

Software de Apoio à Afinação Vocal

Juliana Hermsdorf Henriques dos Santos

Dissertação apresentada à
Escola Superior de Tecnologia e Gestão
para obtenção do grau de Mestre em
Engenharia Industrial – Ramo Engenharia Eletrotécnica

Este trabalho foi efetuado sob orientação de:
Professor Doutor João Paulo Ramos Teixeira

Co-orientado por:
Professor Doutor Roberto Ribeiro Neli

Software de Apoio à Afinação Vocal

Juliana Hermsdorf Henriques dos Santos

Dissertação apresentada à
Escola Superior de Tecnologia e Gestão
para obtenção do grau de Mestre em
Engenharia Industrial – Ramo Engenharia Eletrotécnica

Este trabalho foi efetuado sob orientação de:
Professor Doutor João Paulo Ramos Teixeira
Co-orientado por:
Professor Doutor Roberto Ribeiro Neli

2019

Agradecimentos

Agradeço primeiramente aos meus pais, Rosângela e Manuel, aos meus irmãos, Marcelo e Rodrigo, e a minha avó Wandir por sempre me dar o suporte e o apoio que precisei, independentemente da situação. Nada seria possível sem vocês.

Agradeço ao Instituto Politécnico de Bragança e a Universidade Tecnológica Federal do Paraná, assim como aos meus orientadores João Paulo Teixeira e Roberto Ribeiro Neli pela oportunidade e auxílio necessário para a realização desse trabalho. Assim, como agradeço a paciência e as boas conversas que tive com o pessoal do laboratório, em especial a Mariana Jatobá e Ives Gutierrez.

Agradeço a Luana Diniz e a Larissa Luz por tantos anos de amizade, que sempre me mostrou que podemos estar em cidade ou países diferentes que ainda posso contar com vocês sempre.

Agradeço aos amigos que fiz em Campo Mourão, especialmente ao Alexandre Souto, Eduardo Ximenes, Eric Moretti, Idalberto Cardoso Jr, Gustavo Coelho, João Carlos Vicente, Lucas Sakurada, Matheus Biaggio, Pedro Santos, Thiago Moya e Vitor Amâncio que foram minha família desde do começo da faculdade.

Agradeço ao pessoal do crossfit, em especial ao coach Pedro Maia pelo incentivo de sempre.

Agradeço pela família que Bragança me deu, todos tiveram uma grande importância nesse mais de 1 ano longe de casa, em especial, ao Arthur Casarini, Luiz Miguel Vilche, Letícia Cena, Jean Hilário, Gláucio Kenji, Maria Vertonha, Francieli Graeff, Murilo Esteves, Ighor Amorim, William Molina, Felipe Veloso, Isabelle Corrêa, Lucas D'Avila, Allana Netto, Vitor Hugo, Fabiane Setti, Paulo Zamarchi, Rafael Palmeira, Jessica Sibila, Franciele Fernandes, Letícia Damo.

Resumo

A afinação de voz para um cantor é de extrema importância, podendo comprometer o seu desempenho. O desenvolvimento desse trabalho dá enfoque a ferramentas que auxiliem no aprimoramento da afinação de voz, visando usuários com e sem experiência musical. O *software* realiza a aquisição do sinal da voz determinando a frequência fundamental que está diretamente relacionada com o tom, aplicando a uma interface gráfica que apresente um *feedback* em tempo real. Essa interface dispõe de diversas ferramentas para o apoio à afinação da voz. Utilizou-se neste trabalho o *software* MATLAB na versão 64 bits para o desenvolvimento do sistema. As ferramentas de apoio à afinação são quatro programas: “Classificação Vocal”, “Treino de Escalas”, “Treino de Músicas” e “Compositor de Música”. O programa de “Classificação Vocal” determina por um curto período de tempo se a voz é Soprano, Mezzo-soprano ou Contralto, sendo essas para vozes femininas, ou se a voz é Tenor, Barítono ou Baixo, para vozes masculinas, sendo respectivamente a da voz mais aguda para a mais grave. Essas classificações são utilizadas como base para o programa de “Treino de Escalas”, de acordo com a classificação vocal se tem uma escala de notas para ser treinada. No programa “Treino de Músicas” se tem uma base de dados de músicas sendo possível selecionar uma de acordo com o nível de dificuldade, podendo ser fácil, intermediário ou difícil, além de poder realizar adaptações quanto ao ritmo e a frequência base da mesma. E no programa “Compositor de Música” é possível o utilizador criar uma música adicionando as notas desejadas para que possa treiná-las como ocorre no programa “Treino de Músicas”. Por fim, deseja-se verificar sua funcionalidade com utilizadores, testando o seu funcionamento e analisando o desempenho do *software*. Ao fim, é realizado um inquérito para avaliar a aceitação do programa com aquisição em 64 bits.

Palavras-Chave: Afinação Vocal, Autocorrelação, Frequência Fundamental, MATLAB, Música.

Abstract

Voice tuning for a singer is of the utmost importance and may compromise his performance. The development of this work focuses on tools that help improve voice tuning, targeting users with and without music experience. The software performs voice signal acquisition by determining the fundamental frequency that is directly related to the tone, applying it to a graphical interface that provides real-time feedback. This interface has several tools to support voice tuning. In this work we used the 64-bit MATLAB software for system development. The tuning support tools are four programs: "Vocal Rating", "Scale Training", "Music Training" and "Music Composer". The "Vocal Rating" program determines for a short period of time whether the voice is Soprano, Mezzo-soprano or Contralto, whether these are for female voices, or whether the voice is Tenor, Baritone or Bass, for male voices, respectively, from the highest to the lowest voice. These ratings are used as the basis for the "Scale Training" program, according to the vocal rating if you have a grade scale to be trained. In the "Music Training" program you have a music database and you can select one according to the difficulty level, which can be easy, intermediate or difficult, and you can make adjustments to the rhythm and base frequency of the same. And in the "Music Composer" program it is possible for the user to create a song by adding the desired notes so that he can train them as in the "Music Training" program. Finally, we want to verify its functionality with users, testing its operation and analyzing the performance of the software. Finally, a survey is conducted to evaluate the acceptance of the 64-bit acquisition program.

Keywords: Voice Tuning, Autocorrelation, Fundamental Frequency, MATLAB, Music.

Índice

CAPÍTULO 1: INTRODUÇÃO	1
1.1. OBJETIVOS	2
1.2. ORGANIZAÇÃO DO TRABALHO	3
1.3. ESTADO DA ARTE	4
CAPÍTULO 2: METODOLOGIA.....	9
2.1. VOZ.....	9
2.2. PRODUÇÃO DA VOZ.....	10
2.3. MATLAB.....	11
2.4. FREQUÊNCIA FUNDAMENTAL	11
2.4.1. <i>Método da Autocorrelação</i>	13
2.4.2. <i>Código da Autocorrelação</i>	14
2.5. AQUISIÇÃO DO SINAL.....	17
2.5.1. <i>Frequência de Amostragem</i>	17
2.5.2. <i>Microfone</i>	20
2.6. AQUISIÇÃO DO SINAL NO MATLAB 64 BITS	21
2.6.1. <i>AudioRecorder</i>	21
2.6.2. <i>Data Acquisition Toolbox</i>	23
2.7. PÓS PROCESSAMENTO	27
2.8. TEORIA DE MÚSICA.....	28
2.8.1. <i>Escala de Música</i>	28
2.8.2. <i>Classificação Vocal</i>	30
2.9. AVALIAÇÃO DA PERFORMANCE	33
2.9.1. <i>Erro Quadrático Médio</i>	33
2.9.2. <i>Raiz do Erro Quadrático Médio</i>	33
2.9.3. <i>Erro Absoluto Médio</i>	34
2.9.4. <i>Comparando MSE, RMSE e MAE</i>	34
CAPÍTULO 3: DESENVOLVIMENTO.....	37
3.1. MENU PRINCIPAL.....	37
3.2. CLASSIFICAÇÃO VOCAL	38
3.3. TREINO DE ESCALA.....	40
3.4. TREINO DE MÚSICAS.....	43
3.5. COMPOSITOR DE MÚSICA	46

CAPÍTULO 4: AVALIAÇÃO E RESULTADOS DO SOFTWARE	51
4.1. CARACTERIZAÇÃO DA AMOSTRA	51
4.2. AVALIAÇÃO GERAL	52
4.3. AVALIAÇÃO ESPECÍFICA	55
4.3.1. <i>Classificação Vocal</i>	56
4.3.2. <i>Treino de Escalas</i>	59
4.3.3. <i>Treino de Músicas</i>	60
4.3.4. <i>Compositor de Música</i>	62
CAPÍTULO 5: CONCLUSÃO E TRABALHOS FUTUROS	65
5.1. CONCLUSÕES GERAIS	65
5.2. TRABALHOS FUTUROS	66
BIBLIOGRAFIA.....	67

Lista de Figuras

Figura 1.1: Processo de aprendizagem do pitch no canto. (A) processo de aprendizagem tradicional em andamento; (B) impacto VFT em tempo real no processo de canto. KR (conhecimento da resposta por uma fonte externa) e CP (período de aprendizagem) (adaptada de Hoppe, Sadakata & Desain, 2006).....	4
Figura 1.2. Exemplo de funcionamento do software desenvolvido por Leão(2018).....	6
Figura 1.3: Exemplo de funcionamento do Sing & SEE.	7
Figura 1.4: Exemplo de funcionamento do aplicativo Vocal Pitch Monitor.	8
Figura 2.1: Trato Vocal (Almeida, 2012).....	10
Figura 2.2: Modelo genérico de produção de fala (Teixeira, Ferreira & Moreira, 2011)...	12
Figura 2.3: Parte 1 do método da autocorrelação (a) Sinal completo (b) Segmento de 40 ms.	15
Figura 2.4: Parte 2 do método da autocorrelação (a) Autocorrelação do segmento de 40 ms (b) Busca pelo maior valor da autocorrelação.	15
Figura 2.5: Variando o valor de N1.	17
Figura 2.6: Intervalo de variação de F0 de acordo com a frequência de amostragem.....	19
Figura 2.7: deltaF0 de acordo com a frequência de amostragem.	20
Figura 2.8: Microfone Blue Snowball.....	20
Figura 2.9: Configuração do objeto <i>AudioRecorder</i>	21
Figura 2.10: Funcionamento da aquisição da função <i>getaudiodata</i> (adaptado de MathWorks, 2019a).....	22
Figura 2.11: Fluxograma da aquisição e processamento do método <i>AudioRecorder</i>	23
Figura 2.12: Fluxograma da configuração da aquisição com a <i>Toolbox</i> (adaptado de MathWorks, 2019b).	24
Figura 2.13: Configuração da <i>Toolbox</i>	25
Figura 2.14: Configuração dos parâmetros da <i>Toolbox</i>	25
Figura 2.15: Início da aquisição utilizando a <i>Toolbox</i>	25

Figura 2.16: Fluxograma da aquisição e processamento do método da <i>Toolbox</i>	26
Figura 2.17: Função no MATLAB na detecção de zona de silêncio.....	28
Figura 2.18: Intervalo de um tom (Aleixo e Mansanares,2003).....	29
Figura 2.19: Pauta Geral para as faixas vocais. Na linha superior as vozes femininas e na linha inferior as vozes masculinas (Science Buddies Staff, 2018).	32
Figura 2.20: Caso de deslocamento horizontal.....	35
Figura 2.21: Caso de picos e pequenas variações.....	35
Figura 3.1: Menu principal do software.....	37
Figura 3.2: Fluxograma do funcionamento do programa “Classificação Vocal”.....	38
Figura 3.3: Janela com a classificação final.....	39
Figura 3.4: Interface gráfica do programa “Classificação Vocal”.....	40
Figura 3.5: Janela de informação do programa “Classificação Vocal”.....	40
Figura 3.6: Fluxograma do funcionamento do programa “Treino de Escalas”.....	41
Figura 3.7: Interface gráfica do programa “Treino de Escalas”.....	42
Figura 3.8: Janela de informação do programa de “Treino de Escalas”.....	43
Figura 3.9: Seleção de música por idioma ou nível.....	43
Figura 3.10: Paleta de transposição e de ajuste de ritmo.....	44
Figura 3.11: Interface gráfica do programa de “Treino de Músicas”.....	45
Figura 3.12: Fluxograma do funcionamento do programa “Treino de Músicas”.....	46
Figura 3.13: Paleta de configuração e implementação de música.....	47
Figura 3.14: Interface gráfica do programa “Compositor de Música”.....	48
Figura 3.15: Fluxograma do funcionamento do “Compositor de Música”.....	49
Figura 4.1: Gráfico sobre os alunos participantes.....	51
Figura 4.2: Gráfico de avaliação geral do programa.....	52
Figura 4.3: Gráfico da classificação dos participantes.....	56
Figura 4.4: Gráfico das notas gerais da “Classificação Vocal”.....	57
Figura 4.5: Gráfico das notas da “Classificação Vocal” pela categoria I.....	58

Figura 4.6: Gráfico das notas da “Classificação Vocal” pela categoria II.....	58
Figura 4.7: Gráfico das notas gerais do “Treino de Escalas”.....	59
Figura 4.8: Gráfico das notas do “Treino de Escalas” pela categoria I.	60
Figura 4.9: Gráfico das notas do “Treino de Escalas” pela categoria II.	60
Figura 4.10: Gráfico das notas gerais do “Treino de músicas”.	61
Figura 4.11: Gráfico das notas do “Treino de Músicas” pela categoria I.	62
Figura 4.12: Gráfico das notas do “Treino de Músicas” pela categoria II.....	62
Figura 4.13: Gráfico das notas gerais do “Compositor de Música”.	63
Figura 4.14: Gráfico das notas do “Compositor de Música” pela categoria I.....	64
Figura 4.15: Gráfico das notas da “Compositor de Música” pela categoria II.....	64

Lista de Tabela

Tabela 2.1: Frequência Máxima de acordo com o valor de N1.....	16
Tabela 2.2: Relação entre as notas musicais, intervalos e frequências correspondentes (Aleixo & Mansanares, 2003).....	30
Tabela 2.3: Extensão vocal de acordo com a classificação (Perelló et al., 1982; Junker, 1999).....	32
Tabela 3.1: Escala para treinamento de afinação vocal.	42
Tabela 4.1: Inquérito utilizado para avaliação do software.	53
Tabela 4.2: Avaliação geral do programa de acordo com a categoria I.....	54
Tabela 4.3: Avaliação geral do programa de acordo com a categoria II.	54
Tabela 4.4: Grau de avaliação específica de cada programa.....	55
Tabela 4.5: Notas relacionadas ao programa “Classificação Vocal”.....	57
Tabela 4.6: Notas relacionadas ao programa “Treino de Escalas”.....	59
Tabela 4.7: Notas relacionadas ao programa “Treino de Músicas”.....	61
Tabela 4.8: Notas relacionadas ao programa “Compositor de Música”.....	63

Lista de Abreviaturas

A/D – Conversão Analógico/Digital

CQ – *Larynx Closed Quocient*

IPB – Instituto Politécnico de Bragança

F0 – Frequência Fundamental

Fs – Frequência de Amostragem

GUIDE – *Graphical User Interface Development Environment*

Hz – Hertz

MAE – Mean Absolute Error

MATLAB – *Matrix Laboratory*

ms – milisegundos

MSE – Mean Square Error

PDAs – Pitch Detection Algorithms

RMSE – Root Mean Square Error

st – Semitons

SPL – Sound Pressure Level

T0 – Período fundamental

Ts – Período de Amostragem

UI – User Interfaces

USB – Universal Serial Bus

VFT – *Visual Feedback Technology*

Capítulo 1: Introdução

Para um cantor, a afinação de voz é de extrema importância, o ato de cantar envolve diversos aspectos, dentre eles a fisiologia, a respiração, a concentração e o conhecimento musical. Cantores quando bem treinados tem um controle vocal preciso, podendo ser observado quando comparados com cantores sem treinamento (Oliveira & Lima, 2016; Zaki-Azat, 2016).

O canto precisa ser praticado para desenvolver um controle preciso, assim como qualquer habilidade motora, necessita de tempo e prática, tendo sempre um *feedback* (Zaki-Azat, 2016). De acordo com Wilson, Lee, Callaghan e Thorpe (2008), os usuários que utilizaram um treinamento com o *feedback* visual obtiveram uma melhora significativa na precisão de afinação, diferentemente dos usuários que não o utilizaram, demonstrando nenhuma alteração.

A aquisição de dados pode ser definido como uma ferramenta que mede informações do mundo real, sendo a maioria sinais analógicos. Um sistema de aquisição de dados para aplicação de processamento de sinal e problemas gerais na engenharia, é composto por sensores e transdutores, módulo ou placa de aquisição de dados e unidade de processamento. No caso da aquisição de áudio, a captação dos dados ocorre por meio do microfone (Baptista, 2019).

O áudio processado digitalmente é resultado de um sinal manipulado através de um *software*, sendo que o processo de conversão do sinal analógico para o digital ocorre no *hardware*, isto é, na placa de áudio do computador. O processamento desse sinal é realizado num *software*, como por exemplo no *Matrix Laboratory* (MATLAB).

O objetivo geral deste trabalho é desenvolver um *Visual Feedback Technology* (VFT), para auxiliar no ensino de canto. Dando um *feedback* em tempo real sobre o desenvolvimento do usuário.

Atualmente existem diversos materiais acadêmicos para o treino de afinação vocal assim como aplicativos computacionais para afinação em tempo real. Entretanto, é necessário utilizar diversos *software* para desenvolver diferentes recursos vocais direcionados tanto aos alunos que iniciam o estudo em música, quanto os alunos autodidatas sem experiência em teoria

musical. Assim, surgiu a ideia de desenvolver um algoritmo que realiza a obtenção da frequência fundamental da voz, utilizando um sistema de aquisição de dados e processamento digital de sinal, onde o estudante de música pode de forma autodidata se apoiar a fim de melhorar a sua afinação.

As ferramentas desenvolvidas buscam criar treinos que auxiliem no processo de afinação vocal possibilitando o usuário selecionar o método de treinamento de seu interesse. Quatro ferramentas foram implementadas com o intuito de dar auxílio ao processo de afinação. A primeira é a detecção da classificação vocal, para que o usuário possa aplicar o resultado na segunda ferramenta e treinar uma escala musical na ordem crescente ou decrescente. A terceira ferramenta consiste em músicas pré-inseridas e por fim, a escritura da própria melodia para treinar.

1.1. Objetivos

O objetivo geral deste trabalho é desenvolver um *Visual Feedback Technology* (VFT), baseado no projeto realizado por Leão (2018), para auxiliar no ensino de canto, podendo ser utilizado tanto por iniciantes, sem o devido conhecimento em música como por pessoas que possuem um nível de conhecimento mais avançado nas técnicas de canto.

Para alcançar o objetivo geral é necessário executar algumas etapas intermediárias, sendo os objetivos específicos elencados a seguir:

Realizar uma análise de duas formas de aquisição de voz em tempo real utilizando o MATLAB de 64 bits, sendo uma delas a utilização de uma *Data Acquisition Toolbox*, chamada *Direct Sound*.

Desenvolver um sistema que seja capaz de captar o sinal de um microfone e utilizando o método da autocorrelação faça a determinação da frequência fundamental (F0).

Desenvolver uma interface gráfica interativa utilizando o GUIDE (*Graphical User Interface Development Environment*) do MATLAB para que o usuário tenha apoio para escolher o método de treinamento vocal que deseja.

Desenvolver uma função em que o usuário consiga detetar qual a sua classificação vocal.

Desenvolver uma função no qual o usuário possa realizar o treinamento vocal de acordo com sua classificação.

Desenvolver uma função que possua previamente uma base de músicas para que o usuário possa escolher qual música deseja treinar de acordo com o seu nível.

Desenvolver uma função que possua a inserção de notas musicais para a composição de música que poderá ser salva e treinada pelo usuário a livre modo.

Aproximar as informações obtidas à teoria musical, ao passo de que seja possibilitada a integração desta tecnologia na prática do ensino ao canto e da afinação da voz.

Realizar um questionário baseado no tradicional método de avaliação de usabilidade e técnicas de avaliações heurísticas desenvolvido por Nielsen e Molich (Nielsen, 1994; Nielsen & Molich, 1990).

Testar com utilizadores de diversos grupos para avaliar cada parâmetro do *software* em aspetos gerais e específicos, em cada ferramenta proposta, a fim de avaliar a aceitação do programa.

1.2. Organização do Trabalho

Este trabalho está dividido em cinco capítulos. O primeiro capítulo faz uma introdução descrevendo sobre o tema e os objetivos deste projeto. Apresentando, também, uma revisão bibliográfica das pesquisas acadêmicas realizadas e algumas aplicações comerciais de programas de apoio à afinação da voz atuais.

O segundo capítulo expõe os métodos utilizados, detalhando o funcionamento da aquisição de voz e dando um embasamento teórico para o passo a passo do *software*. Descrevendo assim, uma breve introdução do processo da produção vocal, o funcionamento do *software* MATLAB utilizado como base, a determinação dos parâmetros e do tratamento de dados que é realizado, isto é, aquisição de dados e seu pós processamento. A teoria musical na qual foi utilizada para a construção do *software* também é apresentada neste capítulo.

O funcionamento de cada uma das ferramentas implementadas é descrito no terceiro capítulo. Apresentando a organização geral do *software*, explicando cada uma das aplicações por meio de fluxograma, demonstrando a interface gráfica desenvolvida para auxiliar o usuário e esclarecendo seu funcionamento.

A avaliação geral do *software* e a avaliação específica de cada uma das ferramentas é exibida no quarto capítulo. Descrevendo sua performance avaliada a partir dos resultados obtidos por meio do questionário realizado com os usuários que testaram o *software*.

O último capítulo conclui a eficiência da aplicação através dos resultados obtidos na avaliação feita e apresentada no quarto capítulo. Além de se apontar melhorias na ferramenta e possíveis trabalhos futuros.

1.3. Estado da Arte

De acordo com Welch (1985), o aprendizado de uma aula de canto clássico se tem por meio de um professor que apresenta um modelo e o aluno que tenta imitar. Após isso, o professor fornece o *feedback* sobre o desempenho do aluno e, com isso, o aluno tenta melhorar, como é representado pela Figura 1.1a. A utilização de VFT em tempo real para aprendizado de canto é influenciado pelo esquema exposto na Figura 1.1b, dessa forma o aprendizado ocorre durante a resposta do aluno e no mesmo instante já se tem um retorno, sendo assim apresentado ao aluno um *feedback* durante o tempo de resposta.

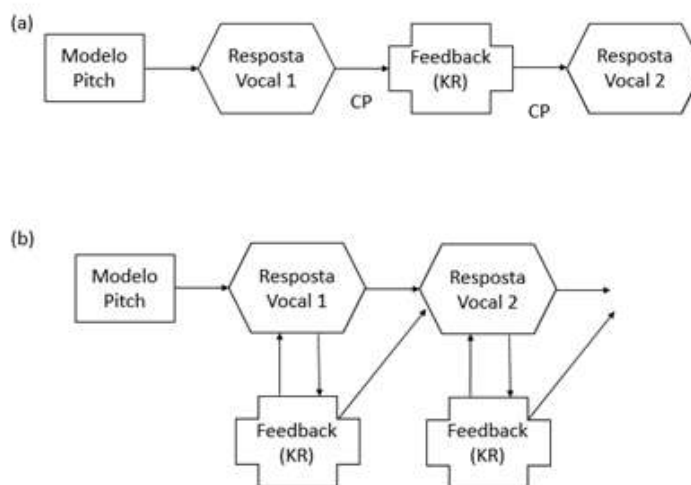


Figura 1.1: Processo de aprendizagem do pitch no canto. (A) processo de aprendizagem tradicional em andamento; (B) impacto VFT em tempo real no processo de canto. KR (conhecimento da resposta por uma fonte externa) e CP (período de aprendizagem) (adaptada de Hoppe, Sadakata & Desain, 2006).

