Universidade de São Paulo
Instituto de Física de São Carlos
Departamento de Física e Informática

William Habaro Ishizawa

Análise das concentrações energéticas no limiar entre
fonemas vozeados e não-vozeados e suas implicações
para fins de reconhecimento de locutores dependente
do discurso

São Carlos - SP

2015

William Habaro Ishizawa

Análise das concentrações energéticas no limiar
entre fonemas vozeados e não-vozeados e suas
implicações para fins de reconhecimento de
locutores dependente do discurso

Dissertação de Mestrado apresentada ao
Programa de Pós-graduação em Física da
Universidade de São Paulo para obtenção do
título de Mestre em Ciências.

Área de concentração: Física Aplicada -
Opção Computacional.
Orientador: Prof. Dr. Rodrigo Capobianco
Guido

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos - SP

2015

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Ishizawa, William Habaro

Análise das concentrações energéticas no limiar entre fonemas vozeados e não-vozeados e suas implicações para fins de reconhecimento de locutores dependente do discurso / William Habaro Ishizawa; orientador Rodrigo Capobianco Guido - versão corrigida -- São Carlos, 2015.

178 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Aplicada Computacional) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2015.

1. Processamento de sinais. 2. Reconhecimento de locutor. 3. Acústica. 4. Energia. 5. *Escala Bark*. I. Guido, Rodrigo Capobianco, orient. II. Título.

Dedico aos meus pais, Adalziria e Robert.

Agradecimentos

A Deus, pela benção e proteção constante a cada passo em minha vida.

A toda minha família, pelo apoio incondicional.

Ao meu orientador, pelo companheirismo, orientação, suporte e muita paciência durante o desenvolvimento deste trabalho.

Aos meus amigos, presença constante e necessária durante este trajeto.

A todos do Cursinho Alternativo, indispensáveis na minha formação e grandes responsáveis por ter chegado até aqui.

Ao Instituto de Física de São Carlos e à Universidade de São Paulo (USP), pelo conhecimento adquirido neste tempo e pela oportunidade que me foi concedida.

“No fim tudo dá certo, e se não deu certo é porque ainda não chegou ao fim.”

Fernando Sabino

Resumo

ISHIZAWA, W.H. *Análise das concentrações energéticas no limiar entre fonemas vozeados e não-vozeados e suas implicações para fins de reconhecimento de locutores dependente do discurso*. 2015. 178p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

Atualmente, diversos trabalhos e aplicações são desenvolvidos com foco na área de reconhecimento computacional de locutores. À medida que o interesse por diversas aplicações reais dentro dessa área emerge, principalmente em biometria, na qual a segurança e a eficácia são de extrema importância, torna-se cada vez mais necessário que estudos sejam feitos, na mesma proporção, visando avaliá-las. Desse modo, a proposta do presente trabalho é a de mensurar a acurácia de um sistema de reconhecimento de locutores baseado em características elementares, isto é, energias de sub-bandas de frequências, em associação com um classificador probabilístico, estudando a viabilidade de extraí-las das transições entre trechos vozeados e não-vozeados (TTVNV) dos sinais. Testes são realizados com diferentes quantidades de locutores e discurso fixado. A acurácia obtida nos testes variam de 20.18% a 92.53%. Os resultados obtidos são comparados e relatados, complementando as afirmações existentes na literatura sobre o uso das TTVNV com dados quantitativos.

Palavras-chave: Processamento de sinais. Reconhecimento de locutor. Acústica. Escala *Bark*.

Abstract

ISHIZAWA, W.H. *Analysis of energy concentrations in the threshold between voiced and unvoiced phonemes and their implications for text-dependent speaker recognition*. 2015. 178p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

Nowadays, many works and applications are developed focusing on computational speaker recognition. As the interest for several real applications within this area emerges, especially in biometrics, where the safety and the efficacy of the applications are extremely important, studies need to be developed in the same proportion, to evaluate the effectiveness of such approaches. Based on that, this work intends to measure the accuracy of a speaker recognition system that uses elementar features, i.e., sub-band frequency energies, associated with a probabilistic classifier, studying the viability of extracting them from the transition between voiced and unvoiced speech tags (TTVNV). Tests are carried out with different numbers of speakers and a text-dependent approach. The accuracy of the tests varies from 20.18% to 92.53%. The results are compared and reported, complementing the existent information on the use of TTVNV with quantitative data.

Keywords: Signal processing. Speaker recognition. Acoustics. Bark scale.

Lista de Figuras

Figura 2.1 -	Exemplo de formantes	41
Figura 2.2 -	Formantes de locutores distintos obtidos a partir de uma mesma locução	41
Figura 2.3 -	Sistema de produção da fala humana	43
Figura 2.4 -	Relação entre as frequências em Hz e na escala <i>Bark</i>	49
Figura 2.5 -	Filtros de banda de passagem	52
Figura 2.6 -	Exemplo numérico da operação de soma de convolução	54
Figura 4.1 -	Teste com 10 locutores: Locutor 10, Gráfico 1	64
Figura 4.2 -	Teste com 10 locutores: Locutor 10, Gráfico 2	65
Figura 4.3 -	Teste com 10 locutores: Locutor 10, Gráfico 3	65
Figura 4.4 -	Teste com 10 locutores: Locutor 10, Gráfico 4	65
Figura 4.5 -	Teste com 10 locutores: Locutor 10, Gráfico 5	65
Figura 4.6 -	Teste com 10 locutores: Locutor 10, Gráfico 6	65
Figura 4.7 -	Teste com 10 locutores: Locutor 10, Gráfico 7	66
Figura 4.8 -	Teste com 20 locutores: Locutor 2, Gráfico 1	67
Figura 4.9 -	Teste com 20 locutores: Locutor 2, Gráfico 2	67
Figura 4.10 -	Teste com 20 locutores: Locutor 2, Gráfico 3	67
Figura 4.11 -	Teste com 20 locutores: Locutor 2, Gráfico 4	68
Figura 4.12 -	Teste com 20 locutores: Locutor 2, Gráfico 5	68
Figura 4.13 -	Teste com 20 locutores: Locutor 2, Gráfico 6	68
Figura 4.14 -	Teste com 20 locutores: Locutor 2, Gráfico 7	68
Figura 4.15 -	Teste com 20 locutores: Locutor 2, Gráfico 8	69
Figura 4.16 -	Teste com 20 locutores: Locutor 2, Gráfico 9	69
Figura 4.17 -	Teste com 20 locutores: Locutor 2, Gráfico 10	69

Figura 4.18 -	Teste com 20 locutores: Locutor 2, Gráfico 11	69
Figura 4.19 -	Teste com 20 locutores: Locutor 2, Gráfico 12	70
Figura 4.20 -	Teste com 30 locutores: Locutor 20, Gráfico 1	71
Figura 4.21 -	Teste com 30 locutores: Locutor 20, Gráfico 2	71
Figura 4.22 -	Teste com 30 locutores: Locutor 20, Gráfico 3	72
Figura 4.23 -	Teste com 30 locutores: Locutor 20, Gráfico 4	72
Figura 4.24 -	Análise global dos resultados de <i>L versus</i> acurácia.	74
Figura 1 -	Teste com 10 locutores: Locutor 1, Gráfico 1	83
Figura 2 -	Teste com 10 locutores: Locutor 1, Gráfico 2	83
Figura 3 -	Teste com 10 locutores: Locutor 1, Gráfico 3	83
Figura 4 -	Teste com 10 locutores: Locutor 1, Gráfico 4	84
Figura 5 -	Teste com 10 locutores: Locutor 2, Gráfico 1	84
Figura 6 -	Teste com 10 locutores: Locutor 2, Gráfico 2	84
Figura 7 -	Teste com 10 locutores: Locutor 2, Gráfico 3	84
Figura 8 -	Teste com 10 locutores: Locutor 3, Gráfico 1	84
Figura 9 -	Teste com 10 locutores: Locutor 3, Gráfico 2	84
Figura 10 -	Teste com 10 locutores: Locutor 3, Gráfico 3	85
Figura 11 -	Teste com 10 locutores: Locutor 3, Gráfico 4	85
Figura 12 -	Teste com 10 locutores: Locutor 3, Gráfico 5	85
Figura 13 -	Teste com 10 locutores: Locutor 3, Gráfico 6	85
Figura 14 -	Teste com 10 locutores: Locutor 4, Gráfico 1	86
Figura 15 -	Teste com 10 locutores: Locutor 4, Gráfico 2	86
Figura 16 -	Teste com 10 locutores: Locutor 4, Gráfico 3	86
Figura 17 -	Teste com 10 locutores: Locutor 4, Gráfico 4	86
Figura 18 -	Teste com 10 locutores: Locutor 4, Gráfico 5	86
Figura 19 -	Teste com 10 locutores: Locutor 4, Gráfico 6	87
Figura 20 -	Teste com 10 locutores: Locutor 4, Gráfico 7	87
Figura 21 -	Teste com 10 locutores: Locutor 4, Gráfico 8	87
Figura 22 -	Teste com 10 locutores: Locutor 4, Gráfico 9	87

Figura 23 -	Teste com 10 locutores: Locutor 4, Gráfico 10	87
Figura 24 -	Teste com 10 locutores: Locutor 5, Gráfico 1	88
Figura 25 -	Teste com 10 locutores: Locutor 5, Gráfico 2	88
Figura 26 -	Teste com 10 locutores: Locutor 5, Gráfico 3	88
Figura 27 -	Teste com 10 locutores: Locutor 5, Gráfico 4	88
Figura 28 -	Teste com 10 locutores: Locutor 5, Gráfico 5	88
Figura 29 -	Teste com 10 locutores: Locutor 5, Gráfico 6	89
Figura 30 -	Teste com 10 locutores: Locutor 5, Gráfico 7	89
Figura 31 -	Teste com 10 locutores: Locutor 5, Gráfico 8	89
Figura 32 -	Teste com 10 locutores: Locutor 5, Gráfico 9	89
Figura 33 -	Teste com 10 locutores: Locutor 6, Gráfico 1	90
Figura 34 -	Teste com 10 locutores: Locutor 6, Gráfico 2	90
Figura 35 -	Teste com 10 locutores: Locutor 6, Gráfico 3	90
Figura 36 -	Teste com 10 locutores: Locutor 6, Gráfico 4	90
Figura 37 -	Teste com 10 locutores: Locutor 7, Gráfico 1	91
Figura 38 -	Teste com 10 locutores: Locutor 7, Gráfico 2	91
Figura 39 -	Teste com 10 locutores: Locutor 7, Gráfico 3	91
Figura 40 -	Teste com 10 locutores: Locutor 7, Gráfico 4	91
Figura 41 -	Teste com 10 locutores: Locutor 7, Gráfico 5	92
Figura 42 -	Teste com 10 locutores: Locutor 8, Gráfico 1	92
Figura 43 -	Teste com 10 locutores: Locutor 8, Gráfico 2	92
Figura 44 -	Teste com 10 locutores: Locutor 8, Gráfico 3	92
Figura 45 -	Teste com 10 locutores: Locutor 8, Gráfico 4	92
Figura 46 -	Teste com 10 locutores: Locutor 8, Gráfico 5	93
Figura 47 -	Teste com 10 locutores: Locutor 8, Gráfico 6	93
Figura 48 -	Teste com 10 locutores: Locutor 8, Gráfico 7	93
Figura 49 -	Teste com 10 locutores: Locutor 9, Gráfico 1	93
Figura 50 -	Teste com 10 locutores: Locutor 9, Gráfico 2	94
Figura 51 -	Teste com 10 locutores: Locutor 9, Gráfico 3	94

Figura 52 -	Teste com 10 locutores: Locutor 9, Gráfico 4	94
Figura 53 -	Teste com 10 locutores: Locutor 9, Gráfico 5	94
Figura 54 -	Teste com 10 locutores: Locutor 9, Gráfico 6	95
Figura 55 -	Teste com 10 locutores: Locutor 9, Gráfico 7	95
Figura 56 -	Teste com 20 locutores: Locutor 1, Gráfico 1	95
Figura 57 -	Teste com 20 locutores: Locutor 1, Gráfico 2	95
Figura 58 -	Teste com 20 locutores: Locutor 1, Gráfico 3	95
Figura 59 -	Teste com 20 locutores: Locutor 1, Gráfico 4	96
Figura 60 -	Teste com 20 locutores: Locutor 1, Gráfico 5	96
Figura 61 -	Teste com 20 locutores: Locutor 1, Gráfico 6	96
Figura 62 -	Teste com 20 locutores: Locutor 1, Gráfico 7	96
Figura 63 -	Teste com 20 locutores: Locutor 1, Gráfico 8	96
Figura 64 -	Teste com 20 locutores: Locutor 1, Gráfico 9	97
Figura 65 -	Teste com 20 locutores: Locutor 1, Gráfico 10	97
Figura 66 -	Teste com 20 locutores: Locutor 1, Gráfico 11	97
Figura 67 -	Teste com 20 locutores: Locutor 1, Gráfico 12	97
Figura 68 -	Teste com 20 locutores: Locutor 3, Gráfico 1	97
Figura 69 -	Teste com 20 locutores: Locutor 3, Gráfico 2	98
Figura 70 -	Teste com 20 locutores: Locutor 3, Gráfico 3	98
Figura 71 -	Teste com 20 locutores: Locutor 3, Gráfico 4	98
Figura 72 -	Teste com 20 locutores: Locutor 3, Gráfico 5	98
Figura 73 -	Teste com 20 locutores: Locutor 3, Gráfico 6	98
Figura 74 -	Teste com 20 locutores: Locutor 4, Gráfico 1	99
Figura 75 -	Teste com 20 locutores: Locutor 4, Gráfico 2	99
Figura 76 -	Teste com 20 locutores: Locutor 4, Gráfico 3	99
Figura 77 -	Teste com 20 locutores: Locutor 4, Gráfico 4	99
Figura 78 -	Teste com 20 locutores: Locutor 5, Gráfico 1	99
Figura 79 -	Teste com 20 locutores: Locutor 5, Gráfico 2	100
Figura 80 -	Teste com 20 locutores: Locutor 5, Gráfico 3	100

Figura 81 -	Teste com 20 locutores: Locutor 5, Gráfico 4	100
Figura 82 -	Teste com 20 locutores: Locutor 6, Gráfico 1	100
Figura 83 -	Teste com 20 locutores: Locutor 6, Gráfico 2	101
Figura 84 -	Teste com 20 locutores: Locutor 6, Gráfico 3	101
Figura 85 -	Teste com 20 locutores: Locutor 6, Gráfico 4	101
Figura 86 -	Teste com 20 locutores: Locutor 6, Gráfico 5	101
Figura 87 -	Teste com 20 locutores: Locutor 7, Gráfico 1	102
Figura 88 -	Teste com 20 locutores: Locutor 7, Gráfico 2	102
Figura 89 -	Teste com 20 locutores: Locutor 7, Gráfico 3	102
Figura 90 -	Teste com 20 locutores: Locutor 7, Gráfico 4	102
Figura 91 -	Teste com 20 locutores: Locutor 7, Gráfico 5	102
Figura 92 -	Teste com 20 locutores: Locutor 7, Gráfico 6	103
Figura 93 -	Teste com 20 locutores: Locutor 8, Gráfico 1	103
Figura 94 -	Teste com 20 locutores: Locutor 8, Gráfico 2	103
Figura 95 -	Teste com 20 locutores: Locutor 8, Gráfico 3	103
Figura 96 -	Teste com 20 locutores: Locutor 8, Gráfico 4	103
Figura 97 -	Teste com 20 locutores: Locutor 8, Gráfico 5	104
Figura 98 -	Teste com 20 locutores: Locutor 8, Gráfico 6	104
Figura 99 -	Teste com 20 locutores: Locutor 8, Gráfico 7	104
Figura 100 -	Teste com 20 locutores: Locutor 9, Gráfico 1	104
Figura 101 -	Teste com 20 locutores: Locutor 9, Gráfico 2	104
Figura 102 -	Teste com 20 locutores: Locutor 9, Gráfico 3	105
Figura 103 -	Teste com 20 locutores: Locutor 9, Gráfico 4	105
Figura 104 -	Teste com 20 locutores: Locutor 9, Gráfico 5	105
Figura 105 -	Teste com 20 locutores: Locutor 10, Gráfico 1	105
Figura 106 -	Teste com 20 locutores: Locutor 10, Gráfico 2	106
Figura 107 -	Teste com 20 locutores: Locutor 10, Gráfico 3	106
Figura 108 -	Teste com 20 locutores: Locutor 11, Gráfico 1	106
Figura 109 -	Teste com 20 locutores: Locutor 11, Gráfico 2	106

