

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA
MESTRADO EM INFORMÁTICA

Clovis Ferreira dos Reis

**Sistema Modular para Detecção e
Reconhecimento de Disparos de Armas de Fogo**

João Pessoa - PB

2015

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA
MESTRADO EM INFORMÁTICA

Clovis Ferreira dos Reis

**Sistema Modular para Detecção e Reconhecimento de
Disparos de Armas de Fogo**

Orientador: Prof. Dr. Ruy Alberto Pisani Altafim

João Pessoa - PB
2015

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA
MESTRADO EM INFORMÁTICA

Sistema Modular para Detecção e Reconhecimento de Disparos de Armas de Fogo

Dissertação de Mestrado apresentada ao Centro de Informática da Universidade Federal de Paraíba por Clovis Ferreira dos Reis, sob a orientação do Prof. Dr. Ruy Alberto Pisani Altafim, como parte dos requisitos para obtenção do título de Mestre em Informática.

João Pessoa, PB
2015

R375s Reis, Clovis Ferreira dos.
Sistema modular para detecção e reconhecimento de
disparos de armas de fogo / Clovis Ferreira dos Reis. - João
Pessoa, 2015.
108f. : il.
Orientador: Ruy Alberto Pisani Altafim
Dissertação (Mestrado) - UFPB/CI
1. Informática. 2. Armas de fogo. 3. Sons impulsivos -
detecção. 4. Reconhecimento - disparos.

UFPB/BC

CDU: 004(043)

Ata da Sessão Pública de Defesa de Dissertação de Mestrado de **CLOVIS FERREIRA DOS REIS**, candidato ao título de Mestre em Informática na Área de Sistemas de Computação, realizada em 04 de dezembro de 2015.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Ao quarto dia do mês de dezembro do ano de dois mil e quinze, às quatorze horas, no Centro de Informática - Universidade Federal da Paraíba (unidade Mangabeira), reuniram-se os membros da Banca Examinadora constituída para julgar o Trabalho Final do **Sr. Clóvis Ferreira dos Reis** vinculado a esta Universidade sob a matrícula 2014108609, candidato ao grau de Mestre em Informática, na área de “*Sistemas de Computação*”, na linha de pesquisa “*Sinais, Sistemas Digitais e Gráficos*”, do Programa de Pós-Graduação em Informática, da Universidade Federal da Paraíba. A comissão examinadora foi composta pelos professores doutores: **Ruy Alberto Pisani Altafim (PPGI-UFPB)**, Orientador e Presidente da Banca, **Leonardo Vidal Batista (PPGI-UFPB)**, Examinador Interno e **Liliane Ventura Schiabel (USP)**, Examinadora Externa à Instituição. Dando início aos trabalhos, o professor Presidente da Banca cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra ao candidato para que o mesmo fizesse, oralmente, a exposição do trabalho de dissertação intitulado “*Deteção de Sons Impulsivos em um Sistema de Reconhecimento de Disparos de Armas de Fogo*”. Concluída a exposição, o candidato foi arguido pela Banca Examinadora que emitiu o seguinte parecer: “*aprovado*”. Assim sendo, eu, Nadja Rayssa Soares de Almeida, Auxiliar em Administração, Secretária do Programa de Pós Graduação em Informática - PPGI, lavrei a presente ata que vai assinada por mim e pelos membros da Banca Examinadora. João Pessoa, 04 de dezembro de 2015.


Nadja Rayssa Soares de Almeida

Prof^o Dr^o Ruy Alberto Pisani Altafim
Orientador (PPGI-UFPB)



Prof^o Dr^o Leonardo Vidal Batista
Examinadora Interna (PPGI-UFPB)



Prof^o Dr^o Liliane Ventura Schiabel
Examinadora Externa à Instituição (USP)



25

Sistema Modular para Detecção e Reconhecimento de Disparos de Armas de Fogo

Resumo

A violência urbana vêm crescendo anualmente em praticamente todos os estados brasileiros e para fazer face a essa ameaça, as autoridades policiais necessitam cada vez mais de ferramentas tecnológicas que os auxiliem na tomada de decisões sobre quando e como empregar os poucos recursos disponíveis à repressão do crime. Neste contexto, é apresentado nesse trabalho uma ferramenta computacional, passível de ser embarcada em dispositivos móveis, que possibilita realizar a detecção e reconhecimento automático de disparos de armas de fogo. Para tanto, são descritos inicialmente os fundamentos básicos sobre sons impulsivos, armas de fogo e características de disparos. Posteriormente, descreve-se uma proposta de um sistema modular de detecção e reconhecimento de disparos. No entanto, devido ao sistema conter diversos módulos complexos, este trabalho teve foco em dois deles: o módulo de detecção de sons impulsivos e o módulo de classificação, que permite distinguir disparos de armas de fogo de outros sons impulsivos. Para o módulo de detecção de sons impulsivos foram analisados três algoritmos amplamente descritos na literatura: o algoritmo da quarta derivada da RMS, o da Conditional Median Filter (CMF) e o Método da Variância (VM). Os algoritmos foram testados com base nas medidas de desempenho da acurácia, precisão, sensibilidade e especificidade. E a para determinar o método mais eficiente, foram realizados testes de cadências, com sons impulsivos sem adição de ruído sonoro, com adição de ruído constante e com ruído variável. Ao final dessa análise, os parâmetros preconizados na literatura para os métodos CMF e VM foram alterados para uma verificação de possibilidade de otimização. Definido o algoritmo de detecção de impulso mais eficiente, iniciou-se o desenvolvimento do módulo de classificação. Para isso, foram propostas duas técnicas para o reconhecimento de disparos de armas de fogo, uma utilizando uma comparação da envoltória do som no domínio do tempo e outra baseada na comparação de frequências dominantes obtidas por meio da transformada de Fourier. Numa comparação entre as duas técnicas observou-se que com a técnica da envoltória é possível identificar 54% dos sons impulsivos, enquanto que com a técnica baseada no domínio da frequência, este percentual foi de 72%.

Palavras-chaves: Sons Impulsivos, Armas de fogo, Detecção de sons Impulsivos, Reconhecimento de disparos.

Modular System for Detection and Recognition of Firearm's Gunshots

Abstract

The urban violence has been increasing in almost Brazilian state and in order to face this threat, new technological tools are required by the police authorities in order to support their decisions on how and when the few available resources should be employed to combat criminality. In this context, this work presents an embedded computational tool that is suitable for detecting gun-shots automatically. To provide the necessary knowledge to understand the work, a brief description about impulsive sounds, fire guns and the gun-shot characteristics are initially presented. Latter, a system based on modules is proposed to detect and recognize impulsive sound, which are characteristics of gun-shots. However, since the system contain several modules in this work we have focus only on two of them: the module for detecting impulsive sounds and the module for distinguish a gun-shot from any other impulsive sound. For the impulsive detection module, three well-known algorithms were analyzed on the same condition: the fourth derivative of the Root Median Square (RMS), the Conditional Median Filter (CMF) and the Variance Method (VM). The algorithms were tested based on four measured performance parameters: accuracy, precision, sensibility and specificity. And in order to determine the most efficient algorithm for detecting impulsive sounds, a cadence test with impulsive sounds, without or with additional noise (constant or increasing) was performed. After this analysis, the parameters employed on the CMF and VM method were tested in a wide range of configurations to verify any possibility of optimization. Once this optimal method was determined, the classification module to recognize gun-shots started to be implemented. For this, two distinguish methods were compared, one based on the signal wrapped over the time and the other based on most relevant frequencies obtained from the Fourier transform. From the comparison between the two methods it was observed that the wrapped method provided 54% of accuracy in the classification of impulsive sounds, while with the frequency analysis this value was 72%.

Key words: Impulsive Sounds, Fire Guns, Impulsive Sound Detection, Gun Shot Recognition.

Sumário

	Lista de ilustrações	iv
	Lista de tabelas	vii
1	INTRODUÇÃO	1
1.1	Objetivos do trabalho	2
1.2	Escopo e organização do trabalho	2
2	REVISÃO BIBLIOGRÁFICA	4
2.1	Sons Impulsivos	4
2.2	Disparos de Armas de Fogo	7
2.3	Conceitos básicos sobre armas de fogo	8
2.4	Métodos de Detecção de Sons Impulsivos	14
2.5	Técnicas no Domínio das Frequências	22
2.6	Técnicas de Aprendizado de Máquina	30
3	O SISTEMA MODULAR - PROPOSTAS E SOLUÇÕES	39
3.1	Sistemas existentes	39
3.2	Módulos componentes do sistema	40
3.3	Módulo de Detecção de Impulso	43
3.4	Módulo de Reconhecimento de Disparos	45
4	METODOLOGIA	53
4.1	Software Utilizado	53
4.2	Módulo de Detecção de Impulso	54
4.3	Módulo de Reconhecimento de Disparo	63
5	RESULTADOS E DISCUSSÃO	69
5.1	Detecção de Impulsos	69
5.2	Reconhecimento de Disparos	79
6	CONCLUSÃO	89
	Referências	92
	APÊNDICE A – ALGORITMO DO K-MÉDIAS	96
	APÊNDICE B – INTRODUÇÃO DE DADOS NO WEKA	97

Lista de ilustrações

Figura 1 – Exemplos de sons impulsivos	4
Figura 2 – Esquema ADSR	5
Figura 3 – Explosão de Friedlander	6
Figura 4 – Sons Impulsivos Gerados no Octave	7
Figura 5 – Componentes de um disparo de arma de fogo.	8
Figura 6 – Tipos de Armamento.	12
Figura 7 – Cartucho e Projetis	13
Figura 8 – Exemplo de RMS com $N = 20$ amostras sobre um sinal impulsivo.	15
Figura 9 – Exemplo de detecção utilizando-se a 4ª derivada do RMS e limiar ajustável.	16
Figura 10 – Exemplo do processo de detecção utilizando-se a RMS.	17
Figura 11 – Exemplo de uma janela contendo $L = 6$ frames $e(k)$	18
Figura 12 – Atraso do Limiar Adaptativo	19
Figura 13 – Exemplo das fases componentes do processo de detecção utilizando-se a CMF.	20
Figura 14 – Janela da variância ($L = 10$) onde nenhum frame possui som impulsivo	21
Figura 15 – Janela da variância ($L = 10$) onde o último frame possui som impulsivo	22
Figura 16 – Exemplo das fases componentes do processo de detecção utilizando-se a VM.	23
Figura 17 – Sinal contendo um disparo de fuzil e a sua FFT.	25
Figura 18 – Sinal contendo um latido de cachorro e a sua FFT.	26
Figura 19 – Transformada de Fourier aplicada sobre um sinal com taxa de amostra- gem de 1kHz, de comprimento igual a 1001 amostras, sem aplicação de <i>Zero Padding</i>	27
Figura 20 – Transformada de Fourier aplicada sobre um sinal com taxa de amostra- gem de 1kHz, de comprimento igual a 1001 amostras, após a aplicação de <i>Zero Padding</i> com 2000 amostras.	28
Figura 21 – Exemplo de sinal gerado antes e depois de ser submetido aos filtro <i>“Spectral Noise Gating”</i>	29
Figura 22 – Exemplo de classificação de armamento quanto ao funcionamento.	32
Figura 23 – Algoritmos de Distância.	34
Figura 24 – Fases do algoritmo K-Médias.	35
Figura 25 – Dados normalizados separados por classe e Hiperplanos possíveis.	36
Figura 26 – Vetores de Suporte, Limite de Classes e Hiperplano Separatório.	37
Figura 27 – Classificação, com base no Hiperplano, de Entidades cuja classe é desconhecida.	37
Figura 28 – Modelo Proposto em [1]	40

Figura 29 – Modelo do Sistema Proposto	41
Figura 30 – Janela da variância ($L = 10$) onde o <i>frame</i> e_7 contém um som impulsivo previamente detectado.	44
Figura 31 – Janela da variância onde o <i>frame</i> e_7 que continha um som impulsivo previamente detectado é substituído pela média dos $L-1$ <i>frames</i> anteriores.	45
Figura 32 – Gráfico comparativo dos resultados de detecção.	46
Figura 33 – Exemplo de Envoltória de um sinal.	47
Figura 34 – Transformação de um sinal em imagem.	48
Figura 35 – Potência do pulso de Friedlander com duração de 0.15 s, simulando um disparo ideal de arma de fogo, transformado em imagem (F).	49
Figura 36 – Exemplo da imagem do som impulsivo analisado subtraída do impulso ideal obtido pela equação de Friedlander.	50
Figura 37 – Exemplo do cálculo do Coeficiente de Variação e Seleção de Frequências.	52
Figura 38 – Representação gráfica dos sons utilizados nos testes.	56
Figura 39 – Exemplos de sinais gerados para os testes comparativos entre métodos.	57
Figura 40 – Diagrama do Teste de Cadência.	58
Figura 41 – Diagrama do Teste de Otimização.	59
Figura 42 – Diagrama do Teste de Validação de Parâmetros	60
Figura 43 – Diagrama do Teste de Generalização do Método.	61
Figura 44 – Diagrama do Teste no domínio do Tempo.	65
Figura 45 – Esquema de geração dos arquivos <i>ARFF</i> no domínio do tempo.	65
Figura 46 – Esquema de processamento do algoritmo K-Médias no WEKA.	66
Figura 47 – Esquema de processamento do algoritmo SVM no WEKA.	66
Figura 48 – Diagrama do Teste no domínio das Frequências.	67
Figura 49 – Esquema de geração dos arquivos <i>ARFF</i> no domínio das frequências.	68
Figura 50 – Esquema do processo de análise de quantidade de frequências.	68
Figura 51 – Matriz de confusão do RMS com parâmetros padrão.	69
Figura 52 – Matriz de confusão do CMF com parâmetros padrão.	70
Figura 53 – Matriz de confusão do VM com parâmetros padrão.	70
Figura 54 – Gráfico comparativo entre M e FP em função de N e L da CMF – Sem ruído (SR).	71
Figura 55 – Gráfico comparativo entre M e FP em função de N e L da CMF – Ruído Constante (R1).	72
Figura 56 – Gráfico comparativo entre M e FP em função de N e L da CMF – Ruído Crescente (R2)	72
Figura 57 – Gráfico comparativo entre M e FP em função de N e L da VM – Sem ruído (SR).	73
Figura 58 – Gráfico comparativo entre M e FP em função de N e L da VM – Ruído Constante (R1).	73

Figura 59 – Gráfico comparativo entre M e FP em função de N e L da VM – Ruído Crescente (R_2).	74
Figura 60 – Matriz de confusão do CMF com parâmetros otimizados.	75
Figura 61 – Matriz de confusão do VM com parâmetros otimizados.	76
Figura 62 – Matriz de confusão do VM para detecção de diversos armamentos.	78
Figura 63 – Matriz de confusão do VM para detecção de diversos sons impulsivos.	79
Figura 64 – Gráfico da distribuição de $\Phi \times \Theta$ – Bateria de Treino.	81
Figura 65 – Gráfico da distribuição de $\Phi \times \Theta$ – Bateria de Teste.	82
Figura 66 – Resultado da clusterização do algoritmo K-Médias sobre a Bateria de Treino utilizando distância Euclidiana.	82
Figura 67 – Resultado da clusterização do algoritmo K-Médias sobre a Bateria de Teste.	83
Figura 68 – Resultado da classificação do algoritmo SVM sobre a Bateria de Treino.	84
Figura 69 – Resultado da clusterização do algoritmo SVM sobre a Bateria de Teste.	84
Figura 70 – Análise da quantidade ideal de frequências a serem utilizadas.	86

Lista de tabelas

Tabela 1 – Tipo de Funcionamento × Cadência de Tiro	10
Tabela 2 – Matriz de Confusão	62
Tabela 3 – Comparativo das medidas de desempenho dos três métodos	71
Tabela 4 – Melhores resultados dos Métodos da CMF e VM	75
Tabela 5 – Medidas de desempenho dos Métodos CMF e VM com os novos parâmetros	76
Tabela 6 – Comparativo das medidas de desempenho dos Métodos CMF e VM com os parâmetros originais e otimizados	77
Tabela 7 – VM com vários armamentos	78
Tabela 8 – VM com vários Sons Impulsivos	80
Tabela 9 – Matriz de Confusão - K-Médias	82
Tabela 10 – Medidas de Desempenho - K-Médias	83
Tabela 11 – Matriz de Confusão - SVM	84
Tabela 12 – Medidas de Desempenho - SVM	85
Tabela 13 – Matriz de Confusão – Bateria de Teste – SVM	87
Tabela 14 – Matriz de Confusão – Bateria de Teste – K-Médias	87
Tabela 15 – Comparativo das medidas de desempenho dos Métodos SVM e K-Médias	87

Lista de Símbolos e Abreviações

ac	Medida de desempenho de Acurácia
ADSR	<i>Attack, Decay, Sustain, Release</i>
ARFF	<i>Attribute-Relation File Format</i> . Formato de arquivo utilizado pelo WEKA
CI	<i>Frame</i> contendo um impulso
CMF	<i>Conditional Median Filter</i>
D	<i>Frame</i> com impulso corretamente detectado
DFT	<i>Discrete Fourier Transform</i>
FFT	<i>Fast Fourier Transform</i>
FP	<i>Frame</i> sem impulso assinalado como tendo ou <i>Falso Positivo</i>
L	Tamanho da janela utilizada no método de detecção
M	<i>Frame</i> com impulso não detectado ou <i>Missed</i>
MP3	<i>MPEG 1 Layer III</i>
N	Número de amostras utilizadas na construção do <i>frame</i>
pr	Medida de desempenho de Precisão
R1	Sinal com ruído constante a 50% da potência máxima
R2	Sinal com ruído crescente, variando de 0 a 100% da potência máxima
RMS	<i>Root Mean Square</i>
se	Medida de desempenho de Sensibilidade
SI	<i>Frame</i> que não contém impulso
SNG	<i>Spectral Noise Gating</i>
SNR	<i>Signal Noise Ratio</i>
sp	Medida de desempenho de Especificidade
SR	Sinal sem ruído
SVM	<i>Support Vector Machine</i>
TN	<i>Frame</i> sem impulso assinalado corretamente ou <i>True Negative</i>
tpm	tiros por minuto
VM	<i>Variance Method</i>
WAV	<i>Waveform Audio Format</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

1 Introdução

É fato inegável a escalada da violência no Brasil [2] [3] [4]. O crime organizado e o tráfico de drogas nas grandes cidades é cada vez mais presente e atuante. Por já viverem à margem da lei, traficantes e bandidos tem o seu acesso facilitado a um número cada vez maior e diversificado de armamentos, onde o limite é apenas o quanto se deseja gastar para ter acesso ao que há de mais avançado em termos de armas de fogo, seja comprando-as ou simplesmente alugando-as de quadrilhas especializadas no negócio cada dia mais lucrativo do tráfico de armas.

A força policial, por conta dos inúmeros entraves burocráticos da máquina governamental, não raramente permanece em defasagem, tentando realizar sua missão sem os meios necessários ao enfrentamento do crime organizado de igual para igual. Em seu embate diário, muitas vezes encontra-se em ampla desvantagem, seja numérica, seja tecnológica, em face aos criminosos [5].

Desta forma, qualquer ferramenta que venha a facilitar o trabalho das autoridades policiais, mesmo que represente apenas uma pequena vantagem, será de extrema valia, podendo representar o fator de equilíbrio em prol do estado de direito.

É dentro deste contexto que se apresenta um sistema de detecção e reconhecimento de disparos de arma de fogo, uma vez que, de posse deste instrumento, a autoridade policial pode antecipar e melhor planejar suas ações com base nas informações oportunas adquiridas quando da ocorrência de um incidente envolvendo armas de fogo.

Assim, um sistema que realiza a detecção de disparos e o reconhecimento da arma autora deste tiro, pode ter diversas aplicações práticas possíveis, a saber:

1. A obtenção de uma detecção confiável de um disparo de arma de fogo possibilita a criação de uma rede de monitoramento e alerta, permitindo que a polícia receba avisos oportunos sobre a ocorrência de tiros ou explosões em uma determinada área;
2. O reconhecimento do tipo de armamento utilizado em dado disparo serve como subsídio para a tomada de decisões por parte do policial sobre como enfrentar a situação apresentada, qual armamento mais adequado, tipo de proteção a ser utilizada na operação, entre outras;
3. Também possibilita a utilização do som como instrumento forense adicional na determinação de responsabilidades durante operações policiais, como por exemplo, a determinação de autoria do primeiro disparo após um tiroteio.

1.1 Objetivos do trabalho

O objetivo geral deste trabalho é permitir a implementação de um sistema leve e de baixo custo computacional, a ser transformado em um sistema embarcado para equipamentos portáteis e de baixo custo, como *smartphones* e *tablets*.

Entretanto, devido à magnitude do sistema em questão, o presente trabalho é limitado ao estudo de apenas dois módulos componentes do sistema completo (vide figura 29, p. 41) e, para tanto, foram definidas as seguintes características gerais:

1.1.1 Módulo de Detecção de Impulso

- Realizar o levantamento de algoritmos de detecção de impulsos;
- Identificar, dentre os algoritmos levantados, aqueles que possuam a eficiência e a simplicidade adequada a um sistema embarcado de detecção em tempo real;
- Realizar o estudo comparativo entre os algoritmos selecionados a fim de determinar sua precisão e performance.

1.1.2 Módulo de Reconhecimento de Disparo

- Propor algoritmos que permitam diferenciar um som impulsivo qualquer de um disparo de arma de fogo, utilizando técnicas no domínio do tempo ou transformadas do sinal para o domínio das frequências, que sejam, ao mesmo tempo, simples, leves e adequados à tarefa;
- Realizar testes de reconhecimento, verificando a viabilidade, eficácia e eficiência destes métodos.

1.2 Escopo e organização do trabalho

No capítulo 2 é apresentada uma revisão dos assuntos pertinentes à presente dissertação. Nele são caracterizados e definidos sons impulsivos. A seguir são caracterizados os disparos de armas de fogo como um caso especial de som impulsivo, definindo suas principais características. Passa-se, então, realizar um pequeno nivelamento de conhecimentos, apresentando definições e conceitos básicos relacionados à armas de fogo, seu funcionamento, os tipos de munição e suas principais aplicações, informações estas consideradas importantes para a correta compreensão dos demais capítulos. Em seguida discorre-se sobre as técnicas utilizadas para identificar, em tempo real, a ocorrência de sons impulsivos em meio a outros sons ambientes, identificando os modelos matemáticos disponíveis para realizar tal detecção.

Nele também serão abordados os fundamentos da Transformada de Fourier e técnicas a ela associadas. Também se discorrerá sobre técnicas de remoção de ruídos em sinais sonoros, assim como fundamentos sobre Data Mining e sobre uma de suas ferramentas, o WEKA, abordando também métodos de aprendizado de máquina como a clusterização via K-Médias e o método de classificação *Support Vector Machine* (SVM).

No capítulo 3 será apresentado o sistema modular em si, o detalhamento de seus módulos e propostas de melhorias nos métodos de detecção tradicionais. Ainda neste capítulo serão propostas técnicas simples para reconhecimento de disparos, em procedimentos baseados no tempo e nas frequências características deste tipo de som, que serão aqui testadas.

O capítulo 4 trata da metodologia empregada no presente trabalho, abordando o software utilizado, a seleção dos sons utilizados na parte experimental e dos testes idealizados para a seleção dos melhores algoritmos.

O capítulo 5 descreve os resultados obtidos nos testes. Em sua seção 5.2 vê-se os resultados referentes a detecção do som impulsivo, onde constam os testes dos métodos com os parâmetros padrão da literatura em 5.1.1, a otimização dos métodos em 5.1.2, os resultados alcançados com os parâmetros já otimizados, em 5.1.3, e os testes de generalização, em 5.1.4. Na seção 5.2 veem-se os resultados obtidos nos testes dos algoritmos de reconhecimento de disparo, separados em métodos de reconhecimento no domínio do tempo, visto em 5.2.1, e em métodos pertencentes ao domínio da frequências, em 5.2.2 .

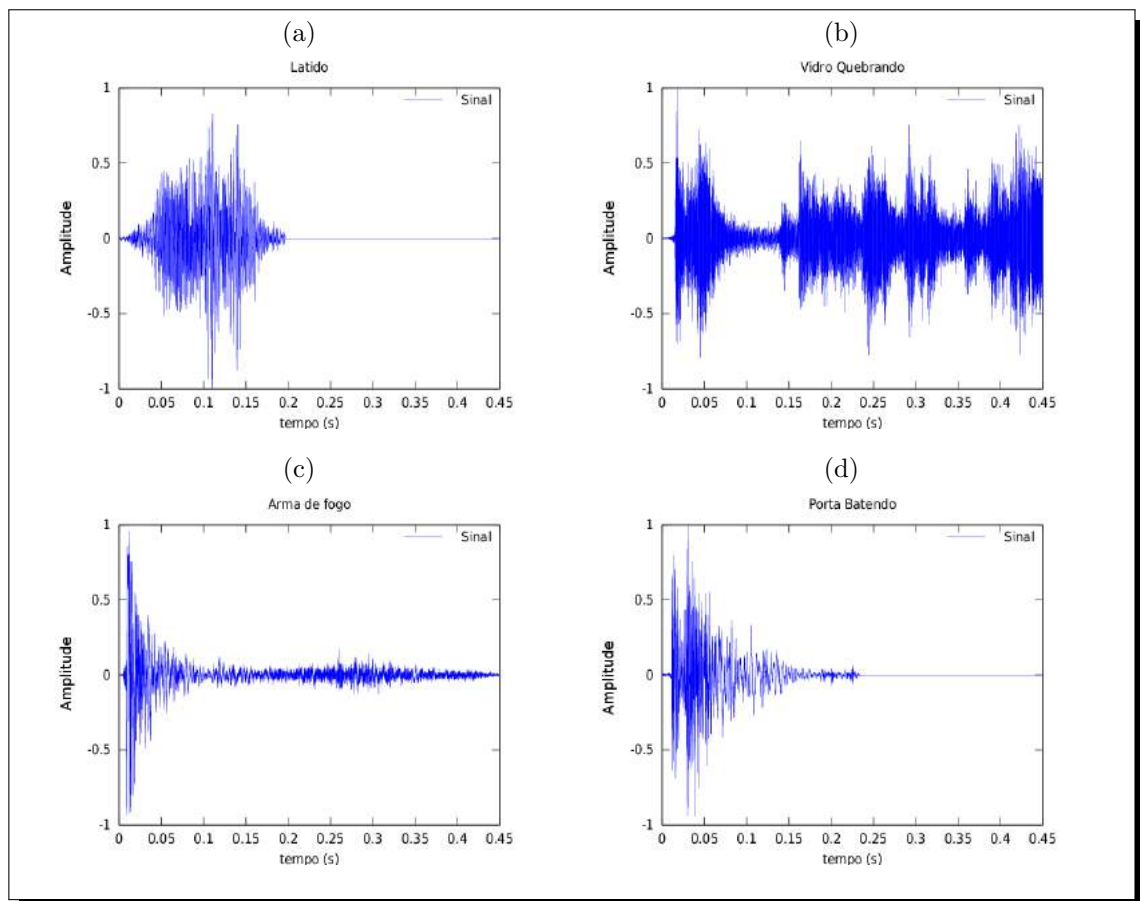
Finalmente, realizar-se-á a conclusão sobre a viabilidade de implementação do sistema aqui descrito em um modelo de produção, discorrendo sobre as técnicas utilizadas, sua eficácia e sua eficiência.

2 Revisão Bibliográfica

2.1 Sons Impulsivos

Sons Impulsivos são, por definição, sons de curta duração, com amplitude e duração aleatórias, ou mais especificamente, são aqueles gerados por uma súbita mudança na pressão do ar, produzindo uma grande amplitude com uma curta duração. Via de regra possuirá uma ascensão na casa dos milissegundos e duração na casa dos décimos de segundo [6, 7]. Eles são decorrentes de eventos que liberam muita energia sonora em um curto espaço de tempo, como é o caso dos instrumentos musicais de percussão, batidas de portas, vidros quebrando, explosões e disparos de armas de fogo. A figura 1 mostra o perfil de alguns sons impulsivos em função do tempo.

Figura 1 – Exemplos de sons impulsivos



(a) Latido; (b) Garrafa quebrando; (c) Tiro de Fuzil; (d) Porta batendo.

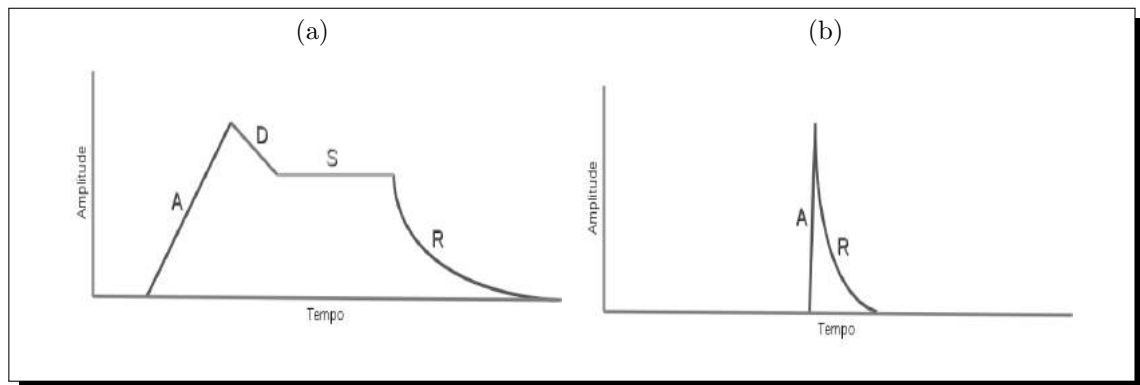
Fonte: o autor

Para um melhor entendimento do que seria um som impulsivo, toma-se por base o conceito ADSR (*Attack, Decay, Sustain, Release*) da música computadorizada [8].

Conforme observa-se na figura 2a, a fase do Ataque (*Attack*) demarca o crescimento e liberação da potência sonora. O decaimento (*Decay*) é representado por uma pequena perda da potência sonora inicial. A seguir a onda sonora passa por um processo de Sustentação (*Sustain*) pelo meio material que rodeia a fonte de emissão da onda sonora, como, por exemplo, uma caixa de ressonância. Finalmente a vibração sonora perde energia até extinguir-se, caracterizando a fase de Relaxamento (*Release*).

Desta forma, caracteriza-se um som impulsivo como sendo aquele onde as fases do Decaimento e Sustentação praticamente não existem, conforme ilustrado na figura 2b, e as fases do Ataque e Relaxamento ocorrem muito rapidamente, possuindo tempos da ordem de décimos de segundos ou menos.

Figura 2 – Esquema ADSR



(a) Esquema ADSR de um som qualquer; (b) Esquema ADSR de um som impulsivo.

Fonte: o autor

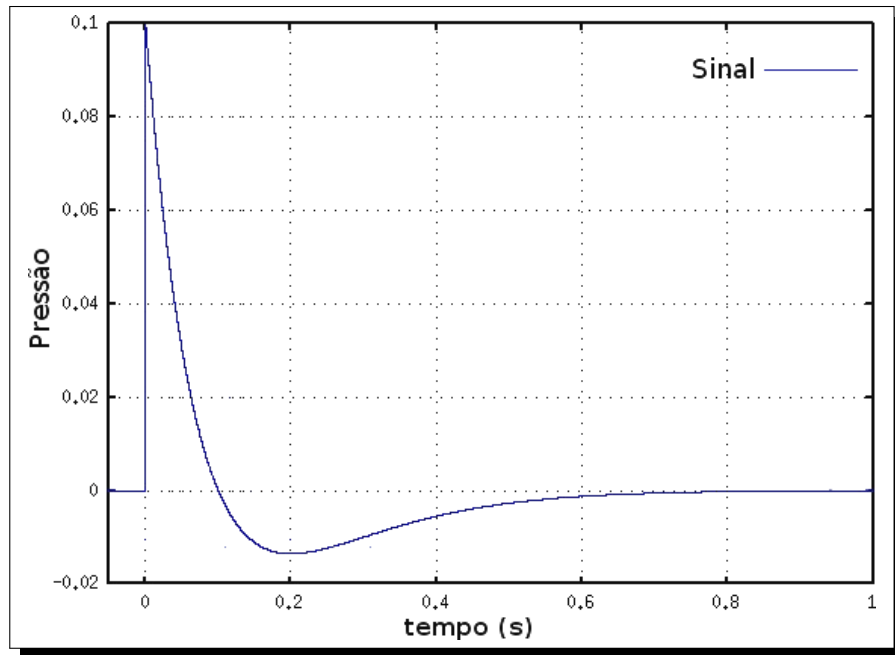
Assim, um som impulsivo ideal possuirá um tempo de ataque muito curto, atingirá o seu máximo de energia muito rapidamente e não se manterá neste estado de alta energia por muito tempo, iniciando o seu decaimento quase que instantaneamente, esvaindo-se rápida e exponencialmente. Sua representação ideal pode ser obtida através da equação de explosão de Friedlander [9], cujo gráfico pode ser visto na figura 3

$$p(t) = \begin{cases} 0 & t < 0 \\ P_0 + P_s \left(1 - \frac{t}{t_{pos}}\right) e^{\frac{-bt}{t_{pos}}} & t \geq 0 \end{cases} \quad (2.1)$$

onde

- P_0 – é a pressão inicial;
- P_s – é o pico de pressão;
- t – é o tempo dado em segundos;
- t_{pos} – é a duração da fase positiva; e

Figura 3 – Explosão de Friedlander



Fonte: o autor

– b – é a razão de decaimento.

Uma representação audível de um som impulsivo simples pode ser conseguida pela equação

$$\mathcal{F}(t) = \begin{cases} 0 & t < t_0 \\ e^{D \cdot (t-t_0)} \cdot \sin(2\pi \cdot f \cdot (t - t_0)) & t \geq t_0 \end{cases} \quad (2.2)$$

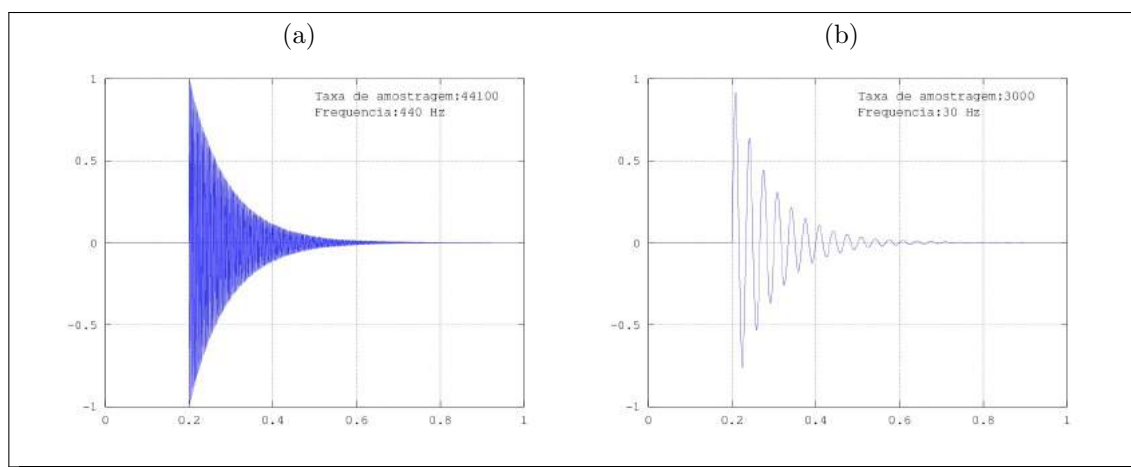
onde:

- f – representa a frequência desejada para o som, em Hz;
- t_0 – define o atraso, em segundos, para o início do som a partir da origem;
- $\sin(2\pi \cdot f \cdot (t - t_0))$ – gerará uma onda senoidal de frequência igual a f e amplitude 1;
- $e^{D(x-t)}$ – irá provocar o decaimento exponencial da onda, e;
- D – é uma constante, responsável por fazer o sinal decair em dado intervalo de tempo. O tempo de decaimento será de aproximadamente $\delta t = 2eD$ para um decaimento de sinal a níveis de potência na casa dos 10^{-5} .

O resultado para uma frequência de 440 Hz pode ser visto na figura 4a.

Para uma melhor visualização das componentes do sinal, reduziu-se a frequência da onda descrita acima para 30 Hz obtendo-se o gráfico exibido na figura 4b.

Figura 4 – Sons Impulsivos Gerados no Octave

(a) Tom Lá Impulsivo; (b) Som Impulsivo com frequência de 30 Hz .

Fonte: o autor

2.2 Disparos de Armas de Fogo

Todo disparo de arma de fogo é um som impulsivo. Porém este apresenta características exclusivas que o diferenciam dos demais sons impulsivos, pois é composto por uma somatória de sinais sonoros, a saber [9, 10]:

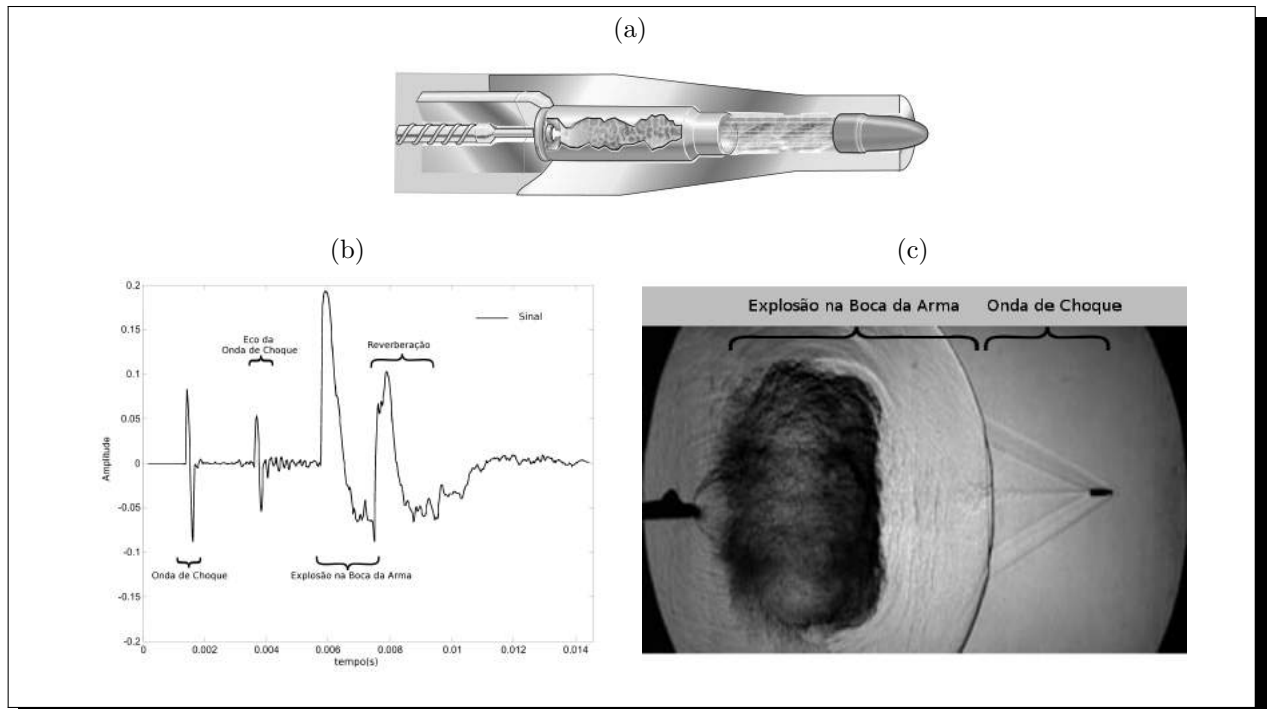
Deflagração do propelente – esta fase se dá dentro da câmara da arma provocando um estampido inicial audível somente quando muito próximo ao local de disparo, conforme ilustrado na figura 5a.

Onda de choque – também conhecida como *N wave*, devido ao seu formato semelhante à letra “N” (figura 5b), a onda de choque é resultado da quebra da barreira do som realizada pelo projétil durante o seu deslocamento, conforme pode ser observado na figura 5c. Sua existência depende do tipo de armamento, já que o projétil tem que ser propelido a velocidades supersônicas para que ela exista e sua detecção depende da posição do observador em relação ao disparo [9].

