

FIA 2020/22

XII CONGRESSO/CONGRESO IBEROAMERICANO DE ACÚSTICA

XXIX ENCONTRO DA SOCIEDADE BRASILEIRA DE ACÚSTICA - SOBRAC

Florianópolis, SC, Brasil

Recuperação de informação musical como ferramenta para análise sonora: uma inicialização

Boratto, T. H. A.¹; Costa, R. A. V.²; Oliveira, J. P. M.³; Schiavoni, F. L.⁴

¹ Programa de Pós-Graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora, Juiz de Fora, MG, Brasil, tales.boratto@engenharia.ufjf.br

² Ciência da Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil, romulo_vieira96@yahoo.com.br

^{3,4} Ciência da Computação, Universidade Federal de São João Del-rei, São João Del-rei, MG, Brasil, joaopedromoliveira1996@gmail.com, fls@ufsj.edu.br

Resumo

A representação do som no domínio digital condensa toda a informação necessária para caracterizá-lo, desde os aspectos temporais até aos relacionados com a atuação de um instrumentista. Contudo, muita desta informação não está disponível explicitamente e este é um fator que dificulta determinadas investigações na área. Por esta razão, os estudos e trabalhos que se concentram na recuperação de informação de áudio tornaram-se cada vez mais relevantes, providenciando melhores parâmetros de análise para categoria de sinais. Assim, este artigo tem como objetivo realizar uma inicialização ao tema, contextualizando a área de Recuperação de Informação Musical (MIR), apresentando os principais conceitos, ferramentas já implementadas, bases de dados disponíveis e, finalmente, as tendências, desafios e possibilidades futuras para a área.

Palavras-chave: MIR, Descritores de Áudio, Acústica.

PACS: 43.10.Ln, 43.75.Xz, 43.60.-c, 43.75.-z, 43.75.Yy.

Music information retrieval as a tool for sound analysis: an initialization

Abstract

The sound's representation in the digital domain condenses all the information needed to characterize it, from temporal aspects to those related to the performance of an instrumentalist. However, much of this information is not available explicitly, and this is a factor that hinders certain research in the area. For this reason, studies and works that focus on audio information retrieval have become increasingly relevant, providing better analysis parameters for this category of signals. Thus, this paper aims to perform an initialization to the subject, contextualizing the area of Music Information Retrieval (MIR), presenting the main concepts, tools already implemented, databases available, and, finally, the tendencies, challenges, and future possibilities for the area.

Keywords: MIR, Audio Descriptors, Acoustics.

1. INTRODUÇÃO

Para iniciar os estudos em acústica é imprescindível que se tenha bem estabelecido o conceito de *som*, bem como o de *sinais de áudio*. O *som* pode ser entendido como sendo resultado de uma perturbação do meio, em termos de oscilação da pressão entre regiões de compressão e rarefação, causada por uma fonte sonora. Assim, o som nada mais é do que uma onda mecânica longitudinal¹ [1]. Quando esta oscilação da pressão é captada por um microfone, surge o conceito *sinais de áudio* [1]. O áudio é a captação do som por um microfone, que atua como um transdutor que converte a energia mecânica do som em um sinal elétrico. Sendo assim, tal conceito pode ser entendido como uma forma de representar sons eletricamente. Para facilitar o entendimento, a Figura 1 ilustra este processo.

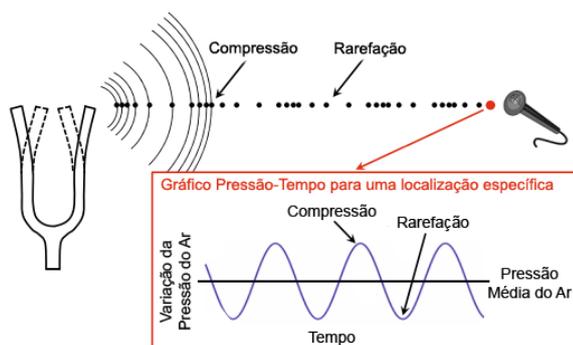


Figura 1: Ilustração da transformação do som em sinal de áudio. Fonte: [1].

Ainda que este processo de representação do som pareça simples, os sinais de áudio conseguem codificar uma enorme quantidade de informações provenientes da fonte sonora. No entanto, recuperar tais informações nem sempre é um processo trivial. A recuperação destas informações pode ser mais simples quando adotamos outra representação do áudio, que é a representação digital. Há algumas formas de representar o áudio digitalmente sendo a mais comum delas a codificação PCM (*Pulse Code Modulation*), que gera amostras digitais a partir do sinal elétrico com uma determinada taxa de amostragem. Tal representação nos permite utilizar algoritmos específicos para retirar informações do áudio, sejam elas musicais ou não.

¹Ondas longitudinais são aquelas que possuem a mesma direção de vibração e propagação.

