são estruturas multilaminadas, constituídas por dobras de músculos e mucosas que se estendem horizontalmente na laringe (ZITTA, 2010). Um exemplo de pregas vocais saudáveis é apresentado na Figura 1.

Figura 1 – Pregas vocais saudáveis.



Fonte: Sulica (2013), apud Vieira (2014).

Patologias laríngeas podem ser consequência de diversos fatores, tais como de uso inadequado da voz, algum tipo de abuso vocal, ou mesmo algum distúrbio neurológico que afete o fluxo aéreo e a pressão aérea adequados para a fonação (BEHLAU, 2001). Neste trabalho, os experimentos realizados consideram três patologias diferentes: paralisia nas pregas vocaias, edema de Reinke e nódulos, as quais são brevemente descritas a seguir.

2.1.1 Paralisia

A paralisia se refere à incapacidade de movimento das pregas vocais, resultante de lesão ou disfunção do nervo laríngeo recorrente[1], que é o principal responsável pela sua mobilidade. Essa condição provoca uma redução na espessura das pregas vocais (COLTON; CASPER; LEONARD, 2006). A paralisia nas pregas vocais pode ser classificada como unilateral ou bilateral. Na paralisia unilateral, a prega vocal afetada não consegue se mover em direção à linha média, o que compromete o fechamento da glote. Por outro lado, na paralisia bilateral, ambos os lados das pregas vocais apresentam uma espessura reduzida e não conseguem se mover completamente para a linha média. Exemplos desses dois tipos de paralisia são apresentados na Figura 2.

[1] O nervo laríngeo recorrente se extende do cérebro para baixo do pescoço e no peito antes de virar para cima de volta para a laringe (SULICA, 2013) (apud (VIEIRA, 2014))

Figura 2 – Paralisia nas pregas vocais: (a) unilateral direita e (b) bilateral.

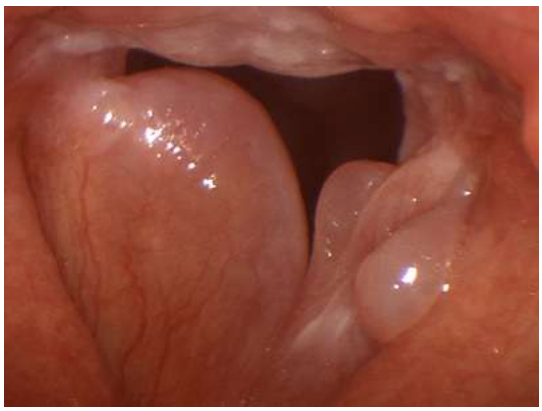
(a) (b)

Fonte: Sulica (2013), apud Vieira (2014).

2.1.2 Edema

O edema, mais conhecido como edema de Reinke, é caracterizado pelo acúmulo de líquido na camada superficial da lâmina própria das pregas vocais, conhecida como espaço de Reinke, levando a um aumento e inchaço das pregas vocais (Figura 2.6). As lesões associadas a essa condição costumam ser bilaterais e apresentam uma assimetria em seu tamanho (KUHL, 1982). Um exemplo de caso de edema de Reinke é apesentado na Figura 3.

Figura 3 – Pregas vocais com edema de Reinke.



Fonte: Sulica (2013), apud Vieira (2014).

2.1.3 Nódulos

Os nódulos vocais são lesões benignas que surgem na borda livre e na superfície inferior das pregas vocais, frequentemente resultantes de abuso vocal. Seus principais sintomas incluem rugosidade e soprosidade na voz. Esses nódulos são sempre bilaterais e podem variar em tamanho, simetria e coloração. Inicialmente, podem ser mais evidentes de um lado, o que pode levar a confusões com pólipos. Com a persistência do trauma, o tecido afetado tende a se tornar mais rígido (COLTON; CASPER; LEONARD, 2006;

BENJAMIN; FIGUEIREDO, 2000). Um exemplo de caso de nódulos é apesentado na Figura 4.

Figura 4 – Pregas vocais com nódulos.

Fonte: Sulica (2013), apud Vieira (2014).

2.2 Análise Acústica de Vozes Saudáveis e Patológicas

A análise acústica envolve métodos computacionais não invasivos, utilizados para extrair características do sinal de voz que são inerentes ao sistema de produção da fala. De maneira geral, essa análise pode ser realizada em um ambiente acusticamente controlado, utilizando um microfone para captar o sinal e um computador processá-lo (COLTON; CASPER; LEONARD, 2006).

A adoção de técnicas de análise acústica não tem como objetivo substituir os exames laringoscópicos. Seu principal objetivo é o auxílio à prática clínica. Considerando que muitos pacientes acham esses exames invasivos e, por vezes, recusam-se a realizá-los, a análise acústica pode ser uma alternativa para reduzir a necessidade de exames laringoscópicos. Além disso, a análise acústica pode ser empregada na terapia vocal para pessoas com dificuldades na fala e por profissionais que utilizam a voz (GODINO-LLORENTE; GOMEZ-VILDA; BLANCO-VELASCO, 2006; COSTA, 2008).

Em relação às técnicas empregadas na análise acústica, duas abordagens são comumente encontradas na literatura: análise linear e análise não linear, que correspondem a diferentes pontos de vista do sistema de produção vocal. Ambas as abordagens são descritas brevemente a seguir. Além disso, é apresentada a análise baseada em não estacionariedade, que compreende o escopo fundamental deste estudo.

2.2.1 Análise Linear

Este tipo de abordagem é baseada no modelo linear de produção da fala, conhecido como Teoria fonte-filtro (FANT, 1981), em que a fonte é o resultado acústico da vibração

da prega vocal, enquanto a função de transferência do trato vocal (filtro) fornece o formato espectral da fala (RABINER; SCHAFER, 1978). Neste trabalho são empregadas, do modelo linear, características acústicas bem estabelecidas na literatura, a saber, LPC, MFCC e GFCC, as quais são brevemente descritas as seguir.

2.2.1.1 LPC

A abordagem LPC tem como premissa a estimação de cada amostra de voz com base em uma combinação linear de p amostras anteriores. Um valor p maior representa um modelo mais preciso. O princípio básico da análise LPC é determinar um conjunto de coeficientes preditores, $\alpha(k)$, diretamente do sinal de fala, para obter uma estimativa adequada das propriedades espectrais dos sinais (OROZCO-ARROYAVE et al., 2015; RABINER; SCHAFER, 1978). Assim, o trato vocal pode ser modelado como um sistema cuja função de transferência, $H(z)$, é dada por:

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} \alpha(k) z^{-k}}, \tag{2.1}$$

em que G é um fator de ganho, que é ajustado para controlar a intensidade de excitação, e p é a ordem do preditor.

No presente estudo, utiliza-se uma quantidade de 29 coeficientes LPC. Tal valor é o resultado da soma dada por $\eta + 4$, em que η é a taxa de amostragem em kHz (no caso desta pesquisa, 25 kHz). Assim, entende-se que é estabelecido um compromisso entre a complexidade computacional e a precisão do modelo de predição, visto que são utilizados 25 polos para representar o trato vocal e, adicionalmente, 4 polos para representar a fonte de excitação (RABINER; SCHAFER, 1978).

2.2.1.2 MFCC

As características MFCC foram propostas e são amplamente utilizadas em tarefas de reconhecimento de fala e de locutor (WANG et al., 2011; WU; FALK; CHAN, 2011; BANSAL; IMAM; BHARTI, 2015; CHOWDHURY; ROSS, 2019), reconhecimento de emoções (FAHAD et al., 2021) e avaliação de distúrbios de voz (TIRRONEN; KADIRI; ALKU, 2022), e estão relacionados à percepção do ouvido humano. A percepção das frequências de tons puros não está em uma escala linear, o que leva ao desenvolvimento da chamada escala mel. Essa escala, por sua vez, aproxima computacionalmente a percepção auditiva (O'SHAUGHNESSY, 1987). Como referência, 1 kHz é equivalente a 1.000 mels, e a transformação de uma frequência f para a escala mel é descrita da seguinte forma:

$$Mel(f) = 1127 \ln\left(1 + \frac{f}{700}\right). \tag{2.2}$$

Na análise mel-cepstral, bancos de filtros são usados para simular a resposta de frequência da membrana basilar do ouvido humano. No caso de sinais de voz, que em muitas aplicações são analisados até aproximadamente uma frequência de 4 kHz, 20 filtros triangulares com uma largura de banda de 300 mel, espaçados 150 mel um do outro, são comumente usados (O'SHAUGHNESSY, 1987).

Para a extração de atributos MFCC, após a etapa de pré-processamento com segmentação de fala, as amostras de cada quadro são convertidas para o domínio de frequência por meio da transformada rápida de Fourier (FFT – *Fast Fourier Transform*), a partir da qual a energia é calculada. O sinal transformado passa então por um banco de filtros em escala mel.

O conjunto de coeficientes MFCC (c_j) é obtido de acordo com (DAVIS; MERMELSTEIN, 1980):

$$c_j = \sum_{k=1}^{F} (\log\ S_k)\ \cos\left[\frac{\pi j}{F}\left(k - \frac{1}{2}\right)\right], \tag{2.3}$$

para $j = 1, 2, ..., D$, em que D é o número de coeficientes, S_k é a energia do k-ésimo filtro, e F é a quantidade de filtros na escala Mel. Neste trabalho, considera-se $D = 12$.

2.2.1.3 GFCC

Similarmente às características MFCC, os atributos GFCC foram propostos para tarefas de reconhecimento de locutor (SHAO; SRINIVASAN; WANG, 2007). Além disso, eles foram usados na última década no contexto de reconhecimento de emoções (SHARMA; SINGH, 2015; MOHANTY, 2016). A ideia geral para obter os coeficientes GFCC também é baseada em uma aproximação computacional do sistema auditivo. Neste caso, filtros *Gammatone* são usados, que estão relacionados ao comportamento da cóclea humana (PATTERSON; HOLDSWORTH; ALLERHAND, 1992).

A resposta ao impulso de um filtro *Gammatone*, $g(t)$, é o produto da função de distribuição Gamma e um sinal senoidal centrado na frequência f_c, de acordo com (SCHLUTER et al., 2007):

$$g(t) = Kt^{(n-1)}e^{-2\pi Bt}\cos(2\pi f_c + \varphi), \quad t > 0, \tag{2.4}$$

em que K é o fator de ganho de amplitude, n é a ordem do filtro, f_c é a frequência central em Hz, φ é a fase e B está relacionado à duração da resposta ao impulso. A extração de características GFCC é semelhante à do MFCC até a estimativa da FFT. Após esta etapa, um banco de filtros *Gammatone* é aplicado ao sinal, seguido pela transformada discreta do cosseno (DCT – *discrete cosine transform*):

$$G_m = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} (\log Y_n) \cos\left[\frac{\pi m}{N}\left(n - \frac{1}{2}\right)\right], \tag{2.5}$$

em que $1 \leq m \leq M$, e Y_n é a n-ésima componente de energia espectral. N é a quantidade de filtros *Gammatone*, e M é o número de coeficientes GFCC. Os trabalhos encontrados na literatura apresentam diferentes valores para a quantidade de coeficientes. Neste trabalho, utiliza-se 22 coeficientes nos experimentos, tal como utilizado em Vieira, Coelho e Assis (2018).

