



# Software de Apoio à Afinação da Voz

**Izabela Ribeiro Leão**

Dissertação apresentada à Escola Superior de Tecnologia e de Gestão de Bragança para  
obtenção do Grau de Mestre em Engenharia Industrial.

Trabalho orientado por:

Prof. Dr. João Paulo Teixeira

Co-orientado por:

Prof. Dr. Roberto Ribeiro Neli

Bragança

2018





# Software de Apoio à Afinação da Voz

**Izabela Ribeiro Leão**

Dissertação apresentada à Escola Superior de Tecnologia e de Gestão de Bragança para  
obtenção do Grau de Mestre em Engenharia Industrial.

Trabalho orientado por:

Prof. Dr. João Paulo Teixeira

Co-orientado por:

Prof. Dr. Roberto Ribeiro Neli

Bragança

2018



A Escola Superior de Tecnologia e de Gestão não se responsabiliza pelas opiniões expressas neste relatório.



# Agradecimentos

Agradeço primeiramente aos meus pais, responsáveis por todos os méritos e objetivos alcançados por mim. O que sou e o que fiz é inquestionável resultado de todos os esforços e ensinamentos transmitidos por eles. Qualquer agradecimento não estará à altura para expressar e retribuir o que fizeram para que eu pudesse chegar até aqui.

Agradeço também ao orientador prof. Dr. João Paulo Teixeira e ao coorientador Prof. Dr. Roberto Ribeiro Neli por toda a paciência, instrução e tempo disponibilizado para me guiar na construção desta dissertação.

Agradeço a todos os meus amigos e pessoas que de alguma forma contribuíram com o projeto, em especial Matthias Tilton pelo companheirismo e Carla Carolina da Cunha pela contribuição e ajuda como prova de consistente amizade.

Agradeço, por fim, a Universidade Tecnológica Federal do Paraná e ao Instituto Politécnico de Bragança pela oportunidade e investimento oferecido na minha trajetória acadêmica.



# Resumo

A criação de um *software* aplicado a uma interface destinada a alunos de música e cantores, nasce da necessidade de um utensílio que reúne diversas ferramentas em apoio à afinação da voz. O software é destinado tanto a utilizadores sem experiência em teoria musical, alunos do tema em específico quanto cantores experientes. O método de aquisição da frequência fundamental da voz é a autocorrelação com *center-clipping*. Na presente pesquisa, utilizou-se o *software* MatLab para o desenvolvimento do sistema. O sinal de voz adquirido é comparado com um vetor de notas que varia de acordo com uma das quatro aplicações do *software*: Ferramenta de apoio à determinação da classificação vocal, treino da escala musical, treino de músicas pré-escritas e finalmente um escritor de melodias com a possibilidade de reprodução da melodia composta, antes do treino. O *software* foi avaliado por três categorias de alunos sendo: sem nenhuma, pouca e considerável experiência com canto. O programa apresentou bons resultados avaliativos nos aspectos geral e específico de cada ferramenta, e um índice evolutivo significativo ao longo dos treinos, com uma diminuição considerável do erro, incluso o grupo com significativa experiência. Desta maneira, verificou-se a validade da ferramenta e a eficiência do método do autocorrelação para a determinação da frequência fundamental, numa aplicação prática de sua determinação em curto espaço de tempo.

**Palavras-chave:** Matlab; Autocorrelação; Frequência Fundamental; Pitch; Afinação vocal.



# Abstract

The creation of a software applied to an interface intended for music students and singers, comes from the need of a new utensil version that reunites different tools to support voice tuning. The software is intended for both users with no experience in musical theory, theme students in specific and experienced singers. The acquisition method of the fundamental frequency is autocorrelation with center-clipping. In the present research, MATLAB was the choosen software for the development of the system. The acquired voice signal is compared to a vector of notes that vary according to the four applications of the software: Vocal classification support tool, musical scale training, prewritten music training and finally a composer of melodies with a possibility of chearing the created melody before training. The software was evaluated by three categories of students: without any, little and good experience with singing. The program presented good evaluative results in the various exercises, and an evolutionary index throughout the training, with a greater diminution of the error, including the group with significant experience. In this way, the projects validates the efficiency of the autocorrelation method for the fundamental frequency determination, and proposes an option for a practical application of aquisition in short time.

**Keywords:** Matlab; Autocorrelation; Fundamental Frequency; Pitch; Voice Tuning.



# Conteúdo

Agradecimentos	vii
Resumo	ix
Abstract	xi
Conteúdo	xiii
Lista de Tabelas	xix
Lista de Figuras	xx
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos . . . . .	4
1.2 Estado da Arte . . . . .	5
1.2.1 Produções e Pesquisas Acadêmicas . . . . .	5
1.2.2 Aplicações e Softwares . . . . .	7
1.3 Estrutura do Relatório . . . . .	10
<b>2 Metodologia</b>	<b>13</b>
2.1 O MatLab . . . . .	13
2.2 Frequência Fundamental . . . . .	13
2.2.1 Método da Autocorrelação . . . . .	15
2.2.2 Autocorrelação com Center-Clipping . . . . .	17

2.3	Aquisição do Sinal de Voz . . . . .	18
2.3.1	Frequência de Amostragem . . . . .	18
2.3.2	Aquisição do sinal em MatLab 32 Bits . . . . .	19
2.3.3	Aquisição do sinal em MatLab 64 Bits . . . . .	21
2.4	Teoria Musical . . . . .	23
2.4.1	Escalas musicais . . . . .	23
2.4.2	Extensão Vocal . . . . .	24
2.5	Pós-Processamento da F0 . . . . .	27
2.5.1	Composição da Melodia . . . . .	29
2.5.2	Erro Quadrático Médio . . . . .	29
<b>3</b>	<b>Aplicações</b>	<b>31</b>
3.1	Organização Geral do Software . . . . .	31
3.1.1	Diagrama de Blocos . . . . .	31
3.1.2	Menu Principal . . . . .	33
3.2	Nota e Classificação Vocal . . . . .	33
3.3	Treino da Escala Musical . . . . .	35
3.4	Treino de Músicas . . . . .	37
3.5	Compositor de Músicas . . . . .	39
<b>4</b>	<b>Performance e Resultados</b>	<b>43</b>
4.1	Caracterização da Amostra . . . . .	43
4.2	Avaliação do Software . . . . .	45
4.2.1	Avaliação Geral . . . . .	45
4.2.2	Avaliações Específicas . . . . .	47
4.3	Resultados Experimentais . . . . .	51
4.3.1	Determinação da Classificação Vocal . . . . .	51
4.3.2	Treino da Escala Musical . . . . .	52
4.3.3	Treino da Música Parabéns . . . . .	53
4.3.4	Treino da Melodia . . . . .	53

4.4	Evolução . . . . .	54
5	Conclusão e Trabalhos Futuros	59
	Bibliografia	60

# Lista de Tabelas

2.1	Relação entre as notas musicais, intervalos e frequências correspondentes. Extraído de (Aleixo e Mansanares, 2003). . . . .	25
4.1	Avaliação dos Aspectos Gerais do Software . . . . .	46
4.2	Resultado da Avaliação dos Aspectos gerais do Software . . . . .	47
4.3	Escala de Avaliação utilizada para os Aspectos Específicos . . . . .	48
4.4	Estatística dos resultados de avaliação da ferramenta Nota e Classificação Vocal	49
4.5	Estatística dos resultados de avaliação da ferramenta Treino da Escala . . . . .	49
4.6	Estatística dos resultados de avaliação da ferramenta Treino de Músicas . . . . .	50
4.7	Estatística dos resultados de avaliação da ferramenta Escritor de Músicas . . . . .	50
4.8	Porcentagem das classificações vocais detectadas . . . . .	51
4.9	Média do MSE por categoria e TD Médio e aAbsoluto do treino da escala musical	52
4.10	Média do MSE por categoria e TD Médio e Absoluto do treino da música Parabéns . . . . .	53
4.11	Média do MSE por categoria e TD Médio e Absoluto do treino da Melodia Composta . . . . .	54



# Lista de Figuras

1.1	<i>Aparelho fonador humano.</i>	1
1.2	<i>Elementos de um sistema de aquisição de dados.</i>	3
1.3	Exemplo de funcionamento do software Canta.	7
1.4	Exemplo de funcionamento do software Sing and See.	8
1.5	<i>Sing True: Exercício de treinamento do acorde Lá Maior (A).</i>	9
1.6	<i>Media Singing Method: Acompanhamento vocal da ópera Ave Maria, Schubert.</i>	10
2.1	<i>Sinal Original.</i>	16
2.2	<i>Janela de observação com um período de 30ms do sinal original.</i>	16
2.3	<i>Resultado da Autocorrelação do Sinal utilizando Certer-clipping.</i>	17
2.4	<i>Análise do período glotal a partir do máximo da Autocorrelação.</i>	18
2.5	<i>intervalo de um tom</i>	23
2.6	<i>Pauta Geral correspondente aos intervalos vocais de cantores. A linha superior corresponde a vozes femininas altas, médias e baixas, e a linha inferior corresponde a vozes masculinas altas, médias e baixas (Science Buddies Staff, 2009).</i>	26
3.1	<i>Diagrama de blocos do funcionamento geral do Software</i>	32
3.2	<i>Menu Principal</i>	33
3.3	<i>Funcionamento da Ferramenta Nota e Classificação Vocal</i>	34
3.4	<i>Janela de Informações da Ferramenta Nota e Classificação Vocal</i>	35
3.5	<i>Funcionamento da Ferramenta Treino da Escala</i>	36

3.6	<i>Janela de Informações da Ferramenta Treino da Escala</i>	36
3.7	<i>Funcionamento da Ferramenta Treino de Músicas</i>	38
3.8	<i>Paleta de Transposição e Ajuste de Velocidade</i>	39
3.9	<i>Inserção da Nota Ré Sustenido da Escala 3</i>	40
3.10	<i>Funcionamento da Ferramenta Treino de Músicas</i>	41
3.11	<i>Janela de Informações da Ferramenta Escritor de Músicas</i>	42
4.1	<i>Gráficos de Estatísticas dos Usuários</i>	44
4.2	<i>Média do resultado Geral do MSE para o Treino da Escala musical</i>	55
4.3	<i>Média do resultado Geral do MSE para o Treino da da Música Parabéns</i>	55
4.4	<i>Média do resultado Geral do MSE para o Treino da Melodia Criada</i>	56

# Acrónimos

**A** Lá Maior. xviii, 9, 27

**A/D** Analógico/Digital. 2, 31

**B** Si Maior. 37

**b** Bemol. 23, 24

**C** Dó Maior. 27, 38

**D** Ré Maior. 26, 40

**E** Mi Maior. 27

**ESE** Escola Superior de Educação. 43

**ESTiG** Escola Superior de Tecnologia e Gestão. 43

**F0** Frequência Fundamental. 5, 11, 13–15, 18–20, 27, 29, 35, 37, 41, 59

**Fs** Frequência de Amostragem. 18, 19

**G** Sol Maior. 26, 27

**Hz** Hertz. 19, 26

**iOS** iPhone Operational System. 8

**MATLAB** Matrix Laboratory. 3, 13

**MIDI** Musical Instrument Digital Interface. 7

**MMQ** Método dos Mínimos Quadrados. 42

**MSE** Mean Square Error. 29, 35, 40, 47, 51–53, 56

**N/A** Nenhuma Alternativa. 45

**SFF** Speaking Fundamental Frequency. 26

**TD** Taxa de Diminuição. 52–54

**Ts** Intervalo de Amostragem. 18

**UI** User Interfaces. 13

**VFT** Visual Feedback Technology. 4–8, 13, 31, 59



# Capítulo 1

## Introdução

A afinação da voz para um cantor é de extrema importância para seu desempenho, dentre todos os aspectos que envolve o canto, como a fisiologia da voz, a respiração, a concentração, dentre outros. (Oliveira e Lima, 2016).

O corpo humano utiliza basicamente dos pulmões, laringe e trato vocal para a produção da voz. A Figura 1.1 ilustra os principais componentes do aparelho fonador humano (Uliani Neto *et al.*, 2013).

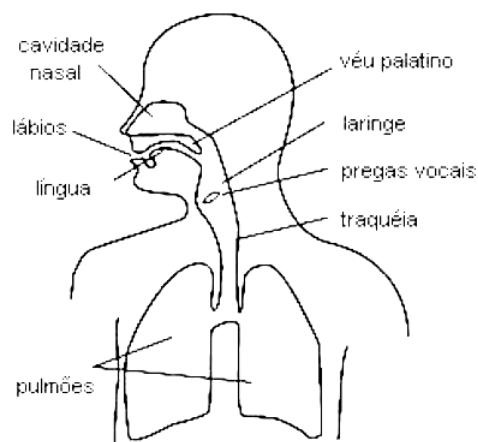


Figura 1.1: *Aparelho fonador humano.*

O aparelho fonador é composto por vários sistemas, tais como o sistema respiratório (fonte de energia, no qual o ar é expelido pelos pulmões), o sistema fonador (fonte

de vibrações, representado pelas pregas vocais), o sistema de ressonância (inclui a cavidade oral e nasal), o sistema articulatorio (inclui a língua, lábios, mandíbula, palato e dentes) e o sistema nervoso central e periférico (como o córtex, responsável pela coordenação)(Almeida, 2010). O som vozeado e filtrado, é amplificado e modulado pelos órgãos do trato vocal para a formação final da voz (Brockmann-Bauser, 2011)(Alves, 2016).

A distinção dos sons vocais subdivide-se em sons produzidos, originados da vibração das pregas vocais através da passagem de ar (sons vozeados) e de sons produzidos por turbulência gerada na passagem do ar por uma secção do trato vocal estrangida (sons não vozeados). Os primeiros possuem periodicidade e uma estrutura regular na frequência, o que não é o caso em sons não vozeados, que apresentam em geral uma estrutura ruidosa, espalhando a sua energia por todo o espectro (Ferreira, 2003).

A maioria dos recursos de extração dos sinais vozeados exigem o sinal transformado para analisar a informação do sinal pós-processado. Como o essencial para o projeto é que a resposta seja apresentada em tempo muito curto, métodos de pós-processamento foram desenvolvidos a fim de alterar seus parâmetros de banda de corte à medida em que a escala musical é selecionada (pois altera-se também as frequências mínimas e máximas admitidas em cada intervalo apresentado)(Rabiner e Schafer, 1978). O objetivo do trabalho não é alterar o sinal de voz obtido, mas apenas trabalhar com o sinal da frequência fundamental de voz.

Um sistema de aquisição de dados pode ser basicamente definido como uma ferramenta que mede informações do mundo real, e portanto em sua maior parte analógica. Um sistema de aquisição de dados típico para aplicações 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 (Baptista, 2017). O processo de aquisição de dados aplicados ao *software* desenvolvido busca a recolha da onda sonora produzida pelo aparelho fonador.

Após o surgimento do áudio processado de maneira digital, o sinal amostrado passou a ser manipulado através de *software*, já que toda a transformação do sinal *Analógico/Digital* (*A/D*) ocorre no *hardware*, ou seja, na placa de áudio do computador. O processamento

e a utilização final dos dados são feitos através do software de processamento, no caso o *Matrix Laboratory (MATLAB)*. O processo com os elementos descritos são exibidos na figura 1.2.

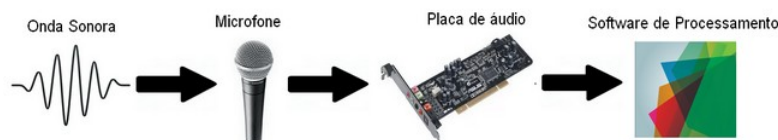


Figura 1.2: *Elementos de um sistema de aquisição de dados.*

A saída tratada é um sinal digital obtido no tempo discreto através de amostras. Esta amostragem é o processo pelo qual o sinal de voz contínuo no tempo é amostrado pela medição da sua amplitude em instantes pré-determinados de tempo discretos. A aproximação dos valores obtidos, para um conjunto finito de níveis, é chamado de quantização. A designação de cada nível quantizado por um dado código é chamado de codificação. A partir das amostras obtidas por este processo, pode-se quantificar as amplitudes, que se apresentam em intervalos contínuo de valores, chamados níveis de quantização (Hamid, 1997) (Panazio, 2009).

