UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE - DOUTORADO

MODELO COM BASE NA LÓGICA FUZZY PARA O RECONHECIMENTO DE DIFERENTES ESTADOS EMOCIONAIS A PARTIR DA VOZ

Alexandra Christine de Aguiar Silva

JOÃO PESSOA-PB
2025

ALEXANDRA CHRISTINE DE AGUIAR SILVA

MODELO COM BASE NA LÓGICA FUZZY PARA O RECONHECIMENTO DE DIFERENTES ESTADOS EMOCIONAIS A PARTIR DA VOZ

Tese apresentada ao Programa de Pós-Graduação em Modelos de Decisão e Saúde – Nível Doutorado, do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba, como requisito para obtenção do título de Doutor.

Linha de Pesquisa: Modelos em Saúde

Orientadores:

Profa. Dra. Anna Alice Figueirêdo de Almeida
Prof. Dr. Ronei Marcos de Moraes

JOÃO PESSOA-PB

2025

Catalogação na publicação
Seção de Catalogação e Classificação

S586m Silva, Alexandra Christine de Aguiar.
Modelo com base na lógica fuzzy para o reconhecimento de diferentes estados emocionais a partir da voz / Alexandra Christine de Aguiar Silva. - João Pessoa, 2025.
147 f. : il.

Orientação: Anna Alice Figueirêdo de Almeida, Ronei Marcos de Moraes.
Tese (Doutorado) - UFPB/CCEN.

1. Inteligência artificial - Voz. 2. Reconhecimento de voz. 3. Emoções expressas. 4. Processamento de voz. 5. Lógica fuzzy. I. Almeida, Anna Alice Figueirêdo de. II. Moraes, Ronei Marcos de. III. Título.

UFPB/BC CDU 004.8(043)

Elaborado por Walqueline da Silva Araújo - CRB-15/514

Ata da Sessão Pública de Defesa de Tese de Doutorado de **ALEXANDRA CHRISTINE DE AGUIAR SILVA**, candidata ao Título de Doutora em Modelos de Decisão e Saúde, realizada em **20 de fevereiro de 2025**.

Aos vinte dias do mês de fevereiro do ano de dois mil e vinte e cinco, às 13h00min, no Auditório Humberto Nóbrega - CCS, instalou-se a banca examinadora de tese de Doutorado da aluna ALEXANDRA CHRISTINE DE AGUIAR SILVA. A banca examinadora foi composta pelos professores Dr. REGIVAN HUGO NUNES SANTIAGO, UFRN, e Dra. ANA CAROLINA CONSTANTINI, UNICAMP, examinadores externos, Dra. LILIANE DOS SANTOS MACHADO, UFPB, e Dr. LEONARDO WANDERLEY LOPES, UFPB, examinadores internos, Dr. RONEI MARCOS DE MORAES, UFPB, como orientador e Dra. ANNA ALICE FIGUEIREDO DE ALMEIDA, UFPB, como orientadora e presidente da banca examinadora. Dando início aos trabalhos, a presidente da banca cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra a candidata para que se fizesse, oralmente, a exposição do trabalho de tese intitulado "MODELO PARA O RECONHECIMENTO DE DIFERENTES ESTADOS EMOCIONAIS A PARTIR DA VOZ COM BASE NA LÓGICA FUZZY". Concluída a apresentação, a candidata foi arguida pela Banca Examinadora que emitiu o parecer **"APROVADA"**. Sendo assim, após a candidata proceder às devidas correções exigidas pela Banca Examinadora no prazo máximo de **30 dias** e depositar as cópias finais da tese, deverá a Coordenação expedir uma declaração de conclusão do Programa e a Universidade Federal da Paraíba deverá expedir o respectivo diploma de Doutora em Modelos de Decisão e Saúde, na forma da lei. E, para constar, o Prof. Hemílio Fernandes Campos Coêlho, Sr, Coordenador do Programa de Pós-Graduação em Modelos de Decisão e Saúde, lavrou a presente Ata, que vai assinada por ele e pelos demais membros da Banca Examinadora. João Pessoa, 20 de fevereiro de 2025.

Prof. Anna Alice Figueiredo de Almeida _____
Orientadora e Presidente

Prof. Ronei Marcos de Moraes_____
Orientador

Profa. Liliane dos Santos Machado _____
Examinadora Interna

Prof. Leonardo Wanderley Lopes _____
Examinador Interno

Profa. Ana Carolina Constantini _____
Examinadora Externa à Instituição

gov.br Documento assinado digitalmente
ANA CAROLINA CONSTANTINI
Data: 21/02/2025 08:12:23-0300
Verifique em https://validar.iti.gov.br

Prof. Regivan Hugo Nunes Santiago_____
Examinador Externo à Instituição

gov.br Documento assinado digitalmente
REGIVAN HUGO NUNES SANTIAGO
Data: 22/02/2025 09:15:12-0300
Verifique em https://validar.iti.gov.br

Emitido em 20/02/2025

ATA Nº 12/2025 - PPGMDS (11.01.14.54)
(Nº do Documento: 12)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 25/02/2025 12:52)
RONEI MARCOS DE MORAES
PROFESSOR DO MAGISTERIO SUPERIOR
337967

(Assinado digitalmente em 28/02/2025 15:09)
LILIANE DOS SANTOS MACHADO
PROFESSOR DO MAGISTERIO SUPERIOR
2379027

(Assinado digitalmente em 27/03/2025 13:01)
LEONARDO WANDERLEY LOPES
PROFESSOR TITULAR-LIVRE MAG SUPERIOR
2634755

(Assinado digitalmente em 26/02/2025 20:32)
ANNA ALICE FIGUEIREDO DE ALMEIDA QUEIROZ
PROFESSOR DO MAGISTERIO SUPERIOR
1668545

Para verificar a autenticidade deste documento entre em https://sipac.ufpb.br/documentos/ informando seu número: **12**, ano: **2025**, documento (espécie): **ATA**, data de emissão: **25/02/2025** e o código de verificação: **b9fa5ed767**

Dedico este trabalho primeiramente a Deus, por me iluminar e me dar força em cada etapa desta caminhada. Aos meus pais, Josete Amorim e Edinaldo Galdino, pelo amor incondicional, apoio e incentivo que sempre me proporcionaram. Ao meu marido Romário Leite, pelo amor, apoio e inspiração. À minha família, que sempre acreditou em meu potencial e foi a base sólida nos momentos de desafio. E a todos aqueles que, de alguma forma, contribuíram para a realização deste sonho.

AGRADECIMENTOS

A **Deus**, por todas as bênçãos concedidas, pela força, sabedoria e graça que me sustentaram em cada etapa desta jornada. Pelo amor e cuidado incondicional, que me guiaram nos momentos de dúvida e me impulsionaram a seguir em frente.

Ao meu marido, **Romário**, por estar ao meu lado em todos os momentos. Pelo incentivo constante, apoio inabalável, companheirismo e amor que me fortalecem diariamente.

À minha mãe, **Josete**, e ao meu pai, **Edinaldo**, por tudo o que fizeram e continuam fazendo por mim. Pelos valores que me transmitiram, pelo apoio incondicional e por sempre acreditarem no meu potencial. Sou profundamente grata pelo amor, pela paciência e pela base sólida que me proporcionaram.

À minha orientadora, **Dra. Anna Alice Almeida**, meu sincero agradecimento! Por sua dedicação, paciência e sabedoria, que me formaram como pesquisadora e me inspiraram como pessoa. Pelos desafios propostos, pelas orientações valiosas e por acreditar no meu trabalho, guiando-me com excelência e generosidade. Sua contribuição para minha formação será sempre lembrada com carinho e admiração.

Ao meu orientador, **Ronei Moraes**, pelo apoio, incentivo e ensinamentos ao longo desta trajetória. Pela paciência em cada etapa do processo e por compartilhar seu conhecimento de forma tão clara e acessível, tornando essa caminhada mais enriquecedora e significativa.

Aos professores **Dr. Leonardo Lopes**, **Dr. Liliane Machado**, **Dra. Ana Carolina Constantini** e **Dr. Regivan Santiago**, por gentilmente aceitarem participar da banca e por suas valiosas contribuições. Suas análises e sugestões enriqueceram profundamente este trabalho.

À família **LIEV**, minha eterna gratidão por todo o auxílio, disponibilidade e atenção.

Aos amigos do **LEAPIG**, por todo o companheirismo e apoio durante esta jornada. Pelas trocas de conhecimentos e pelo estímulo mútuo que fizeram toda a diferença.

Aos meus colegas da **turma de Doutorado 2020**, pelo companheirismo e pela força que compartilhamos ao longo dessa jornada, especialmente durante os desafios enfrentados durante a pandemia. A união e o apoio mútuo fizeram toda a diferença.

Ao **Programa de Pós-Graduação em Modelos de Decisão e Saúde - UFPB**, que me proporcionou um ambiente de crescimento acadêmico e pessoal, fornecendo o apoio e os conhecimentos essenciais para a concretização deste estudo.

À **CAPES**, pelo financiamento da pesquisa, imprescindível para a realização deste trabalho.

A todos que, de alguma forma, contribuíram para a realização deste sonho, deixo aqui minha gratidão. Esta conquista não é só minha, mas de todos que caminharam ao meu lado. Muito obrigada!

“A educação é a arma mais poderosa que você pode usar para mudar o mundo.”

Nelson Mandela

RESUMO

INTRODUÇÃO: A voz é produzida por um complexo processo neurofisiológico que envolve diversas estruturas e sistemas do corpo, pode ser influenciada pelo estado emocional do indivíduo e por características da personalidade. A construção de um modelo que utilize o sinal de voz em diferentes emoções irá servir de base para elaboração e desenvolvimento de sistemas de reconhecimento que sejam robustos na automatização do reconhecimento das emoções. A definição de características acústicas próprias de cada estado emocional poderá permitir a efetiva separação das emoções por meio de técnicas de reconhecimento de padrões de voz. **OBJETIVO:** Desenvolver um modelo baseado na lógica *fuzzy* capaz de reconhecer as emoções a partir da voz de falantes do português brasileiro. **MÉTODO:** Trata-se de um estudo de natureza tecnológica, descritiva, observacional e transversal, que utilizou dados secundários, provenientes do Banco de Vozes Brasileiro nas Variações das Emoções - EmoVox-BR, que conta com um conjunto de dados composto por 182 sinais sonoros associados às emoções básicas: alegria, medo, tristeza, raiva, surpresa, nojo e estado neutro, produzidos por 26 atores profissionais e em formação. Foram extraídos parâmetros acústicos como medidas de frequência fundamental (f_0), *jitter*, *shimmer*, medidas de ruído glótico *Glottal to Noise Excitation Ratio* (GNE) e *Harmonics-to-Noise Ratio* (HNR), medidas cepstrais *Cepstral* Peak Prominence Smoothed (CPPS), *Mel-Frequency* Cepstral *Coefficients* (MFCC) e os parâmetros acústico-prosódicos da f_0, duração e intensidade a partir dos dados do EmoVox-BR. A partir desses, desenvolveu-se o modelo de reconhecimento das emoções. **RESULTADOS:** O estudo apresentou um modelo inovador baseado em lógica *fuzzy* para o reconhecimento dos estados emocionais a partir da voz, com a integração de parâmetros acústicos e prosódicos selecionados de forma criteriosa. O modelo utiliza 18 parâmetros de entrada e 7 de saída. Foram implementadas 23 regras *fuzzy* para discriminação de sete categorias emocionais. O modelo apresentou acurácia de 89,01% e coeficiente Kappa de 87,18%, com sensibilidade variando entre 76,92% para alegria e 100% para tristeza e estado neutro. As especificidades foram superiores a 93% em todas as categorias, indicando alta capacidade de diferenciação emocional. Variáveis sobrepostas, como alguns parâmetros acústicos e mel-cepstrais, foram removidas para simplificar o modelo e refinar seu desempenho. Além de otimizar a simplicidade e eficiência no reconhecimento das emoções, o modelo superou métodos convencionais em robustez em comparação com algoritmos de aprendizado de máquina como o Random

Forest e Kernel SVM. **CONCLUSÃO:** O modelo de reconhecimento emocional, desenvolvido com lógica *fuzzy* a partir da voz humana, alcançou alta acurácia e precisão na diferenciação de emoções. O resultado prevê aplicações promissoras em computação afetiva e tecnologias interativas. A abordagem demonstra eficácia ao capturar nuances emocionais, com adaptação a diferentes contextos para tornar interações humano-máquina mais empáticas e personalizadas.

Palavras-Chave: Voz, Reconhecimento de voz, Emoções Expressas, Processamento de Voz, Inteligência Artificial, Lógica *Fuzzy*.

ABSTRACT

INTRODUCTION: Voice is produced by a complex neurophysiological process that involves several structures and systems of the body, and can be influenced by the individual's emotional state and personality traits. The construction of a model that uses the voice signal in different emotions will serve as a basis for the development and development of recognition systems that are robust in the automation of emotion recognition. The definition of acoustic characteristics specific to each emotional state may allow the effective separation of emotions through voice pattern recognition techniques. **OBJECTIVE**: To develop a model based on fuzzy logic capable of recognizing emotions from the voices of Brazilian Portuguese speakers. **METHOD**: This is a technological, descriptive, observational and cross-sectional study that used secondary data from the Brazilian Voice Bank in Variations of Emotions - EmoVox-BR, which has a data set composed of 182 sound signals associated with the basic emotions: joy, fear, sadness, anger, surprise, disgust and neutral state, produced by 26 professional actors and actors in training. Acoustic parameters such as fundamental frequency (f_o), jitter, shimmer, glottal noise measures Glottal to Noise Excitation Ratio (GNE) and Harmonics-to-Noise Ratio (HNR), cepstral measures Cepstral Peak Prominence Smoothed (CPPS), Mel-Frequency Cepstral Coefficients (MFCC) and acoustic-prosodic parameters of f_o, duration and intensity were extracted from the EmoVox-BR data. From these, the emotion recognition model was developed. **RESULTS**: The study presented an innovative model based on fuzzy logic for the recognition of emotional states from voice, with the integration of carefully selected acoustic and prosodic parameters. The model uses 18 input and 7 output parameters. Twenty-three fuzzy rules were implemented to discriminate seven emotional categories. The model presented an accuracy of 89.01% and a Kappa coefficient of 87.18%, with sensitivity ranging from 76.92% for happiness to 100% for sadness and neutral state. Specificities were higher than 93% in all categories, indicating a high capacity for emotional differentiation. Overlapping variables, such as some acoustic and mel-cepstral parameters, were removed to simplify the model and refine its performance. In addition to optimizing simplicity and efficiency in emotion recognition, the model outperformed conventional methods in robustness compared to machine learning algorithms such as Random Forest and Kernel SVM. **CONCLUSION**: The emotion recognition model, developed with fuzzy logic from the human voice, achieved high accuracy and precision in differentiating emotions. The result predicts promising

applications in affective computing and interactive technologies; the approach demonstrates effectiveness in capturing emotional nuances, with adaptation to different contexts to make human-machine interactions more empathetic and personalized.

Keywords: Voice Recognition, Expressed Emotions, Voice Processing, Artificial Intelligence, Fuzzy Logic.

SUMÁRIO

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE FIGURAS

1 INTRODUÇÃO

A voz é parte integrante da individualidade humana. Por meio dela é possível externalizar sentimentos e pensamentos, além de estabelecer contato com outras pessoas e com o mundo (GUIMARÃES et al., 2010). A produção vocal é influenciada por uma combinação de fatores fisiológicos, psicológicos, ambientais e de aprendizado, como gênero, idade, constituição física, saúde geral, e fatores psicossociais, como personalidade e emoção (CIELO et al., 2009; ALMEIDA et al., 2015; GUZY et al., 2016).

A voz pode atuar como um sinal biológico para diferenciação do estado emocional, uma vez que varia em intensidade, frequência e ritmo. A voz pode expressar os traços de personalidade, sentimentos, estado de saúde física e mental, dentre outros aspectos (COSTA, et al 2013, YEH, et al 2016), com possibilidade de auxiliar em processos de detecção, avaliação, diagnóstico e monitoramento de condições neurológicas, psiquiátricas, cardiovasculares e respiratórias (IDRISOGLU et al., 2023; LOW et al., 2020; TRACEY et al., 2021). Assim, a voz transm ite informações relacionadas aos estados físicos e/ou emocionais do falante e oferece uma abordagem não invasiva e acessível para gerenciar diversas condições clínicas (AMORIM et al., 2021).

A face e a vocalização destacam-se entre os vários canais de expressão das emoções, pois, ao manifestar emoções uma pessoa transmite e comunica a outra o que sente, seja de forma intencional ou inconsciente, que pode impactar em diversos aspectos, na cognição, percepção, aprendizado e comunicação (ADRIANO; ARRIAGA, 2016). Essa expressão poderá surtir efeitos nas emoções e no comportamento do indivíduo que irá receber, a depender da sua sensibilidade e precisão no reconhecimento emocional (VIEIRA, 2018).

Diversos estudos realizados em diferentes culturas contribuíram para que alguns autores propusessem seis emoções básicas, a saber: felicidade/alegria, medo, raiva, tristeza, repugnância/nojo e surpresa (EKMAN, 1999, LEVENSON, 2011; AN et al., 2017). Acredita-se que essas sejam universalmente reconhecidas na face pelos seres humanos, por apresentarem configurações específicas, expressas de forma semelhante em diferentes culturas (HALSZKA, 2016; AN, et al., 2017; WANG, et al., 2019). Essas emoções quando combinadas geram um espectro de estados emocionais (EKMAN, 1999; AN, et al., 2017). Cada estado emocional provoca modificações no trato vocal e alteram momentaneamente a anatomofisiologia de produção da voz, que interfere no controle da respiração, no posicionamento vertical da laringe, no relaxamento relativo das pregas

vocais e no posicionamento e no relaxamento dos músculos da faringe e da língua, que pode resultar em vozes alteradas (METALLINOU, et al., 2008)**.**

Essas modificações no trato vocal ocorrem na variação das emoções e vê-se um maior impacto negativo quando está relacionado a um transtorno mental, que aumenta as chances de instalação de uma alteração vocal, como a disfonia (DIETRICH, VERDOLINI-ABBOTT, 2012; SOUZA; HANAYAMA, 2005). Nessa mesma linha de entendimento, os pesquisadores que estudam a expressão das emoções a partir da voz assumem que essas também possuem padrões distintos nos parâmetros acústicos (MANOHAR, et al., 2019; BHATT, et al., 2021)

A sensibilidade à emoção do outro pode ficar comprometida em algumas pessoas, como nos profissionais expostos a níveis de estresse continuado que, por vezes, culmina em implicações negativas no desempenho profissional e social (NEUMANN et al., 2011; PASSALACQUA; SEGRIN, 2012; WEST et al., 2009). O reconhecimento das emoções é essencial para profissionais que lidam diariamente com outras pessoas e fazem julgamentos sobre o seu estado físico e emocional, além do que o reconhecimento das emoções é essencial para a comunicação e vida em sociedade.

O conhecimento sobre a variação emocional e a construção de bancos de vozes para diferenciação das emoções, vem sendo amplamente estudados. Assim, faz-se necessário o desenvolvimento de bancos de dados de sinais de voz que sejam validados por humanos para que se possa fazer um estudo baseado em emoções identificadas pela voz. Os bancos de dados de vozes utilizados para reconhecimento de emoções podem ser elaborados de forma simulada/atuada, indução de emoções, ou coleta de emoções em situações cotidianas/reais (BURKHARDT et al., 2005; BARRA-CHICOTE et al., 2008; DOUGLAS-COWIE; COWIE; SCHRÖDER, 2000; SNEDDON et al., 2012, KINGESKI, 2019).

Existem alguns bancos de vozes que foram desenvolvidos com a variação das emoções com a população de atores em diferentes línguas, como o *Berlin Database of Emotional Speech* (EMO-DB) (BURKHARDT et al., 2005), *Interactive Emotional Dyadic Motion Capture* (IEMOCAP) (BUSSO et al., 2008), *Sustained Emotionally colored Machine-human Interaction using Nonverbal Expression* (SEMAINE) (MCKEOWN et al., 2012) e *Remote Colaborative and Affective interactions* (RECOLA) (RINGEVAL et al., 2013). De forma geral, esses bancos de vozes foram construídos com informações da análise das variações acústicas emocionais e pouco se tem sobre outros tipos de medidas, como: aprofundamento nas medidas perceptuais por parte de fonoaudiólogos e linguistas

e no impacto do julgamento por ouvintes especialistas em ciências vocais, além da possibilidade de encontrar reconhecimento de padrões de voz comuns a cada uma das emoções.

Dado a escassez de bases de áudio de emoções em português brasileiro (PB) (TORRES NETO et al., 2018; KINGESKI, 2019; GERMANO et al., 2021), recentemente no Brasil foi elaborado um banco de vozes com informações sobre os diferentes estados emocionais que contemplasse dados perceptuais, acústicos e de fala de indivíduos nativos do PB, o banco de vozes nas variadas emoções para o português brasileiro - EmoVox-BR (LIMA, 2022). Os atores participaram das gravações de amostras vocais com a expressão simulada das emoções alegria, medo, tristeza, raiva, surpresa, nojo e a emissão neutra. Na validação, 39 áudios foram hospedados no banco EmoVox-BR por possuírem alta confiabilidade quanto à qualidade dos áudios, identificação das emoções, valência e potência das emoções após julgamento e avaliação de juízes fonoaudiólogos.

Um dos grandes desafios no estudo da expressão das emoções pela voz é identificar qual a melhor tarefa de fala e medidas vocais sensíveis para diferenciar as emoções. Há estudos que utilizam tarefas de fala como a leitura de frases padronizadas, fala espontânea, repetição de palavras, histórias curtas ou narrativas pessoais, imitação de emoções e vocalização sustentada, e que selecionam medidas fisiológicas e vocais para verificar o estado emocional (FITCH, 1990; BURIN, 2017; PATEL et al., 2011; MOON et al., 2012). Pode-se dizer que as medidas vocais mais recorrentes nesses estudos são as medidas acústicas, já que esses parâmetros capturam de forma mais objetiva as variações nas características vocais associadas a diferentes estados emocionais.

As medidas mais utilizadas para o reconhecimento das emoções por meio da voz são as medidas acústicas, como a frequência fundamental (f_0), o *jitter, shimmer*, formantes, coeficientes cepstrais e mel-cepstrais, intensidade do sinal e taxa de elocução da fala (VERVERIDIS; KOTROPOULOS, 2006; BYUN; LEE, 2021; SONG et al., 2024). A definição de características acústicas próprias de cada estado emocional permite a efetiva separação das emoções para o desenvolvimento de sistemas modernos de reconhecimento a partir da voz (LI, et al., 2019; BHATT, et al., 2021). As medidas acústicas tradicionais, como as de perturbação, ruído, cepstrais e espectrais, têm papel crucial na identificação e no monitoramento dos distúrbios vocais, pois fornecem informações sobre parâmetros físicos da voz. No entanto, quando o objetivo envolve diferenciar variações sutis dentro da normalidade vocal, outras análises, como as prosódicas e formânticas, complementam e ampliam esse entendimento. Essas medidas adicionais investigam as

nuances rítmicas e melódicas da fala, essenciais para capturar aspectos transitórios e expressivos, fundamentais para a percepção emocional e a expressividade da voz (BARBOSA; CONSTANTINI, 2020; SANTOS et al., 2022).

A investigação da prosódia emocional envolve tanto a forma como as modulações acústicas são produzidas quanto a maneira como são percebidas (a partir do julgamento perceptivo-auditivo), com possibilidade de uma compreensão mais ampla da relação entre os processos linguísticos e emocionais na comunicação humana, que ultrapassa o conteúdo semântico das palavras (SILVA; BARBOSA, 2017). As medidas acústico-prosódicas comumente utilizadas para o reconhecimento das emoções por meio da voz são f_0, intensidade e duração (FILIPPA et al., 2022; LAUSEN; HAMMERSCHMIDT, 2020). Compreender esses elementos é essencial para desvendar a complexidade da comunicação humana, na qual a dimensão emocional influencia significativamente a interação social. Além de seu impacto teórico, o estudo da prosódia emocional apresenta aplicações práticas em tecnologias avançadas de reconhecimento de emoções, inteligência artificial, e diagnósticos clínicos de transtornos da comunicação e saúde mental (OH et al., 2023; FILIPPA et al., 2022).

Para isto, utiliza-se a combinação de métodos de extração e reconhecimento de características por meio de sistemas de inteligência artificial. Estes sistemas são responsáveis por produzir uma representação do sinal de voz na identificação de emoções, que podem auxiliar mercados como *callcenter*, aplicativos de tutorial de computador, *web* filmes, comunicação móvel, transcrição de discursos gravados, pesquisa em documentos de áudio, comandos de voz, diálogos interativos, entre outros sistemas automáticos de detecção por voz (GALES, 2008; ZHU; WU, 2020; LI; LIN, 2021).

As análises de expressões emocionais vocais demonstraram que a intensidade e a valência variam conforme a emoção expressa. Cada emoção apresenta um perfil vocal específico, o que possibilita sua distinção (Banse, Scherer, 1996). As medidas relacionadas à qualidade vocal contribuem para a identificação de estados emocionais (Nenko et al., 2021; Laukka, Elfenbein, 2012), principalmente nas emoções de valência negativa como tristeza e raiva (Nenko et al., 2021). As variações nas medidas de f_0 e de qualidade vocal foram relevantes para trazer informações emocionais e contextuais mais complexas, o que inclui a intensidade e a potência da emoção (Laukka, Elfenbein, 2012). Os Mel-Frequency Cepstral Coefficients (MFCC) mostraram-se sensíveis à distinção entre vozes de indivíduos deprimidos e saudáveis, o que reforça o potencial da voz como biomarcador para o diagnóstico emocional (Zhao et al., 2022).

Como exposto, há estudos que classificam emoções a partir da voz, sobretudo com base em medidas acústicas, com tarefas de fala sem padronização e em diversas línguas. Desenvolver um modelo que utilize a inteligência artificial, no intuito de procurar fatores preditores para o reconhecimento das emoções básicas é importante para chegar a um conjunto de medidas que caracterize e discrimine as emoções a partir da voz, coletada dentro de um rigor e padrão científico. Portanto, para a efetiva aplicação de reconhecimento de estados emocionais é necessário que haja um sólido estudo com bases de dados que abrange inicialmente a cultura local, de forma que os dados sejam verificados com outras bases de vozes gravadas em outras culturas e, assim, a validação dos métodos e medidas aconteça. Posteriormente, a construção de um modelo de reconhecimento dos estados emocionais a partir da voz.

A construção de modelos que utilizem o sinal de voz em diferentes emoções, servirão de base para elaboração e desenvolvimento de sistemas de inteligência artificial que sejam robustos na detecção por voz e identificação de emoções. O modelo a ser proposto permitirá o desenvolvimento de uma ferramenta automatizada no reconhecimento das emoções com precisão conhecida, o que possibilita ao examinador um avanço potencialmente significativo na identificação do estado emocional. Assim, diversos tipos de mercado podem aplicar este tipo de tecnologia de modo a entender as necessidades do indivíduo de acordo com sua emoção no determinado momento e assim beneficiar e esses sistemas de atendimento.

A lógica *fuzzy* se destaca como uma abordagem robusta para modelar sistemas complexos e incertos, como o reconhecimento de emoções na voz (XIONG, et al., 2024; TON-THAT; CAO, 2019; GRIMM et al., 2007). Diferente dos modelos tradicionais que dependem de limiares rígidos, a lógica *fuzzy* permite uma representação mais flexível e próxima da realidade, incorporando o conhecimento e a experiência humana na análise de padrões contribuindo para a tomada de decisão (MORAES; MELO, 2017). Essa capacidade de lidar com incertezas e subjetividades a torna uma ferramenta promissora para a modelagem preditiva no reconhecimento de estados emocionais.

A aplicação da lógica *fuzzy* neste contexto justifica-se pela necessidade de um modelo que não apenas interprete dados vocais de forma precisa, mas que também seja capaz de captar nuances emocionais que podem escapar a técnicas mais rígidas (MORAES, 2002). Este trabalho pretende explorar o potencial da lógica *fuzzy* para criar um modelo de reconhecimento emocional mais eficaz e adaptável, capaz de melhorar a interação entre humanos e máquinas e contribuir para avanços na área de saúde mental

e tecnologias assistivas. Além de gerar conhecimento que transcende o conjunto de dados específico utilizado para seu desenvolvimento, o modelo proposto busca oferecer uma base metodológica com potencial de generalização e aplicação em outros bancos de dados, ampliando a validade externa e a aplicabilidade dos resultados obtidos.

A identificação automática de estados emocionais a partir da voz têm se tornado uma área de crescente interesse na interseção entre tecnologia, saúde e comunicação (DE LOPE; GRAÑA, 2023, HASHEM et al., 2023). Esse reconhecimento é fundamental em diversas aplicações, como assistentes virtuais, monitoramento de saúde mental, sistemas de atendimento ao cliente e interfaces homem-máquina. Entretanto, a complexidade intrínseca dos sinais de voz, aliada à variabilidade individual e cultural das expressões emocionais, torna essa tarefa desafiadora.

Ao priorizar o conhecimento humano no desenvolvimento do modelo, o sistema proposto torna-se mais intuitivo, confiável e alinhado às complexidades das interações emocionais humanas, para oferecer uma solução inovadora e eficaz para o reconhecimento automático de emoções. Um modelo construído com base nas características vocais e acústicas dos sinais de áudio em cada estado emocional pode proporcionar uma maior taxa de acurácia na diferenciação e reconhecimento das emoções a partir da voz humana em comparação com métodos tradicionais.

Não foram identificados na literatura estudos que desenvolvessem modelos de reconhecimento dos estados emocionais construídos com as características perceptuais, acústicas e prosódicas julgadas por fonoaudiólogos especialistas em voz em uma base de áudios em PB, uma vez que a variação cultural e linguística pode influenciar significativamente a expressão e percepção das emoções. Modelos de reconhecimento que levam em consideração essas limitações são fundamentais para auxiliar de forma eficaz e confiável profissionais e pesquisadores na área de voz. Esses modelos possibilitam a automatização da identificação e diferenciação das emoções a partir da voz humana, sendo capazes de extrair informações suficientes para identificar com precisão a emoção que uma pessoa está sentindo.

2 OBJETIVOS

2.1 Objetivo Geral

- Desenvolver um modelo de reconhecimento das emoções baseado em lógica *fuzzy* a partir da voz de falantes do português brasileiro.

2.2 Objetivos Específicos

- Identificar e selecionar os principais parâmetros vocais que serão utilizados na análise para o reconhecimento das emoções.

- Verificar o desempenho do modelo *fuzzy* nas amostras vocais nas variadas emoções;

- Verificar os parâmetros vocais indicadores para cada emoção por meio de um modelo de decisão baseado em lógica *fuzzy*;

- Implementar o método baseado na lógica *fuzzy* como um modelo preditivo do estado emocional por meio da análise vocal, a partir de regras e funções de pertinência que relacionem as medidas vocais aos estados emocionais;

- Realizar o treinamento do modelo de reconhecimento de emoções utilizando os dados do EmoVox-BR;

- Testar e validar o desempenho do modelo de reconhecimento de emoções, identificando as taxas de acerto para diferentes emoções e ajustando os parâmetros *fuzzy* para otimizar a predição.

- Comparar o modelo baseado em lógica *fuzzy* com outros métodos de reconhecimento de emoções para avaliar as vantagens e limitações dos diferentes modelos.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Voz e Emoção

A voz reflete a personalidade e as emoções do indivíduo, funcionando como um indicador do estado emocional. É na expressão emocional dos gestos e da voz que se encontram os primeiros indícios da linguagem como um fenômeno autêntico, já que a emoção é uma variação de nosso ser no mundo, e manifesta aquele mesmo poder de ordenar os estímulos e situações que estão no seu auge no plano da linguagem (SILVA, 2009). A personalidade e a condição emocional influenciam a níveis de ansiedade do indivíduo, que por sua vez podem ser expressos no comportamento vocal. Por outro lado, este fato significa que as modificações e perturbações vocais podem ser entendidas como uma reação a um estresse emocional (GOMES et al., 2019; TRAJANO et al., 2016; ALMEIDA et al., 2014; COSTA et al., 2013; ALMEIDA et al., 2011).

A voz funciona como um biomarcador, por meio da análise de indicadores e de características típicas da produção vocal. A inteligência artificial permite comparar diversas de vozes de indivíduos saudáveis para estabelecer um padrão de voz de sujeitos com distúrbio. A partir de então, é possível fazer a detecção e diferenciação de algumas características próprias. Esses indicadores são atributos que não são normalmente percebidos pelo ouvido humano, como características das oscilações da prega vocal e flutuações da ressonância do trato vocal (boca, nariz, faringe etc.) que podem variar de acordo com cada estado emocional (MONTICELLI; OTTA 2021). Somente por meio de análise de dados vocais em programas computacionais específicos para esse fim é possível fazer uma investigação detalhada.

Os seres humanos dispõem de cognição social e de inteligência emocional, as emoções presentes em seu cotidiano podem agir diretamente, entre diversos aspectos, na comunicação, cognição, aprendizado e percepção. Um evento inesperado, por exemplo, pode ser motivado por algo que gere a sensação de felicidade, ou pode ocorrer uma surpresa negativa, causando sensações desagradáveis no indivíduo, como medo e estresse. Sem dúvida, esses acontecimentos geram estímulos neurais que afetam, entre outras questões, a tomada de decisão de curto prazo (VIEIRA, 2018). A emoção poderia ser definida como uma condição complexa e momentânea que surge em experiências de caráter afetivo, provocando modificações em várias áreas do funcionamento psicológico e fisiológico, preparando o indivíduo para a ação.

Segundo a literatura, as emoções podem ser divididas em abordagem discreta e abordagem contínua. A abordagem discreta direciona que há um grupo de seis emoções básicas, conhecido como “Big-Six”: raiva, felicidade, surpresa, nojo (ou desgosto), medo e tristeza, que, quando combinadas, geram um espectro de estados emocionais (EKMAN, 1999). Já a abordagem contínua agrupa as emoções em diferentes eixos de dimensões. O modelo mais conhecido, neste sentido, é o modelo tridimensional de Schlosberg (1954). Cada estado emocional pode ser definido como uma combinação linear dos eixos ativação (ou excitação), valência (ou avaliação) e potência (ou poder). Ativação mede o grau de excitação do indivíduo em expressar a emoção. Valência quão positiva ou negativa é a emoção. Potência diz respeito à força da emoção (RUSSELL, 1980).

O comportamento comunicativo dos indivíduos pode ser influenciado pelo estado emocional, sendo possível perceber inúmeras vezes as variações emocionais com base na escuta de suas vozes. As emoções provocam mudanças fisiológicas no corpo humano que se refletem na voz e fala. Variações na velocidade de fala e controle respiratório podem ser influenciados pelo estado emocional, essas variações possuem efeitos na pressão supraglótica que, por sua vez, levarão a modificação de intensidade e até da frequência de fonação (SUNDBERG, 2015), como também pode alterar a velocidade da fala, a coordenação pneumofonoarticulatória, a inteligibilidade de fala e a fluência com a presença de hesitações e repetições de palavras (BARBOSA; FRIEDMAN, 2007). As emoções de valência negativa podem provocar ressecamento da mucosa oral interferindo na ação de fala, afetar o controle da musculatura oral e articulação dos fonemas, e de poder provocar temores. Assim, o estado emocional provoca modificações na atividade comunicativa e não apenas na voz (SUNDBERG, 2015).

A análise do sinal de voz é um importante método para lidar com o reconhecimento do sinal e interpretação das informações contidas nele. Um estudo de revisão observou que o reconhecimento da emoção pela voz aborda questões importantes, como os recursos, características e métodos utilizados nas pesquisas sobre o tema. Foram encontrados 64 artigos e os autores verificaram alguns dados importantes: a maioria dos estudos tinham como voluntários população em geral, atores e estudantes; objetivavam o reconhecimento de emoções; o tipo de amostra vocal era coletada a partir de simulação, seguida de forma natural e eliciada; um quarto dos estudos utilizaram medidas neurofisológicas para mapear o estado emocional dos voluntários de pesquisa; e houve uma maior preocupação em avaliar emoções de valência negativa. De forma geral, verificaram que as medidas mais utilizadas para o reconhecimento das emoções foram as

variações da f_o, formantes, coeficientes mel-cepstrais, intensidade do sinal e taxa elocução da fala (VERVERIDIS, KOTROPOULOS, 2006).

No que se refere às medidas para o reconhecimento das emoções na voz, Ververidis e Kotropoulos (2006) afirmam que o *pitch* (percepção subjetiva da f_o) depende da tensão das pregas vocais e a pressão de ar subglótica e que pode ocorrer a seguinte variação: *pitch* elevado/agudo está relacionado à alegria ou surpresa e *pitch* baixo/grave à raiva ou desgosto. Apontam ainda que os harmônicos são importantes devido ao fluxo de ar não-linear no trato vocal que produz o sinal de fala, com isso verificaram, em espectrograma de banda larga, que a raiva e estresse apresentam fluxo de ar rápido próximo das pregas vestibulares, fornecem sinais de excitação adicional. Complementam que a forma do trato vocal muda devido ao estado emocional e os formantes são a representação de ressonância do trato vocal; em relação a esse tipo de parâmetro, viu-se que em momento de estresse ou depressão comumente não articulam sons das consoantes com o mesmo esforço que o estado emocional neutro. Por fim, direcionaram que as medidas acústicas a curto prazo (média, intervalo e desvio padrão (dp) da f_o, *pitch,* intensidade, *jitter e shimmer*), vão repercutir na taxa de elocução e pausas entre enunciados, assim, encontraram que homens com raiva mostraram níveis mais elevados de energia e falavam mais pausadamente (VERVERIDIS, KOTROPOULOS, 2006). Contudo, compilaram as variações de fala de cinco das seis emoções básicas (alegria, tristeza, medo, raiva, surpresa, nojo): na alegria o *pitch* fica agudo ascendente, intensidade forte e pausas curtas; no medo o *pitch* fica agudo, intensidade forte e pausas curtas; na raiva o *pitch* fica mais agudo, intensidade mais fraca nos homens, forte nas mulheres e velocidade de fala aumentada para mulheres; e no nojo o *pitch* fica grave, intensidade fraca e velocidade de fala reduzida (BURKHARDT et al., 2005; VERVERIDIS, KOTROPOULOS, 2006).

Alguns bancos de vozes internacionais foram desenvolvidos com intuito de analisar variações acústicas emocionais, dentre eles estão a base EMO-DB (BURKHARDT et al., 2005) que foi desenvolvida na Universidade Técnica de Berlim, na Alemanha. Participaram 40 atores das gravações de amostras vocais, em alemão, de 6 emoções: raiva, tédio, nojo, medo, felicidade, tristeza e neutro. A base IEMOCAP (BUSSO et al., 2008) foi desenvolvida na Universidade do Sul da Califórnia, nos Estados Unidos. Os sinais acústicos foram realizados a partir das interações dois a dois, com 10 atores (5 homens e 5 mulheres), onde 7 eram atores profissionais e 3 eram alunos sênior no Departamento de Drama. As gravações foram no idioma inglês, e as emoções gravadas

foram: raiva, felicidade, tristeza e emissão neutra. A base SEMAINE (MCKEOWN et al., 2012) é audiovisual e foi desenvolvida a partir de uma cooperação de Universidades da Inglaterra, da Holanda e da Alemanha. Participaram da coleta 150 estudantes de graduação e pós-graduação de oito diferentes países, todos padronizados no idioma inglês. As gravações foram feitas a partir de interações realizadas dois a dois, onde o controle ficava sempre a cargo de um dos dois. As emoções foram divididas em 27 estilos, dentre raiva e tristeza, além de outros comportamentos. A base RECOLA (RINGEVAL et al., 2013) foi desenvolvida na Suíça, e utilizou-se o idioma francês, apesar de terem participantes nativos de outros idiomas. Um total de 46 participantes foram submetidos a interações entre os dois. Antes de haver a interação entre os participantes, eles foram submetidos a um questionário de autoavaliação emocional conhecido como *Self-Assessment Manikin* (SAM), que está relacionado com a valência das emoções. Os aplicadores do questionário decidiram quais participantes iriam ser induzidos com humor positivo ou negativo, de acordo com o SAM. Nesse, capturaram sinais acústicos e biológicos, onde posteriormente extraíram dados dos sinais acústicos quanto à valência e ativação.

Em relação as bases nacionais, o VERBO (*Voice Emotion Recognition dataBase in Portuguese*) é um banco de vozes emocional desenvolvido em PB, que contou com a amostra de 12 atores brasileiros (6 homens e 6 mulheres), que gravaram 14 frases pré-definidas e adaptadas para o idioma, representando 7 emoções: alegria, nojo, medo, raiva, surpresa, tristeza e estado neutro. No total, foram realizadas 1167 gravações. A validação foi conduzida por três psicólogos, com resultados que apontaram uma concordância de 65% pelo teste Kappa de Fleiss. As emoções mais facilmente reconhecidas foram raiva e alegria, enquanto nojo e surpresa apresentaram maiores desafios de identificação (TORRES NETO et al., 2018).

Desenvolvido pela Universidade Estadual do Rio de Janeiro, o emoUERJ incluiu 8 atores, sendo 4 homens e 4 mulheres, que gravaram um total de 377 áudios. Esses áudios foram baseados em 10 frases, escolhidas livremente pelos participantes, representando quatro emoções: felicidade, raiva, tristeza e estado neutro. O processo de validação incluiu a criação de um conjunto de testes composto por vídeos coletados em uma plataforma de vídeo *online* com amostras que representavam variações linguísticas regionais do Brasil, como baiana, mineira, carioca e gaúcha, organizadas por gênero e emoção (raiva, felicidade e estado neutro) e utilizou como medida vocal o MFCC. Os resultados mostraram melhor desempenho na variação carioca (40% de acurácia) e pior na gaúcha

(7%), com vozes femininas sendo mais bem reconhecidas e maior dificuldade em emoções sutis, como tristeza (GERMANO et al., 2021).

O banco de vozes em PB apresentado no estudo Kingeski (2019) foi desenvolvido com gravação realizada em estúdio, abrangendo seis emoções básicas: felicidade, tristeza, raiva, nojo, medo e surpresa. Além das transmissões causadas por meio de vídeos e atividades, o banco incluiu áudios encontrados de vídeos da internet e de filmes brasileiros e estrangeiros dublados em PB. No processo de validação, 200 frases foram selecionadas e avaliadas por 20 pessoas aleatórias, que identificaram as emoções apresentadas nos áudios, com um desempenho médio de acerto superior a 79%.

Poucas informações são fornecidas por essas bases sobre outros tipos de medidas utilizadas, como: aprofundamento nas medidas perceptuais por parte de juízes, padronização da forma de coleta, além da possibilidade de encontrar reconhecimento de padrões comuns a cada uma das emoções e poder relacionar ao impacto no ouvinte. De toda forma, ainda não há um consenso a respeito de um atributo acústico, de qualidade vocal ou prosódicos que sinalize diferentes estados emocionais. Assim, a definição de um atributo que represente de forma significativa informações relacionadas ao comportamento fisiológico dos estados emocionais é uma busca crucial.

Recentemente um banco de vozes brasileiro nas variações das emoções (EMOVOX- BR) foi desenvolvido e validado com 182 áudios que apresentaram altos níveis de concordância inter-juízes, representando emoções como alegria, medo, tristeza, raiva, surpresa, nojo e estado neutro (LIMA, 2022). Construído remotamente com a participação de 26 atores que simularam emoções em frases foneticamente balanceadas, o EmoVox-BR não apenas estabelece um repositório de vozes nas variações emocionais, mas também fornece um levantamento detalhado das características vocais e prosódicas que distinguem as emoções, além de envolver uma ampla variedade de emoções, essencial para estender o alcance e a aplicabilidade dos bancos de vozes. O julgamento perceptivo-auditivo, conduzido por fonoaudiólogos especialistas, identificou marcadores vocais específicos de cada estado emocional, enriquecendo o entendimento de medidas vocais das emoções na população brasileira. A validação do banco se destaca pela qualidade dos áudios e sua confiabilidade, tornando-o uma ferramenta valiosa para aplicações em contextos clínicos, estudos populacionais e desenvolvimento tecnológico, como sistemas de reconhecimento emocional. Além disso, sua diversidade e aplicabilidade prática ampliam seu impacto nas áreas de fonoaudiologia e computação afetiva, contribuindo significativamente para a pesquisa e inovação em cenários reais.

3.2 Modelos de reconhecimento a partir da Voz

A análise do sinal de voz é um importante método para realizar o reconhecimento das emoções e forte indicador para diversos tipos de transtornos mentais. Embora ainda esteja em estágios iniciais, essa área tem mostrado promessa como uma abordagem não invasiva e objetiva para auxiliar na detecção precoce e no monitoramento de transtornos mentais. Algoritmos de inteligência artificial, como os que utilizam a lógica *fuzzy* e/ou métodos de aprendizado de máquinas são então empregados para tentar identificar os padrões distintos associados a cada transtorno, permitindo que os sistemas compreendam o estado emocional do falante (SINGH; GOEL, 2022).

A importância de técnicas computadorizadas tem sido cada vez mais enfatizada. Para melhorar as habilidades de comunicação dos profissionais de saúde com seus pacientes, existem sistemas que utilizam programas de reconhecimento de padrões que analisam a emoção presente no rosto e voz do profissional de saúde ou do paciente (MARTINS et al., 2020). A avaliação do estado emocional que os profissionais de saúde estão nas consultas pode facilitar o desenvolvimento de empatia com os pacientes.

O reconhecimento a partir da voz é essencialmente um problema de reconhecimento de padrões, realizado a partir de uma sequência de parâmetros que caracterizam o sinal de voz. O reconhecimento de padrões que é naturalmente aprendido pelos humanos vem sendo aprimorado por técnicas de inteligência artificial que tentam representar o padrão como um vetor numérico, chamado de vetor de similaridade. O padrão será catalogado de acordo com a maior semelhança entre ele e o vetor representativo. Isso pode ser alcançado por redes que se ajustam a novas informações, fornecendo boas respostas, mesmo com dados ausentes ou confusos (YU et al., 2019).

Um sistema de reconhecimento que utiliza a voz é capaz de transformar um sinal de voz em uma sequência de dados com a qual uma máquina irá reconhecer por critérios próprios as emoções rotuladas. Esse reconhecimento começa na fase de captação do sinal de áudio. Na fase de pré-processamento, existe a formação de um vetor característico do padrão a ser analisado e na eliminação de sinais redundantes. Na fase de processamento, inicia-se o reconhecimento de padrões (PARTILA et al., 2015). Portanto, o reconhecimento de emoções por meio da voz poderá ser eficaz quando realizado por modelos que considerem o conhecimento humano e permitam a tomada de decisões com base em processos que consideram a incerteza e a complexidade das emoções como a lógica *fuzzy*.

