

## RECONHECIMENTO DE ACORDES MUSICAIS: UMA ABORDAGEM VIA PERCEPTRON MULTICAMADAS

**Pedro Cleto<sup>a</sup>, Renato Ferreira<sup>a</sup>, Rogério Gomes<sup>a</sup> e Marco Túlio Rodrigues<sup>b</sup>**

<sup>a</sup>*Departamento de Computação, Centro Federal de Educação Tecnológica de Minas Gerais, Av. Amazonas 7675, Nova Gameleira, 30180-000, Belo Horizonte, MG, Brasil, rogerio@decom.cefetmg.br, <http://www.decom.cefetmg.br/>*

<sup>b</sup>*Departamento de Recursos em Informática, Centro Federal de Educação Tecnológica de Minas Gerais, Av. Amazonas 5253, Nova Suíça, 30480-000, Belo Horizonte, MG, Brasil, tulio@dri.cefetmg.br, <http://www.cefetmg.br/>*

**Palavras Chave:** Transcrição musical, reconhecimento de padrões, acordes, *perceptron* multicamadas.

**Resumo.** Transcrição Musical Automática (TMA) é o processo de obtenção de uma representação legível de uma música a partir de sua representação audível realizado por meio de um sistema informatizado capaz de realizar essa transformação. A TMA é um tema de constante pesquisa que envolve diversas áreas do conhecimento, tais como a matemática, a física e a tecnologia da informação. Um dos grandes desafios enfrentados por essa área tem sido o reconhecimento de acordes, ou seja, notas musicais tocadas simultaneamente. Este trabalho explora esse problema por meio de uma conhecida técnica utilizada em redes neurais artificiais chamada *Multilayer Perceptron* e foi validada utilizando uma base de dados criada com vários acordes produzidos com diferentes timbres por um sintetizador. A metodologia mostrou-se bastante eficiente alcançando um taxa média de reconhecimento de acordes de 84%.

## 1 INTRODUÇÃO

O uso do computador na obtenção de uma representação escrita de uma música tem diversas utilidades, entre elas a de ajudar no aprendizado de músicos, a de criar aplicativos de produção musical, além de registrar a partitura de músicas sem a necessidade de ser escrita pelo próprio compositor (Diniz, 2009). Diversos sistemas têm sido desenvolvidos com esse objetivo (Diniz, 2009), entretanto, por se tratar de uma tarefa extremamente complexa, a qualidade obtida na transcrição de partituras ainda é bastante limitada. Sendo assim, a maioria das aplicações desenvolvidas não substitui a transcrição feita por humanos e normalmente servem apenas como auxílio ao profissional da área de música.

Para realizar a transcrição automática de uma música vários aspectos devem ser detectados e analisados pelo software: o tempo, que é basicamente a velocidade na qual a música é tocada; os instrumentos utilizados, caracterizados pelos seus timbres; e a altura (frequência), duração e volume de cada nota tocada (Ferreira, 2006). O timbre e a altura das notas se tornam características especialmente complexas de serem identificadas principalmente no caso de músicas polifônicas nas quais várias notas são possivelmente tocadas ao mesmo tempo por diversos instrumentos.

Este trabalho pretende contribuir com esse ramo da computação especificamente no aspecto da identificação da altura das notas tocadas simultâneas (acordes), independente do instrumento utilizado, ao explorar a capacidade de reconhecimento de padrões das redes neurais artificiais. Acordes são, de forma sucinta, várias notas tocadas ao mesmo tempo (Ferreira, 2006). As diferentes variações existentes de um mesmo acorde e as irregularidades das ondas sonoras criadas pelo timbre tornam o problema não trivial, o que justifica o uso de técnicas oriundas da inteligência computacional nessa tarefa como pode ser visto em (Marolt, 1999).

Dessa forma, esse trabalho tem por objetivo o desenvolvimento de um sistema informatizado automático baseado em redes neurais artificiais que seja capaz de identificar os acordes presentes em arquivos de músicas polifônicas previamente gerados. Basicamente, o sistema deverá receber como entrada um arquivo de som contendo um ou mais acordes tocados individualmente, isto é, sem sobreposição de acordes com outros acordes, melodia, percussão ou voz. Como resultado final o sistema deverá fornecer a cifra correspondente ao acorde identificado.

O sistema desenvolvido foi validado por uma série de testes produzidos com diversos acordes e diferentes instrumentos.

## 2 MÉTODO

### 2.1 Os acordes

Os acordes podem ser representados de diversas formas dependendo da notação musical adotada. A notação utilizada nesse trabalho será a notação por cifras textuais e pode ser vista no esquema apresentado na [Tabela 1](#). É importante observar que

somente serão considerados nesse trabalho os acordes maiores naturais e sustentidos de forma a simplificar o campo de pesquisa.