Os recursos para treinamento de voz são diversos e se disponibilizam de diferentes formas de aquisição. O primeiro projeto envolvendo VFT em tempo real em *software* para auxílio educacional para canto foi o de Howard e Welch (1989). O projeto envolveu o desenvolvimento do SINGAD (*SINGing Assessment and Development*), seu objetivo era ser uma ferramenta pedagógica, no qual avaliava e desenvolvia o canto vocal de crianças de uma escola primária no Reino Unido. O SINGAD foi projetado para um microcomputador da BBC e fez uso de um pico-seletor, originalmente projetado para implantes de cocleares, para fazer uma estimativa precisa da frequência fundamental em tempo real. A referência musical do sistema fornecida para o cantor era por meio de bibliotecas musicais com três ajustes de oitavas.

O programa foi dividido em duas fases: uma fase de avaliação e uma fase de desenvolvimento. Na fase de avaliação, uma simples nota era tocada e o usuário tinha de responder cantando em um microfone. A média da frequência fundamental (F0) de toda a resposta cantada foi então comparada com a nota alvo. Na fase de desenvolvimento, o sinal em tempo real da F0 foi traçado em função do tempo e as notas selecionadas podiam ser colocadas na tela para orientar a resposta cantada. Ao fim da pesquisa, foi avaliado um efeito positivo da VFT em tempo real no desenvolvimento da capacidade de *pitching* de crianças do ensino fundamental, com e sem a orientação de um professor de canto quando comparada com a aula de canto tradicional (Howard & Welch, 1989; Hoppe, Sadakata & Desain, 2006).

Outro *software* desenvolvido que utiliza VFT para performance vocal em tempo real foi o ALBERT (*Acoustic and Laryngeal Biofeedback Enhancement in Real-Time*). Visando o aprimoramento do *biofeedback* e laringeo em tempo real, fornecendo ao cantor ferramentas para aprimorar a produção vocal. O ALBERT tem uma característica diferente das outras pesquisas, além de fornecer uma resposta de saída ao usuário, se tem o monitoramento da ação da laringe. O sistema possui os parâmetros de observação de F0, CQ (quociente fechado da laringe), razão espectral, SPL (amplitude), *jitter* e *shimmer*. Para não inserir informações em excesso para o usuário, o sistema foi construído para acrescentar os fatores definidos pelo usuário, combinando qualquer um dos parâmetros primários através de uma janela de visualização. Os parâmetros podem ser visualizados em uma, duas ou três dimensões. Para fins pedagógicos, os atributos de rotulagem dos eixos também podem ser modificados ou até ocultados na tela. Além disso, é possível alterar o tamanho e a cor da tela (Rossiter & Howard, 1996).

Com os avanços tecnológicos e os computadores de alto desempenho acessíveis, muitos programas de computador para treinamento visual foram desenvolvidos e promovidos para aprimorar o controle do tom vocal (Hoppe, Sadakata & Desain, 2006).

A pesquisa de Bartošek e Hanžl (2010) procura encontrar uma resposta proveniente da comparação de diferentes algoritmos de detecção de *pitch* (PDAs – *Pitch Detecting Algorithm*) na área de processamento de sinais de fala. Para avaliação objetiva dos PDAs, a estrutura foi projetada e implementada com base no banco de dados de referência de *pitch* existente. Todos os testes dos algoritmos são brevemente descritos. Os resultados mostram que o ponto mais fraco do algoritmo de correlação normalizada é a fase de distinção dos elementos vozeados e não vozeados.

O projeto de Ferreira (2012), com a ajuda dos *softwares* SingingStudio, Praat e o MATLAB e uma base de dados de diferentes sinais de voz, determina diversos fatores que estão envolvidos em uma aula de canto e os fenômenos acústicos. Os parâmetros para a análise foram o portamento, o falsete, a limpidez, o formante de Cantor e o vibrato. Para cada parâmetro, se teve os resultados por meio de análises estatísticas, tratando diversos parâmetros acústicos como o *jitter*, *shimmer* e a autocorrelação. O teste de Kruskal-Wallis e MannWhitney U foram utilizados para análise dos binômios Falsete/Modal de cantor e Limpidez/Soprosidade, comparando assim os parâmetros acústicos extraídos e evidenciando os que demonstram maior diferença. O único resultado com insucesso foi o binômio Limpidez/Soprosidade.

Neves (2016) desenvolveu uma ferramenta para auxiliar o processamento do sinal de voz após a gravação de uma forma simples para afinação de sinais monofônicos de voz, que possa afinar as notas cantadas automaticamente. O programa de desenvolvimento utilizado foi o MATLAB. Essa ferramenta possui uma interface gráfica que facilita a interação, permitindo alterações livres na altura e duração das notas. A solução adotada para se conseguir alterar a altura do sinal, sem interferir em sua duração é o *vocoder* de fase e de reamostragem. A curva de frequência fundamental ao longo de toda a duração do sinal é calculada aplicando o método YIN. Esse método utiliza método da autocorrelação, posteriormente é verificado o erro e realizado melhorias para reduzir a taxa de erro (De Cheveigné & Kawahara, 2002).



Figura 1.2: Exemplo de funcionamento do software desenvolvido por Leão(2018).

A pesquisa de Leão (2018) se baseia na criação de um *software* que utiliza o programa MATLAB na versão de 32 bits para realizar a aquisição do sinal de voz aplicando VFT.

Buscando reunir ferramentas que forneçam o suporte para auxiliar a afinação vocal. O *software* tem um diferencial que é o desenvolvimento de uma interface gráfica intuitiva de se utilizar, como apresentado na Figura 1.2. A frequência fundamental é calculada por meio da autocorrelação com center clipping. O sinal da voz adquirido é comparado com um vetor de notas que varia de acordo com o exercício desejado. Foram realizados testes com alunos do Instituto Politécnico de Bragança (IPB) com e sem experiência musical.

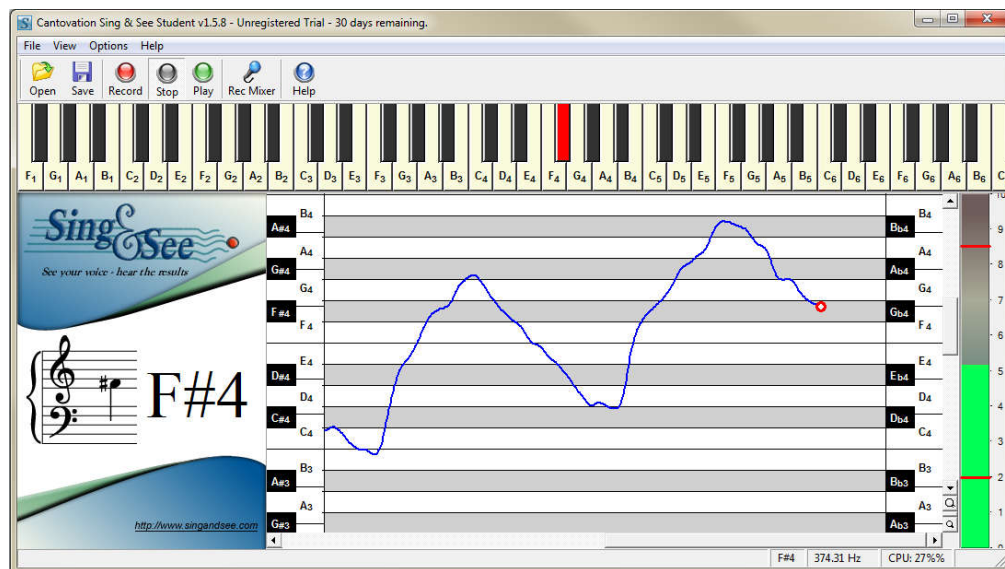


Figura 1.3: Exemplo de funcionamento do Sing & SEE.

O projeto SING & SEE foi apresentado na Conferência de Musicologia Interdisciplinar (Callaghan, Thorpe & Doorn, 2004). Visava o desenvolvimento de nova tecnologia VFT para o estudo de canto. As principais características da pesquisa foram a investigação de técnicas de análise acústica, métodos de exibição VFT de maneira significativa e as abordagens pedagógicas para implementação da tecnologia em prática. Três parâmetros foram distinguidos como relevantes para o uso no estúdio de canto: *pitch*, identidade da vogal (R1, R2) e timbre (espectrograma). O programa é projetado para atender as necessidades dos cantores e professores de músicas de todos os estilos e gêneros. Possui uma ferramenta que se tem o *feedback* em tempo real do tom em que está sendo emitida a nota vocal, assim como um gráfico que representa o timbre da voz. A função *Font of Knowledge* (Fonte de Conhecimento) é uma fonte de informação interativa que ajuda o estudante a entender a voz, utilizando o espectrograma que explora as diferentes estruturas harmônicas e seus respectivos sons.

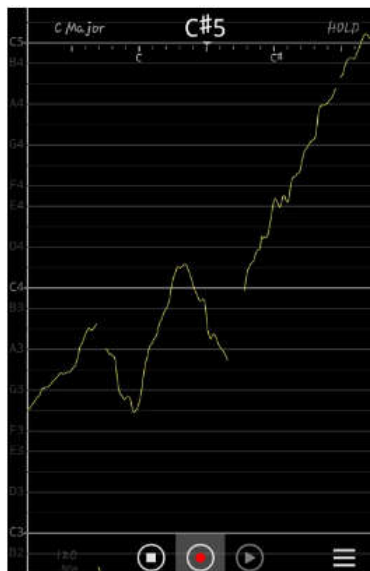


Figura 1.4: Exemplo de funcionamento do aplicativo Vocal Pitch Monitor.

O programa Vocal Pitch Monitor (2019) é um aplicativo que exibe de forma simplificada o *pitch* em tempo real, em outras palavras, um afinador vocal ou instrumental de forma simultânea. A ferramenta exibi a escala atual captada pelo som do microfone na parte superior da tela. O programa permite realizar a gravação do treinamento, sendo possível identificar a extensão vocal do usuário.

Capítulo 2: Metodologia

Equation Section 2

O capítulo a seguir expõe o desenvolvimento do *software*, explicando com base teórica as ferramentas utilizadas. Abordando, como por exemplo, a produção da voz, as ferramentas utilizadas no processo de aquisição de dados, o pós processamento e teoria de música aplicada como suporte para o desenvolvimento dos treinamentos.

2.1. Voz

A voz é uma importante ferramenta de comunicação e um funcionamento inadequado pode trazer profundas implicações na vida social e profissional de uma pessoa (Almeida, 2010).

A voz é resultado de um conjunto de fenômenos fisiológicos que determinam a emissão acústica, sendo ela ligada aos órgãos e ao sistema de respiração (Teixeira, 1995). Ela possui características que variam para cada pessoa, mudando de acordo com o sexo e a faixa etária, podendo refletir na voz o estado e comportamentos laríngeos. (Andrade, 2003; Lopes, 2018).

É proveniente do aumento de pressão do ar nos pulmões causado pelo diafragma, em que o fluxo de ar expelido passa pelas cordas vocais com diferentes padrões de adução (fechamento) e abdução (abertura). A voz é modelada pelas propriedades de reflexão e configuração do trato vocal. Haverá uma resistência na saída do ar se as cordas vocais estiverem aduzidas, causando assim a vibração delas gerando o chamado som vozeado que é a fonte sonora da voz, tratando-se da base da fala e do canto (Almeida, 2012; Lopes, 2008).

A fala e o canto são dois modos de uso do mesmo sistema, contudo, existem várias diferenças entre a voz falada e a cantada, como por exemplo, as durações das vogais que no caso da voz cantada são superiores às da voz falada, sendo a intensidade da cantada ser normalmente superior à voz falada (Almeida, 2012; Sundberg, 1977).

2.2. Produção da Voz

Os órgãos que envolvem a produção da voz são como um gerador de sons, no qual possui três unidades principais: a fonte de alimentação, o oscilador e o ressonador, referindo-se respectivamente aos pulmões, as cordas vocais e ao trato vocal (Sundberg, 1977).

O pulmão, devido ao excesso de pressão, conduz o ar que atravessa a traqueia e passa pela glote, que consiste num espaço na base da laringe entre duas cordas vocais, que são chamados também de pregas vocais, constituído por dobramentos elásticos da membrana mucosa que reveste a laringe (Sundberg, 1977).

Na glote, o ar vindo dos pulmões, é geralmente transformado em vibrações rápidas e audíveis, ocorrendo isso, devido ao fechamento das cordas vocais que causam um aumento gradativo da pressão atrás dela, que acaba por fazer com que elas se abram, liberando a pressão, para então tornarem a se fechar. Além da vibração das cordas vocais, o fluxo de ar pode ser constringido em algum ponto do trato vocal (Almeida, 2010).

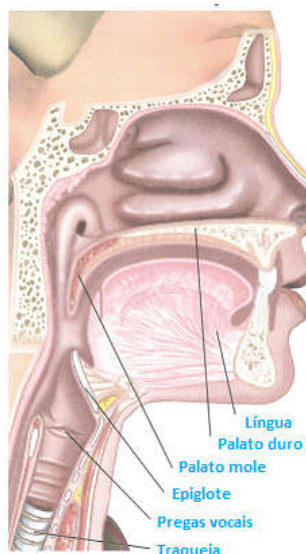


Figura 2.1: Trato Vocal (Almeida, 2012).

O trato vocal é constituído pela laringe, faringe, cavidades nasais e a boca. A forma do trato é determinado pelas posições dos articuladores: os lábios, a mandíbula, a língua e a laringe. Conforme dito anteriormente, o trato vocal funciona como uma caixa de ressonância, na qual transmite as parciais do som laríngeo dependendo das frequências de ressonância, sendo estas parcialmente modelados de forma diferente ao longo do espectro (Almeida, 2012; Sundberg, 1977).

Os sons produzidos podem ser classificados de duas formas, como vozeados ou não vozeados. Nos sons não vozeados, não ocorre vibrações das cordas vocais e o trato vocal recebe um ar turbulento. Enquanto, nos sons vozeados, se tem a vibração das cordas vocais. Existem ainda os sons chamados de mistos, compostos por parte vozeada e não vozeada (Uliani Neto et al., 2013).

2.3. MATLAB

O ambiente de computação do MATLAB é baseado em matrizes, ou seja, seus dados inseridos são armazenados sob a forma de uma matriz ou um *array* multidimensional, possuindo alta eficiência em aplicações de análise numérica e processamento de sinais (MathWorks, 2014a).

O *software* possui uma linguagem de programação de alto nível que serve como ferramenta para aplicações em diversos tópicos, como na teoria de controle, projetos de filtros, inteligência artificial e processamento de sinais, ou melhor, sua estrutura flexível promove um rápido desenvolvimento e análise (Lathi, 2006).

Devida a facilidade que o MATLAB possui em ter funções matemáticas previamente inseridas e um ambiente de criação gráfica que auxilia a projetar interfaces de usuário, o mesmo foi escolhido para a construção do *software* em estudo. O GUIDE (*Graphical User Interface Development Environment*) é uma ferramenta para a criação da interface gráfica para construção de *User Interfaces* (UI), essa interface gera um código em MATLAB que permite a alteração do programa de acordo com o comportamento desejado (MathWorks, 2014b).

2.4. Frequência Fundamental

Existem diversos parâmetros importantes na avaliação anatômica e funcional da voz, como a intensidade e a frequência fundamental da voz (Camargo, Barbosa & Teles, 2007). Entre os parâmetros relacionados à voz, a frequência fundamental (F0) é o que se mantém mais uniforme quando se considera diferentes sistemas de análise acústica e menos sensível às características de gravação de voz. As medidas de variação de frequência e amplitude de ciclo a ciclo, *jitter* e *shimmer* respectivamente, são parâmetros interessantes para serem analisados a questões relacionadas a patologias, por serem mais sensíveis à estas (Felippe, Grillo & Frechi, 2006).

A frequência fundamental corresponde a frequência do sinal de excitação proveniente da glote. Em outras palavras, é a frequência de vibração das cordas vocais por uma unidade de tempo (Almeida, 2012; Teixeira, Ferreira & Moreira, 2011). Dessa forma, a F0 depende diretamente das características morfológicas das cordas vocais e da laringe, como por exemplo, o comprimento das cordas vocais (Sundberg, 1991). A frequência fundamental é diretamente proporcional à tensão e inversamente à massa e a largura das cordas vocais (Andrade, 2003).

As características morfológicas das cordas vocais são diferentes tanto entre adultos do sexo feminino e masculino, quanto para crianças, sendo assim, a frequência fundamental é diferente para cada indivíduo (Sundberg, 1991). A variação de valor desta frequência é de aproximadamente de 80 aos 200 Hz para os homens, para as mulheres entre 200 e 300 Hz e para crianças cerca dos 400 aos 500 Hz. E essa variação nos valores dependem de outros fatores além dos morfológicos, como o período do dia, o estado nervoso, os hábitos de vida (álcool e tabaco), a entonação utilizada, o uso profissional da voz (voz falada e cantada) e à disфонia (Teixeira, 1995; Teixeira, Ferreira & Moreira, 2011)

Na voz cantada, a energia é superior à voz falada, e a frequência fundamental é maior para o primeiro caso (Almeida, 2012). As frequências na voz cantada podem se aproximar a 1024 Hz, no caso da voz feminina do tipo soprano (Vieira, 2005).

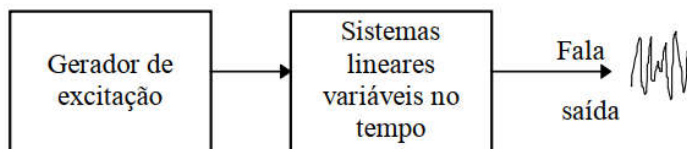


Figura 2.2: Modelo genérico de produção de fala (Teixeira, Ferreira & Moreira, 2011).

Existem vários algoritmos de determinação da frequência fundamental que permitem a extração desse parâmetro do sinal de fala, baseando-se no modelo genérico da produção da fala apresentado na Figura 2.2 (Teixeira, Ferreira & Moreira, 2011). Esses algoritmos são classificados em três categorias básicas, sendo determinação baseada no domínio do tempo, determinação baseada no domínio da frequência ou determinação baseada no domínio de tempo-frequência em conjunto (Upadhyal & Wankhede, 2012). Neste projeto é utilizado o método da autocorrelação para a determinação da frequência fundamental.

2.4.1. Método da Autocorrelação

Como o pós processamento envolve a frequência fundamental para ser projetado as ferramentas para a afinação vocal, a determinação da F0 é a base dos desenvolvimento dos programas que serão descritos no Capítulo 3:

Algoritmos de determinação de F0 utilizam técnicas de curto prazo. Um método normalmente utilizado para a extração da frequência fundamental é baseado em detetar o maior valor da função de autocorrelação na região de interesse (Tan & Karnjanadecha, 2003).

A autocorrelação necessita de uma janela de curta duração para ser calculada, sendo o problema encontrar a janela apropriada. O quadro ideal para análise deve conter no mínimo de dois a três períodos glotais completos. Assim, para vozes masculinas (tom grave), o quadro de análise deve ser longo, enquanto que para vozes femininas (tom agudo) deve ser curto (Pinto & Tomé, 2001).

Dado um sinal de tempo discreto $x(n)$, definido para todo n , a função de autocorrelação (R_x) que é geralmente definida pela Equação (2.1) (Tan & Karnjanadecha, 2003).

$$R_x(m) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n)x(n+m) \quad (2.1)$$

Onde N é o comprimento da sequência analisada. Para a determinação da frequência fundamental, assumindo que $x(n)$ é periódico com um período P , isto é, $x(n)=x(n+P)$. Observando assim que a autocorrelação é periódica para o mesmo intervalo, tendo assim a Equação (2.2) (Tan & Karnjanadecha, 2003; Upadhya1 & Wankhede, 2012).

$$R_x(m) = R_x(m+P) \quad (2.2)$$

Entretanto, para um sinal não estacionário como a fala, a Equação (2.1) não é a mais indicada. Conforme dito anteriormente, neste caso é interessante definir uma função de autocorrelação com segmentos curtos do sinal como mostra a Equação (2.3) (Tan & Karnjanadecha, 2003).

$$R_x(m) = \frac{1}{N} \sum_{n=0}^{N'-1} [x(n+l)w(n)][x(n+l+m)w(n+m)]^2, 0 \leq m \leq M_0 \quad (2.3)$$

Onde $w(n)$ é uma janela apropriada para a análise, podendo ser a janela de Hanning, Hamming ou Blackman. N é o comprimento da seção, N' o número de amostras de sinais usadas no cálculo de $R(m)$, M_0 é o número de pontos de autocorrelação a serem calculados e l é o índice da amostra inicial. E, N' é definido pela Equação (2.4) (Tan & Karnjanadecha, 2003; Teixeira, Ferreira & Moreira, 2011).

$$N' = N - m \quad (2.4)$$

Portanto, a autocorrelação representa a distribuição da magnitude espectral do sinal vocal num domínio de tempo, sendo a Transformação Inversa de Fourier de um espectro de energia desse sinal. Em termos práticos, a autocorrelação fornece uma medida das partes de fala semelhantes repetidas ao longo do sinal com um certo desalinhamento. Quanto maior o valor de autocorrelação, maior é a repetição de eventos semelhantes ao longo do sinal (Ferreira, 2012; Fernandes, 2018). Considerando que para realizar a aquisição em tempo real seja necessário no mínimo dois períodos glotais, e assumindo que a F_0 mínima de 50 Hz. O segmento para realizar o cálculo da autocorrelação corresponde a $2 * \frac{1}{50} = 40$ ms.

2.4.2. Código da Autocorrelação

A frequência fundamental é calculada por meio do método da autocorrelação. Sua implementação tem como base a frequência de amostragem que está sendo feita a aquisição dos dados e na determinação do tamanho do segmento. Neste projeto utiliza a frequência de amostragem de 22050 Hz e o segmento, conforme citado anteriormente, do tamanho de 40 ms. A Figura 2.3 ilustra o resultado de um segundo da vogal ‘/a/’ falada por uma voz feminina e o segmento de 40 ms que será analisado.

Após determinado esses parâmetros é realizado a autocorrelação desse segmento, como mostra a Figura 2.4. O cálculo da F_0 é feito procurando identificar a cada intervalo de N_1 o valor mais alto da função de autocorrelação desse segmento. Neste caso, foi identificado o T_0 de 4,58 ms e utilizando a Equação (2.5), a frequência fundamental encontrada foi de 218,32 Hz.

$$F_0 = \frac{1}{T_0} \quad (2.5)$$

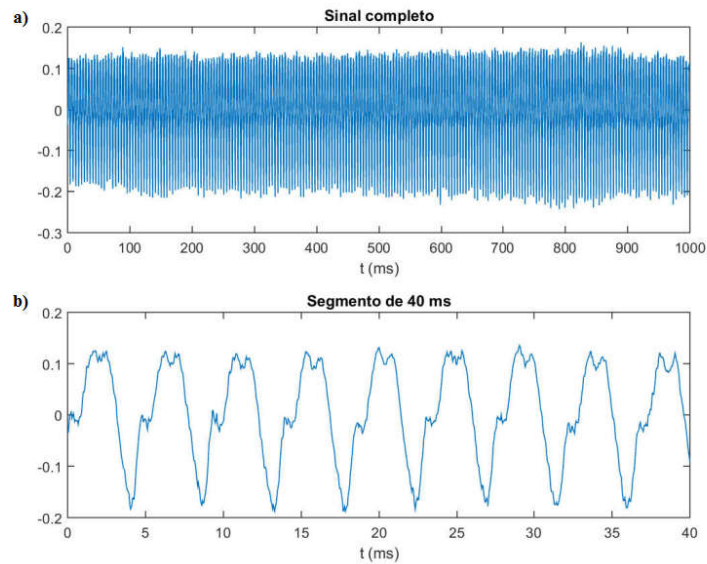


Figura 2.3: Parte 1 do método da autocorrelação (a) Sinal completo (b) Segmento de 40 ms.

