

**PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E PROCESSOS  
INDUSTRIAIS – MESTRADO  
ÁREA DE CONCENTRAÇÃO EM  
CONTROLE E OTIMIZAÇÃO DE PROCESSOS INDUSTRIAIS**

Juliano Gauciniski

**ESTUDO DAS TRANSFORMADAS WAVELETS PARA UTILIZAÇÃO  
EM RECONHECIMENTO E CLASSIFICAÇÃO DE COMANDOS DE  
VOZ**

Santa Cruz do Sul, outubro de 2009

Juliano Gauciniski

**ESTUDO DAS TRANSFORMADAS WAVELETS PARA UTILIZAÇÃO  
EM RECONHECIMENTO E CLASSIFICAÇÃO DE COMANDOS DE  
VOZ**

Dissertação apresentada ao Programa de Pós-Graduação em Sistemas e Processos Industriais – Mestrado, Área de Concentração em Controle e Otimização de Processos Industriais, Universidade de Santa Cruz do Sul – UNISC, como requisito parcial para obtenção do título de Mestre em Sistemas e Processos Industriais.

Orientador: Prof. Dr. Rolf Fredi Molz

Co-orientador: Prof. Dr. Rubén Panta Pazos

Santa Cruz do Sul, outubro de 2009

Juliano Gauciniski

**ESTUDO DAS TRANSFORMADAS WAVELETS PARA UTILIZAÇÃO EM  
RECONHECIMENTO E CLASSIFICAÇÃO DE COMANDOS DE VOZ**

Essa Dissertação foi submetida ao Programa de Pós-Graduação em Sistemas e Processos Industriais – Mestrado, Área de Concentração em Controle e Otimização de Processos Industriais, Universidade de Santa Cruz do Sul – UNISC, como requisito parcial para obtenção do título de Mestre em Sistemas e Processos Industriais.

*Dr. Rolf Fredi Molz*  
Professor Orientador

*Prof. Dr. Rubén Panta Pazos*

*Prof. Dr. Marco Flôres Ferrão*

*Prof. Dr. Eugênio de Oliveira Simonetto*

## **AGRADECIMENTOS**

Agradeço a todos que contribuíram para o desenvolvimento deste trabalho, aos grandes mestres Rolf Fredi Molz e Rúben Panta Pazos, professores doutores do mestrado, pela incansável ajuda, à Patricia Almeida, amiga que me deu forças para continuar e da qual a ajuda foi essencial para que este trabalho se realizasse, ao meu pai, um verdadeiro exemplo de vida e, em especial, à minha mãe, que mesmo à distância, jamais me deixou desistir.

A todos, o meu muito obrigado.

*Mesmo desacreditado e ignorado por todos, não posso desistir, pois para mim, vencer é nunca desistir.*

(Albert Einstein)

## RESUMO

Tecnologias de reconhecimento de voz permitem que computadores equipados com microfones interpretem a fala humana para transcrição ou como método de comando por voz. Por meio dessa tecnologia, pode-se realizar tarefas sem a utilização de determinados dispositivos de entrada, como o mouse e o teclado, deixando livres as mãos do operador para realizar diferentes tarefas. Neste contexto, foi elaborado um projeto que tem como objetivo, através de técnicas de reconhecimento de padrões, desenvolver um estudo das transformadas *wavelets* com a finalidade de definir um melhor método para realizar a classificação do fumo em folha através de comandos de voz, sem a necessidade de nenhum outro dispositivo de entrada de dados. A metodologia empregada foi a aquisição da voz em formato digital e a aplicação das transformadas *wavelets* Haar e Daubechies para análise. Para o processo de classificação e reconhecimento dos padrões, empregaram-se as técnicas de Análises de Componentes Principais e *Power Cepstrum*. Diversos empecilhos ocorreram durante a fase de classificação, principalmente pelo fato do sinal compactado ser de grande porte. Dessa forma, fez-se necessária a utilização do método matemático *Power Cepstrum*, permitindo a redução do número de elementos no sinal e possibilitando obter êxito na classificação dos elementos, neste caso, orientados pela freqüência de cada fonema analisado. Foi possível concluir que as transformadas *Wavelets* utilizadas possibilitaram o reconhecimento de padrões, porém, possuem limitações quando se trata de sinais de grande porte, necessitando, desta forma, de um pré-processamento.

**Palavras-Chaves:** Transformadas *Wavelets*, Análise de Componentes Principais, *Power Cepstrum*, Reconhecimento de Voz, Comandos de Voz.

## ABSTRACT

Voice-recognition technologies allow computers equipped with microphones to interpret human speech for transcription or as a method of voice command. Through this technology, it is possible to perform tasks without using certain input devices such as mouse and keyboard, keeping the operator's hands free to perform different tasks. In this context, a project has been drawn up which aims, through the pattern techniques recognition, to develop a Wavelets Transform study in order to discover the best method to perform tobacco leaf classification through voice commands without the need of any other device for data entry. The methodology used was the voice acquisition in digital form and Haar and Daubechies Wavelet Transform application for analysis. For the classification process and patterns recognition, the Principal Component Analysis and Power Cepstrum techniques have been used. Several setbacks have occurred during the qualifying round, mainly because the compressed signal was large. Thus, the use of the Power Cepstrum mathematical method was necessary, allowing signal element number reduction and making it possible to have success in the elements classification, in this case, guided by the frequency of each phoneme analysis. It was concluded that the used Wavelet Transformed enabled patterns recognition, however, there are limitations when it comes to large signs, thus requiring a pre-processing.

**Keywords:** Wavelet Transform, Principal Component Analysis, Power Cepstrum, Voice Recognition, Voice Commands.

## LISTA DE ILUSTRAÇÕES

Figura 1 – (a) Gráfico de um sinal – amplitude x tempo, (b) Gráfico de um sinal – amplitude x frequência (Transformada de Fourier).....	19
Figura 2 - (a) Gráfico de um sinal com janela temporal - amplitude x tempo, (b) Gráfico de um sinal – frequência x tempo (STFT). ....	20
Figura 3 - (a) Gráfico de um sinal - amplitude x tempo, (b) Gráfico de um sinal – escala x tempo (wavelets). ....	21
Figura 4 - Comparação entre uma senoíde (a) e uma <i>wavelet</i> db10 (b).....	22
Figura 5 - Resolução no plano <i>t-f</i> pela análise STFT .....	24
Figura 6 - Análise Multi-Resolucional (Transformada de <i>wavelet</i> ).....	25
Figura 7 – Lógica do cálculo do primeiro grupo de subsinais de Haar. ....	30
Figura 8 – Lógica do cálculo da primeira flutuação <b>d1</b> .....	31
Figura 9 – Sinal original .....	32
Figura 10 – Sinal após a aplicação da transformada Haar .....	32
Figura 11 – (a) Sinal Original, (b) Transformada Daub4 2-Nível.....	34
Figura 12 – Painel de compra de fumo Fonte: Fotografia tirada na empresa .....	44
Figura 13 – Arquivo de áudio compra.wav aberto no Cool Edit Pro 2.0.....	46
Figura 14 – Ampliação da visualização do sinal de áudio – arquivo compra.wav.....	47
Figura 15 – Seleção da palavra MLJ em apenas um canal (mono).....	48
Figura 16 – Palavra MLJ copiada para uma nova janela.....	48
Figura 17 – Seleção apenas da palavra desejada em uma ampliação maior. ....	49
Figura 18 – Cópia da palavra desejada para salvar arquivo MLJ.wav.....	49
Figura 19 – Gráfico do sinal original da palavra BF1.....	54
Figura 20 – Gráfico do sinal da primeira compactação da palavra BF1 (Haar).....	55
Figura 21 – Gráfico do sinal da segunda compactação da palavra BF1 (Haar).....	56
Figura 22 – Gráfico do sinal da terceira compactação da palavra BF1 (Haar).....	57
Figura 23 – Gráfico do sinal original da palavra BF1.....	59
Figura 24 – Gráfico do sinal da primeira compactação da palavra BF1 (Daub4).....	60
Figura 25 – Gráfico do sinal da segunda compactação da palavra BF1 (Daub4).....	61
Figura 26 – Gráfico do sinal da terceira compactação da palavra BF1 (Daub4).....	62
Figura 27 – Grafico do sinal BF2 de tamanho 38747 pontos.....	63
Figura 28 – Grafico do sinal BF2 de tamanho 92937 pontos (zeros adicionados ao final)....	64
Figura 29 – Grafico da palavra BF2 (terceira transformada Haar).....	65
Figura 30 – Grafico da palavra BF2-2 (terceira transformada Haar).....	65
Figura 31 – Subtração das palavras BF2 e BF2-2 .....	66
Figura 32 - Gráfico da palavra BF2 (com valor máximo no inicio) .....	67



Figura 33 - Gráfico da palavra BF2-2 (com valor máximo no início) .....	68
Figura 34 - Gráfico da subtração das palavras BF2 e BF2-2 (com valores máximos no início) .....	69
Figura 35 – Aplicação do Cepstrum de Potência para as palavras BF1 e BF2.....	70
Figura 36 – Diagrama dos autovalores da matriz de covariância: $y=\log(x+1)$ .....	71
Figura 37 – ACP dos 41 sinais power cepstrum com 4 componentes principais .....	72

## **LISTA DE TABELAS**

Tabela 1 – Pronúncia dos fonemas utilizados para formar a classe interna do fumo. ....	45
Tabela 2 – Classes de fumo selecionadas para testes e suas respectivas pronúncias. ....	52

# SUMÁRIO

INTRODUÇÃO.....	13
1    FUNDAMENTAÇÃO TEÓRICA .....	15
1.1    Wavelets .....	16
1.1.1.    Histórico da Teoria Wavelet.....	17
1.1.2.    Teoria Wavelet .....	18
1.1.2.1.    A Transformada de Wavelet Contínua CWT .....	24
1.1.3.    Wavelets Haar.....	28
1.1.3.1.    Transformada Haar.....	29
1.1.3.2.    A Transformada Haar, 1-Nível .....	31
1.1.4.    Wavelets Daubechies.....	33
1.1.4.1.    As Wavelets Daub4 .....	33
1.2    Análise de Componentes Principais .....	35
1.3    Cepstrum.....	37
1.3.1.    Power Cepstrum.....	38
1.3.2.    Cepstrum Complexo.....	39
1.3.3.    Liftering .....	40
2    METODOLOGIA.....	42
2.1    Aquisição da Voz.....	42
2.2    Processo de Classificação de Fumo.....	42
2.3    Aplicação das Wavelets .....	45
2.3.1.    Aplicação da Transformada Haar .....	53
2.3.2.    Aplicação da Transformada Daub4 .....	57
2.3.3.    Análise, Classificação e Reconhecimento dos Comandos de Voz.....	62
CONCLUSÕES.....	73
REFERÊNCIAS .....	75
ANEXO.....	76

## INTRODUÇÃO

Tecnologias de reconhecimento de voz permitem que computadores equipados com microfones interpretem a fala humana, para transcrição, ou como método de comando por voz. Por meio desta tecnologia, pode-se realizar funções sem a utilização de determinados dispositivos de entrada, como o mouse e o teclado, deixando livres as mãos do operador para realizar diferentes tarefas.

Em uma indústria, em determinados processos, necessita-se ao menos duas pessoas: uma para executar o processo e outra para registrar o resultado em algum sistema informatizado.

No caso de algumas indústrias fumageiras, mais especificamente no processo de compra de fumo, são necessárias duas pessoas para realizar a classificação da folha de fumo: um classificador de fumo, que o faz de acordo com os padrões de qualidade da folha, e para isso necessita das mãos livres para manusear o fumo, e um operador de microcomputador, que registra a classe do fumo no sistema através de um *scanner* de mão e uma tabela de códigos.

Utilizando a tecnologia de reconhecimento de voz, o processo de classificação de fumo pode ser automatizado e realizado apenas por uma pessoa, o classificador de fumo, que, através da fala diretamente no microfone, classifica o fumo ao mesmo tempo em que pode manuseá-lo.

Algoritmos e funções matemáticas, tais como a Transformada de Fourier e a Transformada *Wavelet*, são capazes de analisar e extrair características dos sons, neste caso a fala, possibilitando a classificação deste som (fonema) e o reconhecimento de palavras.

Desta forma, pode-se diminuir erros de classificação de fumo, visto que um operador pode entender de forma errada a classificação falada pelo classificador, além de agilizar o processo de compra de fumo (não será mais necessário digitalizar

códigos para a classificação da folha de fumo). Outra vantagem é a diminuição do custo da mão-de-obra na empresa, sendo necessário apenas um colaborador para exercer a tarefa de classificação, onde antes eram necessárias duas pessoas.

Esta dissertação tem o objetivo de realizar um estudo das técnicas de reconhecimento de padrões de áudio utilizando as transformadas *Wavelets* Haar e Daubechies, definir qual a melhor técnica para ser utilizada dentro de uma indústria fumageira para classificação de comandos de voz no processo de compra de fumo e possibilitar, em trabalhos futuros, desenvolver e aplicar estas técnicas em um sistema de reconhecimento e classificação de comandos de voz utilizando *wavelets*.

O projeto está organizado da seguinte forma: no capítulo 1 é apresentada a fundamentação teórica, na qual são introduzidos conceitos sobre a voz humana, análise de Fourier, *Wavelets*, Análise de Componentes Principais (PCA) e Cepstrum. No capítulo 2 é apresentada a metodologia utilizada para o desenvolvimento do trabalho e, por fim, as conclusões a que se chegou ao término desta dissertação.

# 1 FUNDAMENTAÇÃO TEÓRICA

O reconhecimento de comandos de voz por computador sempre fascinou o homem. Vários filmes apresentam a possibilidade de conversar com um computador, digitar um texto ou emitir comandos utilizando a voz. Algumas empresas dedicam seus esforços para permitir esta tarefa. Um bom exemplo é o *Via Voice®* da IBM, um programa comercial que permite, entre outras coisas, ditar cartas em produtos como o *Microsoft Office®*. Outras empresas como a Microsoft e a Dragon System oferecem ótimos sistemas, ou *engines* (soluções em hardware ou software que trabalham diretamente com a placa de som, capturando as falas ou sintetizando palavras), para tratamento de voz (CASTILHO, 2005).

O reconhecimento da fala consiste em identificar fonemas, sílabas, palavras para formar a mensagem original, ou uma informação onde existe uma seleção mais direta da resposta, sem interpretação dela. Assim, uma ação pode ser executada diretamente quando um padrão falado é reconhecido (PAULA, 2000).

Segundo Barone (1999), citado por Paula (2000, p. 3),

O reconhecimento automático de voz consiste no processo de extrair automaticamente a informação lingüística do sinal da fala, a qual está codificada. Este processo normalmente acontece em três etapas: aquisição do sinal de voz, extração de parâmetros e reconhecimento do padrão.

Para um melhor entendimento sobre reconhecimento de voz, torna-se importante uma breve apresentação sobre a origem da fala, seus sons e fonemas da língua utilizada. Não existe certeza sobre o momento em que o homem começou a falar, mas se supõe que a linguagem tenha começado a desenvolver-se há cerca de 1,5 milhão de anos, com o *Australopithecus*. Um dos últimos passos na evolução da linguagem ocorreu com o *Homo Sapiens spp* há mais de 40.000 anos (BARROS, 1998).

A fala é uma das capacidades ou aptidões que os seres humanos possuem de comunicação, manifestando seus pensamentos, opiniões e sentimentos através de vocábulos que se transladam textualmente quando necessários. Consiste no

principal sinal entre os distintos sinais abordados pela linguagem natural, como por exemplo, ideogramas, gestos, gritos, trejeitos e outros tipos de linguagem corporal (PAULA, 2000).

Os sons da fala não devem ser confundidos com os fonemas da Língua Portuguesa, uma vez que o som é entendido como um evento acústico de cada unidade sonora da fala, enquanto que os fonemas correspondem a percepção eclética e interpretativa realizada pelo falante e ouvinte, respectivamente. O fonema é a menor unidade sonora ou articulatória da palavra (LUFT, 1995). Os fonemas são produzidos pelo aparelho fonador, que é constituído por um conjunto de órgãos que auxiliam no ato da fala. A fala representa a emissão de sons articulados na linguagem oral enquanto que a produção da voz é a fonação (PAULA, 2000).

O objetivo deste trabalho é realizar um estudo das técnicas de reconhecimento de padrões de áudio, utilizando as transformadas Wavelets, em conjunto com a análise de componentes principais e Cepstrum.

Aliado a essa não exaustiva explanação sobre fala e fonema, esse capítulo visa fornecer subsídios teóricos para um melhor entendimento do trabalho realizado. Sendo assim, serão abordados conceitos sobre a Transformada Wavelet, Análise de Componentes Principais e Cepstrum.

## **1.1 Wavelets**

Segundo Daubechies (1992), “A transformada *wavelet* é uma ferramenta que fatia dados ou funções ou operadores em componentes freqüenciais diferentes, e então estuda cada componente com uma resolução casada com sua escala”.

Para que a análise através de *Wavelets* seja compreendida, é importante resgatar seu histórico e fundamentação.

### 1.1.1. Histórico da Teoria Wavelet

A análise *wavelets*, sob um ponto de vista histórico, pode ser considerada um novo método, apesar de sua fundamentação matemática ter sido derivada do trabalho de Joseph Fourier no século XIX, através de suas teorias sobre análise freqüencial.

Gradativamente, a atenção de pesquisadores foi migrando da análise baseada em freqüência para a análise baseada em escala, que é a abordagem da análise por *wavelets*.

As Transformadas de *Wavelets* foram introduzidas pela Escola Francesa (Morlet, Grossmann, Meyer, Battle, Lemarié, Cohen, Mallat, Coifman, Rioul, etc.) originária de estudos de curta duração associados a pacotes de ondas acústicas sísmicas. O procedimento de análise adota uma função protótipo *wavelet*, chamada de *wavelet* analisadora ou *wavelet-mãe*. O termo cunhado originariamente era *ondelettes*, que significa algo como "ondinhas", e *wavelets* corresponde a uma versão anglofônica. (OLIVEIRA, 2002)

A primeira menção ao que hoje é chamado de *wavelet* parece ter sido de 1909, em uma tese de Alfred Haar. O Conceito de *wavelets* na sua forma teórica atual foi proposto pela primeira vez no início da década de 80 por Alex Grossmann (*Université de Marseille*) e Jean P. Morlet (*Elf Aquitaine*). Eles introduziram o conceito de *wavelets* quando analisavam sinais geofísicos destinados à exploração de petróleo e gás, e constataram que, durante as repentinas variações do sinal, a análise de Fourier não estava sendo eficiente (Morlet recebeu o prêmio *Reginald Fessenden Award* 1997).

Em 1985, Stéphane Mallat (França) estabeleceu a ligação desta teoria com o processamento digital de sinais, particularmente com os filtros espelhados em quadratura, algoritmos piramidais. Yves Meyer (França) construiu uma das primeiras *wavelets* não triviais, continuamente diferenciáveis (embora não fossem de suporte



compacto). Ingrid Daubechies (Bélgica) construiu o mais usado conjunto de *wavelets* ortogonais de suporte compacto (tempo-limitada). (OLIVEIRA, 2002)

A descrição matemática para a teoria de *wavelets* discretas foi construída no final da década de 80 (particularmente por Daubechies, Mallat, Meyer), que asseguraram a disseminação dos métodos. O algoritmo principal é da época do trabalho de Stephane Mallat em 1988.

Segundo Wickerhauser (1994), citado por Dias (2003), desde então a pesquisa em *wavelets* se tornou internacional, sendo particularmente ativa nos EUA graças ao trabalho de cientistas como Ingrid Daubechies, Ronald Coifman e Victor Wickerhause.

### **1.1.2. Teoria Wavelet**

Um dos grandes diferenciais da teoria *wavelet* é a capacidade de analisar sinais com espectro variante no tempo. Pode-se entender as *wavelets* como mecanismos para decompor ou quebrar sinais nas suas partes constituintes, permitindo analisar os dados em diferentes domínios de freqüências com a resolução de cada componente amarrada à sua escala (LIMA, 2003).

Segundo Qian; Chen (1996), citado por Dias (2003), tradicionalmente os sinais são estudados como função do tempo ou como função da freqüência. Entretanto, a maioria dos sinais encontrados na prática apresentam espectros variantes no tempo, como por exemplo tons de música. Na Natureza, poucos sinais possuem conteúdo frequencial que não mude ao longo do tempo. Em várias aplicações práticas, caracterizar o sinal simultaneamente nos domínios temporal e frequencial é de grande utilidade, como no processamento de sinais de voz.

Para um melhor entendimento do conceito da teoria de *wavelets* é importante relembrar resumidamente os fundamentos da análise de Fourier, que decompõe um sinal em componentes senoidais de diferentes freqüências. É através dessa

decomposição que se consegue analisar o espectro de freqüências do sinal, conforme mostra a figura 1(b).



Figura 1 – (a) Gráfico de um sinal – amplitude x tempo, (b) Gráfico de um sinal – amplitude x freqüência (Transformada de Fourier)

Embora para muitos sinais a análise de Fourier seja extremamente útil, ela apresenta uma séria deficiência. A transformação espectral acarreta perda total da informação temporal do sinal, e vice-versa com a transformação inversa.

No caso de um sinal onde as propriedades não se modificam substancialmente ao longo do tempo, ou se o sinal for estacionário, esta deficiência não é muito relevante. Porém a maioria dos sinais de interesse contém diversas características transitórias ou não estacionárias como desvios, tendências, mudanças bruscas, começos e términos de eventos. Nestes casos, a análise de Fourier não é adequada para detectá-las.

Para tentar superar a deficiência previamente citada, em 1946 Dennis Gabor propôs uma adaptação à transformada de Fourier, na qual apenas uma pequena seção do sinal em um dado instante era analisada – uma técnica chamada de “janelamento” do sinal. A adaptação de Gabor, chamada Transformada de Fourier em Intervalos Curtos (*Short-Time Fourier Transform – STFT*), mapeia um sinal em uma função bi-dimensional do tempo e da freqüência, conforme ilustrado na figura 2.

O janelamento temporal traz consigo uma incerteza associada à precisão da análise que não havia na transformada Fourier convencional. Essa imprecisão se manifesta em uma relação de compromisso entre a resolução freqüencial e a

capacidade de acompanhar as variações espectrais ao longo do tempo. Além disso, quanto menor a largura da janela, maior o esforço computacional necessário para realizar a análise espectral do sinal (DIAS, 2003).

Na STFT, a largura da janela temporal é fixa para um dado sinal analisado e, em consequência disso, o grau de incerteza associado ao janelamento é o mesmo para todo o espectro conjunto freqüencial-temporal.

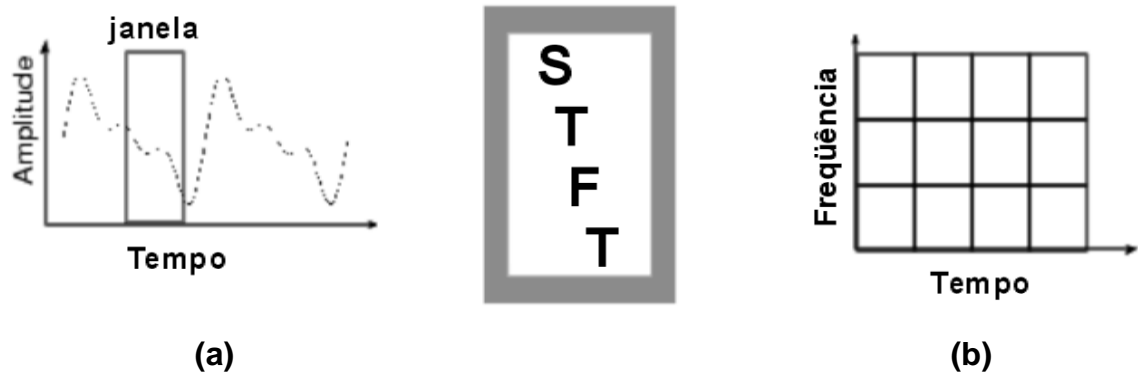


Figura 2 - (a) Gráfico de um sinal com janela temporal - amplitude x tempo, (b) Gráfico de um sinal - freqüência x tempo (STFT).

Existem diversos casos em que a análise de um determinado sinal prioriza faixas específicas do espectro, dando menor importância ao restante do conteúdo freqüencial. Como exemplo, pode-se citar o estudo de sinais sísmicos. Grande parte do tempo, o sismógrafo registra um sinal "de base" e somente na ocorrência de terremotos, um sinal rápido, curto, de freqüência mais elevada, aparece. Analisar estes sinais em tempo real, monitorando via Transformada de Fourier, não é tão eficaz.

Para explorar melhor essa situação, seria interessante uma transformação que flexibilizasse a largura da janela temporal, aplicando janelas mais estreitas às áreas espectrais desejadas e aumentando a largura para o espectro de menor interesse. Uma das técnicas que permite essa abordagem diferenciada é a análise por *wavelets*.

Uma das estratégias de janelamento temporal mais utilizadas na análise por *wavelets* é a que aplica as menores larguras ao conteúdo de alta freqüência

(intimamente associado às variações rápidas), e as maiores larguras ao conteúdo espectral baixo (variações lentas), ou seja, a resolução no tempo torna-se arbitrariamente boa para altas frequências, enquanto que a resolução em frequência torna-se arbitrariamente boa em baixas frequências.