## 2.2 Análise em frequência

A nota musical representada por uma determinada onda é caracterizada, basicamente, pela sua frequência (Ogasawara, 2008) como pode ser observado nas figuras 1 e 2. A [Figura 2](#) mostra como o timbre do instrumento torna a onda irregular em comparação com a onda senoidal correspondente. Sendo assim, torna-se necessária a representação da onda senoidal em outro domínio de forma a identificar as principais componentes que caracterizam essa onda. Nesse trabalho foi aplicada a transformada de Fourier com o objetivo de representar a onda no domínio da frequência.

O matemático francês J. Fourier provou matematicamente que qualquer forma de onda, independente da sua origem, é um somatório de ondas senoidais de diferentes frequências, amplitudes e fases (Gray and Goodman, 1995). Ele mostrou que se a forma de onda se repete periodicamente, então as frequências das componentes senoidais são restritas a valores múltiplos da frequência de repetição da forma de onda. Assim, por meio da transformada de Fourier é possível obter um vetor de amplitudes para cada frequência que compõe a onda sonora original. A [Figura 3](#) mostra como a transformada de Fourier permite destacar as notas que compõem o acorde.

Nota	Acordes	
	Naturais	Sustenidos
Dó	C	C#
Ré	D	D#
Mi	E	-
Fá	F	F#
Sol	G	G#
Lá	A	A#
Si	B	-

Tabela 1: Nomes de acordes na notação por cifras textuais

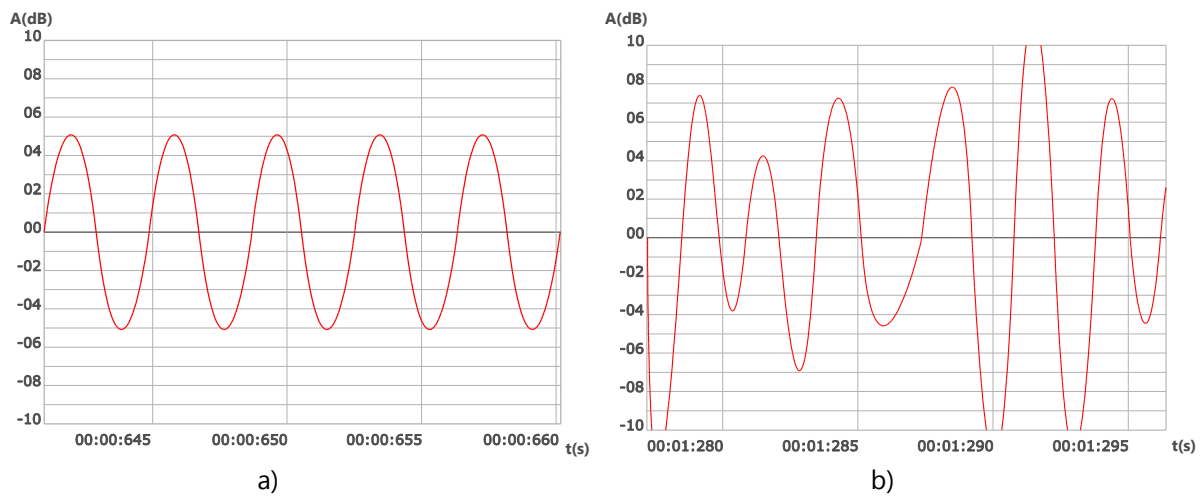


Figura 1: a) Nota Dó formada por uma onda senoidal; b) Acorde C formado por ondas senoidais

