



# Desenvolvimento de Sistema de Avaliação da Qualidade de Reprodução de Áudio em Alto-falantes com Resposta Pobre em Baixas Frequências

Trabalho de Conclusão de Curso

Engenharia de Computação

ALEXANDRE DIEGO SANTOS SILVA

Orientador: Prof. Sérgio Campello Oliveira



Alexandre Diego Santos Silva

**Desenvolvimento de Sistema de Avaliação da Qualidade  
de Reprodução de Áudio em Alto-falantes com Resposta  
Pobre em Baixas Frequências**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Engenharia de Computação  
Escola Politécnica de Pernambuco  
Universidade de Pernambuco

Orientador: Prof. Sérgio Campello Oliveira

Recife - PE, Brasil  
Novembro de 2017

## MONOGRAFIA DE FINAL DE CURSO

### Avaliação Final (para o presidente da banca)\*

No dia 14 de dezembro de 2017, às 10:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente **ALEXANDRE DIEGO SANTOS SILVA**, orientado pelo professor **Sérgio Campello Oliveira**, sob título **Desenvolvimento de Sistema de Avaliação da Qualidade de Reprodução de Áudio em Alto-Falantes com Resposta Pobre em Baixas Frequências**, a banca composta pelos professores:

**José Paulo G. de Oliveira**

**Sérgio Campello Oliveira**

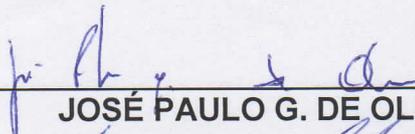
Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada       Aprovada com Restrições\*       Reprovada

e foi-lhe atribuída nota: 9,0 ( nove )

\*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 07 dias para entrega da versão final da monografia a contar da data deste documento.

  
\_\_\_\_\_  
**JOSÉ PAULO G. DE OLIVEIRA**

  
\_\_\_\_\_  
**SÉRGIO CAMPELLO OLIVEIRA**

\* Este documento deverá ser encadernado juntamente com a monografia em versão final.

Alexandre Diego Santos Silva

Desenvolvimento de Sistema de Avaliação da Qualidade de Reprodução de Áudio em Alto-falantes com Resposta Pobre em Baixas Frequências/ Alexandre Diego Santos Silva. – Recife - PE, Brasil, Novembro de 2017-

53 p.

Orientador: Prof. Sérgio Campello Oliveira

Trabalho de Conclusão de Curso – Engenharia de Computação

Escola Politécnica de Pernambuco

Universidade de Pernambuco, Novembro de 2017.

1. Som. 2. Melhoramento de graves. 3. Psicoacústica. 4. Avaliação de qualidade de áudios. 5. Regressão. I. Prof. Sérgio Campello Oliveira. II. Universidade de Pernambuco. III. Escola Politécnica de Pernambuco. IV. Título

*Dedico este trabalho a Deus, que me deu forças para seguir na luta mesmo quando tudo parecia desmoronar e a minha maior fonte de motivação: minha família.*

# Agradecimentos

Em primeiro lugar gostaria de agradecer a Deus, minha fonte de fé, paciência e força, minha fonte de vida e meu refúgio, sem Ele nada seria possível.

Agradeço a todos os familiares que me incentivaram durante esse percurso da minha vida, em especial aos meus pais Moisés e Maria, minha avó Alice, minha irmã Sabrina e meu sobrinho Enzo, por sempre acreditarem em mim e por todo o suporte em todas as horas, por jamais demonstrarem dúvidas de que chegaria onde cheguei. Apesar da distância, sempre ao meu lado.

Também gostaria de dedicar um espaço para agradecer àqueles que estavam ao meu lado nessa trajetória, foram muitos altos e baixos, mas jamais estive só, sempre pude contar com uma mão amiga para me acolher e guiar de volta à estrada. Os que estavam comigo do primeiro *Hello World* até este momento, alguns seguirão em minha trajetória e pretendo reencontrar os demais nos caminhos da vida. Não posso deixar de citar quatro nomes que estão marcados e jamais serão apagados: Airton Soares, Bárbara Mena, Cauanne Linhares e Thaís Melo, mais que amigos, uma família que construí, estes que acompanharam de perto diversas das fases que atravessei. Minha sincera gratidão.

A todos os professores com os quais tive a oportunidade de aprender, em especial aos do Ensino Médio, professores e funcionários da Escola de Referência em Ensino Médio de Gravatá, que me ensinaram além do que esperava, me levando a momentos de reflexão, me fazendo pensar no futuro e buscar o meu caminho. Sem os ensinamentos e conselhos deles não estaria aqui. Aos professores da Escola Politécnica de Pernambuco, por compartilharem seu conhecimento e me motivarem a perseguir mais e mais conhecimento, também quero agradecer o apoio e incentivo de Ana Georgina (in memoriam), com quem tive a honra de trabalhar. Obrigado!

Por fim, meu muito obrigado a meu orientador, o professor Dr. Sérgio Campello Oliveira, por me passar um pouco de seu conhecimento nas áreas em que tive a oportunidade de trabalhar com ele e por me incentivar e acompanhar durante o desenvolvimento deste trabalho, acima de tudo por ter dedicado parte de seu tempo me orientando no fechamento deste ciclo da minha vida acadêmica.

*“(...) I didn’t know my own strength  
Survived my darkest hour  
My faith kept me alive  
I picked myself back up  
Hold my head up high  
I was not build to break  
I didn’t know my own strength (...)”*

***I didn’t know my own strength - Whitney Houston***

# Resumo

A implementação de sistemas sonoros adaptáveis em automóveis vem sendo objeto de estudo de desenvolvedores a muito tempo, inicialmente impossibilitado pela dimensão em espessura e peso dos alto-falantes, que provocaria um grande aumento de peso e espaço utilizado, dado que tais sistemas necessitam de vários alto-falantes. Esse problema inicial foi solucionado com a criação dos alto-falantes de material piezoelétrico, também conhecidos como alto-falantes de filme fino, estes possuem uma espessura e peso menores, porém, devido a estrutura de sua superfície de vibração, eles possuem deficiência na reprodução de sons com baixas frequências (tons graves), inviabilizando sua utilização para os sistemas descrito, uma vez que a reprodução seria prejudicada. Este projeto propõe o desenvolvimento de um sistema de melhoria para reprodução de tons graves em alto-falantes que por natureza não possuem a capacidade de reproduzi-los, para tal serão utilizadas técnicas de processamento de sinal para simular o fenômeno de *missing fundamental*. A proposta é que o sistema possa autoavaliar a qualidade do processamento realizado e determinar se realiza um novo processamento ou retorna o atual, para essa detecção será gerado um processo de predição utilizando análises físicas dos sinais de áudios processados a fim de se obter a qualidade subjetiva dos mesmos. Foi desenvolvida uma ferramenta de avaliação de qualidade para realização de testes a fim de se obter um conjunto de dados para treinar o processo de predição. O resultado obtido para o modelo de predição mostrou-se favorável, mostrando ser possível se determinar a qualidade do áudio a partir de suas características físicas. Os resultados obtidos com os testes realizados mostraram que a nossa proposta é possível e tem potencial de crescimento.

**Palavras-chave:** Processamento de Sinais, Melhoramento de Graves, Psicoacústica, Avaliação de Qualidade de Áudios, Regressão

# Abstract

The implementation of adaptive sound systems in cars has been the object of study by developers for a long time, initially made impossible by the dimension in thickness and weight of the speakers, that would cause a great increase of weight and space used, since such systems need several loudspeakers. This initial problem was solved with the creation of piezoelectric material loudspeakers, also known as thin film loudspeakers, these have a smaller thickness and weight, however, due to the structure of their vibration surface, they are deficient in reproduction of low frequency sounds (bass tones), making it unfeasible for the systems described, since reproduction would be impaired. This project proposes the development of an improvement system for reproduction of bass tones in loudspeakers that by nature do not have the capacity to reproduce them, for this will be used signal processing techniques to simulate the Missing Fundamental phenomenon. The proposal is that the system can self-evaluate the quality of the processing performed and determine whether to perform a new processing or return the current one, for that detection will be generated a process of prediction using physical analysis of the audio signals processed to obtain subjective quality of the same. A quality assessment tool was developed to carry out tests with objective to obtain a set of data to train the prediction process. The result obtained for the prediction model was favorable, showed being possible to determine the quality of the audio from its physical characteristics. The results obtained with the tests performed showed that our proposal is possible and has growth potential.

**Keywords:** Digital Signal Processing, Bass Enhancement, Psychoacoustic, Audio Quality Evaluation, Regression

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
1.1	Estrutura da Monografia	3
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>5</b>
2.1	<i>Bass Enhancement</i>	5
2.2	Psicoacústica	6
2.2.1	O som	6
2.2.2	A orelha humana	7
2.3	Processamento Digital de Sinais	8
2.3.1	Representação de Sinais Digitais	8
2.3.2	Filtros Digitais	9
2.3.3	NLD - <i>Nonlinear Device</i>	10
2.4	Como avaliar a qualidade do áudio	11
2.4.1	Avaliação Subjetiva	12
2.4.2	Avaliação Objetiva	12
2.5	Regressão Linear	12
2.5.1	Redes Neurais	13
<b>3</b>	<b>DESENVOLVIMENTO</b>	<b>16</b>
3.1	Metodologia	16
3.1.1	Os áudios de teste	16
3.1.2	As técnicas de processamento	17
3.1.3	O ambiente de testes	18
3.2	Modelo do sistema	19
3.3	<i>Bass Enhancement</i> e Avaliação de Qualidade	21
3.3.1	Modelo Inicial	22
3.3.2	Modelo Adaptado	25
<b>4</b>	<b>TESTES E RESULTADOS</b>	<b>33</b>
4.1	Avaliações Subjetivas	33
4.2	Avaliações Objetivas	36
4.3	Regressão	37
4.3.1	Regressão utilizando Validação Cruzada com $k = 5$	39
4.3.2	Regressão utilizando Validação Cruzada com $k = 25$	41
<b>5</b>	<b>CONCLUSÕES</b>	<b>43</b>

6	TRABALHOS FUTUROS . . . . .	44
	REFERÊNCIAS . . . . .	46
	ANEXO A – SCRIPT PARA PROCESSAMENTO COM <i>ARC-TANGENT</i> <i>SQUARE ROOT</i> . . . . .	49
	ANEXO B – SCRIPT PARA PROCESSAMENTO COM <i>FULL WAVE</i> <i>INTEGRATOR</i> . . . . .	50
	ANEXO C – SCRIPT PARA PROCESSAMENTO COM <i>SOFT CLIP-</i> <i>PING</i> . . . . .	53

# Lista de ilustrações

Figura 1 – Modelo de sistema proposto para melhoria de graves . . . . .	3
Figura 2 – Exemplo de percepção de grave utilizando conjunto de harmônicas . . .	6
Figura 3 – Os limiares e as regiões de audibilidade. . . . .	7
Figura 4 – Percepção da frequência fundamental não existente . . . . .	8
Figura 5 – Representação gráfica de sinais digitais . . . . .	9
Figura 6 – Diagrama de blocos de inversão espectral para geração de filtro passa-altas	10
Figura 7 – Ilustração do processo de execução do NLD . . . . .	11
Figura 8 – Relacionamento entre variáveis vistos por gráficos de dispersão . . . . .	13
Figura 9 – Ilustração do modelo matemático criado por McCulloch e Pitts para representar o neurônio artificial . . . . .	14
Figura 10 – Diagrama do sistema de testes . . . . .	18
Figura 11 – Alto-falante utilizado nos testes e sua curva de resposta . . . . .	19
Figura 12 – Análise estatística dos voluntários dos testes subjetivos . . . . .	19
Figura 13 – Representação do sistema proposto . . . . .	20
Figura 14 – Diagrama de blocos do processamento de melhoria inicial . . . . .	22
Figura 15 – Ferramenta de avaliação de qualidade . . . . .	24
Figura 16 – Gráfico de análise da média das avaliações utilizando o modelo inicial .	24
Figura 17 – Diagrama de blocos do processamento de melhoria adaptado . . . . .	26
Figura 18 – Armazenamento das avaliações dos voluntários . . . . .	28
Figura 19 – Interface de registro de voluntário e configuração de diretórios na se- gunda versão da ferramenta de avaliação . . . . .	29
Figura 20 – Ferramenta de avaliação de qualidade versão 2.0 . . . . .	29
Figura 21 – Detalhamento da planilha de avaliações . . . . .	30
Figura 22 – Ajuste das avaliações subjetivas por meio de pré-processamentos . . . . .	35
Figura 23 – Análise de pontos discrepantes sobre conjunto de avaliações pré-processados	35
Figura 24 – Configuração de sessão de treinamento no <i>Regression Learner</i> . . . . .	39
Figura 25 – Gráficos de Dispersão utilizando <i>5-fold</i> . . . . .	40
Figura 26 – Gráficos de Dispersão utilizando <i>25-fold</i> . . . . .	42

# Lista de tabelas

Tabela 1 – Frequências de corte para filtros Passa baixa e Passa alta . . . . .	21
Tabela 2 – Nomenclatura e parâmetros dos processamentos utilizando o modelo inicial . . . . .	23
Tabela 3 – Nomenclatura e parâmetros dos processamentos utilizando o modelo adaptado . . . . .	26
Tabela 4 – Resultados obtidos para avaliação objetiva dos áudios processados . . .	31
Tabela 5 – Relação e descrição das MOVs utilizadas . . . . .	32
Tabela 6 – Análise por média das avaliações após aplicação do tratamento de dados - <i>bad</i> . . . . .	36
Tabela 7 – Relação de melhores modelos de regressão por conjunto de dados, determinadas pelo RMSE ( $k = 5$ ) . . . . .	40
Tabela 8 – Relação de melhores modelos de regressão por conjunto de dados, determinadas pelo RMSE ( $k = 25$ ) . . . . .	41

# Lista de abreviaturas e siglas

ATSR	<i>Arc-Tangent Square Root</i>
DSP	<i>Digital Signal Processing</i>
FFT	<i>Fast Fourier Transform</i>
FWI	<i>Full Wave Integrator</i>
FWR	<i>Full Wave Rectifier</i>
GLM	<i>Generalized Linear Model</i>
GPR	<i>Gaussian Process Regression</i>
HPF	<i>High Pass Filter</i>
Hz	Hertz
I.T.U.	<i>International Telecommunications Union</i>
LPF	<i>Low Pass Filter</i>
$MFPD_B$	<i>Maximum Filtered Probability of Detection</i>
MOV	<i>Model Output Variable</i>
NLD	<i>Nonlinear Device</i>
NPS	Nível de Pressão Sonora
ODG	<i>Objective Diference Grade</i>
PEAQ	<i>Perceptual Evaluation of Audio Quality</i>
RMSE	<i>Root Mean Square Error</i>
RNA	Redes Neurais Artificiais
SC	<i>Soft Clipping</i>
SVM	<i>Support Vector Machine</i>

# Lista de símbolos

$\omega$	Frequência angular
$\omega_c$	Frequência de corte (Filtros)
$\alpha$	Intensidade das harmônicas ímpares (ATSR)
$\beta$	Padrões harmônicos das harmônicas ímpares (ATSR)
$\psi$	Intensidade das harmônicas pares (ATSR)
$\zeta$	Padrões harmônicos das harmônicas pares (ATSR)
$K$	Curva de distorção (SC)
$\delta$	Desvio padrão

# 1 Introdução

Atualmente a busca por conforto acústico deixou de ser um item acessório na vida das pessoas, passando a ser decisivo na escolha de um ambiente de vivência, uma vez que ruídos afetam a saúde de quem está inserido no ambiente, não apenas psicologicamente falando, mas também fisicamente. Existe uma degradação da saúde física e psicológica decorrente da submissão dos passageiros, comissários e pilotos, a ruídos e as vibrações do avião durante o voo [1]. No ambiente veicular automotivo não é diferente, a qualidade do som é um dos fatores importantes na avaliação de performance do automóvel, neste tipo de ambiente todos os sons influenciam na saúde física e mental do motorista e passageiros, desde o ruído proveniente do atrito das rodas com o solo até a música executada dentro do veículo.

Com o aumento da produção de veículos e o barateamento dos itens no setor, cada vez mais se aumenta o número de veículos circulando. Dados apontam que só no Brasil, em 2016 cerca de 35,6 milhões de automóveis circulavam [2], sem contar com outros veículos como caminhões, ônibus e comerciais leves. O aumento da circulação de automóveis implica diretamente no aumento do tempo gasto dentro dos mesmos, seja como motorista ou passageiro, e como mencionado anteriormente, a busca por conforto acústico é decisivo na escolha do ambiente que se está inserido. Logo, um automóvel que possibilita ao usuário um ambiente livre de ruídos e sons desagradáveis é preferível.

Além disso, os automóveis atuais apresentam, em alguns casos, tecnologias de facilidade, permitindo que os usuários utilizem o sistema de áudio do automóvel para realizar uma chamada de voz utilizando um sistema *wireless*, por exemplo, porém quando se está com um passageiro e se deseja ter uma conversa privada, não é possível atualmente sem a utilização de fones de ouvido ou utilizando diretamente o fone do aparelho. Para resolução de casos como este são propostos sistemas sonoros adaptáveis que utilizam de mascaramento sonoro para cancelar sons indesejados [3]. Os sistemas de som adaptáveis poderiam desde solucionar o problema apresentado, possibilitando a privacidade da conversa, até anular sons do automóvel permitindo a cada passageiro ouvir uma música diferente.

Os alto-falantes de material piezoelétrico ou de filme fino, como também são chamados, surgiram como solução para problemas antigos ao qual desenvolvedores se deparavam ao tentar realizar a implementação de sistemas sonoros adaptáveis: a espessura e o peso dos alto-falantes. Na construção de sistemas adaptáveis faz-se necessária a adição de vários alto-falantes de forma que possam ser realizadas anulações de som em pontos específicos por meio da geração de frequências destrutivas, sendo assim, acumular alto-falantes espessos e pesados era desfavorável para a construção de tais sistemas.

