

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

ANA CAROLINA FROZZA

**DETECÇÃO DE ACORDES MUSICAIS POR MEIO
DE INFORMAÇÕES ESPECTRAIS**

MONOGRAFIA

CAMPO MOURÃO

2019

ANA CAROLINA FROZZA

**DETECÇÃO DE ACORDES MUSICAIS POR MEIO
DE INFORMAÇÕES ESPECTRAIS**

Trabalho de Conclusão de Curso de Graduação apresentado à disciplina de Trabalho de Conclusão de Curso II, do Curso de Bacharelado em Ciência da Computação do Departamento Acadêmico de Computação da Universidade Tecnológica Federal do Paraná, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Rodrigo Hübner

CAMPO MOURÃO

2019



ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

Às **19:30** do dia **2 de julho de 2019** foi realizada na sala **E101** da UTFPR-CM a sessão pública da defesa do Trabalho de Conclusão do Curso de Bacharelado em Ciência da Computação do(a) acadêmico(a) **Ana Carolina Frozza** com o título **DETECÇÃO DE ACORDES MUSICAIS POR MEIO DE INFORMAÇÕES ESPECTRAIS**. Estavam presentes, além do(a) acadêmico(a), os membros da banca examinadora composta por: **Prof. Dr. Rodrigo Hübner** (orientador), **Prof. Dr. Andre Luiz Satoshi Kawamoto** e **Prof. Dr. Frank Helbert Borsato**. Inicialmente, o(a) acadêmico(a) fez a apresentação do seu trabalho, sendo, em seguida, arguido(a) pela banca examinadora. Após as arguições, sem a presença do(a) acadêmico(a), a banca examinadora o(a) considerou _____ na disciplina de Trabalho de Conclusão de Curso **2** e atribuiu, em consenso, a nota ____ (_____). Este resultado foi comunicado ao(à) acadêmico(a) e aos presentes na sessão pública. A banca examinadora também comunicou ao acadêmico(a) que este resultado fica condicionado à entrega da versão final dentro dos padrões e da documentação exigida pela UTFPR ao professor Responsável do TCC no prazo de **onze dias**. Em seguida foi encerrada a sessão e, para constar, foi lavrada a presente Ata que segue assinada pelos membros da banca examinadora, após lida e considerada conforme.

Observações: _____

Campo Mourão, **2 de julho de 2019**

**Prof. Dr. Andre Luiz Satoshi
Kawamoto**
Membro 1

Prof. Dr. Frank Helbert Borsato
Membro 2

Prof. Dr. Rodrigo Hübner
Orientador

A ata de defesa assinada encontra-se na coordenação do curso.

Resumo

Frozza, Ana. detecção de acordes musicais por meio de informações espectrais. 2019. 34. f. Monografia (Curso de Bacharelado em Ciência da Computação), Universidade Tecnológica Federal do Paraná. Campo Mourão, 2019.

Músicos e cientistas musicais necessitam cada vez mais de ferramentas computacionais que auxiliem em seus trabalhos. Um exemplo disso são ferramentas que reconheçam propriedades de forma automática na música. O intuito deste trabalho é a realização da detecção de acordes por meio de informações espectrais. A estrutura básica do projeto envolve um sinal áudio, que foram convertido para o domínio da frequência, resultando no espectro de potência do sinal. Este então foi utilizado para extração dos valores do vetor de notas, que foram usados posteriormente como entrada para a classificação dos acordes referente ao áudio de entrada, gerando assim a cifra da música tocada como saída. Os resultados acadêmicos foram muito satisfatórios, a fidelidade do Sistema de Detecção alcançou para o melhor cenário 90% de acerto gerando assim uma média de 80% de acerto para todas as amostras testadas. Esta discrepância de resultados não ocorreu devido ao processo de reconhecimento, mas sim ao formato habitual com que os músicos executam estes acorde e o ambiente onde foram gravados. Este trabalho, com seus resultados práticos, poderá servir como referência para outros que envolvam tópicos aqui relacionados, como por exemplo, percepção computacional da música ou processamento digital de sinais.

Palavras-chaves: Detecção de acordes; Análise espectral; Processamento de sinal de áudio.

Abstract

Frozza, Ana. . 2019. 34. f. Monograph (Undergraduate Program in Computer Science), Federal University of Technology – Paraná. Campo Mourão, PR, Brazil, 2019.

Musicians and music scientists are increasingly in need of computational tools to aid their work. An example of this are tools that automatically recognize properties in music. The purpose of this work is to perform the chord detection through spectral information. The basic structure of the project involves an audio signal, which will be converted to the frequency domain, resulting in the power spectrum of the signal. This will then be used to extract the values of the note vector, which will be used later as input for the classification of the chords for the input audio, thus generating the cipher of the music played as output. The academic results were very satisfactory, the reliability of the Detection System achieved for the best case scenario was 90 % of success thus generating an average of 80 % of correctness for all samples tested. This variance of results did not occur due to the recognition process, but rather to the habitual format with which the musicians execute these chords and the environment that were recorded. This work, with its practical results, may serve as a reference for others that involve topics related here, such as, computational perception of music or digital signal processing.

Keywords: Chord detection; Spectral analysis; Processing of audio signal.

Lista de figuras

2.1	Espectrograma de uma nota D4	17
4.1	Fluxograma de etapas da metodologia.	22
4.2	Extração de picos do espectro	24
4.3	Modelo Simples	24
4.4	Modelo Harmônico	25
5.1	Resultados do Grupo 1.	28
5.2	Experimentos X Amostras.	29
5.3	Números de Janelas por Teste.	30

Lista de tabelas

2.1	Tabela de cifras	12
2.2	Tabela de notas e seus acidentes musicais	12
2.3	Tabela de frequências sonoras em Hz	13
2.4	Tabela de formação de Tríade	14
2.5	Tabela Tríade Maior	14
2.6	Tabela Tríade Menor	15
2.7	Tabela Tríade Aumentada	15
2.8	Tabela Tríade Diminuta	15
2.9	Tabela de comparação de parâmetros entre STFT CQT	18
5.1	Porcentagem em Mediana de acerto dos grupos	28

Siglas

- CQT: Constante-Q de Transformação
DSP: Processamento Digital de Sinais
FFT: Transformada Rápida de Fourier
MIDI: *Musical Instruments Digital Interface*
PCP: *Pitch Class Profiles*
STFT: Transformada de Fourier de Curto Tempo

Sumário

1	Introdução	9
2	Fundamentação Teórica	11
2.1	Introdução a Música	11
2.1.1	Escala Musical	13
2.2	Frequências	16
2.3	Transformada de Fourier de Curto Tempo	16
2.4	Constante-Q de Transformação	17
3	Estado da Arte	19
3.1	Sistema Automático de Transcrição Melódica	19
3.2	Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music.	19
3.3	Sistema Especialista Para Reconhecimento De Acordes Musicais Em Tempo Real Para Violão Elétrico Utilizando Técnicas De DSP	20
3.4	Considerações Finais	20
4	Método Proposto	21
4.1	Materiais	21
4.2	Metodologia	21
4.3	Áudio	22
4.4	Análise Espectral	22
4.5	Detecção dos Onsets	23
4.6	Extração de Picos	23
4.7	Classificador de Acordes	24
4.8	Considerações Finais	25
5	Experimentos e Resultados	26
5.1	Experimentos	26
5.2	Resultados	27
5.3	Considerações Finais	29

6	Conclusões	31
	Referências	33

Introdução

Esta dissertação busca descrever os possíveis métodos e aplicações da tecnologia da informação na música, buscando correlacionar conceitos e teorias importantes para a integração da música e suas teorias ao processamento digital de sinais. Tendo isso em vista, foi criado um Sistema de detecção capaz de identificar acordes.

