

A matemática por trás do Shazam: transformada de Fourier e impressão digital de áudio

Roberto Marlon Monteiro Braga | roberto.mmb@aluno.ifsc.edu.br¹

Edmar Sandro Santos da Silva | edmar.s25@aluno.ifsc.edu.br²

Sarah Jane Souza da Silva | sarah.souza@ifsc.ed.br³

Isnaldo Isaac Barbosa | isnaldo@pos.mat.ufal.br⁴

Trabalho de Pesquisa

DT 4 - Inovação Tecnológica, Processos Produtivos e Futuro das Tecnologias

Resumo:

Este trabalho apresenta a base matemática utilizada pelo aplicativo Shazam para reconhecimento musical. O aplicativo é amplamente utilizado em todo o mundo para identificar músicas em segundos, mesmo em ambientes ruidosos, porém poucos usuários conhecem o mecanismo matemático que o sustenta. O objetivo deste artigo é explicar, de forma acessível, como a Transformada Rápida de Fourier (FFT) converte sinais de áudio do domínio do tempo para o domínio da frequência, como o algoritmo principal de busca gera impressões digitais acústicas (áudio fingerprints) a partir de picos espectrais, e como essas impressões são armazenadas e consultadas em um banco de dados de hashes para viabilizar a identificação em tempo real. A metodologia adotada foi a revisão bibliográfica das publicações originais que fundamentam a ferramenta e de literatura técnica sobre processamento digital de sinais. Os resultados demonstram que a combinação de FFT com o mapa de constelação e indexação por hash torna o reconhecimento robusto ao ruído e eficiente em escala global, com milhões de músicas catalogadas. Conclui-se que a matemática aplicada no Shazam é um exemplo notável de como conceitos clássicos de análise de sinais transformam-se em tecnologias do cotidiano.

Palavras-chave: reconhecimento musical; transformada de Fourier; áudio fingerprint; processamento de sinais; Shazam.

Introdução

O Shazam é um aplicativo de reconhecimento musical lançado em 2002 que permite identificar uma música captando apenas alguns segundos de áudio pelo

¹ Estudante do curso superior Tecnólogo em Sistemas Embarcados do Instituto Federal de Santa Catarina.

² Estudante do curso superior em Licenciatura em Matemática do Instituto Federal de Santa Catarina.

³ Servidora da área de matemática e coordenadora do curso superior Tecnólogo em Sistemas Embarcados do Instituto Federal de Santa Catarina.

⁴ Servidor da área da matemática e diretor do Instituto de Matemática da Universidade Federal de Alagoas.

microfone de um smartphone. Em 2023, o aplicativo contava com mais de 2 bilhões de downloads e identificava mais de 50 milhões de músicas por dia (Apple, 2023), consolidando-se como uma das ferramentas de reconhecimento de áudio mais utilizadas no mundo.

A pergunta que naturalmente surge é: como um algoritmo consegue identificar uma música específica em meio a barulho de fundo, distorções e gravações de baixa qualidade, comparando-a com um catálogo de dezenas de milhões de faixas em frações de segundo? A resposta reside em uma combinação elegante de matemática clássica e engenharia de software eficiente.

Este artigo tem como objetivo desmistificar o funcionamento matemático do Shazam, explicando os conceitos de Transformada de Fourier, espectrograma, mapa de constelação e impressão digital acústica. Busca-se demonstrar como ferramentas matemáticas do século XVIII e XX são aplicadas de maneira inovadora em tecnologias do século XXI, tornando o reconhecimento musical rápido, preciso e robusto ao ruído.

Fundamentação teórica

Todo sinal de áudio é, na sua essência, uma onda complexa que representa variações de pressão ao longo do tempo. Para analisar quais frequências compõem esse sinal, utiliza-se a Transformada de Fourier, que decompõe uma função do tempo em suas componentes de frequência. A Transformada de Fourier Discreta (DFT) é definida por:

$$X(k) = \sum_{n=0}^{N-1} X(n) e^{\left(\frac{-i2\pi kn}{N}\right)}$$

onde $X(n)$ representa o sinal no domínio do tempo com N amostras, k é o índice de frequência, n é o índice temporal, e $e^{\left(\frac{-i2\pi kn}{N}\right)}$ é o núcleo complexo da transformada. O resultado $X(k)$ fornece a amplitude e a fase de cada componente de frequência presente no sinal.