Existem atualmente diversos materiais acadêmicos para o treino da afinação da voz em modo literário, assim como aplicações computacionais para a afinação da voz em tempo real. Devido à necessidade de aquisição de vários softwares para desenvolver diferentes recursos específicos, e de uma ferramenta direcionada tanto à alunos que iniciam o estudo em música, quanto a alunos autodidatas sem experiência em teoria musical, nasce a ideia de desenvolver um algoritmo que realiza a obtenção da frequência fundamental da voz, utilizando sistema de aquisição de dados e processamento digital de sinal, onde o estudante de música, em sua fase inicial, pode de forma autodidata se apoiar a fim de melhorar o *pitch*, através da comparação com a frequência das notas de uma melodia monofônica. O *pitch* refere-se à percepção da frequência fundamental e é o inverso da periodicidade das vibrações resultantes das cordas vocais. (Upadhya e Wankhede, 2012) .

Nas ferramentas desenvolvidas busca-se métodos de execução singulares e inovadores do exercício de afinação de voz; Possivelmente o mais significativo deles é a possibilidade

da escolha do usuário utilizar diferentes métodos de treinamento, sendo a detecção da classificação vocal a primeira ferramenta para base do pós treinamento em modo de treino da escala musical crescente ou decrescente. A detecção da nota reproduzida em tempo real é utilizada como afinador e se prova útil como outros instrumentos musicais além do aparelho fonador. O treino de canto com uma base de músicas pré-inseridas trás a facilidade do treino para cantores sem nenhum contato com teoria musical, e por fim, para o auxílio de estudantes de música, na escritura da própria melodia para treino.

## 1.1 Objetivos

São objetivos deste trabalho:

- Desenvolver um sistema de tecnologia *Visual Feedback Technology (VFT)* para realizar um algoritmo de obtenção da frequência fundamental da voz através do método da Autocorrelação;
- Realizar uma análise das técnicas utilizadas para a extração e tratos relevantes do sinal de voz;
- Elaborar um sistema robusto e capaz de captar o sinal utilizando microfones genéricos e mesmo em ambientes de pouco ou médio ruído, distinguindo o silêncio através da análise da energia do espectro de uma janela de observação, que será utilizada também para a aplicação da função da autocorrelação;
- Desenvolver um método de apresentação da informação acústica (Vetor de Notas musicais e sinal de voz) em forma de gráficos.
- Construir uma interface gráfica interativa em apoio à afinação da voz cantada, que permite ao usuário a escolha do método de treinamento vocal através de várias ferramentas desenvolvidas, como a inserção de notas musicais para a composição de melodias que poderão ser salvas e treinadas 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;
- Testar com utilizadores de grupos diversos cada parâmetro do software em diferentes aspectos gerais e específicos, em cada ferramenta proposta, afim de avaliar a eficiência e a utilidade do sistema.
- Realizar um questionário baseado no tradicional método de avaliação de usabilidade e técnicas de avaliações heurísticas desenvolvido por Nielsen e Molich (J. Nielsen, 1993) (R. Nielsen J. M., 1990).

## 1.2 Estado da Arte

Os recursos para treinamento de voz são diversos e se disponibilizam em diferentes formas de aquisição. Nesta seção aborda-se as pesquisas acadêmicas relacionadas à obtenção da frequência fundamental da voz e seu processamento, assim como diferentes programas em formato executável, aplicativo para *smartphone* ou *CD-ROM*.

### 1.2.1 Produções e Pesquisas Acadêmicas

O primeiro projeto a incorporar a *VFT* em tempo real em um *software* com fins educacionais para canto foi o de Howard & Welch (Howard e Welch, 1989). O principal objetivo do projeto foi de desenvolver uma ferramenta pedagógica para o desenvolvimento do canto vocal em uma escola primária do Reino Unido utilizando a tecnologia VFT. O *software* foi projetado para um microcomputador BBC e fez uso de um pico-seletor, originalmente projetado para implantes cocleares para fazer uma estimativa precisa da *Frequência Fundamental (F0)* em tempo real. A referência musical do sistema eram bibliotecas musicais com três ajustes de oitavas. Uma simples nota era tocada para a qual o usuário tinha de responder cantando em um microfone. A média de *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. As notas selecionadas podiam, opcionalmente,

ser colocadas na tela para orientar a resposta cantada (Howard e Welch, 1989)(Hoppe, Sadakata e Desain, 2006).

A pesquisa de Jan Bartošek e Václav Hanžl (2010) procura encontrar uma resposta proveniente da comparação de diferentes algoritmos de detecção de pitch (PDAs) na área de processamento de sinais de fala. Um quadro de avaliação foi desenvolvido utilizando um conjunto de comparações de critérios estabelecidos. 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 realizado por Sá Ferreira (2012) estabelece, com a ajuda dos *softwares SingingStudio, Praat e Matlab* e de uma base de dados de diferentes sinais de voz, a relação entre os parâmetros perceptivos da voz cantada (como altura, brilho, entre outros) e os fenômenos acústicos objetivos, obtidos através das análises das gravações. Os parâmetros escolhidos para análise foram o portamento, o falsete, a limpidez, o formante de Cantor e o vibrato. Os resultados obtidos através das estatísticas tratam, dentre vários parâmetros acústicos, o *Jitter* e *Shimmer* e Autocorrelação. O resultado da autocorrelação média é obtido utilizando o teste de Kruskal-Wallis para a comparação entre o registro de falsete e modal.

Em sua pesquisa, Neves (2016) realizou o processamento do sinal de voz após gravação, a fim de propor uma ferramenta com interface gráfica de afinação de sinais monofônicos, onde é possível alterar a altura e duração das notas. A ferramenta utilizada para a alteração da altura, sem causar interferências na sua duração é o *vocoder* de fase, e de reamostragem. O cálculo de curva de frequência fundamental ao longo de toda a duração do sinal é realizada pelo método YIN.

O estudo de Leong e Cheng (2014) utilizou a tecnologia *VFT* no treinamento vocal, em busca de dois objetivos: averiguar a eficácia do *software* de *feedback* visual em tempo real 'Sing & See' no treinamento vocal de professores de música e analisar a perspectiva destes sobre sua experiência com este tipo de interface. Quarenta participantes de um programa de graduação de professores de música foram designados aleatoriamente para três grupos. Os docentes expressaram uma atitude positiva em relação ao uso do VFT

como um componente de tecnologia de realce do aprendizado em educação musical.

## 1.2.2 Aplicações e Softwares

### Canta

*Canta* (Canta, 2017) é um software livre que possibilita escolher qualquer arquivo monofônico em formato *Musical Instrument Digital Interface (MIDI)*, mostrando assim as respectivas notas a serem cantadas em tempo real. Ele analisa e mede a frequência da voz e mostra a altura em que se canta em forma de uma curva na tela, como pode ser visto na figura 1.3. A curva de altura é sobreposta às notas da melodia escolhida. Por outro lado, o *software* não dispõe de nenhum exercício vocal já registrado, e não calcula o erro, ou seja, a diferença do *pitch* do cantor e das frequências identificadas do arquivo *MIDI*.

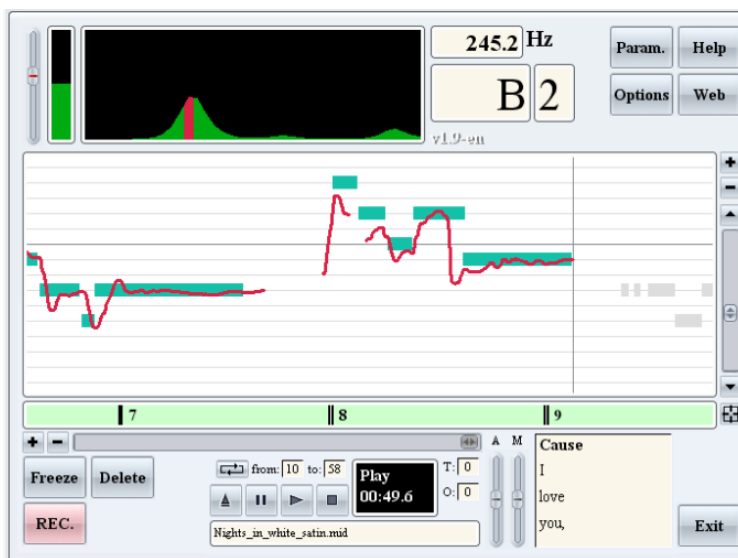


Figura 1.3: Exemplo de funcionamento do software Canta.

### Sing and See

O projeto *Sing and See* foi apresentado na Conferência de Musicologia Interdisciplinar (Callaghan *et al*, 2004). Este visava desenvolver a nova tecnologia *VFT* para o estudo de

canto. As principais características da pesquisa foram a investigação de técnicas de análise acústica, métodos de exibição de *VFT* de forma significativa e as abordagens pedagógicas para implementar a tecnologia em prática.

O programa é projetado para atender às necessidades dos cantores e professores de músicas de todos os estilos e gêneros. Ele possui uma ferramenta chamada *Mirror of the voice* (Espelho da voz), mostrado na figura 1.4, onde se dá o *feedback* em tempo real do tom em que está sendo emitida a nota vocal, assim como o espectrograma (Um gráfico que representa as estruturas harmônicas, ou seja, o timbre da voz). A função *Font of Knowledge* (Fonte de Conhecimento) é uma fonte de informação interativa que ajuda o estudante de música a entender a voz, utilizando o espectrográfico que explora as diferentes estruturas harmônicas e seus respectivos sons.

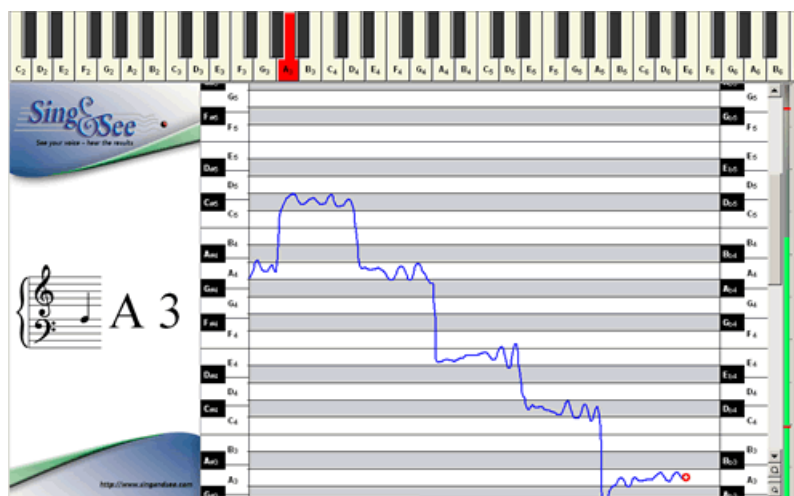


Figura 1.4: Exemplo de funcionamento do software Sing and See.

## Sing True

Na figura 1.5 é possível ver a ferramenta principal do aplicativo *Sing True* (Sing True, 2014). Este aplicativo é disponível para *iPhone Operational System (iOS)*, com três ferramentas principais de aprendizado: Treinamento da escuta, com testes de escolha da altura de cada tom produzido, treinamento da mente (onde ele propõe treino para o controle da respiração) e, por fim, o treinamento da voz, onde mostra-se em tempo real o

*pitch* de cada nota, oferecendo ao fim o resultado final, com as marcações e avaliações do estado de exatidão do exercício. É possível ainda, acompanhar o gráfico de progresso da afinação e do *pitch*.

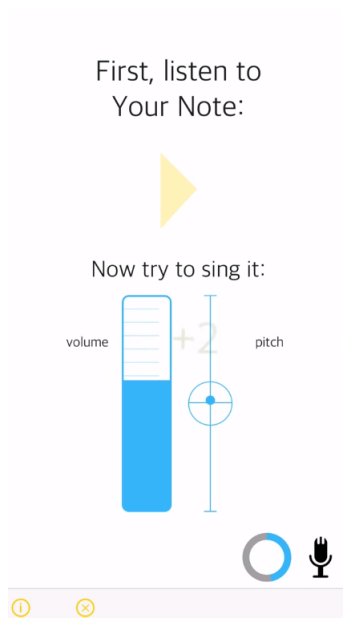


Figura 1.5: *Sing True*: Exercício de treinamento do acorde A.

## Media Singing Method

*Media Singing Method* (Media Singing Method, 2017) é um *software* disponível em CD-ROM, propondo diversas ferramentas para cantar no tempo e no tom correto. As lições propostas ajudam a realçar o tom vocal, melhorar a projeção da voz, e apurar a dicção. O aplicativo também processa o sinal da voz e mostra se está de acordo com o esperado quando se canta uma das 80 músicas disponíveis em sua base de dados (figura 1.6). Além destas músicas, são propostas algumas lições vocais em forma de texto para intensificar o estudo do canto profissional.

Figura 1.6: *Media Singing Method: Acompanhamento vocal da ópera Ave Maria, Schubert.*

## Pitch Perfector

O programa (Pitch Perfector, 2009) é apresentado de forma fácil e possui um propósito: identificar o tom de voz do cantor. A ferramenta dispõe de um guia visual para identificar se o tom cantado condiz com o tom desejado, e ainda permite alterar a escala de acordo com a altura da pauta musical, sendo preciso portanto, identificar o tipo de voz, podendo variar entre Soprano, Contralto, Tenor e Baixo. O programa permite ainda, realizar a gravação do treino de voz realizado.

## 1.3 Estrutura do Relatório

Este trabalho encontra-se dividido em cinco capítulos. O primeiro capítulo introduz o leitor ao assunto, descrevendo as principais definições referentes ao problema, bem como os objetivos principais a serem atingidos durante o desenvolvimento e apresenta, o estado da arte se encontra subdividido entre as produções e pesquisas acadêmicas e em aplicações comerciais que utilizam software de apoio à afinação da voz.

No segundo capítulo são apresentados os métodos utilizados em cada processo. Após uma breve apresentação acerca do funcionamento do *software* MATLAB e da justificativa para a escolha do mesmo, o item 2.2 explicita o processo de aquisição do sinal de voz,

desde a determinação da frequência de amostragem, ao método de aquisição do sinal nas duas versões do MATLAB, 32 e 64 Bits. No item 2.3 é apresentada a determinação da frequência fundamental, utilizando o método da autocorrelação e a opção do uso do *center-clipping* para melhor determinação. No item 2.4 é mostrado o procedimento para o pós-processamento da  $F0$ , como as verificações em relação ao  $F0$  máximo e mínimo estipulados, e a detecção do silêncio ao analisar a energia do sinal. Por fim, no item 2.5 a teoria musical utilizada para a aplicação do *software* em aspecto pedagógico é detalhada, assim como o processo de criação da melodia monofônica para auxiliar na utilização da ferramenta, e o erro quadrático médio mostrado ao final de cada treino.

O terceiro capítulo dispõe sobre a organização geral do *software* ao apresentar o diagrama de blocos do funcionamento geral do programa, assim como o menu principal. As quatro aplicações são mostradas em seguida, providas da explicação do funcionamento teórico e em imagens.

O quarto capítulo descreve a performance do *software*, avaliada a partir dos resultados obtidos em um teste realizado na instituição. Após a apresentação da caracterização da amostra no item 4.1, o item 4.2 traz em detalhes o processo de avaliação do *software* e os resultados gerais e específicos para cada ferramenta. No capítulo 4.3, o resultado dos treinos experimentais são divididos nas quatro aplicações existentes, assim como a discussão e análise da evolução dos resultados em modo geral.

O quinto e último capítulo conclui a eficiência da aplicação do método a partir dos resultados obtidos e da funcionalidade geral do *software*, assim como a utilidade da ferramenta orientada para cada categoria. Além disso, são discutidos os possíveis trabalhos futuros de melhorias nas ferramentas partindo dos desenvolvimentos já realizados.



# Capítulo 2

## Metodologia

### 2.1 O MatLab

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

O *MATLAB* foi escolhido como a ferramenta de construção do *software* pela facilidade na utilização de funções matemáticas pré inseridas na ferramenta, e na construção de interface gráfica utilizando a ferramenta GUI para projetar interfaces de usuário, como as *VFT*. O GUIDE gera automaticamente o código MATLAB para a construção da *User Interfaces (UI)*, permitindo a alteração do programa de acordo com o comportamento do aplicativo (Inc., 2014a.).

### 2.2 Frequência Fundamental

A  $F_0$  é o parâmetro físico resultante da vibração das pregas vocais por unidade de tempo (Sellam e Jagadeesan, 2014) (Lopes, 2008). A respiração criada nos pulmões percorre as cordas vocais e o pitch da voz é dirigido por pulsos glutais. Finalmente, este é filtrado por cavidades nasais (nasal e oral) que atuam como ressonadores com as frequências dos

formantes. Desta maneira, as seções de voz contendo F0 são criadas (Bartošek e Hanžl, 2010).

