



## Detecção de Acordes Musicais por meio de Informações Espectrais

Ana Carolina Frozza, Diego Bertolini, Juliano Henrique Foleiss, Rodrigo Hübner

*Departamento Acadêmico de Computação (DACOM)*

*Universidade Tecnológica Federal do Paraná (UTFPR)*

*Campo Mourão, 2019*

*annafrozza@gmail.com, {diegobertolini, julianofoleiss, rodrigohubner}@utfpr.edu.br*

**Resumo**—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 sinais de áudio convertidos 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 experimentais foram satisfatórios, a fidelidade do sistema de detecção alcançou para o melhor cenário 90% de acerto gerando uma média de 80% de acerto para todas as amostras avaliadas. 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 no qual foram gravados. Este trabalho, com seus resultados práticos, poderá servir como referência para outros que envolvam tópicos relacionados, como por exemplo, percepção computacional da música ou processamento digital de sinais.

**Keywords**-Detecção de acordes; Análise espectral; Processamento de sinal de áudio;

### I. INTRODUÇÃO

De acordo com [4] 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 informações 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, foi adotado neste trabalho a cifra como notação musical, pois considera basicamente a frequência das notas já a análise dos diferentes tipos de instrumentos musicais e ritmos foram ignorados.

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 trabalho 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. Em suma, o principal objetivo deste trabalho é 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 serão usados 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 e que seja capaz de distinguir outros instrumentos.

Este artigo está apresentado como segue. A Seção II apresenta a fundamentação teórica da música e sua relação com a programação. A Seção III apresenta as transformações no domínio do sinal musical. A Seção IV apresenta os materiais e métodos. A Seção V apresenta os experimentos realizados. A Seção VI apresenta os resultados e discussão. E, por fim, a Seção VII apresenta as conclusões e trabalhos futuros.

## II. FUNDAMENTAÇÃO DA 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 [13].

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 para este trabalho é a altura, pois é determinada pela sua frequência fundamental. O ouvido humano pode reconhecer frequências de 20Hz à 20KHz. 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 possui frequência de 440 Hz. No próximo intervalo a frequência é dobrada e pertence à quinta oitava passando a ter 880 Hz [10].

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

Tabela I  
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 (♯), 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♯ é igual a Db e D♯ é igual a Eb. Observe a disposição das notas na Tabela II:

Tabela II  
TABELA DE NOTAS E SEUS ACIDENTES MUSICAIS

-	C♯	-	D♯	-	-	F♯	-	G♯	-	A♯	-
C	-	D	-	E	F	-	G	-	A	-	B
-	Db	-	Eb	-	-	Gb	-	Ab	-	Bb	-

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

Tabela III  
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	587,33	1174,66	2349,32	4698,64
04 D♯	36,70	77,78	155,57	311,13	622,25	1244,51	2489,01	4978,02
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	698,45	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	1567,99	3135,96
09 G♯	25,95	51,91	103,82	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	987,76	1975,53	3951,10

### A. Escala Musical

Segundo [1], 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 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 [12].

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

Tabela IV  
TABELA DE FORMAÇÃO DE TRÍADE

T	2ªm	2ªM	3ªm	3ªM	4ªj	5ªm	5ªj	+5ª	6ªM	7ªm	7ªM
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.

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

Tabela V  
TABELA TRÍADE MAIOR

Acorde	T	3ªM	5ªj
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#

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

Tabela VI  
TABELA TRÍADE MENOR

Acorde	T	3ªm	5ªj
Cm	C	E <sub>b</sub>	G
Dm	D	F	A
Em	E	G	B
Fm	F	A <sub>b</sub>	C
Gm	G	B <sub>b</sub>	D
Am	A	C	E
Bm	B	D	G <sub>b</sub>

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

Tabela VII  
TABELA TRÍADE AUMENTADA

Acorde	T	3ªM	+5ª
C#	C	E	G#
D#	D	F#	A#
E#	E	G#	C
F#	F	A#	C#
G#	G	B	D#
A#	A	C#	F
B#	B	D#	G

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

### B. Frequências

Segundo [11] 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), na qual 1 Hz corresponde a um ciclo de vibração por segundo.