De acordo com Ferreira (1986) sinais de áudio contínuos cuja variação em relação ao tempo é a representação proporcional de outra variável temporal. Transcrever um sinal de áudio significa encontrar qual foi a sequência de movimentos que o músico executou para gerar o sinal. No contexto musical, essa tarefa está associada ao ato de escutar e escrever os acordes tocados, porém nem todos os músicos tem a capacidade de saber exatamente qual acorde foi reproduzido.

A transcrição musical automática é algo que há muito tempo atrai o interesse dos músicos, isso se dá pela realização da transcrição por um dispositivo, sem qualquer interferência humana e pode ser entendida como um processo de extração de informação aplicada sobre um conjunto de dados de sinais correspondentes a música.

Atualmente existem alguns aplicativos que fazem a tradução para o arquivo do tipo *Musical Instruments Digital Interface (MIDI)*, como é o caso dos aplicativos *ScoreCleaner Notes* e o *Chordify*.

A tarefa de tradução de uma música por um software envolve vários aspectos que devem ser avaliados, como: a velocidade em que a música é tocada, os instrumentos utilizados, a frequência das notas fundamentais e a duração de cada nota.

Devido a estes parâmetros serem muito complexos de mensurar, foi adotado neste trabalho a cifra como notação musical, pois considera basicamente a frequência das notas e a análise dos diferentes tipos de instrumentos musicais diferentes e ritmos foram ignorados.

No decorrer deste trabalho três preocupações estavam presentes:

- Evitar que irregularidades na estimativa da frequência de referência ocasionasse notas espúrias.
- Eliminar no estágio inicial o atraso de uma nota.
- Impedir que uma sequência de notas repetidas seja interpretada como uma única nota.

A motivação deste trabalho está na necessidade da criação de um software de detecção de acordes musicais. O público-alvo são músicos em geral, desde auxiliar novos músicos até contribuir com o trabalho dos mais experientes.

Como fonte sonora neste trabalho foi escolhido o violão, por se tratar de um instrumento popular e que normalmente é usado em estudos de outros artigos, dos quais poderá ser comparado com os resultados obtidos neste trabalho .

Este estudo tem como objetivo desenvolver um software capaz de identificar acordes, para isso o primeiro passo é o fracionamento do áudio, para que seja feita a análise do espectro que será gerado por meio da análise espectral. Após a análise será feita a extração dos picos que então será usada na criação do vetor de notas para detecção dos acordes.

Este software poderá ser utilizado futuramente para um sistema de reconhecimento de acordes musicais mais complexos que seja capaz de distinguir outros instrumentos.

O capítulo 2 apresenta alguns conceitos e ferramentas necessárias para o desenvolvimento desse trabalho, além de uma introdução breve sobre os conceitos da música, notas e acordes. Os trabalhos utilizados como referências são abordados no capítulo 3. A metodologia e os materiais utilizados são discutidos no capítulo 4, onde foi apresentado o projeto detalhadamente. No capítulo 5 são abordados os resultados que foram obtidos após aplicado os passos descritos na metodologia. E finalmente, no capítulo 6 é apresentado as conclusões.

Fundamentação Teórica

2.1. Introdução a Música

Para melhor compreensão dos processos computacionais que serão abordados nesse trabalho, é necessário o conhecimento básico sobre psicoacústica. A psicoacústica estuda a relação entre as sensações auditivas e o fenômeno físico da vibração do ar, cuja relação trouxe conceitos que inspiraram a criação de vários algoritmos na área de áudio.

O som é medido fisicamente por:

- **Altura:** que é determinada pela frequência das vibrações, isto é, da sua velocidade (quanto maior for a velocidade da vibração mais agudo será o som).
- **Duração:** sendo a extensão de um som que é determinada pelo tempo de emissão das vibrações.
- **Intensidade:** que representa a amplitude das vibrações que é o grau do volume sonoro.
- **Timbre:** que é a combinação das vibrações, o timbre é a “cor” do som de cada instrumento.

A mais importante das características do som é a altura, pois é determinada pela sua frequência fundamental. O ouvido humano pode reconhecer frequências de 20 Hz à 20K Hz. Sons fora desse intervalo não são percebidos pela maioria das pessoas porque não possuem energia suficiente para vibrar o tímpano, ou porque a frequência é tão alta que o tímpano não consegue perceber.

Cada frequência corresponde a uma nota musical Dó, Ré, Mi, Fá, Sol, Lá e Si, formam as notas que dobram de valor a cada intervalo formando assim as oitavas. Por exemplo, a nota Lá pertencente à quarta oitava que possui frequência de 440Hz. No próximo intervalo a frequência é dobrada e pertence à quinta oitava passando a ter 880 Hz (KOSTKA;

PAYNE, 2004).

A Tabela 2.1 apresenta as notas e suas respectivas cifras, a familiarização com as cifras é importante para compreensão deste trabalho.

Tabela 2.1. Tabela de cifras

Nota	Cifra
Dó	C
Ré	D
Mi	E
Fá	F
Sol	G
Lá	A
Si	B

Os intervalos que compõem uma oitava são chamados de semitons, portanto existem 12 semitons iguais em uma oitava. Dois semitons juntos formam um tom. Um semitom separa uma nota de um acidente musical ou de outra nota. Os acidentes musicais, bemol (b), sustenido (\sharp), alteram o valor da nota em um semitom para baixo ou para cima, respectivamente. As notas em bemóis ou sustenidos podem ter o mesmo som, porém recebem dois nomes diferentes. Sendo assim, $C\sharp$ é igual a $D\flat$ e $D\sharp$ é igual a $E\flat$. Observe a disposição das notas na Tabela 2.2:

Tabela 2.2. Tabela de notas e seus acidentes musicais

-	$C\sharp$	-	$D\sharp$	-	-	$F\sharp$	-	$G\sharp$	-	$A\sharp$	-
C	-	D	-	E	F	-	G	-	A	-	B
-	$D\flat$	-	$E\flat$	-	-	$G\flat$	-	$A\flat$	-	$B\flat$	-

É possível construir uma tabela com as oitavas destinadas ao conceito musical. A Tabela 2.3 apresenta o valor da frequência para uma nota ou acidente musical. As oitavas estão representadas pelas colunas e os intervalos pelas linhas.

Tabela 2.3. Tabela de frequências sonoras em Hz

Notas		1	2	3	4	5	6	7	8
01	C	32,70	65,41	130,82	261,63	523,25	1046,50	2093,00	4186,00
02	C♯	34,65	69,30	136,60	277,20	554,37	1108,73	2217,46	4434,92
03	D	35,71	73,42	146,83	293,66	567,33	1174,66	2349,32	4698,64
04	D♯	36,70	77,78	155,57	311,13	622,25	1544,51	2469,01	4938,02
05	E	20,06	41,20	82,41	164,81	329,63	659,25	1318,51	2637,02
06	F	21,83	43,65	87,31	174,61	349,23	689,45	1396,92	2793,83
07	F♯	23,12	46,25	92,50	184,99	369,99	739,99	1479,98	2959,95
08	G	24,49	48,99	97,99	195,99	391,99	783,99	1567,89	3135,96
09	G♯	25,95	51,91	103,02	207,65	415,30	830,60	1661,22	3322,44
10	A	27,50	55,00	110,00	220,00	440,00	880,00	1760,00	3520,00
11	A♯	29,13	58,27	116,54	233,10	466,16	932,33	1864,65	3729,31
12	B	30,87	61,74	123,48	246,94	493,88	986,76	1975,53	3951,10

2.1.1. Escala Musical

Segundo ALVES (1997), a escala musical é uma sucessão de todas as notas indo de um tom ou semitom até completar uma oitava, por exemplo, de Dó até a oitava nota Dó ou de Si a Si, iniciando e finalizando pela mesma nota.