Figura 110 -	Teste com 20 locutores: Locutor 11, Gráfico 3	106
Figura 111 -	Teste com 20 locutores: Locutor 11, Gráfico 4	107
Figura 112 -	Teste com 20 locutores: Locutor 11, Gráfico 5	107
Figura 113 -	Teste com 20 locutores: Locutor 11, Gráfico 6	107
Figura 114 -	Teste com 20 locutores: Locutor 11, Gráfico 7	107
Figura 115 -	Teste com 20 locutores: Locutor 12, Gráfico 1	107
Figura 116 -	Teste com 20 locutores: Locutor 12, Gráfico 2	108
Figura 117 -	Teste com 20 locutores: Locutor 12, Gráfico 3	108
Figura 118 -	Teste com 20 locutores: Locutor 12, Gráfico 4	108
Figura 119 -	Teste com 20 locutores: Locutor 12, Gráfico 5	108
Figura 120 -	Teste com 20 locutores: Locutor 12, Gráfico 6	108
Figura 121 -	Teste com 20 locutores: Locutor 13, Gráfico 1	109
Figura 122 -	Teste com 20 locutores: Locutor 13, Gráfico 2	109
Figura 123 -	Teste com 20 locutores: Locutor 13, Gráfico 3	109
Figura 124 -	Teste com 20 locutores: Locutor 13, Gráfico 4	109
Figura 125 -	Teste com 20 locutores: Locutor 14, Gráfico 1	109
Figura 126 -	Teste com 20 locutores: Locutor 14, Gráfico 2	110
Figura 127 -	Teste com 20 locutores: Locutor 14, Gráfico 3	110
Figura 128 -	Teste com 20 locutores: Locutor 14, Gráfico 4	110
Figura 129 -	Teste com 20 locutores: Locutor 14, Gráfico 5	110
Figura 130 -	Teste com 20 locutores: Locutor 14, Gráfico 6	110
Figura 131 -	Teste com 20 locutores: Locutor 15, Gráfico 1	111
Figura 132 -	Teste com 20 locutores: Locutor 15, Gráfico 2	111
Figura 133 -	Teste com 20 locutores: Locutor 15, Gráfico 3	111
Figura 134 -	Teste com 20 locutores: Locutor 15, Gráfico 4	111
Figura 135 -	Teste com 20 locutores: Locutor 15, Gráfico 5	111
Figura 136 -	Teste com 20 locutores: Locutor 16, Gráfico 1	112
Figura 137 -	Teste com 20 locutores: Locutor 16, Gráfico 2	112
Figura 138 -	Teste com 20 locutores: Locutor 16, Gráfico 3	112

Figura 139 -	Teste com 20 locutores: Locutor 17, Gráfico 1	112
Figura 140 -	Teste com 20 locutores: Locutor 17, Gráfico 2	113
Figura 141 -	Teste com 20 locutores: Locutor 18, Gráfico 1	113
Figura 142 -	Teste com 20 locutores: Locutor 18, Gráfico 2	113
Figura 143 -	Teste com 20 locutores: Locutor 18, Gráfico 3	113
Figura 144 -	Teste com 20 locutores: Locutor 18, Gráfico 4	113
Figura 145 -	Teste com 20 locutores: Locutor 18, Gráfico 5	114
Figura 146 -	Teste com 20 locutores: Locutor 18, Gráfico 6	114
Figura 147 -	Teste com 20 locutores: Locutor 18, Gráfico 7	114
Figura 148 -	Teste com 20 locutores: Locutor 19, Gráfico 1	114
Figura 149 -	Teste com 20 locutores: Locutor 19, Gráfico 2	114
Figura 150 -	Teste com 20 locutores: Locutor 19, Gráfico 3	115
Figura 151 -	Teste com 20 locutores: Locutor 20, Gráfico 1	115
Figura 152 -	Teste com 20 locutores: Locutor 20, Gráfico 2	115
Figura 153 -	Teste com 30 locutores: Locutor 1, Gráfico 1	116
Figura 154 -	Teste com 30 locutores: Locutor 1, Gráfico 2	116
Figura 155 -	Teste com 30 locutores: Locutor 1, Gráfico 3	117
Figura 156 -	Teste com 30 locutores: Locutor 1, Gráfico 4	117
Figura 157 -	Teste com 30 locutores: Locutor 1, Gráfico 5	118
Figura 158 -	Teste com 30 locutores: Locutor 1, Gráfico 6	118
Figura 159 -	Teste com 30 locutores: Locutor 2, Gráfico 1	119
Figura 160 -	Teste com 30 locutores: Locutor 2, Gráfico 2	119
Figura 161 -	Teste com 30 locutores: Locutor 2, Gráfico 3	120
Figura 162 -	Teste com 30 locutores: Locutor 3, Gráfico 1	120
Figura 163 -	Teste com 30 locutores: Locutor 3, Gráfico 2	121
Figura 164 -	Teste com 30 locutores: Locutor 3, Gráfico 3	121
Figura 165 -	Teste com 30 locutores: Locutor 3, Gráfico 4	122
Figura 166 -	Teste com 30 locutores: Locutor 4, Gráfico 1	122
Figura 167 -	Teste com 30 locutores: Locutor 4, Gráfico 2	123

Figura 168 -	Teste com 30 locutores: Locutor 4, Gráfico 3	123
Figura 169 -	Teste com 30 locutores: Locutor 4, Gráfico 4	124
Figura 170 -	Teste com 30 locutores: Locutor 4, Gráfico 5	124
Figura 171 -	Teste com 30 locutores: Locutor 5, Gráfico 1	125
Figura 172 -	Teste com 30 locutores: Locutor 5, Gráfico 2	125
Figura 173 -	Teste com 30 locutores: Locutor 5, Gráfico 3	126
Figura 174 -	Teste com 30 locutores: Locutor 5, Gráfico 4	126
Figura 175 -	Teste com 30 locutores: Locutor 5, Gráfico 5	127
Figura 176 -	Teste com 30 locutores: Locutor 6, Gráfico 1	127
Figura 177 -	Teste com 30 locutores: Locutor 6, Gráfico 2	128
Figura 178 -	Teste com 30 locutores: Locutor 6, Gráfico 3	128
Figura 179 -	Teste com 30 locutores: Locutor 7, Gráfico 1	129
Figura 180 -	Teste com 30 locutores: Locutor 7, Gráfico 2	129
Figura 181 -	Teste com 30 locutores: Locutor 7, Gráfico 3	130
Figura 182 -	Teste com 30 locutores: Locutor 7, Gráfico 4	130
Figura 183 -	Teste com 30 locutores: Locutor 7, Gráfico 5	131
Figura 184 -	Teste com 30 locutores: Locutor 7, Gráfico 6	131
Figura 185 -	Teste com 30 locutores: Locutor 7, Gráfico 7	132
Figura 186 -	Teste com 30 locutores: Locutor 7, Gráfico 8	132
Figura 187 -	Teste com 30 locutores: Locutor 8, Gráfico 1	133
Figura 188 -	Teste com 30 locutores: Locutor 8, Gráfico 2	133
Figura 189 -	Teste com 30 locutores: Locutor 8, Gráfico 3	134
Figura 190 -	Teste com 30 locutores: Locutor 8, Gráfico 4	134
Figura 191 -	Teste com 30 locutores: Locutor 8, Gráfico 5	135
Figura 192 -	Teste com 30 locutores: Locutor 9, Gráfico 1	135
Figura 193 -	Teste com 30 locutores: Locutor 9, Gráfico 2	136
Figura 194 -	Teste com 30 locutores: Locutor 9, Gráfico 3	136
Figura 195 -	Teste com 30 locutores: Locutor 9, Gráfico 4	137
Figura 196 -	Teste com 30 locutores: Locutor 9, Gráfico 5	137

Figura 197 -	Teste com 30 locutores: Locutor 9, Gráfico 6	138
Figura 198 -	Teste com 30 locutores: Locutor 9, Gráfico 7	138
Figura 199 -	Teste com 30 locutores: Locutor 9, Gráfico 8	139
Figura 200 -	Teste com 30 locutores: Locutor 10, Gráfico 1	139
Figura 201 -	Teste com 30 locutores: Locutor 10, Gráfico 2	140
Figura 202 -	Teste com 30 locutores: Locutor 10, Gráfico 3	140
Figura 203 -	Teste com 30 locutores: Locutor 10, Gráfico 4	141
Figura 204 -	Teste com 30 locutores: Locutor 10, Gráfico 5	141
Figura 205 -	Teste com 30 locutores: Locutor 10, Gráfico 6	142
Figura 206 -	Teste com 30 locutores: Locutor 10, Gráfico 7	142
Figura 207 -	Teste com 30 locutores: Locutor 11, Gráfico 1	143
Figura 208 -	Teste com 30 locutores: Locutor 12, Gráfico 1	143
Figura 209 -	Teste com 30 locutores: Locutor 12, Gráfico 2	144
Figura 210 -	Teste com 30 locutores: Locutor 12, Gráfico 3	144
Figura 211 -	Teste com 30 locutores: Locutor 12, Gráfico 4	145
Figura 212 -	Teste com 30 locutores: Locutor 12, Gráfico 5	145
Figura 213 -	Teste com 30 locutores: Locutor 13, Gráfico 1	146
Figura 214 -	Teste com 30 locutores: Locutor 13, Gráfico 2	146
Figura 215 -	Teste com 30 locutores: Locutor 14, Gráfico 1	147
Figura 216 -	Teste com 30 locutores: Locutor 14, Gráfico 2	147
Figura 217 -	Teste com 30 locutores: Locutor 14, Gráfico 3	148
Figura 218 -	Teste com 30 locutores: Locutor 14, Gráfico 4	148
Figura 219 -	Teste com 30 locutores: Locutor 14, Gráfico 5	149
Figura 220 -	Teste com 30 locutores: Locutor 15, Gráfico 1	149
Figura 221 -	Teste com 30 locutores: Locutor 15, Gráfico 2	150
Figura 222 -	Teste com 30 locutores: Locutor 15, Gráfico 3	150
Figura 223 -	Teste com 30 locutores: Locutor 15, Gráfico 4	151
Figura 224 -	Teste com 30 locutores: Locutor 16, Gráfico 1	151
Figura 225 -	Teste com 30 locutores: Locutor 16, Gráfico 2	152

Figura 226 -	Teste com 30 locutores: Locutor 16, Gráfico 3	152
Figura 227 -	Teste com 30 locutores: Locutor 17, Gráfico 1	153
Figura 228 -	Teste com 30 locutores: Locutor 17, Gráfico 2	153
Figura 229 -	Teste com 30 locutores: Locutor 17, Gráfico 3	154
Figura 230 -	Teste com 30 locutores: Locutor 17, Gráfico 4	154
Figura 231 -	Teste com 30 locutores: Locutor 18, Gráfico 1	155
Figura 232 -	Teste com 30 locutores: Locutor 18, Gráfico 2	155
Figura 233 -	Teste com 30 locutores: Locutor 18, Gráfico 3	156
Figura 234 -	Teste com 30 locutores: Locutor 19, Gráfico 1	156
Figura 235 -	Teste com 30 locutores: Locutor 19, Gráfico 2	157
Figura 236 -	Teste com 30 locutores: Locutor 19, Gráfico 3	157
Figura 237 -	Teste com 30 locutores: Locutor 21, Gráfico 1	158
Figura 238 -	Teste com 30 locutores: Locutor 21, Gráfico 2	158
Figura 239 -	Teste com 30 locutores: Locutor 21, Gráfico 3	159
Figura 240 -	Teste com 30 locutores: Locutor 22, Gráfico 1	159
Figura 241 -	Teste com 30 locutores: Locutor 22, Gráfico 2	160
Figura 242 -	Teste com 30 locutores: Locutor 23, Gráfico 1	160
Figura 243 -	Teste com 30 locutores: Locutor 23, Gráfico 2	161
Figura 244 -	Teste com 30 locutores: Locutor 23, Gráfico 3	161
Figura 245 -	Teste com 30 locutores: Locutor 23, Gráfico 4	162
Figura 246 -	Teste com 30 locutores: Locutor 23, Gráfico 5	162
Figura 247 -	Teste com 30 locutores: Locutor 23, Gráfico 6	163
Figura 248 -	Teste com 30 locutores: Locutor 24, Gráfico 1	163
Figura 249 -	Teste com 30 locutores: Locutor 24, Gráfico 2	164
Figura 250 -	Teste com 30 locutores: Locutor 24, Gráfico 3	164
Figura 251 -	Teste com 30 locutores: Locutor 24, Gráfico 4	165
Figura 252 -	Teste com 30 locutores: Locutor 24, Gráfico 5	165
Figura 253 -	Teste com 30 locutores: Locutor 25, Gráfico 1	166
Figura 254 -	Teste com 30 locutores: Locutor 25, Gráfico 2	166

Figura 255 -	Teste com 30 locutores: Locutor 25, Gráfico 3	167
Figura 256 -	Teste com 30 locutores: Locutor 25, Gráfico 4	167
Figura 257 -	Teste com 30 locutores: Locutor 25, Gráfico 5	168
Figura 258 -	Teste com 30 locutores: Locutor 26, Gráfico 1	168
Figura 259 -	Teste com 30 locutores: Locutor 26, Gráfico 2	169
Figura 260 -	Teste com 30 locutores: Locutor 26, Gráfico 3	169
Figura 261 -	Teste com 30 locutores: Locutor 26, Gráfico 4	170
Figura 262 -	Teste com 30 locutores: Locutor 26, Gráfico 5	170
Figura 263 -	Teste com 30 locutores: Locutor 27, Gráfico 1	171
Figura 264 -	Teste com 30 locutores: Locutor 27, Gráfico 2	171
Figura 265 -	Teste com 30 locutores: Locutor 27, Gráfico 3	172
Figura 266 -	Teste com 30 locutores: Locutor 28, Gráfico 1	172
Figura 267 -	Teste com 30 locutores: Locutor 28, Gráfico 2	173
Figura 268 -	Teste com 30 locutores: Locutor 28, Gráfico 3	173
Figura 269 -	Teste com 30 locutores: Locutor 29, Gráfico 1	174
Figura 270 -	Teste com 30 locutores: Locutor 29, Gráfico 2	174
Figura 271 -	Teste com 30 locutores: Locutor 29, Gráfico 3	175
Figura 272 -	Teste com 30 locutores: Locutor 29, Gráfico 4	175
Figura 273 -	Teste com 30 locutores: Locutor 29, Gráfico 5	176
Figura 274 -	Teste com 30 locutores: Locutor 29, Gráfico 6	176
Figura 275 -	Teste com 30 locutores: Locutor 29, Gráfico 7	177
Figura 276 -	Teste com 30 locutores: Locutor 30, Gráfico 1	177
Figura 277 -	Teste com 30 locutores: Locutor 30, Gráfico 2	178

Lista de Tabelas

Tabela 2.1 - Bandas críticas da escala <i>Bark</i> e seus intervalos de frequências	50
Tabela 4.1 - 10 locutores: acertos e erros obtidos nos testes	64
Tabela 4.2 - 20 locutores: acertos e erros obtidos nos testes	66
Tabela 4.3 - 30 locutores: acertos e erros obtidos nos testes	70
Tabela 4.4 - Acurácia global para todos os testes realizados	74

Lista de Abreviaturas

ABFF	Academia Brasileira de Fonoaudiologia Forense
CHMM	<i>Continuous Hidden Markov Models</i>
DTW	<i>Dynamic Time Warping</i>
DWT	<i>Discrete Wavelet Transform</i>
EER	<i>Equal Error Rate</i>
FIR	<i>Finite Impulse Response</i>
GMM-UBM	<i>Gaussian Mixture Models-Universal Background Models</i>
HMM	<i>Hidden Markov Model</i>
IEEE	<i>Institute of Electrical and Eletronic Engineers</i>
IIR	<i>Infinite Impulse Response</i>
LPCC	<i>Linear Predictive Cepstrum Coefficients</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
MLP	<i>Multiple Layer Perceptron</i>
NIST	<i>National Institute of Standards and Technology</i>
PCM	<i>Pulse Code Modulation</i>
RNA	Rede Neural Artificial
SVM	<i>Support Vector Machine</i>
TIMIT	<i>Texas Instruments and Massachusetts</i>
TTVNN	<i>Transição entre Trechos Vozeados e Não-Vozeados</i>
VQ	<i>Vector Quantization</i>
WAVE	<i>Waveform</i>
WPT	<i>Wavelet Packet Transform</i>
ZCR	<i>Zero Crossing Rate</i>

Sumário

1	Introdução	35
1.1	Considerações iniciais	35
1.2	Objetivos	37
1.3	Organização do trabalho	38
2	Revisão Bibliográfica	39
2.1	Reconhecimento de locutores	39
2.2	O sistema de produção da fala humana	42
2.3	Trabalhos recentes envolvendo a rea de reconhecimento de locutores	42
2.4	Conceitos utilizados no trabalho proposto	48
2.4.1	O formato de arquivos WAVE	48
2.4.2	Escala <i>Bark</i>	48
2.4.3	Filtros digitais	49
2.4.4	Convolução	53
2.4.5	Energia	54
2.4.6	Valores Delta e Delta-delta	55
2.4.7	Estatísticas de primeira ordem e modelamento gaussiano	55
3	Detalhamento do Trabalho Proposto	57
3.1	Considerações iniciais	57
3.2	Extração de características	58
3.3	Reconhecimento de locutores	60
4	Testes e Resultados	63

4.1	Considerações iniciais	63
4.2	Testes utilizando $L = 10$ locutores	63
4.3	Testes utilizando 20 locutores	66
4.4	Testes utilizando 30 locutores	70
4.5	Testes com mais de 30 locutores e análise global	72
5	Conclusões e Trabalhos Futuros	75
	Referências	76
	Apêndice I - Gráficos das características extraídas	83

Capítulo 1

Introdução

1.1 Considerações iniciais

Petry (1) define o reconhecimento de locutores como a ação de identificar uma pessoa analisando um trecho de uma locução sua. De acordo com Campbell, J. P. et al., (2) tem havido um grande interesse por parte da comunidade científica em identificar indivíduos com base na voz, procedimento conhecido como *speaker identification* ou identificação de locutor. Ao longo dos anos, juízes de direito, investigadores de polícia, detetives, entre outros profissionais da área forense, têm desejado utilizar a identificação biométrica por voz para tentar solucionar latrocínios, sequestros ou outros eventos mais sérios, conforme se pode constatar também pelo crescente interesse da Academia Brasileira de Fonoaudiologia Forense (ABFF), (3) que ultimamente tem organizado eventos específicos nessa área, assim como pelas renomadas conferências e periódicos do *Institute of Electrical and Electronic Engineers* (IEEE), da Elsevier e da *World Scientific*, especialmente dedicados a esse propósito. Registre-se ainda a crescente tendência demonstrada pelo setor industrial em construir aplicações baseadas na identificação de locutor.