Explosão na boca da arma – também conhecida como *Muzzle Blast*, esta explosão é causada pela rápida expansão dos gases provenientes da queima do propelente, que ocorre na boca da arma imediatamente após o projétil deixar o cano, conforme visto na figura 5c. Este é o principal componente de um som de disparo, podendo ser ouvido a grandes distâncias.

Reverberações – são todos os ecos dos sons acima descritos resultantes de reflexões do som no solo e em outras superfícies que compõem o ambiente (figura 5b). São o som dominante no sinal sonoro alguns milissegundos após a explosão inicial [9, 10].

Figura 5 – Componentes de um disparo de arma de fogo.



(a) Deflagração do Propelente [11]; (b) Sinais Componentes de um Disparo [10]; (c) Visualização da onda de choque e explosão da boca da arma [12].¹

2.3 Conceitos básicos sobre armas de fogo

Em virtude da nomenclatura empregada na utilização de armas de fogo, faz-se necessária a descrição de alguns conceitos e definições básicos para um correto entendimento dos capítulos posteriores, a começar pela sua própria definição:

Arma que arremessa projéteis empregando a força expansiva dos gases gerados pela combustão de um propelente confinado em uma câmara que, normalmente, está solidária a um cano que tem a função de propiciar continuidade à combustão do propelente, além de direção e estabilidade ao projétil[13].

2.3.1 Quanto ao Transporte

Uma arma é classificada quanto ao seu tamanho e conseqüente forma de transporte como sendo [13]:

¹ As figuras originais foram editadas e modificadas pelo autor.

De Porte: possui dimensões e peso reduzidos que possibilitam o seu transporte em um coldre, podendo ser disparada utilizando-se somente uma das mãos. São exemplos de armas de porte as pistolas, revólveres e garruchas;

Portáteis: possui peso e dimensões que permitam a um único homem transportá-la exigindo ambas as mãos para a realização de um disparo. São exemplos os fuzis, metralhadoras de mão, carabinas e armas de caça;

Não-portáteis: Armas que não podem ser transportadas por um único homem, sendo necessária uma equipe para operá-la. São exemplos deste tipo de armamento os canhões, obuseiros, morteiros, metralhadoras pesadas, dentre outros.

Por ser caracterizada como uma arma de fogo pesado e pouco comum em um ambiente urbano normal, as armas não-portáteis permanecerão fora do escopo deste trabalho.

2.3.2 Fases do Funcionamento

A fim de permitir um melhor entendimento dos mecanismos básicos de operação e de permitir uma correta classificação das armas de fogo, descreve-se a seguir as principais fases do seu funcionamento:

Alimentação: Ato de fornecer munição ao armamento, nela inserindo um carregador ou fita ou ainda inserindo manualmente a munição em compartimento específico da arma destinado a este fim, sem, no entanto inserir a munição dentro da câmara de disparo;

Carregamento: fase na qual um cartucho de munição é inserido na câmara de disparo, deixando-o pronto para utilização;

Trancamento: ação na qual a arma é preparada para o disparo, através do trancamento mecânico de suas partes móveis, de forma a impedir que o estojo da munição seja propulso para trás, diminuindo a eficácia do disparo do projétil para a frente;

Acionamento do gatilho: ato do atirador que irá desencadear o disparo propriamente dito;

Disparo: ação na qual a munição é deflagrada pela queima do propelente e o projétil é efetivamente disparado pelo cano da arma;

Abertura: é a fase na qual a arma é destrancada e aberta de forma que a câmara fique acessível para a introdução de um novo cartucho;

Ejeção: fase na qual o estojo vazio, recém disparado, é arrancado de dentro da câmara para ser, em seguida, descartado.

Note-se, porém que, dependendo do tipo da arma, nem todas as fases do funcionamento estarão presentes.

2.3.3 Classificação quanto ao Funcionamento

Uma arma de fogo é classificada quanto ao seu funcionamento em três categorias distintas [13]:

Arma de Repetição: arma na qual o atirador necessita empregar força muscular para realizar as fases de carregamento, trancamento, abertura e ejeção a cada disparo que for realizar;

Arma Semiautomática: arma que realiza de forma automática todas as fases de seu funcionamento, exceto o acionamento do gatilho. Para tanto ela pode empregar meios elétricos ou aproveitamento de parte dos gases gerados pela deflagração do cartucho como força motriz para abrir, ejetar, carregar e trancar a arma, ficando a espera de um novo acionamento do gatilho para realizar o disparo e recomeçar o ciclo;

Arma Automática: é a arma que realiza o carregamento, trancamento, abertura, ejeção e disparo de forma contínua, enquanto o gatilho da arma for mantido pressionado, numa forma de disparo conhecida como rajada.

2.3.4 Cadência de Tiro

A cadência de tiro encontra-se intimamente relacionada com o tipo de funcionamento da arma de fogo, sendo definida como a quantidade de disparos que tal arma consegue realizar em um minuto, expressa em **tiros por minuto (tpm)** [14].

Embora não haja uma padronização que inter-relacione o funcionamento da arma com sua cadência, para este trabalho considerar-se-á as seguintes cadências de tiro:

Tabela 1 – Tipo de Funcionamento × Cadência de Tiro

Funcionamento	Mínimo	Máximo
<i>Repetição</i>	12 tpm	60 tpm
<i>Semiautomática</i>	60 tpm	120 tpm
<i>Automática</i>	120 tpm	600 tpm

Fonte: o autor

Existem armas pesadas que possuem cadência de tiro superior a 600 tpm, mas estas fogem ao escopo deste trabalho.

2.3.5 Nomenclatura

Quando observado o comprimento e seu funcionamento, uma arma de fogo passa a receber nomes distintos a fim de caracterizá-la. Desta forma temos, dentre as armas mais encontradas nas ruas, os seguintes tipos de armamento [14] [13]:

Pistola – arma de porte geralmente semiautomática, com uma única câmara acoplada ao cano (figura 6a).

Revolver – arma de porte, de repetição, dotada de um cilindro giratório que lhe serve de carregador e câmara (figura 6b).

Rifle – arma portátil, que por definição, significa o mesmo que fuzil. Porém a palavra é costumeiramente utilizada para designar um rifle de repetição, no qual o acionamento do ferrolho é realizado manualmente, por intermédio de uma alavanca (figura 6c) .

Fuzil – arma de fogo portátil, com cano longo que, ao contrário das espingardas de caça, possui uma alma raiada, a fim de imprimir rotação ao projétil, garantindo a precisão do tiro por efeito giroscópico. Modernamente, todos os fuzis são armas semiautomáticas (figura 6d).

Alguns fuzis podem realizar também o tiro automático, possuindo para isto uma chave seletora. Porém, por se tratar de um armamento leve, normalmente este regime de tiro deve ser utilizado apenas em rajadas curtas, de 3 ou 4 tiros, sob o risco um possível derretimento do cano.

Existem muitos outros tipos de armamento com nomes e características próprias, porém toda esta conceituação foge aos objetivos do presente trabalho.

2.3.6 Munições

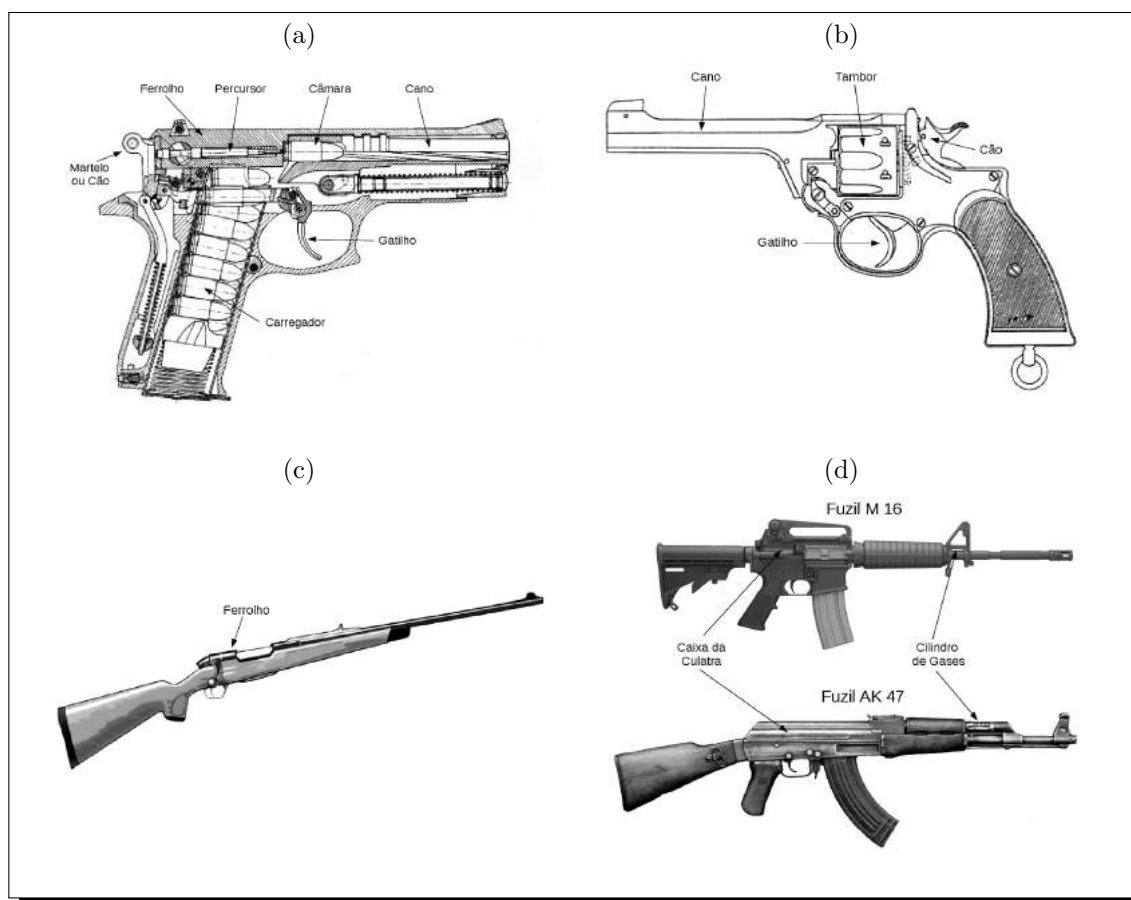
A munição é o artefato completo, pronto para carregamento e disparo de uma arma. É dela que se origina todo o som impulsivo produzido pela arma, sendo pois, importante conhecer-se um pouco a seu respeito.

O Cartucho de Munição

Conjunto composto por estojo, espoleta, propelente e projétil e que compõe a munição (figura 7a):

Estojo: capsula cilíndrica, normalmente feita em latão, plástico ou papelão muitas vezes com o formato de uma pequena garrafa na qual irão se alojar a espoleta e o projétil, sendo preenchida pelo propelente;

Figura 6 – Tipos de Armamento.

(a) Pistola [15]; (b) Revolver [16]; (c) Rifle [17]; (d) Fuzis [18, 19].²

Espoleta: pequena capsula, normalmente fixada na base do estojo, que contém uma substância iniciadora do tipo explosivo, sensível à ação mecânica, que irá inflamar-se quando percutida, inflamando também o propelente;

Propelente: explosivo lento, de queima progressiva, destinado a criar a massa de gases que irão propelir o projétil através do cano da arma. Normalmente composto por algum tipo de pólvora de base simples, dupla ou tripla, a base de nitrocelulose, nitroglicerina e nitroguanina;

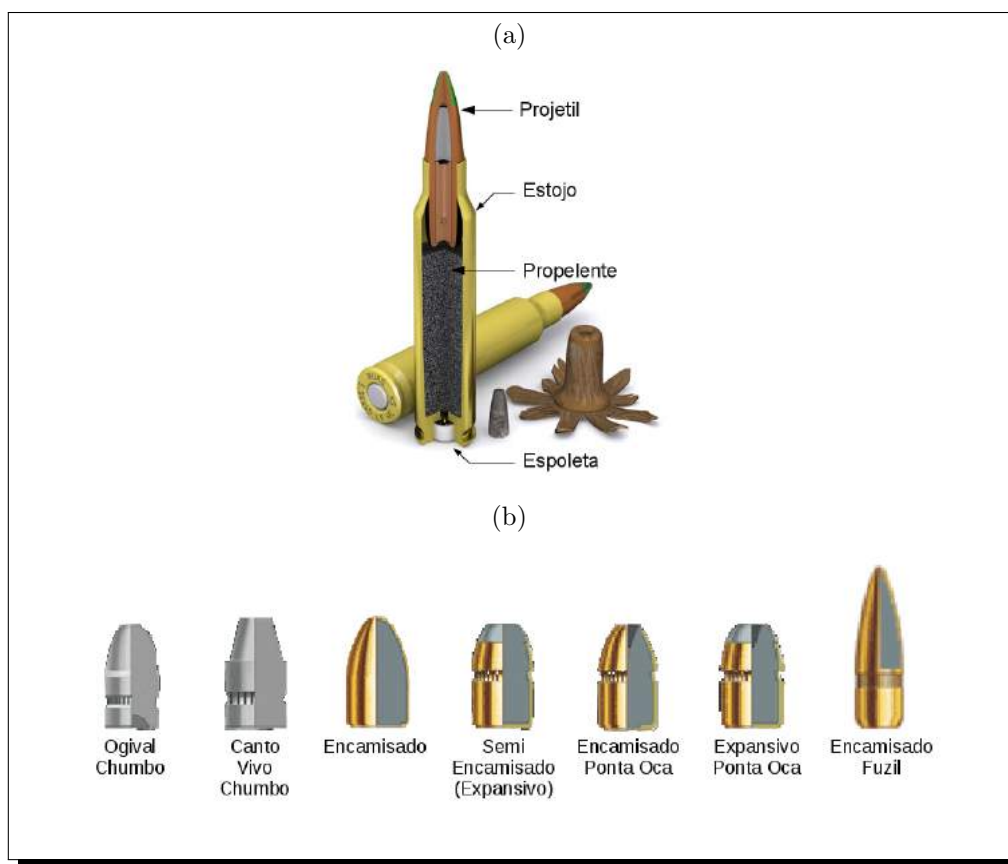
Projétil: peça de metal que é projetada pelo cano da arma em direção ao alvo.

2.3.7 O Calibre da Arma

Calibre pode ser definido como a medida do diâmetro externo de um projétil, que é usada para definir ou caracterizar um tipo de munição ou de arma [13]. Teoricamente uma

² As figuras originais foram editadas e modificadas pelo autor.

Figura 7 – Cartucho e Projéteis

(a) Cartucho de munição [20]; (b) Tipos de projéteis [21].³

arma pode ser criada com qualquer calibre desejado. Porém, com o passar do tempo certos calibres foram consagrados pelo uso e tornaram-se padrão da indústria armamentista.

Como medida, o calibre é definido por um número, que pode ser interpretado de três formas diferentes, a saber:

Sistema de Centésimos de Polegada: popularizado pela indústria norte-americana, define o calibre por dois dígitos. São exemplos desta notação os calibres 45, 38, 32, 22, que representam as dimensões de 0.45”, 0.38”, 0.32” e 0.22”, respectivamente. Muitas vezes tais números são precedido da palavra *ponto*, como em “*ponto 40*”. Uma exceção à esta regra de notação pode ser vista no calibre 380, que, apesar de ter três dígitos, representa 0.380”, calibre também conhecido como 9mm Kurz, Curto ou Short.

Sistema Milimétrico: sistema originário da Europa, representa o calibre da arma medido em milímetros e suas subdivisões. São exemplos desta notação os calibres 5.56, 6.35 Browning, 7.62, 9mm, dentre outros.

³ As figuras originais foram editadas e modificadas pelo autor.

Sistema Gauge: originária da Inglaterra, este sistema de medida de calibre inicialmente identificava o calibre de canhões pelo peso do projétil em libras. Assim o calibre 1 referia-se a um projétil de 1 libra, 2 a um projétil de duas libras e assim por diante. Atualmente é usado para designar calibres de armas de caça e usa como referencia a quantidade de esferas necessárias a compor uma libra. Desta forma uma espingarda calibre doze tem o cano com diâmetro equivalente a uma esfera de chumbo cujo peso é igual a um doze avos de libra ($\cong 20.60mm$). Note-se que neste sistema, quanto maior o número, menor será o calibre.

2.4 Métodos de Detecção de Sons Impulsivos

A primeira etapa a ser realizada no processamento do sinal capturado destina-se a identificar e extrair do som ambiente qualquer som impulsivo que nele ocorra. Devido à massiva quantidade de dados envolvida, faz-se necessário a utilização de um algoritmo de detecção de sons impulsivos que seja simples, eficiente e que exija poucos recursos computacionais.

Desta forma foram elencados três métodos de detecção de sons impulsivos que atendem a estas características, a saber:

- Método da Quarta Derivada do RMS, proposto como ferramenta para detecção de ruídos em gravações de áudio [22];
- Método da Mediana Condicional, amplamente utilizado na detecção de ruído em imagens e sons [23, 24, 25, 26], bem como em telecomunicações [27] ; e
- Método da Variância, método relativamente novo, proposto em [7].

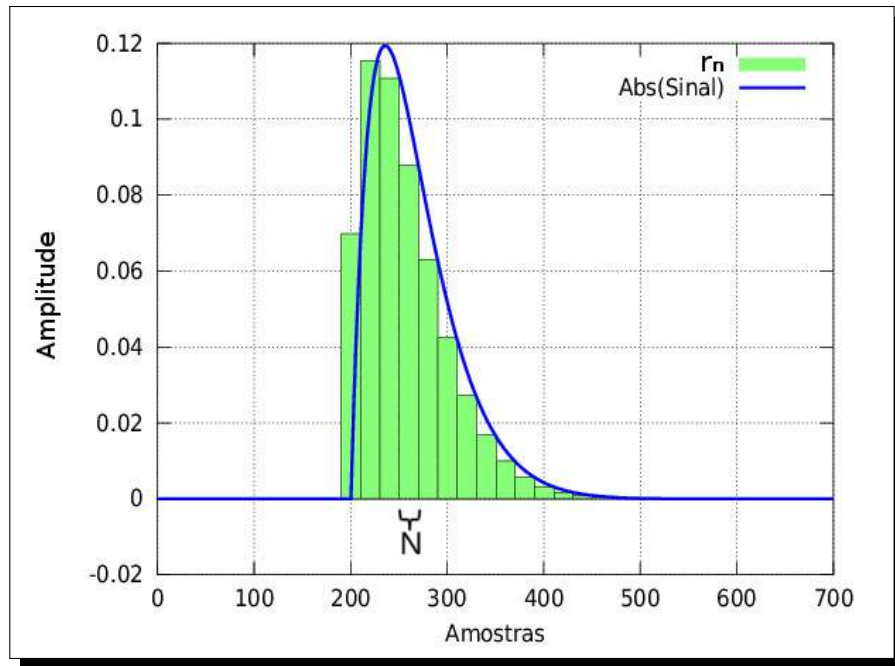
2.4.1 Método da Quarta Derivada do RMS

O *Root Mean Square (RMS)* , Valor Quadrático Médio ou Valor Eficaz é uma medida estatística de magnitude do sinal. Ela pode ser obtida pela equação 2.3 [22]

$$r_n = \sqrt{\frac{1}{N} \sum_{n=1}^N x^2(n)} \quad (2.3)$$

onde N é o número de amostras consideradas para o cálculo do *RMS*, x é o valor de cada amostra do sinal digitalizado e r_n é o valor do RMS. O resultado da transformação da potência em RMS pode ser visto na figura 8

Em um som impulsivo a potência do sinal sobe abruptamente e isto reflete-se também no RMS que aumenta proporcionalmente. Assim, para que se possa detectar uma variação abrupta no sinal, toma-se o valor da derivada do gráfico de RMS para um dado instante.

Figura 8 – Exemplo de RMS com $N = 20$ amostras sobre um sinal impulsivo.

Fonte: o autor

Entretanto, o uso da primeira derivada pode provocar uma série de falsos positivos, já que esta irá considerar apenas dois valores contíguos de RMS [22]. Para evitar estas falsas detecções, pode-se aumentar o número de amostras RMS consideradas, passando a utilizar a quarta derivada discreta, obtida das Séries de Taylor, vista na equação 2.4.

$$f^{(4)}(r_n) = \frac{r_{n-2} - 4r_{n-1} + 6r_n - 4r_{n+1} + r_{n+2}}{(t_{n+2} - t_{n-2})^4} \quad (2.4)$$

onde r_n é o RMS e t_n o tempo correspondente [28].

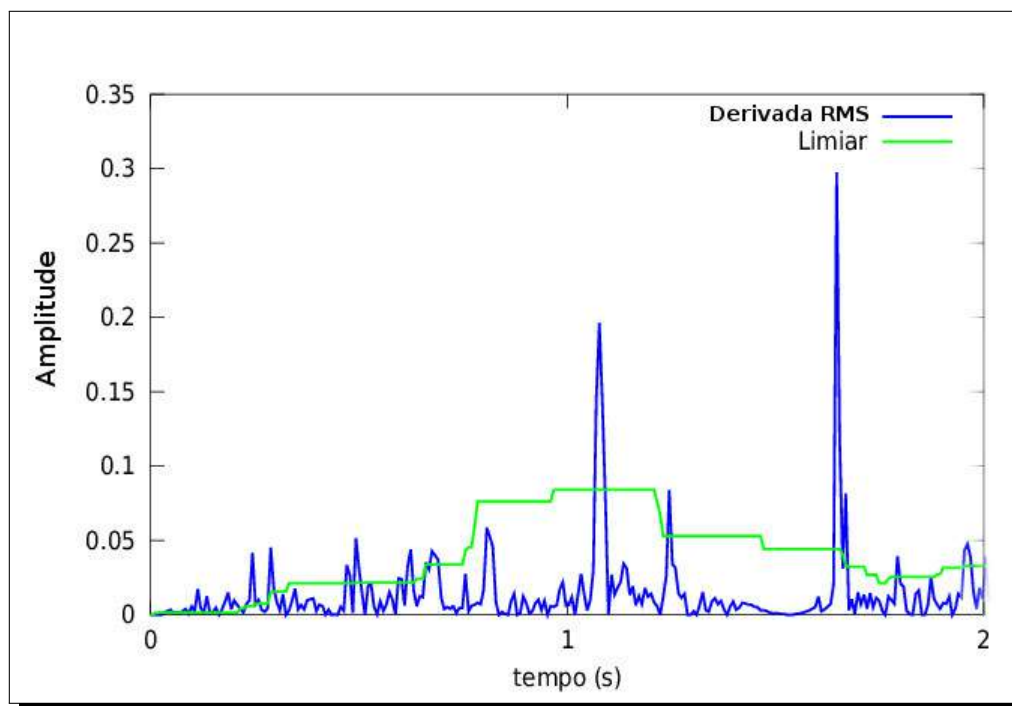
Desta forma leva-se em consideração cinco amostras subsequentes de RMS, conseguindo-se uma maior consistência na detecção impulsos, detectando tendências reais no aumento da potência de sinal. Entretanto, o método ainda utiliza um pequeno número de amostras e um elevado número de falsos positivos pode ser esperado.

Uma vez obtido o valor da quarta derivada do RMS, este é comparado com um valor padrão de limiar e sempre que este valor é ultrapassado considera-se que um impulso foi detectado. Este limiar pode ser fixo, mas, virtude da natureza dinâmica do ambiente sonoro, melhores resultados podem ser obtidos caso este limiar leve em consideração a variação do ruído ambiente[22]. Este limiar adaptativo pode ser obtido levando-se em conta a média da potência do sinal em torno da amostra analisada, utilizando-se este valor como limiar, conforme a equação 2.5,

$$th_{(n)} = \frac{k}{2i + 1} \sum_{m=n-i}^{n+i} r_m \quad (2.5)$$

onde th é o valor do limiar para o intervalo, r_m os valores do RMS considerados, k um fator de multiplicação tomado experimentalmente e $2i + 1$ o número de amostras consideradas [22]. Um exemplo do gráfico de detecção aplicado a um sinal impulsivo pode ser visto na figura 9.

Figura 9 – Exemplo de detecção utilizando-se a 4ª derivada do RMS e limiar ajustável.



Fonte: o autor

Na figura 10 vê-se um exemplo do processo de detecção RMS. A figura 10a mostra um sinal contendo som ambiente, onde foi inserido um som impulsivo na posição assinalada pela seta. A figura 10b representa a comparação realizada entre os valores de RMS e o limiar adaptativo. No exemplo em questão, o método assinalou a ocorrência de 33 sons impulsivos.

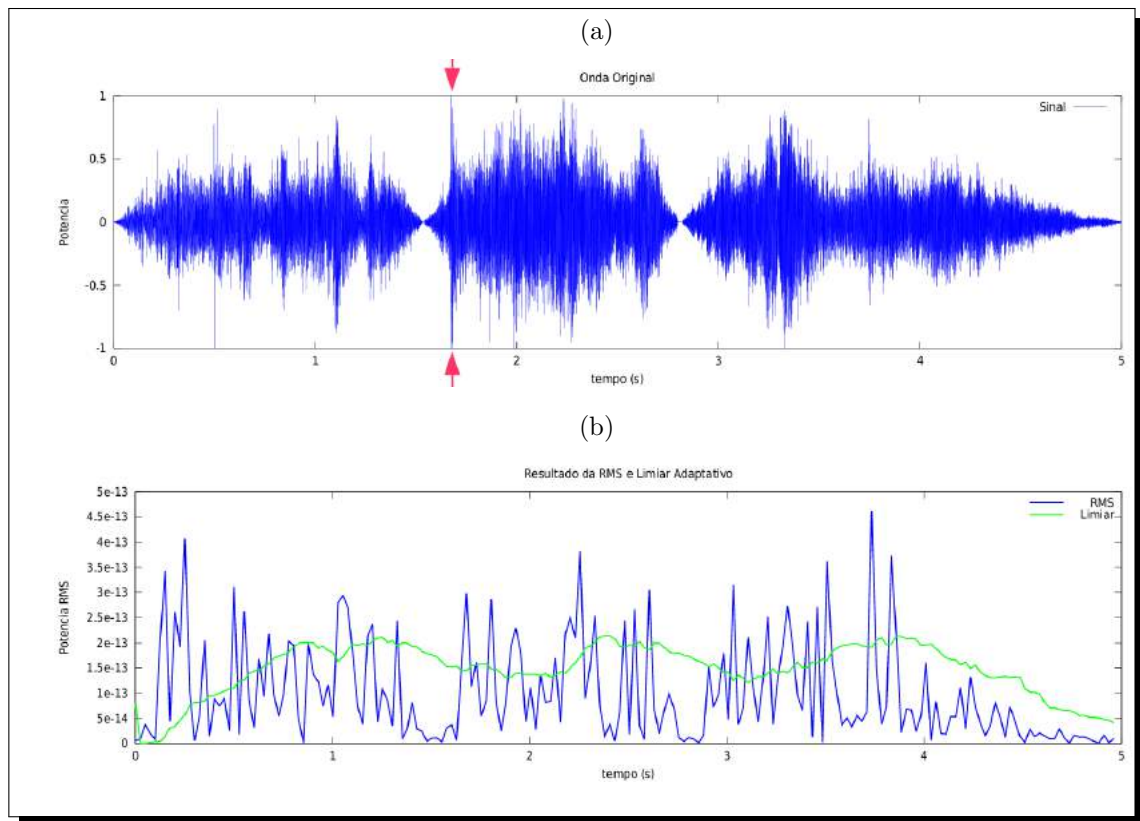
2.4.2 Método do Filtro Condicional de Mediana (CMF)

A *Conditional Median Filter (CMF)* é um método tradicional de detecção de ruídos em sinais sonoros e imagens [25, 24, 26, 23], assim como na área telecomunicações [27], sendo também eficiente na detecção de sons impulsivos [7].

O método consiste em estimar-se o valor da potência $e(k)$ para um determinado número de amostras sucessivas N de um dado sinal de áudio $x(n)$, cujo valor passa a se chamar *frame*, calculado conforme a equação 2.6 [27].

$$e(k) = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n) \quad (2.6)$$

Figura 10 – Exemplo do processo de detecção utilizando-se a RMS.



(a) Sinal Original; (b) RMS comparada ao limiar adaptativo.

Fonte: o autor

Os valores calculados são armazenados em uma janela contendo L valores de $e(k)$, conforme pode ser visto na figura 11.

Com base no cálculo da potência, calcula-se a CMF para L $e(k)$ consecutivos, conforme a equação 2.7.

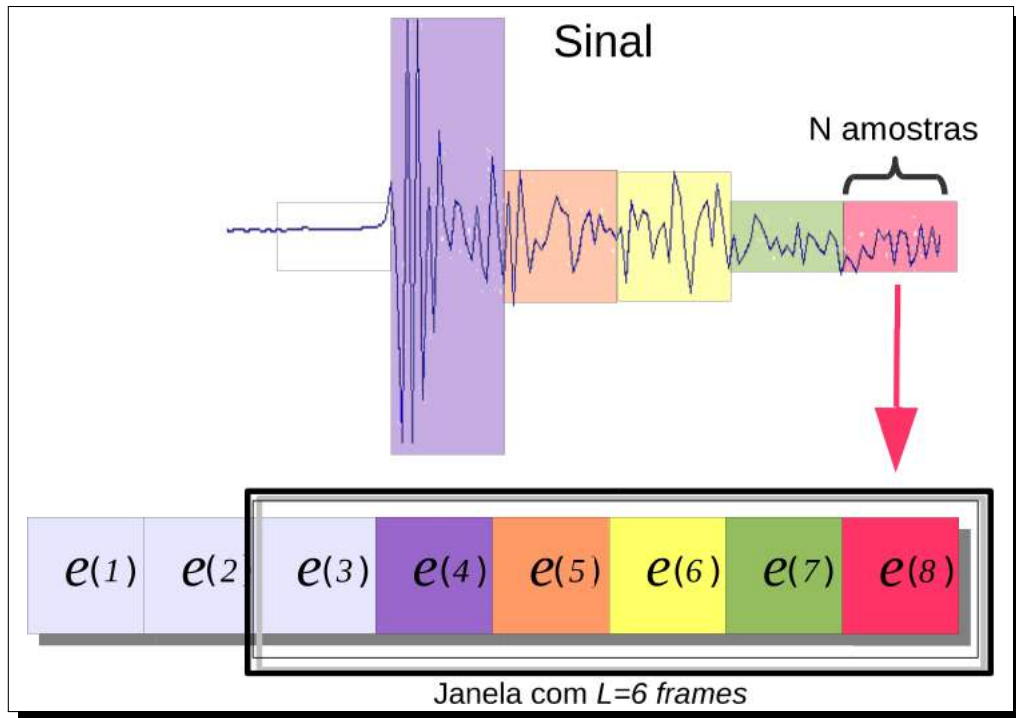
$$M = \{e(k - L + 1) \dots e(k)\}$$

$$mf(k) = \tilde{M}$$
(2.7)

$$CMF(k) = \begin{cases} mf(k) & \text{se } |e(k) - mf(k)| > th \\ e(k) & \text{Caso contrário} \end{cases}$$

onde, L é o número de frames na janela deslizante, mf é o valor da mediana para a janela e th é um valor de limiar de ruído. O método é chamado de “Filtro Condicional de Mediana”, porque o valor da mediana somente será utilizado caso $|e(k) - mf(k)| > th$. Caso contrário será utilizado com valor de CMF o próprio valor de potência do *frame*.

A seguir subtrai-se o valor da CMF do valor de $e(k)$, conforme a equação 2.8. Desta forma obtém-se o coeficiente de detecção $q(k)$ que será sempre zero, exceto quando

Figura 11 – Exemplo de uma janela contendo $L = 6$ frames $e(k)$.

Fonte: o autor

$|e(k) - mf(k)| > th$, ou seja, quando um som impulsivo for detectado.

$$q(k) = e(k) - CMF(k) \quad (2.8)$$

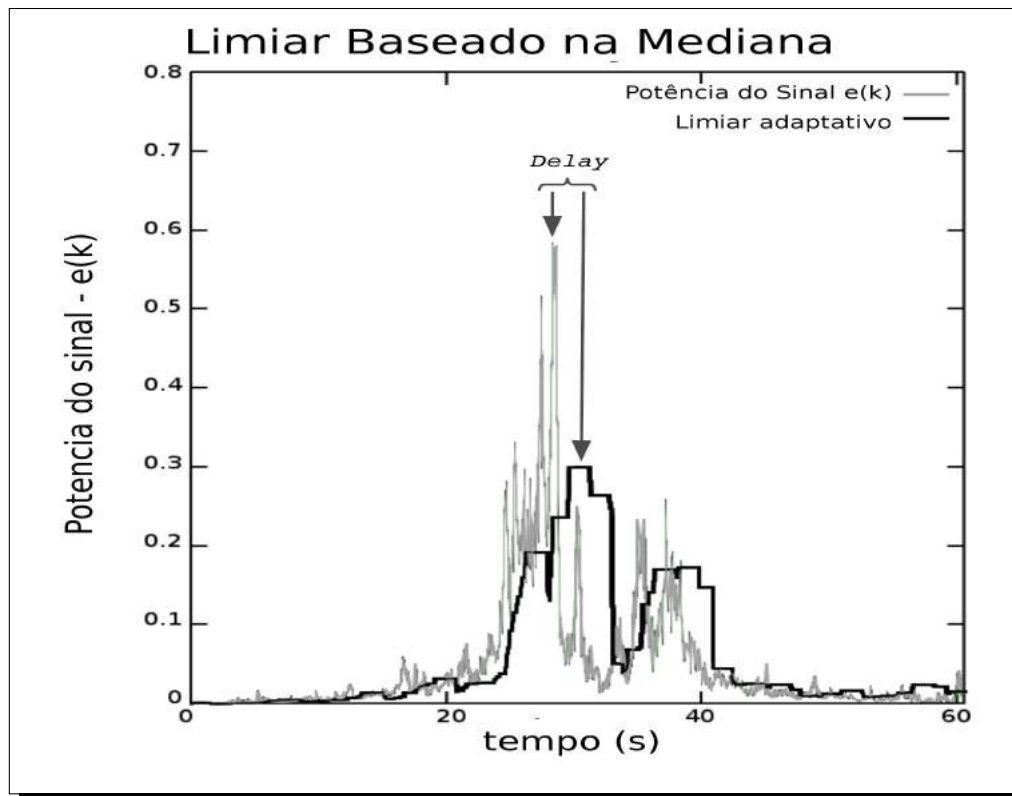
Uma detecção mais apurada pode ser obtida utilizando-se um valor de limiar adaptativo th , que leve em consideração o ruído de fundo. Isto pode ser obtido calculando-se uma segunda CMF detentora de uma janela que abranja uma quantidade maior de frames, de forma a estimar o ruído de fundo minimizando as variações causadas por outros sons impulsivos ocorridos no passado, conforme a equação 2.9

$$th(k) = A \times \max_{i=k-L_2}^k \{CMF_2(i) - mf_2(i)\} \quad (2.9)$$

onde L_2 é o tamanho da janela para CMF_2 , mf_2 a mediana da janela L_2 e A é um multiplicador destinado a manter o limiar 50 % maior que o ruído de fundo [7]. Esta janela L_2 assumirá valores de 3 a 5 vezes maiores que L . Quanto maior a janela L_2 menor será a influencia das variações abruptas ocorridas no ruído de fundo sobre o limiar, porém haverá um maior atraso na estimativa real do ruído de fundo, o que pode resultar numa maior detecção de falsos positivos. Na figura 12 pode-se observar o atraso existente entre a variação do ruído de fundo no sinal e a sua influência na variação do limiar, caracterizado pelo *delay* assinalado.

A figura 13 exemplifica o processo de detecção, onde se vê na figura 13a o mesmo exemplo de som ambiente analisado pela RMS, submetido agora ao método da CMF.

Figura 12 – Atraso do Limiar Adaptativo



Fonte: o autor

A figura 13b representa a comparação realizada entre os valores de $|e(k) - mf(k)|$ e o limiar adaptativo. Já a figura 13c mostra o gráfico dos valores de $q(k)$, no qual $q(k) \neq 0$ representa uma detecção de som impulsivo pelo algoritmo. No exemplo em questão, o método assinalou 26 ocorrências de sons impulsivos.

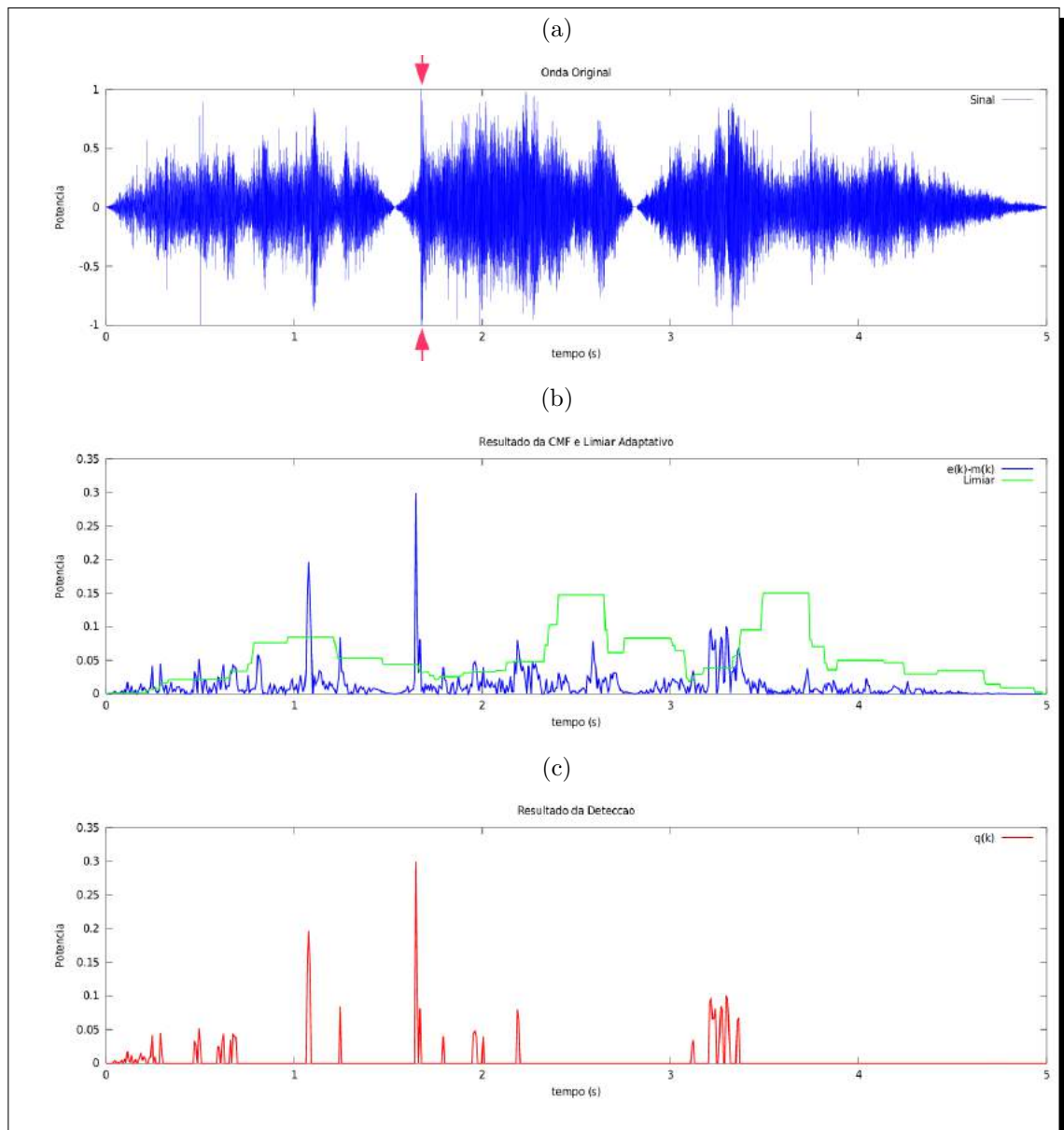
2.4.3 Método da Variância (VM)

De forma similar ao método da CMF, a VM começa com o cálculo da potência do sinal $e(j+k)$ para N amostras, populando uma janela j contendo L frames. O valor j indica a posição absoluta do primeiro frame da janela atual em relação ao sinal como um todo e dará nome a esta janela.

