

Universidade Federal de Campina Grande

Centro de Engenharia Elétrica e Informática

Coordenação de Pós-Graduação em Ciência da Computação

Avaliação do uso de coeficientes Mel-
Cepstrais na Representação das
Características Vocais de um Locutor

Fabício Gutemberg Lélis de Melo

Campina Grande, Paraíba, Brasil

Dezembro - 2014

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Avaliação do uso de Coeficientes Mel- Cepstrais na Representação das Características Vocais de um Locutor

Fabrício Gutemberg Lélis de Melo

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Redes de Computadores e Sistemas
Distribuídos

Elmar Uwe Kurt Melcher (Orientador)

Joseana Macêdo Fachine Régis de Araújo (Orientadora)

Campina Grande, Paraíba, Brasil

Dezembro – 2014

M528a Melo, Fabrício Gutemberg Lélis de.

Avaliação do uso de coeficientes mel-cepstrais na representação das características vocais de um locutor / Fabrício Gutemberg Lélis de Melo. – Campina Grande, 2014.

114 f.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2014.

"Orientação: Prof. Dr. Elmar Uwe Kurt Melcher, Prof.^a Dr.^a Joseana Macêdo Fachine Régis de Araújo".

Referências.

1. Hardware. 2. Reconhecimento de Locutor. 3. Mel-cepstrais. 4. FPGA.
I. Melcher, Elmar Uwe Kurt. II. Araújo, Joseana Macêdo Fachine Régis de.
III. Título.

CDU 004.3(043)

**"AVALIAÇÃO DO USO DE COEFICIENTES MEL-CEPTRAIS NA REPRESENTAÇÃO
DAS CARACTERÍSTICAS VOCAIS DE UM LOCUTOR"**

FABRICIO GUTEMBERG LELIS DE MELO

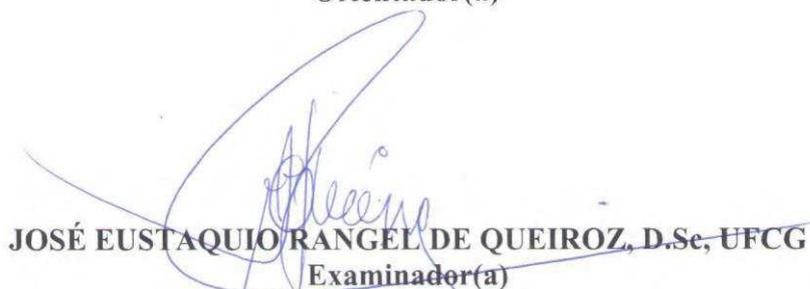
DISSERTAÇÃO APROVADA EM 01/12/2014



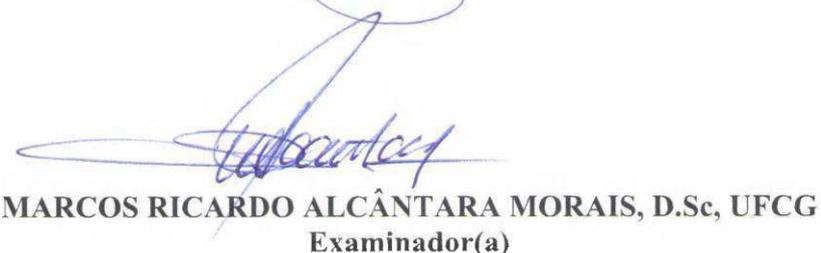
**JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, D.Sc, UFCG
Orientador(a)**



**ELMAR UWE KURT MELCHER, Dr., UFCG
Orientador(a)**



**JOSÉ EUSTAQUIO RANGEL DE QUEIROZ, D.Sc, UFCG
Examinador(a)**



**MARCOS RICARDO ALCÂNTARA MORAIS, D.Sc, UFCG
Examinador(a)**

CAMPINA GRANDE - PB

Resumo

A identificação de indivíduos por meio de biometria vem sendo bastante usada como mecanismo de segurança para o acesso a sistemas computacionais ou ambientes restritos. Os sistemas biométricos têm sido desenvolvidos para realizar a identificação por impressão digital, íridia ou vocal, por exemplo. Usar a voz como meio para a autenticação individual tem sido cada vez mais possível, devido ao avanço significativo na área de Processamento Digital de Sinais de voz. Esta pesquisa tem como finalidade avaliar a eficiência dos coeficientes mel-cepstrais na representação das características de um locutor em um sistema automático de verificação de locutor. As técnicas utilizadas para a construção do sistema automático de verificação de locutor, visando a uma implementação em *hardware*, incluem o uso de: (i) coeficientes mel-cepstrais, na composição do vetor de características; (ii) quantização vetorial, na obtenção de padrões; e (iii) uma regra de decisão, baseada na distância Euclidiana. O sistema utilizado para a avaliação da representação das características vocais de um locutor é uma modificação de outro sistema automático de verificação de locutor que utiliza coeficientes LPC para a representação das características vocais de um locutor. Para tanto, fez-se uso das linguagens C++ (fase de treinamento) e SystemVerilog (fase de verificação). Os resultados utilizando coeficientes mel-cepstrais foram de 99,34% na taxa de acerto, 0,17% para taxa de erros e 0,49% na taxa de respostas desconhecidas, comparados, respectivamente, a 96,52% na taxa de acerto, 0,90% para taxa de erros e 2,58% na taxa de desconhecidos para coeficientes LPC.

Abstract

Biometric identification of individuals has been widely used as a security mechanism for accessing computer systems or restricted environments. Biometric systems have been developed to perform identification through fingerprint, iris, or voice, for example. Using the voice as a biometric identifier has been increasingly possible due to significant advances in digital processing of speech signals area. This research aims to evaluate the efficiency of mel-frequency cepstral coefficients in the representation of the characteristics of a speaker in an automatic speaker verification. The techniques used to construct the automatic speaker verification system aiming at a hardware implementation included the use of: (i) mel-frequency cepstral coefficients, like feature vector; (ii) vector quantization, in patterning modelling; and (iii) a decision rule, based on Euclidean distance. The system used for evaluation in the representation of the characteristics of a speaker is a modification of another automatic speaker verification system using linear predictive coding coefficients for the representation of the vocal characteristics of a speaker. It was implemented using C++ for the training phase, and SystemVerilog for the verification phase. The results using mel-frequency cepstral coefficients were 99.34% in the hit rate, 0.17% to error rate and 0.49% to unknown response rate, compared respectively to 96.52% in success rate, 0.90% to error rate and 2.58% to unknown rate using the linear predictive coding coefficients.

Agradecimentos

Agradeço, em primeiro lugar, a Deus, por sempre estar presente a meu lado e guiar meus passos em todos os dias de minha vida.

Agradeço a meus pais, Targino (*in memoriam*) e Solange, pois a herança mais importante que eles tentaram e continuam tentando deixar tem sido a educação, principalmente minha mãe, que tem feito o papel de mãe e pai em minha vida durante os últimos catorze anos.

Agradeço também a meus avós, por serem exemplos para mim.

Agradeço aos professores Elmar e Joseana, por acreditarem e ajudarem na realização desta pesquisa, sempre com paciência e uma palavra amiga e por contribuírem para minha vida acadêmica e pessoal.

Agradeço aos integrantes do LAD (Leonardo, Nelson, Paulo, Plateny, Thiago e Victor) e do BELADIA (Adalberto, Aninha, Matheus e Samir) que contribuíram com seus tempos e esforços para a realização dos experimentos desta pesquisa e que também proporcionaram contar com a colaboração de outras pessoas (Carla, Erika, Jennifer, Maria Aline, Nathalie, Raissa e Samara) na realização dos experimentos.

Quero agradecer aos professores Heron e Roberto Faria, às irmãs Aline e Alana, ao aluno de doutorado Roberto e à esposa do professor Elmar, Joseilda (Jó), que se disponibilizaram a participar da gravação dos experimentos.

Não posso me esquecer de agradecer também aos(às) amigo(a)s/irmão(ã)s - Danielle, Janine, Laércio, Lenilson, Maria de Lourdes, Michele, Priscylla e Thais, com os quais tive o prazer de compartilhar a presença em vários momentos de minha vida, inclusive neste momento de mais uma conquista. Aos meus irmãos da família Unidos pela Esperança, painho Ricardo e mainha Luciene.

Agradeço a todos da COPIN, pela ajuda prestada durante os anos de realização desta pesquisa.

Agradeço, enfim, ao CNPq, pela concessão da bolsa de estudos.

Sumário

Lista de Siglas e Abreviaturas	x
Lista de Figuras	xi
Lista das Tabelas	xii
1 Introdução	1
1.1 Motivação	4
1.2 Objetivos	5
1.1.1 Objetivo Geral	5
1.1.2 Objetivos Específicos	6
1.3 Estrutura	6
2 Fundamentação Teórica	7
2.1 Pesquisas Relacionadas	7
2.2 Reconhecimento de Padrões de Fala	10
2.2.1 Aquisição	12
2.2.2 Pré-processamento	12
2.2.2.1 Normalização	13
2.2.2.2 Detecção de voz	13
2.2.2.3 Pré-ênfase	14
2.2.2.4 Segmentação e Janelamento	14
2.2.3 Extração de características	15
2.2.3.1 Coeficientes LPC	17
2.2.3.2 Coeficientes Mel-Cepstrais	19
2.2.4 Geração de Padrões	23
2.2.4.1 Quantização Vetorial	24
2.3 Reconhecimento de Locutor em Sistemas Embarcados	25
2.4 Processo de Desenvolvimento de um Sistema Embarcado	27
2.4.1 Especificação do <i>Hardware</i>	27
2.4.2 Especificação da Verificação Funcional	28
2.4.3 Implementação do <i>testbench</i>	28
2.4.4 Implementação do DUV	28
2.4.5 Verificação Funcional	28
2.4.6 Síntese	29
2.4.7 Simulação pós-síntese	29
2.4.8 Prototipação	29
3 Adequação do SPVR para uso de Coeficientes Mel-Cepstrais	30

3.1	Detecção de voz (VD)	31
3.2	Pré-ênfase (PE)	33
3.3	Segmentação e Janelamento (WIN)	33
	Extração de Características (FILTER e MFCC)	34
3.4	Quantização Vetorial (QV)	34
3.5	Comparação (PM)	35
3.6	Regra de Decisão (DM)	35
4	Apresentação e Análise dos Resultados	37
4.1	Metodologia	37
4.1.1	Base de Dados	37
4.1.2	Detecção do Limiars do VD	38
4.1.3	Simulação dos dados	39
4.2	Avaliação dos Experimentos	40
4.2.1	Resultado dos Limiars do Sistema	41
4.2.2	Análise Comparativa entre MFCC e LPC	43
4.2.3	Modelo para implementação do módulo MFCC em FPGA	48
5	Considerações Finais e Sugestões para Pesquisas Futuras	56
5.1	Contribuições	56
5.2	Sugestões para Pesquisas Futuras	57
6	Referências Bibliográficas	58
Apêndice A	Resultados do Processamento dos dados	63
Apêndice B	Análise Estatística	72
Apêndice C	Resultados das Saídas como Desconhecido	80
Apêndice D	Resultados dos Experimentos para o Melhor Caso	84
Apêndice E	Matrizes de Confusão	93
Apêndice F	Verificação Funcional Preliminar	99
Anexo A	Lista das Frases Foneticamente Balanceadas	102
Anexo B	Relatórios Gerados pela Ferramenta <i>dc_shell</i>	103

Lista de Siglas e Abreviaturas

A/D	Analógico/Digital
ASIC	<i>Application Specific Integrated Circuits</i>
CDE	Comparação por Distância Euclidiana
DCT	<i>Discrete Cosine Transform</i>
DM	<i>Decision Maker</i>
DSP	<i>Digital Signal Processor</i>
DTW	<i>Dynamic Time Warping</i>
DV	Detector de Voz
EF	Energia Final
EI	Energia Inicial
FFT	<i>Fast Fourier Transform</i>
FIR	<i>Finite Impulse Response</i>
FPGA	<i>Field Programmable Gate Array</i>
GPS	<i>Global Positioning System</i>
HMM	<i>Hidden Markov Model</i>
LBG	<i>Linde-Buzo-Gray</i>
LI	Limiar Inferior
LPC	<i>Linear Prediction Coding</i>
LS	Limiar Superior
LUT	<i>Look-Up Table</i>
MFCC	<i>Mel-Frequency Cepstral Coefficients</i>
PDSV	Processamento Digital de Sinais de Voz
PE	Pré-ênfase
PM	<i>Parting Matching</i>
QA	Quantidade de Acertos
QD	Quantidade de Desconhecidos
QE	Quantidade de Erros
QV	Quantização Vetorial
RD	Regra de Decisão
SRAM	<i>Static Random-Access Memory</i>
SRF	Sistemas de Reconhecimento de Fala
SRL	Sistemas de Reconhecimento de Locutor
SRV	Sistemas de Resposta Vocal
SVM	<i>Support Vector Machines</i>
TF	Tempo final
TI	Tempo inicial
VLSI	<i>Very Large Scale Integration</i>
WIN	Windowing

Lista de Figuras

Figura 1.1: Classificação geral da área de processamento digital de sinais de voz.	2
Figura 2.1: Representação da tarefa de reconhecimento de padrões.....	10
Figura 2.2: Modelo simplificado para produção de voz baseado no modelo LPC.	16
Figura 2.3: Mapeamento da frequência linear na escala Mel.	20
Figura 2.4: Obtenção dos coeficientes MFCC a partir de banco de filtros.	21
Figura 2.5: Espectro do banco de filtros triangulares de Slaney (1998).	22
Figura 2.6: Fases do processo de desenvolvimento de hardware na ordem proposta por VeriSC. .	27
Figura 3.1: Diagrama em blocos do sistema de verificação de locutor.	30
Figura 3.2: Algoritmo de detecção de início e fim de sentenças.	32
Figura 4.1: Modelagem do extrator da magnitude auditiva.	49
Figura 4.2: Design para extração dos coeficientes MFCC.	52
Figura B.1: Boxplot dos experimentos com as maiores quantidades de erros.	73
Figura B.2: Boxplot dos experimentos com as menores quantidades de erros.	74
Figura B.3: Comparação dos experimentos com extração de características mel-cepstrais sob a influência da variação de TI na métrica QE.	76
Figura B.4: Comparação dos experimentos com extração de características mel-cepstrais sob a influência da variação de TF na métrica QE.	77
Figura B.5: Comparação dos experimentos com extração de características LPC sob a influência da variação de TI na métrica QE.	78
Figura B.6: Comparação dos experimentos com extração de características LPC sob a influência da variação de TF na métrica QE.	79

Lista das Tabelas

Tabela 2.1: Resumo das Pesquisas Relacionadas.....	10
Tabela 4.1: Características da gravação.....	38
Tabela 4.2: Dados estatísticos da energia do conjunto de treinamento.	39
Tabela 4.3: Limiares de energia e de tempo.	39
Tabela 4.4: Descrição dos Experimentos.	40
Tabela 4.5: Limiares de tempo que minimizam QE.	41
Tabela 4.6: Resultados para variação dos limiares de tomada de decisão usando coeficientes MFCC.	42
Tabela 4.7: Índices dos resultados para a melhor configuração de limiares.	43
Tabela 4.8: Índices de falsa aceitação e Repetição da Própria Sentença.	43
Tabela 4.9: Matriz de confusão para a menor quantidade de erros com mel-cepstrais.....	45
Tabela 4.10: Matriz de confusão para a menor quantidade de erros com LPC.	46
Tabela 4.11: Relatório sobre a área dos blocos.	54
Tabela 4.12: Estimativa de consumo dos blocos.....	55
Tabela A.1: Resultados do Processamento 1.....	64
Tabela A.2: Resultados do Processamento 2.....	65
Tabela A.3: Resultados do Processamento 2.....	66
Tabela A.4: Resultados do Processamento 2.....	67
Tabela A.5: Resultados do Processamento 2.....	68
Tabela A.6: Resultados do Processamento 2.....	69
Tabela A.7: Resultados do Processamento 2.....	70
Tabela A.8: Resultados do Processamento 2.....	71
Tabela B.1: Medidas resumo da métrica QE para os 8 experimentos.....	72
Tabela C.1: Resultados do Experimento E1.	80
Tabela C.2: Resultados do Experimento E2.	81
Tabela C.3: Resultados do Experimento E3.	82
Tabela C.4: Resultados do Experimento E4.	83
Tabela E.1: Matriz de confusão para experimento E2 com a menor quantidade de erros.	93
Tabela E.2: Matriz de confusão para experimento E3 com a menor quantidade de erros.	94
Tabela E.3: Matriz de confusão para experimento E4 com a menor quantidade de erros.	95
Tabela E.4: Matriz de confusão para experimento E5 com a menor quantidade de erros.	96
Tabela E.5: Matriz de confusão para experimento E6 com a menor quantidade de erros.	97
Tabela E.6: Matriz de confusão para experimento E8 com a menor quantidade de erros.	98
Tabela AX.1: Lista das sentenças escolhidas.....	989
Tabela BX.1: Relatório de consumo e área do módulo FFT.....	100
Tabela BX.2: Relatório de consumo e área do módulo FILTER.....	103
Tabela BX.3: Relatório de consumo e área do módulo MFCC.....	106
Tabela BX.4: Relatório de consumo e área do módulo LPCA.....	109

1 Introdução

A utilização de características biológicas está sendo amplamente aceita em segurança de sistemas, pois senhas e cartões de identificação podem ser perdidos ou roubados. Aplicações tais, como transações financeiras ou gerenciamento de segurança em redes já estão sendo beneficiadas com o uso da biometria. Dentre as diferentes formas de uso de características biométricas (reconhecimento da face, da íris e de impressão digital, dentre outras), a verificação da identidade vocal de um locutor apresenta uma maneira precisa e eficiente de autenticar a identidade de um indivíduo por meio da análise vocal (EHKAN; ALLEN; QUIGLEY, 2011; RAMOS-LARA et al., 2012).

As primeiras publicações sobre reconhecimento de locutor surgiram na década de 1950 (POLLACK; PICKETT; SUMBY, 1954; SHEARME; HOLMES, 1959). O sinal de voz do locutor era examinado e apresentado a um especialista humano que tomava a decisão sobre a identidade do indivíduo por meio da comparação entre as representações. Desde então, a tecnologia de reconhecimento de locutor tem evoluído significativamente, tornando-se atualmente uma característica biométrica que oferece elevados níveis de confiança e proteção, o que a torna uma solução promissora para o mercado de segurança e verificação pessoal (RAMOS-LARA et al., 2012).

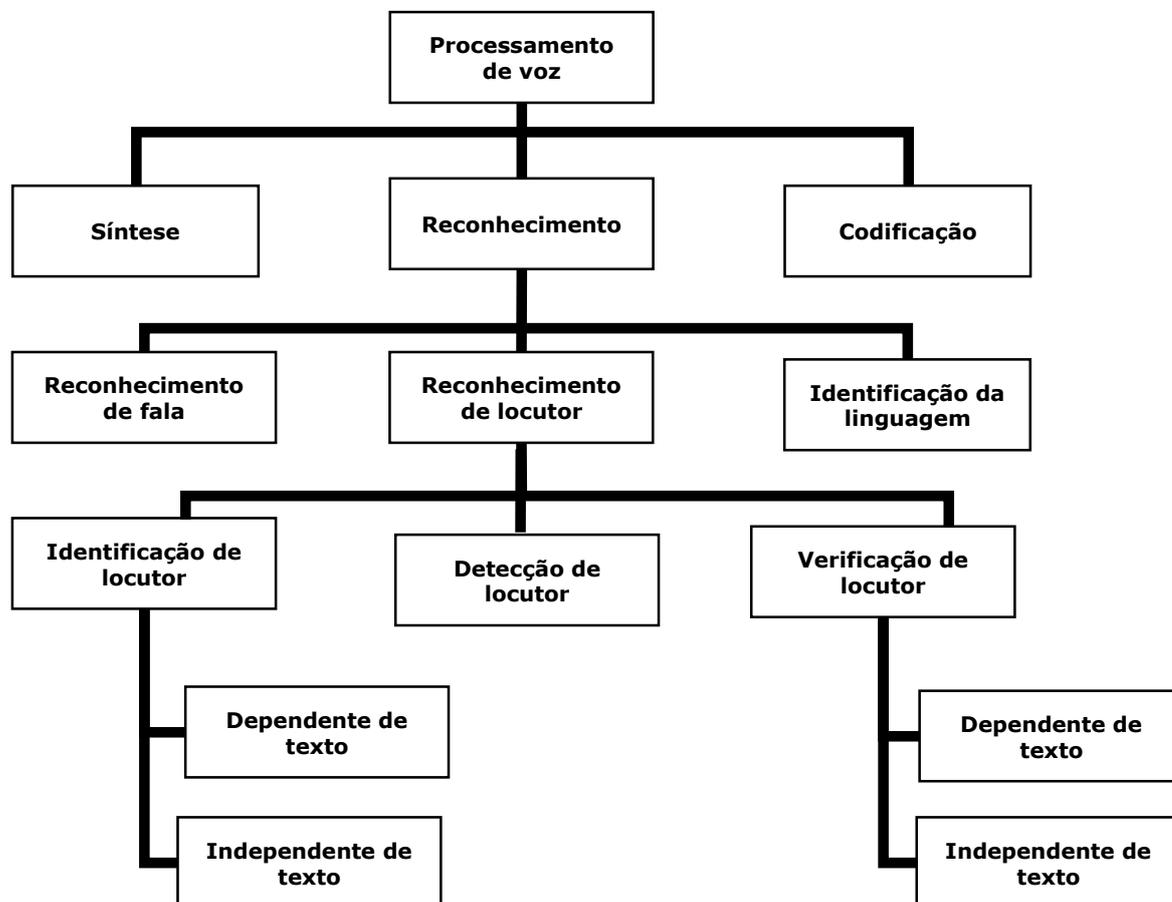
Reconhecimento de locutor é uma área de pesquisa inserida em uma área mais abrangente denominada Processamento Digital de Sinais de Voz (PDSV). A área de PDSV pode ser dividida em três grandes subáreas:

1. Resposta Vocal (síntese);
2. Reconhecimento de Fala e
3. Reconhecimento de Locutor.

Na Figura 1.1, apresenta-se uma classificação geral do processamento de voz, com ênfase no reconhecimento de locutor, objeto de estudo da pesquisa ora descrita (CAMPBELL, 1997).

Figura 1.1: Classificação geral da área de processamento digital de sinais de voz.

Fonte: Adaptado de (CAMPBELL, 1997).



Sistemas de Resposta Vocal (SRV) ou sistemas de síntese de voz são projetados para responder a um pedido de informação utilizando mensagens faladas. Nesses sistemas, a comunicação vocal se faz no sentido homem-máquina (RABINER e SCHAFFER, 1978).

Esses sistemas utilizam sinais de voz armazenados, tais como frases, palavras, fonemas ou segmentos curtos, para compor a mensagem de saída, que é formada pela concatenação dos sinais armazenados, com base em regras linguísticas dos idiomas (regras gramaticais e fonéticas).

Os sistemas de reconhecimento de fala (SRF) têm por objetivo reconhecer uma determinada elocução de uma sentença ou "entender" um texto falado (CHEN, 2005).

Os sistemas de SRF são comumente usados no atendimento telefônico automático, no acesso a menus de celulares ou de outros equipamentos

eletrônicos, na transcrição de fala para texto e em aplicações de comando-controle, dentre outros.

Os sistemas de reconhecimento de locutor (SRL), por sua vez, têm como objetivo reconhecer um locutor por meio da voz. Esses sistemas são classificados como sendo de verificação ou identificação de locutor.

Os sistemas de verificação (validação ou autenticação) irão “confirmar” se o locutor é quem alega ser (LI, 2012). O funcionamento típico de um sistema de verificação de locutor, a exemplo de outros sistemas de reconhecimento de padrões, é dividido em dois cenários: treinamento e teste. O treinamento é a etapa em que o locutor, antes de usar o sistema, precisa cadastrar o modelo vocal representativo do seu sinal de voz, junto com um identificador para esse modelo, que pode ser um número. O locutor repetirá a sentença algumas vezes, permitindo ao sistema construir o modelo com suas características vocais representativas. Na etapa de teste, o locutor fornece sua identificação digitando um número ou outro identificador. Após isso, o sistema solicita que o usuário pronuncie sua sentença. Então, a sentença enunciada é comparada com o modelo armazenado. A tomada de decisão é baseada em dois limiares: um superior e outro inferior. Quando o valor de distância fica abaixo do limiar inferior, o locutor é aceito, ou seja, o locutor é realmente quem ele alega ser. Entretanto, se o valor de distância fica acima do limiar superior, o locutor é rejeitado. Por fim, se o valor de distância ficar entre os dois limiares, é dito que o locutor é desconhecido e então o mesmo repete sua sentença.

Os sistemas de identificação de locutor também são divididos em duas fases: treinamento e identificação. O treinamento desses sistemas ocorre de maneira semelhante àquele dos sistemas de verificação. No entanto, não existe a necessidade de cadastrar um identificador para o locutor. Na fase de identificação, o sistema compara a sentença de teste com todos os modelos vocais armazenados e fornece como resposta o modelo vocal que mais se assemelha aos dados de teste fornecidos.

Os SRL são úteis na realização de operações de autenticação por voz nas áreas de segurança e criminalística. Na área de segurança, esses sistemas atuam na restrição ao acesso, à informação confidencial ou a conteúdo, por exemplo. Na área de criminalística, auxiliam no reconhecimento de indivíduos, uma vez que, as características vocais são únicas para cada indivíduo (BEIGI, 2011; LI, 2012).

1.1 Motivação

A autenticação tem sido usada por seres humanos há milhares de anos para reconhecer uns aos outros, para identificar aliados e inimigos e para proteger informações. Na era do computador, o objetivo da identificação é mais que apenas identificar indivíduos de maneira presencial, mas também identificar indivíduos em locais remotos, computadores em uma rede ou qualquer entidade em uma rede de computadores. Para tanto, a autenticação tem deixado de ser apenas um processo manual, tornando-se um processo automático. Com o visível aumento da preocupação com as questões de segurança e privacidade, a autenticação é um processo cada vez mais comum no dia a dia (LI, 2012).

A verificação de locutor é um método de autenticação que faz uso de características biológicas e que pode ser utilizado em várias situações da vida real. Algumas transações comerciais realizadas por telefone utilizam a verificação de locutor por fornecer a baixo custo segurança e proteção contra a identificação fraudulenta. Além de permitir a autenticação de indivíduos em locais remotos, outra vantagem de utilizar a verificação de locutor é o uso de sensores de baixo custo para a captura da voz (RAMOS-LARA et al., 2012). O presente estudo está inserido na área de verificação de locutor.

Por causa da complexidade presente nos algoritmos dos sistemas biométricos e da necessidade de funcionarem em tempo real, na maioria das aplicações, as implementações desses sistemas geralmente são realizadas em computadores pessoais equipados com microprocessadores de alto desempenho. Processadores de propósito geral contêm unidades de ponto flutuante capazes de realizar milhões de operações por segundo, em frequências na faixa dos GHz, permitindo a resolução de algoritmos complexos em algumas centenas de milissegundos. Entretanto, este tipo de solução pode não ser aceitável em sistemas de baixo custo, pois fatores tais como preço, consumo de energia e tamanho determinam a viabilidade do produto.

Para atender aos requisitos que viabilizam os sistemas de baixo custo, uma possível solução são os ASIC¹ (Application Specific Integrated Circuits) que proporcionam melhor desempenho em relação ao tempo de execução, quando

¹ Circuito projetado para realizar uma aplicação específica em domínios bem definidos.

comparados com microprocessadores de baixo custo ou dispositivos DSP² (*Digital Signal Processor*). Além disso, como os ASIC são desenvolvidos para aplicações específicas, eles apresentam preços acessíveis para a produção em grande quantidade. Outra possível solução é o uso de FPGA (*Field Programmable Gate Array*) que, basicamente, é composta por uma matriz de blocos lógicos configuráveis, interconectados por meio de uma rede de conexões programáveis. Graças ao bom desempenho e à flexibilidade, aliados à facilidade fornecida pelas novas ferramentas de desenvolvimento, que disponibilizam bibliotecas com funções comuns aos DSP, esses requisitos tornam as FPGA úteis em aplicações que precisam da implementação de algoritmos com alta complexidade computacional (EHKAN; ALLEN; QUIGLEY, 2011; RAMOS-LARA et al., 2012; STAWORKO; RAWSKI, 2010).

Para a implementação de um projeto de *hardware* (FPGA, ASIC), rigorosas metodologias devem ser seguidas para evitar que falhas surjam somente depois que o *hardware* prototipado tenha sido integrado a algum sistema. No processo de desenvolvimento de *hardware*, a detecção de problemas funcionais e comportamentais deve ser feita na etapa de verificação funcional³ (OLIVEIRA, 2010; SILVA, 2007). Essa etapa deve ser realizada a partir da comparação de dois modelos, em um ambiente de simulação denominado *testbench*, o modelo sendo desenvolvido (*hardware*) e o modelo ideal que reflete a especificação (modelo de referência) (BERGERON, 2003).

1.2 Objetivos

A seguir, serão apresentados o objetivo geral e os objetivos específicos da pesquisa de mestrado ora descrita.

1.1.1 Objetivo Geral

A presente pesquisa objetiva avaliar a eficiência dos coeficientes mel-cepstrais na representação das características de um locutor. Esta avaliação é feita por meio

² Microprocessadores especializados em processamento digital de sinais usados para processar sinais de áudio e vídeo, por exemplo.

³ Processo usado para demonstrar que o objetivo do projeto é preservado em sua implementação (BERGERON, 2003).

de uma comparação com os coeficientes LPC para a representação das características de um locutor.

1.1.2 Objetivos Específicos

A seguir, são listados os objetivos específicos deste estudo.

- Analisar e adequar as técnicas ao reconhecimento de locutor, implementadas em *software*, visando a um modelo com configuração mais adequada a uma implementação em *hardware*, com exigência de redução do consumo energético.
- Realizar uma análise comparativa do modelo proposto com o modelo implementado por Fachine *et al.* (2010). A escolha desse estudo se deu pelo fato de estar validado em FPGA e prototipado em silício.
- Determinar de forma automática os limiares para o sistema (detecção de início e fim de sentença e tomada de decisão do sistema).
- Construir uma base de dados, com locutores masculinos e femininos, com sentenças para as fases de treinamento e de reconhecimento do sistema.

1.3 Estrutura

O restante da dissertação encontra-se organizado conforme descrição a seguir.

- Capítulo 2 – Neste capítulo, são apresentados os conceitos relacionados à tarefa de reconhecimento de locutor, com a descrição das fases de treinamento e de reconhecimento (teste) e das etapas envolvidas em cada uma delas, como também os conceitos relacionados ao processo de desenvolvimento de um sistema embarcado, além de algumas pesquisas relacionadas à verificação locutor.
- Capítulo 3 – As principais características do sistema de reconhecimento de locutor proposto são apresentadas.
- Capítulo 4 – É descrita a base de dados utilizada e também são apresentados os resultados obtidos a partir das configurações do sistema simulado nesta pesquisa.
- Capítulo 5 – Neste capítulo, são apresentadas as considerações finais e as sugestões para estudos futuros.

2 Fundamentação Teórica

O desenvolvimento de aplicações na área da comunicação vocal envolve a utilização de técnicas de Processamento Digital de Sinais de Voz. Para a aplicação adequada dessas técnicas, é importante que se compreenda como funciona a tarefa de reconhecimento de padrões de fala, bem como os fundamentos do Processamento Digital de Sinais de Voz.