O desafio de estimar a frequência fundamental de sinais de fala ocupa uma posição chave na área de pesquisa de processamento (Pinto e Tomé, 2001). Esta estimação é importante para determinar e comparar as diferentes captações da voz humana. As medidas da  $F_0$  variam de acordo com o gênero e idade da voz humana captada, sendo que em mulheres, o comprimento da prega vocal é mais curto. As medições de frequência da voz cantada ainda podem ser afetadas por fatores como a atividade muscular, a pressão subglótica e a viscosidade dos tecidos da prega vocal (Colton e Woo, 1995) (Woo *et al.*, 1994) .

Existem atualmente dois principais métodos de análise da F0. Os algoritmos de detecção da F0 são classificados nas seguintes categorias básicas: detecção baseada no domínio do tempo, detecção baseada no domínio da frequência ou detecção baseada no domínio de tempo-frequência em conjunto (Upadhye e Wankhede, 2012).

Um dos métodos conhecidos é a análise cepstral. O método do Cepstro consiste em uma operação matemática que extrai a Transformada de Fourier do espectro do sinal em logaritmos. O nome “cepstrum” obtém-se ao inverter a ordem das primeiras quatro letras de “spectrum”. A análise Cepstral do um sinal de fala permite adquirir dois sinais: O sinal da glote (excitação) e o sinal de trato vocal (ressonância) separadamente, pelas suas propriedades homomórficas, facilitando o estudo das alterações nas cordas vocais. O cepstrum é uma função que satisfaz o princípio geral da sobreposição com a convolução como entrada e a adição como a saída. (Kent e Read, 2002) (Teixeira *et al.*, 2011).

Existem várias razões pelas quais o método da autocorrelação é geralmente utilizado. O cálculo da autocorrelação é feito diretamente no forma de onda e é de simples implementação (Rabiner, 1977)(Pinto e Tomé, 2001). Após breves ensaios de desempenho, o método escolhido para a determinação da frequência fundamental foi o método da autocorrelação. Um exemplo de aquisição da frequência fundamental utilizando o método da autocorrelação é descrito no item 2.2.1.

### 2.2.1 Método da Autocorrelação

Para realizar a estimativa do valor de frequência fundamental é necessário obter o máximo da função de autocorrelação (Tan e Karjandecha, 2003). Dado um sinal discreto  $x(n)$  a função de autocorrelação é definida pela eq. (2.1) (Sellam e Jagadeesan, 2014).

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

A autocorrelação é definida como a distribuição da magnitude de um espectro de sinal num domínio de tempo. Ou seja, a autocorrelação de um sinal é a transformação inversa de Fourier de um espectro de energia distribuída do sinal estudado.

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

Assim, para o cálculo do frequência fundamental assume-se que  $x(n)$  é periódico com período  $P$ , isto é  $x(n)=x(n+P)$  para todo o  $n$ , então a função de autocorrelação é também periódica com o mesmo período, eq. 2.2 (Sellam e Jagadeesan, 2014) (Tan e Karjandecha, 2003).

A figura 2.1 mostra a extração do sinal reproduzindo o som "a". O quadro de análise ideal deve conter de dois a três períodos glotais (Pinto e Tomé, 2001) (Teixeira *et al.*, 2011). É selecionada uma janela representativa para observação de pelo menos dois períodos glotais. Assumindo a  $F0$  mínima de 65 Hz, dois períodos correspondem a  $2*1/65=30$ ms, como mostra a figura 2.2.

Em seguida aplica-se a função de autocorrelação que usa a Eq. 2.1, realizada pela função `xcorr`. Do resultado da autocorrelação tem-se a caracterização da similaridade de um sinal com a cópia deslocada de si mesmo, onde o máximo valor é o somatório de todos os outros pontos, localizado sempre ao meio.

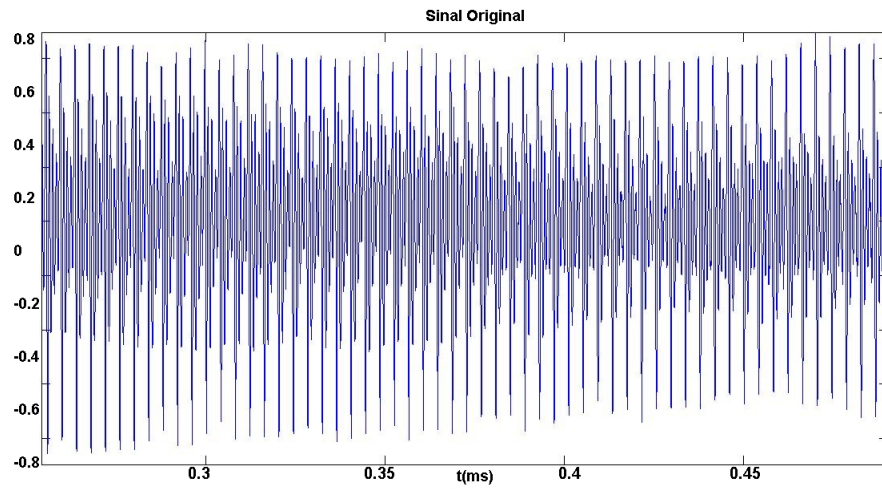


Figura 2.1: *Sinal Original.*

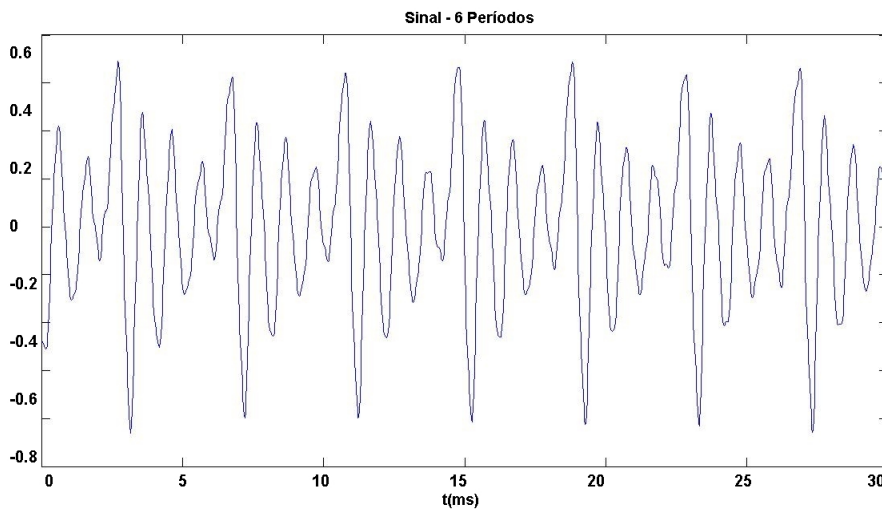


Figura 2.2: *Janela de observação com um período de 30ms do sinal original.*

## 2.2.2 Autocorrelação com Center-Clipping

Uma limitação do método da Autocorrelação é que ele retém muita informação dos sinais de fala e, portanto, a função de autocorrelação apresenta muitos picos, em razão das oscilações amortecidas da resposta do trato vocal, responsáveis pela forma de cada período da onda de fala. Portanto, quando os picos de autocorrelação são maiores do que aqueles devidos à periodicidade da excitação vocal, o procedimento simples de escolher o maior pico na função de autocorrelação falhará (Rabiner e Schafer, 1978) (Upadhyha e Wankhede, 2012).

Afim de melhorar a decisão do máximo da autocorrelação, pode-se recorrer a utilização do chamado "*center-clipping*". Retira-se as amostras negativas de uma certa porcentagem da amplitude máxima para uma melhor evidência dos valores máximos. O resultado pode ser observado na figura 2.3.

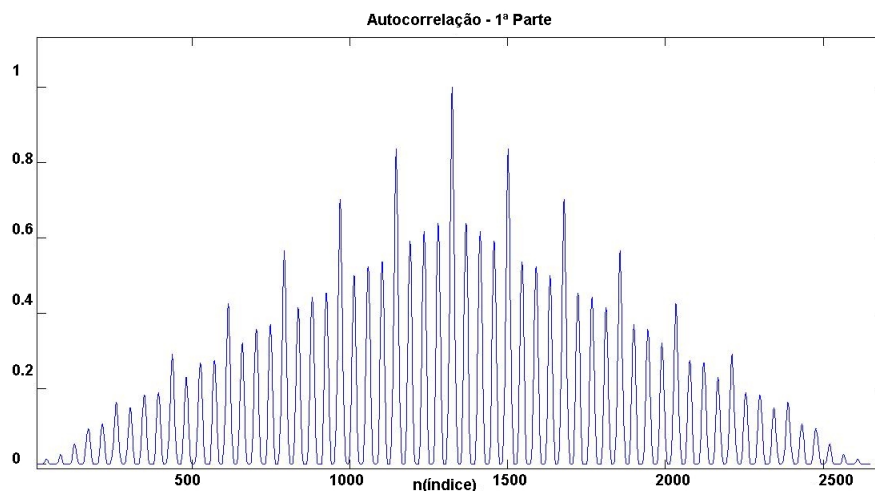


Figura 2.3: Resultado da Autocorrelação do Sinal utilizando Center-clipping.

A determinação do período glotal dá-se pela distância do pico máximo da autocorrelação e o próximo pico mais alto. O inverso do período glotal é a frequência fundamental.

Nota-se na figura 2.4, que o período glotal (eixo x) é de 4.059 ms. Então, o inverso

deste valor é transformado para segundos, resultando na  $F0$  em Hz.

$$T0 = 4.059ms \quad (2.3)$$

$$F0 = (1/4.059) * 1000 \quad (2.4)$$

$$F0 = 246,366Hz \quad (2.5)$$

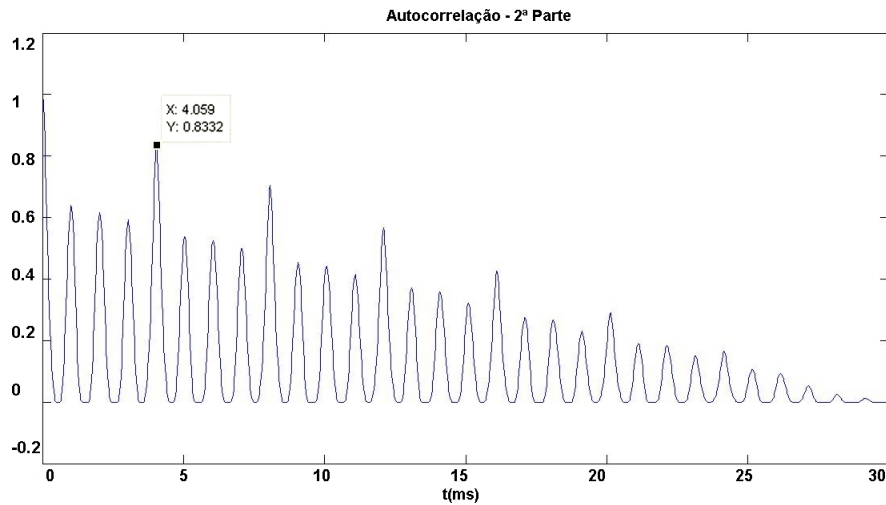


Figura 2.4: Análise do período glotal a partir do máximo da Autocorrelação.

Este valor é apresentado ao usuário final como uma reta. Posteriormente cada valor detectado de  $F0$  é adicionado no gráfico em tempo real, formando o vetor de demonstração do sinal de voz.

## 2.3 Aquisição do Sinal de Voz

### 2.3.1 Frequência de Amostragem

Para um sinal analógico, é necessário recolher amostras em determinado intervalo de tempo. Este intervalo é chamado *Intervalo de Amostragem* ( $T_s$ ). O seu inverso é a denominada *Frequência de Amostragem* ( $F_s$ ), ou seja,  $1/T_s$  amostras por segundo. Para

a reconstrução do sinal analógico, a  $F_s$  precisa ser, no mínimo, o dobro da frequência máxima contida no sinal analógico. Este é o chamado Teorema da Amostragem. A frequência mínima para amostrar um sinal é denominada *Frequência de Nyquist* (Shannon e Weaver, 1949).

O canto lírico feminino apresenta valores maiores em frequência do que a fala. O valor médio da  $F_0$  da fala gira em torno de 220 *Hertz (Hz)*, enquanto as notas emitidas por uma soprano podem estar próximas a 1024 *Hz* (Dó Maior da Escala 5) (Benolken e Swanson, 1990) (Vieira, 2004). Assim sendo, a *Frequência de Nyquist* é calculada em 2048 Hz. Como é um range de frequência considerado baixo, a capacidade de processamento de sinal não é afetada pela extrapolação deste valor, por outro lado, a qualidade das amostras aumenta (Oppenheim e Schaffer, 2009). Para as frequências de voz mais elevadas, é necessário mais amostras para melhor mapeamento do sinal de  $F_0$ , já que para frequências mais altas o mesmo número de amostras provoca resultados de frequência mais distantes do real. Por estas razões, a frequência de amostragem escolhida para as diferentes aplicações foi de 22050Hz.

### 2.3.2 Aquisição do sinal em MatLab 32 Bits

A aquisição do sinal através da placa de som é feita através da inclusão de um canal analógico, que cria um objeto de entrada seguido de um canal de aquisição de áudio.

```
AI=analoginput('winsound');  
chan = addchannel(AI,1);
```

A variável *Size* vista a seguir é uma excessão proveniente do cálculo de regressão linear dos resultados de um experimento de 35 repetições da mesma escala, a fim de calcular a duração de acordo com a variável "ver", que recolhe o valor global de tamanho, e este, por sua vez, varia de acordo com cada aplicação.

```
ver = getGlobalSize;  
Size = char((0.0154*ver)+0.2507);  
duration = Size;
```

Parâmetros como a taxa de amostragem e número de amostras por trigger, são definidos de acordo com a duração da aquisição e a frequência de amostragem, que é configurada de modo distinto para cada aplicação desenvolvida, como apresentado a seguir:

```
set(AI, 'SampleRate', fs);
set(AI, 'SamplesPerTrigger', duration*fs);
preview = round(fs/22);
set(gcf, 'doublebuffer', 'on');
set(findall(gca, 'Type', 'Line'), 'LineWidth', 2);
```

O valor da escala é definido assim como o vetor de notas, que podem vir tanto de matrizes pré-atribuídas quanto de variáveis globais construídas em tempo real, dependendo da aplicação da aquisição escolhida. O *sound* com a primeira nota referente a cada aplicação é evocado em seguida de um *pause* de 1,5 segundos.

Em seguida, a aquisição do canal é iniciada, e um ciclo *while* ocorre enquanto o número de amostras adquiridas não é igual ao número de amostras desejadas. Os dados são recolhidos do canal.

É feito o trato de energia do sinal para definir o silêncio e sons totalmente não-vozeados. Se a energia for baixa o valor é considerado nulo, os valores que contém informações consideradas vozeadas são então tratados em uma função `F0autocorr` que devolve o valor máximo da autocorrelação, como descrita em 2.1.

Afim de diminuir o ruído, um `F0max` e um `F0min` foram estipulados de acordo com o estudo realizado no capítulo 2.4. Se o valor atual (`F0_1`) de autocorrelação for maior que o valor de `F0` máximo estipulado, então `F0_1` é substituído pelo anterior. O mesmo é feito para o valor mínimo.

Ao final de todas as verificações precedentes, o erro quadrático médio é calculado.

```
F0max=1100;
F0min=30;
F0_1=0;
```

```

while AI.SamplesAcquired < duration*fs
    data = peekdata(AI,preview);
    set(P,'ydata',data);
    d2=[d2;data];
    set(P2,'ydata',d2);
    yp = data.^2;
    E= sum(yp(:));
    if E < 1
        FO_1= NaN;
    else
        FO_1=F0autocorr(data,fs);
        if FO_1>=F0max,
            FO_1=FO_0;
        elseif FO_1<=F0min
            FO_1=FO_0;
        end
    end
end
erro=mse(F0(10:end)-nota1(10:end))/div;
Erro=strcat('Erro=',num2str(round(erro)));

```

### 2.3.3 Aquisição do sinal em MatLab 64 Bits

O sistema de aquisição do sinal em 64 bits é similar ao de 32 bits. Porém, em 64 bits é necessária a utilização da estrutura daq, que apresenta diversos dispositivos de aquisição de dados. O primeiro passo é atribuir a uma variável os dispositivos do daq, e estabelecer o canal de áudio para a entrada, no caso recebido pela variável dev. O algoritmo de execução é uma modificação do exemplo original disponível no *MathWorks*. Para acessar a placa de som, é necessário criar uma seção com o canal de entrada de áudio, chamada 'directsound'. Esta seção é atribuída como contínua, conforme apresentado no algoritmo

a seguir:

```
d = daq.getDevices;
dev = d(2);
s = daq.createSession('directsound');
addAudioInputChannel(s, dev.ID, 1:2);
s.IsContinuous = true
```

A criação dos elementos utilizados para setar o plot também são declarados. Em seguida, a função de autocorrelação é chamada, recolhendo das estruturas de base os elementos de dados. Os valores são plotados continuamente depois da atribuição do *listener* e do início da aquisição em *Background*, isto é, o programa continua sua execução enquanto os dados são recolhidos, até que seja pausado com o valor da duração, que neste caso é de 10 segundos.

```
f = figure;
p = plot(zeros(1000,1));
T = title('Detecção da F0');
xlabel('Frequency (Hz)')
ylabel('|Y(f)|')
grid on;
F0=@(src, event) F0autocorr(event.Data, src.Rate, p);
```

O mesmo procedimento pode ser aplicado para a comparação de um vetor de notas e o *plot* em tempo simultâneo, assim como o cálculo final do erro quadrático médio. Ao final da aquisição, a seção deve ser parada e desativada do modo contínuo, e o *listener* deletado.

```
l = addlistener(s, 'DataAvailable', F0);
startBackground(s);
figure(f);
pause(10);
```

## 2.4 Teoria Musical

Para realizar a ligação do *software* com a teoria musical, foi reunida a literatura necessária como o conhecimento da escala musical e da sua transposição, o valor do seu ponto em sustenido e bemol, para então se fazer a escolha do tipo de escala utilizada, além da escolha de um método de atenuação da nota para uma melhor reprodução da melodia escrita.