Tabela VIII  
TABELA TRÍADE DIMINUTA

Acorde	T	3ªm	5ªm
C <sup>o</sup>	C	E <sub>b</sub>	G <sub>b</sub>
D <sup>o</sup>	D	F	A <sub>b</sub>
E <sup>o</sup>	E	G	B <sub>b</sub>
F <sup>o</sup>	F	A <sub>b</sub>	B
G <sup>o</sup>	G	B <sub>b</sub>	D <sub>b</sub>
A <sup>o</sup>	A	C	E <sub>b</sub>
B <sup>o</sup>	B	D	F

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 [11].

De acordo com [8], 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.

## III. TRANSFORMAÇÕES NO DOMÍNIO DO SINAL MUSICAL

### A. Transformada de Fourier de Tempo Curto

O matemático francês [7] provou 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 [6].

Uma das suas representações está descrita na equação 1, que é conhecida como Transformada de Fourier de Tempo Curto (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 1.

$$STFT(t, \omega) = \int [x(\tau) W(\tau - t)] \epsilon^{-j\omega\tau} d\tau \quad (1)$$

O espectrograma é a definição de uma representação tempo por frequência do sinal, que é formado por meio do quadrado do módulo da STFT. Por meio do espectrograma, o sinal representado originalmente em uma dimensão passa a ser representado em duas dimensões: tempo,  $t$ , e frequência,  $\omega$ . 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 1 mostra o espectro aplicado a um nota D4.

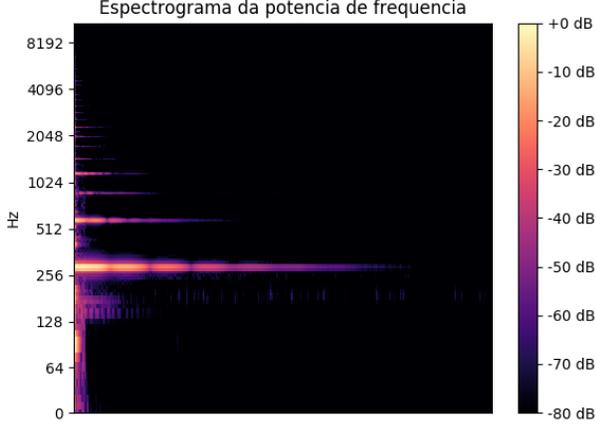


Figura 1. Espectrograma de uma nota D4

Como o espectro referente a uma localização  $\tau$  corresponde a transformada de Fourier de um sinal segmentado pela janela  $W$  em  $\tau$ , 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.

### B. Constante-Q de Transformação

A Constante-Q de Transformação (CQT) segundo [3], 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:

$$f_k = f_{k-1} \left( \frac{1}{Q} + 1 \right) = f_{min} \left( \frac{1}{Q} + 1 \right)^k \quad (2)$$

onde  $f_{min}$  é o menor componente de frequência analisado pela CQT e  $Q$  é o fator de seletividade. Desenvolvendo a equação 2, temos:

$$f_k - f_{k-1} = \frac{1}{Q} + 1 \quad (3)$$

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 4:

$$\frac{f_s}{N_\omega} = \frac{f_{k-1}}{Q} \Rightarrow N_k = \frac{f_s}{f_k} Q \quad (4)$$

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 [3] são apresentados os parâmetros analisados entre as duas representações de tempo e frequência na Tabela IX.

Tabela IX  
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	$f_s / N_\omega$ (Constante)	$f_{k-1} / Q$ (Variável)
Largura da janela	$N_\omega$ (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 e com  $Q = 17$ , teremos  $f_k \approx f_{k-1} 2^{\frac{1}{17}}$ , isto é, as componentes frequenciais 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  $f_k \approx f_{k-1} 2^{\frac{1}{34}}$ , ou seja, um espaçamento de um quarto de tom [3].

## IV. METODOLOGIA

Para a realização dos experimentos foi necessário a colaboração de músicos, que encaminharam os áudios contendo suas cifras para a criação do gabarito que serviu como base de comparação dos resultados na fase de testes.

