UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE ENGENHARIA
CAMPUS DE ILHA SOLTEIRA

Caio Cesar Enside de Abreu

Melhoramento de Sinais de Voz Baseado na Identificação de Padrões Ruidosos

Ilha Solteira
2017

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE ENGENHARIA
CAMPUS DE ILHA SOLTEIRA

Caio Cesar Enside de Abreu

Melhoramento de Sinais de Voz Baseado na Identificação de Padrões Ruidosos

Dr. Francisco Villarreal Alvarado
Orientador

Tese de Doutorado apresentada à Faculdade de Engenharia do Campus de Ilha Solteira - UNESP, como parte dos requisitos para obtenção do Grau de Doutor em Engenharia Elétrica.
Área de Conhecimento: Automação.

Ilha Solteira
2017

FICHA CATALOGRÁFICA
Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

A162m

Abreu, Caio Cesar Enside de .
Melhoramento de sinais de voz baseado na identificação de padrões ruidosos / Caio Cesar Enside de Abreu. -- Ilha Solteira: [s.n.], 2017
120 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia. Área de conhecimento: Automação, 2017

Orientador: Francisco Villarreal Alvarado
Inclui bibliografia

1. Melhoramento de voz. 2. Análise de métodos. 3. Classificação de ruído. 4. Transformada wavelet complexa. 5. Identificação de padrões ruidosos.

unesp

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Ilha Solteira

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Melhoramento de Sinais de Voz Baseado na Identificação de Padrões Ruidosos

AUTOR: CAIO CESAR ENSIDE DE ABREU
ORIENTADOR: FRANCISCO VILLARREAL ALVARADO

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

Prof. Dr. FRANCISCO VILLARREAL ALVARADO
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira

Prof. Dr. JOZUE VIEIRA FILHO
Coordenadoria Executiva / UNESP - Câmpus de São João da Boa Vista

Prof. Dr. CARLOS ROBERTO MINUSSI
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira

Prof. Dr. ROBERTO KAWAKAMI HARROP GALVÃO
Divisão de Engenharia Eletrônica, Departamento de Sistemas e Controle / Instituto Tecnológico de Aeronáutica

Prof. Dr. MÁRCIO ROBERTO COVACIC
Departamento de Engenharia Elétrica / UNIVERSIDADE ESTADUAL DE LONDRINA

Ilha Solteira, 03 de março de 2017

Dedico este trabalho a Deus e à minha família.
Aos jovens estudantes de graduação e de pós-graduação que estão iniciando suas jornadas acadêmicas.
Àqueles que buscam a luz e a libertação por meio do conhecimento.
Àqueles que enfrentam os problemas com paciência, perseverança e gostam de programar.

AGRADECIMENTOS

Agradeço a Deus pelas bênçãos, sabedoria e conhecimentos adquiridos durante esta jornada;

Aos orientadores Francisco Villarreal Alvarado e Marco Aparecido Queiroz Duarte pela oportunidade e confiança;

Às instituições UNEMAT e UNESP;

Aos amigos e companheiros de pós-graduação Arthur Gagg, Gapira, Júlio Cesar Uzinski, Fernando Parra, Simone Frutuoso, Ueslei Fernandes, Diogo Rupolo, Jeferson Vanderlinde, João Zamperin, Stefani Carolline, Rafael Cuerda, Patricia Fernanda, Klayton, Raiane, Riciane, André Domingues, Rafael (Mamute), Ricardo (Crocodilo), Wesley (Madruga), João Ricardo (Mosquito) e Ana Eliza Lima, pelos bons momentos e por marcarem a etapa Ilha Solteira da minha vida;

Aos professores do PPGEE-FEIS Jozue, Minussi, Anna Diva, Tokio e Rubén pelo conhecimento e experiência transmitidos a mim durante as disciplinas e bancas de estudos especiais, qualificações de mestrado e de doutorado;

Aos companheiros que pensam a universidade, a sociedade e a docência junto ao Departamento de Computação - UNEMAT - AIA: Lucas Sperotto, Fernando Obana, Juvenal Silva Neto, Adriana Dias, Carlinho Viana, Max Robert, Sérgio Santos, Carlos Alex, Sérgio Eduardo, Toni Amorin, Gleber Marques, Ubirajara Coelho, Darley Domingos e Wesley Thereza;

Aos meus pais Rosângela e Haroldo pelo apoio e orações direcionadas a mim nos momentos mais difíceis;

À minha irmã Tatiane Abreu e seu esposo Rafael Marçal pela oportunidade de podermos vivenciar momentos descontraídos e alegres, que são tão importantes nessa vida cheia de prazos e compromissos;

Aos meus avós Marina e Romildo por me ensinarem que a simplicidade e a humildade faz a vida mais bela;

À minha querida Grazi, companheira de vida e de lutas por uma educação melhor às nossas crianças e jovens.

"You may lose your faith in us,
but never in yourselves."
(Optimus Prime)

RESUMO

Este trabalho propõe contribuir com pesquisas em melhoramento de voz (MV) por meio do estudo de diversos tipos de algoritmos baseados em Fourier e wavelets, assim como o desenvolvimento de uma ferramenta para a identificação e classificação do ruído, culminando com uma nova metodologia. Denominada "Conjunto de Métodos de Melhoramento de Voz (CMMV)", a metodologia consiste em utilizar um banco de dados com sentenças contaminadas com vários tipos de ruídos reais, ajustando, em modo *off-line*, vários métodos de MV para cada tipo de ruído. Os melhores métodos para cada tipo de ruído são selecionados para compor o conjunto de métodos. Durante a operação, em modo *on-line*, um classificador de ruído prediz o tipo de ruído presente no sinal em processamento e então o melhor método é escolhido dentro do CMMV construído. Seis tipos de ruídos foram utilizados durante as simulações e os métodos que obtiveram melhor desempenho frente a cada tipo foram indicados por meio de análise objetiva. Constatou-se que o desempenho desses métodos pode variar de acordo com o tipo do ruído de fundo, confirmando que o desenvolvimento de algoritmos que trabalham eficientemente em qualquer ambiente ruidoso, incorporando classificação de ruído, é uma tendência. O classificador de ruídos desenvolvido nesta pesquisa tem como base um sistema imunológico artificial e características extraídas por uma análise multiescala fornecida pela transformada wavelet complexa. Com uma acurácia média de 96,29% para os seis tipos de ruído considerados e tempo de resposta médio de 6,9 milissegundos, o classificador desenvolvido se mostrou viável para implementações e utilização em conjunto com outras tecnologias. Explorando algumas das possibilidades e benefícios do processamento baseado na classificação do ruído, a seguinte questão foi levantada: "seria possível realizar uma razoável estimação do ruído a partir do sinal de voz ruidoso por meio de regressão?". Esta questão surgiu durante o desenvolvimento da pesquisa, pois o bom funcionamento de métodos de MV depende de uma boa estimação do perfil do ruído. As simulações mostram que este tipo de estimação de ruído pode gerar resultados satisfatórios com menor custo computacional. Por fim, comparado aos métodos clássicos, o CMMV mostrou-se tão ou mais eficiente quanto.

Palavras-chave: Melhoramento de voz. Análise de métodos. Classificação de ruído. Transformada wavelet complexa. Identificação de padrões ruidosos.

ABSTRACT

This work proposes to contribute with researches in speech enhancement by means of the study of several types of Fourier and wavelet based algorithms, as well as the development of a tool for noise identification and classification, culminating in a new methodology. Named "Set of Speech Enhancement Methods (SSEM)" the methodology consists in using a noisy speech database encompassing several types of real noise, adjusting, in off-line mode, several speech enhancement methods for each noise type. The best methods for each noise type are selected to compose the set of methods. During operation, in on-line mode, a noise classifier predicts the noise type in a noisy sentence, and the best method is chosen within the SSEM. Six noise types have been used for the simulations and the methods that achieved better results were appointed by means of objective analysis. It was found that the methods performance can vary according to the background noise type, confirming that the design of algorithms which work efficiently in any noisy environment, by incorporating noise classification, is a trend. The developed noise classifier in this research is based on an artificial immune system and features extracted from a multiscale analysis provided by the complex wavelet transform. With an average hit rate of 96,29% for the six noise types considered and average response time of 6.9 milliseconds, the developed classifier proved to be feasible for implementation and to be used together with other technologies. In order to explore some possibilities and benefits of noise classification based processing, the following question raised: "is it possible to perform a reasonable noise estimation from corrupted speech sentences by a regression method?" This question arose during the research development, since the proper functioning of speech enhancement algorithms depends on good noise profile estimation. Simulations show that this kind of noise estimation can generate satisfactory results with lower computational cost. Finally, compared to the classical methods, the SSEM proved to be superior.

Keywords: Speech enhancement. Analysis of methods. Noise classification. Complex wavelet transform. Noise patterns recognition.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ASN	Algoritmo de Seleção Negativa
AD	Árvore de Decisão
CART	*Classification and Regression Tree*
ℂWT	*Complex Wavelet Transform*
CMMV	Conjunto de Métodos de Melhoramento de Voz
DAV	Detector de Atividade de Voz
DFT	*Discrete Fourier transform*
DWT	*Discrete Wavelet Transform*
DWPT	*Discrete Wavelet Packet Transform*
DT-ℂWT	*Dual-Tree Complex Wavelet Transform*
ERBR	Estimação de Ruído Baseado em Regressão
ERBRs	Estimação de Ruído Baseado em Regressão utilizando um único modelo de regressão
ERBRm	Estimação de Ruído Baseado em Regressão utilizando vários modelos de regressão
FFT	*Fast Fourier Transform*
Hz	hertz
ITU-T	*International Telecommunications Union*
LLR	*Log Likelihood Ratio*
MB	Máscara Binária
ML	*Maximum Likelihood*
MAP	*Maximum a Posterior Estimator*
MOS	*Mean Opinion Score*
MV	Melhoramento de Voz
MCRA	*Minima Controlled Recursive Averaging*

MMSE	*Minimum Mean-Square Error*
OLA	*overlap–add method*
PLS	*Partial Least Squares*
PESQ	*Perceptual Evaluation of Speech Quality*
PR	*Perfect Reconstruction*
SSNR	*Posteriori Segmental Signal-to-Noise Ratio*
PCA	*Principal Component Analysis*
PCR	*Principal Component Regression*
RAV	Reconhecimento Automático de Voz
RN	Rede Neural
RNPM	Rede Neural Perceptron Multicamadas
SNR	Relação Sinal Ruído
SNR_{post}	Relação Sinal Ruído *a Posteriori*
SNR_{prio}	Relação Sinal Ruído *a Priori*
RGB	Ruído Gaussiano Branco
SIA	Sistema Imunológico Artificial
SIH	Sistema Imunológico Humano
SE	Subtração Espectral
SEP	Subtração Espectral em Potência
SVM	*Support Vector Machines*
WT	*Wavelet Transform*
WSS	*Weighted-Slope Spectral Distance*

LISTA DE SÍMBOLOS

$\mathbb{R}$	Conjunto dos números reais
$\mathbb{Z}$	Conjunto dos números inteiros
$\in$	Elemento de
$\hat{f}(\omega)$	Transformada de Fourier da função $f(t)$
$\psi(t)$	Função wavelet real
$\phi(t)$	Função escala real
$\psi_c(t)$	Função wavelet complexa
$\phi_c(t)$	Função escala complexa
$d(k,n)$	Coeficientes wavelet reais na escala k (também denominado coeficientes de detalhes).
$d_c(k,n)$	Coeficientes wavelet complexos na escala k
$c(k,n)$	Coeficientes de aproximação na escala k
h_0	Filtro passa-baixa
h_1	Filtro passa-alta
f_s	Taxa de amostragem do sinal f
$\mathcal{H}\{.\}$	Transformada de Hilbert
$y[n]$	Sinal de voz contaminado por ruído
$x[n]$	Sinal de voz puro
$w[n]$	Ruído
$Y[l,n]$	n-ésimo componente espectral na l-ésima janela de $y[n]$
$Y[l,k,n]$	Projeção da l-ésima janela sobre a k-ésima escala no domínio wavelet de $y[n]$
Y	Equivalente a $Y[l,k,n]$
$\|.\|$	Valor absoluto
$\angle$	Fase

$\widehat{X}_{l,n}$	n-ésimo componente de magnitude espectral estimado do sinal puro $x[n]$, na l-ésima janela
$\widehat{W}_{l,n}$	n-ésimo componente de magnitude espectral estimado do ruído $w[n]$, na l-ésima janela
$E\{.\}$	Operador Esperança
$\phi^x_{l,n}$	n-ésimo componente de densidade espectral de potência de $x[n]$, na l-ésima janela
$\phi^w_{l,n}$	n-ésimo componente de densidade espectral de potência de $w[n]$, na l-ésima janela
H^w	Filtro de Wiener
γ	Operador SNR_{post}
ξ	Operador SNR_{prio}
$\hat{\gamma}$	SNR_{post} estimada
$\hat{\xi}$	SNR_{prio} estimada
$X^2_{l,n}$	n-ésimo componente do espectro de potência do sinal puro, na l-ésima janela
$\widehat{X}^2_{l,n}$	n-ésimo componente do espectro de potência do sinal puro estimado, na l-ésima janela
$W^2_{l,n}$	n-ésimo componente do espectro de potência do ruído, na l-ésima janela
$\widehat{W}^2_{l,n}$	n-ésimo componente do espectro de potência do ruído estimado, na l-ésima janela
$P[.]$	Retificação de meia onda
$\text{THR}(.,.)$	Função de limiar
λ	Limiar
$sign$	Função *signun*
$c^p_{k,n}$	n-ésimo coeficiente do vetor de diferenças centradas de ordem p, calculado sobre a escala wavelet k
$a^p_{k,n}$	n-ésimo coeficiente do vetor de diferenças avançadas de ordem p, calculado sobre a escala wavelet k

ORSP	Operador relação sinal/ruído *a priori*
$\mathbf{P}$	Conjunto (matriz) de dados próprios
$\mathbf{C}$	Conjunto (matriz) de todos os dados adquiridos
c_j	Elemento de $\mathbf{C}$
$\mathbf{R}$	Conjunto (matriz) de detectores
$\mathbf{P}_*$	Conjunto (matriz) de dados a ser protegido
$\mathbf{C}_*$	Subconjunto (matriz) de $\mathbf{C}$ contendo amostras de ruído do tipo $*$
$\mathbf{c}_j^*$	Elementos de $\mathbf{C}_*$
$\mathbf{R}_*$	Conjunto (matriz) de detectores referente ao ruído do tipo $*$
Ab	Vetor de valores reais que representa os anticorpos
Ag	Vetor de valores reais que representa os antígenos
$d(Ab, Ag)$	Distância entre os vetores Ab e Ag
$\bar{y}_i$	Normalização para $y[n]$
$max(y)$	Valor máximo do vetor y
$\mathbf{B}$	Modelo de regressão
$E_{l,n}$	Erro envolvido no processo de regressão para a janela l
$\mathrm{C_{ovl}}$	Medida que avalia a qualidade geral de um sinal de voz
$\mathrm{C_{bak}}$	Medida que avalia a distorção de fundo em um sinal de voz
$\mathrm{C_{sig}}$	Medida que avalia o nível de distorções de fala em um sinal de voz

SUMÁRIO

1 INTRODUÇÃO

Melhoramento de voz (do inglês: *speech enhancement*) é uma área de pesquisa bem consolidada e amplamente explorada. Métodos de melhoramento de voz (MV) são importantes, pois, muitas vezes, os sinais transmitidos pelos meios de telecomunicações podem conter algum tipo de ruído (DELLER; PROAKIS; HANSEN, 1993). Um sinal de voz pode ser contaminado por ruído ao ser gravado, captado por um microfone ou devido ao meio de reprodução. De um modo geral, algoritmos de redução de ruído são indispensáveis em aplicações como reconhecimento automático de voz (RAV) e comunicação em ambientes ruidosos, sendo fortemente utilizados pela indústria de entretenimento.

Além das aplicações clássicas em telecomunicações, destaca-se também o constante desenvolvimento de smartphones, que estão cada vez mais acessíveis à população em geral. Com eles, uma vasta gama de aplicativos para internet móvel é ofertada pelo mercado. Dentre estes, destacam-se os aplicativos de busca na internet por RAV e a realização de chamadas por comando de voz. Outro fato é o aprimoramento das centrais multimídias em automóveis que, em algumas versões, também integram funções acionadas por comando de voz. De um modo geral, com a oferta de tecnologia a preço acessível, aumenta-se a demanda por métodos que realizem algum tipo de processamento em sinais de voz.

Existem também aplicações voltadas para a área de medicina onde métodos de MV são utilizados em conjunto com aparelhos auditivos. Como um caso específico, citam-se os métodos desenvolvidos em prol dos pacientes que precisam do implante coclear. Nestes casos, ao amplificar o sinal de áudio, estes aparelhos amplificam também o ruído acústico do ambiente. Alguns dos ambientes ruidosos comuns ao quotidiano das pessoas são: ruído de carro (interior ou exterior), vozes (muitas pessoas falando em um mesmo ambiente), restaurantes, shoppings, entre outros. Como exemplo de aplicações de MV na medicina, citam-se os trabalhos em Hu e Loizou (2008b), Bogaert et al. (2009) e Cornelis, Moonen e Wouters (2011).

No campo da robótica e mecatrônica, sistemas de processamento de voz vêm sendo utilizados em aplicações direcionadas à interação humano-robô, como, por exemplo, o reconhecimento de emoções através da fala (BREAZEAL, 2002; KIM et al., 2009), classificação de cenários (RAKOTOMAMONJY; GASSO, 2015), reconhecimento de voz e aplicações dependentes do contexto (DOOSTDAR et al., 2009; MA; SMITH; MILNER, 2003). No entanto, no processo de aquisição do sinal de voz, o ruído acústico do ambiente é também capturado, tornando-se o principal problema para aplicações comerciais. Consequentemente, um pré-processamento é necessário a fim de melhorar a qualidade do sinal capturado. Desta forma, pesquisadores têm combinado métodos de MV com outros sistemas de processamento de voz a fim de obter uma melhor performance.

Quando se propõe o uso de um software para reduzir o ruído de fundo contido

em sinais de voz, o objetivo é melhorar a qualidade psicoacústica, tornando o áudio mais agradável ao ouvinte. Do ponto de vista de aplicações que necessitam de RAV, objetiva-se reduzir qualquer tipo de ruído captado juntamente com a amostra de voz do usuário, pois o ruído pode interferir no desempenho do algoritmo de reconhecimento. Em outras palavras, independente do contexto, o objetivo é recuperar o sinal de voz puro, sem a presença do ruído.

Os métodos clássicos de melhoramento de voz são aqueles baseados na transformada discreta de Fourier (DFT - do inglês: *Discrete Fourier Transform*) (BOLL, 1979; EPHRAIM; MALAH, 1984) e na limiarização wavelet (DONOHO; JOHNSTONE, 1994; DONOHO, 1995). No entanto, em termos mais amplos, tais métodos podem ser divididos em algumas classes:

- **Métodos baseados em subtração espectral:** Esta classe inclui a tradicional Subtração Espectral (SE) e suas versões modificadas, por exemplo, Subtração Espectral Generalizada (BEROUTI; SCHWARTZ; MAKHOUL, 1979), Subtração Espectral em Potência (SEP) e Subtração Espectral Multibanda (KAMATH; LOIZOU, 2002);

- **Métodos baseados em modelos estatísticos:** Nesta classe estão principalmente os métodos baseados em estimadores do Erro Médio Quadrático (MMSE - do inglês: *Minimum Mean Square Error estimators*) (EPHRAIM; MALAH, 1984; EPHRAIM; MALAH, 1985; PALIWAL; SCHWERIN; WÓJCICKI, 2012; ABUTALEBI; RASHIDINEJAD, 2015), algoritmos baseados no filtro de Wiener (SCALART; VIEIRA FILHO, 1996; ALMAJAI; MILNER, 2011; RAO; MURTHY; RAO, 2012; EL-FATTAH et al., 2014; KODRASI; MARQUARDT; DOCLO, 2015), métodos baseados no estimador de Máximo *a Posteriori* (MAP) (GAUVAIN; CHIN-HUI, 1994; LOTTER; VARY, 2005; CHEHREHSA; MOIR, 2016), estimadores de Máxima Verossimilhança (ML - do inglês: *Maximum Likelihood* (MCAULAY; MALPASS, 1980; MENG; RUBIN, 1993; MOHANAPRASAD; ARULMOZHIVARMAN, 2015), estimadores Bayesianos (EPHRAIM, 1992; LOIZOU, 2005; PARCHAMI et al., 2015) e a limiarização wavelet (DONOHO, 1995; LALLOUANI; GABREA; GARGOUR, 2004; GHANBARI; KARAMI-MOLLAEI, 2006; SHEIKHZADEH; ABUTALEBI, 2011; ISLAM et al., 2015; SWAMI et al., 2015; TABIBIAN; AKBARI; NASERSHARIF, 2015; MESSAOUD; BOUZID; ELLOUZE, 2016);

- **Algoritmos subespaciais:** Esta classe inclui métodos que se baseiam na decomposição em valores singulares (SVD - do inglês: *Singular Value Decomposition*) e na decomposição em autovalores (EVD - do inglês: *eigenvalue decomposition*) (EPHRAIM; TREES, 1995; JABLOUN; CHAMPAGNE, 2003).

Em um período mais recente, métodos de MV que atuam no domínio wavelet considerados não limiares foram desenvolvidos com o objetivo de superar algumas deficiências da limiarização. Como exemplo, citam-se os trabalhos em Soares et al. (2011) e Abreu, Duarte e Villarreal (2012). Em Abreu, Duarte e Villarreal (2013), os autores verificaram que os métodos não limiares superam seus antecessores no quesito qualidade psicoacústica, pois realizam uma redução uniforme de ruído durante todo o processamento. Isto se justifica pelo fato de que o processo de busca pelo melhor limiar a cada janela pode gerar estampidos ou sussurros no sinal processado (SOARES et al., 2011).

Apesar das pesquisas estarem bem consolidadas, a grande maioria dos métodos de MV são desenvolvidos sobre o pressuposto de que o ruído de fundo é o Gaussiano Branco. Isto se deve ao fato de que este tipo de ruído possui propriedades conhecidas, sendo desejável em operações analíticas. Em outras palavras, a base matemática que sustenta a grande maioria dos métodos de MV foi desenvolvida considerando a contaminação por ruído Gaussiano Branco (RGB). Por exemplo, a maneira tradicional de se estimar o limiar, proposta por Donoho e Johnstone (1994), se baseia no pressuposto de que o ruído de fundo é o RGB.

Frequentemente utilizado por pesquisadores da área, o RGB é fisicamente irrealizável, ou seja, não existe em situações reais. Este tipo de ruído é dito branco (por analogia a luz) pois contém todas as componentes de frequência, isto é, possui um espectro de frequência bem amplo.

O fato de algum método trabalhar muito bem sobre um sinal contaminado com o RGB não garante que ele seja eficiente para outros tipos de ruído. Pelo contrário, alguns métodos de remoção ou redução de ruído, quando avaliados com algum tipo de ruído colorido e o ruído branco, alcançaram melhores resultados quando o ruído presente no sinal é o ruído branco. Citam-se como exemplo os resultados obtidos em Lu e Loizou (2011) e Soares et al. (2011).

Em contrapartida ao RGB, que é considerado temporalmente homogêneo e estacionário, o que se vê em aplicações reais são os ruídos ditos coloridos. Em alguns casos, o ruído colorido pode apresentar características não estacionárias, tornando difícil a estimação de seu perfil.

Devido ao pressuposto sobre as características do ruído de fundo, para a implementação de um método de limiarização wavelet, um único limiar para todas as sub-bandas wavelet é utilizado. Assim, os resultados podem não ser tão bons em condições de ruído real. Várias estratégias para melhorar a estimação do limiar foram propostas. Nos trabalhos realizados por Ghanbari e Karami-Mollaei (2006) e Tabibian, Akbari e Nasersharif (2015), os autores propuseram métodos de limiarização wavelet baseados na estimação adaptativa do limiar, sobre os coeficientes gerados pela transformada wavelet packet. Dessa forma, um limiar diferente para cada escala wavelet foi estimado. Além de calcular

o limiar de maneira adaptativa, uma nova função para estimação do limiar foi proposta em Tabibian, Akbari e Nasersharif (2015). No entanto, assim como em Donoho e Johnstone (1994), todo o equacionamento foi desenvolvido sobre o pressuposto do RGB.

De um modo geral, nos métodos baseados em modelos estatísticos, é comum o pressuposto de que os coeficientes espectrais do ruído representem uma distribuição Gaussiana (EPHRAIM; MALAH, 1984; LOTTER; VARY, 2005; LU; LOIZOU, 2011). De acordo com Ephraim e Malah (1984), a distribuição de probabilidade da voz e do ruído são desconhecidas e é comum assumir um modelo estatístico razoável, no caso, um modelo estatístico Gaussiano. Talvez, em trabalhos futuros, modelos estatísticos desenvolvidos individualmente para cada tipo de ambiente ruidoso real possam ser propostos. Outra possibilidade seria a de escolher diferentes modelos estatísticos para diferentes tipos de ruído. Assim, o modelo que melhor representasse o tipo de ruído contido em um sinal de voz seria utilizado. É claro que para isso, seria necessário desenvolver classificadores de ruído eficientes.

Uma característica comum a todos os métodos de MV é a necessidade de estimar o perfil do ruído presente no sinal. Na verdade, quanto melhor for a estimação do ruído, maior será a qualidade do sinal processado (EPHRAIM; MALAH, 1984). De acordo com Yuan e Xia (2015), o projeto de algoritmos de MV geralmente não leva em consideração as diferenças em propriedades estatísticas dos diferentes tipos de ruído. Isto pode ser a causa do fraco desempenho de alguns algoritmos frente a condições específicas de ruído, como pode ser visto em Hu e Loizou (2007).

Neste sentido, alguns estudos recentes têm mostrado que o desenvolvimento de métodos específicos para cada tipo de ruído são apropriados, podendo fornecer resultados aprimorados. Em Xia e Bao (2014) e Yuan e Xia (2015), os autores usaram classificação de ruído para predizer o tipo de ruído de fundo em sinais de voz, para então propor métodos que atuam de uma maneira específica para cada tipo de ruído.

Tendo como base os fatos e argumentos apresentados nesta seção, verifica-se que futuras pesquisas em melhoramento de voz tendem a se focar no tratamento e compreensão do ruído de fundo. Além disso, o desenvolvimento de algoritmos que trabalham eficientemente em qualquer ambiente ruidoso, incorporando classificação de ruído, é uma tendência.

Considera-se nesta pesquisa apenas sinais de voz contaminados por ruídos presentes em situações reais. Dessa forma, apresenta-se um estudo que se aproxima o máximo possível das condições reais de operação de um sistema de MV. As condições de ambiente simuladas são aquelas costumeiras ao tipo de vida que a maioria das pessoas levam em tempos contemporâneos.

A metodologia proposta neste trabalho consiste em um processamento que leve em

consideração o tipo de ruído presente no sinal de voz. Para isso, uma avaliação dos principais métodos propostos pela literatura especializada sob condições de ruídos reais será realizada. Dentre os métodos escolhidos, utiliza-se desde os mais clássicos, os pioneiros, até os mais recentes. Além disso, serão considerados métodos que atuam tanto no domínio wavelet como no domínio da frequência. Após esta avaliação, serão identificados os métodos que melhor se adaptaram a determinados tipos de ruído, para então desenvolver um método robusto de MV.

A principal motivação para o desenvolvimento desta pesquisa é a proposta de novas ferramentas e metodologias direcionadas ao melhoramento de voz baseado na classificação do ruído de fundo. Neste sentido, é proposto um sistema para identificação e classificação de ruído baseado em um algoritmo biologicamente inspirado, com características fornecidas por wavelets complexas, que seja facilmente acoplado a outros sistemas de processamento de voz. Trata-se da aplicação de um sistema inteligente com o objetivo de contribuir para o desenvolvimento e o aprimoramento da área clássica de MV. Por isso, acrescenta-se por meio deste trabalho, dois conceitos novos à referida área de pesquisa: o conceito de Sistema Imunológico Artificial (SIA) e o conceito de wavelets complexas.

Com o desenvolvimento de um sistema capaz de identificar e classificar o ruído de fundo em um sinal de voz, além de permitir a escolha do melhor dentre vários métodos, possibilitará o aprimoramento e o desenvolvimento de novas técnicas para a redução e/ou estimação de ruído.

Como um produto secundário, investiga-se também nesta tese o uso de wavelets complexas para a redução de ruído em sinais de voz (ABREU; DUARTE; VILLARREAL, 2015). Salvo engano, não existem na literatura métodos de MV com tais características. O conceito de wavelets complexas vem sendo fortemente explorado no âmbito de processamento e análise de sinais. Como exemplo, citam-se os trabalhos realizados em Khare, Khare e Srivastava (2013), Das, Bhuiyan e Alam (2014), Abreu et al. (2014) e Abreu et al. (2015).

O restante do texto é organizado da seguinte forma: no Capítulo 2 são apresentados conceitos matemáticos fundamentais ao entendimento e desenvolvimento de métodos de MV; no Capítulo 3, apresentam-se os principais métodos de MV propostos na literatura, a análise do desempenho destes métodos frente a vários tipos de ruídos reais é realizada; no Capítulo 4 apresenta-se o sistema de identificação e classificação de ruídos reais proposto, bem como os resultados das simulações e comparações com classificadores comumente utilizados em problemas de reconhecimento de padrões; no Capítulo 5 é apresentado um algoritmo para estimação de ruído em sinais de voz que incorpora classificação de ruído. Além disso, ainda no Capítulo 5, uma nova metodologia para melhoramento de voz que une todas as ferramentas e ideias discutidas nesta tese é proposta.

Ao final deste trabalho, realizam-se as considerações finais e as sugestões para

trabalhos futuros. Logo após as considerações finais, apresentam-se cinco apêndices. O objetivo dos apêndices A, B e C é apresentar alguns conceitos biológicos adicionais, importantes ao entendimento do classificador de ruído apresentado no Capítulo 4. O apêndice D apresenta detalhes adicionais referentes às simulações realizadas no Capítulo 5. Já no apêndice E, apresenta-se a lista de produções resultantes desta pesquisa.

1.1 CONTRIBUIÇÕES DESTE TRABALHO

Embora a área na qual se insere esta pesquisa esteja bem consolidada e explorada, os pesquisadores relutam em apresentar trabalhos onde se avaliam conjuntamente métodos que atuam no domínio da frequência e métodos que atuam no domínio wavelet. Neste sentido, apresenta-se neste trabalho uma análise dos principais métodos de MV, os métodos clássicos. Para isso, considera-se tanto métodos baseados na análise de Fourier quanto em wavelets.

A análise do desempenho de diferentes métodos se faz frente a vários tipos de ruídos reais, por meio de medidas objetivas de qualidade. Dentre outros aspectos, constatou-se que o desempenho dos mesmos pode variar de acordo com o tipo do ruído de fundo, reforçando a tendência em desenvolver algoritmos de MV baseados na classificação do ruído. Sendo assim, no Capítulo 4 é proposto um sistema capaz de identificar e classificar vários tipos de ruídos reais presente em sinais de voz.

Baseado em um SIA, o classificador proposto utiliza um processo de extração de características que é baseado em uma análise multiescala fornecida por wavelets complexas. Dessa forma, contribui-se com a literatura inserindo alguns conceitos e técnicas que antes não faziam parte da mesma. Além disso, o classificador proposto apresenta algumas vantagens sobre os métodos propostos em Ma, Smith e Milner (2003), Xia e Bao (2014) e Yuan e Xia (2015). Dentre elas, destacam-se (ABREU; DUARTE; VILLARREAL, 2017):

- Uma única janela de silêncio (ausência de voz) é necessária para a inicialização do algoritmo;
- Identificação, em primeiro lugar, se o sinal de voz está limpo ou o nível de ruído é tão baixo que nenhum processamento (ex.: melhoramento) é necessário. Somente com a confirmação da presença de ruído, é que o módulo de classificação é acionado, diminuindo o custo computacional.
- Facilidade de acoplamento a outros sistemas de processamento de voz, como por exemplo, melhoramento, RAV, reconhecimento do orador, reconhecimento de emoções e aplicações que incorporam a classificação de cenários;

- Implementação simples, baseada principalmente em simples comparações entre padrões por meio de uma medida de afinidade (ex.: uma medida de distância). Ao contrário de classificadores como redes neurais e máquina de vetores de suporte (SVM - do inglês: *Support Vector Machine*), cujas implementações são baseadas em algoritmos de otimização.

Nas pesquisas apresentadas em Xia e Bao (2014) e Yuan e Xia (2015), são necessárias, respectivamente, 10 e 15 janelas de silêncio para a inicialização dos algoritmos. Além disso, em ambos os métodos, nenhuma distinção entre sinais limpos e contaminados foi feita. Sendo assim, é sempre considerado que os sinais em processamento estarão contaminados.

Outra contribuição desta pesquisa com a área de melhoramento de voz é a proposta de uma nova maneira de estimar o ruído. O que se almejou foi encontrar a resposta para a seguinte questão: "é possível realizar uma razoável estimação do ruído a partir de sinais de voz contaminados por meio de regressão?" Esta questão foi levantada pois, em períodos mais recentes, métodos de MV baseados em redes neurais têm sido explorados (XU et al., 2015; XU et al., 2014).

Nesse sentido, um algoritmo para estimação de ruído baseado no método de regressão por mínimos quadrados parciais foi proposto. Os resultados foram comparados com um algoritmo de estimação de ruído amplamente conhecido e reconhecido, denominado MCRA (do inglês: *Minima Controlled Recursive Averaging*). Inicialmente os resultados não superaram o algoritmo MCRA. Porém, ao incorporar o classificador de ruído desenvolvido, foi possível construir um modelo de regressão para cada tipo de ruído considerado e os resultados melhoraram consideravelmente. Avaliações objetivas de qualidade mostraram que a performance do algoritmo proposto foi ao menos igual ao algoritmo MCRA, em termos de redução de ruído, e superior em termos de qualidade geral dos sinais processados.

A principal vantagem de uma estimação de ruído baseada em regressão se concentra na redução do custo computacional. O ruído pode ser estimado por meio de um simples produto entre um segmento do sinal contaminado e a matriz de coeficientes de regressão, ao invés de se utilizar um complicado algoritmo de estimação de ruído. Assim, os benefícios dessa proposta refletem, principalmente, sobre aplicações de tempo real.

Baseado em todas as simulações e ferramentas desenvolvidas nesta pesquisa, uma nova metodologia para a área de MV pôde ser proposta. Denominada "Conjunto de Métodos de Melhoramento de Voz (CMMV)", a metodologia consiste em utilizar um banco de dados com sentenças contaminadas com vários tipos de ruído reais, onde vários métodos de MV são ajustados para cada tipo de ruído. Estes ajustes são feitos em modo *off-line*, e os melhores métodos para cada tipo de ruído são selecionados. Assim, durante

a operação em modo *off-line*, utiliza-se um classificador de ruído para predizer o tipo de ruído presente no sinal em processamento, para então escolher o melhor método dentro do CMMV construído.

Todas as ferramentas e metodologias apresentadas neste trabalho foram validadas a partir de exaustivas simulações e são contribuições importantes para o desenvolvimento da área clássica de melhoramento de voz.

No próximo capítulo, os fundamentos matemáticos que alicerçam o projeto de métodos de melhoramento de voz serão apresentados. É importante salientar que não será apresentada a teoria dos métodos subespaciais, pois esta classe de métodos é a menos explorada dentre as três classes, além de fugir ao escopo desta pesquisa, que se concentra nos métodos clássicos e amplamente explorados pela comunidade científica.

2 AS TRANSFORMADAS WAVELET E DE FOURIER COMO SUPORTE PARA O DESENVOLVIMENTO DE MÉTODOS DE MELHORAMENTO DE VOZ

No universo do processamento digital de sinais, um fenômeno físico é modelado por meio de uma função, que é chamada de sinal (OPPENHEIM; SCHAFER; BUCK, 1998; GOMES; VELHO; GOLDSTEIN, 1997). Existem, porém, vários tipos de sinais do universo físico cuja modelagem via funções conhecidas torna-se um processo muito complicado, quase impossível. Isto se deve ao fato destes sinais possuírem um alto teor de complexidade, o que reforça ainda mais a necessidade de compreendê-los. A análise desses sinais deve ser feita a partir de ferramentas matemáticas capazes de evidenciar suas principais propriedades. Dentre elas, destacam-se as transformadas integrais (BLATTER, 1998).

As transformadas mais utilizadas no contexto do processamento de sinais de voz são a transformada de Fourier (FIGUEIREDO, 1997; OPPENHEIM; SCHAFER; BUCK, 1998; DELLER; PROAKIS; HANSEN, 1993) e a transformada wavelet (DAUBECHIES, 1992; STRANG; NGUYEN, 1996; MALLAT, 1998).

Neste capítulo, objetiva-se realizar uma breve revisão acerca dos fundamentos da análise de Fourier e da análise wavelet necessários ao entendimento desta tese.

2.1 ANÁLISE DE FOURIER

A análise de Fourier permite fazer uma expansão de funções em termos de polinômios trigonométricos. Em outras palavras, ela permite a decomposição de uma função em uma soma infinita de funções seno e cosseno. A partir daí, torna-se possível uma análise do conteúdo de frequência da função em questão (OPPENHEIM; SCHAFER; BUCK, 1998).

