



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE TECNOLOGIA**  
**DEPARTAMENTO DE TELEINFORMÁTICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA**  
**DOUTORADO EM TELEINFORMÁTICA**

**FÁBIO CISNE RIBEIRO**

**RECONHECIMENTO DE COMANDOS DE VOZ EM PORTUGUÊS BRASILEIRO**  
**EM AMBIENTES RUIDOSOS USANDO LARINGOFONE**

**FORTALEZA**

**2019**

FÁBIO CISNE RIBEIRO

RECONHECIMENTO DE COMANDOS DE VOZ EM PORTUGUÊS BRASILEIRO EM  
AMBIENTES RUIDOSOS USANDO LARINGOFONE

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia Teleinformática. Área de Concentração: Sinais e Sistemas

Orientador: Prof. Dr. Paulo César Cortez

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

R369r Ribeiro, Fábio Cisne.

Reconhecimento de comandos de voz em Português Brasileiro em ambientes ruidosos usando laringofone / Fábio Cisne Ribeiro. – 2019.  
130 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2019.

Orientação: Prof. Dr. Paulo César Cortez.

1. reconhecimento de voz. 2. laringofone. 3. reconhecimento de padrões. 4. redes neurais. I. Título.

CDD 621.38

---

FÁBIO CISNE RIBEIRO

RECONHECIMENTO DE COMANDOS DE VOZ EM PORTUGUÊS BRASILEIRO EM  
AMBIENTES RUIDOSOS USANDO LARINGOFONE

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em: 31 de Janeiro de 2019

BANCA EXAMINADORA

---

Prof. Dr. Paulo César Cortez (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Jose Marques Soares  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Pedro Pedrosa Rebouças Filho  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Ajalmar Rego da Rocha Neto  
Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

---

Prof. Dr. João Paulo do Vale Madeiro  
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

---

Prof. Dr. Francisco Madeiro Bernardino Junior  
Universidade Católica de Pernambuco (UNICAP)

## **AGRADECIMENTOS**

Primeiramente a Deus, pela graça da vida.

A minha esposa, aos meus pais, irmãos e familiares, pelo amor, apoio, paciência e compreensão que sempre me foram fornecidos, tanto no decorrer de minha formação acadêmica, quanto ao longo de minha vida, para que eu aqui chegasse.

Ao Prof. Dr. Paulo César Cortez e ao Prof. Dr. José Marques Soares pelas orientações de vida e acadêmicas, inclusive na produção desta tese de doutorado; e, principalmente, pela grande amizade conquistada.

Agradeço a todos os professores, alunos e colaboradores do Laboratório de Engenharia de Sistemas de Computação (LESC) e do Departamento de Engenharia de Teleinformática (DETI), da Universidade Federal do Ceará (UFC) que me auxiliaram direta e indiretamente no desenvolvimento deste trabalho.

Aos amigos e a todas as pessoas que me ajudaram nesta jornada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Agradeço, também, o apoio financeiro da Financiadora de Inovação e Pesquisa (FINEP).

“Tudo o que um sonho precisa para ser realizado  
é alguém que acredite que ele possa ser reali-  
zado.”

(Roberto Shinyashiki)

## RESUMO

Esta tese tem como objetivo principal o desenvolvimento de um sistema para reconhecimento de comandos de voz em ambientes ruidosos através de palavras isoladas e independentes do locutor, com ênfase no uso do laringofone, que é um sensor de aquisição do sinal da fala mais robusto para ambientes ruidosos. A tecnologia estudada apresenta-se através de dispositivos integrados de hardware e software, que permitem usar a fala como instrumento de operação de equipamentos tecnológicos. Assim, foram pesquisadas quais técnicas que melhor se adéquam para realização do processamento de voz proposto. Como não há outro conjunto de dados com comandos de voz capturados usando o laringofone na língua Portuguesa do Brasil na literatura pesquisada, criamos um conjunto de dados com comandos de voz isolados com elocuições capturadas de 150 pessoas (homens e mulheres). Todas as amostras de voz são capturadas em Português Brasileiro, e são os dígitos “0” a “9” e as palavras “Ok” e “Cancelar”. Para remover os ruídos capturados, dois filtros foram utilizados, o *Least Mean Squares* no espaço temporal e a Transformada Wavelet no espaço em frequência, de forma que esse conjunto permitiu remover os ruídos que são capturados pelo laringofone. O melhor extrator de características testado é o *Perceptual Linear Prediction* e sua melhor configuração é utilizando 9 ou 10 índices na ordem dos seus coeficientes. Para classificação utilizou-se um comitê votador composto por três classificadores, Perceptron Multicamadas, Perceptron Multicamadas Binário e Mapas Auto-Organizáveis para reconhecer o comando de voz. Os resultados mostram que o laringofone é robusto no ambiente de ruído, alcançando 96,6% de taxa de acertos em nosso sistema de reconhecimento de comandos de voz. Foi observado que vogais com baixa intensidade e fricativos presentes nas palavras “3” e “7” em Português confundem o classificador.