A periodicidade da forma de onda, demonstrada na Figura 2.4 do sinal e também a diferença uniforme de tempo decorrido entre os picos da função de autocorrelação explica o fato de o sinal de entrada ser vozeado, pois um sinal não vozeado não possui uma periodicidade como no sinal vozeado (Upadhya & Wankhede, 2012)

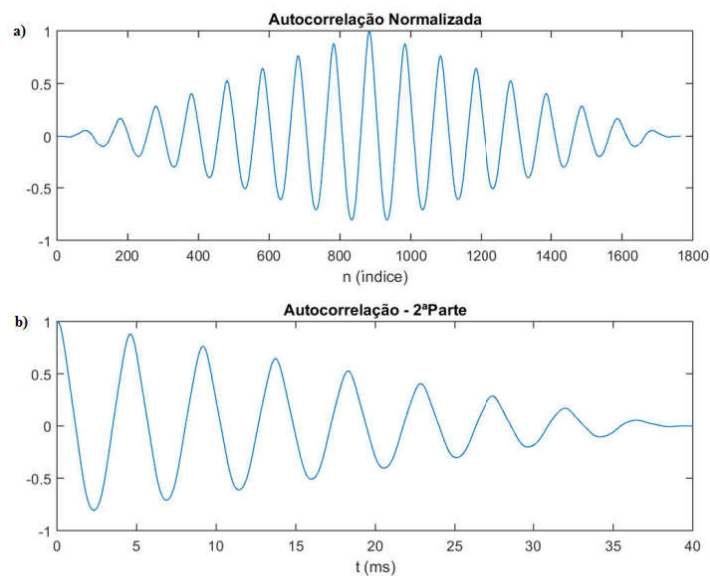


Figura 2.4: Parte 2 do método da autocorrelação (a) Autocorrelação do segmento de 40 ms (b) Busca pelo maior valor da autocorrelação.

A janela de pesquisa N1 é importante avaliar o seu tamanho por determinar o valor máximo da frequência fundamental calculada, por determinar onde se deve iniciar a procura do pico da autocorrelação. A Equação (2.6) apresenta o cálculo desse valor máximo e a Tabela 2.1 exibe a frequência máxima de acordo com quatro valores de N1 aplicando a seguinte equação.

$$F0 \leq \frac{1}{N1} \quad (2.6)$$

Tabela 2.1: Frequência Máxima de acordo com o valor de N1.

Valor de N1	Frequência Máxima de F0
1 ms	1000 Hz
2 ms	500 Hz
3 ms	333 Hz
5 ms	200 Hz

A Figura 2.5 exibe a comparação dos quatro valores de N1 indicados na Tabela 2.1. Como o valor da F0 do sinal da Figura 2.4 é por volta dos 218 Hz, os primeiros valores de N1, 1 ms, 2 ms e a 3 ms, deram o mesmo resultado por estarem abaixo do valor máximo a ser identificado. Enquanto para N1 igual a 5 ms tem o limite abaixo da frequência identificada, por esse motivo, o valor de frequência fundamental calculado foi por volta dos 109,5 Hz, sendo um valor incorreto.

Realizando o teste com um som reproduzido aproximadamente a 170 Hz com os quatro valores de N1, deram o mesmo valor em todos os casos de 170,3 Hz. Por ser um valor abaixo do limite é possível observar que os quatro conseguiram identificar o valor correto. Outro teste realizado, alterando o som produzido para aproximadamente 500 Hz com os mesmos valores de N1 dos casos anteriores, apenas para os valores de 1 ms e 2 ms que reconheceram o valor correto de 501,1 Hz e enquanto para 3 ms o valor encontrado foi de 250,5 Hz e para 5 ms de 167 Hz.

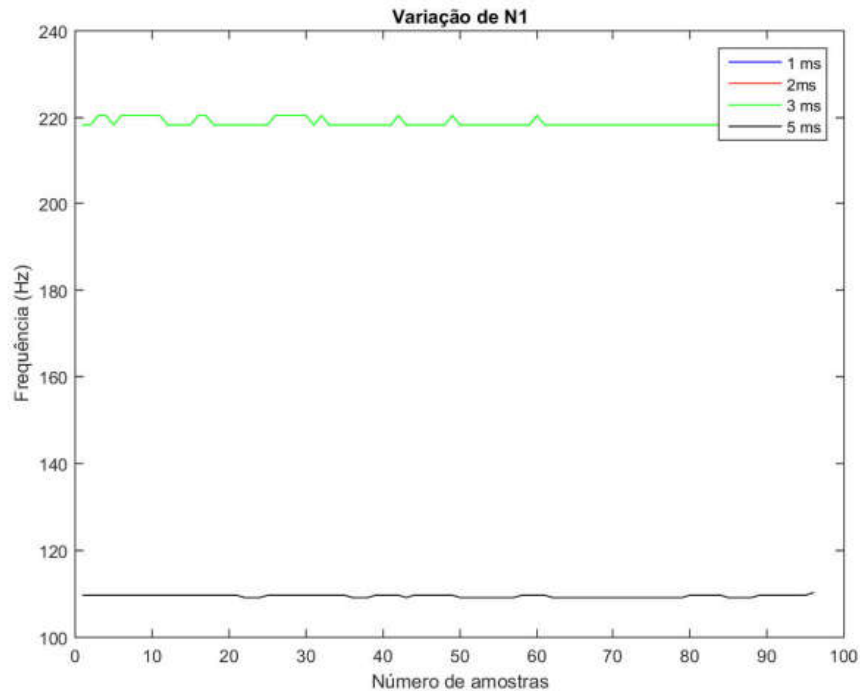


Figura 2.5: Variando o valor de N1.

Portanto, ao realizar todos esses testes observa-se que N1 não pode ser um valor alto e desejando um valor limite de F0 alto, para o caso deste projeto o mais indicado é utilizar o valor de 1 ms para que consiga englobar os valores que são observados na extensão vocal que será discutida na sessão 2.8.2.

2.5. Aquisição do Sinal

2.5.1. Frequência de Amostragem

Um sinal $x(n)$ no tempo discreto consiste em amostras de um sinal $x_a(nT)$ no tempo contínuo, como é mostrado na Equação (2.7). Em outras palavras, é necessário recolher amostras de um sinal contínuo em um determinado intervalo de tempo, sendo chamado de período de amostragem (T_s). E o inverso de T_s é denominado frequência de amostragem (F_s), ou seja, $1/T_s$ amostras por segundo (Diniz, Silva & Netto, 2014).

$$x(n) = x_a(nT) \quad (2.7)$$

A fim de que o sinal seja reconstruído sem erro ou com um erro dentro de uma dada tolerância, é importante manter a taxa de amostragem do sinal suficientemente alta. Para que isso seja possível, é utilizado o Teorema da Amostragem que diz que a F_s tem que ser maior que o dobro da largura de faixa unilateral do sinal no tempo. A frequência mínima para amostrar um sinal conforme descrito no Teorema da Amostragem é chamada de Frequência de Nyquist, como mostra a Equação (2.8) (Oppenheim & Schauer, 2009).

$$F_N = 2F_c \quad (2.8)$$

De acordo com Oppenheim e Schafer (2009), a extrapolação da frequência de amostragem não afeta a capacidade de processamento de sinal, melhorando a qualidade das amostras. Posto isto, ao realizar um mapeamento do sinal de F0, que possui frequências de voz elevadas, é necessário mais amostras, dado que o mesmo número de amostras para frequências mais altas provoca resultados distantes do real. Em vista disso, a frequência de amostragem escolhida para realizar a aquisição do sinal foi de 22050 Hz.

A frequência de amostragem pode influenciar no cálculo da frequência fundamental, ao buscar o valor mais alto da função de autocorrelação. O cálculo da F0 é feito com base na F_s dividido pelo valor do índice é identificado o pico da autocorrelação, como mostra a Equação (2.9).

$$F_0 = \frac{F_s}{I_{máx}} \quad (2.9)$$

Por ser um sinal discreto, ao identificar o pico da autocorrelação é possível ter uma variação de uma amostra pra mais ou menos do valor. Considerando essa variação, o cálculo da F0 seria como é apresentado pela Equação (2.10).

$$F_0 = \frac{F_s}{I_{máx} \pm 1} \quad (2.10)$$

Para avaliar essa influência foram analisados três valores de F_s sendo 11025 Hz, 22050 Hz e 44100 Hz. Essa variação não afeta quando está em frequências baixas, no entanto, em frequências altas essa variação é maior como é possível observar na Figura 2.6.

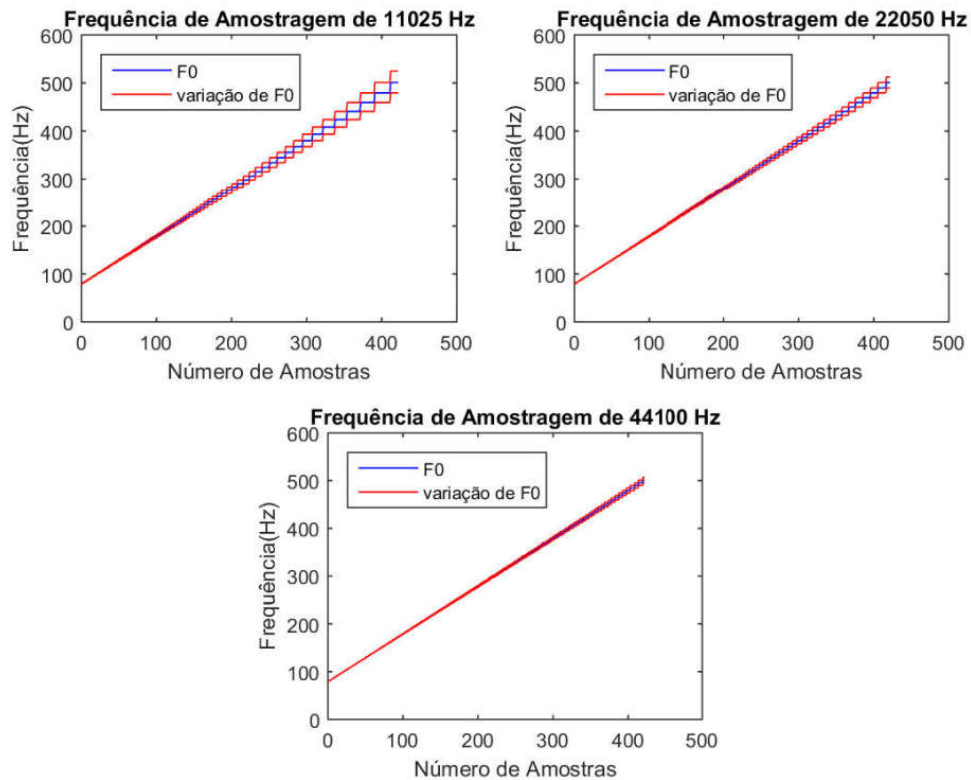


Figura 2.6: Intervalo de variação de F0 de acordo com a frequência de amostragem.

Aumentando a frequência de amostragem utilizada, se tem um número maior de amostras, portanto, ao realizar a reconstrução do sinal, é possível que se tenha uma variação de mais ou menos uma amostra do pico real ao realizar a autocorrelação. A Figura 2.7, demonstra a diferença entre o F0 calculado e o F0 considerando a variação de mais ou menos uma amostra. Conforme aumenta a F0, se tem mesmo aumento com o erro. O número de amostras de um sinal cresce quando é utilizado uma F_s maior, para o mesmo valor de F0 se tem uma menor variação.

Para este trabalho, conforme citado anteriormente, a frequência de amostragem utilizada foi de 22050, apesar do erro ser menor utilizando a frequência de amostragem de 44100, se teria um processamento muito maior do qual, neste caso, não é necessário.

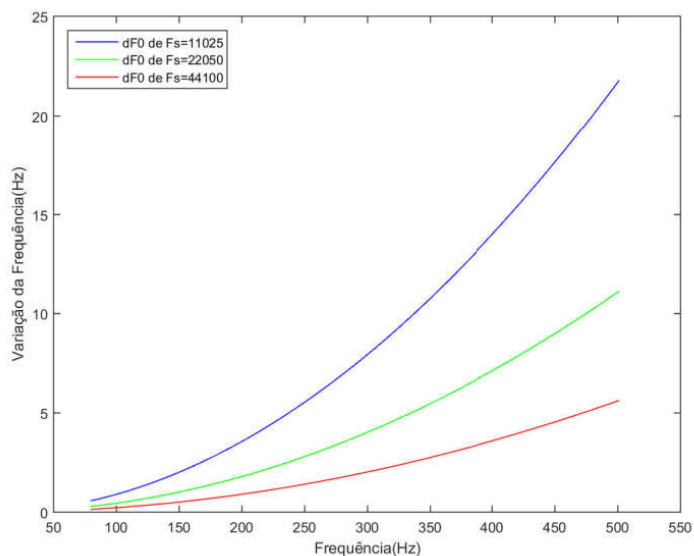


Figura 2.7: deltaF0 de acordo com a frequência de amostragem.

2.5.2. Microfone

O microfone é um dispositivo que converte sinais acústicos em sinais elétricos (Teixeira, 2019). O microfone utilizado neste trabalho foi o Blue Snowball, mostrado na Figura 2.8. Sua conexão é simplificada por ser *Universal Serial Bus* (USB) que alimenta e permite a comunicação dos dados.



Figura 2.8: Microfone Blue Snowball.

Esse microfone possui três configurações. A primeira configuração ocorre a captação direcional, isto é, o som diretamente para o microfone, reduzindo o ruído ambiente. A segunda, capta também o som direcional mas contém mais ruído ambiente. E a última configuração grava ruído omnidirecional, contendo grande parte do som em volta do microfone, sendo ideal para gravar conversas (Blue Design, 2019). Neste trabalho foi utilizado a primeira configuração. Esse dispositivo possui uma resposta em frequência que varia de 40 Hz a 18 kHz.

2.6. Aquisição do Sinal no MATLAB 64 bits

A aquisição em tempo real de voz pode ser realizada de duas maneiras: através do *AudioRecorder* ou *Data Acquisition Toolbox*, ambos os métodos realizam a aquisição através da placa de som.

2.6.1. AudioRecorder

Cria-se um objeto do tipo *AudioRecorder* que vai adquirir os dados usando diretamente o dispositivo de áudio do computador. Antes de realizar a aquisição é necessário que seja feita a configuração de alguns parâmetros. É preciso definir a frequência de amostragem, o número de bits e o número de canais de áudio que será utilizado para realizar a aquisição que neste caso, foi sucessivamente escolhido 22050 Hz, 16 bits e 1 canal. Como deseja-se que a cada 40 ms seja realizada a autocorrelação, foi configurado o *TimerPeriod* da função para esse tempo e chamando a função para que ocorra a aquisição e que seja possível realizar a função de autocorrelação da janela desse período de aquisição. A variável *drt* indica o tempo total que será feita a aquisição, variando para cada aplicação.

```
recorder = audiorecorder(fs, 16, 1);
set(recorder, 'TimerPeriod', timerPer, 'TimerFcn', @VADML);
recordblocking(recorder, drt);

function VADML(recorder, ~)
    pause(0.1)
    y = getaudiodata(recorder, 'single')
end
```

Figura 2.9: Configuração do objeto *AudioRecorder*.

A função *getaudiodata*, realiza a aquisição do tipo *single*, com uma resolução de 32bits (2^{32}), salvando todos os valores num *array*, isto é, esse método armazena em *buffer* os dados de um quadro de dados, como mostra a Figura 2.10, sendo necessário limpar o *buffer* ao final da aquisição.

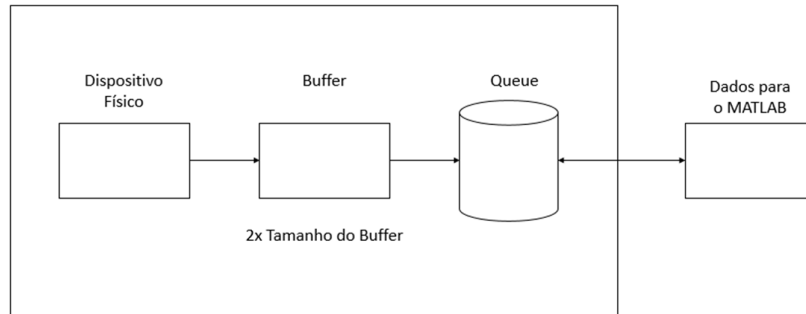


Figura 2.10: Funcionamento da aquisição da função *getaudiodata* (adaptado de MathWorks, 2019a).

A partir do vetor y (*queue*) demonstrado na Figura 2.9, são extraídas as 882 amostras da aquisição que equivale ao segmento de 40 ms desejado para encontrar a frequência fundamental. Esse valor de amostras foi calculado pelo valor inteiro da Equação (2.11), sendo a frequência de amostragem e o *timerPer* os mesmos já citados anteriormente.

$$SampleSize = Fs * timerPer \quad (2.11)$$

O método *AudioRecorder*, armazena em *buffer* toda a aquisição realizada, isto é, a cada aquisição ele concatena os novos dados ao vetor. Por conta disso, se tem um gasto maior de memória por ser necessário armazenar o sinal inteiro, isto é, se o *buffer* for pequeno não é possível realizar toda a aquisição, e assim, terminando o programa. Para exemplificar melhor o funcionamento da aquisição é mostrado a Figura 2.11, onde a cada aquisição ocorre a verificação do *buffer*. Se o tempo de processamento na determinação da F0 e na representação do gráfico da F0 em tempo real for um tempo maior que os 40 ms, ocorrerá um atraso e não será processado todo o sinal inteiro, logo terá um atraso entre os valores de F0 mostrados e o sinal adquirido.

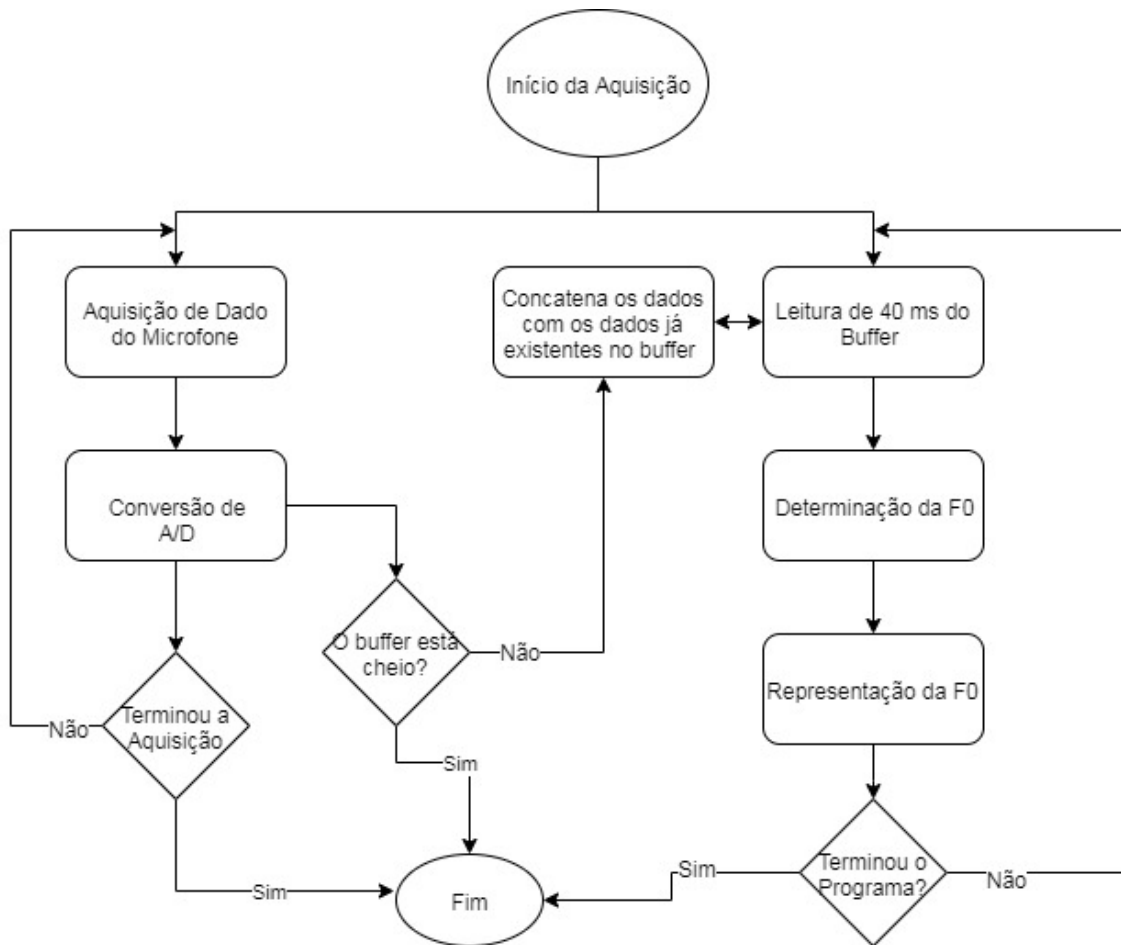


Figura 2.11: Fluxograma da aquisição e processamento do método *AudioRecorder*.

2.6.2. Data Acquisition Toolbox

A *toolbox* auxilia no processo de aquisição realizado no MATLAB, providenciando as ferramentas e os recursos necessários para isso. Neste projeto a *toolbox* utilizada foi a *Direct Sound*, que possibilita: adquirir os dados de vários canais de entrada de áudio, gerir os dados de áudio para vários canais de saída de áudio, configurar a taxa de amostragem da placa de som e outras configurações, além disso analisa os dados da placa de som à medida que ela é adquirida, aciona o início da aquisição de dados e muda facilmente placas de som com alterações mínimas no código.

