

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

Campus Ilha Solteira

EDER PEREIRA NEVES

**MODELO PREDITIVO PESQ-ANFIS/FUZZY
C-MEANS PARA AVALIAÇÃO DE SINAIS DE VOZ
BASEADO EM IMAGENS**

Ilha Solteira

2022

EDER PEREIRA NEVES

**MODELO PREDITIVO PESQ-ANFIS/FUZZY
C-MEANS PARA AVALIAÇÃO DE SINAIS DE VOZ
BASEADO EM IMAGENS**

Tese apresentada à Faculdade de Engenharia de
Ilha Solteira - UNESP, como parte dos requisitos
para obtenção do título de Doutor em Engenharia
Elétrica.

Área de Conhecimento: Automação.

Orientador: Dr. Jozue Vieira Filho.

Ilha Solteira

2022

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

Neves, Eder Pereira.

N518m Modelo preditivo PESQ-ANFIS/Fuzzy C-means para avaliação de sinais de voz baseado em imagens / Eder Pereira Neves. -- Ilha Solteira: [s.n.], 2022
144 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2022

Orientador: Jozue Vieira Filho
Inclui bibliografia

1. Imagem perceptual do sinal de voz. 2. Técnica de extração de fatores. 3. Algoritmo de fuzzificação. 4. Estimação da PESQ Pela ANFIS.


Raiane da Silva Santos

Supervisora Técnica de Seção
Seção Técnica de Referência, Atendimento ao usuário e Documentação
Diretoria Técnica de Biblioteca e Documentação
CRB/8 - 9999

CERTIFICADO DE APROVAÇÃO

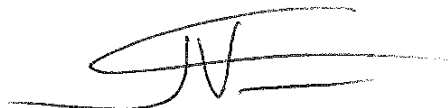
TÍTULO DA TESE: MODELO PREDITIVO PESQ-ANFIS/FUZZY C-MEANS PARA AVALIAÇÃO DE SINAIS DE VOZ BASEADO EM IMAGENS

AUTOR: EDER PEREIRA NEVES

ORIENTADOR: JOZUE VIEIRA FILHO

Aprovado como parte das exigências para obtenção do Título de Doutor em ENGENHARIA ELÉTRICA, área: Automação pela Comissão Examinadora:

Prof. Dr. JOZUE VIEIRA FILHO (Participação Presencial)
Coordenadoria Executiva / Câmpus Experimental de São João da Boa Vista - UNESP



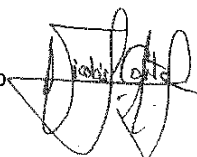
Prof^a. Dr^a. ANNA DIVA PLASENCIA LOTUFO (Participação Presencial)
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira - UNESP



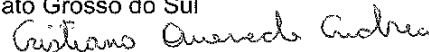
Prof^a. Dr^a. MARA LUCIA MARTINS LOPES (Participação Presencial)
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira - UNESP



Prof. Dr. NICOLAS EUSEBIO CORTEZ LEDESMA (Participação Virtual)
Faculdade de Arquitetura, Engenharia e Tecnologia / Universidade Federal de Mato Grosso



Prof. Dr. CRISTIANO QUEVEDO ANDREA (Participação Presencial)
Centro de Ciências Exatas e Tecnologia / Universidade Federal de Mato Grosso do Sul



Ilha Solteira, 01 de dezembro de 2022

Dedico este trabalho a minha família, minha esposa Irene e as minhas filhas Mariana, Laura e Rafaela.

AGRADECIMENTOS

À Deus por me acompanhar nesta jornada;

Todos meus familiares e em especial a minha esposa Irene que juntamente com nossas filhas Mariana, Laura e Rafaela puderem proporcionar momentos felizes nesta caminhada;

À Universidade Estadual de Mato Grosso do Sul - UEMS e a todos os professores e técnicos administrativos que contribuíram com este processo de formação;

Todos alunos, professores e técnicos administrativos do curso de Matemática da UEMS/Unidade Universitária de Cassilândia pelo apoio;

In memoriam ao querido professor e orientador Francisco Villarreal Alvarado pelas suas contribuições no desenvolvimento desta pesquisa;

Ao orientador Prof. Jozue Vieira Filho que nos atendeu prontamente e não mediu esforços para conclusão deste trabalho;

Ao professor, orientador e amigo Marco Aparecido Queiroz Duarte pela oportunidade e confiança;

À UNESP, em especial a todos os professores e técnicos administrativos do PPGEE-FEIS.

“Já sei olhar o rio por onde a vida passa”
(Ana Carolina/ Totonho Villeroy)

RESUMO

Neste trabalho apresenta-se um método para fazer a predição da pontuação MOS - *Mean Opinion Score* para a medida objetiva intrusiva, ITU-T P.862 - PESQ utilizando uma técnica de processamento de imagem do espectro do sinal de voz. O método foi construído por meio de um processo sistemático simulado por sinais limpos e contaminados com diferentes tipos de ruídos encontrados em situações cotidianas. Cada sinal é convertido para o domínio da frequência, transformado em uma matriz e as bandas críticas são separadas utilizando um modelo perceptual. Em seguida, a imagem gerada é quantificada e o espectro de potência de cada pixel é convertido em um nível de cinza e analisado por meio de uma técnica baseada na matriz de coocorrência de níveis de cinza nas suas quatro principais direções, gerando os 19 fatores de Haralick. Dentre o conjunto de fatores três foram selecionados para constituir as entradas do modelo, proporcionando 969 composições para cada direção da matriz de coocorrência. A proposta utiliza uma abordagem de inferência do tipo Sugeno de primeira ordem, implementada no Sistema de Inferência Neuro *Fuzzy* Adaptativa - ANFIS. Comprovou-se que o desempenho preditivo está ligado à escolha de um algoritmo de fuzzificação do espaço de entrada. Desta forma, foi analisado o comportamento do particionamento *Grid* e os agrupamentos *Subtractive* e *Fuzzy C-Means* - FCM. Constatou-se que o algoritmo FCM proporcionou os melhores resultados para três variáveis linguísticas, sendo avaliado pelas métricas MAPE, RMSE e R^2 . Nas simulações ficou evidente que o modelo atua efetivamente em qualquer ambiente ruidoso, proporcionando resultados satisfatórios independente da quantidade de sinais a serem analisados.

Palavras chave: Imagem perceptual do sinal de voz. Técnica de Extração de Fatores. Algoritmo de Fuzzificação. Estimação da PESQ pela ANFIS.

ABSTRACT

In this work, it is proposed a method to predict of the MOS score - Mean Opinion Score from the intrusive objective measure PESQ - ITU P.862, by analysing the spectrum image of the voice signal. The colorblackmethod was constructed by means of a systematic process simulated by clean signals contaminated with different types of noise found in everyday situations. Each signal is converted to the frequency domain, transformed into a matrix, and the critical bands are separated using a perceptual model. Then the generated image is quantized and the power spectrum of each pixel is converted to a gray level and analyzed using a technique based on the co-occurrence matrix of gray levels in its four main directions, generating the 19 Haralick factors. Among the set of factors, three were selected to constitute the model inputs, which provided 969 compositions for each direction of the co-occurrence matrix. The proposal uses a first-order Sugeno fuzzy inference approach, implemented in the Adaptive Neuro Fuzzy Inference System - ANFIS. Predictive performance has been shown to be linked to the choice of a fuzzification algorithm for the input space. Thus, the behavior of the Grid partitioning and the Subtractive and Fuzzy C-Means - FCM clustering was analyzed. The FCM algorithm was found to provide the best results for three linguistic variables, as evaluated by the MAPE, RMSE and R^2 metrics. In the simulations it became evident that the model works effectively in any noisy environment, providing satisfactory results regardless of the amount of signals to be analyzed.

Key words: Perceptual imaging of the speech signal. Factor Extraction Technique. Fuzzification Algorithm. Estimating PESQ by ANFIS.

LISTA DE FIGURAS

1	Diagrama dos métodos de avaliação: subjetiva e dos métodos objetivos não intrusivos e intrusivos.	36
2	Arquitetura do método intrusivo PESQ.	38
3	Estrutura do ouvido humano.	44
4	Curvas de audibilidade ou Fletcher–Munson.	48
5	Ouvido interno.	49
6	Estrutura coclear no formato linear.	51
7	Bandas críticas ideais para o ouvido humano: a) espectro do sinal de voz na frequência e b) distribuição do sinal em 15 bandas críticas na escala Bark. .	52
8	Ilustração do efeito do mascaramento simultâneo ou em frequência	54
9	Mascaramento temporal.	54
10	Mascaramento simultâneo.	55
11	Combinação de mascaramentos em tempo e frequência.	55
12	Bandas críticas dos modelos: a) Hermansky e b) MFCC.	58
13	Gráfico da função de propagação de banda crítica, com base na equação (17). .	59
14	Fluxograma do modelo perceptual de Hermansky	63
15	Forma de onda do sinal de voz: (a) sinal limpo e (e) sinal contaminado por ruído de multidão com 0 dB; escala espectral do modelo Hermansky: (b) sinal limpo e (f) sinal ruidoso; espectrogramas: (c) sinal limpo e (g) sinal ruidoso; plano quadro $\times$ bandas Bark resposta do modelo Hermansky: (d) sinal limpo e (h) sinal ruidoso.	64

16	Fluxograma do modelo perceptual MFCC.	65
17	Forma de onda do sinal de voz: (a) sinal limpo e (e) sinal contaminado por ruído de multidão com 0 dB; escala espectral do modelo MFCC: (b) para sinal limpo e (f) sinal ruidoso; espectrogramas: (c) sinal limpo e (g) sinal ruidoso; plano quatro $\times$ bandas Mel resposta do modelo MFCC: (d) sinal limpo e (h) sinal ruidoso.	66
18	Diagrama de blocos da extração de fatores.	68
19	Orientações da matriz de coocorrência: a) por células e b) indicam as principais direções.	72
20	a) Uma imagem quantizada e b) Suas matrizes de coocorrência nas suas principais direções	74
21	Arquitetura ANFIS.	79
22	Partição <i>Grid</i>	88
23	Resultados do algoritmo FCM para 90 sinais para os fatores 1, 11 e 13 de Haralick com suas respectivas avaliações PESQ sendo: a), c) e e) três grupos com seus respectivos centros e b), d) e f) funções de pertinência com três variáveis linguísticas.	98
24	Resultados do algoritmo SB para 90 sinais para os fatores 1, 11 e 13 de Haralick com suas respectivas avaliações PESQ sendo: a), c) e e) três grupos com seus respectivos centros e b), d) e f) funções de pertinência com três variáveis linguísticas.	101
25	Diagrama de fluxo do método PESQ-ANFIS/FCM.	102
26	Plano resposta do modelo perceptivo de (a) Hermansky e (b) MFCC para o diretório de sinais com ruído de trem.	105
27	Quantização linear para sinal limpo da base de dados NOIZEUS.	106
28	Quantização linear para sinal contaminado com ruído de multidão com SNR 0 dB da base de dados NOIZEUS.	107

29	(a) Imagem quantizada com 8 níveis de cinza para um sinal limpo da base de dados NOIZEUS, (b) MCNC do mesmo sinal na direção 90° e (c) MCNC normalizada pela equação (35).	108
30	(a) Imagem quantizada com 8 níveis de cinza para um sinal com ruído de multidão com SNR = 0dB da base de dados NOIZEUS, (b) MCNC do mesmo sinal na direção 90° e (c) MCNC normalizada pela equação (35).	108
31	Resultados do MAPE para a divisão percentual de treinamento e teste para o diretório de sinais da base de dados NOIZEUS com ruído de Multidão	111
32	Diagrama de fluxo do algoritmo de fuzzificação FCM/ANFIS.	114
33	Resultados da medida MAPE para as fases de treinamento (azul) e teste (vermelho) nas quatro direções da matriz de coocorrência com ruído de multidão para os algoritmos de fuzzificação: a) FCM, b) <i>Subtractive</i> e c) <i>Grid</i>	116
34	Função de pertinência com três variáveis linguísticas para os algoritmos: a) FCM, b) <i>Subtractive</i> e c) <i>Grid</i>	116
35	Comportamento regressor do modelo na direção 90° para sinais contaminados com ruído de multidão e com os seguintes algoritmos de fuzzificação: a) FCM, b) <i>Subtractive</i> e c) <i>Grid</i>	117
36	<i>Boxplots</i> dos resultados das combinações dos fatores de Haralick para áudios com ruídos de carro, multidão e estação sendo: a), c) e e) dos melhores MAPE e b), d) e f) piores MAPE.	124
37	Regressão dos fatores	126
38	Resultados dos cinquenta valores MAPE representados pelos <i>boxplots</i> obtidos pela ANFIS/FCM para a base de dados TIMIT/NOISEX com ruído de carro.	127
39	Resultados do MAPE para ANFIS/FCM e a rede MLP com algoritmo Levenberg-Marquardt para os 8 sinais da base de dados NOIZEUS.	129

LISTA DE TABELAS

1	Pontuação da recomendação ITU-T P.800.	37
2	Frequências centrais e limítrofes das bandas críticas	52
3	Os fatores de textura de Haralick calculados a partir de MCNC.	77
4	Centros dos grupos mostrados nas figuras 23a, 23c e 23e.	97
5	Centros dos grupos mostrado nas figuras 24a, 24c e 24e.	100
6	Resultados dos 19 fatores de Haralick para sinais limpos e contaminados com ruído de multidão da base de dados NOIZEUS.	110
7	Média (Md) e Desvio Padrão (DP) das fases de treinamento e teste para os 150 sinais da base de dados NOIZEUS com ruído de carro em quatro direções da matriz de coocorrência.	118
8	Média e Desvio Padrão da ANFIS para as fases de treinamento e teste com os sinais da base de dados NOIZEUS com os seus respectivos ruídos na direção 90° da matriz de coocorrência para os algoritmos de fuzzificação FCM, <i>Subtractive</i> e <i>Grid</i>	119
9	Número de Regras - NR - empregadas para cada variável linguística e Método de Fuzzificação - MF - utilizando diretório com ruído de multidão.	120
10	Tempo de execução do programa na fase de teste para 60 sinais dos diretórios carro e multidão da base de dados NOIZEUS.	121
11	Diferença entre os resultados da PESQ e ANFIS/FCM na fase de teste para 60 sinais com ruído de multidão.	122
12	Medidas MAPE, RMSE e R^2 dos fatores.	126

13	Resultados do desempenho entre a ANFIS e as redes MLP para o diretório com ruído de multidão da base de dados NOIZEUS.	128
14	Resultados do desempenho da ANFIS/FCM para os modelos perceptivos de Hermansky e MFCC das fases de treinamento e teste para os oito diretórios de sinais da base de dados NOIZEUS na direção $[-1\ 0]$ e/ou 90° da matriz de coocorrência.	130

LISTA DE ABREVIATURAS E SIGLAS

ACR	<i>Absolute Category Rating</i>
ANFIS	Sistema de Inferência Neuro <i>Fuzzy</i> Adaptativa
Bs	Matriz percepção auditiva modelo de Hermansky
BFGS	<i>Broyden-Fletcher-Goldfarb-Shanno Quasi-Newton</i>
BR	<i>Bayesian Regularization</i>
CGB	Conjugate Gradient with Powell/Beale Restarts
CGF	Fletcher-Powell Conjugate Gradient
CGP	<i>Polak-Ribière Conjugate Gradient</i>
CCR	<i>Comparison Category Rating</i>
CNN	Convolutional Neural Network
CODEC	Codificadores/Decodificadores
dB	Decibel
DE	Distorção Espectral
DC	Distância Cepstral
DCR	<i>Degradation Category Rating</i>
DCT	Transformada Discreta de Cosseno
DFT	Transformada discreta de Fourier
DL	<i>Deep Learning</i>
DMC	Direção Matriz Coocorrências
DMCNC	Direção Matriz Coocorrência Níveis Cinza
DP	Desvio Padrão
M/F	Método de fuzzificação
FB	Voz femininas com ruídos de multidão
FTLM	<i>Feedforward</i> com o algoritmo de treinamento Levenberg-Marquardt
FCM	<i>Fuzzy C-Means</i>
FFT	<i>Fast Fourier Transform</i>
FIS	Sistema de Inferência <i>Fuzzy</i>

FL	Sinais limpos de vozes femininas
Fs	Frequência Amostragem
GB	Grau de pertinência
GDM	Gradient Descent with Momentum
GDX	Variable Learning Rate Gradient Descent
HCM	<i>Hard C-Mean</i>
Hz	Hertz
IS	Itakura-Saito
ITU-T	<i>Telecommunication Standardization Sector</i>
LES	<i>Listening Effort Scale</i>
LF	Lógica <i>fuzzy</i>
FLM	<i>Levenberg-Marquardt</i>
FRB	<i>Resilient Backpropagation</i>
FGD	<i>Gradient Descent</i>
LLR	<i>Log Likelihood Ratio</i>
LPC	Codificação Linear de Previsão
LPS	<i>Loudness Preference Scale</i>
LQO	<i>Listening Quality Objective</i>
LQS	<i>Listening Quality Subjective</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MB	Voz masculinas com ruídos de multidão
MCNC	Cálculo Matriz Coocorrência Nível Cinza
MD	Média
MF	Método de fuzzificação
MFCC	Coeficiente Cepstral Frequência Mel
ML	Sinais limpos de vozes masculinas
MLP	<i>Multi Layer Perceptron</i>

MOS	<i>Mean Opinion Score</i>
MP	Modelo Perceptivo
Ms	Matriz percepção auditiva modelo Mel
ms	milisegundos
MSE	<i>Mean Squared Error</i>
NIS	Nível de Intensidade Sonora
NC	Níveis de Cinza
NR	Número de Regras
OSS	<i>One Step Secant</i>
PCM	<i>Pulse Code Modulation</i>
PESQ	<i>Perceptual Evaluation of Speech Quality</i>
POLQA	<i>Perceptual Objective Listening Quality Analysis</i>
PSQM	<i>Perceptual Speech Quality Measure</i>
RBM	<i>Restricted Boltzmann Machine</i>
SCG	<i>Scaled Conjugate Gradient</i>
RMSE	<i>Root Mean Squared Error</i>
RC	Referência Completa
RNA	Redes Neurais Artificiais
SB	<i>Subtractive</i>
SC	<i>Soft Computing</i>
SF	Sistemas <i>Fuzzy</i>
SNR	<i>Signal Noise Ratio</i>
SNRseg	Relação Sinal Ruído Segmentada
SPL	<i>Sound Pressure Level</i>
SR	Sem Referência
TSK	Takag Sugeno Kang
VAD	Detector de Atividade de Voz
VL	Variáveis Linguísticas
VOIP	<i>Voice Over Internet Protocol</i>

LISTA DE SÍMBOLOS

R^2	<i>Determination Coefficient</i>
N	Número de amostras
$s[i]$	Sinal de voz de referência
$p[i]$	Sinal sob teste
$r[i]$	Ruído de interferência
$a[i]$	Adição sinal ruído
$p(t)$	Sinal sob teste no domínio do tempo
$s(t)$	Sinal de referência no domínio do tempo
$S(\omega)$	Transformada de Fourier do sinal de referência
$P(\omega)$	Transformada de Fourier do sinal sob teste
$C_s(m)$	Coefficiente cepstrum do sinal sob teste
$C_p(m)$	Coefficiente cepstrum do sinal original
$C_p(m)$	Coefficiente cepstral
a_m	Sub-banda do sinal de voz
a_s	Coefficiente de predição linear do sinal original
a_p	Coefficiente de predição linear do sinal degradado
Φ_s^2	Espectro LPC do sinal de referência
Φ_p^2	Espectro LPC do sinal sob teste
$IS(p[i], s[i])$	Pontuação média Itakura-Saito
z	Nota PESQ
$f(z)$	Nota MOS-PESQ
Δ_{fc}	Largura de banda crítica
f_c	Frequência Central
$\Omega(\omega)$	Unidade Bark
ω	Frequência angular
$m(f)$	Unidade Mel da frequência

$\Psi(\Omega_\delta)$	Filtro banda crítica modelo de Hermansky
$M_m(f)$	Filtro banda crítica triangular de escala mel
κ	Elementos do banco de filtros
τ	Bandas críticas
$s(l, k)$	Amostras do sinal de voz
L	Número de pontos da FFT
$s_\omega(l, k)$	Sinal janelado e segmentado
$h(l)$	Função janelas de Hamming
$E(f, k)$	Transformada de Fourier discreta com janela de Hamming
$E(\omega)$	Curva de ponderação a diferentes frequências
$P(\omega)$	Resposta perceptual
$\Theta(\omega)$	Convolução
$S(k)$	Transformada Discreta de Cosseno
L_c	Número de pontos da DCT
$G(n, q)_{Bs}$	Equação de quantização linear em nível inteiro de cinza
N_c	Níveis de cinza
$Bs(\psi, k)$	Resposta perceptual modelo Hermansky na escala Bark
$\mathbb{Z}^+$	Conjunto dos números inteiros positivos
$\mathbb{R}$	Conjunto dos números reais
$G(i, j)$	Tonalidade do pixel
$M(i, j)$	Transição entre elementos da matriz em níveis de cinza
$M(i, j, D, 0^\circ)$	Matriz coocorrência na direção 0°
$M(i, j, D, 45^\circ)$	Matriz coocorrência na direção 45°
$M(i, j, D, 90^\circ)$	Matriz coocorrência na direção 90°
$M(i, j, D, 135^\circ)$	Matriz coocorrência na direção 135°
$P_{(i,j,D,\theta)}$	Matriz coocorrência normalizada
N_{lc}	Número de linhas e colunas
μ	Média
σ	Desvio Padrão

A_i	Conjuntos <i>fuzzy</i> ou variáveis linguísticas de entrada
B_i	Conjuntos <i>fuzzy</i> ou variáveis linguísticas de entrada
C_i	Conjuntos <i>fuzzy</i> ou variáveis linguísticas de entrada
$\min$	Operador mínimo
$\max$	Operador máximo
m_{A_k}	Centro da função gaussiana
σ_{A_k}	O desvio padrão da função gaussiana
O_i^j	Saída do i -ésimo neurônio na j -ésima camada
ω_i	Peso da i -ésima regra
$\bar{\omega}_i$	Grau de ativação de uma regra
θ^*	Estimador do método mínimos quadrados
E_i	Erro médio quadrático
η	Taxa de aprendizado
$Prod(A, B, C)$	Operador produto
$R_{i,j}^K$	Regras <i>fuzzy</i> do método de fuzzificação <i>Grid</i>
M_c	Matriz de partição <i>Hard</i>
U	Partição <i>Hard</i>
η_{mc}	Cardinalidade das partições
$J(U, v)$	Função objetivo do método HCM
d_{ik}	Distância Euclidiana
M_{fc}	Matriz de partição <i>fuzzy</i>
$J_m(\underline{U}, V)$	Função objetivo do método FCM
$\underline{U}$	Partição <i>fuzzy</i>
P_i	Função densidade
r_a	Raio de um grupo
r_b	Raio de eliminação
S_{ri}	Sinal real de entrada
S_{pi}	Sinal predito de saída
$\overline{S_{pi}}$	Média dos sinais predito de saída

SUMÁRIO

1	INTRODUÇÃO	25
2	AVALIAÇÃO DA QUALIDADE DE SINAIS DE VOZ	29
2.1	MEDIDAS OBJETIVAS NÃO PADRONIZADAS PELA ITU-T	30
2.1.1	ERRO QUADRÁTICO MÉDIO	30
2.1.2	RELAÇÃO SINAL RUÍDO	31
2.1.3	DISTORÇÃO ESPECTRAL	32
2.1.4	DISTÂNCIA CEPSTRAL	32
2.1.5	RAZÃO DE PROBABILIDADE LOGARÍTMICA	33
2.1.6	DISTÂNCIA ITAKURA-SAITO	34
2.2	ALGORITMOS DE AVALIAÇÃO DE QUALIDADE DE VOZ PROPOSTOS PELA ITU-T	35
2.2.1	RECOMENDAÇÃO ITU-T P.800	37
2.2.2	RECOMENDAÇÃO ITU-T P.862	37
2.2.3	RECOMENDAÇÃO ITU-T P.863	40
2.2.4	RECOMENDAÇÃO ITU-T P.563	40
2.3	AVALIAÇÃO DA QUALIDADE DE VOZ BASEADAS EM RE- DES NEURAIS ARTIFICIAIS	41
3	AUDIÇÃO HUMANA E MODELOS PSICOACÚSTICOS	43
3.1	A PERCEPÇÃO AUDITIVA: OUVIDO HUMANO	43

3.1.1	PSICOACÚSTICA	46
3.1.2	SELETIVIDADE E BANDA CRÍTICA	49
3.1.3	MODELO PSICOACÚSTICO	56
3.2	IMPLEMENTAÇÃO DO MODELO PSICOACÚSTICO	59
3.2.1	IMPLEMENTAÇÃO DO MODELO PERCEPTIVO DE HERMANSKY	62
3.2.2	IMPLEMENTAÇÃO DO MODELO PERCEPTIVO MFCC	65
4	EXTRAÇÃO DE FATORES DO SINAL DE VOZ	68
4.1	QUANTIZAÇÃO	69
4.2	CÁLCULO DA MATRIZ DE COOCORRÊNCIA DE NÍVEIS DE CINZA - (MCNC)	71
4.3	CONSTRUÇÃO DA MATRIZ DE COOCORRÊNCIA	71
4.4	CÁLCULO DE FATORES DE TEXTURA	74
5	ARQUITETURA DA ANFIS	78
5.1	ETAPA FORWARD : APLICAÇÃO DO MÉTODO DOS MÍNIMOS QUADRADOS NO PROCESSO DE TREI- NAMENTO	82
5.2	ETAPA BACKWARD : APLICAÇÃO DO BACKPROPA- GATION NO PROCESSO DE TREINAMENTO	84
5.3	PARTICIONAMENTO GRID	87
5.4	MÉTODOS DE AGRUPAMENTO	89
5.4.1	AGRUPAMENTO FUZZY C-MEANS - FCM	89
5.4.2	AGRUPAMENTO SUBTRACTIVE - SB	99
6	MÉTODO PROPOSTO	102
6.1	DIRETÓRIO DE SINAIS DE VOZ	102
6.2	DESCRIÇÃO DAS BASES DE DADOS	103

6.3	ANÁLISE ESPECTRAL	104
6.4	MODELO PERCEPTIVO	104
6.5	QUANTIZAÇÃO	106
6.6	CÁLCULO DA MATRIZ DE COOCORRÊNCIA DE NÍVEIS DE CINZA - MCNC	107
6.7	EXTRAÇÃO DOS FATORES	109
6.8	ALGORITMO PESQ	110
6.9	DIVISÃO ALEATÓRIA DOS DADOS	111
6.10	DADOS DE TREINAMENTO	111
6.11	DADOS DE TESTE	112
6.12	ANFIS	112
6.12.1	MÉTRICAS DE AVALIAÇÃO DE DESEMPENHO	112
6.12.2	METODOLOGIA PARA CONFIGURAÇÃO DA ANFIS	113
6.13	SISTEMA FIS	121
7	ANÁLISES E DISCUSSÕES	123
7.1	ESCOLHA DOS FATORES DE HARALICK	123
7.2	COMPROVAÇÃO DO MÉTODO	126
7.3	APLICAÇÕES UTILIZANDO ANFIS E REDES MLP	127
7.4	DESEMPENHO DO MODELO HERMANISKY E MFCC	129
8	CONCLUSÕES	131
	REFERÊNCIAS	133
	APÊNDICES	140

Apêndice A: Resultados da partição *fuzzy* com o algoritmo FCM para os dados de treinamento com áudios da base de dados NOIZEUS contaminados com ruído de bar para os Fatores 1, 11 e 13 de Haralick e suas respectivas avaliações PESQ.141

Apêndice B: As 969 combinações dos fatores de Haralick.	143
---	-----

INTRODUÇÃO

As medidas de avaliação da qualidade de sinais de voz são consideradas importantes ferramentas para o desenvolvimento de codificadores e técnicas de melhoramento de voz. Em um projeto de compressão de voz, por exemplo, uma medida de qualidade é fundamental para indicar a quantidade de distorção introduzida pelo algoritmo, para determinar as configurações dos parâmetros e para otimizar a estrutura do sistema. Nas comunicações móveis do tipo VoIP (do inglês: *Voice Over Internet Protocol*), consideradas telefonia via internet, o sinal de voz é compactado antes da transmissão e reconstruído na outra extremidade. O processo de codificação/decodificação do sinal invariavelmente produz distorções e uma medida de qualidade permite a comparação relativa de técnicas empregadas para a codificação de voz (MEBAWAONDU; DAHUNSI; ADEWALE, 2020).

As medidas de avaliação são consideradas subjetiva ou objetiva e são utilizadas para fornecer a qualidade do sinal de voz. As medidas subjetivas são baseadas nas classificações perceptivas por um grupo de ouvintes, que subjetivamente avaliam um sinal, considerando uma escala preestabelecida. As medidas objetivas utilizam parâmetros físicos extraídos dos sinais e são implementadas computacionalmente, economizam tempo e fornecem resultados altamente seguros em aplicações práticas (BEERENDS *et al.*, 2020). As medidas objetivas podem ser divididas em dois grupos: intrusivas, que necessitam de um sinal, referência (normalmente a entrada de um sistema em avaliação) e do sinal a ser avaliado (normalmente o sinal de saída de algum sistema) e não intrusivas, que avaliam a qualidade diretamente do sinal a ser avaliado (BEERENDS; HEKSTRA; RIX, 2002).

Nas medidas intrusivas os sinais de referência e saída do sistema em teste devem ser sincronizados antes da execução dos cálculos da qualidade de voz. Caso contrário, o cálculo da diferença entre esses dois sinais não fornece resultados satisfatórios. Na prática, uma sincronização perfeita entre sinais de referência e de saída é sempre difícil de obter, como, por exemplo, em comunicação sem fio onde outros elementos podem influenciar na relação.

Além disso, existem muitas aplicações em que o discurso original não está constantemente disponível, como em comunicações móveis, via satélite e VoIP (GAMPER *et al.*, 2019).

Nessas condições, medidas não intrusivas, que avaliam a qualidade da voz baseadas apenas no sinal de saída do sistema em teste são mais adequadas, apesar dos desafios em implementá-las.

Na literatura especializada, muitos estudos têm sido realizados com vistas ao desenvolvimento de modelos perceptivos para extrair características que possam representar a qualidade do sinal de voz. No desenvolvimento de um método conjunto que pudesse resolver problemas inerentes às propostas iniciais, surgiu a medida intrusiva Rec. ITU-T P.862, mais conhecida como PESQ (do inglês - *Perceptual Evaluation of Speech Quality*). Utilizada inicialmente para avaliação e seleção de codificadores, a PESQ passou a ser usada na maioria dos projetos em que a avaliação da qualidade de sinais de voz e áudio é imprescindível (BEERENDS; HEKSTRA; RIX, 2002).

A necessidade do sinal original limita a PESQ para monitoramento de sinais em tempo real (GOUVEIA; ROSA; RODRÍGUEZ, 2018). Devido a esta restrição, alguns esforços foram realizados recentemente para modelar a qualidade da voz de acordo com os parâmetros da rede como perda de pacotes, tipo de codificadores, gênero e idioma do locutor (DEMÓSTENES *et al.*, 2019; DONG; WILLIAMSON, 2019; MEBAWAONDU; DAHUNSI; ADEWALE, 2020). Entre os métodos padronizados pela ITU-T, técnicas inteligentes baseadas em redes neurais têm sido utilizadas com muito sucesso (LIU *et al.*, 2021).

Nos últimos anos muitos pesquisadores têm proposto de medidas baseadas em redes neurais e suas variações (ELKHODR; QUSAY; HASSAN; 2018, GAMPER *et al.*, 2019; MEBAWAONDU; DAHUNSI; ADEWALE, 2020; SONI; PATIL, 2021). Porém, poucas pesquisas foram focadas na combinação entre redes e conceitos de lógica *fuzzy*, que são abordadas com sucesso em outras áreas e não têm sido amplamente aplicadas na composição de métodos que avaliam a qualidade de sinais de voz.

Chen e Parsa (2010) propuseram uma medida não intrusiva para avaliação de sinais de voz utilizando o Sistema de Inferência Neuro *Fuzzy* Adaptativo (ANFIS). O treinamento foi realizado com um banco de dados com sinais avaliados subjetivamente, sendo seu desem-

penho comparado com a medida ITU-T P.563. Embora tenham utilizado uma abordagem semelhante a empregada neste trabalho, existem aspectos fundamentais que não foram explorados, tais como: a escolha de um modelo perceptivo; testes para determinar os melhores fatores de Haralick entre as principais direções da matriz de coocorrência; a escolha de um fuzzificador adequado para a ANFIS; o número de variáveis linguísticas para cada função de pertinência e um banco de dados de sinais de voz contaminados por diferentes ruídos.

McLoughlin (2014) enfatiza que a escolha de um modelo perceptivo é fundamental para uma representação do sinal em bandas críticas. Dois dos modelos por ele destacados são implementados neste trabalho. Fstedt *et al.* (2019) destacam que a escolha do valor de níveis de cinza e a direção tomada na matriz de coocorrência influenciam nos valores dos fatores de Haralick. Os trabalhos de Yeom e Kwak (2018) e Benmouiza e Cheknane (2019) apresentam uma análise comparativa entre os três métodos de fuzzificação: FCM, Subtractive e o Grid implementados na ANFIS. Ambos declaram, em suas aplicações, que o desempenho da ANFIS está estritamente ligado à escolha de um método de fuzzificação e a quantidade de variáveis linguísticas envolvidas em cada função de pertinência, aspectos também comprovados no trabalho de Jafari *et al.* (2021). Ma et al. (2009) enfatizam que para o desenvolvimento de projetos na área de processamento de sinais de voz, a escolha de um banco de dados apropriado é essencial. Desta forma, eles elaboraram um banco de dados com sinais de voz, o NOIZEUS, contendo sinais limpos e contaminados com oito tipos de ruídos presentes no cotidiano.

Com base nestas abordagens, a proposta deste trabalho visa apresentar um modelo de previsão para estimar a MOS obtida pela PESQ baseada nas imagens geradas a partir do modelo perceptivo. O método pode ser sintetizado em dois blocos principais. O primeiro é o bloco de extração de características, que tem a finalidade de encontrar um conjunto de atributos ou fatores para representar efetivamente diferentes propriedades da imagem do sinal, e o segundo é o bloco de modelagem cognitiva que visa construir uma função de mapeamento para combinar os fatores extraídos e fornecer a previsão da PESQ do sinal. Neste bloco um Sistema de Inferência *Fuzzy* - FIS - do tipo Sugeno de primeira ordem com fuzzificadores de agrupamento é aplicado para estimar objetivamente a qualidade do sinal. Este FIS é implementado pela ANFIS.

Para o desenvolvimento desta pesquisa é utilizado um banco de dados de sinais de voz contaminados por ruídos presentes em situações cotidianas. Cada sinal de voz é processado utilizando um modelo perceptivo e a imagem gerada do plano número de bandas críticas $\times$ números de quadros passa por um processamento utilizando uma técnica denominada matriz de coocorrência, e a partir do conjunto dessas imagens são calculados os fatores de Haralick. Dentre o conjunto de fatores são utilizados três para configurar a ANFIS e uma metodologia que busca separar os dados em conjunto de treinamento e teste de forma aleatória foi elaborada. A ideia é repetir o procedimento várias vezes para obter conjuntos de dados diferentes, com a finalidade de treinar a ANFIS para três métodos de fuzzificação: agrupamento *fuzzy C-Means* - FCM, Subtractive e particionamento de *Grid*, modificando o número de variáveis linguísticas para cada fator de entrada, atuando nas quatro principais direções da matriz de coocorrência.

O desempenho da rede será analisado utilizando as medidas *Mean Absolute Percentage Error* - MAPE, *Root Mean Squared Error* - RMSE e o Coeficiente de Determinação - R^2 , indicando quais dos conjuntos de três fatores de entrada apresentam os melhores resultados para o método proposto.

Obtendo uma configuração considerada satisfatória, o modelo passa por uma etapa de comprovação mantendo seu desempenho para outro conjunto de dados que contém um número maior de sinais, com relação a rede *Multi Layer Perceptron* - MLP com seus diferentes algoritmos de aprendizagens e com a troca do modelo perceptivo.

O restante do texto é organizado da seguinte forma: no Capítulo 2, são apresentadas as principais medidas para avaliação da qualidade de voz; no Capítulo 3, apresenta-se uma breve descrição da fisiologia do ouvido humano, os principais conceitos da psicoacústica e dois modelos perceptivos baseados nas escalas Bark e Mel; no Capítulo 4, o método de extração de fatores baseado nas características distributivas da percepção do sinal de voz é abordado; no Capítulo 5 tem-se uma descrição das redes neurais, o tipo FIS Sugeno de primeira ordem implementado pela ANFIS e os três métodos de fuzzificação; no Capítulo 6, são apresentadas, as ferramentas e ideias discutidas para composição do método proposto; No Capítulo 7 são analisados os resultados e por fim, são apresentadas as conclusões e os apêndices.

AVALIAÇÃO DA QUALIDADE DE SINAIS DE VOZ

Sinais estão presentes constatemente no nosso cotidiano. Um sinal é definido como uma unidade que transporta uma informação a respeito do estado ou do comportamento de um fenômeno físico e, geralmente, é sintetizado para a comunicação entre indivíduos ou entre indivíduos e máquinas. Na comunicação telefônica, por exemplo, a voz, um sinal acústico, é convertida em sinais elétricos pelo microfone que são transmitidos por antenas ou satélites para outras regiões onde são convertidos novamente em sinal de voz.

Para McLoughlin (2014), o processamento digital é atualmente o método mais adequado para o tratamento de áudio e voz, pois os aplicativos e os sistemas de áudio são predominantemente de natureza digital.

O avanço das técnicas de processamento de sinais digitais proporcionou um crescente interesse em desenvolver métodos eficientes para tecnologia de voz. As medidas que avaliam a qualidade dos sinais destes métodos se tornaram fundamentais para o desenvolvimento de muitos projetos de comunicação de voz (DEMÓSTENES *et al.*, 2019). Dois conceitos fundamentais para avaliação de sinais de voz são qualidade e inteligibilidade. Com certa frequência são usados indistintamente, mas suas medidas e dependências diferem. A qualidade é a medida de afinidade, ou seja, “correção” da voz. O que faz associar a que nível o sinal de voz se assemelha ao sinal alvo ou de referência. A qualidade é, portanto, altamente subjetiva, medida que depende do ouvido do receptor. Apesar de ser subjetiva, pode ser aproximada usando medidas objetivas, e pode até mesmo ser determinada com algum grau de precisão objetiva.

A inteligibilidade é a medida de quão compreensível é a fala. Em outras palavras, concentra-se no conteúdo portador de informações do discurso, referindo-se ao grau de informação que é transmitida corretamente.

McLoughlin (2014) destaca que apesar de existir uma diferença entre qualidade e inteli-

gibilidade é importante destacar que na maioria das circunstâncias, excelente inteligibilidade implica alta qualidade, enquanto que inteligibilidade insatisfatória está ligada a baixa qualidade.

O conceito de qualidade está condicionado a fatores que variam de acordo com a aplicação pretendida e com o público alvo. Os testes de avaliação de qualidade de voz podem ser realizados seguindo métodos subjetivos, onde há intervenção de indivíduos nos resultados, ou objetivos, que demandam algoritmos matemáticos computacionais para avaliação (HINTERLEITNER, 2017).

Em pesquisas que trabalham com sinais de voz, as medidas são fundamentais para avaliação da proposta. Cada aplicação tem sua forma particular e geralmente utiliza mais de uma medida para analisar o desempenho de seu projeto.

Embora no passado as medidas fossem utilizadas para fins comerciais, hoje elas constituem um campo vasto de pesquisa no meio acadêmico, por isso novas medidas são propostas e avaliadas com certa frequência.

Muitas técnicas atuais usam abordagens com grandes conjuntos de dados para treinar sistemas de aprendizado de máquina que tentam replicar a percepção de qualidade. As redes neurais e os sistemas *fuzzy* também aparecem em diversas pesquisas atuando em projetos particulares ou potencializando as medidas existentes (RAHDARI; EFTEKHARI, 2011; ELKHODR; QUSAY; HASSAN, 2018; MEBAWAONDU; DAHUNSI; ADEWALE, 2020; LIU *et al.*, 2021).

Nas próximas seções serão apresentadas algumas medidas e suas características.

2.1 MEDIDAS OBJETIVAS NÃO PADRONIZADAS PELA ITU-T

2.1.1 ERRO QUADRÁTICO MÉDIO

Uma medida que utiliza o sinal no domínio do tempo e que afere a qualidade pela diferença entre um vetor de áudio processado $p[i]$ sendo $i = 0, 1, 2, \dots, N$ e um vetor de áudio original $s[i]$

é o erro quadrático médio - MSE - (do inglês: *Mean Squared Error*). Ela é calculada amostra por amostra fazendo o quadrado da diferença entre os vetores, ambos de comprimento N (MCLOUGHLIN, 2014):

$$\text{MSE} = \frac{1}{N} \sum_{i=0}^{N-1} \{s[i] - p[i]\}^2 \quad (1)$$

Em projetos que procuram melhorias na qualidade de sinais de voz é possível aplicar uma condição que minimize o MSE.

2.1.2 RELAÇÃO SINAL RUÍDO

Para casos em que os sinais de interesse não estão realmente sendo comparados quanto à semelhança, mas sim quando um sinal é corrompido por outro, as proporções dos próprios sinais podem ser úteis, como a relação sinal ruído - SNR - (do inglês: *Signal Noise Ratio*). Observe que a SNR não é usada para medir o grau de diferença entre os sinais, mas é o logaritmo de base 10 da razão entre um sinal desejado $s[i]$ (sinal de voz de referência) e o vetor de ruído de interferência $r[i]$, sendo dado em decibéis - dB. Assim como o MSE, a SNR é uma medida no domínio do tempo (MCLOUGHLIN, 2014):

$$\text{SNR} = 20 \times \log_{10} \left\{ \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{s[i]}{r[i]} \right) \right\} \quad (2)$$

A Relação Sinal Ruído Segmentada - SNRseg é tomada medida da relação sinal ruído em segmentos do sinal de voz. Estes segmentos têm tipicamente de 20 a 30 milissegundos de duração, normalmente ocorrendo alguma sobreposição. Para um quadro de N amostras e sem sobreposição, para o j -ésimo segmento seria como mostrado na equação (3).

$$\text{SNRseg} = 20 \times \log_{10} \left\{ \frac{1}{N} \sum_{i=jN}^{(j+1)N-1} \left(\frac{s[i]}{r[i]} \right) \right\}. \quad (3)$$

A equação (3) funciona quando se tem acesso aos arranjos separados de sinal e ruído. Para casos em que é apresentado apenas o sinal original, sem o sinal corrompido por ruído $a[i] = s[i] + r[i]$, as equações (2) e (3) serão diferentes, substituindo $\frac{s[i]}{r[i]}$ por $\frac{s[i]^2}{(a[i]-s[i])^2}$.

2.1.3 DISTORÇÃO ESPECTRAL

Com base no fato de que a audição não se limita ao domínio do tempo, as melhores medidas, ou pelo menos as mais relacionadas a análise subjetiva do mundo real, podem ser obtidas no domínio da frequência para muitas tarefas. A primeira medida de diferença no domínio da frequência é a distorção espectral, mostrada na equação (6), medida em dB², como uma comparação entre os sinais $p(t)$ e $s(t)$ para um quadro de tamanho N .

O sinal a ser analisado é convertido do tempo para o domínio da frequência utilizando a transformada de Fourier (MCLOUGHLIN, 2014):

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} s(t)e^{-j\omega t} dt; \quad (4)$$

$$P(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} p(t)e^{-j\omega t} dt. \quad (5)$$

Com ambos os sinais transformados, calcula-se a Distorção Espectral - DE:

$$DE = \frac{1}{\pi} \int_0^\pi (\log(S(\omega)) - \log(P(\omega)))^2 d\omega, \quad (6)$$

a DE é uma medida de diferença perceptualmente relevante para sinais de voz. Contudo, pode ser melhorada ainda mais atuando em regiões de frequências audíveis do sinal. Isso produz uma distorção espectral perceptualmente ponderada e é usada em sistemas práticos que utilizam voz de alta qualidade e análise de sinais de áudio.