O valor de L é importante na sensibilidade do algoritmo *i.e.* quanto maior o valor de L menor será a sua sensibilidade, podendo haver considerável perda na detecção de impulsos. Por outro lado um valor de L pequeno causará um grande aumento no número de falsos positivos detectados. Os valores usualmente aplicados são $L = 30$ para um frame com $\Delta t = 90$ ms conforme descrito por [7, 29].

A seguir os valores dos frames $e(k)$ contidos na janela são normalizados para valores entre 0 e 1, obtendo-se, para cada um deles o correspondente valor de $e_n(k)$, conforme a

Figura 13 – Exemplo das fases componentes do processo de detecção utilizando-se a CMF.



(a) Sinal Original; (b) CMF comparada com o limiar adaptativo; (c) Gráfico de $q(k)$, onde $q(k) \neq 0$ indica a ocorrência de um som impulsivo.

Fonte: o autor

equação 2.10

$$e_n(j+k) = \frac{e^{(j+k)} - e_{min}}{e_{max} - e_{min}} \quad k = 1, 2, \dots, L \quad (2.10)$$

Onde e_{min} e e_{max} representam o menor e o maior valor de potência encontrado na janela j .

Calcula-se, então o valor da variância σ^2 da janela j pela equação 2.11.

$$\sigma^2(j) = \frac{1}{L-1} \sum_{k=1}^{L-1} [e_n(j+k) - \bar{e}_n(j)]^2$$

onde

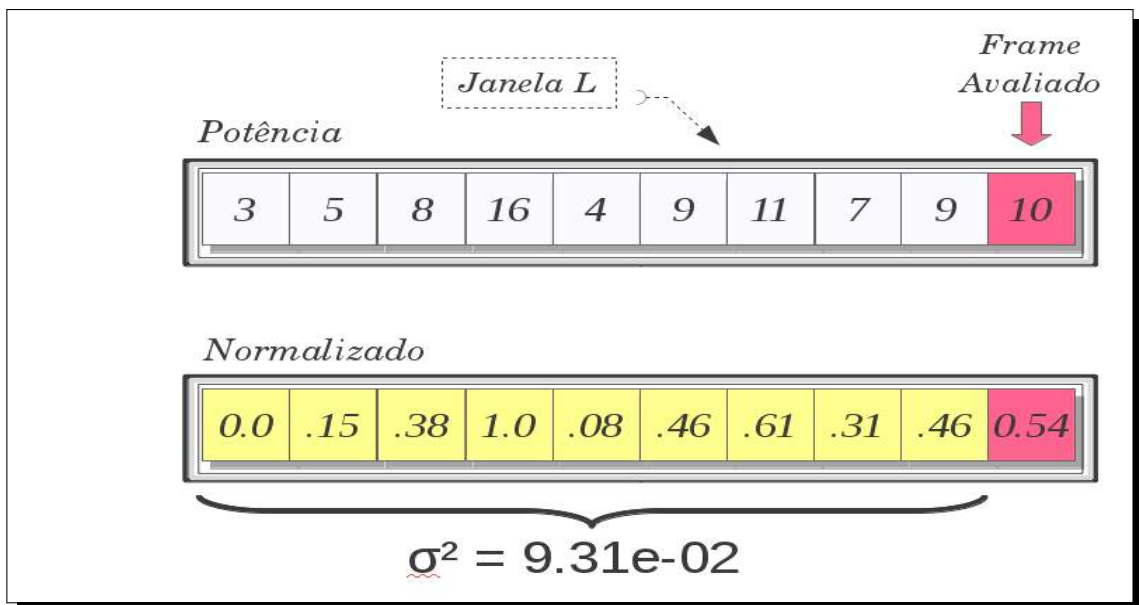
$$\bar{e}_n(j) = \frac{1}{L-1} \sum_{k=1}^{L-1} e_n(j+k)$$
(2.11)

Finalmente o valor da variância é comparado a um limiar fixo $th = 0.0225$, predefinido na literatura, e qualquer valor de variância abaixo deste th indicará a detecção de um impulso em $e(j+L)$ [7]. Isto ocorre porque o valor contido no último frame da janela j é considerado para a normalização dos valores desta janela, mas não é considerado para o cálculo da variância do intervalo. Desta forma, caso o último frame possua valores próximo aos demais, os valores normalizados serão bastante elevados e próximos entre si, mantendo a variância também elevada.

Ao final do processo j é incrementado de 1, deslocando toda a janela um frame para a direita na linha de tempo do sinal.

Na figura 14 vê-se um exemplo onde o valor do *frame* avaliado é bastante próximo dos demais, indicando que a potência sonora do ambiente não sofreu nenhuma variação abrupta. Realizada a normalização e cálculo da variância dos $L - 1$ *frames*, esta última permanece elevada por conta dos elevados valores contidos na janela.

Figura 14 – Janela da variância ($L = 10$) onde nenhum frame possui som impulsivo

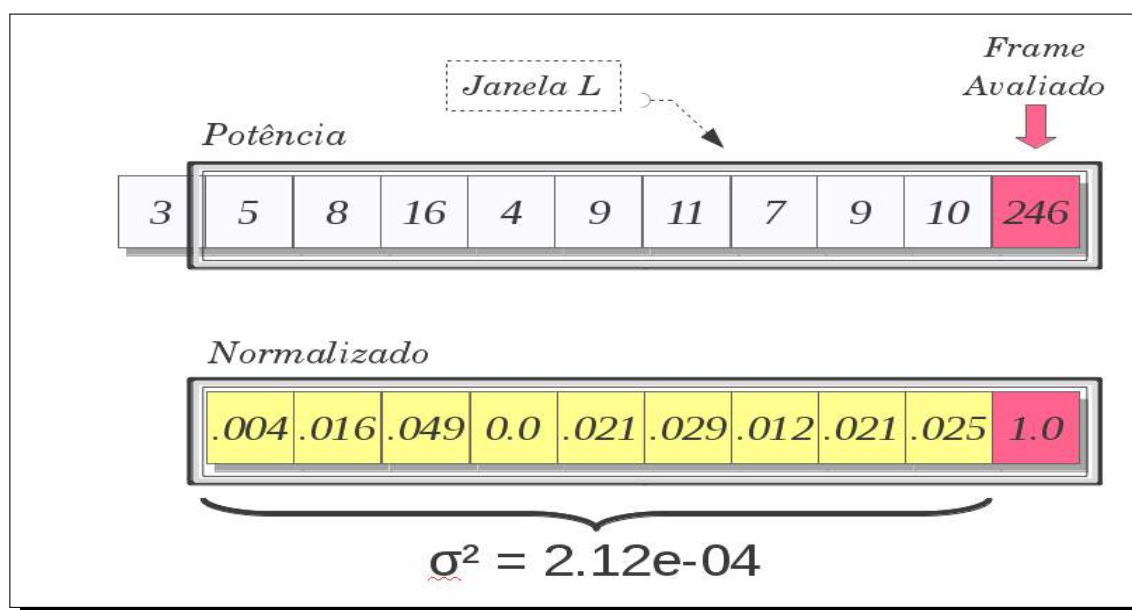


Fonte: o autor

Caso um impulso seja inserido no último *frame*, este assumirá o valor 1 quando normalizado e os demais assumirão valores próximos a zero devido à grande diferença de potência entre eles. Ao ser calculada a variância dos $L - 1$ *frames* obtêm-se um valor significativamente menor.

Isto pode ser visto na figura 15 onde o valor do *frame* considerado possui potência cerca de 27 vezes mais elevada que a média dos demais *frames* da janela. Isto indica um aumento súbito na potência sonora do ambiente, indício de que ocorreu um som impulsivo. Após a normalização e cálculo da variância nota-se que este último valor é significativamente menor que o exibido na figura 14.

Figura 15 – Janela da variância ($L = 10$) onde o último frame possui som impulsivo



Fonte: o autor

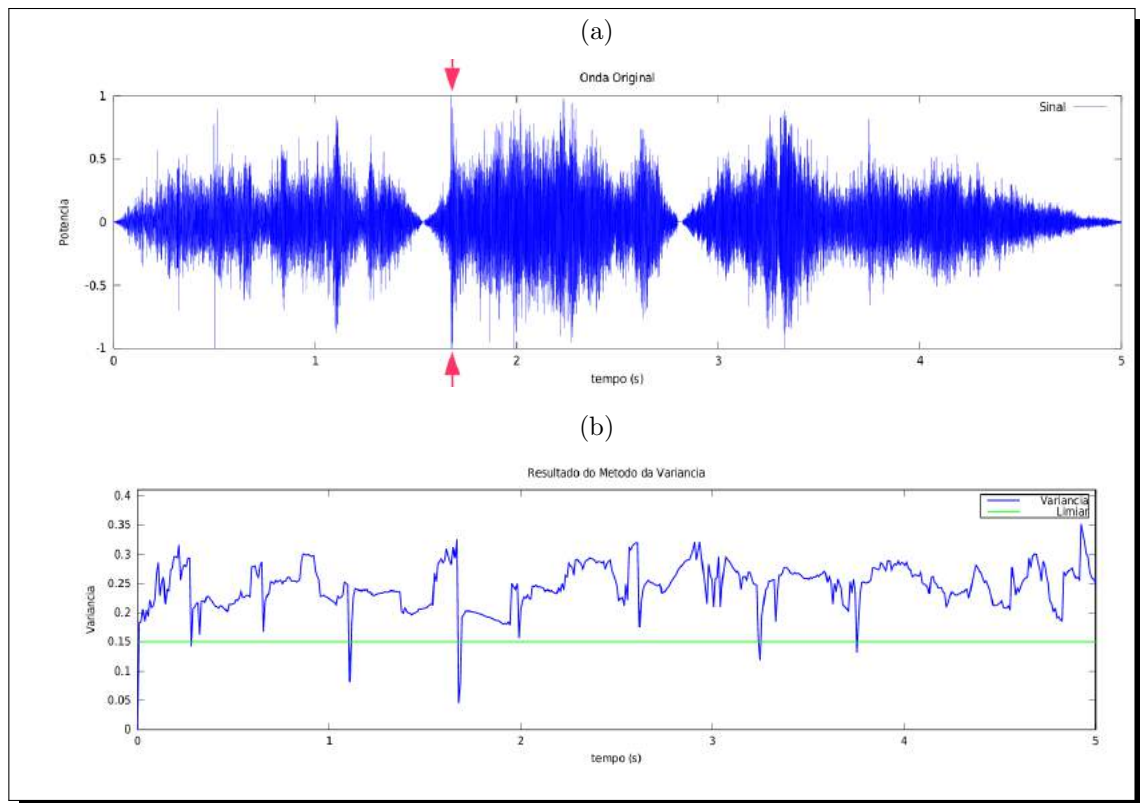
Na figura 16 vê-se o mesmo exemplo de som ambiente previamente analisado agora submetido ao método da VM, onde a figura 16a exibe o sinal original e a figura 16b o resultado apresentado pela VM. Os vales que ultrapassam a linha do limiar representam a ocorrência do som impulsivo. Neste caso o algoritmo acusou cinco ocorrências.

2.5 Técnicas no Domínio das Frequências

2.5.1 Transformada de Fourier

Transformações são uma forma bem conhecida de resolução de problemas em sistemas lineares. Certo problemas envolvendo sinais possuem uma resolução difícil, ou até mesmo impossível no domínio do tempo. Para estes casos, técnicas como Wavelets, transformadas de Laplace, Fourier e Cossenos são uma forma comum de simplificar o problema, realizando transformações matemáticas que permitem converter um sinal, representado por amplitudes em função do tempo, em uma sequência de senos, cossenos ou de outra função base, com frequências e amplitudes diferentes, que, se somados, podem recompor o sinal original convertendo-o novamente ao domínio do tempo caso necessário.

Figura 16 – Exemplo das fases componentes do processo de detecção utilizando-se a VM.



(a) Sinal Original; (b) VM comparada ao limiar constante.

Fonte: o autor

Por intermédio da transformada de Fourier é possível converter um sinal do domínio do tempo no espectro de suas frequências componentes e, posteriormente, recompor o sinal original a partir delas.

As transformadas de Fourier exercem um papel importante em muitos campos da ciência, como óptica, elétrica, acústica, processamento de imagens, dentre inúmeras outras.

Em se tratando de domínios discretos, é comum o emprego das transformadas Discretas de Fourier ou Discrete Fourier Transform (**DFT**), onde, dada uma função unidimensional discreta $s[n]$, $n = 0, 1, \dots, N - 1$, a sua transformada $F[u]$ é dada por [30]

$$F[u] = \frac{1}{N} \sum_{n=0}^{N-1} s[n] e^{-j2\pi un/N}$$

$$\text{para } u = 0, 1, \dots, N - 1$$

(2.12)

onde

$$j = \sqrt{-1} \quad e$$

$$e^{\pm j\theta} = \cos \theta \pm j \operatorname{sen} \theta$$

Em 1965, J.W. Cooley da IBM e J.W. Tukey da Bell Labs criaram a Transformada

Rápida de Fourier ou Fast Fourier Transform (**FFT**). Trata-se de um método eficiente que reagrupa os cálculos dos coeficientes de uma DFT, obtendo-se um menor esforço computacional no cálculo da transformada discreta [30].

Quantidade de frequências obtidas

Para um dado um sinal com duração de Δt , medido em segundos, e com uma taxa de amostragem f_a , dada em Hz , obtém-se $N = (f_a \times \Delta t) + 1$ amostras. A quantidade de valores obtidos no cálculo de uma FFT é dependente desta quantidade de amostras apresentadas, ou seja, para cada valor de N apresentado no cálculo da FFT haverá um valor de $F[u]$ correspondente, conforme visto na equação 2.12.

Range de frequências obtido

Devido às suas características, uma DFT será simétrica e periódica no período N . Desta forma a maior frequência passível de ser representada pode ser encontrada nas transformadas de número $\{\pm\frac{N}{2}, \pm3\frac{N}{2}, \pm5\frac{N}{2}, \dots\}$ que corresponderá ao valor obtido pela Teorema de Nyquist, expresso em Hz , conforme visto na equação 2.13 [30].

$$f_{max} = \frac{f_a}{2} Hz \quad (2.13)$$

Tais frequências estarão distribuídas em um intervalo de $f_1 = \frac{f_a}{N-1}$, valor este conhecido como frequência fundamental.

Nos exemplos das figuras 17 e 18 vê-se o sinal sonoro de um tiro de fuzil e o de um latido que foram digitalizados a $44100 Hz$ e tiveram suas FFT calculadas. Nelas nota-se que somente metade dos valores das componentes de frequências dos sinais são importantes, já que a outra metade é sua imagem especular. Assim, para cada FFT calculada teremos o valor $44100/2 = 22050$ como a frequência máxima possível, de acordo com a equação 2.13.

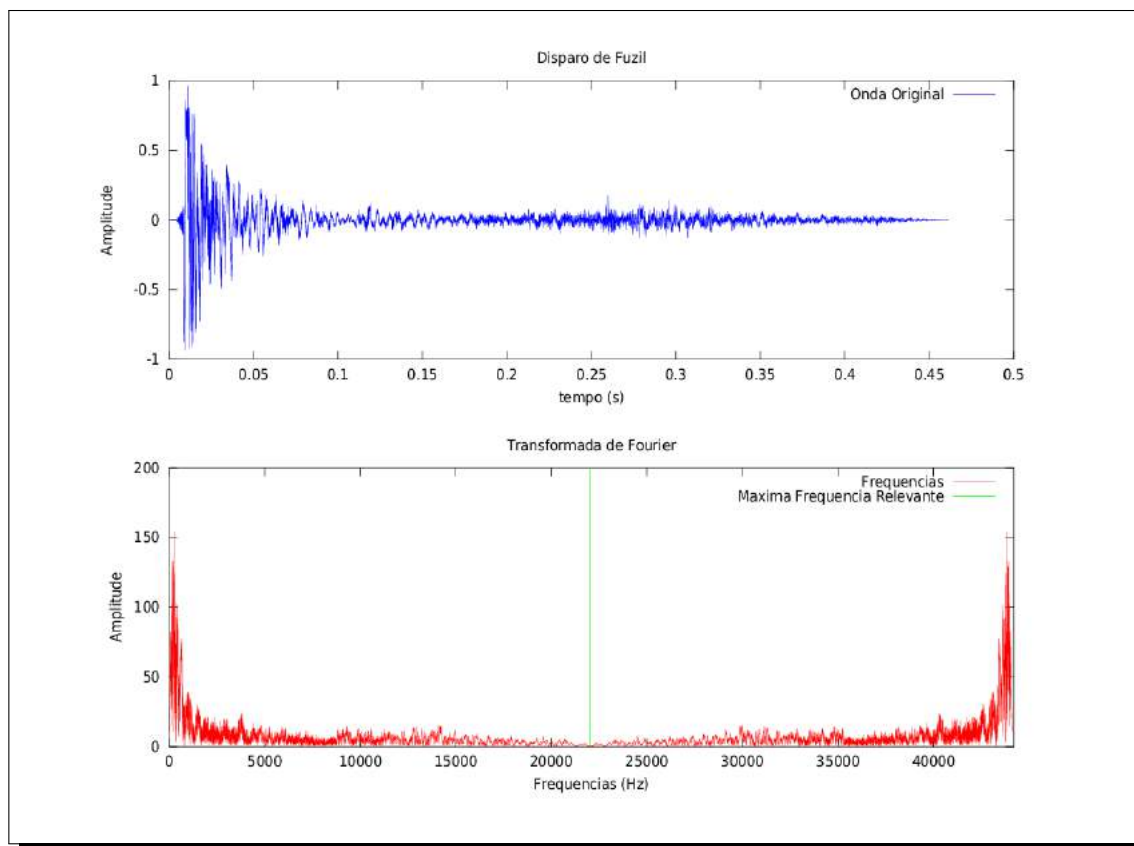
Como nestes exemplos foram utilizadas 20287 amostras em $0.46 s$ de sinal amostrado a $44100 Hz$, obteve-se ao final do processo, um total de 10143 frequências, distribuídas no intervalo entre 1 e $22050 Hz$, espaçadas por um intervalo aproximado de $2.17 Hz$.

Zero Padding

Como visto anteriormente, a quantidade de frequências obtidas ao final de uma FFT é dependente do número de amostras utilizadas.

Como a parte impulsiva de um típico disparo de arma de fogo possui duração aproximada de $0.05 s$, que digitalizado a $44100 Hz$ resulta em 2206 amostras, somente será possível a obtenção de um total de 1102 valores de frequências, distribuídos no intervalo entre 1 e $22050 Hz$, ou seja, um valor para cada grupo de $20 Hz$.

Figura 17 – Sinal contendo um disparo de fuzil e a sua FFT.



Fonte: o autor

Em alguns casos isso pode ser um inconveniente, já que pode ser necessário a obtenção de número específico de valores de frequência, ou ainda, quando deseja-se comparar as frequências de duas amostras de tamanhos diferentes.

Para estes casos utiliza-se a técnica do *Zero padding*, que consiste no acréscimo de um número específico de zeros ao final da amostra de forma a dar-lhe o tamanho adequado à obtenção da interpolação de frequências desejada. Isso fará com que a quantidade de frequências representadas aumente sem que haja alteração no espectro representado [31].

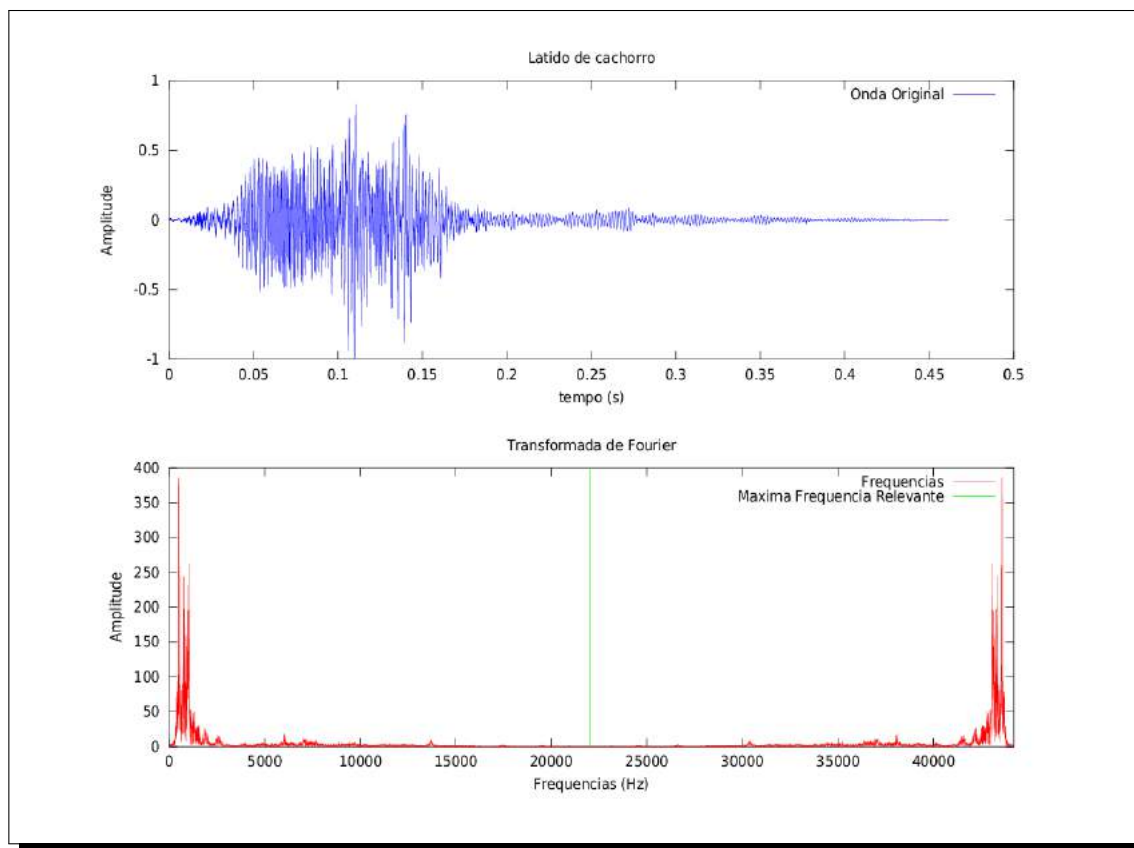
A fim de demonstrar a utilidade desta técnica, foi gerado um sinal definido pela soma de senos exibida na equação 2.14, construída de forma que sua transformada de Fourier apresentasse picos em suas três principais componentes, localizadas aos 200, 201 e 202 Hz .

$$f(t) = \sin(200 \times 2\pi t) + \sin(201 \times 2\pi t) + \sin(202 \times 2\pi t) \quad (2.14)$$

Este sinal foi amostrado a uma taxa de 1kHz durante 1 s, gerando assim 1001 amostras. O resultado da transformada é exibido na figura 19, onde pode-se ver o sinal gerado (figura 19a) e sua FFT (figuras 19b e 19c).

Observando-se o gráfico da figura 19b, nota-se o pico esperado nas circunvizinhanças dos 200 Hz . Porém, uma análise mais detalhada de tal FFT mostra que não existe a

Figura 18 – Sinal contendo um latido de cachorro e a sua FFT.



Fonte: o autor

representação gráfica das três frequências componentes do sinal, conforme pode ser visto na figura 19c.

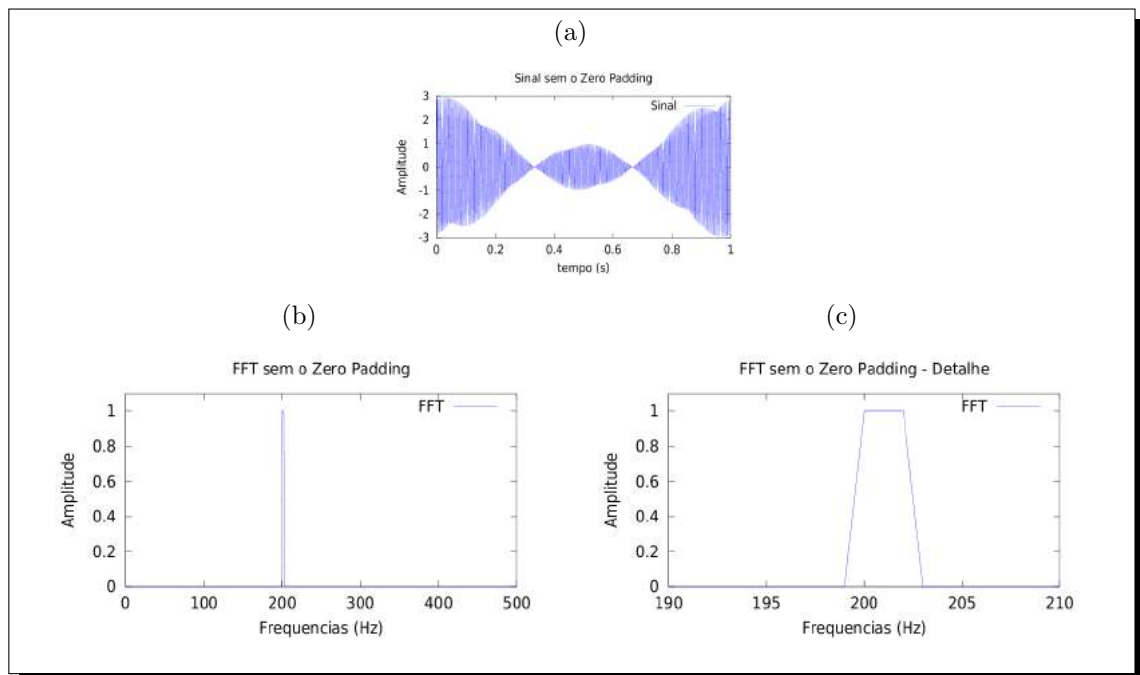
Isto ocorre porque ao apresentar-se 1001 amostras para o cálculo da transformada obtém-se 1001 valores de FFT, distribuídos em intervalos de 1 Hz cada, não havendo assim interpolação suficiente para representar pontos entre os três picos existentes em 200, 201 e 202 Hz , transformando-os em um único platô.

Entretanto é possível melhorar tal interpolação pela simples adição de zeros ao final do sinal original. Este procedimento não altera o sinal em si, mas causa uma grande modificação no resultado obtido na FFT.

Na figura 20a vê-se o mesmo sinal, com a mesma taxa de amostragem, ao qual foram adicionadas mais 2000 amostras, todas contendo zeros. Desta forma passam a ser apresentadas 3001 amostras ao cálculo da transformada, que devolve 3001 valores para as mesmas 500 frequências significativas, já que a taxa de amostragem não foi modificada, resultando em três valores distintos para cada intervalo de 1 Hz .

Como consequência, vê-se na figura 20b o mesmo pico na região dos 200 Hz , que observado no detalhe da figura 20c, permite a visualização dos três picos esperados.

Figura 19 – Transformada de Fourier aplicada sobre um sinal com taxa de amostragem de 1kHz, de comprimento igual a 1001 amostras, sem aplicação de *Zero Padding*.



(a) Sinal sem o *Zero Padding*; (b) FFT do sinal sem o *Zero Padding*; (c) FFT do sinal sem *Zero Padding*, detalhando as frequências entre 190 e 210 Hz.

Fonte: o autor

2.5.2 Remoção de Ruídos

Toda aplicação que se propõe a realizar a gravação de um sinal sonoro em um ambiente não controlado estará sujeita a captura de outros sons indesejáveis. Dentro desta ótica, estes sons devem ser, então, considerados como ruído e, desta forma, serem eliminados a fim de garantir um melhor reconhecimento.

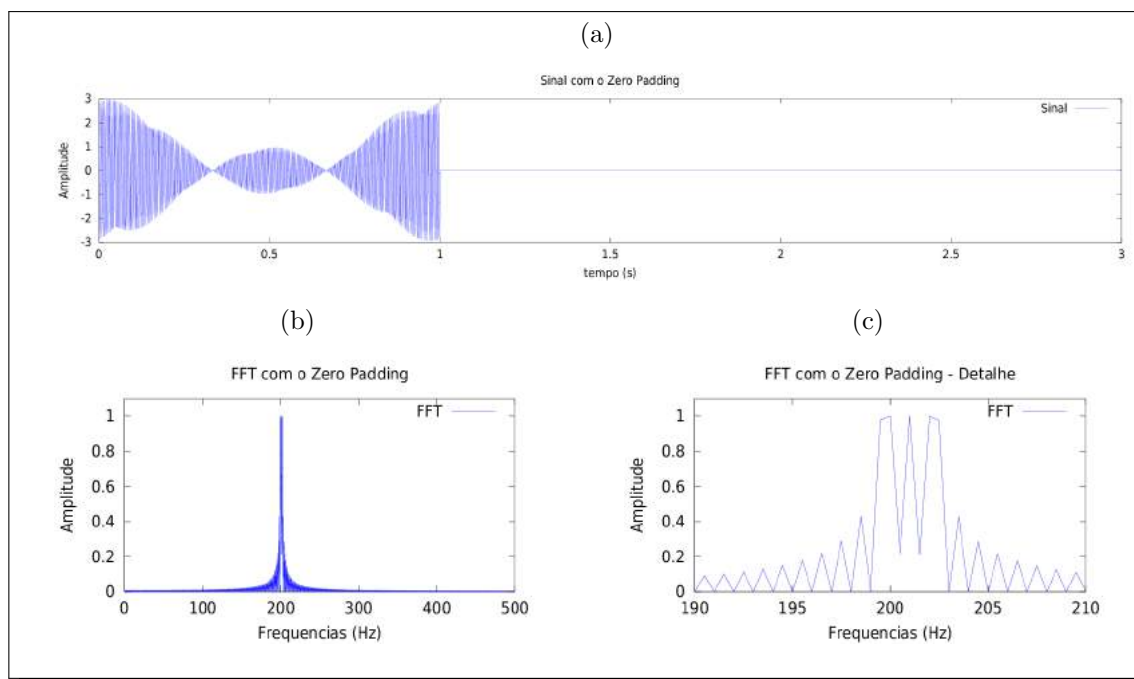
As técnicas normalmente utilizadas para a tarefa podem ser agrupadas em três categorias distintas, a saber:

Beamforming – utiliza um conjunto de sensores distribuídos pelo ambiente que capturam o sinal sonoro de diversas fontes. Com base na posição dos sensores e da fonte sonora desejada, o som produzido pelas demais fontes são considerados como ruído, podendo assim, serem atenuados. Esta técnica é hoje muito utilizada para filtragem de ruídos na telefonia celular.

Suas desvantagens residem na necessidade de utilizar-se vários transdutores separados espacialmente e na necessidade de se conhecer o posicionamento da fonte sonora de interesse [32, 33].

Cancelamento Adaptativo de Ruído – utiliza dois transdutores de forma a um captar prioritariamente o som ambiente e o outro captar a fonte de sinal desejada,

Figura 20 – Transformada de Fourier aplicada sobre um sinal com taxa de amostragem de 1kHz, de comprimento igual a 1001 amostras, após a aplicação de *Zero Padding* com 2000 amostras.



(a) Sinal com *Zero Padding* de 2000 amostras; (b) FFT do sinal com *Zero Padding*; (c) FFT do sinal com *Zero Padding*, detalhando as frequências entre 190 e 210 Hz.

Fonte: o autor

realizando, a seguir o cancelamento do som ambiente no sinal desejado.

Sua principal desvantagem é necessitar de um ambiente controlado para que a eliminação do ruído seja efetiva [32, 33].

Modificação Espectral Adaptativa – utiliza um único microfone para a captura do sinal, removendo-se o ruído através da análise deste sinal, no domínio do tempo ou no domínio das frequências.

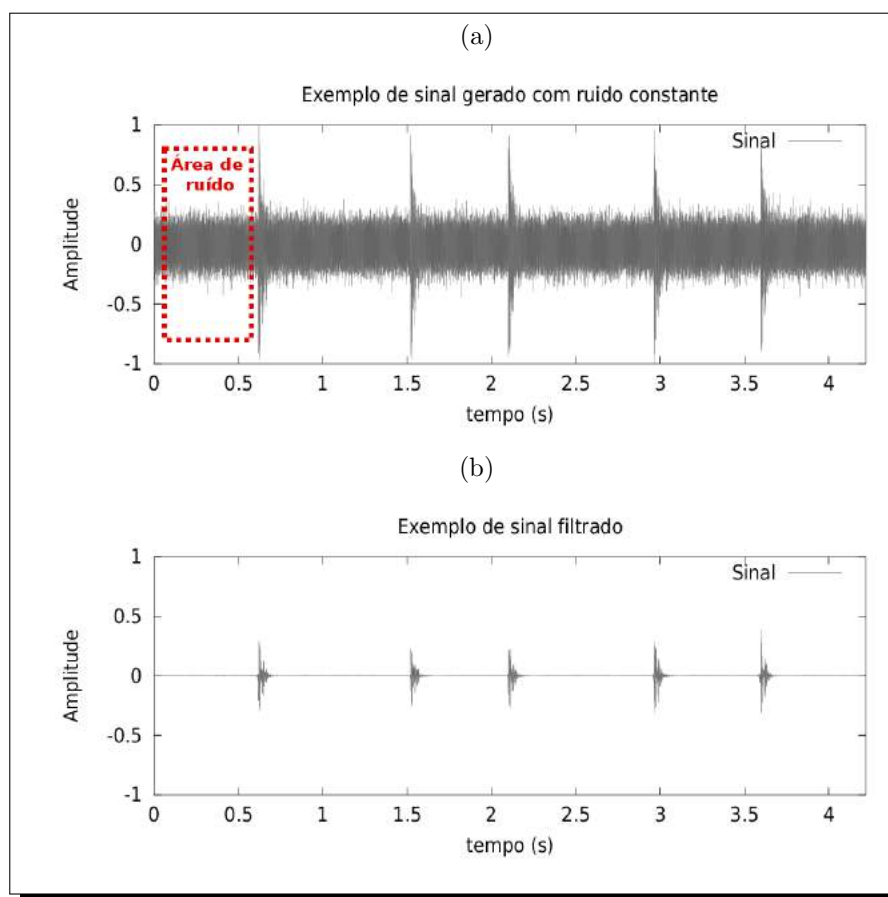
- O método da *Correlation Substraction* (CS) utiliza uma técnica no domínio do tempo, onde encontra-se um *delay* de tempo entre o ruído e o sinal desejado, para, a seguir subtrair aquele primeiro de todo o sinal sonoro [34].
- Já no método do *Spectral Noise Gating* (SNG), seleciona-se uma área do sinal que contenha apenas ruído e nela realiza-se uma análise espectral baseada na FFT, de forma a criar uma assinatura do ruído, denominada *gate*. A seguir remove-se, ainda no domínio das frequências o *gate* de todo o sinal. Após a realização da inversa da FFT para o sinal filtrado, obtém-se o sinal original subtraído do ruído assinalado [34].

A principal desvantagem dos métodos de modificação espectral adaptativa está no

fato de que todos eles realizam uma atenuação em todo o sinal apresentado, seja subtraindo o ruído e também parte do sinal desejado, seja removendo frequências que podem ser importantes na composição de um disparo de arma de fogo, que serão perdidas caso estas estejam presentes no trecho de ruído escolhido.

Em virtude das características do sistema de detecção, da eficiência do método e facilidade na implementação, optou-se pela utilização do método SNG, implementado pelo software SoX. Um exemplo de sua utilização pode ser visto na figura 21, onde vê-se um sinal gerado com ruído constante e a área de ruído utilizada como referência para a filtragem do sinal, identificada pela linha pontilhada na figura 21a, e o sinal resultante obtido depois da filtragem, visto na figura 21b.

Figura 21 – Exemplo de sinal gerado antes e depois de ser submetido aos filtro “*Spectral Noise Gating*”.



(a) Sinal gerado com cinco disparos e ruído branco de amplitude constante; (b) Sinal com ruído branco após passagem pelo filtro “*Spectral Noise Gating*”. A linha pontilhada na figura (a) identifica a região contendo ruído, utilizada na criação do *gate* de filtragem

Fonte: o autor

2.6 Técnicas de Aprendizado de Máquina

2.6.1 Data Mining

A observação de qualquer evento no mundo real pode gerar uma infinidade de dados aparentemente desconexos onde a análise tradicional muitas vezes não produz conclusões satisfatórias, já que pode ser extremamente difícil determinar quais parâmetros possuem influência significativa no resultado final em meio a todos os dados coletados.

Este é o caso de um sistema de reconhecimento de padrões em sinais, já que é necessário localizar padrões e assinaturas sonoras em um rol enorme de frequências que por vezes não tem uma correlação óbvia. Desta forma o uso de ferramentas de Data Mining passa a ser de extrema valia para o presente Sistema.

Data Mining é definido como sendo o processo utilizado para analisar, de forma automática ou semiautomática, dados armazenados e neles encontrar padrões que irão possibilitar a resolução de problemas. Através da identificação de tais padrões em dados previamente estocados, será possível, pela análise de dados similares do presente, realizar previsões e projeções para o futuro [35].

Classificação do Input

Como no caso dos bancos de dados relacionais, quando se trabalha com Data Mining é necessário organizar os dados de forma eficiente. Para tanto, organiza-se a informação utilizando termos como Conceitos, Instâncias, Atributos e Regras [35] [36].

Conceito

Pode ser entendido como o conjunto de dados coletados em uma determinada observação. Em um sistema de Detecção e reconhecimento de Armas de Fogo, pode-se, por exemplo, coletar, além das frequências componentes do sinal sonoro, dados como calibre da arma, tipo de munição utilizada, distância do disparo, duração do disparo, nível de ruído de fundo, a envoltória do sinal, etc ...

Instância

Será cada um dos exemplos do conceito a ser aprendido, tomados de forma individual. No presente caso uma instância seria uma observação contendo todos ou alguns dos elementos acima descritos. Assim, uma instância de disparo de arma de fogo descrevendo o calibre da arma, tipo de munição utilizada, distância do disparo, duração do disparo e nível de ruído de fundo poderia ser representada por

{.380, Ponta_Oca, 100 m, 0.10 s, Silencioso}

Atributo

São os dados propriamente ditos. Eles podem ser numéricos e não numéricos. Dados numéricos podem ser reais ou inteiros. Já os dados não numéricos podem ser:

- *Nominais* – quando os valores do atributo recebem um rótulo escolhido dentro de um rol de possibilidades. Por exemplo, o tipo de munição pode ser considerado um atributo nominal caso o seu valor deva ser escolhido entre um dos valores pertencentes ao conjunto

{Ogival_Chumbo, Ogival_Encamisado, Semi_Encamisado, Ponta_Oca}

- *Ordinais* – quando os valores, apesar de serem também escolhidos em um conjunto predefinido, estes possuem uma certa gradação, podendo ser ordenados. Como exemplo poder-se-ia dizer que nível de ruído de fundo formam o conjunto

{Silencioso, Moderado, Ruidoso, Muito Ruidoso}

onde a quantidade de ruído de fundo encontrada no sinal dá a seguinte gradação:

Silencioso < Moderado < Ruidoso < Muito Ruidoso

- *Intervalos* – quando, ao invés de valores numéricos absolutos, adotam-se faixas de valores. Pode-se tomar, como exemplo, o caso da distância do disparo, onde são definidos intervalos a fim de simplificar a representação dos dados

{10 – 20 m, 20 – 50 m, 50 – 100 m, 100 – 500 m, Acima de 500 m}

Regras

Definem como os diferentes atributos interagem entre si em um universo de atributos predefinidos. Por exemplo, da observação de todos os tipos de munição possíveis e das armas a que elas são destinadas, a seguinte regra pode ser inferida:

Não há munição ogival de chumbo para o calibre 0.50

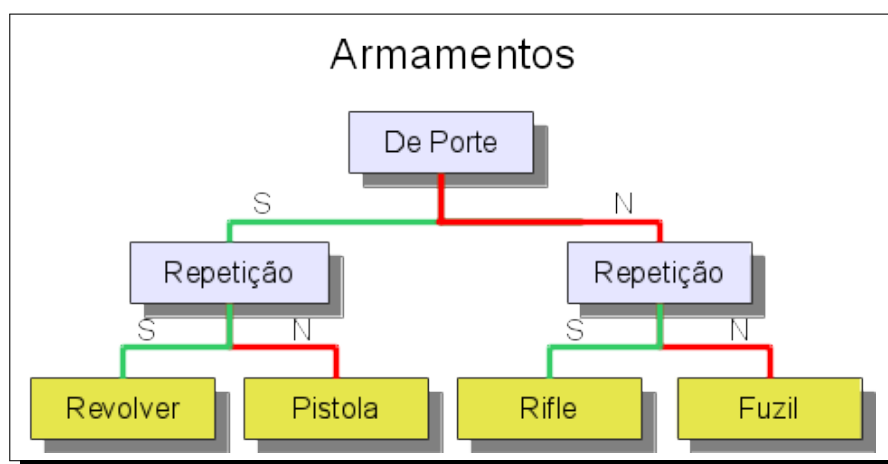
Modelos de aprendizado

Existem basicamente quatro modelos de aprendizado em aplicações de Data Mining [35] [36]:

Classificação

Neste modelo, os dados de aprendizado são agrupados com base em um ou mais atributos de modo a criar um conjunto de classes. Eventos futuros, então, podem ter seus resultados previstos utilizando-se as regras inferidas na classificação de um conjunto de dados analisados previamente. Na figura 22 observa-se um esquema de classificação simplificada de armamento com base em apenas quatro regras, resultando em quatro classes.