Para o desenvolvimento do código foi empregada a linguagem de programação Python, por fornecer varias bibliotecas que contribuíram para este trabalho, em sua maioria a biblioteca LibROSA que foi usada para análise de música e áudio, também foram utilizadas as bibliotecas para auxílio de operações numéricas, vetoriais e plotagens (NumPy, SciPy e Matplotlib, respectivamente).

Entre a obtenção do áudio até a apresentação dos resultados houve uma série de etapas, o diagrama de blocos da Figura 2 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.

### A. 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

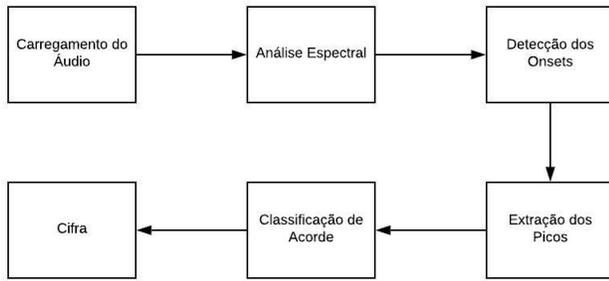


Figura 2. Fluxograma de etapas da metodologia.

consideradas como ruído, em seguida, o áudio precisa ser fracionado para realizar uma análise espectral dessa frequência, e por fim, é realizada a normalização deste sinal.

### B. Análise Espectral

Com base no áudio carregado é realizado 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 Transformada de Fourier de Curto Tempo (do inglês, Short Time Fourier Transform, STFT) e Constante-Q de Transformação (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 incluirmos 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 intermediárias foi utilizado a detecção de onsets.

### C. Detecção dos Onsets

De acordo com [2], 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 [9], 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.

### D. 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 3, 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.

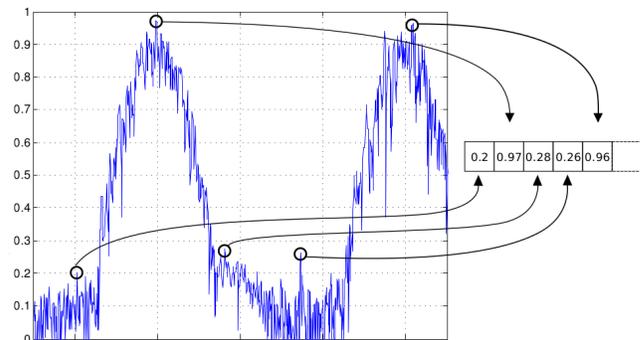


Figura 3. Extração de picos do espectro

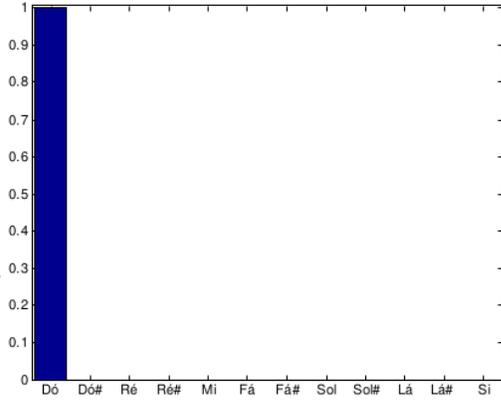
### E. Classificador de Acorde

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 (a). Consequentemente ao juntar as notas para formar o acorde de C, teríamos o modelo da Figura 4 (b).

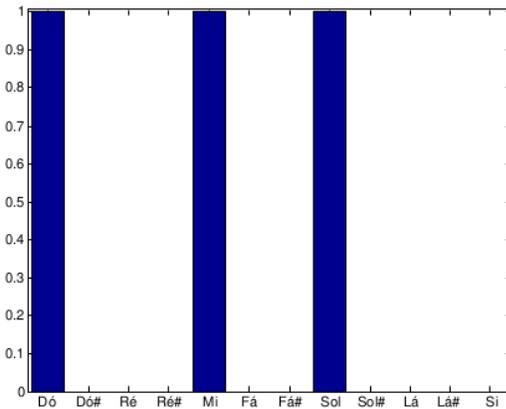
Esse modelo foi utilizado no trabalho de [5]. Outro modelo ocorre ao considerarmos outros harmônicos presentes na nota. A Figura 5 (a) mostra os harmônicos de uma nota Dó tocada no violão.