2.2.2 Análise Não Linear

Nessa abordagem, o sinal de voz é considerado como sendo a saída de um sistema dinâmico não linear, que pode sofrer com interações em vórtices do fluxo de ar induzidas por mudanças na laringe (KUMAR; MULLICK, 1996; JIANG; ZHANG; MCGILLIGAN, 2006). Nas últimas duas décadas, aproximadamente, um conjunto de características não lineares foi estudado para avaliação de distúrbios de voz (VIEIRA et al., 2018), chamadas medidas de quantificação de recorrência (MQRs).

2.2.2.1 MQRs

As MQRs (ZBILUT; WEBBER-JR, 1992; MARWAN et al., 2002; MARWAN, 2003) foram propostas para a análise objetiva de sistemas dinâmicos por meio de uma técnica conhecida como gráficos de recorrência (RP) (ECKMANN; KAMPHORST; RUELLE, 1987). Por definição, um RP é uma representação bidimensional de um sistema multidimensional dinâmico (WEBBER-JR; ZBILUT, 2005; MARWAN; WEBBER-JR, 2015). Exemplos de RPs para sinais saudáveis e patológicos são apresentados na Figura 5. Os RPs para sinais saudáveis têm linhas diagonais sólidas, enquanto a presença de patologia quebra essas estruturas formando outras, e é apresentada como uma combinação de características verticais/horizontais.

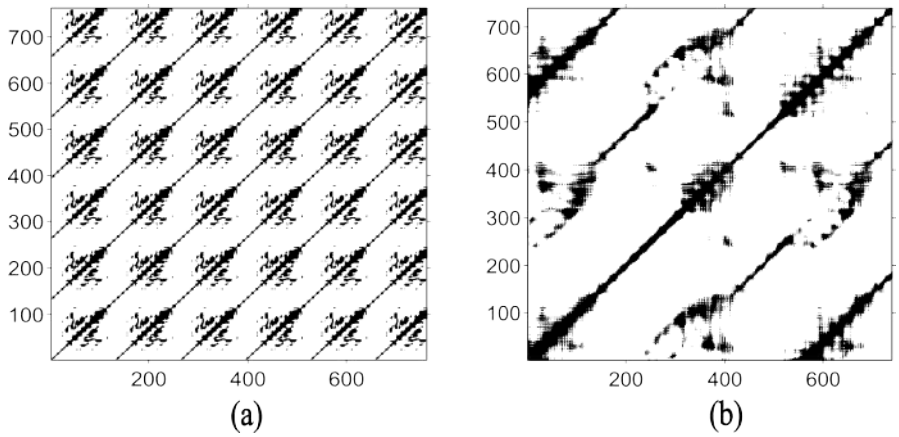
A formulação matemática de um RP é dada por (MARWAN, 2003):

$$\mathcal{R}_{i,j}^{m,\varepsilon} = \Theta(\varepsilon - \|\vec{\xi_i} - \vec{\xi_j}\|), \quad \vec{\xi_i} \in R^m, \quad i, j = 1 \ldots N, \tag{2.6}$$

em que:

- N é a quantidade de estados do sistema dinâmico, $\vec{\xi_i}$;
- ε é o raio de vizinhança centrado no estado $\vec{\xi_i}$;
- $\|\cdot\|$ é a norma Euclidiana;
- $\Theta(\cdot)$ é uma função degrau unitário;
- m é a dimensão de imersão.

Figura 5 – Gráficos de recorrência obtidos de: (a) um sinal de voz de uma laringe saudável; (b) um sinal de voz de uma laringe afetada por nódulos.



Fonte: o autor.

Para a formação de um RP, inicialmente pode-se obter uma representação da evolução do sistema dinâmico por mieo do espaço de fases (WEBBER-JR; ZBILUT, 2005). Tal representação é, em geral, realizada a partir do método de Takens (1981), em que as dimensões do espaço de fase são obtidas como versões defasadas do sinal original. Uma vez obtida a evolução do espaço de fase para o sistema dinâmico, a distância entre os estados $\vec{\xi}_i$ e $\vec{\xi}_j$ é calculada. O limiar ε define se cada estado $\vec{\xi}_j$ é recorrente em $\vec{\xi}_i$ ou não. A função degrau unitário, $\Theta(\cdot)$, mapeia os estados recorrentes para uma matriz $N \times N$ RP. Assim, um ponto preto é marcado quando o estado correspondente é considerado recorrente, e um ponto branco é marcado quando o estado não é recorrente. O número N de estados $\vec{\xi}$, a partir do método de Takens , é dado por:

$$N = T_s - (m-1)\tau, \tag{2.7}$$

em que T_s é o número de amostras do sinal (em outras palavras, o comprimento da série temporal), e τ representa o atraso para a reconstrução do espaço de fase. Como m e τ dependem de cada sinal, eles também podem ser considerados como características acústicas.

A partir dos RPs, a análise objetiva é realizada com as MQRs. A taxa de recorrência (REC) (WEBBER-JR; ZBILUT, 2005) foi a primeira medida estabelecida, pois simplesmente quantifica no RP a razão entre a quantidade de pontos de recorrência e o número de estados. Outras medidas foram escritas em Marwan (2003). Além disso, há medidas que têm sido implementadas em versões de software[2] para extração de MQRs.

Neste trabalho, além de τ e m, outras 13 MQRs são empregadas. Para a extração dessas medidas, foi considerado um percentual de taxa de recorrência de 1%, tal como

[2] https://tocsy.pik-potsdam.de/CRPtoolbox/?q=fnc_crqa [último acesso em 31/08/2024].

descrito em Vieira (2014). Essas medidas podem ser divididas por categorias (MARWAN, 2003):

- Medidas baseadas em estruturas diagonais:
  - determinismo (DET), entropia de Shannon das linhas diagonais ($ENTR_L$), comprimento médio das linhas diagonais (L_{med}), comprimento máximo das linhas diagonais (L_{max}), divergência (DIV), e razão entre DET and REC ($RATIO$).
- Medidas baseadas em estruturas verticais/horizontais:
  - laminaridade (LAM), tempo de aprisionamento ou comprimento médio das linhas verticais (TT), comprimento máximo das linhas verticais (V_{max}) e entropia de Shannon das linhas verticais ($ENTR_V$).
- Outras medidas:
  - tempo de recorrência do tipo 1 (T_1), tempo de recorrência do tipo 2 (T_2), e razão entre LAM and DET ($LAM/DET\ RATIO$).

2.2.3 Análise baseada em não estacionariedade

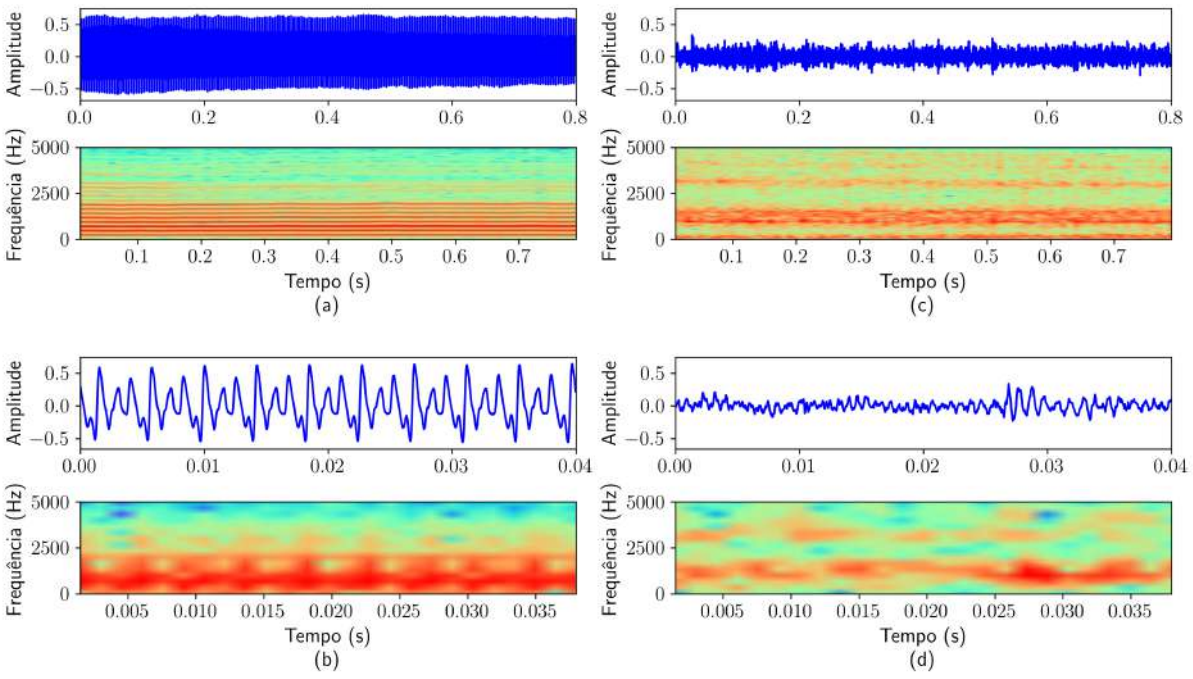
As abordagens de análise acústica mencionadas até então representam diferentes pontos de vista relacionados à produção vocal. Independente das abordagens supracitadas, é possível investigar como o sinal acústico é processado a fim de se obter informações relevantes sobre o estado do sistema de produção vocal. Nesse contexto, a análise relacionada à estacionaridade pode ser levada em consideração, uma vez que a estacionariedade é um aspecto relevante em muitas aplicações de processamento de sinais (BORGNAT et al., 2010). Em Vieira (2018), foi investigado como as variações acústicas não estacionárias no sinal de voz são afetadas por estados emocionais.

De maneira geral, a análise baseada em não estacionaridade dos sinais de voz pode ser conduzida de duas maneiras: 1) objetivando obter uma medida que caracterize a não estacionariedade; e 2) objetivando auxiliar a extração de caractrísticas baseadas nos modelos linear e não linear de produção vocal. Como medida da não estacionariedade dos sinais de voz, utiliza-se o índice de não estacionariedade (INS – *index of non-stationarity*) neste trabalho. Além da descrição dessa métrica, também é discutido a seguir sobre segmentação adaptativa, que é ua maneira de utilizar o INS para extração de outras características acústicas.

2.2.3.1 INS

O INS é uma medida tempo-frequência que analisa de forma objetiva a não estacionariedade de um sinal (BORGNAT et al., 2010). Na proposta original do INS, um sinal é definido estacionário em relação a uma escala de observação se o seu espectro local de tempo curto em diferentes instantes de tempo for estatisticamente similar ao seu espectro global. Na Figura 6 são apresentados espectrogramas globais (800 ms) e locais (40 ms) obtidos de sinais de voz da vogal sustentada /a/ de duas classes: laringe saudável e laringe patológica (paralisia, cuja fonação com menos intensidade resulta em um sinal de menor amplitude). Para ambos os sinais exemplificados, pode-se notar, subjetivamente, diferenças entre os espectrogramas locais e globais. O INS investiga essa diferença de maneira quantitativa. Para verificar se o sinal é não estacionário em uma determinada escala de tempo, é observado o quanto seu espectrograma local (no exemplo, 40 ms) é diferente do espectrograma do sinal completo (no exemplo, 800 ms). Diferentes escalas de tempo são consideradas para o cálculo do espectrograma local. A não estacionariedade é detectada à medida em que o espectrograma local diverge estatisticamente do espectrograma global.

Figura 6 – Sinal de voz da vogal sustentada /a/ e seu espectrograma considerando trechos de 800 ms e 40 ms, respectivamente, para as classes saudável ((a) e (b)) e patológico ((c) e (d)).