**Palavras-chave:** reconhecimento de voz, laringofone, reconhecimento de padrões, redes neurais

## ABSTRACT

This thesis has as main objective the development of a system for voice commands recognition in noisy environments through isolated words spoken independent of a speaker, with emphasis on the use of throat microphone which is a acquisition sensor for speech signal more robust for this type of environment. The technology studied is presented through integrated hardware and software device that allow the use of speech as an instrument for the operation of a technological equipment. Thus, were research which techniques are best to perform the proposed voice processing. There is no other database with voice commands captured using throat microphone in Portuguese language in the researched literature. We created a database with isolated voice commands with captured utterances of 150 people (men and women). All voice samples are captured in Brazilian Portuguese, and are the digits “0” through “9” and the words “Ok” and “Cancel”. To remove the captured noises two filters were used, the Least Mean Squares in the temporal space and the Wavelet Transform in the space in frequency, so that this set allowed to remove the noises that are captured by the laringophone. The best feature extractor tested is the Perceptual LinearPrediction and its best configuration is the use of 9 or 10 indexes in the order of their coefficients. For classification it been used a voting committee composed of three classifiers, MLP, BMLP and SOM to recognize the voice command. For classification a voting committee composed of three classifiers, Multilayer Perceptron, Binary Multilayer Perceptron and Self-Organizing Maps to recognize command of voice. The results show that throat microphone is robust in noise environment, reaching 96,6% of hit rate in our voice command recognition system. It was observed that vowels with low intensity and fricatives present in the words “3” and “7” in Portuguese confuse the classifier.

**Keywords:** speech recognition, throat microphone, pattern recognition, neural networks



## LISTA DE FIGURAS

Figura 1 – Etapas de um sistema de reconhecimento de comandos de voz. . . . .	37
Figura 2 – Posição de utilização do laringofone. . . . .	38
Figura 3 – Modelo <i>Least Mean Squares</i> . . . . .	43
Figura 4 – Cancelamento de interferência. . . . .	43
Figura 5 – Filtro LMS preditivo. . . . .	44
Figura 6 – Decomposição Wavelet. . . . .	45
Figura 7 – Função de limiarização, (a) sinal original, (b) limiarização dura, (c) limiarização suave. . . . .	46
Figura 8 – Diagrama de blocos da filtragem Wavelet. . . . .	47
Figura 9 – Diagrama de blocos do algoritmo <i>endpoints</i> . . . . .	48
Figura 10 – Visão geral da abordagem de análise de curta duração. . . . .	50
Figura 11 – Diagrama de blocos do extrator de características <i>Linear Predictive Coding</i> . . . . .	51
Figura 12 – Representação dos componentes da fala no espectro de amplitude. . . . .	52
Figura 13 – Representação dos componentes da fala no domínio cepstral. . . . .	52
Figura 14 – Diagrama de blocos do extrator de características <i>Linear Prediction Cepstrum Coefficients</i> . . . . .	53
Figura 15 – Diagrama de blocos do extrator de características <i>Weighted Linear Predictive Cepstral Coefficients</i> . . . . .	54
Figura 16 – Diagrama de blocos do conceito do <i>Perceptual Linear Prediction</i> . . . . .	54
Figura 17 – Cálculo dos coeficientes <i>Perceptual Linear Prediction Cepstral</i> . . . . .	55
Figura 18 – Modelo de um neurônio não linear. . . . .	60
Figura 19 – Exemplo de uma rede <i>Multilayer Perceptron</i> / perceptron multicamada (MLP) completa. . . . .	61
Figura 20 – Exemplo de uma das sub-redes <i>Binary Multilayer Perceptron</i> / perceptron multicamada binário (BMLP). . . . .	64
Figura 21 – Exemplo de uma rede BMLP completa usando as abordagens <i>One-Against-All</i> / um-contra-todos (OAA) e <i>One-Against-One</i> / um-contra-um (OAO). . . . .	65
Figura 22 – Conexão dos neurônios com os dados. . . . .	68
Figura 23 – Comportamento das curvas das funções <i>kernel</i> de vizinhança. . . . .	70
Figura 24 – Diagrama de uma rede <i>Learning vector quantization</i> (LVQ). . . . .	71
Figura 25 – Representação gráfica do Algoritmo <i>Bagging</i> . . . . .	75