Seja $f : \mathbb{R} \to \mathbb{R}$ uma função com período $2L$. Segundo a teoria de Fourier, a função f pode ser escrita na forma (FIGUEIREDO, 1997):

$$f(t) \sim \frac{1}{2}a_0 + \sum_{n=1}^{+\infty}\left(a_n cos\frac{n\pi t}{L} + b_n sen\frac{n\pi t}{L}\right). \tag{1}$$

A expressão do lado direito de (1) é dita Série de Fourier da função f. Os termos a_0, a_n e b_n são denominados coeficientes de Fourier, sendo $a_0 = \frac{1}{L}\int_{-L}^{L} f(t)dt$, a_n e b_n são calculados conforme equações (2) e (3) (FIGUEIREDO, 1997):

$$a_n = \frac{1}{L}\int_{-L}^{L} f(t)cos\frac{n\pi t}{L}dt, n \geq 0 \tag{2}$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(t) sen \frac{n\pi t}{L} dt, n \geq 1. \quad (3)$$

Sabe-se que a série de Fourier existe para ser aplicada às funções periódicas. Porém, a maior parte dos sinais de interesse prático não possui período. Na verdade esses sinais, em sua maioria, nem mesmo são estacionários. Sendo assim, nem sempre será possível decompor f como uma soma de funções com frequências bem definidas como em (1). Dessa forma, precisa-se lançar mão de uma ferramenta mais adequada, a transformada de Fourier:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt, \quad (4)$$

onde $\hat{f}(\omega)$ é dita a transformada de Fourier da função $f(t)$ e $j = \sqrt{-1}$.

Analisando a equação (4), a função exponencial é chamada função moduladora ou núcleo da transformada de Fourier. Ela representa uma função periódica de frequência angular ω. Quando f possuir oscilações de frequência ω, essas frequências entram em ressonância com a frequência da função moduladora e $\hat{f}(\omega)$ possuirá valores não nulos. Em outras palavras, $\hat{f}(\omega)$ mensura a ocorrência de frequência ω na função f (GOMES; VELHO; GOLDSTEIN, 1997).

A transformada de Fourier é invertível, ou seja, é possível obter a função f que foi decomposta em (4) da seguinte forma (BLATTER, 1998):

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega) e^{j\omega t} d\omega. \quad (5)$$

A expressão em (5) é denominada transformada inversa de Fourier.

Apesar de ser uma ferramenta poderosa para análise de sinais, a transformada de Fourier pode não ser tão eficiente quando o sinal em questão é não estacionário (OPPENHEIM; SCHAFER; BUCK, 1998). Este fato reflete também sobre a análise de sinais de voz. Neste sentido, é de praxe dividir um sinal de voz em pequenos segmentos por meio da técnica de janelamento, para, então, aplicar a equação (4) sobre cada segmento separadamente. Considera-se que, em pequenos segmentos, o sinal de voz contaminado possua características estacionárias.

Os métodos de melhoramento de voz baseados na equação (4), empregam o algoritmo FFT (do inglês - *Fast Fourier Transform*) para implementar uma transformada de Fourier discreta no tempo (DFT - do inglês: *Discrete Fourier Transform*) (OPPENHEIM; SCHAFER; BUCK, 1998). É comum se referir a estes métodos como aqueles que atuam no domínio da frequência.

Existem várias técnicas de redução de ruído em sinais de voz que atuam no domínio da frequência. Dentre elas, destacam-se: subtração espectral, filtragem de WIENER, métodos baseados em uma máscara binária, métodos baseados na minimização do erro

quadrático médio (MMSE), entre outros. O principal objetivo destes métodos é alterar o espectro de frequência do sinal de voz ruidoso a fim de reduzir o ruído de fundo. Para isso, é necessária uma estimação do espectro de frequência do ruído, de forma que o filtro projetado não cause danos ao espectro de frequência da voz. No Capítulo 3, os conceitos teóricos e práticos de tais métodos serão discutidos e analisados por meio de várias simulações.

2.2 ANÁLISE WAVELET

Em processamento de sinais, a análise wavelet fornece um processo de representação/decomposição de um sinal por meio de uma base de funções que oscilam localmente, essas funções são denominadas wavelets. Em outras palavras, as funções wavelet possuem suporte compacto, ao contrário da análise de Fourier que utiliza um conjunto de funções bases que oscilam infinitamente (as funções seno e cosseno). Uma base de funções wavelet é constituída por dilatações e translações de uma função wavelet real $\psi(t)$ da seguinte forma:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right). \tag{6}$$

Em (6), a função $\psi(t)$ é denominada wavelet mãe, enquanto que a e b são os fatores de escala e translação, respectivamente.

Combinando adequadamente as funções em (6) com translações de uma função escala $\phi(t)$, realiza-se uma expansão do sinal através de uma base ortonormal de funções. Tal base fornece uma análise tempo-frequência do sinal (DAUBECHIES, 1992; LINA; MAYRAND, 1993; SELESNICK; BARANIUK; KINGSBURY, 2005). Sendo assim, qualquer sinal analógico de energia limitada $f(t)$ pode ser expresso em termos de funções wavelet e funções escala da seguinte forma (SELESNICK; BARANIUK; KINGSBURY, 2005):

$$f(t) = \sum_{n=-\infty}^{+\infty} c(n)\phi(t-n) + \sum_{k=0}^{+\infty}\sum_{n=-\infty}^{+\infty} d(k,n)2^{k/2}\psi(2^k t - n), \tag{7}$$

sendo $c(n)$ os coeficientes da função escala e $d(k,n)$ os coeficientes wavelet. Ambos são calculados via produto interno (SELESNICK; BARANIUK; KINGSBURY, 2005):

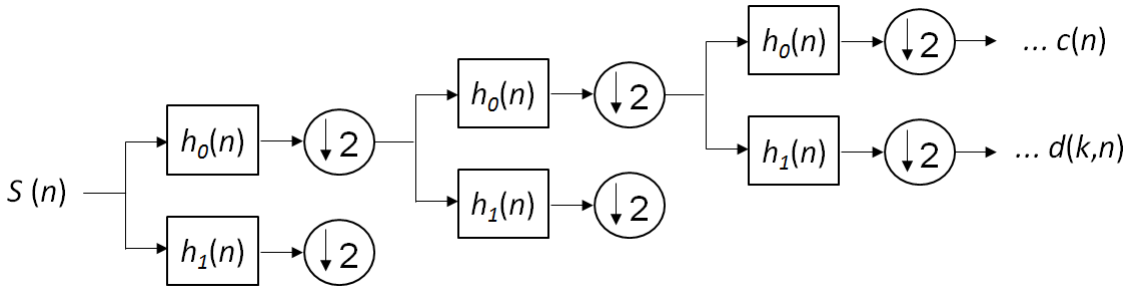
$$c(n) = \int_{-\infty}^{+\infty} f(t)\phi(t-n)dt, \tag{8}$$

$$d(k,n) = 2^{k/2}\int_{-\infty}^{+\infty} f(t)\psi(2^k t - n)dt. \tag{9}$$

Juntas, as equações (7), (8) e (9) fornecem uma análise tempo-frequência do sinal controlado pelo fator de escala k no tempo n.

Existem algoritmos para calcular a função escala e a função wavelet que, na prática, são representadas por um filtro passa-baixa e um filtro passa-alta. Mallat (1998) propôs um algoritmo eficiente para o cálculo dos coeficientes $c(n)$ e $d(k,n)$. Denominado Transformada Wavelet Rápida, este método utiliza um banco de filtros digitais em uma estrutura de árvore que aplica recursivamente um filtro passa-baixa h_0 e um filtro passa-alta h_1, seguidos por operadores de subamostragem (*downsampling*) e de sobreamostragem (*upsampling*) (STRANG; NGUYEN, 1996). Este método é amplamente utilizado para implementar a transformada wavelet discreta (DWT - do inglês: *Discrete Wavelet Transform*). Na Figura 1, ilustra-se o banco de filtros de análise para a DWT, onde $S(n)$ representa um sinal qualquer discretizado.

Figura 1 – Banco de filtros de análise para a DWT.



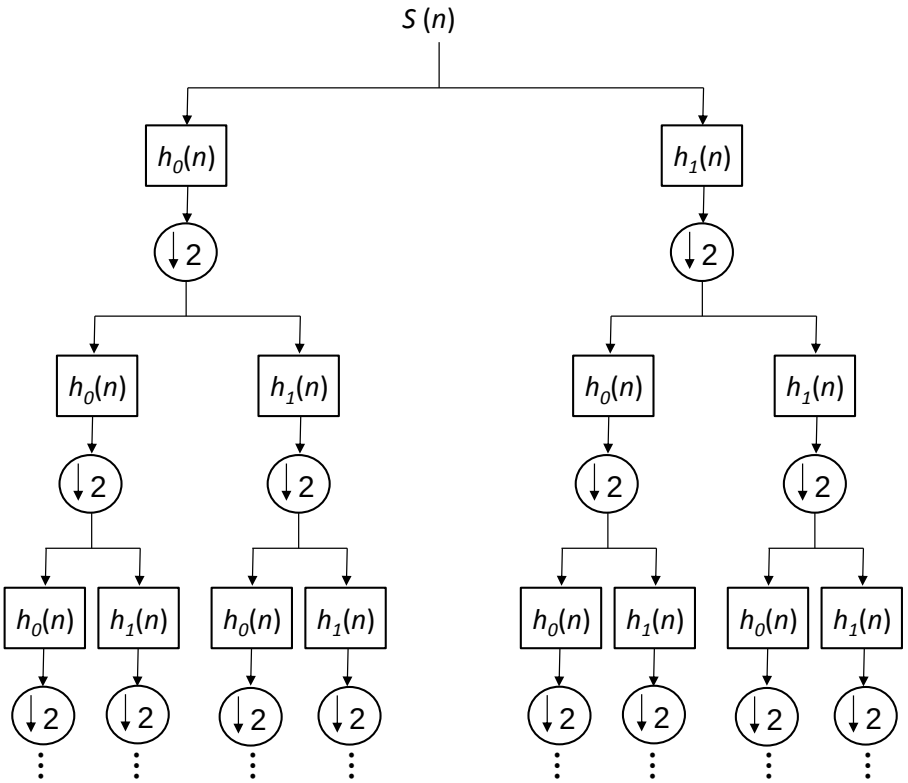
Adaptado de Mallat (1998).

É importante salientar que estes filtros possibilitam a formulação de parametrizações para o projeto de wavelets e funções escala com algumas propriedades interessantes, como por exemplo, suporte compacto e momentos nulos (SELESNICK; BARANIUK; KINGSBURY, 2005; UZINSKI et al., 2013). Para cada nível de decomposição wavelet, os coeficientes de aproximação $c(n)$ e detalhes $d(k,n)$ do sinal no nível $k(k = 1, 2, 3, ...)$ são adquiridos. O nível de decomposição (escala) k, está associado com a quantidade de faixas de frequências obtidas. Quanto maior for k, maior será o número de faixas de frequências, capturando mais detalhes do sinal e minimizando possíveis perdas de informação (STRANG; NGUYEN, 1996; DUARTE, 2005). Isto é possível, pois os coeficientes de detalhes $d(k,n)$ correspondem a frequência situada aproximadamente entre $(2^{-k}f_s, 2^{-k-1}f_s)$, onde f_s é a taxa de amostragem do sinal $S(n)$ (LINA; MAYRAND, 1993).

A DWT possui algumas propriedades desejáveis para o processamento de sinais de voz, são elas: reconstrução perfeita através de filtros com suporte compacto, baixo custo computacional, concentração da energia do sinal transformado em um pequeno número de coeficientes e a obtenção de diferentes sub-bandas de frequência (DUARTE, 2005; ABREU, 2013).

Uma generalização para o conceito de DWT é a denominada transformada wavelet packet (DWPT - do inglês: *Discrete Wavelet Packet Transform*). No processo de decomposição exposto na Figura 1, após o primeiro nível de decomposição ($k = 1$), obtêm-se os coeficientes de aproximação e de detalhes. Para obter o segundo nível, repete-se o mesmo procedimento, porém, apenas sobre os coeficientes de aproximação. Dessa forma, apenas os coeficientes de aproximação são decompostos em aproximação e detalhes. Em contrapartida à análise wavelet, na análise wavelet packet, tanto detalhes como aproximações podem ser decompostos. O banco de filtros de análise para a DWPT é apresentado na Figura 2.

Figura 2 – Banco de filtros de análise para a DWPT.



Adaptado de Mallat (1998).

A principal diferença entre a DWT e a DWPT consiste no número de sub-bandas de frequências obtidas para cada nível k. Note pelas Figuras 1 e 2 que a DWT fornece 2^{k-1} sub-bandas de frequências, enquanto que a DWPT fornece 2^k sub-bandas.

Verifica-se nos trabalhos de Ayat, Manzuri e Dianat (2004), Ghanbari e Karami-Mollaei (2006), Bahoura e Rouat (2006), Soares et al. (2011), que tanto a DWT como a

DWPT são extensivamente utilizadas no problema do melhoramento de voz. A principal motivação está na capacidade de evidenciar o ruído presente no sinal, tendo em vista que ambas realizam uma distinção entre baixas e altas frequências.

Apesar das boas propriedades citadas sobre a DWT e a DWPT, Selesnick, Baraniuk e Kingsbury (2005) destacam alguns inconvenientes:

- Wavelets possuem características passa-banda e seus coeficientes oscilam positiva e negativamente em torno de singularidades. Este fato torna a extração de singularidades ou características do sinal um processo delicado;
- A DWT é variante ao deslocamento; isto significa que um pequeno deslocamento no sinal gera distúrbios no padrão de oscilação dos coeficientes wavelet em torno de singularidades (veja Figura 5). Neste sentido, algoritmos que atuam no domínio wavelet devem ser capazes de lidar com tais características;
- Para dimensões maiores que um, os coeficientes wavelet produzem um padrão que é simultaneamente orientado em várias direções. Essa falta de seletividade direcional pode ser um problema em processamento de imagens (SELESNICK; BARANIUK; KINGSBURY, 2005).

Segundo Selesnick, Baraniuk e Kingsbury (2005), a busca pela solução destes problemas se inicia observando o fato de que a transformada de Fourier não sofre dos mesmos. A análise fornecida pela transformada de Fourier é baseada sobre senoides de valores complexos:

$$e^{j\Omega t} = \cos(\Omega t) + j\,\mathrm{sen}(\Omega t). \tag{10}$$

A parte real (cosseno) e a parte imaginária (seno) formam um par de transformada de Hilbert que produz um sinal analítico $e^{j\Omega t}$. Este sinal possui suporte apenas sobre metade do eixo de frequência ($\Omega > 0$). A magnitude da transformada de Fourier não oscila positiva e negativamente em torno de singularidades. Além disso, ela é perfeitamente invariante ao deslocamento (OPPENHEIM; SCHAFER; BUCK, 1998; SELESNICK; BARANIUK; KINGSBURY, 2005).

Com base nos argumentos acerca da teoria de Fourier e lembrando que tanto a DWT quanto a DWPT são baseadas em funções wavelet reais, Kingsbury (1998) introduziu uma arquitetura de fácil implementação, que culminou com a ampla disseminação da transformada wavelet complexa (ℂWT - do inglês: *Complex Wavelet Transform*). A ℂWT é baseada em funções escala e funções wavelet de valores complexos, apresentadas, respectivamente, nas equações (11) e (12).

$$\phi_c(t) = \phi_r(t) + j\phi_i(t), \tag{11}$$

$$\psi_c(t) = \psi_r(t) + j\psi_i(t). \tag{12}$$

Analogamente à equação (10), ψ_r é real e par, enquanto que ψ_i é imaginária e ímpar. Juntas, elas formam um par de transformadas de Hilbert, o que torna ψ_c um sinal analítico (SELESNICK; BARANIUK; KINGSBURY, 2005). A função ϕ_c é definida de maneira análoga.

Após a decomposição, os coeficientes wavelet obtidos são definidos da seguinte forma:

$$d_c(k,n) = d_r(k,n) + jd_i(k,n). \tag{13}$$

A partir da equação (13), algoritmos baseados em wavelets podem explorar a magnitude, equação (14), e a fase, equação (15), dos coeficientes wavelet complexos:

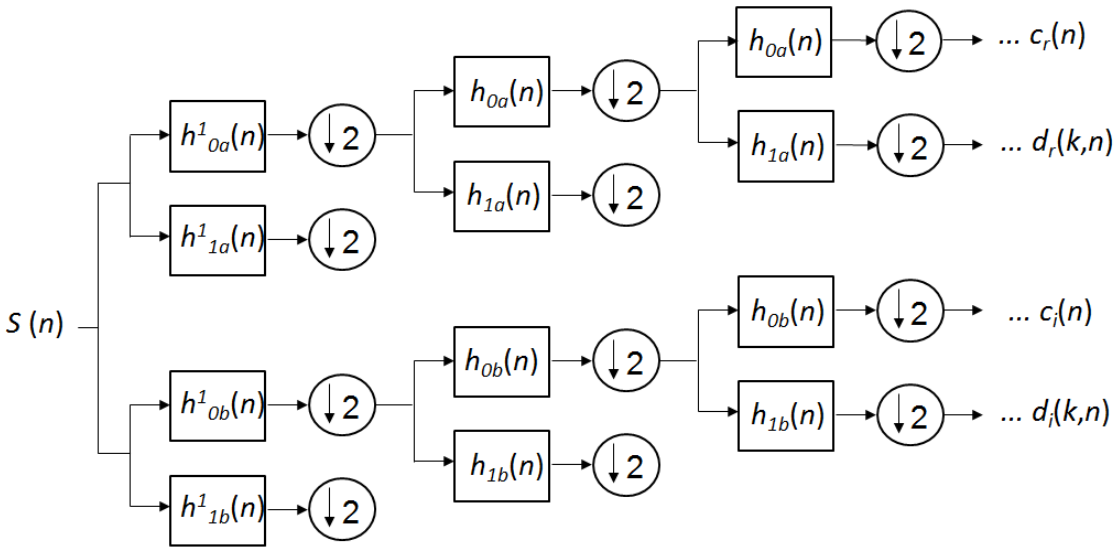
$$|d_c(k,n)| = \sqrt{[d_r(k,n)]^2 + [d_i(k,n)]^2}, \tag{14}$$

$$\angle d_c(k,n) = \arctan\left(\frac{d_i(k,n)}{d_r(k,n)}\right), \tag{15}$$

onde $|d_c(k,n)| \geq 0$.

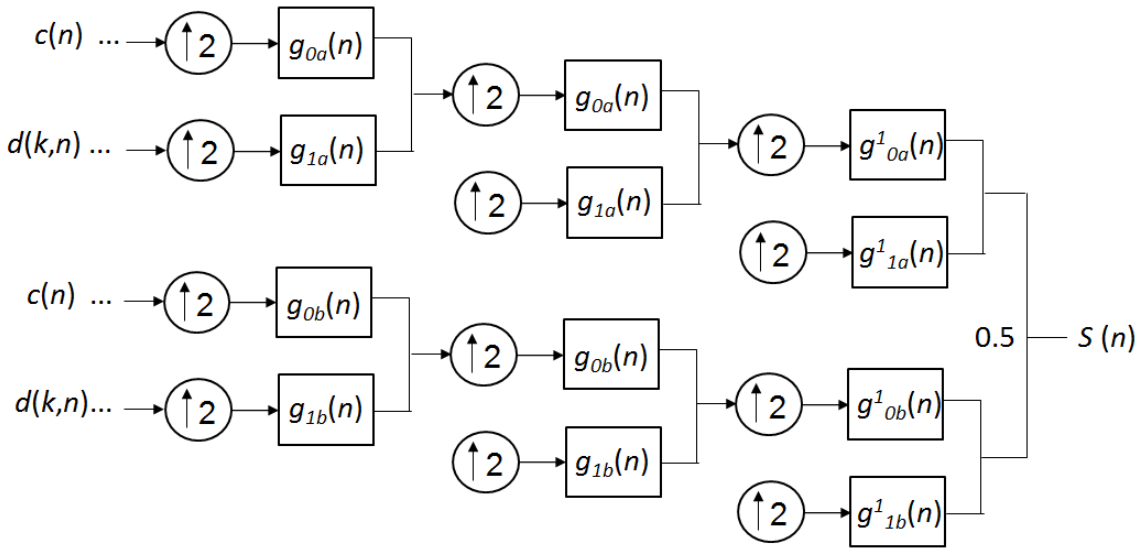
A fim de implementar a $\mathbb{C}$WT e alcançar propriedades como invariância ao deslocamento e seletividade direcional em dimensões superiores, Kingsbury (1998) propôs a arquitetura Dual−Tree. Denominada DT−$\mathbb{C}$WT(do inglês - *Dual− Tree Complex Wavelet Transform*), esta arquitetura utiliza dois bancos de filtros, como o da Figura 1, a fim de calcular a parte real $d_r(k,n)$ e a parte imaginária $d_i(k,n)$ dos coeficientes wavelet. Em outras palavras, a DT−$\mathbb{C}$WT utiliza duas DWT, onde a árvore superior produz a parte real e a árvore inferior produz a parte imaginária dos coeficientes complexos. Os bancos de filtros para análise e síntese na DT−$\mathbb{C}$WT são mostrados nas Figuras 3 e 4.

Figura 3 – Banco de filtros de análise para a DT−$\mathbb{C}$WT



Fonte: Adaptado de Selesnick, Baraniuk e Kingsbury (2005).

Figura 4 – Banco de filtros de síntese para a DT−ℂWT

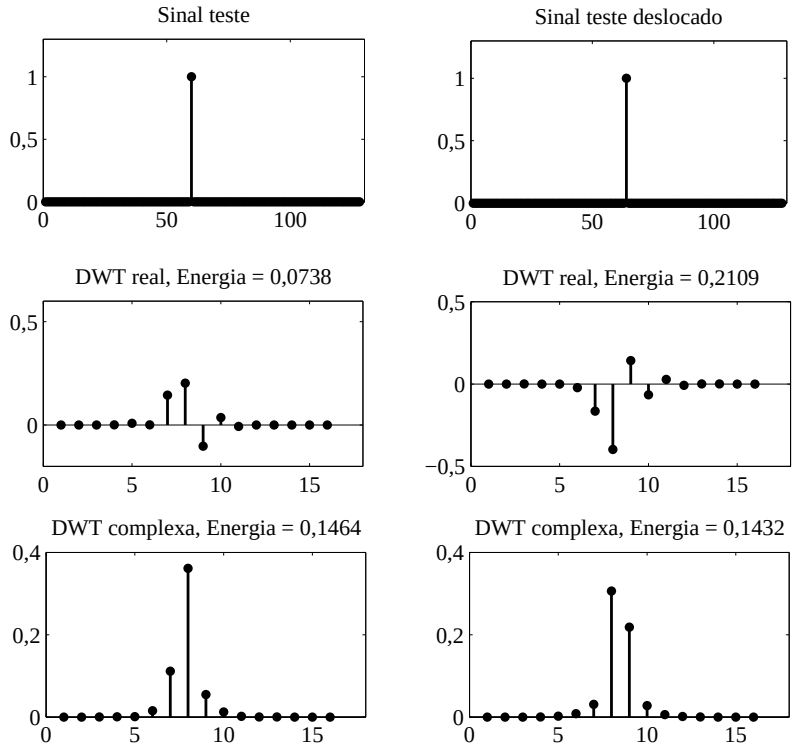


Fonte: Adaptado de Selesnick, Baraniuk e Kingsbury (2005).

Vale ressaltar que cada DWT é composta por dois conjuntos diferentes de filtros reais, onde cada conjunto satisfaz às condições de reconstrução perfeita (PR). Para $k > 1$, os filtros da árvore b são os reversos da árvore a, e os filtros de reconstrução são os reversos dos filtros de análise. Assim, todos os filtros fazem parte do mesmo conjunto ortonormal (SELESNICK; BARANIUK; KINGSBURY, 2005). A única exceção ocorre no primeiro estágio ($k = 1$), onde qualquer conjunto de filtros que satisfaçam a condição de PR podem ser utilizados. Além disso, pode-se utilizar o mesmo par de filtros para ambas as árvores (SELESNICK; BARANIUK; KINGSBURY, 2005). Esta arquitetura foi desenvolvida de forma que a transformada total seja aproximadamente analítica, e assim, $\psi_i \approx \mathcal{H}\{\psi_r\}$. A DT−ℂWT possui a vantagem de ser quase invariante ao deslocamento e possuir boa seletividade direcional.

Uma comparação entre a DWT e a DT−ℂWT com respeito a variança ao deslocamento é apresentada na Figura 5. Para um impulso $f(n) = \delta(n - 60)$ e sua versão deslocada $f_d(n) = \delta(n - 65)$ (a), segue os coeficientes wavelet em uma escala fixa k (b) e (c). Em (b) constam os coeficientes wavelet reais calculados usando a DWT convencional com filtros de Daubechies de comprimento 14 (db7). Já em (c), segue a magnitude dos coeficientes wavelet complexos calculados pela DT−ℂWT utilizando filtros de mesmo comprimento propostos por Kingsbury (2003).

Figura 5 – Quase invariância ao deslocamento da DT−ℂWT



Fonte: Adaptado de Selesnick, Baraniuk e Kingsbury (2005).

Analisando os resultados da decomposição fornecida pela DT−ℂWT, apresentados na Figura 5, verifica-se que além da forma de onda dos sinais teste e teste deslocado sofrerem perturbações semelhantes, a energia dos coeficientes wavelet sofreu uma alteração muito pequena, sendo quase constante. Quando comparados, fica claro um desempenho superior por parte da DT−ℂWT sobre a DWT na análise em questão.

Segundo Selesnick, Baraniuk e Kingsbury (2005), os filtros existentes e frequentemente utilizados para a DWT não podem ser utilizados para implementar ambas as árvores de decomposição da DT−ℂWT. Isso se deve ao fato de que estes filtros não irão satisfazer à condição $\psi_i \approx \mathcal{H}\{\psi_r\}$, e assim, a transformada não irá oferecer as vantagens de wavelets analíticas. Neste sentido, novos conjuntos de filtros devem ser projetados levando em conta alguns requerimentos adicionais aos da DWT. Detalhes sobre o projeto de filtros para a DT−ℂWT podem ser encontrados em Selesnick, Baraniuk e Kingsbury (2005) e Kingsbury (2003).

Salvo engano, não há na literatura algum método de melhoramento de voz que utiliza a DT−ℂWT como uma ferramenta para decompor o sinal. Vale ressaltar que essa abordagem é relativamente recente e, atualmente, vem sendo muito explorada por pesquisadores da área de processamento de sinais. Como exemplo, citam-se os traba-

lhos: Kingsbury (2001), Zhang et al. (2010), Khare, Khare e Srivastava (2013), Mitiche, Adamou-Mitiche e Naimi (2013) e Das, Bhuiyan e Alam (2014).

Por fim, com base no que foi dito nos parágrafos anteriores, conclui-se que:

- A transformada de Fourier fornece uma análise do conteúdo de frequência do sinal. Desta forma, os métodos que utilizam a DFT alteram o espectro de frequência do sinal de voz ruidoso a fim de reduzir o ruído de fundo. Para isso, é de suma importância que uma boa estimação do espectro de frequência do ruído seja realizada, de forma que o filtro projetado não cause danos ao espectro de frequência da voz.

- Utilizando a DWT, DWPT ou a DT$-\mathbb{C}$WT, uma análise tempo-frequência do sinal de voz é realizada. Nestes casos, o conteúdo de frequência do sinal é distribuído pelas escalas wavelet k, seguindo a relação $(2^{-k}f_s, 2^{-k-1}f_s)$. Esta relação fornece k bandas de oitava, tendo como referência a frequência de amostragem do sinal. Vale ressaltar que o domínio em que ocorre o processamento não é o domínio da frequência e sim o domínio tempo-frequência, no contexto, também denominado como domínio wavelet.

No próximo capítulo apresenta-se uma análise dos principais métodos de MV, cujas implementações se baseiam tanto na teoria wavelet, como na teoria de Fourier.

3 UMA ANÁLISE OBJETIVA DOS PRINCIPAIS MÉTODOS DE MELHORAMENTO DE VOZ QUANDO SUBMETIDOS A AMBIENTES RUIDOSOS REAIS

Como mencionado no Capítulo 1, grande parte dos métodos de melhoramento de voz foi desenvolvida com o pressuposto de que o ruído de fundo é o Gaussiano Branco. Isto se justifica, pois o RGB possui propriedades conhecidas, favorecendo o desenvolvimento de operações analíticas para se estimar, por exemplo, a função densidade de probabilidade do ruído como em Tabibian, Akbari e Nasersharif (2015).

Apresenta-se neste capítulo uma breve descrição dos métodos considerados para implementação e avaliação. O objetivo deste capítulo é simular situações reais de operação de sistemas de MV. Os métodos foram escolhidos de acordo com a teoria clássica da área, ou seja, consideram-se desde os primeiros métodos propostos na literatura até os métodos mais atuais. Sendo assim, serão implementados e avaliados tanto os métodos que atuam no domínio da frequência, quanto os que atuam no domínio wavelet.

Além da análise realizada, de um ponto de vista global, entre os nove algoritmos considerados, pode-se destacar alguns objetivos específicos:

1. Investigar a limiarização wavelet tradicional e os atuais métodos wavelet não limiar sob condições de ruído real;
2. Confrontar os resultados obtidos no item anterior com os resultados obtidos pelos métodos que atuam no domínio da frequência;
3. Propor e avaliar um método wavelet não limiar baseado na DT$-\mathbb{C}$WT;
4. Identificar, entre os métodos considerados, aqueles que são mais adequados para cada tipo de ruído.

Devido à grande variedade de métodos de MV disponíveis na literatura, alguns trabalhos têm apresentado estudos comparativos a fim de estabelecer métodos ou classes de métodos que apresentam o melhor desempenho. Neste sentido, em Hu e Loizou (2007) são apresentadas comparações subjetivas englobando as classes de métodos destacadas no Capítulo 1. Na referida pesquisa, os autores concluíram que métodos baseados em modelos estatísticos tiveram o melhor desempenho em termos de qualidade geral e distorções de fala do sinal processado. Apenas um método baseado em subtração espectral obteve resultados semelhantes sob as mesmas condições. Vale destacar que neste estudo, os autores não utilizaram nenhum método baseado puramente em wavelets. Neste sentido, as análises apresentadas neste capítulo têm como objetivo preencher algumas lacunas e apresentar o estado da arte em MV.

3.1 SUBTRAÇÃO ESPECTRAL

Considerando os sinais de voz e ruído como processos aleatórios estacionários e independentes, Boll (1979) propôs uma das primeiras técnicas viáveis para implementação prática. Baseada na subtração espectral (SE), seu principal atrativo é a simplicidade matemática e o baixo custo computacional (VIEIRA FILHO, 1996).

Em problemas de melhoramento de voz, assume-se que o sinal $y[n]$, contaminado por ruído, seja descrito como segue:

$$y[n] = x[n] + w[n], \tag{16}$$

sendo que $x[n]$ representa o sinal de voz puro e $w[n]$ o ruído aditivo. Utilizando a técnica de janelamento e aplicando a DFT sobre cada segmento, escreve-se (16) no domínio da frequência da seguinte forma:

$$Y[l,n] = X[l,n] + W[l,n], \tag{17}$$

sendo n o $n-$ésimo componente espectral obtido pela transformada de Fourier relacionado com a janela $l(l = 0, 1, \ldots, L)$. Na equação (17), $Y[l,n]$, $X[l,n]$ e $W[l,n]$ podem ser escritos na forma polar, como segue:

$$Y[l,n] = Y_{l,n}\exp(j\theta_{Y_{l,n}}), \tag{18}$$

$$X[l,n] = X_{l,n}\exp(j\theta_{X_{l,n}}), \tag{19}$$

$$W[l,n] = W_{l,n}\exp(j\theta_{W_{l,n}}). \tag{20}$$

Sendo $Y_{l,n} = |Y[l,n]|$, $X_{l,n} = |X[l,n]|$, $W_{l,n} = |W[l,n]|$, $\theta_{Y_{l,n}} = \angle Y[l,n]$, $\theta_{X_{l,n}} = \angle X[l,n]$ e $\theta_{W_{l,n}} = \angle W[l,n]$.

A técnica baseada em SE consiste em estimar a magnitude do sinal puro $X_{l,n}$ da seguinte forma (BOLL, 1979):

$$\widehat{X}_{l,n} = Y_{l,n} - \widehat{W}_{l,n}. \tag{21}$$

Note em (21) que esta técnica necessita de uma estimação do espectro do ruído. Tal estimação pode ser realizada durante intervalos de silêncio (ausência de voz). Por fim, o sinal de voz melhorado no domínio da frequência é dado por:

$$\widehat{X}[l,n] = \widehat{X}_{l,n}\exp(j\theta_{Y_{l,n}}), \tag{22}$$

sendo $\widehat{X}[l,n]$ a estimação para $X[l,n]$ (é comum representar esta estimação pelo operador esperança $\widehat{X}[l,n] = E\{X[l,n]\}$).

Um comentário pertinente sobre a subtração espectral é que não há uma recuperação da fase do sinal contaminado. Apesar de alguns estudos mostrarem que o ouvido

humano não discerne bem as distorções de fase, alguns trabalhos mais atuais já propõem tal recuperação, veja Paliwal, Wójcicki e Shannon (2011).

A implementação da SE realizada neste trabalho consiste na estimação do espectro do ruído durante intervalos de silêncio. Além disso, tal estimação é atualizada sempre que uma nova janela de silêncio é detectada:

$$\widehat{W}_{l,n} = \beta.\widehat{W}_{l-1,n} + (1-\beta).\widehat{W}_{l,n}. \tag{23}$$

Trata-se de um método recursivo onde β é responsável por suavizar a estimação de $W_{l,n}$. Valores típicos para β estão no intervalo $[0,89, 0,98]$, podendo ser definido empiricamente pelo usuário (VIEIRA FILHO, 1996).

É evidente que a equação (21) pode gerar valores negativos. Em outras palavras, o método de SE só é realizável se a potência estimada do ruído for menor ou igual à potência do sinal ruidoso. A maneira mais adotada para solucionar este problema denomina-se "retificação de meia onda":

$$\widehat{X}_{l,n} = \begin{cases} Y_{l,n} - \widehat{W}_{l,n}, & \text{se } Y_{l,n} \geq \widehat{W}_{l,n} \\ 0, & \text{caso contrário .} \end{cases} \tag{24}$$

A retificação de meia onda consiste no maior inconveniente da SE. Isto se deve ao fato de (24) gerar uma alteração repentina de amplitudes em frequências onde a segunda condição é satisfeita. O resultado dessa operação é a geração de tons indesejáveis no sinal processado, o ruído Musical (VIEIRA FILHO, 1996).

3.2 SUBTRAÇÃO ESPECTRAL EM POTÊNCIA

A técnica descrita na seção 3.1 é também denominada de subtração espectral em amplitude. Elevando ao quadrado $\widehat{X}_{l,n}$, $Y_{l,n}$ e $\widehat{W}_{l,n}$ na equação (21), segue que

$$\widehat{X}_{l,n} = \sqrt{Y_{l,n}^2 - \widehat{W}_{l,n}^2}. \tag{25}$$

A equação (25) caracteriza o método de subtração espectral em potência (SEP). De maneira análoga à SE, o sinal melhorado no domínio da frequência é adquirido pela substituição da magnitude do sinal ruidoso pela magnitude estimada, conforme (22). No caso da SEP, também se faz necessário a correção de possíveis valores negativos em (25):

$$\widehat{X}_{l,n} = \begin{cases} \sqrt{Y_{l,n}^2 - \widehat{W}_{l,n}^2}, & \text{se } Y_{l,n}^2 \geq \widehat{W}_{l,n}^2 \\ 0, & \text{caso contrário .} \end{cases} \tag{26}$$

Durante a implementação do método baseado na SEP, o espectro de potência do ruído é estimado de maneira análoga ao da seção 3.1:

$$\widehat{W}_{l,n}^2 = \beta.\widehat{W}_{l-1,n}^2 + (1-\beta).\widehat{W}_{l,n}^2. \tag{27}$$

3.3 FILTRAGEM DE WIENER

O filtro de Wiener é baseado no princípio da minimização do erro quadrático médio, e parte do pressuposto de que se tem disponível o espectro do sinal puro e do ruído. Detalhes sobre a estimação realizada a partir do princípio MMSE podem ser encontrados em Ephraim e Malah (1984).

Partindo da equação (17) e considerando as mesmas notações utilizadas nas seções 3.1 e 3.2, segue que a magnitude estimada do sinal puro é resultante de uma operação de filtragem da seguinte forma:

$$\widehat{X}_{l,n} = H_{l,n} Y_{l,n}. \tag{28}$$

Denominado função ganho, $H_{l,n}$ representa um filtro utilizado na redução do ruído.

Considerando que os espectros de potência estimados do sinal puro e do ruído estejam disponíveis, segue que o filtro de Wiener H^w é definido como (VIEIRA FILHO, 1996):

$$H^w_{l,n} = \frac{\phi^x_{l,n}}{\phi^x_{l,n} + \phi^w_{l,n}}, \tag{29}$$

onde $\phi^x_{l,n}$ e $\phi^w_{l,n}$ são as densidades espectrais de potência do sinal de voz puro e do ruído, respectivamente.

Observa-se que para a realização do filtro de Wiener em (29) é necessária uma estimação para o sinal puro. Uma alternativa seria obter esta estimação a partir de (24) ou de (26) (SE ou SEP).

Em Scalart e Vieira Filho (1996), os autores mostraram que é possível unificar os métodos clássicos de melhoramento de voz utilizando o conceito de SNR_{post} (Relação Sinal Ruído *a Posteriori*) e de SNR_{prio} (Relação Sinal Ruído *a Priori*), introduzidos inicialmente por McAulay e Malpass (1980). Além de desenvolver os equacionamentos baseados nestes dois conceitos, os autores mostraram que o método pode ser uma solução para o surgimento do ruído Musical.

