

Clarice Pawlow Torres de Abreu

Music Transcription Bot: Uma aplicação web que utiliza Magenta.js para transpor música

Belo Horizonte, Minas Gerais

2023

Clarice Pawlow Torres de Abreu

Music Transcription Bot: Uma aplicação web que utiliza Magenta.js para transpor música

Relatório apresentado ao
departamento Departamento de Ciência da
Computação para a aprovação no curso de
Atividades Práticas Integradoras

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Orientador: Flávio Figueiredo, Leonardo de Oliveira Gomes

Belo Horizonte, Minas Gerais
2023

Resumo

Após estudar a detecção de onsets e a previsão de frames em música, o objetivo deste trabalho é desenvolver uma aplicação web que utilize essas técnicas para transcrever automaticamente músicas de áudio em partituras. A detecção de onsets é responsável por identificar os momentos de início de eventos musicais relevantes, como notas e acordes, enquanto a previsão de frames gera sequências de frames que representam os eventos detectados.

A aplicação web proposta permitirá que os usuários carreguem arquivos de áudio na plataforma, que serão processados para identificar os onsets e gerar os frames correspondentes. Em seguida, a aplicação irá converter esses frames em partituras musicais, fornecendo uma transcrição precisa e visualmente legível da música. Essa ferramenta terá um papel fundamental na simplificação do processo de transcrição musical, permitindo que compositores, músicos e estudantes explorem e compartilhem suas criações de forma mais eficiente e precisa.

Palavras-chaves: onsets. frames. inteligência artificial. harmonia. música.

Abstract

After studying onset detection and frame prediction in music, the goal of this project is to develop a web application that utilizes these techniques to automatically transcribe audio music into sheet music. Onset detection is responsible for identifying the moments of relevant musical events, such as notes and chords, while frame prediction generates sequences of frames that represent the detected events.

The proposed web application will allow users to upload audio files to the platform, which will be processed to identify the onsets and generate the corresponding frames. Then, the application will convert these frames into musical scores, providing an accurate and visually readable transcription of the music. This tool will play a crucial role in simplifying the process of music transcription, enabling composers, musicians, and students to explore and share their creations more efficiently and accurately.

Keywords: onsets. frames. music.

Sumário

1	INTRODUÇÃO	5
2	REFERENCIAL TEÓRICO	6
2.1	Definições musicais	6
2.1.1	Melodia	6
2.1.2	Intervalo musical	6
2.1.3	Harmonia	6
2.1.4	Midi	6
2.2	Conceitos da computação	7
2.2.1	API	7
2.2.2	TensorFlow.js	7
2.2.3	Tone.js	7
2.3	Trabalhos relacionados à transcrição musical e computação	7
3	ATIVIDADES	8
4	CONCLUSÃO	10
	REFERÊNCIAS	11

1 Introdução

Uma das expressões mais fascinantes e significativas da criatividade humana é a música. Ela transcende fronteiras culturais e históricas, cativando e emocionando pessoas ao redor do mundo. Com o avanço da inteligência artificial e a aplicação de técnicas de aprendizado de máquina, surge a possibilidade de explorar a música de uma maneira inovadora.

A detecção de onsets e a previsão de frames são áreas de pesquisa em constante desenvolvimento que permitem analisar e compreender a estrutura e os elementos fundamentais da música. Essas técnicas possibilitam a identificação dos momentos de início de eventos musicais relevantes e a previsão de sequências de frames que representam esses eventos.

Nesse contexto, o objetivo deste trabalho é criar uma aplicação web que utilize a detecção de onsets e a previsão de frames para transcrever automaticamente músicas de áudio em partituras. Essa aplicação web proporcionará aos usuários uma forma intuitiva e eficiente de converter suas gravações em música escrita, facilitando a análise, a interpretação e o compartilhamento de composições musicais.

2 Referencial Teórico

Diante do objetivo estabelecido foi necessário entender definições e conceitos essenciais sobre música e computação. Essa seção é dedicada a apresentar esses conceitos, sendo que a seção 2.1 é dedicada às definições musicais, a seção 2.2 é dedicada à conceitos de tecnologia e a seção 2.3 cita alguns trabalhos correlatos.

