



Trabalho de Conclusão de Curso

Transcrição Automática de Acordes: Análise Espectral e Classificação.

Por Gabriel Paulino Farias da Silva

Orientado por
Prof. Dr. Thiago Damasceno Cordeiro

Universidade Federal de Alagoas
Instituto de Computação
Maceió, Alagoas
Outubro 20, 2023

UNIVERSIDADE FEDERAL DE ALAGOAS
Instituto de Computação

TRANSCRIÇÃO AUTOMÁTICA DE ACORDES: ANÁLISE ESPECTRAL E CLASSIFICAÇÃO.

Trabalho de Conclusão de Curso submetido
ao Instituto de Computação da Universidade
Federal de Alagoas como requisito parcial
para obter o grau de bacharel em Engenharia
da Computação.

Gabriel Paulino Farias da Silva

Orientador: Prof. Dr. Thiago Damasceno Cordeiro

Banca Examinadora:

Jobson de Araújo Nascimento	Prof. Me., UFAL
Glauber Rodrigues Leite	Prof. Me., UFAL

Maceió, Alagoas
Outubro 20, 2023

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecária: Helena Cristina Pimentel do Vale – CRB4 –661

S586t Silva, Gabriel Paulino Farias da.
Transcrição automática de acordes : análise espectral e classificação / Gabriel Paulino Farias da Silva. – 2024.
28 f : il.

Orientador: Thiago Damasceno Cordeiro.
Monografia (Trabalho de Conclusão de Curso em Engenharia de Computação) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2024.

Bibliografia: f. 28.

1. Engenharia de Computação. 2. Processamento de sinais. 3. Análise espectral. 4. Música. 5. Transcrição automática de acordes. I. Título.

CDU: 004.41:78

Agradecimentos

À minha família, por todo o apoio que me deram ao longo da minha vida.

À minha esposa, Larissa, por está sempre ao meu lado.

Aos professores e toda equipe do Instituto de Computação, onde passei momentos que levarei para toda minha vida.

Ao meu orientador, Thiago Cordeiro, por abraçar a ideia, por sua orientação e compreensão.

Gabriel Paulino Farias da Silva

Resumo

A transcrição de acordes é uma técnica com alto grau de dificuldade, mesmo para alguns músicos experientes. Dito isso, este trabalho possui como objetivo a apresentação do método de transcrição automática de acordes, avaliando diferentes técnicas de análise espectral para a identificação dos componentes de frequência do sinal e por meio da utilização de algoritmos de programação dinâmica para classificação dos acordes. Como resultado, verificou-se que a transformada Q constante e o algoritmo Viterbi, mostraram-se eficientes para a identificação das composições de menor complexidade, compostas por acordes maiores e menores, trazendo resultados satisfatórios para músicas mais complexas que apresentem acordes dissonantes. Por fim, o modelo utilizado, mesmo sem a apresentação de 100% da taxa de acertos para as músicas compostas por acordes dissonantes, mostrou-se como uma ferramenta formidável para auxiliar na transcrição de acordes, tanto para músicas simples quanto para as mais complexas.

Palavras-chave: Processamento de sinais; Análise espectral; Música; Viterbi; Transformada de Fourier; Transformada Q Constante.

Abstract

Chord transcription is a technique with a high degree of difficulty, even for some experienced musicians. That said, this work aims to present the method for automatic transcription of chords, evaluating different spectral analysis techniques for identifying the frequency components of the signal and through the use of dynamic programming algorithms to classify chords. As a result, it was found that the constant Q transform and the Viterbi algorithm proved to be efficient for identifying less complex compositions, composed of major and minor chords, bringing significant results for more complex songs that feature dissonant chords. Finally, the model used, even without presenting a 100% hit rate for songs composed of dissonant chords, proved to be a formidable tool to assist in the transcription of chords, both for simple and more complex songs.

Keywords: *Signal processing; Spectral analysis; Music; Viterbi; Fourier Transform; Constant-Q Transform.*

Lista de Figuras

2.1	Gráficos da nota Dó emitida por diferentes instrumentos. Fonte autoral.	4
2.2	Gráficos da nota Dó emitida por diferentes instrumentos, no domínio da frequência. Fonte autoral.	5
2.3	Figura ilustrativa de duas oitavas de um piano com as 12 notas fundamentais em cada oitava.	6
2.4	Acorde Ré maior no domínio da frequência. Fonte autoral.	9
2.5	Figura ilustrativa da formação do acorde de Ré maior no violão. Fonte: [e Escalas, 2020]	9
3.1	Espectrograma da transformada de Fourier de tempo curto (STFT) de um sinal sonoro musical. Fonte autoral.	12
3.2	Espectrogramas STFT x CQT. Fonte autoral.	15
5.1	Espectrograma do processamento de acordes maiores com STFT x CQT. No eixo y do espectrograma temos a distribuição das 12 notas musicais. Fonte autoral.	23
5.2	Espectrograma com os resultados do algoritmo de Viterbi para os acordes maiores. No eixo y da figura, temos a representação de cada acorde maior e menor. Sendo exibidos somente os que foram identificados na transcrição. Fonte autoral.	24
5.3	Espectrograma das notas de introdução da música Vento no Litoral. Fonte autoral.	24
5.4	Espectrograma dos acordes de introdução da música Vento no Litoral. Fonte autoral.	25
5.5	Espectrograma dos acordes da primeira parte da música O Sol do Jota Quest. Fonte autoral.	25

Lista de Tabelas

2.1	Frequência em Hz de cada nota musical nas 6 primeiras oitavas.	6
2.2	Formação dos acordes maiores e menores.	8
4.1	Tabela de resultados das probabilidades de ocorrência de cada sequência de acordes.	18

Notações e Símbolos

$\in$ pertence a

A Lá

B Si

C Dó

D Ré

E Mi

F Fá

G Sol

Sustenido

b bemol

$\int$ Integral

$\sum$ somatório

Lista de Abreviaturas e Acrônimos

DFT	Discrete Fourier Transform
STFT	Short-time Fourier Transform
CQT	Constant Q Transform
HMM	Hidden Markov Model
Hz	Hertz

Conteúdo

1	Introdução	1
1.1	Justificativa	2
1.2	Objetivos	2
1.2.1	Objetivos específicos	2
1.3	Estrutura do trabalho	2
2	Teoria musical	3
2.1	Aspectos físicos do som	3
2.1.1	Frequência	3
2.1.2	Timbre	3
2.2	As notas musicais	5
2.3	Formação de acordes	7
2.3.1	Acordes maiores	7
2.3.2	Acordes menores	7
2.3.3	Acordes dissonantes	7
2.3.4	Relação acordes e frequências	8
2.4	Escalas Musicais	9
2.4.1	Harmonia	10
2.4.2	Melodia	10
2.5	Considerações Finais	10
3	Análise Espectral de um Sinal Sonoro	11
3.1	Transformada de Fourier	11
3.2	Transformada Q-constante	13
3.3	Considerações Finais	15
4	Classificação de acordes com algoritmo Viterbi	16
4.1	Modelo oculto de Markov	16
4.2	Algoritmo de Viterbi	19
4.3	Considerações Finais	21

5	Resultados Experimentais	22
5.1	Tecnologias utilizadas	22
5.2	Transformada de Fourier X Transformada Q constante	22
5.3	Classificação de acordes com algoritmo Viterbi	23
6	Conclusão	27
	Bibliografia	28

Capítulo 1

Introdução

De acordo com [Kothe, 2006], a análise espectral pode ser entendida como a divisão do sinal em diversas faixas de frequência. Para sinais de áudio, com uma atenção especial para sinais musicais, ao realizar a análise espectral é necessário considerar as particularidades destes tipos de sinais, que vão desde o emissor do sinal até as características musicais, como a escala musical utilizada.

Segundo [Abreu, 2021] a escala mais utilizada na música ocidental, chamada de escala de temperamento igual, utiliza notas musicais com frequências fundamentais espaçadas de acordo com uma progressão geométrica. Por conta disso, as notas presentes em um sinal de áudio musical, podem não ser acuradamente representadas pela transformada discreta de Fourier (DFT), já que esta mapeia as frequências de forma linear.

Para resolver esse problema, foi criado a transformada Q constante (CQT) que possui uma resolução frequencial proporcional a frequência central de cada raia frequencial, permitindo assim um mapeamento mais homogêneo das notas musicais no espectro. [Abreu, 2021]

Para completar o processo, é necessário transcrever os acordes conforme as frequências identificadas a partir da análise espectral, o modelo mais utilizado para estimar a transcrição dos acordes, devido à característica probabilística do problema, é o modelo oculto de Markov (HMM), e para calcular a sequência de estados ocultos (acordes) mais provável, é utilizado o algoritmo de Viterbi, assim como é descrito em [Accioly, , p. 10]

Uma metodologia alternativa para a transcrição automática de acordes, segundo [Cleto et al., 2010], seria a utilização de técnicas de redes neurais como a Multilayer Perceptron, aplicando como entrada as amplitudes obtidas por uma transformada de Fourier do sinal e obtendo como saída os valores que representam os acordes. A metodologia em questão não é objeto de estudo deste trabalho, mas apresenta resultados significativos, o que levanta a possibilidade de um trabalho futuro explorando ambas as metodologias.

1.1 Justificativa

A transcrição automática de acordes é objeto de estudo de diversos trabalhos, sempre visando explorar novas técnicas na busca de formular modelos mais eficientes. Além disso, tem relação direta com a música, a qual é uma das principais formas de manifestação cultural humana, e um dos setores que mais movimentam a economia mundial.