O operador SNR_{post}, denotado por γ, e SNR_{prio}, denotado por ξ, são definidos nas equações (30) e (31), respectivamente (SCALART; VIEIRA FILHO, 1996):

$$\gamma[l,n] = \frac{Y^2_{l,n}}{E\left\{W^2_{l,n}\right\}}, \tag{30}$$

$$\xi[l,n] = \frac{E\left\{X^2_{l,n}\right\}}{E\left\{W^2_{l,n}\right\}}, \tag{31}$$

sendo $E\left\{W^2_{l,n}\right\}$ o espectro de potência do ruído estimado e $E\left\{X^2_{l,n}\right\}$, o espectro de potência do sinal puro estimado.

Note que a estimação de ξ também passa pela estimação do sinal puro. A maneira mais simples de aproximar o cálculo de ξ, denominada ML (do inglês: *Maximum Likelihood estimator*), é definida na equação (32) (VIEIRA FILHO, 1996):

$$\hat{\xi}[l,n] = \hat{\gamma}[l,n] - 1, \tag{32}$$

sendo $\hat{\gamma}$ a SNR_{post} estimada.

Neste trabalho, a implementação do filtro de Wiener é baseada na SNR_{prio}. Segundo Scalart e Vieira Filho (1996), a equação (29) pode ser escrita em termos da SNR_{prio} da seguinte forma:

$$H_{l,n}^{w} = \frac{\hat{\xi}[l,n]}{\hat{\xi}[l,n]+1}. \tag{33}$$

A fim de obter uma melhor estimação para ξ do que a fornecida pelo método ML, utiliza-se nas implementações o método denominado decisão direta (EPHRAIM; MALAH, 1984):

$$\hat{\xi}[l,n] = \beta \frac{X_{l-1,n}^2}{E\left\{W_{l,n}^2\right\}} + (1-\beta)P\left[\hat{\gamma}[l,n]-1\right], \tag{34}$$

onde $X_{l-1,n}^2$ corresponde ao espectro de potência do sinal processado na janela precedente, e o operador $P[.]$ corresponde a retificação de meia onda.

3.4 ESTIMADOR DA MINIMIZAÇÃO DO ERRO QUADRÁTICO MÉDIO (MMSE)

A partir dos resultados apresentados em Ephraim e Malah (1984), um grande número de métodos de melhoramento de voz tem sido derivado a partir do princípio MMSE. Dentre eles, destacam-se os estimadores MMSE do espectro de magnitude (EPHRAIM; MALAH, 1985; ABUTALEBI; RASHIDINEJAD, 2015) e do espectro de potência (WOLFE; GODSILL, 2003; DING; HUANG; XU, 2004; LU; LOIZOU, 2011). Nesta pesquisa, utiliza-se o estimador MMSE do espectro de potência em Lu e Loizou (2011), devido aos melhores resultados apresentados em comparação com seus antecessores e também em comparação com os estimadores do espectro de magnitude. Este estimador é obtido a partir do pressuposto de que as partes real e imaginária dos coeficientes obtidos pela DFT são modelados como variáveis aleatórias Gaussianas independentes. Dessa forma, segundo os autores, a densidade de probabilidade de $X_{l,n}^2$ é conhecida (ver Lu e Loizou (2011)). A função ganho foi então estimada a partir da função densidade de probabilidade *a posteriori* do espectro de potência do sinal de voz puro conforme a equação (35).

$$f_{X_{l,n}^2}(X_{l,n}^2|Y_{l,n}^2) = \frac{f_{Y_{l,n}^2}(Y_{l,n}^2|X_{l,n}^2) f_{X_{l,n}^2}(X_{l,n}^2)}{f_{Y_{l,n}^2}(Y_{l,n}^2)}. \tag{35}$$

Calculando a média da densidade de probabilidade em (35), os autores derivaram a seguinte função ganho (LU; LOIZOU, 2011):

$$G_{\text{MMSE}}(\xi[l,n],\gamma[l,n]) = \begin{cases} \left(\dfrac{1}{v_{l,n}} - \dfrac{1}{\mathrm{e}^{v_{l,n}}-1}\right)^{1/2}, & \text{se} \quad \sigma_x^2[l,n] \neq \sigma_w^2[l,n] \\ \left(\dfrac{1}{2}\right)^{1/2}, & \text{se} \quad \sigma_x^2[l,n] = \sigma_w^2[l,n]. \end{cases} \tag{36}$$

Sendo $\sigma_x^2[l,n] \equiv E\left\{X_{l,n}^2\right\}$, $\sigma_w^2[l,n] \equiv E\left\{W_{l,n}^2\right\}$ e $v_{l,n}$ é definido como (LU; LOIZOU, 2011):

$$v_{l,n} \equiv \frac{1-\xi_{l,n}}{\xi_k}\gamma_{l,n}. \tag{37}$$

Detalhes sobre a estimação de G_{MMSE} podem ser encontrados em Lu e Loizou (2011). No restante deste texto, refere-se ao método acima apenas como MMSE.

3.5 ESTIMADOR MÁXIMO *A POSTERIORI* (MAP)

Assim como o princípio MMSE, vários métodos de MV têm sido propostos baseados em estimadores MAP (do inglês: *Maximum a Posterior estimators*). Durante a mesma pesquisa em Lu e Loizou (2011), os autores derivaram um estimador MAP do espectro de potência, que forneceu melhores resultados do que os tradicionais estimadores MAP do espectro de magnitude em Lotter e Vary (2003), Lotter e Vary (2005). Tal estimador, foi desenvolvido a partir de uma maximização da função densidade de probabilidade *a posteriori* em (35) (LU; LOIZOU, 2011):

$$\hat{X}_{l,n}^2 = \arg\max f_{X_{l,n}^2}(X_{l,n}^2|Y_{l,n}^2). \tag{38}$$

A partir da equação (38), os autores derivaram a seguinte função ganho (LU; LOIZOU, 2011):

$$G_{\text{MAP}}(\xi[l,n]) = \begin{cases} 1, & \text{se } \xi[l,n] \geq 1 \\ 0, & \text{se } \xi[l,n] < 1. \end{cases} \tag{39}$$

Além dos bons resultados apresentados, este estimador MAP foi escolhido pois, ao contrário da função ganho G_{MMSE}, G_{MAP} possui valores binários. Além disso, ela é equivalente às máscaras binárias ideais utilizadas em CASA (do inglês: *Computational Auditory Scene Analysis*) (WANG; BROWN, 2006; WANG, 2011; LU; LOIZOU, 2011). A principal diferença entre G_{MAP} e funções ganho baseadas em máscaras binárias ideais é que as últimas são baseadas na estimação da SNR instantânea, enquanto G_{MAP} é baseada sobre a SNR_{prio} (LU; LOIZOU, 2011). O método de decisão direta é utilizado para estimar ξ. No restante do texto, o termo MB (Máscara Binária), refere-se ao filtro em (39).

A fim de avaliar os métodos MMSE e MB, as implementações fornecidas pelos autores foram utilizadas. É importante destacar que o mesmo banco de dados foi utilizado, assim nenhum ajuste foi necessário.

3.6 LIMIARIZAÇÃO WAVELET TRADICIONAL

A performance dos métodos de limiarização wavelet depende fundamentalmente da estimação do limiar λ. Donoho (1995) propôs o primeiro método de limiarização wavelet. Denominado limiar Universal, este método foi desenvolvido para suprimir o RGB, podendo não ser tão eficiente para ruídos reais.

No domínio wavelet, a projeção da $l-$ésima janela sobre a $k-$ésima escala, a partir da equação (16), gera os coeficientes wavelet representados por $Y[l,k,n]$ ($n = 0,1,\ldots,(N-1)/2^k$). Por simplicidade de notação, em alguns momentos $Y[l,k,n]$ será representado por Y. O cálculo de λ é realizado como segue (DONOHO, 1995):

$$\lambda = \sigma\sqrt{2\log_{10}(N)}, \tag{40}$$

em que N é o comprimento de Y, $\sigma = \text{mediana}(|Y|)/0,6745$.

Após o cálculo de λ, o sinal de voz melhorado no domínio wavelet $\widehat{X}$ é estimado a partir de Y conforme (41). Note que THR(.,.) denota a função de limiar utilizada.

$$\hat{X} = \text{THR}(Y,\lambda). \tag{41}$$

A função utilizada para a implementação da limiarização wavelet tradicional é a *Soft* (DONOHO, 1995):

$$\widehat{X} = \begin{cases} sign(Y)(|Y| - \lambda), & \text{se } |Y| > \lambda \\ 0, & \text{se } |Y| \leq \lambda. \end{cases} \tag{42}$$

Nos restante do texto, refere-se ao método de MV descrito acima como thr.

3.7 LIMIARIZAÇÃO WAVELET ADAPTATIVA

Existem maneiras alternativas para estimar o limiar. Em Ghanbari e Karami-Mollaei (2006), os autores propuseram um método que calcula o limiar de maneira adaptativa sobre os coeficientes wavelet gerados pela tranformada wavelet packet. Seja K o número de níveis de decomposição, para cada janela l a DWPT fornece 2^K sub-bandas. Assim, o limiar é calculado nos intervalos de silêncio da seguinte forma:

$$\lambda_{k,b} = \sigma_{k,b}\sqrt{2\ln(N_k)}, \tag{43}$$

onde $\sigma_{k,b} = \text{mediana}(|Y_{k,b}|)/0,6745$ é estimado sobre a escala k na sub-banda b.

Além de um limiar calculado para cada sub-banda b, o método utiliza a estimação da SNR segmentada *a posteriori* (SSNR) - (do inglês - *Posteriori Segmental SNR*) para cada sub-banda:

$$\text{SSNR}_{K,b} = 10\log_{10}\left[\frac{\sum_{n=0}^{N_k-1} Y_{K,b,n}^2}{\sum_{n=0}^{N_k-1} W_{K,b,n}^2}\right], \tag{44}$$

onde $Y_{K,b,n}$ e $W_{K,b,n}$ denotam o $n-$ésimo coeficiente wavelet packet do sinal ruidoso e do ruído estimado, respectivamente, no nível K sobre a $b-$ésima sub-banda da janela em processamento.

A atualização de $W_{K,b,n}$ é realizada de maneira análoga à equação (23), nos intervalos de silêncio. Os autores propõem atualizar $\lambda_{k,b}$ de maneira similar . Dessa forma, o método pode controlar de uma maneira mais eficiente o nível de redução do ruído em sinais de voz na presença de ruídos reais:

$$\lambda_{k,b,m} = \alpha.\lambda_{k,b,m-1} + (1-\alpha)\sigma_{k,b,m}\sqrt{2\ln(N_k)}. \tag{45}$$

O índice m representa a janela de silêncio identificada; $\lambda_{k,b,m}$ denota o limiar estimado na escala k e na sub-banda b, atualizado na $m-$ésima janela de silêncio.

Por fim, o valor do limiar calculado adaptativamente é determinado como segue:

$$T_{k,b} = \begin{cases} \lambda_{k,b,m} + (B_{k,b,m} - \lambda_{k,b,m})e^{\frac{\text{SSNR}_{K,b}}{\tau}}, & \text{se} \quad \text{SSNR}_{K,b} \geq 0, \\ B_{k,b,m}, & \text{se} \quad \text{SSNR}_{K,b} < 0. \end{cases} \tag{46}$$

Segundo os autores, τ deve ficar no intervalo $[2,4]$ e $B_{k,b,m} = 2\lambda_{k,b,m}$.

Para a avaliação deste método, a função de limiar utilizada será a função que também foi proposta pelos autores:

$$\text{THR}(Y, T_{k,b}) = \begin{cases} Y, & \text{se } |Y| \geq T_{k,b} \\ sign(Y).\dfrac{|Y^p|}{T_{k,b}^{p-1}}, & \text{se } |Y| \leq T_{k,b}, \end{cases} \tag{47}$$

onde o parâmetro p pode ser determinado por otimização (GHANBARI; KARAMI-MOLLAEI, 2006). Na presente implementação, utilizamos $p = 3$. A função wavelet utilizada foi a "db8", para ficar em conformidade com a referência.

O método descrito acima será denotado por thr-adpt.

3.8 MÉTODO WAVELET NÃO LIMIAR

Proposto em Soares et al. (2011), o primeiro método wavelet não limiar consiste na aplicação de três operadores sobre os coeficientes wavelet gerados pela DWT, com

o objetivo de reduzir o ruído. A curva utilizada para a redução do ruído é gerada por uma combinação polinomial destes operadores. A fim de manter a amplitude do filtro no intervalo $[0, 1]$ um ajuste sigmoidal é realizado. As equações (48), (49) e (50) mostram os três operadores propostos em Soares et al. (2011).

Operador média simples:

$$p_n = \frac{|Y_n| + |Y_{n+1}|}{2}, \qquad \text{para } n = 0, \ldots, N-2 \text{ e } p_{N-1} = p_{N-2}. \tag{48}$$

Operador relação sinal/ruído *a priori*:

$$z_n = \frac{|Y_n|}{\alpha + |Y_n|}, \qquad \text{para } n = 0, \ldots, N-2 \text{ e } z_{N-1} = z_{N-2}. \tag{49}$$

Operador relação sinal/ruído *a posteriori*:

$$r_n = \frac{|Y_n|}{\alpha + |Yw_n|}, \qquad \text{para } n = 0, \ldots, N-2 \text{ e } r_{N-1} = r_{N-2}. \tag{50}$$

Neste caso, $Y[l, K, n]_{apr,dtl} = \{Y_0, Y_1, \ldots, Y_n, \ldots, Y_{N-1}\}$ representa o $n-$ésimo coeficiente wavelet de aproximação e detalhe, no último nível de decomposição K, de um sinal ruidoso qualquer obtido pela DWT real. Em (49) e (50), α é um valor entre 0 e 1, escolhido *a priori*, Yw_n é um vetor de mesmo comprimento que Y_n, onde cada componente representa a média do ruído presente nos correspondentes componentes das janelas do último trecho de silêncio. Estes três operadores são responsáveis pela redução do ruído presente no sinal.

A curva usada para a redução do ruído é gerada por uma combinação polinomial dos operadores expostos nas equações (48), (49) e (50), da seguinte forma:

$$f_n = p_n^3 + z_n^2 + r_n. \tag{51}$$

A combinação polinomial proposta em (51), aumenta significativamente a amplitude do sinal processado. A fim de evitar o aumento de amplitude, um ajuste sigmoidal é realizado:

$$g_n = \left| \frac{1 - e^{-\tau f_n}}{1 + e^{-\tau f_n}} \cdot \frac{1 - e^{\tau f_n}}{1 + e^{\tau f_n}} \right|, \tag{52}$$

onde τ controla a inclinação das duas sigmoides envolvidas no processo.

A equação (52) gera os coeficientes do filtro $G_{apr,dtl} = \{g_0, g_1, \ldots, g_n, \ldots, g_{N-1}\}$, que atua no domínio wavelet da seguinte forma: $\hat{X}_n = g_n.Y_n$ $(n = 0, \ldots, N-1)$.

Em Abreu (2013), melhorias para o método de Soares et al. (2011) são propostas, as mesmas são relatadas na próxima seção.

3.9 MÉTODO WAVELET NÃO LIMIAR BASEADO NA DT−ℂWT

Inspirado pelo método da seção 3.8, propõe-se nesta tese um método não limiar baseado em wavelets complexas. A busca pela melhoria do método proposto em Soares et al. (2011) se iniciou em Abreu (2013), onde os autores propuseram o uso de equações de diferenças de diversas ordens para a estimação do filtro, seguido pelo mesmo ajuste sigmoidal em (52). Com o intuito de superar os dois primeiros, apresenta-se, nesta seção, um método que utiliza a transformada wavelet complexa. Além disso, sugere-se a utilização apenas das diferenças de ordem par para a estimação do filtro. Com isso, um novo método para redução de ruído em sinais de voz é derivado.

Os coeficientes wavelet complexos são gerados pela DT−ℂWT. O sinal de voz no domínio wavelet é estimado por equações de diferenças comumente utilizadas na solução numérica de equações diferenciais (SPERANDIO; MENDES; SILVA, 2003):

$$c^2_{k,n} = Y_{k,n+1} - 2Y_{k,n} + Y_{k,n-1}, \tag{53}$$

$$c^4_{k,n} = Y_{k,n+2} - 4Y_{k,n+1} + 6Y_{k,n} - 4Y_{k,n-1} + Y_{k,n-2}, \tag{54}$$

$$a^2_{k,n} = Y_{k,n} - 2Y_{k,n+1} + Y_{k,n+2}, \tag{55}$$

$$a^4_{k,n} = Y_{k,n} - 4Y_{k,n+1} + 6Y_{k,n+2} - 4Y_{k,n+3} + Y_{k,n+4}. \tag{56}$$

Nas equações de (53) a (56), $c^p_{k,n}$ e $a^p_{k,n}$ representam o $n-$ésimo coeficiente dos vetores de diferenças centradas e avançadas de ordem p, respectivamente, calculadas para cada escala $k(k = 1, 2, \ldots, 6)$. O próximo passo do método consiste em aplicar o operador relação sinal/ruído *a priori* (equação (49)) sobre todos os vetores de diferenças obtidos.

O sinal de saída é estimado pela combinação linear do operador relação sinal/ruído *a priori* (ORSP), quando aplicado a todos os vetores de diferenças:

$$s_{k,n} = \text{ORSP}(c^2_{k,n}) + \text{ORSP}(a^2_{k,n}) + \text{ORSP}(c^4_{k,n}) + \text{ORSP}(a^4_{k,n}). \tag{57}$$

Finalmente, os coeficientes do filtro são obtidos conforme o ajuste proposto por Soares et al. (2011):

$$g_{k,n} = \left| \frac{1 - e^{-\tau s_{k,n}}}{1 + e^{-\tau s_{k,n}}} \, \frac{1 - e^{\tau s_{k,n}}}{1 + e^{\tau s_{k,n}}} \right|. \tag{58}$$

O sinal puro estimado no domínio wavelet é obtido pela multiplicação $\widehat{X}_{k,n} = g_{k,n} Y_{k,n}$. Este método consiste em aplicar o esquema não limiar descrito acima sobre os coeficientes de detalhes para todas as escalas $k(k = 1, 2, \ldots, 5)$. Desta forma, é proposto o projeto de dois filtros, um para a parte real e outro para a parte imaginária dos coefecientes wavelet complexos. O filtro final, $g_{k,n}$, é então obtido pela média dos filtros projetados. Vale ressaltar também que o método descrito acima não altera a fase dos coeficientes wavelet complexos. De maneira análoga aos métodos que atuam no domínio da frequência, este método altera apenas a magnitude do sinal contaminado.

É importante salientar que o método wavelet não limiar descrito nesta seção foi desenvolvido após estudos realizados sobre os trabalhos desenvolvidos em Soares et al. (2011) e Abreu (2013). Abreu (2013) avançou na proposta de Soares et al. (2011) e propôs o uso de diferenças finitas para a estimação do filtro. Apesar dos avanços na qualidade do processamento, espera-se que com o uso de wavelets complexas e apenas diferenças de ordem par os resultados sejam ainda melhores. Refere-se ao método descrito acima como comp-wav.

3.10 SIMULAÇÕES

Nesta seção constam as avaliações objetivas de todos os métodos descritos neste capítulo, ou seja, os métodos apresentados nas seções 3.1 a 3.9. O objetivo principal é identificar um ou mais métodos que possam obter melhor desempenho frente a algum tipo de ruído real. Neste sentido, escolheram-se aqueles tipos de ruído que são mais comuns em situações quotidianas.

Algoritmos de redução de ruído em sinais de voz estão presentes em várias aplicações, como descrito no Capítulo 1. Devido às facilidades da internet móvel e à diversidade de aplicações, o usuário pode estar inserido nos mais variados ambientes ruidosos. Neste sentido, utilizam-se nesta pesquisa os ruídos de vozes, cafeteria, carro, salão de exposições, tráfego e trem. Segue abaixo uma breve descrição sobre os seis tipos de ruídos considerados:

- Vozes: áudio captado em ambientes com um grande número de pessoas, onde todos estão falando ao mesmo tempo. Nenhuma das palavras são discerníveis;
- Cafeteria: áudio captado dentro de cafeterias. Som de pratos ao serem postos sobre mesas, talheres ao entrarem em contato com pratos e pessoas conversando. Palavras pouco discerníveis;
- 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;
- Salão de exposições: áudio captado dentro de uma sala muito grande onde pessoas conversam, assoviam e sons de eco são gerados. Pode ser associado também a áreas de lazer no interior de shoppings. As palavras são pouco discerníveis;
- Tráfego: áudio captado durante um congestionamento de carros onde buzinas soam frequentemente;
- Trem: áudio captado no interior de um trem em movimento;

O banco de dados utilizado para as avaliações é o NOIZEUS. Proposto em Hu e Loizou (2007), este banco de dados foi desenvolvido especificamente para avaliar algoritmos de melhoramento de voz. O NOIZEUS é composto por 30 sentenças, 15 na voz masculina e 15 na voz feminina. Todos os sinais possuem uma taxa de amostragem de 8 kHz e receberam uma filtragem para simular as características de frequências recebidas por aparelhos telefônicos. Além do conjunto de sentenças originais (livre da presença de ruído), faz parte do NOIZEUS um conjunto de sinais contaminados com diferentes tipos de ruídos reais, em diferentes níveis de contaminação.

Serão consideradas três medidas de avaliação de qualidade objetivas, a SNR global, coeficiente de correlação de Pearson (FIELD, 2005) e PESQ (do inglês - *Perceptual Evaluation of Speech Quality*) (REC ITUT, 2001). Avaliada em decibéis, a SNR global é calculada como em (59) e mensura o nível de ruído presente no sinal. Dessa forma, apresenta-se neste trabalho o grau de melhoria na SNR do sinal processado, em comparação com a SNR do sinal contaminado (GHANBARI; KARAMI-MOLLAEI, 2006):

$$\text{SNR} = 10\log_{10}\left[\frac{\sum_{n=0}^{N-1} x[n]^2}{\sum_{n=0}^{N-1} |x[n]-\hat{x}[n]|^2}\right], \tag{59}$$

onde N é o comprimento dos sinais original $x[n]$ e melhorado $\hat{x}[n]$.

O coeficiente de correlação de Pearson em (60) varia no intervalo [-1,1], avaliando a relação mútua entre os sinais original e melhorado. Em outras palavras, esta medida mensura quanto os sinais original e melhorados estão correlacionados. Dessa forma, é comum apenas correlações positivas. Quanto mais próximo de um está a correlação, melhor será a recuperação do sinal original, em outras palavras, melhor será a redução do ruído. O coeficiente de correlação de Pearson r é calculado como segue (FIELD, 2005):

$$r = \frac{N\left(\sum x[n]\hat{x}[n]\right) - \left(\sum x[n]\right)\left(\sum \hat{x}[n]\right)}{\sqrt{\left[N\sum x[n]^2 - \left(\sum x[n]\right)^2\right]\left[N\sum \hat{x}[n]^2 - \left(\sum \hat{x}[n]\right)^2\right]}}. \tag{60}$$

Proposta pela ITU-T (do inglês - *International Telecommunications Union*) e padronizada pela recomendação P.862 (02/01), a PESQ é uma das medidas mais complexas a fim de se checar a qualidade de um sinal de voz. A avaliação realizada pela PESQ é baseada em características psicoacústicas do ouvido humano. Ela representa um modelo cognitivo e os resultados são expressos por meio de uma escala que varia de -0,5 a 4.5. Quanto maior a pontuação melhor a qualidade do sinal processado. A PESQ foi proposta inicialmente para predizer a qualidade do sinal de voz, no entanto, em Ma, Hu e Loizou (2009) os autores verificaram que esta medida é razoável na predição da inteligibilidade de sentenças.

Na Tabela 1 consta a melhoria na SNR, denotada por SNRI, e correlação obtidas sobre o banco de sinais NOIZEUS. Para cada tipo de ruído, são utilizados 4 níveis de contaminação (SNR), 15 dB, 10 dB, 5 dB e 0 dB. Para cada nível de contaminação, calcula-se, então, a média das correlações e da SNRI para os 30 sinais. Ao final das simulações, 120 sinais para cada tipo de ruído serão avaliados, totalizando então 840 simulações para cada um dos métodos descritos neste capítulo.

Tabela 1 – Teste de redução de ruído.

		SNRI				Correlação			
Ruído	Método	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
Vozes	SE	4,50	2,66	0,57	-0,00	0,78	0,90	0,95	0,97
	SEP	3,01	1,94	0,50	0,17	0,74	0,88	0,94	0,96
	WIENER	**4,77**	2,60	1,47	0,51	**0,79**	0,90	0,96	0,98
	MB	3,27	2,35	1,48	0,53	0,76	0,90	0,96	0,98
	MMSE	3,91	3,16	**2,43**	**1,48**	**0,79**	0,91	**0,97**	**0,99**
	thr	1,98	1,58	1,16	0,73	0,71	0,88	0,96	0,98
	thr-adpt	3,13	2,21	1,31	0,56	0,73	0,89	0,96	**0,99**
	non-thr	2,95	2,90	1,16	0,26	0,76	0,90	0,95	0,96
	comp-wav	3,03	**3,56**	1,52	0,27	0,76	**0,92**	0,96	0,97
Cafeteria	SE	4,41	2,69	0,43	0,02	0,84	0,91	0,95	0,97
	SEP	2,98	1,93	0,50	0,08	0,80	0,89	0,95	0,97
	WIENER	**5,06**	3,20	2,20	0,66	**0,86**	0,92	0,96	0,98
	MB	3,24	2,85	1,76	0,76	0,82	0,92	**0,97**	**0,99**
	MMSE	3,79	**3,51**	**2,66**	**1,74**	0,84	**0,93**	**0,97**	**0,99**
	thr	1,20	1,09	0,77	0,61	0,75	0,88	0,96	**0,99**
	thr-adpt	2,87	2,11	1,08	0,28	0,79	0,90	0,96	**0,99**
	non-thr	2,19	1,81	0,48	-0,01	0,78	0,89	0,95	0,96
	comp-wav	3,54	2,97	1,06	0,51	0,82	0,92	0,96	0,97
Carro	SE	6,71	4,44	1,07	0,17	0,91	0,95	0,96	0,97
	SEP	4,63	3,28	0,69	0,16	0,86	0,92	0,96	0,97
	WIENER	8,13	6,39	4,94	3,36	**0,94**	**0,97**	0,98	**0,99**
	MB	**8,66**	**7,14**	**5,61**	4,04	**0,94**	**0,97**	**0,99**	**0,99**
	MMSE	7,23	6,51	5,53	**4,38**	0,92	0,96	**0,99**	**0,99**
	thr	0,50	0,48	0,42	0,35	0,74	0,88	0,96	**0,99**
	thr-adpt	3,86	2,74	1,56	0,54	0,82	0,91	0,97	**0,99**

continua na próxima página

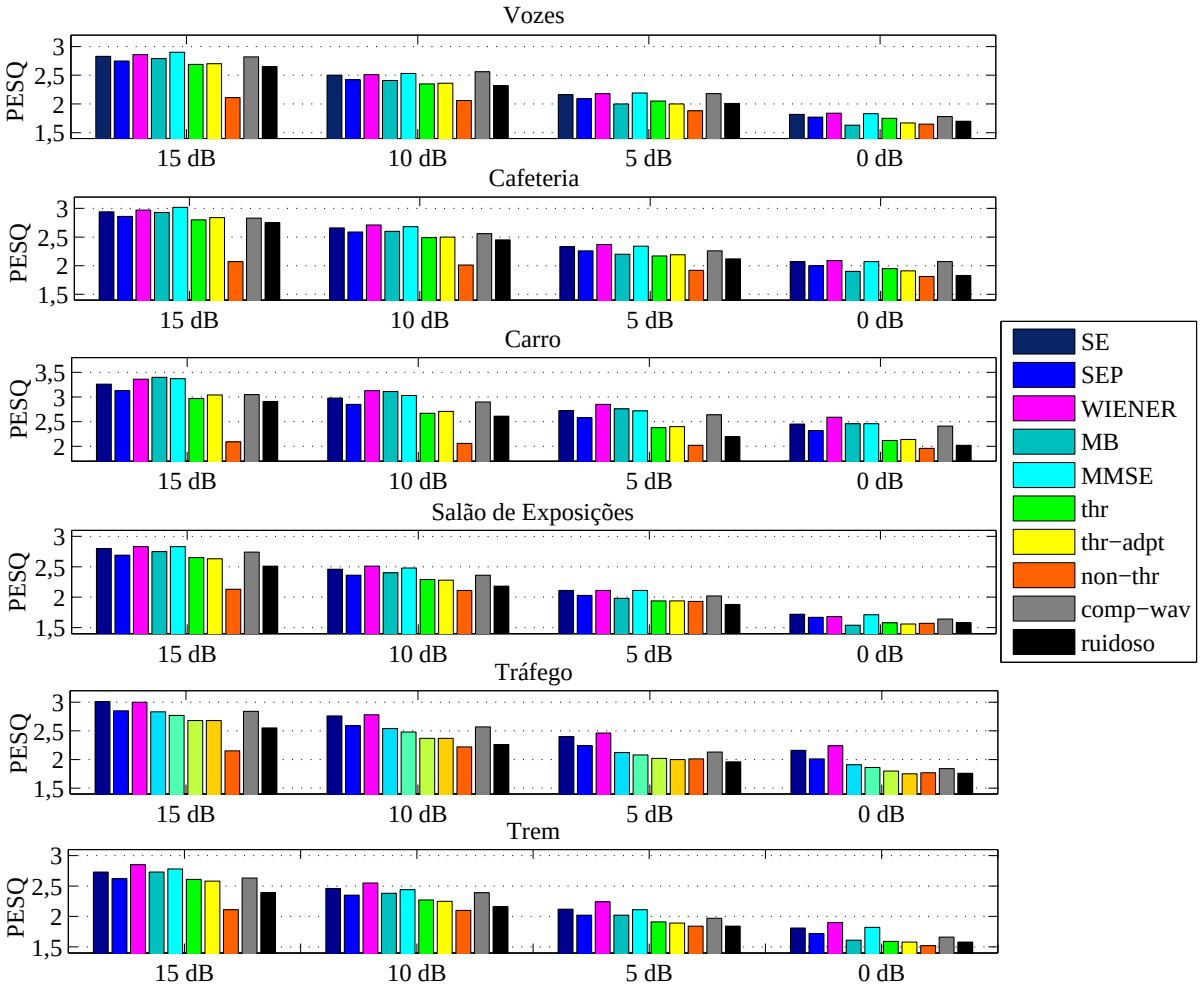
continuação da página anterior

		SNRI				Correlação			
Noise	Method	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
	non-thr	1,47	1,09	0,34	0,04	0,76	0,88	0,94	0,96
	comp-wav	5,59	4,12	0,67	0,03	0,88	0,94	0,96	0,97
	SE	4,40	2,58	0,48	0,01	0,78	0,90	0,95	0,97
	SEP	3,02	1,81	0,30	0,16	0,75	0,88	0,94	0,96
	WIENER	**4,54**	2,81	1,65	0,33	0,78	0,90	0,96	0,98
	MB	3,72	2,50	1,82	0,72	0,78	0,90	0,96	0,98
S. de exposições	MMSE	3,97	3,05	**2,55**	**1,60**	**0,79**	**0,91**	**0,97**	**0,99**
	thr	3,45	2,67	1,90	1,26	0,74	0,90	0,96	**0,99**
	thr-adpt	3,19	2,32	1,69	0,85	0,73	0,89	0,96	**0,99**
	non-thr	3,56	**3,50**	1,28	0,24	0,78	0,89	0,95	0,97
	comp-wav	2,16	2,46	0,70	0,19	0,72	0,89	0,95	0,96
	SE	6,39	4,02	0,72	0,02	0,91	0,94	0,96	0,97
	SEP	5,00	3,35	0,69	0,14	0,87	0,92	0,96	0,97
	WIENER	**6,17**	**4,54**	2,62	1,34	**0,90**	**0,94**	**0,97**	**0,99**
	MB	4,40	3,22	2,20	0,99	0,86	0,92	**0,97**	**0,99**
Tráfego	MMSE	4,01	3,32	**2,75**	**1,77**	0,85	0,93	**0,97**	**0,99**
	thr	3,56	2,81	1,80	1,10	0,81	0,91	**0,97**	**0,99**
	thr-adpt	3,77	2,75	1,74	0,95	0,82	0,91	**0,97**	**0,99**
	non-thr	4,34	3,74	0,73	0,17	0,85	0,93	0,96	0,97
	comp-wav	4,25	3,29	0,48	-0,07	0,84	0,93	0,95	0,96
	SE	5,29	3,14	0,41	-0,11	0,82	0,91	0,95	0,97
	SEP	3,75	2,26	0,23	-0,19	0,77	0,89	0,94	0,97
	WIENER	**5,39**	3,51	1,75	0,52	**0,83**	0,92	0,96	**0,99**
	MB	4,19	3,33	2,27	1,07	0,80	0,92	**0,97**	**0,99**
Trem	MMSE	4,74	**4,06**	**3,14**	**1,93**	0,82	**0,93**	**0,97**	**0,99**
	thr	3,71	2,91	2,05	1,05	0,75	0,90	0,96	**0,99**
	thr-adpt	3,45	2,60	1,82	1,03	0,74	0,90	0,96	**0,99**
	non-thr	3,47	3,57	1,05	0,22	0,77	0,92	0,95	0,97
	comp-wav	0,92	2,47	2,38	0,31	0,71	0,90	**0,97**	0,98

Fonte: Elaborado pelo próprio autor.

Na Tabela 1, valores em negrito destacam a melhor nota para cada nível de contaminação. Devido a sua importância em aplicações práticas, o ruído de vozes será analisado separadamente. De acordo com a Tabela 1, para a contaminação com o ruído vozes, vários métodos trabalharam igualmente bem em termos de SNRI e correlação. Os métodos ligeiramente melhores foram MMSE, WIENER e comp-wav. De um modo geral,

Figura 6 – Resultados em termos de notas PESQ para todas as condições de ruído simuladas.



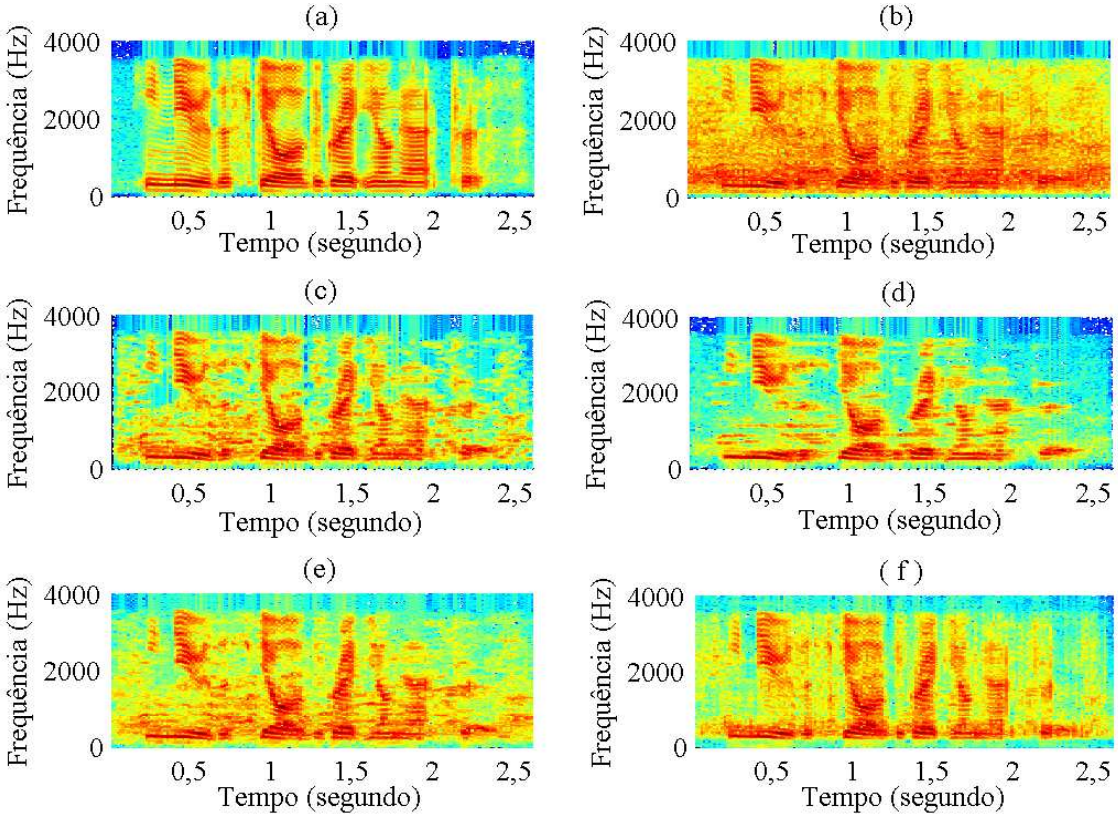
Fonte: Elaborado pelo próprio autor.

eles conseguiram boa SNRI, e consequentemente, boa correlação. Além disso, o método thr trabalhou pior em termos de atenuação do ruído em condições de baixa SNR. Analisando a Figura 6, confirma-se que MMSE, WIENER e comp-wav também obtiveram melhores notas PESQ. Isto significa que estes métodos alcançaram melhor qualidade e inteligibilidade da sentença processada.

De um modo geral, para ambientes com ruído multioradores, os métodos SEP, thr e non-thr foram superados pelos demais. Apesar do método non-thr realizar uma boa redução de ruído, as notas PESQ não foram tão boas. Este fato indica que distorções foram inseridas no sinal processado. Entretanto, comp-wav foi capaz de aliar boa redução de ruído com boa qualidade e inteligibilidade da sentença processada. Vale destacar também que thr-adpt superou o método thr no quesito atenuação do ruído.

Na Figura 7 são mostrados os espectrogramas de um sinal de voz limpo e suas versões contaminada e processadas pelos métodos SE, WIENER, MMSE e comp-wav.

