UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO

MARCOS CORRÊA BARCELOS JUNIOR

IMPLEMENTAÇÃO DE ALGORITMOS DE CORREÇÃO AUTOMÁTICA DE TOM

VITÓRIA
2020

MARCOS CORRÊA BARCELOS JUNIOR

IMPLEMENTAÇÃO DE ALGORITMOS DE CORREÇÃO AUTOMÁTICA DE TOM

Parte manuscrita do Projeto de Graduação do aluno **Marcos Corrêa Barcelos Junior**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Orientador: Prof. Dr. Jorge Leonid Aching Samatelo

VITÓRIA
2020

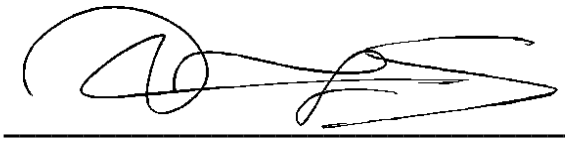
MARCOS CORRÊA BARCELOS JUNIOR

IMPLEMENTAÇÃO DE ALGORITMOS DE CORREÇÃO AUTOMÁTICA DE TOM

Parte manuscrita do Projeto de Graduação do aluno **Marcos Corrêa Barcelos Junior**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

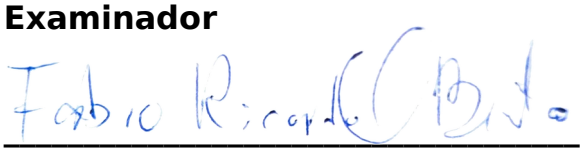
Aprovada em 11 de dezembro de 2020.

COMISSÃO EXAMINADORA:



Prof. Dr. Jorge Leonid Aching Samatelo
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. André Ferreira
Universidade Federal do Espírito Santo
Examinador



Prof. MSc. Fabio Ricardo Oliveira Bento
Instituto Federal do Espírito Santo
Examinador

À minha irmã Paula e a todos que me apoiaram no decorrer do Curso de Engenharia Elétrica da UFES.

AGRADECIMENTOS

À minha irmã, Paula Cristina Lira Corrêa, por fingir que entende minhas divagações sobre engenharia elétrica, por sempre estar disposta a ouvir meus desejos de desistência do curso e me ajudar nas minhas crises existenciais.

Aos meus amigos por terem fé e aguardarem ansiosamente um resultado que parecia não chegar nunca, e por tornarem a minha passagem pela UFES mais enriquecedora e prazerosa.

Ao meus pais, Alessandra Lira Corrêa e Marcos Corrêa Barcelos, por me mostrarem o caminho a se seguir e por me motivarem a ser a melhor versão de mim.

À minha família por estar sempre ao meu lado.

Ao meu professor orientador, Jorge Leonid Aching Samatelo, pela paciência, por todos os conhecimentos e conselhos durante essa jornada.

À banca examinadora pela aceitação do convite e pelo tempo investido para leitura e avaliação desse trabalho.

.

RESUMO

A edição de áudio é bem presente nas gravações de faixas da atualidade. Uma de suas ferramentas é a correção de tom. A fim de que haja um menor tempo de gravação, os editores utilizam essa ferramenta para ajustar desvios nas frequências dos tons, ou seja, desafinações. Assim, como em grandes apresentações se deseja uma correção automática e, em tempo real, alguma ferramenta de autocorreção de tom se torna necessária. É uma tecnologia relativamente recente, considerando que em 1996 foi lançada a primeira faixa comercial utilizando um corretor de tom, que ficou popularmente conhecida como *autotune*. O presente projeto faz uso de teoria musical e de processamento digital de sinais para tentar solucionar o problema da correção automática de tom. Para auxiliar e facilitar a tarefa, foram usadas implementações das técnicas de correção de tom e conjuntos de dados de vozes para testar as mesmas. Além disso, este trabalho inclui material teórico sobre técnicas de correção e detecção de tom para a aprendizagem sobre o tema. Conseguiu-se obter um corretor automático de tom que exerce a função de corrigir a frequência emitida ao próximo tom da tabela de frequências corretas, porém apresentou problemas de ruído. O corretor implementado foi testado usando as faixas da base de dados de cantores (BLACK; LI; MIBLACK, 2014), os resultados obtidos são de um erro médio de 25 hertz e uma qualidade sonora mediana, os quais são valores aceitáveis considerando esta uma implementação inicial.

Palavras-chave: Corretor automático de tom. *Autotune*.

ABSTRACT

Nowadays the process of audio editing is something common on studio records and on live performances. One of its tools is the pitch correction that usually be used to adjust the frequency to the most appropriate pitch when the singer made a mistake. This technology is relatively recent considering that the first commercial song using pitch correction was released in 1996, which became popularly known as autotune. This project makes use of theories of signal processing, musical theory, other pitch correction papers to try to understand the process of an automatic pitch corrector. To facilitate the task, codes from pitch correction and a voice database, (BLACK; LI; MIBLACK, 2014), were used. It was possible to obtain an automatic pitch corrector that corrects the frequency to the next closest pitch, but a characteristic noise of leakage phenomenon appeared on the output signal. This corrector was tested using the voice database and some evaluation metrics, sound quality and error, getting a result an average error of 25 hertz and average sound quality that were considered acceptable values for an initial implementation.

Keywords: Pitch corrector. Autotune.

LISTA DE FIGURAS

Figura 1 – Frequência de 3 e 10 hertz.....15
Figura 2 – Microfone de carvão.....16
Figura 3 – Cordas vocais 17
Figura 4 – Vibração das cordas 17
Figura 5 – Teclas do piano e seus tons 18
Figura 6 – Escala diatônica de dó maior.....19
Figura 7 – Efeito de *leakage*.....20
Figura 8 – Janelamento.....20
Figura 10 – Influência de altas frequências 23
Figura 10 – Sinal de áudio: (a) original e (b) corrigido via a implementação no MATLAB.28
Figura 11 – Sinal de áudio: (a) original e (b) corrigido via a implementação em PYTHON.29
Figura 12 – Autocorretor pseudocódigo 29
Figura 13 – Fluxograma do código implementado.....33

LISTA DE QUADROS

Quadro 1 – Melhores aplicações para cada janela.....21
Quadro 2 – Qualidade sonora 30
Quadro 3 – Resultados para a onda de frequência crescente de 200 a 500 Hz.....31
Quadro 4 – Resultados para uma faixa de áudio do banco de dados.....31

LISTA DE ABREVIATURAS E SIGLAS

AMDF	*Average magnitude difference function*
Colab	*Google Colaboratory*
CD	*Compact disc*
DFT	*Discrete Fourier transform*
FFT	*Fast Fourier transform*
STFT	*Short-time Fourier transform*
UFES	Universidade Federal do Espírito Santo

LISTA DE SÍMBOLOS

A	Lá
B	Si
C	Dó
D	Ré
E	Mi
F	Fá
G	Sol
$a(\tau)$	Função da diferença de magnitude média
τ	Período fundamental (s)
$x(n)$	Sinal de tempo discreto
n	Tempo discreto (s)
$x(n-\tau)$	Sinal deslocado no tempo

SUMÁRIO

1 INTRODUÇÃO 13
1.1 Apresentação e Objeto de Pesquisa 13
1.2 Justificativa 13
1.3 Objetivos 14
1.3.1 Objetivo Geral 14
1.3.2 Objetivos Específicos 14

2 EMBASAMENTO TEÓRICO 15
2.1 Conceitos de Uso na Análise de Áudio Digital 15
2.1.1 O Sinal Digital 15
2.1.2 A Voz e a Música 16
2.1.3 Ferramentas de Uso em Processamento de Sinais 19

