

FIA 2020/22

XII CONGRESSO/CONGRESO IBEROAMERICANO DE ACÚSTICA

XXIX ENCONTRO DA SOCIEDADE BRASILEIRA DE ACÚSTICA - SOBRAC

Florianópolis, SC, Brasil

Utilizando análise espectro-temporal para reconhecimento de tons e acordes musicais

Moraes, R. B.¹

¹Programa de Pós-graduação em Ensino de Biociências e Saúde, Instituto Oswaldo Cruz, Rio de Janeiro, Brasil, rbarmoraes@gmail.com

Resumo

Neste trabalho foi implementado um algoritmo que analisa sinais de áudio de modo a realizar a extração automática de informações sobre a tonalidade e da detecção da sequência de acordes em canções musicais gravadas, através do processamento e análise de informações espectro-temporais com a utilização da Transformada de Fourier de Tempo Curto (STFT). Objetivou-se elaborar um método computacional eficaz de transcrição em notação musical das gravações musicais com intuito de aplicabilidade educacional – tanto profissional quanto amadora, para tal na saída do algoritmo era gerado uma partitura ou tablatura com o tom e os acordes encontrados. Os resultados da acurácia do algoritmo no reconhecimento dos acordes mostraram-se satisfatórios.

Palavras-chave: Processamento de sinais de áudio, Acústica musical, Acústica computacional.

PACS: 43.60.-c, 43.75.Zz.

Using spectral and temporal analysis for recognition of musical tones and chords

Abstract

In this present work, an algorithm that analyzes audio signals in order to perform the automatic extraction of musical information such as tonality and the detection of the chord sequence in recorded musical songs was implemented, through the processing and analysis of spectral and temporal information using the Short Time Fourier Transform (STFT). The objective was to develop an effective computational method of transcribing musical recordings into musical notation with the aim of educational applicability, and for that, at the output of the algorithm, a tablature was generated with the tone and chords found. The results of the algorithm's accuracy in recognizing chords were satisfactory.

Keywords: Audio signal processing, Musical acoustics, Computational acoustics.

1. INTRODUÇÃO

O reconhecimento de acordes – também conhecido como detecção ou estimativa de acordes – é o processo automatizado de atribuição de acordes musicais a segmentos de uma peça musical qualquer, em parte dela ou em sua totalidade. Sua aplicabilidade se encontra de diversas formas, tanto como uma ferramenta para transcrição de partituras (ou tablaturas) musicais, quanto como uma maneira automatizada de preencher bancos de

dados de metadados de música que podem ser pesquisados na internet por usuários diversos, assim como modo computacional de recuperação de informações musicais¹ tais como identificação e classificação de gêneros musicais, consultas por semelhança, dentre outras.

¹ Em tradução livre do inglês, *music information retrieval (MIR)*.



A transcrição automática de acordes geralmente consiste em duas etapas: uma extração de características que captura a informação musical ao longo do tempo e um processo de reconhecimento que produz rótulos de acordes a partir das características calculadas. Os recursos usados para reconhecimento de acordes podem diferir de um método para outro, mas na maioria dos casos são variantes dos perfis de classe tonal de 12-dimensões, conceito introduzido inicialmente por Fujishima [1]. Cada componente representa a energia espectral de um semitom na escala cromática, independentemente da oitava. O cálculo por ser tanto baseado na chamada Transformada Q Constante (CQT), quanto na Transformada de Fourier de Tempo Curto (STFT²), e pode ser realizado em quadros de comprimento fixo ou em quadros de comprimento variável – neste trabalho será utilizada a STFT. A sucessão desses vetores cromáticos ao longo do tempo é chamada de “cromagrama”.

Desta forma, o que convencionou-se chamar unicamente de “reconhecimento de acordes”, na realidade passa necessariamente por três estágios sob qualquer perspectiva: a detecção, a representação e a transcrição. A detecção de acordes lida com todos os tipos de extração computacional de informações de sinais de áudio, onde são extraídas as informações acerca dos harmônicos que ocorrem ao longo do tempo de uma peça musical. A representação de acordes é uma escrita musical compacta que fornece informações sobre o conteúdo harmônico e a estrutura de uma música. A transcrição automática de acordes encontra muitas aplicações no campo da recuperação de informações musicais, como identificação de músicas, consulta por semelhança ou análise de estrutura.

Existem diferentes razões pelas quais o reconhecimento de acordes não seja um assunto trivial. Dentro do campo da psico-acústica, o cérebro humano processa sons de maneiras complexas que podem ser difíceis de modelar em algoritmos computacionais. Por exemplo, o tom percebido de uma única nota

pode não ser representado no espectro de frequência, mas resultar da interferência ou batimento de várias outras frequências parciais (Loy, [2]), além de diferentes indivíduos poderem ouvir/escutar diferentes acordes na mesma amostra de áudio (Burgoyne et al., [3]).