2.1.4 DISTÂNCIA CEPSTRAL

As Medidas de distância entre cepstros são utilizadas como parâmetros discriminantes em diferentes aplicações, principalmente em reconhecimento de voz.

Para obter os coeficientes de dois sinais $s(t)$ e $p(t)$ a serem comparados, procura-se calcular recursivamente os denominados coeficientes cepstrais, $C_s(m)$ e $C_p(m)$ (MCLOUGHLIN, 2014):

$$C(m) = a_m + \sum_{k=1}^{m-1} \frac{k}{m} \{C(k)_{a_{m-k}}\} \text{ para } 1 \leq m \leq P. \quad (7)$$

Obtidos os dois conjuntos de coeficientes, calcula-se a Distância Cepstral (DC), pela equação (8).

$$DC = 10 / \log_{10} \sqrt{2 \sum_{i=1}^P \{C_s(i) - C_p(i)\}^2} \quad (8)$$

2.1.5 RAZÃO DE PROBABILIDADE LOGARÍTMICA

A razão de probabilidade logarítmica - LLR (do inglês: *Log Likelihood Ratio*) também é medida no domínio da frequência. Toma-se um sinal como referência, para medir a proximidade em relação a outro sinal (“degradado”). Em seguida, calcula-se a matriz de autocorrelação do sinal de referência $s[i]$ para calcular a razão logarítmica entre os dois sinais, onde T indica a matriz transposta (MCLOUGHLIN, 2014):

$$LLR = \log \left\{ \frac{a_p s[i] a_p^T}{a_s s[i] a_s^T} \right\} \quad (9)$$

sendo que a_s e a_p são os vetores de coeficientes de predição linear dos sinais original e degradado, respectivamente.

Uma gravação típica consistiria em muitos quadros de análise, que podem ser sobrepostos. O tamanho do quadro, deve ser escolhido para garantir que o sinal dentro de cada quadro seja pseudo estacionário.

No cálculo, uma matriz de autocorrelação pode ser obtida para cada janela de análise e, portanto, uma pontuação LLR seria obtida por quadro. A distância total entre as duas seria

forneida pela média de todas as pontuações LLR.

2.1.6 DISTÂNCIA ITAKURA-SAITO

A distância Itakura-Saito (IS), amplamente utilizada, é calculada de forma semelhante ao LLR, usando os mesmos dados de entrada da seguinte maneira (MCLOUGHLIN, 2014):

$$\text{IS} = \frac{\Phi_s^2}{\Phi_p^2} \left\{ \frac{a_p s[i] a_p^T}{a_s s[i] a_s^T} \right\} \log \left\{ \frac{\sigma_s^2}{\sigma_p^2} \right\} - 1. \quad (10)$$

A principal diferença está em Φ_s^2 e Φ_p^2 que representam os espectros da Codificação Linear de Previsão - LPC (do inglês: *Linear predictive coding*), obtidos por $1/F(e^{j\omega T})$, onde o termo $\omega T = 2\pi k/N$ para $k = 0, 1, \dots, (N-1)$, para uma resolução de frequência de $Fs/2N$ Hz na frequência de amostragem, sendo Fs calculada sobre um segmento de N amostras.

Tal como acontece com a LLR, uma pontuação IS é calculada para cada janela de análise, com a pontuação final sendo a média das pontuações em cada janela. A IS tende a apresentar alguns valores discrepantes que são frequentemente removidos antes do cálculo da média, limitando o intervalo de pontuação a $[0, 100]$. Um problema adicional aqui é que a IS não é simétrica, ou seja, $\text{IS}(a, b) \neq \text{IS}(b, a)$. Portanto, é importante determinar qual é o sinal de referência e o sinal degradado ao obter uma pontuação IS. Se isso não puder ser feito, é melhor calcular a pontuação média em ambas as direções:

$$\text{IS}(p[i], s[i]) = \frac{1}{2} \{ \text{IS}(p[i], s[i]) + \text{IS}(s[i], p[i]) \}. \quad (11)$$

2.2 ALGORITMOS DE AVALIAÇÃO DE QUALIDADE DE VOZ PROPOSTOS PELA ITU-T

Os testes de avaliação de qualidade de voz propostos pela ITU-T podem ser realizados seguindo métodos objetivos, em que o índice MOS (do inglês: *Mean Opinion Score*) é dado por *software*, ou subjetivos, nos quais há intervenção de indivíduos no resultado da avaliação.

Na aplicação de testes subjetivos, é preciso seguir recomendações específicas e selecionar com cautela os sinais, fonte e principalmente os avaliadores tanto em termos qualitativos quanto quantitativos a serem utilizados nos testes. Os testes de entrevista, conversacionais e de audição, constituem as técnicas mais comuns para avaliar subjetivamente um sinal, se destacando pela importância e metodologia empregada (HINTERLEITNER, 2017).

Existem diversos tipos de testes subjetivos da audição, dentre eles destacam-se: O Testes de Qualidade Absoluta o - ACR - (do inglês: *Absolute Category Rating*), Testes de Degradação - DCR - (do inglês: *Degradation Category Rating*) e Testes de Comparação - CCR - (do inglês: *Comparison Category Rating*), todos recomendados pela ITU (do inglês: *International Telecommunication Union*).

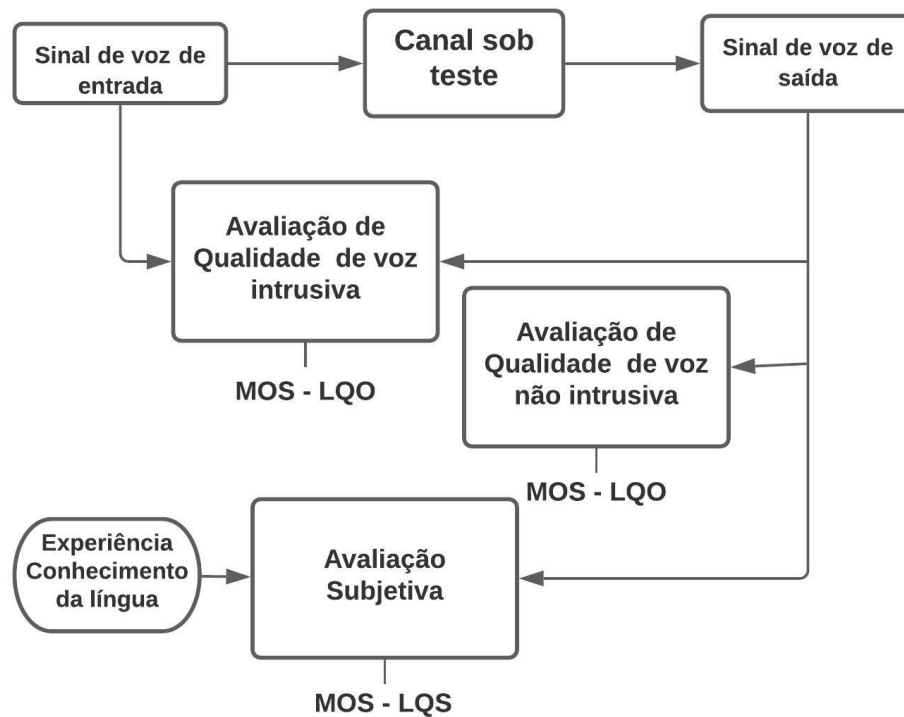
Os métodos objetivos são baseados em algoritmos que tentam predizer a avaliação humana sobre um sinal de voz (ITU-T, 1996). Um método objetivo de qualidade de sinais de voz não é capaz de reproduzir exatamente as pontuações médias obtidas pelos testes subjetivos. Os modelos baseados em sinais de voz são classificados como estimadores de qualidade e definidos como intrusivos ou Referência Completa - RC - e não intrusivos ou Sem Referência - SR.

Os métodos intrusivos precisam do sinal de referência para comparar com o sinal processado e garantir a qualidade da pontuação. Por esta razão, são mais precisos e tidos como referência para a avaliação objetiva. Estes métodos também são utilizados para aplicações *off-line*, incorporando aspectos do sistema auditivo humano considerados relevantes para estimação de qualidade.

Em contrapartida, métodos não intrusivos dependem apenas do sinal a ser avaliado e assim, o que garante simplicidade e rapidez, possibilitando seu uso em serviços de tempo

real, como por exemplo em sistema de rede VoIP (DEMÓSTENES; ROSA; ALMEIDA; MITTAG; 2019). Tais métodos podem ser divididos em dois tipos: objetivos, caso utilize um *software* para a análise e pontuação da qualidade, e subjetivos, caso um ouvinte intervenha diretamente na pontuação dos resultados sobre a qualidade. Uma síntese destes métodos é representada na Figura 1, na qual é ilustrado o MOS-LQS (do inglês: *Listening Quality Subjective*) e a MOS-LQO (do inglês: *Listening Quality Objective*) são os resultados para as medidas subjetivas e objetivas, respectivamente.

Figura 1: Diagrama dos métodos de avaliação: subjetiva e dos métodos objetivos não intrusivos e intrusivos.



Fonte: Elaborada pelo próprio autor.

Como destacado, os métodos objetivos utilizam recursos computacionais para inferir a qualidade da voz submetida a um sistema de transmissão e/ou a um codificador de áudio. Dentre os métodos objetivos, tem-se o Rec. ITU-T G.107, o E-Model e os métodos conhecidos como métodos perceptuais, que se baseiam no sistema auditivo humano para comparar um sinal de referência (trecho de voz previamente gravado) com um sinal degradado (sinal de referência submetido ao sistema de transmissão a ser avaliado).

A seguir serão descritas as recomendações ITU-T: Rec. P.800, Rec. P.862, Rec. P.863 e

Rec. P.563.

2.2.1 RECOMENDAÇÃO ITU-T P.800

A Rec. ITU-T P.800 descreve métodos e procedimentos utilizados na indústria de telecomunicações para condução e avaliação subjetiva da qualidade de transmissão (ITU-T, 1996). Estes testes são conduzidos utilizando apenas ouvintes ou métodos conversacionais que, usando a escala de opinião da Tabela 1, fazem sua avaliação. Uma escala com pontuação média de opinião, a MOS, que varia de 1 a 5, é destacada na Tabela 1. Os valores são obtidos fazendo-se a média aritmética dos pontos atribuídos pela avaliação, sendo que quanto maior o valor obtido melhor a qualidade atribuída à avaliação.

Tabela 1: Pontuação da recomendação ITU-T P.800.

Pontuação(MOS)	P.800 (ACR)
5	Exelente
4	Bom
3	Regular
2	Ruim
1	Péssimo

Fonte: ITU-T P.800.

As escalas de opinião que são frequentemente utilizadas pelas aplicações ITU-T e devem ser adaptadas ao idioma desses indivíduos durante os testes são: LQS, que é utilizada para avaliação da qualidade do sinal, nos testes conversacionais e de audição, (ver Tabela 1), a LES (do inglês: *Listening Effort Scale*) considerada uma escala que mede o esforço necessário para compreender o significado das sentenças e a LPS (do inglês: *Loudness Preference Scale*) que é uma pontuação para a escala de preferência de audibilidade (ITU-T, 1996).

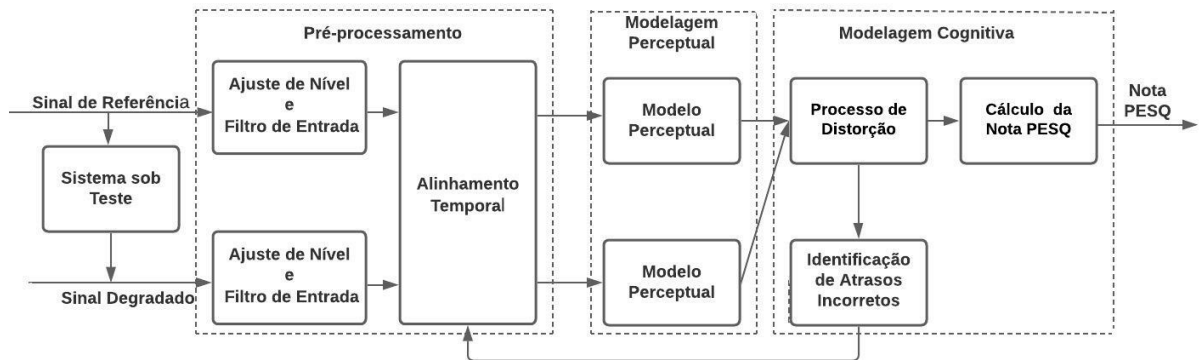
2.2.2 RECOMENDAÇÃO ITU-T P.862

A Rec. ITU-T P.862, intitulada Avaliação Perceptiva da Qualidade de Voz - PESQ, trata-se de um método objetivo e intrusivo elaborado, em forma de algoritmo, por (RIX *et al.*, 2001) com o objetivo de complementar a Rec. ITU-T P.861 - PSQM (do inglês: *Perceptual Speech*

Quality Measure), (1998), incapaz de fazer previsões sob condições de perda de pacotes, ruído de fundo e distorções de codificação.

A primeira versão do modelo PESQ focou em avaliações de qualidade de banda estreita, cobrindo 100 a 3500 Hz. No entanto, uma nova padronização foi realizada, capaz de cobrir distorções em larguras de banda de até 7 kHz.

Figura 2: Arquitetura do método intrusivo PESQ.



Fonte: Adaptado de Rix *et al.* (2001).

O algoritmo PESQ pode ser descrito por meio de três blocos principais: pré-processamento, modelagem perceptual e modelagem cognitiva, conforme a Figura 2. Cada um dos três blocos pode ser subdividido em estruturas específicas, descritas nas seguintes etapas (RIX *et al.*, 2001):

a) Etapa 1 – Pré-Processamento - Os sinais de referência e sob teste são preparados para análise do modelo perceptual, executando as seguintes operações:

1.a) Alinhamento de Nível – Os sinais de referência e sob teste são nivelados para levar em consideração os ganhos e perdas do sistema. Primeiramente, são filtrados para simular os efeitos da banda telefônica (300 Hz a 3400 Hz), então são calculados seus valores de potência média, e finalmente ganhos são aplicados para alinhar ambos os sinais.

2.a) Alinhamento de tempo – Os sinais sob teste são deslocados no tempo para alinhamento com o sinal de referência. Isto é executado em segmentos temporais individuais. Atraso durante a fala e o silêncio é levado em conta neste processo.

b) Etapa 2 – Modelagem Perceptual. A modelagem perceptual transforma os sinais de referência e sob teste em representações perceptíveis ao ser humano.

1b) Transformação tempo-frequência - Divide os sinais em janelas de 256 amostras (considerando sinais com 8000 Hz de taxa de amostragem), aplicando a transformada rápida de Fourier - FFT - (do inglês: *Fast Fourier Transform*) em cada uma dessas janelas.

2b) Transformação na escala de frequência - O espectro uniformemente distribuído resultante da FFT é transformado para a escala Bark, que procura reproduzir a acuidade da percepção de alguns fenômenos na frequência pelo sistema auditivo humano.

3b) Transformação de intensidade - Os sinais na escala Bark passam por uma transformação do espectro de frequência para uma escala de intensidade sonora, que objetiva aproximar a variação de percepção da intensidade sonora pelo sistema auditivo humano.

c) Etapa 3 – Modelagem Cognitiva. O processo de modelagem cognitiva é executado para calcular dois tipos de valores de distúrbios (simétrico e assimétrico). Estes dois valores são combinados no final para produzir uma nota para o sinal.

1c) Densidade de distorção - Diferença entre as representações de ambos os sinais no domínio da frequência. Valores positivos indicam a presença de ruído aditivo, enquanto que valores negativos representam perda de informação.

2c) Mascaramento de pequenas distorções - Um conjunto de limiares ao longo da frequência é calculado para cada uma das janelas do mapeamento tempo-frequência. Em seguida, as distorções que se encontram abaixo de seus respectivos limiares são consideradas inaudíveis por terem sido mascaradas.

3c) Processamento assimétrico - Processando-se apenas as distorções classificadas como ruído aditivo, cada uma delas é multiplicada por um fator distinto, obtido a partir da densidade de distorção em cada uma das bandas/raias na frequência.

4c) Medida de distorção - Realiza-se o cálculo das medidas de distorção simétricas e assimétricas por meio da ponderação dos resultados ao longo do tempo, agregando primeiramente os resultados de 20 quadros com 50% de sobreposição (32 ms) e em seguida unindo todos os resultados obtidos ao longo do sinal.

5c) Identificação de atrasos incorretos - Atrasos incorretos nas janelas são identificados e estimados. Após o realinhamento, as medidas de distorção são recalculadas para as janelas que foram afetadas.

6c) Obtenção da nota PESQ - As medidas de distorção simétrica e assimétrica são combinadas linearmente para gerar a nota de saída fornecidas pelo avaliador. Esta nota é atribuída pelo avaliador equivalente a MOS que o sinal sob teste receberá. Essa conversão de notas PESQ em valores de MOS são fornecidas pela equação (12) (RIX; BEERENDS; HOLLIER; HEKSTRA, 2001):

$$f(z) = 0,999 + \frac{4}{1 + e^{13669z+3,8224}} \quad (12)$$

onde z é a nota PESQ e $f(z)$ representa a nota MOS-PESQ (ITU-T, 2005).

A PESQ foi proposta inicialmente para prever a qualidade do sinal de voz, submetido a distorções e erros de codificação, perda de pacotes, atrasos constantes ou variáveis, filtragem e ruídos de redes analógicas. No entanto, Ma *et al.* (2009) verificaram que esta medida é razoável na predição da inteligibilidade de sentenças.

2.2.3 RECOMENDAÇÃO ITU-T P.863

A Rec. ITU-T P.863 conhecida como POLQA - (do inglês: *Perceptual Objective Listening Quality Analysis*) descreve um método objetivo para predição global da qualidade de voz a partir da banda estreita (300 Hz até 3,4 kHz) até a super banda larga (50 Hz até 14 kHz) em cenários de telecomunicações. A POLQA é considerada uma sucessão da PESQ sem grandes modificações na estrutura do algoritmo (ITU-T, 2011). Para Beerends *et al.* (2020) esta atualização permite prever com precisão o impacto das degradações de técnicas avançadas de processamento de sinais, usando a super banda larga em sinais de voz. Assim, a POLQA incorpora requisitos atuais da indústria e permite a avaliação de voz na super banda larga, em redes e CODEC - (Codificadores/Decodificadores) que introduzem a sincronização temporal.

2.2.4 RECOMENDAÇÃO ITU-T P.563

A Rec. ITU-T P.563 descreve um método objetivo e não intrusivo para a predição da qualidade subjetiva de voz, oriunda de aplicações telefônicas de banda estreita (até 3,1 kHz) e, por isso, é recomendada para aplicações de tempo-real, como o VoIP. Esta medida funciona

como se um ouvinte humano estivesse em um dado ponto da transmissão ouvindo o sinal com um dispositivo convencional de telefonia, atribuindo lhe ao final uma nota (ITU-T, 2004).

A avaliação da qualidade é composta das seguintes etapas: primeiro, o sinal é pré-processado usando o modelo do aparelho receptor; em seguida, um detector de atividade de voz - VAD (do inglês: *Voice Activity Detector*) é usado para identificar as partes do sinal que contêm voz; na sequência o nível de voz é calculado; finalmente, um ajuste de nível de voz é aplicado. O sinal de voz pré-processado a ser avaliado será investigado separadamente, detectando nele um conjunto de parâmetros. Esta análise é inicialmente aplicada a todo o sinal. Então, parâmetros característicos de voz são calculados com base nas restrições de parâmetros-chave. Finalmente, usando tais parâmetros e um modelo de qualidade da voz, é dada uma pontuação como avaliação.

Os parâmetros do sinal, obtidos pelo algoritmo P.563, podem ser divididos em três blocos funcionais, independentes, que correspondem às principais classes de distorção: análise do trato vocal para vozes masculinas, femininas e de forte robotização, análise de forte ruído adicional e interrupções, silenciamentos e recortes de tempo.

O sistema de pontuação utilizado é baseado na escala MOS-LQO, que está de acordo com a Rec. ITU-T P.800 (ITU-T, 1996). A pontuação atribuída a um determinado sinal pode ser comparada com a qualidade percebida por um indivíduo que ouve o mesmo sinal com um dispositivo convencional de telefonia no mesmo ponto, e os valores variam de 1 a 5, de acordo com a Tabela 1.

2.3 AVALIAÇÃO DA QUALIDADE DE VOZ BASEADAS EM REDES NEURAIIS ARTIFICIAIS

O desenvolvimento de vários algoritmos de aprendizado de máquina tais como, as Redes Neurais Artificiais (RNA) e algoritmos de aprendizagem profunda ou DL (do inglês: *Deep Learning*) e suas aplicações no processamento de sinais pavimentam o caminho para o surgimento de modelos de avaliação de qualidade objetivos que proporcionam resultados satisfatórios para diversas aplicações (LIU *et al.*, 2021).

Geralmente estes modelos são criados para resolver um problema específico encontrado nas medidas padronizadas podendo estar presentes nas etapas do processo ou até mesmo transformando uma medida intrusiva em não intrusiva (RAHDARI; EFTEKHARI, 2011). Atualmente, a Rede Neural Convolucional - CNN (do inglês: *Convolutional Neural Network*), a Máquina de Boltzmann Restrita - RBM (do inglês: *Restricted Boltzmann Machine*) e suas variações são métodos bastante difundidos na literatura e apresentam bons resultados (GOUVEIA; ROSA; RODRÍGUEZ, 2018).

Para Liu *et al.* (2021), o principal desafio das novas medidas de avaliação da qualidade de sinais de voz é melhorar a previsão dos métodos não intrusivos ou mesmo superar os métodos intrusivos de avaliação objetiva. Até o momento, a P.563 é o único padrão publicado na ITU-T para avaliar a qualidade do sinal de voz sem referência e sua precisão está longe dos métodos intrusivos (GAMPER *et al.*, 2019).

Nesta perspectiva, este trabalho propõe uma abordagem com o objetivo de estimar a PESQ por meio de um modelo de regressão baseada em uma Rede Neuro *Fuzzy* Adaptativa - ANFIS. A proposta pretende seguir a mesma linha de aplicabilidade da PESQ, medida muito utilizada para avaliar os projetos na área de processamento de sinais de voz.

AUDIÇÃO HUMANA E MODELOS PSICOACÚSTICOS

Um estudo da audição humana e dos processos biomecânicos envolvidos revela várias etapas não lineares, ou estágios, na percepção do sinal acústico. Entretanto, o pré-processamento realizado pelo ouvido sobre um sinal pode ser considerado uma atividade objetiva, pois envolve a transformação do sinal acústico que chega ao ouvido externo em impulsos elétricos nos feixes de neurônios distribuídos ao longo da função da cóclea. O processamento subjetivo será realizado pelas funções superiores do córtex cerebral, baseado neste sinal gerado pelo ouvido. McLoughlin (2014) diz que esses processos, normalmente, ocorrem de maneira semelhante entre os indivíduos, o que possibilita aplicar um modelo único de percepção auditiva em diferentes pessoas. Neste capítulo serão discutidos aspectos fundamentais das estruturas anatômicas do processo de percepção e audição humana, os principais conceitos provenientes do estudo do comportamento auditivo e a modelagem matemática do ouvido.

3.1 A PERCEPÇÃO AUDITIVA: OUVIDO HUMANO

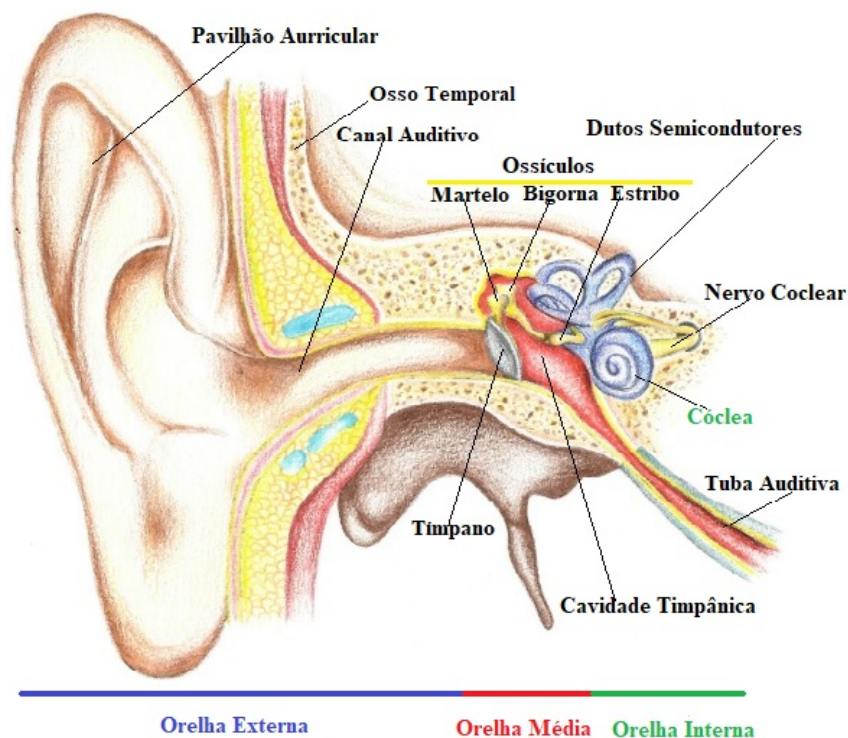
A audição humana é um sistema especializado de comunicação e é formada nos primeiros meses de gestação por meio de estímulos fisiológicos e psicossociais. Gelfand (2010) destaca que por volta dos cinco meses de gestação o sistema auditivo periférico está totalmente formado, porém a porção central do sistema não tem seu desenvolvimento completo no mesmo período.

A percepção auditiva envolve a recepção e a interpretação dos estímulos sonoros através do sentido da audição. A audição é um dos cinco sentidos dos seres humanos, e o termo orelha se refere ao órgão que facilita as audições (ARAS, 2008).

Para fins de estudo, a estrutura do aparelho auditivo humano é constituída de três partes: orelha externa, orelha média e orelha interna, que processam o som em distintas frequências,

esta representação é ilustrada na Figura 3.

Figura 3: Estrutura do ouvido humano.



Fonte: Adaptado de Conn (2017).

O processo da audição começa quando as ondas sonoras entram no conduto da orelha externa. Este conduto condensa as ondas e as leva ao tímpano. A orelha externa é constituída pelo pavilhão auricular e pelo canal auditivo externo, possui funções de filtragem que aumentam a sensibilidade e auxiliam a localização espacial da fonte sonora. O som que chega ao tímpano (orelha média) é afetado pela função de transferência na entrada do canal auditivo e pela função de transferência do conduto auditivo externo (DAVIES, 2016; PFAFF; VOLKOW, 2016).

Na orelha média é feita uma adaptação de impedância acústica (ou seja, da relação entre pressão e deslocamento) entre o ar e o meio aquoso que preenche a orelha interna. Os ossículos (martelo, bigorna e estribo) formam uma ponte, conduzindo as vibrações do tímpano até a orelha interna. As ondas, neste processo, chegam sob a forma de energia mecânica amplificada (CONN, 2017). A função da orelha média é a de ajustar as impedâncias e amplificar o som

(KEENER; SNEYD, 2009).

A orelha interna é composta basicamente da cóclea, uma cavidade cônica, enrolada na forma de um caracol, preenchida por um meio aquoso denominado perilinfa (líquido rico em sódio, Na^+), e pelo labirinto membranoso (bolsas membranáceas comunicantes), preenchido por endolinfa (líquido com alta concentração de potássio, K^+ , e baixa concentração de sódio, Na^+) (CONN, 2017). Apresenta-se dividida em duas câmaras longitudinais através da membrana basilar, sobre a qual estão distribuídos os órgãos de Corti, que são, efetivamente, sensores de vibração ligados a terminações nervosas. O nervo vestibular transmite para o cérebro os sinais elétricos gerados pelo sistema vestibular (responsável pelo equilíbrio) e o nervo coclear transmite os sinais elétricos gerados pela cóclea (responsável pela codificação das informações sonoras) (DAVIES, 2016).

O processamento periférico da informação auditiva começa por meio da onda sonora que inicialmente é captada pela orelha e conduzida à membrana timpânica pelo meato acústico externo. A orelha interna, especificamente a cóclea, recebe a transmissão sonora pela vibração da membrana timpânica e da cadeia ossicular (por via aérea) ou pela vibração direta dos ossos do crânio (por via óssea) (PFAFF; VOLKOW, 2016).

A condução por vias aéreas ocorre quando o som se propaga ao longo do canal auditivo, sendo conduzido para a orelha média e atingindo a orelha interna. Isso acontece seguindo um processo que passa pelas seguintes etapas (CONN, 2017):

1. a entrada de sons no pavilhão auricular e posteriormente no canal auditivo externo faz com que a membrana timpânica se movimente;
2. a membrana timpânica vibra com o som, provocando oscilações;
3. as oscilações sonoras provocadas pelos ossículos estimulam a cóclea;
4. as vibrações sonoras fazem a endolinfa na cóclea se mover;
5. o movimento produzido pela endolinfa causa contração das células ciliadas. Tais células criam sinais neurais que são captados pelo nervo auditivo. As células ciliadas de uma extremidade da cóclea enviam informações de som de baixa frequência e, células ciliadas do outro extremo enviam informações de som de alta frequência;
6. o nervo auditivo envia sinais ao córtex auditivo que os interpretará como sons.

O som, particularmente na faixa de baixa frequência, pode percorrer o canal auditivo rodeado pelo osso craniano, ignorar as orelhas externa e média e atingir diretamente o ouvido interno, sendo esse fenômeno chamado de condução óssea (PFAFF; VOLKOW, 2016). A partir deste processo a audição segue as etapas 5 e 6 supracitadas.

Na próxima seção serão abordados conceitos fundamentais baseados nos princípios da audição humana.

3.1.1 PSICOACÚSTICA

A psicoacústica é o estudo da percepção subjetiva de sinais acústicos. Trata-se das definições de limiares de audibilidade, percepção da intensidade da frequência do som, mascaramento, e os efeitos da audição binaural, localização das fontes, efeito estéreo, *surround* e outros. Inicialmente associada diretamente aos estudos do sistema auditivo humano, a psicoacústica e suas teorias têm sido exploradas em sistema inteligentes artificiais (MCLOUGHLIN, 2014; HERRE; DICK, 2019).

O estudo da psicoacústica teve seu início nas ideias do matemático, médico e físico alemão Hermann L. F. Von Helmholtz - 1821-1894, que afirmava que a música não é mera atribuição mecânica ou matemática oriundas de instrumentos que produzem o sons. Para ele, a música não adivinha de uma força extra humana ou metafísica, mas que era um resultado de como o corpo humano percebia o som e de como o cérebro o interpretava.

Para Rabiner e Schafer (2013), o processamento em níveis mais altos no cérebro forma uma sequência com múltiplas representações seguidas por alguns tipos de reconhecimento de padrões. Isso não é bem entendido e só pode postular os mecanismos utilizados pelo cérebro humano para perceber o som ou a voz. Mesmo assim, uma riqueza de conhecimentos sobre como os sons são percebidos foi descoberta por experimentos cuidadosos que usam tons e sinais de ruído para estimular o sistema auditivo dos observadores humanos em situações muito específicas e controladas. Estas experiências renderam muito conhecimento valioso sobre a sensibilidade do sistema auditivo humano às propriedades acústicas, como intensidade e frequência.

Um fator preponderante na percepção da voz e outros sons é a sonoridade (do inglês:

Loudness). O volume sonoro é uma qualidade perceptual relacionada à propriedade física do nível de pressão sonora. Define-se *Loudness* como a grandeza subjetiva associada à percepção sensorial da intensidade de um som. O volume sonoro é expresso em Phons, sendo que o nível em Phons é numericamente igual à pressão sonora em SPL (do inglês: *Sound Pressure Level*) na frequência de 1 kHz.

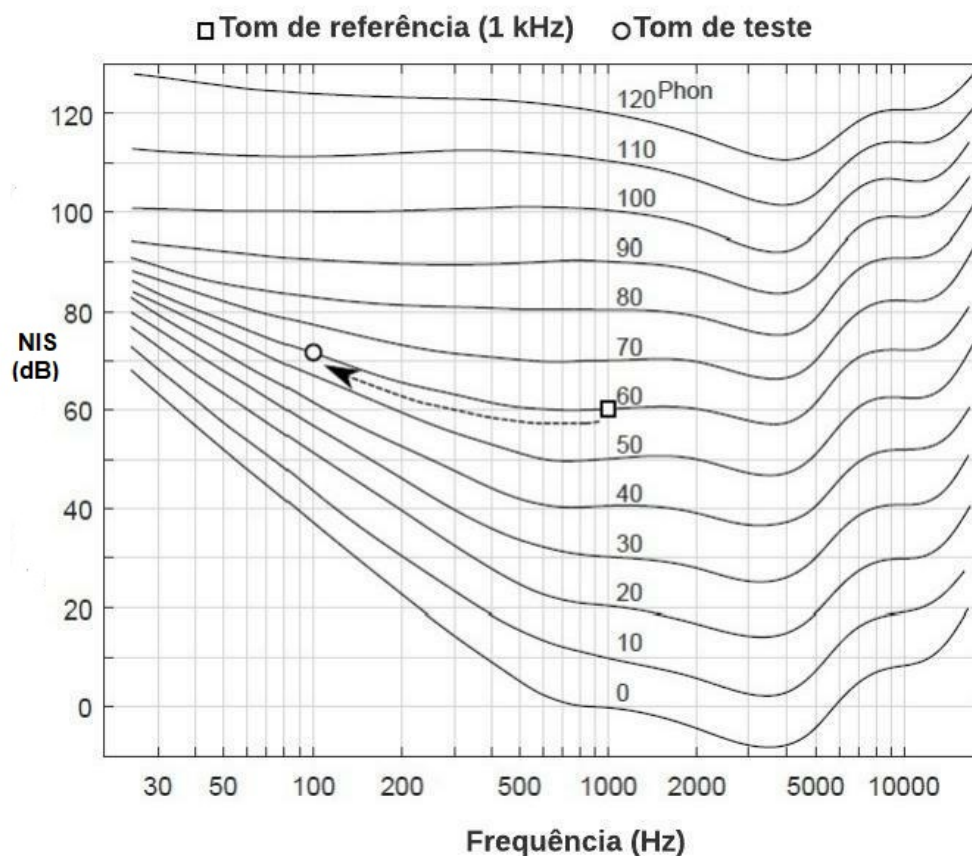
A audibilidade é quantificada relacionando o nível de pressão sonora de um tom puro (em dB em relação a um nível de referência padrão) e a percepção do volume do mesmo tom em Phons ao longo do alcance da audição humana (20 Hz – 20 kHz). A fim de medir a intensidade subjetiva dos sons, foi criada uma grandeza psicoacústica chamada nível de audibilidade. Na Figura 4 são exibidas as curvas de audibilidade ou curvas Fletcher e Munson, que são curvas de *Loudness*, atualmente padronizadas pela regulamentação ISO226:1897.

O primeiro limiar de audição foi determinado por meio de um teste subjetivo na qual é colocado um observador em frente a um alto-falante, com um metro de distância, vibrando a uma frequência de 1kHz em intensidade audível. Na sequência, reduz-se a intensidade do som até que o mesmo não seja mais perceptível, então, substitui-se o observador por um microfone calibrado para medir a intensidade do som. Dessa forma, esta intensidade será o limiar de audição para 1 kHz, que corresponde a 10^{-16} w/cm², ou 0 dB, denominado limiar de silêncio. Sons e ruídos abaixo deste limiar são inaudíveis para os seres humanos. Repetindo o procedimento para outras frequências obtém-se o limiar de audibilidade. O limiar de dor é obtido realizando o mesmo experimento. Primeiramente é fixada a frequência de 1 kHz e posteriormente vai sendo aumentado o nível do som até o limite máximo de 120 dB, podendo ser atingido antes. Os dois procedimentos são repetidos para as demais frequências observadas.

São ilustradas, na Figura 4, as curvas de sonoridade. Nela, destaca-se uma medida 60 dB de NIS (Nível de Intensidade Sonora) a 1 kHz, uma redução na frequência para 100 Hz provoca um aumento no NIS de 70 dB. Rabiner e Schefer (2013) destacam que as curvas de nível de volumes iguais mostram que o sistema auditivo é mais sensível para frequências que variam de cerca de 100 Hz até cerca de 6 kHz com a maior sensibilidade em torno de 3 a 4 kHz. Isto é quase a gama de frequências ocupadas pela maioria dos sons de voz.

São os ouvidos externo e médio, em função da quantidade de energia acústica transmitida em cada frequência, que modelam a curva de limiar de sensibilidade auditiva. O ouvido interno é um detector em que o limiar é constante em função da frequência na quase totalidade da gama auditiva. O ouvido interno funciona como um analisador das frequências mecânicas.

Figura 4: Curvas de audibilidade ou Fletcher–Munson.



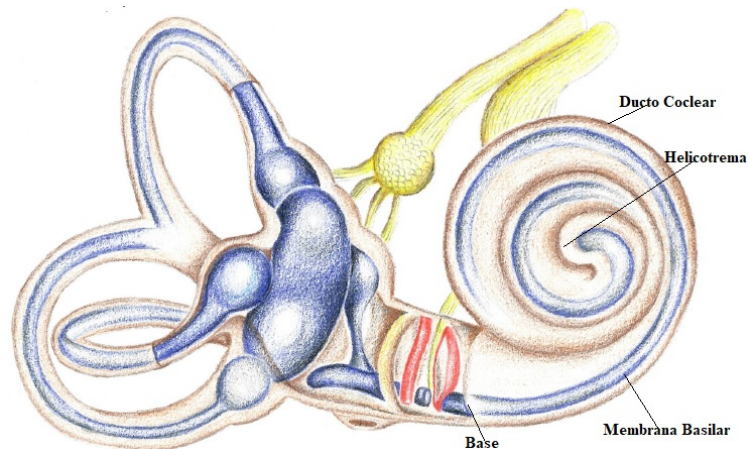
Fonte: Adaptada de Hill (2019).

A maioria dos sons musicais, assim como os sons da voz, têm uma estrutura semelhante quando analisados em intervalos curtos e são percebidos pelo sistema auditivo por meio de uma frequência fundamental denominada *pitch* (ou tom). Assim como a intensidade *Loudness*, o tom é um atributo subjetivo que está relacionado ao som e este por sua vez está relacionado ao atributo físico da forma de onda acústica.

3.1.2 SELETIVIDADE E BANDA CRÍTICA

A seletividade, ou seja, a resposta em frequência de cada célula sensora, varia com a posição da mesma em relação a cóclea. É exibida na Figura 5 a estrutura coclear (formato espiral) constituída por um tubo ósseo dividido em três seções, todas contendo a perilinfa. A primeira, a Scala Vestibuli, está ligada a janela oval, enquanto que a última está ligada a janela redonda. Elas se unem no ápice da cóclea, localizado no helicotrema. Estas secções são separadas por duas membranas: a membrana de Reissner, entre o ducto coclear e a escala vestibular, e a membrana basilar, entre o ducto coclear e a escala timpânica. A membrana basilar é peça fundamental nesta fase, pois funciona como um transdutor que sustenta os órgãos de Corti (em homenagem ao médico italiano Alfonso Corti, 1822-1876) (DAVIES, 2016).

Figura 5: Ouvido interno.



Fonte: Adaptado de Conn (2017).

A força exercida pelo martelo, a bigorna e o estribo sobre a janela oval se converte dentro do ouvido interno, em ondas de pressão sobre a membrana basilar criam ondulações que estimulam os órgãos de Corti.

Os órgãos de Corti possuem uma estrutura semelhante à membrana basilar, constituídos por aproximadamente 4 mil células ciliadas internas e 16 mil externas, dispostas em fileiras. Estas células estão assentadas na membrana basilar e possuem prolongamentos denominados cílios, que se apoiam na membrana tectorial. Elas captam os movimentos da membrana (basilar e tectorial) e passam os sinais às células nervosas (neurônios). Os órgãos de Corti

são vistos como transdutores eletromecânicos que transformam os movimentos vibratórios das membranas basilar e tectorial em sinais elétricos (ALSHAIB; AL-KANDARI; HASAN, 2015). No decorrer deste processo, a vibração dos cílios, provocada pelo deslocamento relativo entre as membranas basilar e tectorial, possibilita a entrada na célula de íons presentes na solução aquosa presente na perilinfa externa. A consequência é uma alteração do potencial elétrico celular, provocando disparos nas terminações nervosas. Isso produz uma representação do nervo auditivo no tempo e na frequência (RABINER; SCHAFER, 2013).

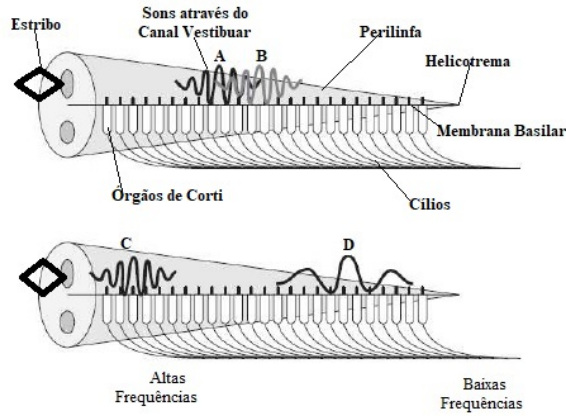
Em resumo, a cóclea opera como um analisador de espectro, ou um banco de filtros seletivos distribuídos. Assim, um ouvido normal é capaz de distinguir sons com frequências entre 20 Hz e 20 kHz, embora a resposta para altas frequências diminua acentuadamente com o aumento da idade de um ser humano.

As células ciliares externas influem na seletividade total da cóclea, através de um processo de realimentação ativa, aumentando a seletividade da resposta neural. A seletividade aparente do ouvido humano é da ordem de 100 Hz para frequências de até 500 Hz, e entre $1/6$ e $1/3$ de oitava para frequências superiores. Essas faixas de seletividade são denominadas bandas críticas, e estão relacionadas com a capacidade do ouvido em discernir sons de amplitudes e frequências diferentes.

A excitação da membrana basilar por sons de diferentes frequências é exposta na Figura 6. As ondas A e B competem pelas mesmas células ciliadas de mesma banda crítica, as ondas C e D excitam células ciliadas de diferentes bandas críticas.

Para Rabiner e Schafer (2013), a análise de frequência não uniforme realizada pela membrana basilar pode ser equivalentemente pensada como um conjunto de filtros de banda não ideais cujas respostas se tornam cada vez mais amplas com o aumento da frequência, como pode ser visto na Figura 6. Georg von Békésy, em 1960, foi naturalmente um defensor convicto da teoria do lugar da audição, sustentando que “o tom é determinado pelo lugar da excitação máxima ao longo da membrana basilar”.

Figura 6: Estrutura coclear no formato linear.



Fonte: Adaptada de Fastl e Zwicker (2007).