Note a forte presença do terceiro harmônico (Sol). Assim o acorde de C tocado no mesmo violão é representado pela Figura 5 (b), no qual está presente a escolha das notas para formação do acorde.

Podemos observar que foi utilizada a Tabela V para a construção dos acordes dos dois modelos apresentados,



(a) Nota Do



(b) Acorde C

Figura 4. Modelo Simples

sendo assim, indiferente dos harmônicos que compõem o vetor, o acorde será o mesmo.

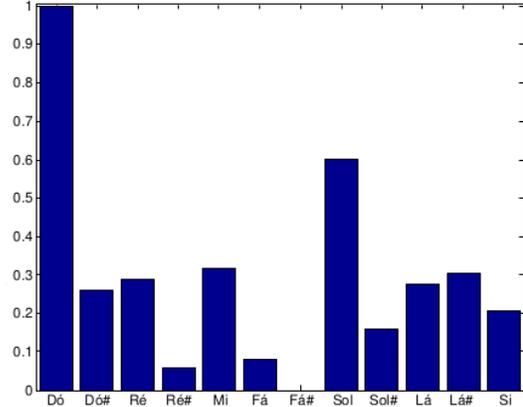
Para este trabalho foi considerado o modelo simples utilizando apenas as notas fundamentais.

### V. EXPERIMENTOS

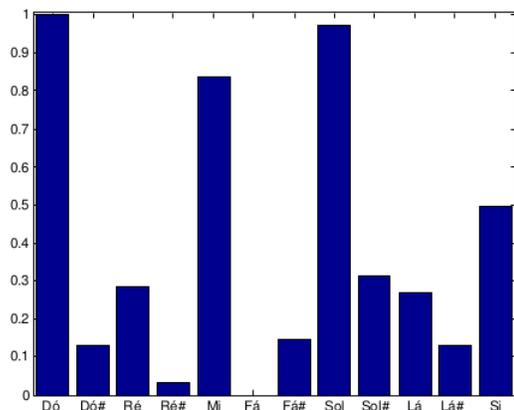
Para a realização dos experimentos foi realizado um processo de obtenção das amostras de áudio para validar o sistema. Essas amostras foram obtidas por meio de colaboradores.

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



(a) Nota Do



(b) Acorde C

Figura 5. Modelo Harmônico

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 avaliados de acordo com cada etapa do processo proposto na metodologia.

Na Etapa de Carregamento do Sinal foram avaliados 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 analisado o 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  $\delta$ , a fim 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 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 foi armazenado:

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

## VI. 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 avaliada nos experimentos.

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

Tabela X  
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 6 apresenta os resultados obtidos por meio dos testes realizados.

Na Figura 6 cada 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%).

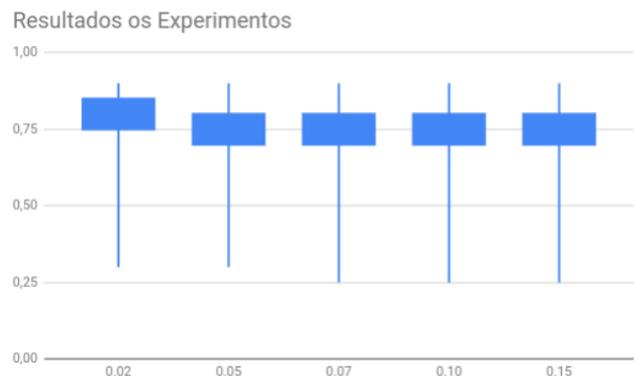


Figura 6. Resultados do Grupo 1.

O Gráfico da Figura 7 apresenta alguns dos resultados das amostras testadas.

Na Figura 7 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%.