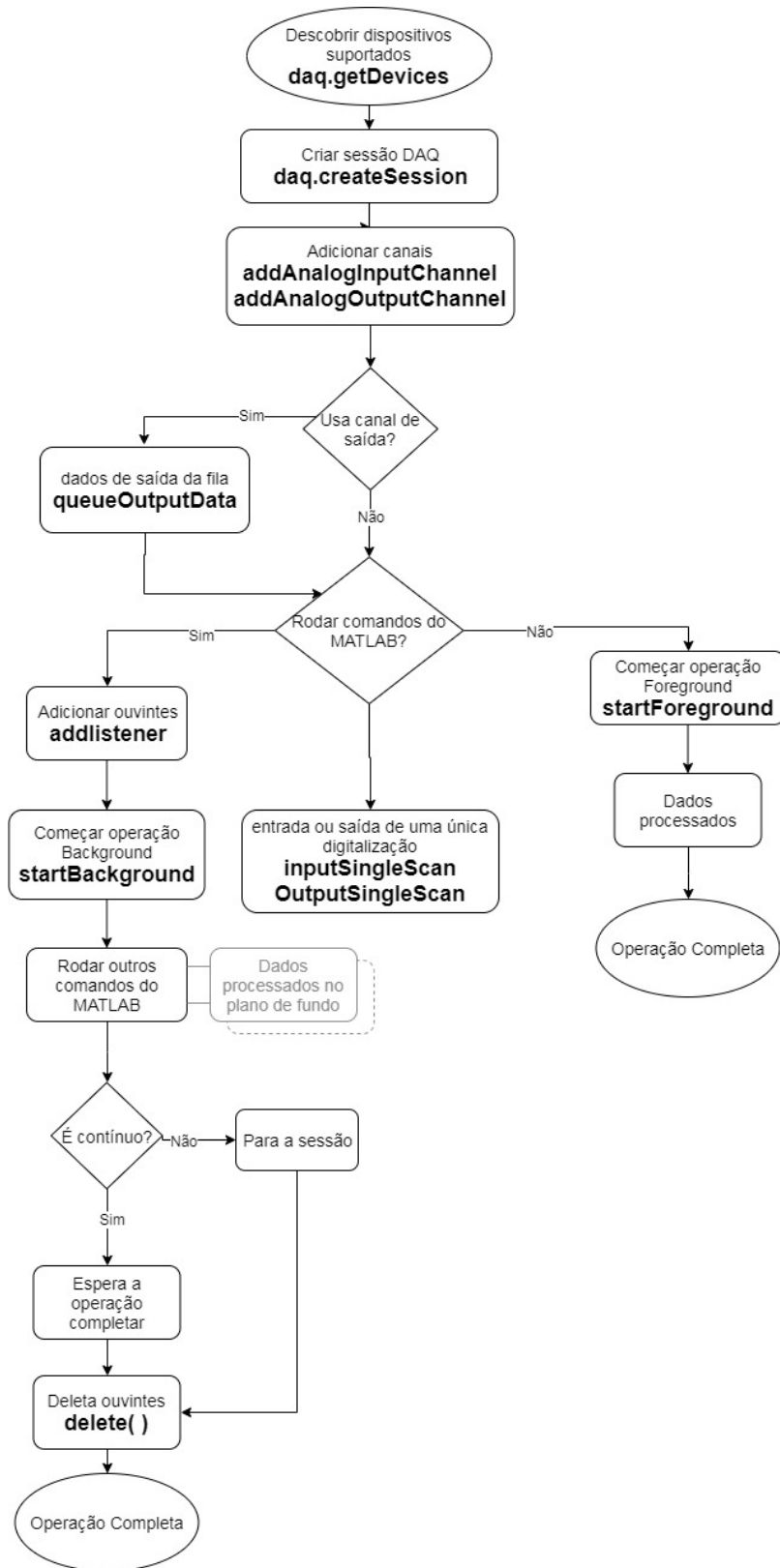


Figura 2.12: Fluxograma da configuração da aquisição com a *Toolbox* (adaptado de MathWorks, 2019b).

A base de configuração da *Data Acquisition Toolbox* no MATLAB segue o fluxograma demonstrado na Figura 2.12. Inicialmente, verifica-se os dispositivos disponíveis para realizar a aquisição utilizando o *daq.getDevices* e é necessário criar uma sessão com o canal de entrada de áudio para acessar a placa de som, que neste caso é o dispositivo *Audio1* que se refere ao microfone.

```
d=daq.getDevices;
s = daq.createSession('directsound');
addAudioInputChannel(s, 'Audio1', '1', 'Audio');
```

Figura 2.13: Configuração da *Toolbox*.

Antes de inicializar a aquisição, é feita a configuração de alguns parâmetros de acordo com a aplicação desejada. A frequência de amostragem e o *sampleSize* são os mesmos utilizados no método do *AudioRecorder*, de 22050 Hz e 882 amostras respectivamente e a duração (*drt*) varia de acordo com a aplicação.

```
s.Rate = fs;
s.DurationInSeconds = drt;
s.NotifyWhenDataAvailableExceeds=sampleSize;
```

Figura 2.14: Configuração dos parâmetros da *Toolbox*.

A função *addlistener* cria um evento que inicializa a entrada do canal com a placa de som, começando a aquisição de dados. O vetor data possui as 882 amostras da janela dos 40 ms desejados para poder realizar a função da autocorrelação.

```
lh = addlistener(s, 'DataAvailable', @plotData);
startBackground(s);

function plotData(src, event)
    x=event.TimeStamps;
end
```

Figura 2.15: Início da aquisição utilizando a *Toolbox*.

A aquisição funciona de duas formas, a primeira forma é *startBackground* que inicia a operação do objeto na sessão sem bloquear o MATLAB. A outra forma é *startForeground* que bloqueia o MATLAB naquela linha de código. Quando utilizado *DataAvailable*, é verificado se está ou não liberado para a entrega dos dados, assim sendo disparado periodicamente durante o processo.

Diferente do método *AudioRecorder*, ao utilizar a *toolbox*, não se tem armazenamento do sinal inteiro no *buffer*, mantendo apenas a janela desejada para análise, em outras palavras, ao sinal ser processado, ele é deletado, mantendo apenas o que não foi processado. Isso faz com que o gasto de memória seja menor. Outro ponto interessante, a quantidade de amostras é sempre constante, independente do processamento do computador. Entretanto, se ocorrer um atraso no processamento da determinação da F0 e da representação do gráfico da F0 em tempo real, isto é, se o tempo de resposta for maior que os 40 ms, o sistema faz o processamento todo atrasado.

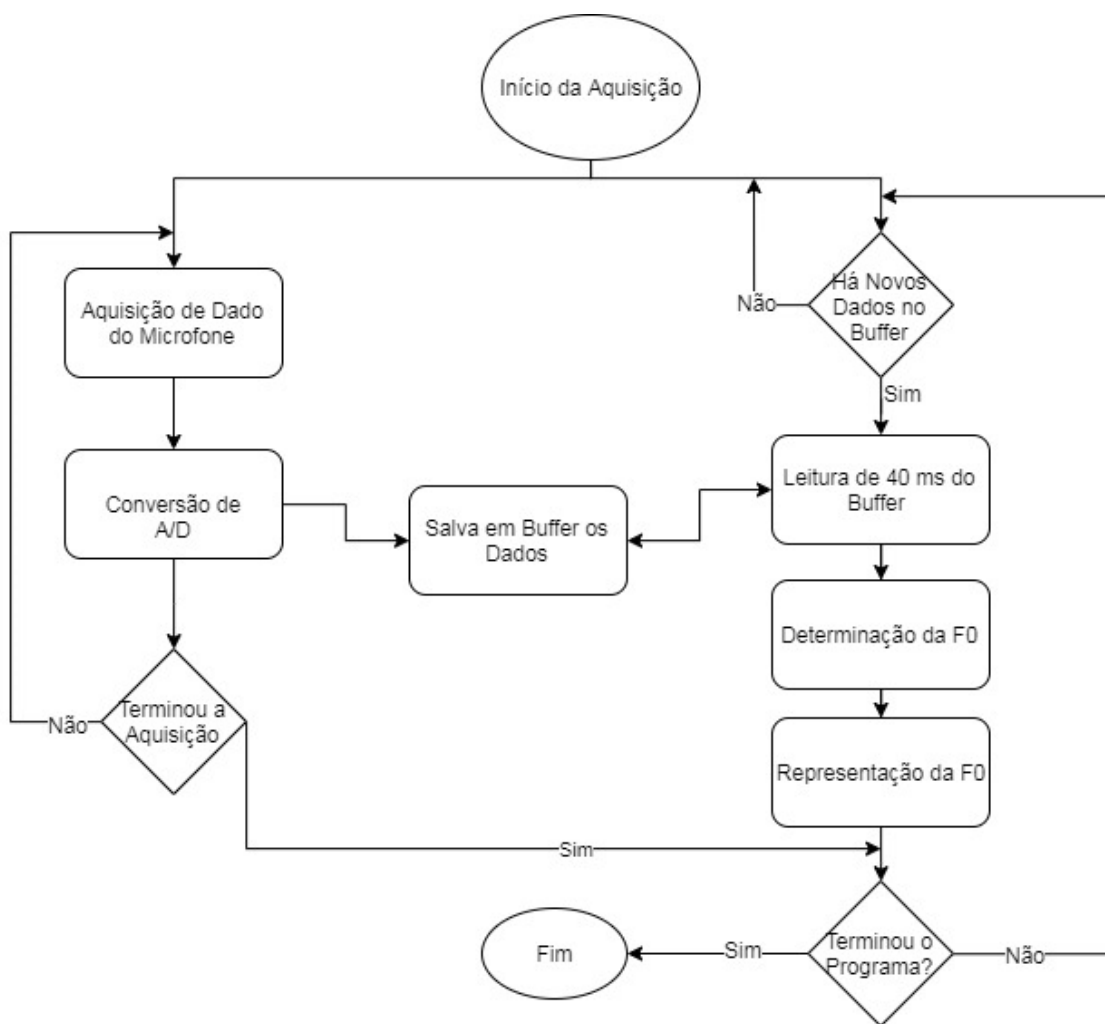


Figura 2.16: Fluxograma da aquisição e processamento do método da *Toolbox*.

Comparando os dois métodos de aquisição num computador com processador Intel Core i7-7500U (CPU 2.70 GHz, 2904 MHz, 2 núcleos e 4 processadores lógicos) e determinando F0 e apresentando o gráfico do sinal, foi observado que uma aquisição de 10 s teve um tempo de

resposta muito curto, o mesmo ocorreu com a de 20 s para ambos os métodos. Entretanto, ao realizar a aquisição de 40 s, ocorreu um leve atraso ao final da mesma, sendo esse atraso mais visível com o método do *AudioRecorder*. Ao implementar o processamento para as ferramentas de afinação vocal, foi observado que o atraso se tornou mais visível com a aquisição acima dos 25 s, isto é, o tempo de resposta se tornou alto, principalmente realizando a aquisição com o *AudioRecorder*.

2.7. Pós Processamento

A fim de evitar anomalias na aquisição da frequência fundamental é realizada a detecção de silêncio em uma janela de amostras vocais antes de ser realizada a função de autocorrelação (Leão, 2018). Assim como um sinal de voz é uma onda, há uma característica importante que auxilia nessa análise. Uma onda transporta energia, sendo assim uma forma de realizar a detecção de silêncio do sinal de voz é a partir de sua energia (Aleixo & Mansanares, 2003).

Dessa forma, é possível analisar a área de baixo de um sinal de voz levando em consideração sua amplitude e sua duração. Entretanto, as áreas negativas podem se cancelar com as áreas positivas, indicando um sinal de tamanho pequeno. Dessa maneira, para corrigir isso, a energia do sinal pode ser calculada como a área debaixo de $[x(t)]^2$, no qual será sempre positiva (Lathi, 2006).

$$Ex = \int_{-\infty}^{\infty} [x(t)]^2 dt \quad (2.12)$$

Se não houver energia na janela do sinal $x(t)$, refere-se a uma zona de silêncio. Desse modo, não tendo a necessidade de realizar a função de autocorrelação para encontrar a frequência fundamental. O código que refere-se a esse procedimento está demonstrado na Figura 2.17.

```
if E<1
    FO_0=0;
else
    FO_1=FO_autocorrelation(data, fs);
    if FO_1-f0ant> 400
        FO_0=f0ant;
    else
        FO_0=FO_1;
    end
end
f0ant=FO_0;
```

Figura 2.17: Função no MATLAB na detecção de zona de silêncio.

Após realizado a Equação (2.12), é verificado se o valor encontrado da energia é menor que 1, caso seja verdade é verificado uma zona de silêncio e a frequência encontrada é zero, caso contrário, é realizado a função da autocorrelação da janela do sinal indicado com a frequência de amostragem de 22050 Hz.

Para evitar sons não vozeados ou picos indesejados é analisado se o valor da autocorrelação é próximo do anterior, caso tenha uma alta variação de F0, o valor de F0 anterior é mantido. Caso contrário, o novo valor é assumido.

2.8. Teoria de Música

O bom desempenho de um coro requer coesão do grupo, equilibrando harmoniosamente as vozes pertencendo a todos os registros vocais. Na voz cantada, utiliza-se a mesma estrutura para a produção da voz falada, entretanto, com alguns ajustes devido às necessidades do canto (Camargo, Barbosa & Teles, 2007).

Ao elaborar a ligação do *software* com a teoria de música, foram reunidas literaturas necessárias para o conhecimento da escala musical, nas extensões vocais para realizar a classificação de forma correta e escolher um método de atenuação da nota para uma reprodução da melodia escrita.

2.8.1. Escala de Música

As escalas musicais são a divisão da sequência de notas contidas dentro de uma oitava, essa divisão é feita em sete notas principais: tom – tom – semitom – tom – tom – tom – semitom.

A divisão é realizada obedecendo a critérios estéticos, ou seja, na melodia que as notas formam e as relações harmônicas entre elas (Aleixo & Mansanares, 2003).

A relação harmônica entre as notas pode ser analisada utilizando a série de Fourier, uma vez que as combinações de formas sonoras cujas frequências não possuem interferência entre si dentro de uma ou duas oitavas (Leão, 2018).

Alguns intervalos foram arranjados para que se tornassem iguais a $9/8$, ou seja, um intervalo de um tom. Esses intervalos são entre as notas Dó - Ré, Fá - Sol e Lá - Si. Já os intervalos de Ré - Mi e Sol - Lá são iguais $10/9$, sendo ligeiramente inferior a um tom. A diferença relativa entre $9/8$ e $10/9$ é imperceptível. O intervalo entre Mi - Fá e Si - Dó ($16/15$) é ligeiramente superior a um semitom. (Aleixo & Mansanares, 2003).

DÓ	RE	MI	FÁ	SOL	LÁ	SI	DÓ
f	$9f/8$	$5f/4$	$4f/3$	$3f/2$	$5f/3$	$15f/8$	$2f$
$9/8$	$10/9$	$16/15$	$9/8$	$10/9$	$9/8$	$16/15$	

Figura 2.18: Intervalo de um tom (Aleixo e Mansanares, 2003).

Para resolver o problema das notas intermediárias foram criadas duas escalas, na escala justa foi introduzido uma alteração nas notas, sendo colocado os sustenidos e os bemóis, formados da multiplicação ou divisão da nota original por $25/24$. A multiplicação cria uma nota sustenida (#) e a divisão cria uma nota bemol (b). Dessa maneira, a escala passava de sete notas para vinte e uma notas e o intervalo entre duas notas sucessivas nunca excede $25/24$, assim sendo perceptível aos ouvidos mais treinados.

Outra escala criada a partir da escala natural para resolver o problema das notas intermediárias foi a escala temperada, em que foi desenvolvida a partir da divisão da afinação, distribuindo o erro causado entre as notas vizinhas. Isto é, dividiu-se a oitava em 12 intervalos iguais à raiz duodécima de 2, desse modo, o Dó sustenido se iguala ao Ré bemol, o Mi sustenido ao Fá e o Fá bemol ao Mi. Portanto, na escala temperada reduz as 21 notas iniciais para 12 (Aleixo & Mansanares, 2003).

A Tabela 2.2 compara as duas escalas, mostrando assim a diferença entre as frequências e as relações entre nota e cada escala. Sendo capaz de observar que os intervalos existentes na escala justa foram eliminados na escala temperada para simplificação.

Tabela 2.2: Relação entre as notas musicais, intervalos e frequências correspondentes (Aleixo & Mansanares, 2003).

Nota musical	Intervalo com a nota fundamental	Afinação natural	Frequência (Hz)	Afinação temperada	Frequência (Hz)
Dó	Dó uníssono	1/1	132.000	1.000	132.000
Dó #	Semitom	25/24	137,544	1,059	139.788
Ré b	Segunda diminuta	27/25	142.560	1,059	139.788
Ré	Segunda maior	9/8	148.500	1.122	148.104
Ré #	Segunda aumentada	76/64	154.704	1.189	156.948
Mi b	Terça menor	6/5	158.400	1.189	156.948
Mi	Terça maior	5/4	165.000	1.260	166.320
Fá b	Quarta diminuta	32/25	168.960	1.260	166.320
Mi #	Terça aumentada	125/96	171.864	1.335	176.220
Fá	Quarta perfeita	4/3	175.956	1.335	176.220
Fá #	Quarta aumentada	25/18	183.348	1.414	186.648
Sol b	Quinta diminuta	36/25	190.080	1.414	186.648
Sol	Quinta perfeita	3/2	198.000	1.498	197.736
Sol #	Quinta aumentada	25/16	206.316	1.587	209.484
La b	Sexta menor	8/5	211.200	1.587	209.484
Lá	Sexta maior	5/3	220.044	1.682	222.024
Lá #	Sexta aumentada	152/72	229.284	1.782	235.224
Si b	Sétima menor	9/5	237.600	1.782	235.224
Si	Sétima maior	15/8	247.500	1.888	249.216
Dó b	Oitava diminuta	48/25	253.440	1.888	249.216
Si #	Sétima aumentada	125/64	257.796	2.000	264.000
Dó	Oitava perfeita	2/1	264.000	2.000	264.000

2.8.2. Classificação Vocal

Vários critérios são levados em consideração na classificação vocal como a estrutura corporal, características anatômicas da laringe, aspectos e a personalidade do falante. Conforme visto anteriormente, esses critérios influenciam diretamente na frequência fundamental de cada indivíduo (Camargo, Barbosa & Teles, 2007).

A intensidade e a frequência fundamental da voz são parâmetros importantes na avaliação anatômica e funcional da voz. Esses dois parâmetros correlacionados possibilitam por meio da fonetografia o registro do perfil de extensão vocal, no qual é um dos principais fatores que são considerados para realizar a classificação vocal (Camargo, Barbosa & Teles, 2007; Costa et al, 2006).

A extensão vocal é denominada como o número de notas que um indivíduo pode emitir, sendo da mais grave à mais aguda, não importando a qualidade ou o esforço para atingi-las. Os sons mais graves relacionam-se com as baixas frequências e os agudos com as altas (Aleixo & Mansanares, 2003; Costa et al, 2006). A extensão vocal, geralmente é expressa em semitons (st), analisada pela diferença existente entre frequência máxima e mínima (Camargo, Barbosa & Teles, 2007).

Ao realizar a conversão de unidades de Hertz (Hz) para semitons (st) foi utilizada a Equação (2.13). Nesta equação, f representa a frequência e r ao valor referencial de 16,352 Hz dado de acordo com Coleman, Mabis e Hinson (1977). Como os semitom é uma medida linear e apresenta um intervalo regular em escala musical de 12 semitons, seus valores tem que ser arredondado para o valor de inteiro mais próximo, sendo representado na equação seguinte como $round()$ (Camargo, Barbosa & Teles, 2007).

$$st = round\left(\log_{\sqrt[12]{2}}\left(\frac{f}{r}\right)\right) \quad (2.13)$$

As vozes femininas possuem uma maior frequência que as vozes masculinas. Habitualmente, os homens possuem a corda vocal mais longa e, enquanto as mulheres, a corda é mais curta (Nadoleczny, 1923). De acordo com Camargo, Barbosa e Teles (2007), as vozes masculinas tem valores médios de frequência por volta dos 120,6 Hz, e as femininas aproximadamente 212,8 Hz.

No canto coral, as vozes são classificadas de acordo com o *range* de frequência e cada uma possui um alcance vocal específico. As vozes femininas no coral podem ser classificadas como Soprano, Mezzo-soprano e Contralto, enquanto as masculinas podem ser Tenor, Barítono e Baixo (Camargo, Barbosa & Teles, 2007).

A Tabela 2.3 apresenta o intervalo em frequência, o intervalo visto em semitons e o intervalo de notas, isto é, a extensão vocal por três visões diferentes para cada tipo de voz.

Tabela 2.3: Extensão vocal de acordo com a classificação (Perelló et al., 1982; Junker, 1999).

	Classificação Vocal	Extensão Vocal (Hz)	Extensão Vocal (st)	Extensão Vocal (notas)
Masculino	Baixo	97 a 147	31 a 38	Sol2 (G2) a Ré3 (D3)
	Barítono	108 a 165	33 a 40	Lá2 (A2) a Mi3 (E3)
	Tenor	65 a 220	24 a 45	Dó2 (C2) a Lá3 (A3)
Feminino	Contralto	98 a 262	31 a 48	Sol2 (G2) a Dó4 (C4)
	Mezzo-soprano	110 a 330	33 a 52	Lá2 (A2) a Mi4 (E4)
	Soprano	130 a 440	36 a 57	Dó3 (C3) a Lá4 (A4)

Entretanto, segundo Camargo, Barbosa e Teles, 2007, ao estudar um grupo de canto, as vozes dentro desse grupo variava. As vozes Soprano apresentaram uma faixa de 36 st (130 Hz) a 72 st (1084 Hz), as vozes Mezzo-soprano de 35 st (124 Hz) a 68 st (878 Hz), as vozes Tenor de 28 st (85 Hz) a 65 st (709 Hz), as vozes Barítono de 26 st (74 Hz) a 59 st (524 Hz) e as vozes Baixo de 26 st (75 Hz) a 59 st (513 Hz).



Figura 2.19: Pauta Geral para as faixas vocais. Na linha superior as vozes femininas e na linha inferior as vozes masculinas (Science Buddies Staff, 2018).

As seis classificações são as mais conhecidas, no entanto, existem mais de três classificações para cada sexo. De acordo com Perelló et al (1982), cada autor apresenta uma extensão vocal para cada uma das seis classificações e para as demais classes que existem. A Figura 2.19 apresenta outra extensão de notas que é possível ser utilizada.

2.9. Avaliação da Performance

O resultado final de cada treino é avaliado de acordo com a comparação entre o vetor de notas para treinamento e o vetor cantado pelo usuário. As distâncias entre as curvas de F0 servem para avaliar o desempenho e são descritas a seguir.

2.9.1. Erro Quadrático Médio

O Erro Quadrático Médio, em inglês *Mean Square Error* (MSE), é definido como sendo a média da diferença entre o valor estimado (X) e o parâmetro (\hat{X}) ao quadrado, como é apresentado na Equação (2.14).

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2 \quad (2.14)$$

O MSE apresenta a média de um conjunto de erros ao calcular a distância entre cada ponto do valor estimado e do parâmetro, e ajusta o erro através da distância entre as duas linhas analisadas. Ao elevar ao quadrado ocorre a remoção de sinais negativos e cada erro influencia o total na proporção de seu quadrado, ao invés de sua magnitude, aumentando o peso de erros maiores (Wackerly, Mendenhall & Scheaffer, 2014).

2.9.2. Raiz do Erro Quadrático Médio

A Raiz do Erro Quadrático Médio, em inglês *Root Mean Square Error* (RMSE), envolve uma sequência de 3 etapas simples para ser calculada. A primeira etapa é o erro quadrático total, obtido por meio da soma dos erros quadráticos individuais. Na seguinte etapa, o erro quadrático total é dividido por N, isto é, gerando o MSE. A última etapa para calcular o RMSE, é realizar a raiz quadrada do MSE (Willmott & Matsuura, 2005).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2} \quad (2.15)$$

A Equação (2.15) apresenta o RMSE, no qual X é o valor estimado e o \hat{X} é o parâmetro. No MSE e no RMSE, os grandes erros tem uma influência maior do que os menores. Em outras palavras, ambos os erros aumentam consideravelmente à medida que essas grandes variações são maiores.

2.9.3. Erro Absoluto Médio

O Erro Absoluto Médio, em inglês *Mean Absolute Error* (MAE), é a medida mais natural para o cálculo de erro de magnitude. O cálculo do MAE é relativamente mais simples que os citados anteriormente. Envolve a soma das magnitudes do erro para obter o erro total, ou seja, é a soma do valor absoluto da diferença entre o valor estimado e do parâmetro. Após encontrado o erro esse valor é dividido por N (Willmott & Matsuura, 2005). A Equação (2.16) representa o cálculo do MAE

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_i - \hat{X}_i| \quad (2.16)$$

Enquanto o RMSE e o MSE que penalizam as maiores variações, dando peso diferente aos erros, o MAE atribui o mesmo peso a todos os erros. (Chai & Draxler, 2014).