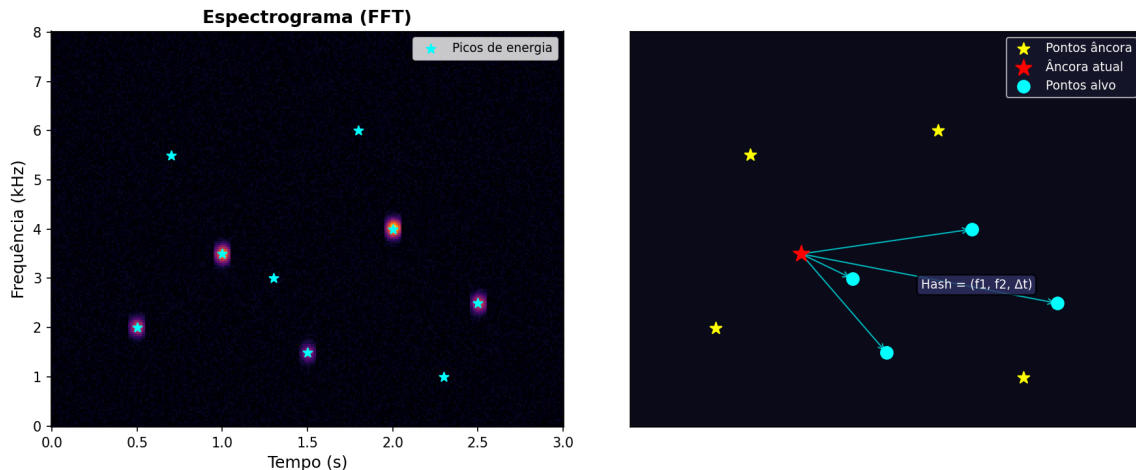
Na prática, o Shazam utiliza a Transformada Rápida de Fourier (FFT — Fast Fourier Transform), um algoritmo eficiente para calcular a DFT desenvolvido por Cooley e Tukey (1965), que reduz a complexidade computacional de $O(N^2)$ para $O(N \log N)$. Isso torna o processamento viável em tempo real em dispositivos móveis.

O espectrograma e o mapa de constelação

O Shazam não aplica a FFT ao sinal inteiro de uma vez. Ao invés disso, divide o áudio em janelas de tempo curtas e sobrepostas (tipicamente de 0,37 a 1 segundo) e aplica a FFT em cada janela. O resultado é um espectrograma: uma representação tridimensional que mostra como a energia de cada frequência varia ao longo do tempo.

A partir do espectrograma, o algoritmo identifica os picos de energia, os pontos de maior amplitude em cada faixa de frequência e janela de tempo. Esses picos formam o que Wang (2003) denomina "mapa de constelação", uma representação esparsa e robusta do conteúdo musical, análoga a um mapa estelar onde cada estrela é um ponto de energia relevante. Na Figura 1 podemos observar o espectrograma com picos de energia (FFT) e o mapa de constelação com fingerprint.

Figura 1 – Espectrograma com picos de energia (FFT) e mapa de constelação com fingerprint



Fonte: elaborado pelo autor (2025).

A geração de hashes: a impressão digital acústica

O passo mais inovador do algoritmo de Wang é a criação de pares de picos a partir do mapa de constelação. Para cada ponto âncora no espectrograma, o algoritmo forma pares com outros picos dentro de uma janela temporal denominada "zona de destino" (target zone). Cada par gera um *hash* acústico da forma:

$$Hash = (f_1, f_2, \Delta t)$$

Onde f_1 é a frequência do ponto âncora, f_2 é a frequência do ponto alvo, e $\Delta t = t_2 - t_1$ é a diferença de tempo entre eles. Esse *hash* é compacto (tipicamente 32 bits) e independente do volume e da tonalidade absoluta, tornando o reconhecimento robusto a variações de gravação.

O *hash* é então armazenado em um banco de dados associado à identidade da música e ao tempo de ocorrência t_1 . No momento do reconhecimento, o Shazam gera os *hashes* da amostra capturada e os consulta no banco de dados, verificando quais músicas possuem combinações de *hashes* consistentes tanto nos valores quanto nos deslocamentos temporais.

Procedimentos metodológicos

Este trabalho adotou como metodologia a revisão bibliográfica sistemática, com foco no artigo original de Wang (2003) publicado na conferência ISMIR, além de fontes técnicas sobre processamento digital de sinais e análise de Fourier. Foram consultados também materiais didáticos sobre FFT e algoritmos de busca por similaridade para aprofundamento dos conceitos matemáticos envolvidos.