Como há doze notas, existe também doze acordes e doze escalas tanto maiores como menores. Três ou mais notas tocadas juntas formam um acorde. Os acordes são formados por intervalos harmônicos dando um sentido na harmonia fundamental. Os acordes menores e maiores são chamados de tríade, pois precisam do agrupamento de três notas separadas por intervalos de terças para compor a escala. O acorde no violão é representado pelas notas que formam a repetição devido a quantidade de cordas que vibram ao mesmo tempo, ser maior que as necessárias para a formação dos acordes. Logo, os acordes menores e maiores do violão são formados pela tríade e pela repetição de alguma destas notas (ROADS, 1996).

Os intervalos com as notas simbolizadas são apresentados na Tabela 2.4, onde temos: **(T)** tônica, **(m)** menor, **(M)** maior e **(+)** aumentada. Com base nesta Tabela 2.4 é possível realizar a formação de acordes **maiores**, **menores**, **aumentados** e **diminutos**.

Tabela 2.4. Tabela de formação de Tríade

T	2 ^a m	2 ^a M	3 ^a m	3 ^a M	4 ^a j	5 ^a m	5 ^a j	+5 ^a	6 ^a M	7 ^a m	7 ^a M
C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B
D	D♯	E	F	F♯	G	G♯	A	A♯	B	C	C♯
E	F	F♯	G	G♯	A	A♯	B	C	C♯	D	D♯
F	F♯	G	G♯	A	A♯	B	C	C♯	D	D♯	E
G	G♯	A	A♯	B	C	C♯	D	D♯	E	F	F♯
A	A♯	B	C	C♯	D	D♯	E	F	F♯	G	G♯
B	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯

Para formar os acordes fundamentais é necessário as notas do primeiro, terceiro e quinto grau da escala. Os acordes são formados a partir da variação da formação básica das tríades.

Acorde Maior

O **Acorde Maior** é formado pela sua nota fundamental, **tônica**, **terça maior** e **quinta justa**, sendo apresentado na Tabela 2.5.

Tabela 2.5. Tabela Tríade Maior

Acorde	T	3 ^a M	5 ^a j
C	C	E	G
D	D	F♯	A
E	E	G♯	B
F	F	A	C
G	G	B	D
A	A	C♯	E
B	B	D♯	F♯

Acorde Menor

O **Acorde Menor** é formado pela sua nota fundamental, **tônica**, **terça menor** e **quinta justa**, sendo apresentado na Tabela 2.6.

Tabela 2.6. Tabela Tríade Menor

Acorde	T	3 ^a m	5 ^a j
Cm	C	E \flat	G
Dm	D	F	A
Em	E	G	B
Fm	F	A \flat	C
Gm	G	B \flat	D
Am	A	C	E
Bm	B	D	G \flat

Acorde Aumentado

O **Acorde Aumentado** é formado pela sua nota fundamental, **tônica**, **terça maior** e **quinta aumentada**, sendo apresentado na Tabela 2.7.

Tabela 2.7. Tabela Tríade Aumentada

Acorde	T	3 ^a M	+5 ^a
C \sharp	C	E	G \sharp
D \sharp	D	F \sharp	A \sharp
E \sharp	E	G \sharp	C
F \sharp	F	A \sharp	C \sharp
G \sharp	G	B	D \sharp
A \sharp	A	C \sharp	F
B \sharp	B	D \sharp	G

Acorde Diminuto

O **Acorde Diminuto** é formado pela sua nota fundamental, **tônica**, **terça menor** e **quinta menor**, sendo apresentado na Tabela 2.8.

Tabela 2.8. Tabela Tríade Diminuta

Acorde	T	3 ^a m	5 ^a m
C $^{\circ}$	C	E \flat	G \flat
D $^{\circ}$	D	F	A \flat
E $^{\circ}$	E	G	B \flat
F $^{\circ}$	F	A \flat	B
G $^{\circ}$	G	B \flat	D \flat
A $^{\circ}$	A	C	E \flat
B $^{\circ}$	B	D	F

2.2. Frequências

Segunda Lacerda (1961) cada período completo por uma unidade de tempo em uma vibração, é chamado de frequência e a sua percepção é conhecida por tom. Quanto maior for a frequência de um som, mais agudo ele será e, quando mais baixa for a frequência, mais grave o som será. Para descrever a frequência, é utilizada a unidade Hertz (Hz), onde 1 Hz corresponde a um ciclo de vibração por segundo.

O som pode ser representado por uma soma de diversas ondas individuais chamadas de componentes de Fourier e cada uma corresponde a uma determinada frequência múltipla da componente inicial. Essas componentes formam uma série conhecida como série harmônica onde o harmônico de ordem zero (F_0) é chamado de frequência natural ou fundamental (LACERDA, 1961).

De acordo com Haykin e Veen (2001), em uma amostra de som onde a frequência fundamental não esteja presente por algum motivo, o valor da diferença absoluta de dois harmônicos adjuntos detectados nessa amostra, determina a frequência fundamental.

2.3. Transformada de Fourier de Curto Tempo

O matemático francês Harris (1978) provou matematicamente que qualquer forma de onda, independente de sua origem, é o somatório de ondas senoidais de diferentes frequências, amplitudes e fases. Ele mostrou que se a forma de onda se repete periodicamente, então as componentes senoidais são restritas a valores múltiplos da frequência da forma da onda. Assim, uma das formas para realizar a análise do sinal no domínio do tempo e da frequência é utilizando a divisão do sinal em pequenos segmentos para podermos determinar o espectro desses segmentos por meio da Transformada Discreta de Fourier (GRAY; GOODMAN, 2012).

Uma das suas representações está descrita na equação 2.1, que é conhecida como Transformada de Fourier de Curto Tempo (do inglês, *Short Time Fourier Transform*, STFT) o onde a seleção dos segmentos no tempo é realizada a partir de uma janela deslizante, representada por W na equação 2.1.

$$STFT(t, \omega) = \int [x(\tau) W(\tau - t)] \epsilon^{-j\omega\tau} d\tau \quad (2.1)$$

O espectrograma é a definição de uma representação tempo por frequência do sinal, que é formado por meio do quadrado do módulo da Transformada de Fourier de Curto Tempo (STFT). Por meio do espectrograma, o sinal representado originalmente em uma dimensão passa a ser representado em duas dimensões: tempo, t , e frequência, ω . O espectro é visualizado como uma imagem, onde a intensidade representa a energia e os eixos x e y representam o tempo e a frequência. A Figura 2.1 mostra o espectro aplicado a um nota D4.