Esses processos aumentam a precisão final na detecção do emocional do indivíduo. Isso permite maior *feedback* sobre como as emoções são manifestadas e uma possível melhora na empatia e relacionamento com o outro, permitindo melhores resultados e adesão ao tratamento, aumentando o grau de satisfação do paciente (HEGDE et al., 2018). Dessa forma, observa-se que, na literatura científica, existe a necessidade de construção de sistemas computacionais que utilizem modelos de inteligência artificial para reconhecer os padrões comuns a cada uma das emoções (WANG; JO, 2007; EDDINS; SHRIVASTAV, 2013). A construção dessas ferramentas poderá auxiliar o clínico nos procedimentos de avaliação e monitoramento de cada estado emocional, o qual pode aumentar as chances de instalação dos desvios da voz.

Do ponto de vista social, a importância da expressão emocional por meio da comunicação falada perpassa por meio da intenção do discurso, é fundamental para um poderoso sistema de reconhecimento por parte do ouvinte (SCHERER, 2003), expressa o estado emocional, sinaliza o quanto se percebe e aceita opiniões diferentes, bem como, como se antecipa o efeito do que se diz e da forma como se transmite a mensagem a quem se ouve. De forma geral, é um importante regulador social, além de, juntamente com a parte verbal, compõe a competência comunicativa (BORREGO; BEHLAU, 2018).

A voz provoca julgamentos, que mesmo acurados ou não, influenciam as interações sociais, a escolha de parceiros, líderes e até mesmo opções de consumidores (KLOFSTAD et al., 2012; TIGUE et al., 2012). Uma primeira impressão ruim pode ser provocada simplesmente por um desvio vocal presente (AMIR; LEVINE-YUNDOF, 2013) e um indivíduo portador de um distúrbio de voz pode ter dificuldades de expressão emocional e relacionais que não tem a ver com a sua personalidade ou com as emoções que está experimentando. Julgamento psicodinâmico, impacto da voz do falante no ouvinte, é um bom exemplo do quanto nos baseamos na voz para inferir os mais variados aspectos do falante.

Com a evolução da inteligência artificial e a recente popularidade do aprendizado de máquinas e do aprendizado profundo, os sistemas de reconhecimento a partir da voz estão se tornando mais poderosos e robustos. O treinamento tornou-se mais rápido e o número de falantes aptos a serem diferenciados tornou-se maior (NGUYEN; PAPRZYCKI; VOSSEN, 2024). Isso abre a oportunidade de treinar um sistema efetivo para reconhecimento das emoções a partir da voz e identificar os diferentes estados emocionais a partir dele, pois sistemas de treinamento podem aprender com uma infinidade de sinais de áudio (MICHALSKI et al., 2013; LECUN et al., 2015).

Para a efetiva separação das emoções, utiliza-se a combinação de métodos de extração e reconhecimento de características do sinal vocal por meio de sistemas de inteligência artificial. Estes sistemas são responsáveis por produzir uma representação do sinal de voz na identificação de emoções, o que pode gerar impacto de inovação tecnológica em diversos tipos de mercado (PARTILA et al., 2015).

Algumas possibilidades ilustram a ampla aplicabilidade desses sistemas na detecção de emoções e promovem avanços em diferentes setores. Em empresas de atendimento ao cliente, é possível identificar o nível de satisfação ou estresse do cliente em atendimento, além de verificar se o funcionário teve influência sobre aquele estado emocional. Em profissões que geram maior pressão emocional, como policiais, bombeiros e militares, esses sistemas podem ajudar a evitar ordens ou atitudes influenciadas pelo estado emocional. Outra aplicação está em sistemas de acesso por voz a ambientes privativos, permitindo identificar se uma pessoa está sendo coagida a acessar o local. Em videoconferências com tradução automática, esses sistemas podem detectar não apenas o que é falado, mas também como é falado, reduzindo desconfortos nas interações. Na área da saúde, o reconhecimento de emoções pela voz pode auxiliar na detecção de sinais de estresse, ansiedade ou depressão em pacientes. Assistentes de voz inteligentes também podem reconhecer as emoções dos usuários, tornando as interações mais personalizadas e empáticas. No campo do entretenimento, a detecção de emoções pela voz pode melhorar a experiência do usuário em jogos, adaptando a narrativa conforme as reações emocionais do jogador. Além disso, essas tecnologias impulsionam o desenvolvimento de dispositivos de tecnologia assistiva, como os de Comunicação Alternativa e Ampliada (CAA), ao adaptar a comunicação ao contexto emocional do usuário, promove maior expressividade e qualidade nas interações sociais e terapêuticas, especialmente em indivíduos com dificuldade severa de comunicação (GALES, 2008; ZHU; WU, 2020; LI; LIN, 2021; IRYAN et al., 2022).

O processo de reconhecimento de emoções a partir da voz envolve as seguintes etapas: Pré-processamento do áudio; Extração de características do sinal; Construção do modelo; Treinamento do modelo; Validação e ajuste e Teste do modelo (CIPRIANO, 2001; ZHANG et al., 2015). Na etapa de pré-processamento acontece a conversão analógico/digital, redução do ruído, filtragem e, se necessário, a segmentação em unidades menores, como frases ou palavras. Na segunda etapa são extraídos parâmetros vocais relevantes, que podem incluir medidas de f_0, de intensidade, medidas de ruído glótico, medidas de intensidade, medidas cepstrais e mel-cepstrais, medidas formânticas

e medidas prosódicas que podem ajudar a representar as características emocionais na voz (VERVERIDIS e KOTROPOULOS, 2006; PATEL et al., 2011). A terceira etapa envolve a construção de um modelo de inteligência artificial que será treinado para reconhecer as emoções presentes na voz com base nas características extraídas. Diferentes algoritmos, como os utilizam a lógica *Fuzzy*, e/ou métodos de aprendizado de máquinas como o *Support Vector Machine* (SVM), Redes Neurais Artificiais (RNA), Redes *Naive Bayes* (RNB), Random Forest, entre outros, podem ser usados para esse fim (DE LOPE; GRAÑA, 2023). Em seguida, o modelo é treinado usando o conjunto de dados rotulados para cada emoção. Durante o treinamento, o modelo ajusta seus parâmetros para aprender a mapear as características acústicas para as emoções correspondentes. Após o treinamento, é essencial validar o desempenho do modelo usando um conjunto de dados de validação separados. Isso ajuda a avaliar a precisão e a eficácia do modelo. Se necessário, o modelo pode ser ajustado e re-treinado para melhorar seu desempenho. Finalmente, o modelo é testado em um conjunto de dados de teste independente para avaliar sua capacidade de reconhecer emoções na voz de novos exemplos (HUANG et al., 2001).

É importante ressaltar que a precisão do reconhecimento de emoções na voz pode variar de acordo com a quantidade e qualidade do conjunto de dados de treinamento, a escolha do algoritmo/modelo e as características acústicas selecionadas. Além disso, as emoções são complexas e podem ser expressas de maneiras sutis e variadas, o que pode representar um desafio adicional na tarefa de reconhecimento. No entanto, com avanços contínuos em tecnologias de processamento de sinais, uso da lógica e do aprendizado de máquina, espera-se que o reconhecimento de emoções na voz continue melhorando e encontre aplicações práticas em diversas áreas (SINGH; GOEL, 2022).

Considerando a importância da análise da voz e suas modificações nos estados emocionais e transtornos mentais, a utilização de técnicas computacionais avançadas permite a investigação dos aspectos vocais como uma ferramenta promissora para o rastreio de forma objetiva dos sintomas das alterações psicológicas, assim, pode contribuir para otimizar o diagnóstico de diversos transtornos.

Um modelo computacional que consiga verificar os efeitos das emoções e a definição de características acústicas próprias de cada estado emocional por meio de técnicas de reconhecimento de padrões possibilita de forma efetiva a separação das emoções. Por meio de um sistema baseado em regras *fuzzy*, é possível traduzir a variabilidade das características acústicas em graus de pertinência para diferentes

emoções. Essas regras, organizadas em um modelo *fuzzy*, agregam informações de forma gradativa e atribui uma classificação emocional ao sinal analisado, mesmo em casos de dados limitados. Além disso, o sistema *fuzzy* torna-se eficiente em termos computacionais e com potencial para ser ampliado com técnicas de pré e pós-classificação (MORAES, 2002). Assim, mesmo diante de incertezas ou sobreposição entre estados emocionais, o sistema oferece uma resposta fundamentada nos graus de compatibilidade calculados, aumenta a qualidade e robustez das classificações e permite uma identificação robusta da emoção predominante em cada sinal vocal.

3.3 Medidas de voz

A extração de medidas acústicas e prosódicas da voz é uma ferramenta que fornece informações quantitativas e qualitativas do sinal sonoro e se torna um importante método para realizar o reconhecimento do sinal vocal e interpretação das informações contidas nele (TANDEL et al., 2020). Elas são amplamente utilizadas em pesquisas devido à sua objetividade e por extrair características do sinal de áudio responsáveis por gerar informações e traçados no formato de onda sonora. O primeiro passo em qualquer técnica de extração de características vocais é extrair e identificar os componentes do sinal de áudio que são adequados para identificar o conteúdo linguístico e descartar todas as outras partes que transportam informações desnecessárias como por exemplo ruído de fundo.

As medidas consideradas mais robustas para a discriminação de falantes e que oferecem uma perspectiva técnica e mensurável dos aspectos emocionais da voz são relativas as medidas de f_0, *Jitter, Shimmer,* HNR, GNE, Medidas Cepstrais (CPPS) e Mel-Cepstrais (MFCC) relacionadas a avaliação acústica, e as medidas de f_0, intensidade e duração, relacionadas a avaliação acústica- prosódicas, (LOPES et al., 2019, BARBOSA; CONSTANTINI, 2020) e a interpretação desses sinais permite entender a fisiologia da produção vocal.

As medidas acústicas tradicionais oferecem uma visão objetiva sobre a qualidade da voz. Esses parâmetros capturam variações na altura vocal, estabilidade e qualidade do som que podem indicar emoções específicas. A f_o é uma medida que fornece os números de ciclos glóticos produzidos em um segundo. Além de fazer parte da avaliação acústica tradicional, a f_o também desempenha um papel crucial na análise acústico-prosódica**.** Suas perturbações ou variabilidades ciclo a ciclo, de frequência e amplitude

são denominadas *jitter e shimmer* respectivamente, os quais apresentam-se alterados em pacientes disfônicos (TEIXEIRA et al., 2013). A HNR, que contrasta com o sinal regular das pregas vocais e do trato vocal, fornece um índice que relaciona o componente periódico (harmônico) versus o componente aperiódico (ruído) da onda acústica. Em outras palavras, essa medida analisa presença de ruído no sinal vocal (FERNANDES et al., 2018). O GNE é responsável pelo cálculo do ruído em uma série de pulsos produzidos pela vibração das pregas vocais na laringe. Portanto, esse parâmetro indica se o sinal vocal é proveniente da oscilação das pregas vocais ou da corrente de ar gerada ao longo do trato vocal. Por fim, as medidas CPPS e MFCC tem a função de evidenciar o quanto os harmônicos advindos da f_o são individualizados e se destacam em relação ao nível de ruído presente no sinal (LOPES et al., 2018; CONSERVA 2022; LOPES et al., 2019). Por esse motivo são consideradas as medidas importantes e confiáveis para avaliação vocal com ampla faixa de desvio e têm sido empregadas como parâmetros de classificação, em sistemas de reconhecimento a partir da voz, sistemas de identificação e verificação de locutor, como também na discriminação entre sinais de vozes afetados por patologias laríngeas (FECHINE, 2000; COSTA, 2008; BARAVIEIRA, 2016).

A análise dos aspectos acústico-prosódicos das emoções constitui uma área de crescente interesse nas ciências da fala e da comunicação, oferecendo novas perspectivas sobre como as emoções são expressas e percebidas por meio da voz (OH, et al., 2023; FILIPPA et al., 2022). A investigação da prosódia emocional envolve tanto a forma como as modulações acústicas são produzidas quanto a maneira como são percebidas, principalmente no comportamento dos parâmetros acústicos de f_o, duração e intensidade, proporcionando uma compreensão mais ampla da relação entre os processos linguísticos e emocionais na comunicação humana, ultrapassando o conteúdo semântico das palavras (SILVA; BARBOSA, 2017).

A f_o desempenha um papel central tanto na análise acústica quanto na análise acústico-prosódica da fala, porém, o foco e a aplicação da f_o em cada tipo de análise são diferentes. Na análise prosódica a f_o é avaliada de maneira mais dinâmica, focando em sua variação ao longo do tempo, em conjunto com outros aspectos prosódicos, como intensidade e duração (ARVANITI, 2020). A intensidade refere-se ao volume ou nível de pressão sonora da fala, ou seja, o quão alto ou baixo um som é produzido. É um aspecto fundamental da prosódia, pois está diretamente relacionado à forma como as emoções são expressas vocalmente (ARVANITI, 2020). A duração refere-se ao tempo que uma unidade de fala, como uma sílaba, palavra ou frase, é mantida. Ela pode ser entendida

como o tempo total gasto em uma produção vocal e também varia conforme o estado emocional (LARROUY-MAESTRI, 2023).

No estudo de Aguiar e colaboradores (2024), os autores investigaram a relação entre medidas acústico-prosódicas e a discriminação de diferentes estados emocionais em falantes do PB. Utilizaram 182 sinais de áudio que compõem o banco EmoVox-BR e que contém dados coletados de falantes nativos com expressão das emoções básicas como alegria, tristeza, medo, raiva, surpresa, nojo e emissão neutra. As análises focaram nos parâmetros de duração, f_o e intensidade. Os resultados indicaram variações significativas, destacou que o nojo apresentou maior duração e taxa de elocução, enquanto a alegria revelou menor duração e maior intensidade. O medo foi associado a menor variabilidade na duração e valores reduzidos de intensidade, enquanto a raiva exibiu a maior intensidade vocal. A surpresa apresentou os maiores valores de f_o. Concluiu-se que tais medidas acústico-prosódicas são ferramentas eficazes para diferenciar as emoções e forneceu suporte para aplicações em tecnologias de reconhecimento emocional e no entendimento da expressividade vocal no PB.

A análise cepstral demonstra ser uma possibilidade para análise de sinais com maior irregularidade, visto que as medidas cepstrais são capazes de determinar a f_0 e produzir estimativas de aperiodicidade e/ou ruído aditivo sem a necessidade de identificar limites de ondas sonoras individuais (LOPES et al., 2019). Os MFCC são considerados uma forma de análise cepstral ajustada para modelar a audição humana com base na escala Mel e têm sido amplamente aplicados em estudos de reconhecimento a partir da voz (MA; FOKOUÉ, 2014; PARTILA et al. 2015; BARAVIEIRA, 2016; FANG et al., 2019; TANDEL et al., 2020). Esses parâmetros são uma representação paramétrica do espectro de frequências do sinal de voz e surgiram devido a estudos que mostraram que a percepção humana das frequências de tons puros ou de sinais de voz que não seguem uma escala linear. Como o sinal de fala consiste em tons com frequências diferentes, para cada tom com uma frequência real f, medida em Hz, um tom subjetivo é medido na escala mel. O mel é uma unidade de medida da frequência percebida de um tom (COSTA et al., 2008).

Um estudo feito por Umapathy e colaboradores (2005) enfatiza a importância das medidas cepstrais como indicadores confiáveis da qualidade vocal, que podem ser mais eficazes do que as medidas acústicas tradicionais e relatou uma precisão geral de 93,4% de acurácia usando uma análise discriminante linear para detectar amostras de fala contínuas do banco de dados de distúrbios de voz do *Massachusetts Eye and Ear*

Infirmary (MEEI). A tecnologia de reconhecimento de voz baseada em MFCC tem sido destacada como importante para aplicativos de inteligência artificial incorporados, aumentando a velocidade e a precisão em dispositivos como alto-falantes inteligentes e smartphones (WANG et al., 2024). A integração do MFCC no processamento em tempo real possibilita o reconhecimento da voz eficiente mesmo em ambientes sem conexão de rede, evidenciando sua versatilidade (WANG et al., 2024). Contudo, embora amplamente reconhecidos por sua eficácia no reconhecimento de a partir da voz, alguns pesquisadores apontam a limitada interpretabilidade dos MFCC como um desafio, restringindo seu uso em diagnósticos clínicos e aplicações específicas (TRACEY et al., 2023).

Apesar do sucesso dos estudos mencionados, os sistemas para reconhecimento das emoções a partir do sinal da voz ainda enfrentam desafios relacionados à definição de quais parâmetros vocais são mais eficazes para capturar as emoções de forma precisa e universal (LAUSEN; HAMMERSCHMIDT, 2020; TRACEY et al., 2023). Além disso, diferenças linguísticas e culturais podem influenciar a manifestação vocal das emoções, tornando necessário um aprofundamento nas pesquisas para identificar as medidas mais adequadas em diferentes idiomas e contextos. Essa variabilidade representa um obstáculo, mas também uma oportunidade para desenvolver modelos mais robustos e adaptáveis, capazes de ajustar suas respostas conforme o estado emocional do usuário para promover interações mais humanas e eficientes.

3.4 Reconhecimento das emoções a partir da voz

O estado emocional da mente humana se expressa por diferentes formas, incluindo sinais de voz. O reconhecimento da emoção pode ser realizado por meio da identificação das características vocais nas variadas classes de emoções. Normalmente, um sistema supervisionado pré-treinado com características vocais de cada emoção como entrada e classe de emoção como saída é usado para determinar a classe de emoções com objetivo de criar um modelo emocionalmente inteligente com capacidades além do ser humano, que entende o sentimento do usuário e gera uma resposta de acordo (SINGH; GOEL, 2022). O reconhecimento automático de emoções por meio da voz é realizado processando-se um arquivo de áudio, identificando as características da voz contidas nos sinais de áudio e classificando-a como pertencente a alguma emoção conhecida (IRIYA, 2014).

Alguns estudos utilizaram métodos de inteligência artificial e as dimensões do espaço de emoções para o reconhecimento por meio da voz. Tato e colaboradores (2002) utilizaram um sistema de reconhecimento a partir da voz de diferentes naturezas para as dimensões da avaliação e da ativação para identificar cinco emoções: felicidade, raiva, tédio, neutro e tristeza. Os resultados mostraram que ao se utilizar um único conjunto de parâmetros de voz em um único sistema, as emoções mais próximas, principalmente na dimensão da ativação, tendem a se confundir frequentemente. No estudo de Lugger e Yang (2008), utilizam seis emoções: felicidade, raiva, tédio, nojo, medo e tristeza, que são reconhecidas por meio de um sistema de reconhecimento hierárquico com três estágios, passando pelas dimensões: ativação, domínio e avaliação. Os autores comparam o desempenho do sistema sequencial com o sistema de um único estágio, encontrando uma melhoria considerável de 14%. Xiao et al (2007) apresentam um sistema hierárquico com as mesmas seis emoções, porém a teoria de emoções é levada em conta apenas para a dimensão da ativação. Os autores ainda utilizam um primeiro estágio adicional para separação prévia do gênero, aumentando o desempenho em cerca de 3% em relação ao cenário independente do gênero.

Eventualmente, com a crescente popularidade das RNA em todos os campos de pesquisa, muitas outras abordagens de aprendizado de máquinas têm sido investigadas, como em Mirsamadi et al. (2017), onde *Convolutional recurrent Neural Network* (CRNN) foram aplicadas para extrair recursos, bem como para tarefas de reconhecimento, alcançando uma taxa de acurácia de 61,8% nas cinco emoções representadas (raiva, felicidade, frustração, tristeza e neutralidade) na base IEMOCAP.

Em uma revisão de literatura realizada por Singh e Goel (2022), alguns modelos foram avaliados pelos pesquisadores para o reconhecimento a partir da voz em busca da melhor precisão. Os sistemas de reconhecimento comumente utilizados foram SVM, *Gaussian Mixture Models* (GMM), *Hidden Markov Models* (HMM), *KNearest Neighbors* (KNN) e RNA. Pode-se concluir que 54,45% dos 152 artigos analisados, utilizaram o SVM como sistema e a precisão de 98% foi alcançada usando recursos MFCC no banco de dados EMODB. A maioria dos artigos utilizaram os MFCC ou MFCC com a combinação de outros recursos acústicos para reconhecimento de emoções.

No estudo de Kerkeni et al. (2019), os autores apresentam dois modelos de aprendizado de máquina para o reconhecimento de sete emoções distintas: alegria, raiva, tristeza, surpresa, medo, desgosto e emissão neutra. O estudo faz uso de atributos baseados em modulação AM-FM e análises não-lineares combinando MFCC baseados

em energia de *Teager* (TEMFCCs) com decomposição do modo empírico. Os sistemas de aprendizado de máquina utilizados foram CRNN e o SVM. Foram utilizadas duas bases de dados, uma em Espanhol, e outra em Alemão. Para a diferenciação do sinal da voz, se torna de grande importância a extração de parâmetros cepstrais e de modulação AM-FM. Baseando-se nos resultados e análises dos experimentos do estudo em questão, o autor conclui que para que um sistema baseado em CRNN tenha bons resultados de reconhecimento de emoções, é necessária uma maior extração de parâmetros e um tempo maior de treinamento do sistema. Já o sistema SVM demonstrou um maior potencial para seu uso prático devido a essas características.

O trabalho apresentado por Jain et al. (2020) aborda um sistema de reconhecimento de emoções a partir da utilização do SVM. Neste estudo, foram utilizados os estados emocionais de tristeza, raiva, medo e felicidade para o reconhecimento das emoções a partir de amostras vocais em uma base de dados em alemão. Os parâmetros extraídos de cada amostra foram energia, *pitch*, MFCC, Coeficientes Cepstrais de Predição Linear (LPCCs) e taxa de fala. O autor, por fim, concluiu que, a partir dos resultados obtidos em seus experimentos, a extração dos MFCC junto ao sistema alcançou a melhor taxa de 93% de acurácia para o reconhecimento de emoções.

Sandhya et al. (2020) propõem em seu estudo métodos para a identificação de locutores sob a influência de emoções na voz. O autor argumenta que sistemas de identificação de locutores funcionam bem em condições neutras, mas se deterioram em condições emocionais, devido ao impacto que as emoções causam ao sinal de voz. Neste estudo foram extraídas diversas características cepstrais da voz como os MFCC e TEMFCCs a fim de analisar quais apresentam melhores resultados entre diversos sistemas. Em seus resultados, o autor obtém uma melhor acurácia de 100% em condições neutras e 87,0967% em condições emocionais.

Em seu trabalho, Iriya (2014) aborda o reconhecimento das emoções nos sinais de voz com diferentes parâmetros como f_0, energia de curto prazo, formantes e coeficientes cepstrais. Foram utilizados diversos modelos para a comparação entre eles, tais como KNN, SVM, GMM e HMM, tendo o GMM como o principal devido ao seu desempenho e custo computacional. Iriya fez o uso de um sistema de reconhecimento de estágio único e um sistema sequencial em três estágios baseado na teoria de emoções da área de psicologia e conseguiu uma taxa de 83% de acurácia na identificação das emoções raiva, tristeza, tédio e neutra.

O trabalho realizado por Jahangir et al. (2021) trata a respeito de aprendizagem profunda no contexto do reconhecimento de emoções a partir da voz. Nele são analisadas diversas características extraídas do sinal vocal e como eles se comportam em diferentes abordagens de aprendizagem profunda. O autor menciona que o MFCC baseado em energia de *Teager* (TEMFCC) é indicado para detectar estresse na voz com base em suas observações apresentadas no estudo. Ele também afirma que os MFCC são as mais eficazes para o reconhecimento das emoções.

Em um estudo de Iqbal e colaboradores (2020) propôs uma técnica eficiente e precisa para o reconhecimento de emoções baseado na voz em quatro emoções básicas (tristeza, raiva, felicidade e neutralidade) usando uma RNA com regularização *bayesiana*. Os experimentos são conduzidos em um banco de dados em Alemão com 1470 amostras vocais contendo quatro emoções básicas com 500 amostras na emoção de raiva, 300 amostras na emoção felicidade, 350 amostras da neutralidade e 320 amostras de emoções tristes. Foram extraídas as medidas de frequência, *pitch*, *loudness* e formantes da voz de cada emoção para reconhecer as quatro emoções básicas da voz. A metodologia proposta alcançou 95% de acurácia no reconhecimento de emoções, que é a mais alta em comparação com outras técnicas de última geração no domínio relevante.

O artigo de Trinh Van et al. (2022) apresentou uma pesquisa sobre reconhecimento de quatro emoções (raiva, felicidade, tristeza e neutralidade) pela voz utilizando redes neurais profundas, como CRNN e *Gated Recurrent Unit* (GRU) no banco de vozes IEMOCAP. Os parâmetros de recursos usados para o reconhecimento incluem os MFCC e outros parâmetros relacionados ao espectro e à intensidade do sinal da voz. Os resultados mostraram que o modelo GRU apresentou a acurácia de reconhecimento de 97,47%.

Sharma (2021) apresentou um estudo comparativo de dois sistemas criados para reconhecimento das emoções calma, felicidade, tristeza, raiva, medo, surpresa, nojo e neutra. Foram criados dois modelos para reconhecimento de emoção por meio da voz na base de dados em inglês *Ryerson Audio-Visual Database of Emotional Speech and Song* (RAVDESS). Os MFCC foram utilizados como parâmetros dos arquivos de áudio. O primeiro modelo foi criado usando uma RNA *Multi-Layer Perceptron* (MLP) que forneceu uma acurácia de 57,29%. O segundo modelo criado foi de uma LSTM que apresentou acurácia de 92,88%.

Gangani et al. (2024) investigaram o reconhecimento das emoções (raiva, calma, nojo, medo, felicidade, tristeza, angústia, surpresa e a emissão neutra) a partir da voz

utilizando CRNN, explorando propriedades acústicas, como MFCC, *pitch* e intensidade. Três modelos foram propostos: redes neurais recorrentes com unidades LSTM, GRU e CRNN. Os modelos foram avaliados com dados dos bancos de vozes em inglês RAVDESS e TESS, e os resultados indicaram que o CRNN superou outros métodos em precisão e eficiência no reconhecimento de emoções. O modelo CRNN integrou a extração de características espectrais e análise temporal, alcançando uma precisão média ponderada de 84%. Os autores concluíram que as CRNN foram promissoras para capturar as variações emocionais complexas.

Com objetivo de implementar um sistema para reconhecimento de quatro estados emocionais a partir de sinais de áudio, Bhakre e Bang (2016) consideraram diferentes características como *pitch*, energia, *Zero Crossing Rate* (ZCR) e MFCC de 2000 enunciados de um banco vozes elaborado. A RNB foi utilizada para reconhecer o sinal de áudio em quatro emoções diferentes. Neste sistema obtiveram a acurácia para raiva, felicidade, tristeza e neutralidade de 81%, 78%, 76% e 77% respectivamente. Os resultados extraídos mostram que o sistema proposto foi capaz de reconhecer emoções em tempo real com um conjunto pequeno de dados.

Embora os métodos de aprendizado de máquina, amplamente utilizados em diversos estudos, apresentem benefícios significativos, como alta precisão e capacidade de identificar padrões complexos em grandes volumes de dados, o uso da lógica se sobressai por sua capacidade única de lidar com a incerteza e imprecisão inerentes ao comportamento humano e aos fenômenos naturais, como as emoções. A lógica *fuzzy* permite uma generalização mais ampla, uma vez que não exige classificações rígidas ou exatas, mas trabalha com conjuntos de possibilidades e graus de pertinência. Essa característica é vantajosa em áreas como o reconhecimento de estados emocionais, onde as fronteiras entre categorias são frequentemente tênues, garantindo uma modelagem mais flexível e adaptável à complexidade do mundo real.

3.5 Teoria dos Conjuntos *Fuzzy* e Lógica *Fuzzy*

A Teoria dos Conjuntos *Fuzzy* (ou *Fuzzy Set Theory*) é uma generalização da Teoria dos Conjuntos Clássica e visa modelar a incerteza sobre a classificação de elementos a um determinado conjunto.

Na Teoria dos Conjuntos Clássica, cada conjunto A de um universo X é definido pela função $X_A : X \rightarrow \{0,1\}$, que é dada por:

$$X_A(x) = \begin{cases} 1, sex \in A \\ 0, sex \notin A \end{cases} \quad (1)$$

onde $x \in X$.

Zadeh (1965) propôs a seguinte definição: seja X um universo, com um elemento genérico de X denotado por x. Um *conjunto fuzzy* A em X é caracterizado por uma função de pertinência $X_A(x)$ que associa a cada ponto em X um número real no intervalo [0,1], onde $X_A(x)$representa para x o seu grau de pertinência em A. Quando ∃ x X, tal que $X_A(x)$) = 1, então A é dito ser normalizado. Quando $X_A(x)$ = 1 diz-se que x é compatível com o conceito expresso por A em X e quando $X_A(x)$ = 0, diz que x é incompatível com A em X.

Essa definição de conjuntos *fuzzy* estabelece a base matemática para a representação de conceitos vagos ou subjetivos que possibilitam a construção de modelos capazes de lidar com informações nebulosas. Como nos conjuntos clássicos, é possível realizar operações sobre esses conjuntos (união, intersecção e complemento). É possível realizar a união e intersecção entre conjuntos *fuzzy* utilizando-se *t-normas* e *t-conormas* aplicados sobre as funções de pertinência dos conjuntos de entrada. A partir de uma *t-norma* (operador binário) pode-se definir a intersecção entre dois conjuntos *fuzzy*, enquanto a *t-conorma* pode-se definir a união entre dois conjuntos *fuzzy*. Algumas *t-normas* e *t-conormas* são apresentadas na Tabela 1:

Tabela 1 - Principais t-normas e t-conormas

t-norma	***t-conorma***
min (a,b)	max (a,b)
a.b	a + b - ab
max (a + b - 1, 0) $\{a, seb = 1b, se\ a = 10, outros\ casos$	min (a + b, 1) $\{a, seb = 0b, se\ a = 01, outros\ casos$

Fonte: MORAES, 1998

A lógica *fuzzy*, também denominada lógica difusa ou nebulosa, é um método amplamente aplicado na solução de problemas complexos, especialmente aqueles que envolvem descrições subjetivas, típicas das interações humanas. Esse método tem por objetivo a modelagem computacional da imprecisão do raciocínio humano, tornando-se útil em cenários onde não existem modelos matemáticos ou teóricos que expressem de forma precisa o comportamento do fenômeno em estudo (ZADEH, 1988). Na lógica *fuzzy*,

as *t-normas* e *t-conormas* modelam, respectivamente, os conceitos de conjunção (E) e disjunção (OU) que permitem lidar com graus de verdade que vão além dos valores binários.

Diferentemente dos métodos tradicionais de aprendizado de máquina, a lógica *fuzzy* oferece uma capacidade única de lidar com incertezas e subjetividades, aspectos frequentemente presentes nas emoções humanas. Enquanto os métodos de aprendizado de máquina, geralmente exigem grandes volumes de dados rotulados e operam com limites rígidos de classificação, a lógica *fuzzy* permite uma abordagem mais flexível (ROSS, 2010). Dessa forma, a lógica *fuzzy* se destaca por garantir maior generalização e por sua capacidade de integrar o conhecimento em situações em que as emoções apresentam características muito próximas, dificultando a distinção clara entre elas com base em medidas extraídas. Nesse sentido, a lógica fuzzy permite modelar essas sutilezas de forma mais intuitiva e eficiente, atribuindo graus de pertencimento a diferentes estados emocionais. Essa abordagem é indicada para o reconhecimento emocional, onde a precisão na identificação de nuances desempenha um papel importante para refletir a complexidade das respostas humanas.

Além disso, a lógica *fuzzy* tem sido considerada como uma ferramenta eficaz na representação de conhecimento e na modelagem de Sistemas de Apoio à Decisão. Diferentemente da lógica clássica, que opera com valores binários, classificando afirmações como verdadeiras ou falsas, a lógica *fuzzy* reconhece que muitas experiências humanas não podem ser reduzidas a essas categorias rígidas (ZADEH, 1965). Em contextos como saúde e doença, por exemplo, as respostas e comportamentos humanos frequentemente não se encaixam em definições absolutas de "sim" ou "não", "verdadeiro" ou "falso". Assim, a lógica *fuzzy* possibilita a consideração de áreas de transição ou granularidade, fornecendo uma modelagem mais precisa para fenômenos que envolvem incerteza e variabilidade (MARQUES et al., 2005).

Enquanto sistemas tradicionais categorizam afirmações de forma binária, a lógica *fuzzy* permite modelar cenários com diferentes níveis de associação, como "Se o copo está quase vazio, então posso matar parte da minha sede". Essa abordagem considera que os limites dos conjuntos não são rígidos, e a manipulação dos elementos em situações de incerteza utilizam o conceito de variáveis linguísticas e valores representados por conjuntos *fuzzy*. Nesse contexto, os valores das variáveis não são precisos ou fixos, mas descritos por termos linguísticos. Essa modelagem permite a caracterização aproximada

de fenômenos complexos e é regida por funções de compatibilidade que avaliam o grau de pertencimento dos elementos a esses conjuntos (ZADEH, 1988).

As variáveis linguísticas surgem como uma extensão prática desse formalismo para traduzir informações qualitativas em representações quantitativas por meio de termos associados a conjuntos *fuzzy*. Segundo Zadeh (1988) cada variável linguística é caracterizada por um quíntuplo (V, T(V), X, G, M), onde V é o nome da variável, T(V) é o conjunto de termos, X é o universo de discurso, G é a regra sintática que gera os termos em T(V), e M é a regra semântica que associa um significado a cada termo, geralmente representado por um subconjunto *fuzzy* de X. Dessa forma, os valores dessas variáveis são palavras com significado de uma linguagem natural ou sintética. Por exemplo, a variável linguística "Intensidade" pode ser descrita em termos linguísticos como: "intensidadeMuitoMuitoPequena", "intensidadeMuitoPequena", "intensidadePequena", "intensidadeMedia", "intensidadeAlta", "intensidadeMuitoAlta" e "intensidadeMuitoMuitoAlta". Cada termo está relacionado a uma função de pertinência que o descreve dentro do universo de discurso.

Uma das maneiras de formalizar esse conhecimento é por meio de regras do tipo condição-ação. Em lógica *fuzzy*, uma regra representa uma relação estabelecida entre variáveis de entrada e saída por meio de operadores lógicos baseados em valores linguísticos e conjuntos *fuzzy*, estruturada no formato "SE... ENTÃO...". A parte antecedente da regra, correspondente à condição "SE", define critérios com base em conjuntos *fuzzy* associados às variáveis de entrada, enquanto a parte consequente, representada pelo "ENTÃO", especifica a ação ou o resultado esperado. Essas regras utilizam os operadores como E, OU e NÃO, que permitem o processamento de graus de pertinência em vez de valores binários, característicos da lógica clássica (ZADEH, 1978). Assim, cada regra *fuzzy* constitui uma unidade de conhecimento que captura informações específicas do domínio em questão, permitindo a modelagem e resolução de problemas de forma mais flexível e adaptativa (PASSINO; YURKOVICH, 1998).

Os conjuntos *fuzzy* são geralmente construídos com base no julgamento de especialistas. Essa abordagem envolve a participação de um ou mais especialistas na área de interesse, cujas opiniões podem ser coletadas individualmente ou em grupo, e agregadas de forma adequada. As funções de pertinência podem ser determinadas por métodos diretos ou indiretos, dependendo da complexidade do conceito que o termo linguístico deve representar (MASSAD et al., 2008). Além da abordagem baseada em especialistas, é possível utilizar bases de conhecimento extraídas de dados para gerar

tanto o conjunto de regras *fuzzy* quanto às funções de pertinência, oferecendo uma alternativa mais automatizada ao processo.

Os sistemas baseados em regras *fuzzy* utilizam conectivos como AND (E) e OR (OU). Esses conectivos podem ser aplicados tanto nas premissas (condições) quanto nas conclusões de uma regra. A premissa representa o argumento da regra, enquanto a conclusão refere-se ao resultado, que neste estudo corresponde aos graus de prioridade. Existem dois sistemas de inferência *fuzzy* amplamente conhecidos: o de Mamdani, que utiliza uma estrutura de operações min-max e regras do tipo "Se x é A e y é B, então z é C", onde A, B e C são conjuntos *fuzzy* (MASSAD et al., 2008); e o de Takagi-Sugeno, o qual é caracterizado por um processo de tomada de decisão simplificado, baseado na lógica *fuzzy*, onde apenas o antecedente das regras é formado por variáveis *fuzzy*. O consequente de cada regra é expresso por uma função linear dos valores observados das variáveis que descrevem o estado do sistema, também conhecidas como variáveis de entrada (TAKAGI, SUGENO, 1983). No presente trabalho, foi utilizado o sistema de inferência de Mamdani, juntamente com o processo de defuzzificação, que converte um conjunto *fuzzy* em um valor numérico real (SAADE; DIAB, 2003).

3.5.1 Lógica *fuzzy* em sistemas de detecção automática

A inteligência artificial com utilização da lógica *fuzzy* fornece uma alternativa à abordagem de parâmetro único ou multidimensional para análise de reconhecimento de padrões. A lógica *fuzzy* tem sido amplamente explorada em diversas aplicações, incluindo sistemas de controle e tomada de decisão, desde sua introdução na década de 1960 (ZADEH, 1965). Embora haja uma vasta pesquisa sobre o uso de sistemas *fuzzy* para modelagem de incertezas e variáveis linguísticas, poucos estudos aplicaram essa tecnologia à análise da detecção automática na variação das emoções (XIONG, et al., 2024; TON-THAT; CAO, 2019; GRIMM et al., 2007). Assim, o levantamento de características acústicas e de fala poderia servir como uma ferramenta eficiente para identificar as variações emocionais em sistemas automatizados de reconhecimento a partir da voz humana, utilizando a lógica *fuzzy* para lidar com as ambiguidades e subjetividades inerentes às expressões emocionais (ESAU, 2007, HEGDE et al., 2019).

A lógica *fuzzy* tem se mostrado uma abordagem promissora para a detecção e reconhecimento automáticos de emoções, principalmente devido à sua capacidade de lidar com incertezas e variabilidades que são inerentes às expressões emocionais

humanas (XU, 2011). Ao contrário de outras técnicas que podem exigir uma categorização rígida das emoções, a lógica *fuzzy* permite que diferentes estados emocionais sejam representados de forma mais precisa, com graus de pertinência que refletem a complexidade das emoções humanas (RAVI; ZIMMERMANN, 2000).

A detecção de emoções por meio da lógica fuzzy aproveita as características únicas dessa abordagem para interpretar sinais vocais, faciais ou fisiológicos que, frequentemente, não são facilmente classificados em categorias binárias (MORAES, 2002). Por exemplo, um sorriso pode ser identificado com graus variados de alegria, ou o tom da voz pode sugerir simultaneamente a surpresa e medo em diferentes intensidades. A teoria de conjuntos *fuzzy* nessas situações é aplicada pela sua capacidade de lidar com situações imprecisas e incertas, resultando em sistemas de reconhecimento emocional que são mais adaptáveis e realistas (ORTEGA, 2001).

Além disso, a lógica *fuzzy* facilita a criação de modelos que incorporam o conhecimento e as regras heurísticas desenvolvidas por seres humanos especialistas, tornando-a uma ferramenta valiosa em contextos onde as emoções precisam ser interpretadas de maneira subjetiva e contextualizada (LILIANA, et al., 2019), tornando o reconhecimento mais sensível em aplicações como assistentes virtuais, sistemas de atendimento ao cliente, e interfaces homem-máquina, onde a interpretação precisa das emoções pode melhorar significativamente a interação e a experiência do usuário.

Pesquisas recentes têm explorado a aplicação da lógica *fuzzy* na análise de sinais de voz e expressões faciais para identificar estados emocionais com maior precisão (LILIANA, et al., 2019; TON-THAT; CAO, 2019; GRIMM et al., 2007). Essas abordagens têm mostrado resultados promissores, especialmente em situações onde outras técnicas, como redes neurais e algoritmos de aprendizado profundo, podem falhar em capturar as sutilezas emocionais ou exigir pequenos volumes de dados para treinamento.

Liliana e colaboradores (2019) desenvolveram um sistema baseado em lógica *fuzzy* para reconhecimento de emoções por meio de expressões faciais, buscando uma abordagem mais natural e alinhada às respostas emocionais humanas. O sistema permitiu lidar com a ambiguidade inerente às emoções humanas, focando em seis emoções principais: felicidade, tristeza, raiva, surpresa, nojo e medo. O estudo demonstrou que o uso da lógica *fuzzy* categorizou de forma mais precisa as emoções, melhorando as taxas de reconhecimento com melhoria de 12% em comparação com modelos tradicionais KNN, RNA e métodos baseados em lógica binária. Os principais resultados afirmaram que o

sistema *fuzzy* teve maior precisão ao reconhecer estados emocionais mistos, como quando as expressões faciais indicavam surpresa e medo simultaneamente.

O estudo de Ton-That e Cao (2019) desenvolveu um sistema de inferência *fuzzy* baseado em memória associativa *fuzzy* (FAM-FIS) para reconhecimento de emoções a partir da voz, com foco na melhoria da precisão do reconhecimento. O sistema utilizou as características MFCC e foi testado em dois bancos de dados: EMO-DB (alemão) e SAVEE (inglês). Os resultados mostraram que o sistema *fuzzy* superou classificadores tradicionais, como Bayes e SVM. Especificamente, o FAM-FIS alcançou uma precisão de 74,31% no EMO-DB e 97,29% no SAVEE, superando o desempenho do SVM. O estudo destacou a capacidade do sistema *fuzzy* de lidar com as complexidades emocionais, aumentando a precisão à medida que mais dimensões das características MFCC foram incluídas. Os autores concluíram que a abordagem *fuzzy* oferece vantagens para o reconhecimento de emoções na voz e outros problemas de reconhecimento de padrões.

Grimm e colaboradores (2007) desenvolveram um sistema para estimar emoções a partir da voz usando lógica *fuzzy* em um espaço multidimensional de emoções. O estudo focou nos eixos de dimensões na abordagem contínua das emoções, de valência (positivo/negativo), ativação (calmo/excitado) e potência (fraco/forte). Utilizando características acústicas como *pitch*, energia e taxa de elocução, o sistema foi testado em dois conjuntos de dados: emoções simuladas e emoções espontâneas de um programa de televisão. Os resultados verificaram que a lógica *fuzzy* apresentou alta correlação com as avaliações humanas. A precisão no reconhecimento das emoções alcançou 83,5% ao mapear os três eixos emocionais para categorias. O sistema foi considerado eficaz em cenários dependentes de locutor, ressaltando a importância de modelar a variação emocional individual. O estudo concluiu que a lógica *fuzzy* oferece um método eficiente para lidar com variações emocionais contínuas.

Hidalgo e colaboradores (2024) desenvolveram um sistema baseado na voz para o reconhecimento de estados emocionais e sua associação a transtornos de humor. O estudo utilizou bases de dados em espanhol e alemão para treinar um modelo de reconhecimento baseado em lógica *fuzzy* integrada a SVM, com características acústicas como *pitch*, intensidade e coeficientes MFCC das 6 emoções básicas. Os resultados demonstraram uma alta precisão na classificação de emoções como felicidade, tristeza, ansiedade e raiva, permitindo a integração dessas informações no projeto Bip4Cast, que visa prever estados de humor, incluindo mania, depressão e eutimias. O uso do sinal vocal foi considerado eficaz na identificação de crises emocionais, com potencial para

complementar outras fontes de dados clínicos, como sensores de movimento. Além disso, os resultados indicaram que a lógica *fuzzy* ofereceu maior flexibilidade e precisão em comparação a métodos tradicionais, permitindo lidar com sobreposições entre estados emocionais e categorizando melhor emoções difíceis de distinguir.

A lógica *fuzzy* oferece uma alternativa eficaz para a detecção e reconhecimento automáticos de emoções, e permite que sistemas de inteligência artificial interpretem as emoções humanas de maneira mais natural e ajustada às incertezas e variações que essas emoções podem apresentar. A contínua pesquisa e desenvolvimento nessa área prometem aprimorar ainda mais a capacidade dos sistemas automatizados de compreender e responder às emoções humanas, abrindo novas possibilidades para a aplicação da lógica *fuzzy* em contextos emocionais e interativos.

3.6 Parâmetros de Acurácia dos Modelos

Os modelos de reconhecimento foram analisados levando em consideração a Matriz de Confusão, a Acurácia Geral (AG) e o coeficiente Kappa, para avaliar a precisão dos resultados obtidos.

A matriz de confusão é uma matriz quadrada de números que expressam a quantidade de unidades amostrais, associada a uma dada categoria durante o processo de reconhecimento efetuado, e à categoria real a que pertencem essas unidades (CONGALTON, 1991). Quando aplicada no contexto da lógica *fuzzy*, essa matriz não apenas detalha os resultados esperados e os resultados obtidos, mas também considera os graus de pertinência das amostras às categorias analisadas. Assim, permite uma análise mais flexível e precisa da performance do modelo e proporciona uma visão mais abrangente das classificações realizadas.

A AG (CONGALTON, 1999) é amplamente utilizada na literatura, mede a concordância geral entre as previsões realizadas por um sistema e os dados de referência, sendo uma métrica fundamental para avaliar a performance de modelos de reconhecimento, incluindo aqueles voltados à identificação de emoções. Esse índice é definido como:

$$AG = \frac{\sum_{i=1}^{M} m_{ii}}{N} \qquad (2)$$

onde $\sum_{i=1}^{M}$ é a soma da diagonal principal da matriz de confusão, M é o número total de

classes e N é a quantidade total de decisões possíveis presentes na matriz de classificação.

Apesar de sua simplicidade e ampla aceitação, o AG é considerado otimista por não ponderar os erros fora da diagonal principal, o que pode mascarar problemas em classes minoritárias ou de difícil reconhecimento (MORAES et al., 2020). Essa abordagem permite incorporar a incerteza inerente às emoções humanas, proporcionando uma análise mais robusta e adequada para aplicações que requerem maior sensibilidade às variações emocionais. Assim, o uso da AG, aliado a métricas complementares, contribui para uma avaliação abrangente e detalhada da performance do sistema de reconhecimento.