Antes de ser introduzida a teoria relacionada à área de reconhecimento de locutor, na Seção 2.1 serão apresentadas pesquisas relacionadas ao estudo. Na Seção 2.2, será descrita a tarefa de reconhecimento de padrões de fala e as subtarefas que a compõe, na Seção 2.3 é apresentada uma teoria para reconhecimento de locutor em sistemas embarcados e na seção 2.4, é descrito o processo de desenvolvimento de sistemas embarcados.

2.1 Pesquisas Relacionadas

Recentemente, várias pesquisas sobre sistemas de reconhecimento de locutor foram apresentadas (CARDOSO, 2009; FECHINE *et al.*, 2010; RAMOS-LARA *et al.*, 2012; SELVAN; JOSEPH; ANISH BABU, 2013; STAWORKO; RAWSKI, 2010; WANG *et al.*, 2011). Essas pesquisas não apresentam apenas soluções que foram realizadas em *software* (CARDOSO, 2009; FECHINE, 2000; SELVAN; JOSEPH; ANISH BABU, 2013), mas também sistemas que utilizam a estratégia de combinar *software* e *hardware* para a implementação do sistema (FECHINE *et al.*, 2010; STAWORKO; RAWSKI, 2010; WANG *et al.*, 2011; RAMOS-LARA *et al.*, 2012). Uma possível abordagem para implementação dessa combinação consiste na divisão entre as fases de treinamento e reconhecimento, de maneira que o treinamento possa ser realizado em *software* e o reconhecimento em *hardware*.

Em Fachine *et al.* (2010), é apresentado um sistema de verificação de locutor que realiza a fase de treinamento em *software* e a fase de verificação em *hardware*. No referido sistema, é implementado um módulo que utiliza a energia do sinal de voz como forma de detectar o início e fim das sentenças. A extração de características é feita utilizando-se os coeficientes LPC e a modelagem dos padrões de referência é feita a partir da utilização de um algoritmo de quantização

vetorial. Foram utilizados 30.819 elementos lógicos de uma FPGA EP2C70F896C6N⁴ da família Cyclone II da Altera, operando na frequência de 785,25 KHz. O *chip* desenvolvido a partir desta pesquisa funciona com tecnologia XFAB⁵ de 0,35 μm e área de aproximadamente de 40 mm^2 , o *chip* foi apresentado no congresso Chip in Aracaju SBCCI 2014⁶.

Em Ramos-Lara et al. (2012), é descrito um sistema de verificação de locutor em que a fase de reconhecimento do sistema é feita em *hardware*. Esta solução foi implementada utilizando uma FPGA Xilinx da família Spartan 3⁷ modelo XC3S2000. Os coeficientes mel-cepstrais são utilizados para a extração de características dos locutores e os padrões gerados na fase de treinamento são armazenados em uma memória do tipo SRAM (*Static Random-Access Memory*). De acordo com os autores, foram utilizados 24% dos recursos da FPGA de um total de 46.080 elementos lógicos; a frequência de operação utilizada foi de 50 MHz, com as operações aritméticas feitas em ponto fixo. Segundo os resultados apresentados, o tempo alcançado relativo à execução do algoritmo é comparável à execução do mesmo algoritmo em um PC Pentium IV.

A abordagem de Jinghong, Yanan e Lijia (2012), apresenta um sistema de reconhecimento de locutor utilizando coeficientes mel-cepstrais para extração de características e quantização vetorial para classificação. O sistema é prototipado em uma FPGA EP2C35F672C6⁸, trabalhando a frequência de 50 MHz. Neste estudo, também é apresentando um módulo para detecção de voz muito similar ao utilizado por Fehine et al. (2010). O módulo apresentado utiliza duas técnicas (limiar de energia e taxa de cruzamento por zero) para detecção de voz.

No estudo de Staworko e Rawski (2010), descreve-se uma solução que utiliza coeficientes cepstrais para extração de características com uma arquitetura especialmente otimizada para realizar a implementação em FPGA. O processador trabalha na frequência de 50 MHz, utilizando uma FPGA EP2C35F484C8. Segundo os autores, a implementação da solução em FPGA ocupa menos de 20% dos seus recursos que tem 33.216 elementos lógicos. Além disso, conforme os autores, por causa da otimização da arquitetura proposta para realizar o cálculo da média espectral, o sistema consegue estimar

⁴ <http://www.altera.com/education/univ/materials/boards/de2-70/unv-de2-70-board.html>

⁵ <http://www.xfab.com/home/>

⁶ <http://www.chip-in-aracaju.ufs.br/>

⁷ <http://www.xilinx.com/support/documentation/spartan-3.htm>

⁸ <http://www.altera.com/education/univ/materials/boards/de2/unv-de2-board.html>

as características dos locutores 10 vezes mais rapidamente que a solução apresentada por Ramos-Lara et al. (2009).

Na tentativa de reduzir a quantidade de recursos utilizados na FPGA, e assim, reduzir os custos de uma possível solução para o mercado consumidor, uma possível abordagem é realizar a fase de treinamento em *software*. Porque quando se trabalha com sistemas de reconhecimento de locutor, o gargalo se encontra na fase de treinamento pois, nela estão contidas as duas etapas (extração de características e geração de padrões) que demandam maior poder computacional. Uma abordagem que tenta minimizar os efeitos provocados pela necessidade de considerável demanda por poder computacional na fase de treinamento é apresentada por Wang et al. (2011).

A solução proposta por Wang et al. (2011), que utiliza coeficientes cepstrais para extrair as características dos locutores, foi desenvolver um módulo VLSI (*Very Large Scale Integration*) dedicado a gerar os padrões representativos dos locutores. As etapas de pré-processamento, extração de características e tomada de decisão são realizadas em *software*, ou seja, o sistema de Wang et al. (2011) não realiza uma fase específica (treinamento ou reconhecimento) em *software* ou *hardware*, mas sim apenas a etapa de geração de padrões que compõe a fase de treinamento, diferentemente dos estudos de Fachine et al. (2010), Staworko; Rawski (2010), Jinghong, Yanan e Lijia (2012) e Ramos-Lara et al. (2012) que apresentam uma fase (treinamento ou reconhecimento) inteiramente implementada em *software* ou *hardware*.

Na Tabela 2.1, é apresentada uma síntese das pesquisas relacionadas ao reconhecimento de locutor, a qual contém uma síntese das características extraídas, da técnica de classificação adotada e da arquitetura utilizada, bem como dos valores obtidos para a taxa de reconhecimento.

Tabela 2.1: Resumo das Pesquisas Relacionadas.

Autor	Detector de Voz	Extração de Características	Classificação	Taxa de Reconhecimento	Arquitetura
FECHINE et al. (2010)	Sim	LPC	QV	--- ⁹	<i>Software + Hardware (FPGA)</i>
STAWORKO e RAWSKI (2010)	Não	Cepstrais com banco de filtros	DTW	98,5%	<i>Software + Hardware (FPGA)</i>
WANG et al. (2011)	Não	Cepstrais	SVM	89,9%	<i>Software + Hardware (VLSI)</i>
JINGHONG, YANAN e LIJIA (2012)	Sim	Mel-Cepstrais	QV	--- ¹⁰	<i>Software + Hardware (FPGA)</i>
RAMOS-LARA et al. (2012)	Não	Mel-Cepstrais	SVM	97,9%	<i>Software + Hardware (FPGA)</i>

A contribuição desta pesquisa é um estudo comparativo do custo de utilizar coeficientes mel-cepstrais como representação de características de um locutor, e a partir dos resultados, decidir se traz benefícios uma implementação dos coeficientes MFCC em RTL, tendo como ponto de partida a pesquisa realizada por Fechine et al. (2010).

2.2 Reconhecimento de Padrões de Fala

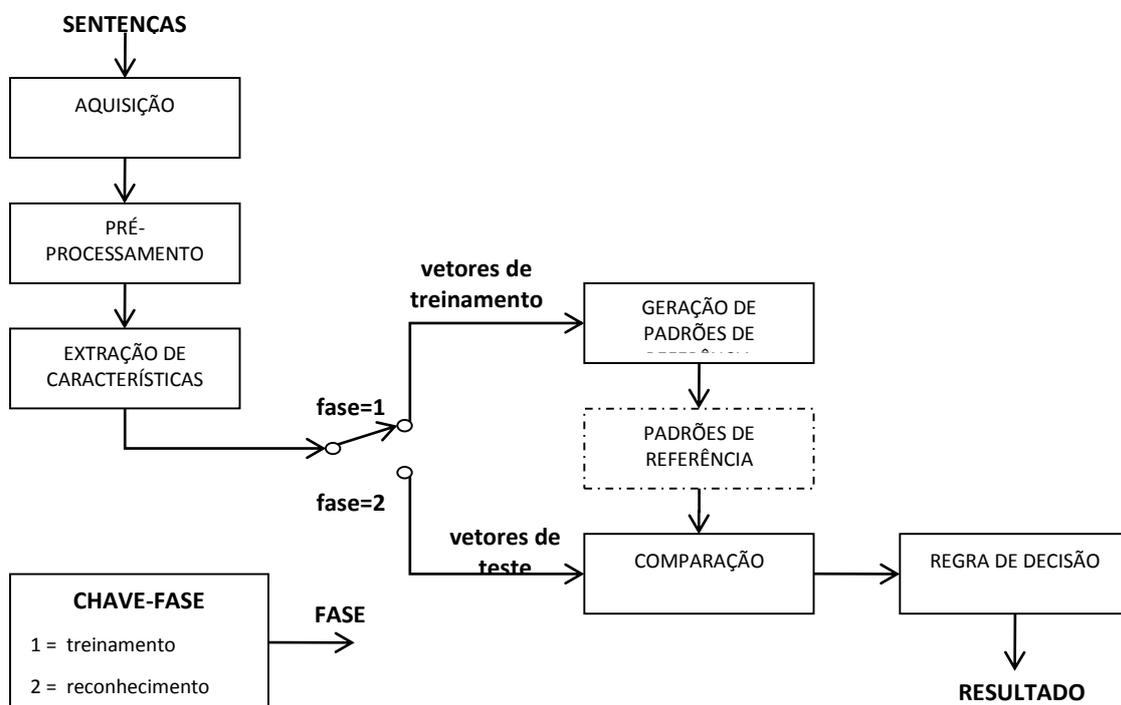
Os sistemas de reconhecimento de locutor têm como propósito solucionar um problema de reconhecimento de padrões de fala e, portanto, consistem em duas fases (RABINER e SCHAFER, 2010): treinamento e reconhecimento (teste). Na Figura 2.1, é apresentada uma representação da tarefa de reconhecimento de padrões.

⁹ Neste estudo não são apresentadas as taxas de reconhecimento do sistema.

¹⁰ Foi enviado um e-mail aos autores, no entanto, não foi obtida resposta com taxa de reconhecimento do sistema.

Figura 2.1: Representação da tarefa de reconhecimento de padrões.

Fonte:(RABINER, LAWRENCE; SCHAFFER, 2010).



Na fase de treinamento, é obtido um conjunto de padrões de referência a partir de sentenças fornecidas pelos locutores (palavras ou frases), sendo atribuídos rótulos que identificam cada locutor. Em SRL, o treinamento não ocorre necessariamente, em tempo real (treinamento *offline*), uma vez que esta fase geralmente demanda um esforço computacional significativo (HERBIG; GERL; MINKER, 2012; O'SHAUGHNESSY, 2000, 2003). Na maioria das estratégias de treinamento (FARAH; SHAMIM, 2013; GUPTA; GUPTA, 2013; WALIA, 2014) atribuem-se valores iniciais aos padrões de referência e, a partir de processos exaustivos de re-estimação, determinam-se os padrões ótimos. Além disso, uma vez estimados, os padrões de referência não necessitam ser atualizados, desde que o conjunto de locutores cadastrados no sistema permaneça inalterado ou o arquivo de treinamento não seja de boa qualidade.

Na fase de reconhecimento (verificação ou identificação), são obtidos padrões de teste, que são comparados aos padrões de referência armazenados obtidos na etapa de treinamento. Essa comparação, e o uso de uma regra de decisão, identificam o padrão que mais se assemelha às características da sentença de voz de teste (voz desconhecida), proporcionando, desta forma, o reconhecimento (RABINER e SCHAFFER, 2010). A maioria dos SRL usa a métrica

taxa de acertos¹¹ (ou taxa de erros¹²) para aferir desempenho (BEIGI, 2011; EHKAN; ALLEN; QUIGLEY, 2011; LI, 2012; RAMOS-LARA *et al.*, 2012; STAWORKO; RAWSKI, 2010).

A seguir, serão descritas as etapas presentes na tarefa de reconhecimento de padrões. As etapas de aquisição, pré-processamento e extração de características são comuns às duas fases (treinamento e reconhecimento).

2.2.1 Aquisição

A etapa de aquisição do sinal de voz compreende a captura e a conversão A/D (*Analógico/Digital*). A captura do sinal é realizada, usualmente, por um microfone, que converte a vibração sonora em sinal elétrico (sinal analógico contínuo no tempo e em amplitude).

Na fase de conversão o sinal de entrada analógico é transformado em um sinal discreto, para permitir seu processamento por qualquer sistema digital. Esse processo de conversão é feito pela amostragem do sinal analógico, $s_a(t)$, a cada T segundos, e na quantização das amostras, para se obter um sinal digital $s(n)$ (RABINER e SCHAFER, 2010), dado por:

$$s(n) = s_a(n \times T), n = 0,1,2, \dots \quad (2.1)$$

2.2.2 Pré-processamento

Conforme Da Silva (2006), a finalidade desta etapa é reduzir os efeitos indesejados incorporados ou presentes no sinal de voz, preparando-o para as etapas seguintes do processo de reconhecimento. Esta etapa é composta das seguintes subetapas: normalização, detecção de voz, pré-ênfase, segmentação e janelamento (CARDOSO, 2009; O'SHAUGHNESSY, 2000; RABINER e SCHAFER, 2010).

¹¹ A taxa de acertos corresponde ao percentual de sentenças de teste que foram associadas corretamente aos seus respectivos padrões de referência.

¹² A taxa de erros corresponde ao percentual de sentenças de teste que foram associadas a outros padrões de referência que não as representavam.

2.2.2.1 Normalização

A etapa de normalização atua na redução da variabilidade do sinal de voz em relação ao ambiente de gravação (ruído de fundo, canal de comunicação). A variabilidade causada pelas diferenças de intensidade de voz dos locutores também é restringida (O'SHAUGHNESSY, 2000), ou seja, a amplitude do sinal de voz dos locutores é limitada a uma dada faixa de valores.

2.2.2.2 Detecção de voz

Avaliar o sinal digital de forma direta, que normalmente possui grande quantidade de dados, além de requerer tempo e processamento consideráveis, pode não proporcionar resultados aceitáveis. Muitos dos pontos avaliados apresentarão informação redundante ou não conterão qualquer informação que possa ser utilizada. Um exemplo de informação desnecessária são os quadros que apresentam apenas ruídos de fundo. Portanto, para realizar uma extração de características dos locutores eficiente, é necessário eliminar trechos do sinal digital que contenham informações desnecessárias ou redundantes (DA SILVA, 2009).

Muitos métodos de detecção de voz utilizam o parâmetro de energia por segmento do sinal (vide Equação 2.2) para identificar os trechos de voz ativa (FECHINE *et al.*, 2010; JINGHONG, YANAN e LIJIA, 2012; PETRY *et al.*, 2007). Na Equação 2.2, E_{seg} , é a energia do segmento, $s(n)$ é a n -ésima amostra do sinal de voz e N_a é o número de amostras por segmento do sinal.

$$E_{seg} = \sum_{n=0}^{N_a-1} [s(n)]^2. \quad (2.2)$$

Para determinar a atividade de voz, outros métodos de detecção utilizam, por exemplo, a taxa de cruzamento por zero, o período da frequência fundamental (*pitch*), a análise espectral e a análise cepstral (LI *et al.*, 2002).

É importante destacar que o desempenho de todo sistema de reconhecimento de locutor está diretamente relacionado à qualidade da informação que é fornecida. Qualquer mecanismo de extração de características, por melhor que seja, apresentará resultados pouco significativos caso as informações utilizadas para treinamento ou reconhecimento não contenham dados

relevantes (CARDOSO, 2009; O'SHAUGHNESSY, 2000). Sendo assim, é evidente a necessidade da utilização de um eficiente método de detecção de voz.

2.2.2.3 Pré-ênfase

A proposta da pré-ênfase é enfatizar as informações úteis das componentes presentes nas altas frequências por meio de uma filtragem com filtro FIR (*Finite Impulse Response*) de primeira ordem. Esta filtragem se faz necessária devido aos efeitos da variação da glote e da impedância de radiação, causados pelos lábios no processo de produção da voz.

A função de transferência da pré-ênfase, no domínio da transformada Z, é dada pela Equação 2.3, sendo α o fator de pré-ênfase, que usualmente apresenta valores entre 0,9 e 1,0 (RABINER e SCHAFER, 1978).

$$H(z) = 1 - \alpha z^{-1}, \quad 0 \leq \alpha \leq 1. \quad (2.3)$$

2.2.2.4 Segmentação e Janelamento

O sinal de voz é fundamentalmente não estacionário e não periódico (LAMAS, 2005). No entanto, em intervalos de tempo curtos (10 a 30 ms), os segmentos de voz possuem caráter estacionário¹³ (CIPRIANO, 2001; LAMAS, 2005; RABINER e JUANG, 1993).

A divisão em segmentos é feita a partir da multiplicação do sinal de voz por uma função janela no domínio do tempo. Os tipos de janelas usualmente utilizados são: janela retangular, janela de Hamming e janela de Hanning.

Na janela retangular (vide Equação 2.4), o sinal é simplesmente particionado em blocos consecutivos de mesmo tamanho N_a . Porém, essa divisão abrupta, causa, no domínio da frequência, fugas espectrais que alteram o espectro do sinal (FECHINE, 2000; HAYVES, 1999).

$$J(n) = \begin{cases} 1, & 0 \leq n \leq N_a - 1 \\ 0, & \text{caso contrário} \end{cases} \quad (2.4)$$

¹³ Para curtos intervalos de tempo (entre 10 a 30 ms), os segmentos de voz apresentam propriedades estatísticas invariantes.

A janela de Hamming, Equação 2.5, proporciona a manutenção das características espectrais do centro do quadro e a eliminação das transições abruptas das extremidades (FECHINE, 2000; HAYVES, 1999).

$$J(n) = \begin{cases} 0,54 - 0,46 \cos \left[\frac{2\pi n}{(N_a - 1)} \right], & 0 \leq n \leq N_a - 1 \\ 0 & , \text{ caso contrário} \end{cases} \quad (2.5)$$

O janelamento de Hanning (Equação 2.6) assemelha-se à janela de Hamming, entretanto gera um esforço menor nas amostras do centro e uma suavização maior nas amostras da extremidade (FECHINE, 2000; HAYVES, 1999).

$$J(n) = \begin{cases} 0,5 - 0,5 \cos[2\pi n/(N_a - 1)], & 0 \leq n \leq N_a - 1 \\ 0 & , \text{ caso contrário} \end{cases} \quad (2.6)$$

Como nas janelas de Hamming e Hanning atribuem-se pesos baixos às amostras da extremidade do quadro, é comum o uso de superposição de quadros adjacentes, a fim de garantir que a variação dos parâmetros entre as janelas adjacentes seja mais gradual e que a análise das amostras localizadas nos extremos das janelas não seja prejudicada (DA SILVA, 2009).

2.2.3 Extração de características

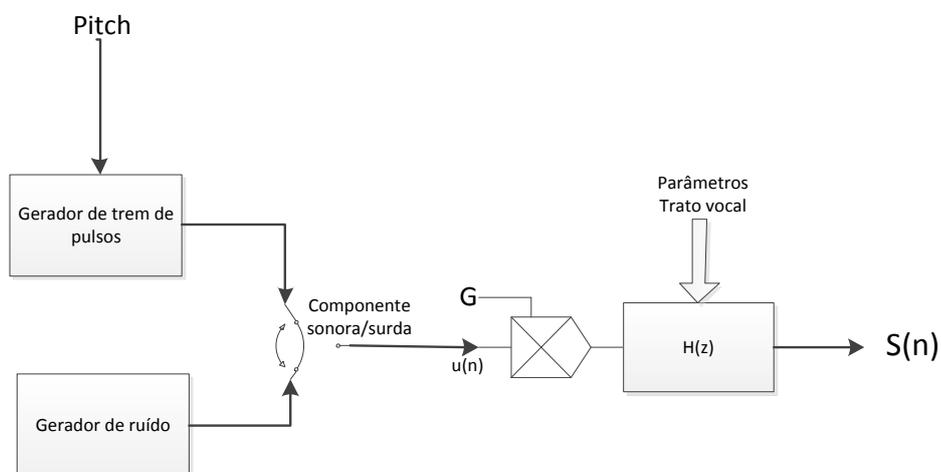
A etapa de extração de características é fundamental em sistemas de reconhecimento, uma vez que nesta etapa são obtidos elementos que possibilitam a geração de padrões. Uma das técnicas mais utilizadas para análise de voz é o método da análise por predição linear LPC (*Linear Predictive Coding*) (LI, PROUDFOOT e KING, 1995; RABINER e SCHAFER, 2010). A relevância desta técnica reside na sua boa representação dos parâmetros da fala, como também, na facilidade de sua implementação (RABINER e SCHAFER, 2010).

A ideia básica dessa análise consiste em se obter uma estimativa da voz amostrada por meio de uma combinação linear de amostras de voz passadas e de valores presentes e passados de uma entrada hipotética de um sistema, sendo, a saída deste sistema o sinal de voz (RABINER e SCHAFER, 2010).

Na Figura 2.2, está representada uma simplificação do modelo de produção da voz baseado na análise por predição linear.

As dobras vocais neste modelo seriam equivalentes a um gerador de pulsos, em que a intensidade do sinal sonoro é determinada por um ganho G . O aparelho fonador é modelado como um filtro linear $H(z)$, também denominado filtro de síntese. Quando um segmento de locução não é sonoro, considera-se que o trato vocal está sendo excitado por uma fonte de ruído (CARDOSO, 2009), o que equivale à chave na Figura 2.2 estar recebendo informação exclusivamente do gerador de ruído.

Figura 2.2: Modelo simplificado para produção de voz baseado no modelo LPC.



O principal problema da predição linear é determinar um conjunto de coeficientes do preditor a partir do sinal de voz, a fim de se obter uma estimativa precisa das propriedades espectrais do sinal de voz. Como o sinal de voz é variante no tempo, os coeficientes do preditor devem ser estimados em curtos intervalos de tempo (de 10 até 30 ms) (CIPRIANO, 2001; LAMAS, 2005; O'SHAUGHNESSY, 2000; RABINER e SCHAFER, 1978).

Esses coeficientes podem ser obtidos diretamente a partir da análise LPC, chamados coeficientes LPC, ou por meio de outras técnicas derivadas dessa análise (DIAS, 2006).

Os coeficientes mais utilizados são: LPC, cepstrais, cepstrais ponderados, delta cepstrais, delta cepstrais ponderados e mel cepstrais (FARAH; SHAMIM, 2013; GUPTA; GUPTA, 2013; RAMOS-LARA *et al.*, 2012; WANG *et al.*, 2011).

2.2.3.1 Coeficientes LPC

No modelo simplificado de produção da voz (Figura 2.2), a função de transferência do trato vocal, $H(z)$, para curtos intervalos de tempo, é definida pela Equação 2.7 (RABINER, LAWRENCE; SCHAFER, 2010).

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{i=1}^p a_i z^{-i}}, \quad S(z) = U(z)H(z), \quad (2.7)$$

na qual:

$S(z)$ – Transformada-z da sequência de voz $s(n)$;

$U(z)$ – Transformada-z do sinal de excitação $u(n)$;

a_i – coeficientes LPC;

p – ordem da predição (número de coeficientes);

G – parâmetro de ganho.

No domínio do tempo, as amostras de voz, $s(n)$, são relacionadas com a excitação sonora ou surda, $u(n)$, e com o parâmetro de ganho G , pela Equação 2.8 (RABINER e SCHAFER, 2010).

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n), \quad (2.8)$$

Uma predição linear com coeficientes de predição, a_i , considerando-se as suas p amostras anteriores, é definida pela Equação 2.9.

$$\tilde{s}(n) = \sum_{i=1}^p a_i s(n-i), \quad (2.9)$$

Vários métodos podem ser utilizados na resolução da Equação 2.9: o método da autocorrelação, o método da covariância, a formulação do filtro inverso, a formulação da estimação espectral, a formulação da máxima verossimilhança e a formulação do produto interno (RABINER e SCHAFER, 2010).

O método da autocorrelação é baseado na minimização do erro de predição, $e(n)$, associado à predição de $s(n)$ pela estimativa $\tilde{s}(n)$. O erro de predição é dado pela Equação 2.10.

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^P a_i s(n-i), \quad (2.10)$$

Para se obter uma solução para o problema de minimização do erro, deve-se selecionar um segmento do sinal de voz por meio de uma janela de comprimento finito e igual a N_a . O comprimento da janela deve estar compreendido em intervalos de curta duração (de 10 até 30 ms), de forma a garantir a estacionariedade do segmento de voz selecionado (RABINER e SCHAFER, 1978).

Nas próximas equações, $x(n)$, corresponderá ao segmento selecionado e ponderado pela janela.

A Equação 2.8, modificada pela janela, será descrita como:

$$x(n) = Gu + \sum_{i=1}^p a_i x(n-i), \quad (2.11)$$

Pela Equação 2.9, tem-se que a predição do segmento $x(n)$ é definida na Equação 2.12.

$$\tilde{x}(n) = \sum_{i=1}^p a_i x(n-i), \quad (2.12)$$

Dado que $\tilde{x}(n)$ é a aproximação de $x(n)$ e que a_i é o i -ésimo coeficiente da predição linear; $\tilde{x}(n)$ é normalmente denominada a estimativa ou a predição de ordem p da amostra $x(n)$.

O erro da predição de cada amostra, $e(n)$, é definido pela Equação 2.13.

$$e(n) = x - \tilde{x}(n) = x(n) - \sum_{i=1}^p a_i x(n-i), \quad (2.13)$$

E o erro quadrático, ε , em todo o segmento é dado por:

$$\varepsilon = \sum_{n=-\infty}^{\infty} e(n)^2 = \sum_{n=-\infty}^{\infty} \left[x(n) - \sum_{i=1}^p a_i x(n-i) \right]^2, \quad (2.14)$$

Como o segmento de voz é nulo para $n < 0$ e para $n > N_a$, o erro de predição (Equação 2.14) é nulo para $n < 0$ e $n > N_a + p - 1$. Logo, obtém-se a Equação 2.15 a partir da Equação 2.14.

$$\varepsilon = \sum_{n=0}^{N_a+p-1} e(n)^2 = \sum_{n=0}^{N_a+p-1} \left[x(n) - \sum_{i=1}^p a_i x(n-i) \right]^2, \quad (2.15)$$

O erro mínimo para o conjunto de coeficientes a_i é obtido fazendo-se:

$$\frac{\partial(\varepsilon)}{\partial(a_i)} = 0, \quad 1 < i \leq p, \quad (2.16)$$

Com a substituição da Equação 2.15 em 2.16 e a realização de derivadas parciais, obtém-se:

$$\sum_{i=1}^p a_i R_r(|j-i|) = R_r(j), \quad 1 \leq j \leq p, \quad (2.17)$$

com

$$R_r(i) = \sum_{n=0}^{N_a-p-1} x(n)x(n+i), \quad (2.18)$$

As Equações 2.17 e 2.18, equações de Wiener-Hopf, podem ser mais bem visualizadas na forma matricial indicada na Equação 2.19 (FECHINE, 2000; RABINER e SCHAFER, 2010).

$$\begin{pmatrix} R_r(0) & R_r(1) & \dots & R_r(p-1) \\ R_r(1) & R_r(0) & \dots & R_r(p-2) \\ R_r(2) & R_r(1) & \dots & R_r(p-3) \\ \dots & \dots & \dots & \dots \\ R_r(p-1) & R_r(p-2) & \dots & R_r(0) \end{pmatrix} x \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_p \end{pmatrix} = \begin{pmatrix} R_r(1) \\ R_r(2) \\ R_r(3) \\ \dots \\ R_r(p) \end{pmatrix}, \quad (2.19)$$

Os coeficientes a_i do preditor são determinados pela solução das Equações 2.17 e 2.18 ou pela Equação 2.19.

A matriz de autocorrelação (vide Equação 2.19) é uma matriz simétrica e, assim, pode ser resolvida de forma eficiente por algoritmos recursivos, como por exemplo, o algoritmo de Levinson-Durbin (CARVALHO, 2007; FECHINE, 2000; MARPLE, 1987).

2.2.3.2 Coeficientes Mel-Cepstrais

Atualmente, os coeficientes mel-cepstrais (*MFCC - Mel-Frequency Cepstrum Coefficients*) são amplamente usados em sistemas de reconhecimento de locutor (DEY *et al.*, 2014; EHKAN; ALLEN; QUIGLEY, 2011; MARTINEZ *et al.*, 2012; SADJADI; HANSEN, 2014; SELVAN; JOSEPH; ANISH BABU, 2013) devido à melhoria proporcionada na precisão do reconhecimento. Aliado a este fato, tem se mostrado mais robustos na presença de ruídos de fundo em comparação com outros coeficientes (RAMOS-LARA *et al.*, 2012).

Os coeficientes MFCC surgiram devido aos estudos na área de psicoacústica (a ciência que estuda a percepção auditiva humana). Esta ciência mostra que a percepção das frequências de tons puros ou de sinais de voz não seguem uma escala linear, estimulando assim, a ideia de criar uma escala, denominada de Mel.

2.2.3.2.1 Escala Mel

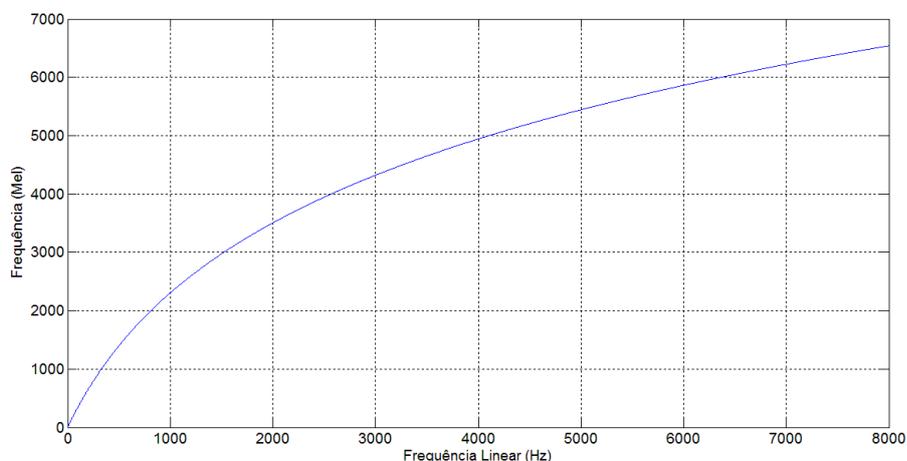
A escala Mel procura aproximar características da sensibilidade do ouvido humano, pois verificou-se que a percepção humana de frequências de tons puros ou de sinais de voz não seguem uma escala linear. Para cada tom com uma determinada frequência, medida em Hz, é associado um valor medido na escala Mel. O Mel é uma unidade de medida de frequência. Como referência, foi definida a frequência de 1 kHz, com potência 40 dB acima do limiar mínimo da audição do ouvido humano, equivalendo a 1000 *mels* (CARDOSO, 2009; DE LIMA, 2001).