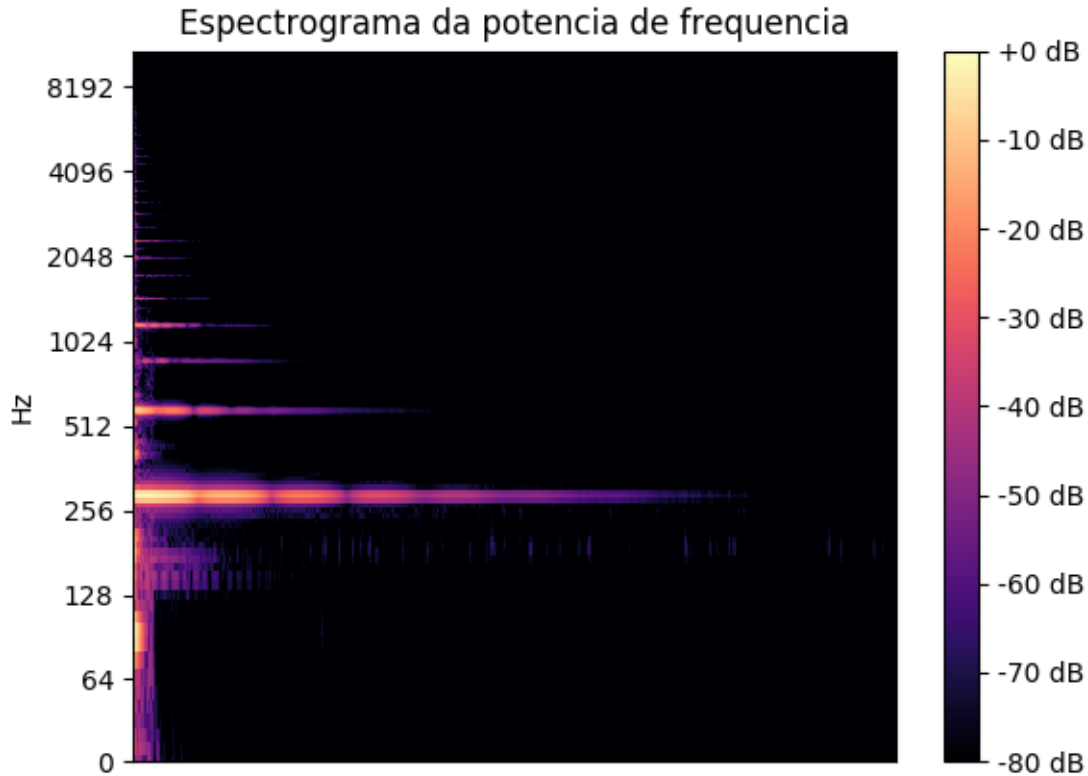


Figura 2.1. Espectrograma de uma nota D4

Como o espectro referente a uma localização τ corresponde a transformada de Fourier de um sinal segmentado pela janela W em τ , conclui-se que o tamanho da janela tem um papel fundamental na resolução em tempo e frequência da análise realizada.

Outra forma de realizar a análise tempo-frequência pode ser obtida decompondo-se adaptativamente o sinal em partes com distribuição tempo-frequência bem conhecidas.

2.4. Constante-Q de Transformação

A Constante-Q de Transformação (CQT) segundo Brown (1991), é uma representação tempo-frequência que, diferente da STFT, apresenta um espectro de frequências com fator de seletividade (também conhecido como fator Q) constante. Devido a isto, as componentes de frequência estão espaçadas geometricamente, ou seja, uma componente f_k é dada por:

$$f_k = f_{k-1} \left(\frac{1}{Q} + 1 \right) = f_{min} \left(\frac{1}{Q} + 1 \right)^k \quad (2.2)$$

onde f_{min} é o menor componente de frequência analisado pela CQT e Q é o fator de seletividade. Desenvolvendo a equação 2.2, temos:

$$f_k - f_{k-1} = \frac{1}{Q} + 1 \quad (2.3)$$

O espaçamento entre componentes de frequência desta transformada é dado pela razão entre a frequência anterior e o fator de seletividade. A relação entre o comprimento da janela e o espaçamento é expresso pela equação 2.4:

$$\frac{f_s}{N_\omega} = \frac{f_{k-1}}{Q} \Rightarrow N_k = \frac{f_s}{f_k} Q \quad (2.4)$$

O comprimento da janela expressado por N_k é dependente de k . Se analisarmos o comportamento da janela, vemos que ele decresce conforme as frequências aumentam. Isso se justifica pelo fato de o senoide de baixa frequência necessitar de mais amostras para completar um ciclo que uma senoide de alta frequência.

Em Brown (1991) são apresentados os parâmetros analisados entre as duas representações de tempo e frequência na Tabela 2.9.

Tabela 2.9. Tabela de comparação de parâmetros entre STFT CQT

	STFT	CQT
Componentes Freqüenciais	$k f_s / N_\omega$ (Linear)	$f_{min} (1/Q + 1)^k$ (Geométrico)
Espaço Freqüencial	f_s / N_ω (Constante)	f_{k-1} / Q (Variável)
Largura da janela	N_ω (Constante)	$Q f_s / f_k$ (Variável)
Freqüência digital	$2k\pi / N_\omega$	$2Q\pi / N_\omega$

Os sinais de natureza musical apresentam espaçamento entre as componentes de frequência geométrico, assim como a CQT é construída. Esse comportamento qualifica a representação como opção ideal para analisar esse tipo de sinal. Por meio da equação 2.2 e com $Q = 17$, teremos $f_k \approx f_{k-1} 2^{\frac{1}{12}}$, isto é, as componentes freqüenciais aproximadamente separadas de um semitom, que é a separação entre as notas na escala musical. Para uma resolução ainda melhor, utiliza-se $Q = 34$ e portanto tem-se $f_k \approx f_{k-1} 2^{\frac{1}{24}}$, ou seja, um espaçamento de um quarto de tom (BROWN, 1991).

Estado da Arte

3.1. Sistema Automático de Transcrição Melódica

Segundo Mitre e Queiroz (2005) o Sistema Automático de Transcrição Melódica tem como entrada um sinal de áudio monofônico e produz uma transcrição de seu conteúdo melódico. O sistema foi projetado como um conjunto de módulos independentes e desacoplados, sendo eles: (1) Sinal de Áudio, (2) Análise Espectral, (3) Detecção de Parciais, (4) Estimção Precisa de Parciais, (5) Triagem de Parciais, (6) Estimção de F0, (7) Produção de Notas e (8) Transcrição Melódica.

A abordagem escolhida por este trabalho utiliza a estimção de frequência fundamental, baseada na hipótese de que o contorno de F0 é tipicamente suficiente para determinar as notas presentes em um sinal de áudio (MITRE; QUEIROZ, 2005).

3.2. Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music.

A proposta de classificação de acordes foi dada por Fujishima (1999), nesse trabalho foi feito uso de vetores croma, que compõem uma técnica para a análise de sinais independentemente da oitava. Essa técnica serviu de base para vários trabalhos posteriores, como os de Stark et al. (2009) e de Oudre et al. (2009).

Por meio deste trabalho, tornou-se comum a utilização de vetores croma, também chamados de *Pitch Class Profiles (PCP)*, como ferramentas para a classificação de acordes. Um vetor croma é uma representação do espectro de um sinal, que é extraído através da Transformada Rápida de Fourier (FFT).

A ideia é criar um vetor de doze posições, cada um relacionado a uma nota da

escala cromática e referente a um acorde da tabela de acordes. Para construir o vetor, primeiramente é relacionado as frequências de todo o espectro às suas notas correspondentes na escala cromática em temperamento igual, independentemente da oitava .

3.3. Sistema Especialista Para Reconhecimento De Acordes Musicais Em Tempo Real Para Violão Elétrico Utilizando Técnicas De DSP

O estudo de Ferreira (2006) busca desvendar a complexidade da interação entre a física, a música e o computador, e apresenta um software capaz de realizar o reconhecimento de acordes musicais maiores e menores do violão elétrico em tempo real. Para atingir os resultados desejados, foram utilizadas a FFT e outras técnicas de Processamento Digital de Sinais (DSP), além de sistemas especialistas para detecção e classificação de acordes maiores e menores.

A aplicação foi desenvolvida e testada utilizando o ambiente `MatLab`, o qual atingiu o resultado esperado, podendo adquirir aprimoramentos para atingir requisitos comerciais. As reflexões e o desenvolvimento deste trabalho levaram a conclusão de que é realmente possível identificar acordes maiores e menores do violão elétrico em tempo real utilizando-se técnicas de DSP.

3.4. Considerações Finais

Por meio do estudo realizado com estes trabalhos, pode-se extrair diferentes abordagens e métodos utilizados para se chegar a uma possível realização da detecção de acordes, sendo assim, foram analisadas as formas de desenvolver este sistema, baseando-se nos pontos positivos e negativos estudos, para então ser desenvolvida a fundamentação e metodologia base deste TCC.