### 2.4.1 Escalas musicais

As escalas musicais são feitas a partir da divisão da sequência de notas contidas dentro de uma oitava. Essa divisão é realizada obedecendo principalmente a critérios estéticos, ou seja, na melodia que as notas formam e as relações harmônicas entre elas (Aleixo e Mansanares, 2003).

A relação harmônica entre as notas pode ser analisada utilizando séries de Fourier, uma vez que combinações de formas sonoras cujas frequências não possuem interferência dentro de uma ou duas oitavas não formarão nenhum tipo de relação harmônica.

Na Figura 2.5, na escala de Dó maior, alguns intervalos foram separados para que se tornassem iguais a  $9/8$ , ou seja, um intervalo de um tom. Porém, os intervalos ré-mi e sol-lá são iguais a  $10/9$ , ligeiramente inferiores a um tom (Aleixo e Mansanares, 2003).

DÓ	RÉ	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 2.5: *intervalo de um tom*

Duas escalas foram criadas a partir da escala natural para resolver o problema das notas intermediárias. A escala justa foi a introdução das notas alteradas, formadas a partir da multiplicação, sustenido ( $\sharp$ ), ou divisão, Bemol ( $b$ ), da nota original por  $25/24$ . Desta maneira, cerca de 1400 intervalos perceptíveis para o ouvido humano se aproximariam de uma das 21 notas da escala natural. A segunda escala foi desenvolvida a partir da divisão

da afinação, mas distribuindo o erro, causado pela forma da divisão da escala, entre notas vizinhas. Dividiu-se a oitava em 12 intervalos iguais à raiz duodécima de 2, ou seja, a razão entre as notas passou a ser de 1,059 entre semitons e 1,122 por tom. Com isso, o dó sustenido se iguala ao ré bemol, o mi sustenido ao fá e o fá bemol ao mi, resultando em 12 notas. Esta escala denomina-se escala temperada.

A distribuição de notas é feita de maneira contínua e, assim sendo, a divisão não altera a análise física dos sons. O que se altera é a posição da frequência fundamental da nota. Por exemplo, a junção do  $Dó\#$  com o  $Réb$  faz com que a frequência fundamental da nota, na escala temperada, seja ligeiramente deslocada para um ponto entre ambas as notas.

O grande problema ligado às escalas diferentes encontra-se quando instrumentos temperados como o piano e o violão tocam com instrumentos não temperados, como o violino, por exemplo. Por isso, para aplicação ao projeto, foram escolhidas as frequências de notas da escala temperada, utilizada em instrumentos mais comumente acompanhados pela voz e mais fácil de ser feita a transposição, caso o instrumento utilize a escala justa (Aleixo e Mansanares, 2003).

A Tabela 2.1 mostra a diferença entre as frequências e as relações entre notas nas escalas justa e temperada. Observa-se, desta maneira, os intervalos suprimidos da escala justa para o reajuste da escala temperada.

### 2.4.2 Extensão Vocal

A extensão vocal abrange todas as notas fisicamente realizáveis pelo cantor. A tessitura abrange um *range* de notas mais apropriadas à qualidade do canto. Dentro dos sistemas de classificação, existem seis tipos básicos de voz, embora os intervalos sejam aproximados e não devam ser aplicados de forma muito rígida (Sadie, 1985)(Peckham, 2005).

Cada extensão abrange uma faixa de frequência adequável ao tipo de voz do cantor. Geralmente homens por possuírem a prega vocal mais longa, apresentam frequência vocal de menor valor, e portanto se encaixam em *tenor*, *barítono* e *baixo*. Análogo às mulheres,

<b>Nota musical</b>	<b>Intervalo com a nota fundamental</b>	<b>Afinação natural (Hz)</b>	<b>Afinação temperada(Hz)</b>
Dó	Dó uníssono	132	132
Dó #	Semitom	137	139.7
Ré b	Segunda diminuta	142.5	139.7
Ré	Segunda maior	148.5	148.1
Ré #	Segunda aumentada	154.7	156.9
Mi b	Terça menor	158.4	156.9
Mi	Terça maior	165	166.3
Fá b	Quarta diminuta	168.9	166.3
Mi #	Terça aumentada	171.8	176.2
Fá	Quarta perfeita	175.9	176.2
Fá #	Quarta aumentada	183.3	186.6
Sol b	Quinta diminuta	190.1	186.6
Sol	Quinta perfeita	198	197.7
Sol #	Quinta aumentada	206.3	209.4
La b	Sexta menor	211.2	209.4
Lá	Sexta maior	220.0	222
Lá #	Sexta aumentada	229.2	235.2
Si b	Sétima menor	237.6	235.2
Si	Sétima maior	247.5	249.2
Dó b	Oitava diminuta	253.4	249.2
Si #	Sétima aumentada	257.7	264
Dó	Oitava perfeita	264	264

Tabela 2.1: Relação entre as notas musicais, intervalos e frequências correspondentes. Extraído de (Aleixo e Mansanares, 2003).

com pregas vocais mais curtas, e então frequência vocal de maior índice, normalmente adequadas aos grupos *soprano*, *mezzo soprano* e *alto* (Nadoleszny, 1923).

A faixa de notas do grupo de menor frequência (baixo), ao grupo de maior frequência (alto), representam as escalas nomeadas do um ao cinco. Este padrão toma como referência o lá central a 440 *Hz*, como recomendado pela *American Standards Association*, sendo esse o lá da escala quatro (Martin, 2008).

Nadoleszny (1923) reuniu um grande número de dados de *Speaking Fundamental Frequency (SFF)* extraídos de diferentes cantores de seis classificações diferentes (11 baixos, 17 barítons, 29 tenores, 18 contraltos, 13 mezzo-sopranos, 55 sopranos). Ao escutar diferentes discursos de cada cantor, ele comparou as diversas *SFF* de cada um com o padrão de *pitch* das amostras utilizando a frequência em Hertz de um diapasão. Cada classificação apresentou diferentes SFF, sendo baixo = 98 *Hz*; barítonos = 116 $Hz$ ; tenores = 134 *Hz*; contraltos = 212 *Hz*; mezzos-sopranos = 230 *Hz* e sopranos = 262 *Hz*.



Figura 2.6: *Pauta Geral correspondente aos intervalos vocais de cantores. A linha superior corresponde a vozes femininas altas, médias e baixas, e a linha inferior corresponde a vozes masculinas altas, médias e baixas* (Science Buddies Staff, 2009).

As classificações foram selecionadas de acordo com o range de frequência e cada uma possui um alcance vocal específico (Camargo *et al*, 2007) (Junker, 1999) :

## Baixo

Voz masculina do tipo mais grave e de menor alcance vocal.

Classificação: *Sol Maior (G)2* a *Ré Maior (D)3* - 97 *Hz* a 147 *Hz*.

## **Barítono**

Voz masculina intermediária, entre as extensões de baixo e tenores.

Classificação: *A2* a *Mi Maior (E)3* - 108 Hz a 165 Hz.

## **Tenor**

Voz masculina mais aguda, produzida naturalmente, sem a utilização de falsete.

Classificação: *Dó Maior (C)2* a *A3* - 65 Hz a 220 Hz.

## **Contralto**

Voz feminina de frequência mais baixa no range vocal.

Classificação: *G2* a *C4* - 98 Hz a 262 Hz.

## **Mezzo Soprano**

Voz feminina intermediária, entre as extensões Contralto e Soprano.

Classificação: *A2* a *E4* - 110 Hz a 330 Hz.

## **Soprano**

Voz feminina mais aguda e apresenta o maior alcance vocal dentre todas as outras classificações vocais.

Classificação: *C3* a *A4* - 130 Hz a 440 Hz.

## **2.5 Pós-Processamento da F0**

Com o valor de  $F0$  em Hz, é feita a comparação com as Frequências Fundamentais máxima e mínima, a fim de filtrar possíveis anomalias de detecção da frequência fundamental. Estes dados são atualizados em tempo real e comparados com o vetor de notas previamente atribuído, e o procedimento é repetido enquanto há aquisição de dados no canal de entrada. O código deste procedimento é expresso a seguir:

```

if FO_1>=FOmax,
    FO_1=FO_0;
elseif FO_1<=FOmin
    FO_1=FO_0;
end
end

```

Onde a  $FO\_0$  é o valor de  $FO$  frequência máxima assumida anteriormente. A detecção de um silêncio em uma amostra de sinal vocal é realizada analisando a energia do sinal a cada extração da janela de observação, antes da aplicação da função de autocorrelação.

Considerando uma janela de observação de um sinal  $x$ , amostrada em  $t$  segundos, a energia deste sinal é descrita por (Lathi, 2006):

$$Ex = \int_{-\infty}^{+\infty} [x(t)]^2 dt \quad (2.6)$$

Se o valor da energia presente na janela é de valor muito baixo, trata-se de um silêncio, e este é, desta forma, representado como um valor indefinido NaN.

A taxa de passagem por zero também é um indicador de silêncio ou fala, e está concentrada no espectro do sinal. O canto é produzido por causa da excitação de trato vocal pelo fluxo periódico de ar na glote e geralmente mostra uma baixa taxa de passagem por zero (Marvan, 2007). Considerando que o som não vozeado é produzido pela constrição do trato vocal estreito o suficiente para causar fluxo de ar turbulento que resulta em ruído apresentando então alta taxa de passagem por zero, que pode ser calculado através da expressão 2.7 (Bachu, Kopparthi e Barkana, 2010).

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]|w(n-m) \quad (2.7)$$

onde

$$sgn[x(n)] = \begin{cases} 1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases}$$

### 2.5.1 Composição da Melodia

A escritura da melodia monofônica como apoio sonoro para o treino das ferramentas, deu-se ao unir as notas em compassos binários de mínima, semi-mínima, colcheia e semi-colcheia (Cardoso e Mascarenhas, 1973). A atenuação do sinal de cada nota representada no domínio das frequências se faz utilizando uma Janela de Hanning.

Esta janela permite de certa forma suprimir o vazamento espectral e aumentar a precisão do sinal de saída (Wen *et al.*, 2009). Cada sinal referente à uma nota é convoluído com a janela Hanning para o efeito de atenuação requerido, e se dá pela expressão 2.8 (Teixeira, 2018). Esta operação é necessária para melhorar a qualidade de reprodução deste som, ocorrendo o melhor acoplamento entre os sinais referentes à cada nota.

$$M(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} |x(m)| \cdot w(N+1) \quad (2.8)$$

### 2.5.2 Erro Quadrático Médio

O *Mean Square Error (MSE)* foi escolhido como resultado dado ao final de cada treino. Ele é representado pela equação 2.9, onde  $\bar{X}$  é o vetor de notas selecionado para treinamento,  $X$  é o vetor  $F0$  e  $n$  é o número de amostras (Wackerly, Mendenhall e Scheaffer, 2008).

O *MSE* realiza a média de um conjunto de erros, ao calcular a distância entre dois conjunto de pontos e ajustar então o erro através da distância entre as duas linhas analisadas. A quadratura ocorre para a remoção dos sinais negativos e ao aumento de peso aos erros maiores. O uso desta função apresenta a vantagem da tratabilidade analítica, uma vez que os valores de erro e o seu peso são bem definidos, embora possa haver a possibilidade de uma certa dificuldade por parte do usuário final na leitura destes dados e na assimilação dos valores com o processo evolutivo da aprendizagem.

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_i)^2 \quad (2.9)$$



# Capítulo 3

## Aplicações

### 3.1 Organização Geral do Software

A organização geral do software pode ser visualizada em dois modos. O diagrama de blocos exibe todas as funções do software de maneira geral e permite a análise dos sistemas de modo simplificado. O menu principal é a maneira com que o usuário tem disponível todas as ferramentas disponíveis para treino.

#### 3.1.1 Diagrama de Blocos

O diagrama de blocos do funcionamento geral do Software é disposto na figura 3.1. Neste modelo, as decisões e processos das ferramenta são disponibilizados, levando ao mesmo processo pré-definido do *VFT* de comparação entre o vetor de notas e o sinal vozeado.

É possível notar os dois processos independentes da escolha do usuário feita no menu, que são a aquisição de dados do Microfone e a conversão *A/D* na placa de áudio. Ao final do *VFT* de cada aquisição é possível, em todas as janelas, retornar ao menu principal ou sair do programa.

**Sistema de Apoio à Afinação da Voz**

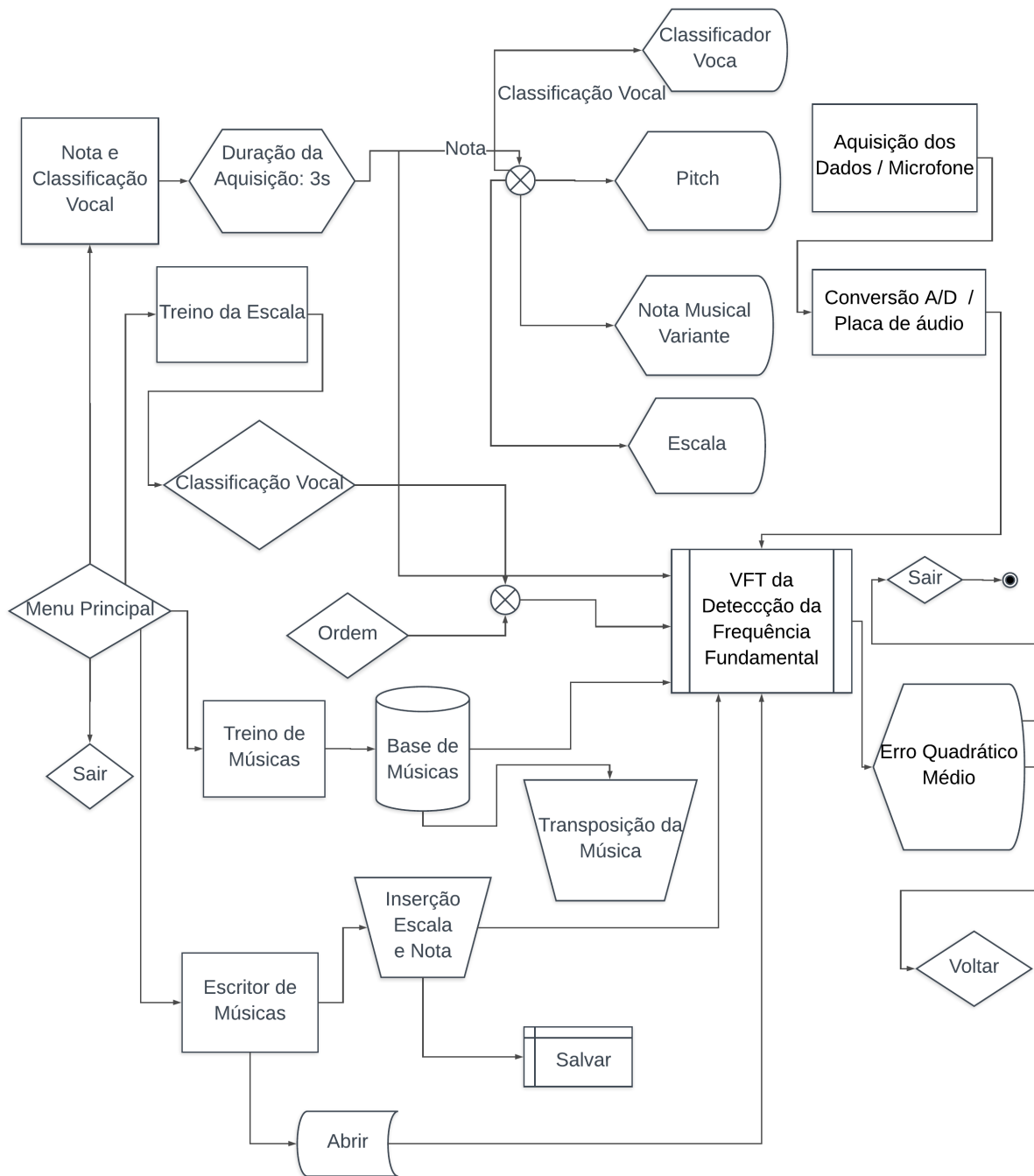


Figura 3.1: Diagrama de blocos do funcionamento geral do Software

### 3.1.2 Menu Principal

O menu principal apresenta de forma clara todas as ferramentas propostas, como mostra a figura 3.2. Ele é a janela de apresentação do software, e possui os botões de entrada para as quatro aplicações propostas.