O índice Kappa (COHEN, 1960) será utilizado para avaliar o desempenho do Modelo. O uso desse coeficiente é satisfatório na avaliação de um sistema de reconhecimento, pelo fato de apresentar uma medida da concordância entre as taxas de acerto e erro alcançadas, usando a matriz de confusão no seu cálculo, inclusive os elementos de fora da diagonal principal, os quais representam as discordâncias no modelo (MORAES, 2020). Seu cálculo procede da seguinte forma.

$$\text{K}= \frac{P_{0-P_c}}{1-P_c}\text{ ,}\quad \text{sendo } P_0 \text{ e } P_c\text{:}\quad P_0=\frac{\sum_{i=1}^{M} m_{ii}}{N} \text{ e } P_c= \frac{\sum_{j=1}^{M} m_{i+m+i}}{N^2}\text{, } (3)$$

em que m_{ii} é o valor total da diagonal principal, m_{i+} é o valor total da linha i, m_{+i} é o total da coluna i, M é o número total de classes, e N é o total de decisões possíveis da matriz de confusão $m_{ii}m_{i+m_{+i}}$.

Segundo Landis e Koch (1977), quanto mais próximo de 100%, maior o grau de concordância do modelo de decisão. A interpretação da magnitude do estimador do Kappa é convencionada como: <0,0% (Ruim), 0,00%–20,00% (Leve), 21,00%–40,00% (Razoável), 41,00%–60,00% (Moderado), 61,00%–80,00% (Substancial) e 81,00%–100,00% (Quase Perfeito).

As medidas de sensibilidade e especificidade são amplamente utilizadas em medicina e epidemiologia e compõem um conjunto de testes diagnósticos (VAN STRALEN, 2009). São medidas estatísticas do desempenho de uma classificação binária, traduzidas por uma tabela de contingência 2 × 2. Também é possível obter essas medidas a partir de uma matriz de confusão n × n.

Quatro medidas básicas são utilizadas nos testes diagnósticos e delas é possível obter todas as outras: Verdadeiro Positivo (VP, que significa identificação correta), Falso

Positivo (FP ou identificação incorreta), Verdadeiro Negativo (VN, que significa rejeição correta) e Falso Negativo (FN ou rejeição incorreta).

A Sensibilidade, ou taxa de Verdadeiro Positivo (VP), é a probabilidade de identificação correta. A TPR é dada por:

$$VP = \frac{VP}{VP+FN} \qquad (4)$$

A Especificidade, ou taxa de Verdadeiro Negativo (VN), é a probabilidade de rejeição correta. A TNR é dada por:

$$VN = \frac{VN}{VN+FP} \qquad (5)$$

Outra medida interessante é a Acurácia (AC), que é a probabilidade de acertos de um classificador, ou seja, a probabilidade de um classificador identificar corretamente ou rejeitar corretamente. A AC é expressa por:

$$AC = \frac{VP+VN}{VP+VN+FP+FN} \quad (6)$$

Essas medidas podem fornecer uma análise detalhada do modelo a partir da matriz de confusão, em oposição aos coeficientes de concordância. Vale salientar, que a análise de testes diagnósticos por meio das métricas de sensibilidade e especificidade é tradicionalmente aplicada a dados binários, nos quais a condição avaliada se encontra claramente categorizada entre presença ou ausência de determinada característica ou desfecho. Essa abordagem, embora amplamente aceita e eficaz em contextos binários, pode apresentar limitações quando aplicada a dados não binários, uma vez que combina diferentes graus de erros e acertos em uma única métrica. Essa simplificação pode comprometer a precisão da análise, especialmente em cenários onde há múltiplas categorias ou níveis de gravidade. No entanto, apesar dessas limitações, a sensibilidade e especificidade podem, ainda assim, ser empregadas para avaliar dados não binários, desde que se reconheça que essa aplicação exige cautela e uma interpretação mais crítica dos resultados. Essa utilização, embora menos precisa, pode fornecer informações relevantes para avaliar o desempenho do modelo.

Um modelo de reconhecimento dos estados emocionais que considera as incertezas e a natureza imprecisa dos dados de entrada, características comuns quando se trata de emoções humanas e que utilize um banco de vozes construído com as medidas perceptuais, acústicas e prosódicas dos sinais de áudio em cada estado emocional permitirá o desenvolvimento de uma ferramenta automatizada no reconhecimento das

emoções e será capaz de permitir maior taxa de acurácia no reconhecimento e diferenciação das emoções a partir da voz humana do que os métodos tradicionais, o que possibilita ao examinador um avanço potencialmente eficiente na identificação do estado emocional, permitindo agilizar, simplificar, facilitar e auxiliar a análise da identificação de cada estado emocional além de gerar impacto de inovação tecnológica em diversos tipos de mercado com ferramentas robustas, não invasivas e de baixo custo.

4 Métodos

4.1 Considerações Éticas

Esta pesquisa faz parte do projeto “Banco de vozes em diferentes estados emocionais: construção, reconhecimento de padrões e validação transcultural”, que foi avaliado e aprovado pelo Comitê de Ética em Pesquisa do Centro de Ciências da Saúde da Universidade Federal da Paraíba (CEP/CCS/UFPB), por meio do parecer nº 3.304.419/2019 (ANEXO 1). Trata-se de uma pesquisa de natureza tecnológica, descritiva, observacional, transversal, que utilizou dados secundários provenientes do EmoVox-BR e que contou com a criação de um modelo de reconhecimento de padrões de voz nas variadas emoções simuladas.

Todos os voluntários foram solicitados a ler e caso concordassem com o conteúdo, assinar um Termo de Consentimento Livre e Esclarecido (TCLE), que expõe os objetivos do estudo, riscos, benefícios e procedimentos de coleta de dados, bem como a garantia de confidencialidade das informações obtidas.

4.2 Área e População do estudo

O estudo foi realizado no Laboratório Integrado de Estudos da Voz (LIEV) do Departamento de Fonoaudiologia e no Laboratório de Estatística Aplicada ao Processamento de Imagens e Geoprocessamento (LEAPIG) do Departamento de Estatística, ambos da Universidade Federal da Paraíba.

Todos os participantes da pesquisa seguiram os seguintes critérios de elegibilidade: ausência de comorbidades que comprometesse a cognição, audição e comunicação que poderia limitar a realização das tarefas solicitadas; não fazer uso de substâncias psicotrópicas; responder aos questionários da pesquisa; ter acesso a internet, microfone

externo, e smartphone; ter realizado as gravações nas 6 variações das emoções e na emissão neutra.

O conjunto de dados para análise foi composto por 182 sinais sonoros, produzidos por 26 atores profissionais e estudantes de Artes Cênicas brasileiros, de ambos os gêneros, com média de idade sendo 26 (±8,86) anos, residentes nos estados da Paraíba, São Paulo, Rio grande do Sul, Ceará, Roraima, Mato Grosso e Distrito Federal. Esses sinais pertencem ao Banco de Vozes Brasileiro nas Variações das Emoções - EmoVox-BR, desenvolvido pelo LIEV (LIMA, 2022).

O EmoVox-BR foi elaborado e validado de acordo com alguns direcionamentos. Os participantes receberam um tutorial com roteiro e procedimentos de gravação e em seguida realizaram a coleta da voz remota. Foram coletadas três tarefas de fala distintas: 1) a emissão da vogal /ɛ/ sustentada; 2) fala automática a partir da contagem de 1 a 10; e 3) fala semi-espontânea por meio da frase “Olha lá o avião azul.”, que compõem o *Consensus Auditory Perceptual Evaluation of Voice* - CAPE-V (BEHLAU et al., 2020). As tarefas foram executadas por cada voluntário nas seis emoções básicas (alegria, tristeza, medo, raiva, surpresa e nojo) e emissão neutra. Todos os voluntários gravaram com as seguintes variações: com e sem fone, por meio de *smartphone* e computador, a partir da plataforma Zoom Meeting e de forma *line in.* Essas variações metodológicas geraram 63 amostras vocais por voluntário (ator profissional ou em formação), o que totalizou 1.638 sinais de áudios (três tarefas de fala, vezes a variação das seis emoções e a emissão neutra, vezes 3 variações de formatação da coleta, vezes 26 voluntários).

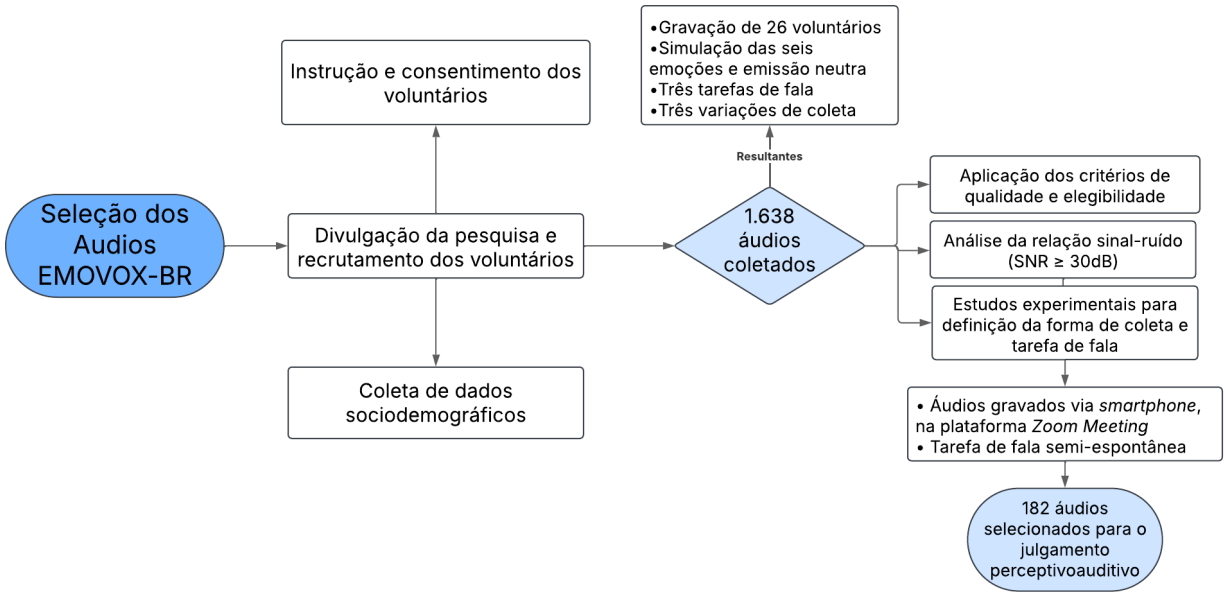
Alguns estudos (ASHA, 2002; MONTEIRO, 2021) buscaram analisar a qualidade de diferentes tarefas de fala nas variações emocionais, como também investigar quais as diferenças entre a qualidade do sinal coletado de forma remota com entrada *line in* e pela plataforma *zoom meeting*, também foram elencados trabalhos, que objetivavam analisar se as vozes coletadas via smartphone possuíam relação sinal-ruído diferente dentre as vozes coletas com o uso de microfone e sem o microfone. Os resultados desses estudos verificaram que a fala semi-espontânea e contagem são as tarefas de fala que apresentam melhor qualidade na coleta remota, a opção de entrada direta no aparelho para coleta da voz de forma *line in* se sobressai quando comparada aos serviços de coleta da plataforma *zoom meeting*, e o *smartphone* é uma opção segura e eficaz para a coleta remota da voz.

O estudo de Monteiro e colaboradores (2021) investigou a eficácia do uso de *smartphones* como estratégia de coleta remota de voz e avaliou a relação sinal-ruído de emissões vocais na variação das seis emoções básicas e emissão neutra, coletadas via

smartphones com e sem microfone externo. Os resultados indicaram que a gravação por *smartphone* apresentou qualidade de áudio satisfatória para todas as emoções, independentemente do uso do microfone externo, mas em local com intensidade controlada. A pesquisa concluiu que o *smartphone*, associado a protocolos adequados de gravação, é uma ferramenta confiável para a coleta remota de voz em contextos clínicos e científicos e garantiu a validade de análises acústicas e perceptivo-auditivas.

Após essas motivações exposta e testes, os pesquisadores optaram por selecionar a tarefa de fala semi- espontânea, coletada via smartphone. Portanto dos 1.638 sinais de áudios coletados, foram selecionados 182 sinais de áudio para a etapa do julgamento perceptivo-auditivo com objetivo de aumentar a robustez e a representatividade da análise (LIMA et al, 2022). Esses mesmos sinais foram utilizados neste estudo como forma de ter acesso a uma amostra mais ampla que permite capturar uma gama mais variada de características e nuances emocionais presentes no banco de dados, o que pode resultar em um modelo mais robusto e com melhor generalização para diferentes contextos de aplicação. A seguir, observa-se o fluxograma detalhando do processo para a obtenção dos áudios finais no EmoVox-BR, desde a seleção inicial até a finalização e validação dos dados (Figura 1).

Figura 1 - Processo de seleção dos áudios para construção do EmoVox-BR



Fonte: Elaborado pelo autor, 2024.

As vozes foram avaliadas por juízes fonoaudiólogos com a tarefa de fala “olha lá o avião azul”, executada por cada voluntário nas seis emoções básicas (alegria, tristeza, medo, raiva, surpresa e nojo) e emissão neutra, e gravadas com uso de smartphone, sem microfone e de forma *line in*. Os sinais de áudio receberam a indicação da emoção correspondente, da potência/intensidade da emoção (o quanto a emoção transmitida pelo áudio), valência (emoção positiva ou negativa), e de quais parâmetros de voz e fala seriam decisivos para a identificar a emoção daquele áudio (*pitch*, *loudness*, articulação, velocidade de fala, incoordenação pneumofonoarticulatória, fluência, qualidade vocal). Os juízes especialistas foram orientados a ouvir 200 áudios. Em 182 áudios mais 18 áudios (10% da amostra) foram apresentados de forma repetida para garantir a confiabilidade intraavaliador na análise perceptivo-auditiva para identificação das emoções, de forma randomizada. Assim, para a avaliação com os juízes foram utilizadas 200 vozes com o objetivo de entender quais vozes eram as mais confiáveis, bem como também a concordância inter juízes. Concluiu-se que os fonoaudiólogos conseguiram um alto percentual de acerto das seis emoções básicas e o estado neutro, com valores de referência superiores a 70% na identificação das emoções, valência e potência, além de indicarem os parâmetros de voz que auxiliam reconhecer as variadas emoções (LIMA et al, 2022).

4.3 Materiais

O presente estudo adota um modelo experimental quantitativo com o objetivo de desenvolver e validar um sistema preditivo baseado na lógica *fuzzy* para reconhecimento de diferentes estados emocionais a partir da análise da voz. Os áudios do EMOVOX-BR foram processados com técnicas de modelamento acústico utilizadas para a extração de vetores de características dos MFCC, e outras medidas acústicas como medidas de f_o, o *jitter*, o *shimmer*, as medidas de ruído glótico GNE, HNR, CPPS e medidas acústico-prosódicas de f_o, duração e intensidade.

Vale destacar que neste estudo, a f_o foi analisada a partir de duas perspectivas distintas: a acústica tradicional e a prosódica. Como medida acústica tradicional, foram considerados os valores de f_o média, mínima, máxima e o dp, que refletem a estabilidade e as variações globais da frequência ao longo do tempo. Já na análise prosódica, os mesmos valores de f_o média, mínima e máxima foram utilizados, porém com a adição do range de f_o, que representa a amplitude entre os valores máximo e mínimo de frequência

fundamental. Essa inclusão do range na análise prosódica destaca o foco dessa abordagem em captar a dinâmica melódica da voz, essencial para compreender aspectos expressivos e entoacionais em contextos comunicativos.

Foram consideradas também como variáveis independentes as dimensões de valência, potência e ativação de cada emoção. Essas, são características fundamentais e inerentes de cada emoção, refletem dimensões essenciais que descrevem como as emoções são experienciadas e diferenciadas. A valência refere-se ao espectro prazer-desprazer e indica se uma emoção é percebida como positiva. A potência, por sua vez, está relacionada ao senso de controle ou força associado à emoção e descreve se ela é percebida como dominante. Já a ativação diz respeito ao nível de energia ou excitação, para diferenciar emoções de alta ativação de baixa ativação (VIEIRA, 2018). Essas dimensões não são construídas a partir de medições, mas representam atributos estruturais de cada emoção, conforme descrito na literatura (GRIMM, 2007; SPEED; BRYSBAERT, 2023, BESTELMEYER et al., 2017).

Todas as variáveis foram coletadas a partir dos dados do EMOVOX- BR, para desenvolver um modelo de reconhecimento das emoções básicas a partir da voz. Cada amostra vocal foi submetida a uma análise detalhada com o objetivo de extrair características acústicas e prosódicas específicas associadas a cada emoção.

4.3.1 Extração das medidas acústico-prosódicas

A extração dos aspectos acústico-prosódicos da voz baseou-se principalmente nos parâmetros de duração, f_0 e intensidade, uma vez que são considerados elementos robustos na identificação de falantes (BARBOSA; CONSTANTINI, 2020). A duração de uma emoção pode variar de momentos breves a períodos prolongados, que depende da intensidade do estímulo emocional, da capacidade de regulação emocional do indivíduo e do contexto em que a emoção ocorre (LARROUY-MAESTRI, 2023). A duração é fundamental para identificar variações no ritmo da fala, na estruturação dos enunciados e na organização temporal das frases (CONSTANTINI; BARBOSA, 2015).

Para análise da duração, todas as amostras de fala foram segmentadas manualmente em unidades Vogal-Vogal (unidade VV), que são unidades do tamanho da sílaba que compreendem o segmento que vai do início de uma vogal até o início da próxima vogal, que inclui as consoantes entre elas (CONSTANTINI; BARBOSA, 2015).

A frase utilizada no presente "Olha lá o avião azul", foi então segmentada foneticamente em unidades VV, resultando nas seguintes divisões: [Al], [auav], [iaNU] e [az]. Essa segmentação é amplamente utilizada na análise acústico-prosódica da fala por se tratar de uma unidade que representa, de forma mais estável, as variações temporais e rítmicas do enunciado. Para evitar interferências que pudessem comprometer a análise dos parâmetros acústicos, os segmentos iniciais e finais da tarefa de fala foram excluídos. Essa exclusão se justifica pelo fato de que esses trechos costumam apresentar maior instabilidade na produção vocal, especialmente devido ao ajuste inicial das pregas vocais e à redução natural da intensidade sonora ao final da frase. Além disso, é comum que essas regiões contenham pausas ou hesitações que não refletem o padrão típico da produção vocal durante a emissão do enunciado. A exclusão desses trechos, portanto, visa garantir que a análise se concentre no núcleo do enunciado, onde os parâmetros de duração se manifestam de forma mais estável e representativa (BARBOSA; CONSTANTINI, 2020).

Para extrair essa medida, com utilização do cálculo da duração normalizada das unidades VV, foi empregado o script SG Detector (BARBOSA, 2006). O script apresenta uma tabela de referência com médias e desvios-padrão dos segmentos fônicos do PB para calcular o valor da duração, z-score e o z-score suavizado das unidades VV ao longo do enunciado, que gerou uma segmentação dos grupos frasais. A segmentação é realizada por meio do cálculo do desvio-padrão das médias de duração das unidades VV, que são normalizadas pelo cálculo do z-score.

O valor do z-score indica o número de desvios-padrão em relação à média, no caso as variações do PB, ou seja, é uma proporção do número de desvios-padrão acima ou abaixo do PB, representando uma pontuação bruta. O z-score é uma pontuação padrão que pode ser posicionada em uma curva de dispersão comum, variando de –3 a +3 desvios (MATTE, 2006).

Os valores de z-score suavizado permitem atenuar variações locais de duração provenientes de quedas de duração em unidades VV pós-tônicas e/ou de fones com durações muito diferentes das relações de duração dos fones do PB (BARBOSA, 2006). Esse valor representa a suavização de cinco pontos aplicada à sequência de dados de z-score, que permite observar com mais precisão as proeminências de duração.

A f_o corresponde ao número de vibrações das pregas vocais por segundo e está diretamente relacionada à percepção de *pitch* (LOPES et al., 2018). Esse parâmetro permite uma análise detalhada da entonação e da variação tonal ao longo dos enunciados,

fornecendo informações sobre o controle e a qualidade vocal (ARVANITI, 2020). Para a análise da f_o, foram medidos os valores médios, máximos, mínimos e o intervalo de variação (f_o range) em cada enunciado.

A intensidade está associada à força ou ao grau de ativação emocional experimentado pelo falante, que se relaciona à energia vocal empregada durante a fala (LOPES et al., 2018). Sua medição ajuda a entender a energia da emissão e a modulação da força ao longo do discurso, contribuindo para estudos de prosódia e expressividade vocal (ARVANITI, 2020). Na análise da intensidade, foram extraídos os valores médios, mínimos e máximos de cada sinal, e os dados foram posteriormente comparados entre as emoções. As medidas de f_o e intensidade foram extraídas com o auxílio do plug-in VoxMore (ABREU et al., 2023).

Utilizou-se o programa PRAAT, versão 5.4.04, para a extração das medidas acústico-prosódicas. O software gera um relatório com informações e imagens referentes aos valores de medidas acústicas relacionadas a f_o, medidas de período, perturbação e ruído. As instruções sobre instalação do VoxMore podem ser encontradas em: https://github.com/abreusamuel/VoxMore.

4.3.2 Extração das medidas acústicas

A extração das medidas acústicas envolveu a análise detalhada de diversos parâmetros essenciais para a avaliação da qualidade vocal e da expressividade emocional dos sinais de áudio. Entre os parâmetros extraídos, a f_o foi medida em termos de seus valores mínimo, máximo, médio e dp, que permite a compreensão das variações tonais ao longo dos enunciados (BROCKMANN-BAUSER DRINNAN 2011). A f_o, é importante para avaliar o *pitch* da voz e entender como diferentes emoções podem influenciar a frequência vocal (LOPES et al., 2018). As variações da f_o entre os diferentes estados emocionais fornecem uma base sólida para a discriminação entre emoções.

O CPPS e seu dp também foram extraídos. O CPPS é uma medida que reflete a suavidade e regularidade das vibrações das pregas vocais, sendo um importante indicador da qualidade vocal e da intensidade emocional. A inclusão do dp do CPPS permite capturar as flutuações na regularidade vocal ao longo do tempo, que pode oferecer *insights* sobre a instabilidade vocal em diferentes estados emocionais (LOPES et al., 2019).

Outro parâmetro relevante foi o *jitter*, que mede a variação na frequência da voz entre ciclos sucessivos de vibração das pregas vocais. O *jitter* é um indicador de perturbações no controle vocal, podendo revelar emoções que impactam diretamente na estabilidade vocal, como medo ou ansiedade (LOPES et al., 2018). O *shimmer*, por sua vez, mede as variações na amplitude das vibrações vocais, sendo fundamental para avaliar a instabilidade na intensidade vocal (LOPES et al., 2018). Emoções que alteram o controle da força vocal, podem ser identificadas por meio do *shimmer*.

As medidas de ruído glótico, como o GNE e o HNR, também foram extraídas. O GNE reflete a proporção de ruído glotal em relação ao som harmônico produzido pelas pregas vocais, sendo uma medida importante para avaliar a quantidade de ruído na voz (BROCKMANN-BAUSER; DRINNAN 2011; GODINO-LLORENTE et al., 2010). Já o HNR quantifica a relação entre os componentes harmônicos e o ruído presente no sinal vocal, sendo utilizado para analisar a qualidade da voz em termos de clareza e pureza (BROCKMANN-BAUSER; DRINNAN 2011; LOPES et al., 2018). Essas medidas são essenciais para distinguir emoções que podem aumentar o ruído vocal, como a tristeza ou o cansaço, em comparação com emoções que resultam em uma voz mais clara e limpa, como a alegria ou a surpresa.

Utilizou-se o programa PRAAT, versão 5.4.04, para a extração dos dados acústicos com o auxílio do plug-in VoxMore (ABREU et al., 2023).

4.3.3 Extração dos Coeficientes Mel-Cepstrais

Os MFCC são parâmetros constantemente utilizados para a construção de modelos de reconhecimento a partir da voz (MA; FOKOUÉ, 2014; PARTILA et al. 2015; FANG et al., 2019; TANDEL et al., 2020), pois são baseados na resposta em frequência do ouvido humano. Sua ideia principal é transformar o sinal do domínio do tempo para o domínio da frequência e mapear o sinal transformado em hertz para a escala Mel devido ao fato de que 1 kHz é o limite da capacidade auditiva humana.

A extração das características MFCC utilizou a escala Mel, desenvolvida por Stevens e Volkmann na década de 1940, com o objetivo de representar como o ouvido humano percebe as frequências presentes no espectro sonoro (MA; FOKOUÉ, 2014). Concluíram que esta relação é linear de 0 a 1000 Hz, e que para frequências superiores a 1000 Hz, a relação pode ser descrita de forma logarítmica.

O MFCC pode ser calculado usando a fórmula apresentada na Equação (7) para converter a frequência em Hertz para a correspondente escala Mel (Kumar et al. 2014)

$$F_{mel} = 2595.log_{10}\left[1 + \frac{F_{linear}(Hz)}{700}\right] \quad (7)$$

onde F_{mel}é a frequência resultante na escala mel medida em mels e $F_{linear}(Hz)$ é a frequência medida em Hertz.

Para realizar o cálculo dos coeficientes mel-cepstrais primeiramente é obtido o módulo ao quadrado da transformada de Fourier do sinal,$x(n)$ para cada segmento do sinal, quando processado a curto intervalo de tempo. Posteriormente. é aplicado um banco de filtros em escala mel com o formato triangular não separados linearmente.

A quantidade de filtros, N_f, é determinada por uma relação com a frequência de amostragem, F_a, sendo $N_f =$. Em seguida, é feito o cálculo do logaritmo da energia de saída de cada filtro para a obtenção do cepstro e, por fim, os coeficientes mel-cepstrais podem ser determinados por meio da equação (8) (COSTA et al., 2008).

$$c_{mel}(n) = \Sigma_{k=1}^{N_f} log\left(S_{FFT}(k)\right).cos\left[n\left(k - \frac{1}{2}\right)\right].\frac{\pi}{N_f}, \quad (8)$$

onde N_f é o número de filtros digitais; $c_{mel}(n)$ corresponde ao n-coeficientes mel-cepstrais e o $S_{FFT}(k)$o sinal de saída do bando de filtros digitais que é obtido por meio da equação (9).

$$S_{FFT}(k) = \Sigma_{k=1}^{N_f} W(j) \cdot X(j), \;\; k = 1, \ldots, N_f, \quad (9)$$

onde $W(j)$ correspondem às janelas de ponderação triangulares associadas a escala mel e $X(j)$ corresponde o espectro da FFT para n pontos (COSTA et al., 2008).

A escala Mel é projetada para refletir a percepção auditiva humana, onde pequenas variações em frequências baixas são mais perceptíveis do que variações em frequências altas. Portanto, os filtros na parte inferior do espectro de frequências são mais estreitos e próximos entre si, enquanto os filtros nas frequências mais altas são mais largos e mais espaçados. Após a aplicação desses filtros, o sinal de saída passa por uma transformação cepstral, que gera 12 coeficientes mel-cepstrais, representando as características acústicas mais relevantes do sinal de voz (MA; FOKOUÉ, 2014).

Para a extração dos MFCC, os sinais de áudio foram pré-processados no formato adequado, garantindo a remoção de ruídos e a normalização da amplitude. Posteriormente, os dados foram importados para o *software* WEKA, versão 3.9.5, onde o

filtro de extração de características foi aplicado. O algoritmo utilizado permitiu a decomposição das amostras de áudio em 12 coeficientes mel-cepstrais, que capturam as nuances de frequência relacionadas às emoções expressas na voz.

4.4 Variáveis

As variáveis independentes são os vetores de características de dados vocais extraídos pela análise acústica: medidas de f_o média, máxima, mínima e dp (Hz), *Jitter* (%), *Shimmer* (dB), CPPS (dB), média e máximo de GNE (Hz), HNR (Hz), coeficientes mel-cepstrais MFCC, medidas acústico-prosódicas de f_o média, máxima, mínima e range (Hz), duração (ms), intensidade (dB), e as dimensões de valência, potência e ativação. Ao total foram extraídas 41 medidas acústicas tradicionais, acústico-prosódicas e mel-cepstrais, descritas no Quadro 1, a seguir.

Quadro 1- Descrição das variáveis independentes

VARIÁVEL	**DESCRIÇÃO**
Duração_Al	Duração da unidade VV [Al]
Duração_auav	Duração da unidade VV [auav]
Duração_iaNU	Duração da unidade VV [iaNU]
Duração_az	Duração da unidade VV [az]
z_Al	Valor do z-score da unidade VV [Al]
z_auav	Valor do z-score da unidade VV [auav]
z_iaNU	Valor do z-score da unidade VV [iaNU]
z_az	Valor do z-score da unidade VV [az]
zsuave_Al	Valor do z-score suavizado da unidade VV [Al]
zsuave_auav	Valor do z-score suavizado da unidade VV [auav]
zsuave_iaNU	Valor do z-score suavizado da unidade VV [iaNU]
zsuave_az	Valor do z-score suavizado da unidade VV [az]
f_o min	Mínima da f_o
f_o max	Máxima da f_o
f_o média	Média da f_o
f_o range	Intervalo f_o (máxima - mínima)
f_o DP	Desvio Padrão da f_o
Intensidade	Intensidade média do som em decibéis
Jitter	Variação de ciclo a ciclo na f_o
Shimmer	Diferença média absoluta entre as amplitudes de períodos consecutivos, dividida pela amplitude média
HNR	Média da relação harmônico-ruído
GNE 1	Taxa de excitação glotal-para-ruído com largura de banda de 1000 Hz
GNE 2	Taxa de excitação glotal-para-ruído com largura de banda de 2000 Hz

	(continuação)
GNE3	Taxa de excitação glotal-para-ruído com largura de banda de 3000 Hz.
CPPS média	Proeminência do pico cepstral suavizado média
CPPS DP	Proeminência do pico cepstral suavizado desvio padrão
x1MFCC	Coeficiente 1 do MFCC
x2MFCC	Coeficiente 2 do MFCC
x3MFCC	Coeficiente 3 do MFCC
x4MFCC	Coeficiente 4 do MFCC
x5MFCC	Coeficiente 5 do MFCC
x6MFCC	Coeficiente 6 do MFCC
x7MFCC	Coeficiente 7 do MFCC
x8MFCC	Coeficiente 8 do MFCC
x9MFCC	Coeficiente 9 do MFCC
x10MFCC	Coeficiente 10 do MFCC
x11MFCC	Coeficiente 11 do MFCC
x12MFCC	Coeficiente 12 do MFCC
Valência	Caráter positivo ou negativo da emoção
Potência	Nível de força ou poder que a emoção provoca
Ativação	Nível de energia ou excitação associado à emoção

As variáveis dependentes são as emoções simuladas que se pretende identificar com base nas variáveis acústicas e prosódicas. As emoções incluem alegria, medo, tristeza, raiva, surpresa, nojo e neutra.

A relação entre as variáveis consiste no uso das variáveis independentes (características vocais) como entradas para o modelo que visa prever reconhecer as variáveis dependentes (emoções). A premissa é que diferentes combinações e valores dessas características vocais estejam associadas a estados emocionais específicos, e permitem que o modelo aprenda essas associações e identifique a emoção a partir dos sinais vocais do EmoVox-BR.

4.5 Estrutura do Modelo *Fuzzy*

Para construir o modelo de reconhecimento automático de emoções a partir de amostras vocais de atores simulando emoções, foi utilizada um sistema baseado em regras *fuzzy*. As amostras de áudio foram previamente rotuladas por fonoaudiólogos especialistas, com identificação da emoção simulada em cada gravação. Esse processo favoreceu ao modelo aprender a reconhecer e diferenciar categorias emocionais (emoções rotuladas) de forma autônoma, utilizando critérios próprios. Na etapa final, foram definidas as variáveis linguísticas de entrada, de saída e suas respectivas funções de pertinência, visando otimizar o desempenho e a precisão do modelo. Essa abordagem

considera as incertezas e a natureza imprecisa dos dados de entrada, aspectos inerentes às emoções humanas, para alcançar uma maior taxa de acerto.

As variáveis linguísticas são aquelas que permitem a descrição de informações de forma qualitativa, sendo expressas qualitativamente por termos linguísticos (fornece o conceito para a variável) e, quantitativamente, por uma função de pertinência. Cada variável linguística tem um conjunto de termos *fuzzy* associados que é o conjunto de valores atribuídos à variável *fuzzy* (ZADEH, 1988).

Para expressar conceitos *fuzzy* é comum o uso de elementos qualitativos e quantitativos. Por exemplo, os termos “pequena”, “média” e “grande” são utilizados pela definição da variável linguística (ZADEH, 1988). Por exemplo, a variável linguística “f_o média ($f_o$md)” admite valores linguísticos: muito pequeno, pequeno, médio, alto, muito alto e muito muito alto. Cada um destes valores admite valores numéricos em um intervalo [0, $f_o md_{MAX}$], podendo também ser representado por valores linguísticos sobre o intervalo [0, $f_o md_{MAX}$] por meio de funções de pertinências.

As regras *fuzzy* são estruturas vastamente utilizadas em várias abordagens da teoria *fuzzy*. Elas descrevem situações específicas que podem ser submetidas à análise de um painel de especialistas, e cuja inferência nos conduz a algum resultado desejado. A inferência baseada em regras *fuzzy* pode também ser compreendida como um funcional que mapeia um conjunto de entradas do sistema para um conjunto de saídas (ORTEGA, 2001).

Essas regras possuem uma estrutura básica definida em duas partes: O antecedente (SE), que é composto pelas variáveis de entrada que descrevem uma condição, e o consequente (ENTÃO), composto pelas variáveis de saída, indicando uma conclusão. Vale salientar que no antecedente é permitido o uso de mais de um conector como o “E” e “OU” e no consequente pode haver mais de uma conclusão (ORTEGA, 2001). De forma genérica, uma regra *fuzzy* pode ser representada da seguinte forma:

$$\text{SE } (x \text{ é } a_i) \text{ E } (y \text{ é } b_i) \text{ OU... ENTÃO } (z \text{ é } c_i) \text{ E } (w \text{ é } d_i) \quad (10)$$

onde x e y são as variáveis linguísticas de entrada, z e w as de saída, e a_i, b_i, c_i *e* d_i são realizações dessas variáveis, medidas a partir da interação do especialista com o sistema (GOMIDE; GUDWIN, 1994)

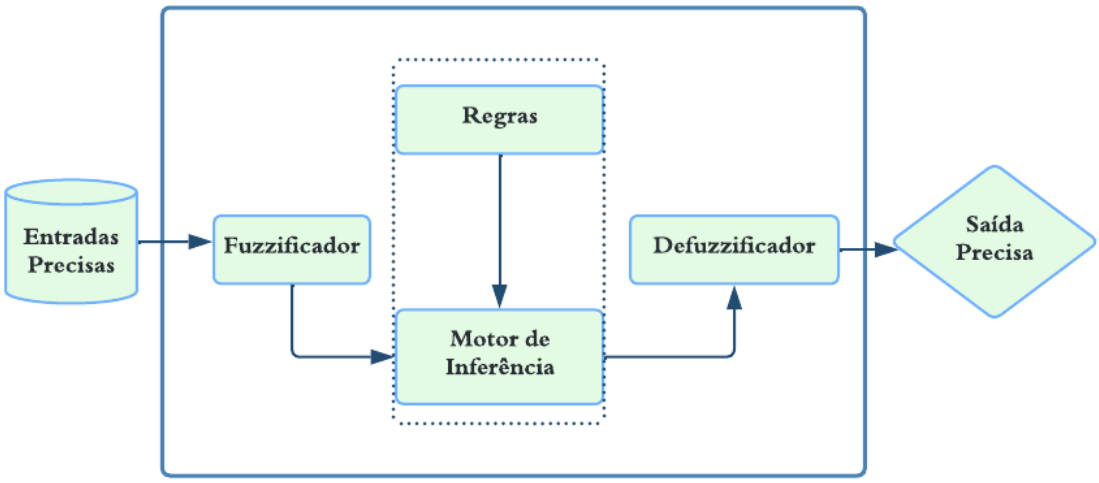
Os antecedentes descrevem uma condição (premissas), enquanto a parte consequente descreve uma conclusão ou uma ação que pode ser esboçada quando as premissas se verificam. A diferença entre os antecedentes de uma regra *fuzzy* e uma regra clássica é que os primeiros descrevem uma condição flexível, ou seja, uma condição que

pode ser parcialmente satisfeita, enquanto os últimos descrevem uma condição rígida (a regra não funciona se os antecedentes não são completamente satisfeitos) (ORTEGA, 2001).

Deste modo, a implementação de um projeto de sistema *fuzzy* pode ser reduzida a um ponto em que problemas anteriormente de difícil tratamento passam a ser factíveis de mais simplificada solução (ZADEH, 1973).

A figura 2 representa a estrutura genérica de um sistema baseado em lógica *fuzzy*, conforme o modelo proposto por Mamdani (1974). Este sistema utiliza entradas precisas, que são convertidas em valores *fuzzy* por meio do processo de fuzzificação, no qual funções de pertinência atribuem graus de pertencimento as variáveis linguísticas. As regras do sistema, estruturadas por meio do conhecimento, são aplicadas pelo motor de inferência, que combina os valores fuzzificados para gerar uma saída. Posteriormente, o defuzzificador transforma essa saída em um valor preciso, pronto para uso em aplicações práticas (Figura 2).

Figura 2 - Estrutura básica de um sistema *fuzzy*



Fonte: Elaborado pelo autor, 2024.

4.5.1 Processador de entrada (Fuzificador)

No processo de fuzzificação, uma ou mais funções de fuzzificação são especificadas para cada variável do modelo, com o objetivo de transformar valores de entrada, representados por números reais, em conjuntos *fuzzy* predefinidos pelo usuário

(KLIR; YUAN, 1995). permitindo que informações contínuas ou discretas sejam analisadas de forma qualitativa.

A fuzzificação é realizada utilizando uma função de pertinência, que calcula o grau de pertinência de cada valor de entrada em relação a um conjunto *fuzzy* previamente definido em uma variável linguística. Esse grau de pertinência é um valor numérico dentro do intervalo [0, 1], e representa a intensidade com que o valor pertence ao conjunto *fuzzy* correspondente. O resultado da fuzzificação, que inclui os conjuntos *fuzzy* relevantes e seus respectivos graus de pertinência, é então encaminhado para o motor de inferência. Este utilizará as informações fuzzificadas em conjunto com regras *fuzzy* previamente definidas para determinar as saídas do sistema, disparando aquelas que forem compatíveis com as condições de entrada (PASSINO; YURKOVICH, 1998). Além disso, o processo de fuzzificação é frequentemente acompanhado de estratégias de pré-processamento, como normalização das entradas ou ajustes dinâmicos das funções de pertinência, para garantir maior precisão e adaptabilidade em aplicações práticas.

4.5.2 Base de conhecimento (Regras)

A base de conhecimento contém informações gerais relativas ao domínio do problema em análise e, juntamente com o motor de inferência, pode ser considerada o núcleo dos sistemas baseados em regras *fuzzy*. As informações existentes na base de conhecimento são expressas através de proposições e regras de produção *fuzzy* (KLIR; YUAN, 1995), as quais compõem a base de regras do sistema.

Uma base de regras *fuzzy* pode ser construída manualmente com o auxílio de um especialista, ou de forma automática através de geradores de regras *fuzzy* a partir de dados tabulados. Independentemente de como foi gerada, a base de regras é composta por uma coleção de proposições condicionais *fuzzy* que devem descrever linguisticamente o conhecimento capturado (ZADEH, 1965). Cada regra *fuzzy* é formada por uma parte antecedente e uma parte consequente (ORTEGA, 2001). O resultado obtido das relações representadas pelas regras será utilizado pelo motor de inferência do sistema. Outras características importantes de uma base de regras *fuzzy* são a consistência e a completude. Uma base de regras *fuzzy* consistente, não possui regras com consequentes conflitantes e que possam ser ativadas simultaneamente (PEDRYCZ; GOMIDE, 1998).

Com base nas características acústicas extraídas, regras *fuzzy* serão desenvolvidas para modelar a relação entre parâmetros vocais e estados emocionais,

utilizando uma combinação de conhecimento empírico e teórico da literatura sobre voz e emoção.

4.5.3 Motor de inferência

Nesta etapa, define-se quais as normas e regras de inferência serão utilizadas na obtenção da relação *fuzzy* que modela a base de regras (KAYACAN; KHANESAR, 2016). Inicialmente, processam-se os dados *fuzzy* de entrada juntamente com as regras, aplicando as *t-normas*, *t-conormas* e as regras de inferência para determinar os conjuntos *fuzzy* de saída e modelar a base de regras (PASSINO; YURKOVICH, 1997).

O processo de inferência de Mamdani (1974) utiliza o operador de implicação mínimo (*t-norma*) que é responsável por modelar a relação lógica "SE... ENTÃO..." em um contexto de incerteza e as regras da base de conhecimento para associar as características vocais às emoções correspondentes, manipulando as incertezas e ambiguidades nas variações emocionais da voz (KAYACAN; KHANESAR, 2016). No processo de inferência *fuzzy*, o operador de implicação é aplicado para calcular o grau de verdade do consequente a partir do antecedente, conforme definido em regras *fuzzy* do tipo "SE intensidade é alta, ENTÃO emoção é raiva". Após a aplicação do operador de implicação para todas as regras relevantes, os resultados são combinados por meio de agregação *fuzzy*, posteriormente, o grau *fuzzy* é convertido em um valor exato no processo de defuzzificação.

O processo de inferência gera um grau de pertinência de ativação para cada regra, onde ocorre a implicação mediante a aplicação do operador escolhido. O último passo do processo é o operador de agregação sobre todos os valores resultantes da implicação de cada regra para a geração de um conjunto *fuzzy* único que será passado para a interface de defuzificação (MORAES, 1998).

Entre os principais operadores de agregação estão o máximo (*t-conorma*), que seleciona o maior grau de pertinência. Outros operadores incluem a média, que balanceia os valores de saída ao calcular a média aritmética, o produto que pondera os resultados de forma mais restritiva, e o mínimo, que avalia o menor grau de pertinência e permite representar de maneira conservadora e intuitiva as relações lógicas "SE... ENTÃO...", evitando superestimações de ativação nas regras *fuzzy* (MEGRI; BOUKEZZOULA, 2008; ZADEH, 1965). Para cada característica vocal, serão definidas funções de pertinência que determinam o grau de correspondência entre os valores das características e as categorias emocionais.

A implicação consiste na formulação de uma conexão entre causa e efeito, ou uma condição e sua consequência. Para a realização da implicação podem ser utilizados vários operadores (Tabela 2) (MORAES, 1998). Neste estudo foi utilizado o operador de Zadeh.

Tabela 2 - Principais operadores de implicação

Operador de Implicação	**Nome**
min (a,b)	*Zadeh*
1 – a + ab	*Reichenback*
min (1 - a + b , 1)	*Lukasiewicz*
max (1-a, b)	*Kleene-Diemes*

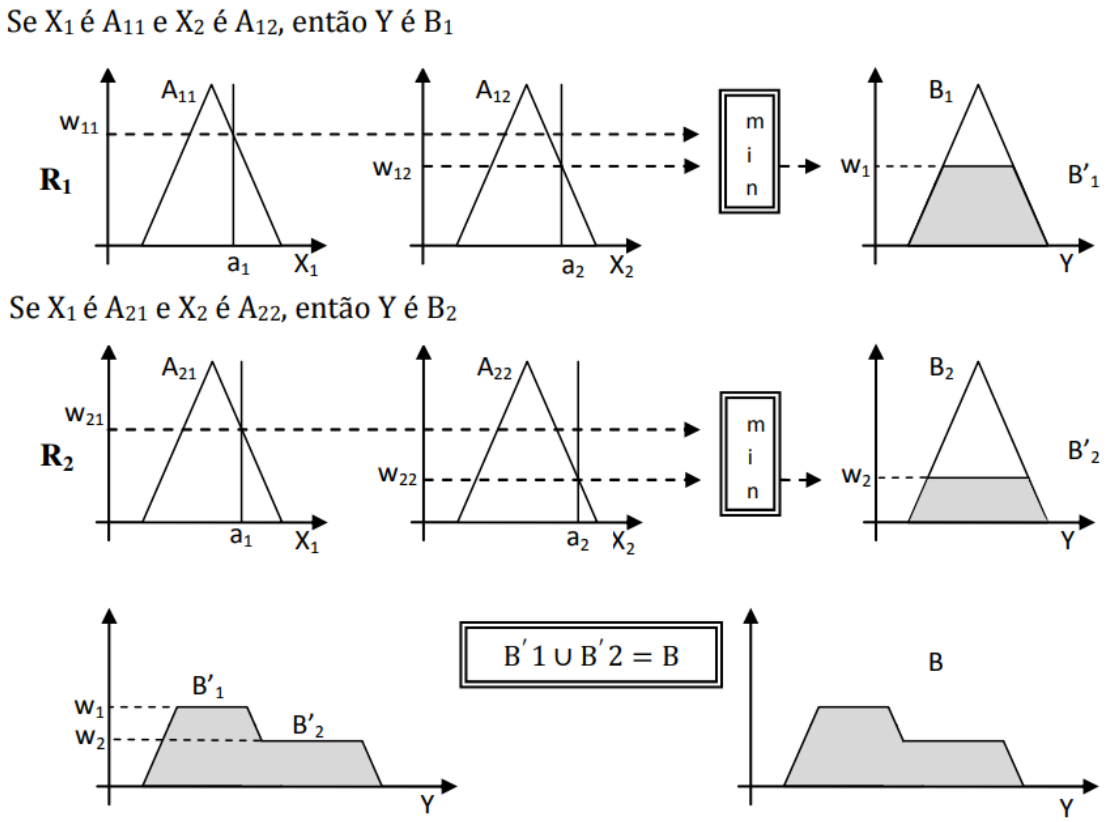
Fonte: MORAES, 1998.

Existem diversos métodos de inferência, deve-se, portanto, escolher aquele que melhor se adapta ao sistema que está sendo modelado. Um dos métodos mais utilizados é o método de Mamdani (1974). O processo de inferência do tipo Mamdani pode ser descrito em quatro etapas principais:

1. Identificação das regras compatíveis e cálculo do grau de pertinência dos valores de entrada para cada proposição no antecedente das regras.
2. Combinação dos graus de pertinência calculados utilizando a *t-norma* padrão de interseção (min).
3. Determinação do grau de ativação do consequente, com base no grau de pertinência resultante, utilizando novamente a *t-norma* padrão de interseção.
4. Agregação dos resultados individuais das regras compatíveis por meio da s-norma padrão de união.

A figura 3 apresenta graficamente um exemplo do modelo de inferência de Mamdani (Figura 3).