Para muitos músicos, o processo de transcrever os acordes de uma música pode ser repetitivo e cansativo, a depender do nível de complexidade da música, e a experiência do músico. Logo, ter uma ferramenta capaz de transcrever os acordes de forma automática e com precisão, é de interesses de toda a comunidade musicista.

1.2 Objetivos

O objetivo geral desse trabalho é apresentar um modelo capaz de auxiliar músicos a realizar automaticamente a transcrição de acordes. Tendo em vista que, em situações onde o modelo não consiga apresentar resultados eficientes, um músico com experiência possa identificar e realizar as devidas correções.

1.2.1 Objetivos específicos

- Apresentar a correlação entre as notas musicais e as componentes de frequência de um sinal proveniente de um instrumento musical.
- Realizar estudo de comparação entre a transformada de Fourier e a transformada Q constante, em busca de validar a superioridade da transformada Q constante para a análise de sinais musicais.
- Utilizar o algoritmo Viterbi para a classificação de acordes, mediante as componentes de frequência identificadas através da análise espectral.

1.3 Estrutura do trabalho

No segundo capítulo serão abordados os conceitos básicos de teoria musical e a relação entre as notas musicais e as frequências fundamentais. No terceiro capítulo serão apresentadas as técnicas de análise espectral que serão objeto de estudo desse trabalho, suas características e particularidades. No quarto capítulo teremos uma breve descrição do modelo oculto de Markov e a explicação de como funciona o algoritmo de Viterbi. No quinto capítulo serão apresentados os resultados do modelo proposto para transcrição automática de acordes. O sexto capítulo serão abordadas as conclusões finais deste trabalho.

Capítulo 2

Teoria musical

No presente capítulo serão abordadas as principais características de um sinal sonoro, com foco na relação entre frequências e notas musicais. Também será apresentado um estudo sobre a definição e formação de acordes.

2.1 Aspectos físicos do som

O som é uma onda capaz de propagar-se pelo ar e por outros meios a partir da vibração de suas moléculas. Os sons são percebidos quando eles incidem sobre o nosso aparelho auditivo, traduzidos em estímulos elétricos e direcionados ao nosso cérebro, que os interpreta. Os seres humanos conseguem ouvir uma faixa de frequências sonoras, chamada de espectro audível, que se estende entre 20Hz até 20.000 Hz. [Dantas and Cruz, 2018]

2.1.1 Frequência

Um dos aspectos do som é a frequência, através dessa grandeza podemos classificar os sons como graves e agudos. As ondas sonoras com maiores frequências são classificadas como mais agudas, enquanto ondas sonoras com menores frequências são classificadas por mais graves. [Dantas and Cruz, 2018]

2.1.2 Timbre

Outro aspecto que devemos abordar é a possibilidade de distinção entre sons emitidos por diferentes instrumentos. Os sons produzidos por um clarinete e por um trombone, por exemplo, são claramente diferentes, mesmo que estes emitam ondas sonoras com a mesma frequência fundamental. Isso deve-se a geometria do instrumento e o material do qual é feito, determinando assim as várias frequências que estarão presentes, e como estarão presentes, na emissão de cada som. A qualidade do som que permite reconhecermos diferenças quanto à origem de emissão é denominada timbre.

Isso acontece porque o som é uma composição de ondas sonoras, denominadas harmônicos, e a presença dos harmônicos e suas intensidades relativas na composição de cada som (de cada onda resultante) difere para sons produzidos por fontes diferentes; o som produzido por uma fonte específica carrega uma “assinatura” que o caracteriza como produzido por aquela fonte; o som tem um timbre específico, tem uma “personalidade”. Descrever esse padrão significa observar o espectro sonoro da onda produzida.[Dantas and Cruz, 2018]

O som pode ser representado por uma função periódica, e conforme a análise de Fourier, qualquer função periódica pode ser escrita como a soma de funções senoidais. Tais ondas senoidais, de forma individual, são os harmônicos do sinal sonoro.

Na Figura 2.1, podemos observar as ondas sonoras formadas por diferentes instrumentos, porém, emitindo a mesma nota musical, a nota Dó.

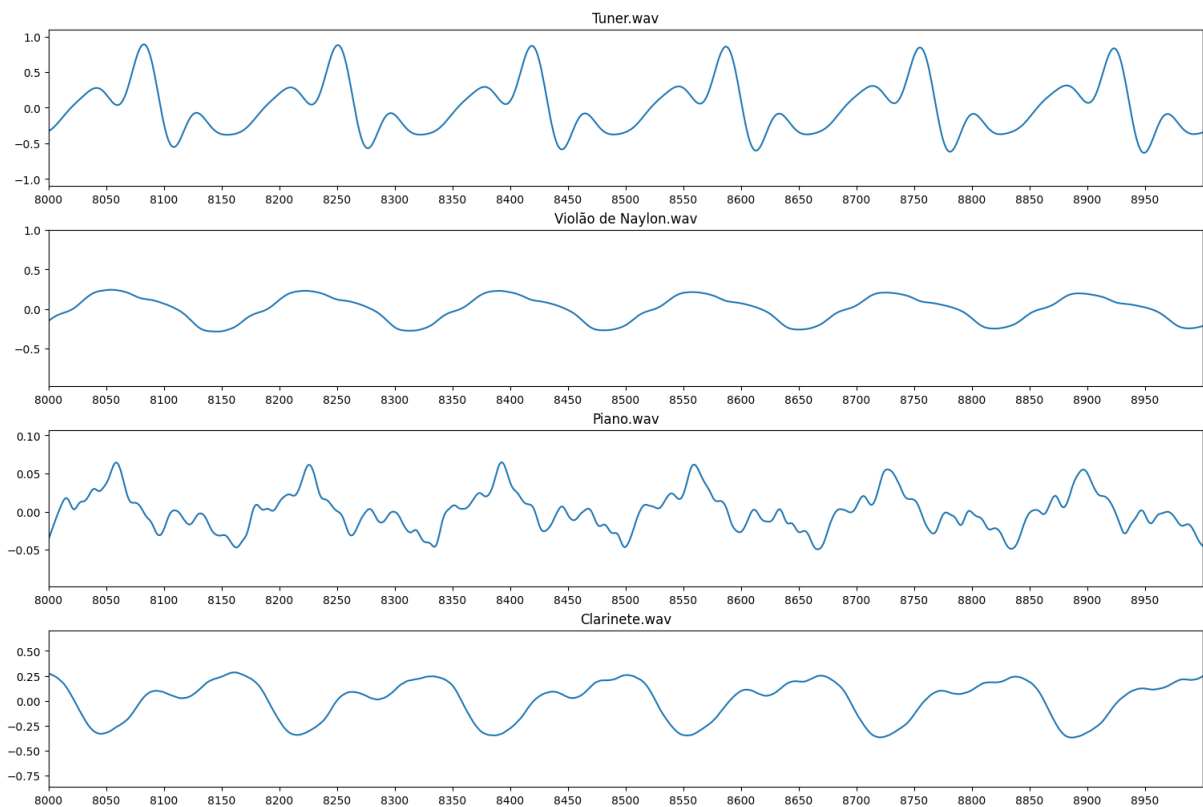


Figura 2.1: Gráficos da nota Dó emitida por diferentes instrumentos. **Fonte autoral.**

É possível visualizar que o formato de onda de todos os gráficos se repetem com o mesmo período, logo é possível afirmar que para todas as ondas temos a mesma frequência fundamental. Para facilitar o entendimento, temos na Figura 2.2 as componentes de cada sinal no domínio da frequência. Nesta figura, podemos averiguar que a frequência fundamental, a de maior potência, é mais ou menos 260Hz, que corresponde a nota Dó. As demais frequências observadas são os harmônicos gerados pelas características individuais de cada instrumento.

