

---

## ANALISANDO OS PROGRAMAS DE RECONHECIMENTO DE MÚSICAS

Diego Bezerra de Santana;  
E-mail: diego\_bk\_201@hotmail.com.  
Valdick B. de Sales Júnior;  
E-mail: valdicksales@gmail.com.

### Resumo

A música nos dias de hoje, graças ao avanço da tecnologia, vem ganhando uma dimensão nunca antes vista. Isto está acontecendo, principalmente, por causa da internet que permite que as pessoas ouçam músicas em formato digital, em qualquer lugar e a qualquer momento; basta ter um celular, um aparelho mp3 ou um computador para ouvir suas músicas preferidas. O crescente uso desse formato de áudio pelas pessoas tem levado muitas empresas a investir nesse ramo, que tem se mostrado altamente lucrativo e, por conta disso, cada vez mais tem surgido serviços de *streaming* de música como: o spotify, o deezer, e o rdio, e programas de reconhecimento de músicas. Esses últimos ainda não atingiram a popularidade que os serviços de *streaming* possuem. Entretanto, eles chamam muito a atenção de estudiosos da área da computação por conta da principal funcionalidade desses programas: a de reconhecer músicas. Fazer com que um computador ou programa seja capaz de identificar músicas é um processo extremamente trabalhoso e complexo que exige do programador um vasto conhecimento matemático, de processamento e armazenamento de dados, de reconhecimentos de padrões, além de ter um bom conhecimento de física. Diante disso, este trabalho irá analisar como funcionam os sistemas de reconhecimento de áudio para que os desenvolvedores que planejam trabalhar na área saibam como é feita a identificação de músicas por um computador, como também saibam identificar quais são os principais *softwares* de reconhecimento de música e quais são suas funcionalidades para que no futuro elas possam ser aprimoradas, para que esses programas apresentem novas funcionalidades.

**Palavras-chaves:** Padrões classificadores; processamento; música digital; reconhecimento de sons.

### Abstract

Music today, thanks to the advancement of technology, has been gaining a dimension never seen before. This is happening mainly because of the internet that allows people to listen to music in digital format, anywhere and anytime; just have a cell phone, an mp3 player or a computer to listen to your favorite songs. The increasing use of this audio format by people has led many companies to invest in this branch, which has been highly profitable and, because of this, more and more music streaming services such as: spotify, deezer, and radio, and music recognition programs. The latter have not yet reached the popularity streaming services have. However, they draw much attention from computer scientists because of the main functionality of these programs: that of recognizing music. Making a computer or program capable of identifying music is an extremely laborious and complex process that requires the programmer to have extensive mathematical knowledge, data processing and storage, pattern recognition, and a good knowledge of physics. Therefore, this work will analyze how the audio recognition systems work so that developers who plan to work in the area know how to identify songs by a computer and also identify the main music recognition software and which are so that in the future they can be improved, as well as for these programs to present new functionalities.

**Keywords:** Classification standards; processing; digital music; recognition of sounds.

## 1. INTRODUÇÃO

A tendência da humanidade para a música é algo inato, essencial em todas as culturas e desde os primórdios da espécie humana vem se desenvolvendo intensamente (CASEY, 2004). A tecnologia vem cada vez mais se tornando uma abordagem de interação com os processos musicais (BERENZWEIG, 2001). Desde sintetizadores eletrônicos até afinadores programados em *software*, a música vem acompanhando o desenvolvimento técnico-científico.

Um exemplo do impacto direto da tecnologia sobre a música é o *software* Auto-Tune que é um editor de áudio em tempo real criado para afinar instrumentos e vozes. Muitos artistas fazem uso desse *software* para poder gravar seus álbuns. Outro exemplo do impacto da tecnologia na indústria musical são os serviços de *streaming* de música como o *spotify* que é um aplicativo para dispositivos móveis Android que permite que os usuários possam ouvir milhares de artistas dos mais diversos gêneros em qualquer lugar, desde haja *internet* no seu telefone *Smartphone* ou *notebook*. *Spotify*, *Deezer*, *Rdio* e *Apple music* utilizam formatos digitais de música como mp3, WMA, FLAC, etc., o que permite que esses serviços de *streaming* possuam uma grande base de dados musicais; o que gera uma necessidade de classificação dessas músicas.