O mapeamento da frequência em Hz para a frequência na escala Mel ocorre segundo a equação 2.23:

$$F_{mel} = 2595 \log \left(1 + \frac{F_{linear}(Hz)}{700} \right), \quad (2.23)$$

Os valores 2595 e 700 são obtidos experimentalmente e, utilizando estes valores, é possível verificar a relação entre a escala linear e escala Mel a partir da Figura 2.3.

Figura 2.3: Mapeamento da frequência linear na escala Mel.

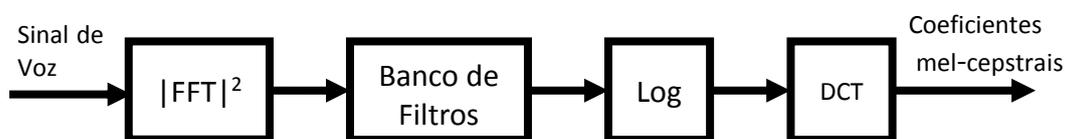


2.2.3.2.2 Estimação dos Coeficientes MFCC

Na Figura 2.4, é descrito o processo de obtenção dos coeficientes mel-cepstrais utilizando um banco de filtros. Antes de extrair esses coeficientes, o sinal de voz passa por uma etapa de pré-processamento, que já foi descrita anteriormente. Após essa etapa, é calculado o módulo da FFT (*Fast Fourier Transform*) sobre o sinal janelado, obtendo-se o espectro do sinal. Posteriormente, os espectros dos segmentos são submetidos ao banco de filtros triangulares, em que as frequências centrais desses filtros obedecem a uma escala acusticamente definida (escala Mel). Segundo Cardoso (2009), o formato triangular do filtro permite enfatizar as componentes presentes nas frequências centrais, atenuando as demais. Em seguida, calcula-se o logaritmo da energia resultante da saída dos filtros, com a aplicação da DCT (*Discrete Cosine Transform*) sobre os valores obtidos do cálculo do logaritmo, chegando-se aos valores dos coeficientes mel-cepstrais.

Figura 2.4: Obtenção dos coeficientes MFCC a partir de banco de filtros.

Fonte: (CARDOSO, 2009).



Ainda de acordo com Cardoso (2009), procura-se utilizar quadros de comprimento de 2^n para agilizar o cálculo da transformada discreta de Fourier com métodos de transformação rápida.

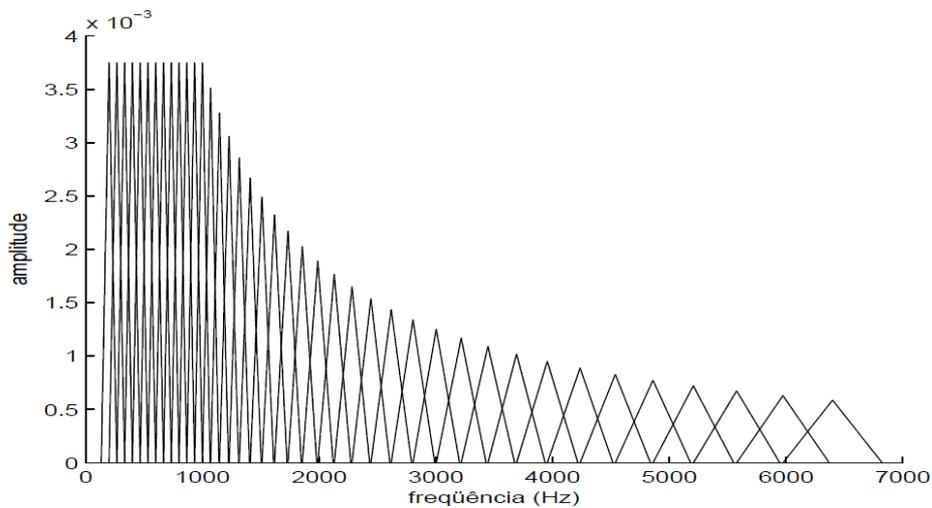
O uso de banco de filtros possibilita reduzir a dimensionalidade da informação extraída do sinal de voz. Deste modo, quanto menor o número de filtros empregados mais compacta é a informação extraída. Como os filtros triangulares estão espaçados linearmente segundo a escala Mel, existe um número maior de filtros nas frequências mais baixas onde está concentrada a energia do sinal de voz. Empregando esse banco de filtros, o que se procura fazer é enfatizar as frequências nas quais ocorrem as mudanças perceptíveis de tom conforme a escala mel.

A construção do banco de filtros realizada neste estudo é baseada na implementação feita por Slaney (1998). De acordo com Cardoso (2009), esse banco de filtros não exige tanto esforço computacional para seu cálculo.

Cada filtro triangular é dimensionado de forma a apresentar uma mesma área, a qual pode ser unitária para fins de simplificação. Percebe-se que os filtros lineares têm a mesma altura. O mesmo não acontece com os filtros logarítmicos cuja altura vai sendo reduzida progressivamente. Na Figura 2.5, é possível visualizar o espectro do banco de filtros.

Figura 2.5: Espectro do banco de filtros triangulares de Slaney (1998).

Fonte: (CARDOSO, 2009).



O processo de obtenção dos coeficientes mel-cepstrais pode ser apresentado matematicamente da seguinte maneira:

$$S(k) = \sum_{k=1}^{N_{FFT}} W_k(j)X(j), \quad k = 1, \dots, M, \quad (2.24)$$

na qual:

$W_k(j)$ – as janelas de ponderação triangulares conforme a escala Mel;

$X(j)$ – o espectro da magnitude da FFT de N pontos;

$S(k)$ – sinal de saída do banco de filtros;

M – número de filtros utilizados.

Agora é aplicada a DCT sobre o sinal de saída do banco de filtros, conforme a Equação 2.25.

$$C_{mel}(n) = \sum_{k=1}^M \log(S(k)) \cos \left[n \left(k - \frac{1}{2} \right) \right] \frac{\pi}{M}, \quad n = 0, 1, \dots, M, \quad (2.25)$$

na qual:

$C_{mel}(n)$ – o n-ésimo coeficiente mel-cepstral.

Os métodos de extração de características que incorporam aos modelos de produção da voz propriedades de percepção auditiva humana, como os coeficientes mel-cepstrais, apresentam melhores taxas de reconhecimento nos SRL (O'SHAUGHNESSY, 2000; RAMOS-LARA *et al.*, 2012). No entanto, eles acarretam em um aumento de complexidade no desenvolvimento do modelo em *hardware*. Este aumento da complexidade deve-se ao cálculo da FFT, do logaritmo e da DCT, como também da aplicação do banco de filtros na escala Mel.

2.2.4 Geração de Padrões

Como mencionado anteriormente, a tarefa de reconhecimento de locutor é um exemplo típico de uma tarefa de reconhecimento de padrões. O principal objetivo desse tipo de tarefa é a comparação realizada entre padrões que são obtidos a partir da representação de características extraídas de um sinal de voz, pertencentes a um locutor desconhecido (fase de reconhecimento), com padrões de referências previamente armazenados (fase de treinamento).

Os métodos frequentemente utilizados para a construção desses padrões são: Modelos de Markov Escondidos (AMUDHA *et al.*, 2007; LIU; CHEUNG, 2012), Máquinas de Vetores de Suporte (WANG *et al.*, 2011; RAMOS-LARA *et al.*, 2012), Modelos de Misturas Gaussianas (CARDOSO, 2009; EHKAN; ALLEN; QUIGLEY, 2011), Alinhamento Dinâmico no Tempo (PHADKE *et al.*, 2004; STAWORKO; RAWSKI, 2010), Quantização Vetorial (FECHINE *et al.*, 2010; JINGHONG; YANAN; LIJIA, 2012) e Redes Neurais Artificiais (AMUDHA; VENKATARAMANI; MANIKANDAN, 2008; GARIMELLA *et al.*, 2012).

A técnica utilizada neste estudo para a geração dos padrões de referência foi a Quantização Vetorial. Esta técnica é indicada por remover a redundância, que geralmente existe, entre os sucessivos vetores de observação (O'SHAUGHNESSY, 2000). Além disso, a quantização vetorial apresenta um baixo custo computacional.

2.2.4.1 Quantização Vetorial

A quantização vetorial (QV) é um método de compressão de dados que mapeia um conjunto de vetores de observação, X , com o objetivo de gerar um conjunto de M vetores-código, C . Os conjuntos X e C são definidos formalmente por (CIPRIANO, 2001; HE, LIU e PALM, 1999):

$$X = \{x_t \in \mathcal{R}^p \mid t = 1 \dots T\}. \quad (2.26)$$

$$C = \{y_i \in \mathcal{R}^p \mid i = 1 \dots M\}. \quad (2.27)$$

O conjunto corresponde ao dicionário (*codebook*) do quantizador que possui níveis e os vetores x_t e y_i , são p -dimensionais. Portanto, tem-se um quantizador vetorial p -dimensional de M -níveis (ou M -partições).

Para realizar o mapeamento do conjunto de observações em um dicionário, o algoritmo LBG (*Linde-Buzo-Gray*) é comumente utilizado. Os passos para geração do dicionário pelo algoritmo LBG são os seguintes (CIPRIANO, 2001; FECHINE, 2000):

1. Definir o número de níveis do quantizador M ;
2. Escolher, de forma aleatória, M vetores de observação, $x_t \in X$, para compor um dicionário inicial C ;
3. Calcular a distorção entre cada vetor de observação $x_t \in X$ e os vetores-código $y_i \in C$, codificando $x_t \in X$ pelo índice i do vetor-código $y_i \in C$ que apresenta a menor distorção, ou seja, se classifica cada vetor de observação $x_t \in X$ em uma das M partições do dicionário;
4. Definir novos vetores-código para as M partições de C , tal que, o novo valor do vetor-código y_i corresponda à média dos vetores de observação, $x_t \in X$, que foram classificados na partição i ;
5. Calcular o valor da distorção percentual e comparar com o limiar. Se este valor for maior que o limiar, retorna-se ao passo 3; caso contrário, finaliza-se o processo com o dicionário C , definido.

Na verificação de locutor, a QV é usada nas duas fases: treinamento e reconhecimento. Na primeira, a quantização vetorial atua na geração de dicionários-modelo (padrões de referência). Na segunda, o vetor de teste é utilizado para o cálculo da distorção em relação aos dicionário-modelos. Uma das técnicas mais utilizadas para a construção do dicionário do quantizador e para a

obtenção da distorção do vetor teste em relação ao dicionário é baseada na distância euclidiana apresentada na Equação 2.28.

$$d(x_t, y_i)^2 = \sum_{j=1}^p (x_{t,j} - y_{i,j})^2, 1 \leq t \leq T, 1 \leq i \leq M, \quad (2.28)$$

A tomada de decisão é baseada em dois limiares: um superior e outro inferior. Quando o valor de distância fica abaixo do limiar inferior, o locutor é aceito, ou seja, o locutor é realmente quem ele alega ser. Entretanto, se o valor de distância fica acima do limiar superior, o locutor é rejeitado. Por fim, se o valor de distância ficar entre os dois limiares, é dito que o locutor é desconhecido e então o mesmo repete sua sentença.

2.3 Reconhecimento de Locutor em Sistemas Embarcados

Sistemas embarcados são sistemas computacionais que executam uma função específica. Eles possuem a mesma estrutura geral de um computador, mas a especificidade de suas tarefas faz com que não sejam utilizados ou percebidos como um computador (DIAS, 2011). Estes sistemas estão presentes em vários aparelhos: *smartphones*, televisores, aparelhos de som, vídeo games, câmeras digitais, forno micro-ondas e outros aparelhos domésticos. Também estão presentes nos carros, por meio do controle do *airbag*, freios ABS, computador de bordo e outros sistemas de controle; nos aviões em sistemas de controle de voo, sistemas anticolisão e outros sistemas de controle.

Em relação à implementação em *hardware* de um sistema embarcado, diferentes abordagens podem ser seguidas. Quando a aplicação requer uma produção em grande escala (mercado de consumo), o mais recomendado é o uso de um SoC (*System-On-Chip*). Este sistema possui os circuitos de processamento de informações incluídos em um único chip. Sua arquitetura pode conter um ou mais processadores, microcontroladores, DSP, memórias, interfaces para periféricos e blocos dedicados. Tais blocos, também chamados de IP Cores (*Intellectual Property Cores*), consistem de blocos em *hardware* que executam tarefas específicas e são definidos de modo a permitir o seu reuso em diferentes sistemas (MARWEDEL, 2011).

Para o desenvolvimento em pequena escala de produção ou de protótipos, a abordagem mais recomendada é em FPGA (*Field Programmable Gate Array*). Este é composto por uma matriz de blocos lógicos configuráveis, interconectados por meio de uma rede de conexões programáveis. A principal vantagem de desenvolver aplicações em FPGA consiste na possibilidade de fácil modificação da estrutura do *hardware* do sistema por meio de um processo denominado de reconfiguração, o que permite o desenvolvimento incremental, correção de erros de projeto, além da adição de novas funções de *hardware* (MARWEDEL, 2011; RAMOS-LARA *et al.*, 2012).

Algoritmos com bom desempenho na área de Processamento Digital de Sinais de Voz, em termos de taxa de reconhecimento, demandam um certo poder computacional, o que já se tem disponível em máquinas tipo PC em tempo real. No entanto, em sistemas embarcados dependentes de baterias, como um telefone celular, o desafio não representa apenas o desenvolvimento de *software*, mas também de *hardware* que se adapte às características inerentes de tais dispositivos, principalmente o baixo consumo de energia (KRISHNA; MAHLKE; AUSTIN, 2004; NEDEVSKI; PATRA; BREWER, 2005).

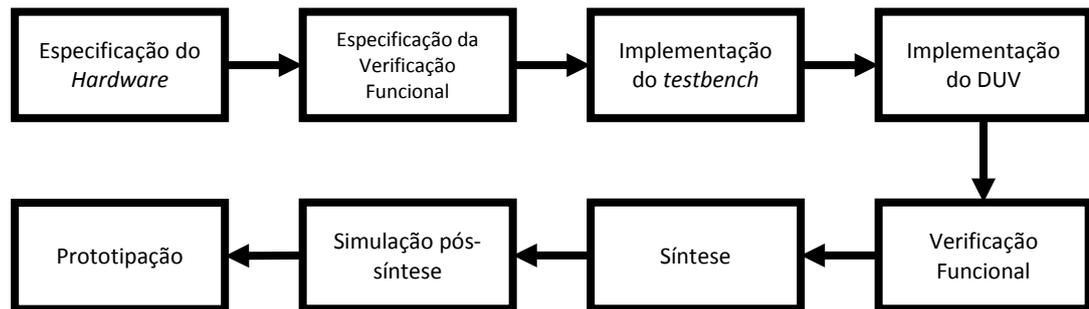
Algumas técnicas de programação permitem a geração de um código mais eficiente no que se refere à quantidade de operações a serem executadas pelo processador (p.ex.: substituição de uma multiplicação por deslocamento de bits), o que conseqüentemente proporciona menor consumo de potência e energia. Sendo assim, a partir da otimização de algoritmos, utilização de estruturas de *hardware* específicas e um *pipeline* de instruções eficiente, é possível reduzir os recursos necessários, sem diminuir a taxa de reconhecimento (DIAS, 2011).

Na próxima seção, é apresentado o processo de desenvolvimento de um sistema embarcado.

2.4 Processo de Desenvolvimento de um Sistema Embarcado

Na Figura 2.6, são apresentadas as etapas necessárias para o desenvolvimento do *hardware* de um sistema embarcado.

Figura 2.6: Fases do processo de desenvolvimento de *hardware* na ordem proposta por VeriSC (DA SILVA, 2007).



A finalidade dessa subdivisão apresentada por Da Silva (2007), é tornar a fase de verificação mais rápida e confiável na eliminação de erros.

2.4.1 Especificação do *Hardware*

A especificação do *Hardware* compreende o estudo e a descrição das funcionalidades que o dispositivo deve executar quando estiver pronto. A descrição é feita por meio de documentos contendo todos os requisitos necessários ao funcionamento do dispositivo. Segundo Da Silva (2007), esta especificação deve possuir alto nível de abstração, no qual ainda não tenham sido tomadas decisões em relação à implementação das funcionalidades, em termos da arquitetura-alvo a ser adotada, nem sobre os componentes de *hardware* ou *software* a serem selecionados. Ela contém detalhes de alto nível, tais como funcionalidades a serem executadas, e também informações de baixo nível, por exemplo, especificação e descrição de pinos a serem utilizados.

2.4.2 Especificação da Verificação Funcional

Nesta etapa, são documentados aspectos importantes que devem ser verificados do dispositivo a ser desenvolvido. Na especificação da verificação funcional é elaborado o plano de verificação, que contém informações sobre como as funcionalidades que serão “cobertas”, tipos de estímulos utilizados e faixa de valores dos estímulos, dentre outras.

2.4.3 Implementação do *testbench*

Segundo Oliveira (2010), esta fase é voltada para a construção do ambiente de verificação funcional do DUV (*Design Under Verification*). Este ambiente é conhecido como *testbench*. O DUV é uma descrição do dispositivo (ou parte dele) a ser desenvolvido, codificado em alguma linguagem de descrição de *hardware* (HDL - *Hardware Description Language*). Uma diferença significava entre as linguagens HDL e as linguagens de programação voltadas para *software* é que a sintaxe e a semântica de linguagens HDL incluem notações explícitas que expressam concorrência e tempo. A etapa de implementação do *testbench* apresenta grande impacto na fase de verificação funcional, visto que é neste ambiente que o DUV será inserido e receberá estímulos, gerando saídas que serão comparadas com as saídas desejadas.

2.4.4 Implementação do DUV

O processo de desenvolvimento do DUV corresponde à concepção do código RTL (*Register Transfer Language*). Este código corresponde ao nível mais baixo de abstração, o nível de sinais. A concepção do código RTL é feita utilizando-se linguagens de descrição de *hardware*, como, VHDL, Verilog, SystemVerilog, dentre outras.

2.4.5 Verificação Funcional

De acordo com Bergeron (2003), a verificação funcional é um processo usado para demonstrar que o objetivo do projeto é preservado em sua implementação. A

verificação funcional constitui umas das etapas do processo de desenvolvimento de *hardware*, a qual é realizada a partir de simulação. Na verificação funcional dois modelos são comparados: o que está sendo desenvolvido (o DUV) e outro que reflete a especificação. Uma maneira de realizar essa comparação consiste em inserir valores pseudoaleatórios na entrada do DUV, o qual será comparado com um modelo ideal, chamado de **modelo de referência**, o qual é uma implementação executável das especificações do projeto.

2.4.6 Síntese

Consiste na conversão da descrição RTL para um conjunto de registradores e lógica combinacional. O código RTL é sintetizado e, a partir deste, é gerada uma *netlist* no nível de portas lógicas.

2.4.7 Simulação pós-síntese

Após a síntese do código RTL, é necessário fazer uma nova simulação. No entanto, esta nova simulação leva em consideração aspectos específicos do dispositivo, como por exemplo, atrasos de porta lógica.

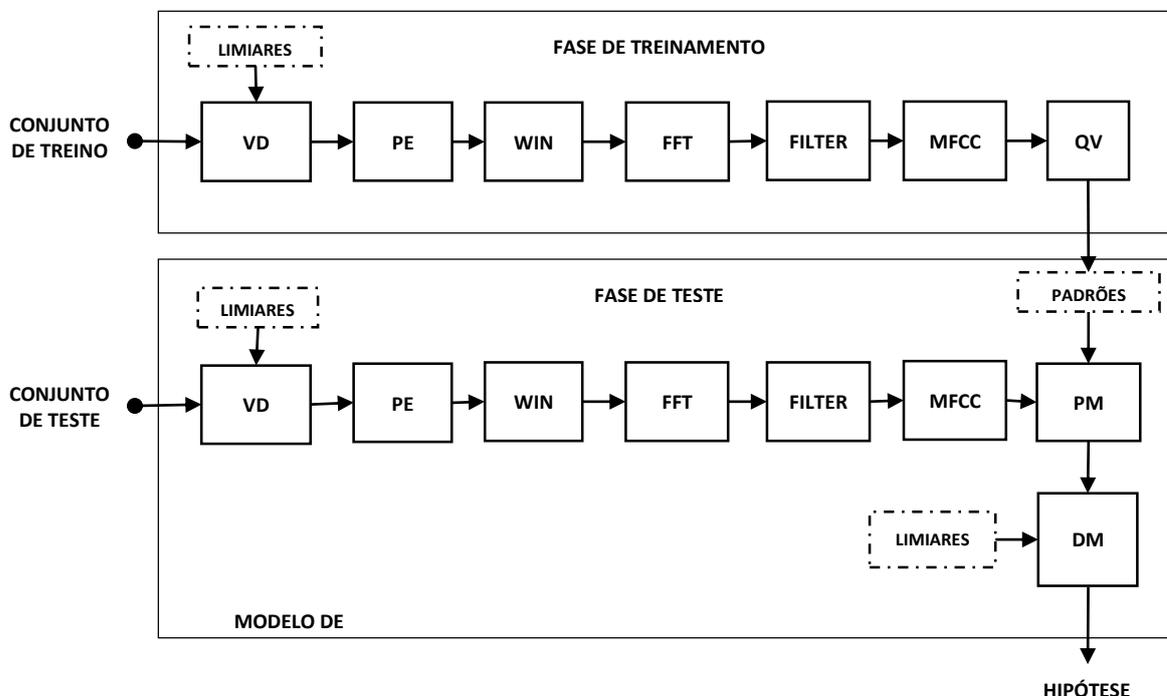
2.4.8 Prototipação

Nesta fase, ocorre a implantação da *netlist* obtida na fase de síntese em algum dispositivo de *hardware*. A prototipação pode ser feita numa FPGA e na sequência, em silício. Ao final de todo processo, caso o protótipo gerado funcione de maneira correta, o IP-core poderá ser integrado com outros IP para criação de um SoC.

3 Adequação do SPVR para uso de Coeficientes Mel-Cepstrais

Na Seção 2.1, foram descritas as principais características de um sistema de reconhecimento de locutor. Neste capítulo, será feita uma descrição do sistema de verificação de proposto. Na Figura 3.1, está representado o diagrama em blocos do sistema, que é constituído das etapas descritas a seguir.

Figura 3.1: Diagrama em blocos do sistema de verificação de locutor.



- Pré-processamento: etapa responsável por extrair do sinal de voz os dados relevantes e desprezar a informação redundante, com a finalidade de repassar a informação útil para a etapa seguinte;
- Extração de características: etapa responsável por estimar parâmetros da fala que irão compor o conjunto de características;
- Quantização vetorial: etapa responsável por gerar os padrões de referência que serão comparados com os padrões de teste na fase de teste;
- Comparação: distância obtida a partir da distância Euclidiana;

- Regra de Decisão: etapa responsável por fornecer um resultado utilizando a distorção obtida na comparação e os limiares utilizados no sistema.

Conforme mencionado anteriormente, a tarefa de reconhecimento de locutor é dividida em duas fases: treinamento e teste. A fase de treinamento é feita *offline* por demandar esforços computacionais significativos para gerar os padrões de referência, que não necessitam ser atualizados com frequência, dado que a base de dados de treinamento permanecerá constante. Assim sendo, há uma diminuição da quantidade de recursos utilizados da FPGA, visando-se a redução dos custos de uma possível solução para o mercado consumidor.

Nesta pesquisa, o *software* para o treinamento foi implementado na linguagem C++, enquanto o *software* para verificação (modelo de referência) foi implementado na linguagem SystemVerilog. Os módulos VD, PE, WIN, DM e PM, utilizados nesta pesquisa, foram codificados pela equipe Brazil-IP-CG (FECHINE *et al.*, 2010).

Os blocos FFT, FILTER e MFCC foram implementados neste estudo e integrados com a pesquisa da equipe BRAZIL-IP-CG, substituindo o bloco LPCA no estudo de Fechine *et al.* (2010) em *software*, no papel do modelo de referência. Em nível RTL, esses três blocos não foram realizados nem tiveram a verificação funcional finalizados e conseqüentemente não foram integrados com os outros módulos implementados pela equipe do Brazil-IP-CG. Entretanto, uma estimativa de área e consumo desses blocos é apresentada no capítulo de resultados.

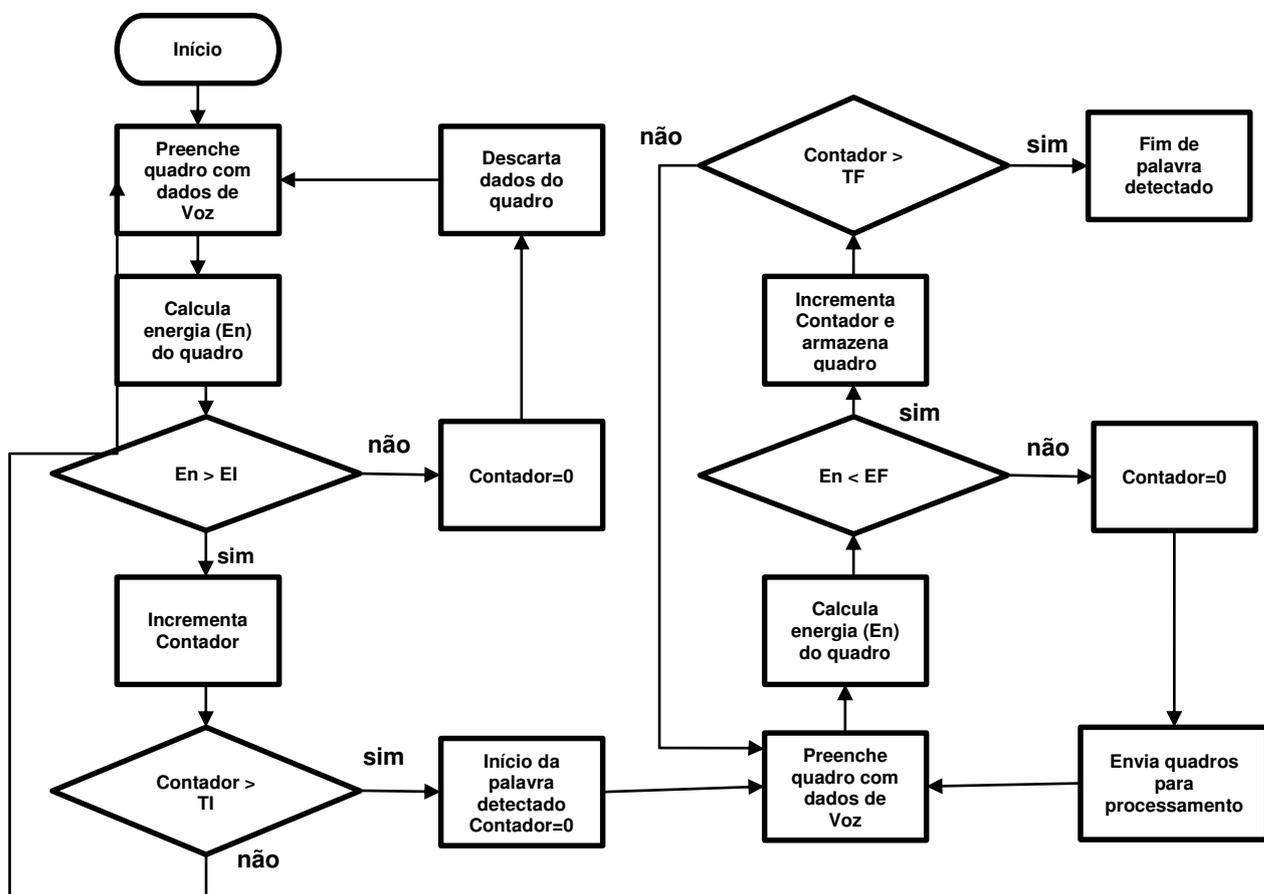
3.1 Detecção de voz (VD)

O método de detecção de voz adotado nesta pesquisa utiliza o parâmetro de energia por segmento (Equação 2.5), com N_a igual a 110 amostras, o que equivale a um trecho de, aproximadamente, 10 ms do sinal de voz para uma frequência de amostragem de 11.025 amostras/segundo.

O algoritmo escolhido para detecção de voz foi o apresentado por Petry et al. (1999), sendo possível visualizar seu fluxograma na Figura 3.2.

Figura 3.2: Algoritmo de detecção de início e fim de sentenças.

Fonte: Adaptado de (PETRY *et al.*, 1999).



Nesse algoritmo, a energia do sinal é calculada para cada quadro de tamanho fixo (segmento de voz de 110 amostras). O início de uma possível sentença é considerado a partir do primeiro quadro que apresenta uma energia maior que um limiar de início (EI) de voz. Para que os quadros seguintes constituam um início de sentença, a energia desses deve permanecer acima do limiar de início durante um período de tempo, chamado de tempo inicial (TI)¹⁴. Durante o período de início, caso algum quadro apresente uma energia menor que o limiar de início (EI), as amostras de voz armazenadas são descartadas e a procura do início da sentença recomeça (PETRY *et al.*, 1999).

¹⁴ Corresponde ao número de quadros sucessivos para a detecção do início de uma palavra. Cada um desses quadros deve apresentar uma energia maior que a energia do limiar de início (EI).

Com o início da sentença detectado, é feita a procura pelo seu fim. O método de detecção do final da sentença é semelhante ao método de detecção de início. A diferença reside na procura por quadros com energia inferior a um limiar de fim (EF). Dessa forma, os quadros de voz devem permanecer com energia inferior ao limiar de fim por um dado período de tempo, chamado de tempo final (TF)¹⁵, para que se considere atingido o fim da sentença. Os quadros inseridos no tempo final (TF) não são considerados componentes da sentença (PETRY *et al.*, 1999).

Na seção 4.1.2, é descrita a abordagem aplicada para determinação dos 4 (quatro) limiares do módulo de detecção de voz.

3.2 Pré-ênfase (PE)

O resultado da pré-ênfase $s_p(n)$ está dependente da entrada $s(n)$ conforme a Equação 3.1 (CIPRIANO, 2001; DIAS, 2006):

$$s_p(n) = s(n) - as(n - 1), \quad (3.1)$$

Nesta pesquisa, foi utilizado um fator de pré-ênfase de 15/16 (que corresponde a 0,9375), uma vez que esse fator facilita a implementação em ponto fixo¹⁶, e como consequência, simplifica a implementação em hardware, já que uma operação de divisão ou multiplicação por potência de 2 pode ser realizada por deslocamento de bits (CIPRIANO, NUNES e BARONE, 2003; DIAS, 2006). Esta relação pode ser vista na Equação 3.2.

$$S_p(n) = s(n) - \frac{15}{16}s(n - 1) = s(n) - s(n - 1) + \frac{s(n-1)}{16}, \quad (3.2)$$

Detalhes do processo de pré-ênfase foram descritos na Seção 2.1.3.3.

3.3 Segmentação e Janelamento (WIN)

Para garantir as condições de estacionariedade do sinal, foram utilizados segmentos de aproximadamente 23 ms, o que equivale a 256 amostras por quadro

¹⁵ Corresponde ao número de quadros sucessivos para a detecção do final da palavra. Cada um desses quadros deve apresentar uma energia menor que a energia do limiar de fim (LF).

¹⁶ Formato numérico em que o ponto decimal ocupa uma posição fixa.

a uma frequência de 11.025 amostras/segundo. A função janela escolhida foi a de Hamming com superposição de 50%. Esta janela mantém as características espectrais do centro do quadro e minimiza as descontinuidades nas extremidades. A superposição de 50%, por sua vez, tende a fazer com que a variação dos parâmetros de janelas adjacentes seja mais gradual, uma vez que a função de Hamming atribui pesos baixos às amostras da extremidade do quadro (HAYVES, 1999).