Figura 22 – Exemplo de classificação de armamento quanto ao funcionamento.



Fonte: o autor

Associação

De forma semelhante, a Associação irá agrupar um ou mais atributos de forma a realizar a previsão de algo, porém, de forma diferente da Classificação, com a Associação é possível prever qualquer um dos atributos e não apenas uma classe específica. Para tanto, os dados são exaustivamente triados e comparados a fim de encontrar relações que tenham relevância na previsão de eventos futuros. Como consequência, haverá um número muito maior de regras de associação que regras de classificação, sendo necessário separar aquelas que se aplicam a um número mínimo de dados do *dataset*, desconsiderando-se as outras.

Clustering

Nos casos onde não se pode estabelecer classes, usa-se a técnica de *Clustering*, onde os dados serão agrupados em faixas, de acordo com sua distribuição natural. Esta técnica é particularmente útil quando os dados coletados são em sua maioria numéricos e de grande diversidade. Assim, determina-se inicialmente em quantos *clusters* a amostra será agrupada e a análise dos dados irá criá-los com base em sua distribuição.

Predição numérica

É aquela na qual o valor a ser previsto não é uma classe discreta, mas uma quantidade numérica.

Neste trabalho utilizou-se o WEKA (Waikato Environment for Knowledge Analysis) como ferramenta destinada a realizar o Data Mining dos dados obtidos, utilizando-se do algoritmo de clusterização K-Médias e do algoritmo de classificação *Support Vector Machine* (SVM), nele contidos.

2.6.2 Método K-Médias

Por ser baseado em clusters, o algoritmo K-Médias ou K-Means ou ainda Lloyd-Forgy, procura fornecer uma classificação baseada nas informações numéricas, contínuas ou discretas, fornecidas pelos próprios dados, sem que nenhuma pré-classificação seja necessária. O nome é composto pelo K , que representa a quantidade de clusters a serem criados, mais a palavra *Média*, que descreve o meio utilizado para a determinação dos centroides [35] [36].

É um algoritmo **não supervisionado**, ou seja, os dados de entrada não são rotulados previamente em classes, sendo necessário, portanto que o número de grupos seja arbitrado ao início do processo.

Ao final do processo, todas as instâncias pertencerão a um cluster, inclusive as distorções. A posição inicial dos centroides pode ser modificada arbitrariamente e influenciará no resultado final da clusterização.

Este método necessita que todos os valores de atributos sejam numéricos, não trabalhando bem com outros tipos de valores.

O primeiro passo para a utilização deste método é definir qual o algoritmo de distância será utilizado. Os mais comuns são a distância Euclidiana, definida pela equação 2.15 e ilustrada na figura 23a e a Manhattan, definida pela equação 2.16 e ilustrada na figura 23b [36]

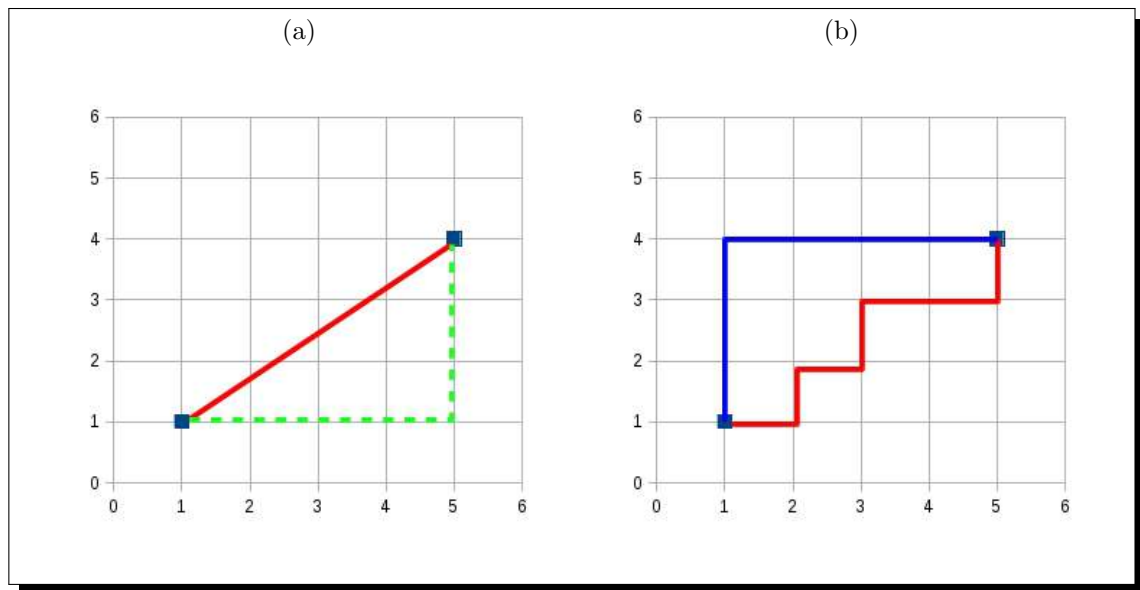
$$D_e = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2.15)$$

$$D_m = |x_1 - x_2| + |y_1 - y_2| \quad (2.16)$$

A seguir, arbitra-se em quantos clusters deseja-se que a informação levantada seja mapeada. Este número é chamado de k .

São arbitrados, então, k pontos quaisquer da distribuição que servirão como centroides temporários (figura 24a), sendo os valores numéricos de cada instância da amostra fornecida a eles comparados utilizando-se o cálculo de distância escolhido (figura 24b).

Figura 23 – Algoritmos de Distância.



(a) Distância Euclidiana; (b) Distância Manhattan.

Fonte: o autor

Criam-se, assim, k clusters provisórios que conterão todas as instâncias próximas a um dado centroide (figura 24c).

A seguir, os centroides serão recalculados com base nos valores das instâncias pertencentes ao cluster, refinando-se as coordenadas dos centroides a cada repetição do processo (figura 24d). O final do processo é alcançado quando obtêm-se k centroides, que serão os pontos centrais de k clusters, cada um contendo os pontos que mais se aproximam do centroide em referência (figura 24e). O algoritmo do K-Médias pode ser visto no apêndice A.

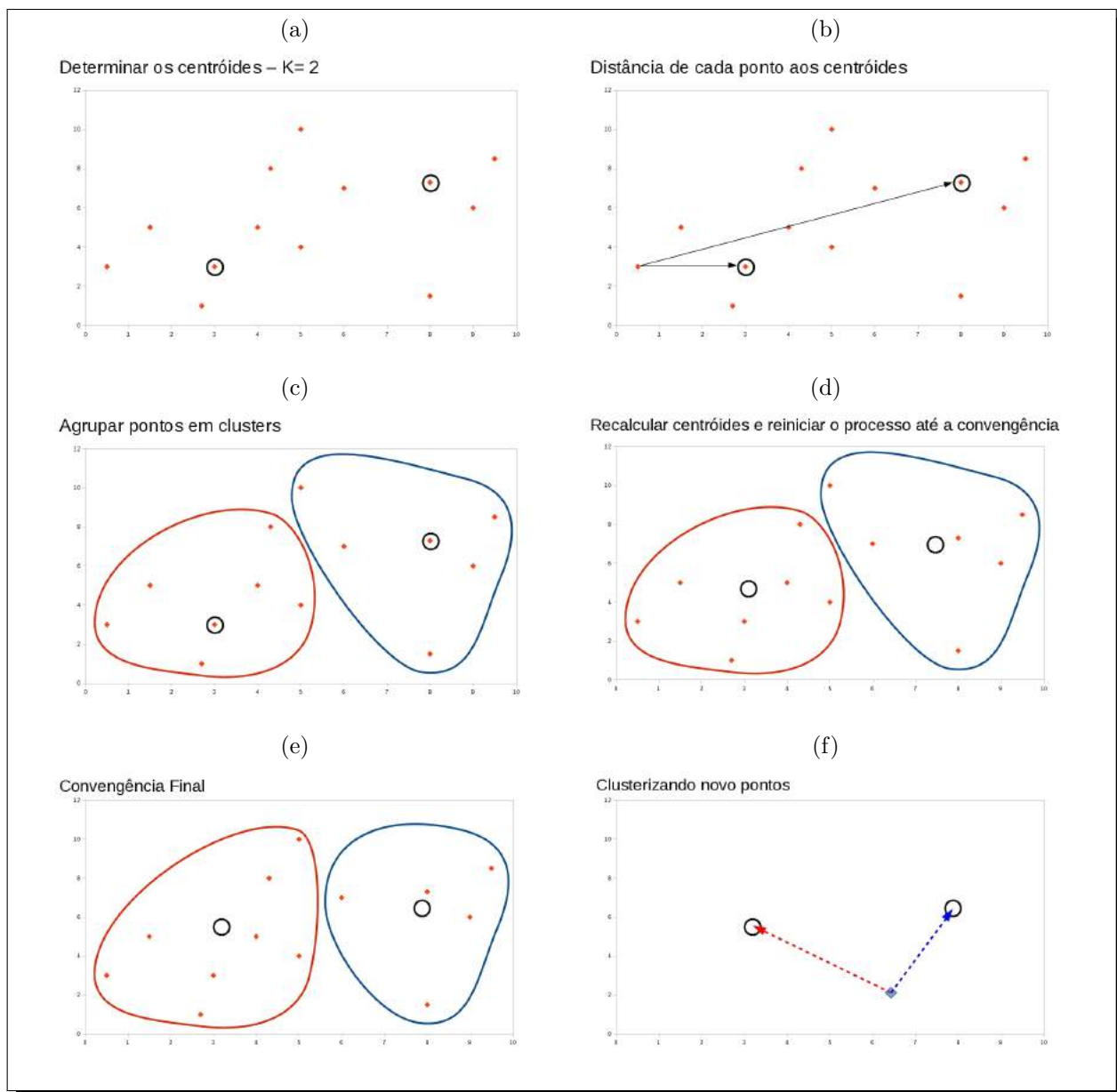
Como vantagem, o algoritmo K-Médias atribui todos os objetos de informação a um cluster de forma automática, sendo necessário pouco ou nenhum processamento prévio da informação. Suas desvantagens advêm da escolha do número de grupos, que deve ser feita antes do início do processo, do fato que todos os objetos de informação são forçados a pertencerem a um cluster e do método possuir baixa eficiência para dados não numéricos.

No presente trabalho, utilizou-se o K-Médias como um algoritmo supervisionado. Assim, após a obtenção dos centroides com um conjunto de treinamento aplicou-se a cada novo ponto apresentado o mesmo algoritmo de distância, colocando-o no *cluster* de centroide mais próximo (figura 24f).

2.6.3 Support Vector Machine

Support Vector Machine (**SVM**) é um método de aprendizado de máquina, baseado na Teoria de Aprendizagem Estatística, possuidor de uma base teórica sólida e um alto desempenho prático. É um algoritmo de classificação supervisionado, ou seja, os dados

Figura 24 – Fases do algoritmo K-Médias.



(a) Determinação dos Centróides iniciais; (b) Cálculo da distância de cada ponto aos Centroides; (c) Agrupando os pontos pelo proximidade ao Centroides; (d) Recalculando os Centroides com base em seus pontos; (e) Clusters e centroides em sua situação final; (f) Utilizando os centroides aprendidos para clusterizar novos pontos.

Fonte: o autor

inseridos como conjunto de treinamento necessitam estar pré-classificados para que o algoritmo possa inferir os critérios de classificação [35] [36].

Desta forma, toma-se um conjunto de dados de treinamento D , pré-classificados com os elementos de um conjunto de rótulos R , a partir do qual se tenta encontrar, por aproximação, uma função $f(D) \mapsto R$.

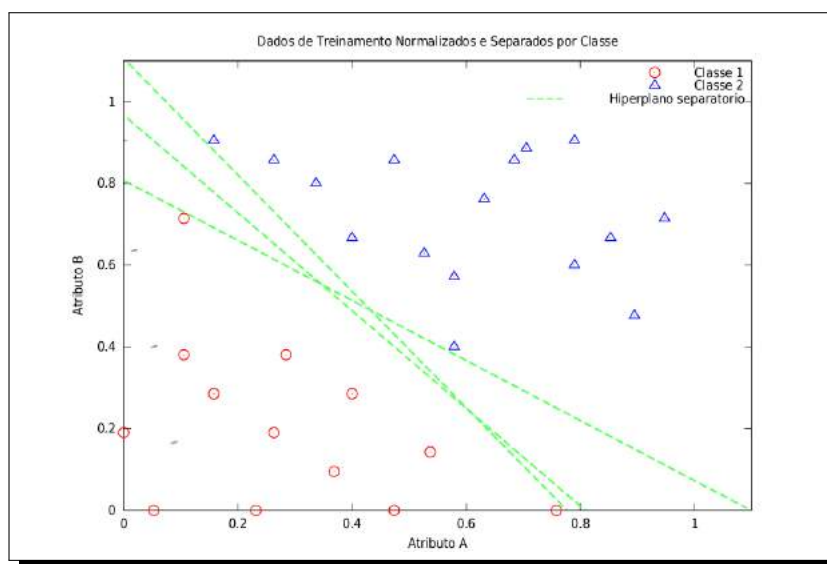
Isto é feito pela utilização de mapeamentos não lineares que transformam os N atributos

de cada entidade em um gráfico N -dimensional e pela busca de hiperplanos de separação que permitam uma separação ideal das C classes previamente mapeadas [36].

O primeiro passo para utilizar-se uma classificação baseada em SVM será converter qualquer valor não numérico em números reais, já que o algoritmo somente pode tratar estes tipo de valores. A seguir, deve ser feita a normalização dos dados, onde cada atributo terá os seus valores colocados no intervalo $[-1, +1]$ ou $[0, 1]$. Isto evita que valores de grandezas diferentes passem a ter um peso diferenciado no cálculo dos limites entre classes.

A figura 25 mostra o gráfico de um conjunto de treinamento contendo 29 entidades com 2 atributos já normalizados, que foram previamente classificadas em 2 classes. Nela vê-se uma área bem definida que separa as duas classes, sendo possível traçar uma reta que sirva como fronteira entre as regiões ocupadas pelas Classes 1 e 2. Porém existem inúmeras retas que desempenham este papel, mas somente uma que o faz de maneira ótima, ou seja, que maximiza a distância entre si e os pontos de cada uma das classes.

Figura 25 – Dados normalizados separados por classe e Hiperplanos possíveis.



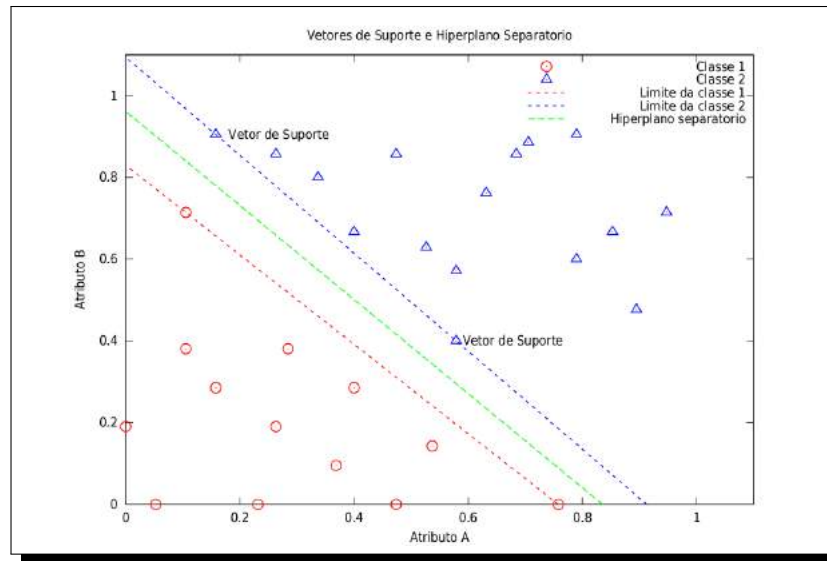
Fonte: o autor

Para definir tal reta determina-se quais são as entidades do espaço de dados de treinamento que definem as fronteiras de uma determinada classe. Estes pontos recebem o nome de *Vetores de Suporte* e podem ser definidos como sendo os únicos pontos de uma classe que, se removidos do *set* de treinamento, irão alterar os limites desta [36].

Neste exemplo chegou-se a conclusão que as entidades $\{Classe1, 0.105, 0.714\}$, $\{Classe1, 0.758, 0.000\}$, $\{Classe2, 0.158, 0.905\}$ e $\{Classe2, 0.579, 0.400\}$ são os vetores de suporte de cada uma das classes, servindo como base para o traçado das retas que definem os limite de cada classe. A partir delas pode-se encontrar uma terceira reta, equidistante de ambas, que definirá o hiperplano de separação das classes, conforme vê-se na figura 26.

De posse deste hiperplano, qualquer nova entidade apresentada será a ele comparada,

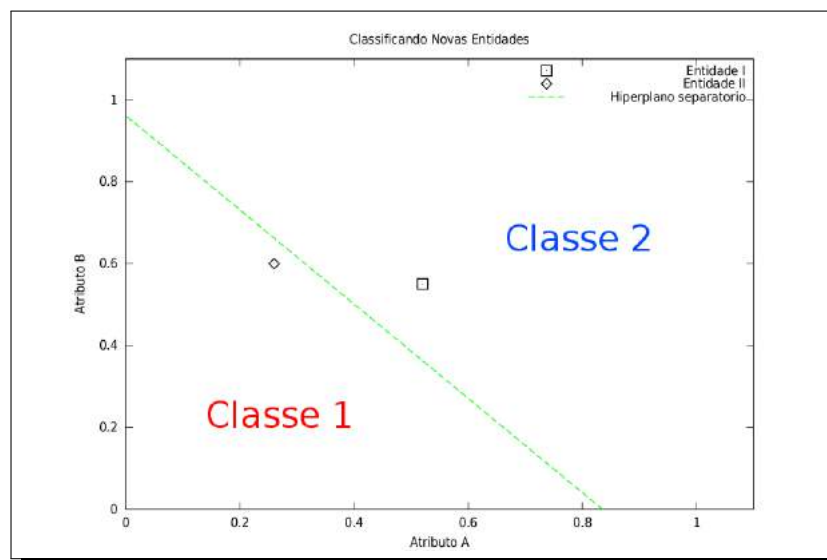
Figura 26 – Vetores de Suporte, Limite de Classes e Hiperplano Separatório.



Fonte: o autor

possibilitando sua classificação em uma das classes definidas, de acordo com o seu posicionamento relativo. Na figura 27 foram apresentadas duas novas entidades cujas classes são ignoradas. Vê-se que a **Entidade I** será classificada como pertencendo à **Classe 2** e a **Entidade II** como pertencente à **Classe 1**.

Figura 27 – Classificação, com base no Hiperplano, de Entidades cuja classe é desconhecida.



Fonte: o autor

Caso não exista a possibilidade de uma separação linear entre as classes, funções polinomiais podem ser utilizadas para definir os hiperplanos de separação não lineares.

Como a identificação de disparos de arma de fogo possui todas as características típicas de uma classificação, espera-se um melhor desempenho do algoritmo SVM sobre o K-Médias na solução do problema.

3 O Sistema Modular - Propostas e Soluções

3.1 Sistemas existentes

Existem diversos sistemas comerciais que realizam a tarefa de detecção e reconhecimento de disparos de armas de fogo, que em sua maioria também realizam a identificação da posição do atirador, dos quais destacam-se os seguintes:

Shoulder-Worn Acoustic Targeting System (Swats) – desenvolvido pela QinetiQ, é o sistema padrão do Exército dos Estados Unidos da América, sendo também utilizado pelo Marine Corps daquele país, além da França e Austrália. É um sistema de detecção de disparos e identificação da posição do atirador, composto por uma mochila contendo microfones, um receptor de GPS, um giroscópio, bússola e acelerômetros [37].

Boomerang – desenvolvido pela Raytheon BBN é um outro sistema de detecção e localização de atirador, composto por uma mochila com 6 Kg de peso, contendo um laptop rodando Windows 7 e o sistema de microfones [38].

Pilar – fabricado pela BSS Defense and Security Solutions fornece ao operador o ponto de origem, distância, elevação e calibre da arma disparada, podendo ser montado em edifícios, veículos, helicópteros e armamento pessoal [39].

PinPoint – o PinPoint *Gunshot Detection System* da Cobham utiliza uma arquitetura baseada no domínio do tempo, aliada a técnicas de inteligência artificial que permitem a detecção rápida e precisa de disparos em ambiente ruidoso [40].

Não existe, porém, informações sobre quais algoritmos tais sistemas utilizam, nem, tampouco, outros detalhes sobre o seu funcionamento, devido a sua natureza comercial.

Entretanto, a literatura propõe diversas técnicas para detecção de disparos, como:

Filtro de mediana – já descrito anteriormente [7] [22] [41];

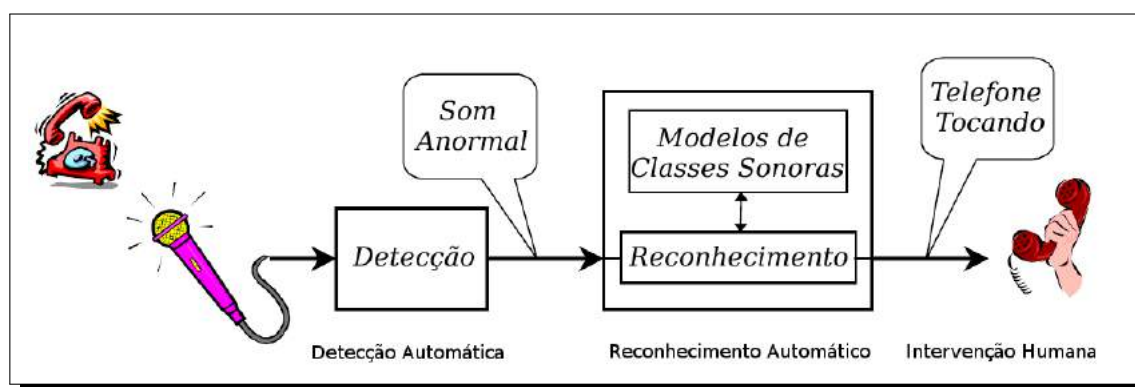
Teager Energy Operator (TEO) – definido como $\Psi(x_n) = x_n^2 - (x_{n-1} \cdot x_{n+1})$ para o caso de sinais discretos, onde x representa cada amostra do sinal, é uma forma de detectar grandes variações na potência sonora do ambiente [41];

Correlação Contra um *Template* – método no qual o sinal a ser analisado é comparado com um modelo conhecido, concluindo-se sobre sua natureza impulsiva [41] [42].

Além destes, ainda são citados na literatura outros métodos que podem ser utilizados em sistemas de detecção e reconhecimento de disparos, como as Transformadas Discretas de Wavelet (DWT), transformadas de Fourier, transformadas de Cossenos, Análise Espectral, Cepstrum e Mel-frequencys, Coeficientes de Predição Linear (Linear Prediction Coefficients - LPC), Transformações lineares, como a Principal Components Analysis (PCA), Hidden Markov Models (HMM), dentre outras.

Em [1] também é proposto um modelo de sistema de reconhecimento, visto na figura 28. Nele o sistema de detecção é subdividido em dois módulos distintos, sendo o **Módulo de Detecção de Som Impulsivo** responsável pela detecção de um som impulsivo qualquer e o **Módulo de Reconhecimentos de Sons**, responsável pela sua identificação, baseada em modelos de classes de sons.

Figura 28 – Modelo Proposto em [1]



Fonte: [1].⁴

3.2 Módulos componentes do sistema

Complementando a descrição sugerida em [1], propõe-se aqui um sistema mais abrangente e operacional para detecção e reconhecimento de disparos de armas de fogo que irá possuir cinco módulos distintos, trabalhando de forma paralela em um refinamento sucessivo do reconhecimento, como pode ser visto na figura 29.

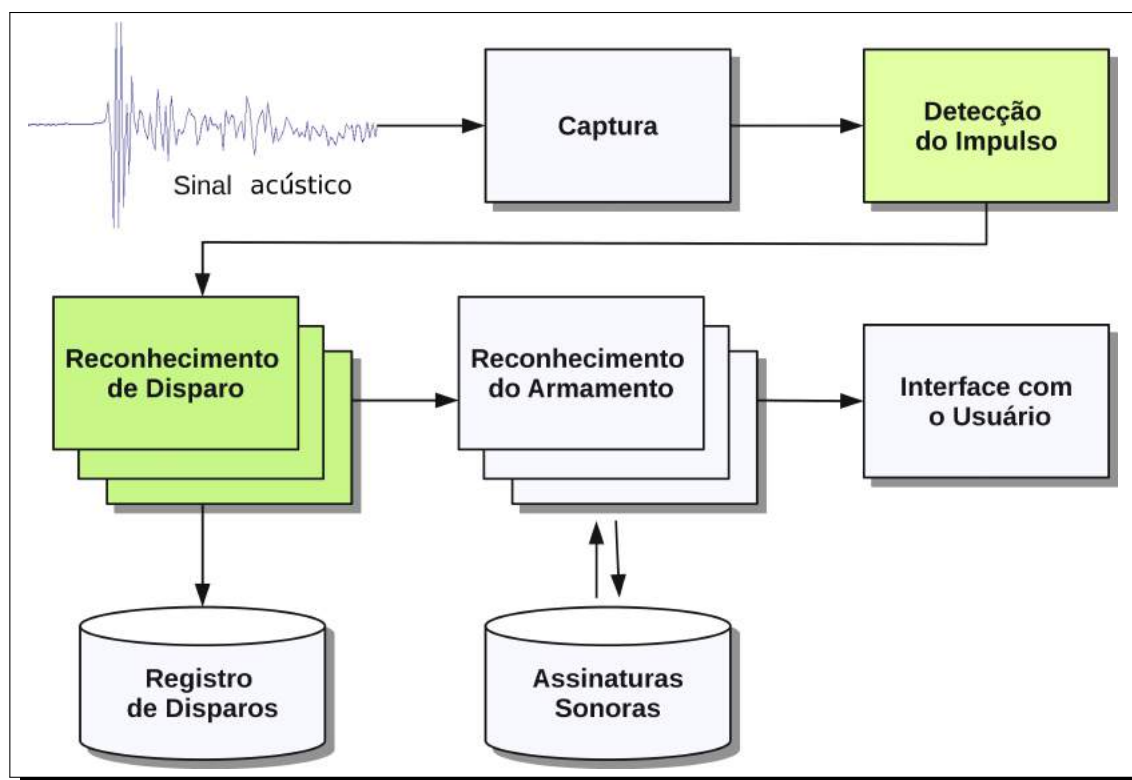
Tais módulos desempenharão as seguintes tarefas:

3.2.1 Módulo de Captura

Tem como finalidade transformar os sinais sonoros em sinais elétricos a fim de que possam ser tratados pelo restante do sistema. Consiste, basicamente de um transdutor acústico-elétrico qualquer, responsável por transformar os sinais sonoros nos sinais elétricos que alimentarão um conversor analógico-digital.

⁴ Traduzido pelo autor.

Figura 29 – Modelo do Sistema Proposto



Fonte: o autor

Sistemas semelhantes já existentes utilizam-se de microfones ou outros sensores acústicos comerciais, como o microfone Sony ECM-MS907 [7], o microfone Bruel Kjaer 4136 [9] e o Sound Meter Bruel Kjaer 2231 [43].

Porém é possível criar funcionalidade semelhante a um custo significativamente mais baixo através do uso de piezoelretos como sensores acústicos, como o proposto em [44], cuja pesquisa vem sendo desenvolvida na UFPB e USP-São Carlos.

3.2.2 Módulo de Detecção de Impulso

Realiza a identificação da ocorrência de sons impulsivos, diferenciando-os dos demais tipos de sons, que passam a ser considerados ruído, dentro do escopo do sistema.

O som ambiente é predominantemente composto por sons que variam de forma menos abrupta, aumentando ou diminuindo a energia acústica local de forma gradativa. Vozes, passos, maquinário, vento e tráfego são sons que não interessam e devem ser filtrados. Assim, o módulo de Detecção deve agir como um pré-processamento do sinal, selecionando sinais candidatos a uma análise mais detalhada, evitando, assim, um sobretrabalho do módulo de reconhecimento, parte mais complexa do sistema e que demanda um maior custo computacional.

Desta forma, o módulo de Detecção de Impulsos monitora continuamente o ambiente

a procura de distúrbios relevantes, baseado no tempo de ataque do som, sua amplitude e a Relação Sinal Ruído (SNR - Signal Noise Ratio). Uma vez encontrado um som que possua características impulsivas este é transmitido para os módulos subsequentes para tratamento, reconhecimento e identificação.

3.2.3 Módulo de Reconhecimento de Disparo

Todo disparo é um som impulsivo, mas nem todo som impulsivo caracteriza um disparo. É neste módulo, portanto, que se realiza uma nova filtragem dos sons capturados, desta vez realizando a diferenciação entre explosões e disparos de armas de fogo e os demais sons impulsivos, baseados na duração do impulso, sua análise de frequência e em características específicas do envelope sonoro [29].

Sons reconhecidos como disparos de armas de fogo podem então ser repassados para o módulo seguinte e/ou serem armazenados em uma base de dados de Registros de Disparos para que, posteriormente, possam ser utilizados por outras aplicações externas ao sistema, como por exemplo, uma análise forense de um disparo específico, onde determinado disparo pode ser novamente analisado a fim de corroborar ou refutar uma dada versão sobre um incidente envolvendo armas de fogo.

3.2.4 Registro de Disparos

Caso o sistema tenha a pretensão de servir como ferramenta forense, será necessário o armazenamento dos sons capturados para uso posterior. É esta, pois a função deste módulo, que armazenará de forma eficiente os intervalos de sons considerados relevantes e a identificação de local e tempo da ocorrência.

Como a digitalização pura dos sons necessita grande volume de armazenamento, é necessário recorrer-se a mecanismos de compactação de áudio para fins de otimização de espaço. Algoritmos de compactação de áudio tradicionais, como o MPEG 1 Layer III (MP3), foram concebidos sob o ponto de vista dos limiares da audição humana, eliminando informações consideradas não relevantes e/ou inaudíveis por conta de sua duração ou de sua frequência.

Tais processos, porém, podem não ser os mais adequados para compactação do áudio no contexto deste sistema, pois durante a compactação, podem ser perdidas informações importantes para futura análise.

Faz-se assim necessário o estudo e, talvez, o desenvolvimento de métodos e algoritmos específicos para a compactação eficiente do sinal de áudio coletado.

3.2.5 Módulo de Reconhecimento de Armamento

Este módulo abriga os algoritmos que irão identificar a assinatura do disparo e compará-la com uma base de dados de assinaturas previamente recolhidas em laboratório que servirão

de referência para a correta identificação do armamento autor do disparo.

3.2.6 Assinaturas Sonoras

É uma base de dados de assinaturas sonoras que servirão como referência para a correta identificação do armamento em questão. Pode conter, também, as outras informações sobre o armamento catalogado que são disponibilizadas ao usuário final.

3.2.7 Interface com o Usuário

É o último estágio do processo, onde o usuário é informado sobre a ocorrência de um disparo e/ou recebe as informações oportunas sobre o tipo de armamento, calibre, *stop-power* ou qualquer outra informação julgada relevante para a situação.

O presente trabalho aborda o estudo dos módulos de Detecção de Impulso 3.2.2 e de Reconhecimento de Disparo 3.2.3, destacados na figura 29.

3.3 Módulo de Detecção de Impulso

Conforme visto na revisão bibliográfica, três diferentes métodos podem ser utilizados nesta etapa, a saber: o RMS, a CMF e a VM. Observou-se, porém, que o método da VM, conforme descrito em [7], apresenta problemas de detecção quando um som impulsivo ocorre dentro de uma janela que já possua outro som impulsivo. Para tanto é apresentada a seguir uma melhoria a este método.

3.3.1 Melhoria ao método do VM

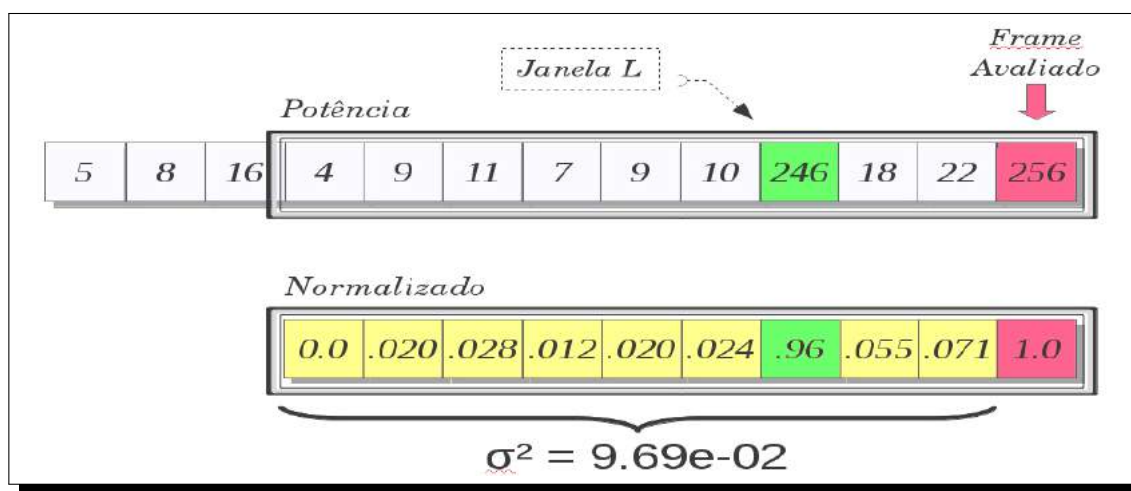
O VM apresenta problemas quando existe mais do que um som impulsivo dentro da janela de detecção, *i.e.* quando ocorrem dois sinais impulsivos em um curto espaço de tempo. Isto ocorre porque um som impulsivo anteriormente detectado estará inserido dentre os $L - 1$ *frames* significativos para o cálculo da variância, elevando assim o valor de σ^2 do intervalo, podendo mascarar uma nova detecção [45].

Para solucionar tal problema, foi proposto neste trabalho aplicar-se um filtro de média, substituindo o valor de $e_n(j + L)$ pela média aritmética dos $L - 1$ valores da janela, após um impulso ter sido detectado, conforme a equação 3.1. Com isto o *frame* contendo o som impulsivo será neutralizado, passando a ter um desvio padrão igual a zero, não mais influenciando no cálculo da variância quando da avaliação do *frame* subsequente.

$$e_n(j + L) = \begin{cases} e_n(j + L) & \text{se } \sigma^2(k) > l \\ \frac{1}{L-1} \sum_{m=1}^{L-1} e_n(m) & \text{Caso contrário} \end{cases} \quad (3.1)$$

Na figura 30 vê-se um exemplo onde o *frame* e_7 , contém um valor elevado de potência decorrente de um som impulsivo anteriormente detectado. Quando é realizada a normalização da janela nota-se que aquele *frame* assume um valor consideravelmente mais elevado que os demais, deixando-o equiparado, e em alguns casos, com valores maiores que aquele contido no *frame* avaliado e_{10} . Como consequência, a variância dos $L - 1$ *frames* permanece elevada o suficiente para poder mascarar uma detecção.

Figura 30 – Janela da variância ($L = 10$) onde o *frame* e_7 contém um som impulsivo previamente detectado.



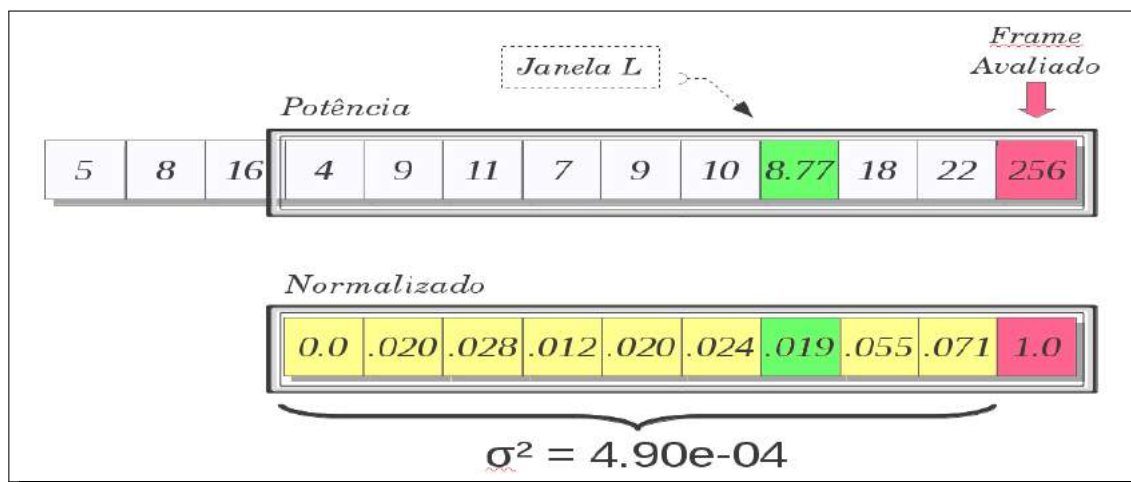
Fonte: o autor

Já na figura 31 é tomada a mesma situação, mas desta vez é aplicada a melhoria ao método, aqui proposta. Neste caso, uma vez detectado o som impulsivo, o *frame* avaliado passa a receber a média aritmética dos $L - 1$ *frames* anteriores, no caso o valor 8.77. Desta forma, quando este *frame* se deslocar para as posições mais internas da janela, ele deixa de influenciar o cálculo da variância, permitindo que esta assuma um valor menor e, como consequência, deixe de mascarar a detecção do novo som impulsivo.

O resultado pode ser visto na figura 32, onde nota-se que o método original apresenta uma queda em sua sensibilidade imediatamente após a detecção de um impulso, representada pela elevação dos valores da variância, visível na área II da figura 32b. Comparada com a mesma área da figura 32c, ali observa-se que, aplicada a melhoria, a linha do gráfico assume valores menos elevados, facilitando novas detecções. O aumento da sensibilidade do algoritmo é demonstrado na área I, onde um segundo vale no gráfico da variância não consegue ultrapassar o limiar no método original, mas que no método otimizado é identificado como sendo um som impulsivo.

Áreas onde nenhum som impulsivo é detectado, como é o caso da área III, permanecem idênticas em ambos os métodos.

Figura 31 – Janela da variância onde o *frame* e_7 que continha um som impulsivo previamente detectado é substituído pela média dos $L - 1$ *frames* anteriores.



Fonte: o autor

3.4 Módulo de Reconhecimento de Disparos

Um som, como qualquer outra onda, possui diversos elementos que podem ser analisados tanto no domínio do tempo, quanto no domínio das frequências.

3.4.1 Análise no Domínio do Tempo

Dentro do domínio do tempo, suas características envolvem a amplitude do sinal, sua duração, sua fase, sua envoltória e a taxa de amostragem utilizada, para sinais digitalizados.

A taxa de amostragem, a pesar de ser uma característica temporal do sinal sonoro, será relevante quando da análise do domínio das frequências.