3 SISTEMA PARA CORREÇÃO AUTOMÁTICA DE TOM 22
3.1 Algoritmos de Detecção de Tom 22
3.1.1 Algoritmos Baseados na Representação Frequencial 22
3.1.2 Algoritmos Baseados na Representação Temporal 22
3.2 Algoritmos de Correção do Tom 24
3.2.1 Algoritmos de Implementação de Forma Automatizada 24
3.2.2 Algoritmos de Implementação Manual 24

4 SOLUÇÃO PROPOSTA 25
4.1 Introdução 25
4.2 Implementação 25

5 RESULTADOS 27
5.1 Recursos Computacionais 27
5.1.1 Recursos de *Software* 27
5.1.2 Recursos de *Hardware* 27
5.1.3 Conjunto de Dados de Sinais de Áudio 27
5.2 Detalhes da Implementação 27
5.2.1 Implementação no MATLAB 28
5.2.2 Implementação em PYTHON 28

5.3 Métricas 30
5.3.1 Avaliação 31
5.4 Código que usa Faixas do Microfone 32
5.5 Comparação de Resultados 33
5.5.1 Primeira Comparação 33
5.5.2 Segunda Comparação 34

6 CONCLUSÕES E PROJETOS FUTUROS 35

REFERÊNCIAS BIBLIOGRÁFICAS 36

1 INTRODUÇÃO

1.1 Apresentação e Objeto de Pesquisa

A música se fez presente em grande parte da história da humanidade, porém, a obtenção do áudio, a fim de que possa ser reproduzido, foi primeiramente conseguido pelo fonógrafo criado por Thomas Edison em 1877 (MILLARD, 1996).

Cem anos depois dessa descoberta, a maioria das casas já tinham um aparelho que reproduziria e gravaria áudio, por meio das chamadas fitas cassetes, que revolucionaram a forma de se reproduzir o som, sendo utilizadas em brinquedos de crianças por ter um tamanho relativamente pequeno. As fitas cassetes foram substituídas por discos compactos (CD, do inglês compact disc), que foi um dos formatos mais difundidos na sociedade (MILLARD, 1996). Posterior a essa criação, desenvolveu-se várias formas de entretenimento a partir da gravação do áudio, como as transmissões de rádio e os telefones (MORTON, 2000). Junto de toda essa evolução da captação de áudio, a música foi totalmente afetada e efeitos começaram a poder ser incluídos nas vozes de cantores e instrumentos, o que elevou as gravações musicais a outro nível. Cantores e músicos são cada vez mais pressionados a não errar, pois deseja-se ouvir a música em perfeita harmonia e afinação. Ser capaz de corrigir no estúdio a emissão vocal de forma a não deixar o som artificial é uma tecnologia recente (BREGTIZER, 2009). Essa ferramenta é cada vez mais usada para corrigir erro de cantores e músicos em programas televisionados, em estúdios e em apresentações importantes. O que faz ser interessante o desenvolvimento de algoritmos relacionados ao meio musical.

Com o avanço dos recursos computacionais, o tratamento de áudio se torna cada vez menos problemático. A criação de um algoritmo que seja capaz de detectar e corrigir a frequência emitida pela voz, ou tom, tem uma grande necessidade e ampla aplicação na área musical.

1.2 Justificativa

Ao abordar a voz humana, sabe-se que errar a afinação é algo comum até mesmo para cantores profissionais e que a autocorreção de tom é utilizada em eventos de importância. Assim, a criação de uma ferramenta que faça este trabalho de correção é interessante para o meio musical,

o que justifica este trabalho, pois embora existam algumas ferramentas que façam essa correção, esses *softwares* tem um custo elevado e muitas vezes tem limitações de correção em tempo real.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo do trabalho atual é desenvolver um algoritmo sobre correção automática de tom. Tal algoritmo foi testado utilizando faixas de uma base de dados de vozes de cantores (BLACK; LI; MIBLACK, 2014) para obter um áudio nas frequências tonais.

1.3.2 Objetivos Específicos

Para que o objetivo geral do trabalho pudesse ser atingido, alguns objetivos específicos precisaram ser alcançados. Dentre estes, destacam-se:

a) Desenvolver um código de correção de tom;
b) Definir métricas de avaliação do código feito;
c) Comparar os resultados com os de outras literaturas para validar o presente trabalho.

2 EMBASAMENTO TEÓRICO

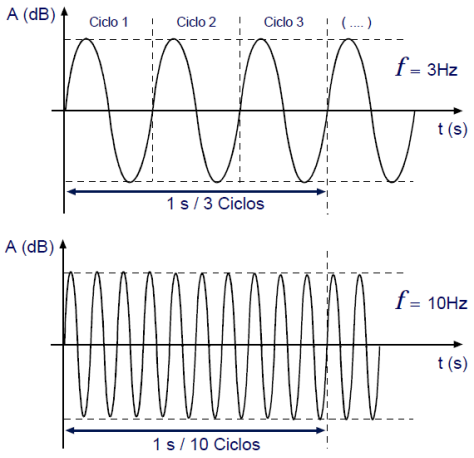
Com o intuito de facilitar a explicação sobre as etapas que compõem um corretor automático de tom, alguns termos da área de processamento de sinais de áudio e de teoria musical serão definidos.

2.1 Conceitos de Uso na Análise de Áudio Digital

2.1.1 O Sinal Digital

Inicialmente, precisa-se desenvolver o conceito de frequência, que é definida como o número de oscilações em um intervalo de tempo determinado (HALLIDAY; RESNICK; WALKER, 2008). Na Figura 1 há uma comparação entre sinais de 3 e 10 oscilações por segundo, ou seja, de 3 e 10 Hz. Em música, cada nota tem uma frequência, nomeada como *pitch*, ou tom.

Figura 1 – Frequência de 3 e 10 hertz



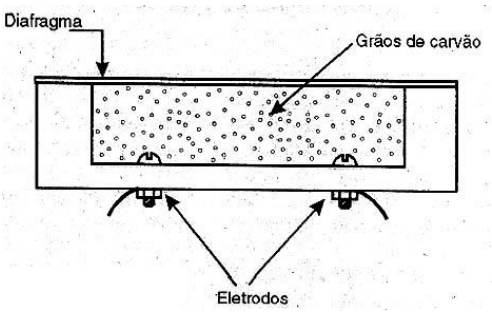
Fonte: Croce (2019).

“O sinal é uma função que representa uma quantidade ou variável física e contém informações sobre o comportamento ou a natureza do fenômeno.” (HSU, 2004). Existem os sinais contínuos e os discretos, que se diferenciam pela forma de como são avaliados, o primeiro em um tempo continuo e o segundo em um tempo discreto. Os discretos podem ser chamados de sinais digitais se o tempo discreto somente puder assumir um conjunto finito de valores distintos (HSU, 2004).

Os seres humanos conseguem ouvir frequências de 20 a 20000 Hz, e para conseguir gravar, reproduzir ou tratar esse som precisasse de algum equipamento que transforme a onda acústica em sinal elétrico, como os microfones (PEIMANI, 2009).

Os conversores ou transdutores são dispositivos que convertem uma forma de energia em outra. Os microfones são conversores eletromecânicos, ou seja, convertem a energia mecânica geralmente provinda de um instrumento musical e transformam em energia elétrica (FALCONE,1979). A Figura 2 mostra um microfone de carvão, ou piezoresistivo, onde o diafragma é uma membrana de metal, plástico ou outro material flexível que oscile quando as ondas sonoras incidam nele. Essa oscilação faz os grãos de carvão serem comprimidos e descomprimidos, mudando a resistência do sistema, e, assim, gera um sinal elétrico correspondente (COSTA, 2002).