Extração de Características (FILTER e MFCC)

Para cada segmento de voz janelado é extraído um conjunto de coeficientes LPC, a partir da análise por predição Linear, e outro conjunto de coeficientes mel-cepstrais. Os coeficientes mel-cepstrais formam vetores de características mais adequados ao desenvolvimento de aplicações de reconhecimento de locutor (RAMOS-LARA *et al.*, 2012) do que os coeficientes LPC (CAEIROS, MIYATAKE e MEANA, 2009). Na Seção 4.2, é apresentada uma análise comparativa de desempenho destes coeficientes com o objetivo de evidenciar este fato.

A escolha da ordem de predição p (número de coeficientes) depende da frequência f_s de amostragem do sinal de voz conforme Equação 3.1 (RABINER e SCHAFER, 1978). Sendo assim, dado que $f_s = 11.025$ amostras/segundo, foi escolhido, por uma aproximação, um valor de $p = 12$.

$$p \cong \frac{f_s}{1000}. \quad (3.1)$$

Para o cálculo dos coeficientes mel-cepstrais, conforme apresentado na Seção 2.1.4.3, foi utilizado um banco de filtros sugerido por Slaney (1998). Conforme Cardoso (2009), esse banco de filtros não exige elevado esforço computacional para seu cálculo.

3.4 Quantização Vetorial (QV)

Na pesquisa ora descrita, foi utilizado um quantizador vetorial p -dimensional de 64 níveis (ou 64 partições). Esta quantidade de níveis é suficiente para uma boa representação do sinal de voz sem geração de um grande volume de dados (FECHINE *et al.*, 2010; GUPTA e GUPTA, 2013; WALIA, 2014). O algoritmo

utilizado para a geração dos padrões de referência foi o LBG descrito na Seção 2.2.4.1.

3.5 Comparação (PM)

Na fase de reconhecimento, a comparação dos vetores de teste p -dimensionais com a identidade vocal que o locutor alega ser, é feita por meio do cálculo da distância euclidiana (vide Equação 2.30), conforme Seção 2.2.5.1. Deste modo, é enviado um valor de distância (ou distorção) para o módulo de decisão.

3.6 Regra de Decisão (DM)

Sendo recebido, do módulo de comparação, o valor de distância (ou distorção), o objetivo do módulo de decisão é indicar se o vetor de teste se assemelha à identidade vocal alegada pelo locutor.

Como citado anteriormente, para se chegar a uma resposta, a regra de decisão é baseada em dois limiares: um superior (LS) e outro inferior (LI). A escolha desses limiares é de fundamental importância para o desempenho do sistema. Na Seção 4.2.1, é descrita a obtenção e quais os valores escolhidos para o sistema proposta. Quando o valor de distância fica abaixo do limiar inferior, o sistema responde com "aceito", ou seja, o vetor de teste se assemelha à identidade vocal alegada. Entretanto, se o valor de distância fica acima do limiar superior, o sistema fornece como resposta "rejeitado". Neste caso, o vetor de teste não se assemelha à identidade vocal alegada. Pode ocorrer do valor de distância ficar entre os dois limiares. Isto acontecendo, o sistema responde com "desconhecido" e solicita que seja fornecido novo vetor de teste. A distância utilizada como parâmetro de decisão foi a distância Euclidiana, conforme Equação 3.2.

$$m_i = \frac{1}{Q} \sum_{j=1}^Q \text{distância}(I), \quad (3.2)$$

sendo:

m_i – Distância média dos vetores teste, que compõem a sentença de teste, em relação à identidade vocal alegada;

Q – Número de quadros da sentença de teste;

I – Identidade vocal alegada;

$distância(I)$ – Distância de um dado vetor de teste de índice j em relação à identidade vocal alegada I .

4 Apresentação e Análise dos Resultados

Neste capítulo, são apresentados e discutidos os resultados obtidos a partir das configurações do sistema simuladas nesta pesquisa. A análise das configurações tem como objetivo sugerir a configuração que maximize as saídas corretas do sistema de verificação de locutor.

4.1 Metodologia

Antes da apresentação e discussão dos resultados, na primeira seção é descrita a metodologia empregada para obtenção dos resultados. Após a descrição do modo de obtenção da base de dados e sua composição, é descrito o algoritmo para a determinação automática dos limiares do VD e as configurações dos experimentos simulados.

4.1.1 Base de Dados

A base de dados é constituída por dois conjuntos distintos, denominados conjunto de treinamento e conjunto de teste. Esses conjuntos são utilizados nas fases de treinamento e de teste, respectivamente, dos SRL.

Na fase de treinamento, o conjunto de treinamento tem por finalidade fornecer elementos para a geração dos padrões de referência. Na fase de teste, o conjunto de elocuições de teste possibilita a avaliação do desempenho do sistema. Essa avaliação se dá por meio de métricas, tais como taxa de acertos e taxa de erros.

A base de dados utilizada para a validação deste estudo foi gravada com uma frequência de amostragem de 11.025 amostras/segundo, utilizando-se uma placa DE2-70. O ambiente de gravação não foi controlado. Desta forma, algumas amostras podem conter ruído. Outras características da gravação estão apresentadas na Tabela 4.1.

Um conjunto com 30 sentenças foi utilizado como senha vocal dos locutores. As sentenças foram escolhidas da pesquisa de MAMEDE et al. (2003). A escolha destas sentenças se deu por serem frases foneticamente balanceadas¹⁷. A lista com as 30 sentenças utilizadas pode ser vista no Anexo A deste documento.

Tabela 4.1: Características da gravação

Microfone	Leson [®] MC-200; Resposta em frequência: 70 Hz a 12 kHz; Sensibilidade a 1 kHz: -49 dB; Impedância: 600 Ω
Codec de som	WM8731 Wolfson [®] Microelectronics
Formato	Wave
Número de canais	1 canal (mono)
Resolução	16 bits por amostra

A base de dados foi gerada a partir de 30 locutores, sendo 15 mulheres e 15 homens. Cada locutor pronunciou cada sentença do conjunto 10 vezes, com exceção da sentença escolhida pelo locutor como senha vocal que foi pronunciada 20 vezes (10 para teste e 10 para treinamento). No total, foram utilizadas, portanto, 9.300 elocuições para a base de dados, 9.000 elocuições para teste e 300 elocuições para o treinamento.

4.1.2 Detecção do Limiars do VD

Para a determinação dos valores dos limiars de energia (EI e EF) e de tempo (TI e TF), os seguintes passos foram seguidos:

1. Determinação de medidas estatísticas da energia do conjunto de treinamento, descrito na seção 4.1.1. Esses dados estão presentes na Tabela 4.2;

¹⁷ Significa que as frases escolhidas têm uma distribuição fonética similar àquela encontrada na fala espontânea.

2. Escolha de medidas estatísticas da energia do conjunto de treinamento para avaliação dos limiares de energia. As medidas escolhidas foram a mediana (Q2) e o terceiro quartil (Q3);
3. Escolha de limiares de tempo. A faixa escolhida para avaliação está compreendida em um período de 1 a 25 quadros, representando 10 ms do sinal de voz;
4. Simulação dos limiares definidos. Na Tabela 4.3, estão representados os limiares avaliados;
5. Análise das métricas de desempenho do sistema para os limiares definidos.

Tabela 4.2: Dados estatísticos da energia do conjunto de treinamento.

Medida Estatística	Energia por Segmento
Mínimo	0,000005
Primeiro Quartil (Q1)	0,000337
Mediana (Q2)	0,002449
Terceiro Quartil (Q3)	0,046433
Máximo	14,706371
Média	0,379271
Variância	0,379271

Tabela 4.3: Limiares de energia e de tempo.

Arranjo	EI	EF	TI	TF
A-I	Q2	Q2	[1,.., 25]	[1,.., 25]
A-II	Q2	Q3	[1,.., 25]	[1,.., 25]
A-III	Q3	Q2	[1,.., 25]	[1,.., 25]
A-IV	Q3	Q3	[1,.., 25]	[1,.., 25]

4.1.3 Simulação dos dados

No total, foram utilizados 2.500 (dois mil e quinhentos) arranjos de detecção de voz, os arranjos são compostos pela junção dos limiares de energia (EI e EF) e os limiares de tempo (TI e TF) (conforme Tabela 4.3) e são usados para os 2 (dois) tipos de métodos de extração de características (mel-cepstrais e LPC). As implementações dos módulos de pré-ênfase, segmentação, janelamento,

quantização vetorial e comparação (distância euclidiana) se repetem para todos os experimentos, conforme apresentado no Capítulo 3.

As configurações foram agrupadas em 8 experimentos¹⁸, conforme descrito na Tabela 4.4. A métrica utilizada para aferir o desempenho de cada configuração foi a quantidade de erros¹⁹ (QE). Os resultados de cada experimento são apresentados no formato de uma matriz de 25x25, sendo que as linhas correspondem aos valores assumidos pelo limiar de tempo inicial (TI) e as colunas correspondem aos valores assumidos pelo limiar de tempo final (TF). As matrizes com esses resultados encontram-se no Apêndice A deste documento.

Tabela 4.4: Descrição dos Experimentos.

Experimento	Limiares do DV	Extração de
E1	A-I	Mel-cepstrais
E2	A-II	Mel-cepstrais
E3	A-III	Mel-cepstrais
E4	A-IV	Mel-cepstrais
E5	A-I	LPC
E6	A-II	LPC
E7	A-III	LPC
E8	A-IV	LPC

É importante destacar que todas as execuções dos experimentos foram realizadas utilizando-se a base de dados descrita na Seção 4.1.1.

4.2 Avaliação dos Experimentos

Antes de avaliar a eficiência dos coeficientes MFCC para representação das características de um locutor por meio da comparação com coeficientes LPC é preciso determinar os limiares de detecção de voz (EI, EF, TI e TF) e tomada de decisão (LI e LS) que minimizam a quantidade de erros dos experimentos.

¹⁸ Cada experimento consistiu de um conjunto formado por 625 elementos, no qual cada elemento corresponde a uma configuração do sistema de verificação automática da identidade vocal dependente de texto.

¹⁹ Número de sentenças de teste que apresentaram medida de distorção abaixo do limiar inferior, não pertencentes ao locutor alegado ou medida de distorção acima do limiar superior pertencentes ao locutor alegado.

4.2.1 Resultado dos Limiares do Sistema

Com base na análise estatística realizada e apresentada no Apêndice B, os melhores pares de limiares de tempo, que tornam mínima a quantidade de erros para cada experimento realizado, são apresentados na Tabela 4.5.

Tabela 4.5: Limiares de tempo que minimizam QE.

Limiares de tempo (TI, TF)	QE Mínima	Experimento
(1,8)	20	E1
(1,9)	20	E1
(1,10)	20	E1
(2,9)	20	E1
(2,17)	20	E1
(2,18)	20	E1
(2,19)	20	E1
(2,20)	20	E1
(2,21)	20	E1
(2,22)	20	E1
(2,23)	20	E1
(2,24)	20	E1
(2,25)	20	E1
(7,7)	24	E2
(7,10)	24	E2
(1,15)	20	E3
(1,17)	20	E3
(2,15)	20	E3
(2,17)	20	E3
(2,18)	20	E3
(3,6)	26	E4
(14,1)	120	E5
(7,2)	120	E6
(2,5)	118	E7
(3,5)	118	E7
(7,4)	127	E8

Conforme informações da Tabela 4.5, as configurações dos experimentos, que apresentam QE mínima, possuem limiares de tempo inicial compreendidos na faixa de 1-14 quadros e limiares de tempo final na faixa de 2-25. A escolha do limiar TI é, portanto, mais restritiva que TF e causa maior impacto na taxa de reconhecimento.

Ainda de acordo com a Tabela 4.5, existe mais de uma configuração que minimiza a QE, então a escolha do par de limiares será baseada na quantidade de resultados desconhecidos fornecidos pelo sistema, conforme a regra de decisão vista na Seção 3.7. As tabelas contendo esses resultados e a quantidade respostas dos locutores desconhecidos podem ser vistas no Apêndice C.

Os valores dos limiares de tomada de decisão, por sua vez são apresentados na Tabela 4.6.

Tabela 4.6: Resultados para variação dos limiares de tomada de decisão usando coeficientes MFCC.

Limiares		Resultados	
LI	LS	Quantidade de Erros	Desconhecidos
0,057	0,0620	15	51
0,057	0,0619	15	51
0,057	0,0618	15	49
0,057	0,0617	15	49
0,057	0,0616	15	48
0,057	0,0615	15	47
0,057	0,0614	15	47
0,057	0,0613	15	46
0,057	0,0612	15	45
0,057	0,0611	15	44
0,057	0,0610	16	42

Portanto, a partir das informações das Tabelas 4.5, 4.6 e C.1, os limiares escolhidos para o sistema usando coeficientes MFCC para a extração de características foram: limiares de energia iguais a 0,002449 (EI) e 0,002449 (EF), limiares de tempo iguais a 1 (TI) e 9 (TF), extração de características mel-cepstrais e limiares de decisão iguais a 0,0611 (LS) e 0,0570 (LI), para o sistema usando coeficientes LPC para extração de características os limiares escolhidos foram: limiares de energia iguais a 0,046433 (EI) e 0,002449 (EF), limiares de tempo iguais a 14 (TI) e 1 (TF), extração de características LPC e limiares de decisão iguais a 0,0380 (LS) e 0,0320 (LI).

4.2.2 Análise Comparativa entre MFCC e LPC

Após a determinação dos limiares, o desempenho geral obtido pelos sistemas pode ser visualizado na Tabela 4.7.

Tabela 4.7: Índices dos resultados para a melhor configuração de limiares.

Índice de Acertos		Índice de Erros		Índice de Desconhecidos	
MFCC	LPC	MFCC	LPC	MFCC	LPC
99,34%	96,52%	0,17%	0,90%	0,49%	2,58%

A partir dos resultados da Tabela 4.7, é possível verificar um aumento de 430% na taxa de erros, quando os coeficientes LPC são utilizados no lugar dos coeficientes MFCC para representação das características de um locutor. Também ocorre um acréscimo no índice de respostas desconhecidas (427%) quando são utilizados coeficientes LPC em vez de coeficientes MFCC, isto significa que o sistema solicitará mais repetições das sentenças aos locutores, o que pode gerar uma insatisfação por parte dos usuários ao utilizar o sistema com coeficientes LPC.

A taxa de falsa rejeição indica a fração de tentativas nas quais usuários são rejeitados de maneira incorreta. Na Tabela 4.8, é possível verificar que a utilização de coeficientes MFCC acarreta uma menor taxa de falsa aceitação.

A taxa de repetição da própria sentença corresponde a quantidade de vezes que usuário é realmente quem ele alega ser e o sistema fornece como resposta “desconhecido” e solicita ao usuário que repita sua senha vocal. A taxa de repetição da própria sentença também é menor quando coeficientes MFCC são utilizados (conforme Tabela 4.8). É possível visualizar esses índices e outras informações individualizadas para cada locutor no Apêndice D.

Tabela 4.8: Índices de falsa aceitação e Repetição da Própria Sentença.

Falsa Rejeição		Repetição da Própria Sentença	
MFCC	LPC	MFCC	LPC
4,34%	16%	3,33%	11%

Para reforçar os resultados apresentados na Tabela 4.7 e 4.8 e apresentar outras justificativas para o uso dos coeficientes MFCC, podemos utilizar as matrizes

de confusão com a menor quantidade de erros para os dois métodos de extração de características. As matrizes podem ser visualizadas nas Tabela 4.9 e Tabela 4.10.

Das Tabelas 4.9 e 4.10, pode-se destacar o resultado para quatro locutores, com relação à taxa de falsa rejeição. Primeiramente, para os locutores L21 e L29, onde a taxa de falsa rejeição ficou em 100% ($r=10$), ou seja, para esse conjunto de teste, estes dois locutores não foram reconhecidos em nenhuma das dez repetições das suas sentenças utilizando coeficientes LPC. No entanto, utilizando-se coeficientes MFCC para o mesmo conjunto de teste o locutor L21 foi aceito em 100% das repetições ($a=10$) e o locutor L29 foi aceito em 60% das repetições ($a=6$). Para os locutores L10 e L28, o índice de falsa rejeição ficou em 80% ($r=8$), e nas duas elocuições que não houve falsa rejeição, o sistema solicita repetição das sentenças ($d=2$). Com coeficientes MFCC os resultados são diferentes para esses dois locutores. O locutor L28 é aceito em 100% das elocuições por ele realizada ($a=10$) e o locutor L10 é aceito em 70% das suas elocuições ($a=7$).

Outro resultado contrário aos coeficientes LPC é o aumento no índice de falsa aceitação, seu problema consiste na possibilidade de usuários mal intencionados conseguirem se passar por locutores cadastrados no sistema. Os resultados que mais chamam a atenção são aqueles do locutor L2 que em todas as repetições da sentença do locutor L3, L2 consegue ser aceito em 100% das tentativas ($a=10$)²⁰. Porém, usando como extração de características os coeficientes mel-cepstrais, em nenhum teste L2 conseguiu se passar por L3 ou vice-versa. Nas elocuições da sentença do locutor L4, L2 foi aceito em 70% das repetições de maneira incorreta ($a=7$). Esses são os resultados para falsa aceitação mais importantes de serem destacados, em outras ocorrências a taxa não passou dos 40%.

Utilizando-se coeficientes mel-cepstrais, verificou-se a ocorrência de falsa aceitação em apenas dois casos. Nas elocuições da sentença do locutor L16, L19 foi aceito em 10% das repetições de maneira incorreta ($a=1$) e nas elocuições da sentença do locutor L4, L23 foi aceito em 10% das repetições ($a=1$). Assim sendo, verificou-se a maior robustez dos coeficientes MFCC para a representação das características de um locutor, quando comparados a coeficientes LPC.

²⁰ Este é o caso onde os locutores são irmãs gêmeas, isto pode ser um atenuante para o resultado.

Tabela 4.10: Matriz de confusão para a menor quantidade de erros com LPC (cont.).

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L16	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	d=7, r=3	r=10	r=10	a=3, d=6, r=1	r=10	r=10	r=10	r=10	r=10	r=10	r=10	d=4, r=6	r=10	r=10	r=10
L17	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=4, d=6	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10
L18	d=1, r=9	r=10	d=3, r=7	a=1, d=8, r=1	d=4, r=6	r=10	d=1, r=9	r=10	d=7, r=3	d=1, r=9	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=10	r=10	r=10	r=10	d=8, r=2	d=6, r=4	r=10	r=10	r=10	r=10	d=5, r=5	r=10	r=10
L19	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	d=2, r=8	r=10	r=10	a=4, d=6	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10
L20	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=1, d=9	r=10	r=10	r=10	r=10	r=10	a=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10
L21	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10
L22	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10
L23	d=2, r=8	r=10	d=1, r=9	a=4, d=5, r=1	d=9, r=1	r=10	r=10	r=10	a=3, d=7	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=10	r=10	r=10	r=10	r=10	d=2, r=8	r=10	r=10
L24	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	d=6, r=4	r=10	r=10	r=10	r=10	r=10	r=10
L25	d=1, r=9	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=1, d=8, r=1	r=10	r=10	r=10	r=10	r=10
L26	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=10	r=10	r=10	r=10	r=10
L27	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	d=3, r=7	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=10	r=10	r=10	r=10
L28	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	d=2, r=8	0	0	
L29	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10
L30	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	r=10	a=10

4.2.3 Modelo para implementação do módulo MFCC em FPGA

Com relação à implementação em *hardware* do módulo responsável para extração dos coeficientes MFCC, será usado como base a pesquisa apresentada por Saambhavi, Rao e Rajalakshmi (2012). Nessa pesquisa é apresentada um *design* de um circuito para extração de características MFCC para reconhecimento de fala. O modelo matemático apresentado no estudo é muito semelhante ao *software* implementado nesta pesquisa cujos resultados foram apresentados na Seção 4.2.2.

A seguir será apresentado o modelo matemático do estudo apresentado por Saambhavi, Rao e Rajalakshmi (2012) e código em C++ equivalente.

Os pesos dos filtros são calculados previamente e armazenados na memória. Cada quadro do sinal de fala é passado pelos filtros, resultando nos valores do espectro de potência na frequência Mel, estes valores são armazenados de volta na memória. A operação de filtragem pode ser representada pela multiplicação entre a matriz que contém os pesos dos filtros e a matriz que contém o quadro de fala, conforme equação 4.1. Sendo $f_{x,y}$, $a_{x,y}$, $emag_{x,y}$, os pesos dos filtros, os valores da magnitude da FFT e a magnitude auditiva, respectivamente.

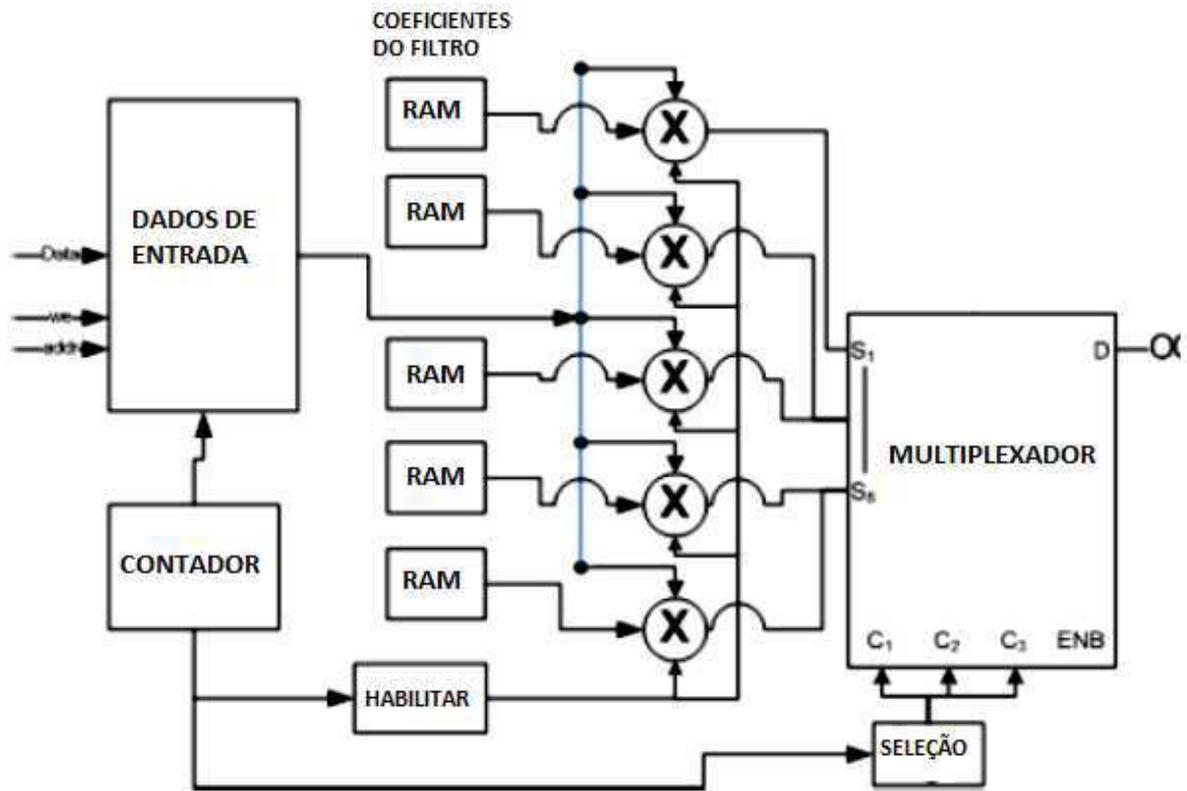
A modelagem da implementação em *hardware* pode ser vista na Figura 4.10.

$$\begin{pmatrix} f_{1,1} & \cdots & f_{1,256} \\ \vdots & \cdots & \vdots \\ f_{27,1} & \cdots & f_{27,256} \end{pmatrix} * \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{256,1} \end{pmatrix} = \begin{pmatrix} emag_{1,1} \\ \vdots \\ emag_{27,1} \end{pmatrix}. \quad (4.1)$$

A RAM com os dados de entrada representa a memória onde os quadros com sinal de voz serão armazenados. Em Saambhavi; Rao e Rajalakshmi (2012) são utilizados 40 (quarenta) filtros, na abordagem proposta nesta pesquisa foram utilizados 27 filtros. Essa quantidade de filtros minimiza a taxa de erros e reduz a quantidade de recursos necessários para a implementação do banco de filtros quando comparado, por exemplo, com a quantidade de filtros utilizada por Saambhavi; Rao e Rajalakshmi (2012).

Figura 4.1: Modelagem do extrator da magnitude auditiva.

Fonte: Adaptado de (SAAMBHAVI; RAO e RAJALAKSHMI (2012)).



Na matriz de peso dos filtros, existem 225 valores diferentes de zero de um total de 6.912 elementos da matriz. Então, serão realizadas 225 multiplicações e algumas adições para cada quadro do sinal de fala. Após a filtragem, cada quadro produz 27 valores que serão armazenados em uma coluna da matriz *emag*. Cada elemento da matriz *emag*, por exemplo $emag_{x,1}$ é produzido pela soma da multiplicação de cada elemento da x -ésima linha da matriz de peso dos filtros com os elementos correspondentes na coluna da matriz de fala. A implementação é feita com os números em ponto fixo.

O código C++ responsável por esta etapa encontra-se no Código Fonte 4.1.

Código Fonte 4.1: Extração do espectro Mel.

```
for(int chan = 0; chan < numberFilters; chan++){

    for(int i = 0; i < numberFftPoints; ++i){

        if((fftFreq[i] > lower[chan]) && (fftFreq[i] <= center[chan])){

            mfccFw[chan][i] = triangleHeigth[chan] * ((fftFreq[i]-lower[chan]) / (center[chan]- lower[chan]));

        }

        if ((fftFreq[i] > center[chan]) && (fftFreq[i] <= upper[chan])){

            mfccFw[chan][i] = triangleHeigth[chan] * ((upper[chan]-fftFreq[i]) / (upper[chan]-center[chan]));

        }

    }

}

void FILTERBANK::applyFilters(FFT fftValues){

    double sum = 0.0;

    for(int i=0; i<numberFilters; ++i){ // aplicação dos filtros (filtros ponderados x espectro da FFT)

        sum = 0.0;

        for(int k=0; k<numberFftPoints; k++){

            sum += (mfccFw[i][k] * fftValues.fMagnitude[k]);

        }

        if(sum != 0.0){

            earMag[i] = log10(sum);

        }

        else earMag[i] = 0.0;

    }

}
```

Para a implementação da operação de logaritmo, a abordagem escolhida por Saambhavi, Rao e Rajalakshmi (2012) foi LUT (*Look-up Table*). Outra possível solução é utilizando uma aproximação com os coeficientes de Chebyshev. Segundo Meyer-Baese (2014), a utilização dos coeficientes Chebyshev possibilita uma

rápida convergência para o valor desejado do logaritmo a custo reduzido de recursos. Esta foi a abordagem escolhida nesta pesquisa.

Os valores obtidos após a passagem pelo banco de filtros e o cálculo do logaritmo estão armazenados na memória e serviram como entrada para o cálculo final do coeficientes MFCC. Os coeficientes são obtidos por meio da aplicação da DCT sobre os valores da matriz *emag*. Isso pode ser expresso pela seguinte equação:

$$melCoef_i = \frac{2}{27} \sum_1^{27} (\log(mag_j)) \cos(\pi * 2 * 27(j - 1) * i), \quad (4.2)$$

na qual :

i – a quantidade de coeficientes MFCC, com $0 \leq i \leq 12$;

j – a quantidade de filtros utilizada, com $0 \leq j \leq 27$.

A Equação 4.2 pode ser reescrita da seguinte maneira:

$$melCoef_i = \frac{2}{27} \sum_1^{27} (\log(mag_j)) * m_{i,j}, \quad (4.3)$$

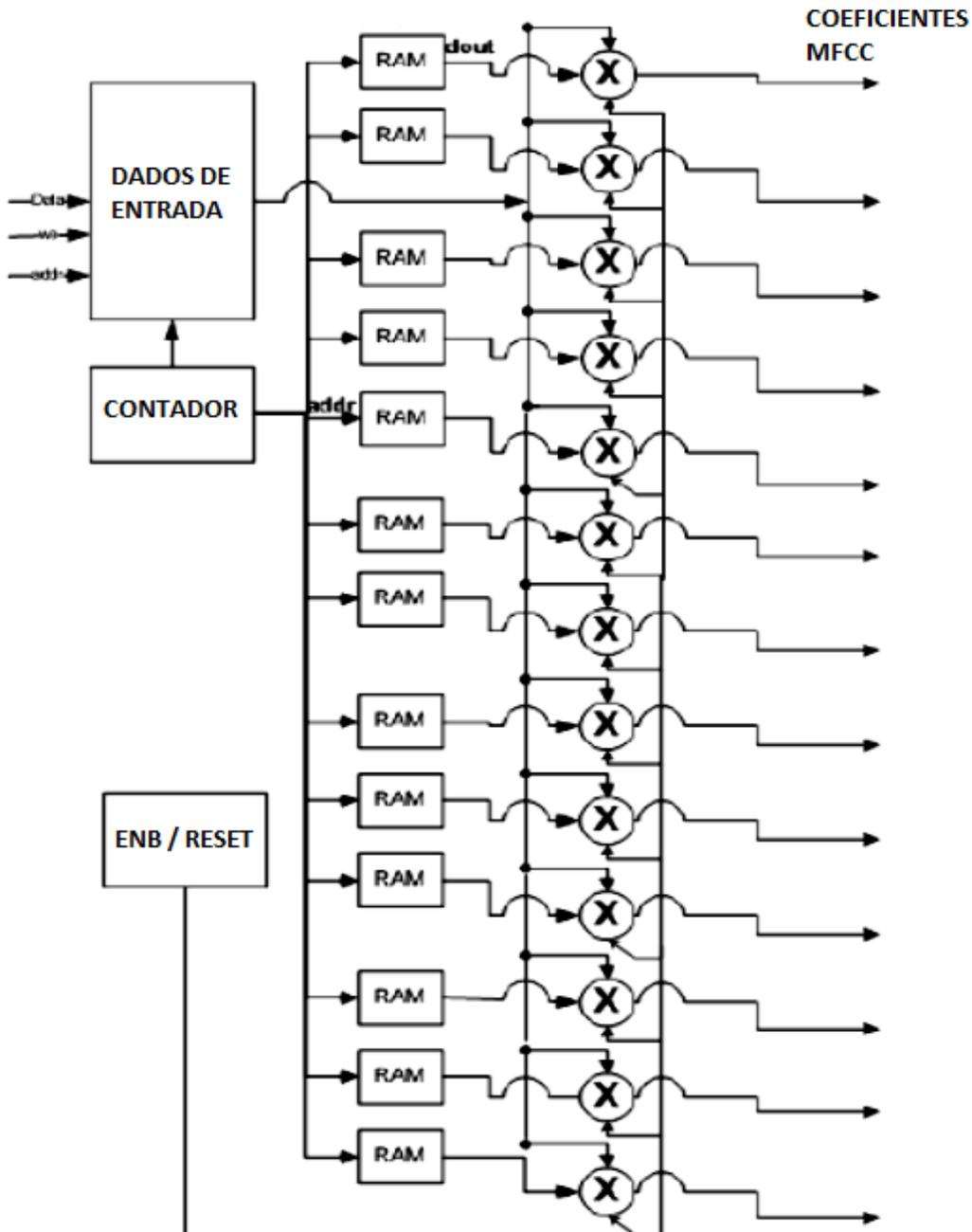
Com $m_{i,j} = \cos(\pi * 2 * 27(j - 1) * i)$. Então para o cálculo dos coeficientes MFCC pode ser modelado da seguinte maneira:

$$\begin{pmatrix} melCoef_1 \\ \vdots \\ melCoef_{13} \end{pmatrix} = \frac{2}{27} * m_{13,27} * \begin{pmatrix} emag_{0,0} \\ \vdots \\ emag_{26,0} \end{pmatrix}, \quad (4.4)$$

Uma matriz de 13 multiplicadores e acumuladores é usada para calcular cada coeficiente cepstral. Isso evita os tempos de acesso de memória repetidas necessárias para computar cada dimensão dos coeficientes MFCC. Os valores de cosseno da equação são armazenados na memória. Uma unidade de controle é implementada para manter o controle dos endereços de memória necessários para acessar os valores de cosseno e a matriz de magnitudes *emag*. Dessa maneira os valores da matriz, *emag*, precisam ser acessados na memória apenas uma vez. A modelagem dessa implementação pode ser vista na Figura 4.2.