Para ilustrar o funcionamento do algoritmo, foram elaboradas representações gráficas do espectrograma e do mapa de constelação utilizando ferramentas de visualização de dados, permitindo visualizar as etapas de extração de picos e formação de pares de *hashes* descritos por Wang (2003).

Resultados e discussões

A partir do estudo do algoritmo proposto por Avery Wang (2003), observa-se que a principal eficiência do sistema de reconhecimento de áudio está na combinação entre processamento matemático robusto e estratégias de redução de dados. A utilização da Transformada Rápida de Fourier (FFT) permite converter o sinal de áudio do domínio do

tempo para o domínio da frequência de forma eficiente, viabilizando o processamento em tempo real, mesmo em dispositivos com limitações de recursos computacionais.

A construção do espectrograma, seguida da identificação dos picos de energia, mostrou-se uma etapa fundamental para a redução da complexidade do sinal. Em vez de analisar todo o conteúdo sonoro, o algoritmo foca apenas nos pontos mais relevantes, o que torna o processo mais rápido e menos sensível a ruídos e interferências. O chamado “mapa de constelação” se destaca por sua capacidade de representar o áudio de forma esparsa, mas ainda assim suficientemente informativa para identificação precisa.

Outro resultado relevante está na geração dos *hashes* acústicos. Ao utilizar pares de frequências e diferenças de tempo ($f_1, f_2, \Delta t$), o algoritmo cria uma “impressão digital” altamente discriminativa e compacta. Essa abordagem reduz significativamente o volume de dados armazenados e permite buscas rápidas em grandes bases de dados. Além disso, a independência em relação ao volume e a certa tolerância a distorções tornam o método robusto em ambientes reais, como locais com ruído ou gravações de baixa qualidade.

A análise também evidencia que o mecanismo de correspondência baseado em alinhamento temporal dos *hashes* é crucial para garantir precisão na identificação das músicas. Mesmo quando apenas pequenos trechos do áudio estão disponíveis, o sistema consegue identificar padrões consistentes, o que demonstra alta eficiência e escalabilidade do método.

De modo geral, os resultados indicam que o algoritmo de Wang (2003) equilibra eficiência computacional, robustez e precisão, sendo altamente adequado para aplicações de reconhecimento de áudio em larga escala. Sua estrutura baseada em espectrograma e *hashing* constitui uma solução elegante para um problema complexo, explicando o sucesso de sistemas como o Shazam.

Considerações finais

Este trabalho demonstrou que o Shazam é sustentado por uma base matemática sólida, combinando a Transformada Rápida de Fourier, a análise de picos espectrais e a indexação por *hashes* acústicos. O objetivo de explicar de forma acessível o mecanismo

matemático por trás do aplicativo foi alcançado, mostrando como conceitos fundamentais de análise de sinais se traduzem em uma tecnologia de uso cotidiano.

O estudo reforça a importância do ensino de matemática aplicada e processamento de sinais, disciplinas diretamente relacionadas a inovações tecnológicas de grande impacto social. A compreensão desses fundamentos capacita estudantes e profissionais a desenvolver e aprimorar sistemas similares de reconhecimento de padrões em áudio, imagem e outros domínios.

Como perspectivas futuras, a implementação de uma versão simplificada do algoritmo de fingerprinting pode ser explorada como atividade prática em disciplinas de matemática aplicada e computação, materializando a indissociabilidade entre ensino, pesquisa e extensão tecnológica.

Referências

APPLE INC. Shazam celebrates 2 billion downloads. Cupertino, 2023. Disponível em: <https://www.apple.com/newsroom>. Acesso em: 10 abr. 2025.

COOLEY, J. W.; TUKEY, J. W. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, v. 19, n. 90, p. 297-301, 1965.

OPPENHEIM, A. V.; SCHAFER, R. W. *Discrete-Time Signal Processing*. 3. ed. Upper Saddle River: Prentice Hall, 2010.

WANG, A. Li-Chun. An industrial strength audio search algorithm. In: *INTERNATIONAL SYMPOSIUM ON MUSIC INFORMATION RETRIEVAL (ISMIR)*, 4., 2003, Baltimore. *Proceedings...* Baltimore: ISMIR, 2003. p. 7-13.