Fonte: o autor.

O teste de estacionariedade é realizado pela comparação de componentes espectrais do sinal com referenciais estacionários, chamados *surrogates*, obtidos do próprio sinal (BORGNAT et al., 2010). Para tanto, os espectrogramas do sinal e dos *surrogates* são obtidos por meio da Transformada de Fourier de Tempo Curto (STFT - *Short Time Fourier Transform*). A distância Kullback-Leibler (KL) (BASSEVILLE, 1989) é aplicada para medir a divergência entre o espectro de tempo curto do sinal analisado e seu espectro

global, bem como a diferença entre cada *surrogate* e seu respectivo espectro global.

Para o cálculo do INS, pode-se assumir que $D_n^{(x)}$ representa a divergência do espectrograma do sinal analisado, $x(t)$, em diferentes escalas de tempo $t_n (n = 1, ..., N)$. $D_n^{(s_j)}$, por sua vez, denota a distância KL medida entre os espectrogramas do $j-$ésimo *surrogate* s_j(t) $(n = 1, ..., N;\ j = 1, ..., J)$. Neste trabalho, são consideradas 10 escalas de observação $(N = 10)$ e 50 *surrogates* $(J = 50)$. Então, a variância calculada a partir dos valores de divergência é dada por:

$$\begin{cases} \Theta_0(j) = \text{var}\left(D_n^{(s_j)}\right)_{n=1,...,N.}, \quad j = 1, ..., J. \\ \Theta_1 = \text{var}\left(D_n^{(x)}\right)_{n=1,...,N.} \end{cases} \tag{2.8}$$

Assim, o INS é dado por:

$$\text{INS} := \sqrt{\frac{\Theta_1}{\langle \Theta_0(j) \rangle}}, \tag{2.9}$$

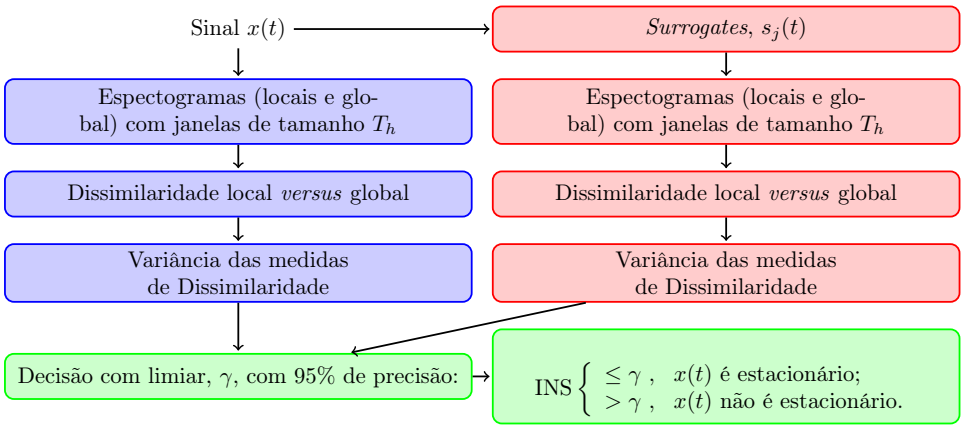
em que $\langle \cdot \rangle$ é o valor médio de $\Theta_0(j)$. Na proposta do INS (BORGNAT et al., 2010), os autores consideram que a distribuição dos valores da distância KL são aproximados por uma distribuição Gamma. Por isso, para cada escala de tempo T_h, um limiar γ, com 95% de precisão, pode ser definido para o teste de estacionariedade. Desta forma, o sinal é considerado não estacionário se o valor de INS estiver acima deste limiar. Ou seja,

$$\text{INS} \begin{cases} \leq \gamma & \text{, sinal estacionário;} \\ > \gamma & \text{, sinal não estacionário.} \end{cases} \tag{2.10}$$

De maneira resumida, a extração do INS é apresentada na Figura 7. Dos sinal original, $x(t)$, são obtidos os *surrogates*. Essas versões estacionárias do sinal original são obtidas por meio da distribuição aleatória da energia do sinal (que originalmente é concentrada em algumas faixas de frequência) em todo o espectro de frequências (BORGNAT et al., 2010). Espectrogramas locais (em diferentes escalas de tempo) e global são obtidos do sinal original e de seus *surrogates*. A medida KL é aplicada para analisar a dissimilaridade, e então a variância dessas medidas em todas as escalas de tempo observadas é calculada. Dos *surrogates*, para cada escala de tempo, é obtida uma distribuição Gamma, na qual é realizado o teste estatístico com a medida da variância dos valores de KL do sinal original, a fim de verificar se, na escala de interesse, o sinal é considerado estacionário.

Exemplos de INS são apresentados na Figura 8. Os valores de INS são obtidos em diferentes escalas de observação, considerando um sinal de voz saudável e três sinais patológicos: paralisia, edema e nódulo. A escala de tempo T_h/T indica a relação entre o

Figura 7 – Fluxograma resumido da extração do INS.



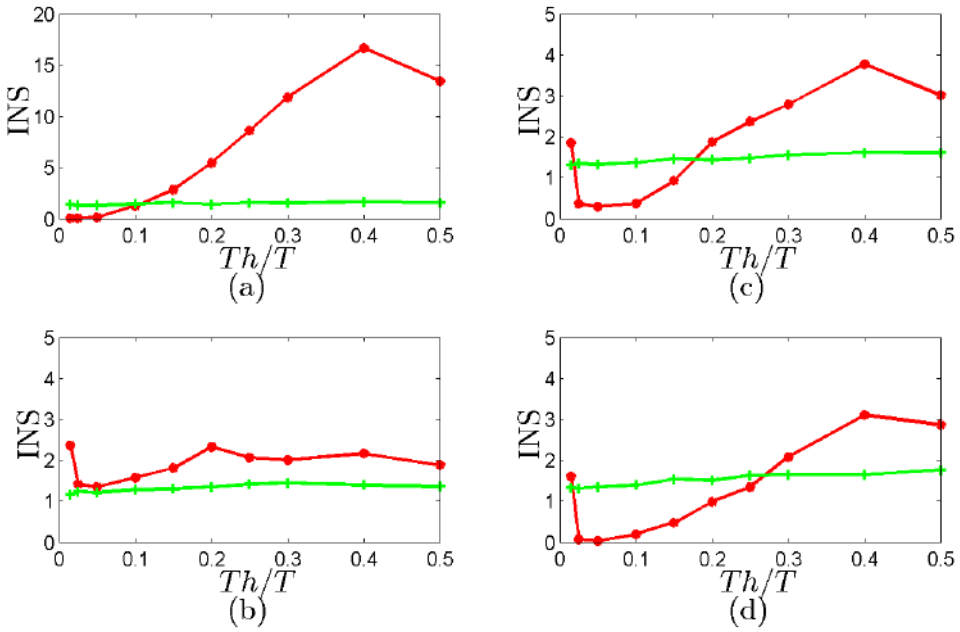
Fonte: o autor.

tamanho adotado para a observação local (T_h) e o tamanho total do sinal ($T = 800$ ms). Os valores de INS são denotados por linhas vermelhas, enquanto as linhas verdes indicam o limiar de estacionariedade. Pode-se notar que essas classes consideradas apresentam diferentes comportamentos baseados em INS, o que responde à primeira questão norteadora desta pesquisa. O comportamento do sinal saudável apresenta um crescimento nos valores de INS após a quarta escala de observação (80 ms, T_h/T=0.1), em que o teste aponta não estacionariedade. O sinal de paralisia mostra-se não estacionário em todas as escalas de observação, o que pode ter sido causado pela severidade da patologia, que é de origem neurológica. Ainda, isto pode responder à segunda questão norteadora: a condição patológica pode levar os sinais a serem não estacionários abaixo de 40 ms As duas patologias de origem orgânica (edema e nódulo) apresentam comportamento semelhante em relação à variação do INS ao longo das escalas. E, comportamento oposto á paralisia também pode ocorrer. Por exemplo, o sinal de nódulos permanece estacionário até 240 ms (T_h/T=0,3). Assim, é razoável supor que o comprimento do quadro baseado em INS seja importante na extração de características. Apesar de serem sinais de vogal sustentada, com pouca variação acústica, diferenças são observadas entre sinais saudáveis e patológicos.

2.2.3.2 Segmentação Adaptativa

A segmentação, em processamento de sinais, envolve a quebra de um sinal de fala contínuo em quadros menores e, sendo assim, menos complexos de processar computacionalmente. Em geral, os objetivos da segmentação são os seguintes: manter a fala como um processo estacionário de sentido amplo (WSS – *wide-sense stationary*) para várias abordagens de extração de características (RABINER; SCHAFER, 1978), otimizar a extração de características e procedimentos de aprendizado de máquina (AGGARWAL et al., 2022) e identificar e isolar os sons ou palavras individuais em um fluxo contínuo de fala (ANU; KARJIGI, 2014).

Figura 8 – INS calculado em diferentes escalas de sinais de voz: (a) saudável; (b) paralisia; (c) edema; (d) nódulo. A linha vermelha indica o valor do INS do sinal original. A linha verde representa o limiar obtido da distribuição Gamma dos *surrogates* para cada escala de tempo.



Fonte: o autor.

A estacionariedade implica que as propriedades estatísticas do sinal permanecem constantes ao longo do tempo. Portanto, essas propriedades não mudam significativamente dentro de cada segmento, permitindo uma análise e processamento mais precisos do sinal, pelo menos quando teorias relacionadas a sinais e sistemas estão sendo consideradas (RABINER; SCHAFER, 1978; O'SHAUGHNESSY, 1987). A segmentação, portanto, é uma técnica de pré-processamento aplicada na análise de sinais de voz, e de maneira tradicional utiliza valores fixos para a extração de características em bases de dados.

A segmentação adaptativa no processamento da voz envolve o ajuste dinâmico do tamanho do quadro de um sinal de acordo com as características associadas ao sistema de produção da fala. Ao contrário das técnicas tradicionais de segmentação de tamanho de quadro fixo, a segmentação adaptativa produz uma análise mais precisa e exata do sinal de fala, garantindo que os segmentos contenham as partes mais relevantes e informativas do sinal em termos de estacionariedade. Além disso, uma vantagem da segmentação adaptativa é sua capacidade de levar em conta a variabilidade e a complexidade do sinal de fala, que pode variar significativamente dependendo das características do falante, como o estado emocional ou a condição patológica nas pregas vocais.

Para selecionar o tamanho de quadro ideal na segmentação adaptativa, a abordagem INS foi aplicada neste estudo para determinar o tamanho (em milissegundos) para o qual acontece um comportamento estacionário.

3 Metodologia

No presente capítulo, é detalhada a metodologia adotada para a realização deste trabalho. Inicialmente, é apresentada a base de dados de voz utilizada. Em seguida, é realizada uma descrição detalhada dos dois estudos de caso que foram conduzidos no contexto da análise baseada em INS. Por fim, é descrita a etapa de classificação, na qual são explicados os métodos utilizaos e as métricas de desempenho adotadas para a validação do estudo.