Figura 2 – Microfone de carvão



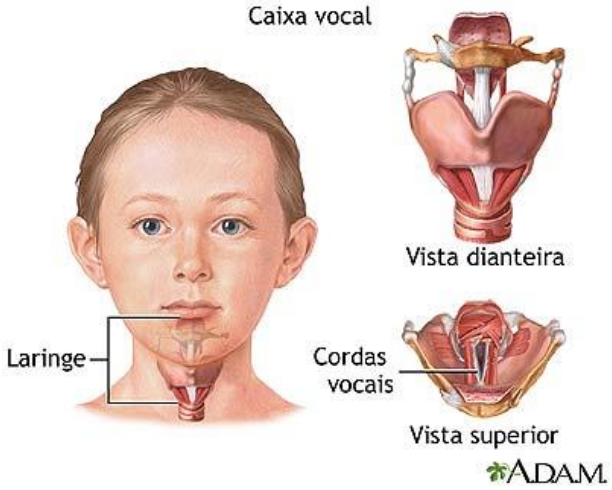
Fonte: Braga (2012).

As ondas sonoras podem ser geradas pela maioria dos seres humanos e essa emissão é chamada de voz.

2.1.2 A Voz e a Música

A emissão vocal é produzida pela pressão de ar que vem dos pulmões e faz vibrar as cordas vocais, ilustradas na Figura 3, gerando um som que é modificado na boca (VALE, 2012).

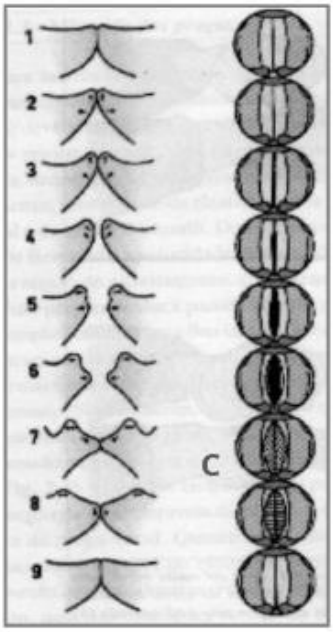
Figura 3 – Cordas vocais



Fonte: *Animated Dissection of Anatomy For Medicine* (2012).

A vibração mostrada na Figura 4 surge, pois os seres humanos têm a capacidade de unir as cordas vocais durante a expiração, produzindo uma série de aberturas e fechos que geram variações de pressão na passagem do fluxo de ar (VALE, 2012).

Figura 4 – Vibração das cordas vocais

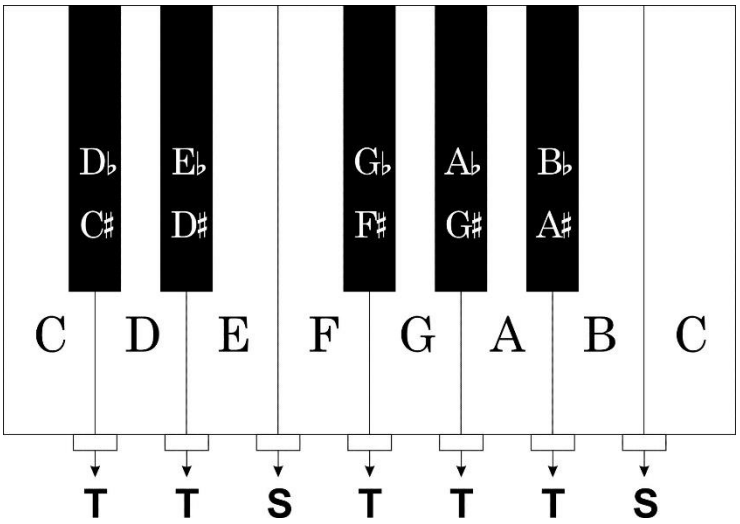


Fonte: Vale (2012).

Cada ser humano emite sons com um conjunto de frequências diferentes e isso forma o timbre. A frequência emitida de maior amplitude pela voz é a frequência fundamental, que traz a informação do tom emitido, e as outras são chamadas de harmônicas (PEIMANI, 2009).

O tom é utilizado na área musical para referenciar frequências emitidas por um instrumento. Na história da música diferentes escalas foram sugeridas e a mais utilizada é a escala de doze tons, onde de 12 em 12 teclas do piano o tom se repete (MULLER, 2015). Pode-se observar na Figura 5 que o tom é obtido por meio do intervalo entre qualquer primeira e terceira tecla do piano e que o semitom consiste na diferença entre qualquer primeira e segunda tecla do piano, onde T é tom e S semitom.

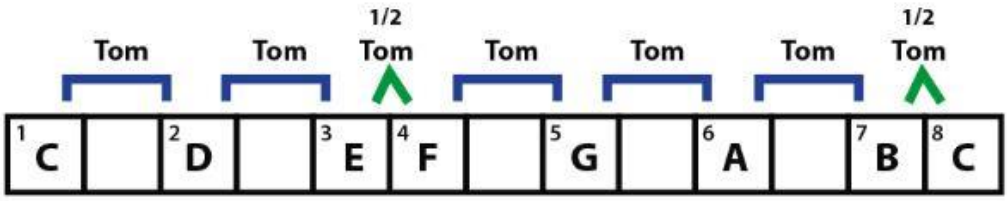
Figura 5 – Teclas do piano e seus tons



Fonte: Songster (2018).

As escalas surgiram a partir de um experimento com o monocórdio, antigo instrumento musical que possui apenas uma corda, onde encontrou-se as notas dó (*C)*, ré (*D)*, mi (*E)*, fá (*F)*, sol (*G)*, lá (*A)* e si (*B)*. Após estudos percebeu-se que essas notas tinham um padrão (PEREIRA; MADRUGA, 2020). Esse padrão é mostrado na Figura 6 e tem a seguinte sequência: tom, tom, semitom, tom, tom, tom, semitom. Por esse padrão começar de *C*, tem-se a escala de dó maior ou escala diatônica de dó maior.

Figura 6 – Escala diatônica de dó maior



Fonte: Bedani (2019).

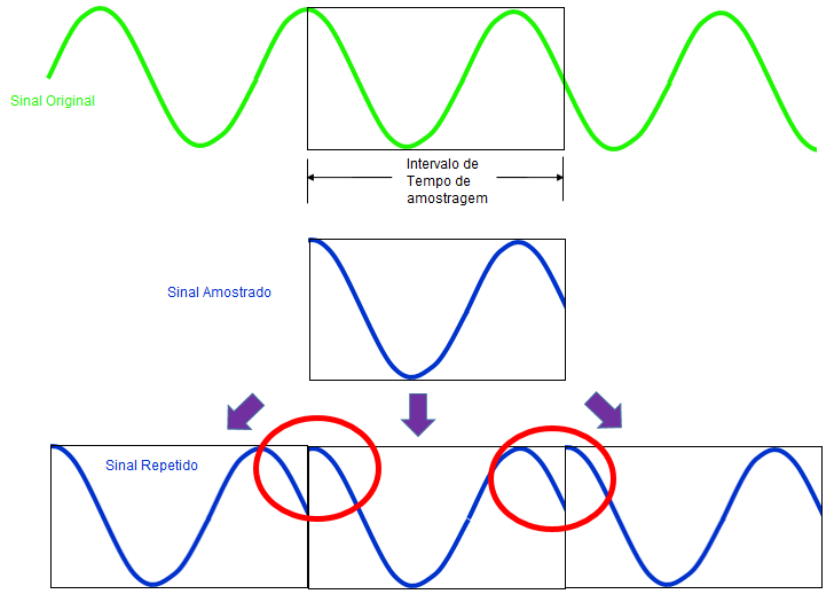
Após este estudo inicial é necessário desenvolver conceitos de algumas ferramentas utilizadas na manipulação do sinal digital.