Figura 4.2: Design para a extração dos coeficientes MFCC.

Fonte: Adaptado de (SAAMBHAVI; RAO e RAJALAKSHMI (2012)).



A memória RAM contém os dados de entrada, a matriz *emag*, as RAM filtro coeficientes contém os valores do cosseno. O resultado do cálculo produz os coeficientes MFCC. O código C++ equivalente a esta modelagem em *hardware*, cujos resultados foram apresentados na Seção 4.2.2, pode ser visto a seguir.

Código Fonte 4.2: Extração dos coeficientes MFCC.

```
void MELCEPCOEFF::calcMelCoeff(FILTERBANK filter){  
  
    double scalefac = (double) (1.0 / sqrt ((double)numberFilters/2.0));  
  
    for(int j=0; j < numCepCoeffs; j++){  
  
        for(int i=0; i < numberFilters; ++i){  
  
            mfccDCT[j][i] = scalefac * cos(j*(2*i + 1)* PI/2/numberFilters);  
  
            if(j==0){  
  
                mfccDCT[j][i] *= (double) (sqrt(2.0)/2.0);  
  
            }  
  
        }  
  
    }  
  
    double sum = 0.0;  
  
    for(int i=0; i < numCepCoeffs; i++){  
  
        sum = 0.0;  
  
        for(int k=0; k < numberFilters; k++){  
  
            sum += (mfccDCT[i][k] * filter.earMag[k]);  
  
        }  
  
        melCoffs[i] = sum;  
  
    }  
  
    for(int i=0; i < melVectSize; i++){  
  
        melCoffs[i]= melCoffs[i+1];  
  
    }  
  
}
```

Utilizando-se a ferramenta *dc_shell*²¹ da Synopsys, foram obtidos relatórios sobre área, como pode ser visto na Tabela 4.11, e consumo de energia, como pode ser visto na Tabela 4.12, dos blocos LPC, FFT, Filter e MFCC. O bloco LPC

²¹ <http://www.synopsys.com/Tools/Implementation/RTLSynthesis/DesignCompiler/>

corresponde a implementação feita por Fachine et al. (2010), o bloco responsável pela FFT é uma implementação que o Matlab²² disponibiliza. Os outros dois blocos foram implementados para esta pesquisa, porém não tiveram a verificação funcional finalizada. Foi feita uma verificação funcional preliminar, para alguns casos escolhidos, com entradas dirigidas. Conforme pode ser visto no Apêndice F.

Tabela 4.11: Relatório sobre a área dos blocos.

Informação	Módulos			
	LPC	FFT	FILTER	MFCC
Quantidade de portas	33	71	38	34
Quantidade de Nets	5590	604	11494	2253
Células Combinacionais	3422	1	10753	1045
Células Sequenciais	1530	0	5839	1044
Área Combinacional em mm ²	177623	30542	123555	26292
Área Total em mm ²	224664	42950	318545	65669

Conforme era esperado, a substituição do LPC por outros 3 (três) blocos acarretará o aumento de 90% da quantidade de recursos e área necessários para a extração das características vocais de um locutor utilizando coeficientes MFCC. Isto acontece pelo fato do aumento da complexidade das operações presentes nesses 3 (três) blocos e o aumento da quantidade de memórias RAM necessárias para armazenar o peso dos filtros e os valores do cosseno para o cálculo da DCT.

²² <http://www.mathworks.com/help/dsp/ref/ffthdoptimized.html>

Outra informação obtida foi uma estimativa do consumo dos blocos que pode ser vista na Tabela 4.12. O consumo irá aumentar em aproximadamente 27,27% com a substituição do LPC pelos outros blocos.

Tabela 4.12: Estimativa de consumo dos blocos.

Informação	Módulos			
	LPC	FFT	FILTER	MFCC
Consumo Dinâmico	18 uW	3 uW	11 uW	11 uW
Consumo <i>Leakage</i>	4 uW	1 uW	1 uW	1 uW

Como também era esperado, a substituição do LPC pelos blocos FFT, FILTER e MFCC acarretará o aumento do consumo de todo o sistema de verificação de locutor utilizando coeficientes MFCC para a extração das características vocais. Isto acontece pelo fato do aumento da quantidade dos recursos a serem utilizados, aumento na complexidade das operações presentes nesses 3 (três) blocos e o aumento da quantidade de memórias RAM necessárias para armazenar o peso dos filtros e os valores do cosseno para o cálculo da DCT.

Mais informações sobre os relatórios gerados pela ferramenta *dc_shell* são apresentadas no Anexo B.

5 Considerações Finais e Sugestões para Pesquisas Futuras

O objeto de estudo desta pesquisa era avaliar a eficiência dos coeficientes MFCC na representação das características de um locutor. Esta avaliação foi feita por meio de uma comparação com os coeficientes LPC para representação das características de um locutor. Em Fehine et al. (2010) existe uma implementação da extração de características com coeficientes LPC, o módulo responsável pela obtenção destes coeficientes foi substituído por 3 (três) blocos que possibilitaram a extração dos coeficientes MFCC.

Neste capítulo, são apresentadas as conclusões obtidas no estudo das configurações do SRL desenvolvido, além de destacadas sugestões para estudos futuros.

5.1 Contribuições

A partir dos resultados obtidos, podem-se destacar as contribuições apresentadas a seguir.

- Estudo inicial comparativo do custo de utilizar coeficientes MFCC como representação de características de um locutor.
- Desenvolvimento de um método para auxiliar na determinação dos limiares do detector de voz para uma dada base de dados.
- Análise dos parâmetros do detector de voz, que possibilitou verificar que o bom desempenho do sistema está associado aos ajustes dos limiares de energia e de tempo. A eficiência na detecção da voz proporcionou também uma redução no consumo energético, uma vez que os períodos de inatividade de voz, tais como, silêncio e ruído de fundo, deixam de ser processados.
- Construção de uma base de dados, composta de 30 locutores (15 indivíduos do gênero masculino e 15 do gênero feminino), constituindo 9.300 elocuições.
- Utilização de menor quantidade de filtros para modelar os coeficientes MFCC. Com a redução do número de filtros necessários, sem diminuição das

métricas de desempenho, é possível uma redução no consumo energético, uma vez que uma quantidade menor de informação precisa ser processada.

- Indicação da configuração de sistema que maximizou a quantidade de acertos, dentre as configurações analisadas.
- Comparação do desempenho do sistema desenvolvido com o *software* apresentado por Fachine et al. (2010). A escolha dessa pesquisa se deu pelo fato de estar validado em FPGA e prototipado em silício.

Apesar do aumento no consumo de energia devido à complexidade adicionada e a quantidade de recursos necessários para realizar a implementação em *hardware* dos coeficientes MFCC, ocorre um ganho na taxa de reconhecimento. E para aplicações que visam a segurança de informações e a restrição de ambientes, a taxa de acertos do sistema deve ser um fator preponderante na escolha do método de extração de característica a ser utilizado.

Diante do exposto, pode-se concluir que as simplificações adotadas com vistas à implementação em *hardware*, aliadas à análise criteriosa para as escolhas de limiares para detecção de voz (eliminação de intervalos de silêncio) e quantidade de filtros para obtenção dos coeficientes MFCC foram eficientes.

5.2 Sugestões para Pesquisas Futuras

Como sugestões para pesquisas futuras, tem-se a finalização da implementação em linguagem de descrição de *hardware* do sistema apresentado. Essa implementação deverá levar em consideração todas as sugestões apresentadas nesta pesquisa e deverá ser realizada a verificação funcional.

Realizar uma investigação com outros métodos de extração de características, tais como, variações dos coeficientes mel-cepstrais e cepstrais; para o cálculo da distorção na construção do dicionário do quantizador, tal como, distância de Manhattan; e para a construção do classificador, a exemplo de HMM e redes neurais artificiais, levando-se em consideração os requisitos de baixo consumo energético. Também é indicada uma análise quantitativa do consumo com a implementação das técnicas propostas.

Uma sugestão adicional para pesquisa futura, é realizar um estudo comparativo com outras implementações de banco de filtros.

6 Referências Bibliográficas

AMUDHA, V et al. SOC Implementation of HMM Based Speaker Independent Isolated Digit Recognition System. In: VLSI DESIGN, 2007. HELD JOINTLY WITH 6TH INTERNATIONAL CONFERENCE ON EMBEDDED SYSTEMS, 20TH INTERNATIONAL CONFERENCE ON, 2007, Bangalore. **Anais...** Bangalore: IEEE, p.848-853.

AMUDHA, V.; VENKATARAMANI, B.; MANIKANDAN, J. FPGA implementation of isolated digit recognition system using modified back propagation algorithm. In: INTERNATIONAL CONFERENCE ON ELECTRONIC DESIGN, 2008, Penang. **Anais...** Penang: IEEE, p.1-6.

BEIGI, H. **Fundamentals of Speaker Recognition**. Boston, MA: Springer US, 2011. 942 p.

BERGERON, J. **Writing Testbenches: Functional Verification of HDL Models**. 2 ed., Boston, MA: Springer, 2003, 512 p.

CAEIROS, A. M.; MIYATAKE, M. N; MEANA, H. P. Isolate Speech Recognition Based on Time-Frequency Analysis Methods. In: CORROCHANO, E. B.; EKLUNDH, J. O. **Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications**. Boston, MA: Springer, 2009. cap. 6, p. 297-304.

CAMPBELL, J. P. **Speaker Recognition: A Tutorial**. Proceedings of the IEEE, v. 85, n. 9, p. 1437-1462, set. 1997.

CARDOSO, D. P. **Identificação de Locutor Usando Modelos de Mistura de Gaussianas**. 2009, 88 f. Dissertação (Mestrado em Engenharia) – Departamento de Engenharia de Sistemas Eletrônicos, Universidade de São Paulo, São Paulo, 2009.

CARVALHO, L. M. P. **Sistema de Verificação do Orador, Baseado em Modelos de Markov, Compactado num Objecto COM para Windows**. 2007, 134 f. Dissertação (Mestrado em Engenharia Electrônica Industrial e Computadores) – Departamento de Electrônica Industrial, Universidade do Minho, Guimarães, Portugal. 2007.

CHEN, Fang. **Designing Human Interface in Speech Technology**. Boston: MA: Springer US, 2005. 382 p.

CIPRIANO, J. L. G. **Desenvolvimento de Arquitetura para Sistemas de Reconhecimento Automático de Voz Baseados em Modelos Ocultos de Markov**. 2001, 125 f. Tese (Doutorado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2001.

CIPRIANO, J. L. G.; NUNES, R. P.; BARONE, D. A. C. Implementation of Voice Processing Algorithms in FPGA Hardware. IX WORKSHOP IBERCHIP, 2003, La Habana. **Anais...** La Habana: IBERCHIP, 2003.

DA SILVA, A. G. **Reconhecimento de Voz para Palavras Isoladas**. 2009, 60 f. Estudo de Conclusão de Curso (Graduação em Engenharia da Computação) - Centro de Informática, Universidade Federal de Pernambuco, Recife. 2009.

DE LIMA, C. B. **Sistema de Verificação de Locutor Independente do Texto baseados em GMM e Ar-Vetorial utilizando PCA**. 2001, 126 f. Dissertação (Mestrado em Engenharia Elétrica) - Programa de Pós-Graduação em Engenharia Elétrica, Instituto Militar de Engenharia, Rio de Janeiro. 2001.

DEY, S. et al. Speech biometric based attendance system In: *TWENTIETH NATIONAL CONFERENCE ON COMMUNICATIONS (NCC)*, 2014, Kanpur. **Anais...** Kanpur: IEEE, 2014 p.1,6.

DIAS, D. **Desenvolvimento de um IP Core de Pré-processamento Digital de Sinais de Voz para Aplicação em Sistema Embutidos**. 2006, 108 f. Dissertação (Mestrado em Ciência da Computação) - Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Campina Grande, Campina Grande. 2006.

DIAS, D. **Reconhecimento de Fala Contínua para o Português Brasileiro em Sistemas Embarcados**. 2001, 178 f. Tese (Doutorado em Engenharia Elétrica) - Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande. 2011.

EHKAN, P.; ALLEN, T.; QUIGLEY, S. FPGA Implementation for GMM-Based Speaker Identification. **International Journal of Reconfigurable Computing**, v. 2011, n. , p. 1-8, 2011.

FARAH, S.; SHAMIM, A. Speaker recognition system using mel-frequency cepstrum coefficients, linear prediction coding and vector quantization. *3RD INTERNATIONAL CONFERENCE ON COMPUTER, CONTROL & COMMUNICATION (IC4)*, 2013, Karachi, **Anais...** Karachi: IEEE, 2013 p.1-5.

FECHINE, J. **Reconhecimento automático de identidade vocal utilizando modelagem híbrida: Paramétrica e Estatística**. 2000, 212 f. Tese (Doutorado em Engenharia Elétrica) - Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande. 2000.

FECHINE et al. SPVR: An IP core for Real-Time Speaker Verification. In: *CONFERENCE PROCEEDINGS OF IP BASED ELECTRONICS SYSTEM CONFERENCE & EXIBITION*, 2010, Grenoble. **Anais...** Grenoble: Conference Proceedings.

GARIMELLA, S. et al. Regularized Auto-Associative Neural Networks for Speaker Verification. **IEEE Signal Processing Letters**, v. 19, n. 12, p. 841-844, out. 2012.

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. **Anais...** Gujarat: IEEE, 2013 p.170-173.

HAYVES, M. **Digital Signal Processing**. New York: McGraw-Hill, 1999. 462 p.

HE, J.; LIU, L.; PALM, G. A discriminative training algorithm for VQ-based speaker identification, In: **IEEE Transactions on Speech and Audio Processing**, v.7, n.3, p. 353-356, maio 1999

HERBIG, T.; GERL, F.; MINKER, W. Self-learning speaker identification for enhanced speech recognition. **Computer Speech and Language**, v. 26, n. 3, p. 210-227, dez. 2012.

JINGHONG, L.; YANAN, T.; LIJIA, Z. Research and implementation of speaker recognition algorithm based on FPGA. In: 24TH CHINESE CONTROL AND DECISION CONFERENCE (CCDC), 2012, Taiyuan. **Anais...** Taiyuan: IEEE, 2012 p. 1155-1158.

KRISHNA, R.; MAHLKE, S.; AUSTIN, T. Memory system design space exploration for low-power, real-time speech recognition. In: INTERNATIONAL CONFERENCE ON HARDWARE/SOFTWARE CODESIGN AND SYSTEM SYNTHESIS, 2004, **Anais...** IEEE publisher, 2004 p. 140-145.

LAMAS, R. **Avaliação de Codificadores de Voz em Ambiente VoIP**. 2005, 90 f. Estudo de Conclusão de Curso (Graduação em Engenharia) – Departamento de Engenharia Eletrônica e de Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

LI, M.; PROUDFOOT, J.; KING, G. ASIC implementation of vector quantization for speaker recognition. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1995, Detroit. **Anais...** Detroit: IEEE, 1995 p. 2667-2670.

LI et al. Robust endpoint detection and energy normalization for real-time speech and speaker recognition. **IEEE Transactions on Speech and Audio Processing**, v.10, n.3, p. 146-157, mar. 2002.

LI, Q. **Speaker Authentication**. Berlin: Springer, 2012. 238 p.

LIU, X.; CHEUNG, Y. A multi-boosted HMM approach to lip password based speaker verification. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2012, Kyoto. **Anais...** Kyoto: IEEE, 2012 p. 2197-2200.

MAMEDE et al. **Computational Processing of the Portuguese Language**. Faro: Springer Verlag, 2003. 272 p.

MARPLE, S. **Digital Spectral Analysis: With Applications**. Englewood Cliffs: Prentice-Hall, 1987. 492 p.

MARTINEZ et al. Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques. In: 22ND INTERNATIONAL CONFERENCE ON ELECTRICAL COMMUNICATIONS AND COMPUTERS (CONIELECOMP), 2012, Cholula. **Anais...** Cholula: IEEE, 2012 p. 248-251.

MARWEDEL, P. **Embedded System Design**. 2nd edition Amsterdã: Springer Netherlands, 2011. 400 p.

MEYER-BAESE, U. **Digital signal processing with field programmable gate arrays**. Berlin: Springer, 2014. 930 p.

NEDEVSKI, S.; PATRA, R.; BREWER, E. Hardware speech recognition for user interfaces in low cost, low power devices. In: 42ND DESIGN AUTOMATION CONFERENCE, 2005. **Proceedings of IEEE**, 2005 p. 684-689.

O'SHAUGHNESSY, D. **Speech Communications: Human and Machine**. 2. ed. Wiley-IEEE Press, 2000. 548 p.

O'SHAUGHNESSY, D. Interacting with computers by voice: automatic speech recognition and synthesis. **Proceedings of IEEE**, 2003 v.91 n.9, p.1272-1305.

OLIVEIRA, H. F. A. **BVM: Reformulação da metodologia de verificação funcional VeriSC**. 2010, 139 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Campina Grande, Campina Grande, 2010.

PETRY et al. SISLOC project: Hardware-based speaker verification system. In: INTERNATIONAL CONFERENCE ON INTELLIGENT AND ADVANCED SYSTEMS, 2007, Kuala Lumpur. **Anais...** Kuala Lumpur: IEEE, 2007 p. 1293-1295.

PETRY, A.; ZANUZ, A.; BARONE, D. A. C. Utilização de Técnicas de Processamento Digital de Sinais para a Identificação Automática de Pessoas pela Voz. In: Simpósio sobre Segurança em Informática, 1999, São José dos Campos. **Anais...** São José dos Campos: SSI99, 1999.

POLLACK, I.; PICKETT, J.; SUMBY, W. On the Identification of Speakers by Voice. **The Journal of the Acoustical Society of America**, v. 26, n. 3 p. 403-406, jul. 1954.

RABINER, L.; SCHAFER, R. **Digital processing of speech signals**. New York: Prentice Hall, 1978. 512 p.

RABINER, L.; JUANG, B. **Fundamentals of Speech Recognition**. New York: Prentice Hall, 1993. 507 p.

RABINER, L.; SCHAFER, R. **Theory and Applications of Digital Speech Processing**. Prentice Hall, 2010. 1056 p.

RAMOS-LARA et al. Real-Time Speaker Verification System Implemented on Reconfigurable Hardware. **Journal of Signal Processing Systems**, v. 71, n.2 p. 89-103, jun. 2012.

RAMOS-LARA et al. SVM speaker verification system based on a low-cost FPGA. In: INTERNATIONAL CONFERENCE ON FIELD PROGRAMMABLE LOGIC AND APPLICATIONS, 2009, Prague. **Anais...** Prague: IEEE, 2009 p.582-586.

SAAMBHAVI, V.; RAO, S.; RAJALAKSHMI, P. Design of feature extraction circuit for speech recognition applications. TENCON 2012 - 2012 IEEE REGION 10 CONFERENCE, 2012, Cebu. **Anais...** Cabu: IEEE, 2012 p. 1-5.

SADJADI, S.; HANSEN, J. Blind Spectral Weighting for Robust Speaker Identification under Reverberation Mismatch. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 22, n. 5, p. 937-945, maio 2014.

SELVAN, K.; JOSEPH, A.; ANISH K. Speaker recognition system for security applications. In: IEEE RECENT ADVANCES IN INTELLIGENT COMPUTATIONAL SYSTEMS, 2013, Trivandrum. **Anais...** Trivandrum: IEEE, 2013, p. 26-30.

SHEARME, J.; HOLMES, J. An experiment Concerning the Recognition of Voices. **Language and Speech Journal**, v. 2, n. 3, p. 123-131, jul./set. 1959.

SILVA, K. Uma Metodologia de Verificação Funcional para Circuitos Digitais. 2007, 121 f. Tese (Doutorado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande. 2007.

SLANEY, M. *Auditory Toolbox*. Palo Alto: Interval Research Corporation. Disponível em: <<https://engineering.purdue.edu/~malcolm/interval/1998-010/>>. , 1998.

STAWORKO, M.; RAWSKI, M. FPGA implementation of feature extraction algorithm for speaker verification. In: 17TH INTERNATIONAL CONFERENCE MIXED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, 2010, Warsaw. **Anais...** Warsaw: IEEE, 2010 p. 557-561.

WALIA, M. Discrete Fractional Fourier Transform and Vector Quantization Based Speaker Identification System. In: FOURTH INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING & COMMUNICATION TECHNOLOGIES, 2014, Rohtak. **Anais...** Rohtak: IEEE, 2014 p. 459-463.

WANG et al. Hardware/software co-design for fast-trainable speaker identification system based on SMO. In: INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS, 2011, Anchorage. **Anais...** Anchorage: IEEE, 2011 p. 1621-1625.

Apêndice A Resultados do Processamento dos dados

Nesta pesquisa, foram simuladas 5.000 configurações do SRL agrupadas em 8 processamentos, cada processamento com 625 configurações. Essas configurações foram obtidas com variações dos limiares do módulo de detecção de voz, como também, da alternância entre o uso de coeficientes LPC e mel-cepstrais. Não houve alterações nos demais módulos do SVL para a realização das simulações, esses módulos foram utilizados conforme descrito no Capítulo 3.

Nas Tabelas A.1, A.2, A.3, A.4, A.5, A.6, A.7 e A.8, são apresentados os resultados para cada um dos 8 processamentos no formato de uma matriz de ordem 25, tal que, cada elemento da matriz corresponde à métrica QE de uma dada configuração. As linhas e colunas destas matrizes representam os limiares TI e TF, respectivamente, do módulo de detecção de voz. Nos 8 processamentos, esses limiares de tempo variam de 1 a 25 quadros²³.

Os resultados apresentados nas Tabelas A.1, A.2, A.3 e A.4 foram gerados com coeficientes mel-cepstrais, nas demais tabelas desse apêndice utilizou-se os coeficientes LPC. Quanto aos limiares de energia do DV, os resultados das Tabelas A.1, A.5, A.4 e A.8 foram gerados com limiares iguais, sendo que, para os resultados das duas primeiras tabelas utilizou-se os limiares EI e EF iguais a 0,02449, enquanto que, para as duas últimas tabelas esses limiares assumiram o valor 0,007760. Para as Tabelas A.2 e A.6 empregou-se limiares de 0,02449 e 0,046433 para EI e EF, respectivamente. Nas Tabelas A.3 e A.7, EI e EF assumiram, respectivamente os valores 0,046433 e 0,02449.

²³ Cada quadro de entrada do módulo DV corresponde a um trecho de voz de aproximadamente 10ms.

Tabela A.1: Resultados do Processamento 1²⁴.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	32	28	29	27	24	24	21	20	14	20	21	24	23	22	23	23	22	22	22	22	23	22	22	22	22
2	31	30	30	27	25	25	21	21	20	22	21	21	25	22	21	21	20	20	20	20	20	20	20	20	20
3	34	30	30	28	24	24	23	22	22	23	20	21	24	22	21	21	21	21	21	21	21	21	21	21	21
4	35	34	34	30	29	26	26	25	25	24	24	27	25	25	25	25	25	25	25	25	25	25	25	25	25
5	36	33	34	31	28	26	28	28	25	26	25	25	28	26	26	26	26	26	26	26	26	26	26	26	26
6	35	32	33	30	28	26	28	29	25	26	25	25	28	26	26	26	26	26	26	26	26	26	26	26	26
7	36	33	33	30	28	26	28	29	25	26	25	25	28	26	26	26	26	26	26	26	26	26	26	26	26
8	36	35	34	31	30	26	28	29	27	27	26	26	29	27	27	27	27	27	27	27	27	27	27	27	27
9	32	32	32	28	26	27	25	27	26	27	26	26	29	27	27	27	27	27	27	27	27	27	27	27	27
10	31	29	32	28	25	26	26	27	26	27	26	27	29	27	27	27	27	27	27	27	27	27	27	27	27
11	31	33	29	28	25	26	27	28	26	27	26	26	29	27	27	27	27	27	27	27	27	27	27	27	27
12	33	32	33	30	26	24	26	27	26	26	25	25	28	26	26	26	26	26	26	26	26	27	27	27	27
13	31	34	34	31	27	26	26	27	25	26	25	25	28	26	26	26	26	26	26	26	26	26	26	26	26
14	40	41	44	39	35	33	35	36	33	35	34	33	36	34	33	33	33	33	33	33	33	33	33	33	33
15	45	44	45	40	36	32	35	36	32	34	33	32	35	33	32	32	32	32	32	32	32	32	32	32	32
16	45	45	45	41	37	33	34	35	32	34	33	32	35	33	32	32	32	32	32	32	32	32	32	32	32
17	42	45	43	42	42	37	38	41	38	39	38	37	40	38	37	37	37	37	37	37	37	37	37	37	37
18	42	44	43	42	42	38	39	42	39	40	37	37	40	38	38	36	36	36	36	36	36	37	37	37	37
19	39	40	39	41	42	39	40	42	39	41	38	38	41	39	39	37	37	37	37	37	37	38	38	38	38
20	38	38	38	39	40	39	40	41	39	41	38	38	41	39	39	37	37	37	37	37	37	38	38	38	38
21	38	39	40	42	41	40	40	41	38	40	38	38	41	39	39	37	37	37	37	37	37	38	38	38	38
22	38	41	41	41	41	41	41	41	38	40	38	38	41	39	39	37	37	37	37	37	37	38	38	38	38
23	42	41	42	43	43	41	41	42	38	40	39	39	43	40	40	38	38	38	38	38	38	39	39	39	39
24	46	46	47	47	45	47	48	47	43	46	44	46	44	43	41	41	41	41	41	41	41	42	42	42	42
25	45	47	48	47	45	47	48	47	45	47	46	43	47	45	44	42	42	42	42	42	42	43	43	43	43

²⁴ Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,02449.

Tabela A.2: Resultados do Processamento 2²⁵.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	29	34	34	34	26	25	26	27	25	27	31	30	29	28	29	28	28	28	28	28	28	28	28	28	28
2	36	34	34	34	26	25	27	28	26	28	31	31	30	29	30	29	29	29	29	29	29	29	29	29	29
3	38	36	35	35	27	26	26	28	26	28	31	31	30	29	30	29	29	29	29	29	29	29	29	29	29
4	38	34	33	32	28	26	26	28	27	28	30	29	30	29	30	29	29	29	29	29	29	29	29	29	29
5	38	33	33	32	27	27	25	27	26	27	29	28	29	28	29	28	28	28	28	28	28	28	28	28	28
6	40	38	33	32	28	26	25	27	26	27	29	28	29	28	29	28	28	28	28	28	28	28	28	28	28
7	38	36	30	29	27	27	24	26	25	24	26	25	26	25	26	25	25	25	25	25	25	26	26	26	26
8	41	37	32	33	27	26	26	27	27	25	27	26	27	26	27	26	26	26	26	26	29	31	31	31	31
9	41	43	41	42	41	38	36	37	36	37	40	40	38	37	37	37	37	37	37	37	38	39	39	39	39
10	43	40	42	45	44	43	41	40	40	40	40	41	41	40	40	40	40	40	40	40	40	41	41	41	41
11	46	44	43	48	47	43	48	45	43	43	42	43	43	42	41	42	42	42	42	42	42	42	42	42	42
12	46	46	50	51	49	51	50	50	47	47	46	47	47	46	46	47	47	47	47	47	46	46	46	46	46
13	68	69	66	66	72	70	68	67	64	62	62	62	62	61	60	61	61	61	61	61	62	62	62	62	62
14	91	92	86	87	90	93	93	93	91	87	87	88	87	85	84	86	86	86	86	86	86	85	85	85	85
15	131	129	129	131	134	134	136	134	130	128	127	127	127	125	123	124	124	124	124	124	124	126	126	126	126
16	206	197	201	195	195	195	195	195	191	190	189	186	185	184	183	184	184	184	184	184	187	191	191	191	192
17	266	262	262	264	256	255	256	256	250	251	251	249	248	248	247	248	247	247	247	247	247	248	248	248	248
18	377	371	369	371	357	354	353	353	350	350	350	348	346	346	344	345	345	345	345	345	344	346	346	346	346
19	496	488	483	488	472	467	465	464	461	461	461	459	458	458	456	457	457	457	457	457	457	459	459	459	459
20	626	616	606	610	598	593	591	590	587	587	587	585	584	584	583	584	584	584	584	584	584	586	586	586	586
21	784	776	764	764	755	747	741	740	737	737	737	735	734	734	733	734	734	734	734	734	734	736	736	736	736
22	950	947	934	935	924	917	910	909	906	906	906	904	903	903	902	903	903	903	903	903	903	905	905	905	906
23	1122	1117	1104	1104	1097	1088	1083	1081	1078	1079	1078	1076	1074	1074	1073	1074	1074	1074	1074	1074	1074	1074	1074	1074	1076
24	1337	1326	1320	1314	1314	1302	1296	1299	1296	1294	1293	1290	1289	1289	1287	1287	1286	1286	1286	1286	1286	1286	1286	1286	1290
25	1575	1566	1558	1556	1558	1545	1541	1542	1538	1537	1536	1534	1532	1532	1531	1531	1530	1530	1530	1530	1530	1530	1530	1530	1533

²⁵ Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,02449 e 0,046433, respectivamente.

Tabela A.3: Resultados do Processamento 2²⁶.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	50	50	41	45	33	31	26	26	29	27	23	29	25	21	20	21	20	22	21	21	21	21	21	24	24
2	46	42	40	40	32	30	27	25	28	24	22	27	23	23	20	21	20	20	21	22	21	21	21	21	22
3	37	39	38	44	36	32	25	26	25	25	24	23	23	22	22	21	22	22	22	23	23	22	22	22	23
4	40	39	42	39	34	28	27	25	27	23	23	23	26	27	25	25	25	26	27	28	27	26	27	25	25
5	43	41	40	42	35	36	31	26	29	25	26	27	27	27	25	25	25	26	29	28	27	27	27	27	27
6	37	38	38	38	34	29	26	24	28	25	27	27	26	27	25	25	25	25	28	28	27	27	27	27	27
7	38	37	34	42	34	32	26	23	29	25	26	27	27	27	26	27	25	25	28	27	27	27	27	27	27
8	34	35	37	42	34	31	25	25	29	26	27	28	27	28	27	27	25	27	28	28	28	28	28	28	28
9	41	37	31	42	33	32	25	28	30	25	24	25	24	24	25	23	21	23	25	25	25	25	25	25	25
10	39	41	36	30	31	28	32	28	30	24	24	25	23	27	23	23	22	23	24	24	24	24	24	24	24
11	41	36	34	33	31	31	29	29	28	24	25	25	25	28	24	24	24	23	25	25	25	25	25	25	25
12	39	40	32	33	35	33	33	31	30	27	27	25	25	26	24	25	24	23	25	25	25	25	25	25	25
13	32	28	33	26	34	32	35	34	31	26	25	24	25	26	26	25	24	24	26	26	26	26	27	27	27
14	41	39	35	40	41	42	45	38	41	34	31	30	30	32	32	30	29	30	32	32	32	33	32	31	31
15	49	46	44	45	48	46	42	41	44	34	34	31	32	34	31	30	29	30	32	32	32	32	32	32	32
16	45	46	40	44	49	44	46	42	42	34	33	31	33	36	32	31	30	31	33	33	33	33	33	33	33
17	45	43	46	42	48	51	47	43	43	40	36	35	36	39	36	35	34	35	37	37	37	37	37	37	37
18	44	44	47	45	48	47	41	40	42	39	37	37	38	40	38	37	35	36	38	38	38	38	38	38	38
19	40	42	45	44	43	45	45	41	45	40	36	36	37	40	41	40	38	39	41	41	41	41	41	41	41
20	39	42	46	43	45	45	43	42	42	38	39	39	36	39	42	40	38	37	39	39	39	39	41	39	39
21	40	42	46	45	42	44	44	44	43	39	36	39	40	41	42	40	39	38	40	40	40	40	41	40	40
22	42	44	42	44	42	42	41	43	43	40	36	38	41	45	43	42	40	38	40	40	40	40	40	40	40
23	42	47	49	44	47	45	42	42	43	40	39	41	43	43	43	43	41	39	41	41	41	41	41	41	41
24	53	57	55	50	49	47	47	51	47	48	42	45	45	46	41	43	42	41	43	43	43	43	43	43	43
25	58	56	51	50	47	45	46	49	48	48	43	42	46	44	46	45	42	43	45	45	45	45	46	45	45

²⁶ Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,046433 e 0,02449, respectivamente.