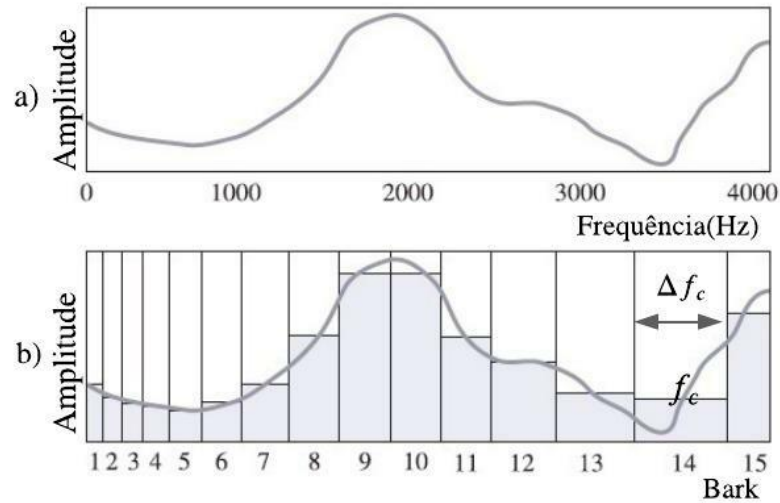
Uma forma de entender como o sistema auditivo funciona é considerar que ele contém um banco de filtros passa-faixa. Tais filtros são produzidos ao longo da membrana basilar e têm a função de aumentar a resolução de frequência da cóclea, ampliando a capacidade de distinguir diferentes sons (MADISETTI, 2010). Este banco de filtros não possui uma configuração linear e a largura de banda, chamada de banda crítica, e a morfologia de cada filtro depende da sua frequência central (MCLOUGHLIN, 2014). Como pode ser visto na Figura 7, há uma largura de banda crítica para cada frequência e as medições empíricas são dadas por (HILL, 2019)

$$\Delta f_c = 25 + 75 \left[1 + 1,4 \left(\frac{f_c}{1000} \right)^2 \right]^{0,69} \quad (13)$$

onde Δf_c é a largura de banda crítica associada à frequência central f_c .

Aproximadamente 25 filtros de banda crítica abrangem a faixa de 0 Hz a 20 kHz para audição humana. Na Tabela 2 é possível observar que as bandas críticas têm larguras constantes para frequências centrais de 100 Hz a 500 Hz. E, partir daí, a largura de banda aumenta a medida que a frequência central aumenta.

Figura 7: Bandas críticas ideais para o ouvido humano: a) espectro do sinal de voz na frequência e b) distribuição do sinal em 15 bandas críticas na escala Bark.



Fonte: Adaptado de McLoughlin (2014).

A unidade Bark, que leva este nome em homenagem a Heinrich Barkhausen, é calculada pela equação (14) (HILL, 2019)

$$\Omega(\omega) = 6 \log \left(\frac{\omega}{1200\pi} + \sqrt{(\omega/1200\pi)^2 + 1} \right) \quad (14)$$

onde Ω é a unidade Bark e ω é a frequência angular.

Tabela 2: Frequências centrais e limítrofes das bandas críticas

Banda (Bark)	Limite Inferior (Hz)	Central (Hz)	Limite Superior (Hz)	Banda (Bark)	Limite Inferior (Hz)	Central (Hz)	Limite Superior (Hz)
1	0	50	100	14	2000	2150	2320
2	100	150	200	15	2320	2500	2700
3	200	250	300	16	2700	2900	3150
4	300	350	400	17	3150	3400	3700
5	400	450	510	18	3700	4000	4400
6	510	570	630	19	4400	4800	5300
7	630	700	770	20	5300	5800	6400
8	770	840	920	21	6400	7000	7700
8	920	1000	1080	22	7700	8500	9500
10	1080	1170	1270	23	9500	10500	12000
11	1270	1370	1480	24	12000	13500	15500
12	1480	1600	1720	25	15500	19500	
13	1720	1850	2000				

Fonte: Elaborada pelo próprio autor.

Tal como acontece com a escala Bark, também existem alguns métodos alternativos para converter Hz em escala Mel. Em geral, todos os métodos adotam 1 kHz equivalente a 1000 mels. Duas escalas Mel são popularmente utilizadas (HILL, 2019):

$$m(f) = 2595 \times \log(1 + f/1000) \quad (15)$$

$$m(f) = (1000 / \log(2)) \times \log(1 + f/1000) \quad (16)$$

sendo a equação (15) o modelo de O'SHAUGHNESSY (1987) e a (16) é o modelo de FAN (1968).

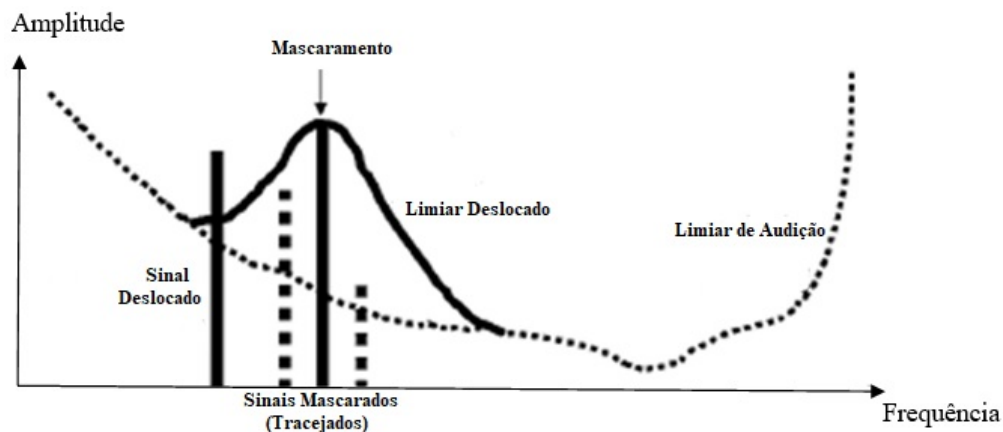
Os modelos de bandas críticas são muito importante na compreensão de fenômenos como percepção de volumes, tons e mascaramento, e, portanto, fornecem motivação para representações digitais do sinal de voz que são baseadas em uma decomposição em frequência.

O fenômeno relacionado ao mascaramento também é atribuído às vibrações da membrana basilar. Quando dois sons puros excitam regiões em sua proximidade, as células ciliadas desta área emitirão os impulsos referentes ao som de maior amplitude sonora. Desta forma, o cérebro interpretará apenas o som de maior amplitude, ou seja, não se ouvirá o som de menor amplitude. Assim, o som de menor amplitude sonora foi mascarado pelo de maior.

A noção de que um som se torna inaudível pode ser quantificada em relação ao limiar de sonoridade definido pelas curvas Fletcher e Munson, representadas na Figura 4. O mascaramento é um fenômeno que altera os limiares de audibilidade pela presença de outros tons, causado pelo deslocamento da membrana basilar.

Observa-se na Figura 8 que um tom intenso (chamado de máscara) tende a aumentar o limiar de sonoridade em torno de sua localização no eixo da frequência, como mostrado pela linha sólida. Todos os componentes espectrais cujo nível está abaixo desse limite elevado são mascarados e, portanto, não precisam ser reproduzidos em um sistema de processamento de voz porque não seriam ouvidos. Porém, qualquer componente espectral cujo nível esteja acima do limite elevado não é mascarado e, portanto, será ouvido (RABINER; SCHAFER, 2013).

Figura 8: Ilustração do efeito do mascaramento simultâneo ou em frequência

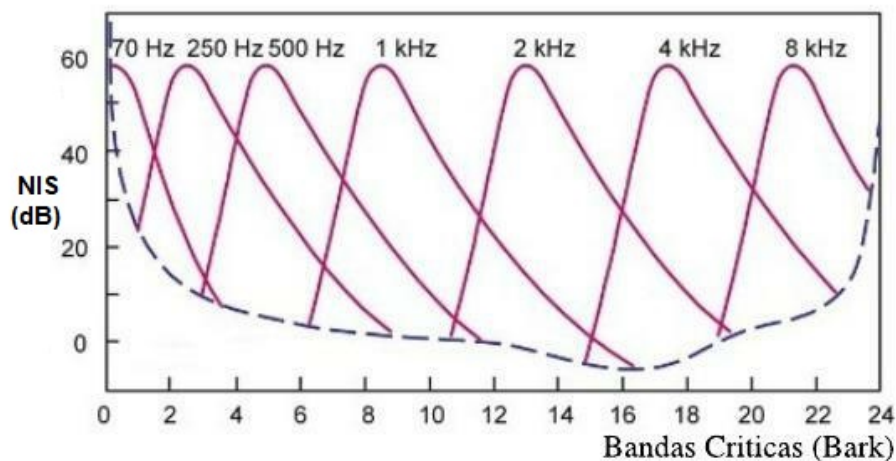


Fonte: Adaptado de Herre e Dick (2019).

Os sons mascarados são todos os sons que se tornam inaudíveis devido a adaptação dinâmica do limiar de sonoridade.

A largura da máscara é dependente da frequência do som que a provoca e aumenta com a frequência. Na Figura 9 é mostrado o ruído de banda estreita com mascarador de 60 dB, para as frequências destacadas, com suas respectivas bandas críticas.

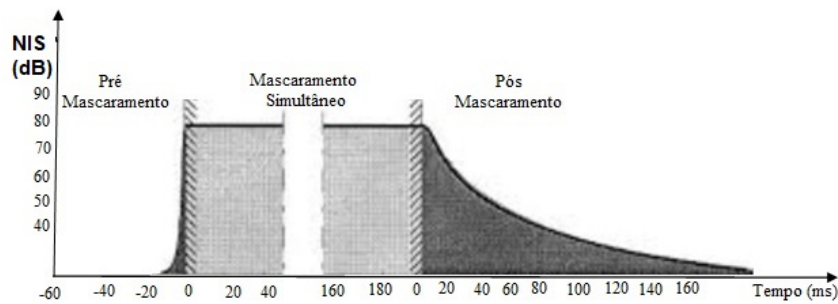
Figura 9: Mascaramento temporal.



Fonte: Adaptado de Gelfand (2010).

Os efeitos de mascaramento podem ser representados pela Figura 10. O mascaramento de sons que chegam ao ouvido humano ao mesmo tempo que um som mascarador é definido como “mascaramento simultâneo”.

Figura 10: Mascaramento simultâneo.

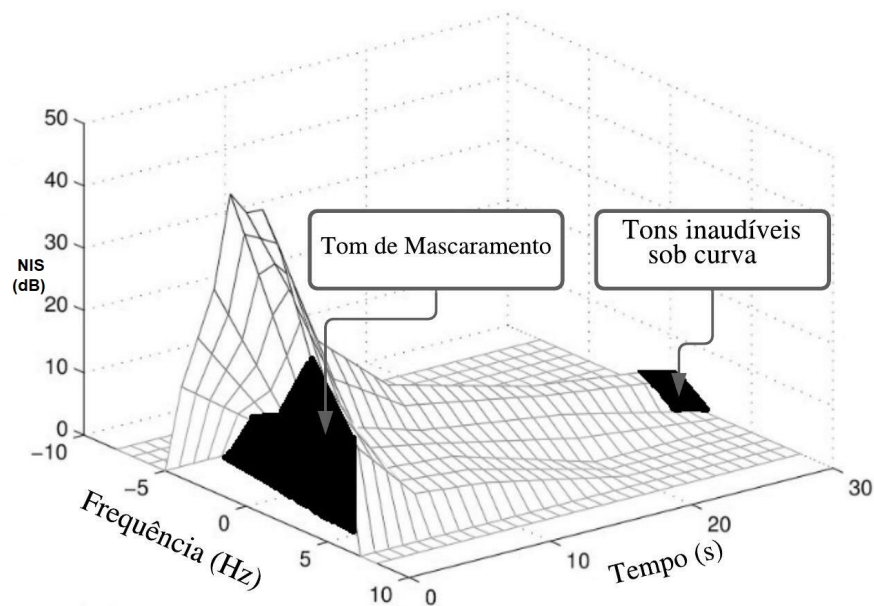


Fonte: Adaptado de Herre e Dick (2019).

Explorando os mascaramentos em tempo e frequência é possível reduzir substancialmente a informação de áudio, sem uma mudança audível, como pode ser visto na Figura 11.

O nível de mascaramento de um som por outro dependerá essencialmente da proximidade dos dois em termos de frequência, como pode ser visto na Figura 11. O efeito de mascaramento diminui mais lentamente para a direita da curva, ou seja, em direção às altas frequências.

Figura 11: Combinação de mascaramentos em tempo e frequência.



Fonte: Adaptado de Necciari (2016).

Quando dois sinais se situam dentro de uma banda crítica, o de maior energia, poderá dominar a percepção e mascarar o outro. Portanto, dependendo dos níveis, dois tons distintos somente serão distinguidos um do outro quando estiverem em bandas críticas diferentes. A detecção de um estímulo por meio de um mascaramento simultâneo dependerá de uma boa resolução espectral, ou seja, da capacidade de perceber e separar a frequência alvo em meio à frequência do estímulo mascarador (ALSHAIB; AL-KANDARI; HASAN, 2015). Já em uma tarefa de mascaramento temporal, há dependência da resolução temporal, ou seja, da capacidade de perceber o intervalo de tempo entre a apresentação do estímulo alvo e do estímulo mascarador, identificando-os como dois estímulos separados, como já apresentado na Figura 10.

O efeito de mascaramento é fortemente influenciado pelo espectro de frequência dos dois sons. Por este motivo, a pura análise por transformadas como Fourier ou *Wavelet* não expressa exatamente como as amostras são percebidas. No entanto, modelagens psicoacústicas podem ser utilizadas para criar representações mais reais da percepção sonora.

3.1.3 MODELO PSICOACÚSTICO

A principal função do modelo psicoacústico, enquanto parte integrante da arquitetura de um codificador perceptivo, é controlar o erro de quantificação introduzido em cada banda crítica do ouvido. O tom mais alto em cada banda crítica é audível, e o efeito de mascaramento é o um ponderador de todos os componentes de som e ruído dentro dessa banda. Embora isso não seja levado em conta para vários fatores auditivos, ele modela a situação geral muito bem, especialmente quando há uma distinção clara entre o sinal desejado e o ruído que interfere na integridade (MCLOUGHLIN, 2014). Existem muitos modelos acústicos que foram criados a partir dos fenômenos descritos na Seção 3.1.2 que procuram quantificar alguns ou todos os efeitos dessas bandas críticas que ficaram conhecidos por muitos nomes, sendo alguns: funções de propagação de banda crítica, curvas de mascaramento de ruído e funções de inibição. Sendo que estes têm sido de particular interesse para as pesquisas envolvendo a psicoacústica. Nesta pesquisa adota-se para análises e simulações o modelo de propagação de banda crítica de Hermansky (1990) e o modelo MFCC - Coeficientes Cepstrais

da Frequência Mel, por se tratarem de duas escalas diferentes mas com o objetivo comum de modelar o sinal de voz. Ambos os modelos serão descritos a seguir.

A curva da função de espalhamento de Hermansky tem uma resposta plana que não apenas aproxima a largura de cada banda crítica, mas é responsável pela propagação com uma dependência na frequência absoluta.

O modelo Hermansky é calculado primeiramente no domínio da frequência, porém em vez de ter unidades em Hz, usa-se uma escala de Bark dada pela equação (14). Em outras palavras, o índice espectral é representado em Barks. Então, o efeito dos filtros de banda crítica pode ser calculado usando uma função de espalhamento no domínio Bark $\Psi(\Omega_\delta)$, definida da seguinte forma (HERMANSKY, 1990):

$$\Psi(\Omega_\delta) = \begin{cases} 0 & \text{se } \Omega_\delta < -1,3 \\ 10^{2,5(\Omega_\delta+0,5)} & \text{se } -1,3 \leq \Omega_\delta \leq -0,5 \\ 1 & \text{se } -0,5 < \Omega_\delta \leq 0,5 \\ 10^{-1,0(\Omega_\delta-0,5)} & \text{se } 0,5 < \Omega_\delta \leq 2,5 \\ 0 & \text{se } \Omega_\delta > 2,5 \end{cases} \quad (17)$$

onde, $\Omega_\delta = \Omega_c - \Omega_i$ é a frequência no domínio Bark, Ω_c é a banda crítica e Ω_i é a frequência de interesse.

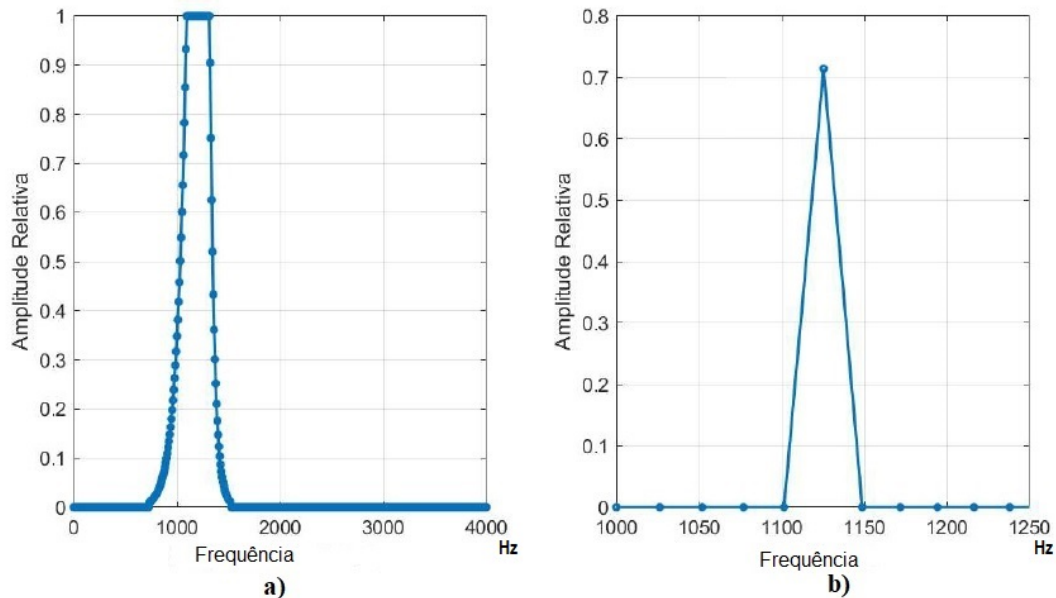
A função de espalhamento no domínio Mel mais comum usa uma forma triangular, onde cada área triangular é 50% sobreposta às suas vizinhas. Cada filtro triangular $M_m(f)$ é calculado como descreve a equação (18)

$$M_m(f) = \begin{cases} 0 & \text{se } f < F(m-1) \\ \frac{f-F(m-1)}{F(m)-F(m-1)} & \text{se } F(m-1) \leq f \leq F(m) \\ \frac{F(m+1)-f}{F(m+1)-F(m)} & \text{se } F(m) < f \leq F(m+1) \\ 0 & \text{se } f > F(m+1) \end{cases} \quad (18)$$

sendo $F(m)$ a frequência em Hertz do m -ésimo filtro, avaliado na frequência f . Assim como no filtro de Hermansky, f é varrido em cada faixa para avaliar a resposta em frequência de cada uma das m bandas.

As funções de espalhamento na escala Bark e Mel dadas pelas equações (17) e (18) são usadas para modelar o efeito de uma única banda crítica como pode ser visto na Figura 12. Normalmente, elas são usadas para ajudar a representar o som em uma forma perceptiva, como um vetor que corresponde às excitações de banda crítica (MCLOUGHLIN, 2014).

Figura 12: Bandas críticas dos modelos: a) Hermansky e b) MFCC.



Fonte: Elaborada pelo próprio autor.

Para justificar este efeito, considera-se que um único segmento de sinal de voz pode ser representado de várias maneiras alternativas. Por exemplo, um único segmento de áudio gravado começa como um vetor de amostras extraídas de uma gravação mais longa. Estes vetores possuem as mesmas informações armazenadas no domínio do tempo.

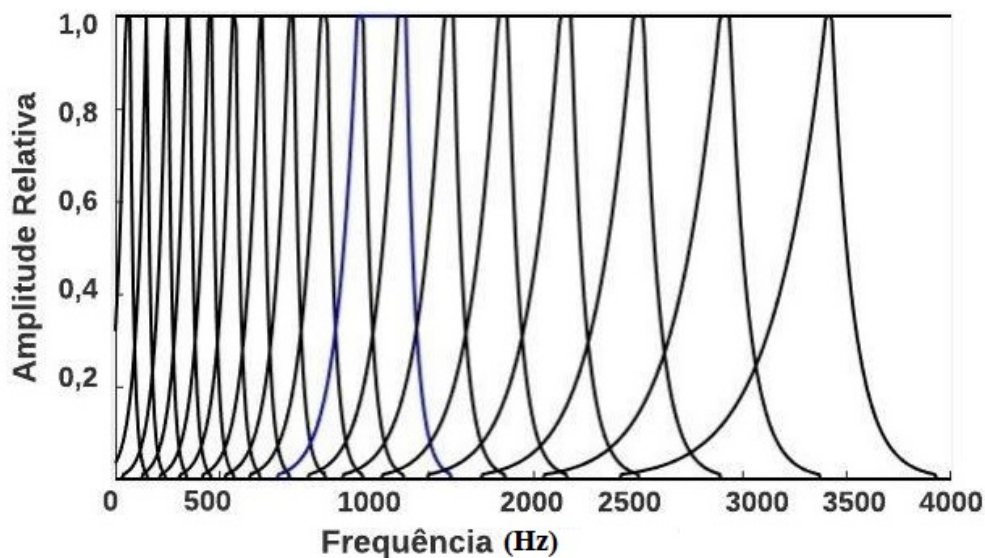
Cabe destacar que o domínio da frequência é mais útil para algumas aplicações, enquanto outras podem usar melhor a representação no domínio do tempo, mas ambos são de natureza física. A ideia de um banco de filtros de banda crítica é fornecer uma representação perceptual sobre as informações extraídas do sinal, ou seja, um vetor de informações que melhor represente o sinal para um ouvinte.

Para construir este vetor perceptual é preciso uma representação do sinal no domínio da frequência, e então identificar quais frequências contribuem para quais bandas críticas.

As bandas críticas são normalmente calculadas utilizando modelos na escala Bark com a

função de Hermansky, ou no domínio Mel como uma função triangular, ilustrada na Figura 12. Para obter κ elementos do banco de filtros, primeiro crie τ bandas críticas sobrepostas e, em seguida, convolva os elementos espectrais de potência nas respectivas bandas, somando suas contribuições por banda. Para fins de ilustração, na Figura 13 é apresentado um gráfico das funções de espalhamento do modelo de Hermansky.

Figura 13: Gráfico da função de propagação de banda crítica, com base na equação (17).



Fonte: Elaborada pelo próprio autor.

Para McLoughlin (2014), um segmento de fala pseudo-estacionária amostrado em 8 kHz seria normalmente representado com 26 a 40 bandas nos respectivos bancos de filtros.

No modelo proposto nesta pesquisa, os sinais de voz da base de dados da NOIZEUS serão explorados para as simulações, correspondentes ao modelo psicoacústico de Hermansky(1990) na escala Bark e os MFCC na escala Mel.

3.2 IMPLEMENTAÇÃO DO MODELO PSICOACÚSTICO

Modelos perceptivos baseados em funções de propagação de banda crítica envolvem vários estágios do processamento auditivo. A maioria dos modelos psicoacústicos, inclusive os utilizados neste trabalho, seguem as seguintes etapas:

1. análise espectral;

- 2. curvatura de banda crítica e convolução da função de espalhamento;
- 3. conversão pré-ênfase de intensidade - volume.

Antes de apresentar os modelos de Hermansky (1990) na escala Bark e o modelo MFCC na escala Mel serão destacadas as etapas deste processo.

1º Etapa: Análise Espectral

Uma vez que o efeito de mascaramento simultâneo é definido em termos de diferenças entre frequências, o primeiro passo é selecionar um quadro de áudio para analisar, criar uma janela e converter para o domínio da frequência.

A solução comum para muitos dos problemas de segmentação em áudio é o janelamento suave com sobreposição, cada janela possui amplitudes iniciais e finais próximas de zero e são deslocadas ao longo do tempo (PODDER; KHAN; KHAN; RAHMAN, 2014). Considere o k -ésimo quadro do sinal de voz, denotado pelas amostras $s(l, k)$ com $l = 1, 2, \dots, L$ sendo $L = 256$. Assim o sinal segmentado e janelado é dado por

$$s_w(l, k) = h(l)s(l, k) \quad (19)$$

onde $h(l)$ é a função janela. Existem diferentes possibilidades de janelas suavizadas (Hamming, Hanning, Kaiser, etc...), mas neste trabalho adotou-se a janela de Hamming dada por:

$$h(l) = 0,54 - 0,46\cos(2\pi l/L), \quad 1 \leq l \leq L. \quad (20)$$

O modelo perceptivo, emprega o algoritmo FFT para implementar uma transformada de Fourier discreta no tempo - DFT (do inglês: *Discrete Fourier Transform*) (OPPENHEIM; SCHAFER; BUCK, 1998).

O sinal de áudio segmentado e janelado é convertido para o domínio da frequência usando um número finito de pontos para a DFT. Neste caso, considerando um número finito L , tem-se:

$$E(f, k) = \frac{1}{L} \sum_{l=1}^L s_w(l, k) e^{-j2\pi kl/L} \quad (21)$$

2º Etapa: Curvatura de banda crítica e convolução da função de espalhamento

Utilizando os modelos de banda crítica baseados nas funções de espalhamento dadas pelas equações (17) e (18), já destacados, os bancos de filtros são convoluídos com a representação espectral do sinal que está sendo analisado para calcular as respostas por banco.

Esse procedimento será utilizado para construir o modelo perceptivo envolvendo o Hermansky(1990) na escala Bark e o modelo MFCC na escala Mel.

3º Etapa: conversão/ pré-ênfase de intensidade - volume

A seletividade de frequência base da audição humana é geralmente aproximada pela curva de ponderação vista na Figura 4, que por si só foi bem aproximada no domínio da frequência. Hermansky (1990) usou $\omega = 2\pi f$ como frequência angular a partir da frequência linear f em seu modelo cuja a equação é dada por

$$E(w) = \frac{\omega^4(\omega^2 + 56,8 \times 10^6)}{(\omega^2 + 6,3 \times 10^6)^2(\omega^2 + 0,38 \times 10^9)}. \quad (22)$$

As curvas de ponderação (22) são ajustadas para corrigir a sensibilidade do ouvido humano a diferentes frequências integradas ao nível de pressão sonora.

Também deve ser observado, a partir da fonte original, que a equação (22) é precisa até cerca de 5 kHz (que está abaixo das frequências utilizadas neste trabalho) (MCLOUGHLIN, 2014). No entanto, para frequências mais altas é aplicada a equação (23), para melhor corresponder à frequência superior na região da curva de ponderação.

$$E(w) = \frac{\omega^4(\omega^2 + 56,8 \times 10^6)10^{27}}{(\omega^2 + 6,3 \times 10^6)^2(\omega^2 + 0,38 \times 10^9)(\omega^6 + 9,58 \times 10^{26})}. \quad (23)$$

Isso é seguido por uma conversão de intensidade-volume para permitir que as bandas calculadas possam se relacionar com as unidades quase arbitrárias do modelo ligadas a uma escala de intensidade percebida, com base na lei de potência da audição (MCLOUGHLIN, 2014). No entanto tal conversão não é necessária quando o modelo é usado para comparar dois sinais diretamente. A resposta perceptual final, é dada pela lei de potência de intensidade sonora $P(\omega)$, representada pela equação volume (MCLOUGHLIN, 2014):

$$P(\omega) = \{E(\omega)\Theta(\omega)\}^{0,33} \quad (24)$$

onde $\Theta(\omega)$ é o espectro resultante da convolução.

Para Mcloughlin (2014), o modelo de domínio Mel utiliza uma conversão logarítmica que é realizada antes da convolução das funções de propagação para definir os MFCC.

3.2.1 IMPLEMENTAÇÃO DO MODELO PERCEPTIVO DE HERMANSKY

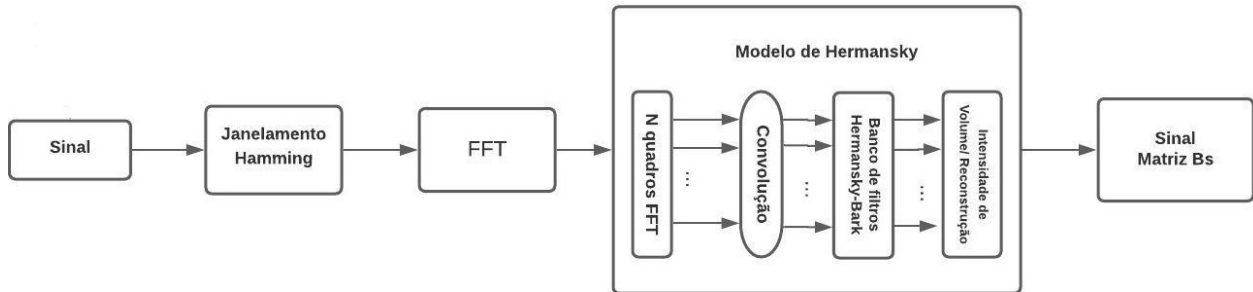
Para obter um modelo perceptivo de Hermansky, como pode ser visto na Figura 14, primeiro se define uma matriz com as frequências centrais. Em seguida, criam-se as funções de propagação convoluindo com o espectro de potência, somando a contribuição dentro de cada banda crítica. As conversões intensidade-volume vêm a seguir para criar uma resposta perceptual no domínio de Bark que pode ser resumida da seguinte forma

$$B_{S(\Psi,k)} = \sum_{t=bm_z}^{bn_z} P(t,k) \quad (25)$$

onde bm_z e bn_z são os limites inferior e superior da banda crítica Ψ , respectivamente.

Dado um segmento de voz, com frequência de amostragem F_s em Hz, a função do modelo perceptual de Hermansky calcula a resposta espectral de N amostras, conforme o esquema do modelo apresentado na Figura 14.

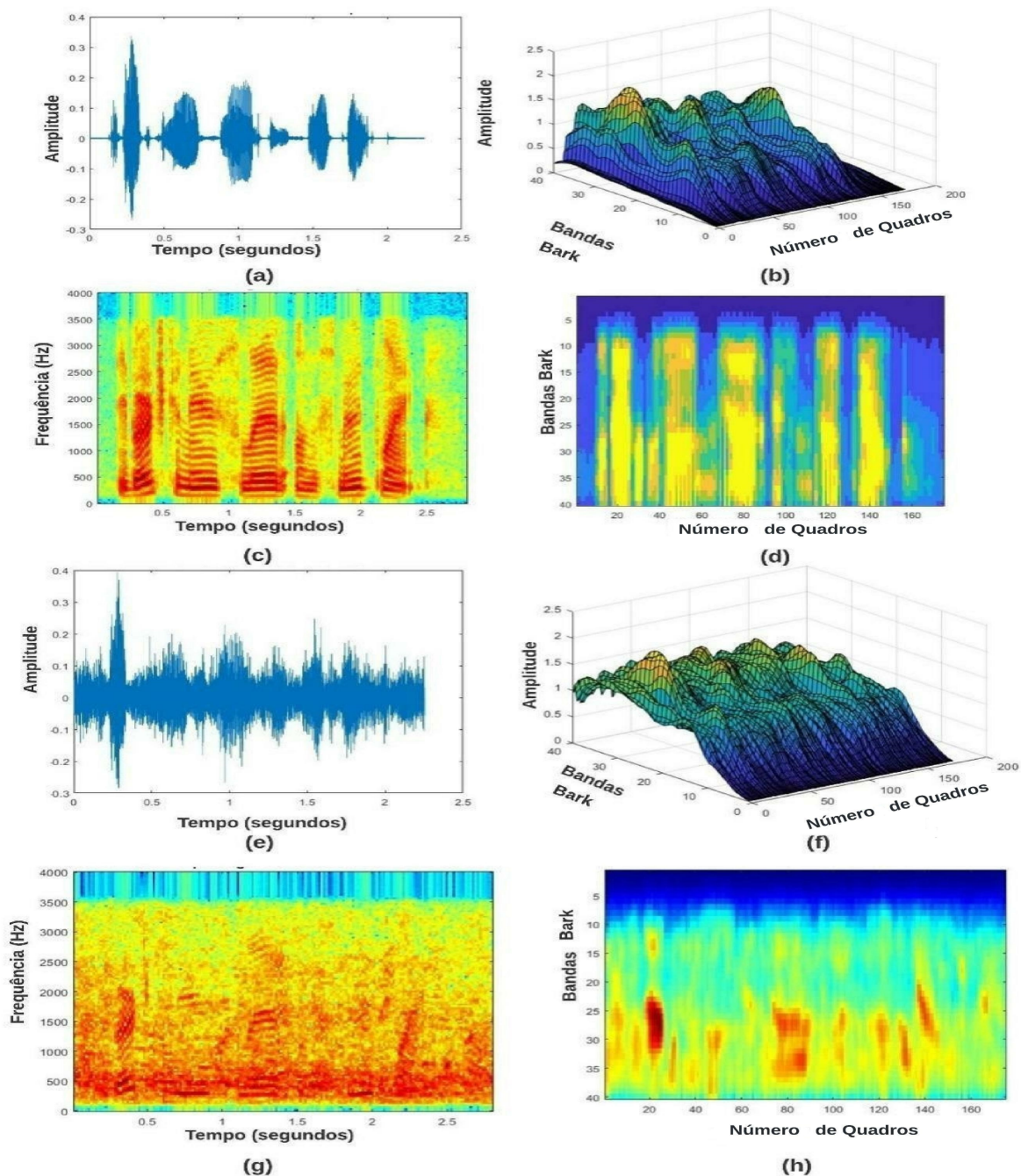
Figura 14: Fluxograma do modelo perceptual de Hermansky .



Fonte: Elaborada pelo próprio autor.

Na prática, a resposta seria calculada quadro a quadro (provavelmente usando quadros sobrepostos) em toda gravação de áudio. Na Figura 15 é apresentada a análise de um sinal limpo e um sinal contaminado com ruído de multidão a 0 dB de SNR, utilizando 40 filtros de banda crítica na escala Bark, seguindo o modelo de Hermansky. As Figuras 15d e h representam o plano Bandas Bark $\times$ Quadros da representação espacial das Figuras 15b e f, respectivamente. As Figuras 15d e h são as representações por imagens das matrizes de percepção auditiva do modelo de Hermansky. Elas são essenciais para o desenvolvimento do método proposto, em ambas, as cores azuladas significam trechos de silêncio, as cores amareladas, atividade de voz e as cores avermelhadas misturadas ao amarelo, como pode visto na Figura 15h, indicam a presença de ruído adicionado ao sinal.

Figura 15: Forma de onda do sinal de voz: (a) sinal limpo e (e) sinal contaminado por ruído de multidão com 0 dB; escala espectral do modelo Hermansky: (b) sinal limpo e (f) sinal ruidoso; espectrogramas: (c) sinal limpo e (g) sinal ruidoso; plano quadro $\times$ bandas Bark resposta do modelo Hermansky: (d) sinal limpo e (h) sinal ruidoso.



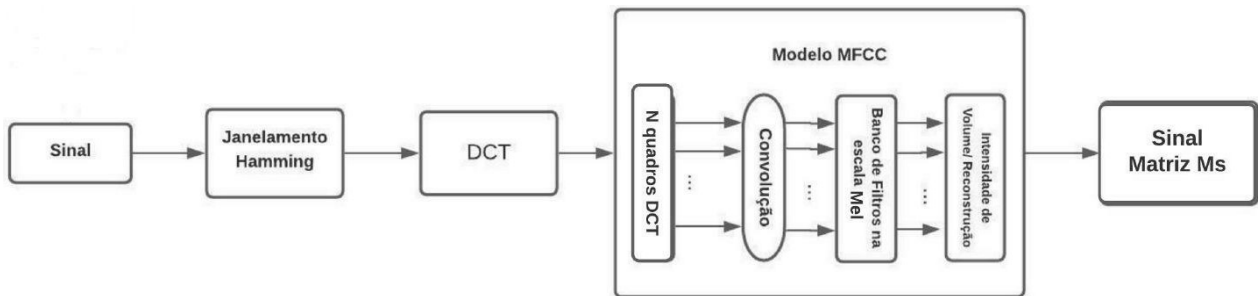
Fonte: Elaborada pelo próprio autor.

3.2.2 IMPLEMENTAÇÃO DO MODELO PERCEPTIVO MFCC

O modelo MFCC representado pela escala Mel é definido por frequências centrais denominadas funções de espalhamento triangular, ou seja, um banco de filtros Mel. Neste, representado pela Figura 16, aplica-se uma convolução com o espectro de potência do segmento de sinal de voz, para que a contribuição de cada banco de filtros seja somada para produzir um vetor de potência. A conversão intensidade-volume é realizada transformando o vetor de potência para a escala logarítmica, antes obtida pela Transformada Discreta de Cosseno - DCT, dada pela equação (26) para produzir os MFCC finais.

$$S(f, k) = \alpha[s] \sum_{l_c=0}^{L_c-1} H(l)s(l) \cos \frac{\pi(2l_c + 1) \cdot k}{2L_c} \quad \text{onde} \quad \alpha[s] = \begin{cases} \sqrt{\frac{1}{l_c}} & \text{para } k = 0 \\ \sqrt{\frac{2}{l_c}} & \text{para } k = 1, 2, \dots, L_c - 1 \end{cases} \quad (26)$$

Figura 16: Fluxograma do modelo perceptual MFCC.

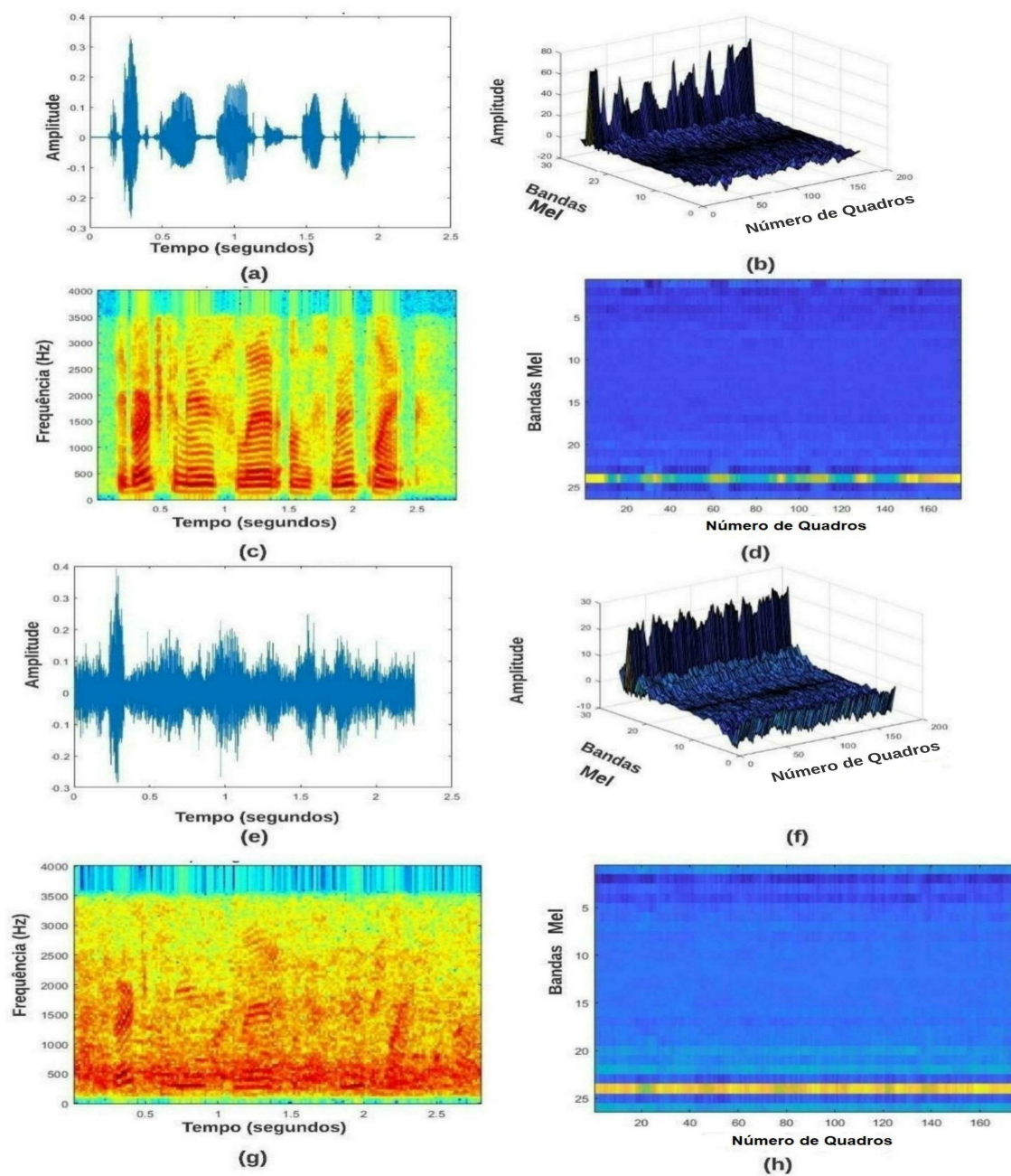


Fonte: Elaborada pelo próprio autor.

Dado um segmento de voz, com F_s dada em Hz, a função do modelo perceptual MFCC calcula a resposta espectral de L_c pontos da DCT.

Na Figura 17 é apresentado o comportamento do modelo MFCC para 40 bandas críticas baseadas na escala Mel analisando um sinal limpo e um sinal contaminado com ruído de multidão a 0 dB de SNR. As Figuras 17d e h são de interesse para esta pesquisa, como destacado na seção anterior.

Figura 17: Forma de onda do sinal de voz: (a) sinal limpo e (e) sinal contaminado por ruído de multidão com 0 dB; escala espectral do modelo MFCC: (b) para sinal limpo e (f) sinal ruidoso; espectrogramas: (c) sinal limpo e (g) sinal ruidoso; plano quatro $\times$ bandas Mel resposta do modelo MFCC: (d) sinal limpo e (h) sinal ruidoso.



Fonte: Elaborada pelo próprio autor.

Uma série de conceitos discutidos aqui serão abordados em capítulos posteriores, especialmente no próximo capítulo que tratará do processo de extração de fatores dos sinais de voz obtidos pela densidade espectral da distribuição no domínio das percepções auditivas que fazem parte do modelo proposto neste trabalho.

EXTRAÇÃO DE FATORES DO SINAL DE VOZ

O objetivo deste Capítulo é apresentar as etapas do processo de extração de fatores do sinal de voz, componente essencial para o método proposto.

As matrizes Bs ou Ms , dadas nas seções 3.2.1 e 3.2.2, resultantes dos modelos perceptivos são consideradas vetores bidimensionais, em que, cada uma de suas células ou pixels contém informações importantes do sinal de voz. Soni e Patil (2021) e Dong e Williamson (2019) apresentam um modelo para avaliar sinais de voz partindo diretamente do espectro, sendo as análises realizadas no domínio dos números complexos. No trabalho de Chen e Parsa (2008) é realizado um processo para gerar matrizes com o objetivo de obter valores reais na sua composição apresentando uma técnica para construção perceptiva do sinal. Para McLoughlin (2014), as informações do espectro de potência (número de quadros $\times$ bandas) geradas pelo modelo perceptivo expressam as características fundamentais do sinal e são utilizadas para análises em diversos tipos de aplicações. Neste trabalho, as imagens são geradas a partir de dois modelos perceptivos, fornecendo valores reais e contundentes na extração de fatores.

Os três processos, quantização, matriz de coocorrência e extração de fatores constituem uma área de pesquisa de tratamento de imagem e serão apresentados, nesta ordem, no decorrer deste capítulo e suas etapas são mostradas no diagrama de blocos da Figura 18.

Figura 18: Diagrama de blocos da extração de fatores.



Fonte: Elaborada pelo próprio autor.

4.1 QUANTIZAÇÃO

Na área de processamento digital de imagens diversos métodos de extração de características ou atributos da imagem são utilizados, dentre eles se destacam: bordas, texturas e vizinhança. Uma técnica estatística para determinação de textura será aplicada neste trabalho com o objetivo de extrair características do modelo perceptivo.