Nesse sentido, “treinar” e avaliar algoritmos de reconhecimento automático de acordes segue como um campo em desenvolvimento na última década. Enquanto algoritmos mais novos e mais sofisticados têm um desempenho cada vez melhor na tarefa, os algoritmos propostos também começam a se tornar consideravelmente complexos e computacionalmente mais dispendiosos. Mesmo com algoritmos aprimorados, a qualidade da estimativa dos algoritmos de melhor desempenho ainda depende muito do tipo de entrada (Benson, [4]).

O presente trabalho se concentra principalmente na etapa do processo de transcrição de acordes, que consiste em atribuir um rótulo de acorde a cada quadro cromatográfico. O algoritmo implementado parte do método proposto por Fujishima [1] utiliza uma espécie de “dicionário de acordes” compostos por *templates* de 12-dimensões constituídos por 1 (para os cromas presentes no acorde) e 0 (para os demais cromas), e incluiu também identificação de tonalidade, com o número de tipos de acordes trabalhados sendo 5 no total: maiores, menores, aumentados, diminutos e com sétima. A transcrição de acordes é então realizada retendo o acorde com maior produto escalar entre os modelos de acordes e os quadros de cromagrama, baseando-se na ideia intuitiva de que para um dado vetor de croma de 12 dimensões, as amplitudes dos cromas presentes no acorde em execução devem ser maiores que aqueles que não aparecem. Ao introduzir modelos de acordes para diferentes tipos de acordes, o presente em um quadro deve, portanto, ser aquele cujo modelo é o mais próximo do *template*.

Este artigo tem como objetivo demonstrar como o processamento de sinais de áudio, a acústica computacional e a teoria musical podem ser aplicados em conjunto para resolver

² Do inglês, *Short-Time Fourier Transform*.

o problema de reconhecimento automático de acordes. Será utilizada a análise espectral de Fourier de modo a extrair informações importantes de frequência de sinais de áudio (musicais), como a frequência se relaciona com o tom na música e, finalmente, como inferir acordes a partir de informações de notas em um cromagrama. Para esse fim, foi elaborado, implementado e avaliado, um algoritmo de detecção de acordes simples, porém razoavelmente preciso. O foco do algoritmo foi propor um design simples que não se baseia em técnicas sofisticadas de *machine learning*, redes neurais ou processamento estatístico e, portanto, não requer uma fase de “aprendizado de máquina” separada – o algoritmo possui um conjunto de parâmetros que podem ser ajustados para um conjunto comparativo, ao serem fornecidos valores padrão que geram bons resultados na maioria dos casos – deixando a qualidade de detecção tão boa quanto possível, mas mantendo a complexidade do algoritmo baixa.

2. TRANSFORMADA DE FOURIER DE TEMPO CURTO (STFT)

Para melhor compreensão dos processos computacionais de processamento de sinais de áudio abordados neste artigo, é necessário retomar os conceitos da análise espectro-temporal e sua aplicação no cálculo de espectrogramas – e que será utilizada no algoritmo proposto.

A expressão “transformada espectro-temporal” se origina do fato de que a dada função transformada não representa mais um sinal em função do tempo, mas codifica como a energia é distribuída no espaço de frequência ao longo de todo o comprimento do sinal. Existem várias formas de transformações espectro-temporais, com diferentes aplicações e propriedades, mas sem dúvidas a mais notável é a Transformada de Fourier. A transformada de Fourier decompõe uma dada função temporal em uma soma de ondas senoidais, cada uma com sua amplitude, fase e frequência específicas. A análise de Fourier e suas aplicações desempenham um papel

fundamental em uma ampla gama de disciplinas, como no processamento de sinais e na compressão digital de dados.

Uma das variações da Transformada de Fourier é a Transformada Discreta de Fourier (DFT) que trabalha com sinais discretos no tempo – que na prática lida melhor com sinais de áudio reais, que possuem um tempo finito e têm amostragens computacionais discretas. E mais do que isso, há o interesse também em saber como o espectro frequencial de um dado sinal se altera ao longo do tempo – como no exemplo de um instrumento musical executando uma sequência de notas musicais ao longo de um dado intervalo de tempo.

Uma maneira de resolver essa limitação é não aplicar a transformação a todo o sinal de uma vez, mas sim calculá-la para curtos períodos de tempo consecutivos, independentemente, por meio de uma janela “deslizante” ao longo do tempo de execução do sinal. Essa abordagem é chamada de Transformada de Fourier de Tempo Curto (STFT).