São inúmeros os casos em que se dispõe apenas da voz como objeto de análise para realizar a identificação de um indivíduo como, por exemplo, em uma gravação telefônica, em um evento de segurança em que a face de determinado suspeito aparece encoberta ou em situações nas

quais as impressões digitais não estão disponíveis. Além da possibilidade de uso da voz como instrumento de análise para fins de perícia, outras aplicações que envolvem segurança física e lógica, tais como o acesso a ambientes restritos, autenticação em aplicações como *phone-banking*, automação residencial, entre outras, podem se beneficiar dessa técnica, como relatam recentes teses, (4–6), livros (7–9) e artigos (10–12) da área.

A identificação biométrica por voz realizada por intermédio computacional é, muitas vezes, uma tarefa relativamente difícil e desafiadora. Entretanto, desde a década de 60, quando o termo *voiceprint identification* foi introduzido em uma edição da conceituada revista *Nature*, (13) o público em geral passou a ter a falsa impressão de que essa tarefa é bastante simples e corriqueira. Além do fato de que as amostras dos sinais de voz podem ser obtidas em condições desfavoráveis, tais como a degradação por ruído intenso, amplitude insuficiente para posterior amplificação, locutor sob o efeito de drogas, álcool, estresse, pânico, pressão psicológica, entre outros fatores, a enorme variabilidade dos sinais faz com que comparações de caráter modesto ou mesmo mais elaborado sejam insuficientes para realizar a identificação. Todos esses fatores têm feito com que, ao longo dos anos, a viabilidade de tal tarefa seja até mesmo questionada por alguns pesquisadores. (14–16)

A partir de 2003, algumas instituições ao redor do mundo, tais como o *National Institute of Standards and Technology* (NIST), (17) têm promovido campanhas e alertado a comunidade científica internacional sobre a necessidade de cuidado com os procedimentos utilizados para reconhecimento de locutores, tendo em vista o fato de que, até o momento, não existe processo computacional cientificamente comprovado que possa, com absoluta certeza, identificar um indivíduo por meio de sua voz. Esse fato gerou uma grande motivação para o prosseguimento das pesquisas envolvendo identificação biométrica por voz nos últimos anos, razão pela qual é possível notar um considerável avanço nessa área, em âmbito internacional. Dentre esses avanços, podem-se mencionar as técnicas que objetivam tratar com mais precisão a variabilidade dos sinais de voz, como as que envolvem estimação de máxima verossimilhança (18) e máquinas de vetores de suporte (*support vector machines* - SVM), (19) utilizadas isoladamente ou em conjunto com os Modelos Ocultos de Markov (*Hidden Markov Models* - HMMs) e os Modelos de Misturas Gaussianas (*Gaussian Mixture Models* - *Universal Background Models* -

GMM-UBM) (8), que podem ser considerados como o estado-da-arte atualmente.

Independentemente da técnica de reconhecimento utilizada, os sinais de voz digitalizados não podem ser diretamente entregues ao classificador devido aos seus tamanhos, isto é, devido às dimensões ou número de amostras serem variáveis e consideravelmente altos. Surgiria, assim, impedimento quanto ao uso da maioria dos classificadores, pois eles trabalham com dimensões de entrada pré-definidas, e também quanto ao custo computacional inviável. Desse modo, é sempre necessário um extrator de características, ou seja, um estágio intermediário entre o sinal de voz digitalizado e o elemento classificador, visando reduzir e fixar a dimensão daquele para possibilitar a avaliação por meio deste. As características extraídas agrupam-se em um vetor designado por vetor de características.

Embora, nos sistemas de classificação em geral, as características extraídas de determinado sinal devam ser avaliadas em conjunto com o elemento classificador, sabe-se que quanto maior o contraste entre os locutores expresso puramente por aquelas, menor o trabalho deste. Desse modo, a proposta deste trabalho está focada justamente nos vetores de características extraídos de sinais de voz, sem maiores preocupações com o elemento classificador, conforme detalhado adiante.

1.2 Objetivos

Motivado pelas considerações anteriores, o objetivo do presente trabalho é mensurar a acurácia de um sistema de reconhecimento de locutores baseado em características elementares, isto é, de energias de sub-bandas de frequências, em associação com um classificador probabilístico, estudando a viabilidade de extraí-las das transições entre trechos vozeados e não-vozeados (TTVNV) dos sinais. Testes são realizados com diferentes quantidades de locutores e discurso fixado. Os resultados obtidos são comparados e relatados, de forma a complementar, com dados quantitativos, as afirmações existentes na literatura sobre o uso das TTVNV para reconhecimento de locutores.

1.3 Organização do trabalho

O texto vindouro do presente trabalho está organizado da seguinte forma:

- No Capítulo 2 apresenta-se uma série de trabalhos publicados envolvendo a área de reconhecimento de locutores, mostrando como são inúmeras as possibilidades de se realizar essa tarefa. Expõe-se, também, os principais conceitos e teorias que estão relacionados com o trabalho que foi desenvolvido.
- No Capítulo 3 apresenta-se, com detalhes, todo o desenvolvimento do trabalho proposto e de que forma os conceitos discutidos no capítulo anterior foram utilizados.
- No Capítulo 4 relatam-se todos os resultados obtidos no trabalho, a partir dos testes de reconhecimento que foram realizados.
- No Capítulo 5 apresentam-se as conclusões sobre o trabalho, bem como propostas para pesquisas futuras.

Capítulo 2

Revisão Bibliográfica

2.1 Reconhecimento de locutores

Reconhecimento de locutores (20) constitui um termo geral que pode englobar diversas classes de problemas específicos, sendo possível destacar:

- **Verificação (ou autenticação) de locutores:** nessa classe de problemas, a verificação da pessoa, ou locutor, não é feita diretamente com a manipulação da fala. Primeiramente, um usuário ou um número de identificação é utilizado para recuperar o modelo representativo (*template*) do locutor dentro de uma base de dados. A partir desse modelo, é feita a sua comparação com uma amostra da fala do locutor, em que é manifestada a sua autenticidade, ou não, pelo sistema. Em uma verificação um pouco mais robusta, também pode haver a comparação da fala com um modelo geral representativo de toda a base de dados. Assim, caso a amostra seja mais parecida com o modelo geral, ela será rejeitada (não-identificada). Caso seja mais parecida com o modelo do locutor em questão, ela será aceita (identificada) pelo sistema.
- **Identificação de locutores:** classe de problemas que pode ser dividida entre *closed-set* e *open-set*. Na abordagem *closed-set*, há a comparação da voz do locutor com todos os modelos presentes na base dados, diferente da verificação de locutores, na qual a comparação

é feita apenas com o modelo do locutor a ser identificado e o modelo geral da base de dados. Não existe rejeição, ou seja, o locutor sob avaliação sempre será identificado de acordo com o modelo mais similar a ele. Já na abordagem *open-set*, pode haver a rejeição da identificação, caso o modelo mais similar não seja considerado representante da amostra do locutor, segundo algum critério imposto. Em outras palavras, podemos dizer que é feita uma identificação *closed-set* e, logo após, uma verificação de locutores.

- **Classificação de locutores:** pode ser feita de várias formas como, por exemplo, classificação de gênero dos locutores (masculino e feminino), de acordo com faixas de idade (jovem, adulto e idoso), e também de vozes saudáveis ou que apresentem algum tipo de patologia. Tal classificação é feita aplicando-se as idéias discutidas previamente de verificação e identificação de locutores.

Adicionalmente, o reconhecimento de locutores pode ser dividido em duas modalidades (20), que são:

- ***Text-dependent*:** é uma forma de reconhecimento na qual utiliza-se apenas uma frase para todos os locutores existentes. Este é o caso da abordagem utilizada no presente trabalho.
- ***Text-independent*:** ao contrário da modalidade *text-dependent*, aqui podem ser utilizadas frases distintas para realizar o reconhecimento, tornando o processo bem mais complexo. Alguns sistemas *text-independent* também podem ser classificados como *language-independent*, em que qualquer frase em qualquer idioma é aceito para o reconhecimento.

As características específicas do trato vocal de cada pessoa fazem com que as vozes possuam suas próprias especificidades, o que permite manipulá-las e extraí-las para realizar o reconhecimento. Essa seria a ideia base para se trabalhar com qualquer classe de problemas dentro da área de reconhecimento de locutores.

Dentre as características mais utilizadas, pode-se destacar a manipulação de sons vozeados e não-vozeados da fala. Sons vozeados são aqueles que, ao serem produzidos, fazem as cordas vocais vibrarem quase periodicamente com a passagem do ar. Logo, sons não-vozeados são aqueles que não fazem as cordas vocais vibrarem periodicamente com a passagem de ar, possuindo, como origem, fluxos turbulentos de ar. As vogais /a/, /e/, /i/, /o/ e /u/ são exemplos de

sons vozeados, enquanto que as consoantes /p/, /t/, /k/, /f/ e /s/ constituem exemplos de sons não-vozeados. Um ponto-chave explorado neste trabalho é que, de acordo com Liñares; Mateo (21) e Weddin; Winther, (22) considerável parte da informação que reflete a identidade de um locutor reside nas regiões transitórias entre trechos vozeados e não-vozeados.

A análise dos formantes também é outra forma muito utilizada de manipulação de características específicas da fala. Os formantes são regiões ressonantes presentes no espectrograma da fala e dependem da forma do trato vocal, que varia muito de acordo com idade e sexo, principalmente, tornando seu uso interessante para ser aplicado nos problemas de reconhecimento. A figura 2.1 contém um exemplo dos formantes obtidos a partir da locução de uma palavra. Já a figura 2.2 ilustra como os formantes obtidos de dois locutores distintos, pronunciando uma mesma frase, podem ser bem diferenciados.

Figura 2.1 – Exemplo de formantes

Fonte: Adaptada de BEIGI, H. (20)

Figura 2.2 – Formantes de locutores distintos obtidos a partir de uma mesma locução

Fonte: Adaptada de BEIGI, H. (20)

Conforme discutido anteriormente, as aplicações são as mais variadas possíveis para o reconhecimento de locutores. Podem-se destacar, entre as mais comuns, principalmente as que envolvem o controle de acesso à informação, as aplicações financeiras e as de auxílio aos deficientes.

2.2 O sistema de produção da fala humana

A estrutura da produção de fala do ser humano (20) é muito complexa e depende de vários órgãos superiores. Narina, boca, faringe e laringe são os órgãos que desempenham o papel mais importante na sua produção, além dos pulmões, responsáveis pelo deslocamento de ar e pela pressão necessária para a geração da fala. Basicamente, o ar vem dos pulmões, passa pela traqueia e pelo esôfago e chega às cordas vocais, fazendo-as vibrar. As cordas vocais, uma vez estando bem fechadas, obstruem consideravelmente a passagem do ar. Por outro lado, podem estar totalmente relaxadas, não oferecendo nenhum tipo de bloqueio para sua passagem. Após as cordas vocais, chega-se à laringe e à faringe, alcançando o trato vocal, onde há todo o controle articulatorio do ser humano para a fala. Com o auxílio da cavidade nasal, narina, língua, dentes e lábios, é gerada a fala de cada pessoa.

A figura 2.3 contém uma ilustração da estrutura do ser humano que é responsável pela produção da fala. Toda essa estrutura, por ser diferenciada em cada pessoa, faz com que diferentes tons de vozes sejam produzidos, o que permite utilizar características específicas da voz para realizar o reconhecimento de locutores.

2.3 Trabalhos recentes envolvendo a rea de reconhecimento de locutores

Diversos trabalhos têm sido desenvolvidos no contexto de reconhecimento de locutores, a fim de observar e relatar a acurácia das diversas técnicas de reconhecimento existentes, combi-

Figura 2.3 – Sistema de produção da fala humana

Fonte: Adaptada de FURUI, S. (23)

nadas com as diversas outras técnicas utilizadas para a extração de características. Tais trabalhos, como pode ser observado a seguir, envolvem a utilização das mais variadas técnicas nos mais diversos cenários de estudo.

Analisando trabalhos relacionados de forma cronológica, pode-se observar que Lung (24) aborda a modalidade *text-independent* de reconhecimento de locutores. A extração de características é feita utilizando a Transformada *Wavelet*, conceito muito empregado para tal tarefa. Uma melhoria é proposta no referido trabalho, somando à Transformada *Wavelet* a ideia de *Kernel Canonical Correlation Analysis*. GMMs, conceito já muito difundido na área, é a técnica utilizada para os testes de identificação. Aqui, duas bases de vozes são utilizadas: *TALUNG*, que contém vozes em Mandarim, com 100 locutores, e *KING*, que possui vozes em Inglês, com 51 locutores. Testes são realizados utilizando 16, 24 e 32 características. Os resultados mostram acurácias na faixa de 90% a 98%.

Wu; Tsai (25) aborda a identificação de locutores para a qual a extração de características é feita utilizando a técnica *Empirical Mode Decomposition* e os testes de identificação são feitos utilizando-se dois modelos distintos de Redes Neurais Artificiais (RNAs), outro conceito largamente utilizado na área de reconhecimento de locutores. A base de dados utilizada contém 36 locutores, na qual foram coletadas amostras a partir de cinco sentenças em Chinês. Os testes

foram realizados individualmente com cada uma das frases, ou seja, foi escolhida a modalidade *text-dependent*. A acurácia média variou de 72,6% a 88,82%.

O trabalho proposto por Rao et al., (26) assim como os anteriores, foca a identificação de locutores, mas neste caso, para uma aplicação específica em dispositivos móveis. Para tanto, a modalidade *text-independent* é utilizada, contaminando os sinais com diversos tipos de ruídos, com o objetivo de melhorar o reconhecimento sob tais adversidades. Três bases de dados são utilizadas: uma é constituída de 50 locutores oriundos da base *Texas Instruments and Massachusetts (TIMIT)*, outra constituída de 50 locutores com amostras coletadas usando um microfone comum e, por fim, a última base constituída de amostras de conversas ao telefone celular. Cinco tipos de ambientes ruidosos foram utilizados nos testes, em combinação com cinco *signal-to-noise ratios*, além do ambiente limpo de ruídos, totalizando 26 diferentes testes para cada base de vozes. As 39 características utilizadas são extraídas com base nos conceitos de *Mel Frequency Cepstral Coefficients* (MFCC), deltas e delta-deltas. Redes neurais auto associativas são utilizadas para os testes de reconhecimento. A acurácia variou de 94% a 98% no ambiente sem ruídos. Já considerando todos as condições de ruídos utilizadas, a acurácia variou muito, chegando ao máximo de 79% entre todos os testes realizados.

No trabalho de Sadic; Gulmezoglu (27) é proposta uma técnica de reconhecimento que combina o conceito de *Common Vector Approach* com o já conhecido GMM, na modalidade *text-independent*, utilizando 20 locutores da base de áudios *TIMIT*. As características extraídas e utilizadas são compostas de 20 MFCCs, refletindo em uma acurácia média de identificação de 93% a 96,6% nos mais variados testes efetuados no trabalho.

O trabalho relatado por Kumari; Nidhyananthan; Anand (28) aborda a identificação de locutores na modalidade *text-independent*. Aqui, mais uma vez é utilizado o conceito de MFCC para a extração das características dos sinais e o GMM para os testes de reconhecimento. Foram utilizados 120 locutores da base de áudios *TIMIT* e a acurácia dos testes variou de 79,59% a 93,88%.