Para Armi e Ershad (2019), a textura é reconhecível de maneiras táteis e ópticas. A textura tátil se refere à sensação tangível de uma superfície e a textura visual se refere a ver a forma ou o conteúdo da imagem. O diagnóstico de textura em um sistema de visão humana é facilmente viável, mas no domínio da visão de máquina e no processamento de imagens tem sua própria complexidade (ARMI; ERSHAD, 2019). No processamento de imagens, a textura pode ser definida em função da variação espacial da intensidade de brilho dos pixels (CHAKI; DEY, 2020). A textura representa as variações de cada nível, medindo características como suavidade, maciez, aspereza e regularidade de cada superfície em diferentes direções e ordens. Para Chaki e Dey (2020), as texturas no processamento e visão de máquina referem-se às imagens nas quais um padrão específico de distribuição e dispersão da intensidade da iluminação do pixel é repetido sequencialmente ao longo da imagem. Assim, um conjunto conectado de pixels que satisfaz uma propriedade específica de nível de cinza que ocorre constantemente em uma região da imagem é um área texturizada.

A matriz de percepção auditiva do sinal de voz é o resultado dos modelos de Hermansky no plano tempo Bark e MFCC no plano tempo Mel, como visto no Capítulo 3. Para que o modelo apresente um desempenho satisfatório, na etapa de extração de fatores, deve-se definir quantos níveis de cinza serão necessários para representar um conjunto de textura com sucesso. Assim, para calcular a matriz de coocorrência dos espectros de potência distribuídos no plano da percepção do tempo, uma quantização linear deve ser realizada para converter os valores dos espectros de potência em níveis de cinza, N_c . No processo de quantização linear proposto, o correspondente $Bs_{(\Psi, K)}$, com elementos em $\mathbb{R} \times \mathbb{R}$, representado pela equação (25) se torna $G(n, q)_{Bs}$ que é considerado uma imagem em níveis de cinza no plano $\mathbb{Z}^+ \times \mathbb{Z}^+$ dada pela equação (27).

$$G(n, q)_{Bs} = \frac{Nc - 1}{\max[Bs_{(\Psi, k)}] - \min[Bs_{(\Psi, k)}]} Bs_{(\Psi, k)} + \min[Bs_{(\Psi, k)}] + 1 \quad (27)$$

A equação (27) também pode ser aplicada na equação (18) do modelo MFCC na escala Mel.

Considere a matriz $Bs_{(\Psi, K)}$ das percepções auditivas, resultado do modelo de Hermansky, representada pela equação (28). Aplicando a equação (27) obtém-se a matriz quantizada dada pela equação (29) com valores aproximados para o inteiro mais próximo e de mesma ordem da matriz original.

$$Bs_{(\Psi, K)} = \begin{bmatrix} B(1, 1) & B(1, 2) & \cdots & B(1, K) \\ B(2, 1) & B(2, 2) & \cdots & B(2, K) \\ \vdots & \vdots & & \vdots \\ B(\Psi, 1) & B(\Psi, 2) & \cdots & B(\Psi, K) \end{bmatrix}_{\mathbb{R} \times \mathbb{R}} \quad (28)$$

$$G(n, q)_{Bs} = \begin{bmatrix} B(1, 1) & B(1, 2) & \cdots & B(1, q) \\ B(2, 1) & B(2, 2) & \cdots & B(2, q) \\ \vdots & \vdots & & \vdots \\ B(n, 1) & B(n, 2) & \cdots & B(n, q) \end{bmatrix}_{\mathbb{Z}^+ \times \mathbb{Z}^+} \quad (29)$$

Chaki e Dey (2020) enfatizam que quanto mais níveis de cinza incluídos no cálculo, mais informações são extraídas das texturas. Porém, isso eleva os custos computacionais quando a proposta for extrair informações de um número considerável de imagens para reconhecimento de padrões.

4.2 CÁLCULO DA MATRIZ DE COOCORRÊNCIA DE NÍVEIS DE CINZA - (MCNC)

Um método estatístico de exame de textura que considera a relação espacial dos pixels é a matriz de coocorrência de níveis de cinza (MCNC), também conhecida como matriz de dependência espacial de níveis de cinza ou matriz de ocorrência simultânea, proposta por Haralick *et al.* (1973), composta de 19 medidas de textura com o objetivo de fornecer as características utilizadas na classificação de conjuntos diferentes de imagens. A extração de fatores tem por finalidade explorar valores das matrizes originais e buscar sintetizar determinados aspectos da imagem ou região nela contida. Nesta pesquisa as características da imagem foram denominadas fatores de Haralick e cada nome tem seu significado quando considerado sob o aspecto sensível.

4.3 CONSTRUÇÃO DA MATRIZ DE COOCORRÊNCIA

As funções da MCNC caracterizam a textura de uma imagem, calculando a frequência com que pares de pixels vizinhos se agrupam numa relação espacial (CLAUSI, 2002). A partir do momento que é criada uma MCNC é possível extrair medidas estatísticas dessa matriz ao longo de diferentes direções (CHAKI; DEY, 2020).

No método MCNC, um percurso é definido como uma sequência colinear de pixels adjacentes que possuem os mesmos tons de cinza. Um percurso pode ocorrer nas direções vertical, horizontal ou qualquer inclinação entre dois extremos. Neste trabalho as direções serão: 0° (horizontal), 45° , 90° (vertical) e 135° , conforme mostradas nas Figuras 19a e b.

Na composição de uma matriz de coocorrência dois conceitos são fundamentais: o ângulo θ e a distância D . Matematicamente, Haralick (1979) definiu θ e D por meio de equações. Dessa forma, seja $M(i, j)$ a Matriz de coocorrência que armazena um pixel com tonalidade (i, j) , para $1 \leq i, j \leq N_c$ onde N_c é considerado o maior nível de cinza. Para representar, define-se G como os pares ordenados (i, j) responsável pela transição entre os pixels i e j . A função $G(i, j)$ representa a tonalidade do pixel na posição (i, j) . $M(i, j)$ é definida pela equação (30) como transição entre elementos, a sua forma com direção e ângulo são dadas

pelas equações (31), (32), (33) e (34), em todas as equações $\#$ denota o número de elementos no conjunto Haralick (1979).

$$M(i, j) = \#\{[(a, b), (c, d)] \in V : G(a, b) = i \text{ e } G(c, d) = j\} \quad (30)$$

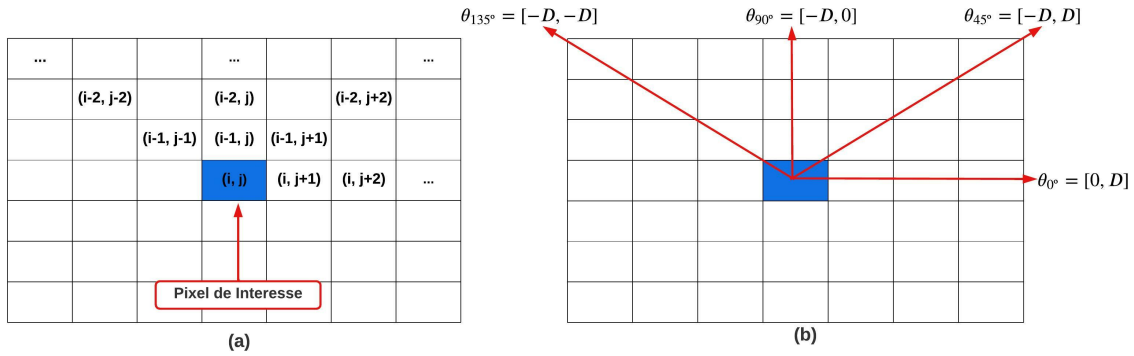
$$M(i, j, D, 0^\circ) = \#\{[(a, b), (c, d)] \in V : |a - c| = D, b - d = 0, G(a, b) = i \text{ e } G(c, d) = j\} \quad (31)$$

$$M(i, j, D, 45^\circ) = \#\{[(a, b), (c, d)] \in V : a - c = -D, |b - d| = D, G(a, b) = i \text{ e } G(c, d) = j\} \quad (32)$$

$$M(i, j, D, 90^\circ) = \#\{[(a, b), (c, d)] \in V : a - c = 0, |b - d| = D, G(a, b) = i \text{ e } G(c, d) = j\} \quad (33)$$

$$M(i, j, D, 135^\circ) = \#\{[(a, b), (c, d)] \in V : a - c = -D, b - d = -D, G(a, b) = i \text{ e } G(c, d) = j\} \quad (34)$$

Figura 19: Orientações da matriz de coocorrência: a) por células e b) indicam as principais direções.



Fonte: Adaptada de Fanizzi *et al.* (2019).

Os fatores de textura normalmente são obtidos por meio da matriz de coocorrência normalizada, definida na equação (35). Cada elemento da matriz calculada pela equação (30) é dividido pela soma de seus componentes.

$$P_{(i,j,D,\theta)} = \frac{M(i,j,D,\theta)}{\sum_{i,j=1}^{N_{lc}} M(i,j,D,\theta)} \quad (35)$$

onde i é número de linhas, j é o número de colunas, $M(i,j)$ é o conteúdo da célula, $P(i,j)$ é a probabilidade da célula (i,j) e N_{lc} é o número de linhas ou colunas, uma vez que M é uma matriz quadrada.

A coocorrência, na sua forma geral, pode ser especificada por uma matriz de frequências relativas $M(i,j;d,\theta)$, na qual dois elementos de textura vizinhos, separados por uma distância d em uma orientação θ ocorrem na imagem, um com propriedade i e o outro com propriedade j . Estendendo essa definição para coocorrência de níveis de cinza, os elementos de textura são pixels e as propriedades são os níveis de cinza. Entretanto, para cada par, distância D e ângulo θ , uma matriz é calculada. Para exemplificar, considere a direção horizontal e a distância unitária. Cada elemento $m_{i,j}$ da matriz $M(1,0^\circ)$ indica quantas vezes a combinação de pixels i e j , vizinhos na horizontal, aparecem na imagem.

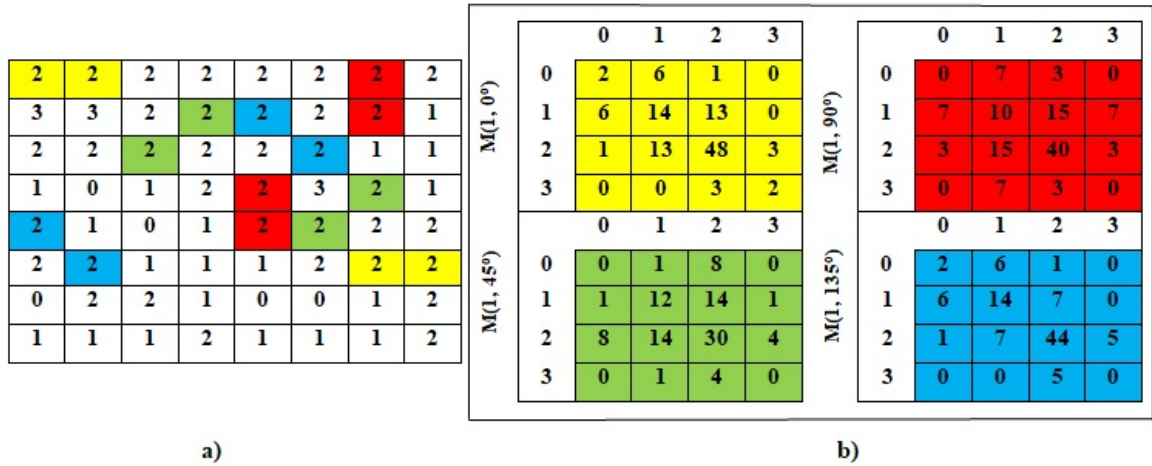
Na Figura 20a são mostrados os valores dos pixels de uma imagem 8×8 . O intervalo de valores para os pixels é $[0,3]$. A dimensão da matriz de coocorrência $M(1,0)$ é 4×4 , já que existem 4 valores de pixels diferentes. Na Figura 20b são apresentados as quatro direções principais da matriz de coocorrência, $M(1,90^\circ)$, $M(1,45^\circ)$, $M(1,0^\circ)$ e $M(1,135^\circ)$ e seu deslocamento unitário.

Os retângulos coloridos da Figura 20a indicam a direção do percurso a ser tomado para compor as matrizes da Figura 20b. A cor amarela indica a direção dos elementos $m(i,j,0^\circ)$ com seus respectivos números de ocorrências, o mesmo ocorre nas direções $m(i,j,45^\circ)$, $m(i,j,90^\circ)$ e $m(i,j,135^\circ)$, com as cores verde, vermelha e azul, respectivamente. Vale ressaltar que as matrizes de coocorrência, por definição, são simétricas, como se pode observar na Figura 20b, as posições (i,j) e (j,i) possuem o mesmo valor, ou seja, para o cálculo da distância não se leva em consideração o sentido em que ela é medida. Entretanto, os elementos da diagonal principal são dobrados na contagem. Algumas propriedades importantes da matriz de coocorrência são:

- toda matriz de coocorrência é uma matriz quadrada;

- a ordem da matriz é igual ao nível de quantização da imagem;
- a matriz de coocorrência é simétrica em torno da diagonal principal.

Figura 20: a) Uma imagem quantizada e b) Suas matrizes de coocorrência nas suas principais direções .



Fonte: Elaborada pelo próprio autor.

Vale ressaltar que a matriz de percepção auditiva proposta neste trabalho apresenta variações nos números de colunas para cada sinal. Mesmo assim, uma MCNC pode ser aplicada, pois sua dimensão não depende do tamanho da matriz sobre a qual ela será calculada (HARALICK, 1979). Ela dependerá unicamente dos níveis de cinza que são definidos para cada aplicação. Uma imagem, por exemplo, que apresenta 8 níveis de cinza produz uma MCNC de dimensão 8×8 .

4.4 CÁLCULO DE FATORES DE TEXTURA

A maioria dos métodos de análise de textura usa estatística de ordem superior e considera a relação entre dois ou mais pixels por vez. Métodos como padrões binários locais, *wavelets* e filtros de Gabor podem ser usados para avaliar a textura da imagem (FSTEDT *et al.*, 2019). Os fatores de Haralick calculados a partir da MCNC são amplamente usados devido à sua praticidade e interpretações intuitivas, e foram aplicados com sucesso em diversas áreas do conhecimento científico (KLEMMT; SEITZ; STRAUB, 2020; GOMES *et al.*, 2020).

É possível medir uma série de fatores de textura da MCNC de duas formas, considerando toda a matriz ou utilizando uma janela que se move ao longo da mesma. Considerando as seguintes notações: N_c é o número de níveis de cinza usados, μ é o valor médio de P , i e j representam linhas e colunas para P , respectivamente. Os μ_x , μ_y , σ_x e σ_y são respectivamente as médias e desvios padrão de P_x e P_y . Estas definições servirão de base para compor as seguintes variáveis:

$$P_x(i) = \sum_{j=0}^{N_c-1} P(i, j) \quad (36)$$

$$P_y(j) = \sum_{i=0}^{N_c-1} P(i, j) \quad (37)$$

$$\mu_x = \sum_{i=0}^{N_c-1} \sum_{j=0}^{N_c-1} i P(i, j) = \frac{1}{N_c} \sum_{i=0}^{N_c-1} i P_x(i) \quad (38)$$

$$\mu_y = \sum_{i=0}^{N_c-1} \sum_{j=0}^{N_c-1} j P(i, j) = \frac{1}{N_c} \sum_{j=0}^{N_c-1} j P_y(j) \quad (39)$$

$$\sigma_x^2 = \sum_{i=0}^{N_c-1} \sum_{j=0}^{N_c-1} (i - \mu_x)^2 P(i, j) = \sum_{i=0}^{N_c-1} (P_x(i) - \mu_x(i))^2 \quad (40)$$

$$\sigma_y^2 = \sum_{j=0}^{N_c-1} \sum_{i=0}^{N_c-1} (j - \mu_y)^2 P(i, j) = \sum_{j=0}^{N_c-1} (P_y(j) - \mu_y(j))^2 \quad (41)$$

$$HXY = - \sum_{i=0}^{N_c-1} \sum_{j=0}^{N_c-1} P_{i,j} \log_2 P_{i,j} \quad (42)$$

$$HXY1 = - \sum_{i=0}^{N_c-1} \sum_{j=0}^{N_c-1} P_{i,j} \log_2 \{P_x(i) P_y(j)\} \quad (43)$$

$$HXY2 = - \sum_{i=0}^{N_c-1} \sum_{j=0}^{N_c-1} p_x(i) p_y(j) \log_2 \{p_x(i) p_y(j)\}. \quad (44)$$

As variáveis e notações dadas pelas equações de (36) a (44), que são utilizadas para compor os fatores de Haralick, são apresentadas na Tabela 3.

Segue uma descrição sucinta de alguns fatores seguindo a ordem mostrada na Tabela 3. A Correlação, fator 5, é utilizada em textura para medir a dependência linear de pixels em relação a sua vizinhança. O Contraste, fator 4, calcula uma estimativa das variações locais ao quadrado dos níveis de cinza entre pares de pixels. Normalmente medidas relacionadas a contraste usam pesos relacionados com a distância da diagonal da MCNC. O fator 9, Energia ou Segundo Momento Angular, determina a uniformidade da textura. Em texturas consideradas ásperas, poucos elementos da MCNC são diferentes de 0 e quando isso ocorre os resultados se aproximam de 1. A Entropia estima a dosagem ou complexidade de uma imagem. A entropia é considerada grande quando a imagem não é texturalmente uniforme e muitos elementos MCNC têm valores muito pequenos. O fator 11, Homogeneidade, é considerado inversamente proporcional aos pesos do contraste, quanto mais distantes da diagonal, menor eles são exponencialmente. A Variância, fator 18, é utilizada na textura se comportando de maneira análoga às aplicações da estatística descritiva, baseando-se na dispersão dos dados numéricos em torno de um valor médio. Dentro da MCNC, ela atua especificamente nas combinações do pixel de referência e de seus vizinhos. Segundo Fstedt *et al.* (2019), estes fatores, destacados na Tabela 3, proporcionaram o surgimento de outras medidas estatísticas a partir da MCNC.

Para o presente estudo foram implementadas ao modelo e realizadas simulações envolvendo todos os 19 fatores de Haralick considerando cada direção da matriz de coocorrência.

Estes fatores são importantes para o modelo proposto, pois identificam os padrões entre sinais de voz limpos e também com seus respectivos sinais ruidosos para diferentes contaminações.

Tabela 3: Os fatores de textura de Haralick calculados a partir de MCNC.

Nº Fatores	Fatores	Expressões
1	Autocorrelação	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i, j) P(i, j)$
2	Tom Central	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \{i + j - \mu_x - \mu_y\}^3 \times P(i, j)$
3	Destaque Central	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \{i + j - \mu_x - \mu_y\}^4 \times P(i, j)$
4	Contraste	$\sum_{i=1}^N \sum_{j=1}^N i - j ^2 P(i, j)$
5	Correlação	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{\{i \times j\} \times P(i, j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$
6	Diferença Entropia	$-\sum_{i=0}^{N-1} P_{x+y}(i) \log(P_{x+y}(i))$
7	Diferença Variância	$\sum_{j=0}^{N-1} (j - \mu_{x-y})^2 P_{x-y}(j)$
8	Dissimilaridade	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i - j \cdot P(i, j)$
9	Energia	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j)^2$
10	Entropia	$-\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j) \times \log(P(i, j))$
11	Homogeneidade	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1 + (i + j)^2} P(i, j)$
12	Medida Informação de Correlação 1	$\frac{HXY - HXY1}{\max\{HX, HY\}}$
13	Medida Informação de Correlação 2	$\sqrt{1 - \exp[-2(HXY2 - HXY)]}$
14	Momento de Diferença Inversa	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{ 1 + (i - j) } P(i, j)$
15	Probabilidade Máxima	$\max_{i,j} P(i, j)$
16	Soma Média	$\sum_{j=0}^{2N-2} j P_{x+y}(j)$
17	Soma Entropia	$-\sum_{i=0}^{2N-2} P_{x+y}(i) \log(P_{x+y}(i))$
18	Soma dos quadrados: Variância	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (1 - \mu)^2 P(i, j)$
19	Soma Variância	$\sum_{j=0}^{2N-2} (j - \mu_{x+y})^2 P_{x+y}(j)$

Fonte: Elaborada pelo próprio autor.

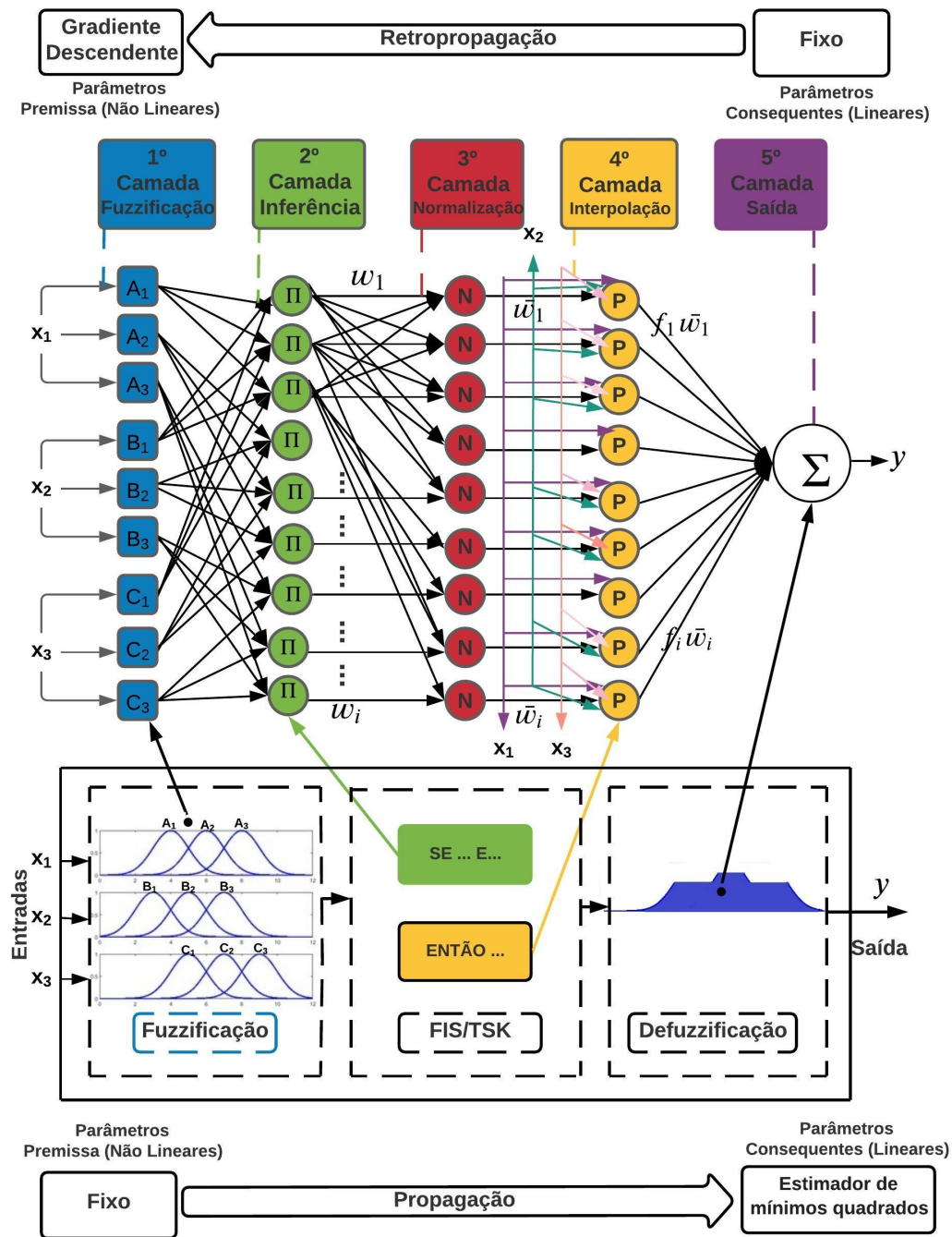
ARQUITETURA DA ANFIS

A ANFIS, proposta por Jang *et al.* (1997) pode ser pensada como uma classe de redes neurais adaptativas que são funcionalmente equivalentes a FIS. Essa característica resulta na vantagem da boa capacidade de aprendizado e apresenta desempenho satisfatório para diversas aplicações.

Na ANFIS, a simulação é realizada por fuzzificação dos valores de entrada via função de pertinência, que mapeiam as entradas no intervalo $[0,1]$ (JAFARI *et al.*, 2021). Os parâmetros das funções de pertinência para entrada e saída das variáveis são calibrados usando retropropagação e/ou método dos quadrados mínimos. Portanto, em contraste com a rede neural Perceptron Multicamadas - MLP - (do inglês: *Multi Layer Perceptron*), que ajusta os pesos, as regras de linguagem *fuzzy* (SE-ENTÃO) são determinadas para criar o modelo ANFIS, sendo implementado um FIS com o método TSK com capacidade de auto aprendizado para obter o mapeamento não linear ideal. A ANFIS consiste em cinco camadas e sua estrutura pode ser vista na Figura 21, onde a saída do i -ésimo neurônio na j -ésima camada é indicada por O_i^j .

Na primeira camada acontece a organização das variáveis de entrada para os conjuntos *fuzzy*, ou seja, é realizado o processo de fuzzificação por meio da aplicação de uma função de pertinência.

Figura 21: Arquitetura ANFIS.



Fonte: Elaborada pelo próprio autor.

Os neurônios da primeira camada são do tipo adaptativo, sendo assim, os parâmetros, que nesse caso dizem respeito às constantes das funções de pertinência, serão estimados via processo de treinamento. As saídas da primeira camada são os graus de pertinência associados

as entradas x_1 , x_2 e x_3 , cujas respectivas saídas são

$$O_i^1 = \mu_{A_k}(x_1), \quad i = 1, 2; \quad (45)$$

$$O_i^1 = \mu_{B_k}(x_2), \quad i = 3, 4;$$

$$O_i^1 = \mu_{C_k}(x_3), \quad i = 5, 6;$$

onde A_k , B_k e C_k são os valores linguísticos associados aos neurônios ($k = 1, 2$). A função de pertinência gaussiana foi usada no método proposto. Para as entradas do sistema, ela é definida como

$$\mu_{A_k}(x_1) = \exp \left[-\frac{(x_1 - m_{A_k})^2}{(2\sigma_{A_k}^2)^2} \right] \quad (46)$$

$$\mu_{B_k}(x_2) = \exp \left[-\frac{(x_2 - m_{B_k})^2}{(2\sigma_{B_k}^2)^2} \right]$$

$$\mu_{C_k}(x_3) = \exp \left[-\frac{(x_3 - m_{C_k})^2}{(2\sigma_{C_k}^2)^2} \right]$$

onde m_{A_k} , m_{B_k} e m_{C_k} são os centros das funções gaussianas, σ_{A_k} , σ_{B_k} e σ_{C_k} são os desvios padrão ou larguras das funções gaussianas ($k = 1, 2$). Estes parâmetros são definidos como os parâmetros da premissa (antecedente) do FIS.

A segunda camada representa os antecedentes das regras *fuzzy* que compõem o FIS e definida como composição que executa as operações “AND” e/ou “OR” (JAFARI *et al.*, 2021). A saída O_i^2 é o produto de todos os sinais de entrada. Cada neurônio que compõe esta camada corresponde a uma regra. A saída para a i -ésima regra é dada por

$$O_i^2 = \omega_i = \mu_{A_k}(x_1) \times \mu_{B_k}(x_2) \times \mu_{C_k}(x_3), \quad k = 1, 2, \dots, I, \quad (47)$$

onde ω_i é a força de disparo da i -ésima regra e I é o número total de regras.

A terceira camada é responsável pela normalização dos valores gerados na segunda camada. A saída dos neurônios que compõem essa camada indica o grau de ativação de uma regra em relação às outras i -ésimas. O valor de saída desses neurônios é dado pela equação

(48).

$$O_i^3 = \bar{\omega}_i = \frac{\omega_i}{\sum_{j=1}^I \omega_j}, \quad i = 1, 2, \dots, I. \quad (48)$$

Na quarta camada encontram-se as composições dos consequentes das regras TSK, cujo formato é dado por uma equação polinomial de primeira ordem. Na Figura 21 é destacada a equação (48). Os valores de saída dos neurônios dessa camada são calculados pela equação (49). Os neurônios dessa camada são do tipo adaptativo e, desta forma, seus parâmetros são determinados um do processo de treinamento, e os parâmetros a serem estimados correspondem aos valores de p_i, q_i, r_i e s_i

$$O_i^4 = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x_1 + q_i x_2 + r_i x_3 + s_i), \quad i = 1, 2, \dots, I \quad (49)$$

onde $\bar{\omega}_i$ é a saída da terceira camada. Os parâmetros f_i , ou seja, $\{p_i, q_i, r_i, s_i\}$, nesta camada são referidos como parâmetros consequentes do FIS.

A camada de saída contém um único neurônio que calcula a saída geral como uma adição de todos os sinais de entrada. Nesta camada é executada a defuzzificação calculada por meio da equação (50) que consiste da soma ponderada dos valores obtidos na quarta camada.

$$O_1^5 = y = \sum_{i=1}^I O_i^4 = \sum_{i=1}^I \bar{\omega}_i (p_i x_1 + q_i x_2 + r_i x_3 + s_i), \quad i = 1, 2, \dots, I. \quad (50)$$

Os parâmetros da premissa ($m_{A_k}, m_{B_k}, m_{C_k}, \sigma_{A_k}, \sigma_{B_k}, \sigma_{C_k}$) e os parâmetros o consequentes (p_i, q_i, r_i, s_i) associados à i -ésima regra constituem o conjunto dos parâmetros que serão configurados na fase de treinamento.

Funcionalmente, a saída y pode ser escrita como

$$y = F(x_1, x_2, x_3, \zeta) = \bar{\omega}_i (p_i x_1 + q_i x_2 + r_i x_3 + s_i), \quad (51)$$

onde ζ é um vetor da premissa e dos parâmetros consequentes.

O processo de treinamento da ANFIS pode ser realizado adotando-se um dentre os três procedimentos: *Backproagation*, *Backpropagation* + Método dos Mínimos Quadrados e *Back-*

propagation + Método dos Mínimos Quadrados uma única vez, propostos por Jang (1993).

Neste trabalho, para a fase de treinamento adotou-se o algoritmo de aprendizado híbrido (*Backpropagation* + Método dos Mínimos Quadrados) usando o método de retropropagação para ajustar os parâmetros dos neurônios da primeira camada, e o método dos mínimos quadrados para ajustar os da quarta camada. Na etapa de propagação adiante (do inglês: *forward*), as entradas x_1, x_2 e x_3 avançam até camada 4, e os parâmetros consequentes são identificados pelos mínimos quadrados, enquanto os parâmetros da premissa são fixos. Na etapa de retropropagação (do inglês: *backward*), as taxas de erro, são definidas como a derivada do quadrado do erro em relação a cada saída do neurônio, e os parâmetros da premissa são atualizados pelo gradiente descendente, enquanto os parâmetros consequentes permanecem fixos.

Por ser um método utilizado para obtenção das bases de regras *fuzzy* para o modelo de regressão deste trabalho, a próxima seção descreve como o algoritmo Backpropagation (etapa “backward”) e o Método dos mínimos quadrados (etapa “forward”) são combinados com a finalidade de gerar o processo de treinamento híbrido de um modelo ANFIS.

5.1 ETAPA *FORWARD*: APLICAÇÃO DO MÉTODO DOS MÍNIMOS QUADRADOS NO PROCESSO DE TREINAMENTO

Com base na arquitetura da ANFIS representada pela Figura 21, pode-se definir a saída f_{s_i} , do i -ésimo conjunto de treinamento, como uma combinação linear dos parâmetros dos

consequentes das regras $(p_i x_1 + q_i x_2 + r_i x_3 + s_i)$ como em (52) (CHEN; YING; PAN, 2010).

$$\begin{aligned}
 f_{s_i} &= \sum \bar{\omega}_i f_j \\
 f_{s_3} &= \bar{\omega}_1 f_1 + \bar{\omega}_2 f_2 + \bar{\omega}_3 f_3 \\
 f_{s_3} &= \bar{\omega}_1 (p_1 x_1 + q_1 x_2 + r_1 x_3 + s_1) \\
 &\quad + \bar{\omega}_2 (p_2 x_1 + q_2 x_2 + r_2 x_3 + s_2) \\
 &\quad + \bar{\omega}_3 (p_3 x_1 + q_3 x_2 + r_3 x_3 + s_3) \\
 f_{s_3} &= x_1 (\bar{\omega}_1 p_1 + \bar{\omega}_2 p_2 + \bar{\omega}_3 p_3) + s_1 (\bar{\omega}_1 + \bar{\omega}_2 + \bar{\omega}_3) \\
 &\quad + x_2 (\bar{\omega}_1 q_1 + \bar{\omega}_2 q_2 + \bar{\omega}_3 q_3) + s_2 (\bar{\omega}_1 + \bar{\omega}_2 + \bar{\omega}_3) \\
 &\quad + x_3 (\bar{\omega}_1 r_1 + \bar{\omega}_2 r_2 + \bar{\omega}_3 r_3) + s_3 (\bar{\omega}_1 + \bar{\omega}_2 + \bar{\omega}_3)
 \end{aligned} \tag{52}$$

$\langle (x_1, x_2, x_3), f_{s_3} \rangle$, constitui o conjunto de treinamento. Considerando as matrizes

$$f_s = \begin{bmatrix} f_{s_1} \\ f_{s_2} \\ f_{s_3} \end{bmatrix}, \quad \theta = \begin{bmatrix} p_1 \\ q_1 \\ r_1 \\ s_1 \\ p_2 \\ q_2 \\ r_2 \\ s_2 \\ p_3 \\ q_3 \\ r_3 \\ s_3 \end{bmatrix}$$

$$\text{e } B = [\bar{\omega}_1 x_1 \quad \bar{\omega}_1 x_2 \quad \bar{\omega}_1 x_3 \quad \bar{\omega}_1 \quad \bar{\omega}_2 x_1 \quad \bar{\omega}_2 x_2 \quad \bar{\omega}_2 x_3 \quad \bar{\omega}_2 \quad \bar{\omega}_3 x_1 \quad \bar{\omega}_3 x_2 \quad \bar{\omega}_3 x_3 \quad \bar{\omega}_3],$$

a equação (52) pode ser expressa por uma forma matricial

$$f_s = B\theta \quad (53)$$

onde θ é uma matriz desconhecida, cujos elementos vêm do conjunto de parâmetros consequentes das regras. O estimador ou método mínimos quadrados θ^* é dado por

$$\theta^* = (B^T B)^{-1} B^T f_s \quad (54)$$

Concluindo esta etapa “forward” de treinamento do modelo *fuzzy*, a rede neural é então utilizada para determinar os parâmetros dos antecedentes das regras de modelagem.

5.2 ETAPA *BACKWARD*: APLICAÇÃO DO *BACKPROPAGATION* NO PROCESSO DE TREINAMENTO

O algoritmo de aprendizado híbrido da ANFIS combina o método do gradiente com o método dos mínimos quadrados para atualizar os parâmetros em uma rede adaptativa. Cada época desse procedimento de aprendizado híbrido é composta de um “forward” e um “backward”. No “forward”, depois que cada vetor de entrada é dado, calcula-se a saída do neurônio correspondente até que as matrizes B e f_s sejam obtidas, de acordo com a equação (53). Os parâmetros consequentes são identificados usando a equação (54) e, desta forma, é calculado o erro médio quadrático dado por

$$E_i = \frac{1}{n} \sum_{i=1}^n (T_i - f_{s_i})^2 \quad (55)$$

onde E_i é a medida de erro para a entrada do conjunto de dados de treinamento, T_i é o resultado estimado e f_{s_i} é o valor esperado. Nesta etapa do treinamento, o erro E é propagado para as demais camadas da ANFIS e, com base nesse valor, os parâmetros dos neurônios adaptativos da primeira camada vão sendo ajustados (CHEN; YING; PAN, 2010).

No “backward”, calcula-se a taxa de erro $\frac{\partial E_i}{\partial \alpha}$ para a entrada do conjunto de dados de

treinamento e para cada saída de neurônio O dado por, $\Delta\alpha$, sendo α um parâmetro do conjunto da premissa de uma função de pertinência da variável *fuzzy* de entrada obtido por $\frac{\partial E}{\partial \alpha}$. Pela regra da cadeia, a derivada da medida de erro geral E em relação a α é dada por

$$\frac{\partial E}{\partial \alpha} = \frac{\partial E}{\partial f} \frac{\partial f}{\partial f_i} \frac{\partial f_i}{\partial \omega_i} \frac{\partial \omega_i}{\partial \mu_{i,j}} \frac{\partial \mu_{i,j}}{\partial \alpha} \quad (56)$$

Então, a fórmula atualizada para o parâmetro α da premissa do método do gradiente é dada por

$$\Delta\alpha = -\eta \frac{\partial E}{\partial \alpha} \quad (57)$$

em que η é uma taxa de aprendizado, expressa como

$$\eta = \frac{s}{\sqrt{\sum_{\alpha} \left(\frac{\partial E}{\partial \alpha} \right)^2}} \quad (58)$$

sendo s o comprimento de cada transição de gradiente no espaço de parâmetros. Geralmente, o valor de s é alterado para variar a velocidade de convergência (ALCALÁ *et al.*, 1999).

Da equação (55) pode-se resolver $\frac{\partial E}{\partial f}$ conforme a equação (59)

$$\frac{\partial E}{\partial f} = (f_i - T_i) \quad (59)$$

Da equação (50), $y = \sum_{i=1}^n \bar{\omega} f_i$ resolve-se o termo $\frac{\partial f}{\partial f_i}$, resultando em

$$\frac{\partial f}{\partial f_i} = 1 \quad (60)$$

Como $f_i = (p_i x_1 + q_i x_2 + r_i x_3 + s_i)$, o termo $\frac{\partial f}{\partial \omega_i}$ resulta em (CHEN; YING; PAN, 2010)

$$\frac{\partial f}{\partial \omega_i} = \frac{(p_i x_1 + q_i x_2 + r_i x_3 + s_i) - f_i}{\sum_{i=1}^n \omega_i} \quad (61)$$

Da segunda camada da rede, considerando o operador produto $Prod(A, B, C)$ aplicado, obtém-se $\omega_i = \prod_{k=1}^I \mu_{i,j}$. Dessa forma, resolvendo-se o termo $\frac{\partial \omega_i}{\partial \mu_{i,j}}$ obtém-se

$$\frac{\partial \omega_i}{\partial \mu_{i,j}} = \frac{\omega_i}{\mu_{i,j}}. \quad (62)$$

Assim, chega-se na última derivada parcial da função de pertinência propriamente dita, representada pelo primeiro termo da equação (56). A função de pertinência gaussiana foi adotada neste trabalho por ser continuamente diferenciável, seus parâmetros de centro m_{A_k} , m_{B_k} e m_{C_k} (posição central da curva no eixo das abscissas) e σ_{A_k} , σ_{B_k} e σ_{C_k} (largura da gaussiana, ou suporte) podem ser aperfeiçoados pelo algoritmo do gradiente. Considerando a equação (46), pode-se estimar o valor dos parâmetros m_{A_k} e σ_{A_k} , assim como os parâmetros m_{B_k} , m_{C_k} , σ_{B_k} e σ_{C_k} , resolvendo simultaneamente as equações (63) e (64)

$$\frac{\partial \mu_{i,j}}{\partial m_{A_k}} = e^{\left(\frac{-(x_1 - m_{A_k})^2}{2\sigma_{A_k}^2}\right)} \frac{(x_1 - m_{A_k})}{\sigma_{A_k}^2} \quad (63)$$

$$\frac{\partial \mu_{i,j}}{\partial \sigma_{A_k}} = e^{\left(\frac{-(x_1 - m_{A_k})^2}{2\sigma_{A_k}^2}\right)} \frac{(x_1 - m_{A_k})^2}{\sigma_{A_k}^3}. \quad (64)$$

O procedimento descrito ocorre para todos os pares de dados disponíveis para treinamento. Uma vez que todos os pares de dados de treinamento são apresentados, uma iteração ou época é caracterizada. O processo de treinamento segue até que o número de épocas determinado seja atingido ou até que um erro objetivo, dado pela equação (55), calculado no final de cada iteração seja alcançado.

Geralmente, cada modelo ANFIS é estruturado para construir seções de treinamento de dados. O número e o tipo de funções de pertinência utilizadas, bem como a divisão dos dados de entrada-saída e pacotes de regras, são obtidos na fase de construção, ou seja, na primeira camada. Portanto, para realizar esta tarefa, os métodos de fuzzificação são usados como poderosas ferramentas para compreender e classificar os fatores de entrada indicando a forma como as regras *fuzzy* estão relacionadas no espaço.

Desta forma, neste trabalho foram escolhidos três métodos de fuzzificação: particionamento *Grid*, agrupamentos *Fuzzy C-Means* (FCM) e *Subtractive* (SC). Benmouiza e Cheknane (2019) enfatizam que estes três métodos de fuzzificação são os mais difundidos na literatura. Entretanto, ajustes de parâmetros e determinação de funções de pertinência corroboram para um bom desempenho do modelo ANFIS em estudo (YEOM ; KWAK, 2018). Nas próximas seções serão apresentados os três métodos de fuzzificação utilizados na fase de implementação, que teve como objetivo apontar aquele que proporciona melhor desempenho frente as métricas escolhidas para avaliar o modelo de regressão de qualidade de sinais de voz.

5.3 PARTICIONAMENTO *GRID*

Neste método, o espaço de dados de entrada é dividido em retângulos que são subespaços usando uma partição paralela aos eixos das abscissas e ordenadas, cada entrada é particionada em funções de pertinência com formato idêntico (BENMOUIZA; CHEKNANE, 2019). Esta abordagem propõe partições independentes de cada variável subsequente (JANG; SUN; MIZUTANI, 1997). Na fuzzificação, usando a abordagem de particionamento em grade (do inglês - *Grid*), o espaço de busca é dividido para formar a estrutura inicial do modelo *fuzzy* (SITOMPUL; NABABAN; ALIM, 2017). A grade formada pela combinação da relação de cada partição irá potencialmente formar as regras, conforme ilustrado na Figura 22.

A regra, que é formada em partições de grade bidimensionais, é definida como:

$$R_{i,j}^K : \text{SE } x_1 \text{ é } A_i^K \text{ ENTÃO } C; CF_{ij}^K.$$

Onde:

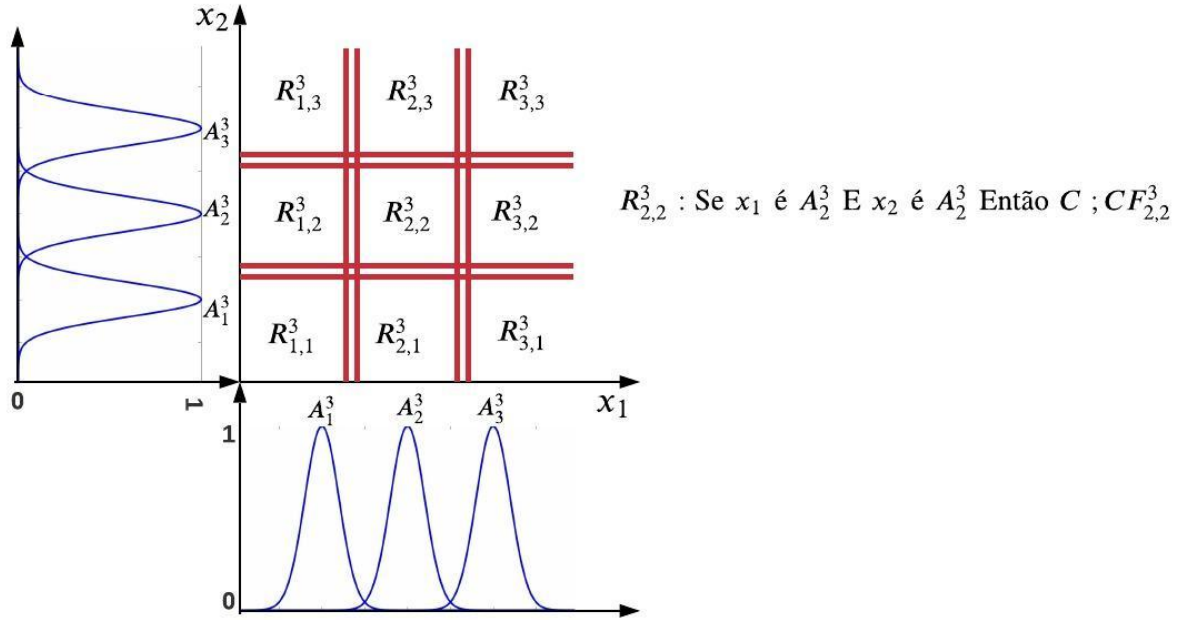
K : é a dimensão da partição;

A : é a premissa;

C : é o consequente;

CF : é o peso da regra;

i, j : são inteiros $1, 2, \dots, K$.