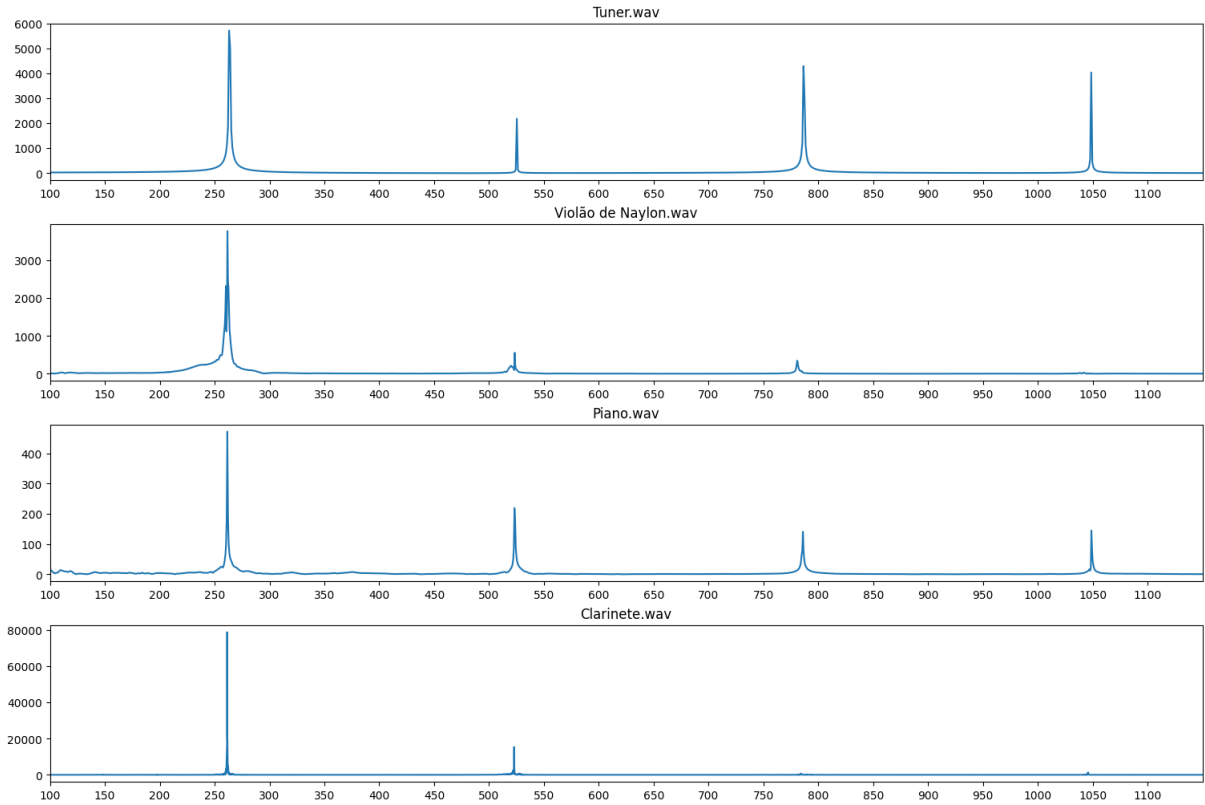


Figura 2.2: Gráficos da nota Dó emitida por diferentes instrumentos, no domínio da frequência. **Fonte autoral.**

2.2 As notas musicais

As notas musicais representam ondas sonoras em uma determinada faixa de frequências. As principais notas músicas que temos são Dó, Ré, Mi, Fá, Sol, Lá, Si, que podem ser representadas respectivamente pelas seguintes letras, C, D, E, F, G, A, B [Gusmão, 2012]. Este conjunto de notas pertence a uma determinada oitava. Após a nota Si, repetimos a sequência novamente e dizemos que estas fazem parte de uma oitava subsequente.

Nesse trabalho, nossa primeira oitava iniciará com a frequência de aproximadamente 16,35Hz, que representa a nota musical Dó. Adotaremos esta frequência, pois esta é a mais próxima do limite inferior das frequências auditivas para o ser humano. Logo, o grupo de frequências que representam todas as notas Dó subsequentes tem o seguinte formato:

$$C_n = C_0 * 2^n = 16,35 * 2^n \quad (2.1)$$

onde C_n é uma nota Dó na n -ésima oitava e C_0 é a nota Dó na primeira oitava, definida neste trabalho com frequência de aproximadamente 16,35 Hz. Como exemplo para neste caso, para calcular qual frequência representa a nota Dó na sua 5ª oitava (C_4), teríamos:

$$C_4 = 16,35 * 2^4 = 16,35 * 16 = 261,60\text{Hz} \quad (2.2)$$

Na Tabela 2.1 temos as frequências-base para cada nota musical em suas respectivas oitavas. [Oliveira, 2002]

Oitava	Dó	Ré	Mi	Fá	Sol	Lá	Si
1 ^a	16,35	18,37	20,60	21,82	24,50	27,50	30,86
2 ^a	32,70	36,74	41,20	43,64	49,00	55,00	61,72
3 ^a	65,40	73,48	82,40	87,28	98,00	110,00	123,44
4 ^a	130,80	146,96	164,80	174,56	196,00	220,00	246,88
5 ^a	261,60	293,92	329,60	349,12	392,00	440,00	493,76
6 ^a	523,20	587,84	659,20	698,24	784,00	880,00	987,52

Tabela 2.1: Frequência em Hz de cada nota musical nas 6 primeiras oitavas.

Como podemos observar na Tabela 2.1, existe uma distância entre as frequências de duas notas musicais subsequentes. Esta distância é representada pela unidade de tom, sendo 1 tom a distância entre as notas Dó e Ré, por exemplo, ou entre as notas Ré e Mi. Porém, como é possível observar, a distância entre as notas Mi e Fá ou entre as notas Si e Dó são menores, mais ou menos a metade do valor em relação às outras. Para estes casos representamos esta distância como sendo de meio-tom ou semitom e podemos denotar que 2 semi-tons equivalem a 1 tom [Gusmão, 2012].

Quando a distância entre duas notas é de 1 tom, temos uma nota intermediária. Para classificar essa nota, utilizamos dois símbolos: sustenido(#) e bemol (b). Um exemplo seria a nota entre Dó e Ré, onde poderíamos ter Dó sustenido (C#) ou Ré bemol (Db) como representação da mesma nota [Frozza, 2019]. Com as notas intermediárias, totalizamos 12 notas musicais fundamentais dentro de uma mesma oitava, como pode ser visto na Figura 2.3.

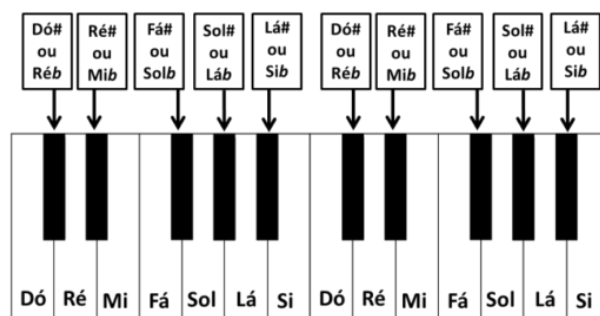


Figura 2.3: Figura ilustrativa de duas oitavas de um piano com as 12 notas fundamentais em cada oitava.

Também é importante observar que a diferença da frequência em Hz entre duas notas não é constante e depende de qual oitava as notas se encontram, seguindo uma progressão geométrica. Logo, a diferença de frequência entre duas notas segue a fórmula a seguir:

$$C_n = C_0 * 2^n$$

$$\begin{aligned}
 Dn &= D0 * 2^n \\
 Dn - Cn &= (D0 * 2^n) - (C0 * 2^n) \\
 Dn - Cn &= (D0 - C0) * 2^n
 \end{aligned} \tag{2.3}$$

onde Cn é uma nota Dó na n -ésima oitava, $C0$ é a nota Dó na primeira oitava, Dn é uma nota Ré na n -ésima oitava, $D0$ é a nota Ré na primeira oitava.

2.3 Formação de acordes

Acordes são a composição de 3 ou mais notas musicais executadas simultaneamente. Na teoria musical existem diversos tipos de acordes, mas os principais, e os abordados nesse trabalho, serão os chamados acordes maiores e os acordes menores.

2.3.1 Acordes maiores

Os acordes maiores também são representados por notas musicais. Logo, temos os acordes Dó maior, Ré maior e assim por diante. Todo acorde maior é formado pela união de 3 notas musicais, sendo a primeira a que dá nome ao acorde, também chamada de nota tônica; a segunda nota é a nota com uma distância de 2 tons da primeira; e a terceira nota está a uma distância de 3,5 tons da primeira nota ou 1,5 tons da segunda. Como exemplo, temos o acorde de Dó maior composto pelas notas musicais Dó, Mi e Sol. De fato, podemos verificar que a distância entre Dó e Mi é de 2 tons e que a distância entre Mi e Sol é de 1,5 tons [Faria, 1999]. Na Tabela 2.2 são apresentadas as notas que compõem acordes maiores e menores.

2.3.2 Acordes menores

O acorde menor também é formado pela união de 3 notas musicais, sendo a primeira a que dá nome ao acorde, também chamada de nota tônica; a segunda nota com uma distância de 1,5 tons da primeira; e a terceira nota uma distância de 3,5 tons da primeira nota, ou 2 tons da segunda. Como exemplo, temos o acorde de Ré menor composto pelas notas musicais Ré, Fá e Lá, onde podemos verificar que a distância entre Ré e Fá é de 1,5 tons e a distância entre Fá e Lá é de 2 tons [Faria, 1999].

2.3.3 Acordes dissonantes

Os acordes dissonantes são variações dos acordes maiores e menores, onde, além das três notas principais, são adicionadas outras notas que não fazem parte do acorde principal. Estes acordes possuem esse nome porque as notas adicionadas destoam das notas principais do acorde. Por exemplo, o acorde Dó menor com sétima ($Cm7$) é formado pelas

notas Dó, Mi bemol, Sol e, por último, Si bemol, representando a sétima para este acorde (este conceito será fornecido a seguir). Acordes dissonantes são bastante utilizados no meio musical, mas seu uso não é obrigatório. Alunos iniciantes muitas vezes substituem estes acordes para executar uma música de forma mais fácil [Faria, 1999].

2.3.4 Relação acordes e frequências

Como podemos observar, para identificarmos um acorde de um instrumento musical em um sinal de onda sonora, é preciso identificar as frequências desse sinal, identificar quais notas representam essas frequências e classificar o acorde conforme as notas encontradas.

Acorde	1ª nota	2ª nota	3ª nota
C	C	E	G
C#	C#	F	G#
D	D	F#	A
D#	D#	G	A#
E	E	G#	B
F	F	A	C
F#	F#	A#	C#
G	G	B	D
G#	G#	C	D#
A	A	C#	E
A#	A#	D	F
B	B	D#	F#
Cm	C	Eb	G
C#m	C#	E	G#
Dm	D	F	A
D#m	D#	Gb	A#
Em	E	G	B
Fm	F	Ab	C
F#m	F#	A	C#
Gm	G	Bb	D
G#m	G#	B	D#
Am	A	C	E
A#m	A#	Db	F
Bm	B	D	F#

Tabela 2.2: Formação dos acordes maiores e menores.