2.9.4. Comparando MSE, RMSE e MAE

Para comparar as três medidas para determinar a distância entre as curvas de F0 alvo e a produzida, é baseado em variações visualizados no programa de Treino de Escalas que será discutido na sessão 3.3. Deixando o vetor F0 alvo fixo e variando o vetor F0 produzido para posteriormente compará-los.

Ao realizar o deslocamento horizontal, é verificado o aparecimento de uma área causado pelo deslocamento de 3 amostras entre o estimado e o parâmetro, como é demonstrado pela Figura 2.20. Neste caso, foi calculado MAE de 1,35 Hz, o MSE de 6,75 Hz² e o RMSE de 2,60 Hz. É possível observar que o peso para o erro é diferente.

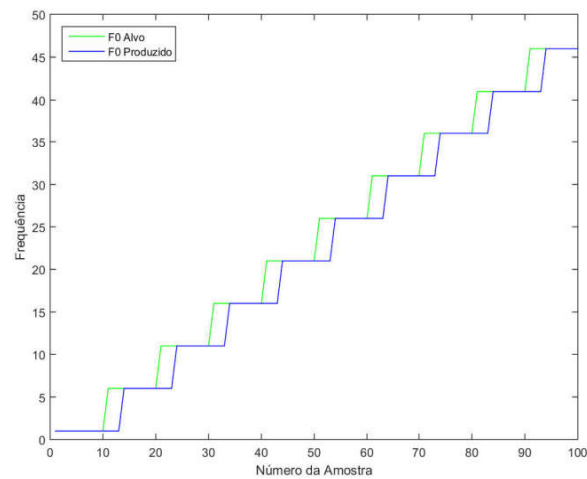


Figura 2.20: Caso de deslocamento horizontal.

Essa análise é interessante pois o usuário ao cantar pode ter um atraso na variação da nota, assim como a resposta do programa pode influenciar nesse atraso, com isso, é possível observar que com o MAE esse erro é muito pequeno, não aumentando o peso para essa variação.

Entretanto, o caso mais interessante para a análise é quando ocorrem alguns picos nas transições, por ser um exemplo mais próximo do real, apresentada pela Figura 2.21. Ao avaliar o cada erro nesse exemplo, o MAE é de 0,84 Hz, para o erro MSE foi de 4,3 Hz² e RMSE de 2,0745 Hz. Sendo assim, é possível observar que os picos durante a transição de notas usando o MAE não compromete significativamente a avaliação do usuário, pois diferentemente dos MSE e do RMSE que dão peso maior para esses picos. Logo, o MAE minimiza essas perturbações que normalmente não são causadas por desafinação.

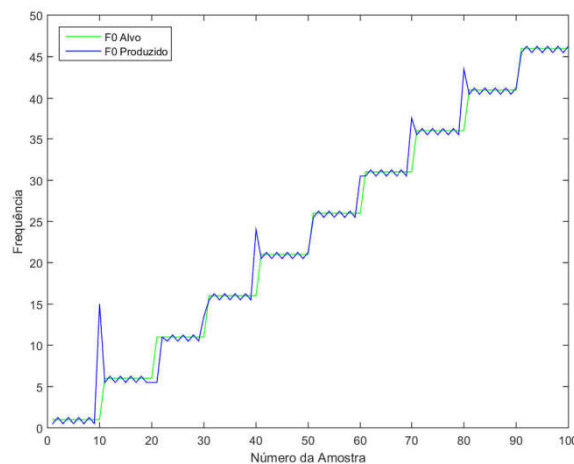


Figura 2.21: Caso de picos e pequenas variações.

Nos casos a analisar é possível observar que o MAE é sempre menor ou igual ao RMSE e por ser elevado ao quadrado a diferença. O MSE é sempre um valor muito maior do que os anteriores. Os dois erros que se mostraram interessantes para este projeto foram o MAE e o RMSE, entretanto, devido à diferença de peso para o tipo de erro que o RMSE apresenta, é utilizado para analisar o desempenho do usuário o MAE por considerar todos os erros com o mesmo peso.

Capítulo 3: Desenvolvimento

Equation Section (Next)

Neste capítulo é descrito o funcionamento e a implementação dos programas que envolvem o *software*. Os quatro programas possuem ferramentas que auxiliaram na afinação vocal, essas ferramentas são para classificar a voz do usuário, para treinar a afinação por meio de uma escala, para treinar utilizando músicas e o compositor de música. Esses quatro programas estão conectados por meio do menu principal do *software*.

3.1. Menu Principal

O menu principal está disposto com todas as ferramentas disponíveis para serem acedidas, apresentando os botões relacionados as quatro aplicações propostas, como mostra a Figura 3.1. Essas ferramentas envolvem a “Classificação Vocal”, o “Treino de Escalas”, o “Treino de Músicas” e um “Compositor de Música”. Ao selecionar a ferramenta, abrirá diretamente o programa desejado.



Figura 3.1: Menu principal do *software*.

3.2. Classificação Vocal

Esta primeira ferramenta busca informar ao usuário qual a sua classificação vocal e os parâmetros extraídos em tempo real, sendo a nota musical, a frequência fundamental em Hz da nota produzida e a escala. As decisões do programa são demonstradas no fluxograma da Figura 3.2.

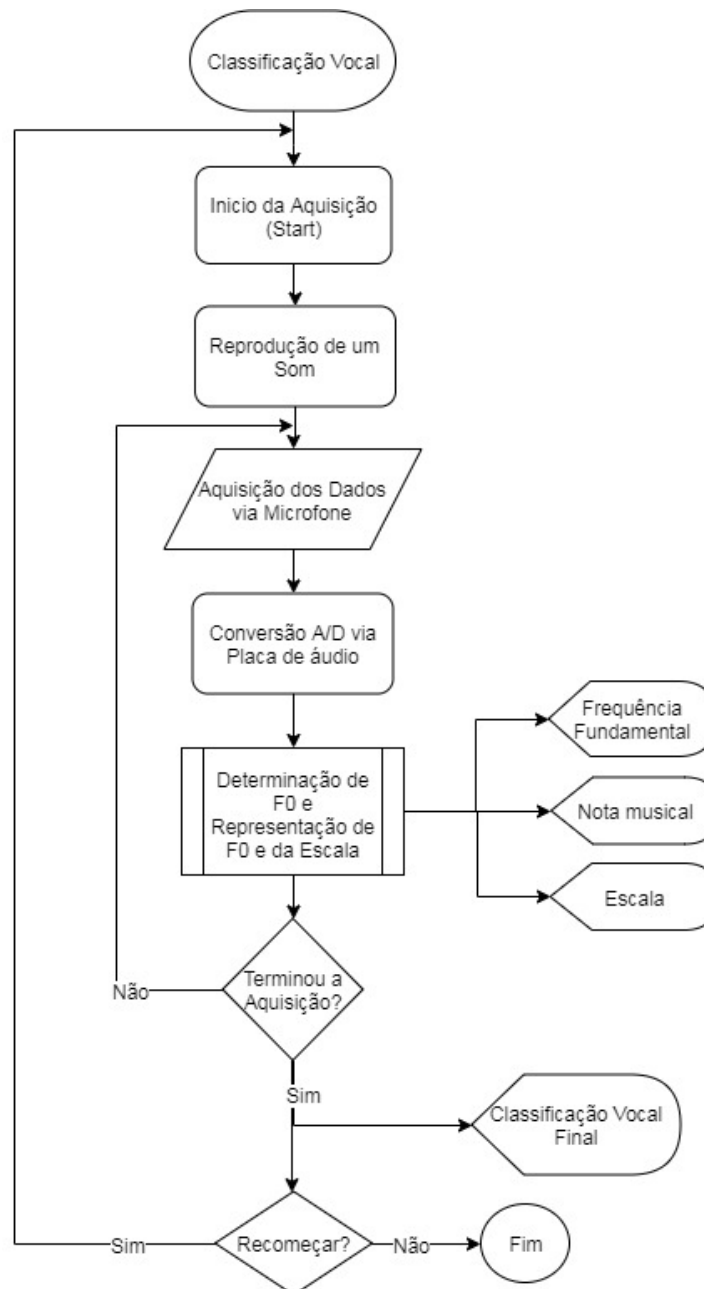


Figura 3.2: Fluxograma do funcionamento do programa “Classificação Vocal”.

Para realizar a classificação, o usuário deve por cinco segundos cantar de forma confortável e preferencialmente variando a voz do mais grave ao mais agudo, para que seja possível identificar qual a extensão vocal que o usuário consegue alcançar. Como mostra a Figura 3.2, é feita a aquisição da voz e são calculados os parâmetros (frequência fundamental cantada, a nota e a escala) que são atualizados a cada novo ciclo. Ao fim da aquisição é apresentada a classificação vocal e o usuário pode decidir se deseja realizar o procedimento novamente, sair do programa ou retornar o menu principal. Identificar a classificação vocal auxilia para que o usuário consiga adequar os treinos para a afinação de forma correta, sem estar a usar escalas inacessíveis ou inadequadas à sua extensão vocal. Por esse motivo, ao utilizar essa ferramenta é preciso ser cuidadoso para que não seja feita uma classificação incorreta.

A cada ciclo de 40 ms é determinada a frequência fundamental desse segmento e, é verificada em qual escala de 1 a 5 à qual pertence a frequência detectada. Por último, a nota é determinada com base na escala encontrada no passo anterior e que sua classificação esteja de acordo com a Tabela 2.2 proposta por Aleixo e Mansanares (2003). A nota é classificada utilizando a escala temperada de notas. Os dados são atualizados durante todo o processo de aquisição. Ao final, é avaliado o vetor das frequências fundamentais calculadas durante todo o processamento de dados, sua média é analisada para que não se pegue valores mínimos incorretos e sua frequência máxima alcançada. Utilizando a Tabela 2.3 como base e verificando em qual das classificações mencionadas o usuário melhor se encaixa. A classificação é apresentada numa janela, demonstrado na Figura 3.3.

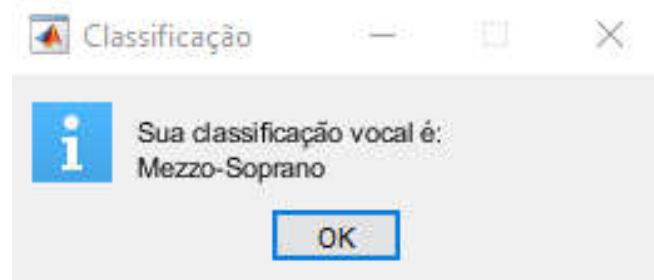


Figura 3.3: Janela com a classificação final.

A interface gráfica do programa na qual o usuário terá acesso é apresentada na Figura 3.4, podendo observar em tempo real a variação de todos os parâmetros citados anteriormente. O botão de “*Start*” inicia a coleta dos dados com a reprodução de um som.



Figura 3.4: Interface gráfica do programa “Classificação Vocal”.

Por ser uma ferramenta de apresentação ao usuário, é necessário que tenha as informações de um tutorial do funcionamento do programa, para que ocorra a aquisição de forma correta. Portanto, esse tutorial está descrito no botão de informação, apresentando as, como demonstrado na Figura 3.5.

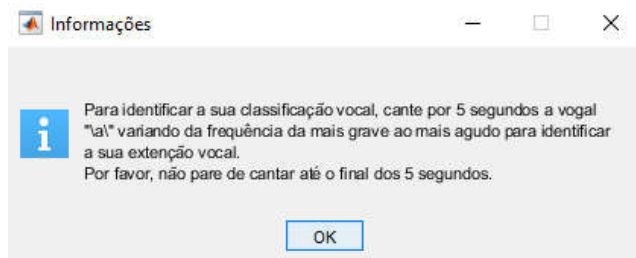


Figura 3.5: Janela de informação do programa “Classificação Vocal”.

3.3. Treino de Escala

Esta aplicação é para auxiliar no processo de afinação, partindo do princípio que o usuário utilizou o programa anterior para obter a estimativa da sua classificação vocal ou que já a conheça, com o intuito de treinar com a sequência mais conveniente ao seu tipo vocal. A Figura 3.6 ilustra o funcionamento desse programa.

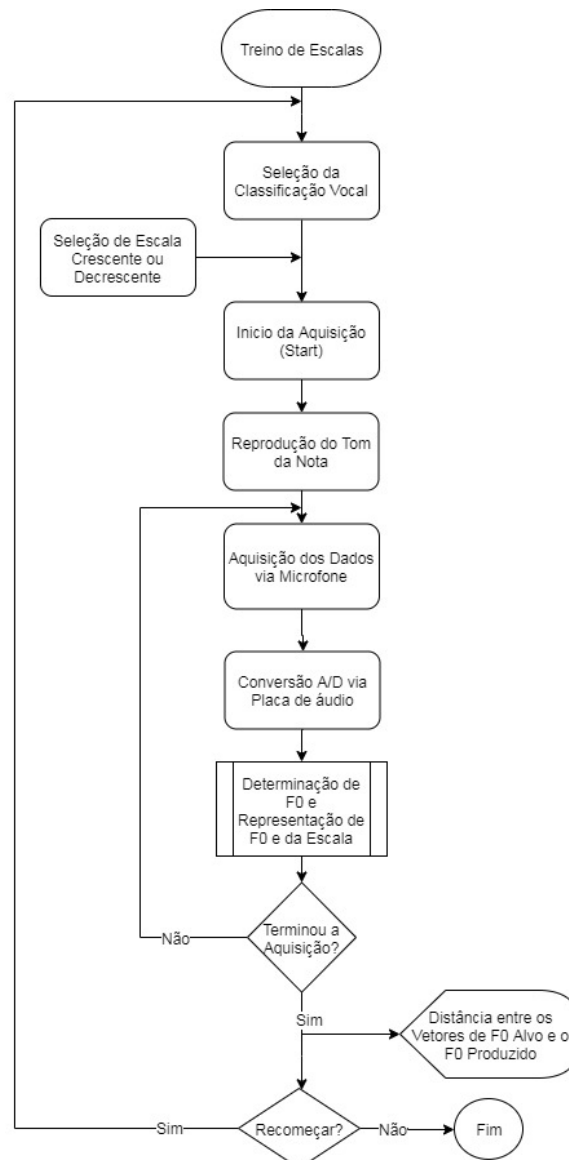


Figura 3.6: Fluxograma do funcionamento do programa “Treino de Escalas”.

A implementação desse programa ocorre selecionando inicialmente qual a classificação vocal, essa seleção é realizada ao clicar em um dos seis botões referentes ao tipo vocal, Figura 3.7, ao selecioná-lo. Uma variável escolhe quais as notas que serão treinadas. Essa sequência pode ser treinada de duas formas diferentes, sendo a frequência crescente ou decrescente, que também pode ser selecionado por outros dois botões. Quando não se opta por nenhuma ordem, por padrão é colocado a forma crescente como a preferencial.

Após escolhida a classificação e a ordem na qual pretende-se realizar o treinamento, uma sequência de notas é selecionada. A Tabela 3.1 apresenta uma sequência de notas

direcionada a cada uma das seis classificações, a sequência apresentada está na ordem crescente.

Tabela 3.1: Escala para treinamento de afinação vocal.

Classificação Vocal	Sequência de Notas
Baixo	Fá2-Sol2-Lá2-Si2-Dó3-Ré3-Mi3
Barítono	Lá2-Si2-Dó3-Ré3-Mi3-Fá3-Sol3
Tenor	Dó3-Ré3-Mi3-Fá3-Sol3-Lá3-Si3
Contralto	Mi3-Fá3-Sol3-Lá3-Si3-Dó4-Ré4
Mezzo-soprano	Fá3-Sol3-Lá3-Si3-Dó4-Ré4-Mi4
Soprano	Dó4-Ré4-Mi4-Fá4-Sol4-Lá4-Si4

O botão de “Start” inicia a aquisição para o treinamento, sendo reproduzido um som com o tom da nota inicial, para indicar o início dessa aquisição. Com esse programa, é possível acompanhar a frequência que o usuário deveria atingir e a que está ocorrendo em tempo real, apresentando ao usuário a captação da sua F0, como mostra a Figura 3.7. As notas que devem ser atingidas estão apresentadas na parte inferior do programa. Ao final da aquisição, a diferença medida pelo MAE entre as notas esperadas e as notas cantadas é apresentado.

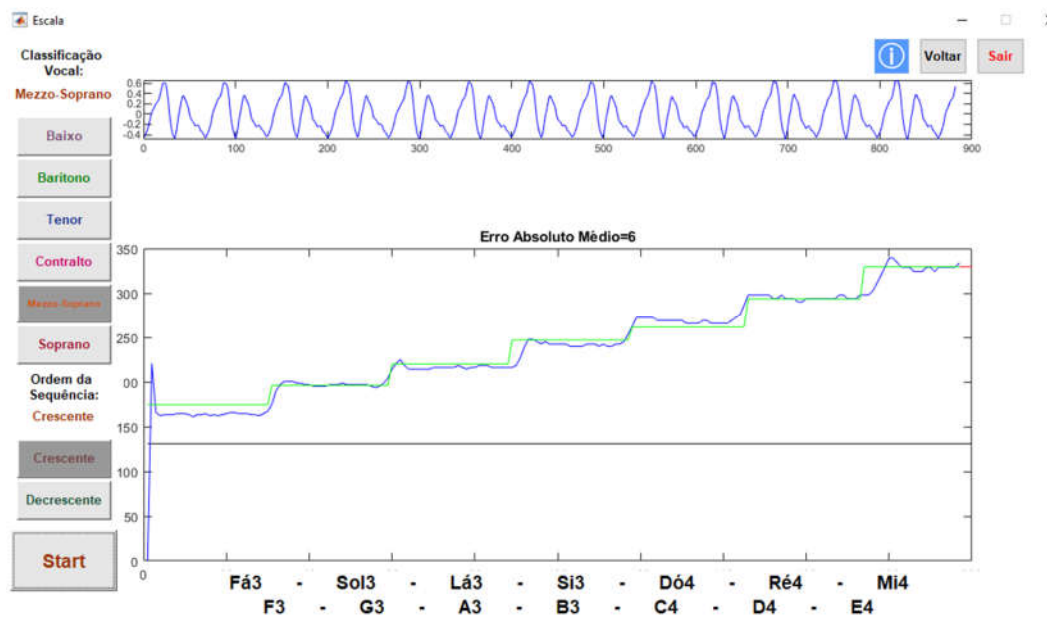


Figura 3.7: Interface gráfica do programa “Treino de Escalas”.

O botão de informação do funcionamento dessa ferramenta é apresentado tal como mostra a Figura 3.8.

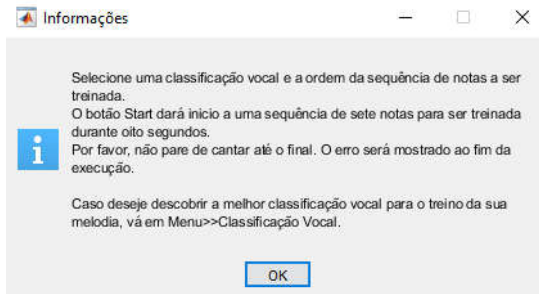


Figura 3.8: Janela de informação do programa de “Treino de Escalas”.

3.4. Treino de Músicas

Neste programa foram implementadas músicas para realizar treinos mais avançados que o de escalas. Esta aplicação procura disponibilizar uma pequena base de dados de músicas previamente escritas, divididas pelo idioma e por níveis.

O programa possibilita selecionar as músicas de três formas diferentes para que aquela que for selecionada esteja de acordo com a desejada pelo usuário. A música pode ser escolhida pelo idioma, apresentando músicas brasileiras, espanholas e inglesas, dado que cada idioma tem nove músicas, sendo três para o nível fácil, três para o nível intermediário e três para o nível avançado. Essa divisão de idiomas apresentadas em dois *pop-up* menu, no qual um tem o idioma, e no outro, as músicas do idioma selecionado, e os níveis divididos em botões, como mostra a Figura 3.9.

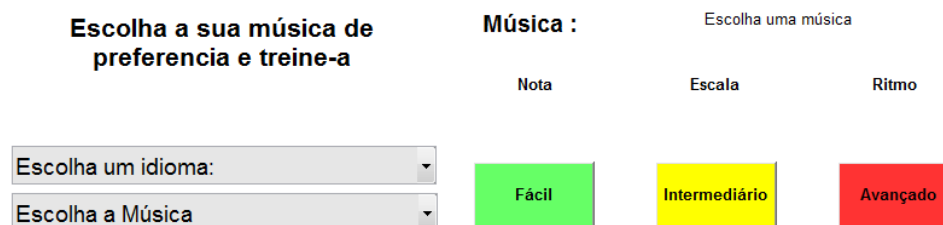


Figura 3.9: Seleção de música por idioma ou nível.

A Figura 3.10 expõe a paleta de botões que realizam a transposição da música, com a intenção de ajustar o semitom e a escala de uma forma que seja favorável ao nível, gênero, idade e tipo vocal do usuário. Além dos botões citados anteriormente, há também um que

aumenta ou diminui o ritmo da música com a finalidade de dificultar ou facilitar o treino de acordo com a intenção didática do usuário. O semitom, a escala e o ritmo podem ser variados, respeitando os limites existentes de cada parâmetro, exibindo mensagens de avisos ou erros quando ultrapassados esses limites. Essas variações no ritmo, na escala e no semitom é chamado de transposição de música.



Figura 3.10: Paleta de transposição e de ajuste de ritmo.

O programa possui a possibilidade de escutar a música no formato monofônico para que o usuário possa observar como ela é. O som é gerado via um sinal senoidal com frequência da nota, para que o som não apresente a variação é realizado o uso da janela de Hanning com sobreposição de 50%, assim se tem um sinal aproximadamente contínuo para ser reproduzido. O som da música fica de acordo com os ajustes feitos pelo usuário, reproduzindo o trecho que será treinado.

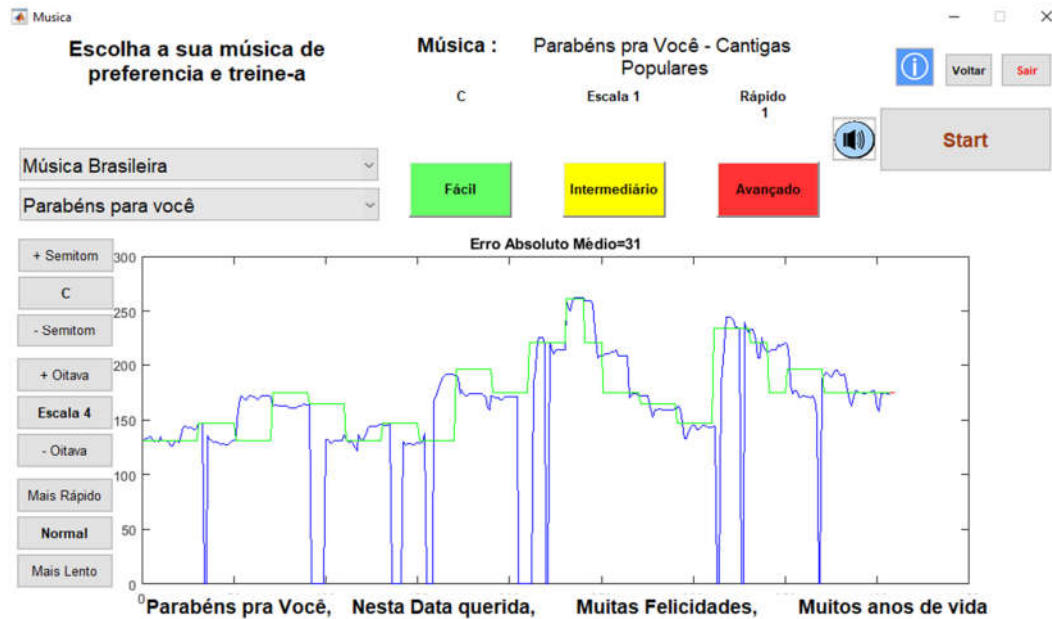


Figura 3.11: Interface gráfica do programa de “Treino de Músicas”.