Figura 22: Partição *Grid*.

Fonte: Elaborada pelo próprio autor.

O número das regras *fuzzy* é igual a R^K , sendo K a dimensão de entrada e R é o número de subconjuntos *fuzzy* para cada variável de entrada.

A grade é construída sem ter qualquer significado físico ou repartição da densidade de dados, e cada parte sua é usada para gerar regras *fuzzy* com base no treinamento de entrada e saída dos dados do sistema, o que permite processos de aprendizagem rápidos e de baixo custo computacional. No entanto, o desempenho deste método depende fortemente do tamanho das entradas e da grade. Por exemplo, quando existem 3 variáveis de entrada e cada variável de entrada possui 3 funções de pertinência, o espaço de dados é dividido em $3^3 = 27$ áreas específicas. Em resumo, uma regra é criada para cada área e o total de regras é igual a 27. Portanto, o método de particionamento *Grid* é usado principalmente quando o número de variáveis de entrada é pequeno. Geralmente, uma grade com menos regras leva a alto desempenho. Uma grade adaptável pode ser usada para otimizar o tamanho e a localização das regiões *fuzzy* (BENMOUIZA; CHEKNANE, 2019).

5.4 MÉTODOS DE AGRUPAMENTO

Para compensar a desvantagem de que o número de regras aumente exponencialmente, na medida que as variáveis de entrada aumentam, os métodos de particionamento de dispersão e/ou agrupamentos em centro (do inglês - *cluster*) foram difundidos na literatura (YEOM; KWAK, 2018). Nos métodos de agrupamento, os espaços de entrada são divididos em n centros por meio de um algoritmo de centralização (do inglês - *clustering*). Uma área específica é dividida em n centros, e cada um deles representa uma regra *fuzzy*. Os métodos de agrupamento são caracterizados em vários tipos, de acordo com seu algoritmo de execução.

Os agrupamentos podem ser classificados como exclusivos ou não exclusivos. Nos agrupamentos exclusivos (do inglês - *hard*), cada elemento pertence exclusivamente a um único grupo. Já no agrupamento não exclusivo, considerado *fuzzy*, um elemento pode pertencer a mais de um grupo (ROSS, 2017).

Para definição de um agrupamento, uma medida de similaridade se torna fundamental, pois define a que grupo o valor do atributo ou fator pertence. Segundo Ross (2017), a medida mais utilizada na literatura para atributos contínuos é a distância euclidiana. Os métodos de agrupamentos estudados neste trabalho são descritos a seguir.

5.4.1 AGRUPAMENTO *FUZZY C-MEANS* - FCM

O método de agrupamento *Fuzzy C-means* (FCM) foi introduzido por BEZDEK (1973) e BEZDEK (1981) e é conhecido como uma melhoria e modificação do agrupamento HCM (do inglês - *Hard C-Means*) e/ou *K-means*, por ter um tratamento *fuzzy* na definição de pertinência de um elemento em determinados grupos.

Bezdek (1981) sugeriu o uso de uma abordagem de função objetivo para agrupar os dados em grupos hipersféricos. A função objetivo é desenvolvida de modo a fazer duas tarefas simultaneamente: minimizar a distância euclidiana entre cada ponto de dados de um grupo com seu centro e maximizar a distância euclidiana entre os centros dos grupos.

O HCM é utilizado para classificar dados em um conjunto crisp. Na teoria clássica, os conjuntos são denominados crisp - $\{0,1\}$ - e se um dado elemento do universo em discurso

(domínio) pertence atribui-se 1 (um), ou 0 (zero) para não pertinência ao referido conjunto. Com isso, pode-se dizer que cada ponto de dados será atribuído a um, e apenas um, grupo de dados. Nesse sentido, esses grupos também são chamados de c partições, ou seja, partições dos dados. O conjunto $\{A_i, i = 1, 2, \dots, c\}$ é uma família com c partições *Hard* de X , onde $X = \{x_1, x_2, x_3, \dots, x_n\}$ é um espaço de conjunto finito que compreende o universo de amostras de dados, sendo c o número de classes, ou partições, ou grupos, em que pretende classificar os dados para $2 \leq c < n$.

Para simplificar a notação, a atribuição de pertinência do j -ésimo ponto de dados no i -ésimo grupo é definida como $\mathcal{X}_{i,j}\mathcal{X}_{A_j}(x_j)$. Assim, é definida uma matriz U compreendendo os elementos $\mathcal{X}_{i,j}$ ($i = 1, 2, \dots, c; j = 1, 2, \dots, n$). U é uma matriz com c linhas e n colunas. Ficando definido o espaço *Hard* de partição c para $\mathcal{X}$ como o seguinte conjunto de matriz:

$$M_c = \left\{ U \mid \mathcal{X}_{i,j} \in \{0, 1\}, \sum_{i=1}^c \mathcal{X}_{i,k} = 1, 0 < \sum_{k=1}^m \mathcal{X}_{i,k} < n \right\} \quad (65)$$

Qualquer matriz $U \in M_c$ é uma partição c *Hard* onde a cardinalidade de M_c é representada por

$$\eta_{M_c} = \left(\frac{1}{c!} \right) \left[\sum_{i=1}^c \binom{c}{i} (-1)^{c-i} \cdot i^n \right] \quad (66)$$

onde a expressão $\binom{c}{i}$ é o coeficiente binomial de c elementos tomados i de cada vez.

A medida de cardinalidade fornecida pela equação (66) define o número de c partições exclusivas para n pontos de dados.

Para selecionar um possível candidato da partição para n amostras de um conjunto de dados dentro do espaço M_c , uma função objetivo ou critério de classificação é adotada para agrupar os dados.

O algoritmo HCM é conhecido como uma soma dentro da classe de abordagem de erros quadráticos usando a norma euclidiana para caracterizar distâncias. Este algoritmo é denotado $J = (U, v)$, onde U é a matriz de partição, e v é um vetor de centro de grupo, e a função

objetivo é dada pela equação (67)

$$J(U, v) = \sum_{k=1}^n \sum_{i=1}^c \mathcal{X}_{i,j}(d_{i,k}), \quad (67)$$

sendo d_{i_k} uma medida de distância euclidiana (em espaço de característica m dimensional, R^m) para a k -ésima amostra de dados representada por x_k e i é o centro do grupo V_i , representada pela seguinte forma

$$d_{ik} = d(x_k - V_i) = \|x_k - V_i\| = \left[\sum_{j=1}^m (x_{kj} - V_{ij})^2 \right]^{\frac{1}{2}} \quad (68)$$

Como cada amostra de dados requerem coordenadas para descrever sua localização no espaço R^m , cada centro do grupo também requer coordenadas m para descrever sua localização neste mesmo espaço. Portanto, o i -ésimo centro do grupo é um vetor de comprimento m ,

$$V_i = [v_{i1}, v_{i2}, \dots, v_{im}], \quad (69)$$

tendo a j -ésima coordenada é calculada pela equação (70).

$$v_{ij} = \frac{\sum_{k=1}^n \mathcal{X}_{i,k} \cdot x_{kj}}{\sum_{k=1}^n \mathcal{X}_{ik}}. \quad (70)$$

Logo, procura-se a partição ótima, U^* , para ser a partição que produz o valor mínimo para a função, J . Isto é,

$$J(U^*, V^*) = \min_{U \in \mu_c} J(U, V). \quad (71)$$

Encontrar a matriz de partição ideal, U^* , é extremamente difícil para problemas práticos porque $\eta_{M_c} \rightarrow \infty$ mesmo para problemas de tamanho modesto. Para resolver este problema foram criados os algoritmos conhecidos como otimização iterativa. Basicamente, estes métodos iterativos começam com uma estimativa inicial na matriz U , estipulando uma quantidade de grupos e um número de tolerância de iterações. A partir desses grupos ou centros,

recalculam-se os valores de pertinência que cada ponto de dados tem no grupo. Os valores são comparados com os assumidos, e assim, o processo continua até que as mudanças de ciclo para ciclo estejam dentro do nível de tolerância estipulado.

O procedimento passo a passo de um método de otimização iterativa é descrito por Bezdek (1981) da seguinte forma:

1º Passo - Fixe c ($2 \leq c < n$) e inicialize a matriz U :

$$U^{(0)} \in M_c$$

Fazendo $r = 0, 1, 2, \dots$

2º Passo - Calcule o c vetores centrais :

$$\{V_i^{(r)} \text{ com } U^{(0)}\}$$

3º Passo - Atualizar $U^{(0)}$ e calcular as funções características atualizadas (para todos os i, k):

$$\mathcal{X}_{i,k}^{(r+1)} = \begin{cases} 1 & \text{se } d_{ik}^{(r)} = \min\{d_{jk}^{(r)}\} \text{ para todo } j \in c; \\ 0, & \text{caso contrário.} \end{cases}$$

4º Passo - Se

$$\|U^{(r+1)} - U^{(r)}\| \leq \epsilon$$

definido como nível de tolerância, for satisfeito, pare; caso contrário, faça $r = r + 1$ e volte ao **2º Passo**.

No **4º Passo**, a notação $\|\cdot\|$ é qualquer norma de matriz, como a norma euclidiana.

O método HCM destacado é considerado suporte para compreensão do método FCM. Assim, o FCM se define como uma família de conjuntos $\{\tilde{A}_i = 1, 2, \dots, c\}$ de partições *c fuzzy* em um universo de pontos de dados, X . Nesta abordagem, cada ponto de dados pertence parcialmente a todos os grupos, no entanto, com diferentes graus de adesão que podem variar entre 0 a 1. Portanto, um único ponto pode ter participação parcial em mais de um grupo. Descreve-se o valor de pertinência que o k -ésimo ponto de dados tem no i -ésimo grupo com a seguinte notação:

$$\mu_{ik} = \mu_{A_i}(x_k) \in [0, 1],$$

como restrição, conforme a definição de conjunto crisp, de que a soma de todos os valores de pertinência para um único ponto de dados em todos os grupos deve ser unitária:

$$\sum_{i=1}^c \mu_{ik} = 1, \text{ para todo } k = 1, 2, \dots, n. \quad (72)$$

Assim como na definição de conjuntos crisp, os grupos não podem ser vazios e nem conter todos os pontos de dados. Essa restrição se manifesta na seguinte expressão:

$$0 < \sum_{i=1}^n \mu_{ik} < n. \quad (73)$$

Como cada ponto do conjunto de dados pode ter pertinência parcial em mais de um grupo, desta forma, seguem os seguintes casos *fuzzy*

$$\sum_{k=1}^c \mu_{A_k} = 1, \text{ para todo } k; \quad (74)$$

$$0 < \sum_{i=1}^n \mu_{A_i}(x_k) < n, \text{ para todos } i. \quad (75)$$

Assim, pode-se definir uma família de matrizes de partição *fuzzy*, M_{fc} , para a classificação envolvendo c grupos em pontos de dados:

$$M_{fc} = \left\{ \underline{U} \mid \mu_{i,j} \in [0, 1]; \sum_{i=1}^c \mu_{i,k} = 1; 0 < \sum_{k=1}^m \mu_{i,k} < n \right\} \quad (76)$$

onde $i = 1, 2, \dots, c$ e $k = 1, 2, \dots, n$.

Qualquer $\underline{U} \in M_{fc}$ é uma partição *c fuzzy*, resultado da sobreposição dos grupos e o número infinito de valores de pertinências possíveis, ou seja, a cardinalidade $\eta_{M_{fc}} \rightarrow \infty$.

O algoritmo FCM também foi criado para resolver o problema do número de partições dadas pela cardinalidade do conjunto de dados. Desta forma, para descrever um método para determinar a matriz c da partição *fuzzy* $\underline{U}$ para agrupar uma coleção de n elementos dentro do conjunto de dados, define-se uma função objetivo J_m dada por:

$$J_m(\underline{U}, V) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^{m'} (d_{ik})^2, \quad (77)$$

onde

$$d_{ik} = d(x_k - V_i) = \left[\sum_{j=1}^m (x_{kj} - V_{ij})^2 \right]^{\frac{1}{2}} \quad (78)$$

sendo μ_{ik} a pertinência do k -ésimo ponto de dados no i -ésimo grupo.

Da mesma forma que acontece na abordagem *Hard*, a função J_m pode ter um grande número de valores, associados ao melhor agrupamento, devido ao grande número de valores possíveis, com cardinalidade infinita dos conjuntos *fuzzy*. No entanto, busca-se encontrar a melhor solução possível, ou ótima, sem recorrer a uma pesquisa exaustiva e com um elevado custo computacional.

A medida de distância d_{ik} é novamente uma distância euclidiana, como vista na equação (78), entre o i -ésimo centro do grupo e o k -ésimo conjunto de dados sendo cada ponto no espaço m . Um novo parâmetro é introduzido na equação (77), chamado de parâmetro de ponderação, m' (BEZDEK, 1981), definido no intervalo $[1, \infty)$, para controlar a quantidade de imprecisão no processo de agrupamento.

Além disso, V_i é o i -ésimo centro do grupo descrito por m coordenadas e pode ser representado pelo de vetor, $V_i = [v_{i1}, v_{i2}, \dots, v_{im}]$.

Cada uma das coordenadas do grupo é obtida da seguinte maneira

$$v_{ij} = \frac{\sum_{k=1}^n \mu_{i,k}^{m'} \cdot x_{ki}}{\sum_{k=1}^n \mu_{i,k}^{m'}} \quad (79)$$

onde j é uma variável no espaço dos fatores, ou seja, $j = 1, 2, \dots, m$.

Como no caso *Hard*, a partição c *fuzzy* ideal será a menor das partições descritas pela equação (80).

$$J_m^*(\underline{U}^*, V^*) = \min_{M_{fc}} J(\underline{U}, V). \quad (80)$$

Como acontece com muitos processos de otimização, a solução para a equação (80) não pode ter a garantia de ser um ótimo global, ou seja, o melhor entre os candidatos. O que se busca é a melhor solução disponível dentro de um nível pré-especificado de precisão. Os passos da adaptação do método de otimização iterativa para o agrupamento *fuzzy* são:

1º Passo: Fixe c ($2 \leq c < n$) e selecione um valor para o parâmetro m' . Inicialize a matriz de partição, $\underline{U}^{(0)}$. Cada etapa neste algoritmo será rotulada por r , onde $r = 0, 1, 2, \dots$

2º Passo: Calcule os centros $c \{V_i^{(r)}\}$ para cada etapa.

3º Passo: Atualize a matriz de partição para a r -ésima etapa, $\underline{U}$, da seguinte maneira:

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^c \left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}} \right)^{\frac{2}{(m'-1)}} \right]^{-1} \quad \text{para } i \in I_k \neq \emptyset \quad (81)$$

$$\text{ou } \mu_{ik}^{(r+1)} = 0, \text{ para todos grupos } i, \text{ com } i \in \tilde{I}_k; \quad (82)$$

$$\text{onde } I_k = \{i | 2 \leq c < n; d_{ik}^{(r)} = 0\}, \quad (83)$$

$$\text{e } \tilde{I}_k = \{1, 2, \dots, c\} - I_k, \quad (84)$$

$$\text{e } \sum_{i \in I_k} \mu_{ik}^{(r+1)} = 1. \quad (85)$$

4º Passo: Se $\|\underline{U}^{(r+1)} - \underline{U}^{(r)}\| \leq \epsilon_L$, pare. Caso contrário, faça $r = r + 1$ e volte ao **2º Passo**.

No **4º Passo**, foi comparada uma norma da matriz de duas partições *fuzzy* sucessivas para um nível de precisão, ϵ_L , para determinar se a solução é ótima o suficiente. No **3º Passo**, há uma considerável quantidade de lógica envolvida nas equações (81) (85). A equação (81) é

direta o suficiente, exceto quando $d_{jk} = 0$. Portanto, os parâmetros I_k e $\tilde{I}_k$ compreendem um sistema de contabilidade para lidar com situações em que algumas das medidas de distância, d_{ij} , são nulas ou extremamente pequenas. Se um valor nulo for detectado, a equação (82) define a pertinência para essa partição e o valor passa a ser zero. As equações (83) e (84) descrevem os parâmetros da contabilidade I_k e $\tilde{I}_k$, respectivamente, para cada um dos grupos. A equação (85) é um somatório indicando que todos os elementos de valores diferentes de zero de cada coluna da partição de agrupamento *fuzzy*, $\tilde{U}_k$, são iguais a 1 (ROSS, 2017).

O algoritmo fornecido na equação (77) é uma função de mínimos quadrados, onde o parâmetro n é o número de elementos do conjunto de dados e c é o número de grupos obtidos. O quadrado da distância, d_{ik}^2 , é então ponderado por uma medida, $(u_{ik})^{m'}$, de pertinência de x_k no i -ésimo agrupamento. O valor de J_m é então uma medida da soma de todos os erros quadráticos ponderados, e é minimizado em relação a duas funções de restrição. Primeiro, J_m é minimizado com relação aos erros quadráticos dentro de cada grupo, ou seja, para cada valor específico de c . Simultaneamente, a distância entre os centros do grupo é maximizada, ou seja, $\max |v_i - v_j|$, $i \neq j$.

Como destacado, o intervalo para o expoente de pertinência é $m' \in [1, \infty)$. Para o caso $m' = 1$, a norma euclidiana do algoritmo FCM se aproxima de um HCM, ou seja, apenas zeros e uns saem do agrupamento. Por outro lado, quando $m' \rightarrow \infty$, $J_m \rightarrow 0$.

Em geral, quanto maior for m' , mais *fuzzy* são as atribuições de pertinência do agrupamento e, para $m \rightarrow 1$, os valores se tornam *Hard*, ou seja, 0 ou 1. O expoente m' , portanto, controla a extensão de pertinência do compartilhamento entre grupos *fuzzy*. Se todos os outros parâmetros do algorítmico forem fixos, aumentando m' resultará na diminuição de J_m . Ross (2017) enfatiza que nenhuma escolha teórica ótima de m' emergiu na literatura. No entanto, a maior parte da literatura parece relatar valores na faixa de 1,25 a 2,0. A convergência do algoritmo tende a ser mais lenta conforme m' aumenta.

Tabela 4: Centros dos grupos mostrados nas figuras 23a, 23c e 23e.

Grupo 1		Grupo 2		Grupo 3	
Figura 23a		Figura 23c		Figura 23e	
Fator 1	PESQ	Fator 11	PESQ	Fator 13	PESQ
38,4989	2,4406	0,8744	2,4484	0,9673	4,4994
48,4848	1,7976	0,8877	4,4994	0,9493	2,4482
28,3552	4,4213	0,8948	1,7487	0,9061	1,7484

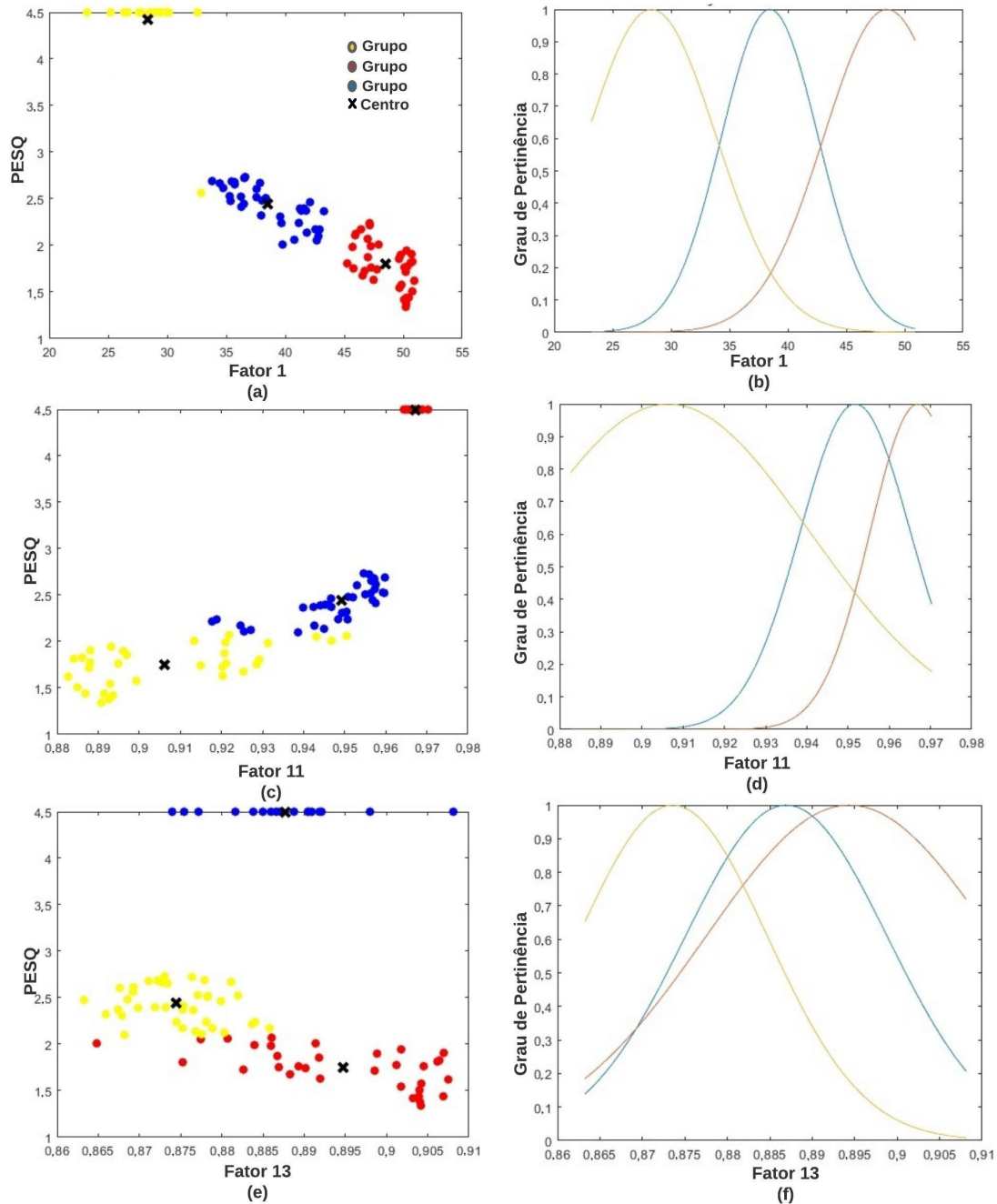
Fonte: Elaborada pelo próprio autor.

Como visto, o algoritmo FCM pode ser extremamente preciso e robusto. Para esta pesquisa sua abordagem foi utilizada como método de fuzzificação dos fatores de entrada, para determinar as funções de pertinência gaussianas. A seguir será mostrado o processo de determinação das funções de pertinência por meio do algoritmo FCM que segue a abordagem de (MATULIN; MRVELJ, 2018).

Utilizando os fatores 1, 11 e 13 de Haralick da Tabela 3 como entrada e a suas respectivas avaliações PESQ como saída, aplica-se o algoritmo FCM para determinar os centros dos grupos. Para este caso foram considerados três centros, ou seja, para cada fator de entrada são definidas três variáveis linguísticas, utilizando o parâmetro de fuzzificação $m' = 2$ e $\epsilon = 10^{-5}$ como critério de parada para o algoritmo. Os resultados são apresentados nas Figuras 23a, 23c e 23e. Nelas $\times$ e os círculos, representam os centros e os grupos respectivamente. As coordenadas dos grupos são destacadas por cores como pode ser visto no Apêndice A, enquanto as coordenadas do centro podem ser encontradas na Tabela 4.

Para determinar as funções de pertinência tipo gaussianas são necessários a média e o desvio padrão das amostras para cada variável linguística dos fatores de entrada. No entanto, valores médios das funções gaussianas representados nas Figuras 23b, 23d e 23f correspondem com as coordenadas dos centros dos grupos apresentados na Tabela 4 e nas Figuras 23a, 23c e 23e, respectivamente. Isso significa que os pontos de dados têm um maior grau de pertinência (μ_{ij}) para um grupo específico j se eles estiverem localizados mais perto do centro desse grupo. Portanto, a mudança na amostra da média da população, impacta no número de grupos e na localização de seus respectivos centros.

Figura 23: Resultados do algoritmo FCM para 90 sinais para os fatores 1, 11 e 13 de Haralick com suas respectivas avaliações PESQ sendo: a), c) e e) três grupos com seus respectivos centros e b), d) e f) funções de pertinência com três variáveis linguísticas.



Fonte: Elaborada pelo próprio autor.

Pode-se observar como as funções de pertinência *fuzzy* estão se sobrepondo devido à dispersão dos pontos de dados em torno dos centros dos grupos. A sobreposição das funções de pertinência são uma característica fundamental do método FCM, uma vez que permite

agrupar um ponto do conjunto de dados em mais de um grupo (MATULIN; MRVELJ, 2018).

5.4.2 AGRUPAMENTO *SUBTRACTIVE* - SB

O método de agrupamento SB (do inglês - *Subtractive*), é aplicado quando não há uma ideia clara sobre o número de centros de distribuição de dados. O SB é uma extensão da classificação do método da montanha proposto por (YAGER; FILEV, 1994; CHIU, 1994).

O SB é um algoritmo não supervisionado, ou seja, não é necessário informar quais são os centros c . Cada elemento do conjunto de dados é considerado um candidato central do grupo. O algoritmo é constituído por um processo iterativo, pressupondo que cada ponto centro do grupo em potencial, aferido pela densidade de pontos ao seu redor, de acordo com sua localização entre os outros pontos de dados (BENMOUIZA; CHEKNANE, 2019). O processo inicia com a informação do raio do grupo, definido como intervalo de influência do elemento central quando o espaço de entrada é considerado um hipercubo unitário e tem um valor entre 0 e 1. Se o raio $r \rightarrow 0$, proporcionará um número pequeno de elementos em cada grupo, aumentando sua quantidade e, conseqüentemente, o número de regras *fuzzy* do sistema. Por outro lado, se $r \rightarrow 1$, resultará em poucos grupos contendo muitos elementos, reduzindo a quantidade de regras *fuzzy* (YEOM; KWAK, 2018).

Note pelo método SB que os grupos são gerados de acordo com raio do grupo, como ilustrado na Figura 24. Os pontos vermelhos representam os dados e os pontos pretos são os centros dos grupos. A seguir, os passos do método SB são descritos.

1º Passo: A identidade de cada elemento no espaço de dados é calculada usando a função de densidade dada por

$$P_i = \sum_{j=1}^m \exp \left(\frac{-||x_i - x_j||}{\left(\frac{r_a}{2}\right)^2} \right) \quad (86)$$

onde, $r_a \in (0, 1]$ é uma constante positiva que representa o raio do grupo previamente definido.

2º Passo: Encontrar x_{c1} , que são os dados com o maior valor de densidade de P_i , tornando-se o elemento central do primeiro grupo.

3º Passo: O centro do grupo encontrado no **2º Passo** é removido pela equação (87),

$$P_i = P_i - P_{c1} \times \sum_{j=1}^m \exp \left(\frac{-||x_i - x_j||}{(\frac{r_b}{2})^2} \right) \quad (87)$$

onde, $r_b = k \cdot r_a$, sendo (k um número positivo, geralmente é adotado $k = 1, 5$, tornando r_b uma constante com valor positivo que representa o raio da função de eliminação) (YAGER e FILEV, 1994). P_{c1} representa o valor de medição de densidade de x_{c1} calculado na **Etapa 2**. Como consequência, todos os pontos próximos do primeiro centro do grupo x_{c1} terão medidas de baixa densidade e, portanto, não serão considerados. O próximo centro de grupo x_{c2} é selecionado a medida que a densidade para cada ponto de dados é recalculada (BENMOUIZA; CHEKNANE, 2019).

4º Passo: Repita o **2º Passo** e o **3º Passo** até que o valor de medição de densidade mais alto seja menor que o valor definido. E, desta forma, são repetidos os processos até que um número suficiente de centros de grupos seja gerado (BENMOUIZA; CHEKNANE, 2019).

De forma análoga à abordagem FCM, o algoritmo SB descreve as funções de pertinência gaussianas por meio das coordenadas centrais de cada grupo. Esta técnica é utilizada quando não há uma ideia clara de quantos grupos devem haver para um determinado conjunto de dados. No entanto, para estimar o número de grupos e os centros no conjunto de dados de entrada é preciso especificar um parâmetro denominado intervalo de influência.

Tabela 5: Centros dos grupos mostrado nas figuras 24a, 24c e 24e.

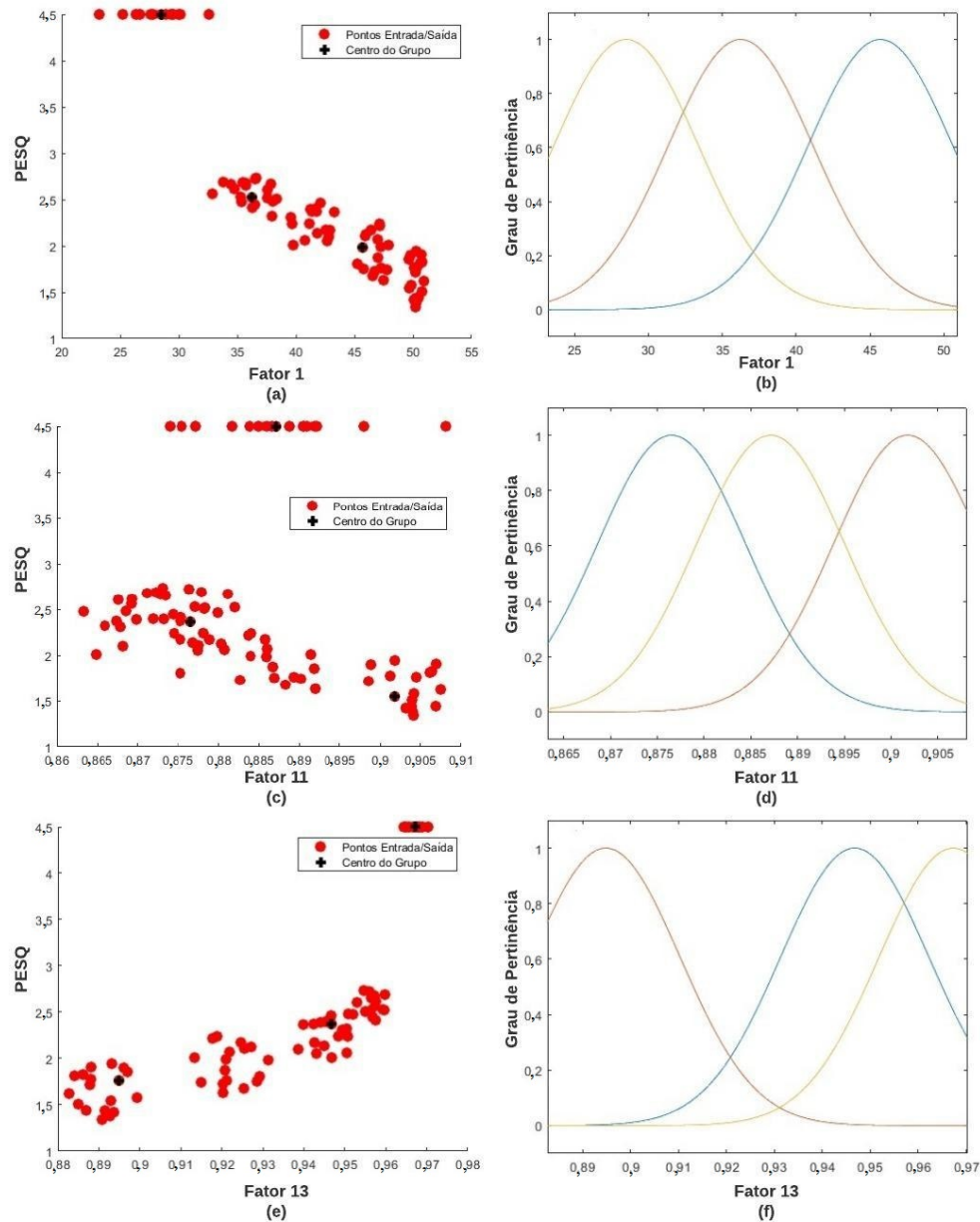
Grupo 1		Grupo 2		Grupo 3	
Figuras 24a		Figura 24c		Figura 24e	
Fator 1	PESQ	Fator 11	PESQ	Fator 13	PESQ
45,7017	1,9833	0,8765	2,3662	0,9468	2,3725
36,2445	2,5232	0,9018	1,5444	0,8949	1,7634
28,4976	4,5000	0,8871	4,5000	0,9673	4,5000

Fonte: Elaborada pelo próprio autor.

Para efeito de comparação de desempenho entre os métodos foi aplicado ao modelo SB um conjunto de dados de entrada contendo os fatores 1, 11 e 13 de Haralick com suas respectivas avaliações PESQ e um raio de aglomeração $r = 0,5$. Esta configuração gerou três centros para cada fator de entrada, como destacados nas Figuras 24a, 24c e 24e. As coordenadas

obtidas pelo algoritmo são apresentadas na Tabela 5. Fixando a coordenada de cada fator como média e calculando o desvio padrão entre os elementos que compõem o grupo são definidas as funções gaussianas destacadas nas Figuras 24b, 24d e 24f.

Figura 24: Resultados do algoritmo SB para 90 sinais para os fatores 1, 11 e 13 de Haralick com suas respectivas avaliações PESQ sendo: a), c) e e) três grupos com seus respectivos centros e b), d) e f) funções de pertinência com três variáveis linguísticas.



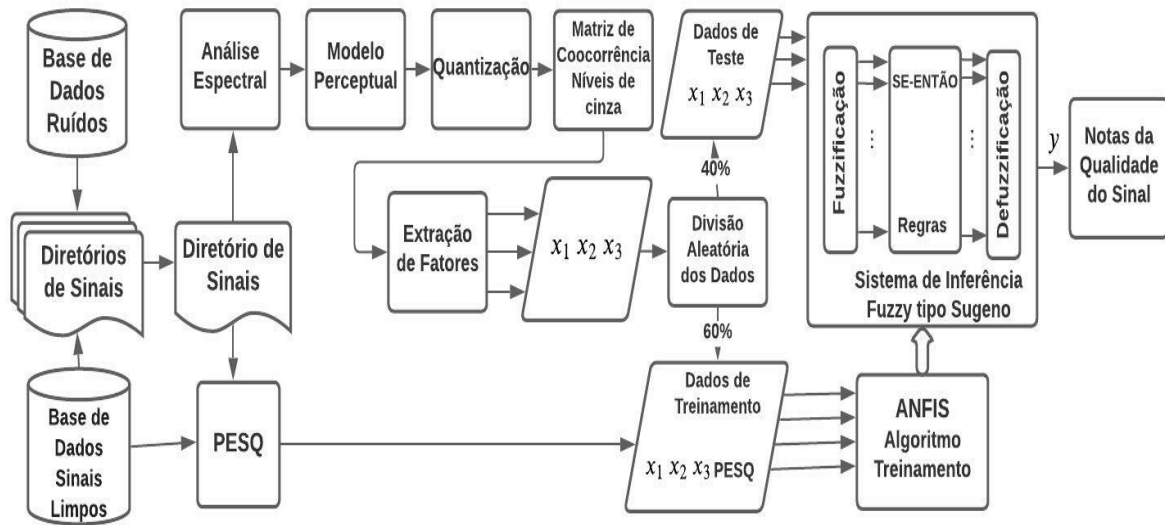
Fonte: Elaborada pelo próprio autor.

MÉTODO PROPOSTO

Neste Capítulo serão descritas as etapas da implementação do método proposto. Devido à quantidade de conceitos abordados para compor o método, diversas figuras serão apresentadas, no intuito de fornecer ao leitor uma ilustração do que ocorre em cada etapa do processo. Todas as implementações apresentadas neste Capítulo foram realizadas em MATLAB® (versão 9.4.0.813654 /R2018a) em um computador com processador Intel® Core™ i7.

Na Figura 25 é mostrada a estrutura do método proposto. A descrição detalhada de cada bloco será dada ao longo das próximas seções.

Figura 25: Diagrama de fluxo do método PESQ-ANFIS/FCM.



Fonte: Elaborada pelo próprio autor.

6.1 DIRETÓRIO DE SINAIS DE VOZ

Como entrada, o algoritmo que implementa o método proposto requer detalhes sobre a base de dados a ser utilizada. Desta forma, os arquivos de áudio são definidos em diretórios

de sinais e para a implementação três informações iniciais são necessárias:

- Os sinais de áudios a serem avaliados devem possuir 8 kHz ou 16 kHz de frequência de amostragem: uma restrição atribuída pelo algoritmo PESQ.
- Cada diretório deve portar nome específico: o que favorece a busca pelo arquivo.
- Em cada diretório os arquivos de áudio devem conter áudios com nomes específicos para: o conjunto de sentenças originais tomados como referência (sinal livre de ruído adicional) e para cada conjunto de sinais contaminados com diferentes tipos de ruídos, em diferentes níveis de contaminação.

Foram utilizadas duas bases de dados, sendo a primeira delas para ajuste dos parâmetros e análise dos resultados e a outra para comprovação do modelo. A seguir, é feita uma breve descrição do conteúdo de cada base.

6.2 DESCRIÇÃO DAS BASES DE DADOS

A base de dados NOIZEUS, proposta em Hu e Loizou (2007), foi desenvolvida especificamente para avaliar algoritmos de melhoramento de voz. Nela os sinais são amostrados a 8 kHz, quantizados com 16 bits por amostra e reproduzidos por um único canal. São trinta sentenças produzidas por três oradores do sexo masculino e três do sexo feminino, de forma a incluir todos os fonemas da língua inglesa americana, e receberam uma filtragem para simular as características de frequências recebidas por aparelhos telefônicos. A base de dados ainda conta com sinais contaminados por diferentes ruídos com relação sinal ruído (SNR) de 0, 5, 10 e 15 dB, constituindo diretórios com 150 sinais salvos em formato WAV, utilizando um codificador de forma de onda - PCM - (do inglês: *Pulse Code Modulation*) da recomendação ITU-T G.711. Para as simulações realizadas nesta pesquisa utilizaram-se ruídos de carro, multidão, aeroporto, restaurante, salão de exposições, estação, rua e trem. Segue abaixo uma breve descrição sobre cada tipo de ruído considerado.

- Carro: ruído captado no interior de um veículo em movimento por rodovias. Este ruído é gerado pelo rolamento dos pneus sobre o asfalto e também pelo som emitido pelo motor.
- Multidão: áudio captado por um número considerável de pessoas, onde todos estão

falando ao mesmo tempo. As palavras não são discerníveis.

- Aeroporto: áudio capturado no saguão de um aeroporto.
- Salão de exposições: áudio captado dentro de uma sala grande contendo um grande número de pessoas conversando. Pode ser similar a áreas no interior de shoppings. As palavras são pouco discerníveis.
- Restaurante: áudio captado dentro de um restaurante. Com sons de talheres em contato com pratos e pessoas conversando. Possui pouco discernimento das palavras.
- Estação: áudio captado em uma estação de trem.
- Rua: áudio captado em uma rua na região central de uma grande cidade, onde sons dos motores e buzinas de veículos aparecem frequentemente.
- Trem: áudio captado no interior de um trem em movimento.

Para a comprovação do modelo, serão utilizados 500 sinais limpos da base de dados TIMIT de Garofolo *et al.* (1993) e adicionados com ruídos de carro da base de dados da NOISEX de Varga e Steeneken (1993), com os mesmos quatro níveis de SNR, formando um diretório com 2500 sinais amostrados a 16 kHz e quantizados com 16 bits.

6.3 ANÁLISE ESPECTRAL

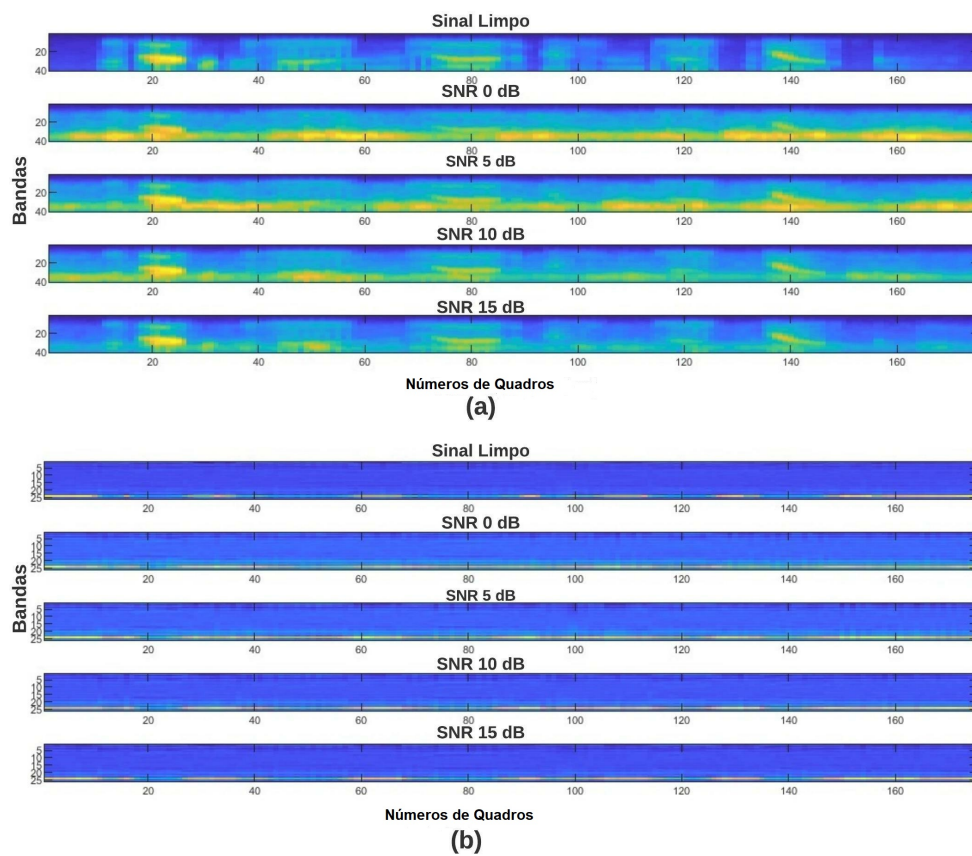
Para compor esta etapa, o modelo executa três etapas: Divisão de sinais em quadros, Janelamento de Hamming e aplicação da FFT descrita na Seção 3.2. Cada sinal de entrada é dividido em quadros e cada quadro é transformado para o domínio da frequência, usando janelas de Hamming calculadas pela equação (20), com comprimento fixado em 256 pontos (32 ms), com sobreposição de 50% e processados pela FFT, equação (21).

6.4 MODELO PERCEPTIVO

Com os sinais no domínio da frequência são utilizados os modelos de banda crítica baseados nas funções de espalhamento dadas pelas equações (17) e (18), destacadas no Capítulo 3, os bancos de filtros são convoluídos com a representação espectral do sinal que está sendo

analisado para calcular as respostas por banco. Para obter um modelo perceptivo, primeiro se define uma matriz com espaçamento igual às frequências das escalas psicoacústicas. Em seguida, cria-se as funções de propagação convoluindo com o espectro de potência, somando a contribuição dentro de cada banda crítica. As conversões intensidade-volume calculadas pela equação (22) vêm a seguir para criar uma resposta psicoacústica gerando uma matriz da percepção auditiva do sinal de voz.

Figura 26: Plano resposta do modelo perceptivo de (a) Hermansky e (b) MFCC para o diretório de sinais com ruído de trem.