A alteração dos alto-falantes presentes atualmente em automóveis pelos de filme fino causariam uma redução em espessura e peso dos sistemas de som veiculares, em contrapartida a resposta em frequência deste tipo de alto-falante não é boa para as baixas frequências (graves) devido a estrutura de sua superfície de vibração [4] e isto impede o desenvolvimento dos sistemas sonoros mencionados utilizando apenas estes alto-falantes. Porém a utilização de alto-falantes menores ainda é útil em diversas outras aplicações, e para contornar o problema de reprodução de graves citado, este trabalho tem por objetivo desenvolver um sistema capaz de: a partir de um áudio de entrada gerar um novo sinal de áudio para reprodução que para o ouvinte soe como o original apresentando os graves.

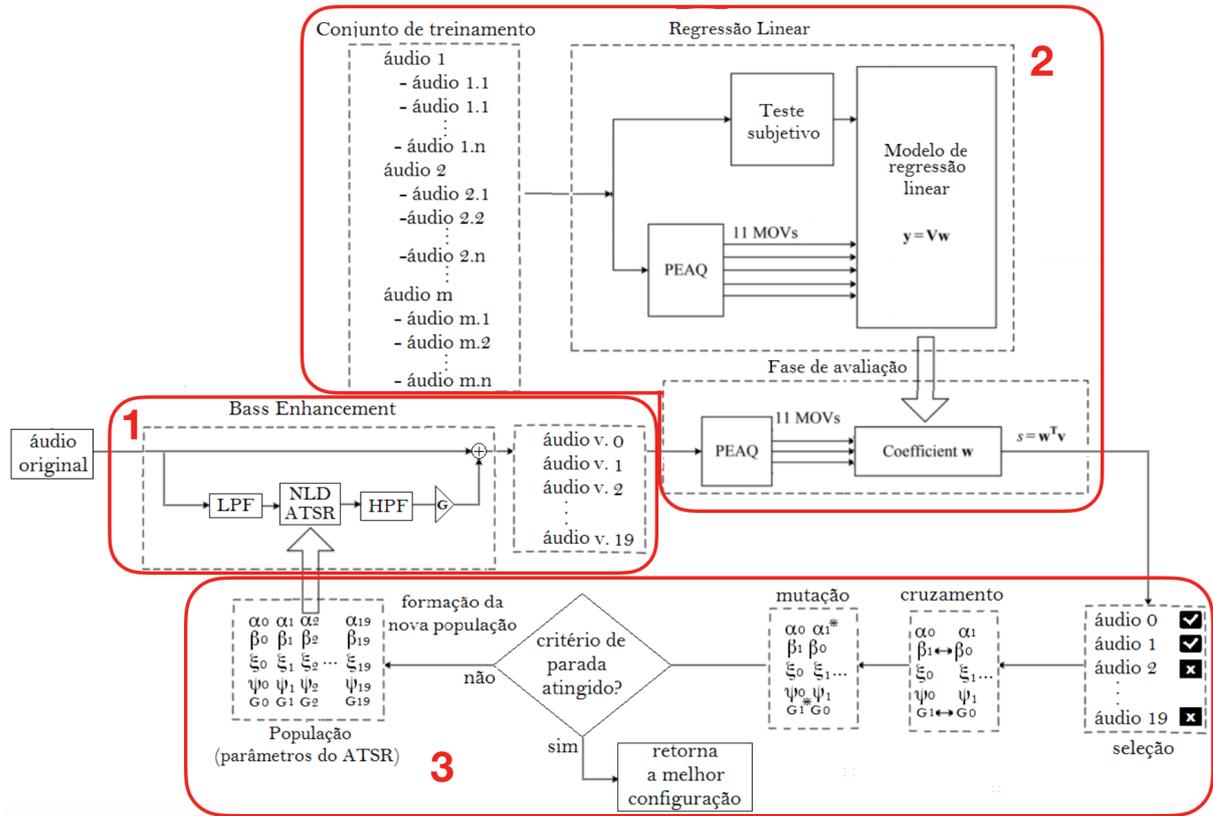
O trabalho realizado se baseia no desenvolvimento de sistema de melhoramento de reprodução de graves em alto-falantes com deficiência de reprodução destes, onde através do estudo da resposta em frequência dos alto-falantes realizar-se-á o aprimoramento dos algoritmos de processamento digital de sinais com o objetivo de realizar o reforço da reprodução de sons graves.

Através de processamentos computacionais dos modelos matemáticos foram gerados conjuntos de áudios com processamentos distintos a fim de determinar os melhores conjuntos de parâmetros para os algoritmos de processamento, para isso foram realizados testes qualitativos e quantitativos para mensurar a melhoria da resposta em frequência dos alto-falantes de filmes finos. E é a partir desses testes que circunda o desenvolvimento deste trabalho.

Pretende-se que o sistema proposto seja capaz de decidir se o processamento gerado tem boa qualidade para a reprodução, caso não tenha ele deve realizar um novo processamento a partir de um conjunto otimizado de parâmetros. Para que isso seja possível foi este trabalho apresenta o desenvolvimento de um sistema de avaliação de qualidade a partir da regressão entre os testes qualitativos e quantitativos a fim de gerar um modelo de predição de qualidade. Deseja-se que, apenas analisando características físicas, se possa determinar se um áudio processado possui boa qualidade para reprodução de graves em alto falantes com deficiência de reprodução destes.

A Fig. 1 apresenta o modelo do sistema completo de melhoramento citado. As áreas em destaque enumeradas por 1 e 3 indicam, respectivamente, o processo de melhoria do áudio (geração do áudio com processamento de melhoria) e o processo de otimização de parâmetros e decisão de envio de áudio para reprodução. A área 2 apresenta o processo desenvolvido por este trabalho, englobando os testes qualitativos (Teste Subjetivo), quantitativos (PEAQ) e a regressão utilizando ambos os testes.

Figura 1 – Modelo de sistema proposto para melhoria de graves



Fonte: extraído de [5] - Adaptado

## 1.1 Estrutura da Monografia

Este trabalho está organizado em 6 capítulos: 1 - Introdução, 2 - Fundamentação Teórica, 3 - Desenvolvimento, 4 - Testes e Resultados, 5 - Conclusões e 6 - Trabalhos Futuros.

O Capítulo 2 apresenta todo o referencial teórico que dá base ao trabalho, abordando assuntos necessários para o seu entendimento, tais como Psicoacústica e Regressão Linear, contando também com as revisões bibliográficas deste.

Os Capítulos 3 e 4 apresentam o sistema proposto para a melhoria da reprodução dos áudios, a metodologia de realização do projeto, onde se expõem as técnicas utilizadas para os processamentos, os áudios utilizados como base de treinamento e o ambiente de testes. Se apresenta também todo o desenvolvimento propriamente dito, com os modelos para os processamentos dos áudios, os resultados obtidos a partir destes, os processos de avaliação de qualidade abordados, descrevendo suas aplicações e resultados obtidos, finalizando com a aplicação do sistema de regressão utilizando estes resultados.

O Capítulo 5 aborda todos os resultados deste trabalho bem como as dificuldades

encontradas durante a sua execução.

Por fim o Capítulo 6 será reservado com o intuito de listar possíveis melhorias e passos a serem executados para a evolução do desenvolvimento e/ou melhoria dos resultados obtidos.

## 2 Fundamentação Teórica

Este capítulo visa a apresentação de conceitos básicos a respeito de melhoramento de grave (*Bass Enhancement*), processamento de sinais, psicoacústica, avaliação objetiva de qualidade de áudio e redes neurais como sistema de regressão. Esses conceitos embasam o desenvolvimento deste trabalho, portanto são necessários para o seu melhor entendimento.

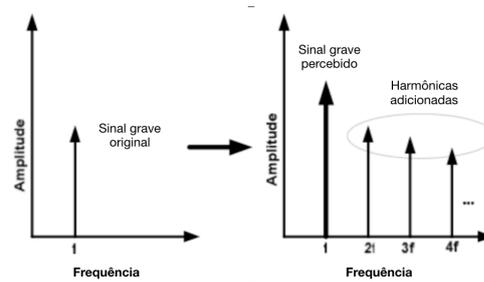
### 2.1 *Bass Enhancement*

Às vezes a utilização de alto-falantes grandes não é viável pelas dimensões do aparelho de reprodução, um celular por exemplo, logo existem restrições relacionadas ao tamanho e/ou custo do sistema sonoro. Portanto, os dispositivos são geralmente pequenos, infelizmente uma das características mais conhecidas dos pequenos alto-falantes é uma resposta fraca de baixa frequência (tons graves). Na prática, isso significa que uma parte significativa do sinal de áudio pode não ser reproduzida (suficientemente) pelo alto-falante, provocando em muitos casos um impacto negativo na qualidade do som como um todo. Porém existem casos, como o fone de ouvido, que não exibem esta deficiência, isso se deve ao revestimento que direciona as ondas sonoras ao tímpano e o lacre que fornece um melhor isolamento do som [6]. Também pode ser provocado pela proximidade da orelha que serve como caixa acústica para o aparelho.

Dado o contexto explicitado acima, o *Bass Enhancement* - Melhoramento de grave, surge com o propósito de ajustar o áudio de forma a melhorar sua reprodução de grave. O trabalho desenvolvido propõe uma metodologia de processamento para melhoramento de áudio utilizando-se da psicoacústica.

A reação do ouvido humano à exposição de uma baixa frequência pode ser simulado utilizando de uma combinação de “altas” frequências, as vibrações internas são semelhantes, dando a sensação de se ouvir a mesma coisa. Aproveitando este fenômeno, o *bass enhancement* é capaz de calcular as ondas de alta frequência (harmônicas ou múltiplas) que geram essa mesma sensação auditiva de uma de baixa frequência, possibilitando a execução em pequenos alto-falantes. A Fig. 2 demonstra um exemplo do funcionamento do *bass enhancement*, onde um conjunto de harmônicas geradas provoca a sensação de audição da frequência fundamental.

Figura 2 – Exemplo de percepção de grave utilizando conjunto de harmônicas



Fonte: extraído de [7] - Adaptado.

## 2.2 Psicoacústica

Como uma subdisciplina da psicofísica<sup>1</sup>, a psicoacústica estuda a relação entre estímulos sonoros e as sensações auditivas decorrentes destes estímulos [8].

No processo de entendimento da psicoacústica é necessário ter como base alguns conceitos relacionados ao som, ao ouvido humano e a relação entre estes.

### 2.2.1 O som

O som se caracteriza por flutuações de pressão em um meio passível de compressão, quando a amplitude destas flutuações e as frequências, com que elas se repetem, estiverem dentro de determinadas faixas de valores [9]. Quando a vibração provocada pelas flutuações estimula o aparelho auditivo, ela é chamada vibração sonora. Assim, o som é definido como qualquer vibração, ou conjunto de vibrações, ou ondas mecânicas que podem ser ouvidas [10].

Quando as variações de pressão e a frequência com qual ocorre a propagação dos sons estiverem dentro dos limites especificados pela fisiologia da orelha humana então acontecerá a percepção desses sons [11].

Alguns aspectos físicos do som, são apresentados [11][12] e seu conhecimento e entendimento são importantes:

- **Amplitude ( $A$ ):** é uma constante positiva, cujo valor depende do modo como o movimento foi produzido, caracterizando a intensidade do som; quanto maior a compressão, maior é o volume do som; tem como unidade o  $N/m^2$  ou  $Pa$ .
- **Comprimento de onda ( $\lambda$ ):** pode ser definido como a distância física no ar entre um pico de onda até o próximo.

<sup>1</sup> Área da ciência que estuda as relações entre as sensações subjetivas e os estímulos físicos e estabelece relações quantitativas entre eles

- **Frequência ( $f$ ):** é o número de vezes que a oscilação (de pressão) é repetida, na unidade de tempo. A frequência é expressa em ciclos por segundo (*cps*) ou em *Hertz* (*Hz*).
- **Harmônica:** são ondas sonoras de frequências múltiplas inteiras da onda fundamental.

## 2.2.2 A orelha humana

O ouvido, atualmente denominado orelha, é um complexo órgão sensorial relacionado a dois sentidos: a audição e o equilíbrio. Os receptores especializados na captação de estímulos mecânicos permitem detectar sons e analisar a posição do corpo [13].

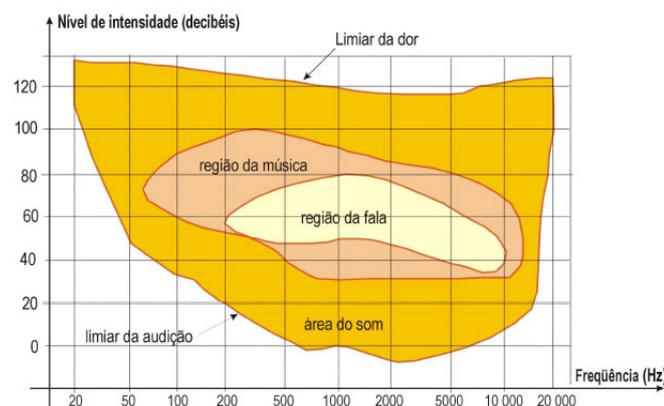
### A orelha e os sons

A orelha humana permite a distinção das qualidades do som e sua direção, ou seja, é possível determinar o timbre, volume, tom e também a posição de origem do som. Porém nem todos os sons são percebidos pela orelha humana, apenas aqueles que possuem frequência compreendida na faixa de 20 *Hz* a 20 *KHz* e dependente de seu nível de intensidade ou nível de pressão sonora (NPS).

A percepção sonora se dá através da oscilação do ar. No ouvido, o tímpano vibra e gera sinais elétricos interpretados pelo cérebro como estímulo auditivo.

Segundo Gerges [9], a percepção auditiva depende de diversas condições subjetivas e de outras variáveis e para avaliar com mais precisão esses aspectos, pesquisadores criaram um índice chamado de nível de audibilidade. Uma representação do campo de audibilidade e seus limiares pode ser visto na Fig. 3, ela mostra as áreas audíveis onde se concentra a região da fala e música.

Figura 3 – Os limiares e as regiões de audibilidade.

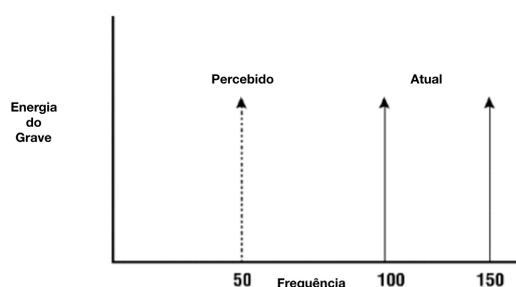


Fonte: extraído de [14].

### Missing Fundamental

É um fenômeno psicoacústico, conhecido como fenômeno da frequência fundamental suprimida, ele consiste na interação entre os sinais emitidos e a orelha humana, onde ao se emitir as harmônicas mais altas do sinal, o ouvido humano pode reconstruir a frequência fundamental, dando a sensação de audição desta [15]. Isso provoca a sensação de audição dos sons graves, melhorando a resposta em frequência, mesmo sem a emissão física dos tons graves. A Fig. 4 representa o princípio do *missing fundamental*, onde as frequências atuais (harmônicas) provocam a sensação de se ouvir a frequência fundamental.

Figura 4 – Percepção da frequência fundamental não existente



Fonte: extraído de [7] - Adaptado.

## 2.3 Processamento Digital de Sinais

O processamento de sinais, ou do inglês DSP (*Digital Signal Processing*) tem por objetivo extrair informações de sinais e/ou torná-los mais apropriados a uma determinada aplicação [16]. O processamento digital de sinais propriamente dito, é em síntese uma série de conceitos que pode ser traduzido na aplicação de algoritmos computacionais ou em um dispositivo eletrônico para realizar operações específicas sobre dados digitais [17]. Diversas são as aplicações possíveis de DSPs, tais como reconhecimento de voz, extração e reconhecimento de padrões em imagens, transmissão de rádio e TV, filtragem digital de sinais, dentre outras.

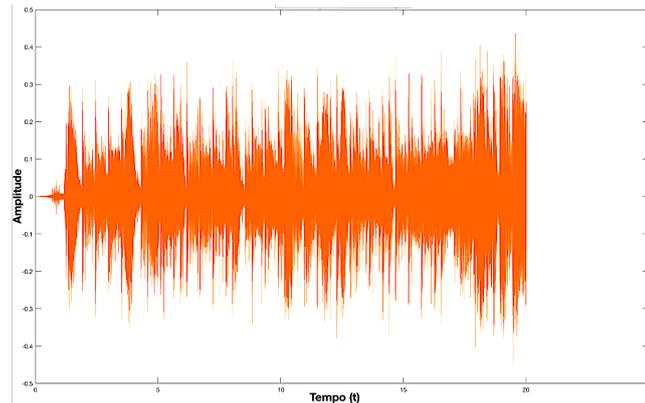
### 2.3.1 Representação de Sinais Digitais

Um sinal de áudio emitido naturalmente é um sinal contínuo, ou seja, ele está definido para qualquer instante do tempo. São várias as formas de representação do sinal contínuo, para o desenvolvimento deste trabalho, apenas duas são relevantes: representação do sinal no domínio do tempo e no domínio da frequência.