A interface gráfica é demonstrada na Figura 3.11, apresentando a paleta de transposição, os botões de nível e os *pop up* para a seleção da música, o nome e a letra da mesma. O tempo de aquisição máximo é de 20 s, entretanto, dependendo se for selecionado um ritmo mais rápido esse tempo diminui e mais notas são possíveis de serem treinadas. A Figura 3.12 ilustra no formato de fluxograma o funcionamento do programa.

O programa inicia escolhendo a música desejada conforme descrito anteriormente, e a sequência de notas é determinada de acordo com um seletor que contém todas as notas referentes a cada música. A sequência de notas depende do ritmo, do semitom e da escala escolhida, caso não seja feita nenhuma alteração, se tem como padrão o ritmo normal, o tom C e a escala 1.

A aquisição é iniciada pelo botão “*Start*”, um som com a primeira nota da música é reproduzido para que o usuário possa identificar como iniciar o treino. O gráfico apresenta a sequência de notas que compõe a música e em tempo real o usuário consegue verificar de forma visual a frequência real, comparando assim com a frequência desejada. Ao final da aquisição é apresentado o erro, resultante da comparação do vetor da frequência fundamental produzida com o da música escolhida.

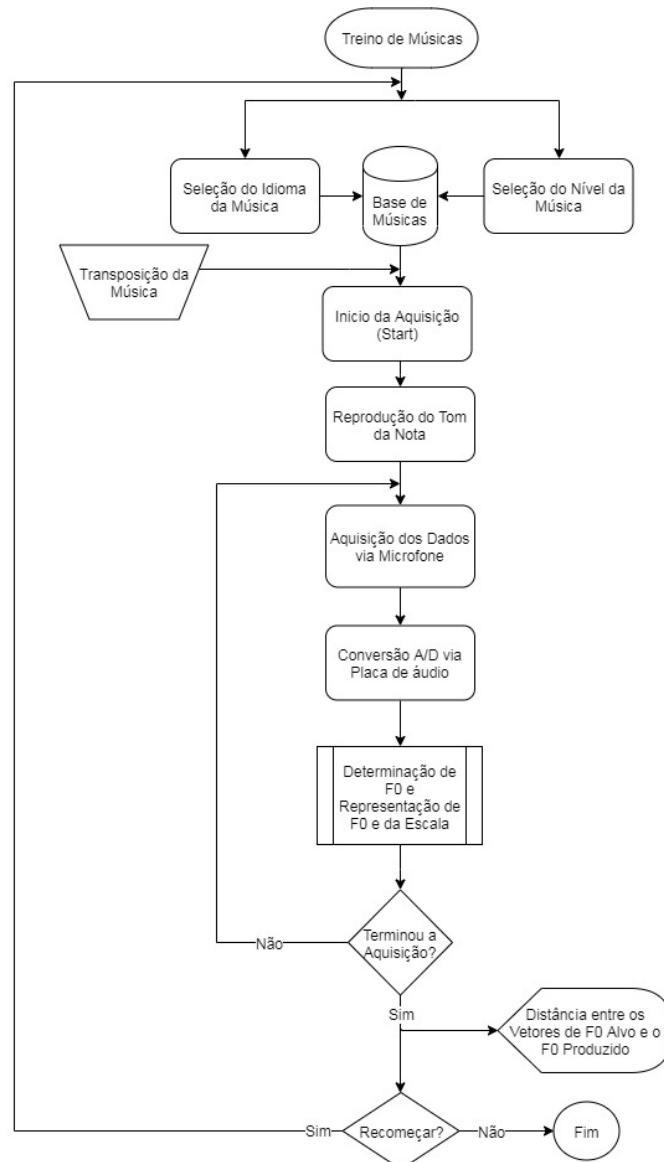


Figura 3.12: Fluxograma do funcionamento do programa “Treino de Músicas”.

3.5. Compositor de Música

Essa ferramenta é para usuários com um conhecimento em música mais avançado, que saiba como escrever uma música utilizando as notas da escala temperada. A escrita da música é realizada pela paleta apresentada na Figura 3.13, a escala e a nota é escolhida com os seus respectivos botões de aumentar e diminuir, podendo visualizar qual a nota e a escala que será inserida na música.

Insira o Nome da
Música:

Escolha a
escala:

1

< >

Escolha a
nota:

C

< >

INSERIR

delete

Limpar

Figura 3.13: Paleta de configuração e implementação de música.

A criação do vetor com as notas desejadas ocorre com o botão de “inserir”, esse concatena as notas conforme são inseridas e apresentando-as na parte inferior do programa. Caso tenha sido introduzida uma nota de forma incorreta, o botão “delete” exclui a última nota inserida na música. O botão “limpar” exclui todo o vetor de notas criado, podendo assim recomeçar a música. Cada nota inserida possui um tempo pré determinado, sendo de 0,48 s.

O nome inserido é colocado na caixa de texto apresentado no canto esquerdo superior. A escrita pode ser feita como o usuário desejar, ao ser salvo os espaços escritos é trocado por um “_” (*underline*) e o formato é o “.mat”. Um *array* é salvo contendo as notas e o nome escrito na caixa de texto.

A música ao ser escrita pode ser escutada pelo botão de ouvir ou pode ser treinada de forma similar ao programa “Treino de Músicas” apertando o botão “Start”. Entretanto, não é possível realizar as alterações com no programa anterior, conforme a música foi implementada ela será treinada, não podendo realizar mudanças em seu ritmo ou escala. O treinamento é feito

como mostra a Figura 3.14 que demonstra a interface gráfica do programa “Compositor de Música”.

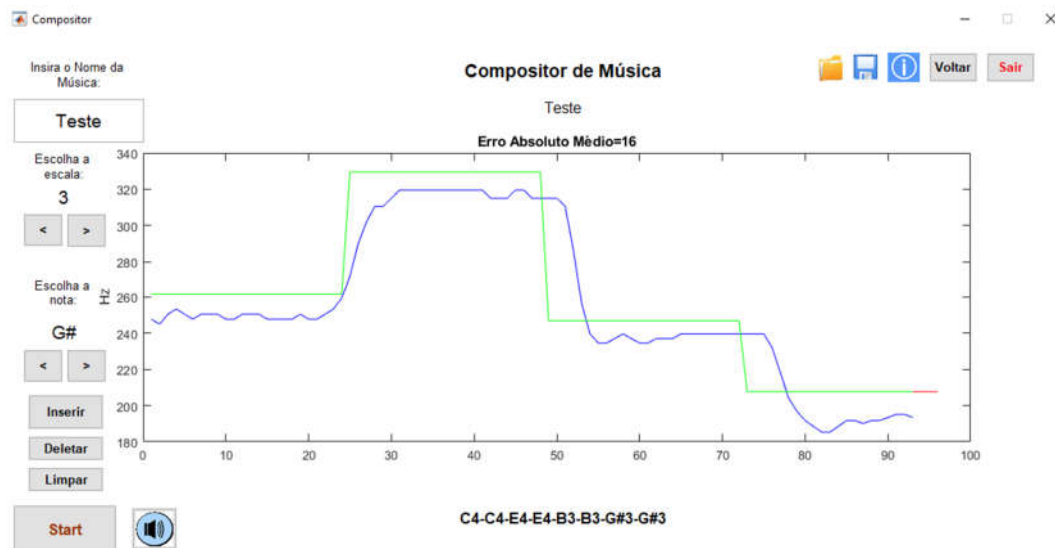


Figura 3.14: Interface gráfica do programa “Compositor de Música”.

A música criada pode ser acessada a qualquer momento. Por meio do botão “abrir”, onde se tem as músicas salvas no computador para serem treinadas. É possível realizar alterações na música carregada, como alterar todas as notas ou fazer modificações parciais, para que isso seja possível, é utilizado a paleta de implementação da Figura 3.13.

Ao criar ou abrir uma música, suas notas são salvas em um vetor que pode ser acessado por todas as funções implementadas no programa, ao realizar as alterações elas serão alteradas para todo o programa, isto é, as alterações são feitas num vetor global. Esse vetor sofre todas as edições realizadas pela paleta de configuração, inserção, troca ou remoção de notas.

A Figura 3.15 ilustra todas as decisões possíveis, apresentando o passo a passo de cada ação descrita anteriormente.

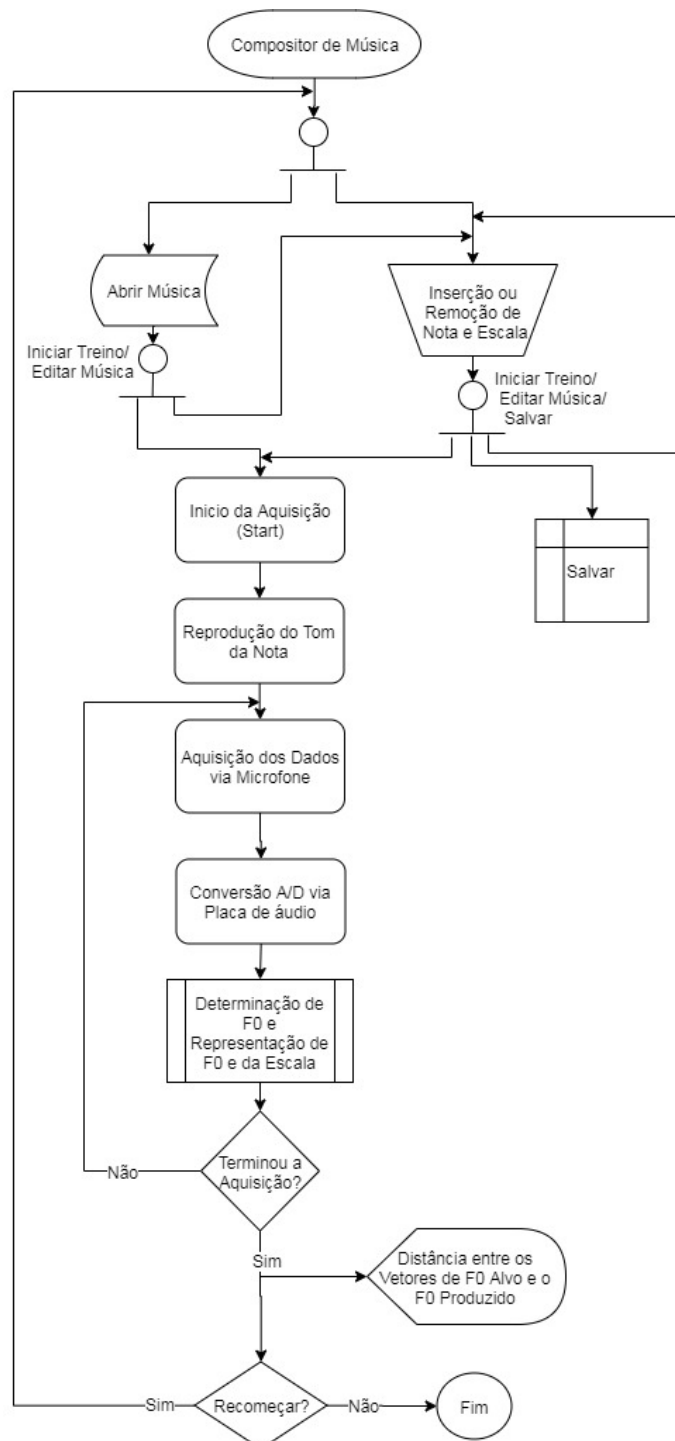


Figura 3.15: Fluxograma do funcionamento do “Compositor de Música”.

Capítulo 4: Avaliação e Resultados do Software

Equation Section (Next)

4.1. Caracterização da Amostra

O teste e avaliação do *software* foi realizado com 13 alunos, sendo todos do Instituto Politécnico de Bragança (IPB) de diferentes cursos e idades. Eles foram divididos em duas categorias são elas:

- I. Nunca canta ou canta com pouca frequência em público ou entre amigos.
- II. Canta com uma certa frequência em público ou entre amigos, podendo ser autodidata ou já estudou/estuda canto.

As características dos estudantes selecionados estão demonstradas na Figura 4.1. Os alunos são todos brasileiros com idade variando de 22 a 27 anos, sendo a maioria de 24 anos (31%) e a média de idade é de 24,4 anos. A maioria dos estudantes são do sexo masculino (69%) e estão na categoria II (54%).

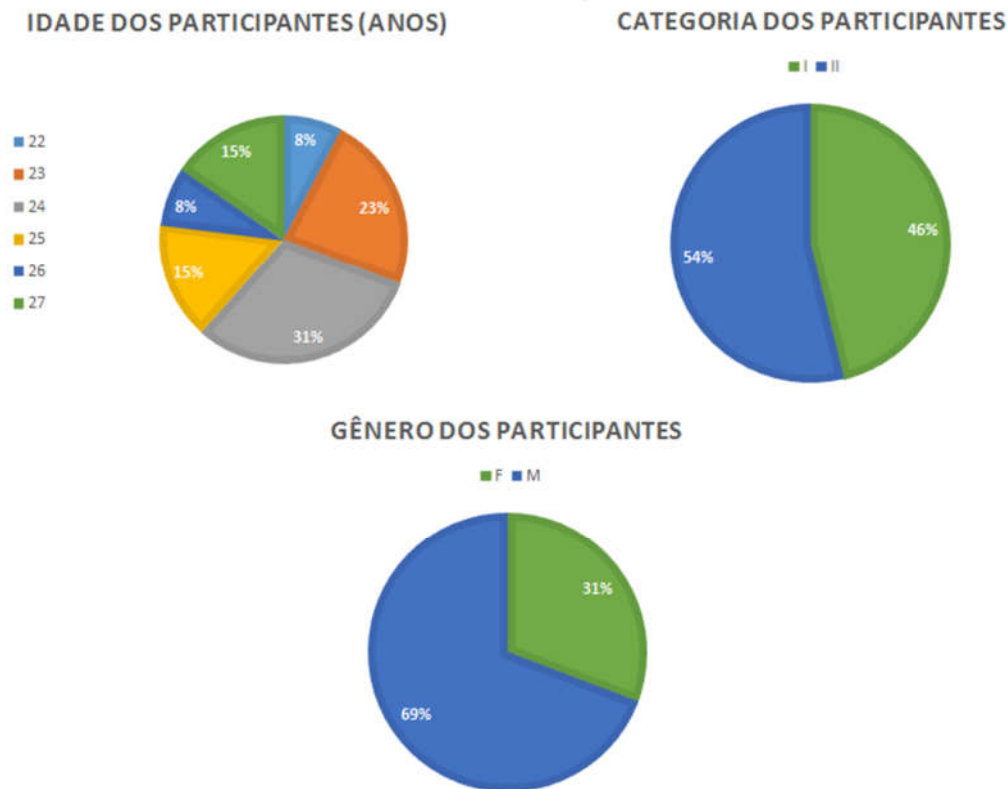


Figura 4.1: Gráfico sobre os alunos participantes

4.2. Avaliação Geral

A avaliação geral do software foi seguindo um inquérito utilizando as heurísticas adaptadas de Nielsen orientadas para *softwares* educacionais com interface voltada ao usuário. Analisando diversos aspetos do programa, como o diálogo simples e natural, a linguagem do usuário, o tempo de resposta, as saídas claramente marcadas, os atalhos, as mensagens de erro, a ajuda e documentação, a eficiência, a confiabilidade e a motivação (Nielsen, 1993; Nielsen e Molich, 1990). Tendo como possíveis respostas “Sim” para quando concordavam, “Não” para quando não concordavam e “Nenhuma Alternativa” (N/A) quando não sabiam opinar ou quando não concordavam completamente com o que estava sendo questionado. A Tabela 4.1 apresenta o inquérito que o usuário deveria preencher após utilizar o programa para avaliá-lo de forma geral.

Ao avaliar as onze heurísticas, foi verificado que o programa de forma geral possui uma linguagem simples e natural, sendo a compatível ao usuário e possuindo um tempo de resposta curto de acordo com os usuários. E, na maioria das heurísticas o programa teve um resultado acima de 80% de aceitação, como é possível ser verificado na Figura 4.2.

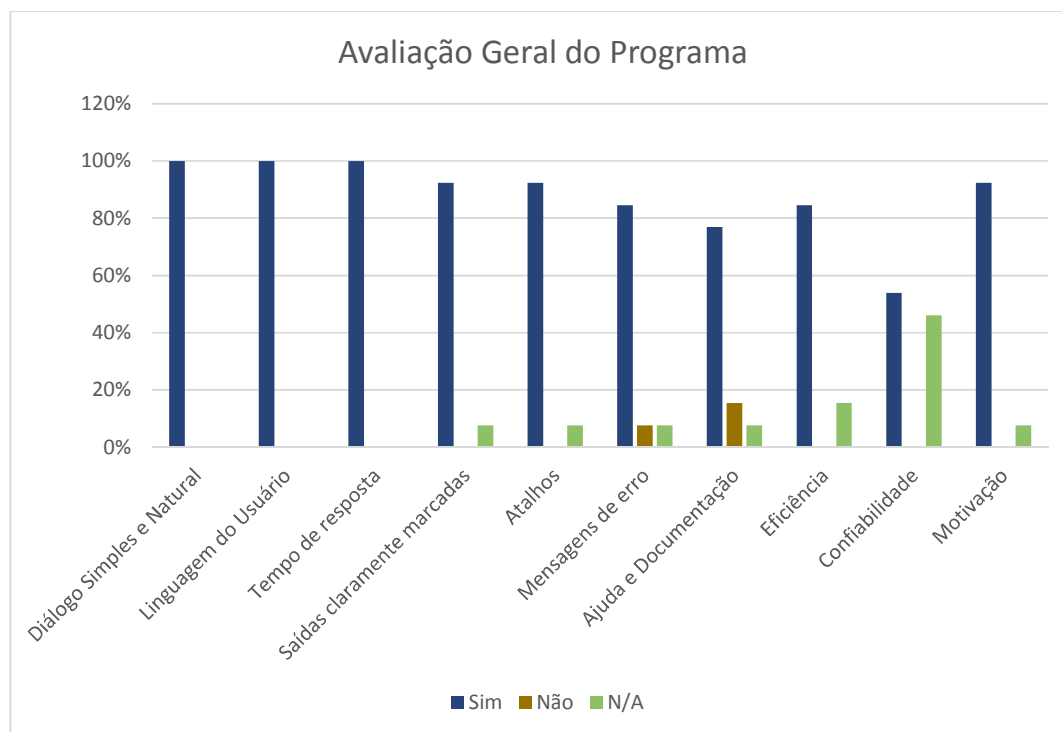


Figura 4.2: Gráfico de avaliação geral do programa.

Tabela 4.1: Inquérito utilizado para avaliação do software.

Heurísticas	Definição	Sim	Não	N/A
Diálogo Simples e Natural	As interfaces do software são simples, práticas e intuitiva?			
Linguagem do Usuário	A interface utiliza uma linguagem compatível?			
Tempo de resposta	O software informa continuamente ao usuário o que está ocorrendo e seu tempo de resposta é curto?			
Saídas claramente marcadas	A interface oferece facilidade para sair das situações?			
Atalhos	A interface permite que operações frequentemente usadas possam ser acedidas e executadas rapidamente?			
Mensagens de erro	As mensagens de erro são claras, precisas e úteis?			
Ajuda e Documentação	O software é fácil de utilizar, de tal forma que dispensa ajuda e consulta à documentação?			
Eficiência	A estrutura pedagógica criada oferece perspectivas de melhoria no processo de aprendizagem?			
Confiabilidade	Os resultados de resposta fornecidos pelo software e o conteúdo apresentado é academicamente confiável?			
Motivação	O software desperta o interesse e incentiva a cumprir os objetivos da aprendizagem?			

Entretanto, duas heurísticas não tiveram a avaliação acima de 80%. Uma das heurísticas, “Ajuda e Documentação”, teve uma aceitação de 77% no geral, devido aos usuários argumentarem que é sempre importante ter uma documentação para auxiliar a entender o

programa ou terem tido a necessidade de acesso à documentação para que entendesse o programa.

A “Confiabilidade” teve uma aceitação de 54%, pois 46% dos usuários colocaram N/A por não se sentirem seguros para afirmar que estava correto sobre ensino de canto, por serem autodidata ou por não possuírem conhecimento em canto. A Tabela 4.2 e a Tabela 4.3 expõem a quantidade de votos e a percentagem de cada heurística segundo cada categoria.

Tabela 4.2: Avaliação geral do programa de acordo com a categoria I.

Heurística	Sim	Não	N/A	Sim	Não	N/A
Diálogo Simples e Natural	6	-	-	100%	0%	0%
Linguagem do Usuário	6	-	-	100%	0%	0%
Tempo de resposta	6	-	-	100%	0%	0%
Saídas claramente marcadas	6	-	-	100%	0%	0%
Atalhos	6	-	-	100%	0%	0%
Mensagens de erro	5	-	1	83%	0%	17%
Ajuda e Documentação	4	1	1	67%	17%	17%
Eficiência	4	-	2	67%	0%	33%
Confiabilidade	4	-	2	67%	0%	33%
Motivação	6	-	-	100%	0%	0%

Tabela 4.3: Avaliação geral do programa de acordo com a categoria II.

Heurística	Sim	Não	N/A	Sim	Não	N/A
Diálogo Simples e Natural	7	-	-	100%	0%	0%
Linguagem do Usuário	7	-	-	100%	0%	0%
Tempo de resposta	7	-	-	100%	0%	0%
Saídas claramente marcadas	6	-	1	86%	0%	14%
Atalhos	6	-	1	86%	0%	14%
Mensagens de erro	6	1	-	86%	14%	0%
Ajuda e Documentação	5	-	2	71%	0%	29%
Eficiência	7	-	-	100%	0%	0%
Confiabilidade	3	-	4	43%	0%	57%
Motivação	6	-	1	86%	0%	14%

4.3. Avaliação Específica

Os usuários testaram cada um dos programas pelo menos cinco vezes para poder avaliá-los. Para analisar cada uma das heurísticas abaixo, é utilizado uma escala que vai de 1 a 5, referindo-se a muito ruim o número um e muito bom ao número cinco, como mostra a Tabela 4.4. As perguntas para avaliar os seguintes aspectos foram baseadas no inquérito de Leão (2018):

1. Funcionalidade: O software faz o que foi proposto de forma correta?
2. Eficiência: O tempo de resposta do software é adequado? O tempo de execução do software é adequado?
3. Confiabilidade: O software apresenta falhas sem frequência? O software devolve os resultados corretos?
4. Usabilidade: É fácil entender o conceito e a aplicação? É fácil aprender a usar?
5. Utilidade: Esta ferramenta é útil?
6. Adequação: O software dispõe de todas as funções necessárias para sua execução? O software é adequado com o conteúdo teórico?
7. Precisão: O software é preciso na execução de suas funções? O software é preciso nos resultados?

Tabela 4.4: Grau de avaliação específica de cada programa.

Grau de Avaliação	Avaliação
5	Muito Bom
4	Bom
3	Médio
2	Ruim
1	Muito Ruim

O resultado da avaliação das heurísticas anteriores foram apresentadas em forma de gráficos para apontar a aceitação dos participantes de forma geral e específica para cada uma das categorias. Esses valores são analisados de acordo com a média, o desvio padrão e a variância.