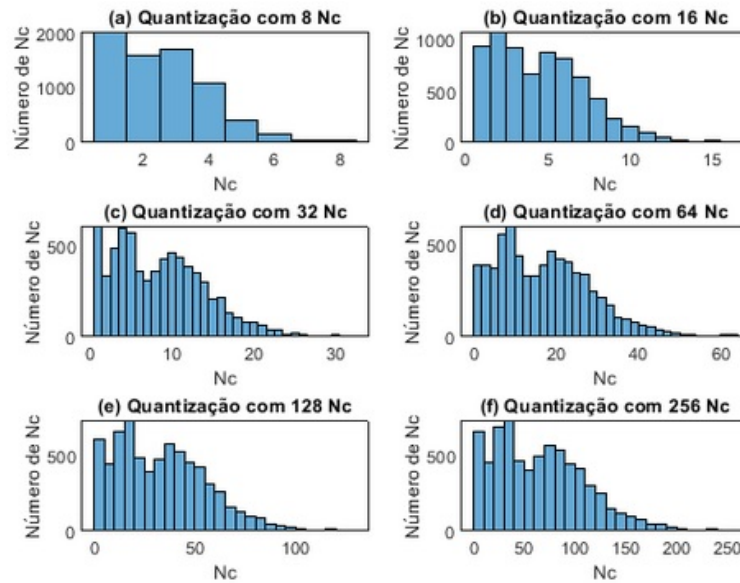
Fonte: Elaborada pelo próprio autor.

As imagens geradas no plano Bandas $\times$ Quadros, dadas pelas Figuras 26a e b, ilustram as matrizes de percepção auditiva para o modelo de Hermansky e o modelo MFCC, respectivamente. Para cada sinal de voz processado pelo modelo, uma nova representação matricial é formada e preparada para execução da próxima etapa.

6.5 QUANTIZAÇÃO

A operação realizada pelo bloco de quantização consiste em converter os valores reais do espectro de potência por banda, a partir da matriz de percepção auditiva, em níveis de cinza. Alguns níveis são representados por histogramas ilustrados nas Figuras 26a, b, c e d em níveis de cinza (Nc) aplicando a equação (27) abordada na Seção 4.1. A matriz quantizada $G(n, q)$ é composta de números inteiros, e a sua estrutura dependerá da escolha do Nc .

Figura 27: Quantização linear para sinal limpo da base de dados NOIZEUS.

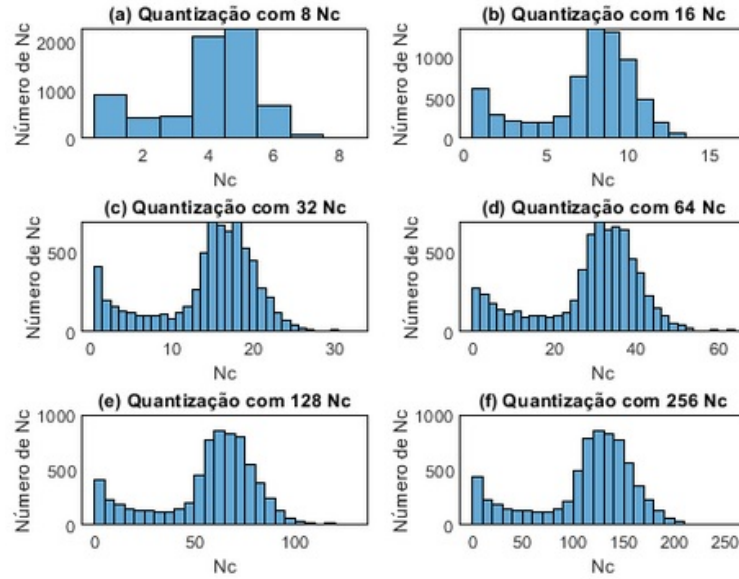


Fonte: Elaborada pelo próprio autor.

Na Figura 27 são apresentados seis níveis de quantização (8, 16, 32, 64, 128 e 256) de $Bs_{(\Psi, K)}$ para um sinal limpo. A mesma simulação é destacada na Figura 28 para um sinal ruidoso. Pode-se perceber a diferença ocorrida na quantização das duas figuras, este efeito é fundamental, para execução da próxima etapa do modelo. O número de elementos da escala de cinza Nc foi definido como $2^3 = 8$, com tons de limites possíveis variando de 0 a 7, e o

número de bits necessários para representação dos pixels igual a 3. Esta escolha foi obtida por meio de análises levando em consideração o custo computacional, pois o modelo opera com um número elevado de sinais.

Figura 28: Quantização linear para sinal contaminado com ruído de multidão com SNR 0 dB da base de dados NOIZEUS.



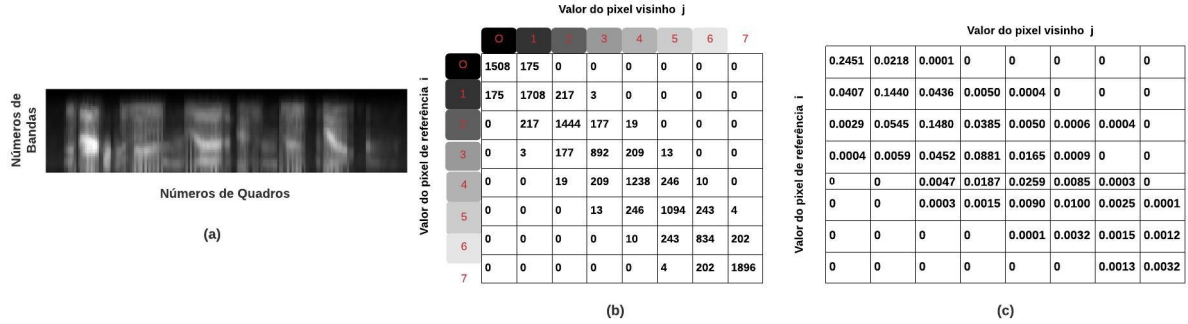
Fonte: Elaborada pelo próprio autor.

6.6 CÁLCULO DA MATRIZ DE COCORRÊNCIA DE NÍVEIS DE CINZA - MCNC

O cálculo dos índices de coocorrência segue o procedimento descrito no Capítulo 4, computacionalmente representado nas Figuras 20a e b. O nível de cinza N_c que aparece em cada pixel para matriz quantizada varia de 0 a N_c . A distribuição pode ser especificada pela matriz de frequências relativas, $M(i, j)$, com dois pixels vizinhos separados pela distância - D ao longo de $G(n, q)$, com um nível de cinza i e o outro com nível de cinza j . Tais matrizes MCNC são obtidas das relações angulares e da distância entre os pixels vizinhos.

Considerando as características do sinal de voz, o deslocamento foi realizado para as quatro direções 0° , 45° , 90° e 135° , representadas pelas respectivas equações (31), (32), (33) e (34).

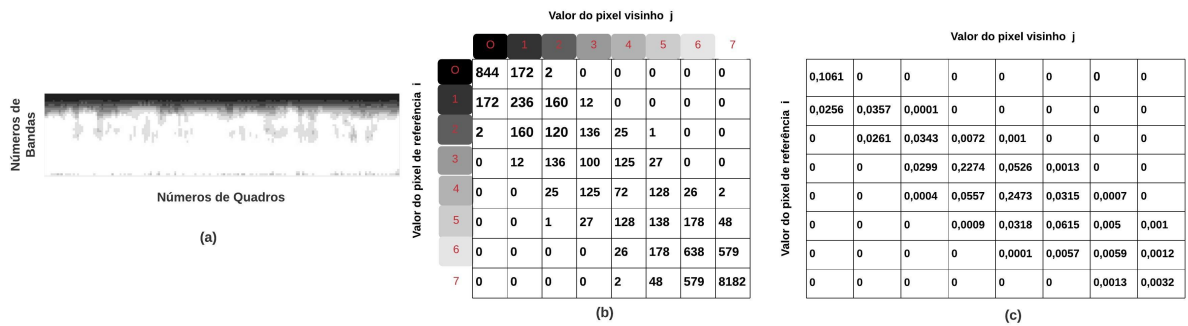
Figura 29: (a) Imagem quantizada com 8 níveis de cinza para um sinal limpo da base de dados NOIZEUS. (b) MCNC do mesmo sinal na direção 90° e (c) MCNC normalizada pela equação (35).



Fonte: Elaborada pelo próprio autor.

Na Figura 29a é apresentada, na forma de imagem, uma quantização em 8 níveis de cinza para uma matriz de percepção auditiva de sinal limpo com 193 quadros e 40 bandas críticas, na Figura 29b está sua correspondente MCNC com a dimensão igual ao nível de quantização e na Figura 29c está a MCNC normalizada pela equação (35). Na Figura 30b é destacada a imagem MCNC para a matriz de percepção auditiva para um sinal contaminado por ruído de multidão com 0 dB de SNR. Nota-se que há diferença entre as matrizes das figuras 29b e 30b comparando os elementos de suas diagonais principais. Estas variações diferem para cada sinal limpo e para cada conjunto de sinais contaminados com ruídos e SNR distintos. Esta distinção é levada em conta para o cálculo dos fatores utilizados no próximo bloco do modelo proposto.

Figura 30: (a) Imagem quantizada com 8 níveis de cinza para um sinal com ruído de multidão com $\text{SNR} = 0\text{dB}$ da base de dados NOIZEUS, (b) MCNC do mesmo sinal na direção 90° e (c) MCNC normalizada pela equação (35).



Fonte: Elaborada pelo próprio autor.

6.7 EXTRAÇÃO DOS FATORES

O bloco de extração de fatores atua na matriz de coocorrência para calcular os 19 fatores de Haralick. Estes são representados por equações destacadas na Tabela 3 e foram implementados para descrever a característica de cada sinal. O modelo opera com um conjunto de três dos 19 fatores, e estes, constituem uma matriz de dados, sendo o número linhas definidas pela quantidade de sinais e três colunas representadas pelos fatores.

Para justificar as medidas dos fatores de Haralick e as variações que ocorrem ao adotá-los, para compor as entradas do modelo de regressão, na Tabela 6 são apresentados os resultados das simulações realizadas para dois sinais limpos com voz masculina e feminina e contaminados com ruído de multidão, com 0, 5, 10 e 15 dB de SNR para as quatro direções da matriz de coocorrência. Estes resultados são fundamentais para o modelo, pois determinam padrões para que a rede ANFIS, que será implementada, reconheça o sinal e o avalie. Fazendo uma análise para o fator 1 na direção 90° da matriz de coocorrência de níveis de cinza - DMCNC da Tabela 6 definido como Autocorrelação, nota-se que os valores dos sinais limpos de vozes masculinas ML e masculinas com ruídos de multidão (MB/15dB, MB/10dB, MB/5dB e MB/0dB) estão na ordem crescente, o mesmo acontece com os sinais limpos de vozes feminina FL e os contaminados (FB/15dB, FB/10dB, FB/5dB e FB/0dB), isto é verificado também para outras DMCNC e para os outros fatores.

Verificou-se por meio das simulações que não existe um único e melhor fator para identificar padrões, mas sim um conjunto de fatores que caracterizam a matriz de percepção auditiva da DMCNC. Estes resultados são motivos de análise e discussão e serão abordados em um capítulo específico, onde serão implementadas todas as combinações para um conjunto contendo três fatores de Haralick.

Tabela 6: Resultados dos 19 fatores de Haralick para sinais limpos e contaminados com ruído de multidão da base de dados NOIZEUS.

DMCNC	Sinais	Fatores de Haralick																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0°	ML	24,72	797,34	6,05	0,93	0,92	0,98	0,65	0,53	0,06	3,15	0,77	-0,48	0,93	0,79	0,12	8,85	2,60	5,62	21,56
	MB/15dB	32,46	835,26	-53,87	0,48	0,95	0,77	0,35	0,36	0,08	2,84	0,83	-0,55	0,94	0,84	0,14	10,58	2,48	4,71	18,37
	MB/10dB	37,51	1061,85	-83,45	0,35	0,97	0,67	0,26	0,30	0,10	2,65	0,85	-0,59	0,94	0,86	0,21	11,44	2,38	5,00	19,64
	MB/5dB	42,34	1355,14	-114,35	0,26	0,98	0,58	0,21	0,24	0,16	2,37	0,88	-0,63	0,94	0,89	0,34	12,18	2,16	5,36	21,17
	MB/0dB	49,15	1795,15	-156,62	0,15	0,99	0,42	0,13	0,14	0,38	1,71	0,93	-0,69	0,91	0,93	0,60	13,22	1,59	5,53	21,96
	FL	28,87	1078,15	-3,54	0,90	0,94	0,87	0,72	0,43	0,10	2,87	0,83	-0,54	0,94	0,84	0,24	9,46	2,44	6,94	26,86
	FB/15dB	37,45	1149,88	-80,53	0,33	0,97	0,63	0,27	0,26	0,13	2,57	0,88	-0,64	0,95	0,88	0,29	11,30	2,33	5,67	22,36
	FB/10dB	42,77	1418,10	-118,10	0,28	0,98	0,59	0,22	0,24	0,18	2,33	0,89	-0,64	0,94	0,89	0,38	12,21	2,12	5,65	22,32
	FB/5dB	47,45	1728,93	-149,69	0,16	0,99	0,43	0,14	0,15	0,34	1,83	0,93	-0,70	0,93	0,93	0,57	12,92	1,71	5,79	22,98
	FB/0dB	50,50	1873,52	-163,48	0,08	0,99	0,27	0,07	0,08	0,50	1,36	0,96	-0,79	0,91	0,96	0,70	13,41	1,31	5,57	22,21
45°	ML	24,93	772,13	4,93	1,06	0,90	1,05	0,67	0,62	0,05	3,26	0,73	-0,42	0,91	0,75	0,11	8,92	2,65	5,56	21,20
	MB/15dB	32,78	792,42	-52,06	0,61	0,93	0,86	0,39	0,47	0,07	2,98	0,78	-0,47	0,92	0,79	0,14	10,68	2,53	4,57	17,67
	MB/10dB	37,90	1004,46	-80,32	0,49	0,95	0,78	0,32	0,41	0,09	2,81	0,80	-0,50	0,92	0,80	0,20	11,55	2,43	4,82	18,79
	MB/5dB	42,79	1281,16	-109,86	0,41	0,96	0,73	0,29	0,36	0,15	2,53	0,83	-0,53	0,91	0,83	0,33	12,30	2,22	5,15	20,17
	MB/0dB	49,62	1681,46	-148,48	0,30	0,97	0,62	0,23	0,26	0,38	1,84	0,88	-0,57	0,88	0,88	0,60	13,34	1,62	5,25	20,69
	FL	29,14	1049,01	-5,27	1,02	0,93	0,97	0,75	0,52	0,09	2,99	0,79	-0,48	0,92	0,80	0,24	9,54	2,49	6,87	26,47
	FB/15dB	37,82	1097,56	-78,56	0,47	0,96	0,76	0,33	0,37	0,12	2,74	0,82	-0,54	0,93	0,83	0,29	11,41	2,40	5,51	21,58
	FB/10dB	43,19	1344,20	-113,81	0,42	0,96	0,73	0,30	0,35	0,17	2,48	0,83	-0,54	0,92	0,84	0,38	12,32	2,17	5,44	21,33
	FB/5dB	47,87	1630,16	-143,04	0,30	0,97	0,63	0,23	0,27	0,33	1,98	0,87	-0,58	0,90	0,87	0,57	13,04	1,75	5,53	21,81
	FB/0dB	50,92	1744,16	-154,19	0,23	0,98	0,53	0,19	0,20	0,50	1,51	0,90	-0,65	0,87	0,91	0,70	13,53	1,35	5,28	20,87
90°	ML	25,22	823,30	9,53	0,24	0,98	0,56	0,19	0,23	0,08	2,73	0,89	-0,68	0,97	0,89	0,14	8,89	2,54	5,57	22,04
	MB/15dB	32,87	804,62	-49,97	0,28	0,97	0,60	0,21	0,26	0,10	2,67	0,87	-0,63	0,96	0,87	0,17	10,67	2,46	4,57	17,98
	MB/10dB	37,96	1010,66	-79,45	0,28	0,97	0,61	0,21	0,27	0,11	2,58	0,87	-0,62	0,95	0,87	0,23	11,54	2,37	4,82	18,99
	MB/5dB	42,81	1283,11	-109,24	0,28	0,97	0,61	0,21	0,27	0,17	2,38	0,87	-0,62	0,94	0,87	0,36	12,30	2,17	5,15	20,30
	MB/0dB	49,64	1684,58	-148,54	0,24	0,98	0,55	0,19	0,22	0,40	1,76	0,89	-0,63	0,90	0,89	0,62	13,34	1,58	5,25	20,77
	FL	29,37	1091,17	-1,64	0,22	0,98	0,53	0,18	0,21	0,12	2,56	0,90	-0,70	0,97	0,90	0,26	9,51	2,39	6,88	27,29
	FB/15dB	37,86	1101,90	-77,32	0,25	0,98	0,57	0,19	0,24	0,14	2,53	0,88	-0,66	0,96	0,88	0,31	11,40	2,33	5,51	21,78
	FB/10dB	43,28	1349,34	-113,51	0,26	0,98	0,58	0,20	0,25	0,20	2,32	0,88	-0,63	0,94	0,88	0,41	12,32	2,12	5,44	21,49
	FB/5dB	47,92	1632,58	-143,03	0,23	0,98	0,55	0,19	0,22	0,35	1,89	0,89	-0,64	0,91	0,89	0,58	13,04	1,72	5,53	21,88
	FB/0dB	50,93	1745,99	-154,32	0,20	0,98	0,50	0,17	0,19	0,50	1,47	0,91	-0,68	0,88	0,91	0,71	13,53	1,72	5,53	21,88
135°	ML	24,94	773,35	5,04	1,04	0,91	1,04	0,67	0,61	0,06	3,25	0,74	-0,43	0,91	0,76	0,11	8,92	2,64	5,56	21,21
	MB/15dB	32,78	792,33	-52,00	0,61	0,93	0,86	0,39	0,46	0,07	2,99	0,78	-0,47	0,92	0,79	0,14	10,68	2,53	4,57	17,67
	MB/10dB	37,90	1004,42	-80,32	0,49	0,95	0,78	0,32	0,41	0,09	2,81	0,80	-0,50	0,92	0,80	0,20	11,55	2,43	4,82	18,79
	MB/5dB	42,78	1280,97	-109,84	0,42	0,96	0,73	0,29	0,36	0,15	2,54	0,83	-0,52	0,91	0,83	0,33	12,30	2,22	5,15	20,16
	MB/0dB	49,61	1680,71	-148,44	0,30	0,97	0,62	0,24	0,26	0,38	1,84	0,87	-0,57	0,88	0,88	0,60	13,34	1,62	5,25	20,69
	FL	29,14	1049,95	-5,27	1,01	0,93	0,96	0,75	0,51	0,10	2,98	0,79	-0,48	0,92	0,80	0,24	9,54	2,48	6,87	26,47
	FB/15dB	37,82	1097,76	-78,55	0,47	0,96	0,76	0,33	0,37	0,12	2,73	0,83	-0,55	0,93	0,83	0,29	11,41	2,39	5,51	21,58
	FB/10dB	43,19	1344,08	-113,86	0,42	0,96	0,74	0,29	0,36	0,17	2,49	0,83	-0,53	0,91	0,83	0,38	12,32	2,17	5,44	21,33
	FB/5dB	47,87	1629,73	-143,04	0,30	0,97	0,62	0,23	0,26	0,33	1,98	0,87	-0,58	0,90	0,87	0,56	13,04	1,76	5,53	21,81
	FB/0dB	50,92	1744,63	-154,22	0,23	0,98	0,53	0,19	0,20	0,50	1,51	0,91	-0,65	0,87	0,91	0,70	13,53	1,35	5,28	20,87

Fonte: Elaborada pelo próprio autor.

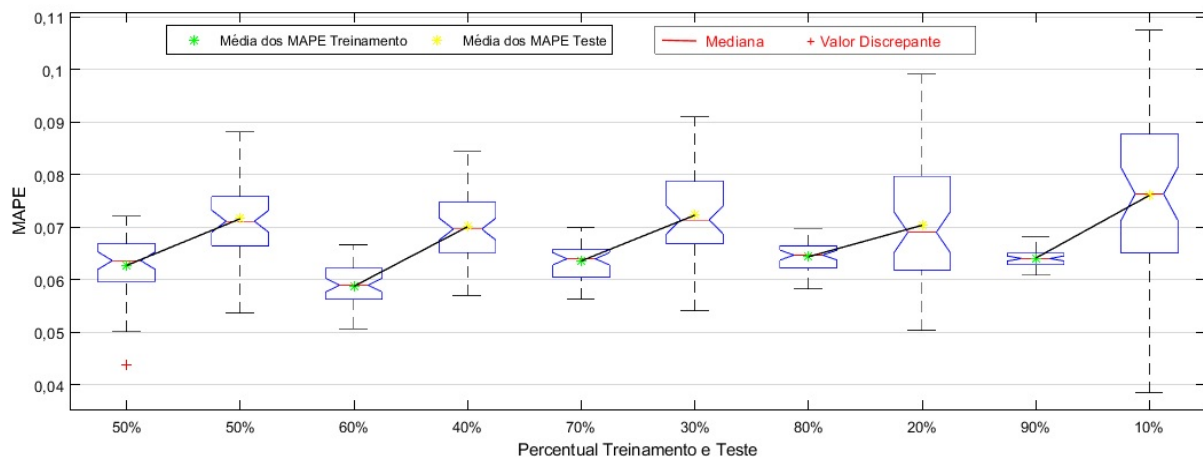
6.8 ALGORITMO PESQ

As avaliações do sinal para compor a fase de treinamento da ANFIS são obtidas pelo bloco que executa o algoritmo PESQ abordado no Capítulo 2. Nele é necessário definir o sinal limpo como referência, que quando comparado consigo próprio gera a maior nota. Comparando-o com outros, dependendo do nível de contaminação do ruído, resulta em notas inferiores.

6.9 DIVISÃO ALEATÓRIA DOS DADOS

O conjunto de dados obtidos da extração de fatores é dividido de forma aleatória sendo que 60% deles são utilizados para o treinamento e 40% para teste da rede. Este processo de divisão aleatória foi aplicado ao mesmo conjunto de dados cinquenta vezes, com a finalidade de obter diferentes composições, evitando ambiguidades nas análises das medidas que avaliam o desempenho do modelo. Esta divisão foi apresentada no trabalho de Nguyen *et al.* (2021) demonstrando bons resultados, confirmados também nas simulações realizadas neste trabalho. Na Figura 31 são ilustrados pares de *boxplot* representando as fases de treinamento e teste contendo os resultados MAPE para os 150 sinais da base de dados NOIZEUS com ruído de multidão. Os valores médios são indicadores de desempenho evidenciando que a escolha de 60% dos sinais para treinamento e 40% para teste proporcionou os melhores resultados entre os demais percentuais.

Figura 31: Resultados do MAPE para a divisão percentual de treinamento e teste para o diretório de sinais da base de dados NOIZEUS com ruído de Multidão .



Fonte: Elaborada pelo próprio autor.

6.10 DADOS DE TREINAMENTO

O bloco de dados para o treinamento é constituído de uma matriz, sendo o número de linhas representado pela quantidade de sinais, e quatro colunas, as três primeiras para os fatores extraídos dos sinais e a última para a avaliação PESQ correspondente.

6.11 DADOS DE TESTE

O bloco de dados de teste é formado por uma matriz contendo somente três colunas com os resultados dos fatores extraídos do sinal.

6.12 ANFIS

O bloco constituído pela ANFIS, algoritmo apresentado no Capítulo 5, treina o sistema FIS utilizando os dados obtidos do bloco divisão aleatória dos dados. A primeira etapa da ANFIS é a fuzzificação dos dados de entrada do sistema, ou seja, definir o algoritmo de fuzzificação, a função de pertinência e as suas respectivas quantidades de variáveis linguísticas é fundamental para o desempenho da rede. Desta forma, foi criada uma metodologia para determinar a configuração da ANFIS.

6.12.1 MÉTRICAS DE AVALIAÇÃO DE DESEMPENHO

Foram realizadas diversas simulações com diferentes parâmetros, utilizando os três algoritmos de fuzzificação e o método Sugeno de inferência (SUGENO; KANG, 1986; SUGENO; TANAKA, 1991). Os resultados preditos pela ANFIS tanto na fase de treinamento quanto para teste passaram por uma análise estatística de regressão utilizando as medidas MAPE (do inglês - *Mean Absolute Percentage Error*), RMSE (do inglês - *Root Mean Squared Error*) e R^2 (do inglês - *Determination Coefficient*) dadas pelas equações:

$$\text{MAPE} = \left(\frac{1}{n} \sum_i^n \left| \frac{S_{r_i} - S_{p_i}}{S_{r_i}} \right| \right) \times 100; \quad (88)$$

$$\text{RMSE} = \sqrt{\frac{\sum_i^n (S_{r_i} - S_{p_i})^2}{n}}; \quad (89)$$

$$R^2 = 1 - \left(\frac{\sum_i^n (S_{r_i} - S_{p_i})^2}{\sum_i^n (S_{r_i} - \overline{S_{p_i}})^2} \right) \quad (90)$$

onde n representa a quantidade de sinais, S_{r_i} é considerado o sinal real de entrada e S_{p_i} e $\overline{S_{p_i}}$ o sinal predito e sua média, respectivamente. Valores de MAPE e RMSE baixos são indicadores de boa precisão, enquanto que quanto mais próximo de 1 for o valor de R^2 melhor será a regressão.

6.12.2 METODOLOGIA PARA CONFIGURAÇÃO DA ANFIS

Para configurar a ANFIS foram utilizados os sinais da base de dados NOIZEUS e escolhidos os fatores 1-Auto Correlação, 11-Homogeneidade e 13-Medida da Informação de Correlação 2, apresentados na Tabela 3. Na Seção 7.1 será apresentada uma discussão sobre a escolha dos três melhores fatores que podem ser aplicados ao modelo para que preserve seu desempenho.

Com o conjunto de dados destinado para treinamento da ANFIS, inicialmente foram aplicados os algoritmos de agrupamento FCM, *Subtractive* e o de partição *Fuzzy Grid* para obtenção das funções de pertinência gaussianas e as suas respectivas variáveis linguísticas e/ou conjuntos difusos para comporem a etapa de fuzzificação, constituindo a primeira camada da ANFIS. Cabe ressaltar que os métodos de fuzzificação fornecem informações para obter as bases de regras iniciais e a estrutura *fuzzy* para geração do sistema de inferência correspondendo as segunda e terceira camadas da ANFIS (JAFARI *et al.*, 2021).

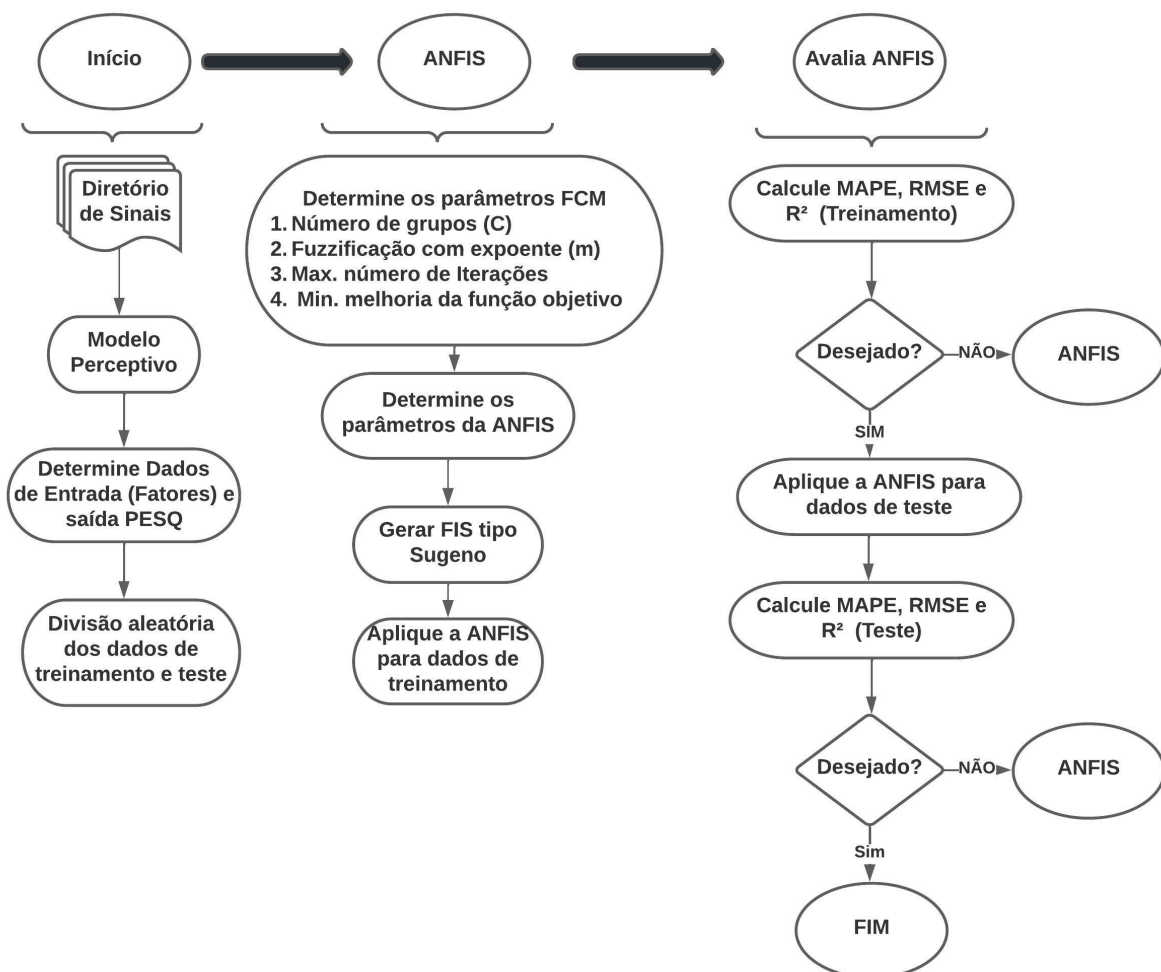
Para cada um dos oito diretórios da base de dados NOIZEUS foram extraídas quatro composições diferentes contendo três fatores baseados nas direções da matriz de coocorrência obtida na fase de extração.

Todas as simulações foram realizadas sob mesmas condições, utilizando cinquenta iterações, frutos das composições aleatórias geradas dos sinais na fase de treinamento e teste, mudando somente as variáveis linguísticas para cada algoritmo de fuzzificação. Para cada iteração foram calculadas as médias e os desvios padrões das medidas utilizadas para avaliação do

modelo. A média geral para cada algoritmo de fuzzificação com suas respectivas variáveis linguísticas será considerada como um indicador de desempenho e o desvio padrão como uma medida de dispersão que retrata o comportamento das medidas em relação a sua média no espaço gerado pelas iterações.

Na Figura 32 é apresentado o fluxograma do algoritmo FCM inserido na primeira camada da ANFIS. O agrupamento FCM foi aplicado utilizando como expoente de fuzzificação $m = 2$, com 100 iterações e critério de parada com erro $\leq 10^{-5}$. Considerando estes parâmetros para o modelo, os resultados das medidas estatísticas, para áudios do diretório com ruídos de multidão, nas quatro direções da matriz de coocorrência variando de duas até a oitava variável linguística, são mostrados na Figura 33a.

Figura 32: Diagrama de fluxo do algoritmo de fuzzificação FCM/ANFIS.



Fonte: Elaborada pelo próprio autor.

Seguindo o mesmo esquema do fluxograma apresentado na Figura 32, foi aplicado na etapa de fuzzificação o algoritmo de agrupamento *Subtractive*, tendo ajustado seus parâmetros para construir funções de pertinência com 2, 3, 4, 5, 6, 7 e 8 variáveis linguísticas. Os resultados, utilizando as medidas estatísticas para o mesmo diretório de áudios, são destacados na Figura 33b.

Adotando o algoritmo de particionamento *fuzzy Grid* na fase de fuzzificação da ANFIS no fluxograma da Figura 32, foi fixada a configuração da função de pertinência para cada variável de entrada com 2 até 8 variáveis linguísticas. Seus resultados utilizando as mesmas medidas avaliadoras e o mesmo diretório de áudios estão expostos na Figura 33c.

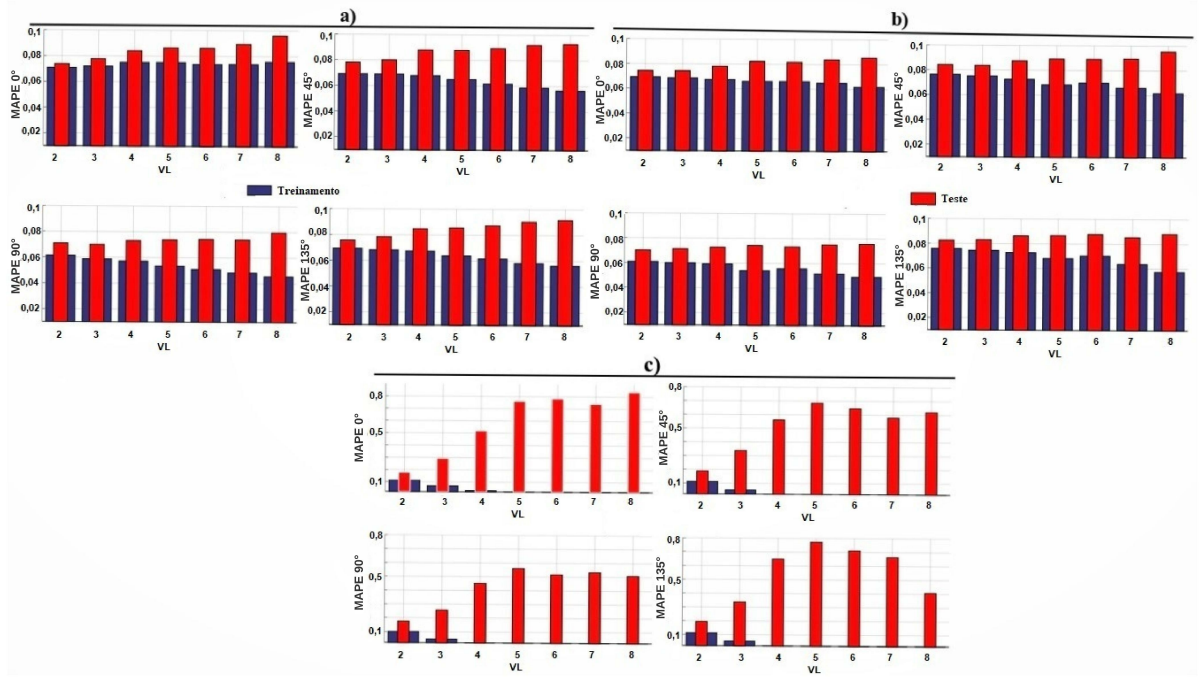
Ao observar a Figura 33 é possível perceber que os algoritmos de agrupamento FCM e *Subtractive* geraram resultados ligeiramente melhores do que o algoritmo de particionamento *Grid* em relação a sincronização entre as fases de treinamento e teste. A Figura 33c traduz o comportamento das medidas estatísticas do algoritmo *Grid*. Pode-se observar que em praticamente todas as direções da matriz de coocorrência, conforme as variáveis linguísticas aumentam, ocorre uma melhora significativa das medidas estatísticas na fase de treinamento e uma piora abrupta na fase de teste.

Esse descompasso entre o aumento das variáveis linguísticas e o comportamento das medidas ocorre de maneira análoga nos algoritmos de agrupamento, mas suave, sendo observado na média entre as fases de treinamento e teste, como pode ser visto nas Figuras 33a e b.

Os resultados observados na Figura 33 apontam para uma configuração do modelo com três variáveis linguísticas na direção 90°, com uma pequena vantagem do algoritmo de agrupamento FCM sobre os demais.

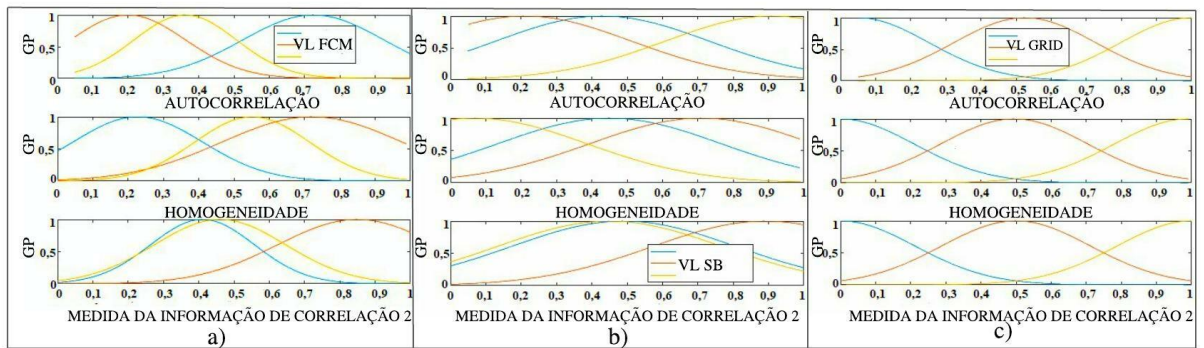
Para fins de ilustração, na Figura 34 são apresentados os comportamentos das funções de pertinência com três variáveis linguísticas para os três algoritmos de fuzzificação na direção 90° e com ruído de multidão. Na Figura 35 é apresentado o comportamento do modelo de regressão também para os três algoritmos de fuzzificação.

Figura 33: Resultados da medida MAPE para as fases de treinamento (azul) e teste (vermelho) nas quatro direções da matriz de coerência com ruído de multidão para os algoritmos de fuzzificação: a) FCM, b) *Subtractive* e c) *Grid*.



Fonte: Elaborada pelo próprio autor.

Figura 34: Função de pertinência com três variáveis linguísticas para os algoritmos: a) FCM, b) *Subtractive* e c) *Grid*.



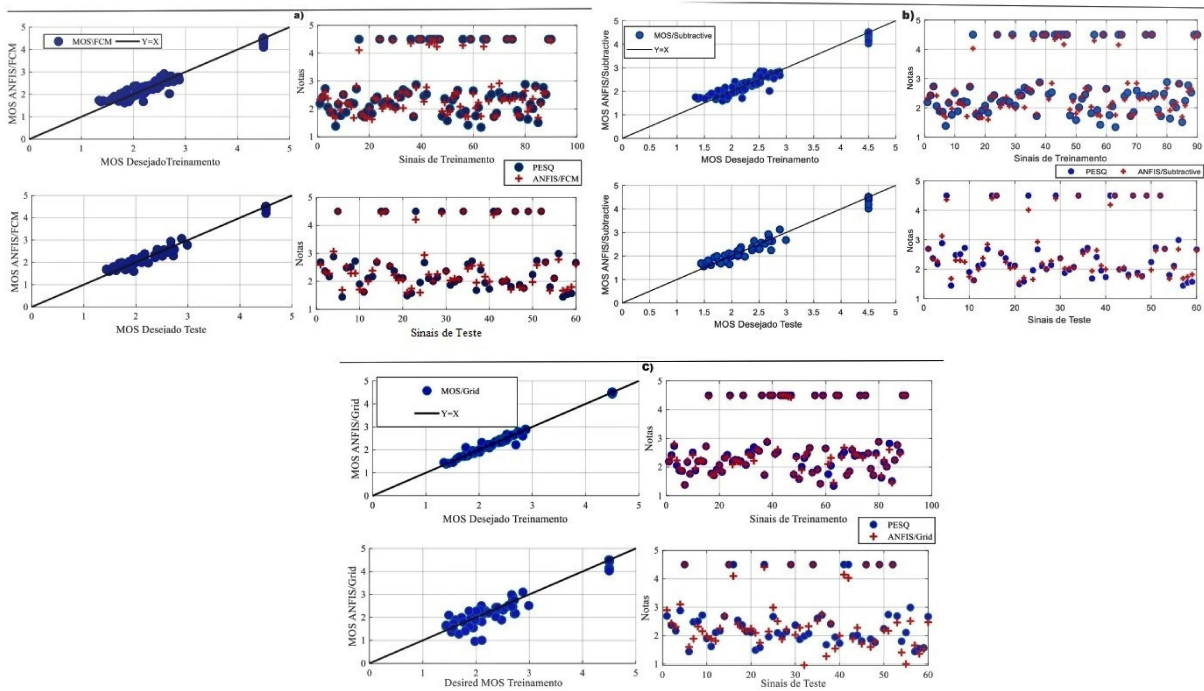
Fonte: Elaborada pelo próprio autor.

Para justificar esta escolha, na Tabela 7, são apresentados os resultados para o diretório com ruídos de carro para os três algoritmos de fuzzificação. Em destaque se encontra o algoritmo FCM na direção 90° com três variáveis linguísticas. Pode-se observar que embora a FCM apresente desvantagens na fase de treinamento para certas medidas estatísticas em

relação aos outros algoritmos, em contrapartida, proporciona os melhores resultados na fase de teste.

Na Tabela 8 são apresentados os resultados das medidas estatísticas obtidos para os algoritmos de agrupamento FCM, *Subtractive* e partição *Grid* com três variáveis linguísticas na direção de 90° para os ruídos de carro, multidão, aeroporto, estação, sala de exibição, restaurante, rua e trem.

Figura 35: Comportamento regressor do modelo na direção 90° para sinais contaminados com ruído de multidão e com os seguintes algoritmos de fuzzificação: a) FCM, b) *Subtractive* e c) *Grid*.



Fonte: Elabora pelo próprio autor.

Analisando a Figura 35, tanto na fase de treinamento quanto na de teste, verifica-se que se a predição fosse exata, ou seja, em uma condição de situação ideal, os pontos plotados no plano seguiriam a reta $y = x$. Porém, em situações reais, isso não é possível e sempre existirá um erro envolvido no processo de regressão. Nesse sentido, nota-se, pelo exemplo apresentado, que as avaliações MOS dos sinais possuem resultados satisfatórios em ambas as fases, uma vez que grande parte dos pontos plotados estão sobre ou muito próximo dessa reta.

Tabela 7: Média (Md) e Desvio Padrão (DP) das fases de treinamento e teste para os 150 sinais da base de dados NOIZEUS com ruído de carro em quatro direções da matriz de coocorrência.

DMC	Nº VL	MF	Md/DP	Treinamento			Teste		
				MAPE	RMSE	R ²	MAPE	RMSE	R ²
0º [0 1]	3	FCM	Md	0,0602	0,1561	0,9745	0,0689	0,1844	0,9649
			DP	0,0047	0,0087	0,0040	0,0072	0,0150	0,0097
	3	<i>Subtractive</i>	Md	0,0598	0,1583	0,9741	0,0712	0,2097	0,9550
			DP	0,0056	0,0148	0,0051	0,0077	0,0246	0,0111
	3	<i>Grid</i>	Md	0,0312	0,0886	0,9921	0,1207	0,3907	0,8336
			DP	0,0048	0,0101	0,0022	0,0253	0,1024	0,0898
45º [-1 1]	3	FCM	Md	0,0626	0,1723	0,9694	0,0721	0,1901	0,9594
			DP	0,0039	0,0125	0,0079	0,0085	0,0228	0,0146
	3	<i>Subtractive</i>	Md	0,0658	0,1901	0,9625	0,0815	0,2277	0,9428
			DP	0,0069	0,0224	0,0118	0,0130	0,0386	0,0232
	3	<i>Grid</i>	Md	0,0275	0,0819	0,9935	0,1449	0,4815	0,7771
			DP	0,0041	0,0086	0,0016	0,0524	0,2562	0,1232
90º [-1 0]	3	FCM	Md	0,0587	0,1512	0,9771	0,0683	0,1748	0,9665
			DP	0,0040	0,0083	0,0030	0,0063	0,0148	0,0078
	3	<i>Subtractive</i>	Md	0,0557	0,1465	0,9784	0,0714	0,1899	0,9600
			DP	0,0045	0,0104	0,0035	0,0075	0,0201	0,0117
	3	<i>Grid</i>	Md	0,0206	0,0670	0,9957	0,1437	0,4864	0,7841
			DP	0,0023	0,0082	0,0011	0,0455	0,1845	0,1148
135º [-1 -1]	3	FCM	Md	0,0617	0,1619	0,9729	0,0712	0,1934	0,9587
			DP	0,0043	0,0136	0,0062	0,0073	0,0221	0,0133
	3	<i>Subtractive</i>	Md	0,0620	0,1704	0,9701	0,0759	0,2231	0,9447
			DP	0,0057	0,0192	0,0080	0,0077	0,0282	0,0174
	3	<i>Grid</i>	Md	0,0254	0,0763	0,9943	0,1502	0,5077	0,7501
			DP	0,0038	0,0096	0,0015	0,0506	0,2011	0,1376

Fonte: Elaborada pelo próprio autor.