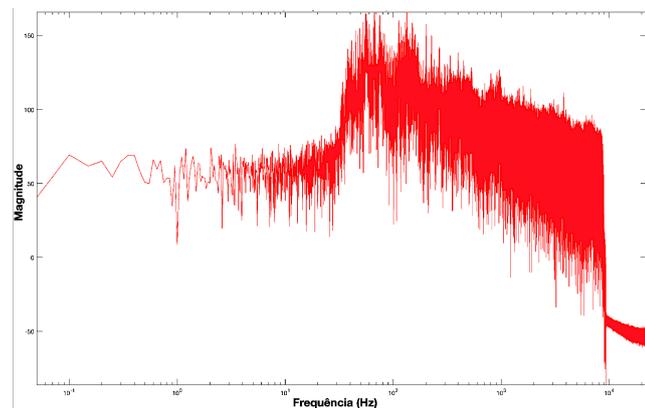
Na matemática, sinais de áudio no domínio do tempo são definidos por  $x(t)$ , onde  $t$  representa o tempo e  $x(t)$  representa a amplitude no instante de tempo  $t$ . A Fig. 5(a)

apresenta uma reprodução deste. Podemos representar sinais de áudio também no domínio da frequência. Um sinal contínuo no domínio da frequência é definido por  $X(\omega)$ , onde  $\omega = 2\pi f$  representa a frequência angular em radianos por segundo,  $f$  o número de vezes que a oscilação se repete, e  $X(\omega)$  representa a magnitude do sinal na frequência  $\omega$  [18]. A Fig. 5(b) representa a relação entre  $f$  e a magnitude de um sinal.

Figura 5 – Representação gráfica de sinais digitais



(a) Representação em tempo



(b) Representação em frequência

No processo de tratamento de sinais, uma ferramenta poderosa e de grande importância são os filtros, tanto os analógicos<sup>2</sup> como os digitais, sendo este último apresentado no tópico seguinte.

### 2.3.2 Filtros Digitais

O processo de filtragem altera o conteúdo de frequência de um sinal. Os filtros digitais são geralmente aplicados com dois propósitos: Separação de sinais ou Restauração de um sinal distorcido.

<sup>2</sup> Utiliza componentes analógicos físicos como resistores, indutores e capacitores para produzir o efeito de filtragem

Ao filtrar um sinal se presume que há uma parte das informações que queremos separar de outra parte. Seja separar o sinal do ruído, ou um sinal de outros sinais. Para isso, várias estratégias podem ser utilizadas, iremos focar em duas: Filtros passa-baixas (*Low Pass Filter - LPF*) e Filtros passa-altas (*High Pass Filter - HPF*).

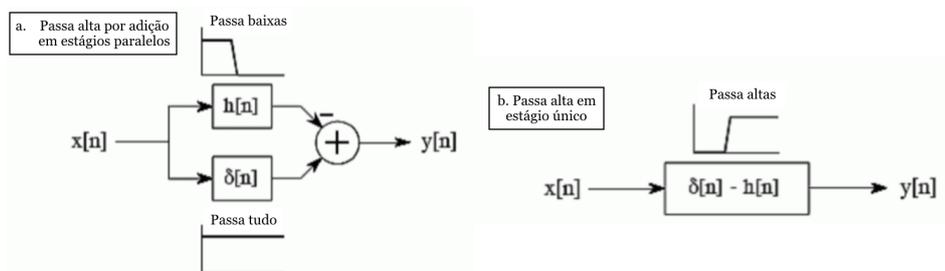
### Filtros passa-baixas

O LPF é o filtro “padrão” em processamento de sinais, com ele podemos dar origem a diversos outros tipos de filtros, tais como o passa-altas, passa-faixa e o rejeita-faixa. Este tipo de filtro tem por finalidade permitir a passagem de sinais com frequências abaixo de um limiar estabelecido, denominado de frequência de corte ( $\omega_c$ ), e atenuar<sup>3,4</sup> sinais com frequências acima desse limiar, fazendo com que seja gerado um sinal que evidencia as frequências abaixo de  $\omega_c$ .

### Filtros passa-altas

O filtro HPF é um filtro derivado do LPF, conforme pode ser visto na Fig. 6. Este filtro é complementar ao LPF, onde sua finalidade consiste em bloquear a passagem de sinais com frequências abaixo de um limiar estabelecido e permitir a passagem dos que possuem frequências acima desse limiar, por fim gerando um sinal que tem foco nas altas frequências (acima de  $\omega_c$ ).

Figura 6 – Diagrama de blocos de inversão espectral para geração de filtro passa-altas



Fonte: extraído de [19] - Adaptado.

### 2.3.3 NLD - *Nonlinear Device*

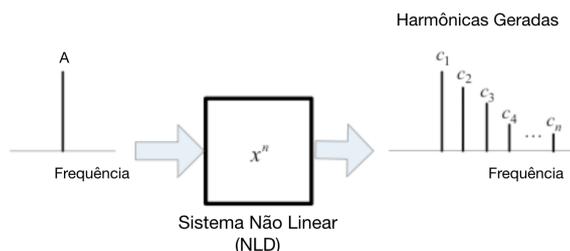
Os dispositivos não lineares (NLDs) são comumente usados para gerar harmônicos específicos no sistema de graves virtuais, eles ‘deslocam’ componentes do sinal de uma faixa de baixa frequência para uma faixa de frequência mais alta. A respeito do tom do sinal, ele é preservado uma vez que os componentes na faixa de frequência mais alta são harmônicos dos componentes originais [20].

<sup>3</sup> Na física, é a diminuição da intensidade de um sinal

<sup>4</sup> A quantidade de atenuação para cada frequência varia de filtro para filtro

Diversas são as funções disponíveis para a geração de harmônicas, desde as mais simples de implementação computacional e de fácil reprodução analógica até as que são mais complexas, solicitando um pouco mais de processamento digital e analógico. A Fig. 7 ilustra o objetivo da execução do NLD, criação de harmônicas do sinal de entrada utilizando uma função não linear.

Figura 7 – Ilustração do processo de execução do NLD



Fonte: extraído de [7] - Adaptado.

## 2.4 Como avaliar a qualidade do áudio

Inicialmente precisa-se compreender por que a avaliação da qualidade do áudio é tão importante. Quando se trata de realizar operações sobre áudios deve-se sempre observar o sinal gerado, como observado na seção anterior, existem certas faixas de frequência/intensidade que podem provocar dor ao ouvinte, em alguns casos podendo provocar danos físicos. Ao se realizar a avaliação da qualidade observa-se as características que definem a audibilidade do sinal, tais como amplitude, frequência, conforto na audição, perda de informação, dentre outras características.

Algumas perguntas que devem ser realizadas ao se analisar uma faixa de áudio são: O tom lhe agrada? Existem distorções no áudio que lhe causam desconforto? Considerando que já se conhece o áudio, nota-se ausência ou acréscimo de algo? Essas perguntas refletem uma **avaliação subjetiva** do áudio, onde a qualidade depende da interpretação e conforto do ouvinte. Por outro lado, pode-se utilizar cálculos matemáticos para analisar os aspectos físicos do áudio, como sua amplitude, frequências, dentre outros, obtendo uma **avaliação objetiva**.

Existem diversas formas de se realizar a avaliação de qualidade de um áudio, a *International Telecommunication Union* - I.T.U. - agência da ONU [21] especializada em tecnologias de informação e comunicação - possui recomendações quanto esse tipo de avaliação.

### 2.4.1 Avaliação Subjetiva

A avaliação subjetiva, como mencionado, depende da percepção de qualidade do ouvinte, logo, faz-se necessária a realização de testes com um conjunto de pessoas com características distintas, de modo a diversificar os casos de análise, neste caso propõe-se a seleção de voluntários e submissão a testes comparativos e/ou quantitativos da qualidade de áudio.

No aspecto apresentado a recomendação ITU BS.1534-1 [22] especifica como deve ser a escolha dos voluntários, os equipamentos, a escala de avaliação, o tratamento dos dados, dentre outras especificações.

### 2.4.2 Avaliação Objetiva

A avaliação objetiva utilizando-se de cálculos matemáticos resulta em um valor que atribui de acordo com a avaliação das características extraídas do áudio uma nota que representa sua ‘qualidade’. A I.T.U. através da recomendação ITU BS.1387-1 [23][24] recomenda a utilização de um sistema matemático que modela várias fases do ouvido humano, gerando variáveis de saída que podem caracterizar distorções originadas do sinal processado. Essas variáveis são denominadas de MOV's - *Model Output Variables* (Variáveis de saída do modelo).

A etapa final do modelo de medição combina os valores de MOV para produzir um único valor de saída que corresponde diretamente a um resultado esperado de uma avaliação de qualidade subjetiva, o ODG - *Objective Difference Grade* corresponde à qualidade percebida subjetivamente.

Quanto aos MOV's, eles compreendem um total de 14 variáveis que determinam características acerca do áudio processado ou da relação entre este e o áudio de referência, das quais, para a realização deste projeto serão analisadas 11, que trazem informações unicamente acerca do áudio processado, as três outras trazem informações conjuntas entre o áudio original e o processamento.

## 2.5 Regressão Linear

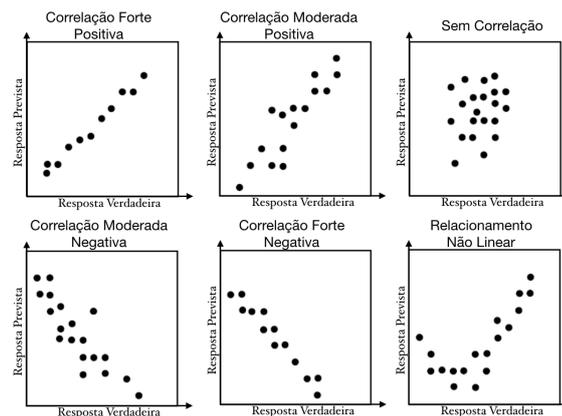
Regressão é uma técnica utilizada com o propósito de estimar uma variável de resposta a partir de outras variáveis específicas, conhecidas como variáveis preditoras, para tal ela gera uma equação matemática capaz de descrever a relação entre uma ou mais preditoras e a variável de resposta, além de prever novas observações.

Algumas ferramentas podem auxiliar no processo de determinação do sistema de regressão, tais como:

**Coefficiente de correlação:** É um número que fornece detalhamento sobre o tipo e a extensão do relacionamento entre variáveis, este coeficiente pode variar de -1.0 a +1.0, onde o sinal indica o tipo de relação entre as variáveis, diretamente proporcional(+) ou inversamente proporcional(-) e o valor numérico indica a extensão da relação, ou seja, quanto maior o valor, maior a relação entre as variáveis, assim o valor 0 (zero) indica ausência de relacionamento **linear** entre as variáveis. Logo, quanto mais próximo dos extremos, mais fácil se torna a especificação de uma equação linear que relacione as variáveis.

**Diagrama de dispersão:** Também denominado de gráfico de dispersão, é uma ferramenta de análise visual dos dados, que representa através de um gráfico cartesiano os valores de um conjunto de dados para duas ou mais variáveis. Pela análise do gráfico de dispersão também pode-se estimar o tipo de relacionamento dentre as variáveis, como pode ser visualizado na Fig. 8.

Figura 8 – Relacionamento entre variáveis vistos por gráficos de dispersão



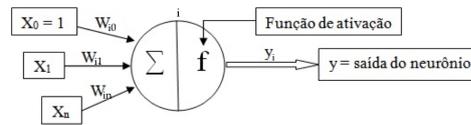
Fonte: extraído de [25] - Adaptado

### 2.5.1 Redes Neurais

As Redes Neurais Artificiais (RNAs) buscam simular o funcionamento do cérebro humano. São técnicas computacionais que se baseiam nas redes neurais biológicas, e são capazes de processar informações simultaneamente utilizando uma rede de neurônios artificiais interconectados, neurônios estes responsáveis pela não-linearidade e pela memória da rede [26].

A primeira proposta de um modelo matemático para representar o neurônio biológico foi proposto por McCulloch e Pitts em 1943, com a denominação de neurônio artificial (Fig. 9).

Figura 9 – Ilustração do modelo matemático criado por McCulloch e Pitts para representar o neurônio artificial



Fonte: extraído de [27] - Adaptado

O modelo consiste na representação do neurônio biológico de forma bastante simples, tomando como base a “lei de tudo ou nada” onde uma ação ocorre se um estímulo ocorrer com intensidade acima de um determinado limiar. Se o sinal tiver intensidade menor que o limiar, o impulso não é disparado. Caso o sinal tiver intensidade maior que o limiar, o impulso é disparado.

Pode-se notar na Fig. 9 que o valor de  $X_0$  é fixado em 1, pois ele representa o limiar excitatório do neurônio [28]. Vale salientar que o índice  $i$  é um identificador que indica o neurônio de análise.

Os neurônios artificiais são compostos por um conjunto de entradas ( $X_j$ ) conectadas através de ligações que possuem seus respectivos pesos ( $W_{ij}$  - onde  $i$  representa o neurônio e  $j$  cada entrada dele). Cada entrada do neurônio é multiplicada pelo seu peso e o somatório de todos esses produtos representa a entrada do neurônio artificial, de acordo com a equação

$$net_i = \sum_{j=0}^n W_{ij} \cdot x_j$$

O valor de  $net_i$  será utilizado para encontrar a saída do neurônio, que deverá ser calculada de acordo com a função de ativação utilizada.

As redes neurais são usadas para modelar relações complexas entre entradas e saídas ou para encontrar padrões em dados. Em muitas situações, a relação funcional entre as covariáveis (também conhecidas como variáveis de entrada ou variáveis independentes) e as variáveis de resposta (também conhecidas como variáveis de saída ou variáveis dependentes) é de grande interesse.

As redes neurais artificiais podem ser aplicadas como aproximação a qualquer relação funcional complexa. Ao contrário dos modelos lineares generalizados (GLM)<sup>5</sup>, não é necessário que o tipo de relação entre variáveis independentes e variáveis resposta seja, por exemplo, uma combinação linear. Isso faz das redes neurais artificiais uma valiosa

<sup>5</sup> Em estatística, GLM - *Generalized Linear Model* é uma generalização flexível da Regressão Linear comum que permite para as variáveis de saída assumam outros modelos de distribuição de erro além da distribuição normal.

ferramenta quantitativa. Esses modelos são, particularmente, extensões diretas dos modelos lineares generalizados e podem ser aplicados de maneira semelhante.

Dados observados são usados para treinar a rede neural, fazendo assim com que a rede neural “aprenda” uma aproximação da relação entre as variáveis independentes e dependentes de forma iterativa por meio da adaptação contínua dos seus parâmetros. Ou seja, os parâmetros do modelo são estimados por meio da amostra.

### Validação Cruzada

É uma técnica utilizada para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados, muito aplicada em problemas cujo objetivo da modelagem é a predição. Busca-se então estimar o quão preciso é este modelo na prática, ou seja, o seu desempenho para um novo conjunto de dados.

O conceito central das técnicas de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, e posteriormente, utiliza-se alguns destes subconjuntos para estimar os parâmetros do modelo (dados de treinamento) e o restante dos subconjuntos (dados de validação ou de teste) são empregados na validação do modelo.

Existem diversas formas de realizar o particionamento dos dados, sendo as três mais utilizadas: o método *holdout*, o *k-fold* e o *leave-one-out*.

O método ***holdout*** consiste em dividir o conjunto total de dados em dois subconjuntos mutuamente exclusivos, um para treinamento (estimação dos parâmetros) e outro para teste (validação). Esta abordagem é indicada quando se tem uma grande quantidade de dados. Caso o conjunto total de dados seja pequeno, o erro calculado na predição pode sofrer muita variação.

O método ***k-fold*** consiste em dividir o conjunto total de dados em  $k$  subconjuntos mutuamente exclusivos do mesmo tamanho e, a partir disto, um subconjunto é utilizado para teste e os  $k - 1$  restantes são utilizados para estimação dos parâmetros e calcula-se a precisão do modelo. Este processo é realizado  $k$  vezes alternando de forma circular o subconjunto de teste.

O método ***leave-one-out*** é um caso específico do *k-fold*, com  $k$  igual ao número total de dados ( $N$ ). Nesta abordagem são realizados  $N$  cálculos de erro, um para cada dado. Este método possui um alto custo computacional, sendo indicado para situações onde se tem poucos dados disponíveis.

## 3 Desenvolvimento

No decorrer deste capítulo é apresentado o modelo proposto para o desenvolvimento do sistema de processamento de áudio que proporciona a melhoria na reprodução em alto-falantes de filme fino, bem como a metodologia e os passos do desenvolvimento, desde o processamento dos áudios de teste<sup>1</sup> até a regressão que proporciona a autoanálise de qualidade do processamento, entre as fases também são apresentadas as ferramentas criadas, e os modelos de avaliação de qualidade utilizados.

### 3.1 Metodologia

O trabalho realizado visa o desenvolvimento ou aprimoramento de um método capaz de melhorar a reprodução de sons graves nos alto-falantes de material piezoelétrico, para tal, faz-se necessário o desenvolvimento computacional do sistema, realização de teste e ajustes para otimizar o resultado a fim de chegar a um resultado final “ótimo”, um sistema estável que atenda as necessidades especificadas.

No contexto explicitado acima, este trabalho apresenta as fases de desenvolvimento computacional do sistema, inicialmente em computadores pessoais objetivando a validação das técnicas desenvolvidas e/ou adaptadas para o melhoramento da reprodução, testes e resultados obtidos com um conjunto de áudios selecionados.

#### 3.1.1 Os áudios de teste

Para a realização do projeto foram selecionados 5 áudios com ocorrência de tons graves, com pouca ou nenhuma reprodução em alto-falantes de filme fino<sup>2</sup>, alguns deles apresentando apenas música instrumental, outros com partes faladas, diversificando as observações e validando a não distorção das altas frequências na tentativa de melhora das baixas. As músicas selecionadas são listadas abaixo, todas elas estão disponíveis em: <http://bit.ly/2neRSR6>

1. **Áudio 1** - denominada nos testes como ‘bad’
2. **Áudio 2** - denominada nos testes como ‘billie-jean’
3. **Áudio 3** - denominada nos testes como ‘led-zeppelin’
4. **Áudio 4** - denominada nos testes como ‘otherside’

---

<sup>1</sup> Áudios utilizados para treinamento do sistema

<sup>2</sup> Recomenda-se a reprodução através dos alto-falantes do computador e depois a reprodução utilizando fones de ouvido

5. **Áudio 5** - denominada nos testes como ‘the-chicken’

### 3.1.2 As técnicas de processamento

Consiste nas técnicas de processamento para geração de harmônicos para as frequências apresentadas pelo sinal de entrada. Para tal as técnicas fazem uso de funções não lineares, dentre outras funções ou processamentos como *Phase Vocoder*, FFT - *Fast Fourier Transform* e Método de Prony.