Matematicamente a STFT pode ser definida como:

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

onde t é o passo do deslocamento da janela “deslizante” e ω é banda de frequência trabalhada e τ denota a variável que caracteriza a distância entre duas janelas consecutivas, x é o sinal de entrada e W é a função da janela de cada “DFT consecutiva” realizada.

Percebe-se portanto que há necessariamente um *tradeoff* entre as diferentes resoluções temporais e frequenciais do sinal amostrado. Janelas maiores de DFT aumentam a densidade de amostragem de frequência, mas, ao mesmo tempo, acarretam uma amostragem mais esparsa no domínio do tempo. Por outro lado, tamanhos de janela menores podem fornecer uma boa precisão temporal, mas podem perder na resolução no domínio da frequência. Na prática, é preciso encontrar um bom equilíbrio entre resolução de tempo e frequência, dependendo da aplicação.



O espectrograma será por definição o quadrado do módulo de amplitude da STFT, visualizado de forma bidimensional, em que o eixo das abscissas (eixo x) corresponde ao tempo do sinal, e o eixo das ordenadas (eixo y) corresponde às frequências dos harmônicos encontrados. Cada ponto da imagem é visualizado com uma cor (como em uma espécie de representação “térmica”) de acordo com a amplitude de um harmônico de uma dada frequência em um dado tempo. A Figura 1 exemplifica um espectrograma de um clarinete executando duas notas musicais em sequência – um E4 seguido de um D5.

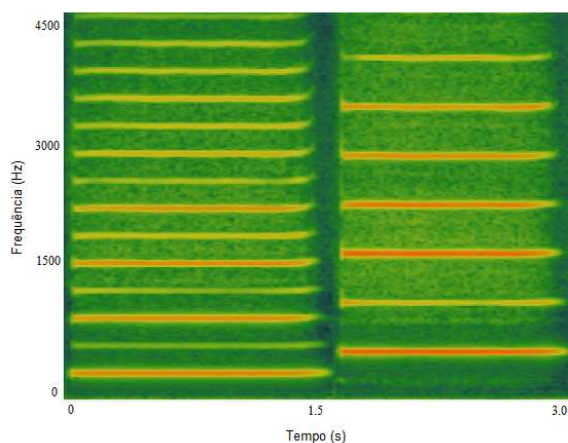


Figura 1: Espectrograma da execução de duas notas musicais em sequência por um clarinete, sendo a primeira um E4 (com fundamental em 328Hz) e a segunda um D5 (com fundamental em 587Hz).

3. FUNDAMENTOS MÚSICAIS

Para se referir às notas musicais será utilizada a notação por cifras, também chamada de “notação anglo-saxônica”, em que as sete primeiras letras do alfabeto (a começar por Lá, denominada de A) são seguidas de um número representativo de sua oitava, por exemplo, C2 (nota Dó, 2ª oitava), G#4 (nota Sol sustenido, 4ª oitava), B5 (nota Si, 5ª oitava).

A teoria dos acordes é melhor descrita em termos de notas e intervalos, mas um algoritmo de estimativa de acordes baseado em STFT necessita funcionar a partir em frequências harmônicas individuais. O mapeamento entre

notas e frequências pode acontecer de diferentes formas, uma delas é um sistema conhecido como “temperamento igual”. Praticamente todas as músicas modernas usam notas temperadas, isto é, a frequência de cada nota e o meio-tem acima dela têm uma proporção fixa. Além de ser mais simples de modelar matematicamente, torna-se uma boa escolha para usar no reconhecimento de acordes de peças musicais modernas. Segundo Howard e Angus [5], a vantagem do temperamento igual é que uma dada peça musical pode ser tocada em diferentes tons sem soar diferente, uma grande desvantagem dos sistemas de afinação anteriores.

Os intervalos que compõem uma oitava são chamados de semitons, existindo portanto 12 semitons iguais em uma oitava temperada. Como existem 12 semitons em uma oitava, e uma oitava caracteriza-se como sendo o dobro da frequência anterior, a taxa fixa será $\sqrt[12]{2} \approx 1,06$. A Tabela 1 apresenta as frequências das notas cromáticas, em temperamento igual, relativas às sete primeiras escalas musicais.