Figura 26 – Representação da tabela de votação do comitê. . . . .	76
Figura 27 – Fluxograma de funcionamento do comitê. . . . .	77
Figura 28 – Diagrama de blocos do sistema automático de reconhecimento de comandos de voz desenvolvido nesta tese. . . . .	82
Figura 29 – Modelo do laringofone utilizado. . . . .	83
Figura 30 – Histograma da quantidade de amostras por classe da base de voz capturada. . . . .	84
Figura 31 – Raspberry Pi Zero W V1.1. . . . .	97
Figura 32 – Dicionário de comandos: “zero”, “um” e “dois”. . . . .	100
Figura 33 – Dicionário de comandos: “três”, “quatro” e “cinco”. . . . .	101
Figura 34 – Dicionário de comandos: “seis”, “sete” e “oito”. . . . .	102
Figura 35 – Dicionário de comandos: “nove”. . . . .	103
Figura 36 – Sinal de voz antes e após o processo de detecção contínua de voz. . . . .	104
Figura 37 – Sinal de voz antes e após o processamento do filtro <i>Least Mean Squares</i> (LMS). . . . .	105
Figura 38 – Sinal de voz referente ao comando “cinco” após o processamento pela transformada Wavelet. . . . .	105
Figura 39 – Sinal de voz capturado e segmentado do comando “cinco”. . . . .	106
Figura 40 – Segmentações do comando de voz “um” e “nove”, através do processo de análise de curta duração. . . . .	107
Figura 41 – Resultados da ordem dos coeficientes do extrator de características <i>Perceptual Linear Prediction</i> (PLP). . . . .	108
Figura 42 – Resultados em <i>boxplot</i> da taxa de acerto do classificador MLP nas ordem dos coeficientes do PLP. . . . .	109
Figura 43 – Resultados da taxa de acerto média da MLP por classe. . . . .	109
Figura 44 – Resultados completos da taxa de acerto média da MLP por classe, e por ordem PLP. . . . .	110
Figura 45 – Resultados em <i>boxplot</i> da taxa de acerto do classificador BMLP nas ordem dos coeficientes do PLP. . . . .	111
Figura 46 – Resultados da taxa de acerto média da BMLP por classe. . . . .	112
Figura 47 – Resultados completos da taxa de acerto média da BMLP por classe, e por ordem PLP. . . . .	113
Figura 48 – Boxplot do resultado das taxas de acerto dos classificadores e do comitê. . . . .	115
Figura 49 – Raspberry Pi Zero W V1.1. . . . .	130