Assim como Rao et al., (26) aqui destaca-se outro trabalho mais específico. Andrei (29) foca seu trabalho na identificação de locutores em tempo real na modalidade *text-dependent*. Mais uma vez são utilizados os coeficientes MFCC e seus respectivos deltas e delta-deltas como

características extraídas. O conceito de *Dynamic Time Warping* (DTW) foi aplicado para os testes de identificação. A base de vozes foi capturada com um microfone comum e consiste de 23 locutores. Os testes foram realizados com diferentes quantidades de coeficientes extraídos, resultando em acurácias entre 80,51% e 96,42%.

Estudos que abordam como estados emocionais afetam o reconhecimento de locutores também são muito desenvolvidos, como é o caso de Ghiurcau; Rusu. (30) Aqui, é feito o estudo da identificação de locutores na modalidade *text-independent*. Novamente, MFCCs e seus deltas foram utilizados como características extraídas e GMM para os testes de reconhecimento. Foi utilizada a *Berlin Database of Emotional Speech*, que contém 10 locutores com falas captadas em Alemão sob 6 diferentes estados emocionais. Diversos testes foram efetuados, variando tanto a quantidade de características extraídas, quanto a base de dados de teste e treinamento, fazendo com que os resultados obtidos tivessem uma grande variação de acurácia, de cerca de 16% até uma porcentagem máxima de 98,57%.

Apesar do Inglês ser a língua utilizada para a maior parte dos estudos na área de reconhecimento de locutores, diversas outras podem ser adotadas, como foi o caso de Ghiurcau; Rusu (30), em que foram utilizados áudios em Alemão. Aqui, apresenta-se também um outro trabalho, agora envolvendo a língua Árabe. (31) A identificação de locutores proposta no trabalho, na modalidade *text-independent*, é feita por meio da junção das ideias contidas nas técnicas *Continuous Hidden Markov Models* (CHMMs) e GMM. A extração das características tem como base os MFCCs, seus deltas e delta-deltas. Um total de 10 locutores foi utilizado nos vários testes realizados, obtendo acurácias que variam de 84,33% a 91,33%.

Em Ajmera; Jadhav; Holambe (32) é apresentada uma nova técnica para a extração de características, combinando a Transformada *Radon* e a Transformada Coseno Discreta, com o objetivo de buscar padrões em imagens de espectograma. Foi testada, na identificação de locutores, a modalidade *text-independent* com a utilização da base de vozes *TIMIT* e técnicas de aprendizado de máquina, atingindo até 96,69% de acurácia.

Zhu; Yang (33) propõe uma nova forma de extração de características para a identificação de locutores na modalidade *text-independent*, baseada em *Linear Predictive Cepstrum Coefficients* (LPCC). A única diferença, se comparado com a utilização usual de LPCC, é que aqui são

atribuídos pesos às 16 características que são extraídas. A técnica de *Vector Quantization* (VQ) é utilizada para os testes de reconhecimento. A base de vozes utilizada foi coletada pelos próprios pesquisadores e é formada por 20 locutores. Foram obtidos 89,33% de acurácia adotando-se pesos iguais para os parâmetros, 91% aplicando-se pesos aos coeficientes em forma sequencial e 94,67% aplicando-se pesos aos coeficientes de forma não sequencial.

Um trabalho recente envolvendo a verificação de locutores é apresentado por Li et al. (34). Aqui, as características extraídas também envolvem o uso de MFCCs e a ideia de contorno do *pitch* com o uso de polinômios cúbicos. Para os testes de reconhecimento, foi utilizada a técnica de SVM. O desempenho é medido por meio do valor *Equal Error Rate* (EER), que atinge 2,7% quando apenas o MFCC é utilizado, 46,9% quando apenas os contornos de *pitch* são utilizados e 1,4% quando ambos são utilizados em conjunto.

No estudo de Shahin, (35) trabalha-se com a identificação de locutores na modalidade *text-independent* considerando diferentes estados emocionais dos locutores. As características utilizadas se baseiam em MFCC e seus deltas. Para os testes de reconhecimento, é utilizado o conceito HMM. A base de áudios utilizada é composta de 40 locutores e seis diferentes emoções. Os resultados obtidos variam de 74% a 95,5% de acurácia.

Abordando o tema de uma forma um pouco diferente, Jawakar; Holambe; Basu (36) trabalham com a voz na forma de sussurros para a realização da identificação de locutores na modalidade *text-independent*. Diferente da maioria dos trabalhos aqui apresentados, vozes de 25 locutores Indianos foram utilizados na pesquisa. Diferentes características foram extraídas: MFCC, *Weighted Instantaneous Frequency*, *Temporal Energy of Subband Cepstral Coefficients* e *Temporal Teager Energy Based Subband Cepstral Coefficients*. A eficiência da representação de cada um desses conjuntos de características foi comparada utilizando-se a técnica GMM para o reconhecimento. Por fim, testes também foram realizados utilizando todas as características extraídas, em conjunto. Considerando todos os testes que foram realizados, as acurácias variaram muito, chegando ao mínimo de 28,96% e ao máximo de 98,81%.

Na pesquisa relatada por Gupta; Gupta (37) são utilizados os conceitos de MFCC e VQ para a extração de características, sendo estas utilizadas para a identificação de locutores na modalidade *text-independent*. Para os testes de reconhecimento foi utilizado o conceito simples

de Distância Euclidiana. As acurácias variaram de 71,1% a 88,9% com diferentes quantidades de características utilizadas.

Neste penúltimo trabalho apresentado, Jayakumar; Antop (38) realizam a identificação de locutores na modalidade *text-independent*. A base de vozes utilizada é própria, coletada de diferentes fontes de captação e é formada por 48 locutores. Duas técnicas são utilizadas para a extração de características e, posteriormente, comparadas: *Wavelet-packet Transform* (WPT) e *Discrete Wavelet Transform* (DWT). Para os testes de reconhecimento utilizou-se uma RNA do tipo *Multiple Layer Perceptron* (MLP). As médias das acurácias foram de 70,5% e 69,34% para WPT e DWT, respectivamente.

O último trabalho aqui exposto, Vikram; Umarani (39) estuda a classificação de locutores entre vozes saudáveis e patológicas na modalidade *text-independent*. Para a extração de características, foram utilizados MFCCs e para os testes de reconhecimento foi utilizado o GMM. A base de vozes foi formada por 142 vozes saudáveis e 147 vozes afetadas por alguma patologia. As acurácias variam de 61,84% a 85,63%, conforme a quantidade de coeficientes extraídos era modificada.

Pelos diversos trabalhos relatados, pode-se perceber que algumas técnicas são largamente utilizadas, seja para a extração de características (MFCC) como para o reconhecimento de locutores (GMM, RNA). Percebe-se, também, como a classe de problemas de identificação de locutores é bastante discutida quando comparada com as outras classes descritas na Seção 2.1.

Apesar dos valores energéticos e suas variantes (deltas e delta-deltas) serem muito utilizados, dificilmente eles são extraídos de pontos específicos dos sinais de voz, como é o caso do trabalho aqui proposto, em que se foca em trechos de transições vozeadas/não-vozeadas. O que comumente ocorre é considerar o sinal como um todo, ou dividido em janelas de iguais larguras, para realizar a extração das características.

2.4 Conceitos utilizados no trabalho proposto

2.4.1 O formato de arquivos WAVE

O formato WAVE (40) é uma abreviação da palavra *waveform*. É um formato de arquivos de áudio desenvolvido pelas empresas *Microsoft* e *IBM*, sendo baseado no *Pulse Code Modulation* (PCM), uma técnica de conversão de sons analógicos para sons digitais. Também é utilizado em formatos de gravação de áudios digitais e na criação de discos de áudio. PCM é um formato descompactado, ou seja, resulta em uma maior qualidade dos arquivos e informações mais detalhadas, mesmo que eles possam ocupar um pouco mais de espaço em seu armazenamento. PCM faz com que o formato WAVE seja perfeito para aplicações de áudio profissionais. No presente trabalho, todos os sinais de voz analisados foram gravados no formato WAVE, com taxa de amostragem igual a 16.000 amostras por segundo e quantização de 16 bits por amostra.

2.4.2 Escala *Bark*

A escala *Bark* (41, 42) foi obtida empiricamente por meio de estudos na área de psicoacústica, e leva em consideração faixas de frequências que são perceptíveis pelo sistema auditivo humano. A essas faixas de frequência, dá-se o nome de bandas críticas, as quais existem por levar-se em consideração as frequências relevantes ao sistema auditivo humano. A escala *Bark* recebe esse nome em homenagem a *Barkhausen*, antigo pesquisador da área.

A figura 2.4 contém um exemplo do comportamento da escala de frequências, em Hz, com a escala convertida, em *Barks*. Abaixo de 400 Hz, a relação é linear entre as frequências. Acima disso, observa-se uma relação logarítmica entre elas.

Figura 2.4 – Relação entre as frequências em Hz e na escala *Bark*

Fonte: Adaptada de HAVELOCK, D.; KUWANO, S.; VORLANDER, M. (41)

Adicionalmente, a tabela 2.1 contém a lista das frequências, em Hz, que delimitam cada uma das 24 bandas críticas da escala *Bark*.

No presente trabalho, concentrações energéticas normalizadas de cada banda *Bark* são utilizadas como características representativas de cada sinal de voz.

2.4.3 Filtros digitais

Filtros digitais (43) são construídos para a manipulação de sinais discretos, diferente de filtros analógicos, que trabalham com sinais contínuos (ou analógicos). Eles são largamente utilizados em aplicações que necessitam que muitos sinais sejam filtrados a partir de uma mesma função.

Existem duas classes principais de filtros digitais (43–45): *Finite Impulse Response* (FIR), utilizada neste trabalho, e *Infinite Impulse Response* (IIR). Filtros FIR são muito aplicados, pois:

Tabela 2.1 – Bandas críticas da escala *Bark* e seus intervalos de frequências

Banda crítica (<i>Bark</i>)	Início (Hz)	Fim (Hz)
1	0	100
2	100	200
3	200	300
4	300	400
5	400	510
6	510	630
7	630	770
8	770	920
9	920	1080
10	1080	1270
11	1270	1480
12	1480	1720
13	1720	2000
14	2000	2320
15	2320	2700
16	2700	3150
17	3150	3700
18	3700	4400
19	4400	5300
20	5300	6400
21	6400	7700
22	7700	9500
23	9500	12000
24	12000	15500

Fonte: Adaptado de BOSI, M.; GOLDBERG, R.E. (42)

- São capazes de manipular um sinal sem adicionar qualquer distorção indesejada no mesmo, tendo em vista que podem ser construídos para possuir resposta em fase linear;
- A fase linear causa um retardo constante em cada amostra do sinal digital filtrado, gerando um alinhamento de tempo bem preciso, característica essencial para aplicações de processamento de fala;
- A fase linear também facilita o processo de aproximação das respostas dos filtros;
- São flexíveis, estáveis e fáceis de serem implementados.

Para projetar um filtro FIR, é necessário definir, basicamente, a sua ordem M e, consequentemente, a quantidade de coeficientes que o compõe, igual a $M + 1$. Além dos filtros utilizados neste trabalho serem digitais e do tipo FIR, eles também são classificados como filtros de banda de passagem (43), que podem ser divididos nos seguintes tipos:

- **Filtro passa-baixas:** a partir de uma frequência-limite, todas as componentes abaixo dela passam pelo filtro, sendo as frequências maiores atenuadas. Filtros passa-baixas também servem de base para a construção de outros filtros;
- **Filtro passa-altas:** a partir de uma frequência-limite, todas as componentes acima dela passam pelo filtro, sendo as frequências menores atenuadas;
- **Filtro passa-faixa:** a partir de duas frequências-limite, todas as componentes que se encontram dentro do intervalo definido entre elas passam pelo filtro, sendo o restante da banda atenuada;
- **Filtro rejeita-faixa:** a partir de duas frequências-limite, todas as componentes que não se encontram dentro do intervalo definido entre elas passam pelo filtro, sendo o restante atenuada.

Na figura 2.5 há uma ilustração da ideia básica e a forma ideal dos filtros citados acima.

Figura 2.5 – Filtros de banda de passagem

Fonte: Adaptada de SANTOS, J.C.V. (43)

O cálculo dos coeficientes, $h[\cdot]$, do filtro passa-baixas, por exemplo, é feito de acordo com a Equação 2.1, onde:

- k é número de coeficientes do filtro;
- M é a ordem do filtro, sempre igual a $(k - 1)$;
- α é a frequência de corte do filtro, obtida multiplicando-se a frequência em questão, em Hz, por $\frac{\pi}{tx_am}$, onde tx_am é a taxa de amostragem do sinal.

$$h_i = \frac{\sin(\alpha \cdot (i - \frac{M}{2}))}{\pi \cdot (i - \frac{M}{2})}, \quad \text{para } i = 1, \dots, k \quad (2.1)$$

No presente trabalho, filtros digitais passa-faixa são utilizados para caracterizar as bandas *Bark* de interesse para cada sinal de voz. Tais filtros podem ser obtidos, sendo este o caso deste

trabalho, por meio da subtração ponto-a-ponto de dois filtros passa baixas: o primeiro com a frequência de corte no final da faixa de interesse e o segundo com frequência de corte no início da faixa de interesse.

2.4.4 Convolução

A operação de convolução (44, 45) é definida como a superposição ponderada de impulsos deslocados no tempo. A convolução possui duas denominações principais:

- **Soma de convolução:** quando manipulam-se sinais discretizados, como é o caso deste trabalho;
- **Integral de convolução:** quando trabalham-se sinais contínuos, isto é, analógicos.

A soma de convolução de dois sinais digitais, $x[m]$ e $y[n]$, é simbolizada pela Equação 2.2. A operação sempre aumenta o tamanho do sinal resultante para $m + n - 1$.

$$C[\cdot] = x[m] * y[n] \quad (2.2)$$

Na prática, pode-se dizer que a convolução é similar ao produto de dois sinais distintos, uma vez que a ideia é semelhante à operação de multiplicação de dois números. A figura 2.6 contém uma ilustração de como são feitos os cálculos da soma de convolução. Primeiro, é efetuado o produto, coeficiente a coeficiente, de ambos os sinais, sendo, depois, feita uma soma superposta dos resultados obtidos, gerando, por fim, o sinal de saída.

No presente trabalho, a soma de convolução é utilizada para filtrar os sinais de voz, sendo equivalente à multiplicação da resposta em frequência do sinal de voz pela resposta em frequência do filtro digital, de acordo com o Teorema da Convolução (44).

Figura 2.6 – Exemplo numérico da operação de soma de convolução

				37	84	8	53	12	←	Sinal 1
					3	5	31	14	←	Sinal 2
				518	1176	112	742	168		
		1147	2604	248	1643	372				
	185	420	40	265	60				+	
111	252	24	159	36						
111	437	1591	3321	1725	1815	1114	168		←	Sinal Resultante

Fonte: Elaborada pelo autor

2.4.5 Energia

A energia de um sinal $x(t)$ (45) é uma das formas existentes de se obter uma medida do potencial de ação do sinal, ou do esforço necessário para realizar determinada tarefa. É definida como a área formada por $x^2(t)$ ao longo de sua duração. Então, para um sinal contínuo, a energia E é dada pela Equação 2.3. Já para sinais discretizados de tamanho n , que constituem o caso do presente trabalho, a energia pode ser calculada de acordo com a Equação 2.4.

$$E = \int_{-\infty}^{\infty} f^2(x) dx. \quad (2.3)$$

$$E = \sum_{k=1}^n x_k^2 \quad (2.4)$$

O valor de energia é uma medida do sinal simples de ser obtida e, conseqüentemente, de ser manipulada. No presente trabalho, a medida de energia de determinados trechos dos sinais de vozes é utilizada como característica descritiva dos mesmos.

2.4.6 Valores Delta e Delta-delta

A partir de um conjunto de valores, no caso deste trabalho, de um vetor de características extraídas das vozes dos locutores, pode-se obter as suas derivadas de primeira e de segunda ordens, usualmente denominadas de deltas e delta-deltas, respectivamente. Tais valores são também conhecidos como *dynamic features*, isto é, características dinâmicas, por serem capazes de representar as características temporais e dinâmicas da fala (46, 47).

De forma simples, pode-se obter os valores delta a partir da razão de dois valores adjacentes que compõem um vetor de características x , como pode-se observar na equação 2.5. De forma análoga, os valores de delta-delta são obtidos a partir da razão de dois valores delta adjacentes, como mostra a Equação 2.6.

$$\Delta_i = \frac{x_i}{x_{i+1}} \quad (2.5)$$

$$\Delta\Delta_j = \frac{\Delta_j}{\Delta_{j+1}} \quad (2.6)$$

São utilizados no presente trabalho, os deltas e delta-deltas das diversas energias extraídas dos sinais de vozes.