Dentre as funções presentes na literatura foram utilizadas as seguintes:

#### ATSR - Arc - Tangent Square Root

Uma das técnicas citadas na literatura que trás bons resultados no processo de reforço de frequências baixas é o ATSR, ele gera os harmônicos do sinal original por meio de uma função não-linear. Através dos parâmetros dessa função pode-se facilmente controlar esses harmônicos [29]. Essa função é descrita por

$$f(x) = \alpha \arctan(\beta x) + \psi \sqrt{1 - (\zeta x)^2} - \psi \quad ,$$

onde:

$\alpha$  - utilizado para o controle de intensidade das harmônicas ímpares;

$\beta$  - utilizado para o controle dos padrões harmônicos das harmônicas ímpares;

$\zeta$  - utilizado para o controle dos padrões harmônicos das harmônicas pares;

$\psi$  - utilizado para o controle de intensidade das harmônicas pares.

Sua implementação em *MATLAB*<sup>3</sup> pode ser vista no Anexo A.

#### FWI - Full Wave Integrator

A escolha dessa técnica é decorrente da simplicidade atrelada ao seu funcionamento, é realizada a integral de cada ciclo do valor absoluto do sinal original, de acordo com a equação

$$A = \int_a^b |f(x)| dx \quad ,$$

gerando os harmônicos do sinal original.

Sua implementação em *MATLAB* pode ser vista no Anexo B.

<sup>3</sup> *MATrix LABoratory* é, dentre outras classificações, uma ferramenta de programação proprietária desenvolvida pela *MathWorks*.

### FWR - Full Wave Rectifier

Essa técnica foi selecionada devido a sua simplicidade de implementação e o seu baixo custo computacional de processamento. O sinal processado é igual ao original, porém com todos os valores positivos.

Ela consiste é executada pela equação

$$f(x) = |x| \quad .$$

### SC - Soft Clipping

Semelhante ao ATSR, é uma técnica para geração de harmônicos que se baseia em uma função, porém, esta utiliza uma abordagem de mais simples implementação [7], conforme a descrita na equação

$$f(x) = \frac{x}{K|x| + 1} \quad ,$$

onde:

$K$  - Representa a curva (*Knee*) de distorção da NLD.

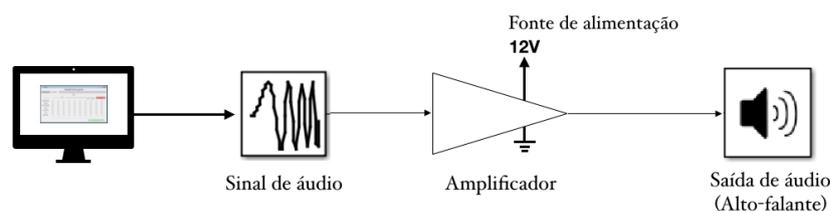
Sua implementação em *MATLAB* pode ser vista no Anexo C.

### 3.1.3 O ambiente de testes

Os testes subjetivos foram realizados em laboratório com ambiente fechado utilizando-se um computador para execução do sistema de avaliação e para a reprodução dos áudios, um sistema sonoro composto de alto-falante de filme fino, módulo amplificador de 2 canais e uma fonte de alimentação para este último.

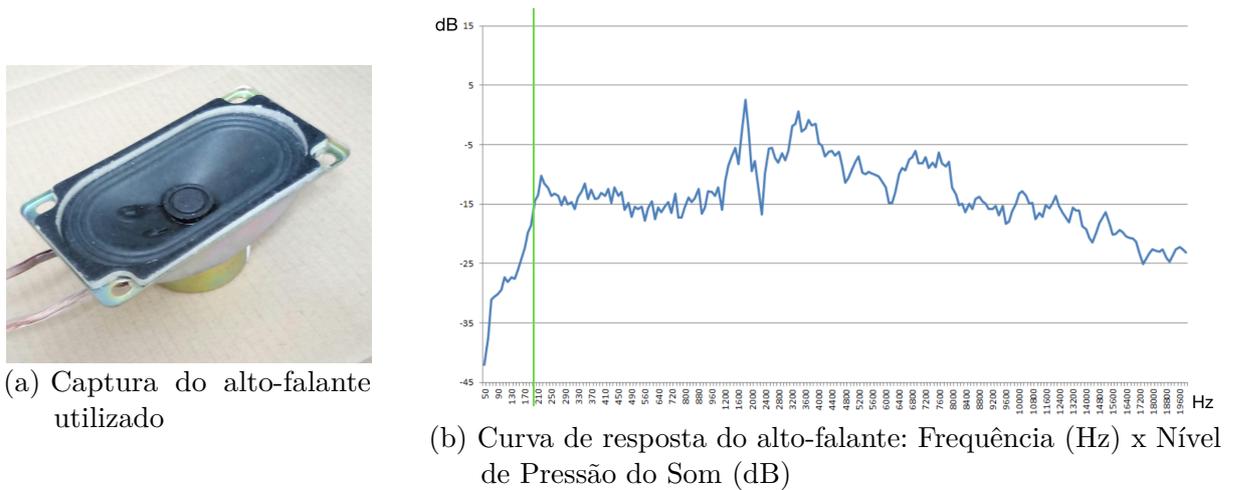
O diagrama de blocos apresentado pela Fig. 10 que exhibe o percurso do sinal de áudio desde a saída do computador onde é realizada a avaliação até o alto-falante. O alto-falante utilizado é apresentado a seguir pela Fig. 11 (a), alto-falante de  $8 \Omega$  e  $1 W$ , cuja curva de resposta às frequências é apresentada pela Fig. 11 (b), sua frequência de corte se encontra em torno dos 200Hz em destaque pelo traço que corta o gráfico.

Figura 10 – Diagrama do sistema de testes



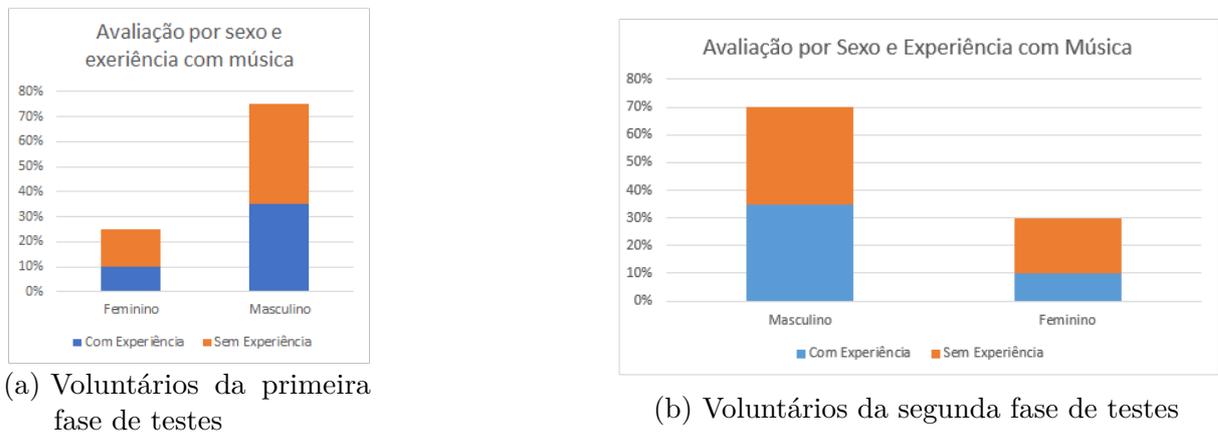
Fonte: elaboração própria.

Figura 11 – Alto-falante utilizado nos testes e sua curva de resposta



A faixa etária dos voluntários encontra-se entre 18 e 40 anos, com maior concentração na faixa de 19 a 24 anos. Em relação ao sexo e experiência com música, esses dados podem ser visualizados pelos gráficos exibidos na Figura 12, onde em (a) é exibida a análise estatística dos voluntários da primeira fase de testes, enquanto (b) para a segunda fase de testes.

Figura 12 – Análise estatística dos voluntários dos testes subjetivos



### 3.2 Modelo do sistema

Para a realização deste sistema, foi proposto um modelo de processamento baseado numa sequência de ações sobre o áudio original, a fim de se obter o áudio processado com a melhor configuração para reprodução.

Para o seu desenvolvimento propôs-se a divisão do projeto em três fases:

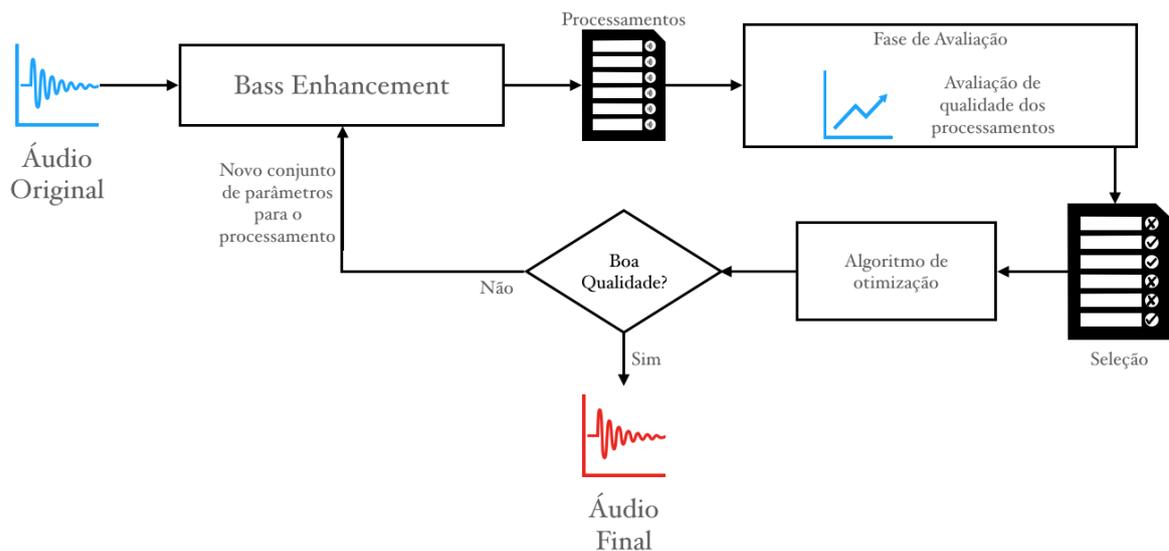
**Treinamento do sistema de autoanálise:** Esta fase consiste na etapa de treinamento do processo de autoanálise do processamento. Dado um conjunto de áudios para treinamento, faz-se necessária a realização de processamentos utilizando técnicas e parâmetros pré definidos. Em posse dos áudios processados, obtêm-se as avaliações objetiva e subjetiva destes, obtendo assim dados para o treinamento de um sistema de regressão, onde se possa definir a qualidade do processamento.

**Determinação do processo de otimização de parâmetros:** Considerando conjuntos de parâmetros utilizados em determinados processamentos e suas respectivas avaliações de qualidade, é possível realizar combinações a fim de determinar o melhor conjunto de parâmetros para o processamento de um determinado áudio. Neste caso necessita-se determinar o sistema de otimização que melhor se adequa a necessidade do projeto.

**Implementar fluxo do sistema:** Em posse dos resultados das etapas citadas acima, esta fase representa a finalização do projeto, contando com um sistema completo capaz de, a partir de um áudio de entrada, gerar processamentos e, através da análise de suas qualidades determinar a melhor combinação de parâmetros para o áudio, gerando um som com boa reprodução em alto-falantes que anteriormente não a possuía.

Uma representação do modelo proposto pode ser visualizada pela Fig. 13.

Figura 13 – Representação do sistema proposto



Fonte: redução do sistema proposto na Fig. 1.

Este modelo consiste basicamente da combinação das etapas propostas, dispostas em três blocos: **Bass Enhancement**, diz respeito ao processamento do áudio, gerando um novo sinal com a melhoria de reprodução nos tons graves; **Fase de Avaliação** é o estágio onde se utilizará do sistema regressor, proposto na etapa 1, para determinar a qualidade dos processamentos; por fim, **Algoritmo de otimização**, onde, como descrito para a etapa 2, será realizada a otimização dos parâmetros para determinar o melhor processamento para o sinal de áudio. O sistema finaliza a execução enviando ao alto-falante um áudio que passe pelo critério de aceitação determinado, cuja qualidade seja adequada para reprodução.

### 3.3 Bass Enhancement e Avaliação de Qualidade

O processamento do áudio gira em torno basicamente do NLD. Ele é o componente chave na maioria dos sistemas de processamento virtual de graves, tendo por objetivo gerar harmônicas do sinal que através do fenômeno de *missing fundamental* provocam a sensação de audição das baixas frequências.

Quanto aos filtros digitais, estes foram utilizados para separação das frequências dos áudios, conforme ilustrado nas Figuras 14 e 17. Utilizando-se as mesmas frequências de corte para todos os processamentos que utilizam os filtros, os valores são apresentados na Tabela 1.

Tabela 1 – Frequências de corte para filtros Passa baixa e Passa alta

	$\omega_{c1}$	$\omega_{c2}$
LPF	250	300
HPF	250	200

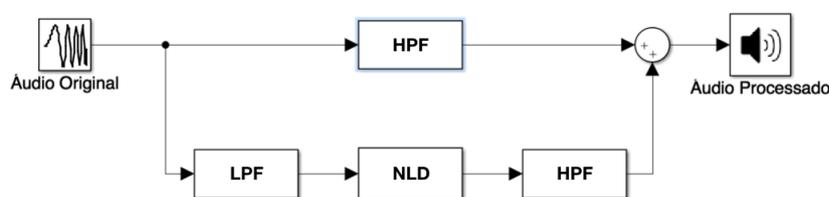
As frequências apresentadas indicam uma faixa de atenuação até a frequência de corte total, como filtros reais não apresentam uma variação abrupta, o que ocorre é uma atenuação gradativa até o corte total. As frequências correspondentes a coluna  $\omega_{c1}$  apresentam o início da atenuação, enquanto a coluna  $\omega_{c2}$  apresenta a frequência de corte total.

Durante a evolução do projeto ocorreram variações do fluxo de processamento com objetivo de se atingir melhorias no resultado final. São apresentados dois destes modelos definidos durante o desenvolvimento.

### 3.3.1 Modelo Inicial

O modelo apresentado na Fig. 14 foi o primeiro modelo de processamento desenvolvido, nele o áudio original é submetido a uma filtragem que retorna o sinal com as frequências mais baixas, este é processado por um NLD, um dispositivo de processamento responsável por gerar frequências altas. Deste sinal se obtêm as frequências mais altas e este é somado com o sinal de frequências mais altas do áudio original. Todas as frequências baixas do processamento são atenuadas a fim de evitar distorções provocadas pela soma com os resultados dos processamentos.

Figura 14 – Diagrama de blocos do processamento de melhoria inicial



Fonte: adaptado com base em sistema proposto em [29].

O NLD deve gerar um mapa de frequências com suas devidas intensidades e fases para que, através da adição destas ao sinal de áudio, a resposta audível percebida pela orelha humana contenha frequências graves, mesmo que elas não tenham sido geradas fisicamente pelo alto-falante.

Utilizando-se deste modelo, os áudios de teste foram submetidos às técnicas de processamento gerando, para cada áudio de entrada, um conjunto de nove áudios processados. Foram utilizadas variações dos parâmetros para os algoritmos ATSR e SC, somando um total de quatro processamentos distintos para o primeiro e três para o segundo. Como os algoritmos FWI e FWR não apresentam parâmetros foi gerado um processamento para cada.

Os processamentos gerados foram categorizados de acordo com a Tabela 2, onde apresentam-se a nomenclatura dada ao processamento, o algoritmo utilizado e seus respectivos parâmetros utilizados. Os valores foram escolhidos aleatoriamente em uma faixa próxima aos valores apresentados na literatura, assim seria possível analisar variações do processamento sem se distanciar dos fatores conhecidos.

Tabela 2 – Nomenclatura e parâmetros dos processamentos utilizando o modelo inicial

Nomenclatura do Processamento	Técnica	$\alpha$	$\beta$	$\zeta$	$\psi$	$K$
g11504_ATSR1	ATSR	1.1	0.5	0.5	0.4	-
g12509_ATSR2	ATSR	1.2	0.5	0.9	2.0	-
g32509_ATSR3	ATSR	3.0	1.5	1.5	3.0	-
g051504_ATSR4	ATSR	2.5	0.9	0.9	2.5	-
g025_FWI3	FWI	-	-	-	-	-
g05_FWR2	FWR	-	-	-	-	-
g0515_SC2	SC	-	-	-	-	1.3
g0525_SC3	SC	-	-	-	-	4.2
g02525_SC4	SC	-	-	-	-	2.5

Os áudios gerados para as cinco músicas de testes podem ser encontrados através do link: <http://bit.ly/2AUb2BX>, eles estão distribuídos em pastas para cada áudio de análise. O nome dos áudios apresentam a nomenclatura do processamento correspondente e a palavra “final” separados por um ‘  ’ (*underline*), uma solicitação da ferramenta de avaliação utilizada.

### Ferramenta de Avaliação

Uma vez que temos um conjunto de áudios, supostamente, com uma reprodução de graves melhor do que o conjunto de áudios originais, necessita-se comprovar esta melhoria. Neste cenário foi desenvolvida uma ferramenta para avaliação de qualidade através da contribuição de voluntários, os quais realizariam a análise dos áudios atribuindo a eles uma nota que reflete a melhora ou piora do áudio processado em relação ao áudio original, apresentado como “Áudio de Referência”. A fim de analisar a distribuição de notas dos voluntários foi introduzido o áudio original dentre os áudios para análise, tendo então para cada faixa de áudio: 1 faixa de referência, 9 processamentos e uma faixa exatamente igual à referência, inserida sem o conhecimento do avaliador. Assim, é possível verificar o desvio na atribuição das notas.