Método Proposto

4.1. Materiais

Para a execução deste trabalho foi utilizado uma máquina com um processador Intel core i5, 8 GB de memória RAM e placa de vídeo Nvidia GeForce MX150 de 2 GB, utilizando um sistema Linux mint em sua versão 18.1.

O projeto foi desenvolvido utilizando a linguagem de programação Python e para auxiliar no desenvolvimento foram utilizadas as seguintes bibliotecas:

- LibROSA
- NumPy
- Matplotlib

4.2. Metodologia

Para a realização dos testes foi necessário a colaboração de músicos, em sua maioria parentes e amigos da autora deste trabalho, que encaminharam os áudios contendo suas cifras para a criação do gabarito que serviu como base de comparação dos resultados na fase de testes.

Entre a obtenção do áudio até a apresentação dos resultados houve uma série de etapas, o diagrama de blocos da Figura 4.1 mostra o processo do fluxograma da metodologia.

Neste processo é necessário uma amostra de áudio para fazer a entrada no sistema, fracionando o sinal que foi convertido para o domínio da frequência, resultando no espectro de potência do sinal do áudio, que será utilizado para a criação do vetor contendo o envelope de força.

Para a realização da escolha dos parâmetros foi preservado os grandes picos, obtendo assim a posição estimada do começo dos **onsets** para segmentação. Este por sua vez será utilizado para extração dos valores do vetor de notas, que foi usado posteriormente como entrada para a classificação dos acordes referente ao áudio de entrada.

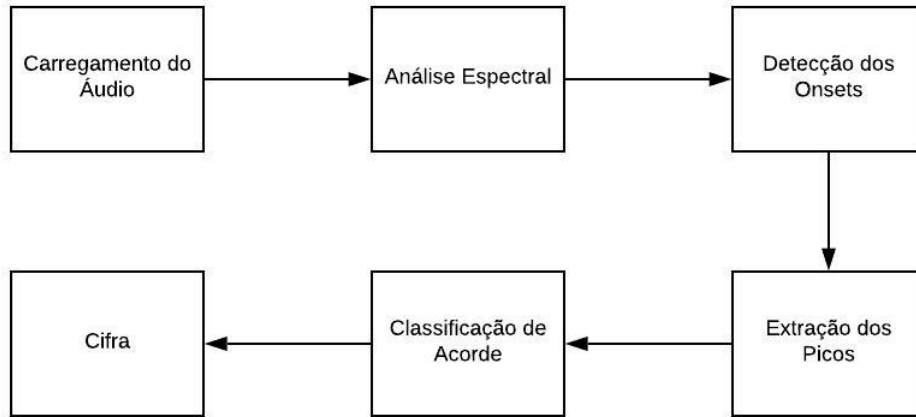


Figura 4.1. Fluxograma de etapas da metodologia.

4.3. Áudio

O Sistema de Detecção faz a leitura das amostras de áudio, o arquivo é lido por inteiro e armazenado em uma variável juntamente com sua taxa de amostragem na memória, onde sofre alguns ajustes: primeiro são removidas as amplitudes consideradas como ruído, em seguida, o áudio precisa ser fracionado e repetido periodicamente. Esse fracionamento do áudio é necessário para realizar uma análise espectral dessa frequência, e por fim, é realizada a normalização deste sinal.

4.4. Análise Espectral

Com base no áudio carregado é feita a análise do sinal no domínio do tempo e da frequência, utilizando a divisão do sinal em pequenos segmentos para podermos determinar o espectro desses segmentos por meio de uma Transformada. Para este trabalho foram utilizados os espectros de tempo-frequência obtidos por meio da **STFT** e **CQT**.

A partir do espectro obtido, separamos as janelas de frequência em torno da intensidade da frequência fundamental (a chamada F_0).

A quantidade de janelas é um ponto importante: quanto menor o número de janelas obtidas, ou seja, quanto mais larga for a janela na extração, maior o risco de englobamento de frequências intermediárias, que não interessam para o estudo, porém, um número muito

grande de janelas acarretaria em inclusão de pequenas desafinações ou ruídos, pois a largura da janela seria muito estreita.

Para controle das frequências intermediarias foi utilizado a detecção de **onsets**.

4.5. Detecção dos Onsets

De acordo com Bregman (1994), **onsets** comuns através da frequência são importantes sinais de agrupamento que são utilizados pelo sistema auditivo humano, para organizar e integrar sons originados da mesma fonte.

O processador localiza os eventos de início de nota escolhendo picos em um envelope de força de início, a partir de um espectrograma pré-computado ao longo do eixo de frequência, esses parâmetros influenciam a saída do processador de início. O início é detectado medindo-se o aumento de energia em dB.

Segundo Klapuri (1999), é possível limitar a força de **onsets** individuais a um limite superior, que por padrão é definido como uma lista de todas as posições estimadas de **onsets** detectados, criando assim um vetor de amplitudes.

4.6. Extração de Picos

Após a obtenção do vetor de amplitudes, com as faixas de frequência, é realizado a detecção de todos os picos de frequências fundamentais válidas deste vetor.

Para atingir o resultado esperado foi elaborada uma técnica de extração de picos, que percorre todo o vetor de amplitude em busca do maior valor em cada região no espectro.

Inicialmente, o primeiro valor do vetor de amplitude é considerado um pico candidato. Caso o elemento seguinte possua um valor de amplitude maior que o atual, o mesmo ocupará a vaga de pico candidato, substituindo o anterior. Esse processo se repete até ser encontrado um valor de amplitude menor que o pico candidato atual.

No exemplo da Figura 4.2, nota-se que o áudio é dividido de tempo em tempo e que a cada intervalo de tempo é referenciado a uma nota. O maior pico desse intervalo é selecionado como um valor do vetor.

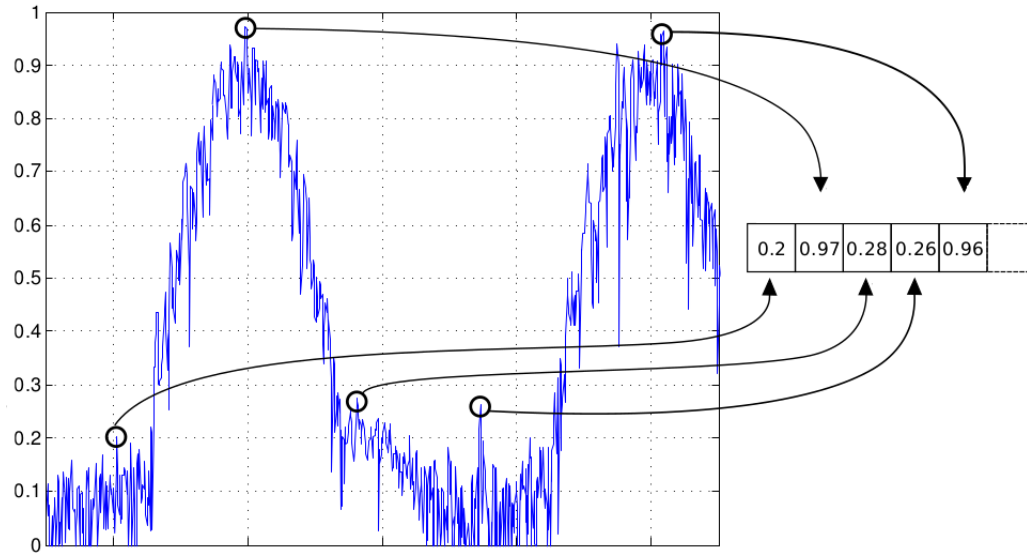


Figura 4.2. Extração de picos do espectro

4.7. Classificador de Acordes

Com a obtenção do vetor de notas, podemos seguir para a etapa de classificação de acordes. O modelo do acorde pode ser definido utilizando a teoria musical e análise espectral. Os modelos mais simples representam as notas fundamentais. A nota Dó é representada pelo modelo da Figura 4.3(a). Conseqüentemente ao juntar as notas para formar o acorde de C, teríamos o modelo da Figura 4.3(b).