2.4.7 Estatísticas de primeira ordem e modelamento gaussiano

A média, μ , e a variância, σ^2 , de um conjunto $x[\cdot]$ de N valores são calculadas de acordo com as Equações 2.7 e 2.8, respectivamente.

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i \quad (2.7)$$

$$\sigma^2 = \frac{1}{N-1} \left(\sum_{i=0}^{N-1} (x_i)^2 - \left(\sum_{i=0}^{N-1} x_i \right)^2 \right) \quad (2.8)$$

No presente trabalho, μ e σ , que são conhecidas como estatísticas de primeira ordem, serão utilizadas na etapa de classificação para formar os modelos de cada locutor matriculado no sistema de reconhecimento proposto. Os modelos em si serão caracterizados por funções Gaussianas de probabilidade, isto é, de acordo com a Equação 2.9, onde σ representa o desvio-padrão de $x[\cdot]$, ou seja, a raiz quadrada da sua variância (50).

$$f(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad (2.9)$$

Capítulo 3

Detalhamento do Trabalho Proposto

3.1 Considerações iniciais

Como primeira etapa para o desenvolvimento do trabalho proposto, foi selecionada a sentença “*She had your dark suit in greasy wash water all year*” dos sub-diretórios das pastas *test* e *train* da base de vozes TIMIT (48), que é pronunciada por todos os 630 locutores. A TIMIT é uma base largamente utilizada em trabalhos de pesquisa envolvendo o reconhecimento de fala e de locutores e é distribuída pelo *Linguistic Data Consortium*. Na sequência, foi realizada a especificação e a implementação da técnica de extração de características, conforme descrito na seção 3.2, com base nas concentrações energéticas normalizadas das sub-bandas *Bark*. Por fim, foi especificada e implementada a técnica de reconhecimento de padrões, conforme a seção 3.3, baseada em um modelamento Gaussiano. A acurácia do sistema, em cada situação experimental, foi então mensurada.

3.2 Extração de características

Como descrito anteriormente, a técnica de extração de características aqui proposta analisa somente as transições vozeadas/não-vozeadas, e vice-versa, entre fonemas. Computacionalmente, é possível identificar uma determinada janela, $w_k[\cdot]$, ($k = 0, 1, 2, \dots$), dos sinais de vozes como sendo vozeada ou não-vozeada observando a sua taxa de cruzamentos por zero (ZCR - *Zero Crossing Rate*). Após retirar o componente de corrente contínua da janela, o que equivale a subtrair, de cada amostra, o valor médio das amplitudes da janela, observa-se que trechos vozeados possuem taxas de cruzamentos por zero cerca de dez vezes menores do que trechos não-vozeados (49). Foram utilizadas janelas compostas de 512 amostras do sinal, com 50% de sobreposição entre janelas consecutivas. Dentro de cada janela, contabilizaram-se os ZCRs e, na sequência, calcularam-se os módulos das diferenças entre os valores. Em linguagem algorítmica, o procedimento é o seguinte:

- **INÍCIO**

double media;

Para cada sinal de voz $s_a[\cdot]$ de tamanho T :

Para cada amostra i de $s_a[\cdot]$, ($0 \leq i \leq T - 1$):

$media \leftarrow media + s[i];$

$media \leftarrow media / T;$

Para cada amostra i de $s[\cdot]$, ($0 \leq i \leq T - 1$):

$s[i] \leftarrow (s[i] - media);$

Definem-se os vetores $ZCR_{k,a}$, com todos os elementos nulos;

Para cada sinal de voz $s_a[\cdot]$:

Para cada janela $w_k[\cdot]$, ($k = 0, 1, 2, \dots$), de $s_a[\cdot]$, isto é, $w_{k,a}[\cdot]$, já com média nula:

Para cada amostra i de $w_{k,a}[\cdot]$, ($0 \leq i \leq 510$):

Se $w_{k,a}[i] \cdot w_{k,a}[i + 1] < 0$

então $ZCR_{k,a} \leftarrow ZCR_{k,a} + 1;$

Definem-se os vetores $ZCR_DIFFERENCES_{k,a}$, com todos os elementos nulos;

Para $k = 0, 1, 2, \dots$:

$ZCR_DIFFERENCES_{k,a} \leftarrow |ZCR_{k+1,a} - ZCR_{k,a}|;$

- **FIM**

A partir de um valor, definido como limiar de transição, o qual foi obtido empiricamente, foram selecionadas as posições de $ZCR_DIFFERENCES_{k,a}$ que apresentam um valor acima

dele, caracterizando uma transição vozeada/não-vozeada do respectivo sinal. Com base em alguns testes que foram realizados, conseguimos observar que sinais vozeados apresentam uma média de 20 cruzamentos por zero, enquanto sinais não-vozeados apresentam cerca de 100 cruzamentos por zero. Logo, subtraindo ambos os valores de cruzamentos, chegamos ao limiar de transição no valor de 80 cruzamentos por zero para identificar uma transição vozeada/não-vozeada no sinal.

Para calcular as energias de cada transição selecionada, que corresponde à duas janelas sobrepostas em 50% totalizando $512 + 512 - \frac{512}{2} = 768$ amostras, primeiramente foram construídos 21 filtros passa-faixa para os 21 primeiros intervalos de frequência da escala *Bark*. Os filtros passa-faixa foram obtidos a partir de dois filtros passa-baixas, sendo um para cada frequência do intervalo em questão. Os coeficientes dos filtros passa-baixas são calculados de acordo com a Equação 2.1. Com os filtros passa-baixas construídos, foram realizadas as subtrações ponto-a-ponto deles, obtendo-se então os filtros passa-faixa. Todos os filtros passa-faixa possuem 100 coeficientes. O processo de convolução foi aplicado entre a transição vozeada/não-vozeada do sinal e o filtro passa-faixa, seguindo os passos descritos na Seção 2.4.4, resultando nos sinais filtrados.

Na sequência, para cada transição, foram calculadas as 21 energias das 21 filtragens, de acordo com a Equação 2.4. Não foi necessário, e nem possível, trabalhar com todos os 24 intervalos de frequências da escala *Bark*, uma vez que a frequência máxima dos sinais de voz da base TIMIT é de $\frac{16000}{2} = 8000\text{Hz}$.

Como último passo, as energias calculadas foram normalizadas para padronizá-las no intervalo de 0 a 1. Além das 21 energias, foram calculados também seus 20 valores delta e 19 valores delta-delta, respeitando as Equações 2.5 e 2.6, respectivamente, totalizando assim, $21 + 20 + 19 = 60$ características que representam cada transição utilizada.

3.3 Reconhecimento de locutores

O número, K , de vetores de características, $c_{k,i}[\cdot]$, de dimensão 60 de cada locutor l_i é variável, pois cada transição vozeada/não vozeada produz um deles. Assim, foram obtidos o vetor de médias, $\mu_i[\cdot]$, e o vetor de variâncias, $\sigma^2_i[\cdot]$, para cada locutor l_i , ($0 \leq i \leq L$), cada um com dimensão 60. O valor de L varia e será discutido no próximo Capítulo. Ambos os vetores, $\mu_i[\cdot]$ e $\sigma^2_i[\cdot]$, foram utilizados para construir o modelo de cada locutor l_i . Assim,

$$\mu_i[j] = \frac{1}{K} \sum_{k=0}^{K-1} c_{k,i}[j] \quad , \quad (0 \leq j \leq 59) \quad , \quad (0 \leq i \leq L-1) \quad (3.1)$$

e

$$\sigma^2_i[j] = \frac{1}{K} \left(\sum_{k=0}^{K-1} (c_{k,i}[j])^2 - \left(\sum_{k=0}^{K-1} c_{k,i}[j] \right)^2 \right) \quad , \quad (0 \leq j \leq 59) \quad , \quad (0 \leq i \leq L-1) \quad (3.2)$$

Com o vetor de médias e o de variâncias de cada locutor calculados, prossegue-se para o reconhecimento. Nesta etapa, trabalha-se com base na tabulação da distribuição Gaussiana de probabilidades com $\mu = 0$ e $\sigma^2 = 1$, isto é, Normal(0,1) ou $\mathcal{N}(0, 1)$, conforme define Meyer (50). Então, a probabilidade $P(Y)$ de uma variável aleatória contínua Y é dada pela Equação 3.3.

$$P\left(\frac{a - \mu}{\sigma} \leq Y \leq \frac{b - \mu}{\sigma}\right) \quad (3.3)$$

Tendo em vista que a distribuição de probabilidades em questão é contínua, a probabilidade de um ponto específico é sempre nula. Assim, adota-se um intervalo de amplitude $0.5 + 0.5 = 1$ para os cálculos das probabilidades em torno do ponto r desejado, como é especificado na Equação 3.4.

$$P\left(\left(\frac{r - \mu}{\sigma}\right) - 0,5 \leq Y \leq \left(\frac{r - \mu}{\sigma}\right) + 0,5\right) \quad (3.4)$$

Particularmente, a determinação da média e da variância de cada característica de determinado locutor faz com o que os parâmetros do seu modelo estejam estabelecidos. Posteriormente, o modelo do locutor l_i , ($0 \leq i \leq L$), é

$$P(l_i) = \prod_{i=0}^{59} P\left(\left(\frac{r_i - \mu}{\sigma}\right) - 0,5 \leq Y_i \leq \left(\frac{r_i - \mu}{\sigma}\right) + 0,5\right) \quad (3.5)$$

pois assume-se independência estatística entre as características (50).

A identificação de um determinado locutor implica no cálculo do seu vetor de características, com 60 elementos, e a posterior substituição de cada um desses valores nos L modelos, conforme a Equação 3.5. O locutor identificado, $\mathcal{L}$, corresponde ao maior valor de $P(l_i)$, isto é,

$$\mathcal{L} = \max\{P(l_i)\} \quad (3.6)$$

No próximo Capítulo, constam os testes realizados variando o valor de L , assim como os comentários pertinentes.

Capítulo 4

Testes e Resultados

4.1 Considerações iniciais

Neste capítulo, os resultados dos testes realizados com diferentes quantidades de locutores foram analisados graficamente, focando nas características extraídas e como elas influenciaram na acurácia global.

4.2 Testes utilizando $L = 10$ locutores

Nos 67 testes realizados com $L = 10$ locutores, correspondentes à quantidade de vetores de características obtidos, foi alcançada uma acurácia média de 92,53%. A tabela 4.1 contém maiores detalhes dos testes realizados. Gráficos foram gerados para todos os 62 casos de acertos positivos dos testes, onde cada reconhecimento positivo, plotado na cor azul, foi comparado com todos os erros obtidos, plotados na cor vermelha.

Tabela 4.1 – 10 locutores: acertos e erros obtidos nos testes

Locutor	Acertos	Locutor	Acertos
l_1	100 %	l_6	100 %
l_2	100 %	l_7	100 %
l_3	85.71 %	l_8	87.50 %
l_4	100 %	l_9	87.50 %
l_5	100 %	l_{10}	77.78 %

Fonte: Elaborada pelo autor

Observando os gráficos, é possível perceber que os 21 valores principais de energias possuem sempre o mesmo comportamento, uma vez que tais valores estão normalizados. Considerando os 20 valores Delta, percebe-se que, pelo menos, quatro deles distoam diversas vezes dos gráficos de erros, como é o caso das características 29, 38, 39 e 40. Já dentro dos 19 valores delta-delta, também percebe-se tal comportamento nas características 41, 42, 49, 55, 56, 57, 58, 59 e 60. Isso acarreta em um certo impacto no cálculo das probabilidades parciais de cada característica e, conseqüentemente, na probabilidade final e nos acertos obtidos nos testes. O maior valor que as características atingem chega próximo de 156, como é observado na característica 59 da figura 4.6. Aqui, destacamos os gráficos de todos os acertos positivos de um mesmo locutor, no caso o locutor de número 10 (da figura 4.1 até a figura 4.7), onde podemos observar a variação existente nos valores das características extraídas destacadas anteriormente.

Figura 4.1 – Teste com 10 locutores: Locutor 10, Gráfico 1

Fonte: Elaborada pelo autor

Figura 4.2 – Teste com 10 locutores: Locutor 10, Gráfico 2

Fonte: Elaborada pelo autor

Figura 4.3 – Teste com 10 locutores: Locutor 10, Gráfico 3

Fonte: Elaborada pelo autor

Figura 4.4 – Teste com 10 locutores: Locutor 10, Gráfico 4

Fonte: Elaborada pelo autor

Figura 4.5 – Teste com 10 locutores: Locutor 10, Gráfico 5

Fonte: Elaborada pelo autor

Figura 4.6 – Teste com 10 locutores: Locutor 10, Gráfico 6

Fonte: Elaborada pelo autor

Figura 4.7 – Teste com 10 locutores: Locutor 10, Gráfico 7

Fonte: Elaborada pelo autor

4.3 Testes utilizando 20 locutores

Nos 139 testes realizados com $L = 20$ locutores, correspondentes à quantidade de vetores de características obtidos, foi alcançada uma acurácia média de 78,41%. A tabela 4.2 contém maiores detalhes dos testes realizados. Gráficos foram gerados para todos os 109 casos de acertos positivos dos testes, onde cada reconhecimento positivo, plotado na cor azul, foi comparado com todos os erros obtidos, plotados na cor vermelha.

Tabela 4.2 – 20 locutores: acertos e erros obtidos nos testes

Locutor	Acertos	Locutor	Acertos
l_1	80 %	l_{11}	100 %
l_2	85.71 %	l_{12}	75 %
l_3	100 %	l_{13}	66.67 %
l_4	80 %	l_{14}	66.67 %
l_5	100 %	l_{15}	71.43 %
l_6	62.50 %	l_{16}	50 %
l_7	85.71 %	l_{17}	100 %
l_8	100 %	l_{18}	87.50 %
l_9	83.33 %	l_{19}	42.86 %
l_{10}	50 %	l_{20}	100 %

Fonte: Elaborada pelo autor

Novamente, como os 21 valores principais de energia se encontram normalizados, o comportamento deles são semelhantes. Observando, então, os 20 valores Delta, percebe-se, graficamente, que as características 22, 38 e 40 se diferenciam de forma mais acentuada, algumas vezes, quando comparadas com os valores de erros destas mesmas características citadas. Dentre os 19 valores de delta-delta, pode-se perceber esse comportamento com maior frequência nas características 42, 58, 59 e 60. O maior valor obtido para as características atinge um valor próximo de 380, como é o caso da característica 42 da figura 4.16. Aqui, destacamos os gráficos de todos os acertos positivos de um mesmo locutor, no caso o locutor de número 2 (da figura 4.8 até a figura 4.19), onde podemos observar a variação existente nos valores das características extraídas destacadas anteriormente.

Figura 4.8 – Teste com 20 locutores: Locutor 2, Gráfico 1

Fonte: Elaborada pelo autor

Figura 4.9 – Teste com 20 locutores: Locutor 2, Gráfico 2

Fonte: Elaborada pelo autor

Figura 4.10 – Teste com 20 locutores: Locutor 2, Gráfico 3

Fonte: Elaborada pelo autor

Figura 4.11 – Teste com 20 locutores: Locutor 2, Gráfico 4

Fonte: Elaborada pelo autor

Figura 4.12 – Teste com 20 locutores: Locutor 2, Gráfico 5

Fonte: Elaborada pelo autor

Figura 4.13 – Teste com 20 locutores: Locutor 2, Gráfico 6

Fonte: Elaborada pelo autor

Figura 4.14 – Teste com 20 locutores: Locutor 2, Gráfico 7

Fonte: Elaborada pelo autor

Figura 4.15 – Teste com 20 locutores: Locutor 2, Gráfico 8

Fonte: Elaborada pelo autor

Figura 4.16 – Teste com 20 locutores: Locutor 2, Gráfico 9

Fonte: Elaborada pelo autor

Figura 4.17 – Teste com 20 locutores: Locutor 2, Gráfico 10

Fonte: Elaborada pelo autor

Figura 4.18 – Teste com 20 locutores: Locutor 2, Gráfico 11

Fonte: Elaborada pelo autor

Figura 4.19 – Teste com 20 locutores: Locutor 2, Gráfico 12

Fonte: Elaborada pelo autor

4.4 Testes utilizando 30 locutores

Nos 176 testes realizados com $L = 30$ locutores, correspondentes à quantidade de vetores de características obtidos, foi alcançada uma acurácia média de 73.29%. A tabela 4.3 contém maiores detalhes dos testes realizados. Gráficos foram gerados para todos os 129 casos de acertos positivos dos testes, onde cada reconhecimento positivo, plotado na cor azul, foi comparado com todos os erros obtidos, plotados na cor vermelha.