A amplitude do sinal é função da distância entre emissor e o receptor, não sendo adequada ao presente propósito, já que por muitas vezes tal distância é desconhecida.

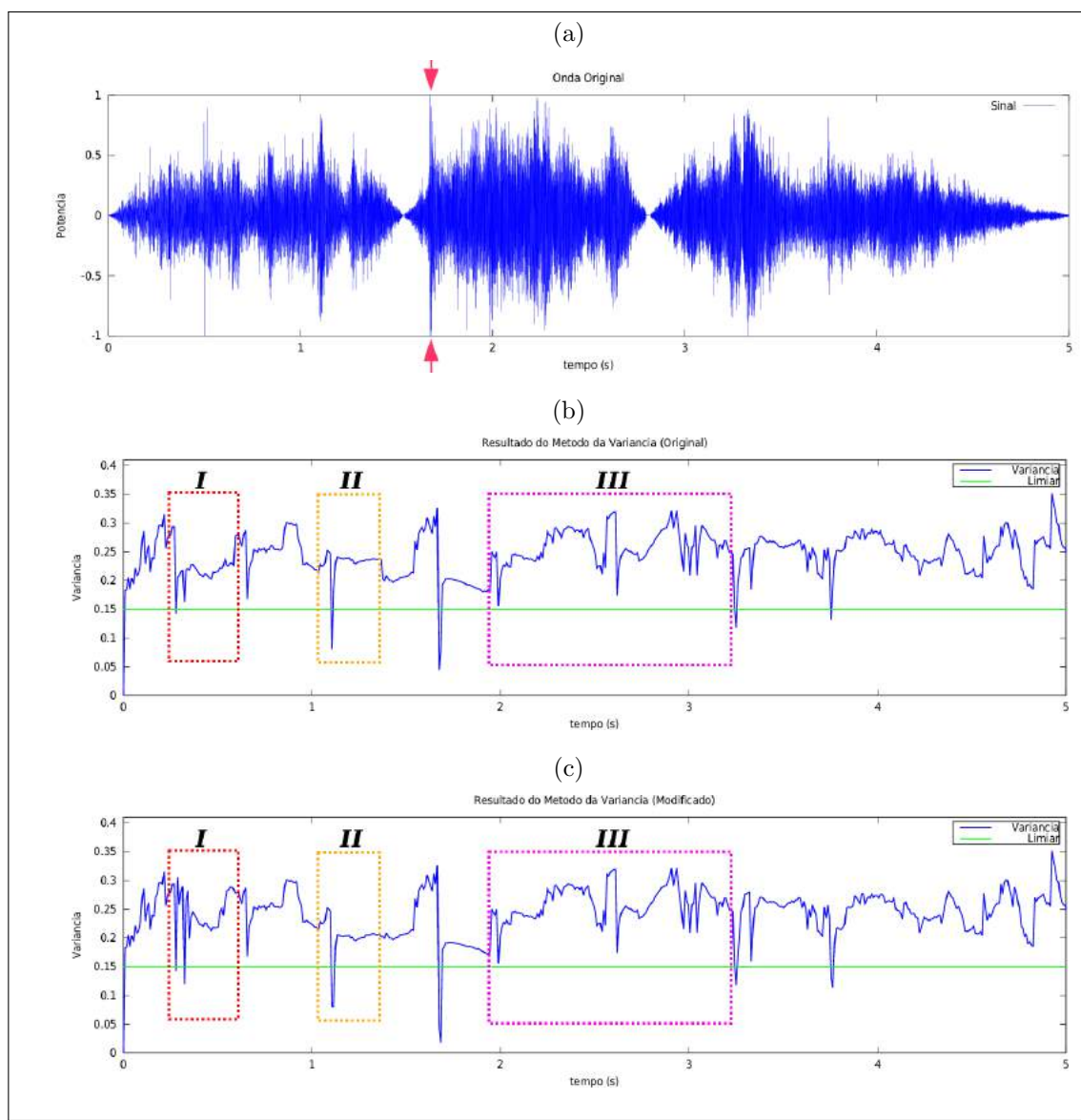
A fase do sinal é igualmente irrelevante para o propósito deste trabalho, já que o som impulsivo completo será considerado como uma unidade individual.

Já o tempo de duração do sinal e a sua envoltória possuem relevância no presente trabalho, uma vez que o som de tiros, quando comparados com outros tipos de sons impulsivos, apresentam forma e uma duração característicos.

A Envoltória

A envoltória de um sinal sonoro pode ser entendida como sendo o gráfico da evolução da intensidade deste sinal ao longo do tempo. Ela pode ser estimada por técnicas no domínio do tempo, como a aplicação de Filtros Passa-baixas ou cálculo do RMS, e por técnicas do domínio das frequências, como é o caso da Predição Linear no Domínio da Frequências (FDLP) ou a True Amplitude Envelope (TAE) [46].

Figura 32 – Gráfico comparativo dos resultados de detecção.



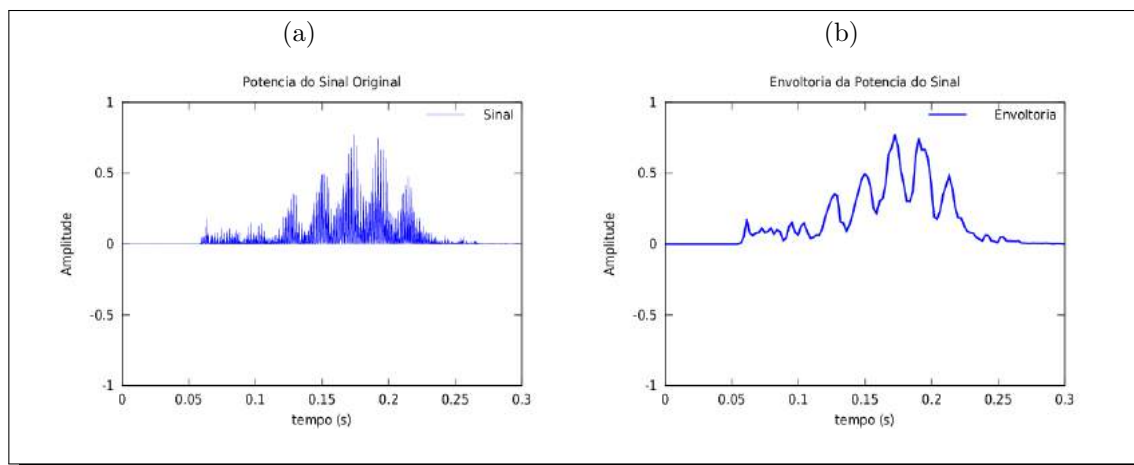
(a) Sinal Original; (b) Resultado obtido pelo método da VM Tradicional; (c) Resultado obtido com o uso da melhoria proposta. (I) e (II) áreas próximas à detecções onde a sensibilidade da melhoria pode ser observada; (III) área onde não houve detecção e o comportamento do algoritmo permanece o mesmo.

Fonte: o autor

Um exemplo de envoltória pode ser visto na figura 33, onde, a partir do gráfico da potência de um sinal qualquer (figura 33a), obteve-se a sua envoltória pela aplicação de um filtro Passa-baixas (figura 33b).

Uma aplicação possível para envoltórias é utilizá-las para estimar, dentro do domínio do tempo, o quanto um determinado sinal se assemelha a outro, determinando os limites das regiões definidas pelos modelos, para, em seguida, compará-los.

Figura 33 – Exemplo de Envoltória de um sinal.



(a) Potência do Sinal original; (b) Envoltória da Potência do Sinal.

Fonte: o autor

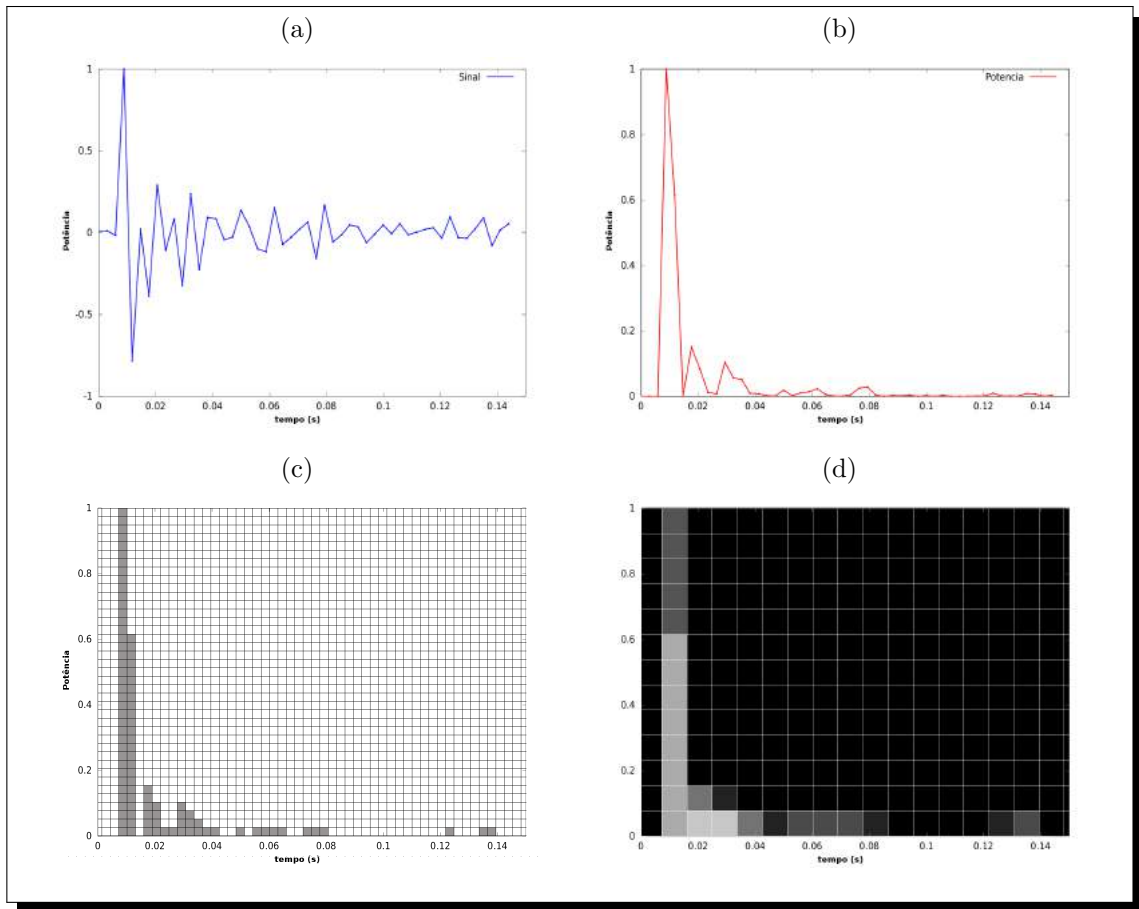
Comparação de sinais via Envoltória

Neste trabalho optou-se por propor uma nova técnica, na qual o sinal sonoro unidimensional será transformado em uma imagem bidimensional, sendo diretamente comparado com um sinal modelo. Este procedimento apresenta como vantagem a possibilidade de realizar a comparação entre duas polinomiais sem que seja necessária a modelagem matemática do sinal introduzido, onde, ao final do processo obtém-se dois valores numéricos que, se iguais a zero, representam um perfeito casamento entre as duas funções.

Para tanto, inicialmente o sinal sonoro é normalizado para valores entre 0 e 1 sendo calculada a sua potência (figura 34b). A seguir, o sinal é transformado em uma matriz onde cada coluna corresponde a uma amostra de sinal e cada linha corresponde a um milésimo da unidade de potência do sinal (figura 34c). Cada coluna da matriz tem suas linhas preenchidas com 1, de baixo para cima, de forma proporcional à intensidade da potência da respectiva amostras do sinal. Assim, amostras com potência 1 terão todas as linhas de sua respectiva coluna preenchidas com 1. Amostras com potência 0.5 terão as primeiras 500 linhas preenchidas com 1 e as demais preenchidas com 0 e assim sucessivamente. Finalmente, arbitra-se a definição da imagem final, sendo as linhas e colunas da matriz de potência agrupadas e somadas de acordo com esta nova definição e normalizadas entre 0 e 1, onde cada valor indica a escala de cinza assumida por um pixel da imagem (figura 34d). A figura 34 ilustra processo aplicado ao sinal contido na figura 34a.

Como no problema em questão a forma do som impulsivo ideal é bem conhecida e pode ser facilmente gerada pela equação de Friedlander (equação 2.1), pode-se criar tal impulso e compará-lo com a envoltória do sinal desejado. Na figura 35 vê-se um pulso de Friedlander gerado com potência máxima de 1 e 0.15 s de duração e a sua transformação em imagem.

Figura 34 – Transformação de um sinal em imagem.



(a) Sinal original; (b) Potência do sinal; (c) Matriz de potência. Os pontos em cinza indicam os elementos da matriz que foram assinalados com 1 e os brancos os assinalados com 0; (d) Potência transformada em escalas de cinza.

Fonte: o autor

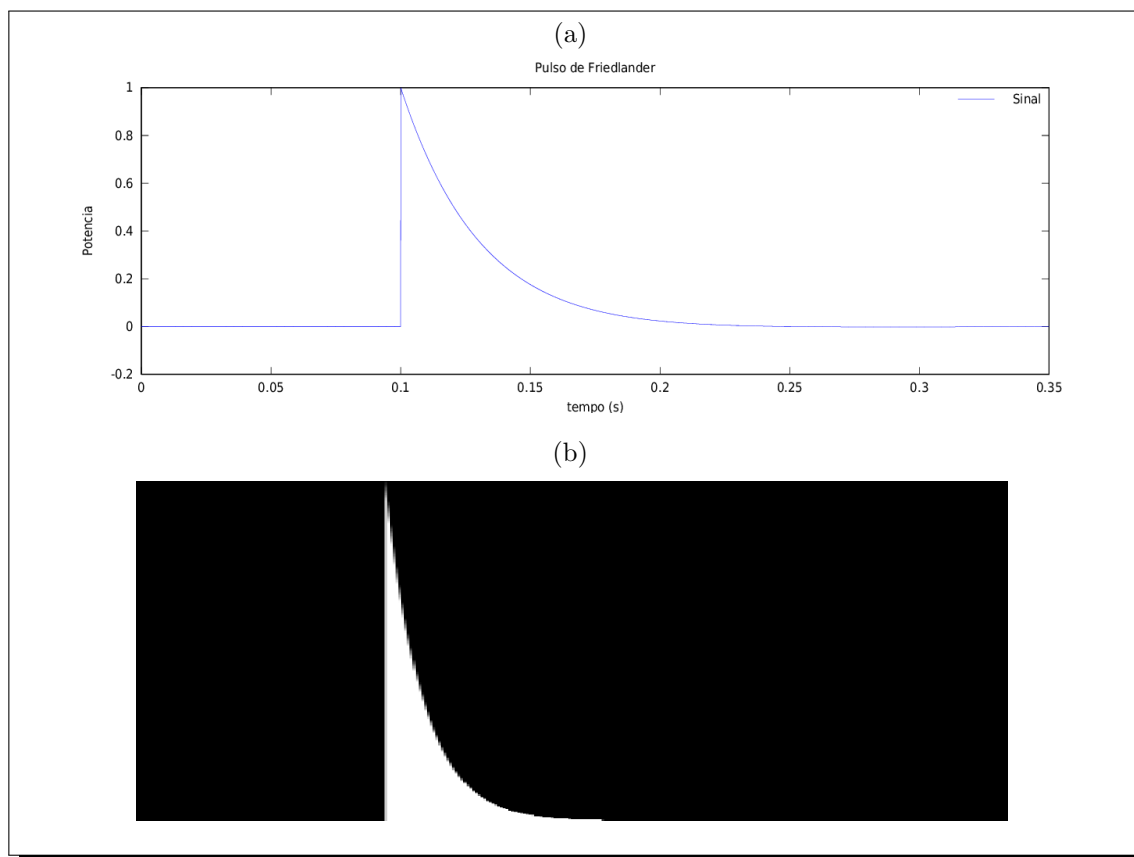
De posse deste sinal base e do sinal a ser analisado, a comparação é realizada pela subtração, célula a célula, de ambos. Considerando-se os dois sinais transformados como duas matrizes de dimensões $l \times c$, onde

$$\begin{aligned}
 S &= [s_{mn}]_{l \times c} & F &= [f_{mn}]_{l \times c} & V &= [v_{mn}]_{l \times c} \\
 e \quad V &= S - F \\
 tem - se \\
 \Theta &= \max[v_{mn}, 0]_{l \times c} \\
 e \\
 \Phi &= \max[-v_{mn}, 0]_{l \times c}
 \end{aligned} \tag{3.2}$$

Onde S é a matriz que compõe a imagem do sinal, F a matriz que compõe a imagem do pulso de Friedlander gerado, q_Θ e q_Φ coeficientes de multiplicação de V para obtenção dos valores Θ e Φ .

Um exemplo do processo pode ser visto na figura 36, onde a figura 36a mostra o sinal a

Figura 35 – Potência do pulso de Friedlander com duração de 0.15 s, simulando um disparo ideal de arma de fogo, transformado em imagem (F).



(a) Gráfico de Potência do Pulso de Friedlander; (b) Pulso de Friedlander transformado em imagem (F).

Fonte: o autor

ser analisado, transformado na imagem vista na figura 36b. Os resultados $q_{\Theta}v_{mn}$ e $q_{\Phi}v_{mn}$ podem ser vistos sob a forma de imagens nas figuras 36c e 36d, respectivamente.

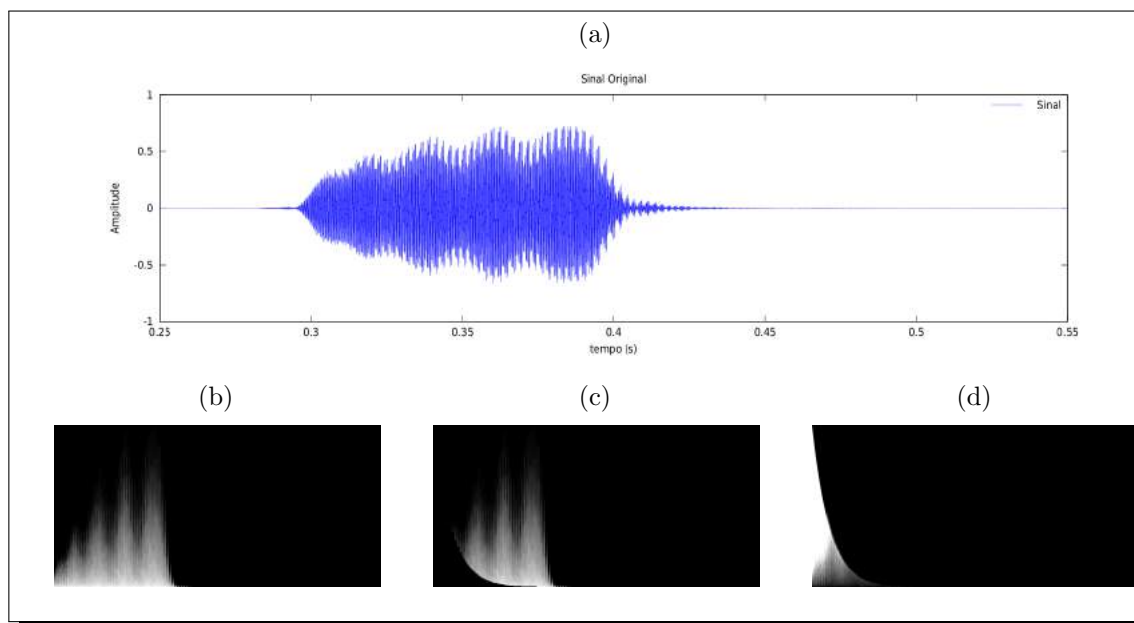
Esta análise resulta em dois números, Θ e Φ que indicam o quanto a envoltória do sinal analisado está próxima do formato ideal de um som impulsivo de duração similar ao gerado pelo disparo de uma arma de fogo.

Um Θ muito grande indica que o sinal analisado possui elevada potência em uma área além dos limites balizados pelo pulso de Friedlander base. Já um Φ muito grande indica que este sinal possui potência aquém da esperada dentro da área delimitada pelo modelo base.

Porém, por se tratar de uma análise no domínio do tempo, anomalias são esperadas. Por exemplo, um disparo de arma de fogo realizado a uma grande distância perderá boa parte de suas características, ficando menos potente e mais curto que outro sinal gravado a uma distância menor, gerando assim um Φ maior e um provável $\Theta = 0$.

Já um disparo realizado a uma curta distância e em um ambiente que provoque muito eco pode causar um aumento significativo em Θ .

Figura 36 – Exemplo da imagem do som impulsivo analisado subtraída do impulso ideal obtido pela equação de Friedlander.



(a) Sinal original; (b) Potência do sinal transformado em imagem (S);
 (c) $\max[v_{mn}, 0]_{l \times c}$; (d) $\max[-v_{mn}, 0]_{l \times c}$.

Fonte: o autor

Há ainda sons impulsivos que possuirão uma duração e envoltória que podem ser facilmente confundidos com um disparo, como é o caso de portas batendo e fogos de artifício.

3.4.2 Análise no Domínio das Frequências

Existem várias técnicas de transformação e análise de um sinal dentro do domínio das frequências, onde se pode optar pela utilização de uma série de ferramentas para tratamento de sinais, como transformadas de Fourier, transformadas de Cossenos, Wavelets, Análise Espectral, Cepstrum e *Mel-frequencys*, Coeficientes de Predição Linear (Linear Prediction Coefficients - LPC), Transformações lineares, como a *Principal Components Analysis* (PCA), *Hidden Markov Models* (HMM), dentre outros.

No presente trabalho optou-se pela proposição de uma técnica simples, onde buscou-se a caracterização de uma assinatura sonora, baseada em um conjunto de frequências dominantes em uma bateria de sons contendo disparos e sons impulsivos, que serve como padrão para o reconhecimento de outros disparos.

Esta técnica é baseada na Transformada de Fourier e consiste, basicamente, em identificar quais são as frequências características de um conjunto de amostras de disparos de armas de fogo, subtraí-las das frequências características de um conjunto de sinais impulsivos e, a partir das frequências obtidas, identificar sua existência em uma outra

amostra qualquer.

Calculo do Coeficiente de Variação

Entende-se por coeficiente de variação (c_v) o valor obtido pela razão entre o desvio-padrão σ e a média μ de uma dada amostragem, cuja fórmula pode ser vista na equação 3.3. Uma vez que duas distribuições normalmente possuem médias e valores bastantes diferentes a simples comparação dos seus desvio-padrão não é viável, já que estes também possuirão grandezas diferentes. O c_v , entretanto, presta-se para este fim, já que é calculado em função da média da amostra, indicando sua dispersão de forma absoluta. Quanto menor o c_v , menor a dispersão dos dados amostrais, independentemente da grandeza dos valores utilizados.

$$c_v = \frac{\sigma}{\mu} \quad (3.3)$$

Assim, para obter-se um conjunto de frequências características de um conjunto de disparos, divide-se um conjuntos de sinais normalizados em dois subconjuntos, **D** contendo os disparos e **I** contendo os sons impulsivos, calculado-se as respectivas FFTs para todos estes sinais.

A seguir, calcula-se a c_v para cada uma das frequências obtidas, tomando-se como base a amostragem composta pelos valores das FFT de mesma frequência, dentro de cada subconjunto, para, em seguida, ordená-las pelo valor da c_v . Ao final desta fase, obtém-se algo semelhante ao exibido na figura 37, onde as colunas $D_1, D_2, D_3, D_4, I_1, I_2, I_3, I_4$ exibem os valores obtidos pelo cálculo da FFT de 4 disparos e 4 sons impulsivos, respectivamente. As linhas indicam os valores obtidos pelo cálculo da FFT para cada frequência, sendo a c_v calculada com base nos valores contidos nestas linhas, conforme ilustra a seta. A coluna σ encontra-se presente para fins de comparação dos valores do desvio padrão e de c_v , demonstrando como os valores do desvio padrão não se prestam à comparação de proximidade entre valores de grandezas muito diferentes. No exemplo em questão as frequências já se encontram ordenadas pelos valores de c_v .

Como os valores da c_v indicam uma menor dispersão absoluta nos dados amostrais, presume-se que quanto menor os valores de c_v assim calculados, mais próximos foram valores obtidos no cálculo da FFT de uma dada frequência em todo o subconjunto **D**, indicando que aquela frequência possui potência similar em todos os disparos testados, sendo-lhes peculiar. O mesmo raciocínio aplica-se ao subconjunto **I**.

Obtenção das $f(N)$ frequências características

De posse do conjunto de frequências ordenadas pelos c_v para disparos e para sons impulsivos, passa-se, então, a determinar as $f(N)$ frequências características de disparos que não sejam características de sons impulsivos.

Figura 37 – Exemplo do cálculo do Coeficiente de Variação e Seleção de Frequências.

Frequência	Disparos					
	D1	D2	D3	D4	σ	C_v
30	72,30	73,10	69,50	77,00	3,096	0,042
20	0,010	0,015	0,012	0,011	0,002	0,180
50	0,370	0,435	0,368	0,276	0,065	0,181
100	601,50	261,82	631,41	798,82	225,13	0,393
60	0,509	0,330	0,621	0,980	0,274	0,450
80	0,650	0,866	0,888	0,095	0,369	0,591
90	23,13	11,57	56,39	45,98	20,543	0,599
40	0,545	0,084	0,991	0,346	0,383	0,778
10	0,950	0,010	0,580	0,370	0,824	0,824
70	0,031	0,548	0,129	0,605	0,290	0,884

Frequência	Sons Impulsivos					
	I1	I2	I3	I4	σ	C_v
10	0,939	0,774	0,840	0,827	0,069	0,081
30	98,130	79,511	96,366	63,328	16,326	0,194
60	0,798	0,928	0,251	0,616	0,294	0,454
70	7,873	2,223	4,290	6,782	2,536	0,479
50	0,188	0,986	0,980	0,519	0,388	0,580
20	20,943	8,450	19,899	40,958	13,507	0,599
90	0,295	0,706	0,849	0,202	0,313	0,611
40	0,706	0,325	0,991	0,120	0,389	0,726
100	0,178	0,302	0,940	0,318	0,343	0,789
80	0,954	0,139	0,312	0,318	0,359	0,832

Fonte: o autor

Para isso, toma-se as N primeiras frequências de menor valor de c_v de cada subconjunto. Desta forma tem-se dois conjuntos de frequências, sendo que o primeiro contém as N frequências características de disparos (F_D) e o segundo contendo as N frequências características de sons impulsivos (F_I).

O conjunto de frequências que será considerado como característica exclusiva de um disparo, contendo $f(N)$ elementos, é dado por

$$F_C = F_D - F_I \tag{3.4}$$

No exemplo da figura 37, os $N = 4$ valores mais característicos dos disparos e sons impulsivos encontram-se destacados e resultam nos seguintes conjuntos de frequências:

$$F_D = \{30, 20, 50, 100\}$$

$$F_I = \{10, 30, 60, 70\}$$

Nota-se a frequência de 30 Hz é comum aos dois conjuntos, não sendo, portanto, útil à identificação de disparos. Logo o conjunto característico de um disparo é, neste caso,

$$F_C = F_D - F_I = \{30, 20, 50, 100\} - \{10, 30, 60, 70\} = \{20, 50, 100\}$$

contendo $f(N) = 3$ frequências.

4 Metodologia

4.1 Software Utilizado

Na execução deste trabalho foram utilizadas as seguintes ferramentas de software:

Fedora™ Linux versão 19 – sistema operacional base para o presente trabalho [47];

Audacity® – é um software livre, destinado à realização de trabalhos de gravação, edição e manipulação de arquivos sonoros. Possui interface gráfica que torna a sua utilização bastante intuitiva. Porém, em algumas distribuições Linux não há o módulo de conversão de arquivos do tipo MPEG 1 Layer III (MP3) incorporado [48].

Foi utilizado para realizar o recorte de sons impulsivos dos arquivos sonoros escolhidos;

Lame – é um codificador/decodificador MP3, responsável por converter arquivos MP3 para arquivos de áudio da Microsoft e IBM, o *Waveform Audio Format* (**WAV**). É um aplicativo em linha de comando, não possuindo interface gráfica[49].

Foi utilizado na conversão de arquivos MP3 para o formato WAV a fim de poder serem manipulados pelo Audacity e Octave.

GNU Octave – é uma linguagem de programação criada especificamente para realizar o processamento de algoritmos matemáticos. É uma poderosa ferramenta destinada ao desenvolvimento de algoritmos, aquisição, análise e visualização de dados, modelagem, simulação e prototipagem, dentre outras atividades no campo acadêmico, sendo particularmente útil para o processamento e manipulação de matrizes [50].

Sua funcionalidade e sintaxe é, em muitos aspectos, idêntica à do MATLAB®, o que o torna uma opção de baixo custo para o desenvolvimento de aplicações. Porém, o octave não possui interface gráfica própria.

Foi utilizado para criar as rotinas de detecção dos diversos algoritmos e na apuração de resultados obtidos.

Qt Octave – é, simplesmente, uma interface gráfica para o GNU Octave [51].

PostgreSQL – é um sistema gerenciador de banco de dados relacional de código aberto desenvolvido pelo PostgreSQL Global Development Group [52].

Foi utilizado na criação de uma base de dados contendo os resultados do teste de otimização de parâmetros, descrito mais a diante, a fim de auxiliar na análise dos resultados.

Libre Office – é uma suite de escritório contendo um editor de planilha, editor de texto, editor de apresentação e editor de desenhos, sendo um software livre e de código aberto [53].

Foi utilizado para planilhar os resultados, assim como para a edição de algumas das figuras apresentadas neste trabalho.

yEd – é um software livre que realiza as tarefas de editor de diagramas [54].

Foi utilizado para a geração dos diagramas de fluxo exibidos no trabalho.

SoX – o *SOund eXchange* (SoX) é um software livre, *cross-platform* para Windows, Linux e MacOS. É um utilitário, em linha de comando capaz de aplicar vários efeitos a um arquivo sonoro.

Neste trabalho ele foi utilizado para realizar a filtragem de ruído de um sinal sonoro pelo método *Spectral Noise Gating*, descrito em 2.5.2.

WEKA – o WEKA (Waikato Environment for Knowledge Analysis) é um pacote de softwares, desenvolvidos em JavaTM na Universidade de Waikato, Nova Zelândia e licenciado pela General Public License (GPL), sendo, portanto, um software livre.

Ele é uma ferramenta destinada a realizar Data Mining, agrupando diversos algoritmos de aprendizado, contendo ferramentas para o pré-processamento, classificação, regressão, clusterização e associação de dados, bem como ferramentas de visualização e análise de dados[55].

Neste trabalho ele foi utilizado para realizar a clusterização de dados, através do algoritmo *K-Médias*, e sua classificação pelo método *Support Vector Machine* (SVM).

4.2 Módulo de Detecção de Impulso

4.2.1 Seleção de sons

Para realizar a detecção de sons impulsivos em meio a um ambiente ruidoso, foram selecionados sons existentes nas bibliotecas de efeitos sonoros “*De Wolfe Classic Sound Effects*” e “*BBC Sound Effects Library*”, citadas em [7] e [1].

Em virtude do recorte realizado no escopo deste trabalho, o módulo de captura de sons não encontra-se entre os objetivos aqui contemplados. Desta forma, para realizar a detecção de sons impulsivos em meio a um ambiente ruidoso, foram selecionados sons existentes nestas bibliotecas de efeitos sonoros.

Delas foram escolhidos arquivos que atenderam aos seguintes critérios:

- Possuam taxa de amostragem igual a 44.100 Hz;

- Arquivos que contenham amostras de sons impulsivos, mas que não tenham sido gerados por armas de fogo, como latidos, vidros quebrando, fogos de artifício, portas batendo, etc;
- Arquivos que contenham amostras de disparos de armas de fogo.

Estes foram convertidos para o formato padrão para arquivos de áudio da Microsoft e IBM, o *Waveform Audio Format* (**WAV**), sendo deles extraídos trechos contendo disparos de armas de fogo e outros sons impulsivos, todos com duração aproximada de 0.5 s. Selecionou-se, a seguir, cinco disparos de armas de fogo diferentes e mais cinco sons impulsivos, sendo um latido de cão, duas explosões de fogos de artifício, um som de vidro quebrando e uma porta batendo, conforme visto na figura 38.

Dos cinco disparos um foi escolhido para realizar o teste de cadência, teste de otimização de parâmetros dos métodos de detecção e teste de validação dos parâmetros otimizados, visto na figura 38a. Os demais foram utilizados nos testes de generalização do método escolhido.

4.2.2 Teste de Cadência

Cadência conforme descrito anteriormente em 2.3.4, pode ser definida como a quantidade de disparos que uma arma pode realizar em um minuto. Desta forma, é necessário que o algoritmo de detecção seja capaz de realizar sua tarefa mesmo quando a cadência de tiro for elevada.

A fim de verificar a validade dos métodos de detecção foram criadas rotinas no Octave rodando sobre uma plataforma Linux, onde foram gerados sinais acústicos compostos por um número aleatório de repetições da amostra de disparo escolhida H , inserida de 5 a 10 vezes, separadas por um intervalo de tempo também aleatório, definido dentro de três faixas distintas:

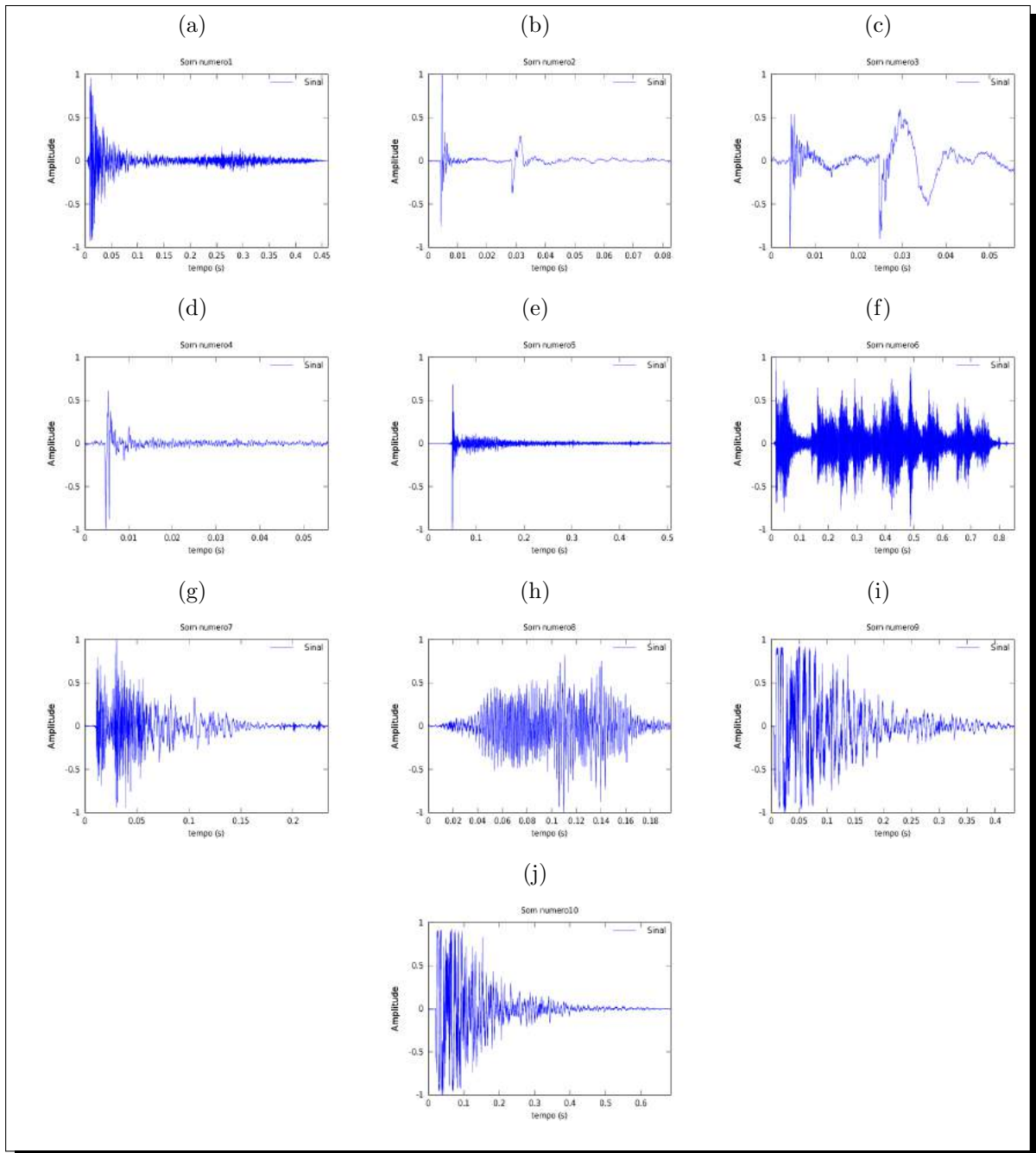
Faixa 1 – o disparo H foi repetido a intervalos aleatórios entre 0.1s e 0.5s, correspondendo a uma cadência entre 600 e a 120 tiros por minutos (tpm), simulando a cadência de armas automáticas;

Faixa 2 – o disparo H foi repetido a intervalos aleatórios entre 0.5s e 1s ou 120 a 60 tpm, simulando a cadência de armas semiautomáticas; e

Faixa 3 – o disparo H foi repetido a intervalos aleatórios entre 1s e 5s ou 60 a 12 tpm, simulando a cadência de armas de repetição. Cobriu-se, assim, toda a faixa de cadência das armas leves, conforme descrito na tabela 1.

Entre cada amostra foram inseridos valores de amplitude próximas a zero. Assim cada sinal gerado possui o perfil genérico que pode ser visto na figura 39a. Com base neste primeiro sinal gerado, denominado **Sem Ruído (SR)**, foram gerados mais dois, o primeiro

Figura 38 – Representação gráfica dos sons utilizados nos testes.



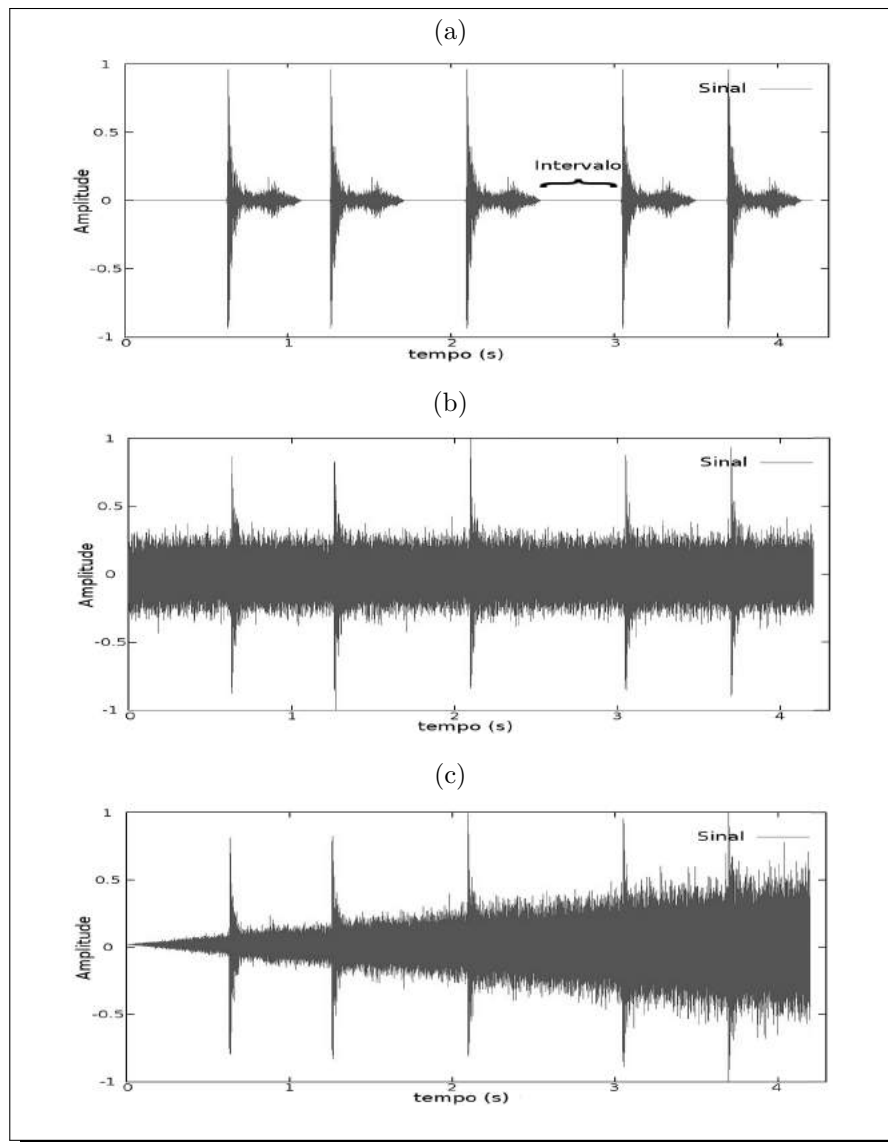
(a), (b), (c), (d) e (e) Disparos de armas de fogo; (f) Vidro quebrando; (g) Porta batendo; (h) Latido de cachorro; (i) e (j) Fogos de artifício.