A ferramenta desenvolvida em MATLAB é exibida na Fig. 15 ela apresenta, para cada fase de avaliação, um botão “REF.” o qual ativa a reprodução do áudio de referência (áudio original), e para cada processamento mais o original é ativado uma barra de avaliação e o botão correspondente (de A até J), distribuídos randomicamente para evitar que o mesmo processamento sempre caia na mesma posição pra todos os áudios e ocorra um direcionamento da avaliação. A avaliação é realizada em uma escala de -100 a 100, indicando uma variação de “Muito Pior” a “Muito Melhor” em relação ao áudio de referência.

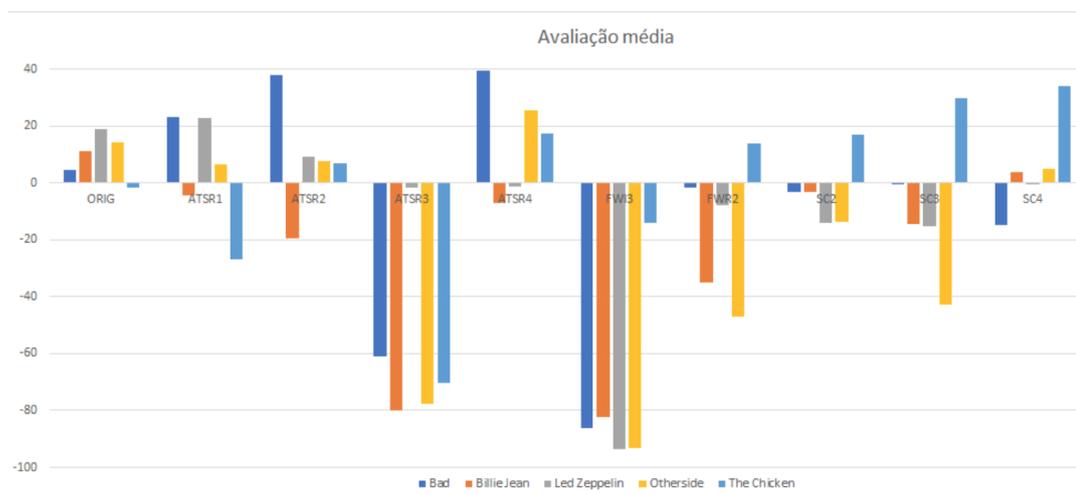
Figura 15 – Ferramenta de avaliação de qualidade



Fonte: elaboração própria.

Nesta fase inicial, após análise dos resultados obtidos (representados graficamente por meio da Fig. 16) foi possível concluir que os algoritmos FWI e FWR não atendem às necessidades do sistema, apresentando resultados negativos a grande maioria dos casos. Em comparação com o SC, o FWR apresenta resultados semelhantes, porém pela possibilidade de alterar o seu parâmetro  $K$  (Knee) o SC ainda se mostra viável, sendo necessária apenas a realização de análises e testes com variações deste parâmetro para uma confirmação de sua eficácia. Para análise completa dos dados brutos, eles podem ser encontrados através do link: <http://bit.ly/2BJPODZ>.

Figura 16 – Gráfico de análise da média das avaliações utilizando o modelo inicial



Fonte: elaboração própria com base nas avaliações coletadas.

Em uma análise mais detalhada dos resultados apresentados no gráfico, podemos notar que as configurações do ATSR apresentaram valores, na maioria dos casos, acima

de 0, ou seja, julgados com reprodução melhor, com exceção da configuração ATSR3 que apresentou resultados ruins, entre -60 e -80. Para os processamentos com FWI e FWR, o primeiro apresentou resultados negativos, não se mostrando favorável em nenhum dos caso de análise, já o FWR mostrou-se desfavorável por não provocar mudanças significativas na reprodução, sendo isto verificado pelas avaliações, que se encontram em torno do valor 0. Quanto ao SC, observa-se que, para as três configurações utilizadas, ele se mostra variável, não conseguindo determinar sua eficácia, uma vez que, para a mesma configuração, apresenta bons resultados em alguns casos e outros indiferentes ou ruins. No entanto, com ajustes no parâmetro do SC é possível que se obtenha melhoria na reprodução.

Através da análise realizada, constata-se que o ATSR apresenta bons resultados, e visto que este algoritmo apresenta a maior possibilidade de combinações para otimização através de seus parâmetros, acredita-se que ele seja o mais promissor para o sistema proposto.

### 3.3.2 Modelo Adaptado

Durante a avaliação de qualidade realizada pelos voluntários, descrita na seção anterior, algumas constatações e recomendações foram realizadas, dentre as quais, as mais mencionadas foram:

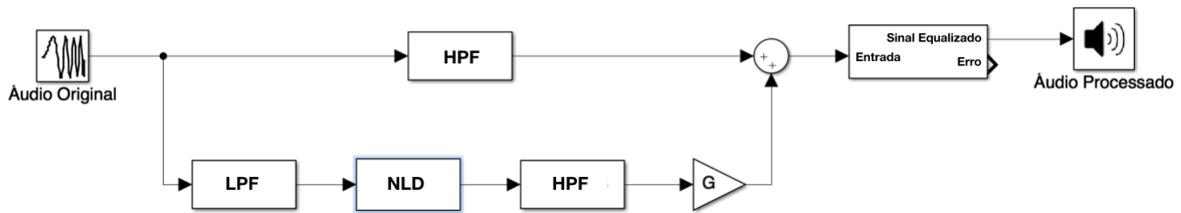
- Alguns áudios apresentavam um volume mais alto que os demais, dificultando a avaliação, pois não se sabia se houve melhora ocasionada pelo processamento ou pelo aumento de volume.
- Alguns áudios são ceifados, causando muita distorção. Isto pode tanto ter sido provocado pelo processamento como pelo aumento de volume citado no item anterior.
- Para alguns áudios haviam partes específicas da faixa que se gostaria de analisar, sendo assim recomendou-se a introdução de um seletor de posicionamento da reprodução, possibilitando o avanço ou volta para um instante de interesse da reprodução.

Em relação aos dois primeiros pontos apresentados, a variação de volume em processamento de áudios é um problema comum com reprodução MP3, com base nesta informação e utilizando da solução aplicada para este cenário, sugere-se a realização da equalização do áudio antes de sua reprodução, além de realizar um ajuste de amplitude do sinal de saída através de um controle de ganho no processamento das baixas frequências. Os ajustes ora propostos provocariam a resolução dos problemas citados.

No que diz respeito a recomendação apresentada no terceiro ponto foram realizados melhoramentos na ferramenta, gerando uma nova versão, denominaremos de versão 2.0, que passou a contar com o seletor recomendado, além de outras melhorias que estão apresentadas mais a frente.

Os ajustes realizados no processamento a fim de sanar os problemas apresentados levaram ao desenvolvimento de um novo modelo, conforme apresentado na Fig. 17. Contando com a introdução da equalização e do ajuste de ganho, esta versão trouxe como resultados áudios mais uniformes no que diz respeito a sua saída, solucionando os problemas apresentados pelos voluntários. Logo, os casos que apresentarem distorção na reprodução são resultantes do processamento, evitando assim dubialidade na avaliação.

Figura 17 – Diagrama de blocos do processamento de melhoria adaptado



Fonte: adaptado com base em sistema proposto em [7].

Assim como apresentado na seção anterior para o modelo inicial, os áudios de teste foram submetidos ao processamento especificado por este modelo adaptado. Para tal se utilizou apenas o algoritmo ATSR, já que, conforme mencionado, ele apresentou melhores resultados e mostra-se mais promissor no desenvolvimento do sistema. Como se passou a utilizar apenas um algoritmo de processamento, fez-se necessária a redefinição de parâmetros. A Tabela 3 apresenta o nome dado aos processamentos e os parâmetros utilizados. Como se pode observar ela inclui um parâmetro  $G$  que indica o Ganho, ele determina o ajuste na amplitude do sinal processado que será somado ao áudio original.

Tabela 3 – Nomenclatura e parâmetros dos processamentos utilizando o modelo adaptado

Nomenclatura do Processamento	$G$	$\alpha$	$\beta$	$\zeta$	$\psi$
ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	1.5	1.5	0.45	0.45	1.5
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	2.0	1.5	0.45	0.45	1.5
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	2.5	1.5	0.45	0.45	1.5
ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	1.5	2.5	0.9	0.9	2.5
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	2.0	2.5	0.9	0.9	2.5
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	2.5	2.5	0.9	0.9	2.5
ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	1.5	3.5	1.8	1.8	3.5
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	2.0	3.5	1.8	1.8	3.5
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	2.5	3.5	1.8	1.8	3.5

Os áudios gerados para este conjunto de treinamento estão disponíveis através do link: <http://bit.ly/2jK3uq1>, assim como os anteriores, eles estão distribuídos em

pastas para cada áudio de análise. O nome dos áudios apresentam uma referência ao áudio original e a nomenclatura do processamento correspondente.

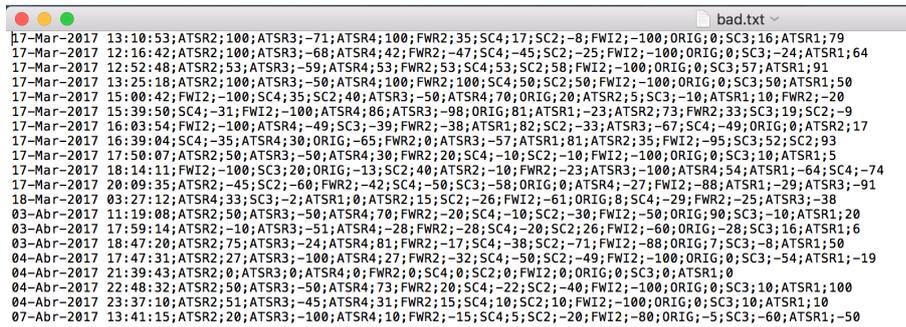
Como já foi mencionado a ferramenta de avaliação sofreu alterações, que provocaram a melhoria de usabilidade e geração de resultados, dentre as alterações realizadas, podemos apontar:

1. Inclusão do seletor de posição para reprodução, possibilitando voltar e avançar a pontos específicos da faixa em reprodução.
2. União do registro de voluntário na interface de avaliação, facilitando posterior identificação durante análise.
3. Alteração na métrica de avaliação, mantendo a ideia de uma avaliação na escala de “Muito Pior” a “Muito Melhor”, porém, em análise numérica, a escala passa a assumir valores entre  $-3$  e  $3$  reduzindo a margem de variação da nota, fixando ao valor  $0$  a igualdade entre o áudio avaliado e a referência, no caso o processamento ser considerado “Muito Pior” que a referência, a nota é mapeada para  $-3$ , caso o oposto ocorra, sendo considerado “Muito Melhor” então a nota é mapeada para  $3$ .
4. Armazenamento dos resultados das avaliações em planilha. Anteriormente os resultados eram armazenados em arquivos de texto separados, um arquivo para cada um dos cinco áudios, na nova versão, os resultados são unificados em uma planilha contendo uma folha para cada um dos cinco áudios. A diferença no armazenamento das avaliações pode ser visualizado na Fig. 18, onde em (a) é mostrado como era o armazenamento para a primeira versão da ferramenta e em (b) para a segunda versão<sup>4</sup>.

---

<sup>4</sup> Apesar de realizar o armazenamento em planilha, as avaliações também são salvas no formato anterior, sendo gerados os arquivos de texto com as avaliações para cada áudio.

Figura 18 – Armazenamento das avaliações dos voluntários



(a) Armazenamento na versão 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		V1804	V2011	V2052	V2790	V2842	V2016	V2317	V3060	V3814	V1609	V1368	V2163	V1502	V1351	V2099	V1750	V1357	V1710	V1816	V1561
2	ORIG	1	0	1	2	-1	1	2	-2	1	0	0	0	1	0	1	-1	0	1	1	0
3	ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	1	1	2	2	-1	1	1	-2	1	-1	0	-1	1	1	1	1	0	0	2	0
4	ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	2	1	0	3	-1	2	-1	-1	1	2	0	-1	-1	2	1	-1	1	2	1	1
5	ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	-1	2	-1	2	-2	-1	-2	-1	-1	1	1	1	-1	1	-1	-1	-1	-1	-1	-1
6	ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	1	1	1	1	1	0	1	-2	1	1	0	-1	0	0	2	1	1	-1	1	0
7	ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	0	2	3	-2	1	0	1	1	2	0	2	2	1	1	1	1	-1	1	0	0
8	ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	-1	-1	-1	-3	-2	-2	-3	-2	-2	-1	1	-2	1	-3	-2	-1	-2	-1	-2	-1
9	ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	1	0	0	-1	0	0	1	0	1	0	-1	2	0	0	2	1	1	0	0	-1
10	ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	1	3	1	2	0	0	1	-1	5	0	1	-2	1	1	-1	1	1	0	5	1
11	ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	-2	-1	-2	-1	-3	-2	-3	-3	-2	-1	-1	-1	-2	-1	-3	-2	-2	-2	-2	-3

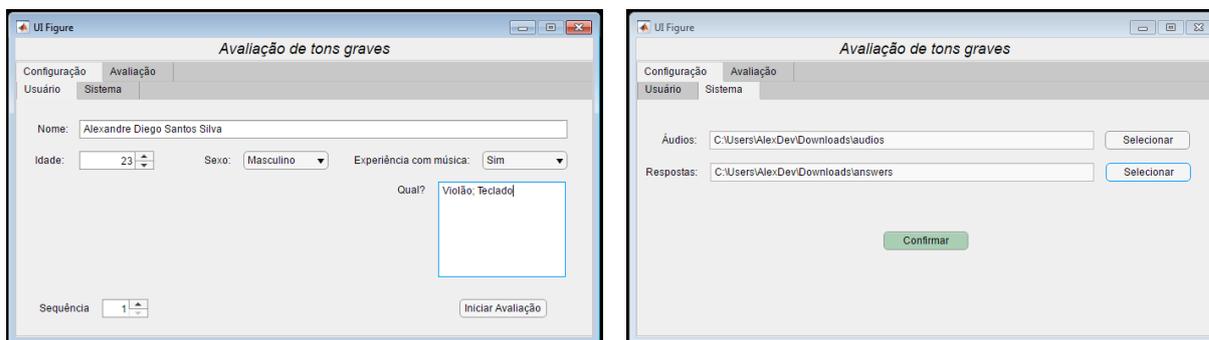
(b) Armazenamento na versão 2

Fonte: arquivos gerados por ferramenta própria.

Nesta nova versão, a tela de avaliações só é liberada após o cadastro do voluntário e a configuração dos caminhos de entrada e saída para o sistema de avaliação. Quanto ao caminho de entrada: é o caminho que indica a localização dos áudios, ele deve indicar a pasta onde estão os áudios originais, nessa mesma pasta deve haver pastas nomeadas com o mesmo nome dos áudios originais contendo os áudios processados. Quanto ao caminho de saída: é qualquer localização onde se deseja armazenar os resultados da avaliação.

Pela Fig. 19(a) vê-se a tela de registro do voluntário. Os dados coletados podem ser utilizados posteriormente para análise estatística dos resultados obtidos, por exemplo como idade, sexo e experiência com música se relacionam com os resultados obtidos; um campo muito importante nessa tela é o campo identificado como **Sequência** no canto inferior esquerdo, ele determina a coluna de armazenamento das avaliações deste voluntário, caso seja utilizado o mesmo número de um voluntário anterior, os resultados deste último serão sobrescritos pelos novos valores atribuídos pelo novo voluntário. A Fig. 19(b) exibe a tela de configuração para a avaliação, onde se determina a localização dos áudios para análise e onde será salva a avaliação.

Figura 19 – Interface de registro de voluntário e configuração de diretórios na segunda versão da ferramenta de avaliação



(a) Tela de registro de voluntário

(b) Tela de configuração de diretórios

Fonte: elaboração própria.

Ao configurar os caminhos de entrada e saída ocorre o redirecionamento automático para a tela de avaliações. A Fig. 20 apresenta uma representação da tela de avaliação durante a reprodução do áudio de referência. Na parte superior é exibido o nome do áudio que está sendo analisado (O nome provém do arquivo original). Pouco mais abaixo a barra de execução que indica o tempo de execução e o estágio atual da reprodução do áudio e ao clicar em sua extensão o áudio segue a execução a partir daquele ponto, ao lado esquerdo dessa barra é exibido o nome do áudio sendo executado (Referência ou Processamentos).

Figura 20 – Ferramenta de avaliação de qualidade versão 2.0



Fonte: elaboração própria.

Como se vê, as avaliações já não exibem o valor numérico correspondente para a nota uma vez que a escala corresponde exatamente com as indicações na lateral esquerda. Este esquema de avaliação remove a indecisão e a avaliação de “meio termo”, pois leva o voluntário a uma avaliação mais fiel, onde tem-se uma análise maior dos detalhes com objetivo de uma classificação adequada.

Os resultados das avaliações subjetivas dessa segunda fase (incluindo arquivos de texto e planilha) podem ser obtidos através do link: <http://bit.ly/2A30K24>, a Fig. 21 e a posterior lista apresentam a estrutura da planilha de resultados.