Na figura 3b, é possível verificar esta estratégia de janelamento, na qual a escala é interpretada como uma grandeza inversamente proporcional à frequência, ou seja, quanto maior a escala, menor a frequência, e, conseqüentemente, maiores são as larguras da janela.



Figura 3 - (a) Gráfico de um sinal - amplitude x tempo, (b) Gráfico de um sinal – escala x tempo (wavelets).

As *wavelets* constituem também transformadas lineares, o que é bastante atrativo do ponto de vista prático e de manipulação matemática. Para exemplificar o problema de escala, pode-se utilizar um exemplo com fotografias, imaginando que se deva usar uma representação para um indivíduo de meia-idade com uma meia dúzia de fotos (0 ano, 5 anos, 10 anos, 15 anos, 20 anos, 25 anos). Ao invés de se tomar uma foto a cada 5 anos, um modelo mais interessante pode levar em conta que as mudanças ocorridas nos 10 primeiros anos são mais significativas que aquelas para o mesmo intervalo, porém com maior faixa etária. Talvez fosse mais interessante considerar um conjunto de seis fotos tomadas a 1 ano, 2 anos, 4 anos, 8 anos, 16 anos e 32 anos. Isto leva em conta que as diferenças entre as fotografias aos 20 e 25 anos são praticamente idênticas, pouco adicionando ao conhecimento. (OLIVEIRA, 2002)

Pode-se definir uma *wavelet* como uma forma de onda de duração limitada e que possui um valor médio igual a zero. São funções obtidas a partir de uma função protótipo, a “*wavelet* mãe”, por meio de dilatações (contrações ou escalamento) e translações (deslocamentos). Inicialmente é possível comparar uma *wavelet* com uma senóide, que é a base da análise de Fourier. Senóides são ilimitadas no tempo, suaves e previsíveis, enquanto *wavelets* tendem a ser irregulares e assimétricas.

A figura 4 ilustra essas diferenças, onde (a) demonstra uma senóide com ondas regulares constantes e (b) com ondas irregulares e assimétricas.

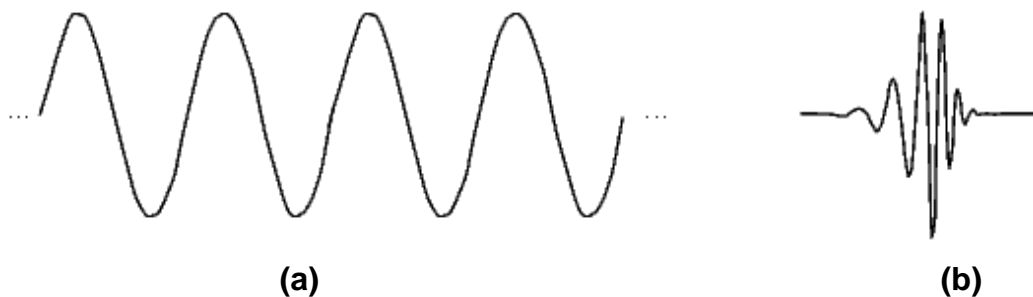


Figura 4 - Comparação entre uma senóide (a) e uma *wavelet* db10 (b).

A análise de Fourier consiste em decompor um sinal em ondas senoidais de várias frequências, assim como, de forma análoga, a análise por wavelets é a decomposição de um sinal em versões “deslocadas” e “escaloadas” da *wavelet* original (ou “*wavelet* mãe”).

A maioria das *wavelets* de interesse são funções ditas “localizadas”, ou seja, com sua energia concentrada nas proximidades de um determinado ponto, tanto no tempo quanto em escala (frequência). É essa característica das *wavelets* que viabiliza aplicações como a compactação do sinal, a focalização da análise para uma região específica de interesse do espectro variante no tempo, ou ainda a localização das áreas de maior concentração de energia.

Ao invés de interpretar os parâmetros nos domínios tempo e frequência ( $f \times t$ ), costuma-se utilizar os domínios escala e deslocamento ( $a \times b$ ). Uma interpretação interessante está associada a lidar com imagens tipo “mapas”. Uma mudança de escala pode permitir, numa escala maior, ter uma visão mais global, mas com menor

precisão. Já em uma escala menor, vê-se detalhes, mas se perde em estudar o comportamento global. A análise via *wavelets* permite, por assim dizer, visualizar tanto a “floresta” quanto as “árvores”. (OLIVEIRA, 2002)

O tratamento analítico para a análise por *wavelets* inclui a transformada contínua de *wavelet*, assim como a discreta, além de suas respectivas transformadas inversas. A transformada contínua traz consigo uma grande redundância de informações sobre o sinal analisado, o que a torna computacionalmente desinteressante. Via de regra, a transformada discreta é utilizada, seja em sua versão mais simples, na chamada “análise de multiresolução”, seja na versão que permite um detalhamento personalizado do espectro, que é a “análise por pacotes”. (DIAS, 2003)

Na Transformada Contínua de *Wavelet* (*Continuous Wavelet Transform* - CWT), todas as respostas ao impulso no banco de filtros são versões escalonadas (expandidas ou comprimidas) de uma mesma  $\psi(t)$ , chamada de *wavelet* básica (*wavelet-mãe*), conforme equação 1.

$$\text{CWT}(\alpha, \tau) := \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{+\infty} f(t) \psi^*\left(\frac{t-\tau}{\alpha}\right) dt. \quad \dots(1)$$

A função  $\psi(t)$  é conhecida como *wavelet* mãe. A partir dela geram-se versões modificadas no decorrer da transformada. O termo mãe vem do fato que funções com diferentes tamanhos são usadas no processo da transformada e todas são originadas de uma *wavelet* principal. Ela é um protótipo para a geração de outras funções janela. Todas as janelas a serem utilizadas são de fato versões dilatadas e comprimidas da mesma "*wavelet* mãe" (GRO; TORR, 2001 e BULT, 1995 citado por OLIVEIRA, 2002)

### 1.1.2.1. A Transformada de Wavelet Contínua CWT

A Transformada de *Wavelet* foi desenvolvida como uma alternativa à STFT para solucionar o problema da resolução.

A análise com *wavelets* é feita similarmente à análise com STFT no que diz respeito à multiplicação do sinal por uma função (que neste caso será a *wavelet* e não mais uma janela, como na STFT). A transformada é calculada separadamente por segmentos diferentes do sinal no domínio do tempo. (OLIVEIRA, 2002)

É importante notar que, uma vez fixada a janela para STFT, a resolução no tempo ( $t$ ) e na frequência ( $f$ ) permanecem constante em todo o plano  $t \times f$ , como mostrado na figura 5.

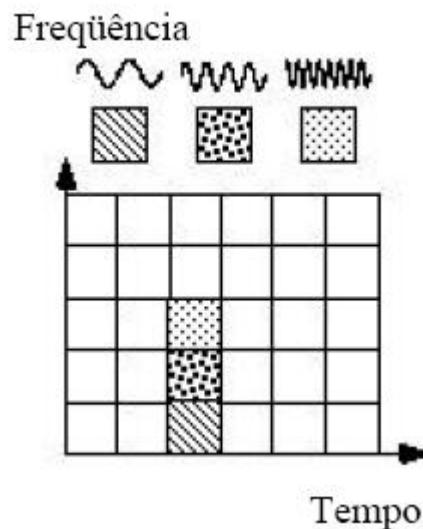


Figura 5 - Resolução no plano  $t$ - $f$  pela análise STFT

Os problemas na resolução do tempo e da frequência são resultados de um fenômeno físico conhecido como *Princípio de Gabor-Heisenberg*. Este fenômeno é indiferente em relação à transformada usada. Existe uma forma alternativa para analisar um sinal, chamada de Análise Multi-Resolucional (**AMR**), que analisa o sinal em frequências diferentes com resoluções diferentes, conforme apresentado na figura 6.

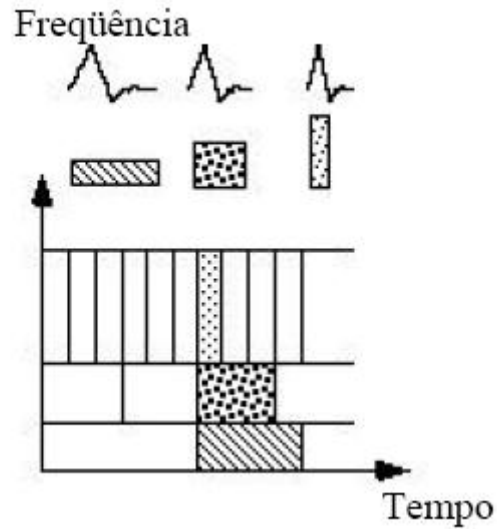


Figura 6 - Análise Multi-Resolucional (Transformada de *wavelet*)

Pode-se definir as *wavelets* como funções matemáticas que separam dados em seus diferentes componentes freqüenciais, e extraem cada componente com uma determinada resolução, de acordo com a sua escala. Comparando-se com a análise de Fourier, a análise por *wavelets* possui vantagens, pois na Fourier a análise é feita como um todo, acarretando em uma representação mais pobre para sinais que possuem descontinuidades e variações bruscas.

A transformada de *Wavelets* Contínua pode ser expressa pela equação 2:

$$\psi(t) \in L^2(\mathfrak{R}). \int_{-x}^{+x} \psi^2(t) dt < +\infty \text{ e } E_\psi = \langle \psi, \psi \rangle. \quad \dots(2)$$

Nela é possível executar operações de escalonamento (equação 3), deslocamento (equação 4) e deslocamento com escalonamento (equação 5):

$$\psi_a(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t}{a}\right), a \neq 0. \quad \dots(3)$$

$$\psi_b(t) = \psi(t - b). \quad \dots(4)$$



$$\psi_{a,b}(t) = \psi_a(t-b) \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \text{ onde}$$

$$\{\psi(t)\} \rightarrow \{\psi_{a,b}(t)\} (\forall a, a \neq 0) (\forall b \in \mathfrak{R}). \quad \dots(5)$$

Em termos matemáticos, dada uma função  $f(t)$ , a função  $f(s.t)$  corresponde a uma versão **comprimida** se  $s>1$  e **expandida** se  $s<1$ . No caso da transformada de *wavelets*, o parâmetro escala aparece no denominador, ou seja, uma versão comprimida da *wavelet* mãe se  $a<1$  e dilatada se  $a>1$ .

O ajuste na amplitude do sinal escalonado foi introduzido visando garantir a isomeria: todas as *wavelets* tem a mesma energia.

Portanto, defini-se a transformada de *wavelet* continua como na equação 6:

$$\text{CWT}(a, b) := \int_{-\infty}^{+\infty} f(t) \psi_{a,b}^*(t) dt = \langle f(t), \psi_{a,b} \rangle. \quad \dots(6)$$

A decomposição de *wavelets* considera a decomposição de  $f$  em sinais  $\{\psi_{a,b}(t)\} \{\psi_{a,b}(t)\}_{a \in \mathfrak{R}_+^*, b \in \mathfrak{R}}$  que constitui um novo conjunto de análise do espaço de sinais.

Esta nova "base" (de fato, este conjunto não é formalmente uma base) é composta por sinais oscilatórios e de "curta duração" (tais ondas não necessitam ter duração estritamente finita, porém devem decair rapidamente). A combinação oscilatória (daí o termo onda) e de curta duração (inha) gera o termo ondinhas, ondeletas, ondelettes no original, ou de forma já consagrada, *wavelets*. (OLIVEIRA, 2002)

A Teoria mais formal considera dois parâmetros  $a$  e  $b$  conhecidos como escala e deslocamento, respectivamente. Para  $a \neq 0$ , define-se conforme equação 7:

$$\text{CWT}(a, b) := |a|^{-1/2} \int_{-\infty}^{+\infty} f(t) \psi^*\left(\frac{t-b}{a}\right) dx = \langle f(t), \psi_{a,b} \rangle. \quad \dots(7)$$

A idéia fundamental da Transformada de *Wavelet* é que ela é uma transformada pontual e proporcional à escala. Ela analisa o sinal em escalas diferentes e se desloca analisando cada trecho do sinal. O parâmetro translação se relaciona com a localização da “janela”. Analisa-se o sinal aos poucos. Este termo corresponde, obviamente, à informação de tempo no domínio da transformada. Processa-se essencialmente o conteúdo que estiver dentro da janela. (GOMES citado em OLIVEIRA,2002)

Existe um grande número de funções que podem ser eleitas como *wavelets* mãe:

Nome da família de *Wavelets*:

- 'haar' Haar wavelet.
- 'db' Daubechies wavelets.
- 'sym' Symlets.
- 'coif' Coiflets.
- 'bior' Biorthogonal wavelets.
- 'rbio' Reverse biorthogonal wavelets.
- 'meyr' Meyer wavelet.
- 'dmey' Discrete approximation of Meyer wavelet.
- 'gaus' Gaussian wavelets.
- 'mexh' Mexican hat wavelet.
- 'morl' Morlet wavelet.
- 'cgau' Complex Gaussian wavelets.
- 'shan' Shannon wavelets.
- 'fbsp' Frequency B-Spline wavelets.
- 'cmor' Complex Morlet wavelets.

Para se ter uma idéia da gama de problemas que podem ser resolvidos utilizando *wavelets*, pode-se citar alguns como:

- remoção de ruído em áudio;
- compressão de sinal;
- detecção de objetos;

- compressão de impressões digitais;
- remoção de ruído em imagens;
- reconhecimento de imagens;
- diagnósticos de problemas de coração;
- reconhecimento de voz.

Devido ao baixo número de coeficientes nos filtros das transformadas Haar e Daubechies (Daub4), estas funcionam muito bem quando se trata de dados, que se traduzem em sinais tipo somatórios de escadas. Para sinais mais complicados ou ainda para uma multiresolução ótima, pode-se trabalhar com outras transformadas *wavelets*, mas, neste trabalho, os estudos foram direcionados especificamente às duas famílias de *wavelets* acima destacadas.

Sendo assim, realiza-se a seguir um estudo mais específico das *wavelets* Haar e Daubechies.

### **1.1.3. Wavelets Haar**

A *wavelet* Haar é o tipo mais simples de *wavelets*. Na forma discreta, as *wavelets* Haar estão relacionadas a operações matemáticas chamadas Transformadas Haar. A Transformada Haar serve como protótipo para todas as outras transformadas *wavelets*. Ao se estudar esta transformada em detalhe, obtém-se uma boa base para o entendimento das transformadas *wavelets* mais complexas.

### 1.1.3.1. Transformada Haar

Primeiramente, para que se possa estudar a Transformada Haar, deve-se definir qual o tipo de sinal que será analisado. Como exemplo, serão utilizados sinais discretos. Um sinal discreto é uma função cujos valores são definidos em instantes discretos do tempo (ocorrem em sucessão) e, portanto, é uma seqüência de números, podendo ser expresso na forma de:  $\mathbf{f} = (f_1, f_2, \dots, f_N)$ , onde  $N$  é um número inteiro positivo referente ao comprimento de  $\mathbf{f}$  (WALKER, 1999).

Como toda *wavelet* transforma um sinal, a transformada de Haar decompõe um sinal discreto em outros dois sinais, os quais pode-se chamar de sub-sinais. Um destes sub-sinais é de média ou de tendência e o outro subsinal é uma diferença ponderada de dois impulsos vizinhos.

Examinando-se primeiramente o subsinal de tendência,  $\mathbf{a}_1 = (a_1, a_2, \dots, a_{N/2})$ , para a função  $\mathbf{f}$  que será processada para se obter a média da seguinte forma: seu primeiro valor,  $a_1$ , é calculado através da média do primeiro par de valores de  $\mathbf{f}$ :  $(f_1 + f_2)/2$ , e então multiplicando por  $\sqrt{2}$ , ou seja,  $a_1 = (f_1 + f_2)/\sqrt{2}$ . Da mesma forma, essa regra se aplica para os dois próximos pares de valores do sinal  $a_2 = (f_3 + f_4)/\sqrt{2}$  e assim sucessivamente até que seja percorrido todo o sinal. A fórmula que define os valores de  $\mathbf{a}_1$  pode ser representada na equação 8:

$$a_m = \frac{f_{2m-1} + f_{2m}}{\sqrt{2}} \quad \dots(8)$$

onde  $m = 1, 2, 3, \dots, N/2$ .

Imaginando-se, como exemplo, que  $\mathbf{f}$  seja definido por oito valores:

$f = (4, 6, 10, 12, 8, 6, 5, 5)$ , o subsinal de tendência, aplicando a equação 1, seria  $a^1 = (5\sqrt{2}, 11\sqrt{2}, 7\sqrt{2}, 5\sqrt{2})$ . Pode-se também calcular conforme mostrado no diagrama da figura 7:

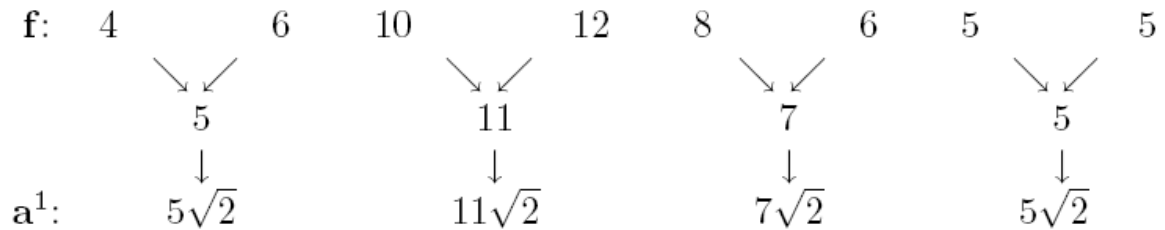


Figura 7 – Lógica do cálculo do primeiro grupo de subsinais de Haar.

O outro subsinal é chamado de primeira flutuação, ou subsinal de diferença. Este subsinal de  $f$  é representado por  $d^1 = (d^1, d^2, \dots, d^{N/2})$ , onde o cálculo é feito levando a metade da diferença do primeiro par de valores de  $f$ :  $(f_1 - f_2)/2$ , e multiplicando por  $\sqrt{2}$ , ou seja,  $d^1 = (f_1 - f_2)/\sqrt{2}$ . Esta mesma regra se aplica para os dois próximos pares de valores do sinal:  $d^2 = (f_3 - f_4)/\sqrt{2}$ .

Desta forma, tem-se para todos os valores do sinal, conforme mostrado na equação 9:

$$d^m = \frac{f^{2m-1} - f^{2m}}{\sqrt{2}} \quad \dots(9)$$

onde  $m = 1, 2, 3, \dots, N/2$ .

Utilizando como exemplo o sinal  $f = (4, 6, 10, 12, 8, 6, 5, 5)$ , a primeira flutuação  $d^1$  é  $(-\sqrt{2}, -\sqrt{2}, \sqrt{2}, 0)$ . Este resultado é obtido através da equação 2. Pode-se também calcular conforme mostrado no diagrama da figura 8:

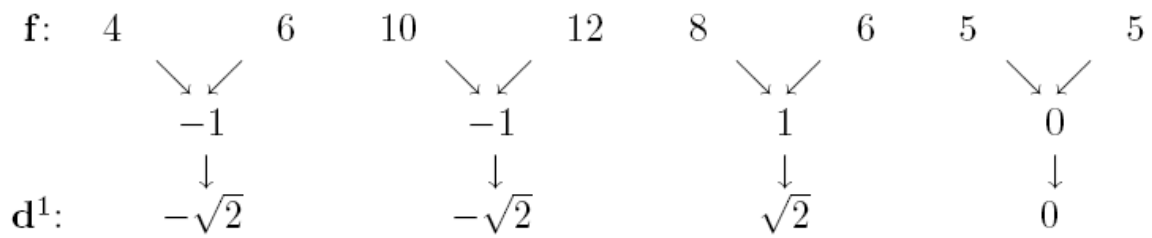


Figura 8 – Lógica do cálculo da primeira flutuação  $d^1$

### 1.1.3.2. A Transformada Haar, 1-Nível

A transformada Haar é desenvolvida em vários estágios, ou níveis. O primeiro nível compreende o mapeamento da transformada  $H1$ , definida pela equação 10:

$$f \xrightarrow{H1} (a^1 | d^1) \quad \dots(10)$$

A função  $f$  é formada por um sinal discreto onde  $a^1$  é o subsinal de tendência e o  $d^1$  é o subsinal de diferença. Como exemplo, tem-se a equação 11:

$$(4, 6, 10, 12, 8, 6, 5, 5) \xrightarrow{H1} (5\sqrt{2}, 11\sqrt{2}, 7\sqrt{2}, 5\sqrt{2} | -\sqrt{2}, -\sqrt{2}, \sqrt{2}, 0) \quad \dots(11)$$

O mapeamento da transformada  $H1$ , apresentado na equação 11, possui uma transformada inversa, onde é possível recompor o sinal original transformando o sinal  $(a^1 | d^1)$  de volta ao sinal  $f$ , expresso pela equação 12:

$$f = \left( \frac{a1+d1}{\sqrt{2}}, \frac{a1-d1}{\sqrt{2}}, \dots, \frac{aN/2+dN/2}{\sqrt{2}}, \frac{aN/2-dN/2}{\sqrt{2}} \right) \quad \dots(12)$$

Em outras palavras,  $f1 = (a1 + d1)/\sqrt{2}$ ,  $f2 = (a1 - d1)/\sqrt{2}$ ,  $f3 = (a2 + d2)/\sqrt{2}$ ,  $f4 = (a2 - d2)/\sqrt{2}$ , e assim por diante.

Como exemplo, na figura 9 tem-se um sinal formado por uma amostragem de 1024 valores gerados através da função  $g(x) = 20x^2(1-x)^4 \cos 12\pi x$ , entre os intervalos (0,1).

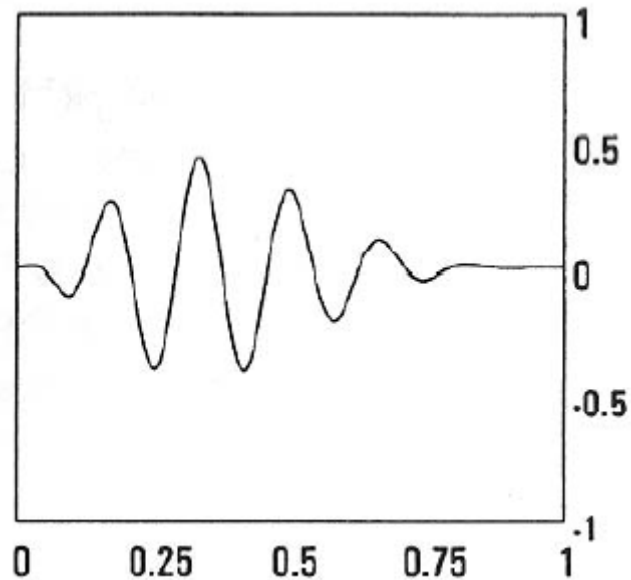


Figura 9 – Sinal original

Na figura 10 é mostrado o mesmo sinal após aplicada a transformada Haar. Pode-se observar que a onda conservou o mesmo número de impulsos em relação ao tempo, porém começou a compactar e aumentar a sua amplitude.

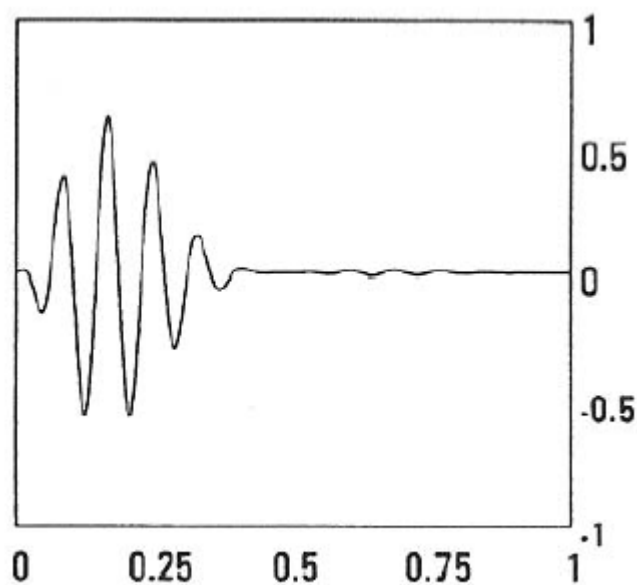


Figura 10 – Sinal após a aplicação da transformada Haar

### 1.1.4. Wavelets Daubechies

Nesta seção será apresentado o conceito sobre uma das transformadas *wavelets* chamada Daubechies. As transformadas de Daubechies são definidas da mesma forma que as Transformadas de Haar – através da execução de médias e diferenças das escalas de sinais e *wavelets* - a única diferença entre elas consiste em como essa escala de sinais e das *wavelets* são definidas.

Para as transformadas *wavelets* de Daubechies, a escala de sinais e as *wavelets* são ligeiramente mais longas, ou seja, elas produzem médias e diferenças usando alguns valores a mais do sinal. Esta pequena diferença, no entanto, provê uma grande melhoria da capacidade dessas novas transformadas. Elas proporcionam diversas e poderosas ferramentas para executar algumas tarefas básicas de processamento de sinal. Estas tarefas incluem compressão e remoção de ruído de sinais de áudio e de imagem, e valorização da imagem e reconhecimento do sinal.

