

# Conversão de músicas para MIDI com redução para 4 canais no contexto do NES

Alex Eduardo  
alexeduardoaeas@ufmg.br

Felipe Zan  
felipezan@ufmg.br

Matheus Flávio  
matheusfgs@ufmg.br

**Resumo**—Este projeto explora a conversão de músicas em formatos MP3/WAV para MIDI, reduzindo a instrumentação para quatro canais, de modo a garantir compatibilidade com o hardware limitado do Nintendo Entertainment System (NES). Com o objetivo de integrar a Recuperação da Informação Musical (MIR), o estudo combina técnicas de processamento de áudio com bibliotecas especializadas em análise musical. Ao adaptar músicas modernas para o ambiente restrito do NES, o projeto enfrenta desafios como a polifonia e a simplificação rítmica, buscando soluções através de métodos como a Transformada de Fourier e a detecção de pitch. Assim sendo, idealiza-se a realização de análise espectral e a transcrição automática, essenciais para a conversão eficiente de áudio para MIDI. Este estudo visa não apenas a conversão técnica, mas também a preservação da essência musical dentro das limitações do NES.

**Index Terms**—Conversão de áudio, MIDI, NES, Transformada de Fourier, Detecção de pitch, Redução de polifonia, Quantização temporal, Música retrô.

## I. INTRODUÇÃO

A transcrição automática de música (AMT) é uma área emergente no campo da Recuperação da Informação Musical (MIR), que visa converter gravações de áudio em representações simbólicas como partituras ou arquivos MIDI.

O Nintendo Entertainment System (NES), lançado originalmente em 1983 (sob o nome Famicom no Japão), revolucionou a indústria de videogames. Seu hardware de áudio é limitado a apenas quatro canais: dois canais de onda *pulse* (geralmente usados para melodias ou harmonias), um canal de onda *triangle* (frequentemente dedicado a linhas de baixo) e um canal de *noise* (usado para efeitos percussivos ou ruídos). Devido a essas restrições, a composição e a conversão de músicas para o NES requerem soluções criativas para contornar o número reduzido de vozes disponíveis.

Este projeto foca então na **adaptação de músicas modernas para o NES**, empregando técnicas de processamento de áudio e bibliotecas de análise musical. O objetivo é explorar a conversão de arquivos MP3/WAV para MIDI, utilizando métodos avançados de processamento de áudio. A natureza particular do hardware do NES impõe restrições que exigem soluções criativas para a transcrição e simplificação musical. Ao integrar métodos de análise espectral e técnicas de transcrição automática, este projeto visa cumprir as limitações técnicas e também manter a integridade musical das composições originais.

## II. TÉCNICAS

Para realizar a conversão de áudio para MIDI com redução para quatro canais, o projeto emprega diversas técnicas descritas no livro *Fundamentals of Music Processing* de M. Müller [1]. Essas técnicas são fundamentais para a análise e manipulação de áudio, permitindo a extração precisa de características musicais e a simplificação necessária para a adaptação ao NES. A seguir, são apresentadas as técnicas principais utilizadas neste projeto, cada uma detalhada em termos de sua aplicação prática e impacto na transcrição musical.

### A. Transformada de Fourier (FFT)

A Transformada de Fourier é uma ferramenta essencial para análise espectral, permitindo a decomposição de um sinal de áudio em suas frequências componentes. No contexto deste projeto, a FFT é utilizada para identificar as frequências dominantes de uma música, facilitando a separação de elementos como melodia e baixo, que são fundamentais para a simplificação para quatro canais. Pretende-se usar essa técnica no que segue:

- **Identificação de Frequências Dominantes:** Identificar as frequências fundamentais de notas, harmônicos, e elementos percussivos em um sinal de áudio.
- **Isolamento de Elementos Musicais:** Separar elementos com características únicas, como timbre e frequência, isolando as frequências baixas (como as do baixo) de frequências médias (melodia e harmonia), para mapeamento eficiente na conversão.

### B. Detecção de Pitch

A detecção de pitch é crucial para identificar a frequência fundamental de notas musicais em uma gravação. Esta técnica ajuda na transcrição de melodias e na simplificação de harmonias complexas, garantindo que as notas mais importantes sejam preservadas no arranjo final para o NES. Nesse contexto, busca-se:

- **Transcrição de Notas:** A detecção de pitch permite associar frequências às notas musicais correspondentes, que por sua vez são convertidas em entradas MIDI.
- **Redução de Complexidade:** Em músicas polifônicas, a detecção de pitch ajuda a escolher as frequências mais importantes, a exemplo da melodia principal.

89 C. Análise Espectral

90 A análise espectral envolve a visualização e interpretação  
91 das características de frequência de uma música ao longo  
92 do tempo. Esta técnica permite identificar as mudanças  
93 dinâmicas em uma peça musical, auxiliando na decisão  
94 de como distribuir diferentes partes da música nos canais  
95 disponíveis do NES. Pretende-se usar no contexto:

- 96 • **Extração de Informações Temporais:** Usar um es-  
97 pectrograma permite identificar como as frequências  
98 mudam ao longo da música. Assim, a análise se torna  
99 mais precisa.
- 100 • **Tomada de Decisões Baseadas na Energia:**  
101 Frequências com maior intensidade são priorizadas no  
102 mapeamento para os canais do NES.