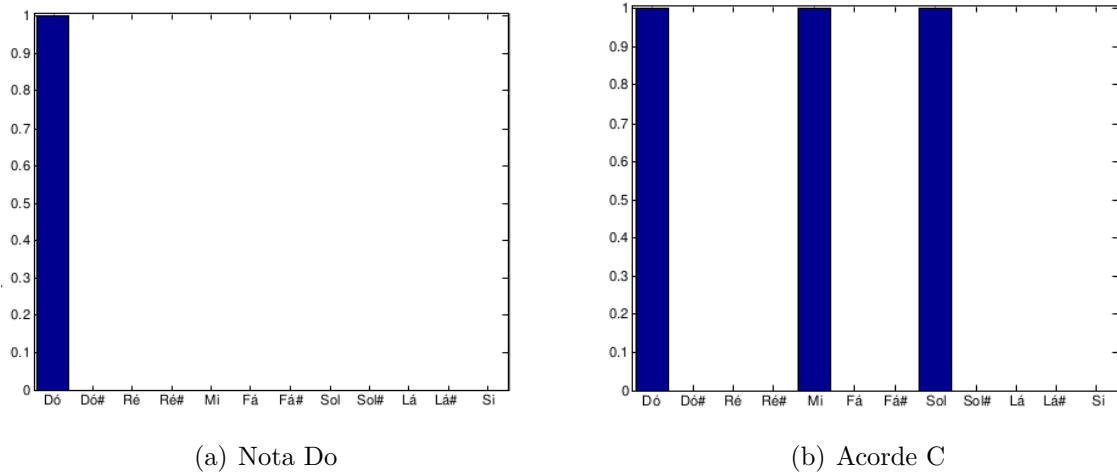


Figura 4.3. Modelo Simples

Esse modelo foi utilizado no trabalho de Fujishima (1999). Outro modelo ocorre ao considerarmos outros harmônicos presentes na nota. A Figura 4.4(a) mostra os harmônicos de uma nota Dó tocada no violão.

Note a forte presença do terceiro harmônico (Sol). Assim o acorde de C tocado no mesmo violão é representado pela Figura 4.4(b),no qual está presente a escolha das notas para formação do acorde.

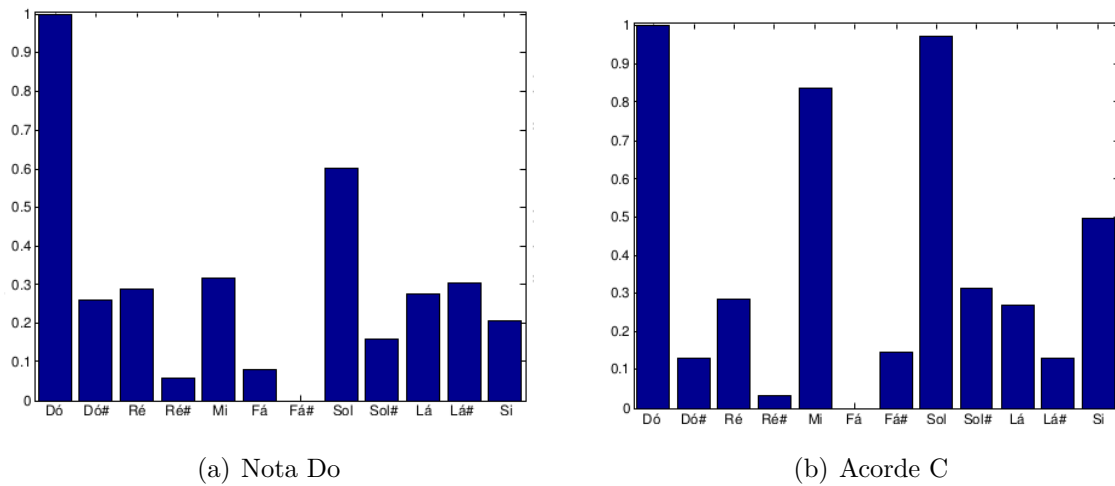


Figura 4.4. Modelo Harmônico

Podemos observar que foi utilizada a Tabela 2.5 para a construção dos acordes dos dois modelos apresentados, sendo assim, indiferente dos harmônicos que compõem o vetor, o acorde será o mesmo.

Para este trabalho foi considerado o modelo simples utilizando apenas as notas fundamentais.

4.8. Considerações Finais

Para construção do algoritmo descrito na metodologia foi utilizada a biblioteca LibROSA, está foi de suma importância para auxiliar na detecção de *onsets*. E para uma melhor análise, os espectros de potência do sinal do áudio foram mostrados em tela por meio das funcionalidades da biblioteca Matplotlib.

O sistema desenvolvido desde trabalho está disponível no GitHub: <https://github.com/AnaFrozza/TCC>.

Depois da conclusão do algoritmo partiu-se para os testes e averiguação dos resultados.

Experimentos e Resultados

5.1. Experimentos

Para a realização dos experimentos foi realizado um processo de obtenção das amostras de áudio para validar o sistema. Essas amostras foram obtidas por meio de colaboradores, em sua maioria familiares e amigos da autora deste trabalho.

Foi estabelecido aos colaboradores que encaminhassem a amostra de áudio juntamente com a cifra da música tocada. Não foi definido quais músicas seriam gravadas, porém foi sugerido aos colaboradores que as amostras contivessem diferentes dificuldades (mudança na quantidade de acordes, contendo acordes de nível básico e avançado). Essas amostras coletadas serviram como base para a comparação dos resultados dos experimentos.

É válido destacar que as amostras foram gravadas em um ambiente não controlado. Coube a cada colaborador a escolha do ambiente onde ocorreria a gravação, bem como o conteúdo que seria gravado, com isso houve ambientes que interferiram na qualidade do áudio, causando em algumas amostras ruídos como interferência de animais, carros ou conversa no fundo do áudio.

Também houve uma quantidade significativa de amostras que foram perdidas, devido a quantidade de ruídos existentes ser tão grande que o sistema não conseguia identificar que o arquivo se tratava de um áudio. Essas amostras em sua maioria pertenciam a um mesmo colaborador, sendo assim, podemos afirmar que a maneira como as amostras são armazenadas e encaminhadas também podem interferir no resultado.

Os primeiros testes foram realizados variando os parâmetros de entrada para encontrar um conjunto de valores mais adequado ao Sistema de Detecção.

Os parâmetros foram testados de acordo com cada etapa do processo proposto na metodologia.

Na Etapa de Carregamento do Sinal foram testados os seguintes parâmetros: Taxa de amostragem e número de bits por oitava.

Na Etapa de Análise Espectral foram alterados os parâmetros das duas transformadas usadas para geração dos espectros de tempo-frequência; e na Etapa de Detecção de **Onsets** foi avaliado o parâmetro correspondente ao tamanho do salto e os espectros gerados por meio da **STFT** e **CQT**.

Após a configuração dos parâmetros foi observado que os valores atribuídos ao **delta** interferem na quantidade de janelas existentes no espectro que estava sendo analisado.

Sendo assim, optou-se pela criação de dois grupos de experimentos, cada um contendo cinco variações do **delta**, afim de melhor avaliarmos a divergência dos resultados.