2.1.3 Ferramentas de Uso em Processamento de Sinais

A primeira ferramenta a ser conceituada é a transformada discreta de Fourier (DFT, do inglês *discrete Fourier transform),* que converte sinais do domínio de tempo para o domínio da frequência. A DFT é a representação de Fourier adequada para uso em computadores digitais porque é discreta e de comprimento finito, nos domínios tanto de tempo como de frequência. A transformada rápida de Fourier (FFT, do inglês *fast Fourier transform*) é um algoritmo que torna o cálculo da DFT mais rápido de ser computado (GONÇALVES, 2004).

A técnica de janelamento é utilizada para analisar de forma mais eficiente pequenos pedaços do sinal, já que, ao cantar, o tom varia junto com o tempo e também por causa de um efeito indesejável chamado *leakage*. Esse efeito é causado pela amostragem da FFT, em que o sinal amostrado apresenta descontinuidades nas suas extremidades, pois dificilmente o ciclo inteiro do sinal se encaixa no tempo de medição (PINTO; SILVA; VANZELA, 2016). Na Figura 7, um tempo de amostragem é obtido e as descontinuidades são geradas.

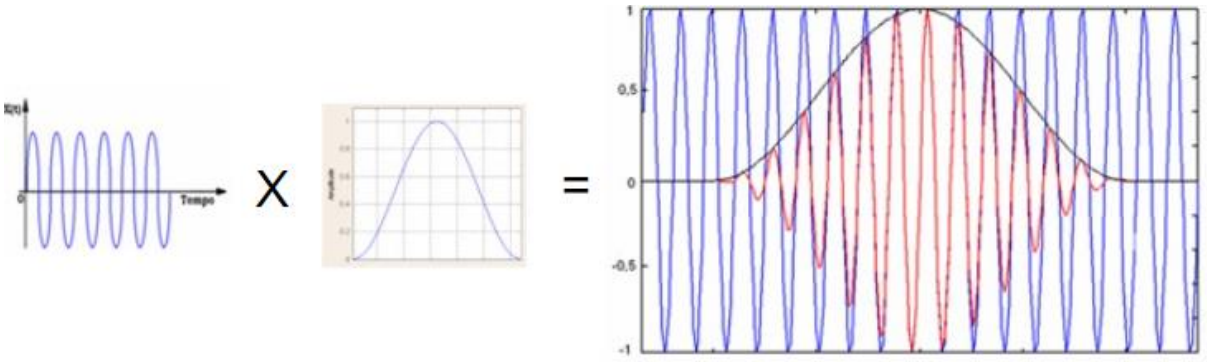
Figura 7 – Efeito de *leakage*



Fonte: Siemens (2019).

Observa-se que na Figura 8 existe um sinal de entrada multiplicado por uma janela e o resultado é uma onda de extremidades atenuadas. Dessa forma, consegue-se diminuir o efeito do fenômeno de *leakage*.

Figura 8 – Janelamento



Fonte: Mathias ([201-?]).

Existem muitos tipos de janelamento que geralmente recebem o nome de seus criadores. Segundo Luis Chioye (2018), a melhor utilização de cada uma dessas técnicas está no Quadro 1.

Quadro 1 – Melhores aplicações para cada janela

CONTEÚDO DO SINAL	JANELA
Caracterização CAD	Blackman Harris
Onda senoidal ou combinação de ondas	Hann
Onda senoidal (Quando precisão de amplitude é importante)	Flat Top
Sinal aleatório de banda estreita	Hann
Banda larga aleatória (Ruído branco)	Retangular
Ondas senoidais próximas	Retangular, Hamming
Sinal de excitação	Force
Sinais de resposta	Exponencial
Conteúdo desconhecido	Hann
Dois tons de frequências próximas, mas de amplitude diferente	Kaiser-Bessel
Dois tons de frequências próximas de amplitudes quase iguais	Retangular
Medições de amplitude de tom único	Flat Top

Fonte: Chioye (2018).
Nota: Adaptado e traduzido pelo autor.

3 SISTEMA PARA CORREÇÃO AUTOMÁTICA DE TOM

Um sistema para correção automática de tom, tal como proposto em Peimani (2009), consiste na execução sequencial de dois tipos de algoritmos, o de detecção e o de correção de tom.

3.1 Algoritmos de Detecção de Tom

O primeiro passo para corrigir o tom de um áudio é a obtenção da frequência fundamental. Primeiro deve-se entender que é necessário dividir o áudio em pequenas janelas para assim conseguir o tom de cada parte do sinal (PEIMANI, 2009). Também é interessante dizer que existem corretores que funcionam melhor para voz cantada e outros para voz falada (GERHARD, 2003). Pode-se estimar essa frequência pelos métodos de detecção de tom baseados na representação frequencial ou temporal. Ambos tipos de estimações serão apresentados.

3.1.1 Algoritmos Baseados na Representação Frequencial

O primeiro algoritmo baseado na representação frequencial que será comentado é o método de detecção do maior pico harmônico. Este método tem duas etapas, a determinação da representação frequencial do sinal de áudio via FFT e a definição da frequência fundamental, obtida como aquela frequência do espectro que apresenta a maior amplitude (GERHARD, 2003).

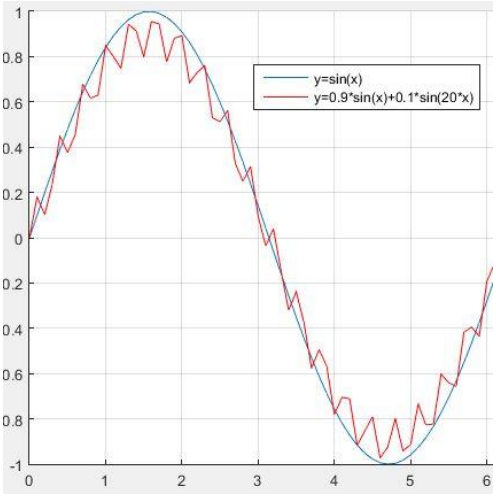
Outro algoritmo a ser discutido é o método de detecção da diferença entre picos harmônicos. Este possui a primeira etapa igual ao método anterior e na segunda, a frequência fundamental é calculada pela diferença entre os picos harmônicos, considerando que as frequências das componentes harmônicas são múltiplos inteiros da frequência fundamental (PEIMANI, 2009).

3.1.2 Algoritmos Baseados na Representação Temporal

A seguir serão comentados quatro métodos para a estimação de tom baseados na representação temporal do sinal de áudio, tendo como objetivo encontrar o período fundamental.

No método de detecção pela taxa de cruzamento por zero encontra-se o período por meio da taxa na qual o sinal de áudio cruza o eixo do tempo (PEIMANI,2009). O problema desse método ocorre quando há presença de altas frequências, que farão o sinal passar pelo zero mais de uma vez por ciclo, como mostra a Figura 9.

Figura 9 – Influência de altas frequências



Fonte: Produção do próprio autor.

A função da diferença de magnitude média (AMDF, do inglês *average magnitude difference function)* analisa a diferença entre ciclos, e parte da premissa que a diferença da magnitude média de um sinal de áudio permanecerá quase a mesma de um período para o outro. Em (1) tem-se a AMDF, onde *x[n]* é o sinal de áudio digital, τ é o período fundamental que é escolhido pelo valor que minimiza *a(τ)* (GERHARD, 2003).

$$a(\tau) = \sum_{n=0}^{N-1} |x[n] - x[n - \tau]| \quad (1)$$

A autocorrelação também analisa a magnitude de um sinal, porém a função é uma versão mais sofisticada da AMDF e as mudanças bruscas não afetarão mais o resultado (KONDOZ, 2004).