2.1 Definições musicais

2.1.1 Melodia

É uma sequência de notas que gera uma coerência musical.

2.1.2 Intervalo musical

É a distância entre duas notas. Para a música ocidental as unidades de medidas são tom e semitom, além do savart e cent, usados para intervalos menores que um semitom.

2.1.3 Harmonia

A harmonia é o resultado de duas ou mais notas musicais tocadas ao mesmo tempo e é um conceito muito importante presente na história da música ocidental. Como campo de estudo, a harmonia faz referência à construção de acordes e como uma ideia musical tem várias possibilidades de ser desenvolvida em torno de uma melodia central.

Com o desenvolvimento do sistema tonal, estruturado por volta do século XV, foram criadas regras para a geração de acordes. No século XIX se iniciou uma degradação gradual da música e harmonia tonal, resultando na existência da música atonal a partir do século XX. Apesar disso, a música popular e comercial ainda é predominantemente tonal e, portanto, segue as regras que formam esse sistema.

2.1.4 Midi

Musical Instrument Digital Interface é um protocolo cujo o objetivo é fazer possível a interação de instrumentos em um ambiente digital.

2.2 Conceitos da computação

2.2.1 API

Uma API é um código que auxiliar dois sistemas de softwares a se comunicarem e trocarem dados sem interferência do usuário.

2.2.2 TensorFlow.js

Uma biblioteca para aprendizado de máquina feita co javascript

2.2.3 Tone.js

Um framework para criar música interativa no browser

2.3 Trabalhos relacionados à transcrição musical e computação

A transcrição musical automatizada tem sido objeto de estudo e desenvolvimento na área de computação e música. Diversos trabalhos têm explorado técnicas e abordagens para extrair informações musicais de gravações de áudio e convertê-las em notações musicais compreensíveis. A seguir, são apresentados alguns trabalhos relevantes nesse campo.

Um trabalho notável é o sistema "AudioScore", desenvolvido pela Neuratron. Esse software utiliza técnicas avançadas de processamento de sinal para realizar a transcrição de áudio em partituras musicais. O AudioScore é capaz de reconhecer uma ampla gama de instrumentos musicais e produzir resultados precisos, tornando-se uma ferramenta popular para músicos e compositores.

Além disso, a aplicação de técnicas de aprendizado de máquina tem se mostrado promissora na transcrição musical. O trabalho de Sigtia et al. apresenta um sistema baseado em redes neurais convolucionais para a transcrição automática de acordes em músicas polifônicas. O sistema foi treinado em um extenso conjunto de dados e alcançou resultados significativos na transcrição precisa dos acordes presentes nas gravações musicais.

Outras abordagens inovadoras incluem o uso de redes neurais recorrentes para a transcrição de melodias e o desenvolvimento de algoritmos de segmentação e classificação para identificar padrões rítmicos e harmônicos em gravações de áudio.

3 Atividades

A primeira atividade foi reler o artigo que apresenta o Onsets and Frames e entender os conceitos utilizados nele. O artigo "Onsets and Frames: Dual-Objective Piano Transcription" aborda o problema da transcrição de piano a partir de gravações de áudio. Os autores propõem um modelo de aprendizado profundo que combina duas tarefas principais: detecção de onsets e previsão de frames musicais. A detecção de onsets é responsável por identificar os momentos de início das notas musicais, enquanto a previsão de frames gera sequências de frames que representam as características das notas.

A Google criou collab que proporciona ao usuário a experiência de executar a rede do onsets and frames e ver os logs de cada etapa. Esse foi o próximo passo para o trabalho.

Com isso, o próximo passo foi ler o repositório do onsets and frames disponibilizado no github e realizar os primeiros testes utilizando o framework React.js para renderizar o projeto e tratar dos eventos do usuário mais facilmente.