Tabela A.4: Resultados do Processamento 2²⁷.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	45	43	43	43	36	31	33	30	30	33	29	28	29	32	29	29	28	30	29	27	27	27	29	31	32
2	42	40	35	43	35	24	34	30	33	28	29	27	29	31	31	30	29	30	29	27	27	27	29	31	31
3	43	41	36	42	34	23	34	30	32	29	31	29	29	31	31	30	29	30	29	27	27	27	29	31	31
4	44	42	39	41	35	24	34	29	32	29	31	29	29	31	30	29	29	30	29	27	27	27	29	31	31
5	43	41	40	43	36	25	31	32	34	29	31	29	28	31	29	29	29	30	29	27	27	27	29	31	31
6	45	44	41	42	36	27	35	34	37	28	26	24	25	28	24	25	25	28	27	25	25	25	27	29	29
7	43	38	39	38	36	30	33	33	40	30	27	26	26	29	24	25	25	27	26	24	24	24	26	28	28
8	43	42	41	35	38	27	31	30	31	29	28	28	28	31	26	27	27	29	28	26	27	27	28	30	30
9	41	45	41	37	41	35	45	44	45	41	39	38	37	38	37	38	38	37	37	35	36	35	34	37	37
10	53	48	45	45	43	40	47	45	46	44	44	43	41	42	43	42	42	40	40	38	40	40	40	42	42
11	71	65	59	54	51	47	52	55	50	47	45	44	41	42	43	45	45	44	44	42	43	43	43	45	45
12	86	79	77	59	56	58	63	61	54	50	48	45	46	46	46	43	43	45	45	43	43	43	44	46	46
13	105	98	89	87	84	84	85	83	73	72	70	70	67	64	65	64	63	62	62	60	59	59	59	61	61
14	119	113	101	101	97	96	101	105	98	97	96	97	96	94	91	91	91	89	89	87	87	87	86	88	88
15	167	162	155	151	152	143	144	143	143	142	140	140	135	132	133	132	133	134	134	132	132	132	133	133	133
16	232	230	223	218	214	212	217	214	213	209	206	206	201	197	198	194	194	196	196	194	195	195	196	196	196
17	310	301	297	286	277	275	287	285	280	271	269	266	267	262	262	261	262	261	260	259	260	260	257	256	256
18	422	412	405	396	385	379	387	388	382	377	374	372	370	363	364	363	362	359	358	356	357	357	355	355	355
19	540	529	524	518	505	497	500	504	497	492	491	487	486	481	482	480	479	477	476	474	473	472	470	469	469
20	667	656	649	641	632	623	621	619	622	622	622	617	613	611	609	606	605	603	601	599	598	597	595	593	593
21	824	810	807	798	789	781	777	777	777	776	776	770	765	763	759	757	755	753	751	749	747	746	744	742	742
22	992	980	976	970	963	955	947	947	946	947	947	941	935	933	930	928	926	924	922	920	918	917	915	914	914
23	1154	1144	1137	1136	1126	1117	1115	1116	1112	1113	1114	1109	1102	1099	1097	1095	1094	1091	1089	1088	1086	1086	1084	1084	1084
24	1363	1353	1348	1346	1339	1332	1330	1327	1324	1326	1322	1318	1316	1311	1307	1307	1306	1306	1304	1302	1300	1299	1297	1297	1297
25	1605	1596	1592	1590	1582	1574	1576	1570	1567	1569	1565	1562	1560	1555	1549	1549	1548	1548	1546	1545	1544	1543	1542	1542	1542

²⁷ Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,046433.

Tabela A.5: Resultados do Processamento 2²⁸.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	146	149	152	158	161	158	172	176	191	188	189	192	198	201	198	199	200	200	197	198	198	197	199	199	201
2	139	148	157	157	160	160	171	178	183	187	187	187	191	192	194	195	194	196	196	195	195	199	202	202	204
3	140	149	155	157	156	159	176	175	180	186	187	188	192	194	196	196	193	195	194	193	193	197	199	199	201
4	138	142	155	155	156	163	174	171	181	185	184	186	191	193	195	194	193	195	195	194	194	198	200	200	204
5	142	140	153	155	158	165	178	176	181	187	186	188	192	195	197	196	195	198	198	197	197	202	204	204	208
6	135	147	153	157	161	163	173	177	182	189	189	191	195	198	198	199	198	199	198	197	198	203	205	205	209
7	133	147	152	156	161	164	172	177	182	189	189	191	195	198	198	199	198	199	198	197	198	203	205	205	209
8	131	146	152	155	158	163	169	174	181	190	189	191	195	198	198	201	200	200	199	198	199	203	205	205	209
9	136	147	154	149	154	160	170	177	180	189	188	189	193	196	196	198	197	198	197	197	198	203	204	204	208
10	128	138	155	151	158	160	173	178	180	189	188	189	193	196	196	198	197	198	197	197	198	203	204	204	208
11	133	136	149	149	158	157	172	177	180	189	188	189	193	196	196	198	197	198	197	197	198	203	204	204	208
12	133	138	151	145	153	156	171	176	180	189	187	188	192	196	196	198	197	198	197	197	198	203	204	204	208
13	128	134	143	142	148	153	172	177	181	188	186	187	191	195	195	197	196	197	196	196	197	202	203	203	207
14	120	125	137	135	140	157	174	187	196	202	202	204	207	213	214	216	215	215	214	214	215	220	221	221	225
15	126	129	138	129	147	160	185	189	196	203	200	205	209	215	216	217	216	218	217	217	218	223	224	224	228
16	128	134	140	135	148	161	179	189	197	205	203	202	204	209	209	210	209	211	209	209	210	215	216	216	220
17	127	138	143	142	151	163	183	192	201	209	206	205	207	211	211	213	213	215	213	213	215	220	221	221	225
18	133	136	135	141	150	159	183	194	206	212	213	211	212	216	214	217	216	217	215	216	218	223	224	224	228
19	131	135	141	143	156	164	185	196	206	212	213	211	211	215	215	216	217	218	216	217	219	224	225	225	229
20	129	134	130	141	150	159	191	191	203	209	211	209	209	212	212	216	217	218	216	217	219	224	225	225	230
21	128	137	129	141	156	160	182	194	205	209	212	215	216	220	222	225	226	227	225	226	227	232	233	233	237
22	134	144	131	140	159	164	183	190	202	204	208	210	210	215	217	218	219	220	219	220	221	226	227	227	231
23	130	138	134	135	151	155	173	188	197	200	204	203	205	207	207	207	208	209	207	208	208	213	214	214	216
24	135	143	133	133	149	155	171	187	190	193	200	203	206	208	211	212	213	214	212	212	212	217	219	219	220
25	138	142	136	139	151	155	174	186	186	190	193	197	199	200	204	204	205	205	203	203	203	208	210	211	213

²⁸ Neste experimento, utilizaram-se coeficientes LPC e limiares de energia de início e fim iguais a 0,02449.

Tabela A.6: Resultados do Processamento 2²⁹.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	127	139	138	146	149	159	171	183	186	190	199	201	204	208	206	205	204	206	207	206	209	216	214	214	212
2	129	137	133	145	145	156	167	184	191	191	200	204	207	211	209	209	208	210	211	210	213	218	216	216	214
3	132	137	132	144	140	151	168	187	190	193	200	204	206	209	207	207	206	208	209	208	211	216	214	215	213
4	129	134	134	141	136	152	166	185	190	190	198	204	206	210	208	207	206	208	209	208	211	216	214	214	212
5	135	141	134	142	142	152	167	185	190	190	198	204	206	210	208	207	206	208	209	208	211	216	214	214	212
6	131	134	132	137	141	150	166	181	187	190	196	202	204	209	209	206	205	207	208	207	210	216	214	214	212
7	127	120	125	135	141	150	164	178	185	188	194	199	201	205	205	202	201	203	204	203	206	212	210	210	208
8	134	133	133	128	139	152	164	173	184	185	192	198	200	205	205	202	201	203	204	203	206	212	210	210	207
9	129	136	133	129	140	156	161	182	190	191	198	201	203	209	209	207	206	208	209	208	211	214	212	212	211
10	125	128	127	128	130	140	152	179	189	193	198	200	202	207	206	207	206	209	210	209	209	213	211	212	211
11	127	126	137	135	138	141	158	186	195	198	199	202	205	211	210	211	210	213	214	213	213	216	214	215	214
12	127	130	134	126	140	145	155	174	185	186	186	187	188	192	194	192	191	193	194	193	199	206	204	205	205
13	145	145	139	136	146	158	165	192	197	199	201	201	201	205	204	204	203	204	205	204	208	216	216	217	217
14	164	164	162	161	171	182	196	222	228	233	237	231	233	238	234	236	234	235	235	236	240	247	245	246	246
15	204	203	207	205	212	227	235	264	271	270	275	283	283	287	285	286	286	288	288	287	292	296	296	297	298
16	266	271	273	274	277	292	299	322	329	329	336	344	344	345	342	345	343	344	343	343	351	356	355	356	358
17	314	311	325	311	320	337	352	372	376	379	382	388	388	393	393	392	392	393	392	391	400	405	406	407	409
18	412	407	405	402	413	429	445	464	473	480	482	481	481	481	483	483	483	483	483	482	487	490	491	492	493
19	528	517	520	516	539	546	560	581	592	596	597	599	600	600	602	602	602	602	601	600	605	609	610	612	613
20	654	642	640	638	660	677	680	701	717	721	721	725	725	726	726	728	728	728	727	726	732	738	739	740	741
21	809	803	803	806	810	827	829	846	862	865	868	874	875	876	875	881	880	881	880	880	883	887	888	889	890
22	992	983	982	989	992	1009	1009	1024	1036	1037	1040	1044	1046	1046	1044	1046	1045	1045	1043	1043	1046	1050	1051	1052	1053
23	1151	1151	1147	1151	1160	1178	1178	1191	1207	1208	1210	1212	1214	1213	1211	1213	1212	1213	1212	1212	1216	1223	1224	1225	1226
24	1366	1368	1369	1371	1375	1386	1391	1402	1417	1421	1423	1424	1426	1424	1422	1424	1424	1425	1424	1424	1427	1434	1435	1436	1437
25	1612	1611	1613	1608	1611	1623	1619	1632	1645	1650	1652	1654	1656	1654	1652	1653	1654	1655	1657	1657	1659	1665	1666	1667	1668

²⁹ Neste experimento, utilizaram-se coeficientes LPC e limiares de energia de início e fim iguais a 0,02449 e 0,046433, respectivamente.

Tabela A.7: Resultados do Processamento 2³⁰.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	167	151	141	131	134	143	153	144	163	170	171	165	168	167	166	173	174	173	173	171	177	171	168	172	169
2	146	136	137	124	118	137	141	137	155	164	171	167	164	163	169	169	167	169	171	166	172	166	171	167	165
3	140	131	127	128	118	135	145	141	154	163	164	169	160	157	167	170	168	166	167	170	163	167	174	172	174
4	140	140	127	123	120	135	142	141	153	166	163	169	163	161	169	170	168	168	167	171	170	172	175	174	175
5	140	131	136	126	128	144	143	140	148	166	170	161	161	161	166	169	164	168	169	172	173	173	175	173	173
6	137	125	125	119	122	136	138	139	157	161	163	162	162	157	168	173	168	167	172	171	173	173	174	173	174
7	145	123	129	122	127	131	140	143	155	162	172	164	160	164	169	174	171	168	170	173	175	174	176	175	175
8	141	141	128	127	126	131	151	148	151	164	173	163	159	167	170	170	165	170	173	172	172	172	174	173	174
9	149	136	136	129	128	138	145	142	152	163	166	165	164	168	170	173	167	171	172	173	174	174	174	174	173
10	141	143	137	130	126	145	139	143	150	164	172	167	166	165	170	173	169	170	172	172	174	173	175	174	174
11	145	135	132	128	132	137	141	142	154	159	171	165	162	167	169	173	169	169	172	171	172	173	174	173	171
12	141	128	130	127	134	141	138	145	155	161	167	162	164	164	170	174	170	169	171	170	172	172	173	170	170
13	134	123	121	126	128	144	139	144	154	157	165	166	162	163	168	174	169	167	169	169	170	170	170	169	168
14	123	124	130	129	132	141	138	147	155	164	174	172	174	175	177	183	178	177	178	177	178	176	178	176	175
15	130	126	134	133	130	136	140	148	156	165	177	177	184	180	182	187	183	179	181	180	179	179	180	178	179
16	130	138	136	137	134	134	140	148	152	163	175	187	177	180	183	186	181	177	177	175	176	174	175	175	175
17	131	144	149	135	129	138	145	151	160	169	183	195	180	183	185	188	184	180	180	179	179	178	180	179	179
18	148	137	138	138	137	140	148	151	163	175	184	188	182	183	185	189	186	179	180	179	179	179	181	180	180
19	144	138	142	139	140	142	151	166	166	179	185	190	185	185	186	191	186	182	182	181	182	181	183	181	181
20	142	144	144	141	148	148	158	168	166	182	185	202	180	182	184	188	185	181	182	181	182	181	180	179	180
21	148	143	139	142	152	145	154	165	162	179	180	190	182	179	187	190	185	183	184	183	184	183	182	181	182
22	144	140	132	133	153	150	148	165	160	174	183	192	184	181	189	190	185	183	184	183	184	184	182	180	182
23	130	130	129	128	149	143	146	156	164	176	177	186	176	179	180	179	175	173	173	172	173	173	171	172	173
24	128	133	132	129	158	146	149	166	167	176	177	183	175	179	181	178	176	172	172	171	172	172	170	171	172
25	144	129	130	137	155	151	154	163	169	176	180	188	180	176	181	179	178	174	174	173	175	174	173	172	171

³⁰ Neste experimento, utilizaram-se coeficientes LPC e limiares de energia de início e fim iguais a iguais a 0,046433 e 0,02449, respectivamente.

Tabela A.8: Resultados do Processamento 2³¹.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	142	133	139	139	143	147	141	157	161	172	171	173	175	175	176	174	169	172	173	172	173	174	176	177	176
2	136	136	141	145	150	154	144	149	159	171	172	180	177	175	176	175	171	176	177	176	177	179	181	181	180
3	130	138	141	147	150	151	136	145	157	168	170	181	178	174	175	175	171	175	176	175	176	176	178	180	179
4	129	136	135	144	152	148	143	146	159	173	175	180	175	175	177	176	171	175	175	174	175	178	180	180	179
5	130	129	136	129	142	143	139	143	161	173	175	180	176	173	176	175	170	175	175	174	175	178	180	180	179
6	133	135	135	130	140	141	140	147	162	173	176	178	179	175	176	179	174	176	177	176	178	177	179	177	176
7	141	145	137	127	138	153	150	153	160	171	171	175	178	173	172	175	169	171	172	171	173	172	173	171	170
8	135	148	137	139	145	145	151	157	160	175	173	176	178	174	173	177	171	172	173	172	174	173	174	172	171
9	138	141	137	138	145	152	150	155	160	173	171	174	175	173	171	175	169	172	173	172	174	173	175	173	172
10	130	134	133	136	134	138	140	149	164	170	175	168	171	168	171	177	172	174	175	174	176	175	176	174	173
11	135	140	146	134	135	142	141	154	160	171	179	171	175	173	170	172	167	169	171	170	171	170	171	169	168
12	154	146	150	149	146	146	144	152	161	174	177	169	175	173	173	174	170	170	172	170	172	171	172	170	169
13	161	163	170	161	159	162	160	168	171	178	183	181	184	180	179	179	177	176	177	175	177	176	177	175	175
14	180	186	183	175	177	182	174	188	197	204	201	195	196	193	196	193	191	191	193	191	193	196	197	197	199
15	225	225	222	214	218	224	220	221	231	240	241	241	248	240	245	242	240	242	243	245	246	250	251	249	251
16	296	289	291	279	286	286	285	285	306	313	309	320	313	306	311	308	308	307	308	310	312	313	314	312	314
17	372	368	354	349	352	347	334	337	346	360	356	362	367	357	358	351	352	353	353	358	360	365	366	366	367
18	460	446	447	446	443	448	426	434	445	441	440	448	456	453	454	452	452	454	455	455	457	459	459	457	459
19	561	545	554	550	548	556	537	540	533	552	561	564	578	578	578	569	572	571	572	571	574	574	573	572	574
20	679	672	671	674	677	677	657	659	666	681	690	696	702	703	703	698	701	699	700	699	702	704	704	703	703
21	835	829	829	830	837	833	827	820	831	845	845	841	850	848	850	849	851	849	851	850	853	853	853	852	851
22	1012	1008	1006	1009	1013	1011	1001	993	1011	1021	1013	1012	1022	1021	1022	1018	1020	1017	1018	1016	1020	1019	1019	1018	1017
23	1179	1177	1176	1175	1171	1176	1169	1169	1186	1193	1181	1181	1192	1193	1195	1189	1189	1187	1187	1185	1188	1188	1188	1187	1186
24	1386	1392	1387	1388	1385	1385	1379	1379	1395	1407	1390	1388	1399	1400	1403	1400	1401	1397	1397	1396	1397	1397	1396	1395	1394
25	1631	1627	1625	1622	1626	1623	1621	1621	1640	1643	1619	1621	1626	1633	1633	1634	1635	1632	1632	1630	1632	1632	1632	1631	1630

³¹ Neste experimento, utilizaram-se coeficientes LPC e limiares de energia de início e fim iguais a iguais a 0,046433.

Apêndice B Análise Estatística

Nesta seção, apresenta-se a análise estatística dos experimentos apresentados na Tabela 4.4. Um dos objetivos desta análise é identificar os limiares de detecção de voz que minimizam a métrica QE.

Na Tabela B.1, é apresentado um resumo das medidas estatísticas dos processamentos. As medidas foram: medidas de tendência central (média e mediana); valores extremos (mínimo e máximo), medidas de espalhamento (quartis: Q1, Q3) e dispersão (desvio padrão, σ).

Tabela B.1: Medidas resumo da métrica QE para os 8 experimentos.

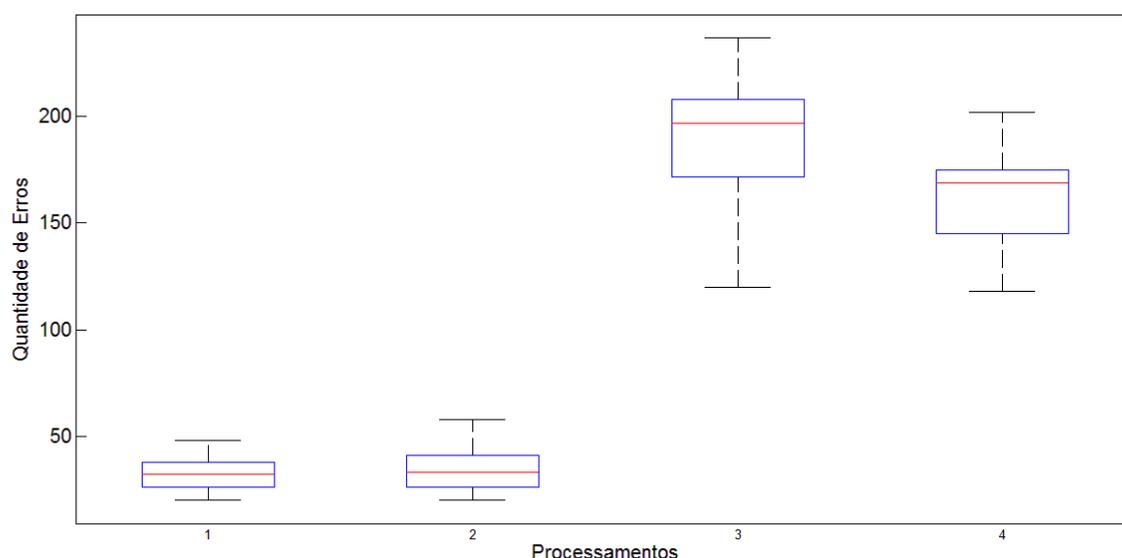
Experimento	Mínimo	Q1	Mediana	Q3	Máximo	Média	σ
E1	20	26	32	38	48	32,27	7,18
E2	24	30	62	465,50	1575	324,09	439,7
E3	20	26	33	41	58	33,99	8,34
E4	23	34	72	501	1605	335,85	447,31
E5	120	171,75	197	208	237	188,20	27,03
E6	120	199	214	602	1668	459,86	426
E7	118	145	169	175	202	162,08	18,37
E8	127	171	180	573,25	1643	439,3	427,82

Na Tabela B.1, as colunas "Mínimo" e "Máximo" representam a menor e a maior métrica QE, respectivamente, dentre as 9.000 elocuições de teste que compõem as 625 configurações de um dado processamento; as colunas "Q1", "Q2" e "Q3" correspondem à realização QE que ocupa a posição um quarto, um meio e três quartos, respectivamente, das 9.000 elocuições de teste das 625 configurações ordenadas de um dado processamento; as colunas "Média" e " σ " apresentam, respectivamente, a média aritmética e o desvio padrão das métricas QE das 9.000 elocuições de teste das 625 configurações de um dado processamento.

A partir dos resultados obtidos pelos processamentos e apresentados na Tabela B.1, foram construídos gráficos do tipo *boxplot*, conforme Figura B.1 e Figura B.2. A partir destes gráficos, é possível evidenciar outras informações e chegar a novas conclusões que não são visíveis por meio de tabelas.

Com as informações da Tabela B.1 e sua representação gráfica na Figura B.1, percebe-se que o processamento com extração de características LPC, com menor quantidade de erros, é E7.

Figura 6.1: Boxplot dos experimentos com as maiores quantidades de erros (E1 – E3 – E5 – E7).

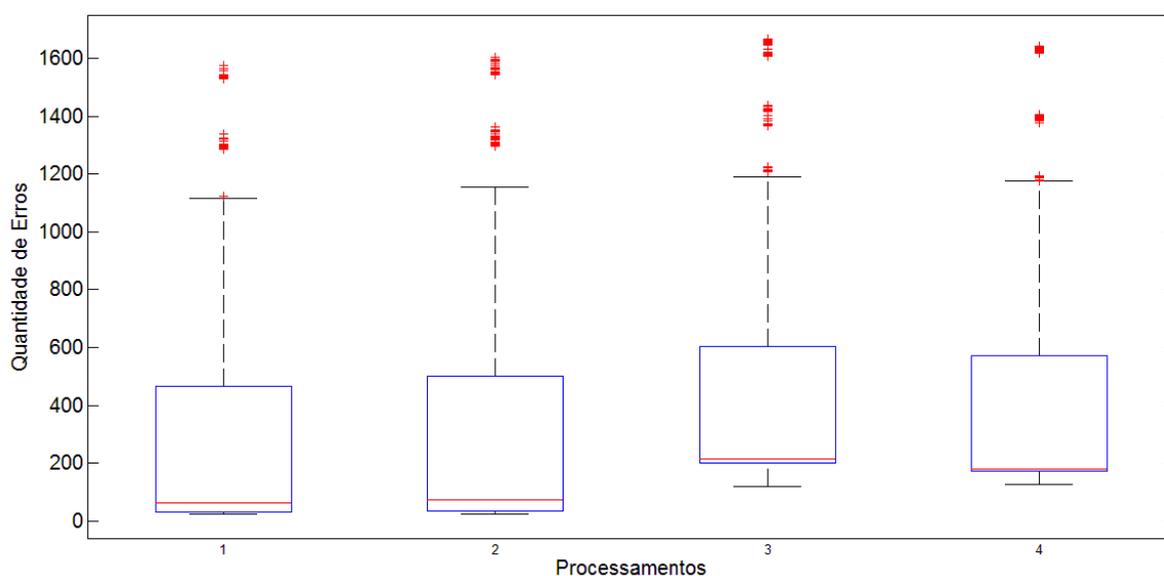


O experimento E7 contém as menores medidas de tendência central, valores extremos e espalhamento, em relação aos outros experimentos com coeficientes LPC. Os experimentos E5 e E6 apresentam o valor mínimo (120) muito próximo do experimento E7 (118), no entanto o valor máximo de E5 (237) e E6 (1.668) são maiores. Além disso, os experimentos E5 ($\sigma=27,03$) e E6 ($\sigma=426$) exibem valores de desvio padrão mais elevados que E7 ($\sigma=18,37$), ou seja, os valores apresentados pelos experimentos E5 e E6 estão mais espalhados por uma gama de valores e, conseqüentemente, mais afastados da média.

Em relação aos experimentos que utilizaram os coeficientes mel-cepstrais como vetor de características, os valores mínimos são bem próximos, E1 (20), E2 (24), E3 (20) e E4 (23). Contudo, os experimentos E2 e E4 apresentam valores máximos elevados 1.575 e 1.605, respectivamente. Assim como descrito anteriormente, valores elevados de desvio padrão E2 ($\sigma=439,7$) e E4 ($\sigma=447,31$) indicam que os valores estão mais afastados da média, ou seja, estão mais dispersos por uma gama de valores. Os experimentos E1 e E3 apresentam resultados semelhantes: Q1 (26;26), Mediana (32;33), Média (32,27;33,99), Q3 (38;41). Entretanto, a quantidade máxima de erros de E1 (48) é menor que a E3

(58) e E1 é constituído por valores menos dispersos ($\sigma=7,18$) quando comparado com E3 ($\sigma=8,34$).

Figura 6.2: Boxplot dos experimentos com as menores quantidades de erros (E2 – E4 – E6 – E8).



A separação dos gráficos em duas figuras (Figura B.1 e Figura B.2), tem como finalidade melhorar a visualização dos gráficos, pois se os 8 *boxplots* estivessem na mesma figura, os que apresentam valores baixos, ficariam achatados.

Conforme a Tabela B.1, e as Figuras B.1 e B.2, os experimentos escolhidos, um com vetor de características mel-cepstrais e outro LPC, foram: E1 e E7. Esses processos apresentam os seguintes arranjos dos limiares de energia para a detecção de voz, 0.002449 (limiar de energia inicial) e 0.002449 (limiar de energia final) para experimento E1 (vide Tabela 4.3 e Tabela 4.4) e 0.046433 (limiar de energia inicial) e 0.002449 (limiar de energia final) para experimento E7 (vide Tabela 4.3 e Tabela 4.4). Após apresentação e análise dos resultados de cada experimento, esses limiares de energia são os mais indicados para detectar os trechos com atividade de voz e, assim, reduzir a métrica QE.

Conforme mencionado anteriormente, os coeficientes mel-cepstrais são mais robustos à presença de ruídos, não necessitando de limiares de energia com valores elevados. Para os coeficientes LPC ocorrem a necessidade de extrair uma maior quantidade de ruído para se iniciar detecção de voz e assim fornecer uma menor quantidade de erros.

Após a definição dos limiares de energia, faz-se necessário analisar a influência dos limiares de tempo no reconhecimento. Conforme descrito na Seção 4.1.2, os limiares TI e TF foram avaliados para períodos de tempo de 10 ms a 0,25 s (ou 1 a 25 quadros). Nas Figuras B.3 e B.4, apresenta-se uma comparação dos experimentos com extração de características mel-cepstrais, sob a influência da variação de TI e TF, respectivamente, na quantidade de erros.

Conforme as informações apresentadas na Figura B.3, percebe-se que a medida que TI aumenta a quantidade de erros também aumenta. Esse fato fica bem evidente para os experimentos E2 e E4. Para os experimentos E1 e E3, a métrica QE apresenta, durante o intervalo de 1 a 13, os valores mínimos e um ligeiro aumento dessa métrica. Para o intervalo de 14-25 acontecem os maiores valores para métrica QE e bem concentrados próximo ao valor máximo da QE para E1(48) e E3(58), conforme Tabela B.1. Então, a escolha para um bom valor de TI deve estar no intervalo compreendido entre 1-13 quadros.

Em se tratando do limiar de tempo final (Figura B.4), não se percebem variações expressivas de QE nos experimentos quando da variação de TF. No entanto, os menores valores para a métrica QE estão concentrados entre dois intervalos 8-11 e 17-25. Portanto, a escolha do limiar de tempo final (TF) deve ser um valor presente em um desses intervalos.

É importante destacar, que a escolha de um bom par de limiares não só vai impactar na taxa de reconhecimento, como também em uma futura implementação de *hardware*. A escolha de um par com valores elevados acarretará em uso de memória adicional para o armazenamento dos sinais de voz.

Figura 6.3: Comparação dos experimentos com extração de características mel-cepstrais sob a influência da variação de TI na métrica QE.