Na Figura 2.4, temos as frequências emitidas por um acorde de Ré maior gerado por um violão, e na figura 2.5, o formato do acorde no braço de um violão. As frequências encontradas foram: 146,5Hz, 220,0Hz, 293,0Hz, 369,4Hz, 440,4Hz, 585,9Hz 660,8Hz, 739,2Hz, 884,6Hz, que representam, respectivamente, as seguintes notas: D3, A3, D4, F#4, A4, D5, E5, F#5, A5.

Como visto na tabela 2.2, o acorde de Ré maior é formado pela notas Ré(D), Fá sustenido(F#) e Lá(A), porém, é possível observar que no acorde de Ré maior da figura

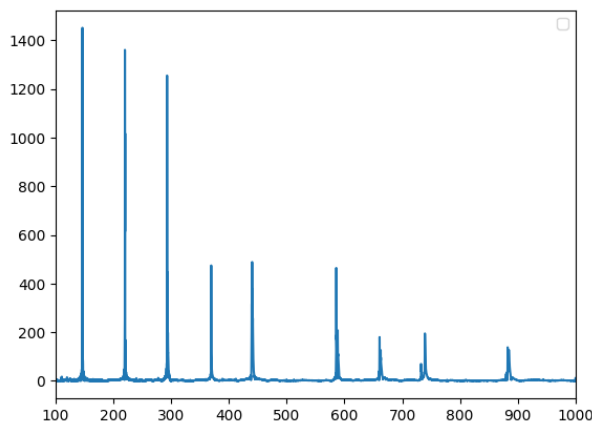


Figura 2.4: Acorde Ré maior no domínio da frequência. **Fonte autoral.**

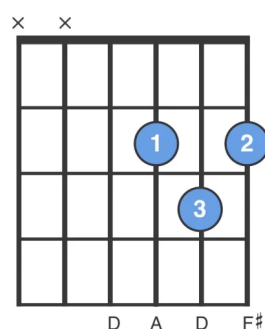


Figura 2.5: Figura ilustrativa da formação do acorde de Ré maior no violão. **Fonte:** [e Escalas, 2020]

2.4 foi obtido também a nota E5(660.8Hz). Isso ocorre devido à soma de harmônicos, resultando em uma frequência com amplitude próxima à amplitude das frequências de pico reais, também havendo a possibilidade do instrumento estar parcialmente desafinado, ou desregulado, além de ruídos e outros fatores. Devido a essas inconsistências comuns, apresentaremos no capítulo 4, uma algoritmo para a classificação de acordes.

2.4 Escalas Musicais

Uma escala é uma série de notas consecutivas que fornece o material para a construção de um trecho musical ou de uma peça musical inteira. Embora exista uma quantidade enorme de diferentes escalas, as mais importantes para o estudo da música são as escalas maiores e menores [Gusmão, 2012].

As escalas são definidas pelas suas distribuições de tons e semitons entre as notas que a compõe. Através das escalas é possível construir uma progressão de acordes com melodia harmônica, e para realizar a composição de uma música, é de suma importância, o conhecimento sobre as principais escalas musicais [Gusmão, 2012].

2.4.1 Harmonia

Segundo [Dantas and Cruz, 2018], a harmonia estuda as diversas combinações de sons emitidos simultaneamente. Trata de como são constituídos os acordes, das consonâncias e dissonâncias e de como os acordes se organizam para o acompanhamento, ou seja, para a harmonização de melodias.

2.4.2 Melodia

Melodia está associada ao fenômeno da combinação de sons de maneira sucessiva e com diferentes características emitidos em sequência. A sequência de sons de diferentes frequências resulta em uma estrutura que podemos denominar de desenho melódico ou melodia [Dantas and Cruz, 2018].

2.5 Considerações Finais

Neste capítulo, fornecemos uma introdução fundamental à teoria musical, um conhecimento essencial para compreender os elementos que compõem a linguagem da música. Exploramos conceitos cruciais relacionados às notas musicais, suas respectivas frequências e como esses elementos formam a base da harmonia musical. Além disso, investigamos a complexa estrutura dos acordes, destacando seus principais tipos e suas funções dentro da música. Ao final deste capítulo, o leitor estará bem preparado para a exploração das práticas e ferramentas que tornarão possível a classificação das notas musicais em qualquer composição sonora. Esta base teórica é essencial para o capítulo seguinte, onde nos aprofundaremos em técnicas avançadas de processamento digital de sinais para determinar as frequências de um sinal sonoro. Essas técnicas são cruciais para a identificação precisa das notas musicais, um passo importante na análise e compreensão dos acordes que compõem uma música.

Capítulo 3

Análise Espectral de um Sinal Sonoro

Neste capítulo discutiremos sobre técnicas de análise espectral de sinais sonoros musicais, realizando um estudo das características de duas técnicas: a transformada de Fourier e a transformada Q-constante.

3.1 Transformada de Fourier

Sinais de áudio são geralmente representados por senoides no domínio do tempo. Porém, com essa representação não é possível identificar algumas informações importantes para a transcrição musical, como a composição em frequências do sinal, necessário para a identificação das notas musicais. Para obtemos essas informações, é necessário utilizarmos ferramentas que forneçam a representação do sinal também no domínio da frequência.

A ferramenta mais utilizada para esta finalidade é a transformada de Fourier, que transforma um sinal no domínio do tempo em um sinal no domínio da frequência. Porém, a transformada de Fourier não considera que sinais provenientes de instrumentos musicais não são estacionários, logo seus componentes frequenciais mudam conforme o tempo. Para contornar esta situação iremos utilizar a transformada de Fourier em pequenos quadros de tempo, e unir essas informações para termos uma representação espectral das componentes de frequência do sinal no domínio do tempo, técnica conhecida como transformada de Fourier de Tempo Curto (STFT, do inglês Short-Time Fourier transform).

A transformada de Fourier de um sinal analógico $x(t)$ é definida como:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (3.1)$$

Como estamos trabalhando com sinais de origem digital, é necessário utilizarmos outro método mais adequado. Para isso, assumindo que o sinal tem duração limitada, o método

mais comum é o da Transformada Discreta de Fourier (DFT, do inglês Discrete Fourier Transform), o qual é definido da seguinte maneira:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j \frac{2\pi}{N} kn} \quad (3.2)$$

onde k é o índice de frequência e N é o número de amostras utilizadas no cálculo da DFT

Para obtermos uma representação tempo-frequencial do sinal analisado é necessário utilizarmos uma janela $w[n]$ que deve multiplicar o sinal e realizar diversas DFT, deslocando a janela no tempo a cada iteração. A partir da utilização da STFT é possível gerar o espectrograma de potência, tomando o módulo da STFT ao quadrado. A representação matemática dessa técnica é definida da seguinte maneira:

$$X[m, k] = \left| \sum_{n=0}^{N-1} x[n] w[n - hm] e^{-j \frac{2\pi}{N} kn} \right|^2 \quad (3.3)$$

onde $m \in N$ é o índice temporal da STFT e $h \in N$ é o tamanho, em amostras, do salto da janela [Abreu, 2021].

Na Figura 3.1, temos um espectrograma gerado a partir do resultado da STFT de um sinal de áudio, através desta figura é possível observar as frequências que compõem a música em cada instante do tempo de duração do sinal analisado.

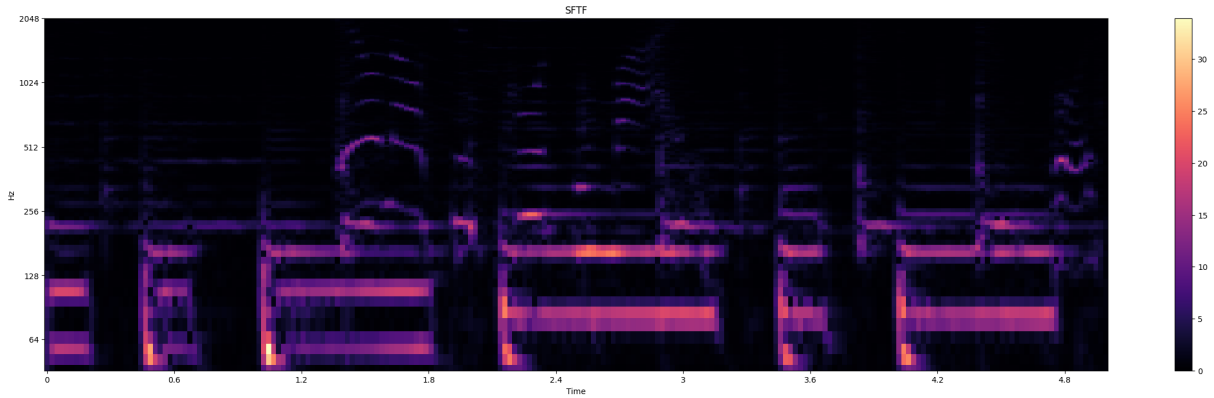


Figura 3.1: Espectrograma da transformada de Fourier de tempo curto (STFT) de um sinal sonoro musical. **Fonte autoral.**

Um parâmetro bastante importante para nossa análise é o tamanho da janela utilizada, pois conforme o tamanho da janela temos uma resolução maior ou menor, tanto para o domínio da frequência quanto para o domínio do tempo. Uma janela maior, com mais amostras, fornece uma melhor resolução para o domínio da frequência, porém fornece uma resolução menor para o domínio do tempo. Por outro lado, uma janela menor, menos amostras por janela, fornece uma melhor resolução no tempo e uma resolução piorada no domínio da frequência.