Figura 7 – Espectrogramas para (a) sinal limpo; (b) sinal contaminado por vozes; (c) melhorado por SE (SNRI=6,65, cor=0,89, PESQ=2,27); (d) melhorado por WIENER (SNRI=6,84, cor=0,90, PESQ=2,20); (e) melhorado por MMSE (SNRI=7,42, cor=0,91, PESQ=2,30); (f) melhorado por comp-wav (SNRI=7,86, cor=0,92, PESQ=2,36).



Fonte: Elaborado pelo próprio autor.

Note que SE e MMSE geraram mais ruído residual do que WIENER. Isto se deve a estimação da SNR_{priori} pelo método decisão direta para WIENER. Em contrapartida, as notas PESQ foram as mais baixas entre os quatro métodos devido a atenuação excessiva do ruído. O método proposto, comp-wav, deixou menos ruído residual quando comparado com os métodos MMSE e SE, além disso, não realizou atenuação excessiva.

Pode ser visto a partir da Tabela 1 e da Figura 6 que MMSE e comp-wav alcançaram os melhores resultados, com uma ligeira vantagem para comp-wav.

As discussões acima destacaram as principais diferenças entre os métodos utilizados sob condições de ruído multioradores. No entanto, comparações envolvendo todas as condições de ruído são interessantes a fim de checar se algum método obteve o melhor desempenho em todas ou em grande parte das condições de ruído simuladas.

Em termos de SNRI, verifica-se que WIENER obteve o melhor desempenho para condições de 0 dB. Em suma, o método WIENER trabalha muito bem em baixas condições de SNR. De um modo geral, os métodos que alcançaram melhor SNRI foram WIENER,

MMSE, MB e comp-wav. Estes métodos trabalharam igualmente bem, de forma que as diferenças nas avaliações objetivas não foram significantes.

Em termos de correlação, não existem diferenças significantes entre todos os métodos simulados. Vale destacar a boa performance dos métodos baseados na limiarização wavelet em condições de 15 dB. Verifica-se, com base nos resultados, que métodos de limiarização trabalham melhor em condições de alta SNR.

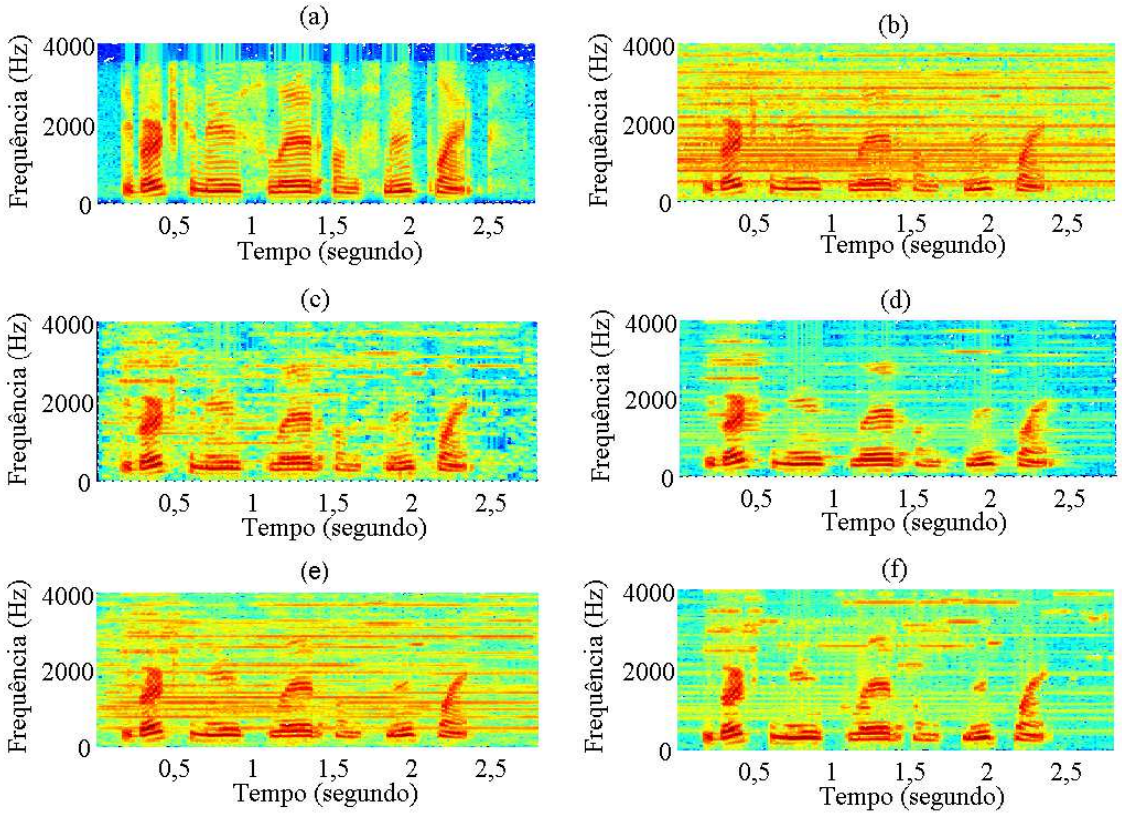
Analisando as notas PESQ na Figura 6, verifica-se que WIENER, MMSE, SE e comp-wav trabalharam igualmente bem na maioria das condições, exceto sob condições de trem e tráfego. Para o ruído de trem, WIENER se sobressaiu em relação aos demais. Isto se deve à alta potência do ruído de trem. Para o ruído de tráfego, SE e WIENER obtiveram resultados melhores para todos os níveis de contaminação, quando comparados aos demais métodos. No caso da SE, isto se deve às buzinas dos carros que soam em frequências específicas, em conjunto com sentenças muito curtas que possibilitam a estimação e subtração destas frequências.

Fica claro, a partir da Figura 6, que não existe um único algoritmo que se destaca, mas vários algoritmos que trabalham igualmente bem sob a maioria das condições de ruído, com pequenas mudanças dependendo do tipo do ruído de fundo. Por exemplo, SE mostrou ser mais adequada para o ruído de tráfego; comp-wav, MMSE e WIENER para o ruído vozes; WIENER para os ruídos de trem e carro; MMSE e WIENER obtiveram ligeira vantagem para o ruído de cafeteria; SE, WIENER e MMSE para salão de exibições.

É importante destacar que, de um modo geral, métodos baseados em modelos estatísticos que atuam no domínio da frequência obtiveram melhor performance para a maioria das condições de ruído (note os resultados dos métodos MMSE, MB e WIENER). O único método baseado em wavelets que obteve desempenho semelhante para algumas condições de ruído foi comp-wav. Métodos com desempenho intermediários foram thr, thr-adpt e SEP, sendo superados pelos outros. SE superou os principais métodos apenas para condições específicas de ruído, onde frequências específicas são facilmente identificadas. Além disso, é conhecido que em condições de SNR muito baixa, o método SE pode danificar a inteligibilidade do sinal (LOIZOU; KIM, 2011). Finalmente, o método non-thr obteve o pior desempenho na maioria das condições simuladas, exceto para vozes, salão de exibição e tráfego com 0 dB de contaminação. Esta performance ruim pode ter sido motivada pelas características do banco de dados utilizados (todos os sinais receberam uma filtragem para simular condições reais de operação de aparelhos telefônicos), levando em consideração que os resultados em Soares et al. (2011) foram muito melhores.

Um fato importante verificado a partir da Tabela 1 e da Figura 6 é que análises objetivas podem variar de acordo com o ambiente ruidoso. Algumas vezes, esta variação pode ser expressiva como no caso do ruído de tráfego: apesar do método MMSE estar entre os melhores métodos para a maioria das condições simuladas, para o ruído de tráfego seu

Figura 8 – Espectrogramas para (a) sinal limpo; (b) sinal contaminado por tráfego; (c) melhorado por SE (SNRI=9,65, cor=0,95, PESQ=2,47); (d) melhorado por WIENER (SNRI=10,95, cor=0,96, PESQ=2,48); (e) melhorado por MMSE (SNRI=9,68, cor=0,94, PESQ=1,97); (f) melhorado por MB (SNRI=10,69, cor=0,95, PESQ=2,25).



Fonte: Elaborado pelo próprio autor.

desempenho foi significativamente pior em comparação com SE, WIENER e MB. Este fato é ilustrado na Figura 8.

O esquema de filtragem proposto pelo método comp-wav melhorou significativamente os resultados, em termos de notas PESQ, alcançados pela limiarização wavelet. Isto se deve principalmente ao filtro sigmoide que gera uma curva de atenuação suave. Além disso, verifica-se mais uma vez que o método thr não é tão eficiente para o caso de contaminação por ruído real.

Outro fato relevante destacado pelas simulações é que métodos de MV melhoram a qualidade do sinal (Tabela 1) mas não melhoram significativamente a inteligibilidade da sentença (Figura 6). Note que na maioria dos ambientes de ruído simulados, a melhoria em notas PESQ foi pobre em comparação com notas PESQ dos sinais sem processamento. Este fato também foi apontado nos estudos realizados em Loizou e Kim (2011) e Hu e Loizou (2007).

Por fim, após as simulações e análises apresentadas, seguem abaixo as respostas

encontradas para cada questão levantada no início deste capítulo, identificada por seu respectivo número:

1. baseado nas simulações e discussões realizadas neste capítulo, verifica-se que a limiarização wavelet tradicional não é tão eficiente sob condições de ruídos reais. Na verdade, ela foi desenvolvida com o objetivo de suprimir o RGB. O método baseado na limiarização wavelet adaptativa (thr-adpt) melhorou significativamente a atenuação do ruído fornecida pela limiarização tradicional. No entanto, os ganhos em termos de PESQ não são significantes. Apesar do método non-thr ter alcançado um desempenho fraco em termos de PESQ, verifica-se que métodos wavelet que dispensam o uso do limiar são promissores. O método comp-wav superou totalmente os métodos de limiarização wavelet em termos de notas PESQ, além de trabalhar igualmente bem no quesito supressão do ruído;

2. os métodos baseados na limiarização wavelet foram superados em termos de PESQ pelos outros métodos, exceto por non-thr. O único método baseado em wavelets que trabalhou igualmente bem aos métodos baseados na DFT, para algumas condições de ruído, foi comp-wav;

3. o método wavelet não limiar baseado na DT$-\mathbb{C}$WT, proposto durante esta pesquisa, apresentou resultados satisfatórios. As avaliações objetivas mostraram que comp-wav superou os resultados alcançados pela limiarização wavelet e pelo método non-thr;

4. de um modo geral, métodos baseados em modelos estatísticos que atuam no domínio da frequência tiveram melhor performance para a maioria das condições de ruído simuladas. Considerando todos os métodos implementados, aqueles que alcançaram os melhores resultados foram: WIENER, MMSE, SE, comp-wav e MB.

Verificou-se por meio das simulações que não existe um único e melhor algoritmo para MV, mas vários algoritmos que trabalham igualmente bem para a maioria das condições, com possíveis mudanças, dependendo das características do ruído de fundo. Como destacado para o ruído de tráfego, essas variações podem ser expressivas, indicando que métodos que incorporam classificação de ruído são uma tendência. Este fato motiva o desenvolvimento do próximo capítulo, onde o problema específico da classificação de ruído é abordado.

4 UMA ABORDAGEM IMUNOLÓGICA BASEADA NO ALGORITMO DE SELEÇÃO NEGATIVA PARA CLASSIFICAÇÃO DE RUÍDOS REAIS EM SINAIS DE VOZ

Neste capítulo será apresentada uma nova abordagem capaz de identificar e classificar o ruído de fundo presente em sinais de voz. O método proposto se baseia no algoritmo de seleção negativa (ASN) e na transformada wavelet complexa dual-tree. O desenvolvimento do ASN foi inspirado em um sistema natural, o sistema imunológico humano (SIH). O surgimento de algoritmos bioinspirados foi impulsionado pela busca por metodologias inteligentes que sejam capazes de resolver problemas do mundo real (DASGUPTA, 1999). Dentre tais algoritmos, destacam-se as redes neurais artificiais e os sistemas imunológicos artificiais. Sendo assim, será exposta a teoria básica na qual o ASN se fundamenta, para então propor um algoritmo que se baseia em seus princípios. Detalhes adicionais sobre os conceitos biológicos utilizados são expostos nos Apêndices A, B e C.

4.1 MELHORAMENTO DE VOZ E CLASSIFICAÇÃO DE RUÍDO

Como destacado no Capítulo 1, métodos clássicos de MV incluem subtração espectral e subtração espectral em potência, métodos baseados em estimadores do erro médio quadrático e estimadores de máximo *a posteriori*, filtragem de Wiener e limiarização wavelet. Apesar de cada um desses métodos possuir sua própria base teórica com objetivos específicos, todos eles requerem uma estimação do perfil do ruído a fim de realizar sua supressão. Na verdade, quanto melhor a estimação do ruído, maior será a qualidade do processamento (EPHRAIM; MALAH, 1984; HU; LOIZOU, 2007).

As principais aplicações de MV incluem RAV e sistemas de comunicação móvel, onde o ruído do ambiente é o principal problema. O ruído pode prejudicar a qualidade da comunicação por fala e o desempenho de algoritmos de RAV (RABINER; JUANG, 1993). De acordo com Yuan e Xia (2015), o projeto de algoritmos de MV geralmente não leva em consideração as diferenças em propriedades estatísticas de diferentes tipos de ruído. Isto pode ser a causa do fraco desempenho de alguns algoritmos em condições de ruído específicas, como pode ser visto nos resultados apresentados em Hu e Loizou (2007) e no Capítulo 3 deste texto. Nesse sentido, com o desenvolvimento de aparelhos cada vez mais poderosos, como smartphones, tablets e aparelhos auditivos, possibilita-se o projeto de algoritmos que trabalham de maneira eficiente em qualquer ambiente ruidoso, com a incorporação de classificação de ruído. Como um exemplo de implementação de tempo real, em Parris, Torlak e Kehtarnavaz (2014), uma implementação para smartphones que inclui transformação para o domínio da frequência, classificação e supressão de ruído foi apresentada.

Um dos primeiros algoritmos de MV viável em aplicações práticas é a SS, proposta por Boll (1979). A ideia básica consiste em estimar o espectro de frequência do ruído para então subtraí-lo do sinal ruidoso. A SS gera resultados satisfatórios quando o espectro de frequência do ruído é uniforme ou quando o ruído é estacionário. Sob condições de ruídos reais, SS gera tons indesejáveis no sinal processado, conhecido como ruído musical. No entanto, o ruído musical não é um problema específico da SS. Métodos baseados em modelos estatísticos também são afetados por este problema. Em Scalart e Vieira Filho (1996), os autores sugeriram que os conceitos de SNR_{post} e SNR_{prio} poderiam evitar ou atenuar o problema do ruído musical por meio de um derivado filtro de Wiener. Simulações em Scalart e Vieira Filho (1996), Hu e Loizou (2007) e no Capítulo 3 desta tese (ver Figura 7) mostraram resultados melhorados; no entanto, para alguns tipos de ruído a performance não foi tão boa. Isto se deve a estimação da SNR_{prio}, que depende fundamentalmente de uma boa estimação do ruído (EPHRAIM; MALAH, 1984). Métodos baseados em wavelets também estão sujeitos a fraca performance devido ao tipo de ruído de fundo. Como mencionado anteriormente, a estimação do limiar é baseada no pressuposto de que o ruído presente no sinal é o RGB.

Como pode ser notado, é natural que futuras pesquisas em MV foquem no tratamento e na compreensão do ruído de fundo. Neste sentido, o desenvolvimento de sistemas que são capazes de predizer o tipo do ruído presente em um sinal de voz ruidoso são indispensáveis para métodos estatísticos ou para a obtenção de parâmetros ótimos em algoritmos. Esta tendência é confirmada em dois trabalhos atuais. Em Yuan e Xia (2015), os autores usaram classificação de ruído para escolher parâmetros de suavização ótimos na estimação do ruído e da SNR_{prio}. Tais parâmetros ótimos foram usados no algoritmo de estimação em amplitude log-spectral. Em Xia e Bao (2014), a classificação de ruído foi utilizada para predizer o tipo de ruído de fundo e um específico modelo autodecodificador ponderado de eliminação de ruído (VICENT et al., 2008) foi utilizado na estimação do espectro de potência do sinal puro, posteriormente empregado na implementação de um filtro de Wiener. Além destas aplicações em MV, a classificação de ruído tem sido utilizada em sistemas de RAV (XU et al., 2005; HOSEINKHANI et al., 2012), classificação de cenas acústicas (MA; SMITH; MILNER, 2003; RAKOTOMAMONJY; GASSO, 2015) e aplicações em aparelhos auditivos (KATES, 1995; SAKI; KEHTARNAVAZ, 2014).

Em Ma, Smith e Milner (2003), os autores inseriram classificação de ruído em aplicações dependentes do contexto. Neste caso, um classificador de ruído baseado em um modelo de Markov oculto foi proposto. Sem qualquer distinção entre segmentos de voz e de silêncio, o algoritmo utiliza os coeficientes cepstrais de frequência mel como características de entrada para o classificador. Contudo, os autores destacaram que segmentos do sinal que contêm apenas ruído seriam mais adequados para a classificação.

O classificador de ruído proposto em Yuan e Xia (2015) é baseado no tradicional

classificador SVM. As características são adquiridas mapeando a energia do ruído a partir da DFT com 256 pontos para o domínio Bark. Em outras palavras, a energia do ruído em cada banda de frequência Bark é calculada e usada como características de entrada para o classificador. A fim de realizar a classificação do ruído contido em um sinal de voz ruidoso, as primeiras 15 janelas são assumidas como segmentos que contêm apenas ruído. Assim, a classificação é realizada apenas nessas 15 janelas da seguinte forma: dentro das primeiras 15 janelas, o ruído é classificado janela por janela e o tipo de ruído com maior número de votos é selecionado para ser o tipo de ruído em toda a sentença.

De um modo similar, em Xia e Bao (2014), os autores usaram o espectro de potência do sinal ruidoso normalizado em sub-bandas de frequência como características de entrada para o classificador. No entanto, neste caso, um modelo de mistura Gaussiana é usado como classificador (REYNOLD; ROSE, 1995). Assim como em Yuan e Xia (2015), as primeiras 10 janelas do sinal ruidoso são assumidas como segmentos que contém apenas ruído e a classificação do ruído é realizada janela a janela. O tipo de ruído com maior número de votos é selecionado para classificação. Além disso, a fim de considerar a possibilidade de mudança do tipo de ruído, os autores utilizaram um detector de atividade de voz (DAV) e a classificação do ruído é atualizada cada vez que uma janela com ausência de voz é identificada.

Os resultados apresentados em Yuan e Xia (2015) e Xia e Bao (2014) mostraram que ajustar o algoritmo de MV para cada tipo de ruído, melhora substancialmente os resultados. Além disso, no segundo trabalho, os autores treinaram um modelo autodecodificador ponderado de eliminação de ruído para cada tipo de ruído. Assim, após a classificação, utiliza-se o modelo treinado especificamente para o tipo de ruído encontrado para a estimação da magnitude do sinal puro.

4.2 MOTIVAÇÃO

A motivação para o desenvolvimento de um novo classificador é a proposição de uma metodologia para classificação de ruídos reais em sinais de voz, janela por janela, que irá contribuir com o desenvolvimento científico de sistemas de MV e outros sistemas de processamento de voz. Ao contrário dos métodos propostos em Yuan e Xia (2015) e Xia e Bao (2014), onde é assumido que os sinais estão sempre corrompidos por ruído, o método proposto será capaz de identificar se o sinal de voz está limpo, ou o nível de ruído é tão baixo que nenhum processamento (por ex. melhoramento) é necessário. Este ajuste, que será feito pelo usuário, permite que o classificador seja facilmente acoplado a outros sistemas de processamento de voz. Além disso, uma vez que o classificador verifica a ausência de ruído, nenhum processamento para classificação é necessário, e passa-se para outras fases do processamento da fala, podendo reduzir também o custo computacional.

Outro ponto a ser destacado é a necessidade de uma única janela para a inicialização do algoritmo proposto.

Além dos classificadores utilizados nos trabalhos mencionados acima, classificadores comumente utilizados em problemas de reconhecimento de padrões incluem redes neurais (RNs) (HAYKIN, 2008) e árvores de decisão (ADs) (BREIMAN et al., 1984). Portanto, simulações evolvendo ambos os classificadores serão apresentadas. Outra motivação deste trabalho é a introdução do conceito de SIA na área de pesquisa relacionada. SIAs constituem uma abordagem relativamente recente no campo da inteligência artificial. Pesquisadores na área de SIAs procuram por inspiração no SIH sobre como resolver problemas em engenharia e ciência da computação (DASGUPTA; NINO, 2008).

Composto por um conjunto de órgãos, células e moléculas, o SIH visa proteger um indivíduo de infecções, eliminando substâncias estranhas (ABBAS; LICHTMAN; PILLAI, 2008). De um modo geral, o SIH é capaz de reconhecer estruturas comum a partir de diferentes classes de microorganismos, a fim de gerar uma resposta imune. A exposição do SIH a um antígeno estranho aumenta sua capacidade de responder mais rapidamente a uma nova exposição do mesmo antígeno, caracterizando o conceito de memória imunológica (DE CASTRO, 2001; ABBAS; LICHTMAN; PILLAI, 2008).

A partir desses conceitos biológicos, vários algoritmos foram propostos. Em Forrest et al. (1994), os autores propuseram um algoritmo baseado na geração de células T no sistema imune que contribuiu com a ampla disseminação de SIAs (DASGUPTA; NINO, 2008). Denominada algoritmo de seleção negativa, esta técnica foi então aplicada ao problema de detecção de vírus em computadores. Apesar de originalmente usados em segurança computacional, SIAs têm sido aplicados em várias áreas; alguns exemplos são: reconhecimento de padrões (FORREST et al., 1993; HUNT et al., 1999), mineração de dados (DE CASTRO; VON ZUBEN, 2000b; KNIGHT; TIMMIS, 2001; PUTEH et al., 2008), otimização (FUKUDA; MORI; TSUKIAMA, 1999; DE CASTRO; VON ZUBEN, 2000a; XIAO; LI; ZHANG, 2015), detecção de falhas e anomalias (LIMA; LOTUFO; MINUSSI, 2014; LI; LIU; ZHANG, 2015), assim como aprendizagem de máquinas (HIGHTOWER; FORREST; PERELSON, 1996).

Neste capítulo, apresenta-se um classificador de ruídos reais para sistemas de processamento do voz, baseado no ASN (FORREST et al., 1994). Entre suas características, ASN é atrativo por sua simplicidade de implementação e alta acurácia em reconhecimento de padrões (DASGUPTA; NINO, 2008). O ASN é baseado principalmente em simples comparações entre padrões por meio de uma medida de afinidade (por ex. uma medida de distância), ao contrário de SVM e de RNA com implementações baseadas em algoritmos de otimização. Além disso, ao contrário dos trabalhos em Yuan e Xia (2015) e Xia e Bao (2014), onde características extraídas no domínio da frequência foram usadas, utiliza-se nesta pesquisa uma análise multiescalas a partir da DT$-\mathbb{C}$WT.

Como destacado previamente, classificação de ruído tem se tornado uma tendência para processamento de voz, tornando possível o desenvolvimento de algoritmos que irão atuar de uma maneira específica para cada ambiente ruidoso real.

4.3 O ALGORITMO DE SELEÇÃO NEGATIVA: ANALOGIAS E DEFINIÇÃO

Para o seu funcionamento adequado e assim prevenir o desenvolvimento de doenças autoimunes, o SIH precisa ser capaz de distinguir entre células e moléculas do próprio organismo, chamadas de moléculas próprias, e moléculas estranhas, que são geralmente denominadas apenas por próprio e não próprio, respectivamente (ABBAS; LICHTMAN; PILLAI, 2008; TIMMIS et al., 2004). De acordo com de Castro (2001), moléculas de anticorpos e receptores de células T possuem a capacidade de reconhecer qualquer molécula própria ou não-própria, incluindo aquelas sintetizadas artificialmente.

A eliminação de qualquer célula com receptores capazes de reconhecer moléculas próprias, denominadas células autorreativas, caracteriza o conceito de seleção negativa (DE CASTRO, 2001; DASGUPTA; NINO, 2008). O princípio de seleção negativa permite o controle dos linfócitos B e T, de forma que aqueles em desenvolvimento que se apresentem como potencialmente autorreativos, sejam eliminados (DE CASTRO; TIMMIS, 2002). A seleção negativa das células T ocorre dentro do Timo, que é responsável por sua maturação. Portanto, apenas células T que não reconhecem moléculas próprias são permitidas sobreviver (TIMMIS et al., 2004).

O ASN foi desenvolvido tendo como base o princípio de seleção negativa de células T que ocorre dentro do Timo. Por esta razão, ele possui a característica intrínseca de reconhecimento de padrões (DE CASTRO; TIMMIS, 2002; DASGUPTA; YU; NINO, 2011). Executado em duas fases, o ASN é definido como segue (FORREST et al., 1994):

1. Defina dois conjunto de dados diferentes $\mathbf{P}$ e $\mathbf{C}$. $\mathbf{P}$ contém apenas padrões próprios e $\mathbf{C}$ contém ambos os tipos de padrões, próprios e não próprios. Calcule a afinidade (*match*) entre cada elemento $c_j \in \mathbf{C}$ e cada elemento em $\mathbf{P}$. Se um elemento em $\mathbf{C}$ é reconhecido por um elemento em $\mathbf{P}$, em outras palavras, se a afinidade entre c_j e um elemento em $\mathbf{P}$ excede um limiar preestabelecido, rejeite c_j. Caso contrário, armazene c_j em um conjunto de detectores $\mathbf{R}$. Esta fase do ASN é comumente denominada de fase de Censoriamento.

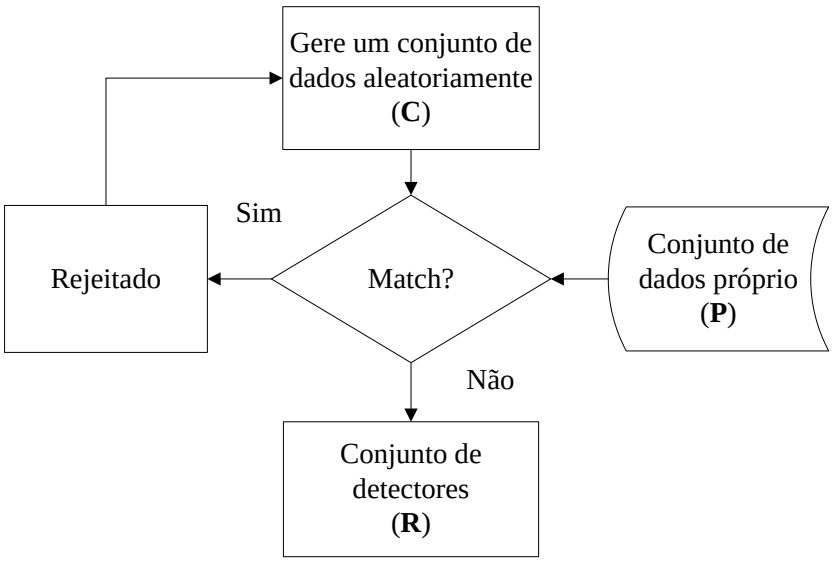
2. Após a geração de $\mathbf{R}$, a atual fase consiste no monitoramento do sistema a fim de detectar padrões não próprios. Um conjunto $\mathbf{P}_*$ é definido para ser protegido e a afinidade entre cada elemento em $\mathbf{P}_*$ e cada $r_j \in \mathbf{R}$ é avaliada. Se uma tal afinidade

é superior a um limiar predefinido, então um elemento não próprio é identificado. Esta fase é denominada fase de Monitoramento.

De acordo com de Castro e TIMMIS (2002), o conjunto $\mathbf{P}_*$ pode ser composto por um subconjunto de **P** com a adição de novos padrões ou um conjunto de dados completamente novos. O conjunto de dados **C** pode ser utilizado tanto na fase de censoriamento quanto na fase de monitoramento. Se assim for, os dados utilizados para gerar os detectores não devem ultrapassar 30% dos dados contidos em **C**, como sugerido em Forrest et al. (1994).

Os fluxogramas da fase de censoriamento e monitoramento do ASN são apresentados, respectivamente, nas Figuras 9 e 10.

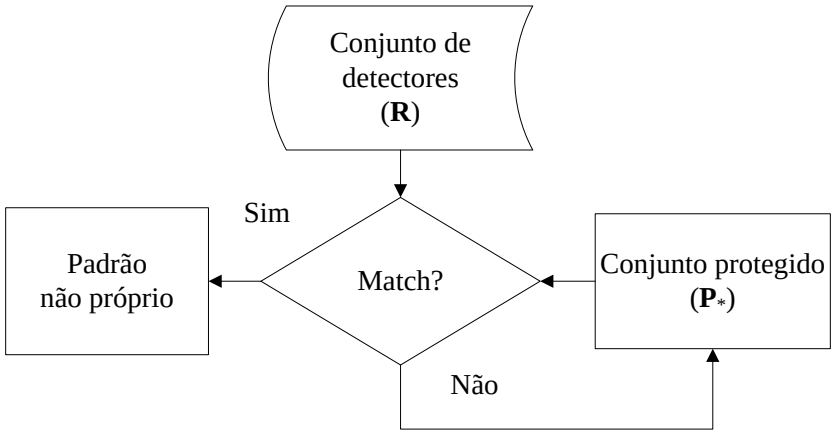
Figura 9 – Fluxograma da fase de censoriamento do ASN



Fonte: Adaptado de Forrest et al. (1994).

No ASN, detectores fazem referência às células T maturadas que podem reconhecer agentes patogênicos. Em outras palavras, detectores são similares a anticorpos (FORREST et al., 1993). Assim como no princípio de seleção negativa no SIH, os anticorpos autorreativos são descartados na fase de censoriamento, que possui como seu único objetivo a geração de um conjunto de detectores capazes de reconhecer padrões não próprios (agentes patogêncicos). A fase de censoriamento é realizada em modo *off-line*.

Figura 10 – Fluxograma da fase de monitoramento do ASN



Fonte: Adaptado de Forrest et al. (1994).

A fase de monitoramento ocorre em tempo real. O objetivo é monitorar e proteger o sistema por meio da identificação de padrões não próprios. A identificação é realizada tendo como base um conjunto de detectores, criados na fase de censoriamento por meio de uma medida de afinidade (DE CASTRO; TIMMIS, 2002; DASGUPTA; NINO, 2008).

4.3.1 Medidas de afinidade

Como mencionado na seção anterior, o ASN foi originalmente proposto para o problema de segurança computacional (FORREST et al., 1994). Em tal problema, padrões são *strings* ou *strings* de valores binários. Sendo assim, existem medidas de afinidades que acessam o número de diferentes caracteres entre duas *strings*, como a distância de Hamming, ou medidas que são baseadas no número relativo de bits que associam ou diferem duas *strings* (DASGUPTA; NINO, 2008). No entanto, as aplicações mais comuns são aquelas onde os padrões possuem valores reais. Na Tabela 2 são apresentadas as medidas de afinidade mais comuns utilizadas na representação por valores reais para o ASN (DASGUPTA; NINO, 2008). Note que Ab e Ag representam um detector e uma molécula estranha, respectivamente.

Na Tabela 2, a distância de Manhattan é também conhecida como distância de *city block*. Além disso, a distância de Minkowski, também conhecida como norma p, se torna a distância de Manhattan se $p = 1$ e a distância Euclidiana se $p = 2$.

Vale destacar que Ab e Ag são vetores de valores reais no espaço $\mathbb{R}^N$, que repre-

Tabela 2 – Principais medidas de afinidade para o algoritmo de seleção negativa a valores reais.

Medidas de afinidade	
medida	equação
Distância Euclidiana	$d(Ab, Ag) = \sqrt{\sum_i (Ab_i - Ag_i)^2}$
Distância de Manhattan	$d(Ab, Ag) = \sum_i \lvert Ab_i - Ag_i \rvert$
Distância de Minkowski	$d(Ab, Ag) = \left(\sum_i \lvert Ab_i - Ag_i \rvert^p \right)^{\frac{1}{p}}$
Distância de Chebyshev	$d(Ab, Ag) = max\{\lvert Ab_i - Ag_i \rvert \text{ para } i = 0, \ldots, n-1\}$
Distância de Canberra	$d(Ab, Ag) = \sum_i \frac{\lvert Ab_i - Ag_i \rvert}{\lvert Ab_i \rvert + \lvert Ag_i \rvert}$

Fonte: Elaborado pelo próprio autor.

sentam anticorpos e antígenos, respectivamente (DASGUPTA; NINO, 2008):

$$Ab = [Ab_0, Ab_1, \ldots, Ab_{i-1}, Ab_i, \ldots, Ab_{N-2}, Ab_{N-1}], \tag{61}$$

$$Ag = [Ag_0, Ag_1, \ldots, Ag_{i-1}, Ag_i, \ldots, Ag_{N-2}, Ag_{N-1}]. \tag{62}$$

A fim de obter um critério de associação (*matching*) e, assim, acessar a similaridade entre dois padrões, uma medida de afinidade adequada precisa ser escolhida de acordo com as características do problema estudado (DASGUPTA; NINO, 2008). O *matching* entre dois padrões pode ser perfeito ou parcial. No *matching* perfeito, os dois padrões são iguais, $d(Ab, Ag) = 0$. No entanto, em aplicações de valores reais o conceito de *matching* parcial é frequentemente empregado (JI; DASGUPTA, 2007; DASGUPTA; NINO, 2008). O *matching* parcial é implementado definindo um limiar de associação λ. Dessa forma, se $d(Ab, Ag) < \lambda$, o *matching* ocorre. Segundo Ji e Dasgupta (2007), o limiar de associação pode ser considerado uma generalização do sistema.

4.4 METODOLOGIA PROPOSTA

Nesta pesquisa, um algoritmo para a identificação e classificação de ruídos reais em sinais de voz baseado no ASN é proposto. O sistema irá operar nos intervalos de silêncio de sentenças de fala. A detecção de intervalos de silêncio e de fala é realizada no domínio wavelet usando o DAV proposto em Duarte, Vieira Filho e Alvarado (2009). Portanto, as

tomadas de decisão serão realizadas toda vez que uma janela de silêncio é detectada. A classificação, realizada janela por janela, é adequada para processamento de tempo real.

Condições normais de operação para o sistema ocorrem quando o sinal de voz em processamento está limpo, ou o nível de ruído é muito baixo. Dessa forma, padrões próprios são características extraídas a partir de sinais limpos, em ausência de voz. Características extraídas a partir de sinais de voz ruidosos via técnica de janelamento em segmentos de silêncio serão consideradas padrões não próprios.

Seis tipos de ambientes ruidosos reais são considerados. O banco de sinais utilizados é o mesmo descrito na seção 3.10. Sendo assim, os tipos de ruído são vozes, cafeteria, carro, salão de exibição, tráfego e trem. A fim de garantir que nenhum par de padrões repetidos sejam utilizados nas simulações, o processo de contaminação das sentenças puras foi iniciado aleatoriamente. Além disso, os dados são divididos em conjuntos de treinamento, desenvolvimento e teste. O conjunto de treinamento é utilizado para a geração de detectores e o conjunto de desenvolvimento é utilizado para a estimação do limiar de associação. Finalmente, o conjunto de teste é utilizado para simular a acurácia. Deste modo, os itens de teste são independentes.

Apesar da classificação ser realizada em segmentos de silêncio, diferentes sentenças de fala pronunciadas por diferentes oradores foram utilizadas para treinamento, desenvolvimento e teste. Metade dos dados, incluindo sentenças geradas por diferentes oradores em todos os níveis de SNR, foi utilizado para teste. O restante dos dados foi dividido em conjuntos de treinamento e desenvolvimento.

