



Universidade Federal de Pernambuco  
Centro de Informática  
Graduação em Engenharia da Computação

## Detecção de acordes em áudio utilizando Redes Neurais

**Aluno:** Livisghton Kleber da Silva Araújo (lksa@cin.ufpe.br)

**Orientador:** Carlos Alexandre Barros de Mello (cabm@cin.ufpe.br)

**Área:** Processamento de Áudio e Aprendizagem de Máquina

13 de Setembro de 2019.

## Resumo

O reconhecimento automático de acordes musicais é uma das áreas com aplicações em recuperação da informação musical (RIM). Esta área de estudo envolve a identificação completa de cada acorde (nota, tipo e duração), bem como sua localização na música, permitindo assim, que essas informações sejam utilizadas por músicos ou por aplicações que necessitem desses dados.

O objetivo deste trabalho é apresentar uma abordagem para reconhecimento de acordes musicais baseada em redes neurais e processamento de áudio.

## 2. Introdução

A cifra é uma notação musical utilizada para indicar os acordes que serão executados por um instrumento (como piano, violão ou ukulele, por exemplo). Ela é bastante utilizada acima das letras, tablaturas ou partituras, indicando o acorde a ser tocado em conjunto a melodia [1]. É bastante comum que músicos amadores ou não profissionais não possuam a habilidade de distinguir os acordes pelos seus próprios ouvidos, fazendo-os recorrerem a sites de cifras como: cifra club <sup>1</sup>, cifras <sup>2</sup> ou pega cifra <sup>3</sup> por exemplo. No entanto, existem várias músicas que não estão cadastradas na base de dados dessas ferramentas, dificultando assim, o aprendizado dessas canções por parte dos músicos menos experientes.

Segundo VELOSO e FEITOSA [2], “...a capacidade de identificar, nomear ou produzir a frequência de um estímulo tonal sem o auxílio de um tom de referência”, caracterizam músicos que possuem “ouvido absoluto”. Similarmente, SANTOS JUNIOR [3] fala sobre “ouvido relativo” em músicos que consiste em; “...não se limita a identificar as notas, mas também as características intervalares que permitem entender a música e as relações entre as notas, acordes e melodias”. Em virtude disto, esses tipos de músicos não apresentam problemas ou dificuldades em identificar notas musicais e reproduzi-las. Por outro lado, a ocorrência de ter um ouvido absoluto é rara, de acordo com VELOSO e FEITOSA, 1 para cada 10.000 pessoas, considerando-se a população da Europa e América do Norte, sendo mais comumente observada em músicos (cerca de 20%) [2]. Já para possuir um ouvido relativo requer treino e isso leva tempo para aperfeiçoar.

Com base nisto, este trabalho tem como objetivo construir um modelo computacional que seja capaz de estimar acordes musicais dado um arquivo de música. Para isso, estão sendo estudados modelos baseados em redes neurais artificiais e técnicas de processamento de áudio para estimar os acordes de uma música. Com isso, espera-se colaborar no auxílio do aprendizado de músicos iniciantes e amadores que não apresentam nenhuma das características mencionadas anteriormente.

---

<sup>1</sup>"cifra club". <https://www.cifraclub.com.br/>. Acessado em agosto de 2019

<sup>2</sup>"cifras". <https://www.cifras.com.br/>. Acessado em agosto de 2019

<sup>3</sup>"pega cifra"<https://www.pegacifra.com.br/>. Acessado em agosto de 2019

### 3. Objetivos

O objetivo **geral** deste trabalho é construir um modelo baseado em técnicas de processamento de áudio e redes neurais que tenha a capacidade de identificar acordes musicais a partir de um arquivo áudio.

São objetivos **específicos** deste projeto:

- Aprimorar conhecimentos na área de inteligência artificial e processamentos de sinais;
- Estudar formas de integração de pesquisas em processamento de áudio e processamento de imagens;
- Entender os problemas relacionados à classificação de áudio;
- Trabalhar em temas relevantes para a área de recuperação da informação musical.

## 4. Metodologia

A metodologia adotada neste trabalho consiste basicamente em quatro fases, onde as três primeiras estão ilustradas na Figura 1 e a última, de classificação, pode ser vista na Figura 3. Estas fases foram baseadas nos trabalhos [4] [5] [6]. Além disso, a base de dados que será usado neste projeto está disponível no site isophonics [7] e contempla 200 músicas no total, onde 20 são músicas da banda Queen e 180 da banda The Beatles, sendo irrelevante essa diferença na quantidade de músicas de cada banda já que o objetivo está no reconhecimento de acordes.