Quando se fala em Informações Musicais, sete aspectos, de acordo com Downie em [2], podem ser considerados. São eles:

1. **Tonal:** Relacionado com a frequência fundamental percebida.
2. **Temporal:** Compreende os elementos rítmicos de um áudio.
3. **Harmônico:** Compreende áudios polifônicos, quando dois ou mais tons são reproduzidos simultaneamente. Este aspecto musical é uma interseção entre a faceta tonal e a temporal.
4. **Timbral:** Está relacionado com a capacidade de um mesmo resultado sonoro poder ser obtido por diferentes formas de onda.
5. **Editorial:** É Normalmente associada às instruções de atuação de um músico, tais como dinâmica, dedilhados, ornamentação, etc.
6. **Textural:** Está associado às letras de músicas, hinos, corais, sinfonias, etc.
7. **Bibliográfico:** Compreende as informações bibliográficas, tais como o título da obra, compositor, editor, letra, data de publicação e artista. É o único dos aspectos em que as informações não estão no áudio.

Estas questões presentes no tratamento de áudios combinados com o uso da tecnologia proporcionaram o surgimento de diversas linhas de pesquisas, como a composição algorítmica, que consiste em um conjunto de técnicas que auxiliam na composição musical sem intervenção humana, resultando em diversos avanços na criação de notação musical (partitura ou MIDI) e nas ferramentas independentes de síntese de som; novas interfaces para expressão musical, reconhecidas como uma nova classe de instrumentos musicais, mas com o diferencial de sua fonte de geração sonora ser um meio digital; Internet das Coisas Musicais (IoMusT), um campo multidisciplinar que surge a partir da extensão dos conceitos de Internet das Coisas (IoT) para a prática musical, e Recuperação de Informação Musical ou de Áudio (*Musical Information Retrieval - MIR* ou *Audio Information Retrieval - AIR*).

Sendo assim, este artigo propõe uma inicialização ao MIR, contextualizando a área, apresentando os principais conceitos, ferramentas já implementadas, bases de dados disponíveis e, finalmente, as tendências, desafios e possibilidades futuras.

Para isso, este trabalho está organizado da seguinte forma. A presente seção tratou de introduzir o conceito de Recuperação de Informação Musical e mostrar a sua importância. A Seção 2 aprofunda nos conhecimentos da área e apresenta algumas ferramentas e base de dados disponíveis para serem utilizadas. Na Seção 3, são abordados alguns desafios e possibilidades futuras de trabalho na área e, por fim, a Seção 4 conclui este trabalho.

2. EXTRAINDO INFORMAÇÕES MUSICAIS

Tradicionalmente, os dados computacionais são tratados como *string*, o que facilita a manipulação e análise dos mesmos. Entretanto, os arquivos de áudio, em especial os musicais, são conjuntos multimodais, contendo representação simbólica (partituras e cifras), texto (letras das músicas), imagens (fotografia de um músico ou capa de álbum), gesto (intérprete), metadados (títulos, artistas, estilos e análises), e claro, os elementos musicais, tais como timbre, tom e andamento rítmico. Um campo de estudo que pode ser útil na extração e representação adequada dessas características é o da Recuperação de Informação Musical, do inglês *Music Information Retrieval* (MIR), que desenvolve um esforço multidisciplinar focado na evolução de mecanismos de reconhecimento e distribuição musical em rede, de modo a tornar o estoque de música no mundo mais acessível e compreensível [3, 4].

Apesar de o termo ter sido cunhado na década de 1960, foi somente no final do século XX que esta área começou a se popularizar. Isso se deve muito ao surgimento de técnicas de compressão de áudio desenvolvidas naquele período, possibilitadas pelo aumento do poder de processamento dos computadores pessoais, que otimizou o tempo de pesquisa e ampliou disponibilidade de dispositivos capazes de tocarem música, como os reprodutores de CD (*Compact Disc*) [3, 4].

Atualmente, com a popularização dos serviços de *streaming* e com a massiva disponibilidade de dados sonoros, este campo de pesquisa vêm sendo bastante explorado, tornando-se cada vez mais relevante.

Dentre suas diversas aplicações desta área, encontram-se atividades como: detecção de música *cover*, reconhecimento de gênero musical, transcrição musical, sistemas de recomendação, verificação de semelhança melódica, detecção do humor, separação dos dados sonoros, classificação de instrumentos, detecção de tom e tempo, análise da estrutura/forma da música, geração musical automática, dentre outros [4].

A extração de informações de um sinal passa pelo uso de descritores de áudio e/ou algoritmos de processamento, com a finalidade de representar tais dados de forma compreensível tanto para as máquinas quanto para os humanos, como será melhor abordado na subseção 2.1 que se segue.

2.1 Descritores de Áudio

Os descritores de áudio são ferramentas analíticas que representam características do sinal musical em curvas unidimensionais. Assim, reduzem a complexidade da informação ao focar em aspectos específicos. São ferramentas úteis para criar uma taxonomia de particularidades do conteúdo dos espectros do sinal musical [5]. Em suma, eles apresentam caráter reducionista, pois costumam extrair apenas uma informação ou atributo do sinal de áudio. Esse atributo pode ser correlacionado com propriedades subjetivas da percepção, como “opacidade” ou “maciez” do som [6, 7].