4.4.1 Fase de Censoriamento para o algoritmo proposto

No algoritmo proposto, a fase de censoriamento visa construir um conjunto de detectores de ruído $\mathbf{R}$. $\mathbf{R}$ deve conter detectores para todos os tipos de ruído considerados para classificação. Para este propósito, é proposto o seguinte procedimento:

1. Construa um conjunto de dados contendo amostras de ruído $\mathbf{C} = \{\mathbf{C_v}, \mathbf{C_c}, \mathbf{C_{ca}}, \mathbf{C_e}, \mathbf{C_t}, \mathbf{C_{tr}}\}$. Em $\mathbf{C}$, os índices representam, respectivamente, os ruídos: vozes, cafeteria, carro, salão de exibição, tráfego e trem. Use a técnica de janelamento;

2. Defina um conjunto de dados próprio $\mathbf{P}$ e carregue um subconjunto $\mathbf{C}_* \subset \mathbf{C}$ contendo amostras de ruído do tipo $*$. Defina o número de detectores para o tipo de ruído $*$ a ser armazenado;

3. Até que o número desejado de detectores seja obtido, escolha uma amostra de ruído em $\mathbf{C}_*$ aleatoriamente, extraia suas características e cheque o *matching* com todos os padrões próprios em $\mathbf{P}$. Se um *matching* não ocorre (não próprio identificado),

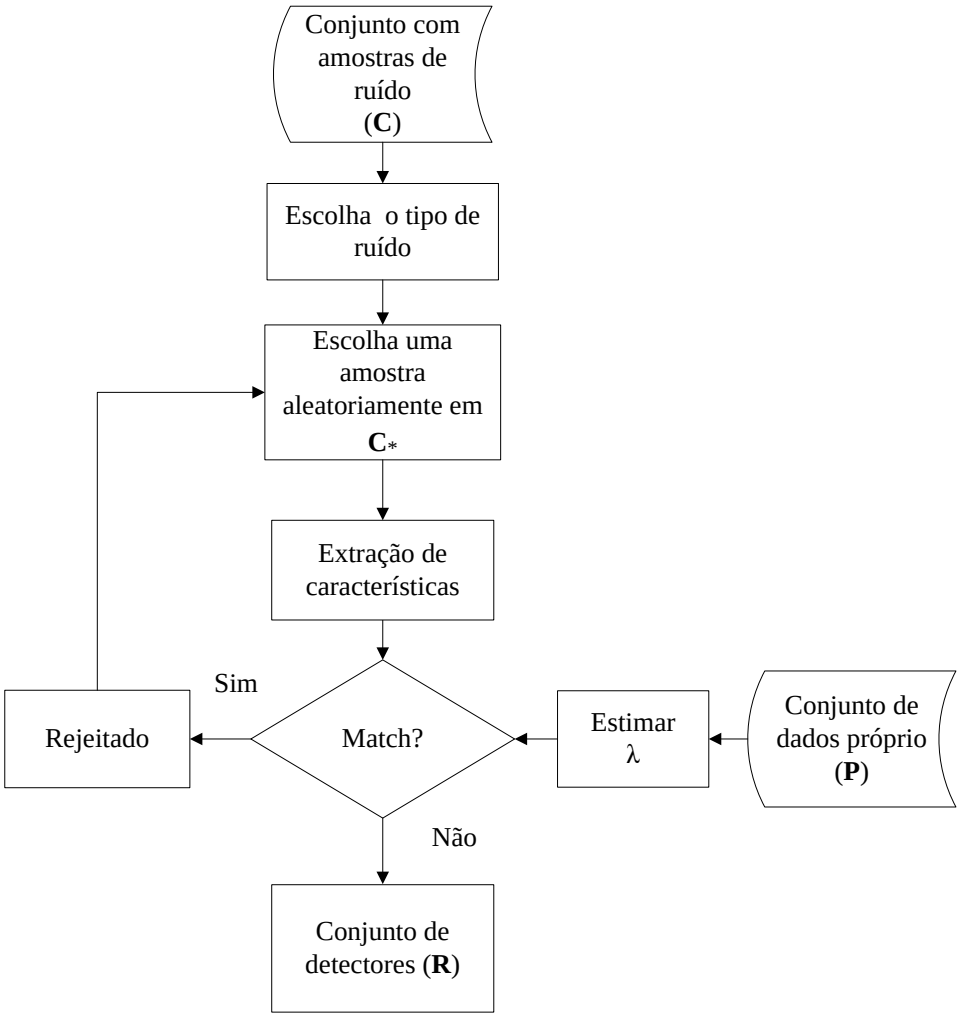
o padrão é armazenado em um conjunto de detectores. Caso contrário, rejeite a amostra de ruído atual e escolha um novo sinal aleatoriamente em $\mathbf{C}_*$;

4. Repita os passos 2 e 3 até que um conjunto de dados **R** seja obtido, contendo padrões para todos os tipos de ruído considerados no passo 1.

O conjunto de dados próprio **P** contém padrões com características extraídas a partir de sinais de voz puros. A fim de realizar a decisão próprio/não próprio (critério de *matching*), um limiar de associação λ é considerado. O nível do ruído em uma sentença pode variar com o tempo. Assim, a fase de censoriamento garante que padrões de ruído similares a padrões próprios não serão armazenados como detectores. Este procedimento garante a qualidade do conjunto de detectores. O limiar λ pode ser definido empiricamente pelo usuário, considerando a acurácia do sistema sobre o conjunto de desenvolvimento.

O fluxograma da fase de censoriamento é apresentado na Figura 11.

Figura 11 – Fluxograma da fase de censoriamento do método proposto.



Fonte: Elaborado pelo próprio autor.

A saída da fase de censoriamento é uma matriz contendo padrões para todos os tipos de ruído. Esta fase ocorre em modo *off-line* e o número de detectores armazenados é determinado pelo usuário. Vale destacar que a fase de censoriamento é executada sobre os dados de treinamento, uma vez para cada tipo de ruído.

4.4.2 Fase de Monitoramento para o algoritmo proposto

Após executar a fase de censoriamento, um conjunto não próprio de padrões de ruído $\mathbf{R} = \{\mathbf{R_v}, \mathbf{R_c}, \mathbf{R_{ca}}, \mathbf{R_e}, \mathbf{R_t}, \mathbf{R_{tr}}\}$ é gerado. $\mathbf{R}$ irá atuar como um conjunto de detectores (anticorpos) para cada tipo de ruído real considerado.

A fase de monitoramento deve atuar da seguinte maneira:

1. Carregue os conjuntos de dados próprio e não próprio $\mathbf{P}$ e $\mathbf{R}$, respectivamente. Considere um sinal de voz em processamento pela técnica de janelamento. É assumido que a primeira janela é um segmento de silêncio;

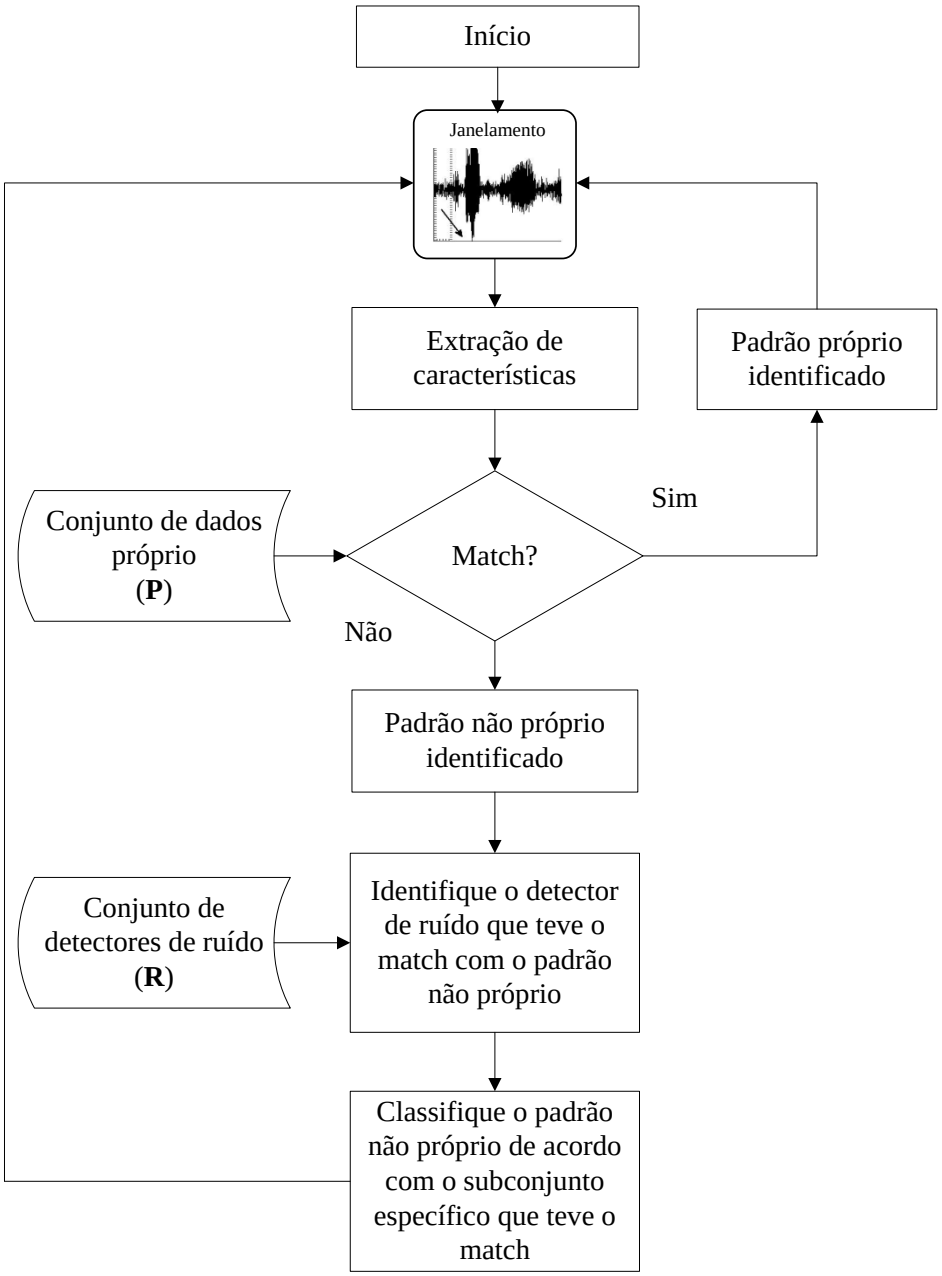
2. Extraia o vetor de características Ag (padrão desconhecido) a partir do conteúdo da janela e cheque o *matching* com $\mathbf{P}$: se $d(Ab, Ag) < \lambda$, para qualquer $Ab \in \mathbf{P}$ (padrão próprio identificado), vá para a próxima janela de silêncio; caso contrário, vá para o passo 3;

3. Calcule a afinidade entre Ag e cada padrão de ruído $r_i \in \mathbf{R}$. Identifique o detector de ruído que teve *matching*: $d(r_i, Ag) < \lambda$.

4. Se mais que um detector de ruído r_i se associa (*matches*) com o padrão Ag a ser classificado, o subconjunto $\mathbf{R}_* \subset \mathbf{R}$ com maior número de detectores ativados é selecionado para representar o padrão não próprio.

É importante destacar que a decisão próprio/não próprio é realizada da mesma maneira para ambas as fases de censoriamento e monitoramento. Se uma amostra própria é identificada, o algoritmo busca pela próxima janela de silêncio. Caso contrário, uma amostra não-própria é identificada e a fase de monitoramento é ativada a fim de classificar o ruído de fundo. Na Figura 12 é apresentado o fluxograma da fase de monitoramento.

Figura 12 – Fluxograma da fase de monitoramento do método proposto.



Fonte: Elaborado pelo próprio autor.

A fase de monitoramento do algoritmo proposto ocorre em tempo real e visa proteger o sistema identificando padrões não próprios. Para um sinal de voz em processamento, toda vez que uma janela de silêncio é identificada, a fase de monitoramento é ativada a fim de verificar se o sinal está corrompido por ruído. Em caso afirmativo, o sistema retorna o tipo do ruído de fundo.

No sistema proposto, o usuário é livre para decidir como extrair as características das amostras de ruído, o comprimento e tipo da janela. O usuário também pode escolher

a medida de afinidade e o limiar de associação λ.

Na seção 4.5, os resultados de várias simulações são apresentados. Além disso, detalhes sobre a extração de características e como acessar a afinidade são dados. Comparações com classificadores clássicos são também realizadas.

4.5 IMPLEMENTAÇÃO E RESULTADOS

Com o objetivo de avaliar o desempenho da metodologia proposta, várias simulações são realizadas nesta seção. O conjunto de dados próprios $\mathbf{P}$ consiste de 30 padrões extraídos a partir de sinais de voz puros, utilizando o janelamento retangular, em intervalos de silêncio. O limiar de associação λ, usado em ambas as fases de censoriamento e de monitoramento, foi definido empiricamente.

Utiliza-se, nesta pesquisa, um processo de extração de características baseado em wavelets complexas. A próxima subsubseção fornece mais detalhes sobre a dual-tree $\mathbb{C}$WT e características de entrada.

4.5.1 Extração de características

Após a extração de uma amostra de ruído $y[n] = [y_0, y_1, \ldots, y_{i-1}, y_i, \ldots, y_{N-2}, y_{N-1}]$ a partir do janelamento retangular com 1024 pontos, a seguinte normalização é realizada:

$$\hat{y}_i = \frac{y_i}{max(|y[n]|)}. \tag{63}$$

Em seguida, a dual-tree $\mathbb{C}$WT é aplicada sobre $\hat{y}_i$ com cinco níveis de decomposição. Assim, o vetor de características $Ab = [Ab_1, Ab_2, \ldots, Ab_5]$ é construído como segue:

$$Ab_k = \sum_{n=1}^{\frac{N}{2^k}} |d_c(k,n)|^2, \quad (k = 1, 2, \ldots, 5), \ \ (N = 1024), \tag{64}$$

sendo $d_c(k,n)$ os coeficientes wavelet complexos na k-ésima escala fornecida pela dual-tree $\mathbb{C}$WT, e N é o comprimento da amostra de ruído.

Note que o valor absoluto em (64) é calculado da seguinte forma:

$$|d_c(k,n)| = \sqrt{[d_r(k,n)]^2 + [d_i(k,n)]^2}. \tag{65}$$

Como mencionado no Seção 2.2, a cada nível de decomposição k, baixas e altas frequência são separadas sucessivamente. Portanto, o vetor de características de cinco dimensões em (64) é baseado na distribuição da energia do ruído através de cinco bandas de oitava. De acordo com Lina e Mayrand (1993), coeficientes de detalhes $d_c(k,n)$, fornecidos pela decomposição wavelet, correspondem a frequência situada aproximadamente

no intervalo $(2^{-k}f_s, 2^{-k-1}f_s)$, sendo f_s a frequência de amostragem do sinal. A Tabela 3 apresenta as faixas de frequência para cada escala wavelet k ($k = 1, \ldots, 5$), considerando o banco de sinais de voz utilizados neste trabalho.

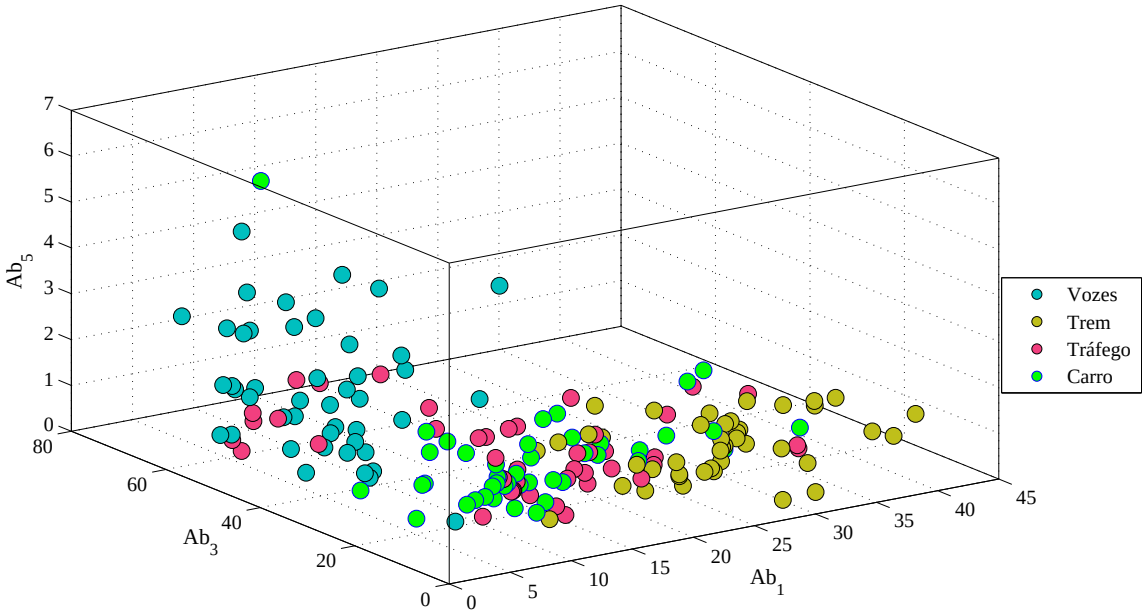
Tabela 3 – Faixas de frequência para cada escala wavelet k.

Escala wavelet	Faixa de Frequência (Hz)
1	2000 - 4000
2	1000 - 2000
3	500 - 1000
4	250 - 500
5	125 - 250

Fonte: Elaborado pelo próprio autor.

Para uma melhor compreensão das características extraídas, na Figura 13 é apresentado o gráfico de dispersão de 180 detectores gerados na fase de censoriamento do algoritmo proposto, divididos igualmente entre os ruídos de vozes, trem, tráfego e carro. Apenas as características Ab_k ($k = 1, 3$ e 5) são usadas.

Figura 13 – Gráfico de dispersão das características Ab_1, Ab_3 e Ab_5.



Fonte: Elaborado pelo próprio autor.

Note, a partir da Figura 13, que as características extraídas são consistentes. Os ruídos de vozes e trem são separados completamente entre si e os demais. Na verdade,

o ruído de vozes possui a maior parte de sua energia concentrada em baixas frequências, enquanto o ruído de trem possui forte concentração de energia em altas frequências (HIRSCH; PEARCE, 2000).

Finalmente, o janelamento com 1024 pontos foi fixado devido à arquitetura do banco de filtros utilizado na implementação da dual-tree $\mathbb{C}$WT, onde o número de pontos N deve ser uma potência de dois (MALLAT, 1998; SELESNICK; BARANIUK; KINGSBURY, 2005). Além disso, com 1024 pontos, um número considerável de coeficientes wavelet é obtido na quinta escala.

4.5.2 Resultados das simulações

Com o objetivo de alcançar a melhor configuração para o sistema proposto, várias simulações envolvendo o limiar de associação e as medidas de afinidade foram realizadas.

Na Figura 14 é mostrada a acurácia na classificação do ruído em termos da medida de afinidade e do limiar de associação λ, avaliada sobre o conjunto de desenvolvimento.

Analisando a acurácia na Figura 14, verifica-se que a melhor performance foi alcançada com a distância de Canberra quando $\lambda = 0,9$. As distâncias Euclidiana e de Manhattan forneceram resultados similares. A pior performance foi alcançada com a distância de Minkowski com $p = 3$.

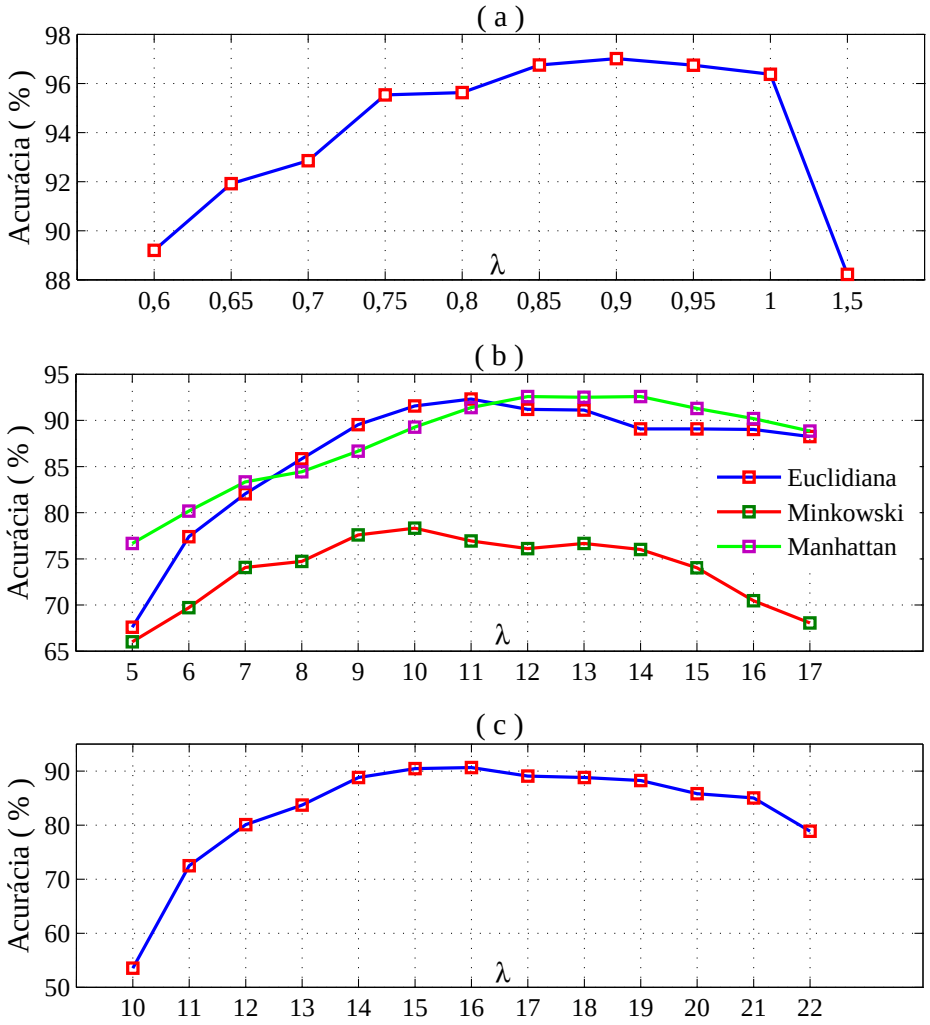
Tendo como base as curvas apresentadas na Figura 14, a distância de Canberra e o limiar de associação $\lambda = 0,9$ são selecionados como medidas de afinidade para os experimentos. A fim de fornecer mais detalhes sobre a performance do método proposto, a Tabela 4 apresenta a matriz de confusão para a classificação do ruído.

Tabela 4 – Matriz de confusão para a classificação do ruído usando a distância de Canberra com limiar de associação $\lambda = 0,9$.

		Classe predita					
	%	vozes	cafeteria	carro	s. exibição	tráfego	trem
Classe atual	vozes	**100**	0	0	0	0	0
	cafeteria	1,66	**98,33**	0	0	0	0
	carro	3,34	0	**91,66**	0	5	0
	s. exibição	0	0	0	**100**	0	0
	tráfego	0	0	7,33	0	**90**	2,67
	trem	0	0	0,49	1,66	0,08	**97,77**

Fonte: Elaborado pelo próprio autor.

Figura 14 – Acurácia na classificação do ruído em termos da medida de afinidade e do limiar de associação λ: (a) Distância de Canberra ($\lambda = 0,9$, acurácia= $97,01\%$); (b) Distância Euclidiana ($\lambda = 11$, acurácia= $92,31\%$), de Minkowski ($p = 3$, $\lambda = 10$, acurácia= $77,33\%$) e de Manhattan ($\lambda = 14$, acurácia= $92,68\%$); (c) distância de Chebyshev ($\lambda = 16$, acurácia= $90,64\%$).



Fonte: Elaborado pelo próprio autor.

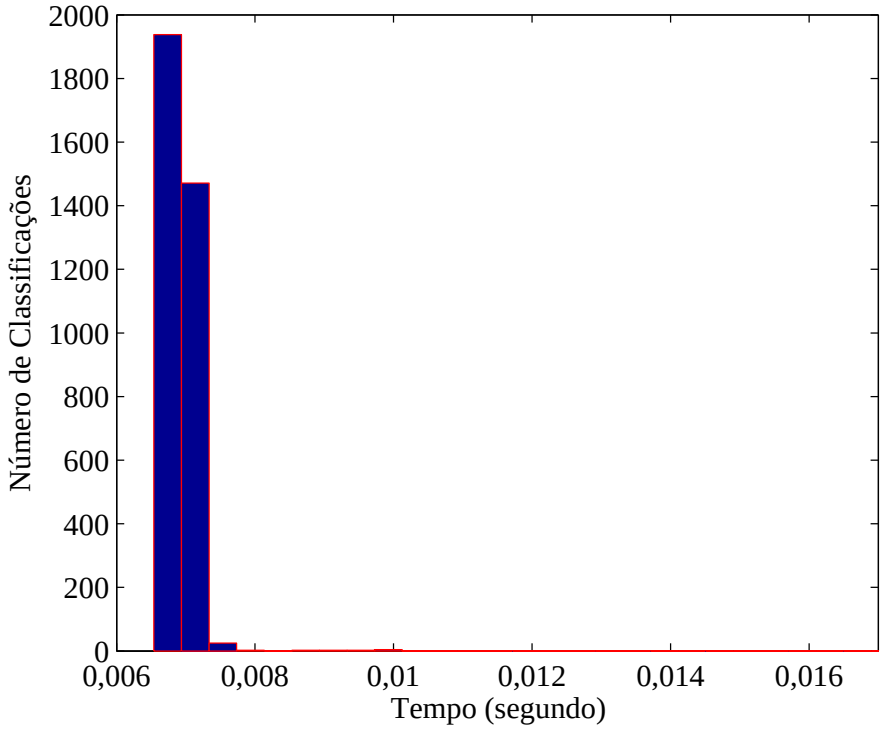
Na Tabela 4, as linhas correspondem à atual classe de ruído, enquanto as colunas correspondem à classe predita. Valores em negrito na diagonal representam a taxa de acerto na classificação do ruído.

De acordo com os dados da Tabela 4, o algoritmo proposto alcançou no mínimo 90% de taxa de sucesso para todos os tipos de ruído. De um modo geral, o classificador de ruídos reais alcançou, para os seis tipos de ruído, uma taxa de acerto de 96,29%. Note que as similaridades entre os padrões de carro e tráfego, destacados na Figura 13, influenciaram diretamente a acurácia do sistema. Juntas, estas duas classes são responsáveis pela maioria das predições incorretas.

Com relação a decisão próprio/não próprio, nenhuma amostra de ruído foi classificada como amostra própria. Portanto, o sistema alcançou uma taxa de acerto de 100% na discriminação próprio/não próprio.

Em aplicações de tempo real, o tempo utilizado em cada classificação deve ser o menor possível. Isto se deve ao fato de que um algoritmo de classificação atuará em conjunto com outros algoritmos, tal como melhoramento ou RAV. A fim de verificar o tempo requerido para cada classificação realizada pelo método proposto, o histograma para o tempo empregado em 3450 classificações é mostrado na Figura 15. Para cada decisão tomada, o processo descrito na Figura 12 foi executado.

Figura 15 – Histograma para o tempo de classificação requerido durante as simulações, totalizando 3450 classificações de ruído.



Fonte: Elaborado pelo próprio autor.

No histograma na Figura 15, 30 *bins* foram utilizados a fim de obter uma análise detalhada. Note que o tempo de execução é menor que oito milissegundos. Na verdade, o tempo de execução médio é 6,9 milissegundos para cada classificação. Portanto, baseado em seu baixo tempo de execução, o sistema proposto é adequado para processamento de temo real.

Na próxima subsubseção, o método proposto é comparado com classificadores clássicos, comumente utilizados em problemas de reconhecimento de padrões.

4.5.3 Comparações com classificadores clássicos

Para fins de comparações, o vetor de características gerado pela equação (64) será utilizado para treinamento e avaliação de métodos clássicos de aprendizagem supervisionada. Os classificadores utilizados são SVM (VAPNIK, 1995), rede neural perceptron multicamadas (RNPM) e árvore de decisão (AD) (BREIMAN et al., 1984). Uma breve descrição de cada abordagem é fornecida nos parágrafos seguintes.

O SVM implementado possui núcleo linear e usa o esquema um-contra-um. Como conhecido, SVMs são classificadores binários e discriminam entre duas classes possíveis. Para problemas com n-classes, o esquema um-contra-um consiste em aplicar $n(n-1)/2$ classificadores binários, onde cada classificador discrimina entre duas classes (BURGES, 1998).

A fim de classificar um padrão de ruído, o objetivo de uma AD é criar um modelo para predizer a classe verdadeira, a partir de um vetor de características, com base em simples regras de decisão inferidas a partir dos dados de treinamento. O algoritmo de árvore de decisão utilizado neste trabalho é o CART (do inglês: *Classification and Regression Tree*) (BREIMAN et al., 1984).

Finalmente, a RNPM possui uma camada de neurônios na entrada e uma camada na saída, podendo ter uma ou mais camadas intermediárias. As camadas intermediárias recebem o nome de camadas ocultas. Assim, o sinal de entrada se propaga através da rede, camada por camada, até que a camada de saída produza a resposta da rede para um determinado estímulo. Na presente implementação, o algoritmo de aprendizagem *backpropagation* e a função de ativação sigmoide foram utilizados (WERBOS, 1974). No estágio de treinamento, a taxa de aprendizagem e número de épocas foram definidas, respectivamente, como 0,3 e 500. Além disso, uma camada oculta com cinco neurônios foi usada. É importante destacar que o aprendizado da RNPM é do tipo supervisionado.

Os conjuntos de treinamento e teste utilizados para a avaliação dos classificadores SVM, RNPM e AD são os mesmos usados para avaliação do SIA proposto. Os resultados, mostrados na Tabela 5, foram adquiridos utilizando 360 sentenças corrompidas em todas as combinações de sinal puro-mais-ruído. As implementações foram realizadas em Python 2.7 e o pacote *scikit-learning* (PEDREGOSA et al., 2011) foi utilizado para as simulações dos classificadores SVM e AD. A RNPM foi implementada em Matlab® e o SIA proposto, em ambas as linguagens de programação.

Na Tabela 5, a última coluna apresenta a acurácia média para os métodos considerados e os valores em negrito indicam as melhores notas. A acurácia média foi calculada tomando a média das acurácias para todos os tipos de ruído.

Tabela 5 – Comparações entre performances considerando os classificadores clássicos e a aborgagem proposta.

Acurácia (%)							
classificador	vozes	cafeteria	carro	s.exibição	tráfego	trem	média
AD	95,00	93,33	90,00	**100,00**	88,33	91,11	92,96
RNPM	95,00	94,16	85,00	99,16	**91,66**	96,66	93,60
SVM	**100,00**	91,66	**91,66**	**100,00**	81,66	97,66	93,77
Proposto	**100,00**	**98,33**	**91,66**	**100,00**	90,00	**97,77**	**96,29**

Fonte: Elaborado pelo próprio autor.

Nota-se, a partir da Tabela 5, que o método proposto forneceu os melhores resultados para a maioria dos tipos de ruído. Apenas para as condições de ruído de tráfego ele foi superado por RNPM. O classificador SVM e o SIA proposto trabalharam igualmente bem para vozes, carro e salão de exibição. Com relação a performance global, o algoritmo proposto foi melhor, como pode ser visto na última coluna.

Baseado nas simulações realizadas neste capítulo, verificou-se que o SIA proposto é eficiente como uma ferramenta para identificação e classificação do ruído de fundo contido em sinais de voz, alcançando uma taxa de acerto médio de 96,29%. Apesar de uma análise multiescala baseada em wavelets ter sido usada no processo de extração de características, qualquer vetor de características pode ser utilizado como dados de entrada para o classificador.

Tendo a disposição um classificador treinado e ajustado, objetiva-se, no capítulo seguinte, a proposição de uma metodologia que leva em consideração o tipo de ruído presente no sinal. Dessa forma, um método robusto para MV será discutido.

5 MELHORAMENTO DE SINAIS DE VOZ BASEADO NA IDENTIFICAÇÃO DE PADRÕES RUIDOSOS

Exploram-se, neste capítulo, algumas das possibilidades e benefícios que o processamento baseado na classificação de ruído pode oferecer. Dessa forma, propõe-se, primeiramente, uma nova maneira de realizar a estimação de ruído para melhoramento de voz. O objetivo é responder a seguinte questão: "é possível realizar uma razoável estimação do ruído a partir do sinal de voz ruidoso por meio de regressão?" O método de regressão denominado mínimos quadrados parciais é utilizado e os resultados são confrontados com o algoritmo de estimação de ruído MCRA, que é amplamente conhecido e utilizado (COHEN; BERDUGO, 2002). A fim de alcançar melhores resultados, a classificação de ruído é incorporada no algoritmo e um modelo de regressão para cada tipo de ruído é construído. O principal benefício do algoritmo proposto incide sobre aplicações em tempo real: após construir um modelo de regressão, o ruído é estimado por meio de um simples produto entre um segmento de fala ruidoso e uma matriz de coeficientes de regressão. Assim, não é necessário usar algoritmos complexos para a estimação do ruído.

Em um segundo momento, e tendo como parâmetro todas as simulações e ferramentas desenvolvidas nesta pesquisa, uma nova metodologia para a área de MV é proposta. Denominada "Conjunto de Métodos de Melhoramento de Voz", esta metodologia consiste, a grosso modo, em utilizar o melhor método ou melhor configuração de um mesmo método de MV para cada tipo de ruído de fundo. Isso será possível por meio do uso de um classificador de ruído. Dessa forma, propõe-se um método robusto, com a capacidade de adaptar-se às características do ruído presente no sinal.

5.1 UM ALGORITMO DE ESTIMAÇÃO DE RUÍDO BASEADO EM REGRESSÃO PARA MELHORAMENTO DE VOZ COM CLASSIFICAÇÃO DO RUÍDO DE FUNDO

Um processo de melhoramento de voz consiste em extrair o sinal de voz limpo a partir do sinal de voz corrompido. Na maioria dos algoritmos de melhoramento de voz é necessária a estimação do perfil do ruído de fundo a fim de projetar um filtro supressor de ruído. Na verdade, quanto melhor a estimação do ruído, maior será a qualidade do sinal melhorado (BOLL, 1979; EPHRAIM; MALAH, 1984).

O método de estimação de ruído mais simples e frequentemente utilizado pela literatura é baseado sobre a média recursiva do espectro do ruído durante períodos de ausência de voz. Neste caso, períodos de ausência de voz são identificados automaticamente por um algoritmo DAV (SOHN; KIM; SUNG, 1999). A principal desvantagem de tal método consiste em rastrear mudanças no perfil do ruído durante segmentos vozeados.

A fim de melhorar a estimação do ruído sob condições de ruídos reais, o espectro do ruído deve ser atualizado continuamente ao longo do tempo. Nesse sentido, vários algoritmos de estimação de ruído têm sido propostos. Como exemplo, existem os algoritmos de estimação de ruído em Malah, Cox e Accardi (1999), Martin (2001), Cohen e Berdugo (2002) e Cohen (2003). O algoritmo proposto em Cohen e Berdugo (2002), denominado MCRA, tornou-se amplamente conhecido devido à sua capacidade de rastrear rapidamente mudanças abruptas no espectro do ruído mesmo em segmentos vozeados. Para este efeito, o ruído é estimado calculando a média dos valores espectrais de potência passados, usando um parâmetro de suavização que é ajustado pela probabilidade de presença da voz em sub-bandas. Quando a presença de voz é considerada fraca, o ruído é atualizado capturando possíveis mudanças no perfil do ruído durante segmentos vozeados. Desse modo, nenhuma distinção entre voz e silêncio é necessária.

Como mencionado anteriormente, em Yuan e Xia (2015), os autores utilizaram classificação de ruído para escolher empiricamente parâmetros de suavização ótimos para a estimação do ruído e da SNR_{prio}. Os resultados indicam que o ajuste do algoritmo de estimação de ruído para cada tipo de ruído de fundo é adequado.

Sendo assim, o principal objetivo desta seção é o desenvolvimento de um algoritmo de estimação de ruído baseado em regressão (ERBR) que trabalha igualmente na presença ou ausência de voz. O uso em tempo real é o principal benefício de um método de ERBR: após construir um modelo de regressão, o ruído pode ser estimado por meio de um simples produto entre um segmento de voz ruidoso e uma matriz de coeficientes de regressão. Assim, não é necessário o uso de algoritmos complicados para a estimação do ruído. Como exemplo, uma técnica de ERBR pode contribuir com aplicações no campo da robótica, onde uma das principais maneiras do usuário interagir com o robô é através de comandos de voz.

Em adição a construção de um único modelo de regressão, investiga-se a possibilidade de treinar vários modelos a fim de endereçar condições específicas de ruído, incorporando classificação do ruído de fundo.

5.1.1 Estimação de ruído baseado na regressão por mínimos quadrados parciais

Considere um sinal de voz ruidoso no domínio da frequência definido por

$$Y[l,n] = X[l,n] + W[l,n], \tag{66}$$

Aqui, o objetivo é estimar o n-ésimo componente de magnitude espectral do ruído na l-ésima janela, representado por $\widehat{W}_{l,n}$, a partir da magnitude do espectro do sinal de voz

contaminado $Y_{l,n}$ da seguinte maneira:

$$\widehat{W}_{l,n} = Y_{l,n}\mathbf{B} + E_{l,n}, \tag{67}$$

sendo $\mathbf{B}$ um modelo de regressão. Portanto, a estimação do ruído é realizada sem qualquer distinção entre presença ou ausência de voz. Note em (67) que $E_{l,n}$ representa o erro envolvido no processo de regressão.

No sentido dos mínimos quadrados, se $Y_{l,n}(n = 0, \ldots, N-1)$ representa uma janela do sinal de voz ruidoso de comprimento N, a magnitude do espectro do ruído é estimada como $\widehat{W}_{l,n} = f(Y_{l,n}; \theta_j)$, e é preciso minimizar

$$E_l^2 = \sum_{n=0}^{N-1} \left[W_{l,n} - \widehat{W}_{l,n}\right]^2 = \sum_{n=0}^{N-1} \left[W_{l,n} - f\left(Y_{l,n}; \theta_j\right)\right]^2, \tag{68}$$

onde θ_j é o j-ésimo parâmetro para a função f a ser estimado. Em outras palavras, a equação $\dfrac{\partial E_l^2}{\partial \theta_j} = 0$ $(j = 0, 1, \ldots)$ precisa ser resolvida.

De acordo com WOLD (1985), a alta dimensionalidade de $Y_{l,n}$ pode levar a um alto erro na predição. Nesse sentido, uma solução para este problema pode ser encontrada na análise de componentes principais (PCA - do inglês: *Principal Component Analysis*) (ABDI, 2003). Usando PCA, é possível projetar $Y_{l,n}$ sobre um novo espaço $\mathbf{M}$, com dimensionalidade reduzida obtendo fatores (ou componentes) $T = Y_{l,n}\mathbf{M}$. Segue que o ruído predito pode ser calculado como (JOLLIFFE, 2002):

$$\widehat{W}_{l,n} = T\mathbf{Q} + E = Y_{l,n}\mathbf{MQ} + E. \tag{69}$$

Note que $\mathbf{B} = \mathbf{MQ}$. O procedimento descrito acima caracteriza o método de regressão por componentes principais (PCR - do inglês: *Principal Components Regression*) (JOLLIFFE, 2002).