Tabela 1: Frequências das notas cromáticas

Escala Cromática	1 (Hz)	2 (Hz)	3 (Hz)	4 (Hz)	5 (Hz)	6 (Hz)	7 (Hz)
C	32,7	65,4	130,8	261,6	523,2	1.046,5	2.093,0
C#	34,6	69,2	138,5	277,1	554,3	1.108,7	2.217,4
D	36,7	73,4	146,8	293,6	587,3	1.174,6	2.349,3
D#	38,9	77,7	155,5	311,1	622,2	1.244,5	2.489,0
E	41,2	82,4	164,8	329,6	659,2	1.318,5	2.637,0
F	43,6	87,3	174,6	349,2	698,4	1.396,9	2.793,8
F#	46,2	92,4	184,9	369,9	739,9	1.479,9	2.959,9
G	48,9	97,9	195,9	391,9	783,9	1.567,9	3.135,9
G#	51,9	103,8	207,6	415,3	830,6	1.661,2	3.322,4
A	55,0	110,0	220,0	440,0	880,0	1.760,0	3.520,0
A#	58,2	116,5	233,1	466,1	932,3	1.864,6	3.729,3
B	61,7	123,4	246,9	493,8	987,7	1.975,5	3.951,0

A nota que ouvimos quando um dado instrumento está em execução dependerá da frequência da vibração que, por sua vez, é determinada por suas características físicas como o comprimento da corda no caso dos cordofones. A noção de que a corda vibra em apenas uma frequência, chamada de fundamental, é um modelo bastante simplificado, e factualmente existem oscilações em múltiplos inteiros da fundamental chamados harmônicos ou parciais, em que o n-ésimo harmônico terá uma

frequência n -vezes a da fundamental. Em geral, a fundamental é o harmônico de maior intensidade em uma nota (mas não necessariamente), com os harmônicos superiores possuindo intensidades subsequentemente menores.

A distribuição exata dos harmônicos em uma dada nota, no entanto, varia muito entre os instrumentos (e seus tipos) e é, de fato, uma parte substancial do que define o timbre de um som (Benson, [4]). Importante ressaltar que há instrumentos (como no caso dos sinos) cujas parciais podem aparecer como não múltiplos da fundamental, e nesse caso o som que produzem é comumente chamado de inarmônico (Loy, [2]).

De forma simplificada, os acordes são várias notas tocadas ao mesmo tempo pelo mesmo instrumento. Eles podem ser caracterizados pelas notas que fazem parte dele, por exemplo, para o acorde Dó Maior (ou simplesmente C), a sequência C-E-G denotará as notas C, E e G tocadas juntas e deste modo formarem o acorde desejado. Alternativamente, os acordes podem ser definidos por uma nota fundamental e os intervalos (em semitons) desta nota fundamental para todas as outras que compõem o acorde, por exemplo, C:1,4,7 especificando o mesmo acorde C-E-G, com C sendo a nota fundamental, E é o quarto semitom, e G é o sétimo semitom a partir da fundamental.

Um grande número de acordes consiste na formação de três (chamados de tríades) ou quatro notas (chamados de tétrades), embora não haja limite superior para o número de notas. Se a nota fundamental (também chamada de “tônica”) não for a nota mais baixa (em escala frequencial) de um acorde, o acorde é chamado de invertido. Os acordes invertidos geralmente soam semelhantes aos não invertidos, pois são construídos a partir das mesmas notas, apenas em uma ordem diferente.

As tríades mais comuns formam os acordes maiores e menores. O acorde *maior* é formado pela tônica, a terça maior (4 semitons acima), e a quinta justa (7 semitons acima). O acorde *menor*, por sua vez, consiste na tônica, na terça

menor (3 semitons acima), e na quinta justa (7 semitons acima).

Outras tríades que foram objeto do presente algoritmo de reconhecimento foram: o acorde *aumentado*, sendo este formado pela tônica, a terça maior (4 semitons acima) e a quinta aumentada (8 semitons acima); e o acorde *diminuto*, formados pela tônica, a terça menor (3 semitons acima) e a quinta menor (6 semitons acima).

Entre as tétrades, uma das mais comuns é o acorde *com sétima*, formado pela tônica, a terça maior (4 semitons acima), a quinta justa (7 semitons acima) e a sétima menor (10 semitons acima).

A Tabela 2 apresenta os templates das tríades e tétrades utilizadas no algoritmo.

Tabela 2: Templates das tríades e tétrades utilizadas.

Maior				Menor			
Intervalo	I	III	V	Intervalo	I	bIII	V
Semitons	t	4	7	Semitons	t	3	7
Acordes				Acordes			
C	C	E	G	Cm	C	D#	G
C#	C#	F	G#	C#m	C#	E	G#
D	D	F#	A	Dm	D	F	A
D#	D#	G	A#	D#m	D#	F#	A#
E	E	G#	B	Em	E	G	B
F	F	A	C	Fm	F	G#	C
F#	F#	A#	C#	F#m	F#	A	C#
G	G	B	D	Gm	G	A#	D
G#	G#	C	D#	G#m	G#	B	D#
A	A	C#	E	Am	A	C	E
A#	A#	D	F	A#m	A#	C#	F
B	B	D#	F#	Bm	B	D	F#