Fonte: o autor

obtido pelo acréscimo de *Ruído Branco* com potencia igual a 50% da potência máxima, denominado de **Ruído Constante (R1)** (figura 39b) e o segundo obtido pelo acréscimo de *Ruído Branco* em volume crescente, variando de 0 a 100% da potência máxima do sinal, denominado **Ruído Crescente (R2)** (figura 39c).

Desta forma, foi criado um conjunto de 450 sinais denominado **Bateria 1**, contendo um total de 3486 sons impulsivos, sendo gerados, para cada Faixa, 50 sinais SR contendo

Figura 39 – Exemplos de sinais gerados para os testes comparativos entre métodos.



(a) Sinal com cinco disparos sem ruído; (b) Sinal com Ruído branco de amplitude constante; (c) Sinal com Ruído branco de amplitude crescente.

Fonte: o autor

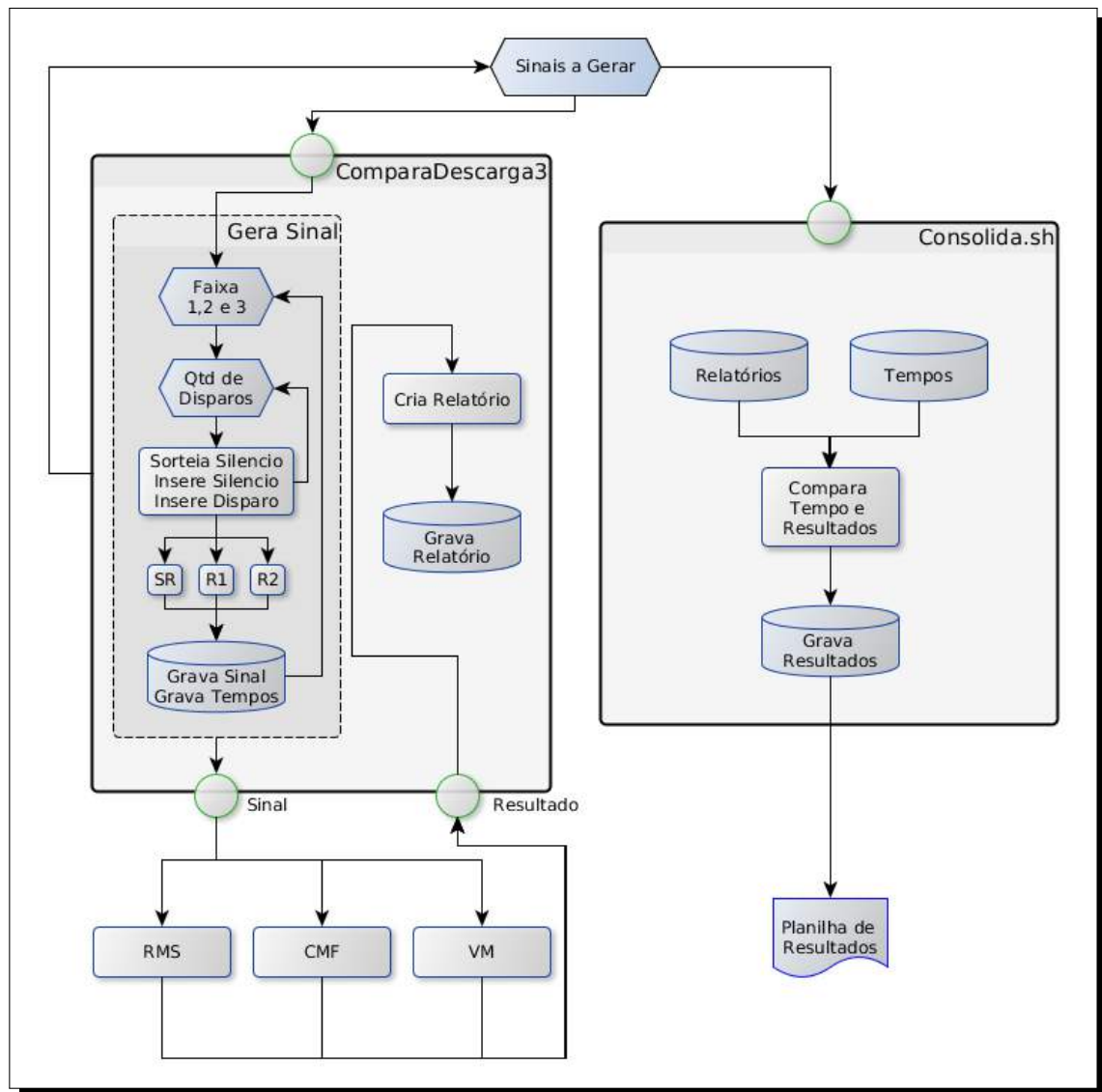
um total de 1155 impulsos para a Faixa 1, 1173 impulsos para a Faixa 2 e 1158 para a Faixa 3. Estes foram replicados e a eles adicionado ruído R1 e R2.

O diagrama lógico para geração dos sinais e algoritmos de detecção pode ser visto na figura 40.

4.2.3 Teste de Otimização de Parâmetros

O tamanho da janela L e a quantidade de amostras N coletadas para a composição do *frame* são parâmetros utilizados nos algoritmos da CMF e VM que possuem um grande impacto no resultado da detecção e desta forma são passíveis de ajustes, visando a

Figura 40 – Diagrama do Teste de Cadência.



Fonte: o autor

otimização da performance dos métodos.

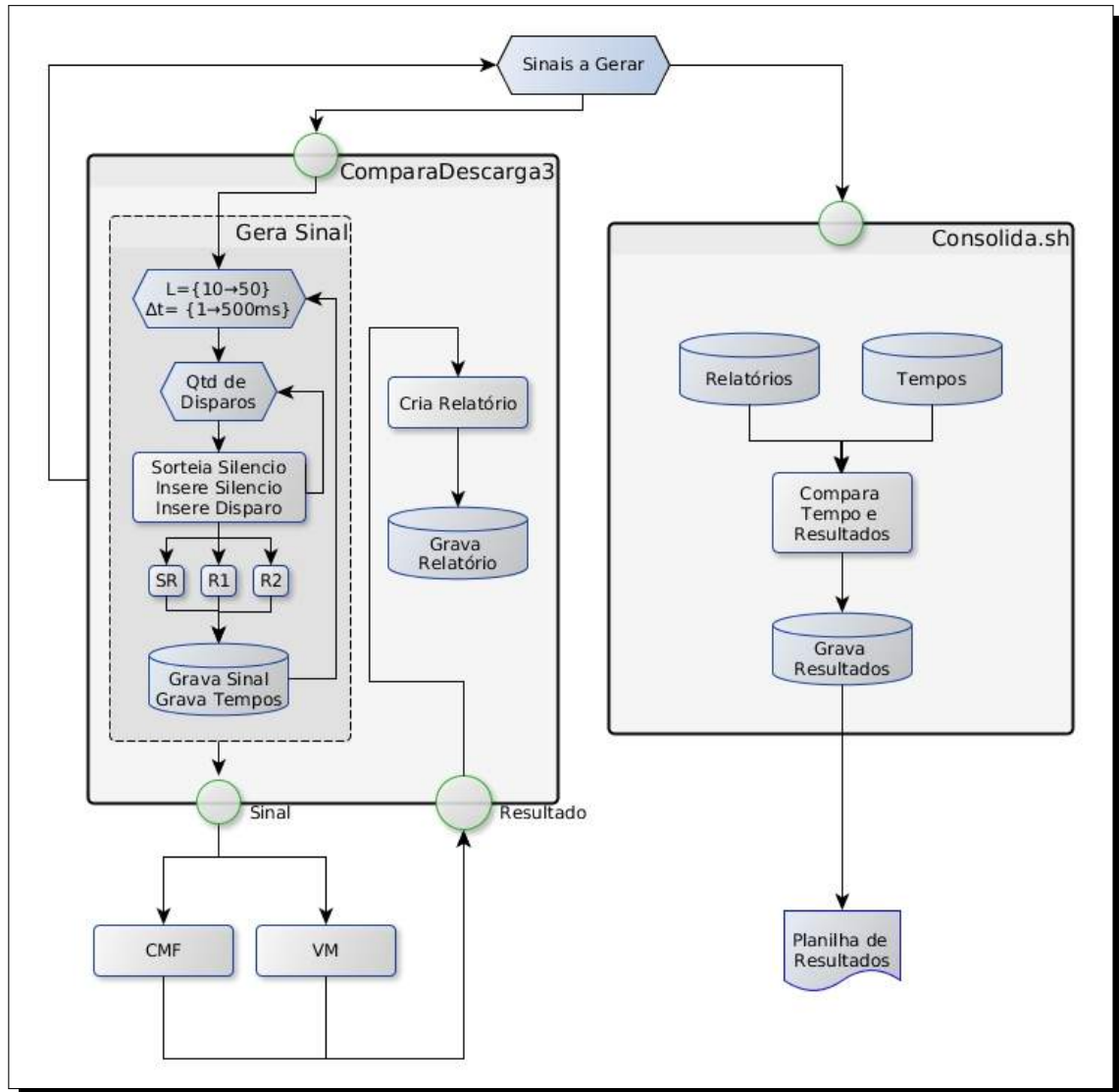
Desta forma, foi gerada outra bateria de sinais, semelhante à **Bateria 1**, contendo 30 amostras **SR**, **R1** e **R2**, a fim de serem submetidas à CMF e VM variando-se os parâmetros de tamanho de janela $L = \{10, 15, 20, 25, 30, 35, 40, 45, 50\}$ e *frame* com $\Delta t = \{1\text{ ms} \rightarrow 500\text{ ms}\}$, ou $N = \{44 \rightarrow 22050\}$ amostras, que foi denominada **Bateria 2**.

Nesta bateria foram gerados apenas sinais com intervalos relativos à **Faixa 1**, já que esta faixa apresenta a maior dificuldade na detecção, por estarem os disparos consecutivos bastante próximos entre si, podendo, em muitos casos, serem considerados como sendo um disparo só.

Para cada sinal gerado também foram criados seus respectivos *R1* e *R2*, sendo as três

amostras submetidas aos algoritmos, conforme mostra o diagrama lógico desta etapa, visto na figura 41.

Figura 41 – Diagrama do Teste de Otimização.



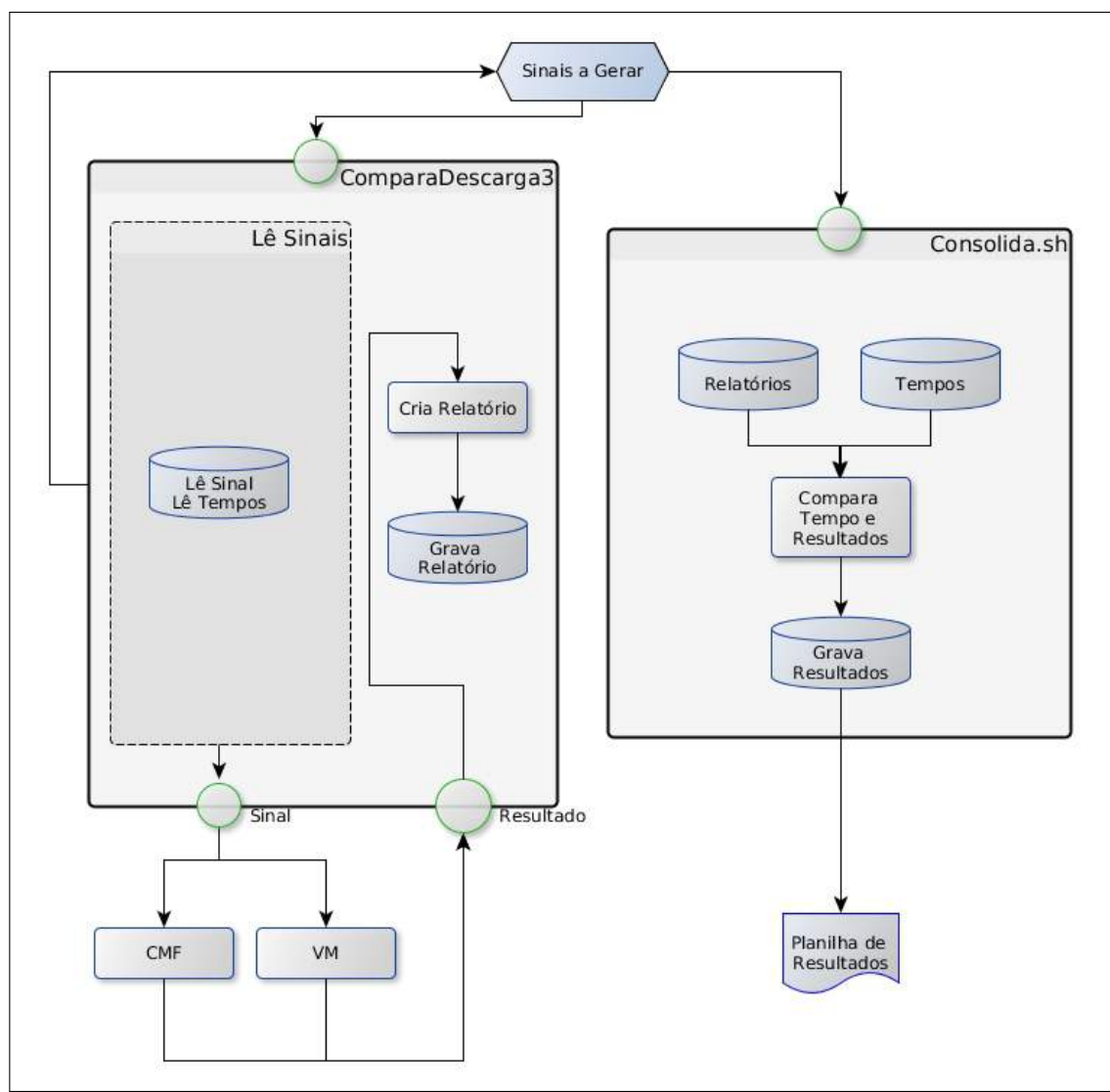
Fonte: o autor

4.2.4 Validação dos Parâmetros

Nesta etapa a **Bateria 1** foi novamente submetida à CMF e VM, desta vez configuradas com os parâmetros considerados de melhor desempenho, tendo os seus resultados comparados com os resultados iniciais. O método que apresentou melhor desempenho nesta etapa foi considerado como sendo o melhor método de detecção.

O diagrama lógico desta etapa pode ser visto na figura 42.

Figura 42 – Diagrama do Teste de Validação de Parâmetros



Fonte: o autor

4.2.5 Teste de Generalização do Melhor Método

A fim de verificar a eficiência do método definido como sendo o melhor, criou-se mais duas baterias de sons, definidas da seguinte maneira:

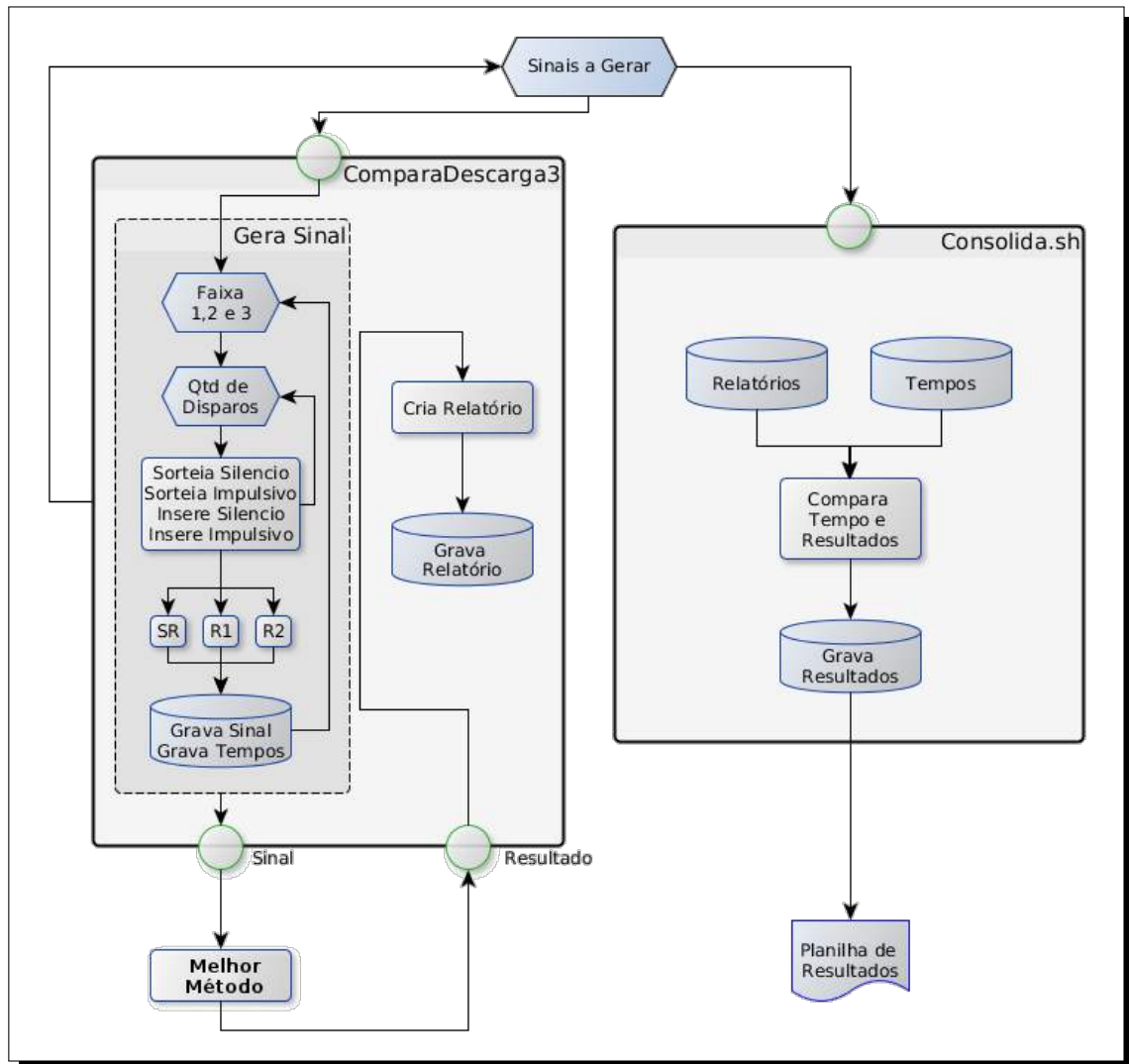
Armas Diversas – criou-se um sinal com intervalos entre sons impulsivos nas faixas 1, 2 e 3, com níveis de ruído **SR**, **R1** e **R2**, semelhantes ao descritos no Teste de Cadência (4.2.2), sendo que, desta vez, foram utilizados todos os cinco disparos selecionados em 4.2.1, sorteados antes de cada inserção no sinal.

Sons Impulsivos Diversos – mais uma vez, criou-se um sinal com intervalos entre sons impulsivos nas faixas 1, 2 e 3, com níveis de ruído **SR**, **R1** e **R2**, também semelhantes ao descritos no Teste de Cadência (4.2.2), sendo que agora foram utilizados os cinco sons impulsivos descritos em 4.2.1, além dos cinco disparos citados no item anterior.

Assim, antes de cada inserção, um entre estes 10 sons foi sorteado, sendo então inserido no sinal.

O diagrama lógico desta etapa pode ser visto na figura 43.

Figura 43 – Diagrama do Teste de Generalização do Método.



Fonte: o autor

4.2.6 Critérios de Avaliação

Durante a geração dos sinais sonoros, o sistema controla e armazena os tempos de inserção de cada som impulsivo. Desta forma é possível classificar previamente um dado *frame* deste sinal sonoro dentro de duas classes distintas: Com Impulso (CI) ou Sem Impulso (SI).

Após a apresentação dos sinais aos diversos métodos de detecção, pode-se comparar a classe de cada *frame* com o resultado assinalado pelos algoritmos, classificando este resultado segundo a matriz de confusão:

Tabela 2 – Matriz de Confusão

<i>Tipo de Frame</i>	<i>Assinalado (A)</i>	<i>Não Assinalado (NA)</i>
Com Impulso (CI)	<i>Detected (D)</i>	<i>Missed (M)</i>
Sem Impulso (SI)	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Com ela, cada resposta dada pelo algoritmo é classificada em quatro categorias distintas [56]:

Detected (D) – identifica que o som impulsivo inserido foi corretamente assinalado pelo método;

Missed (M) – mostra que, embora exista um som impulsivo no *frame* em questão, o método não pôde detectá-lo corretamente, não sendo por ele assinalado;

False Positive (FP) – identifica que, a pesar de não haver nenhum som impulsivo no *frame* avaliado, o algoritmo entendeu que havia;

True Negative (TN) – indica que o *frame* não possuía nenhum som impulsivo e o método identificou esta situação corretamente.

Desta forma, nota-se que as categorias **M** e **FP** indicam os erros cometidos pelo algoritmo durante o processo de análise dos n frames componentes do sinal sonoro e as categorias **D** e **TN** os seus acertos.

De posse desta classificação podemos, então, utilizar as seguintes medidas de desempenho descritas pelas fórmulas 4.1, 4.2, 4.3 e 4.4 para analisar e comparar a performance dos métodos [56]:

acurácia (ac) – mede o quanto o algoritmo prevê corretamente ambas as classes. Quanto mais próximo de um, menor foram os erros cometidos durante a classificação. Se multiplicado por 100, representa o percentual de acertos do método;

$$ac(\hat{f}) = \frac{D+TN}{n} \quad (4.1)$$

precisão (pr) – mostra a exatidão do modelo na detecção de impulsos. Quanto mais próximo de um, menor o índice de falsos positivos detectados;

$$pr(\hat{f}) = \frac{D}{D+FP} \quad (4.2)$$

sensibilidade (se) – indica a taxa de sons impulsivos verdadeiros perdidos. Números próximos a 1 indicam baixas taxas de Missed(M);

$$se(\hat{f}) = \frac{D}{D+M} \quad (4.3)$$

especificidade (sp) – mede a precisão na correta classificação de frames sem impulsos, ou seja, é uma forma alternativa de analisar o número de falsos positivos.

$$sp(\hat{f}) = \frac{TN}{TN+FP} \quad (4.4)$$

Em todos os casos descritos, os resultados variam entre 0 e 1, sendo que valores próximos a 1 significam um melhor desempenho do modelo.

4.3 Módulo de Reconhecimento de Disparo

4.3.1 Seleção de sons

Mais uma vez foram utilizados os sons existentes nas bibliotecas de efeitos sonoros “*De Wolfe Classic Sound Effects*” e “*BBC Sound Effects Library*”, citadas em [7] e [1]. Destes foram selecionados 52 arquivos contendo disparos de armas de fogo e 40 arquivos contendo sons impulsivos diversos.

Estes arquivos foram submetidos a uma detecção de impulsos pelo melhor método selecionado. Para cada impulso detectado, foi realizada a extração do trecho do sinal sonoro conforme o descrito na equação 4.5

$$t_i = D_n - 0.3 \text{ s}$$

$$t_f = \begin{cases} D_n + 0.3 \text{ s} & \text{se } D_{n+1} > D_n + 0.3 \text{ s} \\ D_{n+1} & \text{caso contrário} \end{cases} \quad (4.5)$$

$$\text{para } n = \{1, 2, 3, 4, 5\}$$

onde t_i é o tempo inicial para a extração, t_f o tempo final de extração, D_n é o tempo de detecção dos cinco primeiros sons impulsivos detectados no sinal. Após a extração, o registro foi normalizado para valores entre -1 e 1 e gravado novamente em disco no formato WAV.

Deste processamento foram obtidos 87 registros com disparos e 165 registros com sons impulsivos diversos, todos com a duração máxima de 0.6 s. Estes foram divididos em três baterias de sons, denominadas **Bateria Completa**, contendo todos os 252 sons acima descritos, **Bateria de Treino**, contendo 58 disparos e 110 sons impulsivos, selecionados aleatoriamente da **Bateria Completa**, e **Bateria de Teste**, contendo os demais 29 disparos e 55 sons impulsivos, perfazendo a proporção de 2 sinais de treinamento para cada sinal de teste.

4.3.2 Reconhecimento no Domínio do Tempo

Para a análise no domínio do tempo, cada som foi submetido ao procedimento descrito a seguir, visto de forma esquemática na figura 44.

Remoção de ruído – os 0.3 s iniciais de cada sinal sonoro foi considerado como sendo ruído de ambiente, submetido, junto com todo o sinal, a um *Spectral Noise Gating* (SNG), através da utilização do software *SoX*. Como resultado foi obtido um sinal sonoro livre de ruído que foi novamente normalizado para valores entre -1 e 1.

Início do som impulsivo – de posse do sinal limpo e normalizado, sua potência foi calculada e o início exato do som impulsivo determinado pelo momento em que a potência do sinal atingiu 10% de seu valor máximo. Este ponto ficou definido como sendo o ponto zero para fins de comparação com o pulso de Friedlander padrão.

Transformação em figura – a potência do sinal foi então transformada em figura, conforme descrito em 3.4.1, sendo, a seguir, comparado com um pulso de Friedlander de potência máxima igual a 1, duração igual a 0.15 s e comprimento total igual a 0.6 s, de onde se obteve os valores de Θ e Φ .

Normalização dos valores – os valores de Θ e Φ foram normalizados, para que a grandeza dos dados não influenciasse nos resultados de clusterização e classificação. O valor extremo possível para Θ é 258540, obtido subtraindo-se F de uma figura composta somente por pixels de valor 1. Já o valor máximo para Φ é 6463.3, obtido pela subtração de F de uma figura composta somente por pixels de valor 0. Assim, tais valores normalizados entre 0 e 1, são obtidos pela equação 4.6

$$\Theta_{Norm} = \frac{\Theta}{258540} \quad (4.6)$$

$$\Phi_{Norm} = \frac{\Phi}{6463.3}$$

Tais procedimentos foram aplicados tanto à Bateria de Treino, quanto à Bateria de Teste, sendo os resultados formatados como entrada para o WEKA, sob a forma de dois arquivos *ARFF*, conforme o exibido na figura 45.

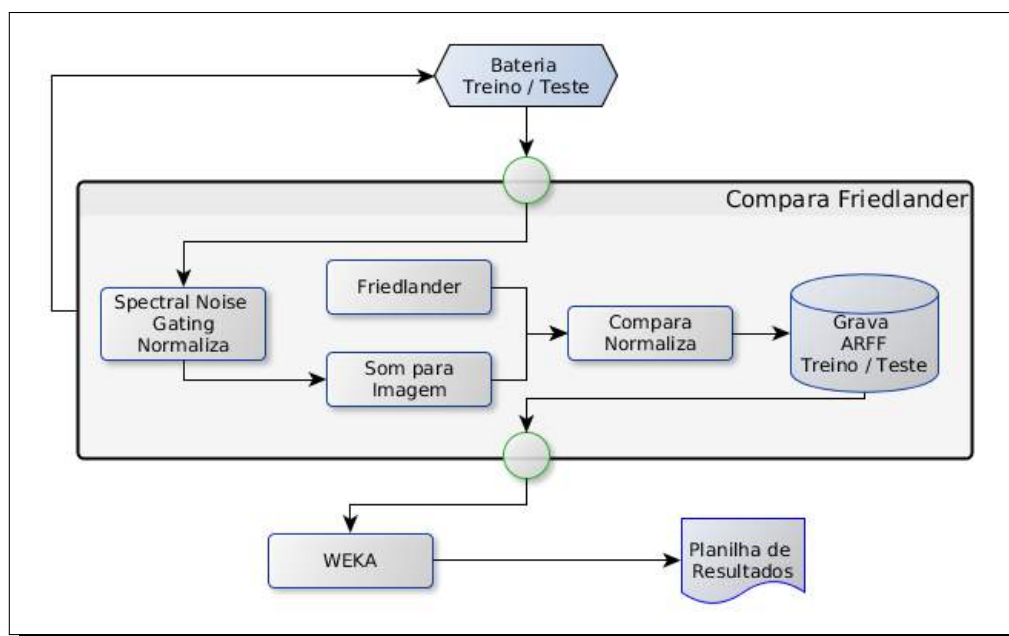
Os arquivos *ARFF* da Bateria de Treino foram submetidos ao WEKA para clusterização e classificação utilizando-se os algoritmos K-Médias e SVM, respectivamente, criando os Centroides e Hiperplanos Separatórios que foram utilizados para a clusterização e classificação dos dados da Bateria de Teste, conforme ilustrado pelas figuras 46 e 47, geradas pelo *WEKA Knowledge Flow*.

4.3.3 Reconhecimento no Domínio das Frequências

Determinação das Principais Frequências

Este procedimento foi efetuado somente sobre a Bateria de Treino visando identificar as frequências que são características de um disparo e que, ao mesmo tempo, não são características de um som impulsivo qualquer. Para tanto foram efetuados as seguintes ações:

Figura 44 – Diagrama do Teste no domínio do Tempo.



Fonte: o autor

Figura 45 – Esquema de geração dos arquivos *ARFF* no domínio do tempo.

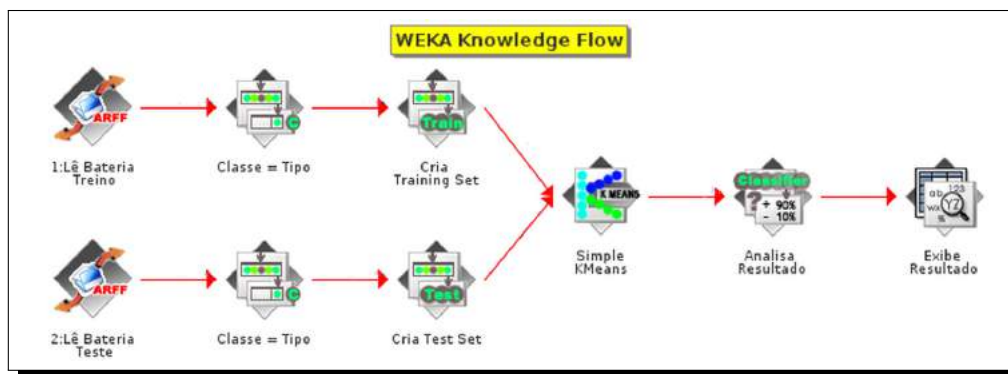
Fonte: o autor

Zero Padding – foi feito o *zero padding* em cada um dos sinais de forma a se obter o número fixo de 44101 amostras. Desta forma o cálculo da FFT retornará sempre 22050 frequências diferentes para cada sinal, separadas pelo intervalo de 1 Hz cada.

Remoção de ruído – de forma similar ao que foi realizado no domínio do tempo, os 0.3 s iniciais de cada sinal sonoro foi considerado como sendo ruído de ambiente, sendo submetido, juntamente com o restante do sinal, ao SNG, cujo resultado foi normalizado para valores entre -1 e 1.

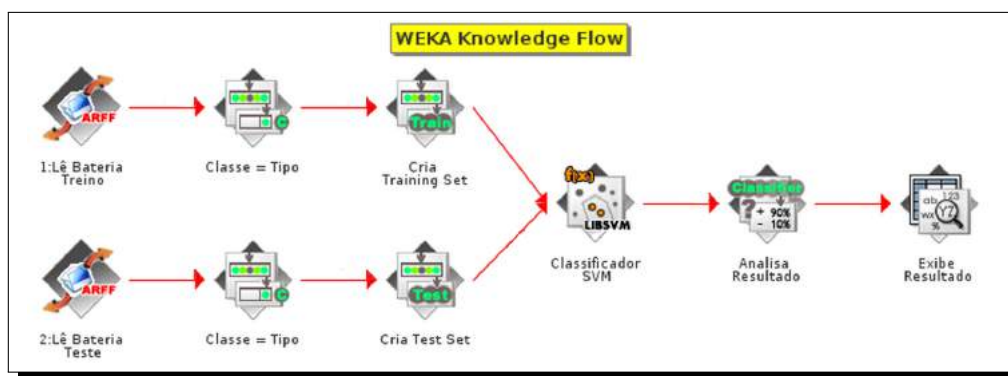
Cálculo da FFT e do Coeficiente de Variação dos sinais – a seguir foi realizado o cálculo da FFT para cada um dos sinais, separando-os em dois grupos: FFTs de Disparos e FFTs de sons Impulsivos. Para cada um destes grupos foi calculado,

Figura 46 – Esquema de processamento do algoritmo K-Médias no WEKA.



Fonte: o autor

Figura 47 – Esquema de processamento do algoritmo SVM no WEKA.



Fonte: o autor

frequência a frequência, o coeficiente de variação dos valores absolutos da FFT, base para determinação das $f(N)$ frequências mais relevantes.

Análise das Frequências

Este procedimento foi realizado sobre a Bateria de Treino e sobre a Bateria de Teste.

De posse das $f(N)$ frequência mais significativas para disparos, tomou-se a bateria em questão e nela foi executado o seguinte procedimento em cada um de seus sinais, a saber:

Zero Padding – de forma a obter-se 22050 frequências para cada sinal;

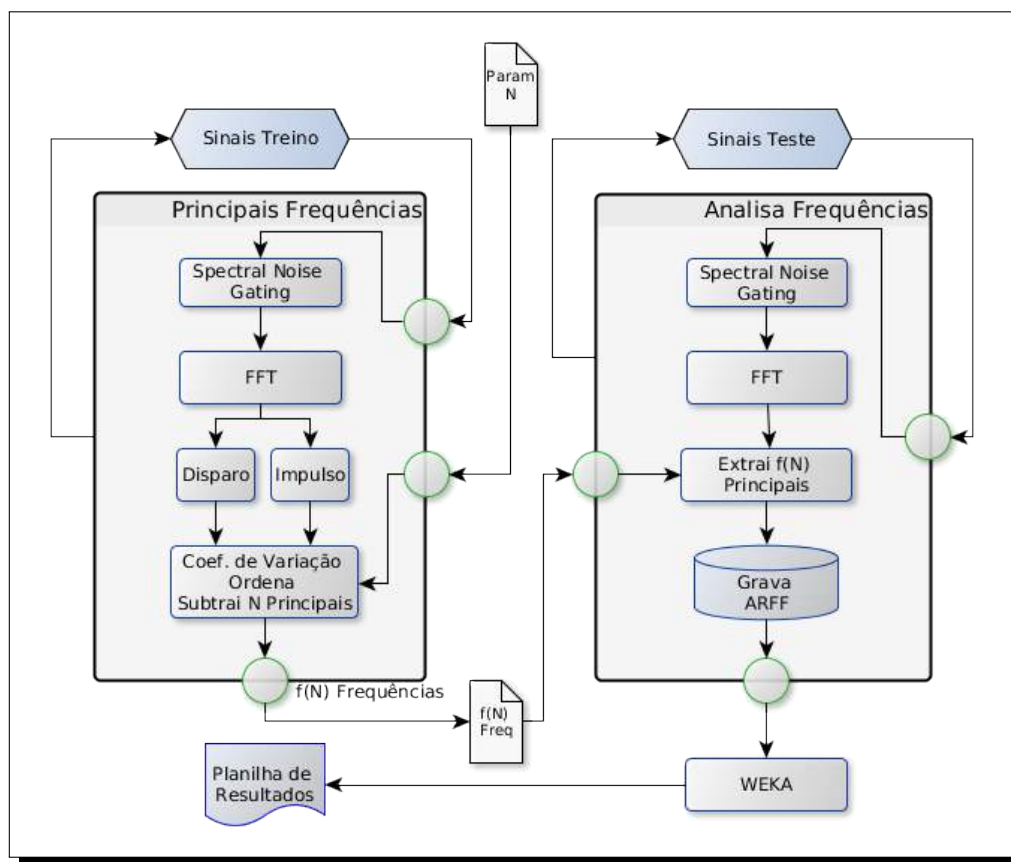
Remoção de ruído – utilizando-se o método SNG;

Normalização do sinal – para valores entre -1 e 1.

Calculo da FFT – das quais os $f(N)$ valores de frequência assinalados como significativos foram extraídos, normalizados e formatados como um arquivo *ARFF* a fim de serem introduzidos no WEKA, onde foram submetidos a uma clusterização K-Médias e classificação SVM.

O esquema completo do processo pode ser visto na figura 48.

Figura 48 – Diagrama do Teste no domínio das Frequências.



Fonte: o autor

Este procedimento, quando aplicado à Bateria de Treino, deu origem a um arquivo *ARFF* que serviu como *Training Set* para obtenção dos centroides e hiperplanos separatórios que foram utilizados como referência para a clusterização e classificação do arquivo *ARFF* da Bateria de Teste, cujos esquemas de geração podem ser vistos na figura 49.

O esquema utilizado pelo WEKA na clusterização e classificação dos dados é o mesmo visto nas figuras 46 e 47 .

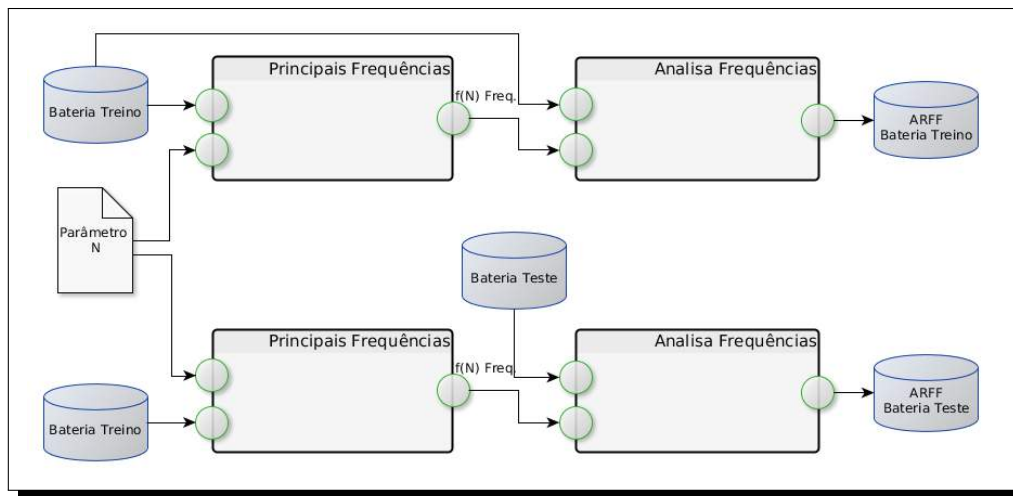
Quantidade de frequências a serem utilizadas

Um parâmetro fundamental nesta análise é a quantidade de frequências N a serem trabalhadas a fim de obter-se máximo desempenho do algoritmo.

Para isto, inicialmente verificou-se qual seria o número máximo de frequências relevantes $f(N)$ que se pode conseguir em uma amostra contendo 22050 frequências possíveis.

Apresentou-se ao algoritmo **Principais Frequências** a Bateria de Treino, solicitando que este trabalhasse com um parâmetro de N frequências variando no intervalo $1 \leq N \leq 22050$ Hz, a fim de estabelecer-se qual é o maior número de frequências $f(N)$ que podem

Figura 49 – Esquema de geração dos arquivos *ARFF* no domínio das frequências.