#### 1.1.4.1. As Wavelets Daub4

Existem diversas transformadas Daubechies, mas todas são muito semelhantes. Será realizado um estudo mais aprofundado na mais simples delas, a transformada *wavelet* Daub4.

A transformada *wavelet* Daub4 é definida em sua essência da mesma forma que a transformada *wavelet* de Haar.

Se um sinal  $\mathbf{f}$  tem um número  $\mathbf{N}$  de valores, então a transformada Daub4 1-Nível é o mapeamento  $\mathbf{f} \xrightarrow{D_1} (\mathbf{a}^1 | \mathbf{d}^1)$  do sinal  $\mathbf{f}$  para o primeiro subsinal de tendência  $\mathbf{a}^1$  e primeiro subsinal de flutuação  $\mathbf{d}^1$ .



Cada valor  $a_m$  de  $a^1 = (a_1, \dots, a_{N/2})$  é igual ao produto escalar de  $a_m = f \cdot V_m^1$  de  $f$  com um sinal escalar 1-Nível  $V_m^1$ . Todavia, cada valor  $d_m$  de  $d^1 = (d_1, \dots, d_{N/2})$  é igual ao produto escalar de  $d_m = f \cdot W_m^1$  de  $f$  com uma *wavelet*  $W_m^1$  1-Nível.

A transformada de *wavelet* Daub4, assim como a transformada de Haar, pode ser estendida em múltiplos níveis, tantas vezes quanto o tamanho do sinal puder ser dividido por 2.

Essa extensão é similar ao método utilizado na transformada de Haar. Por exemplo, aplicando-se a transformada Daub4 1-Nível  $D_1$  à primeira tendência  $a^1$ , tem-se o mapeamento  $a^1 \xrightarrow{D_1} (a^2 | d^2)$  do primeiro subsinal de tendência  $a^1$  para o segundo subsinal de tendência  $a^2$  e segundo subsinal de flutuação  $d^2$ . A transformada Daub4 2-Nível é então definida pelo mapeamento  $f \xrightarrow{D_2} (a^2 | d^2 | d^1)$ . A figura 11 mostra o sinal original (a) e sua transformada Daub4 2-Nível (b).

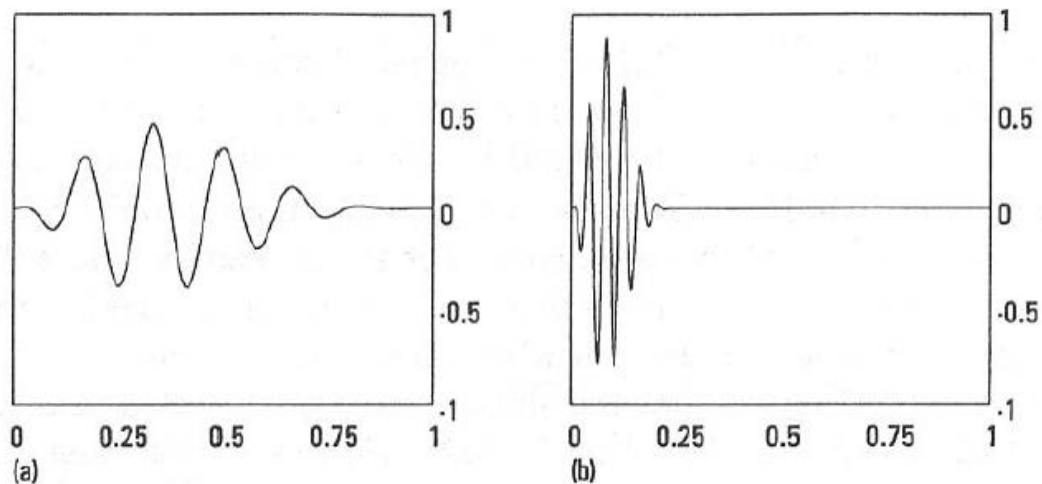


Figura 11 – (a) Sinal Original, (b) Transformada Daub4 2-Nível

Assim como na transformada de Haar, os valores da segunda tendência  $a^2$  e segunda flutuação  $d^2$  podem ser obtidos através de produto escalar com sinais de segundo nível e *wavelets*. Da mesma forma, a definição de uma transformada Daub4  $k$ -Nível é obtida aplicando a transformada 1-Nível do subsinal de tendência do nível anterior, exatamente como realizado na transformada Haar.

A diferença entre a transformada Haar e a transformada Daub4 está na forma em que as escalas dos sinais e as *wavelets* são definidas (WALKER, 1999).

Aliado à utilização das *Wavelets* para reconhecimento de padrões, existe um método chamado Análise de Componente Principais (ACP), ou do inglês, PCA – *Principal Component Analysis*, que pode ser utilizado para identificar as características comuns entre amostras de sinais e foi escolhido pela facilidade de sua implementação no Matlab.

## 1.2 Análise de Componentes Principais

A Análise de Componentes Principais (ACP ou PCA, do inglês *Principal Component Analysis*) é um dos métodos estatísticos mais utilizados quando se pretende analisar dados multivariados que representa a base para o entendimento de muitas outras técnicas de análise de dados. A análise de componentes principais tem por objetivo reduzir a dimensão dos dados originais permitindo a visualização das informações mais importantes em um número menor de fatores, ou componentes principais (RIBEIRO e FERREIRA, 2005).

A análise de componentes principais consiste essencialmente em reescrever as coordenadas das amostras em outro sistema de eixo mais conveniente para a análise dos dados. Em outras palavras, as *n-variáveis* originais geram, através de suas combinações lineares, *n-componentes* principais, cuja principal característica, além da ortogonalidade, é que são obtidos em ordem decrescente de máxima variância, ou seja, a componente principal 1 detém mais informação estatística que a componente principal 2, que por sua vez tem mais informação estatística que a componente principal 3 e assim por diante (MOITA NETO e MOITA, 1997).

As amostras são projetadas num novo sistema de eixos (chamados de componentes principais ou fatores) de dimensão menor, reduzindo assim a

dimensionalidade do espaço do conjunto dos dados. Com isto, as informações mais importantes e relevantes se tornam mais óbvias.

Na PCA, a matriz original dos dados  $X$  é decomposta como descrito na equação 13, em que  $T$  é chamada matriz dos escores e  $L$  é conhecida como a matriz dos pesos. (RIBEIRO e FERREIRA, 2005).

$$X = T * LT \quad \dots(13)$$

Os escores representam a projeção das amostras no novo sistema de coordenadas, e os pesos são os coeficientes que medem a importância de cada variável original em cada componente principal (PC, do termo em inglês, Principal Component ), ou seja, o peso que cada variável tem naquela combinação linear. Estes pesos nada mais são do que o cosseno do ângulo entre o eixo da componente principal e o eixo da variável original, e portanto seu valor estará sempre entre 1 e -1. Quanto mais próximo de  $\pm 1$ , maior a influência que uma determinada variável tem na descrição desta componente principal *i.e.* , quanto mais próximo de zero este coeficiente estiver, menor a sua influência. (RIBEIRO e FERREIRA, 2005).

Este método permite a redução da dimensionalidade dos pontos representativos das amostras, pois embora a informação estatística presente nas  $n$ -variáveis originais seja a mesma dos  $n$ -componentes principais, é comum obter em apenas duas ou três das primeiras componentes principais mais que 90% desta informação.

Um gráfico contendo uma componente principal 1 *versus* a componente principal 2 fornece uma janela privilegiada (estatisticamente) para observação dos pontos no espaço  $n$ -dimensional.

A análise de componentes principais também pode ser utilizada para julgar a importância das próprias variáveis originais escolhidas, ou seja, as variáveis originais com maior peso (*loadings*) na combinação linear dos primeiros componentes principais são as mais importantes do ponto de vista estatístico.

Muito utilizada em experimentos químicos, a Análise de Componentes Principais consiste em interpretar a distribuição dos pontos no gráfico de componentes principais e identificar as variáveis originais com maior peso na combinação linear das componentes principais mais importantes. (MOITA NETO e MOITA, 1997).

Existem pacotes computacionais de estatística que fazem todas as operações necessárias à obtenção de componentes principais e agrupamento hierárquico, inclusive o tratamento prévio de padronização e escalonamento dos dados.

As componentes principais também podem ser obtidas como um dos métodos da análise de fatores (*Factor Analysis*). O procedimento matemático para obtenção de componentes principais pode ser facilmente seguido por aqueles que têm conhecimento de álgebra matricial e pode ser encontrado em diversos textos dirigidos a profissionais do ramo da química.

A utilização de gráficos para visualizar os escores e pesos irá revelar agrupamentos, tendências das amostras analisadas, e a ocorrência de amostras atípicas, que poderiam ser de difícil visualização no caso das variáveis originais. Na literatura podem ser encontrados diversos trabalhos que detalham os fundamentos algébricos da análise de componentes principais (RIBEIRO e FERREIRA, 2005).

### **1.3 Cepstrum**

Operação matemática que consiste em extrair a Transformada de Fourier do espectro do sinal na forma logarítmica, o *cepstrum* foi inicialmente desenvolvido para caracterizar os ecos das ondas tectônicas provenientes de terremotos e explosões. O primeiro estudo envolvendo esse conceito foi realizado em 1963 por Bogert, Healy, e Tukey (STEIGER, 2008).

O nome deve-se a uma reversão das primeiras quatro letras na palavra "spectrum" (assim: c-e-p-s). O *cepstrum* de um sinal permite dar uma medida da taxa de variação das diferentes bandas do espectro.

### 1.3.1. Power Cepstrum

O *Power Cepstrum* é definido como uma técnica heurística para encontrar os tempos de chegada dos ecos de um sinal. Estes atrasos do eco aparecem como um *ripple* no espectro logarítmico. É possível se obter a frequência deste *ripple*, mas o resultado obtido será em unidades de tempo.

Outras unidades relacionadas à obtenção do *cepstrum* sofrem alterações semelhantes. Para evitar confusões, são definidas algumas unidades análogas (STEIGER, 2008):

freqüência → quefrequency

fase → saphe

amplitude → gamnitude

filtragem → liftering

harmônico → rahmonic

período → repiod

O *power cepstrum* é obtido encontrando o *cepstrum* a partir do espectro de potência do sinal. O algoritmo necessário para obtê-lo é dado por:

Sinal → FT → abs() → square → log → FT → abs() → square → Power Cepstrum

Entretanto, a maioria dos sinais que se deseja analisar no mundo físico não possui transformada de Fourier. Para contornar isto, utiliza-se a transformada Z avaliada no círculo unitário. Com isso, os resultados são os mesmos obtidos no caso anterior:

Sinal → X(z) → abs() → square → log → Z<sup>-1</sup> → square → Power Cepstrum

Matematicamente, achar o *power cepstrum* de uma seqüência qualquer significa aplicar a equação 14:

$$x_{pc}(nT) = \left\{ \frac{1}{2\pi j} \oint_c \log |X(z)|^2 z^{n-1} dz \right\}^2 \quad \dots(14)$$

A partir dessa definição, é possível realizar  $n$  manipulações matemáticas para derivar outras equações. Com isso, conclui-se que:

- É possível decompor dois sinais convoluídos nos seus respectivos *power cepstrum* a partir de uma operação de *liftering*;
- O logaritmo da magnitude quadrada da transformada Z contém *ripple* na forma cossenoidal cuja amplitude (ganntude) e quefrequency (freqüência) estão relacionadas com a amplitude do eco (STEIGER, 2008).

### 1.3.2. Cepstrum Complexo

O *cepstrum* complexo é um caso mais geral, no qual mais informações são preservadas. Ele contém informações sobre a magnitude e a fase do espectro inicial, sendo possível reconstruir o sinal original. A forma de calculá-lo é dada pela equação 15 que é algoritmicamente definida por:

Sinal  $\rightarrow$  FT  $\rightarrow$  abs()  $\rightarrow$  log  $\rightarrow$  Ajuste da fase instantânea  $\rightarrow$  FT  $\rightarrow$  Cepstrum

$$\hat{x}(nT) = \frac{1}{2\pi j} \oint_c \log (X(z)) z^{n-1} dz \quad \dots(15)$$

O cálculo do *cepstrum* complexo computacionalmente pode resultar complicado por conta das operações com logaritmos complexos. Isto pode ser contornado seguindo os seguintes quesitos (STEIGER, 2008):

- A parte imaginária de  $\log(X(z))$  deve ser uma função contínua e periódica (quando avaliada ao redor do círculo unitário) de  $\omega$ , com período  $2\pi/T$ ;

- Caso seja necessário obter uma função real do *cepstrum* complexo de uma função real, a parte imaginária de  $\log(X(z))$  deve ser somada com  $\omega$ .

Manipulando-se matematicamente as equações obtidas, observa-se que:

- O *cepstrum* complexo decai, no mínimo, tão rápido quanto  $1/|n|$ ;
- Possui duração infinita, mesmo fixando a seqüência de entrada a uma duração finita.
- É real se a seqüência é real (pólos e zeros são pares complexos conjugados).

Há claramente uma relação entre o *power cepstrum* e o *cepstrum* complexo. Uma forma simples de relacioná-los é expressa pela equação 16:

$$x_{pc}(nT) = \{Z^{-1}(\log X(z) + \log X^*(z))\}^2 \quad \dots(16)$$

### 1.3.3. Liftering

*Liftering* é o nome dado à aplicação de filtros a um determinado *cepstrum*. A filtragem pode ser implementada aplicando uma janela no domínio cepstral para depois transformá-la novamente ao domínio temporal, resultando um sinal mais liso (STEIGER, 2008).

Os métodos de análises cepstrais são poderosas ferramentas no exercício de processamento digital de sinais. Esses métodos conseguem mostrar informações do domínio espectral que não são evidentes sob o olhar de uma simples DTFT. Isso faz dos métodos imprescindíveis em aplicações específicas, especialmente naquelas onde análises de ecos são realizadas.

Para se ter uma idéia da gama de aplicações do *cepstrum*, pode-se citar algumas como:

- Inúmeras aplicações na área de processamento da voz: reconhecimento de fala, estimação dos parâmetros de modelos da fala, detecção de doenças laringeas, etc;

- Medição das propriedades de superfícies refletoras;
- Projeto de falantes;
- Restauração acústica;
- Processamento de informações sísmicas;
- Eletro-encefalograma;
- Processamento dos sinais do sonar;
- Diagnóstico de máquinas.

Com base nos estudos mencionados acima, se torna possível avançar nos experimentos deste trabalho, que visa utilizar *wavelets* para reconhecimento de comandos de voz.

O capítulo a seguir apresenta a metodologia utilizada para justificar a utilização de *wavelets* neste contexto.



## **2 METODOLOGIA**

A metodologia utilizada para desenvolver a pesquisa compreende algumas etapas, que vêm desde a aquisição da voz em formato digital até o seu processamento e análise, permitindo, dessa forma, gerar a saída esperada (reconhecimento da palavra falada).

### **2.1 Aquisição da Voz**

A primeira etapa executada foi a gravação da voz do classificador de fumo. Por um período aproximado de uma hora durante o processo de classificação de fumo, gravou-se a voz através de um microfone conectado ao microcomputador e um software de áudio, de maneira que fosse possível analisar eventuais ruídos no processo e o próprio desgaste da voz do classificador.

Após a aquisição da voz do classificador de fumo em formato digital, essa foi analisada e nela aplicadas transformadas *wavelets*, de forma que fosse possível concluir qual delas utilizar para adquirir as características particulares de cada fonema, que melhor representasse o sinal adquirido.

### **2.2 Processo de Classificação de Fumo**

Atualmente, as classes de fumo subdividem-se em diversas classes internas, de acordo com determinados padrões de qualidade definidos por cada empresa de fumo.

O processo de classificação de fumo inicia quando a carga de fumo do produtor é descarregada e os fardos, por sua vez, passam pelas esteiras da compra de fumo da empresa.

Nessas esteiras estão localizados terminais de compra de fumo (microcomputadores com *scanner* de mão) que são operados pelo apontador de dados (pessoa responsável por ouvir a classe de fumo falada pelo classificador e selecionar o(s) código(s) correspondente(s) no painel de compra de fumo – painel confeccionado para atender à classificação realizada, contendo códigos de barras referentes às diversas classes utilizadas).

O classificador classifica o fumo fardo a fardo, retirando uma manoca (um feixe de folhas) de fumo do fardo e, através da observação da cor, odor e textura da folha (baseado em seu treinamento e experiência), é capaz de classificar a folha do fumo, dizendo a classe que o fumo se enquadra para o apontador de dados. O fardo segue então o fluxo da esteira, onde é pesado e etiquetado.

Após a classificação, ainda existe o processo de confirmação da compra de fumo, onde o produtor, que está vendendo o fumo, pode concordar ou discordar da classificação realizada. Caso o produtor discorde, o fumo vai para o processo de renegociação. Se não houver consenso entre o produtor e o classificador com relação a classe (qualidade do fumo), o mesmo é devolvido ao produtor.

A figura 12 mostra uma fotografia de um painel de compra com códigos de barras que são utilizados para a classificação.

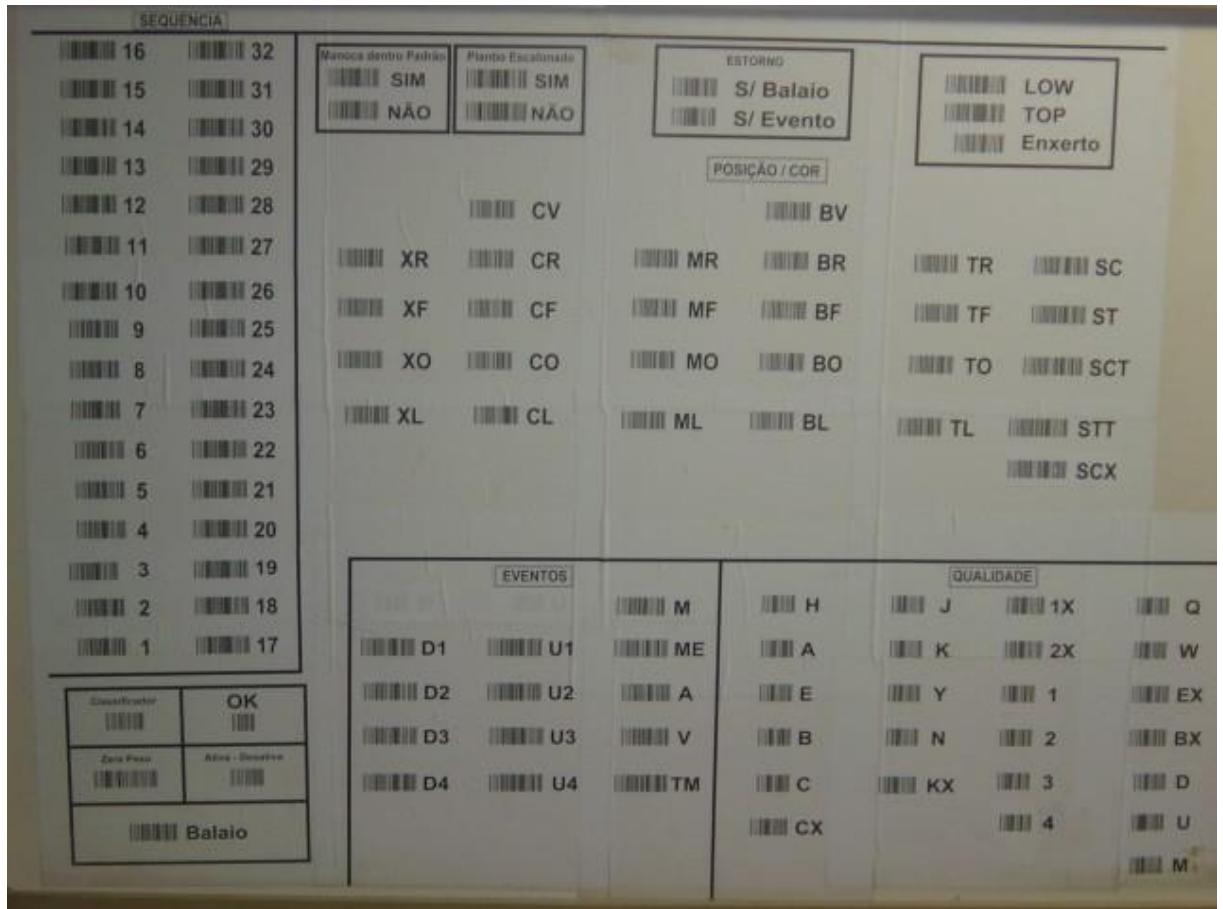


Figura 12 – Painel de compra de fumo  
 Fonte: Fotografia tirada na empresa

Internamente, o classificador de fumo utiliza um padrão diferenciado de pronúncia das palavras (classes “cantadas”), para evitar eventuais erros de entendimento por parte do apontador de dados, pois a pronúncia de alguns fonemas são muito parecidas, como é o caso, por exemplo, da letra P (“pê”) e B (“bê”), e T (“tê”) e D (“dê”).

A tabela 1 apresenta este padrão de pronúncia utilizado pelo classificador para melhorar o entendimento do apontador de dados. É importante que classificadores e apontadores utilizem a mesma linguagem para não ocorrerem erros ou distorções na hora de selecionar a classe no painel de compra.

Tabela 1 – Pronúncia dos fonemas utilizados para formar a classe interna do fumo.

POSIÇÃO	PRONÚNCIA	COR	PRONÚNCIA
P	PRAIM	L	ÉLE
X	XIS	O	Ó
C	CE	F	ÉFE
M	EME	R	RÉ
B	BI	V	VE
T	TU	G	GE

QUALIDADE	PRONÚNCIA
H	AGÁ
A	A
E	E
B	BI
C	CE
1	UM
2	DOIS
J	JÓTA
3	TRÊS
4	QUATRO
Y	YPSILON
K	KA
N	NÉ
Q	QUE
W	DABLIU
X	XIS

EVENTOS	PRONÚNCIA
D	DE
U	ÚMIDO
M	MISTURA
A	ARDIDO
V	VERDE
ME	EME E
AA	AROMA ATÍPICO

### 2.3 Aplicação das Wavelets

Para que fosse possível aplicar as *wavelets*, foi necessário “desmembrar”, ou seja, recortar o arquivo de som, separando individualmente apenas as palavras que se desejava analisar, retirando demais sons que não eram de interesse.

O *software* de áudio utilizado para executar tal tarefa foi o *Cool Edit Pro 2.0*, com o qual foi possível selecionar apenas o intervalo de som desejado (palavra) para analisar.

A figura 13 mostra o arquivo “compra.wav”, que foi obtido através de um microfone ligado a um microcomputador instalado na porta de compra da indústria de fumo, onde foi gravada uma hora ininterrupta de classificação (classificador “cantando” as classes) de fumo durante o processo de compra de fumo da empresa.

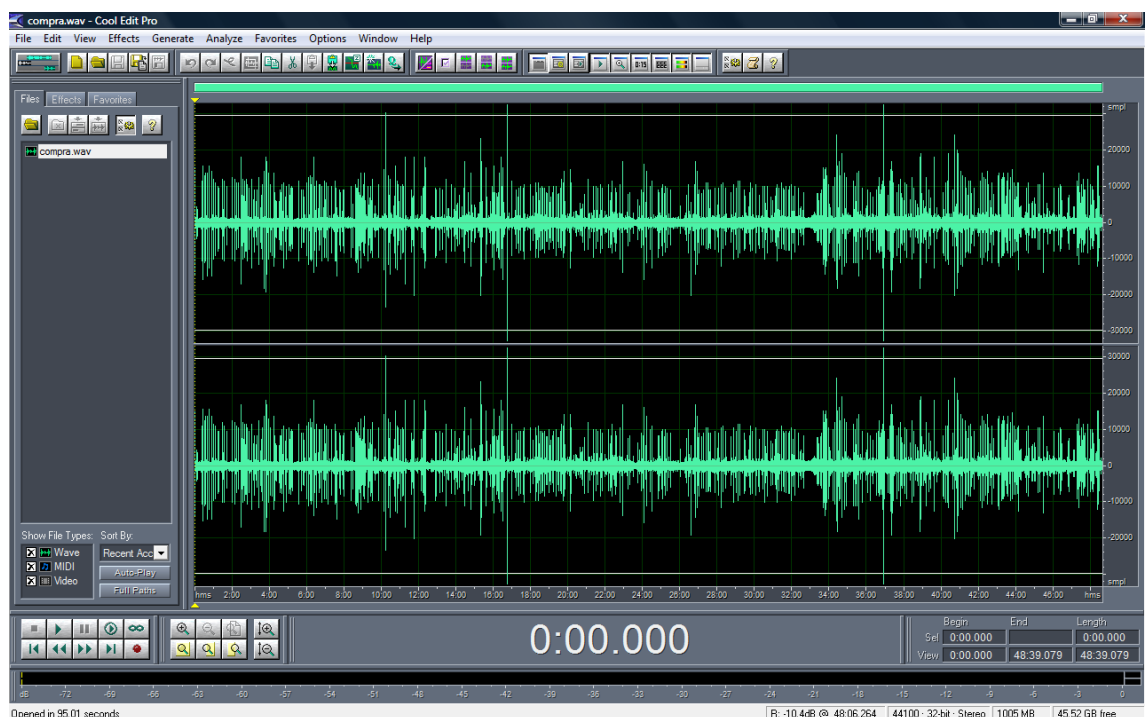


Figura 13 – Arquivo de áudio compra.wav aberto no Cool Edit Pro 2.0.