A variância mede o quanto as notas atribuídas para cada heurística avaliada se diferencia da média geral, como apresenta a Equação (4.1) (Montgomery, Runger & Hunbele, 2004).

$$\sigma^2 = V(x) = \sum \frac{(x_i - \bar{x})^2}{n} \quad (4.1)$$

Da mesma forma, o desvio padrão é calculado como a raiz quadrada da variância, como é demonstrado na Equação (4.2).

$$\sigma = \sqrt{V(x)} \quad (4.2)$$

4.3.1. Classificação Vocal

A classificação dos participantes foi realizada pelo programa, sendo apresentado a percentagem de classe na Figura 4.3. A maioria dos participantes eram do sexo masculino, logo a maioria deles possuem voz do tipo Baixo.

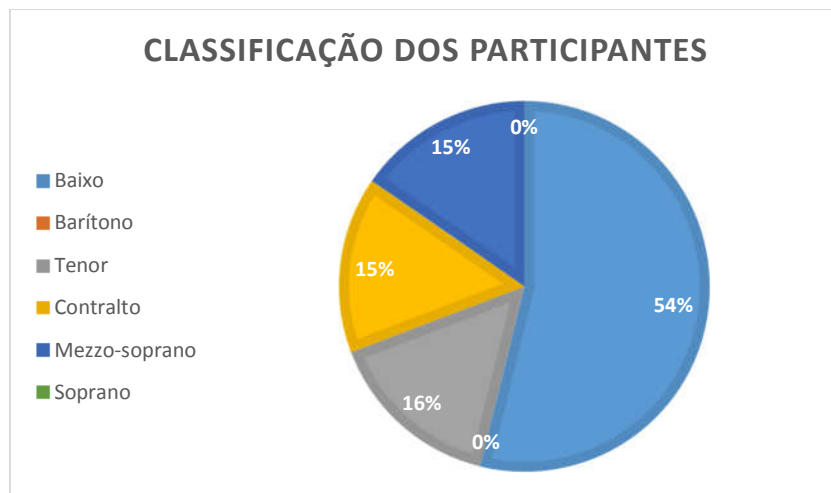


Figura 4.3: Gráfico da classificação dos participantes.

Ao utilizar essa ferramenta, os participantes concordaram com o resultado apresentado, mesmo aqueles que não sabiam exatamente qual a sua classificação, pois ao utilizarem o programa de “Treino de Escalas” acharam mais confortável treinar de acordo com o indicado pelo programa.

Conforme as notas dadas, o programa teve uma média de 4,61. Os valores das notas máxima e mínima, além da média, desvio padrão e variância, estão apresentados na Tabela 4.5.

A média de cada heurística ficou acima de 4,2, tendo uma baixa variância. Apenas três heurísticas tiveram nota mínimas de 4, sendo a funcionalidade, a utilidade e a precisão.

Tabela 4.5: Notas relacionadas ao programa “Classificação Vocal”.

Heurística	Máximo	Mínimo	Média	Desvio padrão	Variância
Funcionalidade	5	4	4,62	0,49	0,24
Eficiência	5	3	4,62	0,62	0,39
Confiabilidade	5	3	4,46	0,63	0,40
Usabilidade	5	3	4,77	0,58	0,33
Utilidade	5	4	4,69	0,46	0,21
Adequação	5	3	4,23	0,70	0,49
Precisão	5	4	4,85	0,36	0,13

A percentagem de cada valor dado nas heurísticas avaliadas são exibidas na Figura 4.4, verificando que a ferramenta teve em grande parte notas 4 e 5 em suas análises. Apenas a “Adequação” foi classificada, pelos usuários, como “bom” em sua maioria.

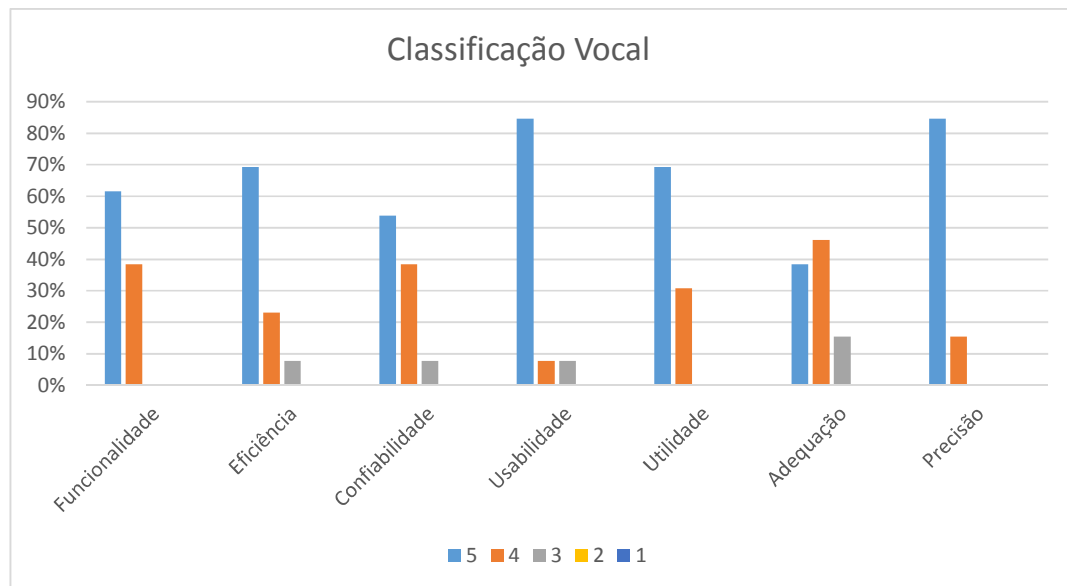


Figura 4.4: Gráfico das notas gerais da “Classificação Vocal”.

A “Usabilidade” e a “Precisão” tiveram uma percentagem de 85% na classificação “muito bom”. Os participantes acharam a ferramenta simples de aplicar e a informação complementar dada pelo programa auxiliou de forma satisfatória.

A “Utilidade” teve apenas notas 4 e 5, por muitos dos usuários não conhecerem a sua classificação, principalmente os participantes da categoria I. A Figura 4.5 e a Figura 4.6 apresenta as notas de acordo com cada categoria dos participantes.

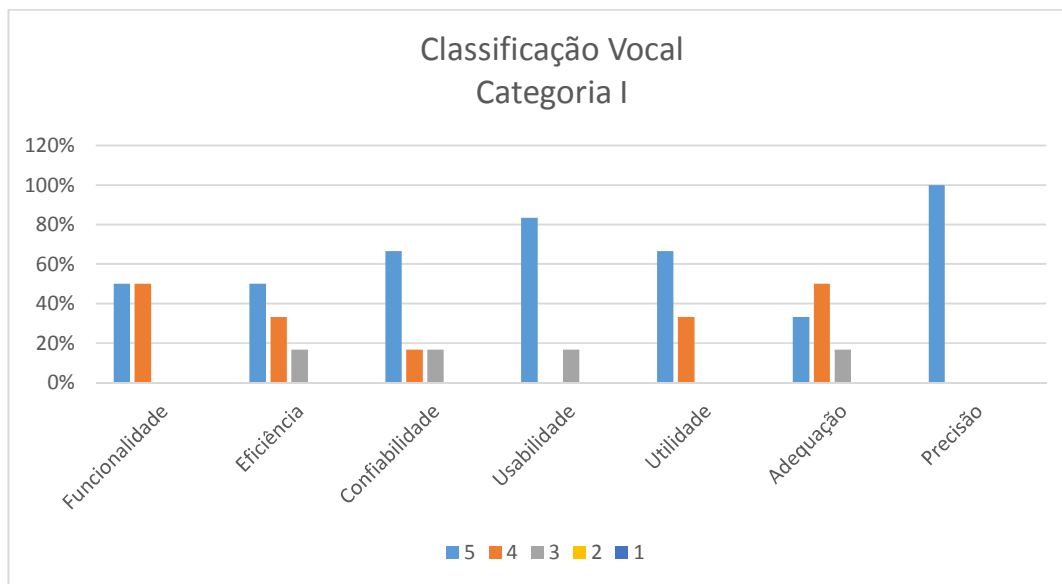


Figura 4.5: Gráfico das notas da “Classificação Vocal” pela categoria I.

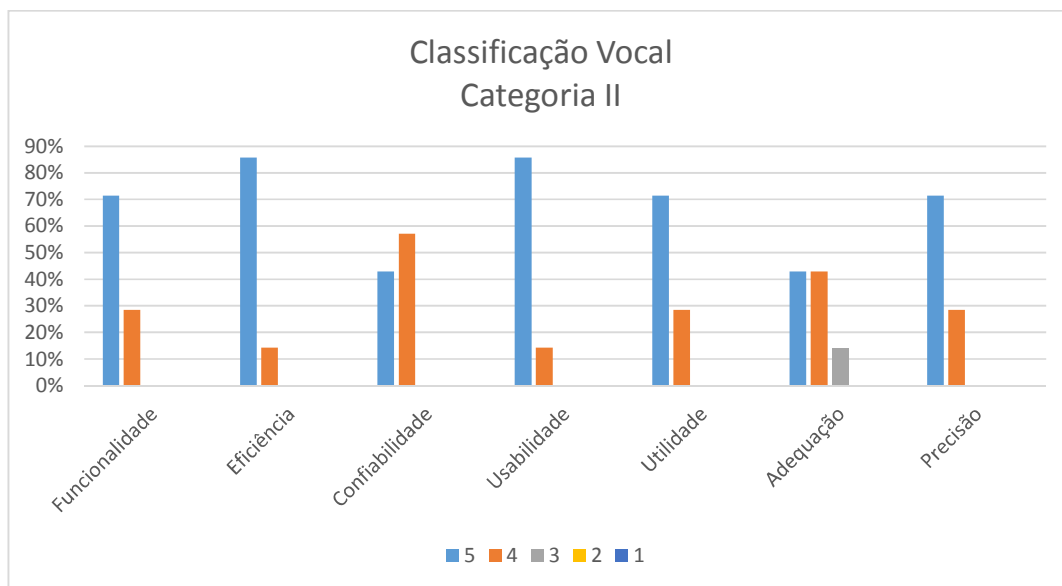


Figura 4.6: Gráfico das notas da “Classificação Vocal” pela categoria II.

4.3.2. Treino de Escalas

O “Treino de Escalas” teve a maior aceitação entre as ferramentas, a média do programa foi de 4,71. A Tabela 4.6 apresentou os menores valores de variância e maiores valores da média de cada heurística.

Tabela 4.6: Notas relacionadas ao programa “Treino de Escalas”.

Heurística	Máximo	Mínimo	Média	Desvio padrão	Variância
Funcionalidade	5	4	4,85	0,36	0,13
Eficiência	5	4	4,77	0,42	0,18
Confiabilidade	5	4	4,85	0,36	0,13
Usabilidade	5	3	4,77	0,58	0,33
Utilidade	5	3	4,85	0,53	0,28
Adequação	5	3	4,31	0,72	0,52
Precisão	5	4	4,54	0,50	0,25

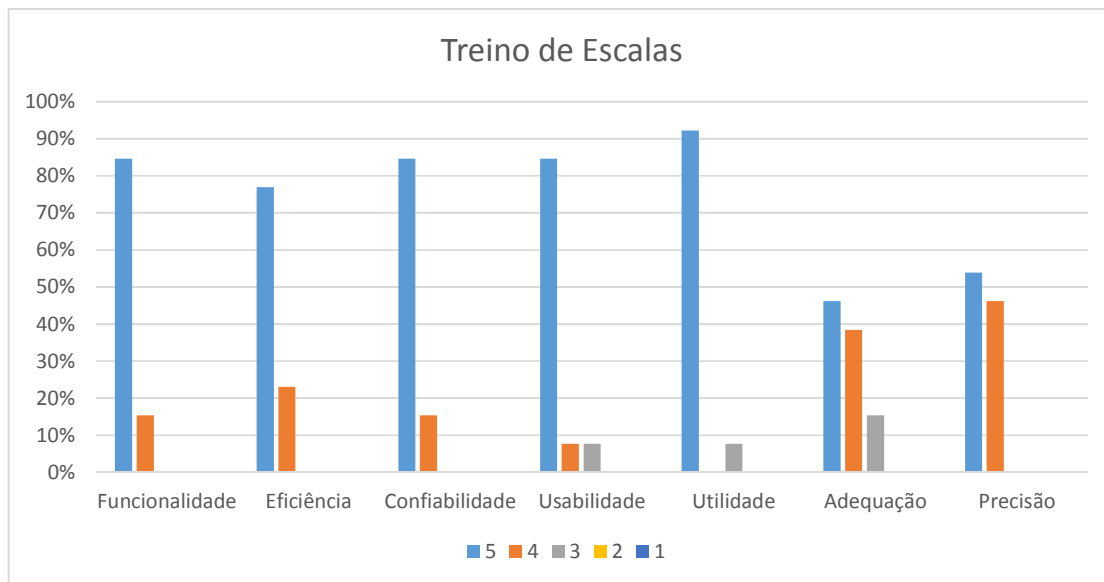


Figura 4.7: Gráfico das notas gerais do “Treino de Escalas”.

A Figura 4.7 aponta que o “Treino de Escalas” recebeu as notas 5 e 4 em quatro heurísticas e sendo que nas outras três, a nota 3 teve uma porcentagem de menos de 20%. A ferramenta, baseado nos dados anteriormente citados, se mostrou funcional, eficiente, confiável, simples e útil.

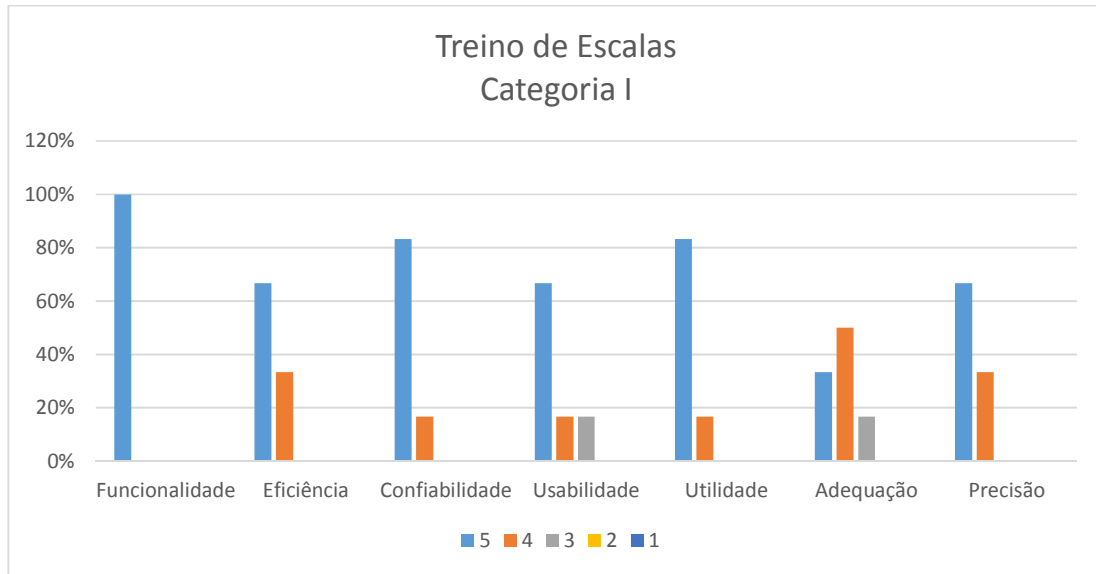


Figura 4.8: Gráfico das notas do “Treino de Escalas” pela categoria I.

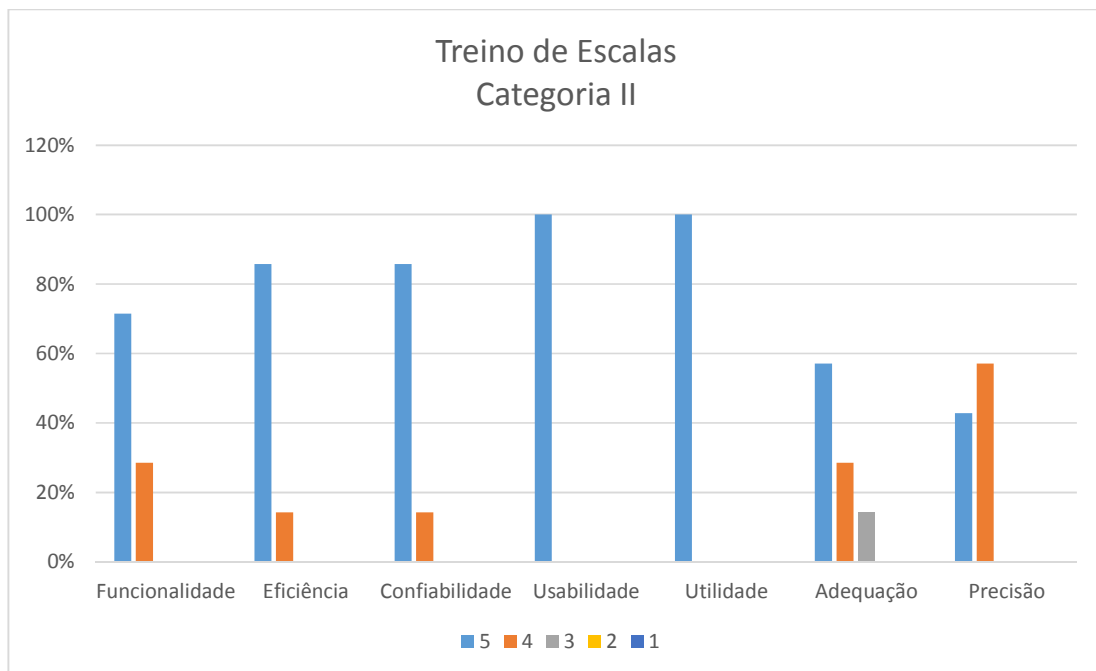


Figura 4.9: Gráfico das notas do “Treino de Escalas” pela categoria II.

4.3.3. Treino de Músicas

O “Treino de Músicas” foi a ferramenta que os usuários tiveram a maior dificuldade. E, conforme a Tabela 4.7, a média dos valores atribuídos foi de 4,40, a mais baixa entre os programas. A variância em quase todas as heurísticas está acima de 0,50, ou seja, a avaliação do programa variou muito em cada categoria.

Tabela 4.7: Notas relacionadas ao programa “Treino de Músicas”.

Heurística	Máximo	Mínimo	Média	Desvio padrão	Variância
Funcionalidade	5	3	4,38	0,74	0,54
Eficiência	5	3	4,77	0,58	0,33
Confiabilidade	5	3	4,23	0,80	0,64
Usabilidade	5	3	4,31	0,82	0,67
Utilidade	5	2	4,62	0,92	0,85
Adequação	5	3	4,15	0,86	0,75
Precisão	5	3	4,31	0,82	0,67

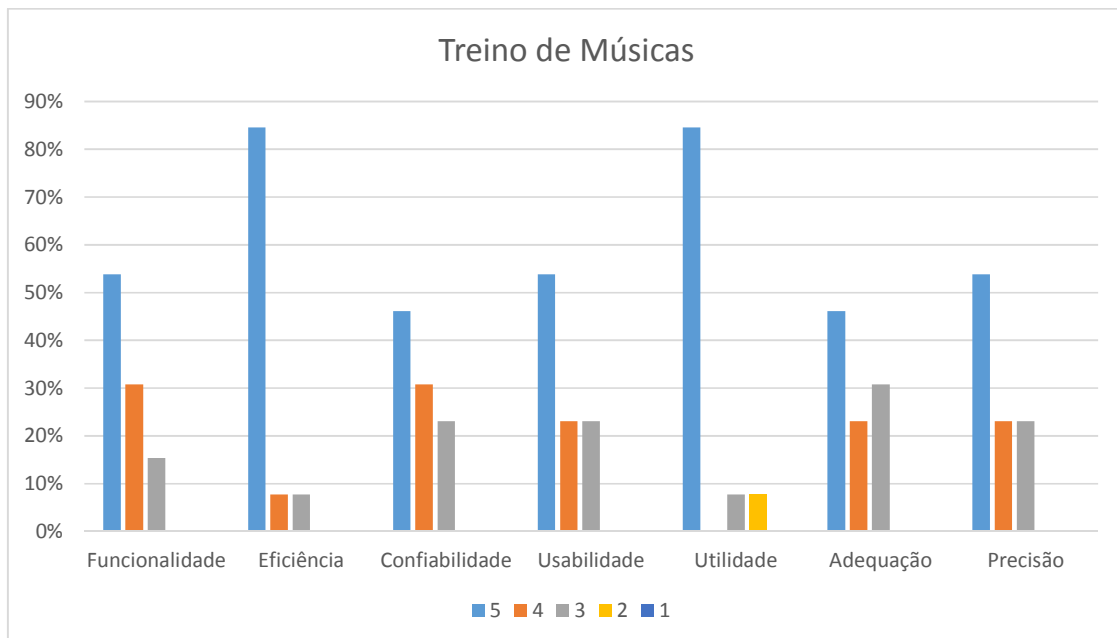


Figura 4.10: Gráfico das notas gerais do “Treino de músicas”.

Apesar de se considerar uma ferramenta difícil pelos usuários, a adaptação foi relativamente rápida. A categoria I não considerou o programa simples e fácil de mexer, determinando a “Usabilidade” como “médio” em 50%. Enquanto a categoria II, considerou 86% como “muito bom”.

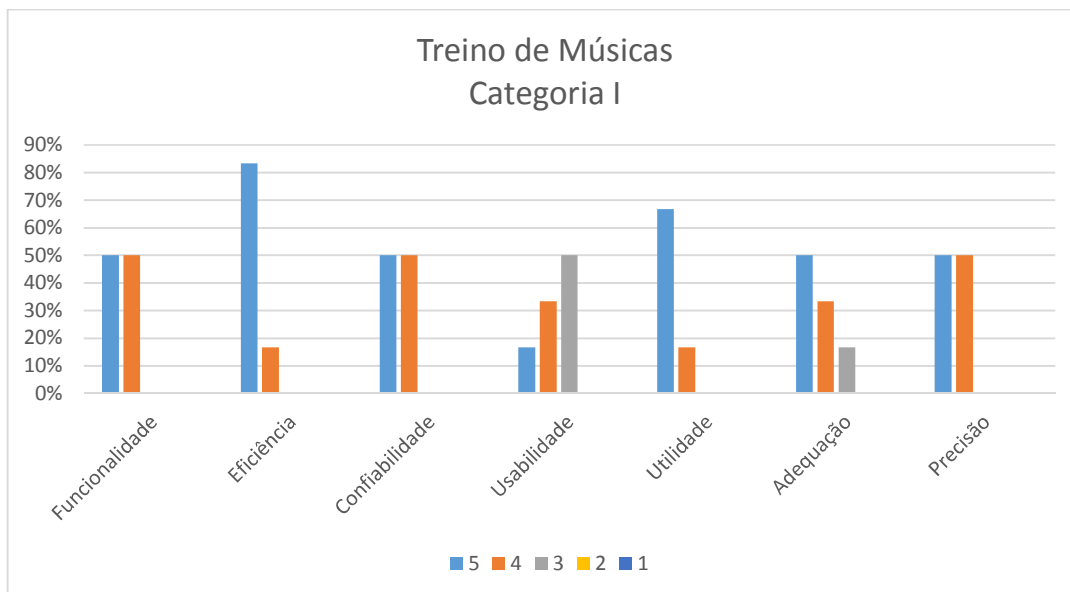


Figura 4.11: Gráfico das notas do “Treino de Músicas” pela categoria I.