Figura 3.2: *Menu Principal*

## 3.2 Nota e Classificação Vocal

A primeira ferramenta busca informar o usuário de forma rápida e em tempo real, os parâmetros extraídos de cada som produzido, nas quais o pitch (Frequência em Hz da nota vocal produzida), a nota musical, a escala e a estimativa da Classificação Vocal. O tempo de aquisição de cada ensaio é de 3 segundos.

Após feita a aquisição de uma janela de observação e a determinação da frequência

fundamental, o valor é analisado por ciclos de controle, o primeiro é uma condicionante *if* com a verificação da classificação vocal. O segundo verifica em qual escala pertence a frequência detectada, entre 1 e 5. Por fim, o último ciclo detecta a classificação da nota, entre sustenido e bemol, mais próxima da nota cantada pelo utilizador. Todos os valores utilizados para o ajuste dos ciclos de controle foram retirados da tabela 2.6. O MatLab atualiza então a representação gráfica com a função *drawnow* e começa então, a análise da próxima aquisição. Os valores são então atualizados durante os próximos 3 segundos de treino. É recomendado que esta ferramenta seja executada pelo menos três vezes, para que a definição da classificação vocal seja confiável e complacente com a realidade.

O funcionamento da aplicação é visto na figura 3.3. Durante o processo de aquisição todos os parâmetros variam de acordo com a nota produzida, e o valor mostrado ao fim da execução foi o da última nota captada, no caso um Lá da escala 4, em uma frequência de 441 Hz, se mostrando de acordo com a tabela 2.1 proposta por Aleixo e Mansanares (2003).



Figura 3.3: *Funcionamento da Ferramenta Nota e Classificação Vocal*

Esta ferramenta informa dados de saída referentes à teoria musical, é a ferramenta de apresentação ao usuário; Portanto, o botão de informações busca tutorar o processo de

utilização desta, como visto na figura 3.4.

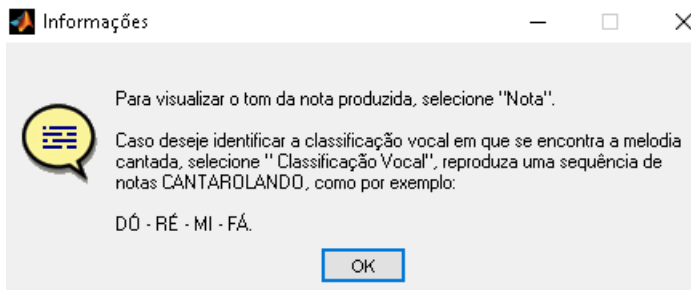


Figura 3.4: *Janela de Informações da Ferramenta Nota e Classificação Vocal*

### 3.3 Treino da Escala Musical

A segunda aplicação parte do princípio que o usuário já utilizou da primeira para obter a estimativa da sua classificação vocal, porém este não é um critério obrigatório para a utilização da ferramenta, já que o usuário pode conhecer qual é a sua classificação afim de treinar a sequência mais propícia ao seu tipo vocal.

As sequências vocais podem ser treinadas de modo crescente ou decrescente, e nesta ferramenta é possível visualizar a amplitude do sinal, para dar a informação ao usuário se há ou não captação do sinal de voz.

O código da escala musical se inicia com a declaração de variáveis globais que alteram a exibição das notas à serem treinadas, a ordem desejada e a variável que será utilizada posteriormente no ciclo de controle *switch* que definirá a sequência de notas no vetor. Ao final do treino, é mostrado o erro quadrático médio proveniente da comparação do vetor de notas previamente selecionado e do vetor de  $F0$  detectada.

A figura 3.5 mostra o funcionamento da ferramenta.

O treino exibido é da classificação vocal "Soprano", em ordem crescente. Ao final o  $MSE$  apresenta um valor de 74, relativamente baixo para esta ferramenta, o que significa que o treino foi bem executado. Nos espaços inferiores da tela é possível notar a exibição das notas em suas duas notações possíveis.

O botão de Informações desta ferramenta se apresenta tal como mostra a figura 3.6.

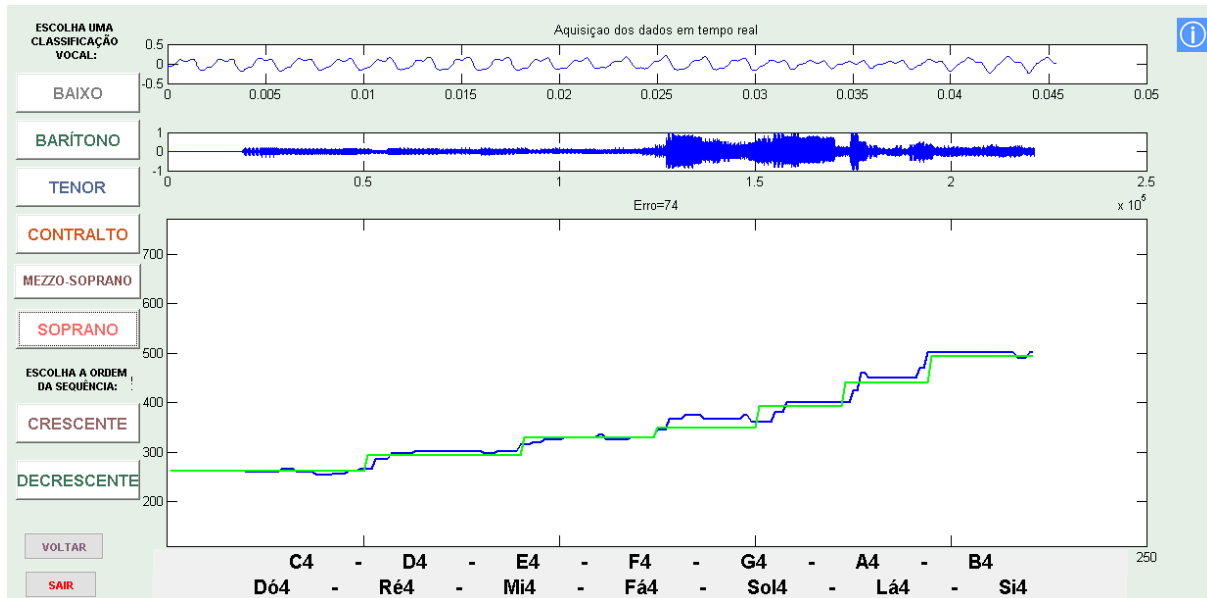


Figura 3.5: *Funcionamento da Ferramenta Treino da Escala*

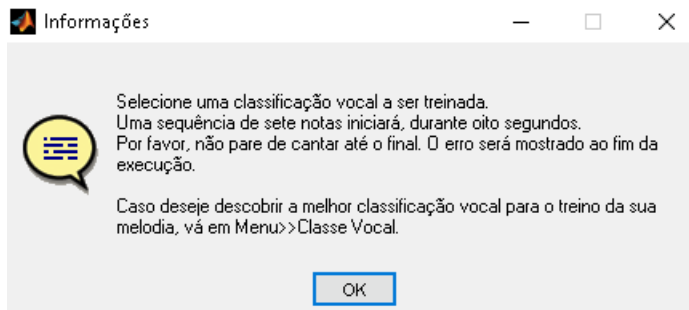


Figura 3.6: *Janela de Informações da Ferramenta Treino da Escala*

## 3.4 Treino de Músicas

Um exemplo do funcionamento da ferramenta Treino de Músicas é exibido na figura 3.7. Esta aplicação procura de forma inteligível disponibilizar o treino de uma pequena base de dados de músicas pré-escritas, divididas em níveis de dificuldade. Não é necessário qualquer pré-ajuste no momento de escolha da música, uma vez que esta é reproduzida em um tom e escala padrão. A ferramenta possui ainda a possibilidade da escuta da música em formato monofônico gerada por um algoritmo que realiza a junção de todas as notas em formato senoidal: A partir da escolha da música a ser treinada, ela realiza o *sound* de cada nota chamada. Esta função faz ainda o tratamento do ruído proveniente desta junção com o uso da janela de Hanning.

Após a declaração de todas as variáveis globais utilizadas na transposição das músicas, como oitava, semitom, velocidade, uma variável com a velocidade temporária para correção de certas anomalias e duração. Ao selecionar a música desejada, se pressionado start em modo padrão, ou seja, sem nenhuma alteração prévia da paleta de transposição, a música se iniciará em Dó da escala mais adequada para cada música. O objeto de entrada analógica é criado e as variáveis globais recolhidas. A variável opção entra em um ciclo de controle *switch* que seleciona o *array* da música desejada. Os limites inferiores e superiores de exibição do gráfico dos vetores de saída são atualizados de acordo com a frequência mínima e máxima possível da escala da música; É então reproduzido um sound com a primeira nota da mesma, durante 1,5 segundo, igual ao tempo de pausa antes do começo da aquisição do sinal. É feita a aquisição da janela de observação, determinação da  $F0$  e o pós-processamento do sinal como nas funções anteriormente descritas. O erro quadrático médio é mostrado ao final do treino, resultante da comparação entre o vetor de  $F0$  e o vetor de música escolhido. São também previstas janelas de informação de erro em relação à manipulação da paleta de transposição, ao tentar diminuir a escala quando se está na escala 1 por exemplo, ou tentar aumentar o tom quando já se está em *Si Maior* ( $B$ ), ou seja, o último semitom do vetor.

A transposição da música pode ser efetuada, com o intuito de ajustar o semitom e a

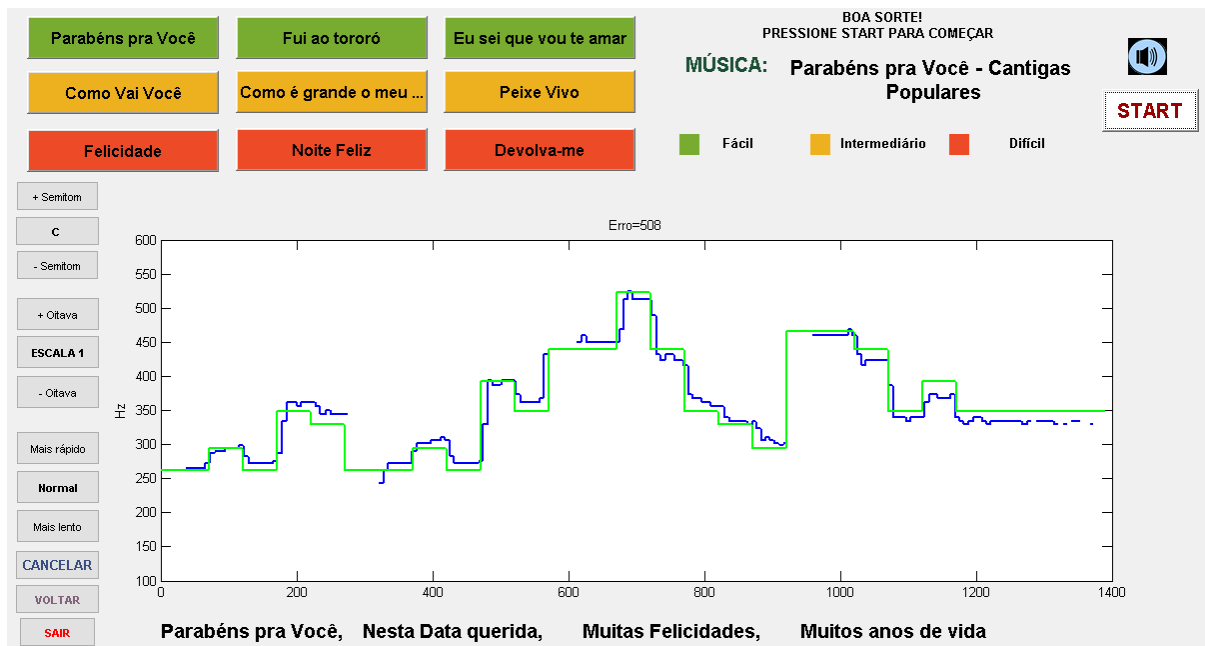


Figura 3.7: *Funcionamento da Ferramenta Treino de Músicas*

escala de forma favorável ao gênero, idade e tipo vocal do usuário. A figura 3.8 mostra a paleta de transposição da música escolhida. É possível notar o tom *C* da escala 1 sendo a configuração padrão para o exemplo mostrado. Nela é possível variar a escala e o semitom, respeitando os limites de escala e semitons possíveis, e exibindo mensagens de erro caso o usuário selecione uma escala ou semitom não-existente.

A velocidade da execução também pode ser manipulada, afim de dificultar ou facilitar o processo de treino não só de acordo com a intenção didática do usuário, mas para corrigir qualquer problema no caso da execução em computadores com hardwares de processamento mais lentos, onde a execução de modo normal ocorre de forma mais lenta que o desejado.

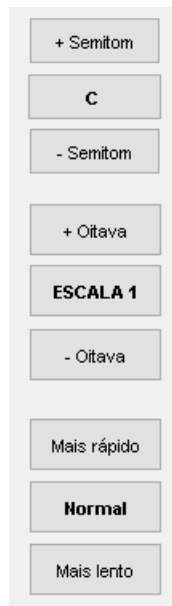


Figura 3.8: *Paleta de Transposição e Ajuste de Velocidade*

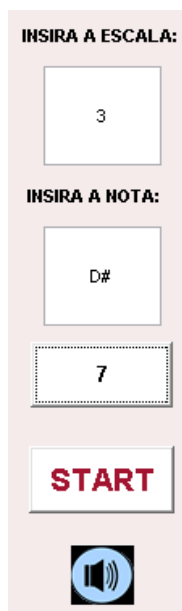
## 3.5 Compositor de Músicas

A quarta e última aplicação proposta é orientada para utilizadores que possuem um conhecimento base em composição melódica, para no caso de uso frequente da ferramenta, dispor de uma base de dados de músicas criadas por si, que podem ser salvas e reutilizadas a qualquer momento do uso da ferramenta.

A estrutura do algoritmo da ferramenta é similar às outras apresentadas, começando com a declaração das variáveis globais utilizadas na ferramenta. Na função acionada ao pressionar o botão inserir (Visualizado na figura 3.9 de demonstração do funcionamento da ferramenta ), os valores das variáveis *handles* criadas automaticamente como variável dinâmica por todos os objetos de interface gráfica da ferramenta GUI do MatLab, são atribuídas por variáveis locais que passarão por verificações afim de inserir a nota desejada pelo utilizador. O primeiro passo é implementar um contador a cada vez que o botão é pressionado, para controlar o número de notas inseridas. A nota é então concatenada em uma string utilizada para visualização e apoio para treino. A variável que corresponde à escala é determinada em um ciclo de controle *switch* que determina a escala de 1 à 5. O segundo ciclo de controle faz o reconhecimento da nota inserida (O reconhecimento

é válido para a terminologia comum, notação alemã, maiúsculas ou minúsculas), que é então multiplicada pelo valor de escala previamente reconhecido. Este valor, expresso em Hz, é atribuído a uma função que constroi a cada inserção o vetor de melodia que será utilizado para o treino. A função executada ao pressionar *start* realiza todas as etapas de aquisição do sinal, determinação da frequência fundamental e pós-processamento do sinal descritos nas funções precedentes.

A inserção das notas pode ser feita em natureza maior, sustenido ou bemol, utilizando a referência das primeiras cinco escalas temperadas do piano, abrangendo as frequências fisicamente realizáveis pelo cantor. A figura mostra o seu funcionamento onde, no exemplo, ocorre a inserção da sétima nota da melodia em valor *D* sustenido da escala 3.