Através do *software*, é possível fazer um *zoom* na visualização do sinal, ampliá-lo até ser possível diferenciar partes do arquivo de áudio que são constituídas de som, ou seja, o momento que é falada a palavra (maior frequência) das partes onde nada é falado. A figura 14 ilustra este procedimento.

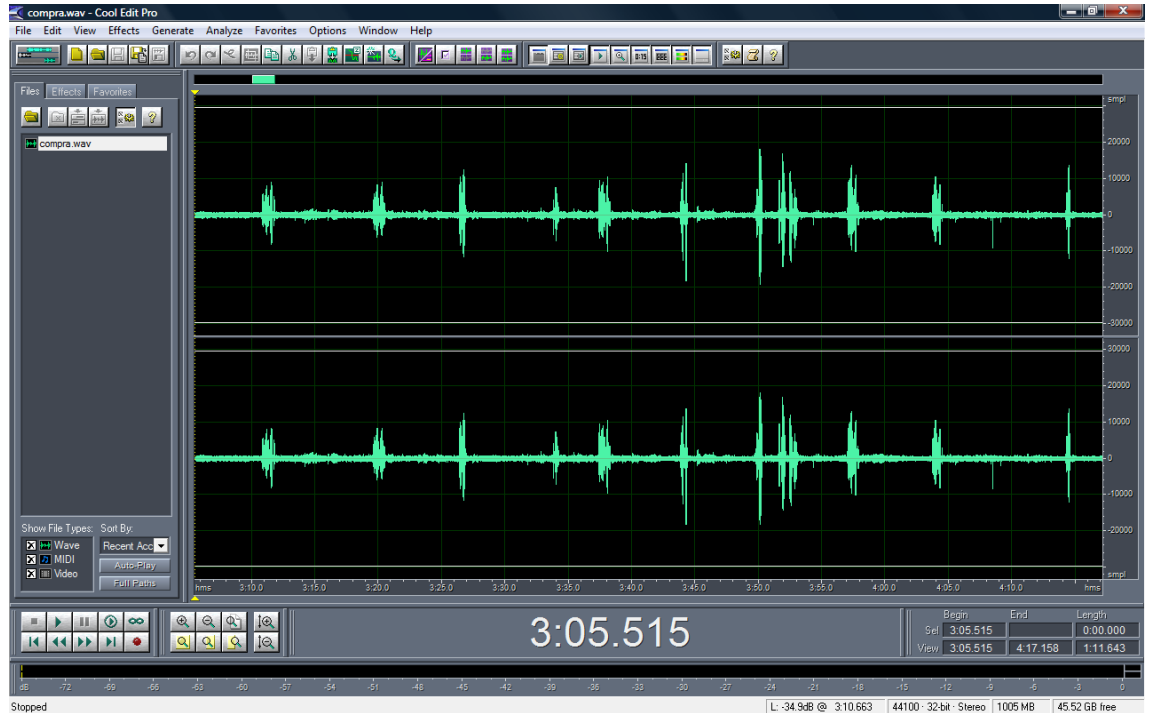


Figura 14 – Ampliação da visualização do sinal de áudio – arquivo compra.wav.

Ampliando o sinal, é possível selecionar apenas o que é desejado para a análise com *wavelets*. A figura 15 apresenta a seleção do sinal de interesse para esta análise. Como exemplo, foi selecionada a palavra (classe) “MLJ” (EME-ÉLE-JÓTA) em apenas um canal (mono) e copiada para uma nova janela – figura 16 – para maior ampliação do sinal e seleção apenas da palavra e não de espaços vazios (silêncio ou ruídos) – figura 17.

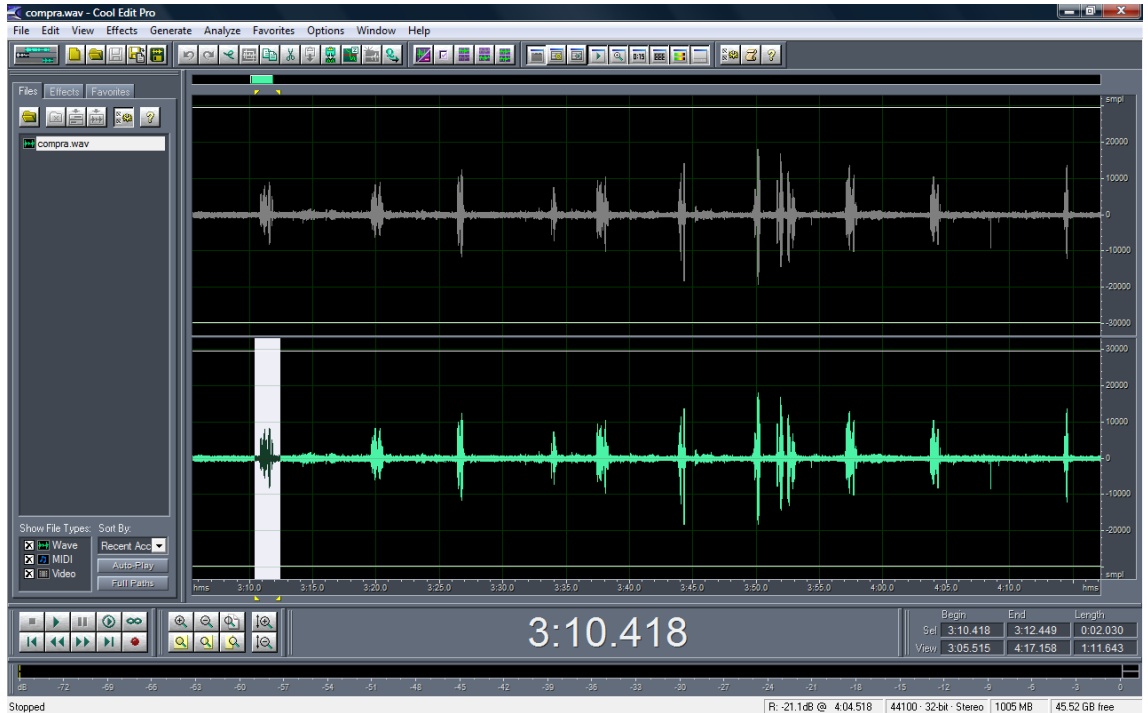


Figura 15 – Seleção da palavra MLJ em apenas um canal (mono).

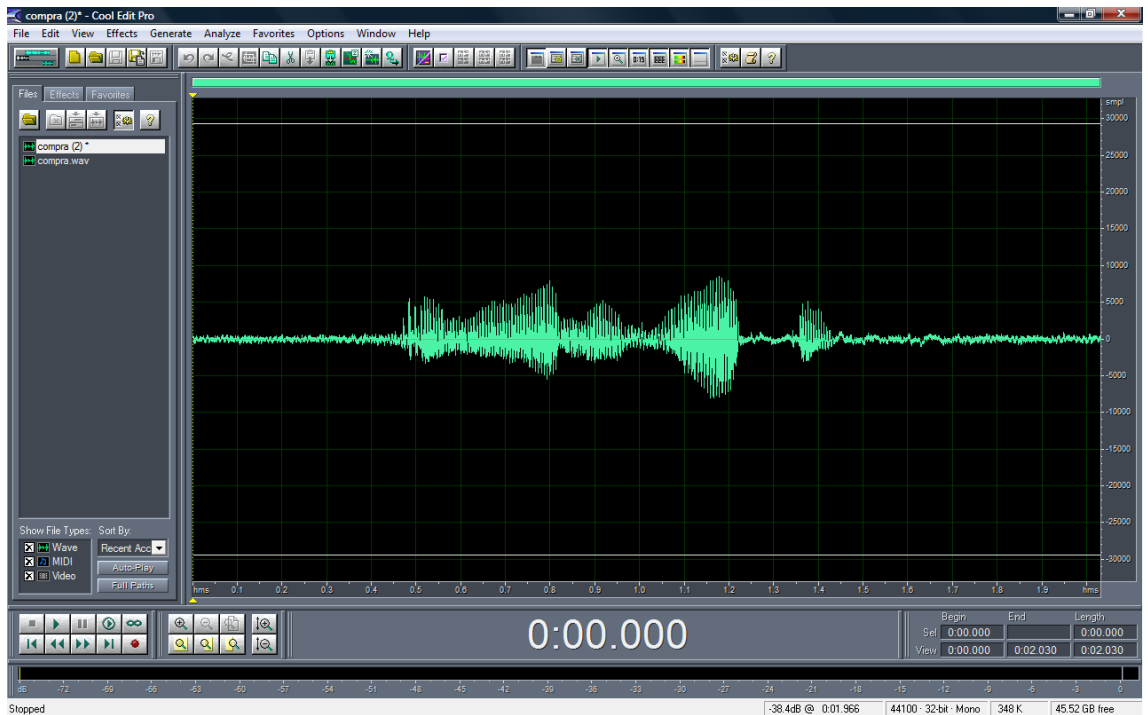


Figura 16 – Palavra MLJ copiada para uma nova janela.

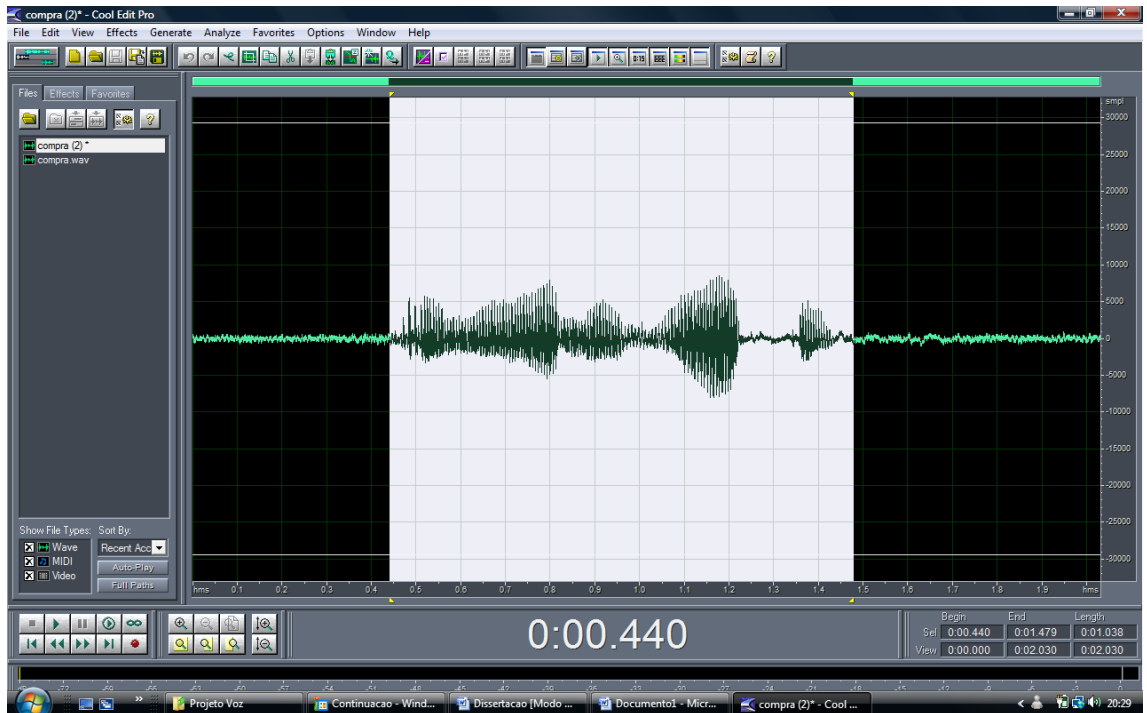


Figura 17 – Seleção apenas da palavra desejada em uma ampliação maior.

Após estes procedimentos, novamente a seleção foi copiada para uma nova janela, conforme mostra a figura 18, e a partir daí foi salvo um novo arquivo de áudio com o nome da palavra adquirida, neste caso “MLJ.wav”.



Figura 18 – Cópia da palavra desejada para salvar arquivo MLJ.wav.



Durante o processo de compra de fumo, muitas classes se repetem. No entanto, foram recortadas todas as palavras faladas, sem exceção, e apenas trocado o nome das palavras que se repetiam, acrescentando um número entre parênteses para as mesmas (exemplo: MLJ(2).wav).

Abaixo é mostrada uma lista de todos os arquivos de áudio adquiridos após a edição do arquivo original compra.wav.

BF1(2).wav	CLC(3).wav	MLJ(8).wav
BF1.wav	CLC(4).wav	MLJ(9).wav
BF2(2).wav	CLC(5).wav	MLJ.wav
BF2(3).wav	CLC(6).wav	MLK(2).wav
BF2(4).wav	CLC(7).wav	MLK.wav
BF2(5).wav	CLC.wav	MO1(2).wav
BF2(6).wav	CLY.wav	MO1(3).wav
BF2.wav	CO1(2).wav	MO1.wav
BF3.wav	CO1(3).wav	MO3(2).wav
BFB(2).wav	CO1(4).wav	MO3(3).wav
BFB(3).wav	CO1(5).wav	MO3(4).wav
BFB(4).wav	CO1(6).wav	MO3.wav
BFB.wav	CO1(7).wav	MOC.wav
BFC(2).wav	CO1.wav	MOJ(2).wav
BFC(3).wav	CO2(2).wav	MOJ.wav
BFC.wav	CO2(3).wav	MOK(2).wav
BFE.wav	CO2.wav	MOK(3).wav
BFJ(2).wav	COB(2).wav	MOK.wav
BFJ(3).wav	COB(3).wav	TF1(2).wav
BFJ(4).wav	COB.wav	TF1(3).wav
BFJ.wav	COM(2).wav	TF1.wav
BFK(2).wav	COM(3).wav	TF2.wav
BFK.wav	COM.wav	TF3.wav
BFM(2).wav	CR2X.wav	TFE(2).wav
BFM(3).wav	D1BF2.wav	TFE.wav
BFM(4).wav	D1TF1.wav	TFJ.wav
BFM(5).wav	D1TF2(2).wav	TL1.wav
BFM.wav	D1TF2.wav	TLJ(2).wav
BFU.wav	D2TR2.wav	TLJ(3).wav
BL3(2).wav	MF1(2).wav	TLJ(4).wav
BL3.wav	MF1(3).wav	TLJ(5).wav
BO2(2).wav	MF1(4).wav	TLJ.wav
BO2.wav	MF1(5).wav	TLK.wav
BO3(2).wav	MF1.wav	TO1(2).wav
BO3.wav	MF2(10).wav	TO1.wav
BOK(2).wav	MF2(11).wav	TO2.wav
BOK(3).wav	MF2(12).wav	TO3.wav
BOK(4).wav	MF2(2).wav	TOJ(2).wav

BOK.wav	MF2(3).wav	TOJ(3).wav
BOM(2).wav	MF2(4).wav	TOJ(4).wav
BOM(3).wav	MF2(5).wav	TOJ(5).wav
BOM.wav	MF2(6).wav	TOJ.wav
BR1.wav	MF2(7).wav	TOK(10).wav
BRE.wav	MF2(8).wav	TOK(11).wav
BV3(2).wav	MF2(9).wav	TOK(12).wav
BV3.wav	MF2.wav	TOK(2).wav
CF1(2).wav	MF3.wav	TOK(3).wav
CF1(3).wav	MFB(2).wav	TOK(4).wav
CF1(4).wav	MFB(3).wav	TOK(5).wav
CF1(5).wav	MFB(4).wav	TOK(6).wav
CF1(6).wav	MFB(5).wav	TOK(7).wav
CF1.wav	MFB.wav	TOK(8).wav
CF2.wav	ML1.wav	TOK(9).wav
CFB(2).wav	ML2(2).wav	TOK(quick).wav
CFB(3).wav	ML2(3).wav	TOK.wav
CFB(4).wav	ML2(4).wav	TRN.wav
CFB.wav	ML2.wav	U1TOJ.wav
CFC(2).wav	MLC.wav	XFB.wav
CFC(3).wav	MLJ(10).wav	XFY.wav
CFC(4).wav	MLJ(11).wav	XL1.wav
CFC.wav	MLJ(12).wav	XL2.wav
CFE.wav	MLJ(13).wav	XO1(2).wav
CFM.wav	MLJ(2).wav	XO1(3).wav
CFU.wav	MLJ(3).wav	XO1(4).wav
CL2(2).wav	MLJ(4).wav	XO1.wav
CL2(3).wav	MLJ(5).wav	XO2.wav
CL2.wav	MLJ(6).wav	XO3(2).wav
CLC(2).wav	MLJ(7).wav	XO3.wav
		XR2X.wav

De acordo com os arquivos de som adquiridos após a edição do arquivo original, foram elencadas algumas classes (palavras) que serão utilizadas para teste e aplicação das *wavelets* nesta dissertação. O critério adotado para a escolha das palavras utilizadas na realização dos testes foi a repetição. Como muitas delas se repetiram, foram escolhidas aquelas que apareceram mais de uma vez no arquivo original “compra.wav”. A tabela 2 apresenta as classes selecionadas para os testes.

Tabela 2 – Classes de fumo selecionadas para testes e suas respectivas pronúncias.

<b>CLASSE</b>	<b>PRONÚNCIA</b>
BF1	BI-ÉFE-UM
BF2	BI-ÉFE-DOIS
BFB	BI-ÉFE-BI
BFC	BI-ÉFE-CE
BFJ	BI-ÉFE-JÓTA
BFK	BI-ÉFE-KA
BFM	BI-ÉFE-EME
BL3	BI-ÉLE-TRÊS
BO2	BI-Ó-DOIS
BO3	BI-Ó-TRÊS
BOK	BI-Ó-KA
BOM	BI-Ó-EME
BV3	BI-VÊ-TRÊS
CF1	CE-ÉFE-UM
CFB	CE-ÉFE-BI
CFC	CE-ÉFE-CE
CL2	CE-ÉLE-DOIS
CLC	CE-ÉLE-CE
CO1	CE-Ó-UM
CO2	CE-Ó-DOIS
COB	CE-Ó-BI
COM	CE-Ó-EME
D1TF2	DE-UM-TU-ÉFE-DOIS
MF1	EME-ÉFE-UM
MF2	EME-ÉFE-DOIS
MFB	EME-ÉFE-BI
ML2	EME-ÉLE-DOIS
MLJ	EME-ÉLE-JÓTA
MLK	EME-ÉLE-KA
MO1	EME-Ó-UM
MO3	EME-Ó-TRÊS
MOJ	EME-Ó-JÓTA
MOK	EME-Ó-KA
TF1	TU-ÉFE-UM
TFE	TU-ÉFE-E
TLJ	TU-ÉLE-JÓTA
TO1	TU-Ó-UM
TOJ	TU-Ó-JÓTA
TOK	TU-Ó-KA
XO1	XIS-Ó-UM
XO3	XIS-Ó-TRÊS

Para aplicar as transformadas *Wavelets* foi utilizado o software MATLAB® versão 7.0.0.19920(R14), utilizando como sinal original cada um dos arquivos de som selecionados para teste.

### 2.3.1. Aplicação da Transformada Haar

Primeiramente foi utilizada a transformada discreta Haar até seu terceiro nível, utilizando como sinal original um dos arquivos adquiridos após “desmembrar” o sinal original “compra.wav”. Após o arquivo ser carregado no MATLAB, foram definidos os subsinais  $a_1$ ,  $a_2$ ,  $a_3$ ,  $d_1$ ,  $d_2$ ,  $d_3$ .

#### Análise Haar do sinal – Arquivo BF1.wav

Abaixo é apresentado o código que foi utilizado no MATLAB para aplicar a compactação do sinal:

```
>> load c:\BF1.dat
>>Signal = BF1(:,1);
>>N = length(Signal);
>>a1 = zeros(1,N/2); d1 = a1;
>>a2 = zeros(1,N/4); d2 = a2;
>>a3 = zeros(1,N/8); d3 = a3;

>>for k = 1:N/2,
a1(k) = 1/sqrt(2)*(Signal(2*k-1) + Signal(2*k));
d1(k) = 1/sqrt(2)*(Signal(2*k-1) - Signal(2*k));
end;

>>for k = 1:N/4,
a2(k) = 1/sqrt(2)*(a1(2*k-1) + a1(2*k));
d2(k) = 1/sqrt(2)*(a1(2*k-1) - a1(2*k));
end;
```

```
>> for k = 1:N/8,
a3(k) = 1/sqrt(2)*(a2(2*k-1) + a2(2*k));
d3(k) = 1/sqrt(2)*(a2(2*k-1) - a2(2*k));
end;
```

Para apresentar os gráficos dos sinais após a compactação, foram definidas matrizes h1 (primeira compactação), h2 (segunda compactação) e h3 (terceira compactação), conforme o código abaixo:

```
>> h1 = [ a1 d1 ];
>> h2 = [ a2 d2 d1 ];
>> h3 = [ a3 d3 d2 d1 ];
```

Através do comando *plot*, no MATLAB, foi possível observar cada uma das compactações realizadas, bem como o sinal original. A figura 19 apresenta gráfico do sinal original da palavra BF1.

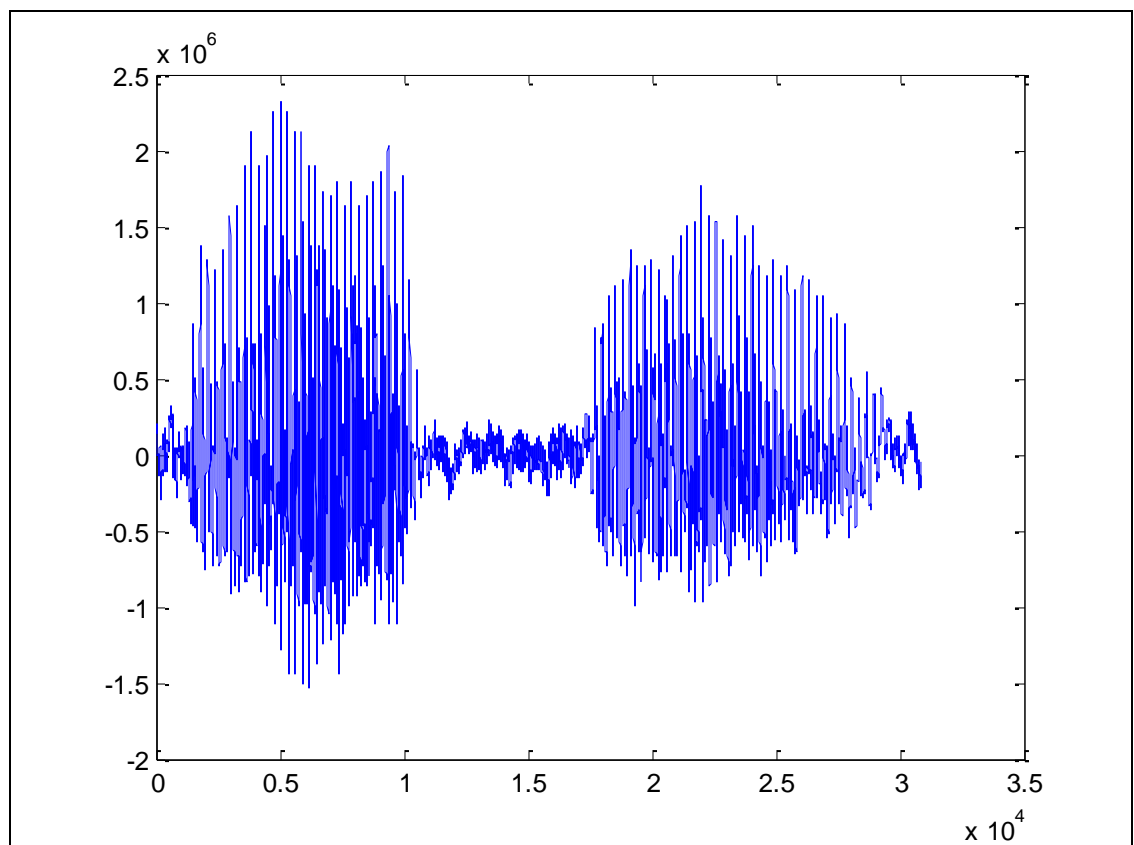


Figura 19 – Gráfico do sinal original da palavra BF1.

A figura 20 apresenta o gráfico do sinal da primeira compactação da palavra BF1 utilizando a transformada Haar.