3.1 Base de Dados

Neste trabalho, foi utilizada a base de voz desenvolvida pelo *Kay Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* (ELEMETRICS, 1994), conhecida como *Disordered Voice Database*, *Model 4337*. Dessa base de dados, que consiste de sinais de vogal sustentada ($/a/$), são analisados 53 casos de vozes saudáveis e 114 vozes afetadas por patologias laríngeas (52 sinais de vozes afetadas por paralisia nas pregas vocais, 44 sinais de vozes afetadas por edema de Reinke e 18 sinais de vozes afetadas por nódulos vocais. A taxa de amostragem considerada na composição da base foi 50 kHz para os sinais saudáveis e 25 kHz para os sinais patológicos. Por convenção, neste trabalho adota-se 25 kHz também para os sinais saudáveis (por meio de *downsampling*) para evitar qualquer influência deste contexto na classificação.

3.2 Estudos de Caso

Para que possa ser realizada uma análise objetiva a respeito da influência da não estacionariedade na produção vocal sob condições patológicas, foram definidos dois estudos de caso:

1. Utilizar o grau de não estacionáriedade como medida acústica para classificação das desordens vocais.

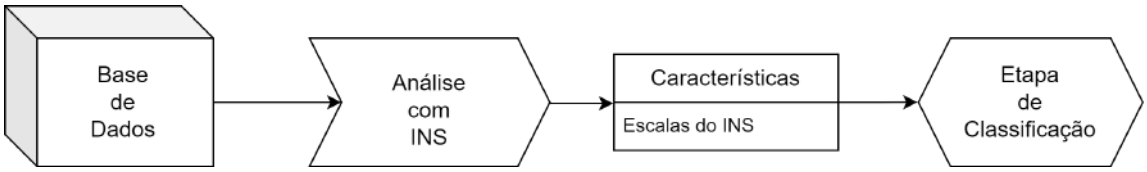
2. Utilizar o grau de não estacionariedade como técnica de pré-processamento em segmentação adaptativa para extração de medidas acústicas clássicas, para a classificação de desordens vocais.

A análise do grau de não estacionariedade é realizada com INS, e esses estudos de caso são abordados com mais detalhe a seguir.

3.2.1 Estudo de Caso 1: INS como característica acústica

Na Figura 9 é apresentado um diagrama de blocos em que é contextualizado o estudo de caso 1. A partir da base de dados de voz, é realizada a análise de não estacionariedade com o INS. Desta análise, as escalas de observação são consideradas como medidas acústicas para serem aplicadas na etapa de classificação.

Figura 9 – Metodologia conduzida no estudo de caso 1, em que as escalas do INS são empregadas como características acústicas.



Fonte: o autor.

De cada sinal da da base de dados, foi utilizado um trecho de 800 ms (a partir do início do sinal) para a extração do INS. Esta escolha tem como razão a padronização das escalas do INS, visto que nem todos os sinais da base de dados têm o mesmo tamanho. Na Tabela 1 são apresentados os valores de fator de escala aplicados ao trecho de 800 ms dos sinais, a fim de se obter o INS de diferentes tamanhos de segmento. Dessa forma, o valor de INS obtido de cada escala (tamanho de segmento) é considerado no processo de classificação como sendo uma característica acústica.

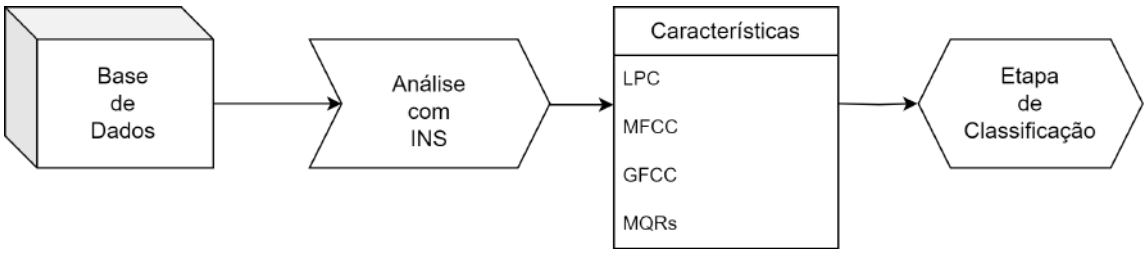
Tabela 1 – Escalas de extração do INS.

Escala	Fator de Escala	Tamanho do Segmento
1	0,015	12 ms
2	0,025	20 ms
3	0,05	40 ms
4	0,1	80 ms
5	0,15	120 ms
6	0,2	160 ms
7	0,25	200 ms
8	0,3	240 ms
9	0,4	320 ms
10	0,5	400 ms

3.2.2 Estudo de Caso 2: INS como base para extração de características acústicas

O diagrama de blocos referente ao estudo de caso 2 é apresentado na Figura 10. Similar ao que acontece no estudo de caso 1, a partir da base de dados de voz, é realizada a análise de não estacionariedade com o INS. Desta análise, no estudo de caso 2 características acústicas são extraídas baseando-se em segmentação adaptativa, de maneira que medidas tradicionais de análise linear (LPC, MFCC e GFCC) e não linear (MQRs) são utilizadas na etapa de classificação.

Figura 10 – Metodologia conduzida no estudo de caso 2, em que as medidas de análise linear (LPC, MFCC e GFCC) e não linear (MQRs) empregadas como características acústicas.



Fonte: o autor.

Para a realização da segmentação adaptativa (ou seja, cada sinal possui seu próprio tamanho de quadro para a extração das características acústicas), foi utilizado um trecho de 800 ms de cada sinal para a estimativa do INS e as dez escalas de observação são as mesmas consideradas no estudo de caso 1 (Tabela 1). Assim, a segmentação adaptativa foi realizada em dois cenários experimentais:

1. Tipo I: Pela menor escala, em que são consierados quadros com o tamanho da menor escala dada como estacionária, com sobreposição de 50%.
2. Tipo II: Pela maior escala, em que são considerados quadros com o tamanho da maior escala dada como estacionária, com sobreposição de 50%.

Para fins comparativos, também foi realizada segmentação tradicional (ou seja, o mesmo tamanho de quadro para todos os sinais da base de dados). Nessa abordagem, foi considerado um tamanho de segmento de 25 ms com 50% de sobreposição.

3.3 Etapa de Classificação

Dois classificadores são empregados para analisar o potencial discriminativo das características acústicas empregadas em ambos os estudos de caso: o LDA (*Linear Discriminant Analysis*) e o QDA (*Quadratic Discriminant Analysis*). Classificadores como esses são úteis para aplicações tabulares com dados de baixa dimensionalidade na entrada (SHWARTZ-ZIV; ARMON, 2022). Tais classificadores são implementados por meio da biblioteca Python scikit-learn[1].

Com o objetivo de dar mais confiabilidade aos resultados, o método *k-fold* de validação cruzada, com $k = 10$, é utilizado, pois tem sido comumente utilizado em classificação de distúrbios vocais (VIEIRA, 2014; LOPES et al., 2017; COSTA, 2012). Para retirar qualquer influência do tamanho amostral nos resultados, no processo de classificação foram selecionados aleatoriamente, entre todas as patologias, 53 sinais, para atingir a mesma quantidade de sinais saudáveis.

[1] https://scikit-learn.org/stable/index.html [último acesso em 22/03/2024].

As medidas de acurácia, sensibilidade e especificidade são utilizadas para analisar o desempenho dos classificadores. Essas medidas estão relacionadas à capacidade de um classificador em diagnosticar uma doença (sensibilidade), diagnosticar um estado saudável (especificidade), bem como medir seu desempenho global (acurácia) (COSTA, 2012), (VIEIRA; COSTA; CORREIA, 2022).

Para analisar o desempenho do classificador empregado neste trabalho, três medidas são utilizadas: acurácia, sensibilidade e especificidade. Essas medidas estão relacionadas à capacidade de um classificador em diagnosticar uma doença em um paciente doente (Verdadeiro Positivo – VP) ou saudável (Falso Positivo – FP), ou, ainda, diagnosticar um estado saudável em um paciente saudável (Verdadeiro Negativo – VN) ou doente (Falso Negativo – FN) (COSTA, 2012).

A acurácia (Ac) representa a taxa global de acerto:

$$Ac = \frac{VP + VN}{VP + VN + FP + FN}. \tag{3.1}$$

A sensibilidade ($Sens$) é a relação entre o número de casos corretamente classificados como presença do distúrbio e a quantidade total de casos com o distúrbio:

$$Sens = \frac{VP}{VP + FN}. \tag{3.2}$$

A especificidade (Esp) mede a relação entre o número de casos corretamente classificados como saudáveis e a quantidade total de casos de estado saudável:

$$Esp = \frac{VN}{VN + FP}. \tag{3.3}$$

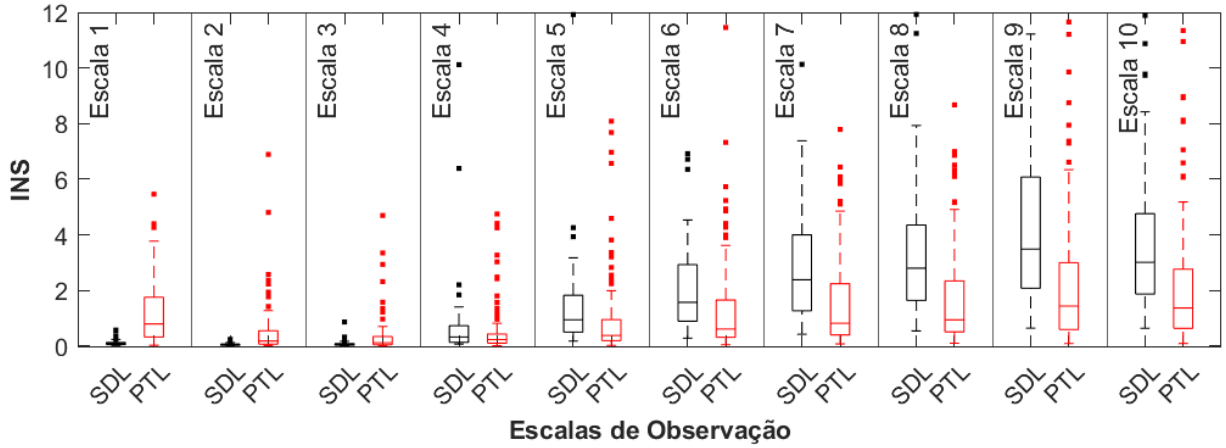
4 Resultados

Nesse capítulo são apresentados os resultados obtidos ao longo do desenvolvimento deste trabalho, considerando os estudos de caso 1 e 2. Além disso, para cada estudo de caso é apresentada uma breve discussão acerca do que foi obtido com os experimentos. Vale ressaltar que, como foi empregada validação cruzada nos experimentos, os resultados de classificação são apresentados em termos de média e desvio padrão, na seguinte configuração: valor da média ± valor do desvio-padrão.

4.1 Resultados do Estudo de Caso 1

Na Figura 11 é apresentada a dispersão das medidas de INS em cada escala de observação, para as classes saudável (SDL) e patológico (PTL). Para as três primeiras escalas de observação, nota-se que os valores de INS para sinais de laringes patológicas são maiores e mais dispersos, com destaque para uma maior diferença observada entre as classes na escala 1. Na escala 4, a dispersão é mais equilibrada entre as classes. Da escala 5 até a escala 10 percebe-se um crescimento e um espalhamento nos valores de INS para a classe saudável. Isto indica que, à medida em que se aumenta a escala de observação (segmento do sinal), sinais de vogal sustentada oriundos de laringes saudáveis possuem um crescimento de INS. Por outro lado, os valores de INS para a classe patológica foram menores que aqueles obtidos para a classe saudável nas maiores escalas de observação. Isto pode ter sido influenciado pela presença de componentes ruidosas (por exemplo, ruído estacionário), que são intrinsecamente adicionadas no processo de produção vocal de laringes patológicas.