Figura 3 - Inferência pelo método de Mamdani



Fonte: Santos (2009)

É importante destacar que o antecedente pode conter mais de duas proposições. A seguir, é apresentado um exemplo do processo de inferência utilizando o método de Mamdani:

Regra1: Se V_1 é A_{11} e V_2 é A_{12}, então U1 é B_1
Regra2: Se V_1 é A_{21} e V_2 é A_{22}, então U2 é B_2
Condição observada: V_1 é a_1' e V_2 é a_2'
Conclusão: U é Y' (11)

$w_{11} = A_{11}(a_1')$ $w_{12} = A_{12}(a_2')$
$w_{21} = A_{21}(a_1')$ $w_{22} = A_{22}(a_2')$

Assim, o grau de ativação das regras 1 e 2 será determinado por w_1 e w_2, calculados respectivamente da seguinte forma:

$$w_1 = w_{11} \wedge w_{12}$$
$$w_2 = w_{21} \wedge w_{22} \quad (12)$$

As saídas das regras serão então calculadas utilizando o operador de mínimo:

$$B_1'(y) = w_1 \wedge B(y)_1 \qquad (13)$$

$$B_2'(y) = w_2 \wedge B(y)_2$$

Por fim, a saída global será obtida por meio da agregação utilizando o operador de máximo:

$$B'(y) = B_1'(y) \vee B_2'(y) \qquad (14)$$

A saída global gerada pelo método de Mamdani deve ser submetida a um processo de defuzzificação, permitindo que o usuário obtenha um valor exato (crisp) (Mamdani, 1974).

4.5.4 Processador de saída (Defuzificador)

A etapa final do processo de inferência *fuzzy*, chamada de defuzzificação, transforma os conjuntos *fuzzy* em números reais, facilitando a obtenção de uma representação numérica que reflete o estado emocional mais provável. Esse processo traduz o conjunto *fuzzy* resultante para valores precisos (KAYACAN; KHANESAR, 2016). Diversos métodos são apresentados para este processo na literatura, sendo os principais: *Centroid or Center of Gravity (*CoG*), Center of Sums (*CoS*), Center of Largest Area (*CoLA*), First of Maxima (*FoM*), Last of Maxima (*LoM*)* e *Mean of Maxima (*MoM*)* (LIU, 2007; JAIN et al., 2022).

No presente estudo, o método de defuzzificação utilizado foi o CoLA, que identifica o valor numérico correspondente ao centro da figura de maior área dos conjuntos fuzzy resultantes. Esse método se baseia na seleção da região mais significativa do conjunto fuzzy e no cálculo do centro dessa área, considerando a importância relativa das variáveis no modelo.

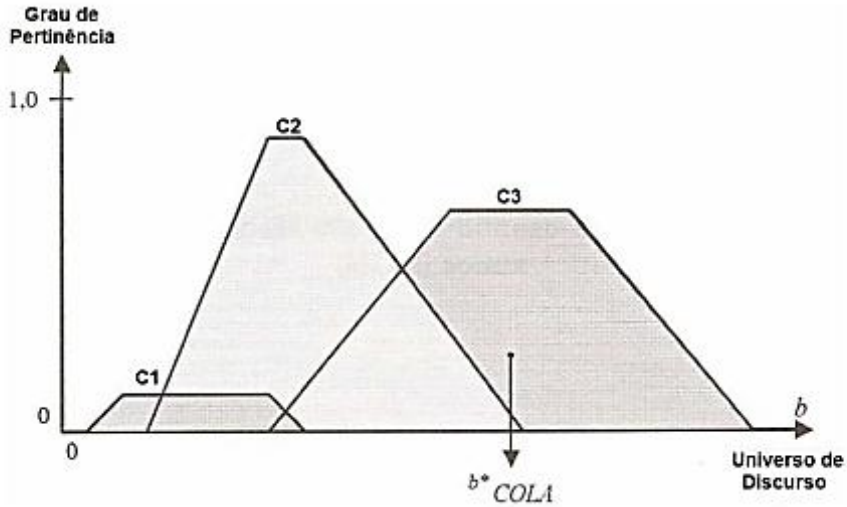
Se existirem pelo menos duas sub-regiões convexas no conjunto *fuzzy* resultante, então o centro de gravidade (isto é, a abordagem do centróide) é utilizado para calcular z∗) da sub-região convexa com a maior área, que é empregada para determinar o valor de z∗) da saída (VAN LEEKWIJCK; KERRE, 1999; SUGENO, 1985)
Essa condição pode ser expressa na forma algébrica como:

$$z^* = \frac{\int \mu\tilde{c}_m(z) z dz)}{\int \mu\tilde{c}_m(z) z dz}$$

onde $\tilde{c}_m$ é o segmento convexo com a maior área que compõe $\tilde{c}_k$. Quando todo o conjunto fuzzy resultante $\tilde{c}_k$ é não convexo, essa condição é válida. Quando $\tilde{c}_k$ é convexo, o valor de z$*$) é igual ao valor obtido utilizando a técnica do centróide, já que há apenas uma única zona convexa. A equação acima pode ser usada para encontrar z$*$) utilizando a abordagem do centro da maior área. Como todo o conjunto *fuzzy* resultante é convexo, a abordagem do centro da maior área gera o mesmo resultado que o método do centróide, resultando em z$*$= 4,9.

O CoLA é indicado quando as saídas do processo de inferência *fuzzy* são formadas por conjuntos de múltiplas áreas (WEBER e KLEIN, 2003). A figura 4 mostra o resultado da defuzzificação pelo método CoLA.

Figura 4 - Defuzzificação pelo método CoLA



Fonte: WEBER e KLEIN, 2003.

O CoLA é indicado para situações onde diferentes regiões apresentam graus de pertinência variados, pois prioriza a área com maior relevância, proporcionando maior precisão e representatividade no processo de defuzzificação (KOVACIC E BOGDAN 2006).

4.5.5 Módulo Fuzzy Rules

Para o desenvolvimento do modelo de reconhecimento dos estados emocionais a partir da voz, utilizou-se o pacote *FuzzyRules*, implementado no *software* R, versão 4.3.3

por Moraes, Machado e Ferreira (2024) disponível em: https://github.com/leapigufpb/FuzzyRules. Este módulo permite a criação de sistemas baseados em lógica *fuzzy*, abrangendo a definição de funções de pertinência, operações em conjuntos *fuzzy* e métodos de defuzzificação.

Na definição das funções de pertinência para as variáveis de entrada, o módulo oferece suporte a diferentes formatos, incluindo triangular, trapezoidal, constante, singleton e gaussiana. No entanto, com base na análise dos histogramas das variáveis utilizadas, foi selecionada a função trapezoidal como a mais adequada. Esse formato apresentou maior correspondência com as distribuições observadas e garantiu uma modelagem mais precisa das características vocais associadas a cada emoção básica.

Adicionalmente, foram implementadas operações em conjuntos *fuzzy*, utilizando *t-normas* e *t-conormas* de Zadeh, Probabilística e de Lukasiewicz. Essas operações possibilitaram a manipulação eficiente dos conjuntos *fuzzy* associados às variáveis de entrada para modelar as interações entre diferentes características acústicas de forma mais robusta.

Em relação à lógica *fuzzy*, foram realizadas operações lógicas como AND, OR e NOT para múltiplos conjuntos *fuzzy*, possibilitando a construção de regras que relacionam as variáveis de entrada aos estados emocionais. O modelo de inferência adotado foi o do tipo Mamdani, amplamente reconhecido por sua habilidade de lidar com incertezas e por fornecer uma representação intuitiva das relações entre as variáveis e os estados emocionais (MAMDANI, 1974).

Para a defuzzificação, o módulo oferece diversos métodos, como CoG*,* CoS*, Center* CoLA*,* FoM, LoM e MoM. Optou-se pelo método CoLA devido à sua eficácia em fornecer resultados representativos em cenários com múltiplos picos nas funções de pertinência, característica presente nas distribuições das variáveis analisadas. O CoLA destacou-se por priorizar a maior área da função de pertinência, proporcionando maior estabilidade nos valores defuzzificados e melhorando a precisão do modelo de reconhecimento das emoções.

A utilização do módulo *FuzzyRules* no R possibilitou a implementação eficiente do modelo, integrando diversas técnicas avançadas de lógica *fuzzy*. Essa abordagem garantiu maior robustez e precisão na identificação das emoções, destacando-se como uma solução inovadora e eficaz no reconhecimento automático das emoções.

4.5.6 Análise de desempenho do modelo *fuzzy*

O modelo *fuzzy* será avaliado de acordo com os seguintes indicadores para validar o modelo lógico:

Acurácia: percentual de emoções corretamente identificadas pelo sistema.

Sensibilidade: capacidade do modelo de detectar corretamente uma emoção presente.

Especificidade: capacidade do modelo de identificar corretamente a ausência de uma emoção específica.

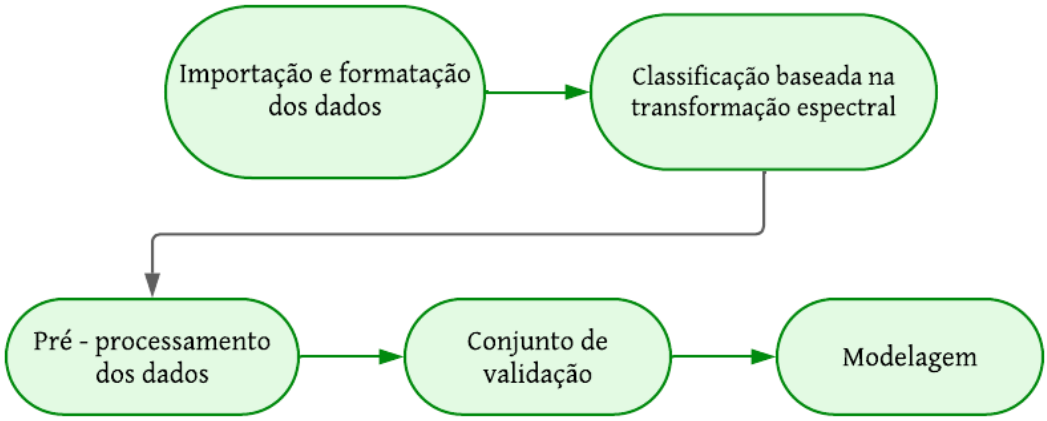
Os resultados apresentados serão analisados levando em consideração a Matriz de Confusão, o coeficiente de Acurácia Geral e o coeficiente Kappa, para avaliar a precisão dos resultados obtidos através do modelo.

4.5.7 Validação do modelo de reconhecimento das emoções *fuzzy*

Realizou-se uma análise comparativa entre o desempenho do modelo *fuzzy* e outros modelos tradicionais de reconhecimento de emoções, como Linear SVM, Random Forest, Kernel SVM, Redes Neurais MLP, Redes Neurais MLP combinadas através de Bagging, Árvore de decisão, KNN, Linear SVM, Redes Naive Bayes, Extreme Gradient Boosting (XGBoost) e Elastic Net (EL Net). As métricas de desempenho foram comparadas entre os diferentes modelos para validar a eficácia do modelo *fuzzy*. Vale destacar que essa comparação visou validar a viabilidade e o potencial do modelo *fuzzy*, sem estabelecer hierarquias rígidas de desempenho, mas ressaltando sua aplicabilidade e contribuição no campo investigado.

O fluxograma a seguir indica as etapas realizadas para o desenvolvimento dos modelos de aprendizado de máquinas (Figura 5).

Figura 5 - Etapas para a construção dos modelos baseados em aprendizado de máquinas



Fonte: Elaborado pelo autor, 2024.

4.5.7.1 Importação e formatação dos dados

Inicialmente foram realizadas as etapas de importação e formatação dos dados. O processamento inicial do conjunto de dados apresentou as medidas vocais e categorias emocionais. Posteriormente, a variável emoção foi convertida em um fator com rótulos descritivos, categorizando as emoções alegria, tristeza, medo, raiva, surpresa, nojo e neutra.

4.5.7.2 Classificação baseada na transformação espectral

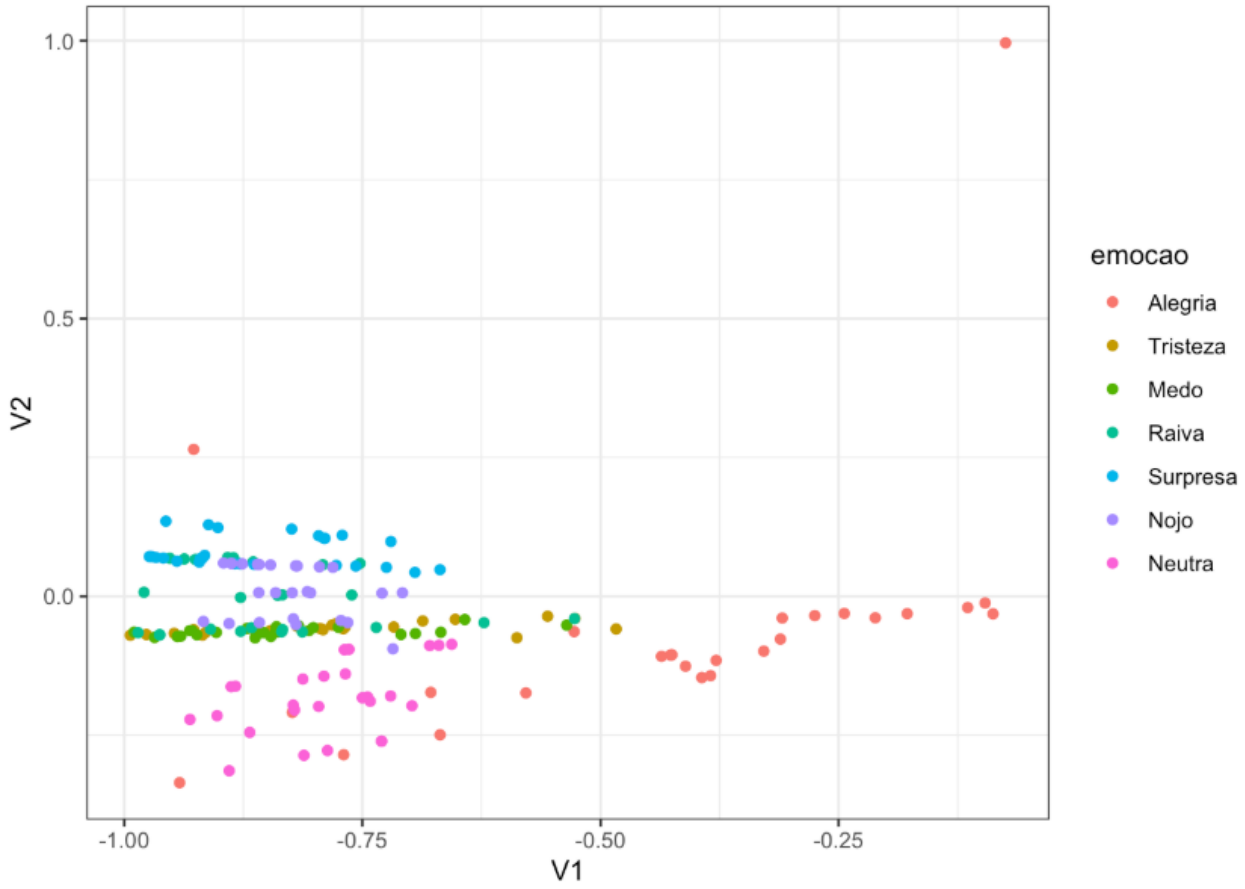
Os testes realizados inicialmente com as variáveis originais não apresentaram resultados satisfatórios em termos de precisão, devido a sobreposição existente entre as variáveis, o que levou à necessidade de aplicar transformações espectrais como a Análise Espectral de Grafos (AEG). Essa transformação permitiu explorar e analisar as relações entre os objetos em um conjunto de dados que podem ser descritos por grafos. O método utiliza a decomposição espectral do Laplaciano do grafo, derivado da matriz de similaridade dos dados, para capturar propriedades estruturais e relacionais. A AEG possibilita a clusterização de objetos com base em relações de similaridade, representando esses objetos como nós em um grafo e conectando-os com arestas ponderadas que refletem o grau de similaridade (BORKOWSKY et al., 2023).

O método espectral facilitou a atribuição de novos dados a clusters já existentes, uma vez que a classificação pode ser realizada com base em critérios de similaridade entre espectros de autovalores, garantindo maior coerência na integração de dados previamente não rotulados. Além disso, o uso de autovalores permite superar limitações de escalabilidade e precisão, comuns em métodos que dependem exclusivamente de autovetores ou de clusterizações estáticas. Essa técnica torna-se eficaz para dados heterogêneos ou com sobreposição de classes (BORKOWSKY et al., 2023). A transformação espectral permitiu decompor o sinal em componentes de frequência, essenciais para o entendimento de características acústicas específicas e para a redução de ruídos ou variações comprometeram a precisão das análises. Essa técnica foi aplicada para facilitar a discriminação entre diferentes classes de sinais e contribuiu para a melhoria da acurácia dos modelos.

Inicialmente, a matriz de adjacência foi construída usando uma métrica de similaridade baseada em distâncias euclidianas, ajustadas por um parâmetro de escala kpar, com valores abaixo de um limiar sendo truncados. Em seguida, foram calculadas as matrizes de grau e o Laplaciano normalizado, que representaram a estrutura relacional dos dados. A decomposição espectral do Laplaciano foi realizada para obter os autovalores e autovetores, sendo os autovetores correspondentes às menores dimensões selecionados para formar a nova representação. Por fim, os dados foram normalizados para garantir a consistência geométrica da transformação, resultando em uma representação reduzida que preserva as características estruturais mais relevantes do conjunto original.

A figura 6 representa o processo de aplicação da transformação espectral em um conjunto de dados para redução de dimensionalidade e visualização. Inicialmente, estimou-se o parâmetro de kernel, que ajustou a escala na construção da matriz de adjacência. A matriz foi criada a partir do conjunto de dados original, excluindo a variável categórica de emoção, e a transformação espectral foi realizada, projetando os dados em um espaço de cinco dimensões reduzidas. Os dados transformados foram então convertidos em um formato tabular e combinados com a variável de emoção original. A visualização em duas dimensões foi gerada utilizando os dois primeiros componentes principais (V1) e (V2) da projeção, com as cores representando as diferentes categorias emocionais. Cada ponto representa uma observação no conjunto de dados, posicionada em relação às dimensões (V1) e (V2), enquanto as cores indicam as emoções associadas (Figura 6).

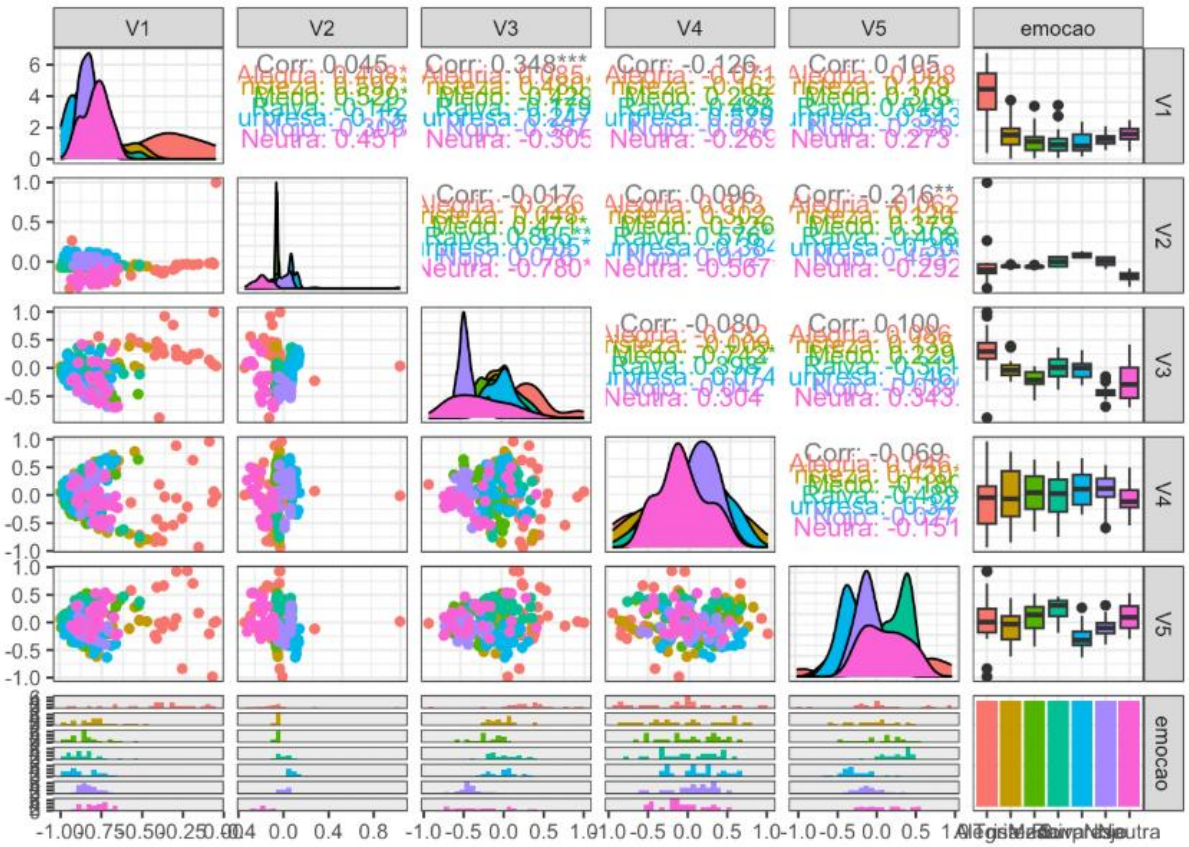
Figura 6 - Transformação espectral aplicada no conjunto de dados



Observa-se como as diferentes emoções se distribuíram no espaço projetado, identificando padrões ou clusters de acordo com a proximidade espacial de observações pertencentes à mesma categoria emocional.

A figura 7 exibe a relação entre as variáveis específicas (V1) a (V5) por meio de diagramas de dispersão, com as cores indicando as diferentes emoções para identificar possíveis clusters ou tendências associadas às emoções. As distribuições univariadas, exibidas nos gráficos diagonais, mostram como as variáveis se distribuíram individualmente para cada emoção (Figura 7).

Figura 7 - Relação entre as variáveis pós transformação espectral



Os painéis laterais, como os *boxplots* para cada variável por emoção, destacaram as diferenças na mediana e dispersão das dimensões reduzidas entre as categorias emocionais. Essas informações são valiosas para observar como as emoções se distinguem em cada dimensão e podem indicar quais variáveis têm maior poder discriminatório. As variáveis V1 a V5, representam os componentes principais do conjunto de variáveis, destacaram-se nos painéis laterais, como os *boxplots* por emoção, evidenciando diferenças na mediana e dispersão entre as categorias emocionais. As correlações apresentadas nos painéis numéricos avaliam a força e direção das relações lineares entre as variáveis (V1) a (V5) que ajudou a identificar redundâncias ou independências entre dimensões.

4.5.7.3 Pré-processamento dos dados

O pré-processamento dos dados incluiu diversas etapas para garantir a qualidade e a padronização das variáveis. Inicialmente, as variáveis numéricas foram normalizadas, ajustando-se para que apresentassem média igual a zero e desvio-padrão igual a um, para melhorar a performance de modelos sensíveis à escala dos dados. Além disso, foram implementadas alternativas para lidar com valores ausentes, como imputação por KNN, considerando k=5, *bagged trees*, média, mediana ou moda, além da possibilidade de exclusão direta de instâncias com valores ausentes. Dessa forma, para cada instância com o valor ausente, o algoritmo K-NN calculou a distância entre as instâncias no conjunto de dados, por meio da métrica de distância euclidiana ao quadrado (IMANDOUS; BOLANDRAFTAR, 2013).

Transformações adicionais, como logaritmo, polinômios de grau até 4 e Yeo-Johnson, também foram previstas para tratar assimetrias e preparar as variáveis para modelagem. Foi realizada a remoção de preditores altamente correlacionados (*threshold* = 0,6, correlação de *Spearman*), reduzindo o risco de multicolinearidade, e a exclusão de variáveis com variância próxima de zero, para melhorar a eficiência computacional.

O conjunto de dados foi então pré-processado e estruturado para garantir que as transformações fossem aplicadas de maneira consistente ao conjunto de treinamento. Esse processo assegurou a preparação adequada dos dados para análises e modelagens subsequentes.

4.5.7.4 Conjunto de Validação

Foi utilizado o método *k-fold cross-validation* para construir um conjunto de validação com *k folds* e considerado um procedimento com k = 10 *folds*, repetido uma vez. Os dados foram particionados em 10 partes utilizando amostragem estratificada com base na variável emoção. Em cada iteração, os modelos foram ajustados em um conjunto de treinamento composto por 9 dessas partes e avaliados em um conjunto de teste composto pela parte restante.

Esse procedimento foi utilizado para determinar os valores otimizados dos hiperparâmetros dos modelos, garantindo representatividade das classes e uma avaliação robusta do desempenho do modelo.

4.5.7.5 Modelagem

Para a construção dos aprendizagem de máquina foram considerados os seguintes modelos: Linear SVM, Kernel SVM, KNN, Árvore de Decisão, Random Forest, Redes Neurais MLP, Redes Neurais MLP combinadas através de Bagging, Redes Naive Bayes, EL Net e XGBoost.

Os critérios de categorização para as métricas de acurácia e sensibilidade dos modelos foram adaptados com base em intervalos de referência: excelente (valores superiores a 0,90), bom (entre 0,80 e 0,90), aceitável (entre 0,70 e 0,80), ruim (entre 0,60 e 0,70) e inaceitável (inferior a 0,60) (HOSMER; LEMESHOW, 2000). Para o coeficiente Kappa, foi considerado os valores segundo Landis; Koch (1977).

Com esses parâmetros, foram avaliados os desempenhos dos modelos de aprendizado de máquina utilizados para o reconhecimento das emoções a partir da voz. Modelos com acurácia, sensibilidade e especificidade inferiores a 0,80 foram considerados insuficientes para este estudo, pois valores abaixo desse limiar indicam uma baixa capacidade de reconhecimento (YER et al., 1991).

Todo o processo de construção dos modelos de aprendizado de máquinas foi realizado utilizando o software R, versão 4.3.3.

5 RESULTADOS

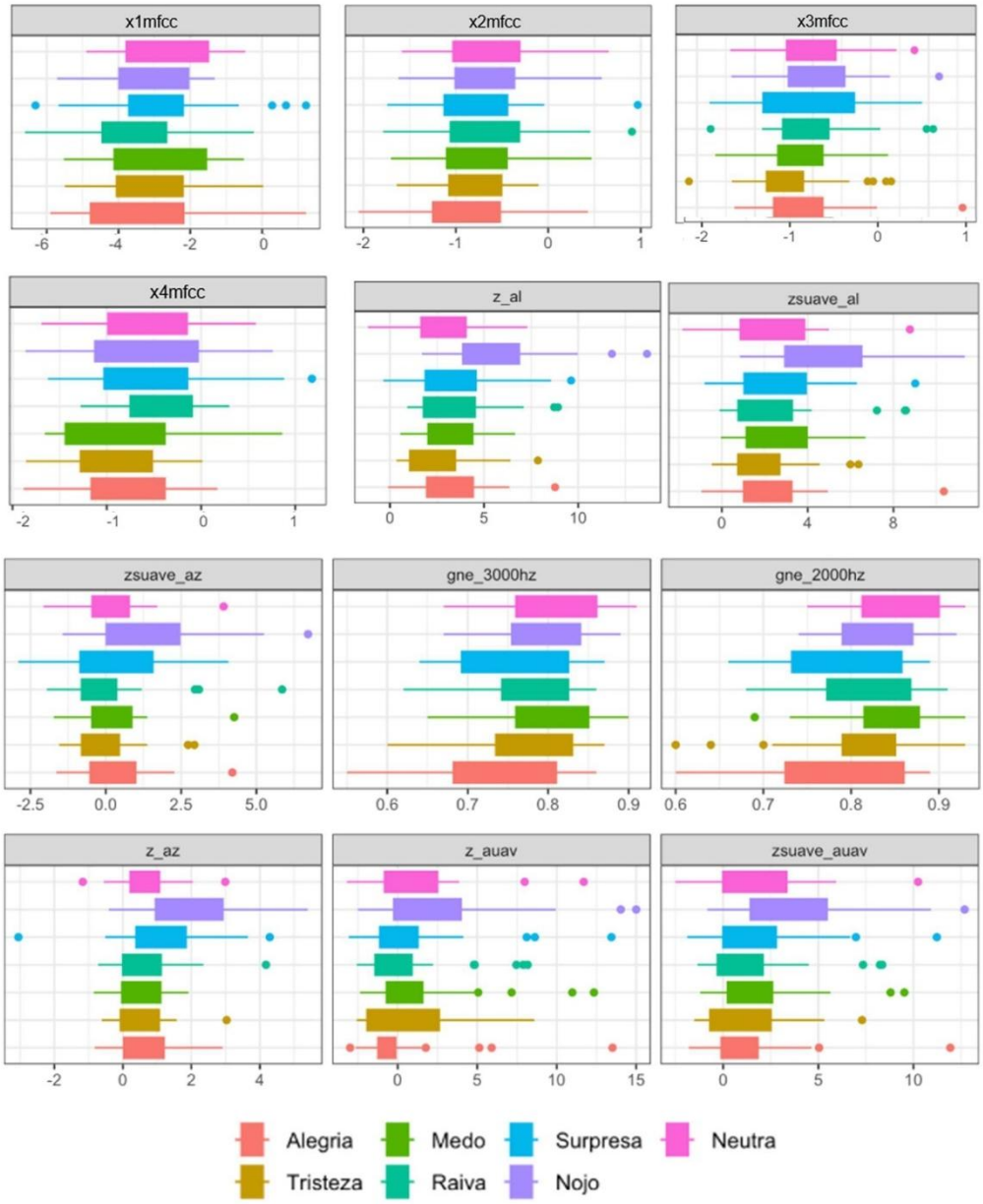
Os resultados a seguir descrevem o desempenho do modelo *fuzzy* baseado em regras desenvolvido para o reconhecimento das emoções a partir da voz. A construção do modelo fundamentou-se em uma abordagem que integra parâmetros acústicos e acústico-prosódicos relevantes, cuidadosamente selecionados para maximizar a precisão no reconhecimento emocional.

A seleção das variáveis que compuseram o modelo *fuzzy* foi realizada de forma criteriosa, considerando tanto a relevância individual quanto a interação entre elas. Utilizou-se inicialmente a construção de *boxplots* para cada variável para realizar a análise das variáveis e identificar possíveis interseções que poderiam comprometer o desempenho do modelo *fuzzy*. Essa abordagem gráfica permitiu observar a distribuição dos valores de forma visual, para facilitar a identificação de padrões similares entre as variáveis. Foram analisados aspectos como a posição da mediana, a amplitude do intervalo interquartil e a presença de *outliers*. Variáveis com distribuições muito semelhantes, que refletiam padrões consistentes em seus *boxplots*, foram consideradas redundantes. Essa análise forneceu subsídios iniciais para avaliar a sobreposição de informações entre variáveis e direcionar etapas subsequentes de seleção ou exclusão, com o objetivo de simplificar o modelo e melhorar sua precisão na diferenciação emocional.

Parâmetros acústicos como os MFCC, os parâmetros de duração z-score e z-suavizado, e de GNE foram excluídos devido ao fato de apresentarem intersecções com outras variáveis. Além disso, a inclusão dessas variáveis piorou o desempenho do modelo, aumentou o ruído no sistema e reduziu sua precisão. Essa exclusão foi fundamental para otimizar a simplicidade do modelo *fuzzy* e aprimorar sua capacidade de diferenciar as emoções de maneira eficiente e robusta, com foco exclusivamente nas variáveis que contribuem de maneira efetiva para a identificação precisa das emoções.

Os *boxplots* de algumas variáveis excluídas destacaram as distribuições dessas variáveis e suas características principais, que oferece uma visão detalhada sobre os dados excluídos do modelo principal (Figura 8).

Figura 8 - *Boxplots* das variáveis excluídas do modelo *fuzzy*



A seleção criteriosa das variáveis no modelo *fuzzy*, verificou as variáveis mais eficazes para capturar diretamente as nuances prosódicas associadas às expressões emocionais humanas. Parâmetros como as variações da f_0, duração e intensidade foram mantidas no modelo, complementados por medidas acústicas adicionais, como HNR, CPPS, *jitter*, *shimmer*, bem como pelos eixos dimensionais das emoções (valência, potência e ativação). Essas variáveis demonstraram um impacto positivo na precisão e no desempenho do sistema, que proporcionou uma análise mais robusta e diferenciada dos estados emocionais.

A inclusão dessas variáveis reforçou a capacidade do modelo de captar nuances emocionais de maneira robusta e com maior exatidão, com diferenciação das emoções de forma mais confiável e detalhada. A exclusão de parâmetros menos relevantes, como as medidas cepstrais, e a incorporação de variáveis acústicas e dimensionais bem fundamentadas não apenas simplificaram o modelo, mas também aumentaram sua robustez e aplicabilidade, assegurando que o sistema responde com precisão às variações emocionais. Dessa forma, a construção do modelo foi orientada para maximizar a eficácia do reconhecimento emocional, com garantia da eficiência no desempenho, além de evitar variáveis que poderiam comprometer a diferenciação entre emoções, o que fortalece sua aplicabilidade em cenários práticos.

5.1 Modelo baseado em regras *fuzzy*

Os parâmetros acústicos e acústico-prosódicos foram usados como variáveis linguísticas de entrada para o sistema baseado em regras *fuzzy*, configurando a última etapa do modelo de reconhecimento dos estados emocionais. As emoções básicas, mais a emissão neutra, constituíram as variáveis linguísticas de saída. A Tabela 3 apresenta as variáveis linguísticas para o conjunto completo de casos, detalhando seus termos linguísticos e universo de discurso.

Tabela 3 - Variáveis linguísticas do sistema *fuzzy* baseado em regras para o modelo de reconhecimento das emoções a partir da voz.

VARIÁVEIS LINGUÍSTICAS	TERMOS LINGUÍSTICOS	DOMÍNIO (SUPORTE)
Variáveis de entrada		
Duração [al]	duracao_alMuitoMuitoPequena	[0, 0, 140, 150]
	duracao_alMuitoPequena	[140, 150, 160, 170]
	duracao_alPequena	[160, 170, 180, 190]
	duracao_alMedia	[180, 190, 200, 210]
	duracao_alAlta	[200, 210, 220, 230]
	duracao_alMuitoAlta	[220, 230, 250, 270]
	duracao_alMuitoMuitoAlta	[250, 270, 380, 380]
Duração [auav]	duracao_auavMuitoMuitoPequena	[0, 0, 500, 510]
	duracao_auavMuitoPequena	[505, 515, 520, 560]

(continuação)

Duração [auav]	duracao_auavPequena	[530, 545, 565, 585]
	duracao_auavMedia	[575, 605, 625, 665]
	duracao_auavAlta	[645, 670, 780, 800]
	duracao_auavMuitoAlta	[790, 810, 1010, 1020]
	duracao_auavMuitoMuitoAlta	[1010, 1040, 1450, 1450]
Duração [ianu]	duracao_ianuMuitoMuitoPequena	[0, 0, 130, 140]
	duracao_ianuMuitoPequena	[130, 150, 170, 180]
	duracao_ianuPequena	[170, 190, 210, 220]
	duracao_ianuMedia	[210, 230, 250, 260]
	duracao_ianuAlta	[250, 270, 290, 300]
	duracao_ianuMuitoAlta	[290, 310, 330, 340]
	duracao_ianuMuitoMuitoAlta	[330, 350, 450, 450]
Duração [az]	duracao_azMuitoMuitoPequena	[0, 0, 180, 200]
	duracao_azMuitoPequena	[190, 200, 210, 220]
	duracao_azPequena	[210, 220, 230, 240]
	duracao_azMedia	[230, 240, 250, 260]
	duracao_azAlta	[250, 260, 270, 280]
	duracao_azMuitoAlta	[270, 290, 305, 320]
	duracao_azMuitoMuitoAlta	[310, 320, 480, 480]
f_o mínimo	$f_o$minMuitoMuitoPequeno	[0, 0, 60, 70]
	$f_o$minMuitoPequeno	[55, 70, 90, 100]
	$f_o$minPequeno	[85, 100, 120, 130]
	$f_o$minMedio	[115, 130, 150, 160]
	$f_o$minAlto	[150, 158, 178, 186]
	$f_o$minMuitoAlto	[178, 186, 210, 225]
	$f_o$minMuitoMuitoAlto	[210, 225, 310, 310]
f_o máximo	$f_o$maxMuitoMuitoPequeno	[0, 0, 180, 210]
	$f_o$maxMuitoPequeno	[190, 200, 230, 250]
	$f_o$maxPequeno	[240, 250, 280, 290]
	$f_o$maxMedio	[280, 290, 320, 330]

(continuação)

f_o máximo	$f_o$maxAlto	[320, 330, 360, 370]
	$f_o$maxMuitoAlto	[360, 370, 415, 425]
	$f_o$maxMuitoMuitoAlto	[415, 425, 530, 530]
f_o médio	$f_o$mdMuitoMuitoPequeno	[0, 0, 90, 110]
	$f_o$mdMuitoPequeno	[100, 115, 140, 150]
	$f_o$mdPequeno	[140, 155, 175, 190]
	$f_o$mdMedio	[180, 195, 230, 245]
	$f_o$mdAlto	[235, 250, 265, 280]
	$f_o$mdMuitoAlto	[270, 285, 300, 315]
	$f_o$mdMuitoMuitoAlto	[305, 320, 400, 400]
f_o range	$f_o$rangeMuitoMuitoPequeno	[0, 0, 30, 50]
	$f_o$rangeMuitoPequeno	[30, 50, 70, 90]
	$f_o$rangePequeno	[70, 90, 110, 130]
	$f_o$rangeMedio	[110, 130, 155, 170]
	$f_o$rangeAlto	[150, 170, 190, 210]
	$f_o$rangeMuitoAlto	[190, 210, 235, 250]
	$f_o$rangeMuitoMuitoAlto	[230, 250, 370, 370]
f_o desvio padrão	$f_o$dpMuitoMuitoPequeno	[0, 0, 13, 15]
	$f_o$dpMuitoPequeno	[14, 16, 18, 20]
	$f_o$dpPequeno	[19, 20, 23.5, 25]
	$f_o$dpMedio	[24, 26, 28, 30]
	$f_o$dpAlto	[29, 31, 33, 35]
	$f_o$dpMuitoAlto	[34, 36, 38, 40]
	$f_o$dpMuitoMuitoAlto	[39, 41, 70, 70]
Intensidade	intensidadeMuitoMuitoPequena	[0, 0, 34, 38]
	intensidadeMuitoPequena	[35, 39, 40, 45]
	intensidadePequena	[42, 46, 47, 52]
	intensidadeMedia	[49, 53, 54, 59]
	intensidadeAlta	[57, 59, 62, 66]
	intensidadeMuitoAlta	[63, 68, 69, 73]

(continuação)

Intensidade	intensidadeMuitoMuitoAlta	[70, 75, 85, 85]
Jitter	jitterMuitoMuitoPequena	[0, 0, 11, 11,5]
	jitterMuitoPequena	[11, 12, 12,7, 13,2]
	jitterPequena	[12,5, 13,4, 14,2, 14,7]
	jitterMedia	[14, 14,9, 15,7, 16,2]
	jitterAlta	[15,5, 16,5, 17,2, 17,7]
	jitterMuitoAlta	[17, 18, 18,7, 19,2]
	jitterMuitoMuitoAlta	[18,5, 19,5, 40, 40]
Shimmer	shimmerMuitoMuitoPequena	[0, 0, 5, 6]
	shimmerMuitoPequena	[5, 6, 7, 8]
	shimmerPequena	[7, 8, 9, 10]
	shimmerMedia	[9, 10, 11, 12]
	shimmerAlta	[11, 12, 13, 14]
	shimmerMuitoAlta	[13, 14, 15, 16]
	shimmerMuitoMuitoAlta	[15, 16, 21, 21]
Hnr	hnrMuitoMuitoPequena	[0, 0, 9, 10]
	hnrMuitoPequena	[9, 10, 11, 12]
	hnrPequena	[11, 11,8, 13, 14]
	hnrMedia	[13, 14, 15, 16]
	hnrAlta	[15, 16, 17, 18]
	hnrMuitoAlta	[17, 18, 19, 20]
	hnrMuitoMuitoAlta	[19, 20, 23, 23]
Cpps	cppsMuitoMuitoPequena	[0, 0, 7,5, 7,8]
	cppsMuitoPequena	[7,7, 8, 8,3, 8,5]
	cppsPequena	[8,4, 8,7, 9, 9,3]
	cppsMedia	[9,2, 9,45, 9,8, 10,1]
	cppsAlta	[10, 10,3, 10,6, 10,9]
	cppsMuitoAlta	[10,8, 11,1, 11,4, 11,7]
	cppsMuitoMuitoAlta	[11,6, 11,9, 15, 15]
Cpps desvio padrão	cppsdpMuitoMuitoPequeno	[0, 0, 2,5, 2,7]

(conclusão)

Cpps desvio padrão	cppsdpMuitoPequeno	[2,6, 2,8, 2,9, 3,2]
	cppsdpPequeno	[3,1, 3,2, 3,3, 3,4]
	cppsdpMedio	[3,3, 3,4, 3,5, 3,6]
	cppsdpAlto	[3,5, 3,7, 3,9, 4]
	cppsdpMuitoAlto	[3,9, 4, 4,4, 4,5]
	cppsdpMuitoMuitoAlto	[4,4, 4,6, 5,3, 5,3]
Valência	valenciaPositiva	[0, 1, 2]
	valenciaNegativa	[1, 2, 3]
	valenciaNeutra	[2, 3, 4]
Potência	potenciaForte	[0, 1, 2]
	potenciaFraca	[1, 2, 3]
Ativação	ativacaoAlta	[0, 1, 2]
	ativacaoBaixa	[1, 2, 3]
Variável de saída		
Emoções	Alegria	[0,5, 1, 1,5]
	Medo	[1,5, 2, 2,5]
	Tristeza	[2,5, 3, 3,5]
	Raiva	[3,5, 4, 4,5]
	Surpresa	[4,5, 5, 5,5]
	Nojo	[5,5, 6, 6,5]
	Neutra	[6,5, 7, 7,5]

A seguir foram apresentados os gráficos das funções de pertinência correspondentes a cada variável utilizada no modelo. As cores nos gráficos foram escolhidas aleatoriamente, e cada cor corresponde a um termo linguístico relacionado às variáveis de entrada e saída. Observa-se como os graus de pertinência são atribuídos a diferentes faixas de valores. Os parâmetros acústicos foram definidos com funções trapezoidais devido à necessidade de representar uma amplitude maior de valores com máxima pertinência. Este formato permite acomodar termos linguísticos de maneira abrangente, apresentam um pico plano que indica a máxima pertinência (grau de pertinência = 1) e bases inclinadas que proporcionam uma transição gradual entre diferentes termos. Essa estrutura possibilita que valores intermediários compartilhem

pertinência com mais de um termo para refletir a suavidade e a incerteza inerentes a lógica *fuzzy*. Por exemplo, um valor intermediário pode pertencer simultaneamente a dois termos linguísticos, como "médio" e "alto", com graus de pertinência distintos. Por outro lado, os eixos de dimensão emocional foram representados por funções triangulares devido à sua simplicidade e capacidade de modelar categorias bem definidas. Cada triângulo possui um único pico, que representa o valor central mais característico de um termo linguístico, como "positivo", "negativo" ou "neutro". As bases inclinadas dessas funções permitem uma transição gradual entre termos adjacentes, assegurando a continuidade entre os estados emocionais.

As funções de pertinência para a variável de saída (emoções) foram representadas no formato triangular, devido à simplicidade para o processo de defuzzificação e à capacidade de representar emoções com um pico central que define o valor mais representativo para cada termo linguístico. As bases do triângulo permitem uma gradação contínua entre as emoções adjacentes, conferindo ao modelo uma transição suave entre diferentes estados emocionais.

Os eixos dos gráficos representam, no eixo X, os valores reais das variáveis dentro de seus respectivos universos de discurso, e, no eixo Y, os graus de pertinência que variam de 0 a 1. As interseções entre as funções de pertinência são importantes para o funcionamento do sistema, pois permitem que um mesmo valor numérico pertença parcialmente a mais de um termo linguístico, influenciando diretamente a construção das regras *fuzzy*. Essas interseções refletem a capacidade do modelo de lidar com informações graduais, possibilitando inferências mais realistas.

Dessa forma, os gráficos apresentados a seguir gerados no pacote *FuzzyRules*, oferecem uma representação clara da categorização das variáveis de entrada e saída no modelo *fuzzy*, evidenciando as relações entre os termos linguísticos, suas transições e os valores centrais. Essas ilustrações, delimitadas por um recorte do universo de discurso, têm como propósito demonstrar de forma visual a atribuição dos graus de pertinência a diferentes faixas de valores, para uma análise mais intuitiva da classificação das emoções (Figura 9–27).