- **Grupo 01:** Foram avaliados os resultados extraídos das amostras de áudio considerando a análise do espectro gerado a partir da **STFT**, produzindo 5 experimentos, cada qual possuindo um **delta** alternado em: 0.02, 0.05, 0.07, 0.1 e 0.15.
- **Grupo 02:** Foram avaliados os resultados extraídos das amostras de áudio considerando a análise do espectro gerado a partir da **CQT**, produzindo 5 experimentos, cada qual possuindo um **delta** alternado em: 0.02, 0.05, 0.07, 0.1 e 0.15.

Para armazenar os experimentos foi gerado um arquivo txt para cada experimento do Grupo 01 e Grupo 02, com a finalidade de comparar o resultado de cada retorno das amostras com o gabarito obtido através dos colaboradores. Em cada arquivo txt foi armazenado:

- Grupo
- Delta
- Lista das amostras de áudio:
 - Nome da amostra de áudio;
 - Número de janelas obtidas;
 - Saída da Cifra.

5.2. Resultados

Para obtenção dos resultados foi aplicado o método apresentado para cada uma das amostras de áudio.

Após a execução do algoritmo de detecção para o Grupo 1 e Grupo 2, os resultados dos experimentos e o gabarito das amostras foram armazenados na forma de um vetor, para então comparar os resultados utilizando um algoritmo de similaridade. Esse algoritmo resulta na porcentagem de similaridade entre os vetores.

Foi então colocado em uma tabela a porcentagem de acerto para cada amostra testada nos experimentos.

A Tabela 5.1 apresenta as medianas das porcentagens de acertos atingidos nos Grupos 01 e 02.

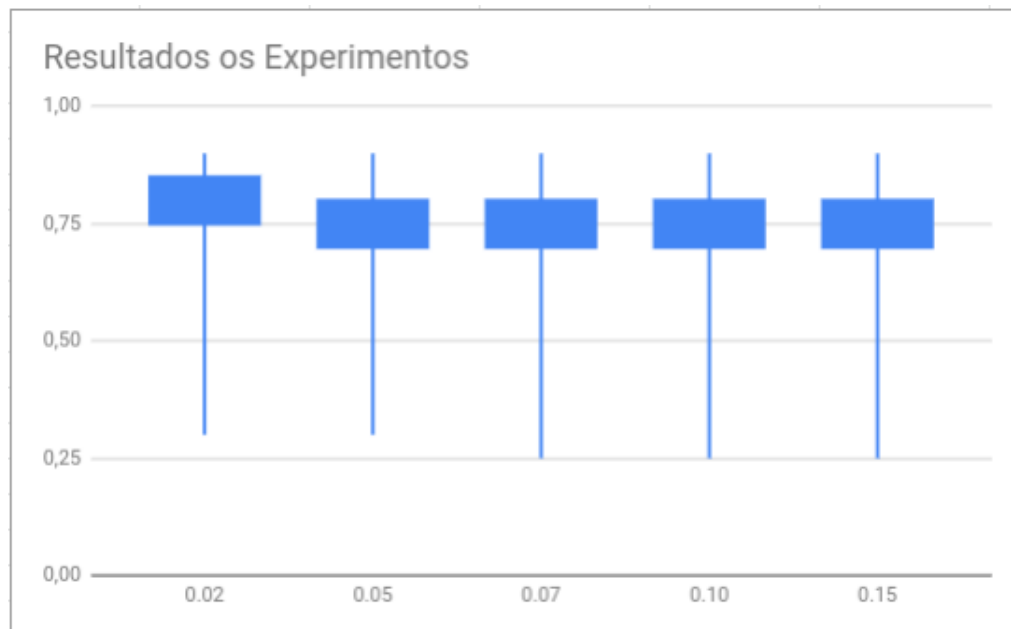
Tabela 5.1. Porcentagem em Mediana de acerto dos grupos

Delta	0,02	0,05	0,07	0,10	0,15
Grupo 01	80%	80%	80%	80%	80%
Grupo 02	80%	80%	80%	80%	80%

Após a análise dos resultados, podemos observar que o Grupo 01 e o Grupo 02 tiveram as mesmas porcentagens de acertos para cada experimento testado. Isso ocorre devido aos parâmetros utilizados para obtenção dos espectros estarem bem ajustados, causando assim o melhor posicionamento possível da janela do espectrograma para ambos os Grupos.

Foram gerados e analisados 880 testes com as 88 amostras válidas coletadas. O Gráfico de Caixas da Figura 5.1 apresenta os resultados obtidos por meio dos testes realizados.

Na Figura 5.1 caixa representa um grupo de testes por experimento. No eixo **x** estão os valores utilizados para o **delta** em cada experimento e o eixo **y** corresponde aos valores de acerto indo de 0,00 (0%) à 1,00 (100 %).

**Figura 5.1.** Resultados do Grupo 1.

O Gráfico da Figura 5.2 apresenta alguns dos resultados das amostras testadas.

Na Figura 5.2 cada linha representa um teste e as cores representam os experimentos realizados em cada Grupo. No eixo **x** estão distribuídas as amostras utilizadas nos testes e o eixo **y** corresponde aos valores de acerto indo de 0 à 100 %.

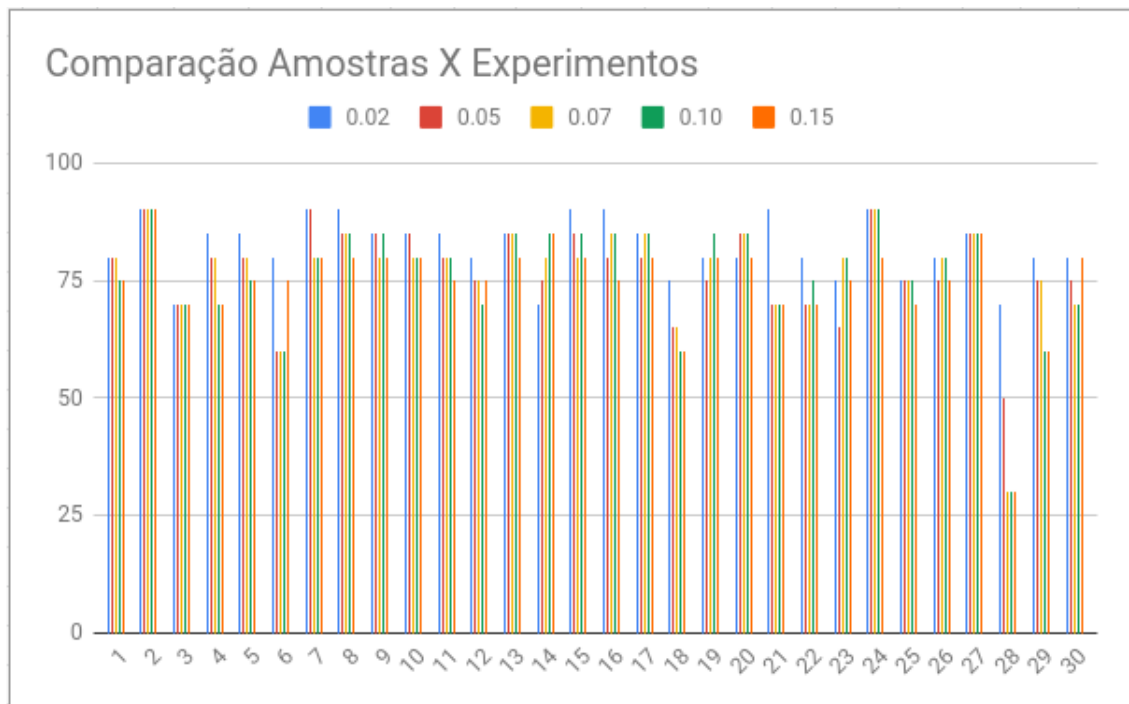


Figura 5.2. Experimentos X Amostras.