Aumentado				Diminuto			
Intervalo	I	III	V#	Intervalo	I	bIII	bV
Semitons	t	4	8	Semitons	t	3	6
Acordes				Acordes			
Caum	C	E	G#	C ⁰	C	D#	F#
C#aum	C#	F	A	C# ⁰	C#	E	G
Daum	D	F#	A#	D ⁰	D	F	G#
D#aum	D#	G	B	D# ⁰	D#	F#	A
Eaum	E	G#	C	E ⁰	E	G	A#
Faum	F	A	C#	F ⁰	F	G#	B
F#aum	F#	A#	D	F# ⁰	F#	A	C
Gaum	G	B	D#	G ⁰	G	A#	C#
G#aum	G#	C	E	G# ⁰	G#	B	D
Aaum	A	C#	F	A ⁰	A	C	D#
A#aum	A#	D	F#	A# ⁰	A#	C#	E
Baum	B	D#	G	B ⁰	B	D	F



Com Sétima				
Intervalo	I	III	V	bVII
Semitons	<i>t</i>	4	7	10
Acordes				
C7	C	E	G	A#
C#7	C#	F	G#	B
D7	D	F#	A	C
D#7	D#	G	A#	C#
E7	E	G#	B	D
F7	F	A	C	D#
F#7	F#	A#	C#	E
G7	G	B	D	F
G#7	G#	C	D#	F#
A7	A	C#	E	G
A#7	A#	D	F	G#
B7	B	D#	F#	A

O termo “acorde” também pode se referir a um conceito maior dentro do contexto de uma peça musical, a harmonia. Nesse sentido, um acorde não é mais entendido apenas como um único conjunto de notas tocadas ao mesmo tempo, mas funciona como uma diretriz global sobre quais notas devem ser tocadas por um conjunto de instrumentos, válida para um número específico de tempos ou compassos, até que é trocado por outro acorde, e assim sucessivamente. De acordo com essa diretriz, os instrumentos responsáveis pelas harmonias tocam algumas ou todas as notas que fazem parte do acorde em execução.

4. IMPLEMENTAÇÃO DO ALGORITMO

O projeto do algoritmo e a execução de seus testes foram desenvolvidos e realizados utilizando *MatLab*, rodando em um sistema operacional *Windows 10*. A ideia central do algoritmo projetado está o conceito de “dicionário de acordes” proposto pela primeira vez por Fujishima [1], para derivar probabilidades de acordes a partir da altura e intensidade de harmônicos. Não foram empregadas técnicas de *machine learning*, modelos estatísticos, ou redes neurais artificiais. A vantagem do algoritmo proposto é que não são necessárias fases de aprendizado separadas, sendo dessa forma menos suscetível a falhas, *bugs* e *loops* infinitos, com implementação mais direta e computacionalmente mais barato e eficiente. A desvantagem é que há não oportunidades para otimizar e “treinar” o algoritmo, o que poderia se traduzir em menor precisão de detecção caso

formações de acordes mais complexas venham a ser implementadas. Sendo que o sistema foi projetado para ser usado em cenários que inicialmente o algoritmo lê todos os dados de uma peça musical inteira, para ao final retornar todos os acordes reconhecidos/encontrados, ou seja, não é um algoritmo realizado em tempo real.

O campo de acordes a serem estimados pelo algoritmo desenvolvido está restrito a 60 acordes: 12 maiores, 12 menores, 12 aumentados, 12 diminutos e 12 com sétima, além do pausa/silêncio (denotado como “N.C.”, *no chord*, em inglês). De todo modo, não é uma variedade tão restrita assim, dado que estes são os acordes mais comuns presentes na música popular ocidental moderna (considerando o período do pós Segunda Guerra Mundial até atualmente).

O algoritmo faz a leitura da amostra do sinal de áudio em formato PCM (*Pulse Code Modulation*, ou seja, arquivos em *wav*) em mono não compactados como entrada, e retorna os acordes estimados/reconhecidos e suas durações em uma tablatura. Os dados brutos de entrada são altamente redundantes, considerando que estamos interessados apenas nos acordes e não há como extrair essa informação diretamente dos dados temporais. A ideia é então transformar os dados de entrada em uma série de etapas de representações progressivamente menores, mais adequadas para extrair as informações dos acordes e descartar dados irrelevantes para a aplicação.

A Figura 2 ilustra a arquitetura geral do algoritmo de reconhecimento automático de acordes. A seguir, serão detalhados seus passos em detalhe.

O passo seguinte à leitura do arquivo de áudio é o cálculo do espectrograma. Para detectar acordes no áudio de entrada, é evidente que é essencial determinar quais notas efetivamente são tocadas. O espectrograma já é um bom ponto de partida para esta tarefa, pois nos dá acesso à distribuição de frequência no sinal ao longo do tempo. Para quantificar a presença de uma nota em um intervalo de tempo. Importante frisar que o sinal de entrada deve

ser amostrado em uma taxa de 44.400Hz ou 48.000Hz. A seguir a STFT é calculada, usando janela de Hamming.