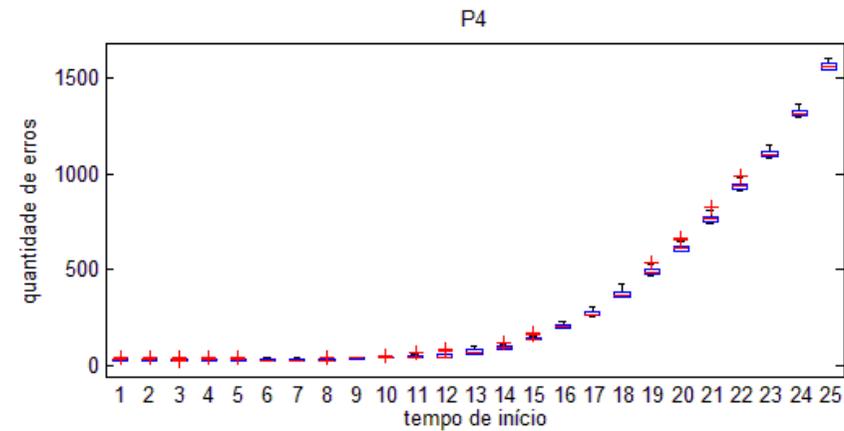
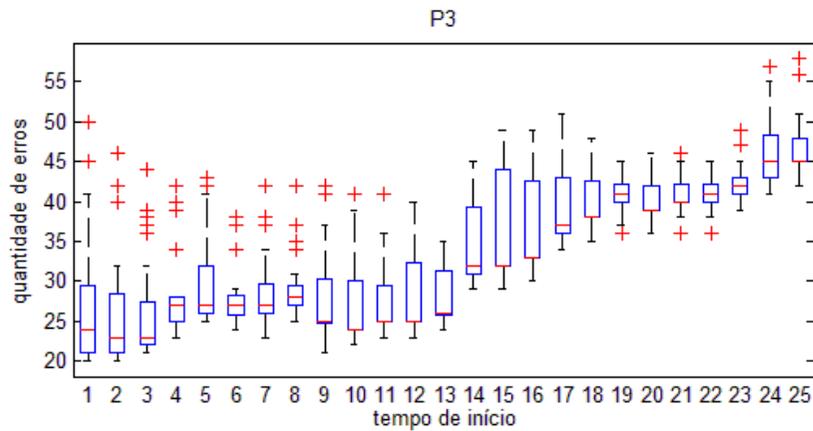
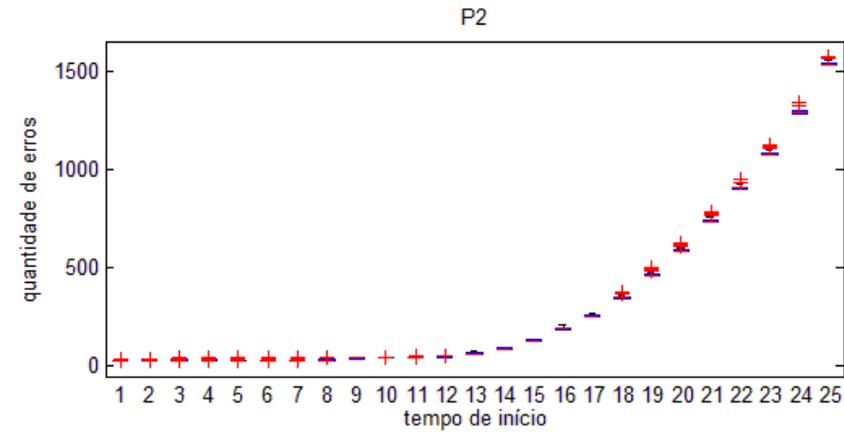
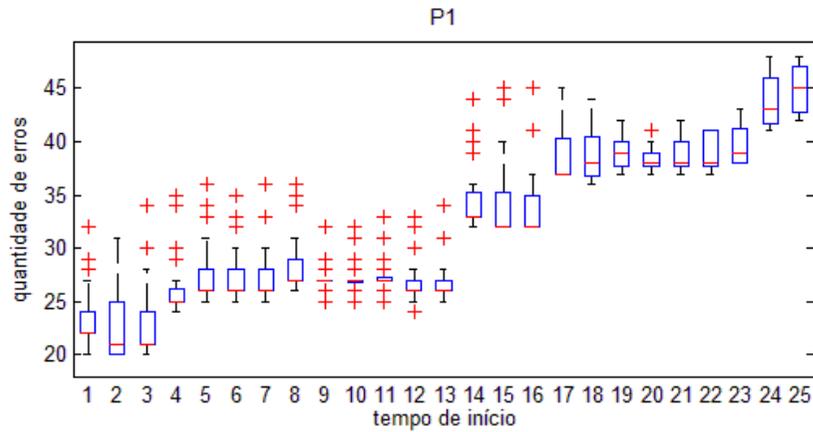


Figura 6.4: Comparação dos experimentos com extração de características mel-cepstrais sob a influência da variação de TF na métrica QE.

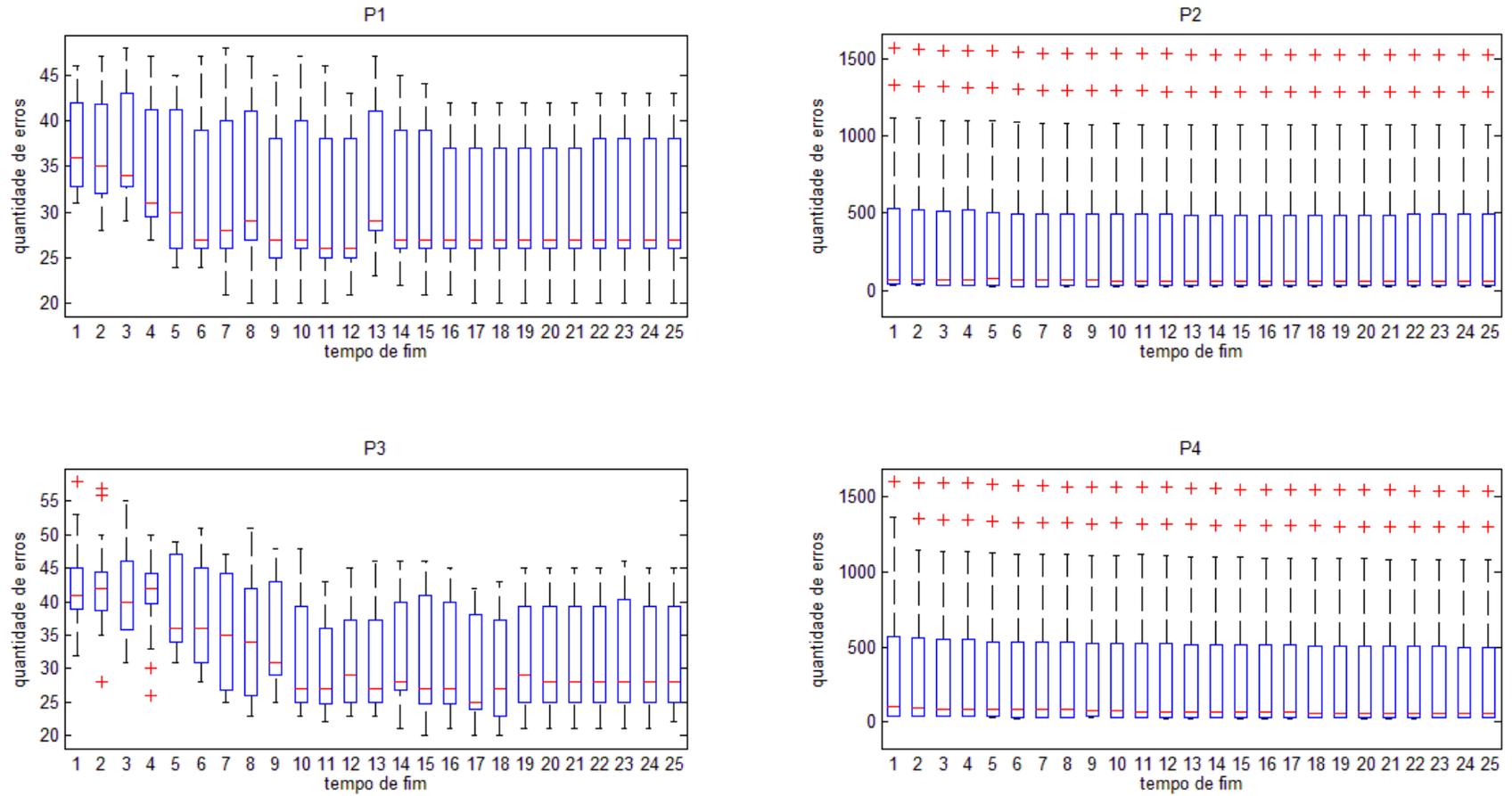


Figura 6.5: Comparação dos experimentos com extração de características LPC sob a influência da variação de TI na métrica QE.

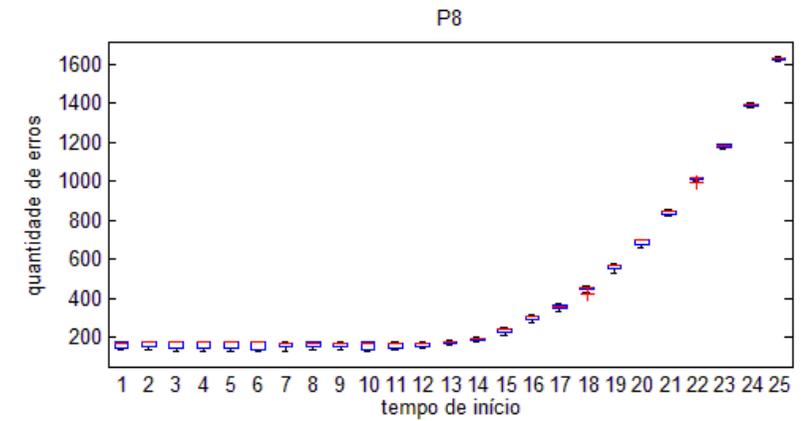
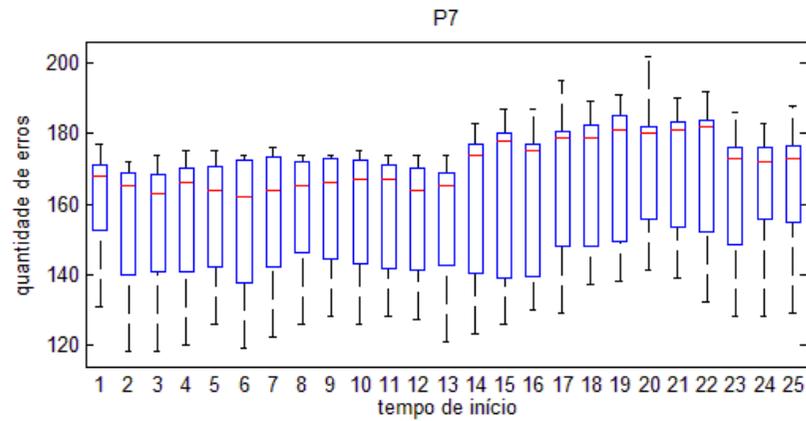
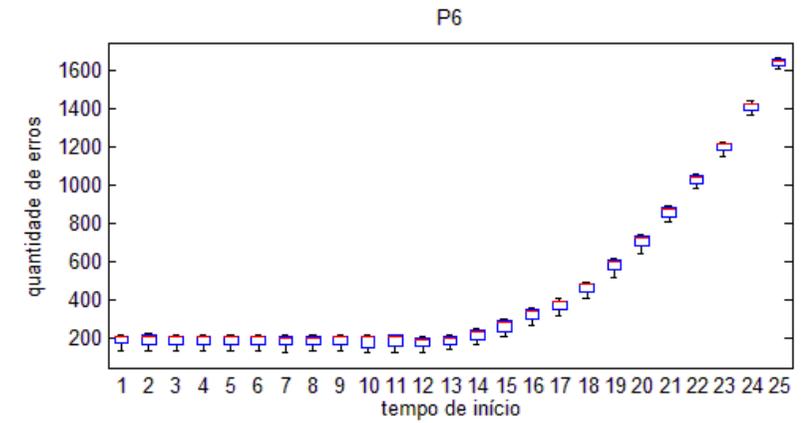
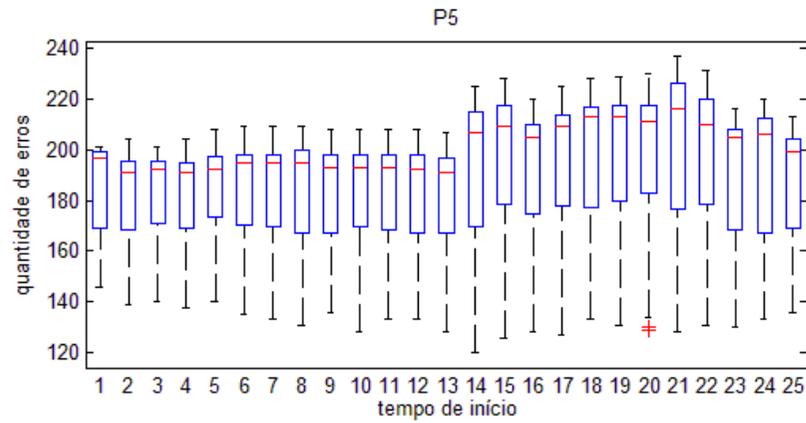
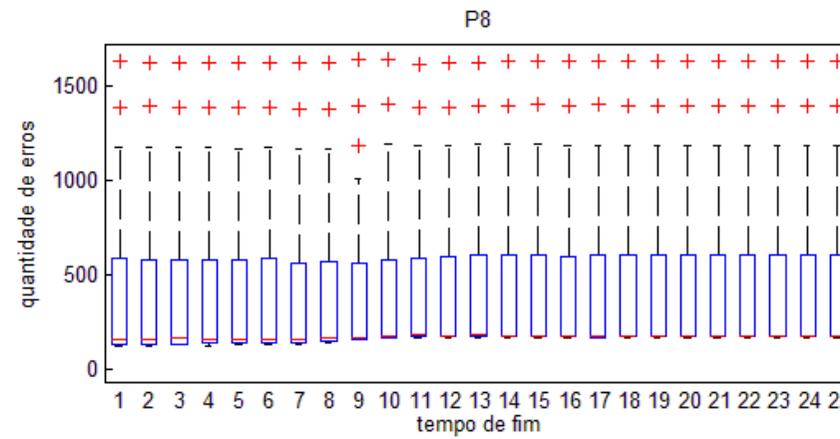
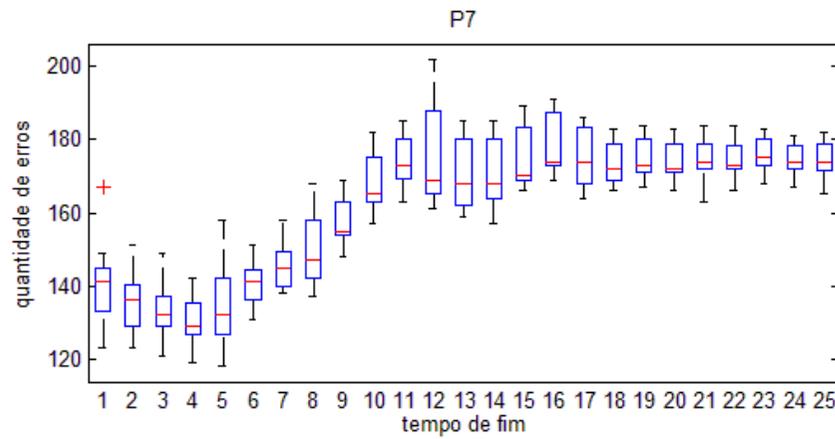
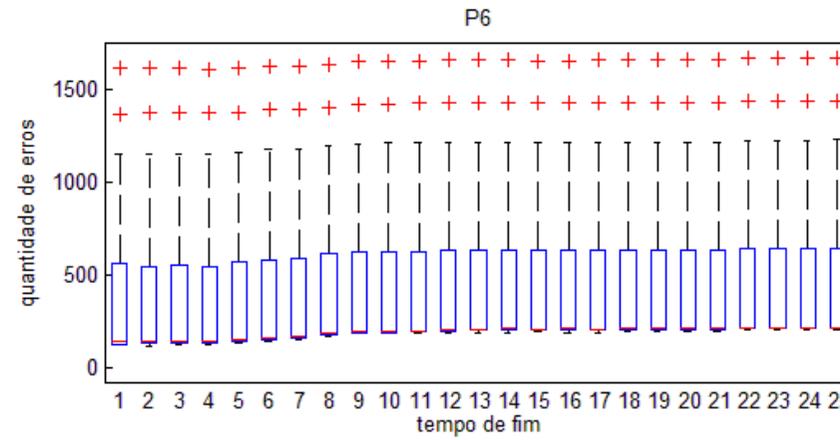
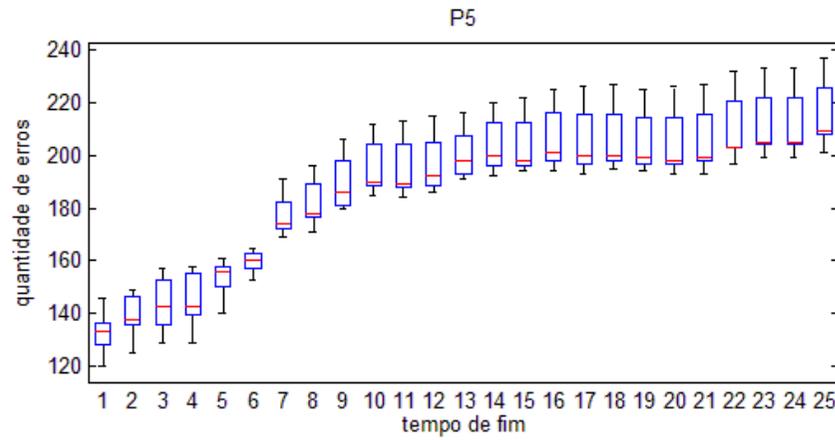


Figura 6.6: Comparação dos experimentos com extração de características LPC sob a influência da variação de TF na métrica QE.



Apêndice C Resultados das Saídas como Desconhecido

A seguir, são apresentadas a quantidade de saídas como desconhecido em cada experimento.

Tabela C.1: Resultados do Experimento E1³².

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	71	72	69	74	75	73	82	72	71	75	72	70	71	71	70	72	73	73	73	73	74	75	76	76	75
2	68	77	69	74	71	69	85	72	75	75	74	74	70	71	74	74	76	75	75	75	78	78	79	79	78
3	68	72	69	72	70	76	82	73	75	75	74	73	69	70	73	73	74	73	73	73	76	76	76	76	76
4	60	75	65	70	68	72	77	68	72	72	73	75	70	70	72	72	73	72	72	72	75	76	76	76	76
5	61	76	63	69	69	72	75	66	72	70	70	72	67	67	69	69	70	69	69	69	72	73	73	73	73
6	64	77	62	72	70	72	78	65	71	70	70	72	67	67	69	69	69	69	69	69	72	73	73	73	73
7	60	77	63	69	69	77	76	65	71	70	70	72	67	67	69	69	69	69	69	69	72	73	73	73	73
8	61	68	65	71	67	68	79	64	69	70	70	72	67	67	69	69	69	69	69	69	72	73	73	73	73
9	65	71	66	70	71	69	79	62	67	68	68	69	64	64	66	66	66	66	66	66	67	69	70	69	69
10	64	72	67	66	73	66	78	62	66	67	67	68	63	63	65	65	65	65	65	66	68	69	68	68	68
11	70	69	70	67	71	65	74	60	65	66	66	67	62	62	64	64	64	64	64	64	65	67	68	67	67
12	68	75	67	67	69	65	75	63	67	69	69	69	64	64	66	66	66	66	66	66	67	69	68	67	67
13	71	70	65	65	67	59	70	58	62	62	62	63	58	58	60	60	60	60	60	61	63	63	62	62	62
14	59	58	63	62	64	60	69	61	65	64	64	66	61	61	64	64	64	64	64	65	67	67	66	66	66
15	52	55	64	59	60	59	62	53	59	58	58	60	55	55	58	58	58	58	58	59	60	60	59	59	59
16	57	52	64	59	60	54	61	56	58	57	57	59	54	54	57	57	57	57	57	58	58	58	57	57	57
17	64	59	69	66	57	54	62	53	53	52	52	54	49	49	52	52	52	52	52	53	54	54	53	53	53
18	62	57	70	67	57	56	62	52	52	51	53	53	49	50	51	53	53	53	53	53	53	51	51	51	51
19	62	57	68	62	57	55	61	52	52	50	52	52	48	49	50	52	52	52	52	52	52	50	50	50	50
20	60	63	70	63	57	57	59	51	53	51	53	53	49	50	50	52	52	52	52	52	52	50	50	50	50
21	61	59	68	64	56	56	63	50	54	51	52	52	49	50	50	52	52	52	52	52	52	50	50	50	50
22	69	63	69	63	59	60	60	55	54	53	53	52	49	50	51	53	53	53	53	53	54	52	52	52	52
23	70	68	73	72	66	66	61	62	62	59	58	60	52	55	56	58	58	58	58	58	59	57	57	57	58
24	67	64	72	67	68	69	58	58	60	59	60	60	58	59	61	64	64	64	64	64	65	63	63	63	63
25	68	68	71	74	69	70	64	63	60	61	59	64	60	60	62	64	65	65	65	65	66	64	64	64	62

³² Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,02449.

Tabela C.2: Resultados do Experimento E2³³.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	68	68	75	75	76	82	80	74	79	82	78	77	78	78	79	79	79	79	79	79	78	79	79	79	78
2	60	63	67	73	74	78	76	76	82	83	80	79	80	80	81	82	82	82	82	82	81	82	82	82	81
3	56	60	63	69	71	76	77	75	81	82	79	78	78	79	80	81	81	81	81	81	80	81	81	81	80
4	53	57	64	69	67	73	74	74	78	80	77	77	76	77	78	79	79	79	79	79	79	80	80	80	79
5	54	63	61	72	63	65	73	74	77	79	76	76	75	76	77	78	78	78	78	78	78	79	79	79	78
6	53	60	70	68	67	69	77	76	79	80	78	78	75	76	77	78	78	78	78	78	78	79	79	79	78
7	50	54	69	69	66	64	76	74	77	82	80	80	77	78	79	80	80	80	80	80	80	80	80	80	79
8	50	57	68	66	65	66	76	75	79	80	78	79	77	78	79	80	81	81	81	81	78	76	76	76	75
9	51	56	56	57	53	56	70	69	72	69	67	68	68	69	71	71	72	72	72	72	71	71	71	71	70
10	62	63	65	63	60	51	64	63	68	66	68	65	64	65	66	66	67	67	67	67	68	67	67	67	66
11	62	62	72	68	62	59	60	61	67	67	69	67	65	66	67	65	65	65	65	65	65	66	66	66	65
12	70	69	72	77	62	55	64	62	69	70	69	70	68	70	70	69	69	69	69	69	71	72	72	72	72
13	68	58	76	73	56	54	67	65	71	68	65	65	66	66	68	68	67	67	67	67	66	68	68	68	67
14	71	67	68	72	66	59	68	61	70	73	77	75	75	77	81	78	78	78	78	78	78	80	80	80	81
15	76	67	68	68	62	66	69	65	72	70	73	72	71	73	76	74	74	74	74	74	74	72	72	72	70
16	71	68	68	74	69	75	79	68	76	75	78	79	79	80	83	81	81	81	81	81	80	77	77	77	76
17	77	71	76	75	85	85	84	79	81	82	84	82	82	81	83	83	84	84	84	84	84	81	81	81	82
18	67	65	67	65	78	81	77	79	78	79	80	78	81	80	82	83	82	82	82	82	83	81	81	81	80
19	70	63	73	70	81	79	82	85	84	84	84	82	84	83	85	86	85	85	85	85	85	83	83	83	82
20	73	59	70	66	76	78	75	79	76	78	77	78	80	78	79	80	79	79	78	78	77	77	77	77	74
21	69	59	72	63	76	72	71	76	74	77	76	73	76	74	75	76	75	75	74	74	73	73	73	73	73
22	75	65	71	63	72	68	63	71	70	70	71	70	71	70	71	73	72	72	71	71	70	70	70	70	68
23	65	58	66	59	57	56	53	64	62	61	63	63	65	64	65	67	66	66	66	66	65	67	67	67	66
24	57	62	58	56	51	57	49	53	53	56	59	60	60	60	63	65	65	65	65	65	64	66	66	66	62
25	58	60	59	63	55	60	44	46	46	50	54	54	55	55	57	59	60	60	60	60	60	60	60	60	57

³³ Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,02449 e 0,046433, respectivamente.

Tabela C.3: Resultados do Experimento E3³⁴.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	107	91	71	74	65	68	73	73	68	62	64	62	65	78	80	80	81	79	81	83	79	78	78	78	79
2	84	85	76	83	76	73	71	77	67	71	64	65	69	74	81	77	83	83	83	83	84	81	78	81	79
3	78	68	83	75	71	72	73	66	74	71	63	62	67	73	78	83	81	84	84	85	82	80	80	80	79
4	73	82	80	69	72	78	74	75	69	71	69	69	70	71	77	82	79	79	79	79	77	77	78	80	80
5	81	80	85	75	65	70	77	76	71	71	67	66	67	74	81	85	82	83	81	83	82	80	80	80	80
6	87	76	76	76	66	82	81	81	79	71	71	68	77	77	83	87	84	86	85	83	82	81	80	80	79
7	81	76	78	72	72	81	88	84	74	69	71	69	73	78	83	83	83	86	83	84	82	81	80	80	78
8	79	77	75	69	71	71	85	83	77	71	66	67	68	76	84	84	84	84	82	84	80	80	78	80	79
9	68	75	78	67	64	78	85	87	72	68	69	67	72	78	84	87	85	82	80	83	80	79	77	80	79
10	74	67	76	76	72	75	83	85	72	70	69	64	72	76	87	85	82	81	82	84	78	78	76	79	78
11	70	76	78	79	69	73	81	82	74	71	67	66	70	74	83	82	82	82	81	82	77	76	76	77	76
12	72	82	74	77	70	72	76	79	73	66	65	69	65	75	76	78	77	78	77	80	76	75	76	77	76
13	84	76	86	79	68	73	69	76	68	64	66	66	67	71	74	81	79	78	76	77	77	75	75	74	73
14	87	79	76	71	71	75	68	74	60	60	64	59	67	71	75	79	77	77	76	77	76	75	77	79	78
15	74	72	71	69	75	70	70	73	57	63	60	62	63	68	73	80	75	74	72	73	75	74	72	74	73
16	73	70	82	77	79	69	58	61	54	59	60	59	67	66	78	78	75	72	70	71	72	71	70	71	70
17	75	75	78	78	83	72	69	70	61	60	61	63	64	66	72	73	69	66	65	66	67	66	65	67	66
18	72	71	77	73	79	74	73	76	66	61	63	66	66	66	70	69	68	66	64	65	65	65	64	65	64
19	70	74	70	75	80	68	62	71	60	61	67	63	65	65	68	67	68	65	63	64	64	62	60	62	61
20	73	70	73	77	68	64	63	71	68	64	63	59	68	69	66	68	68	70	67	68	66	64	62	65	65
21	78	77	77	66	71	69	62	73	65	58	68	62	63	66	65	66	67	69	66	66	64	62	61	63	63
22	76	78	74	76	73	73	71	74	59	57	67	65	58	60	63	65	65	69	66	66	64	62	62	62	62
23	74	72	67	74	70	68	75	70	65	61	68	64	61	64	65	65	68	69	66	66	65	65	65	66	66
24	73	73	72	72	70	70	65	68	64	56	65	62	60	63	72	72	74	75	72	72	70	70	69	69	69
25	65	65	77	75	69	60	64	69	62	60	66	65	63	68	70	69	71	71	68	68	67	67	66	66	66

³⁴ Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,046433 e 0,02449, respectivamente.

Tabela C.4: Resultados do Experimento E4³⁵.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	82	78	70	69	69	77	69	65	55	57	62	60	65	73	77	75	74	71	71	73	72	72	70	68	67
2	87	77	76	74	76	82	71	67	53	61	59	60	67	72	75	72	71	69	69	71	70	70	71	68	68
3	85	77	76	73	75	79	74	66	53	59	56	56	67	71	75	71	69	67	67	70	69	69	69	68	68
4	81	83	83	73	77	73	73	64	53	59	56	57	67	70	75	72	69	68	68	71	70	70	69	67	67
5	83	86	84	76	72	70	68	56	51	62	60	60	70	69	76	75	72	70	69	71	70	70	70	68	68
6	80	81	75	64	77	73	63	55	46	60	66	66	72	74	80	81	78	75	75	77	75	75	75	73	73
7	76	74	71	65	64	66	62	58	43	58	64	65	71	70	79	79	76	74	74	76	74	74	74	72	72
8	78	69	72	72	68	73	77	58	54	60	64	67	69	71	75	76	73	72	72	75	72	72	71	69	69
9	68	66	71	65	68	66	67	60	54	52	56	57	63	68	65	66	64	65	66	68	65	66	68	66	66
10	73	71	81	72	75	76	74	65	61	49	56	59	60	61	62	64	63	63	61	63	60	60	63	59	59
11	83	79	86	82	84	81	81	67	67	61	59	65	68	71	72	70	69	67	67	68	66	66	67	66	66
12	90	78	80	85	84	70	68	80	79	71	69	68	75	76	78	79	74	72	73	74	72	71	69	67	67
13	90	82	82	87	85	76	75	72	72	63	63	63	70	76	81	79	74	72	71	73	72	71	71	66	66
14	95	78	85	83	82	71	68	63	62	64	58	56	62	63	66	64	60	64	62	63	62	61	63	58	58
15	87	77	80	83	79	73	72	73	71	56	53	56	64	64	68	63	60	62	59	61	61	61	59	58	58
16	103	93	94	86	91	75	69	81	78	72	65	70	73	73	73	69	66	69	66	68	66	65	67	66	66
17	93	91	94	84	92	83	69	73	85	83	74	78	73	76	78	75	70	70	71	72	70	67	70	69	69
18	96	94	94	90	95	83	70	72	81	77	75	79	77	86	86	81	78	79	80	81	79	77	78	78	78
19	88	85	89	82	86	76	65	76	82	80	75	80	78	83	82	80	75	75	76	77	77	75	77	77	77
20	93	92	94	86	91	79	67	77	78	74	69	69	69	72	73	71	67	69	69	70	70	68	70	72	72
21	83	89	88	78	83	75	68	72	79	74	70	71	73	75	75	71	67	69	68	69	69	67	69	71	71
22	82	86	84	70	81	71	74	68	71	68	64	66	66	71	70	69	65	66	66	67	67	65	67	67	67
23	68	74	77	60	77	71	69	61	67	64	60	61	62	65	63	63	59	61	62	63	63	61	63	62	61
24	68	74	73	68	79	74	68	56	60	52	55	59	56	59	60	58	56	57	58	60	60	59	62	61	60
25	66	69	68	70	83	70	62	59	65	54	51	57	51	56	57	54	52	53	54	55	54	53	53	52	51

³⁵ Neste experimento, utilizaram-se coeficientes mel-cepstrais e limiares de energia de início e fim iguais a 0,046433.

Apêndice D Resultados dos Experimentos para o Melhor Caso

As taxas de falsa aceitação para todos os locutores foram, de uma forma geral, baixos. Dentre as 10 repetições da sentença do locutor L16 feitas pelo locutor L19, uma foi aceita incorretamente (falsa aceitação, $a=1$), e para cinco foi solicitada nova repetição ($d=5$). Para as repetições da sentença do locutor L4, feitas pelo locutor L23, uma foi aceita incorretamente ($a=1$) e para oito foi solicitada nova repetição ($d=8$).

Com relação apenas à repetição das sentenças, para as dez elocuições da sentença do locutor L3 feitas pelo locutor L2, duas foram solicitadas uma nova repetição ($d=2$), vale destacar que esses locutores são irmãs e gêmeas. Para as elocuições da sentença do locutor L16 feitas pelo locutor L6, duas foram solicitadas uma nova repetição ($d=2$), para o locutor L6 também é solicitado a repetição de quatro sentenças ($d=4$) quando ele tenta se passar pelo locutor L25. Para as dez repetições da sentença do locutor L27 feitas pelo locutor L16 três são solicitadas que sejam repetidas ($d=3$). Nas elocuições da sentença do locutor L25 feitas pelo locutor L19, são solicitadas nove repetições ($d=9$) e nas elocuições da sentença do locutor L3 feitas pelo locutor L23 é solicitada uma nova repetição ($d=1$).

Nas Tabelas D.1, D.2, D.3 e D.4 são apresentados as taxas de falsa aceitação, falsa rejeição e repetição respectivamente, para os locutores utilizando coeficientes MFCC.

Tabela D.1: Taxas de falsa rejeição para cada locutor com coeficientes mel-cepstrais.

Locutor	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15
Falsa rejeição	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	10%	0%	30%

Tabela D.1: Taxas de falsa rejeição para cada locutor com coeficientes mel-cepstrais (cont.).

Locutor	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
Falsa rejeição	0%	0%	0%	0%	0%	0%	0%	10%	0%	20%	0%	0%	0%	40%	20%

Tabela D.2: Taxas de repetição por locutor da sua própria sentença com coeficientes mel-cepstrais.