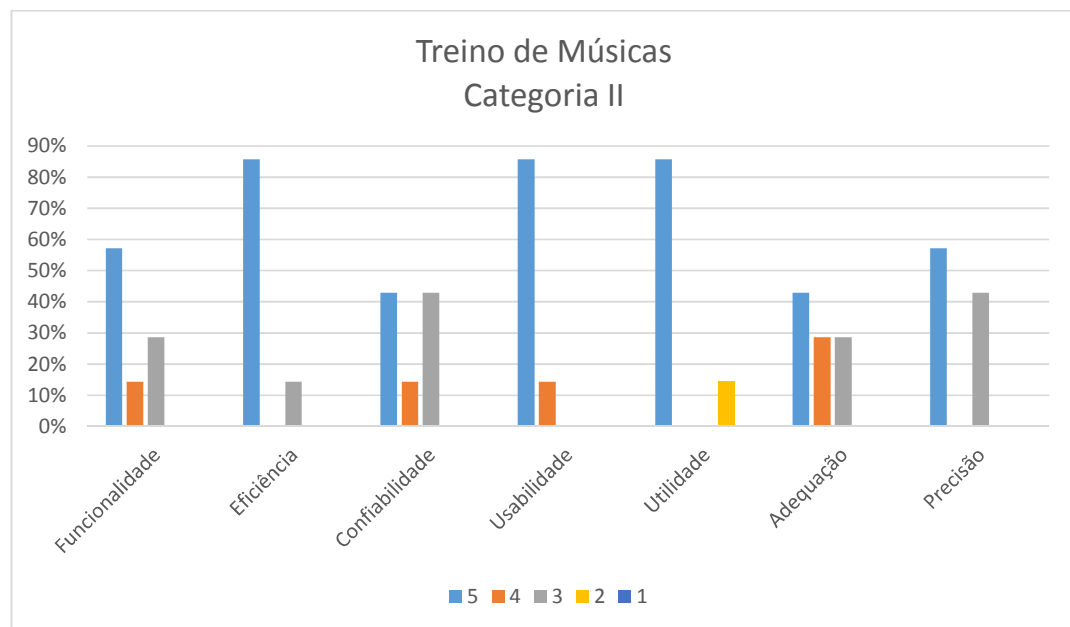


Figura 4.12: Gráfico das notas do “Treino de Músicas” pela categoria II.

4.3.4. Compositor de Música

O “Compositor de Música” é uma ferramenta para quem tem mais conhecimento em música. Os usuários que não possuem conhecimento tiveram dificuldade para entender e sentiram falta de algumas informações para facilitar o entendimento. A média do programa de acordo com a Tabela 4.8 foi de 4,45.

Tabela 4.8: Notas relacionadas ao programa “Compositor de Música”.

Heurística	Máximo	Mínimo	Média	Desvio padrão	Variância
Funcionalidade	5	3	4,62	0,62	0,39
Eficiência	5	3	4,92	0,27	0,07
Confiabilidade	5	3	4,69	0,61	0,37
Usabilidade	5	3	4,15	0,86	0,75
Utilidade	5	2	4,38	0,84	0,70
Adequação	5	3	3,85	0,86	0,75
Precisão	5	3	4,54	0,63	0,40

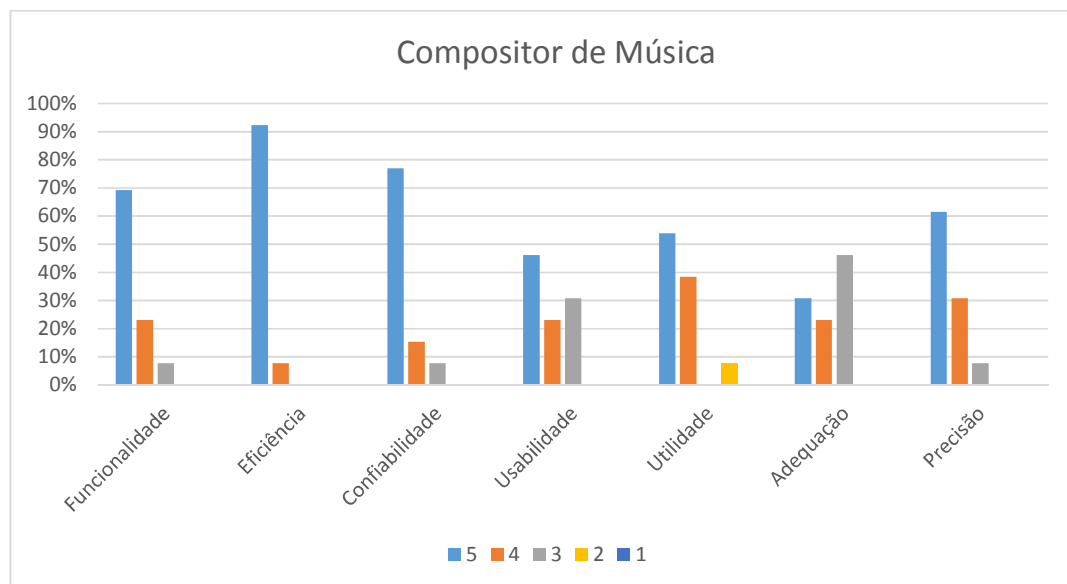


Figura 4.13: Gráfico das notas gerais do “Compositor de Música”.

A Figura 4.13 demonstra que o programa teve notas diversificadas. É possível observar que a “Usabilidade” e a “Adequação” está dividida entre os 3 valores, isto é, entre “muito bom”, “bom” e “médio”.

A Figura 4.14 apresenta que as heurísticas citadas anteriormente, tiveram nota 3 por 50% dos usuários da categoria I, expondo a dificuldade ao utilizar a ferramenta. Os usuários da categoria II sentiram falta de algumas funções para facilitar a composição da música, podendo ser visualizado na Figura 4.15.

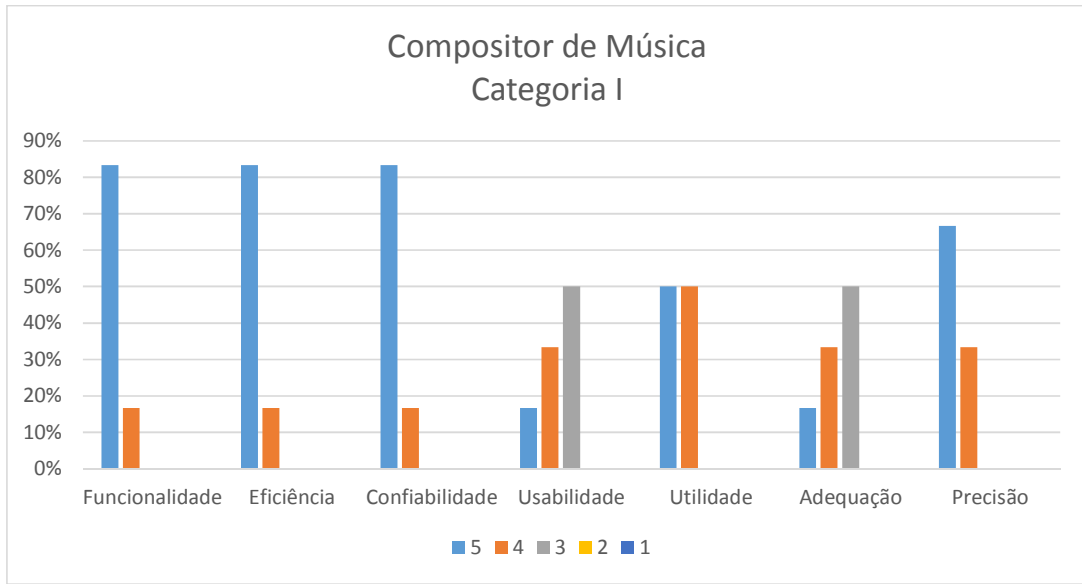


Figura 4.14: Gráfico das notas do “Compositor de Música” pela categoria I.

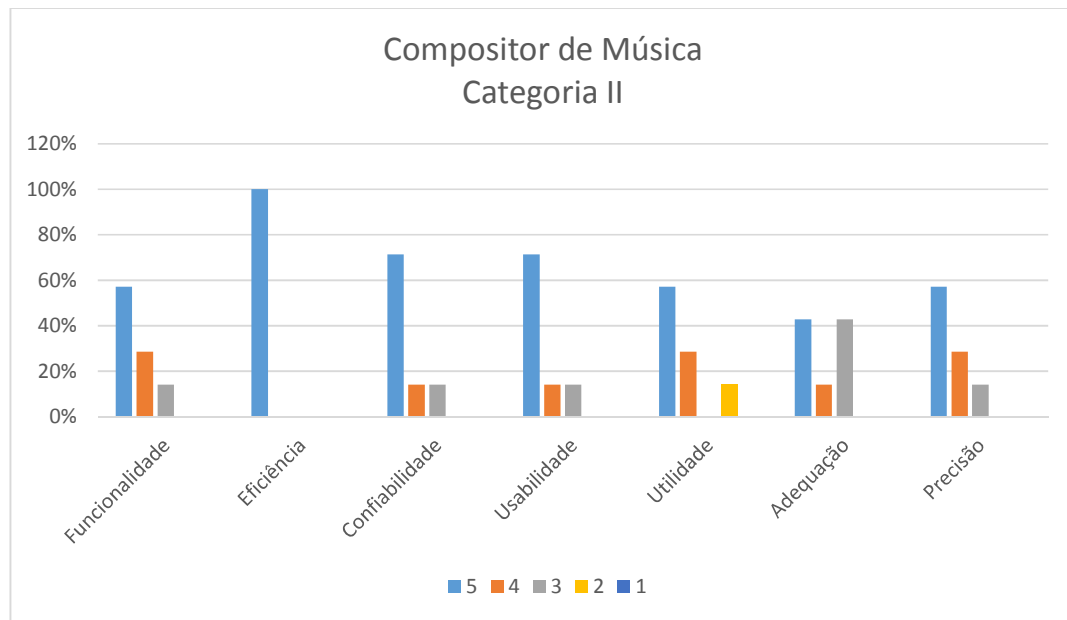


Figura 4.15: Gráfico das notas da “Compositor de Música” pela categoria II.

Capítulo 5: Conclusão e Trabalhos Futuros

Neste capítulo é apresentado a conclusão referente ao desenvolvimento e testes realizados, e discutindo possíveis melhorias a serem realizadas futuramente no *software*.

5.1. Conclusões Gerais

O *software* desenvolvido teve como base o trabalho feito por Leão (2018), realizando aquisição de dados em 64 bits utilizando o MATLAB. A aquisição em 64 bits foi implementado em dois métodos, o primeiro foi a criação de um objeto do tipo *AudioRecorder* que adquiri os dados diretamente do dispositivo de áudio do computador, e o segundo é utilizando a *toolbox Direct Sound* que possui diversas ferramentas para realizar a aquisição de dados, configurando-a de acordo com o desejado. Ambos os métodos se mostraram eficientes. O tempo máximo de aquisição foi de 20 s para não comprometer o funcionamento do programa, isto é, para que o tempo de resposta se mantivesse curto, sendo este considerado adequado pelos participantes.

O tempo de resposta do *software* apresentou ser curto o suficiente para realizar a aquisição do sinal de fala e a determinação da F0, sendo essa calculada utilizando o método da autocorrelação. Esse método demonstrou ser efetivo ao ser analisado um sinal e variando a sua frequência, isto é, o método determinou a F0 de acordo com a frequência do sinal gerado.

A aquisição em 64 bits e a determinação da F0 foram a base para o desenvolvimento das quatro ferramentas desse trabalho, sendo a primeira a “Classificação Vocal” para auxiliar o usuário a definir qual a sua classificação, a segunda o “Treino de Escalas” com uma sequência de notas de acordo com o tipo de voz, a terceira o “Treino de Músicas” possuindo uma base de dados de músicas para poder ser treinada e, por último, o “Compositor de Música” que possibilita a implementação das notas para a criação de uma música desejada.

As ferramentas desenvolvidas foram bem aceitas pelos participantes que realizaram os testes, onde os usuários apresentaram um interesse para que o *software* se tornasse um produto.

5.2. Trabalhos Futuros

Foi verificado após realizado os testes que os usuários acham interessante realizar algumas melhorias para serem implementadas para auxiliar no treino de afinação vocal. Os usuários sem conhecimento de música, acharam que facilitaria a implementação de um tutorial para que, de forma educativa, se entendesse o programa. Na ferramenta “Treino de Músicas” foi sugerido um aumento da base de dados, também foi proposta a possibilidade de selecionar qual parte da música que deseja ser treinada e realizar um acompanhamento da mesma. Na ferramenta “Compositor de Música”, recomendou-se a implementação de poder selecionar o tempo que cada nota pode ter, sem precisar inserir várias vezes a mesma nota. Outro ponto interessante, é a possibilidade do usuário salvar o áudio e o desempenho do teste para que assim, no futuro, consiga realizar comparações do seu processo evolutivo utilizando o *software*.

Outra possível melhoria, seria a tradução do programa para uma linguagem que seja mais independente, ou seja, fazer com que o programa não dependa do *software* MATLAB para realizar sua compilação. Uma alternativa seria a utilização de linguagens, como a Python e a C++ por exemplo, que apresentam ferramentas que auxiliam no desenvolvimento do programa.

Bibliografia

- Aleixo, Giorgia Taiaco & Mansanares, Antonio Manoel (2003). *Propriedades das Ondas numa Corda*. Relatório Final de F809. IFGW/UNICAMP.
- Almeida, Náthalee Cavalcanti de (2010). *Sistema Inteligente para Diagnóstico de Patologias na Laringe Utilizando Máquinas de Vetor de Suporte*. Universidade Federal do Rio Grande do Norte. Dissertação de mestrado em Engenharia Elétrica e de Computação.
- Almeida, Vítor Filipe Maia Baptista Fonseca (2012). *Relação entre Características Objetivas da Voz Cantada e seus Atributos Artísticos e Estéticos*. Faculdade de Engenharia da Universidade do Porto. Dissertação de mestrado em Engenharia Electrotécnica e de Computadores.
- Andrade, Luciana Mara de Oliveira (2003). *Determinação dos Limiares de Normalidade dos Parâmetros Acústicos da Voz*. Universidade de São Paulo. Dissertação de mestrado em Bioengenharia.
- Baptista, Manuel (2019). *Sistemas De Aquisição De Dados. Notas de Aula. Engenharia de Sistemas e Informática - 2.º Ano*. Acessado em: Setembro de 2019.
- Bartošek, Jan & Hanžl, Václav (2010). *Comparing Pitch Detection Algorithms for Voice Applications*. In: Semantic Scholar.
- Blue Design (2019). “Snowball”. URL:<https://www.bluedesigns.com/products/snowball/>
- Callaghan, J., Thorpe, W., & van Doorn, J. (2004, April). *The science of singing and seeing*. In Proceedings of Conference of Interdisciplinary Musicology, Graz, Austria.
- Camargo, Thatiana Francisco de, Barbosa, Daniela Aparecida & Teles, Lídia Cristina da Silva (2007). *Característica da fonetografia em coristas de diferentes classificações vocais*. Revista da Sociedade Brasileira de Fonoaudiologia [online].
- Chai, T.& Draxler, R. R. (2014). *Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature*. Geoscientific model development, 7(3), 1247-1250.

- Costa, P. J. B. M., Ferreira, K. L., de Camargo, Z. A., & Pinho, S. M. R. (2006). *Extensão vocal de cantores de coros evangélicos amadores*. Revista CEFAC, 8(1), 96-106. URL: <http://www.redalyc.org/articulo.oa?id=169320516015>.
- Coton, Raymond H. & Woo, Peak (1995). *Measuring Vocal Function*. In: Silva, Melissa C. & Biase, Noemi de. *Estudo da Frequência Fundamental da voz em Mulheres Jovens com Síndrome Pré-Menstruais*. New York, Ygaku-shoin.
- Coleman, R. F., Mabis, J. H., & Hinson, J. K. (1977). *Fundamental frequency-sound pressure level profiles of adult male and female voices*. Journal of Speech and Hearing Research, 20(2), 197-204.
- De Cheveigné, A., & Kawahara, H. (2002). *YIN, a fundamental frequency estimator for speech and music*. The Journal of the Acoustical Society of America, 111(4), 1917-1930.
- Diniz, Paulo S. R., Silva, Eduardo A. B. da & Netto, Sergio L. (2014). *Processamento Digital de Sinais: Projeto e Análise de Sistemas*. Bookman, 2º edição.
- Fernandes, Joana Filipa Teixeira (2018). *Determinação da Autocorrelação, HNR e Nhr para Análise Acústica Vocal*. Escola Superior de Tecnologia e Gestão no Instituto Politécnico de Bragança. Dissertação de mestrado em Engenharia Industrial.
- Felippe, Ana Clara Naufel de, Grillo, Maria Helena M. M. & Grechi, Thaís Helena (2006). *Standardization of acoustic measures for normal vocal patterns*. Revista Brasileira de Otorrinolaringologia, 72(5), pp. 659-664.
- Ferreira, João Filipe Terleira de Sá (2012). *Tecnologia de Apoio em Tempo Real ao Canto – Relação entre parâmetros preceptivos da voz cantada com o fenômenos acústicos objetivos*. Escola Superior de Música, Artes e Espectáculo do Instituto Politécnico do Porto. Dissertação de mestrado em Música.
- Holzinger, A. (2005). *Usability Engineering Methods for Software Developers*. Communications of the ACM, 48(1), 71-74.
- Hoppe, D., Sadakata, M. & Desain, P. (2006). *Development of Real-Time Visual Feedback Assistance in Singing Training: a Review*. In: Journal of Computer Assisted Learning, vol. 22.4, pp. 308–316. Issn: 1365-2729. DOI: <http://dx.doi.org/10.1111/j.1365-2729.2006.00178.x>.
- Howard, David M. & Welch, Graham F. (1989). *Microcomputer-Based Singing Ability Assessment and Development*. In: Applied Acoustics, vol. 27.2, pp. 89-102. Issn: 0003- 82X. DOI: [https://doi.org/10.1016/0003-682X\(89\)90002-9](https://doi.org/10.1016/0003-682X(89)90002-9).
- Junker, D. B. (1999). *Convenção Internacional de Regentes de Coros*. Anais. Brasília.

-
- Lathi, B.P. (2006). *Sisnais e Sistemas Lineares*. 2ª edição. Bookman. ISBN: 9788577803910.
URL: <https://books.google.pt/books?id=S806DQAAQBAJ>.
- Leão, Izabela Ribeiro (2018). *Software de Apoio à Afinação da Voz*. Escola Superior de Tecnologia e Gestão no Instituto Politécnico de Bragança. Dissertação de mestrado em Engenharia Industrial.
- Lopes, José Manuel dos Santos. (2018). *Ambiente de Análise Robusta dos Principais Parâmetros Qualitativos da Voz*. Faculdade de Engenharia da Universidade do Porto. Dissertação de mestrado em Engenharia Electrotécnica e de Computadores.
- MathWorks, Inc. (2014a). *About MATLAB – Overview*. Massachusetts, United States.
URL: https://www.mathworks.com/products/matlab.html?s_tid=hp_ff_p_matlab.
- MathWorks, Inc. (2014b). *Create Apps with Graphical User Interfaces in MATLAB*. Massachusetts, United States. URL: <https://www.mathworks.com/discovery/matlab-gui.html%20.html>.
- MathWorks, Inc. (2019a). *dsp.AudioRecorder*. Massachusetts, United States.
URL: <https://www.mathworks.com/help/dsp/ref/dsp.audiorecorder-system-object.html>.
- MathWorks, Inc. (2019b). *Analog Input and Output Workflow*. Massachusetts, United States.
URL: <https://www.mathworks.com/help/daq/analog-input-and-output-workflow.html>.
- Montgomery, D. C., Runger, G. C., & Hunbele, N. F. (2004). *Estatística aplicada à engenharia*. LTC.
- Nadoleczny, Max (1923). *Untersuchungen über den kunstgesang*. Berlin: Springer
- Neves, Caio Gomes Fontes Calaça (2016). *Sistema para Afinação de Sinais Monofônicos de Voz com Interface Gráfica*. Universidade Federal do Rio de Janeiro.
- Nielsen, J. (1994). *Usability engineering*. Elsevier.
- Nielsen, J. & Molich, R. (1990). *Heuristic evaluation of user interfaces*. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 249-256). ACM.
- Oliveira, Keroline R. & Lima, Ailen Rose Balog de (2016). *Um Estudo sobre a Relação entre Capacidade de Concentração e Afinação Vocal: um Relato de Experiência*. Revista Integratio, v.2, n.2, ago., pp. 54-60.
- Oppenheim, Alan V. & Schaffer, Ronald W. (2009). *Discrete-Time Signal Processing*. 3rd. Upper Saddle River, NJ, USA. Prentice Hall Press.
- Perelló, J., Enrique Guitart, Santiago Martínez-Fornés & Pamias, Juan A (1982). *Canto-Dicción: Foniatria Estética*. Barcelona. Editorial Científico-Médica.
-

- Pinto, António & Tomé, Ana Maria (2001). *Automatic pitch detection na midi conversion for the singing voice*. In: WSES International Conferences: AITA'01, AMTA'01, MCBE'01, MCBC'01 (pp. 312-317).
- Rossiter, D. & Howard, D. M. (1996). *ALBERT: Real- Time Visual Feedback Computer Tool for Professional Vocal Development*. In: Journal of Voice, vol. 10, pp. 321-336.
- Science Buddies Staff. (2018). *Correlation Between Relative Pitch and Age, Gender, or Musical Background*. URL: https://www.sciencebuddies.org/science-fair-projects/project-ideas/Music_p015/music/relative-pitch-age-gender-musical-background .
- Sundberg, Johan (1977). *The Acoustics of the Singing Voice*. Scientific American, pp. 104-114.
- Tan, Li & Karnjanadecha, Montri (2003). *Pitch Detection Algorithm: Autocorrelation Method and AMDF*. Prince of Songkhla University.
- Teixeira, João Paulo Ramos (1995). *Modelização Paramétrica de Sinais para Aplicação em Sistemas de Conversão Texto-Fala*. Faculdade de Engenharia da Universidade do Porto. Dissertação de mestrado em Engenharia Electrotécnica e de Computadores.
- Teixeira, J. P., Ferreira, D. B. & Moreira, S. (2011). *Análise Acústica Vocal - Determinação do Jitter e Shimmer para Diagnóstico de Patologias da Fala*.
- Teixeira, João Paulo (2018). *Notas de Aula – Microfones: Funcionamento e Características*. Em: Aplicações de Processamento de Sinal.
- Uliani Neto, M., Violatto, R. P. V., Simões, F. O. & Costa, B. R. (2013). *Técnicas de Codificação de Fala Baseadas em Análise por Síntese*. Faculdade de Jaguariúna; Faculdade Politécnica de Campinas; Fundação CPqD.
- Upadhya, S. S. & Wankhede, N (2012). *Pitch Estimation Using Autocorrelation Method and AMDF*. In: International Conference on Advances in Computing and Management (pp. 249-253).
- Vieira, M. N. (2005). *Uma Introdução à Acústica da Voz Cantada*. I Seminário Música Ciência Tecnologia: Acústica Musical. Universidade Federal de Minas Gerais.
- Vocal Pitch Monitor (2019). YAMAOKA Tadoa. Software. Version 1.5.1.
- Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2014). *Mathematical Statistics with Applications*. Cengage Learning, 7º Ed.
- Welch, Graham F. (1985). *A Schema Theory of How Children Learn to Sing in Tune*. In: Psychology of Music, vol. 13, pp. 3-18.

- Willmott, C. J. & Matsuura, K. (2005). *Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance*. *Climate research*, 30(1), 79-82.
- Wilson, P. H., Lee, K., Callaghan, J., & Thorpe, C. W. (2008). *Learning to sing in tune: Does real-time visual feedback help?*. In: *Journal of interdisciplinary music studies*, 2.
- Zaki-Azat, J. N. (2016). *The Influence of Real-Time Visual Feedback Training on Vocal Control*. Wilfrid Laurier University