Ainda na Figura 35, destacam-se os algoritmos FCM e *Subtractive*, com pontos menos dispersos em relação a reta $y = x$. O desempenho do modelo também foi justificado pelos resultados do desvio padrão. Seus baixos valores indicam que os resultados das medidas estatísticas são considerados relevantes para caracterizar o modelo.

Ainda cabe ressaltar que, embora os algoritmos de agrupamento apresentem comportamentos parecidos e resultados muito próximos, o FCM se destaca pelo fato que seus ajustes de parâmetros são padrão para todos os diretórios. O mesmo não ocorre para o algoritmo *Subtractive*, pois seus parâmetros devem ser ajustados para sua melhor configuração quando lhe é apresentado um novo diretório de sinais.

Tabela 8: Média e Desvio Padrão da ANFIS para as fases de treinamento e teste com os sinais da base de dados NOIZEUS com os seus respectivos ruídos na direção 90° da matriz de coocorrência para os algoritmos de fuzzificação FCM, *Subtractive* e *Grid*.

Ruídos	MF	Md/DP	Treinamento			Teste		
			MAPE	RMSE	R ²	MAPE	RMSE	R ²
Carro	FCM	Md	0,0587	0,1512	0,9771	0,0683	0,1748	0,9665
		DP	0,0040	0,0083	0,0030	0,0063	0,0148	0,0078
	<i>Subtractive</i>	Md	0,0557	0,1465	0,9784	0,0714	0,1899	0,9600
		DP	0,0045	0,0104	0,0035	0,0075	0,0201	0,0117
	<i>Grid</i>	Md	0,0206	0,0670	0,9957	0,1437	0,4864	0,7841
		DP	0,0023	0,0082	0,0011	0,0455	0,1845	0,1148
Multidão	FCM	Md	0,0587	0,1601	0,9712	0,0700	0,1877	0,9613
		DP	0,0032	0,0085	0,0031	0,0056	0,0158	0,0073
	<i>Subtractive</i>	Md	0,0602	0,1671	0,9683	0,0713	0,1914	0,9593
		DP	0,0033	0,0095	0,0037	0,0053	0,0153	0,0076
	<i>Grid</i>	Md	0,0241	0,0804	0,9930	0,1286	0,4598	0,8141
		DP	0,0036	0,0133	0,0021	0,0359	0,1824	0,0898
Aeroporto	FCM	Md	0,0574	0,1535	0,9748	0,0658	0,1751	0,9672
		DP	0,0039	0,0090	0,0037	0,0069	0,0215	0,0112
	<i>Subtractive</i>	Md	0,0542	0,1519	0,9742	0,0675	0,1827	0,9611
		DP	0,0037	0,0082	0,0037	0,0075	0,0177	0,0087
	<i>Grid</i>	Md	0,0125	0,0482	0,9974	0,2735	1,0339	0,4599
		DP	0,0029	0,0099	0,0011	0,0678	0,3727	0,1493
Estação	FCM	Md	0,0606	0,1660	0,9707	0,0724	0,1976	0,9572
		DP	0,0044	0,0119	0,0046	0,0067	0,0191	0,0098
	<i>Subtractive</i>	Md	0,0588	0,1575	0,9735	0,0752	0,2066	0,9528
		DP	0,0046	0,0117	0,0045	0,0079	0,0222	0,0118
	<i>Grid</i>	Md	0,0230	0,0809	0,9934	0,2092	0,7462	0,6105
		DP	0,0023	0,0072	0,0012	0,0594	0,3078	0,1669
S.Exibição	FCM	Md	0,0687	0,1763	0,9679	0,0766	0,1978	0,9590
		DP	0,0048	0,0081	0,0049	0,0096	0,0194	0,0097
	<i>Subtractive</i>	Md	0,0659	0,1706	0,9701	0,0781	0,2023	0,9573
		DP	0,0051	0,0079	0,0048	0,0107	0,0190	0,0106
	<i>Grid</i>	Md	0,0332	0,0979	0,9905	0,1578	0,5352	0,5998
		DP	0,0043	0,0103	0,0025	0,1661	0,1303	0,0318
Restaurante	FCM	Md	0,0576	0,1708	0,9666	0,0651	0,1867	0,9590
		DP	0,0045	0,0144	0,0063	0,0077	0,0253	0,0134
	<i>Subtractive</i>	Md	0,0552	0,1621	0,9700	0,0696	0,1993	0,9530
		DP	0,0051	0,0140	0,0057	0,0093	0,0247	0,0155
	<i>Grid</i>	Md	0,0231	0,0789	0,9932	0,2081	0,7923	0,5998
		DP	0,0050	0,0144	0,0026	0,0576	0,2864	0,1370

continua na próxima página

continuação da página anterior

Rua	FCM	Md	0,0737	0,1856	0,9642	0,0849	0,2102	0,9547
		DP	0,0076	0,0141	0,0048	0,0110	0,0161	0,0082
	<i>Subtractive</i>	Md	0,0694	0,1739	0,9679	0,0854	0,2115	0,9536
		DP	0,0088	0,0179	0,0091	0,0107	0,0212	0,0101
	<i>Grid</i>	Md	0,0310	0,0978	0,9905	0,2148	0,7643	0,6192
		DP	0,0054	0,0105	0,0020	0,0671	0,2744	0,1326
Trem	FCM	Md	0,0640	0,1671	0,9727	0,0731	0,1882	0,9647
		DP	0,0052	0,0090	0,0043	0,0099	0,0196	0,0109
	<i>Subtractive</i>	Md	0,0628	0,1653	0,9733	0,0748	0,1890	0,9662
		DP	0,0056	0,0102	0,0048	0,0099	0,0178	0,0091
	<i>Grid</i>	Md	0,0448	0,1214	0,9857	0,1048	0,3289	0,8765
		DP	0,0055	0,0128	0,0035	0,0242	0,1106	0,0994

Fonte: Elaborada pelo próprio autor.

Tabela 9: Número de Regras - NR - empregadas para cada variável linguística e Método de Fuzzificação - MF - utilizando diretório com ruído de multidão.

MF		Número de Variáveis Linguísticas						
		2	3	4	5	6	7	8
FCM	NR	2	3	4	5	6	7	8
<i>Subtractive</i>	NR	2	3	4	4	4	5	9
<i>Grid</i>	NR	8	27	64	125	216	373	512

Fonte: Elaborada pelo próprio autor.

Sobre a quantidade de regras envolvidas no FIS tipo Sugeno utilizado para compor o modelo proposto, é importante destacar que os métodos de agrupamentos para fuzzificação levam vantagem quando comparados ao método de partição *Grid*, como pode ser visto na Tabela 9. Quanto menor a quantidade de regras inseridas no FIS menor será o esforço computacional, pela baixa dimensionalidade da base de regras, pois os parâmetros da premissa e os parâmetros do consequente associados a i -ésima regra constituem um conjunto que configura a fase de treinamento da ANFIS.

Observa-se na Tabela 9 que o FCM produz uma quantidade de regras análogas a quantidade de variáveis, o mesmo não ocorre para o *Subtractive*, pois suas regras dependem dos ajustes dos parâmetros que normalmente não possuem um padrão em relação as variáveis linguísticas. Já o particionamento de fuzzificação *Grid* depende exclusivamente do número de

entradas. Como são três as quantidades de regras seguem uma exponencial, 3^3 gerando 27 regras.

Diante das simulações realizadas, os resultados tomando como base as medidas MAPE, RMSE e R^2 indicam que a configuração da ANFIS com o algoritmo de fuzzificação FCM proporciona o melhor desempenho para o método proposto.

6.13 SISTEMA FIS

Depois de configurada a ANFIS, o sistema FIS ficou treinado para um algoritmo de fuzzificação FCM com função de pertinência gaussiana com três variáveis linguísticas. Desta forma, o FIS está pronto para receber os sinais determinados para teste.

Na Tabela 11 constam os resultados da PESQ, os preditos pela ANFIS/FCM e a diferença - DF - em módulo entre as duas, tomando como exemplo o ruído de multidão. Observa-se em destaque que o maior erro encontrado foi 0,4276, que é considerado aceitável por se tratar de avaliação de sinais de voz (HU; LOIZOU, 2007). Como a nota PESQ implementada varia de -0,5 a 4,5 existe um intervalo de 5,0 pontos a serem analisados neste processo. Todos os valores preditos apontam para uma boa regressão.

Tabela 10: Tempo de execução do programa na fase de teste para 60 sinais dos diretórios carro e multidão da base de dados NOIZEUS.

Ruído		Carro		Multidão		
MF	FCM	<i>Subtractive</i>	<i>Grid</i>	FCM	<i>Subtractive</i>	<i>Grid</i>
Tempo(s)	0,0089	0,0102	0,0183	0,0142	0,0159	0,0265

Fonte: Elaborada pelo próprio autor.

Os resultados indicam que é possível realizar uma boa regressão para as notas PESQ, com baixo tempo computacional, como pode ser observado na Tabela 10. Importante destacar que a divisão aleatória, destacada na Figura 31, realizada para definir os pares de dados de treinamento e teste, proporcionou resultados satisfatórios para o método proposto.

Tabela 11: Diferença entre os resultados da PESQ e ANFIS/FCM na fase de teste para 60 sinais com ruído de multidão.

Valores	Sinais (1 até 60)									
PESQ	2,6900	2,3700	2,1700	2,8800	4,5000	1,4400	2,4800	2,5100	2,7200	1,9000
ANFIS/FCM	2,5895	2,3251	2,2847	3,0633	4,5000	1,6945	2,5237	2,2842	2,2924	1,7082
DF	0,1005	0,0449	0,1147	0,1833	0,0000	0,2545	0,0437	0,2258	0,4276	0,1918
PESQ	1,6200	2,1100	2,1700	2,6800	4,5000	4,5000	2,5400	2,3700	2,1400	2,1200
ANFIS/FCM	1,6363	1,9951	2,3859	2,7181	4,4410	4,5000	2,5575	2,2183	2,0314	2,0715
DF	0,0163	0,1149	0,2159	0,0381	0,0590	0,0000	0,0175	0,1517	0,1086	0,0485
PESQ	1,4900	1,5800	4,5000	1,9600	2,6700	2,1000	2,0000	2,1400	4,5000	2,3700
ANFIS/FCM	1,5963	1,7386	4,2058	1,5961	2,9384	2,2399	2,0032	2,2570	4,4357	2,3483
DF	0,1063	0,1586	0,2942	0,3639	0,2684	0,1399	0,0032	0,1170	0,0643	0,0217
PESQ	1,8700	1,9800	2,0700	4,5000	2,6100	2,7200	1,6800	2,4100	1,9500	1,7300
ANFIS/FCM	2,0327	2,0976	2,0658	4,5000	2,4586	2,5522	1,9548	2,5784	2,1663	1,9895
DF	0,1627	0,1176	0,0042	0,0000	0,1514	0,1678	0,2748	0,1684	0,2163	0,2595
PESQ	4,5000	4,5000	1,9900	2,0100	1,8100	4,5000	1,8800	1,7600	4,5000	2,2400
ANFIS/FCM	4,3729	4,5000	1,9841	2,0572	1,7082	4,5000	1,7865	1,7473	4,5000	1,9743
DF	0,1271	0,0000	0,0059	0,0472	0,1018	0,0000	0,0935	0,0127	0,0000	0,2657
PESQ	2,7400	4,5000	2,6900	1,8000	2,1100	2,9900	1,4400	1,5400	1,5700	2,6700
ANFIS/FCM	2,5790	4,5000	2,6661	1,6705	2,1048	2,7784	1,6686	1,7050	1,7975	2,6122
DF	0,1610	0,0000	0,0239	0,1295	0,0052	0,2116	0,2286	0,1650	0,2275	0,0578

Fonte: Elaborada pelo próprio autor.

ANÁLISES E DISCUSSÕES

Este Capítulo tem como objetivo apresentar as análises dos resultados da implementação computacional do método proposto. Nas próximas seções serão discutidas aspectos relacionados a escolha dos melhores fatores de Haralick, a comprovação do método, comparações de desempenho entre a ANFIS/FCM com as redes MLP e entre os modelos perceptivos de Hermansky e MFCC.

7.1 ESCOLHA DOS FATORES DE HARALICK

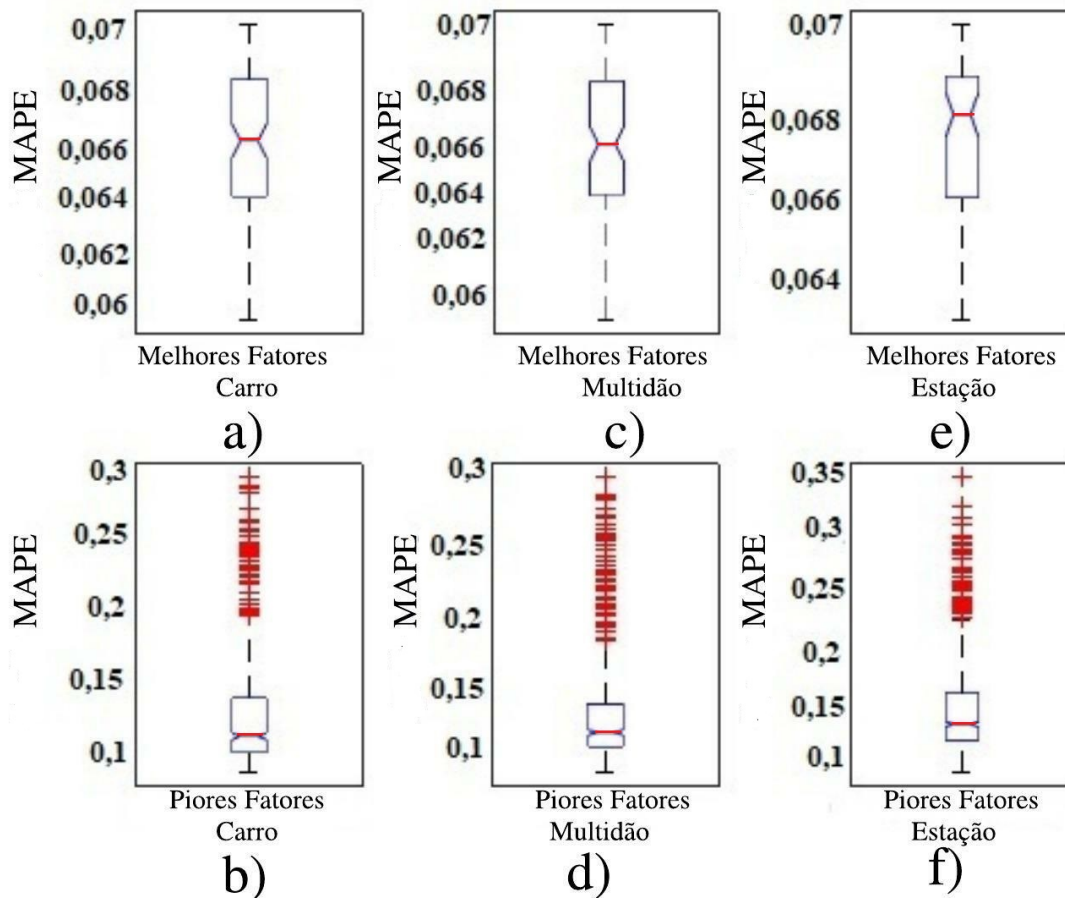
Para um bom desempenho do modelo, a escolha de um conjunto de fatores de entrada faz toda diferença (BENMOUIZA; CHEKNANE, 2019; SITOMPUL; NABABAN; ALIM, 2017). Com relação a quantidade de fatores utilizados na fase de fuzzificação da ANFIS, chegou-se à conclusão, por meio de análises, que três fatores conseguem configurar o modelo. Visto que com menos fatores a rede ANFIS reconhece poucos padrões impactando nos resultados da regressão. Mas, com mais fatores acarretará um esforço maior na fase de fuzzificação, pois a configuração do modelo depende dos ajustes dos parâmetros para compor as suas respectivas variáveis linguísticas envolvidas no processo.

Para eleger os três melhores fatores primeiramente foi realizado um arranjo para obter um conjunto contendo três elementos distintos dos 19 fatores de Haralick. Esta operação gerou 969 composições diferentes, como pode ser visto no Apêndice B. Foram realizadas diversas simulações com o modelo ANFIS/FCM utilizando a base de dados NOIZEUS com intuito de identificar os três melhores fatores que apresentassem uma estimativa confiável para previsão das notas PESQ.

Com base nos resultados obtidos da Tabela 8 com ANFIS/FCM, comprovou-se por meio das medidas, que o modelo preserva seu desempenho para valores representados pelas medidas

$MAPE \leq 0,009$, $RMSE \leq 0,22$ e $R^2 \geq 0,95$ sendo considerados limiares para uma boa predição dos valores PESQ do sinal de voz.

Figura 36: *Boxplots* dos resultados das combinações dos fatores de Haralick para áudios com ruídos de carro, multidão e estação sendo: a), c) e e) dos melhores MAPE e b), d) e f) piores MAPE.



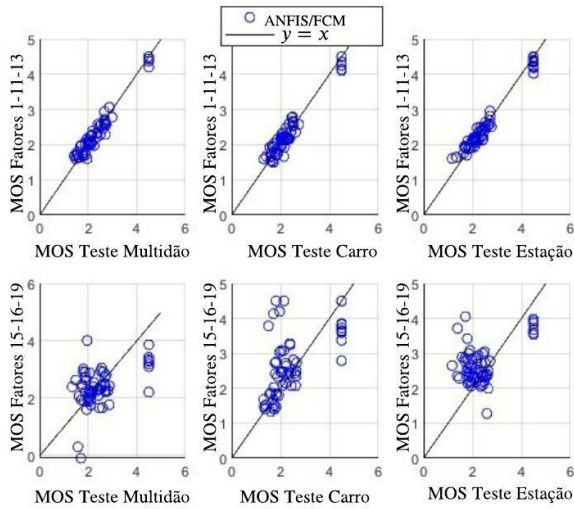
Fonte: Elaborada pelo próprio autor.

A partir das análises realizadas, verificou-se a existência de duas classes de fatores de Haralick, que ficaram definidas pela média dos MAPE entre as fases de treinamento e teste. A primeira contendo as melhores notas das avaliações dos sinais para $MAPE \leq 0,09$, cujos resultados estão destacados no Apêndice B, e a segunda para resultados superiores. Nas Figuras 36a, c, e ; b, d, f são apresentados os diagramas de caixas (do inglês: *boxplots*) para as melhores e piores combinações de fatores para os ruídos de carro, multidão e estação. Em

tais figuras, nota-se que o local onde a haste vertical começa indica o mínimo, executando algum possível valor discrepante (do inglês: *outlier*) e onde a haste termina indica o máximo, também executando algum possível *outlier*. Na caixa retangular principal, existem três linhas importantes: o contorno extremo inferior que indica o primeiro quartil, a linha vermelha interna é o segundo quartil ou mediana e o contorno extremo superior, o terceiro quartil. Comparando os *boxplots* dos diretórios de ruídos, existem conjuntos de fatores que estão presentes em todos e apresentam bons resultados. Nos *boxplots* para os piores MAPE, existem conjuntos de fatores que contribuem para estes baixos valores, visto ainda a presença de *outliers* que indicam a existência de fatores, e se considerados como entradas para processo de fuzzificação não contribuem para um bom desempenho do modelo. Na Figura 37 são apresentados os comportamentos das regressões para os diretórios com ruídos de multidão, carro e estação na fase de teste da ANFIS/FCM. Observa-se que para os fatores 119/1-11-13 o modelo apresentou uma boa regressão, enquanto que a escolha do conjunto de fatores 693/15-16-19 não proporciona uma boa estimativa para a PESQ. Isso pode ser confirmado nos resultados apresentados na Tabela 12.

Cabe ressaltar que é possível deixar o modelo encontrar seu melhor desempenho, percorrendo todas as combinações de fatores, pois cada ruído apresenta características próprias. Porém, devido a quantidade de combinações a determinar, conjuntos de fatores favoráveis a todos os ruídos podem reduzir os custos computacionais. As características a favor dos fatores de Haralick que proporcionam os melhores desempenhos estão ligadas à natureza que eles são constituídos. Variações mínimas ocorridas entre fatores extraídos de sinais caracterizam representações diferentes. Neste contexto, o sistema de inferência *fuzzy* é adequado para este tipo de aplicação.

Figura 37: Regressão dos fatores

Tabela 12: Medidas MAPE, RMSE e R^2 dos fatores.

Ruído	Teste [1-11-13]		
	MAPE	RMSE	R^2
Multidão	0,0700	0,1877	0,9613
Carro	0,0683	0,1748	0,9665
Estação	0,0724	0,1975	0,9572
	Teste [15-16-19]		
	MAPE	RMSE	R^2
Multidão	0,2535	0,5852	0,2409
Carro	0,2660	0,7284	0,4405
Estação	0,2557	0,5811	0,1482

Fonte: Elaborada pelo próprio autor.

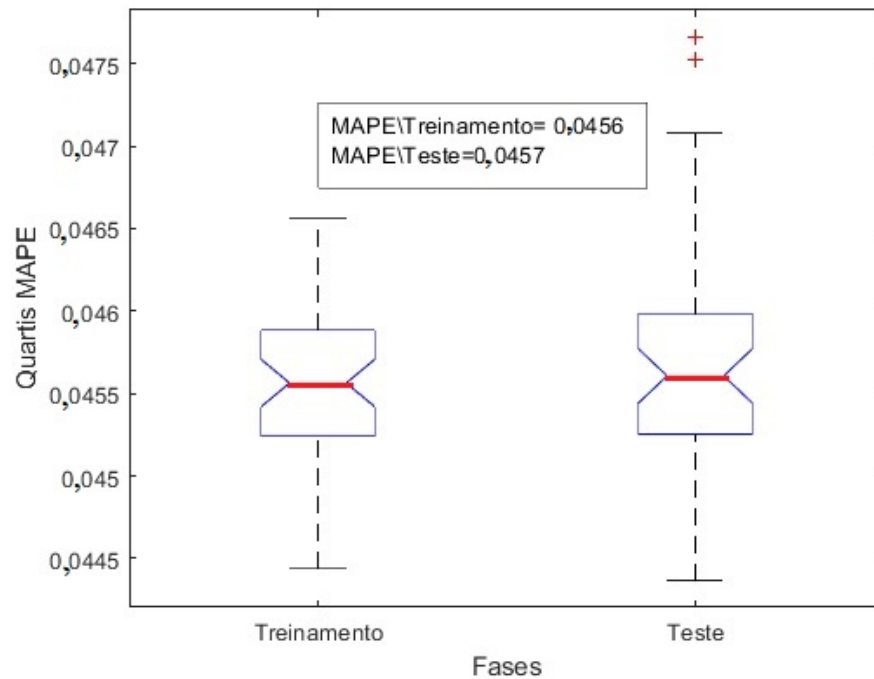
7.2 COMPROVAÇÃO DO MÉTODO

Essa fase tem como objetivo validar o modelo com um conjunto maior de sinais, utilizando uma base de dados diferentes. Primeiramente, a rede ANFIS/FCM na direção 90° com três entradas, para o conjunto de fatores de Haralick 119/1-11-13 de maior desempenho apresentados no Apêndice A e três variáveis linguísticas, foi aplicada para um conjunto de sinais denominado TIMIT/NOISEX, contendo 2500 sinais com ruído de carro. Utilizando os mesmos critérios adotados na divisão do conjunto de dados de treinamento e teste, sendo os sinais divididos aleatoriamente com 60% e 40% destinados de treinamento e teste, respectivamente. Este processo se repetiu 50 vezes, os resultados para medida MAPE são descritos no *boxplot* da Figura 38.

Nota-se na Figura 38, que o modelo proposto possui bons resultados, mesmo aplicando um número considerável de sinais. Observa-se ainda que a linha vermelha que indica a mediana está em conformidade com o MAPE para ambas as fases. No *boxplot* de treinamento, a dispersão dos dados é menor que na fase de teste, não possuindo valores discrepantes, *outliers*. Isto é um fator positivo para a avaliação do modelo, pois tanto nas simulações com a base de

dados NOIZEUS quanto para a simulação com o banco de dados da TIMIT/NOISEX com ruído de carro, os resultados de MAPE para fase de treinamento supera a fase de teste. O fato, é que na fase de treinamento, a rede treina e avalia com os próprios sinais, enquanto que na fase de teste, sinais diferentes são apresentados ao modelo.

Figura 38: Resultados dos cinquenta valores MAPE representados pelos *boxplots* obtidos pela ANFIS/FCM para a base de dados TIMIT/NOISEX com ruído de carro.



Fonte: Elaborada pelo próprio autor.

7.3 APLICAÇÕES UTILIZANDO ANFIS E REDES MLP

O método proposto com a ANFIS/FCM tomando a direção 90° da matriz de coocorrência com fatores Haralick 1, 11 e 13, teve o seu desempenho comprovado em comparação com *Multilayer Perceptron Networks* - MLP com os seus algoritmos de treinamento Levenberg-Marquardt - FLM, Bayesian Regularization - BR, Broyden-Fletcher-Goldfarb-Shanno Quasi-Newton - BFGS, Resilient Backpropagation - FRB, Scaled Conjugate Gradient - SCG, Conjugate Gradient with Powell/Beale Restarts - CGB, Fletcher-Powell Conjugate Gradient - CGF, Polak-Ribière Conjugate Gradient - CGP, One Step Secant - OSS, Variable Learning Rate Gradient Descent - GDX, Gradient Descent with Momentum - GDM e Gradient Descent

- FGD (BISHOP, 2005; LIVINGSTONE, 2008). Para cada uma das redes MLP, foram utilizadas três camadas ocultas com nove neurônios e uma camada de saída, todas com função de ativação sigmoidal. Esta configuração segue a mesma utilizada por Sahin e Erol (2017) que demonstraram bons resultados para o modelo de regressão que desenvolveram. Os resultados do desempenho do modelo ANFIS/FCM em comparação com as redes MLP para um diretório de sinais de voz específico estão destacados na Tabela 13.

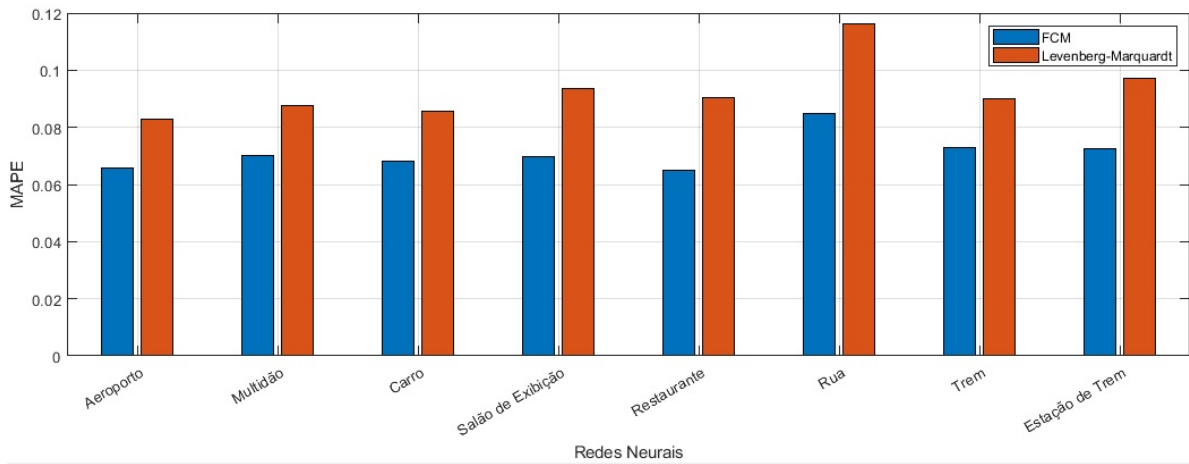
Tabela 13: Resultados do desempenho entre a ANFIS e as redes MLP para o diretório com ruído de multidão da base de dados NOIZEUS.

Redes	Md/DP	Treinamento			Teste		
		MAPE	RMSE	R ²	MAPE	RMSE	R ²
ANFIS/ FCM	Md	0,0587	0,1601	0,9712	0,0683	0,1748	0,9613
	DP	0,0032	0,0085	0,0031	0,0056	0,0158	0,0073
LM	Md	0,0579	0,1778	0,9620	0,0850	0,2464	0,9294
	DP	0,0155	0,0454	0,0269	0,0170	0,0594	0,0425
BR	Md	0,0687	0,1963	0,9522	0,0831	0,2452	0,9283
	DP	0,0186	0,0562	0,0371	0,0176	0,0621	0,0424
BFGS	Md	0,0793	0,2304	0,9328	0,0915	0,2745	0,8946
	DP	0,0245	0,0825	0,0545	0,0281	0,0996	0,1010
RB	Md	0,0782	0,2232	0,9344	0,0911	0,2637	0,9081
	DP	0,0259	0,0738	0,0528	0,0247	0,0709	0,0645
SCG	Md	0,0866	0,2600	0,9154	0,1032	0,3065	0,8823
	DP	0,0248	0,0811	0,0521	0,0276	0,0895	0,0723
CGB	Md	0,0846	0,2413	0,9259	0,0943	0,2670	0,9080
	DP	0,0200	0,0677	0,0451	0,0231	0,0678	0,0548
FPCG	Md	0,0710	0,2092	0,9458	0,0903	0,2769	0,9068
	DP	0,0200	0,0649	0,0411	0,0190	0,0675	0,0532
OSS	Md	0,0904	0,2609	0,9072	0,0995	0,2910	0,8877
	DP	0,0275	0,0797	0,0679	0,0265	0,0859	0,0797
VLRGD	Md	0,1422	0,3683	0,7435	0,1616	0,3937	0,7009
	DP	0,0653	0,1483	0,2240	0,0727	0,1429	0,2288
GDWM	Md	0,3326	0,5549	0,2361	0,3408	0,5721	0,2251
	DP	0,0988	0,2179	0,2353	0,1070	0,2152	0,2150
GD	Md	0,0827	0,2349	0,9168	0,0953	0,2657	0,8965
	DP	0,0279	0,0695	0,0923	0,0310	0,0698	0,0961

Fonte: Elaborada pelo próprio autor.

Na fase de treinamento, o modelo ANFIS/FCM comporta-se de forma semelhante mas com menor desvio padrão entre as medidas MAPE, RMSE e R² em comparação com as redes LM e BR cujos seus resultados são apresentados na Tabela 13.

Figura 39: Resultados do MAPE para ANFIS/FCM e a rede MLP com algoritmo Levenberg-Marquardt para os 8 sinais da base de dados NOIZEUS.



Fonte: Elaborada pelo próprio autor.

Na Figura 39 são apresentados os resultados da medida MAPE na fase de teste para os oito diretórios de sinais. Nela está representada a rede MLP com o algoritmo de aprendizagem Levenberg-Marquardt que apresentou o melhor desempenho entre os outros e também os resultados obtidos para ANFIS/FCM. Nota-se que o método com a ANFIS/FCM tem desempenho superior para todos os conjuntos de sinais em comparação com o MLP/Levenberg-Marquardt (BILSKI *et al.*, 2020).

Um fato importante a observar nesta comparação é a estrutura dos conjuntos de fatores dos elementos de entrada. Cada sinal de voz limpo com seus respectivos ruídos e SNR produzem fatores de Haralick diferentes com variações mínimas e neste contexto de avaliação de sinais de voz a rede que conseguiu captar este nível de variação foi a ANFIS mesmo em diretórios com número reduzidos de sinais.

7.4 DESEMPENHO DO MODELO HERMANSKY E MFCC

O modelo proposto também teve seu desempenho comprovado aplicando os modelos perceptivos de Hermansky e MFCC, seguindo as mesmas condições. Na Tabela 14 são mostrados os resultados da média e o desvio padrão das medidas obtidas dos cinquenta pares aleatórios de treinamento e teste para a ANFIS/FCM. Verificou-se para todos os diretórios de sinais da base de dados NOIZEUS que os resultados obtidos para o método proposto utilizando

o modelo perceptual de Hermansky apresentou os melhores resultados em comparação ao MFCC. O que pode ser levado em conta no desempenho do método com o modelo perceptivo de Hermansky é que na etapa de extração de fatores as matrizes geradas das imagens, como pode ser visto na Figura 26a e b, possuem dados mais contundentes favorecendo a distinção de um sinal para outro, o mesmo não ocorre para o modelo MFCC, Figura 26c e d, pois suas matrizes correspondentes possuem limitações para distinguir os fatores dos sinais.

Um fato ainda em desfavor do MFCC é que a medida MAPE na fase de teste supera o valor 0,09, limite estabelecido para uma boa previsão dos valores PESQ, ou seja, comprometendo sua regressão.

Tabela 14: Resultados do desempenho da ANFIS/FCM para os modelos perceptivos de Hermansky e MFCC das fases de treinamento e teste para os oito diretórios de sinais da base de dados NOIZEUS na direção $[-1\ 0]$ e/ou 90° da matriz de coocorrência.

Ruídos	MP	Md/DP	Treinamento			Teste		
			MAPE	RMSE	R ²	MAPE	RMSE	R ²
Aeroporto	Hermansky	Md	0,0574	0,1535	0,9748	0,0658	0,1751	0,9672
		DP	0,0039	0,0090	0,0037	0,0069	0,0215	0,0112
	MFCC	Md	0,1121	0,3078	0,8741	0,1335	0,3687	0,8155
		DP	0,0068	0,0205	0,0200	0,0153	0,0437	0,0510
Carro	Hermansky	Md	0,0587	0,1512	0,9771	0,0683	0,1748	0,9665
		DP	0,0040	0,0083	0,0030	0,0063	0,0148	0,0078
	MFCC	Md	0,1129	0,3478	0,8468	0,1310	0,3978	0,7993
		DP	0,0098	0,0289	0,0299	0,0186	0,0629	0,0633
Estação	Hermansky	Md	0,0606	0,1660	0,9707	0,0724	0,1976	0,9572
		DP	0,0044	0,0119	0,0046	0,0067	0,0191	0,0098
	MFCC	Md	0,1246	0,3473	0,8428	0,1466	0,4009	0,7767
		DP	0,0089	0,0227	0,0235	0,0176	0,0541	0,0690
Multidão	Hermansky	Md	0,0587	0,1601	0,9712	0,0700	0,1877	0,9613
		DP	0,0032	0,0085	0,0031	0,0056	0,0158	0,0073
	MFCC	Md	0,1280	0,3834	0,8114	0,1479	0,4354	0,7526
		DP	0,0103	0,0271	0,0289	0,0149	0,0481	0,0531
Restaurante	Hermansky	Md	0,0576	0,1708	0,9666	0,0651	0,1867	0,9590
		DP	0,0045	0,0144	0,0063	0,0077	0,0253	0,0134
	MFCC	Md	0,1105	0,3092	0,8765	0,1281	0,3624	0,8255
		DP	0,0080	0,0214	0,0173	0,0149	0,0437	0,0501
Rua	Hermansky	Md	0,0737	0,1856	0,9642	0,0849	0,2102	0,9547
		DP	0,0076	0,0141	0,0048	0,0110	0,0161	0,0082
	MFCC	Md	0,1222	0,3535	0,8294	0,1346	0,3761	0,7963
		DP	0,0101	0,0244	0,0300	0,0182	0,0439	0,0550
S. Exibição	Hermansky	Md	0,0576	0,1708	0,9666	0,0651	0,1867	0,9590
		DP	0,0045	0,0144	0,0063	0,0077	0,0253	0,0134
	MFCC	Md	0,1424	0,4078	0,7852	0,1594	0,4507	0,7382
		DP	0,0134	0,0300	0,0327	0,0198	0,0609	0,0689
Trem	Hermansky	Md	0,0640	0,1671	0,9727	0,0731	0,1882	0,9647
		DP	0,0052	0,0090	0,0043	0,0099	0,0196	0,0109
	MFCC	Md	0,1399	0,4281	0,7590	0,1634	0,4801	0,6911
		DP	0,0083	0,0216	0,0303	0,0153	0,0689	0,0528

Fonte: Elaborada pelo próprio autor.

8 CONCLUSÕES

Neste trabalho, foi proposto um método de previsão baseadas nos fatores da imagem do sinal de voz que teve como objetivo principal estimar a MOS-PESQ por meio de um modelo de regressão aplicado em um sistema *fuzzy* do tipo Sugeno de primeira ordem implementado pela ANFIS. A construção do modelo preditivo foi elaborada a partir de uma abordagem metodológica em três importantes etapas. A primeira constitui na escolha de um modelo perceptivo capaz de gerar as imagens dos sinais de voz, a segunda procura extrair características da matriz perceptiva e a terceira define a melhor configuração da ANFIS tomando como base as medidas MAPE, RMSE e R^2 .

De um modo geral, a partir das simulações realizadas, constatou-se que o modelo apresentou os melhores resultados quando as imagens foram geradas pelo modelo perceptivo de Hermansky sob a análise da combinação de três fatores de Haralick na direção de 90° para a técnica da matriz de coocorrência em níveis de cinza. Com os fatores determinados e os seus respectivos sinais avaliados pela medida intrusiva PESQ, os dados de treinamento e teste foram separados de forma aleatória com o intuito de mesclar e obter composições diferentes na configuração da ANFIS. Na escolha de um método de fuzzificação para ANFIS, o modelo de agrupamento ANFIS com FCM proporcionou os melhores resultados para três entradas e cada uma delas com três variáveis linguísticas. Nas simulações realizadas o modelo ANFIS/FCM conseguiu manter seu desempenho para um conjunto de dados com um grande número de sinais. Comparações foram conduzidas e o modelo de regressão apresentou resultados satisfatórios perante as redes neurais MPL com seus algoritmos de treinamento amplamente difundidos na literatura especializada.

As principais características da proposta PESQ-ANFIS/*Fuzzy C-MEANS* são: avaliação do sinal de voz por meio de um modelo perceptivo, extração de fatores por meio de técnica de processamento de imagem, atuação efetiva em qualquer ambiente ruidoso proporcionando

resultados satisfatórios independentemente da quantidade de sinais a serem analisados.

As perspectivas de pesquisas futuras para o desenvolvimento deste trabalho são: busca de bancos de dados publicados que forneçam avaliações subjetivas do sinal, implementações de outros modelos perceptivos da audição humana, extração de fatores do sinal por meio de outras técnicas de processamento de imagem, implementação da rede ANFIS utilizando outros fuzzificadores e aplicações e simulações com novas redes neurais ligadas ao modelo de regressão.

REFERÊNCIAS

ALCALÁ, R.; CASILLAS, O.; HERRERA, F.; ZWIR, S. J. I. Techniques for learning and tuning fuzzy rule-based systems for linguistic modeling and their application. *Academic Press*, Granada, p. 889-941, 1999.

ALSHAIB, W. B.; AL-KANDARI, M. J.; HASAN, S. M. Classification of Hearing Loss. *INTECH*, London, v.23, p. 29-37, 2015.

ARAS, V. P. Audiometry techniques, circuits, and systems. *Electronic Systems Group*, Alabaster p. 1-17, 2008.

ARMİ, L.; ERSHAD, S. F. Texture image analysis and texture classification methods - a review. *International Online Journal of Image Processing and Pattern Recognition*, India, v. 2, n.1, p. 1-29, 2019.

BEERENDS, J. G.; HEKSTRA, A. P.; RIX, A.W. PESQ, the new ITU standard for objective measurement of perceived speech quality, part II Perceptual model, *J.AudioEng. Soc.*, Geneva, v. 50, p. 765-778, 2002.

BEERENDS, J. G.; NEUMANN, N. M. P.; VAN DEN BROEK, E. L.; CASANOVAS, A. L.; MENENDEZ, J. T.; SCHMIDMER, C.; BERGER, J. Subjective and Objective Assessment of Full Bandwidth Speech Quality. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 28, p. 440-447, 2020.

BENMOUIZA, K.; CHEKNANE, A. Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. *Theoretical and Applied Climatology*, Berlin, v. 137, p. 31-47, 2019.

BEZDEK, J. Cluster validity with fuzzy sets. *J. Cybern.*, [s.l.] v.3, p.58-73, 1973.

BEZDEK, J.C. *Pattern Recognition with fuzzy Objective Function Algorithms*. Plenum Press: New York, 1981. 256 p.

BILSKI, J.; KOWALCZYK, B.; MARCHLEWSKA, A.; ZURADA, J. M. Local levenberg-marquardt algorithm for learning feedforward neural networks. *Sciend*, Warsaw, v.10, n.4.

p. 299-316, 2020.

BISHOP, M. C. *Neural Networks for Pattern Recognition*. USA: Oxford Press, 2005. p 498.

CHAKI, C.; DEY, N. *Texture feature extraction techniques for image recognition*. India: Springer, 2020. p 100.

CHEN, G.; PARSA, V. Objective speech quality evaluation using an adaptive neuro-fuzzy network. *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*, Tallahassee, Springer, p. 97-116, 2008.

CHEN, M.S.; YING, L. C.; PAN, M. C. Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system. *Expert Systems with Applications*, Amsterdam, v. 37, p. 1185–1191, 2010.

CHIU, S. fuzzy model identification based on cluster estimation. *Journal of Intelligent and fuzzy Systems*, Amsterdam, v.2, p. 267–278, 1994.

CLAUSI, D. A. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, New York, v.28, p. 45–62, 2002.

CONN, P. M. *Conn's translational neuroscience*. Amsterdam: Elsevier, 2017. p 776.

DAVIES, R. A. Audiometry and other hearing tests. *HandBook of Clinical Neurology*, Amsterdam, v. 137, n. 3, p. 1-20, 2016.

DEMÓSTENES, R. Z.; ROSA, R. L.; ALMEIDA, F. L; MITTAG, G.; MOLLER, S. S. Speech Quality Assessment in Wireless Communications With MIMO Systems Using a Parametric Model, *IEEE Access*, Piscataway, v.7, p. 35719-35730, 2019.

DONG, X.; WILLIAMSON, D. S. A CLASSIFICATION-AIDED FRAMEWORK FOR NON-INTRUSIVE SPEECH QUALITY. *In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2019, New York , USA. *Proceedings...* New York , USA: EUSFLAT, 2019.

ELKHODR, M.; QUSAY, F. HASSAN, Q. F. *Networks of the future: Architectures, technologies and implementations*. London: CRC Press, 2018. p 513.

FAN. G. Analysis and synthesis of speech processes. *Manual of Phonetics*, Amsterdam, p. 173–177, 1968.

FANIZZI, A.; LOSURDO, L.; BASILE, T. M. A.; BELLOTTI, R.; BOTTIGLI, U.; DE-

LOGU, P.; DIACONO, D.; DIDONNA, V.; FAUSTO, A.; LOMBARDI, A.; LORUSSO, V.; MASSAFRA, R.; TANGARO, S.; FORGIA, D. L. Fully Automated Support System for Diagnosis of Breast Cancer in Contrast-Enhanced Spectral Mammography Images. *Journal of Clinical Medicine*, Basel, v.8, p. 1-19, 2019.

FASTL, H.; ZWICKER, E. *Psychoacoustics*. Berlin: Springer, 2007. p 463.

FSTEDT, T. L.; BRYNOLFSSON, P.; ASKLUND, T.; NYHOLM, T.; GARPEBRING, A. Gray-level Invariant Haralick Texture Features. *Plos One*, San Francisco, v.22, p. 1-18, 2019.