Figura 9 - Função de pertinência da variável de entrada Duração [al]

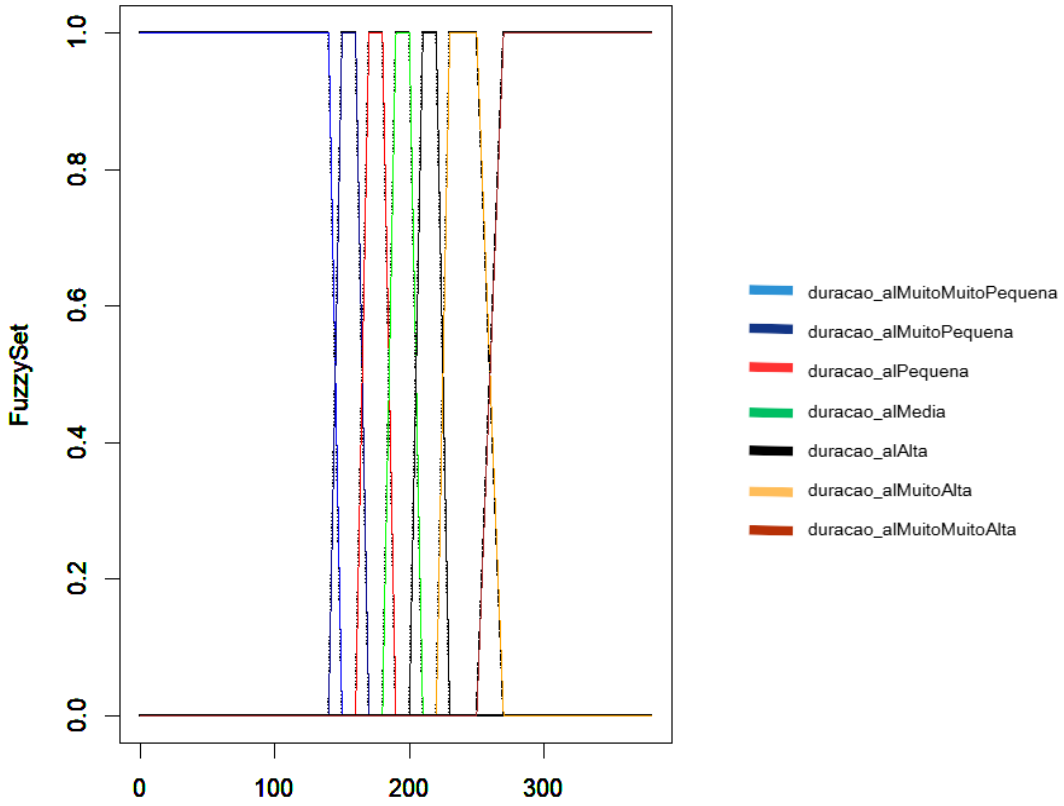


Figura 10 - Função de pertinência da variável de entrada Duração [auav]

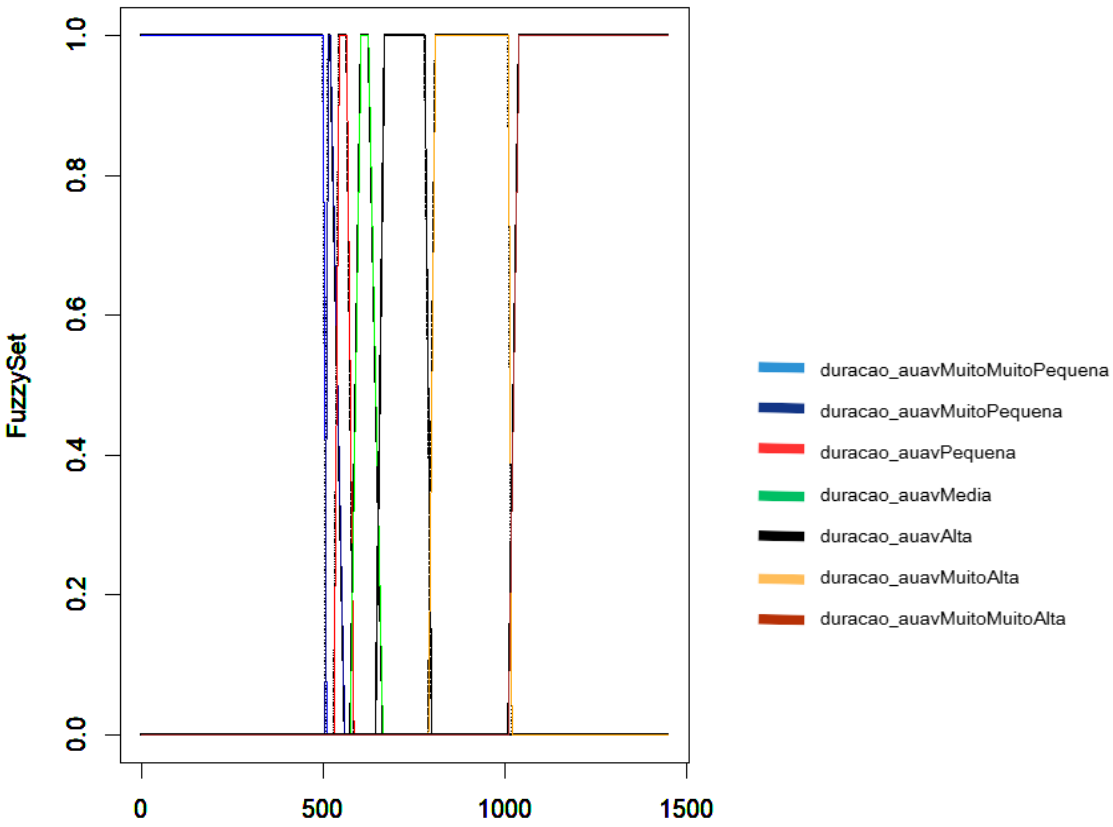


Figura 11 - Função de pertinência da variável de entrada Duração [ianu]

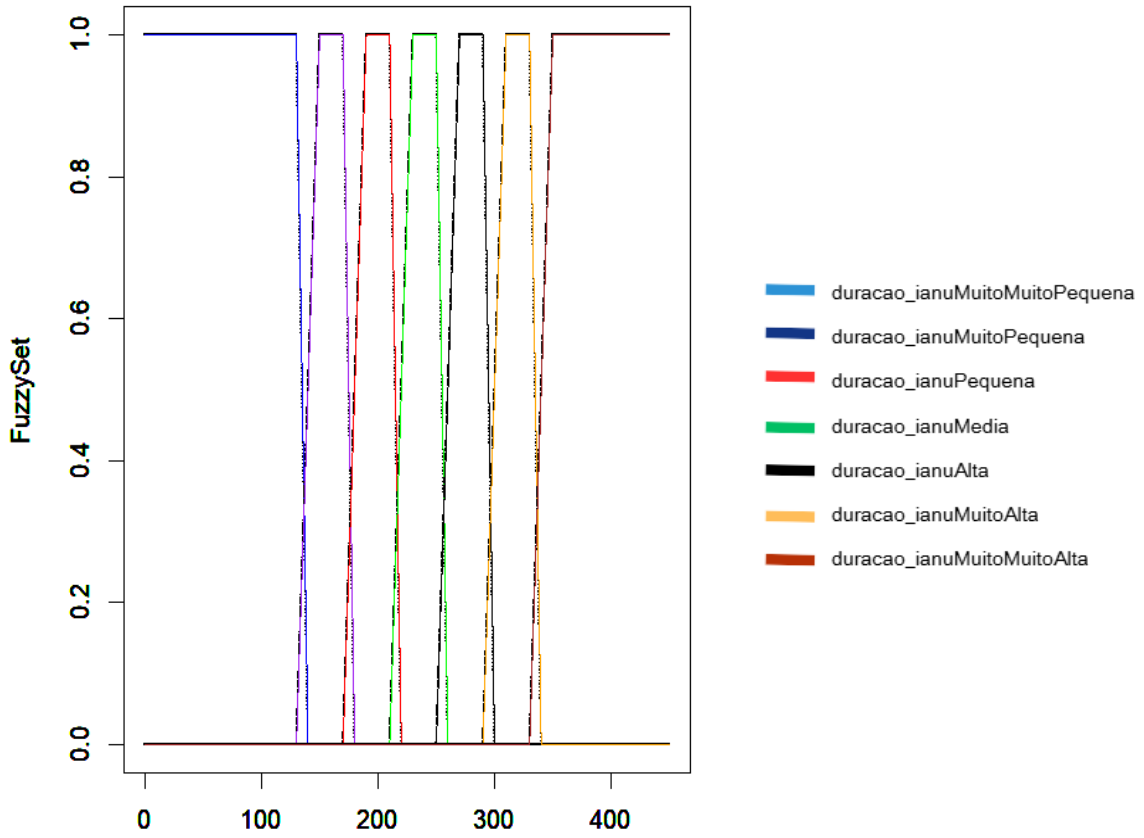


Figura 12 - Função de pertinência da variável de entrada Duração [az]

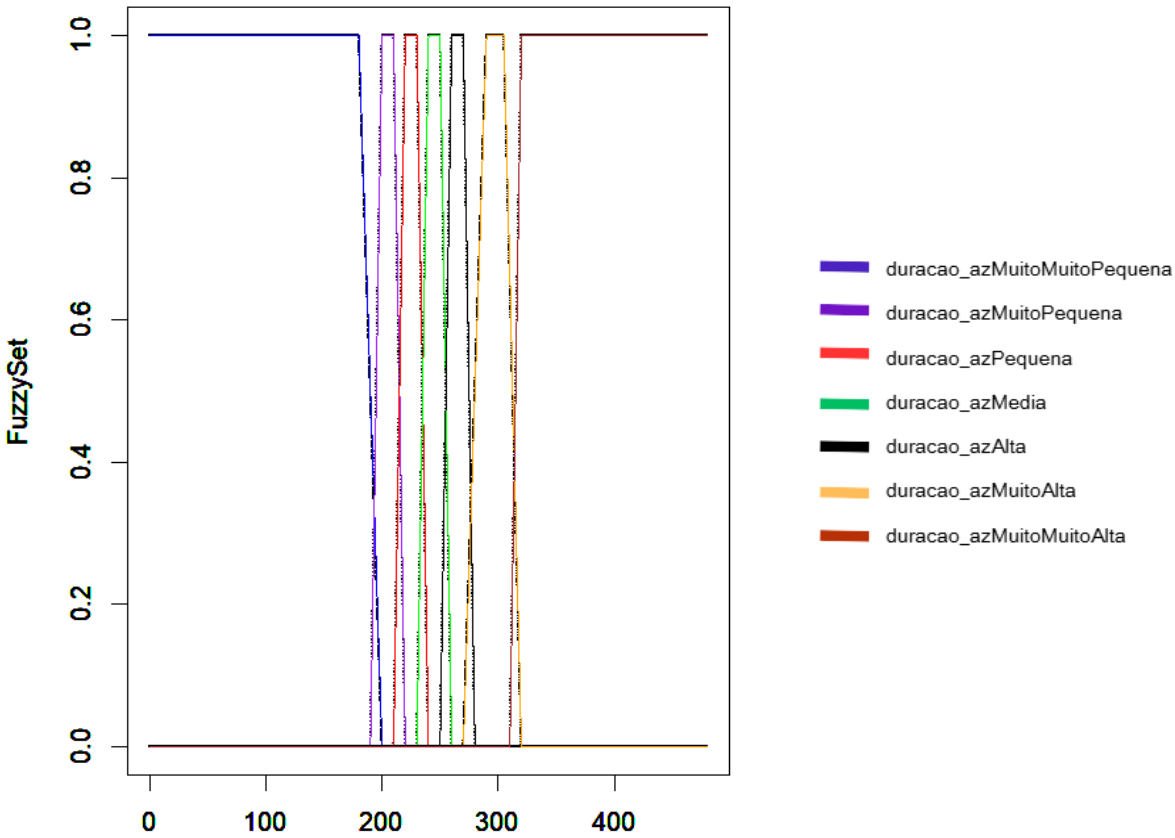


Figura 13 - Função de pertinência da variável de entrada f_o mínimo

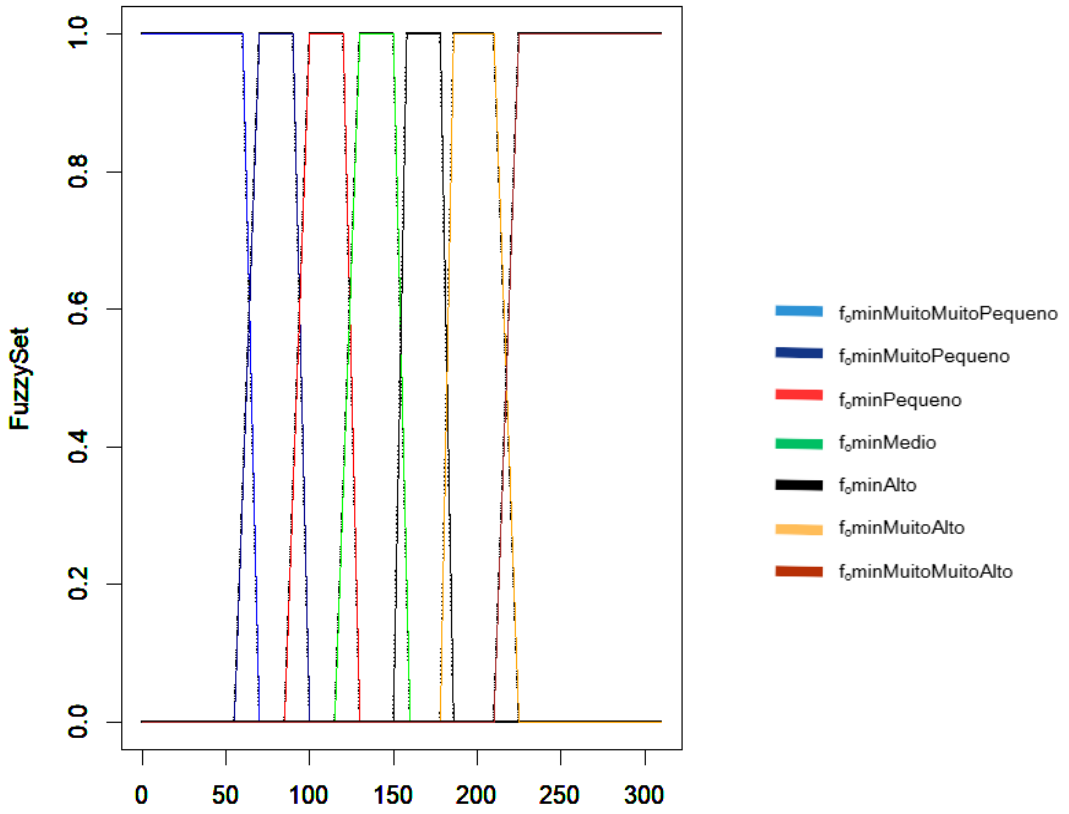


Figura 14 - Função de pertinência da variável de entrada f_o máximo

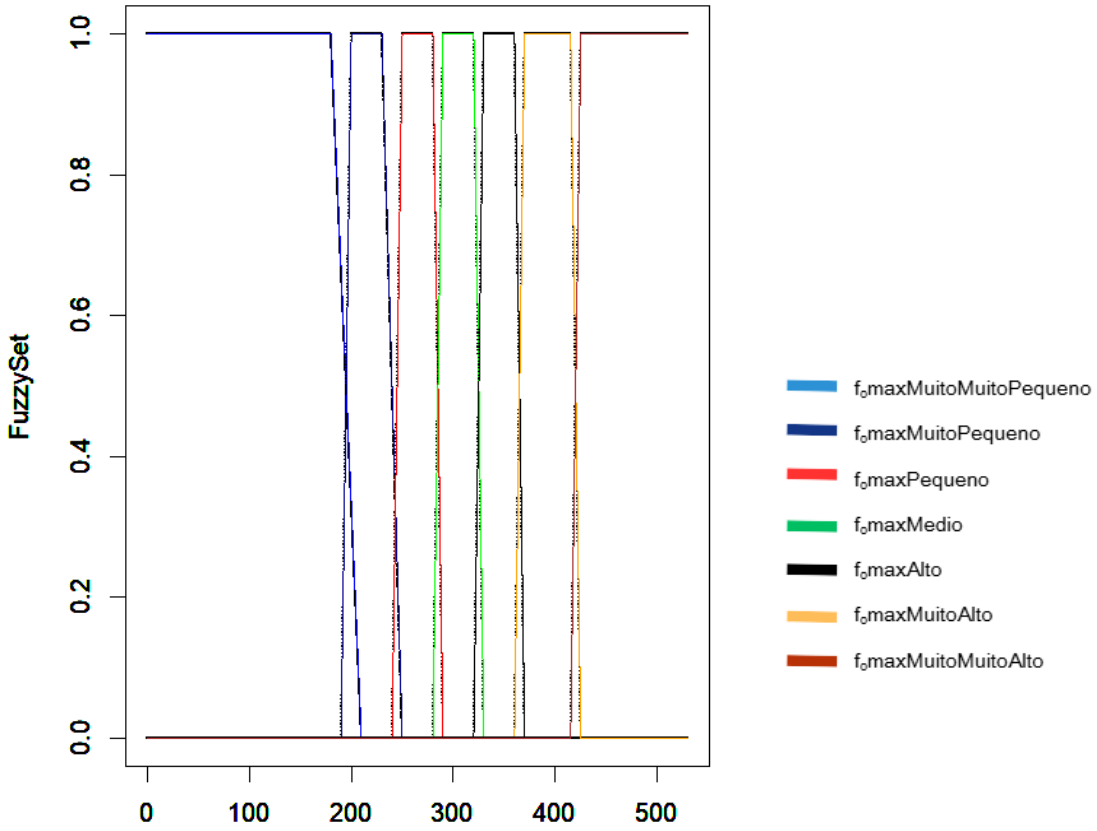


Figura 15 - Função de pertinência da variável de entrada f_o médio

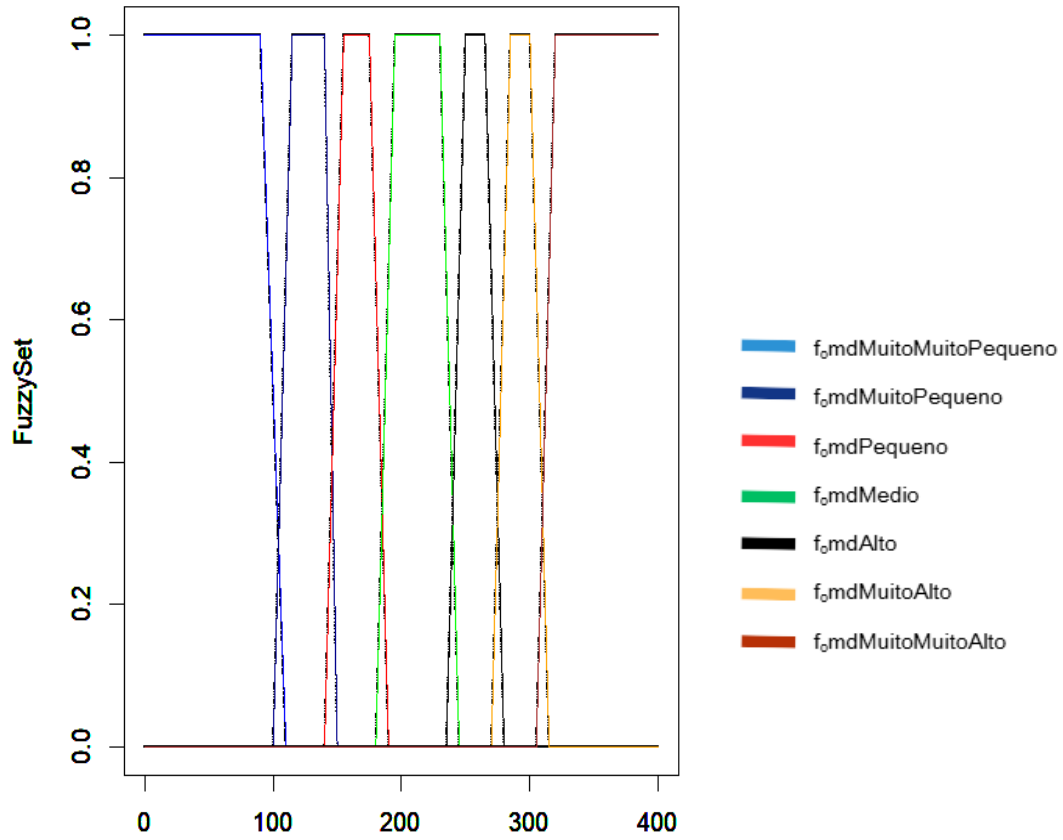


Figura 16 - Função de pertinência da variável de entrada f_o range

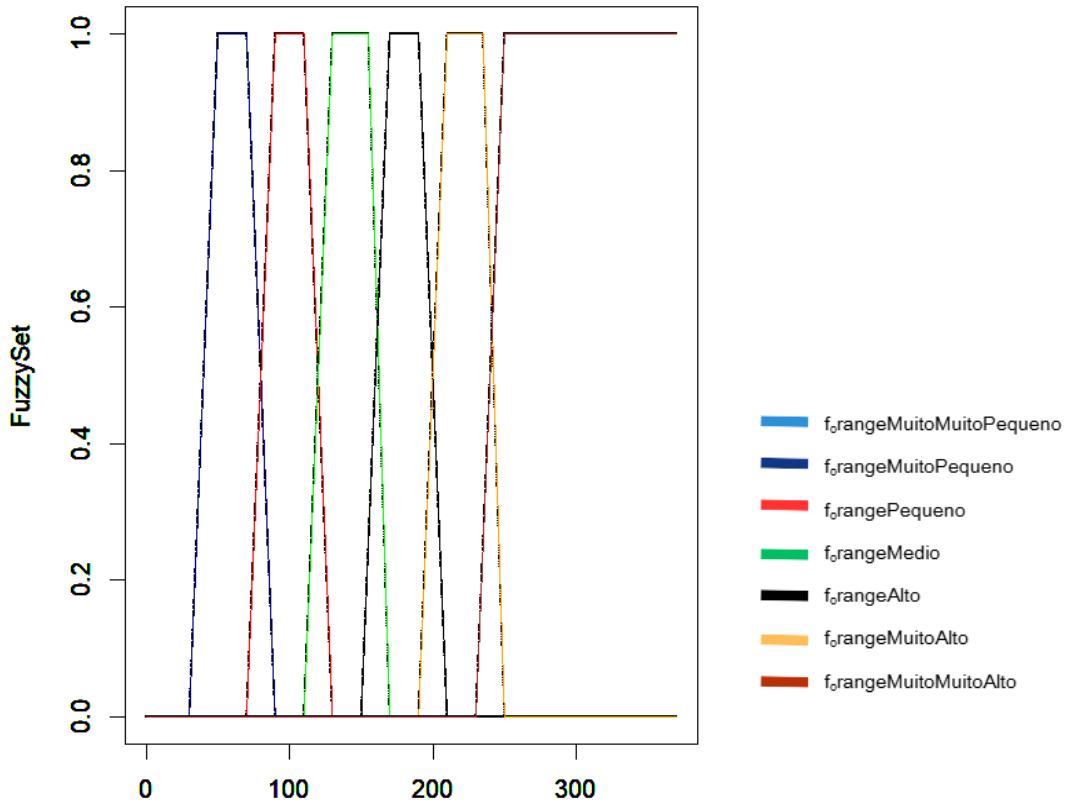


Figura 17 - Função de pertinência da variável de entrada f_o desvio padrão

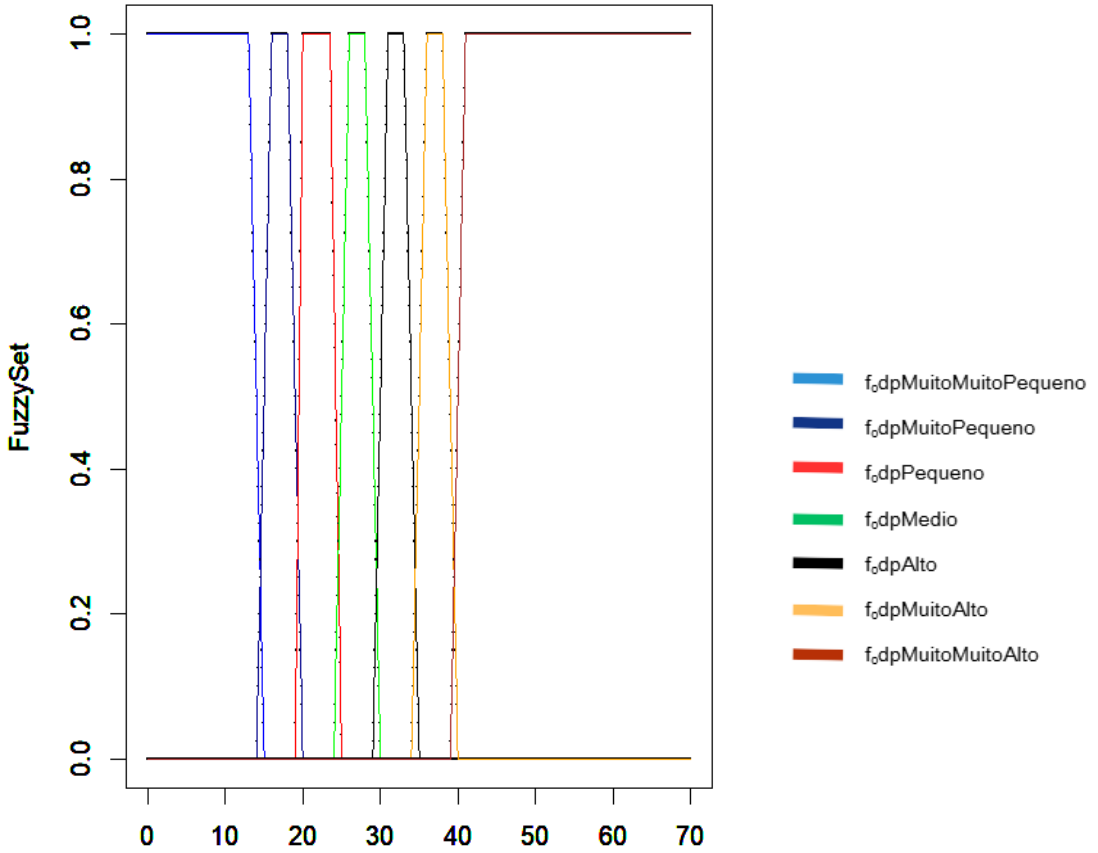


Figura 18 - Função de pertinência da variável de entrada Intensidade

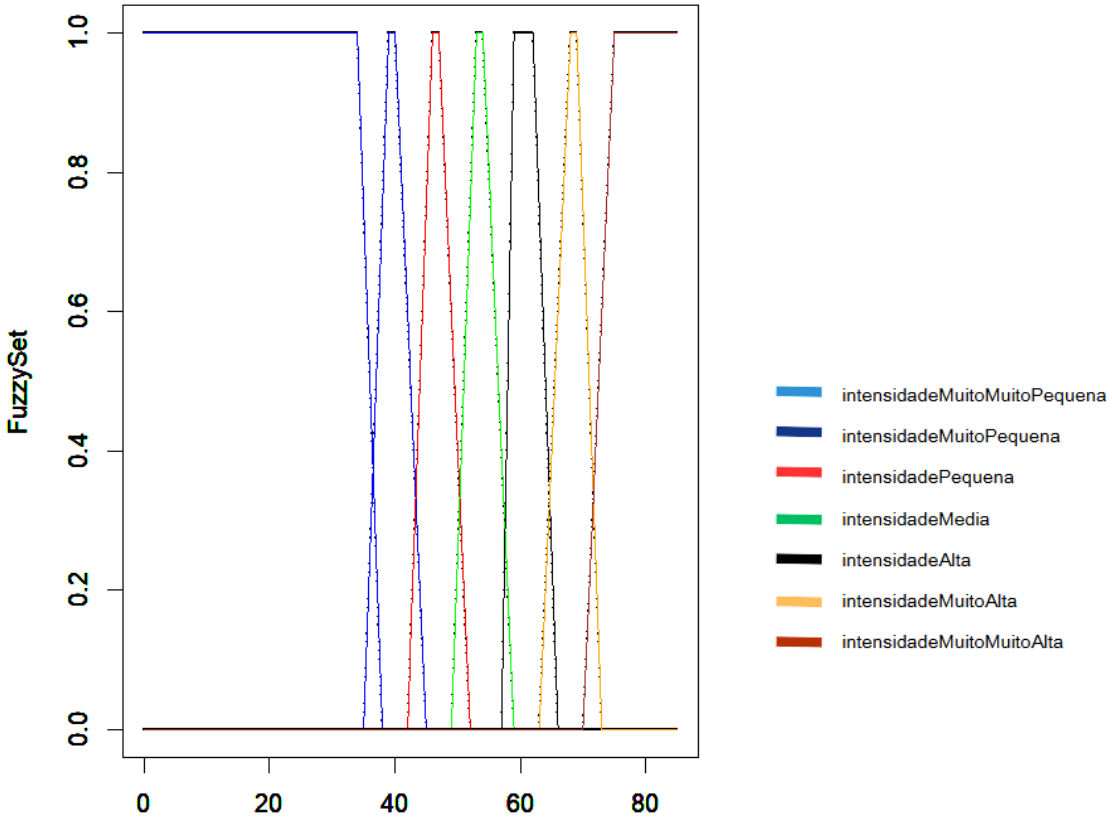


Figura 19 - Função de pertinência da variável de entrada Jitter

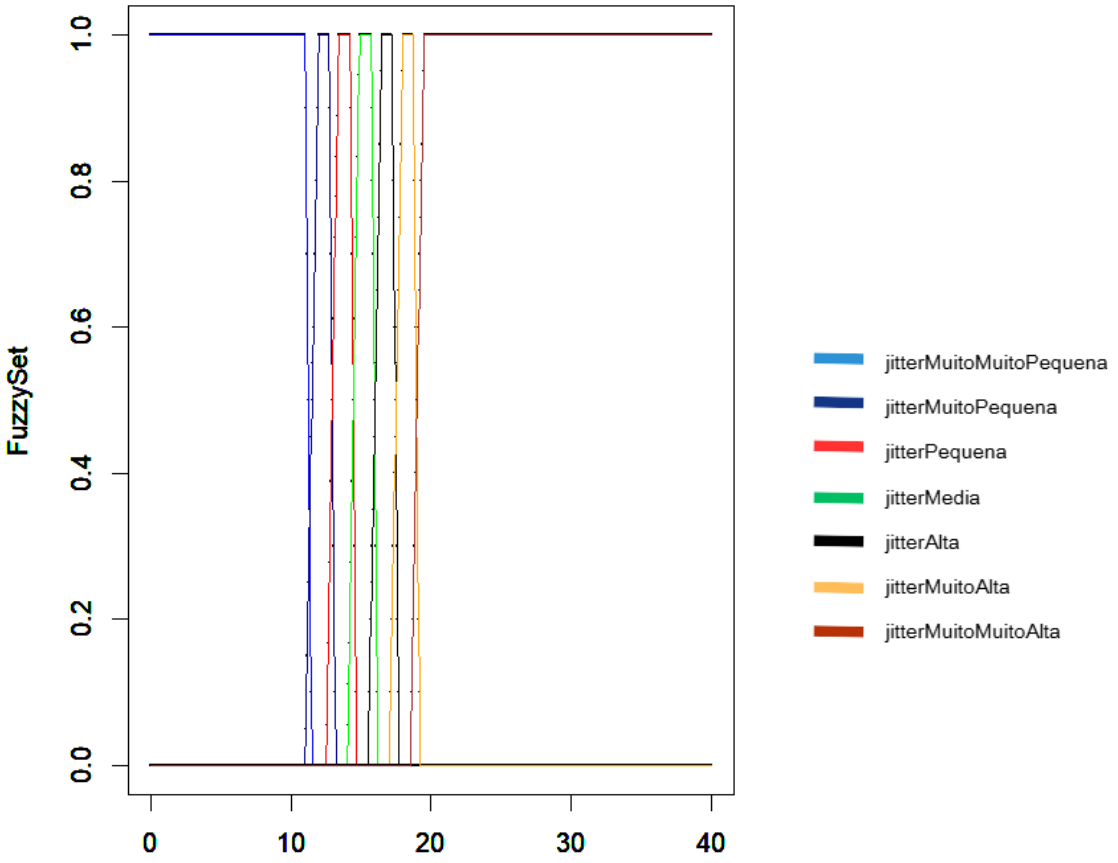


Figura 20 - Função de pertinência da variável de entrada Shimmer

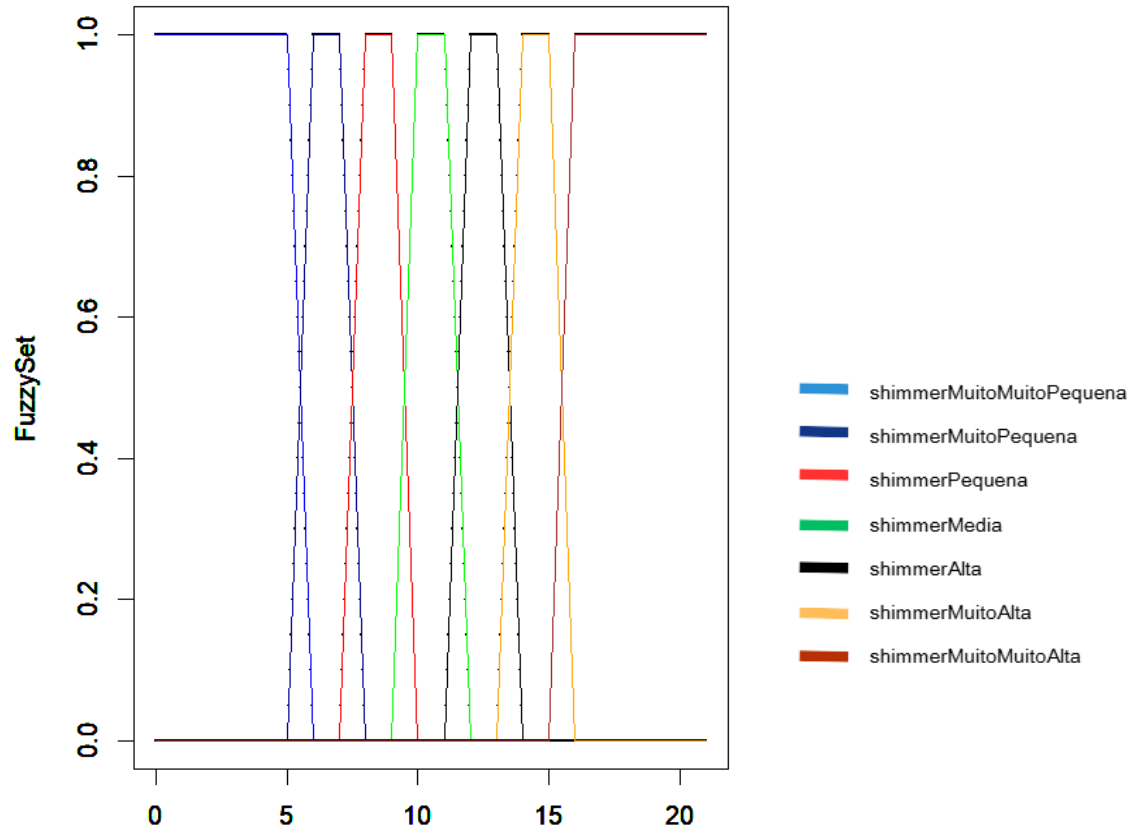


Figura 21 - Função de pertinência da variável de entrada HNR

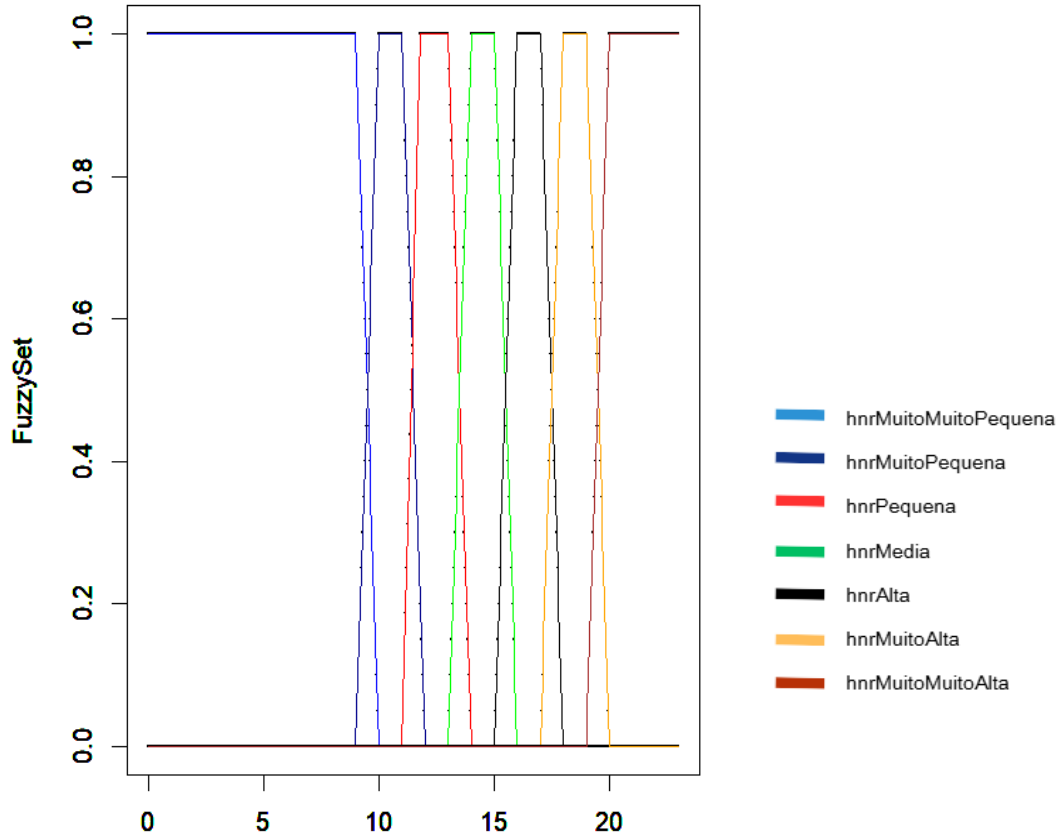


Figura 22 - Função de pertinência da variável de entrada CPPS

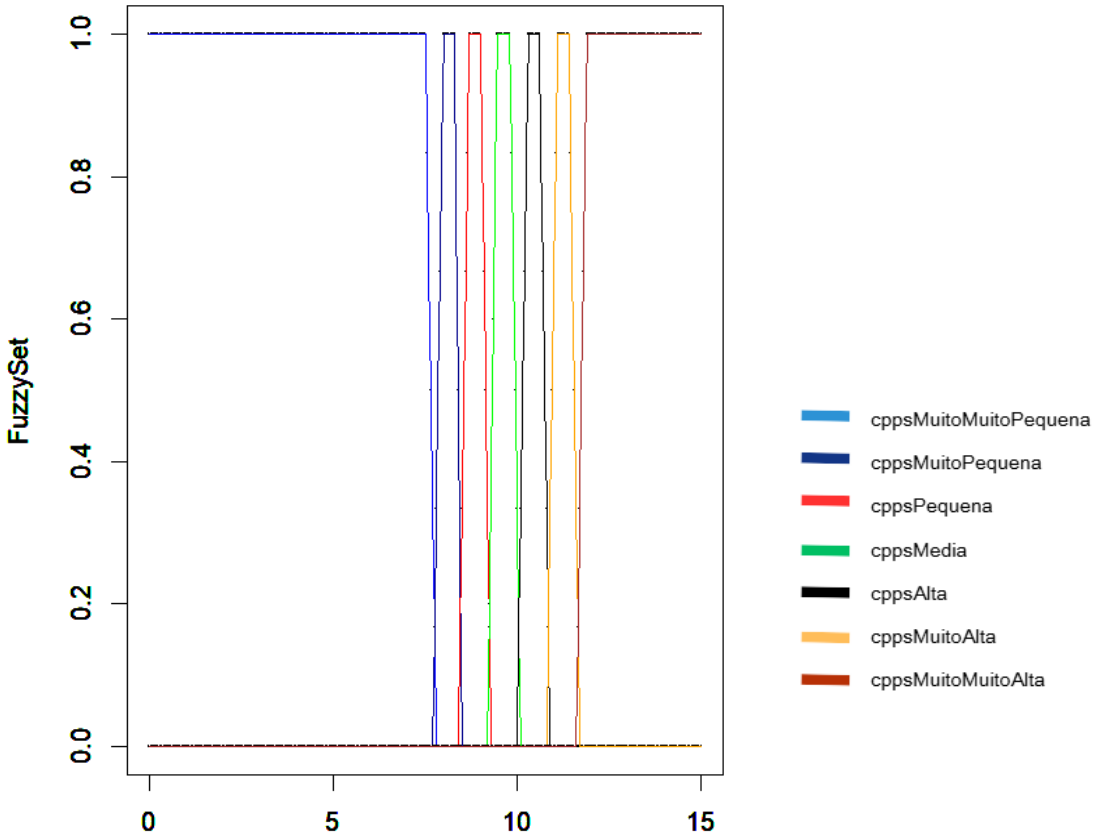


Figura 23- Função de pertinência da variável de entrada CPPS desvio padrão

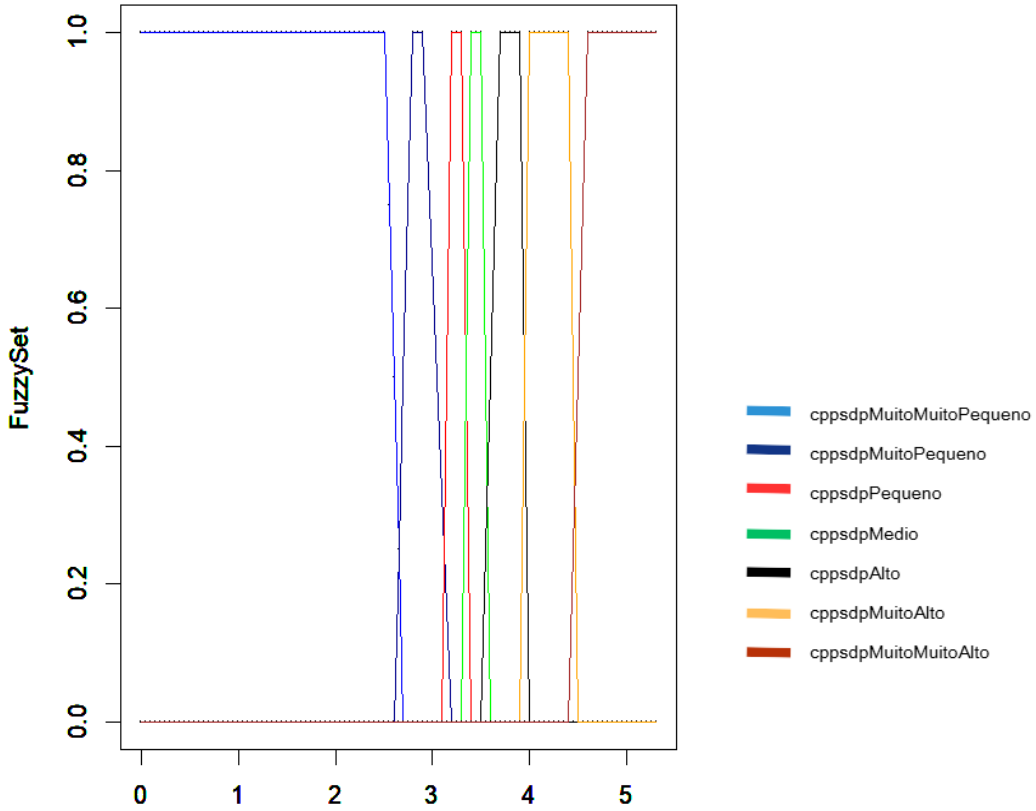


Figura 24 - Função de pertinência da variável de entrada Valência

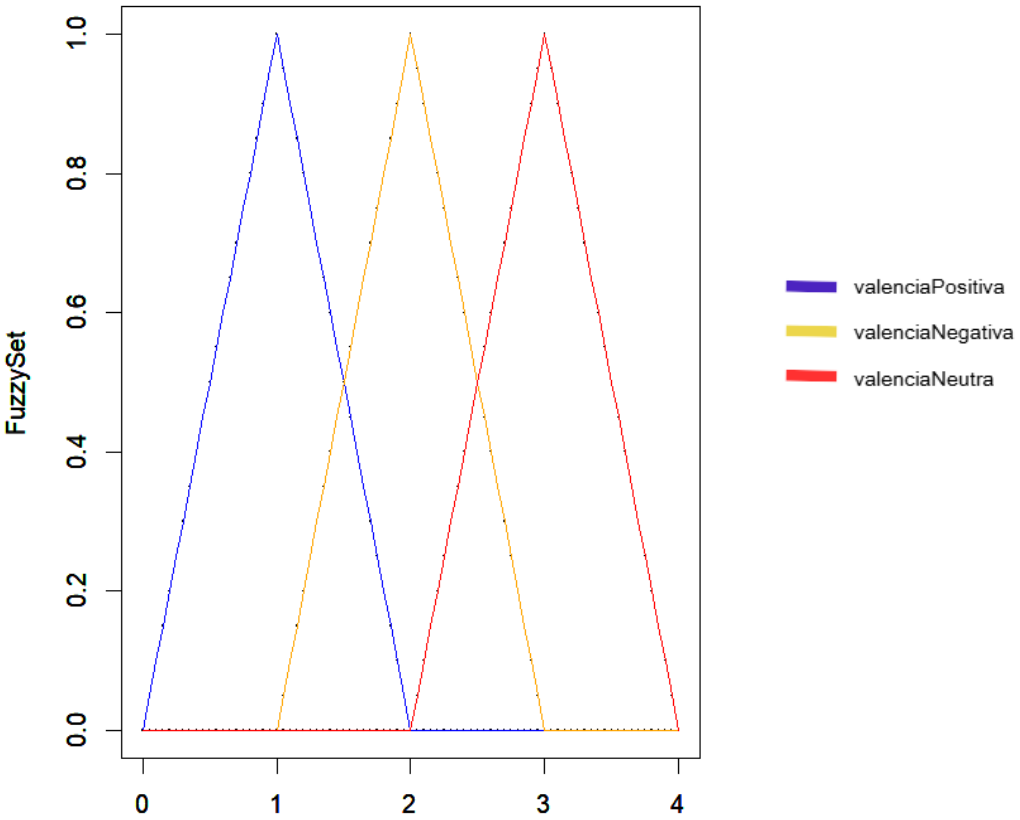


Figura 25 - Função de pertinência da variável de entrada Potência

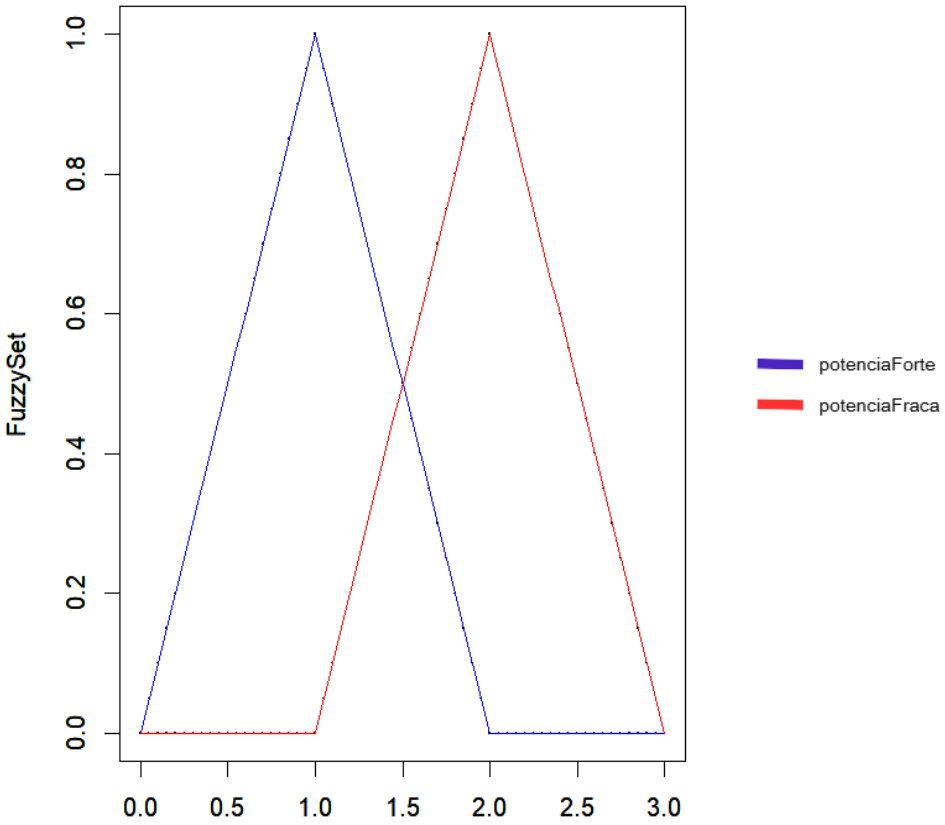


Figura 26 - Função de pertinência da variável de entrada Ativação

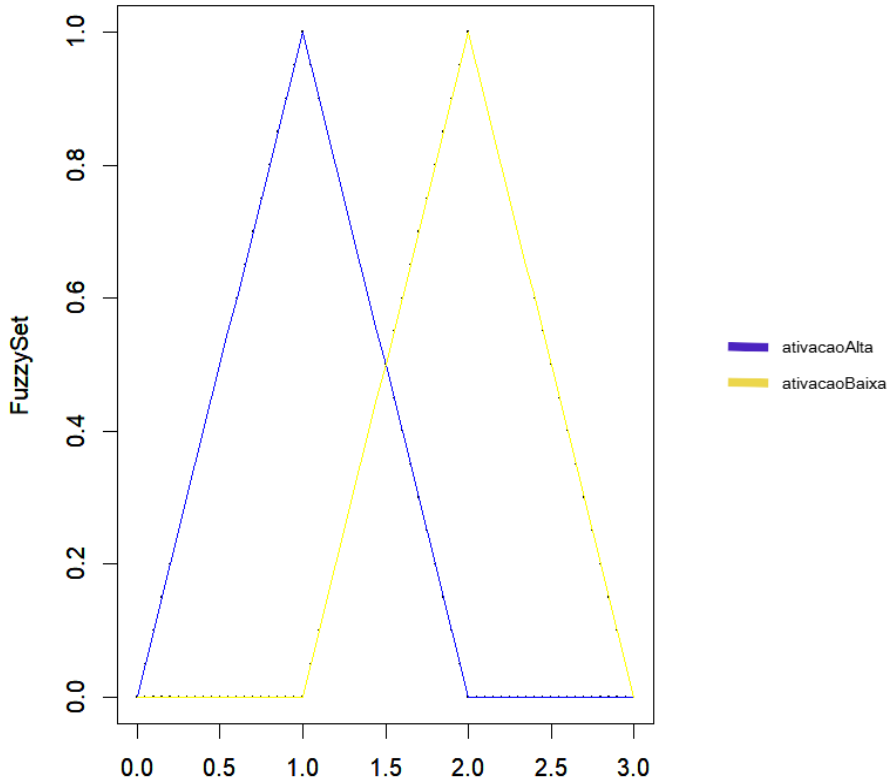
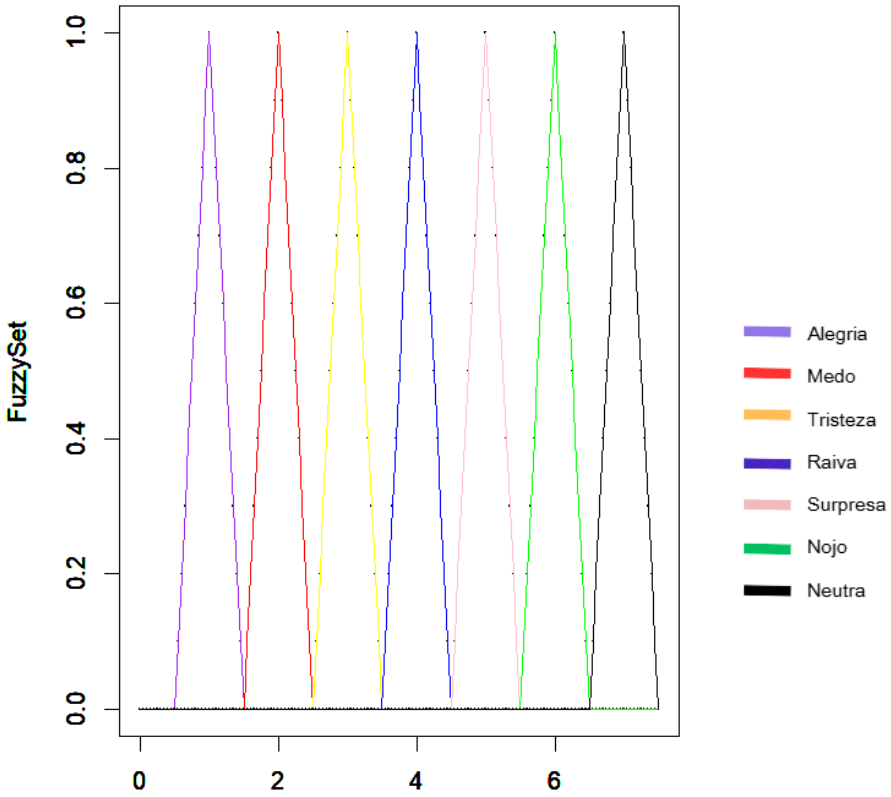


Figura 27 - Função de pertinência da variável de saída Emoções



5.1.1 Construção das regras *fuzzy*

Foram desenvolvidos bancos de regras *fuzzy* para modelar o reconhecimento das emoções com precisão, com a integração de 18 variáveis linguísticas de entrada, representativas de parâmetros acústicos e prosódicos, tais como f_0, intensidade, duração, medidas cepstrais, além da valência, potência e ativação da emoção, e 7 variáveis linguísticas de saída, que incluem as emoções básicas e a emissão neutra. Com base nesses parâmetros, 23 regras foram formuladas e implementadas no sistema *fuzzy*, a fim de garantir uma inferência robusta que permite diferenciar emoções de forma adaptativa e sensível às variações sutis dos sinais de voz.