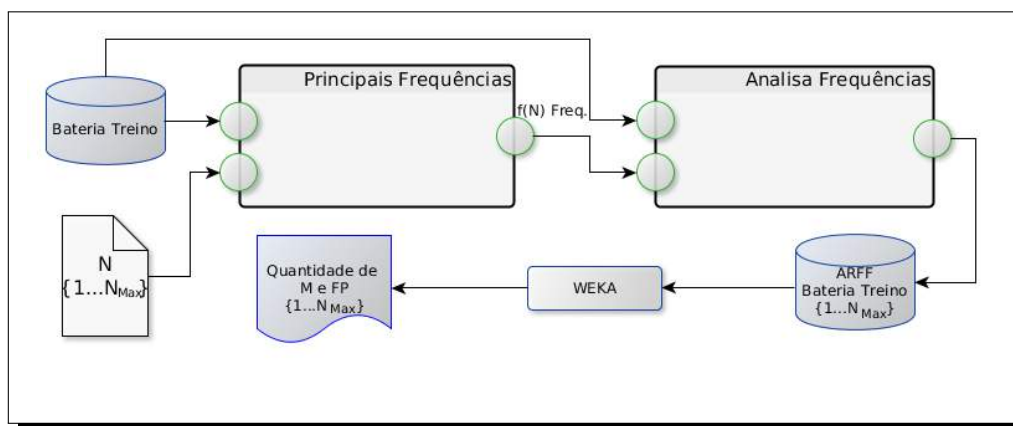


Fonte: o autor

por ele ser devolvidas. Obteve-se assim um range que varia entre 1 e N_{max} , onde N_{max} é o valor do parâmetro N que retorna o maior numero de frequências relevantes $f(N)$.

A seguir, aplicou-se o algoritmo de **Análise de Frequências** na Bateria de Treino, utilizando-se cada um dos conjuntos de frequências obtidos com o parâmetro N , onde $1 \leq N \leq N_{max}$. Os resultados foram inseridos no WEKA para clusterização via K-Médias e classificação via SVM, retornando o número de **M** e **FP** obtidos. O resultado foi analisado e estabeleceu-se como melhor valor para N aquele no qual obteve-se a menor ocorrência de **M** e **FP**, simultaneamente. O esquema do processo completo pode ser visto na figura 50

Figura 50 – Esquema do processo de análise de quantidade de frequências.



Fonte: o autor

5 Resultados e Discussão

5.1 Detecção de Impulsos

5.1.1 Resultados com Parâmetros Padrão

Aqui é apresentada a análise dos métodos que foi realizada utilizando-se os parâmetros preconizados na literatura: frame com $\Delta t = 90 \text{ ms}$ para o RMS; frame com $\Delta t = 90 \text{ ms}$ e $L = 11$ para a CMF; e um frame com $\Delta t = 90 \text{ ms}$ e $L = 30$ para a VM [22, 7]. A **Bateria 1** de sons foi apresentada aos três algoritmos, obtendo-se as matrizes de confusão mostrada nas figuras 51, 52 e 53.

Cada uma delas possui nove matrizes de confusão, resultado da combinação das três faixas de intervalos (**Faixas 1, 2 e 3**) entre disparos pelos três tipos de sinais gerados (**SR, R1 e R2**). Para uma melhor visualização do posicionamento de tais matrizes, destacou-se a matriz de confusão da **Faixa 1** \times **SR** do método RMS na figura 51.

Figura 51 – Matriz de confusão do RMS com parâmetros padrão.

		Faixa 1		Faixa 2		Faixa 3	
		A	NA	A	NA	A	NA
SR	CI	385	0	391	0	386	0
	SI	7786	188842	10340	392237	20508	1379754
R1	CI	384	0	390	0	386	0
	SI	3173	188842	5040	392237	15317	1379754
R2	CI	385	0	391	0	385	0
	SI	4206	188842	6522	392237	16729	1379754

SR – Sem ruído; R1 – Ruído Constante; R2 – Ruído Crescente; CI – *Frame* Com Impulso; SI – *Frame* Sem Impulso; A – Impulso Assinalado; NA – Impulso Não Assinalado.

Fonte: o autor

Aplicando-se as medidas de desempenho aos dados contidos nas matrizes de confusão obtêm-se os resultados descritos na tabela 3.

Ao analisar os números nela contidos, pode-se notar que os três processos possuem uma excelente sensibilidade (*se*), com índices muito próximos a 1, significando um pequeno índice de *Missed (M)*.

O VM também apresenta ótimos resultados em relação ao número de *Falsos Positivos (FP)*, com índices de precisão (*pr*) próximos à 0.8.

O CMF ficou com um escore intermediário para *pr*, na casa dos 0.5, indicando a detecção de um elevado número de **FP**.

Figura 52 – Matriz de confusão do CMF com parâmetros padrão.

		Faixa 1		Faixa 2		Faixa 3	
		A	NA	A	NA	A	NA
SR	CI	381	4	391	0	386	0
	SI	276	1735	386	3988	510	14938
R1	CI	372	0	383	8	385	1
	SI	309	1735	343	3988	382	14938
R2	CI	347	38	354	37	374	12
	SI	464	1735	309	3988	436	14938

Fonte: o autor

Figura 53 – Matriz de confusão do VM com parâmetros padrão.

		Faixa 1		Faixa 2		Faixa 3	
		A	NA	A	NA	A	NA
SR	CI	357	28	391	0	386	0
	SI	52	20613	54	43154	81	152610
R1	CI	372	0	391	0	386	0
	SI	54	20613	61	43154	86	152610
R2	CI	372	0	391	0	386	0
	SI	51	20613	75	43154	139	152610

Fonte: o autor

Já o RMS teve o pior desempenho para *pr*, demonstrando, conforme já era esperado, que este método possui uma tendência a um índice mais elevado de **FP**, em virtude do limitado número de amostras utilizadas na detecção, não sendo, portanto uma opção viável para a tarefa.

5.1.2 Otimização dos Métodos

A **Bateria 2** foi gerada e apresentada à CMF e VM, variando-se o tamanho de $L = \{10, 15, 20, 25, 30, 35, 40, 45, 50\}$ e *frame* com $\Delta t = \{1 \text{ ms} \rightarrow 500 \text{ ms}\}$ conforme descrito na metodologia (4.2.3), a fim de verificar quais seriam os melhores parâmetros para tais algoritmos. Valores ideais são aqueles em que se obtém o menor número de **M** e de **FP**.

Nas figuras 54, 55, 56, 57, 58 e 59 podem ser vistos os gráficos dos resultados dos testes de parametrização para os métodos da CMF e VM, respectivamente. O plano mais claro representa a quantidade de **M** assinalados e o plano mais escuro a quantidade de **FP**.

Tabela 3 – Comparativo das medidas de desempenho dos três métodos

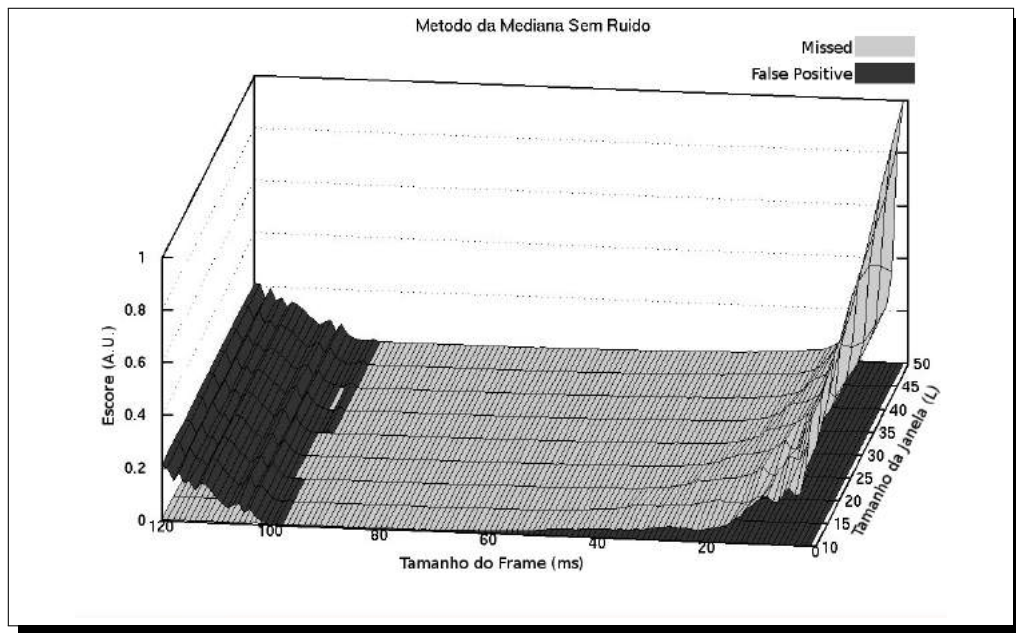
Faixa 1 – De 0.1 a 0.5 s												
Ruído	RMS				CMF				VM			
	ac	pr	se	sp	ac	pr	se	sp	ac	pr	se	sp
SR	1,00	0,05	1,00	0,96	1,00	0,58	0,99	0,86	1,00	0,87	0,93	0,99
R1	1,00	0,11	0,99	0,98	0,99	0,55	0,97	0,85	1,00	0,87	0,97	0,99
R2	1,00	0,08	1,00	0,98	0,98	0,43	0,90	0,79	1,00	0,88	0,97	0,99

Faixa 2 – De 0.5 a 1 s												
Ruído	RMS				CMF				VM			
	ac	pr	se	sp	ac	pr	se	sp	ac	pr	se	sp
SR	1,00	0,04	1,00	0,97	1,00	0,50	1,00	0,91	1,00	0,88	1,00	0,99
R1	1,00	0,07	0,99	0,99	1,00	0,53	0,98	0,92	1,00	0,87	1,00	0,99
R2	1,00	0,06	1,00	0,98	0,99	0,53	0,91	0,93	1,00	0,84	1,00	0,99

Faixa 3 – De 1 a 5 s												
Ruído	RMS				CMF				VM			
	ac	pr	se	sp	ac	pr	se	sp	ac	pr	se	sp
SR	1,00	0,02	1,00	0,99	1,00	0,43	1,00	0,97	1,00	0,83	1,00	0,99
R1	1,00	0,02	1,00	0,99	1,00	0,50	0,99	0,98	1,00	0,82	1,00	0,99
R2	1,00	0,02	1,00	0,99	1,00	0,46	0,97	0,97	1,00	0,74	1,00	0,99

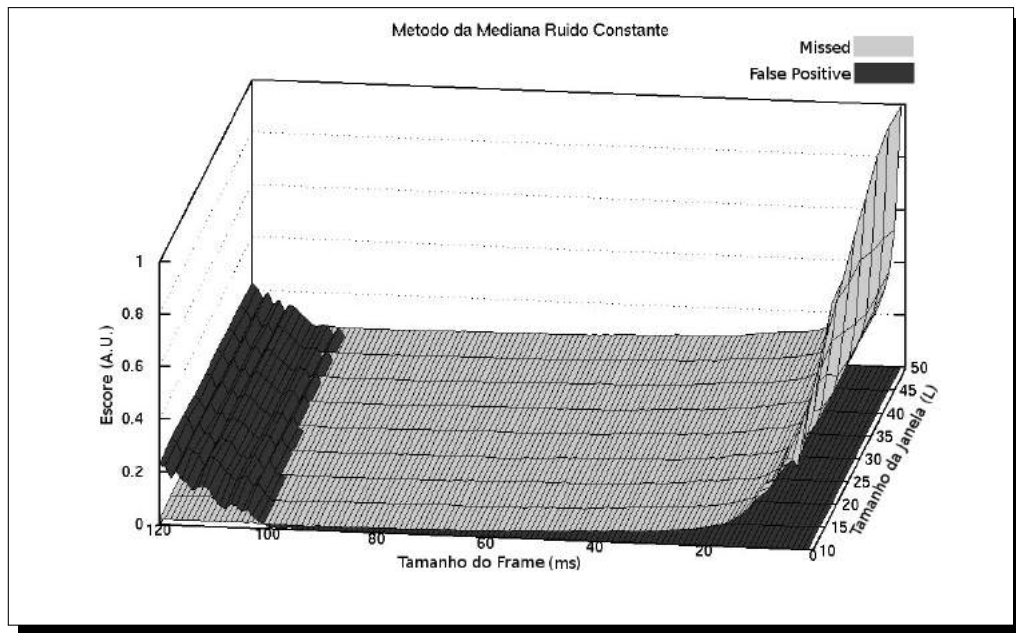
Fonte: o autor

Figura 54 – Gráfico comparativo entre **M** e **FP** em função de **N** e **L** da **CMF** – Sem ruído (SR).



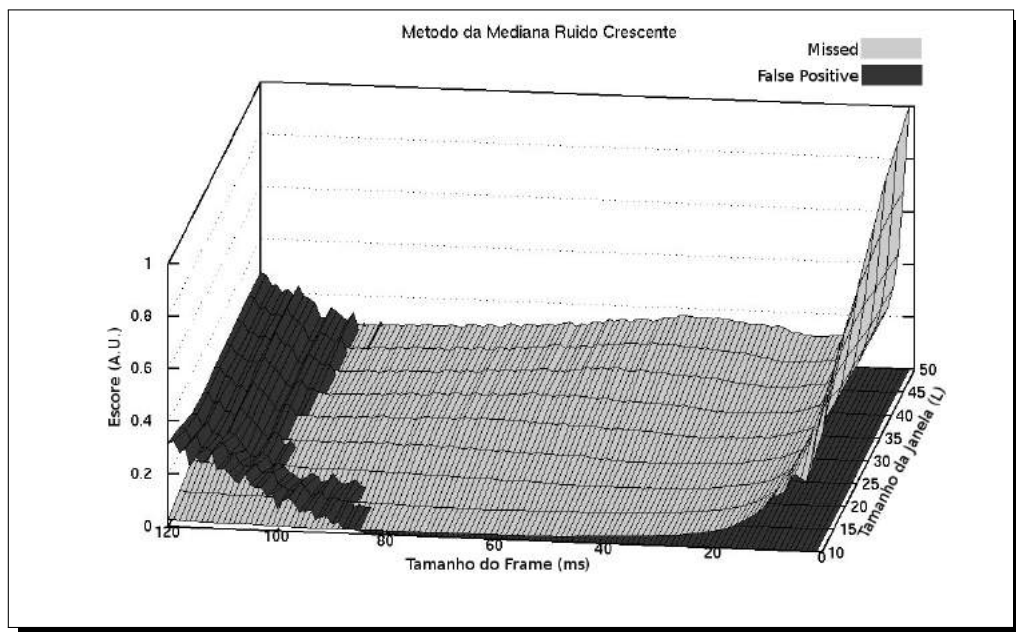
Fonte: o autor

Figura 55 – Gráfico comparativo entre **M** e **FP** em função de **N** e **L** da **CMF** – Ruído Constante (R1).



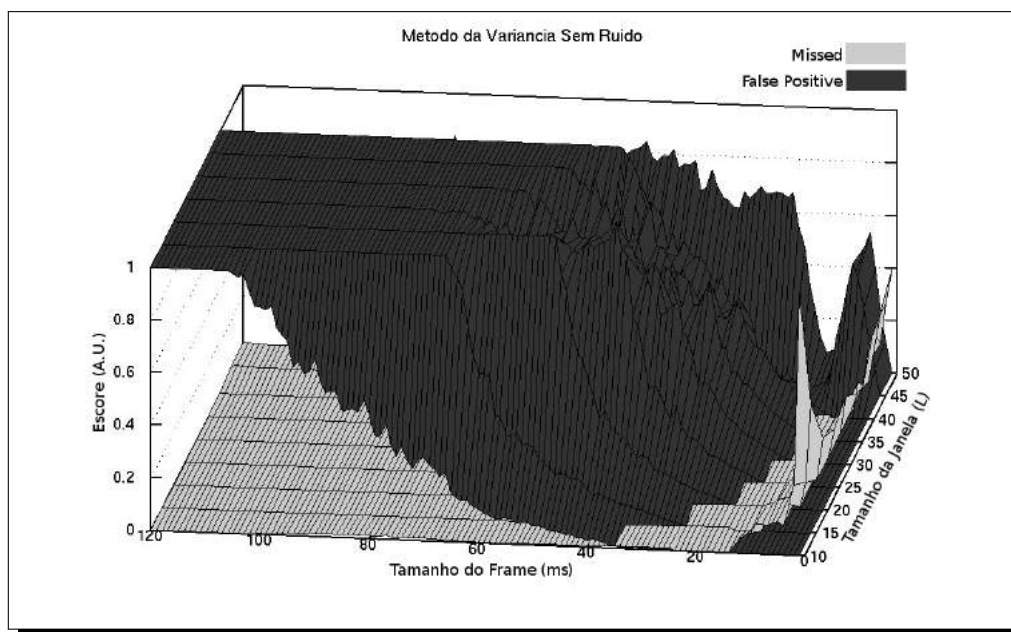
Fonte: o autor

Figura 56 – Gráfico comparativo entre **M** e **FP** em função de **N** e **L** da **CMF** – Ruído Crescente (R2) .



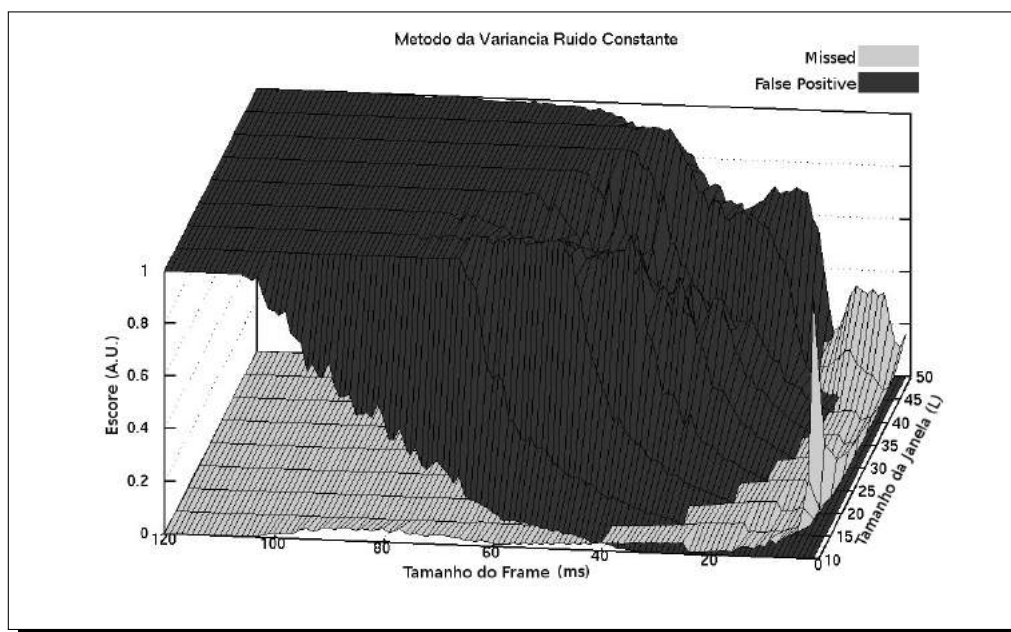
Fonte: o autor

Figura 57 – Gráfico comparativo entre **M** e **FP** em função de **N** e **L** da **VM** – Sem ruído (SR).



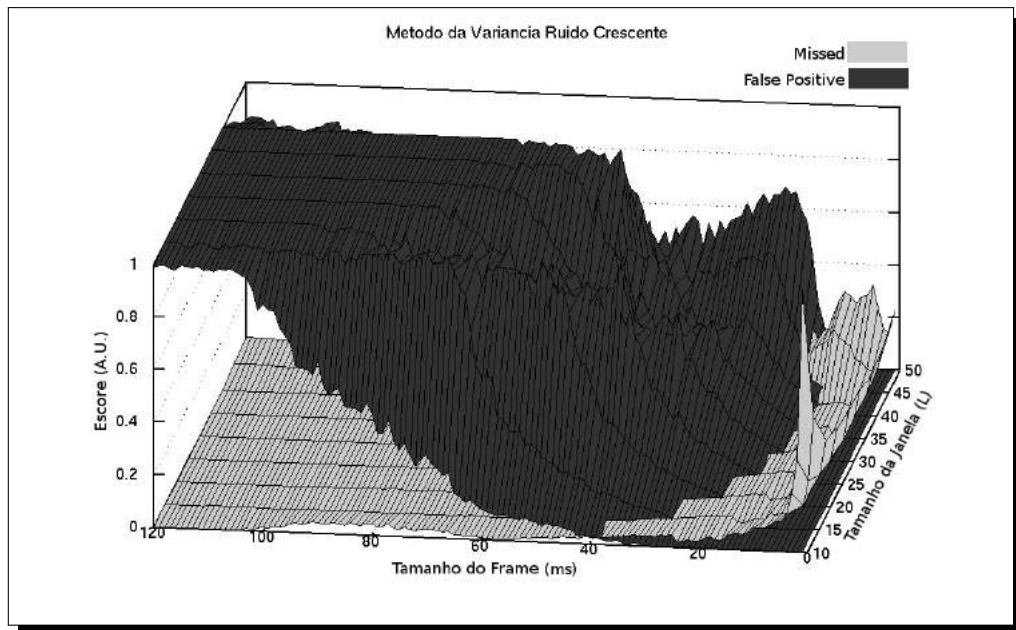
Fonte: o autor

Figura 58 – Gráfico comparativo entre **M** e **FP** em função de **N** e **L** da **VM** – Ruído Constante (R1).



Fonte: o autor

Figura 59 – Gráfico comparativo entre **M** e **FP** em função de **N** e **L** da **VM** – Ruído Crescente (R2).



Fonte: o autor

Os valores do eixo z (Escore) aparecem normalizados para uma melhor comparação entre parâmetros. Os valores ótimos para o tamanho do frame **N** e tamanho da janela **L** podem ser obtidos na intersecção dos planos onde os valores de **M** e **FP** são os menores possíveis.

Observa-se um melhor desempenho da CMF com janelas pequenas, com valores próximos a $L = 10$ e $\Delta t \approx 80$ ms. Já o VM possui uma sensibilidade muito grande às mudanças de L e Δt , não sendo possível uma identificação precisa utilizando-se apenas os gráficos.

Analisando-se os dados numéricos obtidos, com auxílio do PostgreSQL, localizou-se os menores valores de *precisão* e *sensibilidade* para cada algoritmo. Porém não foi possível obter-se um valor ótimo de parâmetros para a CMF já que uma melhora na sensibilidade do método causa um aumento no número de falsos positivos. Optou-se, assim, pelo parâmetro que apresentou a melhor taxa de sensibilidade.

A tabela 4 apresenta os três melhores resultados para cada método, onde são admitidos apenas valores onde a sensibilidade é de 100%, *i.e.*, onde o número de **M** é igual a zero, e a precisão é a maior possível.

Os melhores resultados de CMF foram obtidos com um *frame* de $\Delta t = 73$ ms e janela $L = 10$, onde conseguiu-se $pr = 0.6125$ e $se = 1.0$.

Já no caso da VM, os melhores resultados foram obtidos com um *frame* de $\Delta t = 11$ ms e janela $L = 20$, onde conseguiu-se $pr = 0.9948$ e $se = 1.0$.

Tabela 4 – Melhores resultados dos Métodos da CMF e VM

Método da CMF							
Parâmetros		SR		R1		R2	
L	$\Delta t(ms)$	pr	se	pr	se	pr	se
10	73	0.6125	1.0000	0.5299	1.0000	0.5546	1.0000
10	77	0.6116	1.0000	0.5374	1.0000	0.5359	1.0000
10	76	0.6106	1.0000	0.5420	1.0000	0.5277	1.0000

Método da VM							
Parâmetros		SR		R1		R2	
L	$\Delta t(ms)$	pr	se	pr	se	pr	se
20	11	0.9948	1.0000	0.9896	1.0000	0.9338	1.0000
15	15	0.9948	1.0000	0.9794	1.0000	0.9203	1.0000
15	14	0.9974	1.0000	0.9597	1.0000	0.9293	1.0000

Fonte: o autor

5.1.3 Resultado com os Parâmetros Otimizados

Nesta fase, a **Bateria 1** foi rerepresentada aos métodos da CMF e VM parametrizados com os valores descritos acima, obtendo-se os resultados que podem ser vistos nas matrizes de confusão contidas nas figuras 60 e 61.

Figura 60 – Matriz de confusão do CMF com parâmetros otimizados.

Matriz de Confusão do Método da CMF							
		Faixa 1		Faixa 2		Faixa 3	
		A	NA	A	NA	A	NA
SR	CI	372	0	378	0	393	0
	SI	286	3550	3015	3568	4099	16256
R1	CI	372	0	378	0	393	0
	SI	359	3477	526	6057	758	19597
R2	CI	371	1	378	0	393	0
	SI	384	3452	357	6226	936	19419

Fonte: o autor

Com base nestes dados, foram calculadas as novas medidas de desempenho, apresentadas na tabela 5.

Comparando-se os resultados obtidos no teste preliminar com o atual, conforme pode ser visto na tabela 6, nota-se que foi possível obter uma melhoria na taxa da detecção de ambos os métodos, conseguindo-se um índice de sensibilidade com 100% de acertos para

Figura 61 – Matriz de confusão do VM com parâmetros otimizados.

		Matriz de Confusão do Método da Variância					
		Faixa 1		Faixa 2		Faixa 3	
		A	NA	A	NA	A	NA
SR	CI	372	0	378	0	393	0
	SI	0	27411	11	45651	52	137103
R1	CI	372	0	378	0	393	0
	SI	3	27408	14	45648	51	137104
R2	CI	372	0	378	0	393	0
	SI	19	27392	52	45610	86	137069

Fonte: o autor

Tabela 5 – Medidas de desempenho dos Métodos CMF e VM com os novos parâmetros

Faixa 1 – De 0.1 a 0.5 s								
Ruído	CMF				VM			
	ac	pr	se	sp	ac	pr	se	sp
SR	0,9320	0,5653	1,0000	0,9254	1,0000	1,0000	1,0000	1,0000
R1	0,9147	0,5089	1,0000	0,9064	0,9999	0,9920	1,0000	0,9999
R2	0,9085	0,4914	0,9973	0,8999	0,9993	0,9514	1,0000	0,9993

Faixa 2 – De 0.5 a 1 s								
Ruído	CMF				VM			
	ac	pr	se	sp	ac	pr	se	sp
SR	0,5669	0,1114	1,0000	0,5420	1,000	0,879	1,000	0,999
R1	0,9244	0,4181	1,0000	0,9201	1,000	0,865	1,000	0,999
R2	0,9487	0,5143	1,0000	0,9458	1,000	0,839	1,000	0,998

Faixa 3 – De 1 a 5 s								
Ruído	CMF				VM			
	ac	pr	se	sp	ac	pr	se	sp
SR	0,8024	0,0875	1,0000	0,7986	0,9996	0,8831	1,0000	0,9996
R1	0,9635	0,3414	1,0000	0,9628	0,9996	0,8851	1,0000	0,9996
R2	0,9549	0,2957	1,0000	0,9540	0,9994	0,8205	1,0000	0,9994

Fonte: o autor

todos os casos, exceto para a CMF na **Faixa 1** com **R2**, sendo que mesmo este índice ficou extremamente próximo à unidade.

Também observa-se uma melhora na precisão da VM, onde os índices para a **Faixa 2** permaneceram inalterados, mas havendo melhora entre 5 a 12% nos índices para as faixas 1 e 3, tendo-se alcançado índice de precisão de 100% para a **Faixa 1 SR**.

Já no caso da CMF houve uma queda na precisão em todas as situações, havendo

casos em que o número de **FP** aumentou em 10 vezes. Nota-se, assim, que uma pequena melhora no índice de sensibilidade causa uma enorme perda na precisão, o que induz à conclusão que tal ganho não é vantajoso, sendo preferível manter-se os parâmetros da CMF preconizados na literatura.

Tabela 6 – Comparativo das medidas de desempenho dos Métodos CMF e VM com os parâmetros originais e otimizados

Faixa 1 – De 0.1 a 0.5 s									
Ruído	CMF				VM				
	Original		Otimizado		Original		Otimizado		
	pr	se	pr	se	pr	se	pr	se	
SR	0,58	0,99	0,57	1,00	0,87	0,93	1,00	1,00	
R1	0,55	0,97	0,51	1,00	0,87	0,97	0,99	1,00	
R2	0,43	0,90	0,49	0,99	0,88	0,97	0,95	1,00	

Faixa 2 – De 0.5 a 1 s									
Ruído	CMF				VM				
	Original		Otimizado		Original		Otimizado		
	pr	se	pr	se	pr	se	pr	se	
SR	0,50	1,00	0,11	1,00	0,88	1,00	0,88	1,00	
R1	0,53	0,98	0,42	1,00	0,87	1,00	0,87	1,00	
R2	0,53	0,91	0,51	1,00	0,84	1,00	0,84	1,00	

Faixa 3 – De 1 a 5 s									
Ruído	CMF				VM				
	Original		Otimizado		Original		Otimizado		
	pr	se	pr	se	pr	se	pr	se	
SR	0,43	1,00	0,09	1,00	0,83	1,00	0,88	1,00	
R1	0,50	0,99	0,34	1,00	0,82	1,00	0,89	1,00	
R2	0,46	0,97	0,30	1,00	0,74	1,00	0,82	1,00	

Fonte: o autor

Isto posto, conclui-se que o método mais adequado ao propósito de detecção de um som impulsivo é o método da VM com um *frame* de $\Delta t = 11 \text{ ms}$ e janela $L = 20$, uma vez que este alcançou excelentes índices de *pr* e *se*.

5.1.4 Teste de Generalização

A fim de comprovar a eficiência do método para disparos provenientes de outras armas, bem como para sons impulsivos gerados por outras fontes, foram realizados os testes envolvendo cinco disparos de armas de fogo e outros cinco sons impulsivos diversos, cujos resultados podem ser vistos a seguir.

Para Diversos Armamentos

A figura 62 mostra a matriz de confusão obtida na detecção de disparos de armas de fogo aleatórias.

Figura 62 – Matriz de confusão do VM para detecção de diversos armamentos.

		Faixa 1		Faixa 2		Faixa 3	
		A	NA	A	NA	A	NA
SR	CI	357	16	376	0	354	0
	SI	89	18012	121	35083	148	116011
R1	CI	362	11	376	0	354	0
	SI	139	17962	147	35057	171	115988
R2	CI	358	15	373	3	352	2
	SI	130	17971	152	35052	203	115956

Fonte: o autor

Que resulta nas medidas de desempenho exibidas na tabela 7.

Tabela 7 – VM com vários armamentos

Ruído	ac	pr	se	sp
SR	0,9943	0,8004	0,9571	0,9951
R1	0,9919	0,7226	0,9705	0,9923
R2	0,9922	0,7336	0,9598	0,9928

Ruído	ac	pr	se	sp
SR	0,9966	0,7565	1,0000	0,9966
R1	0,9959	0,7189	1,0000	0,9958
R2	0,9956	0,7105	0,9920	0,9957

Ruído	ac	pr	se	sp
SR	0,9987	0,7052	1,0000	0,9987
R1	0,9985	0,6743	1,0000	0,9985
R2	0,9982	0,6342	0,9944	0,9983

Fonte: o autor

Nela observa-se uma excelente *se*, tendo-se, porém, uma pequena queda na *pr*, particularmente nos resultados da **Faixa 3**.

Isso pode ser explicado pela diferença de potência e duração dos sinais empregados, quando comparados com o som utilizado no teste de cadência, em virtude da distância entre a fonte do disparo e o ponto de gravação do sinal.

Para Sons Impulsivos Diversos

O mesmo experimento foi repetido, só que desta vez utilizando-se todos os 10 sons selecionados, quando foram obtidos os resultados constantes da matriz de confusão apresentada na figura 63.

Figura 63 – Matriz de confusão do VM para detecção de diversos sons impulsivos.

		Faixa 1		Faixa 2		Faixa 3	
		A	NA	A	NA	A	NA
SR	CI	372	11	381	0	395	0
	SI	84	19887	107	37553	145	127825
R1	CI	373	10	381	0	393	2
	SI	87	19884	97	37563	129	127841
R2	CI	372	11	380	1	392	3
	SI	83	19888	132	37528	157	127813

Fonte: o autor

Como decorrência obteve-se as medidas de desempenho apresentadas na tabela 8.

Mais uma vez o algoritmo da VM apresentou excelentes resultados para *se*, mantendo os índices de *pr* variando entre 0.71 e 0.82.

Desta forma conclui-se que o algoritmo é eficiente para todas as faixas e níveis de ruído, independentemente do som impulsivo apresentado.

5.2 Reconhecimento de Disparos

5.2.1 Análise no domínio do tempo

Nesta etapa a Bateria de Treino foi apresentada ao algoritmo de comparação com a explosão de Friedlander. Composta por 58 disparos e 110 sons impulsivos diversos, dela onde foram obtidos 168 valores de Φ e Θ , cuja distribuição, após sua normalização entre 0 e 1, pode ser vista na figura 64, onde a figura 64b exhibe o detalhamento da área de interesse destacada na figura 64a.

Nela pode-se notar que, embora os sons impulsivos estejam distribuídos por uma área um pouco maior do gráfico, ambos, disparos e impulsos, encontram-se concentrados na

Tabela 8 – VM com vários Sons Impulsivos

Faixa 1 – De 0.1 a 0.5 s				
Ruído	ac	pr	se	sp
SR	0,9953	0,8158	0,9713	0,9958
R1	0,9952	0,8109	0,9739	0,9956
R2	0,9954	0,8176	0,9713	0,9958

Faixa 2 – De 0.5 a 1 s				
Ruído	ac	pr	se	sp
SR	0,9972	0,7807	1,0000	0,9972
R1	0,9975	0,7971	1,0000	0,9974
R2	0,9965	0,7422	0,9974	0,9965

Faixa 3 – De 1 a 5 s				
Ruído	ac	pr	se	sp
SR	0,9989	0,7315	1,0000	0,9989
R1	0,9990	0,7529	0,9949	0,9990
R2	0,9988	0,7140	0,9924	0,9988

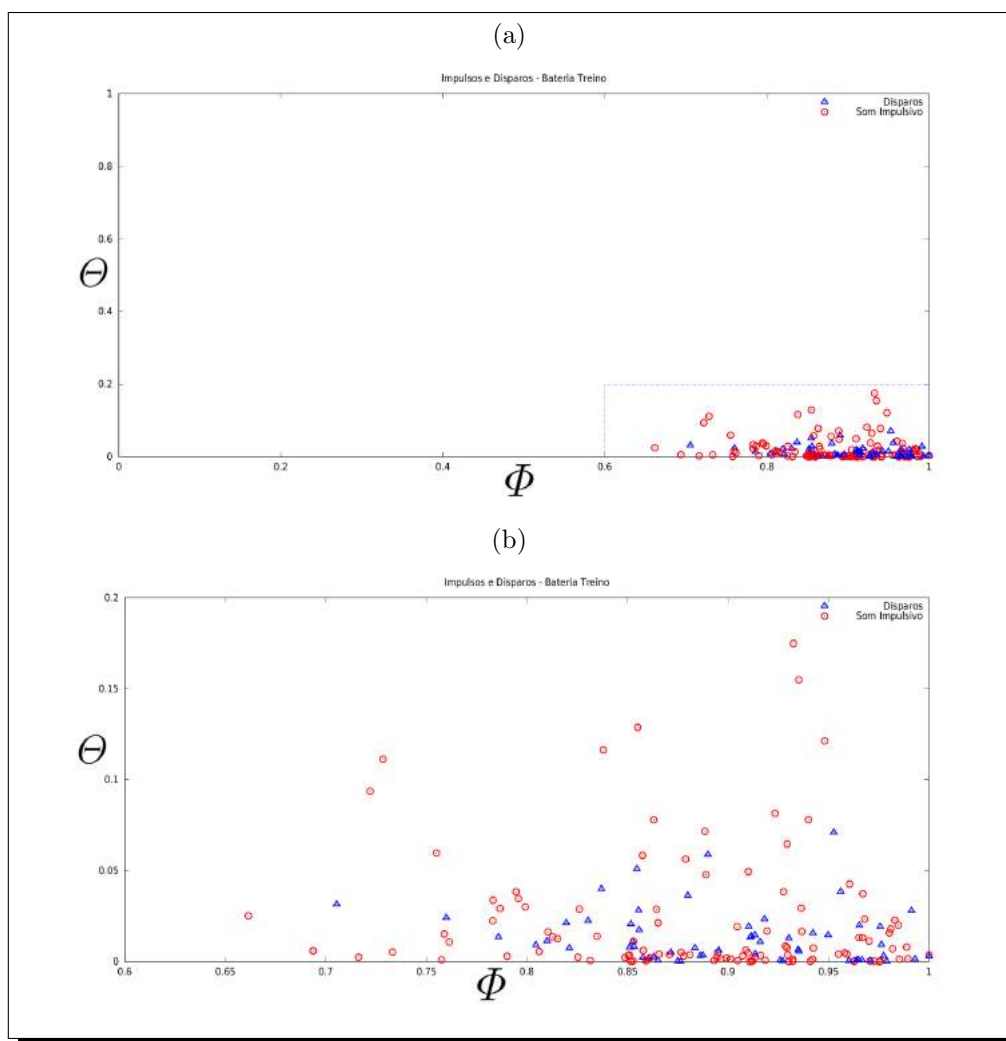
Fonte: o autor

região $\{0.65 < \Phi \leq 1\}$ e $\{0 \leq \Theta < 0.2\}$. Nota-se, ainda, que a maior concentração de instâncias de ambos os tipos localizam-se em uma área ainda menor, onde $\{0.85 < \Phi \leq 1\}$ e $\{0 \leq \Theta \leq 0.05\}$. Isto indica uma predominância de sinais de duração menor ou igual ao pulso de Friedlander, mas que não possuem a mesma morfologia que aquele, conceito este válido tanto para sons impulsivos quanto para disparos.

Como consequência, pode-se estabelecer que, *a priori*, o formato da envoltória de ambos os tipos de sinais são bastante semelhantes e que não há uma fronteira bem visível entre a área ocupada pelos sons impulsivos e a dos disparos. Isso pode comprometer o desempenho dos algoritmos de classificação e clusterização.

Foi, ainda, apresentado ao algoritmo de comparação com a explosão de Friedlander a Bateria de Teste composta por 29 disparos e 55 sons impulsivos diversos, de onde foram obtidos 84 valores de Φ e Θ . Destes 1 disparo e 1 som impulsivo foram descartados por apresentarem valores cerca de 36 vezes maior que a média dos demais.

A distribuição dos 82 valores de Φ e Θ restantes pode ser vista na figura 65. onde se observa uma distribuição semelhante à da Bateria de Treino.

Figura 64 – Gráfico da distribuição de $\Phi \times \Theta$ – Bateria de Treino.

(a) Distribuição de $\Phi \times \Theta$ para Impulsos e Disparos. A linha pontilhada indica a área de interesse do gráfico; (b) Detalhe da área de interesse.

Fonte: o autor

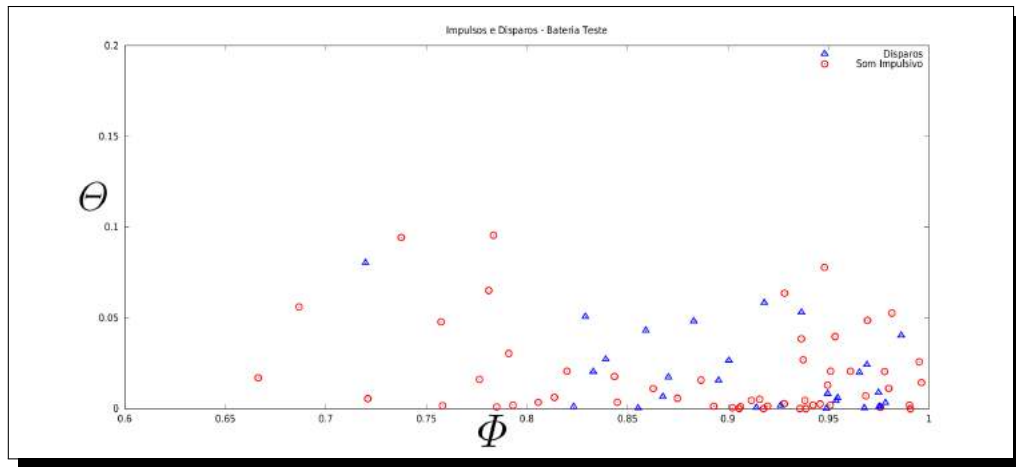
Resultados com o K-Médias

Aplicou-se, então, uma clusterização pelo método K-Médias nos resultados da Bateria de Treino, obtendo-se dois clusters e seus respectivos centroides $c_0(0.8248, 0.0234)$ e $c_1(0.9408, 0.0163)$, vistos na figura 66.