A interface de inserção de nota e escala é apresentada em um painel vertical com fundo rosa claro. No topo, o texto "INSIRA A ESCALA:" precede um campo de entrada contendo o número "3". Abaixo, o texto "INSIRA A NOTA:" precede um campo de entrada contendo "D#". Um campo de entrada contendo o número "7" está destacado com uma borda tracejada. Abaixo disso, há um botão "START" em vermelho. No rodapé, há um ícone de alto-falante azul.

Figura 3.9: *Inserção da Nota Ré Sustenido da Escala 3*

A possibilidade de escuta da melodia composta pode ser efetuada em qualquer momento da composição, com o propósito de averiguar o processo de criação em tempo real e para orientação da sequência.

Na figura 3.10 é possível visualizar a imagem do piano em apoio para a escolha da escala. A seleção da música previamente escrita foi salva no nome de Música Composta e ao final do treino o *MSE* é de 542. Esta ferramenta, devido à liberdade de inserção das

notas torna-se a mais complexa para treino, permitindo ao utilizador realizar sessões de treino com a complexidade desejada.

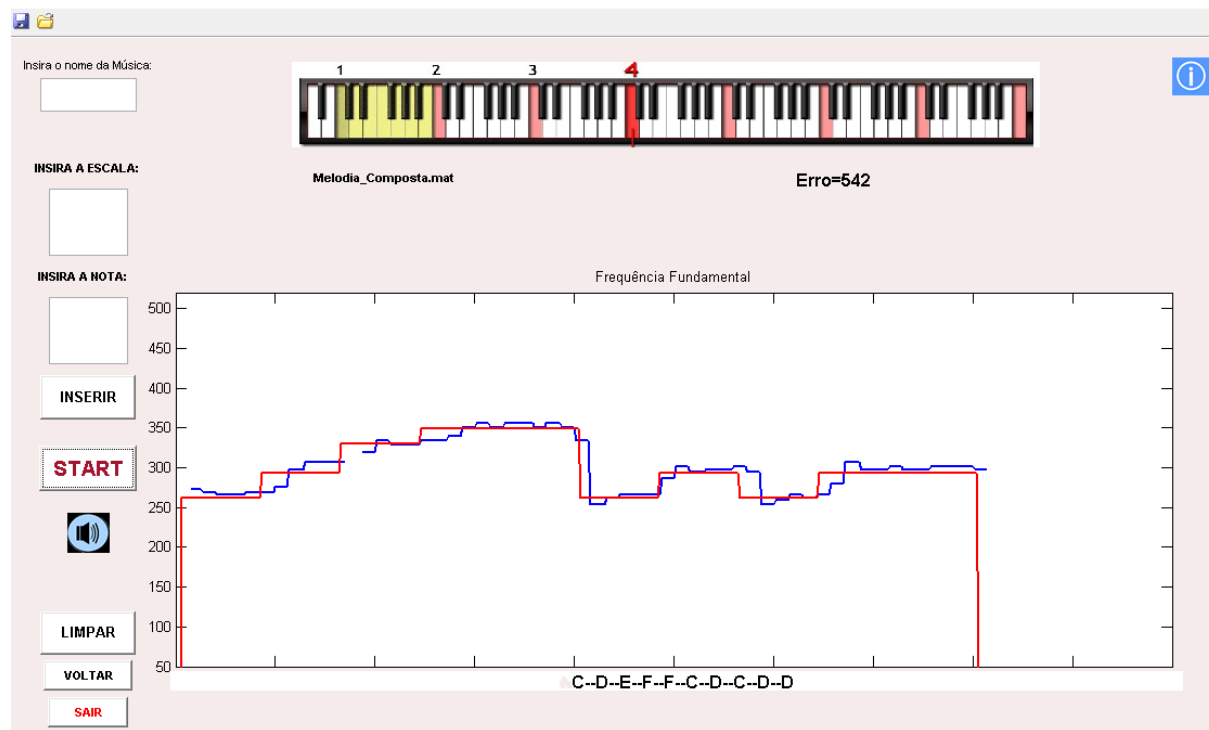


Figura 3.10: *Funcionamento da Ferramenta Treino de Músicas*

A janela de informações 3.11 trás orientações referente à inserção das notas e o processo de "Salvar" e "Abrir" a melodia criada.

Um problema encontrado ao desenvolver esta aplicação, era o controle do tempo de aquisição da  $F0$ , que depende do número de notas inseridas, e portanto, do tamanho da melodia escrita. Para encontrar a relação entre a duração de captação do canal criado e o tempo de duração do treino, foram realizado vinte ensaios em tamanho diferentes de melodias e observado o tempo de duração necessário para cada ensaio. Os valores foram anotados e assim determinado a equação de regressão linear.

A regressão busca mapear a dependência de uma resposta em um preditor (Weisberg, 2005). Afim de prever o valor aproximado da duração, é realizado um diagrama de dispersão, onde observa-se pelo comportamento de Y em relação a X a equação de primeiro

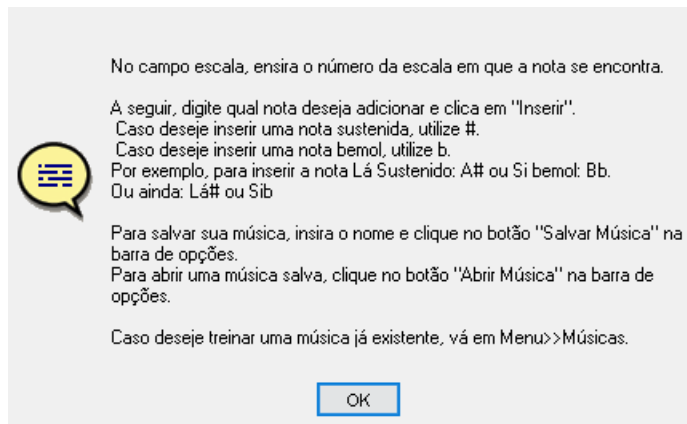


Figura 3.11: *Janela de Informações da Ferramenta Escritor de Músicas*

grau, Sendo a equação da reta  $Y = ax + b$ , a que apresenta o modelo matemático mais próximo para a representação da equação. O método selecionado para se obter a relação funcional foi o *Método dos Mínimos Quadrados (MMQ)*; Este método realiza o somatório do quadrado das distâncias entre os pontos do diagrama de dispersão e os respectivos pontos na curva de estimação é minimizada, representada pela Equação 3.1 (Seber, 1977).

$$S(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (3.1)$$

A relação funcional resultante com o mínimo erro possível é expressa pela equação 3.2:

$$Y(x) = (0.0154 * x) + 0.2507; \quad (3.2)$$

# Capítulo 4

## Performance e Resultados

### 4.1 Caracterização da Amostra

Foram selecionados vinte e quatro utilizadores convidados para teste e avaliação do *software*, sendo estes alunos do IPB, da escola *Escola Superior de Tecnologia e Gestão (ESTiG)* e *Escola Superior de Educação (ESE)*, de diferentes cursos e idades. Eles foram divididos em três categorias, sendo elas:

- I. Nunca cantei em público e/ou entre amigos (Sem experiência de canto).
- II. Canto ocasionalmente em público e/ou entre amigos, mas nunca estudei canto (Autodidata).
- III. Canto frequentemente em público e/ou entre amigos e já estudei/estudo canto (Estudante de Canto).

A estatística dos perfis dos usuários selecionados para o teste segue de acordo com a figura 4.1. Pode-se observar que a idade dos participantes varia de 18 a 25 anos, apresentando uma média de 21,6 anos, dividem-se igualmente entre o gênero masculino e feminino e apresentam uma pequena maioria de nacionalidade brasileira (58%), em relação à portuguesa (42%) e se classificam em categoria de experiência vocal de forma equilibrada, sendo a categoria I (29%), II (38%) e III (33%).

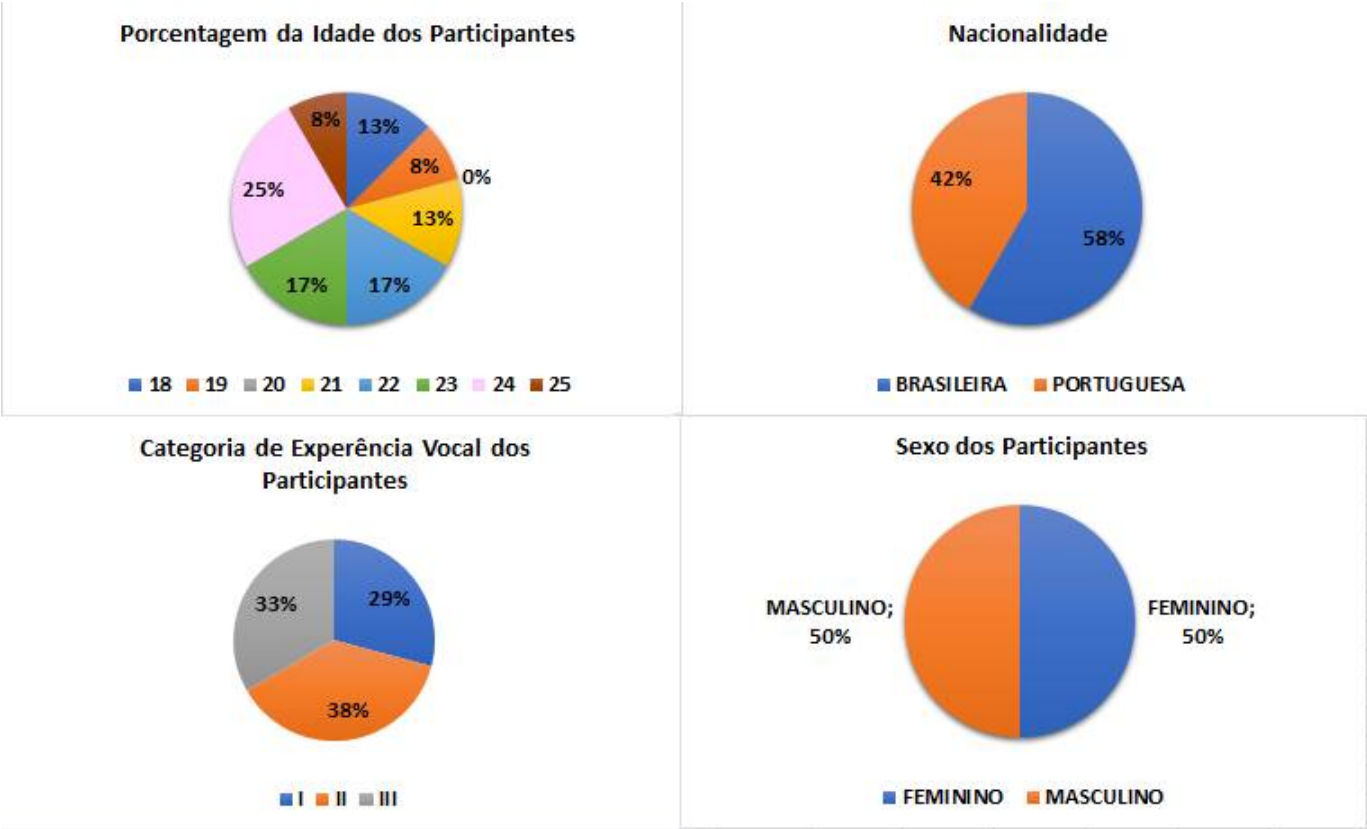


Figura 4.1: Gráficos de Estatísticas dos Usuários

## 4.2 Avaliação do Software

O primeiro teste foi realizado de acordo com a primeira aplicação do *software*, que é o reconhecimento da classificação vocal do cantor, atribuída à melodia cantada. Este valor poderia ser alterado de acordo com a aplicação da escala cantada. Neste caso, foi pedido para o usuário cantar a escala na qual ele se sente mais confortável, para que a resposta de classificação vocal seja aplicável e justa para o treino posterior da segunda ferramenta, o treino da sequência da escala musical correspondente a cada classificação.

As segunda, terceira e quarta ferramentas de treino foram então testadas em etapas sequenciais de cinco vezes, anotando o valor final do erro quadrático médio.

O treino de todas as ferramentas propostas é seguido do questionário com a avaliação dos aspectos gerais do *software*, e utilizou as Heurísticas adaptadas de Nielsen, direcionadas para *softwares* educacionais com interface voltada para o usuário, que busca analisar todos os aspectos 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, 1990), tendo como respostas opções objetivas como sim, não e *Nenhuma Alternativa (N/A)*, conforme mostra a tabela 4.1

### 4.2.1 Avaliação Geral

O resultado da avaliação do aspecto geral do software é apresentado na tabela 4.2. Note-se os resultados em geral altamente positivos, sendo o tempo de resposta aceito como rápido por todos os usuários, e as demais heurísticas com poucas ou nenhuma resposta negativa.

As mensagens de erro se apresentaram claras e úteis para a solução dos problemas para 75% e os outros 25% marcaram *N/A*, visto a não-aparição de mensagem de erro no processo de treino.

O resultado da confiabilidade também apresenta uma taxa de isenção de 25% dos

<b>Heurísticas</b>	<b>Definição</b>	<b>Sim</b>	<b>Não</b>	<b>N/A</b>
Diálogo Simples e Natural	As interfaces do usuário são simples e práticas?			
Linguagem do Usuário	A terminologia da interface usa uma linguagem compatível com o usuário?			
Tempo de resposta	O tempo de resposta em que o sistema informa continuamente o usuário é curto?			
Saídas claramente marcadas	A interface do usuário oferece facilidade para sair das situações?			
Atalhos	A interface permite que operações frequentemente usadas possam ser acessadas e executadas rapidamente.			
Mensagens de erro	As mensagens de erro são claras, sem códigos, precisas e úteis para a solução dos problemas indicados.			
Ajuda e Documentação	O software é fácil de utilizar, de tal forma que dispensa ajuda e consulta à documentação?			
Eficiência	A Estrutura pedagógica criada no software 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 intuitivo, de incentivar a conduzir e a cumprir os objetivos da aprendizagem?			

Tabela 4.1: Avaliação dos Aspectos Gerais do Software

HEURÍSTICAS	SIM	NÃO	N/A	% SIM	% NÃO	% N/A
Diálogo Simples e Natural	23	-	1	96%	-	4%
Linguagem do Usuário	21	1	2	88%	4%	8%
Tempo de resposta	24	-	-	100%	-	-
Saídas claramente marcadas	23	1	-	96%	4%	-%
Atalhos	20	1	3	83%	4%	12%
Mensagens de erro	18	-	6	75%	-	25%
Ajuda e Documentação	22	1	1	92%	4%	4%
Eficiência	22	-	2	92%	-	8%
Confiabilidade	18	1	5	75%	4%	21%
Motivação	22	2	-	92%	8%	-

Tabela 4.2: Resultado da Avaliação dos Aspectos gerais do Software

usuários, que pode ser causado pela incerteza do usuário em relação às informações recebidas, ou pela falta de contato com o recebimento da informação em valores puros de *MSE*.

#### 4.2.2 Avaliações Específicas

A avaliação específica foi concentrada nos fatores específicos para cada uma das quatro ferramentas de treino prático da voz, avaliadas nos seguintes aspectos:

1. Funcionalidade: O software propõe-se a fazer o que é apropriado? 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 raramente apresenta falhas; O software devolve resultados corretos?
4. Usabilidade: É fácil entender o conceito e a aplicação; É fácil aprender a usar?
5. Utilidade: Esta ferramenta é útil?

1	2	3	4	5
Muito mau	Mau;	Nem bom, nem mau;	Bom;	Muito bom

Tabela 4.3: Escala de Avaliação utilizada para os Aspectos Específicos

6. Adequação: O software dispõe de todas as funções necessárias para sua execução? O software é adequado com o conteúdo teórico?
7. Acurácia/Precisão: O software é preciso na execução de suas funções? O software é preciso nos resultados?

Estes parâmetros foram avaliados em uma escala de 1 a 5, conforme representa a escala de classificação da tabela 4.3. O 1 é o pior valor, representando o "Muito mau" funcionamento da heurística analisada, e o 5 é o melhor valor representado por "Muito bom".

Os dados das avaliações específicas encontradas a seguir, são apresentados juntos aos resultados estatísticos como o mínimo valor inserido, o máximo, a mediana, a média aritmética, o desvio padrão e a variância. A variância mede o quanto se diferencia as notas atribuídas para cada parâmetro avaliado em números inteiros da média geral, e é expressa pela equação 4.1 (Montgomery, 2004).