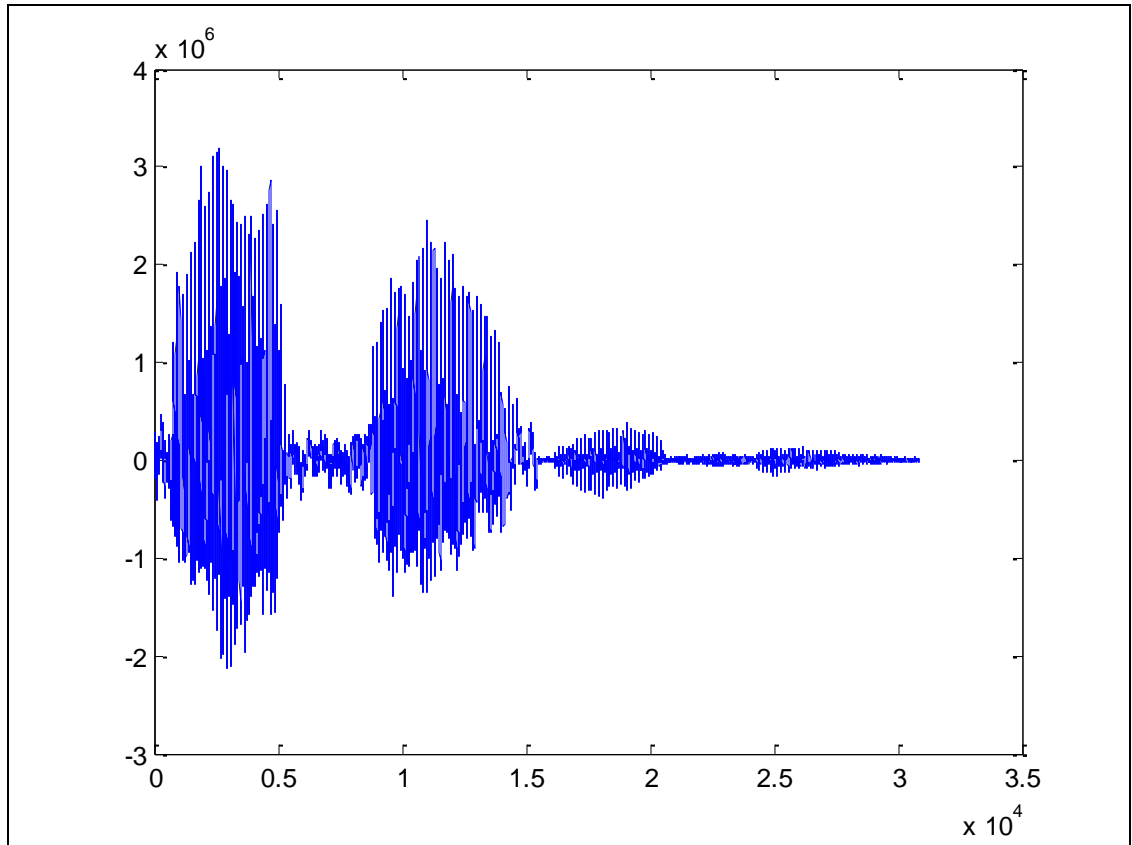


Figura 20 – Gráfico do sinal da primeira compactação da palavra BF1 (Haar)

A figura 21 apresenta o gráfico do sinal da segunda compactação da palavra BF1 utilizando a transformada Haar.

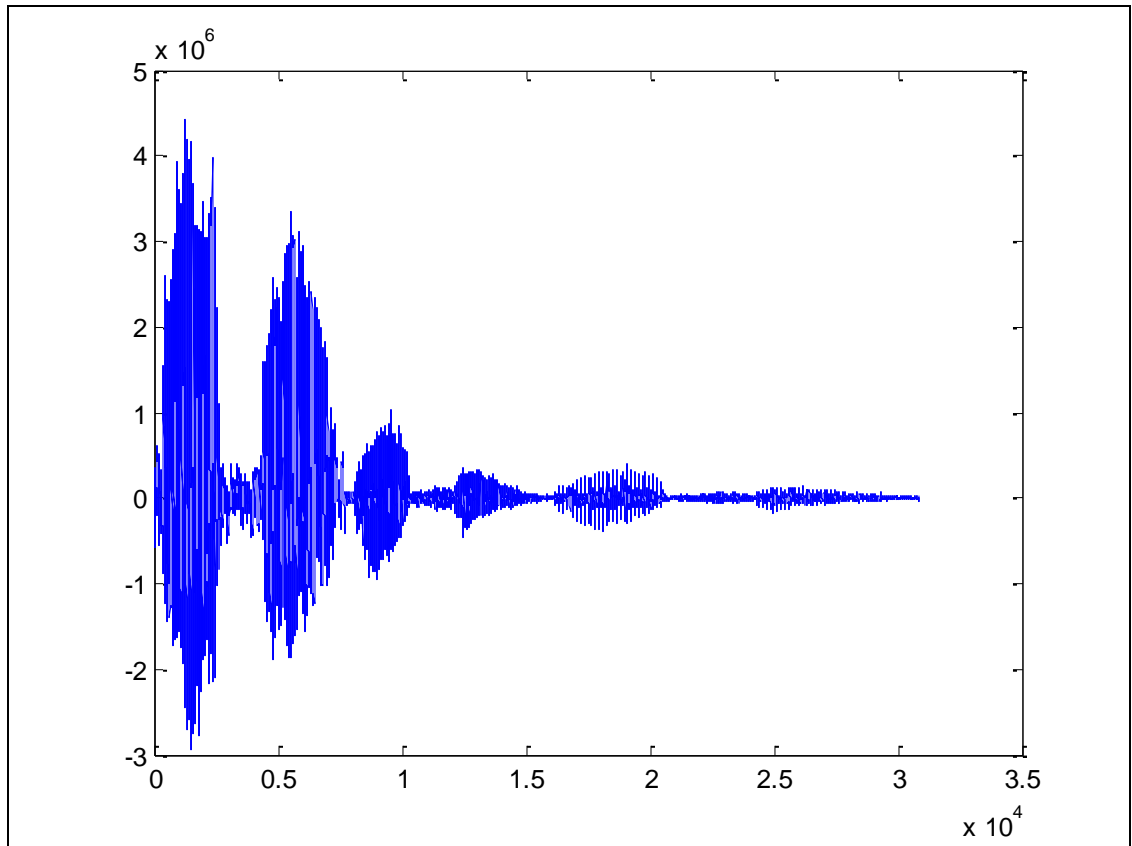


Figura 21 – Gráfico do sinal da segunda compactação da palavra BF1 (Haar)

A figura 22 apresenta o gráfico do sinal da terceira compactação da palavra BF1 utilizando a transformada Haar.

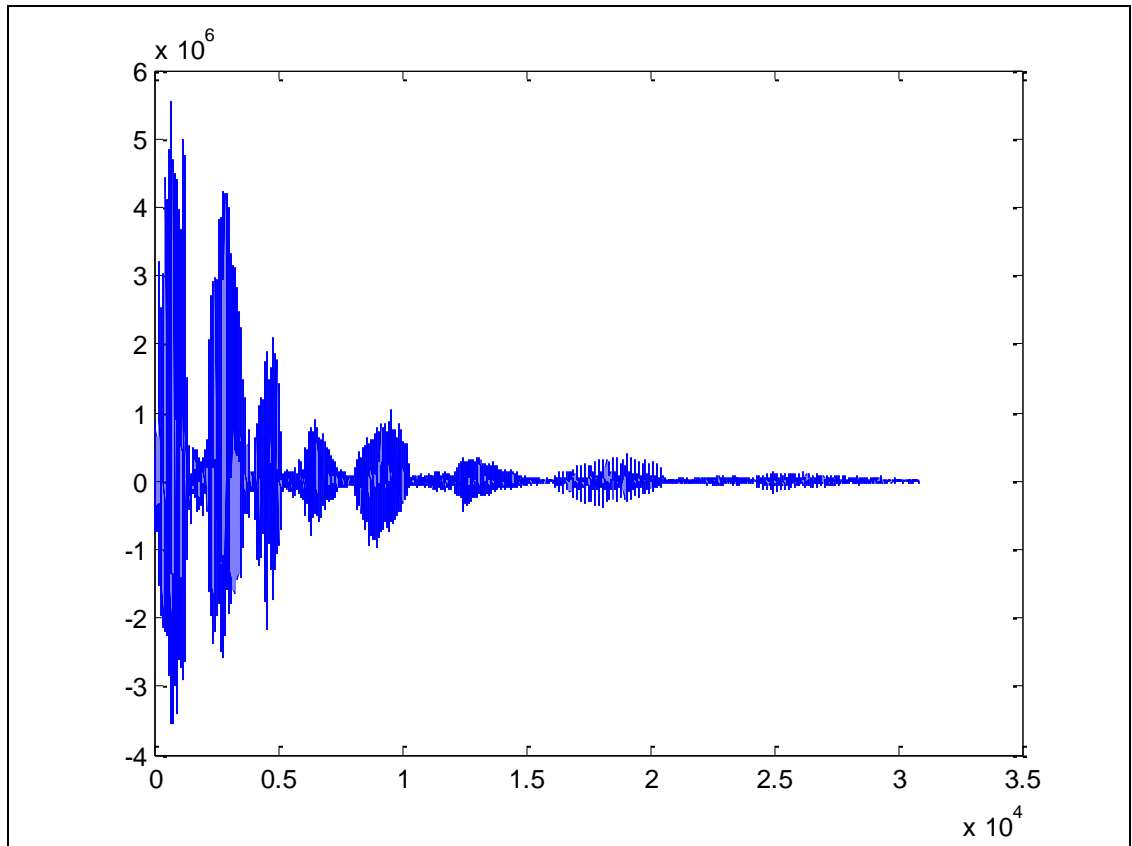


Figura 22 – Gráfico do sinal da terceira compactação da palavra BF1 (Haar)

### 2.3.2. Aplicação da Transformada Daub4

Na transformada Daub4 também foram executadas compactações até o terceiro nível, também utilizando como sinal original um dos arquivos adquiridos após “desmembrar” o sinal original “compra.wav”. Após o arquivo ser carregado no MATLAB, foram definidos os subsinais  $a_1$ ,  $a_2$ ,  $a_3$ ,  $d_1$ ,  $d_2$ ,  $d_3$ .

#### Análise Daub4 do sinal – Arquivo BF1.wav

Abaixo é mostrado o código que foi utilizado no MATLAB para aplicar a compactação do sinal:

```
>> load c:\BF1.dat;
>> s = BF1;
```



```

>> g = [(1+sqrt(3))/(4*sqrt(2)),(3+sqrt(3))/(4*sqrt(2)),(3-sqrt(3))/(4*sqrt(2)),(1-
sqrt(3))/(4*sqrt(2))];
>> h = [(1-sqrt(3))/(4*sqrt(2)),(sqrt(3)-3)/(4*sqrt(2)),(3+sqrt(3))/(4*sqrt(2)),(-1-
sqrt(3))/(4*sqrt(2))];
>> n = length(s);
>> LA1=g(3)*s(1)+g(4)*s(2)+g(1)*s(n-1)+g(2)*s(n);
>> LD1=h(3)*s(1)+h(4)*s(2)+h(1)*s(n-1)+h(2)*s(n);

>> for j = 1:n/2-1,
    A1(j) = g(1)*s(2*j-1)+g(2)*s(2*j)+g(3)*s(2*j+1)+g(4)*s(2*j+2);
    D1(j) = h(1)*s(2*j-1)+h(2)*s(2*j)+h(3)*s(2*j+1)+h(4)*s(2*j+2);
end

>> LA2=g(3)*TD1(1)+g(4)*TD1(2)+g(1)*TD1(n/2-1)+g(2)*TD1(n/2);
>> LD2=h(3)*TD1(1)+h(4)*TD1(2)+h(1)*TD1(n/2-1)+h(2)*TD1(n/2);

>> for j = 1:n/4-1,
    A2(j) = g(1)*TD1(2*j-1)+g(2)*TD1(2*j)+g(3)*TD1(2*j+1)+g(4)*TD1(2*j+2);
    D2(j) = h(1)*TD1(2*j-1)+h(2)*TD1(2*j)+h(3)*TD1(2*j+1)+h(4)*TD1(2*j+2);
end

>> LA3=g(3)*TD2(1)+g(4)*TD2(2)+g(1)*TD2(n/4-1)+g(2)*TD2(n/4);
>> LD3=h(3)*TD2(1)+h(4)*TD2(2)+h(1)*TD2(n/4-1)+h(2)*TD2(n/4);

>> for j = 1:n/8-1,
    A3(j) = g(1)*TD2(2*j-1)+g(2)*TD2(2*j)+g(3)*TD2(2*j+1)+g(4)*TD2(2*j+2);
    D3(j) = h(1)*TD2(2*j-1)+h(2)*TD2(2*j)+h(3)*TD2(2*j+1)+h(4)*TD2(2*j+2);
end

```

Para apresentar os gráficos dos sinais após a compactação, foram definidas matrizes TD1 (primeira compactação), TD2 (segunda compactação) e TD3 (terceira compactação), conforme o código abaixo:

```
>> TD1 = [A1 LA1 D1 LD1];  
>> TD2 = [A2 LA2 D2 LD2 D1 LD1];  
>> TD3 = [A3 LA3 D3 LD3 D2 LD2 D1 LD1];
```

Através do comando *plot*, no MATLAB, foi possível observar cada uma das compactações realizadas, bem como o sinal original. A figura 23 apresenta gráfico do sinal original da palavra BF1.

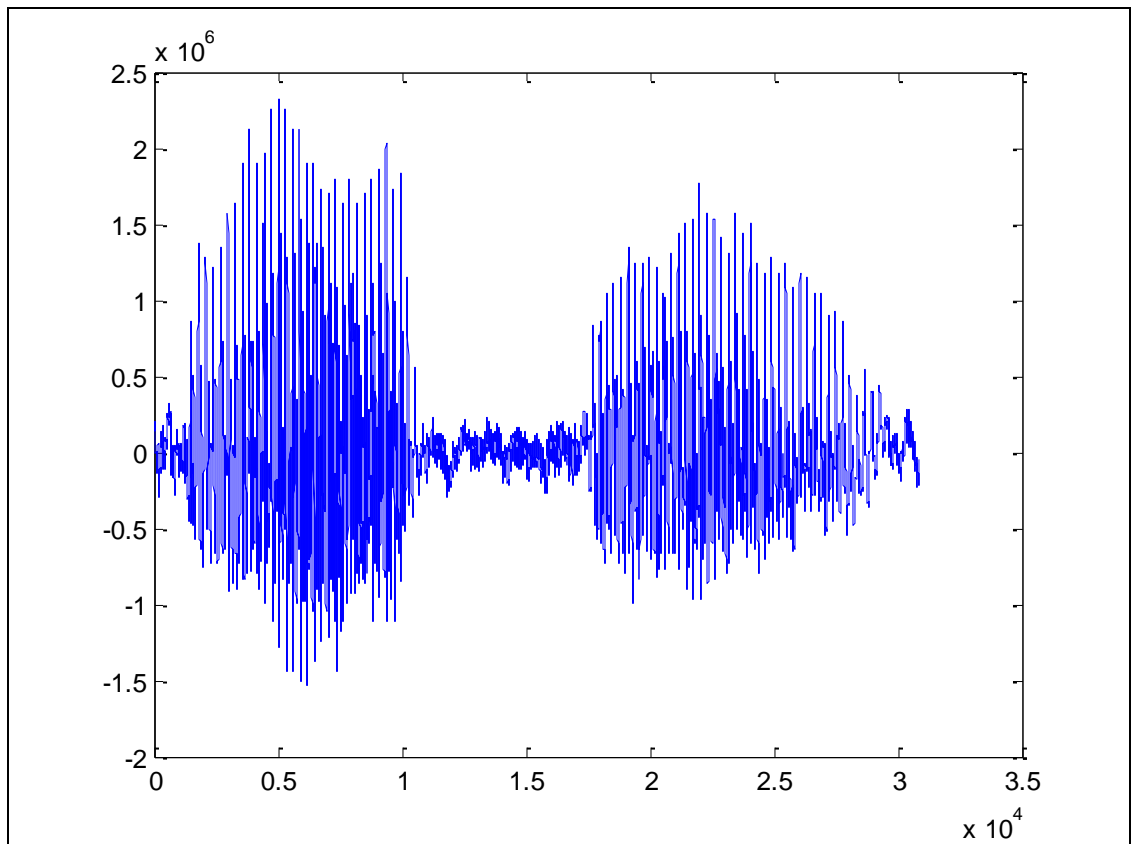


Figura 23 – Gráfico do sinal original da palavra BF1.

A figura 24 apresenta o gráfico do sinal da primeira compactação da palavra BF1 utilizando a transformada Daub4.

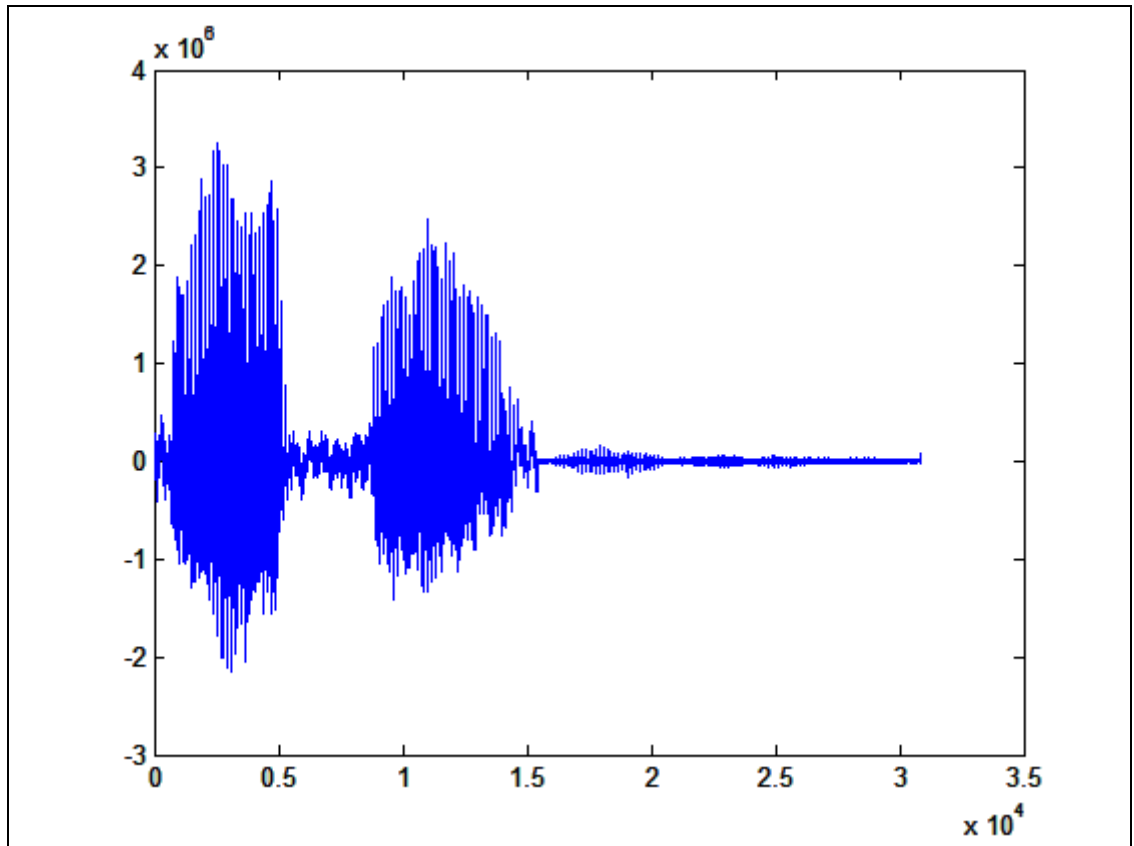


Figura 24 – Gráfico do sinal da primeira compactação da palavra BF1 (Daub4)

A figura 25 apresenta o gráfico do sinal da segunda compactação da palavra BF1 utilizando a transformada Daub4.

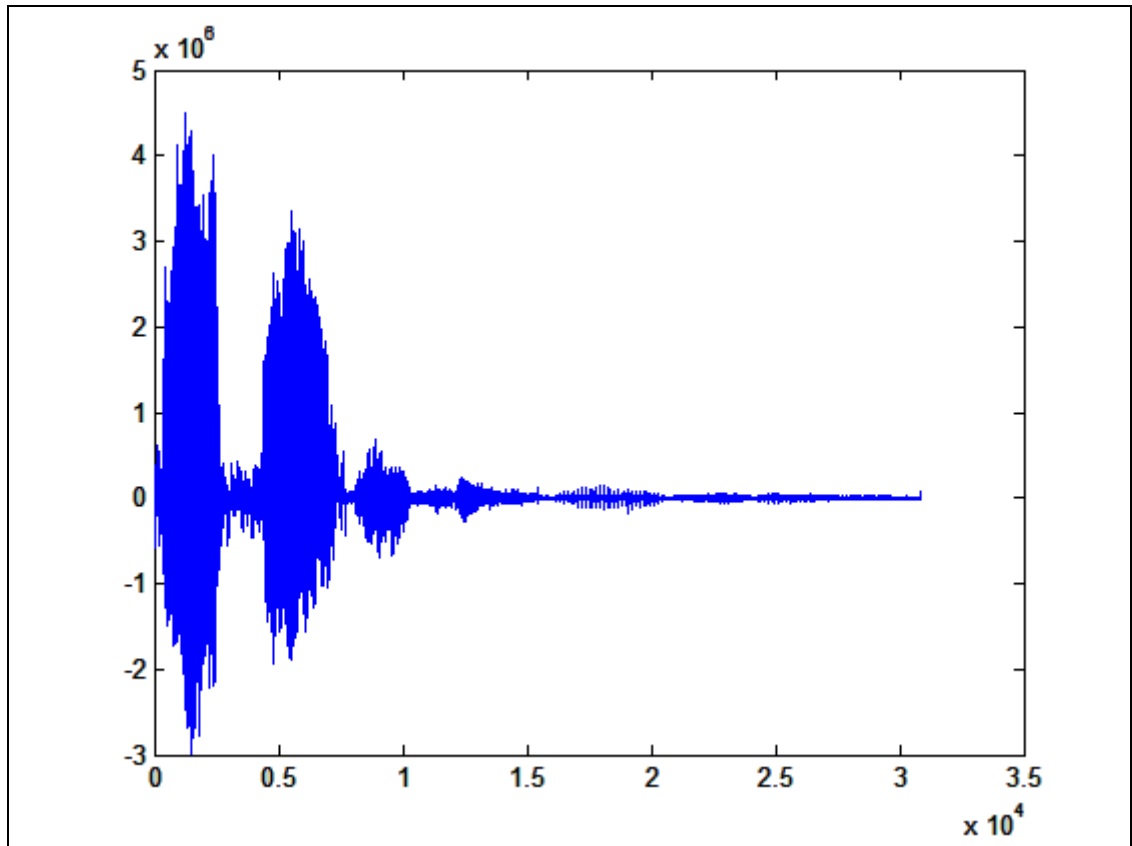


Figura 25 – Gráfico do sinal da segunda compactação da palavra BF1 (Daub4)

A figura 26 apresenta o gráfico do sinal da terceira compactação da palavra BF1 utilizando a transformada Haar.

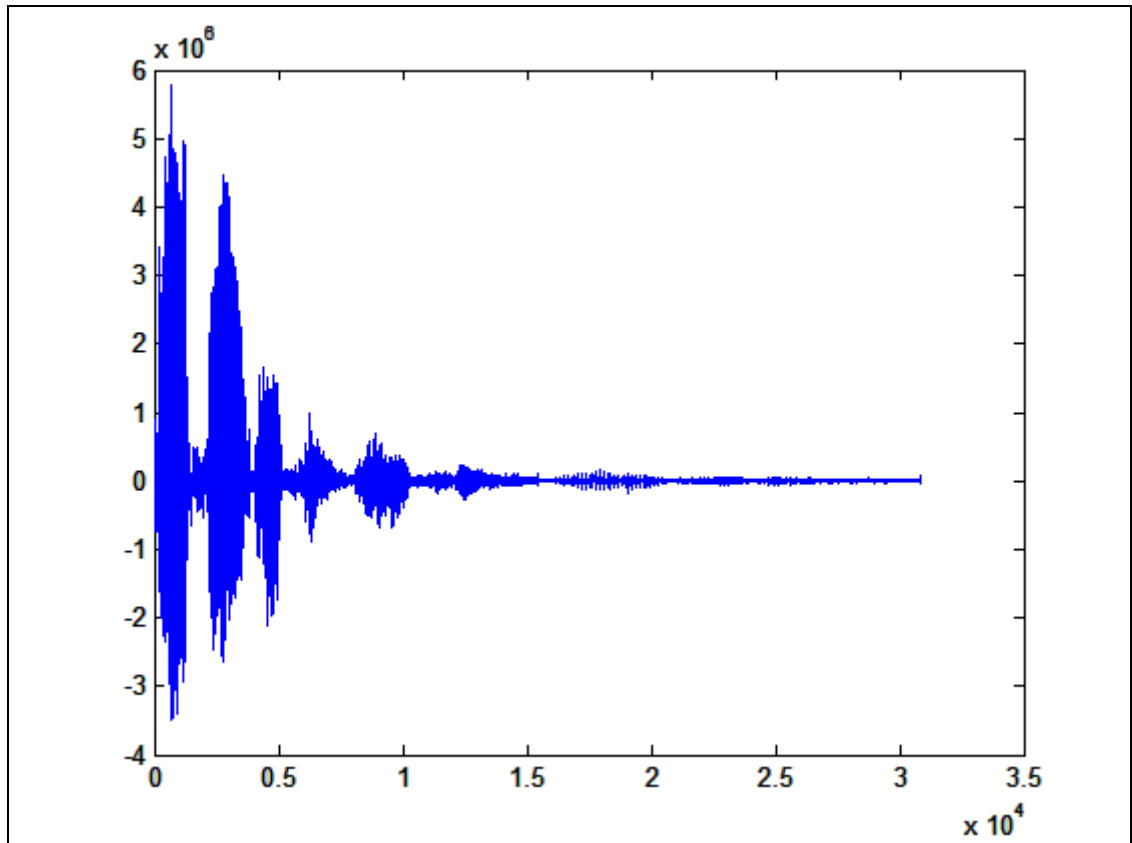


Figura 26 – Gráfico do sinal da terceira compactação da palavra BF1 (Daub4)

### 2.3.3. Análise, Classificação e Reconhecimento dos Comandos de Voz

Para que fosse possível classificar um comando de voz, diversos testes tiveram que ser executados. Primeiramente, partiu-se da hipótese de que após aplicar a terceira transformada *wavelet*, tanto Haar quanto Daub4, bastaria subtrair o sinal resultante da terceira transformada da palavra analisada com o sinal resultante da terceira transformada da palavra original. Dessa forma, a palavra comparada que mais se aproximasse no gráfico do eixo X, ou seja, de zero em seu vetor, seria a palavra reconhecida.