Devido a esta grande quantidade de músicas em formatos digitais existentes nesses serviços de *streaming*, sistemas como o *Shazam* e o *Soundhound* têm ganhado cada vez mais importância, pois eles são capazes de identificar músicas, artistas, gêneros musicais e, no caso do *Shazam*, até programas de TV utilizando sistemas de reconhecimento de áudio, o que tem ajudado bastante aos fãs de música a descobrir novos artistas.

Contudo, muito desses sistemas de reconhecimento nem sempre conseguem identificar as músicas que estão tocando e, em alguns casos, levam muito tempo para identificar as músicas. Isso ocorre porque as técnicas de extração de áudio utilizadas por esses sistemas nem sempre são as mais indicadas para exercer essa função, pois extrair informações musicais como ritmo, melodia, timbre, etc., não é uma tarefa fácil; ela é bem mais complexa do que extrair apenas a voz de uma pessoa como fazem muitos sistemas de reconhecimento de voz.

Como essa é uma área ainda pouco estudada e explorada por quem trabalha na área da tecnologia, algumas técnicas e sistemas de reconhecimento musical serão estudados buscando identificar os problemas e benefícios que cada um possui para que eles possam ser melhorados no futuro, obtendo melhores resultados na busca por músicas, além de implementar funcionalidades para esses sistemas, fazendo com que eles possam ganhar maior importância no mercado.

Diante do exposto, o objetivo deste trabalho é estudar o funcionamento dos principais sistemas de reconhecimento de áudio pela necessidade crescente de investigação neste campo específico, pois cada vez mais surgem sítios especializados ou com componentes especializados na música. Dessa forma, a proposta deste artigo tem como intuito principal descrever como funcionam os sistemas de reconhecimento de música e identificar quais são os principais sistemas do mercado.

## 2. SOM DIGITAL

O som nada mais é que uma onda mecânica que percorre o ar com amplitudes e frequências variadas. Desta forma, ele é usualmente representado por curvas inconstantes que sobem e descem ao longo de um certo tempo.



Fonte: Google.

Sempre que um computador capta algum tipo de som, ele faz inúmeras amostras dessa onda contínua que caracteriza o som, criando um conjunto de diversos pontos: se os pontos ficarem bem próximos um do outro, o computador poderá refazê-lo sem grandes dificuldades. Tendo isto em mente, caso eles fiquem tão perto uns dos outros que possam formar uma reta, eles ficarão bem semelhantes com a onda de origem. Diante disso, quanto mais amostras um computador tiver, mais fácil será para tornar um som digital mais parecido com o som original.

Boa parte das músicas que escutamos hoje na internet, através do computador, do celular e nos CD's são guardadas com mais 40000 amostras de sons por segundo; em DVD, no entanto, esse número duplica. Esse grande número de amostras é necessário para obter uma boa qualidade de som e, principalmente, para que o computador possa ser capaz de refazer – de maneira exata – o som original.

### **3. EXPLICANDO COMO É FEITO O RECONHECIMENTO DE MÚSICAS POR UM COMPUTADOR**

#### **3.1. Como um Computador Consegue Identificar Sons?**

Para reconhecer um som e executar a tarefa desejada, o computador precisa reconhecer uma série de passos: primeiro ele precisa digitalizar o som que se deseja reconhecer, para isso é preciso utilizar um conversor analógico-digital que recebe as vibrações criadas pelo som e converter as ondas em dados digitais; segundo, aplicar uma medida para cada onda captada. Assim o som digitalizado será filtrado e separado de interferências e ruídos. Para isso, efetua-se uma computação das características da frequência contido som. Durante essa fase do processo o som pode precisar ser sincronizado, pois as músicas não utilizam o mesmo tom e nem sempre têm a mesma velocidade (ritmo). Portanto, deve-se realizar um ajuste com modelos de som já armazenados na memória do classificador.



Fonte: Google.

Essa digitalização deve ser separada em pequenas frações, e, em seguida, o programa vai comparar os sons captados com músicas conhecidas e presentes no seu banco de dados que correspondam ao gênero musical da música que está sendo reproduzida. De forma simplificada nada mais é do que um método de busca para associar as saídas com padrões de sons das músicas.

Em suma, o sistema vai analisar o resultado e comparar com músicas, gêneros e ritmos musicais conhecidos, finalizando com a identificação da música reproduzida e mostra o nome da música e nome do artista para o usuário.