Tabela 4.3 – 30 locutores: acertos e erros obtidos nos testes

Locutor	Acertos	Locutor	Acertos	Locutor	Acertos
l_1	66.67 %	l_{11}	20 %	l_{21}	42.86 %
l_2	100 %	l_{12}	83.34 %	l_{22}	100 %
l_3	66.67 %	l_{13}	50 %	l_{23}	100 %
l_4	71.43 %	l_{14}	71.43 %	l_{24}	62.50 %
l_5	100 %	l_{15}	44.45 %	l_{25}	83.34 %
l_6	50 %	l_{16}	60 %	l_{26}	83.34 %
l_7	88.89 %	l_{17}	100 %	l_{27}	60 %
l_8	100 %	l_{18}	75 %	l_{28}	100 %
l_9	80 %	l_{19}	50 %	l_{29}	100 %
l_{10}	87.50 %	l_{20}	66.67 %	l_{30}	100 %

Fonte: Elaborada pelo autor

Neste último teste realizado observa-se, dentre os 20 valores Delta, que as características não possuem uma variação expressiva em relação aos erros dos testes, mas podem ser notadas algumas variações mais acentuadas das características 37 e 38. Considerando os 19 valores

de delta-delta, é possível perceber um comportamento mais diferenciado nas características 57, 58, 59 e 60. A característica 42 apresenta o maior valor extraído, atingindo uma amplitude próxima de 670, como é observado na figura 4.22. Aqui, destacamos os gráficos de todos os acertos positivos de um mesmo locutor, no caso o locutor de número 20 (da figura 4.20 até a figura 4.23), onde podemos observar a variação existente nos valores das características extraídas destacadas anteriormente.

Figura 4.20 – Teste com 30 locutores: Locutor 20, Gráfico 1

Fonte: Elaborada pelo autor

Figura 4.21 – Teste com 30 locutores: Locutor 20, Gráfico 2

Fonte: Elaborada pelo autor

Figura 4.22 – Teste com 30 locutores: Locutor 20, Gráfico 3

Fonte: Elaborada pelo autor

Figura 4.23 – Teste com 30 locutores: Locutor 20, Gráfico 4

Fonte: Elaborada pelo autor

4.5 Testes com mais de 30 locutores e análise global

A partir desse conjunto de testes, com $L > 30$, não serão exibidos gráficos específicos para cada caso, como realizado anteriormente, em função do espaço requerido. Ao contrário, a tabela 4.4 contém os índices globais de acurácias obtidos para alguns testes, com a quantidade de locutores variando até 630. Os valores para $L = 10, 20$ e 30 correspondem aos experimentos

das três seções anteriores, enquanto que os demais casos listados na tabela 4.4 indicam os testes para os demais casos.

Observando graficamente, por meio da figura 4.24, o comportamento global do sistema, considerando-se L versus acurácia, percebe-se uma variação exponencial. Ajustando os dados da tabela 4.4 por uma curva média entre os pontos, objetivando minimizar o erro quadrático médio com o Método dos Mínimos Quadrados (51), foi obtida a seguinte lei de variação, representada pela Equação 4.1, onde:

- x representa a quantidade de locutores utilizada;
- y representa a acurácia.

$$y = 60.76 \cdot e^{-0.002 \cdot x} \quad (4.1)$$

Basicamente, a aproximação dos dados da tabela 4.4 para uma função da forma $y = a \cdot e^{b \cdot x}$ (Equação 4.1), consiste em realizar substituições básicas para que se possa utilizar o Método dos Mínimos Quadrados:

$$y = a \cdot e^{b \cdot x}$$

$$\ln(y) = \ln(a \cdot e^{b \cdot x})$$

$$\ln(y) = \ln(a) + b \cdot x$$

$$Y = a_1 \cdot x + a_0$$

Considerando a minimização do erro quadrático da aproximação, os valores de a_1 e a_0 são calculados de acordo com as Equações 4.2 e 4.3, respectivamente. Sabendo que $b = a_1$ e que $a = e^{a_0}$, pelas substituições feitas anteriormente, a aproximação desejada da função é obtida.

$$a_1 = \frac{n \cdot (\sum_{k=1}^n x_k \cdot Y_k) - (\sum_{k=1}^n x_k) \cdot (\sum_{k=1}^n Y_k)}{n \cdot (\sum_{k=1}^n x_k^2) - (\sum_{k=1}^n x_k)^2} \quad (4.2)$$

$$a_0 = \frac{(\sum_{k=1}^n x_k^2) \cdot (\sum_{k=1}^n Y_k) - (\sum_{k=1}^n x_k \cdot Y_k) \cdot (\sum_{k=1}^n x_k)}{n \cdot (\sum_{k=1}^n x_k^2) - (\sum_{k=1}^n x_k)^2} \quad (4.3)$$

Desse modo, conclui-se que a acurácia decresce exponencialmente em função do número de locutores matriculados no sistema de reconhecimento quando as características são extraídas das transições entre os trechos vozeado/não-vozeado. Para a base TIMIT, que é a mais utilizada internacionalmente em processamento de voz, este estudo, levando em conta características energéticas e classificadores Gaussianos, que são os mais amplamente empregados, ainda não havia sido realizado.

Figura 4.24 – Análise global dos resultados de L versus acurácia.

Fonte: Elaborada pelo autor

Tabela 4.4 – Acurácia global para todos os testes realizados

Locutores	Acurácia	Locutores	Acurácia	Locutores	Acurácia
10	92.53%	90	47.65%	170	36.99%
20	77.85%	100	44.45%	180	36.41%
30	73.29%	110	42.81%	190	35.26%
40	65.08%	120	42.05%	200	33.89%
50	61.14%	130	40.86%	600	20.92%
60	56.76%	140	39.35%	630	20.18%
70	52.39%	150	38.84%		
80	50.40%	160	37.64%		

Fonte: Elaborada pelo autor

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho, após uma etapa de revisão de conceitos sobre processamento de sinais e reconhecimento de locutores, foi apresentado um sistema para reconhecimento de locutores com base em concentrações energéticas de sub-bandas de frequências *Bark* e um classificador Gaussiano. Por meio de tabelas e gráficos, foram apresentados e analisados os testes e resultados.

Os resultados obtidos nos testes efetuados variaram de 92.53% a 20.18% de acurácia, conforme a quantidade de locutores foi sendo aumentada em cada cenário de testes. Para os casos de teste iniciais, gráficos das características extraídas foram gerados a fim de ilustrar seus respectivos comportamentos e analisar de que forma seus valores influenciaram na acurácia dos testes.

A associação das concentrações energéticas com o classificador Gaussiano é uma das mais tradicionais e mais clássicas que pode ser encontrada na literatura e, justamente por isso, foi a abordagem adotada. A base de áudios TIMIT, pelos mesmos motivos citados anteriormente, foi a escolhida para que o trabalho fosse desenvolvido, agregando assim, novos resultados à literatura da área. A intenção, portanto, não foi a de comparar características ou classificadores e nem mesmo de propor qualquer avanço ao estado-da-arte na área. Por outro lado, o objetivo pretendido, que foi o de quantificar a acurácia do reconhecimento dos locutores em função da quantidade deles, extraindo as características das transições entre os trechos vozeados e não-vozeados das respectivas locuções, foi alcançado. Este estudo, refletindo o comportamento *quantidade versus acurácia*, complementou as informações existentes na literatura, quantifi-

cando as informações previamente existentes e comprovando um comportamento exponencial.

Em termos de trabalhos futuros, a intenção é a de estender o estudo para uma abordagem *text-independent*, analisando o respectivo comportamento nesse novo cenário, assim como utilizar outras bases de dados clássicas. Uma extração diferenciada das características também é uma outra abordagem de estudo válida, onde se pode variar a quantidade de características extraídas e a forma com que tal processo de extração é realizado, mas ainda focando apenas nos trechos de transições vozeados e não-vozeados, e sempre visando a divulgação dos resultados obtidos.

Referências

- 1 PETRY, A. *Reconhecimento automático de locutor utilizando medidas de invariantes dinâmicas não-lineares*. 2002. 155 p. Tese (Doutorado em Ciência da Computação)-Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
- 2 CAMPBELL, J. P. et al. Forensic speaker recognition: a need for caution. *IEEE Signal Processing Magazine*, v. 26, n. 2, p. 95-103, 2009. doi:10.1109/msp.2008.931100.
- 3 ACADEFORD. Disponível em: <<http://www.acadefor.com.br/>>. Acesso em: 12 ago. 2014.
- 4 BARBON JUNIOR, S. *Dynamic time warping baseado na Transformada Wavelet*. 2008. 113 p. Dissertação (Mestrado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 5 VIEIRA, L. S. *Conversão de voz baseada na Transformada Wavelet*. 2008. 91 p. Dissertação (Mestrado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 6 FANTINATO, P. C. *Segmentação de voz baseada na análise fractal e na Transformada Wavelet*. 2008. 123 p. Dissertação (Mestrado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 7 REVETT, K. *Behavioral biometrics: a remote access approach*. Chichester: John Wiley & Sons, 2008.
- 8 DENG, L.; O'SHAUGHNESSY, O. *Speech processing: a dynamic and optimization-oriented approach*. New York: Marcel Dekker, 2003.
- 9 CHOU, W.; JUANG, B. H. *Pattern recognition in speech and language processing*. New York: CRC Press, 2003.
- 10 LACERDA, M. A. et al. A wavelet-based speaker verification algorithm. *International Journal on Wavelets, Multiresolution, and Information Processing*, v. 8, n. 6, p. 905-912, 2010.

- 11 GUIDO, R. C.; PEREIRA, J. C.; SLAETS, J. F. W. Introduction to the special issue on advances on pattern recognition for speech and audio processing. *Pattern Recognition Letters*, v. 28, n. 11, p. 1283-1284, 2007.
- 12 GUIDO, R. C. et al. A neural-wavelet architecture for voice conversion. *Neurocomputing*, v. 71, n. 1-3, p. 174-180, 2007.
- 13 KERSTA, L. G. Voiceprint identification. *Nature*, v. 196, n. 4861, p. 1253-1257, 1962.
- 14 BOLT, R. H. et al. Speaker identification by speech spectrograms: a scientists' view of its reliability for legal purposes. *Journal of the Acoustical Society of America*, v. 47, n. 2, p. 597-612, 1970.
- 15 NOLAN, J. F. *The phonetic bases of speaker recognition*. New York: Cambridge University Press, 1983.
- 16 BOE, L. J. Forensic voice identification in France. *Speech Communication*, v. 31, n. 2-3, p. 205-224, 2000.
- 17 NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. Disponível em: <<http://www.nist.gov/>>. Acesso em: 16 ago. 2014.
- 18 KENNY, P.; BOULIANNE, G.; DUMOUCHEL, P. Eigenvoice modeling with sparse training data. *IEEE Transactions on Speech and Audio Processing*, v. 13, n. 3, p. 345-354, 2005. doi:10.1109/TSA.2004.840940.
- 19 CAMPBELL, W. M.; STURIM, D. E.; REYNOLDS, D. A. Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Processing Letters*, v. 13, n. 5, p. 308-311, 2006. doi:10.1109/LSP.2006.870086.
- 20 BEIGI, H. *Fundamentals of speaker recognition*. New York: Springer, 2011.
- 21 LIÑARES, L. R.; MATEO, C. G. Phonetically-trained models for speaker recognition. *Journal of the Acoustical Society of America*, v. 109, n. 1, p. 1-23, 2014. doi:10.1121/1.1331679.
- 22 WEDDIN, M. E. M.; WINTHER, O. *Frame selection for speaker identification*. 2005. Disponível em: <http://www.academia.edu/2720751/Frame_selection_for_speaker_identification>. Acesso em: 16 ago. 2014.
- 23 FURUI, S. *Digital speech processing, synthesis, and recognition*. 2nd ed. New York: CRC Press, 2000.

- 24 LUNG, S. Y. Improved wavelet feature extraction using kernel analysis for text independent speaker recognition. *Digital Signal Processing*, v. 20, n. 5, p. 1400-1407, 2009.
- 25 WU, J. D.; TSAI, Y. J. Speaker identification system using empirical mode decomposition and an artificial neural network. *Expert Systems with Applications*, v. 38, n. 5, p. 6112-6117, 2011. doi:10.1016/j.eswa.2010.11.013.
- 26 RAO, K. S. et al. Robust speaker recognition on mobile devices. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATIONS - SPCOM, 2010, Bangalore, India, *Proceedings...* Bangalore: IEEE Explorer, 2011, p. 1-5. doi:10.1109/SPCOM.2010.5560542.
- 27 SADIC, S.; GULMEZOGLU, M. B. Common vector approach and its combination with GMM for text-independent speaker recognition. *Expert Systems with Applications*, v. 38, n. 9, p. 11394-11400, 2011. doi:10.1016/j.eswa.2011.03.009.
- 28 KUMARI, R. S. S.; NIDHYANANTHAN, S. S.; ANAND, G. Fused mel feature sets based text-independent speaker identification using gaussian mixture model. *Procedia Engineering*, v. 30, p. 319-326, 2011. doi:10.1016/j.proeng.2012.01.867.
- 29 ANDREI, V. Implementation of a real-time text dependent speaker identification system. In: CONFERENCE ON SPEECH TECHNOLOGY AND HUMAN-COMPUTER DIALOGUE - SpeD, 6., 2011, Brasov, Romania. *Proceedings...* Brasov: IEEE Explorer, 2011, p. 1-6. doi:10.1109/SPED.2011.5940739.
- 30 GHIURCAU, M. V.; RUSU, C. A study of the effect of emotional state upon text-independent speaker identification. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING - ICASSP, 2011, Prague, United States. *Proceedings...* Prague: IEEE Explorer, 2011. p. 4944-4947. doi:10.1109/ICASSP.2011.5947465.
- 31 TOLBA, H. A high-performance text-independent speaker identification of arabic speakers using a CHMM-based approach. *Alexandria Engineering Journal*, v. 50, n. 1, p. 43-47, 2011. doi:10.1016/j.aej.2011.01.007.
- 32 AJMERA, P. K.; JADHAV, D. V.; HOLAMBE, R. S. Text-independent speaker identification using Radon and discrete cosine transforms based features from speech spectrogram. *Pattern Recognition*, v. 44, n. 10-11, p. 2749-2759, 2011. doi:10.1016/j.patcog.2011.04.009.
- 33 ZHU, L.; YANG, Q. Speaker recognition system based on weighted feature parameter. *Physics Procedia*, v. 25, p. 1515-1522, 2012. doi:10.1016/j.phpro.2012.03.270.

- 34 LI, Y. G. et al. Multi-layered features with SVM for text-independent speaker verification. In: INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE & SERVICE SYSTEM - CSSS, 2012, Nanjing, China. *Proceedings...* Nanjing: IEEE Explorer, 2012, p. 378-380. doi:10.1109/CSSS.2012.101.
- 35 SHAHIN, I. Speaker identification in emotional talking environments using both gender and emotion cues. In: INTERNATIONAL CONFERENCE ON COMMUNICATIONS, SIGNAL PROCESSING, AND THEIR APPLICATIONS - ICCSPA, 1., 2013, Sharjah, United Arab Emirates. *Proceedings...* Sharjah: IEEE Explorer, 2013, p. 1-6. doi:10.1109/ICCSPA.2013.6487314.
- 36 JAWAKAR, N. P.; HOLAMBE, R. S.; BASU, T. K. Speaker identification using whispered speech. In: INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS AND NETWORK TECHNOLOGIES - CSNT, 2013, Gwalior, India. *Proceedings...* Gwalior: IEEE Explorer, 2013. p. 778-781. doi:10.1109/CSNT.2013.167.
- 37 GUPTA, A.; GUPTA, H. Applications of MFCC and vector quantization in speaker recognition. In: INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS AND SIGNAL PROCESSING - ISSP, 2013, Gujarat, India. *Proceedings...* Gujarat: IEEE Explorer, 2013. p. 170-173. doi:10.1109/ISSP.2013.6526896.
- 38 JAYAKUMAR, A.; ANTOP, B. Comparison of wavelet packets and DWT in a gender based multichannel speaker recognition system. In: IEEE CONFERENCE ON INFORMATION & COMMUNICATION TECHNOLOGIES - ICT, 2013, JeJu Island, South Korea. *Proceedings...* JeJu Island: IEEE Explorer, 2013. p. 847-850. doi:10.1109/CICT.2013.6558212.
- 39 VIKRAM, C. M.; UMARANI, K. Text independent classification of normal and pathological voices using MFCCs and GMM-UBM. In: IEEE CONFERENCE ON INFORMATION & COMMUNICATION TECHNOLOGIES - ICT, 2013, JeJu Island, South Korea. *Proceedings...* JeJu Island, IEEE Explorer, 2013. p. 1215-1220. doi:10.1109/CICT.2013.6558286.
- 40 TRUESDELL, C. *Mastering digital audio production: the professional music workflow with Mac OS X*. Indianapolis: Wiley Publishing, 2007.
- 41 HAVELOCK, D.; KUWANO, S.; VORLANDER, M. *Handbook of signal processing in acoustics*. New York: Springer, 2008.
- 42 BOSI, M.; GOLDBERG, R. E. *Introduction to digital audio coding and standards*. Norwell: Kluwer Academic Publishers, 2003.
- 43 SANTOS, J. C. V. *Introdução à teoria de redes*. Canoas: Ulbra, 2003.
- 44 HAYKIN, S.; VEEN, B. V. *Sinais e sistemas*. Porto Alegre: Bookman, 2001.