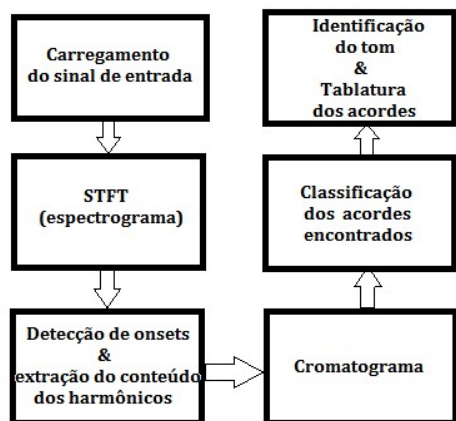


Figura 2: Fluxograma do algoritmo proposto para detecção automática de acordes.

Para mapear as frequências no espectrograma, assume-se que os instrumentos são temperados de acordo com o temperamento igual. Além disso, um rastreamento de ritmo e batida, através da detecção de *onsets* é usado para obter o tempo da música. Em geral, as mudanças de acordes ocorrem no tempo da batida, este é um método eficiente para evitar que janelas de amostragem abranjam dois ou mais mudanças de acordes, perdendo informações importantes no processo. Simultaneamente, varre-se o sinal em busca de picos de amplitudes dos harmônicos, em busca do maior valor em cada região temporal do espectro.

O passo seguinte envolve o cálculo do chamado cromagrama. Assim como os espectrogramas, os cromagramas são gráficos bidimensionais onde uma dimensão representa o tempo, porém enquanto que o eixo-y de um espectrograma descreve o espectro de frequência, no caso de cromagramas, o eixo-y são na realidade vetores de 12-dimensões chamados de vetores cromáticos, propostos pela primeira vez por Fujishima [1], cujos elementos correspondem a 12 semitons.

Cada nota detectada é atribuída a uma dessas classes de semitons (uma espécie de “tom relativo”, onde ignora-se a posição da oitava, por exemplo, G#4 passa a ser somente G#) e os

elementos dos vetores cromáticos podem ser interpretados como indicadores de quão “forte” (em termos conjugados de energia e probabilidade) cada um dos 12 classes de notas cromáticas estão presentes no sinal em um determinado momento.

A Figura 3 exemplifica o cromagrama da execução do exemplo da Figura 1, em que um clarinete executa duas notas musicais – observa-se como no caso do cromagrama é ignorada a posição da oitava relativa, buscando-se apenas a informação do semitom.

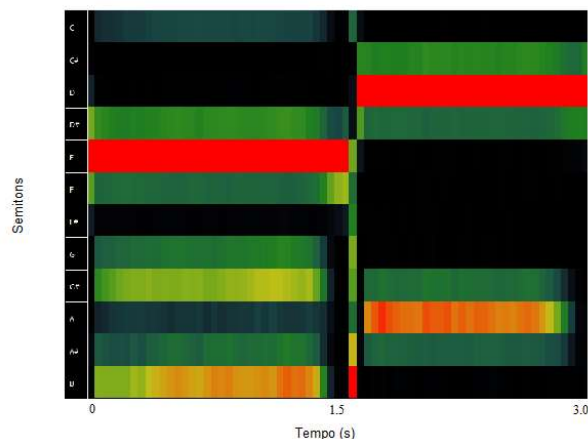


Figura 3: Cromagrama da execução de duas notas musicais em sequência por um clarinete, identificando em vermelho os semitons E e D da exemplo da Figura 1.

Descartar a informação de oitava tomando a média das amplitudes sobre todas as oitavas é usado como uma maneira conveniente de reduzir a dimensionalidade dos dados sem perder características significativas que podem ser úteis para a classificação dos acordes. A posição de oitava absoluta em que um tom é tocado é relativamente sem importância no contexto geral da detecção de qual acorde está sendo executado em um dado momento de uma peça musical.

Tal como os espectrogramas, os cromagramas são melhor visualizados como imagens bidimensionais com o eixo horizontal representando o tempo, o eixo vertical compreendendo aos 12 semitons, sendo os valores dos elementos dos vetores cromáticos codificados por cores, tal qual um gráfico “térmico”.



Na Figura 4, é mostrado o exemplo de um excerto de cerca de 6 segundos da canção “*Eight Day a Week*” dos Beatles, presente em seu álbum *Beatles For Sale*, lançado no Reino Unido em 1964.