Figura 11 – Boxplots obtidos do INS para sinais saudáveis (SDL) e patológicos (PTL) considerando as dez escalas de observação.



Fonte: o autor.

4.1.1 Classificação com as escalas do INS individualmente

Na Tabela 2 são apresentados os valores da classificação individual das medidas de INS, obtidos com classificador LDA. As medidas de INS nas escalas 1, 2 e 10 proporcionam ao classificador LDA uma acurácia média acima de 70%. Como destaque, nota-se que o maior valor de acurácia foi obtido na primeira escala (80%) com um desvio-padrão de 2,64%. Além disso, com a escala 1, o classificador atingiu máximo desempenho (100%) na identificação de patologias (sensibilidade). Por outro lado, o maior valor de especificidade foi obtido com a escala 10, com média 91% e desvio-padrão de 3,02%. Contudo, percebe-se que algumas escalas de observação (1, 2 e 3) têm valores mais elevados de sensibilidade, enquanto outras (escala 4 à 10) contribuem com um incremento nos valores de especificidade.

Tabela 2 – Desempenho da classificação individual com LDA utilizando as medidas de INS de cada escala.

Escala	Ac (%)	Sens (%)	Esp (%)
1	80,00 ± 2,64	100,00 ± 0,00	61,00 ± 5,19
2	70,91 ± 3,26	98,33 ± 1,67	45,33 ± 7,44
3	63,64 ± 3,59	94,67 ± 3,69	34,67 ± 8,92
4	59,09 ± 3,65	32,00 ± 7,03	84,33 ± 3,07
5	64,55 ± 3,44	39,67 ± 6,82	87,67 ± 3,69
6	65,45 ± 3,26	43,33 ± 7,70	85,67 ± 3,48
7	65,45 ± 3,26	43,33 ± 7,70	86,00 ± 3,44
8	69,09 ± 3,37	47,67 ± 7,60	89,33 ± 2,93
9	68,18 ± 3,39	45,00 ± 7,03	89,33 ± 2,93
10	71,82 ± 3,44	51,00 ± 7,03	91,00 ± 3,02

No contexto da classificação individual das escalas de INS empregadas no classificador QDA, tais resultados são apresentados na Tabela 3. Neste cenário, apenas as escalas 1 e 2 proporcionam ao classificador uma acurácia acima de 70%. Como destaque, nota-se que o maior valor de acurácia foi obtido na primeira escala, com mais de 85% de acerto. Outro ponto de destaque, assim como ocorre com o classificador LDA, há escalas em que a sensibilidade é bem maior que a especificidade, e vice-versa.

Tabela 3 – Desempenho da classificação individual com QDA utilizando as medidas de INS de cada escala.

Escala	Ac (%)	Sens (%)	Esp (%)
1	85,45 ± 3,88	96,00 ± 2,67	76,00 ± 6,61
2	78,18 ± 2,42	98,00 ± 2,00	59,67 ± 4,88
3	59,09 ± 3,11	98,00 ± 2,00	23,00 ± 4,51
4	50,91 ± 3,64	17,67 ± 3,14	91,33 ± 5,19
5	55,45 ± 3,16	17,33 ± 3,58	91,67 ± 5,12
6	59,09 ± 3,11	22,00 ± 5,33	93,33 ± 3,69
7	60,00 ± 2,78	22,00 ± 5,33	95,00 ± 2,55
8	60,00 ± 2,78	22,00 ± 5,33	95,00 ± 2,55
9	60,91 ± 1,94	26,00 ± 2,52	93,33 ± 2,72
10	66,36 ± 3,05	37,33 ± 5,09	93,33 ± 2,72

4.1.2 Classificação com as escalas do INS combinadas

Na Tabela 4 são apresentados os resultados da classificação com LDA utilizando a combinação (Combo) das medidas de INS de cada escala. Nota-se que, além da elevação nos valores de acurácia em relação à classificação individual, a combinação das medidas de INS proporciona um maior equilíbrio entre os valores de sensibilidade e especificidade. O maior valor de acurácia (91,82%) é obtido a partir da combinação de nove escalas, com um desvio-padrão de 2,86%. Como foi observado na Figura 11, a escala 4 (a única escala fora da melhor combinação) é a que apresenta valores de INS mais próximos entre as classes envolvidas. Esse resultado indica que a combinação de nove escalas de INS incrementa a acurácia em aproximadamente 11 pontos percentuais (p.p.) em relação ao uso de uma única escala e, ainda proporciona um aumento de aproximadamente 9 p.p. na acurácia em relação à combinação de todas as dez escalas.

Tabela 4 – Desempenho da classificação com LDA por meio das melhores combinações das escalas INS.

Combo	**Ac (%)**	**Sens (%)**	**Esp (%)**	**Escalas**
2 a 2	88,18 ± 3,85	87,00 ± 5,17	89,67 ± 2,83	3 e 5
3 a 3	80,91 ± 4,59	96,67 ± 2,22	66,33 ± 8,26	1, 9 e 10
4 a 4	82,73 ± 4,38	98,33 ± 1,67	67,67 ± 9,34	1, 2, 7 e 10
5 a 5	85,45 ± 2,78	98,00 ± 2,00	73,33 ± 5,92	1, 3, 8, 9 e 10
6 a 6	86,36 ± 2,79	98,33 ± 1,67	75,00 ± 5,40	1, 5, 6, 8, 9 e 10
7 a 7	89,09 ± 2,64	98,00 ± 2,00	81,33 ± 5,24	1, 2, 4, 6, 7, 8 e 9
8 a 8	90,00 ± 2,12	98,00 ± 2,00	82,67 ± 3,54	1, 2, 3, 4, 5, 7, 8 e 9
9 a 9	91,82 ± 2,86	98,00 ± 2,00	85,33 ± 5,73	1, 2, 3, 5, 6, 7, 8, 9 e 10
Todas 10	82,73 ± 3,16	96,00 ± 2,67	70,00 ± 6,65	Todas

Na Tabela 5 são apresentados os resultados da classificação utilizando a combinação das medidas de INS de cada escala, com o classificador QDA. Na combinação 2 a 2, é possível notar que os valores para as medidas de desempenho não diferem muito daqueles obtidos para a primeira escala de observação no cenário da classificação individual. Isto porque a melhor combinação 2 a 2 foi obtida com a escala 1 (melhor individual) e a escala 4 (pior individual). Nenhuma outra combinação 2 a 2 superou este resultado. As demais combinações proporcionaram um incremento nos valores de acurácia em relação à classificação individual, exceto para as combinações 7 a 7, 9 a 9 e todas as 10 juntas. Nota-se, ainda, que a maior taxa de acerto é obtida combinando 5 escalas de observação de INS, cuja acurácia ultrapassa 88%. Assim como nos demais cenários de combinação, neste caso a fusão de características de INS proporcionou um balanceamento nos valores de sensibilidade e especificidade, em detrimento ao comportamento apresentado por estas medidas de desempenho no cenário de classificação individual.

Tabela 5 – Desempenho da classificação com QDA por meio das melhores combinações das escalas INS.

Combo	Ac (%)	Sens (%)	Esp (%)	Escalas
2 a 2	84,55 ± 3,33	90,33 ± 6,09	78,00 ± 6,27	1 e 3
3 a 3	87,27 ± 3,88	94,67 ± 2,73	81,00 ± 6,98	1, 3 e 9
4 a 4	86,36 ± 2,79	92,33 ± 4,33	81,00 ± 5,80	1, 3, 7 e 10
5 a 5	87,27 ± 2,01	94,33 ± 2,90	81,33 ± 3,89	2, 3, 7, 8 e 9
6 a 6	87,27 ± 2,01	92,33 ± 4,33	82,00 ± 5,76	1, 3, 7, 8, 9 e 10
7 a 7	87,27 ± 3,37	94,00 ± 3,06	81,33 ± 5,24	1, 2, 3, 4, 6, 7 e 9
8 a 8	86,36 ± 4,55	90,67 ± 4,00	82,00 ± 7,19	1, 2, 3, 4, 6, 7, 8 e 9
9 a 9	86,36 ± 4,12	93,33 ± 6,05	81,00 ± 4,61	1, 2, 3, 4, 5, 6, 7, 8 e 9
Todas 10	81,82 ± 3,83	85,33 ± 4,61	78,67 ± 6,09	Todas

4.1.3 Discussão dos resultados do estudo de caso 1

No contexto da classificação individual, quando cada escala de observação é analisada, nota-se que o aumento do INS a partir de 80 ms nos sinais saudáveis acaba provocando uma diminuição na taxa de acerto do classificador. No caso em que se considera o classificador LDA, a acurácia atinge mais de 70% novamente em uma janela de observação na décima escala. Isto é um indício de que há escalas de tempo em que são enfatizadas as variações acústicas provocadas por distúrbios da voz.

Nos experimentos de combinação das medidas obtidas das escalas de INS, considerando ambos os classificadores LDA e QDA, foi observado que há um aumento nas taxas de acerto. Em todos os cenários de combinação, o classificador atinge mais de 80% de acurácia, indicando um aspecto complementar das escalas de observação do INS entre si. Isto é, o sistema de classificação é mais robusto quando as informações de INS são analisadas em conjunto.

Este estudo de caso 1 tem duas contribuições relevantes: 1) a aplicação do INS em distúrbios da voz para fins de classificação de patologias na laringe; e 2) a verificação da robustez do INS como medida acústica em sinais sem variações de fala (apenas a emissão da vogal sustentada). Tais resultados indicam que o INS pode ser uma efetiva medida na caracterização e classificação de patologias laríngeas por meio da voz.

4.2 Resultados do Estudo de Caso 2

Nesta Seção são apresentados os resultados obtidos da classificação com LDA e QDA considerando as características baseadas nos modelos linear e não linear de produção da fala. Os cenários experimentais contemplam uma comparação entre as abordagens de segmentação tradicional e adaptativa da voz.

4.2.1 Classificação com as medidas de análise linear de produção da fala

O desempenho da classificação com LDA para LPC, MFCC e GFCC é apresentado na Tabela 6. Com base na acurácia, pode-se notar que a segmentação adaptativa, em ambos os tipos, fornece os melhores resultados de classificação. Comparado à segmentação tradicional, o aumento da acurácia média para LPC, MFCC e GFCC é de aproximadamente 2 p.p., 7 p.p. e 12 p.p., respectivamente. Tal melhoria foi obtida considerando a segmentação adaptativa Tipo I, cujos valores de sensibilidade e especificidade são maiores do que aqueles obtidos com a segmentação tradicional. Por exemplo, para a medida GFCC, o aumento na sensibilidade foi de aproximadamente 18 p.p., enquanto o aumento na especificidade foi de aproximadamente 7 p.p..

Tabela 6 – Resultado da classificação com LDA considerando as medidas de análise linear em diferentes tipos de segmentação.