- 45 LATHI, B. P. *Sinais e sistemas lineares*. 2. ed. Porto Alegre: Bookman, 2007.
- 46 DENG, L. *Dynamic speech models: theory, algorithms, and applications*. Redmond: Morgan & Claypool, 2006.
- 47 KINNUNEN, T.; LI, H. An overview of text-independent speaker recognition: from features to supervectors. *Speech Communication*, v. 52, n. 1, p. 12-40, 2009. doi:10.1016/j.specom.2009.08.009.
- 48 FUJISASKI, H. *Recent research towards advanced man-machine interface through spoken language*. Amsterdam: Elsevier, 1996.
- 49 JALIL, M.; BUTT, F. A.; MALIK, A. Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. In: INTERNATIONAL CONFERENCE ON TECHNOLOGICAL ADVANCES IN ELECTRICAL, ELECTRONICS AND COMPUTER ENGINEERING - TAECE, 2013, Konya, Turkey. *Proceedings...* Konya, IEEE Explorer, 2013. p. 208-212. doi:10.1109/TAECE.2013.6557272.
- 50 MEYER, P. L. *Probabilidade: aplicações à estatística*. 2. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1983.
- 51 FRANCO, N. M. B. *Cálculo numérico*. São Paulo: Pearson Prentice Hall, 2007.

Apêndice I - Gráficos das características extraídas

Figura 1 – Teste com 10 locutores: Locutor 1, Gráfico 1

Fonte: Elaborada pelo autor

Figura 2 – Teste com 10 locutores: Locutor 1, Gráfico 2

Fonte: Elaborada pelo autor

Figura 3 – Teste com 10 locutores: Locutor 1, Gráfico 3

Fonte: Elaborada pelo autor

Figura 4 – Teste com 10 locutores: Locutor 1, Gráfico 4 **Fonte: Elaborada pelo autor****Figura 5 –** Teste com 10 locutores: Locutor 2, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 6 –** Teste com 10 locutores: Locutor 2, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 7 –** Teste com 10 locutores: Locutor 2, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 8 –** Teste com 10 locutores: Locutor 3, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 9 –** Teste com 10 locutores: Locutor 3, Gráfico 2 **Fonte: Elaborada pelo autor**

Figura 10 – Teste com 10 locutores: Locutor 3, Gráfico 3

Fonte: Elaborada pelo autor

Figura 11 – Teste com 10 locutores: Locutor 3, Gráfico 4

Fonte: Elaborada pelo autor

Figura 12 – Teste com 10 locutores: Locutor 3, Gráfico 5

Fonte: Elaborada pelo autor

Figura 13 – Teste com 10 locutores: Locutor 3, Gráfico 6

Fonte: Elaborada pelo autor

Figura 14 – Teste com 10 locutores: Locutor 4, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 15 –** Teste com 10 locutores: Locutor 4, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 16 –** Teste com 10 locutores: Locutor 4, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 17 –** Teste com 10 locutores: Locutor 4, Gráfico 4 **Fonte: Elaborada pelo autor****Figura 18 –** Teste com 10 locutores: Locutor 4, Gráfico 5 **Fonte: Elaborada pelo autor**

Figura 19 – Teste com 10 locutores: Locutor 4, Gráfico 6 **Fonte: Elaborada pelo autor****Figura 20 –** Teste com 10 locutores: Locutor 4, Gráfico 7 **Fonte: Elaborada pelo autor****Figura 21 –** Teste com 10 locutores: Locutor 4, Gráfico 8 **Fonte: Elaborada pelo autor****Figura 22 –** Teste com 10 locutores: Locutor 4, Gráfico 9 **Fonte: Elaborada pelo autor****Figura 23 –** Teste com 10 locutores: Locutor 4, Gráfico 10 **Fonte: Elaborada pelo autor**

Figura 24 – Teste com 10 locutores: Locutor 5, Gráfico 1

Fonte: Elaborada pelo autor

Figura 25 – Teste com 10 locutores: Locutor 5, Gráfico 2

Fonte: Elaborada pelo autor

Figura 26 – Teste com 10 locutores: Locutor 5, Gráfico 3

Fonte: Elaborada pelo autor

Figura 27 – Teste com 10 locutores: Locutor 5, Gráfico 4

Fonte: Elaborada pelo autor

Figura 28 – Teste com 10 locutores: Locutor 5, Gráfico 5

Fonte: Elaborada pelo autor

Figura 29 – Teste com 10 locutores: Locutor 5, Gráfico 6

Fonte: Elaborada pelo autor

Figura 30 – Teste com 10 locutores: Locutor 5, Gráfico 7

Fonte: Elaborada pelo autor

Figura 31 – Teste com 10 locutores: Locutor 5, Gráfico 8

Fonte: Elaborada pelo autor

Figura 32 – Teste com 10 locutores: Locutor 5, Gráfico 9

Fonte: Elaborada pelo autor

Figura 33 – Teste com 10 locutores: Locutor 6, Gráfico 1

Fonte: Elaborada pelo autor

Figura 34 – Teste com 10 locutores: Locutor 6, Gráfico 2

Fonte: Elaborada pelo autor

Figura 35 – Teste com 10 locutores: Locutor 6, Gráfico 3

Fonte: Elaborada pelo autor

Figura 36 – Teste com 10 locutores: Locutor 6, Gráfico 4

Fonte: Elaborada pelo autor

Figura 37 – Teste com 10 locutores: Locutor 7, Gráfico 1

Fonte: Elaborada pelo autor

Figura 38 – Teste com 10 locutores: Locutor 7, Gráfico 2

Fonte: Elaborada pelo autor

Figura 39 – Teste com 10 locutores: Locutor 7, Gráfico 3

Fonte: Elaborada pelo autor

Figura 40 – Teste com 10 locutores: Locutor 7, Gráfico 4

Fonte: Elaborada pelo autor

Figura 41 – Teste com 10 locutores: Locutor 7, Gráfico 5

Fonte: Elaborada pelo autor

Figura 42 – Teste com 10 locutores: Locutor 8, Gráfico 1

Fonte: Elaborada pelo autor

Figura 43 – Teste com 10 locutores: Locutor 8, Gráfico 2

Fonte: Elaborada pelo autor

Figura 44 – Teste com 10 locutores: Locutor 8, Gráfico 3

Fonte: Elaborada pelo autor

Figura 45 – Teste com 10 locutores: Locutor 8, Gráfico 4

Fonte: Elaborada pelo autor

Figura 46 – Teste com 10 locutores: Locutor 8, Gráfico 5

Fonte: Elaborada pelo autor

Figura 47 – Teste com 10 locutores: Locutor 8, Gráfico 6

Fonte: Elaborada pelo autor

Figura 48 – Teste com 10 locutores: Locutor 8, Gráfico 7

Fonte: Elaborada pelo autor

Figura 49 – Teste com 10 locutores: Locutor 9, Gráfico 1

Fonte: Elaborada pelo autor

Figura 50 – Teste com 10 locutores: Locutor 9, Gráfico 2

Fonte: Elaborada pelo autor

Figura 51 – Teste com 10 locutores: Locutor 9, Gráfico 3

Fonte: Elaborada pelo autor

Figura 52 – Teste com 10 locutores: Locutor 9, Gráfico 4

Fonte: Elaborada pelo autor

Figura 53 – Teste com 10 locutores: Locutor 9, Gráfico 5

Fonte: Elaborada pelo autor

Figura 54 – Teste com 10 locutores: Locutor 9, Gráfico 6

Fonte: Elaborada pelo autor

Figura 55 – Teste com 10 locutores: Locutor 9, Gráfico 7

Fonte: Elaborada pelo autor

Figura 56 – Teste com 20 locutores: Locutor 1, Gráfico 1

Fonte: Elaborada pelo autor

Figura 57 – Teste com 20 locutores: Locutor 1, Gráfico 2

Fonte: Elaborada pelo autor

Figura 58 – Teste com 20 locutores: Locutor 1, Gráfico 3

Fonte: Elaborada pelo autor

Figura 59 – Teste com 20 locutores: Locutor 1, Gráfico 4

Fonte: Elaborada pelo autor

Figura 60 – Teste com 20 locutores: Locutor 1, Gráfico 5

Fonte: Elaborada pelo autor

Figura 61 – Teste com 20 locutores: Locutor 1, Gráfico 6

Fonte: Elaborada pelo autor

Figura 62 – Teste com 20 locutores: Locutor 1, Gráfico 7

Fonte: Elaborada pelo autor

Figura 63 – Teste com 20 locutores: Locutor 1, Gráfico 8

Fonte: Elaborada pelo autor

Figura 64 – Teste com 20 locutores: Locutor 1, Gráfico 9

Fonte: Elaborada pelo autor

Figura 65 – Teste com 20 locutores: Locutor 1, Gráfico 10

Fonte: Elaborada pelo autor

Figura 66 – Teste com 20 locutores: Locutor 1, Gráfico 11

Fonte: Elaborada pelo autor

Figura 67 – Teste com 20 locutores: Locutor 1, Gráfico 12

Fonte: Elaborada pelo autor

Figura 68 – Teste com 20 locutores: Locutor 3, Gráfico 1

Fonte: Elaborada pelo autor

Figura 69 – Teste com 20 locutores: Locutor 3, Gráfico 2

Fonte: Elaborada pelo autor

Figura 70 – Teste com 20 locutores: Locutor 3, Gráfico 3

Fonte: Elaborada pelo autor

Figura 71 – Teste com 20 locutores: Locutor 3, Gráfico 4

Fonte: Elaborada pelo autor

Figura 72 – Teste com 20 locutores: Locutor 3, Gráfico 5

Fonte: Elaborada pelo autor

Figura 73 – Teste com 20 locutores: Locutor 3, Gráfico 6

Fonte: Elaborada pelo autor

Figura 74 – Teste com 20 locutores: Locutor 4, Gráfico 1

Fonte: Elaborada pelo autor

Figura 75 – Teste com 20 locutores: Locutor 4, Gráfico 2

Fonte: Elaborada pelo autor

Figura 76 – Teste com 20 locutores: Locutor 4, Gráfico 3

Fonte: Elaborada pelo autor

Figura 77 – Teste com 20 locutores: Locutor 4, Gráfico 4

Fonte: Elaborada pelo autor

Figura 78 – Teste com 20 locutores: Locutor 5, Gráfico 1

Fonte: Elaborada pelo autor

Figura 79 – Teste com 20 locutores: Locutor 5, Gráfico 2

Fonte: Elaborada pelo autor

Figura 80 – Teste com 20 locutores: Locutor 5, Gráfico 3

Fonte: Elaborada pelo autor

Figura 81 – Teste com 20 locutores: Locutor 5, Gráfico 4

Fonte: Elaborada pelo autor

Figura 82 – Teste com 20 locutores: Locutor 6, Gráfico 1

Fonte: Elaborada pelo autor

Figura 83 – Teste com 20 locutores: Locutor 6, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 84 –** Teste com 20 locutores: Locutor 6, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 85 –** Teste com 20 locutores: Locutor 6, Gráfico 4 **Fonte: Elaborada pelo autor****Figura 86 –** Teste com 20 locutores: Locutor 6, Gráfico 5 **Fonte: Elaborada pelo autor**

Figura 87 – Teste com 20 locutores: Locutor 7, Gráfico 1

Fonte: Elaborada pelo autor

Figura 88 – Teste com 20 locutores: Locutor 7, Gráfico 2

Fonte: Elaborada pelo autor

Figura 89 – Teste com 20 locutores: Locutor 7, Gráfico 3

Fonte: Elaborada pelo autor

Figura 90 – Teste com 20 locutores: Locutor 7, Gráfico 4

Fonte: Elaborada pelo autor

Figura 91 – Teste com 20 locutores: Locutor 7, Gráfico 5

Fonte: Elaborada pelo autor

Figura 92 – Teste com 20 locutores: Locutor 7, Gráfico 6 **Fonte: Elaborada pelo autor****Figura 93 –** Teste com 20 locutores: Locutor 8, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 94 –** Teste com 20 locutores: Locutor 8, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 95 –** Teste com 20 locutores: Locutor 8, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 96 –** Teste com 20 locutores: Locutor 8, Gráfico 4 **Fonte: Elaborada pelo autor**

Figura 97 – Teste com 20 locutores: Locutor 8, Gráfico 5 **Fonte: Elaborada pelo autor****Figura 98 –** Teste com 20 locutores: Locutor 8, Gráfico 6 **Fonte: Elaborada pelo autor****Figura 99 –** Teste com 20 locutores: Locutor 8, Gráfico 7 **Fonte: Elaborada pelo autor****Figura 100 –** Teste com 20 locutores: Locutor 9, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 101 –** Teste com 20 locutores: Locutor 9, Gráfico 2 **Fonte: Elaborada pelo autor**

Figura 102 – Teste com 20 locutores: Locutor 9, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 103 –** Teste com 20 locutores: Locutor 9, Gráfico 4 **Fonte: Elaborada pelo autor****Figura 104 –** Teste com 20 locutores: Locutor 9, Gráfico 5 **Fonte: Elaborada pelo autor****Figura 105 –** Teste com 20 locutores: Locutor 10, Gráfico 1 **Fonte: Elaborada pelo autor**

Figura 106 – Teste com 20 locutores: Locutor 10, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 107 –** Teste com 20 locutores: Locutor 10, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 108 –** Teste com 20 locutores: Locutor 11, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 109 –** Teste com 20 locutores: Locutor 11, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 110 –** Teste com 20 locutores: Locutor 11, Gráfico 3 **Fonte: Elaborada pelo autor**

Figura 111 – Teste com 20 locutores: Locutor 11, Gráfico 4

Fonte: Elaborada pelo autor

Figura 112 – Teste com 20 locutores: Locutor 11, Gráfico 5

Fonte: Elaborada pelo autor

Figura 113 – Teste com 20 locutores: Locutor 11, Gráfico 6

Fonte: Elaborada pelo autor

Figura 114 – Teste com 20 locutores: Locutor 11, Gráfico 7

Fonte: Elaborada pelo autor

Figura 115 – Teste com 20 locutores: Locutor 12, Gráfico 1

Fonte: Elaborada pelo autor

Figura 116 – Teste com 20 locutores: Locutor 12, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 117 –** Teste com 20 locutores: Locutor 12, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 118 –** Teste com 20 locutores: Locutor 12, Gráfico 4 **Fonte: Elaborada pelo autor****Figura 119 –** Teste com 20 locutores: Locutor 12, Gráfico 5 **Fonte: Elaborada pelo autor****Figura 120 –** Teste com 20 locutores: Locutor 12, Gráfico 6 **Fonte: Elaborada pelo autor**

Figura 121 – Teste com 20 locutores: Locutor 13, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 122 –** Teste com 20 locutores: Locutor 13, Gráfico 2 **Fonte: Elaborada pelo autor****Figura 123 –** Teste com 20 locutores: Locutor 13, Gráfico 3 **Fonte: Elaborada pelo autor****Figura 124 –** Teste com 20 locutores: Locutor 13, Gráfico 4 **Fonte: Elaborada pelo autor****Figura 125 –** Teste com 20 locutores: Locutor 14, Gráfico 1 **Fonte: Elaborada pelo autor**