Os descritores de áudio podem ser classificados quanto aos seus níveis de complexidade como descritores de baixo nível, nível intermediário e alto nível. De acordo com Knees e Markus em [8] os descritores de alto nível são aqueles que mais se aproximam aos conceitos entendidos pelos humanos e são geralmente formados por uma combinação de descritores dos níveis inferiores. Por outro lado, descritores de baixo nível são aqueles cuja representação é interpretável estritamente por uma máquina.

Atualmente, existem diversas pesquisas em rela-



ção ao desenvolvimento desses descritores, culminando em quatro modelos amplamente utilizados. O primeiro deles é o *Spectral Rolloff*, que extrai o ponto de rolagem, ou seja, os segmentos da onda sonora que estão abaixo de determinada porcentagem. O segundo trata-se do *Spectral Centroid*, responsável por indicar onde está o centro de massa do espectro, isto é, a tendência central da forma de onda. Ele tem uma conexão com a impressão de suas aplicações mais comuns voltadas justamente para a classificação de timbres. Já o terceiro descritor é o chamado *Spectral Flatness*, também conhecido como coeficiente de tonalidade ou Entropia de Wiener. É uma medida usada para quantificar a aparência de um ruído, em oposição ao tom.

Por fim, o último e mais importante descritor é o *Mel Frequency Cepstral Coefficient* (MFCC). Este é um derivado da representação de um espectro de áudio não-linear e pode ser entendido como uma forma de mapear as frequências conforme a capacidade auditiva humana (linear até 1 kHz e logarítmica acima disso) [9]. Apesar de o MFCC não apresentar robustez diante de ruído, necessitando, portanto, de que seus valores sejam normalizados [10], este descritor é muito utilizado em sistemas de reconhecimento de fala, classificação de gêneros musicais e medidas de similaridade de áudio [11].

A Figura 2 ilustra a sequência de etapas utilizadas para a obtenção dos MFCCs. De forma resumida, cinco procedimentos são necessários: Janelamento do sinal; aplicação da Transformada Rápida de Fourier (*Fast Fourier Transform - FFT*); Mapeamento das magnitudes do espectro de acordo com um banco de filtros triangulares; Retificação não linear do mapeamento anterior e, por fim, aplicação da Transformada Discreta do Cosseno (*Discrete Cossine Transform - DCT*)².

Há atualmente uma extensa quantidade de descritores de áudio que podem ser utilizados para recuperar as mais diversas informações de sinais de áudio. Muitos destes descritores estão implementados e disponíveis em bibliotecas e *frameworks*, conforme será visto na Subseção

2.2 seguinte.

2.2 Implementações e Recursos Disponíveis

Diversas bibliotecas e *frameworks* implementam diferentes descritores de áudio para simplificar a utilização destes na análise de sinais, como:

- **Librosa** [14]: a biblioteca *librosa* é um pacote disponível em Python, desenvolvido para áudio e processamento digital na música. A ferramenta tem o objetivo de facilitar o trabalho em MIR e apresenta familiaridade com MATLAB, além de padronização de comandos e característica modular, permitindo que o programador aumente a empregabilidade da biblioteca. Algumas funções merecem destaque, como a ***librosa.beat***, que permite o uso de funções para estimar o andamento rítmico e detectar batidas, ***librosa.core***, que inclui recursos que facilitam o carregamento de arquivos de áudio, representações de espectrograma ou outra informação musical, e ***librosa.display***, responsável por visualizações gráficas integradas [7].
- **Madmom** [15]: biblioteca de código-aberto também desenvolvida em Python, apresenta *design* conciso, é orientada a objetos, apresenta rápida prototipagem, facilmente conversível para programas executáveis autossuficientes e possui integração nativa com algoritmos de aprendizado de máquina (como cadeias de Markov e redes neurais), facilitando a criação e execução de aplicativos MIR. O *madmom* enfatiza conceitos musicalmente relevantes, portanto, atua principalmente na detecção de tempo e tom de uma canção, análise de progressões harmônicas, métrica musical e transcrição de notas.
- **Aubio** [16]: ferramenta de alto nível que segmenta arquivos de áudio para extrair características como o início de um ataque a uma determinada nota, detecção de timbre, tempo e melodia. Além disso, é empregada na criação de filtros (espectrais e lineares), de *softwares* de acompanhamento, etiquetagem de amostras sonoras, e extração de áu-

²Por não ser o foco deste trabalho, tais técnicas não serão abordadas, podendo ser encontradas em livros de processamento de sinais, tal como [12].