Segmentação Tradicional			
Medida	**Ac (%)**	**Sens (%)**	**Esp (%)**
LPC	80,00 ± 2,64	89,00 ± 3,02	72,67 ± 4,96
MFCC	85,45 ± 3,09	88,67 ± 3,98	82,33 ± 5,58
GFCC	76,36 ± 2,01	73,33 ± 5,33	78,33 ± 5,54
Segmentação Adaptativa (Tipo I)			
Medida	**Ac (%)**	**Sens (%)**	**Esp (%)**
LPC	**82,73 ± 3,16**	90,00 ± 6,83	75,33 ± 4,16
MFCC	**92,73 ± 2,27**	92,67 ± 3,01	93,33 ± 3,69
GFCC	**88,18 ± 2,73**	91,00 ± 3,02	85,67 ± 4,58
Segmentação Adaptativa (Tipo II)			
Medida	**Ac (%)**	**Sens (%)**	**Esp (%)**
LPC	81,82 ± 3,83	87,00 ± 4,83	78,00 ± 6,58
MFCC	85,45 ± 2,35	84,67 ± 4,67	78,67 ± 3,62
GFCC	87,27 ± 4,33	85,00 ± 6,27	89,00 ± 4,25

Na Tabela 7 são apresentados os resultados da classificação com QDA considerando as medidas de análise linear. No contexto das medidas LPC e MFCC, a segmentação adaptativa, em ambos os tipos, proporciona valores mais elevados de acurácia. Para a medida LPC, o aumento de acurácia média foi de aproximadamente 14 p.p., tendo como um destaque adicional o aumento significativo na sensibilidade média, com mais de 25 p.p. de melhoria, considerando a segmentação adaptativa Tipo II. Para a medida MFCC, assim como ocorreu com o classificador LDA, os resultados mais elevados de acurácia foram obtidos com segmentação adaptativa Tipo I, tendo como destaque adicional os valores médios de sensibilidade e especificidade ultrapassando o valor de 90%. Em relação à medida GFCC, esta proporcionou ao classificador QDA um desempenho, em termos de acurácia, semelhante no que diz respeito a uma comparação entre a segmentação tradicional e a segmentação adaptativa Tipo I. Por outro lado, nota-se que para o GFCC com QDA, ambas as métricas de sensibilidade e especificidade ultrapassam o valor de 80%.

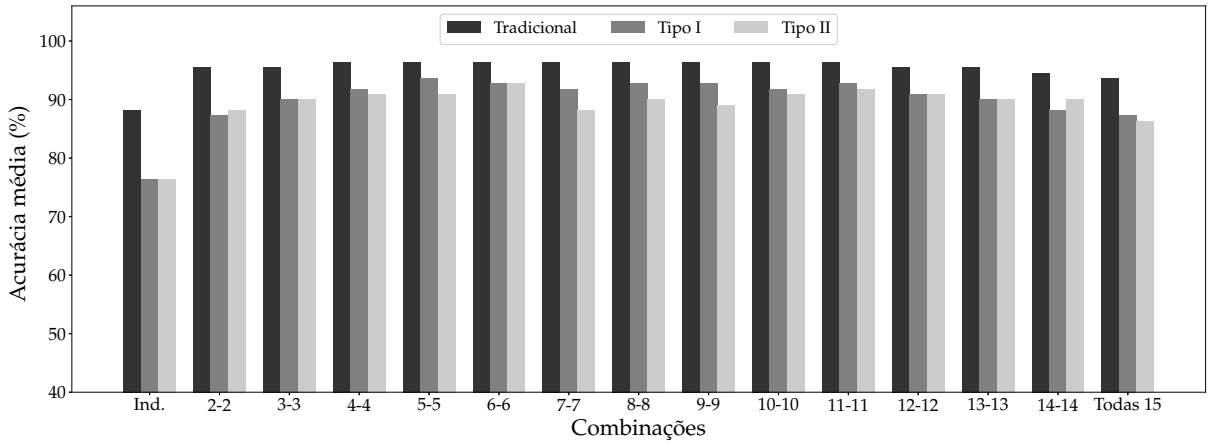
Tabela 7 – Resultado da classificação com QDA considerando as medidas de análise linear em diferentes tipos de segmentação.

Segmentação Tradicional			
Medida	**Ac (%)**	**Sens (%)**	**Esp (%)**
LPC	71,82 ± 3,44	49,67 ± 7,64	93,33 ± 3,69
MFCC	88,18 ± 1,82	88,67 ± 4,31	85,67 ± 4,95
GFCC	**81,82 ± 3,03**	81,00 ± 5,75	83,00 ± 4,98
Segmentação Adaptativa (Tipo I)			
Medida	**Ac (%)**	**Sens (%)**	**Esp (%)**
LPC	79,09 ± 2,73	60,00 ± 4,58	96,67 ± 2,22
MFCC	**91,82 ± 2,52**	92,33 ± 3,14	91,67 ± 3,73
GFCC	**81,82 ± 3,83**	84,33 ± 3,98	79,33 ± 8,86
Segmentação Adaptativa (Tipo II)			
Medida	**Ac (%)**	**Sens (%)**	**Esp (%)**
LPC	**85,45 ± 3,64**	75,00 ± 6,67	95,00 ± 2,55
MFCC	89,09 ± 3,26	90,33 ± 3,24	88,00 ± 4,39
GFCC	70,91 ± 4,45	94,33 ± 4,45	49,00 ± 6,35

4.2.2 Classificação com as medidas de análise não linear de produção da fala

No contexto de medidas baseadas em análise não linear, foram realizados experimentos de classificação considerando cada uma das 15 MQRs, individualmente e em combinação. No contexto do classificador LDA, os melhores resultados em cada caso de combinação, para cada um dos três tipos de segmentação de voz considerados neste trabalho, são apresentados na Figura 12. Na classificação com medidas individuais, o melhor resultado foi obtido com a segmentação tradicional (88,18% ± 2,73%) considerando a medida L_{max}, enquanto a segmentação adaptativa atingiu 76,36% ± 3,88% com DET (Tipo I) e 76,33% ± 3,37% com τ (Tipo II). Em uma análise comparativa da medida L_{max}, o desempenho com segmentação adaptativa foi de 48,18% ± 2,73% (Tipo I) e 69,09% ± 4,92% (Tipo II). Em geral, para todos os cenários de combinação, os melhores

Figura 12 – Melhores valores de acurácia média (%) obtidos com a classificação individual (Ind.) e combinada das MQRs utilizando LDA considerando os diferentes tipos de segmentação.



Fonte: o autor.

resultados de acurácia de classificação foram alcançados com a segmentação tradicional.

De todas as combinações possíveis, os melhores resultados com LDA são detalhados na Tabela 8. Como critério para elencar esses melhores resultados, escolheu-se o compromisso entre o maior valor de acurácia e a menor quantidade de medidas na combinação. Assim, o maior valor de acurácia (segmentação tradicional) foi obtido usando quatro medidas. O valor máximo de acurácia obtido pela segmentação adaptativa considerou cinco medidas para o Tipo I e seis medidas para o Tipo II. Em termos de sensibilidade, tanto a segmentação tradicional quanto a segmentação adaptativa chegaram em um percentual similar (aproximadamente 98%). Por outro lado, no contexto de especificidade, a segmentação tradicional alcançou aproximadamente 3 p.p. a mais em relação à segmentação adaptativa (aproximadamente 94%). No cenário em que a segmentação tradicional obteve tais resultados (τ, m, L_{med} e V_{max}), os valores de acurácia foram 72,73% ± 3,32% e 80,00% ± 2,97% para a segmentação adaptativa dos Tipos I e II, respectivamente.

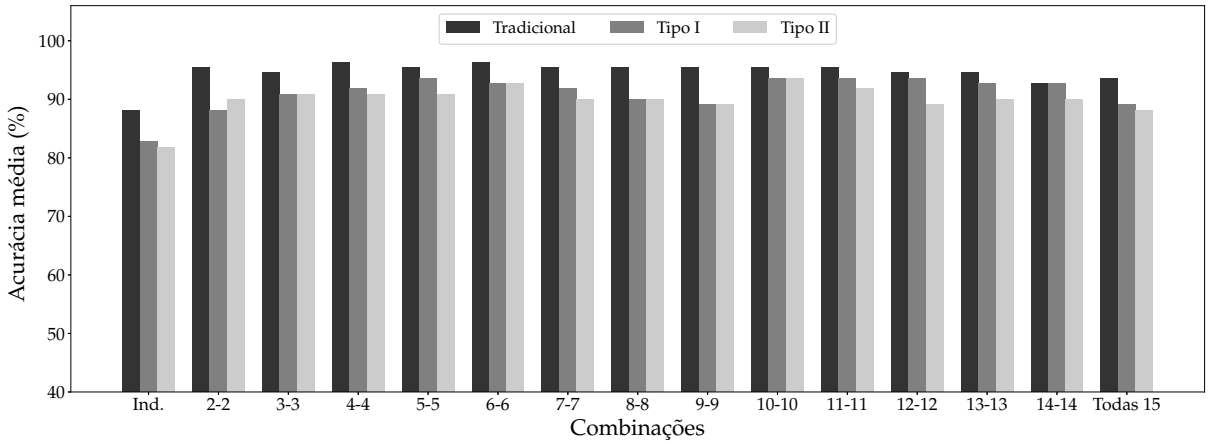
Tabela 8 – Melhores resultados de classificação com LDA considerando as MQRs em diferentes tipos de segmentação.

Segmentação Tradicional			
Medidas	**Ac (%)**	**Sens (%)**	**Esp (%)**
τ, m, L_{med}, V_{max}	**96,36 ± 2,01**	98,00 ± 2,00	94,67 ± 2,73
Segmentação Adaptativa (Tipo I)			
Medidas	**Ac (%)**	**Sens (%)**	**Esp (%)**
τ, LAM, V_{max}, TT, T_2	93,64 ± 1,94	98,00 ± 2,00	89,33 ± 3,84
Segmentação Adaptativa (Tipo II)			
Medidas	**Ac (%)**	**Sens (%)**	**Esp (%)**
τ, DET, $RATIO$, LAM, $ENTR_L$, T_2	92,73 ± 3,26	94,33 ± 2,90	91,00 ± 4,92

Na Figura 13 são apresentados os melhores resultados (em termos de acurácia) obtidos com o classificador QDA comsiderando as MQRs de forma individual e combinada. Assim como ocorreu para o caso com o classificador LDA, os melhores resultados para cada cenário de combinação são obtidos com as MQRs extraídas seguindo a segmentação tradicional. Individualmente, a medida que proporcionou o melhor desempenho do classificador QDA foi L_{max}, com acurácia de 88,18% ± 4,08%. Com esta mesma medida, os valores de acurácia foram 71,82% ± 3,44% e 63,64% ± 4,07% considerando as segmentações adaptativas Tipo I e Tipo II, respectivamente. A medida que, individualmente, proporcinou o maior valor de acurácia ao classificador QDA com segmentação adaptativa Tipo I foi DET (72,73% ± 3,44%) e, para o Tipo II, foi a medida $RATIO$ (65,45% ± 4,85%).

Na Tabela 9 são apresentados, com mais detalhe, os melhores resultados entre

Figura 13 – Melhores valores de acurácia média (%) obtidos com a classificação individual (Ind.) e combinada das MQRs utilizando QDA considerando os diferentes tipos de segmentação.