Locutor	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15
Taxa	0%	40%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	10%

Tabela D.2: Taxas de repetição por locutor da sua própria sentença com coeficientes mel-cepstrais (cont.).

Locutor	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
Taxa	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	20%

Tabela D.3: Taxa de falsa aceitação para cada locutor com mel-cepstrais.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30	
L1	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L2	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L3	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L4	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L5	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L6	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L7	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L8	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L9	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L10	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L11	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L12	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L13	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L14	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L16	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L17	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L18	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L19	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L21	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L22	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%
L23	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%
L24	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%
L25	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%
L26	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%
L27	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%
L28	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%
L29	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%
L30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%

Tabela D.4: Índice de repetição por locutor da sentença de outro locutor com coeficientes mel-cepstrais.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L2	0%	-	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L3	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L4	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L5	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L6	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	40%	0%	0%	0%	0%	0%
L7	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L8	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L9	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L10	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L11	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L12	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L13	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L14	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L16	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	
L17	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L18	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L19	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	0%	0%	-	0%	0%	0%	0%	0%	90%	0%	0%	0%	0%	
L20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	
L21	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	
L22	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	
L23	0%	0%	10%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	
L24	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	
L25	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	
L26	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	
L27	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	
L28	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	
L29	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%
L30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-

D.1 Análise Comparativa entre Coeficientes LPC e Mel-cepstrais

Para alguns casos, os resultados podem não parecer satisfatórios e isto pode ter ocorrido devido a uma série de problemas, tais como: ruído originário do ambiente de gravação (conversas, portas abrindo e fechando, etc), estado emocional e de saúde dos locutores, posição do microfone em relação ao locutor. Além desses, outro fator que afeta o desempenho do sistema são as características inerentes de cada locutor. Por exemplo, a mesma sentença pronunciada por um único locutor apresenta, algumas vezes, características bem diferentes.

No entanto, quando comparados aos resultados apresentados pelos coeficientes LPC, os resultados dos coeficientes mel-cepstrais permitem constatar uma melhoria considerável, como pode ser visto na Tabela D.5.

Tabela D.5: Taxa dos resultados para a melhor configuração de limiares.

Quantidade de Acertos		Quantidade de Erros		Quantidade de Desconhecidos	
MFCC	LPC	Mel	LPC	Mel	LPC
99,34%	96,52%	0,17%	0,90%	0,49%	2,58%

Os resultados da Tabela D.5 mostram um aumento de 440% na quantidade de erros, ou seja, cinco vezes mais erros quando os coeficientes LPC são utilizados em vez dos coeficientes mel-cepstrais. Outro número importante a destacar é o aumento do número de respostas desconhecidas (427%), nesse caso o sistema solicitará mais repetições das sentenças aos locutores, o que pode gerar uma insatisfação por parte dos usuários ao utilizar o sistema com coeficientes LPC.

Para reforçar esses resultados e apresentar outras justificativas para o uso dos coeficientes mel-cepstrais, são apresentadas as Tabelas D.6, D.7, D.8, D.9 e D.10, os índices de falsa aceitação, falsa rejeição e repetição respectivamente, utilizando coeficientes LPC.

O uso de coeficientes LPC acarreta em um aumento significativo da taxa de falsa rejeição. Pode-se destacar o resultado para quatro locutores, em primeiro lugar para os locutores L21 e L29, onde a taxa de falsa rejeição ficou em 100% ($r=10$), ou seja, para esse conjunto de teste, estes dois locutores não foram reconhecidos em nenhuma das dez repetições das suas sentenças. Em seguida, o

resultado dos locutores L10 e L28 cujo índice de falsa rejeição ficou em 80% ($r=8$), o que pode agravar ainda mais esse resultado é que nas duas elocuições que não houve a falsa rejeição, o sistema solicita a repetição das sentenças ($d=2$).

É importante destacar que, a partir do uso dos coeficientes mel-cepstrais a taxa de falsa rejeição entre os locutores não ultrapassou 40%.

Outro resultado agravante é o aumento no índice de falsa aceitação, seu problema consiste na possibilidade de usuários mal intencionados conseguirem se passar por locutores cadastrados no sistema. Os resultados que mais chamam a atenção são os do locutor L2 que em todas as repetições da sentença do locutor L3, L2 consegue ser aceito ($a=10$). Este é o caso no qual os locutores são irmãs gêmeas, o que pode ser um atenuante para o resultado. No entanto, usando-se mel-cepstrais, em nenhum teste L2 conseguiu se passar por L3 ou vice-versa. Nas elocuições da sentença do locutor L4 feitas pelo locutor L2, sete repetições foram aceitas de maneira incorreta ($a=7$). Para as elocuições da sentença do locutor L4 feitas pelo locutor L23, quatro repetições foram aceitas incorretamente ($a=4$). Estes são os resultados para a falsa aceitação mais importantes de serem destacados. Em outras ocorrências, a taxa não passou de 30%.

Ao serem utilizados os coeficientes mel-cepstrais, a taxa de falsa aceitação não foi superior a 10% e este fato só foi observado na elocução de uma sentença de dois locutores.

Com relação à repetição das sentenças, para o primeiro caso, quando o locutor é solicitado a repetir a própria sentença, conforme apresentado na Tabela D.7, os resultados mais significativos são: locutor L25, no qual em dez elocuições, oito foram solicitadas novas repetições ($d=8$); para o locutor L16 das dez elocuições em sete o sistema solicita uma nova repetição ($d=7$); e os locutores L17 e L24. Para estes, das dez elocuições de suas sentenças em seis foram solicitadas novas repetições ($d=6$). Para os outros casos, onde houve a solicitação de repetição das sentenças o índice não foi superior a 20%. Utilizando-se coeficientes mel-cepstrais esse índice não ultrapassou os 40%.

No segundo caso, o locutor é solicitado a repetir a sentença quando tenta se passar por outro locutor. Para o caso no qual o locutor L20 tenta se passar pelo locutor L14, das dez repetições feitas pelo locutor L20 em nove delas o sistema solicita uma nova repetição ($d=9$). Quando o locutor L23 tenta se passar por L4 novamente em nove delas é solicitado a repetição ($d=9$) e as tentativas do locutor

L23 se passar por L8 em sete repetições da sentença, é solicitado uma nova repetição por parte de L23 (d=7). Quando o locutor L18 tenta se passar por L4 ou L22, em oito elocuições foram solicitadas uma nova repetição (d=8) e as elocuições da sentença do locutor L9 feitas pelo locutor L18 em sete foram solicitadas a repetição (d=7). Para os casos nos quais, L3 tenta se passar por L4 e L11 tenta se passar por L10 em ambas foram solicitadas sete repetições das dez elocuições (d=7). Mais informações sobre o índice de repetição das sentenças por locutor são apresentadas na Tabela D.9.

Tabela D.6: Taxas de falsa rejeição para cada locutor com LPC.

Locutor	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15
Falsa rejeição	0%	0%	0%	0%	0%	0%	0%	0%	0%	80%	0%	40%	0%	0%	0%

Tabela D.6: Taxas de falsa rejeição para cada locutor com LPC (cont.).

Locutor	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
Falsa rejeição	30%	0%	0%	0%	0%	100%	0%	0%	40%	10%	0%	0%	80%	100%	0%

Tabela D.7: Taxas de repetição por locutor da sua própria sentença com LPC.

Locutor	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15
Taxa	0%	10%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	10%

Tabela D.7: Taxas de repetição por locutor da sua própria sentença com LPC (cont.).

Locutor	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
Taxa	70%	60%	0%	0%	0%	0%	0%	0%	60%	80%	0%	0%	20%	0%	0%

Tabela D.8: Taxas de falsa aceitação para cada locutor com LPC.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L2	0%	-	100%	70%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L3	0%	10%	-	30%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L4	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L5	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L6	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L7	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%
L8	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L9	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L10	0%	0%	0%	0%	0%	0%	0%	0%	10%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L11	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L12	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L13	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L14	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L16	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L17	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L18	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L19	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L21	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%
L22	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%
L23	0%	0%	0%	40%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%
L24	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%
L25	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%
L26	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%
L27	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%
L28	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%
L29	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%
L30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-

Tabela D.9: Taxas de repetição por locutor da sentença de outro locutor com LPC.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L2	0%	-	0%	20%	10%	0%	60%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	10%	0%	0%	0%	0%
L3	0%	20%	-	70%	0%	0%	60%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%
L4	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L5	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L6	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L7	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%
L8	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L9	0%	0%	20%	0%	0%	0%	10%	0%	-	40%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	0%	0%	40%	0%	10%	0%	0%
L10	0%	20%	0%	0%	0%	0%	0%	0%	40%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L11	0%	0%	0%	0%	0%	0%	0%	0%	40%	70%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	10%	0%	10%	0%	0%
L12	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L13	0%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	-	50%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L14	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L15	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L16	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	60%	0%	0%	0%	0%	0%	0%	0%	40%	0%	0%	0%
L17	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L18	10%	0%	30%	80%	40%	0%	10%	0%	70%	10%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	80%	60%	0%	0%	0%	50%	0%	0%	0%	0%
L19	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	90%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
L21	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%	0%
L22	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%	0%	0%
L23	20%	0%	30%	50%	90%	0%	0%	70%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	20%	0%	0%
L24	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%	0%
L25	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%	0%
L26	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%	0%
L27	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%
L28	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%	0%
L29	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-	0%
L30	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-

Apêndice E Matrizes de Confusão

A seguir, são apresentadas as matrizes de confusão obtidas com a simulação dos experimentos.

Tabela E.1: Matriz de confusão para experimento E2 com a menor quantidade de erros.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L2	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L3	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L11	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L12	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L13	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	1	0	0
L17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
L18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
L19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	10	0	0	0	0	0	0	0	0	0	0	0
L20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0
L22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
L23	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0
L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0
L25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
L26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
L27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
L28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0
L29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
L30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5

Tabela E.2: Matriz de confusão para experimento E3 com a menor quantidade de erros.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30	
L1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
L2	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L3	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L10	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L11	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L12	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L13	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
L17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
L19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
L20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0
L22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
L23	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0
L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0
L25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
L26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
L27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
L28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0
L29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
L30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5

Tabela E.3: Matriz de confusão para experimento E4 com a menor quantidade de erros.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L2	0	8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L3	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L10	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L11	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L12	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L13	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	1	0	0	
L17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	
L18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	
L19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	10	0	0	0	0	0	0	0	0	0	0	
L20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	
L21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	
L22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	
L23	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	
L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	
L25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	
L26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	
L27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	
L28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	
L29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
L30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	

Tabela E.4: Matriz de confusão para experimento E5 com a menor quantidade de erros.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L2	0	10	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L3	0	0	10	10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	
L8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L9	0	0	1	0	0	0	0	0	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L10	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	1	0	
L11	0	0	0	0	0	0	0	0	0	3	10	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	
L12	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L13	0	0	0	0	0	0	0	0	0	0	0	0	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	3	0	0	
L17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	
L18	0	0	1	3	0	0	0	0	3	0	0	0	0	0	0	0	0	10	0	0	2	0	0	0	0	0	0	0	0	
L19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	10	0	0	0	0	0	0	0	0	0	0	
L20	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	
L21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	
L23	0	0	0	7	4	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	1	0	
L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	
L25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	
L26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	
L27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	
L28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	

Tabela E.5: Matriz de confusão para experimento E6 com a menor quantidade de erros.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L2	0	10	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
L3	0	0	10	7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0
L8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L9	0	0	1	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L10	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0
L11	0	0	0	0	0	0	0	0	0	2	10	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
L12	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L13	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
L15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	2	0	0	0	0	0	0	0	3	0	0	0
L17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
L18	0	0	0	4	0	0	0	0	6	0	0	0	0	0	0	0	0	10	0	0	0	1	0	0	0	0	0	0	0	0
L19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0
L20	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
L23	0	0	0	6	4	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	1	0	0
L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
L25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0
L26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
L27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0
L28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10

Tabela E.6: Matriz de confusão para experimento E8 com a menor quantidade de erros.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30
L1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
L2	0	10	10	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
L3	0	4	10	5	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L4	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L5	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L6	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L7	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0
L8	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L10	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0
L11	0	0	0	0	0	0	0	0	1	1	10	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
L12	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L13	0	0	0	0	1	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L14	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
L15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6	0	0	0	0	0	0	0	3	0	0	0
L17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0
L18	0	0	2	2	1	0	1	0	1	0	0	0	0	0	0	0	0	10	0	0	0	1	0	0	0	0	0	0	0	0
L19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0
L20	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
L23	0	0	0	7	4	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0
L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
L25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0
L26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
L27	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0
L28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10

Apêndice F Verificação Funcional

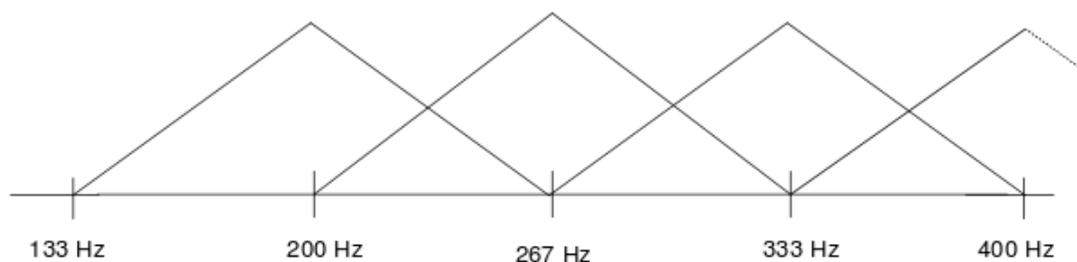
Preliminar

Nesse apêndice são apresentados os casos escolhidos para realizar a verificação funcional preliminar dos blocos FILTER e MFCC.

F.1 Verificação Funcional Preliminar do Bloco FILTER

Na figura F.1, é possível visualizar os primeiros cinco filtros do banco de filtros, junto com as suas frequências centrais e bordas.

Figura F.1: Os primeiros cinco filtros do banco.



Calculou-se as frequências da FFT para os onze primeiros valores. E os resultados podem ser visualizados na Tabela F.1.

Tabela F.1: Os onze primeiros valores da frequência da FFT.

i	FFT_{freq}
0	0
1	43Hz
2	86Hz
3	129Hz
4	172Hz
5	215Hz
6	258Hz
7	301Hz
8	344Hz
9	387Hz
10	430Hz

Esses resultados foram obtidos resolvendo a Equação F.1.

$$FFT_{freq} = \frac{i}{npf} \times f \quad \text{para } 0 \leq i \leq npf, \quad (F.1)$$

na qual:

i – índice da frequência a ser calculada;

npf – número de pontos da FFT;

f – frequência de amostragem.

Para esta pesquisa, $npf = 256$ e $f = 11025$ amostras/segundo. De maneira aleatória foi escolhida a frequência de índice 7 que corresponde ao valor de 301 Hz. Em seguida, foi fornecido como entrada para o módulo FILTER um vetor com 256 elementos no qual o elemento na posição 7, tinha como valor 15.000 e os demais iguais a 0. Então, baseado na Figura F.1 é esperado que apenas o segundo filtro e o terceiro filtro atuem sobre o valor de entrada. Porque o segundo filtro tem como configuração: borda inferior igual a 200 Hz, frequência central igual a 267 Hz e borda superior igual a 333 Hz, o terceiro filtro tem como configuração borda inferior igual a 267 Hz, frequência central igual a 333 Hz e borda superior igual a 400 Hz.

A saída fornecida pelo módulo foi um vetor³⁶ com 0 na primeira posição, 1 na segunda posição, 12 na terceira posição e o restante dos elementos do vetor iguais a 0. Comportando-se assim, conforme o esperado para a entrada fornecida.

F.2 Verificação Funcional Preliminar do Bloco MFCC

O módulo MFCC é responsável pela aplicação da DCT sobre as saídas fornecidas pelo módulo FILTER. Para realizar a verificação funcional preliminar sobre o módulo MFCC, foram analisados dois cenários, baseados no comportamento esperado pela aplicação DCT.

³⁶ Tamanho do vetor de saída: 36 elementos.

No primeiro cenário, foi fornecido um vetor com 36 elementos cujos valores eram iguais a 15.000. E a saída fornecida pelo módulo foi um vetor de doze³⁷ elementos iguais a -1, próximo de zero. Comportamento esperado para este tipo de configuração de entrada.

No segundo cenário, o vetor fornecido como entrada do módulo MFCC tinha a seguinte configuração: o primeiro elemento com valor igual a 15.000 e os demais iguais a 0. A saída fornecida pelo módulo foi um vetor com os seguintes valores: -451, -564, -574, -591, -615, -645, -681, -725, -774, -830, -892, -960 Comportamento esperado para este tipo de entrada.

³⁷ Número de coeficientes MFCC utilizados nessa pesquisa.

Anexo A Lista das Frases Foneticamente Balanceadas

A seguir, são listadas as sentenças escolhidas nesta pesquisa.

Tabela AX.1: Lista das sentenças escolhidas.

Lista das Sentenças
Esse tema foi falado no congresso.
Leila tem um lindo casaco.
O analfabetismo é a vergonha do país.
Agindo com união ainda rende mais.
Recebi meu pai pra almoçar.
Os pesquisadores não acreditam nessa história.
Sei que amanhã atingiremos a meta proposta.
Refleta antes e discuta depois.
Nosso telefone está mudo.
Desculpe se te chamo de velho.
Queremos discutir o orçamento.
Ela não tem fome quando sai de casa.
Uma índia andava na floresta.
Zeca, corra bem rápido pra casa.
Neste caso, dormirei tranquilo.
Entre com seu código e o número da conta.
Ainda faltam seis minutos.
A paixão dele é a natureza.
João caminhou na praia calma.
Você quer me dizer a data?
Ontem não pude fazer minha ginástica.
Ele dorme num leito de palha.
Foi muito difícil entender a canção de natal.
Finalmente o mau tempo deixou o continente.
Quero usar a máquina.
Procurei Maria em casa.
Quero te ver bem quando ele voltar de lá.
Tenho muito orgulho de nossa gente.
Ainda não se sabe o dia da prova.
Será muito difícil conseguir que eu coma.

Anexo B Relatório Gerado pela Ferramenta *dc_shell*

A seguir, são apresentados os relatórios gerados pela ferramenta *dc_shell*.

Tabela BX.1: Relatório de consumo e área do Módulo FFT.

Módulo FFT		

Report : power		
-analysis_effort low		
Design : HDL_FFT		
Version: H-2013.03-SP3		
Date : Thu Nov 6 17:49:52 2014		

Library(s) Used:		
fsc0g_d_generic_core_tt1p2v25c (File: /opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t t1p2v25c.db)		
Operating Conditions: TCCOM Library: fsc0g_d_generic_core_tt1p2v25c		
Wire Load Model Mode: enclosed		
Design	Wire Load Model	Library

HDL_FFT	enG30K	fsc0g_d_generic_core_tt1p2v25c
TwiddleTable	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly	enG10K	fsc0g_d_generic_core_tt1p2v25c
DualPortRam_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
FFTLogic	enG5K	fsc0g_d_generic_core_tt1p2v25c
DualPortRam_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
DualPortRam_2	enG5K	fsc0g_d_generic_core_tt1p2v25c

DualPortRam_3	enG5K	fsc0g_d_generic_core_tt1p2v25c
FFTLogic_DW01_sub_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
FFTLogic_DW01_add_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
FFTLogic_DW01_sub_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
FFTLogic_DW01_add_2	enG5K	fsc0g_d_generic_core_tt1p2v25c
FFTLogic_DW01_inc_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
FFTLogic_DW01_inc_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW01_sub_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW01_sub_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW01_add_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW01_add_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW01_add_2	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW01_sub_2	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW_mult_tc_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW_mult_tc_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW_mult_tc_2	enG5K	fsc0g_d_generic_core_tt1p2v25c
Butterfly_DW_mult_tc_3	enG5K	fsc0g_d_generic_core_tt1p2v25c

Global Operating Voltage = 1.2

Power-specific unit information :

Voltage Units = 1V

Capacitance Units = 1.000000pf

Time Units = 1ns

Dynamic Power Units = 1mW (derived from V,C,T units)

Leakage Power Units = 1pW

Cell Internal Power = 1.9084 uW (68%)

Net Switching Power = 900.7193 nW (32%)

Total Dynamic Power = 2.8091 uW (100%)

Cell Leakage Power = 806.5709 nW

Power Group	Internal Power	Switching Power	Leakage Power	Total Power (%) Attrs
io_pad	0.0000	0.0000	0.0000	0.0000 (0.00%)
memory	0.0000	0.0000	0.0000	0.0000 (0.00%)
black_box	0.0000	0.0000	0.0000	0.0000 (0.00%)
clock_network	3.7817e-04	6.0821e-04	3.5479e+05	1.3412e-03 (37.09%)
register	1.4304e-03	4.7825e-05	1.8666e+05	1.6649e-03 (46.05%)
sequential	0.0000	0.0000	0.0000	0.0000 (0.00%)
combinational	9.9872e-05	2.4468e-04	2.6512e+05	6.0968e-04 (16.86%)
Total	1.9084e-03 mW	9.0072e-04 mW	8.0657e+05 pW	3.6157e-03 mW

Report : area

Design : HDL_FFT

Version: H-2013.03-SP3

Date : Thu Nov 6 17:51:10 2014

Library(s) Used:

fsc0g_d_generic_core_tt1p2v25c (File:
/opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t
t1p2v25c.db)

Number of ports:	71
Number of nets:	604
Number of cells:	8
Number of combinational cells:	1
Number of sequential cells:	0
Number of macros/black boxes:	0
Number of buf/inv:	0
Number of references:	8
Combinational area:	30542.079880
Buf/Inv area:	1813.759960
Noncombinational area:	12408.320044
Macro/Black Box area:	0.000000
Total cell area:	42950.399923

Tabela BX.2: Relatório de consumo e área do Módulo FILTER.

Módulo FILTER
<pre> ***** Report : power -analysis_effort low Design : filter Version: H-2013.03-SP3 Date : Thu Nov 6 17:55:50 2014 ***** Library(s) Used: fsc0g_d_generic_core_tt1p2v25c (File: /opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t t1p2v25c.db) </pre>

Operating Conditions: TCCOM Library: fsc0g_d_generic_core_tt1p2v25c

Wire Load Model Mode: enclosed

Design	Wire Load Model	Library
filter	enG100K	fsc0g_d_generic_core_tt1p2v25c
rom_filter	enG5K	fsc0g_d_generic_core_tt1p2v25c
log	enG10K	fsc0g_d_generic_core_tt1p2v25c
filter_DW01_inc_2	enG5K	fsc0g_d_generic_core_tt1p2v25c
filter_DW01_inc_3	enG5K	fsc0g_d_generic_core_tt1p2v25c
filter_DW01_inc_4	enG5K	fsc0g_d_generic_core_tt1p2v25c
log_DW_mult_tc_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
log_DW_mult_tc_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
log_DW_mult_tc_2	enG5K	fsc0g_d_generic_core_tt1p2v25c
log_DW_mult_tc_3	enG5K	fsc0g_d_generic_core_tt1p2v25c
log_DW01_inc_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
log_DW_mult_tc_4	enG5K	fsc0g_d_generic_core_tt1p2v25c
filter_DW01_add_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
filter_DW_mult_tc_0	enG5K	fsc0g_d_generic_core_tt1p2v25c

Global Operating Voltage = 1.2

Power-specific unit information :

Voltage Units = 1V

Capacitance Units = 1.000000pf

Time Units = 1ns

Dynamic Power Units = 1mW (derived from V,C,T units)

Leakage Power Units = 1pW

Cell Internal Power = 55.5924 uW (93%)

Net Switching Power = 4.3149 uW (7%)

Total Dynamic Power = 59.9073 uW (100%)

Cell Leakage Power = 5.6817 uW

Power Group	Internal Power	Switching Power	Leakage Power	Total Power (%) Attrs
io_pad	0.0000	0.0000	0.0000	0.0000 (0.00%)
memory	0.0000	0.0000	0.0000	0.0000 (0.00%)
black_box	0.0000	0.0000	0.0000	0.0000 (0.00%)
clock_network	0.0000	0.0000	0.0000	0.0000 (0.00%)
register	5.5095e-02	3.3151e-05	3.2760e+06	5.8404e-02 (89.04%)
sequential	0.0000	0.0000	0.0000	0.0000 (0.00%)
combinational	4.9821e-04	4.2818e-03	2.4057e+06	7.1856e-03 (10.96%)

Total	5.5593e-02 mW	4.3149e-03 mW	5.6817e+06 pW	6.5590e-02 mW
-------	---------------	---------------	---------------	---------------

Report : area

Design : filter

Version: H-2013.03-SP3

Date : Thu Nov 6 17:55:59 2014

Library(s) Used:

fsc0g_d_generic_core_tt1p2v25c (File:
/opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t
t1p2v25c.db)

Number of ports: 38

Number of nets: 11494

Number of cells:	10753
Number of combinational cells:	4907
Number of sequential cells:	5839
Number of macros/black boxes:	0
Number of buf/inv:	1822
Number of references:	51
Combinational area:	123555.839254
Buf/Inv area:	11182.079750
Noncombinational area:	194990.072985
Macro/Black Box area:	0.000000
Total cell area:	318545.912239

Tabela BX.3: Relatório de Consumo e Área do Módulo MFCC.

Módulo MFCC	

Report : power	
-analysis_effort low	
Design : mfcc_rtl	
Version: H-2013.03-SP3	
Date : Wed Mar 4 14:48:25 2015	

Library(s) Used:	
fsc0g_d_generic_core_tt1p2v25c (File:	
/opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t	
t1p2v25c.db)	
Operating Conditions: TCCOM Library: fsc0g_d_generic_core_tt1p2v25c	
Wire Load Model Mode: enclosed	

Design	Wire Load Model	Library		
mfcc_rtl	enG30K	fsc0g_d_generic_core_tt1p2v25c		
rom_mfcc	enG5K	fsc0g_d_generic_core_tt1p2v25c		
mfcc_rtl_DW01_inc_1	enG5K	fsc0g_d_generic_core_tt1p2v25c		
mfcc_rtl_DW01_add_0	enG5K	fsc0g_d_generic_core_tt1p2v25c		
mfcc_rtl_DW_mult_tc_0	enG5K	fsc0g_d_generic_core_tt1p2v25c		
Global Operating Voltage = 1.2				
Power-specific unit information :				
Voltage Units = 1V				
Capacitance Units = 1.000000pf				
Time Units = 1ns				
Dynamic Power Units = 1mW (derived from V,C,T units)				
Leakage Power Units = 1pW				
Cell Internal Power = 10.3551 uW (91%)				
Net Switching Power = 1.0650 uW (9%)				

Total Dynamic Power = 11.4200 uW (100%)				
Cell Leakage Power = 1.2616 uW				
Power Group	Internal Power	Switching Power	Leakage Power	Total Power (%) Attrs
io_pad	0.0000	0.0000	0.0000	0.0000 (0.00%)
memory	0.0000	0.0000	0.0000	0.0000 (0.00%)
black_box	0.0000	0.0000	0.0000	0.0000 (0.00%)

clock_network	0.0000	0.0000	0.0000	0.0000 (0.00%)
register	1.0135e-02	2.8341e-05	5.7974e+05	1.0744e-02 (84.72 %)
sequential	0.0000	0.0000	0.0000	0.0000 (0.00%)
combinational	2.1963e-04	1.0366e-03	6.8182e+05	1.9381e-03 (15.28%)

Total	1.0355e-02 mW	1.0650e-03 mW	1.2616e+06 pW	1.2682e-02 mW

Report : area				
Design : mfcc_rtl				
Version: H-2013.03-SP3				
Date : Wed Mar 4 14:48:57 2015				

Library(s) Used:				
fsc0g_d_generic_core_tt1p2v25c (File: /opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t t1p2v25c.db)				
Number of ports:	34			
Number of nets:	2253			
Number of cells:	2093			
Number of combinational cells:	1045			
Number of sequential cells:	1044			
Number of macros/black boxes:	0			
Number of buf/inv:	199			
Number of references:	45			
Combinational area:	31394.560070			
Buf/Inv area:	2009.599955			
Noncombinational area:	34274.558825			
Macro/Black Box area:	0.000000			

Total cell area:	65669.118895
------------------	--------------

Tabela BX.4: Relatório de Consumo e Área do Módulo LPCA.

Módulo LPCA		

Report : power		
-analysis_effort low		
Design : lpca		
Version: H-2013.03-SP3		
Date : Thu Nov 6 18:04:52 2014		

Library(s) Used:		
fsc0g_d_generic_core_tt1p2v25c (File: /opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t t1p2v25c.db)		
Operating Conditions: TCCOM Library: fsc0g_d_generic_core_tt1p2v25c		
Wire Load Model Mode: enclosed		
Design	Wire Load Model	Library

lpca	enG100K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW01_sub_0	enG5K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW01_sub_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW01_inc_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW_mult_tc_0	enG10K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW01_sub_2	enG5K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW_mult_tc_1	enG5K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW01_add_48	enG5K	fsc0g_d_generic_core_tt1p2v25c
lpca_DW_mult_tc_2	enG5K	fsc0g_d_generic_core_tt1p2v25c

lpca_DW01_add_49	enG5K	fsc0g_d_generic_core_tt1p2v25c		
lpca_DW_mult_tc_3	enG10K	fsc0g_d_generic_core_tt1p2v25c		
lpca_DW01_add_50	enG5K	fsc0g_d_generic_core_tt1p2v25c		
lpca_DW_mult_tc_4	enG5K	fsc0g_d_generic_core_tt1p2v25c		
lpca_DW_div_tc_5	enG30K	fsc0g_d_generic_core_tt1p2v25c		
Global Operating Voltage = 1.2				
Power-specific unit information :				
Voltage Units = 1V				
Capacitance Units = 1.000000pf				
Time Units = 1ns				
Dynamic Power Units = 1mW (derived from V,C,T units)				
Leakage Power Units = 1pW				
Cell Internal Power = 15.4750 uW (87%)				
Net Switching Power = 2.2235 uW (13%)				

Total Dynamic Power = 17.6985 uW (100%)				
Cell Leakage Power = 4.1793 uW				
	Internal	Switching	Leakage	Total
Power Group	Power	Power	Power	Power (%) Attrs

io_pad	0.0000	0.0000	0.0000	0.0000 (0.00%)
memory	0.0000	0.0000	0.0000	0.0000 (0.00%)
black_box	0.0000	0.0000	0.0000	0.0000 (0.00%)
clock_network	0.0000	0.0000	0.0000	0.0000 (0.00%)
register	1.4771e-02	7.0999e-05	8.4359e+05	1.5685e-02 (71.69%)
sequential	0.0000	0.0000	0.0000	0.0000 (0.00%)
combinational	7.0429e-04	2.1525e-03	3.3358e+06	6.1926e-03 (28.31%)

Total 1.5475e-02 mW 2.2235e-03 mW 4.1794e+06 pW 2.1878e-02 mW

Report : area

Design : lpca

Version: H-2013.03-SP3

Date : Thu Nov 6 18:05:16 2014

Library(s) Used:

fsc0g_d_generic_core_tt1p2v25c (File:
/opt/rede/umc/fsc0g_d/2009Q2v3.0/GENERIC_CORE/FrontEnd/synopsys/fsc0g_d_generic_core_t
t1p2v25c.db)

Number of ports:	33
Number of nets:	5590
Number of cells:	4965
Number of combinational cells:	3422
Number of sequential cells:	1530
Number of macros/black boxes:	0
Number of buf/inv:	997
Number of references:	75
Combinational area:	177623.039061
Buf/Inv area:	11870.719735
Noncombinational area:	47041.278944
Macro/Black Box area:	0.000000
Total cell area:	224664.318005