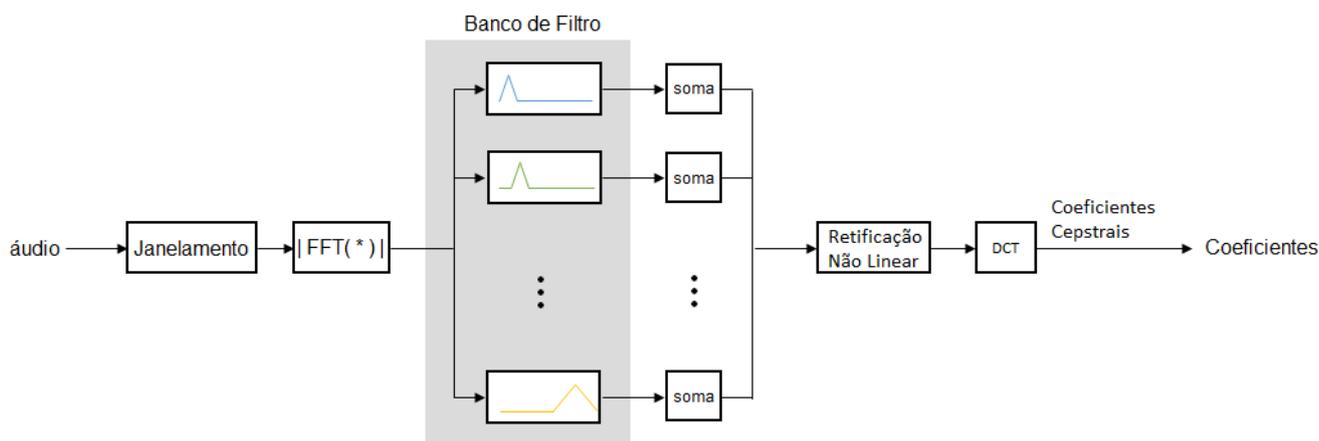


Figura 2: Sequência de etapas utilizadas para a obtenção dos MFCCs. Fonte: Adaptado de [13]

dio em qualquer arquivo de mídia, incluindo vídeos e transmissões remotas. Seu uso é focado em aplicações de tempo real, mas pode ser empregada em editores de som, *plugins* de efeitos e instrumentos virtuais.

- Essentia [17, 18]: biblioteca desenvolvida em C++ e de código-aberto, contém uma ampla coleção de algoritmos que implementam funcionalidade de entrada e saída de áudio, caracterização estatística de dados e aplicação de um grande conjunto de descritores espectrais, temporais e tonais, o que facilita e agiliza sua empregabilidade. A ferramenta foi projetada com foco na robustez, desempenho de tempo e memória, e facilidade de extensibilidade, adequada para aplicações acadêmicas e industriais, como classificação do humor em faixas de áudio, etiquetagem automática de amostras, criação de sistemas de recomendação musical, visualização e interação com a música, indexação do som, detecção de instrumentos musicais, análise acústica, detecção de *covers*, indexação em bancos de dados de grandes escalas, dentre outros.
- Marsyas [19]: *framework* para prototipagem rápida de aplicativos de áudio, desenvolvido priorizando a flexibilidade e extensibilidade de seus serviços. Inicialmente concebido usando uma arquitetura cliente-servidor, onde o cliente foi escrito em Java e continha apenas o código da interface do usuário e se comunicava com o mecanismo de computação através de soquetes,

enquanto o servidor foi desenvolvido em C++ e continha todos os módulos de processamentos de sinais e reconhecimentos de padrões para otimizar o desempenho. Difere-se das demais soluções por apresentar ampla integração, suportando diversos algoritmos e *softwares* para construção de sistemas MIR completos, e pelo tempo de execução baixo, comparado as demais ferramentas. Isso é possível graças à sua arquitetura de fluxo de dados que minimiza a necessidade de alocação de memória, código rápido e otimizado para todas as operações e capacidade de processar grandes coleções de arquivos em memória fixada.

- MIR Toolbox [20]: conjunto integrado de funções escritas em MATLAB, dedicado à extração de características musicais relacionadas ao timbre, tonalidade e ritmo de arquivos musicais. O projeto é baseado em uma estrutura modular, onde os diferentes algoritmos são decompostos em estágios, formalizados usando um conjunto mínimo de mecanismos elementares e integrando diferentes variantes propostas para a área de recuperação da informação musical. Inclui ainda funções para análise estatística, pode se adaptar a diversos formatos de dados como entrada e possui preocupações com seu uso no ensino, de modo a se tornar uma ferramenta fácil e intuitiva.
- YAAFE [21]: outro *framework* para extração de recursos de áudio é o YAAFE, onde sua principal vantagem reside no fato de ser



significativamente menos complexo de ser utilizado do que seus pares, fácil de configurar e podendo ser parametrizado de forma independente. Foi projetado para ter eficiência computacional a partir de uma exploração apropriada de redundâncias dos recursos; simplicidade na sintaxe, que contribui para sua simplicidade de uso, capacidade de processar arquivos muito grande e eficiência de armazenamento. Ainda pode ser extensível, onde cada usuário pode inserir bibliotecas próprias de recursos e componentes, que serão aplicadas ao programa no momento de sua execução.