Figura 21 – Detalhamento da planilha de avaliações

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		V1804	V2011	V2052	V2790	V2842	V2016	V2317	V3060	V3814	V1609	V1368	V2163	V1502	V1351	V209
2	ORIG	1	0	1	2	-1	1	2	-2	1	0	0	0	1	0	1
3	ATSR_A1X5_BOX45_Q0X45_P1X5_G1X5	1	1	2	2	-1	-1	0	-2	1	-1	0	-1	1	1	1
4	ATSR_A2X5_BOX9_Q0X9_P2X5_G1X5	2	1	0	3	-1	2	-1	1	2	0	0	-1	-1	2	1
5	ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	-1	2	-1	2	-2	-1	-2	-1	-1	1	1	1	-1	1	-1
6	ATSR_A1X5_BOX45_Q0X45_P1X5_G2X0	1	1	1	1	0	1	1	1	1	0	-1	0	0	0	0
7	ATSR_A2X5_BOX9_Q0X9_P2X5_G2X0	0	2	2	3	-2	1	0	1	1	2	0	2	2	2	1
8	ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	-1	-1	-1	-1	-3	-2	-2	-3	-2	-2	-1	1	-2	1	-3
9	ATSR_A1X5_BOX45_Q0X45_P1X5_G2X5	1	0	0	0	-1	0	0	1	0	1	0	-1	2	0	0
10	ATSR_A2X5_BOX9_Q0X9_P2X5_G2X5	1	3	1	2	-2	0	0	1	-1	3	0	1	2	1	1
11	ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	-2	-1	-2	-1	-3	-2	-3	-3	-2	-1	-1	-1	-2	-1	-3
12																
13																
14																
15																
16																
17																
18																

Fonte: arquivo gerado por ferramenta própria - Captura adaptada.

1. Folhas de resposta, cada folha apresenta as avaliações dos processamentos do áudio que nomeia a folha.
2. Cada linha da folha, a partir da linha 2, indica as avaliações de um processamento específico, esse processamento é identificado pela primeira célula da linha.
3. Cada coluna da folha, a partir da coluna B, indica as avaliações de um voluntário para os processamentos da música da folha, o voluntário possui um código identificador gerado pela ferramenta de avaliação, esse código é encontrado nas outras folhas também, na respectiva coluna de avaliações daquele voluntário.
4. As avaliações são valores numéricos que variam de -3 a 3, e são atribuídos na folha de acordo com o voluntário avaliador e o processamento avaliado, na respectiva coluna-linha.

Tendo obtido as avaliações subjetivas, partiu-se para a realização da análise objetiva. Utilizando do sistema de geração de MOV's proposto pela I.T.U.. Submetendo para análise cada um dos áudios processados junto com o áudio original. Isto nos gerou para cada um dos 45 áudios processados um conjunto de variáveis e uma nota de qualidade proposta pelo algoritmo. Para nossa análise esta nota não traz acréscimo de informação, logo pode ser desconsiderada. Os resultados obtidos pela extração das variáveis de cada áudio processado foram agrupados em uma planilha que pode ser obtida através do link: <http://bit.ly/2zKBMAG>. A Tabela 4 apresenta a transcrição dos resultados obtidos.

Tabela 4 – Resultados obtidos para avaliação objetiva dos áudios processados

Nome do áudio	BandwidthRefB	BandwidthTestB	Total NMRB	WinModDiff1B	ADBB	EHSB	AvgModDiff1B	AvgModDiff2B	RmsNoiseLoudB	MFPDB	RelDistFramesB
BAD_ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	419.25	417.86	10.4786	91.1284	2.85089	1.4679	73.2305	168.525	11.3726	1	1
BAD_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	419.233	417.863	11.0181	91.1801	2.85676	1.46807	73.1507	169.219	11.637	1	1
BAD_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	419.196	417.813	11.6312	91.3041	2.86546	1.46827	73.1376	169.996	12.0514	1	1
BAD_ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	419.251	417.896	14.8815	92.4045	2.91536	1.46185	73.3352	173.859	14.3919	1	1
BAD_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	419.135	417.801	16.7921	93.3086	2.94495	1.4705	73.5271	176.249	14.7338	1	1
BAD_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	418.657	417.55	18.4303	94.7676	2.97191	1.6631	74.5296	183.921	14.759	1	1
BAD_ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	416.212	415.716	21.5727	98.977	3.05447	2.93039	76.3141	197.141	14.2803	1	1
BAD_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	415.057	414.807	22.6123	99.9109	3.11564	3.70119	75.6759	199.078	13.6501	1	1
BAD_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	414.308	414.037	23.2501	99.0616	3.16585	4.35553	74.9176	194.028	13.0058	1	1
BILLIE_ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	423.092	419.883	29.229	85.7455	3.07038	0.697176	82.4802	144.361	19.0749	1	1
BILLIE_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	423.062	419.869	29.2318	85.3386	3.07449	0.732823	82.3293	143.864	19.3364	1	1
BILLIE_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	422.811	419.684	29.2355	85.1703	3.08124	0.730993	82.3633	143.99	19.5124	1	1
BILLIE_ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	421.249	419.789	29.2716	89.1469	3.13529	2.5188	83.7379	147.224	17.9157	1	1
BILLIE_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	420.146	419.377	29.3177	92.8815	3.17785	3.69277	84.4667	148.913	17.3407	1	1
BILLIE_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	419.006	418.589	29.3854	94.9283	3.2156	4.59283	84.7023	148.068	17.342	1	1
BILLIE_ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	416.8	416.463	29.6976	99.4985	3.31622	7.04539	86.275	148.526	19.3235	1	1
BILLIE_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	416.002	415.582	30.0307	98.749	3.36771	8.23605	84.8607	145.825	21.3471	1	1
BILLIE_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	414.958	414.515	30.4309	98.4435	3.40764	9.05811	83.8459	145.334	22.8364	1	1
LED_ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	701.649	701.649	-11.8029	4.02836	1.67002	0.12178	3.97007	5.97708	0.387497	1	0.764327
LED_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	701.518	701.518	-9.09287	4.32762	1.87077	0.12758	4.31814	6.06961	0.501593	1	0.926316
LED_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	701.702	701.702	-6.80523	4.46765	2.02725	0.128412	4.50685	5.87995	0.607417	1	0.969591
LED_ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	701.386	701.386	0.770866	5.67557	2.43585	0.147113	5.70505	6.14467	1.58907	1	0.997661
LED_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	699.698	699.692	3.8414	8.7727	2.58622	0.215904	7.99656	9.79632	2.49173	1	0.997661
LED_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	696.711	696.697	6.43151	13.9998	2.74722	0.485987	12.9131	18.2133	3.22812	1	0.997661
LED_ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	687.159	686.994	13.1819	28.8925	3.09784	3.03863	28.2172	45.7887	5.5019	1	0.99883
LED_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	682.194	681.77	16.4746	33.7002	3.21627	4.63651	33.5786	56.6727	6.63552	1	0.99883
LED_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	678.518	677.191	18.926	36.8715	3.29007	5.85731	36.8389	63.1762	7.16679	1	1
OTHERSIDE_ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	748.59	748.363	-12.2675	3.64641	1.59904	0.0494001	3.46067	6.63602	0.28174	1	0.760768
OTHERSIDE_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	748.06	748.057	-9.56304	3.89372	1.80411	0.0451293	3.7703	7.05172	0.412857	1	0.864959
OTHERSIDE_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	748.003	748.002	-7.17504	4.09213	1.96597	0.0397814	4.00588	7.43865	0.56071	1	0.968568
OTHERSIDE_ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	747.83	747.83	0.489305	5.75699	2.39165	0.028384	5.60523	10.7862	1.48259	1	0.994761
OTHERSIDE_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	741.374	741.368	3.57514	12.5178	2.55565	0.145315	11.0114	26.8331	1.95403	1	0.995343
OTHERSIDE_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	727.301	727.284	6.23155	25.3319	2.7255	0.408128	21.5599	63.0941	2.22617	1	0.997672
OTHERSIDE_ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	665.561	665.225	16.5041	54.9141	3.14779	0.981573	42.775	159.006	3.94108	1	0.998836
OTHERSIDE_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	624.205	623.568	21.2031	51.061	3.28644	1.13006	42.8837	162.382	4.51691	1	1
OTHERSIDE_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	600.845	599.828	24.0498	48.9856	3.36624	1.19082	44.3506	172.09	4.78972	1	1
CHICKEN_ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	762.249	762.159	-10.9359	3.27459	1.53623	0.0314892	3.22456	3.08748	0.500984	1	0.749129
CHICKEN_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	762.236	762.149	-8.14667	3.57771	1.72434	0.0322825	3.5535	3.43625	0.785323	1	0.872242
CHICKEN_ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	762.194	762.105	-5.87714	3.77679	1.86108	0.0268589	3.771	3.67736	1.09404	1	0.933798
CHICKEN_ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	761.881	761.881	1.27128	4.53123	2.21592	0.0317633	4.53421	4.6716	2.51564	1	0.99187
CHICKEN_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	761.885	761.884	4.15309	4.96826	2.33737	0.0384959	4.97081	5.28321	3.23597	1	0.996516
CHICKEN_ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	762.098	762.098	6.33788	5.40354	2.42505	0.0439441	5.39036	5.94869	3.75347	1	0.997677
CHICKEN_ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	761.673	761.667	11.241	8.40328	2.61763	0.0811713	7.56414	9.61182	4.70767	1	1
CHICKEN_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	760.72	760.71	13.8682	13.4849	2.72526	0.150256	10.4491	15.1428	5.04408	1	1
CHICKEN_ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	758.569	758.516	15.8465	19.0847	2.81343	0.275172	14.2148	22.711	5.12384	1	1

Em posse das avaliações de qualidade subjetivas e objetivas, seguimos para a análise conjunta desses dados para determinar relações entre eles e gerar um sistema para estimar a qualidade subjetiva a partir da objetiva.

Todas as 11 variáveis de saída do modelo (MOVs) usadas para o desenvolvimento deste projeto são nomeadas e descritas pela Tabela 5.

Tabela 5 – Relação e descrição das MOVs utilizadas

Sigla de referência	Descrição
$BandwidthRef_B$	Largura de banda do sinal de referência
$BandwidthTest_B$	Largura de banda do sinal de teste
$TotalNMR_B$	Relação ruído-máscara
$WinModDiff1_B$	Diferença de modulação <i>windowed</i>
$ADB_B$	Distorção média do bloco
$EHS_B$	Estrutura harmônica do erro
$AvgModDiff1_B$	Diferença média de modulação
$AvgModDiff2_B$	Diferença média de modulação
$RmsNoiseLoud_B$	Distorção de volume
$MFPD_B$	Probabilidade filtrada máxima de detecção
$RelDistFrames_B$	<i>Frames</i> relativamente perturbados

## 4 Testes e Resultados

Este capítulo aborda os resultados obtidos com as avaliações realizadas, bem como o tratamento dos dados brutos, tanto para as avaliações subjetivas como para as objetivas. Os resultados foram analisados a fim de se especificar a regressão do sistema, logo são apresentadas as técnicas de tratamento de dados realizadas, os testes de regressão aplicados e os resultados atingidos.

### Avaliações e Tratamento de Dados

Durante o desenvolvimento do projeto foram obtidos diversos processamentos, e para a análise destes foram realizados testes internos, onde se verificou a qualidade de processamento e a complexidade do processo desenvolvido. Assim, como foi citado no capítulo anterior, também foram realizados testes com voluntários, os quais proporcionaram a coleta de avaliações subjetivas<sup>1</sup> para os processamentos.

No que diz respeito ao tratamento dos dados, foram aplicadas algumas técnicas com objetivo de normalizá-los, minimizando problemas oriundos do uso de variáveis com unidades e dispersões distintas. Para tal foi utilizada a Eq. 4.1, que é uma equação básica de normalização que se baseia na amplitude. Ela retorna um valor na escala de 0 a 1 proporcional a escala de  $X_{min}$  a  $X_{max}$ , onde  $X_{min}$  é o menor valor assumido pela variável de análise e  $X_{max}$  é o maior.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.1)$$

Algumas técnicas de pré-processamento foram utilizadas especificamente para as avaliações subjetivas, elas são melhor detalhadas e justificadas na próxima seção.

### 4.1 Avaliações Subjetivas

Quando se trata de análise subjetiva, sempre existe a possibilidade de encontrar disparidade nos dados. Na análise de qualidade de um áudio, apesar de serem especificados os pontos a serem analisados, sempre haverá algum caso de divergência que pode ocasionar a inviabilidade dos dados para análise. Por exemplo, viu-se que em alguns casos os voluntários avaliaram o áudio original (inserido dentre os processados) como sendo diferente do áudio de referência. Para solucionar esses casos existem técnicas de pré-processamento que visam

<sup>1</sup> Os resultados das avaliações subjetivas e objetivas foram disponibilizados na seção 3.3.2, sendo aqui apenas referenciados.

o ajuste do conjunto de dados. Para solucionar o caso apresentado foram propostas algumas soluções para o pré-processamento das notas das avaliações subjetivas:

1. Ajuste das avaliações para os voluntários que apresentarem discrepância na nota do áudio original através do deslocamento da nota do áudio original para zero e para as demais realizando a soma com o inverso da nota atribuída ao áudio original ( $ORIG_{old}$ ). Podemos descrever esse método pelo sistema de equações 4.2, onde  $X$  indica a nota atribuída para o processamento e  $X'$  a nota após o ajuste. As duas primeiras equações do sistema são limitadores, evitando a fuga da escala de -3 a 3.

$$X' = \begin{cases} 3 & , \text{ se } X - ORIG_{old} > 3 \\ -3 & , \text{ se } X - ORIG_{old} < -3 . \\ X - ORIG_{old} & , \text{ c.c.} \end{cases} \quad (4.2)$$

2. Ajuste das avaliações utilizando uma “normalização” para os valores esperados, redistribuindo as notas e escalando-as entre -3 e 3, utilizando-se de uma adaptação da Eq. 4.1, conforme apresentado pela Eq. 4.3

$$X' = -3 + \frac{(X - X_{min}) \cdot 6}{X_{max} - X_{min}} \quad (4.3)$$

3. Remoção de dados “aberrantes”, consiste na eliminação de dados que fogem ao padrão do conjunto de dados. Existem diversas técnicas para encontrar e tratar dados fora do padrão, a título de execução utilizamos a seguinte técnica: Para cada processamento calculou-se a média ( $\bar{X}$ ) e o desvio padrão ( $\delta$ ) das avaliações dadas. São considerados como dados aberrantes aqueles que não se enquadrem na escala  $[\bar{X} - \delta \quad \bar{X} + \delta]$ .

Numa primeira fase de análise foi utilizada a proposta 1, realizando o reajuste das avaliações. Tomando como exemplo o conjunto das avaliações do áudio 1 - “bad” (**P0**), a Fig. 22 apresenta em (a) as avaliações originais enquanto (b) mostra o resultado obtido aplicando-se a proposta de ajuste da avaliação do áudio original para 0.

Figura 22 – Ajuste das avaliações subjetivas por meio de pré-processamentos

	V1804	V2011	V2052	V2790	V2842	V2016	V2317	V3060	V3814	V1609	V1368	V2163	V1502	V1351	V2099	V1750	V1357	V1710	V1816	V1561
ORIG	0	2	1	0	1	-1	0	1	-1	1	0	-2	1	-1	0	1	0	1	-2	0
ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	0	1	0	0	0	-1	1	1	0	1	0	0	2	2	0	-1	-1	0	0	1
ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	1	1	1	-1	-1	-2	-1	2	-1	2	0	2	1	1	1	2	1	1	-1	2
ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	-1	-2	-1	-2	-2	-3	-2	-3	-3	-1	-2	-1	-1	-1	-2	-3	-2	-2	2	-2
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	1	0	1	2	1	-1	0	2	-1	2	0	1	1	1	0	-2	1	2	0	0
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	2	2	-1	-1	-1	-1	-1	-1	-1	3	1	2	3	1	1	1	1	-1	1	1
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	-3	-3	-2	-3	-3	-3	-3	-3	-3	-3	-1	-1	-3	-3	-1	-2	-2	-3	-2	-2
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	1	1	2	1	0	1	0	1	1	0	0	0	0	0	0	-1	1	-2	-2	0
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	3	-1	-1	-1	-1	-2	-1	-1	-2	2	1	2	-1	1	-1	-1	-1	-1	1	-1
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	-2	-3	-2	-3	-3	-3	-3	-3	-3	-2	-3	-2	-2	-2	-2	-3	-2	-3	-3	-3

(a) Avaliações “brutas” sem pré-processamento

	V1804	V2011	V2052	V2790	V2842	V2016	V2317	V3060	V3814	V1609	V1368	V2163	V1502	V1351	V2099	V1750	V1357	V1710	V1816	V1561
ORIG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	0	-1	-1	0	-1	0	1	0	1	0	0	2	1	3	0	-2	-1	-1	2	1
ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	1	-1	0	-1	-2	-1	-1	1	0	1	0	3	0	2	1	1	1	0	1	2
ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	-1	-3	-2	-2	-3	-2	-2	-3	-2	-2	-2	1	-2	0	-2	-3	-2	-3	-2	
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	1	-2	0	2	0	0	0	1	0	1	0	3	0	2	0	-3	1	1	2	0
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	2	0	-2	-1	-2	0	-1	-2	0	2	1	3	2	2	1	0	1	-2	3	1
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	-3	-3	-3	-3	-3	-2	-3	-3	-2	-3	-1	1	-3	-2	-1	-3	-2	-3	0	-2
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	1	-1	1	1	-1	2	0	0	2	-1	0	2	-1	1	0	-2	1	-3	0	0
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	3	-3	0	-1	-2	-1	-1	-2	-1	1	1	3	-2	2	-1	-2	-1	-2	3	-1
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	-2	-3	-3	-3	-3	-2	-3	-3	-2	-3	-3	0	-3	-1	-2	-3	-2	-3	-1	-3

(b) Avaliações ajustadas com base na avaliação do áudio original

Vamos denominar os resultados obtidos com este pré-processamento de **P1**. Mas, se observarmos bem os dados adquiridos verificamos que estes apresentam concentrações em torno do valor mínimo ou máximo devidas ao ajuste. Como forma de solucionar essa concentração das notas, aplicamos a proposta 2, fazendo o ajuste por voluntário, ou seja, cada um dos voluntários passa a apresentar conjuntos de avaliações contendo notas de -3 a 3, conforme o esperado. Esta técnica foi realizada sobre os dois conjuntos de avaliações: o conjunto P0 com as notas brutas, sem nenhum processamento e o conjunto P1 com as notas ajustadas. Denominemos esses dois novos conjuntos de **P2** e **P3**, respectivamente.