## LISTA DE TABELAS

Tabela 1 – Lista de palavras-chave da revisão de literatura. . . . .	24
Tabela 2 – Lista das <i>strings</i> de busca. . . . .	25
Tabela 3 – Resultado das buscas das <i>strings</i> nos indexadores em 30/08/2018. . . . .	25
Tabela 4 – Siglas das aplicações e técnicas identificadas. . . . .	27
Tabela 5 – Lista dos estudos da revisão de literatura e suas referências. . . . .	28
Tabela 6 – Classificação dos estudos por aplicação. . . . .	28
Tabela 7 – Classificação dos estudos por técnica. . . . .	30
Tabela 8 – Comparação da taxa de acerto do reconhecimento de voz dos estudos em ambiente limpo e ruidoso usando microfone acústico (AM) e o laringofone (TM). . . . .	39
Tabela 9 – Tipos de funções de vizinhança. . . . .	70
Tabela 10 – Resultado da avaliação dos extratores de características. . . . .	92
Tabela 11 – Parâmetros da arquitetura do Perceptron Multicamadas. . . . .	94
Tabela 12 – Parâmetros do treinamento coletivo. . . . .	95
Tabela 13 – Resultado da quantidade de amostras do banco de dados. . . . .	98
Tabela 14 – Resultado das taxas de acerto média por classe e geral da rede <i>Self-Organizing Maps</i> / mapa auto-organizável (SOM). . . . .	112
Tabela 15 – Resultado das taxas de acerto de todos os classificadores em conjunto com o comitê. . . . .	114

## LISTA DE ABREVIATURAS E SIGLAS

AM	<i>Acoustic Microphone</i> / microfone acústico
ANN	<i>Artificial Neural Network</i> / redes neurais artificiais
ASI	<i>Automatic Speaker Identification</i> / identificação automática de indivíduos
ASR	<i>Automatic Speech Recognition</i> / sistema de reconhecimento de voz
ASV	<i>Automatic Speaker Verification</i> / verificação automática de indivíduos
AWR	<i>Automatic Word Recognition</i> / sistema de reconhecimento de comandos de voz
BLE	<i>Bluetooth Low Energy</i>
BMLP	<i>Binary Multilayer Perceptron</i> / perceptron multicamada binário
BMU	<i>Best Matching Neuron</i>
CSI	<i>Camera Serial Interface</i>
FFT	<i>Fast Fourier Transform</i> / transformada rápida de Fourier
GOA	Google Acadêmico
GPIO	<i>General Purpose Input/Output</i>
GPU	<i>Graphic Processor Unit</i>
HDMI	<i>High-Definition Multimedia Interface</i>
HMM	<i>Hidden Markov Model</i> / modelos ocultos de Markov
KNN	<i>k-Nearest Neighborhood</i> / <i>k</i> -vizinho mais próximo
LMS	<i>Least Mean Squares</i>
LP	<i>Linear Prediction</i> / predição linear
LPC	<i>Linear Predictive Coding</i> / codificação linear preditiva
LPCC	<i>Linear Predictive Cepstral Coefficients</i> / coeficientes de predição linear representados no domínio cepstral
LVQ	<i>Learning vector quantization</i>
MFCC	<i>Mel-Frequency Cepstrum Coefficients</i>
MLP	<i>Multilayer Perceptron</i> / perceptron multicamada
MSE	<i>Mean Square Error</i> / erro quadrático médio
NN	<i>Nearest Neighborhood</i> / vizinho mais próximo
OAA	<i>One-Against-All</i> / um-contra-todos
OAO	<i>One-Against-One</i> / um-contra-um
OTG	<i>On-The-Go</i>
PLP	<i>Perceptual Linear Prediction</i>

PR	<i>Pattern Recognition</i> / reconhecimento de padrões
RCA	<i>Radio Corporation of America</i>
SNR	<i>Signal-to-Noise Ratio</i> / relação sinal-ruído
SOM	<i>Self-Organizing Maps</i> / mapa auto-organizável
SR	<i>Speech Recognition</i> / reconhecimento de voz
TM	<i>Throat Microphone</i> / laringofone, microfone de garganta ou microfone de contato
USB	<i>Universal Serial Bus</i>
WAV	<i>WAVEform audio format</i>
WLPPC	<i>Weighted Linear Predictive Cepstral Coefficients</i> / coeficientes de predição linear balanceados representados no domínio cepstral
WT	<i>Wavelet Transform</i> / transformada Wavelet
WTA	<i>Winner-Take-All</i>
ZCR	<i>Zero-Crossing Rate</i> / taxa de cruzamento por zero