- Meyda [22]: até aqui, foram apresentadas diversas bibliotecas e estruturas para extração de informação em arquivos de áudio, mas todas elas são focadas em aplicações para *desktops*. Nesse contexto, surge a biblioteca Meyda, primeira metodologia de extração de recursos que funciona em um cliente Web, com a vantagem de ser mais flexível, personalizável, ser de baixo custo e independente da largura de banda disponível. Ela funciona com a *Application Programming Interface* (API) de áudio da Web e é útil para diversas aplicações multimídia, como visualização de ondas sonoras, auxílio na composição musical para jogos eletrônicos e participação em processos criativos.
- Sonic Visualiser [23]: diferentemente das abordagens aqui apresentadas, o *Sonic Visualiser* não é uma biblioteca, mas uma ferramenta focada no usuário final. Portanto, tem uma interface de usuário simples, assemelhando-se a aplicativos conhecidos na edição de áudio e suporte para os formatos de *plugins* mais utilizados no mercado e na academia. Quanto à parte técnica, a ferramenta apresenta recursos de visualização de alta qualidade, conseguindo exibir espectrogramas e parâmetros úteis na análise musical. Também inclui métodos para registrar o tempo das batidas a partir do teclado ou do mouse, permite que diferentes arquivos sejam abertos de forma simultânea, é capaz de transcrever notas e consegue extrair informações em lotes.
- CUIDADO (*Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online*) [24]: O Projeto CUIDADO visa desenvolver uma nova cadeia de aplicações para descritores de áudio e música baseada no padrão MPEG-7. O projeto inclui o *design* de estruturas e o desenvolvimento de ferramentas para extração de informação de alto nível de sinais de áudio. Ainda, há a implementação de duas aplicações: paleta de áudio (*sound palette*) e navegador de música (*music browser*). Essas aplicações incluem recursos para proporcionar aos usuários acesso a conteúdos de grandes catálogos de material de áudio e música. A paleta de sons concentra-se em amostras de áudio e é voltada para usuários profissionais, ao passo que o navegador de música tem um público-alvo mais amplo e fornece títulos de música populares. Neste *framework*, os descritores combinam os bancos de dados com recursos numéricos, gerando metadados que permitem uma busca rápida e precisa em grandes volumes de dados e também permitindo a manipulação de conteúdo sonoro através de alto nível de especificação, combinando a capacidade cognitiva humana e estruturas presentes na percepção auditiva, sendo um contraponto em oposição a estruturas de dados tradicionais que são de baixo nível e orientadas à implementação.

Existem também diversos bancos de dados que ajudam os pesquisadores a criarem descritores, bem como verificar eficácia dos mesmos. A seguir, são listadas e brevemente explicadas algumas coleções.

- MSD (*Million Song Dataset*) [25]: este conjunto de dados é uma coleção de recursos de áudio e metadados composto por 1 milhão de músicas populares e contemporâneas, como sugere o nome. O banco de dados tem 280 GB de tamanho e mais de 44 mil artistas diferentes. Seu objetivo central é lidar com a questão de licenciamento de músicas populares para estudos em recuperação da informação musical. Para sanar tal problema, o MSD contém dados legalmente

disponíveis no *The Echo Nest*. Dessa forma, há um maior incentivo na pesquisa sobre algoritmos que se adaptam a tamanhos comerciais e proporciona novos trabalhos na área de MIR.

- MusiCLEF [26]: banco de dados criado em 2011, junto da *Cross-Language Evaluation Forum* (CLEF), com o objetivo principal de promover o desenvolvimento de novas metodologias de acesso e recuperação da música a partir da combinação de técnicas de extração baseadas em conteúdo com informações contextuais fornecidas pelos usuários via *tags*, comentários ou análises. Além disso, o MusiCLEF tem uma conexão estreita com aplicações reais, de modo a promover o desenvolvimento de técnicas que podem ser aplicadas para resolver problemas comuns aos profissionais da área.
- MTD (*Musical Theme Dataset*) [27]: antes de abordar o banco de dados propriamente dito, é importante compreender o que é um tema musical. Assim, é classificado qualquer elemento, motivo ou pequena peça musical que originou alguma variação. A partir dessa definição, o MTD dispõe de uma representação digital dos 2067 temas, presentes no livro “*A Dictionary of Musical Themes*”. Além das partituras, ele dispõe de codificações simbólicas da música, trechos e gravações musicais, metadados sobre o compositor, obra, gravação e características dos temas. Este conjunto é relevante para o MIR ao permitir recuperação musical cruzada, alinhamento e reconhecimento óptico de música, transcrição e musicologia computacional.
- MedleyDB [28]: em vista das poucas coleções de dados isentas de direitos autorais que podem ser compartilhadas para fins de pesquisa, em especial aquelas com gêneros musicais heterogêneos, surgiu o MedleyDB, um conjunto multifuncional que conta com 122 canções, com aplicabilidade que se estende para diversas atividades da recuperação musical, como reconhecimento de instrumentos, separação de fontes sonoras

e mixagem automática. Foi desenvolvido principalmente para auxiliar pesquisadores.