Em um conjunto de dados quando um dos elementos possui um valor muito distinto dos demais isto pode provocar uma distorção no valor real do conjunto. Sendo assim, a análise de pontos discrepantes proposta pelo terceiro ponto foi aplicada aos conjuntos de dados P2 e P3, gerando os conjuntos **P4** e **P5**. A análise foi realizada como descrito, do conjunto de dados extraiu-se a média e o desvio padrão, a partir dos quais se podem avaliar os intervalos de valores aceitáveis, e para todo elemento que não pertence ao intervalo, ele é então eliminado da análise. A Fig. 23 mostra a tabela com a análise de pontos discrepantes sobre o conjunto de dados P2 para um conjunto de teste.

Figura 23 – Análise de pontos discrepantes sobre conjunto de avaliações pré-processados

	A	B	C	D	E	F	G		Q	R	S	T	U	V	W	X	Y	Z
									Média	Variancia	Desvio Padrão	Min	Max					
14	ORIG	0,000	3,000	1,500	0,600	3,000	0,000		1,800	1,000	1,800	-1,800	0,600	0,785	2,077	1,441	-0,656	2,226
15	ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	0,000	1,800	0,000	0,600	1,500	0,000		-0,600	-1,000	0,600	0,600	1,800	1,005	1,195	1,093	-0,088	2,098
16	ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	1,000	0,600	1,500	3,000	3,000	0,000		-1,800	3,000	3,000	0,600	0,600	1,390	1,575	1,255	0,135	2,645
17	ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	1,000	1,800	3,000	1,800	1,500	3,000		-0,600	3,000	-1,800	-1,800	0,600	1,045	2,115	1,454	-0,409	2,499
18	ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	1,000	1,800	1,500	-0,600	0,000	-1,500		3,000	3,000	1,800	-0,600	3,000	1,385	2,043	1,429	-0,044	2,814
19	ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	2,000	3,000	-1,500	-0,600	0,000	0,000		1,800	3,000	-0,600	1,800	1,800	1,345	2,479	1,575	-0,230	2,920
20	ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	3,000	-0,600	-1,500	-0,600	0,000	-1,500		-0,600	-1,000	-0,600	1,800	-0,600	0,175	2,567	1,602	-1,427	1,777
21	ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	-1,000	-1,800	-1,500	-1,800	-1,500	-3,000		-3,000	-3,000	-1,800	3,000	-1,800	-1,665	1,849	1,360	-3,025	-0,305
22	ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	-3,000	-3,000	-3,000	-3,000	-3,000	-3,000		-1,800	-3,000	-3,000	-1,800	-1,800	-2,495	0,762	0,873	-3,368	-1,622
23	ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	-2,000	-3,000	-3,000	-3,000	-3,000	-3,000		-3,000	-3,000	-3,000	-3,000	-2,790	0,187	0,433	-3,223	-2,357	

Nas colunas V, W, X, Y e Z pode-se ver os valores, extraídos das avaliações de cada processamento, para média, variância, desvio padrão, o valor mínimo e o valor máximo, respectivamente, sendo os dois últimos os indicadores do intervalo das notas aceitáveis. As células destacadas em vermelho indicam os valores descartados para análise.

Uma análise rápida que pode ser realizada em cima destes dados é a verificação das médias das avaliações para os cenários de pré-processamento apresentados, a Tabela 6 nos mostra a relação de médias das avaliações após os pré-processamentos para cada um dos áudios processados.

Tabela 6 – Análise por média das avaliações após aplicação do tratamento de dados - *bad*

Áudio Avaliado	P0	P1	P2	P3	P4	P5
ORIG	0,100	0,000	0,785	0,715	0,906	0,800
ATSR_A1X5_B0X45_Q0X45_P1X5_G1X5	0,300	0,200	1,005	0,915	0,981	0,925
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X0	0,550	0,450	1,390	1,340	1,217	1,067
ATSR_A1X5_B0X45_Q0X45_P1X5_G2X5	0,200	0,100	1,045	0,865	1,008	0,869
ATSR_A2X5_B0X9_Q0X9_P2X5_G1X5	0,550	0,400	1,385	1,280	1,127	0,973
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X0	0,550	0,400	1,345	1,250	1,022	2,073
ATSR_A2X5_B0X9_Q0X9_P2X5_G2X5	-0,300	-0,350	0,175	0,030	-0,600	-0,560
ATSR_A3X5_B1X8_Q1X8_P3X5_G1X5	-1,700	-1,700	-1,665	-1,915	-1,911	-2,174
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X0	-2,450	-2,200	-2,495	-2,515	-2,788	-2,925
ATSR_A3X5_B1X8_Q1X8_P3X5_G2X5	-2,600	-2,400	-2,790	-2,890	-3,000	-3,000

Quando comparamos  $P0$  a  $P1$ ,  $P2$  a  $P3$  e  $P4$  a  $P5$  podemos verificar que a diferença entre seus resultados é pequena, apresentando um único caso em que a diferença se estabelece acima de 0,25. Logo, a utilização de um ou de outro conjunto de dados não provocaria grandes impactos no resultado final.

## 4.2 Avaliações Objetivas

Como apresentado no final do capítulo anterior, as avaliações objetivas foram realizadas utilizando-se a recomendação da ITU tendo como resultado a geração dos MOV's de cada processamento. Cada conjunto de variáveis de saída apresenta informações sobre a estrutura do sinal de áudio e a partir dessas informações espera-se estimar a qualidade dele na reprodução dos tons graves.

Em relação ao tratamento do conjunto de dados obtidos foram realizados dois pré-processamentos para que se diminuísse a complexidade do sistema de regressão e para minimizar problemas ocasionados pela divergência nas unidades das variáveis. Como observamos pela Tabela 4, o conjunto de variáveis de saída da análise objetiva é bem extenso, contando com 11 variáveis.

O primeiro pré-processamento utilizado trata-se da seleção de variáveis, ou seja, selecionamos as variáveis que possam trazer “mais informação” ao sistema, de forma que, tendo um conjunto de variáveis, possamos distinguir a qualidade para este conjunto. A primeira eliminação realizada dentre as variáveis é a análise de qualidade apresentada pelo algoritmo. Ela não tem relação direta com a qualidade que buscamos e por ser um valor dependente das demais variáveis pode causar cálculo desnecessário, de modo que o processo torne-se mais caro computacionalmente. Outra característica que logo se observa é que a variável *MFPDB* da Tabela 4 não nos traz muita informação sobre a qualidade do áudio, uma vez que, para todos os processamentos analisados, o retorno dela é sempre o mesmo. Sendo assim, podemos descartá-la da análise. Se observarmos as demais 10 variáveis, na Tabela 4, podemos notar que, todas possuem variações de seus valores dependendo do processamento observado. Essas variações, que mesmo em alguns casos não sendo apresentadas entre determinados conjuntos de processamentos, podem especificar uma relação com a qualidade esperada.

O segundo pré-processamento aplicado foi a normalização dos dados, aplicando-se a Eq. 4.1 para gerar os novos dados proporcionalmente aos dados anteriores.

Por fim, todos os dados obtidos foram unificados em planilha juntamente com as avaliações subjetivas. São disponibilizadas através deste link: <http://bit.ly/2Au3pB2> seis planilhas que apresentam os resultados obtidos para cada uma das seis propostas apresentadas como alternativas de pré-processamento para as avaliações subjetivas. Cada planilha apresenta sete folhas com dados, sendo as cinco primeiras os resultados das avaliações e seus pré-processamentos, a sexta folha apresenta o conjunto de variáveis resultantes da avaliação objetiva mais uma coluna que apresenta o conjunto de médias das avaliações subjetivas para cada algoritmo de processamento e a sétima folha apresenta os mesmos dados da folha seis, porém normalizados de acordo com a Eq. 4.1, apresentando valores entre 0 e 1 para cada variável.

### 4.3 Regressão

Finalizadas as avaliações de qualidade e em posse dos resultados obtidos, seguimos para a análise de relação entre as variáveis de saída (MOV's) calculadas pela avaliação objetiva de cada processamento e a nota da avaliação subjetiva, calculada pela média das avaliações dos voluntários para cada processamento.

A proposta é utilizar uma rede neural como um modelo de regressão, tendo como entrada as 10 variáveis selecionadas e como saída a nota da avaliação. Para tal, se faz necessária a especificação do modelo da rede, o treinamento e validação da mesma e, por fim, sua aplicação ao sistema completo proposto inicialmente.

Para o treinamento e validação da regressão foi utilizado o MATLAB R2017a,

que provê diversas aplicações para cálculo, estatística e otimização de dados. Dentre as aplicações disponíveis existem algumas específicas para classificação, clusterização, filtragem de distribuição, otimização, dentre outras. A aplicação utilizada chama-se “*Regression Learner*”.

***Regression Learner*** é uma aplicação disponibilizada pelo MATLAB que possibilita ao usuário escolher dentre vários algoritmos para treinar e validar modelos de regressão. Além disso, ele também disponibiliza o erro da validação, ou seja, após treinar vários modelos é possível comparar qual o melhor a partir da comparação dos erros.

Usando esse aplicativo pode-se explorar os dados, selecionar recursos, especificar esquemas de validação, treinar modelos e avaliar os resultados. Ele possibilita também realizar treinamento automatizado para procurar o melhor tipo de modelo de regressão dentre os 19 tipos distintos disponíveis, compreendidos em cinco grupos de acordo com a sua metodologia:

1. Modelos de Regressão Linear
2. Árvores de Regressão
3. Máquinas de Vetor de Suporte (SVM - *Support Vector Machine*)
4. Modelos de Regressão Gaussiana (GPR - *Gaussian Process Regression*)
5. Conjuntos de Árvores de Regressão

É possível escolher executar apenas um tipo de modelo para o conjunto de dados inserido, executar todos os processos de um grupo ou ainda executar todos os 19 processos disponíveis. Durante a realização deste projeto, foi utilizada a terceira opção, fazendo uso de todos os tipos de modelos e realizando-se a comparação através do erro médio quadrático (RMSE em Inglês) dos resultados obtidos com cada um dos modelos.

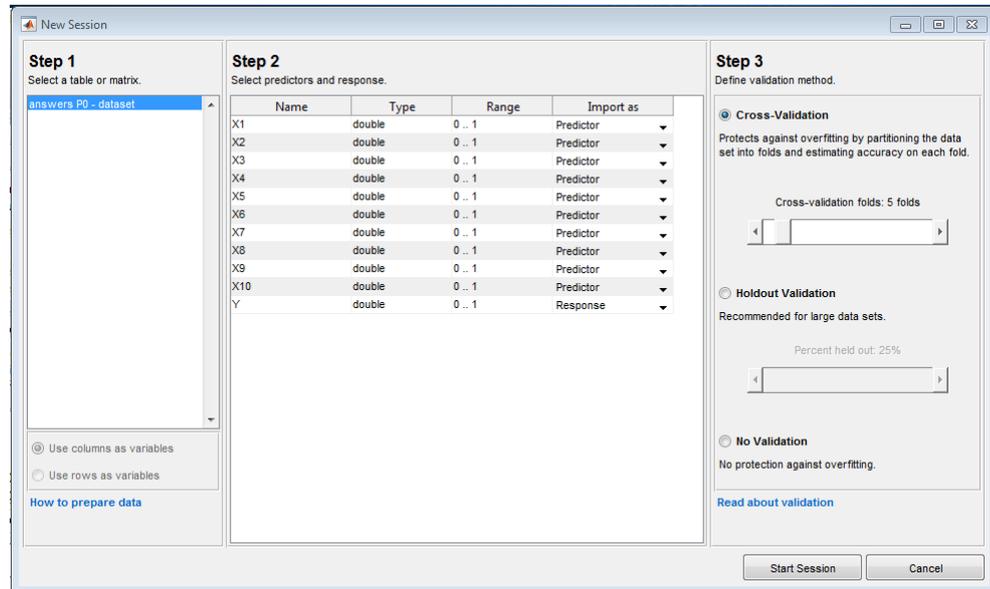
Os seis conjuntos de dados, apresentados anteriormente como  $P_0$ ,  $P_1$ ,  $P_2$ ,  $P_3$ ,  $P_4$  e  $P_5$ , foram submetidos à análise de regressão por meio da metodologia descrita acima.

Os conjuntos de dados coletados proporcionaram um total de 45 amostras de dados. Essas amostras foram utilizadas para o treinamento e a validação da rede. Foi utilizada a validação cruzada através do método *k-fold*.

A Fig. 24 apresenta o início de uma sessão de treinamento e especificação do esquema de validação. Os dados são extraídos de uma planilha, exibidos para seleção e após seleção seguem para esta tela onde se especifica quais variáveis são preditoras, quais são respostas e quais (caso existam) não importam para análise. Também nessa tela se

informa o tipo de validação que será utilizada ou se não haverá validação. Para cada tipo de validação é solicitada a especificação da proporção utilizada.

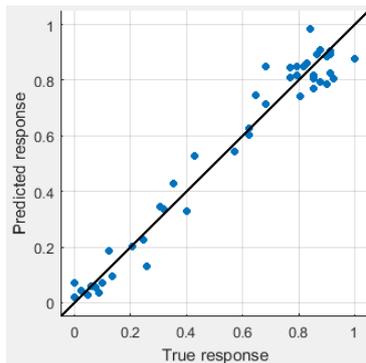
Figura 24 – Configuração de sessão de treinamento no *Regression Learner*



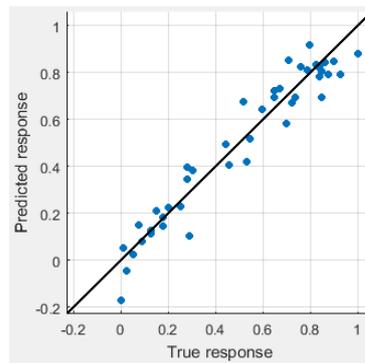
São apresentados a seguir os gráficos de dispersão obtidos para os melhores resultados (menor taxa de erro) de cada um dos conjuntos de dados utilizando a validação cruzada por  $k$ -fold com  $k = 5$  e  $k = 25$ . Lembrando que, quanto maior o valor de  $k$ , maior é o custo computacional da operação.

#### 4.3.1 Regressão utilizando Validação Cruzada com $k = 5$

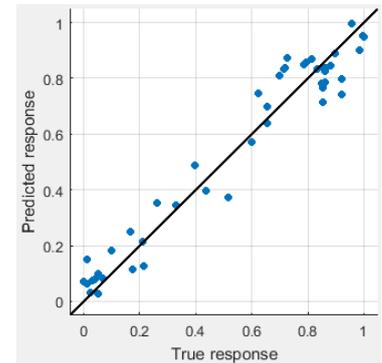
Inicialmente foram realizadas as análises utilizando o valor de  $k$  recomendado pela aplicação ao se optar pelo uso da validação cruzada por  $k$ -fold. Os resultados são apresentados na Fig. 25.

Figura 25 – Gráficos de Dispersão utilizando *5-fold*

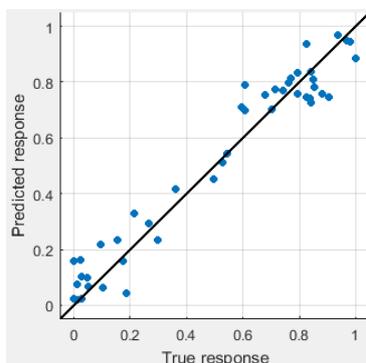
(a) Conjunto de dados P0



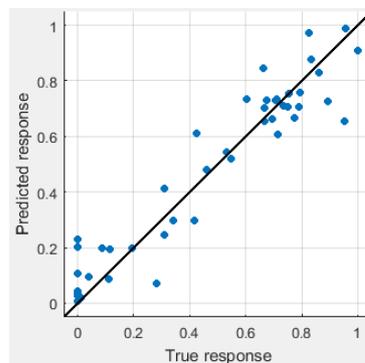
(b) Conjunto de dados P1



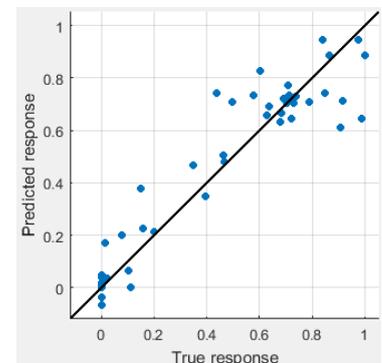
(c) Conjunto de dados P2



(d) Conjunto de dados P3



(e) Conjunto de dados P4



(f) Conjunto de dados P5

A Tabela 7 apresenta a relação de cada conjunto de dados com o menor valor de RMSE obtido pela regressão com os diversos tipos de modelo e o modelo que apresentou este melhor resultado.

Tabela 7 – Relação de melhores modelos de regressão por conjunto de dados, determinadas pelo RMSE ( $k = 5$ )

Conjunto de Dados	RMSE	Tipo de Modelo	Função <i>kernel</i>
P0	0,07	Gaussiano	<i>Matern 5/2</i>
P1	0,08	Máquina de Vetor	Quadrática
P2	0,08	Gaussiano	Exponencial
P3	0,09	Gaussiano	<i>Matern 5/2</i>
P4	0,10	Gaussiano	Exponencial
P5	0,12	Gaussiano	<i>Matern 5/2</i>

O erro médio quadrático (RMSE) é dado na mesma unidade da variável de análise, neste caso sendo apresentada entre 0 e 1, pode-se dizer que a menor taxa de erro encontrada foi de 7%.

Ao observar os gráficos de dispersão apresentados na Fig. 25, nota-se uma diminuição na força da relação entre as variáveis preditoras e a resposta, isto é comprovado na Tabela 7 onde se tem a menor taxa de erro para o conjunto de dados P0 e a maior para o conjunto P5. Uma razão que pode ser responsável por essa diminuição na correlação é o pré-processamento de redistribuição das notas para a escala -3 a 3, este procedimento pode ter provocado a perda das características da avaliação.