Vale ressaltar que na metodologia proposta, o modelo $\mathbf{B}$ é responsável por adquirir todo o conhecimento sobre as características espectrais do ruído. Em outras palavras, a matriz $\mathbf{B}$ irá substituir um algoritmo de estimação de ruído. Portanto, no estágio de treinamento, pares de treinamento $(Y_{l,n}, W_{l,n})$ são considerados. O objetivo é que o modelo possa aprender o máximo possível sobre o perfil do espectro do ruído. Note que $Y_{l,n}$ pode ser tanto um segmento de voz corrompido, quanto um segmento que contém apenas ruído. Desse modo, um modelo que representa a relação mútua entre pares de treinamento $(inst\hat{a}ncia, alvo)$ é adequado.

Considerando um modelo linear $\mathbf{W} = \mathbf{YB} + \mathbf{E}$, onde $\mathbf{W}$ é uma matriz alvo $(p \times m)$, $\mathbf{Y}$ é uma matriz $(p \times n)$ de variáveis preditoras, $\mathbf{B}$ é uma matriz $(n \times m)$ de coeficientes de regressão e $\mathbf{E}$ é uma matriz erro $(p \times m)$. A regressão por mínimos quadrados parciais (PLS - do inglês: *Partial Least Squares*) possui como principal característica predizer $\mathbf{W}$ a partir de $\mathbf{Y}$, descrevendo sua estrutura comum (ABDI, 2003). Além disso, a regressão PLS realiza

um procedimento de decomposição similar a PCA, buscando reduzir a dimensionalidade de **Y**. De acordo com Abdi (2003) e Jolliffe (2002), a regressão PLS busca por um conjunto de componentes (denominados vetores latentes) que realiza uma decomposição simultânea de **W** e **Y**, ao contrário da PCA, onde os componentes são escolhidos para explicar apenas **Y**. Portanto, a principal diferença entre a PCR e a regressão PLS é que em PCR, **M** reflete a variância de **Y**, enquanto na regressão PLS, **M** reflete a covariância entre **Y** e **W**. Com base nas suposições sobre os pares de treinamento e sobre as características da regressão PLS mencionadas acima, o método de regressão PLS é selecionado para construir o modelo **B** em (67).

Existem alguns métodos iterativos para a implementação da regressão PLS, dentre eles, destacam-se os algoritmos NIPALS (WOLD, 1985) e SIMPLS (JONG, 1993). De acordo com Jong (1993), SIMPLS é mais eficiente e menos caro computacionalmente do que NIPALS, e por esta razão será utilizado nessa implementação.

5.1.2 Avaliação da performance

Objetivando avaliar a performance do algoritmo proposto, a ERBR será utilizada para estimar a SNR_{prio}. Para fins de comparação, a mesma estimação para a SNR_{prio} será realizada utilizando algoritmo MCRA. O filtro supressor de ruído escolhido para ambos os algoritmos de estimação de ruído é o de Wiener (SCALART; VIEIRA FILHO, 1996).

Com o objetivo de treinar o modelo **B**, o banco de dados NOIZEUS foi dividido em conjuntos de treinamento e de teste. O conjunto de treinamento é composto por 15 sentenças pronunciadas por três oradores diferentes, em todas as combinações de voz-mais-ruído. As matrizes **W** e **Y** foram construídas sobre o conjunto de treinamento. Com relação a **W**, um DAV foi utilizado para obter apenas janelas de ruído. O comprimento da janela foi fixado em 256 pontos (32 milissegundos) e as amostras de ruído foram armazenadas aleatoriamente. O conjunto de teste é o mesmo utilizado no Capítulo 4. Durante o treinamento de **B**, o número de componentes utilizados na regressão PLS foi determinado a partir de uma validação cruzada com dez pastas.

Como mencionado anteriormente, em adição à construção de um único modelo de regressão, a construção de modelos de regressão específicos para cada condição ruidosa será abordada. Para este objetivo, o classificador de ruído proposto no Capítulo 4 será utilizado.

5.1.2.1 Filtro supressor de ruído

Após construir um modelo para cada tipo de ruído, o método de estimação de ruído proposto foi utilizado para estimar a SNR_{prio} usando o método decisão direta (EPHRAIM; MALAH, 1984). Seja $Y_{l,n}^2$ o n-ésimo componente espectral de potência na l-ésima janela,

ξ é calculado usando as equações (70) and (71) :

$$\hat{\xi}_{l,n} = min\left(\alpha\frac{\widehat{X}^2_{l-1,n}}{\widehat{W}^2_{l-1,n}} + (1-\alpha)P\left[\frac{Y^2_{l,n}}{\widehat{W}^2_{l,n}} - 1\right], \xi_{min}\right) \tag{70}$$

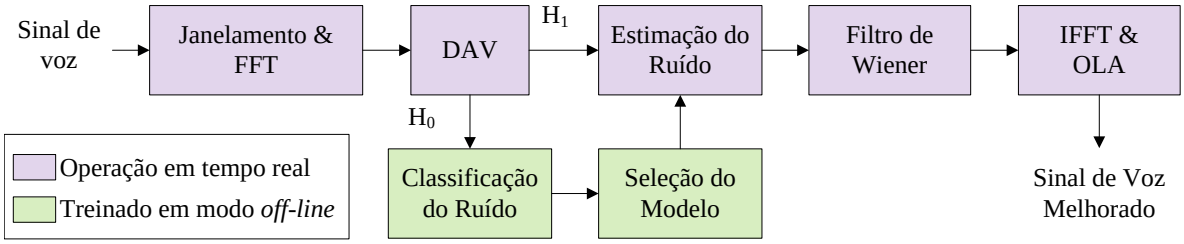
$$\widehat{W}^2_{l,n} = \left[\delta\left(\frac{Y_{l,n}}{max\left(Y_{l,n}\right)}\mathbf{B}\right)\right]^2, \tag{71}$$

sendo que $\widehat{X}^2_{l-1,n}$ é o espectro de potência processado na janela precedente, $\xi_{min} = -20$ dB, $\alpha = 0,9$ é o parâmetro de suavização, $P[.]$ corresponde a retificação de meia onda e δ é a média dos valores máximos para a magnitude espectral do ruído atualizado em intervalos de silêncio. Segue que o filtro Wiener é definido como (EPHRAIM; MALAH, 1984; SCALART; VIEIRA FILHO, 1996):

$$G_{l,n} = \frac{\hat{\xi}_{l,n}}{\hat{\xi}_{l,n} + 1}. \tag{72}$$

Note em (71) que para a estimação do ruído é necessário que $Y_{l,n}$ varie no intervalo $[0, 1]$. A normalização adotada em (71) foi também utilizada no estágio de treinamento. Uma visão geral do esquema de MV proposto é apresentada na Figura 16. Blocos em verde indicam que um treinamento em modo *off-line* é necessário.

Figura 16 – Uma visão geral do esquema de melhoramento de voz com estimação de ruído baseado em regressão e na classificação do ruído.



Fonte: Elaborado pelo próprio autor.

Note na Figura 16 que a saída do bloco DAV é H_0 ou H_1 para a ausência ou presença de voz, respectivamente. É importante frisar que todos os blocos da Figura 16 são utilizados durante o processo de MV. As cores são utilizadas apenas para destacar que o classificador de ruído e o modelo de regressão precisam ser treinados em modo *off-line*.

5.1.2.2 Resultados da simulação

A performance do algoritmo de ERBR proposto é avaliada por meio de medidas de qualidade objetivas que acessam o nível de redução do ruído, a distorção de fundo e a qualidade geral da sentença processada. Para avaliar o nível de ruído na sentença

processada, utiliza-se a SNR segmentada (segSNR). Opta-se nesta seção pela segSNR, ao invés da SNR global, pois a primeira é mais sensível às variações do nível de ruído em todo o sinal. A segSNR é obtida pelo cálculo da SNR convencional em vários períodos de curta duração do sinal, para então tomar a média aritmética desses valores. Para a avaliação da distorção de fundo, ou seja, para avaliar o quanto o ruído residual no sinal processado pode ser irritante (desconfortável) ao ouvinte, utiliza-se a medida composta $\mathrm{C_{bak}}$ em (73). Por fim, para avaliar a qualidade geral da sentença processada, utiliza-se a medida composta $\mathrm{C_{ovl}}$ em (74).

$$\mathrm{C_{bak}} = 1,634 + 0,478\mathrm{PESQ} - 0,007\mathrm{WSS} + 0,063\mathrm{segSNR}, \tag{73}$$

$$\mathrm{C_{ovl}} = 1,594 + 0,805\mathrm{PESQ} - 0,512\mathrm{LLR} - 0,007\mathrm{WSS}, \tag{74}$$

onde LLR and WSS são o logaritmo da razão de verossimilhança (do inglês: *Log Likelihood Ratio*) (QUACKENBUSH; BARNWELL; CLEMENTS, 1988) e a distância espectral de inclinação ponderada (do inglês: *Weighted-Slope Spectral Distance*) (KLATT, 1982), respectivamente.

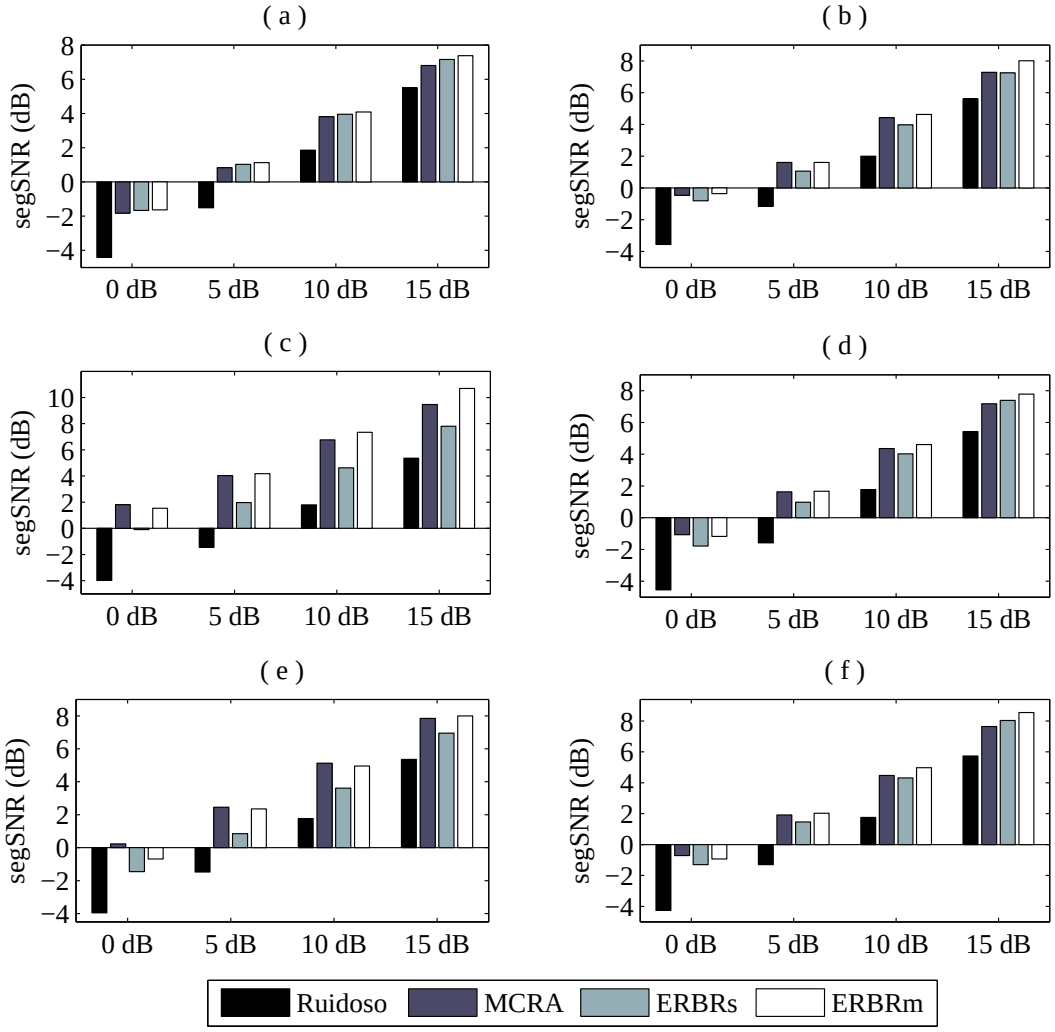
As medidas compostas $\mathrm{C_{ovl}}$ e $\mathrm{C_{bak}}$ foram desenvolvidas em Hu e Loizou (2008a), a partir de uma extensa análise subjetiva de sinais contaminados e processados. Tendo como base a análise subjetiva, verificou-se o quanto as medidas objetivas de qualidade estavam correlacionadas com as avaliações dos ouvintes. Como resultado final dessa pesquisa foram obtidas as medidas em (73) e (74). A medida $\mathrm{C_{ovl}}$ leva em consideração tanto a distorção de fundo quando a distorção da voz. Ambas as medidas compostas emitem uma pontuação que varia no intervalo [1,5], seguindo a escala MOS (do inglês: *Mean Opinion Score*).

Na Figura 17 são mostrados os resultados em termos de segSNR para o filtro de Wiener em (72), incorporando a ERBR e o método já bem estabelecido MCRA. As abreviações ERBRs e ERBRm referem-se a ERBR que utiliza um único modelo de regressão (sem classificação de ruído) e vários modelos (com classificação de ruído), respectivamente. Todos os valores apresentados nesta seção representam a média das medidas de qualidade calculadas sobre as sentenças no conjunto de teste.

Nota-se, a partir da Figura 17, que, para o ruído Vozes, ERBRs e ERBRm geraram resultados ligeiramente melhores do que MCRA, em termos de segSNR, para todas as condições de SNR. No entanto, para o restante dos tipos de ruído ERBRs trabalhou pior do que ERBRm e MCRA para a maioria das condições de SNR, exceto para salão de exibição e trem, ambos com 15 dB. Com relação a ERBRm e MCRA, o primeiro trabalhou ligeiramente melhor para todas as condições de ruído, exceto para o ruído de tráfego. Para a condição de ruído de tráfego, MCRA gerou os melhores resultados para todas as condições de SNR, exceto para 15 dB no qual ambos trabalharam igualmente bem.

Na Tabela 6 são apresentadas as avaliações em termos de $\mathrm{C_{bak}}$ e $\mathrm{C_{ovl}}$ para os sinais

Figura 17 – segSNR para os sinais ruidosos e processados usando os algoritmos de estimação de ruído MCRA, ERBRs e ERBRm sob as condições de ruído (a) vozes, (b) cafeteria, (c) carro, (d) salão de exibição, (e) tráfego e (f) trem.



Fonte: Elaborado pelo próprio autor.

melhorados. Valores em negrito destacam os métodos mais eficientes para cada nível de contaminação.

Analisando as notas C_{bak} e C_{ovl} na Tabela 6, verifica-se que o método ERBRm gerou os melhores resultados para a maioria das condições de ruído, exceto para o ruído de tráfego em todas as condições de SNR e para o ruído de carro 5 dB. Além disso, MCRA trabalhou melhor que ERBRs para todas as condições de ruído, exceto para o ruído Vozes em todos os níveis de SNR.

Verifica-se, a partir da Tabela 6 e da Figura 17, que o método de ERBR incorporando classificação de ruído trabalhou ao menos igual ao método MCRA em termos de atenuação do ruído. Além disso, em adição aos resultados satisfatórios na atenuação

Tabela 6 – Distorção do ruído residual de fundo e qualidade geral das sentenças processadas

		C_{ovl}				C_{bak}			
Ruído	Método	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
Vozes	Ruidoso	2,16	2,59	3,00	3,42	1,79	2,20	2,64	3,09
	MCRA	1,85	2,38	2,90	3,43	1,73	2,19	2,68	3,16
	ERBRs	**1,95**	2,48	2,95	3,45	**1,81**	2,27	2,73	3,21
	ERBRm	**1,95**	**2,50**	**2,98**	**3,50**	1,80	**2,28**	**2,75**	**3,25**
Cafeteria	Ruidoso	2,24	2,60	3,02	3,46	1,84	2,19	2,62	3,10
	MCRA	2,03	2,46	2,97	3,44	1,89	2,27	2,74	3,20
	ERBRs	1,98	2,40	2,88	3,43	1,83	2,20	2,66	3,18
	ERBRm	**2,11**	**2,54**	**3,05**	**3,60**	**1,93**	**2,32**	**2,80**	**3,33**
Carro	Ruidoso	2,39	2,78	3,21	3,62	1,90	2,27	2,71	3,16
	MCRA	**2,83**	3,17	3,62	3,94	**2,47**	2,82	3,26	3,63
	RBNEs	2,40	2,79	3,27	3,74	2,16	2,51	2,95	3,41
	ERBRm	2,72	**3,18**	**3,69**	**4,15**	2,44	**2,86**	**3,35**	**3,82**
S. Exibição	Ruidoso	1,92	2,35	2,78	3,20	1,77	2,15	2,57	3,01
	MCRA	1,76	2,36	2,87	3,29	1,75	2,24	2,70	3,13
	ERBRs	1,75	2,29	2,81	3,24	1,74	2,18	2,66	3,11
	ERBRm	**1,87**	**2,43**	**2,92**	**3,34**	**1,81**	**2,29**	**2,76**	**3,21**
Tráfego	Ruidoso	1,76	2,11	2,56	2,99	1,70	2,01	2,44	2,89
	MCRA	**1,87**	**2,37**	**2,87**	**3,23**	**1,93**	**2,32**	**2,76**	**3,17**
	ERBRs	1,51	1,96	2,35	2,88	1,63	1,98	2,38	2,89
	ERBRm	1,70	2,34	2,68	3,10	1,77	2,29	2,65	3,09
Trem	Ruidoso	1,99	2,34	2,74	3,18	1,83	2,18	2,57	3,03
	MCRA	1,99	2,46	2,91	3,31	1,98	2,40	2,79	3,19
	ERBRs	2,07	2,45	2,85	3,29	1,99	2,38	2,75	3,21
	ERBRm	**2,12**	**2,56**	**3,02**	**3,46**	**2,02**	**2,45**	**2,88**	**3,34**

Fonte: Elaborado pelo próprio autor.

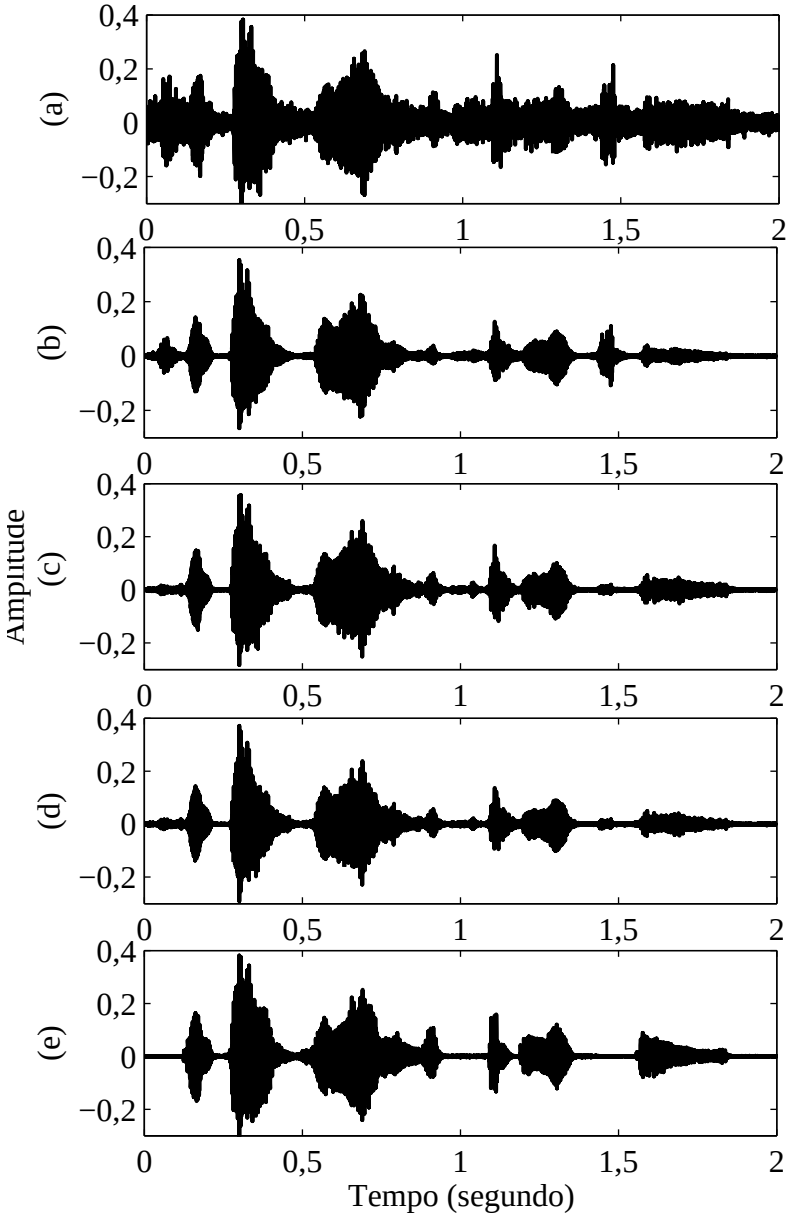
do ruído, o método proposto alcançou melhor qualidade e inteligibilidade de voz. Vale destacar que a medida de qualidade C_{ovl} é baseada principalmente sobre a PESQ, que é razoável na predição da inteligibilidade de sentenças (MA; HU; LOIZOU, 2009).

A superioridade do método MCRA sobre a ERBR para as condições de ruído de tráfego, indicam que a regressão PLS não foi capaz de generalizar sobre este tipo de ruído. Sendo assim, uma maior quantidade de ruído residual foi gerada na sentença processada. Este fato é confirmado na Figura 17. Em trabalhos futuros, este problema

pode ser endereçado por meio do estudo e aplicação de algoritmos de regressão com alto grau de não-linearidade, como em Xu et al. (2015).

Para fins de ilustração, na Figura 18 são apresentadas as formas de onda para um sinal de voz contaminado e sua versão processada utilizando os métodos de estimação de ruído MCRA, ERBRs e ERBRm, respectivamente.

Figura 18 – Forma de onda para (a) sinal de voz corrompido por ruído de trem 5 dB (segSNR=-1,62, C_{bak}=2,28, C_{ovl}=2,58); (b) sinal melhorado com MCRA (segSNR=1,46, C_{bak}=2,35, C_{ovl}=2,48); (c) sinal melhorado com ERBRs (segSNR=1,42, C_{bak}=2,44, C_{ovl}=2,67); (d) sinal melhorado por ERBRm (segSNR=1,90, C_{bak}=2,53, C_{ovl}=2,76); (e) sinal puro.

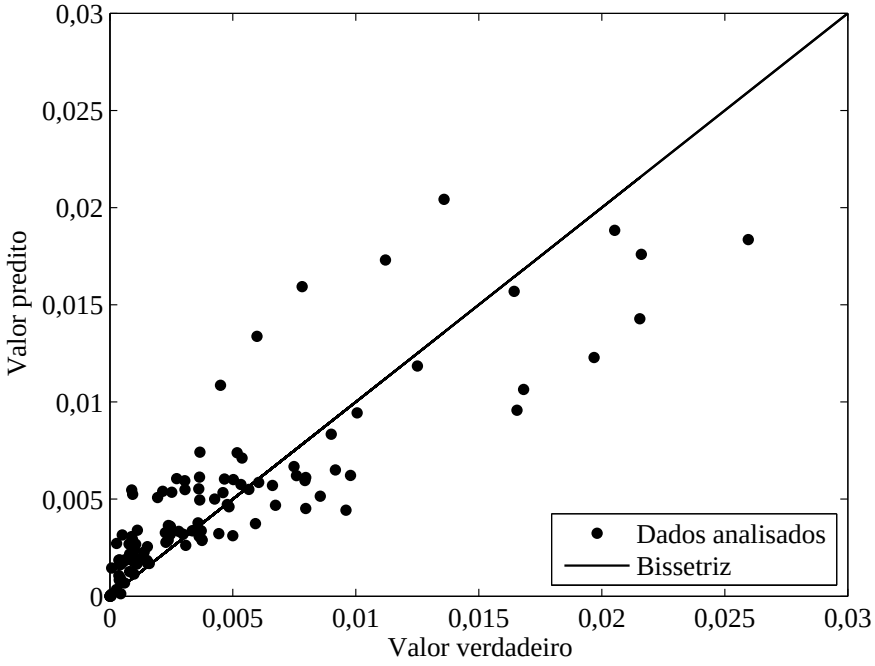


Fonte: Elaborado pelo próprio autor.

Nota-se, a partir da Figura 18, que todos os métodos de estimação de ruído forneceram bons resultados em termos de atenuação do ruído, com ligeira vantagem para ERBRs e ERBRm. No entanto, ambos os métodos baseados em regressão trabalharam melhor sobre segmentos de ruído impulsivo. Em termos de medidas de qualidade objetivas, o melhor método para o sinal em questão foi RBNEm.

É apresentado na Figura 19 um exemplo de regressão efetuada sobre um sinal de voz contaminado. Para a construção desse exemplo, um segmento de um sinal de voz puro foi contaminado com ruído de vozes. Nesse caso, o conhecimento *a priori* do ruído adicionado foi necessário.

Figura 19 – Exemplo de regressão.



Fonte: Elaborado pelo próprio autor.

Analisando a Figura 19, verifica-se que se a predição fosse exata (situação ideal), os pontos plotados no plano seguiriam a bissetriz do quadrante. 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 a estimação de ruído proposta apresenta resultados satisfatórios, uma vez que grande parte dos pontos plotados estão sobre ou muito próximo à bissetriz.

Os resultados apresentados nesta seção indicam que é possível realizar uma razoável estimação do ruído, a partir do sinal corrompido, por meio de regressão. A principal vantagem do método proposto é o baixo custo computacional no processo de estimação do ruído.

5.2 UMA NOVA METODOLOGIA: CONJUNTO DE MÉTODOS DE MELHORAMENTO DE VOZ

Levando em consideração os estudos e as discussões apresentadas nesta pesquisa, assim como as simulações realizadas no Capítulo 3, propõe-se nesta seção uma nova metodologia para a área de melhoramento de voz. Denominada Conjunto de Métodos de Melhoramento de Voz (CMMV), esta metodologia utiliza da premissa de que o desempenho de algoritmos de MV pode variar substancialmente dependendo do tipo do ruído de fundo. Sendo assim, o objetivo é valer-se da possibilidade da classificação do ruído para então propor um método robusto, que trabalhe de maneira eficiente em qualquer ambiente ruidoso. Assim como destacado nos Capítulos 1 e 4, o desenvolvimento de algoritmos que incorporam classificação de ruído é uma tendência.

A metodologia consiste em duas etapas: construção e operação.

- **Construção:** Construa um banco de sentenças contaminadas com vários tipos de ruído e divida este banco de dados em dois conjuntos, um para treinamento e outro para teste. Ambos os conjuntos devem conter sinais contaminados com todos os tipos de ruídos considerados. Para cada tipo de ruído, avalie todos os métodos de melhoramento de voz implementados adotando algum critério de avaliação (objetivo ou subjetivo). Quando possível, ajuste os parâmetros de um mesmo algoritmo de maneira diferente para cada condição de ruído a fim de obter parâmetros ótimos. Diferentes configurações de um mesmo algoritmo de MV são considerados métodos diferentes dos demais. Realize simulações utilizando a validação cruzada sobre o conjunto de treinamento e, com base nas simulações, selecione o melhor método para cada tipo de ruído e construa um conjunto de métodos de MV. Ainda nesta etapa, um classificador de ruído deve ser treinado sobre o conjunto de treinamento. Todas as simulações são realizadas em modo *off-line*.

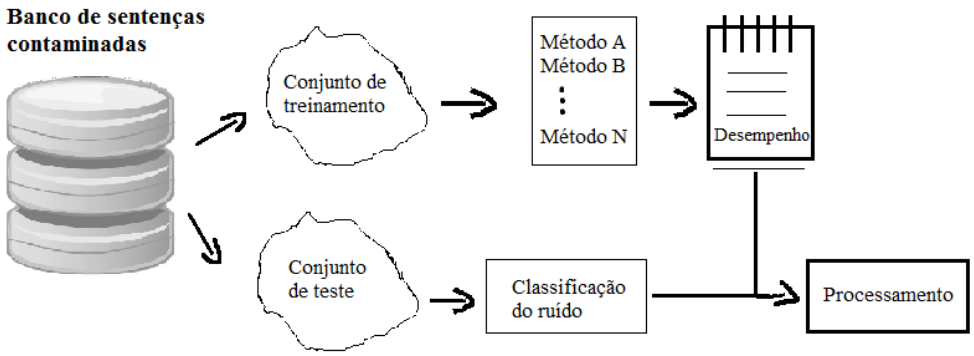
- **Operação:** Para todas as sentenças no conjunto de teste, utilize o classificador de ruído para predizer o tipo de ruído presente no sinal de voz em processamento. A classificação do ruído pode ser realizada no trecho inicial de cada sentença, ou toda vez que um intervalo de silêncio é detectado. Escolha o melhor algoritmo de MV para o tipo de ruído atual e realize o processamento.

É importante ressaltar que, se a classificação do ruído é feita apenas em um segmento inicial do sinal contaminado, assume-se que o tipo de ruído não mudará até o final da sentença. A classificação realizada apenas no início da sentença é coerente do ponto de vista de várias aplicações. É comum, em sistemas de RAV, que o software solicite ao usuário a instrução por comando de voz. Do momento da solicitação, até o usuário emitir

a instrução, existe um atraso onde apenas o ruído acústico do ambiente é captado. No caso do classificador desenvolvido nesta pesquisa, o atraso necessário para o sistema extrair as características é de 32 milissegundos, mais 6,9 milissegundos para emitir a classificação, totalizando aproximadamente 39 milissegundos. A classificação realizada janela por janela em intervalos de silêncio é mais adequada para sistemas de processamento de voz de fluxo contínuo, como por exemplo, sistemas de telecomunicações.

O esquema de melhoramento de voz proposto é ilustrado na Figura 20. Note que durante a fase de construção não é necessário utilizar o classificador de ruído. Neste caso, assume-se a condição ideal onde o tipo do ruído de fundo é conhecido. Cada método de MV recebe uma posição em uma lista segundo seu desempenho. O melhor método para cada tipo de ruído é então selecionado para ser utilizado durante a fase de operação. Nesta fase, uma vez que o tipo de ruído de fundo é identificado, escolhe-se na lista o método mais adequado para o processamento.

Figura 20 – Ilustração do esquema de melhoramento de voz baseado em um CMMV.



Fonte: Elaborado pelo próprio autor.

Nas próximas subseções, explica-se em detalhes como o experimento foi conduzido.

5.2.1 Fase de construção

A fim de realizar simulações e verificar se a abordagem baseada em um CMMV é adequada, o banco de dados NOIZEUS foi dividido em conjuntos de treinamento e de teste. Tais conjuntos contêm as mesmas sentenças utilizadas para treinar e validar a ERBR, na seção 5.1. Além disso, o conjunto de teste contém as mesmas sentenças utilizadas para a validação do classificador de ruído no Capítulo 4.

Baseado na análise objetiva de diferentes algoritmos de MV, apresentada no Capítulo 3, considera-se na fase de construção os 4 melhores métodos: MMSE, WIENER, SS e comp-wav. Sendo assim, o objetivo da fase de construção passa a ser o de encontrar a melhor configuração para cada um dos quatro métodos, frente a condições específicas de ruído.

Com o objetivo de obter parâmetros e melhores configurações para cada algoritmo, o parâmetro α na equação (70), responsável pela suavização da estimação da SNR_{prio}, e o parâmetro τ, responsável por controlar a inclinação da função sigmoide na equação (58), foram ajustados para cada tipo de ruído. O parâmetro α controla a dependência da estimação da SNR_{prio} na janela precedente, para com a estimação da SNR_{prio} na janela atual pelo método ML, sendo o principal parâmetro a ser ajustado para o método decisão direta. Quanto maior for o valor de α, maior será a atenuação do ruído, reduzindo assim o ruído residual. Responsável por controlar a inclinação da função sigmoide, que é utilizado como a função ganho para o método comp-wav, quanto menor o valor de τ, maior será a atenuação do ruído. Como o ruído é distribuído pelas escalas wavelet de acordo com seu conteúdo de frequência, é adequado que se defina diferentes valores de τ para diferentes escalas k.

Sendo assim, várias simulações foram realizadas com os quatro métodos destacados anteriormente, visando obter os melhores valores para α e $\tau_j (j = 1, 2, ..., 5)$. Além disso, cada método foi calibrado com ambos os algoritmos de estimação de ruído, MCRA e ERBRm. A única exceção foi o método comp-wav, que possui um mecanismo próprio de estimação de ruído. Para a calibração de todos os algoritmos, utilizou-se apenas sentenças com um nível de contaminação de 5 dB presentes no conjunto de treinamento. Nas Tabelas 7 e 8 são apresentados os melhores valores para os parâmetros calibrados para cada tipo de ruído.

Tabela 7 – Melhor configuração para o parâmetro α.

Algoritmo	Vozes	Cafeteria	Carro	S. Exibição	Tráfego	Trem
WIENER-MCRA	0,95	0,92	0,94	0,95	0,88	0,92
WIENER-ERBR	0,93	0,93	0,90	0,93	0,95	0,94
MMSE-MCRA	0,97	0,95	0,95	0,95	0,97	0,95
MMSE-ERBR	0,97	0,96	0,95	0,95	0,97	0,95

Fonte: Elaborado pelo próprio autor.

Verifica-se na Tabela 7 que os parâmetros ajustados não se aplicam ao método SE, pois sua implementação não utiliza a estimação da SNR_{prio}. Considerou-se α variando no intervalo $[0,85, 0,97]$. Como o algoritmo MMSE apresenta uma curva de atenuação mais suave que WIENER, a melhor configuração foi encontrada com valores mais elevados para α. A calibração deste parâmetro é importante, pois ele pode levar a uma atenuação excessiva do ruído, podendo inserir distorções no sinal processado.

Após a calibração dos algoritmos, simulações foram realizadas sobre o conjunto de treinamento, para todos os níveis de contaminação, a fim de escolher os melhores métodos para compor o CMMV. As medidas de qualidade escolhidas foram a segSNR e C_{ovl}. Para

Tabela 8 – Melhor configuração para os parâmetros τ_j.

Ruído	τ_1	τ_2	τ_3	τ_4	τ_5
Vozes	160	45	20	18	20
Cafeteria	200	45	10	18	15
Carro	200	45	10	18	15
S. Exibição	160	45	20	18	20
Tráfego	100	20	15	25	25
Trem	70	25	15	25	25

Fonte: Elaborado pelo próprio autor.

fins de construção do CMMV, essas duas medidas fornecem uma análise adequada, pois avaliam os sinais processados em relação a dois aspectos importantes: nível do ruído residual e qualidade geral dos sinais processados. Na Tabela 9 constam os resultados das simulações.

Tabela 9 – Teste de redução de ruído e qualidade geral dos sinais processados pelos algoritmos ajustados para cada tipo de ruído.

		sSNR				C_{ovl}			
Ruído	Método	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
Vozes	WIENER-MCRA	-2,97	0,63	3,48	6,12	1,74	2,31	2,87	3,31
	WIENER-ERBRm	-1,76	**0,77**	**3,70**	**6,76**	1,86	2,39	2,94	3,38
	MMSE-MCRA	-2,97	-0,10	2,85	5,67	1,95	2,47	2,97	3,37
	MMSE-ERBRm	-2,71	-0,19	2,64	5,74	**2,03**	**2,51**	**2,99**	**3,39**
	SE-MCRA	-2,73	-0,82	1,05	2,63	1,47	1,96	2,48	2,84
	SE-ERBRm	**-0,71**	-0,30	0,76	2,03	0,66	1,76	2,59	3,02
	comp-wav	-4,06	-0,28	3,14	5,12	1,90	2,38	2,87	3,13
Cafeteria	WIENER-MCRA	-1,07	1,19	3,85	6,80	1,89	2,31	2,86	3,31
	WIENER-ERBRm	-0,78	**1,36**	**4,26**	**7,41**	1,97	2,38	2,99	3,40
	MMSE-MCRA	-1,97	0,42	3,15	6,12	2,06	2,43	2,94	3,35
	MMSE-ERBRm	-1,85	0,43	3,24	6,52	**2,13**	**2,48**	**3,02**	**3,41**
	SE-MCRA	-2,04	-0,58	1,13	2,71	-2,04	1,94	2,47	2,82
	SE-ERBRm	**-0,36**	0,00	0,96	2,30	1,09	1,86	2,62	3,02
	comp-wav	-3,17	-0,90	1,09	2,27	1,83	2,01	2,41	2,72
Carro	WIENER-MCRA	**1,89**	**3,84**	6,70	9,29	**2,74**	**3,08**	3,50	3,87
	WIENER-ERBRm	1,35	3,80	**6,89**	**10,35**	2,57	3,01	**3,52**	**4,00**
	MMSE-MCRA	-0,44	1,78	4,80	7,81	2,58	2,94	3,36	3,72
	MMSE-ERBRm	-0,88	1,50	4,56	8,21	2,45	2,85	3,31	3,76
	SE-MCRA	-0,80	0,55	2,22	3,53	2,21	2,55	2,94	3,25
	SE-ERBRm	1,81	2,15	2,13	2,91	1,98	2,52	2,94	3,26

continua na próxima página

continuação da página anterior

		segSNR				C_{ovl}			
Ruído	Método	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
	comp-wav	-3,98	-1,83	0,15	1,40	1,75	2,01	2,41	2,77
	WIENER-MCRA	-1,28	**1,00**	3,70	6,17	1,71	2,28	2,81	3,25
	WIENER-ERBRm	-1,32	0,98	**3,93**	**6,86**	1,78	2,35	2,87	**3,33**
	MMSE-MCRA	-2,48	0,09	3,10	5,88	1,86	**2,40**	**2,89**	3,30
S.Exibição	MMSE-ERBRm	-2,56	-0,17	2,89	5,98	**1,89**	2,39	**2,89**	3,32
	SE-MCRA	-2.39	-0,60	1,24	2,70	1,36	1,96	2,38	2,75
	SE-ERBRm	**-1,04**	-0,62	0,68	2,02	0,64	1,59	2,48	2,90
	comp-wav	-4,18	-0,78	2,41	4,71	1,54	1,99	2,47	3,04
	WIENER-MCRA	**0,26**	**1,86**	4,67	6,80	**2,06**	**2,31**	**2,86**	3,12
	WIENER-ERBRm	-0,94	1,10	**4,73**	**7,16**	1,69	2,08	2,85	**3,19**
	MMSE-MCRA	-2,02	-0,08	2,97	5,73	1,69	2,01	2,60	3,00
Tráfego	MMSE-ERBRm	-2,63	-0,54	2,81	5,81	1,67	1,99	2,60	3,06
	SE-MCRA	-0,82	0,26	1,89	3,03	1,63	1,91	2,39	2,71
	SE-ERBRm	-0,55	-0,03	1,48	2,35	1,21	1,76	2,39	2,83
	comp-wav	-2,55	0,34	2,94	4,30	1,46	1,82	2,53	2,95
	WIENER-MCRA	-0,97	1,43	4,14	7,00	2,00	2,51	2,94	3,34
	WIENER-ERBRm	**-0,77**	**1,62**	**4,50**	**7,80**	**2,15**	**2,61**	**3,06**	**3,46**
	MMSE-MCRA	-2,13	0,39	3,19	6,36	2,02	2,49	2,92	3,35
Trem	MMSE-ERBRm	-2,28	0,21	3,19	6,66	2,03	2,50	2,95	3,38
	SE-MCRA	-2,12	-0,52	1,29	2,97	1,47	2,04	2,44	2,83
	SE-ERBRm	-0,79	-0,51	0,81	2,27	0,99	2,02	2,57	2,97
	comp-wav	-3,66	-0,43	2,55	4,85	1,68	2,01	2,47	2,92

Fonte: Elaborado pelo próprio autor.