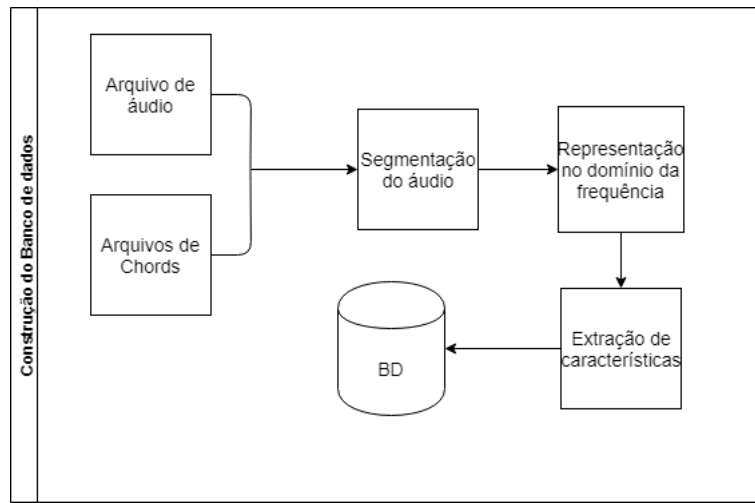


Figura 1: Três primeiras fases do projeto proposto. Fonte: O Autor.

Na primeira fase (segmentação do áudio), terá como entrada dois arquivos onde um será o arquivo de áudio e outro um arquivo de dados. Este arquivo de dados contém informações como: o tempo inicial, final para cada acorde em uma música, como mostrado na figura 2. Para cada transição de acordes ao longo de uma música, será feita a segmentação do áudio com duração especificado no arquivo de dados. Esse passo será repetido para todas as 200 músicas, obtendo-se assim  $m$  arquivos de áudio, onde cada arquivo corresponde a um acorde.

Na fase da representação no domínio da frequência, cada arquivo gerado anteriormente é um vetor de áudio *estéreo*, ou seja, possuem dois canais de áudio. Para que o processamento não seja redundante, será utilizado apenas um dos canais. Em seguida, será aplicada a Transformada de Fourier de tempo curto (*Short-Time Fourier Transform* - STFT) para representar o sinal no domínio da frequência. As configurações da transformada serão definidas no desenvolver do trabalho e de acordo com as usadas na literatura da área.

Na fase de extração de características planeja-se usar uma representação em *chromagram*, que de acordo com M. McVicar[5], é a principal forma para representar o áudio em problemas de detecção automática de acordes. Esta técnica foi introduzida por Fujishima [8] neste tipo de problema. Segundo L. Oudre,

0.000	1.143	N
1.143	12.275	A
12.275	15.886	A
15.886	19.540	B:9
19.540	23.220	D:maj7
23.220	26.922	A
26.922	30.616	A
30.616	34.307	B:9
34.307	37.988	D:maj7
...		

Figura 2: trecho de um arquivo de dados, mostrando o tempo de início (primeira coluna), tempo de fim (segunda coluna) e o acorde (terceira coluna). Fonte: O Autor.

Y. Grenier e C. Fevotte[9], uma nota pode ser dividida em duas perspectivas diferentes: a altura correspondente a oitava na qual a nota pertence e a classe de croma (ou *chroma*), onde é feita a relação com outras notas entre uma oitava. Por exemplo, as notas C4 (261 Hz) e A4 (440 Hz), podem ser decompostas em uma oitava de número 4 com cromas C e A respectivamente. Já um acorde, pode ser definido por no mínimo três notas, onde a primeira é a nota fundamental, a nota mais grave e a que define o nome do acorde; a terceira é chamada de nota modal, que determina se o acorde é menor ou maior; e a quinta, [10]. Há varias representações de modelos de acordes que refletem a estrutura de altura e croma [11] [12]. No entanto, vamos utilizar neste projeto uma representação simples para os acordes. Será criado um vetor de 12 posições, onde cada posição representa a probabilidade de uma nota. Por exemplo, {0.85, 0.2, 0.3, ..., 0.4}. Em seguida, será armazenado em um arquivo texto todos os vetores gerados nesses processos mencionado acima, finalizando assim, todo processo da preparação do banco de dados.

Por fim, na fase de classificação, inicialmente será aplicada uma rede perceptron multicamadas (*Multi-Layer Perceptron* - MLP) simples para fins de verificação da abordagem. Caso os resultados sejam promissores, serão realizados testes com outras variações de redes neurais com o objetivo de obter uma maior taxa de acerto na estimativa dos acordes.