### 4.3.2 Regressão utilizando Validação Cruzada com $k = 25$

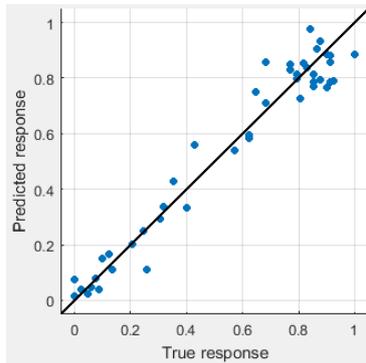
Uma segunda execução para teste de regressão foi realizada utilizando os mesmos conjuntos de dados, porém com um número maior de cruzamentos na validação, utilizando  $k = 25$  espera-se que o sistema possa se aproximar mais do resultado desejado, uma vez que o número de validações aumenta fazendo com que o modelo seja mais ajustado. A Tabela 8 apresenta, assim como no caso anterior, os erros obtidos para o melhor tipo de modelo do conjunto de dados.

Tabela 8 – Relação de melhores modelos de regressão por conjunto de dados, determinadas pelo RMSE ( $k = 25$ )

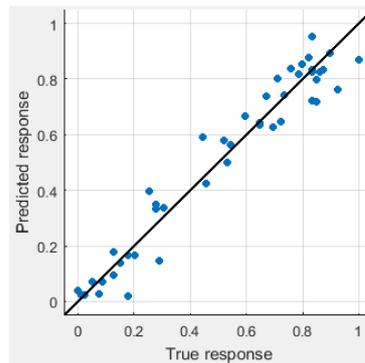
Conjunto de Dados	RMSE	Tipo de Modelo	Função <i>kernel</i>
P0	0,07	Gaussiano	Exponencial
P1	0,07	Linear	<i>Stepwise</i>
P2	0,07	Gaussiano	<i>Matern 5/2</i>
P3	0,07	Gaussiano	<i>Matern 5/2</i>
P4	0,09	Gaussiano	Exponencial
P5	0,11	Gaussiano	<i>Matern 5/2</i>

A seguir, na Fig. 26 são exibidos os gráficos de dispersão desta análise. Conforme esperado, houve melhoria dos resultados para quase todos os conjuntos de dados. No entanto apesar da maioria dos conjuntos de dados conseguirem ter resultados previstos com taxas de erro muito próximas, pode-se notar as diferenças das suas distribuições, mesmo naquelas que tiveram taxas de erro exatamente iguais.

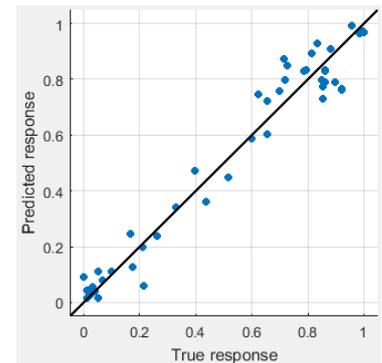
É possível observar que em todos os casos as amostras que apresentam previsão mais distinta do valor real concentram-se nas avaliações de valor mais alto. Isso se deve, muito provavelmente, ao fato de não haver muitos casos de avaliações altas, onde nas avaliações brutas (valores sem pré-processamento), o mínimo e máximo encontrados são -2,950 e 1,150, respectivamente. De certa forma se no treinamento houvesse uma maior distribuição de casos, a estimativa seria mais precisa. Também é possível afirmar que a utilização de uma validação mais robusta não nos trouxe muito efeito, uma vez que o erro persistiu em torno do melhor resultado obtida anteriormente (RMSE = 0,07).

Figura 26 – Gráficos de Dispersão utilizando *25-fold*

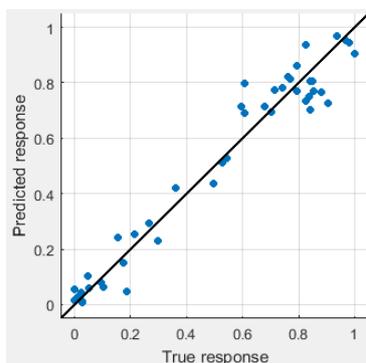
(a) Conjunto de dados P0



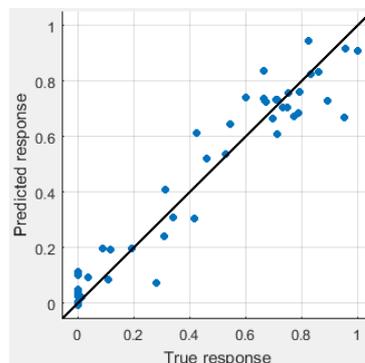
(b) Conjunto de dados P1



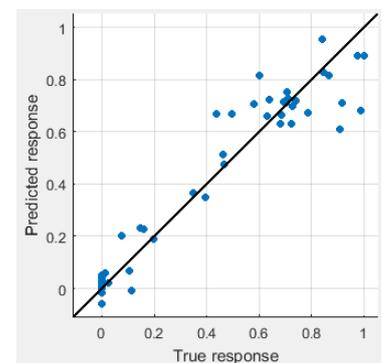
(c) Conjunto de dados P2



(d) Conjunto de dados P3



(e) Conjunto de dados P4



(f) Conjunto de dados P5

Apesar de ser apresentada melhoria nos resultados dos demais conjuntos de dados, a utilização do conjunto original ainda apresenta-se mais favorável, uma vez que se eliminaria a necessidade de pré-processamento dos dados, diminuindo assim, mesmo que com pouco impacto, o custo computacional do sistema.

O *Regression Learner* possibilita a exportação do modelo treinado para execução de testes no ambiente MATLAB. Porém utilizando de outro aplicativo, denominado *MATLAB Coder* é possível converter a função de MATLAB para linguagem C/C++, sendo possível assim o uso externo do modelo de regressão.

## 5 Conclusões

O trabalho realizado teve como objetivo o desenvolvimento de um sistema de melhoria da resposta para baixas frequências (tons graves) em alto-falantes que possuem reprodução pobre destas. Através da utilização de algoritmos de processamento de sinais pretendia-se gerar um novo sinal de áudio a partir do original, e este novo sinal teria uma resposta melhor no alto-falante.

Além disso objetivou-se o desenvolvimento de um sistema de autoanálise da qualidade do processamento, ou seja, após se obter o processamento do áudio é verificada a qualidade deste, no que diz respeito a reprodução de graves, levando-se em consideração a audição humana, como um humano avaliaria o áudio gerado.

O processamento de sinais de áudio é um contexto de aplicação complexo, envolvendo desde a qualidade do sinal gerado até a segurança de que ele não provoque danos ao ouvinte, uma vez que determinados sinais podem provocar dor ou até ferir. Para isso foi necessário um estudo básico do comportamento do ouvido humano em resposta a sinais de áudio, promovendo ampliação de conhecimento. O processamento proposto baseou-se na utilização de quatro algoritmos, os quais apresentam bons resultados na literatura para a aplicação proposta.

Logo, se observou que a utilização do algoritmo ATSR se mostrava favorável aos demais, sendo um algoritmo parametrizável foi possível especificar parâmetros distintos e analisar a eficácia do método. Através da realização de testes com voluntários foi possível obter uma boa massa de dados que a partir de análises estatísticas permitiu a confirmação do acima citado: o ATSR mostra-se promissor para o desenvolvimento do sistema.

Para a realização do sistema de regressão foram utilizados os resultados obtidos por meio da análise dos voluntários e as variáveis geradas como resultado da análise da qualidade objetiva. Como apresentado no capítulo anterior, foi possível especificar uma relação entre as variáveis e a avaliação de qualidade, logo sendo possível ao sistema predizer a qualidade de um áudio a partir de suas variáveis objetivas.

## 6 Trabalhos Futuros

A partir dos resultados apontados no Capítulo 4 constatou-se que é possível a realização do sistema de regressão entre as variáveis objetivas e as subjetivas, como consequência dessa constatação podemos apontar como trabalhos futuros:

1. Aumento da base de treinamento: Os testes realizados provaram que para um conjunto de variáveis de qualidade calculadas objetivamente pode-se realizar uma previsão da qualidade subjetiva na reprodução do sinal de áudio, porém para se obter um sistema de previsão preciso é necessário que este possua um conjunto de treinamento bem diversificado e de grande magnitude a fim de se obter um sistema não direcionado aos casos de análise.
2. Desenvolvimento do sistema de otimização: Dado que temos conjuntos de parâmetros e para cada um deles temos notas que definem sua qualidade para o áudio de referência, necessita-se gerar um modelo de otimização para estes parâmetros, de modo que ao término das iterações seja gerado um áudio cujo processamento apresente o conjunto ótimo de parâmetros.
3. Implementação do sistema proposto: A integração entre os sistemas implementados constituem o modelo de sistema proposto e apresentado pela Fig. 13, a união desses sistemas deve ser capaz de proporcionar a melhoria do áudio de entrada sem a necessidade de influência do usuário.
4. Desenvolvimento físico: Com o modelo montado e implementado podemos realizar a sua implementação física, gerando um sistema sonoro com alto-falantes de material piezoelétrico e com capacidade de melhoria na resposta às baixas frequências.

Os realização dos itens enumerados acima fecham o desenvolvimento do sistema proposto. Porém recomenda-se também a realização de algumas outras ações a título de análise e melhoria no resultado final, são elas:

- Teste com *Soft Clipping*: A primeira fase de testes realizada demonstrou que o algoritmo *Soft Clipping* é um candidato válido para a geração de harmônicas capazes de provocar a melhoria da resposta às baixas frequências, é válido e recomendado a avaliação de sua utilização no processo.
- Remoção do filtro HPF no áudio original: A utilização do filtro antes da soma com o resultado do processamento tem por objetivo evitar a distorção residual provocada pela soma das baixas frequências, porém ao se aplicar o filtro pode ocorrer perda de

informações. Uma vez que se aplica o filtro HPF sobre o áudio processado a distorção residual que se pretendia evitar já é evitada, sendo desnecessário, talvez prejudicial, o uso do filtro no áudio original.

- Separação da percussão: É possível que o processamento para melhoria do baixo na melodia tenha danificado alguns pontos da percussão, fazendo com que as avaliações não refletissem a melhoria, pois os danos na percussão afetaram a reprodução, neste caso propõe-se a separação do áudio em duas faixas, uma contendo a percussão e outra a melodia, o processamento de melhoria seria realizado sobre a melodia e posteriormente esse seria somado a percussão novamente. Para este caso seria necessária a realização de novos testes objetivos, subjetivos e de regressão. (Sendo desenvolvido em por Danilo Oliveira [5])

## Referências

- [1] MELLERT, V. et al. Impact of sound and vibration on health, travel comfort and performance of flight attendants and pilots. *Aerospace Science and Technology*, v. 12, n. 1, p. 18 – 25, 2008. ISSN 1270-9638. Aircraft noise reduction. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1270963807001174>>.
- [2] FROTA&CIA. *Frota de veículos em circulação no Brasil tem estabilidade*. 2017. Disponível em: <<http://www.frotacia.com.br/noticia/frota-de-veiculos-em-circulacao-no-brasil-tem-estabilidade>>.
- [3] NIVEL-SOM. *O mascaramento sonoro*. Disponível em: <<http://www.mascaramentosonoro.com.br>>.
- [4] LARSEN, E.; AARTS, R. Reproducing low-pitched signals through small loudspeakers. v. 50, 01 2012.
- [5] CARVALHO, D. O. \_\_\_\_\_. Dissertação (Dissertação de Mestrado) — Programa de Mestrado em Engenharia de Sistemas - Universidade de Pernambuco, a Publicar.
- [6] ARRUDA, F. *Como funciona um fone de ouvido [infográfico]*. 2011. Disponível em: <<https://www.tecmundo.com.br/infografico/11639-como-funciona-um-fone-de-ouvido-infografico-.htm>>.
- [7] GERSTLE, B. T. Tunable virtual bass enhancement. *University of Miami*, 2009.
- [8] PSICOACÚSTICA. Disponível em: <<http://www.eac.ufsm.br/pesquisa/qualidade-sonora>>.
- [9] GERGES, S. N. Y. *Ruído: fundamentos e controle*. 1. ed. [S.l.]: Editora Imprensa Universitária UFSC, 1992.
- [10] SALIBA, T. M. *Manual Prático de Avaliação e Controle do Ruído - PPRA*. 4. ed. [S.l.]: Editora LTr, 2008.
- [11] SANTOS, U. de P. *Ruído: Riscos e prevenção*. [S.l.]: Hucitec, 1999. v. 64. (Saúde em debate, v. 64).
- [12] HALLIDAY, D.; RESNICK, R.; WALKER, J. *Fundamentos de Física*. 8. ed. [S.l.]: LTC, 2009. v. 2.
- [13] VILELA, A. L. M. *Audição - Anatomia da orelha*. Disponível em: <<http://www.afh.bio.br/sentidos/sentidos3.asp>>.

- [14] CARDOSO, G. R.; SAFT, J. B. Análise das características arquitetônicas para a acústica de salas de concreto. *Congresso de Iniciação Científica e Tecnológica do IFSP*, 2016.
- [15] MATSUWAKI, Y. et al. Evaluation of missing fundamental phenomenon in the human auditory cortex. *Auris Nasus Larynx*, v. 31, n. 3, p. 208 – 211, 2004. ISSN 0385-8146. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0385814604000926>>.
- [16] PROAKIS, J. G.; MANOLAKIS, D. G. *Digital signal processing: principles, algorithms, and applications*. [S.l.]: Prentice Hall, 1996.
- [17] PUHLMANN, H. *Processamento de Sinais - DSP - Parte 1*. 2014. Disponível em: <<https://www.embarcados.com.br/introducao-ao-processamento-digital-de-sinais-dsp-parte-1/>>.
- [18] LATHI, B. P. *Linear Systems and Signals*. 2. ed. [S.l.]: Oxford University Press, 2004.
- [19] SMITH, S. W. *The Scientist and Engineer's Guide to Digital Signal Processing: Chapter 14 - Introduction to Digital Filters*. Disponível em: <<http://www.dspguide.com/ch14/5.htm>>.
- [20] AARTS, R. M.; LARSEN, E. C.; SCHOBEN, D. Improving perceived bass and reconstruction of high frequencies for band limited signals. In: . [S.l.: s.n.], 2002.
- [21] UIT - União Internacional de Telecomunicações. Disponível em: <<https://nacoesunidas.org/agencia/uit/>>.
- [22] METHOD for the subjective assessment of intermediate quality level of coding systems. 2015. Disponível em: <<https://www.itu.int/rec/R-REC-BS.1534-1-200301-S/en>>.
- [23] METHOD for objective measurements of perceived audio quality. 2001. Disponível em: <<https://www.itu.int/rec/R-REC-BS.1387-1-200111-I/en>>.
- [24] KABAL, P. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. [S.l.], 2003.
- [25] A Comprehensive Guide to Data Exploration. 2016. Disponível em: <<https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/>>.
- [26] VALENÇA, M. *Fundamentos das Redes Neurais: exemplos em Java*. 2. ed. Olinda, PE: Livro Rápido, 2010. 386 p.
- [27] COUTINHO, L. D. *Utilizando reservoir computing para realizar diagnóstico precoce de mal de Alzheimer*. Dissertação (Monografia (Graduação em Engenharia da Computação)) — Universidade de Pernambuco - UPE, Recife, 2012.

- 
- [28] HAYKIN, S. *Redes Neurais: Princípios e prática*. 2. ed. [S.l.]: Bookman, 2017.
- [29] OO, N.; GAN, W.-S.; LIM, W.-T. Generalized harmonic analysis of arc-tangent square root (atsr) nonlinear device for virtual bass system. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 301–304, 2010.

# ANEXO A – Script para processamento com *Arc-Tangent Square Root.*

```

function [f] = ATSR(x, alfa, beta, zeta, psi)

    Ne = 12;
    No = 11;

    f1 = 0;
    for i=0:No
        f1 = f1 + (((-1)^i)/(2 * i + 1))*(beta^(2 * i)) * x^(2 * i + 1);
    end
    f1 = alfa * beta * f1;

    f2 = 0;
    for i = 0 : Ne
        f2 = f2 + (((-zeta)^(2 * i) * factorial(2 * i))/((1 - 2 * i) *
        ((factorial(i))^2 * 4^i))) * x^(2 * i);
    end
    f2 = psi * f2;

    f = f1 + f2 - psi;

```



```
function resp = slicer(signal, x1, x2)
    resp = zeros([1 (x2 - x1)]);
    counter = 1;

    for i = x1 : x2
        resp(counter) = signal(i);
        counter = counter + 1;
    end
end

function resp = fill_vector(integral_accumulator, slice, x1)
    for i = 1 : length(slice)
        integral_accumulator((x1 - 1) + i) = slice(i);
    end
    resp = integral_accumulator;
end

function resp = my_integral(signal, temp, dt, x1, x2)
    slice = slicer(signal, x1, x2);
    integ = cumsum(slice.* dt);
    resp = fill_vector(temp, integ, x1);
end

function resp = crossing_points(signal)
    count = 0; % indica quantos pontos de cruzamento existem no sinal

    if length(signal) > 1
        for i = 1 : (length(signal) - 1)
            if signal(i) < 0 && signal(i + 1) >= 0
                count = count + 1;
            end
        end
    end

    if count == 0
        resp = 0;
    else
        resp = zeros([1 count]);
        position_resp = 1;
    end
end
```

```
counted = 0; % indica quantos pontos de cruzamento foram contados

for j = 1 : (length(signal) - 1)
    if signal(j) < 0 && signal(j + 1) >= 0
        resp(position_resp) = j;
        counted = counted + 1;

        if counted == count
            break
        else
            position_resp = position_resp + 1;
        end
    end
end
end
end
end
```

## ANEXO C – Script para processamento com *Soft Clipping.*

```
function [Audio_softclipping] = SoftClipping(signal, knee)
    Audio_softclipping = zeros(1, length(signal));
    for i = 1 : length(signal)
        Audio_softclipping(i) = signal(i)/(knee * abs(signal(i)) + 1);
    end
```