## LISTA DE SÍMBOLOS

$ms$	Milissegundos
$e_{t_1}$ e $e_{t_2}$	Parâmetros relacionados ao limiares do algoritmo de recorte
$Z(N)$	Taxa de cruzamento por zero do N-ésimo quadro de voz
$E_m(n)$	Energia do n-ésimo quadro de voz

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>18</b>
<b>1.1</b>	<b>Objetivos</b>	<b>21</b>
<i>1.1.1</i>	<i>Objetivo Geral</i>	<i>21</i>
<i>1.1.2</i>	<i>Objetivos Específicos</i>	<i>21</i>
<b>1.2</b>	<b>Produção Científica</b>	<b>21</b>
<b>1.3</b>	<b>Estrutura da Tese</b>	<b>21</b>
<b>2</b>	<b>REVISÃO DE LITERATURA</b>	<b>23</b>
<b>2.1</b>	<b>Planejamento da revisão de literatura</b>	<b>23</b>
<i>2.1.1</i>	<i>Formulação dos objetivos</i>	<i>23</i>
<i>2.1.2</i>	<i>Definição das fontes de pesquisa</i>	<i>23</i>
<i>2.1.3</i>	<i>Definição das palavras-chave</i>	<i>24</i>
<i>2.1.4</i>	<i>Definição das strings de busca</i>	<i>25</i>
<b>2.2</b>	<b>Execução da revisão</b>	<b>25</b>
<b>2.3</b>	<b>Avaliação dos estudos</b>	<b>26</b>
<b>2.4</b>	<b>Análise dos estudos</b>	<b>31</b>
<b>2.5</b>	<b>Análise crítica</b>	<b>34</b>
<b>2.6</b>	<b>Conclusão</b>	<b>35</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>37</b>
<b>3.1</b>	<b>Visão geral</b>	<b>37</b>
<b>3.2</b>	<b>Aquisição da voz</b>	<b>37</b>
<i>3.2.1</i>	<i>Laringofone</i>	<i>38</i>
<i>3.2.2</i>	<i>Detecção contínua de voz</i>	<i>40</i>
<b>3.3</b>	<b>Pré-processamento</b>	<b>40</b>
<i>3.3.1</i>	<i>Filtragem</i>	<i>41</i>
<i>3.3.1.1</i>	<i>Filtragem Least Mean Squares (LMS)</i>	<i>42</i>
<i>3.3.1.2</i>	<i>Filtragem por Transformada Wavelet (WT)</i>	<i>44</i>
<i>3.3.2</i>	<i>Detecção de extremos</i>	<i>47</i>
<b>3.4</b>	<b>Extração de características</b>	<b>48</b>
<i>3.4.1</i>	<i>Análise de curta duração</i>	<i>49</i>
<i>3.4.2</i>	<i>Linear Predictive Coding (LPC)</i>	<i>49</i>