“O algoritmo YIN foi desenvolvido por Alain de Cheveigné e Hideki Kawahara, e tem esse nome por se basear na filosofia do yin yang, ou seja, um balanço entre técnicas.” (GERHARD, 2003, tradução nossa). Este código utiliza o método de autocorrelação com algumas melhorias que previnem erros na detecção da frequência. Segundo Cheveigne e Kawahara (2002), estes aprimoramentos diminuem em cerca de três vezes os desvios na detecção.

3.2 Algoritmos de Correção do Tom

É necessário trazer a frequência fundamental para o tom desejado após a detecção, podendo ser ajustado de forma automatizada ou de forma manual.

3.2.1 Algoritmos de Implementação de Forma Automatizada

Essa implementação pode ser feita ao multiplicar a razão entre a frequência correspondente do tom e a frequência a ser corrigida, como também por meio de algoritmos mais complexos. Um exemplo de algoritmos que podem ser utilizados nessa forma de implementação são os chamados PSOLA (do inglês *pitch synchronous overlap and add*), que são algoritmos que visam sintetizar novamente o sinal (CHENG, 2012).

Ao se falar em correção de tom, normalmente o *autotune* é lembrado. Esse *software* tem uma variação, o *autotune live*, que funciona em tempo real com um atraso de processamento e a correção é feita para o tom correto mais próximo (DIAZ, 2009). Por exemplo, se a frequência for de 435 Hz a voz está desafinada, pois não existe um tom nessa frequência, então ajusta-se para o tom mais próximo, que é o *A* de 440 Hz.

3.2.2 Algoritmos de Implementação Manual

O ajuste de forma manual é utilizado em estúdio, onde o operador define para qual tom a frequência será ajustada. Na música contemporânea estes algoritmos são usados tanto para fazer harmonias, fora do conjunto de notas que o cantor consegue emitir, como também na música eletrônica para produzir melodias e uma voz “robotizada”.

4 SOLUÇÃO PROPOSTA

4.1 Introdução

Há um desejo, por parte de programas e eventos de grande porte, de que se tenham apresentações perfeitas, mas o erro faz parte do ser humano, por isso, há um interesse no desenvolvimento de uma ferramenta orientada a corrigir os pequenos erros de entonação por parte dos cantores.

O presente projeto classifica-se por ser do tipo aplicado, por objetivar a solução do problema específico da correção automática de tom. O estudo exploratório da pesquisa é justificado pela pouca informação sobre algoritmos feitos sobre o tema e por ser uma tecnologia recente, utilizando o método experimental para possibilitar a manipulação de variáveis, ou seja, as amostras de sinais de áudio serão tratadas e o algoritmo de correção de tom será aplicado. No trabalho serão testados códigos e, por isso, a pesquisa é classificada como quantitativa, visando a algum resultado inicial.

Assim, a solução proposta é um procedimento feito no MATLAB e em PYTHON que detecte e corrija pequenos erros da emissão vocal, e que futuramente possa ser melhorado e implementado em uma plataforma embarcada a fim de que gere uma autocorreção de tom em tempo real.

4.2 Implementação

O procedimento implementado para obter e corrigir o áudio é composto de quatro etapas, explicadas a seguir:

a) Áudio de entrada – Será utilizado um sinal de frequência crescente e faixas obtidas pelo conjunto de dados de sinais de áudio e pelo microfone.
b) Preparando a análise – Para preparar a análise do sinal se faz necessário uma tabela com as frequências dos tons, podendo escolher uma escala musical, como a escala de dó maior, ou apenas colocar todos os semitons, respeitando o limite de frequência escolhido.

c) Detecção de frequências – A detecção é feita pelo método do maior pico harmônico. Utiliza-se a função *spectrogram* que gera como resultado a transformada de Fourier de curto tempo (STFT, do inglês *short-time Fourier transform*). Trata-se da transformada discreta de Fourier aplicada a uma janela de tempo. O janelamento Hann foi utilizado pois, segundo Chioye (2018), é a técnica mais adequada para combinação de ondas senoidais.
d) Correção de Frequências – O ajuste acontecerá por meio da multiplicação da razão entre frequências e depois aplica-se a transformada inversa de Fourier de curto tempo para obter o áudio corrigido.

A seguir, tem-se os resultados obtidos por meio do procedimento descrito, maiores detalhes sobre as implementações utilizadas e comparação com os resultados de outras literaturas.

5 RESULTADOS

5.1 Recursos Computacionais

5.1.1 Recursos de *Software*

A implementação dos algoritmos de correção de tom é efetuada utilizando o *software* MATLAB, que tem a sua versão de estudante disponível aos alunos de graduação da Universidade Federal do Espirito Santo (UFES), e o ambiente de desenvolvimento do *Google Colaboratory* (*Colab*), que usa a linguagem PYTHON.

5.1.2 Recursos de *Hardware*

Além disso, a máquina utilizada nos experimentos apresenta a seguinte configuração:

a) Sistema operacional Windows 10;
b) Processador AMD FX–8120, 3.10 GHz com 8 núcleos físicos;
c) Memória RAM de 8 GB;
d) Unidade de armazenamento de 2 TB (disco rígido);
e) Placa de vídeo AMD Radeon HD 6670, com 5 GB de memória dedicada.

5.1.3 Conjunto de Dados de Sinais de Áudio

Para a avaliação dos métodos implementados de correção automática de tom foi efetuado um estudo sobre os conjuntos de dados de sinais de áudio orientados ao problema e foi selecionado o conjunto de Black, Li e Miblack (2014). Tal conjunto contém mais de setenta gravações de vozes de vinte e oito cantores profissionais, semiprofissionais e amadores, cantando o estilo musical ópera. Todas as gravações estão com 44,1 kHz de taxa de amostragem.

5.2 Detalhes da Implementação

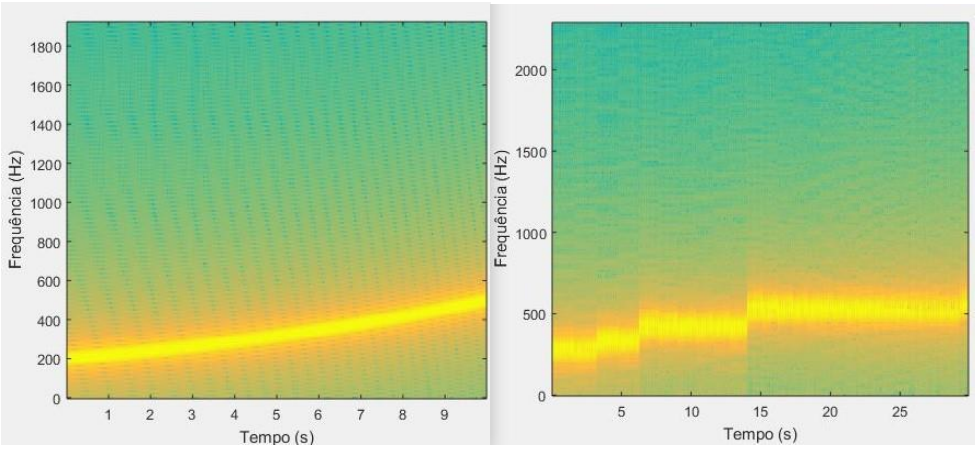
Foram estudadas diferentes implementações e desenvolveu-se um código base em MATLAB. Esse algoritmo foi transcrito para PYTHON a fim de que futuramente o trabalho possa ser melhor compreendido e que possíveis melhorias sejam mais fáceis de implementar. Além de

transcrever o código houve uma melhoria, pois na primeira versão existia um erro que ajustava a frequência para um tom mais alto do que o correto.