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

Da mesma maneira, o desvio padrão é a raiz quadrada do valor  $V(x)$ .

$$\sigma = [V(x)]^{1/2} \quad (4.2)$$

### Nota e Classificação Vocal

Os valores estatísticos de avaliação da primeira aplicação são descritos na tabela 4.4. Os valores de desvio padrão e variância apresentados a seguir possuem valores baixos, devido à todas as médias apresentarem valores muito próximos entre si, isto poderá ser observado para todas as outras ferramentas. 4.4

A média geral dos critérios avaliados foi de 4.24, com um desvio padrão de 0.12,

Definição	Mín	Máx	Mediana	Média	Desv. Pad.	Variância
Funcionalidade	2	5	4.5	4.38	0.77	0.59
Eficiência	2	5	4	4.38	0.71	0.51
Confiabilidade	2	5	4	4.08	0.78	0.6
Usabilidade	2	5	4	4.29	0.81	0.65
Utilidade	3	5	4	4.08	0.5	0.25
TOTAL	-	-	-	4.24	0.12	0.03

Tabela 4.4: Estatística dos resultados de avaliação da ferramenta Nota e Classificação Vocal

Definição	Mín	Máx	Mediana	Média	Desv. Pad.	Variância
Funcionalidade	3	5	5	4.54	0.59	0.35
Eficiência	3	5	4	4.29	0.75	0.56
Confiabilidade	3	5	4	4.21	0.66	0.43
Usabilidade	2	5	5	4.29	0.95	0.91
Utilidade	2	5	4	3.96	0.86	0.74
Adequação	3	5	4	4.29	0.55	0.30
Acurácia/Precisão	3	5	4	3.96	0.69	0.48
TOTAL	-	-	-	4.22	0.15	0.05

Tabela 4.5: Estatística dos resultados de avaliação da ferramenta Treino da Escala

variando assim de 4.08 a 4.38, o que permanece entre as classificações "Bom" e "Muito Bom", da tabela 4.3.

### Treino da Escala

Os valores estatísticos de avaliação do Treino da Escala são descritos na tabela 4.5. A média das avaliações foi de 4.22 com um desvio padrão de 0.15. variando de 3.96 a 4.54, no limite superior da classificação "Bom" da tabela 4.3. Nota-se que o critério "Funcionalidade" exibe a melhor média do grupo, obtendo uma mediana de 5, o que mostra que a maioria dos valores de classificação ficaram divididos igualmente entre 4 (Bom) e 5 (Muito bom), evidenciando que em geral o programa fez o que foi proposto de forma apropriada.

Definição	Mín	Máx	Mediana	Média	Desv. Pad.	Variância
Funcionalidade	3	5	4	4.08	0.65	0.43
Eficiência	3	5	4	4.17	0.70	0.49
Confiabilidade	3	5	4	4.08	0.72	0.51
Usabilidade	2	5	4	4.17	0.92	0.84
Utilidade	2	5	4	4.21	0.83	0.7
Adequação	3	5	4	4.05	0.75	0.56
Acurácia/Precisão	3	5	4	4.52	0.75	0.56
TOTAL	-	-	-	4.11	0.09	0.02

Tabela 4.6: Estatística dos resultados de avaliação da ferramenta Treino de Músicas

Definição	Mín	Máx	Mediana	Média	Desv. Pad.	Variância
Funcionalidade	3	5	5	4.42	0.72	0.51
Eficiência	2	5	4	4.21	0.89	0.78
Confiabilidade	3	5	4	3.90	0.65	0.43
Usabilidade	2	5	4	4.00	0.88	0.78
Utilidade	2	5	4	3.92	1.06	1.12
Adequação	1	5	4	4.00	0.93	0.87
Acurácia/Precisão	3	5	4	4.25	0.73	0.54
TOTAL	-	-	-	4.10	0.14	0.06

Tabela 4.7: Estatística dos resultados de avaliação da ferramenta Escritor de Músicas

## Treino de Músicas

Os valores estatísticos de avaliação do Treino de Músicas são descritos na tabela 4.6. A média das avaliações foi de 4.11 com um desvio padrão de 0.09. variando de 4.05 a 4.52, no limite superior da classificação "Bom", da tabela 4.3.

## Escritor de Músicas

Os valores estatísticos de avaliação do Escritor de Músicas são descritos na tabela 4.7. A média das avaliações foi de 4.10 com um desvio padrão de 0.14. variando de 3.9 a 4.42, entre os limites inferiores e superiores da classificação "Bom", da tabela 4.3. Esta ferramenta apresentou maior variância total entre os parâmetros, porque ela apresentou maior usabilidade e utilidade para as categorias II e III, e pouca para a categoria I.

<b>CLASSIFICAÇÃO VOCAL DETECTADA</b>	<b>QTD</b>	<b>%</b>
BAIXO	4	17%
BARÍTONO	5	21%
TENOR	4	17%
CONTRALTO	2	8%
MEZZO-SOPRANO	3	13%
SOPRANO	6	25%

Tabela 4.8: Porcentagem das classificações vocais detectadas

## 4.3 Resultados Experimentais

Na tabela 4.8 é possível ver todas as classificações vocais levadas em consideração para os resultados experimentais.

O resultado de média citado nas tabelas 4.9, 4.10 e 4.11 foi dado a partir da média aritmética do valor de cada *MSE* dos usuários, em modo geral e separados por categoria.

A taxa de diminuição média foi calculada pela média aritmética de cada valor negativo da taxa de crescimento, ou seja, a razão do valor presente menos o valor passado pelo valor passado. A taxa de diminuição absoluta foi calculada pelo mesmo princípio, mas sem considerar a média de todos os valores, mas apenas do resultado do primeiro treino com o último.

### 4.3.1 Determinação da Classificação Vocal

A classificação vocal abrangeu as classificações baixo, barítono, tenor, mezzo-soprano e soprano. O exercício foi feito seguindo a documentação de ajuda disponibilizada no botão de informação da janela, incluída no software, sendo o exercício de uma sequência vocal, canonicamente sendo adotada a sequência Dó - Ré - Mi - Fá, por um período de três segundos. A maior classificação se deu pela classificação soprano, mais comum entre as mulheres, e a classificação mais comum entre os homens foi a barítono (Tabela 4.8).

TREINO DA ESCALA	MÉDIA GERAL	MÉDIA CAT I	MÉDIA CAT II	MÉDIA CAT III
1	11022	19124	9205	5975
2	5463	7337	5113	4218
3	2360	1881	2202	2955
4	1315	1285	1471	1166
5	995	510	1183	1208
TD MÉDIA	45%	60%	39%	29%
TD ABSOLUTA	91%	98%	87%	80%

Tabela 4.9: Média do MSE por categoria e TD Médio e aAbsoluto do treino da escala musical

### 4.3.2 Treino da Escala Musical

O treino da escala vocal foi feito de acordo com a classificação apontada, e os valores do *MSE* levam ao cálculo da *Taxa de Diminuição (TD)* médio e absoluto, são dados de acordo com a tabela 4.9. É possível perceber que o treino da escala apresentou ótimos resultados de TD médio e absoluto em aspecto geral, em sua maior eficiência para a categoria I, que apresentou alto erro de treino inicial pela falta de prática. A taxa de diminuição do erro apresenta aumento progressivo de acordo com a falta de experiência do canto, o que mostra a eficiência didática da ferramenta. Desta maneira, é notável que a *TD* do erro média da categoria III foi menor, visto que em tese esta categoria possui certa prática prévia do treino de escala musical, e portanto apresentou uma grandeza menor de *MSE* desde o primeiro treino realizado.

O resultado mais inesperado deste experimento se encontra nos resultados de média geral de valores mais baixos para a categoria I em relação a II e III. Este fenômeno pode ser por conta de uma maior adaptação da categoria I à ferramenta, que pelo fato de nunca ter realizado este tipo de exercício antes, seguiu mais especificamente a instrução da ferramenta; Por outro lado, muitos utilizadores pertencentes à categorias II e III seguiram a velocidade e a escala em que estavam acostumados para este tipo de exercício.

TREINO DA MÚSICA	MÉDIA GERAL	MÉDIA CAT I	MÉDIA CAT II	MÉDIA CAT III
1	14814	30794	6719	9937
2	12575	20762	9109	9311
3	5802	13015	3012	2631
4	5032	11986	1396	3039
5	3001	6142	668	2878
TD MÉDIA	31%	32%	34%	17%
TD ABSOLUTA	80%	80%	90%	71%

Tabela 4.10: Média do MSE por categoria e TD Médio e Absoluto do treino da música Parabéns

### 4.3.3 Treino da Música Parabéns

A tabela 4.10 mostra o resultado da média geral e por categorias do *MSE* dos cinco treinos da música selecionada "Parabéns". Neste treino, os usuários da categoria II apresentaram melhores resultados de *TD* do erro, e também valores de *MSE* em geral menores em relação às demais categorias. Isso se deu provavelmente pelo hábito dos usuários da categoria II aprenderem autodidaticamente, e conseguirem absorver mais rapidamente à metodologia de treino, em relação à categoria I, que precisaria de mais tempo e mais números de treinos para evoluir e melhorar a afinação.

### 4.3.4 Treino da Melodia

No Escritor de Músicas, foi pedido para o usuário compor uma melodia de dez notas, dentre notas maiores, sustenidas e bemóis, das cinco escalas disponíveis, em intervalos de semitons. O resultado foi ouvido com o auxílio do algoritmo de composição de melodia monofônica, e salvo em arquivo .mat.

As *MSE* em geral das três categorias apresentaram os valores de treino inicial bem menores do que as demais ferramentas testadas, possivelmente pelo fato da familiarização da metodologia de uso do *software*. Pelo maior conhecimento em teoria musical da categoria III, e conseqüentemente maior interesse pelo despertar do âmbito criativo, esta

TREINO DA MELODIA	MÉDIA GERAL	MÉDIA CAT I	MÉDIA CAT II	MÉDIA CAT III
1	4005	3544	5427	2807
2	2331	2104	3027	3007
3	1080	1399	1793	1237
4	924	1624	519	1085
5	411	859	533	406
TD MÉDIA	41%	26%	38%	32%
TD ABSOLUTA	90%	76%	90%	85%

Tabela 4.11: Média do MSE por categoria e TD Médio e Absoluto do treino da Melodia Composta

ferramenta apresentou o melhor índice de evolução na aprendizagem, ao analisar os valores de *TD* média e absoluta das três ferramentas testadas. Evidentemente, o contrário acontece com a categoria I, pela falta de conhecimento teórico, e portanto, a dificuldade por parte deste grupo de utilizadores na utilização desta ferramenta.

## 4.4 Evolução

A evolução geral de cada categoria é mostrada nas figuras 4.2, 4.3 e 4.4. O resultado do MSE para cada treino apresenta uma taxa de diminuição positiva e portanto, ocorre a diminuição do erro em todos os casos, o que se traduz para uma melhora na afinação da voz do utilizador.

O resultado do treino da escala musical apresentou displicência em relação à categoria 1, pelo alto índice de erro nos dois primeiros ensaios; Isto se deve pelo fato do primeiro contato com o metologia aplicada na ferramenta para o treino, e pela falta de experiência de controle do aparelho fonador. Entretanto, em termos gerais esta ferramenta apresentou o menor erro ao final dos cinco ensaios, comprovando a maior eficiência para a afinação vocal.

O resultado do treino da música "Parabéns" mostrou-se difícil para a categoria I pelo alto índice de erro ao longo de todos os ensaios, embora haja uma evolução significativa. A

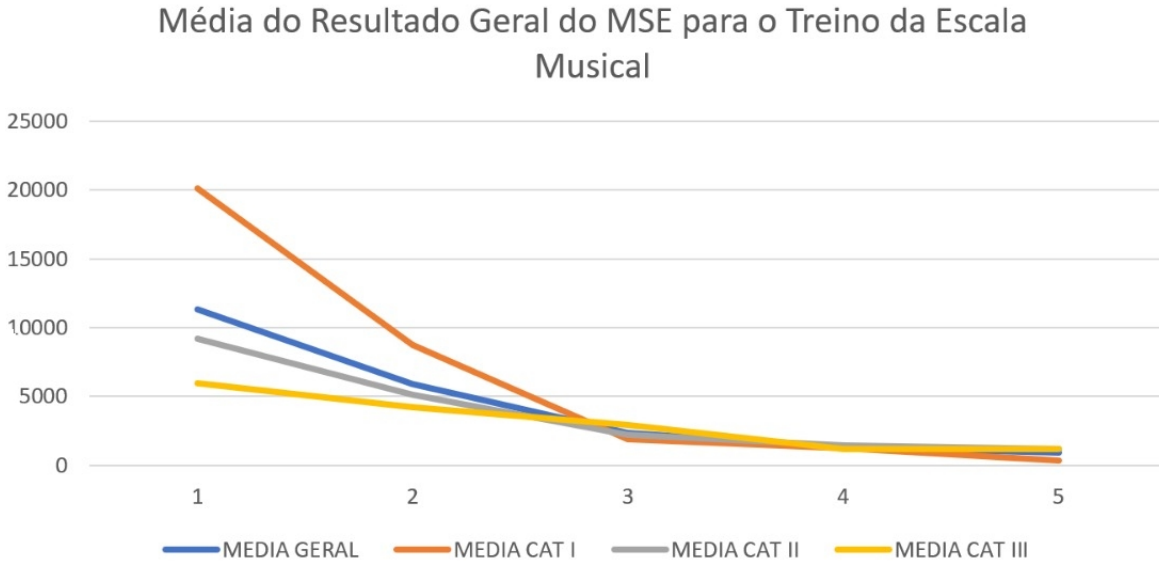


Figura 4.2: Média do resultado Geral do MSE para o Treino da Escala musical

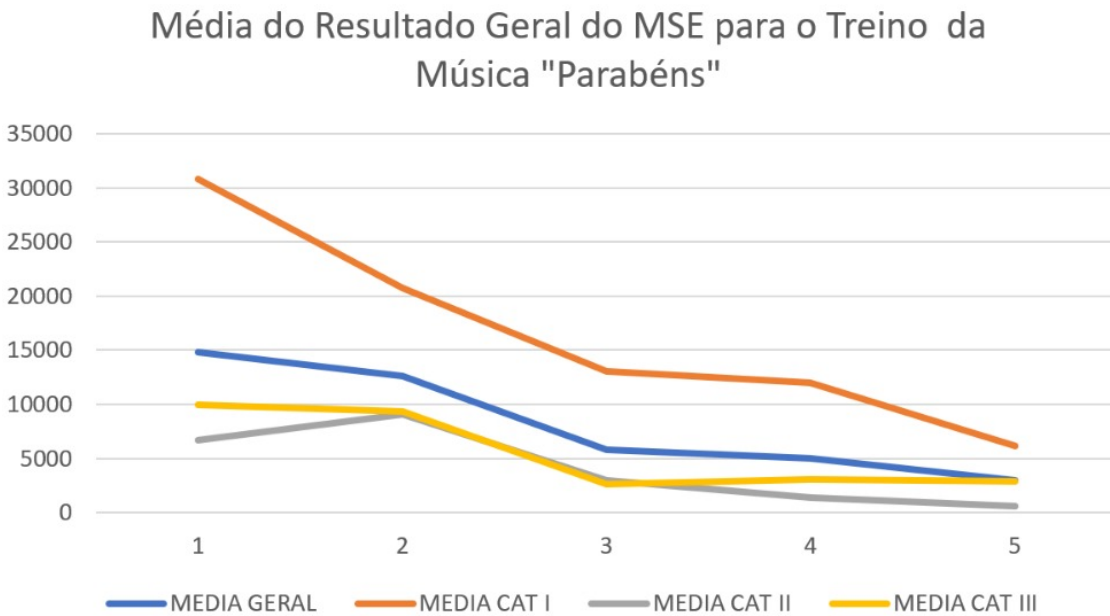


Figura 4.3: Média do resultado Geral do MSE para o Treino da Música Parabéns

categoria III apresenta resultados constantes de erro após o terceiro ensaio, o que mostra que esta ferramenta não é a mais adaptada para este grupo em termos de evolução.

Devido ao fato de ser a última ferramenta treinada, o que produziu com treino agregado conhecimento do mecanismo do *software*, apresentou, desde o primeiro ensaio, resultados de *MSE* mais baixos em relação às outras atividades.