A seguir realizou-se a clusterização da Bateria de Teste, aplicando-se os mesmos centroides $c_0(0.8248, 0.0234)$ e $c_1(0.9408, 0.0163)$, obtendo-se o resultado visto na figura 67.

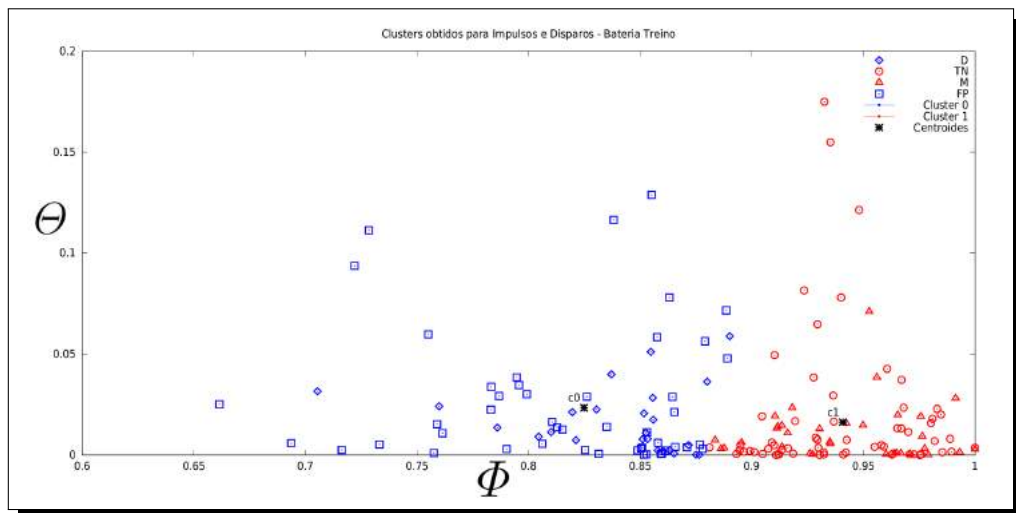
A tabela 9 exhibe a matriz de confusão obtida, onde nota-se que houve uma taxa elevada de disparos não identificados, ou 18 Missed (M), e uma grande taxa de Falsos Positivos (FP), com 19 sons impulsivos sendo identificados como Disparos. Isso já era esperado pela simples análise visual da figura 64, já que uma grande quantidade de sons impulsivos se superpõe à mesma área ocupada pelos disparos.

Figura 65 – Gráfico da distribuição de $\Phi \times \Theta$ – Bateria de Teste.



Fonte: o autor

Figura 66 – Resultado da clusterização do algoritmo K-Médias sobre a Bateria de Treino utilizando distância Euclidiana.



Fonte: o autor

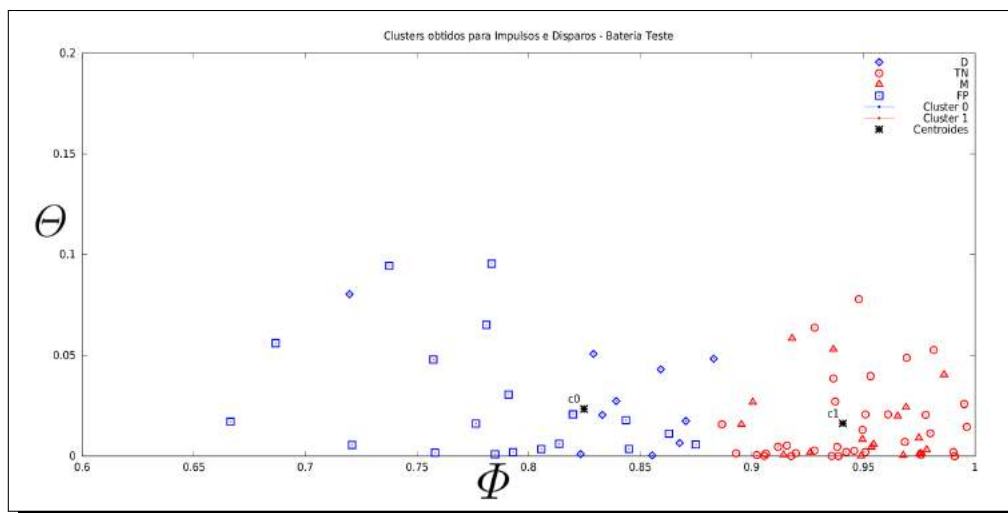
Tabela 9 – Matriz de Confusão - K-Médias

Amostras	Agrupado como	
	Disparos	Sons Impulsivos
Com Disparo	10 (<i>D</i>)	18 (<i>M</i>)
Sem Disparo	19 (<i>FP</i>)	35 (<i>TN</i>)

Fonte: o autor

A tabela 10 mostra as medidas de desempenho do método, onde nota-se que todos os

Figura 67 – Resultado da clusterização do algoritmo K-Médias sobre a Bateria de Teste.



Fonte: o autor

índices alcançaram escores muito baixos.

Tabela 10 – Medidas de Desempenho - K-Médias

ac	pr	se	sp
0.54878	0.34483	0.35714	0.64815

Fonte: o autor

Desta forma verifica-se que, o método é pouco preciso e pouco sensível, clusterizando corretamente apenas 54.9% dos sinais apresentados.

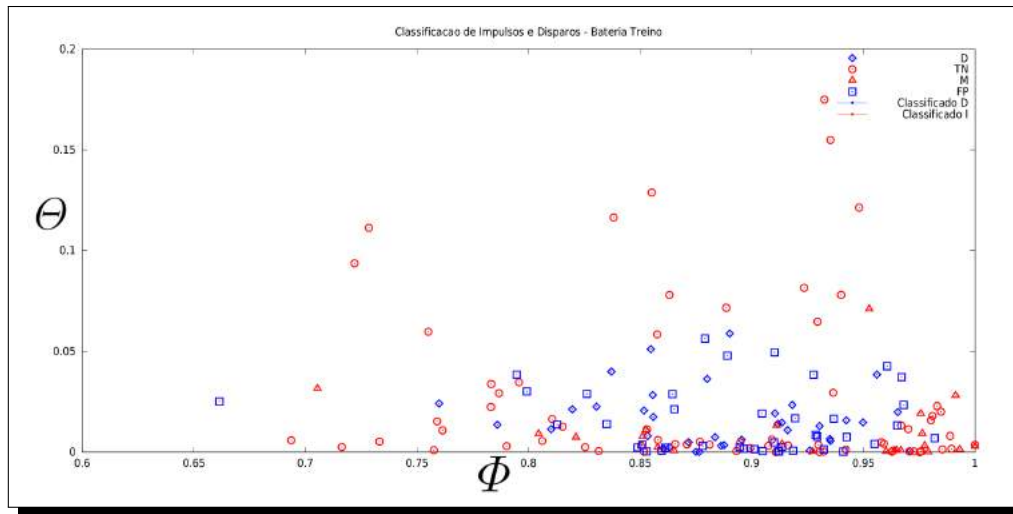
Resultados com o SVM

De forma similar, utilizou-se o SVM na Bateria de Treino obtendo-se a classificação vista na figura 68.

A seguir realizou-se a classificação da Bateria de Teste, aplicando-se o hiperplano encontrado, obtendo-se o resultado visto na figura 69. A tabela 11 exhibe a matriz de confusão obtida, e a tabela 12, suas medidas de desempenho.

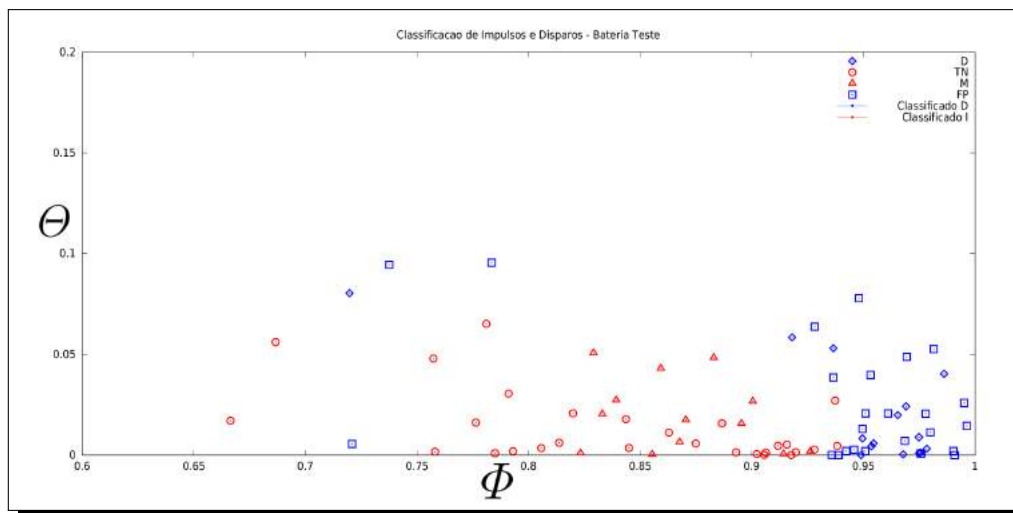
Nota-se que os índices, a pesar de melhores que os obtidos pelo K-Médias, ainda assim não são satisfatórios. Este baixo rendimento pode ser explicado pela ausência de uma fronteira bem definida entre os dois tipos de sinais, já que a distribuição de ambos encontra-se superposta, dificultando o trabalho de determinação do hiperplano separatório adequado.

Figura 68 – Resultado da classificação do algoritmo SVM sobre a Bateria de Treino.



Fonte: o autor

Figura 69 – Resultado da clusterização do algoritmo SVM sobre a Bateria de Teste.



Fonte: o autor

Tabela 11 – Matriz de Confusão - SVM

Amostras	Classificado como	
	Disparos	Sons Impulsivos
Com Disparo	15 (<i>D</i>)	13 (<i>M</i>)
Sem Disparo	26 (<i>FP</i>)	28 (<i>TN</i>)

Fonte: o autor

Tabela 12 – Medidas de Desempenho - SVM

ac	pr	se	sp
0.52439	0.36585	0.53571	0.51852

Fonte: o autor

5.2.2 Análise no domínio das frequências

Quantidade de Frequências a serem utilizadas

Inicialmente determinou-se qual seria o número máximo de frequências relevantes $f(N)$ que poderia ser obtido pelo método aqui proposto. Para tanto utilizou-se a rotina **Principais Frequências**, que recebeu como parâmetros a Bateria de Treino e a quantidade de frequências a serem utilizadas N , onde $1 \leq N \leq 22050 \text{ Hz}$. O resultado pode ser visto na figura 70a, onde o eixo das abscissas representa a quantidade de frequências N a serem usadas na obtenção do conjunto de frequências relevantes e o eixo das ordenadas a quantidade de frequências consideradas relevantes $f(N)$ que foram retornadas pelo algoritmo.

Pode-se notar um crescimento $f(N)$ até $N = 6916$, que retornou $f(N) = 3452$ frequências. A partir deste ponto $f(N)$ passa a decair até chegar em zero.

Isto ocorre porque o conjunto de frequências relevantes é igual ao conjunto das N frequências com menor Coeficiente de Variação dos Disparos (FD) menos o conjunto das N frequências com menor Coeficiente de Variação dos Impulsos (FI). Desta forma, quanto maior for N , maior também será o número de frequências coincidentes entre FD e FI , retornando um número cada vez menor de frequências no conjunto ($FD - FI$). Com efeito, se $FD = \{1, 2, 3, \dots, 22050\}$ e $FI = \{1, 2, 3, \dots, 22050\}$ tem-se $(FD - FI) = \emptyset$.

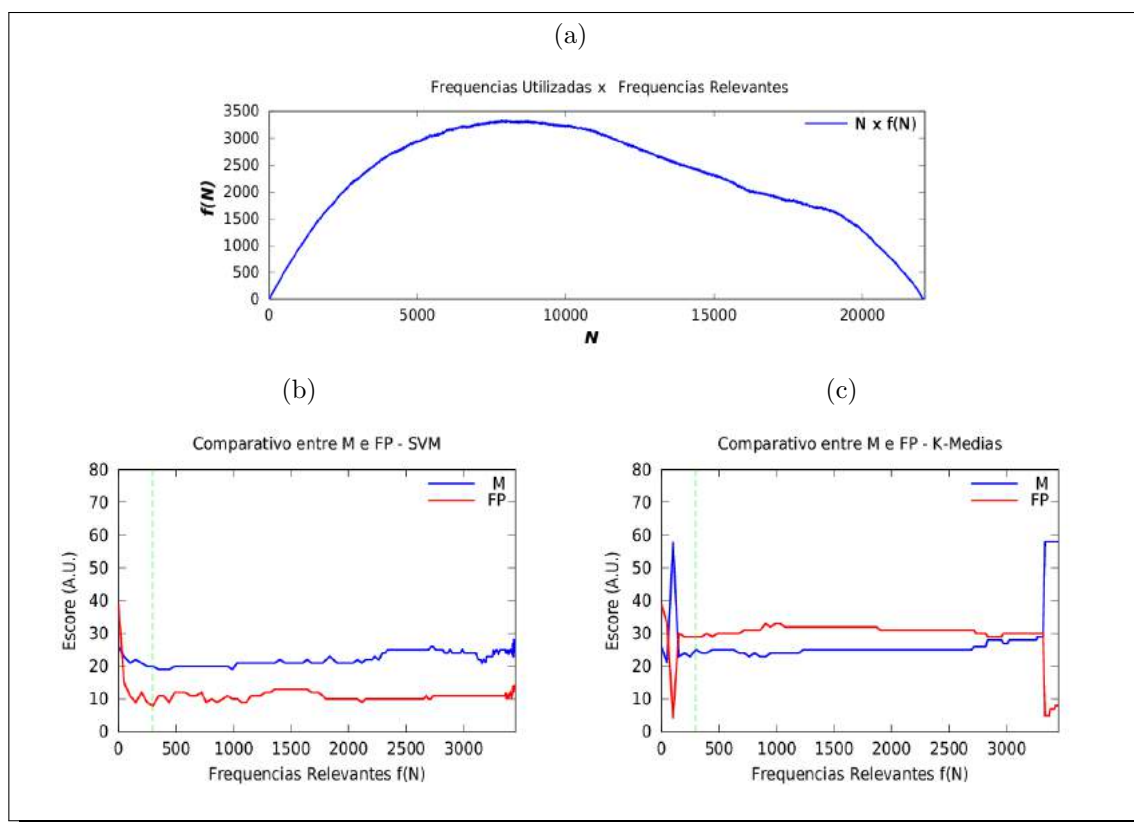
Assim, passou-se a considerar o N_{max} , a ser utilizado na próxima fase, como o ponto de inflexão da curva da figura 70a, onde para $N = 6916$ tem-se $f(N) = 3452$.

A seguir passou-se a alimentar a rotina **Analisa Frequências** com as N frequências obtidas na etapa anterior, onde $1 \leq N \leq 6916$, mais os sons constantes da Bateria de Treino, aplicando os resultados a uma classificação SVM e uma clusterização K-Médias no WEKA, utilizando-se como treinamento uma *cross-validation* com 10 *folds*. Obteve-se, assim, a quantidade de **M** e **FP** registrados em cada método, sendo os resultados vistos na figura 70b e 70c, respectivamente.

Observa-se que a SVM apresenta um melhor desempenho com um menor número de **M** e de **FP**, quando comparada ao K-Médias, tendo também a menor variação em função do número de frequências utilizadas.

Os melhores resultados foram obtido na faixa de 290 a 300 frequências relevantes para ambos os métodos, sendo que o melhor resultado foi obtido pela SVM utilizando 298

Figura 70 – Análise da quantidade ideal de frequências a serem utilizadas.



(a) Quantidade de frequências utilizadas N versus quantidade de frequências relevantes $f(N)$; (b) Comparativo entre M e FP obtidos pelo SVM, em função de $f(N)$; (c) Comparativo entre M e FP obtidos pelo K-Médias, em função de $f(N)$; A linha pontilhada nas figuras (b) e (c) indica a marca das 298 frequências.

Fonte: o autor

frequências, indicado pela linha pontilhada nas figuras 70b e 70c, onde obteve-se 18 M e 9 FP .

A seguir, alimentou-se a rotina **Analisa Frequências** com a Bateria de Teste e com o conjunto das 298 frequências mais relevantes descritas. Ao resultado foi aplicada a classificação SVM e clusterização K-Médias. Os resultados podem ser vistos nas tabelas 13 e 14, respectivamente, e as medidas de desempenho decorrentes podem ser vistas na tabela 15.

Comparando os índices de forma absoluta percebe-se uma melhor performance no método SVM na acurácia, precisão e especificidade, porém o índice de sensibilidade alcançado foi muito pequeno, com um número muito grande de M . Por outro lado o K-Médias apresentou uma sensibilidade bem mais elevada que o SVM, contando com um número de M três vezes menor que os disparos não identificados pelo SVM. Assim, não há como concluir de forma inequívoca sobre um benefício evidente no uso de um ou de outro método, havendo, porém, uma ligeira vantagem para o K-Médias, por conta dos números mais consistentes.

Tabela 13 – Matriz de Confusão – Bateria de Teste – SVM

Amostras	Classificado como	
	Disparos	Sons Impulsivos
Com Disparo	5 (<i>D</i>)	23 (<i>M</i>)
Sem Disparo	0 (<i>FP</i>)	54 (<i>TN</i>)
Total	5	77

Fonte: o autor

Tabela 14 – Matriz de Confusão – Bateria de Teste – K-Médias

Amostras	Agrupado como	
	Disparos	Sons Impulsivos
Com Disparo	20 (<i>D</i>)	8 (<i>M</i>)
Sem Disparo	17 (<i>FP</i>)	37 (<i>TN</i>)
Total	37	45

Fonte: o autor

Tabela 15 – Comparativo das medidas de desempenho dos Métodos SVM e K-Médias

SVM				K-Médias			
ac	pr	se	sp	ac	pr	se	sp
0.7195	1.0	0.1786	1.0	0.6951	0.5405	0.7143	0.6852

Fonte: o autor

Porém, observando-se as totalizações dos resultados nas tabelas 13 e 14, nota-se uma tendência da SVM em classificar os sinais como sons impulsivos diversos. Já o K-Médias tem a tendência oposta, onde nota-se que ele classificou 45% dos sinais como disparos. Como consequência, poder-se-ia especular que o K-Médias coloca sinais com uma interpretação dúbia no cluster dos disparos e o SVM os classifica como impulsos. Isto não pode ser confirmado, nem, tampouco, os fatores que levam a esta interpretação dúbia podem ser levantados, em virtude da escassa informação que a base sonora utilizada possui sobre as circunstâncias em que os sinais foram gravados, não havendo informações sobre o transdutor utilizado, distância fonte/transdutor e, em muitos casos, nem informações sobre o armamento utilizado.

A baixa sensibilidade do método SVM pode ser prejudicial ao desempenho do sistema como um todo. Porém considerações mais profundas a respeito do que seria mais eficiente devem ser feitas.

Em virtude da natureza do sons envolvidos, onde disparos realizados a uma grande

distância podem ser facilmente confundidos com outros sons impulsivos provenientes de áreas mais próximas, em decorrência da perda de potência sonora, absorção de frequências importantes em sua identificação e interferências do ambiente ruidoso na qual o sistema como um todo está imerso, pode vir a ser uma vantagem um sistema que possua uma menor sensibilidade em prol de uma maior precisão, identificando como positivos apenas disparos que possuam características inconfundíveis.

A base sonora utilizada apresentava todas estas características, contemplando disparos de muito boa qualidade e outros gravados a uma grande distância, sendo que os resultados obtidos sob estas condições podem ser considerados satisfatórios, não cabendo um descarte imediato do método sem que este possa ser testado com uma base sonora gravada sob condições mais controladas.

6 Conclusão

Para um adequado funcionamento de um Sistema de Detecção e Reconhecimento de Disparos de Armas de Fogo, é necessário que o seu Módulo de Detecção de Sons Impulsivos conte com um algoritmo eficiente e preciso.

No presente trabalho foram apresentados três métodos para detecção de sons impulsivos, sendo os do RMS e CMF bastante conhecidos e documentados e o relativamente novo método da VM. Submetidos a testes de detecção com os parâmetros descritos na literatura, os métodos da CMF e VM mostraram uma boa performance mesmo sob condições de ruído ambiente acentuado e cadências elevadas. Já o RMS não foi eficiente o bastante, apresentando um elevado número de falsos positivos em todas as situações.

Submetidos a uma otimização de parâmetros, a CMF mostrou que um pequeno aumento na sensibilidade causa uma grande queda na precisão, demonstrando que os parâmetros preconizados na literatura são os mais adequados, uma vez que é necessário admitir-se alguma perda de detecção em detrimento a um número razoável de falsos positivos.

Já o VM apresentou um melhor desempenho, com índices de sensibilidade perfeitos e índice de falsos positivos bastante reduzidos. Desta forma conclui-se que, apesar do método da CMF ser bastante eficiente na detecção de sons impulsivos, ele deixa a desejar na sua precisão, não sendo possível conseguir um número que otimize tanto sensibilidade quanto precisão, ao contrário do que acontece com o VM, onde a máxima sensibilidade alcança níveis de precisão entre 0.82 e 1.00.

A fase seguinte sofreu algum prejuízo em sua avaliação face à natureza da biblioteca de sons disponíveis para os testes, onde muitos dos disparos utilizados foram gravados a uma grande distância, tornando-os semelhantes à outros sons impulsivos provenientes de áreas mais próximas. Além disso, as informações adicionais trazidas pelas bibliotecas de sons utilizadas eram escassas, não havendo dados sobre parâmetros vitais da gravação, como as circunstâncias em que os sinais foram gravados, o transdutor utilizado, distância fonte/transdutor e informações sobre o armamento utilizado.

No reconhecimento de disparos, cujo objetivo era diferenciá-los de outros sons impulsivos, optou-se por uma abordagem mais simples, visando economia de poder computacional, já que um dos objetivos do sistema aqui descrito é possuir a capacidade de rodar em sistemas embarcados e de baixo custo. Assim, escolheu-se uma análise no domínio do tempo baseado na envoltória do sinal sonoro filtrado pela técnica do *Spectral Noise Gating*, que passa a sofrer uma transformação de sinal unidimensional em uma imagem bidimensional, comparando-a com um padrão teórico para sons impulsivos baseado na equação de Friedlander.

Esta técnica no domínio do tempo não alcançou bons índices de desempenho, nem com a clusterização do K-Médias, nem com a classificação do SVM, uma vez que ambos os

tipos de sons, Disparos e Sons Impulsivos apresentaram características muito similares de envoltória, quando gravados em ambiente ruidoso e sujeito à reverberações.

Apesar de seus resultados não terem sido os melhores, acredita-se que o desempenho deste método possa ser melhorado, valendo-se de novos testes onde a duração do impulso de Friedlander e tamanho do sinal a ser analisado possam ser modificados e aplicados a uma base sonora mais consistente, de forma a determinar se existem valores ótimos para estes dois parâmetros.

Ainda no módulo de reconhecimento, elegeu-se uma abordagem simplificada das técnicas de análise no domínio das frequências, onde procurou-se identificar o conjunto de frequências características comuns a todos os disparos da base sonora levantada e que não pertenciam ao conjunto de frequências relevantes de um sinal impulsivo qualquer, pela análise do Coeficiente de Variação das FFT de um conjunto de sinais.

Submetidos a uma classificação, via SVM, e a uma clusterização, via K-Médias, obteve-se resultados medianos, onde a índice de acertos alcançou 71,95% e 69,51%, respectivamente. O SVM alcançou ótimos índices de precisão e especificidade, mas um índice de sensibilidade muito baixo. Já o K-Médias obteve índices medianos de precisão e especificidade, com um melhor desempenho de sensibilidade.

Notou-se, ainda, uma tendência da SVM em classificar os sinais sonoros como sons impulsivos diversos e o K-Médias em classificá-los como disparos. Disso poder-se-ia inferir que o K-Médias assinala como disparos sinais que possuam uma interpretação dúbia, ao contrário do SVM, que os classifica como impulsos. Não foi possível confirmar tal suspeita em virtude da escassa informação que a base sonora utilizada possui, ficando esta análise prejudicada.

Apesar de tal prejuízo, os resultados obtidos sob estas condições podem ser considerados satisfatórios, não cabendo um descarte imediato do método sem que este possa ser testado com uma base sonora gravada sob condições mais controladas, o que permitiria a confirmação da tendência da SVM em classificar sons dúbios como sons impulsivos.

Caso isto se mostre verdadeiro, a baixa sensibilidade do método SVM pode vir a ser uma vantagem, dando ao sistema uma menor sensibilidade em prol de uma maior precisão, identificando como positivos apenas disparos que possuam características inconfundíveis.

Novos estudos que possam contar com uma base sonora mais consistente, gravada sob condições e parâmetros conhecidos, que não tenha sofrido compactação de áudio **MP3** e com taxas de amostragem maiores que 44100, podem permitir uma melhor caracterização das frequências próprias de um disparo, identificação de quais frequências são absorvidas em função da distância, qual a influência do ambiente e da posição do transdutor em relação à fonte face aos resultados de reconhecimento e identificação filtros mais adequados à remoção de ruídos, ecos e reverberações.

Técnicas de reconhecimento de fala também podem ser úteis na fase de reconhecimento dos disparos, onde métodos como a análise espectrográfica, *Cepstrum*, *Mel-Frequency*,

Wavelets, *Hidden Markov Models*, dentre outras, podem alcançar resultados muito mais significativos que as técnicas aqui abordadas, devendo, porém, o seu custo computacional ser avaliado.

Por fim, nota-se que, mesmo empregando técnicas que primaram pela simplicidade e uso de um baixo poder computacional, é plenamente possível criar um Sistema de Detecção e Reconhecimento de Disparos de Armas de Fogo, baseado em um software leve e eficiente, permitindo o seu uso como software embarcado em dispositivos de baixo custo e com resultados satisfatórios.

Referências

- 1 BESACIER, L. et al. Automatic sound recognition relying on statistical methods, with application to telesurveillance. *Proceedings of International Workshop on Intelligent Communication Technologies and Applications, with Emphasis on Mobile Communications*, Institute of Microtechnology University of Neuchatel, Neuchatel Switzerland, p. 116–120, 1999.
- 2 WAISELFISZ, J. J. *Mapa da violência 2012: A cor dos homicídios no Brasil, 1a Edição*. Rio de Janeiro: CEBELA, FLACSO, Brasília, SEPP/PR 2012, 2001.
- 3 MURRAY, J.; CERQUEIRA, D. R. de C.; KAHN, T. Crime and violence in brazil: Systematic review of time trends, prevalence rates and risk factors. *Aggression and Violent Behavior*, v. 18, n. 5, p. 471 – 483, 2013. ISSN 1359-1789. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1359178913000566>.
- 4 ADORNO, S. Densidade de ocorrências de mortes violentas: Temporalidade e espaços dos homicídios dolosos na capital paulista. *Olhar São Paulo - Violência e Criminalidade*, Núcleo de Estudos da Violência, USP, São Paulo, SP, p. 27 – 34, 2008.
- 5 NEME, C. Projeto: O policiamento que a sociedade deseja : Análise das discussões em grupo com oficiais da polícia militar do estado de são paulo. Núcleo de Estudos da Violência, USP, São Paulo, SP, 2003.
- 6 VASEGHI, S. V. *Advanced Digital Signal Processing and Noise Reduction, Second Edition*. Chichester, UK: John Wiley & Sons Ltd, 2000.
- 7 DUFAUX, A. *Detection and Recognition of Impulsive Sound Signals*. Neuchatel Switzerland: D. Sc. thesis, Institute of Microtechnology. University of Neuchâtel, Neuchâtel, Switzerland, 2001.
- 8 DODGE, C.; JERSE, T. *Computer Music*. New York: Schirmer Books, 1997.
- 9 BECK, S. D.; NAKASONE, H.; MARR, K. W. Variations in recorded acoustic gunshot waveforms generated by small firearms. *Journal of the Acoustical Society of America*, Melville, NY, v. 129, n. 4, p. 1748–1759, 2011.
- 10 MAHER, R. C.; STEVEN; SHAW, R. *DECIPHERING GUNSHOT RECORDINGS*. Denver, CO, USA: AES 33rd International Conference, 2008.
- 11 FIRING Sequence. Disponível em: http://www.hunter-ed.com/washington/studyGuide/Firing-Sequence/20105001_700046703. Acesso em: 30 abr. 2015.
- 12 UNIVERSITY, T. P. S. *Gas Dinamyc Laboratory - Facilities - Explosives*. Disponível em: <http://www.mne.psu.edu/psgdl/Fac-Explosives.html>. Acesso em: 30 abr. 2015.
- 13 EXÉRCITO BRASILEIRO. *R-105*: Regulamento para a fiscalização de produtos controlados. Brasília - DF, 2000, 58 p.
- 14 PIMENTEL, R. B. *Dicionário de termos técnicos da área de armas e munições*. São Paulo: Ed. Magnun, 1994.

- 15 POLICE Departament. Disponível em: <http://policedepartament.files.wordpress.com/2008/11/corte-trans.jpg>. Acesso em: 06 ago. 2013.
- 16 ARMAS On Line. Disponível em: <http://cfpaula.files.wordpress.com/2011/05/enfield-revolver-diagram-100-dpi.jpg>. Acesso em: 06 ago. 2013.
- 17 ARMED Personal Defense. Disponível em: [http://www.arizonaccwpermit.com/store/media/ccp0/prod/rg/gun_parts_bolt_action_rifle_flash\[1\].gif](http://www.arizonaccwpermit.com/store/media/ccp0/prod/rg/gun_parts_bolt_action_rifle_flash[1].gif). Acesso em: 06 ago. 2013.
- 18 MISFITS Architecture. Disponível em: <http://betterarchitecture.files.wordpress.com/2010/06/ak-47-type-ii-part-dm-st-89-01131-5440.jpg>. Acesso em: 06 ago. 2013.
- 19 BUSHMASTER Firearms. Disponível em: http://www.bushmaster.com/img/firearms/90289_XM15.jpg. Acesso em: 06 ago. 2013.
- 20 MUNICAO 5.56x45 Ball da RUAG Ammotec. Disponível em: <http://jardimgrandearora.blogspot.com.br/2011/10/municao-556x45-ball-da-ruag-ammotec.html>. Acesso em: 30 abr. 2015.
- 21 CBC - Municoes para Pistolas. Disponível em: <http://www.cbc.com.br/municoes-para-pistolas-subcat-1.html>. Acesso em: 30 abr. 2015.
- 22 KAUPPINEN, I. Methods for detecting impulsive noise in speech and audio signals. *Proc. 14th Int. Conf. Digit. Signal Process*, Turku, Finland, v. 2, n. DSP2002, p. 967–970, 2002.
- 23 ARAKAWA, K.; NOMOTO, K. Nonlinear denoising filter for images with interactive evolutionary computing considering the subjective assessment. *in: Proc. IEEE Conference on Soft Computing in Industrial Applications Recognition (SMCia/08)*, v. 248, p. 264–268, 2008.
- 24 ARAKAWA, K. Median filter based on fuzzy rules and its application to image restoration. *Fuzzy Sets and Systems*, v. 77, p. 3–13, 1996.
- 25 SHAH P. S. TSAI, K. K. M.; KASPARIS, T. Cyclic motion detection for motion based recognition. *Pattern Recognition*, v. 21, p. 1591–1603, 1994.
- 26 BETTGE, D.; STARCEVIC, J. Quantitative description of wear surfaces of disc brakes using interference microscopy. *Wear Recognition*, v. 248, p. 121–127, 2001.
- 27 NELSON, R. S.; KASPARIS, T. Digital processing for non-stationary narrow-band interference suppression in fading channels. *Southeastcon Creative Technology Transfer - A Global Affair Recognition*, p. 408–412, 1994.
- 28 LEVY, D. *Introduction to Numerical Analysis*. [S.l.]: Scientific Department of Mathematics and Computation and Mathematical University of Maryland Modeling (CSCAMM), 2010. 87-89 p.
- 29 PAGE, E.; SHARKEY, B. Secures: System for reporting gunshots in urban environments. *Proc. SPIE*, v. 2497, n. May, p. 160–172, 1995.
- 30 BRACEWELL, R. *The Fourier Transform and its Applications*. Burr Ridge Parkway, Burr Ridge, Illinois, USA: McGraw-Hill Higher Education, 2000.

- 31 RUST, B. The fast fourier transform for experimentalists. part i. concepts. *Computing in Science Engineering*, v. 7, n. 2, p. 80–88, March 2005. ISSN 1521-9615.
- 32 LEVITT, H. Noise reduction in hearing aids: An overview. *Journal of rehabilitation research and development*, v. 38, n. 1, p. 111–121, 2001.
- 33 KRIM, H.; VIBERG, M. Two decades of array signal processing research: the parametric approach. *Signal Processing Magazine, IEEE*, v. 13, n. 4, p. 67–94, Jul 1996. ISSN 1053-5888.
- 34 INOUYE, J. M.; BLEMKER, S. S.; INOUYE, D. I. Towards undistorted and noise-free speech in an mri scanner: Correlation subtraction followed by spectral noise gating. *The Journal of the Acoustical Society of America*, v. 135, n. 3, p. 1019–1022, 2014. Disponível em: <http://scitation.aip.org/content/asa/journal/jasa/135/3/10.1121/1.4864482>).
- 35 WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining - Practical Machine Learning Tools and Techniques, Third Edition*. is an imprint of Elsevier Burlington, MA, USA: Morgan Kaufmann Publishers, 2011.
- 36 HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques, Second Edition*. 500 Sansome Street, Suite 400, San Francisco, CA: Morgan Kaufmann / Elsevier, 2006.
- 37 QINETIQ. *Swats*. Disponível em: <https://www.qinetiq-na.com>). Acesso em: 07/08/2015.
- 38 RAYTHEON. *Boomerang*. Disponível em: http://www.raytheon.com/capabilities/products/boomerang_warriorx/index.html). Acesso em: 07/08/2015.
- 39 BSS. *Pilar*. Disponível em: http://www.bssholland.com/#!en/p/gunshot-detection-system_2.html). Acesso em: 07/08/2015.
- 40 COBHAM. *Pinpoint*. Disponível em: <http://www.cobham.com/about-cobham/advanced-electronic-solutions/about-us/integrated-electronic-solutions/lowell/products-and-services/pinpoint%E2%84%A2-gunshot-detection-system.aspx>). Acesso em: 07/08/2015.
- 41 CHACON-RODRIGUEZ, A. et al. Evaluation of gunshot detection algorithms. *EEE Transactions on Circuits and Systems*, v. 58, n. 2, p. 363–373, February 2011.
- 42 FREIRE, I. L.; JR., J. A. A. Gunshot detection in noisy environments. *7th International Telecommunications Symposium*, 2010.
- 43 GOLDHOR, R. Recognition of environmental sounds. *ICASSP'93 International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, USA, I, p. 149–152, 1993.
- 44 ALTAFIM, R. et al. Piezoelectric sensor based on electret thermoforming technology. In: *Solid Dielectrics (ICSD), 2010 10th IEEE International Conference on*. [S.l.: s.n.], 2010. p. 1–3.
- 45 REIS, C. F. et al. Comparative analysis between impulsive detection methods applied on partial discharge acoustic signals. in: *2014 IEEE International Instrumentation and Measurement Technology Conference, Montevideo, Uruguay*, p. 418–421, 2014.

- 46 CAETANO, M.; BURRED, J. J.; RODET, X. Automatic segmentation of the temporal evolution of isolated acoustic musical instrument sounds using spectro-temporal cues. *13th Int. Conference on Digital Audio Effects (DAFx-10)*, p. 6–10, September 2010.
- 47 FEDORA 19. Disponível em: <https://fedoraproject.org>. Acesso em: 30 abr. 2015.
- 48 AUDACITY. Disponível em: <http://audacity.sourceforge.net>. Acesso em: 11 mai. 2014.
- 49 LAME. Disponível em: <http://lame.sourceforge.net>. Acesso em: 11 mai. 2014.
- 50 GNU Octave. Disponível em: <http://www.gnu.org/software/octave>. Acesso em: 11 mai. 2014.
- 51 QT Octave. Disponível em: <http://sourceforge.net/projects/qt octave.berlios>. Acesso em: 11 mai. 2014.
- 52 POSTGRESQL. Disponível em: <http://www.postgresql.org>. Acesso em: 30 abr. 2015.
- 53 LIBRE Office. Disponível em: <https://pt-br.libreoffice.org>. Acesso em: 30 abr. 2015.
- 54 YED. Disponível em: <http://www.yworks.com/en/products/yfiles/yed>. Acesso em: 30 abr. 2015.
- 55 WEKA. *Weka 3: Data Mining Software in Java*. 2013. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/index.html>. Acesso em: 27/01/2014.
- 56 FAWCETT, T. An introduction to roc analysis. *Pattern Recog Lett*, v. 27, p. 861–874, 2006.
- 57 PRASS, F. S. *Algoritmo de k-means*. Disponível em: <http://fp2.com.br/blog/index.php/2013/algoritmo-de-k-means/>. Acesso em: 07/08/2015.

APÊNDICE A – Algoritmo do K-Médias

```
1  Especifique k
2  Selecione os k objetos que serão os centróides dos agrupamentos
3  para todos os objetos restantes faça
4      Calcule a distância entre o elemento e os centróides
5      Adicione o elemento ao agrupamento que possuir a menor
        distância
6      Recalcule o centróide do agrupamento
7  fim para
8  para todos os k agrupamentos faça
9      Calcule a Soma de Quadrados Residual
10 fim para
11 repita
12     para todos os n elementos faça
13         Mova o elemento para os outros agrupamentos
14         Recalcule a Soma de Quadrados Residual
15         se soma dos Quadrados Residual diminuiu então
16             O objeto passa a fazer parte do grupo que produzir maior
                ganho
17             Recalcule a Soma de Quadrados Residual dos grupos
                alterados
18         fim se
19     fim para
20 até Número de interações = i ou Não ocorra mudança de objetos
```

Fonte: [57]

APÊNDICE B – Introdução de dados no WEKA

Para introduzir dados no WEKA é necessário formatá-los, colocando-os em um arquivo do tipo ARFF (Attribute-Relation File Format). Este é um arquivo de texto que irá descrever os diversos atributos de um dado conceito, para em seguida listar todas as instancias levantadas.

Inicia-se um arquivo ARFF com a instrução **@relation** *<relation-name>* que identificará todo o conjunto de dados. A seguir, cada um dos atributos é descrito por uma instrução do tipo **@attribute** *<attribute-name>* *<datatype>* onde *<datatype>* é uma dentre as seguintes palavras chaves:

- **NUMERIC** – para identificar atributos Reais ou Inteiros;
- **STRING** – para atributos que contenham caracteres de texto;
- **DATE** – para atributos do tipo data. O formato da data de ser formatado para cada atributo. O formato padrão será “yyyy-MM-dd’T’HH:mm:ss”.

Para valores nominais, os valores aceitáveis para o atributo devem ser passados entre chaves. A instrução **@data** demarca o início da seção de dados do arquivo ARFF. Cada linha representa uma instância de dados, onde os atributos serão separados por uma virgula. A ordem de cada atributo na linha deve ser a mesma da definição de atributos realizada anteriormente.

Assim, o arquivo ARFF da Bateria de Teste tem a seguinte estrutura:

```
@relation BateriaTeste
@attribute Tipo I,D
@attribute 98Hz numeric
@attribute 100Hz numeric
@data
D,82.77,122.239
I,161.133,112.029
?,21.714,24.861,
```

Valores desconhecidos são representados por um ponto de interrogação, como na terceira linha de dados, onde não se sabe se o sinal em questão é um disparo ou som impulsivo.