2.3 MIR e Aprendizado de Máquina

Dada a quantidade de dados de áudio oriundos de sistemas computacionais, faz-se necessária a aplicação de técnicas capazes de reconhecer padrões e atribuir aos algoritmos a capacidade de aprender e classificar esses dados da maneira mais eficiente possível. Nesse contexto, existe um ponto de convergência entre recuperação da informação e aprendizado de máquina. No âmbito musical, os problemas de classificação mais recorrentes aparecem no reconhecimento de gênero musical, na detecção de emoções em músicas e em sistemas de recomendação [29].

A categorização do gênero musical é um importante componente na área de MIR e consiste em classificar diversas peças musicais em uma única classe a partir de uma análise computacional baseada nas representações de seus recursos. As técnicas mais comuns de aprendizado de máquina utilizadas para resolver esta categoria de problema são Modelo de Mistura Gaussiana (*Gaussian Mixture Model - GMM* [30]), *k*-vizinhos mais próximos (*k-Nearest Neighbors - k-NN* [30]) e máquinas de vetores de suporte (*Support Vector Machines - SVM* [30]).

Já a detecção de emoções na música e as relações entre os sons e seus impactos nos ouvintes têm sido estudadas há décadas. Inicialmente, estes experimentos consistiam em registrar os adjetivos informados pelos elementos presentes no experimento ao ouvirem determinadas músicas. Hoje em dia, essa investigação utiliza principalmente classificações multirrótulos.

Por fim, existem os sistemas de recomendação, que consistem em entregar a um ouvinte uma lista de peças musicais que provavelmente será de seu gosto. Esta tarefa tem recebido uma crescente quantidade de atenção recentemente e seu objetivo é satisfazer dois requisitos: alta precisão de recomendação, onde em uma lista curta de músicas, boa parte das faixas presentes são recomendadas, e novidade nas recomendações, onde uma rica variedade de músicas e artistas se faz presente. Para criar estes agrupamentos, são usadas abordagens colaborativas (principal-



mente o *tagging*, concebidos a partir da técnica *Word2Vec*), gráfico Laplaciano e métodos semi-supervisionados.

3. DISCUSSÕES, DESAFIOS E POSSIBILIDADES

Diferente de problemas de recuperação de informação textual em que se usa uma abordagem mais objetiva para a resolução dos problemas, isto é, utiliza-se o próprio texto para as avaliações, a recuperação de informação em áudio necessita de um tratamento um pouco mais profundo, dado que as informações de um áudio não estão expressas de maneira direta quando o áudio é representado digitalmente em seu formato PCM.

Esse tratamento mais profundo é, na verdade, a aplicação das técnicas de MIR que buscam identificar certas propriedades dos áudios, transformando-os em estruturas com conteúdos mais significativos. Portanto, essa transformação permite recuperar diversas informações presentes nos sinais, sendo importante para tratar algumas categorias de problemas tanto no contexto musical, como vem sendo abordado neste trabalho, quanto em um âmbito geral de outras áreas de conhecimento. O trabalho de Pauline Mouawad, Tammuz Dubnov e Shlomo Dubnov [31] é um exemplo de uma importante aplicação fora do âmbito musical, visto que utilizam técnicas de MIR para a detecção de COVID-19 em pacientes através da análise do sinal de áudio com a gravação de uma tosse.

Embora seja uma área de pesquisa relativamente nova, existem organizações, concursos e desafios comprometidos com o avanço do conhecimento na área, como é caso da *International Society for Music Information Retrieval (ISMIR)*³ e do *Music Information Retrieval Evaluation eXchange (MIREX)*⁴.

Além dos avanços propostos pela comunidade, há também possibilidades de pesquisa nos novos desafios da área. A investigação da influência dos parâmetros de fabricação no som de um instrumento musical ou até mesmo dos parâmetros

de um compressor no processamento do som ainda são estudos em desenvolvimento e demandam muitos esforços, inclusive para a criação de base de dados. Sendo assim, estes esforços aliados às bases de dados já existentes, como mencionado na Subseção 2.2, possibilitam o surgimento de novos descritores, bem como a validação dos já existentes em novas abordagens.

Por fim, conhecer as características e as informações fornecidas por uma fonte sonora também são úteis para outras linhas de pesquisa, como no caso da modelagem física de instrumentos musicais.

4. CONCLUSÃO

O computador vem sendo entendido como uma ferramenta fundamental capaz de ampliar a capacidade humana e de nos trazer dados determinísticos para informações que nem sempre nossos sentidos são capazes de absorver. No caso do som, as ferramentas computacionais para a análise sonora nos permite superar os limites de nossa audição e de extrair informações que apenas profissionais com muitos anos de experiência conseguiriam ouvir. Tais ferramentas, aqui apresentadas brevemente dentro do domínio da Recuperação de Informação Musical, podem permitir que a análise sonora seja feita a partir de metodologias bem definidas onde a extração de características do som e a medição destas características podem guiar pesquisadores de diversas áreas em seus projetos.