Cada regra estabelece condições específicas que acionam classificações precisas no sistema, promovendo uma interpretação ajustada às variações vocais associadas a cada estado emocional. A seguir, exemplificam-se as regras aplicadas nos bancos, ilustrando como as variáveis de entrada interagem para gerar as inferências de saída conforme as características emocionais detectadas.

Regras emoção Alegria

SE (valencia é valenciaPositiva) E (duracao_az é duracao_azMuitoMuitoPequena OU duracao_az é duracao_azMuitoPequena OU duracao_az é duracao_azPequena) E (cpps é cppsPequena OU cpps é cppsMedia OU cpps é cppsAlta OU cpps é cppsMuitoAlta OU cpps é cppsMuitoMuitoAlta) ENTÃO (Emoção é Alegria)

SE (valencia é valenciaPositiva) E (f0min é f0min é F0minAlto OU f0min é F0minMuitoAlto OU f0min é F0minMuitoMuitoAlto) E (hnr é hnrMuitoMuitoPequena OU (hnr é hnrMuitoPequena OU hnr é hnrPequena OU hnr é hnrMedia OU hnr é hnrAlta OU hnr é hnrMuitoAlta) E (duracao_ianu é duracao_ianuPequena OU duracao_ianu é duracao_ianuMedia OU duracao_ianu é duracao_ianuAlta OU duracao_ianu é duracao_ianuMuitoAlta OU duracao_ianu é duracao_ianuMuitoMuitoAlta) ENTÃO (Emoção é Alegria)

Regras emoção Medo

SE (valencia é valenciaNegativa) E (f0md é F0mdMedio OU f0md é F0mdAlto OU f0md é F0mdMuitoAlto OU f0md é F0mdMuitoMuitoAlto) E (hnr é hnrMuitoMuitoPequena OU hnr é hnrMuitoPequena OU hnr é hnrPequena OU hnr é hnrMedia OU hnr é hnrAlta OU hnr é hnrMuitoAlta) E (cppsdp é cppsdpMedia OU cppsdp é cppsdpAlta OU cppsdp é cppsdpMuitoAlta OU cppsdp é cppsdpMuitoMuitoAlta) E (duracao_az é duracao_azMuitoMuitoPequena OU duracao_az é duracao_azMuitoPequena OU duracao_az é duracao_azPequena) E (potencia é potenciaForte) ENTÃO (Emoção é Medo)

SE (valencia é valenciaNegativa) E (cpps é cppsPequena OU cpps é cppsMedia OU cpps é cppsAlta OU cpps é cppsMuitoAlta OU cpps é cppsMuitoMuitoAlta) E (hnr é hnrAlta OU hnr é hnrMuitoAlta OU hnr é hnrMuitoMuitoAlta) E (duracao_al é duracao_alMuitoMuitoPequena OU duracao_al é duracao_alMuitoPequena OU duracao_al é duracao_alPequena) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Medo)

SE (valencia é valenciaNegativa) E (cpps é cppsMuitoMuitoPequena) E (potencia é potenciaForte) E (ativacao é ativacaoAlta) E (duracao_az é

duracao_azMuitoMuitoPequena OU duracao_az é duracao_azMuitoPequena OU duracao_az é duracao_azPequena) ENTÃO (Emoção é Medo)

SE (valencia é valenciaNegativa) E (f0min é F0minMuitoMuitoAlto) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Medo)

SE (valencia é valenciaNegativa) E (f0min é F0minPequeno OU f0min é F0minMedio OU f0min é F0minAlto) E (f0md é F0mdMuitoPequeno OU f0md é F0mdPequeno OU f0md é F0mdMedio) E (hnr é hnrMuitoPequena OU hnr é hnrPequena OU hnr é hnrMedia) E (cpps é cppsMuitoMuitoPequena OU cpps é cppsMuitoPequena) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Medo)

Regras emoção Tristeza

SE (valencia é valenciaNegativa) E (potencia é potenciaFraca) E (ativacao é ativacaoBaixa) ENTÃO (Emoção é Tristeza)

Regras emoção Raiva

SE (valencia é valenciaNegativa) E (f0md é F0mdMedio OU f0md é F0mdAlto OU f0md é F0mdMuitoAlto OU f0md é F0mdMuitoMuitoAlto) E (cppsdp é cppsdpMuitoAlta OU cppsdp é cppsdpMuitoMuitoAlta) E (hnr é hnrMuitoMuitoPequena OU hnr é hnrMuitoPequena OU hnr é hnrPequena OU hnr é hnrMedia OU hnr é hnrAlta OU hnr é hnrMuitoAlta) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Raiva)

SE (valencia é valenciaNegativa) E (f0min é F0minMuitoMuitoPequeno OU f0min é F0minMuitoPequeno OU f0min é F0minPequeno OU f0min é F0minMedio OU f0min é F0minAlto OU f0min é F0minMuitoAlto) E (f0max é F0maxMedio OU f0max é F0maxAlto OU f0max é F0maxMuitoAlto OU f0max é F0maxMuitoMuitoAlto) E (duracao_auav é duracao_auavMuitoMuitoPequena OU duracao_auavMuitoPequena OU duracao_auavPequena OU duracao_auavMedia OU duracao_auavAlta OU duracao_auavMuitoAlta) E (intensidade é intensidadeAlta OU intensidade é intensidadeMuitoAlta OU intensidade é intensidadeMuitoMuitoAlta) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Raiva)

SE (valencia é valenciaNegativa) E (duracao_al é duracao_alMuitoMuitoPequena OU duracao_al é duracao_alMuitoPequena) E (duracao_ianu é duracao_ianuMuitoMuitoPequena OU duracao_ianu é duracao_ianuMuitoPequena OU duracao_ianu é duracao_ianuPequena) E (cppsdp é cppsdpMuitoMuitoPequena OU cppsdp é cppsdpMuitoPequena OU cppsdp é cppsdpPequena OU cppsdp é cppsdpMedia OU cppsdp é cppsdpAlta OU cppsdp é cppsdpMuitoAlta) E (f0min é F0minMuitoMuitoPequeno OU f0min é F0minMuitoPequeno OU f0min é F0minPequeno OU f0min é F0minMedio OU f0min é F0minAlto OU f0min é F0minMuitoAlto) E (hnr é hnrMuitoMuitoPequena OU hnr é hnrMuitoPequena OU hnr é hnrPequena) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Raiva)

SE (valencia é valenciaNegativa) E (cppsdp é cppsdpPequena OU cppsdp é cppsdpMedia OU cppsdp é cppsdpAlta OU cppsdp é cppsdpMuitoAlta OU cppsdp é cppsdpMuitoMuitoAlta) E (f0range é f0rangeMuitoPequeno OU f0range é f0rangePequeno OU f0range é f0rangeMedio OU f0range é f0rangeAlto OU f0range é f0rangeMuitoAlto OU f0range é f0rangeMuitoMuitoAlto) E (hnr é hnrMuitoMuitoPequena OU hnr é hnrMuitoPequena) E (duracao_auav é duracao_auavMuitoMuitoPequena OU duracao_auav é duracao_auavMuitoPequena OU duracao_auav é duracao_auavPequena OU duracao_auav é duracao_auavMedia OU duracao_auav é duracao_auavAlta OU duracao_auav é duracao_auavMuitoAlta) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Raiva)

SE (valencia é valenciaNegativa) E (hnr é hnrMuitoMuitoPequena OU hnrMuitoPequena) E (f0max é F0maxMuitoMuitoPequeno OU f0max é F0maxMuitoPequeno OU f0max é F0maxPequeno OU f0max é F0maxMedio OU f0max é F0maxAlto OU f0max é F0maxMuitoAlto) E (ativacao é ativacaoAlta) ENTÃO (Emoção é Raiva)

Regras emoção Surpresa

SE (valencia é valenciaPositiva) E (cpps é cppsMuitoMuitoPequena OU cpps é cppsMuitoPequena OU cpps é cppsPequena) E (cppsdp é cppsdpMuitoMuitoPequena OU cppsdp é cppsdpMuitoPequena OU cppsdp é cppsdpPequena OU cppsdp é cppsdpMedia OU cppsdp é cppsdpAlta OU cppsdp é cppsdpMuitoAlta) ENTÃO (Emoção é Surpresa)

SE (valencia é valenciaPositiva) E (f0md é F0mdAlto OU f0md é F0mdMuitoAlto OU f0md é F0mdMuitoMuitoAlto) E (duracao_auav é duracao_auavMuitoMuitoPequena OU duracao_auav é duracao_auavMuitoPequena OU duracao_auav é duracao_auavPequena) E (cpps é cppsMuitoMuitoPequena OU cpps é cppsMuitoPequena OU cpps é cppsPequena) E (intensidade é intensidadeMuitoMuitoPequena OU intensidadeMuitoPequena OU intensidadePequena OU intensidadeMedia OU intensidadeAlta OU intensidadeMuitoAlta) ENTÃO (Emoção é Surpresa)

SE (valencia é valenciaPositiva) E (f0md é F0mdMedio OU f0md é F0mdAlto OU f0md é F0mdMuitoAlto OU f0md é F0mdMuitoMuitoAlto) E (cppsdp é cppsdpMuitoPequena OU cppsdp é cppsdpPequena OU cppsdp é cppsdpMedia OU cppsdp é cppsdpAlta) E (cpps é cppsMedia OU cpps é cppsAlta OU cpps é cppsMuitoAlta OU cpps é cppsMuitoMuitoAlta) E (hnr é hnrMedia OU hnr é hnrAlta OU hnr é hnrMuitoAlta OU hnr é hnrMuitoMuitoAlta) E (duracao_auav é duracao_auavMuitoMuitoPequena OU duracao_auav é duracao_auavMuitoPequena OU duracao_auav é duracao_auavPequena OU duracao_auav é duracao_auavMedia) ENTÃO (Emoção é Surpresa)

SE (valencia é valenciaPositiva) E (f0md é F0mdMuitoMuitoAlto) E (cppsdp é cppsdpMuitoPequena OU cppsdp é cppsdpPequena OU cppsdp é cppsdpMedia OU cppsdp é cppsdpAlta OU cppsdp é cppsdpMuitoAlta) E (cpps é cppsMedia OU cpps é cppsAlta OU cpps é cppsMuitoAlta) E (hnr é hnrMedia OU hnr é hnrAlta OU hnr é hnrMuitoAlta OU hnr é hnrMuitoMuitoAlta) E (f0min é F0minMuitoMuitoAlto) ENTÃO (Emoção é Surpresa)

Regras emoção Nojo

SE (valencia é valenciaNegativa) E (duracao_az é duracao_azMuitoAlta OU duracao_az é duracao_azMuitoMuitoAlta) E (intensidade é intensidadeMuitoMuitoPequena OU intensidade é intensidadeMuitoPequena OU intensidade é intensidadePequena OU intensidade é intensidadeMedia OU intensidade é intensidadeAlta OU intensidade é intensidadeMuitoAlta) E (ativacao é ativacaoBaixa) ENTÃO (Emoção é Nojo)

SE (valencia é valenciaNegativa) E (duracao_al é duracao_alMedia OU duracao_al é duracao_alAlta OU duracao_al é duracao_alMuitoAlta OU duracao_al é duracao_alMuitoMuitoAlta) E (duracao_ianu é duracao_ianuMuitoMuitoPequena OU duracao_ianu é duracao_ianuMuitoPequena OU duracao_ianu é duracao_ianuPequena) E (intensidade é intensidadeMuitoMuitoPequena OU intensidade é intensidadeMuitoPequena OU intensidade é intensidadePequena OU intensidade é intensidadeMedia OU intensidade é intensidadeAlta) E (ativacao é ativacaoBaixa) ENTÃO (Emoção é Nojo)

SE (valencia é valenciaNegativa) E (duracao_az é duracao_azMedia OU duracao_az é duracao_azAlta OU duracao_az é duracao_azMuitoAlta OU duracao_az é duracao_azMuitoMuitoAlta) E (f0dp é f0dpMuitoMuitoPequeno OU f0dp é f0dpMuitoPequeno OU f0dp é f0dpPequeno) E (ativacao é ativacaoBaixa) E (potencia é potenciaForte) ENTÃO (Emoção é Nojo)

SE (valencia é valenciaNegativa) E (f0dp é f0dpMedio OU f0dp é f0dpAlto OU f0dp é f0dpMuitoAlto OU f0dp é f0dpMuitoMuitoAlto) E (jitter é jitterMuitoMuitoPequena OU jitter é jitterMuitoPequena OU jitter é jitterPequena) E (ativacao é ativacaoBaixa) ENTÃO (Emoção é Nojo)

SE (valencia é valenciaNegativa) E (cpps é cppsMuitoMuitoPequena) E (potencia é potenciaForte) E (ativacao é ativacaoBaixa) ENTÃO (Emoção é Nojo)

Regras Estado Neutro

SE (valencia é valenciaNeutra) ENTÃO (Emoção é Neutra)

Após a construção das regras, o modelo de decisão baseado na lógica *fuzzy* foi aplicado para a categorização das emoções, com utilização de cada conjunto de regras para representar as diferentes emoções. Essa aplicação prática permitiu validar a funcionalidade do modelo, que se mostra promissor como um método de análise emocional de baixo custo, capaz de ser integrada a sistemas que demandam processamento eficiente.

O quadro 2 corresponde a matriz de confusão gerada no *FuzzyRules*, que avalia o desempenho do modelo em um problema de reconhecimento emocional, sendo que as classes estão associadas a diferentes emoções. A análise da matriz revela que os valores dispostos na diagonal principal indicam as predições corretas, enquanto os valores fora da diagonal representam os erros de classificação.

Na posição correspondente à emoção alegria (linha e coluna 1), observa-se que 79,9% (n=20) foram corretamente classificadas, mas há erros, como em 19,2% (n=5) das instâncias de alegria classificadas incorretamente como pertencentes à emoção surpresa e 3,8% (n=1) classificada como neutra. Da mesma forma, para a emoção medo, 88,5% (n=23) das instâncias foram corretamente classificadas, enquanto 11,5% (n=3) foram erroneamente atribuídas à emoção raiva. A emoção tristeza foi totalmente classificada de forma correta, sem erros, com 100% (n=26) das ocorrências. Para raiva, 84,6% (n=22) das instâncias foram corretamente classificadas, enquanto 15,4% (n=4) foram equivocadamente atribuídas à emoção medo. Na emoção surpresa, houve 80,8% (n=21) das classificações corretas, mas 19,2% (n=5) das instâncias foram classificadas como pertencentes à emoção alegria. O nojo apresentou 92,3% (n=24) das classificações corretas, com 7,7% (n=2) erros atribuídos à emoção medo. Por fim, a neutra foi corretamente classificada em sua totalidade, com 100% (n= 26) das ocorrências (Quadro 2).

Quadro 2 - Matriz de confusão do modelo *fuzzy*

Classe	Alegria	Medo	Tristeza	Raiva	Surpresa	Nojo	Neutra
Alegria	20	0	0	0	6	0	0
Medo	0	23	0	3	0	0	0
Tristeza	0	0	26	0	0	0	0
Raiva	0	4	0	22	0	0	0
Surpresa	5	0	0	0	21	0	0
Nojo	0	2	0	0	0	24	0
Neutra	0	0	0	0	0	0	26

Observa-se, na tabela 4, o desempenho do modelo de reconhecimento *fuzzy*. O coeficiente Kappa alcançou 87,18%. Este valor indica um nível substancial de concordância, próximo ao considerado excelente, conforme os critérios de avaliação padrão (LANDIS E KOCH, 1977). O coeficiente de AG forneceu a acurácia do modelo de

89,01% e o número absoluto de decisões corretas foi de 162, de 180 observações (Tabela 4).

Tabela 4 - Índices da estatística para o ajuste do modelo *fuzzy*.

Kappa	Acurácia	Decisões corretas
87,179	89,01	162

A tabela 5 apresenta os valores de sensibilidade e especificidade para cada emoção avaliada pelo modelo de reconhecimento. A sensibilidade apresentou valores elevados para a maioria das emoções, variando de 76,92% para alegria a 100% para tristeza e neutra. A especificidade também apresentou índices elevados, com valores acima de 93% para todas as classes de emoções. A acurácia global, de 89,01%, reforça o bom desempenho geral do modelo (Tabela 5).

Tabela 5 - Sensibilidade e especificidade por emoção do modelo *fuzzy*.

Emoção	VP	FN	FP	VN	Sensibilidade (%)	Especificidade (%)
Alegria	20	6	5	151	76,92	96,79
Medo	23	3	6	150	88,46	96,15
Tristeza	26	0	0	156	100,0	100,0
Raiva	22	4	3	153	84,62	98,08
Surpresa	21	5	6	150	80,77	96,15
Nojo	24	2	0	156	92,31	100,0
Neutra	26	0	0	156	100,0	100,0

Legenda: VP-Verdadeiro Positivo FN-Falso Negativo FP- Falso Positivo VN-Verdadeiro Negativo

O quadro 3 apresenta a descrição dos parâmetros acústico-prosódicos associados às emoções com base na construção das regras *fuzzy*. A emoção alegria apresenta curta duração, f_o mínima alta e velocidade de fala acelerada, enquanto os demais parâmetros não mostram informações que a discriminam. O medo é caracterizado por curta duração, valência negativa, f_o elevada em todos os aspectos (média, mínima e máxima) e ativação elevada. A tristeza demonstra valência negativa, intensidade baixa, ativação reduzida e potência fraca, com *jitter* baixo e HNR elevado, enquanto os demais parâmetros permanecem sem informações que a discriminam. A raiva possui curta duração, valência

negativa, f_o elevada (média, mínima e máxima), ativação alta, intensidade alta e potência forte. Já a surpresa é marcada por curta duração, valência positiva, f_o máxima e média altas, intensidade variável (baixa a alta) e ausência de informações que discriminam a emoção nos demais parâmetros. O nojo apresenta duração longa, valência negativa, intensidade que varia de muito baixa a muito alta, ativação baixa e potência forte, enquanto os outros parâmetros não exibem informações importantes de descriminação. O estado neutro não apresenta características marcantes em nenhum dos parâmetros analisados, sendo descrito apenas com aspecto de valência neutra (Quadro 3).

Quadro 3 - **Descrição dos parâmetros vocais das emoções baseados nas regras *fuzzy***.

	Duração	**F_o Média**	**F_o Mínima**	**F_o Máxima**	**F_o range**	**F_o dp**	**Jitter**	**HNR**	**CPPS**	**CPPS dp**	**Intensidade**	**Valência**	**Ativação**	**Potência**
Alegria	Curta	AD	Alta	AD	AD	AD	AD	Baixo a alto	Baixo a alto	AD	AD	AD	AD	AD
Medo	Curta	Alta	Alta	AD	AD	AD	AD	Baixo	Baixo	Alto	AD	Negativa	Alta	AD
Tristeza	AD	AD	AD	AD	AD	AD	AD	AD	AD	AD	Baixa	Negativa	Baixa	Fraca
Raiva	Curta	Alta	Baixa a alta	Alta	Baixo a alto	AD	AD	Baixo a alto	Baixo a alto	Alto	Alta	Negativa	Alta	Forte
Surpresa	Curta	Alta	Alta	AD	AD	Baixo	Baixo	Alto	Baixo	Baixo a alto	Baixa a alta	Positiva	AD	AD
Nojo	Alta	AD	AD	AD	AD	AD	AD	AD	Baixo	AD	Baixa a alta	Negativa	Baixa	Forte
Neutra	AD	AD	AD	AD	AD	AD	AD	AD	AD	AD	AD	Neutra	AD	AD

Legenda: AD = Ausência de discriminação

5.2 Modelos de Aprendizado de Máquinas

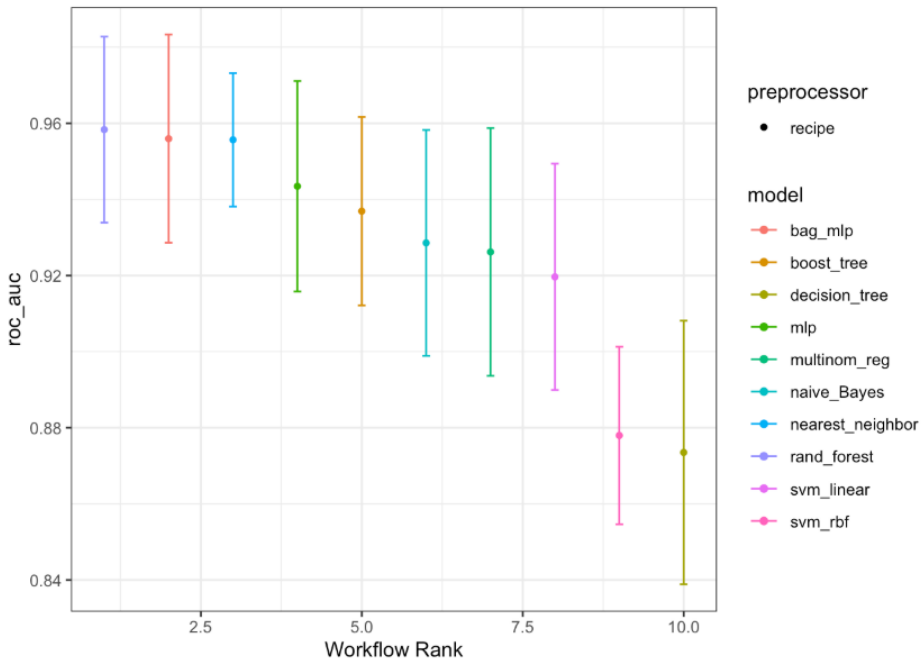
Foi realizado o comparativo de desempenho do modelo de reconhecimento das emoções *fuzzy* com outros métodos que realizaram a classificação básica das emoções.

Após o pré-processamento dos dados e definição do conjunto de validação, os hiperparâmetros dos modelos foram otimizados no processo de validação cruzada repetida. A busca pelos valores ótimos dos hiperparâmetros se deu a partir de um processo de busca em uma grade aleatória de valores definida através de um esquema de hipercubo latino, com vistas a preencher adequadamente o espaço de valores dos hiperparâmetros.

Para a análise dos dados, foi calculada a área sob a curva ROC (AUC-ROC) no conjunto de validação, de acordo com o melhor conjunto de hiperparâmetros ajustados para cada modelo. Esse cálculo permite avaliar o desempenho de cada modelo em distinguir corretamente entre as categorias da variável-alvo. Embora a área sob a curva ROC (AUC-ROC) tenha sido calculada para os modelos tradicionais como forma de avaliar sua capacidade de distinguir corretamente entre as categorias da variável-alvo, essa métrica não é adequada para o modelo *fuzzy*. Isso ocorre porque a geração da curva ROC no modelo *fuzzy* exige a realização de múltiplas interações, o que não é coerente com a natureza desse modelo, tornando a análise por essa métrica pouco significativa. Assim, a avaliação da acurácia no conjunto de validação, com base nos mesmos hiperparâmetros otimizados, foi adotada como medida complementar para compreender o desempenho geral dos modelos.

Na Figura 28, observou-se os resultados da AUC-ROC para cada modelo de aprendizado testado no conjunto de validação, destacando como cada abordagem se comportou utilizando os hiperparâmetros ajustados. A visualização ilustra as diferenças de desempenho entre os modelos, evidenciando aqueles que apresentaram maior capacidade de discriminar corretamente entre as classes, com base nas características dos dados analisados.

Figura 28 - Área sob a curva ROC dos modelos de aprendizado de máquinas



Legenda: bag_mlp: Redes neurais multi-layer perceptron com bagging; - **boost_tree:** Extreme gradient boosting; decision_**tree**: Árvore de decisão; **mlp**: Redes neurais multi-layer perceptron; **multinom_reg**: Elastic net; **naive_Bayes**: Redes naive bayes; **nearest_neighbor:** KNearest neighbors; **rand_forest:** Random Forest; **svm_linear**: Linear support vector machine; **svm_rbf**: Kernel support vector machine.

Dessa forma, os modelos de aprendizado de máquinas foram ordenados de acordo com o resultado da AUC-ROC (Quadro 4).

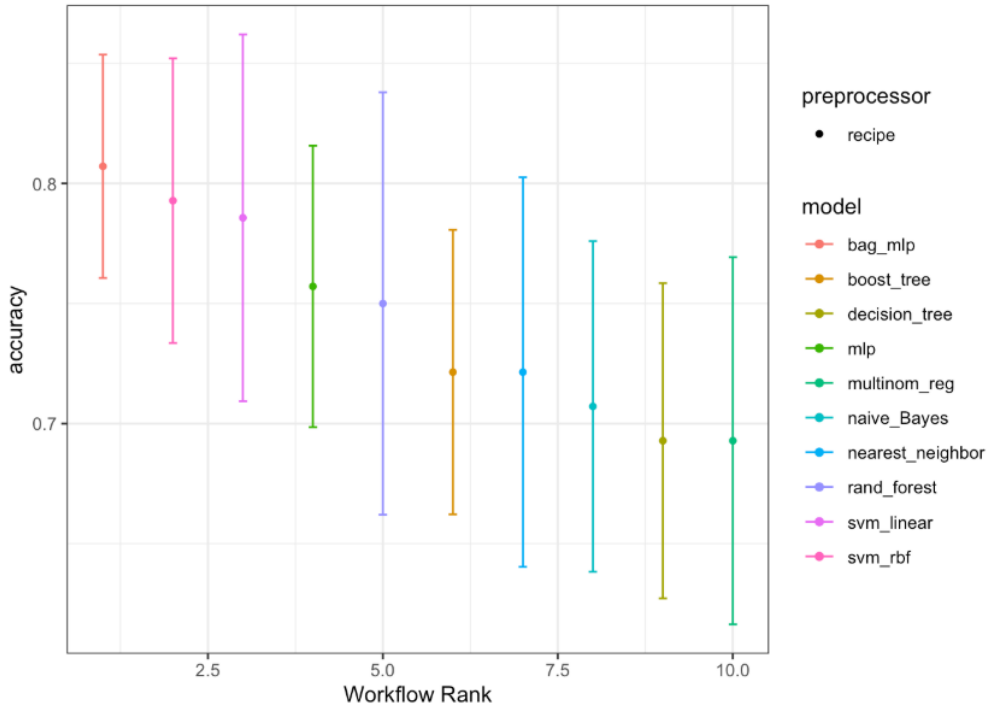
Quadro 4 - Ranqueamento dos modelos de aprendizado de máquinas investigados na etapa de validação de acordo com o valor da área sob a curva ROC.

RANKING	MODELO	Média	DP
1	Random Forest	0.958	0.014
2	Redes Neurais MLP com Bagging	0.955	0.016
3	KNN	0.955	0.010
4	Redes Neurais MLP	0.943	0.016
5	XGBoost	0.936	0.015
6	Redes Naive Bayes	0.928	0.018
7	EL Net	0.926	0.019
8	SVM Linear	0.919	0.018
9	Kernel SVM	0.877	0.014
10	Árvore de Decisão	0.873	0.021

Legenda: MLP: Multi-layer Perceptron - **KNN:** KNearest Neighbors; **SVM**: Support Vector Machine; **XGBoost:** Extreme gradient boosting; **EL Net**: Elastic net; **RBF**: Radial Basis Function; **DP**: desvio-padrão.

Em seguida, foi extraída a acurácia no conjunto de validação de acordo com o melhor conjunto de hiperparâmetros obtidos para cada modelo.

Figura 29 - Acurácia dos modelos de aprendizagem



Legenda: bag_mlp: Redes neurais multi-layer perceptron com bagging; - **boost_tree:** Extreme gradient boosting; decision_**tree**: Árvore de decisão; **mlp**: Redes neurais multi-layer perceptron; **multinom_reg**: Elastic net; **naive_Bayes**: Redes naive bayes; **nearest_neighbor:** KNearest neighbors; **rand_forest:** Random Forest; **svm_linear**: Linear support vector machine; **svm_rbf**: Kernel support vector machine.

A tabela 6 contém os resultados das medidas de desempenho de reconhecimento obtidas pelos diferentes modelos ajustados com seus respectivos conjuntos de hiperparâmetros. Os modelos foram avaliados utilizando diversas métricas de classificação, que inclui acurácia, AUC-ROC, sensibilidade, especificidade e índice kappa (Tabela 6).

Tabela 6 - Valores dos modelos ajustados com os conjuntos de hiperparâmetros selecionados de acordo com a acurácia

Método	Acurácia	AUC	Sensibilidade	Especificidade	Kappa
Random Forest	0.8333	0.8355	0.9722	0.9669	0.8056
Kernel SVM	0.8333	0.8338	0.9722	0.8505	0.8056
Redes Neurais MLP Combinadas por Bagging	0.7857	0.7829	0.9643	0.9735	0.75
Árvore de decisão	0.7619	0.7762	0.9603	0.9246	0.7222
KNN	0.7619	0.7587	0.9603	0.9689	0.7222
Redes Neurais MLP	0.7619	0.7576	0.9603	0.957	0.7222
Linear SVM	0.7381	0.7402	0.9563	0.8995	0.6944

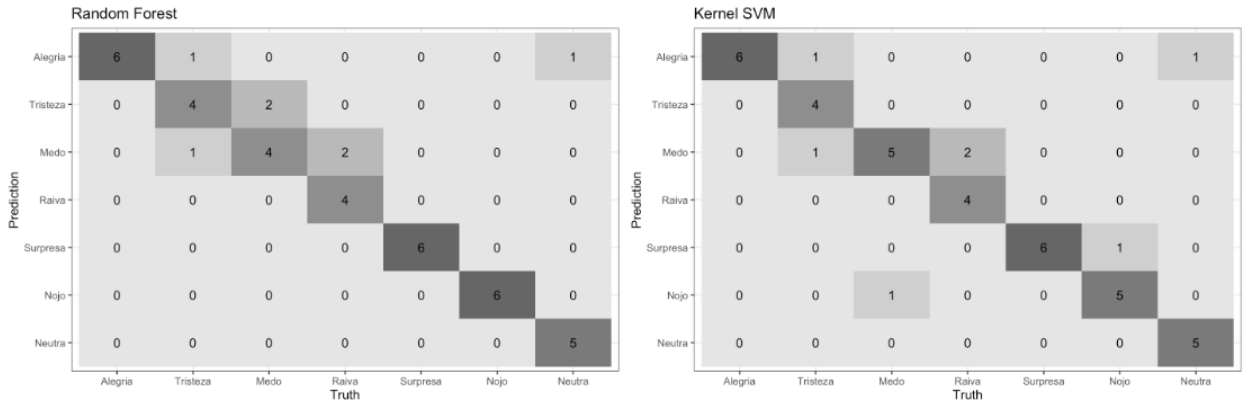
Redes Naive Bayes	0.7381	0.7257	0.9563	0.9471	0.6944
XGBoost	0.6905	0.687	0.9484	0.9193	0.6389
EL Net	0.6429	0.647	0.9405	0.916	0.5833

MLP: Multi-layer Perceptron - **KNN:** KNearest Neighbors; **SVM**: Support Vector Machine; **XGBoost:** Extreme gradient boosting; **EL Net**: Elastic net; **RBF**: Radial Basis Function**; AUC**: Area Under Curve.

Apenas os modelos Random Forest e Kernel SVM foram selecionados ao considerar os critérios de desempenho estabelecidos no presente estudo. Ambos alcançaram uma acurácia de 0,8333, AUC ≥ 0.80, e apresentando valores de Kappa de 0,9669 e 0,8505, respectivamente. Esses modelos também demonstraram alta sensibilidade (0,9722), o que indica que conseguem identificar corretamente uma alta proporção de positivos.

Observa-se as matrizes de confusão para os modelos Random Forest e Kernel SVM na classificação das emoções (Figura 30). Ambos os modelos apresentaram bom desempenho na alegria, tristeza, surpresa, nojo e neutro, com seis a cinco classificações corretas em geral, de sete possibilidades. No entanto, o Kernel SVM mostrou menor confusão na emoção medo, o que indica desempenho ligeiramente superior em relação ao Random Forest nesse aspecto.

Figura 30 - Matriz de confusão dos modelos Random Forest e Kernel SVM



Considerando o exposto, os resultados deste trabalho evidenciam o potencial das técnicas baseadas na lógica *fuzzy* no reconhecimento das emoções a partir de dados vocais. Os valores de acurácia, sensibilidade, especificidade alcançados indicam a relevância dessa abordagem, que preserva a continuidade dos parâmetros vocais e evita segmentações rígidas nas variáveis, características observadas nas abordagens tradicionais de aprendizado de máquina.

6 DISCUSSÃO

A voz humana é um fenômeno complexo que reflete aspectos fisiológicos, psicológicos e emocionais e pode ser considerada como biomarcador sensível das emoções. Com os avanços tecnológicos, a inteligência artificial tem se destacado como uma ferramenta inovadora na identificação de estados emocionais a partir da análise vocal, que pode ser aplicada para auxiliar no diagnóstico, monitoramento e acompanhamento de diversas condições de saúde (MILLING et al., 2022; HANSEN et al., 2022; KHAN et al., 2023; CANSEL et al., 2023).

Estudos indicam que parâmetros vocais, como f_o, intensidade, qualidade vocal e taxa de elocução, sofrem alterações em resposta a diferentes estados emocionais (SUNDBERG, 2015; ALHINTI et al., 2021). As mudanças que ocorrem na fisiologia da produção vocal durante a expressão das emoções são mediadas por ajustes no sistema nervoso autônomo, que influenciam a tensão muscular laríngea, o fluxo respiratório e a coordenação pneumofonoarticulatória (BARBOSA; FRIEDMAN, 2007; JOHNSTONE et al., 2017; KIM et al., 2020). A análise da voz, associada a modelos computacionais, tem demonstrado potencial na identificação e predição de estados emocionais, além de oferecer uma ferramenta não invasiva e de baixo custo para o monitoramento da saúde mental e emocional.

O conhecimento das medidas vocais contribui para construção da definição sobre as variações emocionais comuns nos sinais vocais (SILVA; BARBOSA, 2020). No âmbito clínico, a análise vocal facilita o diagnóstico de condições relacionadas à expressão emocional e à saúde mental, como transtornos de ansiedade e depressão, contribuindo para uma abordagem mais direcionada e eficaz (BAIRD et al., 2021; HANSEN et al., 2021). Além disso, o mapeamento detalhado dessas medidas vocais favorece o desenvolvimento de modelos de reconhecimento automático que identificam padrões emocionais com base na voz (JING, et al., 2018; ASIYA; KIRAN, 2021; HASHEM et al., 2023). Essa integração de tecnologia e ciência vocal amplia as aplicações práticas, desde assistentes virtuais mais empáticos até ferramentas para monitoramento emocional em tempo real, com a finalidade de promover avanços significativos na interface entre comunicação humana e inteligência artificial.

A análise das variações emocionais manifestadas na voz e a construção de bancos de dados vocais que capturam essas nuances tem sido amplamente estudada em diferentes línguas/culturas e incluem gravações associadas a diferentes estados

emocionais, coletadas tanto em condições controladas quanto em cenários espontâneos (BURKHARDT et al., 2005; BUSSO et al., 2008; MCKEOWN et al., 2012; RINGEVAL et al., 2013). No entanto, ainda existem lacunas importantes em relação à exploração de medidas acústico-prosódicas e perceptuais mais aprofundadas. Além disso, poucos estudos incorporam o impacto do julgamento de ouvintes especialistas em voz ou investigam padrões acústicos comuns a cada emoção de forma sistemática. Esses bancos são fundamentais para pesquisas que buscam compreender melhor as relações entre emoção e voz (SWAIN et al., 2018).

Apesar do crescente interesse na construção de bancos de vozes para análise de variações emocionais, observa-se uma escassez de acervos voltados para o PB, o que limita a generalização e a aplicabilidade dos resultados em contextos locais. Nesse cenário, o EmoVox-BR se destaca como um banco de vozes que adota critérios rigorosos de controle na coleta dos dados, utiliza tarefas de fala cuidadosamente balanceadas para captar diferentes estados emocionais, que permitem uma análise mais detalhada de parâmetros acústicos e prosódicos (LIMA, 2022). Além disso, o banco integra avaliações realizadas por juízes especialistas em voz, que garante uma validação perceptual das emoções expressas. Outro diferencial importante é a consideração dos aspectos culturais e linguísticos inerentes ao PB, que desempenham um papel importante na manifestação emocional.

Esses fatores permitem incluir variações prosódicas, diferenças na entonação e padrões de expressão emocional que podem variar quando comparado com outras línguas (LIU; PELL, 2014; KAUR, et al., 2018; MUÑETÓN-AYALA et al., 2017). O EmoVox-BR, ao abordar essas particularidades, não apenas preenche uma lacuna na pesquisa vocal em PB, mas também estabelece uma base sólida para o desenvolvimento de tecnologias e modelos preditivos mais alinhados à realidade sociocultural brasileira.

Diferentes tarefas de fala são utilizadas para diferenciar o estado emocional devido à sua capacidade de explorar aspectos específicos da produção vocal, como prosódia, entonação e intensidade, com o objetivo de maximizar a distinção entre emoções (AL-TALABANI et al., 2015; SEAL et al., 2018). Vogais sustentadas, por exemplo, têm sido empregadas para analisar parâmetros acústicos como frequência fundamental e qualidade vocal, permitindo identificar variações relacionadas a emoções básicas como felicidade, tristeza ou raiva (WAARAMAA et al., 2010). Frases balanceadas, com conteúdo semântico, são úteis para avaliar a entonação e o ritmo (JOHNSON et al., 1986). A leitura de textos padronizados possibilita a captura de mudanças emocionais na prosódia de

forma consistente, que exclui influências do conteúdo (PELL; KOTZ, 2011). Outras abordagens, como sons não verbais e interjeições, têm sido usadas para analisar a prosódia emocional (BOSTANOV; KOTCHOUBEY, 2004). Essa ausência de padronização reforça a necessidade de procedimentos mais consistentes, que possam aumentar a precisão na identificação de estados emocionais e aprimorar o desenvolvimento de tecnologias e modelos preditivos nessa área.

A escolha de frases padronizadas ou leituras controladas são consideradas estratégias eficazes para analisar os parâmetros vocais, pois destacam as nuances vocais relacionadas a emoções, oferecem dados consistentes e estruturados para a extração de características emocionais e possibilitam melhorar o treinamento de modelos de reconhecimento. Alguns estudos têm demonstrado a eficácia do uso de frases padronizadas no reconhecimento de emoções (RINGEVAL et al., 2014; HERACLEOUS; YONEYAMA, 2019; ER, 2020; LI et al., 2021). Contudo, ainda se observa uma ausência de padronização na forma de coleta e nas tarefas de fala em pesquisas que utilizam bancos de vozes para o reconhecimento emocional, o que compromete a uniformidade dos dados e dificulta comparações entre pesquisas.

A tarefa de fala "olha lá o avião azul" utilizada neste estudo, contém uma variedade de sons que abrangem diferentes articulações e posições vocálicas, que permite a análise ampla de aspectos acústicos e articulatórios da fala do PB (BEHLAU et al., 2020). Essa característica proporciona um recurso valioso para investigações sobre emoções, uma vez que oferece um estímulo de fala consistente e diversificado, que favorece a identificação de sutis mudanças prosódicas e de timbre que podem refletir estados emocionais. Além disso, a uniformidade de uma frase padronizada contribui para a comparabilidade entre estudos, reduz a variabilidade introdutória na coleta dos dados e aumenta a confiabilidade dos resultados (SANDAGE et al., 2015, PINHEIRO et al., 2022).

As medidas vocais representam outro importante atributo no reconhecimento emocional, pois fornecem pistas acústicas que auxiliam na identificação de emoções a partir da voz (VERVERIDIS; KOTROPOULOS, 2006; SAKURAI; KOSAKA 2021; HASHEM et al., 2023). Entre os achados principais, destacam-se os aspectos prosódicos, como variações de f_o, intensidade e duração, que diferenciam as emoções (PERVAIZ; KHAN, 2016; AGUIAR et al., 2024). Além disso, medidas de qualidade vocal, como o HNR e medidas cepstrais, aumentam a precisão, especialmente para emoções como felicidade e raiva (LEE et al., 2013). A intensidade da emoção impacta a acurácia, com emoções de alta intensidade, como raiva e medo, sendo reconhecidas mais rapidamente que emoções

de baixa intensidade, como alegria (MORNINGSTAR et al., 2021). Dinâmicas temporais e características de falantes e ouvintes, incluindo diferenças de gênero, também são relevantes. De modo geral, abordagens automatizadas, baseadas em inteligência artificial, demonstram potencial para aprimorar sistemas de reconhecimento emocional.