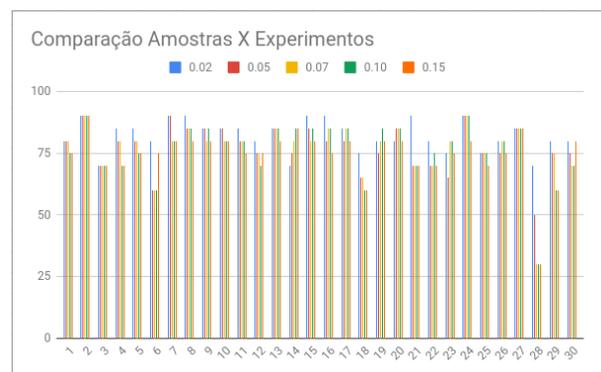


Figura 7. Experimentos X Amostras.

### A. Considerações Finais

Com base na Figura 6 e na Tabela X podemos observar que para os experimentos realizados neste trabalho, quanto

menor o  $\delta$  maior o número de janelas geradas para realização da análise, como demonstrado na Figura 8. Assim, conclui-se que o melhor resultado das amostras foi gerado a partir do menor  $\delta$  avaliado.

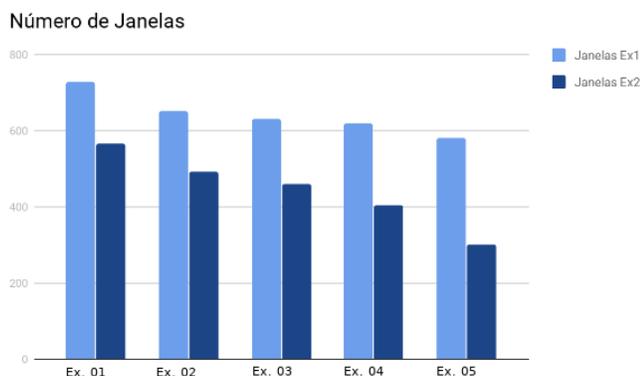


Figura 8. Número de Janelas por Teste.

## VII. CONCLUSÃO

O desafio de integrar técnicas de processamento digital de sinal às teorias de tecnologia da informação relacionados à 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 foram 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. Assim, podemos concluir 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 diferentes ambientes, 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.

## REFERÊNCIAS

- [1] ALVES, LUCIANO. *Escalas para improvisação*. São Paulo: Irmãos Vitale, 1997.
- [2] BREGMAN, Albert S. *Auditory scene analysis: The perceptual organization of sound*. London: MIT press, 1994.
- [3] BROWN, Judith C. Calculation of a constant  $q$  spectral transform. *The Journal of the Acoustical Society of America*, SA, v. 89, n. 1, p. 425–434, 1991.
- [4] FERREIRA, Aurélio Buarque de Holanda, Novo dicionário da língua portuguesa, *In: Novo dicionário da língua portuguesa*. Rio de Janeiro: Nova Fronteira, 1986.
- [5] FUJISHIMA, Takuya. ealtime chord recognition of musical sound: a system using commonlisp music *In: ICMC*. EUA: Journal of the Autonomic Nervous System, 1999. p. 464–467.
- [6] GRAY, Robert M; GOODMAN, Joseph. Fourier transforms: an introduction for engineers. *In: .* EUA: Springer Science & Business Media, 2012. v. 322, p. 53–113.
- [7] HARRIS, Fredric J. On the use of windows for harmonic analysis with the discrete fouriertransform. *Proceedings of the IEEE*, IEEE, v. 66, n. 1, p. 51–83, 1978.
- [8] HAYKIN, Simon S; VEEN, Barry Van. *Sinais e sistemas*. Porto Alegre RS: Bookman, 2001.
- [9] KLAPURI, Anssi. ound 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.
- [10] 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.
- [11] LACERDA, Osvaldo. *Teoria elementar da música São Paulo: Ricord Brasileira*, 1961.
- [12] ROADS, Curtis. The computer music tutorial. In: . Cambridge, MA, USA: MIT Press, 1996.p. 15–300. ISBN 0262680823.
- [13] Zwicker, E., and H. Fastl. *"Psychoacoustics Facts and Models Springer Heiderberg."* Technische Universität München. München: Germany (1999).