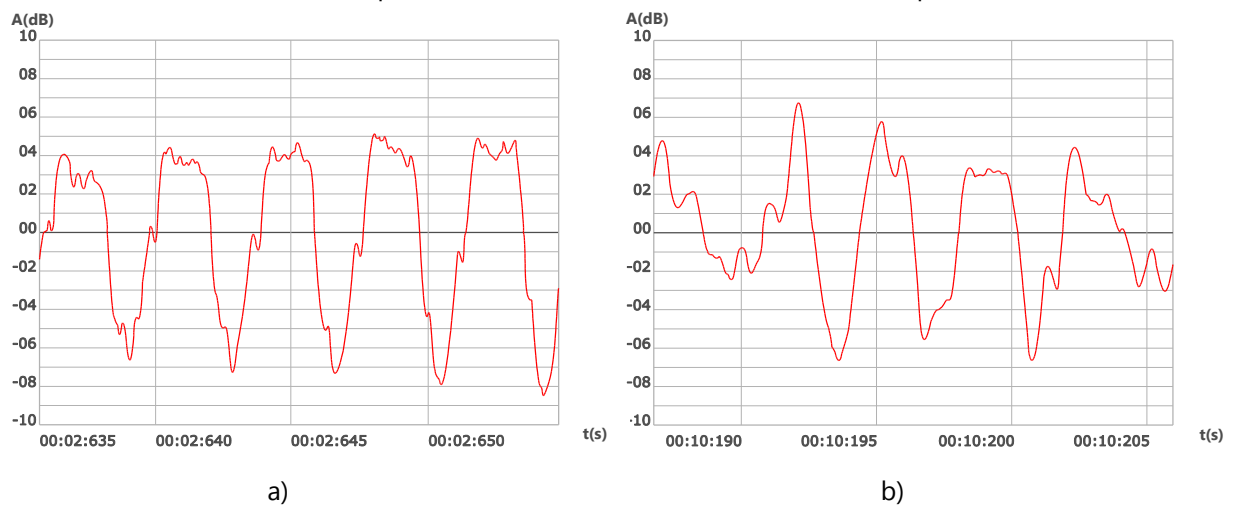


Figura 2: Nota Dó formada por um piano; b) Acorde C formado por um piano

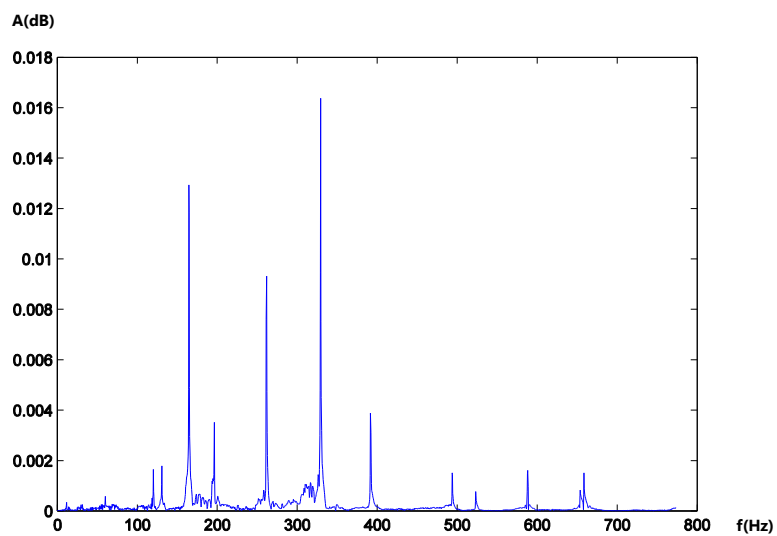


Figura 3: Transformada de Fourier de um acorde C

## 2.3 Modelagem em redes neurais

Um perceptron multicamadas (*Multilayer Perceptron* - MLP) é um modelo de rede neural artificial *feedforward* que mapeia conjuntos de dados de entrada para um conjunto de saída apropriada (Haykin, 1999). O MLP é composto por várias camadas de nós em um grafo direcionado sendo totalmente ligado a partir de uma camada para a próxima. Exceto para os nós de entrada, cada nó é um neurônio (ou elemento de processamento) com uma função de ativação não-linear. Um MLP utiliza uma técnica chamada de aprendizado supervisionado *backpropagation* para treinamento da rede (Rosenblatt, 1961) (Rumelhart et al., 1986). MLP é uma modificação do perceptron linear, que pode distinguir os dados que não são linearmente separáveis (Cybenko, 1989).

O problema de reconhecimento de acordes proposto nesse trabalho pode ser modelado para ser usado em redes MLP aplicando como entrada as amplitudes obtidas pela transformada de Fourier e como saídas os valores que representam os acordes.

## 3 RESULTADOS EXPERIMENTAIS

Para obter os sons utilizados no trabalho, foi usado um teclado musical arranjador Yamaha PSR-E413 de 61 teclas que sintetiza sons de vários instrumentos. Foram escolhidos para produzir os registros sonoros utilizados como entrada do sistema proposto, os seguintes instrumentos: Piano, Cravo, Órgão e Violão.

Na etapa de treinamento do sistema 144 amostras de acordes foram gravadas com as seguintes características:

- Para cada um dos 12 acordes citados na [Tabela 1](#) foram escolhidas 3 variações (formas diferentes de se tocar um mesmo acorde), totalizando 36 combinações;
- Para cada um dos 4 instrumentos escolhidos foram gravadas as 36 combinações, totalizando 144 amostras.

O espectro de frequências foi dividido em faixas correspondentes a cada uma das 61 notas musicais que foram consideradas no trabalho (foram consideradas apenas as notas das 61 teclas do teclado). A frequência de maior amplitude de cada faixa, obtida por meio da transformada de Fourier, foi coletada e utilizada como variável de entrada no MLP configurado com 61 nós na sua camada oculta, conforme [Figura 4](#).