5.2.1 Implementação no MATLAB

Inicialmente, utiliza-se a função *spectrogram* que faz a FFT para a janela do tipo Hann do sinal de entrada. Com a função *find*, encontra-se o pico de maior amplitude entre um conjunto de frequências. Esse valor é comparado com os da tabela de frequências corretas, assim o tom com a próxima frequência é encontrado. Cria-se então uma relação entre o pico de maior amplitude, que é a frequência fundamental, e o tom mais próximo. Por fim, todas as frequências dessa janela são multiplicadas pela relação criada. A Figura 10 mostra o sinal de áudio de frequência crescente, em (a), e esse sinal corrigido para o tom mais próximo, em (b).

Figura 10 – Sinal de áudio: (a) original e (b) corrigido via a implementação no MATLAB



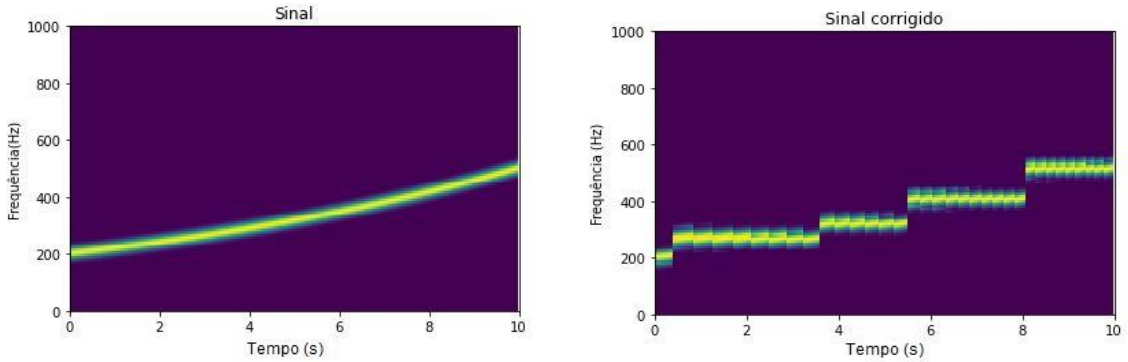
(a) (b)

Fonte: Produção do próprio autor.

5.2.2 Implementação em PYTHON

O algoritmo começa pela função *spectrogram* da biblioteca *signal* para obter a FFT com a janela do tipo Hann do sinal de entrada, e por meio da função *where* da biblioteca *numpy*, encontra-se o pico de maior amplitude entre o conjunto de frequências. Esse pico é utilizado para obter o próximo tom da escala escolhida e a razão entre esses dois valores é calculada e utilizada para multiplicar todas as frequências da janela. A Figura 11 apresenta o sinal de frequência crescente, em (a), e o sinal corrigido para o tom mais próximo, em (b).

Figura 11 – Sinal de áudio: (a) original e (b) corrigido via a implementação em PYTHON



(a) (b)

Fonte: Produção do próprio autor.

A partir das Figuras 10 e 11 pode-se concluir que houve uma melhor resposta em PYTHON e grande parte disso é motivada pelos aprimoramentos feitos na correção do tom no primeiro código que fazia o sinal ser ajustado para uma frequência errada. Ao analisar a Figura 10, observa-se que uma frequência de 200 Hz, em (a), foi ajustada para 250 Hz, em (b), e gerou um sinal com um tempo maior que o dobro do original. Já na Figura 11, essa correção foi mais adequada e encontra um tom próximo de 200 Hz, além de manter o tempo do áudio. E assim, por esses motivos, as métricas desenvolvidas na próxima seção foram aplicadas no código em PYTHON.

O pseudocódigo do algoritmo implementado em MATLAB e PYTHON é mostrado na Figura 12.

Figura 12 –Pseudocódigo do autocorretor

```
Autocorretor_Pseudo-Código

inicio
    Onda, fs := Importar_audio("audio.wav")
    S, f, t := Espectograma(Onda,Tamanho_janela,Tamanho_FFT)
    Plotar(t,f,S)
    S_corrigido := Corretor_Tom(S, f, t,Tabela_de_tons)
    Plotar(t,f,S_corrigido)
    Erro := Erro(S_corrigido, f, Tabela_de_tons)
    Onda_corrigida := Transf_Fourier_Inversa(S_corrigido,fs,Janela)
    Ouvir_audio(Onda_corrigida,fs)

fimalgoritmo

funcao Espectograma (onda, Tamanho_janela, Tamanho_FFT)
inicio
    Janela := Janelamento(Tipo,Tamanho_janela)
    f , t , S := Gerar_Espectograma(onda,fs ,Janela, Tamanho_FFT)
    return [S, f, t]
fimfuncao
```

Fonte: Produção do próprio autor.

5.3 Métricas

Considerando que a correção de tom dos sinais é um problema de alteração de frequência, a avaliação aqui utilizada será feita via as seguintes métricas:

a) Qualidade sonora (Qs), uma métrica avaliada pelo próprio autor ao escutar a faixa de áudio corrigida, podendo ser classificado conforme o Quadro 2.

Quadro 2 – Qualidade sonora

CLASSIFICAÇÃO	CARACTERÍSTICA
A	Faixa de áudio sem ruído e afinado
B	Faixa de áudio com ruído e afinado
C	Faixa de áudio sem ruído e com alguns erros de afinação
D	Faixa de áudio com ruído e com alguns erros de afinação

Fonte: Produção do próprio autor.

b) Erro (Er), calculado pela diferença entre a frequência do tom correto e a frequência ajustada.

A afinação será considerada como a manutenção em um tom e a desafinação, como a oscilação entre tons.

O código terá 3 tabelas a fim de verificar as métricas já explicadas. As duas primeiras utilizam os tons entre o *C* de 16,352 Hz e o *F* de 22351 Hz, segundo Iazzetta ([200-?]), e diferenciam-se nas escalas escolhidas, já que uma usa todos os semitons e a outra a escala de dó maior. O terceiro conjunto de valores é de frequências espaçadas para avaliar melhor se os ajustes estão acontecendo e, por isso, só será aplicado na onda de frequência crescente.

5.3.1 Avaliação

No Quadro 3 pode-se observar os erros, em hertz, ocasionados principalmente pelo fato da função *spectogram* gerar valores fixos e linearmente espaçados de frequências, e a qualidade sonora do áudio que apresentou ruídos no fim de cada janela, mas teve uma afinação desejada.

Quadro 3 – Resultados para a onda de frequência crescente de 200 a 500 Hz

METODOLOGIA	Qs	Er (Hz)
PRIMEIRA CONFIGURAÇÃO	B	13,6518
SEGUNDA CONFIGURAÇÃO	B	30,7185
TERCEIRA CONFIGURAÇÃO	B	18,8245

Fonte: Produção do próprio autor.

Os resultados para a faixa de áudio são mostrados no Quadro 4 e, ao comparar com o Quadro 3, conclui-se que os erros foram maiores e a classificação de qualidade sonora pior, ocasionados pela maior variação tonal do sinal de entrada.

Quadro 4 – Resultados para uma faixa de áudio do banco de dados

METODOLOGIA	Qs	Er (Hz)
PRIMEIRA CONFIGURAÇÃO	D	31,2203
SEGUNDA CONFIGURAÇÃO	D	40,2507
TERCEIRA CONFIGURAÇÃO	D	Não se aplica

Fonte: Produção do próprio autor.