O primeiro impasse se deu no tamanho dos vetores. Cada palavra variava seu tamanho de vetor e para que fosse possível subtrair vetores no Matlab, esses deveriam ser do mesmo tamanho. Para resolver o problema, foi necessário igualar o tamanho de todos eles. A estratégia utilizada foi analisar todas as palavras

digitalizadas, selecionar a de maior tamanho e utilizar este tamanho como padrão, ou seja, para palavras com tamanho de vetor inferior, foram adicionados zeros no restante do vetor (após aplicada a transformada *wavelet*).

O tamanho padrão do vetor para que todas sejam equivalentes foi de 92937 pontos. A Figura 27 apresenta o gráfico do sinal da palavra BF2 (terceira transformada Haar) de tamanho 38747 pontos e a figura 28 mostra o gráfico do sinal da palavra BF2 (terceira transformada Haar) com zeros no final para igualar os valores de todas as palavras analisadas e permitir a subtração.

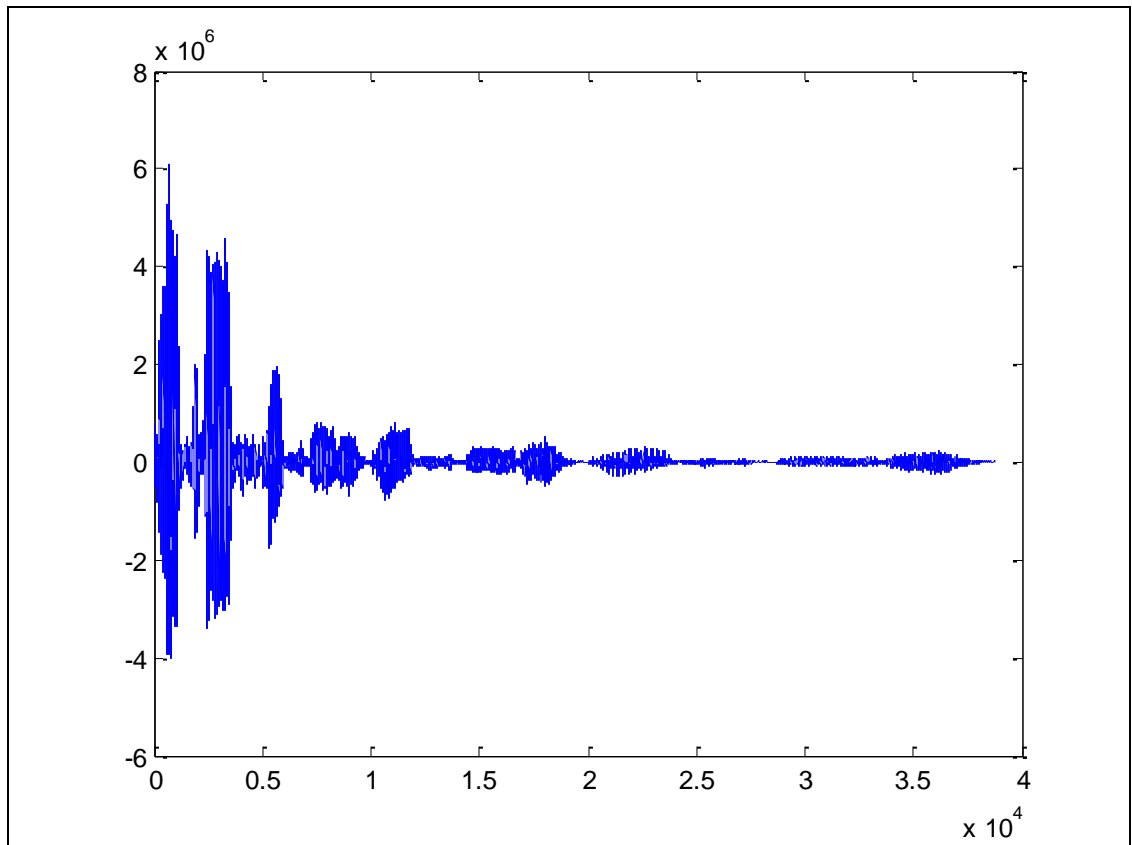


Figura 27 – Gráfico do sinal BF2 de tamanho 38747 pontos

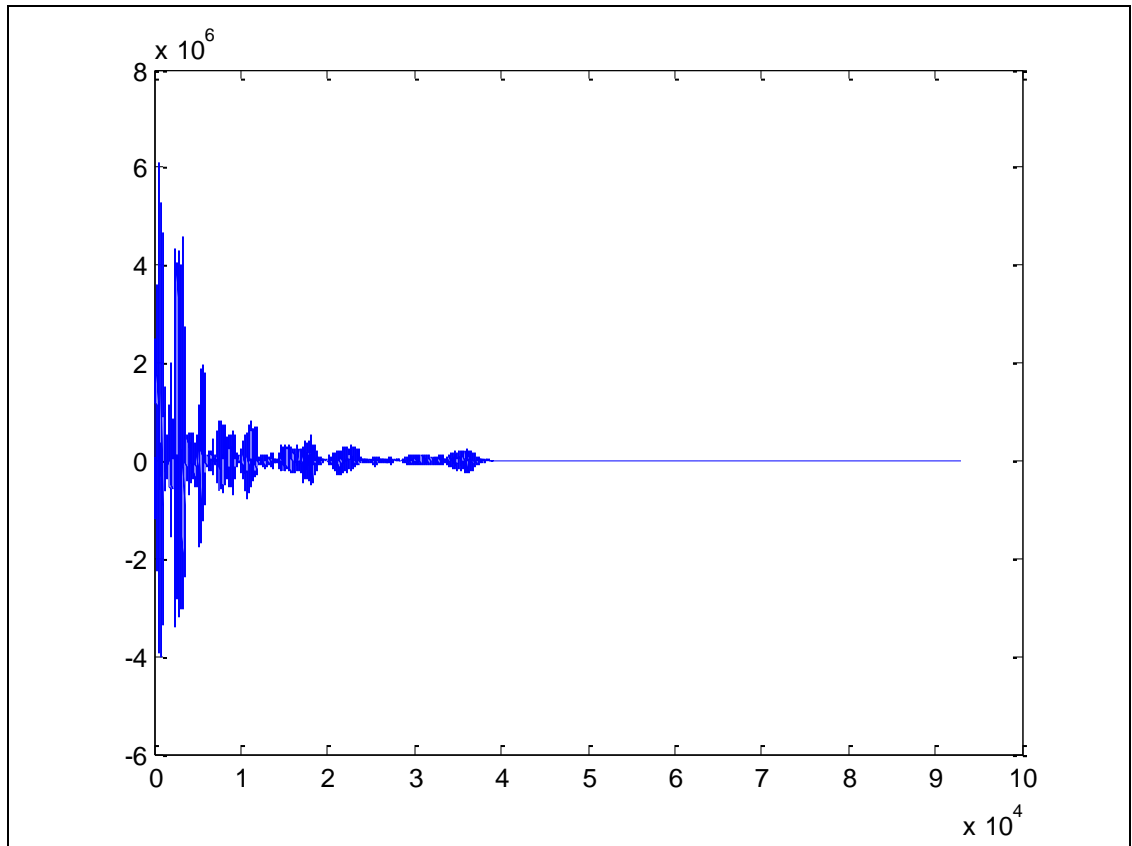


Figura 28 – Gráfico do sinal BF2 de tamanho 92937 pontos (zeros adicionados ao final)

Após os vetores das palavras analisadas estarem de mesmo tamanho, surgiu um segundo impasse: mesmo subtraindo um vetor pelo outro, não foi possível chegar a alguma conclusão definitiva, pois a maior dificuldade nesta fase dos testes foi de alinhar as palavras em relação ao tempo x frequência, ou seja, realizar um pré-processamento do sinal, onde foi necessário recortar o som da gravação original. Conforme apresentado nas figuras abaixo, este desalinhamento resulta em um aumento de determinados valores do vetor da palavra analisada, ao invés da aproximação de zero desejada.

A figura 29 mostra a palavra BF2, já na terceira transformada Haar, a figura 30 mostra a BF2-2, também já na terceira transformada Haar e por fim a figura 31 apresenta a subtração de um vetor pelo outro.

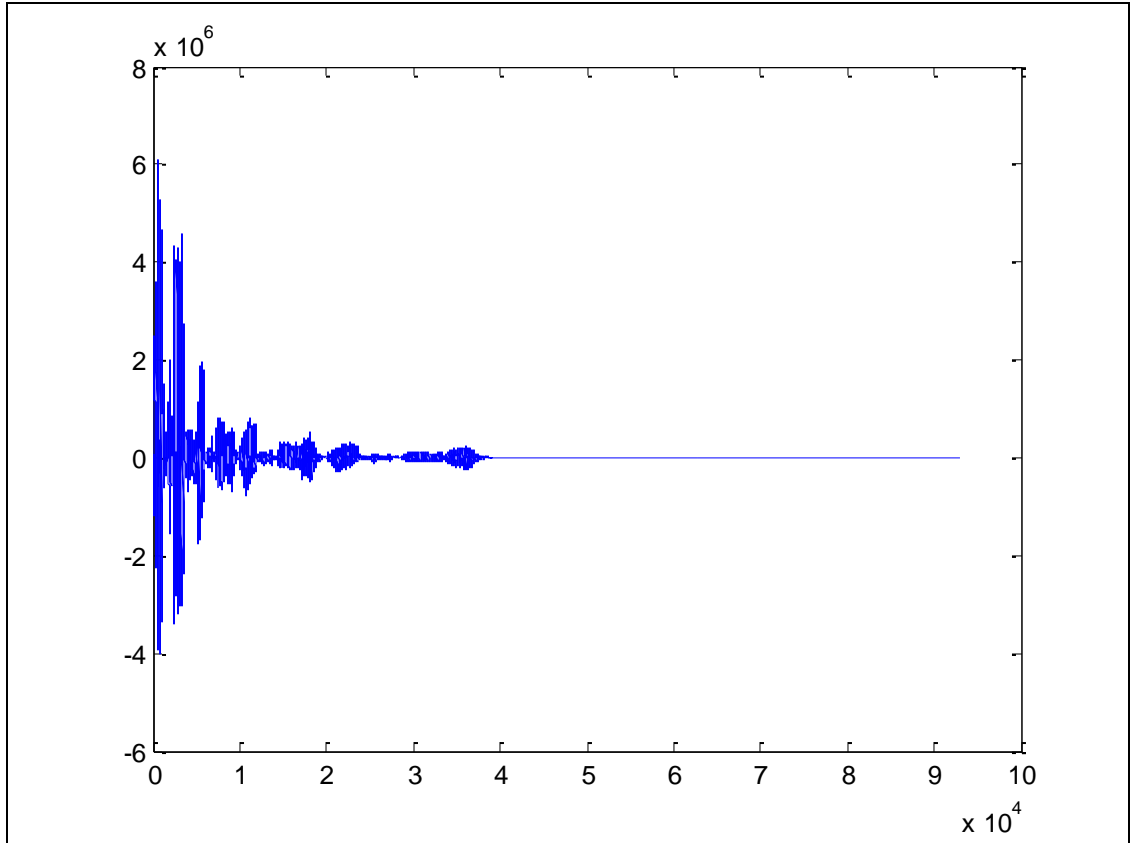


Figura 29 – Grafico da palavra BF2 (terceira transformada Haar)

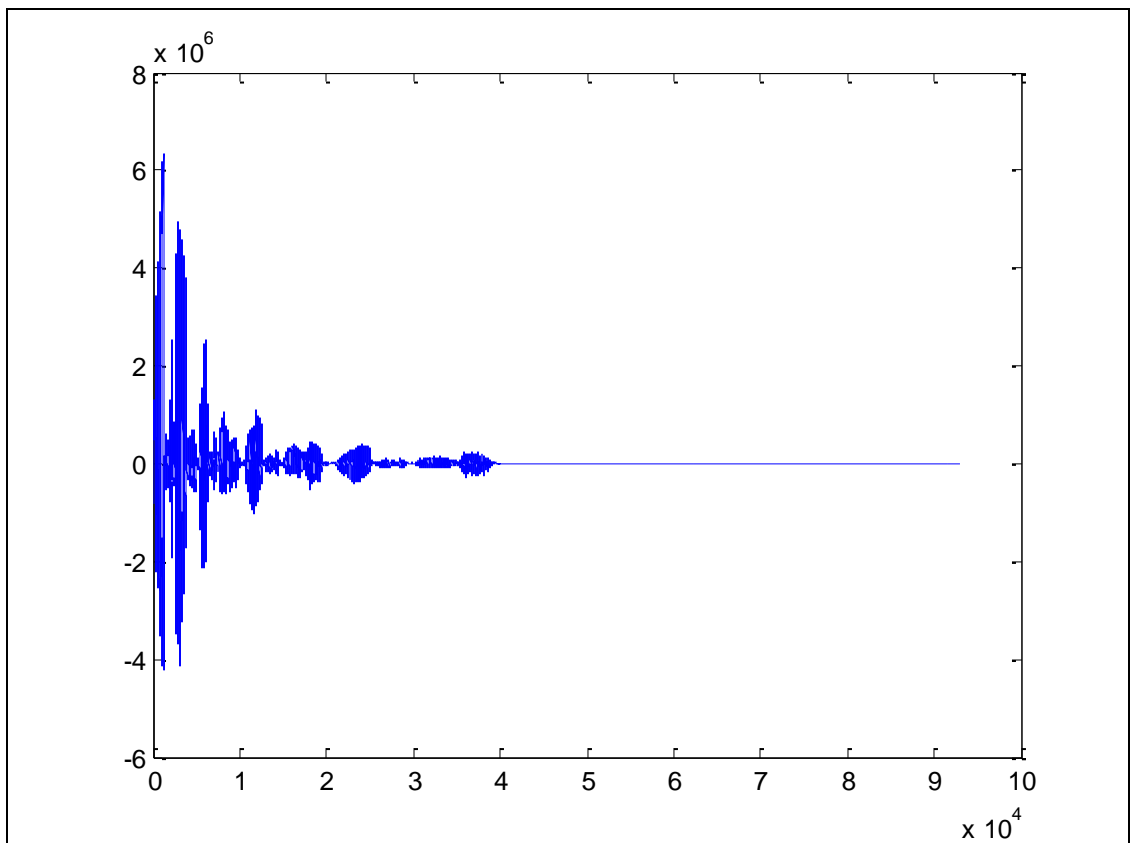


Figura 30 – Grafico da palavra BF2-2 (terceira transformada Haar)



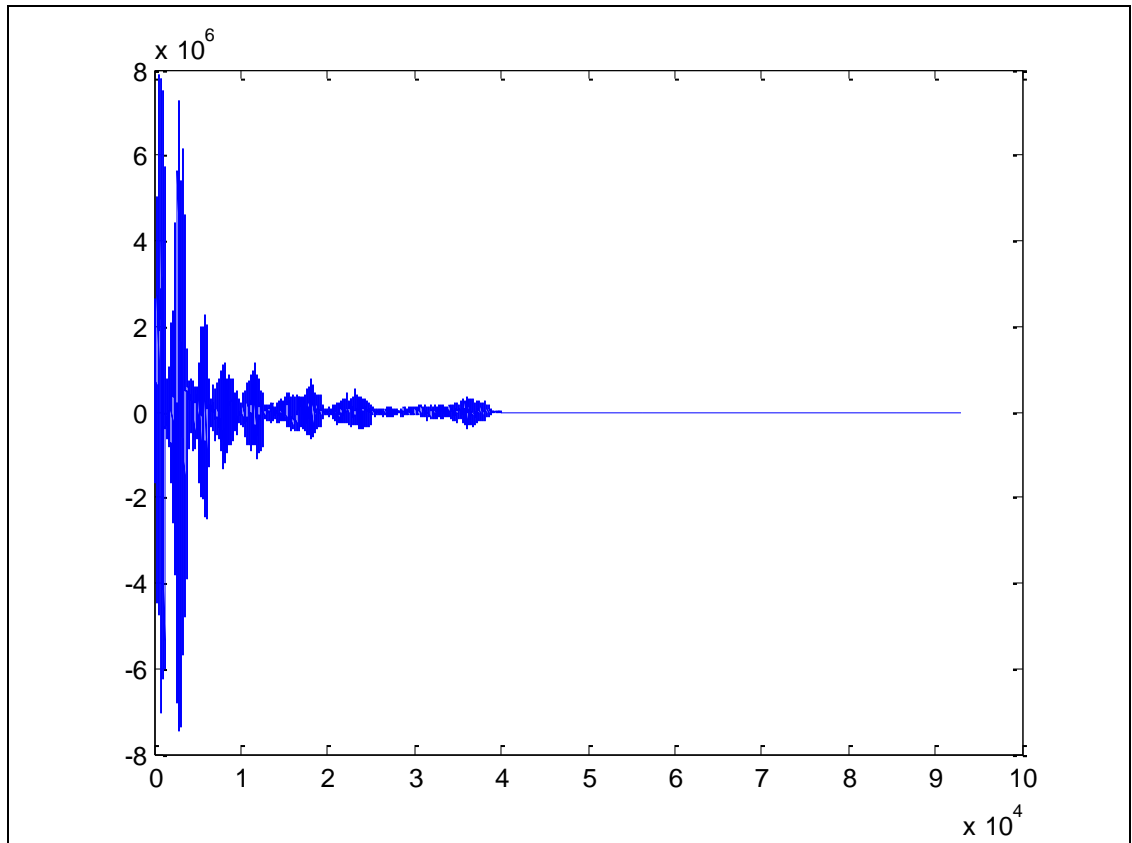


Figura 31 – Subtração das palavras BF2 e BF2-2

Uma solução proposta foi de desenvolver um algoritmo que buscasse os valores máximos dos vetores analisados para que fosse possível fazer uma translação do sinal para alinhar tempo x frequência do vetor analisado com o vetor “dicionário” – este vetor “dicionário” foi desenvolvido consolidando em uma matriz de 41 elementos, onde existe uma linha para cada som analisado e compactado até a terceira transformada Haar. A idéia foi buscar o primeiro valor máximo do vetor e desconsiderar todos os valores anteriores ao valor máximo. Para manter o vetor com seu tamanho padrão, foram adicionados zeros no final do vetor. Dessa forma, foi realizada uma translação do vetor, sendo que seu primeiro valor foi seu primeiro máximo (maior valor do vetor).

A figura 32 apresenta o gráfico da palavra BF2 com seu valor máximo já na posição inicial.

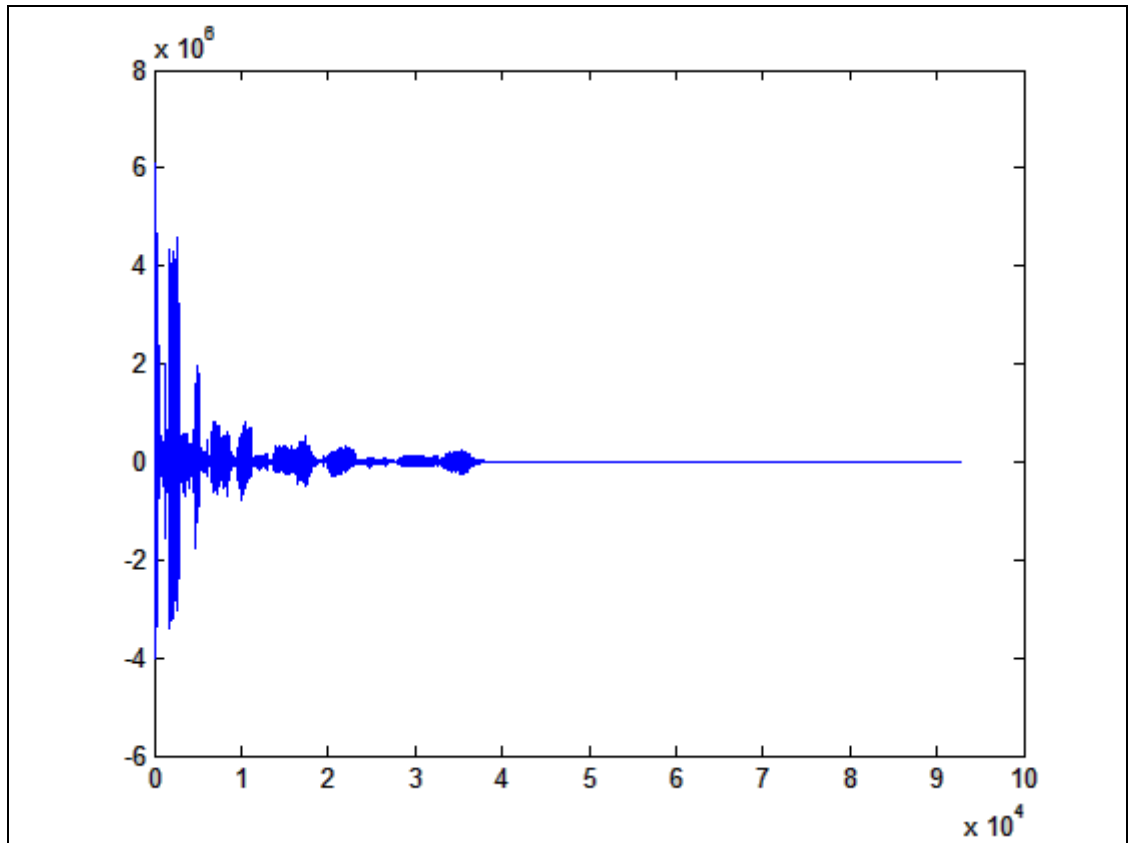


Figura 32 - Gráfico da palavra BF2 (com valor máximo no início)

O mesmo procedimento foi realizado para a palavra BF2-2 (a mesma palavra BF2 falada novamente). A figura 33 mostra o gráfico da palavra.

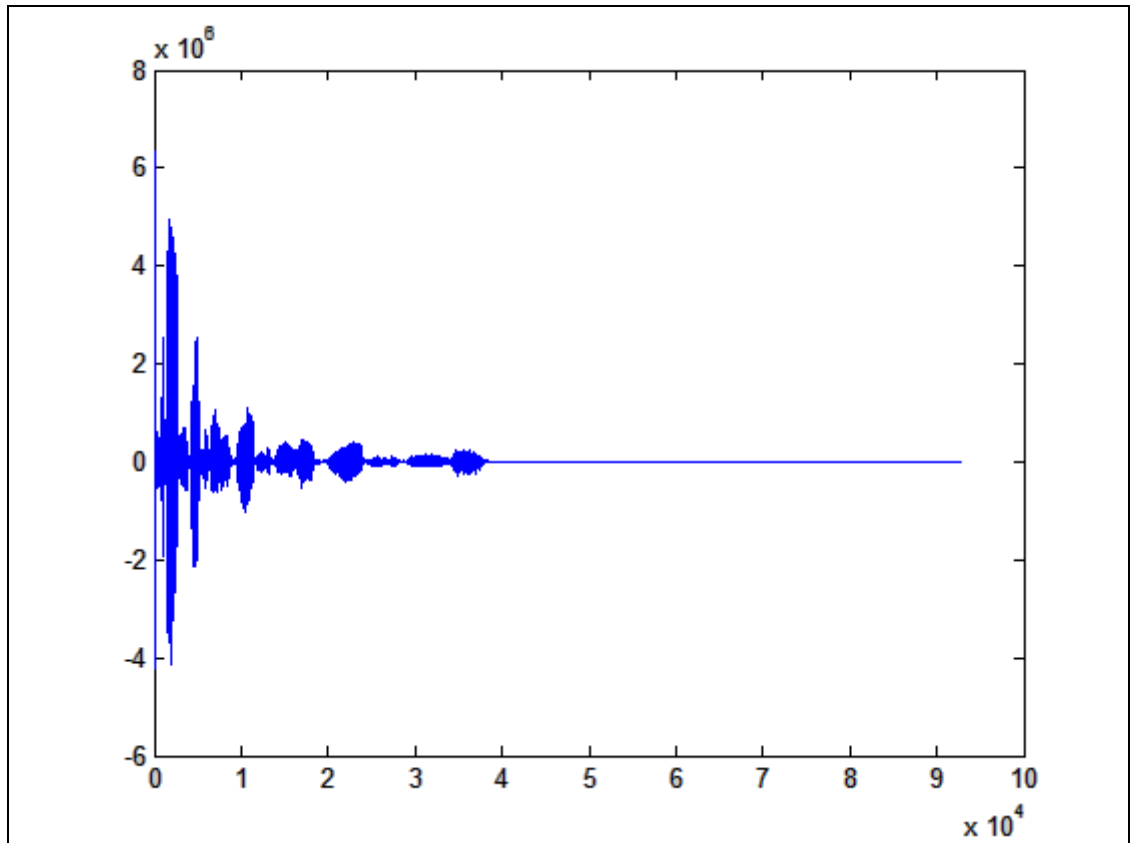


Figura 33 - Gráfico da palavra BF2-2 (com valor máximo no início)

Após a translação dos vetores com seus valores máximos no início do vetor, foi realizada a subtração dos vetores. Mesmo com a translação dos valores máximos para o início do vetor para a tentativa de alinhamento dos vetores, não foi possível chegar a um resultado conclusivo, conforme apresenta a figura 34, onde foi realizada a subtração das palavras BF2 e BF2-2 com a finalidade de aproximar o vetor resultante de zero.