Nota-se, então, uma relação inversa entre as resoluções em função do tamanho da janela. Tal relação é formalmente definida pelo Princípio da Incerteza, que mostra que é impossível obter, simultaneamente, uma precisão arbitrariamente alta no tempo e na frequência.

Poderíamos então buscar um valor para o tamanho da janela onde satisfizesse uma resolução satisfatória de acordo com critérios previstos. Uma janela com tamanho não ótimo, para ambos os domínios. Porém, as frequências observadas em sinais provindos de instrumentos musicais possuem uma variação relevante de frequências ao longo do tempo, logo uma janela com tamanho constante não é a melhor solução.

3.2 Transformada Q-constante

Como visto no capítulo 2, utilizamos notas musicais com frequências espaçadas segundo uma progressão geométrica, por essa razão, as notas musicais de um sinal de áudio de um instrumento podem não ser bem representados por uma DFT, já que a DFT mapeia as frequências de forma linear. Para resolver este problema, foi criada uma variação da STFT: A Transformada Q-constante (CQT, do inglês Constant Q-Transform), que possui uma resolução frequencial proporcional à frequência central de cada intervalo frequencial, permitindo assim um mapeamento mais homogêneo das notas musicais no espectro. O termo “Q constante” está atribuído ao fator da qualidade Q do filtro associado, definido como a razão entre cada frequência e a resolução frequencial [Abreu, 2021].

$$Q = \frac{f_k}{\Delta_f} \quad (3.4)$$

As frequências da CQT são distribuídas da seguinte maneira:

$$f_k = f_{min} * 2^{\frac{k}{B}} \quad (3.5)$$

onde $f_{min} = f_0$ é uma frequência mínima estabelecida, B é o número de componentes por oitava e f_k é a k -ésima componente, sendo $f_{min} < f_k < f_{max}$ e a frequência f_{max} é menor do que a frequência de Nyquist. Assim, a resolução frequencial, definida como a distância entre componentes adjacentes, é calculada como:

$$\begin{aligned} \Delta f_{CQT} &= f_{k+1} - f_k \\ \Delta f_{CQT} &= f_{min} * 2^{\frac{k+1}{B}} - f_{min} * 2^{\frac{k}{B}} \\ \Delta f_{CQT} &= f_{min} * 2^{\frac{k}{B}} * (2^{\frac{1}{B}} - 1) \\ \Delta f_{CQT} &= f_k * (2^{\frac{1}{B}} - 1) \end{aligned} \quad (3.6)$$

Logo, se utilizarmos a equação 3.4, teremos o seguinte resultado para Q:

$$Q = \frac{f_k}{f_k * (2^{\frac{1}{B}} - 1)}$$

$$Q = (2^{\frac{1}{B}} - 1)^{-1} \quad (3.7)$$

o que prova que Q é uma constante [Abreu, 2021].

Como visto na seção 3.1, a resolução frequencial da STFT possui uma relação inversa com o tamanho da janela. De forma mais específica, pode-se demonstrar que seu valor é:

$$\Delta f_{STFT} = \frac{F_s}{N} \quad (3.8)$$

onde F_s é a frequência de amostragem do sinal e N é o tamanho da janela, em amostras. Para que o fator Q se mantenha constante, é necessário que cada componente espectral seja associada a um tamanho de janela. Portanto, na CQT, o parâmetro N passa a ser um conjunto de N_k tamanhos de janela, calculados ao substituímos a resolução frequência da STFT na Equação 3.4:

$$N_k = Q \frac{F_s}{f_k} \quad (3.9)$$

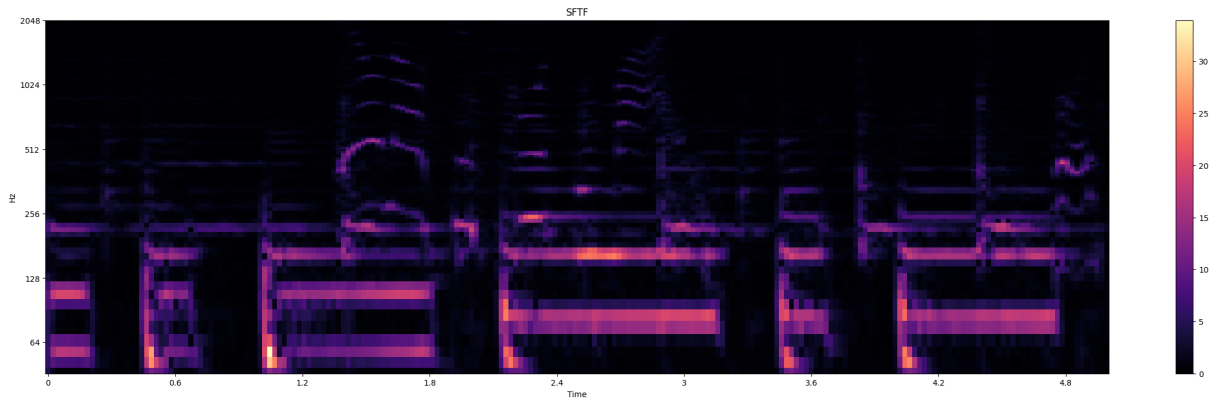
Fazendo as devidas substituições, chegamos à definição da CQT, dada por:

$$X_{CQT}[k] = \frac{1}{N_k} \sum_{n=0}^{N_k-1} x[n] w_k[n] e^{-j2\pi n \frac{Q}{N[k]}} \quad (3.10)$$

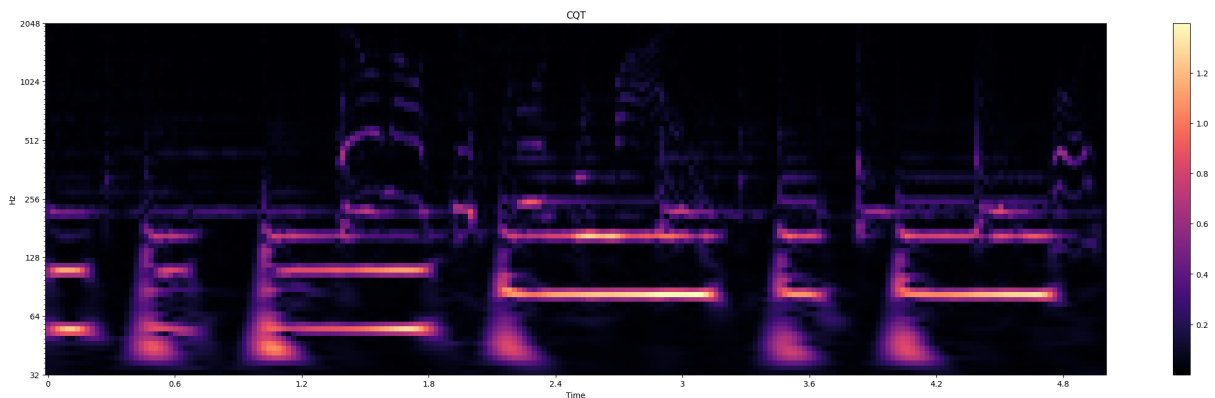
onde $w_k[n]$ é uma janela de tamanho N_k

Na Figura 3.2, temos dois espectrogramas do mesmo sinal de áudio, o primeiro gerado pela STFT e o segundo gerado pela CQT. Como esperado, podemos observar que a CQT fornece uma melhor resolução no domínio da frequência, dessa forma, possibilitando a identificação das frequências fundamentais com uma margem menor de erro.

É possível notar, no entanto, que essa representação possui um alto custo computacional, já que é necessário repetir o cálculo da DFT para cada tamanho de janela em cada segmento do sinal. Além disso, o uso de janelas diferentes resulta em uma resolução temporal diferente em cada faixa de frequências, já que, para cada componente, será analisado um trecho do sinal de tamanho específico. Essas duas características da CQT devem ser consideradas na escolha dos parâmetros de resolução e quantidade de frequências utilizadas na análise [Abreu, 2021].



(a) Espectrograma da transformada de Fourier de tempo curto(STFT).



(b) Espectrograma da transformada Q constante(CQT).

Figura 3.2: Espectrogramas STFT x CQT. **Fonte autoral.**

3.3 Considerações Finais

Através desse capítulo concluímos que para a análise espectral de sinais musicais, a transformada Q-constante se mostra mais eficiente no que se diz respeito a resolução das frequências no espectrograma, o que é de extrema importância para o resultado desse trabalho. Agora que definimos qual técnica utilizar para obter as frequências de um sinal, e como visto no capítulo 2, quais frequências representam cada nota musical, abordaremos no próximo capítulo qual modelo utilizaremos para classificar os acordes.

Capítulo 4

Classificação de acordes com algoritmo Viterbi

Neste capítulo será apresentado o modelo e algoritmo para solucionar o problema da classificação de acordes, conforme as frequências identificadas durante a análise espectral. Tendo em vista que, precisamos analisar a sequência completa dos acordes de um sinal musical, e não somente, cada acorde individualmente.