Analisando os resultados apresentados na Tabela 9, verifica-se que para o ruído Vozes o algoritmo WIENER-ERBRm foi ligeiramente superior em termos de segSNR. Em termos de C_{ovl}, o algoritmo que obteve o melhor desempenho foi MMSE-ERBRm. Apesar do algoritmo WIENER-ERBRm fornecer uma efetiva atenuação do ruído, os sinais processados foram afetados pela presença indesejável do ruído musical. Isto se deve a alta complexidade envolvida no processo de supressão do ruído Vozes. Neste sentido, o algoritmo MMSE-ERBRm foi mais eficiente: apesar dos sinais processados conterem um pouco mais de ruído residual, este ruído é menos desconfortável do que o ruído musical. Este fato refletiu sobre a avaliação C_{ovl}. Sendo assim, o algoritmo MMSE-ERBRm será selecionado para processar sinais contaminados com o ruído Vozes.

Com relação ao ruído Cafeteria, novamente o algoritmo WIENER foi superior em termos de segSNR, com ligeira vantagem para WIENER-ERBRm. Em termos de C_{ovl} o melhor algoritmo foi MMSE-ERBRm. Considerando que o ruído Cafeteria é similar ao ruído Vozes, com exceção de eventos aleatórios produzidos pelo som dos talheres e pratos, o algoritmo MMSE-ERBRm será selecionado para processar o ruído em questão.

Em se tratando do ruído Carro, fica evidente o desempenho superior do algoritmo

WIENER; WIENER alcançou melhor segSNR e qualidade geral para todos os níveis de contaminação. Considerando a ligeira vantagem para condições de baixas SNR, a configuração WIENER-MCRA será selecionada para compor o CMMV.

Para o caso do ruído Salão de Exibição, WIENER-ERBRm alcançou um melhor equilíbrio entre segSNR e C_{ovl}, sendo então selecionado para processar sinais contaminados com este tipo de ruído. Note que existe uma diferença significativa a favor do algoritmo WIENER-ERBRm termos de segSNR, o que não ocorre em termos de qualidade geral.

Como já destacado anteriormente, para o ruído Tráfego tanto a ERBR quanto o método MMSE não forneceram bons resultados. Este fato se confirma, novamente, na Tabela 9. Note que para o ruído de tráfego o algoritmo WIENER-MCRA se destacou tanto em termos de segSNR quanto em termos de C_{ovl}, sendo então selecionado para processar este tipo de ruído.

Finalmente, para o ruído Trem, a combinação WIENER-ERBRm se destacou em relação aos demais algoritmos, fornecendo melhor segSNR e C_{ovl} para todos os níveis de contaminação. Este fato reflete o bom desempenho do algoritmo WIENER em situações onde a potência do ruído é elevada, juntamente com o bom desempenho da ERBR sobre o mesmo tipo de ruído.

Na Tabela 10 estão, em síntese, os algoritmos selecionados para compor o CMMV e os respectivos tipos de ruído a ser processados.

Tabela 10 – Lista de métodos de melhoramento de voz criada e os respectivos tipos de ruído a ser processado.

Método	Ruído de fundo
MMSE-ERBRm	Vozes
MMSE-ERBRm	Cafeteria
WIENER-MCRA	Carro
WIENER-ERBRm	Salão de exibição
WIENER-MCRA	Tráfego
WIENER-ERBRm	Trem

Fonte: Elaborado pelo próprio autor.

5.2.2 Fase de operação

Para a execução da fase de operação, utiliza-se o CMMV construído e o classificador de ruído. Os métodos considerados para comparação utilizarão uma única configuração padrão para todos os tipos de ruído: $\alpha = 0,92$, MCRA para a estimação do ruído e

$\tau_k = [\tau_1, \tau_2, \tau_3, \tau_4, \tau_5] = [160, 45, 20, 18, 20]$ para o método comp-wav. As simulações serão realizadas sobre o conjunto de teste.

A fim de realizar simulações realísticas, todos os sinais contaminados do conjunto de teste serão apresentados para os 4 melhores métodos de MV destacados nas seções anteriores (SE, WIENER, MMSE e comp-wav), mais o CMMV proposto, sem o conhecimento prévio do tipo do ruído de fundo. Para cada tipo de ruído escolhido aleatoriamente, um grupo de 15 sinais é apresentado para o processamento. Repete-se este procedimento até que se tenha realizado o processamento dos seis tipos de ruído e para os 4 níveis de contaminação. A ordem com que as 15 sentenças são apresentadas será mantida, pois precisa-se do sinal de referência para calcular as medidas objetivas de qualidade.

Na Tabela 11 estão os resultados das avaliações objetivas realizadas sobre os sinais processados. A fim de obter uma análise mais rigorosa, além das medidas que já vinham sendo utilizadas nesta pesquisa, acrescenta-se nesta avaliação a medida $\mathrm{C_{sig}}$. Descrita na equação (75), a medida $\mathrm{C_{sig}}$ foi proposta por Hu e Loizou (2008a), juntamente com as medidas $\mathrm{C_{ovl}}$ e $\mathrm{C_{bak}}$, e tem por finalidade avaliar o nível de distorções de fala introduzidas durante o processamento. Quanto maior a nota $\mathrm{C_{sig}}$, menos distorções foram inseridas na fala e, consequentemente, melhor é o processamento.

$$C_{\mathrm{sig}} = 3,093 - 1,029\mathrm{LLR} + 0,603\mathrm{PESQ} - 0,009\mathrm{WSS}. \tag{75}$$

Os resultados apresentados na Tabela 11 representam a média das respectivas medidas de qualidade, quando avaliadas sobre todos os sinais processados, sem qualquer distinção sobre o tipo do ruído de fundo. Em outras palavras, cada valor apresentado na Tabela 11 é resultado da média aritmética das notas recebidas por 360 sinais processados, e que englobam todas as seis condições de ruído simuladas.

Verifica-se, a partir da Tabela 11, que de um modo geral a metodologia proposta se mostrou robusta e apresentou melhor desempenho frente a condições adversas de ruído. Observando as medidas $\mathrm{C_{ovl}}$ e $\mathrm{C_{bak}}$, o CMMV alcançou o melhor desempenho para todos os níveis de contaminação, com exceção para $\mathrm{C_{ovl}}$ em 0 dB, onde a diferença para o método MMSE é muito pequena. Analisando a medida $\mathrm{C_{sig}}$, os métodos que obtiveram o melhor desempenho foram MMSE e CMMV. No caso do método MMSE, a atenuação mais suave do ruído teve grande contribuição para com este bom desempenho. Nesse sentido, analisando a avaliação segSNR para ambos os métodos, verifica-se que o método proposto alcançou uma atenuação do ruído mais efetiva do que MMSE, além de manter as distorções de fala em um nível muito semelhante, sendo inclusive superior para as condições de 5 dB e 10 dB.

Tabela 11 – Avaliações objetivas dos sinais processados sem o conhecimento *a priori* do tipo do ruído de fundo.

Medida	SNR	SE	WIENER	MMSE	comp-wav	CMMV
C_{ovl}	0 dB	1,75	2,05	**2,20**	1,88	2,19
	5 dB	2,19	2,53	2,62	2,19	**2,66**
	10 dB	2,63	3,03	3,05	2,57	**3,12**
	15 dB	3,01	3,44	3,48	2,98	**3,51**
C_{bak}	0 dB	1,68	1,97	1,99	1,82	**2,03**
	5 dB	2,03	2,38	2,37	2,19	**2,45**
	10 dB	2,39	2,82	2,80	2,55	**2,88**
	15 dB	2,72	3,24	3,24	2,86	**3,30**
C_{sig}	0 dB	2,04	2,39	**2,61**	2,11	2,58
	5 dB	2,56	2,94	3,11	2,43	**3,13**
	10 dB	3,05	3,49	3,60	2,83	**3,62**
	15 dB	3,46	3,93	**4,06**	3,28	4,03
segSNR	0 dB	-1,45	**-0,29**	-1,69	-3,00	-0,59
	5 dB	0,22	**2,08**	0,97	0,05	1,94
	10 dB	1,96	**4,76**	4,05	2,64	4,73
	15 dB	3,49	7,60	7,29	4,22	**7,81**
PESQ	0 dB	1,90	2,06	2,04	1,93	**2,07**
	5 dB	2,19	2,38	2,33	2,19	**2,41**
	10 dB	2,50	2,75	2,64	2,50	**2,76**
	15 dB	2,78	3,07	2,98	2,85	**3,08**

Fonte: Elaborado pelo próprio autor.

Com relação a medida segSNR, fica claro que o filtro Wiener fornece uma atenuação mais expressiva do ruído, seguido pelo método CMMV. É importante ressaltar que a segSNR não deve ser olhada como único parâmetro para avaliação de algoritmos de MV, uma vez que esta medida, sozinha, não possui muita correlação com avaliações subjetivas de qualidade.

Por fim, em termos da avaliação PESQ, o método com melhor desempenho foi o CMMV, seguido por WIENER e MMSE.

A análise apresentada nesta seção confirma que o processamento de sinais de voz baseado na classificação do ruído de fundo e em um CMMV fornece resultados aprimorados, pois é possível calibrar os algoritmos para condições específicas de ruído, ou até mesmo escolher o melhor algoritmo para cada condição. No entanto, é importante salientar que o desempenho aprimorado depende fundamentalmente do bom desempenho do

classificador de ruído.

Verificou-se, durante as simulações, que frente a algum erro de classificação, apenas uma maior quantidade de ruído residual é gerada no sinal processado e nenhum tipo de distorção mais grave é inserida. Isto se deve ao fato de que a maior parte dos ajustes dos algoritmos se concentram na estimação do perfil do ruído, quando a estimação é pobre, o algoritmo não consegue removê-lo adequadamente. Para verificar este fato, no Apêndice D apresentam-se os mesmas avaliações apresentadas na Tabela 11, porém, individualizadas para cada tipo de ruído. Dessa forma, é possível verificar que não houve discrepâncias nas avaliações objetivas em relação ao tipo do ruído de fundo. Além disso, baseando-se nos resultados apresentados no Apêndice D, verifica-se que o CMMV proposto se mostrou eficiente para a remoção de todos os seis tipos de ruído.

6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O principal objetivo desta pesquisa foi investigar o desempenho dos principais métodos de melhoramento de voz quando submetidos a ambientes ruidosos reais, bem como explorar suas principais características e propor novas metodologias. Os estudos realizados e as metodologias sugeridas neste trabalho contribuem com aspectos teóricos e práticos para a área clássica de melhoramento de sinais de voz, sendo eles: (a) avaliação de diferentes métodos de melhoramento de voz, incluindo aqueles que trabalham no domínio da frequência, bem como os que utilizam o domínio wavelet, sob condições de ruído real; (b) desenvolvimento de um método de melhoramento de voz baseado em wavelets complexas; (c) desenvolvimento de um classificador de ruído real inspirado em um sistema imunológico artificial; (d) proposta de um algoritmo de estimação de ruído para MV baseado em regressão e classificação de ruído; (e) proposta de uma nova metodologia que se baseia na construção de um conjunto de métodos de MV, e que é capaz de explorar o desempenho máximo de algoritmos em condições específicas de ruído.

Durante a avaliação de diferentes métodos de MV, um dos pontos investigados foi o desempenho da limiarização wavelet sob condições de ruído real. Verificou-se que a limiarização wavelet tradicional, não é tão eficiente na remoção de ruídos reais. Na verdade, ela foi desenvolvida com o objetivo de suprimir o RGB. Apesar das melhorias propostas para a limiarização wavelet, como a limiarização adaptativa, o desempenho ainda não é satisfatório em termos de qualidade do processamento. Dessa forma, verificou-se um desempenho fraco tem termos da avaliação PESQ. Na busca por melhorar a qualidade do processamento, e baseado em um método wavelet não limiar, foi desenvolvido um esquema não limiar baseado na DT$-\mathbb{C}$WT. As simulações mostraram que este esquema alcançou melhor qualidade no processamento, quando comparado a limiarização, aliada a uma satisfatória atenuação do ruído.

De um modo geral, a partir das simulações realizadas, constatou-se que métodos baseados em modelos estatísticos que atuam no domínio da frequência tiveram melhor desempenho para a maioria das condições de ruído simuladas. O único método baseado em wavelets que trabalhou igualmente bem aos métodos baseados na DFT, para algumas condições de ruído, foi o proposto comp-wav. Os melhores resultados foram obtidos pelos métodos WIENER, MMSE, SE, comp-wav. No entanto, apesar desses quatro métodos se sobressaírem, verificou-se por meio das simulações que existem variações no desempenho dos mesmos, a depender do tipo de ruído contido no sinal. Como destacado para o ruído de tráfego, essas variações podem ser expressivas, indicando que um processamento baseado no tipo do ruído de fundo é adequado. Essa premissa se confirmou durante análise de literatura realizada na seção 4.1, onde apresentou-se alguns trabalhos recentes em MV, que são baseados na classificação do ruído. Apesar de ainda existirem poucos trabalhos

neste sentido, pois trata-se de uma abordagem recente, os resultados apresentados pela literatura mostraram que ajustar o algoritmo de MV para cada tipo de ruído melhora substancialmente os resultados.

O desenvolvimento de métodos de MV que atuam de diferentes maneiras para diferentes tipos de ruído depende, fundamentalmente, de um bom classificador de ruído. Nesse sentido, foi apresentado um método baseado no ASN para classificação de ruídos reais em sinais de voz. Além do bom desempenho na tarefa de classificação, o método desenvolvido possui um estágio de identificação do ruído: caso o sinal de voz esteja limpo, ou o nível de ruído seja tão baixo que nenhum processamento (por ex. melhoramento) é necessário, o sistema não aciona o módulo de classificação. Assim, se nenhum processamento para classificação é necessário, passa-se para outras fases do processamento da fala, reduzindo também o custo computacional. Este dispositivo permite o fácil acoplamento do classificador a outros sistemas de processamento de voz. Comparações foram conduzidas e o classificador proposto apresentou melhores resultados do que alguns dos classificadores clássicos que são amplamente utilizados em problemas de reconhecimento de padrões, alcançando um taxa de acerto médio de 96,29%. A presente pesquisa também reportou sobre a extração de características baseada na energia dos coeficientes wavelet complexos através das escalas wavelet. Cinco escalas foram utilizadas e um vetor de características com apenas cinco elementos foi proposto.

Tendo como base as análises realizadas e o classificador de ruído desenvolvido, foram exploradas as possibilidades e benefícios que o processamento baseado na classificação do ruído pode oferecer. Nesse sentido, apresentou-se duas metodologias para a área de MV que constituem contribuição científica, fornecendo subsídios teóricos e práticos que podem nortear o desenvolvimento e proposição de futuras pesquisas. Num primeiro momento, o objetivo foi o de investigar a seguinte problemática levantada: é possível realizar uma razoável estimação do ruído a partir do sinal de voz ruidoso por meio de regressão? O método de regressão PLS foi utilizado para estimar o perfil do ruído, que posteriormente foi utilizado para a implementação de um filtro de Wiener baseado na estimação da SNR_{prio}. A princípio, os resultados não foram satisfatórios, e o algoritmo proposto não se equiparou ao desempenho do já bem consolidado MCRA. Optou-se, então, por incorporar a classificação de ruído e um modelo de regressão para cada tipo de ruído foi treinado. Os resultados foram novamente confrontados com o algoritmo MCRA e, desta vez, a ERBR apresentou um desempenho semelhante ao método MCRA no quesito atenuação do ruído. Além disso, avaliações objetivas indicaram uma melhor qualidade e inteligibilidade dos sinais processados para a maioria das condições de ruído. Uma das principais vantagens de um método de ERBR consiste na substituição de complicados algoritmos de estimação de ruído, por uma matriz de coeficientes de regressão, diminuindo assim o custo computacional.

Após os resultados promissores de uma estimação de ruído, que também é realizada de maneiras diferentes de acordo com o tipo do ruído de fundo, uma nova metodologia para a área de MV foi proposta. Denominada Conjunto de Métodos de Melhoramento de Voz, esta metodologia incorpora tudo que foi discutido e desenvolvido neste trabalho. A ideia principal consiste em utilizar o melhor algoritmo, ou melhor configuração de um mesmo algoritmo, para cada tipo de ruído presente nos sinais contaminados. Os resultados das simulações mostraram que, de um modo geral, o CMMV apresentou melhor desempenho frente a condições adversas de ruído, quando comparado aos métodos clássicos.

É importante destacar que durante a avaliação do CMMV proposto, não havia o conhecimento *a priori* do tipo de ruído contido nos sinais contaminados. Assim, aproximou-se o máximo possível de situações reais, apresentando uma avaliação que não depende do tipo do ruído processado.

Como sugestão para trabalhos futuros, seria interessante aplicar o classificador de ruído desenvolvido nesta pesquisa para a tarefa de classificação de cenas acústicas. Esta área de pesquisa vem se consolidando devido ao amplo leque de aplicações, que vai desde aplicações em robótica, envolvendo a interação humano-robô, a aplicações FORENSE. Neste sentido, existem alguns bancos de dados desenvolvidos especificamente para este fim e que podem ser utilizados. Além disso, um estudo direcionado especificamente para a extração de características e seleção de atributos que envolva, além das características tradicionais, características extraídas por meio de wavelets complexas seria relevante e pode melhorar a acurácia na classificação do ruído.

No caso específico da área de MV, os resultados da ERBR propiciam futuras investigações sobre modelos de regressão mais poderosos e com alto grau de não-linearidade, a fim de melhorar a estimação do ruído. Os benefícios de uma estimação de ruído com estas características são muito atraentes.

REFERÊNCIAS

ABBAS, A. K.; LICHTMAN, A. H. *Imunologia básica:* funções e distúrbios do sistema imunológico. Rio de Janeiro: Elsevier, 2007. 354 p.

ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. *Imunologia celular e molecular*. Rio de Janeiro: Elsevier, 2008. 564 p.

ABDI, H. Partial least square regression (pls regression). In: ENCYCLOPEDIA for research methods for the social sciences. Thousand Oaks: Sage: [s.n.], 2003. p. 792–795.

ABREU, C. C. E. *Uso de equações de diferenças na obtenção de filtros para redução de ruído em sinais de voz no domínio wavelet.* 2013. 96 f. Dissertação (Mestrado em Engenharia Elétrica) — Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2013.

ABREU, C. C. E.; CHAVARETTE; DUARTE, M. A. Q.; VILLARREAL, F. Analysis of the structural integrity of a building by complex wavelets. *International Journal of Applied Mathematics*, Sofia, v. 28, n. 2, p. 159–164, 2015.

ABREU, C. C. E.; CHAVARETTE; VILLARREAL, F.; DUARTE, M. A. Q.; LIMA, F. P. A. Dual-tree complex wavelet transform applied to fault monitoring and identification in aeronautical structures. *International Journal of Pure and Applied Mathematics*, Sofia, v. 97, n. 1, p. 89–97, 2014.

ABREU, C. C. E.; DUARTE, M. A. Q.; VILLARREAL, F. Uso de equações de diferenças para obtenção de filtros na redução de ruído em sinais de voz no domínio wavelet. In: CONGRESSO DE MATEMÁTICA APLICADA E COMPUTACIONAL DO NORDESTE (CMAC - NORDESTE). Natal. *Anais...*: [s.n.], 2012. p. 376–379.

ABREU, C. C. E.; DUARTE, M. A. Q.; VILLARREAL, F. Analysis of the evolution speech enhancement methods in wavelet domain. In: CONGRESSO DE MATEMÁTICA APLICADA E COMPUTACIONAL DO SUDESTE (CMAC-SUDESTE). Bauru. *Anais...*: [s.n.], 2013. p. 555–560.

ABREU, C. C. E.; DUARTE, M. A. Q.; VILLARREAL, F. Dual-tree complex wavelet transform in the problem of speech enhancement. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, São Carlos, v. 3, n. 1, p. 555–560, 2015.

ABREU, C. C. E. de; DUARTE, M. A. Q.; VILLARREAL, F. An immunological approach based on the negative selection algorithm for real noise classification in speech signals. *AEÜ - International Journal of Electronics and Communications*, Muenchen, v. 72, p. 125 – 133, 2017.

ABUTALEBI, H. R.; RASHIDINEJAD, M. Speech enhancement based on β-order mmse estimation of short time spectral amplitude and laplacian speech modeling. *Speech Communication*, Amsterdam, v. 67, p. 92 – 101, 2015.

ALMAJAI, I.; MILNER, B. Visually derived wiener filters for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 19, n. 6, p. 1642–1651, 2011.

AYAT, S.; MANZURI, M.; DIANAT, R. Wavelet based speech enhancement using a new thresholding algorithm. In: INTERNATIONAL SYMPOSIUM ON INTELLIGENT MULTIMEDIA, VIDEO AND SPEECH PROCESSING, 2004. Hong Kong. *Proceedings...* Hong Kong: IEEE, 2004. p. 238–241.

BAHOURA, M.; ROUAT, J. Wavelet speech enhancement based on time-scale adaptation. *Speech Communication*, Amsterdam, v. 48, n. 12, p. 1620–1637, 2006.

BEROUTI, M.; SCHWARTZ, R.; MAKHOUL, J. Enhancement of speech corrupted by acoustic noise. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING - ICASSP'79, 1979. Washington. *Proceedings...* Washington: IEEE, 1979. p. 208–211.

BHATTACHARYA, J. S. . S. *A text book of immunology*. Kolkata: Academic Publishers, 2006. 487 p.

BLATTER, C. *Wavelets:* a primer. A.K.: Peters, 1998. 212 p.

BOGAERT, T. V. D.; DOCLO, S.; WOUTERS, J.; MOONEN, M. Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids. *The Journal of the Acoustical Society of America*, Melville, v. 125, n. 1, p. 360–371, 2009.

BOLL, S. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Piscataway, v. 27, n. 2, p. 113–120, 1979.

BREAZEAL, C. Recognition of affective communicative intent in robot-directed speech. *Autonomous Robots*, New York, v. 12, n. 1, p. 83–104, 2002.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. *Classification and regression trees*. [S.l.]: Wadsworth and Brooks, 1984. 368 p.

BURGES, C. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, New York, v. 2, n. 2, p. 121–167, 1998.

CHEHREHSA, S.; MOIR, T. J. Speech enhancement using maximum a-posteriori and gaussian mixture models for speech and noise periodogram estimation. *Computer Speech & Language*, Londres, v. 36, p. 58 – 71, 2016.

COHEN, I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 11, n. 5, p. 466–475, 2003.

COHEN, I.; BERDUGO, B. Noise estimation by minima controlled recursive averaging for robust speech enhancement. *IEEE Signal Processing Letters*, Piscataway, v. 9, n. 1, p. 12–15, 2002.

CORNELIS, B.; MOONEN, M.; WOUTERS, J. Performance analysis of multichannel wiener filter-based noise reduction in hearing aids under second order statistics estimation errors. *IEEE Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 19, n. 5, p. 1368–1381, 2011.

DAS, A.; BHUIYAN, M.; ALAM, S. Statistical parameters in the dual tree complex wavelet transform domain for the detection of epilepsy and seizure, 2014. In: 2013 INTERNATIONAL CONFERENCE ON ELECTRICAL INFORMATION AND COMMUNICATION TECHNOLOGY. Khulna. *Proceedings...* Khulna: [s.n.], 2014. p. 1–6.

DASGUPTA, D. *Artificial immune systems and their applications*. New York: Springer-Verlag, 1999. 306 p.

DASGUPTA, D.; NINO, F. *Immunological computation:* theory and applications. Boca Raton: CRC, 2008. 296 p.

DASGUPTA, D.; YU, S.; NINO, F. Recent advances in artificial immune systems: Models and applications. *Applied Soft Computing*, Amsterdam, v. 11, n. 2, p. 1574–1587, 2011.

DAUBECHIES, I. *Ten lectures on wavelets.* Philadelphia: SIAM Books, 1992. 357 p.

DE CASTRO, L. *Engenharia imunológica:* desenvolvimento e aplicação de ferramentas computacionais inspiradas em sistemas imunológicos artificiais. 2001. 286 p. Tese (Doutorado em Engenharia Elétrica) — Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2001.

DE CASTRO, L.; TIMMIS, J. Artificial immune systems: A novel paradigm to pattern recognition. In: ARTIFICIAL NEURAL NETWORKS IN PATTERN RECOGNITION, 2002. Scotland. *Proceedings...* Scotland: University of Paisley, 2002. p. 67–84.

DE CASTRO, L. N.; VON ZUBEN, F. J. The clonal selection algorithm with engineering applications. In: GENETIC AND EVOLUTIONARY COMPUTATION CONFERENCE - GECCO, 2000. Nevada. *Proceedings...* Nevada: [s.n.], 2000. p. 36–39.

DE CASTRO, L. N.; VON ZUBEN, F. J. An evolutionary immune network for data clustering. In: BRAZILIAN SYMPOSIUM ON NEURAL NETWORKS, 6, 2000. Rio de Janeiro. *Proceedings...* Rio de Janeiro: IEEE: [s.n.], 2000. v. 1, p. 84–89.

DELLER, J. L.; PROAKIS, J. G.; HANSEN, J. H. L. *Discrete-time processing of speech signals*. New York: Macmillam, 1993. 908 p.

DING, G. H.; HUANG, T.; XU, B. Suppression of additive noise using a power spectral density mmse estimator. *IEEE Signal processing letters*, Piscataway, v. 11, n. 6, p. 585–588, 2004.

DONOHO, D. L. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, Piscataway, v. 41, n. 3, p. 613–627, 1995.

DONOHO, D. L.; JOHNSTONE, I. M. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, Oxford, v. 81, n. 3, p. 425–455, 1994.

DOOSTDAR, M.; SCHIFFER, S.; LAKEMEYER, G.; IOCCHI, L.; MATSUBARA, H.; WEITZENFELD, A.; ZHOU, C. A robust speech recognition system for service-robotics applications. In: *RoboCup 2008:* XII robot soccer world cup. Suzhou: Springer Berlin, 2009. p. 1–12.

DUARTE, M. A. Q. *Redução de ruído em sinais de voz no domínio wavelet.* 2005. 105 p. Tese (Doutorado em Engenharia Elétrica) — Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2005.

DUARTE, M. A. Q.; VIEIRA FILHO, J.; ALVARADO, F. V. Um simples e eficiente detector de atividade de voz utilizando a transformada wavelet. In: BRAZILIAN CONGRESS OF COMPUTATIONAL AND APPLIED MATHEMATICS - CNMAC, 2009. Cuiabá. *Proceedings...* Cuiabá: [s.n.], 2009. p. 1022–1028.

EL-FATTAH, M. A.; DESSOUKY, M.; ABBAS, A.; DIAB, S.; EL-RABAIE, E.-S.; AL-NUAIMY, W.; ALSHEBEILI, S.; EL-SAMIE, F. A. Speech enhancement with an adaptive wiener filter. *International Journal of Speech Technology*, New York, v. 17, n. 1, p. 53–64, 2014.

EPHRAIM, Y. A bayesian estimation approach for speech enhancement using hidden markov models. *IEEE Transactions on Signal Processing*, Piscataway, v. 40, n. 4, p. 725–735, 1992.

EPHRAIM, Y.; MALAH, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Piscataway, v. 32, n. 6, p. 1109–1121, 1984.

EPHRAIM, Y.; MALAH, D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Piscataway, v. 33, n. 2, p. 443–445, 1985.

EPHRAIM, Y.; TREES, H. L. V. A signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 3, n. 4, p. 251–266, 1995.

FIELD, A. *Discovering statistics using SPSS*. England: Sage, 2005. 915 p.

FIGUEIREDO, D. G. *Análise de Fourier e equações diferenciais parciais.* Rio de Janeiro: IMPA, 1997. 274 p.

FORREST, S.; JAVORNIK, B.; SMITH, R. E.; PERELSON, A. S. Using genetic algorithms to explore pattern recognition in the immune system. *Evolutionary computation*, Cambridge, v. 1, n. 3, p. 191–211, 1993.

FORREST, S.; PERELSON, A.; ALLEN, L.; CHERUKURI, R. Self-nonself discrimination in a computer. In: IEEE COMPUTER SOCIETY SYMPOSIUM ON RESEARCH IN SECURITY AND PRIVACY, 1994. Oakland. *Proceedings...* Oakland: [s.n.], 1994. p. 202–212.

FUKUDA, T.; MORI, K.; TSUKIAMA. Parallel search for multi-modal function optimization with diversity and learning of immune algorithm. In: DASGUPTA, D. (Ed.). *Artificial immune systems and their applications*. [S.l.]: Springer, 1999. p. 210–220.

GAUVAIN, J.; CHIN-HUI, L. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 2, n. 2, p. 291–298, 1994.

GHANBARI, Y.; KARAMI-MOLLAEI, M. R. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Communication*, Amsterdam, v. 48, n. 1, p. 927–940, 2006.

GOMES, J.; VELHO, L.; GOLDSTEIN, S. *Wavelets:* teoria, software e aplicações. Rio de Janeiro: IMPA, 1997. 216 p.

HAYKIN, S. *Neural networks and learning machines*. Upper Saddle River: Prentice-Hall, 2008. 936 p.

HIGHTOWER, R.; FORREST, S.; PERELSON, A. S. The baldwin effect in the immune system: Learning by somatic hypermutation. In: ADAPTIVE INDIVIDUALS IN EVOLVING POPULATIONS, 2000. Santa Fe. *Proceedings...* Santa Fe: Addison-Wesley Longman, 1996. p. 159–167.

HIRSCH, H. G.; PEARCE, D. The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In: ASR2000-AUTOMATIC SPEECH RECOGNITION: CHALLENGES FOR THE NEW MILLENIUM, 2000. Paris. *Proceedings...* Paris: [s.n.], 2000.

HOSEINKHANI, F.; PARCHAM, E.; POURNAZARY, M.; BORZUE, N. Speech recognition by classifying speech signals based on the fire fly and fuzzy. In: INTERNATIONAL CONFERENCE ON ADVANCED COMPUTER SCIENCE APPLICATIONS AND TECHNOLOGIES- ACSAT, 2012. Kuala Lumpur. *Proceedings...* Kuala Lumpur: [s.n.], 2012. p. 187–191.

HU, Y.; LOIZOU, P. C. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, Amsterdam, v. 49, n. 7–8, p. 588 – 601, 2007.

HU, Y.; LOIZOU, P. C. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 16, n. 1, p. 229–238, 2008.

HU, Y.; LOIZOU, P. C. A new sound coding strategy for suppressing noise in cochlear implants. *The Journal of the Acoustical Society of America*, Melville, v. 124, n. 1, p. 498–509, 2008.

HUNT, J.; TIMMIS, J.; COOKE, E.; NEAL, M.; KING, C. Jisys: the envelopment of an artificial immune system for real world applications. In: DASGUPTA, D. (Ed.). *Artificial immune systems and their applications*. [S.l.]: Springer, 1999. p. 157–186.

ISLAM, M. T.; SHAHNAZ, C.; WEI-PING, Z.; AHMAD, M. O. Speech enhancement based on student t modeling of teager energy operated perceptual wavelet packet coefficients and a custom thresholding function. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 23, n. 11, p. 1800–1811, 2015.

JABLOUN, F.; CHAMPAGNE, B. Incorporating the human hearing properties in the signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 11, n. 6, p. 700–708, 2003.

JI, Z.; DASGUPTA, D. Revisiting negative selection algorithms. *Evolutionary Computation*, Cambridge, v. 15, n. 2, p. 223–251, 2007.

JOLLIFFE, I. *Principal component analysis*. [S.l.]: Wiley Online Library, 2002. 488 p.

JONG, S. D. Simpls: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, Amsterdam, v. 18, n. 3, p. 251–263, 1993.

KAMATH, S.; LOIZOU, P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING- ICASSP, 2002. Orlando. *Proceedings...* Orlando: [s.n.], 2002. p. IV–4164–IV–4164.

KATES, J. M. Classification of background noises for hearing-aid applications. *The Journal of the Acoustical Society of America*, Melville, v. 97, n. 1, p. 461–470, 1995.

KHARE, A.; KHARE, M.; SRIVASTAVA, R. Dual tree complex wavelet transform based multiclass object classification. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS - ICMLA, 2003. Miami. *Proceedings...* Miami: [s.n.], 2013. v. 2, n. 12, p. 501–506.

KIM, E. H.; HYUN, K. H.; KIM, S. H.; KWAK, Y. K. Improved emotion recognition with a novel speaker-independent feature. *IEEE/ASME Transactions on Mechatronics*, Piscataway, v. 14, n. 3, p. 317–325, 2009.

KINGSBURY, N. G. The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters. In: IEEE DSP WORKSHOP, 1998. Utah. *Proceedings...* Utah: [s.n.], 1998. p. 9–12.

KINGSBURY, N. G. Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, Maryland Heights, v. 10, n. 3, p. 234–253, 2001.

KINGSBURY, N. G. Design of q-shift complex wavelets for image processing using frequency domain energy minimization. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2003. Barcelona. *Proceedings...* Barcelona: [s.n.], 2003. p. 1013–1016.

KLATT, D. Prediction of perceived phonetic distance from critical-band spectra: A first step. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING - ICASSP'82, 1982. Paris. *Proceedings...* Paris: [s.n.], 1982. v. 7, p. 1278–1281.

KNIGHT, T.; TIMMIS, J. Aine: an immunological approach to data mining. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING - ICDM, 2001). San Jose. *Proceedings...* San Jose: [s.n.], 2001. p. 297–304.

KODRASI, I.; MARQUARDT, D.; DOCLO, S. Curvature-based optimization of the trade-off parameter in the speech distortion weighted multichannel wiener filter. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING - ICASSP, 2015. South Brisbane. *Proceedings...* South Brisbane: [s.n.], 2015. p. 315–319.

LALLOUANI, A.; GABREA, M.; GARGOUR, C. S. Wavelet based speech enhancement using two different threshold-based denoising algorithms. In: CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING, 2004. Ontario. *Proceedings...* Ontario: [s.n.], 2004. p. 315–318.

LI, D.; LIU, S.; ZHANG, H. Negative selection algorithm with constant detectors for anomaly detection. *Applied Soft Computing*, Amsterdam, v. 36, p. 618–632, 2015.

LIMA, F. P. A.; LOTUFO, A. D. P.; MINUSSI, C. R. Disturbance detection for optimal database storage in electrical distribution systems using artificial immune systems with negative selection. *Electric Power Systems Research*, Amsterdam, v. 109, p. 54 – 62, 2014.

LINA, J.; MAYRAND, M. Parametrizations for daubechies wavelets. *Phys. Rev. E*, College Park, v. 48, n. 6, p. R4160–R4163, 1993.

LOIZOU, P.; KIM, G. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 19, n. 1, p. 47–56, 2011.

LOIZOU, P. C. Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 13, n. 5, p. 857–869, 2005.

LOTTER, T.; VARY, P. Noise reduction by maximum a posteriori spectral amplitude estimation with supergaussian speech modeling. In: INTERNATIONAL WORKSHOP ON ACOUSTIC ECHO NOISE CONTROL - IWAENC'03, 2003. Kyoto. *Proceedings...* Kyoto: [s.n.], 2003. v. 3, p. 83–86.

LOTTER, T.; VARY, P. Speech enhancement by map spectral amplitude estimation using a super-gaussian speech model. *EURASIP Journal on Applied Signal Processing*, Heidelberg, v. 2005, n. 1, p. 1110–1126, 2005.

LU, Y.; LOIZOU, P. Estimators of the magnitude-squared spectrum and methods for incorporating snr uncertainty. *IEEE Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 19, n. 5, p. 1123–1137, 2011.

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.