Fonte: o autor.

todas as combinações possíveis de MQRs utilizando o classificador QDA. Assim como destacado para a Tabela 8, o critério para elencar esses melhores resultados é o compromisso entre o maior valor de acurácia e a menor quantidade de medidas na combinação. Como principais pontos de destaque desses resultados, pode-se elencar: 1) os maiores valores de acurácia, sensibilidade e especificidade são obtidos por meio da segmentação tradicional; 2) tais melhores resultados são obtidos com uma menor quantidade de MQRs na combinação, em comparação às segmentações Tipo I (com cinco medidas) e Tipo II (com dez medidas); e 3) O classificador foi capaz de acertar todos os casos patológicos nos cenários de validação cruzada (sensibilidade média igual a 100%). Além disso, em termos de acurácia, a melhor combinação na segmentação tradicional, com as medidas m, L_{med}, $ENTR_L$ e $ENTR_V$, representou uma melhora de aproximadamente 8 p.p. em relação à melhor medida individual, L_{max}.

Tabela 9 – Melhores resultados de classificação com QDA considerando as MQRs em diferentes tipos de segmentação.

Segmentação Tradicional			
Medidas	**Ac (%)**	**Sens (%)**	**Esp (%)**
m, L_{med} $ENTR_L$, $ENTR_V$	**96,36 ± 2,01**	100,00 ± 0,00	93,33 ± 3,68
Segmentação Adaptativa (Tipo I)			
Medidas	**Ac (%)**	**Sens (%)**	**Esp (%)**
m, LAM, TT, $ENTR_V$, T_2	93,64 ± 3,60	98,00 ± 2,00	89,67 ± 5,15
Segmentação Adaptativa (Tipo II)			
Medidas	**Ac (%)**	**Sens (%)**	**Esp (%)**
τ, DET, L_{med}, LAM, $RATIO_{LAM/DET}$, $ENTR_L$, DIV, TT, $ENTR_V$, T_2	93,64 ± 1,94	100,00 ± 0,00	87,33 ± 4,09

4.2.3 Discussão dos resultados do estudo de caso 2

Os experimentos neste segundo estudo de caso objetivaram a verificação do desempenho de medidas acústicas conhecidas na literatura, com e sem segmentação adaptativa. Os resultados obtidos no contexto das medidas lineares mostraram que a segmentação adaptativa é mais robusta no sentido de proporcionar ao classificador valores mais elevados de acurácia. No caso dos coeficientes LPC (que faz parte de um contexto espectral), o melhor resultado foi obtido com a segmentação pela maior escala. Por outro lado, no contexto de medidas cepstrais (MFCC e GFCC), o melhor resultado foi obtido com a segmentação pela menor escala. Isto é um indício de que medidas espectrais e cepstrais são sensíveis às variações acústicas não estacionárias de maneiras diferentes.

No contexto das medidas não lineares, o comportamento das MQRs foi diferente se comparado com as medidas lineares. A segmentação convencional proporcionou melhores resultados que a segmentação adaptativa. Entende-se que este resultado é coerente, pois as MQRs são medidas projetadas para lidar com sistemas dinâmicos não lineares e não estacionários (WEBBER-JR; ZBILUT, 2005; MARWAN, 2003). Isso significa que MQRs podem usar a natureza não estacionária dos sinais como uma característica do fenômeno. Tal comportamento pode afetar os resultados obtidos, caso em que a segmentação tradicional fornece maior acurácia do que aquela obtida com segmentação adaptativa. Assim, medidas que não têm um requisito de estacionariedade, como MQRs, são mais promissoras sem segmentação adaptativa.

Este estudo de caso 2 tem duas contribuições relevantes: 1) implementação de segmentação adaptativa baseada no INS para a extração de medidas acústicas; e 2) a proposta de uma nova metodologia de pré-processamento para extração de características de sinais de voz, baseando-se no tipo de medida acústica.

5 Considerações Finais

A classificação de distúrbios da voz por meio de uma medida acústica que efetivamente caracterize um estado patológico ainda é um desafio. A avaliação de patologias laríngeas por meio do sinal de voz é, em geral, realizada por meio da vogal sustentada. Para este tipo de sinal, a detecção de variações acústicas não estacionárias pode representar padrões relacionados a disfonias. Esta pesquisa foi realizada considerando patologias de diferentes naturezas. Enquanto a paralisia nas pregas vocais tem origem neurológica, tanto o edema de Reinke quanto os nódulos têm origem funcional (abuso vocal, tabagismo e alcoolismo, entre outros fatores). Além disso, uma mesma patologia pode apresentar variações em diferentes pacientes. Por exemplo, há casos de paralisia unilateral direita, unilateral esquerda e bilateral.

Quanto às questões norteadoras deste estudo, pôde-se observar que, mesmo com essa heterogeneidade de sinais de voz, em relação a diferentes patologias, as classes saudável e patológica se comportaram de forma diferente quanto à estacionariedade. Isso significa que sinais de voz de laringes saudáveis apresentaram diferentes graus de não estacionariedade, para diferentes tamanhos de segmento, quando comparados a sinais de laringes patológicas. Portanto, segmentos de tamanho curto podem não manter a estacionariedade, de acordo com a patologia. Por exemplo, um segmento de 20 ms pode ser estacionário se for um sinal saudável, mas pode ser não estacionário dependendo da gravidade da patologia laríngea. Além disso, os resultados deste trabalho verificaram que o grau de não estacionariedade do sinal, obtido com INS, pode afetar o procedimento de extração de características e, ainda, servir como característica acústica para a classificação de desordens vocais.

Em relação aos classificadores empregados em ambos os estudos de caso, a abordagem LDA apresentou resultados de acurácia ligeiramente mais elevados em comparação ao classificador QDA. Isto significa que LDA pode ser uma escolha razoável de classificação em ferramentas de reconhecimento de padrões que utilizem as características empregadas neste trabalho. Por exemplo, para o estudo de caso 1, a classificação, realizada com LDA, mostrou que há escalas de observação em que a medida de INS proporciona acurácia de mais de 70%. Além disso, com a combinação de medidas, foi verificado que o desempenho do classificador atinge mais de 91% de acerto. Estes resultados indicam que o INS pode ser uma efetiva medida na caracterização e classificação de patologias laríngeas por meio da voz.

No conetxto do estudo de caso 2, foi observado que a segmentação adaptativa contribuiu para a melhoria do desempenho do classificador LDA com as características baseadas no modelo linear de produção da fala. Em comparação com a segmentação

tradicional, o melhor resultado foi alcançado ao considerar GFCC, que forneceu melhorias de 12 p.p. em acurácia, 18 p.p. em sensibilidade e 7 p.p. em especificidade. Como os MQRs não exigem estacionariedade, a segmentação tradicional proporcionou os melhores resultados para características não lineares. Assim, a segmentação adaptativa baseada em não estacionariedade pode ser empregada para extrair características como LPC, MFCC e GFCC para executar tarefas de classificação de distúrbios de voz de maneira aprimorada.

Por fim, após a análise dos resultados obtidos nesta pesquisa e apresentados neste trabalho, pode-se surgir a seguinte questão: "em meio às medidas utilizadas nesta pesquisa, se pudéssemos escolher apenas uma abordagem, qual seria ela?". A resposta é: depende. Para colocar alguma dessas abordagens em uma aplicação para o dia-a-dia, deve-se levar em consideração quais os seus requisitos. Nesse contexto, se a aplicação requerer baixa latência e acurácia entre 70% e 80%, a segmentação tradicional pode ser a escolha razoável, uma vez que não haverá uma etapa de processamento a mais (relacionada à análise com INS). Ou ainda, o próprio valor de INS em duas escalas de observação pode ser útil, uma vez que ele mesmo é a medida acústica e não se faz necessário percorrer diferentes escalas de tempo para escolher um tamanho de segmento ideal. Por outro lado, se a necessidade for alta acurácia, e os requisitos de latência não forem tão exigentes, pode-se escolher a análise linear com segmentação adaptativa, ou mesmo a análise não linear com segmentação tradicional.

Como sugestão para trabalhos futuros, pode-se elencar:

- Combinção das características empregadas no estudo de caso 1 com aquelas empregadas no estudo de caso 2, ou seja, realizar uma investigação do potencial discriminativo da combinação entre as melhores escalas do INS com os melhores casos de medidas lineares e não lineares.
- Investigação com outros tipos de classificadores, tais como árvores de decisão e redes neurais artificiais.
- Validação dos resultados encontrados neste trabalho utilizando outras bases de dados.
- Investigação da segmentação adaptativa utilizando outras características acústicas baseadas nos modelos linear e não linear de produção da fala.

5.1 Publicações desta Pesquisa

- VIEIRA, Vinicus J D; COSTA, Silvana C.; CORREIA, Suzete E N. Índice de Não Estacionariedade aplicado à Classificação de Desordens Vocais. Anais do XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT22), 2022.

  [Disponível em http://dx.doi.org/10.14209/sbrt.2022.1570818066].

- VIEIRA, Vinicius J D; COSTA, Silvana C.; CORREIA, Suzete E N. Non-Stationarity-Based Adaptive Segmentation Applied to Voice Disorder Discrimination. IEEE Access, v. 11, p. 54750-54759, 2023.

  [Disponível em http://dx.doi.org/10.1109/ACCESS.2023.3281191].

REFERÊNCIAS

AGGARWAL, S. et al. Audio segmentation techniques and applications based on deep learning. *Scientific Programming*, Hindawi, v. 2022, 2022.

AKBARI, A.; ARJMANDI, M. K. An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features. *Biomedical Signal Processing and Control*, Elsevier, v. 10, p. 209–223, 2014.

ANU, J.; KARJIGI, V. Sentence segmentation for speech processing. In: IEEE. *2014 IEEE National Conference on Communication, Signal Processing and Networking (NCCSN)*. [S.l.], 2014. p. 1–4.

BAI, Z.; ZHANG, X.-L. Speaker recognition based on deep learning: An overview. *Neural Networks*, Elsevier, v. 140, p. 65–99, 2021.

BANSAL, P.; IMAM, S. A.; BHARTI, R. Speaker recognition using mfcc, shifted mfcc with vector quantization and fuzzy. In: IEEE. *2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI)*. [S.l.], 2015. p. 41–44.

BASSEVILLE, M. Distance measures for signal processing and pattern recognition. *Signal processing*, Elsevier, v. 18, n. 4, p. 349–369, 1989.

BEHLAU, M. *Voz: O Livro do Especialista*. [S.l.]: Revinter, 2001. v. 1.

BENJAMIN, B.; FIGUEIREDO, J. E. F. de. *Cirurgia endolaríngea*. [S.l.]: Revinter, 2000.

BENZEGHIBA, M. et al. Automatic speech recognition and speech variability: A review. *Speech communication*, Elsevier, v. 49, n. 10-11, p. 763–786, 2007.

BORGNAT, P. et al. Testing stationarity with surrogates: A time-frequency approach. *IEEE Transactions on Signal Processing*, IEEE, v. 58, n. 7, p. 3459–3470, 2010.

CAPPONI, L. et al. Non-stationarity index in vibration fatigue: Theoretical and experimental research. *International Journal of Fatigue*, Elsevier, v. 104, p. 221–230, 2017.

CHERN, A. et al. A smartphone-based multi-functional hearing assistive system to facilitate speech recognition in the classroom. *IEEE Access*, IEEE, v. 5, p. 10339–10351, 2017.

CHOWDHURY, A.; ROSS, A. Fusing mfcc and lpc features using 1d triplet cnn for speaker recognition in severely degraded audio signals. *IEEE transactions on information forensics and security*, IEEE, v. 15, p. 1616–1629, 2019.

COLTON, R. H.; CASPER, J. K.; LEONARD, R. *Understanding voice problems: A physiological perspective for diagnosis and treatment*. [S.l.]: Wolters Kluwer Health, 2006.