4.1 Modelo oculto de Markov

Um sistema de classificação de acordes envolve fases de treinamento e teste. Na fase de treinamento, é aplicado o modelo oculto de Markov (HMM, do inglês hidden markov model), onde cada estado representa um único acorde, e a progressão de acordes é modelada como uma série de transições entre estados. Os elementos centrais do HMM são probabilidade de transição de estado e distribuição de observações. A probabilidade de transição de estado representa a probabilidade de um acorde ser seguido por outro acorde, e a distribuição de observações descreve a probabilidade de um vetor observações ser gerado por um determinado estado [Cheng et al., 2008].

O modelo de classificação de acordes, baseado no modelo oculto de Markov, que adotaremos neste trabalho, possui a seguinte estrutura $M = (Q, S, I, T, E)$, onde:

- Q é um conjunto finito de estados ocultos, ou, conjunto de todos os acordes possíveis. Nesse trabalho consideraremos somente os acordes maiores e menores, totalizando 24 acordes.
- S é o conjunto finito de observações. Para o modelo escolhido, S será um conjunto de vetores de notas musicais, onde cada vetor conterá 3 notas musicais distintas, conforme as frequências de maior potência identificadas em cada trecho da análise espectral. Utilizaremos vetores de tamanho 3, pois, todos os acordes maiores e

menores são formados apenas por 3 notas musicais, e para facilitar a implementação do modelo.

- I é o vetor, de tamanho Q , de probabilidades iniciais, que defini a probabilidade de cada acorde, entre os 24 possíveis, ser o primeiro da sequência mais provável. Porém, adotaremos uma distribuição uniforme, ou seja, cada acorde terá a mesma probabilidade de ser o acorde inicial.
- T é uma matriz $Q \times Q$ que representa as probabilidades de transições entre os acordes.
- E é uma matriz $Q \times S$ que representa as probabilidades de emissão de cada vetor de notas gerar um acorde.

De acordo com o modelo proposto, dado um conjunto de n vetores de notas musicais, é possível calcular a probabilidade de ocorrência de uma sequência de n acordes.

Vale ressaltar que poderíamos usar um modelo mais simples, baseado somente na matriz de emissão, tendo em vista que a mesma fornece a probabilidade de ocorrência de cada acorde dado um conjunto de notas musicais. Porém, essa abordagem não seria tão eficaz quanto o modelo proposto devido à natureza da composição musical, onde é geralmente utilizado uma escala musical para definir quais acordes serão utilizados para formar a melodia, e quais as progressões de acordes possuem melhor harmonia. Logo, podemos concluir que, o acorde $n-1$ sempre terá influência no acorde n , e dessa forma um modelo de classificação de acordes baseado no modelo oculto de Markov é mais eficaz, por considerar as transições entre estados(acordes).

Para melhorar o entendimento, ilustraremos um exemplo simples com o modelo $M = (Q, S, I, T, E)$ onde:

- $Q = \{C, G\}$
- $S = \{[G, B, D]; [C, G, F]; [C, E, G]\}$
- $I = [0,5; 0,5]$

C	G
0,5	0,5

- $T = [[0,3; 0,7]; [0,8; 0,2]]$

	C	G
C	0,3	0,7
G	0,8	0,2

	s1	s2	s3
C	0,1	0,7	0,9
G	0,9	0,3	0,1

- $E = [[0,1; 0,7; 0,9]; [0,9; 0,3; 0,1;]]$

E o conjunto de vetores $V = \{[G, B, D]; [C, G, F]; [C, E, G]\}$. Qual a probabilidade de ocorrência da sequência de acordes $\{G, G, C\}$?

Primeiramente simplificaremos V para $V = \{v_1, v_2, v_3\}$

Para o primeiro acorde temos:

$$I(G) * E(G, v_1) = 0,5 * 0,9 = 0,45$$

Para o segundo acorde temos:

$$T(G, G) * E(G, v_2) = 0,2 * 0,3 = 0,06$$

Para o terceiro acorde temos:

$$T(G, C) * E(C, v_3) = 0,8 * 0,9 = 0,72$$

Logo, a probabilidade de ocorrência da sequência de acordes $\{G, G, C\}$ é:

$$0,45 * 0,06 * 0,72 = 0,0194$$

Analogamente, podemos calcular a probabilidade de cada sequência de acordes possíveis para o modelo de exemplo, assim teríamos os seguintes resultados:

C,C,C	0,0028
C,C,G	0,0007
C,G,C	0,0076
C,G,G	0,0002
G,C,C	0,0680
G,C,G	0,0176
G,G,C	0,0194
G,G,G	0,0005

Tabela 4.1: Tabela de resultados das probabilidades de ocorrência de cada sequência de acordes.

Conforme a tabela 4.1, a sequência com maior probabilidade de ocorrência é $\{G, C, C\}$, porém, calcular a probabilidade de todas as sequências para encontrar a sequência

mais provável, tem alto custo computacional. Para resolver esse problema utilizaremos o algoritmo de Viterbi.

4.2 Algoritmo de Viterbi

O algoritmo de Viterbi é um algoritmo de programação dinâmica usado para obter a sequência mais provável de estados, dado um conjunto de observações, em um modelo oculto de Markov, assim como afirma [Chen et al., 2012, P. 447].

Para calcularmos a probabilidade de cada estado oculto q , em cada posição n da sequência mais provável, utilizaremos as seguintes fórmulas:

$$Vit(S_n, Q) = \max\{Vit(S_{n-1}, i) * T(i, Q) * E(Q, S_n) : i = q_1 \dots q_m\}; \quad (4.1)$$

$$Vit(S_1, Q) = I(Q) * E(Q, S_1) \quad (4.2)$$

Para melhorar o entendimento, encontraremos a sequência mais provável do exemplo anterior, utilizando o algoritmo de Viterbi. Identificaremos como $(Q - 1)_{max}$, o estado oculto anterior que gerou a máxima probabilidade em cada passo.

$$Vit(v_1, C) = I(C) * E(C, v_1) = 0,5 * 0,1 = 0,05$$

$$Vit(v_1, G) = I(G) * E(G, v_1) = 0,5 * 0,9 = 0,45$$

$$Vit(v_2, C) = \max\{Vit(v_1, C) * T(C, C) * E(C, v_2); Vit(v_1, G) * T(G, C) * E(C, v_2)\}$$

$$Vit(v_2, C) = \max\{0,05 * 0,3 * 0,7; 0,45 * 0,8 * 0,7\}$$

$$Vit(v_2, C) = \max\{0,0105; 0,252\}$$

$$Vit(v_2, C) = 0,252$$

$$(Q - 1)_{max} = G$$

$$Vit(v_2, G) = \max\{Vit(v_1, C) * T(C, G) * E(G, v_2); Vit(v_1, G) * T(G, G) * E(G, v_2)\}$$

$$Vit(v_2, G) = \max\{0,05 * 0,7 * 0,3; 0,45 * 0,2 * 0,3\}$$

$$Vit(v_2, G) = \max\{0,0105; 0,0270\}$$

$$Vit(v_2, G) = 0,0270$$

$$(Q - 1)_{max} = G$$

$$Vit(v_3, C) = \max\{Vit(v_2, C) * T(C, C) * E(C, v_3); Vit(v_2, G) * T(G, C) * E(C, v_3)\}$$

$$Vit(v_3, C) = \max\{0,252 * 0,3 * 0,9; 0,0270 * 0,8 * 0,9\}$$

$$Vit(v_3, C) = \max\{0,0680; 0,0194\}$$

$$Vit(v_3, C) = 0,0680$$

$$Q_{max} = C$$

$$Vit(v_3, G) = \max\{Vit(v_2, C) * T(C, G) * E(G, v_3); Vit(v_2, G) * T(G, G) * E(G, v_3)\}$$

$$Vit(v_3, G) = \max\{0,252 * 0,7 * 0,1; 0,0270 * 0,2 * 0,1\}$$

$$Vit(v_3, G) = \max\{0,0176; 0,0005\}$$

$$Vit(v_3, G) = 0,0176$$

$$(Q - 1)_{max} = C$$

Logo 0,0680 é a probabilidade máxima. Conforme o resultado acima, o último acorde da sequência será C. Verificando os $Q - 1_{max}$ recursivamente, obteremos a seguinte sequência: C, C, G. Colocando na ordem do primeiro ao último acorde, temos {G, C, C} como a sequência mais provável, assim como visto na tabela 4.1

Através do exemplo acima, é possível entender como funciona a classificação de acordes através do modelo oculto de Markov e o algoritmo Viterbi. Também conseguimos observar que para um exemplo completo, onde teríamos um conjunto de estados ocultos com 24 acordes, e um conjunto de observações proporcional ao número de componentes da análise espectral, o número de iterações seria enorme.

Para possibilitar a construção da matriz de transição e a matriz de emissão para um caso real, utilizaremos o auxílio da linguagem de programação Python e a biblioteca Librosa. Tendo em vista que a construção dessas matrizes, para um caso real, considera vários fatores da teoria musical que não são objeto de estudo desse trabalho, como exemplo: formato das escalas musicais, padrões de progressão de acordes, entre outros.

4.3 Considerações Finais

Nesse capítulo realizamos a modelagem do problema da classificação de acordes dado um conjunto de notas musicais, e propomos um algoritmo para a sua resolução. Dessa forma, possibilitamos a identificação da sequência de acordes mais provável para a reprodução de uma música. No próximo capítulo apresentaremos os resultados obtidos através da junção da transformada Q-constante e do algoritmo Viterbi para implementação da transcrição automática de acordes.