3.4.3	<i>Coeficientes Cepstrais</i> . . . . .	51
3.4.4	<i>Linear Prediction Cepstrum Coefficients (LPCC)</i> . . . . .	53
3.4.5	<i>Weighted Linear Predictive Cepstral Coefficients (WLPPC)</i> . . . . .	53
3.4.6	<i>Perceptual Linear Prediction (PLP)</i> . . . . .	54
3.5	<b>Classificadores</b> . . . . .	56
3.5.1	<i>Bayes</i> . . . . .	57
3.5.2	<i>Vizinho mais Próximo</i> . . . . .	58
3.5.3	<i>k-Vizinhos mais próximo (KNN)</i> . . . . .	58
3.5.4	<i>Redes Neurais Artificiais (ANN)</i> . . . . .	59
3.5.5	<i>Perceptron Multicamadas (MLP)</i> . . . . .	59
3.5.5.1	<i>Aprendizagem</i> . . . . .	62
3.5.6	<i>Perceptron Multicamadas Binário (BMLP)</i> . . . . .	62
3.5.6.1	<i>Abordagem um-contra-todos (OAA)</i> . . . . .	64
3.5.6.2	<i>Abordagem um-contra-um (OAO)</i> . . . . .	66
3.5.6.3	<i>Função de decisão</i> . . . . .	67
3.5.7	<i>Mapas Auto-Organizáveis (SOM)</i> . . . . .	67
3.5.7.1	<i>Algoritmo Competitivo</i> . . . . .	68
3.5.7.2	<i>Algoritmo Cooperativo</i> . . . . .	69
3.5.7.3	<i>Classificação supervisionada</i> . . . . .	71
3.5.7.3.1	<i>Learning Vector Quantization (LVQ)</i> . . . . .	71
3.5.7.3.2	<i>Métrica de classificação</i> . . . . .	72
3.5.8	<b>Comitê</b> . . . . .	72
3.5.8.1	<i>Técnicas de treinamento dos classificadores base</i> . . . . .	74
3.5.8.2	<i>Técnicas de composição da saída do comitê</i> . . . . .	75
3.5.8.2.1	<i>Seleção por votação da saída dos classificadores</i> . . . . .	75
3.5.8.2.2	<i>Fusão da saída dos classificadores</i> . . . . .	76
3.5.8.2.3	<i>Seleção de classificadores</i> . . . . .	78
3.5.8.3	<i>Estatísticas</i> . . . . .	78
3.5.8.3.1	<i>Probabilidade a priori e a posteriori aplicado a redes neurais</i> . . . . .	79
3.5.8.3.2	<i>Normalização linear aplicada a redes neurais</i> . . . . .	79
3.5.8.3.3	<i>Normalização exponencial aplicada a redes neurais</i> . . . . .	80
4	<b>SISTEMA DE RECONHECIMENTO DE COMANDOS DE VOZ</b> . . .	81



<b>4.1</b>	<b>Aquisição</b>	81
4.1.1	<i>Laringofone</i>	81
4.1.2	<i>Base de dados de comandos de voz</i>	81
4.1.3	<i>Detecção contínua de voz</i>	84
<b>4.2</b>	<b>Pré-processamento</b>	85
4.2.1	<i>Filtro Least Mean Squares (LMS)</i>	86
4.2.2	<i>Filtro baseado na Transformada Wavelet (WT)</i>	87
4.2.3	<i>Detecção de extremos</i>	88
<b>4.3</b>	<b>Extração de características</b>	91
4.3.1	<i>Análise de curta duração</i>	91
4.3.2	<i>Avaliação dos extratores de características</i>	92
<b>4.4</b>	<b>Classificadores</b>	93
4.4.1	<i>Perceptron Multicamadas (MLP)</i>	93
4.4.2	<i>Perceptron Multicamadas Binário (BMLP)</i>	93
4.4.3	<i>Self-Organizing Maps (SOM)</i>	94
4.4.4	<i>Comitê</i>	95
<b>5</b>	<b>RESULTADOS E DISCUSSÕES</b>	97
<b>5.1</b>	<b>Hardware</b>	97
<b>5.2</b>	<b>Aquisição</b>	98
5.2.1	<i>Base de dados de voz</i>	98
5.2.2	<i>Detecção contínua de voz</i>	103
<b>5.3</b>	<b>Pré-processamento</b>	104
5.3.1	<i>Filtragem Least Mean Squares</i>	104
5.3.2	<i>Filtragem por Transformada Wavelet</i>	104
5.3.3	<i>Detecção de extremos</i>	105
<b>5.4</b>	<b>Extração de características</b>	106
5.4.1	<i>Análise de curta duração</i>	106
5.4.2	<i>Perceptual Linear Prediction (PLP)</i>	107
<b>5.5</b>	<b>Classificadores</b>	108
5.5.1	<i>Perceptron Multicamadas (MLP)</i>	108
5.5.2	<i>Perceptron Multicamadas Binário (BMLP)</i>	111
5.5.3	<i>Mapas Auto-Organizáveis (SOM)</i>	112

5.5.4	<i>Comitê</i> . . . . .	114
6	<b>CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS</b> . . .	116
	<b>REFERÊNCIAS</b> . . . . .	119
	<b>ANEXOS</b> . . . . .	129
	<b>ANEXO A</b> – Especificações do Raspberry Pi Zero W V1.1 . . . . .	130