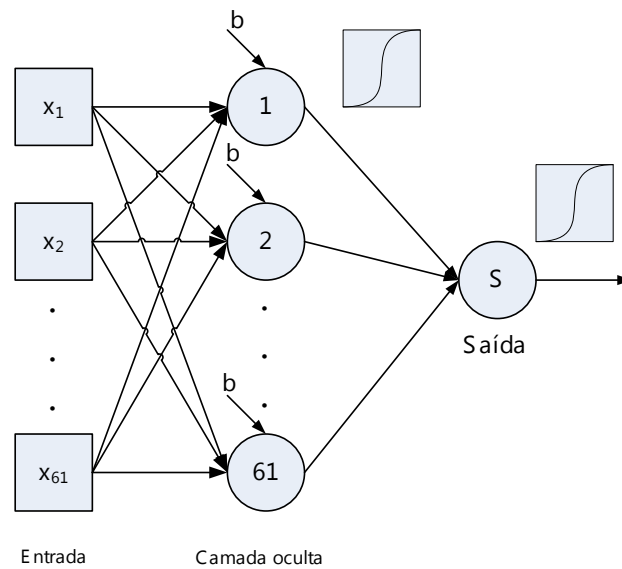


Figura 4: Arquitetura da rede usada

Na etapa de validação do sistema 90 novas amostras de acordes foram gravadas com as seguintes características:

- Teste 1: Cada um dos 12 acordes tocados com o mesmo timbre (Piano);
- Teste 2: O acorde C com 11 timbres não usados no treinamento;
- Teste 3: Três variações do acorde C com 4 timbres semelhantes aos treinados;
- Teste 4: Os 12 acordes com cada um dos 4 timbres usados no treinamento;
- Teste 5: Sete acordes tocados em um violão real.

Observou-se que o treinamento da rede realizado com as 144 amostras de treinamento obtidas convergiu por volta da iteração 300 e que o erro quadrático médio, após 1000 iterações, atingiu um valor de 0,00061.

Os resultados dos testes são apresentados na [Tabela 2](#) e é possível verificar que o timbre influencia bastante no reconhecimento já que os testes realizados com timbres iguais aos usados no treinamento (testes 1 e 4) foram os que tiveram a maior taxa de acerto.

	Número de acordes	Número de acordes não reconhecidos	Taxa de acerto
Teste 1	12	0	100%
Teste 2	11	3	73%
Teste 3	12	4	67%
Teste 4	48	5	90%
Teste 5	7	2	71%
Total	90	14	84%

Tabela 2: Resultados dos testes com a rede neural

## 4 CONCLUSÃO

Este artigo apresenta uma arquitetura de rede neural tradicional, o perceptron multicamadas, no reconhecimento de acordes. Os resultados obtidos revelam que essa metodologia pode ser usada com sucesso, proporcionando taxas de reconhecimento médio de 84%. Dessa forma, apesar da grande quantidade de limitações impostas, dentre elas o baixo número de acordes (12) e instrumentos (4) considerados e o baixo número de amostras utilizadas no treinamento (144), podemos considerar que os resultados foram promissores.

Como trabalhos futuros, pretende-se investigar a eficácia do sistema com a utilização de uma base de treinamento e validação mais representativa do problema, além da incorporação de métodos de treinamento mais efetivos aplicados às redes MLP.

## REFERÊNCIAS

- Cybenko G. Approximations by superpositions of a sigmoidal function. Technical Report CSR-856, Center for Supercomputing Research and Development, University of Illinois, 1989.
- da Costa Beltrão Diniz F.C. *Transcrição Musical Automática Usando Representação Freqüencial Eficiente por Banco de Filtros de Alta Seletividade*. Tese de doutorado em engenharia elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.
- Gray R. and Goodman J. *Fourier Transforms: An introduction for engineers*. Kluwer Academic Publishers, Boston, 1995.
- de Souza Ferreira S.A. *Sistema Especialista Para Reconhecimento De Acordes Musicais Em Tempo Real Para Violão Elétrico Utilizando Técnicas de DSP*. Dissertação de mestrado em engenharia elétrica, Universidade Federal da Bahia, Bahia, 2006.
- Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.
- Marolt M. A comparison of feed forward neural network architectures for piano music transcription. 2002.
- Ogasawara A.S. *Reconhecedor de Notas Musicais em Sons Polifônicos*. Trabalho de conclusão de curso, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- Rosenblatt F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington DC, 1961.
- Rumelhart D.E., Hinton G.E., and Williams R.J. Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland, and the PDP research group., editors, *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations*. MIT Press, 1986.