### **3.2 Reconhecimento de Padrões Através de Padrões Classificadores**

Para sistemas de reconhecimento de padrões existem dois tipos de classificadores: um que armazena uma grande quantidade de músicas, em seu banco de dados, e os que possuem sons, frequência e ritmos restritos.

Os sistemas que possuem uma quantidade de som mais limitado, armazenados em seu banco de dados, são geralmente aplicados para a utilização de um grande número de usuários. O computador usado para isto é programado para ser utilizado de uma forma mais geral e, embora exista uma grande variedade de sons, ele possui uma boa capacidade de reconhecimento. Um bom exemplo disto são os sistemas de reconhecimento de música como: *Shazam* e *SoundHound*.

Os programas com uma biblioteca de sons mais completa são indicados para um número pequeno de usuários. O computador deste tipo de programa é mais adequado para um pequeno grupo de pessoas que irão utilizá-lo. Portanto, esse sistema deve ser treinado para adaptar-se ao som de cada música e o nível de precisão de seu reconhecimento é bem maior do que um que possua menos sons. Os programas que identificam notas musicais são um exemplo desse tipo de sistema.

#### 4. DIFICULDADES

Alguns dos problemas mais conhecidos com relação aos sistemas de reconhecimento de músicas estão relacionados com a reprodução contínua. Para as pessoas é fácil ouvir uma canção e rapidamente fazer a junção entre os sons e os ritmos. Contudo, para um computador, ele compreende de forma fácil se os sons forem reproduzidos de forma separada e pausada durante uma música. Há alguns anos essa era uma grande dificuldade para os programas fazerem reconhecimento de músicas, embora ainda hoje isso possa acontecer em certos sistemas.

Para um classificador isso demonstra uma significativa diferença e ele pode acabar identificando um som completamente diferente do que aquele que foi reproduzido. Diferentes tipos de músicas como: *rock*, *pop*, *jazz*, *blues*, *folk*, eletrônica, etc., podem modificar bastante como certos sons serão reproduzidos, fazendo com que a interpretação do sistema também seja diferente para certas músicas.

Outra dificuldade existente nesses sistemas é a de separar sons simultâneos de vários instrumentos musicais. Quando usados para criar uma cifra, eles têm problemas para identificar os sons que foram sobrepostos porque dois instrumentos foram tocados ao mesmo tempo durante uma música. Isso ocorre porque o programa precisa ouvir os sons de maneira correta para que possa distingui-las. Diversos sons reproduzidos ao mesmo tempo acabam gerando barulho em excesso e faz com que o programa perca o foco e não funcione corretamente.

Qual o motivo de tanta complexidade? O classificador necessita fazer o tratamento de sons separando-os. Para isso, deve reconhecer em que momento cada som começa e termina. A partir disso ele cria a cadeia dos sons enfileirando os

acordes e construindo os ritmos. Se às vezes até nós não conseguimos compreender o som de algum instrumento, imagine uma máquina.

## 5. SISTEMAS DE RECONHECIMENTO DE MÚSICAS

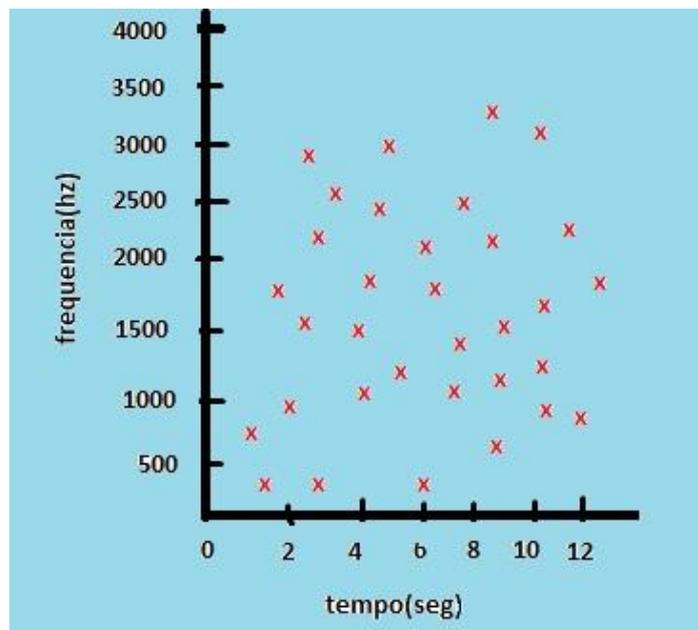


Fonte: Google.