Capítulo 5

Resultados Experimentais

Nesta seção abordaremos os resultados obtidos através do modelo computacional proposto para realizar a transcrição musical.

5.1 Tecnologias utilizadas

Foi utilizada a linguagem de programação python na sua versão 3.7.11, com auxílio das bibliotecas, Matplotlib, para gerar os gráficos e espectrogramas dos sinais analisados, Numpy, para facilitar a criação de funções, series, e por possuir implementações da FFT, Librosa, para realizar os cálculos da STFT e CQT. A Librosa também possui alguns métodos que auxiliaram na implementação do algoritmo Viterbi.

Todos os códigos desenvolvidos nesse trabalho estão disponível em:

https://colab.research.google.com/drive/191W_WSUCyg5gha1Md6iy7oN-eN4Elb-T#scrollTo=z7WVukHLcK2J.

5.2 Transformada de Fourier X Transformada Q constante

Realizamos um processamento do sinal de um arquivo de áudio contendo todos os acordes maiores, produzidos por um violão, começando de Dó maior até Si maior, e plotamos os resultados em um espectrograma para verificar as notas encontradas durante a execução de cada acorde. Realizamos o processamento do sinal com STFT e posteriormente com CQT e obtivemos o resultado apresentado na figura 5.1.

Ambos espectrogramas apresentam as notas identificadas através da coloração mais clara. Logo as regiões mais escuras são as notas que não estavam presente no sinal naquele instante no tempo, algumas regiões com a coloração em tons médios são causadas por harmônicos e ruídos em sua maior parte.

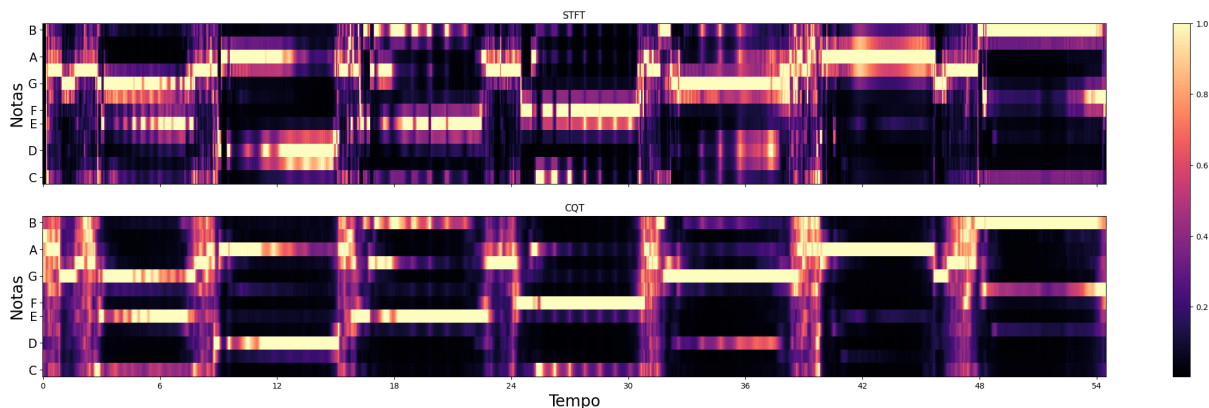


Figura 5.1: Espectrograma do processamento de acordes maiores com STFT x CQT. No eixo y do espectrograma temos a distribuição das 12 notas musicais. **Fonte autoral.**

Como esperado, a transformada Q constante nos fornece informações mais precisas para todos os acordes, devido à janela dinâmica, mantendo a qualidade da resolução. Como podemos observar nos espectrogramas, as notas tônicas de cada acorde são mais destacadas, com maior amplitude, e as transições entre os acordes geram uma região onde o processamento de sinal não consegue identificar com exatidão as notas musicais, isso ocorre por que a troca de uma acorde para outro é normalmente feita de forma muito rápida, logo nesse instante no tempo, temos a presença de várias notas de ambos os acordes, além do ruído causado durante a mudança de um acorde para o outro, logo não é possível obter informações precisas nesses momentos.

5.3 Classificação de acordes com algoritmo Viterbi

Na última parte da transcrição foi realizado a classificação de acordes através do algoritmo Viterbi, fornecendo as notas encontradas em cada janela de amostragem. No primeiro caso de estudo dessa seção foi fornecido o mesmo arquivo de áudio com todos os acordes maiores, e utilizamos a CQT para processar o sinal, com o algoritmo de Viterbi conseguimos os resultados apresentados na figura 5.2.

O espectrograma apresenta a probabilidade de cada acorde em cada instante do tempo, marcando de rosa o acorde com maior probabilidade de ocorrência, e, em geral, com tonalidades mais claras proporcionalmente a probabilidade de ocorrência de cada um dos acordes.

Como é possível observar, o algoritmo de Viterbi conseguiu identificar os acordes assim como esperado. E como observamos na seção anterior, durante as mudanças de acorde temos uma região onde é gerado bastante ruído, sendo possível observar, nessas regiões, que o algoritmo tem uma certa “indecisão” ao calcular o acorde com maior probabilidade de ocorrência.

Posteriormente, executamos o processamento de sinal e a classificação de acordes em

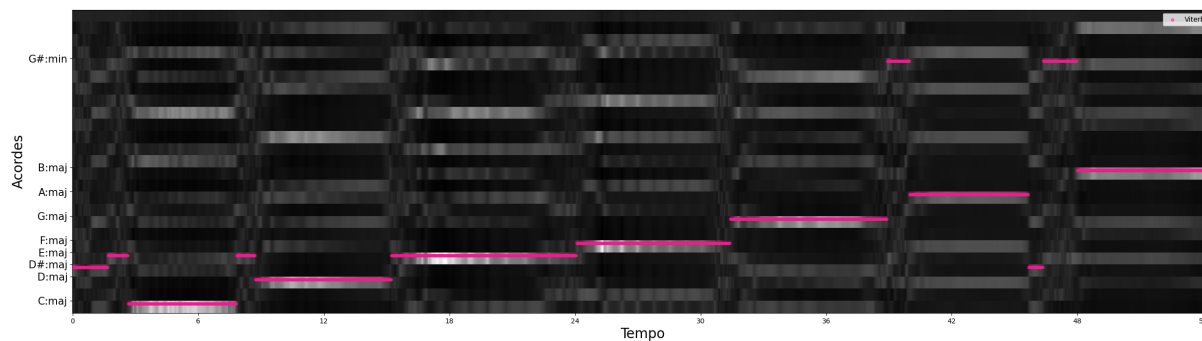


Figura 5.2: Espectrograma com os resultados do algoritmo de Viterbi para os acordes maiores. No eixo y da figura, temos a representação de cada acorde maior e menor. Sendo exibidos somente os que foram identificados na transcrição. **Fonte autoral.**

um arquivo de áudio de uma música, Vento no litoral de Legião Urbana, nesse contexto temos alguns pontos que dificultam a obtenção bons resultados, como estamos processando o sinal de áudio de uma música, teremos vários instrumentos produzindo sons ao mesmo tempo, por esse motivo utilizamos um filtro passa faixa, para eliminar frequências fora do intervalo que um violão pode gerar, é aconselhável executar esse processo para cada instrumento que deseja realizar a transcrição.

Outro fator que dificulta os resultados, é que essa música usa acordes dissonantes, mesmo assim gostaríamos de verificar como o algoritmo de Viterbi iria classificar esses acordes, o resultado está presentes nas figuras 5.3, 5.4.

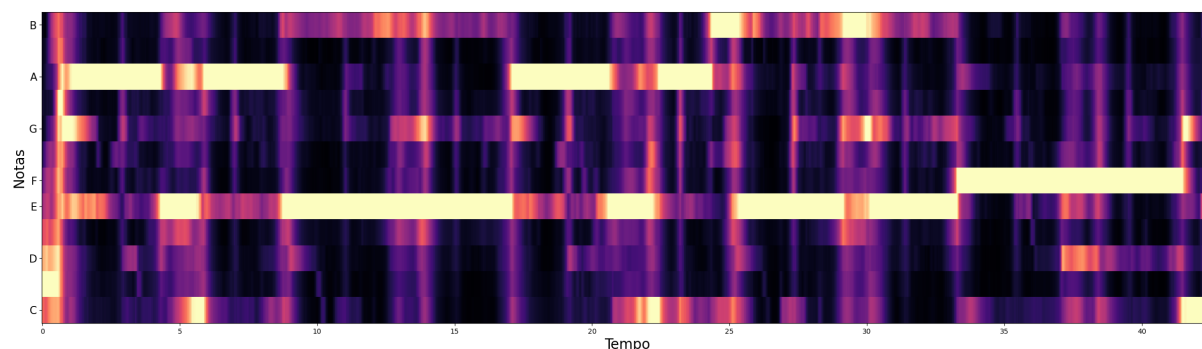


Figura 5.3: Espectrograma das notas de introdução da música Vento no Litoral. **Fonte autoral.**

Os acordes esperados, de acordo com o site Cifras Club seriam: Lá menor(Am), Mi menor(Em), Lá menor(Am), Mi menor(Em), Fá(F), e Dó(C), até o trecho recortado para essa análise. Como podemos observar no segundo espectrograma, nos 4 primeiros acordes tivemos 100% de exatidão, porém, ao classificar o acorde F, o Viterbi teve uma dificuldade, isso ocorre por que, na versão da música utilizada para essa análise(arquivo de áudio), o acorde utilizado nesse momento é o Fá com 6.^o, que seria um acorde dissonante, a formação desse acorde é dado pelas seguintes notas musicais: Fá(F), Lá(A), Dó(C), Ré(D); Essa formação se aproxima muito da formação do acorde de Ré menor(Dm): Ré(D),