A emergente área de recuperação da informação musical trata da recuperação e organização de grandes coleções ou dados musicais, recebendo crescente interesse por conta de suas variadas aplicações, permeando desde serviços de streaming, reconhecimento de gêneros e instrumentos musicais, sistemas de recomendação e detecção de plágio e até mesmo na área de saúde, onde auxilia no diagnóstico de COVID-19 a partir de uma análise da tosse dos pacientes e também na área da agricultura, onde pragas e insetos nocivos são detectados a partir dos sons gerados por eles próprios.

Apesar de sua grande aplicabilidade, esta área ainda é pouco explorada fora dos círculos que estudam e pesquisam computação musical. Sendo

³<https://ismir.net/>

⁴https://www.music-ir.org/mirex/wiki/MIREX_HOME

assim, o presente artigo faz um compilado de técnicas, descritores de áudio e bancos de dados que são comuns a este campo de estudo, com o objetivo de sintetizar em um único lugar informações que podem auxiliar futuros desenvolvedores a criarem seus próprios projetos.

Ainda, é apresentado um panorama da área, abordando desde seus primórdios até informações sobre o conceito de som e quais informações podem ser abstraídas a partir deles. Os autores acreditam que essa pesquisa pode ser um ponto de partida para o fortalecimento e até mesmo crescimento dessa área, em especial entre os pesquisadores lusófonos.

5. AGRADECIMENTOS

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001 - pelo apoio financeiro para o desenvolvimento do presente trabalho, e as colaborações do Grupo ALICE (*Arts, Lab in Interfaces, Computers, and Else*), Grupo de Modelagem Computacional Aplicada (GMCA) e Grupo de Pesquisa em Ciência de Dados Aplicada à Engenharia (CIDENG).

REFERÊNCIAS

- [1] Müller, Meinard. *Fundamentals of Music Processing*. Springer International Publishing, 2015. doi:10.1007/978-3-319-21945-5. Disponível em: <https://doi.org/10.1007/978-3-319-21945-5>.
- [2] Downie, J. Stephen. Music information retrieval (Chapter 7). Em *Annual Review of Information Science and Technology 37*, págs. 295–340. Information Today, Medford, NJ, 2003. Disponível em: http://music-ir.org/downie_mir_arist37.pdf.
- [3] Burgoyne, John; Fujinaga, Ichiro e Downie, J. *Music Information Retrieval*, págs. 213–228. 11 2015. ISBN 9781118680599. doi:10.1002/9781118680605.ch15.
- [4] Schedl, Markus; Gómez, Emilia e Urbano, Julián. Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, 8:127–261, 01 2014. doi:10.1561/15000000042.
- [5] Simurra, Ivan Eiji. A utilização descritores de áudio à análise e composição musical assistidas por computador: um estudo de caso na obra labori ruinae, 2016.
- [6] Simurra, Ivan Eiji. Análise musical assistida por descritores de Áudio: um estudo de caso da obra reflexões de jônatas manzoli, 2015.
- [7] Vieira, Rômulo; Araújo, João Teixeira; Paulo; Batista, Edimilson e Schiavoni, Flávio. Automatic classification of instruments from supervised methods of machine learning. Em *Proceedings of the 18th Brazilian Symposium on Computer Music*, págs. 28–34, Recife - PE - Brasil, Outubro 2021. Sociedade Brasileira de Computação.
- [8] Knees, Peter e Schedl, Markus. Music retrieval and recommendation: A tutorial overview. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015.
- [9] Logan, Beth. Mel frequency cepstral coefficients for music modeling. *Proc. 1st Int. Symposium Music Information Retrieval*, 11 2000.
- [10] T. Ganchev, N. Fakotakis, G. Kokkinakis. Comparative evaluation of various mfcc implementations on the speaker verification. *10th International Conference on Speech and Computer (SPECOM 2005)*, 1: 191–194, 2005.
- [11] T. Ganchev, N. Fakotakis e G. Kokkinakis. Comparative evaluation of various mfcc implementations on the speaker verification task. *10th International Conference on Speech and Computer (SPECOM 2005)*, 2005.
- [12] Alan V. Oppenheim; Ronald W. Schaffer e John R. Buck. *Discrete-Time Signal Processing*. 2nd ed^a ed. Prentice Hall, 1999. Prentice-Hall Signal Processing Series. ISBN 9780137549207; 0137549202.
- [13] MATLAB. Extract cepstral features from audio segment. Disponível em: <https://www.mathworks.com/help/audio/ref/cepstralfeatureextractor-system-object.html>. Last accessed 25 may 2022.
- [14] McFee, Brian; Raffel, Colin; Liang, Dawen; Ellis, Daniel PW; McVicar, Matt; Battenberg, Eric e Nieto, Oriol. *librosa: Audio and music signal analysis in python*. Em *Proceedings of the 14th python in science conference*, volume 8, págs. 18–25. Citeseer, 2015.
- [15] Böck, Sebastian; Korzeniowski, Filip; Schlüter, Jan; Krebs, Florian e Widmer, Gerhard. Madmom: A new python audio and music signal processing library. Em *Proceedings of the 24th ACM international conference on Multimedia*, págs. 1174–1178, 2016.
- [16] Brossier, Paul M. The audio library at mirex 2006. *Synthesis*, 2006.
- [17] Bogdanov, Dmitry; Wack, Nicolas; Gómez, Emilia; Gulati, Sankalp; Herrera, Perfecto; Mayor, Oscar; Roma, Gerard; Salamon, Justin; Zapata, José e Serra, Xavier. Essentia: an open-source library for sound and music analysis. Em *Proceedings of the 21st ACM international conference on Multimedia*, págs. 855–858, 2013.