Há alguns anos, quando ouvíamos uma música na televisão ou no rádio, era necessário aguardar o nome do artista e da música aparecer na televisão ou ser dito no rádio para então procurarmos na *internet* para poder escutar na hora em que nos desse vontade. Hoje, por causa dos sistemas de reconhecimento de áudio, isso mudou. Basta colocar o celular ou computador para o local de onde estava vindo a música e, em alguns segundos, se consegue saber o nome da música, do artista e até a letra dela.

Para realizar essa tarefa os programas de reconhecimento de música necessitam de um banco de dados extremamente vasto, e, além disso, devem ser atualizados constantemente com músicas de diversos artistas de todo globo terrestre. Esses aplicativos criam um tipo de impressão digital em gráfico 3D para identificar as músicas. Esta impressão é denominada de espectograma que é uma peça fundamental para identificar as canções.

O Ponto X mostra a duração da música e no Y é mostrada a frequência; no ponto Z é mostrada a intensidade da música. Dessa maneira, dá para converter uma canção em um gráfico escaneável pelo programa.



Fonte: Imagens do google

Dentro de uma música existem diversas frequências sonoras. Portanto, quando o programa receber os dados iniciais, irá buscar os pontos mais altos da frequência, para então iniciar o restante do gráfico que é criado através do tempo da música.

Quando o usuário ativar esses programas para reconhecer uma canção o aplicativo vai dar início à criação do gráfico 3D mencionada anteriormente e, após 10 segundos de música, o programa vai começar uma busca no banco de dados para encontrar uma música que possua exatamente os mesmos padrões de frequência e intensidade da música que está sendo reproduzida. Se por acaso a canção não for encontrada no banco de dados, ou caso a música contenha excesso de ruídos, não poderá ser reconhecida por esses aplicativos de identificação de música.

### 5.1 Shazam

O *Shazam* foi criado em 1999, mas só em 2002 o aplicativo foi disponibilizar a função de reconhecer músicas. Nessa época, apenas os ingleses tinham acesso ao

---

programa que funcionava através de uma ligação telefônica. O usuário deveria discar para um número telefônico e por o telefone ou celular próximo de onde estava sendo reproduzida a música. A ligação era feita de forma autônoma e quando o *Shazam* recebia os dados da música, a ligação era encerrada e pouco tempo depois era entregue uma mensagem ao usuário com o nome da música e do artista que estava tocando.

Entretanto, nos dias de hoje o *Shazam* é o principal sistema de reconhecimento de músicas, disponível para android, iOS e PC, e pode ser utilizado para quem quer descobrir que música está tocando colocando o computador ou celular para identificá-la, independentemente de onde a música está tocando; ela pode estar tocando em um show, no rádio ou na televisão, e o *Shazam* consegue reconhecer a música.

Além de reconhecer a música, também permite que o usuário possa entrar em contato direto com lojas virtuais de música para que possa adquirir a música que está ouvindo ou, ainda, com serviços de *streaming* de música como o *Spotify* para que possa escutá-la.

## **5.2 MusicID**

O *MusicID* é outro programa de reconhecimento de músicas. Quando uma música é reconhecida, o aplicativo carrega na tela algumas informações sobre ela e sobre o artista, como: biografia, discografia e gênero musical. Além disso, o usuário possui contato direto com maior número de informações relacionadas à música identificada, como a letra da música e vídeos do YouTube. O *musicID* está disponível para Android e iOS.

## **5.3 SoundHound**

O *SoundHound* assim como o *Shazam* consegue reconhecer uma música em qualquer lugar. Depois de reconhecer a canção, o programa oferece a letra da canção, o título da música, do álbum e nome do artista, além de outros dados do artista.

Além disso, o aplicativo mostra um mapa musical da sua região e de todo o mundo, onde aparecem outros usuários do programa que o utilizaram para identificar uma música recentemente. Para isso basta ampliar ou reduzir o mapa para ver as localizações exatas e verificar se existem pessoas próximas de você ouvindo a mesma música e artista. O aplicativo também fornece a função de procurar por localização, basta por o nome de um local que o programa irá mostrar uma lista de músicas ouvidas nessa região e, por último, o *software* permite que você geoclassifique as músicas para que elas apareçam no mapa e crie nos serviços de *streaming* como o *Spotify playlists* com as músicas identificadas pelo programa que está disponível para Android, iOS e PC.