Um estudo prévio demonstrou que medidas acústico-prosódicas de duração, f_0 e intensidade discriminam as emoções alegria, tristeza, medo, raiva, surpresa, nojo e emissão neutra em falantes do PB pertencentes ao EmoVox-BR (AGUIAR, et al., 2024). Os resultados indicaram que as emoções influenciam as medidas acústico-prosódicos de maneira distinta. A emoção alegria foi caracterizada por uma fala mais acelerada, com menor duração e maior intensidade média, enquanto nojo apresentou maior duração e taxa de elocução, indicando uma fala mais lenta. A raiva destacou-se pela maior intensidade vocal máxima, e surpresa registrou os valores mais altos de frequência fundamental. Em contraste, tristeza e medo apresentaram as menores intensidades e frequências fundamentais. O medo exibiu também os menores valores de assimetria positiva de z-score, refletindo menor alongamento nos segmentos analisados. Os achados confirmam que as medidas acústico-prosódicas são ferramentas eficazes para discriminar estados emocionais, e o destacou o potencial dessas análises para o desenvolvimento de tecnologias de reconhecimento de emoções (AGUIAR, et al., 2024).

A subjetividade das emoções, caracterizada pela variação contínua e pela interpretação individual de cada estado emocional, representa um desafio significativo para o desenvolvimento de sistemas que buscam identificar padrões precisos na voz TALELE; JAIN, 2023). Nesse contexto, a lógica *fuzzy* se apresenta como uma ferramenta eficaz, ao permitir a modelagem de emoções como estados graduais, em vez de categorias fixas e absolutas (VASHISHTHA et al., 2023). Essa abordagem reconhece que emoções podem coexistir em diferentes graus, capturando melhor a complexidade das variações presentes nos sinais vocais. A lógica *fuzzy* utiliza regras e variáveis linguísticas para interpretar parâmetros vocais e possibilita uma análise mais precisa e adaptativa (ZADEH, 1965). Quando aplicada em bancos de vozes como o EmoVox-BR, essa metodologia incorpora julgamentos perceptuais de especialistas e considera as particularidades culturais e linguísticas do PB, tornando o reconhecimento de emoções mais contextualizado e sensível. Nesta pesquisa, essa modelagem foi utilizada para construir um modelo preditivo de reconhecimento emocional, e os principais resultados observados nesse processo foram discutidos a seguir.

A utilização de um modelo baseado na lógica *fuzzy* permite lidar com a natureza ambígua e subjetiva das emoções e fornece uma representação flexível e adaptável das variações vocais associadas a cada estado emocional (TON-THAT; CAO, 2019; BARROS et al., 2023). Diversos estudos reforçam o potencial da lógica *fuzzy* em lidar com variabilidade e desafios de generalização no reconhecimento emocional (ALI et al. 2020; FAROOQUE; MUÑOZ-HERNÁNDEZ 2016; ZHANG et al. 2017; LILIANA et al., 2019; ROVETTA et al. 2020).

Os resultados a seguir apresentam uma análise detalhada da eficácia do modelo *fuzzy* para o reconhecimento emocional a partir de dados vocais, destaca os acertos, limitações e potenciais aplicações no reconhecimento automático das emoções em contextos reais.

No presente estudo, cada emoção foi caracterizada por parâmetros vocais específicos que a diferenciam de forma singular das demais. A identificação de parâmetros exclusivos para cada emoção é essencial para a lógica *fuzzy*, pois permite a formulação de regras claras e distintas que capturam as nuances emocionais, utilizando variáveis como medidas de f_0, duração, intensidade e medidas cepstrais (CAO et al., 2014; DHAMYAL et al., 2023). Esses parâmetros específicos criam perfis acústicos e prosódicos para cada emoção, ativando conjuntos particulares de regras e variáveis linguísticas que são exclusivos a cada estado emocional (ZADEH, 1975).

A presença desses parâmetros diferenciais oferece ao modelo *fuzzy* a capacidade de reconhecer emoções com precisão. Ao vincular variáveis acústicas específicas a cada emoção, o sistema *fuzzy* consegue não apenas detectar a presença de uma emoção, mas também captar variações sutis entre diferentes estados emocionais (MING et al., 2015; TON-THAT; CAO, 2019; ROVETTA et al., 2019). Essa diferenciação detalhada amplia a capacidade do modelo de capturar as nuances emocionais e garantir que o reconhecimento seja fiel aos perfis de emoção analisados.

Para a construção do modelo *fuzzy*, a exclusão de determinadas variáveis foi uma etapa inicial, fundamentada na identificação de redundâncias informacionais que comprometiam a capacidade de discriminação precisa entre os estados emocionais. A presença de dados sobrepostos dificultou a definição de padrões únicos para cada emoção e aumentou a possibilidade de ambiguidade nos processos de classificação.

Algumas variáveis, como os MFCC, parâmetros de duração z-score e z-suavizado, e parâmetros acústicos de GNE, foram excluídas do modelo *fuzzy* devido ao fato de apresentarem intersecções com outras variáveis. As medidas apresentaram baixo poder

de discriminação entre as emoções e poderiam introduzir inconsistências nos resultados. Esse efeito levou à decisão de eliminá-los, pois a presença de variáveis que não contribuem para a diferenciação emocional clara pode adicionar complexidade desnecessária e tornar o modelo mais suscetível a erros (GUPTA, 2018; ZHANG; HAN, 2021).

Além disso, a inclusão dessas variáveis piorou o desempenho do modelo, e aumentou o ruído no sistema e reduziu sua precisão. Quando apresentam alto grau de interseção ou redundância, essas variáveis tendem a carregar informações sobrepostas, o que dificulta a distinção clara entre categorias ou estados emocionais. No contexto de um modelo *fuzzy*, essa redundância pode gerar ruído no sistema, aumentando a complexidade sem agregar informações úteis, o que compromete a clareza das inferências geradas pelas regras *fuzzy* (BRANCO; DENTE, 2000; LUGHOFER et al., 2011). Somado a isso, variáveis interseccionadas inflacionam o número de regras necessárias no modelo, tornando-o mais denso e dificultando tanto sua interpretação quanto sua manutenção. Esse cenário também prejudica a precisão do modelo, uma vez que variáveis redundantes priorizam características com baixo impacto para a diferenciação emocional, levando a decisões menos robustas. Outro efeito negativo é a redução da capacidade de generalização do sistema, já que a redundância favorece o ajuste excessivo aos dados de treinamento (*overfitting*), comprometendo seu desempenho em novos conjuntos de dados (JUANG; TSAO, 2008).

Embora tenha inicialmente reduzido o número de parâmetros analisados, a simplificação resultante foi benéfica para refinar o conjunto de variáveis mais relevantes, ou seja, aquelas com maior poder preditivo sobre as emoções. Essa abordagem permitiu identificar os parâmetros mais robustos para a discriminação dos estados emocionais, como variações da f_0 (média, mínima, máxima, range e dp), duração, intensidade, HNR, CPPS, *jitter*, *shimmer*, bem como os eixos dimensionais das emoções (valência, potência e ativação), reduzindo a redundância e aumentando a eficiência do sistema. Além disso, a exclusão de variáveis sobrepostas levou à geração de um menor número de regras *fuzzy*, o que tornou o sistema com maior simplicidade e menor custo computacional, preservando a precisão necessária para as aplicações propostas.

Essa simplificação contribuiu para um modelo mais direto e focado nas variáveis realmente determinantes, facilitando sua implementação. A exclusão desses parâmetros também contribuiu para o desenvolvimento de um modelo com maior eficiência e de menor custo computacional, com operações mais simples e processamento mais rápido. Assim,

a simplificação do conjunto de variáveis manteve apenas os parâmetros com relação clara às emoções e garantiu um reconhecimento mais eficiente e preciso.

Neste estudo, observou-se que os MFCC, amplamente utilizados em sistemas de reconhecimento das emoções a partir da voz (DURUKAL; HOCAOGLU, 2015, LIKITHA et al., 2017; GOEL, 2018, PATNAIK, 2023), apresentaram limitações no modelo para o PB. Essa medida prioriza frequências médias e baixas, relevantes para o conteúdo linguístico, mas pode perder nuances emocionais em frequências altas ou transições rápidas. A compressão espectral dos MFCC gerou sobreposições entre características cruciais, como variações de entonação, intensidade e qualidade vocal. No PB, cuja prosódia é fundamental na expressão emocional, os coeficientes baseados na escala mel não captaram adequadamente as interações dinâmicas entre *pitch*, *loudness* e ritmo.

O estudo de Sharif (2023) destacou uma limitação dos MFCC na determinação do número de características, o que pode causar sobreajuste e perda de distinção entre aspectos cruciais, como entonação, intensidade e qualidade vocal. A pesquisa envolveu o reconhecimento de sotaques ingleses utilizando *Recurrent Neural Network LSTM* a partir da extração de características MFCC, para treinamento e teste do modelo. Os resultados revelaram uma acurácia máxima de 64,96%. O estudo concluiu que a atribuição dos MFCC aumentou a dimensão das características e diminuiu o desempenho do modelo, fato que evidenciou uma limitação no uso desse atributo.

Portanto, o processamento de variáveis que não explicam bem o modelo elevaria o custo computacional, sem ganhos proporcionais na performance e diminuiria sua eficiência (MAGDALENA, 2018). A exclusão dessas medidas foi fundamental para otimizar a simplicidade do modelo *fuzzy* e aprimorar sua capacidade de diferenciar as emoções de maneira eficiente e robusta, com foco exclusivamente nas variáveis que contribuíram de maneira efetiva para a identificação precisa das emoções.

Foram realizadas análises criteriosas para definir os parâmetros de entrada e saída para a construção das regras *fuzzy*, baseadas em parâmetros acústicos e acústico-prosódicos que integram a etapa final do modelo de reconhecimento de emoções. As variáveis linguísticas de entrada foram elaboradas a partir de medidas relacionadas à f_o, intensidade, duração das unidades VV, medidas cepstrais e características emocionais como valência, potência e ativação. Essas variáveis foram representadas por funções de pertinência cuidadosamente ajustadas, que permitiram associar diferentes intervalos de valores às nuances vocais específicas de cada emoção. O sistema *fuzzy* foi estruturado de forma a integrar essas variáveis em regras que traduzem as interações entre os

parâmetros vocais e os estados emocionais, a fim de promover inferências sensíveis às sutilezas do sinal de voz.

As regras *fuzzy* para a emoção alegria utilizam parâmetros acústicos e atributos prosódicos para identificar padrões que correspondem a esse estado emocional. A primeira regra estabelece que a emoção é classificada como alegria quando a valência é positiva, a duração da unidade VV [az] é muito muito pequena, muito pequena ou pequena, e o CPPS varia entre níveis pequenos e muito muito altos. A segunda regra complementa essa análise ao identificar a emoção alegria quando a valência também é positiva, a a f_o mínima é alta, muito alta ou muito muito alta, o HNR varia entre muito muito pequeno e muito alto, e a duração da unidade VV [ianu] abrange desde valores pequenos até muito muito altos. Esses critérios demonstram que a alegria é caracterizada por parâmetros acústicos que refletem alta energia vocal, f_o elevada, consistência harmônica e padrões de fala acelerada, indicadores típicos de maior intensidade emocional.

Para a emoção medo, as regras *fuzzy* utilizam uma combinação de valência negativa, parâmetros acústicos e prosódicos para identificar padrões característicos dessa emoção. A primeira regra define que o medo é identificado quando a valência é negativa, a f_o média varia de média a muito muito alta, o HNR varia de muito pequeno a muito alto, o dp de CPPS é médio ou alto, a duração da unidade VV [az] é curta, e a potência vocal é forte. A segunda regra complementa essa análise ao indicar que o CPPS é de pequeno a alto, o HNR é alto ou muito alto, a duração da unidade VV [al] é curta, e a ativação é alta. Regras adicionais reforçam a identificação do medo em contextos de valência negativa quando há combinações específicas de parâmetros, como f_o mínimo muito muito alto ou CPPS reduzido combinado com potência vocal forte e ativação elevada. As regras reforçam que o medo ocorre em contextos com valência negativa e valores intermediários de f_o mínimo e f_o médio, associados a HNR reduzido, CPPS baixo e alta ativação. Essas combinações refletem padrões de alta tensão vocal, curta duração entre emissões e características de alta energia e esforço vocal, comuns no estado emocional de medo.

Quanto as regras *fuzzy* para a emoção tristeza, demonstram que sua identificação ocorre especialmente com base na valência negativa, potência vocal fraca e baixa ativação. As características acústicas e acústico-prosódicas apresentaram ambiguidades ao diferenciar tristeza do estado neutro. Alguns estudos indicam que a tristeza apresenta características menos expressivas e compartilha propriedades acústicas semelhantes ao estado neutro, o que pode causar confusão entre os dois estados emocionais (PEREIRA; WATSON, 1998; YILDIRIM et al., 2004; SUNDBERG et al., 2019).

Portanto, a análise pelos eixos dimensionais de valência, potência e ativação permitiu uma separação mais clara entre esses estados emocionais. A tristeza é marcada por baixos níveis de energia vocal, níveis baixos de frequência e intensidade emocional, refletindo um padrão distinto em relação ao estado neutro, que apresenta maior estabilidade e equilíbrio nos parâmetros prosódicos (YILDIRIM et al., 2004). Essa distinção destaca a importância de considerar os eixos dimensionais para a classificação precisa de emoções que compartilham semelhanças acústicas, como tristeza e neutralidade, especialmente em contextos que exigem alta sensibilidade na identificação de estados emocionais.

As regras *fuzzy* para a emoção raiva utilizam múltiplos parâmetros acústicos e prosódicos para caracterizar esse estado emocional, enfatizando a interação entre valência negativa e alta ativação. A primeira regra identifica a raiva por meio de f_o média elevada, alto dp de CPPS, níveis de ruído harmônico (HNR) variados e ativação alta, sugerindo forte energia vocal e tensão emocional. A segunda regra considera a amplitude mínima da f_o variando de muito baixa a muito alta, frequência máxima predominantemente elevada, curta duração da unidade VV [auav] e intensidade vocal alta, refletindo um padrão vocal enérgico e instável. Outras regras reforçam a associação da raiva com curta duração nas unidades VV [al] e [ianu], variações no alcance tonal (f_o range) e presença de altos níveis de ruído harmônico. Portanto, a raiva é caracterizada por sinais acústicos com altos níveis de intensidade, variabilidade tonal, alta ativação e elementos que indicam alta excitação emocional, que a distingue de outras emoções negativas.

Quanto as regras *fuzzy* para a emoção surpresa, combinou informações de valência positiva, parâmetros acústicos e prosódicos para sua identificação. A surpresa é caracterizada por f_o média geralmente altas ou muito altas, associadas a baixos níveis de suavidade espectral (CPPS) e de seu dp. A curta duração de unidades VV [auav] é uma característica marcante, refletindo um padrão vocal breve e abrupto que enfatizam a natureza inesperada dessa emoção. Além disso, os valores de intensidade vocal variam amplamente, desde muito baixa até muito alta, que pode indicar flexibilidade na projeção vocal. O HNR, que representa o equilíbrio harmônico, tende a ser médio ou alto, enquanto a f_o mínima pode atingir valores muito altos, reforçando o elemento de excitação. Observa-se que a surpresa combina elementos de alta variabilidade acústica, excitação emocional e de velocidade de fala acelerada, refletindo reatividade imediata e padrões vocais distintivos associados a esse estado emocional.

Na emoção nojo, as regras *fuzzy* destacam a valência negativa como característica central, associada a parâmetros acústicos que refletem baixa energia vocal, excitação reduzida e velocidade de fala lenta. O nojo é identificado por ativações consistentemente baixas em todas as regras, associadas a durações maiores nos segmentos vocais, como [az] e [al], que variam de muito alta a muito muito alta, reforçando a característica de prolongamento vocal e ênfase emocional. A duração da unidade VV [ianu], em contrapartida, apresenta valores mais curtos, refletindo uma dinâmica contrastante dentro do padrão vocal. A intensidade vocal apresenta ampla variabilidade, desde valores muito baixos até níveis altos, enquanto a potência vocal pode ser forte em determinados cenários, sugerindo momentos de maior ênfase emocional. O dp da f_o é predominantemente pequeno ou muito pequeno, indicando estabilidade tonal, e baixos valores de *jitter* que refletem regularidade nos ciclos vocais. Em algumas condições, valores muito baixos de CPPS refletem uma voz menos harmônica, alinhando-se ao aumento de ruído e tensão (LOWELL et al., 2013) frequentemente associado ao nojo. Esses critérios revelam padrões acústicos que enfatizam uma reação aversiva, marcada por baixa excitação, longa duração e ajustes específicos na projeção vocal.

A regra *fuzzy* para o estado neutro definem exclusivamente pela valência neutra, sem especificar outros parâmetros acústicos ou prosódicos, o que evidencia a dificuldade de análise desse estado emocional. O estado neutro apresenta desafios interpretativos porque frequentemente compartilha características acústicas com outras emoções de baixa intensidade, como a tristeza, tornando-se menos distinto nos parâmetros vocais (GUZMÁN et al., 2013). A ausência de variações marcantes em atributos como f_o, intensidade, duração e suavidade espectral dificulta a identificação de padrões exclusivos. Além disso, sua classificação depende fortemente do contexto e da percepção subjetiva, pois o neutro funciona como um ponto de equilíbrio entre estados emocionais opostos, o que pode gerar ambiguidades durante a análise acústica. Verifica que a neutralidade carece de valência, ou seja, não possui carga afetiva positiva ou negativa, e por esse motivo apresenta desafios interpretativos em análises vocais devido à sua natureza menos distinta (GASPER et al., 2019). No contexto do modelo *fuzzy*, que lida com categorias não rigidamente delimitadas, a valência surge como a métrica mais objetiva e operacionalizável para diferenciar o neutro dos estados emocionais valenciados.

A ambiguidade entre as emoções tristeza e neutra foi um exemplo claro da dificuldade em modelar categorias emocionais que apresentam características vocais pouco definidas ou sobrepostas. No modelo desenvolvido, essa sobreposição gerou

desafios para a diferenciação entre essas duas emoções, uma vez que muitos parâmetros acústicos analisados apresentaram valores semelhantes para ambas. No entanto, ao considerar a variável de valência, que reflete a carga emocional positiva ou negativa associada à emoção, foi possível identificar uma distinção clara entre os estados (BESTELMEYER et al., 2017; SPEED, BRYSBAERT, 2023). Essa abordagem permitiu a criação de uma única regra *fuzzy* para a emissão neutra, suficiente para resolver essa limitação e assegurar a precisão do modelo no reconhecimento dessas categorias. Essa simplificação demonstrou a importância de selecionar variáveis altamente discriminativas para lidar com ambiguidades emocionais em sistemas de reconhecimento de emoções.

Assim, a utilização de regras com características exclusivas para cada emoção oferece melhor precisão do modelo *fuzzy*, uma vez que a categorização detalhada das emoções vai além da identificação genérica e promove uma distinção sensível a características acústicas específicas. Os resultados fortalecem a aplicabilidade do modelo *fuzzy* em cenários práticos de reconhecimento emocional, como na análise de interações em atendimento ao cliente, no monitoramento de saúde mental, ou em sistemas de análise emocional para mídia e entretenimento, onde a diferenciação precisa de estados emocionais é fundamental para a qualidade do serviço.

A análise da matriz de confusão revelou que o modelo *fuzzy* classificou corretamente a maioria das instâncias, com destaque para as emoções tristeza e neutra, que não apresentaram erros de categorização. Por outro lado, a emoção alegria exibiu maior número de confusões, principalmente com surpresa, o que refletiu a dificuldade do modelo em distinguir emoções de valência positiva com características acústicas semelhantes. Apesar disso, a predominância de valores na diagonal principal da matriz evidenciou um bom desempenho geral, indicando que o modelo manteve coerência na categorização emocional, evitando classificações aleatórias. As confusões observadas ocorreram majoritariamente dentro do mesmo espectro de valência, sugerindo sobreposição de traços vocais entre emoções positivas, como alegria e surpresa, ou entre estados de baixa ativação, como neutra e tristeza. Esse padrão pode estar associado a mecanismos biomecânicos da produção vocal, uma vez que emoções de valência negativa tendem a apresentar marcadores vocais mais consistentes e distintos (JAYWANT; PELL, 2012), o que facilita sua identificação. A precisão na classificação da tristeza reforçou esse achado. Adicionalmente, os valores de especificidade confirmaram a capacidade do modelo em evitar falsos positivos, o que fortaleceu sua eficácia na exclusão de emoções incorretas. Assim, os resultados indicam um bom desempenho do

modelo na distinção emocional, com margem para aprimoramentos que aumentem a discriminação entre emoções com características acústicas mais próximas.

O coeficiente Kappa indicou uma concordância substancial no desempenho o que reflete consistência do modelo na classificação correta das emoções. Além disso, a acurácia geral do modelo reforçou a robustez do sistema em diferenciar corretamente as emoções. Os índices de sensibilidade e especificidade para cada emoção confirmam o desempenho positivo do modelo. A sensibilidade, que reflete a capacidade do modelo em identificar corretamente as instâncias de cada emoção, variou de níveis altos para alegria, a valores perfeitos para tristeza e neutra. Esses resultados sugerem que o modelo foi eficaz em detectar corretamente emoções com padrões bem definidos. A especificidade, por sua vez, foi elevada para todas as emoções, demonstrando que o modelo também foi eficiente em evitar falsos positivos. Esses resultados indicam que o modelo *fuzzy* apresenta uma solução promissora para problemas de reconhecimento emocional, com excelente acurácia e alta capacidade de discriminação entre diferentes emoções.

Os resultados encontrados neste estudo destacaram a lógica *fuzzy* ao aprimorar os modelos de suporte à decisão para o reconhecimento de emoções, especialmente no tratamento das ambiguidades e complexidades inerentes às emoções humanas. Diferentemente dos métodos tradicionais, que dependem de classificações rígidas e exigem relações discretas de entrada-saída, frequentemente enfrentando dificuldades com a natureza sobreposta dos estados emocionais, a lógica *fuzzy* demonstrou grande eficácia no gerenciamento de incertezas e imprecisões.

Diante da efetividade dessa modelagem, diversos estudos que empregaram a lógica *fuzzy* para o reconhecimento das emoções obtiveram resultados superiores aos métodos de aprendizado de máquinas (CHATCHINARAT et al., 2016; LILIANA et al., 2019; TON-THAT; CAO, 2019). A lógica *fuzzy* se destaca como uma abordagem mais intuitiva e aplicada devido ao seu benefício potencial de compreensão humana. Ademais, exibe menor tempo de computação, sendo mais rápida do que outras técnicas de aprendizado de máquina, especialmente quando o número de regras é gerenciável (PAREEK et al., 2023). Essas características tornam a lógica *fuzzy* uma solução robusta, adaptável e eficiente em cenários desafiadores envolvendo a diferenciação de estados emocionais.

Essa abordagem processou de forma eficaz pistas emocionais diferenciadas, para que emoções com características vocais sobrepostas fossem representadas por funções de pertinência que definiram os graus de pertencimento (ANAAM et al., 2023). Esse

método resultou em um mapeamento preciso dos estados emocionais a partir de padrões vocais, mesmo em condições de dados ambíguos. Além disso, ao incorporar variáveis linguísticas que simulavam o raciocínio humano, o sistema tornou-se altamente interpretável em comparação com os modelos baseados em aprendizado de máquina, fato que possibilita a integração de diversos recursos acústicos em regras *fuzzy* que refletiam a experiência humana.

Foi realizada uma análise de desempenho em relação a métodos de aprendizado de máquinas para o reconhecimento das emoções para validação do modelo baseado em lógica *fuzzy*. Essa análise foi importante para contextualizar a eficiência do modelo *fuzzy* no cenário mais amplo de reconhecimento emocional, que permitiu avaliar como a abordagem utilizada ofereceu vantagens em termos de modelagem, acurácia, sensibilidade e especificidade em relação a outras técnicas. Vale destacar que a finalidade não foi identificar o modelo com o melhor desempenho absoluto, considerando as diferenças entre as abordagens e aplicabilidades das metodologias analisadas. O foco residiu na apresentação de uma proposta inovadora que leva em conta a incerteza das variações emocionais, fundamentada na lógica *fuzzy*, explorada como uma nova perspectiva metodológica em vez de uma alternativa competitiva direta. A comparação com outras metodologias teve como propósito contextualizar e demonstrar a eficácia do modelo desenvolvido em cenários práticos, oferecendo um ponto de referência para análise dos resultados. Além disso, o comparativo forneceu uma base para identificar as vantagens e limitações do modelo, confirmar sua aplicabilidade prática e destacar possíveis melhorias em relação às abordagens existentes.

Os testes realizados inicialmente com as variáveis originais não apresentaram resultados satisfatórios em termos de acurácia, o que levou à necessidade de aplicar transformações espectrais no conjunto de variáveis (BORKOWSKI et al., 2023). Essas transformações permitiram converter os dados do domínio do tempo para o domínio da frequência, destacando padrões relevantes para a classificação emocional que não eram evidentes nas variáveis originais.

A transformação espectral nos dados foi realizada para capturar relações não lineares e preservar tanto a estrutura global quanto local das amostras no espaço original (WEN et al., 2021). Essa abordagem realizada devido a característica complexa e sobreposta do conjunto de dados, pois permite reduzir a dimensionalidade enquanto mantém informações essenciais para a análise subsequente. Além disso, a laplaciana normalizada empregada na transformação reduz os impactos de ruído e

desbalanceamento entre as amostras, o que garante maior robustez na representação dos dados. No dos dados do EmoVox-BR, a transformação espectral foi utilizada para revelar padrões subjacentes e facilitar a separação entre grupo para estabelecer uma análise mais precisa e visualização clara das relações entre as classes de interesse.

Porém, essa transformação adiciona complexidade computacional e pode não capturar integralmente as características emocionais sutis nos sinais, além de apresentar dificuldade ao interpretar os resultados em termos das variáveis originais, nesse sentido, dificulta a compreensão direta do comportamento dos dados e dos fenômenos subjacentes no sinal de entrada, tornando o modelo menos explicável (NADLER; GALUN, 2006; ZHU et al., 2017). O estudo de Noguchi (2020), analisou a precisão do agrupamento espectral em redes com estruturas sobrepostas e evidenciou suas limitações teóricas. O autor destacou que a incapacidade do método de explicar a sobreposição de informações leva à perda de informações estruturais no autovetor principal. A pesquisa revelou como essas informações estruturais são comprometidas em redes com sobreposição, ressaltando a necessidade de abordagens específicas para detecção de informações sobrepostas.

Após análise com base em critérios de desempenho, os modelos de aprendizado de máquinas Random Forest e Kernel SVM foram selecionados para o presente estudo de acordo com a performance em comparação a outros métodos avaliados para o reconhecimento de emoções a partir de dados vocais. Ambos apresentaram resultados consistentes em métricas globais, que evidenciou a capacidade de realizar classificações precisas e equilibradas. A comparação entre os modelos Random Forest e Kernel SVM destacou o desempenho consistente de ambos no reconhecimento das emoções. Esses métodos demonstraram alta acurácia, particularmente nas emoções alegria, tristeza, surpresa, nojo e neutra, evidenciando sua capacidade de reconhecer padrões emocionais em diferentes cenários. Embora os dois modelos tenham apresentado resultados similares em termos de métricas globais, como sensibilidade e especificidade, o Kernel SVM exibiu uma leve vantagem ao lidar com a emoção medo, mostrando menos confusões em comparação ao Random Forest. Essa diferença, ainda que discreta, indicou uma capacidade superior do Kernel SVM em captar nuances específicas desta emoção.

A análise entre os modelos de aprendizado de máquinas tradicionais e o modelo baseado em lógica *fuzzy* evidenciou diferenças importantes em desempenho e abordagem metodológica. Os modelos de aprendizado de máquinas necessitaram de etapas adicionais, como a transformação espectral das variáveis além de apresentarem

desempenho inferior em termos de acurácia. Apesar da eficácia das transformações espectrais em melhorar o desempenho dos modelos de aprendizado de máquina, a perda de conexão com as variáveis originais representou uma limitação importante em cenários de reconhecimento emocional, que exigem explicabilidade e alinhamento prático com os dados reais.

A utilização do modelo baseado em lógica *fuzzy* apresentou também limitações importantes. A definição das regras *fuzzy*, bem como a adaptação e modificação dessas regras, tornou-se desafiadora e dependeu fortemente do conhecimento, domínio e da experiência dos especialistas. Esse processo é subjetivo e muitas vezes requer múltiplas iterações, o que pode tornar o desenvolvimento do modelo demorado e suscetível a inconsistências. Além disso, a escolha adequada das funções de pertinência e dos parâmetros associados pode influenciar diretamente o desempenho do modelo, sendo difícil alcançar um equilíbrio ideal entre complexidade e precisão. A escalabilidade do modelo também pode ser comprometida à medida que o número de variáveis ou estados emocionais aumenta, levando a uma explosão combinatória no número de regras necessárias.

Por outro lado, a lógica *fuzzy* teve a vantagem de operar diretamente com as variáveis originais e com variáveis linguísticas derivadas, mantendo uma relação mais intuitiva e transparente com os dados de entrada, além de se destacar pela capacidade generalização e de lidar com incertezas e ambiguidades intrínsecas às emoções humanas, dispensando a necessidade de transformações nas variáveis. Esses fatores explicam o melhor desempenho em acurácia e a eficiência do modelo *fuzzy* na classificação de estados emocionais, especialmente em cenários onde as variáveis apresentam sobreposição ou imprecisão, tornando-a essencial para o reconhecimento de emoções humanas. Deve-se ressaltar que a lógica *fuzzy* pode também ser integrada a técnicas de aprendizado de máquina e outras metodologias de inteligência artificial, para elevar o nível de adaptabilidade e eficácia desses modelos (CHRYSAFIADI, 2022).

De forma geral, o modelo baseado em lógica *fuzzy* não apenas supera as limitações dos modelos tradicionais, mas também amplia a capacidade dos sistemas de identificar e responder adequadamente a complexidade dos estados emocionais. Ao preservar a conexão com as variáveis originais e incorporar a flexibilidade das variáveis linguísticas, essa abordagem oferece uma combinação única de precisão, explicabilidade e generalização. Essa generalização é possível no âmbito de diferentes contextos culturais e linguísticos, bem como na aplicação em outras bases de dados, com características

distintas. A lógica *fuzzy* se adapta a padrões emocionais extraídos de diferentes idiomas, e permite que o modelo identifique nuances emocionais específicas de cada língua. Além disso, ao integrar dados de múltiplas fontes, o modelo demonstra capacidade de aprendizado robusto, suporta análises mais amplas e inclusivas, com potencial para atender demandas em contextos globais. Portanto, sua habilidade de lidar com incertezas e sobreposições reforça seu potencial como ferramenta robusta para aplicações práticas e avançadas no reconhecimento emocional contribuindo para o desenvolvimento de sistemas mais eficientes, humanos e inclusivos.

7 CONCLUSÃO

A construção de um modelo baseado na lógica *fuzzy* para o reconhecimento de estados emocionais a partir de medidas acústicas e acústico-prosódicas da voz representa uma importante contribuição para o campo da análise emocional automatizada. Foi possível desenvolver um sistema robusto e flexível, capaz de lidar com a complexidade inerente às variações emocionais humanas ao utilizar características vocais específicas associadas a emoções básicas, como alegria, medo, tristeza, raiva, surpresa, nojo e a emissão neutra.

A lógica *fuzzy* demonstra ser uma abordagem eficaz para tratar incertezas e nuances nas expressões vocais, resultando em um modelo com maior capacidade de diferenciação e maior taxa de acurácia no reconhecimento emocional. A análise criteriosa e a exclusão de variáveis redundantes, como os MFCC e os parâmetros acústicos de GNE, foram passos fundamentais para otimizar o desempenho do modelo. Assim, essa abordagem permitiu a simplificação da estrutura analítica e destacou as medidas vocais com maior poder preditivo para a identificação dos estados emocionais, como as variações da f_0, duração, intensidade, HNR, CPPS, *jitter*, *shimmer*, além dos eixos dimensionais das emoções (valência, potência e ativação). Essas medidas apresentaram potencial para atuar como biomarcadores digitais de estados emocionais, oferecendo subsídios objetivos para a caracterização automatizada das emoções.

O modelo *fuzzy* mostrou-se eficiente em integrar múltiplas medidas vocais, preservando a simplicidade e a interpretabilidade das regras que orientam as decisões do modelo. Garantiu a precisão no reconhecimento das emoções, bem como viabilizou a generalização do modelo para diferentes contextos e populações.

Os resultados obtidos evidenciam que o uso de sistemas *fuzzy*, aliados à seleção criteriosa de variáveis acústicas, superou as limitações dos métodos tradicionais de aprendizado de máquinas na análise de emoções, especialmente em cenários onde as fronteiras entre estados emocionais são difusas. Esse trabalho reafirma o potencial da inteligência artificial e da lógica *fuzzy* na compreensão e no reconhecimento automático das emoções humanas, abrindo perspectivas para sua aplicação em áreas como saúde, comunicação, entretenimento e inovação para ser aplicada em sistemas humano-máquina.

Por fim, o modelo desenvolvido representa um avanço na interface entre tecnologia e cognição humana, com destaque pela sua capacidade de capturar e interpretar as

sutilezas emocionais a partir da voz. As contribuições deste estudo reforçam a importância de abordagens interdisciplinares e inovadoras para compreender as complexidades da expressão emocional, estabelecendo bases sólidas para futuros aprimoramentos e aplicações práticas em sistemas inteligentes.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU SR, MORAES RM, MARTINS PN, LOPES LW. VOXMORE: Artefato tecnológico para auxiliar a avaliação acústica da voz no processo ensino-aprendizagem e prática clínica. *CoDAS*, v. 35, n. 6, 2023.

ADRIANO T, ARRIAGA P. Exaustão emocional e reconhecimento de emoções na face da voz em médicos. *SPPS*, 17(1), p. 97-104, 2016.

AGUIAR AC, MORAES RM, CONSTANTINI AC, ALMEIDA AAA. Medidas acústico-prosódicas discriminam as emoções de falantes do português brasileiro. *Revista CoDAS.* 2024. (No prelo).

AHMAD K, MESIAROVÁ-ZEMÁNKOVÁ A. Choosing t-norms and t-conorms for fuzzy controllers. In: Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007). IEEE. 2007;2:641-646.

ALHINTI L, CHRISTENSEN H, CUNNINGHAM S. Acoustic differences in emotional speech of people with dysarthria. *Speech Commun.* 2021;126:44-60. doi: 10.1016/j.specom.2020.11.005.

ALI YM, ABD RAHIM AF, NOORSAL E, MAT YASIN Z. Fuzzy-based voiced-unvoiced segmentation for emotion recognition using spectral feature fusions. *Indonesian J Electr Eng Comput Sci.* 2020;19(1):196-206. doi: 10.11591/ijeecs.v19.i1.pp196-206.

ALMEIDA AAF, BEHLAU M, LEITE JR. Correlação entre ansiedade e performance comunicativa. Revista da Sociedade Brasileira de Fonoaudiologia. 16(4), p.384-386, 2011.

ALMEIDA LNA, LOPES LW, COSTA DB, GONÇALVES SILVA EG, CUNHA GMS, ALMEIDA AAF. Características vocais e emocionais de professores e não professores com baixa e alta ansiedade. Scielo, 19(2), p. 179-85, 2014.

ALMEIDA, AAF, Fernandes LR, Azevedo EHM, Pinheiro RSA, Lopes LW. Características vocais e de personalidade de pacientes com imobilidade de prega vocal. 27(2), p. 178–185, 2015.

AL-TALABANI A, SELLAHEWA H, JASSIM S. Emotion recognition from speech: Tools and challenges. In: Mobile Multimedia/Image Processing, Security, and Applications, 2015, Baltimore, Maryland, United States. Proceedings of SPIE - The International Society for Optical Engineering. 2015;9497. doi: 10.1117/12.2191623.

AMIR O, LEVINE-YUNDOF R. Listeners' attitude toward people with dysphonia. J Voice. 27(4) p. 524.e1-10, 2013.

AMORIM M, ANIKIN A, MENDES AJ, LIMA CF, KOTZ SA, PINHEIRO AP. Changes in vocal emotion recognition across the life span. Emotion. 21(2) p.315-325, 2021.

AN S, JI L, MARKS M, ZHANG Z. Two sides of emotion: Exploring positivity and negativity in six basic emotions across cultures. *Front Psychol.* 2017;8. doi: 10.3389/fpsyg.2017.00610.

ANAAM E, HAW SC, NG KW, et al. Utilizing fuzzy algorithm for understanding emotional intelligence on individual feedback. *JIWE.* 2023;2(2). doi: 10.33093/jiwe.2023.2.2.19.

AOUANI, H.; AYED, Y. BEN. Speech Emotion Recognition with deep learning. Procedia Computer Science, v. 176, p. 251–260, 2020.

ARIAS-LONDONO JD, GODINO-LLORENTE JI, SAENZ-LECHON N, et al. Automatic detection of pathological voices using complexity measures, noise parameters, and Mel-cepstral coefficients. IEEE Trans Biomed Eng. 58(2) p. 370–379, 2011.

ARVANITI A. The phonetics of prosody. In: ARONOFF M, CHEN Y, CUTLER C, eds. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press, 2020. DOI: 10.1093/acrefore/9780199384655.013.411.

ASHA, American Speech-Language-Hearing Association. Consensus auditoryperceptual evaluation of voice (CAPE-V). Rockville: ASHA Special Interest Division 3, Voice and Voice Disorders; 2002.

ASIYA UA, KIRAN VK. Speech emotion recognition - A deep learning approach. In: 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), Palladam, Índia, 2021. IEEE.

BAIRD A, TRIANTAFYLLOPOULOS A, ZÄNKERT S, et al. An evaluation of speech-based recognition of emotional and physiological markers of stress. *Front Comput Sci.* 2021;3.

BĄK H. Emotional prosody processing in nonnative English speakers. In: *Emotional prosody processing for non-native English speakers*. [S.l.]: [s.n.], 2016. p. 141-169. DOI: 10.1007/978-3-319-44042-2_7.

BANSE R, SCHERER KR. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, v. 70, n. 3, p. 614–636, 1996. DOI: 10.1037/0022-3514.70.3.614.

BARAVIEIRA P. B. Aplicação de uma rede neural artificial para avaliação da rugosidade e soprosidade vocal. 101 f. Tese de doutorado, Universidade de São Paulo, SP, 2016.

BARBOSA PA, CONSTANTINI AC. Editorial: prosódia e fonética acústica no Brasil. J Speech Sci. 2020;8(2):1-1.

BARBOSA PA. Incursões em torno de ritmo da fala. Campinas: Editora Pontes, 2006.

BARBOSA RA, FRIEDMAN S. Emoção: efeitos sobre a voz e a fala na situação em público. *Distúrbios da Comunicação*, v. 19, n. 3, p. 325-336, 2007.

BARRA-CHICOTE R, MONTERO J, MACIAS-GUARASA J, LUFTI S, LUCAS JM, FERNANDEZ F, D'HARO L, SAN-SEGUNDO R, FERREIROS J, CORDOBA R, PARDO J. Spanish Expressive Voices: Corpus for emotion research in Spanish. *Proceedings of the 6th International Conference on Language Resources and Evaluation*, 2008.

BARROS LC, BASSANEZI RC, LODWICK WA. Fuzzy sets theory and uncertainty in mathematical modeling. *Studies in Fuzziness and Soft Computing*, p. 1-21, 2023. DOI: 10.1007/978-3-031-50492-1_1.

BEHLAU, M.; ROCHA, B.; ENGLERT, M.; MADAZIO, G. Validation of the brazilian portuguese CAPE-V instrument—br CAPE-V for auditory-perceptual analysis. J voice. 36(4), p. 586.e15-586.e20, 2020.

BESTELMEYER PEG, KOTZ SA, BELIN P. Effects of emotional valence and arousal on the voice perception network. *Soc Cogn Affect Neurosci.* 2017;12:1351-1358. doi: 10.1093/scan/nsx059.

BHAKRE, SK; BANG, A. Emotion recognition on the basis of audio signal using Naive Bayes classifier. 2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016, p. 2363–2367, 2016.

BHATT S, JAIN A, DEV A. Feature Extraction Techniques with Analysis of Confusing Words for Speech Recognition in the Hindi Language. *Wireless Personal Communications*, Dordrecht, v. 118, n. 4, p. 3303, 2021. DOI: 10.1007/s11277-021-08181-0.

BORKOWSKI P, KŁOPOTEK MA, STAROSTA B, WIERZCHOŃ ST, SYDOW M. Eigenvalue-based spectral classification. *PLoS ONE*, v. 18, n. 4, p. e0283413, 2023. DOI: 10.1371/journal.pone.0283413.

BORKOWSKI P, KŁOPOTEK MA, STAROSTA B, WIERZCHOŃ ST, SYDOW M. Eigenvalue-based spectral classification. *PLoS ONE.* 2023;18(4):e0283413. doi: 10.1371/journal.pone.0283413.

BORREGO MCM, BEHLAU M. Mapeamento do eixo condutor da prática fonoaudiológica em expressividade verbal no trabalho de competência comunicativa. *CoDAS*, v. 30, n. 6, p. 4-7, 2018. DOI: 10.1590/2317-1782/20182018054.

BOSTANOV V, KOTCHOUBEY B. Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology.* 2004;41(2):259-268. doi: 10.1111/j.1469-8986.2003.00142.x.

BRANCO P, DENTE J. Noise effects in fuzzy modelling systems. *ArXiv.* 2000;cs.NE/0010002.

BROCKMANN-BAUSER M, DRINNAN MJ. Routine acoustic voice analysis: time to think again? Curr Opin Otolaryngol Head Neck Surg. 2011;19(3):165- 70

BURIN L. Accommodation of L2 speech in a repetition task: Exploring paralinguistic imitation. *Res Lang.* 2017;16:377-406. doi: 10.2478/rela-2018-0019.

BURKHARDT F, PAESCHKE A, ROLFES M, SENDLMEIER W, WEISS B. A Database of German Emotional Speech. INTERSPEECH. pp. 1517–1520, 2005.

BUSSO C, BULUT M, LEE CC, KAZEMZADEH A, MOWER E, KIM S, CHANG JN, LEE S, NARAYANAN SS. IEMOCAP: Interactive Emotional Dyadic Motion Capture Database. Language Resources and Evaluation. 42(4), p. 335–59, 2008.

BYUN, S.W.; LEE, S.P. A Study on a Speech Emotion Recognition System with Effective Acoustic Features Using Deep Learning Algorithms. Appl. Sci. 2021, 11, 1890. https://doi.org/10.3390/ app11041890

CANSEL N, ALCIN ÖF, YILMAZ ÖF, ARI A, AKAN M, UCUZ İ. A new artificial intelligence-based clinical decision support system for diagnosis of major psychiatric diseases based on voice analysis. *Psychiatr Danub.* 2023;35(4):489-499. doi: 10.24869/psyd.2023.489.

CAO H, BEŇUŠ Š, GUR RC, VERMA R, NENKOVA A. Prosodic cues for emotion: Analysis with discrete characterization of intonation. *Speech Prosody.* 2014;2014:130-134. doi: 10.21437/SpeechProsody.2014-14.

CHAMOVITZ, I.; ELIA, M. F; COSENZA, C. A. N., Fuzzy Assessment Model for Operative Groups in Virtual Educational, Science and Information Conference (SAI), 2015. IEEE, 2015. p. 395-405.

CHATCHINARAT A, WONG KW, FUNG CC. Fuzzy classification of human emotions using fuzzy C-mean (FCFCM). In: *2016 International Conference on Fuzzy Theory and Its Applications (iFuzzy).* IEEE; 2016.

CHRYSAFIADI K. The role of fuzzy logic in artificial intelligence and smart applications. *Learning and Analytics in Intelligent Systems.* 2022;25-29. doi: 10.1007/978-3-031-44457-9_2.

CIELO CA, BEBER BC, MAGGI CR, KORBES D, OLIVEIRA CF, WEBER DE. Disfonia funcional psicogênica por puberfonia do tipo muda vocal incompleta: aspectos fisiológicos e psicológicos. Estud Psicol (Campinas). 26(2), p. 227-36, 2009.

CIPRIANO. Desenvolvimento de Arquitetura Para Sistemas de Reconhecimento Automático de Voz Baseados em Modelos Ocultos de Markov. 123 f. Tese (Doutorado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001.

COHEN J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20(1):37-46.

CONGALTON RG, GREEN K. Assessing the Accuracy of Remotely Sensed Data Principles and Practices. Lewis Publishers, Boca Raton, 1999.

CONGALTON RG, GREEN K. Assessing the accuracy of remotely sensed data: Principles and practices. New York: Lewis Publishers; 1999.

CONGALTON RG. A review of assessing the accuracy classifications of remotely sensed data. *Remote Sensing of Environment*, New York, v. 49, n. 12, p. 1671-1678, Dec. 1991.

CONSERVA KCF, FRANÇA FP, LOPES LW. Medidas perceptivo-auditivas e acústicas de mulheres com e sem nódulos vocais. *Audiology - Communication Research*, v. 27, 2022. DOI: 10.1590/2317-6431-2022-2655pt.

CONSTANTINI AC, BARBOSA PA. Prosodic characteristics of different varieties of Brazilian Portuguese. *Revista Brasileira de Criminologia*, v. 4, p. 44-53, 2015.

COSTA DB, LOPES LW, SILVA EG, CUNHA GMS, ALMEIDA LNA, ALMEIDA AAF. Fatores de risco e emocionais na voz de professores com e sem queixas vocais. Rev. CEFAC. 15(4), p. 1001-1010, 2013.

COSTA SC, NETO BGA, FECHINE JM. Pathological voice discrimination using cepstral analysis, vector quantization, and hidden Markov models. In: IEEE International Conference on Bioinformatics and Bioengineering. 1–5, 2008.

COWIE R, DOUGLAS-COWIE E, TSAPATSOULIS N, VOTSIS G, KOLLIAS S, FELLENZ W, TAYLOR JG. Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine. v. 18(1), p 32-80, 2001.

DE LOPE J, GRAÑA M. An ongoing review of speech emotion recognition. *Neurocomputing.* 2023;528:1-11. doi: 10.1016/j.neucom.2023.01.002.

DE LOPE, J; GRAÑA, M. An ongoing review of speech emotion recognition. Neurocomputing, v. 528, p. 1-11, 2023.

DHAMYAL H, ELIZALDE B, DESHMUKH S, WANG H, RAJ B, SINGH R. Prompting Audios Using Acoustic Properties for Emotion Representation. *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 11936-11940, 2023.