- [18] Bogdanov, Dmitry; Wack, Nicolas; Gómez Gutiérrez, Emilia; Gulati, Sankalp; Boyer, Herrera; Mayor, Oscar; Roma Trepat, Gerard; Salamon, Justin; Zapata González, José Ricardo; Serra, Xavier *et al.* *Essentia: An audio analysis library for music information retrieval*. Em Britto A, Gouyon F, Dixon S, editors. 14th Conference of the International Society for Music Information Retrieval (ISMIR); 2013 Nov 4-8; Curitiba, Brazil.[place unknown]: ISMIR; 2013. p. 493-8. *International Society for Music Information Retrieval (ISMIR)*, 2013.
- [19] Tzanetakis, George e Cook, Perry. *Marsyas: A framework for audio analysis*. *Organised sound*, 4 (3):169-175, 2000.
- [20] Lartillot, Olivier; Toivainen, Petri e Eerola, Tuomas. *A matlab toolbox for music information retrieval*. Em Preisach, Christine; Burkhardt, Hans; Schmidt-Thieme, Lars e Decker, Reinhold, editores, *Data Analysis, Machine Learning and Applications*, págs. 261-268, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. ISBN 978-3-540-78246-9.
- [21] Mathieu, Benoit; Essid, Slim; Fillon, Thomas; Prado, Jacques e Richard, Gaël. *Yaafe, an easy to use and efficient audio feature extraction software*. Em ISMIR, págs. 441-446. Citeseer, 2010.
- [22] Rawlinson, Hugh; Segal, Nevo e Fiala, Jakub. *Meyda: an audio feature extraction library for the web audio api*. Em The 1st web audio conference (WAC). Paris, Fr, 2015.
- [23] Cannam, Chris; Landone, Christian e Sandler, Mark. *Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files*. Em Proceedings of the 18th ACM international conference on Multimedia, págs. 1467-1468, 2010.
- [24] Vinet, Hugues; Herrera, Perfecto e Pachet, Francois. *The CUIDADO Project*. Em International Conference on Music Information Retrieval, págs. 197-203, Paris, France, Outubro 2002. Disponível em: <https://hal.archives-ouvertes.fr/hal-01250799>. cote interne IRCAM: Vinet02b.
- [25] Bertin-Mahieux, Thierry; Ellis, Daniel; Whitman, Brian e Lamere, Paul. *The million song dataset*. págs. 591-596, 01 2011.
- [26] Orio, Nicola; Rizo, David; Miotto, Riccardo; Schedl, Markus; Montecchio, Nicola e Lartillot, Olivier. *Musiclef: a benchmark activity in multimodal music information retrieval*. págs. 603-608, 01 2011.
- [27] Zalkow, Frank; Balke, Stefan; Arifi-Müller, Vlora e Müller, Meinard. *Mtd: A multimodal dataset of musical themes for mir research*. *Transactions of the International Society for Music Information Retrieval*, 3:180-192, 10 2020. doi:10.5334/tismir.68.
- [28] Bittner, Rachel; Salamon, Justin; Tierney, Mike; Mauch, Matthias; Cannam, Chris e Bello, Juan. *Meddleydb: A multitrack dataset for annotation-intensive mir research*. 10 2014.
- [29] Li, Tao; Ogihara, Mitsunori; Shao, Bo e Dingding Wang. *Machine Learning Approaches for Music Information Retrieval*. 01 2009. ISBN 978-953-7619-55-4.
- [30] Geron, Aurelien. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2nd^a ed. O'reilly, 2019. 978-1-492-03264-9. ISBN 9781492032649; 1492032646. Disponível em: libgen.li/file.php?md5=40ca3f6e08498377145117d8b48bfd1b.
- [31] Mouawad, Pauline; Dubnov, Tammuz e Dubnov, Shlomo. *Robust detection of COVID-19 in cough sounds*. 2(1), Janeiro 2021. doi:10.1007/s42979-020-00422-6. Disponível em: <https://doi.org/10.1007/s42979-020-00422-6>.