MA, L.; SMITH, D.; MILNER, B. Environmental noise classification for context-aware applications. In: INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS APPLICATIONS - DEXA, 2003. Prague. *Proceedings...* Prague: [s.n.], 2003. p. 360–370.

MALAH, D.; COX, R.; ACCARDI, A. Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING - ICASSP, 1999. Phoenix. *Proceedings...* Phoenix: [s.n.], 1999. v. 2, p. 789–792.

MALLAT, S. *A wavelet tour of signal processing*. San Diego: Academic, 1998. 700 p.

MARTIN, R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on speech and audio processing*, Piscataway, v. 9, n. 5, p. 504–512, 2001.

MCAULAY, R.; MALPASS, M. Speech enhancement using a soft-decision noise suppression filter. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Piscataway, v. 28, n. 2, p. 137–145, Apr 1980.

MENG, X.; RUBIN, D. B. Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, Oxford, v. 80, n. 2, p. 267–278, 1993.

MESSAOUD, M. anouar B.; BOUZID, A.; ELLOUZE, N. Speech enhancement based on wavelet packet of an improved principal component analysis. *Computer Speech & Language*, London, v. 35, p. 58 – 72, 2016.

MITICHE, L.; ADAMOU-MITICHE, A.; NAIMI, H. Medical image denoising using dual tree complex thresholding wavelet transform. In: IEEE JORDAN CONFERENCE ON APPLIED ELECTRICAL ENGINEERING AND COMPUTING TECHNOLOGIES, 2013. Amman. *Proceedings...* Amman: [s.n.], 2013. p. 1–5.

MOHANAPRASAD, K.; ARULMOZHIVARMAN, P. Wavelet-based ica using maximum likelihood estimation and information-theoretic measure for acoustic echo cancellation during double talk situation. *Circuits, Systems, and Signal Processing*, Basel, v. 34, n. 12, p. 3915–3931, 2015.

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

PALIWAL, K.; SCHWERIN, B.; WÓJCICKI, K. Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. *Speech Communication*, Amsterdam, v. 54, n. 2, p. 282–305, 2012.

PALIWAL, K.; WÓJCICKI, K.; SHANNON, B. The importance of phase in speech enhancement. *Speech Communication*, Amsterdam, v. 53, n. 4, p. 465–494, 2011.

PARCHAMI, M.; ZHU, W.-P.; CHAMPAGNE, B.; PLOURDE, E. Bayesian stsa estimation using masking properties and generalized gamma prior for speech enhancement. *EURASIP Journal on Advances in Signal Processing*, Heidelberg, v. 2015, n. 1, p. 1–21, 2015.

PARRIS, S.; TORLAK, M.; KEHTARNAVAZ, N. Real-time implementation of cochlear implant speech processing pipeline on smartphones. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY - EMBC, 2014. Chicago. *Proceedings...* Chicago: [s.n.], 2014. p. 886–889.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Cambridge, v. 12, p. 2825–2830, 2011.

PUTEH, M.; HAMDAN, A. R.; OMAR, K.; BAKAR, A. A. Flexible immune network recognition system for mining heterogeneous data. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL IMMUNE SYSTEMS - ICARIS, 2008. Phuket. *Proceedings...* Phuked: [s.n.], 2008. p. 232–241.

QUACKENBUSH, S.; BARNWELL, T.; CLEMENTS. *Objective measures of speech quality*. [S.l.]: Englewood Cliffs: Prentice-Hall, 1988. 377 p.

RABINER, L.; JUANG, B. *Fundamentals of speech recognition*. USA: Prentice hall, 1993. 496 p.

RAKOTOMAMONJY, A.; GASSO, G. Histogram of gradients of time-frequency representations for audio scene classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 23, n. 1, p. 142–153, 2015.

RAO, C. R.; MURTHY, M. R.; RAO, K. S. Speech enhancement using sub-band cross-correlation compensated wiener filter combined with harmonic regeneration. *AEÜ - International Journal of Electronics and Communications*, Muenchen, v. 66, n. 6, p. 459–464, 2012.

REC ITUT. *P. 862 Perceptual Evaluation of Speech Quality: an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. Geneva: International Telecommunication Union, p. 749–752., 2001.

REYNOLD, D. A.; ROSE, R. C. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 3, n. 1, p. 72–83, 1995.

SAKI, F.; KEHTARNAVAZ, N. Background noise classification using random forest tree classifier for cochlear implant applications. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING - ICASSP, 2014. Florence. *Proceedings...* Florence: [s.n.], 2014. p. 3591–3595.

SCALART, P.; VIEIRA FILHO, J. Speech enhancement based on a priori signal to noise estimation. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING - ICASSP'96. Atlanta. *Proceedings...* Atlanta: [s.n.], 1996. v. 2, p. 629–632.

SELESNICK, I. W.; BARANIUK, R. G.; KINGSBURY, N. G. The dual-tree complex wavelet transform—a coherent framework for multiscale signal and image processing. *IEEE Signal Processing Magazine*, Piscataway, v. 22, n. 6, p. 123–151, 2005.

SHEIKHZADEH, H.; ABUTALEBI, H. R. An improved wavelet-based speech enhancement system. In: EUROSPEECH 2001. Aalborg. *Proceedings...* Aalborg: [s.n.], 2011. p. 1855–1858.

SOARES, W. C.; VILLARREAL, F.; DUARTE, M. A. Q.; FILHO, J. V. Wavelets in a problem of signal processing. *Novi Sad Journal of Mathematics*, Novi Sad, v. 41, n. 1, p. 11–20, 2011.

SOHN, J.; KIM, N.; SUNG, W. A statistical model-based voice activity detection. *IEEE signal processing letters*, Piscataway, v. 6, n. 1, p. 1–3, 1999.

SPERANDIO, D.; MENDES, J. T.; SILVA, L. H. M. *Cálculo Numérico:* Características Matemáticas e Computacionais dos Métodos Numéricos. São Paulo: Prentice Hall, 2003. 368 p.

STRANG, G.; NGUYEN, T. *Wavelets and filter banks*. Wellesley: Cambridge, 1996. 500 p.

SWAMI, P. D.; SHARMA, R.; JAIN, A.; SWAMI, D. K. Speech enhancement by noise driven adaptation of perceptual scales and thresholds of continuous wavelet transform coefficients. *Speech Communication*, Amsterdam, v. 70, p. 1 – 12, 2015.

TABIBIAN, S.; AKBARI, A.; NASERSHARIF, B. Speech enhancement using a wavelet thresholding method based on symmetric kullback–leibler divergence. *Signal Processing*, Amsterdam, v. 106, n. 0, p. 184–194, 2015.

TIMMIS, J. *Artificial Immune Systems:* A novel data analysis technique inspired by the immune network theory. 2001. Tese (PhD in Computer Science) — University of Wales, Aberystwyth, 2001.

TIMMIS, J.; KNIGHT, T.; DE CASTRO, L. N.; HART, E. An overview of artificial immune systems. In: *Computation in cells and tissues*. Heidelberg: Springer Berlin, 2004. p. 51–91.

UCHÔA, J. Q. *Algoritmos imunoinspirados aplicados em segurança computacional:*utilização de algoritmos inspirados no sistema imune para detecção de intrusos em redes de computadores. 2009. 245 f. 245 p. Tese (Doutorado em Bioinformática) — Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

UZINSKI, J. C.; PAIVA, H. M.; VILLARREAL, F.; DUARTE, M.; GALVÃO, R. Additional constraints to ensure three vanishing moments for orthonormal wavelet filter banks. In: CONGRESSO DE MATEMÁTICA APLICADA E COMPUTACIONAL - CENTRO-OESTE, 2003. Cuiabá. *Proceedings...* Cuiabá: [s.n.], 2013. p. 16–19.

VAPNIK, V. N. *The nature of statistical learning theory*. New York: Springer-Verlag, 1995.

VICENT, P.; LAROCHELLE, H.; BENGIO, Y.; MANZAGOL, P. A. Extracting and composing robust features with denoising autoencoders. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2008. [S.l.: s.n.], 2008. p. 1096–1103.

VIEIRA FILHO, J. *Redução de ruído em sinais de voz nos sistemas rádio móveis veiculares. 1996*. Tese (Doutorado em Engenharia Elétrica) — Faculdade de Engenharia Elétrica e de Computação Universidade Estadual de Campinas, Campinas, 1996.

WANG, D. On ideal binary mask as the computational goal of auditory scene analysis. In: DIVENYI, P. (Ed.). *Speech separation by humans and machines*. [S.l.: s.n.], 2011. p. 137–140.

WANG, D.; BROWN, G. J. *Computational auditory scene analysis:* Principles, algorithms, and applications. [S.l.]: Wiley-IEEE, 2006. 395 p.

WERBOS, P. J. *Beyond regression:* new tools for prediction and analysis in the behavioral sciences. 1974. Tese (Doutorado) — - Harvard University, 1974.

WOLD, H. Partial least squares. In: ENCYCLOPEDIA of statistical sciences. [S.l]. [S.l.]: Wiley Online Library, 1985.

WOLFE, P. J.; GODSILL, S. J. Efficient alternatives to the ephraim and malah suppression rule for audio signal enhancement. *EURASIP Journal on Advances in Signal Processing*, Heidelberg, v. 2003, n. 10, p. 1–9, 2003.

XIA, B.; BAO, C. Wiener filtering based speech enhancement with weighted denoising auto-encoder and noise classification. *Speech Communication*, Amsterdam, v. 60, p. 13–29, 2014.

XIAO, X.; LI, T.; ZHANG, R. An immune optimization based real-valued negative selection algorithm. *Applied Intelligence*, New York, v. 42, n. 2, p. 289–302, 2015.

XU, H.; TAN, Z. H.; DALSGAARD, P.; LINDBERG, B. Robust speech recognition based on noise and snr classification-a multiple-model framework. In: INTERSPEECH. [S.l.: s.n.], 2005. p. 977–980.

XU, Y.; DU, J.; DAI, L. R.; LEE, C. H. An experimental study on speech enhancement based on deep neural networks. *IEEE Signal Processing Letters*, Piscataway, v. 21, n. 1, p. 65–68, 2014.

XU, Y.; DU, J.; DAI, L. R.; LEE, C. H. A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 23, n. 1, p. 7–19, 2015.

YUAN, W.; XIA, B. A speech enhancement approach based on noise classification. *Applied Acoustics*, Oxford, v. 96, p. 11–19, 2015.

ZHANG, S.; TANG, T.; WU, C.; XI, N.; WANG, G. A novel image denoising method using independent component analysis and dual-tree complex wavelet transform. In: INTERNATIONAL CONFERENCE ON WIRELESS COMMUNICATIONS NETWORKING AND MOBILE COMPUTING, 2010. Chengdu City. *Proceedings...* Chengdu City: [s.n.], 2010. p. 1–4.

Apêndices

APÊNDICE A – O SISTEMA IMUNOLÓGICO BIOLÓGICO

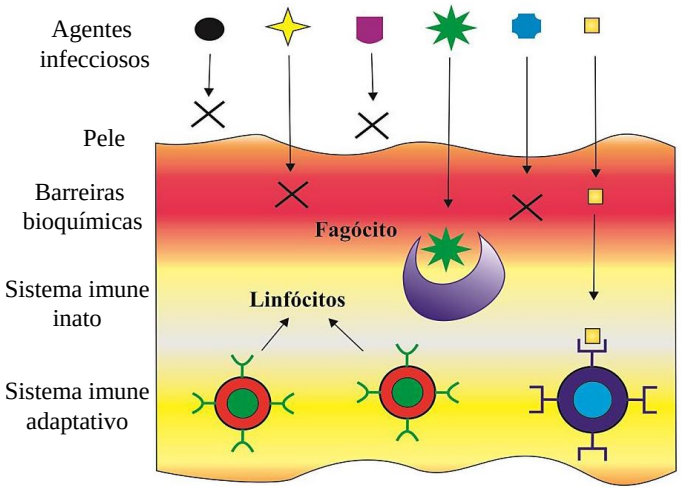
O sistema imunológico biológico pode ser interpretado como um mecanismo de defesa do organismo contra agentes infecciosos, os patógenos. Os patógenos possuem como componentes os antígenos, que podem ser quaisquer moléculas capazes de serem reconhecidas pelo sistema imunológico, podendo ser passivas ou agressivas (DE CASTRO, 2001; UCHÔA, 2009). A maneira mais adequada de referir-se a moléculas passivas ou agressivas seria a introdução do termo Próprio ou não Próprio.

O estudo do sistema imunológico biológico se concentra, quase que exclusivamente, sobre os vertebrados, e de uma forma mais particular, sobre mamíferos (DE CASTRO, 2001). Neste contexto, os conceitos expostos neste apêndice estão sob a perspectiva do sistema imunológico humano.

Constituído por células, tecidos e moléculas que agem de uma maneira coletiva e ordenada, o SIH produz uma resposta imunológica sempre que detectado algum tipo de agente infeccioso externo (tais como bactérias, vírus, fungos ou parasitas) (ABBAS; LICHTMAN; PILLAI, 2008). Para isso, é necessário o reconhecimento dos antígenos do próprio corpo e dos antígenos externos. Quando, por engano, uma resposta imunológica é desencadeada contra antígenos do próprio corpo, têm-se a caracterização das doenças autoimunes. Alguns exemplos de doenças autoimunes são: Diabetes Mellitus tipo 1; Tireoide Autoimune; Vitiligo; Retocolite Ulcerativa; Esclerose Múltipla; Anemia Hemolítica; Hepatite Autoimune, entre outras (ABBAS; LICHTMAN, 2007).

O sistema imunológico biológico é disposto em uma estrutura multicamadas, sendo que a primeira delas é a pele, que atua como uma espécie de escudo protetor contra quaisquer tipos de invasores, sendo estes maléficos ou não (DE CASTRO, 2001). A segunda camada é denominada camada bioquímica, onde a temperatura e o pH corporais tornam o ambiente desfavorável para a sobrevivência de organismos estranhos (ABBAS; LICHTMAN; PILLAI, 2008). Também fazem parte da camada bioquímica, a saliva, o suor e os ácidos estomacais. Estes últimos são capazes de eliminar grande parte dos microorganismos ingeridos com o alimento e a água (DE CASTRO, 2001). Na Figura 21 é ilustrada a atuação de todas as camadas do SIH.

Figura 21 – Estrutura multicamadas do sistema imunológico biológico



Fonte: Adaptado de de Castro (2001).

Caso as duas primeiras barreiras sejam vencidas, os agentes infecciosos conseguirão penetrar no corpo e então serão combatidos por outras duas camadas, são elas o sistema imune inato e o sistema imune adaptativo.

O sistema imunológico inato é responsável por fornecer uma resposta rápida e eficiente contra uma grande variedade de patógenos. Suas células estão disponíveis para uma intervenção imediata sem exigir uma prévia exposição a alguma variedade específica de patógenos. Uma característica deste sistema é que ele atua de maneira muito semelhante em todos os indivíduos (DE CASTRO, 2001). Penetrando nos tecidos ou na circulação, os patógenos são atacados pelos Fagócitos, cuja função principal é identificar, ingerir e destruir microorganismos. O sistema imune inato é caracterizado pela carga genética que o indivíduo herda ao nascer e não se altera ao longo do tempo e/ou exposições patogênicas (ABBAS; LICHTMAN, 2007).

Devido à incapacidade de aprendizagem, muitos microorganismos patogênicos aos seres humanos, ou seja, microorganismos capazes de causar doenças, evoluíram para resistir aos mecanismos da imunidade inata. Estes agentes infecciosos serão combatidos pelo sistema imunológico adaptativo (ou adquirido) (ABBAS; LICHTMAN, 2007).

O sistema imunológico adaptativo é responsável por uma resposta imune específica, ou seja, é responsável pela produção de anticorpos a um determinado agente infeccioso. Nesta etapa, os anticorpos são produzidos pelos linfócitos, que são responsáveis por reconhecer e eliminar o agressor. Este sistema proporciona uma resposta imune dura-

doura, sendo estimulada por meio da vacinação ou a exposição a uma doença (DE CASTRO, 2001). O sistema imune adaptativo tem como característica a memória imunológica e a produção de respostas especializadas. Neste sentido, a resposta imune adaptativa aperfeiçoa-se a cada encontro com um antígeno (DE CASTRO, 2001; ABBAS; LICHTMAN; PILLAI, 2008).

Os linfócitos são capazes de reconhecer um antígeno caso o mesmo entre novamente em contato com o organismo, tornando a resposta imune mais eficiente e rápida, caracterizando o conceito de memória imunológica e evitando o reestabelecimento da doença. Neste processo, destacam-se os linfócitos T (ou células T) e os linfócitos B (ou células B). As células B são as únicas capazes de produzir anticorpos. Elas atuam reconhecendo antígenos de microorganismos extracelulares, diferenciando-se em células secretoras de anticorpos. Já as células T, atuam reconhecendo os antígenos de microorganismos intracelulares, destruindo-os ou destruindo as células infectadas (ABBAS; LICHTMAN, 2007). Existem milhões de células B e T circulando pelo corpo através do sistema linfático, caracterizando-se como detectores móveis e independentes, trabalhando em conjunto no processo de detecção e eliminação de invasores (ABBAS; LICHTMAN; PILLAI, 2008).

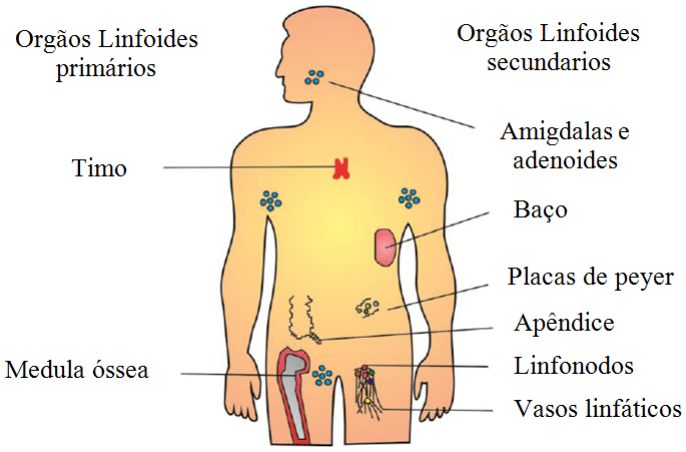
APÊNDICE B – ANATOMIA DO SISTEMA IMUNOLÓGICO

Os órgãos e tecidos que compõem o sistema imunológico humano estão distribuídos por todo o corpo. O conjunto desses órgãos forma o sistema linfático, pois estão diretamente relacionados com a produção, crescimento e o desenvolvimento de linfócitos (DE CASTRO, 2001).

De um modo geral, são nos órgãos linfoides que os linfócitos interagem com outros tipos de células durante o processo de maturação e, também, durante o início de uma resposta adaptativa. Classificados em primários e em secundários, os órgãos linfoides são responsáveis pela produção e maturação de linfócitos (órgãos primários), e também por estimularem a produção de anticorpos (órgãos secundários)(DE CASTRO, 2001; ABBAS; LICHTMAN; PILLAI, 2008).

O sistema linfático está distribuído pelo corpo humano conforme a Figura 22.

Figura 22 – Anatomia do Sistema Imunológico Humano



Fonte: Adaptado de de Castro (2001)

Os órgãos linfoides primários são o timo e a medula óssea:

- **Timo**: localizado na porção superior do tórax, é responsável pelo desenvolvimento das células T. Algumas células migram para o timo a partir da medula óssea, onde se multiplicam e amadurecem, transformando-se em células T (DE CASTRO, 2001; BHATTACHARYA, 2006).
- **Medula óssea**: local onde ocorre a geração dos elementos celulares do sangue (hematopoese), incluindo as hemácias, os monócitos, os leucócitos polimorfonucleares (granulócitos), os linfócitos B e as plaquetas. Nos mamíferos, é o local onde se de-

senvolvem as células B e as células-tronco que dão origem aos linfócitos T após a migração para o timo (DE CASTRO, 2001; BHATTACHARYA, 2006).

Os órgãos linfoides secundários são as amígdalas e adenoides, baço, placas de peyer, apêndice, linfonodos e vasos linfáticos:

- **Amígdalas e Adenoides**: constituem grandes agregados de células linfoides organizadas como parte do sistema imune associado a mucosas ou ao intestino (DE CASTRO, 2001).
- **Linfonodos**: Atuam como regiões de convergências de um extenso sistema de vasos que atuam na coleta do fluido extracelular dos tecidos, fazendo-o retornar para o sangue. Este fluido é denominado linfa e é produzido de maneira constante por filtragem do sangue. É também o ambiente onde ocorre a resposta imunológica adaptativa (DE CASTRO, 2001; BHATTACHARYA, 2006).
- **Apêndice e Placas de Peyer**: linfonodos especializados contendo células imunológicas destinadas à proteção gastrointestinal (DE CASTRO, 2001).
- **Baço**: único órgão linfoide entreposto na corrente sanguínea, sendo o local onde linfócitos combatem os organismos que invadem a corrente sanguínea. É responsável pela remoção de células sanguíneas envelhecidas e também por responder aos antígenos levados ao baço pelo sangue (DE CASTRO, 2001).
- **Vasos linfáticos**: rede de canais que transporta a linfa para o sangue e órgãos linfoides. Os vasos aferentes drenam o líquido dos tecidos e carregam as células portadoras dos antígenos dos locais de infecção para os órgãos linfáticos (linfonodos). Neste estágio, as células apresentam o antígeno aos linfócitos que estão circulando, os quais elas ajudam a ativar. Após passarem por processos de proliferação e diferenciação, estes linfócitos deixam os linfonodos como células efetoras através dos vasos linfáticos eferentes (DE CASTRO, 2001; ABBAS; LICHTMAN; PILLAI, 2008).

APÊNDICE C – RECONHECIMENTO DE PADRÕES NO SIH

As células B e T possuem moléculas receptoras (reconhecedoras) em suas superfícies capazes de reconhecer antígenos com características distintas. As células B possuem receptores capazes de reconhecer os antígenos livres em solução (como no sangue), enquanto que os receptores das células T reconhecem antígenos apresentados por células acessórias (DE CASTRO, 2001).

Cada célula B produz um tipo específico de anticorpo, com a capacidade de reconhecer e ligar-se a um determinado antígeno. É uma forma de sinalizar a outras células para que façam a ingestão, processamento ou remoção da substância identificada. Já o receptor da célula T reconhece apenas antígenos já processados (TIMMIS, 2001; DE CASTRO, 2001).

As células T desempenham um papel fundamental no SIH, elas precisam identificar e discriminar sobre o que é próprio e o que é não próprio, ou seja, elas identificam os antígenos do próprio organismo para que estes não sejam confundidos com agentes infecciosos. Para isso, o SIH utiliza alguns mecanismos, dentre eles podemos citar os processos de seleção negativa, seleção positiva e seleção clonal (DE CASTRO, 2001; ABBAS; LICHTMAN, 2007).

O principal objetivo do processo de seleção positiva é selecionar (identificar) aquelas células capazes de atuar em uma resposta imune adaptativa. Em outras palavras, o processo de seleção positiva realiza a identificação daquelas células capazes de reconhecer um determinado estímulo antigênico (DE CASTRO, 2001).

A teoria da seleção clonal está associada às características básicas de uma resposta imune adaptativa a um estímulo antigênico. Apenas células capazes de reconhecer um determinado estímulo antigênico irão se proliferar, sendo, portanto, selecionadas em detrimento das outras. Utilizando o processo de seleção clonal, o SIH produz células efetoras específicas em quantidade suficiente para combater uma determinada infecção (DE CASTRO, 2001; ABBAS; LICHTMAN; PILLAI, 2008).

APÊNDICE D – AVALIAÇÃO DETALHADA DO CMMV PROPOSTO

Neste apêndice, uma análise mais detalhada sobre a metodologia proposta nesta pesquisa é apresentada. A avaliação realizada no Capítulo 5 é pertinente, pois avalia e compara o CMMV proposto frente a diferentes condições de ruído, sem o conhecimento prévio do tipo de ruído de fundo contido nos sinais. De um modo geral, a metodologia proposta se mostrou robusta e apresentou melhor desempenho frente a condições adversas de ruído. No entanto, a fim de verificar os resultados para tipos de ruído específicos, seguem as Tabelas 12, 13, 14, 15, 16 e 17. O objetivo é verificar se houve discrepâncias em relação ao tipo de ruído, ou seja, verificar se para algum tipo de ruído, o método proposto se mostrou ineficiente. Novamente, os resultados dos algoritmos tradicionais também serão apresentados, a fim de que se tenha parâmetros para julgar se o processamento é adequado ou não.

Tabela 12 – Avaliações objetivas dos sinais processados para o ruído Vozes.

Medida	SNR	MMSE	WIENER	SE	comp-wav	CMMV
C_{ovl}	0dB	2,17	1,84	1,66	**2,23**	2,21
	5dB	2,63	2,37	2,13	2,60	**2,70**
	10dB	3,07	2,90	2,59	2,87	**3,11**
	15dB	3,56	3,42	3,06	3,17	**3,58**
C_{bak}	0dB	1,89	1,73	1,56	1,92	**1,94**
	5dB	2,32	2,19	1,94	**2,37**	**2,37**
	10dB	2,76	2,68	2,32	2,73	**2,79**
	15dB	3,24	3,15	2,69	3,03	**3,25**
C_{sig}	0dB	2,66	2,2	2,01	2,74	**2,75**
	5dB	3,20	2,82	2,55	3,1	**3,30**
	10dB	3,69	3,41	3,07	3,31	**3,72**
	15dB	4,20	3,97	3,57	3,55	**4,23**
segSNR	0dB	-2,54	**-1,77**	-2,32	-3,11	-2,37
	5dB	0,23	**0,84**	-0,40	0,50	0,38
	10dB	3,42	**3,76**	1,45	3,29	3,27
	15dB	**6,69**	**6,69**	3,20	4,97	6,53
PESQ	0dB	**1,92**	1,85	1,75	1,91	1,88
	5dB	**2,25**	2,21	2,07	**2,25**	**2,25**
	10dB	2,57	**2,59**	2,40	2,55	2,58
	15dB	**2,97**	3,00	2,75	2,90	2,96

Fonte: Elaborado pelo próprio autor.

Tabela 13 – Avaliações objetivas dos sinais processados para o ruído Cafeteria.

Medida	SNR	MMSE	WIENER	SE	comp-wav	CMMV
C_{ovl}	0dB	2,25	2,03	1,75	1,91	**2,34**
	5dB	2,65	2,46	2,15	2,14	**2,72**
	10dB	3,09	2,98	2,62	2,37	**3,16**
	15dB	3,52	3,44	3,04	2,75	**3,62**
C_{bak}	0dB	1,97	1,90	1,63	1,78	**2,03**
	5dB	2,34	2,28	1,95	2,06	**2,39**
	10dB	2,78	2,74	2,35	2,35	**2,81**
	15dB	3,23	3,20	2,69	2,66	**3,30**
C_{sig}	0dB	2,68	2,35	2,03	2,12	**2,80**
	5dB	3,13	2,84	2,49	2,36	**3,21**
	10dB	3,63	3,43	3,04	2,56	**3,73**
	15dB	4,10	3,92	3,49	2,96	**4,23**
segSNR	0dB	-1,41	**-0,41**	-1,46	-3,10	-1,27
	5dB	0,89	**1,62**	-0,07	-0,74	0,94
	10dB	3,88	**4,39**	1,72	1,75	3,78
	15dB	7,05	7,20	3,21	3,26	**7,21**
PESQ	0dB	2,13	2,11	1,97	2,04	**2,17**
	5dB	2,42	2,41	2,23	2,23	**2,44**
	10dB	2,73	**2,74**	2,52	2,46	**2,74**
	15dB	3,06	**3,11**	2,83	2,79	3,09

Fonte: Elaborado pelo próprio autor.

Tabela 14 – Avaliações objetivas dos sinais processados para o ruído Salão de Exibição.

Medida	SNR	MMSE	WIENER	SE	comp-wav	CMMV
C_{ovl}	0dB	**2,03**	1,76	1,53	1,92	1,89
	5dB	**2,53**	2,36	2,08	2,35	2,45
	10dB	**2,99**	2,87	2,51	2,86	2,93
	15dB	**3,41**	3,28	2,86	3,22	3,35
C_{bak}	0dB	**1,91**	1,76	1,54	1,89	1,83
	5dB	2,34	2,25	1,95	**2,35**	2,31
	10dB	2,77	2,71	2,30	**2,78**	**2,78**
	15dB	3,20	3,12	2,61	3,04	**3,22**
C_{sig}	0dB	**2,46**	2,12	1,83	2,25	2,32
	5dB	**3,04**	2,79	2,47	2,71	2,92
	10dB	**3,57**	3,34	2,95	3,26	3,42
	15dB	**4,02**	3,79	3,30	3,63	3,86
segSNR	0dB	-2,19	**-1,00**	-1,98	-2,87	**-1,00**
	5dB	0,74	1,64	-0,04	0,83	**1,74**
	10dB	3,83	4,31	1,75	3,67	**4,62**
	15dB	7,05	7,06	3,31	5,00	**7,75**
PESQ	0dB	**1,81**	1,75	1,66	1,76	1,77
	5dB	2,17	2,19	2,05	2,13	**2,21**
	10dB	2,52	2,59	2,37	2,56	**2,61**
	15dB	2,86	2,91	2,67	2,90	**2,94**

Fonte: Elaborado pelo próprio autor.

Tabela 15 – Avaliações objetivas dos sinais processados para o ruído Carro.

Medida	SNR	MMSE	WIENER	SE	comp-wav	CMMV
C_{ovl}	0dB	2,77	**2,83**	2,36	1,72	2,82
	5dB	3,13	**3,16**	2,7	1,94	3,15
	10dB	3,51	**3,63**	3,08	2,28	3,62
	15dB	3,88	**3,94**	3,42	2,71	3,93
C_{bak}	0dB	2,29	2,48	2,01	1,65	**2,49**
	5dB	2,66	**2,82**	2,31	1,93	**2,82**
	10dB	3,09	**3,26**	2,65	2,24	3,25
	15dB	3,50	**3,62**	2,95	2,57	3,61
C_{sig}	0dB	3,26	**3,28**	2,79	1,76	3,27
	5dB	**3,67**	3,64	3,15	1,99	3,62
	10dB	4,08	**4,11**	3,56	2,42	4,09
	15dB	**4,45**	4,43	3,91	2,89	4,41
segSNR	0dB	-0,45	1,88	-0,46	-3,91	**1,92**
	5dB	2,13	**4,05**	0,99	-1,6	3,99
	10dB	5,31	**6,71**	2,59	0,78	6,59
	15dB	8,44	**9,38**	3,89	2,37	9,23
PESQ	0dB	2,52	**2,64**	2,32	2,08	**2,64**
	5dB	2,78	**2,89**	2,57	2,26	**2,89**
	10dB	3,08	**3,29**	2,85	2,46	3,28
	15dB	3,39	**3,54**	3,10	2,79	3,53

Fonte: Elaborado pelo próprio autor.

Tabela 16 – Avaliações objetivas dos sinais processados para o ruído Tráfego.

Medida	SNR	MMSE	WIENER	SE	comp-wav	CMMV
C_{ovl}	0dB	1,79	**1,86**	1,62	1,62	1,68
	5dB	2,23	**2,37**	2,05	2,09	**2,37**
	10dB	2,69	**2,86**	2,48	2,65	2,82
	15dB	3,13	**3,22**	2,82	3,13	3,09
C_{bak}	0dB	1,80	**1,93**	1,71	1,77	1,80
	5dB	2,16	**2,32**	2,00	2,17	**2,32**
	10dB	2,61	**2,75**	2,36	2,62	**2,75**
	15dB	3,06	**3,15**	2,67	2,93	3,07
C_{sig}	0dB	1,98	**2,01**	1,72	1,70	1,76
	5dB	2,53	**2,66**	2,28	2,28	**2,66**
	10dB	3,10	**3,22**	2,76	2,87	3,15
	15dB	3,59	**3,61**	3,14	3,41	3,43
segSNR	0dB	-1,74	**0,22**	-0,66	-1,81	-0,21
	5dB	0,76	2,41	0,79	1,38	**2,49**
	10dB	3,99	4,99	2,49	3,61	**5,15**
	15dB	7,17	**7,70**	3,79	4,76	7,66
PESQ	0dB	1,95	**2,13**	2,01	1,90	2,06
	5dB	2,22	**2,40**	2,24	2,24	**2,40**
	10dB	2,51	**2,75**	2,55	2,68	**2,75**
	15dB	2,83	3,01	2,76	**3,05**	2,96

Fonte: Elaborado pelo próprio autor.

Tabela 17 – Avaliações objetivas dos sinais processados para o ruído Trem.

Medida	SNR	MMSE	WIENER	SE	comp-wav	CMMV
C_{ovl}	0dB	**2,20**	2,00	1,57	1,89	2,15
	5dB	2,55	2,47	2,04	2,02	**2,59**
	10dB	2,96	2,92	2,48	2,37	**3,05**
	15dB	3,37	3,32	2,88	2,91	**3,48**
C_{bak}	0dB	**2,06**	1,99	1,64	1,93	**2,06**
	5dB	2,41	2,40	2,01	2,26	**2,48**
	10dB	2,79	2,79	2,35	2,59	**2,90**
	15dB	3,21	3,19	2,68	2,92	**3,35**
C_{sig}	0dB	**2,62**	2,37	1,84	2,09	2,60
	5dB	3,07	2,91	2,40	2,15	**3,08**
	10dB	3,54	3,42	2,89	2,56	**3,59**
	15dB	4,00	3,86	3,36	3,25	**4,03**
segSNR	0dB	-1,78	-0,66	-1,80	-3,20	**-0,63**
	5dB	1,08	1,91	0,06	-0,04	**2,09**
	10dB	3,86	4,42	1,73	2,76	**4,98**
	15dB	7,31	7,58	3,51	4,94	**8,45**
PESQ	0dB	**1,92**	1,86	1,67	1,86	1,90
	5dB	2,14	2,19	2,00	2,03	**2,24**
	10dB	2,45	2,54	2,31	2,27	**2,61**
	15dB	2,77	2,85	2,58	2,65	**2,98**

Fonte: Elaborado pelo próprio autor.

APÊNDICE E – ARTIGOS PUBLICADOS

Segue abaixo a lista dos trabalhos publicados e que são frutos da pesquisa desenvolvida. Todos os trabalhos estão diretamente relacionados às ferramentas estudadas e investigações realizadas ao decorrer do desenvolvimento do Doutoramento. Os trabalhos com numeração [5] e [7] foram desenvolvidos durante o estudo da DT−ℂWT, realizado no Capítulo 2. Os estudos realizados Capítulo 2 geraram, inclusive, o trabalho de iniciação científica orientado pelo autor desta tese em [3], onde a escolha dos filtros wavelet utilizado na DWT foi abordado. A análise de métodos realizada no Capítulo 3 resultou nas publicações [2] e [8]. Ainda como fruto do Capítulo 3, a publicação número [6] relata o método de melhoramento de voz baseado na DT−ℂWT. A publicação de número [1] é a principal publicação referente ao Capítulo 4, seguido pelo trabalho de iniciação científica em [4]. É importante destacar que os trabalhos relacionados ao Capítulo 5 ainda estão sendo preparados e serão submetidos para apreciação por parte de revistas a serem selecionadas.

[1] ABREU, C. C. E.; DUARTE, M. A. Q.; VILLARREAL, F. An immunological approach based on the negative selection algorithm for real noise classification in speech signals. *AEU. International Journal of Electronics and Communications*,Muenchen, v. 72, p. 125-133, 2017.

[2] ABREU, C. C. E; TRAVASSOS, N. C. L.; DUARTE, M. A. Q.; VILLARREAL, F. A comparative study between recent wavelet nonthresholding methods and the well-established spectral subtractive and statistical-model-based algorithms for speech enhancement under real noisy conditions. In: IEEE/IAS INTERNATIONAL CONFERENCE ON INDUSTRY APPLICATIONS - INDUSCON, 12, 2016, Curitiba. *Proceedings...* Curitiba: IEEE/IAS, 2016. p.1-8.

[3] MOURÃO, E. S. ; ABREU, C. C. E. Sobre a escolha dos filtros wavelet e da função janela para sistemas de melhoramento de voz baseados na DWT. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT 2016, 34, 2016, Santarém. *Anais...* Santarém:[S.n], 2016. p. 635-636.

[4] PETRIKICZ, D. H.; OLIVEIRA, B. R.; ABREU, C. C. E. Aplicação do algoritmo de seleção negativa na detecção de arritmias cardíacas. SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT 2016, 34, 2016, Santarém. *Anais...* Santarém:[S.n], 2016. p. 627-628.

[5] ABREU, C. C. E.; CHAVARETTE, F. R. ; DUARTE, M. A. Q. ; VILLARREAL, F. Analysis of the structural integrity of a building by complex wavelets. *International*

Journal of Applied Mathematics, Sofia, v. 28, n. 2, p. 159-164, 2015.

[6] ABREU, C. C. E.; DUARTE, M. A. Q.; VILLARREAL, F. Dual-tree Complex Wavelet Transform in the problem of speech enhancement. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, São Carlos, v. 3, n. 1, 2015.

[7] ABREU, C. C. E.; CHAVARETTE, F. R.; VILLARREAL, F.; DUARTE, M. A. Q.; LIMA, F. P. A. Dual-tree complex wavelet transform applied to fault monitoring and identification in aeronautical structures. *International Journal of Pure and Applied Mathematics*, Sofia, v. 97, n. 1, p. 89-97, 2014.

[8] ABREU, C. C. E.; DUARTE, M. A. Q. ; VILLARREAL, F. Analysis of the evolution speech enhancement methods in wavelet domain. In: CONGRESSO DE MATEMÁTICA APLICADA E COMPUTACIONAL - CMAC,2, Bauru, 2013. *Proceedings...* Bauru: [S.n], 2013. p. 555-560.