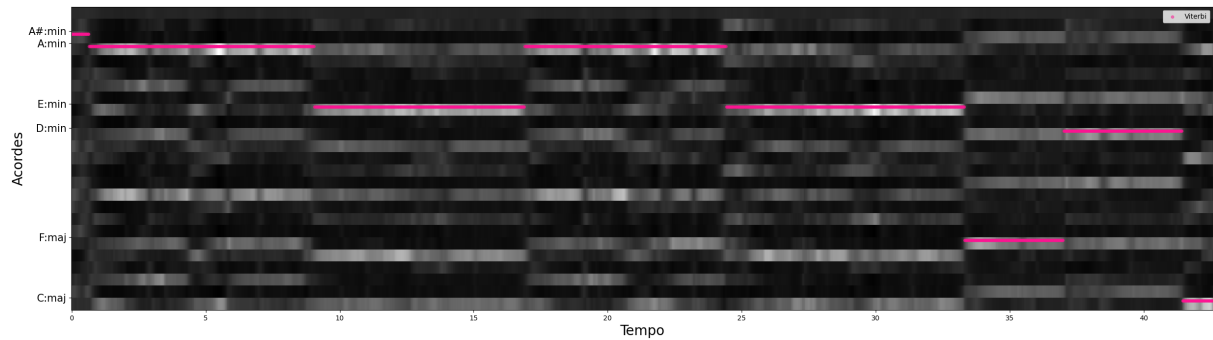


Figura 5.4: Espectrograma dos acordes de introdução da música Vento no Litoral. **Fonte autoral.**

Fá(F), Lá(A); E por isso que o Viterbi acaba em um segundo momento apresentando o acorde de Ré menor no lugar de Fá, provavelmente nesse momento a nota Dó do acorde já estava com amplitude baixa assim o Viterbi calculou com maior probabilidade de ocorrência o acorde de Ré menor. Para corrigir esse problema seria necessário implementar o algoritmo de Viterbi para identificar acordes dissonantes, mas para isso passaríamos de 24 possibilidades, acordes maiores e menores, para mais de 500 possibilidades de acordes, e essa implementação não é objeto de estudo desse trabalho.

No terceiro caso de estudo, utilizamos um trecho da música O Sol da banda Jota Quest, essa música tem uma particularidade interessante, durante o trecho analisado nenhum instrumento executa os acordes base. Tanto o instrumento contrabaixo quanto a guitarra, executam uma progressão de notas, que neste caso, serve como base para a melodia.

Para executar a melodia da música somente com acordes, segundo o site Cifras Club, seriam necessários os acordes Lá(A), Mi(E), Sol(G) e Ré(D).

Novamente utilizamos o filtro passa faixa, para eliminar frequências fora do intervalo desejado, e aplicamos a CQT junto ao algoritmo Viterbi e obtemos o resultado apresentado na figura 5.5.

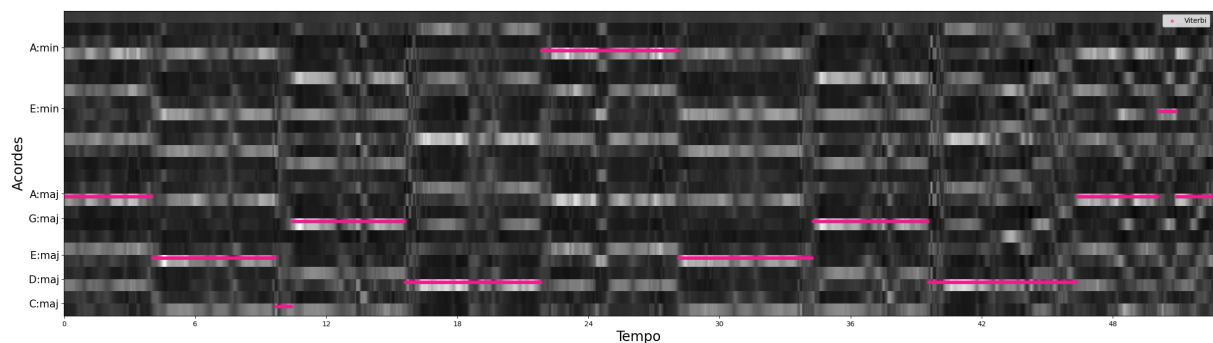


Figura 5.5: Espectrograma dos acordes da primeira parte da música O Sol do Jota Quest. **Fonte autoral.**

Como é possível observar na figura 5.5, durante a execução do segundo acorde de Lá

maior, mais ou menos no tempo 24, o modelo apresentou, equivocadamente, o acorde de Lá menor(Am). Por se tratarem de acordes muito semelhantes, onde diferem somente pelas notas de Dó sustenido(C#), para o acorde de Lá maior, e Dó(C), para o acorde de Lá menor, é um erro esperado. Vale ressaltar a particularidade dessa música, onde os acordes base não estão presentes na versão analisada, e mesmo assim o modelo apresentou um resultado satisfatório.

Mesmo com os problemas encontrados, chegamos a um resultado de 93,8% de acerto em média. Com as informações obtidas através do processamento de sinal e classificação, foi possível obter os acordes necessários para reproduzir os trechos das músicas estudadas, sendo necessários pequenos ajustes, onde um músico experiente atuaria.

Capítulo 6

Conclusão

Por fim, foi possível concluir com o presente estudo que, para o processamento de sinais provindos de instrumentos musicais, sinais não estacionários, a transformada Q constante se mostrou mais eficiente do que a transformada de Fourier de tempo curto, devido ao cálculo dinâmico para o tamanho da janela de amostragem, garantindo assim uma boa resolução do sinal no domínio da frequência e no domínio do tempo para todo o sinal.

Mesmo com a transformada Q constante, em muitos momentos houve a identificação de frequências que geravam notas musicais que não faziam parte do acorde em execução naquele instante do tempo, para contornar essa situação o algoritmo de Viterbi se mostrou bastante eficiente em calcular o acorde com maior probabilidade de ocorrência, conforme as notas encontradas no processamento do sinal, obtendo uma taxa de acerto bastante satisfatória.

Vale ressaltar que fatores como o ritmo da música e a presença de vários instrumentos podem impactar os resultados. Entretanto, podemos minimizar esses impactos através de alguns ajustes na transformada Q constante, diminuindo a janela de amostragem para aumentar a resolução no domínio do tempo e utilizando filtros passa-faixa para separar a parte do sinal referente ao instrumento analisado.

Embora este estudo tenha alcançado seus objetivos, estudos futuros podem utilizar técnicas de rede neural, como o Multilayer Perceptron, para melhorar a fase de classificação de acordes. Além disso, é importante abordar estudos com acordes dissonantes.

Podendo assim concluir o processo de transcrição musical, de um sinal de áudio para os acordes que devem ser executados para reproduzir a música em algum instrumento musical.

Bibliografia

- [Abreu, 2021] Abreu, B. P. d. (2021). Aplicações de processamento de sinais em extração de informação musical.
- [Accioly,] Accioly, A. L. P. Transcrição automática de acordes com fontes híbridas.
- [Chen et al., 2012] Chen, R., Shen, W., Srinivasamurthy, A., and Chordia, P. (2012). Chord recognition using duration-explicit hidden markov models. In *ISMIR*, pages 445–450. Citeseer.
- [Cheng et al., 2008] Cheng, H.-T., Yang, Y.-H., Lin, Y.-C., Liao, I.-B., and Chen, H. H. (2008). Automatic chord recognition for music classification and retrieval. In *2008 IEEE International Conference on Multimedia and Expo*, pages 1505–1508. IEEE.
- [Cleto et al., 2010] Cleto, P., Ferreira, R., Gomes, R., and Rodrigues, M. T. (2010). Reconhecimento de acordes musicais: Uma abordagem via perceptron multicamadas. *Mecânica Computacional*, 29(93):9169–9175.
- [Dantas and Cruz, 2018] Dantas, J. D. and Cruz, S. d. S. (2018). Um olhar físico sobre a teoria musical. *Revista Brasileira de Ensino de Física*, 41.
- [e Escalas, 2020] e Escalas, A. (2020). O acorde de ré maior. Disponível em: <https://acordescala.com/o-acorde-de-re-maior>. Acesso em: 17/10/2023.
- [Faria, 1999] Faria, N. (1999). *Acordes Arpejos E Escalas*. Irmãos Vitale.
- [Frozza, 2019] Frozza, A. C. (2019). Detecção de acordes musicais por meio de informações espectrais. B.S. thesis, Universidade Tecnológica Federal do Paraná.
- [Gusmão, 2012] Gusmão, P. (2012). Teoria elementar da música. *Centro de Artes e Letras. Universidade Federal de Santa Maria*.
- [Kothe, 2006] Kothe, I. (2006). Técnicas de análise espectral de linhas musicais. *Rio de Janeiro*, 127.
- [Oliveira, 2002] Oliveira, L. M. (2002). Frequências fundamentais das notas musicais (hz). Disponível em: <https://www.if.ufrgs.br/cref/ntef/som/oitavas.html>. Acesso em: 10/08/2023.