GAMPER, H.; REDDY, C. K. A.; CUTLER, R.; TASHEV, I. J.; GEHRKE, J. Intrusive and non-intrusive perceptual speech quality assessment using a convolutional neural network. In: IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS - 2019. New Paltz. *Proceedings...* New Paltz: IEEE, 2019. p. 85-89.

GAROFOLO, J.; LAMEL, L. F.; FISHER, W. M.; FISCUS, J. G.; PALLETT, D. S.; DAHLGREN, N. L.; ZUE, V. TIMIT Acoustic-phonetic continuous speech corpus. *Linguistic Data Consortium*, Philadelphia, 1993. Disponível em: <https://catalog.ldc.upenn.edu/LDC93S1>. Acesso em: 14 jul. 2021.

GELFAND, S. A. *Hearing: an introduction to psychological and physiological acoustics*. New York: Informa, 2010. 322 p.

GOMES, J. C.; BARBOSA, V. A. F.; SANTANA, M. A.; BANDEIRA, J.; VALENÇA, M. J. S.; SOUZA, R.; MASOOD, A. I.; SANTOS, W. P. IKONOS: an intelligent tool to support diagnosis of COVID-19 by texture analysis of X-ray images. *Biomedical Engineering*, Berlin, p. 1-14, 2020.

GOUVEIA, D. J. S.; ROSA, R. L.; RODRÍGUEZ, D. Z. Avaliação da qualidade da voz em serviços de comunicação usando deep learning. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 36. , 2018, Campos do Jordão. *Anais [...]* Porto Alegre: Sociedade Brasileira de Computação, 2018 . Disponível em: <https://sol.sbc.org.br/index.php/sbrc/article/view/2415>. Acesso em: 18 jul. 2021.

HARALICK, R. M. Statistical and structural approaches to texture. *Proceedings of IEEE*, Piscataway, v. 67, p. 786-804, 1979.

HARALICK, R. M.; SHANMUNGAM, K.; DINSTEN, I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, Piscataway, v. 3, n. 6, p. 610-620, 1973.

HERMANSKY, H. Perceptual linear predictive (PLP) analysis of speech. *Acoustical Society of America*, [s. l.], n. 4, v. 87, p. 1738-1752, 1990.

HERRE, J.; DICK, S. Psychoacoustic models for perceptual audio coding — A tutorial review. *Journal Applied Sciences*, Basel, v.9, p. 1-22, 2019.

HILL, P. R. *Audio and Speech Processing with MATLAB*. London: CR PRESS, 2019. 354 p.

HINTERLEITNER, F. *Quality of synthetic speech: Perceptual dimensions influencing factors, and Instrumental assessment*, Springer: Berlin, 2017. 157 p.

HU, Y.; LOIZOU, P. C. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, Amsterdam, v. 49, n. 7, p. 588–601, 2007.

HUA, C. Y.; CHENA, S. R.; TZENGB, G. H. Discovering fuzzy association rules using fuzzy partition methods. *Knowledge-Based Systems*, Amsterdam, v.16, n.3 p. 137–147, 2003.

ITU-T Recommendation P.862. *Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*. International Telecommunication Union, Genebra, Suíca, 2005. Disponível em: <https://www.itu.int/rec/T-REC-P.862>. Acesso em: 15 nov. 2022.

ITU-T Recommendation P.863. *Perceptual objective listening quality Analysis (POLQA)*. International Telecommunication Union, Genebra, Suíca, 2011. Disponível em: <https://www.itu.int/rec/T-REC-P.863-201803-I/en>. Acesso em: 20 nov. 2022.

Recommendation ITU-T P.861. *Perceptual Speech Quality Measure (PSQM)*. International Telecommunication Union, Genebra, Suíca, 1998. Disponível em: https://www.itu.int/rec/T-REC-P.861/_page.print . Acesso em: 15 nov. 2022.

ITU-T Coded-Speech Database, ITU-T Suppl. 23 to P.863, International Telecommunication Union, Geneva, Switzerland, 1998. Disponível em: <https://www.itu.int/rec/T-REC-P.Supp23-199802-I>. Acesso em: 20 jun. 2021.

ITU–T. Recommendation P.563. *Single ended method for objective speech quality assessment in narrowband telephony applications*. International Telecommunication Union, Genebra, Suíca, 2004. Disponível em: <https://www.itu.int/rec/T-REC-P.563/en>. Acesso em: 10 jul. 2021.

ITU–T. Recommendation P.800. *Methods for Subjective Determination of Transmission Quality*. International Telecommunication Union, Genebra, Suíca, 1996. Disponível em: <https://www.itu.int/rec/T-REC-P.800/en>. Acesso em: 10 jul. 2021.

ITU-T Recommendation G.107. *Computational Model for use in Transmission Planning (E-model)*. International Telecommunication Union, Genebra, Suíca, 2014. Disponível em :

<https://www.itu.int/rec/TREC-G.107>. Acesso em: 10 jul. 2021.

ITU-T Recommendation G.711: *Pulse code modulation (PCM) of voice frequencies*. International Telecommunication Union, Genebra, Suíça, 1999. <https://www.itu.int/rec/T-REC-G.711/>. Acesso em: 20 jul. 2022.

JAFARI, M. M.; OJAGHLOU, H.; ZARE, M.; SHUMANN, G. J. P. Application of a novel hybrid Wavelet-ANFIS/fuzzy C-Means clustering model to predict groundwater fluctuations. *Journal Atmosphere*, Basel, n. 9, v. 12, p. 1-15, 2021.

JANG, J. S. R.; SUN, C. T.; MIZUTANI, E. *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence*. London: Prentice-Hall, 1997. 640 p.

JANG, J.S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, Piscataway, v.23, p. 665–685, 1993.

KEENER, J.; SNEYD, J. *Mathematical physiology*. New York: Springer, 2009. p 574.

KLEMMT, H. J.; SEITZ, R.; STRAUB, C. Application of Haralick's texture features for rapid detection of windthrow hotspots in orthophotos. *Journal Florests*, Basel, v.11, p. 2-17, 2020.

LEVENBERG, K. A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, v.5, p. 164–168, 1944.

LIU, M.; WANG, J.; YI, W.; LIU, F. Neural network-based non-intrusive speech quality assessment using attention pooling function. *Journal on Audio, Speech, and Music Processing*, Berlin, v.20, p. 1-10, 2021.

LIVINGSTONE, J. D. *Artificial Neural Networks: Methods and Applications*. Hatfield: Humana Press, 2008. 258 p.

MARQUARDT, D. An algorithm for least-squares estimation of nonlinear parameters, SIAM. *Journal on Applied Mathematics*, v.11, p. 431–441, 1963.

MA, J.; HU, Y.; LOIZOU, P. C. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *The Journal of the Acoustical Society of America*, Melville, v. 125, n. 5, p. 3387–3405, 2009.

MADISETTI, V. K. *The Digital Signal Processing Handbook*. CRC Press: California, 2010. 2400 p.

MATULIN, M.; MRVELJ, S. Modelling user quality of experience from objective and subjective data sets using fuzzy logic. *Multimedia Systems*, Berlin, n.6, v.24, p. 645–667, 2018.

MEBAWAONDU, J. O.; DAHUNSI, F. M.; ADEWALE, O. S. Hybrid intelligent model for real time assessment of voice quality of service. *Scientific African*, Amsterdam, v.9, p.1-16, 2020.

MCLOUGHLIN, I. V. *Speech and audio processing*. Cambridge: Cambridge University Press, 2014. 403 p.

NGUYEN, Q. H.; LY, H.; HO, L. S.; AL-ANSARI, N.; LE, H. V.; TRAN, V. Q.; PRAKASH, I.; PHAM, B, T. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering* , London, v.2021, p. 1-15, 2021.

NECCIARI, T.; LABACK, B.; SAVEL, S.; YSTAD, S.; BALAZS, P.; MEUNIER, S.; KRONLAND-MARTINET, R. Auditory time-frequency masking for spectrally and temporally maximally - compact stimuli. *Journal Plos One* , Austrian, v.11, p. 1-23, 2016.

OPPENHEIM, A. V.; SCHAFER, R. W.; BUCK, J. R. *Discrete-time signal processing*. New Jersey: Prentice Hall, 1998. 1120 p.

O'SHAUGHNESSY, D. *Speech Communication: Human and Machine*. Addison-Wesley, Piscataway, 2000. 548 p.

PFAFF, D. W.; VOLKOW, N. D. *Neuroscience in the 21st Century*. New York: Springer, 2016. 4155 p.

PODDER, P.; KHAN, T. Z.; KHAN, M. H.; RAHMAN, M . M. Comparative Performance Analysis of Hamming, Hanning and Blackman Window. *International Journal of Computer Applications*, Karlsruhe, n.18, v.96, p. 0975–8887, 2014.

RABINER, L. R.; SCHAFER, R. W. *Introduction to digital speech processing*. Boston: Editorial Board, 2007. 2013 p.

RAHDARI, F.; EFTEKHARI, M. Modeling the perceived voice quality for VOIP system based on neuro-fuzzy: A comparative study. *In: IEE INTERNATIONAL ECONFERENCE ON COMPUTER AND KNOWLEDGE ENGINEERING (ICCKE)*, 2011. Mashhad. /it Proceedings... Mashhad: [s.n.], 2011. p. 60-65.

RIX, A. W.; BEERENDS, J. G.; HOLLIER, M. P.; HEKSTRA, A. P. Perceptual evaluation of speech quality (PESQ) a new method for speech quality assessment of telephone networks

and codecs. *In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*, 2001. Salt Lake City. *Proceedings...* Salt Lake City: 2001. p. 749–752.

ROSS, T. J. *fuzzy logic with engineering applications*. New Mexico: John Wiley & Sons, Lda, 2017. 573 p.

SAHIN, M.; EROL, R. A Comparative Study of Neural Networks and ANFIS for Forecasting Attendance Rate of Soccer Games. *Mathematical and Computational Applications*, Basel, v.22, p. 1-12, 2017.

SITOMPUL, O. S.; NABABAN, E. B.; ALIM, Z. Adaptive distributed grid-partition in generating fuzzy rules. *In: INTERNATIONAL CONFERENCE ON INFORMATION & COMMUNICATION TECHNOLOGY AND SYSTEM*, 2017. Surabaya *Proceedings...* Chengdu City: [s.n.], 2017. p. 1–6.

SONI, H. M.; PATIL, H. A. Non-intrusive quality assessment of noise-suppressed speech using unsupervised deep features. *Speech Communication*, Amsterdam, v.130, p. 27-44, 2021.

SUGENO, M.; KANG, G. T. Structure identification of fuzzy model. *fuzzy Sets and Systems*, Amsterdam, v.28, p. 329–346, 1986.

SUGENO, M.; TANAKA, K. Successive identification of a fuzzy model and its applications to prediction of a complex system. *fuzzy Sets and Systems*, v.42, p. 315-334, 1991.

VARGA, A.; STEENEKEN, H.J.M. Assessment for automatic speech recognition II: NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication*, Amsterdam, n.3, v.12, p. 247-251, 1993.

YAGER, R.R.; FILEV, D.P. Generation of fuzzy rules by mountain clustering. *Journal of Intelligent & fuzzy Systems: Applications in Engineering and Technology*, Amsterdam, v. 2, p. 209–219, 1994.

YEOM, C.U.; KWAK, K.C. Performance comparison of ANFIS models by input space partitioning methods. *Symmetry*, Basel, v.10, p. 1-25, 2018.

APÊNDICES

Apêndice A: Resultados da partição *fuzzy* com o algoritmo FCM para os dados de treinamento com áudios da NOIZEUS contaminados com ruído de bar para os Fatores 1, 11 e 13 de Haralick e suas respectivas avaliações PESQ.

Legenda: ● grupo 1, ● grupo 2 e ● grupo 3 das partições *fuzzy*.

Fator 1	PESQ	$\bar{U}$ - Fator 1/PESQ			Fator 11	PESQ	$\bar{U}$ - Fator 11/PESQ			Fator 13	PESQ	$\bar{U}$ - Fator 13/PESQ		
25,215	4,500	0,051	0,017	0,932	0,887	4,500	0,000	0,000	1,000	0,969	4,500	1,000	0,000	0,000
49,645	1,856	0,011	0,986	0,003	0,892	1,856	0,967	0,032	0,002	0,897	1,856	0,002	0,032	0,967
37,958	2,321	0,994	0,003	0,003	0,866	2,321	0,046	0,950	0,003	0,950	2,321	0,003	0,951	0,046
32,872	2,563	0,407	0,053	0,540	0,869	2,563	0,019	0,977	0,003	0,957	2,563	0,003	0,977	0,019
42,811	2,100	0,600	0,348	0,052	0,868	2,100	0,489	0,500	0,011	0,939	2,100	0,011	0,500	0,489
30,134	4,500	0,041	0,009	0,951	0,882	4,500	0,000	0,000	1,000	0,966	4,500	1,000	0,000	0,000
50,275	1,944	0,023	0,971	0,006	0,902	1,944	0,864	0,131	0,005	0,893	1,944	0,005	0,130	0,865
41,364	2,373	0,828	0,133	0,039	0,875	2,373	0,014	0,984	0,001	0,947	2,373	0,001	0,984	0,014
36,610	2,734	0,928	0,024	0,048	0,873	2,734	0,076	0,901	0,024	0,955	2,734	0,024	0,901	0,076
47,150	2,239	0,026	0,969	0,005	0,884	2,239	0,153	0,839	0,007	0,919	2,239	0,007	0,837	0,156
27,797	4,500	0,003	0,001	0,997	0,875	4,500	0,000	0,000	1,000	0,967	4,500	1,000	0,000	0,000
47,213	1,716	0,021	0,973	0,006	0,899	1,716	0,998	0,002	0,000	0,888	1,716	0,000	0,003	0,997
39,573	2,309	0,977	0,014	0,009	0,868	2,309	0,058	0,938	0,004	0,949	2,309	0,004	0,939	0,058
34,746	2,615	0,717	0,053	0,229	0,869	2,615	0,036	0,957	0,008	0,958	2,615	0,008	0,957	0,036
45,702	1,983	0,127	0,851	0,022	0,886	1,983	0,791	0,202	0,007	0,931	1,983	0,007	0,204	0,789
28,498	4,500	0,000	0,000	1,000	0,892	4,500	0,000	0,000	1,000	0,967	4,500	1,000	0,000	0,000
50,709	1,907	0,032	0,959	0,009	0,907	1,907	0,917	0,079	0,003	0,888	1,907	0,003	0,079	0,918
42,100	2,463	0,724	0,228	0,049	0,880	2,463	0,001	0,999	0,000	0,947	2,463	0,000	0,999	0,000
36,531	2,723	0,922	0,025	0,052	0,876	2,723	0,072	0,906	0,022	0,956	2,723	0,022	0,906	0,072
47,166	2,217	0,025	0,970	0,005	0,884	2,217	0,195	0,797	0,008	0,918	2,217	0,008	0,794	0,198
26,679	4,500	0,019	0,006	0,975	0,891	4,500	0,000	0,000	1,000	0,969	4,500	1,000	0,000	0,000
50,618	1,814	0,030	0,962	0,009	0,906	1,814	0,989	0,011	0,001	0,884	1,814	0,001	0,012	0,988
41,606	2,390	0,796	0,161	0,043	0,870	2,390	0,008	0,991	0,001	0,944	2,390	0,001	0,991	0,008
35,465	2,686	0,815	0,044	0,141	0,872	2,686	0,059	0,925	0,016	0,957	2,686	0,016	0,925	0,059
46,986	2,070	0,031	0,963	0,006	0,886	2,070	0,574	0,416	0,010	0,922	2,070	0,010	0,415	0,575
26,355	4,500	0,026	0,008	0,967	0,892	4,500	0,000	0,000	1,000	0,968	4,500	1,000	0,000	0,000
50,782	1,827	0,033	0,957	0,010	0,906	1,827	0,983	0,016	0,001	0,886	1,827	0,001	0,016	0,983
41,573	2,398	0,800	0,157	0,042	0,872	2,398	0,006	0,993	0,001	0,945	2,398	0,001	0,993	0,006
35,700	2,679	0,843	0,040	0,117	0,871	2,679	0,057	0,928	0,015	0,957	2,679	0,015	0,928	0,057
46,422	2,173	0,065	0,923	0,012	0,886	2,173	0,294	0,696	0,010	0,925	2,173	0,010	0,695	0,295
26,346	4,500	0,026	0,008	0,966	0,884	4,500	0,000	0,000	1,000	0,968	4,500	1,000	0,000	0,000
49,791	1,899	0,013	0,983	0,004	0,899	1,899	0,927	0,070	0,003	0,896	1,899	0,003	0,070	0,927
38,034	2,481	0,996	0,002	0,002	0,869	2,481	0,002	0,998	0,000	0,951	2,481	0,000	0,998	0,002
34,457	2,666	0,671	0,056	0,273	0,873	2,666	0,053	0,934	0,013	0,957	2,666	0,013	0,934	0,052
45,921	2,108	0,106	0,876	0,019	0,878	2,108	0,467	0,523	0,011	0,925	2,108	0,011	0,521	0,468
27,763	4,500	0,003	0,001	0,996	0,887	4,500	0,000	0,000	1,000	0,968	4,500	1,000	0,000	0,000
50,105	1,763	0,019	0,976	0,005	0,905	1,763	0,999	0,001	0,000	0,895	1,763	0,000	0,001	0,999
41,247	2,395	0,842	0,121	0,037	0,873	2,395	0,007	0,993	0,001	0,946	2,395	0,001	0,993	0,007
35,726	2,654	0,846	0,040	0,114	0,873	2,654	0,049	0,940	0,012	0,956	2,654	0,012	0,940	0,048
45,977	2,125	0,101	0,881	0,018	0,880	2,125	0,420	0,570	0,011	0,927	2,125	0,011	0,569	0,420
27,601	4,500	0,005	0,001	0,994	0,877	4,500	0,000	0,000	1,000	0,965	4,500	1,000	0,000	0,000
50,472	1,440	0,027	0,965	0,008	0,907	1,440	0,906	0,085	0,009	0,887	1,440	0,009	0,085	0,906
39,773	2,009	0,964	0,023	0,013	0,865	2,009	0,730	0,262	0,008	0,947	2,009	0,008	0,263	0,728

continua na próxima página

continuação da página anterior

35,365	2,477	0,805	0,046	0,149	0,863	2,477	0,002	0,998	0,000	0,952	2,477	0,000	0,998	0,002
45,252	1,806	0,180	0,792	0,028	0,875	1,806	0,990	0,009	0,001	0,929	1,806	0,001	0,009	0,990
29,522	4,500	0,016	0,004	0,981	0,874	4,500	0,000	0,000	1,000	0,965	4,500	1,000	0,000	0,000
50,369	1,777	0,024	0,969	0,007	0,901	1,777	0,998	0,002	0,000	0,888	1,777	0,000	0,002	0,997
41,767	2,372	0,773	0,182	0,045	0,867	2,372	0,015	0,984	0,001	0,942	2,372	0,001	0,984	0,015
37,575	2,606	0,983	0,007	0,010	0,868	2,606	0,032	0,961	0,007	0,953	2,606	0,007	0,961	0,032
47,252	1,992	0,020	0,976	0,004	0,884	1,992	0,773	0,220	0,007	0,921	1,992	0,007	0,220	0,773
30,031	4,500	0,035	0,008	0,957	0,890	4,500	0,000	0,000	1,000	0,968	4,500	1,000	0,000	0,000
50,771	1,507	0,034	0,957	0,010	0,904	1,507	0,933	0,061	0,006	0,885	1,507	0,006	0,061	0,932
42,905	2,171	0,584	0,364	0,052	0,879	2,171	0,298	0,693	0,010	0,943	2,171	0,010	0,693	0,297
37,564	2,518	0,983	0,007	0,010	0,878	2,518	0,008	0,991	0,001	0,956	2,518	0,001	0,991	0,008
47,278	1,762	0,018	0,978	0,004	0,889	1,762	0,999	0,000	0,000	0,921	1,762	0,000	0,001	0,999
29,981	4,500	0,033	0,007	0,960	0,890	4,500	0,000	0,000	1,000	0,967	4,500	1,000	0,000	0,000
49,856	1,577	0,015	0,981	0,004	0,904	1,577	0,960	0,037	0,003	0,899	1,577	0,003	0,037	0,960
41,829	2,137	0,762	0,192	0,046	0,877	2,137	0,386	0,604	0,010	0,945	2,137	0,011	0,605	0,384
36,244	2,523	0,900	0,030	0,070	0,882	2,523	0,009	0,989	0,001	0,960	2,523	0,001	0,989	0,009
47,007	1,874	0,029	0,965	0,006	0,887	1,874	0,952	0,046	0,002	0,921	1,874	0,002	0,046	0,952
29,375	4,500	0,012	0,003	0,985	0,898	4,500	0,000	0,000	1,000	0,967	4,500	1,000	0,000	0,000
50,934	1,621	0,037	0,952	0,011	0,907	1,621	0,975	0,023	0,002	0,883	1,621	0,002	0,023	0,974
43,282	2,366	0,516	0,431	0,052	0,877	2,366	0,017	0,981	0,001	0,940	2,366	0,001	0,981	0,017
37,864	2,670	0,991	0,004	0,005	0,881	2,670	0,054	0,932	0,014	0,957	2,670	0,014	0,933	0,054
47,916	2,009	0,004	0,995	0,001	0,891	2,009	0,734	0,258	0,008	0,913	2,009	0,008	0,257	0,735
28,888	4,500	0,003	0,001	0,996	0,889	4,500	0,000	0,000	1,000	0,968	4,500	1,000	0,000	0,000
50,220	1,340	0,023	0,971	0,006	0,904	1,340	0,868	0,118	0,014	0,891	1,340	0,015	0,118	0,868
42,671	2,054	0,624	0,324	0,052	0,877	2,054	0,619	0,371	0,010	0,943	2,054	0,010	0,374	0,616
36,279	2,414	0,904	0,030	0,067	0,875	2,414	0,003	0,997	0,000	0,958	2,414	0,000	0,997	0,003
47,489	1,630	0,012	0,985	0,003	0,892	1,630	0,978	0,020	0,002	0,920	1,630	0,002	0,021	0,978
32,575	4,500	0,298	0,045	0,657	0,885	4,500	0,000	0,000	1,000	0,964	4,500	1,000	0,000	0,000
50,256	1,379	0,023	0,970	0,007	0,904	1,379	0,882	0,105	0,012	0,893	1,379	0,012	0,105	0,883
42,559	2,172	0,645	0,303	0,052	0,875	2,172	0,294	0,696	0,010	0,943	2,172	0,010	0,697	0,293
38,358	2,508	1,000	0,000	0,000	0,878	2,508	0,006	0,993	0,001	0,955	2,508	0,001	0,993	0,006
47,783	1,743	0,006	0,993	0,001	0,890	1,743	1,000	0,000	0,000	0,915	1,743	0,000	0,000	1,000
23,205	4,500	0,097	0,036	0,868	0,908	4,500	0,000	0,000	1,000	0,970	4,500	1,000	0,000	0,000
50,265	1,438	0,023	0,970	0,007	0,904	1,438	0,905	0,085	0,009	0,891	1,438	0,009	0,085	0,905
41,154	2,241	0,852	0,112	0,036	0,878	2,241	0,150	0,843	0,007	0,948	2,241	0,007	0,844	0,149
33,802	2,690	0,562	0,057	0,381	0,878	2,690	0,061	0,923	0,016	0,960	2,690	0,016	0,923	0,061
46,559	1,677	0,053	0,937	0,010	0,888	1,677	0,991	0,009	0,001	0,925	1,677	0,001	0,009	0,990
29,356	4,500	0,011	0,003	0,986	0,886	4,500	0,000	0,000	1,000	0,967	4,500	1,000	0,000	0,000
50,082	1,420	0,019	0,975	0,005	0,903	1,420	0,898	0,092	0,010	0,894	1,420	0,010	0,092	0,898
40,770	2,061	0,891	0,079	0,030	0,881	2,061	0,600	0,390	0,010	0,950	2,061	0,010	0,395	0,596
36,496	2,446	0,922	0,026	0,053	0,874	2,446	0,000	1,000	0,000	0,957	2,446	0,000	1,000	0,000
46,736	1,727	0,043	0,949	0,008	0,883	1,727	0,999	0,001	0,000	0,920	1,727	0,000	0,001	0,999
27,788	4,500	0,003	0,001	0,997	0,891	4,500	0,000	0,000	1,000	0,969	4,500	1,000	0,000	0,000
49,685	1,544	0,012	0,985	0,003	0,902	1,544	0,947	0,048	0,005	0,893	1,544	0,005	0,048	0,947
39,681	2,238	0,971	0,018	0,011	0,874	2,238	0,153	0,839	0,007	0,951	2,238	0,007	0,840	0,153
35,288	2,529	0,794	0,047	0,159	0,877	2,529	0,011	0,988	0,002	0,959	2,529	0,002	0,988	0,011
45,783	1,753	0,118	0,862	0,020	0,887	1,753	1,000	0,000	0,000	0,929	1,753	0,000	0,001	0,999

Fonte: Elaborada pelo próprio autor.

Apêndice B: As 969 combinações dos fatores de Haralick.

1/1-2-3	52/1-5-9	103/1-9-14	154/2-3-4	205/2-6-13	256/2-11-14	307/3-5-8	358/3-9-13	409/3-18-19	460/4-9-10
2/1-2-4	53/1-5-10	104/1-9-15	155/2-3-5	206/2-6-14	257/2-11-15	308/3-5-9	359/3-9-14	410/4-5-6	461/4-9-11
3/1-2-5	54/1-5-11	105/1-9-16	156/2-3-6	207/2-6-15	258/2-11-16	309/3-5-10	360/3-9-15	411/4-5-7	462/4-9-12
4/1-2-6	55/1-5-12	106/1-9-17	157/2-3-7	208/2-6-16	259/2-11-17	310/3-5-11	361/3-9-16	412/4-5-8	463/4-9-13
5/1-2-7	56/1-5-13	107/1-9-18	158/2-3-8	209/2-6-17	260/2-11-18	311/3-5-12	362/3-9-17	413/4-5-9	464/4-9-14
6/1-2-8	57/1-5-14	108/1-9-19	159/2-3-9	210/2-6-18	261/2-11-19	312/3-5-13	363/3-9-18	414/4-5-10	465/4-9-15
7/1-2-9	58/1-5-15	109/1-10-11	160/2-3-10	211/2-6-19	262/2-12-13	313/3-5-14	364/3-9-19	415/4-5-11	466/4-9-16
8/1-2-10	59/1-5-16	110/1-10-12	161/2-3-11	212/2-7-8	263/2-12-14	314/3-5-15	365/3-10-11	416/4-5-12	467/4-9-17
9/1-2-11	60/1-5-17	111/1-10-13	162/2-3-12	213/2-7-9	264/2-12-15	315/3-5-16	366/3-10-12	417/4-5-13	468/4-9-18
10/1-2-12	61/1-5-18	112/1-10-14	163/2-3-13	214/2-7-10	265/2-12-16	316/3-5-17	367/3-10-13	418/4-5-14	469/4-9-19
11/1-2-13	62/1-5-19	113/1-10-15	164/2-3-14	215/2-7-11	266/2-12-17	317/3-5-18	368/3-10-14	419/4-5-15	470/4-10-11
12/1-2-14	63/1-6-7	114/1-10-16	165/2-3-15	216/2-7-12	267/2-12-18	318/3-5-19	369/3-10-15	420/4-5-16	471/4-10-12
13/1-2-15	64/1-6-8	115/1-10-17	166/2-3-16	217/2-7-13	268/2-12-19	319/3-6-7	370/3-10-16	421/4-5-17	472/4-10-13
14/1-2-16	65/1-6-9	116/1-10-18	167/2-3-17	218/2-7-14	269/2-13-14	320/3-6-8	371/3-10-17	422/4-5-18	473/4-10-14
15/1-2-17	66/1-6-10	117/1-10-19	168/2-3-18	219/2-7-15	270/2-13-15	321/3-6-9	372/3-10-18	423/4-5-19	474/4-10-15
16/1-2-18	67/1-6-11	118/1-11-12	169/2-3-19	220/2-7-16	271/2-13-16	322/3-6-10	373/3-10-19	424/4-6-7	475/4-10-16
17/1-2-19	68/1-6-12	119/1-11-13	170/2-4-5	221/2-7-17	272/2-13-17	323/3-6-11	374/3-11-12	425/4-6-8	476/4-10-17
18/1-3-4	69/1-6-13	120/1-11-14	171/2-4-6	222/2-7-18	273/2-13-18	324/3-6-12	375/3-11-13	426/4-6-9	477/4-10-18
19/1-3-5	70/1-6-14	121/1-11-15	172/2-4-7	223/2-7-19	274/2-13-19	325/3-6-13	376/3-11-14	427/4-6-10	478/4-10-19
20/1-3-6	71/1-6-15	122/1-11-16	173/2-4-8	224/2-8-9	275/2-14-15	326/3-6-14	377/3-11-15	428/4-6-11	479/4-11-12
21/1-3-7	72/1-6-16	123/1-11-17	174/2-4-9	225/2-8-10	276/2-14-16	327/3-6-15	378/3-11-16	429/4-6-12	480/4-11-13
22/1-3-8	73/1-6-17	124/1-11-18	175/2-4-10	226/2-8-11	277/2-14-17	328/3-6-16	379/3-11-17	430/4-6-13	481/4-11-14
23/1-3-9	74/1-6-18	125/1-11-19	176/2-4-11	227/2-8-12	278/2-14-18	329/3-6-17	380/3-11-18	431/4-6-14	482/4-11-15
24/1-3-10	75/1-6-19	126/1-12-13	177/2-4-12	228/2-8-13	279/2-14-19	330/3-6-18	381/3-11-19	432/4-6-15	483/4-11-16
25/1-3-11	76/1-7-8	127/1-12-14	178/2-4-13	229/2-8-14	280/2-15-16	331/3-6-19	382/3-12-13	433/4-6-16	484/4-11-17
26/1-3-12	77/1-7-9	128/1-12-15	179/2-4-14	230/2-8-15	281/2-15-17	332/3-7-8	383/3-12-14	434/4-6-17	485/4-11-18
27/1-3-13	78/1-7-10	129/1-12-16	180/2-4-15	231/2-8-16	282/2-15-18	333/3-7-9	384/3-12-15	435/4-6-18	486/4-11-19
28/1-3-14	79/1-7-11	130/1-12-17	181/2-4-16	232/2-8-17	283/2-15-19	334/3-7-10	385/3-12-16	436/4-6-19	487/4-12-13
29/1-3-15	80/1-7-12	131/1-12-18	182/2-4-17	233/2-8-18	284/2-16-17	335/3-7-11	386/3-12-17	437/4-7-8	488/4-12-14
30/1-3-16	81/1-7-13	132/1-12-19	183/2-4-18	234/2-8-19	285/2-16-18	336/3-7-12	387/3-12-18	438/4-7-9	489/4-12-15
31/1-3-17	82/1-7-14	133/1-13-14	184/2-4-19	235/2-9-10	286/2-16-19	337/3-7-13	388/3-12-19	439/4-7-10	490/4-12-16
32/1-3-18	83/1-7-15	134/1-13-15	185/2-5-6	236/2-9-11	287/2-17-18	338/3-7-14	389/3-13-14	440/4-7-11	491/4-12-17
33/1-3-19	84/1-7-16	135/1-13-16	186/2-5-7	237/2-9-12	288/2-17-19	339/3-7-15	390/3-13-15	441/4-7-12	492/4-12-18
34/1-4-5	85/1-7-17	136/1-13-17	187/2-5-8	238/2-9-13	289/2-18-19	340/3-7-16	391/3-13-16	442/4-7-13	493/4-12-19
35/1-4-6	86/1-7-18	137/1-13-18	188/2-5-9	239/2-9-14	290/3-4-5	341/3-7-17	392/3-13-17	443/4-7-14	494/4-13-14
36/1-4-7	87/1-7-19	138/1-13-19	189/2-5-10	240/2-9-15	291/3-4-6	342/3-7-18	393/3-13-18	444/4-7-15	495/4-13-15
37/1-4-8	88/1-8-9	139/1-14-15	190/2-5-11	241/2-9-16	292/3-4-7	343/3-7-19	394/3-13-19	445/4-7-16	496/4-13-16
38/1-4-9	89/1-8-10	140/1-14-16	191/2-5-12	242/2-9-17	293/3-4-8	344/3-8-9	395/3-14-15	446/4-7-17	497/4-13-17
39/1-4-10	90/1-8-11	141/1-14-17	192/2-5-13	243/2-9-18	294/3-4-9	345/3-8-10	396/3-14-16	447/4-7-18	498/4-13-18
40/1-4-11	91/1-8-12	142/1-14-18	193/2-5-14	244/2-9-19	295/3-4-10	346/3-8-11	397/3-14-17	448/4-7-19	499/4-13-19
41/1-4-12	92/1-8-13	143/1-14-19	194/2-5-15	245/2-10-11	296/3-4-11	347/3-8-12	398/3-14-18	449/4-8-9	500/4-14-15
42/1-4-13	93/1-8-14	144/1-15-16	195/2-5-16	246/2-10-12	297/3-4-12	348/3-8-13	399/3-14-19	450/4-8-10	501/4-14-16
43/1-4-14	94/1-8-15	145/1-15-17	196/2-5-17	247/2-10-13	298/3-4-13	349/3-8-14	400/3-15-16	451/4-8-11	502/4-14-17
44/1-4-15	95/1-8-16	146/1-15-18	197/2-5-18	248/2-10-14	299/3-4-14	350/3-8-15	401/3-15-17	452/4-8-12	503/4-14-18
45/1-4-16	96/1-8-17	147/1-15-19	198/2-5-19	249/2-10-15	300/3-4-15	351/3-8-16	402/3-15-18	453/4-8-13	504/4-14-19
46/1-4-17	97/1-8-18	148/1-16-17	199/2-6-7	250/2-10-16	301/3-4-16	352/3-8-17	403/3-15-19	454/4-8-14	505/4-15-16
47/1-4-18	98/1-8-19	149/1-16-18	200/2-6-8	251/2-10-17	302/3-4-17	353/3-8-18	404/3-16-17	455/4-8-15	506/4-15-17
48/1-4-19	99/1-9-10	150/1-16-19	201/2-6-9	252/2-10-18	303/3-4-18	354/3-8-19	405/3-16-18	456/4-8-16	507/4-15-18
49/1-5-6	100/1-9-11	151/1-17-18	202/2-6-10	253/2-10-19	304/3-4-19	355/3-9-10	406/3-16-19	457/4-8-17	508/4-15-19
50/1-5-7	101/1-9-12	152/1-17-19	203/2-6-11	254/2-11-12	305/3-5-6	356/3-9-11	407/3-17-18	458/4-8-18	509/4-16-17
51/1-5-8	102/1-9-13	153/1-18-19	204/2-6-12	255/2-11-13	306/3-5-7	357/3-9-12	408/3-17-19	459/4-8-19	510/4-16-18

continua na próxima página

continuação da página anterior

511/4-16-19	562/5-10-12	613/6-7-15	664/6-13-15	715/7-11-13	766/8-10-17	817/9-11-15	868/10-13-17	919/12-13-19
512/4-17-18	563/5-10-13	614/6-7-16	665/6-13-16	716/7-11-14	767/8-10-18	818/9-11-16	869/10-13-18	920/12-14-15
513/4-17-19	564/5-10-14	615/6-7-17	666/6-13-17	717/7-11-15	768/8-10-19	819/9-11-17	870/10-13-19	921/12-14-16
514/4-18-19	565/5-10-15	616/6-7-18	667/6-13-18	718/7-11-16	769/8-11-12	820/9-11-18	871/10-14-15	922/12-14-17
515/5-6-7	566/5-10-16	617/6-7-19	668/6-13-19	719/7-11-17	770/8-11-13	821/9-11-19	872/10-14-16	923/12-14-18
516/5-6-8	567/5-10-17	618/6-8-9	669/6-14-15	720/7-11-18	771/8-11-14	822/9-12-13	873/10-14-17	924/12-14-19
517/5-6-9	568/5-10-18	619/6-8-10	670/6-14-16	721/7-11-19	772/8-11-15	823/9-12-14	874/10-14-18	925/12-15-16
518/5-6-10	569/5-10-19	620/6-8-11	671/6-14-17	722/7-12-13	773/8-11-16	824/9-12-15	875/10-14-19	926/12-15-17
519/5-6-11	570/5-11-12	621/6-8-12	672/6-14-18	723/7-12-14	774/8-11-17	825/9-12-16	876/10-15-16	927/12-15-18
520/5-6-12	571/5-11-13	622/6-8-13	673/6-14-19	724/7-12-15	775/8-11-18	826/9-12-17	877/10-15-17	928/12-15-19
521/5-6-13	572/5-11-14	623/6-8-14	674/6-15-16	725/7-12-16	776/8-11-19	827/9-12-18	878/10-15-18	929/12-16-17
522/5-6-14	573/5-11-15	624/6-8-15	675/6-15-17	726/7-12-17	777/8-12-13	828/9-12-19	879/10-15-19	930/12-16-18
523/5-6-15	574/5-11-16	625/6-8-16	676/6-15-18	727/7-12-18	778/8-12-14	829/9-13-14	880/10-16-17	931/12-16-19
524/5-6-16	575/5-11-17	626/6-8-17	677/6-15-19	728/7-12-19	779/8-12-15	830/9-13-15	881/10-16-18	932/12-17-18
525/5-6-17	576/5-11-18	627/6-8-18	678/6-16-17	729/7-13-14	780/8-12-16	831/9-13-16	882/10-16-19	933/12-17-19
526/5-6-18	577/5-11-19	628/6-8-19	679/6-16-18	730/7-13-15	781/8-12-17	832/9-13-17	883/10-17-18	934/12-18-19
527/5-6-19	578/5-12-13	629/6-9-10	680/6-16-19	731/7-13-16	782/8-12-18	833/9-13-18	884/10-17-19	935/13-14-15
528/5-7-8	579/5-12-14	630/6-9-11	681/6-17-18	732/7-13-17	783/8-12-19	834/9-13-19	885/10-18-19	936/13-14-16
529/5-7-9	580/5-12-15	631/6-9-12	682/6-17-19	733/7-13-18	784/8-13-14	835/9-14-15	886/11-12-13	937/13-14-17
530/5-7-10	581/5-12-16	632/6-9-13	683/6-18-19	734/7-13-19	785/8-13-15	836/9-14-16	887/11-12-14	938/13-14-18
531/5-7-11	582/5-12-17	633/6-9-14	684/7-8-9	735/7-14-15	786/8-13-16	837/9-14-17	888/11-12-15	939/13-14-19
532/5-7-12	583/5-12-18	634/6-9-15	685/7-8-10	736/7-14-16	787/8-13-17	838/9-14-18	889/11-12-16	940/13-15-16
533/5-7-13	584/5-12-19	635/6-9-16	686/7-8-11	737/7-14-17	788/8-13-18	839/9-14-19	890/11-12-17	941/13-15-17
534/5-7-14	585/5-13-14	636/6-9-17	687/7-8-12	738/7-14-18	789/8-13-19	840/9-15-16	891/11-12-18	942/13-15-18
535/5-7-15	586/5-13-15	637/6-9-18	688/7-8-13	739/7-14-19	790/8-14-15	841/9-15-17	892/11-12-19	943/13-15-19
536/5-7-16	587/5-13-16	638/6-9-19	689/7-8-14	740/7-15-16	791/8-14-16	842/9-15-18	893/11-13-14	944/13-16-17
537/5-7-17	588/5-13-17	639/6-10-11	690/7-8-15	741/7-15-17	792/8-14-17	843/9-15-19	894/11-13-15	945/13-16-18
538/5-7-18	589/5-13-18	640/6-10-12	691/7-8-16	742/7-15-18	793/8-14-18	844/9-16-17	895/11-13-16	946/13-16-19
539/5-7-19	590/5-13-19	641/6-10-13	692/7-8-17	743/7-15-19	794/8-14-19	845/9-16-18	896/11-13-17	947/13-17-18
540/5-8-9	591/5-14-15	642/6-10-14	693/7-8-18	744/7-16-17	795/8-15-16	846/9-16-19	897/11-13-18	948/13-17-19
541/5-8-10	592/5-14-16	643/6-10-15	694/7-8-19	745/7-16-18	796/8-15-17	847/9-17-18	898/11-13-19	949/13-18-19
542/5-8-11	593/5-14-17	644/6-10-16	695/7-9-10	746/7-16-19	797/8-15-18	848/9-17-19	899/11-14-15	950/14-15-16
543/5-8-12	594/5-14-18	645/6-10-17	696/7-9-11	747/7-17-18	798/8-15-19	849/9-18-19	900/11-14-16	951/14-15-17
544/5-8-13	595/5-14-19	646/6-10-18	697/7-9-12	748/7-17-19	799/8-16-17	850/10-11-12	901/11-14-17	952/14-15-18
545/5-8-14	596/5-15-16	647/6-10-19	698/7-9-13	749/7-18-19	800/8-16-18	851/10-11-13	902/11-14-18	953/14-15-19
546/5-8-15	597/5-15-17	648/6-11-12	699/7-9-14	750/8-9-10	801/8-16-19	852/10-11-14	903/11-14-19	954/14-16-17
547/5-8-16	598/5-15-18	649/6-11-13	700/7-9-15	751/8-9-11	802/8-17-18	853/10-11-15	904/11-15-16	955/14-16-18
548/5-8-17	599/5-15-19	650/6-11-14	701/7-9-16	752/8-9-12	803/8-17-19	854/10-11-16	905/11-15-17	956/14-16-19
549/5-8-18	600/5-16-17	651/6-11-15	702/7-9-17	753/8-9-13	804/8-18-19	855/10-11-17	906/11-15-18	957/14-17-18
550/5-8-19	601/5-16-18	652/6-11-16	703/7-9-18	754/8-9-14	805/9-10-11	856/10-11-18	907/11-15-19	958/14-17-19
551/5-9-10	602/5-16-19	653/6-11-17	704/7-9-19	755/8-9-15	806/9-10-12	857/10-11-19	908/11-16-17	959/14-18-19
552/5-9-11	603/5-17-18	654/6-11-18	705/7-10-11	756/8-9-16	807/9-10-13	858/10-12-13	909/11-16-18	960/15-16-17
553/5-9-12	604/5-17-19	655/6-11-19	706/7-10-12	757/8-9-17	808/9-10-14	859/10-12-14	910/11-16-19	961/15-16-18
554/5-9-13	605/5-18-19	656/6-12-13	707/7-10-13	758/8-9-18	809/9-10-15	860/10-12-15	911/11-17-18	962/15-16-19
555/5-9-14	606/6-7-8	657/6-12-14	708/7-10-14	759/8-9-19	810/9-10-16	861/10-12-16	912/11-17-19	963/15-17-18
556/5-9-15	607/6-7-9	658/6-12-15	709/7-10-15	760/8-10-11	811/9-10-17	862/10-12-17	913/11-18-19	964/15-17-19
557/5-9-16	608/6-7-10	659/6-12-16	710/7-10-16	761/8-10-12	812/9-10-18	863/10-12-18	914/12-13-14	965/15-18-19
558/5-9-17	609/6-7-11	660/6-12-17	711/7-10-17	762/8-10-13	813/9-10-19	864/10-12-19	915/12-13-15	966/16-17-18
559/5-9-18	610/6-7-12	661/6-12-18	712/7-10-18	763/8-10-14	814/9-11-12	865/10-13-14	916/12-13-16	967/16-17-19
560/5-9-19	611/6-7-13	662/6-12-19	713/7-10-19	764/8-10-15	815/9-11-13	866/10-13-15	917/12-13-17	968/16-18-19
561/5-10-11	612/6-7-14	663/6-13-14	714/7-11-12	765/8-10-16	816/9-11-14	867/10-13-16	918/12-13-18	969/17-18-19

Fonte: Elaborada pelo próprio autor.