Figura 126 – Teste com 20 locutores: Locutor 14, Gráfico 2

Fonte: Elaborada pelo autor

Figura 127 – Teste com 20 locutores: Locutor 14, Gráfico 3

Fonte: Elaborada pelo autor

Figura 128 – Teste com 20 locutores: Locutor 14, Gráfico 4

Fonte: Elaborada pelo autor

Figura 129 – Teste com 20 locutores: Locutor 14, Gráfico 5

Fonte: Elaborada pelo autor

Figura 130 – Teste com 20 locutores: Locutor 14, Gráfico 6

Fonte: Elaborada pelo autor

Figura 131 – Teste com 20 locutores: Locutor 15, Gráfico 1

Fonte: Elaborada pelo autor

Figura 132 – Teste com 20 locutores: Locutor 15, Gráfico 2

Fonte: Elaborada pelo autor

Figura 133 – Teste com 20 locutores: Locutor 15, Gráfico 3

Fonte: Elaborada pelo autor

Figura 134 – Teste com 20 locutores: Locutor 15, Gráfico 4

Fonte: Elaborada pelo autor

Figura 135 – Teste com 20 locutores: Locutor 15, Gráfico 5

Fonte: Elaborada pelo autor

Figura 136 – Teste com 20 locutores: Locutor 16, Gráfico 1

Fonte: Elaborada pelo autor

Figura 137 – Teste com 20 locutores: Locutor 16, Gráfico 2

Fonte: Elaborada pelo autor

Figura 138 – Teste com 20 locutores: Locutor 16, Gráfico 3

Fonte: Elaborada pelo autor

Figura 139 – Teste com 20 locutores: Locutor 17, Gráfico 1

Fonte: Elaborada pelo autor

Figura 140 – Teste com 20 locutores: Locutor 17, Gráfico 2

Fonte: Elaborada pelo autor

Figura 141 – Teste com 20 locutores: Locutor 18, Gráfico 1

Fonte: Elaborada pelo autor

Figura 142 – Teste com 20 locutores: Locutor 18, Gráfico 2

Fonte: Elaborada pelo autor

Figura 143 – Teste com 20 locutores: Locutor 18, Gráfico 3

Fonte: Elaborada pelo autor

Figura 144 – Teste com 20 locutores: Locutor 18, Gráfico 4

Fonte: Elaborada pelo autor

Figura 145 – Teste com 20 locutores: Locutor 18, Gráfico 5 **Fonte: Elaborada pelo autor****Figura 146 –** Teste com 20 locutores: Locutor 18, Gráfico 6 **Fonte: Elaborada pelo autor****Figura 147 –** Teste com 20 locutores: Locutor 18, Gráfico 7 **Fonte: Elaborada pelo autor****Figura 148 –** Teste com 20 locutores: Locutor 19, Gráfico 1 **Fonte: Elaborada pelo autor****Figura 149 –** Teste com 20 locutores: Locutor 19, Gráfico 2 **Fonte: Elaborada pelo autor**

Figura 150 – Teste com 20 locutores: Locutor 19, Gráfico 3

Fonte: Elaborada pelo autor

Figura 151 – Teste com 20 locutores: Locutor 20, Gráfico 1

Fonte: Elaborada pelo autor

Figura 152 – Teste com 20 locutores: Locutor 20, Gráfico 2

Fonte: Elaborada pelo autor

Figura 153 – Teste com 30 locutores: Locutor 1, Gráfico 1

Fonte: Elaborada pelo autor

Figura 154 – Teste com 30 locutores: Locutor 1, Gráfico 2

Fonte: Elaborada pelo autor

Figura 155 – Teste com 30 locutores: Locutor 1, Gráfico 3

Fonte: Elaborada pelo autor

Figura 156 – Teste com 30 locutores: Locutor 1, Gráfico 4

Fonte: Elaborada pelo autor

Figura 157 – Teste com 30 locutores: Locutor 1, Gráfico 5

Fonte: Elaborada pelo autor

Figura 158 – Teste com 30 locutores: Locutor 1, Gráfico 6

Fonte: Elaborada pelo autor

Figura 159 – Teste com 30 locutores: Locutor 2, Gráfico 1

Fonte: Elaborada pelo autor

Figura 160 – Teste com 30 locutores: Locutor 2, Gráfico 2

Fonte: Elaborada pelo autor

Figura 161 – Teste com 30 locutores: Locutor 2, Gráfico 3

Fonte: Elaborada pelo autor

Figura 162 – Teste com 30 locutores: Locutor 3, Gráfico 1

Fonte: Elaborada pelo autor

Figura 163 – Teste com 30 locutores: Locutor 3, Gráfico 2

Fonte: Elaborada pelo autor

Figura 164 – Teste com 30 locutores: Locutor 3, Gráfico 3

Fonte: Elaborada pelo autor

Figura 165 – Teste com 30 locutores: Locutor 3, Gráfico 4

Fonte: Elaborada pelo autor

Figura 166 – Teste com 30 locutores: Locutor 4, Gráfico 1

Fonte: Elaborada pelo autor

Figura 167 – Teste com 30 locutores: Locutor 4, Gráfico 2

Fonte: Elaborada pelo autor

Figura 168 – Teste com 30 locutores: Locutor 4, Gráfico 3

Fonte: Elaborada pelo autor

Figura 169 – Teste com 30 locutores: Locutor 4, Gráfico 4

Fonte: Elaborada pelo autor

Figura 170 – Teste com 30 locutores: Locutor 4, Gráfico 5

Fonte: Elaborada pelo autor

Figura 171 – Teste com 30 locutores: Locutor 5, Gráfico 1

Fonte: Elaborada pelo autor

Figura 172 – Teste com 30 locutores: Locutor 5, Gráfico 2

Fonte: Elaborada pelo autor

Figura 173 – Teste com 30 locutores: Locutor 5, Gráfico 3

Fonte: Elaborada pelo autor

Figura 174 – Teste com 30 locutores: Locutor 5, Gráfico 4

Fonte: Elaborada pelo autor

Figura 175 – Teste com 30 locutores: Locutor 5, Gráfico 5

Fonte: Elaborada pelo autor

Figura 176 – Teste com 30 locutores: Locutor 6, Gráfico 1

Fonte: Elaborada pelo autor

Figura 177 – Teste com 30 locutores: Locutor 6, Gráfico 2

Fonte: Elaborada pelo autor

Figura 178 – Teste com 30 locutores: Locutor 6, Gráfico 3

Fonte: Elaborada pelo autor

Figura 179 – Teste com 30 locutores: Locutor 7, Gráfico 1

Fonte: Elaborada pelo autor

Figura 180 – Teste com 30 locutores: Locutor 7, Gráfico 2

Fonte: Elaborada pelo autor

Figura 181 – Teste com 30 locutores: Locutor 7, Gráfico 3

Fonte: Elaborada pelo autor

Figura 182 – Teste com 30 locutores: Locutor 7, Gráfico 4

Fonte: Elaborada pelo autor

Figura 183 – Teste com 30 locutores: Locutor 7, Gráfico 5

Fonte: Elaborada pelo autor

Figura 184 – Teste com 30 locutores: Locutor 7, Gráfico 6

Fonte: Elaborada pelo autor

Figura 185 – Teste com 30 locutores: Locutor 7, Gráfico 7

Fonte: Elaborada pelo autor

Figura 186 – Teste com 30 locutores: Locutor 7, Gráfico 8

Fonte: Elaborada pelo autor

Figura 187 – Teste com 30 locutores: Locutor 8, Gráfico 1

Fonte: Elaborada pelo autor

Figura 188 – Teste com 30 locutores: Locutor 8, Gráfico 2

Fonte: Elaborada pelo autor

Figura 189 – Teste com 30 locutores: Locutor 8, Gráfico 3

Fonte: Elaborada pelo autor

Figura 190 – Teste com 30 locutores: Locutor 8, Gráfico 4

Fonte: Elaborada pelo autor

Figura 191 – Teste com 30 locutores: Locutor 8, Gráfico 5

Fonte: Elaborada pelo autor

Figura 192 – Teste com 30 locutores: Locutor 9, Gráfico 1

Fonte: Elaborada pelo autor

Figura 193 – Teste com 30 locutores: Locutor 9, Gráfico 2

Fonte: Elaborada pelo autor

Figura 194 – Teste com 30 locutores: Locutor 9, Gráfico 3

Fonte: Elaborada pelo autor

Figura 195 – Teste com 30 locutores: Locutor 9, Gráfico 4

Fonte: Elaborada pelo autor

Figura 196 – Teste com 30 locutores: Locutor 9, Gráfico 5

Fonte: Elaborada pelo autor

Figura 197 – Teste com 30 locutores: Locutor 9, Gráfico 6

Fonte: Elaborada pelo autor

Figura 198 – Teste com 30 locutores: Locutor 9, Gráfico 7

Fonte: Elaborada pelo autor

Figura 199 – Teste com 30 locutores: Locutor 9, Gráfico 8

Fonte: Elaborada pelo autor

Figura 200 – Teste com 30 locutores: Locutor 10, Gráfico 1

Fonte: Elaborada pelo autor

Figura 201 – Teste com 30 locutores: Locutor 10, Gráfico 2

Fonte: Elaborada pelo autor

Figura 202 – Teste com 30 locutores: Locutor 10, Gráfico 3

Fonte: Elaborada pelo autor

Figura 203 – Teste com 30 locutores: Locutor 10, Gráfico 4

Fonte: Elaborada pelo autor

Figura 204 – Teste com 30 locutores: Locutor 10, Gráfico 5

Fonte: Elaborada pelo autor

Figura 205 – Teste com 30 locutores: Locutor 10, Gráfico 6

Fonte: Elaborada pelo autor

Figura 206 – Teste com 30 locutores: Locutor 10, Gráfico 7

Fonte: Elaborada pelo autor

Figura 207 – Teste com 30 locutores: Locutor 11, Gráfico 1

Fonte: Elaborada pelo autor

Figura 208 – Teste com 30 locutores: Locutor 12, Gráfico 1

Fonte: Elaborada pelo autor

Figura 209 – Teste com 30 locutores: Locutor 12, Gráfico 2

Fonte: Elaborada pelo autor

Figura 210 – Teste com 30 locutores: Locutor 12, Gráfico 3

Fonte: Elaborada pelo autor

Figura 211 – Teste com 30 locutores: Locutor 12, Gráfico 4

Fonte: Elaborada pelo autor

Figura 212 – Teste com 30 locutores: Locutor 12, Gráfico 5

Fonte: Elaborada pelo autor

Figura 213 – Teste com 30 locutores: Locutor 13, Gráfico 1

Fonte: Elaborada pelo autor

Figura 214 – Teste com 30 locutores: Locutor 13, Gráfico 2

Fonte: Elaborada pelo autor

Figura 215 – Teste com 30 locutores: Locutor 14, Gráfico 1

Fonte: Elaborada pelo autor

Figura 216 – Teste com 30 locutores: Locutor 14, Gráfico 2

Fonte: Elaborada pelo autor

Figura 217 – Teste com 30 locutores: Locutor 14, Gráfico 3

Fonte: Elaborada pelo autor

Figura 218 – Teste com 30 locutores: Locutor 14, Gráfico 4

Fonte: Elaborada pelo autor

Figura 219 – Teste com 30 locutores: Locutor 14, Gráfico 5

Fonte: Elaborada pelo autor

Figura 220 – Teste com 30 locutores: Locutor 15, Gráfico 1

Fonte: Elaborada pelo autor

Figura 221 – Teste com 30 locutores: Locutor 15, Gráfico 2

Fonte: Elaborada pelo autor

Figura 222 – Teste com 30 locutores: Locutor 15, Gráfico 3

Fonte: Elaborada pelo autor

Figura 223 – Teste com 30 locutores: Locutor 15, Gráfico 4

Fonte: Elaborada pelo autor

Figura 224 – Teste com 30 locutores: Locutor 16, Gráfico 1

Fonte: Elaborada pelo autor

Figura 225 – Teste com 30 locutores: Locutor 16, Gráfico 2

Fonte: Elaborada pelo autor

Figura 226 – Teste com 30 locutores: Locutor 16, Gráfico 3

Fonte: Elaborada pelo autor

Figura 227 – Teste com 30 locutores: Locutor 17, Gráfico 1

Fonte: Elaborada pelo autor

Figura 228 – Teste com 30 locutores: Locutor 17, Gráfico 2

Fonte: Elaborada pelo autor

Figura 229 – Teste com 30 locutores: Locutor 17, Gráfico 3

Fonte: Elaborada pelo autor

Figura 230 – Teste com 30 locutores: Locutor 17, Gráfico 4

Fonte: Elaborada pelo autor

Figura 231 – Teste com 30 locutores: Locutor 18, Gráfico 1

Fonte: Elaborada pelo autor

Figura 232 – Teste com 30 locutores: Locutor 18, Gráfico 2

Fonte: Elaborada pelo autor

Figura 233 – Teste com 30 locutores: Locutor 18, Gráfico 3

Fonte: Elaborada pelo autor

Figura 234 – Teste com 30 locutores: Locutor 19, Gráfico 1

Fonte: Elaborada pelo autor

Figura 235 – Teste com 30 locutores: Locutor 19, Gráfico 2

Fonte: Elaborada pelo autor

Figura 236 – Teste com 30 locutores: Locutor 19, Gráfico 3

Fonte: Elaborada pelo autor

Figura 237 – Teste com 30 locutores: Locutor 21, Gráfico 1

Fonte: Elaborada pelo autor

Figura 238 – Teste com 30 locutores: Locutor 21, Gráfico 2

Fonte: Elaborada pelo autor

Figura 239 – Teste com 30 locutores: Locutor 21, Gráfico 3

Fonte: Elaborada pelo autor

Figura 240 – Teste com 30 locutores: Locutor 22, Gráfico 1

Fonte: Elaborada pelo autor

Figura 241 – Teste com 30 locutores: Locutor 22, Gráfico 2

Fonte: Elaborada pelo autor

Figura 242 – Teste com 30 locutores: Locutor 23, Gráfico 1

Fonte: Elaborada pelo autor

Figura 243 – Teste com 30 locutores: Locutor 23, Gráfico 2

Fonte: Elaborada pelo autor

Figura 244 – Teste com 30 locutores: Locutor 23, Gráfico 3

Fonte: Elaborada pelo autor

Figura 245 – Teste com 30 locutores: Locutor 23, Gráfico 4

Fonte: Elaborada pelo autor

Figura 246 – Teste com 30 locutores: Locutor 23, Gráfico 5

Fonte: Elaborada pelo autor

Figura 247 – Teste com 30 locutores: Locutor 23, Gráfico 6

Fonte: Elaborada pelo autor

Figura 248 – Teste com 30 locutores: Locutor 24, Gráfico 1

Fonte: Elaborada pelo autor

Figura 249 – Teste com 30 locutores: Locutor 24, Gráfico 2

Fonte: Elaborada pelo autor

Figura 250 – Teste com 30 locutores: Locutor 24, Gráfico 3

Fonte: Elaborada pelo autor

Figura 251 – Teste com 30 locutores: Locutor 24, Gráfico 4

Fonte: Elaborada pelo autor

Figura 252 – Teste com 30 locutores: Locutor 24, Gráfico 5

Fonte: Elaborada pelo autor

Figura 253 – Teste com 30 locutores: Locutor 25, Gráfico 1

Fonte: Elaborada pelo autor

Figura 254 – Teste com 30 locutores: Locutor 25, Gráfico 2

Fonte: Elaborada pelo autor

Figura 255 – Teste com 30 locutores: Locutor 25, Gráfico 3

Fonte: Elaborada pelo autor

Figura 256 – Teste com 30 locutores: Locutor 25, Gráfico 4

Fonte: Elaborada pelo autor

Figura 257 – Teste com 30 locutores: Locutor 25, Gráfico 5

Fonte: Elaborada pelo autor

Figura 258 – Teste com 30 locutores: Locutor 26, Gráfico 1

Fonte: Elaborada pelo autor

Figura 259 – Teste com 30 locutores: Locutor 26, Gráfico 2

Fonte: Elaborada pelo autor

Figura 260 – Teste com 30 locutores: Locutor 26, Gráfico 3

Fonte: Elaborada pelo autor

Figura 261 – Teste com 30 locutores: Locutor 26, Gráfico 4

Fonte: Elaborada pelo autor

Figura 262 – Teste com 30 locutores: Locutor 26, Gráfico 5

Fonte: Elaborada pelo autor

Figura 263 – Teste com 30 locutores: Locutor 27, Gráfico 1

Fonte: Elaborada pelo autor

Figura 264 – Teste com 30 locutores: Locutor 27, Gráfico 2

Fonte: Elaborada pelo autor

Figura 265 – Teste com 30 locutores: Locutor 27, Gráfico 3

Fonte: Elaborada pelo autor

Figura 266 – Teste com 30 locutores: Locutor 28, Gráfico 1

Fonte: Elaborada pelo autor

Figura 267 – Teste com 30 locutores: Locutor 28, Gráfico 2

Fonte: Elaborada pelo autor

Figura 268 – Teste com 30 locutores: Locutor 28, Gráfico 3

Fonte: Elaborada pelo autor

Figura 269 – Teste com 30 locutores: Locutor 29, Gráfico 1

Fonte: Elaborada pelo autor

Figura 270 – Teste com 30 locutores: Locutor 29, Gráfico 2

Fonte: Elaborada pelo autor

Figura 271 – Teste com 30 locutores: Locutor 29, Gráfico 3

Fonte: Elaborada pelo autor

Figura 272 – Teste com 30 locutores: Locutor 29, Gráfico 4

Fonte: Elaborada pelo autor

Figura 273 – Teste com 30 locutores: Locutor 29, Gráfico 5

Fonte: Elaborada pelo autor

Figura 274 – Teste com 30 locutores: Locutor 29, Gráfico 6

Fonte: Elaborada pelo autor

Figura 275 – Teste com 30 locutores: Locutor 29, Gráfico 7

Fonte: Elaborada pelo autor

Figura 276 – Teste com 30 locutores: Locutor 30, Gráfico 1

Fonte: Elaborada pelo autor

Figura 277 – Teste com 30 locutores: Locutor 30, Gráfico 2

Fonte: Elaborada pelo autor