DIETRICH M, VERDOLINI ABBOTT K. Vocal function in introverts and extraverts during a psychological stress reactivity protocol. *Journal of Speech, Language, and Hearing Research*, v. 55, p. 973–987, 2012. DOI: 10.1044/1092-4388(2011/10-0344).

DURUKAL M, HOCAOGLU A. Performance analysis of MFCC features on emotion recognition from speech. *Int J Sci Technol Res.* 2015;1:1-7.

EDDINS, DA,; SHRIVASTAV, R. Psychometric properties associated with perceived vocal roughness using a matching task. The Journal of the Acoustical Society of America, 134 (4), p. 294-300, 2013.

EKMAN P. An argument for basic emotions. Cognition and emotion. 6, 169-200, 1992.

EKMAN P. The Handbook of Cognition and Emotion, ch. Basic Emotions. Wiley Online Library, pp. 45–60, 1999.

ELSAYED T. Fuzzy inference system for the risk assessment of liquefied natural gas carriers during loading/offloading at terminals. *Applied Ocean Research*, 2009. Disponível em: https://doi.org/10.1016/j.apor.2009.08.004.

ER M. A novel approach for classification of speech emotions based on deep and acoustic features. *IEEE Access.* 2020;8:221640-221653. doi: 10.1109/ACCESS.2020.3043201.

ESAU, Natascha; WETZEL, Evgenija; KLEINJOHANN, Lisa; KLEINJOHANN, Bernd. Real-Time Facial Expression Recognition Using a Fuzzy Emotion Model. In: *2007 IEEE International Fuzzy Systems Conference*, 2007, London. Proceedings [...]. London: IEEE, 2007. p. 1-6. DOI: 10.1109/FUZZY.2007.4295451.

FANG SH, FEI YX, XU ZZ, et al. Learning transportation modes from smartphone sensors based on deep neural networks. *IEEE Sens J*. 17:6111–6118, 2017.

FANG, S. H. et al. Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. Journal of Voice, 33(5), p. 634–641, 2019.

FAROOQUE M, MUÑOZ-HERNÁNDEZ S. Easy fuzzy tool for emotion recognition - prototype from voice speech analysis. *International Joint Conference on Computational Intelligence.* 2016.

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

FERNANDES J, TEIXEIRA F, GUEDES V, JUNIOR A, TEIXEIRA JP. Harmonic to Noise Ratio Measurement - Selection of Window and Length. *Procedia Computer Science*, v. 138, p. 280-285, 2018. DOI: 10.1016/j.procs.2018.10.040.

FERREIRA R.P, et al. Study on daily demand forecasting orders using artificial neural network. IEEE Latin America Transactions, 14(3), p.1519-1525, 2016.

FILIPPA M, LIMA D, GRANDJEAN A, LABBÉ C, COLL SY, GENTAZ E, GRANDJEAN DM. Emotional prosody recognition enhances and progressively complexifies from childhood to adolescence. *Scientific Reports*, v. 12, n. 1, p. 17144, 2022. DOI: 10.1038/s41598-022-21554-0. PMID: 36229474; PMCID: PMC9561714.

FITCH J. Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *J Speech Hear Disord.* 1990;55(2):360-363. doi: 10.1044/JSHD.5502.360.

GALES SYM. The Application of Hidden Markov Models in Speech Recognition. 2008.

GANGANI A, ZHANG L, JIANG M. Speech Emotion Recognition Using Convolutional Recurrent Neural Networks. In: *Intelligent Management of Data and Information in Decision Making: Proceedings of the 16th FLINS Conference on Computational Intelligence in Decision and Control; the 19th ISKE Conference on Intelligence Systems and Knowledge Engineering (FLINS-ISKE 2024)*. World Scientific Proceedings Series on Computer Engineering and Information Science, p. 283-290, 2024. DOI: 10.1142/9789811294631_0036.

GASPER K, SPENCER LA, HU D. Does Neutral Affect Exist? How Challenging Three Beliefs About Neutral Affect Can Advance Affective Research. *Frontiers in Psychology*, v. 10, p. 2476, 8 nov. 2019. DOI: 10.3389/fpsyg.2019.02476.

GERMANO RB, TCHEOU MP, HENRIQUES FR, GOMES JUNIOR SP. Emouerj: an emotional speech database in Portuguese. Zenodo, Sep. 2021.

GODINO-LLORENTE JI, 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. IEEE Trans Biomed Eng. 53:1943–1953, 2006.

GODINO-LLORENTE JI, OSMA-RUIZ V, SÁENZ-LECHÓN N, GÓMEZ-VILDA P, BLANCO-VELASCO M, CRUZ-ROLDÁN F. effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. J Voice. 2010;24(1):47- 56

GOEL M. A novel technique for speech recognition using modified MFCC. *J Emerg Technol Innov Res.* 2018.

GOMES VEFI, COSTA DB, LOPES LW, ARAUJO RA, ALMEIDA AAF. Symptoms and vocal risk factors in individuals with high and low anxiety. *Folia Phoniatrica et Logopaedica*, v. 71, n. 1, p. 7-15, 2019. DOI: 10.1159/000494211.

GOMIDE FAC, GUDWIN RR. Modelagem. Controle, Sistemas e Lógica Fuzzy, 1994.

GRIMM, M; KROSCHEL, K; MOWER, E; NARAYANAN, S. Primitives-based evaluation, and estimation of emotions in speech. *Speech Communication*, v. 49, n. 10, p. 787-800, 2007.

GUIMARÃES DS. Symbolic objects as sediments of the intersubjective stream of feelings. *Integrative Psychological; Behavioral Science*, v. 44, n. 3, p. 208-216, 2010.

GUPTA S, GUPTA A. Handling class overlapping to detect noisy instances in classification. *Knowledge Eng Rev.* 2018;33(8). doi: 10.1017/S0269888918000115.

GUZMÁN M, CORREA S, MUÑOZ D, MAYERHOFF R. Influence on spectral energy distribution of emotional expression. *J Voice.* 2013;27(1):129.e1-129.e10. doi: 10.1016/j.jvoice.2012.08.008.

GUZY A. Psychosocial factors and problems with voice production. *The European Health Psychologist*, v. 18, p. 990, 2016.

HALSZKA B. Emotion universals—Argument from nature. In: *Emotional Prosody Processing for Non-native English Speakers.* Cham: Springer; 2016. p. 27-51. doi: 10.1007/978-3-319-44042-2_2.

HANSEN L, ZHANG YP, WOLF D, et al. A generalizable speech emotion recognition model reveals depression and remission. *Acta Psychiatr Scand.* 2022;145(2):186-199.

HANSEN L, ZHANG YP, WOLF D, SECHIDIS K, LADEGAARD N, FUSAROLI R. A generalizable speech emotion recognition model reveals depression and remission. *Acta Psychiatr Scand.* 2022;145(2):186-199. doi: 10.1111/acps.13388.

HASHEM, A; ARIF, M; ALGHAMDI, M. Speech emotion recognition approaches: A systematic review. Speech Communication, v. 154, p. 102974, 2023.

HEGDE S., SHETTY S., RAI S., DODDERI T. A survey on machine learning approaches for automatic detection of voice disorders. Journal of Voice. 33(6) p. 947.e11–947.e33, 2019.

HERACLEOUS P, YONEYAMA A. A comprehensive study on bilingual and multilingual speech emotion recognition using a two-pass classification scheme. *PLoS ONE.* 2019;14.

HIDALGO R, LÓPEZ V, URGELÉS D. Voice-based Mood Recognition: An Application to Mental Health. In: *Intelligent Management of Data and Information in Decision Making*. World Scientific Proceedings Series on Computer Engineering and Information Science, p. 41-48, 2024. DOI: 10.1142/9789811294631_0006.

HOSMER DW, LEMESHOW S. Applied logistic regression. New York: Wiley; 2000. doi: 10.1002/0471722146.

HUANG X, ACERO A, HON H. Spoken Language Processing. Prentice Hall, Upper Saddle River, NJ, 2001.

IDRISOGLU A, DALLORA A, ANDERBERG P, BERGLUND J. Applied machine learning techniques to diagnose voice-affecting conditions and disorders: Systematic literature review. *J Med Internet Res.* 2023;25. doi: 10.2196/46105.

IMANDOUS SB, BOLANDRAFTAR M. Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background. *International Journal of Engineering Research and Applications*, v. 3, n. 5, p. 605–610, 2013.

IQBAL, M. et al. Artificial Neural Network-based Emotion Classification and Recognition from Speech. International Journal of Advanced Computer Science and Applications, v. 11, n. 12, p. 434–444, 2020.

IRIYA, R. Análise de sinais de voz para reconhecimento de emoções. Tese de Doutorado, Universidade de São Paulo, 2014.

IRVAN M, JUNAIDI A, DEWANTORO D, PUSPITASARI E, PRAMESTI Y, RAMADHANI V. Emotional Expressions Recognize Ability in Children with Autism: Intervention Using Emotion Recognition Assistance Application. *2022 8th International Conference on Education and Technology (ICET)*, p. 87-90, 2022. DOI: 10.1109/ICET56879.2022.9990898.

IYER R, HOSMER DW, LEMESHOW S. Applied Logistic Regression. *The Statistician*, v. 40, p. 458, 1991.

JACOBSON BH., et al. The voice handicap index (VHI): development and validation. Am J Speech Lang Pathol., v. 6, p. 66–70. 1997.

JAHANGIR, R. et al. Deep learning approaches for speech emotion recognition: state of the art and research challenges. Multimedia Tools and Applications, Springer, p. 1–66, 2021.

JAIN D, SHARMA S, DHIMAN P. Comparative analysis of defuzzification techniques for fuzzy output. *Journal of Algebraic Statistics*, v. 13, n. 2, p. 874–882, 2022. DOI: 10.52783/jas.v13i2.234.

JAIN, M. NARAYAN S., BALAJI P., BHOWMICK A., MUTHU R. K. Speech Emotion Recognition using Support Vector Machine. 2020.

JAYWANT A, PELL MD. Categorical processing of negative emotions from speech prosody. *Speech Communication*, v. 54, n. 1, p. 1-10, Jan. 2012.

JERMSITTIPARSERT, K. et al. Pattern recognition and feature selection for speech emotion recognition model using deep learning. International Journal of Speech Technology, 23(4), p. 799–806, 2020.

JING S, MAO X, CHEN L. Prominence features: Effective emotional features for speech emotion recognition. *Digit Signal Process.* 2018;72:216-231. doi: 10.1016/j.dsp.2017.10.016.

JOHNSON WF, EMDE RN, SCHERER KR, KLINNERT MD. Recognition of emotion from vocal cues. *Arch Gen Psychiatry.* 1986;43(3):280-283. doi: 10.1001/archpsyc.1986.01800030098011.

JOHNSTONE T, VAN REEKUM C, BÄNZIGER T, et al. The effects of difficulty and gain versus loss on vocal physiology and acoustics. *Psychophysiology.* 2007;44(5):827-837. doi: 10.1111/j.1469-8986.2007.00552.x.

JOKINEN, E.; ALKU, P. Estimating the spectral tilt of the glottal source from telephone speech using a deep neural network. The Journal of the Acoustical Society of America, 141 (4), p. EL327–EL330, 2017.

JUANG C, TSAO Y. A self-evolving interval type-2 fuzzy neural network with online structure and parameter learning. *IEEE Trans Fuzzy Syst.* 2008;16:1411-1424. doi: 10.1109/TFUZZ.2008.925907.

KAI-JUN X. Speech Emotion Recognition Based on Fuzzy Logic Theory. *Informatization Research*, 2011.

KAUR J, JUGLAN K, SHARMA V. Role of acoustic cues in conveying emotion in speech. *J Forensic Sci Crim Invest.* 2018;11(1).

KAUR, J., SINGH, A., KADYAN, V.: Sistema de reconhecimento automático de fala para idiomas tonais: levantamento de última geração. Arco. Computar. Metanfetamina Eng. 28 (3), p. 1039–1068, 2021.

KAYACAN E, KHANESAR MA. Fuzzy Neural Networks for Real Time Control Applications: Concepts, Modeling and Algorithms for Fast Learning. Chapter 2 - Fundamentals of Type-1 Fuzzy Logic Theory, p. 13-24, 2016. DOI: 10.1016/B978-0-12-802687-8.00002-5.

KERKENI, L. et al. Automatic speech emotion recognition using an optimal combination of features based on emd-tkeo. Speech Communication, Elsevier, v. 114, p. 22–35, 2019.

KHAN H, ULLAH M, AL-MACHOT F, CHEIKH FA, SAJJAD M. Deep learning-based speech emotion recognition for Parkinson patients. In: *Proceedings of the Electronic Imaging 2023 Conference*. Society for Imaging Science and Technology; 2023. doi: 10.2352/EI.2023.35.9.IPAS-298.

KIM J, TOUTIOS A, LEE S, NARAYANAN S. Vocal tract shaping of emotional speech. *Comput Speech Lang.* 2020;64. doi: 10.1016/j.csl.2020.101100.

KINGESKI, R. Desenvolvimento de um banco de dados de voz com emoções em idioma português brasileiro. Dissertação (Mestrado em Engenharia Elétrica) – Joinville (SC): Universidade do Estado de Santa Catarina, 2019.

KLIR GJ, YUAN B. Fuzzy sets and fuzzy logic: Theory and applications. Prentice-Hall; 1995.

KLOFSTAD CA, ANDERSON RC, PETERS S. Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society of London B*, v. 297, p. 2698–2704, 2012.

KOVACIC Z, BOGDAN S. Fuzzy controller design: Theory and applications. CRC Press; 2006.

KRISHNAN PT, ALEX NOEL JR, RAJANGAM V. Emotion classification from speech signal based on empirical mode decomposition and non-linear features. Complex; Intelligent Systems, 7, p.1919–1934, 2021.

KUMAR J., PRABHAKAR O., SAHU N. Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems: A Review. International Journal of Innovative Research in Computer and Communication Engineering, 2 (1), 2260-2269, 2014.

LANDIS JR, KOCH GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.

LARROUY-MAESTRI P, POEPPEL D, PELL MD. The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science*, 2023.

LAUKKA P, ELFENBEIN HA. Emotion appraisal dimensions can be inferred from vocal expressions. *Social Psychological and Personality Science*, v. 3, n. 5, p. 529–536, 2012. DOI: 10.1177/1948550611428011.

LAUSEN A, HAMMERSCHMIDT K. Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications*, v. 7, p. 2, 2020.

LAUSEN A, HAMMERSCHMIDT K. Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications*, v. 7, p. 2, 2020. DOI: 10.1057/s41599-020-0499-z.

LECUN, Y, BENGIO Y, HINTON G. Deep learning. Nature. 521, p. 436-444, 2015.

LEE J, CHOI J, KANG H. Analysis of voice quality features and their contribution to emotion recognition. *JBE.* 2013;18:771-774. doi: 10.5909/JBE.2013.18.5.771.

LEVENSON R. Basic emotion questions. *Emotion Rev.* 2011;3:379-386. doi: 10.1177/1754073911410743.

LI D, ZHOU Y, WANG Z, GAO D. Exploiting the potentialities of features for speech emotion recognition. *Inf Sci.* 2021;548:328-343. doi: 10.1016/j.ins.2020.09.047.

LI, J., et al.: Jasper: an end-to-end convolutional neural acoustic model. INTERSPEECH. Graz, Austria, pp. 71-75, 2019.

LI, X, LIN, R. "Speech Emotion Recognition for Power Customer Service", *2021 7th International Conference on Computer and Communications (ICCC)*, pp. 514-518, 2021.

LI, X.; NEIL, D.; DELBRUCK, T.; et al. Lip reading deep network exploiting multi-modal spiking visual and auditory sensors. In: *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. [Anais…] IEEE, 2019. p. 1-5.

LI, Z.; HUANG, J.; HU, Z. Screening and diagnosis of chronic pharyngitis based on deep learning. International Journal of Environmental Research and Public Health, v. 16, n. 10, 2019.

LIKITHA MS, GUPTA SR, HASITHA K, RAJU AU. Speech-based human emotion recognition using MFCC. In: *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE; 2017.

LILIANA, DY; BASARUDDIN, T.; WIDYANTO, MR; ORIZA, IID. Fuzzy emotion: a natural approach to automatic facial expression recognition from psychological perspective using fuzzy system. Cognitive Processing, v. 20, n. 4, p. 391-403, nov. 2019.

LIMA LMM. de, RIBEIRO KSQS., MORAES RM. Modelo de suporte à decisão aplicado ao acesso a fisioterapia de indivíduos com Acidente Vascular Encefálico. Revista Eletrônica Acervo Saúde, 11(15), e1391, 2019.

LIMA, HMO.; ALMEIDA, AA.; ALMEIDA, LNA. Elaboração e validação do Banco de Vozes Brasileiro nas Variações das Emoções (EMOVOX-BR). In: 30º Congresso Brasileiro de Fonoaudiologia, 2022, João Pessoa. Anais do 30º Congresso Brasileiro de Fonoaudiologia. São Paulo: Sociedade Brasileira de Fonoaudiologia, 2022. v. 1. p. 4298-4302.

LIU P, PELL MD. Processing emotional prosody in Mandarin Chinese: A cross-language comparison. In: Proceedings of the International Conference on Speech Prosody 2014. Dublin, Ireland; 2014.

LIU X. Parameterized defuzzification with maximum entropy weighting function – another view of the weighting function expectation method. *Mathematical and Computer Modelling*, v. 45, p. 177-188, 2007.

LOPES LW, SOUSA ESS, SILVA ACF, SILVA IMD, PAIVA MAA, VIEIRA VJD, et al. Cepstral measures in the assessment of severity of voice disorders. CoDAS. 2019;31(4):e20180175

LOPES, LW et al. Acurácia das medidas acústicas tradicionais e formânticas na avaliação da qualidade vocal. Codas. 30(5), p. 1-10, 2018.

LOW DM; GHOSH SS; BENTLEY KH. Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investigative Otolaryngology. 5, p.96–116, 2020.

LOWELL S, COLTON R, KELLEY R, MIZIA S. Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *Journal of Voice*, v. 27, n. 4, p. 393-400, 2013. DOI: 10.1016/j.jvoice.2013.02.005.

LUGGER, M.; YANG, B. Cascaded emotion classification via psychological emotion dimensions using a large set of voice quality parameters. In: IEEE. IEEE International Conference on Acoustics, Speech and Signal Processing, [S.1.], 2008. p. 4945-4948, 2008.

LUGHOFER E, BOUCHOT J, SHAKER A. On-line elimination of local redundancies in evolving fuzzy systems. *Evolving Syst.* 2011;2:165-187. doi: 10.1007/s12530-011-9032-3.

MA Z.C.; FOKOUÉ E. A Comparison of Classifiers in Performing Speaker Accent Recognition Using MFCCs. Open Journal of Statistics, 4, p. 258-266. 2014.

MAGDALENA L. Designing interpretable hierarchical fuzzy systems. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE; 2018. doi: 10.1109/FUZZ-IEEE.2018.8491452.

MAMDANI EH. Application of fuzzy algorithms for control of simple dynamic plant. *Proc IEEE.* 1974;121(12):1585-1588.

MANOHAR, V., CHEN, S.-J., WANG, Z., FUJITA, Y., WATANABE, S., KHUDANPUR, S.:Acoustic modeling for overlapping speech recognition: JHU CHiME-5 challenge system. In: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6665–6669, 2019.

MARQUES IR, BARBOSA SF, BASILE AL, MARIN HD. Guia de Apoio à Decisão em Enfermagem Obstétrica: aplicação da técnica da Lógica Fuzzy. *Revista Brasileira de Enfermagem*, v. 58, p. 349-354, 2005.

MARTINS, TGS; SILVA, RUFINO; MENDES, LGA; SCHOR, P. Use of Artificial Intelligence to Assess Human Emotion. The Open Psychology Journal , v. 13, p. 14-16, 2020.

MASSAD E, ORTEGA NRS, BARROS LC, STRUCHINER C. Fuzzy Logic in Action: Applications in Epidemiology and Beyond. *Studies in Fuzziness and Soft Computing*, v. 232. Springer, 2008. DOI: 10.1007/978-3-540-69094-8. ISBN: 978-3-540-69092-4.

MATTE ACF. Taxa de elocução, grupo acentual, pausas e fonoestilística: Temporalidade na prosa e na poesia com interpretação livre. *Estudos Linguísticos*, v. XXXV, p. 276-285, 2006.

MCKEOWN G, VALSTAR M, COWIE R, PANTIC M, SCHRODER M. The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. IEEE Transactions on Affective Computing. V.3(1), p. 5–17, 2012.

MEGRI F, BOUKEZZOULA R. MIN and MAX operators for trapezoidal fuzzy intervals Part II: Proof of analytical expressions. In: 2008 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2008, Hong Kong, China. IEEE. 2008;2339-2344. doi: 10.1109/fuzzy.2008.4630695.

METALLINOU A, SUNGBOK L, NARAYANAN S. Audio-Visual Emotion Recognition using Gaussian Mixture Modelsfor Face and Voic. IEEE Computer society, p. 15-17, 2008.

MICHALSKI, RS, CARBONELL, JG E MITCHELL, TM. MACHINE learning: An artificial intelligence approach. Springer Science; Business Media, 2013.

MILLING M, BAIRD A, BARTL-POKORNY KD, LIU S, ALCORN AM, SHEN J, TAVASSOLI T, AINGER E, PELLICANO E, PANTIC M, CUMMINS N, SCHULLER BW. Evaluating the impact of voice activity detection on speech emotion recognition for autistic children. *Front Comput Sci.* 2022;4. doi: 10.3389/fcomp.2022.837269.

MING Z, FENG Z, ZHENGBIAO J. A study on speech emotion recognition based on fuzzy K nearest neighbor. *Int J Multimedia Ubiquitous Eng.* 2015;10(10):57-66.

MINGUILLON, Jesus; LOPEZ-GORDO, M. Angel; PELAYO, Francisco. Detection of attention in multi-talker scenarios: A fuzzy approach. Expert Systems with Applications, v. 64, p. 261–268, 2016.

MIRSAMADI, S.; BARSOUM, E.; ZHANG, C. Automatic speech emotion recognition using recurrent neural networks with local attention. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2227-223*,* 2017.

MONTEIRO, G F P. Será que o smartphone é uma boa estratégia de coleta de voz de forma remota. Iniciação Científica - Pró-Reitoria de Pesquisa. Universidade Federal da Paraíba, João Pessoa, 2021.

MONTICELLI PF, OTTA E. Acoustic Communication. São Paulo: USP, 2021. DOI: 10.11606/9786587596198. ISBN: 978-65-87596-19-8.

MOON K, CHUNG S, PARK H, KIM H. Materials of acoustic analysis: Sustained vowel versus sentence. *J Voice.* 2012;26(5):563-565. doi: 10.1016/j.jvoice.2011.09.007.

MORAES R, MACHADO L, FERREIRA J. GitHub Leapig-UFPB. [online]. Disponível em: https://github.com/leapigufpb/FuzzyRules. Acesso em: 17 out. 2024.

MORAES R.M., MELO A.C.O. Sistemas de Suporte à Decisão Espacial e Aplicações. Comunicações em Informática, 1(1) (2017), 2-5.

MORAES RM, BANON GJF, SANDRI SA. Fuzzy expert systems architecture for image classification using mathematical morphology operators. *Inf Sci.* 2002;142(1-4):7-21. doi: 10.1016/S0020-0255(02)00132-6.

MORAES RM, MACHADO L. Psychomotor skills assessment in medical training based on virtual reality using a weighted possibilistic approach. *Knowl Based Syst.* 2014;70:97-102.

MORAES RM. Uma arquitetura de sistemas especialistas nebulosos para classificação de imagens utilizando operadores da morfologia matemática. 114 f. Tese (Doutorado) – Pós-graduação em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998.

MORAES, R. M, MACHADO L.S. Gaussian Naive Bayes for Online Training Assessment in Virtual Reality-Based Simulator. Mathware Soft Comput. pp. 123-132, 2009.

MORAES, R. M.; FERREIRA, J. A.; MACHADO, L. S. A New Bayesian Network Based on Gaussian Naive Bayes with Fuzzy Parameters for Training Assessment in Virtual Simulators. International Journal of Fuzzy Systems, v. 23, p. 849-861, 2021.

MORAES, R. M.; SOARES, E. A. M. G.; MACHADO, L. S. A Double Weighted Fuzzy Gamma Naive Bayes Classifier. Journal of Intelligent; Fuzzy Systems, v. 38 n. 1, p. 577-588. ISSN: 1064-1246. 2020. DOI: 10.3233/JIFS-179431.

MORAES, R. M; MACHADO, L. S. Assessment systems for training based on virtual reality: A comparison study. In:. [S.l.]: SBC Journal on 3D Interactive Systems, 2012.

MORAES, RM, SILVA ILA, MACHADO LS. Online Skills Assessment in Training Based on Virtual Reality Using a Novel Fuzzy Triangular Naive Bayes Network. WSPC, 2020.

MORNINGSTAR M, GILBERT A, BURDO J, LEIS M, DIRKS M. Recognition of vocal socioemotional expressions at varying levels of emotional intensity. *Emotion.* 2021. doi: 10.1037/emo0001024.

MUÑETÓN-AYALA M, DE VEGA M, OCHOA-GÓMEZ JF, BELTRÁN D. The brain dynamics of syllable duration and semantic predictability in Spanish. *Brain Sci.* 2022;12(4):458. doi: 10.3390/brainsci12040458.

NADLER B, GALUN M. Fundamental limitations of spectral clustering. In: *Advances in Neural Information Processing Systems.* MIT Press; 2006. p. 1017-1024. doi: 10.7551/mitpress/7503.003.0132.

NENKO AV, LOSKUTOVA OA, BERG YA, BOROVIKOVA DV. Voice Changes as a Result of Psychoemotional Impact. *2021 XV International Scientific-Technical Conference on Actual Problems Of Electronic Instrument Engineering (APEIE)*, Novosibirsk, Russian Federation, 2021, p. 105-110. DOI: 10.1109/APEIE52976.2021.9647583.

NEUMANN M, EDELHÄUSER F, TAUSCHEL D, FISCHER MR, WIRTZ M, WOOPEN C, HARAMATI A, SCHEFFER C. Empathy decline and its reasons: A systematic review of studies with medical students and residents. *Academic Medicine*, v. 86, n. 8, p. 996-1009, Aug. 2011. DOI: 10.1097/ACM.0b013e318221e615.

NOGUCHI C, KAWAMOTO T. Fragility of spectral clustering for networks with an overlapping structure. *Phys Rev Res.* 2020;2(4):043101. doi: 10.1103/physrevresearch.2.043101.

NGUYEN NT, PAPRZYCKI M, VOSSEN G. Intelligent Information and Database Systems: Proceedings of the 16th Asian Conference on Intelligent Information and Database Systems. Singapore: Springer Nature Singapore; 2024.

OH C, MORRIS R, WANG X, RASKIN MS. Analysis of emotional prosody as a tool for differential diagnosis of cognitive impairments: A pilot research. *Frontiers in Psychology*, v. 14, 2023.

ORTEGA NRS. Aplicação da teoria de conjuntos Fuzzy a problemas da biomedicina. 2001. Tese (Doutorado em Física) - Instituto de Física, Universidade de São Paulo, São Paulo, 2001. DOI: 10.11606/T.43.2001.tde-04122013-133237. Acesso em: 2025-01-09.

PAOLIELLO K, OLIVEIRA, G. BEHLAU M. Desvantagem vocal no canto mapeado por diferentes protocolos de autoavaliação. CoDas, v. 1, 2012.

PAREEK S, GUPTA H, KAUR J, KUMAR R, CHOHAN JS. Fuzzy Logic in Computer Technology: Applications and Advancements. *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, p. 1634-1637, 2023. DOI: 10.1109/ICPCSN58827.2023.00273.

PARTILA, P.; VOZNAK, M.; TOVAREK, J. Pattern Recognition Methods and Features Selection for Speech Emotion Recognition System. Scientific World Journal, v., 2015.

PASSALACQUA SA, SEGRIN C. The effect of resident physician stress, burnout, and empathy on patient-centered communication during the long-call shift. *Health Communication*, v. 27, n. 5, p. 449–456, 2012. DOI: 10.1080/10410236.2011.606527.

PASSINO KM, YURKOVICH S. Fuzzy control [Internet]. Menlo Park: Addison-Wesley; 1998

PATEL S; SCHERER KR; BJORKNER E; SUNDBERG J. Mapping emotions into acoustic space: The role of voice production. Biological psychology. 87(1):93-8, 2011.

PATNAIK S. Speech emotion recognition by using complex MFCC and deep sequential model. *Multimed Tools Appl.* 2023;82:11897–11922. doi: 10.1007/s11042-022-13725-y.

PEDRYCZ W, GOMIDE F. An introduction to fuzzy sets: Analysis and design. MIT Press; 1998.

PELL MD, KOTZ SA. On the time course of vocal emotion recognition. *PLoS One.* 2011;6(11):e27256. doi: 10.1371/journal.pone.0027256.

PEREIRA C, WATSON C. Some acoustic characteristics of emotion. In: *5th International Conference on Spoken Language Processing (ICSLP)*. 1998. doi: 10.21437/ICSLP.1998-148.

PERVAIZ M, KHAN TA. Emotion recognition from speech using prosodic and linguistic features. *Int J Adv Comput Sci Appl (IJACSA).* 2016;7(8):84-89.

PINHEIRO MMC, VIEIRA MG, VIEIRA LM, KOERICH I, ROSSETO I, LAZZAROTTO-VOLCÃO C, PAUL S. Updating sentences lists for assessment speech perception. *CoDAS.* 2022;34(1):e20200301. doi: 10.1590/2317-1782/20202020301.

RAMONIA, M.; SEBASTIANI, P. Robust bayes classifiers. In: Artificial Intelligence. [S.l.: s.n.], 2001. p. 209–226

RAVI V, ZIMMERMANN H. Fuzzy rule based classification with FeatureSelector and modified threshold accepting. *European Journal of Operational Research*, v. 123, p. 16-28, 2000.

RINGEVAL F, AMIRIPARIAN S, EYBEN F, SCHERER K, SCHULLER B. Emotion recognition in the wild: Incorporating voice and lip activity in multimodal decision-level fusion. *Proceedings of the 16th International Conference on Multimodal Interaction.* 2014. doi: 10.1145/2663204.2666271.

RINGEVAL F, SONDEREGGER A, SAUER J, LALANNE D. Introducing the recola multimodal corpus of remote collaborative and affective interactions. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–8, 2013.

ROSS, T. J., Fuzzy logic with engineering applications, West Sussex, England: John Wiley; Sons Ltd, 2010.

ROVETTA S, MNASRI Z, MASULLI F, CABRI A. Emotion recognition from speech: An unsupervised learning approach. *Int J Comput Intell Syst.* 2020;14(1). doi: 10.2991/ijcis.d.201019.002.

ROVETTA S, MNASRI Z, MASULLI F, CABRI A. Emotion recognition from speech signal using fuzzy clustering. *EUSFLAT.* 2019;120-127. doi: 10.2991/EUSFLAT-19.2019.19.

RUSSELL JA. A circumplex model of affect. *Journal of Personality and Social Psychology*, v. 39, n. 6, p. 1161, 1980.

SAADE JJ, DIAB H. Defuzzification methods and new techniques for fuzzy controllers. *Iran J Electr Comput Eng.* 2003;3(2):161-174.

SAKURAI M, KOSAKA T. Emotion recognition combining acoustic and linguistic features based on speech recognition results. In: *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*. IEEE; 2021. doi: 10.1109/GCCE53005.2021.9621810.

SALHI L, MOURAD T, CHERIF A. Voice disorders identification using multilayer neural network. Int Arab J Inf Technol. 7:177 185, 2010.

SANDAGE MJ, PLEXICO LW, SCHIWITZ A. Clinical utility of CAPE-V sentences for determination of speaking fundamental frequency. *Journal of Voice.* 2015;29(4):441-445.

SANDHYA, P. et al. Spectral features for emotional speaker recognition. In: 2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC). [S.l.: s.n.], p. 1–6, 2020.

SANTOS AJ, ROTHE-NEVES R, PACHECO V, BALDOW VS. Emotional speech prosody: How readers of different educational levels process pragmatic aspects of reading aloud. *DELTA*. 2022;38(3).

SANTOS FJJ. Sistemas de apoio à decisão em grupo multicritério: uma abordagem baseada em regras fuzzy. 2009. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de São Carlos, Centro de Ciências Exatas e de Tecnologia, Programa de Pós-Graduação em Ciência da Computação, São Carlos, 2009.

SCHERER KR. Vocal Communication of Emotion: A Review of Research Paradigms. Speech Communication.V.40(1), p 227–56, 2003.

SCHLOSBERG H. “Three Dimensions of Emotion,” Psychological Review. V.61(2):81,1954.

SEAL D, ROY UK, BASAK R. Sentence-level emotion detection from text based on semantic rules. In: Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018. Advances in Intelligent Systems and Computing. Springer. 2019;423-430. doi: 10.1007/978-981-13-7166-0_42.

SHARIF A, SITOMPUL OS, NABABAN EB. Analysis of variation in the number of MFCC features in contrast to LSTM in the classification of English accent sounds. *J Inform Telecommun Eng.* 2023;6(2):587-601. doi: 10.31289/jite.v6i2.8566.

SHARMA S. Emotion Recognition *from* Speech using Artificial Neural Networks and Recurrent Neural Networks*.* 11th International Conference on Cloud Computing, Data Science; Engineering (Confluence), p. 153-158, 2021.

SHASHANK P, GUPTA H, KAUR J, et al. Fuzzy logic in computer technology: Applications and advancements. *IEEE.* 2023. doi: 10.1109/icpcsn58827.2023.00273.

SILVA EF. A voz dentro da relação psíquico-orgânica: estudo sobre a influência das emoções na voz do ator. Rev Cient /FAP. 4(1), p.1-19, 2009.

SILVA JUNIOR LJ, BARBOSA PA. Speech rhythm of English as L2: An investigation of prosodic variables on the production of Brazilian Portuguese speakers. *J Speech Sci.* 2020;8(2):37-57.

SILVA W, BARBOSA PA. Perception of emotional prosody: Investigating the relation between the discrete and dimensional approaches to emotions. *Rev Estud Linguagem.* 2017;25(3):1075-1102.

SILVER D, HUANG A, MADDISON CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 529,p.484–489, 2016.

SIM J, WRIGHT CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 85(3) p.257-68, 2005.

SINGH, YB, GOEL S. A systematic literature review of speech emotion recognition approaches. Neurocomputing 492, p. 245–263, 2022.

SNEDDON I, MCRORIE M, MCKEOWN G, HANRATTY J. The Belfast Induced Natural Emotion Database. *IEEE Transactions on Affective Computing*, v. 3, n. 1, p. 32-41, 2012. DOI: 10.1109/T-AFFC.2011.26.

SOBIN C, ALPERT M. Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, New York, v. 28, n. 4, p. 347-365, 1999.

SONG H., KANG K., WEI Y., ZHANG H., ZHANG L. Speech Emotion Recognition based on Multi COATTENTION Acoustic Feature Fusion. 2024 doi: 10.1109/icdsis61070.2024.10594517

SOUZA OC, HANAYAMA EM. Fatores psicológicos associados a disfonia funcional e a nódulos vocais em adultos. *Rev CEFAC.*, 2005; 7(3):388-97.

SPEED LJ, BRYSBAERT M. Ratings of valence, arousal, happiness, anger, fear, sadness, disgust, and surprise for 24,000 Dutch words. *Behav Res Methods.* 2023. doi: 10.3758/s13428-023-02239-6.

STORY, M. AND CONGALTON, R.G. Accuracy Assessment: A User's Perspective. Photogrammetric Engineering and Remote Sensing, 52, 397–399, 1986.

SUNDBERG J, SALOMÃO G, SCHERER K. Analyzing emotion expression in singing via flow glottograms, long-term-average spectra, and expert listener evaluation. *J Voice.* 2019. doi: 10.1016/j.jvoice.2019.08.007.

SUNDBERG J. A ciência da voz. Editora da Universidade de São Paulo, 2015.

SUNNY, S., DAVID PETER, S.; JACOB, K.P. Combined feature extraction techniques and Naive Bayes classifier for speech recognition. Computer Science; Information Technology (CS; IT), pp. 155– 163, 2013.

SWAIN M, ROUTRAY A, KABISATPATHY P. Databases, features and classifiers for speech emotion recognition: A review. *Int J Speech Technol.* 2018;21(1):93-120. doi: 10.1007/s10772-018-9491-z.

Takagi, T. e Sugeno M. (1983) Derivation of fuzzy control rules from human operator's control action. IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis, Marseille, p. 55- 60.

TALELE M, JAIN R. Complex facial emotion recognition: A systematic literature review. In: *Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*. IEEE; 2023. doi: 10.1109/ICAECT57570.2023.10117836.

TANDEL, N. H.; PRAJAPATI, H. B.; DABHI, V. K. Voice Recognition and Voice Comparison using Machine Learning Techniques: A Survey. 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020, p. 459–465, 2020.

TATO, R. et al. Emotional space improves emotion recognition. In: INTERSPEECH. S.1.: s.n.J, 2002.

TEIXEIRA, JP; FERREIRA, D; CARNEIRO, S. Análise acústica vocal - determinação do Jitter e Shimmer para diagnóstico de patologias da fala. In: CONGRESSO LUSO-MOÇAMBICANO DE ENGENHARIA, 6., 2011, Maputo, Moçambique. Anais [...]. Maputo: [s.n.], 2011.

TIGUE CC, BORAK DJ, O'CONNOR JJM, SCHANDL C, FEINBERG DR. Voice pitch influences voting behaviour. *Evolution and Human Behavior*, v. 33, p. 210–216, 2012.

TON-THAT, An H.; CAO, Nhan T. Speech emotion recognition using a fuzzy approach. Journal of Intelligent; Fuzzy Systems, v. 36, n. 2, p. 1587-1597, 2019.

TORRES NETO JR, FILHO GPR, MANO LY, UEYAMA J. VERBO: Voice Emotion Recognition database in Portuguese language. *Journal of Computer Science*, v. 14, n. 11, p. 1420-1430, nov. 2018. DOI: 10.3844/jcssp.2018.1420.1430.

TRACEY B, PATEL S, ZHANG Y, et al. Voice biomarkers of recovery from acute respiratory illness. *IEEE J Biomed Health Inform.* 2021;26:2787-2795. doi: 10.1109/JBHI.2021.3137050.

TRACEY B, VOLFSON D, GLASS JR, HAULCY R. Towards interpretable speech biomarkers: exploring MFCCs. *Scientific Reports*, Springer Nature, v. 13, n. 1, dez. 2023. DOI: 10.1038/s41598-023-49352-2. Disponível em: https://doi.org/10.1038/s41598-023-49352-2.

TRAJANO FMP, ALMEIDA LNA, ARAUJO RA, CRISOSTOMO FLS, ALMEIDA AAF. Níveis de ansiedade e impactos na voz: uma revisão da literatura. *Distúrbios da Comunicação*, v. 28, p. 423-433, 2016.

TRINH VAN, L. et al. Emotional Speech Recognition Using Deep Neural Networks. Sensors (Basel, Switzerland), v. 22, n. 4, 2022.

TZIRAKIS, P.; ZHANG, J.; SCHULLER, W. End-to-end speech emotion recognition using deep neural networks. Department of Computing, Imperial College London, London, UK Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany. Icassp 2018, p. 5089–5093, 2018.

UMAPATHY K, KRISHNAN S, PARSA V, et al. Discrimination of pathological voices using a time-frequency approach. *IEEE Trans Biomed Eng*. 52:421–430, 2005.

VAN STRALEN KJ, STEL VS, REITSMA JB, et al. Diagnostic methods I: Sensitivity, specificity, and other measures of accuracy. *Kidney Int.* 2009;75(12):1257-1263.

VASHISHTHA S, GUPTA V, MITTAL M. Sentiment analysis using fuzzy logic: A comprehensive literature review. *Wiley Interdiscip Rev Data Min Knowl Discov.* 2023;13(5). doi: 10.1002/widm.1509.

VERVERIDIS D, KOTROPOULOS C. Emotional speech recognition: Resources, features, and methods. Speech Communication, V.48, p. 1162–1181, 2006.

VIEIRA VJD. Análise de Variações Acústicas Não Estacionárias e seu Efeito na Detecção de Múltiplas Emoções e Condições de Estresse. [Tese em Engenharia Elétrica]. Campina Grande/PB: Universidade Federal de Campina Grande, 2018.

WAARAMAA T, LAUKKANEN A, AIRAS M, ALKU P. Perception of emotional valences and activity levels from vowel segments of continuous speech. *J Voice.* 2010;24(1):30-38. doi: 10.1016/j.jvoice.2008.04.004.

WANG L, MINAMI K, YAMAMOTO K, NAKAGAWA S. Speaker identification by combining MFCC and phase information in noisy environments. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, p. 4502-4505, 2010. DOI: 10.1109/ICASSP.2010.5495586.

WANG Y, ZHU Z, CHEN B, FANG F. Perceptual learning and recognition confusion reveal the underlying relationships among the six basic emotions. Cogn Emot. 33(4):754-767, 2019.

WANG, J., JO, C. Vocal folds disorder detection using pattern recognition methods. Annu Int Conf IEEE Eng Med Biol Soc. 2007:3253-6, 2007.

WEBER L, KLEIN PAT. Aplicação da lógica fuzzy em software e hardware. Canoas: Ed. ULBRA; 2003. 112p.

WEN G, ZHU Y, CHEN L, ZHAN M, XIE Y. Global and local structure preservation for nonlinear high-dimensional spectral clustering. *Comput J.* 2021;64:993-1004. doi: 10.1093/comjnl/bxab020.

WEST CP, TAN AD, HABERMANN TM, SLOAN JA, SHANAFELT TD. Association of resident fatigue and distress with perceived medical errors. *JAMA*, v. 302, p. 1294-1300, 2009.

XIAO, Z, DELLANDREA, E, DOU, W, et al. Hierarchical classification of emotional speech. IEEE Transactions on Multimedia, 2007.

XIONG, Yu; CAI, Ting; ZHONG, Xin; ZHOU, Song; CAI, Linqin. Fuzzy speech emotion recognition considering semantic awareness. Journal of Intelligent; Fuzzy Systems, v. 46, n. 3, p. 7367-7377, 2024.

XU D. Applications of Fuzzy Logic in Bioinformatics. Series on Advances in Bioinformatics and Computational Biology, v. 9. London: Imperial College Press, 2008. 225 p. ISBN: 1848162588.

YEH PW, GEANGU E, REID V. Coherent emotional perception from body expressions and the voice. Neuropsychologia. 91:99-108, 2016.

YILDIRIM S, BULUT M, LEE CM, et al. An acoustic study of emotions expressed in speech. In: *Eighth International Conference on Spoken Language Processing.* 2004.

YU, K, KOHANE, IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 28.3, 238-241, 2019.

ZADEH LA. Fuzzy Logic. *IEEE Computer*, Apr. 1988, p. 83-92.

ZADEH LA. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1978.

ZADEH LA. Fuzzy sets. Inf Control. 1965;8(3):338–53. Disponível em: http://dx.doi.org/10.1016/S0019- 9958(65)90241-X.

ZADEH LA. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, p. 28-44, 1973. DOI: 10.1109/TSMC.1973.5408575.

ZADEH LA. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning-1. *Information Sciences*, v. 8, p. 199-249, 1975. DOI: 10.1016/0020-0255(75)90036-5.

ZHANG B, ESSL G, PROVOST EM. Recognizing emotion from singing and speaking using shared models. *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Xi'an, China, p. 139-145, 2015. DOI: 10.1109/ACII.2015.7344563.

ZHANG Y, HAN F. An improved ensemble classification algorithm for imbalanced data with sample overlap. *Springer.* 2021;454-468. doi: 10.1007/978-981-19-6135-9_34.

ZHAO Q, FAN HZ, LI YL, LIU L, WU YX, ZHAO YL, TIAN ZX, WANG ZR, TAN YL, TAN SP. Vocal Acoustic Features as Potential Biomarkers for Identifying/Diagnosing Depression: A Cross-Sectional Study. *Frontiers in Psychiatry*, v. 13, p. 815678, 28 abr. 2022. DOI: 10.3389/fpsyt.2022.815678. PMID: 35573349; PMCID: PMC9095973.

ZHU W, NIE F, LI X. Fast spectral clustering with efficient large graph construction. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE; 2017. p. 2492-2496. doi: 10.1109/ICASSP.2017.7952605.

ZHU, J., WU, D. Application of New Artificial Intelligence Technology in the Voice Recognition and Analysis System of Electric Power Information Customer Service. 2020 International Conference on Computers, Information Processing and Advanced Education (CIPAE 2020), October 16-18, 2020.

W. van Leekwijck and E. E. Kerre, “Defuzzification: criteria and classification,” Fuzzy Sets and Systems, vol. 108, no. 2, pp. 159–178, Dec. 1999, doi: 10.1016/S0165-0114(97)00337-0

M. Sugeno, “An introductory survey of fuzzy control,” Information Sciences, vol. 36, no. 1–2, pp. 59–83, Jul. 1985, doi: 10.1016/0020-0255(85)90026-X

ANEXOS

ANEXO 1 - PARECER CONSUBSTANCIADO DO CEP

UFPB - CENTRO DE CIÊNCIAS DA SAÚDE DA UNIVERSIDADE FEDERAL DA PARAÍBA

Plataforma Brasil

Continuação do Parecer: 3.304.419

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1306092.pdf	03/04/2019 23:31:56		Aceito
Outros	Carta_resposta_CEP_UFPB.pdf	03/04/2019 23:31:25	Anna Alice Figueiredo de Almeida	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE2.pdf	03/04/2019 23:28:03	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Pesquisadores	Atestado_MBehlau.pdf	03/04/2019 23:27:46	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Instituição e Infraestrutura	Autoriza_UNIFESP.pdf	03/04/2019 23:26:42	Anna Alice Figueiredo de Almeida	Aceito
Declaração de Instituição e Infraestrutura	Autoriza_DepFonoUFPB.pdf	03/04/2019 23:25:29	Anna Alice Figueiredo de Almeida	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_Universal_Banco_de_vozes.pdf	01/03/2019 22:57:48	Anna Alice Figueiredo de Almeida	Aceito
Declaração do Patrocinador	Projeto_Universal2018_ContratoCNPq.pdf	01/03/2019 22:57:22	Anna Alice Figueiredo de Almeida	Aceito
Folha de Rosto	folha_de_rosto_CEP_assinada.pdf	01/03/2019 22:56:54	Anna Alice Figueiredo de Almeida	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

JOAO PESSOA, 06 de Maio de 2019

Assinado por:
Eliane Marques Duarte de Sousa
(Coordenador(a))