#### **5.4 MusiXmatch**

O *musiXmatch* conta com um catálogo considerável de letras de música, sendo também uma alternativa para as pessoas que estão escutando uma música e desejam saber o nome dela e quem é o artista da música. O aplicativo ainda permite que você salve e classifique as músicas, dividindo-as da forma que o usuário achar melhor, para que possa mais tarde rever as informações sobre as músicas reconhecidas.

O programa permite que o usuário entre em contato direto com serviços *Streaming* e lojas virtuais para que possa escutar a música. O aplicativo pode ser obtido pelos usuários de Android e iOS.

#### **5.6 TrackID**

O *TrackID* é um aplicativo da Sony desenvolvido para Android que traz um sistema preciso para reconhecimento de música e ainda disponibiliza a biografia e outras canções dos artistas identificados. O programa armazena suas pesquisas em um histórico, permitindo que o usuário possa salvar as músicas reconhecidas para ouvi-las em outro momento utilizando em serviços como *Spotify*, *YouTube* e *GooglePlaymusic*.

---

## 6 CONCLUSÃO

Seguindo a linha do tempo que essa tecnologia percorreu e o quanto ela cresceu nos últimos anos, quando nos propomos a falar no aumento considerável no número e variedade de aplicações observamos que à medida que o tempo passa ela se aperfeiçoa e se amplia infinitamente mostrando que, num futuro próximo, surgirão novos programas para celular que façam uso do reconhecimento de sons.

Algo mais surpreendente será a melhoria dos programas de acessibilidade, conversão de fala em texto e interpretação de voz por atendimento eletrônico. Visualizamos, num futuro bem próximo, que poderemos ter acesso a um tradutor simultâneo, que transmute palavras de um idioma para outro e um programa que consiga criar cifras automaticamente através do reconhecimento dos acordes, do tipo de instrumento e do ritmo da música que está sendo reproduzida.

## REFERÊNCIAS

- AGOSTINI, G.; LONGARI, M. & POLLASTRI, E. 2003. “**Musical instrument timbres classification with spectral features**”, EURASIP Journal on Applied Signal Processing, 2003:1, 1-11.
- ALSABTI, K.; RANKA, S. & SINGH, V. 1998. “**An efficient K-Means clustering algorithm**”, 11th International Parallel Processing Symposium-IPPS 1998.
- BERENZWEIG, A. & ELLIS, D. 2001. “**Locating singing voice segments within music signals**”, Proceedings of Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA 2001, p. 119-123, Mohonk, NY.
- BHATTACHARJEE, S.; GOPAL, R.D.; SANDERS, G.L. **Digital music and online sharing: software piracy 2.0?** Commun. ACM, ACM, New York, NY, USA, v. 46, p. 107–111, July 2003. ISSN 0001-0782.
- BISHOP, C.M. 1995. “**Neural Networks for Pattern Recognition**”, Oxford University Press. 114 BIBLIOGRAFIA.
- BRADLEY, P.; FAYYAD, U. & Reina, C. 1998. “**Scaling clustering algorithms to large databases**”, Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining – AAI 1998, p. 9-15, NY. [Broomhead & Lowe, 1988].
- BROOMHEAD, D.S. & LOWE, D. 1988. “**Multivariable function interpolation and adaptativo networks**”, Complex Systems, v. 2, p. 321-355.

CASEY, M., et al. **Content-based music information retrieval**: Current directions and future challenges. Proceedings of the IEEE, v. 96, n. 4, p. 668-696, april, 2008. ISSN 0018-9219.

CHANG, S.F.; SIKORA, T.; PURL, A. **Overview of the mpeg-7 standard**. **Circuits and Systems for Video Technology**, IEEE Transactions on, v. 11, n. 6, p. 688-695, jun, 2001. ISSN 1051-8215.

CHELLA, M.T. **Sistema para classificação e recuperação de conteúdo multimídia baseado no padrão mpeg7**. 2004.

CHIH-CHUNG, Chang e CHIH-JEN, Lin. **LIBSVM: a library for support vector machines**, 2001. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. Acesso em: 25 de abril de 2018.

CYBENKO, G. **Approximations by superpositions of a sigmoidal function**. Technical Report CSRD-856, Center for Supercomputing Research and Development, University of Illinois, 1989.

GRAY, R. and Goodman J. **Fourier Transforms**: An introduction for engineers. Kluwer Academic Publishers, Boston, 1995.

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.