Diante do que foi estudado sobre o que seria possível fazer com o modelo, a próxima atividade realizada foi idealizar uma interface para a aplicação web, isso foi feito por meio da ferramenta online Figma e o link para o acesso é o seguinte: <https://www.figma.com/file/Kpl9zCh9w1fOHjTranscription-Bot?type=designnode-id=0-1mode=designnt=wQMPAVs8kLtAA9Z6-0>

Após esse desenvolvimento, foi necessário ir para a parte da prática e implementação do código. Essa atividade foi desafiadora, já que exigiu o uso de algumas bibliotecas que tratam o áudio, além do estudo do framework React.js. Foram alguns tópicos que precisaram ser estudados, antes de a transcrição ser realizada.

Ainda nessa atividade, o método utilizado para transpor a música foi o `transcribeFromAudioFile`. Esse método é um recurso fornecido pela classe `OnsetsAndFrames` no `Magenta.js`, uma biblioteca de JavaScript para música baseada em aprendizado de máquina. Esse método é usado para transcrever automaticamente um arquivo de áudio em uma sequência de eventos musicais, como notas e acordes.

Ao chamar o `transcribeFromAudioFile`, você fornece o caminho ou URL do arquivo de áudio como entrada. A biblioteca `Magenta.js` processará o áudio e retornará uma transcrição que contém informações sobre os eventos musicais detectados, incluindo as notas, durações, intensidades e outros atributos relevantes.

Além disso, o upload de dados é realizado através da gravação de áudio pelo microfone do dispositivo do usuário. A função `startRecording` é responsável por iniciar a gravação de áudio, que é realizada utilizando a API `navigator.mediaDevices.getUserMedia`. Essa API solicita permissão ao usuário para acessar o microfone e retorna uma stream de áudio.

Uma vez que a gravação é iniciada, o áudio capturado é armazenado em um objeto `MediaRecorder`, que é configurado para capturar os dados em formato de áudio webm com o codec Opus. Durante a gravação, o evento `dataavailable` é acionado, permitindo que os chunks de dados do áudio sejam coletados em um array.

Quando a gravação é encerrada com a função `stopRecording`, o áudio gravado é convertido em um blob e passado para a função `transcribeAudio`. Essa função realiza o processamento do áudio usando o modelo `OnsetsAndFrames` do `Magenta.js` para transcrever o áudio em uma sequência de eventos musicais, comportamento explicado acima.

4 Conclusão

Neste trabalho, foi desenvolvida uma aplicação web que utiliza técnicas de detecção de onsets e previsão de frames para transcrever automaticamente músicas de áudio em partituras. Através do uso da biblioteca Magenta.js e do modelo OnsetsAndFrames, foi possível realizar a transcrição precisa e eficiente das músicas.

A aplicação permite que os usuários gravem áudio através do microfone de seus dispositivos, que é processado e transcrito em uma sequência de eventos musicais. Além disso, a transcrição é convertida em formato MIDI, possibilitando o download da partitura e sua reprodução em softwares musicais.

No entanto, é importante ressaltar que ainda há espaço para melhorias e aprimoramentos nessa área. A aplicação pode ser expandida para lidar com diferentes gêneros musicais, explorar a detecção de outros elementos musicais além dos onsets e frames, e aprimorar a interface do usuário para uma experiência mais intuitiva.

No geral, este trabalho demonstra o potencial da inteligência artificial aplicada à música, permitindo a automação de tarefas complexas e proporcionando novas possibilidades criativas. A transcrição musical automatizada tem o potencial de impulsionar a colaboração e o compartilhamento de músicas, facilitando a aprendizagem, a criação e a apreciação da arte musical.

Referências

[Neely]NEELY, A. *There are no RULES in music, only CONTEXT!*

[Team]TEAM, D. G. B. *Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset*. <https://arxiv.org/abs/1810>, (Acessado em 15/03/2023).

[Team]TEAM, G. B. *Onsets and Frames: Dual-Objective Piano Transcription*. <https://arxiv.org/abs/1710.11153>, (Acessado em 15/03/2023).