Pode-se concluir que, independente da configuração, o código de autocorreção ajusta a frequência para o próximo tom. O problema relacionado aos ruídos sempre esteve presente, ruído esse ocasionado pelo fenômeno de *leakage* e por causa da codificação feita que gera valores fixos e linearmente espaçados de frequências, que muitas vezes não são iguais às frequências dos tons. Ao utilizar uma escala especifica a implementação apresenta uma melhor afinação em relação a melodia da música, porém é necessário algum aprimoramento a fim de identificar em que escala a pessoa está cantando.

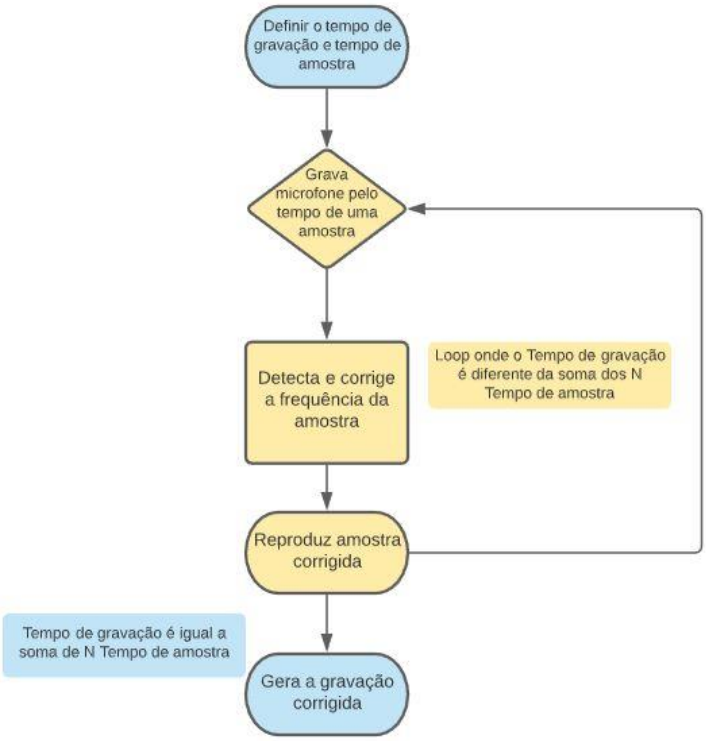
5.4 Código que usa Faixas do Microfone

O código que usa faixas do microfone tenta se aproximar de um código em tempo real, foi feito em PYTHON e tem o sinal de entrada obtido pela gravação do microfone, funcionando da seguinte forma:

a) Define-se o tempo de gravação e o tempo da amostra;
b) Cada amostra irá passar individualmente pelo processo de detecção e correção e reproduzirá o resultado até que se atinja o tempo de gravação;
c) No fim, é gerado o áudio corrigido de toda gravação.

Ou seja, ao escolher um tempo de gravação de 20 segundos com um tamanho de amostra de 0,5 segundos, o microfone irá gravar o tempo de uma amostra, corrigir a frequência e reproduzir a faixa de saída, após esses itens a próxima amostra do tempo estabelecido é gravada e o ciclo se repete, como mostrado no fluxograma da Figura 13. A princípio pensava-se em aprimorar esse código ao ponto de diminuir os espaços, perdas de áudio, gerados pelo tempo de processamento da correção de cada amostra e assim conseguir se aproximar de uma correção em tempo real, mas decidiu-se que antes de continuar a desenvolver o código é necessário melhorar a implementação inicial, que ainda apresenta ruídos.

Figura 13 – Fluxograma do código implementado



Fonte: Produção do próprio autor.

5.5 Comparação de Resultados

5.5.1 Primeira Comparação

A primeira comparação será feita com o trabalho de Gerhard (2003), que faz a implementação de algumas técnicas de detecção de frequência. Nesta literatura, as faixas de entrada utilizadas foram áudios de conversação e a detecção é baseada tanto na representação temporal, implementando o algoritmo YIN, quanto na representação frequencial. Também faz o uso de um *software* de análise de fala denominado Colea, desenvolvido por Philip Loizou (LOIZOU, 2003), e traz consigo a métrica de erro médio entre os métodos utilizados, existindo erros entre 11,11 Hz e 43,75 Hz. Comparando com os valores do trabalho aqui desenvolvido, 31,2203 Hz e 40,2507 Hz, conclui-se que são resultados aceitáveis, porém precisam de melhorias. A métrica qualidade sonora não pôde ser analisada, pois o trabalho somente faz a detecção das frequências.

5.5.2 Segunda Comparação

O trabalho de Cheng (2012) utiliza como sinal de entrada faixas de músicas pop conhecidas, divide o sinal em 3 blocos, um dos blocos passa por processos de detecção de frequência por autocorrelação, por maior pico harmônico e usa algumas ferramentas de aprimoramento. A correção de frequência é feita pela ferramenta PSOLA.

Neste trabalho a métrica de erro não é utilizada, porém, segundo Corey Cheng (2012, tradução nossa), “Embora não tendo realizado testes formais de escuta, nós comparamos informalmente nosso sistema com a saída da linha de produtos Antares Autotune, e descobriu que os dois sistemas produziram resultados comparáveis.”. Então, a obtenção de uma saída comparável com itens comerciais será considerada com uma qualidade sonora do tipo A, ou seja, o sinal de saída foi afinado e sem a presença de ruídos, mostrando ser o trabalho com as melhores métricas, por apresentar muito mais ferramentas para a correção do áudio e a diminuição de ruídos.

6 CONCLUSÕES E PROJETOS FUTUROS

Neste trabalho são efetuados estudos e uma análise inicial das implementações de algoritmos de detecção e correção de tom. Especificamente, foi possível utilizar a técnica de detecção do maior pico harmônico e uma correção inicial.

A implementação desenvolvida faz de forma automática a alteração da frequência, mas o corretor automático de tom não leva em consideração a melodia em si, então, mesmo corrigindo para um tom, não significa que será o correto. Nesse contexto, se levar em consideração a utilização do código para gravações de cantores profissionais, será necessário desenvolver algo que diminua os ruídos e que possibilite a correção para uma frequência específica.

Em relação aos resultados do algoritmo de correção de tom, conseguiu-se obter um corretor automático que exerce a função de corrigir a frequência emitida ao próximo tom da tabela de frequências corretas, porém apresentou ruídos próprios do fenômeno de *leakage*. O corretor implementado foi testado usando faixas de teste e da base de dados de cantores de Black, Li e Miblack (2014), sendo avaliado levando em consideração as métricas de qualidade sonora e erro, obtendo um erro médio de 25 hertz e uma qualidade sonora mediana, resultados considerados aceitáveis para uma implementação inicial.

Por fim, sugere-se como trabalhos futuros o estudo de maneiras de diminuir o ruído por meio de filtros ou pela implementação de outros métodos de detecção e de correção de frequência.

REFERÊNCIAS BIBLIOGRÁFICAS

ANIMATED DISSECTION OF ANATOMY FOR MEDICINE. **Laringe.** 2012. 1 ilustração. Disponível em: https://ssl.adam.com/content.aspx?productid=125&pid=70&gid=19708&site=bestdoctors.adam.com&login=BEST4545. Acesso em:10 nov. 2020.

BEDANI, F. **O que é uma escala musical?.** 2019. 1 ilustração. Disponível em: https://blog.planetamusica.net/o-quee-uma-escala/. Acesso em: 24 set. 2020

BLACK DAWN A. A.; LI, M.; TIAN, M. Automatic Identification of Emotional Cues in Chinese Opera Singing. *In:* INTERNATIONAL CONFERENCE ON MUSIC PERCEPTION AND COGNITION AND ASIA-PACIFIC SOCIETY FOR THE COGNITIVE SCIENCES OF MUSIC, 13., 2014, Seoul. **Proceedings** [...]. Seoul: Yonsei University, 2014. Disponível em: http://isophonics.net/SingingVoiceDataset. Acesso em: 18 out. 2018.