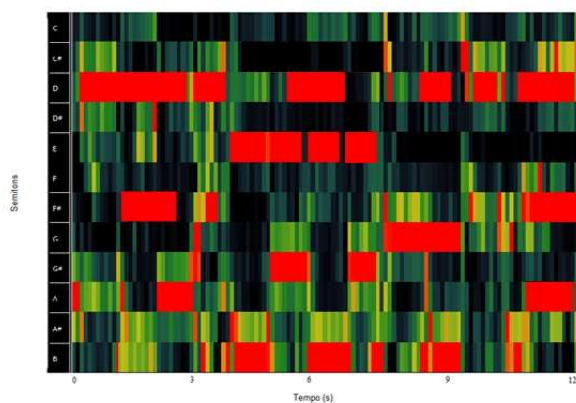


Figura 4: Cromagrama de um excerto de 6 segundos da canção “*Eight Days a Week*” dos Beatles.

Tendo reduzido o sinal de entrada para uma série de vetores cromáticos, o próximo passo é derivar as probabilidades de acordes a partir desses vetores. Primeiro, define-se o conjunto de acordes que o algoritmo pode detectar: todos os 60 acordes maiores, menores, aumentados, diminutos e com sétima, além do “*no chord*” (pausa/silêncio).

Para cada vetor cromático no cromagrama e para cada um desses acordes detectáveis, é computada uma pontuação que expressa a probabilidade (ou possibilidade) do acorde detectado estar correto. A ideia geral é que, ao assumir que um acorde específico é executado, o vetor cromático aponta para uma direção que pode ser capturada em um “dicionário de acordes” que, por sua vez, pode ser usado para reconhecer esse acorde específico em execução. Em outras palavras, a distribuição da energia nos vetores cromáticos sobre seus 12 elementos/semitons é indicativo do acorde que está sendo executado e ao compará-lo com as distribuições típicas de todos os acordes suportados, o acorde correto pode ser estimado.

Os modelos, ou *templates*, também são vetores de 12 dimensões, assim como os vetores cromáticos do sinal, e também se referem aos mesmos 12 semitons. A forma de comparação

entre o vetor do sinal e o vetor do template é através do cálculo do produto escalar entre ambos.

Reunindo todas as informações sobre os cromagramas e o tempo (tanto da batida quanto da duração dos acordes) da peça musical, pode-se então estimar a sequência dos acordes na peça musical.

Ainda sobre o exemplo apresentado do excerto de cerca de 6 segundos da canção “*Eight Day a Week*” dos Beatles, a Figura 5 mostra as linhas em azul identificando o passo da música, o algoritmo então identificou na primeira parte o vetor cromático [D F# A], na segunda parte o vetor cromático [E G# B D], na terceira parte o vetor cromático [G B D], e na última parte novamente o vetor cromático [D F# A], identificando ao fim a seguinte sequência de acordes: D | E7 | G | D .

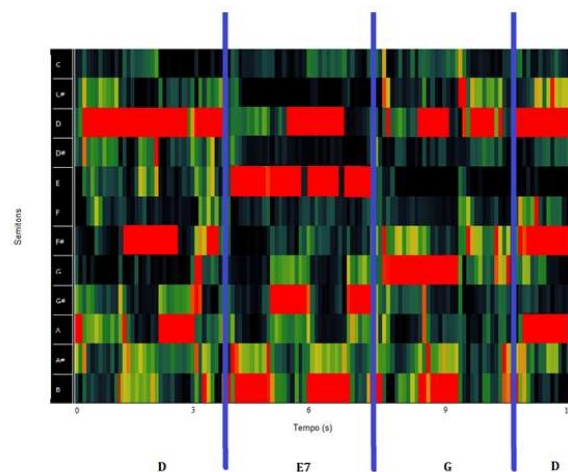


Figura 5: Cromagrama de um excerto de 6 segundos da canção “*Eight Days a Week*” dos Beatles, com a marcação em azul das linhas do compasso e os acordes identificados a partir dos vetores cromáticos.

A última parte detecção do algoritmo será então o tom da peça musical. Importante notar que o algoritmo não foi projetado para detectar mudança de tom no decorrer da peça musical, ou seja, caso haja alguma modulação no decorrer da canção, o sistema não irá detectar.

A Figura 6 apresenta como o resultado da tablatura encontrada para o exemplo da canção “*Eight Days a Week*” dos Beatles seria apresentada ao usuário:

||: D | E7 | G | D :||

Figura 6: Apresentação para o usuário da tablatura do exemplo do excerto da canção “*Eight Days a Week*” dos Beatles.

5. ACURÁCIA DO ALGORITMO

Ser capaz de medir (ou comparar) quantitativamente o quão um algoritmo consegue detectar acordes com fidelidade é essencial. Além de permitir eventuais ajustes, também permite uma comparação objetiva de diferentes algoritmos e métodos.