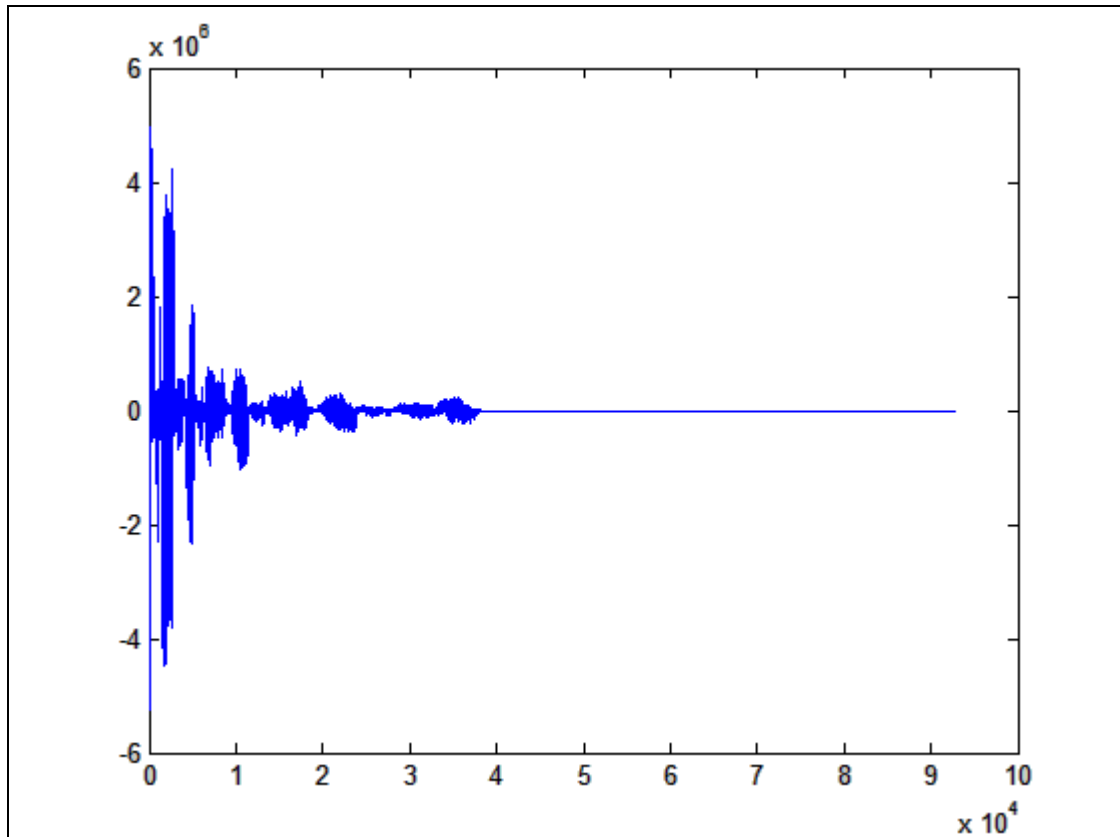


Figura 34 - Gráfico da subtração das palavras BF2 e BF2-2 (com valores máximos no início)

Em virtude de todos os testes realizados não permitirem uma conclusão definitiva, ou seja, o reconhecimento da palavra analisada, optou-se por utilizar o método Análise de Componentes Principais (ACP) .

Porém, devido ao fato de os sinais originais serem de grande porte, e, mesmo ao aplicar as transformadas *wavelets*, os mesmos continuaram desta forma, foi necessário aplicar o conceito de *cepstrum* de potência (*power cepstrum*), visando reduzir o número de elementos e, desta forma, aplicar o método ACP.

Primeiramente aplicou-se a Transformada de Fourier aos sinais analisados para calcular o valor absoluto, elevou-se ao quadrado, obteve-se o logaritmo e novamente a Transformada de Fourier foi aplicada. Finalmente, retirou-se o valor absoluto desse resultado, elevou-se este resultado ao quadrado e aplicou-se o *cepstrum* de potência, conforme apresentado no algoritmo abaixo:

Sinal  $\rightarrow$  FT  $\rightarrow$  abs() $\rightarrow$  square  $\rightarrow$  log  $\rightarrow$  FT  $\rightarrow$  abs() $\rightarrow$  square  $\rightarrow$  Power Cepstrum

A figura 35 apresenta esse procedimento para as palavras BF1 e BF2:

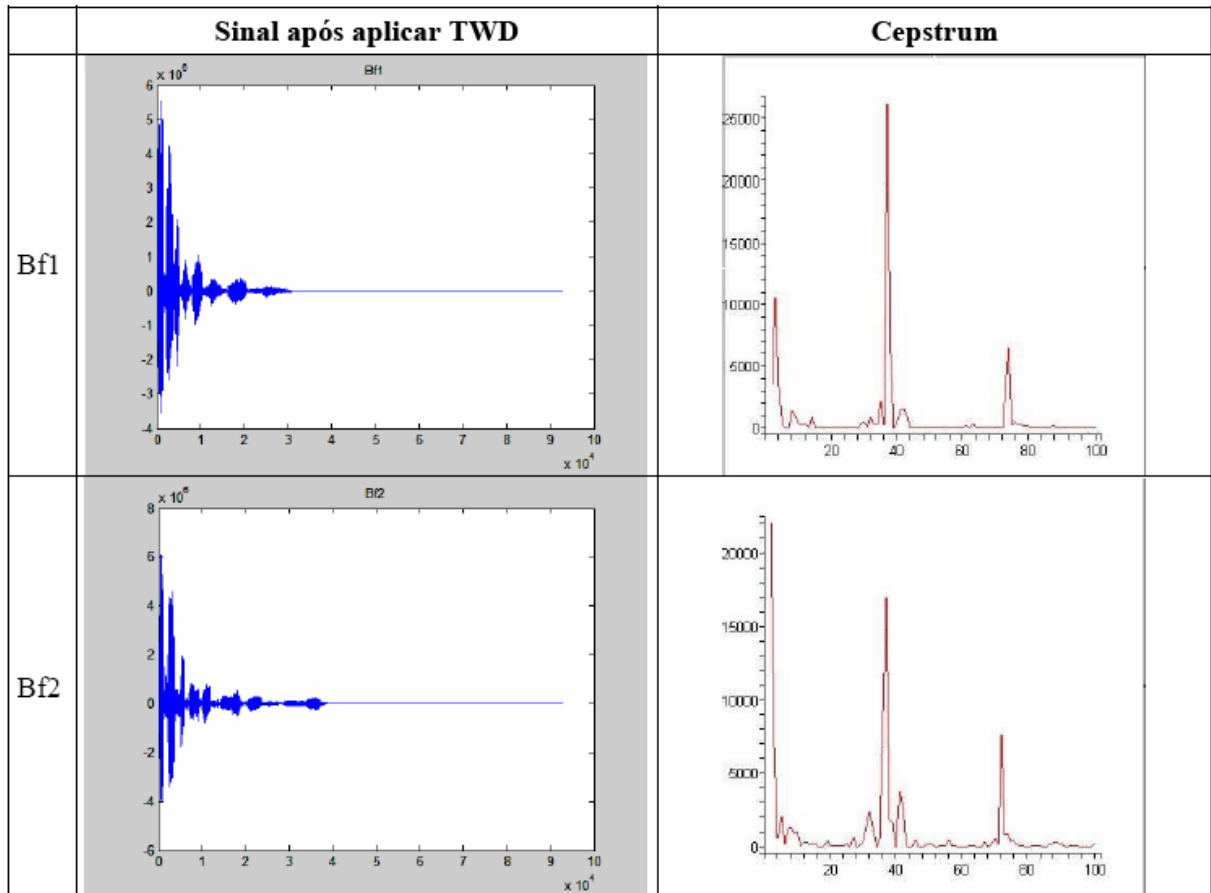


Figura 35 – Aplicação do Cepstrum de Potência para as palavras BF1 e BF2

Desta forma, obteve-se uma combinação de espectros com uma quantidade menor de pontos (99 pontos para os sinais analisados), possibilitando aplicar o método ACP.

Este mesmo procedimento foi aplicado para cada um dos 41 sons analisados. Todas as imagens dos sinais se encontram anexas nesta dissertação.

A partir destes 99 pontos de cada uma das 41 amostras, criou-se um arquivo com extensão .dat com cada um desses "vetores", e gerou-se uma matriz de 41 x 99 elementos, com a qual se calculou a matriz de covariância, representando o

*cepstrum* de potência de cada sinal compactado. Esta matriz contém 9801 elementos (99 x 99). A figura 36 ilustra o gráfico realizado apresentando o logaritmo do autovalor acrescentado em uma unidade, ou seja,  $y=\log(x+1)$ .

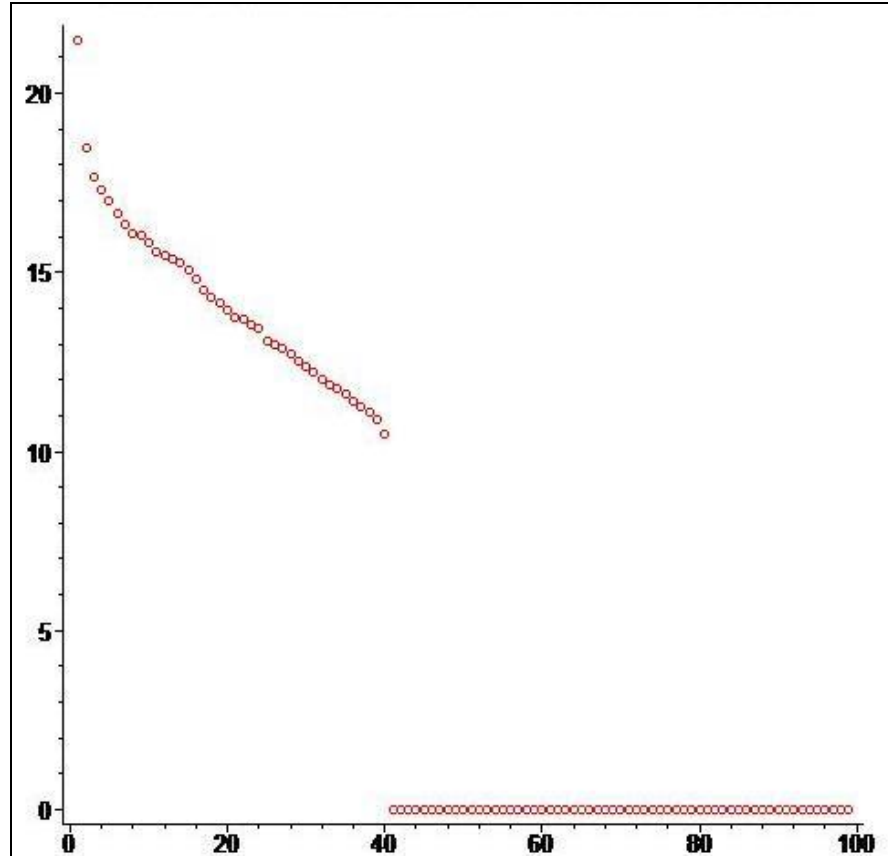


Figura 36 – Diagrama dos autovalores da matriz de covariância:  $y=\log(x+1)$ .

Após a aplicação do *cepstrum* de potência nos sinais analisados, foi possível utilizar o método de Análise de Componentes Principais (ACP) para classificação das palavras. A figura 37 apresenta um gráfico com uma análise ACP com 4 componentes principais, onde se pode observar 4 grupos distintos de sinais *power cepstrum*.

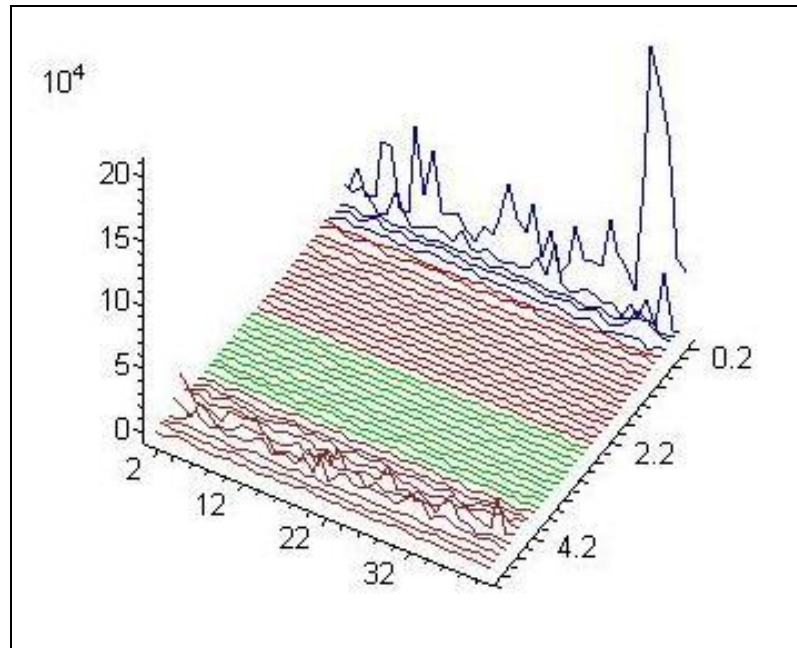


Figura 37 – ACP dos 41 sinais power cepstrum com 4 componentes principais

A associação dos sinais no gráfico apresentado na figura 37 é feita através de cores, no qual 10 sinais estão em primeiro plano na cor marrom, 11 sinais na cor verde, 14 sinais na cor vermelha e 5 sinais na cor azul. Uma análise detalhada deve ser feita para os principais autovalores para verificar como cada *cepstrum* de um sinal está associado ao *cepstrum* de outro.

É possível verificar, de acordo com o *cepstrum* dos sinais, que as palavras MFB, MLJ, MOK e MOJ estão contidas no grupo azul. Se for uma identificação de tipo gráfico, deve-se analisar cada grupo de *cepstrum* dos sinais para que seja possível fazer uma associação com os próprios sinais originais.

## CONCLUSÕES

O presente trabalho objetivou apresentar um estudo das técnicas de reconhecimento de padrões de áudio utilizando as transformadas *Wavelets* Haar e Daubechies, definir qual a melhor técnica a ser utilizada dentro de uma indústria fumageira para classificação de comandos de voz no processo de compra de fumo e possibilitar, em trabalhos futuros, desenvolver e aplicar estas técnicas em um sistema de reconhecimento e classificação de comandos de voz utilizando *wavelets*.

Durante a execução dos testes, observou-se que os arquivos de áudio coletados para o experimento eram de grande porte, mesmo após a aplicação das transformadas *Wavelets*, tanto Haar, quanto Daubechies, dificultando a aplicação do método ACP (Análise de Componentes Principais). Em virtude disso, fez-se necessário aplicar o conceito de *cepstrum* de potência (*power cepstrum*), visando reduzir o número de elementos.

De posse desses dados, verificou-se diferenças consideráveis entre alguns gráficos gerados em relação aos fonemas analisados, conforme se pode observar na seção Anexo deste documento. Desta forma, foi possível reduzir o número de elementos e aplicar o método ACP para classificação das palavras.

A partir dos dados gerados, notou-se quatro grupos distintos de sinais *power cepstrum*, obtendo-se êxito na classificação dos elementos, neste caso, orientados pela frequência de cada fonema analisado.

Foi possível concluir que as transformadas *Wavelets* utilizadas, neste caso, Haar e Daubechies, possibilitaram o reconhecimento de padrões. Porém, possuem limitações quando se trata de sinais de grande porte, necessitando, assim, de um pré-processamento antes da sua utilização, visto que mesmo após a aplicação dessas transformadas, os sinais se mantiveram de grande porte.

As transformadas *Wavelets*, associadas a outros métodos matemáticos, como a ACP e o *cepstrum* de potência, tornam-se ferramentas poderosas no que diz



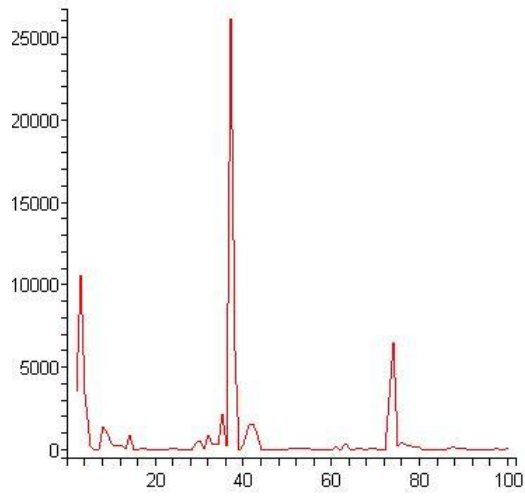
respeito ao reconhecimento e classificação de padrões, podendo ser utilizadas na indústria, em diversos segmentos.

Esta dissertação pode servir de insumo para futuros trabalhos de pesquisa como eliminação de ruídos, determinação de erros do método e possíveis filtros alternativos, abrindo novas oportunidades de pesquisa.

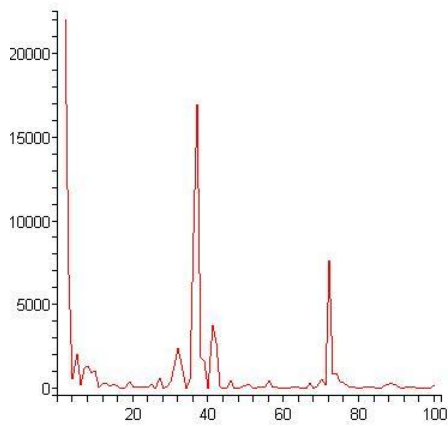
## REFERÊNCIAS

- BARROS, C.; PAULINO, W. *O Corpo Humano*. [S.L.]: Editora Ática, 1998.
- CASTILHO, Marzon Pereira de. JavaSpeech: Reconhecimento e sintetização de voz com Java. *Revista Java Magazine*. [S.l.] n. 4, p. 41-44, [ca. 2005]
- DAUBECHIES, Ingrid. *Ten lectures on wavelets*. Philadelphia: SIAM, 1992.
- DIAS, Maurício Henrique Costa. Estimação das respostas do canal real de propagação rádio-móvel nos domínios espacial e temporal Análise da supressão de ruído por decomposição Wavelet como técnica complementar de processamento. Tese – Pontifícia Universidade Católica do Rio de Janeiro – PUC-RIO, 2003.
- LIMA, Paulo Cupertino de. *Wavelets: uma introdução*. ICEX-UFMG, 2003
- LUFT, Celso Pedro. *Novo Manual de Português*. [S.L.]: Editora Globo S.A., 1995.
- MOITA NETO, J.M.; MOITA G.C. Uma Introdução a Análise Exploratória de Dados Multivariados, Teresina: [s.n], 1997.
- OLIVEIRA, Hélio Magalhães de. *Wavelets: Entrando na Onda*. Recife: Universidade Federal de Pernambuco, 2002.
- PAULA, Maurício Braga de. *Reconhecimento de palavras faladas utilizando Redes Neurais Artificiais*. Rio Grande do Sul: UFPEL, 2000.
- RIBEIRO, F.A.L.; FERREIRA M.M.C. Análise de Componentes Principais como Ferramenta para a Investigação de Contaminação Ambiental: Um Estudo de Caso. 1997.
- RODRIGUES, P.C.; BRANCO, J.A. A Análise de Componentes Principais sobre dados dependentes, 2006. p. 3-26.
- STEIGER, W.R.G.: *Cepstrum*. Curitiba: Universidade Federal do Paraná, 2008.
- WALKER, James S. *A Primer on Wavelets and Their Scientific Applications*, Boca Raton, 1999.

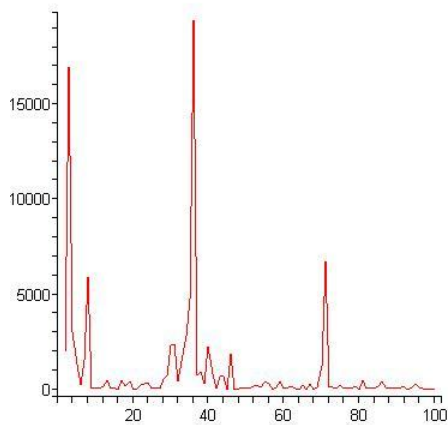
## ANEXO



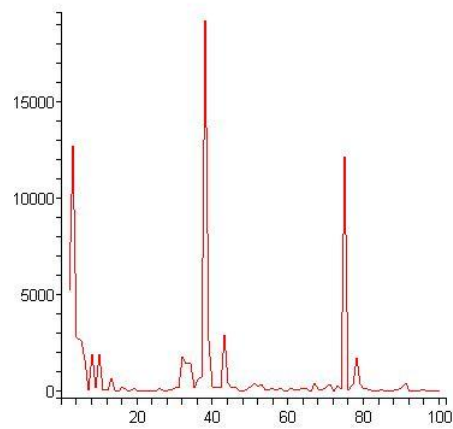
Aplicação do Power Cepstrum  
para a palavra BF1



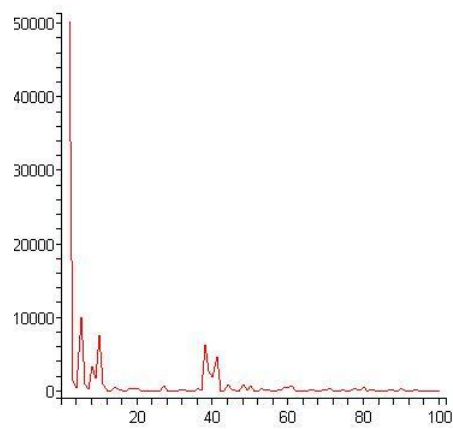
Aplicação do Power Cepstrum  
para a palavra BF2



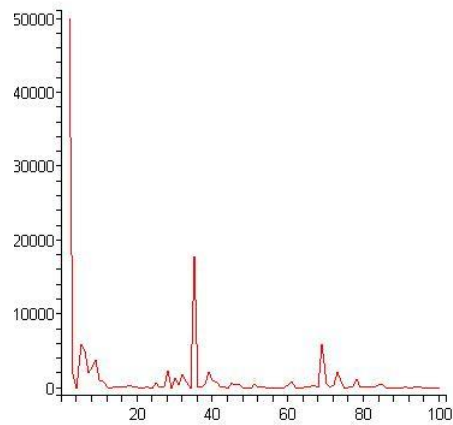
Aplicação do Power Cepstrum  
para a palavra BFB



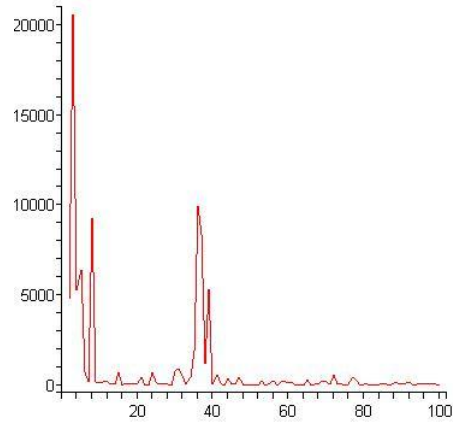
Aplicação do Power Cepstrum  
para a palavra BFC



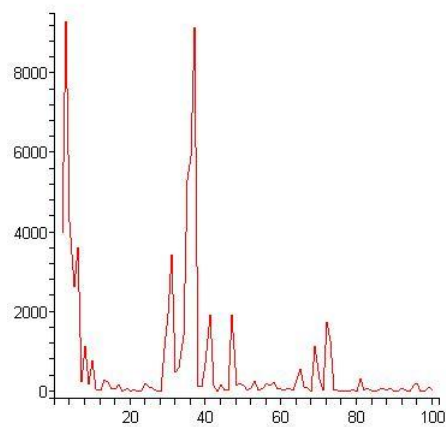
Aplicação do Power Cepstrum  
para a palavra BFJ