5.3. Considerações Finais

Com base na Figura 5.1 e na Tabela 5.1 podemos observar que para os experimentos testados neste trabalho, quanto menor o δ maior o número de janelas geradas para realização da análise, como demonstrado na Figura 5.3. Assim, conclui-se que o melhor resultado das amostras foi gerado a partir do menor δ avaliado.

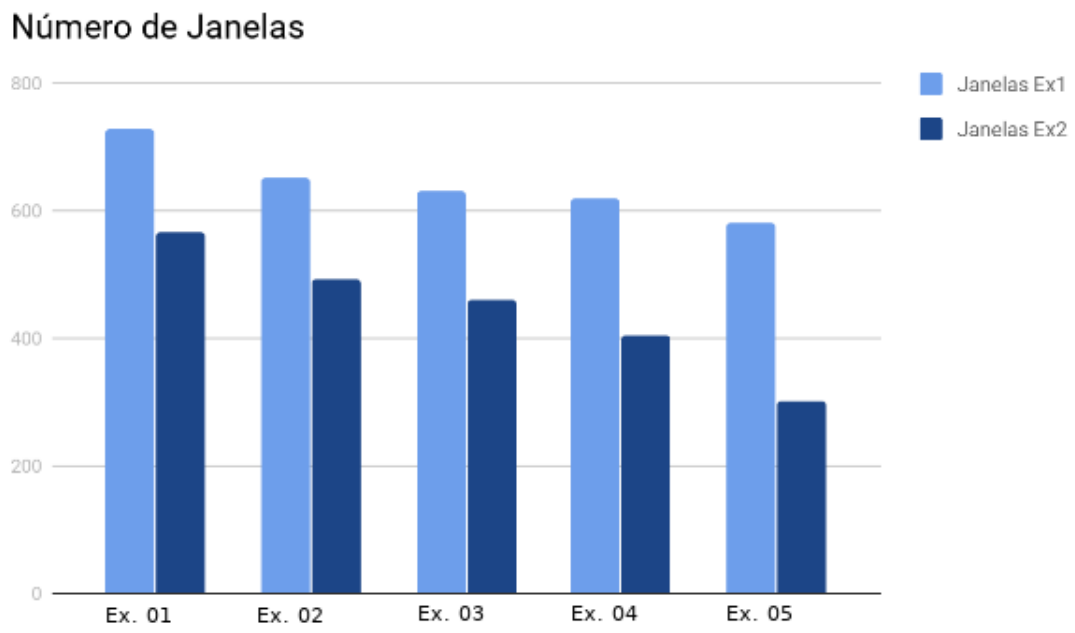


Figura 5.3. Números de Janelas por Teste.

Conclusões

O desafio de integrar técnicas de processamento digital de sinal às teorias de tecnologia da informação e física da música foi uma das motivações desde trabalho, que teve como resultado o desenvolvimento de um Sistema de Detecção de acordes por meio de informações espectrais.

A primeira parte do trabalho dedicou-se a explicar de forma sucinta a teoria musical necessária para a compreensão do funcionamento do Sistema. Posteriormente, foram explicados os conceitos da física da música e a aplicação da Transformada de Fourier de Curto Tempo (STFT) e da Constante-Q de Transformação (CQT).

Depois do conceito teórico estar fundamentado explicou-se o funcionamento do Sistema e os detalhes da sua implementação. Em seguida, todos os testes realizados foram descritos e analisados para obter resultados verdadeiros, ainda que delimitados apenas ao violão.

Os resultados acadêmicos foram muito satisfatórios, porém observou-se que a fidelidade do Sistema de detecção não realizou em sua maioria os 100% dos acertos. Esta discrepância de resultados não ocorreu devido ao processo de reconhecimento, visto que em testes com amostras gravadas em ambiente controlado e bem armazenadas o sistema teve acertos próximos de 100%, sendo assim o formato habitual com que os músicos executam os acordes e o ambiente que foram gravados alteram os resultados do sistema, logo é conceitual que em um ambiente com acústica controlada alcançaríamos melhores resultados.

Este trabalho não considerou utilizar um ambiente controlado pois o objetivo do sistema é ser utilizado por músicos em qualquer ambiente, não se fazendo necessário um ambiente próprio para as gravações.

Este trabalho, poderá servir como referência para outros que envolvam tópicos aqui relacionados, tais como: percepção computacional da música, processamento digital de sinais, entre outros.

Como sugestão, a partir do conhecimento adquirido com a evolução desta monografia, é possível ampliar o Sistema de Detecção de acordes para outros instrumentos. É possível também ampliá-lo para realizar o reconhecimento de outros tipos de acordes.

Este trabalho atendeu os fins acadêmicos, e a evolução deste Sistema pode levar à sua aceitação na área comercial.

Diante dos resultados obtidos concluiu-se que é possível realizar o reconhecimento de acordes musicais por meio de de um Sistema Especialista totalmente desenvolvido em Python.

Referências

- ALVES, LUCIANO. *Escalas para improvisação*. São Paulo: Irmãos Vitale, 1997.
- BREGMAN, Albert S. *Auditory scene analysis: The perceptual organization of sound*. London: MIT press, 1994.
- BROWN, Judith C. Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, ASA, v. 89, n. 1, p. 425–434, 1991.
- FERREIRA, Aurélio Buarque de Holanda. Novo dicionário da língua portuguesa. In: *Novo dicionário da língua portuguesa*. Rio de Janeiro: Nova Fronteira, 1986.
- FERREIRA, Sandro Alex de Souza. *Sistema Especialista Para Reconhecimento De Acordes Musicais Em Tempo Real Para Violão Elétrico Utilizando Técnicas de DSP*. Tese (Doutorado) — Dissertação de mestrado em engenharia elétrica, Universidade Federal da Bahia, Bahia, 2006.
- FUJISHIMA, Takuya. Realtime chord recognition of musical sound: a system using common lisp music. In: *ICMC*. EUA: Journal of the Autonomic Nervous System, 1999. p. 464–467.
- GRAY, Robert M; GOODMAN, Joseph. Fourier transforms: an introduction for engineers. In: . EUA: Springer Science & Business Media, 2012. v. 322, p. 53–113.
- HARRIS, Fredric J. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, IEEE, v. 66, n. 1, p. 51–83, 1978.
- HAYKIN, Simon S; VEEN, Barry Van. *Sinais e sistemas*. Porto Alegre RS: Bookman, 2001.
- KLAPURI, Anssi. Sound onset detection by applying psychoacoustic knowledge. In: IEEE. *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*. Phoenix, AZ, USA, USA: IEEE, 1999. v. 6, p. 3089–3092.
- KOSTKA, S.M.; PAYNE, D. *Tonal Harmony, with an Introduction to Twentieth-century Music*. 4.ed. ed. Universidade de Michigan: McGraw-Hill, 2004. 309–315 p. ISBN 9780072852608.
- LACERDA, Osvaldo. Teoria elementar da música. *São Paulo: Ricord Brasileira*, 1961.
- MITRE, Adriano; QUEIROZ, Marcelo. Um sistema automático de transcrição melódica. *Simpósio Brasileiro de Computação Musical*, 2005.
- OUDRE, Laurent; GRENIER, Yves; FÉVOTTE, Cédric. Template-based chord recognition: Influence of the chord types. In: *ISMIR*. Paris, France: Institut TELECOM, 2009. p. 153–158.
- ROADS, Curtis. The computer music tutorial. In: . Cambridge, MA, USA: MIT Press, 1996. p. 15–300. ISBN 0262680823.

STARK, Adam M; DAVIES, Matthew EP; PLUMBLEY, Mark D. Real-time beat-synchronous analysis of musical audio. In: *Proceedings of the 12th Int. Conference on Digital Audio Effects, Como, Italy*. London, United Kingdom: Centre for Digital Music, 2009. p. 299–304.