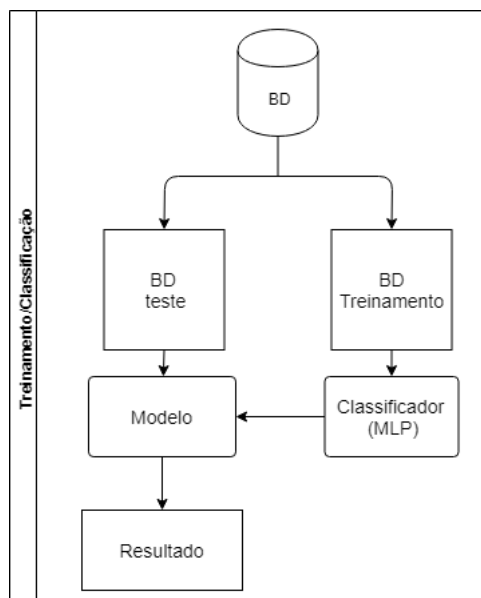


Figura 3: Fase de classificação e treinamento dos acordes - quarta fase do projeto proposto.

## 5. Cronograma

Atividade	Ago	Set	Out	Nov	Dez
Revisão bibliográfica	X				
Implementação	X	X			
Simulação			X		
Análise dos resultados				X	
Escrita da monografia	X	X	X	X	
Defesa					X



## Referências

- [1] *cifra definição de cifra em música*, [https://pt.wikipedia.org/wiki/Cifra\\_\(música\)](https://pt.wikipedia.org/wiki/Cifra_(música)), Accessed: 2019-09-7.
- [2] F.-z. Veloso e M. Ã. G. Feitosa, “O Ouvido Absoluto: bases neurocognitivas e perspectivas”, pt, *Psico-USF*, vol. 18, pp. 357–362, dez. de 2013, ISSN: 1413-8271. endereço: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-82712013000300002&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-82712013000300002&nrm=iso).
- [3] P. S. Junior, *Ouvido absoluto e ouvido relativo: sua natureza e relevância para a educação musical*, 2014. endereço: [http://www.abemeducacaomusical.com.br/conferencias/index.php/regional\\_sul/regional\\_sul/paper/view/468](http://www.abemeducacaomusical.com.br/conferencias/index.php/regional_sul/regional_sul/paper/view/468).
- [4] Yu-Lun Hsu, Chi-Po Lin, Bo-Chen Lin, Hsu-Chan Kuo, Wen-Huang Cheng e Min-Chun Hu, “DeepSheet: A sheet music generator based on deep learning”, em *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, jul. de 2017, pp. 285–290. DOI: 10.1109/ICMEW.2017.8026272.
- [5] M. McVicar, R. Santos-Rodríguez, Y. Ni e T. D. Bie, “Automatic Chord Estimation from Audio: A Review of the State of the Art”, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, n° 2, pp. 556–575, fev. de 2014, ISSN: 2329-9290. DOI: 10.1109/TASLP.2013.2294580.
- [6] J. P. d. Santana Neto. (2015). Solução computacional para reconhecimento de harmonias musicais Música - solução computacional para reconhecimento de harmonias, endereço: <http://bdm.unb.br/handle/10483/11336> (acesso em 09/09/2010).
- [7] *isophonics Conjunto de dados*, <http://www.isophonics.net/datasets>, Acessado: 2019-09-03.
- [8] T. Fujishima, “Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music”, em *ICMC*, 1999.
- [9] L. Oudre, Y. Grenier e C. Fevotte, “Chord Recognition by Fitting Rescaled Chroma Vectors to Chord Templates”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, n° 7, pp. 2222–2233, set. de 2011, ISSN: 1558-7916. DOI: 10.1109/TASL.2011.2139205.
- [10] *Acorde definição de acordes*, <https://pt.wikipedia.org/wiki/Acorde>, Accessed: 2019-09-5.
- [11] E. Gómez, “Tonal Description of Polyphonic Audio for Music Content Processing”, *INFORMS J. on Computing*, vol. 18, n° 3, pp. 294–304, jan. de 2006, ISSN: 1526-5528. DOI: 10.1287/ijoc.1040.0126. endereço: <http://dx.doi.org/10.1287/ijoc.1040.0126>.
- [12] H. Papadopoulos e G. Peeters, “Large-Scale Study of Chord Estimation Algorithms Based on Chroma Representation and HMM”, em *2007 International Workshop on Content-Based Multimedia Indexing*, jun. de 2007, pp. 53–60. DOI: 10.1109/CBMI.2007.385392.

## 6. Possíveis Avaliadores

São possíveis avaliadores do trabalho os professores:

- Tsang Ing Ren (tir@cin.ufpe.br)
- Geber Lisboa Ramalho (glr@cin.ufpe.br)
- Giordano Ribeiro Eulalio Cabral (grec@cin.ufpe.br)

## 7. Assinaturas

---

Carlos Alexandre Barros de Mello  
**Orientador**

---

Livisghton Kleber da Silva Araújo  
**Orientando**