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

COSTA, W. C. de A. *Análise Dinâmica Não Linear de Sinais de Voz para Detecção de Patologias Laríngeas*. Tese (Doutorado) — Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, 2012.

DAVIS, S.; MERMELSTEIN, P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, IEEE, v. 28, n. 4, p. 357–366, 1980.

DIAS, L. *Detecção de patologias laríngeas por meio da análise de sinais de voz utilizando Deep Neural Network*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, 2020.

ECKMANN, J.-P.; KAMPHORST, S. O.; RUELLE, D. Recurrence plots of dynamical systems. *Europhys. Lett*, World Scientific, v. 4, n. 9, p. 973–977, 1987.

ELEMETRICS, K. *Kay Elemetrics Corp. Disordered Voice Database*. 1994. Model 4337, 03 Ed.

FAHAD, M. S. et al. Dnn-hmm-based speaker-adaptive emotion recognition using mfcc and epoch-based features. *Circuits, Systems, and Signal Processing*, Springer, v. 40, p. 466–489, 2021.

FANT, G. The source filter concept in voice production. *STL-QPSR*, v. 1, n. 1981, p. 21–37, 1981.

GODINO-LLORENTE, J. I.; GOMEZ-VILDA, P.; BLANCO-VELASCO, M. Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. *Biomedical Engineering, IEEE Transactions on*, IEEE, v. 53, n. 10, p. 1943–1953, 2006.

GOY, H. et al. Normative voice data for younger and older adults. *Journal of Voice*, Elsevier, v. 27, n. 5, p. 545–555, 2013.

ISMAIL, A.; ABDLERAZEK, S.; EL-HENAWY, I. M. Development of smart healthcare system based on speech recognition using support vector machine and dynamic time warping. *Sustainability*, Multidisciplinary Digital Publishing Institute, v. 12, n. 6, p. 2403, 2020.

JIANG, J. J.; ZHANG, Y.; MCGILLIGAN, C. Chaos in voice, from modeling to measurement. *Journal of Voice*, Elsevier, v. 20, n. 1, p. 2–17, 2006.

KINNUNEN, T.; LI, H. An overview of text-independent speaker recognition: From features to supervectors. *Speech communication*, Elsevier, v. 52, n. 1, p. 12–40, 2010.

KNIGHT, E. J.; AUSTIN, S. F. The effect of head flexion/extension on acoustic measures of singing voice quality. *Journal of Voice*, Elsevier, v. 34, n. 6, p. 964–e11, 2020.

KUHL, I. A. *Manual prático de laringologia*. Porto Alegre: Editora da Universidade UFGRS, 1982. v. 11.

KUMAR, A.; MULLICK, S. K. Nonlinear dynamical analysis of speech. *The Journal of the Acoustical Society of America*, v. 100, p. 615, 1996.

LOPES, L. W. et al. Accuracy of acoustic analysis measurements in the evaluation of patients with different laryngeal diagnoses. *Journal of Voice*, Elsevier, v. 31, n. 3, p. 382–e15, 2017.

MARTIN, N.; MAILHES, C. A non-stationary index resulting from time and frequency domains. In: *CM 2009-MFPT 2009-6th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*. [S.l.: s.n.], 2009.

MARWAN, N. Encounters with neighbours. *University of Potsdam. Tese de Doutorado. 159 p.*, 2003.

MARWAN, N.; WEBBER-JR, C. L. Mathematical and computational foundations of recurrence quantifications. *Recurrence quantification analysis: Theory and best practices*, Springer, p. 3–43, 2015.

MARWAN, N. et al. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, APS, v. 66, n. 2, p. 026702, 2002.

MOHANTY, S. Language Independent Emotion Recognition in Speech Signals. *International Journal*, v. 6, n. 10, 2016.

MUNIER, C. et al. Relationship between laryngeal signs and symptoms, acoustic measures, and quality of life in finnish primary and kindergarten school teachers. *Journal of Voice*, Elsevier, v. 34, n. 2, p. 259–271, 2020.

OROZCO-ARROYAVE, J. R. et al. Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases. *IEEE journal of biomedical and health informatics*, IEEE, v. 19, n. 6, p. 1820–1828, 2015.

O'SHAUGHNESSY, D. *Speech communication: human and machine*. [S.l.]: Universities press, 1987.

PATTERSON, R. D.; HOLDSWORTH, J.; ALLERHAND, M. Auditory models as preprocessors for speech recognition. *The Auditory Processing of Speech: from Auditory Periphery to Words*, Mouton de Gruyler, Berlin, p. 67–89, 1992.

QUEIROZ, G. K. L. P. *Análise Dinâmica não linear e Análise de Quantificação de Recorrência aplicadas na Classificação de Desvios Vocais*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Elétrica, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, 2018.

RABINER, L. R.; SCHAFER, R. W. *Digital processing of speech signals*. [S.l.]: Prentice-Hall Signal Processing Letters, 1978.

RABINER, L. R.; SCHAFER, R. W. *Introduction to digital speech processing*. [S.l.]: Now Publishers Inc, 2007. v. 1.

RANGARATHNAM, B. et al. Telepractice versus in-person delivery of voice therapy for primary muscle tension dysphonia. *American Journal of Speech-Language Pathology*, ASHA, v. 24, n. 3, p. 386–399, 2015.

SCHLUTER, R. et al. Gammatone features and feature combination for large vocabulary speech recognition. In: IEEE. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.* [S.l.], 2007. v. 4, p. IV–649.

SHAO, Y.; SRINIVASAN, S.; WANG, D. Incorporating auditory feature uncertainties in robust speaker identification. In: IEEE. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.* [S.l.], 2007. v. 4, p. IV–277.

SHARMA, S.; SINGH, P. Emotion Recognition based on Audio Signal using GFCC Extraction and BPNN Classification. *International Journal of Computational Engineering Research*, p. 39–42, 2015.

SHWARTZ-ZIV, R.; ARMON, A. Tabular data: Deep learning is not all you need. *Information Fusion*, Elsevier, v. 81, p. 84–90, 2022.

STROHL, M. P. et al. Implementation of telemedicine in a laryngology practice during the covid-19 pandemic: lessons learned, experiences shared. *Journal of Voice*, Elsevier, 2020.

SULICA, M. L. *Specialized Care for the Voice*. 2013. Online: http://voicemedicine.com/ [Acesso em 18 de novembro de 2013].

TACHBELIE, M. Y.; ABATE, S. T.; BESACIER, L. Using different acoustic, lexical and language modeling units for asr of an under-resourced language–amharic. *Speech Communication*, Elsevier, v. 56, p. 181–194, 2014.

TAKENS, F. Detecting strange attractors in turbulence. In: *Dynamical systems and turbulence, Warwick 1980*. [S.l.]: Springer, 1981. p. 366–381.

TAVARES, R.; COELHO, R. Speech enhancement with nonstationary acoustic noise detection in time domain. *IEEE Signal Processing Letters*, IEEE, v. 23, n. 1, p. 6–10, 2015.

TIRRONEN, S.; KADIRI, S. R.; ALKU, P. The effect of the mfcc frame length in automatic voice pathology detection. *Journal of Voice*, Elsevier, 2022.

VENTURINI, A.; ZAO, L.; COELHO, R. On speech features fusion, α-integration gaussian modeling and multi-style training for noise robust speaker classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, IEEE, v. 22, n. 12, p. 1951–1964, 2014.

VIEIRA, V.; COELHO, R.; ASSIS, F. M. de. Hilbert–huang–hurst-based non-linear acoustic feature vector for emotion classification with stochastic models and learning systems. *IET Signal Processing*, IET, v. 14, n. 8, p. 522–532, 2020.

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

VIEIRA, V. J. D. *Análise de Variações Acústicas Não-Estacionárias e seu Efeito na Detecção de Múltiplas Emoções e Condições de Estresse*. Tese (Doutorado) — Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, 2018.

VIEIRA, V. J. D.; COELHO, R.; ASSIS, F. M. Classificação de variações acústicas emocionais com atributos da fonte e do trato vocal. In: SBRT 2018. *Anais do XXXVI Simpósio Brasileiro de Telecomunicações*. [S.l.], 2018.

VIEIRA, V. J. D.; COSTA, S. C.; CORREIA, S. E. N. Índice de não estacionariedade aplicado à classificação de desordens vocais. In: SBRT 2022. *Anais do XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. [S.l.], 2022.

VIEIRA, V. J. D. et al. Análise de quantificação de recorrência a curto e a longo intervalo de tempo na avaliação de patologias laríngeas. In: SOCIEDADE BRASILEIRA DE ENGENHARIA BIOMéDICA. *Anais do XXIV Congresso Brasileiro de Engenharia Biomédica*. [S.l.], 2014.

VIEIRA, V. J. D. et al. Exploiting nonlinearity of the speech production system for voice disorder assessment by recurrence quantification analysis. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, AIP Publishing LLC, v. 28, n. 8, p. 085709, 2018.

VIEIRA, V. J. D. et al. Discriminação de sinais de voz com análise de quantificação de recorrência e redes neurais mlp. In: *Anais do XXXI Simpósio Brasileiro de Telecomunicações (SBrT 2013)*. [S.l.: s.n.], 2013.

WANG, K. et al. Speech Emotion Recognition using Fourier Parameters. *IEEE Transactions on Affective Computing*, IEEE, v. 6, n. 1, p. 69–75, 2015.

WANG, N. et al. Robust Speaker Recognition using Denoised Vocal Source and Vocal Tract Features. *IEEE Transactions on Audio, Speech, and Language Processing*, IEEE, v. 19, n. 1, p. 196–205, 2011.

WEBBER-JR, C. L.; ZBILUT, J. P. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, v. 94, n. 2005, p. 26–94, 2005.

WU, S.; FALK, T. H.; CHAN, W.-Y. Automatic speech emotion recognition using modulation spectral features. *Speech Communication*, Elsevier, v. 53, n. 5, p. 768–785, 2011.

ZBILUT, J. P.; WEBBER-JR, C. L. Embeddings and delays as derived from quantification of recurrence plots. *Physics letters A*, Elsevier, v. 171, n. 3, p. 199–203, 1992.

ZITTA, S. M. *Análise perceptivo-auditiva e acústica em mulheres com nódulos vocais*. Dissertação (Mestrado) — Centro Federal de Educação Tecnológica do Paraná, 2010.

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
Campus João Pessoa - Código INEP: 25096850
Av. Primeiro de Maio, 720, Jaguaribe, CEP 58015-435, Joao Pessoa (PB)
CNPJ: 10.783.898/0002-56 - Telefone: (83) 3612.1200

Documento Digitalizado Ostensivo (Público)

TCC com ficha catalográfica e folha de aprovação

Assunto:	TCC com ficha catalográfica e folha de aprovação
Assinado por:	Vinicius Vieira
Tipo do Documento:	Dissertação
Situação:	Finalizado
Nível de Acesso:	Ostensivo (Público)
Tipo do Conferência:	Cópia Simples

Documento assinado eletronicamente por:

- **Vinicius Jefferson Dias Vieira**, **ALUNO (20201610001) DE BACHARELADO EM ENGENHARIA ELÉTRICA - JOÃO PESSOA**, em 01/10/2024 22:03:57.

Este documento foi armazenado no SUAP em 01/10/2024. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse https://suap.ifpb.edu.br/verificar-documento-externo/ e forneça os dados abaixo:

Código Verificador: 1264936
Código de Autenticação: ecf29628a3