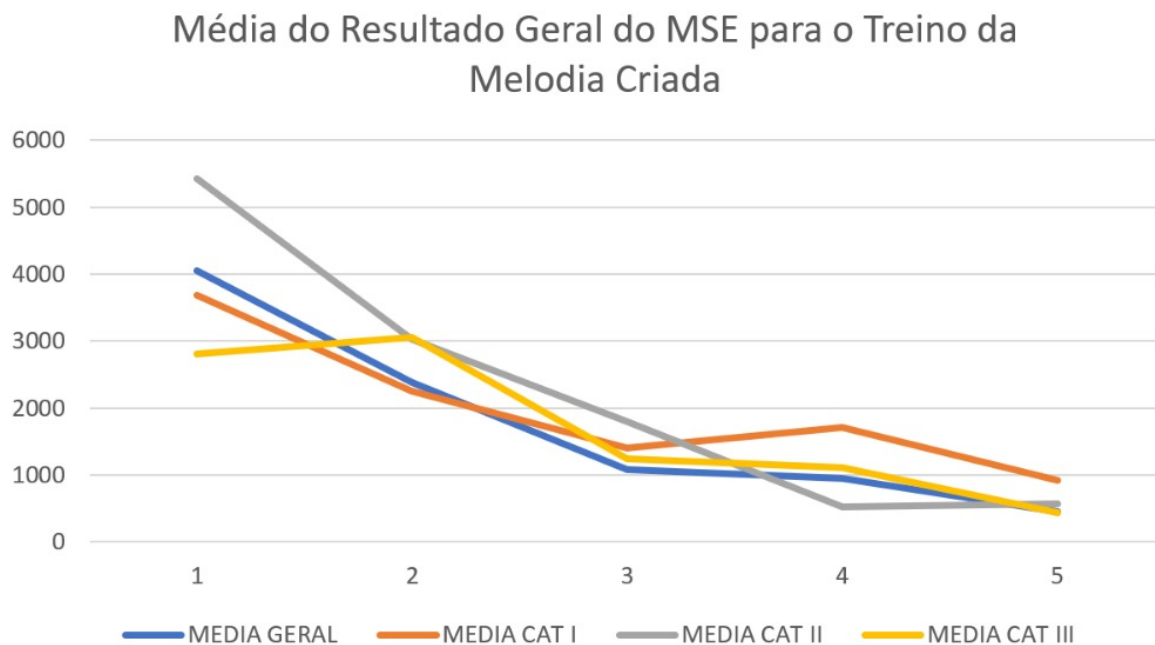


Figura 4.4: Média do resultado Geral do MSE para o Treino da Melodia Criada

As categorias II e III obtiveram um resultado linear de melhoria no terceiro exercício, o que prova uma maior eficiência pedagógica ao inserir a própria melodia para este grupo de utilizadores, ao contrário da categoria I, que embora apresente resultados de evolução, possui um pequeno aumento do erro entre o ensaio 3 e 4, o que demonstra a dificuldade do treinamento desta ferramenta para esta categoria.

Por estas razões, se mostrou válida a disposição de diferentes métodos de treino em um só sistema, para que de alguma maneira o *software* apresente um método adequado de melhoria da afinação da voz para todos os diferentes utilizadores.

Em alguns casos, o alto valor do erro final representa alguma falha no processo de aquisição do sinal, seja pela aquisição do ruído sendo confundida com um elemento de

alta frequência, seja pela não-distinção de alguns elementos vocais que contém informações vozeadas e não vozeadas em um curto período de tempo.



# Capítulo 5

## Conclusão e Trabalhos Futuros

O sistema desenvolvido em apoio à afinação da voz propõe inovações no âmbito do treinamento vocal e na inovação didática em relação ao aprendizado. O método da autocorrelação com *center-clipping* se mostra altamente efetivo. A escolha de desenvolver um sistema *VFT* foi uma medida eficaz para aplicar o método de aquisição da frequência fundamental, veja-se a evolução linear da afinação vocal dos três grupos de utilizadores selecionados para os testes.

O uso da detecção da classificação vocal mostrou-se interessante aos utilizadores, visto que, a partir do resultado desta ferramenta, o uso do treino de escala tornou-se mais fácil. O menu com músicas pré-criadas que realiza a transposição instantânea foi eficaz para se adaptar aos diferentes tipos vocais dos utilizadores.

Os resultados obtidos demonstram que o sistema elaborado mostra-se robusto para realizar a detecção do sinal não vozeado (mesmo em lugares pouco ruidosos) e a distinção do silêncio onde não há sinal de fala. O tempo de resposta do *software* é curto o suficiente para que a aquisição da  $F_0$  se faça simultânea ao treino, obtendo resultados próximos ao tempo real.

As ferramentas utilizadas na presente pesquisa possuem diversas possibilidades de continuidade, que incluem: a tradução para uma linguagem que possua maior liberdade de exportação e criação de um *software* executável sem dependências como C++ ou python, a transcrição das ferramentas para a versão do MATLAB 64 bits, tais fatores

podem expandir o projeto para além do âmbito acadêmico.

A melhora do sistema de paragem de aquisição do sinal de voz produziria potenciais diminuições do erro final ao término do treino. O erro quadrático médio pode ser ajustado a um sistema de exibição mais adaptado a alguns utilizadores mais familiares com a percentagem de acertos no lugar da diminuição do erro.

A inserção de mais ferramentas de treino, ou a adição de uma base de dados mais ampla na ferramenta de treino de músicas, aliadas à criação de um acompanhamento sonoro de fundo que se adapte ao timbre de um instrumento musical, e que efetue a transposição de suas notas simultaneamente à transposição do vetor de treino, apresentam-se como medidas interessantes para auxiliar o utilizador durante o uso da ferramenta, de forma que o sistema seja otimizado e possibilite resultados cada vez mais satisfatórios.

# Bibliografia

1. Aleixo, G. T. e A. M. Mansanares (2003). *Propriedades das Ondas Numa Corda*. Relatório final de F809, IFGW / UNICAMP.
2. Almeida, N. C. (2010). *Sistema Inteligente para Diagnóstico de Patologias na Laringe Utilizando Máquinas de Vetor de Suporte*. Dissertação de Pós-Graduação em Engenharia Elétrica e de Computação. UFRN - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia.
3. Alves, Nuno Filipe Ribeiro (2016). *Diagnóstico inteligente de patologias da laringe*. *Dissertação de Mestrado em Tecnologia Biomédica*. Escola Superior de Tecnologia e Gestão - IPB.
4. Bachu, R.G., B. Kopparthi S.and Adapa e B.D. Barkana (2010). “Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy”. Em: ed. por Khaled Elleithy.
5. Baptista, Manuel (2017). *Sistemas De Aquisição De Dados. Notas de Aula. Engenharia de Sistemas e Informática - 2.º Ano*. Acessado em: Dezembro de 2017.
6. Bartošek, Jan e Václav Hanžl (2010). “Comparing Pitch Detection Algorithms for Voice Applications”. Em: *Semantic Scholar*.
7. Benolken, M.S. e C.E. Swanson (1990). “The effect of pitch-related changes on the perception of sung vowels”. Em: 87, pp. 1781–1785. DOI: 10.1121/1.399426.

8. Brockmann-Bauser, M. (2011). “Improving jitter and shimmer measurements in normal voices”. Em: Institute of Cellular Medicine, Medical School, Newcastle University.
9. Callaghan, Jean, William Thorpe e Jan van Doorn (2004). “The science of singing and seeing”. Em: *Proceedings of the Conference on Interdisciplinary Musicology*.
10. Camargo, Thatiana Francisco de, Barbosa Daniela Aparecida e Lídia Cristina da Silva. Teles (2007). *Características da fonetografia em coristas de diferentes classificações vocais*. *Revista da Sociedade Brasileira de Fonoaudiologia [online]*.
11. Canta (2017). <http://www.singintune.org/>. Software. Acessado em: Out, 2017. Chaumet Software. France.
12. Cardoso, B. e M. Mascarenhas (1973). *Curso Completo de Teoria Musical e Solfejo - 1o Vol.* vol. 1. Irmãos Vitale. ISBN: 9788585188177. URL: [https://books.google.pt/books?id=%5C\\_zQCnNwc3vMC](https://books.google.pt/books?id=%5C_zQCnNwc3vMC).
13. Colton, R.H. e P. Woo (1995). *Measuring vocal funcion*. In: Silva, M. C. Biase, N. D. *Estudo da Frequência Fundamental da voz em mulheres jovens com síndrome Pré-Menstruais*. New York, Ygaku-shoin.
14. Ferreira, Ricardo (2003). “Codificação de voz em GSM. Geração de voz humana”. Em: Notas de Aula. Instituto Superior Técnico, Lisboa.
15. Hamid, Oppenheim Alan V; Willsky Alan S; Nawab Syed (1997). *Signals and systems. 2 ed.* Upper Saddle River N.J: Prentice Hall.
16. Hoppe, D., M. Sadakata e P. Desain (2006). “Development of real-time visual feedback assistance in singing training: a review”. Em: 22.4, pp. 308–316. ISSN: 1365-2729. DOI: 10.1111/j.1365-2729.2006.00178.x. URL: <http://dx.doi.org/10.1111/j.1365-2729.2006.00178.x>.

17. Howard, David M. e Graham F. Welch (1989). “Microcomputer-based singing ability assessment and development”. Em: vol. 27.2, pp. 89–102. ISSN: 0003-682X. DOI: [https://doi.org/10.1016/0003-682X\(89\)90002-9](https://doi.org/10.1016/0003-682X(89)90002-9). URL: <http://www.sciencedirect.com/science/article/pii/0003682X89900029>.
18. Inc., Mathworks (2014a.). *About MATLAB - Overview*. Massachusetts, United States. URL: [https://www.mathworks.com/products/matlab.html?s\\_tid=hp\\_ff\\_p\\_matlab](https://www.mathworks.com/products/matlab.html?s_tid=hp_ff_p_matlab).
19. Junker, David Bretanha (1999). *A importância do canto coral*. Convenção Internacional de Regentes de Coros. Anais. Brasília.
20. Kent, R.D. e C. Read (2002). *The Acoustic Analysis of Speech*. Singular/Thomson Learning. URL: <https://books.google.fr/books?id=0mhqAAAAMAAJ>.
21. Lathi, B.P. (2006). *Signals and Systems - 2.ed.*: Bookman. ISBN: 9788577803910. URL: <https://books.google.pt/books?id=S806DQAAQBAJ>.
22. Leong, S. e L. Cheng (2014). “Effects of Real-time Visual Feedback on Pre-service Teachers’ Singing”. Em: *J. Comp. Assist. Learn.* 30.3, pp. 285–296. ISSN: 0266-4909. DOI: 10.1111/jcal.12046. URL: <http://dx.doi.org/10.1111/jcal.12046>.
23. Lopes, J. M. (2008). *Ambiente de Análise Robusta dos Principais Parâmetros Qualitativos da Voz. Dissertação de Mestrado em Engenharia Eletrotécnica e de Computadores*. Faculdade de Engenharia da Universidade do Porto.
24. Martin, George (2008). *The Opera Companion, A Guide for the Casual Opera-Goer. An imprint of Hal Leonard Corporation, Paper*, N.Y.:Amadeus Press,
25. Marvan, Jaber (2007). “Voice Activity detection Method and Apparatus for voiced/unvoiced decision and Pitch Estimation in a Noisy speech feature extraction”. Em: *Jaber Associates LLC*.

26. MathWorks, Inc. (2014). *Create Apps with Graphical User Interfaces in MATLAB*. Massachusetts, United States. URL: <https://www.mathworks.com/discovery/matlab-gui.html> (Access%20Date:%2020th%20June%202018).
27. Media Singing Method (2017). *eMedia Music Corporation. Software. Version 1.2.2*. CD-ROM.
28. Montgomery Douglas C.; Runger, George C.; Hubele; Norma F. (2004). *Estatística Aplicada à Engenharia*. Segunda Edição - LTC Editora S.A. Rio de Janeiro.
29. Nadoleszny, M. (1923). *Untersuchungen uber den kunstgesang*. Berlin: Springer.
30. Neves, Caio Gomes Fontes Calaça (2016). *Sistema para Afinação de Sinais Monofônicos de Voz com Interface Gráfica*. Universidade Federal do Rio de Janeiro.
31. Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.
32. Nielsen J.; Molich, R. (1990). "Heuristic evaluation of user interfaces". Em:
33. Oliveira, Keroline R. e Ailen Rose Balog de Lima (2016). *Um estudo sobre a relação entre capacidade de concentração e afinação vocal: um relato de experiência*. *Revista Integratio*, v. 2, n. 2, ago. p. 54-60.
34. Oppenheim, Alan V. e Ronald W. Schafer (2009). *Discrete-Time Signal Processing*. 3rd. Upper Saddle River, NJ, USA: Prentice Hall Press.
35. Panazio, Tales Gouveia Fernandes; Aline Neves (2009). *Do Analógico Ao Digital: Amostragem, Quantização e Codificação*. Modelagem e Ciências Sociais Aplicadas (CECS). II SIC - UFABCII.
36. Peckham, Anne (2005). *Vocal Workouts for the Contemporary Singer*. Erkle Press Publications.

37. Pinto, António Sá e Ana Maria Tomé (2001). “Pitch Detection in Vocal Music Monophonic Signals”. Em:
38. Pitch Perfector (2009). *PIERCE, Stephen. Software. Version 3.8 Training.*
39. Rabiner, L. R. (1977). “On the use of Autocorrelation Analysis for Pitch Detection”. Em:
40. Rabiner, Lawrence R. e Ronald W. Schafer (1978). *Digital processing of speech signals / Lawrence R. Rabiner, Ronald W. Schafer.* English. Prentice-Hall Englewood Cliffs, N.J, xvi, 512 p. : ISBN: 0132136031.
41. Sá Ferreira, João Filipe Terleira de (2012). “Tecnologia de Apoio em Tempo-Real ao Canto - Relação entre parâmetros perceptivos da voz cantada com fenómenos acústicos objectivos. Dissertação de Mestrado.” Em: *Instituto Politécnico do Porto Escola Superior de Música, Artes e Espectáculo.*
42. Sadie, Stanley (1985). *The New Grove Dictionary of Musical Instruments.* Macmillan Press.
43. Science Buddies Staff (2009). *Correlation Between Relative Pitch and Age, Gender, or Musical Background. 28 July 2017. Web. 28 Nov. 2017.* <[https://www.sciencebuddies.org/science-fair-projects/project-ideas/Music\\_p015/music/relative-pitch-age-gender-musical-background](https://www.sciencebuddies.org/science-fair-projects/project-ideas/Music_p015/music/relative-pitch-age-gender-musical-background)>.
44. Seber, G.A.F. (1977). “Linear Regression Analysis”. Em: Wiley series in probability and mathematical statistics.
45. Sellam e J.I Jagadeesan (2014). *Classification of Normal and Pathological Voice Using MVS and RBFNN. Journal of Signal and Information Processing. 5, 1-7.*
46. Shannon, Claude E. e Warren Weaver (1949). *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press. ISBN: 978-0-252-72548-7.

47. Sing True (2014). *Easy Ear Training Ltd. / Musical U. Version 1.2.2*. Surbiton, United Kingdom.
48. Tan, L. e M. Karjandecha (2003). *fundamental Frequency Detection Algorithm: Autocorrelation Method and AMDF*. M. , vol. 2, pp. 551–556, Proceedings of the 3rd International Symposium on Communications e Information Technology.
49. Teixeira, João Paulo (2018). “Notas de Aula - Produção e Análise do Sinal de Fala”. Em:
50. Teixeira, João Paulo, Débora Ferreira e Susana Moreira (2011). “Análise acústica vocal - determinação do Jitter e Shimmer para diagnóstico de patologias da fala”. Em: 6º Congresso Luso-Moçambicano de Engenharia. Maputo, Moçambique.
51. Uliani Neto, M., R. Violato, F. Simões e B. Costa (2013). *Speech Codification Techniques Based On Analysis-by-synthesis. Técnicas de Codificação de Fala Baseadas Em Análise por Síntese*. Faculdade de Jaguariúna; Faculdade Politécnica de Campinas; Fundação Cpqd.
52. Upadhyya, Savitha S e Nilashree Wankhede (2012). “Pitch Estimation using Autocorrelation Method and AMDF”. Em: *International Conference on Advances in Computing and Management*.
53. Vieira, Maurílio Nunes (2004). *Uma introdução à acústica da voz cantada. I Seminário Música Ciência Tecnologia n.1*.
54. Wackerly, D., W. Mendenhall e Richard L. Scheaffer (2008). *Mathematical Statistics with Applications*. 7 ed. Thomson Higher Education Belmont, CA, USA.
55. Weisberg, S. (2005). *Applied Linear Regression*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9780471704089. URL: <https://books.google.pt/books?id=xd0tNdF00jcC>.

56. Wen Heand Teng, ZhaoSheng, SiYu Guo, JingXun Wang, BuMing Yang, Yi Wang e Tao Chen (2009). “Hanning self-convolution window and its application to harmonic analysis”. Em: 52.2, pp. 467–476. ISSN: 1862-281X. DOI: 10.1007/s11431-008-0356-6. URL: <https://doi.org/10.1007/s11431-008-0356-6>.
57. Woo, P., J. Casper, R. Colton e D. Brewer (1994). *Aerodynamic and stroboscopic findings before and after microlaryngeal phonosurgery*. J Voice.