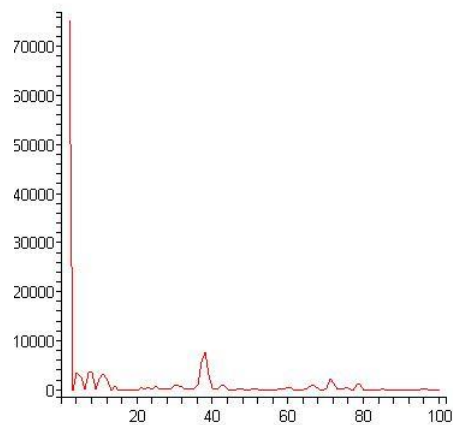
Aplicação do Power Cepstrum  
para a palavra BFK



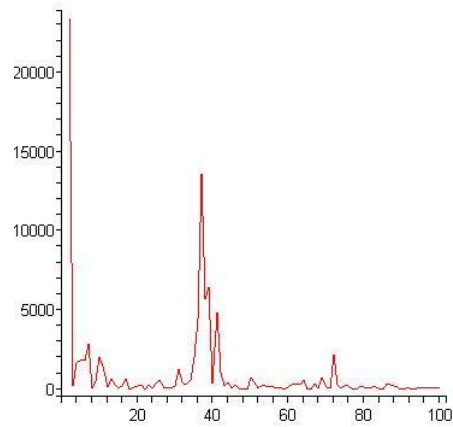
Aplicação do Power Cepstrum  
para a palavra BFM



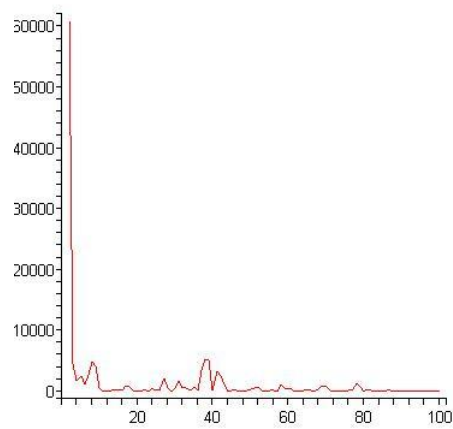
Aplicação do Power Cepstrum  
para a palavra BL3



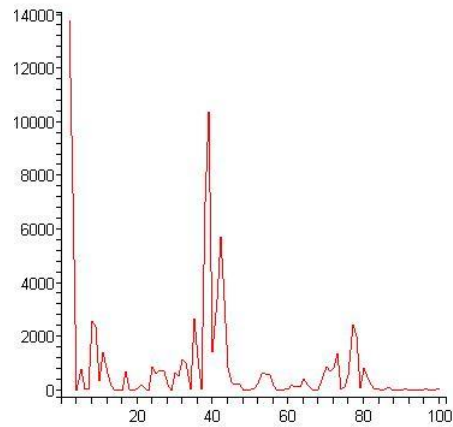
Aplicação do Power Cepstrum  
para a palavra BO2



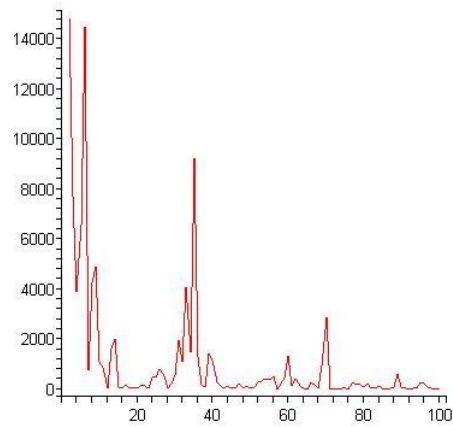
Aplicação do Power Cepstrum  
para a palavra BO3



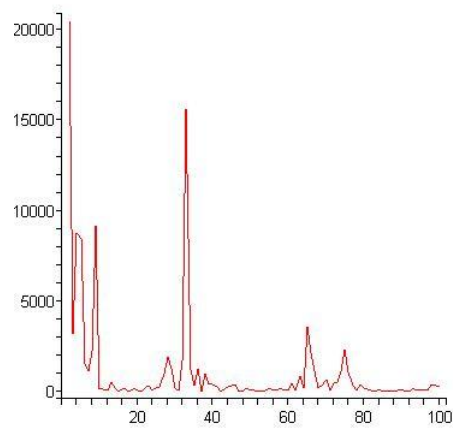
Aplicação do Power Cepstrum  
para a palavra BOK



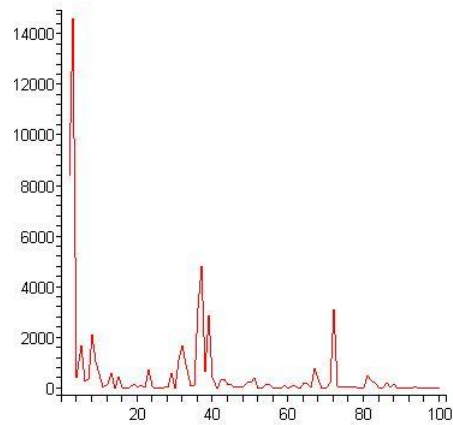
Aplicação do Power Cepstrum  
para a palavra BOM



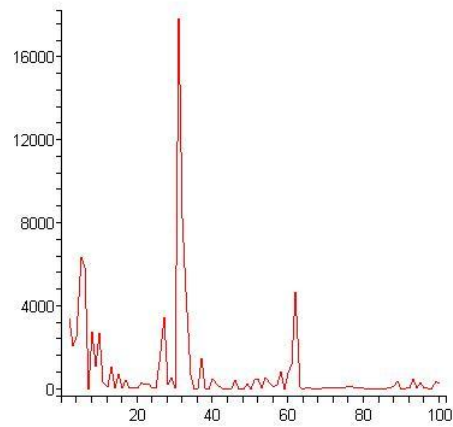
Aplicação do Power Cepstrum  
para a palavra BV3



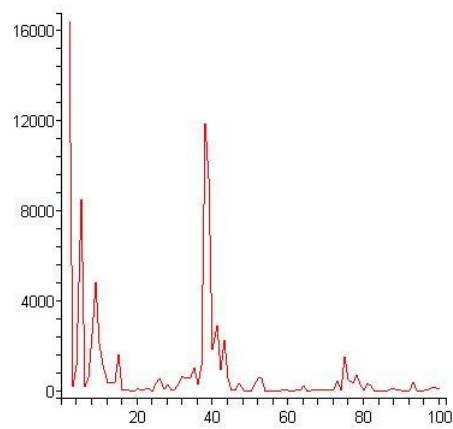
Aplicação do Power Cepstrum  
para a palavra CF1



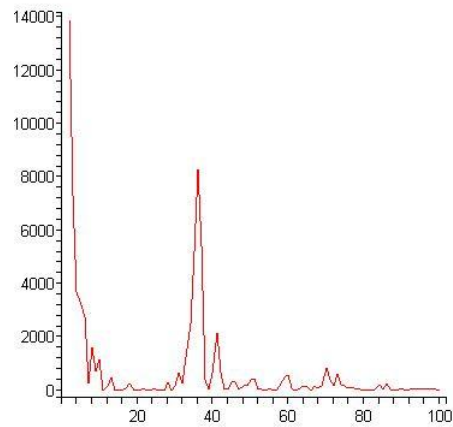
Aplicação do Power Cepstrum  
para a palavra CFB



Aplicação do Power Cepstrum  
para a palavra CFC

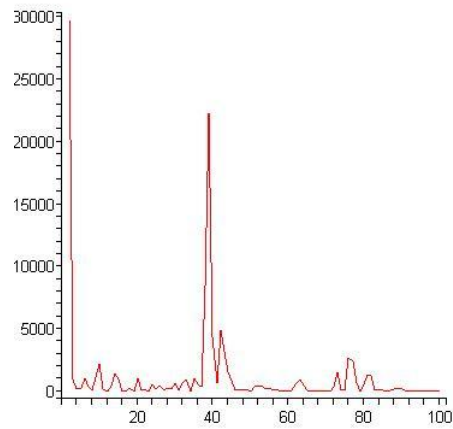


Aplicação do Power Cepstrum  
para a palavra CL3

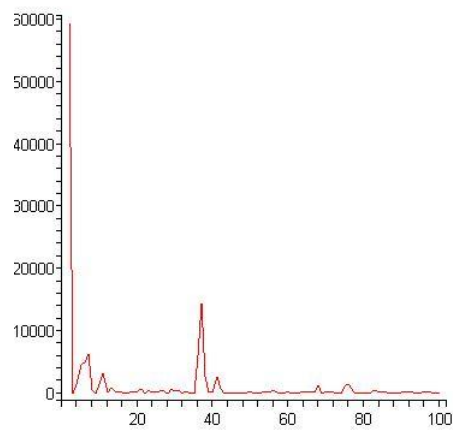


Aplicação do Power Cepstrum  
para a palavra CLC

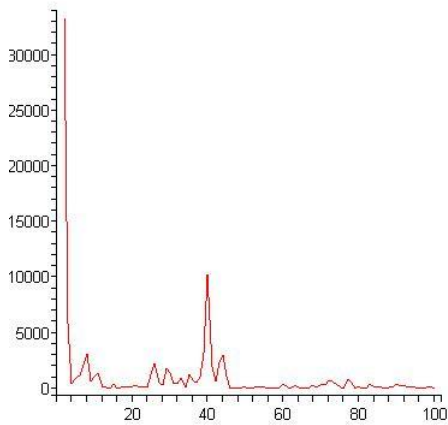




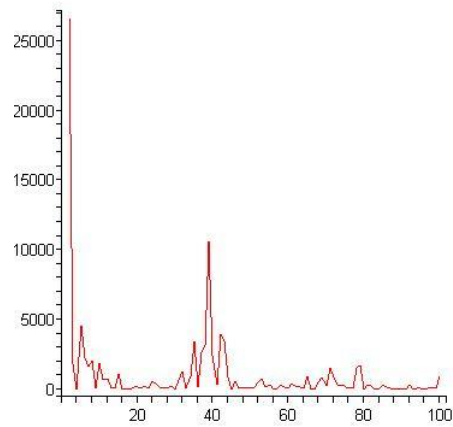
Aplicação do Power Cepstrum  
para a palavra CO1



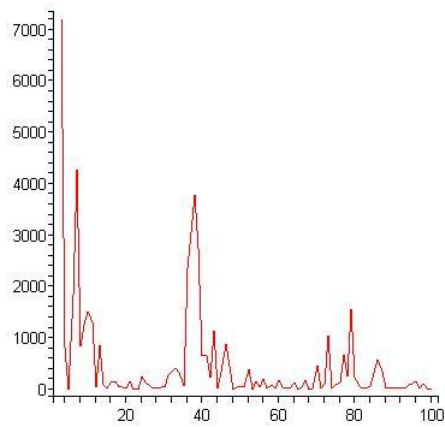
Aplicação do Power Cepstrum  
para a palavra CO2



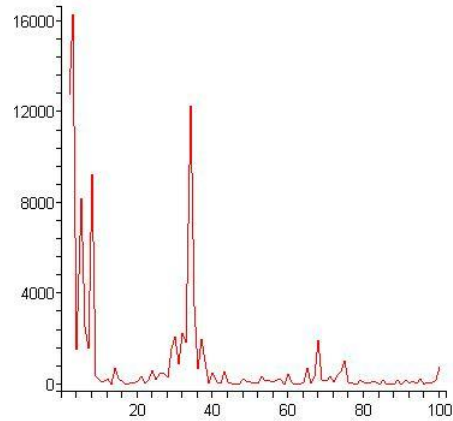
Aplicação do Power Cepstrum  
para a palavra COB



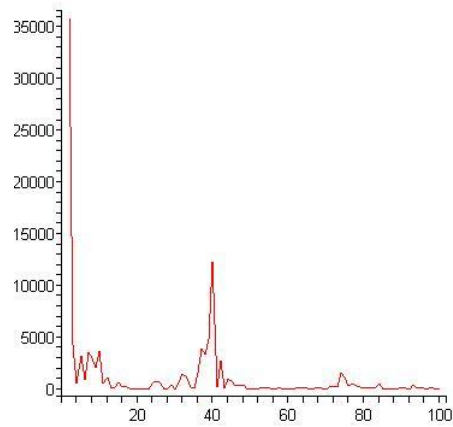
Aplicação do Power Cepstrum  
para a palavra COM



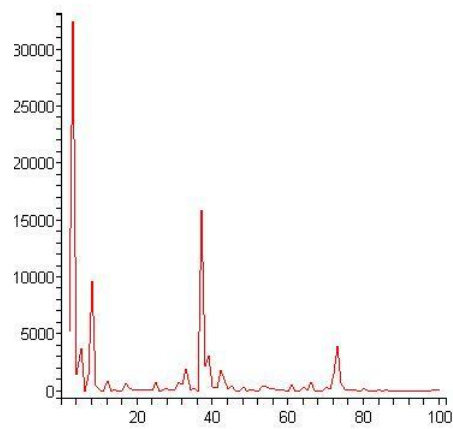
Aplicação do Power Cepstrum  
para a palavra D1TF2



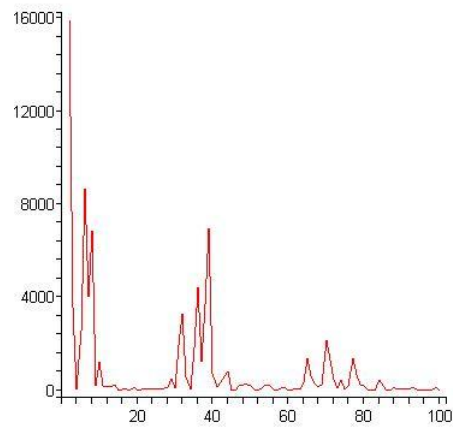
Aplicação do Power Cepstrum  
para a palavra MF1



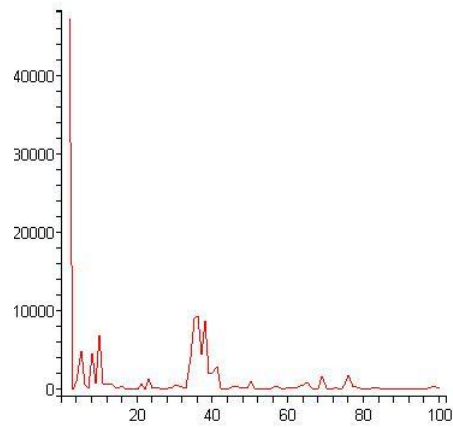
Aplicação do Power Cepstrum  
para a palavra MF2



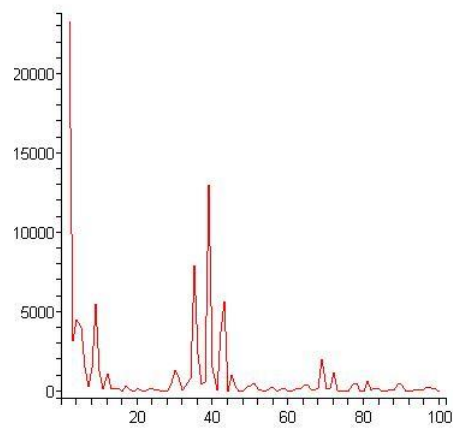
Aplicação do Power Cepstrum  
para a palavra MFB



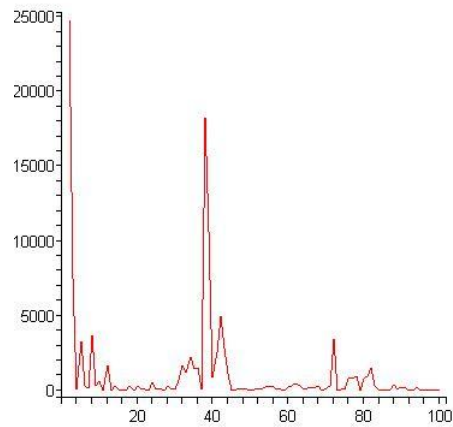
Aplicação do Power Cepstrum  
para a palavra ML2



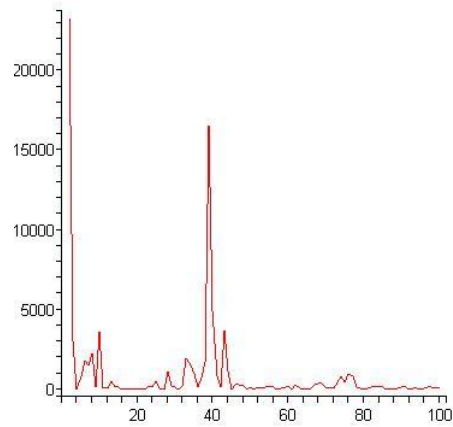
Aplicação do Power Cepstrum  
para a palavra MLJ



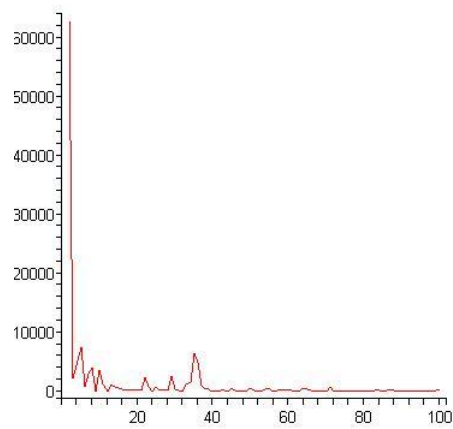
Aplicação do Power Cepstrum  
para a palavra MLK



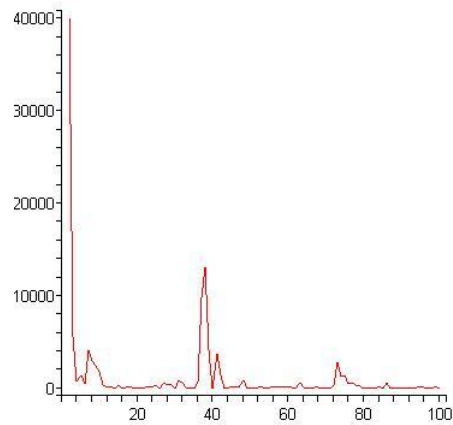
Aplicação do Power Cepstrum  
para a palavra MO1



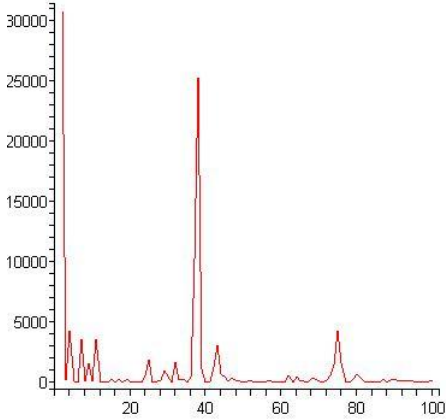
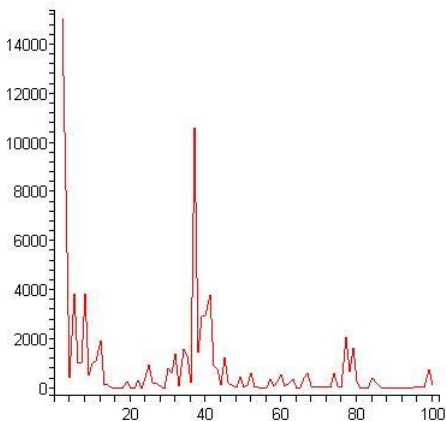
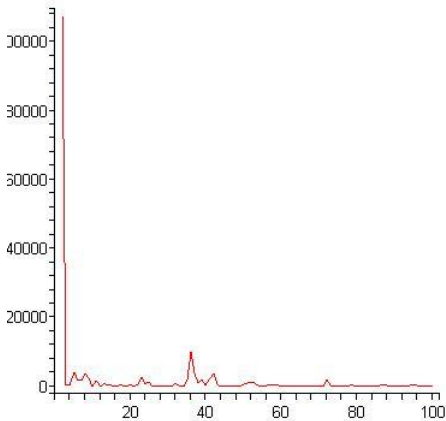
Aplicação do Power Cepstrum  
para a palavra MO3

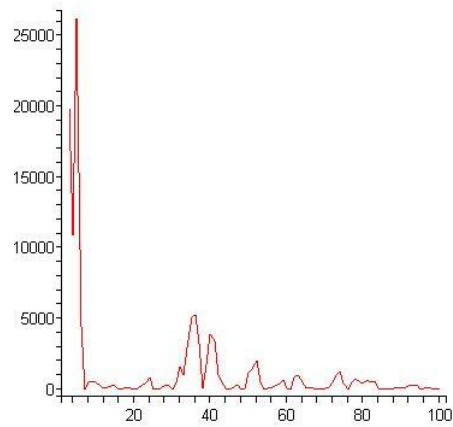


Aplicação do Power Cepstrum  
para a palavra MOJ

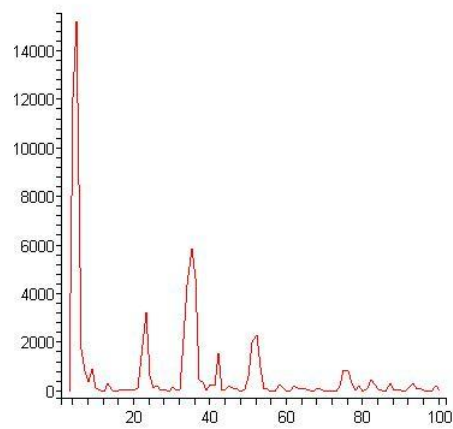


Aplicação do Power Cepstrum  
para a palavra MOK

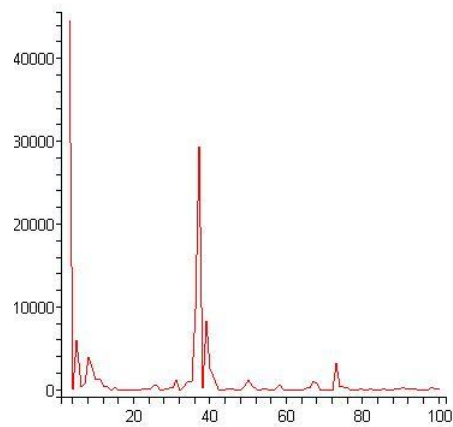
 <p>A Power Cepstrum plot for the word 'TF1'. The vertical axis (y-axis) represents power, ranging from 0 to 30,000 with major ticks every 5,000. The horizontal axis (x-axis) represents frequency in Hertz, ranging from 0 to 100 with major ticks every 20. The plot shows a very high initial peak at 0 Hz (around 30,000), followed by several smaller peaks. A prominent peak is visible at approximately 38 Hz, reaching a power of about 25,000. Another significant peak is at approximately 75 Hz, reaching a power of about 4,000.</p>	<p>Aplicação do Power Cepstrum para a palavra TF1</p>
 <p>A Power Cepstrum plot for the word 'TFE'. The vertical axis (y-axis) represents power, ranging from 0 to 14,000 with major ticks every 2,000. The horizontal axis (x-axis) represents frequency in Hertz, ranging from 0 to 100 with major ticks every 20. The plot shows a very high initial peak at 0 Hz (around 14,000), followed by several smaller peaks. A prominent peak is visible at approximately 38 Hz, reaching a power of about 10,000. Another significant peak is at approximately 75 Hz, reaching a power of about 2,000.</p>	<p>Aplicação do Power Cepstrum para a palavra TFE</p>
 <p>A Power Cepstrum plot for the word 'TLJ'. The vertical axis (y-axis) represents power, ranging from 0 to 50,000 with major ticks every 10,000. The horizontal axis (x-axis) represents frequency in Hertz, ranging from 0 to 100 with major ticks every 20. The plot shows a very high initial peak at 0 Hz (around 50,000), followed by several smaller peaks. A prominent peak is visible at approximately 38 Hz, reaching a power of about 10,000. Another significant peak is at approximately 75 Hz, reaching a power of about 2,000.</p>	<p>Aplicação do Power Cepstrum para a palavra TLJ</p>



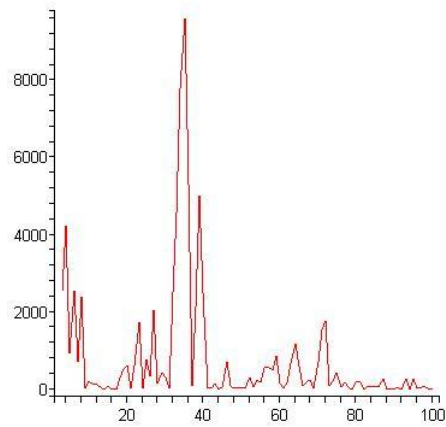
Aplicação do Power Cepstrum  
para a palavra TO1



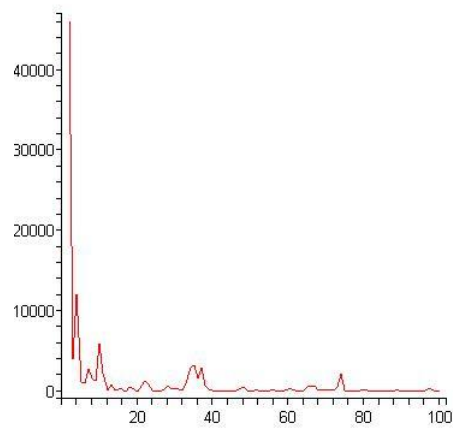
Aplicação do Power Cepstrum  
para a palavra TOJ



Aplicação do Power Cepstrum  
para a palavra TOK



Aplicação do Power Cepstrum  
para a palavra XO1



Aplicação do Power Cepstrum  
para a palavra XO3