Compilar um conjunto de “treinamento” (ou comparação) do zero pode consumir muito tempo, pois os acordes corretos para cada música devem ser determinados por alguém que tenha a capacidade de reconhecê-los de ouvido. Esta pode ser a razão pela qual o número de conjuntos de anotações de acordes publicamente disponíveis é muito baixo, infelizmente. Além disso, as músicas em que se baseiam geralmente são protegidas por direitos autorais, difíceis de obter ou muito específicas em seu gênero. As transcrições variam em qualidade e precisão e podem obviamente conter erros humanos.

No caso deste trabalho foi usado um *dataset* dos Beatles disponível *online* contendo 180 canções da banda, correspondendo aos seus 13 álbuns de estúdio. De todo esse conjunto de dados, foram utilizadas ao todo 50 canções como objeto de análise.

A qualidade de detecção de acordes geralmente é medida em termos de “taxa de sobreposição”. Quase todas as peças musicais analisadas obtiveram acurácias que variaram entre 60% e 90%, com apenas 4 canções possuindo acurácias inferiores a 50%, e 4 canções atingindo acurácias superiores a 90%. A máxima alcançada foi de 92,7%, e a mais baixa de 15,8%.

Alcançar acurácias próximas a 100% é praticamente impossível, e nesse sentido acurácias acima de 80% já podem ser consideradas muito boas, e suficientes para a maioria das aplicações. Por outro lado, estimativas com taxas de sobreposição abaixo de 50% podem ser consideradas como falhas do sistema. A Figura 7 apresenta a distribuição normal das porcentagens de acurácia obtidas através do algoritmo utilizado.

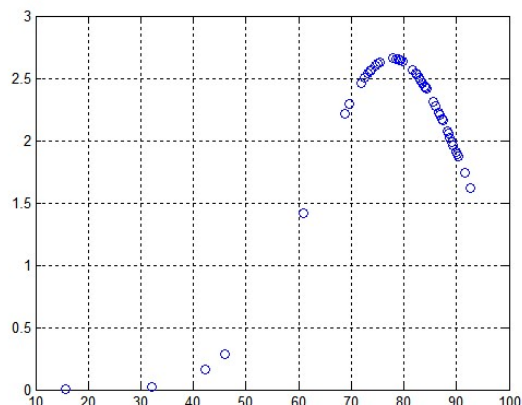


Figura 7: Distribuição normal das porcentagens de acurácia obtidas nos testes realizados com 50 canções dos Beatles.

Ao observar quais músicas obtiveram acurácias baixíssimas, as razões para o mau desempenho delas são na maioria dos casos evidentes, como no caso da acurácia mais baixa de 15,8%, nominalmente a canção “*The End*” dos Beatles, presente em seu álbum *Abbey Road* de 1969, que possui um longo solo de bateria, além de vários solos de guitarras ao longo da canção. No geral, porém, o sistema se comportou de forma bastante satisfatória e os casos de acurácia baixa foram poucos.

6. CONSIDERAÇÕES FINAIS

O objetivo desse trabalho foi através de técnicas de processamento digital de sinais áudio desenvolver um sistema de detecção automática de acordes por meio de informações espectrais de peças musicais usadas como sinais de entrada. Os resultados acadêmicos foram muito satisfatórios, porém entende-se que o algoritmo pode passar por melhorias para aumentar a acurácia de seus acertos.



Parte da discrepância de resultados não ocorreu devido ao processo de reconhecimento em si, mas do fato que há outras tríades possíveis, tais como as *sus2* e *sus4*, além de uma grande quantidade de tétrades que não foram implementadas, dentre outras possibilidades de formação de acordes (como as tétrades com tensões).

Entende-se portanto que a partir da ampliação do “dicionário de acordes” o sistema poderá ser melhor treinado para a identificação de uma gama maior de vetores cromáticos, o que possivelmente acarretará em uma maior acurácia geral.

Pretende-se ainda passar a implementação de *Matlab* para a linguagem de programação *Python*, onde pretende-se ainda elaborar uma interface gráfica de usuário. O objetivo final deste projeto é que o sistema seja utilizado como ferramenta didática transdisciplinar de Física e Música para estudantes em geral.

REFERÊNCIAS

- [1] Fujishima, T. Realtime chord recognition of musical sound: a system using Common Lisp Music. In: Proc. Int. Computer Music Conference, pp 464–467, 1999.
- [2] Loy, G. Musimathics: the mathematical foundations of music. MIT Press. 2006.
- [3] Burgoyne, J. A.; Wild, J.; Fujinaga, I. An expert ground-truth set for audio chord recognition and music analysis. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), pp. 633–638. 2011.
- [4] Benson, D. J. Music: A Mathematical Offering. Cambridge University Press. 2006.
- [5] Howard, D. M.; Angus, J. A. S. Acoustics and Psychoacoustics. Focal Press. 2009.