103 D. Segmentação e Detecção de Notas

104 Segmentar uma faixa de áudio em notas individuais é  
105 essencial para a criação de arquivos MIDI precisos. Este  
106 projeto utiliza algoritmos de detecção de notas para extrair  
107 eventos musicais discretos, que são então atribuídos aos  
108 canais do NES de maneira a maximizar a fidelidade musical  
109 dentro das limitações do hardware. Essa prática envolve  
110 diretamente:

- 111 • **Detecção de Onsets:** Busca realizar a identificação de  
112 pontos onde novas notas comecem.
- 113 • **Tomada de Decisões Baseadas na Energia:** Além do  
114 pitch, a duração e a intensidade das notas são essen-  
115 ciais para preservar a expressividade. Assim sendo,  
116 é útil manter essa informação a fim de expressar  
117 corretamente a transcrição do áudio da maneira mais  
118 fidedigna que as limitações do NES possibilitam.

119 E. Redução de Polifonia

120 Dado que o NES suporta apenas quatro canais de áudio,  
121 a redução de polifonia é uma etapa crítica. Técnicas de  
122 agrupamento e simplificação harmônica são empregadas  
123 para reduzir acordes complexos a intervalos ou notas  
124 únicas, assegurando que a música permaneça compreensível  
125 e agradável. Assim sendo, entende-se como necessário:

- 126 • **Clustering para Agrupamento:** Pretende organizar  
127 as notas em categorias, por exemplo, classificando-as  
128 em melodia, harmonia, baixo e percussão, fazendo uso  
129 de algoritmos como K-means.
- 130 • **Simplificação Harmônica:** Em virtude das limitações  
131 técnicas, é necessário reduzir acordes complexos a  
132 tríades ou notas únicas, preservando a base harmônica  
133 da música.
- 134 • **Filtragem por Energia:** Visa usar a amplitude da FFT  
135 para priorizar frequências dominantes.

136 F. Quantização Temporal para o NES

137 Uma vez que o *timing* (precisão rítmica) no NES é  
138 determinado por variações de CPU e divisores de clock  
139 do console, a **quantização temporal** torna-se essencial  
140 para garantir que as notas e eventos musicais estejam

141 alinhados com a capacidade do hardware. Neste projeto,  
142 são consideradas:

- 143 • **Definição de Resolução Temporal:** Ajuste do número  
144 de *ticks* ou *frames* por batida para adequar-se aos  
145 limites do NES.
- 146 • **Acomodação de Variações de BPM:** Garantir que  
147 os eventos MIDI sejam mapeados de forma estável  
148 mesmo em músicas com variações de andamento.
- 149 • **Simplificação de Passagens Complexas:** Em tre-  
150 chos com alta densidade de notas, procedimentos de  
151 quantização podem fundir ou eliminar notas para evitar  
152 sobrecarga.

153 Essas técnicas, combinadas, formam a base do processo  
154 de conversão de áudio para MIDI no contexto deste projeto,  
155 permitindo uma transcrição que respeita tanto às limitações  
156 técnicas quanto à integridade musical.

157 III. RESULTADOS ESPERADOS

158 Esse projeto visa como resultado alcançar um método  
159 automatizado para conversão de músicas para MIDI com  
160 4 canais, garantindo a preservação da identidade musical,  
161 mesmo com a redução de complexidade. Em outras pa-  
162 lavras, é visado desenvolver uma ferramenta acessível e  
163 reproduzível para criação de músicas compatíveis com o  
164 NES.

165 IV. TECNOLOGIAS E FERRAMENTAS

- 166 • **Bibliotecas:** Pydub, Librosa, Mido, NumPy.
- 167 • **Algoritmos Clássicos:** FFT (Transformada Rápida  
168 de Fourier), Algoritmos de detecção de pitch (ex.:  
169 Autocorrelação, HPS), Modelos de agrupamento para  
170 redução de canais, Métodos de quantização temporal.

171 V. CRONOGRAMA

Tabela I  
CRONOGRAMA DE DESENVOLVIMENTO DO PROJETO.

Semana	Atividade
1-2	Pesquisa e definição da metodologia.
3-4	Pipeline de conversão.
5	Validação e ajustes.
6	Escrita e submissão do paper.

172 VI. REFERÊNCIAS

173 REFERÊNCIAS

- 174 [1] M. Müller, *Fundamentals of Music Processing*, Germany: Erlangen,  
175 International Audio Laboratories Erlangen, 2015.
- 176 [2] B. McFee, C. Raffel, D. Liang, D. P.W. Ellis, M. McVicar, E.  
177 Battenberg, O. Nieto, “librosa: Audio and Music Signal Analysis in  
178 Python”, *Proc. of the 14th Python in Science Conf. (SciPy 2015)*,  
179 2015.
- 180 [3] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, A. Klapuri,  
181 “Journal of Intelligent Information Systems 41,” *Automatic Music*  
182 *Transcription: Challenges and Future Directions*, 2013.
- 183 [4] C.-W. Wu, C. Dittmar, C. Southall, R. Vogl, G. Widmer, J. Hockman,  
184 M. Müller, A. Lerch, “IEEE/ACM Transactions on Audio, Speech,  
185 and Language Processing,” *A Review of Automatic Drum Transcrip-*  
186 *tion*, 2017.

- 187 [5] K. Collins, *Game Sound: An Introduction to the History, Theory, and*  
188 *Practice of Video Game Music and Sound Design*. Cambridge, MA:  
189 The MIT Press, 2008.  
190 [6] *NESDev Wiki*, disponível em: <https://www.nesdev.org>, acesso em:  
191 10/01/2025.