BRAGA, N. C. **Como funciona o microfone**. 2012.1 ilustração. Disponível em: http://www.newtoncbraga.com.br/index.php/como-funciona/5679-art734. Acesso em: 15 maio 2020.

BREGTIZER, L. **Secrets of Recording**: Professional Tips, Tools Techniques. Burlington, Massachusetts: Focal Press, 2009. p. 105.

CHENG, C. Design of a pitch quantization and pitch correction system for real-time music effects signal processing. *In:* ASIA-PACIFIC SIGNAL AND INFORMATION PROCESSING ASSOCIATION ANNUAL SUMMIT AND CONFERENCE, 4., 2012 California. **Proceedings** [...]. California: APSIPA. 2012. Disponível em: http://www.apsipa.org/proceedings_2012/papers/375.pdf. Acesso em: 24 set.2020.

CHEVEIGNE, A. D.; KAWAHARA, H. K. Yin, a fundamental frequency estimator for speech and music. **The Journal of the Acoustical Society of America**, [s. l.], v. 111, n. 4, p. 1917-1930, 2002. Disponível em: http://audition.ens.fr/adc/pdf/2002_JASA_YIN.pdf. Acesso em: 15 out. 2020.

CHIOYE, B. **TI Precision Labs – ADCs: Fast Fourier Transforms (FFTs) and Windowing**. Apresentado por Peggy Liska. [S. l.]: Texas Instruments Precision Labs, 2018. 1 vídeo (11 min). Publicado pelo canal Texas Instruments. Disponível em: https://www.youtube.com/watch?v=dCeHOf4cJE0. Acesso em: 13 set. 2020.

CROCE, C. **Isolamento acústico – Lei da Massa**. 2019. 1 ilustração. Disponível em: https://portalacustica.info/isolamento-acustico-lei-da-massa/. Acesso em: 10 nov. 2020.

COSTA, G. D. Microfones Características e Aplicações. *In:* SEMINÁRIO DE ENGENHARIA DE ÁUDIO DA UFMG, 1., 2002. Belo Horizonte. **Anais** [...]. Belo Horizonte, MG: UFMG, 2002. Disponível em: https://dgcaudio.com.br/wp-content/uploads/2016/06/microfones_-_caracteristicas_e_aplicacoes.pdf. Acesso em: 22 nov. 2020.

FALCONE, G. A. **ELETROMECÂNICA.** 1 ed. São Paulo: Edgar Blücher, 1979. p. 1.

GERHARD, D. **Pitch extraction and fundamental frequency**: History and current techniques. Regina, Canadá: University of Regina, 2003. 22 p. ISBN 0-7731-0455-0. Disponível em: https://www.researchgate.net/publication/233778945_Pitch_Extraction_and_Fundamental_Frequency_History_and_Current_Techniques. Acesso em: 15 jun. 2020.

GONÇALVES, L. A. **Um estudo sobre a transformada rápida de Fourier e seu uso em processamento de imagens.** 2004. Dissertação (Mestrado em Matemática Aplicada) – Programa de Pós-Graduação em Matemática Aplicada, Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2004. Disponível em: https://www.lume.ufrgs.br/bitstream/handle/10183/6748/000446124.pdf?sequence=1. Acesso em: 20 nov. 2018.

HALLIDAY, D.; RESNICK, R.; WALKER, J. **Fundamentos de Física.** 8. ed. Rio de Janeiro: LTC, 2008. v. 2. p. 86-171.

HSU, H. P. **Teoria e Problemas de Sinais e Sistemas**. Porto Alegre: Bookman, 2004. p. 13, 14.

IAZZETTA, F. **Tabela de Frequências, Períodos e Comprimentos de Onda**. [200–?]. Disponível em: http://www2.eca.usp.br/prof/iazzetta/tutor/acustica/introducao/tabela1.html. Acesso em: 22 out. 2018.

KONDOZ, A. M. **Digital Speech**: Coding for Low Bit Rate Communication Systems. 2. ed. England: John Wiley Sons, 2004. p. 152-155.

LOIZOU, P. **Colea:** A Matlab software tool for speech analysis. Dallas, Texas: University of Texas. 2003. Disponível em: ecs.utdallas.edu/loizou/speech/manual.pdf. Acesso em: 05 nov. 2020

MATHIAS, M. H. **Ferramentas de diagnóstico de máquinas.** [201-?]. 1 ilustração. Disponível em: https://slideplayer.com.br/slide/379943/. Acesso em: 8 nov. 2020.

MILLARD, A. **America on Record:** A History of Recorded Sound. 2. ed. Cambridge: Press Syndicate of the University of Cambridge, 1996. p. 1, 2.

MORTON, D. **Off the record:** The technology and culture of sound recording in America. Burlington, MA: Focal Press, 2000. p. 5.

MULLER, M. **Fundamentals of music processing**. 1. ed. Germany: Springer International Publishing Switzerland, 2015.

PEIMANI, M. A. **Pitch correction for the human voice**. 2009. Trabalho de Conclusão de Curso (Bachelor of Science in Physics) – University of California Santa Cruz, Santa Cruz, 2009. Disponível em: http://dave.ucsc.edu/physics195/thesis_2009/m_peimani.pdf. Acesso em: 17 set.2018.

PEREIRA, A. A.; MADRUGA, Z. E. F. Música e Modelagem Matemática: Representação de uma escala musical por meio de modelo matemático. **Ensino da Matemática em Debate,** [s. l.], v. 7, n. 1, p. 1-25, 2020. Disponível em: https://revistas.pucsp.br/emd/article/view/45029. Acesso em: 24 nov.2020.

PINTO; SILVA; VANZELA. Processamento de sinais aplicado à análise comparativa do timbre de instrumentos produzidos com madeiras nacionais e tradicionais. *In:* SIMPÓSIO DE MECÂNICA COMPUTACIONAL, 12., 2016. Diamantina. **Anais** [...]. Diamantina, MG: ABMEC, 2016. Disponível em: https://dgcaudio.com.br/wp-content/uploads/2016/06/microfones_-_caracteristicas_e_aplicacoes.pdf. Acesso em: 22 nov.2020.

SIEMENS. **Windows and Spectral Leakage.** 2019. 1 ilustração. Disponível em: https://community.sw.siemens.com/s/article/windows-and-spectral-leakage. Acesso em: 8 nov. 2020.

SONGSTER. **Quais as diferenças entre sustenido e bemol?.** 2018. 1 ilustração. Disponível em: http://blog.songster.com.br/2018/01/12/quais-as-diferencas-entre-sustenido-e-bemol-2/. Acesso em: 10 nov. 2020.

TREKHLEB, O. **Playing with Discrete Fourier Transform Algorithm in JavaScript.** 2018. 1 ilustração. Disponível em: https://dev.to/trekhleb/playing-with-discrete-fourier-transform-algorithm-in-javascript-53n5. Acesso em: 10 nov. 2020.

VALE, S. M. M. D. **Emissão vocal**: Uma visão física, fisiológica e psicológica das pregas vocais. 2012. Trabalho Acadêmico (Mestrado Especialização em Canto) – Escola das Artes, Universidade Católica Portuguesa, Porto. 2012. Disponível em: https://www.meloteca.com/wp-content/uploads/2018/11/emissao-vocal.pdf. Acesso em: 15 maio 2020.